

• 国外护理 •

基于 Python 语言美国国家科学基金会护理-信息-工程学相关立项情况分析

苑秋辰,杨浩杰,麻盛淼,沈晨,梁涛

(中国医学科学院 北京协和医学院 护理学院,北京 100144)

【摘要】 **目的** 分析美国国家科学基金会中护理-信息-工程学相关科研项目资助情况,为促进我国护理-信息-工程学科发展提供参考。**方法** 使用 Python 语言的网络爬虫和隐含狄利克雷分布(latent Dirichlet allocation,LDA)模型对美国国家科学基金会中护理-信息-工程学相关的项目数据进行采集、预处理和分析。**结果** 本研究共纳入 265 个相关科研项目,时间跨度为 1988 年至 2023 年,立项数量和资助金额整体呈上升趋势。纳入项目共包含 2153 个词,基于 LDA 模型分为 6 个主题,分别为智能技术在患者安全管理中的研究、面向患者的护理大数据系统的研发、护士健康及辅助决策工具的研发、虚拟现实技术在护理教育中的应用、传感器和无线辅助设备的研发和机器人在医院工作流程中的应用。**结论** 大数据和信息化时代让信息学和工程学融入到护理学科的建设当中,研究者可结合国情继续深入探索我国护理-信息-工程学的发展前景。

【关键词】 护理学;信息学;工程学;美国国家科学基金会;Python;机器学习;网络爬虫

doi:10.3969/j.issn.2097-1826.2023.06.027

【中图分类号】 R47 **【文献标识码】** A **【文章编号】** 2097-1826(2023)06-00112-05

Analysis of Nursing-Information-Engineering-Related Programs Funded by the National Science Foundation of America Based on Python

YUAN Qiuchen, YANG Haojie, MA Shengmiao, SHEN Chen, LIANG Tao (School of Nursing, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100144, China)

Corresponding author: LIANG Tao, Tel: 010-88771002

【Abstract】 Objective To analyze the nursing-information-engineering-related scientific research programs funded by the National Science Foundation of the United States, and to provide a reference for the research in China. **Methods** The Python web crawler and machine learning technology were used to collect, pre-process, and analyze the data of the nursing-information-engineering-related programs funded by the National Science Foundation of the United States. **Results** A total of 265 related scientific research programs were included in the study, covering the period from 1988 to 2023. The overall number of projects and the funding amount showed an increasing trend. The total number of words included in the included project was 2 153, which was divided into 6 topics based on the LDA model, including research on intelligent technology in patient safety management, research and development of patient-oriented nursing big data system, research and development of nurse health and auxiliary decision-making tools, application of virtual reality technology in nursing education, research and development of sensors and wireless auxiliary devices, and application of robots in hospital workflow. **Conclusions** With the background of big data and informatization, informatics and engineering have been integrated into the construction of nursing sciences. The researchers can continue to explore the development prospect of nursing-information-engineering in China according to China's national conditions.

【Key words】 nursing; informatics; engineering; National Science Foundation of America; python; machine learning; web crawler

[Mil Nurs, 2023, 40(06): 112-116]

2021 年我国工业和信息化部提出要加强基础

学科和前沿学科建设,促进新兴交叉学科发展以提高社会建设水平^[1],因此“医工交叉”已成为推进护理学科发展的新思路和必然趋势^[2]。美国早在 1988 年和 2008 年分别开展护理信息学和护理工程

【收稿日期】 2022-12-19 **【修回日期】** 2023-03-09

【作者简介】 苑秋辰,本科在读,电话:010-88771002

【通信作者】 梁涛,电话:010-88771002

学培养计划^[3],是第一个落实护理交叉学科培养的国家。美国国家科学基金会(National Science Foundation,NSF)的资助领域覆盖广,在推动美国科学发展中扮演关键角色^[4],因此其资助项目能反映美国科学研究的政策引领和发展方向,同时,促进跨学科研究也是NSF的重要优先事项^[5]。近年来Python已成为最受欢迎的编程语言之一,其代表性的网络爬虫、机器学习和数据挖掘等技术已初步应用到医疗卫生领域中^[6-7]。本研究旨在以创新的研究方法借助Python的网络爬虫和自然语言处理技术对NSF中护理-信息-工程学科相关资助项目进行分析,探索护理-信息-工程学未来的发展趋势,为后续研究提供参考。

1 数据来源与研究方法

1.1 数据来源 计算机检索NSF官网(<https://nsf.gov/awardsearch/advancedSearch.jsp>),选择高级检索,在补充信息中按关键词分别检索“nursing”“nurse”“care”“caring”,选择活跃资助和到期资助,不设时间范围。纳入标准:(1)项目主题与护理领域相关;(2)项目使用了信息学或工程学相关的技术。排除标准:(1)项目主题为动物或植物;(2)所得数据的关键字段(项目编号、项目标题、摘要等)有缺失值。

1.2 研究工具 Python是一种易于学习,功能强大,免费、开源、灵活的解释型编程语言^[8]。网络爬虫是指通过编程实现程序模拟人操作浏览器上网,并在网站中批量获取数据信息的过程^[9]。隐含狄利克雷分布(latent Dirichlet allocation,LDA)是一类贝叶斯层次主题模型,属于无监督机器学习算法中的一种^[10],作为高效的自然语言处理模型之一,在医疗卫生领域已被应用于社交媒体平台中相关主题的文本分析^[11-12]。

1.3 研究方法

1.3.1 数据采集与存储及筛选 使用Python的requests模块对上述网站的公开数据进行采集,包括每个项目的编号、标题、NSF资助机构、摘要、开始日期、项目负责人、资助金额等;使用csv模块将爬取的数据保存至本地。根据纳入和排除标准对数据进行筛选。

1.3.2 数据清洗及格式规范 使用Python的pandas模块读取csv文件,将每个项目的标题和摘要拼接,用re模块中的正则表达式去掉数据中的网页标签,将字符串类型的数据全部转换为小写,将复数词和动词转化为原形。使用python的jieba模块对拼接后的数据进行分词,并去除停用词。

1.3.3 词袋模型建立及主题数量确定 使用

Python中gensim模块的corpora和doc2bow方法将每个项目数据的分词列表转化为词袋模型^[13]。使用gensim、math和matplotlib模块根据传入的词袋模型绘制困惑度(perplexity,P)-主题数量(number of topic,NT)关系图以确定主题数量,当主题数量为某一值时P出现极大值拐点,且此后P-NT曲线逐步趋于平缓,则认为此时的这个值为该词袋模型的最佳主题数量^[13]。

1.3.4 LDA模型建立及可视化分析 使用Python中gensim模块的LDAmodel方法对上述词袋模型进行主题聚类分析,并使用LDAvis模块绘制主题间距离图进行可视化效果评估,图中灰色气泡表示该主题的词袋,气泡越大则词袋中单词数量越多,当各气泡间均无重叠区域时,认为此LDA模型达最有效,主题的相关性(relevance,r)最好。

1.3.5 各主题代表词数量确定 借助文献计量学方法中的齐夫定律^[14]选定词袋中高频词的数量。

2 结果

2.1 NSF资助项目概况 经检索和筛选,共资助护理-信息-工程学项目265项,时间跨度为1988年至2023年,累计资助金额约合6.87亿元人民币。每年立项数量及资助金额整体呈上升趋势,详见图1。统计NSF资助金额最高的前10个项目,涵盖护理管理、老年护理、社区护理、孕产妇护理等,技术层面包括增强现实、自动化分析、智能互联、活动感知分析、可穿戴设备、新型引流装置等,详见表1。NSF资助学部方面,分别有科技创新与合作学部(91项,34.3%),计算机信息科学与工程学部(81项,30.5%),社会行为与经济学部(54项,20.3%),工程学部(27项,10.1%),科学、技术、工程和医药(science, technology, engineering and medicine,STEM)教育学部(3项,1.1%),教育与人力资源学部(3项,1.1%),主任办公室(3项,1.1%),生物科学学部(2项,0.7%),数学与物理科学学部(1项,0.3%)。

2.2 主题数量确定 使用Python建立词袋模型并绘制P-NT关系图。当NT为6时P出现明显的极大值拐点,且此后P-NT曲线变化逐渐减小且逐步趋于平缓;使用Python绘制主题间距离图如图2所示,发现当NT为6时,各灰色气泡间均无重叠区域。因此本研究将模型拟抽取的最终主题数量定为6个。

2.3 各主题命名及所含高频词 使用Python对本研究词袋中出现的词进行统计,共涌现2153个词,依据上述齐夫定律公式计算得出,前65个出现频次最多的词为高频词,由于本研究已确定分6个主题,因此平均每个主题纳入出现频次最多的前11个词,详见表2。

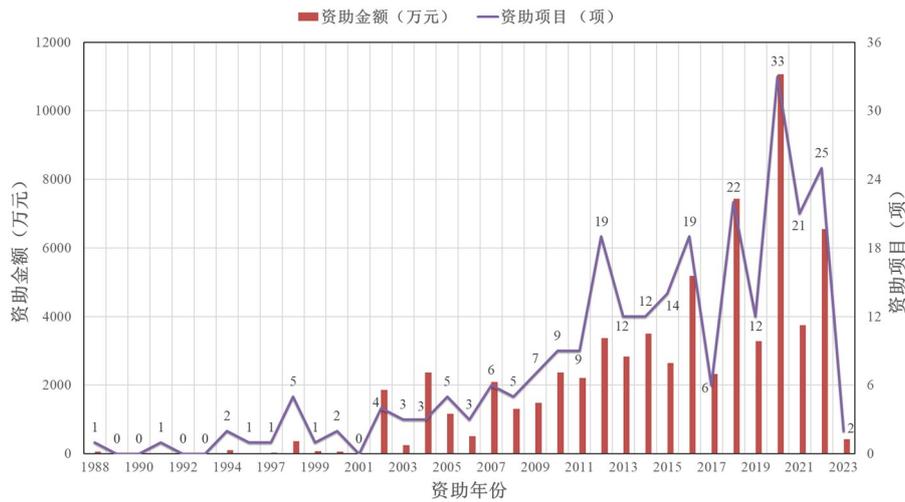


图1 NSF 每年资助项目数量和资助金额

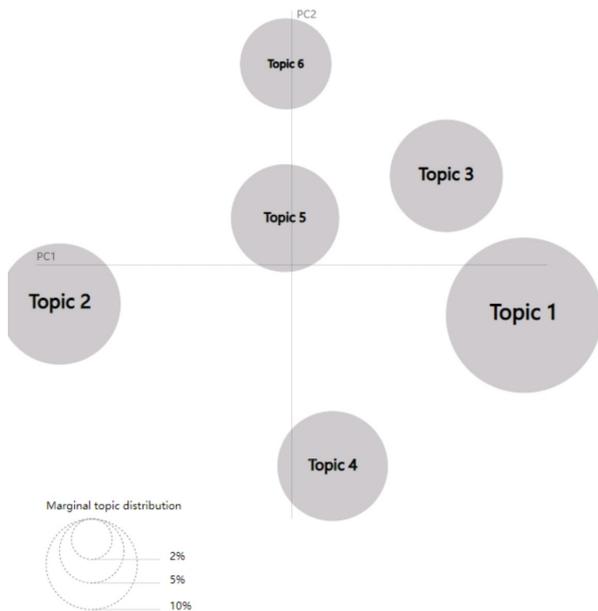
表1 NSF 资助护理-信息-工程项目金额前10个项目

项目编号(#)	项目标题	所属领域	资助年份	资助金额 (万元)
2026498	使医疗工作者能够完善以安全为关键的医院工作的学习系统	护理管理/信息系统	2020	2078.58
2225890	通过协作增强现实技术减少长期护理老年人的孤独感	老年护理/增强现实	2022	1558.94
0205219	用于老年护理的自动化视频和传感器分析	老年护理/自动化分析	2002	1476.48
1831918	为服务不足的社区提供智能、互联和协调的孕产妇护理	孕产妇护理/社区护理/智能互联	2018	1454.58
0427071	提高医疗流程的安全性和效率	护理管理/人机交互	2004	1207.13
2026584	通过协作式人机界面促进护理专业人员的未来工作	护理管理/人机交互	2020	1050.17
2026577	家庭护理工作的未来:为信任、隐私和授权设计技术	家庭(社区)护理/多媒体 异步通讯	2020	1039.21
1028195	通过异构活动感知分析转变基于社区的老年护理	老年护理,社区护理/活动感知分析	2010	1037.13
2221315	基于生物感应运动集成的自适应运动辅助可穿戴机器人	康复护理/可穿戴设备	2022	1021.62
1660238	心胸外科术后小口径胸管新型持续引流装置	术后护理/新型引流装置	2017	976.93

3 讨论

3.1 NSF 资助护理-信息-工程学项目力度整体上升

本研究显示,NSF 资助护理-信息-工程学项目数量和资助金额整体呈上升趋势。美国马里兰大学最早于1988年成立护理信息学硕士培养类型^[3],自此护理-信息交叉项目资助开始起步,而2008年麻省大学阿默斯特分校创立护理工程学博士培养类型后^[3],护理-工程学交叉项目融入NSF 资助计划中。2019年至今,随着范德堡大学、约翰霍普金斯大学等先后开展护理工程学博士或硕士培养类型后^[3],NSF 对于护理-信息-工程学项目资助力度也随之达到高峰,4年间资助项目93项(35%),资助金额3600余万美金(36%)。在NSF 资助学部方面,科技创新与合作学部和计算机信息科学与工程学部共涵盖172项,占比达65%,同时工程学部单独27项,反映出计算机信息技术为工程学技术在护理领域中的初步发展提供了基础。建议我国护理领域保持信息化技术发展的同时,不断拓宽视野,与工程专业人员合作,借助现有信息化技术和平台搭载相关工程技术,为护理专业的高水平快速发展提供帮助。



Topic 1: 智能技术在患者安全管理中的研究; Topic 2: 面向患者的护理大数据系统的研发; Topic 3: 护士健康及辅助决策工具的研发; Topic 4: 虚拟现实技术在护理教育中的应用; Topic 5: 传感器和无线辅助设备的研发; Topic 6: 机器人在医院工作流程中的应用

图2 主题间距离图

表 2 各主题名称及所含高频词

主题序号	主题名称	高频词
Topic 1	智能技术在患者安全管理中的研究	care, health, monitoring, patient, robot, mobile, intelligent, assistive, automated, management, technology
Topic 2	面向患者的护理大数据系统的研发	project, care, patient, research, digital, data, echnology, nursing, assessment, system, software
Topic 3	护士健康及辅助决策工具的研发	health, healthcare, nurse, support, decision, technology, clinical, training, work, future, improve
Topic 4	虚拟现实技术在护理教育中的应用	care, technology, virtual, patient, digital, student, education, mobile, objective, nursing, future
Topic 5	传感器和无线辅助设备的研发	healthcare, monitoring, sensor, device, development, approach, wireless, data, patient, evaluation, disease
Topic 6	机器人在医院工作流程中的应用	care, health, covid, monitoring, system, robot, hospital, unit, design, evaluation, work

3.2 NSF 资助护理-信息-工程学项目对我国的启示

3.2.1 护理信息化改善护理服务质量 护理信息平台和健康干预 App 已然改善了护理服务质量^[15-16]。如远程会诊平台能带给患者长期信任并为护理干预提供支持^[17], 健康干预 App 能促进患者的健康行为, 同时减轻护士工作压力, 改善护理服务质量^[15]。但与此同时, 护理信息平台不可避免地为患者信息安全带来挑战^[17], 健康干预 App 的设计缺少临床专家的直接参与也是一大问题^[18]。Jaensson 等^[19]基于“信息技术-电子健康-护理专家”的跨学科团队开发了一种用于患者术后恢复监测和评估的网络应用程序, 为患者的使用提供了便利, 提高了护士评估患者术后恢复水平的效率和质量。此外, Sara 的项目 (#2026577) 提出了一个获取患者护理相关数据的新机制以保护患者隐私, 并提出一种异步通信工具将家庭健康助理与护理团队的成员直接关联, 通过多时空维度的数据收集和反馈机制实现协同护理工作。Ephraim 和 Sylvia 的项目 (#0428420、#0703692、#1035565) 也加入了隐私保护程序, 维持了患者基本隐私问题和健康需求的相互平衡。建议今后加大护理学与信息科学的合作力度, 从患者隐私安全等实际问题入手, 鼓励临床护理专家直接参与平台设计以促进护理信息平台的建设, 改善护理服务质量。

3.2.2 计算机图形优化虚拟现实场景革新护理教育范式 虚拟现实(virtual reality, VR)至今已发展至沉浸交互的仿真技术, VR 的出现让一种跨时空的网络虚拟学习空间正初步形成^[20], 推动了医学教育事业的发展^[21]。研究^[22-23]证实, VR 在护理教育中是可行的, 并且能显著提升护生和护士的学习、认知和精神活动水平以及护理技能熟练度、应对事件时的自信心和同理心。但在虚拟场景中, 学生缺少了对现实患者真实痛苦感受的人文关怀^[21]。Ephraim 的项目 (#1564065) 开发了一种物理虚拟病床, 使用动态计算机图形及各种传感器将出汗、呼吸、脉搏跳动、身体各部分温暖或寒冷的感觉表现出来, 将虚拟

患者的灵活性与人体模型的物理性相结合, 使人类特征充分融入到 VR 模型中, 让护生更好地感受患者所感。Ephraim 的项目 (#0803652) 提出了混合现实人类, 集成触觉、物理对象和生理监控以形成虚拟患者, 提高护生对于患者隐私部位(乳房、骨盆、前列腺)的护理评估和人际沟通技巧。建议今后学院教师-临床护理人员-VR 技术厂家三方通力合作, 针对不同疾病特点, 结合临床经验总结不同疾病患者的临床真实感受, 构建更加真实和个性化的虚拟患者模型, 让护生更加全面、真实地学习不同场景下的操作和沟通技巧。

3.2.3 智能辅助设备优化护理资源配置与护理决策效率 在护士短缺的背景下, 智能辅助设备协助护士解决临床问题的发展思路愈发受到重视^[24]。智能辅助机器人多存在于医院和老年护理机构中, 能帮助患者进行转移、行走、力量练习等, 高效减轻护士的工作负担, 降低护士职业相关疾病发生风险, 逐步发展为护士的潜在合作伙伴^[25-26]。护理决策需要护士收集患者信息, 综合分析后做出判断, 而护理团队的年轻化和临床繁重的工作加重了护理决策的困难^[27]。这样的背景下, 以辅助机器人为代表的智能辅助设备可以监测患者实时行为, 一定程度上降低了护士的工作量和职业相关疾病的患病风险, 但其无法监测患者所处环境的状态, 由此非接触式环境智能传感技术应运而生, 更加全面地协助护士满足患者需求^[28], 达到不良事件预警的作用。Sara 的项目 (#2026498) 开发了一款环境智能监控设备, 包括开发硬件(相机、深度传感器和麦克风)捕获患者受隐私保护的数据、开发算法识别患者临床活动行为的程度、医护人员输入互动信息、临床实验探究设备可用性。建议护理专业人员联合科技部门人员协同研究, 以更优的算法处理从环境中监测到的冗杂临床数据, 为优质护理决策提供保障, 实现患者-环境-护士一体化的智能护理路径, 通过智能技术赋能提升护理资源配置与决策效率。

3.3 本研究的局限性 本研究基于 Python 语言实

现了对 NSF 中护理-信息-工程学科相关立项的数据采集和分析,同时存在一定的局限性:(1)NSF 中每个检索词最多只开放 3000 条检索结果,这可能导致数据采集不完整。(2)截至目前,尚未有护理学专用的中英文停用词表和分词库,英文文本的分词多数是以空格和连接符为区分,可能会导致分词结果存在部分偏倚。(3)LDA 模型多用于对文本的情感得分进行分析,但对于政策类文本数据的分类效果有待纳入更大样本的数据进一步验证。

4 小结

本研究使用 Python 语言的网络爬虫和机器学习技术对 NSF 资助的护理-信息-工程学项目进行分析,从资助项目概况和项目主题两方面进行论述,为促进我国护理-信息-工程学科发展提供参考。在后续研究中,护理学者可根据国家最新政策,结合国情不断拓宽护理-信息-工程学交叉发展的深度和广度,推动我国护理事业的发展。

【参考文献】

- [1] 中华人民共和国工业和信息化部. 国务院关于落实《政府工作报告》重点工作分工的意见[EB/OL]. [2023-02-22]. https://www.miit.gov.cn/xwdt/szyw/art/2021/art_4b2496c6ed3844aeb2a8f0a3f0bef03b.html.
- [2] 周兰姝. 护理学科发展现状与展望[J]. 军事护理, 2023, 40(1): 1-4.
- [3] 王峥, 周滢, 王怡丹, 等. 美国护理工程学与护理信息学研究生培养机制及特征分析[J]. 中华护理教育, 2022, 19(5): 414-420.
- [4] 靳嘉林, 王曰芬, 巴志超, 等. 基金项目研究的主题挖掘与动态演化分析——以美国 NSF 数据中 AI 领域为例[J]. 情报学报, 2022, 41(9): 967-979.
- [5] NICHOLS L G. A topic model approach to measuring interdisciplinarity at the National Science Foundation[J]. *Scientometrics*, 2014(100): 741-754.
- [6] 刘海虹, 刘海宁, 车佳郡, 等. 不同代际居住模式老年人社交活跃度现状及影响因素——基于 Charls 数据的随机森林分析[J]. 军事护理, 2023, 40(2): 54-57, 62.
- [7] 曹煜隆, 单娇, 包小源, 等. 中国医院感染控制电子信息服务系统热点分析——基于 SIFIC 论坛的数据挖掘[J]. 中国感染控制杂志, 2020, 19(2): 164-168.
- [8] PYTHON SOFTWARE FOUNDATION. History and license [EB/OL]. [2023-01-21]. <https://docs.python.org/3/license.html>.
- [9] 周江杰, 王胜锋, 李立明. Python 爬虫技术在信息流行病学中的应用[J]. 中华流行病学杂志, 2020, 41(6): 952-956.
- [10] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation [J]. *J Mach Learn Res*, 2003, 3(4-5): 993-1022.
- [11] GONZALWZ G, VACULIK K, KHALIL C, et al. Experiences of women with interstitial cystitis/bladder pain syndrome; what can we learn from women's online discussions? [J]. *J Urol*, 2023, 209(1): 208-215.
- [12] TAMANG S, JIN Z, PERIYAKOIL V S. Revelations from a machine learning analysis of the most downloaded articles published in *Journal of Palliative Medicine* 1999-2018[J]. *J Palliat Med*, 2023, 26(1): 13-16.
- [13] 杨和稳. 人工智能算法研究与应用[M]. 江苏: 南京东南大学出版社, 2021: 248.
- [14] DONOHUE J C. Understanding scientific literatures—a bibliometric approach[M]. Cambridge: The MIT Press, 1973: 31-32.
- [15] 吴碧瑜, 戴亚端, 高玉玲, 等. 基于网络的 Cox 健康行为互动模式在初产妇产晚期质量管理中的应用[J]. 军事护理, 2023, 40(2): 63-66.
- [16] ROMSKAUG R, SKOVLUND E, STRAAND J, et al. Effect of clinical geriatric assessments and collaborative medication reviews by geriatrician and family physician for improving health-related quality of life in home-dwelling older patients receiving polypharmacy: a cluster randomized clinical trial[J]. *JAMA Intern Med*, 2020, 180(2): 181-189.
- [17] 肖雅婧, 邵圣文. 移动医疗信息平台在延续护理中的应用现状[J]. 护理研究, 2021, 35(7): 1229-1231.
- [18] ARNHOLD M, QUADE M, KIRCH W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older[J]. *J Med Internet Res*, 2014, 16(4): 34-51.
- [19] JAENSSON M, DAHLBERG K, ERIKSSON M, et al. The development of the recovery assessments by phone points (RAPP): a mobile phone app for postoperative recovery monitoring and assessment[J/OL]. [2023-02-21]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4704964/>. DOI: 10.2196/mhealth.4649.
- [20] 张静, 杨龙飞, 刘宇, 等. 基于虚拟现实技术的居家康复训练在脑瘫患儿中应用的范围综述[J]. 军事护理, 2023, 40(2): 71-75.
- [21] 王佳弘, 席文杰, 董丽丽, 等. 虚拟仿真技术在护理教育中的应用与挑战[J]. 中华护理杂志, 2020, 55(3): 401-404.
- [22] CHOI J, THOMPSON C E, CHOI J, et al. Effectiveness of immersive virtual reality in nursing education: systematic review [J]. *Nurse Educ*, 2022, 47(3): E57-E61.
- [23] KARDONG-EDGREN S, BREITKREUZ K, WERB M, et al. Evaluating the usability of a second-generation virtual reality game for refreshing sterile urinary catheterization skills [J]. *Nurse Educ*, 2019, 44(3): 137-141.
- [24] MAALOUF N, SIDAOUI A, ELHAJJ I H, et al. Robotics in nursing: a scoping review[J]. *J Nurs Scholarsh*, 2018, 50(6): 590-600.
- [25] YAMANAKA M, KAJIWARA K, KAKO J, et al. Nurses' attitudes toward the use of robots in healthcare[J]. *Nurs Open*, 2023, 10(3): 1936-1937.
- [26] 王开, 金爱萍, 李燕, 等. 脑卒中患者机器人辅助步态训练体验的质性研究[J]. 军事护理, 2023, 40(2): 15-19.
- [27] 刘晓娜, 潘红英. 护理决策支持系统的应用进展[J]. 中华护理杂志, 2018, 53(6): 735-739.
- [28] PHAM C T, VISVANATHAN R, STRONG M, et al. Cost-effectiveness and value of information analysis of an ambient intelligent geriatric management (AmbIGeM) system compared to usual care to prevent falls in older people in hospitals[J]. *Appl Health Econ Health Policy*, 2023, 21(2): 315-325.

(本文编辑: 王园园)