

通过实施推荐系统强化图书馆目录的实用性

——以德国卡尔斯鲁厄大学图书馆为例

□ 王艳翠 / 聊城大学图书馆 聊城 252059

摘要: 2002年至2007年间, DFG资助德国卡尔斯鲁厄大学图书馆开发了BibTip推荐系统, 它是一种有益于增加图书馆目录参考分量的工具。BibTip应用了Web 2.0技术, 有利于读者向导门户——图书馆目录的开发。BibTip是在统计模式的基础上建立起来的, 它可以引导读者用户从检索一条书目记录而附带检索到目录中拥有的相似文献记录。文章对卡尔斯鲁厄大学的BibTip进行了介绍, 并对其在图书馆中的应用进行了研究。

关键词: BibTip技术, 推荐系统, 图书馆目录

DOI: 10.3772/j.issn.1673-2286.2010.09.011

1 推荐系统概述

推荐系统在互联网上得到了蓬勃发展。互联网上很难找到一个不提供推荐建议的网店。推荐系统为客户感兴趣的其他产品提供推荐建议, 从而促进销售量的增长。或许, 图书馆员熟知的最好的例子就是亚马逊网上书店的推荐系统, 几乎每个标题都包含一个链接: “购买这本书的顾客也购买了……。”这些推荐链接都是对亚马逊客户的购买模式做出分析和评价后得出的结论。

互联网出现之前, 我们获得的推荐建议要么是从我们所熟识的人处获得, 要么是从出版的评论性文章中获得。今天的电子推荐系统具备不需人为干预而自动生成建议的功能。推荐系统编制的自动推荐功能基本上是利用对用户的统计数据而自动产生的。这一过程与下列因素相关: 客户、产品及购物车。为了生成推荐建议就需要观察用户选择的、放在购物车中的商品。推荐

系统能否识别出用户在一次购物机会中实际一起购买的商品对于推荐质量是至关重要的。无意识一起购买的商品或许不会形成互惠式的推荐建议; 有意识一起购买的商品有利于生成推荐建议。推荐系统使用统计学的方法区分选择的商品汇编是有意识的还是无意识的。推荐系统所使用的基本统计程序决定生成推荐的质量, 统计程序依据他们所应用的领域做出相应的变化。推荐效率取决于系统能否从大量的点击中区分出相关点击及非相关点击的能力。

图书馆目录非常适合采用推荐系统, 尤其是那些支持用户查找文献目录的服务更适合应用推荐系统。从图书馆员的角度来看, 可以把推荐系统看作是一种强化了(或浓缩了的)目录格式或者是一种传统主题分类法的替代品, 推荐系统协助图书馆员建立并保持最新的馆藏信息。从更专业的角度来看, 由于推荐系统直接使用了图书馆用户间接产生的数据并通过内容

聚合的方式在目录中融入了Web 2.0应用程序的两个方面, 因而可把推荐系统看作是Web 2.0在图书馆中的应用。

2 BibTip推荐系统

2002年至2007年, 由DFG资助的数个项目(这些项目都与为图书馆界开发推荐系统有关)在德国卡尔斯鲁厄大学图书馆实施。最初项目名为“卡尔斯鲁厄大学推荐系统”, 但在2007年的时候项目改名为“BibTip”。该项目是由卡尔斯鲁厄大学图书馆与情报服务及电子市场研究院(The Institute for Information Services and Electronic Markets)的教授Andreas Geyer-Schulz博士合作完成的。情报服务及电子市场研究院负责为系统开发运算法则及科学依据; 卡尔斯鲁厄大学图书馆负责把该系统融入图书馆目录、收集统计数据。最近, 该大学图书馆把BibTip作为一种服务进行开发并实际运转实施, 事实证

明, BibTip项目是非常成功的, 并由DFG于2007年12月在华盛顿召开的网络信息联盟(CNI)秋季工作会议上介绍给与会各国。

BibTip推荐系统是基于用户与图书馆目录互动模式而建立起来的一种“隐式”服务。这种所谓的“隐式”服务是基于对用户模式的观察及对所用数据的统计计算而得出的。所有的数据存储及处理都是匿名的。从技术水平来看, 可认为BibTip架构是一个包含三个软件代理的代理架构。这三个代理分别是: OPAC观察研究数据代理、集体代理、推荐代理。OPAC研究数据代理负责研究限定时间段内OPAC书目数据的选择, 这些数据被传输至集体代理, 集体代理再根据统计资料计算出一份推荐名单, 最后, 推荐代理把推荐名单传送给用户(如图1)。

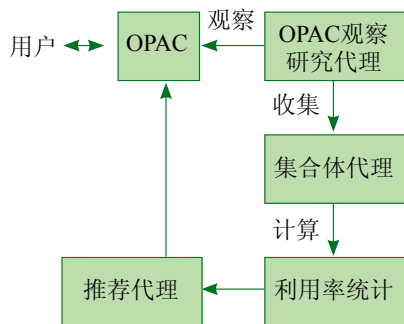


图1 BibTip的架构示意图

互联网电子商务领域根据购买交易数或所链接网页的点击数生成用户数据; 而BibTip则是根据多个时间段内汇集的需要显示完整联机书目的数量生成数据。用户关于目录中的完全书目显示的观点也被计算在内。如果两条书目至少在一个时间段内被认为是一致的, 就会

据其建立起多对书目之间的共现矩阵模式, 然后更进一步地对这种矩阵进行评价, 从而生成推荐建议。在这一步中所使用的运算法则发挥着至关重要的作用。BibTip所基于的算法过程有着强烈的抗干扰性, 其开发的算法尤其适用于图书馆目录数据。BibTip在Andrew Ehrenberg“重复购买理论”^①的基础上, 使用了泊松框架中独立交易过程基本理论分布以及对数及数分布算法, 使得推荐系统实现了完全自动化、高度准确并具备稳定的防误差性能。

系统经过计算后就会在OPAC的完全书目显示中生成超链接推荐。这些链接把用户引导至具备相关内容的书目。BibTip根据普遍使用度列出推荐建议, 这就意味着最顶部显示的书目就是最经常检索的书目。

3 BibTip推荐系统的技术层面

BibTip系统在技术方面的基本要求如下: (1) OPAC与网络的快速、有效连接。(2) OPAC有可识别的全文显示(OPAC操作者可以变更全标题表示的HTML代码)。(3)通过ISBN或书目标识符实现标题与全文显示之间的永久链接。

BibTip有2种形式的集成(表1): 通过OPAC的书目标题标识符实现的标准集成; 以及使用ISBN简化版的集成。简化了的ISBN集成允许ISBN在全标题显示中只有2个额外的“静态”HTML代码行(此处的“静态”指的是在每个完全书目代理中这些代码行总是相同的)。通过书目标识符实现的集成包含了根据原文显示而变化的动态部分,

它需要另外三个或更多书目标识符的HTML代码。BibTip通过这2种形式的集成实现OPAC网络与互联网的快速、可靠的连接, 从而实现从书名到完全代理的永久链接。

3.1 ISBN集成

ISBN集成的情况下, 需要在以下几点中限定书目标识的HTML源代码: (1)要从BibTip服务器中插入加载了JavaScript代码的脚本标记。(2)为推荐表的插入提供一个Div(分层)标记。

下面是一个简单的用HTML源代码表示的ISBN集成^②:

```

<body>
  Content Full title
  <div style="display:none"
  id="bibtip_reclist"></div>
  Content Full title
  <script src="http://
  recommender.ubka.uni-karlsruhe.de/js/
  bibtip_xxx.js"
  type="text/javascript"></script>
</body>
    
```

3.2 BibTip标准集成

在标准集成情况下, 一个OPAC全标题显示要求如下扩展: (1)在BibTip服务器上插入一个脚本标记以加载JavaScript代码。(2)为推荐表的插入提供Div(分层)标记。(3)为书目标识、ISBN及短标题提供Div标记。简例如下^③:

```

<body>
  Content Full title
  <div id="bibtip_isxn"
  style="display:none">3-8266-1762-2,9
    
```

^① “重复购买理论”是荷兰著名的市场营销学家Ehrenberg, Andrew A.S.C.在1972年出版的Repeat-buying: theory and applications中初次提出的。内容是: 在影响消费者行为的1001个可变因素中, 其中的999个因素并不重要。消费者行为的许多方面可以从平均购买率和产品普及程度这两个变量进行预测(详见文后参考文献[9])。

^② http://www.bibtip.org/bibtip_flyer_en.pdf

```
78-3-8266-1762-1</div>
<div id="bibtip_shorttitle"
style="display:none">Ajax ge-packt /
Seeboerger-Weichselbaum,
Michael; 2007</div>
<div id="bibtip_id"
style="display:none">26214927</div>
Content Full title
<script src="http://
recommender.ubka.uni-karlsruhe.de/js/
bibtip_XXX.js"
type="text/javascript"></script>
Content Full title
<div style="display:none"
id="bibtip_reclist"></div>
Content Volltitel
</body>
```

4 卡尔斯鲁厄大学图书馆使用BibTip的经验

德国卡尔斯鲁厄大学图书馆建立的OPAC汇总了德国境内12所图书馆的网上书目（12所图书馆书目又是各个州书目的汇总）、奥地利和瑞士德语区内的书目，它还提供世界上一些其他重要图书馆的书目（如“大英图书馆”、“美国国会图书馆”和几个不同语种国家的图书书目）。读者可通过“卡尔斯鲁厄大学图书馆网上书目”检索到几乎全德国乃至世界范围内图书馆的图书。显然，BibTip为OPAC的图书馆用户提供了增值服务。

对于图书馆来说，BibTip的实施有几大优点：（1）推荐建议永远不会过时。与传统分类表的标记法不同的是，BibTip推荐正在不断地进行重估并根据用户利用文献的变化作出动态调整。如果用户在与以前不同的背景下利用文

表1 ISBN集成与标准集成形式的概况

时间	ISBN集成	BibTip标准集成
识别码	ISBN/ISSN	OPAC内部书目标识符
推荐的可获得性	仅可获得带有ISBN/ISSN的标题	所有的标题
推荐表的定制	否	是
静态集成	是	否

献，就会在BibTip提供的推荐名单上反映出来。（2）由于推荐数目代表了一种书目使用率的测试方法，因而BibTip有助于发展图书馆的馆藏。如果一条书目没有任何的推荐建议，说明用户对它似乎没有多大的兴趣；而另一方面，对某一特殊书目的大量推荐则表明对该书的使用甚多，从而可根据推荐数作为调整馆藏方向及馆藏结构的重要参考。（3）推荐建议完全是中立的。BibTip可以为目录中的任一条目（无论该条目是书籍、视频资料、音频资料还是期刊）创建一份推荐清单。在所有含有参考文献资料数据的数据库中均可使用BibTip。（4）最后值得一提的是：由于BibTip推荐是一个完全自动化进程，因而非常符合成本效益，一旦建立，不需要任何人去操作。

BibTip根据对统计数据的分析及评价（而不是根据流通数据）对目录的使用作出推荐。这是由于根据流通数据作出推荐的方法有如下几个缺点：（1）不能对未借出的参考书籍及其他文献做出有效的观察。（2）由于通常书目检索数量多于图书馆实际借出书的数量，因而使用目录数据积累的数据量更高一些。（3）此外，如果根据流通计算得出推荐书目，那么推荐书

目就会受到欲借阅资料的可获得性影响。如，由于用户倾向于借出他们目前易获取的书籍，因而复本量较少而需求量较高的书籍的流通量就会远远排在那些复本多、趣味性较低、较易获取到的书籍之后。

（4）用户数据的安全性和隐私性是推荐系统需要考虑的一个重要方面。德国在这方面的立法极其严格。但由于BibTip仅收集和储存那些OPAC机构匿名数据、书目标识部分及书籍标识部分的数据，因而BibTip没有法律方面的难题。如果BibTip使用流通数据，就必须匿名使用这些数据以满足数据安全性与隐私性方面的考虑。

由于系统根据大量的业务处理数量而计算得出统计数据，而且这类统计数据富有意义，因而BibTip与其他的推荐服务一样：数据库越大，推荐效果越好。因此，在展示首次推荐之前，对于收集并分析的数据必须有一个观察期。观察期较短并且在此期间存储了更多的数据（即目录被更为集中地使用）。这些数据被充分利用的必要条件是由大多数图书馆（起码是那些大学图书馆及学院图书馆）来完成的，此外，数据的数量及利用率、OPAC时间段内提出的书目数量，以及书目的多样性都影响观察

③ http://www.bibtip.org/bibtip_flyer_en.pdf

期的持续时间。需要指出的是,没必要为OPAC中的所有书目进行推荐,对于整个馆藏来说,目录数据的援引比率通常会服从帕累托最优分布:大约80%的检索影响了目录中仅20%的书目。由此,足可为这20%的书目提供充分的推荐目录名单。卡尔斯鲁厄大学图书馆目录中共包含了1,000,000条目录,而该大学图书馆目录大约有200,000份推荐名单,达到了80%的覆盖范围,符合帕累托最优法则。然而,正确预测观察期的时间跨度则是相当的困难。其中一个加快数据建立生成推荐的方法是从多个图书馆中汇集统计数据。这种办法对于那些在书目选择上有共同兴趣并有相似用户及相似用户行为的图书馆(如公共图书馆)来说尤其可行。

5 用BibTip作为一项网络服务

BibTip是德国卡尔斯鲁厄大学

图书馆OPAC富有吸引力的重要特征之一,它在提高对用户的服务质量方面大有潜力。2005年至2006年间,卡尔斯鲁厄大学图书馆在读者用户中进行了一项调查并要求用户对推荐服务进行评价打分(1分表示最差;5分表示最优),结果显示,大多数的用户发现推荐服务非常有益,用户评价的平均分达到了4.21。卡尔斯鲁厄大学决定在DFG结束对其资助后仍然继续推荐服务的开发,并把BibTip提供给其他有兴趣为用户提供推荐服务的图书馆。为此,卡尔斯鲁厄大学图书馆对软件进行了部分重新设计,使它能作为一项网络服务,使其更灵活并可进行升级。迄今为止,包括德国国家图书馆、弗莱堡大学图书馆、布伦瑞克大学图书馆、柏林大学图书馆、汉堡联邦军事大学图书馆、斯图加特图书馆、巴登州图书馆以及巴登-符腾堡及柏林地区的高校图书馆等相当多的图书馆早已应用了BibTip推荐,BibTip在图书馆团体中越来越广泛的应用为其将来的

进一步发展奠定了坚实的基础。

6 未来与展望

随着使用BibTip图书馆数量的增长,BibTip的进一步发展有了更多的方法与选择。其中之一就是从多个图书馆中汇集数据的机会,以及在多个图书馆中生成跨图书馆推荐建议。这样就会吸引具有类似用户概况的图书馆(如公共图书馆、具有相似学科设置或类似研究重点的高校图书馆)为用户提供推荐服务。如此一来,只有少量目录或目录使用率低的图书馆也可为他们的用户提供推荐服务。

BibTip未来发展的另一个领域是为图书馆的IT部门提供先进的管理工具,该工具能够自定义配置并根据推荐代理生成文本而不受图书馆工作人员的干扰。BibTip进一步的科学研究包括:分析并降低有效推荐生成需要的形成时间;或者把推荐建议群进行聚类从而为目录创建学科检索切入点。

参考文献

- [1] MÖNNICH M, SPIERING M. Adding Value to the Library Catalog by Implementing a Recommendation System [J/OL]. D-Lib Magazine, 2008(14):5-6 [2010-03-16]. <http://www.dlib.org/dlib/may08/monnich/05monnich.html>.
- [2] DFG, German Research Foundation [EB/OL]. [2010-03-16]. <http://www.dfg.de/en/index.html>.
- [3] KIT Library : Homepage [EB/OL]. [2010-03-16]. http://www.ubka.uni-karlsruhe.de/index_engl.html.
- [4] Information Services and Electronic Markets [EB/OL]. [2010-03-16]. <http://www.em.uni-karlsruhe.de/index.php?language=en>.
- [5] CNI Fall 2007 Task Force Meeting [EB/OL]. [2010-03-16]. <http://www.cni.org/tfms/2007b.fall/>.
- [6] Recommender Systems for Meta Library Catalogs [EB/OL]. [2010-03-26]. <http://www.em.uni-karlsruhe.de/research/projects/reckvk/index.php?language=en>.
- [7] KIT-Bibliothek: Bibliothekskataloge : KIT-Katalog [EB/OL]. [2010-03-26]. <http://www.ubka.uni-karlsruhe.de/katalog/>.
- [8] BibTip flyer in English [EB/OL]. [2010-03-26]. http://www.bibtip.org/bibtip_flyer_en.pdf.
- [9] EHRENBURG A S. Repeat-buying: theory and applications [M]. Amsterdam: North-Holland Publ., 1972.

作者简介

王艳翠,女,馆员,1998年山东大学信息管理学系毕业,现聊城大学图书馆资源建设部工作。通讯地址:山东聊城大学图书馆 252059。E-mail: wangyancui007@163.com

Adding Value to the Library Catalog by Implementing a Recommendation System – Example of the Library in Karlsruhe University

Wang Yancui / Liaocheng University, Liaocheng, 252059

Abstract: Recommender systems are useful tools for adding a reference component to a library catalog, and they help develop library catalogs that serve as customer-oriented portals, deploying Web 2.0 technology. Recommendation systems are based on statistical models, and they can lead users from one record to similar literature held in the catalog. In this article we describe the recommendation system BibTip, developed in Karlsruhe University, and we discuss its application in libraries.

Keywords: BibTip technology, Recommendation System, Library catalog

(收稿日期: 2010-05-17)