

# 高校图书馆学科服务小数据系统的构建与优化<sup>\*</sup>

张凤斌

(东北林业大学图书馆 哈尔滨 150040)

**〔摘要〕** 小数据思维与高校图书馆学科服务“以人为本”理念高度契合,小数据应用系统以个性化特征为导向,推动学科服务从普适性服务向个性化、精准化服务转变。本文在分析学科服务用户类型及信息需求特点的基础上,研究探讨用户小数据构成及特征,在实践层面上探索了高校图书馆学科服务小数据系统的构建与优化。

**〔关键词〕** 高校图书馆 小数据 学科服务 系统构建

**〔中图法分类号〕** G250.7

**〔引用本文格式〕** 张凤斌.高校图书馆学科服务小数据系统的构建与优化[J].图书馆,2021(11):64—68.

## 1 引言

近年来,“小数据”应用成为国内图书馆界的一个研究热点。本文通过“中国知网”以主题词“小数据+图书馆”进行检索,从2015年1月到2020年12月,共计获得76篇文献,经过筛查后获得相关文献52篇。通过梳理分析发现,研究主题和内容主要集中于理论研究方面,如小数据概念及内涵<sup>[1-2]</sup>、图书馆小数据定义及内涵<sup>[3-4]</sup>、小数据应用策略<sup>[5-6]</sup>、小数据应用服务模式<sup>[7-8]</sup>等;实践研究方面比较薄弱,缺乏对小数据应用涉及的实际问题及解决方案的深入探讨,如小数据在图书馆中的技术角色定位、小数据在技术层面与图书馆现有技术系统的衔接等。从总体情况来看,目前还没有形成比较成熟的理论体系和可以普遍推广的应用模式。

小数据是在大数据环境下衍生出来的一类新兴数据,是以个人为中心的全方位数据集,具有鲜明的个体独特性,也称之为“量化自我(Quantified Self)”。美国学者德波哈尔·艾斯汀(D. Estrin)在2013年首次提出小数据概念,他认为小数据是“全部关于‘我’的数据(small data where n=me)”<sup>[9]</sup>,“个体用户的小数据是其日常行为活动

的全部表征”<sup>[10]</sup>,他指出“从个体用户数据中提取出的具有个性化特征的数据将有效地揭示个体用户行为模式规律”<sup>[11]</sup>。由此可见,小数据思维与图书馆“以人为本”服务理念高度契合,这也是小数据应用成为图书馆界研究热点的原因所在。但是目前研究成果主要集中于理论层面,在实践层面探索较少。

小数据的产生源于实践,具有很强的实践性,小数据研究应从图书馆业务工作中寻找切入点,针对某项实际业务工作把理论知识与实践研究相结合,探索构建符合实际业务需要的、具有可操作性的小数据应用系统。高校图书馆学科服务工作是一项核心业务,在一定程度上代表了图书馆服务能力和服务水平。目前学科服务存在的主要缺陷是服务精准程度不足,主要原因在于用户提出的需求是显性需求,只是实际需求的一部分;而另一部分是用户没有表达出来的隐性需求,这种潜在需求与其学习工作经历、专业知识结构等因素密切相关。在传统思维下,个人隐性需求存在于其潜意识中,外界无法识别。但是在小数据思维下,这个难题具有破解的可能性,即通过个人行为数据发现和预测个人需求特征,使隐性需求显性化,从而为学科服务实现精准化创造前提条件。

<sup>\*</sup> 本文系黑龙江省哲学社会科学研究规划项目“面向新型智库的高校图书馆信息服务策略研究”(项目编号:18TQD369)、黑龙江省统计科学研究项目“高校科研大数据管理与共享平台的建设策略研究”(项目编号:2020B07)研究成果之一。

2 高校图书馆学科服务用户类型及小数据构成分析

2.1 学科服务用户类型分析

高校图书馆学科服务用户是在校师生，从群体特征来看，可以划分为三个类型：教学型用户、科研型用户和学习型用户。因为学科用户类型和层次存在差异，其信息需求和服务需求方面也呈现出不同特点。

(1)教学型用户：教师以授课为主，在教案准备和课程教学中，需要与教学相关的学科资源导航、课程建设相关资源及精品课件推荐、在线教学、学科最新信息推送，以及教参服务、教学支持服务、馆际互借、文献传递服务等。

(2)科研型用户：科研人员需要把握学科前沿发展态势，跟踪学科研究动态，了解借鉴他人研究成果、先进方法和技术，包括文献调研、查新查引、文献传递、资源推荐、专利分析、深度学科咨询服务、深层次学科信息加工服务等，以及科研成果发表、鉴定评价和获奖等信息。

(3)学习型用户：在校学生任务是学习专业知识和完成毕业论文，本科生的需求主要在学习支持方面，包括教学参考书、工具书、考级考证辅导资料，以及信息素养培训、在线课程和在线阅读等；研究生则对专业期刊、学位论文及投稿指导等的需求比较多。

2.2 高校图书馆学科服务用户小数据构成分析

小数据应用具有明确的目的性，首先需要明确有待解决的实际问题，其次确定需要哪些数据支撑。就高校图书馆学科服务而言，小数据应用是为了解决准确发现用户信息需求的问题，其相关数据蕴含在学习、教学、科研利用的各类信息管理系统中。通过文献梳理与分析，学科服务用户数据构成主要包括两个方面：基本特征数据、各类行为特征数据，其具体内容如表 1 所示。

3 高校图书馆学科服务小数据系统构建

3.1 学科服务小数据系统开发应用的价值

小数据系统是围绕个体的全方位数据及其配套的收集、处理、分析和对外交互的综合系统<sup>[12]</sup>，通过采集、分析学科用户个人特征数据，动态地识别学科用户个体的显、隐性信息需求，在此基础上提供与学科用户需求匹配的服务和资源，实现个性化、精准化的学科服务目标。进一步而言，高校图书馆学科服务小数据系统是大数据环境下的一

表 1 高校图书馆学科服务用户小数据构成

数据类型	数据范围	数据来源
基本特征数据	年龄、性别、院系、专业、学历、学位、职称、学习年级、工作年限等	教学信息管理系统 科研信息管理系统 学生信息管理系统 图书馆信息管理系统 移动图书馆 App 等
教学行为数据	教学科目、课时数、学生人数、在线教学情况、培养研究生情况、图书期刊借阅与荐购、教学服务咨询、文献传递、在线阅读、访问图书馆网页及移动图书馆 App 等	
科研行为数据	发表论文及被引情况、完成课题及获奖情况、在研课题及论文、申请课题、查新查重、科研服务咨询、文献传递、在线阅读、访问图书馆网页及移动图书馆 App 等	
学习行为数据	课堂学习、在线学习与阅读等、培训讲座、图书期刊借阅、参与阅读推广、访问图书馆网页及移动图书馆 App 等	
社交行为数据	参加学术会议、学术博客、学术论坛、参加课堂展示与演讲、学习小组内部交流、课外活动、成果展示、访问社交网站等	

个子系统，设计开发的根本目的是更加精准地描绘、预测与分析用户需求，同时根据不同阶段用户行为的变化，通过聚类分析、协同过滤等技术手段，不断调整推荐、决策参考方案，进而实现为用户提供个性化强、精准度高的服务效能。此外，小数据系统是以“机器为主，人力为辅”的运行模式，各种应用系统和智能设备成为数据采集、存储、传输和处理的主体，人力只在模型设计、参数设置、编辑矫正等环节发挥作用。因而，学科服务小数据系统将提高学科服务的精准度，并减轻学科馆员在学科服务中的工作强度。

3.2 学科服务小数据系统的基础组织结构

(1)数据采集、数据预处理、数据存储。这一过程主要是利用 Hadoop2.0 分布式平台的数据提取、转换与加载的 ETL (Extract-Transform-Load) 工具与 Hive 数据仓储工具实现操作处理。其中，Hadoop2.0 是 Apache 软件基金会基于 Java 语言开发的开源分布式计算平台，主要包括数据存储 HDFS、作业调度与资源管理框架 YARN、并行处理数据计算框架 MapReduce。而 Hive 是 Hadoop 的一个数据仓储工具，可将结构化数据文件映射为数据库表并存储到 HDFS 上，并提供了类似 SQL 查询语言，可把 SQL 语句转换为 MapReduce 程序运行。

数据采集是基于 Hadoop2.0 的学科服务小数据应用系统通过 API 接口或 Heritrix 网络爬虫采集获取用户的注册信息、使用日志、咨询交互等数据，依据采集路径可分为静态数据采集和动态数据采集：①静态数据是存储在高校各类信息管理系统中的个人信息，如基本特征数据（年龄、专

业、学历等)、教学行为数据(教学科目、课时数、学生人数等)、科研行为数据(发表论文情况、科研项目情况等)。(2)动态数据是高校图书馆学科服务用户与图书馆的一切关联行为数据,通过图书馆信息集成管理系统提取出来,如图书期刊借阅、图书馆网页查询、查新查引、文献传递、访问移动图书馆 App 等。

数据预处理、数据存储是基于 Hadoop2.0 的存储工具 HDFS 联合 HBase、DataWrangler 等对采集的数据进行分类、去重、标引、关联等,进而去除无关数据、平滑噪声、填补遗漏数据,实现数据标准化、规范化,将其集成归约并分布式存储于用户属性库、用户行为库,或者根据用户画像进行用户需求分析,抽取特定资源存储到基础资源库。

(2)标签模型建立、数据挖掘分析、画像模型建立与质量评估:①标签模型建立,在数据存储的基础上结合数据的分类特征建立用户、资源的标签库,构建标签体系,实现用户与资源特征标签化。②数据挖掘分析,运用 Hadoop2.0+Mahout+RapidMiner 数据挖掘工具进行分类、聚类、回归、关联规则等多算法多维度分析用户需求、行为与资源标签体系,建立相互关联的基础资源画像与用户画像模型。③画像模型建立与质量评估,画像建模需要根据相关数据进行质量评估,不断反馈、修正、优化与完善。而画像质量评估包括服务收益率  $P_{ser}$ 、用户满意度  $D_{sat}$  两个指标,其中  $P_{ser}$  权重比例为  $r$ ,  $D_{sat}$  权重比例为  $1-r$ ,那么评估结果是  $W=rP_{ser}+(1-r)D_{sat}$ ,  $0<W\leq 100$ <sup>[13]</sup>。

### 3.3 学科服务小数据应用系统的主要功能

(1)个性化检索与推荐:①个性化检索,借助用户画像提供的信息需求、偏爱主题数据进行挖掘与可视化分析,为个体量身定制检索式以实现用户个性化检索。在检索结果排序与反馈上,根据用户画像对需求度的预判,实现检索结果按相关度从大到小降序排列,并在每个检索结果页面设有“需要”“一般”“不需要”的快速选择反馈模块,动态修正用户需求度,以持续优化服务效果。②个性化推荐,通过画像模型完善 Mahout 协同过滤与推荐功能,解决评分难、稀疏性与冷启动的问题,实现个性化推荐。用户画像深度挖掘分析、动态预测出用户需求、兴趣习惯,资源画像对学科资源进行语义化、标签化描述实现内容特征提取。两类画像相互匹配、关联映射,辅以应用情境、协同等要素特征,优化完善子系统精准推荐性能,并构建了深度学习召回算法策略,有效提高了用户服务满意度和黏合度。

(2)精准营销:①用户分析与群组划分,追踪并采用数据挖掘工具对用户行为、资源特征标签及画像数据进行多维度分析,实现群体需求差异化分类,把有共同背景、目的、兴趣习惯的学者、专家与科研团队精细划分,形成不同的领域和主题组别,并对群组内成员的参与度和活跃程度进行分析。②服务精准推送及反馈,将学科服务产品或宣传内容传递给指定用户或群体,避免有关信息对其他用户造成困扰。同时注重用户评价与反馈意见,实现“传递—

反馈—再传递—再反馈”的交互营销过程,动态优化用户画像,并融合移动化、社交化与本地化 SoLoMo 服务内涵,发挥营销效应,深化图书馆服务策略,树立图书馆优质服务品牌。

(3)决策参考:①图书馆资源与服务,决策参考子系统对学科用户、资源、业务服务等产生的数据进行全面融合、画像智慧分析,揭示发现隐藏在数据背后的模式和规律,形成大量数据分析报告,可为图书馆进行资源采购配置、业务类型调整、规章制度等提供决策依据。②用户认知,基于用户画像对用户行为需求、兴趣爱好等主题标签进行长期跟踪、变化分析解读,可为学科用

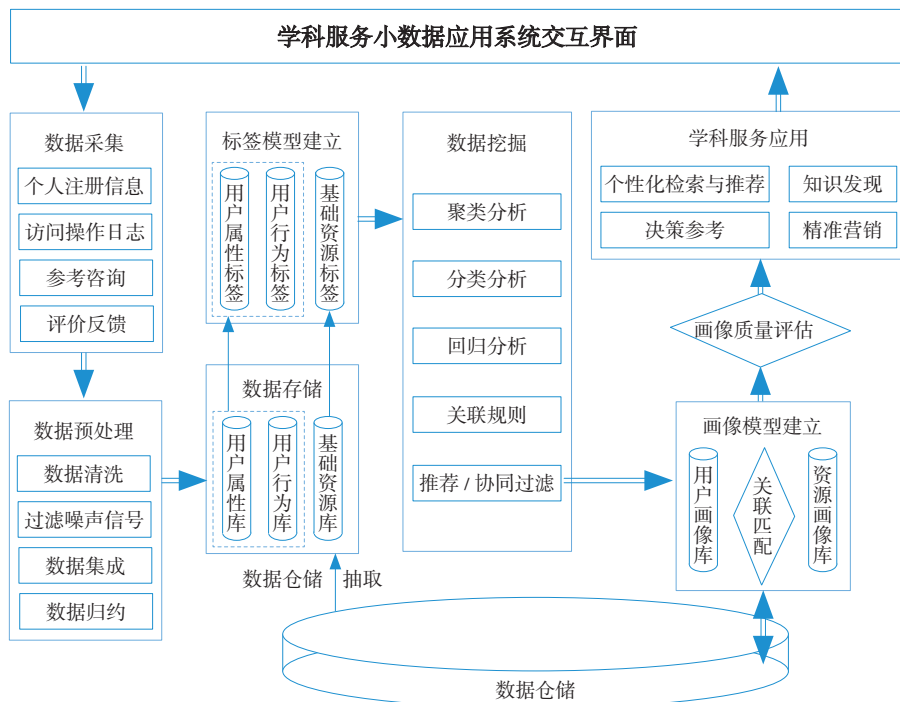


图1 学科服务小数据应用系统与组织结构图



户研究、教学、学习的发展方向提供参考依据,丰富用户自我认知途径。

(4)知识发现:①数据准备阶段是数据采集、数据预处理、数据存储环节。②数据挖掘阶段,运用 Mahout、RapidMiner 数据挖掘系统以及 SAS、SPSS、Matlab 等统计与可视化工具,根据任务选择合适算法对数据进行科学分析与知识发现。③结果评价与解释阶段,利用画像模型分析用户知识需求,关联匹配去除无关和冗余内容,将所需成果展示给学科用户。

## 4 高校图书馆学科服务小数据系统的优化

### 4.1 精确定位用户,建立用户需求模型

高校图书馆学科服务的逻辑起点是用户信息需求,准确识别用户信息需求是实现精准化学科服务的前提条件。目前高校图书馆在满足学科服务用户信息需求方面主要存在两个问题:一是面向用户群体的普适性服务,难以满足学科用户个性化的信息需求;二是仅能识别用户提出的显性信息需求,对于用户未表述出来的隐性信息需求很难识别,用户信息需求识别不完整,导致学科服务精准程度不高。

针对上述存在的问题,小数据应用系统从以下两个方面提供解决方案:①精确定位个体用户。通过小数据应用系统用户入口进行注册,个体用户即可获取系统唯一定位标记,同时授权高校图书馆采集其个人特征数据;②建立用户信息需求模型。小数据应用系统通过个人特征数据分析发现其隐性信息需求,从而使其转化为显性信息需求,使用户全部信息需求完整呈现出来。

### 4.2 强化技术导向,提升学科馆员工作效能

学科馆员是高校图书馆学科服务工作的主体,参与学科服务工作的各个环节,不仅包括分析用户信息需求、制定服务策略、为用户提供服务等,还要负责对口院系联络、学科服务平台维护、学科信息资源建设等。由此可见,学科馆员既要具有某个专业的学科知识背景又要精通图书馆业务。目前高校图书馆学科服务普遍存在学科馆员数量配置不足的问题,由于人力资源成本限制,只有少数高校图书馆能够做到按照院系配置学科馆员,大多数高校图书馆学科馆员都需兼顾两个或以上的院系的学科服务工作。由此可见,学科馆员面临的工作任务比较繁重,导致工作效能降低。

针对上述存在的问题,从强化技术导向出发,可以采

取的解决方案包括两个部分:①小数据应用系统。通过用户个体特征数据整合、分析,发现和预测其个性化信息需求特征,为精准化学科服务提供依据;②智能化学科服务平台。根据小数据应用系统提供的用户个体信息需求,自动进行信息资源挖掘与整合,然后精准推送给用户。上述解决方案的最大特点是以“机器为主,人力为辅”,各应用系统和智能设备成为数据采集、存储、传输和处理的主体,人力只在模型设计、参数设置、编辑矫正等环节发挥作用。学科馆员可以充分利用数据技术发展成果,融入学科服务的业务流程中,提高学科服务工作的智能化、自动化水平,从而提升工作效能。

### 4.3 优化学科服务平台功能,为用户提供个性化智能服务

目前国内多数高校图书馆建立了学科服务平台,但对其承担的任务与实现功能的定位尚不够清晰,大部分以信息资源组织与揭示为主体,以学科信息资源导航方式为依托,虽然具有一定的体系与功能,但缺乏精准的用户划分,没有将用户的知识背景、专业与信息行为、阶段性需求等融为一体作为信息需求参照。小数据应用系统就是为了准确发现用户信息需求,解决学科服务精准化程度低、针对性不强的问题。

针对上述存在的问题,从优化学科服务平台功能出发,小数据应用系统的优势体现在:①个性化检索与推荐。通过引入用户画像模型、推荐技术,构建学科服务智能个性化检索与推荐系统,过滤大数据通用检索系统相同处理过程和无差别结果所包括的与需求无关的信息;②精准营销。通过运用 SoLoMo 的社交网络精准营销服务模块设计,实现用户分析与群组划分、服务精准推送及反馈,解决当前学科服务平台专指度低、精准性差的问题;③决策参考。通过用户行为特征标签化、可视化、具象化呈现个性化的用户需求,挖掘资源与用户两类画像数据以指导图书馆资源、服务发展方向与用户认知路径,可对学科资源质量、服务内容评估提供定量到定性的验证依据;④知识发现。通过数据准备、数据挖掘与结果评价揭示隐性科研知识,应用数据挖掘算法按特定方式和阈值挖掘抽取有价值的知识和模式,并进行评价和解释<sup>[14]</sup>。

## 5 结语

随着时代发展和技术进步,高校图书馆学科服务理念也在不断演变,从以图书馆为中心到以用户为中心,从文

献服务到知识服务,从普适服务到精准服务。高校图书馆精准化学科服务是“在普适服务基础上面向用户个性化需求而建立的以用户问题为导向的服务模式”<sup>[15]</sup>,学科服务实现精准服务需要“准确识别和定位用户,并紧贴用户特征和环境,更细致而动态地识别用户显性、隐性需求,提供对接需求的更优质服务和资源配置方案”<sup>[16]</sup>。

小数据思维与学科服务精准化理念高度契合,本文选择学科服务作为小数据应用实践研究的切入点,在技术层面以用户数据为核心元素,探索构建具有收集、处理、分析和对外交互功能的小数据应用系统。从技术角度来看,在高校图书馆现有技术设施和信息资源基础上,学科服务小数据应用系统具有可行性和可操作性;从实施角度来看,主要取决于两个因素,一是资金投入和技术人员配备,二是需要整合分散在高校不同管理部门的用户数据。由此可见,学科服务小数据系统应该纳入高校建设发展规划,在高校统筹协调下分阶段逐步实施。

( 来稿时间:2021 年 4 月 )

## 参考文献:

1. 马晓亭,陈臣. 基于可信小数据的图书馆个性化服务研究[J]. 图书情报工作, 2015, 59 (4):70-75.
2. 陈臣. 基于小数据的图书馆个性化服务 QOS 保证研究[J]. 图书馆, 2015 (11):32-35.
3. 陈臣. 基于小数据决策支持的图书馆个性化服务[J]. 图书与情报, 2015 (1):82-86.
4. 陈廉芳. 大数据环境下图书馆用户小数据的采集、分析与应用[J]. 国家图书馆学报, 2016, 25 (3):69-74.
5. 王昊. 基于小数据的图书馆个性化服务研究[J]. 图书馆学报, 2016, 38 (6):102-104.
6. 刘庆麟. 基于小数据的图书馆精准服务研究[J]. 图书馆工作与研究, 2017 (5):45-50.
7. 王敏. 小数据思维驱动下的公共图书馆信息服务模式研究[J]. 图书馆学报, 2015, 37 (12):77-79.
8. 刁羽. 基于小数据的高校图书馆智库型信息咨询服务模式研究[J]. 图书馆工作与研究, 2019 (8):82-86.
9. Estrin D. Small data, where n=me[J]. Communications of the ACM, 2014, 57 (4):32-34.
10. Estrin D. Thompson C W. Internet of you: data big and small[J]. IEEE Internet Computing, 2015, 19 (6):8-10.
11. Estrin D. Small, n=me, data [EB/OL]. [2020-09-15]. [https://www.nsf.gov/events/event\\_summ.jsp?cntn\\_id=129138&org=CISE](https://www.nsf.gov/events/event_summ.jsp?cntn_id=129138&org=CISE).
12. 陈辉. 小数据之美:精准捕捉未来的商业小趋势[M]. 北京: 中信出版社, 2019:4.
13. 王欣, 张冬梅. 大数据环境下基于高校读者小数据的图书馆个性化智能服务研究[J]. 情报理论与实践, 2018, 41 (2):132-137.
14. 章新友. 药学文献检索:新世纪第2版[M]. 北京: 中国医药出版社, 2017:195.
15. 牛勇. 图书馆精准服务研究[J]. 图书馆学研究, 2016 (5):50-52.
16. 张汝昊, 杨志萍. 国内外学术图书馆精准服务前沿研究综述[J]. 图书馆学研究, 2019 (24):2-12.

## Construction and Optimization of Subject Service Small Data System in University Library

Zhang Fengbin

( Library of Northeast Forestry University )

**[Abstract]** Small data thinking is highly compatible with the “people-oriented” concept of university library subject services, and the small data application system is guided by individual characteristics to promote the transformation of subject services from universal services to personalized and accurate services. Based on the analysis of the types of subject service users and the characteristics of information needs, this paper studies the composition and characteristics of user small data, and explores the construction and optimization of subject service small data system in university libraries at the practical level.

**[Keywords]** University library Small data Subject service System construction

**[作者简介]** 张凤斌(1969—),男,博士,东北林业大学图书馆馆长,副研究馆员,研究方向:智慧图书馆、信息系统构建与知识发现。