

数字图书馆个性化推荐系统研究

李娜

(山西大学商务学院, 山西 太原 030031)

摘要: 互联网技术的迅猛发展使数字图书馆成为现代图书馆建设的一个重要维度, 为读者提供更加优质的信息加工和组织服务方式成为数字图书馆发展的重要方向, 基于读者兴趣和使用习惯的个性化推荐系统应运而生。数字图书馆个性化推荐服务主要包括个性化定制服务、个性化推送服务和数据挖掘及决策服务。数字图书馆个性化服务需要通过相应的系统来实现, 该系统一般包括用户行为数据收集模块、用户兴趣偏好数据分析模块和算法实现及推荐模型模块。现代智能搜索技术、数据挖掘技术和信息推送技术为数字图书馆个性化推荐系统的实现提供了技术保证。

关键词: 数字图书馆; 个性化服务; 个性化推荐系统; 现代智能技术

中图分类号: G250

文献标识码: A

文章编号: 1002-1248 (2015) 09-0108-03

Personalized Recommendation System of Digital Library

LI Na

(Business College of Shanxi University, Shanxi Taiyuan 030031, China)

Abstract: The rapid development of Internet technology has made digital library become an important dimension of modern library. To provide readers with more quality information processing and organization service has become an important direction for the development of library. Therefore, personalized recommendation system based on readers' interests and habits came into being. The personalized recommendation service of digital Library included customization service, personalized push service and data mining and decision-making service. The personalized service required appropriate system to achieve its functions, including users' behavior data collection module, users' interests and preference data analysis module, algorithm and recommendation model module. Modern intelligent search technology, data mining technology and information push technology provided technical supports for the implementation of digital library's personalized recommendation system.

Keywords: Digital library; Personalized service; Personalized recommendation system; Modern smart technology

计算机技术和网络技术的发展为传统图书馆带来了颠覆性的变革, 它从多个角度影响了图书馆信息资源的存在和组织形式, 建设越来越多的数字化资源成为图书馆服务的重要维度。随着图书馆资源的加工和存储方式的改变, 传统的图书馆开发和服务模式已经无法满足现代图书馆对于穿越时空综合集成多种资源服务的要求。依托因特网进行信息收集、组织加工、存储和传递的数字图书馆概念便应运而生, 并迅速成为信息时代图书馆发展的方向^[1]。

图书馆建设重点的改变使读者对图书馆的需求也正在从购买、保存、借阅书刊开始向读者提供更有效

的信息(包括纸本和数字化信息)服务转变^[2]。同时, 读者也面临着身处信息泛滥的海洋却依旧对知识倍感迷茫的困惑当中。搜索引擎中一个关键词的输入导致的是成千上万的搜索结果, 对于信息的驾驭我们依旧无法笃定。在商业领域中, 向客户销售大量的产品, 通过给其提供推荐服务来实现, 这种技术也为数字图书馆的发展提供了新的思路, 建设带有个性化推荐服务系统成为变革数字图书馆发展的一个重要方面。

1 数字图书馆个性化推荐服务的形式

数字图书馆实现个性化服务的关键是以读者为中

收稿日期: 2015-04-14

作者简介: 李娜 (1984-), 女, 山西大学商务学院助理馆员, 学士, 研究方向: 数字图书馆。

心,把用户的需求和个人偏好作为个性化推荐服务的前提依据和实现目标。只有从用户出发,才能真正满足用户的要求,做到更加准确和有效地推荐服务。就目前的市场情况看,数字图书馆的个性化推荐服务主要包括以下3种形式:

1.1 个性化定制服务

这是一种提前做好产品供用户选择定制的服务方式,即需要以个人的主动性选择为前提。作为服务方,首先以大量采集的用户兴趣和偏好作为初始依据,制作相应的定制网页,或者推送相关的内容邮件,将分类全面且有效的信息导航发送给选择定制的目标用户。这种推送方式面对的首先是用户自己的选择,服务比较有针对性。但由于用户本身可能对内容不够了解,他们很难将抽象的个人需求转化为可以推荐的具体内容,因此,这种服务方式相对粗糙,属于比较浅层次的个性化推荐服务。

1.2 个性化推送服务

与第一种被动的依据用户定制进行信息推送不同,这种服务方式属于主动推荐,它主要采用智能的方式自动收集大量的用户信息,包括个人基本特征、使用行为习惯、兴趣爱好、专业水平等,通过对其进行更进一步的处理,进而在数据分析的支撑下主动为用户推送相关内容,或者主动地报送一些与其数据分析大致接近的相关信息。这种个性化推荐服务以一种相对隐形的方式对用户进行信息推送,是一种依据相对科学,操作更加便捷的服务方式。这样的推送服务由于有前期的数据分析为依据,属于比较深层次的推荐服务形式。

1.3 数据挖掘及决策服务

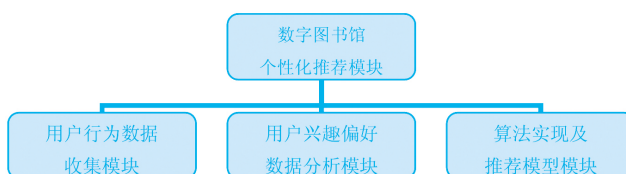
这种服务形式和第二种个性化推送服务有相似的地方,即都需要在对用户信息进行数据分析的基础上进行推送。但该种服务方式的数据分析更加先进,多使用大数据挖掘、算法自动发现、语义网技术等现代智能手段。通过对已经分析后的数据进行进一步的拓展和延伸,进而用于支持用户进行决策和查询。这样的推荐服务需要更加强大的技术支撑,同时服务也更加地精准和有效,属于深层次个性化推荐服务。

2 数字图书馆个性化推荐系统的组成

个性化推荐系统最早且使用最好的是在商业领域,特别是大型购物网站的兴起和它对人们生活产生的巨大影响。该系统的完成需要综合多个研究领域的最新成果,其中涉及计算机技术、网络信息技术、物理学、行为学、心理学等多个学科,是一项精细且综合

的科学技术集合体。对于数字图书馆的个性化推荐系统而言,它属于个性化推荐系统应用的一个分支,以通用的个性化推荐系统为标准,结合数字图书馆自身的特点和读者这个特殊群体的需求,进而形成具有个人风格的独特样貌。

数字图书馆个性化推荐系统一般由3个基本模块组成,如附图所示。



附图 数字图书馆个性化推荐系统的基本模块组成

2.1 用户行为数据收集模块

推荐系统的实现需要以用户的先前行为作为参考依据,此模块主要用于实现数字图书馆中读者行为数据的收集。为了跟踪读者的行为数据,需要为每一位用户建立单独的描述性文件,该文件包括读者的个人基本信息、用户特征和用户数据等。该描述性文件在制定前需要充分考虑该系统的数据分析脉络,理清不同元素之间的网络关系和相互连接情况,进而保证数据和最终目标的实现相一致。

描述性文件的具体形式虽然千差万别,但其类型基本相似,主要分为基于兴趣的描述和基于用户行为的描述^[9]。以兴趣为出发点的描述性文件主要以层次结构和语义网络连接等结构化较强的形式进行,这样的方式层级递进,逻辑严密。以用户行为为主的描述性文件多以浏览化的形式表示,导入便捷,查看清晰。在个性化系统的实现中,大多结合两种方式同时进行,进而充分利用每一种表示方式的优点,实现效率的最大化利用。对于描述性文件的存储形式,大多依据常规的计算机数据表示方式,即以文件描述和关系数据库进行存储。

对于数据的收集,现在大多采用比较便捷的基于Web页面的行为数据挖掘方式。通过集合和分析,比较用户在相关页面的停留时间、点击次数、访问顺序等来获得用户的行为习惯和使用偏好数据。进而在对其分析的基础上,比较发现相似人群,建立用户访问方式模型。

2.2 用户兴趣偏好数据分析模块

该模块的主要作用是依据已经收集的用户数据信息来推断和发现读者下一步的可能行为。对于用户行为的分析一般采用基于内容的方法和基于分类的方法。基于内容的方法是指以信息内容本身出发来抽取资源,

需要以加权的方式对关键词进行特征向量的提取。关键词的选取是保证该方法有效的重要前提,即关键词既要恰当地代表信息源又能进行拆割和细化。基于分类的数据分析方法大多以提前制定的类别为参考标准,然后依据信息的所属范畴对号入座,进而将用户的兴趣和偏好划拨于不同的位置,以便于下一步有针对性地信息推荐。

2.3 算法实现及推荐模型模块

该模块的任务是依据用户兴趣分析结果,采用一定的算法将已有的相关合适内容推送到图书馆的读者手里。每一个推荐算法的采用和实现大都有相关的模型作为理论和前提指导,这种相对科学的经典模型进而保证了推荐的有效性和实用性。其中比较成熟的模型有基于关联规则的推荐算法、基于内容的推荐算法、基于社会网络关系的推荐算法和协同过滤模型等^[4]。这些模型大都在商业领域应用较广,作为数字图书馆同样需要参考这些模型,以便结合自己的实际情况,进而更加有目的性的选择自己的实现算法和参考模型。

3 实现数字图书馆个性化推荐系统所需的技术支撑

数字图书馆个性化推荐系统的实现离不开现代先进技术的支撑,从图书馆服务对象的特殊性出发,基本功能的实现需要在以下3种技术的配合。

3.1 智能搜索技术

数字时代,人们常常迷失在信息海洋中,正因为如此,萌生了各种各样的搜索引擎技术,不同的搜索技术侧重点和关注内容不同,有的偏向于类别的搜索,有的主要用于商业领域,有的则更加关注知识的获取和呈现。智能搜索技术是一种实现了多种搜索功能的综合性搜索技术,以人工智能技术为参考,可以实现信息的快速定位、基于角色的信息呈现、语义识别、用户个人偏好自动识别等功能,是一种更加高效有用的搜索技术。

智能搜索技术可以为数字图书馆的用户提供更加精准的信息服务,进而提高整个服务系统的运行效率,使用户获得的服务更能满足他们自身个性化的、有针对性的需求。该技术是数字图书馆个性化服务系统的关键和核心,它的实现大多以数据库的资源作为依据进行信息的抽取和判断。

3.2 数据挖掘技术

大数据时代,如何充分利用数据挖掘技术来发现更多地信息成为各大领域关注的重点。数字图书馆的个性化服务系统作为新时代图书馆的典型及关键代表,自然离不开数据挖掘技术的支持。简要的说,数据挖

掘技术主要是指从大量的数据中通过算法来搜索隐藏于其中信息的过程。数据挖掘可以借助统计学、搜索算法、信息论、数据库系统等多学科领域的知识来处理海量的数据,进而发现数据中包含的更加深层次及更为重要的东西。

在数字图书馆的个性化推荐系统中,通过数据挖掘技术,可以发现读者更多的使用偏好和兴趣特征,进而更好的预测读者未来的可能行为习惯,做出更加前射的决策。同时,可以对用户的使用信息进行更加准确的关联分析,找到不同数据之间存在的规律性,也可以获得更多聚类集合,为相似用户做出决策提供参考和指导。

3.3 信息推送技术

客户—服务器模式需要靠用户主动提出请求进而使服务器做出响应来送达相关信息,是一种双向配合的信息传输模式。而推送技术则是一种广泛的媒体传播技术,它不管客户端是否提出申请仍然进行信息的发送,例如发送电子邮件,实质上是一种发布/查看模式。推送大多依据用户之前的使用习惯和喜好,将及时更新的相关信息发送到用户手中。对于图书馆的个性化推荐系统来说,推送技术实现的功能必不可少,它可以依据读者的图书浏览方式和已查看的图书信息进行相关资源的推荐,这样可以方便地为用户去发现他们自己都不够清楚但却可能真正有用的图书和资料^[5]。

4 结语

个性化推荐系统是数字图书馆读者服务的一个重要维度,它通过对用户的使用习惯和兴趣偏好进行深层次的分析,可以相对准确地为读者推荐更多图书信息资源,进而为读者的下一步行为决策提供指导建议。在现代信息技术和人工智能技术的支撑下,个性化推荐系统的实现和数字图书馆的建设将向着更加完善的方面发展。

参考文献:

- [1] 郭晖.我国高校数字图书馆建设研究[D].武汉:华中师范大学,2004.
- [2] 陈卓辉.当代图书馆的个性化信息推荐考[J].科技文慧,2014,(1):195-196.
- [3] 彭国莉.信息推送技术与信息推送服务[J].中国信息导报,2001,(8):49-50.
- [4] 宋丽哲,牛振乐,宋瀚涛.数字图书馆个性化服务用户模型研究[J].北京理工大学学报,2005,(1):58-62.
- [5] 侯雅楠.个性化服务系统功能设计与实现—泰达图书馆个性化服务的实践[J].现代图书情报技术,2005,(3):16-19.