数据驱动的图书馆用户画像构建

与信息资源精准服务研究

汤丽媛1 王 俏2

(1.江苏财经职业技术学院图书馆 淮安 223003；

2.江苏护理职业学院图书馆 淮安 223005)

**摘 要** [目的/意义]对图书馆用户画像的构建策略和实际应用进行研究，以期为该领域的研究和实践提供参考。[方法/过程]从用户信息的产生类型出发，利用大数据挖掘技术和相关算法，形成了用以描述用户特征的画像标签体系，并在此基础上引入3W2H五维导向法模型，在图书馆与用户双向沟通的前提下设计出图书馆用户画像的基础构建模型。[结果/结论]结合工作实际，提出基于读者画像的信息资源精准服务模式在图书馆各项工作中的实现途径：分析用户需求，提供精准信息推送；细化读者群体，采用多种推广模式；感知用户情境，营造特色服务场景；预测读者需求，优化馆藏资源建设；引入反馈机制，搭建用户交流平台。

**关键词** 大数据 图书馆 用户画像 信息资源服务

**分类号** G250

**Research on Library ReaderPortrait Construction and Information Resource Accurate Service Sriven by Data**

Tang Liyuan1, Wang Qiao2

1. Jiangsu Polytechnic of Finace and Economics Library, Huaian 223003, China;

2. Jiangsu College of Nursing Library, Huaian 223005, China

**Abstract** [Purpose/significance] This paper studies the construction strategy and practical application of library user portrait in order to provide references for the research and practice in this field. [Method/process] Firstly, starting from the generation type of user information, the big data mining technology and related algorithms are used to form a portrait label system to describe user characteristics. On this basis, the 3W2H thought-oriented five-dimensional model is introduced to design the basic construction model of library user portrait. [Result/conclusion] combined with the actual work, this paper puts forward the realization way of accurate service mode of information resources based on reader portrait in various works of the library, which are analyzing users' needs and providing accurate information push; refining readers' groups and adopting various promotion modes; perceiving users' situations and creating special service scenes; predicting readers' needs and optimizing collection resource construction; introducing feedback mechanism and building user communication platform.

**Key words** Big data Library User portrait information Resource service

**前言**

计算机技术和移动互联网的飞速发展，为人们的学习、工作和生活带来了前所未有的变革。计算机、智能手机前的每个人在遨游网络时，总会在时间或者空间上发生一定程度的交互行为，留下一定的“痕迹”：小至每一位用户的注册信息、登陆账号、浏览行为和阅读喜好，大到用户的隐性兴趣、社交行为、参考咨询和学术研究等多元化信息，这些信息和数据均以超乎我们想象的速度在产生和增长，形成了用户大数据，在大数据技术的助力下，我们可以分析和挖掘这些隐性的、动态的信息，将大量无序的数据转化为有价值的信息，提取不同的特征标签，从而更为精确的勾画用户，使图书馆的个性化信息资源服务更为精准，在此基础上做到更好的提升用户体验，满足用户需求[1]。

本文对现有的理论研究和实践情况进行了归纳，结合国内几大数据库中查阅到的学术文献，目前国内图情领域中与用户画像构建相关的研究成果总体较少,截至2020年2月在知网中仅查阅到70余篇学术论文，且起步较晚。首先在概念研究方面，吴国芳[2]在《绍兴图书馆图书检索垂直引擎研究及应用》中首次提出了“通过建立用户画像实现个性化垂直引擎研究及应用”的设想，但并未对“用户画像”的概念进行阐述，直至2017年以后的专题研究中基本上都包含了对“图书馆用户画像”这一概念的各种解析，且都将研究重点落在了图书馆用户画像的构建研究和实际应用上，如胡媛在《基于用户画像的数字图书馆知识社区用户模型构建》[3]一文中对“图书馆用户画像”概念进行阐述后，采用了层次分析法评价动态精准画像对信息服务的影响，刘速通过案例分析介绍图书馆用户画像框架模型的构建方法[4]，陈慧香分别从概念研究、模型设计、实践应用等层面分析归纳了国外图书馆用户画像的研究现状，为国内图书馆界的研究和实践提供了新的思路[5]；其次在构建策略方面，朱白提出利用用户数据来绘制读者的“脸谱”[6]，张海涛分别尝试利用“概念格”的方法构建用户画像[7]，梁欣在梳理国内外数字图书馆用户画像的基础上提出了一种基于Hadoop的数字图书馆用户画像系统[8]；最后在实践研究层面，李晓旭探索了用户画像在数字图书馆智能阅读方面的创新模式[9]，裘惠麟[10]提出了应该利用大数据搭建图书馆精准服务架构系统，徐海玲等学者研究了基于群体用户画像的图书馆智能服务[7]……足以可见，近三年来，随着大数据技术的日益成熟和广泛使用，图书馆用户画像也逐渐受到了国内外学者的重点关注和研究，可广泛运用在图书馆的科研服务、精准信息服务、智能化阅读推广和场景化服务。

笔者认为，图书馆用户画像际上就是对符合一定特征的读者进行模型化的数据描述。用户画像避免了传统的主观分析读者的行为，转而从读者本身的行为数据出发，以需求为导向，带动图书馆信息资源服务模式的精准化。通过前期的文献回顾可知，基于各个馆的实际情况和用户群体的差异，研究人员对图书馆用户画像的构建策略和实践应用是通过不同角度、不同层面来展开的。笔者对前人的研究成果进行了总结，发现用户画像的构建和应用处于一个不断发展、变化中的环境，但多数学者将用户画像的构建策略仍聚焦在以“图书馆”为主体、以“读者”为客体的单向构建研究上，鲜少引入用户的互动参与和反馈机制，也未形成用户画像“构建—应用—完善”的闭环效应。是否可以通过引导用户来参与“画像”的自我修正，使其更加客观、准确的反应出用户的主观意向呢？是否可以引入有效的反馈和评价环节，变传统的单项沟通为图书馆与用户协同发展的双向机制呢？基于这些问题的提出，笔者拟从用户信息数据的类型出发，利用大数据挖掘技术，引入3W2H五维导向法模型，从分析用户的属性标签、行为标签、活跃度标签、喜好特征标签、满意度标签设计出图书馆用户画像的基础构建模型，并在此基础上对图书馆信息资源的精准服务模式进行探索，以实现图书馆与用户的双向沟通为前提，以搭建协同、长效的沟通机制为目标，确保用户画像在具体实践中发挥出其应有的价值，以求进一步提升图书馆信息资源精准服务的质量和效率。

**1 图书馆读者用户信息的获取与利用**

**1.1 用户画像信息获取**

一般情况下，用户画像数据基本上按照两个维度来划分：静态维度和动态维度。静态数据即读者身上较为显性或者固化的信息，如姓名、性别、年龄、专业、学历等信息；而动态数据则指的是隐藏在用户行为下面的隐性数据，如登陆频次、借阅信息、研究方向、兴趣爱好、检索数据、教育培训、浏览痕迹、常用平台等一系列可以反映出读者某一特性或者某一属性的动态数据(如图1所示)。对图书馆来说，通过用户管理系统即可获得用户在注册时输入的基本信息，这些基本信息会帮助馆员对用户的身份进行初步的识别和判断，在此基础之上，用户通过网站系统、数字资源平台和移动终端产生的数据，如浏览历史、登陆频次和时间段、借阅历史、下载、订阅、收藏以及搜索的关键词等可以描述读者行为轨迹的动态数据，通过这些行为数据组成了用户画像中标签描述，再利用用户画像构建模型通过对用户标签的聚类形成其宏观画像。

**图1 图书馆用户静态信息(左)与动态信息对比(右)**

**1.2 用户数据挖掘与处理**

用户画像(persona)的概念最早由交互设计之父Alan Cooper提出，“Personas are a concrete representation of target users”，是指真实用户的虚拟代表，是建立在一系列属性数据之上的目标用户模型[11]。由此可见，数据的挖掘与处理是实现用户画像的重要支撑，通过前面的章节可知，图书馆可以通过用户信息系统和智慧管理终端收集到海量的用户数据，但这些数据多是重复出现且杂乱无章的，要将这些价值密度较低的大量数据重新加工处理、并从中得到有用的信息，就要运用特定的算法进行分析了，基于大数据技术在图书馆中的应用日益普及，我们也可以利用大数据挖掘算法处理读者数据，大数据挖掘算法较多，其中关联规则、神经网络、支持向量机、K-Means聚类算法等均可用于分析用户行为大数据的规律和模式[12]，从而发现用户行为特征、阅读兴趣偏好、活跃度等信息，从中提取画像标签。

**2 图书馆读者用户画像模型构建**

**2.1 读者用户画像标签建立**

用户画像是真实用户的虚拟代表，其实就是在数据驱动环境下建立起来的用户模型，如果将用户画像比作真实的人脸，那么这张“脸”的骨相就是我们所要用到的模型，“脸”上的五官特征便由读者用户画像的标签体系组成，而用户数据则像血和肉一样，让这张脸变得生动具体起来，无论选择何种“手法”进行“勾画”，其画像的“轮廓”都离不开用户数据的支持，其画像的“一颦一笑”皆由数据驱动下的读者标签决定，其画像的“色调”也应与读者的信息资源个性化需求相契合。其中的用户画像标签体系需要馆员结合图书馆的馆藏特色，再根据本馆用户的实际情况，进行人工总结与概括，如高校图书馆需要从教师和学生这两类不同的读者群体出发、科研院所图书馆需要将科研情况作为读者用户的重要指标，而公共图书馆则要面对多元化的读者和多样化的读者需求……虽然存在着这些差异，时刻面对着千变万化的读者，但我们仍然可以按照读者属性标签、读者行为标签、读者活跃度标签、读者喜好特征标签[13]、读者满意度标签这五个不同指标体系来确定用户的基本画像，其中，读者属性标签是通过采集读者的基本信息来呈现对象的基本特征，读者行为标签是通过获取读者使用图书馆系统这一行为中留下的“足迹”信息来描述用户现阶段获取信息资源的各项特征，读者活跃度主要是指将读者行为的时间路径和空间路径进行整合，用以描述读者获取信息资源的时间跨度和长度特征，读者信息偏好标签主要通过对用户数据的挖掘，预测性的对用户潜在的信息资源需求做出描述，读者满意度主要是从用户的角度出发，对图书馆的信息资源服务模式、用户画像的全面性和准确性进行一个全面的评估，这样既为图书馆的决策提供了一个科学依据，也让读者自身也参与了进来，极大的发挥了用户价值，形成读者用户画像构建闭环。

**2.2 图书馆读者用户3W2H五维画像模型**

再做进一步研究，我们就会发现上文提出的五个用户画像标签正好对应了商业营销领域中的3W2H思维导向法，即以“who、what、when、how to、how many”为核心的五维导向法。本文所将3W2H思维导向法变得具体化、实践化，并且进一步解答了该“如何做”的问题。同理，在对图书馆读者用户画像建模时，可在界定数据源和清洗数据后采用 3W2H的基础模型给用户的标签赋予权重，提炼关键要素。who对应了读者属性标签，解答了“用户是谁”的问题，what对应了读者行为标签，解答了“用户做了什么”的问题，when对应了读者活跃度标签，解答了“何时何地构建服务场景”的问题，how to对应了读者喜好特征化标签，解答了”如何实现信息资源的精准服务“的问题，how many对应了读者满意度标签，解答了”如何有效利用反馈机制“的问题，这五个维度相互作用、相互影响，在图书馆读者大数据资源层的支持下，连同不断发展变化的读者动态数据及其标签支撑体系，一起勾勒出了一个多级的、立体的读者画像，如图2所示。

宣传敏感度

资讯关注度

画像准确度

资源使用率

服务满意度

……

访问时段

在线时长

到馆时间

借阅频次

空间利用

……

资源偏好

阅读兴趣

内容喜好

活动参与

常用服务

……

用户登录

页面浏览

资源下载

书籍借阅

文献求助

……

性别

年龄

学历

专业

地域

……

**图2 基于数据挖掘的图书馆读者画像3W2H建构图**

由此可见，在读者标签的内容展现与推荐过程中，可以参考利用公式：标签权重=时间衰减因子×行为权重×接触点权重[14]来确定用户的最终标签，再将计算出来的标签进行排序，并且考虑到权重值的变化和读者的反馈实时推理用户的兴趣，主要向读者推送与其画像模型中权重值高的标签相符的资源，从而实现读者的信息资源精准服务。

**3基于读者画像的信息资源精准服务**

商业领域的用户画像构建主要遵从以下两个原则：一是以解决实际问题为目的，二是为了进一步提升用户体验。对图书馆来说，构建读者画像的核心目标即实现信息资源服务的精准化，这既是我们研究问题的出发点，也是解决实际问题的落脚点。一般来说，图书馆可从以下几个方面入手，利用用户画像模型实现信息资源的精准服务。

**3.1分析用户需求，提供精准信息推送**

用户画像的构建和利用从本质上解决了当代读者面临的信息过量和知识迷航的问题。通过用户画像中展示出来的静态信息，如用户的年龄、学历、专业、职业、兴趣爱好、研究方向等帮助馆员充分的洞察用户“可能需要何种信息资源”，进而有针对性的主动推送，如高校图书馆可以根据教师用户的画像特征将科研资讯、国内外研究热点作为主要的信息资源推广主体，而面对学生用户时，就要根据当代大学生的画像特征，选择一些新生必读、信息素养或者是职业规划、考研创业类的书籍和资讯进行推送；公共图书馆在对本馆用户进行深入分析的基础上，可以将亲子阅读、社会热点、大众科学、执业考试、继续教育等不同类型的信息资源推荐给不同的读者，这样就做到了变过去的“被动式”服务为新形势下的“主动型”服务，变过去的“经验主义”模糊服务模式为“数据驱动”的精确服务模式，进一步节省了用户的时间和精力，在浩瀚如烟的信息爆炸场景中帮助用户摆脱信息过载的困境，从而进一步提高图书馆信息资源服务的使用率。

**3.2细化读者群体，采用多种推广模式**

在引入读者行为标签体系划分时，就可以利用大数据算法关联到邻近用户和相似用户，因此，图书馆可以利用用户画像汇聚具有类似行为特征的用户群体，在阅读推广实际工作中，就可以在掌握特定群体需求的基础上，选择最合适的推广模式，设计与用户需求相匹配的宣传内容，从而提高推广的针对性。如根据同一群体呈现出来的阅读行为特征，可以变过去的畅销书目推荐为主题书目推荐，将不同主题、不同学科的书目信息细化到相应的受众群体中；在契合目标群体行为习惯的基础上，再采取丰富多彩的推广形式呈现给读者，如很多馆在面向年轻读者时，广泛运用新媒体推广手段极大的调动了读者的参与性；超星学习通APP面向在校大学生组织“云舟杯”线上主题共读活动，取得了较好的推广效果；北上广深等一线城市公共馆面向社区、服务站点推出的“24小时图书馆”、“城市书房”等项目，助力全民阅读推广；北京大学、南京大学等面向高层次读者推出了“人与书”互联互通的智慧化服务模式。在公共图书馆的阅读推广活动中，也可以根据读者画像的群体特征分门别类的制定和实施推广方案，如针对阅读写作爱好人群开展专题讲座、针对亲子阅读人群制定长效共读计划、针对有继续教育、出国考试需求人员提供专业知识方面的咨询与辅导、针对本地特色文化资源打造全民阅读品牌项目。这种细化读者、分类推广的模式不仅满足了特定人群的信息需求，而且进一步扩大了图书馆阅读推广的覆盖面，保障了图书馆各项推广活动的效率和效果。在进行学科服务时可根据科研用户的行为特征将其分为新用户、普通用户和核心用户，然后根据不同的群体开展个性化服务：对新用户开展信息资源技能培训和科研选题、查新服务，鼓励其掌握更多的信息资源获取方法；对普通用户进行深度挖掘，提供嵌入式科研资源信息服务，使其转化为核心用户；邀请核心用户参与图书馆的读者荐购和资源配置活动，助力馆藏资源建设工作。此外，图书馆还可以根据协同过滤技术，在同一画像群体中采用“读这本书的人还读了……”的推荐模式，将用户Ａ频繁使用而用户Ｂ尚没有发觉的资源推荐给用户Ｂ，或者根据用户B当前的资源使用特点推测用户A下一步的资源需求进行推荐[15]，从而进一步体现了图书馆信息服务的人性化原则。

**3.3 感知用户情境，营造特色服务场景**

提取用户的到馆频率、访问时段、借阅频次、空间利用信息等可以显现读者活跃度的标签信息，不仅让用户画像“动”了起来，也让读者本人在不断流动着的数据之中“活”了起来，让我们更加真切的感知到用户使用图书馆的情境，再利用数据处理技术和智能化手段，为用户提供场景化服务。如燕山大学曾利用用户画像中的情境化标签，通过分析某高校图书馆情境相似度、活跃相似度较高的50位读者，将传统模式下的书单和基于情境化推荐和算法的两种书单分别推荐给这50位读者，结果后者获得了更高的用户满意度[16]；吉林交通职业技术学院在分析微信公众平台用户活跃时段的基础上推出了“八点阅读”栏目，获得了较好的推广效果[17]；在网络媒介的环境下，国内众多高校图书馆都推出了面向用户需求情境的大学信息共享空间(UCIC),通过物理空间的再造增设了自主区域、创客空间、多媒体学习区、科研小间等区域，通过虚拟空间的构建形成了以学科馆员和用户共同参与的科研服务平台、“产教研”相互融合的知识服务平台，共同为读者提供了一个一站式服务的信息资源学习中心。

**3.4预测读者需求，优化馆藏资源建设**

用好用户画像模型，可以更好的了解读者用户对信息资源的需求，为图书馆信息服务及馆藏资源建设提供科学决策。比如，通过用户画像中的喜好特征标签体系发现读者用户对电子资源的使用频率远远高于纸质资源，这说明图书馆在下一步的资源建设中要提高电子资源的比例；如果读者用户画像显示，到馆读者对图书馆空间的使用需求呈上升趋势，那么图书馆要考虑是否将物理空间和功能布局的优化作为下一步工作的重点；如果可以对本馆学科服务的核心用户进行深度挖掘和学科追踪，预测其未来阶段所需要的科研资源，就能在此基础上主动完善馆藏资源，更大程度的满足科研用户，赋能本校的学科建设。在条件成熟的情况下，图书馆还可以尝试建立用户画像与信息资源匹配模型，同时将用户画像的数据标签和馆藏资源的标签进行交互匹配，生成用户相似资源集和主题的匹配[14]，这样不仅进一步提高了用户获取信息资源的精准度，同时也加大了图书馆本身的资源建设力度，提高了图书馆的核心竞争力。

**3.5引入反馈机制，搭建用户交流平台**

海量数据标签化的图书馆用户画像模型相比于传统的用户模型更具柔性，更注重用户需求的变化，因此必须建立起相应的评估测试方案和用户满意度反馈机制，使用户画像及其反应出来的需求更加精准。对用户来说，因其动态数据一直处在一个不断变化的过程中，如果因为他的“不慎”或者“偶尔”的检索行为，就有可能被打上不太合适的属性标签，这时就可以引导读者通过满意度标签来进行自我修正，从而避免了其真正需要的信息资源和服务受到规避；对图书馆来说，可以将读者满意度标签体系作为图书馆服务绩效评价的主要标准，通过用户的反馈信息更加有效的评估自身提供服务的准确率和利用率，激励馆员查漏补缺，提升业务能力，从而更好的推动信息服务的效能。这种开放、共享、沟通的理念也再次印证了将以“FOLIO(Future of Libraries is Open)为主体的下一代开放型图书馆发展新模式” [18]，在这一过程中，一座可视化、直观化、数量化的共享平台在读者和馆员之间逐渐搭建，有效的弥补了长期以来竖立在用户需求满足和馆员提供服务的“信息鸿沟”，既因为读者的参与使其自身的画像更为生动精确，又打通了图书馆利用用户画像实现精准服务的“最后一个环节”，初步形成了“以用户为中心”的新时期图书馆信息资源精准服务的良性闭环。

**4 结语**

互联网时代和大数据技术为图书馆的发展带来了无限可能。新时期的图书馆员应勇于面对时代发展的挑战，准确判断新时期图书馆信息资源面临的环境改变，把握好科技革命带来的机遇，善于利用现代信息技术和互联网发展理念创新信息资源服务的模式。本文基于数据驱动的图书馆用户画像模型的构建和利用，摒弃了传统的信息资源推荐方式，通过对用户数据的挖掘、标签的识别和3W2H模型来构建用户画像，并且增设了用户参与环节，从而进一步保证了图书馆用户画像的科学性和准确性，助力新时代的图书馆信息资源工作变“大众适用”为“群体划分”，变“模糊推荐”为“精准服务”，变“单向沟通”为“双向交流”，在对用户需要或潜在的信息需求做出精准判断的基础上，为图书馆的阅读推广工作提供一个有力的参考，为图书馆的发展决策提供科学依据，为图书馆实现信息资源的精准服务提供一个有效的路径。本文对用户画像的构建研究和图书馆信息资源精准服务模式的探索，可为行业内的相关研究者和图书馆信息服务人员提供一个理论上的参考和解决问题的思路。

**参考文献：**

[1]陈 丹,罗 烨,吴智勤.基于大数据挖掘和用户画像的高校图书馆个性化服务研究[J].图书馆研究与工作,2019(4):50-53,59.

[2]吴国芳.绍兴图书馆图书检索垂直引擎研究及应用[J].绍兴文理学院学报(自然科学), 2014,34(1):18-21.

[3]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(4):82-85,97.

[4]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.

[5]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.

[6]朱 白.图书馆读者用户“脸谱”绘制研究[J].商洛学院学报,2017,31(5):87-90,96.

[7]徐海玲,张海涛,张枭慧,魏明珠.基于概念格的高校图书馆群体用户兴趣画像研究[J].情报科学,2019,37(9):153-158,176.

[8]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.

[9]李晓旭.基于用户画像的数字图书馆智能阅读推荐模式[J].图书馆学刊,2019,41(08): 120-124.

[10]裘惠麟,邵 波.基于用户画像的高校图书馆精准服务构建[J].高校图书馆工作, 2018,38(2):70-74.

[11]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.

[12]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,2018,40(4):68-71.

[13]李 雅.基于读者用户画像的高校图书馆精准化服务研究[J].农业图书情报学刊, 2018,30(12):108-111.

[14]戴陈熙.基于读者用户画像的图书馆资源个性化推荐服务研究[J].各界,2018(22):171-172.

[15]杨 燕.图书馆用户画像及其应用研究[J]. 四川图书馆学报, 2018(01):69-71.

[16]刘海鸥,姚苏梅,黄文娜,等.基于用户画像的图书馆大数据知识服务情境化推荐[J].图书馆学研究,2018(24):57-63,32.

[17]2018年中国高职图书馆发展论坛案例分析[EB/OL].[2020-02-17]. https://wenku.baidu.com/view/0715d53f842458fb770bf78a6529647d26283463.html.

[18]肖 铮,林俊伟.用微服务构架下一代图书馆服务平台——以FOLIO为例[J].图书馆杂志, 2018,37(11):63-69.

**作者简介** 汤丽媛，女，1987年生，图书馆学学士，工商管理硕士，研究方向为图书馆读者服务、阅读推广、图书馆信息资源营销，发表论文6篇；王 俏，女，1989年生，图书馆学硕士，研究方向为图书馆读者服务、学科资源建设等，发表论文1篇。