

基于 CiteSpace 的在线学习者画像 研究现状与可视化分析

贾礼远¹, 董克²

(1. 安徽开放大学 教务处, 安徽 合肥 230022; 2. 蚌埠学院 经济与管理学院, 安徽 蚌埠 233030)

摘要:在人工智能、大数据技术赋能教育数字化转型的背景下,探究学习者画像研究现状、研究热点及其发展趋势,对推动在线教育领域学习者画像研究的深入开展和实际应用具有重要意义。以中国知网(CNKI)学习者画像研究论文为研究对象,运用 CiteSpace 软件对样本文献的发文机构、科研合作网络、关键词共现、关键词聚类等进行可视化分析。研究表明,有关学习者画像研究呈增长态势,作者机构合作网络逐渐成形,研究主题主要集中在算法研究、画像系统构建和画像应用与服务三个方面,未来需加强对数据采集的隐私保护、画像系统标准化建设和画像输出的评价利用等方面的研究。

关键词:CiteSpace;学习者画像;文献计量分析;可视化分析

中图分类号:G727;G434

文献标识码:A

文章编号:2097-0625(2024)02-0036-07

一、引言

用户画像概念最早是由 A. Cooper 于 1999 年提出,他将用户画像定义为“基于用户真实数据的虚拟代表”^[1]。用户画像作为大数据时代提供精准服务的关键工具,其应用已经扩展到电子商务、金融和公共服务等众多行业。目前,用户画像在教育领域,尤其是在在线教育领域已得到广泛应用。在大数据环境下,把学习者作为用户展开研究,基于教育大数据构建学习者画像体系,精准描绘学习者人群特征,探索学习者对在线学习的需求或潜在需求,全面提高学习者的学习效率,有很大的学术发展空间和参考价值,且具有较强的实践指导意义^[2]。

选取 2015 年至 2023 年中国知网全文数据库收录的主题为“学习者画像”的期刊论文为研究对象,利用 CiteSpace 软件构建共词网络,绘制学习者画像领域的作者机构共现图谱、高频关键词聚类图谱和时间线图谱,旨在揭示国内学习者画像研究的热点主题、发展前沿及趋势走向。

二、数据来源与研究方法

(一)数据来源

以中国知网为文献数据源,设定检索条件为“主题(精确)”,检索词为“学生画像+学习者画像”“用户画像*学生+教学+学习”,来源类别为“学术期刊”,时间范围不做限制,检索时间为 2024 年 1 月 3 日,共检索出 603 篇相关文献,经过筛选,剔除如文件、征稿启事及报道等非研究类内容,共获取有效文献 375 篇,构成本文进行文献计量分析和可视化分析的基础数据,所收集的数据包含作者、标题、摘要、关键词、发表年份以及作者单位等关键信息。为确保分析的准确性和一致性,对关键词和作者单位名称进行了标准化处理,包括合并同义关键词、核实并统一同一作者单位的不同表述方式等。

(二)研究工具与方法

1. 研究工具

利用 CiteSpace 软件对样本文献进行可视化研究分析,具体步骤为:首先,利用 CiteSpace 软件的数

收稿日期:2024-02-28

基金项目:安徽省高校省级质量工程教学研究项目“基于学习者画像的远程教育个性化学习支持服务研究”(项目编号:2020jyxm0264);安徽省高校自然科学研究项目“基于学习者画像的在线学业监控与预警机制研究”(项目编号:KJ2020A1217);全国教育科学规划教育部重点项目“成人学习者网络学习行为数据挖掘与网络教学质量监控机制构建”(项目编号:DKA220345)

作者简介:贾礼远(1983—),男,山东滕州人,讲师,硕士。研究方向:计算机教育应用。

据格式转换工具将从 CNKI 导出的 Refworks 格式文件转化为软件可识别的数据格式;其次,建立项目并进行参数设置,主要包括确定时间跨度、选择时间切片(Time Slicing),设置分析网络类型、关联强度和节点阈值筛选参数;最后,采用聚类静态(Cluster View-static)和展示整个网络(Show Merged Network)的可视化方式呈现最终分析图谱^[3]。

2. 研究方法

依托 CiteSpace 软件采用文献计量学方法识别并可视化学习者画像研究趋势与动态,文献计量学是借助文献出版数量,采用统计学相关方法来描述、评价和预测科学技术现状与发展趋势的图书情报分支学科^[4]。通过提取文献的发表年份、期刊名、作者、

机构、关键词等信息,研究某学科发文数量、分布及变化情况,从而揭示某学科的发展现状及未来研究热点^[5]。从年度发文量、研究机构分布与研究合作情况、高频关键词等多维度进行文献计量分析和可视化分析,展现 2015—2023 年学习者画像研究的发展情况。

三、学习者画像研究现状分析

(一)年度发文量

通过分析 2015—2023 年的发文数量,可以揭示学者们关于学习者画像研究的关注度变化及其发展趋势。本研究对纳入分析的 375 篇文献按照年度发文量进行了统计(见图 1)。

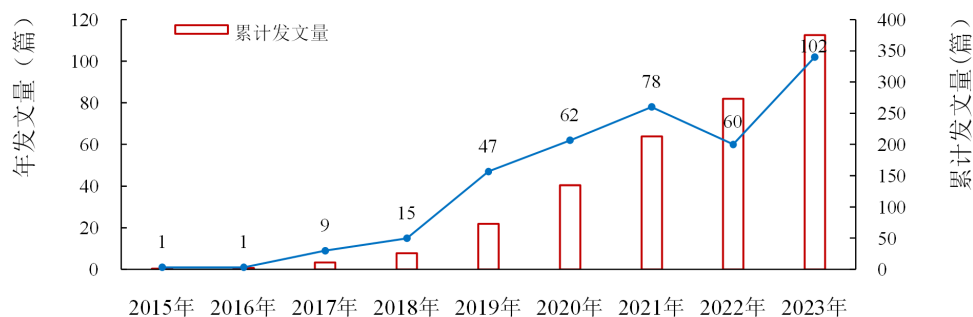


图1 学习者画像期刊论文年度发文量

由图1可见,国内对于学习者画像的研究起步较晚,CNKI收录的我国最早的一篇关于学习者画像的期刊文献为《大数据时代高校学籍预警机制的探索与研究》(2015年发表),该领域的发文量在2018年攀升至两位数,并在2023年增长至102篇。利用SPSS软件对纳入文献的年累计发文量与时间进行多项式拟合,得出年累计发文量与时间之间的关系式为: $y=6.996x^2-22.923x+16.31$ (y 表示年累计发文量, x 表示时间),调整后 $R^2=0.996$,模型拟合度佳,可以预见,未来学习者画像领域的研究将会得到进一步的深入与拓展,发文量也会进一步增加。

2018年,教育部印发《教育信息化2.0行动计划》,文件指出要将教育信息化作为教育系统性变革的内生变量,支撑引领教育现代化发展,推动教育理念更新、模式变革、体系重构,使我国教育信息化发展水平走在世界前列,发挥全球引领作用,为国际教育信息化发展提供中国智慧和方案。相关政策文件的出台显著助推了该领域的发展,近五年来,学习者画像研究保持了持续的高热度,这表明其在学术和实践层面受到了广泛的关注。

(二)发文机构

为揭示学习者画像研究领域的核心机构,以发文第一作者所在单位为依据,对发文数量在3篇及以上的机构进行了统计(见图2)。华东师范大学教育学部、东北师范大学信息科学与技术学院、华中师范大学人工智能教育学部和西南大学教育学部等12所机构发文量均在3篇以上,12所机构共计发文51篇。从机构类型来看,主体为师范类大学,表明师范类大学在学习者画像研究方面具有较大的潜力和发展空间,间接体现出学习者画像研究需要有较强的教育技术、教育科学、信息技术等相关专业背景支持。

(三)科研合作网络分析

CiteSpace提供了三个层次的科学合作网络分析,分别为微观的作者合作网络(Co-Author Analysis),中观的机构合作网络(Co-Institutions Analysis)和宏观的国家/地区的合作(Country/Territory)。通过研究作者之间合作关系可以反映出学习者画像的主要研究力量,利用CiteSpace绘制了作者和机构共现图谱(见图3)。如图3所示,排名前列的作者合作网络分别是东北师范大学信息科学与技术学院和教

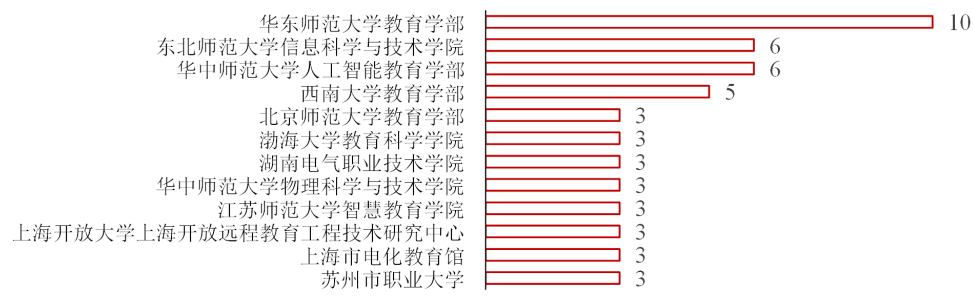


图 2 学习者画像机构发文量

育部数字化学习支撑技术工程研究中心的合作网络、华中师范大学的合作网络、成都中医药大学的合作网络以及华东师范大学教育学部与上海市电化教育馆和上海开放大学上海开放远程教育工程技术研究中心等的合作网络。其中前 3 个合作网络均是校内不同机构之间的合作,第 4 个则是校际及社会机构之间的合作。在当前的学习者画像研究领域中,作者间的合作主要集中于同一学校内部,而校际和社会机构之间的合作模式更能为学习者画像研究注入新的活力,从而加速学习者画像的理论探索和技术实现。

(四)文献的期刊分布

对纳入分析的 375 篇文献所涉及的来源期刊进行了统计分析,结果显示文献来自 221 种不同的期刊,显示出较为分散的分布状况。具体来看,有 12 种期刊的发文量达到或超过 4 篇(见表 1),如表 1 所示,其中 5 种为核心期刊。特别是《中国教育信息化》《电化教育研究》《中国电化教育》等,它们在学习者画

像领域进行了深入研究和讨论,显示出该研究方向的重要性。此外,对 2015 年至 2023 年间各期刊发文量的时序分析(动态迷你图,横轴代表年份,纵轴代表文章数量)显示,《电脑知识与技术》《中国教育信息化》《现代信息科技》《远程教育杂志》《现代教育技术》等期刊在过去三年的发文量呈现上升趋势或保持较高稳定状态,这一态势表明当前学习者画像研究是一个备受关注且具备较高研究价值的主题。



图 3 学习者画像研究核心作者合作网络图谱

表 1 发文 4 篇及以上期刊发文趋势

序号	期刊名称	发文篇数	是否为核心期刊	2015—2023 年发文趋势
1	《电脑知识与技术》	14	否	
2	《中国教育信息化》	12	否	
3	《电化教育研究》	11	是	
4	《现代信息科技》	9	否	
5	《中国电化教育》	9	是	
6	《无线互联科技》	9	否	
7	《远程教育杂志》	7	是	
8	《信息与电脑》	6	否	
9	《软件导刊》	6	否	
10	《现代教育技术》	5	是	
11	《开放教育研究》	4	是	
12	《电子技术与软件工程》	4	否	

(五)高被引文献

被引量是反映论文影响力大小的一项重要指标。论文被引量越大,说明论文受关注的程度越大,论文的影响力就越大^[6]。在 2015 年至 2023 年期间,有关学习者画像研究的被引用文献次数最多的前十篇文献如表 2 所示,从中可以看出,排名前十被引用最频繁的论文发表在《中国电化教育》《电化教育研究》《开

放教育研究》等数字教育研究的核心期刊上,发表时间集中在 2017—2020 年,研究方向主要为大数据环境下学习者画像构建与实证研究、基于学习者画像的个性化教学及评价研究和基于学习者画像的精准支持服务研究,以及个性化学习路径生成和学习资源推送等。

表 2 学习者画像高被引文献(排名前十)

序号	篇名	作者	刊名	发表时间	被引频次	下载频次
1	开放式教学下的学习者画像及个性化教学探讨	陈海建等	《开放教育研究》	2017-06-05	115	3 909
2	基于大数据的精准教学:生成路径与实现条件	任红杰	《黑龙江高教研究》	2017-09-05	101	3 191
3	大数据环境下在线学习者画像的构建	肖君等	《开放教育研究》	2019-08-05	83	4 415
4	基于学生画像的项目式学习评价指标体系研究	余明华等	《电化教育研究》	2021-03-01	81	6 283
5	基于可视化学习分析的研究性学习学生画像构建研究	余明华等	《电化教育研究》	2020-12-09	73	5 306
6	泛在学习环境下基于活动理论的移动 MOOC 设计及实证研究	肖君等	《中国电化教育》	2017-11-01	64	2 633
7	基于大数据深度画像的个性化学习精准服务研究	刘海鸥等	《图书馆学研究》	2019-08-15	63	3 687
8	基于学习者画像建模的个性化学习路径规划研究	唐烨伟等	《电化教育研究》	2019-09-24	61	4 275
9	基于 xAPI 的在线学习者画像的构建与实证研究	肖君等	《中国电化教育》	2019-01-03	60	3 071
10	基于学习画像的精准个性化学习路径生成性推荐策略研究	师亚飞等	《中国电化教育》	2019-05-09	60	4 556

四、学习者画像研究热点分析

(一)关键词共现分析

通过提取某研究领域内能够表达文献核心内容的关键词并分析主题词频次的高低分布,可以研究该领域发展动向和研究热点。本文运用 CiteSpace 进行关键词共现分析,网络类型选择“Keyword”,按照“TOP 50 per slice”提取节点,生成的关键词图谱共计 498 个节点(关键词)、1 064 条连线,网络密度为 0.008 6,说明学习者画像研究领域内的子领域之间的关联不是非常紧密,不同的研究机构或作者更侧重

于各自的专题。在排除了“学习者画像”以及与检索策略直接相关的关键词后,频次大于 5 的关键词见表 3。

由表 3 可见,“大数据”“数据挖掘”“数据分析”“机器学习”“深度学习”等反映了学习者画像领域在多模态数据处理和分析方面的需求,以及采用算法从数据中提取学习者特征进而构建画像的重要性;“智能教育”“在线学习”“混合式教学”“人工智能”和“智慧校园”这些关键词体现出学习者画像应用在教育领域向数字化、网络化、智能化转型过程中的适切性;“精准教学”“精准思政”“高职院校”“开放大学”则体

现了学习者画像应用的深度和广度。这些关键词表明学习者画像领域正快速发展,且越来越多地依赖于

技术特别是数据分析和 AI 来驱动教育的数字化,强化学生的学习体验,提升教学的质量和效率。

表 3 学习者画像研究热点关键词

序号	关键词	频次	中心性	序号	关键词	频次	中心性
1	大数据	74	0.50	12	学习行为	9	0.04
2	用户画像	45	0.29	13	精准思政	8	0.04
3	数据挖掘	19	0.08	14	开放教育	8	0.03
4	智慧校园	13	0.07	15	深度学习	8	0.04
5	精准教学	13	0.07	16	在线教育	7	0.04
6	人工智能	12	0.07	17	高职院校	6	0.02
7	在线学习	12	0.04	18	聚类分析	6	0.01
8	学习分析	11	0.08	19	数据驱动	5	0.04
9	智能教育	10	0.05	20	机器学习	5	0.02
10	知识图谱	10	0.05	21	开放大学	5	0.02
11	数据分析	9	0.03	22	混合式教学	5	0.01

(二) 关键词聚类分析

CiteSpace 的关键词聚类功能可以明确某研究领域的热点和发展趋势^[7]。在关键词共现图谱的基础上,采用对数似然比算法(Log-Likelihood Ratio, LLR)提取聚类标签,进而构建关键词聚类图谱(图 4)和时间线图(图 5)。该聚类图谱的模块性指标(Modularity, Q 值)为 0.658 1,而轮廓系数(Silhouette, S 值)为 0.902 9,两项指标的高值说明划分出来的各个聚类具有很好的内聚性和间隔性。进一步梳理图 4 和图 5 中的关键词聚类标签,可以发现当前学习者画像研究的 3 个主要热点方向。

1. 学习者画像算法研究,该方向的关键词聚类名称包含:大数据、数据挖掘、数字化、数据驱动等,其时间跨度为 2017—2023 年。算法是数据分析过程中的重要环节,涵盖从数据收集到分析的不同阶段。杨长春等提出了一种基于随机森林的双向特征选择算法(RFBFS),用于解决构建智慧校园学生画像时的数据缺失和高维特征问题,提高了画像的分类精度^[8]。翟鸣宇等优化了一种 K-prototype 算法,结合汉明距离和欧氏距离的混合度量来进行更有效的聚类分析,进一步助力管理人员进行教育决策^[9]。凌玉龙等从学生消费画像和精准资助两个角度对校园消费数据进行挖掘研究,证明了改进 K-means 算法在精准资助中的应用价值^[10]。

2. 学习者画像系统构建研究,该方向的关键词聚类名称包含:学习者画像、用户画像、标签模型、知识图谱等,其时间跨度为 2015—2023 年。此类研究聚焦构建可操作的学习者画像系统,如画像系统架构、技术实现和用户界面设计等方面。肖君等提出 ODAS 开放学习分析概念模型,并设计了一套基于 ODAS 的在线学习者画像构建流程,为学习者画像教学应用与评估提供了方法^[11]。余明华等提出了基于可视化学习分析的研究性学习学生画像流程框架,并借助数据可视化技术呈现了面向学生的个人画像和面向教师的群体画像^[12]。张治等从指标体系层、数据实践层以及数字画像层设计了基于数字画像的综合素质评价框架,并对数字画像的具体应用策略进行了阐述^[13]。

3. 学习者画像应用与服务研究,该方向的关键词聚类名称包含开放大学、在线学习、学习分析、学习预警等,时间跨度为 2015—2023 年。此类研究聚焦学习者画像的应用场景、服务模式创新等,而非平台搭建或算法开发。韩凤霞将基于学生特征画像方法应用于学籍状态动态监测,实现了对学籍状态异常学生进行预警和帮扶指导^[14]。陈海建等基于学习者画像对个性化教学研究进行了新的探索,开辟了个性化教学研究的新思路^[15]。伍智鑫基于学习者画像优化高校贫困生的认定体系,为高校贫困生精准扶贫工作提

供“量化”“可视”的决策支持^[16]。黄文林提出依托教育大数据构建服务高校精准思政工作的学生画像方法和工作路径^[17]。开放教育办学体系内学习者画像的研究基本涵盖了自适应学习、学习预警、个性化学习资源推荐等多个应用层面。

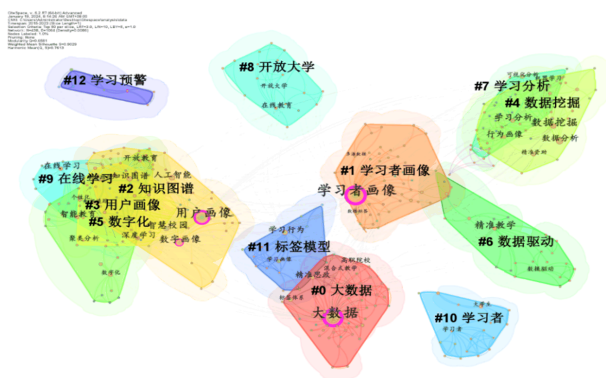


图4 学习者画像研究关键词聚类图谱

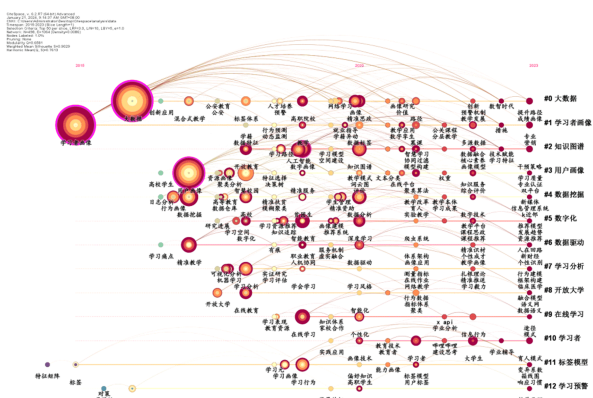


图5 学习者画像研究关键词聚类时间线图

五、结论与讨论

学习者画像领域的研究在我国起步较晚,但近年来该领域在国内的关注度不断上升。一方面,期刊载文量逐年增加、作者机构合作网络逐渐成形;另一方面,从最初利用多维数据源、技术和算法来提取学习者的多维特征形成画像标签,到基于不同理论视角的画像框架搭建,再到画像系统被应用到在线教育的各个层面,研究脉络逐渐清晰,研究成果日渐丰富。国内学习者画像的研究取得了一定进展,但在学习行为数据采集、学习者画像系统标准化建设、画像结果评价利用等方面仍有进一步研究的空间。

(一)数据采集的隐私保护

数据采集和挖掘是学习者画像构建的基础,实时

数据更新是动态画像时效性的关键,画像的准确度和实用性在很大程度上取决于数据采集的维度和质量。随着深度学习和实时数据分析的发展,画像构建可采集的数据种类不再局限于在线学习平台的学习行为数据,学习者的肢体反应、面部表情、眼动追踪等通过特定设备所收集的生物反应数据以及通过学习者的穿戴设备、移动设备或学习环境中的传感器收集的位置、环境互动等信息都可能被用作画像建构的数据源。无处不在的数据挖掘势必会威胁到主体的隐私安全和数据控制权^[18],因此,在画像构建过程中必须杜绝数据权力滥用倾向,确保数据的安全和合理利用,这就要求遵循数据最小化、透明度和知情同意等原则。此外,还需要加强与政府、学校、其他教育机构等相关方的合作,共同制定数据管理、隐私保护的政策和标准,建立健康、安全的数据环境。

(二)学习者画像系统缺乏标准化框架

当前,种类繁多的在线学习平台共存,平台间的数据异质性导致了所谓的“数据孤岛”现象,再加上不同研究机构和研究者采用的算法和建模方法各不相同,加上对于学习者画像系统有效性的实证研究较少,这都在很大程度上阻碍了这一技术的普及和规模化应用。因此,有必要通过行业合作,学术研究和政策指导等方式,构建一个广泛认可的标准化框架,逐步推动在线教育领域学习者画像的标准化发展。

(三)画像输出及评价反思

基于学习者画像的可视化输出结果是一种通过数据分析和图形表示来描述和理解学习者性格特征、学习行为和学习成效的手段。作为一种工具,其输出结果首先依赖于数据采集的准确性和真实性,数据收集错误、不完整或者过时,都可能导致错误的评价和决策。另外输出结果的质量在很大程度上还取决于画像系统本身的科学水平,如果画像系统设计存在缺陷,则会导致输出结果不符合实际或结果解读困难,难以为教育工作者所理解,更难以以为学习者提供有效支撑。基于客观事实、数据分析所呈现的教育评价结果固然有利于教育评价标准化,但偏执地追求客观理性,人特有的主体气质将难以评判诠释^[19]。因此,教育工作者使用可视化输出结果时,应当使用它们作为促进教育公平和效率的有益工具,而不是一个冰冷的评判标准。

参考文献:

- [1] ALAN C. About Face: The Essentials of Interaction Design[M]. US: Wiley, 2014.
- [2] 谈松英, 杨敏. 大数据分析视野下的用户画像及其在开放教育中的运用研究综述: 基于知网 2010—2017 年全文收录文献的数据分析[J]. 中国教育信息化, 2018(16): 1-6.
- [3] 李璐. 基于 CiteSpace 的国内图书馆用户画像研究现状与可视化分析[J]. 江苏科技信息, 2023, 40(27): 13-17.
- [4] 李雯, 姜仁贵, 解建仓, 等. 基于文献计量学的城市洪涝灾害研究可视化知识图谱分析[J]. 西安理工大学学报, 2020, 36(4): 523-529.
- [5] 王娟, 陈世超, 王林丽, 等. 基于 CiteSpace 的教育大数据研究热点与趋势分析[J]. 现代教育技术, 2016, 26(2): 5-13.
- [6] 宗淑萍. 基于普赖斯定律和综合指数法的核心著者测评: 以《中国科技期刊研究》为例[J]. 中国科技期刊研究, 2016, 27(12): 1310-1314.
- [7] 段春雨, 蔡建东. 国际泛在学习领域知识图谱研究[J]. 现代远程教育研究, 2016(1): 85-95.
- [8] 杨长春, 徐筱, 宦娟, 等. 基于随机森林的学生画像特征选择方法[J]. 计算机工程与设计, 2019, 40(10): 2827-2834.
- [9] 翟鸣宇, 程建, 王苏桐, 等. 基于 K-prototype 聚类的大学生教育画像分析[J]. 大连理工大学学报(社会科学版), 2021, 42(6): 22-31.
- [10] 凌玉龙, 张晓, 李霞, 等. 改进 kmeans 算法在学生消费画像中的应用[J]. 计算机技术与发展, 2021, 31(10): 122-127.
- [11] 肖君, 乔惠, 李雪娇. 大数据环境下在线学习者画像的构建[J]. 开放教育研究, 2019, 25(4): 111-120.
- [12] 余明华, 张治, 祝智庭. 基于可视化学习分析的研究性学习学生画像构建研究[J]. 中国电化教育, 2020(12): 36-43.
- [13] 张治, 刘小龙, 徐冰冰, 等. 基于数字画像的综合素质评价: 框架、指标、模型与应用[J]. 中国电化教育, 2021(8): 25-33.
- [14] 韩凤霞. 大数据时代高校学籍预警机制的探索与研究[J]. 中国教育信息化, 2015(19): 46-49.
- [15] 陈海建, 戴永辉, 韩冬梅, 等. 开放式教学下的学习者画像及个性化教学探讨[J]. 开放教育研究, 2017, 23(3): 105-112.
- [16] 伍智鑫. 基于一卡通数据挖掘的高校贫困生精准扶贫应用研究[J]. 湖南科技学院学报, 2019, 40(4): 122-123.
- [17] 黄文林. 基于学生画像分析的高校精准思政探索[J]. 东北大学学报(社会科学版), 2021, 23(3): 104-111.
- [18] 鹿星南, 高雪薇. 人工智能赋能教育评价改革: 发展态势、风险检视与消解对策[J]. 中国教育学报, 2023(2): 48-54.
- [19] 吴龙凯, 程浩, 张珊, 等. 智能技术赋能教育评价的时代内涵、伦理困境及对策研究[J]. 电化教育研究, 2023, 44(9): 19-25.

Current Situation and Visual Analysis of Online Learner Portrait Research Based on CiteSpace

JIA Liyuan¹, DONG Ke²

- (1. Department of Teaching Management and Affairs, Anhui Open University, Hefei Anhui 230022, China;
2. School of Economics and Management, Bengbu University, Bengbu Anhui 233030, China)

Abstract: In the context of artificial intelligence and big data technology empowering the digital transformation of education, it is of great significance to explore the current situation, research hotspots, and development trends of learner portrait research to promote the in-depth development and practical application of learner portrait research in the field of online education. Taking the related research papers of CNKI on learner portraits as the research object, CiteSpace software was used to conduct visual analysis of the publishing institutions, scientific research cooperation network, co-occurrence of keywords, and keyword clustering in the included literature. The research results show that scholars are showing a growing trend in learner portrait research, the cooperation network of author institutions is gradually taking shape, and the research is mainly focused on algorithm research, portrait system construction, and portrait application and services. In the future, it is necessary to strengthen research on privacy protection for data collection, standardization of portrait system construction, and evaluation and utilization of portrait outputs.

Keywords: CiteSpace; learner portrait; bibliometric analysis; visual analysis

[责任编辑 许炎]