

·信息工作·

个性化关键技术研究综述

苏玉召 (中国科学院国家科学图书馆 北京 100190)

赵妍 (郑州航空工业管理学院 河南郑州 450015)

摘要: 个性化研究的关键技术主要包括有三个方面的内容: 用户建模、推荐系统和评价。首先, 用户建模是收集用户数据, 目的是为推荐系统提供用户的兴趣、偏好等建立模型; 其次, 推荐系统的作用是根据用户模型为用户推荐个性化内容, 是个性化的核心内容; 最后, 评价系统根据用户对个性化应用的满意程度, 反馈给推荐系统调整建模策略。个性化应用是一个不断反馈修正的过程。

关键词: 个性化 Web 挖掘 推荐系统 用户建模 个性化评价

中图分类号: G350; TP311

文献标识码: A

文章编号: 1003-6938(2011)01-0059-07

Survey of Key Technologies in Personalization Application

Su Yuahao (National Science Library, CAS, Beijing, 100190)

Zhao Yan (Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou, Henan, 450015)

Abstract: The key technology of Personalization Application includes three aspects: User Modeling, Recommendation System, and Personalization Evaluation. First, User Modeling collects user personal data, and provides user's interests and preference to Recommendation System. Secondly, Recommendation System is the most important part of Personalization, which recommends personalized content for individual. Finally, Evaluation provides the feedback information of user's evaluation for Recommendation System to help it modulating the strategy of User Modeling. Personalization application is a process with continuous modulating its recommend strategy according to user's feedback.

Key words: personalization; web mining; recommendation system; user modeling; personalization evaluation

CLC number: G350; TP311

Document code: A

Article ID: 1003-6938(2011)01-0059-07

1 引言

个性化最初起因于基于 Web 应用的商务智能(BI)发展, 获得商业利润最大化的需要。电子商务(e-commerce)和客户关系管理(CRM)网站采用数据挖掘的方法发现顾客与商品之间的关联和分析用户购物行为, 销售更多的商品和吸引顾客而不断调整营销策略。随着数据挖掘理论及其技术进步不断深化, 计算机网络和网络信息管理技术的发展, 基于客户/服务器模式为用户和网站提供更多的交互机会。于是针对每个用户的个性化服务便迅速发展起来。

关于个性化概念, 不同的学者和研究人员, 在不同的时期对其理解, 及其所持观点也各不尽相同。有些从事 IT 行业的专家认为个性化是一种能力, 例如, 前剑桥福雷斯特研究公司的分析师 Paul Hagen 在 1999 年的一次电子商务会议上说, “个性化是一种能力, 根据每个用户的爱好和行为知识, 对其提供简洁的内容和服务。”长期从事 IT

行业实践和研究的专家 Jill Dyche 女士, 2002 年由艾迪生-韦斯利出版公司出版的“CRM 指南”, 认为“个性化是一种定制用户交流的能力, 这种能力的实现是借助于用户进行交流时的爱好和行为知识的获得。”而有些计算机技术专家认为, 个性化是一种服务。例如, 前 IBM 沃森研究中心的研究人员 Doug Riecken 在 2000 年美国计算机协会(ACM)通信年会上提出, “个性化是关于通过建立一种有意义的一对一关系, 从而建立用户的忠诚度。同时, 通过理解每一个用户特殊需要的内容, 为其提供高效、有价值的服务。”一些从事个性化技术研究的专家学者认为, 个性化是一种技术。例如, 在 2003 年的个性化智能技术国际研讨会上, 各国的学者和研究人员就个性化概念达成共识, 认为“个性化是一种技术, 这种技术根据用户的信息定制商业和每个用户电子商务的交互。用户的信息可以是以前获得的, 也可以是实时方式获得信息。根据可用的用户信息, 商业不同部门之间进行变更以适应用户的需求。”还有的一些从事信息技术研究的专家和学者认

收稿日期 2010-08-22, 责任编辑 宋戈

为,个性化是一种过程。例如,Kwon 分别在 2007 年和 2009 年美国信息技术协会(ITAA)年会上分别提出“个性化是一个过程,在电子服务应用方面,是一个应用用户信息为其发送目标方案的过程。”这些关于个性化的概念重点是从个性化是什么和个性化能够做什么方面定义,但是,这些观点没有从系统化、全面性定义个性化。

美国芝加哥德保罗大学的 Bamshad Mobasher 教授从 20 世纪 90 年代就开始个性化的研究,他对个性化的特点、历史、现状和未来的发展趋势具有全面和独到的见解。因此, Mobasher 对个性化的定义具有重要的参考价值,他认为:“在 Web 上下文中,个性化意味着动态内容的发送,例如文本元素、链接、广告和产品推荐等,这些内容专门为特定用户或者一部分用户的需要或者兴趣定制的过程”。^[1]他把个性化过程看作是一个包含数据挖掘循环所有阶段中的一个典型应用。这些阶段包括数据收集、预处理模式发现、性能评价和在用户和 Web 网站之间应用实时发现的知识。

主要研究个性化的 ITWP 国际会议重点内容是个性化理论和技术,其每一次会议的成果都体现了当时的研究方法和应用领域。典型的个性化研究代表是 Mobasher 教授和其他个性化研究的专家学者等研究人员,同时, Mobasher 教授一直作为 ITWP 组委会主席,许多他本人及其合作者的研究成果及文献可以通过其个人网站获得。因此,他们的个性化研究成果具有权威性和代表性。研究他们的成果对于我们未来个性化研究具有借鉴意义。总结近几年个性化研究内容,其关键技术主要包括推荐系统、用户建模、评价等三个方面。

2 个性化推荐系统

个性化推荐是 Web 挖掘结果呈现给用户的应用,其使用的数据主要来自 Web 网站与用户行为交互生成的数据。其依据原理是数据挖掘理论基础。因此,个性化推荐使用的算法主要来自数据挖掘理论的算法。通常,个性化推荐分为基于规则过滤、基于内容过滤、基于协作过滤的方法,以及这三种方法混合的推荐方法。根据不同的个性化推荐方法,采用的挖掘算法也各不相同,常用的算法类型包括关联规则发现、聚类分析、序列模式挖掘和回归分析等。

2.1 Web 挖掘

许多学者认为, Etzioni 是第一个提出 Web 挖掘(Web Mining)技术的人,^[2]他在 1996 年的描述中认为, Web 挖掘是应用数据挖掘技术,从 Web 文档和服务内容里自动发现并抽取有用信息。一般情况, Web 挖掘技术可以分为

3 种类型^[3]: Web 内容挖掘(Web Content Mining)、Web 结构挖掘(Web Structure Mining)和 Web 使用挖掘(Web Usage Mining)(Kosala&Blockeel,2000)。Srivastava 等人对 Web 使用挖掘的定义是^[4]: Web 使用挖掘是应用数据挖掘技术,为了更好的理解和服务基于 Web 应用的需要,发现 Web 数据的使用模式。

Web 内容挖掘包括应用数据挖掘技术从 Web 对象中抽取模型, Web 对象内容包括普通文本、半结构化文档(如 HTML 和 XML)、结构化文档(如数字图书馆)、动态文档和多媒体文档。Web 结构挖掘的目标是发现隐藏在 Web 对象之间相互连接的拓扑结构建模,对 Web 网站进行分类和分级,以发现对象之间的相似性。Web 使用挖掘是应用数据挖掘技术,从 Web 数据中发现用法模式。进行挖掘的数据通常来自用户与 Web 对象的交互行为,比如, Web 服务器或者代理服务器日志、用户查询、注册数据。Web 使用挖掘工具发现并预测用户行为,帮助设计者进行统计分析、改进 Web 网站、吸引访问者或者为有规律的用户提供个性化服务。Web 使用挖掘与 Web 结构挖掘和 Web 内容挖掘的区别在于, Web 使用挖掘反映的是人与网络交互的行为,对用户行为与 Web 网站交互的分析可以深入理解定制与个性化的用户 Web 体验。

Web 使用挖掘使用的数据可能来自 Web 服务器访问日志、代理服务器日志、引用页日志、浏览器日志、错误日志、用户资料、注册数据、用户会话、用户交易、cookies、用户查询或者是用户书签数据(Gunduz,2003)。通过分析这些文件和文档,我们可以获得用户感兴趣的使用模式和信息。Web 使用挖掘最初在电子商务领域, Web 使用挖掘吸引了销售决策者和电子商务专家的强烈兴趣。^[5]随着数据挖掘理论和网络信息管理技术的发展,个性化研究现在已经渗透到各个行业和领域,例如,电信行业、银行金融机构、移动设备和反恐领域等。

2.2 基于规则过滤的技术

基于规则过滤方法的典型例子就是“购物篮”算法,通过查找购买物品之间的联系为用户推荐。其特点在于采用关联规则、回归算法,找到用户对相关联物品之间的兴趣度、发现用户偏好,预测用户未来行为。

Peng 等人提出的一种个性化推荐模型采用 apriori 算法和用于统计的 tf-idf 加权技术,包括三部分:资源描述、用户偏好抽取和个性化推荐。首先,通过分析挖掘用户 Web 访问日志获取的资源信息,生成资源文本空间向量;然后,采用 apriori 算法对这些向量进行计算,获得兴趣集;最后,根据推荐模型,把推荐过滤和存储的资源内容推荐给用户。^[6]2003 年华墨西哥阿卡普尔科举办的第 2

届个性化智能技术(ITWP2003)研讨会上^[7]Mobasher等人基于关联规则和序列模式的个性化推荐模型对网站的影响进行了研究。^[8]Forsati等人提出一种用于个性化基于权重的关联规则算法,该算法是对传统的关联规则算法的扩展,允许交易中的每一个项目分配一个权重以反映用户对该项目的兴趣度。在结果关联规则集里每一个项目都对应一个权重参数,根据用户的兴趣程度,为每个用户访问的Web页面分配一个时间权重和访问频率权重。这种方法能够客观地、更有效的表示预测结果,对推荐系统效率有很大改进。^[9]

2.3 基于内容过滤的技术

基于内容的过滤推荐技术特点是^[10]根据用户过去选择项目的特点,从项目描述、终端数据库里的项目属性关联抽取项目内容特点,系统为其推荐相似的项目。基于内容的过滤系统最大的缺点是用户模型的建立过度依赖于用户以前选择和点击的具体项目。此外,基于内容过滤技术要求项目能够有效代表抽取的文本特点,Web数据异构的特性决定了这种方法并不总是有效。更详细的基于内容的过滤系统内容参见文献。^[11]

2.4 基于协作过滤的技术

维归约、关联挖掘、聚类 and 贝叶斯学习是协作过滤系统采用的一些技术。协作过滤推荐系统利用了用户的相似性进行推荐,而基于内容过滤的推荐系统根据商品内容的相似性进行推荐。

Amazon是使用协作过滤推荐系统的先驱,作为市场战略的一部分,提供“针对每位顾客的个性化商店”,这种方法即有益于顾客又有益于公司。公司拥有顾客更准确的模型,可以对顾客的需求有更好的了解。而服务于这些需求则可在产品的交叉销售、提升销售、产品亲和力、一对一促销、大购物篮和顾客忠诚度方面获得更大的成功。2001年美国西雅图举办的第1届个性化智能技术(ITWP2001)研讨会上^[12]Mobasher等人的根据匿名Web使用数据提高协作过滤效率的研究,根据数据挖掘和Web技术实现个性化协作过滤系统。Aghabozorgi等人提出一种基于Web使用挖掘的离线个性化系统动态模型,该模型采用聚类算法。其特点是,根据用户交易,把离线模型周期性地转变为动态模型,解决了离线模型随着时间的流逝,一些新用户的加入和原有用户行为的改变引起的个性化用户行为预测精度下降问题。^[13]

协作过滤推荐系统^[14]寻找与目标顾客历史吻合的顾客群组(称为近邻),比如他们购买相似的商品或对某种商品的评价。协作过滤推荐系统面临两个主要挑战:可伸缩性和确保对顾客推荐的质量。可伸缩性重要原因是,

电子商务系统必须能够实时搜索数以百万计的潜在近邻。如果网站使用浏览模式作为产品偏爱的指示,则对某些顾客可能会有数以千计的数据点。保证推荐质量是赢得顾客信任的基本要素。如果顾客听从了系统的推荐但是最终并不喜欢这些推荐的商品,那么他们就会对该推荐系统产生怀疑,甚至不用。有些推荐系统采用基于内容过滤和协作过滤的推荐技术,进一步改善系统的推荐。

2.5 基于混合过滤的技术

基于规则过滤和基于协作过滤技术的缺点是,通常需要离线学习训练样本生成关联规则或者用户组的分类,这种方法具有可扩展性差和动态变化缓慢的特点。因此,不适于在线的迅速响应,尤其不适合实时查询的响应。基于内容过滤技术的优点是及时响应快,总是处于在线状态,但是,其缺点是依赖于用户的历史操作记录,如果用户的兴趣偏好变化或者历史记录缺失,推荐给用户的结果将会产生严重的失真。因此,结合这三种方法,将会有较快的响应速度、较好的推荐准确性和较高的预测精度。

当前解决基于内容过滤和协作过滤不足的研究热点是采用混合推荐算法,目的是提高推荐的精度。Burke提出的方法是混合基于内容和协作过滤技术,通过丰富变量的方法生成推荐系统,旨在提高推荐的质量。^[15]有的混合推荐系统,例如,Ardissono等人的用户建模和个性化推荐技术研究^[16]通过收集多种用户偏好的信息,采用多种异构推荐技术的方法实现。这种方法越来越多地被用于各种个性化服务中,例如,Nima等人关于Q-learning的Web推荐系统研究^[17]Chen等人的手机新闻混合推荐系统的普适访问研究。^[18]Mobasher等人提出一种框架,把在线用户访问Web页面的会话活动过程与在线数据准备和数据挖掘任务分开处理,基于聚类的技术把Web站点使用和模型采用统一的表示方法,用于实时的个性化操作。这种方法的优点在于Web站点的使用和内容特点集成到Web挖掘框架里,为推荐引擎提供统一的访问方式,从而提高了个性化的效率。^[19]2009年美国帕萨迪纳举办的第7届个性化智能技术和推荐系统(ITWP2009)研讨会上^[20]Rosenthal等人的提出的研究旨在提高推荐系统在线预测精度。

3 个性化用户建模

个性化推荐进行Web使用挖掘的基础是用户建模,根据用户模型进行挖掘,发现用户的兴趣和偏好,预测其未来的访问行为,并为其推荐个性化信息。个性化用户建模需要根据推荐系统的需求收集Web使用数据。所有个

个性化推荐方法的好坏在很大程度上依赖于收集的数据质量。如果收集到的数据能够准确地反映出用户兴趣和偏好,则有助于个性化推荐系统性能的提高,否则,个性化系统推荐给用户的信息不能满足要求,用户可能逐渐失去个性化推荐的兴趣,也可能有糟糕的推荐会激怒用户,最终会导致用户根本不用该个性化推荐系统。

用户模型与用户偏好和兴趣的数据相关,这些数据的收集不应该过于强调用户自身数据。收集用户数据常用的方法有两个^[21]—一是显式询问他们的偏好,二是隐式监测用户的行为推到他们的偏好。

显示询问用户偏好信息的方法有几种:具体兴趣的调查问卷,例如,喜欢体育新闻或者科技新闻;编辑用户关键词,或者直接询问用户具体的兴趣。这些方法都有其局限性,首先,因为这些方法都无法准确地动态反应用户兴趣和偏好的变化。并且,通过这些方法收集到的数据可能会引起对用户偏好错误的判断,因为并不是所有的用户提供可靠的信息。其次,这些方法重点都是依赖于来自用户的信息,因此,收集的用户模型数据片不全面。最后,并不是所有的用户都愿意显式地提供他们的偏好信息。

隐式收集用户偏好是通过一种特殊的智能代理监测用户行为,从收集到的用户信息推导用户的偏好。用户行为有几个方面可以被监测到,例如一篇文档操作的滚动、停留在该文档上的时间。但是,这些方法可能会引起误导,因为这些方法并不能反映用户是否真的感兴趣。

建立用户模型需要的信息可以归纳为三个方面:①用户个人信息,例如,用户名、性别、出生日期和地址等人口统计学信息;②用户历史记录,包括过去交易成功的记录、浏览的记录和收藏的记录,例如,在电子商务网站上购买喜爱的商品类别或者购买数量等;③用户定制的兴趣偏好,例如,iGoogle定制的音乐盒、谷歌翻译/字典、背单词、博客主题订阅、网站导航、BBS导航等。这些信息大部分包含个性化应用收集到的事实数据,也包括从过去交易中推导出来的事实数据。个性化应用程序收集这些信息完成后,记录到Web日志文件,可以进行日志清洗、处理,并存储到数据仓库中以便进一步为推荐系统进行数据挖掘使用。

如2.1节所述。在个性化推荐的Web使用挖掘研究中,数据的收集可以在服务器端、客户端、代理服务器端,或者是集成Web服务器的数据库进行。Web使用挖掘的数据源主要包括用户查询、注册的数据、Web服务器访问日志和Web应用服务器日志记录的服务器日志文件^{[22][23]}。收集Web使用数据,特别是Web日志,对于个性化预测和推荐的精度具有非常重要的作用。Lathia等人对个性化

协作过滤算法进行研究^[24]发现数据的质量和数量对推荐系统的精度有重要影响。在他们的自适应信息源的协作过滤研究中,发现推荐系统关注数据质量而不是算法本身,能提高预测的精度。

近年来,关于个性化用户建模的研究成果不断出现。例如,2008年美国芝加哥举办的第6届个性化智能技术和推荐系统(ITWP2008)研讨会上,^[25]Gemmis等人的基于用户概要的个性化搜索信息抽取模型^[26]Hung等人的基于标签用户建模的社会媒体推荐系统^[27]Kirmemis等人的基于内容的用户模型生成和电影推荐系统优化方法^[28]这些研究都促进了个性化用户建模研究的发展。

当前关于个性化用户建模不足之处在于,关于Web日志数据需要记录什么样的数据缺乏理论研究。因为不同的应用环境中评价个性化推荐性能的指标不同,所以,Web日志记录采用的格式也不相同。例如,在线购物网站想发现商品关联规则实现捆绑销售策略,那么,Web日志记录顾客在买了书籍之后还买了什么关键因素;对于移动服务供应商来说,要实现增值服务,Web日志记录顾客对什么类型的资费套餐感兴趣是关键因素。

4 个性化评价

在个性化应用被部署之前及应用过程中,评价其准确性和效率至关重要,需要不断的调整个性化策略以实现性能更好的个性化推荐和预测结果。早期的个性化评价研究的重点是推荐系统采用算法的精度提高,但是,好的推荐系统还应该同时采用其它的方法。Franc等人认为评价方法还包括^[29](1)覆盖范围,测量推荐系统数据集为用户推荐包含内容的广泛程度;(2)信任值,帮助用户更有效的决定是否采用推荐的内容;(3)计算时间,测量推荐算法生成好的推荐内容的时间;(4)新颖,测量推荐系统是否采用新技术,具有发展前途;(5)健壮性,测量在有噪声干扰和稀疏数据情况下,推荐算法是否做出好的预测;(6)采用协作推荐方法,当前个性化推荐系统研究主要采用多种方法相结合的协作过滤方法。研究结果证明,综合采用这些方法能实现较好的个性化推荐质量。

许多个性化推荐系统的评价要求用户参与,对个性化推荐系统提供的产品或者服务质量做出明确评价。这样的方法的实际效果并不明显,因为系统提供的信息列表可能占用用户很多时间,有些用户可能不耐烦,敷衍了事。所以,个性化评价应该采用尽可能消耗用户时间少,较少的操作、用户反馈的信息能够体现用户兴趣偏好的方法。^[30]Kwon等人提出的如何最好体现电子商务个性化特征的方案^[31]个性化策略的效果可以通过用户的满意

度和忠诚度来体现,系统对用户发送调查问卷,回答其对个性化的满意度并判断其忠诚度。

当前,个性化模型的评价工作面临诸多挑战。首先,不同的建模方法和推荐算法可能需要不同的评价因子(如用户的兴趣、偏好、是否选择推荐的内容、个性化评分等)。其次,个性化活动需求可能在不同的应用领域、特定的应用和数据收集的不同而有很大的区别。最后,研究者和开发者对于个性化系统中什么因素对服务质量的影响最大缺乏共识。基于以上原因,个性化评价标准体系还需要进一步深入研究。

5 个性化应用和发展趋势

随着个性化研究理论的不断深化,个性化应用逐渐发展到各种行业和领域。这些行业和领域包括(1)电子商务(如 Amazon, www.amazon.com)(2)公共服务(如英国的国家健康卫生网站, www.nhs.uk)(3)搜索引擎(如 iGoogle, www.google.com.hk)(4)图书馆(剑桥杂志在线, journals.cambridge.org 美国国家医学图书馆, www.ncbi.nlm.nih.gov)等,还有一些网站正在准备建立自己的个性化应用(中国数字图书馆, www.d-library.com.cn 中国科学院国家科学数字图书馆 CSDL, 159.226.100.28)(5)生物医学(5)电子学习(E-learning)(6)移动设备(如 3G 手机个性化)(7)电子地图(8)居家生活等。

近几年的研究也出现了一些新的研究方向。在 2005 年苏格兰爱丁堡举办的第 3 届个性化智能技术(ITWP2005)研讨会上^[32]出现了个性化安全方面的研究,例如 Burke 等人的协作过滤系统中欺骗攻击研究。^[33]

研究发现,基于记忆的标准协作过滤算法,诸如 K 最近邻算法,易于受到入侵攻击,但是,基于模型的技术比 K 最近邻算法有较强的健壮性。Sandvig 等人采用数据挖掘技术的关联规则算法,发现采用基于模型的 Apriori 算法在稳定性和健壮性方面比 K 最近邻算法更好。此外,实验结果还显示 Apriori 算法比 K 最近邻算法的推荐精度更高。^[34]

Mobasher 于 2005 主持,由美国国家科学基金支持的“个性化安全”研究项目在 Microsoft 的推荐系统上实现^[35],重点是通过攻击建模分析对各种推荐算法的影响。研究结果显示,基于用户和项目的算法在特定的攻击模型下很脆弱,但是,采用混合算法的推荐系统具有较强的健壮性。

在 2006 年美国马萨诸塞州波士顿举办的第 4 届个性化智能技术(ITWP2006)研讨会上^[36]开始出现了语义和本体技术在个性化协作过滤方面的研究。例如,Toivonen 等人的本体角色在个性化内容过滤应用需求的研究,^[37]

Symeonidis 等人的基于潜在语义标引的可扩展协作过滤研究。^[38]

在 2008 年美国芝加哥举办的第 6 届个性化智能技术和推荐系统(ITWP2008)研讨会上^[25]个性化研究开始采用社会标签的社会网络等新技术。例如, Hung 等人的基于标签用户建模的社会媒体推荐系统^[39] Sandvig 等人的分析社会标签系统攻击的框架研究。^[40]

移动设备促进了普适计算的发展,这使得不同的设备在任何时间、任何地点访问服务成为可能。^[41]个性化服务的提供也应该根据需要进行定制,特别是应该满足移动用户在城市中步行或者开车的特定需要。一些鼓吹者认为,在未来,除了普通人群需要把普适计算集成到服务中,并且普适计算将会在学术和公司的实验室里发挥重要作用。这种新环境要求服务提供者考虑新的策略适应新特点,例如用户的位置、移动速度、环境条件(光线、噪音)特别是不同特性的移动设备访问个性化服务的需求也不一样。^[42]

在未来的普适计算环境中,一些影响服务质量的必要数据可能由于以下两个原因而无法获得。一是数据不存在,二是数据虽然存在却无法获得。例如,用户不想泄露私人信息,服务提供者也不想暴露其提供服务功能技巧的数据。因此,个性化服务存在着获得涉及到个人隐私的数据的风险,尤其是在一些敏感行业,例如银行金融、移动服务等。Kasai 等人进行私有数据保护功能的分布式个性化服务研究^[43]在不侵犯用户隐私的情况下,既能保护用户的隐私信息,又能保证获得的有效数据不失真。

6 结语

个性化是一种 Web 智能技术,在未来几年的研究中其关键技术依然是推荐系统。^[44]个性化推荐是基于 Web 挖掘理论和技术而不断发展。Web 挖掘是数据挖掘应用的一个分支,涉及到数据挖掘、人工智能、机器学习等多个领域的理论。并且随着 Web 技术和普适计算的发展,个性化推荐系统采用的挖掘技术越来越复杂,由单一技术向混合技术发展。例如,个性化推荐系统逐渐趋向于使用混合几种挖掘算法;由离线或在线状态学习趋向于二者的结合,这样便于个性化推荐的可扩展性。

用于个性化用户建模的数据收集也不再是单纯的显式或隐式的方法,而是采用更加智能的方法,既保护用户的隐私,又能够保证收集到数据的有效性。同时,个性化推荐进行数据挖掘需要的数据源越来越多样化,综合了日志数据、历史交易记录、用户搜藏、用户注册信息、个人评价反馈信息等。当前,关于 Web 日志格式的设计原则,

还缺乏相关的理论研究。因此，我们可以考虑将来进行 Web 日志格式设计的研究。

当前关于个性化评价的研究存在的问题比较多。因为个性化评价的因子选择与推荐系统的精度相关，因此，如何选择评价因子具有重要的影响作用。一般情况下，影响个性化推荐系统精度的因素有两个：一是个性化用户模型建模是否准确的体现了其偏好，这个因素影响推荐系统精度的原因在于收集的用户数据是否全面、准确；另一个因素是推荐系统算法，一些算法的先进性直接影响着推荐系统的精度。当前，一些研究者为了提高个性化推荐系统的精度，把研究重点放在提高推荐系统的算法优化方面。但是，有针对性的定制收集用户数据进行用户模型建模，对于个性化推荐系统精度的提高有着根本性的促进作用。因为，这些用户数据是根据用户操作定制收集，能够体现出其实际需求。所以，这些定制收集到的用户数据能准确的反映出用户的偏好，并且，用户模型建模收集到的数据越准确，就越有利于推荐系统的精度提高。同时，经过定制收集到的数据还能够有助于减轻推荐系统计算冗余数据的负载，从而提高个性化系统整体性能。

简言之，一个好的个性化应用是一个不断完善的过程。首先，个性化系统根据搜集到的用户建模信息，把经过挖掘的推荐结果呈现给用户；其次，用户对推荐的内容进行评价，系统收集到这些反馈信息后，调整个性化策略，再次进行用户建模；最后，把以更满足用户需求的挖掘结果再次呈现给用户。这个过程不断重复，即保证了系统的性能提高，又吸引了用户的使用。

参考文献：

- [1] Bamshad Mobasher. Data Mining for Web Personalization [C]. The Adaptive Web, Lecture Notes in Computer Science, 2007:90-135.
- [2] Resul Das, Ibrahim Turkoglu. Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method[J]. Expert Systems with Applications, 2009, (36):6635-6644.
- [3] F. Zhang and H. Chang. "Research and development in web usage mining system- key issues and proposed solutions: a survey" [C]. In First IEEE Int. Conf. on Machine Learning and Cybernetics Proceedings, 2002:986-990.
- [4] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan, Web usage mining: discovery and applications of usage patterns from web data[C]. SIGKDD Explore, 2000,1(12):12-23.
- [5] Z. Markov, and D. T. Larose. Data mining the Web : uncovering patterns in Web content, structure, and usage [M]. Hoboken, N.J.: Wiley-Interscience/John Wiley & Sons, 2007.
- [6] Xueping Peng, Yujuan Cao, Zhendong Niu. Mining Web Access Log for the personalization Recommendation[C]. International Conference on MultiMedia and Information Technology, 2008 :72-75.
- [7] ITWP2003[C/OL]. [2010-09-26]. <http://maya.cs.depaul.edu/~mobasher/itwp03/>.
- [8] Miki Nakagawa and Bamshad Mobasher. Impact of Site Characteristics on Recommendation Models Based On Association Rules and Sequential Patterns [C/OL]. [2010-09-26]. <http://maya.cs.depaul.edu/~mobasher/itwp03/papers/itwp03-06.pdf>.
- [9] Forsati R., Meybodi M. R., Neiat A.G.. Web Page personalization Based on Weighted Association Rules [C]. International Conference on Electronic Computer Technology, 2009 :130-135.
- [10] Billsus.Pazzani.A personal news agent that talks, learns and explains [C]. Proc. 3rd Int. Conf. on Autonomous Agents (Agents '99), Seattle, WA, 1999 268-275.
- [11] Michael J. Pazzani and Daniel Billsus. The Adaptive Web:Content-Based Recommendation Systems [C]. Lecture Notes in Computer Science,2007 325-341.
- [12] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa. Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data. Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01), 2001 [C/OL].[2010-09-26]. <http://maya.cs.depaul.edu/~mobasher/itwp01/>.
- [13] Aghabozorgi, S.R., Teh Yang Wah. Dynamic Modeling by Usage Data for personalization Systems[C]. 13th International Conference Information Visualisation, 2009 : 450-455.
- [14] (加) Jiawei Han, (加) Micheline Kamber. 范明, 孟小峰译. 数据挖掘概念与技术[M].北京 机械工业出版社.2006.
- [15] Burke.Hybrid Web recommender systems[C]. Brusilovsky, P., Kobsa, A., Nejdl, W., eds.: The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, 2007 377-408.
- [16] Ardissono, L., Gena, C., Torasso, P., Bellifemine, F.,

- Difino, A., Negro, B. User modeling and recommendation techniques for personalized Electronic Program Guides [C]. Personalized Digital Television. Targeting Programs to Individual Users. Kluwer Academic Publishers, 2004 :3-26.
- [17] Nima Taghipour, Ahmad Kardan. A hybrid web recommender system based on Q-learning [C]. Proceedings of the 2008 ACM symposium on Applied computing, 2008 :1164-1168.
- [18] Wei Chen, Li-jun Zhang, Chun Chen, Jia-jun Bu. A Hybrid Phonic Web News Recommender System for Pervasive Access [C]. International Conference on Communications and Mobile Computing, 2009 :122-126.
- [19] Bamshad Mobasher, Honghua Dai, Tao Luo, Yuqing Sun, Jiang Zhu. Integrating Web Usage and Content Mining for More Effective Personalization [C]. Lecture Notes in Computer Science, Electronic Commerce and Web Technologies, 2000 :165-176.
- [20] ITWP2009 [C/OL]. [2010-09-27]. <http://www.dcs.warwick.ac.uk/~ssanand/itwp09/>.
- [21] Talia Lavie, Michal Sela, Ili Oppenheim, Ohad Inbar, Joachim Meyer. User attitudes towards news content personalization [J]. International Journal of Human-Computer Studies, 2010 :483-495.
- [22] WWW Committee Web Usage Characterization Activity [EB/OL]. [2010-09-27]. <http://www.w3.org/>.
- [23] M. Baglioni¹, U. Ferrara², A. Romei¹, S. Ruggieri¹, and F. Turini¹. Preprocessing and Mining Web Log Data for Web Personalization [C]. Advances in Artificial Intelligence, Lecture Notes in Computer Science Pages, 2003 :237-249.
- [24] Neal Lathia, Xavier Amatriain, Josep M. Pujol. Collaborative Filtering With Adaptive Information Sources [C]. The 7th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems (ITWP2009), 2009 :63-74.
- [25] ITWP2008 [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~Emobasher/itwp08/program.html>.
- [26] Marco de Gemmis, Giovanni Semeraro, Pasquale Lops, Pierpaolo Basile. A Retrieval Model for Personalized Searching relying on Content-based User Profiles [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~Emobasher/itwp08/program.html>.
- [27] Chia-Chuan Hung, Yi-Ching Huang, Jane Yung-jen Hsu. Tagging-Based User Profiling for Social Media Recommendation [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~Emobasher/itwp08/program.html>.
- [28] Ozgur Kirmemis, Aysenur Birturk. A Content-Based User Model Generation and Optimization Approach for Movie Recommendation [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~Emobasher/itwp08/program.html>.
- [29] Francois Fous, Marco Saerens. Evaluating performance of recommender systems: An experimental Comparison [C]. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008 :735-738.
- [30] Lorraine Mc Ginty, Barry Smyth. Evaluating Preference-Based Feedback in Recommender Systems [C]. Lecture Notes in Computer Science, Artificial Intelligence and Cognitive Science, 2002 :209-214.
- [31] Kwiseok Kwon, Jinhyung Cho, Yongtae Park. How to best characterize the personalization construct for e-services [J]. Expert Systems with Applications, 2010, 37(3): 2232-2240.
- [32] ITWP2005 [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~mobasher/itwp05/>.
- [33] Robin Burke, Bamshad Mobasher, Runa Bhaumik. Limited Knowledge Shilling Attacks in Collaborative Filtering Systems [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~mobasher/itwp05/>.
- [34] J. J. Sandvig, Bamshad Mobasher, Robin Burke. Robustness of Collaborative Recommendation Based On Association Rule Mining [C]. Proceedings of the 2007 ACM conference on Recommender systems, 2007 :105-111.
- [35] Bamshad Mobasher, Robin Burke, Runa Bhaumik, Chad Williams. Toward Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness [J]. ACM Transactions on Internet Technology, 2007, 7(4):231-269.
- [36] ITWP2006 [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~mobasher/itwp06/>.
- [37] S. Toivonen, O. Riva. The Necessary But Not Sufficient Role of Ontologies in Applications Requiring Personalized Content Filtering [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~mobasher/itwp06/>.
- [38] P. Symeonidis, A. Nanopoulos, A. (下转第 98 页)

有了自己的教育权利,具有一定识字阅读水平的女性能进入公共图书馆,享受到一定的读者权利,无疑也为她们个人的良好发展铺平道路。部分知识女性在时代的催生中进入知识文化圈,得以享受到知识文明的权利,这对女性的进步,文明的进步无疑是一大幸事。

但反思一下,女性天生就与男性共同生活在社会之中,把这样一个自然而然的文明现象视为不寻常,在图书馆这样面向社会开放的公共场所,专设女子阅览室,把她们与男性读者分开,难免有视女子为“另类”之嫌。虽然是照顾方便女子入馆借阅,但也不难看出男性把持话语权,制定图书馆制度来区分男性和女性借阅区,甚至可以说视女性为弱者的意图,何况当时女子受教育的还在极少数,能和男子一样求学受教,利用好图书馆资源、享受到较多读者权利的女子更在少数。真正的性别平等不是哪一方压倒另一方,由一方来规定另一方,而是双方平等、和谐地共同进步。但不管怎么说,部分近代知识女性能享受到知识权利,享有十分有限却从古未有的读者权利,也确是历史的进步。

参考文献:

- [1]韩松涛.人权入宪与信息自由[J].图书馆,2004,(6):15-16.
- [2]耿翠花.图书馆为弱势群体服务中的人文精神[J].图书馆,2004,(5):41-42.
- [3]陈方劲.谈我国图书馆立法与读者权利[J].山东图书馆季刊,2007,(2):13-14.
- [4]刘向.列女传·母仪·邹孟轲母[M].沈阳:辽宁教育出版社,1998:52.

- [5]罗婷.女性主义文学批评在西方与中国[M].北京:中国社会科学出版社,2004:11.
- [6]胡文华.略论中国早期妇女报刊[J].安徽师范大学学报,2003,(5):340.
- [7]王惠君.图书馆文化论[M].长沙:湖南大学出版社,2004:12.
- [8][9]黄建国.中国古代藏书楼研究[M].北京:中华书局,1999:23,364.
- [10]刘英.妇女解放与中国图书馆的变迁[J].内蒙古师范大学学报哲社版,2002,(8):107-108.
- [11]郑观应.盛世危言·学校(上)[A].夏远东.郑观应集(上册)[M].上海:上海人民出版社,1982:247.
- [12][13][15][16]程焕文.晚清图书馆学术思想史[M].北京:北京图书馆出版社,2004:154,195,222,238.
- [14][21][22][23][24]李希泌,张椒华.中国古代藏书与近代图书馆史料(春秋至五四前后)[M].北京:中华书局,1996:98,156,221,321,333.
- [17]李桂林.中国近代教育史资料汇编(普通教育)[G].上海:上海教育出版社,1995:956.
- [18]贾春.浅析梁启超兴办女学的思想主张[J].中华女子学院学报,1997,(4):30-31.
- [19]梁启超.变法通议[M].北京:华夏出版社,2002:92.
- [20]梁启超.变法通议·论女学(饮冰室合集11册)[M].北京:中华书局,1941.

作者简介:谢咏芬(1977-),女,湖南商学院图书馆馆员;唐淑香(1968-),女,湖南商学院图书馆副研究员。

- (上接第65页) Papadopoulos, Y. Manolopoulos. Scalable Collaborative Filtering based on Latent Semantic Indexing [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/~mobasher/itwp06/>.
- [39] Chia-Chuan Hung, Yi-Ching Huang, Jane Yung-jen Hsu. Tagging-Based User Profiling for Social Media Recommendation [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/%7Emobasher/itwp08/program.html>.
- [40] J.J. Sandvig, Runa Bhaumik, Maryam Ramezani, Robin Burke, Bamshad Mobasher. A Framework for the Analysis of Attacks Against Social Tagging Systems [C/OL]. [2010-09-27]. <http://maya.cs.depaul.edu/%7Emobasher/itwp08/program.html>.
- [41] Sadeh, N. Mobile commerce: new technologies, services

- and business models [M]. Wiley & Sons, Ltd. 2002.
- [42] Herder, E., van Dijk, B.: Personalized adaptation to device characteristics [C]. Berlin Heidelberg New York, Springer-Verlag, 2002, 598-602.
- [43] Hiroyuki Kasai, Wataru Uchida, Shoji Kurakake. A service provisioning system for distributed personalization with private data protection [J]. The Journal of Systems and Software, 2007, 80: 2025-2038.
- [44] ITWP2010 [C]. [2010-09-27]. <http://ls13-www.cs.uni-dortmund.de/homepage/ITWP2010/index.shtml>.

作者简介:苏玉召(1975-),男,中科院国家科学图书馆博士研究生,研究方向:计算机网络、网络信息管理技术与信息系统、数据挖掘;赵妍(1979-),女,郑州航空工业管理学院助教。