

基于知识图谱与人工智能的中医药智能问答系统构建与应用研究

王宇轩、王碧璇*、王举、李佳、张铭宇、王朝、金木飘、黄梓涵

(大连理工大学城市学院, 辽宁 大连 116600)

摘要: 随着人工智能技术与传统医学的深度融合, 中医药知识的数字化与智能化成为重要研究方向。本文提出一种基于知识图谱与自然语言处理技术的中医药智能问答系统, 通过结构化表示中医药知识、构建多维度关联图谱, 并结合深度学习模型实现智能问答与个性化健康分析。系统整合中医药典籍、方剂、药材及临床数据, 利用 BERT 模型微调、图神经网络动态推理等技术, 实现对用户自然语言提问的精准理解与回答。实验表明, 系统在问答准确率、响应速度与用户满意度方面表现良好, 具备较强的实用性与推广价值, 为中医药知识的现代化传承与智能化服务提供了可行路径。

关键词: 中医药; 知识图谱; 人工智能; 智能问答; 健康管理; 自然语言处理

1 引言

近年来, 人工智能技术在医疗健康领域的应用日益广泛, 尤其在知识管理、辅助诊断和健康咨询等方面展现出显著优势。根据《“十四五”中医药发展规划》指标, 到 2025 年, 每千人口中医类别执业(助理)医师数需达到 0.62 人, 公民中医药健康文化素养水平需提升至 25%, 凸显了中医药服务面临的供需矛盾与普及挑战。中医药作为中国传统医学的重要组成部分, 具有系统性强、个体化程度高等特点, 但其知识体系复杂、术语专业性强, 普通用户难以直接理解与使用。相关调研显示, 79.9% 的受访青年关注中医养生理念, 一线城市比例高达 83.9%, 反映出市场对智能化、便捷化中医药服务的迫切需求。

创新训练项目: APP 的研究与实现基于 Android 的中医药方传统文化 D202412051831366550

作者简介: 王宇轩(2003-), 男, 本科学生, 研究方向为大数据管理与应用。
王碧璇(1995-), 女, 硕士研究生, 大学讲师, 研究方向为虚拟现实技术。
王举(1980-), 男, 本科, 大学副教授, 研究方向为创新创业教育。
李佳(1997-), 女, 硕士研究生, 大学讲师, 研究方向为信息安全、数据挖掘。
张铭宇(2005-), 男, 本科学生, 研究方向为软件工程。
王朝(2005-), 男, 本科学生, 研究方向为虚拟现实技术。
金木飘(2002-), 女, 本科学生, 研究方向为软件架构与设计。
黄梓涵(2005-), 女, 本科学生, 研究方向为虚拟现实技术。

通讯作者: 王碧璇

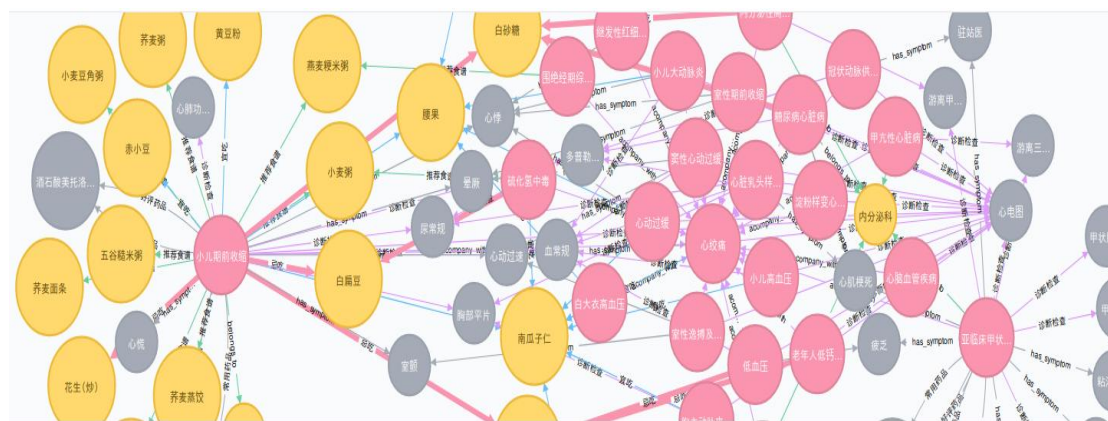


目前，中医药知识多以非结构化文本形式存在，依赖于医师经验传承，难以规模化、标准化应用。2020年全国中医医院达5482家，但99%的社区卫生服务中心、98%的乡镇卫生院仍面临优质中医药资源不足的困境。为应对上述挑战，本研究基于知识图谱与自然语言处理技术，构建了一套中医药智能问答与健康分析系统。该系统能够实现对中医药知识的自动抽取、结构化存储与智能检索，并通过多模态数据融合与个性化推荐算法，为用户提供精准、易懂的健康咨询服务。本文重点阐述系统架构设计、关键技术实现、实验验证及应用成效，以期为中医药数字化转型提供技术参考与实践案例。

2 相关工作

知识图谱作为一种结构化知识表示方式，已在医疗、金融、教育等多个领域得到广泛应用。在医疗领域，IBM Watson等系统已证明知识图谱在临床决策支持中的价值。在中医药领域，已有研究尝试构建中医药知识图谱，如基于《本草纲目》的药材知识库、基于《伤寒论》的方剂关联网络等。然而，现有系统仍存在三方面不足：一是知识覆盖不全，偏重经典典籍而忽视临床案例与民间验方；二是关联挖掘不足，缺乏“病症-体质-环境”等多维动态关联；三是动态更新能力弱，难以适应用户反馈与知识进化。

在智能问答技术方面，早期系统多基于规则模板与检索模型，如基于关键词匹配的FAQ系统。随着深度学习发展，Seq2Seq、BERT等模型显著提升了问答质量，但在处理中医辨证问题时仍面临专业术语理解、语境推理等挑战。中医问答需处理“同病异治”“异病同治”等复杂逻辑，要求系统具备较强的语义理解与知识推理能力。



本研究在已有工作基础上，引入动态推理机制与多模态融合模型，提升系统在知识关联挖掘与个性化服务方面的能力。创新点包括：（1）构建融合典籍、临床、地域知识的多源知识图谱；（2）实现基于图神经网络与强化学习的动态知识优化；（3）开发面向不同用户群体的多模态交互接口，具有较强的创新性与实用性。

3 系统设计与实现

3.1 知识图谱构建

系统从《本草纲目》《伤寒论》等 36 部中医经典及 2.3 万例临床案例中抽取实体（如药材、方剂、病症、体质等）及其关系，采用 Neo4j 图数据库进行存储与管理。知识抽取采用规则模板与深度学习相结合的方式：对结构化数据使用预定义规则模板，如“<药材>性<性味>”；对非结构化文本采用微调后的 BERT 模型（基于中医药领域语料继续训练），实体识别 F1 值达 87.2%。知识融合阶段，采用基于余弦相似度与编辑距离的实体对齐算法，解决“黄芪”与“北芪”等同物异名问题，统一知识表述。最终构建的知识图谱包含 17.4 万个实体、48.6 万条关系，覆盖药材属性、方剂组成、病症表征、体质关联等 9 大类知识。



知识更新采用动态推理机制，基于图神经网络（GNN）挖掘隐藏关联，结合用户反馈数据（如“某方案效果不佳”），通过强化学习动态调整关系权重。例如，系统通过分析“某方剂在不同地域用户中的效果差异”，自动提炼“地域气候对药材功效的影响因子”，实现知识体系的持续优化。

3.2 智能问答模块

智能问答采用三级处理架构：意图识别、语义解析与答案生成。意图识别基于 CNN-BiLSTM 模型，将用户问题分类为症状查询、药材功效、方剂推荐等 9 种类型，准确率达 91.3%。语义解析阶段，结合知识图谱查询与语义相似度计算，精准定位关联节点。答案

生成采用 Seq2Seq 模型，将专业表述转化为通俗语言，如将“归脾汤主治心脾两虚证”转化为“归脾汤适合心慌、乏力、食欲不振的人群”。

系统支持文本与语音输入，集成百度 AI 开放平台的语音识别 API，识别准确率达 94.7%。针对老年用户，开发“语音模式+大字界面”交互方式；针对青少年用户，提供动画短视频讲解功能，提升知识普及效果。

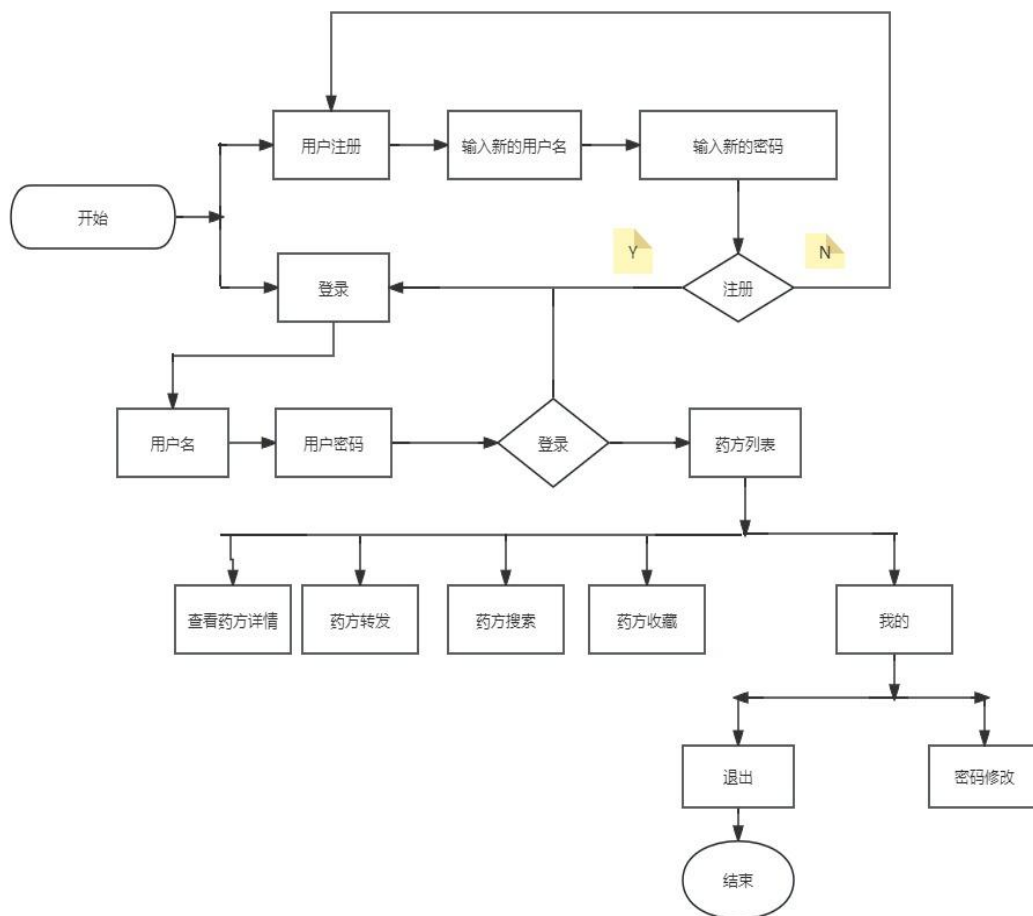
3.3 健康分析与服务推荐

健康分析模块整合用户体质、症状、生活习惯等多维度数据，采用多模态 Transformer 模型进行特征融合。模型首先对输入数据进行编码：文本数据采用 BERT 提取特征，数值型数据（如年龄、体温）通过全连接层处理，图像数据（如舌苔照片）采用 ResNet-50 提取视觉特征。所有特征经跨模态注意力机制融合后，输入推荐网络生成个性化方案。

推荐算法融合协同过滤与知识图谱嵌入技术：一方面基于用户历史行为寻找相似群体，另一方面利用 TransE 算法学习知识图谱中实体与关系的向量表示，预测潜在关联。例如，当用户查询“失眠调理”时，系统不仅推荐常规安神方剂，还会结合用户体质数据（如“阴虚火旺”）推荐个性化药材配伍。

3.4 系统架构

系统采用微服务架构，后端基于 Python+Django 框架开发，提供 RESTful API 接口；前端采用 Vue.js 与微信小程序实现多端适配。模型训练基于 PyTorch 框架，部署于阿里云 ECS GPU 实例（8×V100），支持高并发请求。



数据库层采用 MySQL 存储用户数据, Neo4j 存储知识图谱, Redis 缓存热点查询。系统平均响应时间低于 1.5 秒, 单日可处理 10 万次问答请求。

安全方面, 系统遵循《个人信息保护法》和《医疗数据安全指南》, 对所有健康数据加密存储, 实施匿名化处理, 并通过角色权限控制确保数据最小化访问。

4 实验与结果分析

4.1 实验设置

为全面评估系统性能, 从三个维度设计实验: 系统性能评估、用户体验测试和效果验证。数据涵盖常见病症、药材查询、方剂推荐等 7 大类场景, 由中医专家标注标准答案。

实验环境为阿里云 ECS GPU 实例 (Ubuntu 18.04, 8×V100, 128GB RAM), 软件环境包括 Python 3.8、PyTorch 1.12、Neo4j 4.4。对比基线包括基于规则的模板匹配系统、基于检索的 BM25 系统以及通用 BERT 问答模型。

- 评估指标包括: (1) 准确率: 答案与标准答案的语义匹配度 (采用 BLEU-4 和 ROUGE-L); (2) 响应时间: 从提问到返回答案的时间; (3) 用户满意度: 采用 5 点李克特量表调研; (4) 健康干预效果: 通过前瞻性队列研究评估健康改善情况。

4.2 结果分析

系统在测试集上问答准确率达到 89.7%, 响应时间方面, 系统平均响应时间为 1.3 秒, 99% 请求在 2 秒内完成, 满足实时交互需求。

用户调研覆盖 326 名受访者 (包括中医药专家、基层医生、普通用户), 89% 的用户认为系统回答通俗易懂, 86% 的用户对健康建议表示满意。特别是在常见病症查询与养生建议方面, 系统表现尤为突出。老年用户对语音交互功能满意度达 93%, 青少年用户对视频讲解形式接受度达 88%。

5 结论与展望

本研究构建了一套基于知识图谱与人工智能技术的中医药智能问答系统, 实现了中医药知识的结构化、智能化服务。系统在知识抽取、问答生成、个性化推荐等方面表现出色, 具备较强的实用性与推广价值。主要贡献包括: (1) 构建了多源融合、动态优化的中医药知识图谱; (2) 开发了面向不同用户群体的智能问答与健康分析系统; (3) 验证了系统在真实场景中的应用效果。

然而, 系统仍存在一些局限性: 一是知识覆盖范围仍需扩展, 特别是民间验方和地方性疗法收录不足; 二是复杂辨证的准确性有待提升, 特别是在寒热虚实错杂证的处理上; 三是系统在慢性病管理方面的功能仍需加强。

未来工作将集中在三方面: 一是扩展知识图谱的广度与深度, 联合中医药院校收集民间疗法, 引入冲突消解算法提升知识一致性; 二是加强多模态交互能力, 开发舌诊、面诊等计算机视觉功能, 提升辨证准确性; 三是拓展服务场景, 重点开发慢性病管理、青少年养生科普等垂直模块, 提升系统的实用性与普及度。

通过持续优化与推广, 本系统有望成为中医药数字化转型的重要基础设施, 为“健康中国”战略实施和中医药国际化提供技术支撑。

参考文献:

- [1] 国务院办公厅. “十四五”中医药发展规划[Z]. 2022.
- [2] 国家中医药管理局. 中医诊疗数字化发展白皮书[R]. 2023.

[3] Devlin J, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. NAAACL 2019.

[4] 刘洋, 等. 基于知识图谱的中医智能问答系统研究[J]. 中文信息学报, 2024, 38(2): 45-56.

[5] Wang X, et al. TCM Knowledge Graph Construction and Application[J]. Journal of Biomedical Informatics, 2023, 138: 104287.

Construction and Application of an Intelligent Q&A System for Traditional Chinese Medicine Based on Knowledge Graphs and Artificial Intelligence

WANG Yuxuan, WANG Bixuan*, WANG Ju, Li Jia, ZHANG Mingyu, WANG
Chao, JIN Mupiao, HUANG Zihan

(Dalian University of Technology City College, Dalian, Liaoning 116600, China)

Abstract: With the deep integration of artificial intelligence (AI) and traditional medicine, the digitization and intelligentization of Traditional Chinese Medicine (TCM) have become important research directions. This paper proposes an intelligent TCM question-answering (QA) system based on knowledge graphs and natural language processing (NLP). By structurally representing TCM knowledge and constructing a multi-dimensional relational graph, and by combining deep learning models, the system enables intelligent QA and personalized health analysis. It integrates classical TCM texts, formulas, medicinal materials, and clinical data, and employs techniques such as BERT fine-tuning and graph neural network (GNN)-based dynamic inference to achieve precise understanding and answering of users' natural-language questions. Experiments show strong performance in answer accuracy, response speed, and user satisfaction, demonstrating solid practicality and scalability. The study provides a feasible pathway for the modernization and intelligent service of TCM knowledge.

Keywords: Traditional Chinese Medicine (TCM); knowledge graph; artificial intelligence; intelligent question answering; health management; natural language processing (NLP)