

【技术·平台】

基于多源数据的高校画像构建与应用场景研究*

●任福兵 王朋 (华东理工大学 上海 200237)

【摘要】文章在借鉴已有研究成果的基础上,探讨高校画像的内涵和基本特征,分析高校画像的构建流程,并阐述画像应用场景的具体实践。研究发现:高校画像具有很强的实用性,可以应用于高校、院系、师生和社会等场景,能更好地服务于高校智库校园建设。

【关键词】高校画像 内涵 基本特征 构建流程 应用场景

【中图法分类号】G252

【文献标识码】A

【文章编号】1003-7845(2022)02-0034-07

【引用本文格式】任福兵,王朋.基于多源数据的高校画像构建与应用场景研究[J].高校图书馆工作,2022(2):34-40

随着信息技术的快速发展,国家对高校建设提出了更高要求。2016年2月2日,教育部办公厅发布《教育部办公厅关于印发〈2016年教育信息化工作要点〉的通知》,提出“推动各级各类学校数字校园建设与应用。充分调动学校的积极性,引导学校围绕教育教学和教师、学生、教务、后勤组织管理等方面,开展利用信息技术改变教学模式、进一步落实教学管理的数字校园/智慧校园应用”^[1]。通过数据化智慧校园建设,服务国家“双一流”高校建设,构建舒适便捷的智慧化校园环境,成为当前高校重要的战略任务。

在建设“双一流”高校进程中,高校面临着如何及时准确地从学校内部和外部发现自身的优势和不足、机会和挑战,以及教学、科研和管理决策等方面存在的问题。大数据技术的出现,对高校师生在科研、教学和学习中产生的大量数据进行挖掘,发现这些数据隐藏的价值,并将直观描绘高校的画像运用于许多场景,为服务高校教学科研和管理决策提供了便利。因此,本文将用户画像的思维运用于高校画像的构建,以期通过这一研究为高校管理者在营造智慧校园环境、服务教学科研、辅助教育决策等方面提供有效建议,助力高校“双一流”建设。

1 研究现状

目前,国内外画像研究主要围绕用户画像展开,用户画像的研究涉及电子商务^[2]、图书馆^[3]、健康

医疗^[4]、旅游^[5]等领域,主要研究对象包括图书馆用户^[6]、Web站点用户^[7]、Twitter用户^[8]、大学生群体^[9]、旅游用户^[10]等。在用户画像研究的基础上,国内外学者开始将画像研究向城市画像领域、教育画像领域和机构画像领域扩展。

(1)城市画像。Wong C等利用Trip旅游网站上的在线用户评论数据,使用文本挖掘技术提取澳门城市画像,借助可视化技术展现澳门城市画像的演化^[11]。Liu L等运用深度学习技术对Flickr网站上的文本、图像数据集进行分析,探讨了7个城市的空间分布特征,并分析各城市之间的相关性和差异性^[12]。毕崇武等基于“知乎”平台中的用户问答数据,建立标签关注度、新颖度和创新度3个评价指标,采用线性加权和法构造指标评价模型识别热点城市特征^[13]。毕崇武等还利用标签相似度和凝聚式层次聚类算法,提出一种层级结构的城市画像描述框架,借鉴LSA潜在语义挖掘思想,筛选出具有较高城市画像解释度的标签集合^[14]。任思源等基于大量城市历史快递数据,对快递公司的数据进行汇聚和预处理,提出寄递频次、寄递时间、寄递地址、寄递物品4个分析指标进行城市画像^[15]。

(2)教育画像。徐畅等提出高校信息素养教育用户画像模型框架,阐述了用户画像在高校图书馆信息素养教育中的作用^[16]。陈会等使用大数据技术构建学生学习画像基础模型框架,研究学生学习

* 本文系安徽省教育厅高校人文社会科学研究重点项目“大数据驱动下高校图书馆知识增值服务体系构建研究”(SK2016A0254)研究成果之一。

画像在个性化学习、问题预警及辅助学校决策等方面的应用,为提升学生的培养质量提供参考^[17]。

(3)机构画像。孟琳首次提出机构画像,从核心成员发现、社团发现、关系抽取及兴趣发现方面研究机构画像的动态属性^[18]。

从上述关于画像构建的研究来看,其研究方法主要有:基于用户行为、用户兴趣偏好、主题、人格特性与用户情绪等方法^[19-22]。目前高校画像的研究相对较少,即使已有的研究也多涉及高校的某一领域或某一群体,基本没有对高校整体画像进行系统性研究。因此,本文以高校为研究对象,概括高校画像的概念和基本特征,探讨高校画像的构建流程,并分析高校画像在具体场景的运用。

2 高校画像的概念界定及基本特征

2.1 高校画像的概念界定

高校画像源于用户画像研究。用户画像(User Profile)是由交互设计之父 Cooper A 最早提出的概念,作为真实用户的虚拟代表,用户画像是建立在一系列真实数据上的目标用户模型^[23]。随着画像研究的不断深入,学者们开始向机构画像领域研究拓展。20世纪60年代,Lynch K 就依据市民的心理形象衡量城市的视觉质量,并由此引出了城市画像的概念^[24]。此后,国内学者毕崇武等将城市画像定义为公众对城市特征的认知、评价和情感的综合体现^[13]。孟琳在城市画像的基础上,指出研究机构画像需要将机构看作一个整体^[18]。本文在借鉴用户画像、城市画像和机构画像定义的基础上,对高校画像的概念进行推定,认为高校画像是利用大数据技术将高校相关的各项数据进行整合、存储、融合和分析,对高校进行知识展示和多元应用场景服务的过程。

高校画像和传统用户画像存在不同之处。高校画像中的高校不是现实生活中真实存在的用户,而是由许多成员和不同群体组成,因此高校画像较传统用户画像更为复杂。但是,高校画像同样可以借鉴用户画像的思维,将其看作一个较为复杂的群体用户,根据高校自身的属性特征进行画像的构建。

2.2 高校画像的基本特征

Travis D 认为,用户画像一般具有7个特征:基本性(primary research)、真实性(realistic)、目标性(objectives)、移情性(empath)、独特性(singular)、应用性(applicable)和数量(number)^[25]。作为一个整体用户的高校,其画像在数据驱动环境下具有多元化特征。从技术视角观察,高校画像具有数字化特征;从

教育视角观察,高校画像具有共享性、交互性和个性化特征;从管理视角观察,高校画像具有人本性和整体性特征。

(1)数字化。通过大数据技术将高校各类信息进行整合、存储在数据库中,实现数字化。一方面极大地提高了信息查询和传播的速度;另一方面打破了学院间的信息孤岛,实现了数据交流,方便了教学科研与知识交流。

(2)共享性。数字化为便捷共享提供了前提。高校画像的共享性包含两个层面:一是高校内部的知识共享;二是高校间的知识共享。通过高校画像实现资源互通,打破高校间的资源封闭,为师生提供更多教学科研信息、知识和资源。

(3)交互性。信息共享为高校师生交互奠定了基础。高校画像不仅便利了师生之间的知识交流,而且提高了师生间互动的频率,使教学和科研等之间的交流由过去的单向灌输式转变为双向互动式,有利于通过多层次交流互动提升学生独立思考的能力。

(4)个性化。个性化特征基于大量个人行为数据分析。依据高校的科研与教学行为等数据,分析师生的学习行为特点、研究方向、经常访问的数据库和查阅的杂志等,为师生提供个性化的资源推荐、教学科研服务等。

(5)人本性。高校画像的根本目的是提供便捷精准的服务,即人本性是高校画像的本质特征。高校画像的出发点以高校师生为中心,更精准高效地服务于师生的科研、教学、学习和深造等,体现以人为本的发展理念。

(6)整体性。高校内部成员复杂、群体众多,因而高校画像专注于多个相互联系的要素或子系统的联系,从交互、共享和发展的角度体现高校画像的整体性。

3 面向多维数据的高校画像构建流程

在参考已有研究的基础上,将高校画像的构建分为以下五个步骤。

3.1 高校属性分析

高校属性是对高校发展特点和高校成员行为特征的描述,是建立高校画像标签体系的前提。根据高校画像的概念,将高校属性定义为{基本属性,院系科研属性,院系培养属性}的集合形式,基本属性从高校角度描述,院系科研属性和院系培养属性从高校院系角度描述。

首先,基本属性表示高校的基本信息,通过该属性可以使人们了解高校的整体情况。其次,院系科研属性表示高校每个院系的科研水平,该属性可以了解到高校每个学院的科研产出情况及每个教师的科研特点;院系培养属性表示高校每个院系的人才培养情况,通过该属性可以了解到各学院学生的招生情况、学习培养情况、深造情况和就业情况等。

3.2 高校画像的数据获取

在分析高校属性的基础上,梳理高校画像的数据来源、数据内容、采集方法和存储方式,如表 1 所示。

(1)基本数据。基本属性的相关数据来源于高校网站和百度百科,从高校官方网站和百度百科中通过人工采集相关数据。

(2)院系科研数据。院系科研属性的相关数据

来源于高校机构知识库、个人学术网站和学术社交平台,从高校机构知识库中导出高校每个院系的科研成果数据;从个人学术网站和学术社交平台中编写爬虫程序采集每个院系教师的学术看法和研究关系等数据。

(3)院系培养数据。院系培养属性的相关数据来源于高校网站、院系管理系统和校内学生管理系统,从高校网站和院系管理系统通过人工采集该校往年的招生数据、深造数据和就业数据;从校内学术管理系统导出每个院系学生的学习数据。

将三种属性的数据按照表 1 指定的采集方法获取后存储到 MySQL 数据库中,三个属性的数据存放在三张表中,分别命名为基本属性表、院系科研属性表和院系培养属性表。

表 1 高校画像数据来源

数据类别	数据内容	数据来源	采集方法	存储方式
基本数据	高校的基本情况数据,如学校地址、院系组成、教师队伍等。	高校官方网站	人工采集	MySQL 数据库
		百度百科		
院系科研数据	高校每个院系及院系教师从事科研活动的研究成果等信息。	机构知识库	人工采集	
		个人学术网站	编写爬虫程序采集	
		学术社交平台		
院系培养数据	高校每个院系专业的招生数据(录取分数线、录取人数等)、学生学习数据、学生深造数据和就业数据。	高校官方网站	人工采集	
		各院系管理网站		
		校内学生管理系统		

3.3 数据预处理

数据预处理是画像构建的必经环节。采集的源数据包含杂乱数据,存在数据不一致、数据缺失和数据重复等情况。数据不一致,可通过数据集成的方式将多个来源的数据整合到一起;数据缺失,可通过编写程序填充特殊值或空值解决;数据重复,可通过 SQL 语句中的 distinct 函数进行过滤。在解决以上数据问题后,还需要通过人工筛查的方式删除研究价值低的数据,获得所需的最终研究数据。

3.4 建立高校画像标签体系

使用数据挖掘技术对高校画像进行数据分析,利用分类、聚类、关联规则、回归分析等算法进行标签关联^[10],形成高校画像标签体系,如图 1 所示。从图 1 可以看到,高校画像由三个一级标签和十三个二级标签进行描述。其中,“基本属性”标签由学校名称、创办时间、办学性质、学校类别、学校特色、学校地址和院系构成七个二级标签组成,介绍高校

的基本信息;“院系科研属性”标签从科研成果和教师研究特点两个维度描述高校每个院系的科研水平和教师的科研特征;“院系培养属性”标签由招生、学习、深造和就业四个二级标签构成,分别描述高校每个院系学生的招生情况、学习情况、毕业深造情况和就业去向情况。

3.5 高校画像可视化

高校画像可视化是高校画像呈现的最后环节,可以帮助高校管理者理解高校画像的标签含义、分析高校每个院系的发展情况及全面了解高校的基本情况。随着技术的发展,画像可视化的方式越来越丰富,词云图、折线图、脸谱图等是画像研究者常用的可视化方式。可视化的实现难易及复杂程度不是评价画像结果的关键,笔者认为应根据画像的内容选择合适的可视化方式。本文分别从高校画像的三个属性介绍可视化的方式,高校画像可视化如图 2 所示。

(1)基本属性标签可视化。学校名称、创办时

间、办学性质、学校类别、学校特色和学校地址使用树图可视化;院系构成单独使用树图展示名称、专业、教师构成和学生构成;教师构成和学生构成使用饼状图可视化。

(2)院系科研属性标签可视化。高校每个院系的科研成果使用表可视化;院系每个教师的研究特点使用文本标签可视化。

(3)院系培养属性标签可视化。通过使用折线图可视化展示高校每个学院的往年招生情况;每个院系学生的学习特征使用文本标签可视化;深造情况使用条形图可视化;就业情况使用饼图可视化。

4 高校画像多维应用场景分析

高校画像是以高校属性的相关数据为基础的特征描绘视图,能够服务于高校决策管理与高质量发展,提升高校的社会形象。高校画像可以运用到智慧校园建设,全面支持学校管理决策,服务国家与地方发展等实际场景,实现高校画像研究的最终目的。

4.1 服务高校:支持管理决策

高校画像能够将高校的基本特征直观展示给高校管理者。高校管理者可以从内部和外部不同的视角了解学校,全面客观地反映高校的整体情况,发现

高校存在的问题,有利于快速做出科学决策。同时,高校画像通过对内部校园数据进行处理和建模,发现高校成员的行为特点,呈现高校教师和学生的研究偏好和学习偏好等,确立学校的专业发展特色和合理的学科发展定位。

4.2 服务院系:推动学科特色发展

高校院系师生画像可以直观分析各学院教师的研究偏好,精准分析各学院专业的特点、发展优势及研究热点和趋势。通过横向对比各院系学科发展特点,明确学校的优势学科和短板专业、发展潜力和不足之处,服务学校合理地配置整合资源,推进学科群相互支持,形成具有优势竞争力的学科,提高高校整体实力,实现高质量发展。

4.3 服务教师:保障教学科研

教学和科研是高校的主要任务。基于高校教学和科研画像,管理决策层可以精准获得高校科研资源的使用情况,包括实验室环境、设备使用效率、文献资源需求与保障情况等;可以及时精准地发现学校管理和运行中存在的问题,进而有效地解决问题,为师生的教学和科研活动构建良好的软硬件环境。

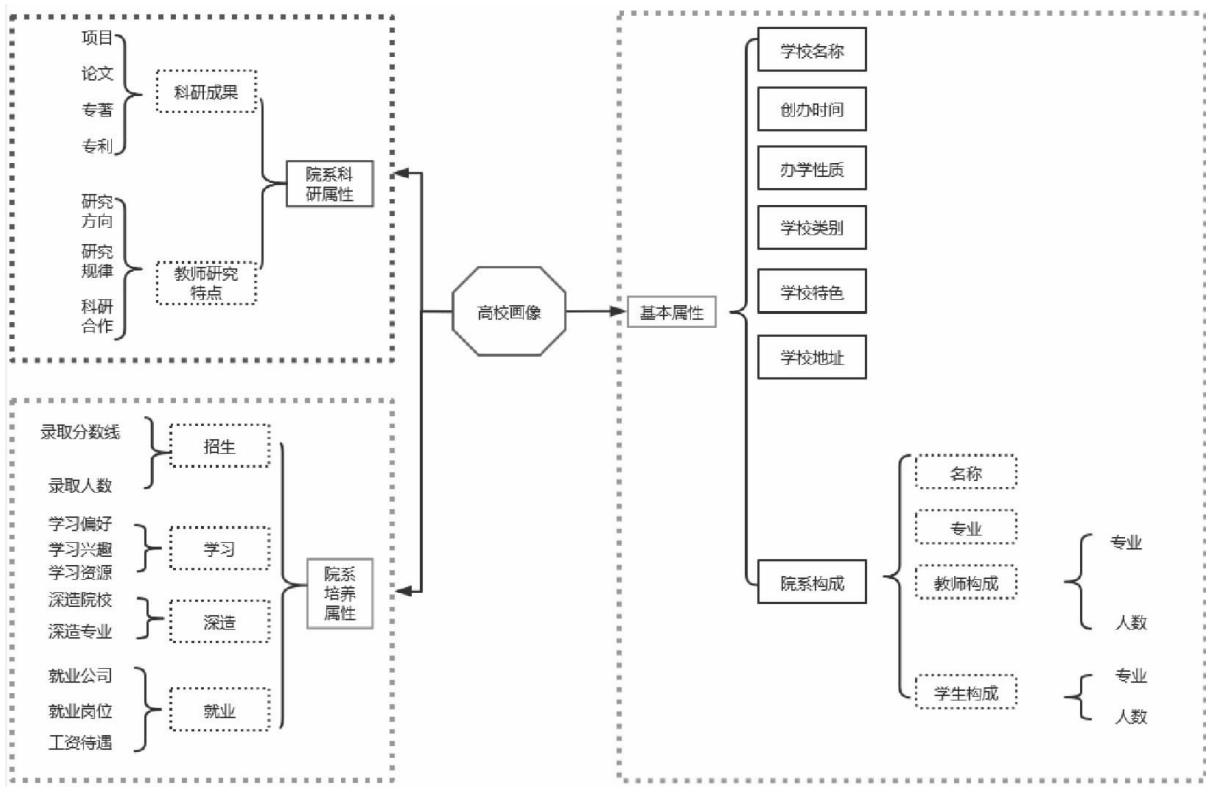


图1 高校画像标签体系

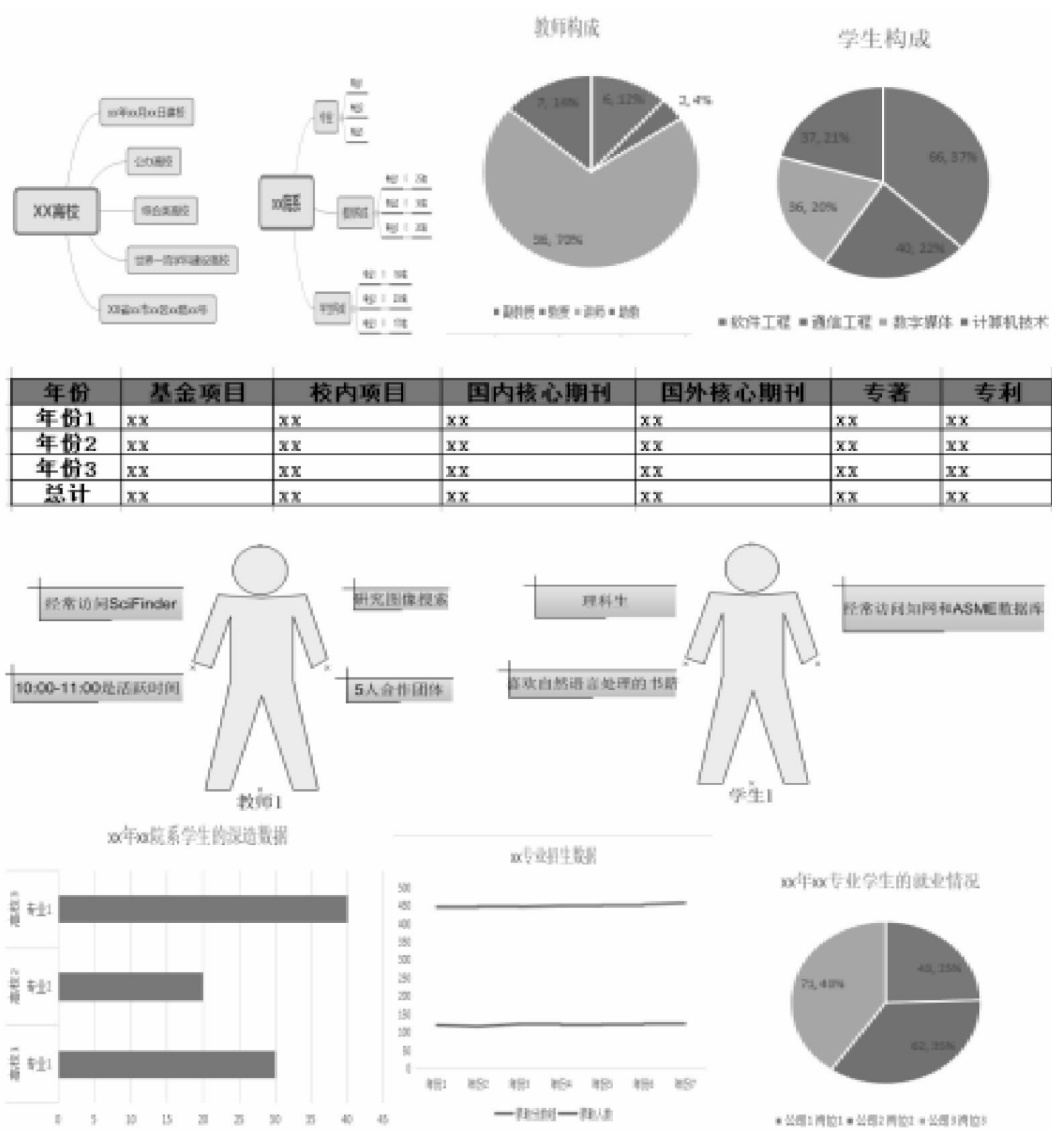


图2 高校画像可视化

高校院系教师画像可以从研究方向、研究规律和研究团体三个维度评价教师的科研行为,为其精准提供学术资源,帮助其优化团队建设,开展交叉学科研究。基于校园学习行为数据的高校学生画像,能够多角度呈现学生的学习偏好和学习行为特征,按照学生偏好进行群体聚类,有针对性地为其推荐学习资源。同时,教师和学生知识交流中实现精准关联,形成从“教”到“学”的闭环式联动反馈,提升学习、科研和管理等效率。

4.4 服务学生:拓展成长空间

(1)有效提升报考成功率。许多高考学生存在填报志愿时不了解学校、进入大学后不满意专业等问题。针对这些问题,高校招生画像能从高校的综合实力、专业实力和录取难度三个维度进行描述,其

中综合实力维度体现学校类型、学科数量和综合排名;专业实力维度展示优势专业、专业数量和知名度;录取难度维度呈现历年报考人数和录取分数线。通过以上三个维度的分析,高考生可以直观获悉高校基本情况,结合自身特点选择适合自己的高校,解决高考生填报志愿的烦恼。

(2)科学化规划职业。高校画像将往年毕业生的信息进行汇总分析,构建高校学生就业发展画像,为毕业生科学规划职业提供帮助。通过定期调查每个专业毕业生的就业单位、工作情况、工作感受等信息,利用大数据技术实现分类统计并将其可视化,形成有效的分析反馈机制,便于高校学生及时了解所属专业的就业情况,帮助学生科学开展职业规划,避免在择业中走弯路,更好地实现职业发展。

(3) 高效化招聘应聘。高校毕业生画像可以清晰地呈现学生的基本情况:专业课成绩、社会实践能力、奖学金获得情况及兴趣偏好等,校招企业可以根据画像结果更直观地了解学生的基本情况,按照自身的招聘需求,快速精准地筛选合适的毕业生;基于高校的学科特点和毕业生表现,高校能借助画像更准确地向企业推荐毕业生。同时,毕业生借助画像可以获得较准确的就业行情。可见,高校毕业生画像能极大地提高招聘和应聘效率。

(4) 精准化择校深造。选择深造是许多高校毕业生必经的成长路径。随着选择深造的学生人数不断攀升,如何选择合适的学校、专业和导师,一直困扰着许多学生。高校画像通过汇总往届各专业学生的在校成绩、录取院校,以及高校的录取难度、科研水平、历年报录比、研究生在校情况等数据,结合教师科研画像,为学生深造提供针对性参考,帮助学生根据自身学习实力和兴趣、专业水准和导师特点,精准选择适合的高校。

4.5 服务社会:助力发展转型

服务社会是高校的价值所在。高校画像不仅是高校内部的形象呈现,也是社会对高校外部形象的重要评判。高校形象的评判基于高校的社会服务、社会引领和社会贡献等指标而定。这些指标具体表现为高校的人才培养、研究成果的运用和推广、服务国家重大战略、推动地方经济社会发展和为人类社会所作出的贡献。高校画像涵盖了人才培养数量、学生就业状况、科技成果转化、引领社会发展、服务国家决策、支持重大产业发展以及服务社区等,能体现高校师生受社会欢迎的程度,准确反映高校的社会价值和社会贡献。

5 结论

随着数据研究和数据分析在高校建设和发展中作用的日益明显,本文以高校多源数据为研究对象,提出高校画像的概念和基本特征,按照界定高校属性、获取高校画像数据、数据预处理、建立高校画像标签体系和高校画像可视化的顺序分析了高校画像的构建流程。最后对高校多维应用场景进行了分析,提出了高校画像在高校管理决策、特色发展、精准化服务、科学研究、学生发展和社会服务等方面的创新作用,旨在通过智慧校园建设增加学校发展的服务供给。总之,本文研究的高校画像能为学校治理提供有效的服务和决策支持工具,帮助提高高校管理效率和治理水平,推动高校朝着更高目标发展。

参 考 文 献

- [1] 教育部办公厅关于印发《2016年教育信息化工作要点》的通知[EB/OL]. [2016-02-28]. http://www.cac.gov.cn/201602/28/c_1118180280.htm.
- [2] 洪亮,任秋圆,梁树贤. 国内电子商务网站推荐系统信息服务质量比较研究——以淘宝、京东、亚马逊为例[J]. 图书情报工作,2016(23):97-110.
- [3] Thompson S. Watching the movie using personas as a library marketing tool [J]. Reference & User Services Quarterly, 2017(1):17.
- [4] 唐晖岚,文庭孝,罗爱静,等. 网络健康信息精准服务模式研究[J]. 现代情报,2019(7):109-114.
- [5] Ravi Logesh, Vairavasundaram Subramaniaswamy. A collaborative location based travel recommendation system through enhanced rating prediction for the group of users [J]. Computational intelligence and neuroscience, 2016(9):20-26.
- [6] 于兴尚,王迎胜. 面向用户认知需求的图书馆用户画像研究[J]. 图书馆,2021(2):57-62.
- [7] 王菊艳. 基于WEB日志的用户画像及可视化分析[D]. 西安:西安理工大学,2019:11-32.
- [8] 韩月辉. Twitter数据采集与用户画像系统的设计与实现[D]. 苏州:苏州大学,2019:23-27.
- [9] 陈会,余馨,李琳琳,等. 基于大数据技术的大学生学习画像构建[J]. 软件工程,2021(3):36-38.
- [10] 刘海鸥,孙晶晶,苏妍嫒,等. 基于用户画像的旅游情境化推荐服务研究[J]. 情报理论与实践,2018(10):87-92.
- [11] Wong C, Qi S. Tracking the evolution of destination's image by text-mining online reviews - the case of Macau [J]. Tourism Management Perspectives, 2017(13):19-29.
- [12] Liu L, Bolei Z, Jinhua Z, et al. C-IMAGE: city cognitive mapping through geo-tagged photos [J]. GeoJournal, 2016(6):817-861.
- [13] 毕崇武,叶光辉,胡婧岚,等. 城市画像视角下的热点城市特征识别方法研究[J]. 现代情报,2020(4):13-22.
- [14] 毕崇武,叶光辉,李明倩,等. 基于标签语义挖掘的城市画像感知研究[J]. 数据分析与知识发现,2019(12):41-51.
- [15] 任思源,郭斌,张曼,等. 寄递大数据城市画像[J]. 浙江大学学报(工学版),2019(9):1779-1787.
- [16] 徐畅,毛桂芳,周吟剑. 基于用户画像的高校图书馆信息素养教育[J]. 大学图书馆学报,2021(1):28-31.
- [17] 陈会,余馨,李琳琳,等. 基于大数据技术的大学生学习画像构建[J]. 软件工程,2021(3):36-38.
- [18] 孟琳. 多源信息融合的机构画像的方法研究[D]. 北京:北京邮电大学,2018:6-34.
- [19] Iglesias J A, Angelov P, Ledezma A, et al. Creating evolving user behavior profiles automatically [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(5):854-867.
- [20] 张钧. 基于用户画像的图书馆知识发现服务研究[J]. 图书与情报,2017(6):60-63.
- [21] 马超. 基于主题模型的社交网络用户画像分析方法[D]. 合肥:中国科学技术大学,2017:10-15.
- [22] Zhou M X, Wang F, Zimmerman T, et al. Computational discovery

of personal traits from social multimedia [C]//IEEE. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). 2013; 1-6.

- [23] Cooper A, Reimann R, Cronin D. About Face 3: The essentials of interaction design[J]. Technical communication, 2008, 55(2): 199-200.
- [24] Lynch K. The Image of the City [M]. Cambridge, Massachusetts: The MIT Press, 1960; 3-5.

[25] Travis D. E-Commerce usability: tools and techniques to perfect the onLine experience [M]. Oxford: Routledge, 2002; 6-7.

[作者简介]任福兵,华东理工大学图书馆研究馆员;王朋,华东理工大学科技信息研究所图书情报与档案管理专业2019级硕士研究生。

[收稿日期]2021-12-15

(宋小华 编发)

Research on University Portrait Construction and Application Scenarios Based on Multi-Source Data

Ren Fubing Wang Peng

(East China University of Science and Technology, Shanghai 200237, China)

Abstract Based on the results of existing research, this study discusses the connotation and basic characteristics of university portraits, analyzes the process of constructing university portraits, and elaborates the specific practice of application scenarios of the portrait. This study found that university portraits have strong practicability, can be applied to universities, departments, teachers and students, and the society, etc., and can better serve the construction of university think tanks.

Keywords University portrait. Connotation. Basic characteristics. Construction process. Application scenarios.

(上接第 23 页)

- [6] 吴建中. 从量的发展到质的提升: 研究图书馆的新课题[J]. 图书馆杂志, 2017(10): 4-10.
- [7] 蔡迎春. 基于专业思维的“类 PDA”采购模式实践优化[J]. 图书馆杂志, 2021(1): 39-47.
- [8] 孔青青. 科研人员电子资源需求调查分析[J]. 图书情报工作, 2016(10): 47-54.
- [9] 王群, 王琼. ATO: “获取即拥有”电子书采购模式研究[J]. 大学图书馆学报, 2020(4): 34-40.
- [10] 邱葵. 后电子书时代的美国高校图书馆馆藏建设[J]. 图书馆论坛, 2020(10): 165-175.

[11] 陈磊, 陈惠兰, 詹司徒, 等. 学科竞争力视角下的馆藏资源建设探析——基于 Scival 跨学科科研评价工具[J]. 图书馆, 2015(5): 83-86.

[12] 张恬, 刘凤红. 数据出版新进展[J]. 中国科技期刊研究, 2018(5): 453-459.

[13] 苗美娟, 刘兹恒. 国内图书馆开放存取出版服务探析[J]. 国家图书馆学报, 2016(2): 60-65.

[作者简介]马越, 郑州大学图书馆副研究馆员。

[收稿日期]2021-09-14

(肖蔚 编发)

Digital Resource Construction of University Libraries Driven by the "Double First-Class" Construction

Ma Yue

(Zhengzhou University, Zhengzhou, Henan 450001, China)

Abstract Driven by the goals and tasks of the "Double First-Class" construction, the teaching, scientific research and management of universities put forward new requirements for the digital resource construction. This paper analyses the demand changes and existing problems of digital resource construction and proposes strategies and measures for the construction, aiming to provide references for the digital resource construction of university libraries under the context of "Double First-Class" construction.

Keywords "Double First-Class" construction. Digital resource construction. University library.