



A knowledge-guided and traditional Chinese medicine informed approach for herb recommendation*

Zhe JIN[†], Yin ZHANG^{†‡}, Jiaxu MIAO, Yi YANG, Yueting ZHUANG, Yunhe PAN

College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

[†]E-mail: 11521043@zju.edu.cn; yinzh@zju.edu.cn

Received Dec. 25, 2022; Revision accepted Feb. 28, 2023; Crosschecked Sept. 19, 2023

Abstract: Traditional Chinese medicine (TCM) is an interesting research topic in China's thousands of years of history. With the recent advances in artificial intelligence technology, some researchers have started to focus on learning the TCM prescriptions in a data-driven manner. This involves appropriately recommending a set of herbs based on patients' symptoms. Most existing herb recommendation models disregard TCM domain knowledge, for example, the interactions between symptoms and herbs and the TCM-informed observations (i.e., TCM formulation of prescriptions). In this paper, we propose a knowledge-guided and TCM-informed approach for herb recommendation. The knowledge used includes path interactions and co-occurrence relationships among symptoms and herbs from a knowledge graph generated from TCM literature and prescriptions. The aforementioned knowledge is used to obtain the discriminative feature vectors of symptoms and herbs via a graph attention network. To increase the ability of herb prediction for the given symptoms, we introduce TCM-informed observations in the prediction layer. We apply our proposed model on a TCM prescription dataset, demonstrating significant improvements over state-of-the-art herb recommendation methods.

Key words: Traditional Chinese medicine; Herb recommendation; Knowledge graph; Graph attention network
<https://doi.org/10.1631/FITEE.2200662>

CLC number: TP391

1 Introduction

Traditional Chinese medicine (TCM) carries the experience and theoretical knowledge of the ancient Chinese people's struggle against diseases and has played an indispensable role in health care for Asian people for thousands of years as the foundation of the Chinese medical system. It constantly combines the basic principles and methods of ancient Chinese philosophy with long-term specific practice. In recent decades, TCM is increasingly accepted worldwide (Cheung, 2011), and with the continuous development of various medical specialities such as microbi-

ology and immunology, more studies have begun to focus on the importance of TCM as a complementary and alternative therapy (Lu et al., 2022; Mao et al., 2022), providing new ideas and references for clinical practice.

Traditionally, TCM consultation has been mainly offline. Doctors obtain the symptoms of patients by four basic diagnostic methods called *wang-wen-wen-qie* and develop a prescription for treatment. The whole process is based mainly on the doctor's experience in treating patients and clinical practice, which is highly subjective. Therefore, several studies have tried to learn such prescribing patterns and knowledge through the topic model and other machine learning algorithms (Yao et al., 2018; Zhao GS et al., 2018). For example, Yao et al. (2018) proposed a topic model to characterize the treatment process by regarding herbs and symptoms as

[‡] Corresponding author

* Project supported by the China Knowledge Centre for Engineering Sciences and Technology (CKCEST) and the National Natural Science Foundation of China (No. 62037001)

ORCID: Zhe JIN, <https://orcid.org/0000-0002-4221-4547>; Yin ZHANG, <https://orcid.org/0000-0001-6986-4227>

© Zhejiang University Press 2023

observed variables, while treating syndromes, treatment methods, pathogeny, and other elements as hidden variables. However, the topic model has several limitations in such information extraction tasks of TCM prescription: (1) Most TCM texts are short and deficient in contextual information, which leads to data sparsity. The topic model is a bag-of-words model, which ignores the order and semantic relationships among words, and is therefore not suitable for semantic analysis. (2) The topics in TCM often overlap, which might confuse the topic model.

With the rapid development of artificial intelligence and deep learning, graph convolutional networks (GCNs) are gradually being used in tasks such as social network analysis and knowledge graph mining due to their high performance and interpretability. GCNs excel in leveraging the graph structure and aggregating node information from neighborhoods; hence, they have great expressive power to learn graph representations and have achieved superior performance in a wide range of tasks and applications (Zhang S et al., 2019). Because of the excellent performance of GCNs on graphs, researchers have suggested that the mutual relationship between the herbs and symptoms in TCM prescriptions can also be represented as a graph (Jin et al., 2020; Yang et al., 2022). For example, a TCM prescription in *Dictionary of Traditional Chinese Medicine Prescriptions* (Fig. 1) shows the name, composition, usage, function, and treatment. The herbs, such as Chinese Thorowax Root and *Scutellaria baicalensis*, are in the composition part, while symptoms, such as no appetite and vexation and vomiting, are listed in the treatment part. Jin et al. (2020) constructed a symptom–symptom synergy graph (S-S graph), an herb–herb synergy graph (H-H graph), and a symptom–herb synergy graph (S-H graph) based on prescriptions according to the frequencies of co-occurrence between herbs and symptoms. This method could be further improved by adding external knowledge and graph structure (Yang et al., 2022; Zhao W et al., 2022).

Even though some of the previous studies have incorporated some external TCM knowledge, they still lack more comprehensive TCM knowledge to guide herb recommendation. The first type of knowledge concerns the TCM prescription process. In general, modern medicine understands and treats diseases mainly from the microscopic perspective of

09318 小柴胡汤 (《伤寒论》)

(09318 Minor Bupleurum Decoction (from *Treatise on Cold Damage Diseases*))

【组成】柴胡半斤 黄芩三两 人参三两 半夏半升 甘草 生姜各三两 大枣十二个

(Composition: 125 g Chinese Thorowax Root, 47 g *Scutellaria baicalensis*, 47 g Ginseng, 65 g Pinellia Tuber, 47 g Licorice Root, 47 g Fresh Ginger, 12 Chinese Dates)

【用法】以水一斗二升，煮取六升，去滓，再煎取三升，温服一升，日三服 (Usage: Add 2.4 L water and boil until 1.2 L remains, and then filter and remove the residue. Boil again until 600 mL remains. Take 200 mL decoction with warm water three times a day)

【功用】和解表里

(Function: harmonizing exterior and interior)

【主治】伤寒少阳病，寒热往来，胸胁苦满，不思饮食，心烦喜呕，口苦咽干，目眩头痛，舌苔薄白，脉弦数，或妇人伤寒，热入血室

(Treatment: typhoid and lesser yang disease, alternating chills and fever, fullness and discomfort in chest and hypochondrium, no appetite, vexation and vomiting, bitter taste in mouth and dry pharynx, dizziness and headache, thin and white coating, wiry and frequent pulse, or typhoid and heat entering the blood chamber for women)

Fig. 1 An example traditional Chinese medicine (TCM) prescription “Minor Bupleurum Decoction” including composition, usage, function, and treatment. The herbs marked in red, orange, blue, and green are the *jun* (monarch), *chen* (minister), *zuo* (assistant), and *shi* (envoy) ones, respectively. This classification is obtained based on the TCM formulation principle and reveals the different role that each herb plays in the prescription. References to color refer to the online version of this figure

molecules, genes, and cells, while TCM is a holistic concept and uses evidence-based treatment. In TCM, the human body is an organic whole and the various components that make up the human body are structurally inseparable, functionally coordinated, and pathologically influenced by each other. Therefore, in the TCM therapeutic process, in addition to the doctor’s initial approach of collecting symptoms and the conclusion of prescribing treatment, there will be a syndrome induction stage between them. In this part, doctors will judge which syndrome the patient has on the basis of all symptoms and physical signs. In fact, one syndrome may encompass multiple symptoms and one symptom may appear in multiple syndromes, which makes the syndrome induction process complex. Inspired by Jin et al. (2020), we try to reproduce this prescription process and add the syndrome induction part with the possibility of multiple syndromes.

The second type of knowledge concerns the path interaction relationships from external knowledge graphs. Wang XY et al. (2019) used TransE (Bordes et al., 2013) to obtain knowledge graph embeddings, which are integrated into the topic model. Yang et al. (2022) constructed an herb knowledge graph, which is composed mainly of the herb’s effect, flavor, and tropism. Only herb’s one-hot embedding was

used in the model by combination with the output of the H-H graph. In our study, we construct a much larger TCM knowledge graph that covers most of the mainstream TCM knowledge and provides richer and more comprehensive information. We also generate a subgraph based on the path interaction relationships between herbs and symptoms in the knowledge graph, which will be used to obtain the discriminative feature vectors of symptoms and herbs via a graph attention network (GAT).

The third type of knowledge is TCM-informed observations and patterns. Recently, a physics-informed neural network (PINN) (Raissi et al., 2019) was presented to enable the synergistic combination of mathematical models and data using principled physical laws as prior knowledge to act as a regularization agent. Similarly, during the long period of practice and development of TCM, many TCM-informed observations and patterns have been summarized. The most important and influential TCM-informed principle of prescription composition is *jun-chen-zuo-shi* from *Shennong Bencao Jing (Variorum of Shennong's Classic of Materia Medica)*, which symbolizes monarch, minister, assistant, and envoy, four types of people in government. This is a metaphor indicating that different herbs have distinct roles in the prescription. The monarch herbs called *jun* are the most irreplaceable and indispensable, and will target the main cause or primary symptoms of a disease. Then *chen* herbs will assist *jun* herbs in enhancing their effects, just like the minister. The *zuo* herbs are ancillary herbs that are used mainly to improve the effects of *jun* and *chen* and relieve or eliminate the toxicities or side effects of these herbs. The *shi* herbs harmonize other herbs and guide them to target organs or disease locations. In Fig. 1, a famous prescription “Minor Bupleurum Decoction” is composed of several herbs. Chinese Thorowax Root marked in red is the *jun* herb and relieves exterior and fever. The main symptoms treated by Chinese Thorowax Root are alternating chills and fever, and chest fullness or pain, which can be found in the treatment part of the whole prescription. *Scutellaria baicalensis*, marked in orange, is the *chen* herb because it clears heat and reconciles lesser yang. Ginseng, Pinellia Tuber, Fresh Ginger, and Chinese Dates, marked in blue, are the *zuo* herbs. They can reduce vomiting, nourish the spleen and stomach, provide stomach nutrients, and promote

gastrointestinal digestion. Licorice Root, marked in green, is used to regulate all herbs and plays the role of the *shi* herb. Note that different herbs play multiple roles in different prescriptions.

Based on these three types of knowledge, in this work we propose a knowledge-guided and TCM-informed approach for herb recommendation. We construct a new large-scale TCM knowledge graph, TCMKG. Based on path interactions and co-occurrence relationships among symptoms and herbs from the knowledge graph, we determine the discriminative feature vectors of symptoms and herbs via a GAT. We also present a multi-syndrome prediction layer to simulate the possibility of multiple syndromes motivated by the first knowledge type, and introduce the third knowledge type, the TCM-informed principle of prescription composition, called *jun-chen-zuo-shi*, into the prediction layer.

The main contributions of this paper are summarized as follows:

1. We construct a large-scale TCM knowledge graph, TCMKG, which consists of herbs, prescriptions, pieces, diseases, symptoms, syndrome, tropism, properties, flavor, and functions, and covers most of the common knowledge in TCM.
2. We propose a knowledge-guided and TCM-informed approach combining path interactions and co-occurrence relationships from the knowledge graph, multi-syndrome prediction, and the TCM-informed principle of prescription composition.
3. Experimental results on a TCM prescription dataset show that our proposed approach has better performance than state-of-the-art TCM herb recommendation models.

2 Related works

2.1 Graph neural network

Graph neural networks (GNNs) are deep learning based methods that deal with arbitrary graph-structured data. Compared with convolutional neural networks (CNNs) (LeCun et al., 1998), which have achieved remarkable results in both computer vision (Girshick et al., 2014) and natural language processing (Gehring et al., 2017), GNNs are applied to a graph in a non-Euclidean space, whereas CNNs can operate only on regular Euclidean data

like images and texts. Based on CNNs and graphs, variants of GNNs are proposed to collectively aggregate information from graph structure (Xu et al., 2019). GNNs have been used in a wide variety of fields, such as side effect prediction (Zitnik et al., 2018), traffic state prediction (Guo et al., 2019; Zheng et al., 2020), and chemical reaction prediction (Do et al., 2019).

Graph convolutional networks (GCNs) are variants of GNNs that generalize convolutional operation on the graph-structured data. GCNs fall into two categories: spectral information based methods and spatial information based methods. Spectral information based methods work with a spectral representation of the graphs and define convolution from the perspective of graph signal processing. For example, Bruna et al. (2014) extended convolution to general graphs, and Kipf and Welling (2017) proposed a localized first-order approximation of spectral graph convolutions. Spatial information based methods directly operate on the graph and propagate node information from groups of spatially close neighbors. Atwood and Towsley (2016) treated graph convolutions as a diffusion operation, and Hamilton et al. (2017) introduced GraphSAGE, which generates embeddings by sampling and aggregating features from a node's local neighborhood.

Inspired by attention mechanisms (Vaswani et al., 2017), GAT (Veličković et al., 2018) assumes that the contributions of neighbor nodes to the central node are different and uses a multi-head attention mechanism to fuse the importance of neighbors. Zhang JN et al. (2018) introduced a self-attention mechanism to compute an additional attention score for each head. Wang X et al. (2019) proposed a heterogeneous GNN based on hierarchical attention, including node- and semantic-level attention, to fully use the information of different types of nodes and links.

2.2 Herb recommendation

Because medical records contain the core ideas of TCM, the prescriptions in them play an important role in the inheritance and practice of TCM clinical experience. Herb recommendations based on prescriptions include two main types of approaches.

2.2.1 Topic models

The topic model, which is usually used in natural language processing, provides a more principled approach for developing novel models for latent variable discovery and data analysis. TCM prescriptions are documents containing herbs and symptoms as words, and the topic model may assume that herbs and symptoms occurring under the same topic are similar. Ma and Wang (2016) used a symptom-syndrome model to discover the correlation between symptoms and latent syndrome topics. Lin et al. (2016) considered "pathogenesis" as the latent topic to connect symptoms and herbs. Wang S et al. (2016) designed an asymmetric probability generation model to model symptoms, herbs, and diseases simultaneously. Ji et al. (2017) used a multi-content topic model for prescription recommendation by regarding pathogenesis as the latent topic to associate symptoms and herbs. Yao et al. (2018) proposed a prescription topic model inspired by *li-fa-fang-yao*, which represents principles, methods, prescriptions, and Chinese herbs, indicating the basic steps of diagnosis and treatment. Zhao GS et al. (2018) proposed a double-end fusion recommendation framework, including methods of adjusting weight proportion and similarity remapping. Wang XY et al. (2019) proposed the knowledge graph embedding enhanced topic model (KGETM), which adds co-occurrence information in TCM cases and comprehensive semantic relatedness of symptoms and herbs in the TCM knowledge graph. Chen et al. (2018) introduced TCM domain knowledge into a topic model to capture the herb compatibility regularities.

These models exceed traditional non-topic models in terms of generalizability and herb recommendation. However, the topic model is based on the bag-of-words concept, ignores the order and semantic relationships between words, and fails to adequately simulate the multiple relationships in TCM knowledge. Moreover, the topics in TCM often overlap, which might confuse the topic model.

2.2.2 Neural network models

Deep learning based methods have proven their effectiveness in several recent studies. Li and Yang (2017) adopted bi-directional recurrent neural networks (BRNNs) to do text representation learning of the herb words in the TCM literature

for the treatment complement task. Ruan et al. (2019a, 2019b) integrated the auto-encoder model with meta-path to mine the TCM heterogeneous information network. Li and Yang (2019) proposed a novel sequence-to-sequence model whose decoder is designed with a coverage mechanism and a soft loss function. As mentioned above, GNNs were proposed to leverage the graph structure and aggregate node information from the neighborhoods. Based on GNNs, which can fully learn graph representations, several methods have been proposed using the prescriptions' own graphs and external graphs for herb recommendation. Zhou et al. (2021) fused a CNN with phenotype and molecule information. Jin et al. (2020) obtained low-dimensional embedding of symptoms and herbs through a GCN, and then proposed a syndrome-aware mechanism to obtain embedding of symptom sets in the prescription. Yang et al. (2022) constructed the herb knowledge graph and proposed multi-layer information fusion based on a GCN for knowledge-driven herb recommendation, using the properties of herbs in the knowledge graph as auxiliary information to help the model better fit the feature representation of herbs. Zhao W et al. (2022) considered the n -ary relationships among symptoms, state elements, syndrome types, and herbs, and proposed a prescription recommendation model based on a multi-graph convolutional network. Although some research has incorporated additional TCM knowledge, such as a fixed vector of herbs trained through knowledge graphs, there still lacks more comprehensive TCM knowledge to guide herb recommendation. In this work, we use a GAT to dynamically train both path interaction and co-occurrence relationship graphs to ensure that the representation of each node is also continuously updated during training.

3 Proposed methodology

We formulate the definition of herb recommendation in Section 3.1, and introduce our TCM knowledge graph and the overall architecture of our approach in Sections 3.2 and 3.3, respectively.

3.1 Problem definition

Herb recommendation aims to learn the recommended route of prescribing from symptoms to herbs based on a large training corpus of prescrip-

tions; we formulate this problem as follows. Let $S = \{s_1, s_2, \dots, s_M\}$ represent a symptom set containing M different symptoms and $H = \{h_1, h_2, \dots, h_N\}$ an herb set containing N different herbs. Each prescription $p = \{\{s_1, s_2, \dots, s_m\}, \{h_1, h_2, \dots, h_n\}\}$ consists of at least one set of herbs and one set of symptoms (an example is shown in Fig. 1). The introduced directed TCM knowledge graph $G = (V, E)$ contains n TCM-related nodes $V = \{V_1, V_2, \dots, V_n\}$ and edges $E \subseteq V \times V$, where $(V_j, V_i) \in E$ denotes an edge from a node V_j to a node V_i . We assume that every node $V_i \in V$ is represented by a d -dimensional vector.

Given a symptom set s_{set} , the task of herb recommendation is to recommend an herb set h_{set} . Usually this task is converted to learning a prediction function $\hat{y}_{s_{\text{set}}} = f(s_{\text{set}}, H|\theta)$, where $\hat{y}_{s_{\text{set}}}$ represents the probability matrix and θ indicates the trainable parameters. The value of dimension i of $\hat{y}_{s_{\text{set}}}$ represents the probability of the i^{th} herb in the recommendation.

3.2 Knowledge graph

Because there is no authoritative and public structured TCM knowledge graph, we have constructed our TCM knowledge graph by Neo4j. Texts from TCM textbooks, dictionaries, standards, and other officially released data are extracted into triples as the basic unit of the knowledge graph. Triples are 3-tuples consisting of a subject, a predicate (or relation), and an object, usually expressed as (s, r, o) . Both the subject and the object are regarded as entities, and the triples contain the relationship between entities.

We pre-define 10 types of entities based on the field of TCM: herb, prescription, pieces, disease, symptom, syndrome, tropism, property, flavor, and function. The pre-defined relations between them are shown in Fig. 2. Property, flavor, and tropism often appear in the description text of the herb, showing the understanding of the herb from the perspective of Chinese medicine. The property refers to the four natures, or four characteristics of cold, heat, warm, and cool, which reflect the body's semi-interior phase change after the herb acts on the body. Flavor means the herb's localization in Yin and Yang. In the five flavors, pungent and sweet belong to Yin, whereas sour, bitter, and salty belong to Yang. Tropism describes the therapeutic effect of the herb on the pathology of certain viscera and

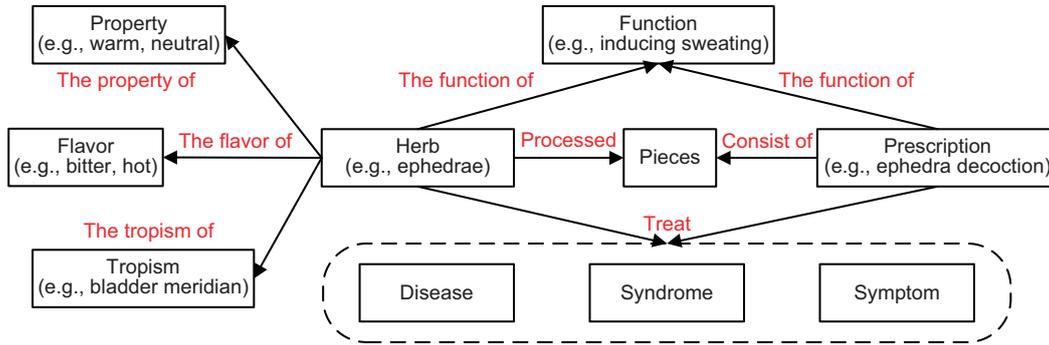


Fig. 2 Pre-defined types of entities and relations of TCMKG (Rectangles represent entities and arrows indicate relationships between two entities)

meridians. The remaining entities are common and can be easily defined.

All entities and relationships are stored in the knowledge graph as triples. We have constructed 128 357 entities and 1 138 718 triples in TCMKG. Detailed information about TCMKG can be found in the Appendix.

3.3 Proposed model

Our proposed approach, called MKMI-GAT, is shown in Fig. 3, and the details of the graph embedding module are shown in Fig. 4. The graph embedding module takes symptom list $S = \{s_1, s_2, \dots, s_M\}$, herb list $H = \{h_1, h_2, \dots, h_N\}$, and TCM knowledge graph $G = (V, E)$ as input, and will update all embeddings of herbs and symptoms during each training iteration. The whole model MKMI-GAT will take a symptom set $s_{\text{set}} = \{s_1, s_2, \dots, s_m\}$ as input in both the training and testing stages, and outputs the predicted probability vector $\hat{y}_{s_{\text{set}}}$ in which the value of dimension i represents the probability of herb i in the recommendation.

3.3.1 Graph embedding module

We construct two graphs for learning discriminative feature embeddings of herbs and symptoms (Fig. 4). The first graph is the path interaction graph. Distinct from other research efforts that have not fully used knowledge graph information, we incorporate TCMKG into the training of the entire model. For each herb and symptom that have occurred in prescriptions, we use TCMKG to take side information into account.

Assuming that a TCM prescription $p = \{\{s_1, s_2\}, \{h_1, h_2\}\}$ has two symptoms and two herbs,

all the subgraphs of these four nodes are shown in the medical knowledge graph part of Fig. 4. Based on Neo4j, we use the all-shortest-paths function in Cypher to find the most relevant relationships or paths between two nodes while avoiding excessive computational burden. Every two nodes need to be queried once by the all-shortest-paths function. We regard these paths as their path interaction and collect all paths to construct a path interaction graph. For instance, as we see in Fig. 4, the shortest path between h_1 and h_2 is via a green node above them. In TCMKG, we can know that different herbs might have the same functions, properties, and flavors, which is exactly the possible category of this green node. The shortest path between h_1 and s_1 can pass through a yellow node (syndrome entity) or red node (disease entity), and both paths will be counted. The path interaction graph is defined as follows:

$$G(\text{PI}) = \begin{cases} 1, & \text{if } (V_i, V_j) \in \bigcup f_{\text{asp}}(n_j, n_k), \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $n_j, n_k \in p$, and f_{asp} is the all-shortest-paths function in Cypher and is applied in TCMKG.

The second graph was inspired by Jin et al. (2020, 2022). We construct a co-occurrence relationship graph based on the co-occurrence frequency in prescriptions, which represents the intrinsic therapeutic correlations between herbs and symptoms.

Taking a prescription $p' = \langle \{s_1, s_2, \dots, s_m\}, \{h_1, h_2, \dots, h_n\} \rangle$ as an example, we can obtain a new treating relationship edge according to the co-occurrence of them in the form of $\{(s_1, h_1), \dots, (s_1, h_n), \dots, (s_m, h_1), \dots, (s_m, h_n)\}$. The whole co-occurrence relationship graph is defined as

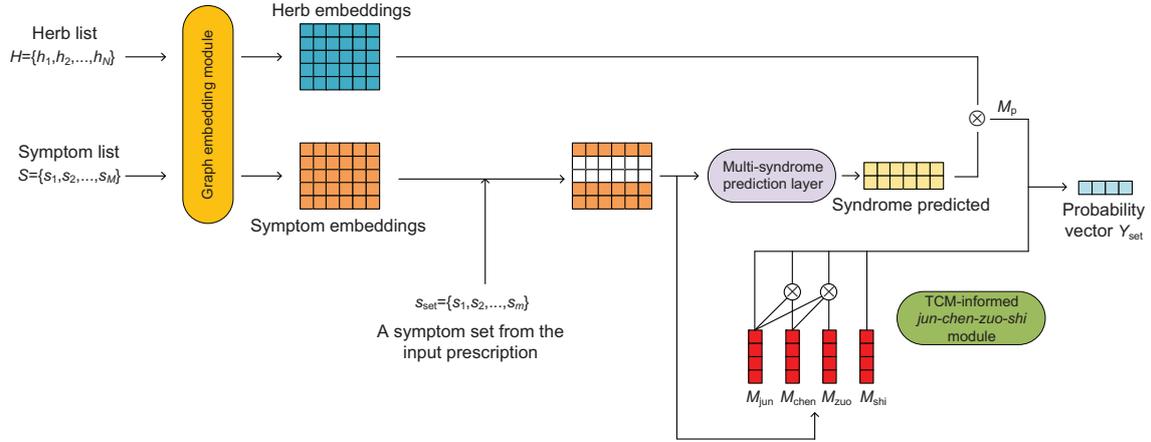


Fig. 3 The overall architecture of our proposed approach, MKMI-GAT

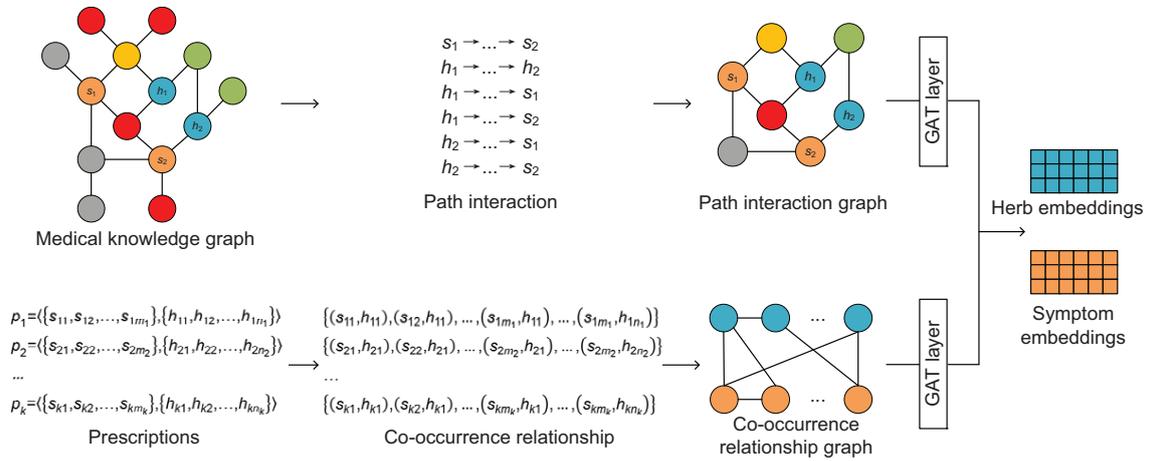


Fig. 4 The graph embedding module from MKMI-GAT. The top half shows the process of constructing the path interaction graph. It uses the all-shortest-paths function to generate paths between two nodes from the medical knowledge graph and construct a path interaction graph. The bottom half shows the process of constructing the co-occurrence relationship graph. It generates a new treatment relationship edge in the co-occurrence relationship graph according to the co-occurrence of each pair of symptom and herb. References to color refer to the online version of this figure

follows:

$$G(\text{CO}) = \begin{cases} 1, & \text{if } (s, h) \text{ co-occur in } p', \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

3.3.2 Graph attention network

After the construction of the path interaction graph and co-occurrence relationship graph, the embeddings of nodes in the embedding layer are summarized as follows:

$$E = \{e_h, e_s, e_o | h, s, o \in V_{G(\text{PI})} \cup V_{G(\text{CO})}\}, \quad (3)$$

where h , s , and o refer to herbs, symptoms, and other nodes, respectively, and $e_h \in \mathbb{R}^{N \times d}$, $e_s \in \mathbb{R}^{M \times d}$, and

$e_o \in \mathbb{R}^{|O| \times d}$, where $|O|$ is the number of all other nodes besides herbs and symptoms. V_G represents all nodes in graph G , and all the embeddings are randomly initialized and trained during optimization.

With the initial representation of all nodes, according to GATv2 (Brody et al., 2022), the scoring function for every edge (j, i) used to indicate the importance of the features of neighbor j to node i is formulated as follows:

$$e(h_i, h_j) = a^T \text{LeakyReLU}(W \cdot \text{concat}(h_i, h_j)), \quad (4)$$

where $a \in \mathbb{R}^{\mathbb{D}'}$, $W \in \mathbb{R}^{\mathbb{D}' \times 2\mathbb{D}}$, \mathbb{D} is the dimension of the hidden layers, \mathbb{D}' is an internal hyperparameter, and $\text{concat}()$ denotes vector concatenation.

The attention function that calculates the attention scores across all neighbors $j' \in \text{Neigh}(i)$ is defined as

$$\alpha_{ij} = \text{softmax}(e(h_i, h_j)) = \frac{\exp(e(h_i, h_j))}{\sum_{j' \in \text{Neigh}(i)} \exp(e(h_i, h_{j'}))}, \quad (5)$$

where $\text{Neigh}(i) = \{j \mid j \in V, (j, i) \in E\}$. We obtain the new representation of i by computing the weighted average of itself and the transformed features of its neighbor nodes as follows:

$$h'_i = \alpha_{ii} \cdot Wh_i + \sum_{j \in \text{Neigh}(i)} \alpha_{ij} \cdot Wh_j. \quad (6)$$

3.3.3 Representation fusion

We have obtained herb and symptom features from the path interaction (PI) graph and co-occurrence relationship (CO) graph as $h^{\text{PI}}, h^{\text{CO}}, s^{\text{PI}}$, and s^{CO} , respectively, after each epoch's neighbor information aggregation. Combined with herb embedding e_h and symptom embedding e_s , we have

$$\hat{h}_f = W_{hf} \text{concat}(e_h, h^{\text{PI}}, h^{\text{CO}}), \quad (7)$$

$$\hat{s}_f = W_{sf} \text{concat}(e_s, s^{\text{PI}}, s^{\text{CO}}). \quad (8)$$

Input symptom set $s_{\text{set}} = \{s_1, s_2, \dots, s_m\}$ will be transformed into one-hot form s_{onehot} and used as a lookup mechanism to obtain the full input symptom representation:

$$s_{\text{in}} = s_{\text{onehot}} \cdot \hat{s}_f, \quad (9)$$

where $s_{\text{in}} \in \mathbb{R}^{M \times \mathbb{D}}$.

3.3.4 Multi-syndrome prediction layer

As mentioned above, one syndrome may encompass multiple symptoms and one symptom may appear in multiple syndromes. Therefore, after obtaining the representation of the current input set of symptoms, we add a multi-syndrome prediction layer to predict the possible syndromes by introducing the multi-head self-attention mechanism with N heads:

$$M_{\text{syn}} = \text{concat}_{l=0}^N (A(W_q^{(l)} s_{\text{in}}, W_k^{(l)} s_{\text{in}}) W_v^{(l)} s_{\text{in}}), \quad (10)$$

$$A(W_q^{(l)} s_{\text{in}}, W_k^{(l)} s_{\text{in}}) = \text{softmax}\left(\frac{(W_q^{(l)} s_{\text{in}})^T (W_k^{(l)} s_{\text{in}})}{\sqrt{\mathbb{D}}}\right), \quad (11)$$

where $W_q^{(l)}, W_k^{(l)}$, and $W_v^{(l)}$ are trainable parameters and the number of heads N decides the number of possible syndromes.

3.3.5 TCM-informed *jun-chen-zuo-shi* module

To add constraints of TCM-informed principles in deep learning, we propose a TCM-informed *jun-chen-zuo-shi* module to control the final prediction. We assume that there are four matrices $M_{\text{jun}}, M_{\text{chen}}, M_{\text{zuo}}$, and $M_{\text{shi}} \in \mathbb{R}^{M \times N}$ to represent the relationship between symptoms and four types of herbs. We collect each herb's *jun-chen-zuo-shi* occurrence for every symptom and herb tuple in *Dictionary of Traditional Chinese Medicine Prescriptions* and initialize $M_{\text{jun}}, M_{\text{chen}}, M_{\text{zuo}}$, and M_{shi} separately after L1 normalization.

In each epoch, the one-hot form of symptoms, s_{onehot} , will be calculated with the *jun-chen-zuo-shi* matrix to obtain their first-step *jun-chen-zuo-shi* probability vectors as follows:

$$\begin{cases} M'_{\text{jun}} = s_{\text{onehot}} \cdot M_{\text{jun}}, \\ M'_{\text{chen}} = s_{\text{onehot}} \cdot M_{\text{chen}}, \\ M'_{\text{zuo}} = s_{\text{onehot}} \cdot M_{\text{zuo}}, \\ M'_{\text{shi}} = s_{\text{onehot}} \cdot M_{\text{shi}}. \end{cases} \quad (12)$$

We integrate TCM-informed principles of prescription composition into the subsequent steps. As we have illustrated above, *jun* herbs target the main cause and are most irreplaceable and indispensable; thus, the input symptom list could directly determine the probability vector of *jun*:

$$\hat{M}_{\text{jun}} = M'_{\text{jun}}. \quad (13)$$

Chen herbs are used to assist *jun* herbs in enhancing their effects. Thus, the choice of *chen* herb will be influenced by the priority decision of the *jun* herb:

$$\hat{M}_{\text{chen}} = W_{\text{chen}} M'_{\text{jun}} + M'_{\text{chen}}. \quad (14)$$

Zuo herbs are assisting herbs, and are used mainly to improve the effects of *jun* and *chen* herbs and relieve or eliminate the toxicities or side effects of these herbs. To implement this principle, the model will learn to adjust the parameters of \hat{W}_{zuo} in training according to M'_{jun} and M'_{chen} :

$$\hat{M}_{\text{zuo}} = W_{\text{zuo}} (M'_{\text{jun}} + M'_{\text{chen}}) + M'_{\text{zuo}}. \quad (15)$$

There are generally only a few fixed *shi* herbs in datasets, and not all prescriptions contain *shi* herbs. In our statistics, only 55.4% of the prescriptions have *shi* herbs and 71.1% of them are Licorice Roots. As

most of *shi* herbs are unrelated to other herbs, \hat{M}_{shi} will not have adjustments:

$$\hat{M}_{shi} = M'_{shi}. \quad (16)$$

3.3.6 Herb recommendation

After determining the representation of herbs \hat{h}_f , the probability matrix of syndromes M_{syn} , and herb prediction from the TCM-informed *jun-chen-zuo-shi* module \hat{M}_{jun} , \hat{M}_{chen} , \hat{M}_{zuo} , \hat{M}_{shi} , the prediction vector of the herb set is shown as follows:

$$M_p = \text{ReLU}(W_{syn}M_{syn} + b_{syn})\hat{h}_f, \quad (17)$$

$$\hat{y}_{set} = \alpha M_p + (1 - \alpha)(\hat{M}_{jun} + \hat{M}_{chen} + \hat{M}_{zuo} + \hat{M}_{shi}), \quad (18)$$

where W_{syn} and b_{syn} are trainable parameters, and α is a hyperparameter used to balance the two parts of the predictions.

We use the probability of TOP@K as the recommendation result and BCELoss (Braga-Neto, 2020) as the loss function.

4 Experiments and analysis

4.1 Dataset

We use the benchmark dataset collected by Yao et al. (2018). After pre-processing there are 33 765 prescription pieces containing 390 symptoms and 811 herbs.

4.2 Evaluation

Given a symptom set, our proposed model outputs a set of herbs to relieve the symptoms. To evaluate the performance of our model, we adopt the following measures, which are commonly used in recommendation systems:

$$\text{Precision@K} = \frac{|\text{Top}(\hat{M}_p, K) \cap \text{herb_set}|}{K}, \quad (19)$$

$$\text{Recall@K} = \frac{|\text{Top}(\hat{M}_p, K) \cap \text{herb_set}|}{|\text{herb_set}|}, \quad (20)$$

$$\text{F1@K} = \frac{2 \cdot \text{Precision@K} \cdot \text{Recall@K}}{\text{Precision@K} + \text{Recall@K}}, \quad (21)$$

where herb_set is the ground-truth herb set in the prescription, and $\text{Top}(\hat{M}_p, K)$ is the top K herbs with the highest prediction scores given \hat{M}_p .

4.3 Baselines

We compare the proposed MKMI-GAT with the following baselines:

KDHR (Yang et al., 2022) proposes a multi-layer information fusion based on a GCN using the properties of herbs in the knowledge graph as auxiliary information to help the model better fit the feature representation of herbs.

SMGCN (Jin et al., 2020) develops a series of GCNs to simultaneously learn the symptom-herb, symptom-symptom, and herb-herb graphs, and recommends herbs using the syndrome induction process.

HC-KGETM (Wang XY et al., 2019) integrates the TransE embeddings obtained from a TCM knowledge graph into the topic model, and adds co-occurrence information in TCM cases and comprehensive semantic relatedness of symptoms and herbs in a TCM knowledge graph.

GCNConv (Kipf and Welling, 2017) introduces the degree matrix of the node to solve the problem of self-loops and the normalization of the adjacency matrix, and updates the node by adding embedding of neighbor nodes.

GATConv (Veličković et al., 2018) proposes a shared attention mechanism to distinguish the importance of different neighboring nodes. Every node updates its representation by attending to its neighbors.

4.4 Implementation details

We implement our model MKMI-GAT and the comparative methods using PyTorch. For the topic model HC-KGETM, we follow the parameter settings in Wang XY et al. (2019). For other neural network models, we perform fivefold cross-validation on the training set. For SMGCN and KDHR, the initial learning rate is 0.0003. For MKMI-GAT and other baseline models, the initial learning rate is 0.0001. The learning rate is attenuated every seven epochs. The dropout ratio of MKMI-GAT is 0.5. The number of epochs is set to 200. An early stopping strategy is applied, which means that the model will stop training when the loss of the validation set does not decrease within the seven consecutive epochs. We use the Adam optimizer (Kingma and Ba, 2015) to train MKMI-GAT with a batch size

of 256. The batch size is 512 for other methods. For MKMI-GAT, the dimensions of the embedded layer and the hidden layer are both 256. α is initialized with 0.5. GATConv, GCNConv, HC-KGETM, and SMGCN do not use any information from the knowledge graph. KDHR uses herb embeddings from TCMKG based on TransE.

4.5 Results and analysis

The results of MKMI-GAT and the baseline models are shown in Table 1.

MKMI-GAT achieves satisfactory performance. It outperforms KDHR, whose performance ranks first among the baseline approaches, by 18.0%, 19.0%, and 18.6% in terms of F1@5, F1@10, and F1@20 respectively, showing that by integrating the path interaction and co-occurrence relationship graphs, the multi-syndrome prediction layer, and the TCM-informed *jun-chen-zuo-shi* module, our model can learn a better representation of herbs and symptoms and output a more accurate prediction of herbs.

Among GCNConv, GATConv, HC-KGETM, SMGCN, and KDHR, the GATConv and HC-KGETM models have the worst performances. For GATConv, the reason might be that the network, fused only with the attention mechanism, might lose too much global information. The problem with HC-KGETM (Wang XY et al., 2019) is that it ignores symptom set information and is weaker in collecting high-order information from the graph compared with GCN and GAT. With the construction of symptom–herb, symptom–symptom, and herb–herb graphs, SMGCN (Jin et al., 2020) learns better symptom embedding and herb embedding, and performs better than HC-KGETM. KDHR (Yang et al., 2022) uses a multi-layer information fusion mechanism to integrate the different levels of information obtained from each layer of GCN, adds external herb information from an herb knowledge graph, and outperforms the baseline methods, which shows the importance of external knowledge.

4.5.1 Ablation study

To better understand and investigate the individual components of our proposed MKMI-GAT model, we record the results of ablation analysis (Ta-

ble 2) and have the following findings: (1) From a whole perspective, all the knowledge integrated into our proposed model, including the graph embedding module, multi-syndrome prediction layer, and TCM-informed *jun-chen-zuo-shi* module, is verified to be effective based on the better performance of the MKMI-GAT model. (2) Without the TCM-informed *jun-chen-zuo-shi* module, the accuracy and recall both drop slightly, indicating that the module improves the model and finds one or two correct herbs based on principles of prescription composition. (3) Both the path interaction and co-occurrence relationship graphs are necessary for the representation learning of herbs and symptoms. (4) The most effective part is the multi-syndrome prediction layer. Without this layer, there is a significant drop in precision. The reason for this is that there is often more than one symptom in the input prescription, and there may be primary and secondary relationships between them, as well as concomitant and other relationships. Through the multi-head self-attention mechanism in the multi-syndrome prediction layer, the model can learn multiple symptom-focused patterns to decide which symptom should be prioritized and to predict the output list of herbs. When we go further and remove the whole syndrome layer, the performance of Recall@20 drops severely, which indicates that for those prescriptions with a small number of symptoms, the generalization ability of the model becomes weaker when it lacks the attention mechanism.

4.5.2 Impact of the syndrome number

We investigate the impact of the number of heads in the multi-syndrome prediction layer when using self-attention. The number of heads corresponds to the number of syndromes we simulate and induct in the prediction layer. When there is only one head, or a single syndrome, the results are recorded in Table 2. Considering that the dimension of embedding must be divisible by the number of heads, we test two cases of two and four symptoms. As shown in Fig. 5, the model performs better when the syndrome number is 2. This may be because when the number of pre-defined syndromes is too large, the complexity of the model increases and the prediction accuracy decreases in the case of single symptom input.

Table 1 Performance evaluation of different models

Model	Precision			Recall			F1-Score		
	<i>P@5</i>	<i>P@10</i>	<i>P@20</i>	<i>R@5</i>	<i>R@10</i>	<i>R@20</i>	<i>F1@5</i>	<i>F1@10</i>	<i>F1@20</i>
GATConv (Velicković et al., 2018)	0.192	0.151	0.114	0.132	0.183	0.288	0.156	0.165	0.163
GCNConv (Kipf and Welling, 2017)	0.211	0.169	0.124	0.138	0.218	0.320	0.167	0.190	0.178
HC-KGETM (Wang XY et al., 2019)	0.181	0.142	0.106	0.120	0.199	0.291	0.145	0.166	0.155
SMGCN (Jin et al., 2020)	0.193	0.157	0.115	0.127	0.203	0.303	0.153	0.177	0.167
KDHR (Yang et al., 2022)	0.237	0.188	0.140	0.182	0.283	0.416	0.206	0.226	0.210
MKMI-GAT	0.288	0.229	0.169	0.210	0.327	0.474	0.243	0.269	0.249

The best values are in bold

Table 2 Ablation study results for MKMI-GAT

Model	Precision			Recall			F1-Score		
	<i>P@5</i>	<i>P@10</i>	<i>P@20</i>	<i>R@5</i>	<i>R@10</i>	<i>R@20</i>	<i>F1@5</i>	<i>F1@10</i>	<i>F1@20</i>
MKMI-jczs	0.2848	0.2265	0.1677	0.2081	0.3247	0.4722	0.2405	0.2669	0.2475
MKMI-PI	0.2811	0.2233	0.1663	0.2053	0.3208	0.4707	0.2373	0.2633	0.2458
MKMI-CO	0.2759	0.2191	0.1627	0.2021	0.3145	0.4591	0.2334	0.2583	0.2402
MKMI-multi-syn	0.2785	0.2218	0.1645	0.2031	0.3186	0.4649	0.2349	0.2615	0.2430
MKMI-syn	0.2757	0.2181	0.1625	0.2000	0.3112	0.4565	0.2318	0.2564	0.2397
MKMI-GAT	0.2884	0.2288	0.1687	0.2098	0.3266	0.4744	0.2429	0.2691	0.2489

The best values are in bold

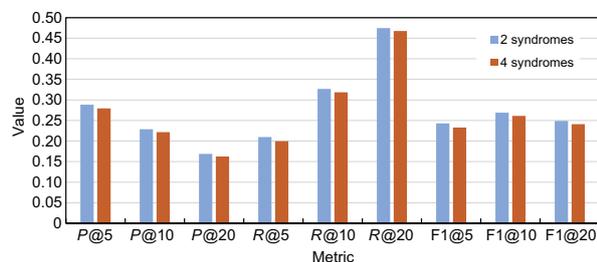


Fig. 5 Performance at different syndrome numbers in the multi-syndrome prediction layer

4.5.3 Study of the TCM-informed *jun-chen-zuo-shi* module

We randomly select an example to visualize the four symptom–herb mapping matrices in the TCM-informed *jun-chen-zuo-shi* module after training, and the results are shown in Fig. 6. In this example, five symptoms on the vertical axis from top to bottom are asthma, opisthotonus, unconsciousness, fever, and epilepsy. The herbs in the horizontal axis are arranged from left to right according to the probability values in the final output matrix. This figure allows us to investigate the contribution of the *jun-chen-zuo-shi* module to each symptom in the top 20 herbs recommended in the final output.

The first correct herb with the highest prediction probability is cinnabar, which is a common *jun* herb in TCM for epilepsy, as can be seen in the dark grid in the lower left corner of Fig. 6a. The same effect of the *jun* prediction matrix is reflected in the second predicted herb, musk, because musk is a famous herb to open the orifices and induce resuscitation, which is specific for unconsciousness. The third and fourth correctly predicted herbs are the sixth and eighth herbs in the graph respectively, and it can be seen that it is the matrix of *chen* playing its role. The herb with the fifth highest prediction is tall gastrodia tuber. Although this prediction is wrong, this herb is described in *Pharmacopoeia of the People's Republic of China* as a suitable herb for treating pediatric convulsions, epileptic convulsions, hand and foot disorders, and numbness of the limbs. This also shows that this symptom–herb prediction module is of high value in practical scenarios to assist doctors in prescribing. The *shi* graph in Fig. 6d illustrates that the three other types of herbs are given much higher priority than a *shi* herb in prediction, which correctly reflects the status and priority of the four herbs in the principles of prescription composition.

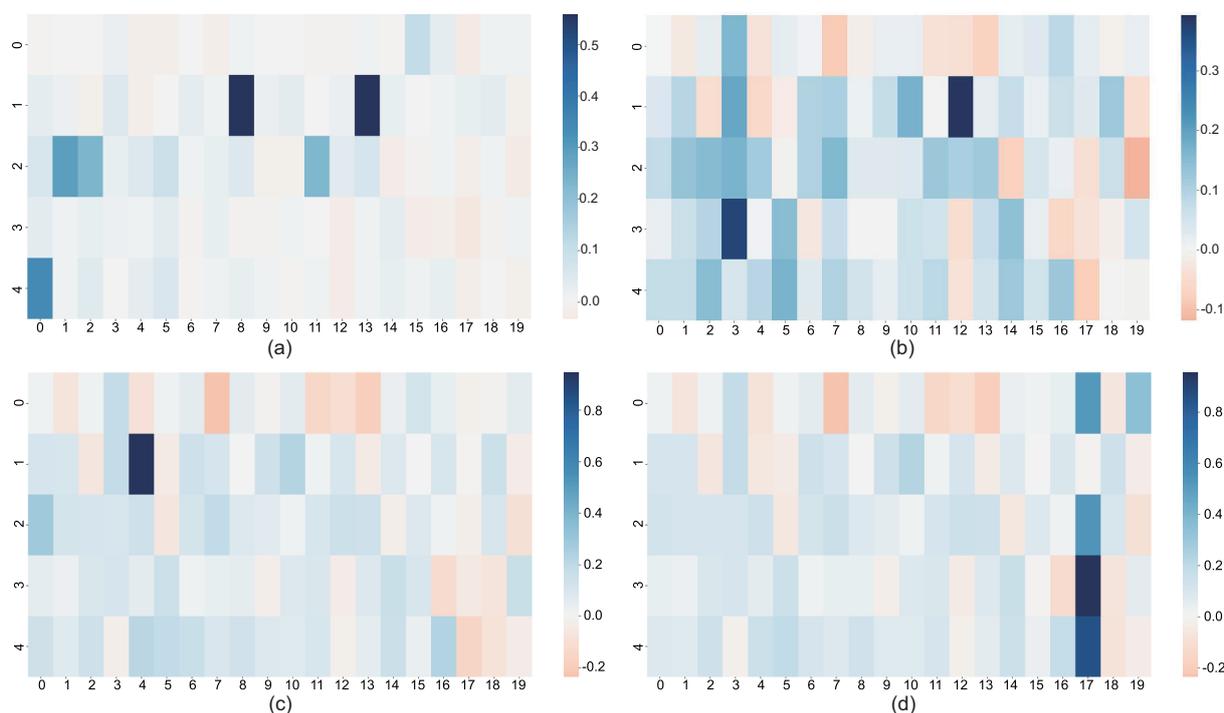


Fig. 6 A visualization example for symptom–herb mapping matrices in the *jun-chen-zuo-shi* module: (a) *jun*; (b) *chen*; (c) *zuo*; (d) *shi*. The five symptoms on the vertical axis from top to bottom are asthma, opisthotonus, unconsciousness, fever, and epilepsy. The herbs in the horizontal axis are arranged from left to right according to the probability values in the final output matrix

Based on the above analysis, we believe that TCM-informed observations and principles are effective in allowing the model to add the constraint of *jun-chen-zuo-shi* knowledge, thus boosting the performance of herb recommendation.

5 Conclusions

In this study, we propose a knowledge-guided and TCM-informed approach based on GATs for herb recommendation. We construct a large-scale TCM knowledge graph, TCMKG, which covers most of the common knowledge in the TCM field. The TCMKG path interaction graph and co-occurrence relationship graph based on the treatment relationship from prescriptions are trained by GATs and output the discriminative representation of herbs and symptoms. After the fusion of these two representations, the multi-syndrome prediction layer will generate a possible representation of syndromes through multi-head attention. The model also introduces the TCM-informed *jun-chen-zuo-shi* module, which provides knowledge about the principles of prescription composition. Experiments on a TCM dataset con-

firm the effectiveness of our method, and the ablation study affirms the significance of integration with each external knowledge type. In future work, we shall focus more on representation learning of TCMKG, because the current all-shortest-paths method does not make good use of multi-hop information among herbs, pieces, and prescriptions. We shall also try to modify the TCM-informed *jun-chen-zuo-shi* module to allow the *jun-chen-zuo-shi* principle to be added to the model holistically.

Contributors

Zhe JIN and Yunhe PAN designed the research. Zhe JIN and Jiayu MIAO processed the data and drafted the paper. Yin ZHANG helped organize the paper. Yi YANG, Yueting ZHUANG, and Yin ZHANG revised and finalized the paper.

Compliance with ethics guidelines

Yi YANG, Yueting ZHUANG, and Yunhe PAN are editorial board members for *Frontiers of Information Technology & Electronic Engineering*, and they were not involved with the peer review process of this paper. All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Atwood J, Towsley D, 2016. Diffusion-convolutional neural networks. Proc 30th Int Conf on Neural Information Processing Systems, p.2001-2009.
- Bordes A, Usunier N, Garcia-Duran A, et al., 2013. Translating embeddings for modeling multi-relational data. Proc 26th Int Conf on Neural Information Processing Systems, p.2787-2795.
- Braga-Neto U, 2020. Fundamentals of Pattern Recognition and Machine Learning. Springer, Cham, Switzerland.
- Brody S, Alon U, Yahav E, 2022. How attentive are graph attention networks? Proc 10th Int Conf on Learning Representations.
- Bruna J, Zaremba W, Szlam A, et al., 2014. Spectral networks and locally connected networks on graphs. Proc 2nd Int Conf on Learning Representations.
- Chen XT, Ruan CY, Zhang YC, et al., 2018. Heterogeneous information network based clustering for categorizations of traditional Chinese medicine formula. Proc IEEE Int Conf on Bioinformatics and Biomedicine, p.839-846. <https://doi.org/10.1109/BIBM.2018.8621431>
- Cheung F, 2011. TCM: made in China. *Nature*, 480(7378): S82-S83. <https://doi.org/10.1038/480S82a>
- Do K, Tran T, Venkatesh S, 2019. Graph transformation policy network for chemical reaction prediction. Proc 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.750-760. <https://doi.org/10.1145/3292500.3330958>
- Gehring J, Auli M, Grangier D, et al., 2017. A convolutional encoder model for neural machine translation. Proc 55th Annual Meeting of the Association for Computational Linguistics, p.123-135. <https://doi.org/10.18653/v1/P17-1012>
- Girshick R, Donahue J, Darrell T, et al., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.580-587. <https://doi.org/10.1109/CVPR.2014.81>
- Guo SN, Lin YF, Feng N, et al., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proc 33rd AAAI Conf on Artificial Intelligence, p.922-929. <https://doi.org/aaai.v33i01.3301922>
- Hamilton WL, Ying R, Leskovec J, 2017. Inductive representation learning on large graphs. Proc 31st Int Conf on Neural Information Processing Systems, p.1025-1035.
- Ji WD, Zhang Y, Wang XL, et al., 2017. Latent semantic diagnosis in traditional Chinese medicine. *World Wide Web*, 20(5):1071-1087. <https://doi.org/10.1007/s11280-017-0443-3>
- Jin YY, Zhang W, He XN, et al., 2020. Syndrome-aware herb recommendation with multi-graph convolution network. Proc IEEE 36th Int Conf on Data Engineering, p.145-156. <https://doi.org/10.1109/ICDE48307.2020.00020>
- Jin YY, Ji WD, Zhang W, et al., 2022. A KG-enhanced multi-graph neural network for attentive herb recommendation. *IEEE/ACM Trans Comput Biol Bioinform*, 19(5):2560-2571. <https://doi.org/10.1109/TCBB.2021.3115489>
- Kingma DP, Ba J, 2015. Adam: a method for stochastic optimization. Proc 3rd Int Conf on Learning Representations.
- Kipf TN, Welling M, 2017. Semi-supervised classification with graph convolutional networks. Proc 5th Int Conf on Learning Representations.
- LeCun Y, Bottou L, Bengio Y, et al., 1998. Gradient-based learning applied to document recognition. *Proc IEEE*, 86(11):2278-2324. <https://doi.org/10.1109/5.726791>
- Li W, Yang Z, 2017. Distributed representation for traditional Chinese medicine herb via deep learning models. <https://arxiv.org/abs/1711.01701>
- Li W, Yang Z, 2019. Exploration on generating traditional Chinese medicine prescriptions from symptoms with an end-to-end approach. Proc 8th CCF Int Conf on Natural Language Processing and Chinese Computing, p.486-498. https://doi.org/10.1007/978-3-030-32233-5_38
- Lin F, Xiahou J, Xu ZX, 2016. TCM clinic records data mining approaches based on weighted-LDA and multi-relationship LDA model. *Multim Tools Appl*, 75(22):14203-14232. <https://doi.org/10.1007/s11042-016-3363-9>
- Lu YT, Liu HY, Yang Kuo, et al., 2022. A comprehensive update: gastrointestinal microflora, gastric cancer and gastric premalignant condition, and intervention by traditional Chinese medicine. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 23(1):1-18. <https://doi.org/10.1631/jzus.B2100182>
- Ma JL, Wang ZJ, 2016. Discovering syndrome regularities in traditional Chinese medicine clinical by topic model. Proc 11th Int Conf on P2P, Parallel, Grid, Cloud and Internet Computing, p.157-162. https://doi.org/10.1007/978-3-319-49109-7_15
- Mao YJ, Meng LK, Liu HY, et al., 2022. Therapeutic potential of traditional Chinese medicine for vascular endothelial growth factor. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 23(5):353-364. <https://doi.org/10.1631/jzus.B2101055>
- Raissi M, Perdikaris P, Karniadakis GE, 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving non-linear partial differential equations. *J Comput Phys*, 378:686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- Ruan CY, Ma JG, Wang Y, et al., 2019a. Discovering regularities from traditional Chinese medicine prescriptions via bipartite embedding model. Proc 28th Int Joint Conf on Artificial Intelligence, p.3346-3352.
- Ruan CY, Wang Y, Zhang YC, et al., 2019b. Exploring regularity in traditional Chinese medicine clinical data using heterogeneous weighted networks embedding. Proc Int Conf on Database Systems for Advanced Applications, p.310-313. https://doi.org/10.1007/978-3-030-18590-9_35
- Vaswani A, Shazeer N, Parmar N, et al., 2017. Attention is all you need. Proc 31st Int Conf on Neural Information Processing Systems, p.6000-6010.

- Veličković P, Cucurull G, Casanova A, et al., 2018. Graph attention networks. Proc 6th Int Conf on Learning Representations.
- Wang S, Huang EW, Zhang RS, et al., 2016. A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional Chinese medicine patient records. Proc IEEE Int Conf on Bioinformatics and Biomedicine, p.411-418. <https://doi.org/10.1109/BIBM.2016.7822553>
- Wang X, Ji HY, Shi C, et al., 2019. Heterogeneous graph attention network. Proc World Wide Web Conf, p.2022-2032. <https://doi.org/10.1145/3308558.3313562>
- Wang XY, Zhang Y, Wang XL, et al., 2019. A knowledge graph enhanced topic modeling approach for herb recommendation. Proc 24th Int Conf on Database Systems for Advanced Applications, p.709-724.
- Xu KYL, Hu WH, Leskovec J, et al., 2019. How powerful are graph neural networks? Proc 7th Int Conf on Learning Representations.
- Yang Y, Rao YL, Yu MH, et al., 2022. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neur Netw*, 146:1-10. <https://doi.org/10.1016/j.neunet.2021.11.010>
- Yao L, Zhang Y, Wei BG, et al., 2018. A topic modeling approach for traditional Chinese medicine prescriptions. *IEEE Trans Knowl Data Eng*, 30(6):1007-1021. <https://doi.org/10.1109/TKDE.2017.2787158>
- Zhang JN, Shi XJ, Xie JY, et al., 2018. GaAN: gated attention networks for learning on large and spatiotemporal graphs. Proc 34th Conf on Uncertainty in Artificial Intelligence, p.339-349.
- Zhang S, Tong HH, Xu JJ, et al., 2019. Graph convolutional networks: a comprehensive review. *Comput Soc Netw*, 6(1):11. <https://doi.org/10.1186/s40649-019-0069-y>
- Zhao GS, Zhuang XT, Wang XM, et al., 2018. Data-driven traditional Chinese medicine clinical herb modeling and herb pair recommendation. Proc 7th Int Conf on Digital Home, p.160-166. <https://doi.org/10.1109/ICDH.2018.00037>
- Zhao W, Lu WK, Li ZY, et al., 2022. TCM herbal prescription recommendation model based on multi-graph convolutional network. *J Ethnopharmacol*, 297:115109. <https://doi.org/10.1016/j.jep.2022.115109>
- Zheng CP, Fan XL, Wang C, et al., 2020. GMAN: a graph multi-attention network for traffic prediction. Proc 34th AAAI Conf on Artificial Intelligence, p.1234-1241. <https://doi.org/10.1609/aaai.v34i01.5477>
- Zhou WA, Yang K, Zeng JY, et al., 2021. FordNet: recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacol Res*, 173:105752. <https://doi.org/10.1016/j.phrs.2021.105752>
- Zitnik M, Agrawal M, Leskovec J, 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457-i466. <https://doi.org/10.1093/bioinformatics/bty294>

Appendix: TCMKG

We list the main data sources of TCMKG entities as follows:

Herb: *Chinese Materia Medica* (《中华本草》) and *Dictionary of Traditional Chinese Medicine* (《中药大辞典》).

Prescription: *Dictionary of Traditional Chinese Medicine Prescriptions* (《中医方剂大辞典》).

Disease: *Clinic Terminology of Traditional Chinese Medical Diagnosis and Treatment—Diseases* (《中医临床诊疗术语——疾病部分》).

Syndrome: *Clinic Terminology of Traditional Chinese Medical Diagnosis and Treatment—Syndromes* (《中医临床诊疗术语——证候部分》).

Symptom: *Traditional Chinese Medical Symptom Differential and Diagnosis* (《中医症状鉴别诊断学》).

Entities like property, flavor, tropism, function, and pieces are extracted from the description context of the main entity.