

基于数据挖掘的读者流量与图书馆开放能力分析系统

杜慧敏

(南京审计学院图书馆 江苏 211815)

摘要 提出了建立基于数据挖掘的读者流量与图书馆开放能力优化分析系统的设想,并通过对系统需求和系统可行性的分析,探讨了系统设计、数据挖掘过程及系统功能等问题。

关键词 数据挖掘 读者流量 开放成本 开放能力 图书馆

中图分类号:G250.7

文献标识码:A

文章编号:1005-8095(2009)07-0030-03

1 读者流量与图书馆开放能力分析

高校图书馆读者流量具有明显的潮汐性、阶段性的特点。所以,高校图书馆如果一味地追求开放时间的延长、开放部门的扩大,其投入和利用容易失去平衡,服务效益也难以实现。因此,图书馆的开放既应尽量满足读者需求,也需要根据读者使用图书馆的情况做出合理安排,这就需要对读者流量的特点及变化规律进行研究,从而优化图书馆的开放能力,为读者提供最佳开放方案。

2 数据挖掘技术运用于读者流量与图书馆开放能力分析系统的可行性

2.1 数据挖掘技术

数据挖掘技术主要有:

(1)分类分析技术:通过分类挖掘对数据库中的各类数据挖掘出其数据的描述或模型,或者通过建立的统计模型对历史数据进行预测分析。

(2)聚类分析技术:在分类的基础上,根据对象的特征宏观上控制同一类的数据,发现数据的分布模式,以及可能的数据属性之间的相互关系。

(3)关联分析技术:在个性化的检索中,对信息的需求是强调兴趣信息的相关性,而非相异性,根据信息需求心理确定适当的规则可提高数据挖掘的准度和深度。

(4)时间序列分析技术:时序模式是指通过时间序列搜索出的重复发生概率较高的模式,能够根据数据随时间变化的趋势预测将来的值。

(5)预测分析技术:预测是通过分类或估值得出模型,该模型用于对未来未知变量的预测,而这种预测是需要时间来验证的。

在数据挖掘过程中,各项功能不是独立存在而是互相联系的,共同发挥作用。图书馆日常工作中会产生大量的统计数据,如果应用数据挖掘技术对其中的数据进行分析挖掘,就可以为图书馆管理工作提供决策支持。

2.2 数据挖掘技术在图书馆流通管理中的应用

随着图书馆信息化建设程度的不断加深,图书馆管理信息系统得到广泛应用,管理系统对图书馆借阅数据做了非常详细的记录;同时,图书馆门禁系统的使用,也为图书馆准确而有效地获得图书馆资源的利用情况,如每天进馆人数、进馆高峰期等提供了数据,为了解、研究和分析图书馆读者流量提供了可靠的依据。由此图书馆开放能力的研究有了实际操作的平台,也就有了对图书馆开放方案及成本效益进行预测和评价的基础。若以图书馆开放能力和读者流量相匹配以及成本效益最佳为优化目标,建立基于数据挖掘的读者流量与开放能力分析系统,利用数据挖掘技术对图书馆读者流量信息进行挖掘,对读者流量的变化规律进行分析,将有助于发现读者流量的趋势,掌握读者使用图书馆的特点,评价现有的开放方案,并给出优化的开放方案及建议,从而促进图书馆开放策略的灵活调整,更合理地安排有限的工作人员和开放空间。

3 读者流量与图书馆开放能力分析系统的设计与功能分析

3.1 数据挖掘分析系统的设计

要实现图书馆开放能力与读者流量最佳匹配,不仅需要现有管理信息系统及技术的支持,同时也离不开优化的流程设计。图书馆的开放能力与优化的流程设计如图1所示。

图1中流程设计的具体内容包括以下几个步骤:

(1)确定优化目标和挖掘对象。优化过程要考虑两个因素:一是读者减少或增加对人员配置的影响;二是人员配置对开放成本的影响。所以,要找到一个合理的开放时间段,就要对与开放能力有关的信息有所了解,并据此制定目标。

(2)数据挖掘分析。这是图书馆优化开放能力的最重要步骤,它以读者流量、开放成本、部门信息的数据仓库为挖掘对象。这些数据仓库通过对图书馆开放活动中的各类信息进行采集、处理和传输,并按照一定的格式存储,记录了包括读者人数、借阅次

收稿日期:2008-08-19

作者简介:杜慧敏(1962—),女,本科学历,副研究馆员,已发表论文10余篇。

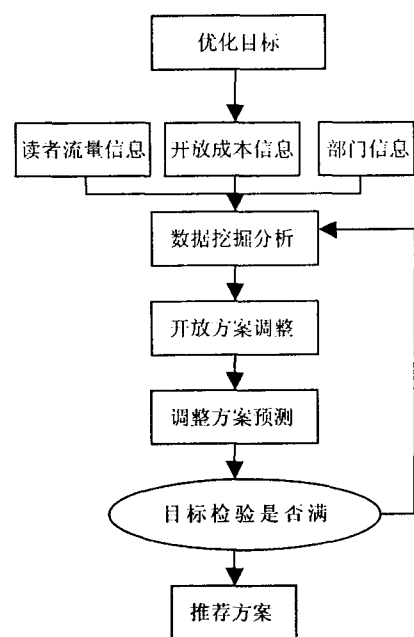


图1 图书馆的开放能力与优化的流程设计

数、阅览室上座率、某一时段读者流量密度、人员工资支出、水电费支出、物业管理费支出、工作人员配备数等在内的数据挖掘所需要的各类信息。

(3) 开放方案调整。依据数据挖掘出的有用信息对图书馆开放方案进行调整。开放方案调整包括人员编组调整、开放时间调整、部门调整、阅览座位数调整、图书类型及数量调整等。

(4) 调整方案预测。这是对调整后的方案效果进行预测。它由开放能力预测、读者流量预测、开放成本预测等组成。其中开放能力预测和读者流量预测用于检验开放能力和读者流量的匹配程度,根据不同的情况可以设置一定的弹性系数;开放成本预测用来进行图书馆开放成本效益的预测。

(5) 目标检验。这是以开放能力和读者流量匹配为优化目标,根据不同的开放方案调整情况,检验能否达到优化目标。目标检验采用人机交换的模式,对于不满足要求的方案可以返回数据挖掘层继续迭代分析;对于满足要求的方案作为推荐开放方案。

3.2 分析系统数据挖掘的实现过程

(1) 数据准备。数据挖掘是由可以获取的数据驱动的,其成功率如何很大程度上取决于所收集数据的数量和质量。因此,数据准备对于数据挖掘的成功至关重要。数据准备大致可以分为如下三步:数据集成、数据选择、数据转化。

① 数据集成。这是集成各数据库中的数据,这些数据可以源于实际工作中相应的数据库,如从图书馆管理数据库、读者流量统计数据库(门禁系统)、财务管理系统或以往遗留的系统中提取,并使之转化为标准的结构化数据库。

② 数据选择。这是从系统中选择出需要分析的与开放能力有关的数据集合,以缩小挖掘范围,避免盲目搜索,提高数据挖掘的效率和质量。选择的数据包括:读者流量信息(这些信息可以通过图书馆管理系统及门禁系统获得或根据预测目标确定);开放成本信息(这些信息可通过学校的财务管理系统获

得);部门信息(主要包括部门类型、数量、阅览座位数、工作人员数等信息)。

③ 数据转化。这是对选定的数据在挖掘前根据挖掘目标加以精炼处理,最后把收集得到的有关数据汇总到数据仓库,为下一步的数据挖掘做好准备。

(2) 数据挖掘。建立数据仓库后,就可以基于这个数据仓库平台进行数据挖掘工作了。根据开放过程中出现的问题及特点,确定数据挖掘的任务和目的,选择适当的挖掘技术和算法。

(3) 结果解释和评估。这是对数据挖掘发现的结果进行解释和评估,并根据评估结果建立相关分析模型,如开放时间与读者流量关系模型、读者流量与开放部门关系模型、开放部门与成本关系模型等,把实际数据(读者流量、成本和开放时间等)作为输入信息,通过分析模型的计算获得预测结果,根据分析结果给出优化的开放方案及建议。其过程如图2所示。

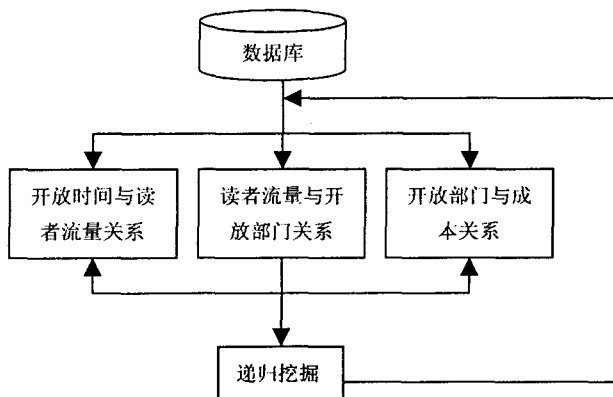


图2 结果解释与评估

3.3 系统功能

基于数据挖掘的读者流量与开放能力分析系统具有以下功能:

(1) 可以对读者流量数据等进行统计分析,按时间集统计读者人数、借阅次数、阅览室里上座率、某一时段读者流量密度,发现读者使用图书馆的高频时间及高频部门,从而有针对性地调整开放时间和人员岗位、优化资源配置。如在读者高峰时间段可以采取加大人手、增加服务窗口,加快图书借阅速度、加开阅览室、延长开放时间等方法来满足读者需求,而在平时或读者低峰阶段则可减少在岗工作人员,适当减少服务窗口或阅览部门的开放等。

(2) 可以利用数据挖掘技术,对读者流量数据进行分析,并据此定期出一个读者使用图书馆情况的分析报告,帮助读者了解图书馆的使用情况;还可对读者流量信息进行即时统计,可视化地展现给读者,给读者提供借阅分流指导,使读者错开高峰、节省时间,提高图书馆开放部门的使用效率。

(3) 根据优化的开放方案,预测部门开放数量及人员的配备情况,测算图书馆开放成本,使图书馆的

图书馆读者阅读倾向研究综述

赵 应

(杭州学军中学图书馆 浙江 310012)

摘 要 对2000年以来有关读者阅读倾向的研究论文进行了综述,认为其研究内容主要涉及读者角度和图书馆角度两个方面。读者角度方面,主要研究读者阅读倾向、影响读者阅读倾向的因素、读者阅读倾向实证方法等;图书馆角度方面,研究了图书馆针对读者阅读倾向中存在的问题所采取的对策。

关键词 图书馆 读者 阅读倾向

中图分类号:G252.17

文献标识码:A

文章编号:1005-8095(2009)07-0032-03

我国对读者学的研究始于20世纪80年代,阅读倾向的研究是读者学的主要研究内容之一,它是指读者为实现阅读愿望,将阅读需要付诸阅读行为,表现阅读特点的一种趋势。图书馆读者阅读倾向的研究能够反映某一历史时期,某些读者群的阅读心理及特点,对图书馆文献资源建设和读者获得知识的便利有理论和实践的指导意义。

基于此,笔者查阅了2000~2004年图书情报学的核心期刊,以及2005年至今的期刊,获取了图书馆读者阅读倾向研究的相关论文,通过对这些论文的分析,得知,其研究内容主要涉及读者和图书馆两个角度、四个方面的研究。

1 读者阅读倾向内容的研究

近几年读者阅读倾向内容的研究主要集中在以下几个方面:(1)不同读者群体的阅读倾向;(2)读者阅读倾向的特征;(3)读者阅读倾向差异性。

1.1 不同读者群体的阅读倾向

1.1.1 大学生读者的阅读倾向

(1)快餐式获取信息型阅读倾向。随着数字化环境的发展,越来越多的读者花更多的时间在浏览式阅读和关键词式的暂时性阅读上,而忽视了深层次的注解式阅读。毕素清认为大学生有浮光掠影、浅尝辄止地获取信息取向,在海量信息面前,大学生只能通过各种媒体,浅层地了解 and 关注信息的大概,只有当与自身功利相结合时,才不得不作深入的钻研。

服务效益和开放成本效益尽可能达到最佳比值。

(4)对历年读者流量数据进行分析,建立读者流量分析模型,预测读者流量的变化规律及趋势,根据预测趋势提前进行开放方案制定、人员准备、空间预留或现有方案的预调整等。

总之,随着图书馆信息化水平的不断提高,运用数据挖掘技术来分析读者流量与图书馆开放能力的最佳匹配关系,制定符合读者需求的开放方案,可以及时对人员配置和开放部门进行调整,最大限度地满足读者需求,提高读者满意度。此外,在提高资源

(2)功利性阅读倾向。李苏媚通过对西安理工大学图书馆2002年图书流通统计表数据分析得出大学生有以下阅读倾向:①专业性倾向。大学生的阅读目的就是为学好专业,提高自身技能,为将来作准备。但计算机、股票、房地产、现代管理等实用书刊走俏校园,使得大学生的阅读常常带有急功近利的色彩。②“外语热”倾向。大学生要经常阅读各类外语书籍以应付各类考试;为了解国际科技发展前沿和动态,再加上出国留学风甚行,大学生要阅读大量的外语书刊。③经济、商贸、金融类书刊也是大学生阅读热点之一。当前大学生的竞争意识和市场意识越来越强,他们想在获取本专业知识的同时,也拓宽自身的知识面,以便将来走向社会,能更好地表现自己的实力。

(3)时尚娱乐性阅读倾向。毕素清认为大学生个性张扬,喜欢追逐潮流,于是,大学生往往会随大流,更多浏览报刊,网络介绍的最热门读物。另外,大学生往往选择图文并茂或网络读物取代原先的纯文字阅读,用视觉上的享受以缓解学业和就业带来的压力。

1.1.2 中学生读者的阅读倾向

(1)轻松娱乐性阅读引领导潮流。夏叶对中学生阅读倾向进行了调查,在15种类型的读物中,中学生最喜欢的读物分别是侦探历险读物、卡通漫画、幽默读物、名著读物和武侠言情读物。青少年最不喜欢

利用效率的同时,还可降低开放成本,使图书馆服务效益和成本效益最大化。

参考文献

- 1 于光,李文峰.数据挖掘技术在图书馆用户管理中的应用.图书情报工作,2005(1)
- 2 田春燕.高校图书馆开放时间探微.图书情报工作,2006(4)
- 3 王真.数据挖掘及其在图书馆的应用.情报探索,2007(12)

(责任编辑:黄建)