

数据驱动的高校图书馆用户画像构建研究^{*}

陈添源 吴锦辉 杨思洛

摘要 用户画像作为高校图书馆数字化智能化转型过程中实现精准服务的重要手段,近年来受到了普遍关注,但当前研究大多面向单一的业务类型采集用户历史行为数据建模,缺乏具有一定通用性、能切合高校图书馆多种服务场景、基于动态实时的全量数据构建用户画像的理论框架。本研究通过梳理高校图书馆用户画像的理论内涵,结合高校图书馆业务情境探究数据驱动的用户画像新特征,以具体服务场景下的业务流程和与之相关的用户行为路径为主线,借鉴 OSM 模型自上而下逐层构建适应不同业务场景的用户画像标签体系,提出适应不同数据规模且实时更新的用户画像系统框架,并从宏观、中观和微观层面阐述高校图书馆用户画像的应用实践。图 3。表 1。参考文献 47。

关键词 高校图书馆 用户画像 数据驱动 研究框架

Research on the Construction of University Library User Profile Based on Data-Driven

Chen Tianyuan Wu Jinhui Yang Siluo

Abstract: As an important means to achieve accurate services in the process of digital and intelligent transformation of university libraries, user profiling has attracted widespread attention in recent years. However, most of the current research focuses on the collection of user historical behavior data modeling for a single business type, and lacks a theoretical framework with certain generality, which can meet a variety of service scenarios of university libraries and is based on dynamic real-time full data. This paper combs the theoretical connotation of user profile in university libraries, and explores the new characteristics of user profile based on data driving in combination with the business situation of university libraries; then taking the business process and related user behavior path in specific service scenarios as the main line, using OSM model for reference, building a top-down user profile label system that adapts to different business scenarios and proposing a data processing framework that adapts different data scales and updates in real time. Finally, this paper expounds the application of user profile in university libraries from macro level, medium level and micro level. 3 figs. 1 tab. 47 refs.

Keywords: University Library; User Profile; Data-Driven; Framework

国家“十四五”规划纲要明确提出加快数字化发展、建设数字中国的战略思想,用户画像作为新兴数字产业激活数据要素赋能业务创新、实现精准化营销的重要应用之一,被广泛应用于多个行业领域。在数字化转型驱动服务转型和管理方式变革的新形势下,高校图书馆也迫切需要借助用户画像构建精准服务模式,通过数据赋能实现服务增值。而在数字图书馆、智能图书馆和智慧图书馆的实践发展过程中,表征用户行为特征的数据量日益增多,为构建用户画像提供了丰

富的标签数据来源。近年来,聚焦于知识发现与服务、阅读推广、学科服务和个性化推荐等主题,图书馆界开展了大量的用户画像研究与实践。在画像标签的构建上,有研究结合具体服务场景^[1]或目标受众需求^[2]直接预设标签维度,还有研究对用户人口统计属性进行统计分析或算法建模预测后再汇聚画像标签体系^[3],但都缺乏对标签体系构建理论基础的阐述,且大多面向较为单一的服务场景采集历史行为数据,甚少将之与数据驱动、动态实时和全量数据等关联^[4],可复制性

^{*} 本文系教育部人文社科基金项目“基于用户画像的高校图书馆精准服务模式构建及实证研究”(项目编号:20YJC870002)的研究成果之一。

不强,尚未形成契合高校图书馆多种服务场景的通用性系统框架,如何融合服务场景系统化构建和推动用户画像助力服务增值也未能得到有效开拓。在画像建模过程中,各种定量方法的应用层出不穷,但对于方法的适应场合却缺乏系统性梳理。有鉴于此,本文尝试对高校图书馆范畴的用户画像概念内涵、标签体系和研究范式开展基础性研究,构建具有普适性的标签体系框架与系统架构,以期突破以往基于历史行为数据的研究模式,提升用户画像分析、挖掘、建模、预测和提炼的广度,从而辅助业务决策和提升精准服务质量,激活数据要素推动用户画像赋能高校图书馆服务。

1 高校图书馆用户画像理论内涵

1.1 用户画像研究溯源

早期研究中的用户画像主要称为“用户角色”,最初是基于调研问卷数据将用户群体划分为不同类型,是一种描述用户特征、直观展示用户核心诉求的定性研究方法,被称为“Persona”。随着自动化数据采集技术和数据科学工具的快速发展,数据驱动的用户画像研究成为可能。Williams^[5]在2006年提出动态收集用户的服务需求、心理偏好、行为特征等多维数据组成画像标签体系,继而借助数据科学工具理解和预测用户需求差异。这是一种实现用户聚类 and 精准化服务的定量研究方法,被称作“User Profile”^[6],广泛运用在软件开发^[7]、社交媒体^[8,9]等领域。

用户画像理论极大契合了图书馆聚焦用户需求的个性化服务理念。纵观既有的研究综述及述评^[10-12],在国外,典型的应用有获取机构知识库的用户需求^[13]、科学数据管理的用户群体聚类^[14]及图书馆用户体验提升^[15]等;在国内,最早的研究始于2014年^[16],涉及图书馆个性化服务^[17]、移动阅读^[18]、智慧图书馆^[19]等众多主题。

标签数据是用户画像的基石。不同于公共图书馆的用户范围广流动性大、用户数据较为分散,高校图书馆的用户群体基本稳定,便于对特定用户群体的行为进行长期跟踪。对单一用户而言,高校图书馆用户在文献、空间、数字资源利用等方面的使用频率远高于公共图书馆用户,在不同服务场景均能形成大规模高密度的用户数据;对群体用户而言,高校图书馆用户类型明确,在学科、年级、师生类别等方面具有明显的分群差异,极有利于构建用户画像标签维度。因此,高校图书馆领域的相关研究成果更为丰富,在空间服务^[20]、阅读推广^[1]、年度阅读报告^[21]、学科服务^[22]、参考咨询^[23]、知识服务^[24]等业务场景中,用户画像都得到了广泛的应用实践。

融合业务场景将异构多源的用户行为数据汇聚起来刻画出一个来自于真实数据的虚拟用户全貌,比较好地消除了用户错综复杂的信息行为及潜在的心理状态与产品或服务之间的信息鸿沟,并能按照画像描述的特征分群开展精准营销。此种研究范式逐渐成为当前高校图书馆探索业务转型创新和实现用户增长的新洞察模式。

1.2 用户画像的概念内涵

用户画像需要借助多维度的特征聚类来组织与虚拟人物有关的真实数据,通过数据的不断交集叠加产生行为标签,呈现用户的行为特征和使用偏好,以便更好地为目标用户提供信息或服务。用户画像的核心在于“打标签”,而标签来源于数据抽取,所需数据处于高层次的用户概念化和分析级的用户粒度表示之间。因此,数据的获取维度和处理方式决定了用户画像的有效性和精准程度。通过对近几年高校图书馆领域用户画像研究相关文献的梳理发现,业界对不同业务情境下用户画像构成要素有着各自不同的理解,典型举例如表1所示。

表1 高校图书馆领域的用户画像构成要素及研究范式

构成要素	研究方式	业务范畴	提出者
基本信息标签、行为偏好标签、互动标签、情景标签、活跃度标签	改进的 RFM 模型	图书借阅活跃度用户细分	乐承毅、王曦 ^[25]
基本属性、阅读状态、学习风格、阅读偏好	聚类分析、关联分析、相似度分析、特征匹配分析	个性化资源推荐	李宝 ^[26]
访问时间偏好、用户操作偏好、访问模式偏好、检索偏好	Session 访问识别、序列模式挖掘算法	电子资源访问	肖敖夏等 ^[27]
信息行为标签、研究兴趣标签	相似度检测、关联推荐	学术资源推荐	王仁武、张文慧 ^[28]
自然维度、偏好维度、情境维度、社交维度	情境识别与推理	参考咨询	程秀峰等 ^[23]
用户维度(基本属性、行文特征、兴趣特征)、图书维度	本体建模法、聚类与关联算法	阅读推广	刘漫 ^[1]
用户基本信息、用户行为信息、用户需求信息	卡方分析	空间服务	龙泉 ^[29]
基本信息、兴趣偏好、行为特征、合作关系、综合评估	文献计量法、主题模型及主题发现法、社会网络分析法	学科服务	刘邦国 ^[22]
身份属性、需求领域、需求偏好、需求层次	可视化统计工具、聚类分析	知识服务	卢思佳、王凤姣 ^[24]
基本信息、行为信息、内容信息、互动信息、情景信息	亲和图法、本体论	智慧图书馆	孙守强 ^[19]

由表1可以看出,用户画像研究已经遍及高校图书馆的众多服务场景,尽管从不同业务视角出发,对用户画像的数据要素及特征汇聚维度的分析方法存在差异,但均是以服务场景为驱动实现用户画像特征的标签化,借助算法科学工具构建标签体系,实现对不同业务的智能决策支撑。所谓标签化,就是“给用户打标签”,无论结构化还是非结构化的数据,均是将真实业务数据抽象为画像标签,使其易观察和可量化;所谓支撑业务决策,即用户画像不再停留于简单的数字化和可视化,而是将它与具体的产品或者服务高度关联,为定位目标用户群体、开展精准推荐、提升用户转化和实现用户增长等提供数据支撑。

综上所述,高校图书馆范畴的用户画像是基于数据驱动在高校图书馆服务场景中构建用户基本属性、行为特征和兴趣偏好等标签指标体系,借助数据科学工具展现用户个体或群体特征,为业务决策和精准服务提供有客观依据的动

态标签化用户模型。它是一类聚焦目标受众基础属性、偏好、行为、目标、观点等相关指标的用户研究方法。

1.3 用户画像的隐私保护

从前述概念内涵界定来看,用户画像是在一系列真实数据之上抽象概括的虚拟用户代表,由于其数据来源于用户个人信息,在构建和使用过程中不可避免地涉及到用户隐私保护问题。按照《网络安全法》和《信息安全技术个人信息安全规范》的要求,经授权同意和去标识化处理方可共享和转让用户数据。故此,在获取构建用户画像所需的相关数据的过程中,需要遵循合法收集、最小化收集、明示同意原则等相关法律规范,即在数据采集环节应遵循政策法规要求向用户告知采集范围和用途,获得用户同意;在数据存储、处理、挖掘等环节应广泛运用各种隐私保护技术,进行数据脱敏处理,也可以结合使用隐匿性标签等^[30];在用户画像可视化展示环节,应对单用户

画像进行匿名处理或仅展示群体画像;在服务推送环节,经用户授权许可,可以与用户的人口统计属性、常用终端信息、使用行为等其他标签信息结合识别,将资源或服务信息精准推送给用户;同时也要加强数据馆员的用户信息安全保护意识和责任意识,确保用户隐私信息的内部安全。

1.4 数据驱动的高校图书馆用户画像特征

伴随着高校图书馆的各类自动化业务系统和资源服务平台朝着数据耦合、业务联动及高度协同方向快速发展,每日产生的业务数据、资源

描述数据、服务交互数据、用户行为数据、内生数据等急剧增长。在多源异构数据整合和数据科学工具的赋能下,数据驱动的用户画像研究不仅能复现具备可测量性、可重复性的用户习惯与兴趣爱好,也能利用各种情境感知技术、自然语言处理技术等识别用户情绪、情感倾向、人格特征等心理和认知状态^[31]。有鉴于此,本文提出从服务场景层、标签数据层和业务决策层三个层面以及“数据化+服务场景分析”“服务场景分析+算法”两个阶段共同描述数据驱动的高校图书馆用户画像特征框架,如图1所示。

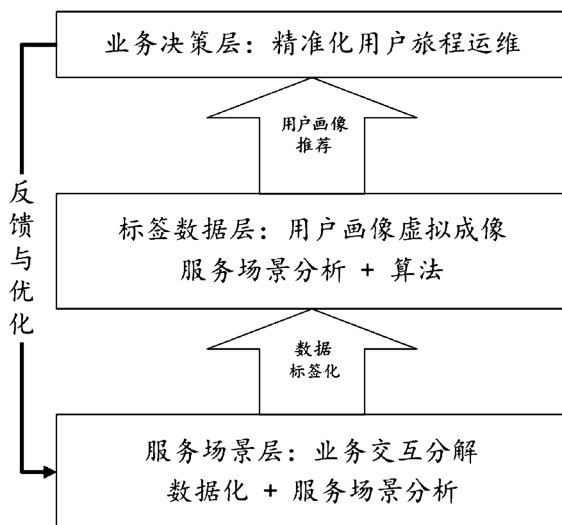


图1 数据驱动的高校图书馆用户画像特征框架

1.4.1 服务场景层面

数据驱动的用户画像构建面向服务场景实现业务赋能,聚焦用户触达的关键业务要素逐层汇聚画像标签,深度融合服务场景是其核心特征。贯穿于用户旅程全过程,高校图书馆构建用户画像的终极目标是实现精准化营销和个性化推荐,这就决定了其构建过程必然要以高校图书馆的资源服务、空间服务、阅读推广、参考咨询和学科服务等不同服务场景的用户决策流程与核心要素为主线,既要反映服务场景的用户与要素之间产生的可测量指标,也要对特定业务状态的发生与转化环节进行过程指标量化。

因此,这是一个服务场景要素的数据化过程,也是构建用户画像实施“数据化+服务场景分析”的阶段。通过对服务场景业务流程的还原,不断细化用户画像数据粒度,充分识别关键价值节点,使服务场景要素的“像素”显著提升,从而勾勒出在业务目标决策下服务场景各要素的发展轨迹和演变方向,实现业务精准赋能,加快用户转化。

1.4.2 标签数据层面

数据科学工具的广泛运用极大提升了构建用户画像的精准程度,自然语言处理对定性文本数据的量化表达,使用户画像从使用定性数据向使用定量数据发展,继而加速向定性 with 定量分析

相结合的转变^[32]。在画像标签数据的采集与汇聚过程中,服务场景和业务策略共同确立用户画像所需的数据粒度,标签来源既包括人口统计属性的静态数据,也涵盖用户与业务触点交互产生的动态数据。依托数据治理和连通技术,采集的标签数据既能体现用户业务旅程特征数据的路径连接,也支持根据决策目标对标签数据的抽取、切分、上卷、下钻、回溯等分析处理。借助统计思维开展建模预测与智能学习,不断刻画和迭代出用户标签、行为聚合标签和场景标签等,从而展现用户各类特征的动态演化过程。

在根植于服务场景勾勒高像素化用户画像的过程中,成像算法也成为画像产生和推荐应用的支撑重点,更是转型期高校图书馆面对新业态、新模式、新用户群体的发展策略和竞争优势。实践中,早期为基于调研数据展开统计分析,主要运用聚类或者分类方法实施用户细分,目前则较多运用神经网络、语义分析、情感分析和时序分析等算法绘制用户画像。用户画像的绘制是一个结合服务场景进行算法分析的过程,也是实施“服务场景分析+算法”的阶段,画像标签提取与重组、标签权重分配、特征因子降维、群体聚类分群、画像可视化描述及情境感知推荐等,都离不开各种数据科学工具的支持。高校图书馆既要根据不同的服务场景创建适宜的数据分析算法,也要根据获取的规则标签或预测标签创新服务场景的营销策略,持续深化和扩展用户画像辅助业务决策的功能。

1.4.3 业务决策层面

数据驱动的用户画像以面向业务决策为终极目标,画像构建过程就是对画像标签做信息转化和增益,并通过算法预测业务场景调整营销策略。它能够发现关联服务场景下用户特征数据构成要素之间的因果关系,厘清服务场景的业务逻辑和用户生命周期的状态转换机理。数据驱动构建的用户画像实现了全场景、全触点、全旅程的成像覆盖与测量,快速洞察用户需求和预测用户行为,有效延伸用户业务触点的同时也反馈与优化用户体验,提升服务的精准化,让业务战

略更具前瞻性,对用户流失的风险洞见更加敏锐,最终使用户画像能够针对用户拉新、用户促活、价值提升和用户留存等转化环节进行情境映现、趋势预判和风控评估。

2 高校图书馆用户画像标签体系构建

2.1 标签体系构建思想

画像标签是用户特征的符号表示,是对用户数据进行过滤、挖掘、聚合等处理后的抽象概括与提炼,也是用户画像发挥其实际应用价值的基础,其体系构建是伴随着用户与服务场景的行为交互而产生的。数据驱动的用户画像标签体系不仅要能精准契合服务场景下的业务目标定位,更要聚焦业务目标与用户行为相融合的关键节点,定量且实时呈现用户行为数据。具体到高校图书馆业务实践中,面向图书借还、空间服务、参考咨询、学科服务等不同的服务场景,需要采集的画像数据和构建的标签体系存在明显差异。与此同时,用户画像标签数量并非越多越好,只有提炼出体现业务价值、能影响用户关键决策的数据要素,才能精准反映用户需求。然而,虽然既有研究普遍认为标签体系框架对用户画像至关重要,但大部分都未明确阐述其标签体系构建的理论基础,这样的标签体系固然能勾勒特定研究对象的画像特征,但不具备移植性。在高校图书馆的业务实践中,由于信息化建设和业务系统成熟度的差异,需要根据业务数据的获取难易和数据治理情况,选择具备构建用户画像条件的服务场景灵活构建用户画像。因此,一个能面向高校图书馆每个业务场景且对不同数据规模具有普适性的用户画像标签体系构建框架显得尤为重要。

构建用户画像的根本目的是激活数据要素提升精准服务,因此,用户画像标签体系的构建必然与具体场景的每一个服务环节和用户行为匹配,所选取的标签数据既要能反映用户属性和行为特征,也要能从业务流程与用户的接触点洞察业务痛点和用户转化节点,这样才能有助于达成该服务场景不同阶段的业务目标。而对业务流程

中的用户行为进行逐层分解,寻找对应的业务目标和可拆解的度量指标,进而逐层转化标签构建用户画像,则能较好地实现此目的。故此,本文以高校图书馆具体服务场景的业务流程和与之相关的用户行为路径为主线,聚焦用户关键转化节点,借鉴 OSM 模型^[33]从战略目标 (Objective)、营销策略 (Strategy) 和业务度量 (Measurement) 等环节自上而下逐层解构和细化度量指标;在按需确立画像标签数据的相关域后,根据画像标签的维度组织方式,在用户洞察中实现用户行为数据的归纳、提炼和萃取,最终汇聚形成画像标签。

2.2 标签维度组织

在确立用户画像标签体系的组织思想之后,如何将所有度量指标组织起来形成逻辑分明的体系取决于画像标签的组织维度。从表 1 对高校图书馆已有用户画像构成要素的举例来看,虽然这些用户画像根据各自的使用场景对具体的标签维度表述方式不一,但基本都使用了用户基础属性、行为特征、兴趣偏好等标签类型,少数包含用户心理特征、用户活跃度等标签,部分涉及到业务应用的还有用户行为预测标签^[34]。而用户画像在其他领域的应用,也有按照事实类、规则类、机器学习挖掘类等进行标签组织的^[35]。

在参考上述标签组织维度的基础上,本文认为,数据驱动的用户画像以服务场景和用户业务触点为依据拆解和抽取了与业务强关联的标签数据,这些标签数据将被应用于高校图书馆领域的战略目标决策辅助、情景感知推荐和精准化营销服务等。因此,用户画像标签的组织维度既要包含直接体现用户属性的标签类型,也要有依据业务决策需要对用户行为数据进行统计分析、数据挖掘后形成的标签类型,为了使标签维度对高校图书馆更具针对性,本文倾向于选取事实类、行为类、规则类、预测类等四种维度的标签,组成高校图书馆领域的用户画像标签指标体系。其中,事实类标签泛指高校图书馆用户的人口统计属性以及学科专业、身份类型等基础属性指标;行为类标签指服务使用时长、服务使用次数、

服务使用数量、近期活跃度等用户在高校图书馆各个服务场景的行为数据指标;规则类标签指根据用户属性、用户行为在确定的规则下通过数据挖掘得出的用户流失率、使用偏好、用户生命周期等与服务场景对应的具有规律性的用户特征指标;预测类标签是指基于现有事实类及规则类标签无法得出的画像指标,需要运用集成算法、机器学习、自然语言处理等进行建模并实施数据挖掘与训练,从而得出可以对用户后续行为做出预测的标签,此类标签开发周期长且成本高,所占比例较少。

2.3 标签体系构建框架

首先明晰需建立用户画像的服务场景,对该场景下的用户行为路径及所处的生命周期进行还原拆解,然后确立这一具体服务场景的可度量业务战略目标;根据此目标在每个业务场景触点中制定营销策略,也即形成相匹配的二级衡量指标;最后针对每一项营销策略确立可采取的业务度量指标,并将其与用户画像标签维度进行对应,自上而下逐层汇聚服务场景中的用户画像数据。

标签体系的构建离不开具体的服务场景,为了更直观地阐述基于上述构建思想的用户画像标签体系构建过程,本文以高校图书馆的空间服务为例,构建如图 2 所示的空间服务用户画像标签体系框架。

标签体系构建初期,本文以 CJM (Customer Journey Map, 用户旅程地图) 模型^[36]为依托拆解高校图书馆空间服务过程中用户所产生的接触点,从这些关键点识别用户转化节点。用户旅程地图是客户体验全程映射的一种常用方法,结合它绘制用户生命周期阶段和路径,可以洞察和掌握每个用户在每个阶段的独特心态、需求和关注点。在拉新阶段,聚焦用户的账号绑定、空间预约和签到服务等流程的转化,提升用户体验的便捷性和易用性;在促活阶段,多途径保障用户在空间服务过程中的履约状态切换,提升履约完成率,从而形成空间服务依赖,加速用户忠诚度的转化;在

留存阶段,通过优化业务功能和开发新功能模块等提高用户满意度,例如减少预约等候时长、入

馆自动激活预约及常用空间快速预约等,对沉没用户则可以定向推送服务攻略减少用户流失。



图2 高校图书馆空间服务的用户画像标签体系框架

在基于CJM模型拆解和还原业务流程之后,接下来可以在OSM模型中逐层构建用户画像标签数据。业务目标方面,高校图书馆空间服务的战略目标(O)是提升整体的空间使用效益,根据服务场景、用户关键需求和服务绩效指标,可将该指标分解为用户使用流量、转化率、使用时长和服务复用率。据此在营销策略(S)方面围绕上述关键指标进一步拆解和扩展相适宜的营销策略目标,例如提高预约使用率、履约完成率、用户活跃率和用户留存率等。在业务度量(M)方面,在对营销策略目标充分理解的基础上汇聚相应场景中的用户行为数据依次组织用户画像度量指标。例如图2中的策略3为“提高用户活跃率”,可以以使用时长、自动激活率和使用频率等

组织用户画像标签。在对所有营销策略完成业务度量的画像指标组合后,即可在用户画像层按前述的维度组织归类所有标签。由此,通过复盘服务场景和拆解用户旅程地图,从业务定位和营销策略框架中自上而下分解核心指标相对应的用户画像标签。

3 高校图书馆用户画像系统架构

3.1 系统架构设计

在前述以服务场景为依托自上而下拆解和构建用户画像标签体系之后,结合高校图书馆所处的信息化建设环境、异构业务系统的数据协同模式和交互平台的用户行为数据特征,本文借鉴

数据仓库技术和目前已有高校图书馆用户画像标签数据获取的系统架构模式^[37-39],提出图3所示的高校图书馆用户画像系统架构,以服务场景的业务实体要素为核心实现相关域的数据链接(即数据源层)、标签汇聚(即数据汇聚层)及画像标签萃取(即标签萃取层),最终形成各个维度的单用户画像和群体画像(即画像主题层),用于支撑高校图书馆战略规划和业务决策、个性化推荐及面向不同服务场景和用户生命周期不同阶段的精准化营销。

用户在服务场景的动态转化必然要求数据驱动的用户画像能够实时更新数据。在图3所

示的系统架构中,小样本数据的用户画像可通过BI平台连接高校图书馆业务系统或者规范化的数据模板导入后实时更新。大数据规模则通过ETL管理平台实现标签数据抽取的任务管理、调度和监控,例如采用Airflow统一调度Spark、Hive、Hadoop等计算引擎对各类标签数据在对应的更新周期分区中插入日增量或者截止到当日为止用户在高校图书馆服务场景的全量数据。而对于通过网络调研、面对面访谈和网上第三方权威数据等方式收集的定性数据,根据服务场景的发展阶段确立更新间隔周期,并转换为心理量表数据。



图3 高校图书馆用户画像系统架构

3.2 系统架构的各层功能

在数据源层,根据数据规模大小按需选择BI平台或者Hive离线与Spark实时处理组合等模式采集、汇聚来自关联域不同数据源层的异构多源数据。例如以空间服务的用户行为数据为核心,集成用户人口属性数据、与空间服务关联的入馆

行为数据、馆藏资源借还数据和自助服务行为数据、用户外部行为数据等,同时包含因定性分析需要调研获取的用户心理偏好数据。本层属于数据准备区,经过API接口读取或Sqoop数据抽取、数据清洗、Flink同步等ETL调度和定期更新上述日志数据、前端埋点数据和业务行为数据后,为数据

汇聚层做好数据源基础。

在数据汇聚层,小样本数据集可通过BI平台进行不同逻辑表的关联链接操作形成数据宽表。大规模数据集则采用ID-Mapping技术、GraphX图算法对ODS层的多种异构数据实现要素识别和全域数据的融合打通,消除数据孤岛、避免数据歧义。在遵守合理使用用户数据的法律规范框架内,高校图书馆可以采用身份证号、学工号、手机号、Cookie等ID识别实体,通过一系列Oneid数据清洗流程最终形成一张统一用户标识的相关域数据宽表,为标签萃取层的用户画像标签体系提供高质量的基础数据。这也是用户画像的“拼图”过程。

在标签萃取层,根据图2的画像标签体系构建方法,以高校图书馆空间服务为例,在事实类主题标签方面,借助BI平台或者数据仓库的聚合算法工具从Oneid数据宽表中获取涵盖性别、年级、学科专业、身份类型、成长积分、注册渠道等的高校图书馆用户基础信息数据;在行为类主题标签方面,可使用统计算法获取用户的履约次数、使用时长、登录次数、违约次数、服务复用间隔周期均值和常用空间使用次数等;在规则类主题标签方面,结合业务实际和统计时间窗口,可以以RFM模型计算用户价值等级,数据分析人员和管理人员制定规则和计算用户生命周期层级、流失风险等级和活跃度等;在预测类主题标签方面,可根据服务场景需求通过机器学习算法挖掘产生,例如判断用户使用偏好、心理偏好和流失意向等。此类标签与数据集规模有关,小数据产生预测类标签的应用场景较少。

在画像主题层,主要有三大类型的用户画像。可以根据应用场景需要,在标签萃取层的基础之上建立面向不同宏观层面分析和决策的主题宽表,提供不同画像主题的聚合标签数据支持,例如从学科专业、年级、身份类型、时段偏好和资源偏好等切面进行群体画像特征聚合分析;在获取用户授权许可且合理使用用户数据的前提下可建立单用户画像,例如输入用户OneID返回其特定业务场景的全息画像,然后根据标签开展

精准化的馆藏资源推送、业务服务消息推送以及实时画像分析;根据群体画像通过特征工程提取关键画像标签,确立标签权重计算方法,进而采用聚类算法实现用户细分生成差异群体的服务需求特征,精细化辅助学科专业建设的资源保障和学科服务。

4 高校图书馆用户画像的应用实践

4.1 宏观层面:战略制定及业务定位

基于用户数据的深入挖掘和分析而建立的不同主题画像,使高校图书馆能明晰当前业务痛点和不同群体的用户需求,合理制定战略目标,为高校图书馆用户分群识别和定位、创新服务方式、提升服务效能等提供决策支持。

在资源配置方面,可以通过学科分类、资源分级与用户分类分级标签的匹配情况完善馆藏资源结构,兼顾学科发展需要和用户个性化需求^[40];在阅读推广方面,可以基于用户画像模型挖掘教师、科研人员和不同阶段大学生群体的阅读需求、阅读动机与活动类型偏好,制定差异化阅读推广方案^[1];在学科服务方面,学科服务对象的学者画像、团队画像和学科画像,可以成为高校科研团队评价、学科建设评估和图书馆制定学科服务政策、深化学科服务领域的重要依据^[22];在智慧图书馆建设方面,将用户特征扎根于产品设计中,有利于智慧图书馆产品优化和服务创新^[41]。

数据驱动模式的完整、多维和精准化,可以有效提升高校图书馆决策的科学性和预见性。面向不同服务场景建立的用户画像标签体系,在动态展示特定情境下用户全貌的同时,也能使高校图书馆更清晰地认识当前业务所处的层次、状态、优势与薄弱环节,推动图书馆业务转型升级和服务创新。

4.2 中观层面:服务过程优化

通过数据科学工具汇聚和配置的用户标签,可以支持高校图书馆在服务过程中更为便捷地实现潜在用户发现、服务场景精准触达和渠道优化等。特别是基于用户行为路径逐层分解形成的

用户画像标签体系,能快速定位业务流程中存在问题的用户触点,对于指导业务系统改进或服务过程优化具有天然的优势。资源优化方面,可以通过用户外借、续借、还回、预约等环节的指标,挖掘热门外借图书、热门预约图书、不同流通周期图书等标签构成的用户资源使用情况画像,据之调整相关图书的采购比例,如根据受欢迎程度调整相关作者、出版社的图书圈选率,根据用户预约量、流通周期长短调整相关图书的复本量等。移动图书馆系统、资源检索与发现系统、空间预约系统等业务系统的优化方面,可以通过用户菜单点击、页面浏览、页面检索、系统智能推荐内容点击等环节的指标,结合用户心理量表数据、调研访问数据等,形成系统模块使用率、用户检索行为习惯、页面停留时长等标签画像,据之调整系统页面结构、改进系统功能,优化系统访问路径,实现与用户在资源和服务上的高效多维交互。由于系统设计的复杂性,目前相关实践案例较少,但也不乏先行者,如赵杨等^[42]通过构建基于用户画像的移动图书馆用户体验地图,发现业务流程中的用户痛点和服务机会点,进而提出服务优化策略。

除了对既有业务系统和服务过程的优化外,移动信息网络的普及和移动阅读技术的飞速发展,也拓展了高校图书馆的业务范围和服务方式。为了更好地嵌入用户环境,提高服务效能,基于用户画像进行新的业务系统设计或重构也逐渐受到重视。陈月娟^[43]引入用户画像技术设计了高校图书馆数字阅读平台框架;闫舟舟等^[44]基于用户画像技术主动预测用户咨询需求,重构了媒介融合环境下的高校图书馆参考咨询服务流程;杨群等^[38]结合大学生用户画像特征数据和服务需求构建了高校图书馆微知识服务体系。

4.3 微观层面:精准化情景感知推荐

数据驱动的用户画像为高校图书馆长期以来倡导的个性化、精准化服务提供了有力的数据支撑。通过对用户行为数据的动态追踪,高校图书馆可以实现对用户需求的实时感知和用

户画像标签的迭代更新,从而精准匹配用户的最新需求和关注点,实时调整个性化服务内容,实现智能推送、关联查询和个性化推荐等。如基于借还和检索行为的用户画像,个性化推送符合用户当前阅读偏好的图书文献,或在用户检索和数字阅读过程中动态展示关联资源^[45];基于用户数据库使用习惯、学科属性和兴趣爱好,针对性地推荐信息素养培训或学术讲座信息^[46];根据科研人员个人或团队画像,推送符合用户学科及其当前研究方向的最新科研成果、学科领域核心作者与机构的研究动态以及学科发展态势分析等^[47]。当前,基于用户画像的差异化内容展示和推送已经在各种社会化软件平台、购物平台、资讯平台等得到了广泛应用,与之相比,高校图书馆的用户对象范围更加明确,用户基本属性信息更加完善,在利用用户画像感知用户需求、实现精准推荐、提升用户体验方面还有很大的发展空间。

用户画像对用户多元化、个性化需求的精准识别,也为提高用户黏性、降低用户流失率提供了精准服务的方向,高校图书馆可以通过用户流失风险等级等规则类标签发现目标用户群体,继而根据群体内不同用户个体的使用偏好精准把握推送时间和推送内容,实现对用户的有效召回。以前文的空间服务为例,对于服务复用周期较长的用户,可以根据空间服务系统的动态使用情况,在其常用的入馆时段实时推荐符合其使用偏好的空余座位,吸引用户持续使用。

5 研究局限及展望

伴随着智慧校园、智慧图书馆和大数据技术的稳步推进和拓展应用,高校图书馆亟需具有普遍参考价值、能适应不同服务场景、不同业务系统和数据规模的用户画像构建方法和系统框架,本研究利用 OSM 模型和用户行为路径解析具体服务场景中各业务流程的关键节点,构建数据驱动的用户画像标签体系,契合高校图书馆应需而变的精准服务需求;依托数据仓库 ID-Mapping 技术和数据科学工具汇聚和产生的各类动态实时的

用户画像主题,不受系统数量和规模的限制,能有效适应不同数据治理水平的高校图书馆的需求,加快高校图书馆的数字化转型发展。但这些仍具有一定的局限性:一是本文提出的用户画像标签构建方法及系统架构,未来还需要通过高校图书馆的业务实践验证模型的普适性,更需要汇聚大量的用户行为数据增强用户画像在支撑精准服务上的可用性;二是数据驱动的定量分析与定性分析相结合的用户画像构建模式,需要通过更多的服务场景触达、持续的行为数据积累、画像标签迭代修正和反复的数据验证等检验其合理性和有效性。

参考文献

- 刘漫.基于用户画像的高校图书馆阅读推广模式构建[J].图书馆理论与实践,2019(7):21-26.
- 过仕明.数字图书馆用户画像及场景重构研究[J].情报科学,2019(12):11-18.
- 陈丹,等.基于大数据挖掘和用户画像的高校图书馆个性化服务研究[J].图书馆研究与工作,2019(4):50-53,59.
- 吴文瀚.搜索引擎全量数据的用户画像模型研究——设计与实证[J].图书情报工作,2022(4):129-141.
- Williams K L. Personas in the Design Process: A tool for understanding others[J]. Georgia Institute of Technology, 2006(8):69-78.
- 赵雅慧,等.大数据背景下的用户画像研究综述:知识体系与研究展望[J].图书馆学研究,2019(24):13-24.
- Jantunen S, et al. Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering[C]//15th IEEE International Requirements Engineering Conference, RE 2007, October 15-19th, 2007. New York: IEEE Computer Society, 2007:185-194.
- Dang-Pham D, et al. Demystifying online personas of Vietnamese young adults on Facebook: A Q-methodology approach[J]. Australasian Journal of Information Systems, 2015, 19:1-22.
- 万震,等.基于ISM-MICMAC的移动社交媒体倦怠影响因素与关联路径研究[J].信息资源管理学报,2022(1):46-55.
- 宋美琦,等.用户画像研究述评[J].情报科学,2019(4):171-177.
- 阳广元,白美程.国内图书馆领域用户画像研究综述[J].图书馆理论与实践,2021(3):95-101.
- 徐芳,应洁茹.国内外用户画像研究综述[J].图书馆学研究,2020(12):7-16.
- Maness J M, et al. Using Personas to Understand the Needs and Goals of Institutional Repository Users[EB/OL][2022-09-18]. <https://www.dlib.org/dlib/september08/maness/09maness.html>.
- Lage, et al. Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder[J]. Portal: Libraries and the Academy, 2011(4):915-937.
- Brigham, Tara J. Personas: Stepping into the Shoes of the Library User[J]. Med Ref Serv Q, 2013, 32(4):443-450.
- 李业根.基于大数据的图书馆信息营销策略[J].图书馆学刊,2014(10):7-9.
- 何胜,等.基于用户行为建模和大数据挖掘的图书馆个性化服务研究[J].图书情报工作,2017(1):40-46.
- 冯龄莹,魏群义.高校移动图书馆流失用户画像模型构建[J].国家图书馆学刊,2021(1):14-24.
- 孙守强.基于用户画像的智慧图书馆个性化服务研究[J].图书馆工作与研究,2019(7):60-65.
- 尹相权,等.基于系统日志的高校图书馆研究间用户利用行为分析[J].现代情报,2018(1):115-120.
- 都蓝.基于用户画像的高校图书馆年度阅读报告研究[J].图书馆杂志,2019(4):27-33,40.
- 刘邦国.面向学科服务的用户画像研究[D].镇江:江苏大学,2020.

- 23 程秀峰,等.基于用户画像的图书馆智慧参考咨询服务模式研究[J].图书馆学研究,2021(2):86-93,101.
- 24 卢思佳,王凤蛟.基于画像分析的高校图书馆精准知识服务研究[J].大学图书馆学报,2021(5):55-62,86.
- 25 乐承毅,王曦.基于改进RFM聚类的高校图书馆用户画像研究[J].图书馆理论与实践,2020(2):75-79,93.
- 26 李宝.基于用户画像的高校图书馆个性化资源推荐服务设计[J].新世纪图书馆,2021(4):68-75.
- 27 肖教夏,等.基于Session识别的高校图书馆电子资源用户访问行为画像分析[J].图书馆杂志,2022(1):98-105.
- 28 王仁武,张文慧.学术用户画像的行为与兴趣标签构建与应用[J].现代情报,2019(9):54-63.
- 29 龙泉.高校图书馆实体空间用户画像构建实证研究[J].图书馆学研究,2021(10):70-80.
- 30 周林兴,徐承来.用户画像视域下的图书馆用户隐私问题研究[J].图书馆学研究,2020(3):26-34.
- 31 徐恩,等.基于移动感知数据的用户画像系统[J].郑州大学学报(理学版),2019(4):30-36.
- 32 Mijac T, et al. The potential and issues in data-driven development of web personas[C]//2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). New York: IEEE, 2018:1237-1242.
- 33 郑银巧,施灿涛.基于OSM模型和金字塔原理构建钢铁企业数据驱动指标体系研究[J].冶金经济与管理,2022(3):38-41.
- 34 张晗,等.基于用户画像的数字图书馆精准推荐服务体系构建研究[J].情报理论与实践,2019(11):69-74,51.
- 35 赵宏田.用户画像:方法论与工程化解决方案[M].北京:机械工业出版社,2019:23-25.
- 36 谢克生,王韞.服务设计中的接触点:定义、研究现状与趋势[J].包装工程,2022(8):137-148.
- 37 王庆,赵发珍.基于“用户画像”的图书馆资源推荐模式设计与分析[J].现代情报,201838(3):105-109,137.
- 38 杨群,等.面向用户画像的大学图书馆微知识服务体系构建分析[J].图书馆,2020(3):81-87.
- 39 陈丹,等.基于用户画像的图书馆个性化智慧服务模型框架构建[J].图书馆工作与研究,2019(6):72-78.
- 40 朱会华.基于用户画像的高校图书馆学科采访模式研究[J].图书馆学研究,2020(19):36-40,49.
- 41 廖运平,等.大数据视域下智慧图书馆用户画像研究[J].国家图书馆学报,2020(3):73-82.
- 42 赵杨,班蛟蛟.移动图书馆用户体验地图构建与服务优化研究[J].图书情报工作,2021(24):61-70.
- 43 陈月娟.大数据时代高校图书馆数字阅读平台研究与实践[J].计算机工程与科学,2019(S1):245-248.
- 44 闫舟舟,詹庆东.媒介融合视角下高校图书馆参考咨询服务流程再造研究[J].图书情报工作,2021(3):61-66.
- 45 王茹芳,宁璐.基于用户画像的图书馆推荐系统研究[J].图书馆建设,2020(S1):100-102.
- 46 杨传斌,楼应凡.基于用户画像的讲座信息精准推送服务研究[J].数字图书馆论坛,2021(10):60-65.
- 47 谢靖,等.科研学术大数据的精准服务架构设计[J].数据分析与知识发现,2019(1):63-71.
- (陈添源 副研究馆员 闽南师范大学图书馆, 吴锦辉 副研究馆员 闽南师范大学图书馆, 杨思洛 教授 武汉大学信息管理学院)

投稿日期:2022-09-29