

基于兴趣图谱的图书馆泛在服务模式构建研究*

闫霏

(中南民族大学图书馆, 武汉 430074)

摘要: 基于兴趣图谱的图书馆泛在服务能集成用户信息行为数据, 全面挖掘用户的兴趣节点, 为用户推荐资源, 解决传统图书馆信息服务中的数据稀疏性与数据需求不完全满足等问题。鉴于此, 本文设计了一种算法来绘制图书馆信息用户的兴趣图谱, 从而加强图书馆与信息用户的联系, 完善一种相对精准的图书馆泛在服务, 满足信息用户对于“无处不在”的图书馆信息服务的强烈需求。

关键词: 兴趣图谱; 图书馆; 泛在服务

中图分类号: G252

DOI: 10.3772/j.issn.1673-2286.2015.12.007

1 引言

“泛在服务”意为无处不在的服务。图书馆“泛在服务”是国外近几年提出的一个新概念, 指在泛在信息环境下, 图书馆利用自身信息密度高、知识存储总量大的优势, 针对教育、科研及生产等大众化信息需求所开展的各种服务性活动的总称。目前, 信息用户对于泛在服务的需求越来越强烈, 图书馆需要在短时间内整理出信息用户的最强需求并给出相应的对策。因此, 兴趣图谱的方法与图书馆泛在信息服务之间形成了一个共同的连接点, 可以作为构建图书馆泛在服务模式的突破口, 而以下几种图书馆信息用户需求与信息服务现状也反映了引入兴趣图谱的必要性。

其一, 传统的信息用户管理理念不能满足现有需求。通过关注信息用户之间的兴趣相关性, 对信息用户及其感兴趣的主题进行聚类, 可将传统的图书馆对信息用户的一对一维护转化为图书馆对信息用户的一对多维护, 同时也强化了图书馆在信息用户群体中的核心角色。

其二, 图书馆具备收集用户信息行为数据的条件, 并可用于构建完整的图书馆泛在服务模式。利用数据挖掘技术, 绘制信息用户的兴趣图谱, 再基于此兴趣图谱构建完整的图书馆泛在服务方式、服务内容、服务途径等要素, 这从根本上不同于传统的图书馆信息服务,

同时, 也充分发挥了图书馆已有的庞大用户信息行为数据的价值。

其三, 图书馆在服务方式上, 亟需变“被动式”服务为“主动式”服务。以兴趣相同的资源为联系来聚合信息用户群体, 图书馆不仅可以更高效地提供可信赖的建议, 而且可将传统的“被动式”图书馆信息服务改变为以“兴趣”为基点的“主动式”信息服务, 最终可构建以图书馆为核心的泛在服务模式。

2 国内外研究现状及发展趋势

泛在信息环境下, 如果图书馆不改变信息服务方式与理念, 发展之路可以说是危机重重^[1]。其中, 图书馆传统信息服务不适应大数据时代泛在信息环境的主要表现为: ①服务方式及理念落后, 无法满足读者需求; ②图书馆创新型信息服务方法的普及性较差; ③缺乏高素质的复合型人才。这三方面都导致了信息用户需求与图书馆信息服务之间矛盾的不断扩大, 也使传统图书馆信息服务遭遇“检索代价大”和“信息服务不精准”两大难题。

目前, 越来越多的用户从网络搜寻所需信息, 甚至已经完全取代了其他信息获取方法。如何从海量信息中找到用户所需信息, 如何从同一信息的众多提供者中辨别值得信赖的提供者, 这必将成为图书馆信息服

* 本研究得到湖北省高校图工委项目“兴趣图谱视域下的图书馆泛在服务模式构建研究”(编号: 2014-YB-08)资助。

务发展过程中亟需解决的问题,引起国内外学者的强烈关注。

(1) 兴趣图谱的概念与内涵研究领域: 认同兴趣图谱的理论价值, 界定兴趣图谱的特征与构成要素

目前, 兴趣图谱在学界公认的含义有两点: ①可标识个人身份的、特定的、多样的兴趣; ②试图基于这些兴趣将人们连接起来^[2]。另外, 信息用户的兴趣通常由数学描述模型来表示, 前提是构建出信息用户的兴趣概念层次模型, 明确兴趣概念的界定范围及相互关系, 探索兴趣分类方法, 并最终建立兴趣分类标准^[3]。这就意味着兴趣图谱不是一个通用的、相对静态的兴趣领域本体, 而是反映了用户个体兴趣概念的关联程度以及喜爱程度的动态变化, 具有个性化、动态化的特征。通过用户兴趣图谱的动态演化可以发现、预测信息用户的未知兴趣, 及时有效地更新推荐结果^[4]。另外, 通过信息用户兴趣图谱所推荐的资源是否采纳等行为来进一步调节信息用户对于相关主题感兴趣程度, 可以再次更新用户的兴趣图谱^[5]。

(2) 信息用户资源推荐算法研究领域: 以用户的信息行为为主线, 模拟信息用户兴趣图谱的变化过程

现有信息推荐算法研究大多是基于信息用户的特征或信息行为进行推荐的, 基于兴趣图谱的信息推荐也由此成为国内外近年来一个研究热点^[6]。有些学者认为建立用户兴趣图谱是实现智能代理与主动服务的基础, 在这之前要充分理解用户信息行为数据中所反映出的“兴趣”, 并由此建立用户的描述模型。Lynne 则提出了信息用户兴趣图谱的构建方法, 即通过集合、筛选社交网站上的用户信息行为数据, 定义兴趣图谱的关键元素及筛选、聚类的方法, 从而实现对用户感兴趣资源的自动归类、全面挖掘与分析。Fabrizio认为, 除了对上述数据挖掘, 仍应该进一步地从语义层面对信息用户的兴趣图谱进行跨网站的分析, 最终实现用户新兴趣的发现以及已有兴趣的正确性判断, 从而绘制出更加完整、精准的信息用户兴趣图谱, 并采用混合链路预测和基于内容的扩散激活方法对资源进行推荐^[7]。在兴趣图谱的实践层面, 企业界一直走在前端, 例如, Gravity 公司通过绘制自身企业网站的兴趣图谱, 实现了对用户信息行为的实时追踪, 并将兴趣图谱反映出的结果应用于企业发展战略之中, 比如对合作伙伴的选择等等。除此之外, 国内的腾讯公司、新浪公司也非常重视用户信息行为数据的分析, 以及信息用户兴趣点的漂移, 腾讯公司特别提供了兴趣图谱的API开放接口, 实现了一

键转播热门排行、同话题热门转播排行和通过标签搜索用户等。

但是, 目前的兴趣图谱研究与实践还处于起步阶段, 缺乏统一的用户兴趣建模标准。由于用户兴趣图谱建模的质量直接关系到所推荐信息资源的种类, 因此, 目前也有大量的研究着重于探讨采用何种算法、统计方法对用户兴趣进行建模^[8]。

总体上看, 国外关于兴趣图谱的资源推荐系统已有很大进展, 基于跨网站数据集的信息行为数据挖掘和推荐理论与方法研究也开始被关注, 国内学者在基于兴趣图谱的信息推荐应用方面也有探索。但是, 在兴趣图谱视域下的图书馆泛在服务模式构建研究方面, 国内目前还是空白。

(3) 图书馆泛在服务模式研究领域: 尚属于初步探索阶段, 并且服务方式以传统服务为主, 在服务途径多样化方面稍有进步

国内泛在图书馆的研究由陈维军于2006年首次引入, 这标志着我国泛在图书馆研究的起步; 随后在2008年得到发展, 之后则处于持续升温的状态, 并于2012年达到学术界关注度的最高峰^[9]。但是, 目前图书馆泛在服务模式仍处于初步探索阶段, 这缘于泛在服务模式是泛在图书馆研究内容的延伸与升华, 服务方式依旧以常见的传统服务为主, 虽然在此过程中也增加了以信息技术为特色的服务途径。

综上所述, 泛在服务模式涉及到的层面多、主体多, 以国内泛在图书馆服务模式为例, 需要考虑的要素存在于制度层面、管理层面、不同类型的信息用户、不同内容的特色服务等, 如何融合这些要素, 并最终形成一个科学、合理、高效的图书馆泛在服务模式, 这是值得深思的问题。国外已经将兴趣图谱引入图书馆泛在服务模式的具体工作流程之中, 国内图书馆可以根据各类用户的信息行为来绘制、生成个性化的兴趣图谱, 既可以作为图书馆业务的展现, 也可以优化现有的图书馆业务与各类服务, 将兴趣图谱应用于图书馆的需求, 显得十分迫切。

3 泛在服务模式下的图书馆信息用户兴趣图谱构建

兴趣图谱可以将各类信息用户的需求可视化呈现, 并且从一定程度上发现信息用户的潜在需求, 从而提高图书馆的信息服务质量与服务响应速度。本文将在深

入挖掘、分析图书馆用户信息行为数据的前提下,对比现有图书馆信息服务的现状,探索基于兴趣图谱的图书馆泛在服务方式,并通过兴趣图谱建立不同信息用户以及不同主题之间的相互关联。根据绘制出的兴趣图谱,不仅可以减少图书馆信息服务的成本,还能更最大限度地满足信息用户的信息需求,从而构建出区别于图书馆传统信息服务的图书馆泛在服务模式,具体步骤如下。

(1) 信息用户兴趣图谱的绘制

兴趣图谱的绘制是一个对用户信息行为和所关注主题动态描述的过程。在本研究中,对兴趣图谱绘制过程中的关键元素进行了数学定义,比如用户关注的主题,最频繁出现的信息行为,主题、用户之间的关联度等,从而将用户信息行为转化成了一个可量化的信息搜寻过程,为图书馆泛在服务模式的良好运行提供了科学依据。

准确地生成信息用户的兴趣图谱是开展图书馆泛在服务的前提条件,而较受关注的主题为用户信息行为的描述贡献了大部分的内容和交互信息方面的数据。本研究首先分析了较受关注主题所具备的一些基本特征,在此基础上为这些主题建立了集成动态关联图,并基于此模型生成了针对不同推荐算法的内容兴趣图谱和协同兴趣图谱。下面是信息用户兴趣图谱绘制过程中关键要素的算法表达。

最小支持度 ($\min_support$) 将其作为一个阈值,判定信息用户经常出现的行为有哪些,并通过二元组 ($User_ID, Activity$) 中的 $Activity$ 计算用户关注度;

关注度高的主题 $Active_topic = \{ID | Activity \geq \min_support\}$;

关联度 $W(u_i, u_j)$ 表示两个用户出现在同一个主题中的次数,其中, $u_i, u_j \in User_ID$;

高关注度主题的关联图将用户行为抽象成一个带权无向图 $G = \langle V, E \rangle$, 其中, $V = \{v | v_i \in Active_Users\}$, $E = \{(v_i, v_j) | Active_Users, v_i \in Active_Users, W(u_i, u_j) \geq \min_support\}$;

高频率的用户信息行为 $G = \langle V, E \rangle$ 中的连通子图;

高关注度主题特征词提取数据集 高关注度主题特征词提取过程中,把用户信息行为数据抽象为 $\{n_i, U(n_i), Topic\}$, 其中 n_i 表示主题编号, $U(n_i)$ 是主题中包含的所有用户名, $Topic$ 是主题 n_i 的标题。

本文主要对图书馆信息用户的信息行为数据进行了分析与挖掘,从而发现频率较高的用户信息行为,以

及此类信息行为下所关注的主题是否存在某方面的特别偏好。

(2) 基于兴趣图谱的图书馆泛在服务资源推送算法

首先根据用户信息行为的聚类结果得到用户感兴趣的信息资源集合,然后根据领域本体中概念间的关系分别计算不同信息用户所感兴趣资源间的相似度,最后根据得出的两个相似度值,将该算法应用到图书馆泛在服务模式的构建过程中,从而实现图书馆面向信息用户的泛在服务的精准度。

实现泛在服务的前提是计算出信息用户所关注主题之间的关联度,那么,需要寻找关联规则算法中的频繁2项集。寻找频繁2项集的算法有很多,比如Apriori算法、频繁模式树算法等。为了提高算法的运行效率,本文将使用Apriori算法发现频繁2项集。继而,在得到信息用户所关注主题间关联度的基础之上,构造所有高频率信息行为之间的关联图 $G = \langle V, E \rangle$, 其中, G 是一个有权无向图,顶点 V 是发生频率较高的信息行为,如果在 V 有两个信息用户行为 v_1, v_2 , 他们之间关联度大于或等于最小支持度,即在 E 中存在一条无向边 $v_1 \rightarrow v_2$, 这条边的权重就代表这两种用户信息行为的关联度。具体细节算法如下。

- ①初始化图 $G = \langle V, E \rangle$ 为空, $V = \emptyset, E = \emptyset$;
- ②for 频繁2项集 $(v_1, v_2) \in L2$;
- ③对于图 $G = \langle V, E \rangle$, 添加节点 $v = v \cup \{v_1, v_2\}$;
- ④添加边 $E = E \cup \{v_1 \rightarrow v_2\}$, 2项集 (v_1, v_2) 的支持度作为权重;
- ⑤end;
- ⑥返回强连通图 G 。

(3) 基于兴趣图谱的泛在服务模式构建

①泛在服务方式构建

泛在服务方式即为信息用户兴趣导向下的嵌入式立体化服务,可以引导用户在短时间内获取有用的信息以及利用信息。可对用户提供面对面的咨询服务,也可将参考咨询工具嵌入图书馆网页。这项服务充分体现了图书馆泛在服务无所不在的特征,也是信息用户所亟需的。

②泛在服务内容构建

泛在服务内容不是固定的,而是动态的、变化的,发生变化的依据是信息用户的动态兴趣图谱。当信息用户兴趣图谱发生变化时,图书馆作为拥有大量信息载体的服务机构,也应该调整图书馆的服务方向,提升信

息用户对图书馆服务的“同步感”，从而实现泛在服务的核心意义。

③ 泛在服务途径构建

泛在服务途径是指图书馆可通过移动终端、通讯设备、网络、电视等多种途径开展信息用户的泛在服务，以方便信息用户、满足信息用户的需求为前提，并通过分析用户的兴趣图谱做到准确充分的信息挖掘与传递，从而促使图书馆与信息用户之间形成有效沟通。比如，可通过兴趣图谱的聚类结果为信息用户提供信息分类定制服务，建立信息用户档案库和信息资源库，从而分析、追踪用户的信息需求方向及可能感兴趣的文献资源，实现信息定制服务、推送服务、检索服务、预定事件响应、个人帮助、智能代理等，为实现即时、高效的泛在服务提供有力保障。

4 结语

本文旨在探索一种基于信息用户兴趣图谱的泛在服务模式，并使其具有可实施性、高效性、科学性以及可持续性。系统研究了图书馆信息用户兴趣图谱的形成机理与基础理论，并对图书馆信息用户的“兴趣”范畴进行了科学界定，为泛在服务模式的构建提供了良好的基础；设计了图书馆信息服务的信息资源推送算法，可以主动、批量、精准地满足信息用户的需求，大幅度提

升了信息服务质量；探寻出具有可实施性的图书馆泛在服务模式，将用户的动态兴趣图谱贯穿于图书馆泛在服务的始终，从而有利于图书馆开展针对式服务，解决信息用户在大数据时代搜寻信息的困境。

参考文献

- [1] 王丽丹.图书馆泛在服务探讨[J].科技情报开发与经济,2015(19):30-31.
- [2] 岳春杰,欧阳洪.图书馆泛在服务的构建与读者泛在学习的实现[J].大学图书情报学刊,2015(02):83-85.
- [3] 苏慧.浅析兴趣图谱网站中的社交——以堆糖网为例[J].新闻世界,2014(08):165-166.
- [4] 张佳琦.基于兴趣图谱的社交网站成功要素研究[D].上海:上海师范大学,2014.
- [5] 戴盈.数字化时代读者的选书智慧和阅读兴趣图谱构建[J].中国出版,2014(05):40-44.
- [6] 王静.高校图书馆泛在服务刍议[J].科技情报开发与经济,2015(02):48-49,54.
- [7] 陈媛媛,聂规划,陈冬林等.全网兴趣图谱下电子商务推荐理论研究现状[J].武汉理工大学学报(信息与管理工程版),2014(01):109-113.
- [8] 农秋色.后数字图书馆视阈下高校图书馆泛在知识服务探究[J].兰台世界,2015(20):104-105.
- [9] 孙玉芝.面向用户偏好的高校图书馆泛在服务研究[J].情报探索,2015(09):111-114.

作者简介

闫霏，女，1984年生，研究方向：信息资源管理，E-mail: yanfei_1231@163.com。

Research of Library's Ubiquitous Service Based on Interest Graph

YAN Fei

(Library of South-Central University for Nationalities, Wuhan 430074, China)

Abstract: The ubiquitous library services, which based on interest map could integrate the users' behavior information, deeply mining user interest nodes and recommending the required resources. In this paper, we design an algorithm to draw the library users' interest graph of the information, thereby, a relatively "accurate" ubiquitous library services will be improved, and also will meet the strong demand of users for the need of "ubiquitous" information services.

Keywords: Interest Graph; Library; Ubiquitous Services

(收稿日期: 2015-11-30)