第 27 卷 第 5 期
Vol. 27 No. 5

控 制 与 决 策
Control and Decision

2012 年 5 月
May 2012

文章编号: 1001-0920 (2012) 05-0673-08

面向数据删除的核属性更新算法

葛 浩[1a], 李龙澍[2], 杨传健[1b]

(1. 滁州学院 a. 电子信息工程系，b. 计算机系，安徽 滁州 239012；2. 安徽大学 计算机学院，合肥 230039)

摘 要: 针对决策表存在数据删除的情况, 首先提出决策表等价类链表存储结构, 并引入基于该存储结构的简化决策表定义和基于简化决策表核属性定义, 同时证明了该核属性与原始决策表核属性是等价的; 然后, 分别从删除指定对象和删除指定信息两个方面研究核属性更新理论, 并给出相应的算法实现; 最后, 通过实例验证了所提出算法的有效性.

关键词: 粗糙集；正区域；可分辨矩阵；核属性；更新算法

中图分类号: TP181 **文献标识码:** A

Updating algorithms of core attribute based on deleting data

GE Hao[1a], *LI Long-shu*[2], *YANG Chuan-jian*[1b]

(1a. Department of Electronic and Information Engineering，1b. Department of Computer Science，Chuzhou University, Chuzhou 239012，China；2. School of Computer Science，Anhui University，Hefei 230039，China．Correspondent: GE Hao，E-mail：togehao@126.com)

Abstract: For the cases of data deleted, the storage structure of equivalence class list table of decision table is presented, and the definitions of the simplified decision table and core attributes based on the simplified decision table are proposed. It is proved that the core attributes acquired from the definition is equivalence to core attributes of original decision table. Then, the theories and algorithms of updating core attribute are researched and described on two aspects of deleting the object appointed and deleting the objects specified information. Finally, the example is given to verify the effectiveness of the algorithms.

Key words: rough set；positive region；discerniblity matrix；core attribute；updating algorithm

1 引 言

粗糙集理论是波兰数学家Pawlak教授[1]于1982年提出的一种处理含糊和不精确知识的数学工具, 它能有效地分析和处理不精确、不一致以及不完备的信息, 从海量数据中发现隐含的知识. 属性约简是粗糙集理论研究主要内容之一, 备受研究者关注[2-5], 而属性约简常以核属性为初始约简集, 因此求核是该类方法的关键. 现有的核属性求解方法有: 基于可分辨矩阵的求核方法, 基于正区域的求核方法和基于信息熵的求核方法.

由于基于可分辨矩阵的求核方法简单、易实现, 许多学者对此作了大量研究. 文献[2]根据Skowron可分辨矩阵[3]提出一种基于可分辨矩阵的求核方法; [6]举例说明了[2]方法中存在的问题, 对可分辨矩阵进行改进, 并提出新的求核算法; [7]指出[6]虽然给出了修正方法, 但并没发现产生错误的根本原因是由于不一致规则的比较导致的; [8]提出了一种改进的可分辨矩阵及其求核方法, 有效地解决了[2]方法的错误; [9]提出了基于冲突域的核属性求解算法, 该方法不需创建可分辨矩阵, 因此大大节省了存储空间.

上述求核问题是针对静态信息系统, 而现实世界中, 信息系统的对象在不断变化, 这就需要对核属性进行动态更新. 文献[10]给出了在添加对象情况下的增量式核属性求解算法, 该方法需要建立可分辨矩阵, 在处理海量信息系统时, 需要占用很大的空间, 这是制约求核效率的瓶颈. [11]讨论了在删除对象情况下

收稿日期: 2010-11-14；**修回日期:** 2011-05-30.

基金项目: 安徽省自然科学基金项目(090412054)；安徽高校省级自然科学研究项目(KJ2012A212, KJ2011Z276)；安徽省高等学校省级优秀青年人才基金项目(2010SQRL138, 2011SQRL123)；滁州学院科学研究项目(2010KJ014B, 2011KJ003Z).

作者简介: 葛浩(1976−), 男, 副教授, 硕士, 从事数据挖掘和粗糙集的研究；李龙澍(1956−), 男, 教授, 博士生导师, 从事不精确信息处理和智能软件的研究.

核属性更新算法, 由于没有考虑决策表中对象存在取值相同的情况, 在某些情况下无法获得正确的结果; 另外, 数据删除存在删除指定对象和删除指定信息两种情况, 但 [11] 并没有明确地指出处理的是哪种情况; 再有, [11] 的核属性更新算法 (FUAC 算法) 仍需依赖于可分辨矩阵, 具有较高的空间复杂度.

为此, 本文讨论在删除指定对象和指定信息两种情况下的核属性更新方法. 首先, 提出一种决策表等价链表存储结构, 并引入基于该存储结构的简化决策表定义和基于简化决策表核属性定义, 同时证明了该核属性与原始决策表核属性是等价的; 然后, 分别从删除指定对象和指定信息两种情况讨论核属性更新理论和相应的算法实现. 由于不需要依赖于可分辨矩阵, 算法空间和时间效率均有很大提升.

2 等价类链及求核算法

决策表 $S=(U,A,V,f)$. 其中: U 为论域, 是对象的有限集; A 为属性集, $A=C\bigcup D$ 且 $C\bigcap D=\phi$, 这里: C 为条件属性集, D 为决策属性集; $V=\bigcup\limits_{a\in A}V_a$, V_a 是属性 a 的值域; $f:U\times A\to V$ 为信息函数, 对于 $\forall a\in A,x\in U$, 有 $f(x,a)\in V_a$. 不失一般性, 也为了便于操作, 假设 D 仅有一个决策属性值, 其取值范围为从 1 到 n 的整数.

决策表 S 中, 若 $\exists x_i,x_j\in U(i\neq j)$, 有

$$f(x_i,C)=f(x_j,C)\bigwedge f(x_i,D)\neq f(x_j,D),$$

则称 S 为不一致决策表, x_i 与 x_j 称为不一致对象; 否则, 称 S 为一致决策表.

关于粗糙集其他一些概念请参见文献 [1].

定理 1 给定决策表 $S=(U,C\bigcup D,V,f)$, $a\in C$ 为核的充分必要条件为[12]

$$\mathrm{POS}_{C-\{a\}}(D)\neq \mathrm{POS}_C(D).$$

性质 1 决策表 $S=(U,C\bigcup D,V,f)$ 中, 设 $P\subseteq C$, $\mathrm{POS}_P(D)=\{x_i|\forall x_j\in[x_i]_P(i\neq j)$, 有 $f(x_i,P)=f(x_j,P)$ 且 $f(x_i,D)=f(x_j,D)\}$.

证明 设 $U/D=\{Y_1,Y_2,\cdots,Y_m\}$, 则有

$$\mathrm{POS}_P(D)=P\_Y_1\bigcup P\_Y_2\bigcup\cdots\bigcup P\_Y_m.$$

设 $\exists x_i\in U,x_i\in \mathrm{POS}_P(D)$, 则 $\exists Y_k\in U/D(1\leqslant k\leqslant m)$, 有 $x_i\in P\_Y_k$. 由下近似集定义 $P\_Y_k=\{x\in U|[x]_P\subseteq Y_k\}$, 可得 $[x_i]_p\subseteq Y_k$, 则对于 $\forall x_j\in[x_i]_p(i\neq j)$, 有 $f(x_i,P)=f(x_j,P)$ 且 $f(x_i,D)=f(x_j,D)$. □

定义 1 在决策表 $S=(U,C\bigcup D,V,f)$ 中, 设

$$U/C=\{[x_1]_C,[x_2]_C,\cdots,[x_t]_C,[x_{t+1}]_C,\cdots,[x_l]_C\}.$$

其中: 对于 $\forall x_i,x_j\in[x_u]_C$, $i\neq j,1\leqslant u\leqslant t$, 有 $f(x_i,D)=f(x_j,D)$; $\exists x_i,x_j\in[x_v]_C$, $i\neq j,t<v\leqslant l$, 有 $f(x_i,D)\neq f(x_j,D)$; 记 $U_1'=\{x_1,x_2,\cdots,x_t\}$, $U_2'=\{x_{t+1},x_{t+2},\cdots,x_l\}$, $U'=U_1'+U_2'=\{x_1,x_2,\cdots,x_l\}$. 将 U/C 中同一个等价类的元素组成一个单链表, 该链表称为等价类链; 取每个等价类链的第一个元素组成一维数组, 数组元素为 $U'=\{x_1,x_2,\cdots,x_l\}=U_1'+U_2'$; 将 U_1' 中具有相同决策值的元素放在一起, 即 $\{Y_1,Y_2,\cdots,Y_{|U_1'/D|}\}$, 并将 U_2' 记为 Y_ξ, 则 U' 也可以表示为 $U'=\{Y_1,Y_2,\cdots,Y_{|U_1'/D|},Y_\xi\}$. 这种存储结构称为决策表 S 的等价类链表结构, 并记 $S_1'=(U',C\bigcup D,V,f)$ 为简化决策表.

性质 2 对于简化决策表 S_1', 若 $\exists x\in U_1'$, 则 $x\in\mathrm{POS}_C(D)$; 若 $\exists x\in U_2'$, 则 $x\notin\mathrm{POS}_C(D)$.

由性质 1 和定义 1 可以得证.

定义 2 简化决策表 S_1' 的可分辨矩阵为 M, $M=\{m_{ij}|1\leqslant i,j\leqslant|U'|\}$, 其中

$$m_{ij}=\begin{cases}\{a|a\in C\},\\ \quad f(x_i,a)\neq f(x_j,a)\bigwedge f(x_i,D)\neq f(x_j,D)\bigwedge\\ \quad (x_i,x_j\ \text{all}\ \in U_1');\\ \{a|a\in C\},\\ \quad f(x_i,a)\neq f(x_j,a)\bigwedge(x_i,x_j\ \text{only one}\in U_2');\\ \phi,\ \text{otherwise}.\end{cases}$$

定义 3 在简化决策表 S_1' 中, 其核 $\mathrm{Core}'(C)$ 定义为

$$\mathrm{Core}'(C)=\{m_{ij}\mid |m_{ij}|=1\text{ 且 }m_{ij}\in M\}.$$

定理 2 设决策表 S 的核为 $\mathrm{Core}(C)$, 简化决策表 S_1' 的核为 $\mathrm{Core}'(C)$, 则有 $\mathrm{Core}(C)=\mathrm{Core}'(C)$.

证明 首先证明 $\mathrm{Core}(C)\subseteq\mathrm{Core}'(C)$. 只要证明 $\forall a\in\mathrm{Core}(C)$, 便有 $a\in\mathrm{Core}'(C)$.

若 $a\in\mathrm{Core}(C)$, 则 $\exists x_i\in U$, 有 $x_i\in\mathrm{POS}_C(D)$, 但 $x_i\notin\mathrm{POS}_{C-\{a\}}(D)$, $\exists x_j\in U(i\neq j)$, 有

$$f(x_i,C-\{a\})=f(x_j,C-\{a\})\bigwedge f(x_i,a)\neq f(x_j,a)\bigwedge f(x_i,D)\neq f(x_j,D).$$

设 $x_i\in[x_i']_C$, $x_j\in[x_j']_C$, 则有

$$f(x_i',C-\{a\})=f(x_j',C-\{a\})\bigwedge f(x_i',a)\neq f(x_j',a).$$

由于 $x_i\in\mathrm{POS}_C(D)$, 则 $x_i'\in\mathrm{POS}_C(D)$. 对于 x_j 存在两种情况: ① $x_j\in\mathrm{POS}_C(D)$, 则 $x_j'\in\mathrm{POS}_C(D)$ 且 $f(x_i',D)\neq f(x_j',D)$. 由于 $x_i',x_j'\in\mathrm{POS}_C(D)$, 由性质 2 有 $x_i',x_j'\in U_1'$, 再由定义 2 和定义 3 可得 $m_{ij}=\{a\}$, 则 $a\in\mathrm{Core}'(C)$. ② $x_j\notin\mathrm{POS}_C(D)$, 则 $x_j'\notin\mathrm{POS}_C(D)$, 由于 $x_i'\in\mathrm{POS}_C(D)$, 由性质 2 有 $x_i'\in U_1',x_j'\in U_2'$, 再由定义 2 和定义 3 可得 $m_{ij}=\{a\}$, 则 $a\in\mathrm{Core}'(C)$. 由此可得 $a\in\mathrm{Core}'(C)$.

然后证明 $\text{Core}'(C) \subseteq \text{Core}(C)$. 只要证明 $\forall a \in \text{Core}'(C)$, 便有 $a \in \text{Core}(C)$.

若 $a \in \text{Core}'(C)$, 则 $\exists m_{ij} \in M$, 有 $m_{ij} = \{a\}$, 又 $\exists x_i, x_j \in U'$, 有

$$f(x_i, a) \neq f(x_j, a) \bigwedge f(x_i, C - \{a\}) = f(x_j, C - \{a\}).$$

其中: x_i, x_j 至少一个属于 U_1', 不妨设 $x_i \in U_1'$, 由性质 2 有 $x_i \in \text{POS}_C(D)$. 对于 x_j 存在两种情况: ①$x_j \in U_1'$, 由定义 2 可得

$$f(x_i, a) \neq f(x_j, a) \bigwedge f(x_i, C - \{a\}) = f(x_j, C - \{a\}) \bigwedge f(x_i, D) \neq f(x_j, D),$$

则 $x_i \notin \text{POS}_{C-\{a\}}(D)$. ②$x_j \in U_2'$, 设 $\exists x_j' \in [x_j]_C$, 有 $f(x_i, D) \neq f(x_j', D)$, 同理可得 $x_i \notin \text{POS}_{C-\{a\}}(D)$. 由此可得 $\text{POS}_{C-\{a\}}(D) \neq \text{POS}_C(D)$, 则 $a \in \text{Core}(C)$, 因此 $\text{Core}(C) = \text{Core}'(C)$. □

定义 4　给定决策表 S 和相应的简化决策表 S_1', 设 $x \in U'$, 用 $x.\text{count}N$ 表示 $[x]_{C \bigcup D}$ 中元素的个数(与 x 值相同元素的数目); 用 $x.\text{count}C$ 表示 $[x]_C$ 中元素的个数; 用 $x.\text{count}E$ 表示 $[x]_C$ 中等价类 $[x]_{C \bigcup D}$ 的个数.

本文的核属性更新运算, 只需先计算出 U_1' 中对象的 $\text{count}N$ 值和 U_2' 中对象的 $\text{count}E$ 值. 另外, 为了便于进行核属性更新, 对每个核属性添加一个计数器 count, 用于标记 M 上三角部分对应的核属性数目, 即 $\text{Core}(C) = \{(a, \text{count})\}$, 其中 $a \in \text{Core}'(C)$. 在核属性更新过程中, 如果 count 一旦为 0, 该核属性将从核属性集中删除. 下面给出核属性一般求解算法.

算法 1　核属性求解算法.

输入: 决策表 $S = (U, C \bigcup D, V, f)$, C 为条件属性集, D 为决策属性集.

输出: 核属性集 $\text{Core}(C)$.

Step 1: 根据定义 1 对 S 创建等价类链表, 并求得 U_1' 中对象的 $\text{count}N$ 和 U_2' 中对象的 $\text{count}E$;

Step 2: for i=1 to $|U_1'|$ Do

for $j = i + 1$ to $|U'|$ Do

{ len=0;

$\forall a_k \in C$, if $((x_j \in U_1' \bigwedge f(x_i, D) \neq f(x_j, D) \bigwedge f(x_i, a_k) \neq f(x_j, a_k)) \bigvee (x_j \in U_2' \bigwedge f(x_i, a_k) \neq f(x_j, a_k)))$ then

{ len++ ;

$a = a_k$;

if(len>1) break; }

if(len==1)

Core(C)=Core(C) + $(a, 1)$;

}

Step 3: Output Core(C).

算法 1 中, Step 1 采用基于计数的基数排序方法[9]实现, 其时间复杂度为 $O((|C| + |D|)|U|)$, 由于 $|D| = 1$, Step 1 的时间复杂度为 $O(|C||U|)$; Step 2 的时间复杂度为 $O(|C||U'|^2)$. 因此, 算法 1 的时间复杂度为 $O(|C||U'|^2)$. 由于没有存储可分辨矩阵 M, 算法 1 主要空间开销为核属性存储空间和链表指针空间开销, 所以算法 1 空间复杂度为 $O(|C|) + O(|U|)$.

表 1 为一个决策表 S. 其中: $\{a, b, c, d\}$ 为条件属性集 C, D 为决策属性, 共有 12 个样本对象.

表 1　决策表 S

U	a	b	c	d	D
x_1	1	1	0	1	2
x_2	1	0	1	1	1
x_3	0	1	0	0	1
x_4	1	1	1	0	2
x_5	1	0	1	1	2
x_6	0	1	1	1	1
x_7	0	1	0	0	1
x_8	1	0	1	1	3
x_9	1	1	1	1	5
x_{10}	1	1	1	1	5
x_{11}	1	1	1	1	2
x_{12}	1	0	1	1	2

通过分析表 1 可知

$$U/C = \{\{x_3, x_7\}, \{x_6\}, \{x_1\}, \{x_4\}, \{x_2, x_5, x_{12}, x_8\}, \{x_{11}, x_9, x_{10}\}\}.$$

其中: $\{x_2, x_5, x_{12}, x_8\}$ 和 $\{x_{11}, x_9, x_{10}\}$ 为不一致对象的集合, 决策表 S 为不一致决策表. 根据算法 1, 对决策表 S 生成等价类链表, 如图 1 所示.

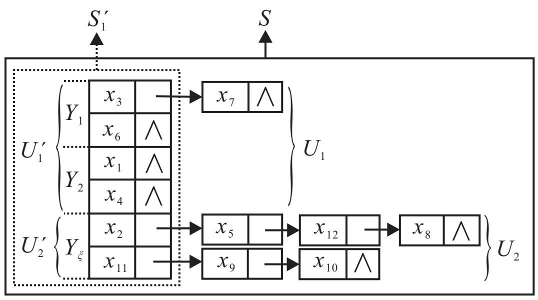

图 1　等价类链表

图 1 中: $U_1 = \{x_3, x_6, x_1, x_4\}$, $U_2 = \{x_2, x_5, x_{12}, x_8, x_{11}, x_9, x_{10}\}$; $Y_1 = \{x_3, x_6\}, Y_2 = \{x_1, x_4\}, Y_\xi = \{x_2, x_{11}\}$, 故有 $U_1' = \{x_3, x_6, x_1, x_4\}$, $U_2' = \{x_2, x_{11}\}$, 则 $U' = \{x_3, x_6, x_1, x_4, x_2, x_{11}\}$; 进而可得 $x_3.\text{count}N = 2$, $x_6.\text{count}N = x_1.\text{count}N = x_4.\text{count}N = 1$, $x_2.\text{count}E = 3$, $x_{11}.\text{count}E = 2$. 执行 Step 2 可得 $\text{Core}(C) = \{(a, 1), (c, 1), (d, 1)\}$.

3 删除指定对象情况下核更新方法

文献[11]在讨论删除情况下的核属性更新问题时, 没有考虑决策表中存在重复对象的情况, 因此在处理存在取值相同对象的决策表时, 其核更新结果不能完全保证是正确的; 另外, 文献[11]并没有严格区分删除的是指定对象还是指定信息(具有指定信息的所有对象). 为此, 本文将分别研究删除指定对象情况下和删除指定信息情况下的核属性更新问题. 本节首先研究前一种情况.

3.1 核属性更新理论

对于决策表 S, 当删除指定对象 x 时, 存在如下 4 种情况:

Case(1): $x \in U_1$ 且 x.count$N \geqslant 2$.

Case(2): $x \in U_1$ 且 x.count$N = 1$.

Case(3): $x \in U_2$ 且 x.count$E > 2$.

Case(4): $x \in U_2$ 且 x.count$E = 2$, 则存在以下 3 种情况:

Case(4.1): x.count$N \geqslant 2$;

Case(4.2): x.count$N = 1$ 且 $[x]_C - x$ 为 U_1' 中不存在的类;

Case(4.3): x.count$N = 1$ 且 $[x]_C - x$ 为 U_1' 中已存在的类. 设为 Y_k;

定理 3 对于 Case(1), Case(3)和 Case(4.1), 核属性集保持不变.

证明 当 x 符合 Case(1), Case(3) 和 Case(4.1) 时, 对象 x 删除后, 根据定义 1 更新等价类链表, 可得对应的简化决策表, 设为 S_2', 可以发现 S_2' 和 S_1' 是相同的. 因此, Case(1), Case(3) 和 Case(4.1) 的核属性集保持不变. □

定理 4 对于 Case(2), 设 x 属于类 Y_k, 核属性集 Core(C) 满足: $\exists y \in U' - Y_k, \exists a \in C$, 有

$$f(x,a) \neq f(y,a) \bigwedge f(x, C-\{a\}) = f(y, C-\{a\}),$$

则 Core(C) =Core(C) $-(a,1)$.

证明 对于 Case(2), 对象 x 删除后, 还需要将 x 与原来 $U' - Y_k$ 中对象产生的核属性剔除. 根据条件, 如果 $\exists y \in U' - Y_k, f(x,a) \neq f(y,a) \bigwedge f(x, C-\{a\}) = f(y, C-\{a\})$, 又由于 $x \in Y_k, y \in U' - Y_k$, 则根据定义 3 可知 $a \in$ Core(C), 这个 a 需要从 Core(C) 中剔除, 即 Core(C) =Core(C) $-(a,1)$. □

定理 5 对于 Case(4.2), 对象 x 删除后, $[x]_C - x$ 中的剩余对象构成一个等价类且为一个新类, 记为 $[z]_C$, 则核属性集 Core(C) 满足: $\exists y \in Y_\xi - x, \exists a \in C$, 有

$$f(x,a) \neq f(y,a) \bigwedge f(x, C-\{a\}) = f(y, C-\{a\}),$$

则 Core(C) =Core(C) $+(a,1)$.

证明 对于这种情况, $[z]_C$ 作为一个等价类从 U_2 转移到 U_1, 则 $z \in U_1'$. 那么新的核属性可能在 z 与 $Y_\xi - x$ 中的对象比较时产生. 根据条件, 如果 $\exists y \in Y_\xi - x$, 则有

$$f(x,a) \neq f(y,a) \bigwedge f(x, C-\{a\}) = f(y, C-\{a\}),$$

由于 $z \in U_1', y \in U_2'$, 由定义 3 可知 $a \in$Core(C). 因此, Core(C) = Core(C) $+(a,1)$. □

定理 6 对于 Case(4.3), 对象 x 删除后, $[x]_C - x$ 中的剩余对象构成一个等价类, 记为 $[z]_C$. 设 z 属于 U_1' 中已存在类 $Y_k (1 \leqslant k \leqslant |U_1'/D|)$, 则核属性集 Core($C$) 满足以下两个条件:

1) 如果 $\exists y \in Y_\xi - x, \exists a \in C$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}),$$

则 Core(C) = Core(C) $+(a,1)$;

2) 如果 $\exists y \in Y_k, \exists a \in C$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}),$$

则 Core(C) = Core(C) $-(a,1)$.

证明 对于这种情况, 当 $[z]_C$ 作为一个等价类从 U_2 转移到 U_1 时, $z \in U_1'$, 存在以下两种情况:

1) z 加入 Y_k 后, 可能会与 $Y_\xi - x$ 中对象产生新的核, 条件 1) 就是针对这种情况的. 根据条件 1), 如果 $\exists y \in Y_\xi - x$, 则有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}).$$

由于 $z \in U_1', y \in U_2'$, 且由定义 3 可知 $a \in$ Core(C). 因此, Core(C) = Core(C) $+(a,1)$.

2) 还需要将 z 在加入 Y_k 之前与 Y_k 中对象产生的核属性剔除, 条件 2) 就是针对这种情况的. 根据条件 2), z 加入 Y_k 之前 $z \in U_2'$, 如果 $\exists y \in Y_k$, 则有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}).$$

由于 $y \in U_1'$, 由定义 3 可知 $a \in$ Core(C). a 需要从 Core(C) 中剔除, 因此 Core(C) = Core(C) $-(a,1)$. □

3.2 核属性更新算法

假设被删除的指定对象存在. 根据上述理论分析, 删除指定对象情况下的核属性更新算法的思路是: 设待删除对象为 x, 先在 S_1' 中, 以及在条件属性 C 下, 寻找与 x 具有相同条件属性值的对象. 若在 U_1' 中找到这样的对象, 假设为 y, 则进入 $[y]_C$ 等价类链, 找到 x 并删除, 同时更新核; 否则, 若在 U_2' 中找到这样的对象, 假设为 y, 则进入 $[y]_C$ 等价类链, 删除 x, 并更新核. 下面给出具体算法描述:

算法 2 删除指定对象情况下的核属性更新算法.

输入: 决策表 $S_1' = (U', C \bigcup D, V, f)$, S_1' 对应的

核属性集 Core$(C)=\{(a,\text{count})\}$, 删除对象 x;

输出: 新的核属性集 Core(C).

Step 1: Core$'(C)=$ Core(C);

Step 2: 遍历 U_1', if($\exists x'\in U_1'$, 有 $f(x',C)==f(x,C)$) then

Step 2.1: { if (x'.count$N\geqslant 2$) then　// Case(1)

{ if (x 即为 x') then {取 $y\in[x']_C$ 且 $y\neq x'$, 将 x' 与 y 位置互换;}　else 将 x' 同时令为 y;

y.count$N=--x'$.count$N$;

在 $[y]_C$ 中找到 x 并 Del x ;

Core$'(C)=$ Core(C); }　// end_Step 2.1

Step 2.2: else　// Case(2)

{ for i=1 to $|U'-Y_k|$ Do　//设 $x'\in Y_k$

$\forall a_j\in C$, if (Condition1) then

Core$'(C)=$Core$'(C)-(a_j,1)$;

Del x ;　　}　//end_Step2.2

} //end_Step2

Step 3: else 遍历 U_2', if ($\exists x'\in U_2'$, 有 $f(x',C)==f(x,C)$) then

Step 3.1: { if (x 即为 x') then {取 $y\in[x']_C$ 且 $y\neq x'$, 将 x' 与 y 位置互换;} else 同时令 x' 为 y;

Step 3.2: if (x'.count$E>2$) then　// Case(3)

Step 3.2.1: { for i=1 to $|[y]_C|$ Do

{if ($f(x_i,D)==f(x,D)$) then num++; 标识 x 的位置; }

Step 3.2.2: if (num $\geqslant 2$) {y.count$E=x'$.count$E$; Del x; Core$'(C)=$Core(C);}

Step 3.2.3: if (num $==$ 1) {y.count$E=--x'$.count$E$; Del x; Core$'(C)=$Core(C);}

goto Step 4 ; } //end_Step3.2

Step 3.3: if (x'.count$E$==2) then

Step 3.3.1:{ for i=1 to $|[y]_C|$ Do

{ if ($f(x_i,D)==f(x,D)$) then num++;

if ($f(x_i,D)==f(y,D)$) then num$y++$;

标识 x 的位置; }

Step 3.3.2: if (num $\geqslant 2$) {y.count$E=x'$.count$E$; Del$x$; Core$'(C)=$Core(C); } // Case(4.1)

Step 3.3.3: if (num==1)

Step 3.3.3.1: { if ($f(y,D)$ 不属于 U_1' 中任何决策值) then　// Case(4.2)

for i=1 to $|Y_\xi-y|$ Do

$\forall a_j\in C$, if (Condition1) then

Core$'(C)=$ Core$'(C)+(a_j,1)$;

Step 3.3.3.2: else if ($f(y,D)$ 属于 U_1' 中某一已存在的类 Y_k) then　// Case(4.3)

{ for i=1 to $|Y_\xi-y|$ Do

$\forall a_j\in C$, if (Condition1) then

Core$'(C)=$ Core$'(C)+(a_j,1)$;

for i=1 to $|Y_k|$Do

$\forall a_j\in C$, if (Condition1) then

Core$'(C)=$ Core$'(C)-(a_j,1)$; }

Step 3.3.3.3: y.count$N=$ num$y$; Del x; 将 $[y]_C$ 并入 U_1, y 并入 U_1'; } //end_Step 3.3.3

} //end_Step 3.3

} //end_Step3

Step 4: Core$(C)=$Core$'(C)$;

Step 5: Ouput Core(C)

其中: Condition 1 为

$$f(y,a_j)\neq$$

$$f(x_i,a_j)\bigwedge f(y,C-\{a_j\})==f(x_i,C-\{a_j\}),$$

$|[y]_C|$ 为等价类 $[y]_C$ 中元素的个数.

对算法 2 按照最坏情况进行分析: Step 2 最坏情况下需遍历完 U_1', 其时间复杂度为 $O(|C||U_1'|)$, 则 Case(1) 的时间复杂度为 $O(|C||U_1|)+O(|[y]_C|)$; Case(2) 的时间复杂度为 $O(|C||U_1'|)+O(|C||U'-Y_k|)$. 当执行到 Step 3 时, 最坏情况下需遍历完 U_2' 的时间开销为 $O(|C||U'|)$, 则 Case(3) 的时间复杂度为 $O(|C||U'|)+O(|[y]_C|)$; Case(4.1) 的时间复杂度为 $O(|C||U'|)+O(|[y]_C|)$; Case(4.2) 的时间复杂度为 $O(|C||U'|)+O(|[y]_C|)+O(|C|(|Y_\xi|+|Y_k|))$; Case(4.3) 的时间复杂度为 $O(|C||U'|)+O(|[y]_C|)+O(|C|(|Y_\xi|+|Y_k|))$. 因此, 算法 2 最坏情况下的时间复杂度为

$$O(|C||U'|)+O(|[y]_C|)+O(|C|(|Y_\xi|+|Y_k|))<$$
$$2O(|C||U'|)+O(|[y]_C|).$$

算法 2 的空间开销为核属性和指针空间开销之和, 故空间复杂度为 $O(|C|)+O(|U|)$.

由于文献 [11] 的核属性更新算法 (FUAC 算法) 是建立在可分辨矩阵的基础上, 其时间和空间复杂度分别为

$$O(|C|(|U_1|+|U_2'|))+O(|U_1|(|U_1|+|U_2'|)),$$
$$O(|C||U_1|(|U_1|+|U_2'|)).$$

文献 [11] 在分析复杂度时, 没有考虑条件属性 C 的基. 如果不创建可分辨矩阵, 采用核属性计数方法, 文献 [11] 核属性更新算法在 $x\in U_1$ 时, 最坏情况下的时间复杂度为 $O(|C|(3|U_1|+|U_2'|))$; 在 $x\in U_2$ 时, 最坏情况下的时间复杂度为 $O(|C|(2|U_1|+2|U_2'|))$. 可以发现, 无论在何种情况下, 算法 2 的时间复杂度均小于

文献 [11], 并且算法 2 无需创建可分辨矩阵, 这样大大降低了空间开销, 有利于大数据集的处理.

3.3 实例分析

为了说明算法 2 的方法, 下面通过 4 个实例, 分别分析 Case(2), Case(3), Case(4.1) 和 Case(4.2).

1) 对于决策表 S, 删除对象 $x_6(0, 1, 1, 1, 1)$.

遍历 U'_1, $\exists x_6 \in U'_1$, 有 $x_6.\mathrm{count}N = 1$ 且 $x_6 \in Y_1$, 则执行 Step 2.2, 将 x_6 逐一与 $U' - Y_1 = \{x_1, x_4, x_2, x_{11}\}$ 中的对象比较, $\exists x_{11} \in U'_2$, 有

$$f(x_6, a) \neq f(x_{11}, a) \bigwedge f(x_6, C - \{a\}) = f(x_{11}, C - \{a\}),$$

则有 $a \in \mathrm{Core}'(C)$, 故

$$\mathrm{Core}'(C) = \mathrm{Core}'(C) - (a, 1) = \{(c, 1), (d, 1)\},$$

并删除 x_6. 最终 $\mathrm{Core}(C) = \{(c, 1), (d, 1)\}$.

2) 对于决策表 S, 删除对象 $x_8(1, 0, 1, 1, 2)$.

遍历 U'_1, 不存在 $x \in U'_1$, 使得 $f(x, C) = f(x_8, C)$; 遍历 U'_2, $\exists x_2 \in U'_2$, 有 $f(x_2, C) = f(x_8, C)$, 由于 $x_2.\mathrm{count}E = 3 > 2$, 执行 Step 3.2.1, 发现和 x_8 同值对象的数目 $\mathrm{num} = 1$, 则执行 Step 3.2.3, 通过计算可得 $x_2.\mathrm{count}E = x_2.\mathrm{count}E - 1 = 2$, 并删除 x_8. 最终可得 $\mathrm{Core}(C)$ 不变, 仍为 $\{(a, 1), (c, 1), (d, 1)\}$.

3) 对于决策表 S, 删除对象 $x_9(1, 1, 1, 1, 5)$.

遍历 U'_1, 不存在 $x \in U'_1$, 使得 $f(x, C) = f(x_9, C)$; 遍历 U'_2, $\exists x_{11} \in U'_2$, 有 $f(x_{11}, C) = f(x_9, C)$ 且 $x_{11}.\mathrm{count}E = 2$. 执行 Step 3.3.1, 对 $[x_{11}]_C$ 等价类链计数, 有 $\mathrm{num} = 2, \mathrm{num}y = 2$; 由于 $\mathrm{num} = 2$, 则执行 Step 3.3.2, 有 $x_{11}.\mathrm{count}E = 2$, 删除 x_9. 最终可得 $\mathrm{Core}(C)$ 不变, 仍为 $\{(a, 1), (c, 1), (d, 1)\}$.

4) 对于决策表 S, 删除对象 $x_{11}(1, 1, 1, 1, 2)$.

遍历 U'_1, 不存在 $x \in U'_1$, 使得 $f(x, C) = f(x_{11}, C)$; 遍历 U'_2, $\exists x_{11} \in U'_2$ 且 $x_{11}.\mathrm{count}E = 2$, 因为被删除对象 x_{11} 为 U'_2 中的对象, 则用 $[x_{11}]_C$ 链表中另一个对象 x_9 与 x_{11} 互换; 再执行 Step 3.3.1, 对 $[x_9]_C$ 等价类链进行统计, 有 $\mathrm{num} = 1, \mathrm{num}y = 2$; 由于 $\mathrm{num} = 1$, 则执行 Step 3.3.3, 又由于 x_9 的决策类不存在于 U'_1 中, 是一个新类, 故执行 Step 3.3.3.1. $\exists x_2 \in U'_2$, 有

$$f(x_2, b) \neq f(x_9, b) \bigwedge f(x_2, C - \{b\}) = f(x_9, C - \{b\}),$$

则有 $b \in \mathrm{Core}'(C)$, 故 $\mathrm{Core}'(C) = \mathrm{Core}'(C) + (b, 1)$. 接着执行 Step 3.3.3.3, 有 $x_9.\mathrm{count}N = \mathrm{num}y = 2$. 删除 x_{11} 后, 将 $[x_9]_C = \{x_9, x_{10}\}$ 并入 U_1, x_9 并入 U'_1. 最终有

$$\mathrm{Core}(C) = \{(a, 1), (b, 1), (c, 1), (d, 1)\}.$$

4 删除指定信息情况下核更新方法

本节将研究删除指定信息情况下的核属性更新问题.

4.1 核属性更新理论

对于决策表 S, 删除指定信息 x, 设 $\exists x' \in U'$, 有 $f(x', C) = f(x, C)$, 存在以下 3 种情况:

Case(1): $x' \in U'_1$.

Case(2): $x' \in U'_2$, 且 $x'.\mathrm{count}E > 2$.

Case(3): $x' \in U'_2$, 且 $x'.\mathrm{count}E = 2$, 则存在以下两种情况:

Case(3.1): $x'.\mathrm{count}N = 1$, 且 $[x']_C - [x']_{C \bigcup D}$ 为 U'_1 中不存在的类;

Case(3.2): $x'.\mathrm{count}N = 1$, 且 $[x']_C - [x']_{C \bigcup D}$ 为 U'_1 中已存在的类, 设为 Y_k.

定理 7 对于 Case(1), 核属性集 $\mathrm{Core}(C)$ 满足: $\exists y \in U' - Y_k, \exists a \in C$, 有

$$f(x', a) \neq f(y, a) \bigwedge f(x', C - \{a\}) = f(y, C - \{a\}),$$

则 $\mathrm{Core}(C) = \mathrm{Core}(C) - (a, 1)$.

证明 对于这种情况, 将与 x 具有相同信息的对象删除后, 同时需要将 x' 对象与原来 $U' - Y_k$ 产生的核属性剔除, 那么需先将这样的核属性找出来. 根据条件, 如果 $\exists y \in U' - Y_k$, 有

$$f(x', a) \neq f(y, a) \bigwedge f(x', C - \{a\}) = f(y, C - \{a\}).$$

由于 $x' \in Y_k, y \in U' - Y_k$, 由定义 3 可知 $a \in \mathrm{Core}(C)$, a 需从 $\mathrm{Core}(C)$ 中剔除, 即

$$\mathrm{Core}(C) = \mathrm{Core}(C) - (a, 1). \quad \square$$

定理 8 对于 Case(2), 核属性集保持不变.

证明 删除指定信息 x 后, 根据定义 1 更新等价类链表, 获得对应的简化决策表, 设为 S'_2, 可以发现 S'_2 和 S'_1 是相同的. 因此, 核属性集保持不变. $\square$

定理 9 对于 Case(3.1), 指定信息 x 删除后, $[x']_C - [x']_{C \bigcup D}$ 中的剩余对象构成一个等价类且为一个新类, 记为 $[z]_C$, 则核属性集 $\mathrm{Core}(C)$ 满足: $\exists y \in Y_\xi - x', \exists a \in C$, 有

$$f(z, a) \neq f(y, a) \bigwedge f(z, C - \{a\}) = f(y, C - \{a\}),$$

则 $\mathrm{Core}(C) = \mathrm{Core}(C) + (a, 1)$.

证明 对于这种情况, $[z]_C$ 作为一个新等价类从 U_2 转入 U_1, 则 $z \in U'_1$, 那么新的核属性可能在 z 与 $Y_\xi - x'$ 中的对象比较时产生. 根据条件, 如果 $\exists y \in Y_\xi - x'$, 有

$$f(z, a) \neq f(y, a) \bigwedge f(z, C - \{a\}) = f(y, C - \{a\}).$$

由于 $z \in U'_1$, $y \in U'_2$, 由定义 3 可知 $a \in \mathrm{Core}(C)$. 因此, $\mathrm{Core}(C) = \mathrm{Core}(C) + (a, 1)$. $\square$

定理 10 对于 Case(3.2), 指定信息 x 删除后, $[x']_C - [x']_{C \bigcup D}$ 中的剩余对象为一个等价类, 记为

$[z]_C$. 设 z 属于 U_1' 中已有类 $Y_k(1 \leqslant k \leqslant |U_1'/D|)$, 则核属性集 Core($C$) 满足以下两个条件:

1) 如果 $\exists y \in Y_\xi - x', \exists a \in C$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}),$$

则 $\mathrm{Core}(C) = \mathrm{Core}(C) + (a,1)$.

2) 如果 $\exists y \in Y_k, \exists a \in C$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}),$$

则 $\mathrm{Core}(C) = \mathrm{Core}(C) - (a,1)$.

证明　对于这种情况, $[z]_C$ 作为一个等价类由 U_2 转入 U_1, 则 $z \in U_1'$, 存在以下两种情况:

1) z 加入 Y_k 后, 可能会与 $Y_\xi - x$ 中的对象产生新的核, 条件 1) 就是针对这种情况的. 由条件 1) 可知, $z \in U_1', y \in U_2'$, 如果 $\exists y \in Y_\xi - x'$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}),$$

又由于 $z \in U_1'$, 则由定义 3 可知 $a \in \mathrm{Core}(C)$. 因此, $\mathrm{Core}(C) = \mathrm{Core}(C) + (a,1)$.

2) 还需要将 z 在加入 Y_k 之前与 Y_k 中对象产生的核属性剔除, 条件 2) 就是针对这种情况的. 根据条件 2), z 加入 Y_k 之前 $z \in U_2'$, 如果 $\exists y \in Y_k$, 有

$$f(z,a) \neq f(y,a) \bigwedge f(z, C-\{a\}) = f(y, C-\{a\}).$$

又由于 $y \in U_1'$, 由定义 3 可知 $a \in \mathrm{Core}(C)$, 需将 a 从 Core(C) 中剔除, 即 $\mathrm{Core}(C) = \mathrm{Core}(C) - (a,1)$.

综合以上两种情况, 定理 10 得证. □

4.2　核属性更新算法

假设被删除的指定信息存在. 根据上述理论分析, 删除指定信息情况下核属性更新算法的思路是: 设待删除指定信息为 x, 先在 S_1' 中, 在条件属性 C 下, 查找与 x 具有相同条件属性值的对象, 若在 U_1' 中找到这样的对象, 假设为 y, 则删除 $[y]_C$ 等价类链, 并更新 Core$'(C)$; 否则, 若在 U_2' 中找到这样的对象, 假设为 y, 则进入 $[y]_C$ 等价类链, 删除与 x 具有相同信息的对象, 并更新核. 具体算法描述如下:

算法 3　删除指定信息情况下核属性更新算法.

输入: 决策表 $S_1' = (U', C \bigcup D, V, f)$, S_1' 对应的核属性集 $\mathrm{Core}(C) = \{(a, \mathrm{count})\}$, 删除指定信息 x;

输出: 新的核属性集 Core(C).

Step 1: $\mathrm{Core}'(C) = \mathrm{Core}(C)$;

Step 2: 遍历 U_1', if ($\exists x' \in U_1'$, 有 $f(x', C) == f(x, C)$) then　//Case(1)

Step 2.1: {将 x' 同时令为 y;

Step 2.2: for $i = 1$ to $|U' - Y_k|$ Do　//设 $x \in Y_k$

$\forall a_j \in C$, if (Condition2) then

$\mathrm{Core}'(C) = \mathrm{Core}'(C) - (a_j, 1)$;

Step 2.3: 从 U_1 中删除 $[y]_C$;　}　//end_step2

Step 3: else 遍历 U_2', if ($\exists x' \in U_2'$, 有 $f(x', C) = f(x, C)$) then

Step3.1：{ if ($f(x', D) == f(x, D)$) then

{取 $y \in [x']_C$ 且 $f(x', D) \neq f(y, D)$, 将 x' 与 y 位置互换;}　else 将 x' 同时令为 y;

Step 3.2: if (x'.count$E > 2$) then　//Case(2)

{ y.count$E = --x'$.count$E$;

for $i = 1$ to $|[y]_C|$ Do

if ($f(x_i, D) == f(x', D)$) then Del x_i ;

goto Step 4;　} //end_Step 3.2

Step 3.3: if (x'.count$E == 2$) then

Step 3.3.1: { if ($f(y, D)$不属于U_1'中任何决策值) then　//Case(3.1)

for i=1 to $|Y_\xi - y|$ Do

$\forall a_j \in C$, if (Condition2) then

$\mathrm{Core}'(C) = \mathrm{Core}'(C) + (a_j, 1)$;

Step 3.3.2: else

if ($f(y, D)$ 属于 U_1' 中某一已存在的类 Y_k) then　//Case(3.2)

{ for i=1 to $|Y_\xi - y|$ Do

$\forall a_j \in C$, if (Condition2) then

$\mathrm{Core}'(C) = \mathrm{Core}'(C) + (a_j, 1)$;

for i=1 to $|Y_k|$ Do

$\forall a_j \in C$, if (Condition2) then

$\mathrm{Core}'(C) = \mathrm{Core}'(C) - (a_j, 1)$;} //end_Step 3.3.2

Step 3.3.3: for i=1 to $|[y]_C|$ Do

if ($f(x_i, D) == f(x', D)$) then Del x_i;

Step 3.3.4: 将 $[y]_C$ 并入 U_1, y 并入 U_1';

} //end_Step 3.3

} //end_Step3

Step 4: $\mathrm{Core}(C) = \mathrm{Core}'(C)$;

Step 5: Ouput Core(C)

其中: Condition 2 为

$$f(x', a_j) \neq f(x_i, a_j) \bigwedge f(x', C - \{a_j\}) == f(x_i, C - \{a_j\}).$$

对算法 3 按照最坏情况进行分析: Step 2 最坏情况下需遍历完 U_1', 其时间复杂度为 $O(|C||U_1'|)$, 则 Case(1) 的时间复杂度为

$$O(|C||U_1'|) + O(|C||U' - Y_k|) + O(|[y]_C|).$$

当执行到 Step 3 时, 最坏情况下需遍历完 U_2', 其时间开销为 $O(|C||U'|)$, 则 Case(2) 的时间复杂度为 $O(|C||U'|) + O(|[y]_C|)$; Case(3.1) 的时间复杂度为 $O(|C||U'|) + O(|[y]_C|)$; Case(3.2) 的时间复杂度为

$O(|C||U'|)+O(|[y]_C|)+O(|C|(|Y_{\xi|}+|Y_k|))$. 因此, 算法3最坏情况下的时间复杂度小于$2O(|C||U'|)+O(|[y]_C|)$. 算法3空间开销为核属性和指针空间开销之和, 故空间复杂度为$O(|C|)+O(|U|)$.

同样, 算法3的时间复杂度也小于文献[11]的核属性更新算法的时间复杂度.

4.3 实例分析

对于算法3, 下面通过3个实例分别分析Case(1), Case(3.1)和Case(3.2).

1) 对于决策表S, 删除指定信息x(0, 1, 0, 0, 1).

遍历U_1', $\exists x_3\in U_1'$, 有$f(x,C)=f(x_3,C)$, 则执行Step 2.2, 将x_3与$U'-Y_1=\{x_1,x_4,x_2,x_9\}$中的对象作比较, 无核产生; 从$U_1$中删除$[x_3]_C$等价类链. 最终Core$(C)$保持不变, 仍为$\{(a,1),(c,1),(d,1)\}$.

2) 对于决策表S, 删除指定信息x(1, 0, 1, 1, 2).

遍历U_1', 不存在$x\in U_1'$, 使得$f(x,C)=f(x',C)$; 遍历U_2', $\exists x_2\in U_2'$, 有$f(x_2,C)=f(x,C)$, 但$f(x_2,D)\neq f(x,D)$; 由于$x_2.\mathrm{count}E=3>2$, 执行Step 3.2, 更新$x_2.\mathrm{count}E=3-1=2$; 遍历$[x_2]_C$, 将与$x$相同信息的对象$x_5,x_{12}$删除. 最终核属性保持不变, 即$\mathrm{Core}(C)=\{(a,1),(c,1),(d,1)\}$.

3) 对于决策表S, 删除指定信息x (1, 1, 1, 1, 5).

遍历U_1', 不存在$x'\in U_1'$, 使得$f(x,C)=f(x',C)$; 遍历U_2', $\exists x_{11}\in U_2'$, 有$f(x_{11},C)=f(x,C)$且$f(x_{11},D)\neq f(x,D)$, 由于$x_{11}.\mathrm{count}E=2$, 执行Step 3.3; 因为x_{11}的决策类为U_1'中已存在的类Y_2, 故执行Step 3.3.2, 将x_{11}与$Y_\xi-x_{11}=\{x_2\}$中的对象作比较, $\exists x_2\in U_2'$, 有

$$f(x_2,b)\neq f(x_{11},b)\bigwedge f(x_2,C-\{b\})=f(x_{11},C-\{b\}),$$

则$b\in\mathrm{Core}'(C)$, 故

$$\mathrm{Core}'(C)=\mathrm{Core}'(C)+(b,1)=\{(a,1),(b,1),(c,1),(d,1)\};$$

再将x_{11}分别与$Y_2=\{x_1,x_4\}$中的对象比较, 有核属性$(c,1)$和$(d,1)$存在, 但这两核属性需要被删除, 即

$$\mathrm{Core}'(C)=\mathrm{Core}'(C)-(c,1)-(d,1)=\{(a,1),(b,1)\};$$

最后执行Step 3.3.3, 删除$[x_9]_{C\bigcup D}=\{x_9,x_{10}\}$, 将$[x_{11}]_C=\{x_{11}\}$并入$U_1$, x_{11}并入U_1'. 最终$\mathrm{Core}(C)=\{(a,1),(b,1)\}$.

5 结　论

本文主要研究对象动态删除情况下核的更新问题. 首先, 提出一种决策表等价链表存储结构, 并引入基于该存储结构的简化决策表定义和基于简化决策表核属性定义, 同时证明了该核属性与原始决策表核属性是等价性的; 然后, 分别从删除指定对象和指定信息两个方面讨论核属性更新理论, 并给出相应的核属性更新算法; 最后, 通过实例验证了本文算法的正确性.

由于本文提出方法是事先在简化决策表中寻找对应信息, 再进一步深入等价类链中搜寻待删除的对象, 这样便避免了在整个决策表中无目的地查找待删除对象, 缩小了寻找范围, 从而提升了算法求解效率. 另外, 对简化决策表按决策属性划块, 核更新时仅执行必要的计算, 避免了对可分辨矩阵进行整行或整列计算, 从而最大限度地降低了核属性更新时的计算量. 此外, 由于采用核属性计数的方法, 无需存储即可分辨矩阵, 降低了空间复杂度, 更有利于大数据集的处理.

参考文献(References)

[1] Pawlak Z. Rough sets[J]. Int J of Computer and Information Science, 1982, 11(5): 341-356.

[2] Hu X H, Cercone N. Learning in relational databases: A rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-337.

[3] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[C]. Intelligent Decision Support-handbook of Applications and Advances of the Rough Sets Theory. Dordrecht: Kluwer Academic Publisher, 1991: 331-362.

[4] Jelonek J, Krawiec K, Slowinski R. Rough set reduction of attributes and their domains for neural networks[J]. Computational Intelligence, 1995, 11(2): 339-347.

[5] 刘少辉, 盛秋戬, 吴斌, 等. Rough集高效算法的研究[J]. 计算机学报, 2003, 26(5): 524-529.
(Liu S H, Sheng Q J, Wu B, et al. Research on efficient algorithms for Rough set methods[J]. Chinese J of Computers, 2003, 26(5): 524-529.)

[6] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086-1088.
(Ye D Y, Chen Z J. A new discernibility matrix and the computation of a core[J]. Acta Electronica Sinica, 2002, 30(7): 1086-1088.)

[7] 王国胤. 决策表核属性的计算方法[J]. 计算机学报, 2003, 26(5): 611-615.
(Wang G Y. Calculation methods for core attributes of decision table[J]. Chinese J of Computers, 2003, 26(5): 611-615.)

[8] 杨明, 孙志挥. 改进的差别矩阵及其求核方法[J]. 复旦学报: 自然科学版, 2004, 43(5): 865-868.
(Yang M, Sun Z H. Improvement of discernibility matrix and the computation of a core[J]. J of Fudan University: Nature Science Edition, 2004, 43(5): 865-868.)

(下转第685页)