

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.

DOI: 10.3969/j.issn.1671-3982.2018.03.007

· 研究与探讨 ·

大数据环境下用户画像在图书馆的应用

韦良珍

[摘要] 图书馆的用户数据分析能从总体上把握读者对图书馆的需求及利用情况, 而深入分析、标签用户数据, 构建用户画像, 提高服务个性化、精准化则是图书馆服务的新趋势。用户画像在图书馆中的应用将会深化图书馆变革, 促进图书馆进一步发展。根据国内外关于用户画像在图书馆界及其他领域的研究现状, 分析了用户画像应用于图书馆的可行性与作用, 并在大数据环境下构建了图书馆用户画像, 基于用户的信息需求提供精准化、个性化服务, 提高图书馆的服务水平及质量。

[关键词] 大数据; 用户画像; 模型框架建构; 信息需求; 图书馆

[中图分类号] G250.76; G252

[文献标志码] A

[文章编号] 1671-3982(2018)03-0033-05

Application of user profile in library under the big data environment

WEI Liang-zhen

(Guangxi University of Nationalities, Nanning 530006, Guangxi Zhuang Autonomous Region, China)

[Abstract] Analysis of library user data can show their needs and use of library. Deep analysis of tagged user data, establishment of user profile, improvement of individualized and accurate service are the new trends in library service. Application of user profile will promote the reform and further development of library. The feasibility to apply user profile in library and its role in library service were analyzed according to the domestic and foreign studies on the application of user profile in library and other fields. A library user profile was established under the big data environment, which can provide accurate and individualized service for users according to their needs and improve the library service.

[Key words] Big data; User profile; Model framework construction; Information needs; Library

大数据环境下图书馆的变革进一步深化, 图书馆服务形式也不断创新, 如知识发现、学科服务、Lib 2.0、移动终端服务等。同时图书馆用户信息需求更加多元化、个性化, 而目前图书馆信息系统主要统计每天读者数量、借阅数据等整体数据, 与用户需求产生了矛盾。大数据时代, 图书馆服务的发展趋势逐渐个性化、精准化, 用户画像为解决这个问题提供了

思路。

用户画像旨在深入了解用户, 推测用户的真实需求和潜在需求, 精准化定位个人或者人群特点, 以便为用户提供个性化、精准化的服务。用户画像为商家和企业利用大数据进行精准营销提供了营销机会、运营方向, 最大程度把用户的消费行为进行“可视化”。因此, 笔者认为可以借鉴电子商务, 在掌握读者对图书馆的现实需求下, 通过深入分析读者检索行为、阅读行为等数据, 将用户画像应用于图书馆的服务中, 实现图书馆的精准服务。

1 用户画像及其国内外研究现状

经过文献检索发现国外图书馆领域对用户画像的研究比较多, 且应用实践也相对成熟。相比之下, 目前国内图书馆界对用户画像的研究很少。

[基金项目] 广西研究生教育创新计划项目“广西公共图书馆面向老年人群体创新服务研究”(gxun-chxzs2017096)研究成果之一

[作者单位] 广西民族大学, 广西南宁 530006

[作者简介] 韦良珍(1990-), 女, 壮族, 广西来宾人, 在读硕士研究生, 研究方向为图书馆阅读服务。

1.1 用户画像概述

用户画像来自人类对自我需求的探索。大数据环境下,信息技术发展迅速,用户画像作为一种勾画目标用户、联系用户诉求与设计方向的有效工具^[1],广泛应用于电子商务、金融、社交网络等领域。交互设计鼻祖 Alan Cooper 较早提出了用户画像概念,认为用户画像是真实用户的虚拟代表,是建立在一系列真实数据之上的目标用户模型^[2]。

David Travis 提出满足用户画像的 7 个条件:基本性 (Primary research)、移情性 (Empathy)、真实性 (Realistic)、独特性 (Singular)、目标性 (Objectives)、数量 (Number)、应用性 (Applicable)^[3]。用户画像是构建在多元系列真实数据上勾画用户需求和喜好的用户模型,其意义在于深度了解用户,推测用户的现实需求和潜在需求,进而推送精准化、个性化信息。而这些多元真实数据来源于用户本身,包含用户的基本信息、兴趣、习惯、行为等特征,将这些数据立体化、可视化,从而勾画用户画像。

1.2 国外用户画像模型构建研究

用户画像技术的算法是国外学者研究的主要内容之一。构建、完善及应用用户画像都需要算法与技术的支持,如向量空间模型算法、朴素贝叶斯分类法、加权算法、过滤算法等。运用标签向量和标签组描述用户画像时因用户的唯一性导致每个标签对不同用户的重要程度不同,因此需要对标签加权处理。Billsus D 和 Pazzani MJ 提出针对用户的短期喜好应采用最邻近分类算法,而长期喜好则采用朴素贝叶斯分类法^[4]。

由于对用户画像的理解角度不同,如用户画像有个人用户画像与群体用户画像,因此学者研究的出发点有所差异,构建用户画像的模型和方法存在差别。Amato G 等学者提出用户画像的构建需明确描述相关的用户信息及如何描述两个步骤,同时他们认为用户画像构建的信息应该包含个人信息、采集数据(采集的内容、类型和来源)、数据的提交时间与方式、用户的行为数据(特别是与隐私相关的安全数据)^[5]。Rumpler B 认为应通过以下 4 种方法构建用户画像模型:采用调查或访谈的方式采集用户信息,运用记忆算法和程序挖掘用户信息,通过用户对自身、对使用系统、对特定领域的认知,通

过案例推理用户行为^[6]。Henczel S 提出用户画像构建的 6 个步骤:明确需要采集用户的信息,分析现有数据以及需补充的数据,通过调查或者访谈收集数据,分析系统数据特征并按照个人或具有相同特征的群体进行用户画像,评估阶段注意用户行为变化制定方案,通过反馈机制完善及维护用户画像^[7]。

1.3 国内用户画像研究现状

学者通过用户的群体特征对移动互联网和互联网的数据构建用户画像,从而为精准营销提供数据支撑。如通过对新浪微博数据进行采集分析,构建用户画像模型,描述企业用户群体的行为特征,支撑精准营销^[2];在分析移动用户的基站轨迹基础上,采用频繁模式挖掘、构建概率矩阵等方法,构建包含地理位置信息的用户画像模型^[8];采用贝叶斯网络构建用户兴趣模型,以用户兴趣模型刻画用户画像^[9];借助大数据平台分析用户的潜在需求和目标,帮助企业营销和品牌形象建设^[10];以大量的电信移动用户数据为基站,从移动用户的活动频率、行为偏好和移动速率 3 方面进行用户画像构建^[11]。以三枪集团的营销数据库作为案例,通过 4C 理论构建了用户细分模型,力求重构用户的需求,精确定位消费者目标群体,为企业实施精准营销方案提供了重要支撑数据^[12]。

图书馆服务也可以借鉴电商及其他领域应用先进技术的经验,为读者提供更为准确的信息服务。目前图情领域的学者已经开始关注用户画像。我国图情领域有关用户画像的研究较少,有的从知识社区的角度构建图书馆知识社区用户模型,运用层次分析法评价用户画像中标签对服务能力的影响^[13]。有的在实际案例中从资源层、数据采集层及数据挖掘层构建图书馆用户画像^[14]。把用户画像应用到阅读推广的智慧推荐系统中,从而实现用户个性化服务及打破无差别推广现状^[15]。也有学者从国外图书馆界对用户画像研究的现状和热点进行分析为国内图书馆应用用户画像提供了思路^[16]。用户画像在智慧图书馆服务中也展现了很大作用,高校图书馆智慧服务在可视化用户画像基础上实现了精准阅读服务^[17]。用户画像技术在特殊读者群体中的应用具有现实意义及社会价值,如通过用户画像构

建读者抑郁情感词典,分析读者的抑郁情感指数,进而针对特殊读者提供其所需的阅读服务^[18]。笔者认为,国内图书馆可以借鉴国外经验构建用户画像模型。由于国内外图书馆发展程度不同,服务类型和用户群体也存在许多差异,图书馆在构建用户画像时应注意差异,对模型进行调整和借鉴,分析用户需求,真正实现精准服务和有效服务。

2 图书馆应用用户画像的可行性与作用

2.1 可行性分析

我国图书馆正处于快速发展阶段,图书馆除了为用户提供知识信息资源保障外,也应构建图书馆用户画像,为用户提供精准信息服务。用户画像已在电商、手机用户、微博等各领域应用广泛,为图书馆应用用户画像提供了技术保障。虽然目前国内图书馆尚未广泛应用该技术,但是天津图书馆、武汉大学图书馆社区知识用户构建的案例,证明了将用户画像应用于图书馆服务是可行的。再者,国外图书馆用户画像的实践应用也已经成熟,典型的应用有信息推荐、信息过滤与图书馆的服务设计。如用户在利用图书馆在线系统检索时,通常会检索到很多不相关或者相关性不大的信息,通过用户画像,了解到用户真实需求便能解决这个问题,为用户提供精准信息。

2.2 用户画像应用于图书馆的作用

2.2.1 有助于重视用户信息需求,推动知识服务

用户画像最初的目的是深度挖掘用户的真实需求与潜在需求。图书馆服务人员研究用户画像,从用户需求角度设计服务系统,在用户画像中实现用户需求的“可视化”,针对用户的基本属性信息、用户搜索行为习惯及兴趣喜好精准推送信息,进而促进知识服务、智能服务。

2.2.2 有助于实现精准服务,提高服务质量

辅助数据挖掘是用户画像的显著特征。通过深度挖掘用户的环境背景,实现图书馆服务系统个性化推荐和检索。目前,用户真正所需信息与检索结果之间的“信息鸿沟”是图书馆检索系统面临的问题之一,而图书馆检索系统可以通过用户画像完善个性化检索。即使用户检索词表述不精准,系统依然能根据用户画像的数据反馈出与用户相关度较高的结果。通过利用用户画像的关联原则和聚类算法

进行个性化推荐,不仅可以实现精准服务,还能提高服务质量。

2.2.3 有助于更新优化系统,提升用户体验

当前图书馆服务系统种类繁多,功能相似。数量庞大的数据库虽然给用户带来了大量数据,但是面对种类繁多、功能各异的数据库,用户难免会产生困扰。图书馆应引入用户画像,构建特征用户画像群体。在服务系统的设计中设置感知变化,便于及时适应用户需求,根据用户画像及时更新优化服务系统,从而提升用户体验。

3 图书馆用户画像构建

目前图书馆业务中用户数据分析主要是从图书馆信息系统中统计图书馆每天借阅量、读者流量等整体用户数据情况。信息系统中还存在大量个体用户的行为痕迹,个体用户或者具有相似信息行为数据的用户数据更需要深入挖掘。用户画像是基于目前图书馆读者数据分析的大平台上更深入挖掘各大信息系统的数据库。针对个体用户数据挖掘、分析及标签最大限度定位用户需求画像,才能提供精准化、个性化、多元化服务。

用户画像是用户信息标识,目的是了解用户(如知识背景、教育程度、倾向领域、信息行为等)并确定如何向各类用户提供精准服务。如某博士用户在搜索引擎的选择上倾向于国外软件,检索的关键词与“教育”相关,那就基于上述信息描述标签,进行用户画像。用户画像的核心即为个人用户或群体用户描述特征标签,这些标签建立在相应的数据仓库上,数据仓库的建立则需要用户各方面数据集合。

3.1 用户画像的数据来源

用户数据统计与分析是用户画像的基础。用户数据可分为用户属性数据与用户行为数据,即通常所说的静态数据与动态数据。图书馆用户在注册时已将个人基本属性信息输入图书馆用户管理系统,用户属性数据包含性别、年龄、学历、住址、身份证号、工作单位等静态数据。根据上述数据,图书馆员可对个人信息进行统计进而为用户画像提供用户个人数据集,并根据这些数据集对用户进行初步识别判断。

用户的行为数据则是通过网站系统、数字资源平台及移动终端产生的数据,如浏览历史记录、咨

询、阅读、下载、订阅、收藏及搜索的关键词或主题等数据信息,甚至可以查询用户登录次数和浏览时间。这些行为数据是用户画像十分重要的标签描述,在此基础上通过聚类形成用户的宏观画像。

用户数据的来源不仅仅是一个系统,用户画像所需的数据在各类系统或数据库中,静态数据和动态数据源自不同的系统。因此需要对用户进行整合。在图书馆用户管理系统中采集用户基本属性信息,用户检索、下载、订阅、阅读等行为数据则通过资源服务平台系统、移动终端系统、网站系统采集。用户画像数据来源由静态数据与动态数据构成。

3.2 用户数据分析与处理

用户画像数据来源中,用户的注册信息和浏览日志的数据需经过数据挖掘技术分类、聚类、关联分析等预处理后才能得到用户画像的有效数据。因为数据源中包含一些与用户特征无关的冗余信息,分析处理此信息十分必要。通过信息过滤系统筛选出有效数据,才能确保用户画像与用户之间的真实性、有效性。

3.3 用户画像数据标签

标签通常是人为规定的高度精炼的特征标识,如年龄、性别、地域、用户偏好等,最后将所有标签综合起来,就可以勾勒出该用户的“画像”^[2]。用户画像信息标签是模型构建的最后一步。图书馆用户画

像是对用户属性数据、行为数据等进行语义化、短文本化标签。如用户登录账号、姓名、性别、职业、工作单位、工作性质等标签主要是对用户属性特征画像;而行为数据标签则通过浏览、检索、下载主题、关键词、登录时间、点击数据库类别等历史记录数据形成用户行为特征画像。通过用户基本特征标签和用户行为数据标签,逐步构建分类标签、多级标签用户画像标签结构。

3.4 用户画像模型构建

通过图书馆用户管理系统、网站系统、移动终端系统、资源服务平台系统采集用户的动、静态数据,经过分析处理后对用户属性数据和用户行为数据进行标签化工作,形成分类、多级标签用户画像标签结构,用户画像模型在此基础上便可实现“可视化”构建。用户画像中个人用户画像与传统的数据统计不同。个人用户画像以“用户”为本,围绕用户进行数据分析,运用可视化方式描述用户需求、动态分析用户的行为习惯,实现精准化、个性化服务。“可视化”的用户画像从多维度标签、分类标签反应用户大量信息和变化,化抽象为形象,逐步展示画像数据结构。同时根据用户行为的动态及时更新、优化、完善已有用户画像。

用户画像框架模型(图 1)建立在数据来源层、数据分析处理层及数据标签层上。

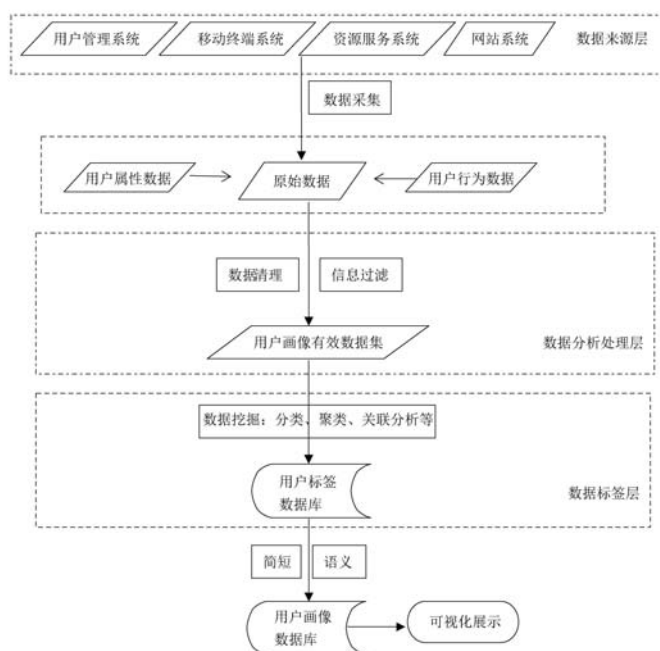


图 1 用户画像框架模型

数据来源层是用户画像的基础,通过图书馆用户管理系统和其他三大系统采集用户属性数据和行为数据,将数据序列化后形成原始数据库,同时注意用户行为数据的变化,以便及时更新原始数据库数据;数据分析处理层则是通过信息过滤、数据清理等技术对原始数据进行预处理,建立用户画像有效数据集;数据标签层是运用数据挖掘技术进行分类、聚类、关联分析等对用户画像有效数据集进行语义化标签,建立用户数据标签数据库,最后建立用户画像,并通过可视化形式展现。

4 结语

大数据环境下,图书馆庞大的用户群体为图书馆服务工作带来了机遇与挑战。基于图书馆业务的用户数据分析为图书馆员了解用户需求提供数据支撑和服务方向,并有助于掌握图书馆总体被利用情况。同时图书馆也需不断提高读者服务水平和服务质量,因此,了解用户现实需求和挖掘潜在需要对服务工作的重要性不言而喻。用户画像技术是在用户数据分析的基础上深化挖掘个体用户或者相似群体用户的信息行为,通过对用户行为数据的分析、标签等用户画像的构建,实现图书馆服务的个性化、精准化。通过分析用户画像数据来源、数据分析处理、数据标签化 3 个层次搭建用户画像模型,能使图书馆用户需求可视化,也能让图书馆馆员了解用户的真实需求及推测潜在需求,从而提高服务质量。

【参考文献】

- [1] 张 哲. 基于微博数据的用户画像系统的设计与实现[D]. 武汉:华中科技大学,2015.
- [2] 曾 鸿,吴苏倪. 基于微博的大数据用户画像与精准营销[J]. 2016(16):306-308.
- [3] Travis D. E-Commerce Usability: tools and techniques to perfect

the On-Line experience[M]. CRC Press,2002.

- [4] Billsus D, Pazzani MJ. A Hybrid User Model for News Story Classification[M]. Sprin Vienna; UM99 User Modeling, 1999:99-108.
- [5] Amato G, Straccia U. User Profile Modeling and Applications to Digital Libraries [C]. Springer-Verlag; European Conference on Research and Advanced Technology for Digital Libraries, 1999:184-197.
- [6] Rumpel B. A Study of the impact of the user profile in documentary systems[J]. Online information review, 2001(6):359-365.
- [7] Henczel S. Creating User Profiles to Improve Information Quality [J]. Online, 2004(3):30-33
- [8] 黄文彬,徐山川,吴家辉,等. 移动用户画像构建研究[J]. 现代情报,2016,36(10):54-61.
- [9] 王庆福. 贝叶斯网络在用户兴趣模型构建中的研究[J]. 无线互联科技,2016(12):101-102.
- [10] 张丽娟. 基于大数据分析的用户画像助力精准营销研究[J]. 电信技术,2017(1):61-62,65.
- [11] 黄文彬,徐山川,吴家辉,等. 移动用户画像构建研究[J]. 现代情报,2016,36(10):54-61.
- [12] 刘 海,卢 慧,阮金花,等. 基于“用户画像”挖掘的精准营销细分模型研究[J]. 丝绸,2015,52(12):37-42,47.
- [13] 胡 媛,毛 宁. 基于用户画像的图书馆知识社区用户模型构建[J]. 图书馆理论与实践,2017(4):82-85.
- [14] 刘 速. 浅议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.
- [15] 王顺管. 以用户画像构建智慧阅读推荐系统[J]. 图书馆学研究,2018(4):92-96.
- [16] 陈慧香,邵 波. 国外图书馆领域用户画像的研究现状及启示[J]. 图书馆学研究,2017(20):16-20.
- [17] 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.
- [18] 韩梅花,赵景秀. 基于“用户画像”的阅读疗法模式研究:以抑郁症为例[J]. 大学图书馆学报,2017,35(6):105-110.

[收稿日期:2018-02-20]

[本文编辑:刘 娜]