基于遗传算法与KNN融合的中医体质量表简化研究*

管树桃[1]，李红岩[1]，郎许锋[1]，李 灿[1]，周作建[1**]，胡孔法[1,2]，战丽彬[3]

（1. 南京中医药大学人工智能与信息技术学院 南京 210023；2. 江苏省中医药防治肿瘤协同创新中心 南京 210023；3. 辽宁中医药大学中医药创新工程技术中心 沈阳 110847）

摘 要：目的 针对中医体质量表在评估个人体质时条目多和填写时间久的问题，研究利用人工智能技术进行属性选择，帮助构建中医体质量表简短版本。方法 分析由江苏省中医院体检科提供的中医体质数据，其中有特定的目标变量作为体质类型的分类。采用遗传算法的特征选择、交叉验证和KNN分类算法作为过滤器筛选问题，并通过问题子集规模、KNN分类准确率和填写时间评估效果。结果 该方法选择出具有31个问题的中医体质量表简短版本，且在模型中的分类平均准确率为86.16%，时间提快了48.5%。结论 该算法可以有效地找出较好的问题子集，实现降维并有一定的准确性，从而帮助简化中医体质量表。

关键词：中医体质 中医体质量表 遗传算法 KNN算法

doi: 10.11842/wst.20221118011 中图分类号：R-058 文献标识码：A

体质是人体生命过程中，在先天禀赋和后天获得的基础上形成的形态结构、生理功能和心里状态方面综合的、相对稳定的固有体质[1]。掌握个体的体质异同点与体质特性，从“体质可分、体病相关、体质可调”三方面管理全生命周期健康[2]，在临床实践中发挥效用，且有助于将疾病从“治病”到“防病”的转变。目前各大医院和机构广泛根据不同体质表现以及疾病发展过程中的体质表现变化进而分析中医体质类型与疾病之间存在的关联，而所依据的体质类型判定标准是王琦教授提出的中医体质九分法。九分法[3]将中医体质划分为平和质、阳虚质、阴虚质、气虚质、痰湿质、湿热质、血瘀质、特禀质、气郁质。且王琦[4]依据体质九分法编制了相应的有平和质、气虚质等9个亚量表，各亚量表含7-10个条目共60个条目且每个条目有5个选择项的《中医体质量表》。自2006年在国内外发表后一直备受关注，更是成为中医诊断和治疗的标准，更是被广泛运用在健康管理和临床实践中。随着中医体质学的发展，王琦教授团队以及相关学者在《中医体质量表》基础上进一步研究，开发出适合不同人群和语言的量表[5-9]，婴幼儿中医体质量表也在思考构建中[10-13]。为了不影响体质评价或判定效率，在量表应用中需要注意使用不规范、研究样本量不足、混淆量表与《中医体质分类与判定》等问题[14]。但量表条目较多，应用者在填写时花费时间长，甚至填写过程中失去耐心进而影响选择[15]，也是值得注意与解决的一个问题。因此，如何缩短量表条目数且减少填写时间是关键。

针对量表的制定与简化，传统研究中一般采用德尔菲法、统计学法和结合经典测试理论与项目反应理论两种方法来指导对量表条目的筛选。杨江等[16]基于改良德尔菲法筛选支气管哮喘急性发作风险因素且形成调查表；白一帆等[17]采用德尔菲法筛选学龄前儿童中医体质类型与定义指标，给建立学龄前儿童中医体质量表提供依据。刘竞男等[18]应用统计学方法对血

收稿日期：2022-11-18

修回日期：2023-05-29

* 国家科学技术部重点研发计划中医药现代化研究专项（2018YFC1704400）：阴虚证辨证标准的系统研究，负责人：周作建；国家自然科学基金委员会面上项目（82074580）：基于知识图谱的现代名老中医诊治肺癌用药规律及其机制研究，负责人：胡孔法。

** 通讯作者：周作建，研究员，硕士研究生导师，主要研究方向：服务计算，物联网，医疗卫生信息化。

虚中医疗效评价量表从条目的可行性、敏感性、内部一致性等方面进行筛选剔除，结合专家意见，形成最终量表。经典测试理论与项目反应理论这两种方法属于心理测量领域的重要理论，但二者偏重点有所差异。经典测试理论的信效度的精确度不高，测验的参数会受到样本的干扰，其侧重于宏观；项目反应理论侧重微观、标准客观、信效度高，受到参与者的影响小，一定程度上弥补了经典测试理论的不足，在简化条目数量时优势明显[19]。朱燕波等[20]在2017年通过经典测试理论与项目反应理论方法对原60条目中医体质量表进行分析，形成41条目简短版；2018年从兼顾干预反应、辨识效率和准确性角度进一步简化，形成30条目简短版[21]。刘四军等[22]通过收集到的4000份调查问卷进行相关性检验分析，剔除原标准中无相关性的条目，最终简化成29个条目。不同条目版本中医体质量表总体上信度、效度较好，但量表的信度会受到条目数量影响[23]。随着对问卷或量表开发简短版本的深入研究，遗传算法也逐渐被应用于选择属性与优化组合问题。

本文提出了应用遗传算法结合KNN算法的属性选择方法在体质量表60个条目中筛选代表性问题，以帮助简缩减条目数量且减少时间，助力中医体质量表被各人群或机构广泛使用，为体质辨识与“治未病”、健康管理提供依据，帮助制定有针对性的预防、治疗计划。

1 相关工作

1.1 遗传算法

遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型[24]。其基本思想就是模仿自然进化过程，利用自然选择、交叉和变异的过程对群体中具有某种结构形式的个体进行遗传操作，进而生成新的群体，逐渐淘汰掉适应度函数值低的解，最后寻得最优解。迪盼祺等[25]研究中医智能问诊系统中，基于协同过滤算法和遗传算法构建症状获取模块，有目的性提问患者可能存在的症状，且利用随机森林算法建立分类器完成中医辨证，只需经过简单的几次问答便可获得需要的核心症状，大大地提高了问诊效率。Eisenbarth等[26]在1590个总样本上使用遗传算法生成了40个条目的简短版本，且进行测试后，发现收敛性和辨别效度较好。Sahdra等[27]在具有全国代表性的美国大样本中使用遗传算法简化多维经验回避问卷，把62项的问卷缩短一半，且维持6个维度的准确性，没有太多的信息损失。Rachmani等人[28]基于印尼1029份数据对具有47个问题的欧洲健康素养调查问卷使用遗传算法和Knn方法选择测量中最重要的特征，最终简化为拥有10个条目的简短问卷。

遗传算法一般通过编码、初始化、迭代进化、解码4个操作实现。首先是编码，每个染色体代表着问题的候选解，实现解空间向编码空间的映射过程，一般采用二进制编码方式，用0和1组成的数字串模拟染色体，以便于实现基因交叉与变异等操作。接着是种群初始化操作，产生代表问题可能潜在解集的一个初始群体。那么产生初始化种群通常是经过随机方法产生或是根据先验知识设置一组必须满足的条件，再依据条件生成初始样本这两种方法生成。然后便是利用适应度函数对每代种群的个体计算适应度大小，并且按照适应值大小对所有个体依次排序，依据优胜劣汰的原则淘汰掉适应度值低的个体，且通过交叉算子或变异算子对留下的个体进行遗传操作。最后是经过多次的进化过程，得到的最优解是在种群中求解得到的适应度值最高的某个体。

1.2 KNN算法

KNN算法由Conver和Hart提出，是机器学习中一种较为经典的监督算法[29]。其基本思想是一个样本与数据集中的k个样本最相似，如果属于k个样本中的大多数样本都被归为某一类，那么该样本也是归到此类中。在该算法中，k值的大小在一定程度上影响着异常点的处理与分类的确定。k值设置较大时，比较容易排除异常点，但包含的样本较多，模型将更加简单，容易形成欠拟合；k值设置较小时，可以解决欠拟合问题，但模型将变得复杂，给排除异常点带来困难，并会造成过拟合。因此使用KNN算法时，需要选取合适的k值。

本文是将遗传算法中原有的适应度函数改为KNN分类算法，把分类的准确率作为每个个体的适应度值，然后选择适应值大的个体，进而继续遗传算法中的交叉、变异操作，得到合适的属性子集，最后再使用KNN算法与交叉验证分类，查看效果。

1.3 评价指标

本文提出的结合遗传算法和KNN算法对中医体质量表进行问题选择的方法，是通过遗传算法进行问题子集的搜索，搜索得到最优的问题子集。那么需要考虑到子集规模即问题数量和该问题子集在分类算

法上的准确度[30]，还有简化后的量表填写时间。分类准确率计算公式为：

$$acc = \frac{n}{N} \quad (1)$$

其中，n是正确分类的样本个数，N是样本总数。

2 实验与结果分析

2.1 数据准备

本文实验数据是由江苏省中医院体检科提供，共1644条。原始数据集有用户编号、姓名、性别、问诊问题、所属体质、体质得分6种属性。本文量表简化工作是在王琦院士编制的9个亚量表、60个问题的《中医体质量表》上进行。原始数据集中，问诊问题符合王琦院士编制的《中医体质量表》中问题描述且数量满足60个，则纳入该条数据；相反，若不符合实验要求，将其排除。最终共纳入1298条数据，其中9种体质的构成情况见表1。对被纳入的数据集进行处理，每条数据包含60个问题的回答选项和1个体质类型。将每条数据的60个自测问题视为60个属性，属性值则依据回答选项“没有（根本不）”“很少（有一点）”“有时（有些）”“经常（相当）”“总是（非常）”分别赋予分值1、2、3、4、5；每条数据的体质类型值若有缺失，则按照《中医体质分类与判定》里的标准计算相应的条目分数与转化分，从而确定为9种体质类型中的一种，体质类型视为1个标签。实验数据结构如表2所示。

表1 实验数据集构成

体质类型	数量（条）	百分比（%）
平和质	50	3.85
阳虚质	193	14.87
阴虚质	219	16.87
气虚质	176	22.42
痰湿质	291	13.56
湿热质	112	8.63
血瘀质	108	8.32
特禀质	65	5.01
气郁质	84	6.47
总数	1298	100

由于实验数据构成不平衡，为了能够获得更好的效果，利用过采样方法获取更多的比例少的数据。通过复制数据来平衡数据集，将平和质标签的数据重复5次，阳虚质、气虚质、湿热质和血瘀质各重复2次，特禀质重复4次，气郁质重复3次，直到达到平衡组合，最终有2450条数据组成。

2.2 实验描述

2.2.1 实验过程

对王琦院士编制的中医体质量表60个问题，依次用Q_1，Q_2……Q_{59}，Q_{60}代表，在对中医体质量表选择问题过程中，首先是在遗传算法进行特征选择过程时，把KNN分类准确率作为遗传算法中的个体适应度值，选择出合适的问题子集，详细的选择过程如下：

①输入体质量表数据作为数据集X，60个问题的集合$Q=\{Q_1,Q_2,\cdots\cdots,Q_{59},Q_{60}\}$，标签$Y=\{0,1,\cdots\cdots,7,8\}$，选择的问题个数$n$，最大进化代数为$M$；

②参考二进制编码方案，把中医体质量表的不同问题组合全部编码为个体的基因序列，生成一组初始个体，形成初始种群；

③采用KNN算法对数据集X分类，计算各个体种群之间的距离，并以KNN算法的分类准确率$acc(Q_n)$来作为个体的适应度值*fitness*；

④判断是否已处理某代所有个体，如果是，则进行下一代；否则将个体按照适应度值大小排序并执行选择操作，2个个体被选择来执行交叉或变异操作，接着将生成的个体加入到新群体中；

⑤判断是否已经达到最大进化次数M，如果达到则输出当前选择出的问题集合；否则将返回③操作。

实现选择问题子集后，对筛选出的包含28-32个问题的5个问题子集，从数据集X中选出降维后的新数据集X_Q，使用KNN算法对数据集X_Q进行分类。将数据集X_Q按8∶2划分为训练集和测试集，训练时使用10折交叉验证法，将训练集随机划分成10份，其中9份用于模型的训练，1份用于模型验证，以此来观察问题子集的分类性能。且通过调整KNN算法中的参

表2 实验数据结构

编号	您手脚发凉吗？	您胃脘部、背部或腰膝部怕冷吗？	……	您容易忘事（健忘）吗？	体质类型
1	2	3	……	2	阴虚质
2	3	3	……	2	血瘀质
……	……	……	……	……	……
1298	2	1	……	2	痰湿质

数k，寻找到模型最佳k值，并运用准确率和问题子集规模两个评价指标以比较效果。

2.2.2 实验参数设置

采用Python算法来编制遗传算法，设置种群大小为30，进化代数为60。在多次运行实验后，发现被选择的属性子集在测试集上拥有比较好的分类准确率时，遗传算法中设置的交叉概率为P_c=0.6，变异概率为P_m=0.05，且KNN算法中k取值为1时，可以获得最佳分类结果。

2.2.3 结果分析

本次实验主要是观察体质量表的问题数量简化到28-32个问题时，被选择的不同问题子集经过KNN算法分类后，比较不同问题子集的性能效果以获得到具有较好的体质量表简短版本。从图1可以看出，随着问题个数的增加，问题子集规模的增大，在模型中分类的准确率越高。在实验模型中，当问题个数从28增加到29时，准确率从84.34%增至85.41%，提高了1.07%；再增加一个问题时，平均准确率提高了0.67%；当问题个数增加到31时，平均准确率相比选择30个问题上升了0.05%；再加入一个问题后，平均准确率仅提高了0.03%的精度。那么即使在31个问题的模型上逐渐添加一个问题，精度也不会有太大的变化。因此，本研究提出了一个拥有31个问题的模型作为中医体质量表简短版本，具体问题参考表3。

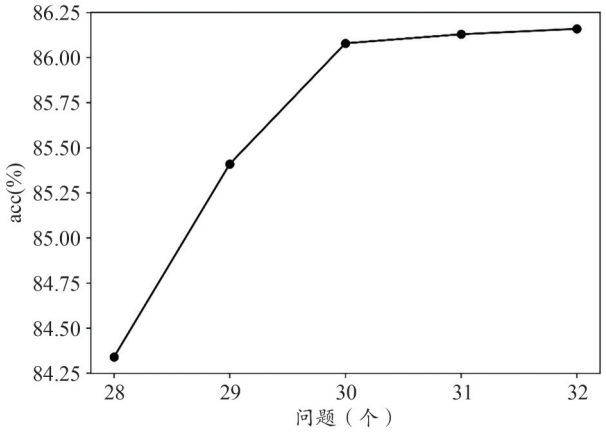

图1 不同问题子集规模的KNN分类准确率

由表4可了解到，每种体质类型在具有31个问题的中医体质量表简短版本上的问题分布情况，阳虚质类型最少，只有2个问题；痰湿质有3个问题；剩下的类型都各有4个问题，分布均匀。

为了对比原60条目与简化后的31条目中医体质量表填写时长，随机抽取了60名18-40岁的对象，分2组各30人，在安静无干扰的环境下填写体质量表并计时。表5显示原60条目的体质量表平均花费时间是355 s，简化后的31条目量表版本平均花费183 s，填写占用时间减少51.5%。

以上说明，本文提出的方法对中医体质量表进行问题选择研究后，选择出的问题考虑并覆盖了每种体质类型，且在评价指标中表现良好，那么该模型对于解决问题数量多、填写时间长问题有现实意义。

表3 中医体质量表——31条目简短版本

序号	问题
Q_2	您胃脘部、背部或腰膝部怕冷吗？
Q_3	您感到怕冷、衣服比别人穿得多吗？
Q_8	您感到手脚心发热吗？
Q_{10}	您皮肤或口唇干吗？
Q_{12}	您容易便秘或大便干燥吗？
Q_{13}	您面部两潮红或偏红吗？
Q_{16}	您容易疲乏吗？
Q_{17}	您容易气短（呼吸短促，接不上气）吗？
Q_{21}	您说话声音无力吗？
Q_{22}	您活动量就容易出虚汗吗？
Q_{25}	您腹部肥满松软吗？
Q_{26}	您有额部油脂分泌多的现象吗？
Q_{28}	您嘴里有黏黏的感觉吗？
Q_{33}	您感到口苦或嘴里有异味吗？
Q_{34}	您大便黏滞不爽、有解不尽的感觉吗？
Q_{35}	您小便时尿道有发热感、尿色浓（深）吗？
Q_{36}	您带下色黄（白带颜色发黄）吗？（限女性回答）/您的阴囊部位潮湿吗？（限男性回答）
Q_{37}	您的皮肤在不知不觉中会出现青紫瘀斑（皮下出血）吗？
Q_{38}	您两颧部有细微红丝吗？
Q_{40}	您面色晦黯或容易出现褐斑吗？
Q_{41}	您容易有黑眼圈吗？
Q_{44}	您没有感冒时也会打喷嚏吗？
Q_{47}	您容易过敏（对药物、食物、气味、花粉或在季节交替、气候变化时）吗？
Q_{48}	您的皮肤容易起荨麻疹（风团、风疹块、风疙瘩）吗？
Q_{50}	您的皮肤一抓就红，并出现抓痕吗？
Q_{53}	您多愁善感、感情脆弱吗？
Q_{54}	您容易感到害怕或受到惊吓吗？
Q_{55}	您胁肋部或乳房腹痛吗？
Q_{57}	您咽喉部有异物感，且吐之不出、咽之不下吗？
Q_{58}	您精力充沛吗？
Q_{59}	您能适应外界自然和社会环境的变化吗？

表4 中医体质量表——31条目简短版本的问题分布

阳虚质	阴虚质	气虚质	痰湿质	湿热质	血瘀质	特禀质	气郁质	平和质
Q_2	Q_8	Q_{16}	Q_{25}	Q_{33}	Q_{37}	Q_{44}	Q_{53}	Q_{58}
Q_3	Q_{10}	Q_{17}	Q_{26}	Q_{34}	Q_{38}	Q_{47}	Q_{54}	Q_{59}
	Q_{12}	Q_{21}	Q_{28}	Q_{35}	Q_{40}	Q_{48}	Q_{55}	Q_{16}
	Q_{13}	Q_{22}		Q_{36}	Q_{41}	Q_{50}	Q_{57}	Q_{22}

表5 中医体质量表60条目与31条目填写时间比较

量表条目	填写时间(s)
60条目	355±67
31条目	183±28

3 总结

中医认为体质是在遗传的基础上，受到缓慢与潜在的环境条件影响，在生长发育过程中逐步形成的个体特性。把握体质的特征和变化可以判断疾病发生和发展的进程，那么辨识体质在疾病的治疗与预防规划中显得格外重要。目前普遍应用的中医体质量表由于条目数较多，填写耗时长会给填写者的耐心带来挑战，从而增加判断的不稳定性。因此需要采取方法来减少问题数量和缩短填写时间。本文采用将KNN的分类准确率作为遗传算法中的适应度值对中医体质量表进行问题选择，以KNN分类的准确率和问题数量作为问题子集的评价标准，并采用10折交叉验证训练模型，最终模型的KNN分类准确率与问题性数量解释了结合遗传算法和KNN算法的方法运用在中医体质量表属性选择上有一定的可行性及有效性，且当KNN中近邻值为1时的模型效果最佳。同时表明了运用该方法筛选出的属性子集是帮助简化了中医体质量表，减少了题目数量，缩短了填写时间且一定程度上能够帮助辨识体质，为中医体质辨识提供更快速便捷的方式，帮助制定有针对性的预防、治疗计划。现今中医体质量表的相关研究仍主要集中在“体质可分”“体病相关”领域[31]，对量表的简化研究较少，未来可采取更加先进与严格的方法对量表进行简化与评价工作，极力推动量表应用与体质学发展。

参考文献

1 王琦. 中医体质学: *2008*. 北京: 人民卫生出版社, 2009:406-407.
2 王济. 基于中医体质学的三个关键问题探讨全生命周期健康管理. 北京中医药大学学报, 2023, 46(3):51.
3 王琦. 9种基本中医体质类型的分类及其诊断表述依据. 北京中医药大学学报, 2005, 28(4):1-8.
4 王琦, 朱燕波, 薛禾生, 等. 中医体质量表的初步编制. 中国临床康复, 2006, 10(3):12-14.
5 井慧如, 王济, 王琦. 英文版《中医体质量表》的初步编译. 安徽中医药大学学报, 2015, 34(5):21-25.
6 李炳皎. 韩文版中医体质量表开发与韩国人群中医体质流行病学调查研究. 北京: 北京中医药大学博士学位论文, 2015.
7 柳璇, 王琦. 《中医体质分类与判定》标准修改建议及分析. 北京中医药大学学报, 2013, 36(5):300-304.
8 马书鸽, 陈凤媚, 邓雪梅, 等. 1000例广州地区儿童中医体质调查研究. 南京中医药大学学报, 2015, 31(1):87-89.
9 杨寅. 《7-14岁儿童中医体质量表》的编制研究. 北京: 北京中医药大学博士学位论文, 2015.
10 刘卓勋, 黄斌, 郑燕霞, 等. 中国儿童中医体质分类辨识研究现状及展望. 世界科学技术-中医药现代化, 2016, 18(12):2182-2187.
11 李竹青, 张维, 孙鹏程, 等. 婴幼儿中医体质量表的研发思路. 中华中医药杂志, 2021, 36(10):5984-5987.
12 王雪峰, 王琦, 许华, 等. 《中医体质量表(0~1岁儿童试行版)》的评价与修订研究. 中华中医药学刊, 2022, 40(5):8-13.
13 刘卓勋, 杨京华, 许尤佳. 儿童中医体质量表研制的思考及建议. 中医杂志, 2022, 63(22):2122-2126.
14 朱燕波. 《中医体质量表》应用中的问题及其使用规范. 中华中医药杂志, 2022, 37(9):5066-5070.
15 Huan E Y, Wen G H, Zhang S J, *et al*. Deep convolutional neural networks for classifying body constitution based on face image. *Comput Math Methods Med*, 2017, 2017:9846707.
16 杨江, 王明航, 李建生, 等. 基于改良德尔菲法的支气管哮喘急性发作风险预警因素调查表及条目筛选研究. 中国全科医学, 2022, 25(35):4425-4432.
17 白一帆, 李敏, 艾浩楠, 等. 基于德尔菲法构建学龄前儿童中医体质类型及定义指标. 中医杂志, 2021, 62(12):1027-1031.
18 刘竞男, 张会永, 于莉, 等. 血虚证中医疗效评价量表条目筛选. 中华中医药杂志, 2021, 36(8):4583-4586.
19 朱燕波, 虞晓含, 王琦, 等. 简短版中医体质量表的初步设置与考评. 中国全科医学, 2017, 20(7):879-885.
20 朱燕波, 王琦, 虞晓含, 等. 中医体质量表-41条目简短版的结构效度和反应度评价. 中国全科医学, 2017, 20(26):3282-3286.
21 朱燕波, 王琦, 史会梅, 等. 中医体质量表-30条目简短版的制定与

评价. 中医杂志, 2018, 59(18):1554-1559.
22 刘四军, 周成成, 林秋姗, 等.《中医体质分类与判定表》的简化研究. 广州中医药大学学报, 2021, 38(8):1734-1739.
23 朱燕波, 史会梅, 虞晓含. 不同条目版本的中医体质量表在健康人群中应用的性能比较. 中国全科医学, 2019, 22(35):4381-4387.
24 周明, 孙树栋. 遗传算法原理及应用. 北京: 国防工业出版社, 1999: 1-64.
25 迪盼祺, 夏春明, 王忆勤, 等. 基于协同过滤算法的中医智能问诊系统研究. 世界科学技术-中医药现代化, 2021, 23(1):247-255.
26 Eisenbarth H, Lilienfeld S O, Yarkoni T. Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory-Revised (PPI-R). *Psychol Assess*, 2015, 27(1):194-202.
27 Sahdra B K, Ciarrochi J, Parker P, *et al*. Using genetic algorithms in a large nationally representative American sample to abbreviate the multidimensional experiential avoidance questionnaire. *Front Psychol*, 2016, 7:189.
28 Rachmani E, Hsu C Y, Nurjanah N, *et al*. Developing an Indonesia's health literacy short-form survey questionnaire (HLS-EU-SQ10-IDN) using the feature selection and genetic algorithm. *Comput Methods Programs Biomed*, 2019, 182:105047.
29 Zhang S C. Cost-sensitive KNN classification. *Neurocomputing*, 2020, 391:234-242.
30 崔正斌, 汤光明. 基于遗传算法和KNN的软件度量属性选择研究. 计算机工程与应用, 2010, 46(30):57-60.
31 白明华, 李倩茹, 李竹青, 等. 中医体质量表的国内外研究概述. 中华中医药杂志, 2021, 36(10):5993-5996.

Simplified Study of Constitution in Chinese Medicine Questionnaire Based on Genetic Algorithm and KNN Method

Guan Shutao[1], Li Hongyan[1], Lang Xufeng[1], Li Can[1], Zhou Zuojian[1], Hu Kongfa[1,2], Zhan Libin[3]

(1. School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China; 2. Jiangsu Collaborative Innovation Center of Traditional Chinese Medicine in Prevention and Treatment of Tumor, Nanjing 210023, China; 3. Center for Innovative Engineering Technology in Traditional Chinese Medicine, Liaoning University of Traditional Chinese Medicine, Shenyang 110847, China)

Abstract: Objective Aiming at the problems of many items and long time to fill in the Constitution in Chinese Medicine Questionnaire (CCMQ) when evaluating individual constitution, the research uses artificial intelligence technology to select attributes, and to help construct a short version of the CCMQ. Methods Analyzing the constitution data provided by the Physical Examination Department of Jiangsu Province Hospital of Traditional Chinese Medicine, there are specific target variables as the classification of constitution types. Feature selection of genetic algorithm, cross-validation and KNN classification algorithm are used as filters to select problems, and the effect is evaluated by problem subset size, KNN classification accuracy and filling time. Results The method selected a short version of the CCMQ with 31 problems, and the average classification accuracy in the model was 86.16%, and the time was improved by 47.7%. Conclusion The algorithm can effectively find a better problem subset, achieve dimensionality reduction and have certain accuracy, thus helping to simplify the CCMQ.

Keywords: Constitution of traditional Chinese medicine, CCMQ, Genetic algorithm, KNN algorithm

（责任编辑：刘玥辰）