

教育资源共享网络体系结构及其关键策略

刘方爱 邢长明

(山东师范大学信息科学与工程学院 济南 250014)

摘要 针对中小学的教育资源共享问题,设计了教育资源共享网络体系结构,并对各层节点的功能进行了定义。从网络拓扑结构和通信效率两个角度来探讨降低网络延迟时间和提高网络带宽利用率的方法和措施。首先,在管理节点之间构建逻辑 RP(k)网络,然后基于 RP(k)网络给出了一系列提高系统服务质量的关键策略,包括节点的加入/离开策略、分布式资源检索策略以及节点数据的协同策略。通过理论分析比较,证实了采用 RP(k)互连网络的优势和相关策略的有效性。

关键词 分布式,资源管理,资源共享,体系结构,策略

Architecture of Education Resource Sharing Network and its Key Strategies

LIU Fang-ai XING Chang-ming

(Information Science and Technology College, Shandong Normal University, Jinan 250014, China)

Abstract For the primary and secondary education resource sharing issues, an education resource sharing network architecture was designed and the functions of nodes in each layer were defined in this paper. To reduce the network delay and improve the network utilization, we explored the methods and measures from the two angles of the network topology and communication efficiency. Firstly, we constructed a logic RP(k) network among the management nodes, and then, to improve the QoS of education resource sharing system, a series of key strategies based on the RP(k) were proposed. These strategies included the node joins or leaves strategy, distributed resource retrieval strategy and the data cooperative strategy. Finally, the advantages of the logic RP(k) network and the effectiveness of correlated strategies were confirmed by theoretical analysis and comparison.

Keywords Distributed, Resource management, Resource sharing, Architecture, Strategy

1 引言

因特网的诞生为不同机构的资源共享提供了一种途径。到目前为止,现存的技术仅仅为集中管理的资源提供了一种有效的资源管理途径。在很多场合下,资源都分布存放在不同的地点,需要实施分布管理的方法和技术。针对这一问题观察中小学的教育资源。目前,各地中小学都建立了自己的校园网络,并且每个学校已经积累了较多的网络资源,例如各学科的教学课件、实验案例、教学视频和教学管理软件等。这些资源的分布范围较广,应用面十分普及,怎样在保护知识产权的前提下共享使用这些资源,避免盲目地重复开发,是迫切需要解决的问题。

近年来,围绕网络资源集成共享的研究十分活跃,Web服务、网格计算和对等计算等新技术在网络资源虚拟化、协同、共享和集成方面均进行了有益的尝试。然而,这些研究都处在探索阶段,其在具体应用上还存在着诸多问题。例如,基于P2P的数据共享系统,根据网络拓扑结构主要可分为结构化和非结构化P2P网络。结构化P2P网络(如Chord, CAN, Pastry, Viceroy等)^[1-3],基于分布式哈希函数,数据严格地映射到节点,即如果一个文件 f 的标识符的哈希值与某个节点

P 的标识符的哈希值最相近,则由节点 P 维护文件 f 。由于数据在系统中的放置是严格控制的,因此,这类系统的查询路由比较高效,然而它需要改变现有资源的存储模式,不能实现用户资源的自治性,此不利于资源知识产权的保护。非结构化P2P网络(如Gnutella和Freenet等)^[4],并不将数据放到某个特定的节点,网络中的节点加入或离开所要遵循的规则也比较松散,然而其资源的查询方法采用泛洪方式或流言传播机制,这些机制容易产生大量的垃圾通信,浪费宝贵的网络带宽,降低了互连网络的效率^[5]。

针对上述问题,主要做了如下工作:针对中小学的教育资源共享问题,设计了教育资源共享网络的体系结构,介绍了系统中节点的相互关系和具体功能;在教育资源共享网络中引入RP(k)网络,提高了互连网络的效率;基于RP(k)结构给出了提高系统服务质量的关键策略,并从理论上分析比较了相关策略的性能,结果表明了教育资源共享网络采用RP(k)逻辑网络拓扑的有效性。

2 教育资源共享网络

为了有效地实现中小学教育资源共享,教育资源共享系统应遵循如下原则:为了保护知识产权,各校的资源都保存在

到稿日期:2008-10-27 返修日期:2009-05-22 本文受国家自然科学基金资助项目(60373063,90612003),山东省自然科学基金资助项目(Y2007G11)资助。
刘方爱(1962-)男,博士,教授,主要研究方向为并行处理和互连网络;邢长明(1983-)男,博士研究生,主要研究方向为互连网络和网格计算,
E-mail: xingchm@tom.com.

各自的局域网上;系统中的节点具有自治性,各节点只需维护自己的资源,并且能够方便、灵活地加入与离开;各个节点应该高效地感知整个虚拟组织中的资源,系统能够自动协同维护所有资源,实现资源的共建共享。

2.1 教育资源共享网络体系结构

我国中小学的教育资源组织在行政区和资源所有权方面有其独特性,因此,考虑资源特点、通信效率、组织模式等多种因素,设计教育资源共享网络的体系结构,如图 1 所示。

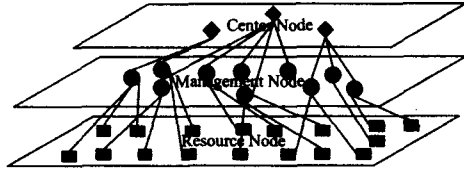


图 1 教育资源网络体系结构

该体系结构为教育资源网络的逻辑结构。该结构共分为 3 层,自下而上依次是:

1)资源节点(resource node)。由分布在各个学校动态、异构的计算机组成。资源节点存储管理本地资源,包括资源的添加、删除和修改,实时地向管理节点报送本地资源的动态变化;存储和维护本地资源的目录信息;响应本地用户的资源请求,缓存本地用户频繁访问的资源信息;资源节点只需维护和查找自己关注的文件,独立自主地设定共享文件范围以及共享文件权限,消除了结构化 P2P 系统忽略节点自治性和数据严格放置的局限性。另外,节点在不同的环境中还可以承担不同的角色工作,既可作为教育资源共享系统中的节点,又可以用作独立自主的资源管理系统,还可以作为其它网络(如 Internet)的节点,实现了资源共享系统节点功能的多样化。

2)管理节点(management node)。由各个学校的一台或多台服务器组成。每个管理节点具有全局唯一的标识,记为 MNID。管理节点存储和维护其辖域内资源节点的元数据信息(关于资源元数据信息结构本文不进行探讨);资源节点请求资源时,管理节点依据资源检索策略查询资源;管理节点还可以依据副本创建策略,存储其辖域内用户频繁访问的资源副本。为了提高服务质量,管理节点之间在逻辑上构成一个结构化网络,避免了无结构化 P2P 系统中由于消息泛洪而引起的大量垃圾通信,提高了系统的通信效率。在实际系统中,管理节点和资源节点可以合二为一,即一台计算机可以同时具有资源节点和管理节点的功能。

3)中心节点(center node)。中心节点根据其管辖的区域可以是省、市、县教育管理部门专用服务器。中心节点作为其辖域内的全局名称服务器,是一个具有固定 IP 的节点,它主要有两个功能:一是为每一个管理节点产生一个全局标识(MNID),另一个是记录管理节点的信息,例如节点的 IP 地址、是否在线等。由于中心节点功能简单,整个系统不会对中心节点有很高的要求。另外,中心节点的数量不受限制,它可以根据自身的能力指定成员的数量,当中心节点的成员数达到上限后,中心节点可以拒绝分配新的 MNID,以保证自身的效率。当然,被拒绝的节点可以去其它中心节点上申请 MNID,这样避免了系统对集中服务器要求很高,而且容易引起单点失效的问题。

从上述功能描述可以看出,教育资源共享系统其实质是一种分布式数据资源共享系统。在这类系统中,数据请求的

响应时间和网络带宽利用率是衡量系统性能的主要指标。针对这类系统,目前主要有 3 种提高系统性能的方法:拓扑优化方法、路由优化方法和基于缓存的方法。本文主要从网络拓扑结构和通信效率两个角度来探讨降低网络延迟时间和提高网络带宽利用率的方法。下面首先介绍一种实用的 RP(k)互连网络。

2.2 RP(k)互连网络

拓扑结构优化是指通过改变节点间的逻辑连接来减少请求响应时间,提高信息资源发现的效率。本文借助于 RP(k)互连网络,在管理节点之间构建一个逻辑连接,从而提高资源检索的效率和实现系统中资源的有效协同。

RP(k)^[6]是被提出的一种网络拓扑结构,它建立在 Petersen 图的基础之上,该网络的结构如下,RP(k)由 $10 \times k$ 个节点组成,每 10 个节点组成一片,片内按照 Petersen 图互连在一起,其节点按照 0, 1, ..., 9 编号,如图 2 所示。 k 个 Petersen 图按照以下办法连接:不同片内编号相同的顶点按照环的结构互连在一起,共 10 个环,称为环 0, 环 1, ..., 环 9。

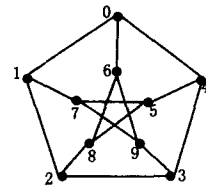


图 2 Petersen 图

RP(k)网络不仅继承了许多 Petersen 图和环的性质,而且还有自身独特的性质,如 RP(k)中含有 $10 \times k$ 个节点,各个节点的度为 5,直径是 $\lfloor k/2 \rfloor + 2$ 等。该网络具有短的网络直径、简单的拓扑结构和方便的路由策略。下文将充分利用这些优良的性质构造教育资源共享网络,并给出提高网络服务质量的关键策略。

2.3 教育资源共享网络的构建

在教育资源共享网络中,如果同层节点内部采用无结构化进行管理,那么在消息传递的过程中容易引起大量的垃圾通信。目前中小学的网络带宽有限,特别是农村中小学的校园网络带宽更低。在这种背景下,合理地选择拓扑简单的结构化网络可以提高网络的可用性。下面以管理层节点为例,给出利用 RP(k)在同层节点间构建逻辑网络的方法。

教育资源网络中的节点是可以动态加入与离开的,所以在动态环境下需要建立管理节点与 RP(k)节点间的映射关系。RP(k)采用如下的编址方式,每一个节点由两部分 (m, n) 决定,其中 m 是片的编号 $(0 \leq m \leq k-1)$, n 是片内编号 $(0 \leq n \leq 9)$ 。管理节点的 MNID 同样可以由片的编号 (m) 和片内编号 (n) 共同标识。对于一个管理节点,片的编号可由中心节点依据该管理节点兴趣值进行分配,片内编号可由中心节点从该片内未划分的编号内随机分配。这样兴趣度相同的若干节点便分配在相邻的片内。为了使节点之间构成一个逻辑 RP(k)网络,每个节点有 3 个片内邻居和 2 个环上邻居,总共 5 项,来维持与系统其它节点的连接。图 3 是一个 RP(k)中节点 $(4, 0)$ 的路由表,它对应的连接如图 4 所示。这样不难看出,当网络中所有的节点都活着时,该网络就成为一个标准的 RP(k)互连网络。

节点标号(4,0)
路由表
片内小邻居:(4,1)
片内邻居:(4,6)
片内大邻居:(4,4)
环上左邻居:(3,0)
环上右邻居:(5,0)

图3 节点(4,0)的路由表(3个片内邻居,2个环上邻居,共5项)

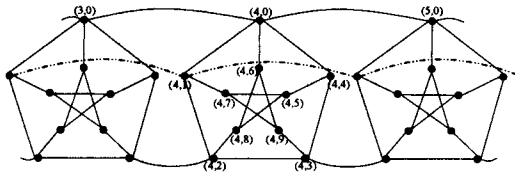


图4 RP(k)部分节点连接方式

3 关键策略

本节论述教育资源共享系统中节点的加入/离开策略、资源检索策略以及数据协同策略。

3.1 节点的加入/离开

教育资源共享网络中的节点是自治的,它们有可能频繁地加入或离开,整个系统处于不断地动态变化中。在 RP(k)网络中用分布式的方式很好地解决了节点加入与离开的问题,无需信息在整个网络中传播。

3.1.1 节点加入

当一个节点加入系统,它需要构建自己的路由表和通知相关节点它的到来。

如果一个节点是首次登录共享网络,将进行以下操作:

(1)该节点在中心节点上登记。这类似用户在 Internet 环境中的邮件服务器上登记。

(2)节点登记后,中心节点为节点分配全局唯一的节点标识,记为 ID。中心节点并不考虑节点当前的 IP 地址,而是通过 ID 来确认节点。ID 由一对标识组成(CNID, MNID),CNID 是节点登记的中心节点服务器的 IP 地址,MNID 是节点在 RP(k)网络中的 ID。

(3)在节点登记的同时,中心节点记录一条由(ID, IP)对组成的记录。其中 IP 为当前登记节点的 IP。

在节点获取 MNID 后,登录节点成为共享网络的一员,为了减轻中心节点的负载,通过以下操作进入共享网络。

(1)节点申请进入网络,将自己当前的 IP 地址发给中心节点。如果它的 IP 地址发生变化,中心节点会自动进行记录更新。

(2)与 Chord 和 Viceroy 一样,假设新节点加入之前要知道一个活节点。假设新节点是 $X=(m, n)$,活节点是 $A=(m_1, n_1)$,新节点 X 路由表中的邻居节点为 $B=(m_2, n_2)$ 。节点 X 可以通过活节点 A 将消息发送给节点 B,发送过程为:若节点 (m_1, n_1) 和 (m_2, n_2) 在同一片内,如果它们直接相连则直接发送,否则先发送到公共邻接节点,由后者再发送到目的节点。当两个节点不在同一片时,按照最短距离,沿着环 n_1 ,将消息送到片 m_2 上的 (m_2, n_1) 节点,然后在同一片内将消息从 (m_2, n_1) 送到 (m_2, n_2) 。

(3)节点 (m_2, n_2) 收到消息后,更新自己的路由表,并将 IP 地址发送给 (m, n) 。

在最坏情况下,活节点 (m_1, n_1) 与节点 (m_2, n_2) 间的消息

传递需要 $\lfloor k/2 \rfloor + 2$ 次中转。由于该消息在传递过程中沿最短路径传送,因而,该策略是最优策略。

3.1.2 节点离开

教育资源共享系统中,节点的离开给整个网络的可生存性带来了挑战。例如,节点 P 须经 X 将消息传递给 Q,如果 X 恰好离线或出错,则 Q 就不能接受到相关消息。因而节点的动态性影响到结构化网络的连通性及性能。解决该问题可采用代理机制,对于系统中的任一节点 X,可邀请一个或数个节点形成一个虚拟节点并互为代理,当自己离线时代替自己收发相关消息。这些节点的行为必须满足以下条件:

$$\cup Online_{proxy} \geq Offline_x \text{ 并且 } \cup Online_{proxy} \cup Online_x = 7 * 24$$

上式中, $\cup Online_{proxy} \geq Offline_x$ 表示代理节点在线时间大于等于节点的离线时间;而 $\cup Online_{proxy} \cup Online_x = 7 * 24$ 表示节点及其代理每时每刻至少有一个节点在线,即代理节点的在线时间必须覆盖节点 X 的离线时间。

3.1.3 容错性

在分布式系统中,由于节点的自治性,节点很可能意外地崩溃或离开,低节点度的结构化网络的连通性及其性能因此会变得很差。通常,系统通过维护其它节点的路由表来解决此类问题,例如 Chord 的后继列表(successor list)和 Viceroy 的桶(bucket)。本文中,通过代理机制解决这一问题,这样容错性问题便同路由设计问题分开处理,这与 Koorde^[7] 的主张相同。

3.2 资源检索策略

在教育资源共享系统中,用户的检索方式通常有两种,一种是全局检索,另一种是局部检索。全局检索是指搜索整个系统内部的所有节点,而局部检索是指只检索系统内部特定节点的资源,例如每个学校都有自己的优势学科,用户在进行资源检索时,可能只选择该学科较强的几所学校进行检索。下面针对这两种检索方式,给出具体的检索策略。

3.2.1 全局检索

对于全局检索,用户提交的查询首先由本地节点 L 解析后,提交给其管理节点 (m, n) ,然后以并行方式进行处理:如果节点 (m, n) 可能有答案,则搜索本地元数据文件(否则不检索本地元数据),同时把该查询消息发送到整个虚拟组织内的所有管理节点,其消息传播过程分为两个阶段。在第一个阶段,节点 (m, n) 将消息发送到 m 片内的所有节点;在第二个阶段, m 片内的节点将消息按照 10 个环向两个方向进行扩散。这样实现整个系统的全局检索需要 $\lfloor k/2 \rfloor + 2$ 次通信。

3.2.2 局部检索

对于局部检索,其实质是提交查询的用户所有区域的管理节点 (m, n) 和其查询目标节点所在区域的管理节点 (m_2, n_2) 之间检索消息的传递过程。该过程可按照上述管理节点加入时两节点间的消息传递过程进行。在最坏情况下,该过程同样需要 $\lfloor k/2 \rfloor + 2$ 次通信。

3.3 数据一致性策略

在教育资源共享网络中,为了提高系统的可扩展性、容错性、可用性和减少资源传输的时间,可以引入副本机制,即管理节点为其辖域内用户频繁访问的资源建立副本。以往的研究认为资源共享网络(如 P2P 系统)中共享的是静态资源,没有经常需要更新的内容,对于这些资源的访问通常是只读形

式,如 MP3,Video 等。然而,在教育资源共享系统中,资源以只读的形式转换为可读写方式,确保其一致性是保证其正确运行的关键。

对于分布式系统而言,数据更新时机和一致性策略的选择至关重要。由于节点间的数据传输必然消耗带宽,如果在查询处理时执行数据更新操作,必然会影响到系统的响应时间。为了消除该影响,系统可选择大部分节点比较空闲的时机(如午夜)更新数据。实验证明该机制是有效的。

选择好数据更新时机后,接下来就要进行数据的一致性操作。实现整个虚拟组织内所有节点的数据的一致性,就是要求每个节点都向其它所有节点发送一个更新消息。可以采用一种简单策略,即每个节点都做一次广播操作。在 RP(k)网络上所需通信时间为: $10 \times k \times (\lfloor k/2 \rfloor + 2)$ 次。然而,这样的算法效率太低,现在考虑另外一种策略。

该策略分 3 步进行。

第一步,同一片内的节点实现消息的相互传递。其算法如下:

- (1) 各个节点并行地向其相邻的 3 个节点发送消息;
- (2) 收到消息的节点向其相邻的节点发送如下两个消息:这个消息既不是发送节点的消息也不是目的节点的消息。

算法的正确性可以验证,按照图 2 的标号,经过第 1 步后,各个节点所含数据为 0: {0,1,4,6} 1: {0,1,2,7} 2: {1,2,3,8} 3: {2,3,4,9} 4: {0,3,4,5} 5: {4,5,7,8} 6: {0,6,9,8} 7: {1,5,7,9} 8: {2,5,6,8} 9: {3,6,7,9}

经过算法的第 2 步后,每个节点上都含有 10 个消息。在节点间不可重叠发送消息的情况下,该算法第一步需要两次通信操作,第二步需要 4 次通信操作,因而共需 6 次通信操作。

第二步,将每个节点上的 10 个消息打包,即将 10 个消息封装在一起。

第三步,每个节点沿它所在的环,按顺时针方向,向相邻节点发送数据包,并接收到达的数据包,重复 $k-1$ 次;在每个节点上,解开所接收的数据包,这样就完成了整个协同过程。

整个过程的第一步需要 6 次通信操作,第三步需要 $(k-1)$ 次操作,因此整个过程需要 $(k+5)$ 次操作。

4 相关研究及性能分析

寻找良好的拓扑结构是实现网络资源共享的基础。其研究集中在结合共享资源的特征,寻找适合其特点的网络拓扑结构。到目前为止,其代表性的工作有 Grid 和 P2P。Grid 技术主要侧重计算资源的共享,而 P2P 系统集中在数据资源的共享。文献[8,9]对结构化和非结构化的 P2P 网络进行了大量的研究。对无结构化的 P2P,例如 Gnutella, KaZaA 文件的位置是随机的,与网络的拓扑无关。在这类系统中利用“泛洪”机制进行资源的查询请求,因而带来了效率低下的问题,如通信消耗大、反应时间慢等。据文献[10]统计,在当前 Internet 主干网络的通信开销中,有 40% 以上被 P2P 应用消耗。对于结构化的 P2P,例如 Chord, CAN, Pastry 和 Viceroy 分别基于环、Tours、Hypercube 和 Butterfly 网络实现了结构化 P2P 系统,从根本上改变了 P2P 无结构的状态。然而,这些系统需要借助于分布式 Hash 表,这不仅限制了资源的存放位置,而且不适合复杂和模糊的检索。

本文将系统分为 3 个层次,给出了教育资源分布式共享的系统架构,为了提高系统的服务质量,在管理层节点构建 RP(k)逻辑网络,以减少垃圾通信的数量。影响网络通信效率的因素主要有节点度、网络直径^[11,12]。表 1 是基于这些具体指标,对管理节点采用 RP(k)与几个常用互连网络的性能比较。

表 1 RP(k)与其它互连网络的比较

互连网络	RP(k)	ring	2d-Tours	Hypercube
正规性	是	是	是	是
节点数	$10 * k$	n	n	n
节点度	5	2	4	\log_2^n
网络直径	$\lfloor k/2 \rfloor + 2$	$n/2$	$2 * \lfloor k/2 \rfloor$	\log_2^n

在表 1 中,节点个数表明互连网络的规模。节点度,指节点连接到其它节点边的数目,它决定了每个节点需要维护的邻居的个数。网络直径,是网络中两个节点间的最远距离。网络的距离越大,两个节点间的通信延迟就越大,通信效率越低。通过比较可以看出,RP(k)网络作为一个常数度的拓扑结构,具有较强的扩展性,其网络直径小于环,而且当网络节点数小于 300 时,RP(k)的直径小于 2d-Tours 的直径。因此,采用 RP(k)拓扑可以降低网络的通信开销,提高系统的服务质量。

结束语 本文针对教育资源分散性的特点,设计了教育资源共享网络的体系结构,并给出了系统中的关键策略,实现了教育资源的分布式管理。为了提高系统的服务质量,降低网络的通信开销,本文将整个系统分为 3 个层次,并在管理层构建结构化的 RP(k)逻辑网络。针对教育资源共享网络,设计了系统中的节点的加入/离开策略、资源检索策略和数据一致性策略,并给出了相应策略的理论性能分析。通过具体参数的比较,表明了管理层采用 RP(k)拓扑结构比环、Mesh 和 Hypercube 结构的优势。另外,针对教育资源管理问题的研究,依据上述理论,设计了教育资源共享原型系统,证实了上述理论的可行性。下一步将选择具体学校进行系统实际性能的分析。

参考文献

- [1] Stoica I, Morris R, Karger D, et al. Chord A Scalable Peer-to-Peer Lookup Service for Internet applications[C]// Proceeding of ACM SIGCOMM, San Diego, California, USA, 2001
- [2] Rowstron A, Druschel P. Pastry: Scalable, Decentralized Object Location and Routing for Large-Scale Peer-to-Peer systems[C] // Lecture Notes in Computer Science 2218. Heidelberg: Springer-Verlag, 2001: 329-350
- [3] Malkhi D, et al. Viceroy: A scalable and dynamic emulation of the butterfly[C]// Proc. PODC. New York: ACM Press, 2001
- [4] Gnutelliums L L C. Gnutella protocol specification version 0. 4. [OL]. 2008. http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf
- [5] 冯国富,张金城,姜玉泉,等. 无结构 P2P 覆盖网络的拓扑优化[J]. 软件学报, 2007, 18(11): 2819-2829
- [6] Liu Fang-Ai, Liu Zhi-Yong, Qiao Xiang-Zhen. A practical interconnection network RP(k) and its routing algorithms[J]. Science in China (Serial F), 2001, 44(6): 461-473
- [7] Kaashoek M F, Karger R. Koorde: A simple degree optimal distributed hash table[C]// Proceedings of the 2nd International Workshop on P2P Systems, Berkeley, CA, 2003: 98-107

(下转第 160 页)

定义 7(约简) 目标信息系统中删除一个属性元素得到一个新的目标信息系统,如果新目标信息系统的近似质量与原目标信息系统的近似质量相同或接近,则新目标信息系统是原目标信息系统的一个约简。

从定义中可以看出:在约简中删除的是对一个目标信息系统中近似质量没有影响或影响不大的一个元素;一个目标信息系统可能存在多个约简,但每个约简可能都不是最终的约简。

(3) 广义变精度粗糙模糊集模型中的近似约简算法

算法:近似约简

输入:一个目标信息系统。

输出:目标信息系统的约简。

步骤 1 初始化。求原目标信息系统的近似质量 γ_R ;

步骤 2 求原目标信息系统中删除一个属性元素的关系 R' 、由 R' 确定的划分 $U/R' = \{C_1, C_2, \dots, C_t\}$ 和新目标信息系统的近似质量 $\gamma_{R'}$;

步骤 3 比较 γ_R 和 $\gamma_{R'}$,如果相同或接近则新目标信息系统是原目标信息系统的一个约简;

步骤 4 依次按步骤 2、步骤 3 对目标信息系统中的各个属性进行试探,看所得到的新目标信息系统是否是原目标信息系统的约简。

从上面的算法可以看出:原目标信息系统的约简可能不止一个;该算法是对目标信息系统中的各个属性进行试探,每次可能都会得到原目标信息系统的一个约简,但这个约简又可能不是最终的约简;另外,步骤 3 中的接近程度要根据具体的应用要求来确定达到什么值时才算是得到目标信息系统的约简。

8 应用举例

用下面的例子来分析一个目标信息系统。如表 1 所列,其中 $U = \{x_1, x_2, x_3, x_4, x_5\}$ 为论域, (c_1, c_2, c_3) 为条件属性, (d) 为决策属性。设 $A = 0.6/x_1 + 0.4/x_2 + 0/x_3 + 0.5/x_4 + 0.7/x_5$ 是论域 U 上的模糊集。

表 1 目标信息系统

U	c_1	c_2	c_3	d
x_1	A_2	B_2	C_2	D_1
x_2	A_1	B_2	C_2	D_2
x_3	A_3	B_1	C_1	D_2
x_4	A_1	B_1	C_1	D_1
x_5	A_1	B_1	C_2	D_1

根据前面的约简算法可知目标信息系统的近似质量 $\gamma_R = 0.44$;当去掉属性 c_3 时,新的近似质量 $\gamma_{R'} = 0.4$;而当去掉属性 c_1 或 c_2 时,新的近似质量 $\gamma_{R'} = 0.44$ 。这说明去掉属性 c_3 时会严重影响新目标信息系统的近似质量,而去掉属性 c_1 或 c_2 时不会影响新系统的近似质量,这足以表明属性 c_1 或 c_2 在系统中是可以约简的,故原目标信息系统的约简为 (c_1, c_3) 或 (c_2, c_3) ,即 (c_1, c_3) 或 (c_2, c_3) 可以和条件属性集 (c_1, c_2, c_3)

产生同样的决策能力。

结束语 广义变精度粗糙集模型在处理含有不完备信息的目标系统中是非常有用的,粗糙模糊集模型是在等价关系下对模糊数据进行处理,把这两种方法结合起来处理目标信息系统中含有不完备的数据具有广阔的应用前景。本文就是基于这一目的对广义变精度粗糙模糊模型进行了详尽的探讨并给出了它的应用理论与方法。但是,当该模型中的一般关系变成论域上的模糊关系时,如何进行数据处理将是我们要研究的内容。

参考文献

- [1] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177: 3-27
- [2] Pawlak Z, Skowron A. Rough sets: Some extensions[J]. Information Sciences, 2007, 177: 28-40
- [3] Pawlak Z, Skowron A. Rough Sets and Boolean Reasoning[J]. Information Sciences, 2007, 177: 41-73
- [4] Ziarko W. Variable precision rough set model [J]. Journal of computer system science, 1993, 46(1): 39-59
- [5] 陶志, 许宝栋, 汪定伟, 等. 基于变精度粗糙集理论的粗糙规则挖掘算法[J]. 信息与控制, 2004, 33(1): 18-22
- [6] Beynon M. Reducts within the variable precision rough sets model: A further investigation[J]. European journal of operational research, 2001, 134: 592-605
- [7] 张文修, 梁怡, 吴伟志, 等. 信息系统与知识发现[M]. 北京: 科学出版社, 2003: 56-67
- [8] Inuiguchi M. Structure-Based Approaches to Attribute Reduction in Variable Precision Rough Set Models[C]//Proceeding of 2005 IEEE International Conference on Granular Computing, Beijing China, IEEE GrC2005, 2005: 34-39
- [9] Mieszkowicz - Rolka A, Rolka L. Remarks on approximation quality in variable precision fuzzy rough sets model[C]// Rough Sets and Current Trends in Computing: 4th International Conference, RSCTC 2004. Heidelberg: Springer-Verlag, 2004: 402-411
- [10] 巩增泰, 孙秉珍, 邵亚斌, 等. 一般关系下的变精度粗糙集模型[J]. 兰州大学学报: 自然科学版, 2005, 41(6): 110-114
- [11] 张亚军, 王艳平. 基于覆盖的变精度粗糙集模型[J]. 辽宁工学院学报, 2006, 26(4): 274-276
- [12] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences, 2003, 152: 217-230
- [13] Mieszkowicz - Rolka A, Rolka L. Fuzziness in Information Systems[J]. Electronic Notes in Theoretical Computer Science, 2003, 82(4): 1-10
- [14] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2003: 41-131
- [15] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766
- [8] Kaya O S. A glance at Peer to Peer systems[R]. TR-CTIT-05-21, University of Twente, 2005, 1381-3625
- [9] 王勇, 云晓春, 李奕飞. 对等网络拓扑测量与特征分析[J]. 软件学报, 2008, 19(4): 981-992
- [10] Foster I, Iamnitchi A. On death, taxes, and the convergence of peer-to-peer and grid computing [C]//The 2nd Int'l Workshop on Peer-to-Peer Systems (IPTPS'03). Berkeley, USA, 2003
- [11] 陈贵海, 须成忠, 沈海英, 等. 一种新的常数度数的 p2p 覆盖网络[J]. 计算机学报, 2005, 28(7): 1084-1095
- [12] 刘方爱, 刘志勇. 环、mesh 嵌入 RP(k) 网络[J]. 中国科学 E 辑, 2004, 34(8): 919-929