

· 综述 ·

量表简化版研制的方法技术研究*

肖正琴¹ 刘钰曦² 全 鹏¹ 傅崇辉² 万崇华^{1△}

【提 要】 随着量表的简洁性和准确性成为量表研制的焦点,成熟量表的简化也日益受到重视。近些年,许多研究者尝试用不同的方法研制原始量表的简短量表。本文除了报道一些常用的基于测量学理论的简化方法(比如因子分析法、相关系数法、克隆巴赫系数法等)外,还报道一些近些年发展的简化方法,包括最优测试集合法、蚁群优化算法、机器学习算法等,对各种方法在量表简化研究中的应用及其优缺点进行介绍,为量表简化版本的研制提供一定的方法技术支持。

【关键词】 量表 简化版 最优测试集合法 蚁群优化算法 机器学习算法

【中图分类号】 R195.1 【文献标识码】 A DOI 10.11783/j.issn.1002-3674.2024.05.033

随着量表在医学研究中的广泛应用,研制简洁量表在现代医学研究中具有越来越重要的意义。量表条目数较长,不仅会降低受测者配合度,还会导致受测者产生疲劳效应,进而降低测量数据的准确性。条目数较多也不太适用于日常评估、快速动态监测等临床应用,一定程度上阻碍量表的推广与应用。因此,如何研制简单明了、易于操作、测量学性能较好的简短版量表成为研究者关注的热点问题。本文将介绍常用的量表简化统计学方法,以期研究者提供参考。

基于测量学理论的量表简化方法

量表简化版研制的常用方法有基于经典测量学理

论(classical test theory, CTT)、项目反应理论(item response theory, IRT)和概化理论(generalizability theory, GT)的传统方法。

1. 经典测量学理论

CTT 也称为“真分数理论”,是假设观测分数 X 是由真分数 T 及随机测量误差 E 所组成,即 $X = T + E$; 误差 E 的平均数等于 0; 误差 E 与真分数 T 间的相关为 0^[1]。基于 CTT 有多种简化方法,分别从不同角度和侧重点筛选条目,进而保证简表条目的代表性好、独立性强、区分度高^[2-3]。常用的筛选方法有离散趋势法、相关系数法、因子分析法、克隆巴赫系数法、临界比值法、回归分析等,概括总结于表 1。

表 1 基于经典测量学理论的条目筛选方法比较

筛选方法	简要说明
离散趋势	从敏感性角度筛选条目。若条目的变异系数 CV 值 < 15% 或标准差 < 1 ^[2] , 则表明该条目区分度不佳, 需考虑删除
相关系数	从代表性和独立性角度筛选条目。若条目-量表总分相关系数或条目-本维度相关系数 < 0.3/0.4 ^[4] , 则条目代表性较差, 需考虑删除; 若条目与多个维度相关系数 > 0.4, 则条目独立性较差, 需考虑删除
因子分析	从代表性角度筛选指标。若条目因子载荷 < 0.3/0.4, 交叉因子载荷 < 0.2/0.3 ^[5-6] , 则条目代表性较差, 需考虑删除
克隆巴赫系数	从内部一致性角度筛选条目。计算某侧面总克隆巴赫系数, 比较去除该条目后的系数变化。若某一条目去除后, 克隆巴赫系数有较大上升, 则说明该条目降低该侧面内部一致性, 需考虑删除
临界比值	又叫 t 检验法, 从敏感性、区分度、重要性角度筛选条目。将量表得分排名前 27% 的个体定义为低分组, 排名后 27% 定位为高分组, 通过比较两组条目的统计学差异, 若条目差异无统计学意义, 则条目质量较差, 需考虑删除
天花板/地板效应	从难易程度筛选条目。若答案的分布显示天花板或地板效应 > 20%, 则条目过于容易或简单, 其测量精度不足, 需考虑删除
逐步回归法	从重要性角度筛选条目。将量表总分作为因变量, 各条目作为自变量, 并进行多重逐步回归分析, 需考虑删除因变量影响较小的条目 ^[3]

2. 概化理论

GT 是 CTT 的扩展, 通过评估多个方差源及其相互作用来评估量表的可靠性^[7]。GT 研究主要包括概化研究(generalizability study)和决策研究(decision study)。概化研究, 又称“G 研究”, 是利用方差分析估计所有潜在误差来源的研究。决策研究, 又称“D 研

究”, 是基于 G 研究结果, 观察不同测量条件下可靠性系数的变化, 进而探索最佳决策的研究。在量表简化研究中, 概化理论的 D 研究可以用于^[7-8]: ①确定最佳条目数, 通过观察信度系数(概化系数、可靠性指数)和误差系数(相对误差、绝对误差)随着条目数变化的变动情况, 确定最佳条目数; ②确定最优条目集, 在条目数固定的情况下, 通过观察不同条目集的信度和误差系数的变化, 确定最优条目集。

3. 项目反应理论

IRT 是通过评估条目反应与潜在特质关系, 进而评估条目质量的统计方法^[9]。IRT 模型分为单维性项

* 基金项目: 国家自然科学基金资助项目(71974040; 71373058)

1. 广东医科大学生命质量与应用心理研究中心/生命质量与心理测评干预重点实验室(523808)

2. 广东医科大学卫生法制与政策研究所

△通信作者: 万崇华, E-mail: wanchh1964@qq.com

目反应(unidimensional item response theory, UIRT)模型和多维性项目反应(multidimensional item response theory, MIRT)模型。单维模型主要包括评级量表(rating scale model, RSM)模型、等级反应(graded response model, GRM)模型和拓广分布评分(generalized partial credit model, GPCM)模型;多维模型主要包括 MRSM(multidimensional rating scale model), MGRM(multidimensional grade response model), MG-PCM(multidimensional generalized partial credit model)等^[10-12]。IRT 方法主要根据数据特点,选择适当模型,进而通过模型评估条目质量,确定简表的具体条目。IRT 方法较常用的筛选标准有^[9-12]:①拟合统计量,反映项目与模型的拟合情况,需要考虑删除条目拟合度较差的条目;②区分度参数,反映条目对受测者的区分程度,需考虑删除区分度较差的条目;③阈值/难度参数,反映条目的难易程度,需考虑删除阈值参数不合理的条目;④信息函数,反映条目测量潜在特质的信息量,需考虑删除平均信息量较低或条目信息函数曲线平缓的条目;⑤项目功能差异(DIF),反映条目测量不同亚组(性别、年龄、职业、地区)潜在特质的公平性,需考虑删除存在 DIF 的条目。

新发展的量表简化方法

除了传统的基于经典或现代测量学理论的简化方法,近些年也提出了最优测试集合法(optimal test assembly, OTA)、蚁群优化算法(ant colony optimization, ACO)、机器学习算法(machine learning, ML)等新方法。

1. 最优测试集合法

OTA 是一种分支定界混合整数规划方法,它依赖 IRT 模型的项目参数估计,选择最能满足客观、可重复和预定标准的最优项目子集^[13]。近年,该方法被用于研制量表简化版本^[14-16]。OTA 研制简化版本主要包括以下两个阶段^[14-16]:第一步,在固定条目的情况下,系统搜索所有可能的简化版本,选出信息量最大的候选版本;第二阶段,依据预定标准,确定满足预定标准的最少项目集合。预定标准包括信度和效度标准,如简表的克朗巴赫系数至少为原量表量的 95%,且得分相关性至少为 0.95。因此,在给定相同标准的前提下,OTA 获得的简化形式是可重复性且测量学性能等效于原始量表^[14-16]。

2. 蚁群优化算法

ACO 是一种现代启发式(modern heuristic)算法,就是利用蚂蚁觅食最优路径原理寻求待解决问题的最优解。蚂蚁在觅食时会在觅食路径上释放一定的信息素,随着搜索次数的增加,路径信息素累积的越多。蚂蚁群在正反馈机制的作用下逐渐集中于信息素最多的

路径,即为最优路径。ACO 算法被广泛应用于解决各种领域的最优化问题,现被用于研制量表简化研究。在量表简化研究中,ACO 可以同时根据原表的概念框架和简表的心理测量学特性,如模型拟合、测量不变性、与原表的相关性等,选择最优项目集,进而研制出量表简化版本^[17]。首先,ACO 随机选择几个简表,并根据一组自定标准(例如模型拟合和测量不变性等)对几个简表进行评估^[18]。然后,更加符合预定标准方案的条目将获得更高的(虚拟)信息素值。这些信息素反过来增加了条目在后续迭代过程中被选择的可能性^[18]。最后,在多次迭代过程中重复进行这种“选择-评估-信息素增加”的过程,直到预定标准不能进一步优化或信息素水平达到某个阈值,因所选解决方案变得过于相似而结束进程^[19]。研究表明,蚁群算法是一种有目的、有效的条目选择方法^[17-19]。

3. 机器学习算法

机器学习算法是人工智能领域的核心技术之一,是大数据预测和分类领域的强大工具,已得到十分广泛的应用^[20-21]。ML 在量表简化条目的研究中已被证实具有较大的潜力,是一种有效的量表简化方法,尤其在医学量表研究中^[20-21]。Lin^[22]等认为不仅可以通过 ML 方法选择最重要的条目来提高临床评估效率,还可以通过该方法进一步提高量表的心理测量学特性。随着机器学习的不断发展,机器学习算法逐步应用于量表简化的研究,并提出了一些基于机器学习算法的具体方法,例如梯度提升决策树、随机森林、支持向量、 k 近邻、线性回归模型等^[23-25],其基本特征总结于表 2。

量表简化方法的比较

不同理论下的统计学方法简化的角度和侧重点各异。CTT、IRT 是量表简化最常用的条目选择方法;GT 常为条目选择的补充策略方法,常与 CTT、IRT 等条目选择策略一起使用。CTT、IRT 是从条目质量的局部角度,通过评估所有条目的质量,进而选出质量最优的条目组成简表。OTA、ACO、ML 从简表的整体角度出发,基于一些标准,选择某些性能等效于原表的简表。这些基于不同理论的统计学方法各有其优缺点,详见表 3。

鉴于不同理论下统计方法的局限性,较多研究者建议综合使用多种理论的统计方法研制量表简化版本。较为常用的是采用 CTT 和 IRT 结合的模式^[6,28]。这种结合多理论方法的模式不仅可以克服单独使用某种测量理论的不足,还可以更全面把握条目特征,提高简化结果的客观性和准确性^[3,28-29]。然而,这些方法选择的条目是主观的,而不是通过系统地建立截断值或使用可重复的标准,可能会导致同一措施存在多个

简化版本^[30-31]。

表 2 基于机器学习算法简化条目各种方法基本特征

算法	优点	缺点
梯度提升回归树	能处理混合类型的数据,不需要复杂的特征工程和特征变换;具有很强的预测能力;具有强大的损失函数,在输出空间上对异常值具有很强的鲁棒性	在可扩展性能方面,由于提升的时序性原因,模型不可以并行处理
随机森林	在处理异常值时很少出现过拟合的现象,并且对异常值具有优异的稳健性,具有良好的性能	国内研究较少,仍需进一步研究,进而提出更多切实有效的优化方案
支持向量	对于小样本数据优秀的训练能力和突出的泛化能力,对新增数据的预测具有很好的泛化能力	对大规模训练样本难以实施;解决多分类问题困难;对参数和核函数选择敏感
k 近邻	流程简单,算法容易理解,预测精度高,理论较为成熟等优点,既可以用来做分类研究也可以用来做回归,同时兼备了时间复杂度低的优点	计算、空间复杂性较高;需要大量内存,难以给出数据的内在含义
线性回归模型	思路简单,实现容易,建模迅速,对小数据及简单的关系很有效;是许多强大的非线性模型的基础;且容易理解,结果具有很好的可解释性	对于非线性数据或者数据特征间具有相关性多项式回归难以建模;难以很好地表达高度复杂的数据

表 3 量表简化研制方法及其比较

理论/方法	优点	缺点
经典测量理论	简单易理解、体系完整、适用性广、数学模型简单、计算量小、实施条件要求不严格、多数情况下计算准确 ^[1]	测验结果拓广的有效性、样本依赖性、误差含糊和信度估计的不精确性、能力与难度量尺的不一致性 ^[1]
概化理论	可对各种误差来源进行分解与控制,控制误差方差,信度估计比 CTT 更为细致和准确 ^[7]	未改良 CTT 的微观结构,因此在 CTT 的主要局限性依然存在 ^[7]
项目反应理论	测量理论从整体宏观转向了细致微观领域;被试对项目的反应与其潜在特质间关系,更符合事实;且不依赖于测验题目;使测验达到预先规定的满意精度 ^[9]	单维性假定有时难以满足,数学模型更为复杂,测验条件要求严格,计算量更大,甚至可能会增大测量误差,对被试给出不够准确的评价等 ^[9]
最优测试集合法	具有可复制性和可重复性,并且能产生最短的简短版本,简化过程变得更快、更标准和更透明 ^[15,26]	不适用于多维性量表;其次对预定标准非常敏感,且基于不同完整的项目库会产生不同的简化结果
蚁群优化算法	简化结果更具有客观性和可重复性,且优于传统的简化方法 ^[18,27]	算法的计算复杂性更高,且不太适用于样本量较少(< 500)的简化研究 ^[18]
机器算法	直观、好理解,既节约答题时间,又提高了测量的精确程度 ^[25]	对研究者要求更高,国内应用较少 ^[25]

相比之下,OTA, ACO, ML 的可重复性和可复制性可能更胜一筹。但这些方法计算过程复杂,对研究者提出要求更高,且在我国简化研究中少有报道。因此,如何更好应用与结合这些简化方法,如何在保证简化结果科学性的情况下提高简化程序的可重复性等问题上,仍需进一步的探索。

总的来说,目前量表简化方法众多,各种方法各有优缺点,且研究者各自为营,尚无统一标准,简化程序的可重复性较差。因此,未来仍需探讨多方法的结合应用,为可重复性的量表简化研制流程提供科学有效的依据。

结 语

本文简单介绍了常用的量表简化研制方法及其思想,总结了不同方法的优缺点和进展情况。虽然这些方法在简化研究中应用越来越广,但是部分方法在国内却少有报道。因此,在使用基于测量学理论的统计方法时,建议研究者可以再考虑结合 OTA、ACO、ML 等方法,进而提高简化程序的可复制性以及简化量表的有效性。

参 考 文 献

[1] 戴海崎, 张锋, 陈雪枫, 等. 心理与教育测量 [M]. 广东: 暨南大学出版社, 1999, 58-68.

[2] 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 [J]. 中医杂志, 2018, 59(18): 1554-1559.

[3] 路桃影, 吴大嵘. 简短版量表研究中条目筛选方法概述 [C] // 第七届中医/中西医结合循证医学方法研讨会会议材料. 广州: 中国中西医结合学会, 2013, 95-103.

[4] 陈姝, 万崇华, 杨铮, 等. 基于经典测量理论与项目反应理论的乳腺癌生命质量测定量表 QLICP-BR (V2.0) 条目分析 [J]. 中国卫生统计, 2021, 38(6): 864-869.

[5] 苗春霞, 刘慎军, 卓朗, 等. 大学生生命质量评价简量表研制与信效度分析 [J]. 中国学校卫生, 2019, 40(6): 865-869.

[6] Li M, He S, Wang J. Development and validation of a new short form of the self-management ability questionnaire for patients with chronic periodontitis [J]. Community Dent Oral Epidemiol, 2022, 50(3): 171-179.

[7] 杨志明, 张雷. 测评的概化理论及其应用 [M]. 北京: 教育科学出版社, 2003: 40-50.

[8] 张昊, 汤青霞, 江逊, 等. 基于概化理论学龄前儿童饮食行为量表的评价和修订 [J]. 中国儿童保健杂志, 2019, 27(11): 1179-1182+1187.

[9] 王伟梁, 周郁秋. 项目反应理论在健康相关量表中的应用现状及展

- 望[J].中国卫生统计,2018,35(4):633-636.
- [10] Kaspar R, Gabrian M, Brothers A, et al. Measuring Awareness of Age-Related Change: Development of a 10-Item Short Form for Use in Large-Scale Surveys[J]. *Gerontologist*,2019, 59(3):e130-e140.
- [11] Liang MZ, Tang Y, Chen P, et al. New resilience instrument for family caregivers in cancer: a multidimensional item response theory analysis[J]. *Health Qual Life Outcomes*,2021,19(1):258.
- [12] Feng JY, Chen CC, Chang YT, et al. A psychometric analysis of a short form of the Chinese version of the ISPCAN child abuse screening tools~ Children's home version (SC-ICAST-CH) using multidimensional item response theory [J]. *Child Abuse Negl*,2020, 109:104693.
- [13] Linden WJ. *Linear Models for Optimal Test Design*. Springer; New York, NY, USA; 2006,81-82.
- [14] Harel D, Mills SD, Kwakkenbos L, et al. Shortening patient-reported outcome measures through optimal test assembly: application to the Social Appearance Anxiety Scale in the Scleroderma Patient-centered Intervention Network Cohort [J]. *BMJ Open*, 2019, 9(2):e024010.
- [15] Harel D, Baron M. Methods for shortening patient-reported outcome measures[J]. *Stat Methods Med Res*,2019,28(10-11):2992-3011.
- [16] 肖正琴,黄聿明,万崇华,等.基于最优测试集合法的良性前列腺增生量表特异模块简化版的研制与考评[J].中国卫生统计,2022,39(6):819-823.
- [17] Olaru G, Schroeders U, Hartung J, et al. Ant colony optimization and local weighted structural equation modeling: A tutorial on novel item and person sampling procedures for personality research[J]. *European Journal of Personality*,2019,33(3):400-419.
- [18] Olaru G, Danner D. Developing Cross-Cultural Short Scales Using Ant Colony Optimization[J]. *Assessment*,2021,28(1):199-210.
- [19] Schultze M, Eid M. Identifying measurement invariant item sets in cross-cultural settings using an automated item selection procedure [J]. *Methodology*,2018,14(4):177-188.
- [20] Abbas H, Garberson F, Glover E, et al. Machine learning approach for early detection of autism by combining questionnaire and home video screening [J]. *J Am Med Inform Assoc*,2018,25(8):1000-1007.
- [21] Thabtah F, Kamalov F, Rajab K. A new computational intelligence approach to detect autistic features for autism screening [J]. *Int J Med Inform*,2018,117:112-124.
- [22] Lin GH, Huang CY, Lee SC, et al. A 10-item Fugl-Meyer Motor Scale Based on Machine Learning[J]. *Phys Ther*,2021,101(4):pz-ab036.
- [23] Gonzalez O. Psychometric and Machine Learning Approaches to Reduce the Length of Scales[J]. *Multivariate Behav Res*,2020,4:1-17.
- [24] Lötsch J, Sipilä R, Dimova V, et al. Machine-learned selection of psychological questionnaire items relevant to the development of persistent pain after breast cancer surgery [J]. *Br J Anaesth*,2018, 121(5):1123-1132.
- [25] 刘金铭.基于机器学习的症状自评量表简化与应用[D].青岛:青岛大学,2021:1-54.
- [26] Gerdessen JC, Souverein OW, van 't VP, et al. Optimising the selection of food items for FFQs using Mixed Integer Linear Programming[J]. *Public Health Nutr*,2015,18(1):68-74.
- [27] Leite WL, Huang IC, Marcoulides GA. Item selection for the development of short forms of scales using an ant colony optimization algorithm [J]. *Multivariate Behavioral Research*,2008, 43(3):411-431.
- [28] 吾尔肯·朱马江,林晓,王莹,等.慢性 HBV 感染者歧视测量量表简化版的研制[J].中华疾病控制杂志,2021,25(6):716-721.
- [29] 薛红红,杨云滨,万崇华,等.基于经典测量理论与项目反应理论的老年 COPD 患者多维健康测定量表条目分析[J].中国卫生统计,2019,36(1):32-35.
- [30] Goetz C, Coste J, Lemetayer F, et al. Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales [J]. *J Clin Epidemiol*, 2013,66(7):710-718.
- [31] Kruyen PM, Emons WHM, Sijtsma K. On the Shortcomings of Shortened Tests: A Literature Review [J]. *International Journal of Testing*,2013,13(3):223-248.

(责任编辑:邓妍)

(上接第 780 页)

- [9] Linden WJVD. *Constrained Adaptive Testing with Shadow Tests* [M].//Linden WJVD, Glas CAW. *Elements of Adaptive Testing*. NY, USA; Springer New York, 2010:31-55.
- [10] 简小珠,戴海琦,张敏强,等. CAT 选题策略分类概述[J].心理学探新,2014,34(5):446-451.
- [11] Brown JM, Weiss DJ. *An Adaptive Testing Strategy for Achievement Test Batteries*[R]. USA:ERIC Clearinghouse, 1977.
- [12] Barrada JR, Olea J, Ponsoda V, et al. A Method for the Comparison of Item Selection Rules in Computerized Adaptive Testing[J]. *Applied Psychological Measurement*,2010,34(6):438-452.
- [13] 王超,陆宏.计算机自适应测验的研究策略与应用实践[J].现代教育技术,2017,27(12):44-49.
- [14] Nydick SW, Weiss DJ. A Hybrid Simulation Procedure for the Development of CATs [C].//Weiss DJ. *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Minneapolis, MN, USA, 2009.

(责任编辑:邓妍)