

# 用户画像在图书馆信息服务中的应用

贾蔷

(首都图书馆, 北京 100021)

**摘要:** 随着大数据技术的应用及用户需求的多元化, 图书馆信息服务逐渐向精准方向发展, 重视通过用户和资源数据的分析, 了解用户需求, 拓展服务模式。本文介绍用户画像的概念, 探讨构建用户画像的方法与手段, 并对构建过程中的数据来源、用户数据的分析与处理、数据标签和模型构建过程深入阐述, 最后对用户画像在信息服务中的作用和后续发展提出思考。

**关键词:** 图书馆; 用户画像; 信息资源; 数据标签; 模型构建

**中图分类号:** G252

**DOI:** 10.3772/j.issn.1673-2286.2019.07.009

近十年, 我国图书馆事业处于快速发展中。在大规模兴建各级图书馆的同时, 图书馆也在认真思考未来发展的方向。图书馆承担以满足知识信息资源需求为基础的信息服

务, 如何提供更为精准与定制化的服务, 成为业界日益关注的焦点。随着大数据技术的应用及用户需求的多元化, 图书馆服务逐渐向集中、开放的网络平台转移, 对用户和资源数据的分析为关注重点, 了解用户需求, 采用现代信息技术手段, 拓展新型服务模式, 从而满足用户精细、精准和个性化需求, 有助于信息服务价值的体现。这就为大规模收集用户数据、开展以用户为中心的信息服

## 1 用户画像的概念

Cooper最早提出用户画像的概念, 指出用户影像是真实用户的虚拟代表, 是建立在一系列真实数据上的目标用户模型<sup>[1]</sup>。Travis<sup>[2]</sup>更进一步提出用户画像的7个条件, 即基本性、移情性、真实性、独特性、目标性、数量、应用性。

用户影像是用户基本属性、生活习惯和消费行为等因素信息, 通过信息识别, 完成贴“标签”的核心工作。标签通常是精练的特征标识(如年龄、性别、兴趣等), 将所有标签集合起来, 可以勾勒出该用户的“画像”<sup>[3]</sup>。

用户画像最先应用于电子商务和门户平台, 如阿里巴巴、京东、百度等科技公司已在多年前就开始着手用户画像的建构, 基于已有信息数据, 分析用户关注点、兴趣点, 推测用户潜在需求, 从而制定精准、优化的营销策略, 对推送个性化服务产生了积极作用, 引领与带动潜在消费、拓展用户, 亦收到很好的经济效益。

查询CNKI可知, 截至2018年12月, 共检索到相关论文205篇, 学科分布集中在计算机、工商管理、商业经济等领域。图书档案情报学科共计43篇, 其中2014年李业根<sup>[4]</sup>在《图书馆学刊》发表的《基于大数据的图书馆信息营销策略》, 标志着用户画像的到来。纵览图书档案情报学科用户画像相关论文, 发现研究多集中于资源推荐、知识发现和数据模型与算法等内容, 而用户画像在图书馆服务应用中所能达到的深度与广度尚在探讨与摸索阶段。

## 2 图书馆用户画像的构建

在大数据全面推广的社会背景下, 图书馆所拥有的

信息数据显现出多维度、多层次交叉重叠的特性,其中个体信息数据和行为信息数据尤为重要。用户画像是基于图书馆用户信息数据分析平台,运用统计分析与数据挖掘等技术集群,完成各类异构系统的数据挖掘技术。通过对信息数据的挖掘、分析及显现,精准勾画出用户需求画像,从而提供精准化、定制化的信息服务。

用户画像是在海量数据基础上产生的,通过确认数据来源、用户数据的分析与处理、用户画像数据标签和用户画像模型构建等过程,对个体用户和用户群准确地标签化,从而细致描绘出用户特征。

## 2.1 获取异构系统中用户数据,建立关联关系

作为构建用户画像的基本要素,图书馆需要大量的可用于数据处理的基础数据,这些基础数据可以最大限度地显现个体用户信息、行为信息;基础数据的全面获取,可以提高用户画像的精准度。

用户画像数据被保存在不同的异构系统中,数据之间没有融合甚至没有关联,且数据获取时需要从不同系统中逐项单独获取,整个获取过程的进度与系统集群的规模有直接关系。获取的数据来源可分为基本数据和变化数据。基本数据相对稳定,主要由用户自身数据组成,包括用户基本信息,如性别、年龄、职业、联系方式和地址等,其更新点大多是联系方式和地址。

变化数据是相对复杂的用户行为痕迹信息,行为日志数据、兴趣数据、交互数据等都是变化数据,其来自用户在文献利用时产生的使用数据、获取知识库的交互数据、移动端的访问数据,甚至用户的登录次数、登录设备、页面浏览时长、页面点击位置信息和页面复制粘贴内容等<sup>[5]</sup>,这些数据会随着用户环境、焦点兴趣、社会趋势等不断变化。

基本数据与变化数据分别从各种异构系统中采集出来,借助统一认证和用户管理手段,识别与获取用户行为数据<sup>[6]</sup>,把握和跟踪用户需求变化<sup>[7]</sup>。当某位用户的借还记录与上网行为记录完成同一比对认证后,其自身的基本数据和变化数据就构成了个体画像数据;当大量与该用户相似的个体共性数据被识别,这位用户所属的群体数据即形成,无形中群体关联关系得以建立。关联关系具有动态属性,个体变化数据的不断改变,决定并影响了个体数据在群体数据归属的持续重组与更新,此时这位用户所在的群体会发生改变。关联关系的

建立是提高信息服务有效性和专指性的必要条件,是用户画像模型构建的基础。

## 2.2 全面准确地获取多类型用户数据,整合得到有效数据

用户数据具有采集来源广、数量大的特点,并且由结构化、半结构化和非结构化的数据组成,其中半结构化与非结构化数据超过大数据总量的85%<sup>[8]</sup>。结构化数据主要是个体数据、借阅数据和日志数据,虽然依自身结构利用的需要存储于不同的系统模块中,但存储要求基本统一,存储位置相对固定,数据较为规范、差错率可控。相对而言,这类数据容易获取和分析。

半结构化和非结构化数据大多是行为数据,产生于不同的系统模块中,如书目检索系统中的检索数据、电子阅览管理系统的登录数据、网页界面的浏览数据、数据库操作数据等。由于数据产生的形式不同,标准与规范各不相同,存储方式与数据格式具有一定的独享性,获取与分析相对复杂。以整合系统、统一管理数据为基点,使采集各类数据、完成数据清洗操作更加便捷和可操作<sup>[9]</sup>。

对用户数据进行整合、清洗、认证、管理,从而得到用户画像的有效数据。无用数据是与用户数据无关的数据,包括干扰数据和噪声数据,如登录数据中的进程中产生的错误数据、读者数据中修改前的个人口令数据、检索数据中输入的误操作数据等。剔除这些数据需要运用数据挖掘技术手段,从而达到清洗的目的,有利于数据的精准和有效,对之后开展信息服务的准确性提供有效数据的保证。

## 2.3 通过个体标签描述,建立用户标签体系

用户画像的核心就是完成用户标签描述,即给用户打标签。用户画像标签体系的建立,以人为干预为前提,梳理与排除数据挖掘过程中产生的数据偏差,概括与制定精练的标签特征标识。标签通常是年龄、性别、兴趣等,将所有标签综合后,可以勾画出该用户的“画像”<sup>[3]</sup>。

用户数据标签化的目的,是将处于模糊状态下的用户个体形象化、生动化,并通过打标签的方式,将其个体特征信息有效揭示。具体实施过程是采集获取用户数据,包括用户基本信息、用户行为信息、互动上传操

作信息等数据,提炼关键信息点,利用主题词表或者其他可作为标签词表的参照词表,基本形成与用户画像需求相匹配的标签词表。标签词表的维度不宜过粗或过细,过粗起不到区分的作用,过细不具有典型性和代表性,维度的粗细可在打标签中不断完善。

每个用户的个体标签的数量取决于用户个体产生的行为和生成数据的多少。用户标签的可识别性为信息服务的定制化和专属化提供依据。其中,对用户基础特征的描述数据包括性别、年龄、职业数据等;对用户在全类系统的操作行为描述,用来反映用户潜在特征的数据包括借阅、点击登录、网页浏览、资源下载等行为数据。依据相邻归类、聚类算法,形成聚合后的上位类标签<sup>[10]</sup>,利用现有的标签词表以简单直观地打标签方式将用户画像的特征逐一显现出来,当标签描述与分析维度契合后,进而汇合成为标签群和标签集群,最终建立用户画像数据标签体系,为信息服务的用户画像模型构建提供指引和帮助。

## 2.4 利用大数据及其算法,搭建用户画像模型

围绕用户自身及日常行为来构建用户画像模型,提供个性化服务,以用户的需求、习惯和兴趣爱好为精准推送点。

构建用户画像模型过程中,可以利用文本挖掘与分析、过滤与推荐算法、自然语言处理、聚类算法、关联规则等大数据技术,进行用户数据建模,结合机器学习来完成数据模型的不断修正与精准,随着用户变化数据的更新,模型也会同步更新和优化,从而得到精准有效的数据模型,为后续数据处理提供保障。

用户画像模型的构建可分为三层,即数据来源层、数据分析整合层及数据挖掘层。

数据来源层是用户画像的原始基础数据层,其数据形式由原始数据服务的系统功能决定,按照一定的顺序排列并存储在相应的系统中。

数据分析整合层是实现跨系统数据获取、整合的数据层,其通过数据清洗等技术手段对原始数据进行前期处理,运用专业技术进行数据分析,建立用户画像的有效数据集群;同时,数据分析整合层还要兼顾用户变化数据的时时更新,以确保数据分析过程中产生的新数据或者更新数据,可以及时划入数据分析范围。

数据挖掘层是用户画像构建过程中的关键层,它

是在利用数据分析整合层的前期处理结果,结合用户画像标签体系,完成对用户的比对与识别,采用数据挖掘等技术对用户进行聚类分析,建立数据模型,从而实现用户画像的直观展现。数据挖掘层需要从多角度多维度组合进行数据挖掘,包括行为规律的挖掘、知识库的挖掘、空间利用的挖掘、利用周期的挖掘和需求采集的挖掘等多种组合方式。

最后,通过用户画像的模型预测,可以对用户未来行为进行预测判断,使用户特征得以具体化、形象化,从而完成精准推送服务,为用户提供全方位、个性化的特定服务,增强与用户的互动交流,最终提高信息资源的有效利用率。

## 3 用户画像在信息服务中的作用

用户画像已经进入公众生活的方方面面,包括电子商务、APP、微信、微博等营销手段和新媒体在内的各个领域,对图书馆用户画像的构建起到“试金石”的作用。用户画像的应用将是图书馆精准服务的基本条件,甚至可以看作前提条件,为有效的信息服务提供价值帮助。基于图书馆对于知识用户构建的深入研究,使图书馆领域的用户画像应用成为趋势。

目前,国内的专业图书馆和高校图书馆已经在深入研究与构建用户画像,公共图书馆也开始了用户画像的构建工作,特别是经历2017年第6次县级以上公共图书馆评估定级工作后,服务数据的公众显现度被强行列入公共图书馆的服务范畴,但将大数据体系混同为大数据发布系统是知识体系中狭义且片面的认知错误。用户画像作为大数据体系的重要组成部分,将引领图书馆服务的创新与变革。

### 3.1 提升信息资源利用效果

深入研究用户画像,挖掘与获取用户的关注点、兴趣点与潜在需求。基于用户需求设计构建系统,精准显现用户需求,通过对个体信息、行为痕迹及关注点的获取,了解用户的资源利用行为,精准推送用户需求资源,收集用户画像对信息资源利用的效果,评估阶段时间内资源推送和信息服务精准率的实际利用效果,进而促进知识服务、智能服务<sup>[11]</sup>。

图书馆信息服务资源中,信息资源繁多各异,用户选择难度大、查全率和查准率不高,是用户利用效果有

限的主要障碍。用户画像的数据挖掘是完善个性化信息推送与检索的重要手段,通过深度挖掘用户的个体画像信息与关联信息,结合周期感知的变化与更新,利用数据关联原则和聚类算法等方式,构建特征用户画像群体,完成个体个性化信息推送,提升信息服务质量与有效性,实现对用户的精准服务。

用户画像使得用户个体信息清晰化,其兴趣与行为数据也趋于扁平化,信息资源的精准推送成功率不断提高,信息资源的实际利用效果显而易见。以用户为主体的信息服务,与图书馆信息服务发展方向相吻合,通过对用户画像的不断“绘制”与更新,建立个体用户标签库和标签库集群,分析用户个体与用户群体的需求动向与趋势,规划图书馆的未来发展方向和建设思路,完善资源建设体系、调整空间布局和功能等,从而为科学决策提供依据与参考。

### 3.2 促进信息服务方式创新

现有的信息服务分为被动推送和主动推送,被动推送一般是根据用户的文献借阅与检索记录,提供主题词相关性的信息资源推送,准确性有限、局限性大,还会由于关键信息不足导致信息服务推送的偏差。主动推送多是依据借阅记录和预约记录,结合主题关键词提供的初级推送服务。

用户画像构建完成后,用户个体画像已经形成,用户个体行为的预测会以用户属性数据与行为数据为依据,结合用户自身的多种“变化”,得到用户个体行为预测的结论。信息资源的服务依据预测结论,可达到用户信息服务完全的定制化和个性化,提高信息服务的效能。

用户通过标签完成聚类与联合,也通过标签查找和确定用户。当信息资源需要完成定向推送时,标签作为决定性要素,从用户群中精准选择有效用户。例如,摄影类资源的推送,可推送的用户一般具有的标签包括摄影摄像、户外运动、旅行、发烧友等,由于标签之间存在联合关系,推送范围可扩大至地理学、生物学、野外生存等标签。较之大面积推送,有针对性的推送可以获得更好的效果,也可以节省人财物力。同理,如果开展主题为户外运动类的读者活动时,通过标签查找,可以精准定位到拥有运动达人、野外探险、旅行家、登山爱好者等标签的用户,结合相关宣传完成主题类信息资源的梳理,将包含主题信息资源和本次活动内容

一并推送给这部分群体,活动的关注程度将达到最大化,亦完成了主题信息资源的推送服务。

## 4 推进图书馆用户画像应用的思考

国内用户画像的研究与应用正在悄然兴起。借鉴用户画像在电子商务和门户平台成功的应用经验,结合图书馆用户画像的研究现状,提出以下思考。

### 4.1 做好用户数据的隐私保护工作

用户画像的研究与应用正在兴起,特别是在大数据时代,虽然社会背景与技术手段等为用户画像原始数据的获取提供了基础,但尚未进入成熟阶段,用户个体信息的隐私忧虑尤为突出。用户画像数据中,涵盖有用户大量的私人数据,个性化推送服务越精准,则表示其个体数据越全面、数据类型越丰富,这对隐私安全带来挑战。Henczel<sup>[12]</sup>提出可以从只收集必要信息、尽量利用现有信息和确保受访用户充分了解收集信息的原因和用途的3个方面,来解决涉及用户隐私的问题。

图书馆应建立健全信息安全和隐私保护规范,作为安全管理机制的组成部分;加大相应的防火墙模块功能建设,从技术手段力求杜绝信息威胁;提高技术人员的业务素质与防隐私泄露意识,在数据调取与使用上实行分级管理,严格操作流程和手续,确保用户信息安全。

### 4.2 更新优化用户标签库以提升用户体验

将用户信息运用立体化交叉分析的描述方法逐项打标签,一个标签代表一个维度。以标签和标签群的描述为基础,针对个体用户的标签各异且侧重不同,用户标签库需对集中特征与重要属性的标签予以加权操作和更新加权,不断优化标签结构。一方面,用户画像处于动态变化中,其个体标签作为动态变量产生新生标签,为保证精准推送信息资源,就必须完成动态标签的实时更新,将动态变化同步至动态标签库;另一方面,减少冗余标签可以帮助建立用户画像与用户数据的映射关系,筛选有效信息。改进用户画像的标签库是提高用户画像准确度的有效方式。

基于个体动态标签的更新,发现与关注个体用户

的动态变化,及时掌握由于个体动态变化引发的群体动态变化,关注用户群体的特征,及时了解和预测个体用户及用户群的需求,可以提高信息资源的利用率,完成信息资源精准推送,提升用户体验。

## 5 结语

大数据时代下,用户画像不断应用于各领域各行业,但其侧重点和用途不同。同样,当用户画像被大众普遍接受时,图书馆也在积极探索用户画像的应用潜能,特别是在图书馆信息服务要求日益精准化、个性化的当下,用户画像的发展前景已经被业界广泛认同。图书馆还需进一步研究如何有效利用大数据分析手段,充分运用数据挖掘技术和分析处理方法,精准分析和预测用户行为趋势,从而完成知识数据转化为有效信息,为信息资源精准推送提供支撑。

## 参考文献

- [1] 张小可,沈文明,杜翠凤. 贝叶斯网络在用户画像构建中的研究[J]. 移动通信, 2016(22): 22-26.
- [2] TRAVIS D. E-Commerce Usability: Tools and Techniques to Perfect the On-Line Experience [M]. Boca Raton: CRC Press, 2002.
- [3] 曾鸿,吴苏倪. 基于微博的大数据用户画像与精准营销[J]. 现代经济信息, 2016(16): 306-308.
- [4] 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.
- [5] 程全. 基于用户画像的数字图书馆信息服务模式研究[J]. 图书馆学刊, 2018(4): 68-71.
- [6] 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, 2018(3): 105-109, 137.
- [7] 曾建勋. 精准服务需要用户画像[J]. 数字图书馆论坛, 2017(12): 1.
- [8] 刘速. 浅议数字图书馆知识发现系统中的用户画像——以天津图书馆为例[J]. 图书馆理论与实践, 2017(6): 103-106.
- [9] 潘宇光. 高校智慧图书馆读者信息需求的用户画像[J]. 合肥工业大学学报(社会科学版), 2018, 32(2): 113-119.
- [10] 薛欢雪. 高校图书馆学科服务用户画像创建过程[J]. 图书馆学研究, 2018(13): 67-71.
- [11] 韦良珍. 大数据环境下用户画像在图书馆的应用[J]. 中华医学图书情报杂志, 2018, 27(3): 33-37.
- [12] HENCZEL S. Creating user profiles to improve information quality[J]. Online, 2004(3): 30-33.

## 作者简介

贾蔷,女,1979年生,副研究馆员,研究方向:图书馆服务、数字图书馆,E-mail: 22857782@qq.com。

Application of User Portrait in Library Information Service

JIA Qiang

(Capital Library of China, Beijing 100021, China)

Abstract: With the application of big data technology and the diversification of user demand, library information services are gradually developing to precision. Technical personnel pay more attention to understand user demand and expand service mode through the analysis of user and resource data. This paper introduces the concept of user portrait, discusses the methods and means of constructing user portrait, and elaborates on the data sources in the construction process, the analysis and processing of user data, data labels and the process of model construction. Finally, it puts forward some thoughts on the function and subsequent development of user portrait in information service.

Keywords: Library; User Portrait; Information Resource; Data Tags; Model Construction

(收稿日期: 2019-06-16)