

项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	高职院校智慧图书馆用户画像体系构建与可视化表达-以北京农业职业学院为例				
主持人	付天新	职务/职称	工程师/信息系统项目管理师		
所在单位	(加盖单位公章) 北京农业职业学院图书馆				
专 家 意 见	<p style="text-align: center;">课题为新技术在图书馆工作中的应用研究。课题研究与发展方向和趋势相一致，选题与目前高职院校智慧图书馆建设的实际联系密切，对推动图书馆的建设有积极的现实意义。</p> <p>①课题在大数据视域下，梳理、整合、吸收前人经验成果，面向高职院校图书馆，统筹学院全域资源，详细探讨大数据视域下高职院校图书馆用户画像可视化构建与表达的数据体系、构建流程、系统框架与技术实现，通过研究能够快速支持高校图书馆用户画像可视化构建与表达的智能系统，提高图书馆用户画像可视化智能构建、表达与精准服务的速度与能力。</p> <p>②课题充分发挥大数据时代用户画像技术在智慧图书馆领域的技术优势，不断挖掘用户真实需求和潜在需要，不断探索新的应用场域，不断提高用户粘度与系统智慧化，以期提供更加精准的、智能的信息服务。选题新颖、紧密结合图书馆建设与发展实际，理论联系实际，具有一定的创新性。</p> <p style="text-align: center;">专家组认为该研究项目完成了计划任务书规定的各项研究任务，达到预期目标，一致同意通过课题验收。</p> <p style="text-align: right;">(如需要可增加页数)</p>				
专家签字	石进朝	刘诚斌	肖丽华	诸继红	王晶
职务/职称	教授	副教授	副研究员	副教授	副研究员



项目编号：2021008

注：项目编号请查看立
项通知，也可缺省

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称：高职院校智慧图书馆用户画像体系构建
与可视化表达-以北京农业职业学院为例

项目关键词：大数据；图书馆；用户画像；可视化；

项目单位（盖章）：北京农业职业学院图书馆

通信地址：(详细地
址含邮编)北京市房山区稻田南里5号 102442

项目主持人：付天新

联系电话：18510105848

电子邮件：fu_tianxin@163.com

提交日期：2022-05-10



高职院校智慧图书馆用户画像体系构建与可视化表达

— 以北京农业职业学院为例

付天新、张清华、辛力春、吕巧枝等

摘要:随着信息技术的不断发展,社会智慧化进程不断加快,用户画像被引入到智慧图书馆的建设中,并已应用到图书馆的智慧化服务领域。课题面向高职院校图书馆,统筹学院全域资源,详细探讨大数据视域下高职院校图书馆用户画像可视化构建与表达的数据体系、构建流程、系统框架与技术实现,通过研究能够快速支持高校图书馆用户画像可视化构建与表达的智能系统,提高图书馆用户画像可视化智能构建、表达与精准服务的速度与能力,进一步提高高职院校智慧图书馆建设水平。

关键字:大数据 图书馆 用户画像 可视化

Construction of User Portrait System and Visual Expression of Intelligent Library in Higher Vocational Colleges - Take Beijing Agricultural Vocational College as an example

Fu Tian-xin, Zhang Qing-hua, Xin Li-Cun, Lv Qiao-zhi, etc

Abstract: With the continuous development of information technology and the accelerating process of social intelligence, user portrait has been introduced into the construction of Smart Library and has been applied to the intelligent service field of library. The paper, Facing the university library, plans the overall resources of the University and discusses deeply in the data system, construction process, system framework and technical implementation of the visual construction and expression for the user portrait of the University Library from the perspective of big data, studies and constructs an intelligent system that can quickly support the visual construction and expression of the university library user portrait, and then improves the speed and ability of the visual intelligent construction and expression of the library user portrait, and further improves the construction level of University Smart Library.

Keywords: Big Data; Library; User Portrait; Visual Expression;

【中图法分类号】 G250.7

目 录

第一章 课题概述	1
1.1 研究背景	1
1.2 研究目的及意义	1
1.3 研究内容	2
1.3.1 构建大数据视域下智慧图书馆用户画像理论体系与应用框架	2
1.3.2 探索高职院校智慧图书馆用户画像框架应用	2
1.3.3 开展图书馆用户画像可视化表达与应用分析	2
第二章 图书馆用户画像的内涵、分型与内容描述	2
2.1 用户画像内涵	2
2.2 用户画像分型	3
2.2.1 用户画像本体视角分型	3
2.2.2 用户画像群落视角分型	4
2.2.3 用户画像功能视角分型	5
2.3 用户画像内容	5
第三章 图书馆用户画像的体系、框架与构建实践	6
3.1 数据体系	6
3.2 总体框架	9
3.2.1 基础层 (IAAS 层)	10
3.2.2 数据层 (DAAS 层)	10
3.2.3 平台层 (PAAS 层)	10
3.2.4 应用层 (SAAS 层)	10
3.3 构建实践	11
3.3.1 数据采集	11
3.3.2 数据管理	11
3.3.3 模型运算与数据计算	12
3.3.4 可视化构建与表达	13
第四章 研究结论与展望	18
第五章 项目成果	19
参考文献	21

第一章 课题概述

1.1 研究背景

随着信息技术的不断发展，社会智慧化进程不断加快，图书馆也逐渐由“数字”迈向“智慧”。智慧图书馆这一概念最早由芬兰奥卢大学图书馆的艾托拉提出，他认为“智慧图书馆”是一个不受空间限制并且能够随时被感知的移动图书馆^[1]。与传统图书馆注重馆藏、数字图书馆注重读者不同，智慧图书馆更加注重技术，结合与运用智能化技术，全面整合图书馆内、外信息资源，通过感知、捕捉、记录、挖掘、分析来提供更加智慧的服务^[2]。然而，信息技术高速发展的时代下，智慧图书馆建设并非一帆风顺。大数据环境下高校图书馆的智慧化服务面临着一系列挑战，面对丰富的信息资源和广泛的读者类型，现有服务模式与不同用户的多样需求之间的矛盾日益突出。高校图书馆要想更好地服务用户，就需要了解和满足用户需求，为用户提供个性化、精细化服务。“用户画像”这一最早应用于产品交互设计的概念被引入到智慧图书馆的建设中，并已应用到图书馆的智慧化服务领域。

在国外，以Ebsco作为检索数据库，“user profile”作为关键词检索项，近十年间共检索到5149篇文献。在国内，以中国知网（CNKI）作为检索数据库，把“用户画像”作为关键词为检索项，时间跨度选取2010-2020年，共检索到1335篇文献。通过分析检索国内结果可以看出：2015年以前“用户画像”极少受到国内学者的关注，2015-2016年属于起步阶段，2017-2019属于快速发展阶段。特别是2018、2019年这两年，国内关于用户画像的研究成果显著增多，用户画像引起了更多学者的关注。但从整体来看，用户画像整体发文量不多，研究热度有增长的趋势^[3]。同时，通过国内用户画像检索结果不难看出：国内图情领域涉及用户画像的研究成果大多面向某一具象服务，如：个性化推荐、智能推送、学科服务、阅读推广、资源发现、图书馆空间重构等^[4-6]。对于用户画像应用于智慧图书馆建设的理论、框架、流程、模型与可视化等方面的研究目前并没有形成一个完整框架体系。以基础框架体系为基础，面向具体某一类型的图书馆开展用户画像数据采集与处理方法、数据标签定义与选择、用户画像体系构建、模型策略与可视化等中微观层面的研究较少。国内图书馆界对用户画像的研究仍有很大的研究空间。

1.2 研究目的及意义

将用户画像技术引入高职院校图书馆，梳理、整合、吸收前人经验成果，从智慧图书馆用户画像定义与内涵、分型与内容、框架体系、构建过程以及应用实践等多个方面出发，构建大数据视域下智慧图书馆用户画像理论体系与应用框架。并以北京农业职业学院（以下简称北农职）为例，以用户、资源为主线，以框架体系为支撑，融合教育教学、科学研究、智慧服务等多项要素，深入剖析与研究用户画像数据采集与处理方法、数据标签定义与选择、用户画像体系构建与可视化。进而充分发挥大数据时代用户画像技术在智慧图书馆领域的技术优势，不断挖掘用户真实需求和潜在需

要，不断探索新的应用场域，不断提高用户粘度与系统智慧化，进而提供更加精准的、智能的信息服务。

1.3 研究内容

1.3.1 构建大数据视域下智慧图书馆用户画像理论体系与应用框架

在前人研究成果的基础上，从智慧图书馆用户画像定义与内涵、分型与场域、框架体系、构建过程、应用实践等角度出发，探索构建大数据视域下智慧图书馆用户画像理论体系与应用框架。

1.3.2 探索高职院校智慧图书馆用户画像框架应用

以北农职为例，建立大数据环境下高职院校图书馆用户画像的数据体系。结合高职院校实际情况，面向“不同场景、不同层次、不同维度”灵活的组装与运用，构建覆盖高职院校图书馆的用户画像标签体系与集成应用方法，促进用户画像成果的深入应用。

1.3.3 开展图书馆用户画像可视化表达与应用分析

对高等职业院校图书馆多维用户画像及其场域模型进行可视化表达，分析与挖掘各类用户画像的服务场景，形成一套完整的可视化表达策略与形式。围绕场景内容，结合用户画像的重要描述参量，研究场景化精准服务模型的构建方法和策略，面向用户提供延伸服务，提高高职院校智慧图书馆服务能力与信息水平。

第二章 图书馆用户画像的内涵、分型与内容描述

2.1 用户画像内涵

在国外，根据关注用户侧重点不同，通常从信息标签、用户需求和用户行为等维度来表达用户画像。研究学者通过信息标签刻画用户特征，反映用户全貌；通过用户需求挖掘用户典型特征，辅助产品设计；通过用户行为洞察用户偏好，提供智能服务^[9, 10]。交互设计之父 Alan Cooper 在 1998 年首次提出用户画像的概念，他指出：用户画像是现实生活中真实用户的虚拟代表，是建立在真实数据之上的目标用户模型^[11]。Djilali 等^[12]将用户画像定义为包含用户特征、目标 and 需求等信息的用户描述模型。Jisun 等^[13]认为，用户画像是有着共同行为特征的用户群的虚拟代表。Amato 等^[14]认为用户画像是从海量信息中获取的、由用户信息构成的形象集合，可以描述用户的兴趣偏好、用户的信息需求及用户的预期特征值。

在国内，随着信息技术的高速发展，大数据、云计算、物联网、人工智能、5G 通信等技术的广泛应用给图情领域的研究与发展带来了新的机遇。2014 年李业根首次将用户画像应用到图书馆领域，打开了国内图书馆领域用户画像研究的大门。吴智勤等^[15]认为用户画像是数据分析基础上用户标签的集合，标签刻画用户特征并能精准提供个性化服务和洞察用户行为规律。朱静等^[16]认为用户画像

是以大数据技术为基础,提取出包含静态信息和动态信息,并应用用户信息勾勒出的用户全貌。兰冰等^[17]认为用户画像是通过收集用户相关信息,采用定量与定性相结合的分析方式,建构符合用户特点的标签化模型,以完整细致地勾勒与表达出用户信息,进而更好地预测用户需求的工具。宋美琦等^[18]认为,用户画像的实质是标签化的用户全貌。余传明等^[19]认为,用户画像是根据用户人口统计学信息(Demographic Data)、社交关系(Social Network Relationships)和行为模式(Behavioral Patterns)等信息而总结、抽象和挖掘出来的标签化用户模型。王丽艳^[20]等认为大数据时代下图书馆用户画像即用户信息标签化,是以图书馆服务系统及其相关平台一系列真实数据为基础,借助数据预处理、机器学习、数据挖掘、可视化等技术自动提取出个体或群体用户标签,从而建立起联系用户诉求与设计方向的目标用户模型。陈慧香等^[21]认为用户画像是建立在一系列真实数据之前的描述用户需求和偏好的目标用户模型,该模型可全方位、立体化地反映用户特征。胡媛等^[22]认为数字图书馆将知识社区用户信息抽象化并运用聚类、关联规则及分类等数据挖掘方法汇制所得的用户可视化画像即为用户画像。张洁等^[23]认为,数字图书馆用户画像主要指面向真实读者用户,以用户的静态属性(人口统计特征、科研属性特征、空间和地理特征等)和动态属性(访问行为、资源检索及获取行为、学术社交行为、学术成果发表行为等)数据为基础,综合应用文本挖掘、机器学习等方法提炼出的具有显著特征的用户标签集合。

由以上可以看出,关于用户画像的定义,随着用户画像研究的设计思路、实现技术、主题、场域的不同,对其定义与内涵也有不同理解与侧重。但在各自不同的表达中又体现着对用户画像的统一理解,即用户画像是实现图书馆智慧化的新理念、新技术和新模型。与用户画像应用之初不同的是,在大数据时代,可供采集和挖掘的用户数据维度更多、数量更大,目标用户的特征更具体,生成的用户画像也更清晰真实。

2.2 用户画像分型

随着图书馆用户画像研究的作主体、群落和用途的不同,在各自的视角下有不同的分型。本课题通过总结前人的研究成果,梳理与归纳出图书馆画像在本体、群落和功能等视角下的3种分类。即:以本体视角出发,将其划分为读者画像、机构画像、资源画像和场域画像;以群落视角出发,将其划分为个人用户画像和群组用户画像;以功能和用途视角出发,将其划分为描述型用户画像、服务型用户画像和评价型用户画像。具体描述如下:

2.2.1 用户画像本体视角分型

①读者画像

以图书馆读者/用户为中心,围绕读者/用户的人口统计学特征、社会关系和行为模式等数据进行系统分析、归纳和总结,构建能够清晰描述读者/用户全貌、精准刻画读者/用户行为特征、深入挖掘与提炼读者/用户信息标签的模型集合或可视化系统。人口统计学特征信息主要描述读者/用户的姓名、性别、年龄、职业、教育背景和联系方式等信息。行为和偏好信息主要描述的是读者/用户系统信息检索、页面浏览、内容下载等行为。社会关系特征信息主要描述的是读者/用户阅读过程中产生的分享、评论和收藏^[24]。

②机构画像

以图书馆内、外相关机构为中心。机构画像是把机构和内部成员视为统一整体而非单一个体，用以研究机构的固有属性、行为动态和发展规律，归纳总结与其他机构的差异性，从而构建标签化机构模型。机构是不同个体的组合体，机构画像所包含的信息是不同用户画像信息的组合体，它主要包括静态特征(如机构名称、成立时间、地理位置和组织架构等)和动态特征(如核心成员和社团、社交网络、科技项目、科技论文、科技奖项、知识产权和排行等)。通过机构画像能够全面、精准、动态地呈现机构的不同特征，不但能够辅助图书馆了解掌握机构的科研动态，而且能够为机构的不同需求提供优质的知识服务，满足机构的创新发展需求^[25]。

③资源画像

以图书馆纸本、电子资源为中心，以图书馆资源的基础属性、媒体类型、内容属性和资源热度等方面入手，建立图书馆资源数据体系，构建能够清晰描述资源全貌、全面表达资源特征、深入挖掘与提炼资源标签的模型集合和可视化系统。基本属性主要描述的是图书资源的作者、出版社、出版时间、资源分布等数据信息。媒体类型主要描述的是图书资源的媒体类型，如文本、图片、音频和视频等数据信息。内容属性表达图书资源的主要内容、知识主题和研究领域等。资源热度则是图书资源被用户检索、浏览、下载次数等数据的综合体现^[24]。

④场域画像

又可称之为“主题用户画像”或“专题用户画像”。其主要是以图书用户画像的应用场域(主题或专题)为中心，从图书馆用户画像各类应用场景和智能服务入手，融合图书馆读者/用户、机构、资源等多个主体信息，面向特定场域/场景构建能够清晰描述场域(主题或专题)信息全貌、精准刻画应用场域特征、深入挖掘与提炼场域应用标签的模型集合和可视化系统。用户画像为图书馆提升服务效能和精准服务提供了广阔的空间。目前用户画像已经应用在图书馆智慧化的各个方面，如：基于用户画像的知识服务、基于用户画像的阅读推广、基于用户画像的个性化推荐、基于用户画像的学科服务等^[26]。

2.2.2 用户画像群落视角分型

①独立用户画像

是以个体为主要研究中心，对研究个体(如:读者、课程、品目、技术、工艺等独立实体)的全貌、特征进行精准的表达与刻画，深入挖掘与提炼研究个体场域应用标签的模型集合或可视化系统。独立用户画像用来表达研究群落中最核心和重要的对象个体，如:构建图书馆读者/用户画像;构建核心课程画像;构建某个品目(生猪、牛、养、鸡、果树、花等)在画像等。

②群体用户画像

是以组织为主要研究中心，对研究群组(如:机构、院系、团队、社团等)的全貌、特征进行精准的表达与刻画，深入挖掘与提炼研究个体场域应用标签的模型集合或可视化系统。群体用户画像以群组为单元来表达研究群体的信息全貌和特征。

2.2.3 用户画像功能视角分型

①描述型用户画像

目标在于理解、掌握研究主体的信息全貌。通过构建描述型用户画像直观、全面、准确的表达研究主体属性、特征、行为规律等，进而使人们能够在充分了解主体信息的情况下，结合实际开展各项服务，如：通过构建图书馆读者用户画像了解并掌握学生的年级、专业、行为习惯等信息，为学生提供图书资源个性化推荐或知识咨询服务；通过构建院系/机构用户画像了解、掌握其信息，向学院/提供更好的学科服务。

②服务型用户画像

目标在于挖掘信息、提供智能化的服务。通过构建服务型用户画像，深入挖掘潜在信息，发现规律与痛点问题。并通过数据、模型、系统、服务等维度的前后关联与协作，完成服务场景描述、构建与实现。如：通过构建服务性画像实现图书的智能推介、馆内空间的预测以及学科的智能咨询等。

③评价型用户画像

目标在于整合数据信息、提供智能分类、定级与评价。通过构建评价型用户画像，梳理、分析、整合目标主体的数据信息，经数据挖掘、场景匹配、信息归一化处理等过程，最终形成能够清晰、准确表述场景特征的评价信息。如：通过图书馆读者的到馆次数、借阅信息等评价读者属于主动型、潜力型、激发型、游离型中的哪一种。

2.3 用户画像内容

科学文献的主题分类、学科分类与图书馆用户画像研究内容和应用场域直接相关，用户画像具有较强的领域性^[27]，其研究不再是大范围地粗略关注用户，而是进一步聚焦从用户身上所折射出来的各类标签以及这些标签所抽象呈现的各类栩栩如生的图书馆场景^[28]。随着图书馆用户画像研究的不断深入，国内图书馆用户画像已经积累大量研究文献，研究内容基本覆盖理论应用、模型构建、系统研究、实践运用、服务创新^[29-32]等诸多方面。课题以中国知网 CNKI 数据库为实证库，并以“主题 = 图书馆 * 用户画像”为检索式，检索时间范围为 2014 年 1 月 1 日至 2021 年 12 月 31 日之间的科学文献。选取实证样本中文献量排名前 20 的主题(见表 1)进行内容与场域的分析发现：图书馆用户画像的研究内容可划分为本体、场域、技术和服务四个方面(见图 1)。

表 1 国内图书馆用户画像科学文献来源主题分布

主题分类	文献量(篇)	主题分类	文献量(篇)
用户画像	158	构建研究	13
图书馆	78	公共图书馆	11
高校图书馆	65	学科服务	11
大数据	30	知识服务	11

智慧图书馆	24	阅读推广	11
精准服务	21	模式研究	10
数字图书馆	17	数据驱动	9
推荐服务	16	个性化服务	8
图书馆用户	16	个性化推荐	8
移动图书馆	14	大数据环境	7

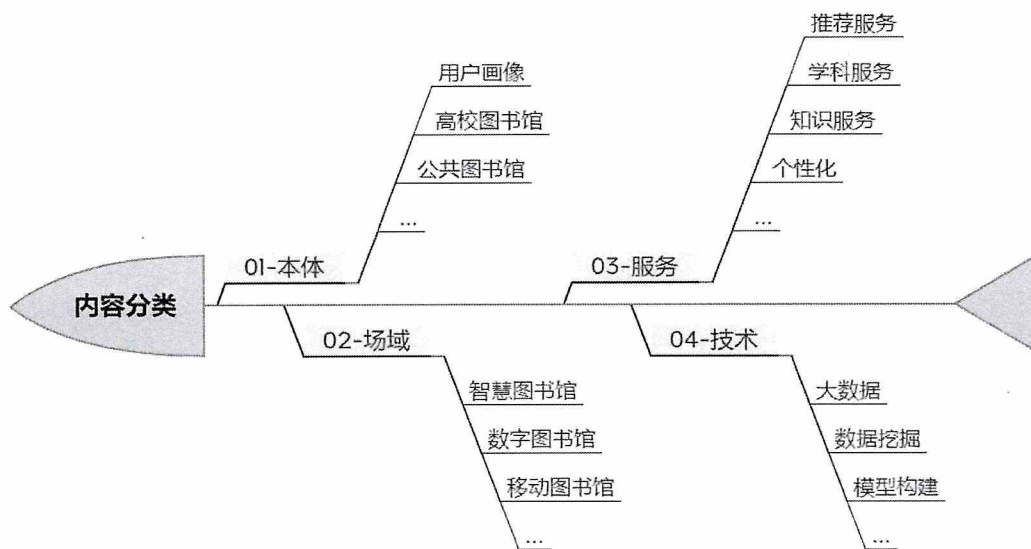


图 1 国内图书馆用户画像研究内容与场域分析

具体描述如下：

①本体：是图书馆用户画像研究主要面向的对象，是与其他主题具有高度联系的研究主体，如：“用户画像”、“高校图书馆”、“公共图书馆”等；

②场域：是图书馆用户画像研究主要面向的场景，是图书馆用户画像研究具体的细分领域，如：智慧图书馆、数字图书馆、移动图书馆等；

③技术：是图书馆领域用户画像研究的技术手段与途径，是图书馆用户画像研究得以实施的“加工厂”，如：大数据、模式研究、构建研究、数据驱动等；

④服务：是图书馆领域用户画像研究定制能力的输出，是图书馆用户画像研究面向用户应用的“百宝箱”，如：推荐服务、学科服务、知识服务、阅读推广、个性化服务等。

第三章 图书馆用户画像的体系、框架与构建实践

3.1 数据体系

图书馆领域的用户画像构建与应用离不开数据的支撑。数据是图书馆用户画像构建的基础，也是精准刻画用户特征、真实表达用户需求、直观洞察用户偏好和行为的关键。以北京农业职业学院为例，图书馆用户画像数据可以通过图书馆智慧门户、移动图书馆、图书馆管理系统、门禁系统、科研系统、教务系统、一卡通系统、后勤保障系统、以及公共开放服务等系统平台获取。学者朱静^[33]将构建图书馆用户画像的数据划分为静态数据和动态数据两大类，并将这两大类数据细化为用户属性数据、用户人格数据、内容偏好数据、互动数据、会话数据和情境数据 6 个子类来描述图书馆用户画像的数据体系。学者许鹏程等^[34]将构建图书馆用户画像的数据划分为用户人口统计学数据、用户行为数据、用户社交数据、用户其它数据，提出图书馆用户画像数据体系是由结构化数据、半结构化数据和非结构化数据组成，结构化数据容易采集、便于标签化。半结构化数据和非结构化虽然采集和处理难度较大，但数据量庞大，也是用户画像的主要数据。

课题以大数据技术为支撑，以图书馆用户画像研究为本体，突破传统图书馆资源对构建用户画像的数据限制，打破数据壁垒，从学院整体出发，融合多部门数据资源，并从图书情报、科研、教务、机构、保障和情境 6 方面建立学院图书馆用户画像数据体系，形成图书馆用户画像多维六面的数据体系（见表 2）。数据体系以“校园一卡通”中的用户基础信息为关联纽带，关联图书馆用户画像数据体系中的 6 类数据，并分层表达个人、群体、资源等不同类别的用户画像构建应用的数据群集（见图 2）。

表 2 图书馆用户画像构建数据体系与维度对照

序号	数据维度		数据内容	来源
	一级分类	二级分类		
1	自然维度	基本属性	用户 ID、姓名、性别、年龄、学历、职业、籍贯、邮箱、电话等	校园一卡通系统 图书馆集成管理系统 门禁管理系统
		科研属性	职称、部门、研究专业与研究方向	人事系统 科研系统 农职文库
2	兴趣偏好维度	行为特征 (显性)	检索、浏览、借阅、下载、试读、评论、分享、复制、收藏、点击、跳转、咨询、预约	门户网站 图书馆集成管理系统 馆藏目录检索系统 (OPAC) 空间预约系统 数字资源管理平台 智能咨询机器人

序号	数据维度		数据内容	来源
		行为特征 (隐性)		登录次数、浏览时长、频次、关注度、资源下载量、咨询量、浏览路径
	阅读偏好		作者、内容、学科、主题	
3	情境感知维度	自然情境	自然条件, 用户周围的天气、温度、湿度、风级和噪声	智能终端设备 传感器
		时间情境	时间、月份、时间段	Wi-Fi 蓝牙 RFID、GPS 北斗、无线网络基站
		位置情境	所在位置和位置类型	
		访问情境	电脑型号、使用的操作系统、客户端版本、网络类型	国资管理系统 移动图书馆
		活动情境	报到、竞赛、考试、出国、实习、答辩	门户网站
4	网络社交维度	互动	点赞、分享、评论、讨论、互动、关注、引用、被引、情感态度等	门户网站 移动图书馆 数字资源管理平台
5	实体联系维度	图情	类型、名称、作者、编号、主题、关键词、ISBN/ISSN、中图分类号、摘要	图书馆集成管理系统
		机构	院系(专业、学生、教师)、行政、教辅	统一机构与认证平台
		教务	学生、课程、成绩、奖励	教务管理系统 网络教学平台
		科研	教研、科研、考核评价	科研系统 农职文库
		保障	宿管、餐饮、交通	后勤管理平台

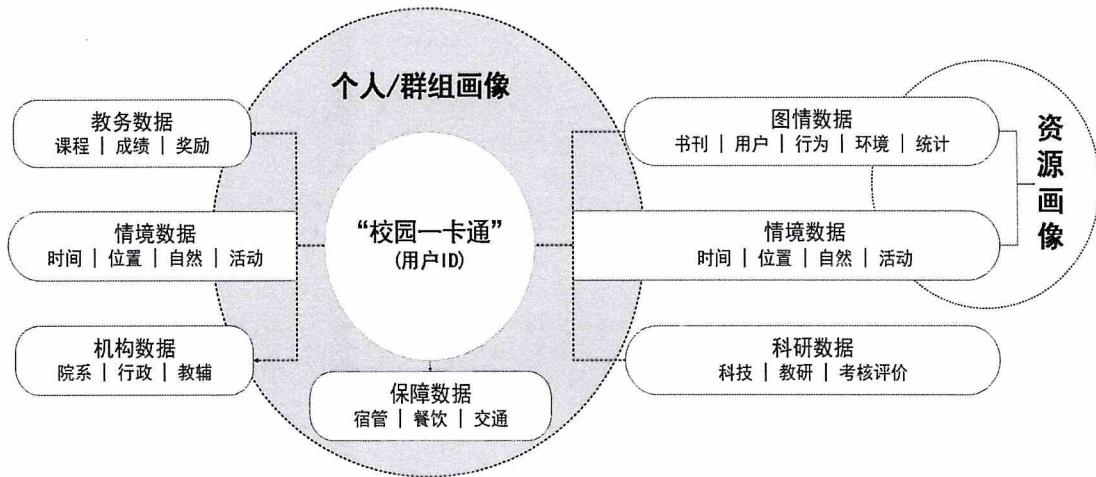


图2 图书馆用户画像可视化构建数据体系

3.2 总体框架

国内图图书馆用户画像在理论应用、模型构建、系统研究、实践运用与服务创新等诸多方面积累了大量的研究成果^[35-45]。信息技术的发展与大数据时代的到来给图书馆用户画像研究带来了新发展和新内容。利用大数据的技术优势不断升级系统框架、不断提高运行效率、深入挖掘用户真实需求和潜在需要、探索新的应用场域、提高用户粘度与系统智慧化，为图书馆用户画像研究保驾护航。尹婷婷等^[46]将大数据背景下高校图书馆用户画像信息服务应用的建模步骤总结为服务场景抽取、标签描绘体系建立、标签抽取与映射、用户数据集生成、数据挖掘与建模以及可视化呈现。许鹏程等^[34]认为图书馆画像模型构建流程大致可分为数据采集、数据预处理、数据存储、数据挖掘、形成用户画像、可视化及应用。宫兆阳等^[47]将城市运行体征概念引入到高校图书馆智慧服务中，通过研发图书馆数据综合分析系统智能分析图书馆“健康状态”，最终形成包含运行体征数据可视化、数据分析报告、数据智慧化分析的三项数据服务内容。



图3 图书馆用户画像可视化构建总体框架

课题在吸取前人研究成果的基础上，以学院图书馆为核心，统筹学院全域信息资源，融合图情、教务、科研、保障、学工等多个机构部门数据，形成覆盖高校全域、统筹利用、统一接入、数据共享的学院图书馆用户画像可视化系统框架，实现数据、系统、资源的有机整合。课题采用大数据的4层体系结构将图书馆用户画像可视化构建框架划分为基础层(IAAS层)、数据层(DAAS层)、平台层(PAAS层)和应用层(SAAS层)，框架如图3所示。

3.2.1 基础层(IAAS层)

是构建框架的基础。提供存储、计算、网络与虚拟化等资源和核心服务与计算能力。基础层结合学院信息化建设情况，采用公有云、混合云、私有云和自主虚拟化等不同的方式和策略搭建基础环境，通过提供存储、服务、管理和桌面等服务，满足上层对基础设施层的要求。学院全域的图情、教务、科研、保障等信息系统平台统一管理其中，并部署在大数据基础环境里。

3.2.2 数据层(DAAS层)

是构建框架的“原材料”。通过ETL技术(数据仓库技术)对来自学院全域、多源、异构的信息资源进行采集、清洗、加工转换与加载，实现结构复杂、标准不一、分散零乱、语义模糊的多源异构数据的整合。数据层通过ETL技术完成图情、教务、科研、保障、机构、情境6类数据的清洗转换，将其处理加工成为能够被学院图书馆用户画像可视化构建识别的“干净”数据。数据层建立统一数据访问引擎，提供支持Oracle、SqlServer、MySQL、Hive、MongoDB以及接口服务等灵活多样的数据访问策略，实现研究数据模型的封装，进而为图书馆用户画像可视化构建提供统一、规范、一致的数据能力。

3.2.3 平台层(PAAS层)

是构建框架的“配件厂”。平台层采用微服务的架构设计，以Hadoop生态、Spark、Kafka、Sqoop等大数据技术为平台底座，向下承接数据层，加载与传输数据；向上支撑构建耦合度低、灵活通用的可视化组件集和模型库。平台层包含数据视图、图表、报表、报告、仪表盘、词云等20余类可视化组件和4大类30余个小类的模型组件，组件与模型统一注册、集成与管理在可视化引擎之中，并通过可视化引擎实现各类组件的灵活组装、调度、监控与集成应用。

3.2.4 应用层(SAAS层)

是构建框架的“设计师”。图书馆用户画像可视化构建的业务逻辑、场域应用和可视化呈现均在应用层。应用层通过用户可视化构建的场域和对象来确定采用画像树表、图表、热词云图、书面报告和可视化展板等可视化表达的呈现形式。大数据视域下的图书馆用户画像可视化构建有效融合计算机的计算能力和人的认知能为，通过灵活、直观的可视化用户界面，将大规模复杂逻辑实现和可视化呈现封装在组件之中，用户只需要通过简单的拖拽和参数配置，即可完成用户画像的自动生成与表达并面向个体、机构、系统等提供服务。

3.3 构建实践

从用户画像可视化构建流程的视角，可将大数据视域下图书馆用户画像构建划分为数据获取、数据预处理、数据挖掘、标签提取与数据可视化五个阶段^[48]。其中，数据获取是“原料”，数据处理是“工艺”，数据挖掘是“大脑”，标签提取是“场域”，可视化是“加工厂”。而基于大数据视域下图书馆用户画像可视化构建总体框架研建的数据综合分析平台则是完全统筹五项内容的“智者”，是用户画像可视化构建的核心，是一套融合学院全域系统、整合多源数据，集数据服务、智慧化分析于一身的综合服务平台。从用户画像可视化构建的技术实现视角，数据综合分析平台包含数据采集、数据管理、模型运算与数据计算和数据可视化四大功能模块。各模块既联系紧密又彼此独立，模块之间的交互遵循统一数据与服务标准，各模块功能的不仅可以在平台内部使用，也可以集成对接符合统一数据与服务标准的第三方数据和系统，使得平台灵活、智能。

3.3.1 数据采集

数据采集是构建用户画像的基础，精准、全面的数据采集是用户画像能够精准挖掘用户需求的前提^[49]。数据采集实现结构复杂、标准不一、分散零乱、语义模糊、跨系统、跨机构的异构数据的融合与处理，是数据综合分析平台数据来源的唯一入口。数据采集模块集成统一数据访问引擎（DataStore），支持丰富的数据形式，提供支持各种数据库（Oracle、SqlServer、MySQL、Hive、MongoDB 等）、文件、API 接口服务等数据访问与读取策略，提供批处理采集和实时数据采集两种模式。批处理采集模式可对跨系统、跨机构的异构源数据进行全量、增量和自定义抽取，实时采集模式支持文件、互联网数据、第三方数据与系统的实时采集。数据采集的通过任务管理、清洗规则、数据加载、执行监控等功能实现从“源库”到“目标库”的清洗、转换与加载。数据采集过程通过统一数据访问引擎访问与读取数据；通过设置清洗规则和转换策略，对数据进行过滤、去重、更正、剔除、补充和脱敏，实现数据归约，达成多源异构数据的高效整合^[48、50]；通过统一的清洗转换调度与执行监控策略（如：任务调度、资源调度、执行监控与异常处理恢复机制等），对采集任务进度实时监控，保障数据采集任务的顺利执行。

3.3.2 数据管理

数据管理以 Hadoop 生态及 Spark 技术为基础，对数据采集、清洗、转换后加载至“目标库”中的数据资源进行统一的管理与维护。数据管理包括数据资源目录、元数据管理、数据稽核与质量管理、数据仓库管理等功能。平台通过数据资源目录的建设，摸清数据资源家底、明确各类数据的来源、内容、质量、权限与更新要求等数据特征，进行数据资源编目和分类标签，构建学院图书馆用户画像数据目录体系和管理分类体系，实现数据资源的快速定位与查询。通过元数据管理对元数据的分类及术语进行维护管理，使用元数据管理工具对数据的具体含义、数据字典、数据处理流程等进行管理，提高数据使用的标准化与可用性，使组织混乱、难于应用的零散数据变为标准统一、井然有序、具有核心价值的资产，进而支撑大数据视域下高校图书馆用户画像可视化构建对数据的调用需要，产生越来越多的增值效用。通过数据稽核与质量管理对构建图书馆用户画像的数据进行质

量管控，包含数据稽核规则配置管理、质量模型管理、稽核任务调度配置、稽核问题管理和稽核结果分析等功能，其重点是实现数据的完整性和一致性检查，提升数据质量。数据稽核是包含数据采集、预处理、比对、分析、预警、通知、修复在内的完整数据质量管控链条。用户根据预先配置的规则、算法和质量检查度量，对数据的准确性、合理性、一致性、安全性进行多维度的检查，及时发现问题和解决问题。数据仓库是大数据视域下图书馆用户画像构建全域数据存储与管理的主阵地。平台按照数据仓库的分层思想，结合用户画像可视化构建过程和阶段，将其数据划分为三个层次进行组织管理，即 ODS(数据运营层)、DW(数据仓库层)和 DM(数据服务层)。ODS 存储与管理的是采集数据，采集数据是经过 ETL 处理、清洗、转换后得到符合标准、具有业务含义的“纯净数据”，可存储在 Oracle、SqlServer、MySQL、Hive、Hbase、HDFS、MongoDB 等数据库建立的数据仓库中，并利用 Kafka、Spark 流式处理技术保持数据的持续更新与同步。DW 数据存储与管理的是宽表数据。宽表数据承上启下，向下对接 ODS，面向某一主题域进行不同维度的数据汇总、公共指标计算。向上对接 DW，构建面向主题和场域应用的主题库和标签库。DM 数据存储与管理的是场域数据集(主题库、标签库、数据集市)，大部分是来自于 DW 层中具有维度信息的各类宽表数据，数据内容本身不再具有明细数据，是数据挖掘或模型运算后直接应用于画像可视化表达与呈现的结果数据，具有高度的主题性和决策指导价值。数据仓库贯穿于图书馆用户画像构建与表达的全过程，既是大数据视域下高校图书馆用户画像构建与应用数据的“来源”，也是“归处”，其逻辑关系如图 4 所示。

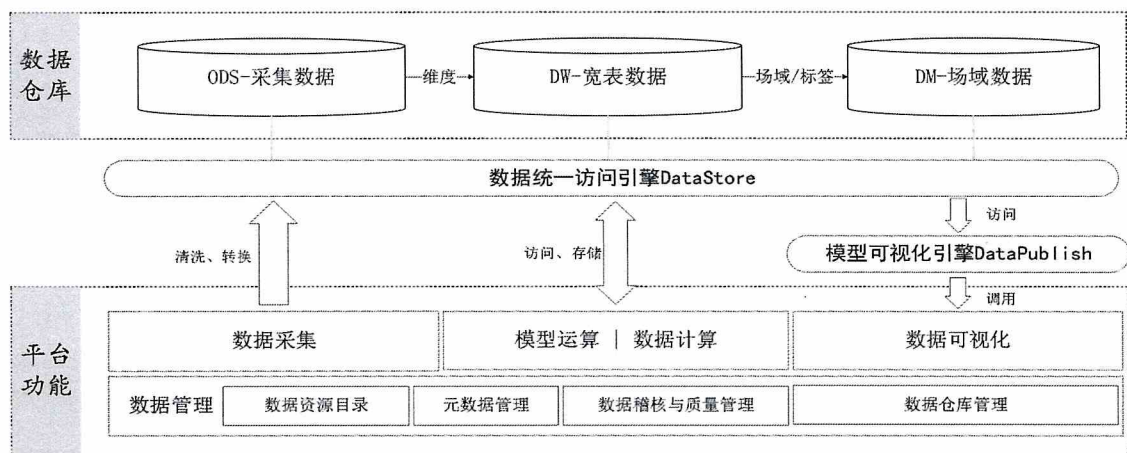


图 4 图书馆用户画像可视化构数据-功能逻辑关系

3.3.3 模型运算与数据计算

模型运算与数据计算是数据综合分析平台的核心功能，是以数据仓库中的采集数据、宽表数据为基础，通过数据挖掘技术发现图书馆用户画像构建的价值数据。模型运算与数据计算内置涵盖分类、回归、聚类等 4 大类 30 余个小的模型组件（见表 3），提供从数据源选择、数据合规性检测、模型环境参量配置、模型构建与执行、模型评估等全流程、一站式的可视化构建功能，并在分布式内存计算和分布式离线计算的支持下，完成模型的构建、训练与验证。模型运算与数据计算实现模型与数据算法的封装，将各类模型算法抽象成为可操作性、独立的图形组件，图书馆用户画像构建人员按照模型/数据的构建过程搭建工作流程，通过拖拽模型/数据图形组件、建立模型/数据关系、配

置模型/数据环境参量、触发模型/数据执行，完成数据模型运算与数据计算。在一次用户画像构建工作流程搭建中支持混合模型/数据组件应用，即：可以在一个工作流程中应用多个模型/数据图形组件共同完成一个用户画像的构建，模型/数据图形组件的分析结果具备前后关系，前面模型/数据图形组件的分析结果，是后面模型/数据图形组件数据来源。模型运算与数据计算形成面向一定场域用户画像的场域(标签)数据，最终分门别类的注册在数据资源目录中，并以统一数据访问引擎 DataStore 或接口服务的方式向外提供。全流程、可视化、一站式的模型运算与数据计算让用户画像构建过程像搭积木一样简单灵活，极大的缩短用户、数据、模型/算法与模型构建的距离，真正实现了不同场域、不同层次、不同维度用户画像构建的自定义和组配。

表 3 图书馆用户画像模型运算与数据计算模型

类别	分类	回归	聚类	推荐
子项			划分法	内容法 TF-IDF COSINE 协同法 皮尔森相关系数 余弦相似性 调整余弦相似性 关联规则 效用协同 知识协同
			K-MEANS	
			K-MEDOIDS	
			CLARANS	
			层次法	
			BIRCH	
			CURE	
			CHAMELEON	
			密度法	
			DBSCAN	
			OPTICS	
			DENCLUE	
			网络法	
			STING	
			CLIQUE	

3.3.4 可视化构建与表达

数据可视化的素质与能力直接决定着人们对数据的形态认知、价值挖掘以及呈现等方面的透视与洞察^[51, 52]。数据可视化因其可拓性、可及性以及交互性，逐渐成为增强数据理解、洞悉新兴趋势以及推动数据驱动决策的重要工具。温芳芳等^[53]在多伦多大学图书馆数据可视化工具分类研究成果的基础上，将图书馆数据可视化工具划分为编程语言和 JavaScript 库、图表工具、GIS 可视化工具以及统计分析工具四种。钱爱娟等^[54]使用 D3.js、Tableau、Echarts 和 RAWGraphs 四种可视化工具，以可视化及交互的方式构建图书馆用户画像，直观展示用户基本信息，精准刻画用户行为特征。可视化构建与表达系统完美将大数据图书馆用户画像可视化构建与表达的五个阶段串联在一起，形成集数据、场域、平台、服务四位一体的、完整的业务链条与生态系统。

课题研究的用户画像可视化构建与表达系统具备专业的可视化设计器和 UI 编排功能，选取

EChart、HighChart 为可视化编程组件，内置柱状图、折线图、饼状图、气泡图、仪表盘、雷达图、瀑布图、词云、GIS 地图等多种可视化组件，提供任务管理、可视化设计与 UI 编排、数据加载、可视化验证与预览以及发布功能。描述如下：

①系统支持空白、单值、树表、图表、报表、热词云图、报告与可视展板 8 种可视化表达呈现方式(见表 4)，根据构建用户画像的主体、类型、场域、功用和主题等自行选择任务类型，画像类型、标签、模型、可视化展现方式详见图 5。

②可视化编排设计器集成大量的可视化组件，用户完成任务创建后，通过系统提供的可视化设计和 UI 编排功能对用户画像呈现的页面结构进行设计，页面设计应结合用户画像的表达主题和内容，确定页面的结构层次、业务关系和组件元素。

③应用数据统一访问引擎 DataStore 加载模型运算与数据计算或第三方系统中具有场域特征的数据，实现数据到可视化组件元素的绑定。

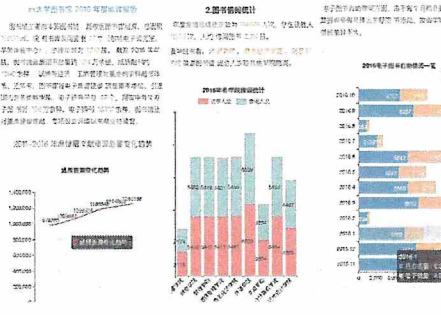

④进行可视化验证与预览，通过系统提供的可视化验证功能验证页面中的可视化组件绑定数据的完整性、规范性、一致性、兼容性以及配置参数和环境参量的准确性；通过在线预览功能验证用户画像可视化表达的场域、逻辑和结构层次。

⑤系统提供独立 Web 页面、可视化组件、API 接口和标准文件交换 4 种发布方式。生成的用户画像不仅能以独立 Web 页面的形式沉淀为图书馆的业务资产，同时也能以可视化组件、API 接口和标准文件交换的方式与其他应用或第三方进行集成，为图书馆更好的实现智慧、精准、快速的服务奠定基础。

表 4 图书馆用户画像可视化呈现方式

序号	呈现形式	组件/元素	描述
1	空白 (接口/服务)		建立空白任务，用户可以按照预设的业务逻辑，组织模型/算法数据，并通过数据/模型可视化引擎以数据接口或服务的方式向第三方系统提供共享数据信息。
2	单值		建立单值任务，表达用户画像中读者、资源、机构等主体身份、等级、数量、偏好、趋势等特征，并以主体图形和单独值域或集合的方式来呈现(示例如左图)。例如：描述读者阅读情况，可以将读者等级划分为：青铜、白银、黄金、钻石等。

序号	呈现形式	组件/元素	描述																																																																																																																
3	树表		<p>建立树表任务，呈现面向某一主体或服务场景的集合数据，数据本身除了具有主体或场景的概要信息以外，还会有主体和场景相关的列表集合。例如：向某一读者提供个性化图书服务时，即可建立树表任务，在其可视化呈现中，不仅包含读者的概要信息（如性别、偏好、用户等级等），还包含通过模型运算后得到的个性化推荐图书列表。</p>																																																																																																																
4	图表		<p>建立图表任务，以可视化图形要素的方式来展现用户画像的内容。可视化要素包含柱状图、折线图、饼状图、气泡图、仪表盘、雷达图、瀑布图、词云、GIS 地图等。图表任务主要应用于维度统计和轻量汇总后的可视化呈现，对于用户画像中模型运算后获得的具有维度的数据，也可使用此方式进行呈现。</p>																																																																																																																
5	报表	<table border="1" data-bbox="504 1279 943 1570"> <thead> <tr> <th>年份</th> <th>课程人数</th> <th>课程总人数</th> <th>学生人数</th> <th>课程完成率</th> <th>课程有效率</th> <th>平均成绩</th> <th>合格率</th> </tr> </thead> <tbody> <tr><td>600</td><td>448</td><td>5924</td><td>1820</td><td>154.02%</td><td>11.65%</td><td>37.91%</td><td>24.62%</td></tr> <tr><td>655</td><td>412</td><td>3009</td><td>1454</td><td>158.90%</td><td>21.33%</td><td>44.91%</td><td>28.34%</td></tr> <tr><td>189</td><td>311</td><td>6774</td><td>1126</td><td>60.77%</td><td>2.79%</td><td>16.79%</td><td>27.62%</td></tr> <tr><td>151</td><td>269</td><td>1054</td><td>903</td><td>96.13%</td><td>14.33%</td><td>16.72%</td><td>29.79%</td></tr> <tr><td>445</td><td>239</td><td>7973</td><td>1319</td><td>171.61%</td><td>5.58%</td><td>23.74%</td><td>19.64%</td></tr> <tr><td>194</td><td>395</td><td>2120</td><td>1700</td><td>90.26%</td><td>9.15%</td><td>11.41%</td><td>22.71%</td></tr> <tr><td>535</td><td>395</td><td>8462</td><td>1346</td><td>136.96%</td><td>6.31%</td><td>39.79%</td><td>28.00%</td></tr> <tr><td>574</td><td>353</td><td>3396</td><td>1722</td><td>162.61%</td><td>15.96%</td><td>33.33%</td><td>20.50%</td></tr> <tr><td>146</td><td>262</td><td>5752</td><td>1639</td><td>96.67%</td><td>2.89%</td><td>9.09%</td><td>15.09%</td></tr> <tr><td>332</td><td>359</td><td>6944</td><td>1409</td><td>94.05%</td><td>4.78%</td><td>23.56%</td><td>25.05%</td></tr> <tr><td>489</td><td>456</td><td>7898</td><td>1037</td><td>107.24%</td><td>6.22%</td><td>47.16%</td><td>43.97%</td></tr> <tr><td>750</td><td>497</td><td>3254</td><td>1148</td><td>150.91%</td><td>23.09%</td><td>55.33%</td><td>42.28%</td></tr> <tr><td>665</td><td>328</td><td>6316</td><td>1336</td><td>202.66%</td><td>10.85%</td><td>50.52%</td><td>24.93%</td></tr> </tbody> </table>	年份	课程人数	课程总人数	学生人数	课程完成率	课程有效率	平均成绩	合格率	600	448	5924	1820	154.02%	11.65%	37.91%	24.62%	655	412	3009	1454	158.90%	21.33%	44.91%	28.34%	189	311	6774	1126	60.77%	2.79%	16.79%	27.62%	151	269	1054	903	96.13%	14.33%	16.72%	29.79%	445	239	7973	1319	171.61%	5.58%	23.74%	19.64%	194	395	2120	1700	90.26%	9.15%	11.41%	22.71%	535	395	8462	1346	136.96%	6.31%	39.79%	28.00%	574	353	3396	1722	162.61%	15.96%	33.33%	20.50%	146	262	5752	1639	96.67%	2.89%	9.09%	15.09%	332	359	6944	1409	94.05%	4.78%	23.56%	25.05%	489	456	7898	1037	107.24%	6.22%	47.16%	43.97%	750	497	3254	1148	150.91%	23.09%	55.33%	42.28%	665	328	6316	1336	202.66%	10.85%	50.52%	24.93%	<p>建立报表任务，以报表的形式展现用户画像的内容。报表呈现方式具有数据结构统一、成果格式固定等特征，一般用于有明确成果要求的用户画像可视化。</p>
年份	课程人数	课程总人数	学生人数	课程完成率	课程有效率	平均成绩	合格率																																																																																																												
600	448	5924	1820	154.02%	11.65%	37.91%	24.62%																																																																																																												
655	412	3009	1454	158.90%	21.33%	44.91%	28.34%																																																																																																												
189	311	6774	1126	60.77%	2.79%	16.79%	27.62%																																																																																																												
151	269	1054	903	96.13%	14.33%	16.72%	29.79%																																																																																																												
445	239	7973	1319	171.61%	5.58%	23.74%	19.64%																																																																																																												
194	395	2120	1700	90.26%	9.15%	11.41%	22.71%																																																																																																												
535	395	8462	1346	136.96%	6.31%	39.79%	28.00%																																																																																																												
574	353	3396	1722	162.61%	15.96%	33.33%	20.50%																																																																																																												
146	262	5752	1639	96.67%	2.89%	9.09%	15.09%																																																																																																												
332	359	6944	1409	94.05%	4.78%	23.56%	25.05%																																																																																																												
489	456	7898	1037	107.24%	6.22%	47.16%	43.97%																																																																																																												
750	497	3254	1148	150.91%	23.09%	55.33%	42.28%																																																																																																												
665	328	6316	1336	202.66%	10.85%	50.52%	24.93%																																																																																																												
6	热词云图		<p>建立词云任务，以热词云图的形式展现用户画像的内容。词云一般用来直观表达某一维度特征，如热度、活跃度、数量级等。图书馆用户画像可视化呈现中，一般用来表示关键词、主题/内容、学科等维度的热度。</p>																																																																																																																

序号	呈现形式	组件/元素	描述
7	报告		<p>建立报告任务，以书面报告的形式对图书馆在一定时间范围内(年、季、月、日)的各个维度信息进行数据抽象与综合呈现。书面报告支持学院报告、图书馆报告和个人报告3类。报告任务在充分挖掘数据的基础上，灵活组配与抽象各维度的特征信息，为学院、图书馆和个人提供整体、综合、宏观的画像呈现。</p>
8	可视展板		<p>建立可视展板任务，以大屏幕展现的形式，综合展示某一画像主题(如：读者、资源、机构、主题等)的不同维度特征。可视展板的呈现形式具有极佳的展示效果与用户体验，用户通过可视展板展示的信息，可直观、清晰了解信息全貌。</p>

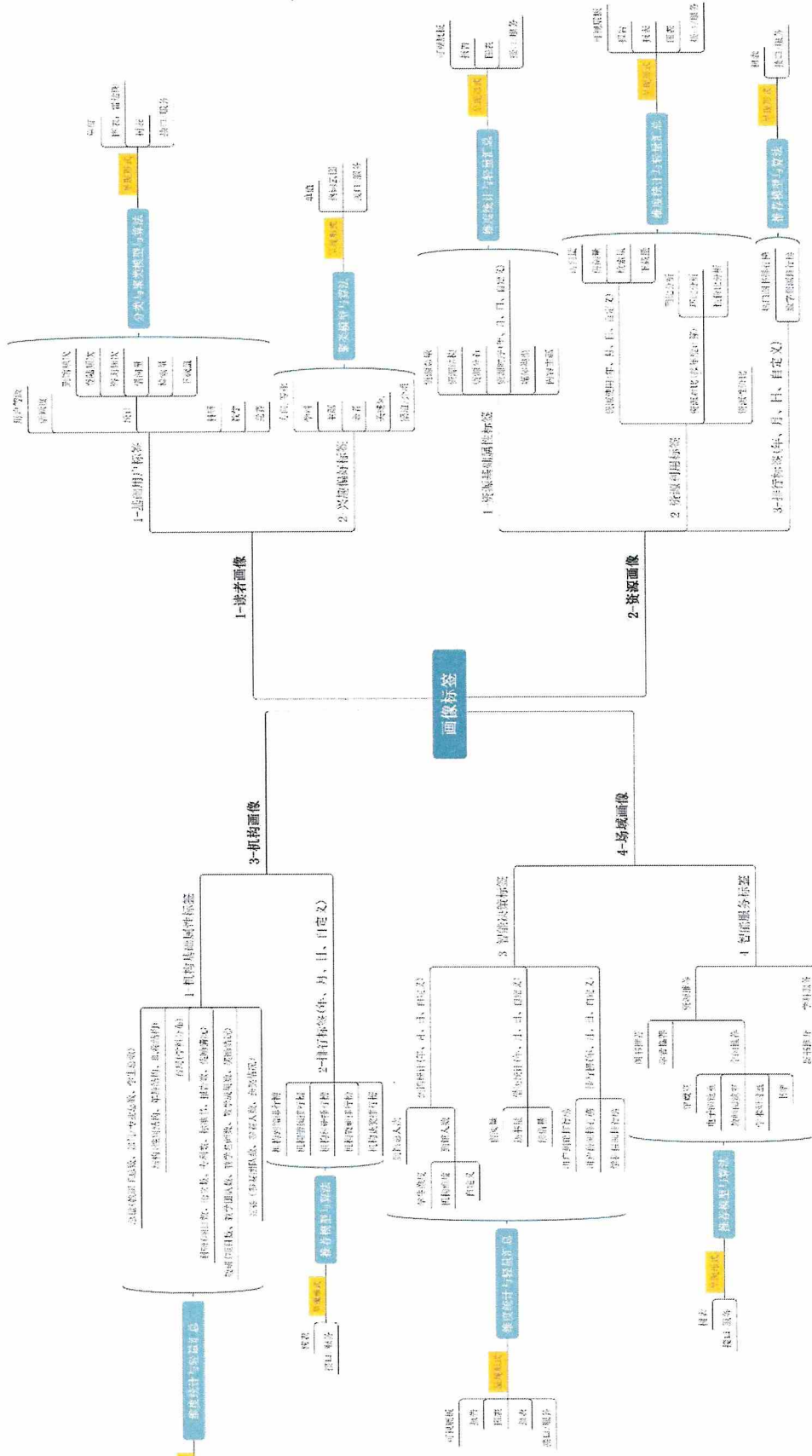


图 5 图书馆用户画像标签/场域-模型-可视化关系对照

第四章 研究结论与展望

信息技术的快速发展给高校智慧图书馆用户画像研究带来了新的发展和新的机遇。用户画像已然成为高校图书馆实现智慧化发展的全新理念、先进技术和场域模型。大数据充分利用其技术优势搭建系统框架、提高运行效率、挖掘用户需求、探索应用场域、构建画像模型、提供精准服务，这为图书馆用户画像研究、探索、应用与实践提供强大的数据空间、技术保障。课题面向高职图书馆，统筹高校全域资源，以图书馆用户画像的可视化构建与表达为研究主题，以 Hadoop 生态、Spark、Kafka、Sqoop 等大数据技术为平台底座，详细探讨了大数据视域下高职图书馆用户画像可视化构建与表达的数据体系、系统框架、构建流程与技术实现；形成了一条全面覆盖数据获取、数据预处理、数据挖掘、标签提取与数据可视化五个阶段紧密完整的业务链条；提出了大数据视域下能够快速支持高校图书馆用户画像可视化构建与表达的一体化解决方案。通过数据综合分析平台实践，提高图书馆用户画像可视化智能构建、表达与精准服务的速度与能力，进一步推进高校智慧图书馆建设和大数据情境下高职数据可视化素养的研究与教育。同时，在课题研究的大数据视域下高职图书馆用户画像可视化构建与表达的解决方案和智能系统中还需在平台云化、网络数据融合、安全策略与授权机制、场景动画以及机器人可视化表达与融合呈现等方面深入研究、持续改进。

第五章 项目成果

[1]付天新. 大数据视域下高校图书馆用户画像可视化构建与表达[J]. 新世纪图书馆 (已投稿)

大数据视域下高校图书馆用户画像可视化构建与表达

付天新 (北京农业职业学院, 北京 102442)

摘要: 随着信息技术的不断发展, 社会智能化进程不断加快, 用户画像被引入到智慧图书馆的建设中, 并已应用到图书馆的智慧化服务领域。本文面向高校图书馆, 统筹高校全域资源, 详细探讨大数据视域下高校图书馆用户画像可视化构建与表达的数据体系、构建流程、系统框架与技术实现, 研究能够快速支持高校图书馆用户画像可视化构建与表达的智能系统, 提高图书馆用户画像可视化智能构建、表达与精准服务的速度与能力, 进一步提高高校智慧图书馆建设水平。

关键字: 大数据 图书馆 用户画像 可视化

Research on Library User Portrait Visualization System from the Perspective of Big Data

FU Tian-xin (Beijing Vocational College of Agriculture, Beijing 102442, China)

Abstract: With the continuous development of information technology and the accelerating process of social intelligence, user portrait has been introduced into the construction of Smart Library and has been applied to the intelligent service field of library. The paper, Facing the university library, plans the overall resources of the University and discusses deeply in the data system, construction process, system framework and technical implementation of the visual construction and expression for the user portrait of the University Library from the perspective of big data, studies and constructs an intelligent system that can quickly support the visual construction and expression of the university library user portrait, and then improves the speed and ability of the visual intelligent construction and expression of the library user portrait, and further improves the construction level of University Smart Library.

Keywords: Big Data; Library; User Portrait; Visual Expression;

【中图法分类号】 G250.7

作者简介: 付天新(1986—), 男, 硕士研究生, 工程师。研究方向: 信息管理系统与空间信息技术。

基金项目: 2021 年度 CALIS 全国农学文献信息中心研究项目 (课题编号 2021008)

1 引言

随着信息技术的不断发展, 社会智能化进程不断加快, 图书馆也逐渐由“数字”迈向“智慧”。智慧图书馆这一概念最早由芬兰奥卢大学图书馆的艾托拉提出, 他认为“智慧图书馆”是一个不受空间限制并且能够随时被感知的移动图书馆^[1]。与传统图书馆注重馆藏、数字图书馆注重读者不同, 智慧图书馆更加注重技术, 结合与运用智能化技术, 全面整合图书馆内、外信息资源, 通过感知、捕捉、记录、挖掘数据、分析读者需求来提供更加智慧的服务^[2]。然而, 在信息技术高速发展的时代下, 图书馆的智慧化服务面临着一系列挑战, 面对丰富的

信息资源和广泛的读者类型, 现有服务模式与不同用户的多样需求之间的矛盾日益突出^[3]。为了更好地服务用户、了解和满足用户需求、提供个性化、精细化服务, 图书馆将用户画像这一理念和技术引入到智慧图书馆的建设中, 并已应用到图书馆的智慧化服务领域。本文以智慧图书馆为研究本体, 以用户画像的可视化表达为研究主题, 结合可视化表达的场域内容, 研究在大数据视域下能够快速支持图书馆用户画像可视化构建与表达的智能系统, 提高用户画像场景化精准服务与智能表达能力, 为图书馆更好的实现智慧、精准、快速的服务与创新奠定基础 and 提供借鉴。

2 图书馆用户画像技术

[2]付天新,许昱苹,刘诚斌,辛力春.国内图书馆用户画像研究现状实证与维度分析[J].河南图书馆学报(已投稿)

国内图书馆用户画像研究现状实证与维度分析

付天新¹,许昱苹²,刘诚斌¹,辛力春¹

(1.北京农业职业学院,北京 102442;2.内蒙古电子信息职业技术学院 呼和浩特 150105)

摘要:随着新时代信息技术的高速发展和图书馆领域用户画像研究的不断深入,国内图书馆用户画像研究内容日渐完备、研究方法推陈出新、研究领域渗透融合、研究生态稳定多样。本文以中国知网数据库为实证库,采用文献计量法和内容分析法,结合科研文献的发文数量、主题、学科、刊源、机构、作者与基金等属性要素,对国内图书馆用户画像研究的内容场域、质量规模、潜力趋势、发展空间、时空特征多个维度进行分析,深度挖掘数据信息,厘清研究现状,进而为国内图书馆用户画像研究的有序发展、深入运用和服务创新提供有力保障。

关键词:图书馆 用户画像 生态群落 文献计量

Empirical and Dimensional Analysis on the Research Status of Library User Portrait in China

FU Tian-xin¹, XU Yu-ping², LIU Cheng-bin¹, XIN Li-chun¹

(1.Beijing Vocational College of Agriculture, Beijing 102442; 2.Inner Mongolia Electronic Information Vocational Technical College, Hohhot 010070)

Abstract: With the rapid development of information technology in the new era and the in-depth research on user portrait in the field of library, the research content of user portrait in domestic library is becoming more and more complete, the research methods are innovating, the research fields are infiltrated and integrated, and the research ecology is stable and diverse. the paper, taking the CNKI database as the empirical database, uses the methods of bibliometrics and content analysis, and combines with the factors such as the number of documents, theme, discipline, journal source, community and fund of scientific research documents, makes an in-depth analysis of the content field, quality scale, potential trend, development space and temporal and spatial characteristics of domestic library user portrait research, and excavates data information, clarify the research status, and then provide a strong guarantee for the orderly development, in-depth application and service innovation of the library user portrait research.

Keywords: library; User portrait; Ecological community; Bibliometrics;

作者简介:付天新(1986—),男,硕士研究生,工程师,研究方向:信息系统管理与空间信息技术。

基金项目:2021、2022年度CALIS全国农学文献信息中心研究项目,2022年度北京农业职业学院科研项目(课题编号:XY-SK-22-07)

1 引言

1998年交互设计之父Alan Cooper首次提出用户画像概念,并指出用户画像是现实生活中真实用户的虚拟代表,它是新时代用户画像技术广泛应用于各个领域的理论基础与实践基石^[1]。2010年郑宝鑫等在广东通信青年论坛中分享的《基于用户画像、信令挖掘技术的手机游戏产品推广》一文中应用用户画像技术进行手机游戏产品推广使得营销

成功率提高了近50倍^[2],这一成功案例不仅让业界同胞为之一振,同时也揭开了“用户画像”神秘的面纱,使得用户画像这一全新的理念和技术成为新时代各个领域争相探索、研究和应用的宠儿。2014年中共江西省委党校图书馆的毕业根在《基于大数据的图书馆信息营销策略》一文中,首次应用用户画像实现用户与资源的精准匹配,开启了大数据时代国内图书馆用户画像研究的大门。至此,掀起了国内将用户画像运用于图书馆研究的热潮^[3]。

参考文献

- [1] Aittola M, Ryhanen T, Smart Library: Location-Aware Mobile Library Service[C]//Proceedings of the International symposium on Human Computer Interaction with Mobile Device sand Services (Mobile HCI). Udine: Springer, 2003: 411-415.
- [2] 李晓敏, 熊回香等. 智慧图书馆中基于用户画像的图书推荐研究[J]. 情报科学, 2021, 39(07): 15-22.
- [3] 谭晓晓. 基于用户画像的图书馆资源个性化推荐研究[D]. 杭州师范大学, 2020.
- [4] 李慧芳. 图书馆资源发现系统用户画像研究——以东南大学图书馆资源发现系统为例[J]. 新世纪图书馆, 2021(03): 38-43.
- [5] 毛玉兰. 基于用户画像的阅读推广服务创新[J]. 大学图书情报学刊, 2020, 38(01): 23-26.
- [6] 张晗, 毕强, 李洁, 丁梦晓. 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, 42(11): 69-74+51.
- [7] 蒋南. 基于用户画像的高校图书馆精准化学科服务模式构建[J]. 图书情报导刊, 2021, 6(08): 17-24.
- [8] 孔繁超. 基于数字孪生技术的智慧图书馆空间重构研究[J]. 情报理论与实践, 2020, 43(08): 146-151.
- [9] ATOTE B S, SAINI T S, BEDEKAR M, et al. Inferring emotional state of a user by user profiling[C]//Proceedings of the 2016 2nd international conference on contemporary computing and informatics. Noida: IEEE, 2016: 530-535.
- [10] 郭宇, 孙振兴, 刘文晴, 于文倩. 基于数据驱动的移动图书馆 UGC 用户画像研究[J]. 情报理论与实践, 2022, 45(01): 30-37.
- [11] (美) 库珀, 等. 软件观念革命: 交互设计精髓 [M]. 詹剑锋, 等, 译. 北京: 电子工业出版社, 2005: 59.
- [12] Idoughi D, et al. Adding user experience into the interactive service design loop: a persona-based approach [J]. Behaviour & Information Technology, 2012, 31(3): 287-303.
- [13] An J, et al. Validating social media data for automatic persona generation [C]//2016 IEEE/ ACS 13th International Conference of Computer Systems and Applications (AICCSA). IEEE, 2016: 1-6.
- [14] AMATO G, STRACCIA U. User profile modeling and applications to digital libraries[C]// Research and Advanced Technology for Digital Libraries, Third European Conference, ECDL'99. Heidelberg: Springer, 1999.
- [15] 吴智勤, 李萍. 大数据情境下高校图书馆科研用户画像构建策略研究[J]. 江苏理工学院学报, 2019, 25(06): 117-121.
- [16] 朱静, 徐一然, 吴颖, 孙叙伦, 任娟. 基于大数据的高校图书馆用户画像研究[J]. 科技风, 2021(31): 74-77.
- [17] 兰冰. 基于群组用户画像的图书馆信息智能推送服务[J]. 图书馆学刊, 2019, 41(07): 109-112.

- [18]宋美琦,陈焯,张瑞.用户画像研究述评[J].情报科学,2019,37(04):171-177.
- [19]余传明,田鑫,郭亚静,安璐.基于行为-内容融合模型的用户画像研究[J].图书情报工作,2018,62(13):54-63.
- [20]王丽艳,郭春侠.图书馆用户画像研究现状及趋势[J].图书馆界,2021(02):65-70+94.
- [21]陈慧香,陈婷.高校网络舆情用户画像构建与应用研究[J].长江工程职业技术学院学报,2021,38(04):58-61.
- [22]胡媛,毛宁.基于用户画像的数字图书馆知识社区用户模型构建[J].图书馆理论与实践,2017(04):82-85.
- [23]张洁,仲跻亮,岳怡然,寇远涛.数字图书馆用户画像建模与应用实践[J].数字图书馆论坛,2020(03):51-58.
- [24]刘海鸥,李凯,姜波.移动图书馆推荐系统中的用户画像与资源画像情境化融合研究[J].图书馆,2021(06):66-71+93.
- [25]徐冉.面向机构画像的高校图书馆嵌入式学科服务体系研究[J].图书馆界,2022(01):7-10.
- [26]陈丹,柳益君,罗焯,钱秀芳,吴智勤.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(06):72-78.
- [27]何娟.基于用户个人及群体画像相结合的图书个性化推荐应用研究[J].情报理论与实践,2019,42(01):129-133+160.
- [28]王丽艳,郭春侠.图书馆用户画像研究现状及趋势[J].图书馆界,2021(02):65-70+94
- [29]范晓玉,窦永香,赵捧未,周潇.融合多源数据的科研人员画像构建方法研究[J].图书情报工作,2018,62(15):31-40.
- [30]姚远.基于本体的用户画像构建方法.中国计算机用户协会网络应用分会.中国计算机用户协会网络应用分会2018年第二十二届网络新技术与应用年会论文集[C]//中国计算机用户协会网络应用分会:北京联合大学北京市信息服务工程重点实验室,2018:7.
- [31]刘漫.基于用户画像的高校图书馆阅读推广模式构建[EB/OL].<https://doi.org/10.14064/j.cnki.issn1005-8214>.
- [32]李晓敏,熊回香,杜瑾,景紫薇.智慧图书馆中基于用户画像的图书推荐研究[J].情报科学,2021,39(07):15-22.
- [33]朱静,徐一然,吴颖,孙叙伦,任娟.基于大数据的高校图书馆用户画像研究[J].科技风,2021(31):74-77.
- [34]许鹏程,毕强,张晗,牟冬梅.数据驱动下数字图书馆用户画像模型构建[J].图书情报工作,2019,63(03):30-37.
- [35]刘海鸥,孙晶晶,陈晶,张亚明.用户画像模型及其在图书馆领域中的应用[J].图书馆理论与实践,2018(10):92-97.
- [36]朱会华,肖海清,梁红烂.基于用户画像的智能荐购模式研究[J].图书馆学研究,2020(07):34-38.
- [37]何胜,冯新翎,武群辉,熊太纯,李仁璞.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,2017,61(01):40-46.
- [38]胡媛,毛宁.基于用户画像的数字图书馆知识社区用户模型构建[J].图书馆理论与实践,2017(04):82-85+97.
- [39]陈添源.高校移动图书馆用户画像构建实证[J].图书情报工作,2018,62(07):38-46.

- [40] 于兴尚, 王迎胜. 面向精准化服务的图书馆用户画像模型构建[J]. 图书情报工作, 2019, 63(22):41-48.
- [41] 梁欣, 王晓辉. 基于 Hadoop 的数字图书馆用户画像系统研究[J]. 情报探索, 2019(11):6-12.
- [45] 赵杨, 杨彬, 董姝仪, 王森. 多源大数据驱动的移动图书馆个性化推荐系统设计与实现[J]. 图书馆学研究, 2021(11):20-31.
- [42] 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.
- [43] 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.
- [44] 宋建玮. 基于 AISAS 模型的公共图书馆研学旅行服务路径与策略研究[J]. 图书情报导刊, 2021, 6(12):1-6.
- [45] 周莉. 基于区块链技术的移动图书馆用户画像数据管理策略研究[J]. 图书馆工作与研究, 2021(07):49-57.
- [46] 尹婷婷, 曾宪玉. 用户画像技术在高校数字图书馆信息服务中的研究与应用[J]. 图书馆理论与实践, 2021(06):106-111.
- [47] 宫兆阳. 基于运行体征分析的图书馆数据服务创新研究[J]. 大学图书情报学刊, 2020, 38(05):75-80.
- [48] 卢思佳, 王凤姣. 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, 39(05):55-62+86.
- [49] 蒋南. 基于用户画像的高校图书馆精准化学科服务模式构建[J]. 图书情报导刊, 2021, 6(08):17-24.
- [50] 莫君兰, 窦永香, 开庆. 基于多源异构数据的科研团队画像的构建[J]. 情报理论与实践, 2020, 43(09):100-106.
- [51] 孟祥保, 常娥, 叶兰. 数据素养研究:源起、现状与展望[J]. 中国图书馆学报, 2016, 42(02):109-126.
- [52] 张晨. 大数据时代的图书馆与数据素养教育[J]. 图书与情报, 2014(04):117-119.
- [53] 温芳芳, 冯玲玲, 王春迎, 曹芬芳, 李斯, 魏银珍. 国外一流大学图书馆数据可视化服务实践与启示[J]. 大学图书馆学报, 2021, 39(06):78-86.
- [54] 钱爱娟, 董笑菊, 沈绮文, 李晨璐, 施晓华. 高校图书馆用户画像与行为可视化分析[J]. 图书馆杂志, 2020, 39(10):82-88.