DOI:10.13193/j.issn.1673-7717.2025.01.007

基于改进遗传算法的 2 型糖尿病中医药有效处方推荐方法研究

刘红萍[1,2],杨杰[1,2],庞国明[3],李鹏辉[3],邢颖[1,2],吴敏[1,2],温宵宵[1,2],李洪皎[4],文天才[1,2]
(1. 中国中医科学院中医药信息研究所,北京 100700;2. 中国中医科学院中医药数据中心,北京 100700;
3. 开封市中医院,河南 开封 475000;4. 中国中医科学院中医临床基础医学研究所,北京 100700)

摘要:目的 改进遗传算法创新推荐度模型,以 2 型糖尿病(T2DM)电子病历数据为基础进行核心处方挖掘及有效处方推荐。方法 基于真实世界电子病历数据,构建 T2DM 患者有效处方集及不同证型原始处方集数据库;对遗传算法进行改进,优化适应度函数的构建,使提取出的核心有效处方朝着有效处方集的方向进化;基于核心有效处方与原始处方集的相似度关系,创新推荐度模型,通过遍历不同证型原始处方集进行处方推荐度挖掘。结果 共使用有效诊疗标准的处方 17 712 条,提取的核心有效处方中包含中药 37 种;最终挖掘出满足"推荐度≥85%"的处方 26 条,最大推荐度为 97.26%。结论 研究改进遗传算法提高了对中药处方集的特征提取和全局搜索能力,并通过提出新的推荐度模型进行临床处方决策推荐,提高了电子病历数据的利用率。

关键词:遗传算法;核心处方;处方推荐;真实世界数据;2 型糖尿病
中图分类号:R259.871 **文献标志码**:A **文章编号**:1673-7717(2025)01-0039-06

Study on Effective Prescription Recommendation Method of Traditional Chinese Medicine for Type 2 Diabetes Based on Improved Genetic Algorithm

LIU Hongping[1,2],YANG Jie[1,2],PANG Guoming[3],LI Penghui[3],XING Ying[1,2],WU Min[1,2],WEN Xiaoxiao[1,2],LI Hongjiao[4],WEN Tiancai[1,2]
(1. Institute of Information on Traditional Chinese Medicine,China Academy of Chinese Medical Sciences,Beijing 100700,China;
2. Traditional Chinese Medicine Data Center,China Academy of Chinese Medical Sciences,Beijing 100700,China;
3. Kaifeng Traditional Chinese Medicine Hospital,Kaifeng 475000,Henan,China;
4. Institute of Basic Research in Clinical Medicine,China Academy of Chinese Medical Sciences,Beijing 100700,China)

Abstract:*Objective* To improve the genetic algorithm and innovate the recommendation model,mine the core prescriptions and recommend the effective prescriptions based on the electronic medical records data of type 2 diabetes mellitus(T2DM). *Methods* Based on the real-world electronic medical records data,the database of effective prescriptions of T2DM patients and original prescriptions of different syndromes was constructed. The genetic algorithm was improved and the construction of fitness function was optimized,so the extracted core effective prescriptions can evolve towards the collection of effective prescriptions. Based on the similarity relationship between the core effective prescriptions and the collection of original prescriptions,the recommendation model was innovated,and the prescription recommendation was mined by traversing the collection oforiginal prescriptions of different syndromes. *Results* A total of 17 712 prescriptions with effective diagnosis and treatment standards were used,and 37 kinds of traditional Chinese medicines were extracted. Finally,26 prescriptions with "recommendation value≥85%" were excavated,and the maximal recommendation value was 97.26%. *Conclusion* In this study,the improved genetic algorithm has improved the ability of feature extraction and global search for prescription sets of traditional Chinese medicine,and proposed a new recommendation model for clinical prescriptions decision-making,thus improving the utilization rate of electronic medical records data.

Keywords:genetic algorithm;core prescription;prescriptions recommendation;real world data;type 2 diabetes

长期以来,中医药累积了大量的临床电子病历数据[1],这些来源于临床的庞大数据中蕴含着丰富的患者有效处方信息,但因碎片性、复杂性等特点难以被有效利用而应用于临床。

随着现代计算机技术的不断发展[2],许多研究人员开始采用复杂网络、遗传算法等方法在中医药核心处方策略方面进行了大量的研究。方旭鹏等[3]运用中药网络药理学分析系统(V1.0)软件,通过复杂网络的极大团求解算法(BK 算法)从 2372 张有效中药处方中提取出了 5 张核心处方,但该方法过程较为繁琐。YAN M 等[4]基于 60 例慢性荨麻疹患者,利用遗传算法(GA)构建了中西医结合治疗的优化模型,可根据患者症状积分进行诊疗方案的优化,但该研究样本量较小,并且没有对核心处方的方剂组成进行研究。李嘉旗等[5]基于 300 例肺癌患者的门诊电子病历数据,利用 GA 发现核心处方,并通过动物实验对核心处方的有效性进行验证,但该研究中其适应度函数是向着随机生成的新处方的方向进化的,没有进行电子病历数据的再利用。虽然上述证据显示在肺癌、慢性荨麻疹等领域已有关于核心处方挖掘的研究,但目前针对糖尿病的相关研究仍然较少。2 型糖尿病(T2DM)在所有糖尿病个体中,占

基金项目:国家自然科学基金面上项目(82374624);北京市自然科学基金面上项目(7232306);中国中医科学院自主选题项目(ZZ16-XRZ-107-SJ)
作者简介:刘红萍(1998-),女,安徽六安人,硕士在读,研究方向:2 型糖尿病中医药真实世界研究。
通讯作者:文天才(1977-),男,四川绵阳人,正高级工程师,研究员,博士研究生导师,博士,研究方向:智慧医学与真实世界研究。E-mail:wtcsnake@163.com。

95%以上[6-7]。近年来糖尿病的患病人数大大增加,国际糖尿病联合会(IDF)预计 2045 年全球糖尿病患病人数将达到 6.29 亿[8],因此 T2DM 临床有效处方策略的研究十分重要。

本文提出了一种改进 GA 适应度函数,并创新了一种推荐度模型。基于 T2DM 患者电子病历数据挖掘其核心有效处方,并对原始数据再利用,根据患者中医证型进行处方推荐度分析。以期提高真实世界电子病历数据的利用率,同时为临床处方决策提供一定的参考。

1 资料与方法

1.1 整体研究设计 模型的构建主要有以下几个步骤:(1)根据有效诊疗情况评价标准,对全部患者的所有中药处方做初始筛选,得到初始有效处方集;(2)利用遗传算法的并行全局搜索寻优能力,对初始有效处方集进行优化搜索,挖掘出最具有患者有效处方集特征的核心有效处方(CEP);(3)基于寻优得到的 CEP,根据患者中医证型的不同对临床处方做推荐度分析。具体的研究思路见插页Ⅸ图 1。

1.2 数据来源及预处理 (1)数据来源:选取 2017 年 1 月—2021 年 7 月于河南省开封市中医院内分泌科就诊的患者,回顾性采集患者入院及住院期间临床资料,包括病历资料、中药处方、空腹血糖(FBG)及餐后血糖(PBG)数据,并基于 Execl 2019 软件构建 T2DM 患者中西医临床信息数据库。电子病历数据筛选标准如下:①西医首要诊断明确为 2 型糖尿病;②病历中药处方、中医证型、FBG 及 PBG 数据完整;③年龄≥18 岁。本研究已获得开封市中医院伦理委员会批准(伦理批件号:2022 - ky - 006)。

(2)数据预处理:对于中药处方数据,使用 R 4.2.3 软件,参照《中华人民共和国药典》(2010 版)[9]中所规定的中药名进行标准化处理,将原始数据使用的中药俗名、炮制加工名等全部规范化为上述标准中的正规名称,如黑顺片标准化为附子;醋延胡索标准化为延胡索;党参片标准化为党参等。为降低真实世界数据的不完整性、含噪声及杂乱性对挖掘结果的影响,对数据进一步清洗,提取满足“中药频数≥500 AND 处方药数≥3”的数据。对于中医证型数据,参考《糖尿病中医药临床循证实践指南(2016 版)》[10]《糖尿病中医诊疗标准》[11]《消渴病(糖尿病)中医分期辨证与疗效评定标准》[12],由本领域专家进行规范化,提取满足“证型频数≥200”的记录,最终得到气阴两虚证、脾虚痰湿证、肝郁气滞证、气虚血瘀证、湿热内蕴证 5 组中医证型。对于数据库中的计量资料,使用 Python 软件计算出治疗后患者 FBG 及 PBG 降幅等研究所需数据。

(3)有效诊疗情况评价:空腹血糖是 2 型糖尿病诊断的主要标准之一[13],且空腹血糖与餐后血糖因其方便、可重复性强的特点常被用作疗效评定的标准。根据 T2DM 真实世界数据特点及临床专家意见,参照《中国 2 型糖尿病防治指南》(2020 年版)[11]《中药新药临床研究指导原则》(2002 版)[14],拟定有效诊疗情况评价方法如下:①有效:空腹血糖或餐后血糖值下降超过治疗前的 20%;②无效:空腹血糖或餐后血糖值无下降,或下降未达到有效标准。

1.3 基于改进遗传算法的核心有效处方挖掘模型构建 在有效诊疗情况评价的基础上,筛选达“有效”标准的患者中草药处方,并进一步利用遗传算法挖掘出核心有效处方。GA 是一种模拟达尔文生物进化论的自适应全局优化算法,它通过模拟种群进化过程中的选择、交叉、变异等遗传操作,根据适应度的原则,优胜劣汰,经过不断地循环迭代,最终产生期望的最优解或近似最优解[15]。遗传算法通过模仿生物进化的过程反复地进行遗传迭代,它的进化以种群为单位,而种群则由一系列包含着不同基因的染色体组成。基于遗传算法的核心有效处方挖掘模型主要可以分为以下 6 个步骤进行构建。

(1)随机产生初始处方集。在本研究中,处方集对应算法中的种群,处方对应染色体,中药对应基因。如插页Ⅸ图 2 所示为处方集、处方与中药之间的关系。在 GA 算法开始时,没有对原始数据集进行分析,不能判断出哪些处方是更加有效的。因此在模型最开始构建时,将通过算法随机生成一个初始处方集,它将作为算法的初代种群。该处方集由一系列的二进制数串表示,这些二进制数串上的 1 代表该位置的中药存在,0 代表该位置的中药缺失。假设用二进制数串的第一位代表黄芪,那么如果处方 1 对应的二进制数串第一位数字为 1,则代表处方 1 中包含黄芪这味中药,反之则不包含。

(2)处方集的适应度计算。适应度是处方集中的处方被选择进入下一代的唯一标准,它代表了处方的优劣。在遗传算法中,适应度越大的处方具有更大的概率在迭代过程中存活下来,因此适应度的设计和计算对遗传算法的性能具有很大的影响。本研究在传统遗传算法的基础上,创新了处方的适应度计算,计算公式如公式(1):

$$\text{fitness}\ \frac{1}{N}\sum_{i=1}^{n}\frac{\text{count}(\text{core}\cap R_i)}{\text{count}(R_i)} \tag{1}$$

式中,core 表示种群中的任意一个处方中药的集合,N 为原始数据集中处方的数量,R_i 代表原始数据集中的第 i 张处方中药的集合,count( ) 为计数函数,统计集合的中药味数。fitness 函数表明了当前计算的处方与原始数据集的关联程度,适应度值越大,当前处方与原始数据集的关联性越高。从公式(1)可以看出,适应度值的计算是通过遍历原始数据,将当前处方与原始数据集中的每一个处方都做比对,计算两者的吻合度。因此,经过上述适应度函数迭代出的核心处方将是最能展现原始数据集特征的全新处方。

(3)选择操作。选择操作是通过指定的选择算子从父代处方集中,按照各处方适应度值的大小,选出适应度高的处方进入交配池中。在 GA 算法迭代的过程中,通过不断地进行这一操作,淘汰适应度低的处方,选择适应度高的处方,从而促使整个处方集朝着更优势的方向进化。选择操作的基本原理就是对于适应度更强的处方,其被选中的概率也就越大。本研究采取的方法为“轮盘赌”选择法,该方法原理简单并且具有很好的实用效能,它是一种基于比例的选择方法,与选择算子的基本原理相吻合。在该方法中,中药处方被选中的概率与其自身的适应度值呈正相关。假设在一个处方集中有 NP 个中药处方,其中任意一处方 i 的适应度值为 f_i,那么该处方被选中进入下一次遗传迭代的概率为 p_i,其计算公式如公式(2):

$$p_i = \frac{f_i}{\sum_{i=1}^{NP} f_i}\ (i = 1,2,\cdots,NP) \tag{2}$$

从上式可以看出,适应度越高的处方被选中进入下一代的概率就越大。基于上式,可以建立如插页Ⅸ图 3 所示轮盘。

该轮盘被分割为 NP 个部分,每一个部分对应处方集中的一个处方,每个处方在轮盘上所占的区域面积将根据上式计算的概率成比例表达出来。当进行选择操作时,可以让该轮盘旋转起来,然后将固定指针指向的那一片区域选为第一个亲本,按照这种方法,继续选择进入下一代遗传迭代的其他亲本。

(4)交叉操作。交叉操作是指从交配池中随机选择成对的处方组合,然后再随机选择这两个处方的多个或者单个相同位置上的中药,以指定的交叉概率进行交换。引入交叉操作的目的是为了提高 GA 算法的全局搜索能力,从而保证处方集的多样性,否则处方集将无法进化。如插页Ⅸ图 4 所示分别为单点交叉和多点交叉的示意图,可以看出,通过交叉操作可以使

处方集产生全新的处方，保证了处方集一直处于进化的过程中，而不是处于停滞状态。

(5)变异操作。变异操作是对处方集中的每一个处方，以指定概率(即变异概率)将二进制数串处方上的某一位置或者某一些位置上的中药改变为其他的等位值，即将该位置上的数值 1 更改为 0，或者将数值 0 更改为 1。将对应位置上的数值取 1 或者取 0 即是表示将该位置对应的中药从处方中添加或者是去除，从而实现处方的更新。变异操作通常发生的概率较低，当处方集中所有处方的适应度值趋于一致时，通过选择和交叉操作很难再产生新的差异化较为明显的处方，而此时通过变异操作可以保证新处方的产生，能够尽可能地避免 GA 算法最后收敛到局部最优解(见插页Ⅸ图 5)。但是，需要注意的一点是，太大的变异概率也会导致 GA 算法的进化过程趋于随机搜索。

(6)终止条件判断。遗传算法终止的条件是适应度值达到期望值，或者迭代次数达到算法指定的最大迭代次数。在本研究中，由于并不存在已知的最优处方，因此无法对适应度值的优良阈值进行计算，从而无法通过适应度值的实际大小判断遗传算法是否达到期望的要求。因此，本文通过选择设定一个合适的最大迭代次数来实现结束算法的目的，当算法迭代次数达到最大值时，选择相对大的适应度值对应的处方作为本次算法优化的最优处方。通常最大迭代次数越大，算法收敛的效果越好。但是当迭代超过一定次数时，算法将不再进化，即已经收敛到最优解。因此，本文在经过基于蒙特卡洛方法的实验后，设定最大迭代次数为 500。

1.4 *基于电子病历数据的处方决策推荐算法设计* 以核心有效处方为依据，利用 Python 语言计算电子病历数据中的所有原始处方(OP)与核心有效处方的相似度(SV)，计算公式如公式(3)：

表 1 不同证型处方记录筛选

中医证型	处方数/(条)	患者数/(例)
气阴两虚证	11 184	1838
脾虚痰湿证	4083	715
气虚血瘀证	3022	319
肝郁气滞证	1811	351
湿热内蕴证	1384	276
总计	21 484	3499

表 2 核心有效处方组成及平均剂量 单位：g

序号	中药	平均剂量	序号	中药	平均剂量
1	黄芪	41	20	北沙参	19
2	首乌藤	37	21	延胡索	19
3	琥珀	34	22	麦芽	18
4	白茅根	34	23	杜仲	18
5	丹参	32	24	麦冬	17
6	牛膝	31	25	枸杞子	15
7	山慈菇	30	26	川芎	15
8	金樱子	29	27	化橘红	14
9	生地黄	29	28	牡丹皮	13
10	山茱萸	27	29	豆蔻	13
11	连翘	25	30	桃仁	13
12	党参	24	31	郁金	13
13	白芍	24	32	黄芩	11
14	石决明	23	33	枳实	11
15	黄连	23	34	槟榔	10
16	龟甲	22	35	皂角	9
17	赤芍	22	36	干姜	8
18	防己	22	37	沉香	8
19	酒大黄	21			

表 3 T2DM 患者推荐处方中药组成

证型	序号	推荐处方中药组成	RV/(%)
气阴两虚证	RP1	红芪，党参，茯苓，大黄，生地黄，生石膏，山茱萸，乌梅，山楂，鸡内金，知母，山药	90.56
	RP2	枸杞子，泽泻，山茱萸，生地黄，茯神，葛根，决明子，党参，白术，连翘，天花粉，红芪	90.56
	RP3	桃仁，党参，生地黄，当归，葛根，黄精，白芍，桂枝，淫羊藿，甘草，红芪	90.56
	RP4	瓜蒌，姜半夏，熟地黄，红花，当归，木香，茯苓，川贝母，化橘红，牡蛎，甘草，黄芪	88.50
	RP5	熟地黄，太子参，丹参，五味子，山茱萸，葛根，茯苓，黄芪，牡丹皮，泽泻，山药，玄参，麦冬	85.48
脾虚痰湿证	RP6	苍术，陈皮，白术，猪苓，泽泻，茯苓，厚朴，甘草，姜半夏，桂枝	97.26
	RP7	茯苓，苍术，桂枝，白术，升麻，牛膝，佩兰，陈皮，丹参，葛根，姜半夏，薏苡仁，泽泻，厚朴，猪苓	95.10
	RP8	升麻，甘草，桂枝，陈皮，泽泻，苍术，茯苓，白术，薏苡仁，皂角，猪苓，姜半夏	90.57
	RP9	陈皮，猪苓，厚朴，苍术，土茯苓，桂枝，茯苓，干姜，泽泻，甘草，白术，麦冬，四季青	88.47
	RP10	薏苡仁，白术，泽泻，厚朴，火麻仁，甘草，茯苓，天花粉，黄连，猪苓，姜半夏，牛膝，升麻，陈皮，苍术	86.46
气虚血瘀证	RP11	甘草，鸡血藤，生地黄，黄芪，太子参，钩藤，泽泻，山药，天麻，牡丹皮，地龙，茯苓，水蛭，山茱萸	95.10
	RP12	生地黄，大黄，玄参，厚朴，桃仁，川芎，当归，酒苁蓉，牵牛子，白芍，莱菔子，麦冬	90.57
	RP13	白术，大黄，桂枝，猪苓，厚朴，牛膝，柴胡，泽泻，苍术，茯苓，姜半夏，薏苡仁，陈皮，枳实，升麻	88.47
	RP14	红芪，天花粉，茯苓，赤芍，当归，桃仁，党参，葛根，山茱萸，红花，川芎，白术	86.46
	RP15	川芎，桃仁，白芍，白术，当归，赤芍，鬼箭羽，太子参，葛根，牛膝，甘草，生地黄，黄芪，山药，桂枝	86.46
肝郁气滞证	RP16	当归，茯神，薄荷，知母，酸枣仁，白术，白芍，仙鹤草，柏子仁，山药，牡丹皮，栀子，丹参，柴胡，首乌藤	95.10
	RP17	枸杞子，菊花，丹参，太子参，麦冬，葛根	90.57
	RP18	丹参，白术，当归，柴胡，升麻，茯苓，牛膝，苍术，白芍，甘草，砂仁，檀香，牡丹皮，栀子，薄荷，生姜	88.47
	RP19	水蛭，枸杞子，菊花，太子参，麦冬	87.34
	RP20	木香，甘草，桂枝，白芍，鸡内金，党参，百合，四季青，茯神	86.46
	RP21	鱼腥草，葛根，山楂，鸡内金，白芍，黄芩，党参，大黄，四季青，柴胡，蒲公英	86.46
湿热内蕴证	RP22	黄连，生石膏，姜半夏，薏苡仁，苍术，淡豆豉，枳实，石菖蒲，芦根，栀子，黄柏，厚朴，牛膝，泽泻	94.76
	RP23	淡豆豉，石菖蒲，栀子，黄柏，黄连，姜半夏，薏苡仁，厚朴，升麻，牛膝，生石膏，芦根，天麻，白术	91.38
	RP24	黄连，栀子，石菖蒲，芦根，升麻，白术，党参，薏苡仁，姜半夏，黄柏，苍术，牛膝，丹参，茯苓，酸枣仁	90.46
	RP25	牛膝，芦根，枳实，石菖蒲，升麻，厚朴，姜半夏，薏苡仁，苍术，栀子，黄柏，黄连	88.45
	RP26	淡豆豉，石菖蒲，芦根，姜半夏，薏苡仁，厚朴，牛膝，黄连，麦冬，栀子，黄柏，天花粉	87.80

注：RP：推荐处方，RV：推荐度。

表 4 不同证型处方集中药平均剂量 单位:g

中药名	气阴两虚证	脾虚痰湿证	气虚血瘀证	肝郁气滞证	湿热内蕴证	均值	中药名	气阴两虚证	脾虚痰湿证	气虚血瘀证	肝郁气滞证	湿热内蕴证	均值
仙鹤草	60	70	69	68	74	68	黄连	22	23	21	25	19	22
黄芪	41	39	41	38	39	40	莱菔子	21	21	21	19	28	22
首乌藤	35	41	32	39	43	38	地龙	20	21	22	20	19	21
牡蛎	35	40	37	34	38	37	泽泻	15	25	16	21	22	20
芦根	30	34	29	29	37	32	菊花	19	20	21	19	20	20
薏苡仁	30	33	29	32	33	31	麦冬*	18	19	22	19	19	19
鸡血藤	29	32	31	36	27	31	百合	20	18	13	20	23	19
丹参	29	35	31	29	29	31	淫羊藿	18	21	15	23	15	19
牛膝	27	33	28	31	33	31	乌梅	17	25	13	22	13	18
太子参	29	29	31	29	30	30	钩藤	16	20	19	18	17	18
玄参	27	29	34	27	29	29	当归	18	17	18	14	21	18
白术*	28	31	29	26	31	29	牡丹皮	14	19	15	13	23	17
鱼腥草	30	30	28	26	30	29	枸杞子	16	18	18	16	14	17
酒苁蓉	28	36	29	30	21	29	柴胡	14	14	17	16	17	16
葛根	27	29	29	26	31	28	牵牛子	15	15	15	15	15	15
蒲公英	26	30	29	27	29	28	红花	12	13	12	20	15	14
黄精	25	29	28	27	30	28	天麻	14	15	15	12	13	14
生地黄	28	27	28	28	29	28	川芎	13	13	14	15	13	14
猪苓	28	28	29	26	28	28	桃仁	14	13	12	13	16	14
土茯苓	28	30	22	28	30	28	大黄	12	14	15	14	13	14
决明子	28	31	26	25	28	28	化橘红	12	16	10	20	10	14
四季青	27	28	25	29	28	27	栀子	11	13	18	11	10	12
茯神	24	28	27	28	29	27	佩兰	12	10	11	12	11	12
天花粉	27	26	28	27	27	27	桂枝	12	8	14	12	12	12
火麻仁	25	26	32	26	28	27	厚朴	13	10	13	10	11	11
红芪	27	28	27	29	25	27	五味子	12	11	11	13	10	11
茯苓*	23	30	24	27	30	27	黄芩	11	12	11	11	11	11
山茱萸	25	27	27	26	27	26	枳实	11	11	12	10	10	11
生石膏	27	26	26	25	26	26	姜半夏	10	11	10	10	11	10
酸枣仁	26	22	30	24	26	26	陈皮	11	10	11	10	11	10
山药	24	27	24	26	27	26	黄柏	10	10	11	11	10	10
连翘	24	24	26	24	27	25	石菖蒲	11	10	10	11	8	10
柏子仁	24	25	29	23	22	24	檀香	9	9	10	9	10	9
白芍	22	25	25	24	25	24	薄荷	10	9	7	10	10	9
山楂	21	25	19	29	25	24	皂角	9	8	9	10	9	9
鬼箭羽	23	23	24	24	22	23	干姜	7	7	8	7	8	8
熟地黄	23	27	20	24	23	23	川贝母	7	8	10	6	7	7
知母	19	25	19	26	26	23	生姜	9	7	8	7	6	7
党参	26	22	24	24	21	23	砂仁	6	6	7	6	7	6
淡豆豉	22	24	16	26	29	23	升麻	7	7	6	7	5	6
鸡内金	21	26	23	22	22	23	甘草	7	5	7	6	7	6
赤芍	20	23	22	26	21	23	木香	6	7	5	6	7	6
苍术	22	27	21	22	19	22	水蛭	6	5	6	6	6	6
瓜蒌	22	26	24	18	19	22							

注:* 表示 5 组证型推荐处方共同包含的中药种类。

$$SV_i = \frac{count(CEP \cap OP_i)}{count(CEP) + count(OP_i) - count(CEP\&OP_i)} \quad (3)$$

其中,SV_i 代表第 i 条原始处方与 CEP 的相似度,CEP 为核心有效处方,OP_i 为第 i 条原始处方。相似度值越大表示与核心有效处方吻合程度越高。

考虑到推荐度(RV)的临床实际应用价值,对相似度做归一化处理[16],如公式(4):

$$RV_i = \frac{SV_i - Min(SV)}{Max(SV) - Min(SV) + (1e-3)} \times 100\% \quad (4)$$

其中,Max( )为求集合的最大值函数,Min( )为求集合的最小值函数,加上 1e-3 是为了避免出现 RV = 1 的情况[17]。推荐度值越大表示原始处方与核心有效处方的相似度排名越高,在临床医生处方决策时推荐排名越靠前。

2 结果

2.1 *有效诊疗情况及患者处方记录* 原始数据中满足纳入标准的处方记录共计 25 588 条,涉及患者 4684 例,其中达有效诊疗标准的处方记录共 17 712 条,涉及患者 2667 例,包含中药 185 种。

根据上述中医证型提取标准,最终得到 T2DM 患者不同中医证型的处方记录共计 21 484 条,涉及患者 3499 例。见表 1。

2.2 *核心有效处方组成及平均剂量* 利用改进遗传算法对上述满足有效诊疗标准的处方进行寻优,得到核心有效处方,包

含中药共 37 种。表 2 列出了核心有效处方所包含的中药种类及其平均剂量(Ad)。其中平均剂量最大的中药为黄芪(Ad = 41 g);其次为首乌藤(Ad = 37 g)。

2.3 不同中医证型处方推荐度分析 分别将不同证型的所有原始处方通过推荐度(RV)公式进行遍历,计算每条处方的推荐度,输出"RV≥85%"的记录,最终得到满足条件的推荐处方记录共 26 条,最大推荐度为 97.26%(RP6)。其中气阴两虚证 5 条,脾虚痰湿证 5 条,气虚血瘀证 5 条,肝郁气滞证 6 条,湿热内蕴证 5 条。5 组证型共同涉及的中药为白术、茯苓、麦冬,其平均剂量在不同证型中略有不同。见表 3、表 4。

3 讨论

3.1 基于改进遗传算法的处方推荐模型架构 中医方剂处方的构成与遗传算法染色体的组成结构具有相似的数学特点,各个方剂的药味加减操作与遗传算法染色体的基因交叉、变异操作类似。因此可以按照遗传算法的种群进化过程来对真实世界的处方集进行迭代寻优,最终产生的处方不仅能够体现处方集合的特征,还可能是未被发现的遣方用药新规律。本研究对遗传算法进行改进,优化了适应度函数的构建,使新生成的核心有效处方向着有效处方集的方向进化。并进一步基于核心有效处方与原始处方集的相似度关系,搭建了一个全新的处方推荐模型,利用上述 CEP 对原始处方逐个进行推荐度分析。虽然本研究患者电子病历数据来源较为单一,尚不足以代表 2 型糖尿病全部患者群体的药物使用情况,但是本研究所提出的中药处方推荐模型在对真实世界电子病历数据的处理方法上是创新的,在发挥核心有效处方价值的同时,通过推荐度模型更加充分且高效地利用了现有的海量电子病历数据,为临床处方决策提供了参考。

3.2 有效诊疗情况评价及核心有效处方分析 研究显示[18-19],FBG 和 PBG 可以在一定程度上对糖化血红蛋白(HbA1c)水平进行评估而进行疗效评价,二者也是临床常用的血糖评价指标[20]。基于原始数据特点及专家意见,结合 FBG、PBG 数据对 T2DM 患者进行有效诊疗情况评价,最终得到满足标准的处方记录共17 712条,占初始处方记录的 69.45%,数据量足以支撑研究所用。核心有效处方中包含的黄芪、丹参、牛膝、生地黄、麦冬等与庞国明教授"序贯三法"[21-22]中所用专证专方一致;调气血之党参、丹参、黄芪与调脾胃之麦芽、枳实等与孙光荣教授所提出的"平腑调代方"[23-24]组成原理相似;黄连平均剂量为 23 g,与仝小林教授治疗肺胃热盛型糖尿病"态靶"同调[25]使用剂量一致(15 ~ 30 g)。

4 结论

本文基于 2 型糖尿病真实世界电子病历数据,对遗传算法进行改进,优化适应度函数,并结合推荐度算法构建了中药处方推荐模型。该模型所具备的中药处方集的特征提取和全局搜索能力,大大提高了真实世界电子病历数据的利用率。将预处理后的 2 型糖尿病患者电子病历数据投入模型当中,对有效处方进行挖掘,并进一步分析不同证型下患者处方的推荐度。结果显示核心有效处方与临床中医专家所开具的处方在药物及剂量上相符,以此为基础的患者处方推荐度最大可达 97.26%。研究表明该模型处理数据效率高,且易于推广,可广泛应用于其他疾病领域,从而为中医药处方辅助决策的发展提供思路和方法。

参考文献

[1] 吴宗友, 白昆龙, 杨林蕊, 等. 电子病历文本挖掘研究综述[J]. 计算机研究与发展, 2021, 58(3): 513 - 527.
[2] SHEN Y, ZHOU J, HU G. Practical use of electronic health records among patients with diabetes in scientific research[J]. Chin Med J, 2020, 133(10): 1224 - 1230.
[3] 方旭鹏,徐振晔,王中奇,等. 基于复杂网络分析肺癌术后优势人群有效核心处方的发现及实验验证[J]. 世界科学技术 - 中医药现代化,2023,25(4):1224 - 1229.
[4] YAN M, YE F, ZHANG Y, et al. Optimization model research on efficacy in treatment of chronic urticaria by Chinese and Western medicine based on a genetic algorithm[J]. J Tradit Chin Med, 2013, 33(1): 60 - 64.
[5] 李嘉旗, 杨铭, 焦丽静, 等. 基于遗传算法的中医药治疗肺癌核心有效处方的发现及实验研究[J]. 中华中医药杂志, 2018, 33(9): 4143 - 4146.
[6] XU L, LI Y, DAI Y, et al. Natural products for the treatment of type 2 diabetes mellitus: Pharmacology and mechanisms[J]. Pharmacol Res, 2018, 130: 451 - 465.
[7] THOMAS C C, PHILIPSON L H. Update on diabetes classification[J]. Med Clin North Am, 2015, 99(1): 1 - 16.
[8] PANG G M, LI F X, YAN Y, et al. Herbal medicine in the treatment of patients with type 2 diabetes mellitus[J]. Chin Med J (Engl), 2019, 132(1): 78 - 85.
[9] 国家药典委员会. 中华人民共和国药典:一部[S]. 北京: 中国医药科技出版社, 2010: 42 - 182, 197 - 361.
[10] 仝小林. 糖尿病中医药临床循证实践指南: 2016 版[M]. 北京: 科学出版社, 2016: 22 - 26.
[11] 仝小林,倪青,魏军平,等. 糖尿病中医诊疗标准[J]. 世界中西医结合杂志,2011,6(6): 540 - 547.
[12] 吕仁和, 张洁荣, 高彦彬. 消渴病(糖尿病)中医分期辨证与疗效评定标准[J]. 中国医药学报, 1993(3): 54 - 56.
[13] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南(2020 年版)[J]. 中华糖尿病杂志, 2021, 13(4): 315 - 409.
[14] 郑筱萸. 中药新药临床研究指导原则[M]. 北京: 中国医药科技出版社, 2002: 233 - 237.
[15] LIU P, EL BASHA M D, LI Y, et al. Deep evolutionary networks with expedited genetic algorithms for medical image denoising[J]. Med Image Anal, 2019, 54: 306 - 315.
[16] 汪敏, 徐英豪, 朱习军. 基于改进遗传算法优化 BP 神经网络的糖尿病并发症预测模型[J]. 计算机与现代化, 2022(11): 69 - 74.
[17] 包子阳, 余继周, 杨杉. 智能优化算法及其 MATLAB 实例[M]. 2 版. 北京: 电子工业出版社, 2018: 20 - 31.
[18] 黄双双, 冉孟婷, 吕艳春. 仙鹤草对糖尿病小鼠血糖的影响研究[J]. 遵义医学院学报, 2017, 40(4): 378 - 382, 388.
[19] 邓志军, 李阿荣, 罗永佳, 等. 荔枝核改善胰岛素抵抗治疗 2 型糖尿病的网络调控作用研究[J]. 广东药科大学学报, 2020, 36(6): 834 - 839.
[20] 宋聪琳, 张威娜. 黄连解毒汤加减治疗 2 型糖尿病的临床疗效[J]. 辽宁中医杂志, 2022, 49(4): 110 - 113.
[21] 庞国明, 张芳, 王凯锋, 等. 中药"序贯三法"治疗 2 型糖尿病患者 252 例临床疗效的真实世界研究[J]. 中医杂志, 2021, 62(18): 1612 - 1616, 1621.
[22] 庞国明,张芳,李鹏辉,等. 基于真实世界研究探析纯中药"序贯三法"对 252 例 2 型糖尿病患者胰岛功能的影响[J]. 中华中医药学刊,2021,39(10):253 - 258.
[23] 庞国明, 高言歌, 张侗, 等. 国医大师孙光荣教授学术思想探析[J]. 中医研究, 2022, 35(5): 1 - 4.
[24] 陈弘东,王耀献,刘伟敬,等. 平腑调代方对肥胖 2 型糖尿病小鼠脂代谢和炎症及氧化应激水平的影响[J]. 中国医药,2022,17(8):1238 - 1241.
[25] 刘晟, 苟筱雯, 赵林华, 等. 态靶辨证在肺胃热盛型糖尿病中的运用——白虎汤加黄连、桑白皮[J]. 辽宁中医杂志, 2020, 47(3): 4 - 6.

基于改进遗传算法的 2 型糖尿病中医药有效处方推荐方法研究

（正文见 39－43 页）

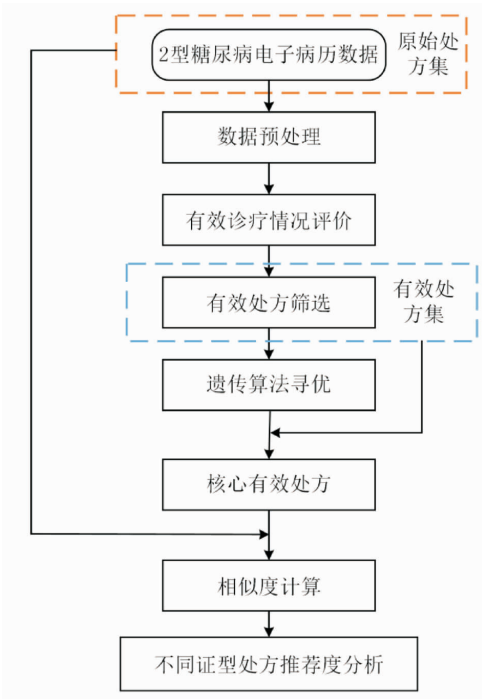


图 1　研究思路

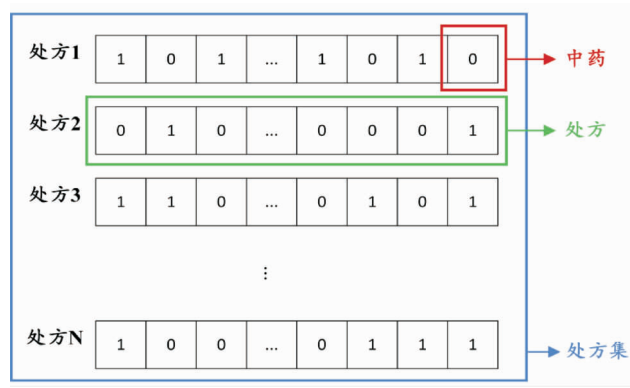


图 2　遗传算法中处方集、处方与中药之间的关系

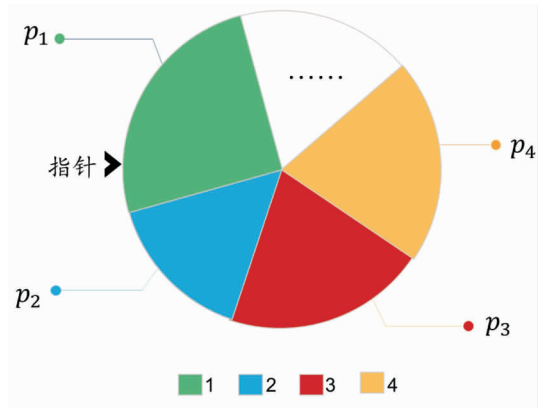


图 3　轮盘赌选择法示意图

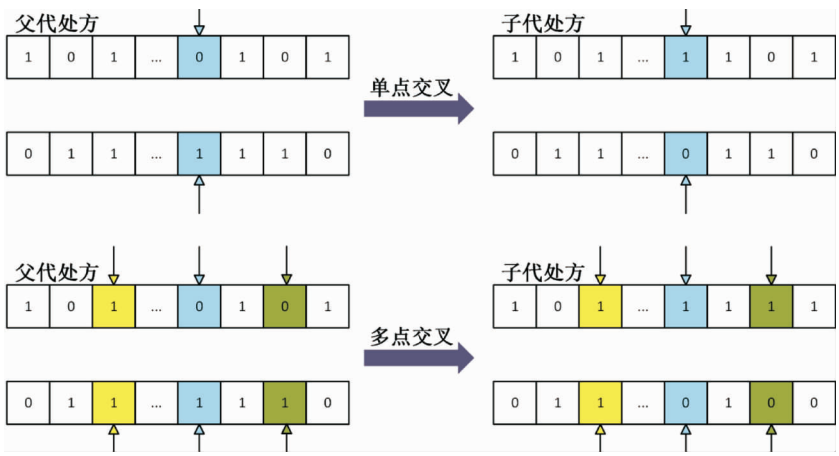


图 4　交叉操作原理



图 5　变异操作原理