

Agent 社会:一种个性化智能信息发现环境^{*})

Agent Society: A Personalized Intelligent Information Retrieval Environment

许晓春 杜晓晨 梅卫锋 徐永森 陈俊良

(南京大学计算机软件新技术国家重点实验室 南京210093)(南京大学网络中心 南京210093)

Abstract Agent Society is a WWW-based information retrieval environment. It is composed of diverse information Agents. They can provide intelligent information discovery and retrieval services independently to individual users, which also have the ability to communicate and collaborate with each other.

Keywords Information retrieval, Intelligent Agent, Mobile Agent, Agent society

1. 引言

随着万维网的迅速发展,如何在纷繁复杂的信息海洋中寻找相关主题的信息变得越来越困难。以搜索引擎为代表的传统信息发现系统由于体系结构上的固有弱点已经跟不上信息膨胀的步伐^[1]。同时,现有的基础网络在速度、稳定性等方面还远不能满足用户的需求。

绝大多数的传统型搜索引擎都是基于客户/服务器模型的在线搜索模式^[2],并且在服务端后台维护着预定义、海量的信息数据库。这种结构对信息数据库和网络存在着过多的依赖性。首先,搜索引擎采用一套预定义的分类手段,并不考虑用户的特殊选择,因此用户得到的搜索结果经常会与期望结果大相径庭;第二,搜索引擎通常不具备学习功能,无法从广阔的信息空间动态地收集信息;第三,由于中心数据库过于庞大,很难对其进行及时的维护和更新;第四,搜索引擎不支持结果的异步传送,要求在提交请求和回送结果的过程中保持良好的网络连接。最后,传统的搜索引擎强调以引擎为中心,而忽视了不同背景和不同兴趣的用户的需求,因此迫切需要有一种个性化智能化的新型信息发现系统。

Agent 技术是当今计算机科学的一个研究热点,它的不断发展为我们提供了一个将信息发现个性化和智能化的契机^[3]。有鉴于此,本文提出了一个基于 Agent 技术的个性化智能化的信息发现环境——Agent 社会。Agent 社会是一个以用户为中心的信息发现环

境,它采用了先进的智能 Agent 和移动 Agent 技术。在 Agent 社会中,用户根据个人的兴趣爱好进行个性化配置,并通过多种 Agent 间的协作来实现信息发现。

本文首先介绍 Agent 社会的基本组成和运作机制,然后对 Agent 社会的个性化、智能化特点做具体的阐述,最后将 Agent 社会与传统的搜索引擎做了比较。

2. Agent 社会的组成及运行机制

2.1 基本运作机制

Agent 社会中主要有信息服务 Agent (ISA)、信息移动 Agent (IMA) 和个人智能 Agent (PIA) 三类 Agent。PIA 是用户的虚拟“搜索引擎”。它负责组织用户的请求,发送 IMA,并将搜索结果进行智能化的整理,最后提交给用户。IMA 具有网络智能性和通讯协作智能性,它携带着用户的个性化请求,自主地寻找 ISA,汇总搜索结果后返回到 PIA。ISA 一方面为 IMA 提供信息服务,另一方面它还可以和其他 ISA 协作,及时维护和更新资源信息。

Agent 社会的信息发现过程是对手工进行信息浏览和发现过程的一种模拟。我们可以将其分成两个阶段:学习阶段和服务阶段。在学习阶段,从 WWW 网搜索信息需要四个步骤:

1. 用户给出起始 URLs,含有权重的关键字表和电子货币数,并选择信息匹配协议和站点搜索控制策略。PIA 根据上述信息组织 IMA。

^{*})本课题为江苏省应用基础计划资助项目。许晓春 研究生,主要研究方向为 Internet 上的信息发现技术和软件体系结构,杜晓晨 研究生,主要研究方向为 Internet 上的信息发现技术和面向对象软件度量与测试,梅卫锋 研究生,主要研究方向为 Internet 上的信息发现技术和形式化方法,徐永森 教授,主要研究方向为软件工程和软件理论,陈俊良 研究员,主要研究方向为计算机网络。

2. IMA 按照旅行计划,结合对网络状况的动态分析,移动到预定的 ISA,向其提交搜索请求并等待搜索结果。

3. ISA 根据搜索请求在本地进行搜索,将符合要求的结果递交给移动信息 Agent。

4. 移动信息 Agent 完成任务返回到 PIA,PIA 对结果进一步删选整理,也可由用户手动参与进行。

上述四个步骤可以反复进行,直到基本满足用户的要求。在完成了学习阶段后,PIA 帮助用户构建了一个“虚拟信息站点”。这个逻辑意义上的站点包含了被用户认可的 Web 页面,种子 URLs,关键字表,用户的个性化分类,以及 Web 页面的物理地址和更新时间。然后在服务阶段,PIA 动态地检索认可页面的最新版本,如果需要的话,还可以进一步拓展虚拟站点的信息覆盖范围。

用户不仅可以给出初始的 URLs 和关键字表等来创建虚拟站点^[4],还可以对搜索结果手工给出评价来进一步调整虚拟站点。在获得用户的反馈之后,PIA 可以自动调整以适应用户的实际需求,并将用户的偏好存储在用户个人信息库中。

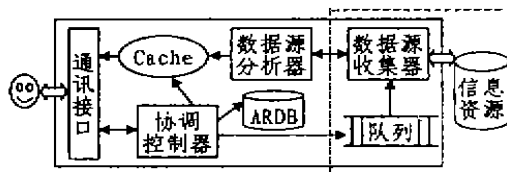


图1 信息服务 Agent

2.2 信息服务 Agent(ISA)

ISA 主要为 IMA 提供各种信息服务,同时还可以和其他 ISA 协作,及时维护和更新资源信息。ISA 由七部分组成,如图1所示,协调控制器通过通信接口管理 IMA。它负责提取搜索请求,往 ARDB(Agent 注册信息库)中填写 Agent 注册信息,最后将任务提交队列。数据收集器从队列中不断提取任务,并在数据源分析器的协作下,按照所拥有的电子货币量,反复地在本地资源库中进行搜索。数据源分析器按照用户选择的信息匹配协议对信息进行过滤,将适用结果存储在 Cache 中,同时将分析出的可能有效链接回送给数据收集器。搜索完成后,协调控制器将 Cache 中的最终搜索结果传送给 IMA,并删除 ARDB 中的相关注册信息。

2.3 信息移动 Agent(IMA)

IMA 是携带着用户个性化请求的网络智能体。它自主地在复杂的网络环境中寻找 ISA,最后将搜索结果返回给 PIA。IMA 主要由四部分组成,如图2所示。

智能决策器是核心控制模块,它可以通过通讯接口获取网络的实时状况,并根据基本信息库和旅行计划的支持,决策自身下一步的移动方向。通讯接口提供 IMA 与 PIA、ISA 的交互。信息库最初存放着用户在客户端配置时提供的电子货币量、带权重的关键字表。随着旅行的深入,搜索信息也将存放其中。IMA 可以部分克隆,以避免自身因携带过多的搜索信息而过度膨胀。

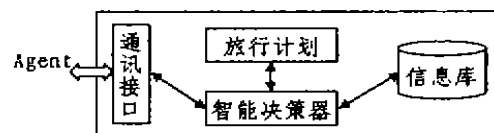


图2 信息移动 Agent

旅行计划描述了 IMA 的搜索策略,Mogent 系统^[5]提出的旅行计划后组装机制很好地支持了本系统个性化、智能化的要求。用户在配置旅行计划时,从 PIA 的搜索策略库中选择适当的逻辑搜索策略,并给出一定顺序的初始 URLs。PIA 将上述两部分进行自动组装,生成最终的旅行计划。用户可以预先设计若干常用的逻辑搜索策略放入搜索策略库,实际使用时只需考虑初始 URLs。

2.4 个人智能 Agent(PIA)

PIA 是用户的虚拟“搜索引擎”。它负责组织用户的请求,发送 IMA,并将搜索结果进行智能化的整理,最后提交给用户。PIA 主要由四部分组成,如图3所示。用户可以通过用户接口按照自己的兴趣爱好来配置搜索引擎,浏览搜索的最终结果,给出反馈意见。用户个人信息库则包含了若干用户感兴趣的虚拟站点、信息匹配策略库、同义词词典和用户基本信息库。信息管理器是 PIA 进行本地整理和再优化的智能处理模块。

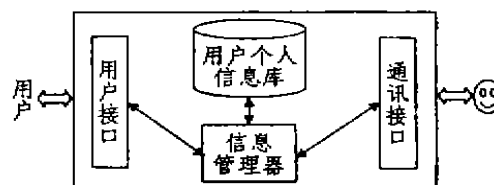


图3 个人智能 Agent

2.5 代理 Agent

在 Agent 社会中我们还设置了一种代理 Agent,它实现 Agent 社会中 Agent 间的介绍和协调功能。在 Agent 社会中,ISA 是唯一直接与信息资源相连接的 Agent,它具体分析网站的组织和信息结构,并通过知识共享的形式将这些信息提供给远端的用户,而代理

Agent 正是为建立共享提供了一种途径。以下将具体介绍 Agent 社会的协作关系。

2.6 Agent 社会的协作关系

为了更有效地改善信息共享、提高搜索效率, Agent 社会采用了多 Agent 协作发现的机制,这在一定程度上解决了单 Agent 搜索带来的一些问题。由于每一个信息 Agent 的信息覆盖面是有限的,对用户来说,特别是那些对一定领域不很熟悉的用户而言,如何选择起始 URLs 是一个很令人头疼的问题。如果在其他信息 Agent,特别是在领域专家的协作帮助下,信息发现的成果就可以很容易被有相同兴趣的用户所共享。

Agent 社会提供了两种手段实现 Agent 之间的协作。

1. 通过代理 Agent 实现协作关系 代理 Agent 的责任就是帮助服务 Agent 和客户 Agent 的相互通讯。代理 Agent 首先向服务 Agent 发出查询,将其基本属性注册在代理 Agent 的服务注册库。然后一旦客户 Agent 提出帮助请求时,代理 Agent 在注册库中找到合适的服务 Agent,并为客户 Agent 和服务 Agent 建立通讯连接。

2. 通过 Agent 俱乐部实现协作关系 Agent 俱乐部是由一些具有相同或相近兴趣爱好的 Agent 组成的团体,俱乐部内的 Agent 经常性地进行交流,使关注信息得到最大限度的共享。

在 Agent 社会中,众多的 PIA 和 ISA 利用上述机制协作发现信息资源,大大拓展了单 Agent 的信息覆盖面,在很大程度上改善了信息共享,提高了信息发现的效率和准确率。

3. 主要特点

作为一个以用户为中心的信息发现环境, Agent 社会与以搜索引擎为代表的传统信息发现工具相比较,在个性化和智能化方面有着显著的特色。以下我们将对这些特色做逐一的介绍。

3.1 自我学习功能

传统的搜索引擎一般不具备自学习能力,而本系统则具有两方面的学习能力。一方面,通过不断地扩展和更新用户个人信息库,以及 Agent 之间的协作,系统能够逐步学习如何来分析网站的结构和信息构成。另一方面,PIA 也能够逐步地学习用户特殊的兴趣爱好,按照不同的受关注程度给出一定的顺序,并提供一系列的辅助关键字表,随着 Agent 的经验积累到一定程度,它就可以相当有效地代表用户完成 WWW 信息发现的任务,并且能不断地适应用户兴趣爱好的变化,同时不断地提高分析网站的能力。

3.2 较强的网络适应性

移动信息 Agent^[6-7]的引入使系统具有较强的网络适应性。移动信息 Agent 能实时分析网络状况,动态地决定下一步动作,有效地克服了网络的不稳定性。另一方面,移动信息 Agent 的引入,避免了网络上大量信息的频繁传输,因而大大减轻了网络负载。

3.3 异步传送功能

使用传统搜索引擎时,用户在发出请求与接收结果之间必须保证网络的正常连接。移动信息 Agent 能分离发送请求和接收结果,使它们异步进行。用户配置完请求后可以完全脱开 PIA,PIA 会进一步组织并发送移动信息 Agent。移动信息 Agent 返回本地后向用户发出提示信息,用户再次通过 PIA 查看最后结果。总的来说,用户只需与本地的 PIA 打交道,而实际搜索完全由三类 Agent 协作进行。

3.4 便利的扩充功能

传统的搜索引擎一般只能针对某种协议进行信息搜索,本系统的数据源分析器提供了多协议知识库支持,因而很容易实现协议的扩充。另外针对传统搜索引擎的结构,ISA 可以在传统引擎的部分内核(如图1右侧虚线框所示)的基础上扩充实现。

3.5 动态监控

PIA 给用户提供了动态监控支持。它自动地监测网站,并将站点的最新信息定期地汇报用户,而在传统搜索引擎时,用户为了获得最新的信息需要经常地进行查询。用户个人信息库中储存着每一个 Web 页面的更新时间。PIA 在执行动态监控中,首先从网站获取 Web 页面的最近更新时间,并将其与用户个人信息库中暂存的 Web 页面进行比较,只在确定 Web 页面已经被更新后,才真正请求获取页面并再一次分析判断是否仍然相关。如果 PIA 掌握了网站的更新特点,利用上述提到的自我学习功能,例如分析网站的更新频度等,从而动态监控功能就能够更为有效和便利地运作。

3.6 同义词词典

同义词词典具有很强的个性化色彩。如我们所知,在传统的搜索系统中,信息的分类都是由机器或相关专家预先完成,它无法按照用户对领域知识的特殊兴趣和理解来过滤信息,这也是导致信息发现效率较低,准确率较差的主要原因之一。

同义词词典则为用户提供了可编辑的自行分类机制^[4],使用户拥有了信息分类的权利,在输入搜索关键字时,同义词词典为用户提供相应的参考关键字表,用户可从中选择也可加入或删除新关键字。用户还能给不同的关键字进行某种排序。

3.7 多层次配置

传统的搜索引擎是一种典型的集中控制式服务,缺乏对个人用户不同需求的考虑,所以用户只能是被动地接收信息,而无法从个人需求、效率等方面对信息加以过滤。本系统提出的个性化多层次配置机制能很好地改变这一状况。用户可以预先选择 URLs 范围,确定最大连接时间、动态监控频度和最大搜索层数等,这种个性化机制可为用户提供更为满意的服务。

表1 Agent 社会与传统搜索引擎的比较

属性	Agent 社会	传统搜索引擎
网络智能	移动 Agent 智能选择移动策略	完全依赖于网络性能
动态监控	自动监测并更新	需要手动更新
异步传送	发送请求和接收结果异步分离	必须保证连接正常
信息分类	用户按照个人的兴趣爱好进行分类	资源信息被预先分类,不考虑用户选择
学习能力	有一定程度上的机器学习能力,并能自动拓展知识	手动支持的人工学习
主动搜索	根据用户的兴趣主动搜索、过滤信息	当用户提交查询请求时才提供服务,无主动搜索
成果共享	Agent 间的多层次共享	信息存放于集中式数据库

结论 Agent 社会是一个具有个性化智能化的多 Agent 信息发现环境。它以用户为中心,从用户的角度出发进行 WWW 搜索。比较传统的 Web 搜索引擎,Agent 社会在个性化和智能化信息发现方面有很多特

色和优势。具体比较见表1。

从上述比较不难发现,Agent 社会能够为用户提供更高效的信息服务,对用户更具亲和力。可以预见,随着智能 Agent 和移动 Agent 相关技术的进一步发展,Agent 社会在未来的信息发现应用中会起到越来越大的作用。

参考文献

- 1 Bowman C M, et al. Scalable Internet Resource Discovery: Research Problems and Approaches. Communications of the ACM, 1994
- 2 Palmas D. Programmung Bots, Spiders, and Intelligent Agent in Microsoft Visual C++ Beijing Hope Electronic Press, October 1999
- 3 Hubns M N, Singh M P. Software Agents and Resource Discovery IEEE Internet Computing, July & August, 1998
- 4 Chen Xin, Chen Junliang, Zhong Yutao, Liu Guodong, Xu Yongsen. VINE. A User-oriented and Cooperative Model in Internet Information Retrieval. Proceedings of ISFST 99, October 1999
- 5 Lu Jian, Zhang Min. Research on the Componentware Framework Based on the Mobile Agent Technology: [Technical Report of State Key Lab. for Novel Software Technology]. Nanjing University, 1999
- 6 Straber M, Baumann J. Mole-A Java Based Mobile Agent System, IPVR. University of Stuttgart, German, 1996
- 7 White J. Telescript Technology: Mobile Agents In Software Agents, J. Bradshaw (Ed.), MIT Press, 1996
- 8 Zhong Yutao, Huang Weiyun, Chen Xin, Xu Yongsen, Chen Junliang. A Multi-Agent Cooperation Model in Internet Information Retrieval. Proceedings of TOOLS Asia'99 and OOT China'99, September 1999

(上接第55页)

而使期望的结论无法得到。在 NBU-CALA+ 中,因果诊断与因果解释一起发挥作用。与因果解释相对照可以看到,正常的一条解释因果链在冲突的地方断开了,而这个冲突处由因果诊断推理机制给出。基本算法框架如下:

(1) 检查当前问题领域属于某一因果关系的作用域 L, 根据索引标记得到可用的默认规则库 D。转(2)。

(2) 根据当前目标库 G, 寻找 D 中的所有默认规则 $R = c_1 \wedge \dots \wedge c_m : s_1 \wedge \dots \wedge s_r / e_1 \wedge \dots \wedge e_n$, 使得它的结论 $e_1, \dots, e_n$ 与 G 中的事实相冲突, 而它的前提 $c_1, \dots, c_m$ 被 G 中的事实所匹配。转(3)。

(3) 对于每一条规则 R, 检查它的判据 $s_1, \dots, s_r$, 得到与事实相冲突的判据条件 s_i 并返回, 结束。

参考文献

- 1 Gan Honghua. A Computational Analysis of Event Causa-

tion, Theory and Applications: [Research Report]. Department of Computer Science, University of Exeter, UK, 1996

- 2 Mackie J L. The Cement of the Universe: A Study of Causation. Oxford University Press, 1974
- 3 Reiter R. A Logic for Default Reasoning. Artificial Intelligence, 1980, 13: 81~132
- 4 Savasdisara P. No 1 Computer-Aided Law Analysis System. A Source of Computer-Aided Support Systems Computer and Law, 1994, 5(2): 28~30
- 5 Shoham Y. Nonmonotonic Reasoning and Causation. Cognitive Science, 1990, 14: 213~252
- 6 全国人大常委会法制工作委员会. 中华人民共和国法律汇编(行政[经济类]卷). 法律出版社, 1996
- 7 干红华, 潘云鹤. 一个基于事件的因果关系实体框架. 计算机科学, 2000, 27(11, 增刊): 63~65
- 8 史济建, 俞瑞钊. 专家系统实现技术. 浙江大学出版社, 1995