

一种基于属性值粗化的决策表正域约简算法

景运革^{1,2}

(1 山西运城学院 公共计算机教学部, 山西 运城 044000;

2 西南交通大学 信息科学与技术学院, 四川 成都 610031)

摘 要: 目前粗糙集模型属性约简大多是基于静态信息系统, 而实际决策表的数据信息都是动态变化的, 为了有效地对这些数据集进行处理, 在信息系统属性值粗化时, 首先介绍了关系矩阵增量机制, 提出了一种粗糙集增量式约简的矩阵方法并构造其相应的算法, 最后, 通过实例分析来说明属性约简的具体操作方法和算法的有效性与可行性。

关键词: 粗糙集; 增量式更新; 正域约简; 关系矩阵; 属性值粗化

中图分类号: TP 311

文献标识码: A

文章编号: 1000-7180(2015)02-0047-05

A Reduction Algorithm of Positive Domain for Decision Table Based on Values Coarsening

JING Yun-Ge^{1,2}

(1 Department of Public Computer Teaching, Yuncheng College, Yun cheng 044000, China;

2 School of Information Science and Technology, South Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Most methods for attribute reduction in Rough Set model is based on static information system nowadays. However, data in real decision table changing dynamically. For dealing with such data effectively and efficiently, we first introduce incremental mechanisms for relation matrix and then develop an incremental algorithm for attribute reduction based on modified equivalence relation matrix, by which minimal attribute reduction of new decision table can be obtained quickly when attribute values are Coarsened. Finally, Example confirms the feasibility and effectiveness of proposed operation and method of attribute reduction.

Key words: rough set; incremental update; positive domain reduction; relation matrix; values coarsening

1 引言

随着计算机网络技术的发展, 各种各样的信息系统积累了大量的数据, 现有的属性约简方法大多是针对静态的信息系统, 而信息系统中的属性值经常随时间的变化而变化, 当属性值发生粗化时, 必然会引起信息系统的约简发生变化. 如何在属性值粗化的条件下快速高效地找到决策表的属性约简, 已成为基于粗糙集的动态知识发现领域中的一个研究热点. 当面临数据不断变化的动态环境时, 用传统的知识获取方法对复杂数据进行挖掘, 需要耗费大量的时间, 很难及时、高效地从动态数据中提取出有用

的信息及知识. 增量式更新方法通过更新部分变化的数据, 不必对全部知识进行重新操作, 能提高操作效率, 科学工作者已运用该方法在动态知识挖掘领域做了大量的研究工作. Liang 等分析了一些对象增加到决策表情况下三种信息熵的变化机制, 提出了一种利用粗糙集技术的增量式特征选择的方法^[1]; 杨明提出了基于改进差别矩阵的属性约简增量式算法, 当一些对象增加到决策表时, 用动态的方法求属性的核并更新差别矩阵, 充分利用了更新前决策表的属性约简, 能快速找出新的约简^[2]; 刘洋等在分析了对象集增量与差别矩阵关系的基础上, 提出一种增量式属性约简的完备算法^[3]; 刘薇等提出

收稿日期: 2014-05-13; 修回日期: 2014-06-25

基金项目: 国家自然科学基金联合项目(U1230117; 61175047)

了一种基于条件熵的快速增量约简方法^[4]; Chan 针对决策表中单个属性增加和删除情况, 提出了粗糙集近似集的增量式更新和快速提取规则的算法^[5]; Wang 等分析了一些属性值变化的情况下三种信息熵的变化机制, 提出了一种利用属性值动态变化的属性约简方法^[6]; 王磊等提出了信息系统的属性值发生变化时变精度粗糙集模型中概念上、下近似集增量式更新的矩阵方法并构造出相应的算法^[7]. 在不完备有序决策系统中, Chen 等讨论了属性值变化时优势特性关系粗糙集模型下近似集更新的原理, 设计了向上向下的近似集增量式高效算法^[8], 季晓岚等在优势关系粗糙集模型中, 给出数据库中属性值粗化细化的概念, 提出了属性值发生粗化细化时近似集的增量更新算法^[9].

李抒音等定义一种新的带权差别集合模型, 提出了一种增量式属性约简方法, 当有新的数据对象被加入决策表, 可有效提高属性约简的效率^[10], 但是当属性值发生粗化, 通过增量式更新等价关系矩阵, 并通过等价关系矩阵相乘的约简方法却鲜有报道.

矩阵理论是处理大量复杂数据的有效工具, 已被广泛应用于科学研究和工程应用等领域. 用矩阵方法对粗糙集属性约简进行研究具有一定的理论意义. 本文在信息系统的属性值发生粗化时, 运用矩阵作为表达和计算工具, 利用既有的约简知识, 通过增量式学习方法, 能够在较短的时间内得到一个属性约简, 在理论分析的基础上给出了实例, 说明所给出的算法是可行和高效的.

2 粗糙集基础知识

2.1 基本概念

定义 1 信息系统是一个四元组 $S = (U, A = C \cup D, V, f)$, $U = \{x_1, x_2, \dots, x_n\}$ 为论域; C, D 分别表示条件属性集和决策属性集, $V = \bigcup_{a \in C \cup D} V_a$, 其中 V_a 是属性 a 的值域, $f: U \times C \cup D \rightarrow V$ 是一个信息函数, 即 $a \in C \cup D, x \in U$ 有 $f(x, a) \in V_a$ ^[11].

定义 2 设知识库 $K = (U, R), P, Q \subseteq R$, $POS_P^Q = \bigcup R$, 知识 Q 依赖于 P 定义为: $\gamma_P(Q) = \frac{|POS_P^Q|}{|U|}$.

定义 3 矩阵乘法: 如果 $A = (a_{ij})_{m \times n}, B = (b_{ij})_{n \times p}$, 则 $A \cdot B = (c_{ij})_{m \times p}$, 其中 $(c_{ij})_{m \times p} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$.

定义 4 信息系统 $S = (U, A = C \cup D, V, f)$, T_C 是论域 U 上的等价关系, 则其等价关系矩阵 $M_{n \times n}^{T_C} = (m_{ij})_{n \times n}$ 的元素为^[12]

$$(m_{ij})_{n \times n} = \begin{cases} 1, & (u_i, u_j) \in T_C \\ 0, & (u_i, u_j) \notin T_C \end{cases}$$

定义 5 信息系统 $S = (U, A = C \cup D, V, f)$, 决策属性 D 的等价类 $U/D = \{D_1, D_2, \dots, D_m\}$, D_i 是 U/D 的任意子集. 则决策属性的矩阵 $M_{n \times m}^{D \times m}$ 表示为

$$M_{n \times m}^{D \times m} = (m_{ij})_{n \times m} = \begin{cases} 1, & u_i \in D_j \\ 0, & u_i \notin D_j \end{cases}$$

2.2 非增量式矩阵约简方法

针对静态信息系统, 文献^[13]中给出了利用决策条件属性的等价关系矩阵和诱导矩阵来计算决策属性的正域, 并提出了矩阵的属性约简算法 1.

决策表 S 中的等价关系为 T_C , 其等价关系矩阵 $M_{n \times n}^{T_C} = (m_{ij})_{n \times n}$, 矩阵 $M_{n \times n}^{T_C}$ 的诱导矩阵 $\Lambda_{n \times n}^{T_C}$ 表示为:

$$\Lambda_{n \times n}^{T_C} \triangleq \text{diag}(1/\sum_{k=1}^n m_{1k}, 1/\sum_{k=1}^n m_{2k}, \dots, 1/\sum_{k=1}^n m_{nk})$$

定理 1 信息系统 $S = (U, A = C \cup D, V, f)$, 令 U/C 和 U/D 为 U 中的两个划分, 其中 $U/D = \{D_1, D_2, \dots, D_m\}$, D_i 是 U/D 的任意子集, 决策属性 D 关于条件属性 C 的正域矩阵表示: $POS_C(D) = (\Lambda_{n \times n}^{T_C} \cdot (M_{n \times n}^{T_C} \cdot M_{n \times m}^D))_1 = (\Lambda_{n \times n}^{T_C} \cdot C_{n \times m})_1$ 且 $(\Lambda_{n \times n}^{T_C} \cdot (M_{n \times n}^{T_C} \cdot M_{n \times m}^D))_1$ 是矩阵 $\Lambda_{n \times n}^{T_C} \cdot (M_{n \times n}^{T_C} \cdot M_{n \times m}^D)$ 的 1-截矩阵.

定理 2 设决策信息系统 $S = (U, A = C \cup D, V, f)$, $C = \{a_1, a_2, \dots, a_n\}$ 是条件属性集, $B \subseteq C$ 且 $B \neq \emptyset$, $a_i \in C - B$, 属性 B, a_i 的等价关系矩阵分别为 $M_{n \times n}^{T_B} = (m_{ij})_{n \times n}, M_{n \times n}^{T_{a_i}} = (m_{ij}')_{n \times n}$, 则把属性 $\{a_i\}$ 添加到属性集 B 后, B 的等价关系矩阵更新为 $M_{n \times n}^{T_{B \cup \{a_i\}}}$ 的元素为: $(m_{ij})^+ = (\min(m_{ij}, m_{ij}'))$.

定理 3 设信息系统 $S = (U, A = C \cup D, V, f)$, $B \subseteq A, B \neq \emptyset, a_i \in A - B, \Lambda_{n \times n}^{T_B} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) (\lambda_i = \sum_{j=1}^n m_{ij})$ ^[12].

设

$$\Lambda_{n \times n}^{T_{B \cup a_i}} = \text{diag}(\lambda_1^+, \lambda_2^+, \dots, \lambda_n^+) (\lambda_i^+ = \sum_{j=1}^n m_{ij}^+),$$

则有

$$\lambda_i^+ = \lambda_i - \sum_{j=1}^n (1 - \omega_{i,j}) \cdot m_{ij}.$$

2.3 决策信息系统属性约简算法的描述

由定理 1 ~ 3 设计了矩阵的属性约简算

法 1^[13].

输入: 决策信息系统 $S = (U, A = C \cup D, V, f)$.

输出: 决策表的属性约简 RED.

Setp1: 计算 D 关于 C 的正域 POS_C^D ;

Step2: 计算 D 关于 C 的依赖度并降序排列, 假设顺序为 $\gamma_{C_1}(D) > \gamma_{C_2}(D) > \dots > \gamma_{C_m}(D)$;

Setp3: RED = $\varnothing$;

Setp4: for $i = 1:m$

4.1 把属性 $\{C_i\}$ 添加到 RED, 利用定理 2 计算属性 RED $\cup \{C_i\}$ 的等价关系矩阵 $M_{RED \cup \{C_i\}}^T$;

4.2 把属性 $\{C_i\}$ 添加到 RED, 利用定理 3 计算 RED $\cup \{C_i\}$ 的对角矩阵为

$$\Lambda_{n \times n}^{RED} = \text{diag}(\lambda_1^+, \lambda_2^+, \dots, \lambda_n^+) (\lambda_i^+ = \sum_{j=1}^m m_{ij})$$

4.3 把属性 $\{C_i\}$ 添加到 RED, 利用定理 3 计算 D 关于 RED $\cup \{C_i\}$ 的正域 $POS_{RED}(D)$

if $POS_C(D) = POS_{RED}(D)$

Then

RED = RED $\cup C_i$

End

For ($a' \in \text{RED}$)

If ($POS_{RED-(a')}(D) = POS_{RED}(D)$) then

RED = RED - a'

End if

End

Setp5: 输出决策表的约简 RED.

3 属性值粗化的矩阵约简方法

3.1 属性值粗化的概念

定义 6 信息系统^[7] $S = (U, A = C \cup D, V, f), B \subseteq A$ 且 $B \neq \varnothing, a_l \in B$, 设 $f(x_i, a_l) \neq f(x_k, a_l)$, $[x_i]_{a_l} = \{x_i \in U \mid f(x_i, a_l) = f(x_j, a_l)\}$, $[x_k]_{a_l} = \{x_k \in U \mid f(x_k, a_l) = f(x_l, a_l)\}$, 令 $f(x_i, a_l) = f(x_k, a_l), \forall x_i \in [x_i]_{a_l}$, 则称属性值 $f(x_i, a_l)$ 粗化为 $f(x_k, a_l)$.

设 $X = [x_i]_{a_l}, Y = [x_k]_{a_l}, X$ 是元素 x_i 等价类, Y 是元素 x_k 的等价类. $I_X = \{i \mid u_i \in X, X \subseteq U\}, I_Y = \{i \mid u_i \in Y, Y \subseteq U\}$, 则 I_X 是集合 X 中的元素下标组成的集合, I_Y 是集合 Y 中的元素下标组成的集合.

3.2 属性值粗化更新机制

假设有一个动态决策表, 当属性值随着时间而发生变化, 重新计算属性等价关系需要花费大量的时间, 为了克服这个缺陷, 增量式学习方法在粗化前

的等价关系矩阵基础上, 通过快速更新原来的等价关系矩阵而得到粗化后的等价关系矩阵, 当决策表属性值发生粗化时, 定理 4 ~ 5 的机制介绍了等价关系矩阵更新机制, 然后利用定理 1 得到新的正域.

定理 4 信息系统 $S = (U, A = C \cup D, V, f), \{a_i\} \subseteq B$, 属性 a_i 的等价关系矩阵为 $M_{n \times n}^{T_B}$, 属性集 $B - \{a_i\}$ 的等价关系矩阵为 $M_{n \times n}^{T_{B-\{a_i\}}}$, 当属性 a_i 的值粗化后, 其对应的等价关系矩阵 $M_{n \times n}^{T_B \downarrow}$ 中元素表示为^[7]:

$$(m_{ij}^\downarrow)_{n \times n} = \begin{cases} m_{B-a_i}(i, j) & i \in I_X \text{ 且 } j \in I_Y \\ m_{B-a_i}(i, j) & i \in I_Y \text{ 且 } j \in I_X \leq i, j \leq |U| \\ m_B(i, j) & \text{其他} \end{cases}$$

定理 5 设决策信息系统 $S = (U, A = C \cup D, V, f)$, 条件属性 $a_i \in C, T_C$ 是论域 U 上的等价关系, 则其等价关系矩阵为 $M_{n \times n}^{T_C} = (m_{ij})_{n \times n}$, 从删除属性 $\{a_i\}$ 后, $T_{C-\{a_i\}}$ 是论域 U 上的等价关系, 则其对应的等价关系矩阵 $M_{n \times n}^{T_{C-\{a_i\}}}$ 的元素表示为:

$$m_{ij}^\downarrow = \begin{cases} 0, m_{ij} = 0 \wedge (x_i, x_j) \notin M_{C-\{a_i\}} \\ 1, m_{ij} = 1 \vee (x_i, x_j) \in M_{C-\{a_i\}} \end{cases}$$

注释: 当从属性集 C 中删除属性 $\{a_i\}$ 时, 如果 $m_{i,j} = 1$ 时, 更新后的 $m_{i,j}$ 是不变的^[12].

3.3 属性值粗化时约简算法

根据等价关系矩阵增量式更新机制, 当决策表中属性值发生变化(粗化)时, 我们依据定义 1 ~ 5 和定理 1 ~ 4 提出了增量式矩阵约简算法 2, 所提算法是在原有约简的基础上, 通过快速更新等价关系矩阵, 能在较短的时间内找到一个新的属性约简.

粗糙集增量式矩阵约简算法:

输入: 粗化前等价关系 $M_{n \times n}^{T_C}$, 约简的等价关系矩阵 $M_{n \times n}^{T_{RED}}$ 和属性的一个约简 RED, 属性 $\{b\}$ 的值发生粗化.

输出: 输出粗化后的最小属性约简 RED $^\downarrow$.

Setp1: 根据定理 5 计算删除属性 $\{b\}$ 的等价关系矩阵 $M_{n \times n}^{T_{C-(b)}}$;

Setp2: 根据定理 4 计算等价关系矩阵 $M_{n \times n}^{T_C}$ 和 $M_{n \times n}^{T_{RED}}$;

Step3: RED $^\downarrow$ = RED;

Step4: 根据定理 1 计算属性值粗化后的 $POS_C(D)^\downarrow$ 和 $POS_{RED}(D)^\downarrow$, 如果 $POS_C(D)^\downarrow = POS_{RED}(D)^\downarrow$ 转到 Setp6, 否则执行 Setp5;

Setp5: for $i = 1: |C - RED|$

5.1 按照属性依赖度的大小来增加属性 C_i ;

5.2 根据定理 2 计算增加属性后的等价关系矩

阵 $M_{n \times n}^{T_{RED \cup \{C_i\}}}$;

5.3 根据定理 3 计算增加属性更新后的对角矩

阵 $\Lambda_{n \times n}^{T_{RED}^\downarrow} = \text{diag}(\lambda_1^+, \lambda_2^+, \dots, \lambda_n^+) (\lambda_i^+ = \sum_{j=1}^n m_{ij})$;

5.4 根据定理 1 计算增加属性后 D 关于 $RED \cup \{C_i\}$ 的正域 $\text{POS}_{RED}(D)^\downarrow$;

If $\text{POS}_C(D)^\downarrow = \text{POS}_{RED}(D)^\downarrow$

Then

$RED^\downarrow = RED \cup C_i$

End

For $(a' \in RED)$

If $(\text{POS}_{RED-\{a'\}}(D) = \text{POS}_{RED}(D))$

Then

$RED^\downarrow = RED^\downarrow - a'$

End if

End

Setp6: 输出粗化后的最小属性约简 $RED^\downarrow$.

4 实例

决策表如表 1 所示, 论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, 条件属性 $C = \{a, b, c\}$, 决策属性 $D = \{d\}$. 决策表的属性约简为 $\{a, b\}$, 如果属性值粗化表示为 $f(x_i, b) = f(x_k, b), 4 \leq i \leq 6, 1 \leq k \leq 3$, 则属性粗化后的属性约简.

表 1 决策表

U	a	b	c	d
x_1	2	2	0	1
x_2	1	2	0	0
x_3	1	2	0	1
x_4	0	0	0	0
x_5	1	0	1	0
x_6	2	0	1	1

4.1 计算决策属性的列矩阵 $M_{n \times m}^D$

根据表 1 和定义 5: 可得决策属性的列矩

阵 $M_{n \times m}^D$:

$$M_{n \times m}^D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

4.2 增量式更新等价矩阵 $M_{n \times n}^{T_C}$ 和 $M_{n \times n}^{T_{RED}}$

条件 C 的等价关系矩阵 $M_{n \times n}^{T_C}$ 从文献[13] 可得, 即:

$$M_{n \times n}^{T_C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

当属性 $\{b\}$ 被删除后, 根据定理 5 可计算出删除属性后的等价关系矩阵 $M_{n \times n}^{T_{C-\{b\}}}$:

$$M_{n \times n}^{T_{C-\{b\}}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

当属性 $\{b\}$ 的值发生粗化后, 根据定理 4 计算粗化后的等价关系矩阵 $M_{n \times n}^{T_C^\downarrow}$:

$$M_{ij}^\downarrow = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

条件属性 $\{a\}$ 和 RED 的等价关系矩阵 $M_{n \times n}^{T_a}$ 和 $M_{n \times n}^{T_{RED}}$ 可从文献[13] 得, 即:

$$M_{n \times n}^{T_a} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M_{n \times n}^{T_{RED}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

当属性值粗化后, 根据定理 4 计算粗化后的等价关系矩阵 $M_{n \times n}^{T_{RED}^\downarrow}$:

$$m_{ij}^\downarrow = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

4.3 属性值粗化矩阵约简的方法

利用定理3分别计算条件属性 C 和约简 RED 的诱导矩阵 $\Lambda_{n \times n}^{T_C \downarrow}, \Lambda_{n \times n}^{T_{RED} \downarrow}$:

$$\Lambda_{n \times n}^{T_C \downarrow} = \text{diag}(1/1, 1/2, 1/2, 1/1, 1/1, 1/1)$$

$$\Lambda_{n \times n}^{T_{RED} \downarrow} = \text{diag}(1/2, 1/3, 1/3, 1/1, 1/3, 1/2)$$

属性值粗化后,根据定理1分别计算 D 关于条件属性 C 和约简 RED 的正域:

$$\text{POS}_C(D)^\downarrow = (\Lambda_{n \times n}^{T_C \downarrow} \cdot (M_{n \times n}^{T_C \downarrow} \cdot M_{n \times m}^D))_1$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix},$$

则

$$\text{POS}_C(D)^\downarrow = \{x_1, x_2, x_3, x_4, x_5, x_6\}.$$

$$\text{POS}_{RED}(D)^\downarrow = (\Lambda_{n \times n}^{T_{RED} \downarrow} \cdot (M_{n \times n}^{T_{RED} \downarrow} \cdot M_{n \times m}^D))_1$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix},$$

则

$$\text{POS}_{RED}(D)^\downarrow = \{x_1, x_3, x_4\}.$$

综上: $\text{POS}_C(D)^\downarrow \neq \text{POS}_{RED}(D)^\downarrow$.

所以根据属性依赖度大小,增加属性 c 到 RED 中,由等价关系矩阵 $M_{n \times n}^{T_{RED}}$ 和 $M_{n \times n}^{T_c}$ 增量式求增加属性后的等价关系矩阵 $M_{n \times n}^{T_{(RED) \cup \{c\}}}$:

$$M_{n \times n}^{T_{(RED) \cup \{c\}}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

增加属性后,根据定理3计算其诱导矩阵 $\Lambda_{n \times n}^{T_{(RED) \cup \{c\}} \downarrow}$:

$$\Lambda_{n \times n}^{T_{(RED) \cup \{c\}} \downarrow} = \text{diag}(1/1, 1/2, 1/2, 1/1, 1/1, 1/1).$$

$$\text{POS}_{RED}(D)^\downarrow = (\Lambda_{n \times n}^{T_{(RED) \cup \{c\}} \downarrow} \cdot (M_{n \times n}^{T_{(RED) \cup \{c\}} \downarrow} \cdot M_{n \times m}^D))_1$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\text{POS}_{RED}(D)^\downarrow = \{x_1, x_2, x_3, x_4, x_5, x_6\},$$

$$\text{所以: } \text{POS}_C(D)^\downarrow = \text{POS}_{RED}(D)^\downarrow,$$

则

$\{a, b, c\}$ 是表1的一个属性约简.

5 结束语

当决策表中一些属性值发生粗化时,常常会引起属性约简的变化,而重新计算属性约简,将会造成大量重复计算,效率不高.为此给出属性值发生粗化条件下矩阵约简的方法.该算法充分利用了更新前决策表的等价关系矩阵和属性约简等信息,能快速找到新的约简,缩短了属性约简的时间,保证了属性约简的实时性,在理论分析的基础上给出了实例,说明了基于属性值粗化的决策表正域约简算法是可行和有效的,下一步工作将探讨属性值细化的决策表正域约简算法.

参考文献:

- [1] Liang J, Wang F, Dang C, et al. A group incremental approach to feature selection applying rough set technique[J]. IEEE Transactions on Knowledge and Data Engineering, 2012(9): 1-31.
- [2] 杨明. 一种基于改进差别矩阵的属性约简增量式更新算法[J]. 软件学报, 2007, 30(5): 815-822.
- [3] 刘洋, 冯博琴, 周江卫. 基于差别矩阵的增量式属性约简完备算法[J]. 西安交通大学学报, 2007(2): 158-161.
- [4] 刘薇, 梁吉业, 魏巍, 等. 一种基于条件熵的增量式属性约简算法[J]. 计算机科学, 2011(1): 229-231.
- [5] Chan C C. A rough set approach to attribute generalization in data mining [J]. Information Science, 1998, 107(1/4): 169-176.
- [6] Wang F, Liang J, Qian Y. Attribute reduction for dynamic data sets [J]. Applied Soft Computing, 2012 (18): 1-18.
- [7] 王磊, 洪志全, 万旋. 属性值变化时变精度粗糙集模型中近似集动态更新的矩阵方法研究[J]. 计算机应用研究, 2013, 30(7): 2011-2013.
- [8] Chen H, Li T, Ruan D. Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining [J]. Knowledge-Based Systems, 2012(31): 140-161.
- [9] 季晓岚, 李天瑞, 邹维丽, 等. 优势关系下属性值粗化细化时近似集分析[J]. 计算机工程, 2010, 36(12): 33-35.

(下转第55页)

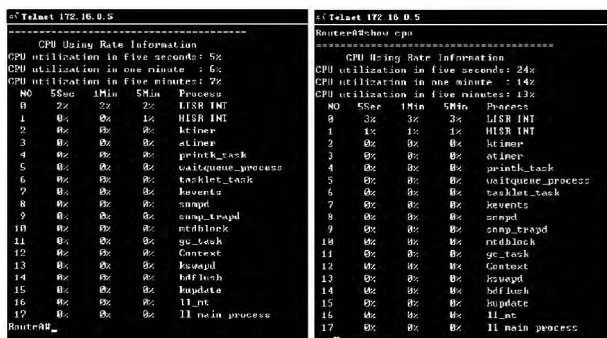


图 6 DD 报文伪造攻击效果图

将 DD 报文中 LSA 的数量固定为 71, Link State ID 字段值采用线性增加, 测试 CPU 占用率与链路状态数据库大小的关系。经过测试发现 CPU 占用率随着链路状态数据库的增大而提高。

通过上面的测试可以发现路由器查找 LSA 的算法对需要查找的 LSA 的数量以及 LSA 的 Link State ID 值不敏感,但是查找的容量大小对算法效率有影响,同时验证了 DD 报文伪造攻击模型对于思科 3900 和锐捷 5750 交换机有效。另外在测试的过程中发现思科和锐捷的路由器在 DD 报文交互流程方面的一点差异。思科路由器要求收到的第一个 DD 报文中必须有一个 LSA 的 Link State ID 值为接收者本身,锐捷的路由器没有这个要求。

6 结束语

OSPF 路由协议由于其自身的复杂性,虽经多次改进(V2/V3),但仍然存在一些脆弱性,单纯依靠协议本身较难以防范各类攻击给网络带来的安全隐患,因此本文建议采用积极的多方式主动防护和检测机制来增加网络安全管理^[10]。

参考文献:

- [1] Hartman S, Zhang D. Analysis of OSPF security according to KARP design guide[S]. IETF; draft-ietf-karp-ospf-analysis-02, 2011.
- [2] JAVVIN. TCP/IP network vulnerability and security [EB/OL]. [2014-05-15]. <http://www.javvin.com/networksecurity/tcpipnetwork.html>.
- [3] Jones E, Moigne O. OSPF security vulnerabilities analysis[S]. IETF; draft-ietf-rpsec-ospf-vuln-02 (work in progress), 2006.
- [4] Barbir A, Murphy S, Yang Y. Threats to routing protocols [S]. IETF; RFC 4593, 2006.
- [5] Moy J. OSPF Version 2[S]. IETF; RFC2328, 1998.
- [6] 陈海燕, 季仲梅, 李鸥, 等. OSPF 路由协议安全性分析及其攻击检测[J]. 微计算机信息, 2005(5): 234-235.
- [7] 蔡朝权. OSPF 路由协议的攻击分析与安全防范[J]. 计算机工程与技术, 2007, 28(23): 5618-5620.
- [8] 高峥, 李林涛. OSPF 协议安全性分析[J]. 黑龙江科技信息, 2011(11): 78.
- [9] Vetter B, Wang Feiyi, Wu S F. An experimental study of insider attacks for OSPF routing protocol[C] // IEEE 1997 International Conference on Network Protocols. Atlanta; Georgia, 1997.
- [10] 孔伟桐, 杨家海. 基于 OSPF 的实时网络监控[J]. 微电子学与计算机, 2006, 23(9): 10-12.

作者简介:

安宁刚 男, (1977-), 硕士, 工程师. 研究方向为计算机网络管理. E-mail: annng@mail. xjtu. edu. cn.

李国栋 男,(1975-),硕士,高级工程师.研究方向为计算机网络安全.

覃尊颖 女,(1984-),硕士,工程师.研究方向为计算机网络管理.

李卫男, (1968-), 博士, 副教授, 博士生导师. 研究方向为计算机网络安全、网络管理.

(上接第 51 页)

- [10] 李抒音, 姚崇怀. 基于增量式更新带权差别集合的属性约简算法[J]. 微电子学与计算机, 2008, 25(11): 68-71.
- [11] 刘清. Rough set 及 Rough 理[M]. 北京: 科学出版社, 2001: 7-16.
- [12] Zhang J, Li T, Ruan D, et al. Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems[J]. International Journal

of Approximate Reasoning, 2012, 53(4):620–635.

- [13] 景运革, 李天瑞. 一种基于关系矩阵的决策表正域约简算法[J]. 计算机科学, 2013, 40(11):261-264.

作者简介:

景运革 男,(1970-),博士研究生,讲师.研究方向粗糙集理论. E-mail:jyg701022@163.com.