

●胡昌平 孙高岭

# 个性化网络协同推荐服务的扩展及其实现\*

**摘要** 传统的网络协同推荐服务对用户行为数据进行统一挖掘处理,但在处理分散数据时,难以开展协同服务。由此,推荐服务系统的可扩展问题日益凸显。实际上,用户信息行为与需求层次结构的对应关系已为构建可扩展的个性化协同推荐服务系统提供了一种新思路。可扩展的个性化协同推荐服务模型从需求层出发,通过行为层、过滤层、推荐层、交互层和资源库的扩展,最终实现个性化协同推荐系统的可扩展性。图6。参考文献11。

**关键词** 网络协同推荐服务 信息服务 用户需求 扩展性

**分类号** G250.72

**ABSTRACT** The authors analyze the disadvantages of traditional online cooperative recommendation services, and point out the extensibility issue. Actually, the mapping between the hierarchical structures of user behaviors and user needs has provided a new idea for the personalized online cooperative recommendation services with high extensibility. The authors also proposed a design of the model for the realization of the extensible personalized online cooperative recommendation services. 6 figs. 11 refs.

**KEY WORDS** Online cooperative recommendation services. Information services. Extensibility. User needs.

**CLASS NUMBER** G250.72

在海量信息的互联网世界中,如何迅速而准确地发现并满足用户的信息需求是信息服务面临的一个挑战,而个性化协同推荐服务在一定程度上解决了该问题。然而,随着系统用户数量和服务项目的增多,传统的推荐服务难以满足扩展要求。因此,必须提高其协同过滤的可扩展性,以实现全面满足用户个性化需求的、可扩展的个性化网络协同推荐服务。

## 1 基于个性化需求信息过滤的推荐服务演进

用户的信息行为受用户的主体工作和外在的信息所激励,是一种与需求直接相联系的信息目标活动。处在一定环境下的用户,在社会、自然和个体因素作用下必然产生某种信息需求,信息需求的内容和形式不断引起用户主体的信息心理活动,继而产生为实现某一目标的认知行为,这就是体现个性化需求的用户个体特有的信息行为<sup>[1]</sup>。网络协同推荐服务应针对体现不同状态需求的信息行为进行分层过滤,以在每个层次上找到用户的最近邻集合。对用户行为数据的分层处理,必然使计算量大幅下降,因而可以保证协同推荐服务的可扩展性。

### 1.1 基于客观状态需求的过滤

一般情况下,用户不愿意为自己没有需求的信息付出时间或货币成本,自然也不会引发对这些信息或服务利用的行为<sup>[2]</sup>。因此,客观状态信息需求最有可能发生的行为路径是:客观状态需求——人机交互——信息利用(消费)行为(见图1)。如果一次没有完全满足用户的客观需求,用户可以将这个过程重复进行多次,直至自身的客观需求得到完全满足为止。个性化协同推荐系统将用户的信息利用和消费行为转化为用户对所利用(消费)信息产品或服务的评价,因此在客观需求层面找到用户的相似用户群,然后针对用户群进行信息利用推荐是可行的。

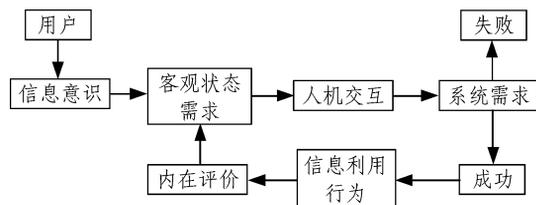


图1 用户客观状态需求行为路径

### 1.2 基于认识状态需求的过滤

在用户现有的客观需求得到满足的同时,用户通

\* 本文系教育部哲学社会科学重大课题攻关项目“创新型国家的知识信息服务体系研究”(项目编号:06JZD0032)的成果之一。

过内在评价或外在引导可能会意识到自身的某些潜在需求。处于认识状态的潜在信息需求在转变为实际需求的过程中最有可能发生的行为路径是:认识状态需求——信息交流行为——作用于其他用户(见图2)。在这条路径上,用户在满足客观需求的基础上,进行内在评价,继而激发自身的潜在需求。此时用户不一定能确认其潜在需求的真实性和必要性,也无法给出准确的需求描述,因此也不会向信息服务系统发出信息利用的请求,但可以通过系统交流平台与其他用户进行交流,以对自己处于认识状态的需求进行进一步的确认。同时,具有相似需求的用户会通过信息传递行为,对其他用户的信息交流行为进行响应,实现认识状态信息需求的发现和确认<sup>[3]</sup>。个性化协同推荐系统将用户的信息传递和交流行为数据转化为用户对相关信息产品或服务的期望数据。据此,可以在认识需求层面找到相似用户群,然后在用户进行信息交流的过程中给出系统的相关推荐信息。

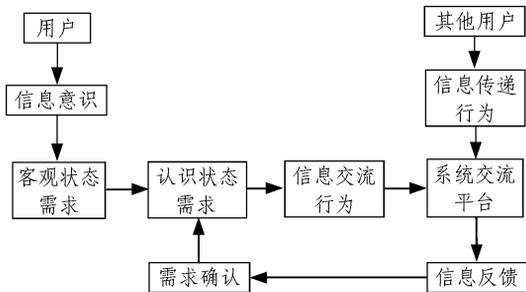


图2 用户认识状态需求行为路径

### 1.3 基于表达状态需求的过滤

在各种需求状态对应的信息行为中,用户虽然都与系统发生了交互,但需求信息的表达常是隐式的。只有当用户的需求提升到表达状态时,用户才会主动通过信息行为显性地向服务系统表达自己的个性化需求请求。威尔逊(Wilson)在信息行为模型分析中将信息查询行为细化为4种:被动参与行为、被动查询行为、主动查询行为和持续查询行为<sup>[4]</sup>。其中的主动查询行为和持续查询行为,是用户表达状态需求对应的行为(见图3)。用户除了被动地获取信息外,主动表达信息需求的行为十分强烈。用户主动表达信息需求,表现为网络社会中的需求信息发布行为,个性化协同推荐系统将根据此进行用户信息查询行为或发布行为的数据转化,同时利用用户对指定信息产品或服务的评价数据,在认识需求层面上找到用户群,在用户信息查询过程中给出用户可能感兴趣的信

息产品或服务推荐。

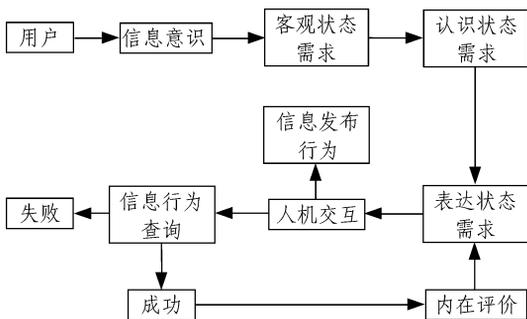


图3 用户表达状态需求行为路径

## 2 可扩展的个性化协同推荐服务模型构建

传统的网络协同推荐服务对用户行为数据进行统一挖掘处理,虽然具有针对性,但是处理数据分散,难以开展协同服务。面对这一现实,推荐服务系统的可扩展性问题日益凸显。实际上,用户信息行为与需求层次结构的对应关系已为构建高可扩展性的个性化协同推荐服务系统提供了一种新的思路。首先,将可以表达用户个性化需求的用户行为信息进行全面、准确的收集,然后依据行为信息所体现的信息需求层次的不同对用户行为数据进行分离,以在各自的层次上找到相应的最近邻集合,最后借助资源库和方法库为用户提供不同层次的个性化推荐服务。由此,我们提出了可扩展的网络协同推荐服务模型(如图4所示)。

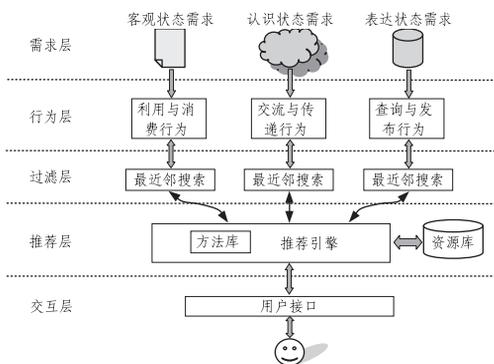


图4 可扩展的个性化协同推荐服务模型

可扩展的个性化协同推荐服务模型从需求层出发,通过行为层、过滤层、推荐层、交互层和资源库的扩展,最终实现个性化协同推荐系统的可扩展性。

行为层关注用户的需求,是个性化协同推荐服务实现的基础。行为层通过将用户行为数据按需求所

体现的层次分解,实现协同推荐服务范畴的扩展。依据用户不同需求状态的行为路径分析,用户客观状态需求可以映射为信息利用(消费)行为,认识状态需求可以映射为信息交流(传递)行为,表达状态需求可以映射为信息查询(发布)行为,推荐服务模型对这些行为信息的全面收集将使协同推荐服务的服务范畴全面覆盖客观需求、认识需求和表达需求三种需求状态。

过滤层关注推荐的效率,是个性化协同推荐服务性能的瓶颈<sup>[5]</sup>。经历了基于用户的协同过滤、基于模型的协同过滤和基于项目的协同过滤的算法改进<sup>[6-7]</sup>,协同推荐服务的性能可扩展性有了很大提高<sup>[8]</sup>,但推荐的准确性有所下降<sup>[9]</sup>。提高推荐服务性能可扩展性的关键,在于降低推荐算法处理的数据量和计算的复杂度,因此,将行为数据按所体现的需求层次进行分离,选用合适的算法在不同的需求层次上分别找到相应的最近邻集合,可以进一步提高推荐服务性能的可扩展性。

推荐层关注推荐的效果,是个性化协同推荐服务实现的核心。推荐层通过引入方法库来实现推荐服务功能的可扩展性。方法库可以吸纳最新的推荐方法和剔除效率低下的推荐方法,是开放的动态的推荐方法集合。通过对信息行为的分层过滤,我们可以得到不同需求状态下的相似用户群,推荐引擎根据用户行为数据的特点选择合适的推荐方法,对目标用户的兴趣进行预测,生成个性化的推荐结果。根据推荐目的的不同,也可以将多种推荐方法进行综合,开展多种形式的推荐,实现各种不同的推荐服务功能,满足用户不同的个性化信息需求。

交互层关注用户本身,是个性化协同推荐服务实现的关键。交互层通过良好的用户接口设计实现推荐服务对象的可扩展性。个性化协同推荐服务起于用户止于用户,人机交互系统是紧密联系用户和系统的桥梁。流程顺畅、功能完善的交互系统将吸引并留住大量的用户,使协同推荐服务的成功具有保障<sup>[10]</sup>。推荐服务的交互层不只是人机交互,也包括系统与系统的交互。在用户数量众多且地理位置分散的情况下,可以利用代理模式实现服务对象更大程度的扩展。推荐服务系统将推荐结果传递给一级代理,一级代理再推荐给二级代理,依次传递直到推荐给最终的用户。

此外,个性化协同推荐服务通过统一的资源库建设实现推荐服务内容的可扩展性。现有的网络信息

资源大多是异构的且分布在网络不同节点上,查询利用都很不方便。要实现推荐服务内容的可扩展性,必须对现有的分布式异构信息资源进行整合,建设一个格式统一、内容丰富且质量优秀的信息资源库,通过推荐服务系统进行集成信息推荐服务。

### 3 个性化网络协同推荐服务的扩展实现

在个性化协同推荐的扩展实现中,我们以湖北省农业信息集成服务平台的个性化服务为实证,进行了协同推荐的扩展研究。农业信息集成服务平台是一个面向全省的县市级农业信息服务网站的综合型信息服务平台,用户和系统服务项目数量都很大,应用传统的协同推荐方法系统必然存在严重的可扩展性问题。我们根据可扩展的个性化协同推荐服务模型,由人机交互子系统、数据收集模块、数据过滤子系统和协同推荐模块架构了农业信息集成服务平台的个性化协同推荐系统(如图5所示)。

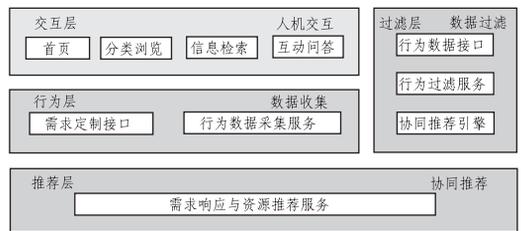


图5 平台个性化协同推荐服务功能架构

农业信息集成服务平台个性化协同推荐系统的模块根据功能的不同将分布在系统的不同层面,实现推荐服务不同方面的可扩展性(如图6所示)。

(1)推荐服务对象的扩展。推荐系统的交互层采取一级代理的形式实现,信息资源集成服务平台将资源推荐给各县市级的信息服务代理,信息服务代理再通过集成在自己的信息门户中的人机交互子系统为具体的用户开展个性化的信息服务。推荐服务系统利用代理模式成功地解决了用户数量众多且地理位置分散的问题,使服务对象从传统的单一县级信息服务网站的用户群扩展到全省的所有农业信息服务用户。推荐服务从直接推荐变为间接推荐,通过信息服务代理将推荐服务对象线性扩大。

(2)推荐服务范畴的扩展。推荐系统的行为层通过分布式信息服务代理上的行为数据收集模块实现。每个信息服务代理节点上都有行为数据采集服务负责收集该节点上用户所发出的各种信息行为,并

对行为数据按表达的需求状态不同进行分离,确保推荐服务的范畴会全面覆盖客观状态需求、认识状态需求和表达状态需求。依据需求与行为的对应关系将行为数据进行分离,使推荐服务由针对某种单一需求状态进行推荐扩展到面向所有需求状态进行全面推荐,推荐服务的范畴得到了扩展。

(3) 推荐服务性能的扩展。推荐系统的过滤层是系统性能可扩展性的瓶颈,因此数据过滤系统将作为平台的独立子系统单独实现。协同推荐服务子系统通过行为数据接口接收由信息资源集成服务平台传递过来的行为数据,然后由行为过滤服务模块在不同需求层面上搜索到用户的最近邻集合,传递给协同推荐引擎。最近邻搜索的核心是计算两个用户的相似度<sup>[11]</sup>。在每个需求层面上,最近邻搜索需要处理的数据量都有所降低,再根据数据集的不同特点选择合适的算法,最近邻搜索的复杂度将大幅下降,因此推荐系统的性能会得到充分的扩展。

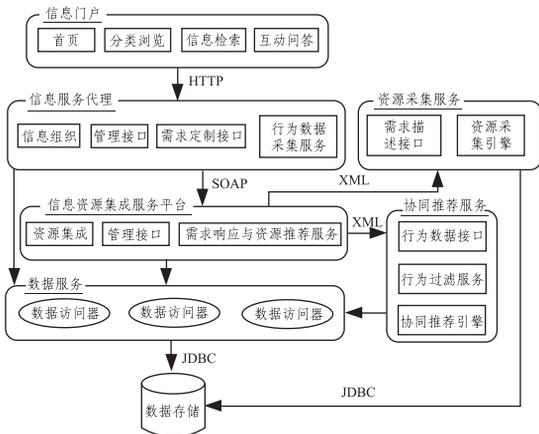


图6 湖北省农业信息集成服务平台系统架构

(4) 推荐服务功能的扩展。推荐系统的推荐层通过协同推荐模块在信息资源集成服务平台中与需求响应模块一并实现,以达到系统信息传递效率最优。根据推荐目的的不同,个性化协同推荐服务可进行多种形式的推荐。信息资源集成服务平台的推荐结果通常有两种:Top-N推荐和关联推荐。Top-N推荐是针对单个信息服务产生的,它对每个站点是不一样的。系统通过对该站点用户的最近邻用户进行统计,选择出现频率最高且在信息服务平台的资源定制需求中不存在的项目作为推荐结果。关联推荐方法仅对最近邻用户的需求定制记录进行关联规则挖掘。如果用户曾经定制过关联规则左边的资源项目,

而没有定制过关联规则右边的资源项目,那么就把右边的这个资源推荐给该站点。随着推荐系统方法库的不断完善,推荐服务功能的扩展性也不断的增强。

(5) 推荐服务内容的扩展。推荐服务内容的扩展主要通过平台统一的资源库建设实现。内容扩展包括两方面:内容深度的扩展和内容广度的扩展。内容深度的扩展体现在资源库将为推荐服务提供深度支持,使推荐服务的内容从传统的目录、线索、文摘服务深入到信息内容的全文服务,并可以对信息内容进行深层次的知识挖掘,以便开展知识信息服务。内容广度的扩展体现在平台资源库将本地资源与网络资源进行整合。在进行自身资源建设的同时,平台利用资源采集引擎对互联网上的各类分布式信息源进行内容抓取,再通过统一的格式转化使巨大而分散的网络信息资源成为自身资源库的一部分。通过资源库在深度和广度两方面对推荐服务的支持,推荐服务内容必然得到扩展。

### 参考文献:

- [1] 胡昌平,晏浩. 知识管理活动创新性研究之协同知识管理[J]. 中国图书馆学报, 2007(3):95-98.
- [2] Bruce, H., Jones, W. & Dumais, S. Information behaviour that keeps found things found[J]. Information Research, 2004, 10(1).
- [3] Niedzwiedzka, B. A proposed general model of information behaviour[J]. Information Research, 2003, 9(1).
- [4] Wilson, T. D. Human Information Behavior[J]. Informing science, 2000, 3(2):49-55.
- [5] Goldberg D, Nichols D, Oki B M, Terry D. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
- [6] Greg Linden, Brent Smith, Jeremy York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering[J]. IEEE Internet Computing, Jan/Feb, 2003, 7(1): 76-80.
- [7] G. D. Linden, J. A. Jacobi, and E. A. Benson. Collaborative Recommendations Using Item-to-Item Similarity Mappings[P], US Patent 6, 266, 649 (to Amazon.com), Patent and Trademark Office, Washington, D. C., 2001.
- [8] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-based collaborative filtering recommender algorithms[C]. 10th Int'l World Wide Web Conference, ACM Press, 2001, 285-295.

(下转第45页)

究越来越多,主要涉及知识经济背景下的知识发现、知识管理和知识管理系统。也可能正是由于知识管理作为信息资源管理的一个新的研究方向出现,才导致 1998~2006 年信息资源管理论文数以线性方式增长而非指数方式增长,这一观点与 S. H. Liao 的观点相同<sup>[6]</sup>。图书馆作为信息资源管理的一个核心机构,逐渐由高校图书馆、公共图书馆等实体图书馆的研究向虚拟图书馆、数字图书馆的研究方向倾斜。无论是信息资源管理系统还是知识管理系统,国内外学者对“系统”的研究都比较多,从技术实现角度探讨信息资源的管理。电子商务、电子政务是近几年新兴起的研究热点,另外,服务、用户、政策法规也是国内外学者共同关注的方向。对专业人才的教育、培养是国内的第六大研究热点,但却不是外文文献中研究的重点,这与我国正在进行信息化建设,十分重视人才培养与教育有关。“信息检索”在外文关键词中排第六名,在国内排名则相对靠后,这是前十大中英文关键词中的唯一不同之处。此外,信息共享、知识产权等问题也是国内外学者研究的重点问题。

### 3 结论

目前,国内外信息资源管理领域的科技文献均以线性方式增长,该领域发展已相对成熟,知识管理成为信息资源管理新的研究方向。国内文献的适用时间相对较短,文献利用速度相对较慢,有较大的改善提升空间。国内刊载信息资源管理领域研究论文的期刊比较集中,该领域的研究人员主要由图书情报档案领域的学者构成,国外则相对分散。国内作者合作智能发挥还不够充分,有待进一步提高。国内外研究

热点问题比较一致,网络、知识、技术、用户等问题是共同的研究重点。总之,国内外信息资源管理研究情况存在一定差异,但都取得了很大的发展。

### 参考文献:

- [1] 严贝妮. 1998 - 2002 年我国信息资源管理领域研究论文计量分析[J]. 图书情报工作, 2004(5): 109 - 112.
- [2] 周群. 1994 - 2003 年信息资源管理研究进展及发展趋势[J]. 图书情报工作, 2004(11): 115 - 118.
- [3] 邱均平. 信息计量学[M]. 湖北: 武汉大学出版社, 2007.
- [4] 孙峰. 科技论文参考文献的作用及引用中存在的问题[J]. 燕山大学学报(哲学社会科学版), 2005(3): 92 - 94.
- [5] 党亚茹, 王莉亚. JCR 自然科学版期刊半衰期指标的区间变化分析[J]. 情报科学, 2007(6): 804 - 810.
- [6] S. H. Liao. Knowledge management technologies and applications-literature review from 1995 - 2002 [J]. Expert System with Application, 2003(25): 155 - 164.

**邱均平** 武汉大学博士生导师, 武汉大学中国科学评价研究中心主任。通讯地址: 武汉大学。邮编 430072。

**苏金燕** 武汉大学信息管理学院博士研究生。通讯地址同上。

**熊尊妍** 武汉大学信息管理学院硕士研究生。通讯地址同上。

(收稿日期: 2008-01-11)

(上接第 64 页)

- [9] Prem Melville, Raymond J. Mooney, Ramadass Nagarajan. Content-Boosted Collaborative Filtering for Improved Recommendations[C]. Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI - 2002), pp. 187 - 192, Edmonton, Canada, July 2002.
- [10] 胡昌平, 邓胜利. 基于用户体验的信息资源整合分析[J]. 情报学报 2006(2): 231 - 235.
- [11] Mobasher B, Cooley R, Srivastava J. Automatic Person-

alization Based on Web Usage Mining[J]. Communications of the ACM, 2000, 43(8): 142 - 151.

**胡昌平** 武汉大学信息资源研究中心教授, 博士生导师。通讯地址: 武汉大学信息管理学院。邮编 430072。

**孙高岭** 武汉大学信息管理学院信息资源管理专业硕士研究生。通讯地址同上。

(收稿日期: 2007-09-05)