

从数字化到数据化

——关于“图书馆大脑”的思考

刘柏嵩 豆洪青 杨春艳

(宁波大学图书馆与信息中心, 宁波 315211)

摘要: 随着数字图书馆走向智慧图书馆, 图书馆正历经从数字化向数据化变革。在此过程中, 智慧图书馆发展需要构建“图书馆大脑”, 而人工智能的发展为“图书馆大脑”的产生提供充足的技术支持。数据驱动的“图书馆大脑”体现在数据图书馆、智慧情报服务、智慧空间服务等方面。

关键词: 智慧图书馆; 图书馆大脑; 数据图书馆; 人工智能

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2018.03.001

1 图书馆因“大脑”产生智慧

1.1 图书馆当前主要面临的挑战

图书馆由自动化的传统模式向数字图书馆发展, 到如今的智慧图书馆, 发展原因主要是知识传播方式的变化引发图书馆的自身变革。

随着出版数字化的发展, 以及“互联网+技术”在图书馆的应用, 图书馆资源呈现海量、动态化的发展趋势, 来自互联网的文献资源比重越来越大。图书馆虽拥有海量数字化资源, 但面对大数据时代越来越明显的数字鸿沟, 文献资源的跨学科因子和现有知识组织方式已不能满足跨学科知识服务问题, 图书馆的文献资源结构需要改变。

用户信息行为一直随外部环境的变化而实时改变, 在融合线上线下行为数据、社交网络数据等后, 用户个性化需求的显现变化越发巨大。信息服务不是简单地为用户提供文献, 而是需要辨识用户个性化需求的特点, 并进行深度服务, 如对概念关系、主题脉络、知识网络进行分析与呈现。深度、可视化、个性化信息服务是大数据环境下智慧图书馆发展的方向, 数据使用户行为具有“可知性”, 给图书馆与用户的交互带来机遇和挑战。不受时间、空间限制的多终端线上服务需

求不断涌现, 与线下服务需求同等重要。

空间作为图书馆重要的服务重地之一, 不断被赋予新的变化与功能。图书馆的主动服务与可供用户选择的被动服务同样重要, 图书馆除利用场馆与资源向大众提供公平、普适的服务外, 还需摸索如何在有限的人力基础上提供更多个性化、精准化、实时性、贴心化、机器化、自助化的服务, 由固定时间服务转变为全天候服务。这是图书馆在大数据时代面临的一个重要挑战, 需要图书馆以资源优势为基础, 推进服务的优势转变, 提供“智慧”服务。

从数字人文研究视野看, 数据的分析、利用已成为图书馆提升其服务能力的一个重要表现形式, 如基于数据的研究热点、趋势分析等。同时, 在数字人文研究中, 对研究对象的计量化及深入全文的本地化处理, 也将生成大量数据, 通过对其充分地分析与挖掘以满足用户在内容层级的数据服务需求^[1-2]。

1.2 智慧图书馆发展需要“大脑”

王世伟^[3]、刘兹恒^[4]认为, 智慧图书馆是互联、高效、便利的图书馆, 主要特点是工作网络化、服务泛在化、内容数字化、用户自主化、阅读移动化、功能智慧化。其功能特点表明, 智慧图书馆的发展需要一个智慧

的“大脑”。

智慧图书馆所谓的“智慧”，指图书馆对其资源、空间及用户的认识、辨析与判断处理，并具有发明创造的能力，而这种能力非“大脑”不能完成。图书馆在活动过程中产生大量认知数据（如用户借阅行为数据、空间预约数据、入馆频率与时间数据等），这些数据如果不经过“大脑”的整理、分析、归类，就不会生成知识与智慧，也不能指导图书馆服务的“应激”行为。

智能化的自助服务并不等于智慧服务。图书馆虽然通过引进物联网、机器学习、图像识别等技术构建智能化服务，但诸如RFID定位与自助借还、RFID图书盘点、人脸识别入馆等智能化自助服务仅是对业务流程的机器化，并未对图书馆业务流程产生根本性的变革，不能主动回应用户需求，难以发现用户的新需求。

“互联网+大数据”的发展，给图书馆装上“大脑”。“图书馆大脑”是一个数据驱动的智能体，是对图书馆各类运行系统产生的直接数据及关联数据的集成与融合，并对图书馆所有资源与服务进行优化与调配，实现图书馆智慧服务与智慧治理。“大脑”是智慧图书馆的中枢系统，其基于对用户行为与资源、服务的全面感知，建立人、资源、服务三者间的立体互联，形成立体、多维的用户画像，通过对数据的感知、收集、分析、挖掘，指导图书馆用户服务。

2 人工智能发展催生“图书馆大脑”

技术变革催生新经济形态的加速形成，移动互联网、云计算、大数据、物联网、人工智能等新一轮科技革命和产业变革席卷全球，网络信息技术与社会各领域深度融合，AlphaGo、Libratus、无人驾驶汽车等新型应用是集云计算、大数据与人工智能等信息技术的知识产物，在改变图书馆知识管理与信息服务的同时，锻造知识服务结构和服务主客体行为，给图书馆未来发展带来较大影响^[5-6]。

人工智能有两个基础，分别是大数据和人工智能算法。深度学习兴起、大数据发展、计算能力上升、计算成本下降及自然语言理解提高是人工智能实现突破的关键，也是推动和实现“图书馆大脑”的技术支撑。

2.1 图书馆“大脑感知”所需技术

传统图书馆的感知数据，主要来源于用户刷卡所产

生的数据。越来越多的新型感知类数据获取技术被用于图书馆，如移动定位、语音识别、人脸识别等技术。

目前，语音识别技术已从实验室走向市场。2012年，微软发布同声传译产品，采用深度神经网络方法，一举将语音识别错误率降低近30%，大力推动语音识别技术产品化的进程^[7]；2017年，微软文字转录超过人类转录员平均水平，语音到文字转录的词错率为5.1%，低于人类转录员平均词错率（5.9%）^[8]。语音识别技术的发展，意味着图书馆可提供基于机器的智能化现场咨询服务，如天津市、杭州市、宁波市等地图书馆展出的“会说话的机器人”，可根据用户过往的阅读行为习惯，为用户提供参考咨询服务^[9-11]。

人脸识别技术让用户感知与识别应用变得更精准、更方便。2017年9月，苹果公司发布的产品Face ID采用深度神经网络法，在真实生活场景中人脸识别错误率降低至百万分之一^[12]。上海交通大学、浙江理工大学已将人脸识别技术应用于图书馆，让图书馆告别“刷卡”时代，可自动感知、识别用户，使用户进图书馆再无“门槛”^[13-14]。运动轨迹识别技术为图书馆识别用户运动行为，提供多样化的技术支持，包括基于移动端的运动轨迹识别技术、基于视频的运动轨迹识别技术等。目前RFID、蓝牙技术、NFC、WiFi AP、红外信号、iBeacon等单元情景感知技术，已经成功应用于图书馆，并融入导航和定位服务。该类技术结合图书馆空间功能，向用户提供基于场景的个性化服务，如在用户借阅、空间预约过程中，根据空间管理系统、RFID定位系统等积累大量用户空间使用数据，面向不同服务场景，提供线上委托和线下执行服务、自助服务（自助借阅、预约）及个性化推荐服务。

2.2 图书馆“大脑思考”所需技术

（1）机器学习。机器学习是利用经验改善系统自身性能，是从人工智能中产生的一个重要学科分支，也是实现智能化的关键。学习是人类智能的主要标志和获得智慧的基本手段，机器学习的研究是希望计算机能像人类那样具有从现实世界获取知识的能力，这是智慧大脑产生的基本条件。

（2）深度神经网络的演进，优化机器学习的速度。近十年来，人工智能领域最重要的突破是产生了深度学习。深度学习适应大数据发展的需要，这种灵活的建模语言对不同问题可构建不同模型，有效地模拟大

脑行为。

如杭州市构建的“城市大脑”，通过对车流、事故、违章等数据与公共服务数据、运营商数据、互联网数据等进行关联分析，结合视频识别、图像搜索等技术，认识并“懂得”车辆运行状态和轨迹。这些数据成为“城市大脑”智慧的起源，辅助“城市大脑”对城市交通进行全局的即时分析，并根据机器学习的迭代，不断优化算法模型，优化道路调度，随着运行时间增加可计算出更“聪明”的方案^[15]。

3 基于大数据驱动的“图书馆大脑”

3.1 数据图书馆

数据图书馆是数字图书馆的升级，是从资源的数字化向资源、空间、用户的数据化转变与丰富。数据图书馆服务是基于多源数据识别、保存、认知、交互的智慧化服务，如决策支撑、知识发现支撑、服务融合支撑等^[16]。

当前，图书馆服务过程中主要产生三类数据：一是GB级书目数据，主要包括书目信息、特色数据库等；二是TB级数字资源，主要有馆藏数字化、商业学术数据库、MOOC课程资源等；三是PB级用户数据，用于记录用户偏好，主要来源于空间管理系统、图书借阅系统，以及图书馆网站访问、数字资源访问、网络学习平台访问等所产生的数据。数据图书馆不仅提供数据资源层面的服务，而且须从海量文献数据中挖掘出可能影响用户阅读行为的焦点、热点问题，进而通过数据图书馆的互操作向用户提供深层次、个性化的服务，进一步推动实现智慧图书馆。

在智慧图书馆中，数据从四个维度产生作用：①决策，通过数据而非图书馆管理者的经验，更好地作出理性决策；②优化，通过数据分析，进行管理和服务的优化；③推广，基于用户数据创建用户画像，进行精准推荐与阅读推广；④创新，在数据服务上，形成新的服务与业务增长点，如科研数据服务、数字人文中的数据服务等。图书馆一直在实践和探索如何更全面、更方便地收集用户的兴趣，思考如何让用户参与到服务中，产生更完整、全面的数据。但数据的完整获取，营造智慧环境，需要通过感知技术，增加其与用户的接触点，记录每一位用户的行为数据；需要更多地从到馆和不到馆的服务中感知用户并与其互动。如通过分布在网站和

移动APP中不到馆的服务接触点和实体场馆的到馆服务接触点，收集和分析每一位用户的行为数据，洞察用户，发现其需求并为其提供个性化服务；通过建立图书馆“数据集市”，实时呈现其整体运行状况和用户需求。如文献采购过程中，结合网上书城的销售数据和评论数据、书商的销售数据、用户的借阅数据等进行分析，并对用户阅读的需求进行预测。

3.2 智慧空间服务

智慧图书馆在空间上体现为交互性、在线化、感知化。交互性使图书馆更懂用户；在线化，使图书馆“永不打烊”；知识化，即利用数据分析的智能算法和数据驱动策略，使图书馆“更聪明”^[17]。

空间服务智慧化体现为动态的空间推荐服务，是基于用户当前的显式和隐式空间需求，以用户对空间的使用轨迹数据、使用目的数据（如研讨主题获取）和时段分布与偏好数据等，构建用户与空间的联系，从而进行空间推荐。图书馆中的研讨空间和座位分布于各个楼层，传统的研讨空间和座位管理系统通过用户预约，为其展示空间使用情况，以二维的形式展现空间里的各种设备，缺乏对空间的立体描述，是一种静态的被动式选择。而“大脑”管理下的空间推荐服务，是对现有空间资源的智能调配，是与用户交互的动态推荐服务。其以用户数据为导向，综合考虑用户的历史空间数据、学科背景、使用时间、使用评价，以及其对现有空间的利用情况，为用户动态推荐定点空间（如靠窗、带插座的座位，有投影仪的研讨室）、关联空间（如接近用户学科和阅读兴趣的区域）等。

3.3 智慧情报服务

智慧情报服务，即从海量数据中获取对用户有价值的信息，辅助用户决策。

（1）基于数据的个人学术助理，包括个性化学术资源推荐、投稿智能引导、学者圈推荐等服务。“图书馆大脑”通过记录、分析、理解用户个人的学术行为（包括用户的检索词、检索过程、成果产出等），自动向用户推荐学术资源；投稿是用户展现学术成果的渠道，“图书馆大脑”通过基于主题、学科与用户投稿的匹配，分析相关期刊的学术研究热点、评审专家等，对投稿信息进行自动分析和匹配，引导用户一键式投稿；学

者圈是开启用户间相互关注和交流的钥匙，“图书馆大脑”汇集用户的众多数据（包括资源使用、研究方向、兴趣点等），以此发现用户间的关联，从而进行社会关系的推荐（如相同兴趣用户推荐、相同研究方向推荐、学术圈推荐等），以便为用户提供学术交流资源。“图书馆大脑”以数据为核心提供学术服务，根据用户和学术团队的物联网、视频图像、语音、文本等输入行为数据，以及学科背景、研究方向等建立用户画像，构建互动的学者关联图谱，打造“资源找人”的智能服务模式，实现资源智能地匹配用户画像，让静态资源推荐转变为实时动态推荐，完成学术助理的使命。

（2）基于机器人的自助参考咨询服务。机器人是“图书馆大脑”和用户交互的触角，机器人除能提供用户咨询、借还书指引、扫码找书、用户引路等服务外，还能与用户、工作人员进行日常交流，推进图书馆与用户的智能交互。机器人对用户的历史数据进行分析，通过迭代学习，不断丰富用户画像，从而精准定位用户问题焦点与需求，完成智能荐书、智慧引导等多项服务^[18]。

（3）基于海量文献的动态资讯生成。“图书馆大脑”利用算法程序，从海量数据库收集各类题材数据，建立数据间的联系，将重要资讯和解读呈现给用户。如2009年，美国西北大学开发名为“Stats Monkey”的软件，利用此软件可以自动从网页中抓取大学棒球比赛的数据信息，并在12秒内生成一篇新闻报告^[19]。

4 结语

人工智能发展可分为感知、认知与自主三个阶段。感知属于弱智能阶段，听、说、读方面的发展都很成熟，触觉感知较弱，如图像识别、语音识别能力不强；认知属于智能阶段，能够理解、思考、交流，如深度语义解析、对话式交互；自主属于强智能阶段，是让机器有自主智能^[20]。

“图书馆大脑”的构建与发展，是由弱智能向强智能的发展，目的是帮助传统图书馆，形成图书馆的智慧生态，走向智慧图书馆。通过对个体的画像，由基于文献单元的服务向基于数据单元服务的转变；通过提升场景感知与计算能力，从被动适应用户需求向主动预测并应对需求的转变。让图书馆“知道更多，做到更多”，让用户“体验更多”^[21]。

参考文献

- [1] 刘炜, 叶鹰. 数字人文的技术体系与理论结构探讨 [J]. 中国图书馆学报, 2017, 43 (5): 32-41.
- [2] 柯平, 宫平. 数字人文研究演化路径与热点领域分析 [J]. 中国图书馆学报, 2016, 42 (6): 13-30.
- [3] 王世伟. 论智慧图书馆的三大特点 [J]. 中国图书馆学报, 2012, 38 (6): 22-28.
- [4] 刘兹恒. 智慧图书馆——资源建设新篇章 [C] // 安徽省数字图书馆建设研讨会. 2017.
- [5] ARLITSCH K, NEWELL B. Thriving in the age of accelerations: a brief look at the societal effects of artificial intelligence and the opportunities for libraries [J]. Journal of Library Administration, 2017, 57 (7): 789-798.
- [6] FERNANDEZ P. “Through the looking glass: envisioning new library technologies” how artificial intelligence will impact libraries [J]. Library Hi Tech News, 2016, 33 (5): 5-8.
- [7] 微软计算机“同声传译”用本人声音说中文 [EB/OL]. (2012-11-01) [2018-01-22]. <http://tech.sina.com.cn/it/csj/2012-11-09/16527785340.shtml>.
- [8] 新里程碑! 微软语音到文字转录已经达到人类水平 [EB/OL]. (2017-08-21) [2018-01-22]. <http://www.chinaz.com/news/2017/0821/797863.shtml>.
- [9] 机器人现图书馆 [EB/OL]. (2017-06-29) [2018-01-22]. http://www.tianjinwe.com/hotnews/wlr/201706/t20170629_1169269.html.
- [10] 图书馆来了个机器人管理员 能听能说能借还 [EB/OL]. (2016-10-14) [2018-01-22]. <http://zj.people.com.cn/n2/2016/1014/c228592-29144907.html>.
- [11] 宁波大学园区图书馆来了个机器人管理员 能听能说能借还 [EB/OL]. (2017-12-19) [2018-01-22]. <http://news.cnnb.com.cn/system/2017/12/29/008713452.shtml>.
- [12] Sergio De Simone. Apple Details Face ID Security [EB/OL]. (2017-09-29) [2018-01-22]. <https://www.infoq.com/news/2017/09/apple-faceid-white-paper>.
- [13] 上海交通大学图书馆人脸识别服务正式启用 [EB/OL]. (2018-01-16) [2018-01-22]. <http://www.lib.sjtu.edu.cn/index.php?m=content&c=index&a=show&catid=212&id=1834>.
- [14] 百度云人脸识别落地浙理工图书馆 [EB/OL]. (2017-04-11) [2018-01-22]. <http://www.chinanews.com/business/2017/04-11/8196746.shtml>.
- [15] 耿晓军. 阿里城市大脑1.0发布 献礼世界城市发展 [J]. 物联网技术, 2017, 7 (10): 6-8.

- [16] 谭影虹. 从数字图书馆到数据图书馆——大数据时代的图书馆服务范式转变 [J]. 图书与情报, 2016 (3): 75-78.
- [17] OTTERLO M V. Project BLIIPS: making the physical public library more intelligent through artificial intelligence [EB/OL]. [2017-12-21]. <http://qqml-journal.net/index.php/qqml/article/view/341>.
- [18] 章玲. 成都高校图书馆首台机器人上岗 [EB/OL]. [2018-01-01]. http://news.youth.cn/sh/201709/t20170926_10785179.htm.
- [19] 人工智能观察室. 写稿机器人25秒完成一篇地震消息稿 [EB/OL]. [2018-01-01]. http://www.sohu.com/a/164580979_443184.
- [20] OYELUDE A A. What's trending in libraries from the internet cybersphere—artificial intelligence and other emerging technologies [J]. Library Hi Tech News, 2017, 34 (2): 11-12.
- [21] FERNANDEZ P. “Through the looking glass: envisioning new library technologies” understanding artificial intelligence [J]. Library Hi Tech News, 2016, 33 (3): 20-23.

作者简介

刘柏嵩, 男, 1971年生, 博士, 研究馆员, 研究方向: 知识组织与信息管理, E-mail: lbs@nbu.edu.cn。
豆洪青, 男, 1975年生, 硕士, 副研究馆员, 研究方向: 智慧图书馆, E-mail: douhongqing@nbu.edu.cn。
杨春艳, 女, 1990年生, 硕士, 馆员, 研究方向: 信息组织, E-mail: yangchunyan@nbu.edu.cn。

Digitalization to Datalization: About the Library Brain

LIU BaiSong DOU HongQing YANG ChunYan
(Library and Information Center, Ningbo University, Ningbo 315211, China)

Abstract: Library is transforming from digital to data-centered with the development that Digital Library move to Smart Library. In this process, the development of Smart Library needs to build “Library Brain”, and the development of AI provides enough technologies to support the generation of “Library brain”. The data-driven “Library Brain” is embodied in the data library, intelligent information service, intelligent space service and so on.

Keywords: Intelligent Library; Library Brain; Data Library; Artificial Intelligence

(收稿日期: 2018-02-12)

■ 书 讯 ■

《中国高被引分析报告2016》

《中国高被引分析报告2016》按理、工、农、医、人文、社科等领域划分为50个学科, 综合分析了各个学科的高影响力论文、研究热点与前沿、高影响力期刊、高影响力作者和高影响力科研机构, 并以关联图谱的方式展现了多种学术关系, 有助于科研人员及时发现并跟踪研究热点, 有利于期刊编辑部监测本刊学术影响力, 有利于科研管理机构评估科研能力, 是高等院校、科研院所及期刊编辑部等相关单位和人员的参考工具书。

该书以“中国知识链接数据库”为依托, 数据覆盖我国6 000余种期刊的论文及引文。书中分学科揭示了高影响力的学者、研究机构(大学、研究所、医院等)、地区(省/自治区/直辖市)、学术期刊、图书、外文期刊和会议录, 并采用共词分析、共被引分析和合著分析等方法绘制出各学科的前沿主题分布以及作者、机构和期刊间关联的知识图谱。

《中国高被引分析报告2016》由中国科学技术信息研究所编制, 曾建勋主编, 科学技术文献出版社出版。《中国高被引分析报告2015》也已同步发行, 欢迎业界同仁鉴阅订购。