

当前互联网几种信息推送技术及其评析*

张月天¹, 邬伟², 曾欣³

(1. 四川大学图书馆, 四川成都 610065; 2. 绵阳市图书馆, 四川绵阳 621000;
3. 美国密西根州立大学 概率统计系, 美国 兰辛)

摘要: 随着计算机技术和网络技术的飞速发展, 信息推送作为一种新兴的信息传播模式, 正在被越来越多的信息服务者所重视。信息推送技术的出现从根本上改变了人们获取信息的传统方式, 极大地提高了用户获取信息的效率。文章对信息推送进行概述, 阐述 Agent、RSS、网格、协同过滤等几种信息推送技术的内涵、特点、基本结构和流程, 并对各自的优势和不足进行分析和评价, 最后提出未来的研究方向和展望。

关键词: 信息推送; 推送技术; RSS; 协同过滤; Agent; 网格

中图分类号: G 2 **文献标识码:** A **文章编号:** 1004 - 390X (2009) 02 - 0116 - 05

Analysis of the Information Push Technology

ZHANG Yue-tian¹, WU Wei², ZENG Xin³

(1. Library of Sichuan University, Chengdu 610065, China; 2. Library of Mianyang, Mianyang 621000, China;
3. Department of Statistics and Probability of Michigan State University, Lansing, USA)

Abstract: With the development of internet, the Information Push Technology, as a revolutionary information transmitting mode, is increasingly being paid more and more attention by the information services. The appearance of the "Information Push Technology" has tremendously changed the traditional accesses to information, resulting in a revolutionary high efficiency in obtaining information. This thesis mainly introduces several kinds of Information Push Technology, such as Agent, RSS, Grid and Collaboration Fiterin, at the same time, analyses and evaluates their merits and demerits, which lays out the direction of future research.

Key words: information push; push technology; RSS; collaboration fiterin; agent; grid

互联网是人们进行信息共享的一个最有效的工具。但随着网络信息资源容量急剧膨胀, 让用户开始困惑的已经不是有没有所需要的信息, 而是如何能够获取自己真正所需要的信息。信息推送技术的出现从根本上改变了人们获取信息的传统方式, 极大地提高了用户获取信息的效率, 对信息的采集、加工、传递和利用产生了重大影响。

信息推送技术最先由美国 PointCast Network 公司于 1996 年提出, 目的是为了提提高基于计算机网络的信息获取效率。随着 RSS、Agent、协同过滤的成功推出, 信息推送技术被迅速应用到互联网、图书馆、数据库、电子商务、广播电视系统

及通讯系统等领域, 尤其是近年来国际上兴起的网格 Grid 技术, 它将改变传统的 Client/ Server 和 Client/ Cluster 结构, 形成新的 Pervasive/ Grid 体系结构, 这种体系结构将使得用户把整个网络视为一个巨大的计算机, 并从中享受一体化的、智能的、协作式信息服务。^[1]

一、信息推送概述

(一) 信息推送的定义

信息推送 (Information push) 是通过一定的技术标准或协议, 在 Internet 上通过定期传送用户需要的信息来减少信息过载的一项新技术。准

收稿日期: 2008 - 12 - 05 修回日期: 2009 - 03 - 20

* 基金项目: 四川省科技厅软科学项目 (2006R16 - 047)。

作者简介: 张月天 (1956 -), 女, 四川江津人, 主要从事图书馆学、信息技术研究。

确地说, 它属于第三代浏览器的核心技术, 其关键是能够主动地根据用户的需求, 将最新的信息分门别类地传送到相应的用户设备中, 有效地改变人们获取信息的方式, 因而较大幅度地提高了互联网信息使用效率。

(二) 信息推送的实现方式

目前, 信息推送的实现方式主要有消息、代理、频道三种方式。其中消息方式是根据用户提交的信息需求, 利用电子邮件或其它消息系统将有关信息发送给用户。代理方式是通过使用代理服务定期或根据用户指定的时间间隔在网上搜索用户感兴趣的信息内容, 将结果推送给用户。频道方式则需提供完整的 Push 服务器、客户端部件及相关开发工具等一整套集成应用环境, 将某些站点定义为浏览器中的频道, Push 服务器负责收集信息形成频道内容后推送给用户。

二、信息推送技术介绍

(一) 基于网络的信息推送技术

1. 网络的概念

网络的概念来源于电力网。它是一个集成的计算与资源环境, 它的目标是基于网张技术, WEB 技术和高性能计算等技术, 采用开放标准, 实现网张虚拟环境的资源共享和协同工作, 消除信息孤岛。

2. 网络的体系结构

目前网络主要有两种体系结构: 一种是五层沙漏结构。这是 Foster 等在早些时候提出的模型。

五层沙漏从下到上依次为构造层、连接层、资源层、汇聚层和应用层, 层层为递进关系。第二种是开放网格服务结构, 它是以服务为中心的网格体系结构, 是继沙漏模型之后最重要, 也是目前较新的一种网格体系结构。^[1]图 1 为基于网络的网格信息推送模式。

3. 网格技术的优势^[1]

(1) 虚拟性: 网格与外界的交流是通过网格与用户的交互界面来实现的。用户所使用的资源分布于世界各地, 用户的行为被虚拟化为网格用户和网格资源, 实际物理用户是不可见的。

(2) 共享性: 这是网格的一个重要特性, 共享性指的是网格中的各种资源都能够被用户共同使用, 网格就是一个共享的场所。

(3) 协同性: 协同性包括资源共享的协同性和问题解决的协同性。资源协同主要是指各种资源被利用的无冲突, 而问题解决的协同性指的是虚拟组织在解决某一个共同的问题时可以协同解决。

4. 网格技术的不足

(1) 自治性与管理的多重性。网格允许资源拥有者对他的资源自主管理, 这就是网格的自治性。但是网格资源也必须接受网格的统一管理, 否则不同的资源就无法建立起相互之间的联系, 无法实现共享和互操作, 因此网格的管理具有多重性, 给网格管理带来混乱和麻烦。

(2) 标准化问题。网格需要依赖标准协议才能共享和互通。标准问题的根源不在于技术, 而

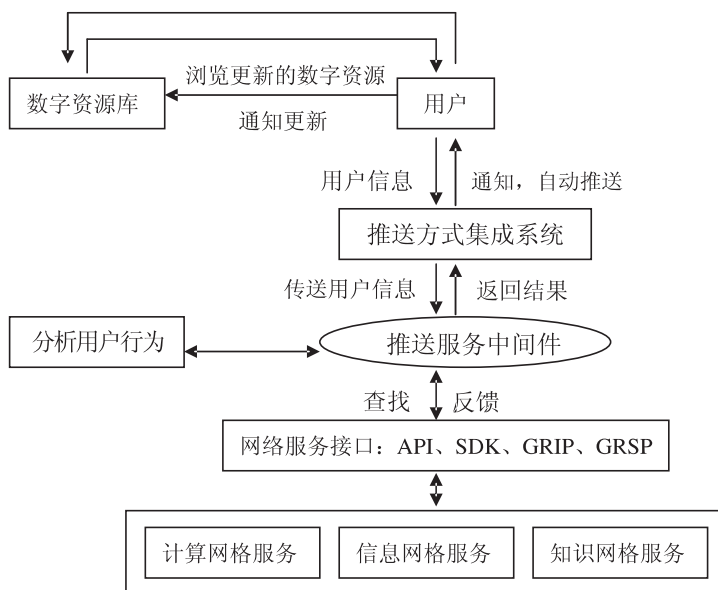


图 1 基于网络的网格信息推送模式

在于成本巨大和利益分割,所以虽然目前包括全球网络论坛 GGF 在内的诸多团体都在积极的努力,网络计算的开放标准却仍在发展之中。^[2]

(3) 安全性问题。网络安全服务并不能完全解决网络计算环境的安全问题,网络安全问题比网络安全问题更为复杂。从某种角度上说网络计算,甚至是一种无政府状态的计算模式,因而如何进行有序的控制和管理就显得十分重要。

(二) Agent 推送技术

1. Agent 概念

Agent 是人工智能领域发展起来的一个概念,最早出现在上个世纪七十年代。随着计算机网络以及基于网络的分布式计算技术的发展,有关 Agent

的研究逐步成为人工智能领域的一个新的研究热点。如今,Agent 技术已经逐渐渗透到计算机领域的方方面面,为解决各类问题提供了一个崭新的角度,被广泛应用于信息检索、信息过滤、网络管理、电子商务、远程教育、工业控制等诸多领域。

2. Agent 工作流程

Agent 工作流程如图 2 所示。当用户使用系统时,会产生对应于这个用户的五个 Agent。用户 Agent 提供用户与系统的接口;过滤 Agent 对信息进行过滤筛选,产生推荐信息;检索 Agent 进行信息检索;学习 Agent 根据信息反馈和用户浏览行为调整兴趣模型;监测 Agent 监视用户兴趣和相似用户兴趣的改变。^[3]

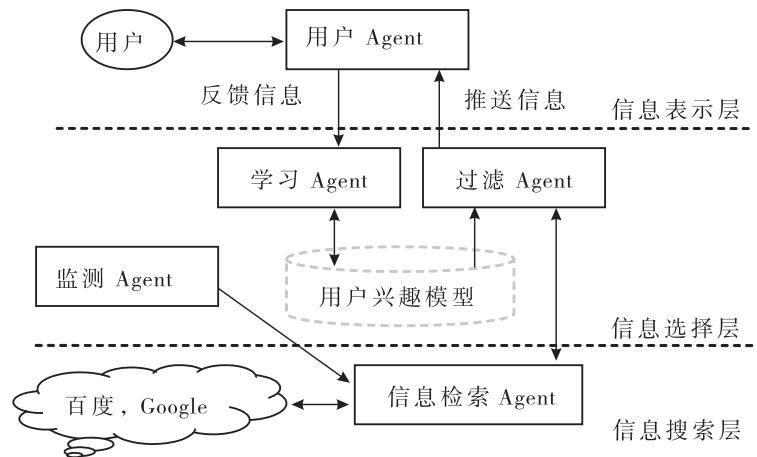


图 2 基于 Agent 的信息推送系统结构

3. Agent 技术的优势^[3]

(1) 自治性 (Autonomy): 一个 Agent 应该具有独立的局部于自身的知识和知识处理方法,在自身的有限计算资源和行为控制机制下,能够在没有人类或其它 Agent 的直接干涉和指导的情况下持续运行,以特定的方式响应环境的要求和变化,并能根据其内部状态和感知到的环境信息自主决定和控制自身的状态和行为。

(2) 反应性 (Reactivity): Agent 能够感知外界环境,并对外界环境的变化适时作出反映,表现出受目标驱动的自发行为。

(3) 社会行为能力 (Social Ability): Agent 往往不是独立存在的,如同现实世界中的生物群体一样,在环境中经常有很多 Agent 同时存在,形成一个社会性的群体。Agent 不仅能够自主运行,同时应该具有和外部环境中其它 Agent 相互协作的能力,而且在遇到冲突时能够通过协调进行解决。

(4) 主动性 (Pro-Activeness): Agent 具有主动地启动某种行为,执行某个操作的特点。

4. Agent 技术不足

在分布式开发环境中,各个 Agent 往往是异构的、动态的和不可预测的,这些特征使得 Agent 之间的合作变得困难和复杂。不同的系统拥有各自的运行环境,访问控制、安全策略、命名规则和指令系统等,相互之间不能兼容。目前尚缺乏对 Agent 成员的行为、协作、通信等进行描述、设计、编程和调试的广泛可用且强大的 MAS 工具。另外 Agent 系统的安全性和可信赖性也是一个需要研究的问题。^[4]

(三) 协同过滤 (Collaborative Filtering) 推送技术

1. 协同过滤的概念

协同过滤推送是当前最成功的推送技术。它是一个基于存储推理的变形,能够发现内容上完全不相似的资源,用户对推荐的内容可能事先是

预料不到的,却是用户潜在的兴趣,它特别适合于提供个性化推荐方面的应用。

2. 协同过滤推送系统结构

图3为基于用户聚类的协同过滤推送技术典型结构图。分为两个部份:第一个是离线处理部份,

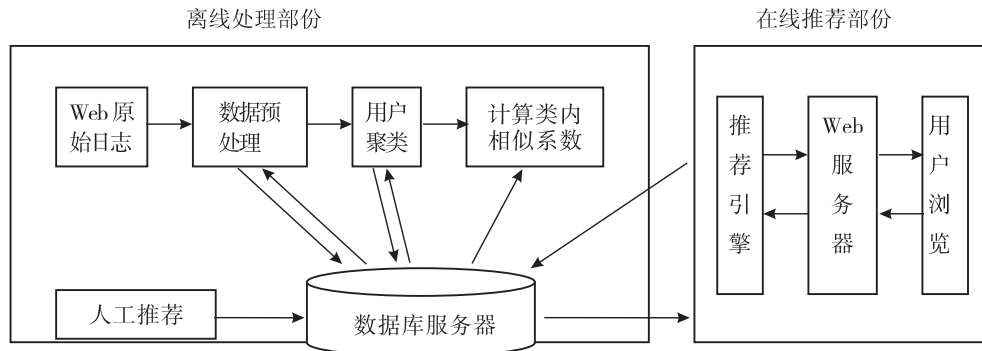


图3 协同过滤推送系统结构图

3. 协同过滤技术的优势

协同过滤主要分为两类:基于用户的过滤和基于项目的过滤。目前应用较多的是基于用户的过滤。它先使用统计技术,找出和当前用户兴趣相似的用户群,综合相似用户群中的用户对某一项信息的评价形成当前用户对该项信息的喜好,这种方法可以推荐各种类型的信息,例如图像、视频、音乐等等。

4. 协同过滤技术的不足

(1) 首先是数据的稀疏性问题。在许多采用这种技术的推送系统中,每个用户涉及的信息量相当有限,在大型电子商务系统中,用户评分的信息项数一般不会超过信息项总数的1%~2%,经两个用户共同评分的信息项就更少。如果两个用户没有对相同的项目进行打分,即使这两个用户的兴趣爱好都相同,系统也无法得出他们之间的相似度,从而难以找到相似用户集,导致推送效果大大降低。^[6]

(2) 其次就是扩展性问题。基于用户预测算法中,计算用户相似性时需要对所有用户进行相似性计算,如果评价矩阵规模很大,那么带来的计算复杂度将使协同过滤失去实际意义。

(3) 第三是冷开始。我们可以把它看成是数据稀疏问题的极端情况,当一个信息项没有被任何用户评价时,那么单纯的基于用户的协同过滤无法对其进行预测。

(4) 自动化程度问题。目前多数的协同过滤推荐算法所需要的项目评分都是需要采用显式的评分输入方式,也就是用户必须显式的输入对项

主要完成WEB日志的预处理以及对用户进行聚类和相似系数的计算;第二部份是在线推荐部分,利用离线阶段的处理结果,预测目标用户对未访问项的兴趣度,把兴趣高的项推荐给用户。^[5]

目的数值评分。这种方式虽然最能直接了解用户的兴趣点,却使得用户在使用上很不方便。

(四) RSS 推送技术

1. RSS 概念

RSS 是 Really Simple Syndication 的缩写,中文称作“简易信息聚合”,是基于XML技术的互联网内容发布和集成技术。号称“网络信息拯救者”的RSS有强大的信息聚合和推送功能,它能够用于共享各种各样的信息,包括新闻、简讯、Web站点更新、事件日历、软件更新、特色内容集合和基于Web进行拍卖的商品等。

2. 基于RSS信息推送技术模型

RSS种子以XML格式文件为载体,所以可供提取的数据来源有3种:XML文件;关系型数据库;纯XML数据库即Native-XML数据库。网站提供的信息内容可以划分为不同的频道,为每个频道制定单独的RSS文件链接,并将其发布到网站上提供给用户访问。^[7]其技术模型如图4所示。

3. RSS技术的优势

(1) 聚合性:在用户端,RSS阅读器软件的作用就是按照用户的需要,有选择性地将用户感兴趣的内容聚合到该软件的界面中,为用户提供多来源信息的“一站式”服务。^[8]

(2) 实时性:RSS搭建了信息迅速传播的一个技术平台,发布一个RSS Feed后,这个RSS Feed中包含的信息就能直接被其它站点调用。

(3) 即时性:RSS使用了规范的XML文本格

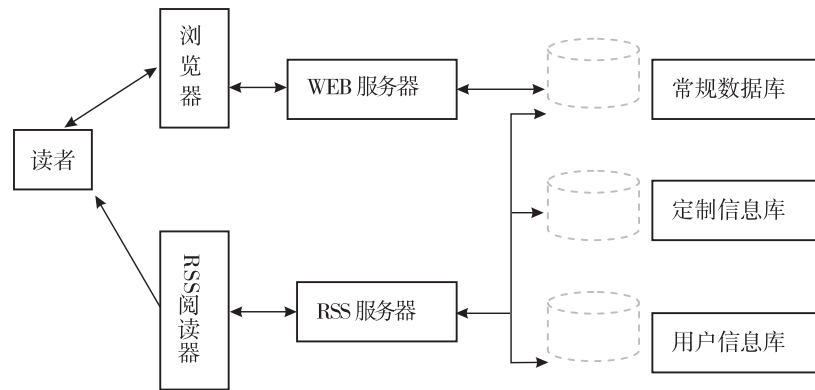


图4 RSS信息推送技术模型

式，信息的传递、接收处理都非常方便。RSS桌面聚合工具也提供了将RSS信息转化成HTML信息，直接在页面上点击媒体链接，就可以获得最新信息的HTML的页面显示。

(4) 成本低：服务器端内容的RSS包装在技术上实现极为简单，而且是一次性的工作，使用长期的信息发布边际成本几乎为零，是传统的电子邮件、卫星传输、网络浏览等发布方式所无法比拟的。^[9]

(5) 无垃圾信息：RSS用户端阅读软件的特点是完全由用户以频道的形式订阅值得信任的内容来源，同时RSS信息本身具有很好的分类特性。^[9]

4. RSS技术的不足

RSS技术已经形成了一套自己的理论和方法，包括用户模型和推荐技术，但是在个性化推送方式上有不足的地方。在信息推送方式上缺乏交互性，没有考虑到用户的个性化需求。

信息推送技术的出现从根本上改变了人们获取信息的传统方式，极大地提高了用户获取信息的效率。但因为这项技术应用时间不长，系统还不完善，还存在着一些不尽人意的地方。

比如在大型的信息推送系统中，信息推送的伸缩能力和实时性要求越来越难以保证，大部分推送技术都是在保证实时性要求的同时，以牺牲推送质量为前提的。而推荐系统的单一，也是影响信息推送效果的重要原因。当前大部分的推送系统还都只是一个单一的工具，只能提供一种推送模型，而由于系统本身的应用复杂性，难以应对不同类型的推送。

未来的研究还需要融入人工智能、知识发现技术及数据库技术，形成真正的“智能信息推

拉”(IIPP)技术。面对复杂的应用环境，还需改变单一系统的模式，研究新型的复合推送技术与框架，有效结合多种推送技术，收集多种类型的数据，满足不同类型的推送需求，从而提供更加有效的推送服务。^[10]

总之，无论如何，随着对网络信息推送技术研究的不断深入，我们相信网络信息推送技术将会逐渐完善和成熟。可以预期，未来的互联网将精彩纷呈，八仙过海名显神通，人们获取信息将获得极大的便利。

[参考文献]

- [1] 范迎秋. 基于网格的信息推送服务模式[J]. 情报探索, 2007, (3): 58-60.
- [2] 卜世波, 肖晓旦. 新技术、新思维—网格(Grid)研究综述[J]. 现代情报, 2005, (5): 82-86.
- [3] 余静, 段隆振, 熊必成. 基于Agent的信息推送系统的研究[J]. 计算机与现代化, 2007, (10): 18-21.
- [4] 肖佳. 基于Agent的个性化信息推送系统的研究[D]. 苏州: 苏州大学, 2006: 1-20.
- [5] 王辉, 高利军, 王听忠. 个性化服务中基于用户聚类的协同过滤推荐[J]. 计算机应用, 2007, 27(5): 1225-1228.
- [6] 顾勇镛. 一种混合型信息推送技术研究及其在企业情报服务平台中的应用[D]. 杭州: 浙江大学, 2008.
- [7] 于金海, 郭军成. 基于RSS的图书馆网络信息推送服务[J]. 科技情报开发与经济, 2007, 17(29): 38-39.
- [8] 王正元. 基于RSS的数据库资源信息推送服务[J]. 农业图书情报学刊, 2007, (7): 36-39.
- [9] 黄艳娟. 基于RSS的图书馆个性化信息推送服务[J]. 情报科学, 2006, 24(7): 123-126.
- [10] 刘迎清. 国内信息推送研究综述[J]. 长沙大学学报, 2006, 20(5): 82-86.