数字中医中药

DOI:10.13193/j.issn.1673-7717.2025.06.004

基于深度学习和医案相似度的古代方剂智能推荐方法研究

郭凤英[1],高晨曦[1],王一凡[1],杜鹏程[2],王若佳[1],翟兴[1],田甜[2]
(1.北京中医药大学管理学院,北京 100105;2.北京中医药大学中医学院,北京 100105)

摘要:目的 解决中医古籍中的中药计量规则复杂、方剂查找不便和缺乏针对古籍的现代智能应用的问题。方法 总结历代计量转换标准和特殊计器计量转换标准,提出一种基于"主证+方剂名称+中药组成"特征融合的医案相似度计算方法,采用基于深度学习的知识增强语义表示模型(Enhanced Representation from Knowledge IntEgration,ERNIE)模型进行模型训练;利用MySQL数据库和Django框架,设计开发一个基于深度学习推荐引擎的智能化古代方剂推荐和计量换算系统。结果 采用该方法得到的总体推荐准确率达到了89.55%、精确率为94.51%、召回率为83.97%,取得了较好的效果,系统能够实现方剂智能检索和中药计量单位转换,根据用户输入的症状推荐主治功效匹配度较高的系列古方信息,并将古方按不同朝代的换算规则显示方剂组成及符合现代标准计量的中药用量。结论 利用该方法进行方剂智能推荐和中药计量转换,达到了较为满意的效果,为古代方剂的智能推荐提供了一种新的思路,对中药用量的标准化具有一定的借鉴意义和临床用药参考意义。

关键词:深度学习;医案相似度,古代方剂;智能推荐;计量转换
中图分类号:R2-03;R289.5 **文献标志码**:A **文章编号**:1673-7717(2025)06-0024-07

Research on Intelligent Recommendation Method of Ancient Prescription Based on Deep Learning and Medical Case Similarity

GUO Fengying[1],GAO Chenxi[1],WANG Yifan[1],DU Pengcheng[2],WANG Ruojia[1],ZHAI Xing[1],TIAN Tian[2]
(1. School of Management,Beijing University of Chinese Medicine,Beijing 100105,China;
2. School of Traditional Chinese Medicine,Beijing University of Chinese Medicine,Beijing 100105,China)

Abstract:*Objective* To solve the problems of complicated rules of measurement of Chinese medicines in ancient books,inconvenience in finding formulaeand lack of modern intelligent applications for ancient books. *Methods* It summarized the measurement conversion standards and special meter measurement conversion standards in the past dynasties,proposed a similarity calculation method for medical cases based on the fusion of the features of" mainevidence + formulaname + composition of traditional Chinese medicine",and adopted Enhanced Representation from Knowledge IntEgration(ERNIE)model based on deep learning for model training. Itdesigned and developed an intelligent ancient formula recommendation and measurement conversion system based on deep learning recommender engine,using MySQL database and Django framework. Using MySQL database and Django framework,it designed and developed an intelligent ancient prescription recommendation and measurement conversion system based on deep learning recommendation engine. *Results* The overall recommendation accuracy rate obtained by this method reached 89.55%,the precision rate was 94.51%,and the recall rate was 83.97%,which achieved better results,and the system was able to achieve the intelligent retrieval of formulae and the conversion of Chinese medicine measurement units,recommend a series of ancient formulae with a high degree of matching of the main treatment efficacy according to the symptoms inputted by the user,and display the composition of formulas and the Chinese medicine dosage in accordance with the conversion rules of different dynasties. It also displayed the composition of the formula and the dosage ofChinese medicine according to the conversion rules of different dynasties,which conformed to the modern standard measurement. *Conclusion* The method of intelligent recommendation of prescriptions and conversion of Chinese medicine measurements achieves satisfactory results,provides a new way of thinking for the intelligent recommendation of ancient prescriptions,and has certain reference significance for the standardisation of Chinese medicine dosage and the reference significance for the clinical use of medicines.

Keywords:deep learning;similarity of medical records;ancient formula;intelligent recommendations;measurement transformation

基金项目:国家自然科学基金青年科学基金项目(82204963);教育部中国高校产学研创新基金—蓝点分布式智能计算项目(2021LDA12004)
作者简介:郭凤英(1979-),女,河南兰考人,副教授,硕士研究生导师,硕士,研究方向:机器学习、自然语言处理,中医药大数据。
通讯作者:田甜(1983-),女,山东淄博人,副教授,硕士研究生导师,博士,研究方向:中医基础理论。E-mail:tt8324@163.com。

随着中医药数字化、信息化管理程度越来越高，积累了海量的古籍数据，对于使用者而言，从海量的古籍中找出治疗疾病的有效方剂变得异常困难，随之产生了信息过载问题。有研究发现，智能推荐系统则是解决信息过载问题的一个有效工具，可以帮助用户更加便捷地查询和获取有效信息[1]，因此，利用智能推荐技术，设计开发古代方剂智能推荐和计量转换系统，将会推动古籍的有效利用。

目前，关于方剂智能推荐的研究以基于传统机器学习算法的方剂相似度计算为主。方剂相似度计算通常包含 4 类方法。(1)基于文本相似度的方法[2-4]；(2)基于成分相似度的方法[4-5]；(3)基于网络相似度的方法[3,6]；(4)基于深度学习的方法[7-9]。如，李园白等[10]提出利用 Jaccard、夹角余弦对中药组成、剂量和功效三层相似度计算比较法分析处方之间的相似度；朱志鹏等[5]利用 LDA 主题模型和 KL 距离分别挖掘了“方剂 - 证型 - 中药组成”的隐含关系来计算方剂相似度；郭文龙等[2]利用 TF - IDF 算法计算方剂属性及方剂之间的相似度，并构建了中药方剂数据挖掘系统；李新龙等[6]通过对方剂相似度的方法进行研究后发现，药物量效关系、药物功效与方剂功效关系等难以有效评估及量化。

此外，不同朝代的计量换算规则复杂、难以记忆，使用者即使查找到了古代方剂，但如果不熟悉当时的计量单位换算规则，也可能导致在遣方用药时对其中的中药用量难以准确把握。目前，计量标准化研究局限在产出学术著作和学术论文，缺乏数字化和标准化研究。如，王安妮等[11]以方剂资料为依据，分析了中药用量在不同历史时期的概貌及其演变规律；黄璐明等[12]提出和探讨了从临床到实验再到临床的中药用量规范标准化研究的新模式；徐凤凯[13-14]调研了《伤寒杂病论》中的特殊计器，同时对《伤寒论》和《金匮要略》中的特殊计量换算进行了整理。

综上发现，目前的方剂智能推荐研究主要以方剂相似度计算为主，且大多参考方剂说明中的药物组成、功效和剂量信息，而在中医古籍中的方剂信息大多包含在一个临床医案中，这些医案中包含患者的症状、证候、既往病史、服药后反馈、方剂名称、药物组成和古代剂量等信息，因此，可以假设，在古籍中，相似的医案包含着相似的患者主证和相似的方剂，根据中医“方证相应”的思想，方剂与主证是对应的，若要向现代使用者推荐方剂，则需要考虑该使用者的实际症状和某医案中的主证是否相似，才能推荐该医案中所包含的方剂，而在现有的研究中，没有在方剂推荐中考虑主证信息。在计量研究方面，学者们虽然对部分古籍中的经典著作的计量换算和特殊计量有深入的研究，对于中药用量标准化研究模式进行了探索，但缺乏对历代计量和特殊计量的整体研究和数字化研究，不方便查询和计算。

针对以上问题，本文研究设计基于深度学习的古方智能推荐和计量转换系统。利用自先秦至明清时代的中医古籍经典著作的医案文本和计量换算规则建立数据库，实现根据用户输入的疾病症状，基于多维特征融合的相似医案计算方法采用 ERNIE 模型进行建模，并将相似医案中的方剂向用户推荐，可视化显示方剂信息和中药用量标准换算结果。

1 古代方剂智能推荐和计量转换系统设计

1.1 系统概述

古代方剂智能推荐和计量转换系统充分挖掘中医古籍经典著作中的古代医案资源，根据用户输入的症状信息，结合历代中药用量计量转换规则，运用本研究设计的融合主证 + 方剂名称 + 药物组成的多维特征的医案相似度计算方法，将符合条件的相似医案中的方剂信息推荐给用户，辅助医生诊疗，提高医师工作效率。

1.2 系统功能设计

本系统主要包括方剂智能推荐和中药用量计量转换两大功能模块。

方剂智能推荐模块的主要功能是根据用户输入的症状推荐适宜方剂，在本研究中，实现方剂智能推荐功能的本质是计算医案之间的相似度。系统可以根据用户输入的症状，根据内置的古代医案相似度计算算法，从数据库中找到匹配度较高的系列相似医案，提取其中的方剂信息，输出包括方剂来源、方剂名称、中药组成及功能主治等信息，并根据计量转换规则将中药组成中的原始中药用量及计量单位标准化为“克(g)”显示，便于用户选择合适的方剂。

中药用量计量转换模块适用于用户查阅古籍方剂及其计量信息。系统可以方便用户查询各朝代经典著作中包含的古代方剂信息，并可以自动将古方剂中的中药用量及其计量单位根据计量转换规则进行转换，统一转换为现代标准化计量单位为“克(g)”。另外，系统可以方便用户查询特殊计量单位，如“附子大者”换算为 20 ~ 30 g；可以按照不同朝代查询计量单位，并实现计量单位标准化换算。

1.3 系统结构设计

1.3.1 *系统模块结构* 本系统综合考虑安全性和使用维护便捷性，采用 B/S 系统架构和 MySQL 数据库，服务器端采用 MVT(Model View Template，MVT)3 层结构，分别是 Model 层(模型层)、View 层(表现层)和 Template 层(模板层)。其中，View 层接收用户需求，根据请求内容通过 Model 层访问数据库，完成数据库数据交互后，将数据返回给 View 层，接着，由 Template 层的模板渲染后生成 HTML 页面返回给浏览器供用户浏览。

本系统共有 2 个模块，分别是计量转换模块和方剂智能推荐模块，实现以古医案为基础的多朝代中药用量计量的转换和方剂推荐。计量转换模块包括特殊计量查询、按朝代计量转换计算器、古代方剂的中药用量查询和计量转换 3 个功能。方剂智能推荐模块包括医案相似度计算功能，根据症状智能推荐古代方剂功能。具体的各层次任务划分如图 1 所示。

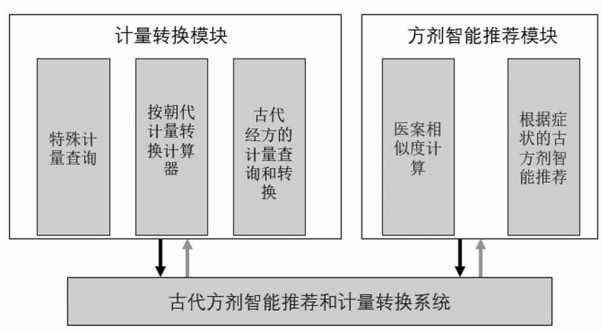


图 1 系统逻辑架构图

1.3.2 *模块接口定义* 古代方剂智能推荐和计量转换系统信息的采集是通过外部用户输入实现的。首先，对用户输入的症状文本进行预处理，包括数据清洗、数据去重、数据标准化、数据堆叠等，将用户输入文本处理成可供计算机处理的数据；然后，根据本系统构建的 ERNIE 模型对输入的症状和系统内训练的医案进行相似度预测，并将判定为相似的医案所对应的方剂信息及其中药用量计量标准化信息返回给智能推荐模块，供

用户参考使用。

2 古代方剂计量转换方法

中医古籍隶属不同历史时期和朝代,时期不同、朝代不同度量衡的计量差异较大,甚至同一朝代计量标准也不同,因此,将古籍中的方剂中药物按照计量转化标准转换成现代计量单位克,对于古籍的临床使用提供科学依据。

2.1 古代计量转换规则

本研究根据《经方剂量揭秘》[15]和《经方本原剂量问题研究》[16]两本书籍及若干参考文献归纳总结形成的中国历代衡制标准化计量转换规则表,见表 1,包括自战国至今共计 12 个朝代的常用度量衡单位,如:石、斤、两等,折合成现代标准化计量克(g)的数值。由于朝代存在差异性,若此朝代不存在某计量单位,则在表中以"/"表示。战国时期及南北朝时期各个国家所使用计量标准也不尽相似,故采用"时期 - 朝代"方式进行展现,而隋朝由于其存在大、小两种计量标准,故在表内进行标注。

表 1 中国历代衡制标准化计量转换规则

时代	衡制统一换算为现代标准化计量克/(g)								
	石	斤	两	铢	镒	斮	钧	钱	分
战国 - 楚	/	250	15.6	0.65	/	/	/	/	/
战国 - 赵	30 000	250	15.6	0.65	/	/	/	/	/
战国 - 魏	/	/	/	/	315	31.5	/	/	/
战国 - 秦	30 360	253	15.8	0.69	/	/	7590	/	/
秦	30 360	253	15.8	0.69	/	/	7590	/	/
汉	29 760	248	15.5	0.65	/	/	7440	/	/
三国	26 400	220	13.8	0.57	/	/	6600	/	/
两晋	26 400	220	13.8	0.57	/	/	6600	/	/
南北朝 - 北魏	/	440	/	/	/	/	/	/	/
南北朝 - 北齐	/	440	/	/	/	/	/	/	/
南北朝 - 北周	/	660	/	/	/	/	/	/	/
南北朝 - 梁	/	220	/	/	/	/	/	/	/
南北朝 - 陈	/	220	/	/	/	/	/	/	/
南北朝 - 南齐	/	330	/	/	/	/	/	/	/
隋 - 大计量	79 320	661	41.3	/	/	/	19 830	/	/
隋 - 小计量	26 400	220	13.8	/	/	/	6600	/	/
唐	79 320	661	40.0	/	/	/	/	4.13	0.4
宋	75 960	633	40.0	/	/	/	/	4.00	0.4
元	75 960	633	40.0	/	/	/	/	4.00	0.37
明	70 800	590	36.9	/	/	/	/	3.69	0.37
清	70 800	590	36.9	/	/	/	/	3.69	0.37

2.2 古代特殊计器转换规则

除了标准重量单位计量外,在中医经典文献中还存在大量非标准重量单位,如:枚、鸡子大等,由于朝代变迁及地理位置因素,此类单位往往不便于准确识别,故本研究通过上述两本书籍及若干参考文献,归纳总结非标准重量单位计量药物量化折算参考值表。见表 2。

表 2 特殊计器药物量化折算参考值表

药名	原剂量	折算计量/(g)
蜀椒	1 合	4.2
吴茱萸	1 升	约 70
酸枣仁	1 升	112
杏仁	1 升	112
葶苈子	1 升	124
赤小豆	1 升	150
麻子仁	1 升	100
麦冬	1 升	90
芒硝	1 升	124
五味子	1 升	76
半夏	1 升	84
粳米	1 升	160
香豉	1 升	124
大枣	1 枚	2.5
栝蒌实	1 枚	70;大者 120
诃子	1 枚	4.0
杏仁	1 枚	0.3
乌头	1 枚	3;大者 7
附子	1 枚	15;大者 30
枳实	1 枚	20
乌梅	1 枚	2.3
栀子	1 枚	0.5
鸡子黄	1 枚	12.5
厚朴	1 尺	20.0
竹叶	1 把	10.0
石膏	如鸡子大	约 50 - 60

3 古代医案相似度计算方法

中医古籍中的单条医案属于短文本,要计算医案之间的相似度,实际上是计算医案文本之间的相似度。文本相似度计算常常作为信息检索、智能问答、智能推荐等任务的基础,可以有效改进任务效果[17]。本研究选择的古代医案相似度算法即是基于双向编器表示模型(Bidirectional Encoder Representations from Transformers,BERT)模型优化后的 ERNIE 模型,是一种文本相似度算法。

3.1 古代医案相似度算法概述

抓主证是中医辨证施治中的一种重要方法,即在疾病表现出的众多症状中,抓住最主要、最典型的症状进行分析和治疗[18],在中医古籍经典著作的古代医案描述中,医生通常会详细记录患者的症状、体征、舌苔或脉象、所使用的方剂及中药组成等信息,虽然在临床上每个病案的症状不完全一样,只要符合主证要素,则可以使用相同的方剂或方剂的加减方,如桂枝汤的主证以发热、汗出、恶风为主,在伤寒论的很多医案中都使用了桂枝汤。

本文提出一种融合"主证 + 方剂名称 + 中药组成"的多维特征古代医案相似度计算方法,并依据该方法对古代医案文本进行结构化处理,使用 ERNIE 模型训练后获得医案相似度预测结果,对比仅使用主证相似度和仅使用药物组成相似度的预测结果的差异,以准确率、精确率及召回率作为模型评价标准,分析不同医案相似度计算标准下的古代方剂智能推荐算法的效果。

3.1.1 医案文本结构化提取 由于中医古籍医案的原文文本为文言文,包含大量的疾病症状、体征、治疗前后症状等信息,也包含所用方剂名称、中药及其剂量和制备方法等医学术语,需要分别把各类信息进行提取并结构化。其中,方剂名称和中药组成等信息采用使用 Python 编写程序自动化完成,而医案中所涉及到疾病的主证,则需要中医专业人员进行判断、识别、提取,并参考根据主证、中药组成、方剂名称,将医案条文两两标识为"相似"或"不相似"两个类别。

①方剂名称提取:依据《中国中医药学主题词表》《伤寒论》《金匮要略》《方剂学》等方剂名称标准,提取中医古籍医案的条文文本中包含的方剂名称,如:桂枝汤、半夏汤等,若未明

确提及方剂名称，则删除此条原文。

②中药组成提取：依据《中国中医药学主题词表》，提取古籍医案条文文本中包含的中药，仅保留药物名称，去除药物剂量、炮制方法等。

③主证提取：由 2 名中医专业人员分别独立进行主证提取，提取方式为：根据中医经典著作中的古代医案条文的描述，人工提取医案中的主证，如：恶寒、恶风、奔豚等；再由 1 位中医专业人员对两份结果对比校正，如发现两份文档存在分歧，则通过三方讨论达成一致意见后录入，以保证正确性。

3.1.2 医案相似度计算方法 本文设计的医案相似度计算方法如下：两个医案条文中，若使用同一方剂，则表明医案具有相似性，赋值为 1；若方剂名称不同，则再考察主证是否完全相同或部分相同，若主证完全相同，则表明医案具有相似性，赋值为 1；若主证的某部分相同，则计算出两个主证的不重复的主证数量和重复出现的主证数量，将重复主证数占主证总数的比例记为主证相似度，若主证相似度高于 0.5（包括 0.5），则判定两医案相似，赋值为 1。若方剂名称不相同且主证相似度低于 0.5，则再考察中药组成是否完全相同或部分相同，计算出每行数据两医案不重复的中药种类数量和两医案重复出现的中药种类数量，将重复中药数占中药总数的比例记为药物相似度，根据异病同治的原理，若药物相似度高于 0.5（包括 0.5），则判定两医案相似，赋值为 1；其余均赋值为 0，即不相似。多维特征融合的医案相似度计算方法流程如图 2 所示，预处理后的样例数据见表 3。

3.2 ERNIE 模型

本研究在计算医案相似度时，采用的是一种深度学习模型——ERNIE，ERNIE 模型是一种基于预训练的语义表示模型[19]，ERNIE 模型和 BERT 在模型结构和模型原理上类似，并在 BERT 的基础上做了优化。从结构上来看，两个模型在预训练阶段都增加了实体等先验知识，并在大型中文语料上进行了训练；不同的是，ERNIE 模型首先用预训练参数对模型初始化，并利用文本分类、问题问答等任务的标注数据对所有参数进行微调[20]。从模型原理上来看，两个模型都使用隐码语言模型进行任务训练，但 ERNIE 模型更多的是参考命名实体或词语关系等后再对字词之间的逻辑关系进行掩盖，因此，相对于 BERT 模型保留了词语的完整性和逻辑性[21]。因此，ERNIE 模型相比于 BERT 模型更具有全局性，预测效果也更优于 BERT 模型，在文本相似度领域具有优异的表现。

在短文本相似度计算任务中，ERNIE 可以通过对两个文本进行编码，然后计算它们之间的相似度得分来判断它们是否相似[20,22]。步骤如下：首先对短文本进行分词，并将分词后的数据转换为 BERT 模型的输入格式；然后，将分词后的文本进行向量化处理，此步骤主要使用预训练的 BERT 模型对两个文本进行编码将其向量化；接着，利用余弦相似度等方法计算两个文本向量之间的相似度得分；最后，根据预先设定的相似度得分的阈值来判断两个短文本是否相似。

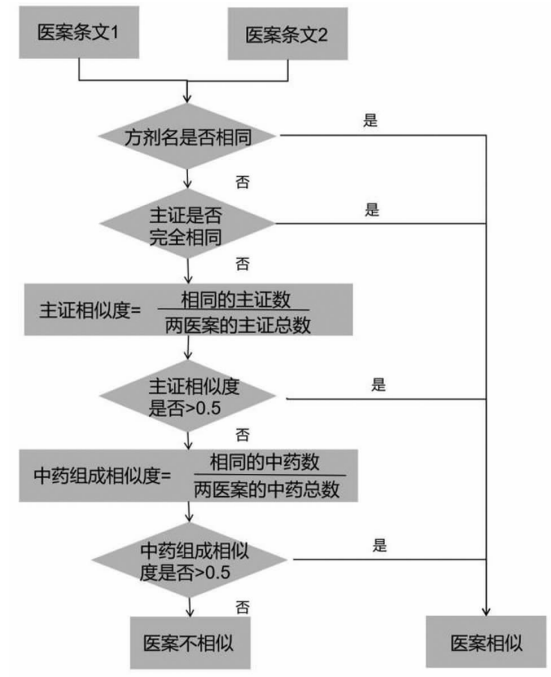


图 2 多维特征融合的医案相似度计算方法

3.3 模型训练和预测

设定 ERNIE 模型训练的 epoch 为 45，将数据样本组合成 Batch 数据，对于不等长的数据进行 Padding 操作，统一长度后训练模型。基于 ERNIE 模型，对已标注相似度的数据集进行训练，首先通过对输入数据与数据集内数据进行编码，然后依次计算用户输入文本与数据集中各数据之间的相似度，并输出模型判定为“相似”的医案文本。模型训练结果主要为二分类，即将用户输入数据与数据库实际数据进行预测，由模型给出 0 和 1 的预测结果，0 代表两文本不相似，1 代表两文本相似。

3.4 模型的评价指标

本研究采用二分类预测任务模型的评价指标，包括准确率（Accuracy）[23]、精确率[24]（Precision）、召回率[25]（Recall）和 F1 值（F1 - score）。公式见式 1 ~ 式 4。由于训练样本不均衡的特点，单独用某个评价指标不能评估模型性能，因此，本研究综合准确率、精确率、召回率 3 个评价指标衡量模型预测结果。同时，本研究使用最小二乘法设计损失函数，实际值和预测值之间的偏差越小，说明两医案越相似，代表模型效果越好[26]，

表 3 预处理后的样例数据

药物总数	相似药物数	药物组成 1	药物组成 2	总主证数	相似主证数	主证 1	主证 2	方剂名称 1	方剂名称 2	主证相似度	药物组成相似度	最终相似度
6	4	桂枝、芍药、甘草、生姜、大枣	桂枝、芍药、甘草、大枣、生姜	7	0	恶寒、恶风、发热、鼻鸣、干呕	腹满、腹痛	桂枝汤	桂枝加芍药汤	0.00	0.67	1
7	4	桂枝、芍药、甘草、生姜、大枣	桂枝、芍药、甘草、大枣、生姜、大黄	7	0	恶寒、恶风、发热、鼻鸣、干呕	腹痛、便秘	桂枝汤	桂枝加大黄汤	0.00	0.57	1
9	3	桂枝、芍药、甘草、生姜、大枣	柴胡、黄芩、人参、甘草、生姜、大枣、	9	1	恶寒、恶风、发热、鼻鸣、干呕	胁下硬满、干呕、不能食、往来寒热、脉沉紧	桂枝汤	小柴胡汤	0.11	0.33	0
9	1	桂枝、芍药、甘草、生姜、大枣	猪[illegible]、泽泻、白术、茯苓、桂枝	9	0	恶寒、恶风、发热、鼻鸣、干呕	心下痞、渴、口燥、小便不利	桂枝汤	五苓散	0.00	0.11	0

见式 5。

$$A = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 * \frac{P * R}{P + R} \quad (4)$$

$$Lossfunction = \min \sum_{i=1}^{n} (x_i - y_i)^2 \quad (5)$$

4 实验分析和系统测试

4.1 实验环境和数据准备

本研究医案文本数据选自《外科证治》284 首、《金匮要略》245 首、《温病条辨》208 首、《小儿药证直诀》135 首、《伤寒论》113 首、《脾胃论》63 首和《仙授理伤续断秘方》46 首，共 7 本中医古籍。依据上文提到的医案文本信息提取，提取后的数据集包括方剂 ID、方剂名、方剂组成、主治、来源，共计 1094 首方剂。根据多维度融合的医案相似度判定标准，对数据进行预处理，共得到医案文本数据 27 495 条，其中，存在不同医案中出现方剂名称相同的方剂，但中药组成和剂量不完全相同的情况，为保证数据的可靠性，本数据集参照医案中描述，遵从医案提供的中药组成及剂量，不做修改。但由于数据偏倚较大，正样本（相似医案）与负样本（不相似医案）占比为 0.93∶0.07，正样本占比数量偏少，故选择从总数据中随机抽取与正样本数量相同负样本（1897 条）构成模型训练数据集，本实验每条样本数据均由 2 条内容不同的医案文本及相似度二分类标签组成，即正、负样本各包含 1897 个医案对，共计 3794 条数据。将处理好的数据按照 7∶2∶1 的比例将数据集划分为训练集、测试集和验证集，使用百度开源 paddlepaddle 框架及其中包含 ERNIE 模型进行模型训练。

4.2 实验测试结果

根据专家意见，分别以太阳少阳证的“发热、恶风，汗出，鼻鸣，干呕，不欲饮食，胸满，胁痛”和太阳阳明证的“发热、恶风，汗出，鼻鸣干呕，腹痛，腹满，不大便”作为输入症状，分别得到仅使用主证作为判定标准、仅使用药物组成作为判定标准、多维度特征融合判定标准三类判定标准预测结果及结局指标。其中，仅使用主证作为判定标准的方法，由于判定为相似的数据量过少（仅含 126 条），在与输入医案进行相似度匹配时，输出结果为空，即表示模型判定在数据库中，不存在与输入医案相似语句。综合准确率、精确率及召回率 3 个评价指标发现，当使用不同医案相似度计算方法进行判定，模型推荐结果相差较大，使用本研究提出的多维特征融合计算方法的医案相似度评价指标效果最优，具体表现为准确率（89.55%）、精确率（94.51%）、召回率（83.97%）。仅使用主证及仅使用药物组成评价标准效果均较差，精确率及召回率均为 0，见表 4。

表 4 不同医案相似度计算方法的 ERNIE 模型评估指标结果

单位：%

计算方法	准确率	精确率	召回率
仅使用主证的计算方法	95.64	0.00	0.00
仅使用药物组成的计算方法	54.86	0.00	0.00
使用多维特征融合的计算方法	89.55	94.51	83.97

在推荐结果部分，仅使用药物组成判定方法针对“太阳阳明证”推荐 92 条数据，针对“太阳少阳证”推荐 91 条数据；使用多维度融合相似判定方法针对“太阳阳明证”推荐 40 条数据，针对“太阳少阳证”推荐 15 条数据。见表 5。其中，两判定方法推荐结果存在部分重合方剂，如在“太阳阳明证”部分，两方法均推荐《伤寒论》－桂枝汤方、《伤寒论》－桂枝加芍药汤方、《伤寒论》－大承气汤方等；在“太阳少阳证”部分，两方法均推荐《伤寒论》－桂枝汤方、《伤寒论》－小柴胡汤方、《伤寒论》－麻黄汤方等。

表 5 不同医案相似度计算方法的模型推荐结果和专家评价

输入医案	计算方法	推荐方剂条数	专家认可条数	百分比/（%）
发热、恶风，汗出，鼻鸣干呕，腹痛，腹满，不大便	仅使用主证的计算方法	0	0	0
	仅使用药物组成的计算方法	92	3	3.2
	使用多维特征融合的计算方法	40	20	50
发热、恶风，汗出，鼻鸣，干呕，不欲饮食，胸满，胁痛	仅使用主证的计算方法	0	0	0
	仅使用药物组成的计算方法	91	0	0
	使用多维特征融合的计算方法	15	12	80

据专家意见，发现在推荐结果中，使用多维度特征融合的判定方法推荐的方剂，专家认可度更高。在推荐结果中，对于“太阳阳明证”推荐，使用多维度特征融合的判定方法推荐的方剂，总推荐 40 条，专家认可条数为 20，占比为 50%；仅使用组成判定方法总推荐 92 条，专家认可条数为 3 条，占比仅达到 3%；仅使用主证的推荐方剂数为 0。对于“太阳少阳证”推荐，使用多维度特征融合的判定方法推荐的方剂，总推荐 15 条，专家认可条数为 12，占比为 80%；仅使用药物组成判定方法总推荐 91 条，专家认可条数为 0 条，不存在专家认可条目；仅使用主证的推荐方剂数为 0。可以发现，仅使用药物组成作为判定医案是否相似的标准无论是从模型指标及专家意见效果均较差，虽然在推荐结果时输出较多文本，但输出结果并不具有参考价值。

根据实验结果发现，仅使用主证和仅使用药物组成判定标准只从单一维度进行判定，忽略了临床实际的复杂性，所以推荐结果欠佳，而使用多维度融合相似判定方法较有临床参考意义。针对“太阳阳明证”推荐的方剂，涵盖了主治“恶风、汗出、鼻鸣、干呕”等太阳中风证的桂枝汤及其类方，同时涵盖了主治“腹痛、腹满、不大便”等阳明腑实证的大承气汤及其类方；针对“太阳少阳证”推荐的方剂，涵盖了主治“恶风、汗出、鼻鸣、干呕”等太阳中风证的桂枝汤及其类方，同时涵盖了主治“干呕、默默不欲饮食、胸满、胁痛”等少阳证的小柴胡汤。推荐的方剂基本符合临床辨证的选方思路，说明多维特征融合的方法比较智能地推荐了相关的方剂，不过这是在主治症状与条文症状输入比较一致的情况下。

针对同时出现“恶风、汗出、鼻鸣、干呕、默默不欲饮食、胸满、胁痛”等症状的太阳少阳证，系统并没有直接推荐出符合太阳少阳合病的“柴胡桂枝汤”这一方剂，原因是“柴胡桂枝汤”在原文条文中的描述与输入的症状不符合，但其病机是一致的。针对这种情况，需要将“柴胡桂枝汤”的主治证与上述症状相关联，以便系统可以推荐出更为恰当的方剂。另外，在“太阳少阳证”部分，推荐方剂中还出现了主治太阳伤寒的麻黄汤，属于干扰项，说明系统未对太阳中风和太阳伤寒的选方加以区分，表明目前系统模型对于更深层次的病机相似识别尚存在一定的不足，仍具有改进空间。

4.3 系统测试

从古代方剂智能推荐和计量转换系统首页进入方剂智能推荐页面，如图 3 所示。

首先，用户可以勾选页面左侧的复选框选择相关症状，也可以在页面上方文本搜索框中输入疾病症状，单击推荐方剂按

钮，即可获取多个相似的方剂信息，包括方剂来源、方剂组成、方剂剂量转换结果、方剂主治。例如，输入上文所用症状文本“发热、恶风，汗出，鼻鸣干呕，腹痛，腹满，不大便”，系统在网页上显示输出本研究所构建的特征融合的医案相似度判定模型的最优推荐结果：两种不同主治的桂枝汤方 -《伤寒论》，桂枝加芍药汤方 -《伤寒论》，大承气汤方 -《伤寒论》，小承气汤方 -《伤寒论》。推荐结果页面如图 4 所示。

然后，从古代方剂智能推荐和计量转换系统首页进入古代方剂的中药用量查询页，选择经方所属朝代、古籍和方剂名称，单击页面中的提交按钮后，即可以查询到古籍中的古代方剂原文信息和中药用量计量转换后的方药组成信息。如图 5 所示。

5 结语

本研究收集了先秦两汉至明清的古籍医案文本数据、各朝代的特殊计器及计量标准化转换规则数据，提出了一种融合主证 + 方剂名称 + 药物组成的多维特征医案相似度计算方法，在此基础上，建立数据库，设计并开发了一个古代方剂智能推荐和计量转换系统，该系统可以根据用户输入的症状，能够推荐匹配度较高的系列方剂，并按照系统的计量转换规则，将古代方剂的中药用量进行标准化转换，综合模型评价指标及专家意见，该方法的推荐结果更具有临床参考价值。本研究还存在一



图 3 古代方剂智能推荐和计量转换系统



图 4 系统推荐的方剂结果页面—太阳阳明证

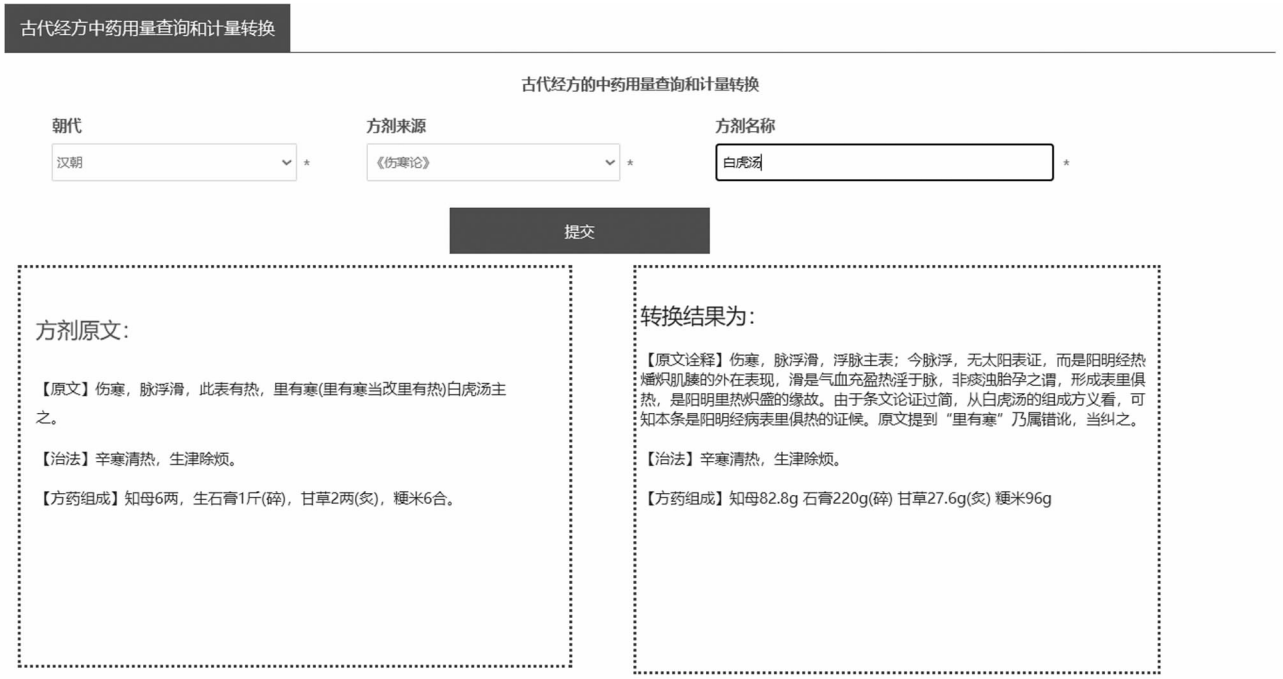


图 5 系统古代方剂查询和计量转换的结果页面

些不足，未来研究中将在以下方面进行改进：(1)将在输入的症状信息做同义词匹配和主证标准化转换；(2)在医案相似判定算法中将主证信息加权处理，以此提高模型的准确性；(3)由于本文实验主要依据中医古籍医案文本进行分析，存在方剂在古籍中出现的频次不等的情况，因此对于出现频次较少的医案文本存在模型推荐相似医案较少的情况，一定程度上会影响本文的研究结果，后续可考虑通过全面筛选补充数据集，使数据更均衡。

参考文献

[1] 董云薪，林耿，张清伟，等. 基于 Apriori 算法填充数据及改进相似度的推荐算法[J]. 计算机科学，2022，49(S2)：307－311.
[2] 郭文龙，罗熊，姜惠娟，等. 基于 TF－IDF 算法的方剂构成相似度可视化研究[J]. 中国中医药信息杂志，2019，26(7)：104－108.
[3] 操牡丹，何前锋，王柏. 中医药方剂相似度模型[J]. 计算机工程，2009，35(16)：275－277.
[4] 周晟，佟旭. 中医医案方剂多维相似度算法设计与探索[J]. 世界科学技术－中医药现代化，2021，23(8)：2864－2870.
[5] 朱志鹏，杜建强，刘英锋，等. 基于 LDA 主题模型的中医药方剂相似度计算[J]. 计算机应用研究，2017，34(6)：1668－1670，1676.
[6] 李新龙，刘岩，王宁，等. 方剂间相似性的量化表征方法及模型应用[J]. 中华中医药杂志，2022，37(4)：2120－2124.
[7] 邸剑，刘骏华，曹锦纲. 利用 BERT 和覆盖率机制改进的 HiNT 文本检索模型[J]. 智能系统学报，2024，19(3)：719－727.
[8] 刘栋梁，张嵩，张宁康，等. 基于 BERT 的航天术语标准化[J]. 南京航空航天大学学报，2022，54(S1)：109－114.
[9] 顾铮. 基于文本分类技术计算中医方剂相似度[J]. 微计算机信息，2010，26(12)：199－201.
[10] 李园白，崔蒙，杨阳，等. 方剂相似度探析[J]. 中华中医药学刊，2012，30(5)：1096－1097.
[11] 王安妮，于晓，张瑜，等. 基于方剂数据分析的中药用量历史变迁规律研究[J]. 世界科学技术－中医药现代化，2019，21(3)：342－347.
[12] 黄璐明，杨洪军，唐仕欢，等. 中药用量标准研究的思路与方法[J]. 中医杂志，2009，50(7)：651－654.
[13] 徐凤凯，吴汇天，曹灵勇.《伤寒杂病论》特殊计量药物换算考证[J]. 中华中医药杂志，2017，32(9)：4159－4162.
[14] 徐凤凯.《伤寒论》、《金匮要略》特殊计量药物的换算[D]. 杭州：浙江中医药大学，2016.
[15] 范吉平，程先宽. 经方剂量揭秘[M]. 北京：中国中医药出版社，2009.
[16] 傅延龄，宋佳，张林. 经方本原剂量问题研究[M]. 北京：科学出版社，2015.
[17] 王仲林. 基于孪生 NLP 预训练模型的 FAQ 检索式问答[D]. 镇江：江苏科技大学，2021.
[18] 施庆武，曾妮，吴承艳.《百病问对辨疑》学术思想探析[J]. 中国中医基础医学杂志，2023，29(9)：1434－1437.
[19] 章韵，房辉. 一种基于 ERNIE 和多特征融合的多模态情感分析方法[P]. CN202211589442.5[2023－10－03].
[20] 程强，杜中敏. 基于 ERNIE 预训练的中医临床病历分类[J]. 南阳师范学院学报，2023，22(1)：37－42.
[21] 杨文阳，孔科迪. 基于 ERNIE－BiLSTM 的社交网络文本情感分析[J]. 中国电子科学研究院学报，2023，18(4)：321－327.
[22] 常俊豪，武钰智. 基于 ERNIE_BiGRU 模型的中文医疗文本分类[J]. 电脑知识与技术，2022，18(1)：101－104.
[23] 崔博，贾兆年，姬鹏，等. 基于改进 EfficientNetV2 网络的脑肿瘤分类方法[J]. 吉林大学学报(理学版)，2023，61(5)：1169－1177.
[24] 蔡翟源，陈杰，奚雪峰，等. 电子病历的关系语义实体识别[J]. 苏州科技大学学报(自然科学版)，2023，40(3)：62－70.
[25] 董家慧子，谢忠，邱芹军，等. 融合容错机制的基于 Attention－Mask RCNN 地质表格信息抽取方法[J]. 地质科学，2023，58(3)：1147－1163.
[26] 梁博尧. 浅析最小二乘法及其在数学建模中的应用[J]. 中国新通信，2018，20(3)：197－199.