A KG-Enhanced Multi-Graph Neural Network for Attentive Herb Recommendation

Yuanyuan Jin, Wendi Ji, Wei Zhang, Xiangnan He, Xinyu Wang, and Xiaoling Wang

Abstract—Traditional Chinese Medicine (TCM) has the longest clinical history in Asia and contributes a lot to health maintenance worldwide. An essential step during the TCM diagnostic process is syndrome induction, which comprehensively analyzes the symptoms and generates an overall summary of the symptoms. Given a set of symptoms, the existing herb recommenders aim to generate the corresponding herbs as a treatment by inducing the implicit syndrome representations based on TCM prescriptions. As different symptoms have various importance during the comprehensive consideration, we argue that treating the co-occurred symptoms equally to do syndrome induction in the previous studies will lead to the coarse-grained syndrome representation. In this paper, we bring the attention mechanism to model the syndrome induction process. Given a set of symptoms, we leverage an attention network to discriminate the symptom importance and adaptively fuse the symptom embeddings. Besides, we introduce a TCM knowledge graph to enrich the input corpus and improve the quality of representation learning. Further, we build a KG-enhanced Multi-Graph Neural Network architecture, which performs the attentive propagation to combine node feature and graph structural information. Extensive experimental results on two TCM data sets show that our proposed model has the outstanding performance over the state-of-the-arts.

Index Terms—Herb recommendation, graph neural network, knowledge graph, attention mechanism

1 Introduction

Traditional Chinese Medicine (TCM) is an effective clinical tool employed by Asia people for thousands of years, and contributes a lot to health maintenance for people around the world [1]. The TCM academic system is based on the holism philosophy, in which the human body interacts with natural environments as a whole[2]. Fig. 1 takes the classic composition *Effective Integration Decoction* as an example to demonstrate the three-step clinical practice in TCM: (1) *Symptom Collection*. First the doctor observes the patient's symptoms. Here sc includes "hypochondriac pain", "swallowing bitterness", "red and dry tongue" and "small and weak pulse". (2) *Syndrome Induction*. Relevant syndromes are summarized through an overall analysis of symptoms. In this instance, the critical syndrome in the solid circle is "yin deficiency of both liver and kidney" . As "stagnation of liver qi" can also lead to "hypochondriac pain" and "small and weak pulse", there exists another syndrome "stagnation of liver qi", shown in a dotted circle. (3) *Treatment Determination*. The doctor organizes multiple herbs into a remedy to cure the syndromes. The herb compatibility is also taken into consideration in this step. As the example shows, the composition hc consists of "szechwan chinaberry fruit", "dwarf lilyturf tuber", "coastal glehnia root" and "dried rehmannia root". Notably, although these herbs cure the symptoms as a whole, every herb has its main functions and complements each other. Back to Fig. 1, the same color indicates that there are strong correlations between the symptoms and herbs. Taking the blue color as an example, the key functions of herb "dwarf lilyturf tuber" and "coastal glehnia root" are both to relieve the symptoms "swallowing bitterness", "red and dry tongue". From the above instance, we can observe that syndrome induction serves as a bridge to connect symptoms and herbs. However, due to the ambiguity attribute of TCM academic foundation, each doctor has the space to develop his own experiences. Facing the same symptom set, different syndromes may be induced by various doctors. As such, there exist no standard syndromes.

The key difference between the traditional item recommendation task and herb recommendation task is that item recommendation is mainly founded on a single user, while recommending herbs needs to consider the co-occurred symptoms simultaneously, as shown in Fig. 1. Because no standard syndromes can act as labels, existing herb recommendation studies [4], [5], [6], [7] treat the syndrome as latent topics. In particular, they only capture the relation between a single symptom and a single herb via the intermediate latent topics, while the symptom set information is omitted. Recently, GNNs are introduced to tackle the herb recommendation task. [8] takes the implicit syndrome induction into account, by leveraging an MLP to fuse multiple symptom embeddings into the syndrome representation. Although [8]

- *Yuanyuan Jin, Wendi Ji, Wei Zhang, and Xinyu Wang are with the Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai 200062, China. E-mail: 2547124038@qq.com, {wendyg8886, zhangwei.thu2011}@gmail.com, xinyuwang@stu.ecnu.edu.cn.*
- *Xiangnan He is with the School of Data Science, University of Science and Technology of China, Hefei 230052, China. E-mail: xiangnanhe@gmail.com.*
- *Xiaoling Wang is with the Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai 200062, China, and also with the Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China. E-mail: xlwang@cs.ecnu.edu.cn.*

Manuscript received 13 Apr. 2021; revised 22 Sept. 2021; accepted 22 Sept. 2021. Date of publication 27 Sept. 2021; date of current version 7 Oct. 2022.
This work was supported in part by the NSFC under Grants 61972155, 61702190, 61972372, 62072182, and U19A2079, in part by the Science and Technology Commission of Shanghai Municipality under Grant 20DZ1100300, and in part by the Open Project Fund from Shenzhen Institute of Artificial Intelligence and Robotics for Society.
(Corresponding author: Xiaoling Wang.)
Digital Object Identifier no. 10.1109/TCBB.2021.3115489

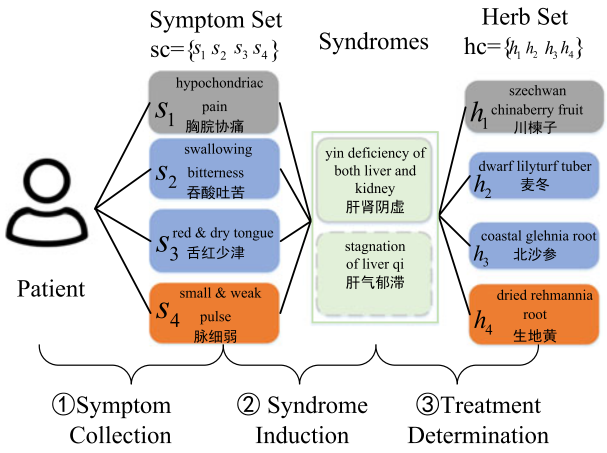

Symptom Set
sc={s_1 s_2 s_3 s_4}
Syndromes
Herb Set
hc={h_1 h_2 h_3 h_4}
s_1 hypochondriac pain 胸脘协痛
s_2 swallowing bitterness 吞酸吐苦
s_3 red & dry tongue 舌红少津
s_4 small & weak pulse 脉细弱
Patient
yin deficiency of both liver and kidney 肝肾阴虚
stagnation of liver qi 肝气郁滞
h_1 szechwan chinaberry fruit 川楝子
h_2 dwarf lilyturf tuber 麦冬
h_3 coastal glehnia root 北沙参
h_4 dried rehmannia root 生地黄
①Symptom Collection
② Syndrome Induction
③Treatment Determination

Fig. 1. An TCM prescription example “Effective Integration Decoction”. The same colors indicate the strongly correlated symptoms and herbs.

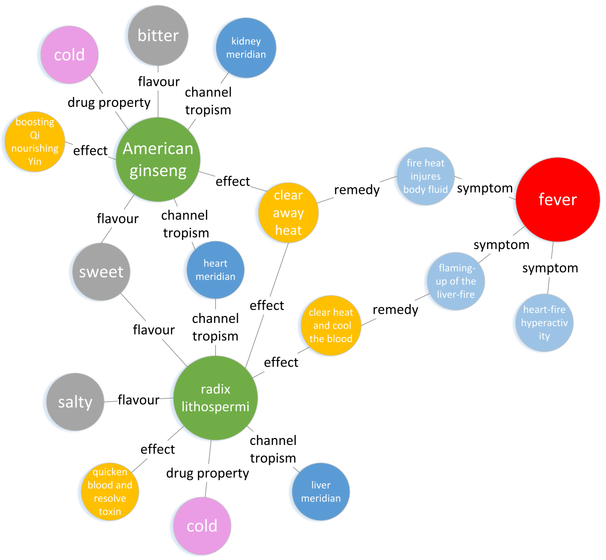


Fig. 2. The schematic diagram of TCM KG [3].

preserves the symptom set information, it treats each symptom equally when aggregating the symptom embeddings together and generates one syndrome representation for all herbs. As such, the main functions of each herb, *a.k.a.* the underlying symptom-herb mapping depicted in Fig. 1, is erased.

In this paper, we argue that for each herb, it is critical to discriminate its different contribution degrees on symptoms to capture the fine-grained symptom-herb mapping. Besides, it is necessary to go beyond prescriptions and take more side information into consideration. In this way, the ambiguity level of the TCM recommendation system can be degraded. To achieve the above goals, we devise an approach, *Syndrome-aware KG-enhanced Attentive Multi-Graph Neural Network* (KG-ASMGNN), which suggests herbs with their mapping to the given symptoms. Technically, in the prediction layer, we leverage an attention-based Multi-Layer Perceptron (MLP) to capture the underlying herb-symptom mapping. For each herb, the attention mechanism helps discriminate its different degrees of contribution on symptoms. After the symptoms are assigned different weights, the MLP adaptively aggregates the symptom embeddings into an individual syndrome representation for every single herb. Through this way, the fine-grained clinical practice of TCM doctors could be depicted. Despite the symptom-herb interactions in TCM prescriptions, the symptom-herb mappings also hide in the attributes of symptoms and herbs. Inspired by [9], [10], [11], [12], [13], we introduce a TCM knowledge graph constructed by [3] (as shown in Fig. 2) to enrich the semantic relations among symptoms and herbs and further enhance embedding learning quality. The knowledge graph is a semantic network consists of triplets. For instance, the triple (American ginseng, effect, clear away heat) means that the herb “American ginseng” can clear away heat. In this way, the embedding layer can obtain symptom and herb embeddings encoding both graph structure and node features. In particular, we extract the side information from the TCM-KG as node features of symptoms and herbs in advance. Following [8], besides the bipartite symptom-herb graph, we also add the synergy graphs, including symptom-symptom and herb-herb graphs extracted from prescriptions. As for the GNN design, we operate the Graph Attention Networks on the above multiple graphs to encode the graph structure information. By employing attention-based propagation to discriminate the importance of neighbors, we can obtain representations with better quality. Finally, we fuse the node feature embeddings and the multi-graph structure embeddings to generate the final symptom and herb representations. The experiments are conducted on two TCM datasets, validating the efficacy of the KG-ASMGNN model and the rationality of every single module. The key contributions of this work are as follows:

- We emphasize the significance of considering the fine-grained symptom-herb mappings when conducting syndrome induction for herb recommendation.
- We propose KG-ASMGNN, which benefits from the attention mechanism and MLP to depict the underlying herb-to-symptom mapping in fusion modeling. In addition, we also incorporate KG-based node features and apply attentive propagation on multi-graphs to enhance the representation learning quality.
- The experimental results based on two TCM datasets demonstrate the proposed method’s effectiveness.

A preliminary version of this article is published in [8]. In this work, we extend the previous research in two aspects. First, we introduce the side information from a TCM knowledge graph to incorporate graph structure and node features. Further we employ graph attention networks to discriminate the importance of different neighbors during propagation. Second, as for syndrome induction, the attention mechanism is also employed to model the underlying herb-symptom mapping, thereby making a more detailed induction from symptoms. The extensive results demonstrate the effectiveness of integrating KG-based node features and the attention mechanism.

2 Related Work

2.1 Herb Recommendation

Topic Model Based Herb Recommendation. Topic models aim to regard the TCM prescriptions as documents and the contained herbs and symptoms as words. Wang *et al.* [14] hire an asymmetric probabilistic generation process to consider symptoms, herbs, and diseases together. Ma *et al.* [15] devise a "symptom-syndrome" framework to capture relations between symptoms and implicit syndrome topics. Ji *et al.* [5] adopt "pathogenesis" as the latent topics. Wang *et al.* [3] and Chen *et al.* [16] bring TCM knowledge into topic models to extract the herb compatibility patterns. To mimic the generative process of TCM prescriptions, Yao *et al.* [4] regard TCM concepts like "syndrome", "treatment", and "herb roles" as latent topics. However, the above topic models suffer from the data sparsity problem [6]. Further, they fail to fully explore the complex correlations among various TCM entities.

Graph Based Herb Recommendation. Graph representation learning based herb recommendation focuses on encoding the TCM graph into a low-dimensional space and then suggests herbs based on the learned embeddings. Li *et al.* [17] employ the BRNN [18] to represent herb words for treatment complement. [6], [7] leverage a meta-path based autoencoder to process the TCM heterogeneous information network. Li *et al.* [19] implement a multi-label classification method with the attentive Seq2Seq [20] module, so as to automatically generate prescriptions. However, the applied deep learning techniques in the above graph-based models are initially operated on the euclidean space, and not equipped with the reasoning ability on the non-euclidean graph data.

2.2 GNNs-Based Recommender Systems

GNNs are founded on the information aggregation mechanism, integrating both edge structures and node features simultaneously. GNNs are employed to various kinds of graphs: 1) *User-item Interaction Graphs*: more generally, the bipartite graph structure. Berg *et al.* [21] devise an autoencoder based graph convolutional matrix completion model. Wang *et al.* [22] utilize GNN to encode the high-order collaborative filtering signals effectively; However, the above methods ignore different node types in bipartite graphs. To this end, [23] proposes ABCGraph, which permits multiple feature dimensions in GCN encoder to learn representations for nodes in two different domains and bipartite structures. [24] designs the Cascade Bipartite Graph Neural Networks, which consists of inter-domain message passing (IDMP) and intra-domain alignment (IDA) to represent the node features within different domains into a single representation. 2) *User Social Networks*: Wu *et al.* [25] and Fan *et al.* [26] utilize GCNs to mimic the social diffusion process for capturing users' preferences in social networks; 3) *User Sequential Behavior Graphs*: Wu *et al.* [27] and Wang *et al.* [28] perform GNN on user behavior sequences to capture complex transition relations among items; 4) *Knowledge Graphs*: Wang *et al.* [29] design the Ripple Network, which passes users' potential interests along edges in a knowledge graph to capture users' preferences. Wang *et al.* [30] capture the multi-hop connectivity in a collaborative knowledge graph, where the core attentive embedding propagation layer adaptively updates each node's representation.

However, the above researches are mostly restricted to the collaborative filtering scenario, where node features and context information are not available [10]. To address this limitation, Ying *et al.* [13] initialize node embeddings by pre-trained image and textual annotation embeddings. Hamilton *et al.* [12] utilize node feature information to produce embeddings for unseen nodes. Wei *et al.* [9] enrich the representation of each node with the topological structure and multi-modal features of its neighbors in the multimedia recommendation. Fu *et al.* [11] design a node content extraction component to encode the node attributes. Wu *et al.* [10] extend the advantages of GNNs to context-aware recommender systems, which take the attributes of users and items as node features and the contexts as edge features.

To notice that, Heterogeneous Information Network (HIN) [31] can also model various objects and their complex relations in recommender systems. Recently, some researchers also incorporate GNNs for HIN-based recommendation. HAN [32] aggregates the meta-path based neighbors with the hierarchical attentions in the GNN propagation. MEIRec [33] also designs meta-path guided heterogeneous GNN layers, which adopt LSTM and CNN as the aggregation function.

3 Problem Definition

Herb recommendation models usually mine the complex relations between entities from the TCM prescriptions. Let $S{=}\{s_1, s_2, \ldots, s_M\}$ and $H{=}\{h_1, h_2, \ldots, h_N\}$ represent all symptoms and herbs, respectively. Each prescription $p = \langle\{s_1, s_2, \ldots\}, \{h_1, h_2, \ldots\}\rangle$, consisting of a symptom set and a herb set. In addition, we introduce a TCM knowledge graph *TCM-KG* [3] to take side information into account. By aligning the symptoms and herbs with the KG node, we extract the node feature $G = \{g_{s_1}, \ldots, g_{s_M}, g_{h_1}, \ldots, g_{h_N}\}$ for all symptoms and herbs from the TCM-KG ahead of time. As for syndrome induction, given a symptom set, first an overall syndrome presentation is adaptively learned for each herb; second the syndrome embedding interacts with the herb embedding to compute the prediction score. Here *symptom set* and *herb set* are denoted by $sc = \{s_1, s_2, \ldots\}$ and $hc = \{h_1, h_2, \ldots\}$, respectively. Thus, each prescription is indicated by $p{=}\langle sc, hc\rangle$.

The herb recommendation task is defined as given a symptom set sc, our approach aims to obtain an N-dimensional probability vector for all candidate herbs. Specifically, the probability vector $\hat{y}_{sc}$ is formulated by $\hat{y}_{sc}{=}f(sc, H; \theta)$, wherein θ denotes the trainable parameters. The input and output are described as:

- Input: Herbs H, Symptoms S, Prescriptions P, pre-extracted node feature embeddings G.
- Output: A learned function $f(sc, H; \theta)$, which suggests the probability vector $\hat{y}_{sc}$ for all candidate herbs in H, given the symptom set sc.

4 Overview of Proposed Model

Our proposed *Syndrome-aware Attentive KG-enhanced Multi-Graph Neural Network* (KG-ASMGNN) approach is shown in Fig. 3. KG-ASMGNN regards a symptom set $sc = \{s_1, s_2, \ldots, s_k\}$, all herbs $H = \{h_1, \ldots, h_N\}$ and the pre-extracted node feature embeddings $G{=}\{g_{s_1}, g_{s_2}, \ldots, g_{s_M}, g_{h_1}, g_{h_2}, \ldots, g_{h_N}\}$ as input, and outputs the predicted probability vector $\hat{y}_{sc}$ in dimension $|H|$. Each dimension i in $\hat{y}_{sc}$ denotes the probability that h_i is suitable to treat sc. In particular, it

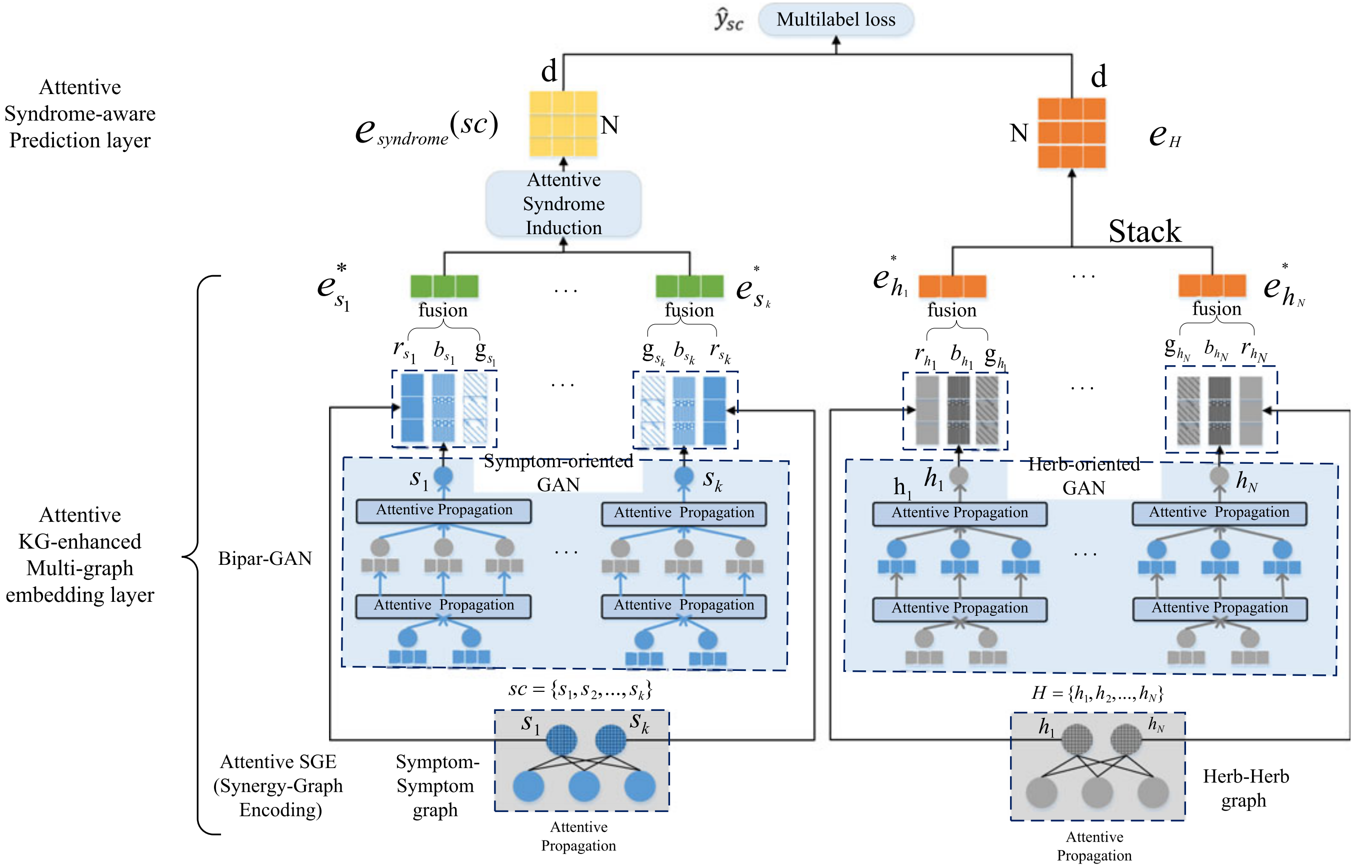

$\hat{y}_{sc}$
Multilabel loss
d
d
Attentive Syndrome-aware Prediction layer
$e_{syndrome}(sc)$
N
N
e_H
Attentive Syndrome Induction
Stack
$e^*_{s_1}$
fusion
$e^*_{s_k}$
$e^*_{h_1}$
fusion
$e^*_{h_N}$
r_{s_1} b_{s_1} g_{s_1}
g_{s_k} b_{s_k} r_{s_k}
r_{h_1} b_{h_1} g_{h_1}
g_{h_N} b_{h_N} r_{h_N}
Symptom-oriented GAN
Herb-oriented GAN
s_1
s_k
h_1
h_N
Attentive Propagation
Attentive KG-enhanced Multi-graph embedding layer
Bipar-GAN
$sc = \{s_1, s_2, ..., s_k\}$
$H = \{h_1, h_2, ..., h_N\}$
Attentive SGE (Synergy-Graph Encoding)
Symptom-Symptom graph
Herb-Herb graph
Attentive Propagation

Fig. 3. An illustration of our proposed model. Symptom nodes are in blue and herb nodes are in gray, and the nodes with oblique lines are target nodes.

includes two parts: Attentive KG-enhanced Multi-Graph Embedding Layer and Attentive Syndrome-aware Prediction Layer.

Attentive KG-Enhanced Multi-Graph Embedding Layer. This layer focuses on obtaining comprehensive representations for all herbs H and all symptoms S. To integrate both topological structure and node features, there are three components: for the graph structure information, we use the Bipartite Graph Attention Network (Bipar-GAN) and Attentive Synergy Graph Encoding (ASGE) modules following [8]; for the node feature information, we processed the *TCM-KG* in advance to obtain the node feature representations for symptoms and herbs. Technically, the motivation of the Bipar-GAN module is that there exists a natural difference between different types of nodes, symptoms and herbs. Thus Bipar-GAN split the unified model into type-aware parts, symptom-oriented GAN and herb-oriented GAN, to develop the model flexibility. To this step, we obtain the herb embedding b_h and symptom embedding b_s. Next, considering the strong synergy relations in symptom pairs and herb pairs, the ASGE module performs GANs on the symptom-symptom graph and herb-herb graph, respectively, in order to distill the synergy embedding r_s and r_h. Up to now, the Bipar-GAN and ASGE modules have captured the graph topology information in prescriptions. As for the node feature distillation, we extracted node embedding g_s and g_h from the *TCM-KG* as node features. Through this manner, the semantic relations linked by node attributes in the knowledge graph can be captured to enrich the node representations. Finally, given each symptom (herb), three types of embeddings b, r and g generated by the Bipar-GAN, ASGE and *TCM-KG* modules are aggregated into the representation e^*.

Attentive Syndrome-Aware Prediction Layer. As the result of the syndrome induction guides the composition principles of herbs, it is necessary to mimic the fine-grained syndrome determination process. As shown in Fig. 1, each herb has its main target symptoms, and the underlying herb-symptom mappings need to be handled carefully. To achieve the above goal, we employ the attention mechanism to learn the beneath symptom-herb mappings automatically. In particular, for each herb, all symptom embeddings in sc are fed into an attention network to be distributed with different weights. Finally, for all herbs in H, we obtain the syndrome representation matrix $e_{syndrome}(sc)$ in dimension $N \times d$, where d is the herb embedding dimension. The syndrome embedding matrix $e_{syndrome}(sc)$ later interacts with e_H to compute $\hat{y}_{sc}$, the probability vector for all herbs H. We adopt a multi-label loss function to train the proposed model. All notations in this paper are included in Table 1.

5 Methodologies

5.1 Bipartite Graph Attention Network

Although recent studies have verified the convincing performance of GNNs on the user-item graph [22], they treat different types of nodes equally and omit the natural difference between various types of nodes in the bipartite user-item

TABLE 1
Summary of All Notations

e_h, e_s	initial embeddings for herbs, symptoms
S, H	symptom collection and herb collection
SC, HC	collection of symptom sets and collection of herb sets
P	prescription collection
N_s, N_h	neighborhood of symptom, herb on the bipartite graph
SH	symptom-herb graph
SS, HH	symptom-symptom graph and herb-herb graph
x_s, x_h	threshold for constructing SS and HH
N_s^{SS}, N_h^{HH}	neighborhood of symptom on SS neighborhood of herb on HH
T_s^k, T_h^k	message construction function for symptom, herb at kth Bipar-GAN layer
W_s^k, W_h^k	message aggregation function for symptom, herb at kth Bipar-GAN layer
V_s, V_h	aggregation function for symptom on SS aggregation function for herb on HH
$b_{N_s}^k, b_{N_h}^k$	symptom, herb neighborhood embedding at kth Bipar-GAN layer
b_s^k, b_h^k	symptom, herb output embeddings at kth Bipar-GAN layer
r_s, r_h	symptom output embeddings on SS herb output embeddings on HH
g_s, g_h	symptom embeddings from TCM-KG herb embeddings from TCM-KG
M_s, M_h	the transformation matrix for g_s the transformation matrix for g_h
e_h^*, e_s^*	herb, symptom final embedding after fusion
W^{mlp}, b^{mlp}	the MLP weight matrix and bias parameter
$e_{syndrome}(sc, h)$	the implicit syndrome embedding of symptom set sc given herb h
$e_{syndrome}(sc)$	the induced syndrome embedding matrix for symptom set sc given all herbs H
$\hat{y}(sc)$	the predicted probability vector for sc

graph. Different types of nodes share the common aggregation and transformation functions, which may limit the model flexibility and hurt the embedding expressiveness to some degree. Thus, we continue to utilize Bipar-GAN, shown in Fig. 4, which splits the unified GNN framework into two type-aware parts, *Symptom-oriented GAN* and *Herb-oriented GAN*. These two modules share the same symptom-herb graph structure but utilize different aggregation and transformation functions.

5.1.1 Symptom-Herb Graph Construction

For instance, in a TCM prescription p=$\langle sc$=$\{s_1, s_2, \ldots, s_k\}$, hc =$\{h_1, h_2, \ldots, h_m\}\rangle$, the symptoms and herbs are correlated to each other. $\{(s_1, h_1), \ldots, (s_1, h_m), \ldots, (s_k, h_1), \ldots, (s_k, h_m)\}$ denote graph edges. Here the symptom-herb graph is constructed by,

$$SH_{s,h}, SH_{h,s} = \begin{cases} 1, & \text{if } (s, h) \text{ co-occur in prescriptions;} \\ 0, & \text{otherwise,} \end{cases}$$

wherein SH represents the symptom-herb graph.

5.1.2 Message Construction

To pass information from neighbors to the target node, we need to determine two steps: 1) how to define the message that each neighbor transfers to the target node; 2) how to summarize messages from multiple neighbors.

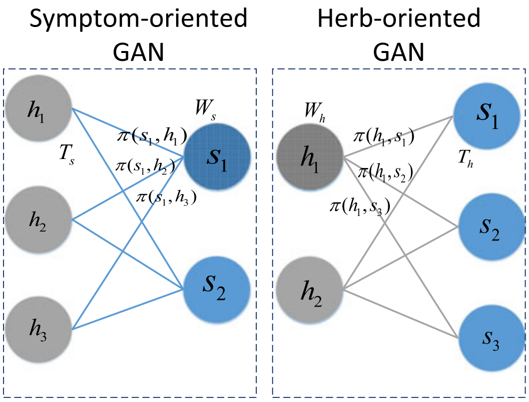


Fig. 4. Bipartite GAN. Blue edges and gray edges indicate different graph convolution functions. The nodes with oblique lines are target nodes.

For symptom s, the message its one-hop neighbor herb h brings to it is formulated as m_h,

$$m_h^0 = e_h \cdot T_s^1, \tag{1}$$

wherein e_h is herb h's initial embedding. T_s^1 is the transformation matrix in the first-layer (symptom). In the bipartite symptom-herb graph, taking symptoms as example, given a symptom, its direct herb neighbors will relieve the symptom in various degrees. Thus, to discriminate the importance of the neighbors, we employ the attention mechanism to summarize the messages from all neighbors as

$$b_{N_s}^0 = tanh\left(\sum_{h \in N_s} \pi(s, h) m_h^0\right), \tag{2}$$

where N_s is the one-hop neighbor set of s. We implement $\pi(s, h)$ as follows:

$$\pi(s, h) = \frac{exp(e_s e_h^T)}{\sum_{h' \in N_s} exp(e_s e_{h'}^T)}. \tag{3}$$

Similarly, for herb h, the summarized massage from one-hop neighbors is denoted by,

$$b_{N_h}^0 = tanh\left(\sum_{s \in N_h} \pi(h, s) m_s^0\right), \tag{4}$$

wherein $m_s^0 = e_s \cdot T_h^1, \pi(h, s) = \frac{exp(e_h e_s^T)}{\sum_{s' \in N_h} exp(e_h e_{s'}^T)}$.

5.1.3 Message Aggregation

To recursively update the embedding for the target node, we utilize the *GraphSAGE Aggregator* [12]. Thus the first-layer symptom representation b_s^1 and herb representation b_h^1 are defined as

$$b_s^1 = tanh(W_s^1 \cdot (e_s || b_{N_s}^0)), \tag{5}$$

$$b_h^1 = tanh(W_h^1 \cdot (e_h || b_{N_h}^0)), \tag{6}$$

wherein $||$ means the concatenation of two vectors. W_s and W_h indicate the aggregation matrices for symptoms and herbs, respectively.

5.1.4 High-Order Propagation

Furthermore, at the kth layer, the representation of herb h is recursively formulated as,

$$b_h^k = tanh(W_h^k \cdot (b_h^{k-1} || b_{N_h}^{k-1})), \tag{7}$$

where the message from kth layer neighbors of h is formulated as,

$$b_{N_h}^{k-1} = tanh\left(\sum_{s \in N_h} \pi(h, s) b_s^{k-1} \cdot T_h^k \right), \tag{8}$$

For symptom s, the definitions are similar,

$$b_s^k = tanh(W_s^k \cdot (b_s^{k-1} || b_{N_s}^{k-1})), \tag{9}$$

where the message propagated within kth layer for s is formulated as

$$b_{N_s}^{k-1} = tanh\left(\sum_{h \in N_s} \pi(s, h) b_h^{k-1} \cdot T_s^k \right). \tag{10}$$

5.2 Attentive Synergy Graph Encoding Layer

5.2.1 Synergy Graphs Construction

The beneath motivation to construct the synergy graphs is that the herb pairs and symptom pairs with strong relations will co-occur in prescriptions frequently. Thus, as for the herb synergy graph, we first count the frequency of all herb-herb pairs in prescriptions: if herb h_m and herb h_n co-occur in the same hc, the frequency of pair (h_m, h_n) is increased by 1. Then we set a threshold x_h, which is a hyperparameter, to filter the entries in the frequency matrix. It is defined as,

$$HH_{h_m,h_n}, HH_{h_n,h_m} = \begin{cases} 1, & \text{if frequency } (h_m, h_n) > x_h; \\ 0, & \text{otherwise,} \end{cases}$$

wherein HH means the herb-herb graph. By obeying the above process, we also built the symptom-symptom graph.

5.2.2 Information Propagation

We utilize an one-layer graph attention network to distinguish the importance of various herb pairs or symptom pairs based on the herb-herb graph HH and symptom-symptom graph SS, respectively

$$\begin{aligned} r_s &= tanh\left(\sum_{k \in N_s^{SS}} \pi(s, k) e_k \cdot V_s \right), \\ r_h &= tanh\left(\sum_{q \in N_h^{HH}} \pi(h, q) e_q \cdot V_h \right), \end{aligned} \tag{11}$$

where e_k and e_q are initial embeddings of symptom k and herb q respectively. N_s^{SS} denotes the neighbor set of s in SS. N_h^{HH} represent the neighbor set of h in HH. V_s and V_h are weight parameters for SS and HH, respectively. $\pi(s, k)$ and $\pi(h, q)$ are formulated as follows: $\pi(s, k) = \frac{exp(e_s e_k^T)}{\sum_{k' \in N_s^{SS}} exp(e_s e_{k'}^T)}$, $\pi(h, q) = \frac{exp(e_h e_q^T)}{\sum_{q' \in N_h^{HH}} exp(e_h e_{q'}^T)}$.

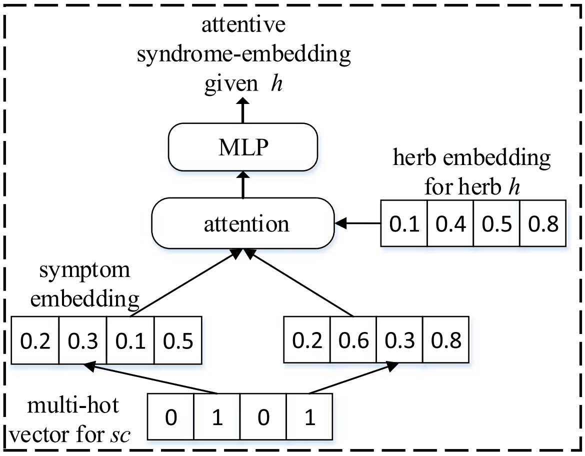


Fig. 5. The attention network for syndrome induction.

5.3 KG-Enhanced Information Fusion

However, the above two types of embeddings only encode the topological structure information and ignore node features. To address this limitation, we introduce the *TCM-KG* [3] to enhance the node embedding quality. The entity types in the *TCM-KG* include herbs, symptoms, herbal properties, efficiencies, pathogeny *et al.* The relation set consists of “hasChemical”, “treat”, “channel tropism”, “hasNature” *et al.* We extract 50-dimension symptom embedding g_s and herb embeddings g_h through HIN2vec [34] method based on the TCM-KG, as the pre-extracted node features. Following [9], the above three types of embeddings are merged by:

$$\begin{aligned} e_s^* &= tanh(b_s + r_s + g_s M_s), \\ e_h^* &= tanh(b_h + r_h + g_h M_h), \end{aligned} \tag{12}$$

wherein M_s is the transformation matrix to project g_s into the same dimension with r_s and b_s. M_h is the transformation matrix to project g_h into the same dimension with r_h and b_h. e_s^* and e_h^* are the integrated embeddings for s and h, respectively.

5.4 Attentive Syndrome Induction

Based on the herbal compatibility in TCM theory, a set of herbs often cooperate with others as a remedy. Every herb in a remedy has its primary roles. For instance, in Fig. 1, “szechwan chinaberry fruit” mainly treats the symptom “hypochondriac pain”. “dried rehmannia root” focuses more on the symptom “small and weak pulse”. To distinguish the different contribution degrees of each herb on multiple symptoms, we propose an attention network to depict the fine-grained syndrome induction, which models the underlying herb-symptom mapping and adaptively obtain an implicit syndrome representation for each herb. In Fig. 5, for the symptom set sc, given each candidate herb h, the attention network generates an adaptive syndrome embedding for h. Similar to [8], first we obtain the symptom set embedding matrix $e_{sc} \in R^{|sc| \times d}$ for sc, and the herb embedding $e_h \in R^{1 \times d}$. Second, the attention weights are defined as a softmax operation on the inner product of e_{sc} and e_h,

$$a_h = Softmax(e_{sc} e_h^T). \tag{13}$$

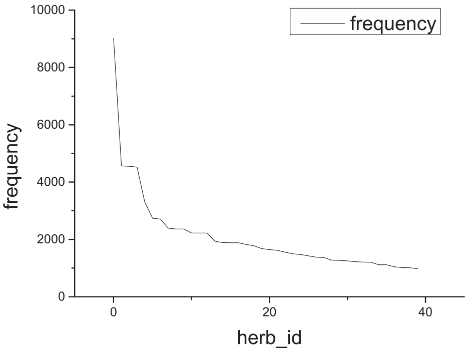


Fig. 6. Frequency distribution for the top 40 most frequent herbs.

Finally, considering the complexity of syndrome induction, we feed the weighted sum of symptom embeddings in a symptom set into an MLP to obtain the syndrome embedding for the herb h,

$$e_{syndrome}(sc,h) = ReLU(\mathbf{W}^{mlp} \cdot Sum(e_{sc} \odot a_h^{'}) + \mathbf{b}^{mlp}), \tag{14}$$

wherein $a_h^{'}$ is the repeated a_h in dimension $|sc| \times d$, $\odot$ indicates the element-wise multiplication and $e_{syndrome}(sc,h)$ indicates the syndrome embedding of sc for h.

5.5 Training and Inference

To be consistent with [8], we also formulate the herb recommendation problem as a multi-label classification task with label imbalance shown in Fig. 6. Specifically, we adopt the objective function defined in Eq. (15) to address the above problems, where $e_H \in R^{|H| \times d}$ denotes the learned embedding matrix of all herbs H, $e_{syndrome}(sc) \in R^{|H| \times d}$ is the stacked $e_{syndrome}(sc,h)$ vectors for all herbs H and the $rowsum$ operation sums the matrix entries along the row axis.

$$\begin{gathered} Loss = \arg\min_{\theta} \sum_{(sc,hc^{'}) \in P} WMSE(hc^{'}, f(sc,H)) + \lambda_{\Theta} ||\Theta||_2^2 \\ f(sc,H) = sigmoid(rowsum(e_{syndrome}(sc) \odot e_H)). \end{gathered} \tag{15}$$

For the symptom set sc, a multi-hot vector $hc^{'}$ in dimension $|H|$ denotes the true herb set hc. $f(sc,H)$ is the output probability vector for all herbs. WMSE [35] *(weighted mean square loss)* is formulated as,

$$WMSE(hc^{'}, f(sc,H)) = \sum_{i=1}^{|H|} w_i (hc_i^{'} - f(sc,H)_i)^2, \tag{16}$$

the dimensions of $hc^{'}$ and $f(sc,H)$ equal $|H|$. w_i for each herb i aims to tackle the label imbalance problem, defined by

$$w_i = \frac{max_k freq(k)}{freq(i)}, \tag{17}$$

where $freq(i)$ denote herb i's frequency in prescriptions. The weight design is to balance the contribution of herbs with different frequencies. Specifically, the more frequently herb i appears, the smaller its weight is. We adopt Adam [36] to do optimization and update the model parameters in a mini-batch fashion.

Inference. To be consistent with the setting in [35], given sc, we regard the top k herbs with the highest probabilities in $f(sc,H)$ as the recommended herb set.

6 Experiments

We conduct experiments on the public TCM dataset in [4] and the lung cancer prescriptions used in [5] from Longhua Hospital Shanghai University of Traditional Chinese Medicine. The experiments are designed to answer the following research questions:

RQ1: How does KG-ASMGNN perform over state-of-the-art herb recommendation approaches and GNNs-based recommendation models?

RQ2: How do modules (AKG-MGE and ASI) influence KG-ASMGNN?

RQ3: How do different hyper-parameter settings affect our model performance?

RQ4: Can our proposed KG-ASMGNN provide reasonable herb recommendations with herb-symptom mapping?

6.1 Dataset

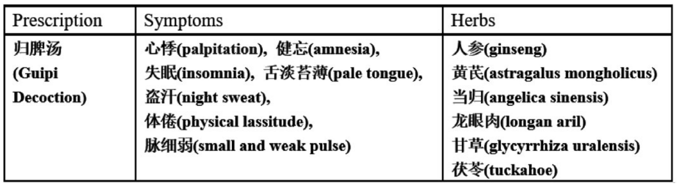
We utilize the processed data set by [3] based on the public TCM data set [4] and the lung cancer prescriptions in [5] to conduct experiments. Fig. 7 demonstrates a prescription example including multiple symptoms and the corresponding herbs. The data set is introduced in Table 2. In addition, we also obtain the TCM KG from [3] to extract the node feature embeddings for symptoms and herbs.

6.2 Evaluation Metrics

We leverage three commonly used evaluation metrics [3], [22]: Precision at rank K (Precision@K), Recall at rank K (Recall@K), and Normalized Discounted Cumulative Gain at rank K (NDCG@K). Specifically, we adopt $Precision@5$ to determine the optimal parameters. We cut off the ranked list at length 20 for all three measures.

6.3 Baselines

- MP: The most frequent herbs in the train dataset are recommended.



Prescription	Symptoms	Herbs

**归脾汤
(Guipi
Decoction)**
	**心悸(palpitation)，健忘(amnesia),
	失眠(insomnia)，舌淡苔薄(pale tongue),
	盗汗(night sweat),
	体倦(physical lassitude),
	脉细弱(small and weak pulse)**
		**人参(ginseng)
		黄芪(astragalus mongholicus)
		当归(angelica sinensis)
		龙眼肉(longan aril)
		甘草(glycyrrhiza uralensis)
		茯苓(tuckahoe)**

Fig. 7. The prescription ”Guipi Decoction”.

TABLE 2
Statistics of the Data Sets

Dataset		#prescriptions	#symptoms	#herbs
TCM	All	26,360	360	753
	Train	21,755	360	753
	Valid	1,162	199	468
	Test	3,443	254	558
Lung cancer	All	947	77	336
	Train	761	77	336
	Valid	90	71	175
	Test	96	73	159

Topic Model

- HC-KGETM [3]: the state-of-the-art topic model to tackle herb recommendation. It benefits from the TransE [37] embeddings extracted from a TCM KG, to consider the semantic correlations among entities in TCM prescriptions and knowledge graph simultaneously.

GNNs-Based Models

- GC-MC [21]: It applies GCN [38] to learn the representations for users and items. There is one graph convolution layer, and the hidden dimension equals the embedding size.
- PinSage [13]: It performs GraphSAGE [12] on item-item graph. In our implementation, it is performed on the symptom-herb graph. There are two graph convolution layers, and the hidden dimension equals the embedding size.
- NGCF [22]: It builds a user-item graph to capture the high-order relations between users and items.
- SMGCN [8]: SMGCN is our previously published approach with a multi-graph convolution network. It introduces multiple graphs and exploits GCNs on them to learn the TCM entity embedding. Besides, it aggregates the symptom set to make the syndrome-aware herb recommendation.

Our Proposed Model

- SMGAN: SMGAN is the extended version of our previously published approach SMGCN in [8] with attentive embedding propagation layers. It introduces multiple graphs and exploits Graph Attention Networks on them to learn the TCM entity embedding. Besides, it aggregates the symptom set to make the syndrome-aware herb recommendation.
- KG-ASMGNN: The proposed approach builds a KG-enhanced Attentive Multi-Graph GNN framework to incorporate both graph structure and node feature information and provide herb recommendations with the underlying herb-symptom mapping. It extracts the side information from a TCM KG as node features. In the prediction layer, for each herb, we leverage an attention network to generate the adaptive syndrome embedding of the symptom set. Notably, after removing the KG-based node feature embedding in the embedding layer and attention mechanism in the prediction layer, our KG-ASMGNN model degrades to the simpler SMGAN model.

TABLE 3
Optimal Values of Parameters

Dataset	Methods	Best parameter settings
TCM	HC-KGETM	$\alpha = 0.05$ $\beta = 0.01$ $lambda = 0.5$
	GC-MC (w/SI)	lr = 9e-4 dropout = 0.2 λ = 1e-6
	PinSage (w/SI)	lr = 5e-4 dropout = 0.1 λ = 1e-3
	NGCF (w/SI)	lr = 3e-3 dropout = 0.1 λ = 1e-5
	SMGCN	lr = 1e-4 dropout = 0.2 λ = 1e-2
	SMGAN	lr = 3e-4 dropout = 0.2 λ = 0
	KG-ASMGNN	lr = 1e-4 dropout = 0.3 λ = 0
Lung cancer	HC-KGETM	$\alpha = 0.05$ $\beta = 0.01$ $lambda = 0.0$
	GC-MC (w/SI)	lr = 5e-3 dropout = 0.0 λ = 1e-4
	PinSage (w/SI)	lr = 3e-2 dropout = 0.0 λ = 1e-3
	NGCF (w/SI)	lr = 1e-2 dropout = 0.1 λ = 1e-3
	SMGCN	lr = 5e-4 dropout = 0.0 λ = 1e-5
	SMGAN	lr = 8e-4 dropout = 0.1 λ = 1e-3
	KG-ASMGNN	lr = 1e-3 dropout = 0.4 λ = 9e-5

6.4 Parameter Settings

We choose Tensorflow as the experimental platform. To notice that, for the fair comparison, we equip the GNN-based methods GC-MC, PinSage and NGCF with the MLP-based SI part and multi-label loss, which are represented by “GC-MC (w/SI)”,“PinSage (w/SI)” and “NGCF (w/SI)”. The parameter setting for HC-KGETM is consistent with [3], which utilizes log-loss and is without the SI part. The model parameters are initialized by Xavier initializer [39], and we choose Adam optimizer [36] with the batch size of 1024. Specifically, We apply the same grid search strategy as [8] for learning rate lr, the regularization coefficient λ_Θ, and the dropout ratio. For GNN baselines, the embedding size and latent dimension equal 64. As for the proposed KG-ASMGNN, the embedding size is 64, and the first output layer is in dimension 128. Table 3 demonstrates the optimal parameter settings. Without explanation, to be consistent with our previously published work [8], the following KG-ASMGNN performance is with 2 GAN layers, the last layer dimension of 256, threshold x_s of 5 and threshold x_h of 40. For HC-KGETM, the KG-based embeddings are extracted through TransE[37] in dimension 100 following [3]. For our proposed KG-ASMGNN, the KG-based node feature embeddings g_s and g_h are extracted by HIN2vec [34] in dimension 50.

6.5 Performance Comparison

6.5.1 Overall Comparison (RQ1)

The empirical results are reported in Table 4. We find that:

- KG-ASMGNN consistently outperforms all the comparative approaches on both datasets. Taking the TCM dataset for example, KG-ASMGNN outperforms HC-KGETM in terms of p@5 by 7.45%, r@5 by 8.02%, and ndcg@5 by 5.97%. Besides, KG-ASMGNN achieves improvements over the second best SMGAN in terms of p@5 by 0.18%, r@5 by 0.36%, and ndcg@5 by 0.56%. We conduct the paired t-tests between KG-ASMGNN and SMGAN, which shows that the improvements of KG-ASMGNN over SMGAN are statistically significant.
- MP performs poorest for all metrics, which indicates that just recommending the top-k most frequent

TABLE 4
The Overall Performance Comparison

dataset	Approaches	p@5	p@10	p@20	r@5	r@10	r@20	ndcg@5	ndcg@10	ndcg@20
TCM	MP	0.2243	0.1665	0.1264	0.1537	0.2283	0.3472	0.3467	0.4127	0.5074
	HC-KGETM	0.2709	0.2154	0.1593	0.1925	0.3044	0.4428	0.3702	0.4488	0.5495
	GC-MC (w/SI)	0.2878	0.2295	0.1681	0.2057	0.3239	0.4666	0.3896	0.4690	0.5705
	PinSage (w/SI)	0.2867	0.2301	0.1671	0.2035	0.3251	0.4649	0.3895	0.4723	0.5711
	NGCF (w/SI)	0.2856	0.2297	0.1675	0.2027	0.3255	0.4675	0.3831	0.4671	0.5680
	SMGCN	0.2878	0.2287	0.1664	0.2070	0.3229	0.4631	0.3917	0.4728	0.5715
	SMGAN	0.2906	0.2307	0.1689	0.2072	0.3249	0.4697	0.3901	0.4691	0.5710
	KG-ASMGNN	**0.2911***	**0.2334***	**0.1692***	**0.2079***	**0.3301***	**0.4698***	**0.3923***	**0.4739***	**0.5734***
	%Improv. to HC-KGETM	7.45%	8.34%	6.20%	8.02%	8.45%	6.09%	5.97%	5.60%	4.35%
	%Improv. to GC-MC (w/SI)	1.15%	1.67%	0.65%	1.06%	1.94%	0.68%	0.68%	1.04%	0.51%
	%Improv. to SMGAN	0.18%	1.17%	0.16%	0.36%	1.61%	0.02%	0.56%	1.04%	0.42%
Lung cancer	MP	0.8563	0.7104	0.5255	0.2222	0.3701	0.5430	0.8809	0.7990	0.9318
	HC-KGETM	0.8710	0.7214	0.5474	0.2259	0.3744	0.5653	0.8904	0.8012	0.9355
	GC-MC (w/SI)	0.8938	0.7646	0.5849	0.2318	0.3974	0.6055	0.9091	0.8281	0.9439
	PinSage (w/SI)	0.8938	0.7563	0.5797	0.2321	0.3928	0.6004	0.9067	0.8230	0.9405
	NGCF (w/SI)	0.8896	0.7740	0.5781	0.2306	0.4020	0.5985	0.9026	0.8345	0.9405
	SMGCN	0.8958	0.7740	0.5839	0.2325	0.4025	0.6037	0.9102	0.8376	0.9452
	SMGAN	0.8958	0.7792	0.5885	0.2324	0.4049	0.6092	0.9125	0.8358	0.9468
	KG-ASMGNN	**0.9104***	**0.7948***	**0.5990***	**0.2365***	**0.4126***	**0.6198***	**0.9191***	**0.8457***	**0.9476***
	%Improv. to HC-KGETM	4.52%	10.18%	9.41%	4.67%	10.20%	9.64%	3.22%	5.55%	1.30%
	%Improv. to GC-MC (w/SI)	1.87%	3.95%	2.40%	2.01%	3.83%	2.36%	1.09%	2.12%	0.40%
	%Improv. to SMGAN	1.63%	2.00%	1.77%	1.75%	1.90%	1.74%	0.72%	1.18%	0.08%

*HC-KGETM Utilizes log-loss but without SI. The Other Models Are with SI and Multi-label Loss. The Second Best Results are Underlined. * Means the Significant Paired t-test Result.*

herbs cannot treat different symptoms appropriately. For the TCM dataset, the second poorest model is HC-KGETM. Some possible reasons are 1) at the prediction phase, the candidate herbs are ranked at the single symptom level, and thus the symptom set information is omitted; 2) for representation learning, it only employs translation-based TransE [37] to process the TCM-KG. As for graph representative learning, the graph neural networks are better than the translation-based methods by explicitly capturing the high-order connectivity.

- GC-MC (w/SI) slightly outperforms NGCF (w/SI) and PinSage (w/SI) in some metrics, which suggests that the multiple graph convolution layers of NGCF (w/SI) may lead to overfitting. As for propagation functions, GC-MC (w/SI) sums the representations of the target node and its neighbors, PinSage (w/SI) concatenates these two embeddings. In NGCF (w/SI), the neighbor's information propagates to the target node includes both its own representation and the element-wise product between the target node and neighbor node. Compared with the concatenation operations and element-wise product, the sum function is more suitable to capture the complex interactions in prescriptions.
- SMGCN performs slightly better than GC-MC (w/SI) in some metrics, which validates that introducing the synergy graphs can enrich the semantic relations. Further, SMGAN is slightly superior to SMGCN in terms of the precision metrics, which verifies that applying attention mechanism in embedding propagation can help improve the recommendation performance. KG-ASMGNN still reaches better performance than SMGAN, which verifies that incorporating both graph structure and node feature can learn a more effective model, and the attentive syndrome representation is more suitable for capturing the underlying herb-symptom mapping and mimic the fine-grained syndrome induction step.

6.5.2 Ablation Analysis (RQ2)

To verify the impact of the KG-based node feature embedding and the attention mechanism, the ablation study is conducted on the two variants of KG-ASMGNN. In particular, we disable the attention mechanism in the prediction layer of KG-ASMGNN, termed KG-SMGAN; we disable the KG-based node feature embedding in the embedding layer, termed ASMGNN. The experimental results are reported in Table 5. The paired t-test between KG-SMGAN and SMGAN shows that the improvements of KG-SMGAN over SMGAN are statistically significant. We have the following observations:

- For both SMGAN and ASMGNN, integrating the KG-based embedding part leads to further improvement, which shows that introducing node feature embedding helps representative learning and herb prediction.
- For KG-SMGAN, adding the attention mechanism in the prediction layer also brings slight improvement. We observe that compared with the equal weights on symptoms, the attention network can mimic the complex syndrome induction process in a more fine-grained level.
- KG-ASMGNN, the combination of KG-based node feature and attention mechanism performs best, showing that capturing the herb-symptom mapping and incorporating both the graph structure and node feature benefits the herb recommendation task.

In addition, we add several baselines to verify the effectiveness of MLP-based syndrome induction in Table 6. PinSage (w/SI) is equipped with MLP to induce syndrome, while the others adopt average pooling. We can find that the proposed MLP-based syndrome induction performs slightly better than others. Besides, the PinSage (w/KG) performs

TABLE 5
Effect of Different Modules on TCM Dataset

Approaches	p@10	r@10	ndcg@10
SMGAN	0.2307	0.3249	0.4691
ASMGNN	0.2312	0.3269	0.4737
KG-SMGAN	0.2328	0.3272	0.4713
KG-ASMGNN	**0.2334**	**0.3301**	**0.4739**

TABLE 6
Effect of Syndrome Induction on TCM Dataset

Approaches	p@10	r@10	ndcg@10
PinSage	0.2272	0.3197	0.4622
PinSage (w/SI)	**0.2301**	**0.3251**	**0.4723**
PinSage (w/SGE)	0.2290	0.3213	0.4656
PinSage (w/KG)	0.2300	0.3249	0.4679
PinSage (w/attention network)	0.2253	0.3181	0.4610

slightly better among the methods with average pooling. In addition, PinSage (w/attention network) shows slightly worse performance than PinSage. The possible reason may be that the embedding layer of PinSage (w/attention network) does not incorporate the rich semantic information from SGE and KG modules, which restricts the representation quality of symptoms and herbs, and hurts the attention network performance to some extent.

6.5.3 Influence of Hyperparameters (RQ3)

- Effect of TCM-KG representation ways

Our TCM-KG [3] is a kind of Heterogeneous Information Network (HIN) containing multiple entities and relations. The HIN-based representative learning approaches and the KG-based representative learning methods like TransE can both be applied to extract the entity embeddings. Here we compare TransE [37] with HIN2vec [34]. TransE models the (head, relation, tail) triples by translation-based constraints, while HIN2vec learns both the node embedding and the meta-path embedding by neural networks. First, we obtain the TransE embeddings with dimension 100. Then we conduct the comparison experiments on SMGAN. Table 7 demonstrates the results, where KG(TransE)-SMGAN means that the TransE embeddings are leveraged in Section 5.3, and KG(HIN2vec)-SMGAN indicates that the HIN2vec embeddings are utilized in Section 5.3. We have the following observations.

- Incorporating TransE or HIN2vec embedding can both improve the recommendation performance, which indicates that considering both graph structure and node feature information helps generate representations with higher quality and further improve the recommendation accuracy.
- The performance of HIN2vec is slightly better than TransE, which implies that HIN-based representative learning outperforms the KG representation learning method on the TCM-KG for our herb recommendation task.

- Effect of different attention network implementations

The attention network is essential in our syndrome induction design. Inspired by the success of memory networks to address collaborative filtering [40], we adopt the memory network [41] to be compared with the attention implementation in Section 5.4. Memory network contains two parts: an external memory and a controller operating on the memory, such as read, write and erase [40]. Similar to Equations (13) and (14), the memory network formulations are as follows,

TABLE 7
Comparison of Different Representation Ways on TCM Dataset

Approaches	p@10	r@10	ndcg@10
SMGAN	0.2307	0.3249	0.4691
KG(TransE)-SMGAN	0.2303	0.3259	0.4708
KG(HIN2vec)-SMGAN	**0.2328**	**0.3272**	**0.4713**

TABLE 8
Comparison of Different Attention Network Implementations on TCM Dataset

Approaches	p@10	r@10	ndcg@10
KG(HIN2vec)-ASMGNN(attention)	**0.2334**	**0.3301**	**0.4739**
KG(HIN2vec)-ASMGNN(memory)	0.2297	0.3244	0.4706
KG(TransE)-ASMGNN(attention)	0.2304	0.3262	0.4710
KG(TransE)-ASMGNN(memory)	0.2302	0.3254	0.4693

$$a_h = Softmax((e_{sc}M_A)(e_hM_B)^T) \tag{18}$$

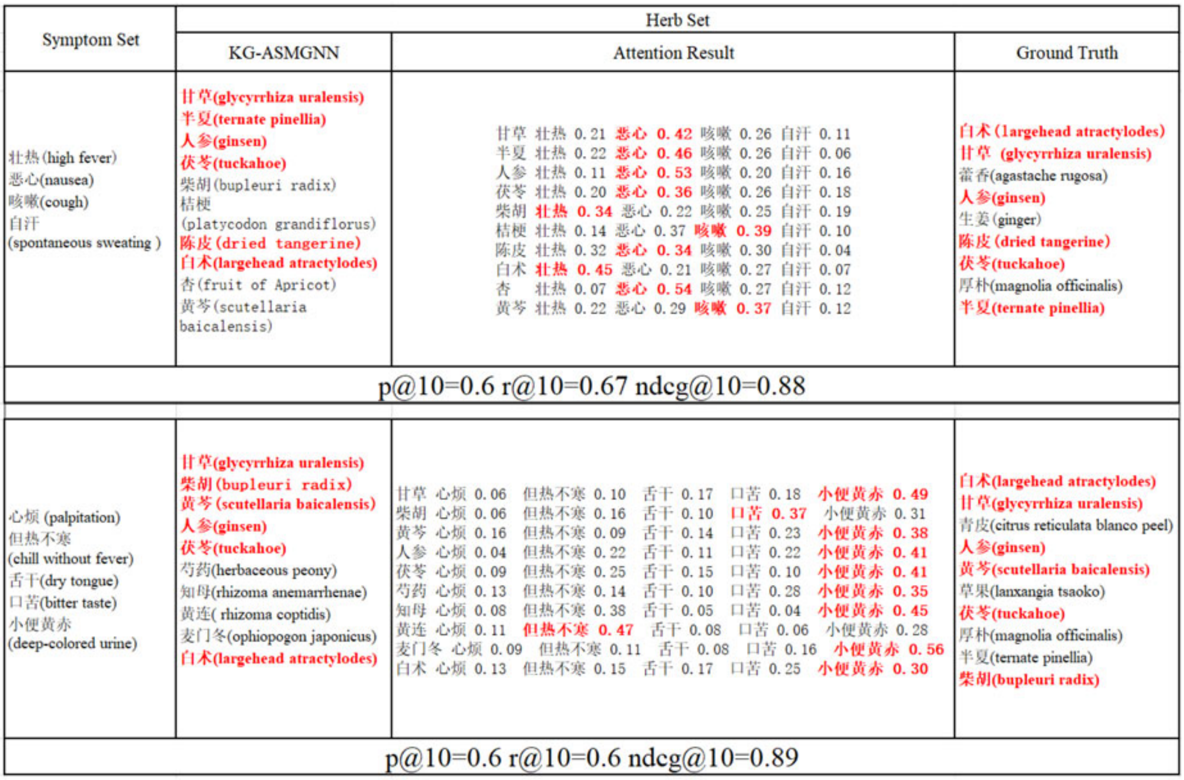
$$e_{syndrome}(sc, h) = ReLU(\mathbf{W}^{mlp} \cdot Sum((e_{sc}M_C) \odot a_h^{'}) + \mathbf{b}^{mlp}), \tag{19}$$

wherein M_A, M_B and M_C are matrices in dimension $d \times d$ to convert the syndrome embedding e_{sc} and herb embedding e_h into memory vectors. Considering that the embedding quality may influence the performance of attention network at the prediction layer, we conduct the comparison experiments with both TransE and HIN2vec node feature embeddings. In Table 8, the suffix ”attention” means the attention implementation in Section 5.4 and suffix ”memory” indicates the implementation defined in Equations (18) and (19). We have the following observations:

- For both the TransE and HIN2vec node feature embeddings, the attention implementation in Section 5.4 achieves slightly better performance than the memory network implementation.
- For the attention implementation in Section 5.4, incorporating HIN2vec feature is slightly superior to TransE.
- From the whole perspective, the combination of HIN2vec method and attention network in 5.4 achieves slightly better performance than other combinations. It seems that the memory vector in the memory network helps little in our herb recommendation task.

6.5.4 Case Study (RQ4)

Here we show some cases to analyze the rationality of our KG-ASMGNN model. In Fig. 8, we select two recommendation examples from the test dataset. In the *Herb Set* column, the hit herbs are marked by the bold red font. The *Attention Result* column demonstrates the learned attention weights in the prediction layer. For instance, in the first sample, as for the herb



Symptom Set	Herb Set		
	KG-ASMGNN	Attention Result	Ground Truth

壮热(high fever)
恶心(nausea)
咳嗽(cough)
自汗
(spontaneous sweating)
	甘草(glycyrrhiza uralensis)
	半夏(ternate pinellia)
	人参(ginsen)
	茯苓(tuckahoe)
	柴胡(bupleuri radix)
	桔梗
	(platycodon grandiflorus)
	陈皮(dried tangerine)
	白术(largehead atractylodes)
	杏(fruit of Apricot)
	黄芩(scutellaria baicalensis)
		甘草 壮热 0.21 恶心 0.42 咳嗽 0.26 自汗 0.11
		半夏 壮热 0.22 恶心 0.46 咳嗽 0.26 自汗 0.06
		人参 壮热 0.11 恶心 0.53 咳嗽 0.20 自汗 0.16
		茯苓 壮热 0.20 恶心 0.36 咳嗽 0.26 自汗 0.18
		柴胡 壮热 0.34 恶心 0.22 咳嗽 0.25 自汗 0.19
		桔梗 壮热 0.14 恶心 0.37 咳嗽 0.39 自汗 0.10
		陈皮 壮热 0.32 恶心 0.34 咳嗽 0.30 自汗 0.04
		白术 壮热 0.45 恶心 0.21 咳嗽 0.27 自汗 0.07
		杏 壮热 0.07 恶心 0.54 咳嗽 0.27 自汗 0.12
		黄芩 壮热 0.22 恶心 0.29 咳嗽 0.37 自汗 0.12
			白术(largehead atractylodes)
			甘草 (glycyrrhiza uralensis)
			藿香(agastache rugosa)
			人参(ginsen)
			生姜(ginger)
			陈皮(dried tangerine)
			茯苓(tuckahoe)
			厚朴(magnolia officinalis)
			半夏(ternate pinellia)

p@10=0.6 r@10=0.67 ndcg@10=0.88			

心烦 (palpitation)
但热不寒
(chill without fever)
舌干(dry tongue)
口苦(bitter taste)
小便黄赤
(deep-colored urine)
	甘草(glycyrrhiza uralensis)
	柴胡(bupleuri radix)
	黄芩(scutellaria baicalensis)
	人参(ginsen)
	茯苓(tuckahoe)
	芍药(herbaceous peony)
	知母(rhizoma anemarrhenae)
	黄连(rhizoma coptidis)
	麦门冬(ophiopogon japonicus)
	白术(largehead atractylodes)
		甘草 心烦 0.06 但热不寒 0.10 舌干 0.17 口苦 0.18 小便黄赤 0.49
		柴胡 心烦 0.06 但热不寒 0.16 舌干 0.10 口苦 0.37 小便黄赤 0.31
		黄芩 心烦 0.16 但热不寒 0.09 舌干 0.14 口苦 0.23 小便黄赤 0.38
		人参 心烦 0.04 但热不寒 0.22 舌干 0.11 口苦 0.22 小便黄赤 0.41
		茯苓 心烦 0.09 但热不寒 0.25 舌干 0.15 口苦 0.10 小便黄赤 0.41
		芍药 心烦 0.13 但热不寒 0.14 舌干 0.10 口苦 0.28 小便黄赤 0.35
		知母 心烦 0.08 但热不寒 0.38 舌干 0.05 口苦 0.04 小便黄赤 0.45
		黄连 心烦 0.11 但热不寒 0.47 舌干 0.08 口苦 0.06 小便黄赤 0.28
		麦门冬 心烦 0.09 但热不寒 0.11 舌干 0.08 口苦 0.16 小便黄赤 0.56
		白术 心烦 0.13 但热不寒 0.15 舌干 0.17 口苦 0.25 小便黄赤 0.30
			白术(largehead atractylodes)
			甘草(glycyrrhiza uralensis)
			青皮(citrus reticulata blanco peel)
			人参(ginsen)
			黄芩(scutellaria baicalensis)
			草果(lanxangia tsaoko)
			茯苓(tuckahoe)
			厚朴(magnolia officinalis)
			半夏(ternate pinellia)
			柴胡(bupleuri radix)

p@10=0.6 r@10=0.6 ndcg@10=0.89			

Fig. 8. The real herb recommendation examples in test dataset.

"bupleuri radix", the corresponding row contains the attention weight of each symptom given "bupleuri radix", which can be treated as the mapping from "bupleuri radix" to the symptoms. We can observe that "bupleuri radix" mainly cure the symptom "high fever" with the highest weight 0.34. Similarly, it can be seen that "platycodon grandiflorus" mainly relieves "cough". "dried tangerine" can lighten "nausea". Though this manner, we can conclude that KG-ASMGNN can provide reasonable herb recommendations with their mapping to symptoms.

7 Conclusion

In this work, we aim to depict the fine-grained syndrome induction in the herb recommendation task. We devise multiple GANs to learn the comprehensive embeddings for symptoms and herbs. Besides, to incorporate the graph structure and node feature information simultaneously, we introduce the node feature embedding extracted from a TCM KG to enhance entity representations. To depict the syndrome induction step, we merge multiple symptom embeddings by an attentive MLP layer, which adaptively generates the syndrome embedding for each herb with the underlying herb-symptom mapping. The experiments conducted on two TCM datasets verify the rationality and effectiveness of our model. In the future, we will consider the herb dosages to offer more practical herb recommendations.

References

[1] F. Cheung, "TCM: Made in China," *Nature*, vol. 480, no. 7378, pp. S82–S83, 2011.

[2] Y. Wang and A. Xu, "Zheng: A systems biology approach to diagnosis and treatments," *Science*, vol. 346, no. 6216, pp. S13–S15, 2014.

[3] X. Wang, Y. Zhang, X. Wang, and J. Chen, "A knowledge graph enhanced topic modeling approach for herb recommendation," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2019, pp. 709–724.

[4] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional chinese medicine prescriptions," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1007–1021, Jun. 2018.

[5] W. Ji, Y. Zhang, X. Wang, and Y. Zhou, "Latent semantic diagnosis in traditional Chinese medicine," *World Wide Web*, vol. 20, no. 5, pp. 1071–1087, 2017.

[6] C. Ruan, J. Ma, Y. Wang, Y. Zhang, and Y. Yang, "Discovering regularities from traditional Chinese medicine prescriptions via bipartite embedding model," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 3346–3352.

[7] C. Ruan, Y. Wang, Y. Zhang, and Y. Yang, "Exploring regularity in traditional Chinese medicine clinical data using heterogeneous weighted networks embedding," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2019, pp. 310–313.

[8] Y. Jin, W. Zhang, X. He, X. Wang, and X. Wang, "Syndrome-aware herb recommendation with multi-graph convolution network," in *Proc. IEEE 36th Int. Conf. Data Eng.*, 2020, pp. 145–156.

[9] Y. Wei, X. Wang, L. Nie, X. He, R. Hong, and T-S. Chua, "MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 1437–1445.

[10] J. Wu, X. He, X. Wang, Q. Wang, W. Chen, J. Lian, and X. Xie, "Graph convolution machine for context-aware recommender system," 2020, *arXiv:2001.11402*.

[11] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, 2020, pp. 2331–2341.

[12] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[13] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 974–983.

[14] S. Wang *et al.*, "A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional Chinese medicine patient records," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2016, pp. 411–418.

[15] J. Ma and Z. Wang, "Discovering syndrome regularities in traditional Chinese medicine clinical by topic model," in *Proc. Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput.*, 2016, pp. 157–162.

[16] X. Chen, C. Ruan, Y. Zhang, and H. Chen, "Heterogeneous information network based clustering for categorizations of traditional Chinese medicine formula," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2018, pp. 839–846.

[17] W. Li and Z. Yang, "Distributed representation for traditional Chinese medicine herb via deep learning models," 2017, *arXiv: 1711.01701*.

[18] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[19] W. Li, Z. Yang, and X. Sun, "Exploration on generating traditional Chinese medicine prescription from symptoms with an end-to-end method," 2018, *arXiv:1801.09030*.

[20] Y. Zhang, M. Yu, N. Li, C. Yu, J. Cui, and D. Yu, "Seq2Seq attentional siamese neural networks for text-dependent speaker verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 6131–6135.

[21] R. v. d. Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," 2017, *arXiv:1706.02263*.

[22] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 165–174.

[23] C. He *et al.*, "Adversarial representation learning on large-scale bipartite graphs," 2019, *arXiv:1906.11994*.

[24] C. He *et al.*, "Cascade-BGNN: Toward efficient self-supervised representation learning on large-scale bipartite graphs," 2019, *arXiv: 1906.11994*.

[25] L. Wu, P. Sun, R. Hong, Y. Fu, X. Wang, and M. Wang, "SocialGCN: An efficient graph convolutional network based model for social recommendation," 2018, *arXiv:1811.02815*.

[26] W. Fan *et al.*, "Graph neural networks for social recommendation," in *Proc. World Wide Web Conf.*, 2019, pp. 417–426.

[27] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 346–353.

[28] W. Wang *et al.*, "Beyond clicks: Modeling multi-relational item graph for session-based target behavior prediction," in *Proc. Web Conf.*, 2020, pp. 3056–3062.

[29] H. Wang *et al.*, "RippleNet: Propagating user preferences on the knowledge graph for recommender systems," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 417–426.

[30] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 950–958.

[31] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 1, pp. 17–37, Jan. 2017.

[32] X. Wang *et al.*, "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.

[33] S. Fan, J. Zhu, X. Han, C. Shi, L. Hu, B. Ma, and Y. Li, "Metapath-guided heterogeneous graph neural network for intent recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2478–2486.

[34] T.-Y. Fu, W.-C. Lee, and Z. Lei, "HIN2Vec: Explore meta-paths in heterogeneous information networks for representation learning," in *Proc. 2017 ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 1797–1806.

[35] H. Hu and X. He, "Sets2Sets: Learning from sequential sets with neural networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1491–1499.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.

[37] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 2787–2795.

[38] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Representations*, 2017.

[39] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[40] T. Ebesu, B. Shen, and Y. Fang, "Collaborative memory network for recommendation systems," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2018, pp. 515–524.

[41] S. Sukhbaatar *et al.*, "End-to-end memory networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2440–2448.

Yuanyuan Jin is currently working toward the PhD degree with the School of Software Engineering, East China Normal University. Her research interests include recommender system and data engineering.

Wendi Ji is currently a postdoctoral researcher with East China Normal University. Her research interests include deep learning, recommender system, and data mining.

Wei Zhang received the PhD degree. He is currently an associate researcher. His main research focuses on data mining.

Xiangnan He is currently a professor with the University of Science and Technology of China. His research interests include information retrieval, data mining, and multi-media analysis.

Xinyu Wang received the BSc degree in software engineering from East China Normal University, China, in 2020. His research interests include recommender system and data mining.

Xiaoling Wang research interests include knowledge graph, personalized recommendation, and privacy protection.