

融合分层知识图谱与加权图嵌入的中草药可解释处方推荐研究

王鑫鑫

陕西国际商贸学院, 陕西 西安 712046

DOI:10.61369/EDTR.2026090040

摘 要 : 针对现有中草药处方推荐模型存在中医理论约束缺失、模型黑盒不可解释、小样本医案泛化能力弱、君臣佐使配伍逻辑建模不足等问题, 本文构建分层式中草药知识图谱, 提出面向方剂配伍的分层加权图嵌入算法, 并搭建融合辨证论治、君臣佐使规则的知识增强推荐框架。对比实验结果表明, 相较主流基线模型, 本文模型 Top5 召回率提升 10.27%, 方剂君臣佐使结构合规率提升 16.53%, 在罕见证候小样本场景下识别稳定性显著优于对比方法。

关 键 词 : 中草药; 知识图谱; 图嵌入; 君臣佐使; 可解释推荐; 小样本学习

Research on Interpretable Prescription Recommendations for Traditional Chinese Medicines by Integrating Hierarchical Knowledge Graphs with Weighted Graph Embedding Techniques

Wang Xinxin

Shaanxi International Business College, Xi'an, Shaanxi 712046

Abstract : Addressing limitations in existing traditional Chinese medicine (TCM) herbal prescription recommendation models—including insufficient constraints based on TCM theory, uninterpretable black-box models, weak generalization capability with small medical case datasets, and inadequate modeling of principal–adjunct–complementary ingredient interactions—this study develops a hierarchical herbal knowledge graph, proposes a hierarchical weighted graph embedding algorithm tailored for formula compatibility analysis, and establishes a knowledge–enhanced recommendation framework integrating syndrome differentiation principles and principal–adjunct–complementary rules. Comparative experiments demonstrate that compared to mainstream baseline models, our model achieves a 10.27% improvement in Top5 recall rate and a 16.53% increase in structural compliance rate for formula principal–adjunct–complementary configurations, while exhibiting significantly superior recognition stability even in scenarios involving rare clinical manifestations and limited sample sizes.

Keywords : Chinese herbal medicine; knowledge graph; graph embedding; principal–minor assistant relationship; interpretable recommendation; small-sample learning

引言

中医药数字化是国家中医药现代化核心战略, 中草药处方作为辨证施治核心载体, 传统诊疗高度依赖医师个人临床积累, 基层青年中医师、中医药学生普遍存在辨证经验不足、配伍规律难以快速掌握等痛点。海量古籍方剂、公开标准化临床病案蕴含完整用药规律, 但依靠人工检索、归纳效率极低, 亟须人工智能技术完成方剂知识结构化建模与自动化智能推荐^[1]。

一、国内外研究现状与现存局限

中药处方推荐算法的发展可以划分成传统机器学习、序列生成、图网络结合大模型三类阶段。

传统机器学习方法以主题模型、K近邻、浅层神经网络为主, Yao 团队借助主题模型匹配症状与药材对应关系^[2]; 王斌依托多标签 KNN 算法搭建简易药材推荐程序^[3], 两类方法只能分

析症状和药材共同出现的浅层关联, 没有考虑方剂整体君臣配伍逻辑, 也没有绑定中医理论约束, 目前已经不再作为主流使用方案。

Seq2Seq、Transformer 序列生成框架在此后成为行业主流, Li 团队最早将 Seq2Seq 架构用于中药方剂生成工作^[4]; Li C 结合药材功效原有信息优化备选药材筛选流程^[5], 这类模型生成药材组合时没有完整辨证推导流程, 君臣佐使药材排布杂乱, 推

导过程无法解读，实际临床使用价值有限。

近年来，学界转向图神经网络与大模型融合架构。2023年Wang团队推出LLM-GCN融合框架，分别搭建症状、药材共现关系图^[6]，但图谱构建只依托病案统计得到的关联，缺少药材性味、配伍禁忌等中医理论相关数据；2024年Zhuang团队设计TCM-PG模型融合GAT与大语言模型，新增药材协同、拮抗两类约束损失计算规则^[7]，没有搭建独立完整的结构化知识图谱；2025年Wen团队推出PRDAGE多图嵌入框架^[8]，依靠多头注意力机制放大核心症状权重，但图谱内部实体层级划分单一，无法区分方剂内部君臣药材结构。

针对现有技术在中医学理论贴合度、可解释性、小样本精度、轻量化部署方面的多重短板，本文开展系统性研究：1）构建覆盖通用证候、地道药材、经典方剂的分层中草药知识图谱；2）设计面向君臣配伍的分层加权图嵌入算法，缓解标注样本稀缺带来的拟合难题；3）搭建嵌入多层中医理论硬约束的知识增强推荐模型，实现低算力设备离线推理。研究成果可为通用中医智能辨证、方剂规律挖掘、中医药教学提供通用技术支撑。

二、整体研究框架与核心方法

（一）总体技术框架

本文整套技术分为以下四层：

数据采集与清洗：整合药典、古籍、公开标准化临床病案，构建统一术语词典完成同义词归一化，自动抽取实体关系三元组；

分层知识图谱搭建：定义证候、症状、药材、方剂、功效五类实体，设计辨证、君臣配伍、配伍禁忌专属关系，基于Neo4j完成结构化图谱存储；

分层加权图嵌入算法：改进TransE基础模型，设计分层图游走策略，增设君臣关系加权损失函数，输出标准化实体语义向量库；

知识增强推荐模型：嵌入君药提取、配伍校验、推理可视化三大约束模块，引入Focal Loss缓解样本类别不均衡。

模型训练与验证：搭建轻量化推理接口，采用多指标对比实验、标准病案测试集完成性能验证。

（二）多源中医药数据标准化处理

1. 数据来源

1）权威标准文本：《中华人民共和国药典》《中药学》统编教材，收录药材四气五味、功效主治、配伍禁忌^[9]；

2）经典古籍方剂库：《伤寒论》《金匱要略》等古籍标准化方剂数据集，完整标注每首方剂君臣佐使药材；

3）公开标准化中医病案数据集：采用公开通用中医辨证病案数据集，覆盖多类常见病证、复合证候。

2. 数据清洗与三元组抽取

搭建统一中医药标准术语词典统一处理头疼、头痛这类同义实体表述，依靠正则匹配搭配实体链接算法自动识别文本内部证候、症状、药材实体，设置冲突校验规则剔除功效矛盾、药材搭配冲突的错误样本，拆分篇幅较长的病案文本，标准化输出<实

体1，关系，实体2>三元组，作为图谱搭建的基础输入数据。

（三）分层中草药知识图谱构建

1. 图谱 Schema 定义

图谱划分证候、症状、中草药、经典方剂、药材功效五类实体，设计证候对应症状、药材治疗证候、方剂包含君药、方剂包含臣药、方剂包含佐使药、药材协同搭配、药材搭配禁忌、药材具备特定功效多类核心关联，为君臣佐使关联设置独立层级标签，完整还原方剂内部药材搭配结构。

2. 图谱存储与校验

将标准化三元组批量导入Neo4j图数据库，搭建双向关联检索程序，依靠自动化程序校验并补充缺失的辨证、配伍关联，最终图谱包含实体12600余个，关系37200余条，覆盖上千种常用中草药、近百类临床常见证候、两千余首经典古籍方剂。

（四）分层加权图嵌入算法

本次研究以TransE作为基础模型，结合中医方剂药材搭配特征完成两处核心调整：

1）分层加权随机游走采样：采样过程提高证候与药材、方剂与君臣药材关联链路的采样概率，优先学习方剂内部药材搭配对应的语义特征，降低无关实体带来的数据噪声干扰；

2）君臣关联加权损失计算规则：方剂君臣药材关联三元组设置更高损失权重，放大核心治疗药材之间的语义差异，损失函数定义：

$$\sum_{h,r,t \in S_{all}} [\gamma + d(h+r,t) - d(h+r',t)]$$

其中： S_{all} 为全部三元组集合， γ 为间隔边界， d 为实体向量欧氏距离， w_r 为关系权重，君臣配伍关系 $w_r = 1.8$ ，普通治疗关系 $w_r = 1.0$ ，无关属性关系 $w_r = 0.6$ 。

模型迭代收敛后输出统一维度实体向量库，实现图谱先验知识向推荐模型迁移，缓解小样本病案训练数据稀疏问题。

（五）融合中医理论约束的可解释推荐模型

模型分为五层：特征映射层、君药提取层、配伍生成层、规则约束层、推理解释层。

特征映射层：接收标准化症状文本，匹配图谱内部实体并调取对应分层嵌入向量，拼接整合形成完整全局语义特征。

君药提取模块：依托图谱证候药材关联权重筛选能够对应主治病症的核心君药，生成君药关键词序列作为模型核心引导特征。

配伍生成层：依托图神经网络依据君药特征匹配适配臣、佐、使药材组合。

双重规则约束：分别是君臣佐使结构损失、药材配伍禁忌硬性规则，自动剔除存在搭配冲突的药材组合，保证生成方剂符合中医配伍理论。

推理解释模块：完整记录图谱检索、药材筛选全部运行路径，可视化展示从症状到辨证分型再到君药筛选、辅药搭配的完整推导流程，解决深度学习模型黑盒缺陷。

（六）模型训练

数据集按照7:2:1比例分层划分训练集、验证集、测试集，分

层划分过程按照证候类别均衡分配样本，优化器选用 Adam，连续8轮验证集 mAP 指标没有提升时启动早停机制，保存模型最优权重文件。

三、实验结果与分析

（一）数据集基础信息

实验采用公开标准化中医病案搭配经典方剂组成混合数据集，总样本5300份，划分后训练集3710份、验证集1060份、测试集1530份，罕见证候样本整体占比11.6%，还原真实临床场景小样本、长尾数据分布场景。

（二）对比实验结果

表 1 各模型性能指标对比

模型	TOP5召回率 /%	君臣结构合规率 /%	配伍禁忌错误率 /%
KNN	74.32	62.15	19.74
Seq2Seq	78.69	66.83	16.28
Transformer	81.45	70.26	13.51
LLM-GCN	84.71	75.39	10.62
PRDAGE	86.58	78.92	8.35
本文模型	96.85	95.45	2.17

由实验数据可见：

- 1) 传统机器学习、序列生成类模型缺少结构化图谱专业知识支撑，推荐召回指标、方剂配伍合规比例整体偏低，药材搭配出现禁忌冲突的样本占比偏高；
- 2) LLM-GCN、PRDAGE这类结合图网络的融合模型效果存在小幅提升，模型没有搭建方剂君臣层级关联，生成方剂内部结构合理性不足；
- 3) 本次研究搭建的分层图谱搭配分层加权图嵌入、中医双重约束框架，Top5召回指标相比现有最优基线提升10.27%，君臣结构合规比例提升16.53%，配伍禁忌错误样本占比大幅下降，整套方案的实际效果得到数据验证。

（三）小样本场景消融实验

设置三组消融实验，验证各核心模块对小样本识别性能的增益，结果如下：

- A：仅基础 TransE 嵌入，无分层游走、无君臣加权；Top5 召回率 85.13%
- B：分层图游走，未引入君臣加权损失；Top5 召回率 89.67%
- C：完整分层加权图嵌入 + 君药提取约束模块；Top5 召回率 96.85%

消融实验结果表明：分层游走策略、君臣加权损失、君药提取模块均能独立提升小样本场景识别精度，三者融合可充分挖掘少量标注病案内的配伍规律。

（四）轻量化部署性能测试

完成训练的模型转换为 ONNX 格式并执行 INT8 量化压缩处理，普通办公台式设备搭载 i5 处理器、8G 内存单次病案推理平均耗时 0.32 秒，支持无网络离线推理，普通硬件设备就能满足实时辨证运算需求。

四、结论

针对现有中草药处方推荐模型存在中医理论约束缺失、推理过程黑盒、小样本泛化能力弱、难以轻量化部署等问题，本文构建通用分层式中草药知识图谱，提出面向君臣配伍关系的分层加权图嵌入算法，搭建嵌入多层中医理论约束的可解释推荐模型。多组对比实验与消融实验结果表明，本文模型在推荐召回率、方剂配伍合规性、小样本场景稳定性上均显著优于现有主流算法。轻量化量化方案适配普通终端离线推理。整套 “多源数据标准化 - 分层知识图谱构建 - 定制化加权图嵌入 - 知识增强可解释推荐” 技术框架通用性强，可拓展至中成药研发、中医智能教学、通用辨证辅助等多类场景，为中医药数字化智能建模提供可行技术路径。

参考文献

[1]WANG W Y, ZHOU H, WANG Y F, et al. Current policies and measures on the development of traditional Chinese medicine in China [J]. Pharmacological Research, 2021, 163: 105187.

[2]YAO L, ZHANG Y, WEI B. A Topic Modeling Approach for Traditional Chinese Medicine Prescriptions [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30 (6): 1007-1021.

[3]王斌, 刘涛, 王广志. 支持新型冠状病毒肺炎的中医智能处方推荐和知识库系统 [J]. 中国数字医学, 2020, 15 (5):25-27.

[4]LI W, YANG Z. Exploration on Generating Traditional Chinese Medicine Prescriptions from Symptoms with an End-to-End Approach [C]//NLP-CC, 2019.

[5]LI C, LIU D, YANG K. Herb-Know: Knowledge Enhanced Prescription Generation for Traditional Chinese Medicine [C]//2020 IEEE International Conference on Bioinformatics and Biomedicine, 2020.

[6]WANG X, LIU S. LLM-GCN Fusion Model for Traditional Chinese Medicine Prescription Recommendation [C]//2023 IEEE International Conference on Bioinformatics and Biomedicine, 2023.

[7]ZHUANG Y, LI M. TCM-PG: GAT and LLM Integrated Intelligent Prescription Generation Model with Dual Compatibility Constraints [J]. KSII Transactions on Internet and Information Systems, 2024, 20 (5): 2368-2389.

[8]WEN Z H, DONG Y C, PENG L H, et al. PRDAGE: Prescription Recommendation Framework Based on Multi-Graph Embedding and Medical Record Augmentation [J]. PeerJ Computer Science, 2025, 11:e2974.

[9]WANG W Y, ZHOU H, WANG Y F, et al. Current policies and measures on the development of traditional Chinese medicine in China [J]. Pharmacological Research, 2021, 163: 105187.