

DRSA 中基于支配矩阵的类合集近似的新方法

李明, 张保威, 赵丽

(兰州理工大学计算机与通信学院, 兰州 730050)

摘要: DRSA(Dominance-based Rough Set Approach)是经典粗糙集理论在支配关系下的扩展, 而类合集近似是 DRSA 中非常重要的概念与方法。传统的方法是利用原始的定义来求取类合集的上下近似, 针对其计算复杂的缺点, 该文提出了一种基于支配矩阵的类合集近似新方法, 不仅将计算复杂度降到约 $1/k$ 倍 (k 为决策类的个数), 而且简洁直观。最后的实例证明了新方法的有效性及其可行性。

关键词: 粗糙集理论; DRSA; 类合集近似; 支配矩阵

A New Method for Approximation of Class Unions Based on Dominance-matrix Within DRSA

LI Ming, ZHANG Baowei, ZHAO Li

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050)

【Abstract】 Dominance-based rough set approach (DRSA) is an extension of classical rough set theory (RST). Approximation of class unions is a very important approach in DRSA. Aiming at the disadvantage of the classical method, this paper presents a new methodology for approximation of class unions based on dominance-matrix. It only needs to calculate the dominance-matrix and does not need to consider the preference relations one by one. Thus it greatly simplifies the process and decreases the computing complex by k times (where k is the number of the decision classes). Besides it is intuitive and efficient. The example illustrates its feasibility and efficiency.

【Key words】 Rough set theory; Dominance-based rough set approach (DRSA); Approximation of class unions; Dominance-matrix

1 概述

粗糙集理论^[1]是一种新型的处理模糊和不确定问题的数学工具, 已经被证明在处理不一致问题方面有其独特的优势, 它基于个体之间的等价关系利用上下近似、约简及依赖等基本概念获取决策规则。然而当我们引入个体之间的序关系的时候, 经典的粗糙集理论对处理不一致等问题显然缺乏有效性。为了更好地说明这一问题, 举一个简单的例子, 例如保险公司用一系列指标对 A、B 两家公司进行破产风险评估, 包括产品质量、市场占有率、贷款比率等衡量标准, 如果 A 公司的贷款比率高于 B 公司, 而其他指标上述两家公司都相同的话, 这种情况下从破产风险的角度来看, A 公司的破产风险明显要高于 B 公司, 可是 B 公司却被决策者归为破产风险较高的类, 这就出现了不一致。如果用经典的粗糙集理论来处理这一问题, A、B 两公司会被认为是可以分辨的而加以分离, 这样挖掘出来的规则显然区分出不一致。

针对经典粗糙集理论在处理不一致等问题上的缺点, 文献 [2-4] 提出了基于支配关系的粗糙集理论 DRSA (Dominance-based Rough Set Approach)。用支配关系替换了经典粗糙集中不可辨识关系, 对经典粗糙集理论进行了扩展, 利用上下近似将决策规则分为确定规则 (certain rules) 和不确定规则 (doubtful rules), 这样就增加了决策规则的适应性及鲁棒性。类合集的上下近似是 DRSA 中非常重要的概念与方法, 只有得出类合集的上下近似, 才能在此基础上挖掘规则。但 DRSA 提出以后并没有提出有效的求取近似的算法, 只能利用最原始的定义来进行求解, 复杂而繁琐, 针对这一缺点我们对其进行了分析探讨, 构造了支配矩阵 (dominance-

matrix), 提出了基于支配矩阵的类合集的近似方法, 无需逐个比较个体之间的序关系, 只用将原始决策表转化为支配矩阵, 利用支配矩阵可以快速直观地获取所有类合集的上下近似, 而且计算复杂度降低了 k 倍 (其中 k 为决策类的个数)。

2 基于支配关系的粗糙集理论

2.1 DRSA 中数据的表示

在 DRSA 中一般用信息表来表示数据, 信息表的列表示属性, 这里称之为标准 (criteria), 行表示个体 (object), 标准的决策信息表是一个四元组

$$S = (U, C \cup d, V, f)$$

其中 U 为论域, C 为条件属性, d 为决策属性, $V = \bigcup_{a \in C \cup d} V_a$, V_a 为属性 a 的值域, f 为由论域到值域之间的信息函数 $f: U \times A \rightarrow V$, 它指定论域中每一个对象 x 的属性值, 即对 $x \in U, a \in C \cup d$ 有 $f(a, x) \in V_a$ 。

对任意的 $q \in Q, \succeq_q$ 表示在标准 q 上的一种序关系, $x \succeq_q y$ 表示在标准 q 上个体 X 至少与 Y 一样优。对决策标准 d 来说, 其值域 $V_d = \{v'_t, t \in T\}, T = \{1, \dots, n\}$ 在 \succeq_q 下对论域 U 产生一划分 $Cl(d) = \{Cl_t, t \in T\}$, 其中 $Cl_t = \{x \in U : f(x, q) = v'_t\}$, 我们称 Cl_t 为类, 对任意的个体 $x \in U$ 属于且只能属于一个类 $Cl_t \in Cl(d)$, 而且我们认为类之间也是有序关系的, 对所有的

基金项目: 甘肃省自然科学基金资助项目(3ZS042-B25-014); 甘肃省教育厅科研基金资助项目(0416B-04)

作者简介: 李明(1959—), 男, 教授, 主研方向: 数据库, 数据挖掘; 张保威、赵丽, 硕士生

收稿日期: 2005-09-14 **E-mail:** bwzh1980@163.com

$r, s \in T, r > s$ 有: if $x \in Cl_r, y \in Cl_s$, then $x \succeq y$ 。为了使挖掘出来的规则更具适用性,类合集的概念被引入,类的上合集和下合集的定义如下:

类的上合集 (upward unions of class)

$$Cl_t^{\succeq} = \bigcup_{s \geq t} Cl_s, t \in T;$$

类的下合集 (downward unions of class)

$$Cl_t^{\preceq} = \bigcup_{s \leq t} Cl_s, t \in T。$$

$x \in Cl_t^{\succeq}$ 表示 x 至少属于类 Cl_t , $x \in Cl_t^{\preceq}$ 表示 x 至多属于类 Cl_t 。

2.2 支配关系及类合集的近似

对一个给定的决策信息表 S ,我们称个体 x 在标准 $P \subseteq C$ 上支配 y ,如果在每一个标准 $q \in P, x \succeq_q y$,记为 $x D_P y$ 。 $\forall P \subseteq C$,支配关系 D_P 满足自反性和传递性。给定 $P \subseteq C$,所有支配 x 的个体定义为 $D_P^+(x) = \{y \in U : y D_P x\}$,所有被 x 支配的个体为 $D_P^-(x) = \{y \in U : x D_P y\}$ 。

对类 Cl_t 的上合集 Cl_t^{\succeq} ,其上下近似分别定义如下:

$$\underline{P}(Cl_t^{\succeq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\succeq}\}$$

$$\overline{P}(Cl_t^{\succeq}) = \{x \in U : D_P^-(x) \cap Cl_t^{\succeq} \neq \emptyset\}$$

同理 Cl_t 的下合集 Cl_t^{\preceq} 的上下近似分别定义为

$$\underline{P}(Cl_t^{\preceq}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\preceq}\}$$

$$\overline{P}(Cl_t^{\preceq}) = \{x \in U : D_P^+(x) \cap Cl_t^{\preceq} \neq \emptyset\}$$

与传统粗糙集理论类似 Cl_t^{\succeq} 与 Cl_t^{\preceq} 的边界分别定义为

$$Bn(Cl_t^{\succeq}) = \overline{P}(Cl_t^{\succeq}) - \underline{P}(Cl_t^{\succeq}), Bn(Cl_t^{\preceq}) = \overline{P}(Cl_t^{\preceq}) - \underline{P}(Cl_t^{\preceq})$$

2.3 决策规则的获取与表示

DRSA 中将决策规则分为:确定规则 (certain rules),可能规则 (possible rules) 与近似规则 (approximation rules)。确定规则是从类合集的下近似中得到的,可能规则是从类合集的上近似中得到的,而近似规则则是从类合集的边界中得到的。

(1) D_{\geq} -decision rules

$$\text{if } f(x, q_1) \geq r_{q_1} \text{ and } f(x, q_2) \geq r_{q_2} \text{ and}$$

$$\dots f(x, q_p) \geq r_{q_p}, \text{ then } x \in Cl_t^{\succeq}$$

其中: $P = \{q_1, q_2, \dots, q_p\} \subseteq C$,

$$(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}, t \in T。$$

我们称之为“至少”规则,它是从 Cl_t^{\succeq} 的下近似中得到的。

(2) D_{\leq} -decision rules

$$\text{if } f(x, q_1) \leq r_{q_1} \text{ and } f(x, q_2) \leq r_{q_2} \text{ and}$$

$$\dots f(x, q_p) \leq r_{q_p}, \text{ then } x \in Cl_t^{\preceq}$$

其中: $P = \{q_1, q_2, \dots, q_p\} \subseteq C$,

$$(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}, t \in T。$$

我们称之为“至多”规则,它是从 Cl_t^{\preceq} 的上近似中得到的。

(3) D_{\leq} -decision rules

$$\text{if } f(x, q_1) \geq r_{q_1} \text{ and } f(x, q_2) \geq r_{q_2} \text{ and}$$

$$\dots f(x, q_k) \geq r_{q_k} \text{ and } f(x, q_{k+1}) \leq r_{q_{k+1}} \text{ and}$$

$$\dots f(x, q_p) \leq r_{q_p}, \text{ then } x \in Cl_t \cup Cl_{t+1} \cup \dots \cup Cl_s$$

其中: $O' = \{q_1, q_2, \dots, q_k\} \subseteq C, O'' = \{q_{k+1}, q_{k+2}, \dots, q_p\} \subseteq C,$

$P = O' \cup O'', O'$ 与 O'' 并不一定是必须相交的。

$$(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}, s, t \in T \text{ such that } t < s。$$

这些规则是从 Cl_t^{\succeq} 与 Cl_t^{\preceq} 的边界中得到 d 的。

可以看到基于支配关系所挖掘出来的规则在语义上更具表达性和适用性。它用“ \geq ”和“ \leq ”代替“ $=$ ”,使得仅用更少、更简单的规则就可以表达原有信息所蕴含的决策。

3 基于支配矩阵的类合集近似新方法

3.1 问题的形式化描述

类 Cl_t 的上下合集的上下近似的传统定义如下:

$$\underline{P}(Cl_t^{\succeq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\succeq}\}$$

$$\overline{P}(Cl_t^{\succeq}) = \{x \in U : D_P^-(x) \cap Cl_t^{\succeq} \neq \emptyset\} = \bigcup_{x \in Cl_t^{\succeq}} D_P^+(x)$$

$$\underline{P}(Cl_t^{\preceq}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\preceq}\}$$

$$\overline{P}(Cl_t^{\preceq}) = \{x \in U : D_P^+(x) \cap Cl_t^{\preceq} \neq \emptyset\} = \bigcup_{x \in Cl_t^{\preceq}} D_P^-(x)$$

$$Bn(Cl_t^{\succeq}) = \overline{P}(Cl_t^{\succeq}) - \underline{P}(Cl_t^{\succeq})$$

$$Bn(Cl_t^{\preceq}) = \overline{P}(Cl_t^{\preceq}) - \underline{P}(Cl_t^{\preceq})$$

以前只能用以上定义来求取类合集的近似,需要逐个考虑个体之间的序关系,并计算每一个个体所支配的及支配该个体的所有元素,计算繁琐而复杂,极易产生错误,其计算复杂度为 $2kn^2$,其中 n 为决策表中个体元素的个数, k 为决策类的个数。为了简化计算过程,降低计算的复杂度,我们对类合集近似的计算过程进行了探讨,构造支配矩阵,提出了一种基于支配矩阵的类合集近似的新算法。

3.2 支配矩阵

$S = \langle U, Q, V, f \rangle$ 是一个决策系统,假设它有 n 个个体元素,其中 $U = \{x_1, x_2, \dots, x_n\}$, $Q = C \cup d$, c 表示条件属性, d 表示决策属性。则支配矩阵的定义如下:

支配矩阵分为两类,上支配矩阵 $M(Cl_t^{\succeq})$ 及下支配矩阵

$M(Cl_t^{\preceq})$ 。

$$M(Cl_t^{\succeq}) = [C_{i,j}]_{n \times n}$$

$$\text{其中: } C_{i,j} = \begin{cases} 1 & x_i D_c x_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

D_c 表示属性 c 上的支配关系,矩阵可被描述为

$$M(Cl_t^{\succeq}) = [C_{i,j}]_{n \times n} = [b_1, b_2, \dots, b_n], b_i \text{ 是矩阵 } M(Cl_t^{\succeq}) \text{ 的第 } i \text{ 列,}$$

$$b_i = [C_{1,i}, C_{2,i}, \dots, C_{n,i}]^T, \tau \text{ 表示向量的转置。}$$

类似可以定义 $M(Cl_t^{\preceq}) = [C'_{i,j}]_{n \times n} = [b'_1, b'_2, \dots, b'_n]$

$$\text{其中 } C'_{i,j} = \begin{cases} 1 & x_j D_c x_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

3.3 基于支配矩阵的类合集近似方法

根据 3.2 节中的定义可以将原始的决策信息表中转化为两个二元支配矩阵,为了简便起见在这里仅考虑类的上合集的近似,下合集近似的求法类似。

$M(Cl_t^{\succeq}) = [C_{i,j}]_{n \times n} = [b_1, b_2, \dots, b_n]$,对 Cl_t^{\succeq} 首先找出所有出现在 Cl_t^{\succeq} 中的个体元素,将这些元素在矩阵 $M(Cl_t^{\succeq})$ 中对应的行向量置零得到 $M'(Cl_t^{\succeq})$,考察 $M'(Cl_t^{\succeq})$,如果存在第 i 列为零向量,则 $x_i \in \underline{C}(Cl_t^{\succeq})$,找出所有零向量的列所对应的个体元素就得到了 $\underline{C}(Cl_t^{\succeq})$ 。为了求取类合集的上近似,同样首先要找出所有出现在 Cl_t^{\preceq} 中的元素及其所对应的列向量,将所有的列向量相加得到一个向量,该向量中所有不为零的元素

所对应的个体所组成的集合即为 $\overline{C}(CI_i^s)$ 。求得 CI_i^s 的上下近似后很容易得到其边界 $Bn(CI_i^s) = \overline{C}(CI_i^s) - \underline{C}(CI_i^s)$ 。

类似的也可以得到 CI_i^s 的上下近似。根据以上步骤可以求得所有类合集的上下近似,其计算复杂度为 $2n^2 + 2k$ 。其中 n 为决策表中个体元素的个数, k 为决策类的个数。与传统的方法相比,计算复杂度是原来的 $(2n^2 + 2k) / 2kn^2$, 即计算复杂度降低了将近 k 倍。

4 实例分析

下面用文献[5]中的实例来对基于支配矩阵的类合集近似方法进行验证分析,如表 1,它给出了用以评价学生的 3 个标准,要将学生分为两个类:“good”和“bad”。

表 1 决策信息表

student	A1	A2	A3	A4
S1	good	good	bad	good
S2	medium	bad	bad	bad
S3	medium	bad	bad	good
S4	bad	bad	bad	bad
S5	medium	good	good	bad
S6	good	bad	good	good

注: A1, level in Mathematics; A2, level in Physics; A3, level in Literature; A4, global evaluation (decision class)

这里认为“good”优于“medium”,“medium”优于“bad”,从表中知道在这里有两个类:类“good” $CI_1^s = \{S1, S3, S6\}$, 类“bad” $CI_2^s = \{S2, S4, S5\}$, 由定义得到 $CI_1^s = CI_1^s$, $CI_2^s = CI_2^s$, 因为 CI_1^s 、 CI_2^s 都等于全域 U , 考察它没有意义,所以只需考虑 CI_1^s 及 CI_2^s 。

利用原始的定义可以求得

$$\begin{aligned} \underline{C}(CI_1^s) &= \{S4\}, \quad \overline{C}(CI_1^s) = \{S2, S3, S4, S5\}, \\ Bn_c(CI_1^s) &= \{S2, S3, S5\}, \quad \underline{C}(CI_2^s) = \{S1, S6\}, \\ \overline{C}(CI_2^s) &= \{S1, S2, S3, S5, S6\}, \quad Bn_c(CI_2^s) = \{S2, S3, S5\}. \end{aligned}$$

下面用基于支配矩阵的方法进行求解。

首先从将原始信息表转化为如下的上支配矩阵 $M(CI^s)$ 、下支配矩阵 $M(CI^s)$:

$$M(CI^s) = \begin{matrix} & S1 & S2 & S3 & S4 & S5 & S6 \\ \begin{matrix} S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \end{matrix} & \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \\ \begin{matrix} S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

下面先求类合集的下近似:

对类合集 $CI_1^s, CI_2^s = \{S1, S3, S6\}$ 故对支配矩阵 $M(CI^s)$ 进行处理,将 $M(CI^s)$ 第 1、3、6 行置零,得到 $M'(CI^s)$ 。同理对于类合集 CI_2^s , 将 $M(CI^s)$ 第 2、4、5 行置零得到 $M'(CI^s)$ 。

$$M'(CI^s) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

考察 $M'(CI^s)$, 因为第 1、第 6 列为零向量,所以对应的个体 $S1, S6$ 肯定属于类合集 CI_1^s 的下近似, $\underline{C}(CI_1^s) = \{S1, S6\}$ 。同理考察 $M'(CI^s)$, 可得 $\underline{C}(CI_2^s) = \{S4\}$ 。

接下来求类合集的上近似:

对类合集 $CI_1^s, CI_2^s = \{S1, S3, S6\}$, 将 $M(CI^s)$ 的第 1、第 3、第 6 列加起来得到列

$$M'(CI^s) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

第 1、第 2、第 3、第 5、第 6 个元素不为零,其对应的个体 $S1, S2, S3, S5$ 和 $S6$ 即为类合集 CI_1^s 的上近似,即 $\overline{C}(CI_1^s) = \{S1, S2, S3, S5, S6\}$; 同理可得 $\overline{C}(CI_2^s) = \{S2, S3, S4, S5\}$ 。

求得了类合集的上下近似,很容易可得到

$$\begin{aligned} Bn_c(CI_1^s) &= \overline{C}(CI_1^s) - \underline{C}(CI_1^s) = \{S2, S3, S5\} \\ Bn_c(CI_2^s) &= \overline{C}(CI_2^s) - \underline{C}(CI_2^s) = \{S2, S3, S5\} \end{aligned}$$

通过与用原始定义所得结果比较可知,基于支配矩阵的方法所得的结果完全正确,我们也在其它数据上进行了实验,结果亦可行有效。从计算的过程可以看出,基于支配矩阵的类合集近似方法简化了计算步骤,简单快捷,而且大大降低了计算复杂度。

5 结束语

类合集的近似是 DRSA 中重要的概念和方法,同时它也是一项复杂的计算工作,本文针对目前只能用原始定义进行求解的缺点,我们对计算过程进行了分析和探讨,构造了支配矩阵,提出了基于支配矩阵的类合集近似的新方法,不需要逐个地比较个体之间的序关系,只需将原始信息表转化为支配矩阵,对其进行简单的操作即可以得到类合集的所有近似,不仅将计算复杂度大约降低了 k 倍(其中 k 为决策类的个数),而且简单直观。实例分析也表明了该方法的可行性和有效性。

参考文献

- 1 Pawlak Z. Rough sets[J]. International Journal of Information & Computer Science, 1982, 11(5): 341-356.
- 2 Greco S, Matarazzo B, Slowinski R. Rough Approximation of a Preference Relation By Dominance Relations[J]. European Journal of Operational Research, 1999, 117(2): 62-83.
- 3 Polkowski L, Skowron A. A New Rough Set Approach to Multicriteria and Multiattribute Classification[C]. Proc. of the 1st Rough Sets and Current Trends in Computing International Conference, Warsaw, Poland, 1998: 60-78.

(下转第 114 页)

每个 Archive 在目录服务中对应的是一条记录。目录服务支持对记录的添加、更新、删除等操作。当有一个新的 Archive 向目录服务注册时,目录服务就相应地添加一条记录;当某个 Archive 的状态发生改变时,目录服务就更新相应的记录。目录服务中的每条记录都服从软状态生命周期管理。也就是每条记录都有一个生命期,如果在整个生命期之内都没有接到 Archive 的“keep-alive”信号,那么这条记录就被视为过期。目录服务会自动清除过期的记录。这种机制使得失效的 Archive 不会出现在目录服务的记录中。

2.4 目录服务的分层管理与备份机制

网格中资源的数量众多,这是网格的一个特征。在这种情况下,单一的目录服务可能会因为所存储的数据量过于庞大而导致查询效率低下。另外从可靠性方面考虑,单一目录服务也会成为资源信息服务系统的性能瓶颈。因此有必要采取分层次的目录服务管理,这样一方面可以避免因单个目录服务器崩溃而导致整个网格资源信息服务系统失灵的情况,另一方面也更加符合实际应用中资源在一个组织机构中的组织方式。我们根据虚拟组织的概念来划分目录服务的管理范围。虚拟组织是为了完成某一个确切的业务而动态形成的资源的集合体。每个完成特定任务的虚拟组织中配置一个目录服务(可能还有一个备份目录服务)。多个虚拟组织又可以聚合成一个更大的虚拟组织以完成更复杂的任务,同时它们中的目录服务也聚合成更高层的目录服务。这个更大的虚拟组织可以称为“高层虚拟组织”,所拥有的目录服务称为“高层目录服务”,而组成它的小的虚拟组织可以称为“低层虚拟组织”,相应的目录服务称为“低层目录服务”。这就是目录服务的分层管理,可以用图 2 做一个简单的说明,图中,每个虚线方框部分划分为一个虚拟组织。

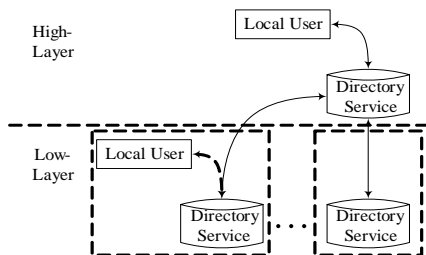


图 2 目录服务的分层管理

如图 2 所示,低层目录服务向高层目录服务注册,每个目录服务可以有自己本地用户,本地用户可以查询本地目录服务下所有的节点的性能信息。

目录服务是网格资源信息服务的核心组件,一方面使用者只有通过它才能发现所要的资源信息,另一方面资源也只有向它注册才能被使用者所发现。一旦目录服务器瘫痪或不能访问,那么使用者将无法发现和定位任何想要的性能数据。为了解决这个性能瓶颈问题,一方面可以采用分层次的目录服务管理,这样即使某个目录服务器出现了故障,其它目录

服务器也可以照常工作;另一方面,要建立备份机制。就是定期(或当目录服务的内容发生改变时)对目录服务的内容进行冗余备份。当主目录服务器发生故障时,备份目录服务器将接替主目录服务器的工作,所有用户的请求也将被重定向到备份目录服务器上。如果这时有新加入的 Archive(或低层目录服务),则将它注册到备份目录服务器上。待主目录服务器恢复正常后,主、备份目录服务器同步注册信息,然后重定向使用者的请求到主目录服务器上去。这样,只有当两台目录服务器同时出现故障时才会影响整个监控系统的正常运行。这样的概率是非常小的,因此采用这种备份方式能够大幅度提高整个网格监控系统的可靠性。

综上,本系统可以由下到上层次化为资源节点层、信息归档层、目录服务层,其中目录服务层又可以根据不同的虚拟组织规模细分为多级层次。至此,我们实现了一个基于 GMA 的层次化资源信息服务系统模型。网格资源信息服务系统采用这种层次化的管理模式,使得网格资源信息服务系统层次分明、便于管理,符合实际应用背景下计算节点的组织形式,且具有较强的可扩展性。

我们已应用上述的系统模型和工具,在面向国家地质调查(NGG)的应用中实现了一个原型应用系统——国家地质网格资源信息服务系统(NGGRIS)(<http://www.ngg.cgs.gov.cn/>)。其资源信息服务的核心即是层次化网格资源信息服务。

3 结论

在网格应用当中,为了聚合大量动态、自治和异构的资源,网格资源信息服务具有很关键的作用。本文参考了国内外一些网格资源信息服务的实现方案,结合 NGG 网格应用的实际情况,基于 GMA 体系结构开发了一个层次化网格资源信息服务的原型系统。实践证明,该系统运行情况稳定,信息采集准确,易于管理,查询方便,可扩展性好,能适应整个 NGG 网格应用的需求。今后将增加更多的实验节点,进一步完善目录服务的分层次管理和备份的机制。

参考文献

- 1 Zhang Xuehai, Freschl J L, Schopf J M. A Performance Study of Monitoring and Information Services for Distributed Systems[C]. Proceedings of HPDC-12, 2003.
- 2 Tierney B, Aydt R, Gunter D, et al. A Grid Monitoring Architecture[Z]. <http://www.didc.lbl.gov/GGF-PERF/GMA-WG/papers/GWD-GP-16-1.pdf>.
- 3 Ribler R L, Vetter J S, Simitci H, et al. Autopilot: Adaptive Control of Distributed Applications[C]. Proceedings of the 7th IEEE Symposium on High Performance Distributed Computing, Chicago, 1998-07.
- 4 查礼,徐志伟,林国璋等.基于 LDAP 的网格监控系统[J].计算机研究与发展,2002,40(8):930-936.
- 5 褚瑞,肖依,卢锡城.一种基于 GMA 的开放式网格资源信息服务.计算机研究与发展,2004,42(12):2114-2122.

(上接第 102 页)

- 4 Greco S, Matarazzo B, Slowinski R. Rough Sets Methodology for Sorting Problems in Presence of Multiple Attributes and Criteria[J]. European Journal of Operational Research, 2002, 138(2): 247-259.

- 5 Greco S, Matarazzo B, Slowinski R. Rough Sets Theory for Multicriteria Decision Analysis[J]. European Journal of Operational Research, 2001, 129(1): 1-47.