

Internet

数据挖掘

数据挖掘

数据挖掘

(10)

# 数据挖掘在 Internet 中的应用

Application of Data Mining in Internet

44-49

陈宁 周龙骧

TP393

TP311.13

(中国科学院数学研究所 北京 100080)

(中国科学院管理决策与信息系统开放研究实验室)

**Abstract** Mining Information and knowledge from Large Database has been recognized by many researchers as a key topic in Database systems. Internet is a huge space of information and knowledge. How to discover new knowledge from it is a important question in the information times. The paper gives some feature and application of data mining in Internet.

**Keywords** Data Mining, Association Rule, World Wide Web(WWW)

近年来,计算机网络的普及使 Internet 成为世界上最大的信息网,目前已有两万多个 WWW 服务器,且每天还在以两百个以上的速度增加,其蕴藏的数据已无法计量,因此如何从这些巨量的数据中发现有用的知识是知识工程研究面临的新课题。数据挖掘就是为满足这种要求而产生并迅速发展起来的,可用于开发信息资源的一种新的数据处理技术。简单地说,数据挖掘是从大量的数据中采掘出隐含的、先前未知的、对决策有潜在价值的知识和规则<sup>[1]</sup>,这些规则蕴含了数据库中一组对象之间的特定关系,揭示出一些有用的信息,为经营决策、市场策划、金融预测等方面提供依据<sup>[1]</sup>。

Internet 上的数据挖掘有如下特点:(1)算法必须有很高的效率:由于数据量非常庞大,而且每天都在迅速增长和更新,从如此巨量的数据中有效地提取有用的信息要求数据挖掘速度必须有很高的效率;(2)有强大的并行性:分布在网络上各个站点的资源通过 Internet 连成一个大分布的数据库,数据的巨大规模和广泛分布要求并行性很高;(3)具有动态性:Internet 中数据更新非常迅速,有些信息可能很快过时,针对当前状态的信息能快速更新知识,提供准确的决策支持要求数据挖掘的动态性;(4)有效地组织和管理数据:目前数据挖掘多应用于关系和面向对象数据库,它们有完美的结构,按照预先定义的模式进行组织、存储和存取,而 Internet 的信息

往往具有半结构化或非结构化特性,难以映射到一个固定的模式,使传统数据模型和数据库系统难以支持 Internet 上的信息管理。因此如何有效地组织和管理数据,从而为数据挖掘提供其所需的源数据也是一个重要问题。

目前 WWW 信息容量已超过 Gopher 和 WAIS 而成为全球最大的信息系统。它是一个由许多被称为 Web 页的超媒体文档组成的集合,用 HTML 书写,包含多种媒体对象和指向其它文档的指针(超链),Web 页面散布在世界各地的 Web 服务器上,每个服务器自主管理自己的资源,没有统一的管理机制,这为数据挖掘带来了更高的难度。以下,本文将着重讨论数据挖掘在 WWW 上的研究方向。

## 一、提高网络响应速度

几乎每一个在 WWW 上浏览的用户都感觉到网络的响应速度太慢了,下载一个 Web 页面需要等待很长时间。这是因为 Web 页面由各种类型数据组成,必须分别下载,特别是当含有大量图象等多媒体数据时下载的速度就更慢,因此如何提高 Web 响应速度是一个重要问题。解决这一问题传统上一般有两种途径:

(1)优化传输,减少阻塞。例如可以将 HTTP 和 TCP 结合起来传输,通过客户端和服务器的链路控制和优化传输。HTTP/TCP 优化方式的好处是被

陈宁 博士生,主要研究方向为数据挖掘。周龙骧 研究员,博士生导师,

阻塞的结点和传输的数据减少,服务器负载减轻,响应迅速,特别是在线路因忙碌而发生 Internet 包丢失时更为有效。另外还可以通过降低图形等多媒体数据的传输质量来提高传输速度。

(2)根据预测,预先传输某些页面。最简单、通用的基于链路的方法是当浏览器装入一个 Web 页时,沿着页面上每一条链路,在后台取出那些 Web 页,以后,当用户点击下一个条链路时,浏览器只需从本地硬盘的高速缓冲区取出,无须访问 Internet。为避免无选择地下载所有链接的 Web 页,记录页面中每个超链接被访问的次数并按频率排序,频率越高说明被访问的可能性越大,选择频率高的页面预先下载。这种方法的缺点是只限于对一个页面中超链的简单统计分析,有时会浪费许多时间和带宽传送不需要的资源;另一种基于历史的方法是记录用户经常访问的站点,并只预取那些站点。例如可以在第一次访问某个站点时,为其建立一个档案,不产生实际的加速效果,当下次再启动浏览器时,它利用空闲时间预装认为用户感兴趣的站点。

以上两种方法都基于客户端,通过分析用户的习惯预测下一步可能的动作,达到加速的目的,但对新用户不能作出任何预测,而且不能减轻网络代理的负担。目前已有 ISP 在自己的服务器上存储流行页面的拷贝,以缓解那些热门站点的拥塞情况。

将关联规则挖掘应用在提高网络响应速度,有效地调度网络代理的缓存上有相对较好的结果。关联规则挖掘<sup>[2]</sup>是数据挖掘研究领域的一个重要方面,由 Agrawal 等人提出,用来从记录交易内容的数据(篮子数据)中挖掘出有用的知识;它把每种商品作为一个项目,顾客每次购买的商品集作为一个事务(交易),关联规则挖掘的目的是找到这样的规则:如果一个顾客在一次交易中购买了 A 和 B,那么他在同一笔交易中也很可能购买 C,表示为  $A, B \Rightarrow C$ 。目前关联规则挖掘在不断改进和扩展,以适应广泛的要求,例如:多层次关联规则,数值型关联规则,带约束的关联规则,模糊关联规则,时序规则等。关联规则,虽然最初针对篮子数据库提出,但同样适合于其它类型的数据库,在 WWW 中关联规则也有广泛的应用。

在服务器端利用关联规则挖掘,不仅可以提高网络的响应速度而且还能有效地调度网络代理的缓存。方法是把 WWW 上每一个站点的 URL 作为一个项目,用户的一次浏览过程作为一个事务,它由该次浏览所访问的所有 URL 构成,记录所有用户的

每一次浏览过程,构成事务库。对事务库进行关联规则挖掘,可从中找到被频繁访问的 URL 集合,得到 URL 的关联规则,存储在服务器的知识库中。当用户在浏览某个页面时,网络代理可根据关联规则预先下载与该页面相关联的页面,即用户很可能访问到的页面,从而提高网络的响应速度。因为关联规则是基于统计规律的,它反映了大多数用户的兴趣,因此能有效地调度网络代理的缓存。

现在有些索引工具把信息按照主题分类,构成层次树,它对研究网络上的多层次关联规则挖掘具有实际意义。另外,网上的信息发布经常具有连续性,例如对某次会议或重大比赛的连续报道,或者一部小说的连载,读者对这些页面的浏览具有时序性。如果记录用户每次浏览过程的时间及用户标志(例如客户端 IP 地址),就可以把所有浏览过程组织成一个时序数据库,利用时序规则挖掘算法找到时序规则。

法国一家日报社 QUEST-FRAME,采用 Client/Server 模式提供有关“面向对象数据库”资料的电子报刊,包含 40 个不同的专题,读者可从中选择感兴趣的专栏阅读。为保证每个用户都有可忍受的响应时间,一种方法是在客户端根据用户的浏览行为进行预测,预先传输可能阅读的页面,但只适用于单个用户;另一种方法是在服务器端利用关联规则把具有相似浏览模式的客户组织到一个服务器上,从而减少服务器缓存和传输页面的数量。首先收集客户端每次阅读的专题集合(浏览模式),作为一条事务存储在数据库中,然后利用关联规则挖掘算法找到访问频率超过给定阈值的专题(项目)集,最后把客户的浏览模式与频繁项目集进行相似匹配,将客户分组到不同的服务器上。

目前大多数关联规则算法针对不变的数据库,数据库一改变,就必须重新运行算法,而关联规则的增量算法也多用于新事务的增加。但 Internet 是一个开放的信息网,网上的任何一个人都能访问各 Web 站点。例如上例中读者的不断加入、退出和阅读专题的经常改变造成数据库的频繁更新(插入、删除、修改),Charpiot 等人针对这种进化的数据库提出了动态关联规则算法<sup>[3]</sup>,利用树结构存储频繁项目集及其关系,当数据库更新时,可迅速修改树结构,找到新的关联规则。这种算法的优点是使发现的知识基于当前数据,把执行代价分布到整个数据收集阶段,而其它静态算法基于不变的历史数据,需要阶段性重复执行。实验表明,即使对不变的数据库,

动态算法的执行效率也优于 Apriori 算法<sup>[2]</sup>,缺点是树结构必须常驻内存,而且由数据项产生关联规则的步骤较复杂。

## 二、路径模式采掘

目前已有的 Web 分析工具仅能提供单个站点的统计信息(例如统计页面被访问次数和用户的职业、年龄调查结果等),而不是整个 Web 中的路径信息,因此不能满足信息服务的飞速发展对用户浏览行为越来越强烈的要求。理解用户浏览路径不仅有助于改进系统设计,而且有助于做出更好的市场决策,例如在适当的页面处增加广告。捕捉用户的浏览路径称为路径模式采掘(Mining Path Pattern)。在 WWW 中,文档通过超链相连便于用户浏览,用户为寻找感兴趣的信息经常通过超链从一个页面跳到另一个页面,用户沿着超链寻找所需要的信息时,有些页面不包含用户感兴趣的内容,而仅仅由于所在位置被访问到,这就增加了从浏览路径中抽取有用信息的难度。

路径模式采掘有三个步骤<sup>[4]</sup>:

(1)把每条原始路径(由浏览中的每个站点构成)转换成浏览子序列集合,每个子序列表示从用户访问起始点开始的最大向前引用,这一步骤删除了由于访问失败或未找到所需信息造成的向后引用;

(2)从上一步得到的最大向前引用中获取大引用序列(large reference sequence),即在全部浏览过程中出现次数超过给定阈值的序列;

(3)确定最大引用序列,即不包含在其它任何最大引用序列中的大引用序列。一个最大引用序列对应于 Web 中一条频繁出现的浏览路径。

如图 1,原始路径为{A,B,C,D,C,B,E,G,H,G,W,A,O,U,O,V},则最大向前引用集为{ABCD, ABEGH, ABEGW, AOU, AOV}。假设

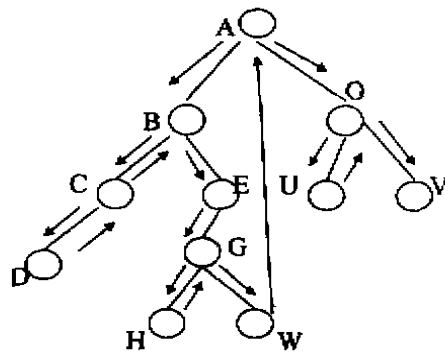


图 1

{AB, BE, AD, CG, GH, BG}是二维大引用序列, {ABE, CGH}是三维大引用序列,则最大引用序列为{ABE, CGH, AD, BG}。如果把 Web 的每个站点看作一个项目,寻找最大引用序列与寻找关联规则中的大项目集有相似之处,它们都满足在事务库中出现足够次数,但前者必须是连续的,而后者仅仅是项目组合。

## 三、WWW 中的智能查询

在过去几年里,Internet 尤其是 WWW 得到了飞速的发展,已发布的主页估计达五千万以上,而且这个数字每天都在增加,然而人们越来越感觉到这个数字时代的图书馆并不象真正的图书馆那样支持有组织的信息管理和检索。恰恰相反,它只是一个又一个杂乱无章的信息仓库。在这个仓库中有书刊、论文、科研资料、会议记录、广告、录像、录音等,转瞬即逝的信息和具有持久意义的重要资料混杂在一起。WWW 中的智能查询包括以下三个方面:

(1)资源发现:其重点是自动生成可查找的索引。Web Robots 浏览数百万个 Web 文档,为之建立索引,通过查找索引,得到包含用户指定关键词的文档。单个索引系统的数据库很难覆盖所有的 Web 资源,因此人们不得不分别检索多个索引系统。为了满足这种要求,出现了独立于索引系统的查询工具 Meta Search,它没有自己的索引数据库,只给用户提供集成的查询界面,用户的查询经过处理后转发给相应的索引信息系统,查询结果返回给用户。另外,Web 查询技术不仅要尽可能地把所有满足查询条件的资源全找出来,不要有遗漏,还要尽可能保证结果的准确性,不要把关系不大的东西返回用户。大多数搜索引擎能按文件的相关程度排列,最相关的排在前面。例如概率方法根据关键字在文件中出现的频率判定相关程度,关键字出现得越多越早相关程度越大。搜索引擎的相关性判断功能越强,对用户的帮助就越大。

利用分类和聚集技术,资源发现系统可把自动文本分类技术用到 Web 文档的分类上,通过识别符合其分类标准的文档,自动发现 Web 目录的结构,它还可以用于过滤查询结果。例如在象“Find product reviews of Encarta”的搜索中,一个发现系统可以通过查找索引找到包含“Encarta”的所有文档,并从中抽取与 product reviews 相关的文档。

(2)信息抽取:发现 Web 资源后,下一个任务就要进行信息的自动抽取。以往需要手工编码从资源

中获取信息,但是当前 Web 的迅速发展要求系统能从各种类型资源中自动抽取信息以减少或避免手工编码。HTML 文档有一定的结构,若能充分利用结构信息,将显著提高查询的准确性;除 HTML 本身的结构外,Web 文档还有更深层次的模式结构。如何识别并利用这些模式是提高查询质量的关键。在这方面已有了一些努力,例如,Ahoy 可以输入用户名,根据其掌握的关于个人主页的模式,找出用户的个人主页;Harvest 系统利用半结构化文档的模式实现信息的自动抽取。它虽然不能发现新文档,也不能获得文档结构的新模式,但它能处理某种已知类型的文档。例如,它能从 Latex 文档中抽取作者和题目信息,FAQ-Finder 能从常用问题文件中抽取问题的答案。它把用户用自然语言提出的问题自动匹配到 FAQ 文件中的相应问题,然后返回对应的答案。以上几种系统都侧重于对 Web 文档的处理,依赖对某些文档类型预先给出的描述。Internet Learning Agent (ILA)根据测试查询和域知识,自动获得对 Web 服务(可查询的表格、私人目录等)的描述。例如 ILA 能从 Internet whois 服务数据库和私人目录中成功地抽取电话号码和 E-mail 地址。但是如何自动获取文档的结构信息,并将它与目录组织结合起来,是今后索引系统重点探索的问题。

(3)信息归纳:由于 Web 太大,又没有数据库那样良好的结构,信息的组织杂乱无章,虽然有些索引系统例如 Yahoo 把搜集到的索引信息按主题分类,组成一个层次型的目录系统,但对资源的分类和组织主要还是依赖于人工,也有一些利用查询操作来分类与组织文档的尝试。利用分类技术可自动组织和管理数据,也可以用于发现用户感兴趣的模式。但是 Web 上的数据量很大,且没有标记,除了聚集技术不需要标记数据外,许多数据采掘技术(例如分类)要求输入的概念事例标记为正例或反例。虽然不确定抽样可减少所需的标记数据数量,但不能解决这个问题。Ahoy 系统利用自身的优势吸引了大批的用户,请求他们的反馈,根据反馈信息较好地解决了标记问题。

#### 四、Web 智能工具

现在 Internet 上有 275,600 个站点,3,000,000 个 Web 页。每天都有大量的新信息涌入网络,同时旧信息以很高的速度被刷新,因此用户必须借助于软件系统来抽取、定位和管理 Web 文档,才能跟上信息的改变速度。这种软件系统被称为 Web 工具。

现有的 Web 工具能力有限,特别是缺乏识别和使用深层语义的能力;Web 用户的信息需求通常比较模糊,有时在寻找过程中才慢慢形成清晰的目标;再加上现有的 Web 查询语言描述能力有限,因此现有的搜索引擎没有一个可以完全有效地搜索网络资源。输入一项检索请求的网络用户经常被数以千计的回答弄得不知所措,检索结果常常涉及一些无关的网址,却漏掉了那些存有重要资料的其它网址,可以说我们现在已经拥有了一个桌面上的图书馆,但却无法有效地用它。一般地根据智能化程度和功能,Web 工具可分为五个层次:

1 根据用户指定的 URL 直接抽取文档,如 Netscape, Mosaic;

2 根据用户给出的初始查询条件找到相关的 Web 页,如 Alta vista。大多数查询工具为 Web 文档建立索引数据库,通过查询索引定位到相关的 Web 页;

3 保存用户的档案(Profile),当发现新的相关信息时,及时地告知用户。例如,Sift 能根据用户的兴趣自动地查找 Web 上的新文档,并通过 E-Mail 传达给用户;

4 对用户信息具有学习和演绎功能的 Web 工具,如 Diff-agent, Letizia。它们把用户的兴趣主题和浏览模式保存在用户档案中;

5 对用户信息和资源信息都具有学习能力的工具,它能帮助系统满足信息交换的要求,从而更好地发现用户最新和相关的信息。

Internet 上的人工智能产品<sup>[5]</sup>有智能搜索引擎,智能浏览器,学习智能体,知识共享智能体等。

Arthur Andersen 的 FSA (Financial Statement Analyzer) 和 Elose (English Language Oriented System For Edgar) 专门用于搜索美国证券交易委员会的 Edgar 商业数据库,这两个系统都内嵌了特定领域的商业知识,并使用了推断-证明式的自然语言理解技术。智能浏览器是基于智能学习理论设计的系统,经过一定的训练后,它可以成为某个领域熟练的搜索专家,帮助用户在网络中查找信息。智能浏览器两个较成功的实验原型是 Web Watcher 和 Letizia。智能体是一个具有控制问题求解机理的计算单元,网络中的智能体通常是一个专家系统、一个过程、一个模块,或一个求解单元。支持智能体独立工作的技术主要包括:模式匹配和逻辑比较;基于知识的快速推理系统;继承其它智能体知识的能力;当数据不完整时,进行缺省推理的能力。

人工智能(AI)可以在 Internet 中引导用户,不仅在用户进行搜索浏览时给予直接的支持,而且能够提供具有独立搜索功能的 Agent(智能体)的幕后支持。

David W. Cheung 设计的智能 Web 工具<sup>[6]</sup>,利用智能 Agent 帮助用户发现相关的和新的信息。它具有以下特点:(1)能够自动地获取用户的兴趣主题,用于发现用户相关信息;(2)能发现用户的浏览模式和信息资源的修改模式,当资源改变时,系统能自动地获取文档的最新版本;(3)能更有效地利用网络资源。通过把多个用户的查询要求按相似性聚集成组,一起执行,从而减少查询次数,避免网络拥挤;(4)将抽取的文档及其全文索引保存在文档数据库中,并对其进行数据采掘,发现各种有用的模式;(5)适用于各种 Web 浏览器,除了标准 http 协议外,不需要额外要求。

#### 用户兴趣模式发现算法<sup>[7]</sup>:

**STEP1** 处理用户访问 log 中的每个文档,用一个二元组集合(关键字,权值)表示用户对该文档感兴趣的关键字和程度。例如<(basketball, 30)(NBA, 50)>;

**STEP2** 调整权值消除噪音干扰。例如有些 Web 文档不提供任何有用信息,而仅仅由于包含许多超链被访问,因此必须调整每个关键字的权值,从而确定文档的相关性。对此,有以下原则:(1)如果某个 Web 页包含 URLs 超链的个数超过给定的阈值,则该页可看作是目录页,因此降低从中抽取的关键字的权值;(2)如果某个 Web 页的访问时间小于给定阈值,则该页可看作是引用页,因此降低从中抽取的关键字的权值;(3)访问路径末端的页面一般包含用户感兴趣的内容。首先根据访问 log 建立浏览图,其中每个文档作为一个结点。如果用户从一个文档进入另一个文档,则在图中有一条边将对应结点相连。从用户的起始页面开始到向后引用的页面为止构成一条浏览路径,位于路径末端且访问时间大于给定阈值的页面即为内容丰富的页面,因此提高从中抽取的关键字的权值;(4)浏览图中结点的扇出表示该页面包含的浏览分支的数目,如果超过给定阈值,则降低从中抽取的关键字的权值;(5)根据关键字在文档中的位置调整权值,例如提高出现在<title>中的关键字权值,其次是<H1>,依次类推;(6)提高出现在 URL 中的关键字权值。

**STEP3** 利用聚集算法从上述元组集合产生用户的兴趣主题,元组间的距离根据相似性计算,相似性越高,距离就越小。当一个新的元组加入时,计算它与每个已有聚集中心点的距离,得到最小距离。如果小于给定的阈值,则将其加入到该聚集中,否则产生一个新的聚集。调整阈值可改变最终得到的聚集数目,每个聚集对应一个兴趣模式,可用聚集的中心元组表示。

系统各模块见图 2,其功能如下:

**Web 浏览器:**用户访问 Internet 的工具,与代理服务器相连;

**代理服务器:**用户和系统交流的媒介。它接受用户发出的 http 请求,抽取 Web 文档,并通过文档管理器存入文档数据库中。当提交一个 http 请求时,先由文档管理器查找数据库,如果所需文档已经存在于数据库中,则返回该文档的备份,否则访问 Internet。另外,代理服务器为每个用户的每次访问建立一条记录,包括用户 id,访问时间,文档内容及大小等属性,存入访问 log 中。

**文档管理器:**是访问文档数据库的中介,负责文档的存取。数据库不仅存储文档的内容及超链结构,还存储文档的全文索引。

**学习 Agent:**通过分析存取 log,发现用户的访问模式和兴趣主题。它为每个用户建立一份档案,包括两种信息:一类是兴趣主题,即关键字及权值的集合,例如:<Basketball, 0.7; Chicago Bulls, 0.4>,说明用户对 Basketball 和 Chicago Bulls 的兴趣分别为 0.7 和 0.4。这种知识用于帮助索引引擎发现所需的信息;另一类是时序访问模式,记录被阶段性访问的文档(如股票和报纸)的地址和时间,这些信息用于监测 Agent 预先取出相关文档。

**搜索引擎:**沿 Web 查找,把找到的文档存入数据库中。它可以利用用户的兴趣主题,从而避免访问与目标无关的页和站点。

**监测 Agent:**阶段性地访问用户感兴趣的页和站点,得到文档的修改模式,从而及时地刷新数据库中的文档,还能预先抽取用户阶段性访问的文档。

**建议 Agent:**把抽取的文档按与用户兴趣的紧密程度排序。

其中,学习 Agent 和监测 Agent 是两个重要的部分。在客户端,学习 Agent 分析访问 log 获取用户的访问模式,即兴趣主题,在服务器端,它能发现

Web 页的修改模式。监测 Agent 监测用户感兴趣的 Web 页的修改。

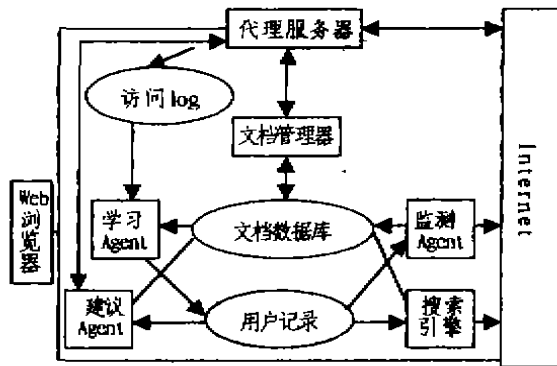


图 2

**结束语** 当前数据采掘和知识发现 (DMKD) 研究正方兴未艾, 预计在 21 世纪将形成更大的高潮, 由于 Internet 用户迅速增加, 为了快速高效地找到网上的知识, 研究在网络下的数据采掘, 特别是在 Internet 上建立 DMKD 服务器与数据库服务器配合, 实现数据采掘, 加强对非结构化数据如文本数据, 图形图像数据, 多媒体数据的采掘, 将是近期数

据采掘的重要课题。

参考文献

- 1 Chen M-S, et al. Data Mining: An Overview from a Database Perspective IEEE Transactions on Knowledge and Data Engineering, 1996, 8(6): 866~883
- 2 Agrawal R, Srikant R. Fast Algorithm for Mining Association Rules. In: Proc. of the 20<sup>th</sup> VLDB Conf Santiago, Chile, 1994. 487~499
- 3 Charpiot B. Dynamic Clustering Using Associative Data Mining. ftp. research. microsoft. com/pub
- 4 Chen M-S, et al. Data Mining for Path Traversal Patterns in a Web Environment. In: Proc. of the 16<sup>th</sup> Intl. Conf. of Distributed Computing System. 1996. 385~392
- 5 伯晓晨. Internet 与 Intranet 中的人工智能技术 计算机世界报, 1998-5-18
- 6 Cheung D W, et al. Discovering User Access Patterns on the World-wide Web; Knowledge Based Systems. Journal Elsevier Science, 1998, 10(7)
- 7 陈宁, 周龙骧. 数据采掘技术. 计算机世界报, 1998-5-24, 1998-6-1

(上接第 36 页)

从表中可以看出: OrbixWeb 远程调用所需时间比 RMI 平均约多 40%, RMI 又比 Visibroker 约多 20%。RMI 居然比某些 CORBA 产品速度还要慢, 这多少让人感到意外。从这个测试结果看不出这两种技术在速度上有什么必然的差异, 具体的差异取决于不同系统的实现。另一方面, 也说明了 RMI 还有很多潜力可以挖掘。

**结论** 本文对开发基于 Java 的分布式应用程序之两种可用的分布对象技术——Java RMI 和 CORBA 的功能和速度进行了对比和分析。CORBA 具有良好的语言中立性, 定义有功能较强的各种服务, 适合于规模较大, 需要与其他语言集成的系统, 其缺点是开发过程较为繁琐, Java 的一些先进特性无法在远程调用的界面上使用。RMI 与 Java 语言结合得十分紧密, 开发方便, 支持对象传值, 但它不太适合需要异种语言集成的场合。RMI 的性能高于一些 CORBA 实现, 又低于另外一些 CORBA 实现。

参考文献

- 1 JDK 1. 1. 6 Document, SUN Microsystems. Available at: http://www.javasoft.com/products/jbk/1.1/docs/index.html
- 2 DCE1. 1: Remote Procedure Call Specification: [Technical Report]. 1995, Available at: http://www.rdg.opengroup.org/public/pubs/catalog/c706.htm
- 3 OMG, X/Open. The Common Object Request Broker: Architecture and Specification, Revision 2. 2. Feb. 1998, Available at: http://www.omg.org
- 4 Morgan B. Java 1. 2 extends Java's distributed object capabilities Java World, April 1998. Available at: http://www.javaworld.com/javaworld/jw-04-1998/jw-04-distributed.html
- 5 OMG, X/Open. CORBA Services; common Object Services Specification, Revised Edition, Updated at Nov 1997, Available at: http://www.omg.org
- 6 Visgenic VisiBroker for Java. Available at: http://www.visgenic.com
- 7 IONA Technologies Orbix Web. Available at: http://www.IONA.com