

基于用户数据挖掘的医院图书馆文献采访和服务实践

商慧兰 周珏榕 李 晶 陶慧娟 于海燕

(同济大学附属杨浦医院图书馆 上海 200090)

〔摘要〕 介绍数据挖掘应用于图书馆的主要功能,分析医院图书馆及其数据特点,探讨数据挖掘在医院图书馆的应用实践及其困境与对策,指出用户数据挖掘对于文献采访决策、图书馆服务创新具有指导作用。

〔关键词〕 数据挖掘;医院图书馆;文献采访;服务创新;实践

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.05.018

Literature Acquisition and Services in Hospital Library on the Basis of User Data Mining SHANG Huilan, ZHOU Juerong, LI Jing, TAO Huijuan, YU Haiyan, Library of Yangpu Hospital Affiliated to Tongji University, Shanghai 200090, China

〔Abstract〕 The paper introduces the main functions of data mining applying in library, analyzes the characteristics of hospital library and its data, discusses the application of data mining in hospital library as well as its plight and countermeasures. It points out that user data mining can serve as a guideline for making decisions on literature acquisition and library service innovation.

〔Keywords〕 data mining; hospital library; literature acquisition; service innovation; practice

1 引言

随着数字图书馆理论与实践的发展,医院图书馆开始步入数字化时代。图书馆管理系统中书目数据、读者流通借还数据、Web 及图书馆平台访问记录、资源使用数据等为实现数据挖掘技术在图书馆的应用提供基础。流通数据的挖掘帮助图书馆更好地了解读者阅读的内容、习惯、兴趣等行为特征^[1-2],不仅能为馆藏建设、资源配置提供参考依据,而且有助于了解个性化服务需求。在图书管理系统、Web、数字平台(镜像)电子资源、数据库资源智能分析环境中利用关联分析、序列分析、聚

类相关功能、聚类分析方法进行数据挖掘研究,从而指导图书馆文献采访决策和服务创新。

2 数据挖掘应用于图书馆的主要功能

2.1 关联分析

图书馆数据挖掘中应用最为广泛的方法,属于描述型模式,挖掘隐藏在数据间的相互关系^[2-3]。在图书馆应用中可用来发现读者阅览、借阅行为中的规律,分析用户的需求类型、识别用户特点和兴趣等^[1-2]。

2.2 序列分析

通过分析样本特点,发现其规则和方法,在图书馆服务系统中可对点击率、阅读、检索记录数据进行采集并整理、排序,从中获知用户热门阅读文献和图书^[3],分析读者类别和属性特征。关注数据

〔修回日期〕 2018-12-11

〔作者简介〕 商慧兰,副高级职称,发表论文 20 余篇。

检索、下载、浏览频次,按序排列统计,分析数据利用率等。

2.3 聚类相关功能

利用聚类相关功能挖掘数据的因果关联,如数据库历史检索中主题词、关键词的频次等,可以了解医院热门研究方向,临床疾病、诊断、治疗技术的研究重点;掌握用户资源需求,预测科研发展态势。也可聚焦学科频数,筛选出临床优势学科,从而为图书馆资源配置和个性化服务提供依据。

2.4 聚类分析

聚类的方法很多,有划分、层次、核聚类以及基于网络、数据库的方法等。如层次方法,将数据划分到不同组,尽量保持类层的同质性,将数据组成类分层结构。在图书馆应用中可以对用户分别聚类,识别同类子群,提供同类资源配置与个性化服务。也可利用数据库聚类分析功能,聚焦分类用户、科室科研的数据集,分析专家学者、学科研究的发展态势,预测未来趋势,为决策提供参考依据。

3 医院图书馆及其数据特点

3.1 医院图书馆特点

为医疗、教学、科研提供信息服务是医院图书馆基本职能,资源的结构具有医院专业和学科特色。服务对象主要分为临床医疗、医技用户;科研、教师(导师)用户;行政管理用户;规培学员、医学院校学生、实习用户和进修人员。每类用户专业不同,又可细分为医疗、护理、药学、检验、放射、超声、管理等学科。医学用户的学科专业性非常强,各类人员对信息的需求差异很大,信息主要集中在专业领域紧密相关的生物医学文献上。因此医院图书馆在资源构建、配置、使用习惯模式、用户体验、服务方式等既要兼顾到不同层次人员的需求,又要兼顾到医院网络环境、馆址面积、功能布局、资源分布、采访资金等。

3.2 数据特点

医院图书馆作为文献情报中心,可通过多个渠道获取用户数据,如结构化的用户流通借阅、预约

数据。用户登录、资源使用中产生的记录数据,如用户基本信息、借阅历史、访问记录、资源浏览、下载、拒绝数据;文献检索历史记录、条件检索数据等。只有经过清理、统计、归类、变换、集成的数据才可提供增值服务。

4 数据挖掘在医院图书馆的应用实践

4.1 管理系统集成数据应用

图书馆利用书目管理系统中的统计功能对用户的流通借阅^[1]、检索、预约数据进行跟踪统计^[3]。通过分类聚类方法对收集的数据进行分析,了解用户的阅读倾向与需求^[4]进行馆藏配置和预测新一轮文献采访的重点。如近年来随着图书馆数字化建设的推进,图书流通借阅数据量呈逐年下降趋势,同济大学附属杨浦医院图书馆逐年调整纸版书刊的采购量,增加电子类书刊的采访,不仅节省资金,而且极大丰富资源,将有限经费投入资源最大化。跟踪用户借阅习惯与周期,进行关联分析,调整借阅规则,使制度符合用户实际需求。根据阅览室流通数据的用户分层聚类分析并结合阅览室规模,确立移动图书馆的建设方案,解决一院四区都具备使用图书馆资源的条件,同时确保有限空间用户使用最优化。

4.2 用户检索数据挖掘分析

图书馆一站检索平台实现资源数据集成。首先,图书馆对用户检索过程中产生的记录数据进行捕捉、关联、整理并按序排列,从中分析用户关注较高的热门检索图书及文献需求,为后续图书馆的采选种类、数量决策提供依据^[4]。其次,对数据库的访问量、检索数据、期刊数据、检索历史等进行挖掘排序,可以发现用户习惯,了解其关注的研究热点和方向。聚焦 IP 站点数据分析站点需求、使用情况、培训需求,为图书馆开展后续服务提供依据。数据库流量、并发数、拒绝数据分析也为数据库续订的内容、价格提供采访依据。第三,为资源建设、信息服务提供方向^[5]。如临床指南检索数据促使馆员开展最新临床指南的资源建设;专家专著发表数据促使图书馆开辟专家专著栏目;用户咨询数据

(如核心期刊、影响因子、论文发表、培训)等为平台读者指南服务内容和项目提供建设方向,相继提供前沿信息导读、信息编译、最新到刊等服务。最后,历年临床订阅文献偏好数据归类分析(如使用情况、数据库收录情况、出版商、资源分析)为图书馆采选数据库提供决策方向。文献传递服务数据(如电子书、文献传递)拓展了解用户需求的窗口,也为用户信息数据的建立和应用奠定基础。

4.3 建立用户数据库

图书馆作为知识信息服务中心和中转站,对数据库海量数据信息进行专业化加工处理和应用,对图书馆知识服务、创造、组织和存储有十分重要的作用和意义。图书馆对搜集、检索的用户数据进行分析与处理正逐步成为各图书馆的工作重点。无论是现存信息的检索,还是检出信息的处理分析都促使图书馆在资源建设、管理、服务及发展模式的转变。大数据已从根本上改变人们对以往数据的观念,通过数据信息可以看到其潜在价值,建立一个整合、结构化的用户数据仓库并及时更新已迫在眉睫。因此现阶段主要发现、挖掘数据并进行筛选和积累,建立用户数据资源库(包括用户、使用、资源、检索、科研学术数据等),加强结构性数据库资源建设,为后续数据转化、分析、增值及决策提供基础保障。用户基本信息按聚类可以了解不同层次用户的信息需求,也为个性化服务提供依据。

4.4 数据挖掘在采访策略中的应用

最初会采用非结构化或半结构化数据挖掘,如同行、不同医院、机构数据库功能咨询、评价;相似数据库的功能、数据分析比较;同行使用反馈体验等。医院结构性数据挖掘分析,如数据库的收录范围与医院、学科发展方向、专家的偏好相关度;能否满足不同层次用户的需求(或满足何层群体需求);用户使用情况数据分析、习惯和体验、标准检索方法等。其次,界定采访流量数据,流量数据决定数据库价格。如中国知网拒绝数据分析,可以了解医院并发数能否满足用户需求,指导续订采访决策。数据库(镜像)用户信息数据与网络数据用

数据群组分类分析可以发现一院四区用户数据库的使用情况,预测未来提高数据库利用率的切入点,提高采访资源利用率和全面性。最后,评估采访价格与合同约定,确立数据库性价比。如万方医学数据库价格数据指数可以估算采访价格。迈特思创外文文献检索平台文献传递数据分析可界定限量数据的约定与分配。最大用户流量数据及数据更新趋势分析可为预测平台服务器硬件更新升级时间和硬件配置要求提供依据。通过使用数据的趋势分析可以了解用户的习惯偏好和数据库对用户的影响力。总之,图书馆在基于用户数据挖掘进行采访时要将数据信息与用户信息有序筛选、提炼、积累、转化,才能指导文献采访的方向和重点^[6],优化资源配置和资金分配。

4.5 图书馆数据挖掘扩展空间

图书馆除用户信息数据外还具有其他数据,而通过多维度数据分析可实现以下目标:营造第二课堂,提升信息素养;保障文献需求,推送前沿信息;辅助课题申报,防范学术不端;跟踪学科动态,辅助科研决策;聚焦人才发展,参与人才评估。如整体科研实力检索可预测医院科研发展态势;通过历史检索数据分析重点学科及热门研究方向。序列分析科研成果高产作者和科室,筛选优势学科。聚焦行业领军人物,了解医院科研前沿性,聚类分析研究热点趋势。还可跟踪个体学者,挖掘其科研发展轨迹,评估其研究的相关性、领域的地位;分析学者科研成果的顶峰期、瓶颈期或衰退期,预测未来潜能等。图书馆则可根据多元化数据挖掘分析开展个性化服务(如前沿信息推送、相关领域最新文献更新推送、优势学科和热点研究的资源采访等),实现文献采访资源、资金向优势学科、领先领域研究、热门研究倾斜配置。

5 医院图书馆数据挖掘的困境及对策

5.1 困境

第一,医院图书馆不同于高校图书馆,其资源有限,数据不够完整,缺乏专业软件。部分教授、硕博士可以通过校园网登录使用大学资源,这部分

数据图书馆无法采集。第二,目前数据挖掘限于管理软件、图书馆平台集成数据以及数据库自带的数据统计,缺乏整体集成和专业软件支持,馆员对数据的处理需要筛选、导出、整理、甄别、归类、统计。数据挖掘方法主要采用常规(SPSS、EXL等)工具处理,数据准备费时耗力,不能进行更深层次挖掘利用来提升价值。第三,数据挖掘精确问题。源数据的准确性对数据挖掘分析有很大影响。系统数据来自不同平台,数据的异构性以及没有统一标准导致信息整合、利用困难重重。第四,馆员和管理者对推进图书馆数字化建设认识不足,对医院图书馆定位及功能服务停滞在传统的非数字化时代,缺乏数字化意识。第五,图书馆在医院属行政辅助科室,研究项目少,资金缺乏,人才引进与培训缺乏保障,如果不能顺利解决存在于用户需求、管理决策和图书馆之间的矛盾必将制约图书馆在大数据时代的发展。

5.2 对策

第一,加强馆员及管理者对大数据的认识,信息资源整合是推进数据挖掘利用的首要问题。图书馆作为医院文献信息汇集地,要重视与领导沟通,为临床医教研发挥信息枢纽作用,争取领导最大支持。第二,展示信息成果,扩大影响力。建设信息共享平台,形成特色馆藏,推动档案馆、图书馆、院史馆的融合发展^[7]。第三,培养复合型专业人才。人才是推进资源价值升级的关键,数据挖掘技术门槛较高,馆员不但需要有计算机、信息、图书文档、档案管理知识,还必须具有良好的统计学、外文基础,熟悉数据挖掘技术与工具;而医院图书馆专业性强,又需要馆员最好具有医学背景,对抓取信息有较强的专业敏感度。加强专业化人才队伍建设,既要领导重视,又要馆员努力,通过招聘、送修、培训、外聘兼职专家顾问、对接临床专业馆员等方式,提高馆员技术、能力和专业性。第四,推进智慧图书馆一体化建设。即集智慧管理、大数据分析决策、智能采选、数据应用和智慧办公于一体。建立大学与附属医院联盟,共享数据库资源,

资源互补、完善数据。第五,汇集不同数据源优势与数据商协调解决方案。与数据集成商协作改良、开发有针对性的数据挖掘工具包,嵌入图书馆服务平台,智能抓取、集成分析,注重智慧图书馆集成建设方案和数据挖掘工具的开发,即专业软件开发。第六,促进行业标准的建立。建立统一的数据挖掘标准或模型,探索数据仓库的内涵、范围及评价体系指标。

6 结语

数据挖掘是提取隐含在数据中的规律,为医院图书馆文献采访决策工作提供潜在、有价值的信息指导。通过数据的挖掘、筛选、分析可以发现和获取用户需求,指导采访资源和资金分配。多维度用户数据挖掘为图书馆创新服务提供更加广阔的想象空间,创新服务模式,预测服务方向;提高文献采访的精确性,提供精准化的个性服务。图书馆应结合本馆结构、特色、资源、人力等情况,合理分配有限资源、资金,在实践中不断完善用户数据的挖掘、构建,为开拓更广泛的读者服务领域提供技术支撑和决策指导。

参考文献

- 1 彭仪普,熊拥军. 关联挖掘在文献借阅历史数据分析中的应用[J]. 情报杂志, 2005, 24(8): 40-41, 44.
- 2 王姝. 数据挖掘要医院图书馆用户信息需求研究中的应用[J]. 预防医学情报杂志, 2008, 24(12): 996-998.
- 3 金松根,王琳琳,石丹. 关联规则在流通库读者数据挖掘中的应用[J]. 医学信息学杂志, 2015, 36(9): 74-79.
- 4 徐杰. 基于用户数据挖掘的图书馆文献采访决策模式研究[J]. 中国中医药图书情报杂志, 2018, 42(2): 28-31.
- 5 尹斌. 数据挖掘技术在图书馆创新服务中的应用探讨[J]. 电子技术与软件工程, 2016(2): 195.
- 6 杨杰. 大数据环境下的公共图书馆采访策略[J]. 图书馆学刊, 2015, 37(3): 44-46.
- 7 黄山,李晶,商慧兰. 从创客空间展望医院图书馆转型发展[J]. 医学信息学杂志, 2017, 38(12): 79-82.