

近十年我国学习分析领域研究热点及发展趋势

——基于CNKI (2013~2023)核心期刊文献分析

刘睿, 向磊*

西华师范大学教育学院, 四川 南充

收稿日期: 2024年3月2日; 录用日期: 2024年4月1日; 发布日期: 2024年4月7日

摘要

本文的目的为系统梳理学习分析研究复杂脉络, 探索学习分析前沿热点主题, 追踪学习分析研究发展趋势。本文通过文献梳理对国内学习分析实践进行了探索性描述研究, 采用CiteSpace软件基于中国知网数据库中权威期刊绘制了2012~2023年学习分析领域研究图谱, 并通过高频热点词挖掘及聚类分析对该领域的研究现状、热点主题、发展趋势进行可视化整理。研究表明: 我国学习分析领域以数据挖掘、学习投入、精准教学、学习科学等为研究热点, 呈现出初步萌芽、快速发展、深入拓展三个主要阶段。结论: 我国目前学习分析研究正处于拓展应用领域与规范研究伦理阶段, 未来研究将趋向于更精准学习分析方法、更广泛的应用领域、更完善的分析机制及更符合伦理的研究规范。

关键词

学习分析, CiteSpace, 知识图谱

Research Hotspots and Development Trends of Learning Analytics in China in the Past Ten Years

—Literature Analysis of Core Journals Based on CNKI (2013~2023)

Rui Liu, Lei Xiang*

School of Education, China West Normal University, Nanchong Sichuan

Received: Mar. 2nd, 2024; accepted: Apr. 1st, 2024; published: Apr. 7th, 2024

*通讯作者。

Abstract

The purpose of this paper is to systematically sort out the complex vein of learning analytics research, explore the hot topics at the frontier of learning analytics, and track the development trend of learning analytics research. An exploratory descriptive study of domestic learning analytics practice was conducted through literature combing, and CiteSpace software was used to draw a research mapping of learning analytics field from 2012 to 2023 based on authoritative journals in the CNKI database, and visualize and organize the current research status, hot topics, and development trend of the field through high-frequency hot words mining and clustering analysis. The study shows that the field of learning analytics in China takes data mining, learning input, precise teaching, and learning science as the research hotspots, and presents three main stages: initial germination, rapid development, and in-depth expansion. Conclusion: China's current learning analytics research is in the stage of expanding application fields and standardizing research ethics, and future research will tend to be more accurate learning analytics methods, broader application fields, more perfect analysis mechanisms and more ethical research norms.

Keywords

Learning Analytics, CiteSpace, Knowledge Graphs

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自《地平线报告》在 2011 年提出学习分析将成为未来五年革新教育的技术后, 受到了我国教育领域学者的密切关注。近十年来, 学习分析已经从一个研究假想发展为一个成熟的研究领域。学习分析是指测量、收集、分析和报告有关学生的学习行为以及学习环境的数据, 用以理解和优化学习及其产生的环境的技术[1]。在教育研究领域, 学习分析通常用来揭示学习者的学习模式、学习策略和学习成果, 以及评估教学策略和教育政策的实施效果。随着人工智能赋能课堂, 产生了大体量、多模态的过程性学习数据, 这为学习分析领域带来了前所未有的数据支撑。然而, 如何针对复杂多源的课堂数据, 在前人基础上进一步挖掘潜在的学习规律, 是现阶段研究者面临的主要挑战。基于此, 本文通过 CiteSpace 文献计量分析方法, 系统梳理学习分析领域研究脉络、分析研究现状、追踪发展趋势, 为学习分析的未来研究提供参考。

2. 研究方法

2.1. 样本选择

本研究以“学习分析”、“学习分析技术”为检索关键词, 检索中国知网北大中文核心与 CSSCI 来源期刊, 时间跨度为 2012~2023 年, 文献检索截止 2023 年 12 月, 共检索文献 457 篇, 通过题录手动摘除与本研究相关度较低的 37 篇文献, 最后获得 420 篇有效文献。

2.2. 研究工具

本研究运用 CiteSpace6.2 文献计量软件绘制学习分析研究的知识图谱, CiteSpace 是一种基于科学引

文网络的可视化分析工具, 主要用于探索和分析学术文献之间的关系和趋势, 能帮助研究人员识别领域内的热点话题、重要作者、关键文献以及研究领域的发展趋势[2]。

2.3. 研究过程

研究首先在 CNKI 核心数据库中以时间轴为 2012~2023 进行主题检索, 筛选目标数据得到有效研究样本。其次将筛选所得数据经过 CiteSpace 自带的数据格式转换器进行预处理, 在导入 CiteSpace 对研究文献的作者、机构、关键词等不同节点类型信息进行可视化呈现。通过共现图谱、聚类视图和时区视图等, 解析我国学习分析研究领域的进展、热点及前沿, 希望从宏观上把握学习分析在我国的发展脉络。

3. 研究结果与分析

3.1. 文献年度分析

文献发表量的年度变化可以一定程度反映研究者对于某领域的关注度。了解学习分析研究的发文变化趋势, 可以帮助研究者洞悉发展趋势与发展水平。使用 Pathfinder 算法以年为单位切片统计, 将 2012~2023 年我国学习分析研究总体发文量绘制成年发文量图(见图 1)。可以看出, 学习分析研究自 2012 年开始呈现脉冲式上升趋势, 到 2016 年达到第一个峰值, 其后有短期的下降, 在 2019 年达到第二个峰值, 此后便呈现出快速下降趋势。这可能由于自 2012 学习分析技术引入我国便受到学者强烈关注, 2016 达到初步饱和, 2017 后人工智能技术在教育研究领域的广泛应用给学习分析研究带来了突破瓶颈的机遇, 于 2019 发文量再次达到峰值。此后学习分析研究迎来深入拓展阶段, 发文量迅速下降。

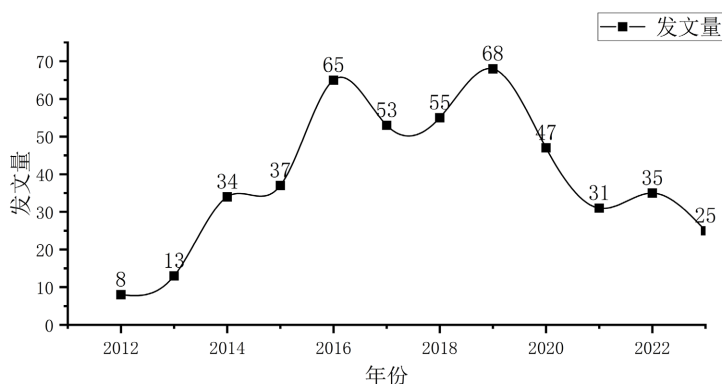


Figure 1. Chart of annual number of publications

图 1. 年发文量图

3.2. 作者分布

作者合作网络图谱见图 2。图中节点大小代表作者发文多寡; 连线的粗细则表示作者之间的合作强度[3]。在作者合作图谱中, 节点数共计 291, 连线数共计 181, 学习分析共被引网络密度为 0.0042。由图可知, 国内学习分析领域研究呈现出“部分集中、整体分散”状态, 仅有少量明显的学术共同体。其中三个较大合作群体为华东师范大学顾小清、东北师范大学赵蔚、北京师范大学陈丽, 其他的研究者合作范围小, 主要为同机构合作。

3.3. 研究机构分布

设置 CiteSpace 分析参数, 将 Institute 选项中阈值设定为 5, 表示学习分析研究发文数量 5 篇及以上

的机构, 以此确定对该领域关注度较高的研究单位, 如图 3 所示。从研究机构的发文量上看, 华东师范大学教育信息技术学系、北京师范大学教育技术学院、北京师范大学远程教育研究中心排名前三, 表明这三所大学是学习分析研究领域的领跑者。其中, 发文量最多的研究机构是华东师范大学教学信息技术学系, 频次为 31, 发文最早出现在 2012 年, 表明该研究机构对我国学习分析研究关注度最高, 具有重要的奠基作用。从研究机构的属性来看, 已有研究机构主要为国内 6 所部属师范大学。这表明我国学习分析研究形成了以北京师范大学教育技术学院、华东师范大学教育信息技术系为核心机构、师范类高校为重要研究力量的研究格局。研究机构在图中均为同单位间连线, 这表明同一机构内部合作紧密, 呈现出“自发自为”的研究特点, 而各研究机构之间的合作仅限于同单位, 没有跨校合作。



Figure 2. Author collaboration network mapping
图 2. 作者合作网络图谱



Figure 3. Mapping of collaborative networks of research institutions
图 3. 研究机构合作网络图谱

3.4. 关键词分析

3.4.1. 关键词频次统计分析

分析 420 篇研究文献的关键词, 在知识图谱中以圆圈来表示, 频次越高, 节点越大, 结果如图 4 所

示。如图可知,“学习分析”、“大数据”、“数据挖掘”、“在线学习”是出现频率较高的关键词,表明这些领域是近十年来学习分析领域研究的热点问题。



Figure 4. Keyword co-occurrence network mapping
图 4. 关键词共现网络图谱

关键词的中心度是一个用以测量节点在网络中地位重要性的指标,可以显示领域的结构和动态本质,中心性 > 0.1 即表明该关键词在文献中起重要作用[4]。“学习分析”的中心性数值为 1.58,是知识图谱中最核心的节点,是联系其他关键词的中心枢纽。“大数据”“学习分析技术”“人工智能”“个性化学习”的中心性数值均在 0.2 以上,是次要的联系节点,在该领域研究中具有重要影响。

3.4.2. 突显词分析

图 5 通过 CiteSpace 关键词突显分析展示了近十年中相关高频词汇在文献中被采用的时间轴,加粗的线条表示该高频词汇爆发的年度,例如第一高频词“大数据”高密度地出现在 2013~2015 年,说明在这个时间轴内,有很多文献资料都在研究大数据在学习分析领域的应用。到了 2018~2023 年,文献材料的高爆发词汇变成了学习投入、协作学习、多模态数据,说明研究者的关注度从早期的大数据如何应用逐步发展到了关注具体的数据取向与分析方向。

Keywords	Year	Strength	Begin	End	2013-2023
大数据	2013	3.52	2013	2015	
学习行为	2013	2.5	2015	2017	
个性化学习	2014	2.72	2016	2017	
教育大数据	2014	3.84	2017	2018	
数据挖掘	2013	2.77	2018	2019	
学习投入	2019	3.87	2020	2023	
协作学习	2017	2.94	2020	2023	
多模态数据	2018	3.08	2021	2023	

Figure 5. Keywords highlighted mapping
图 5. 关键词突显图谱

3.4.3. 关键词聚类图谱分析

采用 LLR 算法生成高频关键词聚类图谱, 显示出前 9 个聚类(见图 6)。其中节点数为 385, 连线数为 526, 网络密度为 0.0071。聚类模块值(Q 值) = 0.661, 大于 0.3, 说明聚类结构显著, 聚类平均轮廓值(S 值) = 0.913, 大于 0.7, 说明聚类的结果具有较高信度。下面结合前文高被引论文分析、高频关键词分析, 对学习分析领域研究主题进行系统梳理。



Figure 6. Keyword clustering mapping
图 6. 关键词聚类图谱

1) 学习分析实施技术研究。包含#0 学习分析技术, #3 人工智能, #8 认知网络分析。主要包含学习分析具体实施技术, 如语义分析、认知网络分析、社交学习与社交网络分析等。其中认知网络分析产生于教育大数据与学习分析快速发展的大背景下, 用于表征学习者认知网络结构的一种日益重要的研究方法[5]。认知网络分析将人类认知活动通过节点和连线组成的图形表示, 节点代表认知实体(如人、概念、思维过程等), 连线代表这些认知实体之间的关系或交互。通过分析认知网络的拓扑结构、节点属性和连接模式, 能揭示学习者认知活动中的模式和规律。社交网络分主要研究学习者在在线社交环境中的合作、交流及影响, 用以揭示学习者之间的社会关系、知识流动模式以及社交学习的效果。

2) 学习分析数据挖掘研究。包含#1 大数据, #2 教育大数据, #5 数据挖掘, 主要围绕着数据分析与数据挖掘展开。在线教学中产生了大量的学习数据, 如学习日志、在线讨论、作业和测验等, 通过开发有效的数据挖掘技术充分利用多模态学习数据, 能揭示学习过程数据的潜在模式、关联性, 进而挖掘学习规律, 深层理解学习发生机制。

3) 学习分析应用方向研究。包含#7 智慧教育, #9 学习科学, #10 精准教学。精准教学主要围绕个性化学习与自适应学习技术展开, 个性化学习旨在根据学习者的特征和需求, 提供定制化的学习体验和支持。研究者关注如何利用学习分析来推动个性化学习的实现, 并开发智能教育系统, 如智适应教学和推荐系统, 为个性化的学习提供资源和反馈。智慧教育则主要围绕数据驱动决策和可视化交互设计技术展开, 智慧教育倡导个性化学习, 而学习分析则提供了个性化学习支持的关键技术。

学习分析为智慧教育提供了数据支持和决策依据, 通过收集和分析学习数据, 学习分析可以揭示学

生的学习需求、行为模式和学习成果, 为教育工作者和决策者提供重要的信息, 帮助他们制定针对性的教学策略和教育政策。

而学习科学主要围绕学习者模型与学习轨迹分析展开, 学习分析则借鉴学习科学的理论和概念, 将其应用于实际的学习环境中。通过学习科学的理论支持, 学习分析可以更好地解读和解释学习数据, 揭示学习者的行为和学习过程的内在机制。学习者模型是对学习者特征、行为和表现进行建模和分析的方法。通过学习者模型来预测学习者的学习结果、个性化学习支持和诊断学习问题。学习轨迹分析旨在识别学习者在学习过程中的路径和模式, 以帮助改善教学和学习策略。

3.4.4. 演进阶段

时序分析图谱(图 7)能够用来展示文献的时序演化过程, 运用文献的时序演化, 可以发现关键概念的引入和发展, 重要研究方向的变化, 以及领域内的研究热点的兴起和衰退。通过对学习分析研究的关键词时区图谱的分析, 将学习分析研究划分为三个阶段。

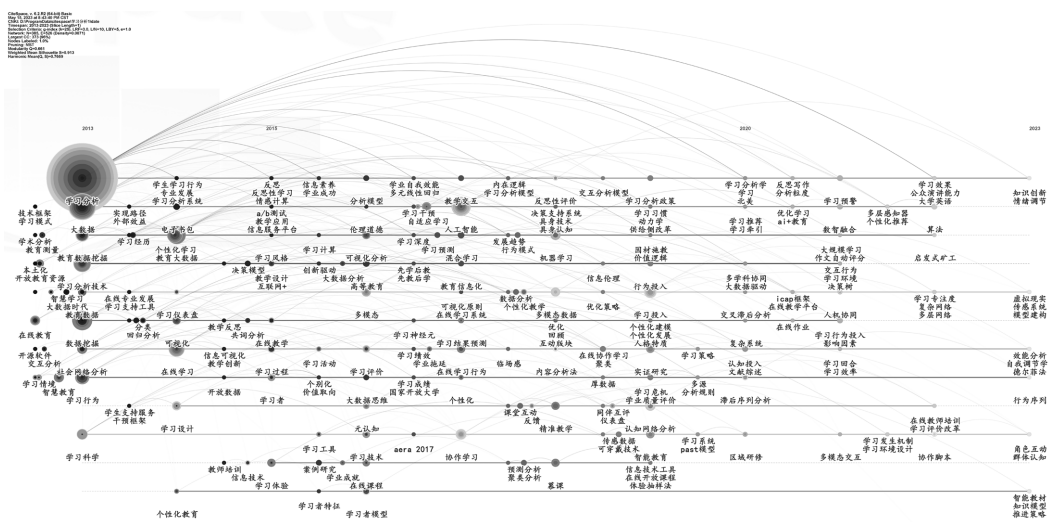


Figure 7. Timing analysis mapping
图 7. 时序分析图谱

- 1) 初步萌芽阶段: “学习分析”方法探索阶段(2012~2015 年)
本阶段的热点关键词有“学习分析”“大数据”“数据挖掘”“学习模式”“教育数据”, 重点关注的是学习分析的基本过程及理论基础。该阶段发表的文章开启了我国教育学习分析的先河, 被引频次较高。数据挖掘是学习分析中的一种主要的技术手段, 用于从学习数据中发现隐藏的模式、关联和知识。数据挖掘可以帮助揭示学习者的行为模式、学习特征和学习趋势, 从而提供对学习过程和学习者的深入理解。“学习分析”“数据挖掘”呈现出了较高的突现度, 且持续时间长达十年。
- 2) 快速发展阶段: “学习分析”技术创新阶段(2015~2020 年)
本阶段的热点关键词有“情感计算”“可视化分析”“多模态交互”, 重点关注的是学习分析的方法与技术, 涉及学习者模型、认知网络分析、社会网络分析、多模态学习分析和时序性学习分析等。许多研究者将其他领域的技术引入到学习分析中, 形成了学习分析技术热潮, 促进学习分析领域研究飞速发展。
- 3) 深入拓展阶段: “学习分析”实践规范阶段(2020~2023 年)
这一阶段的热点关键词有“滞后序列分析”“学习专注度”“多学科协同”“学习评价改革”, 重

点关注的是学习分析的应用领域, 涉及自我调节、协作学习、教育会话、情感分析等具体研究问题。这充分展现出学习分析是一个涉及多学科领域的综合性研究领域, 它需要多学科的协同合作来实现对学习过程和学习者的全面理解和支持。在学习分析技术热潮后, 研究者们进入了更深入的研究, 开始关注学习分析的机制、规范、政策及普适性。

4. 研究结论与展望

4.1. 研究结论

采用 CiteSpace 对 2012~2023 年国内学习分析文献进行计量研究和知识图谱可视化分析, 系统梳理了国内学习分析的研究现状与热点。文献计量分析结果表明: 国内学习分析研究成果颇丰, 大致经历三个阶段。发文量在 2012~2015 年整体呈现迅速上升态势, 属于初步萌芽期, 2015~2020 年呈现稳定波动趋势, 属于快速发展期, 2020 年后发文相对下降, 但研究更加深入, 属于深入拓展期。

作者及研究机构网络图谱显示: 我国学者以及机构已建立相对稳定的合作交流网络, 呈现出了“部分集中、整体分散”状态, 仅形成以华东师范大学顾小清、东北师范大学赵蔚、北京师范大学陈丽为主的三个较大学术共同体。其次同一机构内部合作紧密, 几乎没有跨校合作, 呈现出“自发自为”的研究特点。

关键词聚类及时序分析图谱结果显示: 学习分析的研究热点从早期的方法机制探索、技术实证研究发展到后期关注于教育领域的高质量的学习分析制度规范。近年来, 随着人工智能技术快速发展, 自然语言处理、多模态学习分析、自适应学习等学习分析领域取得了较多研究成果, 是当前该领域的研究热点。

学习分析在发展过程中仍存在一些局限。研究主要关注于学习者的学习行为数据, 较少探索精准化的多模态学习分析。表现为学习分析研究中较多关注学习者外显行为数据频次, 缺少关注学习者的生理指标、行为模式与环境交互等隐含信息。对学习投入数据的研究不足, 同时学习分析收集大量学习者数据, 涉及隐私及伦理问题, 目前少有研究者关注相关问题。

4.2. 研究展望

通过以上文献知识图谱可视化分析, 研究认为我国学习分析技术的未来发展应关注以下研究方向。

4.2.1. 更精准的学习分析方法

目前学习分析技术采用的数据具有行为偏向性, 基于残缺学习数据的学习分析与学习干预, 会导致学习分析缺乏精准性和适切性, 使得学习者长期处于狭隘的“信息茧房”中, 无法得到全面发展, 甚至会出现显著的认知偏见[6]。因此, 研究者可深化对学习过程性数据分析研究, 深入剖析学习过程中的关键因素和机制, 包括认知过程、情感因素、社交互动等要素。其次通过整合多源数据和采用先进的数据挖掘技术, 揭示更细粒度的学习过程和学习者行为之间的关联, 从而提供更准确的个性化支持及教学建议。此外可融合多模态数据分析方法, 多模态学习分析通过多维度全面采集和分析学习过程动态数据, 使面向过程与真实学习情境的学习评价成为可能[7], 融合多种数据类型如学习日志、视频观看记录、眼动数据等, 进行多模态数据联合分析, 全面了解学习者的学习过程, 同时挖掘不同数据之间的潜在关联和影响, 为个性化学习和智慧教育提供更全面的视角和支持。

4.2.2. 更全面的应用领域

目前学习分析聚焦于数据挖掘及精准教学。未来的研究可更加关注学习者之间的社会互动、合作和知识共享, 以深入揭示社会学习的效果和机制。社交学习分析还可以探索如何设计有效的社交学习环境

和促进社会互动, 以提高学习者的合作能力和学习成果。同时研究者也可将学习分析应用到教育政策及决策方面, 学习分析可以为教育政策制定者和决策者提供重要的数据和见解。未来的研究应更加关注将学习分析应用于实际的教育政策制定和教学决策中, 从而提高教育质量、促进公平性和推动教育改革。

4.2.3. 更完善的分析机制

确保学习分析机制的完善性是未来学习分析研究的重要主题, 学习分析机制的完善性需要考量学习分析过程的公平公正性。目前研究中学习分析算法的不完善和数据的不全面是影响学习分析公平和公正的主要因素。未来学习分析研究中需采取多种措施促进学习分析公平, 主要包括以下几个方面。

数据收集的公正性, 包括避免数据的偏见和歧视, 确保数据采集的透明度和合法性, 以及保护学习者的隐私权。数据收集过程应遵守伦理原则和相关法规, 确保学习者的权益和利益得到保护。

数据分析的透明性, 在学习分析过程中, 应提供对数据分析方法和算法的透明度。这意味着清楚地描述数据分析的过程、使用的算法和模型, 以及分析结果的解释和推断。透明的数据分析有助于减少偏见和歧视的可能性, 使决策过程更加公正和可理解。

多样化的数据源, 确保学习分析使用多样化的数据源, 以减少数据偏向性的影响。依赖单一的数据源可能会导致数据的片面性和不完整性, 从而影响学习分析的公正性。使用多样化的数据源可以提供更全面和客观的视角, 更好地理解学习者的多样性和差异性。

反思和修正偏见, 学习分析团队应反思和修正可能存在的偏见和歧视。审查和评估分析方法和结果, 关注可能存在的偏见来源, 并及时采取措施纠正或修正。通过不断的反思和改进, 提高学习分析的公正性和准确性。

持续监控和评估, 对学习分析的公平性和公正性进行持续的监控和评估。定期审查分析方法和结果, 评估其对学习者的影响和效果, 以及及时修正任何发现的学习分析过程的不公正或偏见。持续的监控和评估是确保学习分析公平公正的关键步骤。

4.2.4. 更符合伦理的研究规范

以人工智能、大数据分析和自适应学习等技术为基础的学习分析为教学提供了支持, 但也无可避免涉及许多隐私及伦理问题。学习分析过程涉及大量的学习者数据, 研究者应关注如何确保学习者数据的安全和隐私保护, 确保在学习分析中收集、存储和处理数据时遵守隐私和数据保护原则。这些原则包括获得学习者的知情同意, 保护其个人隐私和数据安全, 以及仅使用数据用于明确的研究目的。学习分析团队应遵循相关法规和伦理准则, 制定和执行严格的数据隐私和安全政策, 并获得伦理审查委员会的同意。在科学研究中, 对个人身份的保护至关重要, 采用匿名和去标识化处理方法, 可减少个人身份的暴露和数据关联的风险。通过采用随机标识符或加密技术, 可以确保学习者的身份得到保护, 同时仍能够进行有效的研究分析。

基金项目

西华师范大学博士启动项目“农村小规模学校信息化发展路径及对策研究”(20E009)。

参考文献

- [1] 顾小清, 张进良, 蔡慧英. 学习分析: 正在浮现中的数据技术[J]. 远程教育杂志, 2012, 30(1): 18-25.
- [2] 陈悦, 陈超美, 胡志刚, 王贤文. 引文空间分析原理与应用: CiteSpace 实用指南[M]. 北京: 科学出版社, 2014: 16, 76.
- [3] 张雅俐, 卢红. 基于 CiteSpace 的“双创”教育研究文献分析[J]. 教育理论与实践, 2019, 39(30): 21-24.
- [4] 李征, 周小勇. 近 20 年国内外语教育技术研究评述与未来展望: 基于 CiteSpace 的可视化分析[J]. 中国电化教育,

2018(6): 91-96+122.

- [5] 王志军, 杨阳. 认知网络分析法及其应用案例分析[J]. 电化教育研究, 2019, 40(6): 27-34+57.
- [6] 李海峰, 王炜. 学习分析研究: 基本框架和核心问题——2022 版《学习分析手册》的要点与思考[J]. 开放教育研究, 2022, 28(6): 60-71.
- [7] 张家华, 胡惠芝, 黄昌勤. 多模态学习分析技术支持的学习评价研究[J]. 现代教育技术, 2022, 32(9): 38-45.