# 大数据环境下在线学习者画像的构建

# 肖君 乔惠 李雪娇

【摘 要】在线学习中,学习者特征不明、学习支持服务针对性不足、学生流失率高等问题是教学利益相关者关注的重点。大数据环境下在线学习者画像的构建对识别、分析学习者特征并据此提供针对性学习支持服务具有重要意义。本研究以开放学习分析架构为基础,从目标(Object)、数据(Data)、分析(Analysis)、服务(Service)四个角度提出"ODAS"开放学习分析概念模型,并设计了一套具体可行的、可操作的基于ODAS的在线学习者画像构建流程,包括画像构建目标、数据收集、标签分析、画像服务输出四个阶段,从而为学习者画像教学应用与评估提供方法。本研究进而以上海开放大学的在线学习者为对象,从学习者基本特征、在线学习行为和学习路径三个维度构建高风险学习者画像,开展画像标签分析。研究表明,学习者画像能反映高风险学习者的学习能力、行为表现、学习路径等特征,并具有教学情境性,能为针对性教学设计及在线学习支持服务设计提供数据支撑。

【关键词】教育大数据;开放学习分析;在线学习者画像;精准支持服务

【作者简介】肖君,上海开放大学上海开放远程教育工程技术研究中心,研究方向:在线学习、学习分析、学习者画像(xiaoj@sou.edu.cn)(上海 200433);乔惠,东航技术应用研发中心有限公司,研究方向:学习分析、学习者画像(上海 200131);李雪娇,华东师范大学教育信息技术学系,研究方向:学习分析、学习者画像(上海 200237)。

【原文出处】《开放教育研究》(沪),2019.4.111~120

【基金项目】全国教育科学"十三五"规划 2016年度国家一般课题"大数据下在线学习用户画像的构建及其应用研究"(BCA160053)。

#### 一、引言

随着互联网的发展,开放、共享式在线学习模式日益被教育领域认可,然而在线学习领域普遍存在学生流失率高、课程完成度低等问题。受师生分离影响,现有在线教学设计及各类学习支持服务如电话、邮件、社交软件等不能针对不同学习者提供精准的支持服务,有必要精准分析学习者特征,并提供针对性学习支持服务(曾建勋,2017)。

用户画像(Persona)最初由库珀(Cooper)提出(穆德,2007),又被称为用户角色,常作为勾画目标用户、产品定位、联系用户诉求与设计方向的有效工具,广泛应用于商业领域(余孟杰,2014)。阿里、京

东、百度等企业使用用户画像凝聚用户,提升用户体验(肖君等,2019)。用户画像可基于大数据规模存储和机器学习算法定期对全量数据进行计算和挖掘构建用户,提供用户标签的使用和查询服务。例如,有学者(Iglesias et al.,2012)对Web站点的网络日志数据进行深度挖掘,通过对用户行为模式进行聚类,创建不同群体的用户画像。王顺箐(2018)在数据采集基础上构建用户画像,以用户需求分析为核心构建图书馆智慧推荐系统,把握用户的个人喜好和动态需求,实现无差别推荐。学习者画像构建主要指基于某种个性化学习目的建立的原型用户和教学代理角色(Van Mulken,1998),通过建立多维度的画像标

签展示体系,为更精准的学习支持服务设计提供数据支撑。学习者画像借助用户画像的应用潜力,以学习分析为基础,对教育大数据进行标签化处理,多维度分析学习者特征,深入理解开放学习者,从而提供精准的学习支持服务,提升学习者的学习体验。本研究将理论研究和案例研究相结合,首先基于开放学习分析方法设计了基于ODAS的在线学习者画像构建流程;然后,以上海开放大学在线学习者为例,从学习者基本特征、在线学习行为和学习路径环境三个层面输出了高风险学习者画像;最后,提出了针对教学干预的措施,从而帮助教师提高在线教学效果。

# 二、研究综述

# (一)开放学习分析

大数据使海量学习数据挖掘成为可能,学习分析作为分析学习者学习过程的重要方法得到了广泛关注。然而在实践中,人们逐渐意识到传统的学习系统或平台在数据采集、分析、处理、展示方面有专有形式,通用化程度低,不利于向不同利益相关者推广。因此,以满足不同利益相关者需求为目的的开放学习分析研究成为该研究领域的趋势。西门子等(Siemens et al.,2011)认为开放学习分析通过开放学习分析过程、算法和技术可满足不同利益相关者(学习者、教育工作者、管理人员和决策者)的需求(Siemens et al.,2011)。沙提等(Chatti et al.,2017)将开放学习分析看作是学习分析的深入,代表了向新学习分析模型的转变。新模型将"开放性"考虑其中,包括开放的目标、技术、实践等,同时也满足学习分析参考模型的四个维度,即"What""Why""How""Who"。

开放学习分析架构可指导开放学习分析的实践与应用,是开放学习分析研究的重要内容。国际上较典型的开放学习分析架构包括 SoLAR、Apereo、JISC等机构提出的架构模型。SoLAR基于利益相关者(学习者、教育工作者、管理人员以及研究人员或数据分析师)的场景,构建了整合式模块化学习分析系统(Siemens et al.,2012),强调满足不同利益相关者的需求。Apereo基金会的艾伦(Aaron)等学者为 Apereo LAI项目设计了开放学习分析钻石模型,其中定义了由"沟通"主导的四个主要领域:收集、存储、分

析和行动(Apereo, 2017);利用 xAPI 数据标准进行多平台数据的收集和存储,在大型数据集的基础上开展数据挖掘、处理、建模等,形成开放的仪表盘可应用于不同系统。英国联合信息系统委员会(JISC, 2015)提出的开放学习分析服务架构与 Apereo 开放分析架构有很多共同点,并发布了每个组件的 API,使用者可以在架构之上构建额外的服务,也可以选择使用部分或全部组件,很好地体现了学习分析服务的开放性。本研究通过对已有架构的组成分析发现,开放学习分析架构有一定的共性:架构设计以目标为导向,以满足不同利益相关者的需求;和学习分析一样,学习者数据是开放学习分析开展的基础,但更强调多来源的数据;使用开放的技术、算法等进行数据分析,以满足利益相关者需求;分析结果以服务的形式可接入不同系统。

综合已有文献,本研究认为开放学习分析是"开放"的"学习分析",是面向多维度、多渠道数据来源的学习分析,即为了适应如今开放在线学习环境,满足不同利益相关者的学习分析需求,利用开放的标准、技术、方法,整合异构的学习分析技术,收集和分析更全面的大数据,提供可视化等分析服务形式,以帮助优化在线学习过程。开放学习分析在学习分析研究的基础上更强调目标、技术、服务等的开放性,适用于当今开放的在线学习环境。

#### (二)学习者画像

学习者画像是用户画像在教育领域中的应用,也是学习分析的一个新应用领域。学习者画像和学习者特征的根本区别是它以群体特征为核心,将学习者进行群体分类描述,为不同学习者提供精准的支持服务(肖君等,2019)。大数据环境下,我们可收集多样性的全局教育大数据,多维分析学习者的特征并以标签的形式呈现,形成学习者画像,为分析学习者特征、设计精准支持服务等提供依据。学习者画像分析方法包括描述性分析(What happened)、诊断性分析(Why did it happen)、预测性分析(What will happen)以及规定性分析(How can we make it happen)(王小雪等,2017)。

国内外有关大数据环境下学习者画像的研究主要聚焦在通过学习者画像识别不同群体的学习者,

从而为不同学习者提供相应的教学服务。例如,美 国普渡大学的"Signals"项目通过分析学习者的行为 画像做出预警,帮助师生更好地完成教学活动,促进 有效学习并提高学习者学习成绩(刘艳华等,2014)。 英国开放大学对学习者学习数据进行全方位的采集 分析并开展学习质量评估,不断完善学习支持服务 并为学生提供个性化指导,以提高学习者的学习成 果,降低中途退学率(肖君等,2015)。国外研究人员 还对学习者画像构建进行了研究。霍利等(Holley & Oliver, 2010) 通讨风险预测模型构建学习者画像, 识 别高危、中危、低危学习者,为不同学习者提供针对 性学习支持服务。国内学者认为学习者画像基于教 育大数据,通过分析学习者数据记录标签化学习者 特征,更精准地描述学习者需求以及预测学习者行 为,从而提供更精准的信息服务(肖君等,2019;陈海 建等,2017)。

当前国内大数据环境下的学习者画像研究尚未 形成开放的流程和方法,以应对日益复杂的在线学 习者特征分析。如何基于开放学习分析方法形成在 线学习者画像构建流程,以面对开放化的在线学习 环境、多样化的学习需求、多来源的学习大数据,从 而满足不同利益相关者的多样化需求?如何通过规 模化的实践研究,对我国在线学习者画像、高风险学 习者画像开展研究,从而提出针对性教学干预服务 以帮助教师提高在线教学效果?这些问题都亟需教 育研究者关注。

#### 三、基于开放学习分析的在线学习者画像构建

# (一)"ODAS"开放学习分析框架

开放学习分析架构是开放学习分析研究的重要 内容,指导开放学习分析在实践中的应用。

基于对典型开放学习分析架构的分析,本研究提炼出开放学习分析在教学实践中应用的四个核心要素:1)目标(Objective)。虽然早期的开放学习分析架构没有明确地将目标作为架构的一部分,但均强调利益相关者的重要性,开放学习分析的实践需满足利益相关者的需求。沙提等(Chatti et al., 2017)设计的开放学习分析生态系统明确将目标作为系统的一部分,强调目标的重要性。2)数据(Data)。不管是学习分析还是开放学习分析,其实施基础均是数

据。开放学习分析强调数据及其获取的开放性。数 据的开放性表现在数据可以对利益相关者开放,例 如学习者可以访问自己的数据。xAPI数据标准的兴 起,使跨平台、跨终端的大数据收集成为可能,开放 学习分析中的多维大数据得到了丰富。3)分析 (Analysis),即对数据的分析,包括数据分析相关的分 析方法、算法、模型等,其具体内容由目标决定。开 放学习分析强调数据分析的开放,分析方法、算法等 可嵌入开放学习分析平台,也可以由利益相关者自 定义。4)服务(Service)。为了实现互操作性、增加重 用性等目标,IT领域提出了"面向服务的架构"(service-orientation architecture)。服务作为基本结构,可 为更多解决方案提供指导方针(Thomas, 2007)。开放 学习分析结果输出也具有同样的作用,可接入多种 学习系统,以实现结果输出的开放性。因此,服务也 是开放学习分析实践应用的重要内容,包括仪表盘、 预警服务等。

本研究进而提出了开放学习分析概念模型 "ODAS"(由四个要素的英文首字母构成,见图1)。整个模型以目标(O)为导向,在数据(D)收集的基础上,进行数据分析(A),其分析结果以服务(S)的形式输出,并应用于教学,最终又指向目标(O),判断目标(O)是否达成,若没有则进行下轮迭代,直到完成目标(O)。



图 1 "ODAS"开放学习分析概念模型

(二)大数据环境下基于开放学习分析的在线学 习者画像构建流程

基于"ODAS"开放学习分析概念模型,本研究提出了大数据环境下基于开放学习分析的在线学习者画像构建流程(见图2)。

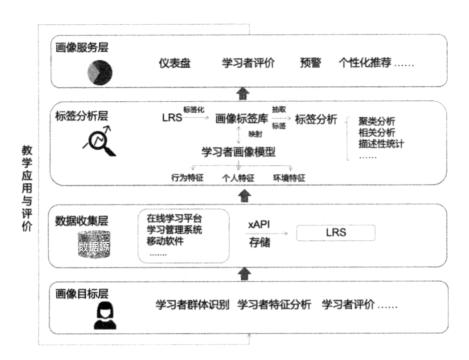


图2 基于ODAS的开放学习者画像构建流程

本研究所设计的在线学习者画像构建流程是一 个以目标为导向的循环系统框架,包括明确画像构 建目标、数据收集、标签分析、画像服务输出等四个 关键阶段,分别对应ODAS模型中的目标、数据、分析 和服务,最终在教学应用与评估中确定目标是否达 到。1)画像目标层:学习者画像围绕学习者特征分 析展开,可以为在线教学设计、学习支持服务、效果 评估等提供支持。因此其目标可聚焦到学习者群体 识别、学习者特征分析和学习者评价等方面,以满足 学习者、教师、管理人员等利益相关者的需求。2)数 据收集层:针对不同画像目标进行大数据收集,包括 学习者基本信息、各类行为日志以及调查数据等,其 中在线学习者画像的行为数据来源于开放学习环境 下所有学习者的各种学习经历,其产生的数据以 "Statement"的形式收集跨平台、跨终端的xAPI格式 的数据(Bautista, 2013), 最终以 xAPI活动流的形式存 储于学习记录库。3)标签分析层:指根据画像目标 抽取画像标签库中的标签进行数据分析。画像标签 库的形成由学习者画像模型决定,根据画像模型中 的标签定义对收集到的数据进行标签化处理,即通 过对底层数据的筛选、组合得到标签,并存储于画像 标签库。基于标签的数据分析方法主要有聚类分 析、相关分析等,其分析方法的选择由画像目标决定。4)画像服务层:是标签分析结果的输出,根据不同的目标其输出形式也不一样,例如有关学习者特征的分析报告将由仪表盘输出,而以学习支持为目标的画像输出将以个性化推荐、预警等服务形式输出。基于开放学习分析的画像服务都是可重复使用的,可接入不同的平台系统。5)应用与评价层:指将画像服务应用于教学,并通过评估判断是否达到了最初的画像目标,如果没有则进行下一轮的画像分析应用,由此整个流程构成一个闭环。

在线学习者画像模型是标签分析层中画像标签库形成的依据,是整个学习者画像流程的主要内容。本研究基于社会认知理论进行在线学习者画像模型的设计。美国心理学家班杜拉基于环境、人和行为特征的互动关系考察学习者的心理活动以及社会表现,把人的心理活动看成是个体、环境及其行为三者之间的互动系统(蔡耀杰,2011)。在社会认知理论的指导下,本研究从"个人层面一个人特征""行为层面一行为特征""环境层面一学习环境特征"三个维度构建在线学习者画像。在线学习者个人特征主要指成人学习者在在线学习情境下的学习特征,包括基本信息、问题解决能力、自我效能感、时间管理

能力和计算机水平等二级标签。学习者行为特征主要指学习者在与学习环境的交互过程中产生的学习行为,既与学习活动有关,又与学习者的社会化特征有关,可以分为在线操作性行为、课程或资源学习行为、总结反思行为和交流互动行为(乔惠,2018)。在此基础上,我们可分析学习者画像的各种标签,如投入程度、活跃程度等。学习者环境特征主要指学习者所使用的在线学习平台、课程资源支持服务等相结合的技术应用环境。在线学习环境下常见的学习方式有异步在线学习以及以直播课为主的同步学习等,在此基础上我们可以分析学习者在线学习轨迹、学习路径等各种环境特征。

# 四、在线学习者画像构建案例

# (一)研究目的

本研究以上海开放大学的在线学习者为研究对象,围绕在线学习者画像模型的个人特征、行为特征和环境特征三个维度,通过基于ODAS的开放学习者画像构建流程探究高风险学习者群体的相关特征,从学习者基本特征、在线学习行为和学习路径环境三个层面输出高风险学习者画像,并进行针对性的教学干预措施设计。

# (二)数据收集

本研究从基本特征、行为特征、学习环境特征三个维度开展数据收集。

#### 1.学习者基本特征维度的数据收集

本研究通过上海开放大学的在线学习平台发放成人学习者特征调查问卷,具体包括性别、年龄、最高学历、时间管理能力(Britton & Tesser,1991)、问题解决能力(蔡慧英,2016)、计算机水平(蔡慧英,2016)、自我效能感(蔡慧英等,2017)。最终收集的问卷有1611份,其中,男性学习者有966人,占60%;专科学历的学习者最多,占44.6%,其次是高中及以下学历,占33.7%,本科和其它学历分别占16.9%和2.74%;20~30岁学习者最多,占40.2%,其次是30~40岁和41~50岁学习者,分别占35.9%和11.4%。可以看出,男性学习者占多数,本科以下学历超过80%,多数学习者年龄在20~40岁之间。

2. 在线学习行为特征维度的数据收集 本研究收集了2018春季学期47513名在线学习 者的学习平台大数据日志,得到了有关资源访问行为的1048576条数据记录、有关发帖行为的290211条数据记录以及有关作业行为的40105848条数据记录。根据等效交互理论,远程教育中只要有一种以上的交互处于较高水平就有可能带来更满意的教育体验(李晓兰等,2015)。由于属于"学生一学生""学生一教师"交互的BBS活动参与人数相对较少,本研究将"学习者—内容"的交互作为学习平台中的主要交互形式,包括学生与平台资源、作业的交互。

因不同学习者修读的课程数量不同,为保证一致性,本研究将所有数据以课程为单位进行平均值处理,利用xAPI标准对不同数据类型、跨系统的数据进行标准化处理,最终得到47513名学习者的ID、性别、年龄段、成绩、平均作业时长、平均资源访问时长、平均资源访问次数、平均答题次数、平均资源访问率、平均作业完成率。为了方便计算,我们以量纲一致为原则进行数据标签化处理,通过平均作业时长与平均资源访问时长相加得到投入程度标签,由平均资源访问次数得到活跃程度标签,由平均答题次数得到努力程度标签,由平均资源访问率与平均作业完成率相加得到完成度标签。

# 3. 学习路径环境特征维度的数据收集

在对于学习环境特征维度的实验研究中,本研究选取上海开放大学"经济法概论"在线课程。该课程包含 28 个主题资源,共 317 位学生参与在线学习。本次实验选取学习者在该在线学习环境中的学习行为数据"学习者的成绩、学习课程资源的时间、学习次数"等,共得到 14533 条数据记录,作为学习者学习路径的分析依据。其中,学习行为数据来源于学生在线学习过程中对课程内容的访问日志,本研究对每个学生的日志按访问时间进行排序,建立学生对于这门课程不同章节的访问序列(序列中的每个节点表示一个章节),并将日志中访问时间长度赋予节点,作为节点属性,从而构建每个学生的学习路径。

#### (三)高风险学习者画像标签

# 1.基本特征维度下高风险学习者的画像标签

在线学习者基本特征主要指成人学习者在在线 学习情境下的学习特征。诺尔斯的成人教育学理论 指出成人学习者在学习过程中倾向于问题中心或任 务中心为主的学习,并且具有独立的自我概念,学生的自主性和独立性较大,能够自己选择学习内容、制定学习计划(田山俊等,2011),因此本研究选取学习者基本特征问卷调查得到的"时间管理能力、问题解决能力、计算机水平、自我效能感"四个标签,作为学习者基本特征的评价依据。基于SPSS19.0的问卷可靠性分析结果表明,各要素的Cronbach's Alpha大于0.85,可靠性较高,可以进行数据分析。

为了识别在基本特征上存在高风险的学习者,本研究以时间管理能力、问题解决能力、计算机水平、自我效能感为指标,利用SPSS19.0进行 K-means 聚类分析(见表一和表二)。表一中四个指标的 Sig 值 均小于 0.01,说明聚类结果有效。表二则描述了聚类后的四类学习者数量和各特征的均值,问卷采用5点式李克特量表,分数越高,代表越差。其中,高风险学习者人数占比为 0.87%,表现为问题解决能力、时间管理能力较差,计算机水平低,学习不自信且不适应线上学习方式,学习态度消极。

2. 在线学习行为特征维度下高风险学习者的画 像标签

本研究选取投入程度、活跃程度、努力程度、完

成度作为学习者行为特征的评价依据,利用 SPSS19.0进行描述性分析和相关分析(见下页表 三)。由于61岁及其以上的学习者较少,不具有普遍 性,本研究不予分析。41~50岁的学习者整体表现 最佳,其投入程度、活跃程度、努力程度和完成度整 体上表现最好,成绩最高。成绩最差的是20岁及其 以下的学习者,其投入程度、活跃程度、努力程度和 完成度均最低,是行为特征维度下的高风险学习者 群体。综合来看,高风险学习者呈现以下特征:人数 占比11.68%、年轻活泼、时间管理能力差、成绩差、 线上学习不活跃、不按时完成作业、所花费的时间和 精力均较少。

为了充分理解影响学习者成绩的特征因素,本研究选取投入程度、努力程度、完成度、成绩进行Persona相关分析,相关分析结果表明成绩与投入程度(r=0.269,p<0.01)、活跃程度(r=0.056,p<0.01)、努力程度(r=0.227,p<0.01)、完成度(r=0.752,p<0.01)在0.01水平上均显著相关,其中完成度与成绩是强相关关系,而活跃程度表现出弱相关关系。教师可着重督促高风险学习者完成作业和资源学习,进而改善学习效果。

	聚类		误差		E	G:
	均方	df	均方	df	Г	Sig.
问题解决能力	191.696	3	0.113	1607	1690.430	0.000
自我效能感	180.876	3	0.126	1607	1439.661	0.000
计算机水平	149.312	3	0.180	1607	828.188	0.000
时间管理能力	195.941	3	0.147	1607	1335.107	0.000

F检验应仅用于描述性目的,因为选中的聚类将被用来最大化不同聚类中的案例间的差别。观测到的显著性水平并未据此进行更正,因此无法将其解释为是对聚类均值相等这一假设的检验。

表二

#### 最终聚类分析结果

	聚类						
	1. 中危学习者	2. 高风险学习者	3. 低危学习者	4.优秀学习者			
问题解决能力	2.70	3.93	1.95	1.13			
自我效能感	2.78	3.43	2.09	1.23			
计算机水平	2.55	3.89	1.88	1.18			
时间管理能力	2.83	3.96	2.09	1.24			
案例数	296	14	728 57				
缺失	0	0	0	0			

#### 表三

#### 描述性统计结果

	人数	投入程度	活跃程度	努力程度	完成度	成绩
20及其以下	5548	3572.84	5.93	243.11	0.87	75.84
21 ~ 30	21543	5151.59	7.55	260.56	0.95	82.51
31 ~ 40	16781	7055.88	9.88	291.64	0.99	86.71
41 ~ 50	3374	9167.21	12.85	310.17	1.02	87.63
51 ~ 60	222	11117.96	16.60	314.02	0.92	76.89
61 及其以上	23	6127.15	2.08	483.02	0.42	39.81

# 3.环境特征维度下高风险学习者的画像标签

环境特征维度包含学习者学习方式、学习终端 等多种标签。学习路径是其中一种综合性标签,作 为学习者的学习行为轨迹,能反映学习者资源学习 的顺序。本研究中学习路径是指在线学习者学习课 程资源的先后顺序。本研究选取学习者的成绩、学 习课程资源的时间、学习次数等数据进行学习路径 标签分析,并针对学习者与"经济法概论"课程主题 之间的交互行为进行路径分析,将学习者按总成绩 分为四类:不及格(0~60)为高风险学习者、中等 (60~75)为中风险学习者、良好(75~90)为低风险学 习者、优秀(90~100)为优秀学习者。在生成学习路 径之后,本研究对节点行为进行压缩,控制模型的复 杂度,得到学习路径结果(见图3)。图3左图中优秀 学习者群体学习过程顺序性和条理性较强,基本按 照课程的前后章节顺序进行学习,在课程后期,对前 序章节的复习也均有规律地进行:右图中高风险学 习者群体的学习路径明显杂乱,章节的访问也无明 显规律,学习过程中随意性较强。经过对学习路径 的对比和分析,本研究发现高风险学习者群体成绩 差、深度学习能力差、资源访问行为随机、学习习惯 差、深度学习不够、学习态度消极等。

# 五、结论与讨论

(一)学习者画像能反映高风险学习者的学习能力 在基本特征维度的画像分析中,高风险学习者 人数占比0.87%,其问题解决能力(3.93)、自我效能感 (3.43)、时间管理能力(3.96)、计算机水平(3.89)较差(值 越高代表越差),反映了该类学习者学习不自信且不 适应在线学习方式,学习态度消极。教学设计者在 设计教学时,要考虑该类学习者的学习能力特征,针 对性地降低学习难度,并及时提供学习支持服务,减

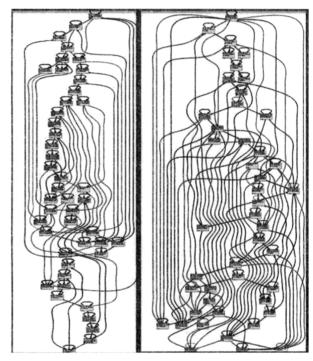


图3 学习者群体学习路径分析

注:左图是优秀学习者群体的路径;右图是高风险学习者 群体的路径。

少该类学习者流失的可能性。具体策略可包括: 1)做好学习风格测评,比如通过平台设置相关问卷评测学生的学习风格,以了解学习者的学习方式、思维活动方式等;2)做好学习者学习进度比较,即结合教学内容测评点分析学生的学习进度,并和标准进度、同班学习进度进行对比;3)开展学习者与同班同学的比较和学习者的阶段性比较;4)做好课程的形成性评价,帮助教师在教学过程的中后期及时发现问题并进行调整。

本研究邀请上海开放大学教师从画像标签的选择和画像服务的价值等视角对输出的高风险学习者

画像进行访谈。调查发现,教师普遍认为时间管理能力和问题解决能力是影响在线学习效果最重要的因素。此外,教师反映高风险学习者画像能够帮助其理解学习者的能力特征,对于针对性地设计不同学习难度的教学材料和教学活动具有重要参考价值。有效的自主学习可以提高学习效率(Paris & Paris,2001; Gvasser & Mc Namara,2010),教师可通过为学习者提供信息和机会等支持服务,使学习者成为有策略、有动力、独立的学习者。本研究也强调教师应该关注学生的问题解决和时间管理等自主学习能力,并强调教师应该为学生提供针对性学习支持服务,提高学习者自主学习能力,降低流失可能性。

(二)学习者画像能反映高风险学习者的行为表现 通过对学习者行为特征的分析,本研究发现20 岁及以下年龄段的学习者群体在学业成绩表现上最 不理想,投入程度值仅为3572.84(最高值为 11117.96),活跃程度值仅为5.93(最高值为16.6),努 力程度值仅为243.11(最高值314.02),完成度值仅为 0.87(最高值1.02),均排在最后且与最高值相差较 大。相关分析结果表明,上述四个特征均是学业成 绩的影响因素,特别是完成度与成绩的相关性最 高。教师可对该类学习者提供学习方法的指导,在 学习过程中及时发送学习任务完成情况的反馈和预 警,以促进该类学习者自我反思。具体策略可包括: 1)加强对学习者学习意愿的引导(方旭,2015);2)加强 个性化学习任务提醒,提供多种方式的信息推送,如 提醒参加教学活动、完成作业等;3)提供微信群、直 播课堂等多种社会化移动教学互动模式,实现课堂 临场感(王广新等,2010;腾艳杨,2013);4)提供学习 情况预警,如截止日期预警、成绩预警等。

本研究邀请上海开放大学教师对输出的高风险学习者画像进行评价,教师反馈本研究输出的学习者画像能够帮助其更好地监督教学过程,提高教学效果。本研究借助上海开放大学的在线学习平台针对高风险学习者实施个性化预警,如网页提醒、短信提醒、微信提醒等,通过前后两个学期的数据对比,学生的作业完成率、活动参与率、按时率等都有较大提升,平均提高5%。顾静相等(2007)认为远程教育学习支持服务应该考虑学习者的总体差异、分散程

度和不同需要;姜蔺等(2013)认为设计有效的学习指导并提供及时的学习支持服务可以提高学习者的学习积极性和学习意愿。可见,在提供及时的学习预警支持服务的同时,针对不同学习者提供不同的学习预警服务,可提高学习者的学习积极性和学习意愿。

(三)学习者画像能反映高风险学习者的学习路 径特征

画像分析结果表明优秀学习者和高风险学习者的学习路径具有明显差异。优秀学习者的课程主题学习路径比较有规律,具有"学习一复习巩固"等可循环的特征。高风险学习者则未按课程安排的进度进行学习,学习活动比较随机,在不同单元之间无序跳转。优秀学习者的学习路径对于优化在线学习路径、合理安排在线学习单元序列具有参考价值,其学习方法值得高风险学习者借鉴。具体策略包括:1)将优秀学习者的学习路径作为高风险学习者的参考学习路径。2)针对学习者个人特征和学习进度,为高风险学习者提供基于作业或学习目标的学习资源推荐等。

本研究邀请相关教育研究者就高风险学习者画像的可用性进行访谈。访谈结果表明,基于数据分析输出的学习者画像更有说服力,为认识高风险学习者的学习行为提供了实践依据。同时,学习者画像具有可视化的特征,能够更生动地理解学习者的学习过程,教学设计人员的注意力应从学习内容转向学习活动和学习路径。杜兰德等(Durand et al., 2013)认为学习设计可根据学习知识的特点和学习材料的顺序,为学习者推荐合适的学习路径。王楠等(2009)基于学习活动框架理论指出学习活动是一个依据学习者能力变化而变化的实体,具有多方面的可能性,是学习设计的焦点。本研究也证明了在线学习过程中学习活动和学习路径的重要性。

#### (四)学习者画像具有教学情境性

学习者画像的生成因教学情境而异。本研究生成的三类高风险学习者画像既有共性特征,也有个性特点,需要采取针对性干预措施以提升教学效果。

本研究从学习者基本特征、在线学习行为和学习路径等层面输出了高风险学习者画像,不同层面的高风险学习者画像有共同点也有不同点。例如,

高风险学习者时间管理能力差、学习过程中时间投 入少、学习路径混乱,均表现出学习能力差等特点, 这类学习者不适应在线学习方式,容易流失。不同 点体现在不同教学情境下的学习者表现出独特的特 点,需要针对性干预措施。比如基本特征维度下的 高风险学习者是问题解决能力、时间管理能力、计算 机水平和自我效能感比较低的学习者,对这类学习 者的教学干预主要从课程资源的难易度设计以及学 习进度的监控方面开展。行为特征维度下的高风险 学习者是20岁及以下的学习者群体,这群学习者年 轻、活泼,对新奇事物比较感兴趣,教师可通过监控 学习进度开展教学干预。学习路径分析是从学习 成绩和学习路径角度分析高风险学习者与学习环 境之间的交互特征,对这类学习者的教学干预措施 包括课程资源编排、个性化推荐设计以及资源重难 点设计。郑红苹等(2018)认为教学情境是基于数据 的个性化教学实施的重要支撑条件,强调教师需 要将教学经验与教学情境相结合,以促进个性化 教学实施。以往的学习者画像构建和应用局限于 单一情境,例如,顾小清等(2016)将学习者画像用 于学业评价, 肖君等(2017)主要以评价学习效果为 目的开展学习者画像模型的构建和分析。本研究 发现通过不同维度的高风险学习者画像构建,可适 应不同教学情境下精准化的教学服务设计和学习支 持服务设计。

综上,本研究对大数据环境下的在线学习者画像构建结果表明,高风险学习者画像可以反映该群体在学习能力、行为表现、学习路径上的综合特征及其教学情境性。本研究中的学习者画像构建偏重阶段性的分析和教学优化,但在线学习过程中,学习者学习心理状态、学习行为数据会不断变化,如何应对这种动态化学习仍面临很多挑战。中共中央、国务院印发的《中国教育现代化2035》指出加快信息化时代教育变革,需实现规模化教育与个性化培养的有机结合。学习者画像主要以群体为单位开展学习者特征分析,不同学习者群体之间具有差异化特征。学习者画像既能为个性化学习提供依据也有助于规模化教育的开展,借助学习者画像开展人才培养模式的变革将是未来的研究方向。

#### 参考文献:

[1]Apereo(2017). Apereo open analytics[EB/OL]. https://www.apereo.org/sites/default/files/projects/Brochures/ApereoAnalyticsBriefing26Apr16.pdf.

[2]Bautista, R. G.(2013). The reciprocal determinism of online scaffolding in sustaining a community of inquiry in physics [J]. Journal of Technology & Science Education, 3(2): 89–97.

[3]Britton, B. K., & Tesser, A.(1991)Effects of time-management practices on college grades[J]. Journal of Educational Psychology, 83(3): 405–410.

[4]蔡慧英(2016). 语义图示工具支持的协作问题解决学习的研究[D]. 上海: 华东师范大学博士学位论文.

[5]蔡慧英,乔惠,顾小清(2017). 协作概念图工具在协作学习情境中对个体认知发展的实证研究[J]. 现代远程教育研究, (6):93-102.

[6]蔡耀杰(2011). 论思想政治教育知行转化的困境与对策 [D]. 上海: 复旦大学硕士学位论文.

[7]Chatti, M. A., Muslim, A., & Schroeder, U.(2017). Toward an open learning analytics ecosystem[M]. Springer International Publishing: 195–219.

[8]陈海建,戴永辉,韩冬梅,冯彦杰,黄河笑(2017).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,23 (3):105-112.

[9]Durand, G., Belacel, N., & LaPlante, F.(2013). Graph theory based model for learning path recommendation[J]. Information Sciences, (251): 10–21.

[10]方旭(2015).MOOC学习行为影响因素研究[J]. 开放教育研究,21(3):46-54.

[11] Graesser, A., & McNamara, D.(2010). Self–regulated learning in learning environments with pedagogical agents that interact in natural language [J]. Educational Psychologist, 45(4): 234–244.

[12]顾静相,方慕真(2007). 远程教育学习支持服务体系构建探索[J]. 中国电化教育,(6):39-43.

[13]顾小清,舒杭(2016).信息技术的作用发生了吗:用学习分析技术刻画学习行为印记[J].现代远程教育研究,(5):10-19.

[14]Holley, D., & Oliver, M.(2010). Student engagement and blended learning: Portraits of risk[J]. Computers & Education, 54 (3): 693–700.

[15]Iglesias, J. A., Angelov, P., Ledezma, A., & Sanchis A. (2012). Creating evolving user behavior profiles automatically[J]. IEEE Transactions on Knowledge and Data Engineering, 24(5): 854–867.

[16]姜蔺,韩锡斌,程建钢(2013).MOOCs学习者特征及学习效果分析研究[J].中国电化教育,(11):54-59+65.

[17]JISC(2015). Who we are and what we do[EB/OL]. [2015-

12-28]. https://www.jisc.ac.uk/about/who-we-are-andwhat-we-do.

[18]刘艳华,徐鹏(2014). 大数据教育应用研究综述及其典型案例解析: 以美国普渡大学课程信号项目为例[J]. 软件导刊(教育技术),13(12):47-51.

[19]李晓兰, 张萍(2015). 等效交互理论相关研究与反思 [J]. 中国医学教育技术,(5):490-493.

[20]穆德(2007). 赢在用户: Web任务角色创建和应用实践指南[M]. 北京: 机械工业出版社.

[21]Paris, S. G., & Paris, A. H.(2001). Classroom applications of research on self-regulated learning[J]. Educational Psychologist, 36(2): 89–101.

[22]乔惠(2018). 基于 xAPI 的开放学习者画像的构建与应用研究[D]. 上海: 华东师范大学硕士学位论文.

[23]腾艳杨(2013). 社会临场感研究综述[J]. 现代教育技术,23(3):64-70.

[24]Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K., & Baker R. S. J. d.(2011). Open learning analytics: An integrated & modularized platform[EB/OL]. Open University Press.

[25]Thomas, E.(2007). SOA principles of service design[M]. Boston: Prentice Hall, (37): 71–75.

[26]田山俊,杨桂梅(2011).成人学习者学习特征及其教育价值:马尔科姆·诺尔斯成人教育思想解析[J].中国职业技术教育,(6):71-74

[27] Van Mulken, S., André, E., & Müller, J.(1998). The per-

sona effect: How substantial is it?[A]. In People and computers XIII(pp. 53–66)[C]. Springer, London.

[28]王楠, 乔爱玲(2009). 在线学习活动本质及理论基础探究[J]. 中国远程教育,(1): 36-40+78.

[29]王广新,宫淑红(2010). 网络学习情境中社会临场感研究回溯[J]. 开放教育研究,16(5);41-46.

[30]王小雪,刘菁,许涛,陈蕙若,刘炬红(2017).开放创新理念与"颠覆式技术"在教育教学中的实践与研究:AECT 2016年会评述与思考[J].远程教育杂志,35(1):3-16.

[31]王顺箐(2018). 以用户画像构建智慧阅读推荐系统[J]. 图书馆学研究, (4):92-96.

[32]肖君,姜冰倩,陈海建(2015).基于学习分析的体验式 开放在线教学环境研究:英国开放大学的实践与探索[J].中国 远程教育,(10):64-70.

[33]肖君, 乔惠, 李雪娇(2019). 基于 xAPI 的在线学习者画像的构建与实证研究[J]. 中国电化教育,(1): 123-129.

[34]肖君, 乔惠, 许贞, 徐颖(2017). 泛在学习环境下基于活动理论的移动 MOOC 设计及实证研究[J]. 中国电化教育,(11): 87-94.

[35]余孟杰(2014).产品研发中用户画像的数据模建:从具象到抽象[J].设计艺术研究,(6):60-64.

[36]曾建勋(2017). 精准服务需要用户画像[J]. 数字图书馆论坛,(12):1.

[37]郑红苹,崔友兴(2018)."互联网+教育"下循证教学的理念与路径[J].教育研究,39(8):101-107.

# Construction and Empirical Study of Online Learners' Persona in the Big Data Environment Xiao Jun Qiao Hui Li Xuejiao

Abstract: In the online learning environment, there are many challenges such as unclear characteristics of learners, inadequate learning support services, and high student attrition rate. Online learner persona in the context of big data is of great significance in identifying and analyzing the characteristics of learners and providing targeted learning support services. Based on open learning analytics framework, this study proposed the "ODAS" open learning analysis conceptual model from the four perspectives of Object, Data, Analysis and Service, and designed concrete and workable online learner persona construction process based on ODAS conceptual model. The process includes four stages: Persona goal, data collection, label analysis, and persona service output. It provides a method for learning application and evaluation of learner persona. This study took the online learners of Shanghai Open University as research participants, and constructed high-risk learner persona from the three dimensions of learners' basic characteristics, online learning behavior, and learning path environment. The study showed that learner persona could reflect the learning ability, behavioral performance, learning path characteristics of high-risk learners, which is contextual and provide data support for targeted learning design and online learning support service design, thus improving learners' participation and learning effect.

Key words: big education data; open learning analysis; online learner persona; precise support service