文章编号: 1001-0920(2006)07-0787-05

基于描述子的信息系统属性约简及决策规则优化

管延勇[1,2], 史开泉[2], 薛佩军[2]
(1. 济南大学 理学院, 济南 250022; 2. 山东大学 数学与系统科学学院, 济南 250100)

摘 要: 在不完备信息系统中, 利用描述子来表达决策规则. 引入描述子简化和约简的概念, 给出描述子的区分函数, 利用布尔推理理论求取描述子的约简, 并由此获取信息系统的优化决策规则. 从保持原系统决策信息的角度出发, 定义信息系统的 G_S-约简和 D_S-约简, G-约简和 D-约简. 通过对这几种约简及其关系的讨论, 从不同角度考察了信息系统中属性的重要性. 以上结果对于完备信息系统同样有效.

关键词: 信息系统; 决策规则; 属性约简; 粗糙集; 描述子
中图分类号: TP18 **文献标识码**: A

Attributes Reduct and Decision Rule Optimization Based on Descriptors in Information Systems

GUAN Yan-yong[1,2], *SHI Kai-quan*[2], *XUE Pei-jun*[2]
(1. School of Science, Ji'nan University, Ji'nan 250022, China; 2. School of Mathematics and System Science, Shandong University, Ji'nan 250100, China. Correspondent: GUAN Yan-yong, E-mail: jn_gyy@163.com)

Abstract: Descriptors are used to compute optimal decision rules and evaluate significance of attributes in incomplete information systems. The concepts of reduct of descriptor and its discernibility function are introduced, and an approach to compute the reducts of descriptor by using of Boolean reasoning techniques is given. From the reducts of descriptors, all the optimal decision rules contained in information systems can be derived. In order to evaluate significance of attributes, G_S-reduct, D_S-reduct, G-reduct and D-reduct of information systems are proposed. Their properties are discussed, and the relationship between them are revealed. The results are also valid for complete information systems.

Key words: Information system; Decision rule; Attribute reduct; Rough set; Descriptor

1 引 言

Pawlak 教授提出的粗糙集理论[1], 已成功地应用于信息系统的决策规则获取以及属性约简[2]. 从数学表示形式上看, 完备信息系统是含有属性空值的不完备信息系统的特例, 而含有属性空值的不完备信息系统则是含有不完全确定属性值的不完备信息系统的特例. 在完备信息系统中, 每个对象关于各个属性的属性值是已知的, 所以可在论域上定义不可分辨关系, 并可利用等价类定义上下近似以及正域等概念, 进而定义信息系统的相对约简[3]. 这样定义的约简只能使确定性决策规则得到简化, 但这种简化并不一定是最优的.

张文修等[4]讨论了协调信息系统的约简及其区分函数求法, 在此基础上对决策规则进行优化, 但没有用此方法讨论不协调信息系统. Kryszkiewicz[5]定义了对象的相对约简, 并由此求取优化确定性决策规则. 在不完备信息系统中, 某些对象关于某些属性的属性值是未知的或不完全确定的, 所以无法在论域上定义不可分辨关系. 可以采用不同的方法将其转化为完备信息系统来处理, 但这些方法可能引起信息丢失或信息失真(参见文献[6]的总结); 也可通过在论域上定义相容关系, 利用经典粗糙集的扩展

收稿日期: 2005-06-07; **修回日期**: 2005-12-20.
基金项目: 国家自然科学基金项目(70271048).
作者简介: 管延勇(1964—), 男, 山东潍坊人, 副教授, 博士生, 从事粗糙集理论及其应用的研究; 史开泉(1945—), 男, 山东莱州人, 教授, 博士生导师, 从事粗系统与粗分析、系统安全等研究.

模型, 在不改变系统原始数据的基础上, 讨论其属性约简以及决策规则获取问题(如文献[7, 8]).

Leung 等[9]利用属性描述子的概念, 提出了几种不同的决策规则, 并讨论了它们的几种不同的定量度量 本文在此基础上, 讨论含有不完全确定属性值的不完备信息系统的一般决策规则优化以及属性重要性评价问题 不同于 Kryszkiewicz[5,7]在含有属性空值的不完备信息系统中, 定义相容关系以及关于对象的相对约简, 本文利用描述子约简的概念, 通过定义描述子的区分函数, 利用布尔推理理论[10,11]求取描述子的约简, 并由此获取相应的优化决策规则 从保持原系统决策信息的角度出发, 定义信息系统的 G_S-约简和 D_S-约简, G-约简和 D-约简 通过对这几种约简及其关系的讨论, 从不同角度评估信息系统中属性的重要性

2 基本概念

2 1 不完备信息系统

信息系统可视为一个序对 (U, A_T), 其中 U 是研究对象的非空有限集, 称为论域; A_T 是属性的非空有限集合 每个对象关于每个属性有唯一的属性值 如果信息系统中每个对象关于每个属性的属性值都是已知的, 即对于任意 $a \in A_T$, 存在映射 $a: U \to V_a$, $a(x) \in V_a$(V_a 为属性 a 的值域), 则这样的信息系统称为完备信息系统 如果信息系统中某些对象关于某些属性的属性值是未知的或不完全确定的, 则这样的信息系统称为不完备信息系统 在不完备信息系统中, 未知属性值 $a(x)$ 可用 a 的所有可能的取值来表示, 即 $a(x) = V_a$, 或表示为 $a(x) = *$, 称为空值 如果 $a(x)$ 是不完全确定的, 例如已知 $a(x)$ 不是 V_a 中的 b 和 c, 则可记 $a(x) = V_a \setminus \{b, c\}$, 意味着对象 x 关于属性 a 的属性值取 $V_a \setminus \{b, c\}$ 中的某一个. 显然, 未知属性值是不完全确定属性值的特例 本文讨论含有不完全确定属性值的不完备信息系统

一个决策表是一个信息系统 $(U, A_T \cup \{d\})$, 其中 $d \notin A_T$, 称为决策属性, 并且 d 关于每个对象的属性值是已知的; A_T 是条件属性的集合. 如果 (U, A_T) 是完备信息系统, 则称 $(U, A_T \cup \{d\})$ 为完备决策表; 如果 (U, A_T) 是不完备信息系统, 则称 $(U, A_T \cup \{d\})$ 为不完备决策表 表1即为一个不完备决策表

表 1 不完备决策表 $(U, A_T \cup \{d\})$

U	a	b	c	e	d
x_1	{0, 1}	{0}	{0, 1}	{2}	1
x_2	{0, 2}	{1}	{0}	{0}	2
x_3	{2}	{1}	{0}	{1}	0
x_4	{2}	{1}	{0}	{0, 1}	0
x_5	{1}	{0}	{0}	{2}	2
x_6	{1}	{0}	{0, 1}	{2}	1

2 2 描 述 子

文献[9]给出了属性描述子的概念 对于 $A \subseteq A_T$, $a \in A$, $v \in V_a$, 称 (a, v) 为 A - 原子性质; 任何一个 A - 原子性质或几个不同的 A - 原子性质的合取称为 A - 描述子. 设 t 是一个 A - 描述子, 以 $\wedge$ 表示逻辑合取, 则记 $t = \wedge (a, v)$; 所有出现在 t 中的属性的集合记为 $A_T(t)$. 如果 (a, v) 是出现在 t 中的原子性质, 则记为 $(a, v) \in t$ 形式上具有 A - 描述子 t 所表达的属性特征的对象的全体称为 t 的支持集, 记为 $\|t\|$. 即 $\|t\| = \{x \in U: v \in a(x), \forall (a, v) \in t\}$; 记 $d(t) = \{d(x): x \in \|t\|\}$. 以 $\vee$ 表示逻辑析取, 称 $s = \bigvee_{\omega \in W \subseteq V_d} (d, \omega)$ 为决策描述子. 类似地, 记 $d(s) = \{\omega: \omega \in V_d, (d, \omega) \in s\}$. 对于 $\omega \in V_d$, 记 $D_\omega = \|(d, \omega)\| = \{x \in U: d(x) = \omega\}$; $\|s\| = \cup \{D_\omega: \omega \in d(s)\}$. 决策属性 d 确定了论域的划分 $U / \mathrm{IND}\{d\} = \{D_\omega | \omega \in V_d\}$.

注 1 在完备决策表中, $\|t\|$ 为由属性集确定的不可分辨类(等价类); 在含有属性空值的不完备决策表中, $\|t\|$ 为一个最大协调块[8].

定义 1[9] 对于 $A \subseteq A_T$, 记 $\mathrm{DES}(A) = \{t: t$ 是 A - 描述子且 $\|t\| \neq \varnothing\}$, $\mathrm{FDES}(A) = \{t: t \in \mathrm{DES}(A), A_T(t) = A\}$. 对于 $t \in \mathrm{DES}(A)$, 如果 $|d(t)| = 1$(其中 $|d(t)|$ 表示集合 $d(t)$ 的基数), 则称描述子 t 是协调的; 否则, 称为不协调的 记协调 A - 描述子的集合为 $\mathrm{CDES}(A) = \{t: t \in \mathrm{FDES}(A), |d(t)| = 1\}$. 如果 $\forall t \in \mathrm{FDES}(A_T)$, t 都是协调的, 则称信息系统 $(U, A_T \cup \{d\})$ 是协调的; 否则, 称为不协调的

设 t 和 t' 是两个描述子, 如果对于任意 $(a, v) \in t'$ 必有 $(a, v) \in t$, 则称 t' 比 t 粗或称 t 是 t' 的加细, 记为 $t' \succ t$ 或 $t \prec t'$. 显然, 由 $t' \succ t$ 可得 $\|t'\| \supseteq \|t\|$. 对于 $B \subseteq A \subseteq A_T$, $t \in \mathrm{FDES}(A)$, 记 $t|_B = \bigwedge_{(b, v_b) \in t, b \in B} (b, v_b)$, 则 $t|_B \in \mathrm{FDES}(B)$, 称为 t 在 B 上的限制

2 3 一般决策规则与确定性决策规则

$(U, A_T \cup \{d\})$ 中蕴含的知识可表达为决策规则的形式: $t \to s$ 其中: $t = \wedge (a, v)$, $a \in A \subseteq A_T$; $s = \vee (d, \omega)$, $\omega \in W \subseteq V_d$. 分别称 t 和 s 为决策规则的条件部分和决策部分. 本文简化文献[9]中有关决策规则的概念, 给出如下定义:

定义 2 对于 $A \subseteq A_T$, $t \in \mathrm{DES}(A)$, $s = \vee (d, \omega)$, 如果 $d(t) = d(s)$, 则称 $t \to s$ 为一般决策规则 如

果 $|d(t)|=1$, 则称 $t\to s$ 是确定性的; 否则, 称为非确定性的.

显然, 一般决策规则 $t\to s$ 是文献[9]中的 true 决策规则, 记为 $t\Rightarrow s$. 对于任意 $t\in\mathrm{DES}(A)$, 可以诱导一般决策规则, $t\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$.

2.4　上下近似

对于 $A\subseteq A_T$, 可在论域上定义相容关系 R: $xRy\Leftrightarrow\forall a\in A, a(x)\cap a(y)\neq\phi$. 对于 $t\in\mathrm{FDES}(A)$, $\|t\|$ 是一个 R 相容类. 由于决策表数据的局限性, $\|t\|$ 一般不是最大相容类(即文献[8]中的最大协调块). 因此 $\{\|t\|: t\in\mathrm{FDES}(A)\}$ 是论域的覆盖而非完全覆盖.

定义 3　在信息系统 $(U, A_T\cup\{d\})$ 中, 对于 $A\subseteq A_T, X\subseteq U$, 令

$$\underline{A}(X)=\cup\{\|t\|: t\in\mathrm{FDES}(A),\ \|t\|\subseteq X\},$$

$$\overline{A}(X)=\cup\{\|t\|: t\in\mathrm{FDES}(A),\ \|t\|\cap X\neq\phi\},$$

分别称为 X 的 A - 下近似和 A - 上近似. 令 $\mathrm{POS}_d^A(U)=\bigcup_{\omega\in V_d}\underline{A}(D_\omega)$, 称为信息系统的 A - 正域. 易证以下命题:

命题 1　$\mathrm{POS}_d^A(U)\subseteq\mathrm{POS}_d^{A_T}(U)$, $\forall D_\omega\in U/\mathrm{IND}\{d\}$, $\overline{A_T}(D_\omega)\subseteq\overline{A}(D_\omega)$.

3　描述子的约简和优化决策规则的求法

3.1　初等决策规则与优化决策规则

对于一个决策规则 $t\to s$, $|d(s)|$ 越小, 说明此决策规则的确定性程度越好. 特别地, 当 $|d(s)|=1$ 时, $t\to s$ 为确定性决策规则. 对于一个描述子 t', 如果在 t' 中添加新的原子性质将它扩充成 t, 则 $t\prec t'$, $d(t)\subseteq d(t')$, 说明 $t\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$ 比 $t'\Rightarrow\bigvee_{\omega\in d(t')}(d,\omega)$ 的确定性程度高. 显然, $\mathrm{FDES}(A_T)$ 中的描述子 t 导出的决策规则 $t\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$ 的确定性水平最高, 称为由 t 诱导的初等一般决策规则. 进一步, 若有 $|d(t)|=1$, 则称为初等确定性决策规则.

例 1　表 1 所示不完备决策表中的初等一般(确定性)决策规则如下:

$t_1=(a,0)\wedge(b,0)\wedge(c,0)\wedge(e,2)\Rightarrow(d,1)$,

$t_2=(a,0)\wedge(b,0)\wedge(c,1)\wedge(e,2)\Rightarrow(d,1)$,

$t_3=(a,0)\wedge(b,1)\wedge(c,0)\wedge(e,0)\Rightarrow(d,2)$,

$t_4=(a,1)\wedge(b,0)\wedge(c,0)\wedge(e,2)\Rightarrow(d,1)\vee(d,2)$,

$t_5=(a,1)\wedge(b,0)\wedge(c,1)\wedge(e,2)\Rightarrow(d,1)$,

$t_6=(a,2)\wedge(b,1)\wedge(c,0)\wedge(e,0)\Rightarrow(d,0)\vee(d,2)$,

$t_7=(a,2)\wedge(b,1)\wedge(c,0)\wedge(e,1)\Rightarrow(d,0)$.

本文试图在不改变初等决策规则确定性程度的前提下, 简化其条件部分的描述形式.

定义 4　对于 $t\in\mathrm{DES}(A_T)$, 如果有 $t'\in\mathrm{DES}(A_T)$, 满足 $t'\succ t$ 且 $d(t')=d(t)$, 则称 t' 是 t 的一个简化; 进一步, 若有 $\forall t''\succ t'$, $d(t'')\neq d(t)$, 则称 t' 是 t 的一个约简, 此时 $A_T(t')$ 称为 A_T 关于 t 的约简. 记 t 的约简的集合为 $\mathrm{red}(t)$, A_T 关于 t 的约简的集合为 $\mathrm{red}A_T(t)$.

定义 5　如果 $t\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$ 是一般决策规则, t' 是 t 的简化, 则称 $t'\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$ 是简化决策规则. 如果 t' 是 t 的约简, 则称 $t'\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$ 是优化决策规则.

令

$$\mathrm{core}A_T(t)=\bigcap_{t'\in\mathrm{red}(t)}A_T(t'),$$

$$\mathrm{nec}A_T(t)=\bigcup_{t'\in\mathrm{red}(t)}A_T(t'),$$

$$\mathrm{redun}A_T(t)=A_T-\mathrm{nec}A_T(t).$$

称 $\mathrm{core}A_T(t)$ 为 A_T 关于 t 的核, 称 $\mathrm{nec}A_T(t)$ 为 A_T 关于 t 的必要属性集, 称 $\mathrm{redun}A_T(t)$ 为 A_T 关于 t 的冗余集.

3.2　描述子约简的区分函数求法

对于 $t\in\mathrm{FDES}(A_T)$, 记 $\|d(t)\|=\bigcup_{\omega\in d(t)}\|(d,\omega)\|$, $\alpha(x,t)=\{a: a\in A_T, (a,v_a)\in t, v_a\notin a(x)\}$. 可证如下定理:

定理 1　对于 $t\in\mathrm{FDES}(A_T)$, $t'\succ t$, 则有

$$d(t')=d(t)\Leftrightarrow\forall x\in U-\|d(t)\|,\ \alpha(x,t)\cap A_T(t')\neq\phi$$

定义 6　对于 $t\in\mathrm{FDES}(A_T)$, 令 $\Delta_d(t)=\bigwedge_{x\in U-\|d(t)\|}\bigvee\alpha(x,t)$, 称为描述子 t 的区分函数.

根据布尔推理理论[10,11], 由定理 1 可得如下命题:

命题 2　对于 $t\in\mathrm{FDES}(A_T)$, 有:

1) $B\in\mathrm{red}A_T(t)\Leftrightarrow\wedge B$ 是区分函数 $\Delta_d(t)$ 的极小析取范式的一个合取项;

2) $\mathrm{core}A_T(t)=\{c: \exists x\in U-\|d(t)\|, \alpha(x,t)=\{c\}\}$.

3.3　求取优化决策规则的步骤

1) 计算 $\mathrm{FDES}(A_T)$, 得到所有初等一般决策规则 $t\Rightarrow\bigvee_{\omega\in d(t)}(d,\omega)$;

2) 对于每个 $t\in\mathrm{FDES}(A_T)$, 计算 $\|d(t)\|$; 对

于 $\forall x \in U - \| d(t) \|$, 计算 $\alpha(x, t)$;

3) 计算区分函数 $\Delta_d(t)$ 的极小析取范式, 求出 $\mathrm{red}\, A_T(t)$;

4) $\forall B \in \mathrm{red}\, A_T(t)$, 得出由 t 诱导的优化决策规则 $t|_B \Rightarrow \bigvee_{\omega \in d(t)} (d, \omega)$.

例 2 求解表 1 所示不完备信息系统 $(U, A_T \cup \{d\})$ 的优化决策规则

1) $\| t_1 \| = \{x_1\}, U - \| d(t_1) \| = \{x_2, x_3, x_4, x_5\}, \alpha(x_2, t_1) = \{b, e\}, \alpha(x_3, t_1) = \{a, b, e\}, \alpha(x_4, t_1) = \{a, b, e\}, \alpha(x_5, t_1) = \{a\}$; $\Delta_d(t_1) = \bigwedge_{x \in U - \| d(t_1) \|} \vee \alpha(x, t_1) = (a \wedge b) \vee (a \wedge e)$. 由 t_1 诱导的优化确定性决策规则为

$$t_1^1 = (a, 0) \wedge (b, 0) \Rightarrow (d, 1),$$
$$t_1^2 = (a, 0) \wedge (e, 2) \Rightarrow (d, 1).$$

2) 对于 t_4, $\| t_4 \| = \{x_1, x_5, x_6\}, \Delta_d(t_4) = a \vee b \vee e$. 由 t_4 诱导的优化一般决策规则为

$$t_4^1 = (a, 1) \Rightarrow (d, 1) \vee (d, 2),$$
$$t_4^2 = (b, 0) \Rightarrow (d, 1) \vee (d, 2),$$
$$t_4^3 = (e, 2) \Rightarrow (d, 1) \vee (d, 2).$$

3) 同理, 可得出由 t_2, t_3, t_5, t_6, t_7 诱导的优化决策规则(略).

4 信息系统中属性重要性的评估

4 1 不完备信息系统的 G_S- 约简和 D_S- 约简

定义 7 在不完备信息系统 $(U, A_T \cup \{d\})$ 中, 对于 $A \subseteq A_T$, 若 A 是满足对于 $\forall t \in \mathrm{FDES}(A_T)$, $d(t|_A) = d(t)$ 成立的极小条件属性子集, 则称 A 为信息系统的 G_S- 约简, 称所有 G_S- 约简的交集为信息系统的 G_S- 核; 若 A 是满足对于 $\forall t \in \mathrm{CDES}(A_T)$, $d(t|_A) = d(t)$ 成立的极小条件属性子集, 则称 A 为信息系统的 D_S- 约简, 称所有 D_S- 约简的交集为信息系统的 D_S- 核

定义 8 令

$$\Delta^G = \bigwedge_{t \in \mathrm{FDES}(A_T)} \bigwedge_{x \in U - \| d(t) \|} \vee \alpha(x, t),$$

$$\Delta^D = \bigwedge_{t \in \mathrm{CDES}(A_T)} \bigwedge_{x \in U - \| d(t) \|} \vee \alpha(x, t).$$

分别称 Δ^G 和 Δ^D 为信息系统的一般区分函数和确定性区分函数

命题 3 A 为信息系统的 G_S- 约简 $\Leftrightarrow \wedge A$ 是区分函数 Δ^G 的极小析取范式的一个合取项; A 为信息系统的 D_S- 约简 $\Leftrightarrow \wedge A$ 是区分函数 Δ^D 的极小析取范式的一个合取项

注 2 1) 信息系统的 G_S- 约简(D_S- 约简) 是保证使原系统的每个初等一般(确定性) 决策规则都得到简化的极小属性子集, 即它是任一 A_T- 描述子(协调 A_T- 描述子) 的简化 对于某个 A_T- 描述子(协调 A_T- 描述子) 而言, G_S- 约简(D_S- 约简) 仅是简化之一, 但一般不一定是约简 2) 系统的一个 D_S- 约简必是某个 G_S- 约简的子集; D_S- 核是 G_S- 核的子集

例 3 对于表 1 所示不完备决策表, 由例 2 的结果计算可得

$$\Delta^G = \bigwedge_{t \in \mathrm{FDES}(A_T)} \bigwedge_{x \in U - \| d(t) \|} \vee \alpha(x, t) = a \wedge c \wedge e,$$

$$\Delta^D = \bigwedge_{t \in \mathrm{CDES}(A_T)} \bigwedge_{x \in U - \| d(t) \|} \vee \alpha(x, t) = a \wedge c \wedge e.$$

可见不完备决策表的 G_S- 约简为 $\{a, c, e\}$, D_S- 约简亦为 $\{a, c, e\}$.

命题 4 1) 对于不完备信息系统 $(U, A_T \cup \{d\})$, 若 A 为信息系统的 G_S- 约简, 则 A 为使 $\overline{A}(D_\omega) = \overline{A_T}(D_\omega)$ 对于 $\forall \omega \in V_d$ 都成立的极小条件属性子集 若 A 为信息系统的 D_S- 约简, 则 A 为满足 $\mathrm{POS}_d^A(U) = \mathrm{POS}_d^{A_T}(U)$ 的极小条件属性子集; 反之, 一般不成立

2) 对于完备信息系统 $(U, A_T \cup \{d\})$, A 为 G_S- 约简 $\Leftrightarrow A$ 为使 $\overline{A}(D_\omega) = \overline{A_T}(D_\omega)$ 对于 $\forall \omega \in V_d$ 都成立的极小条件属性子集; A 为 D_S- 约简 $\Leftrightarrow A$ 为满足 $\mathrm{POS}_d^A(U) = \mathrm{POS}_d^{A_T}(U)$ 的极小条件属性子集

注 3 在不完备信息系统中, 对于 $t_1, t_2 \in \mathrm{FDES}(A)$, 即使 $t_1 \neq t_2$, 但 $\| t_1 \|$ 与 $\| t_2 \|$ 也可能相交, 从而一般由 $\| t \| \subseteq \overline{A}(X)$ 得不到 $\| t \| \cap X \neq \Phi$ 这正是命题 4 之 1) 的充分性不成立的原因

注 4 由命题 4 可知, 对于完备信息系统, 其 D_S- 约简即为文献[3] 中定义的相对约简 一般而言, 在不完备信息系统中, D_S- 约简的定义比要求正域不变更加严格, 因此利用要求正域不变来定义系统的相对约简没有实际意义

4 2 不完备信息系统的 D- 约简和 G- 约简

定义 9 令

$$\mathrm{dred}(A_T) = \bigcup_{t \in \mathrm{CDES}(A_T)} \bigcup_{B \in \mathrm{red}\, A_T(t)} B,$$

$$\mathrm{dcore}(A_T) = \bigcap_{t \in \mathrm{CDES}(A_T)} \mathrm{core}\, A_T(t).$$

分别称为信息系统 $(U, A_T \cup \{d\})$ 的 D- 约简和 D- 核 令

$$\mathrm{gred}(A_T) = \bigcup_{t \in \mathrm{FDES}(A_T)} \bigcup_{B \in \mathrm{red}\, A_T(t)} B,$$

$$\mathrm{gcore}(A_T) = \bigcap_{t \in \mathrm{FDES}(A_T)} \mathrm{core}\, A_T(t).$$

分别称为信息系统 $(U, A_T \cup \{d\})$ 的 G- 约简和 G-

核

显然, $\mathrm{dred}(A_T)$ ($\mathrm{gred}(A_T)$) 是能保持原系统关于确定性(一般) 决策规则信息量的极小条件属性子集 相对于确定性(一般) 优化决策规则的获取而言, $\mathrm{dcore}(A_T)$ ($\mathrm{gcore}(A_T)$) 中的属性是绝对必要的, $\mathrm{dred}(A_T) - \mathrm{dcore}(A_T)$ ($\mathrm{gred}(A_T) - \mathrm{gcore}(A_T)$) 中的属性是相对必要的, 而 $A_T - \mathrm{dred}(A_T)$ ($A_T - \mathrm{gred}(A_T)$) 中的属性是绝对不必要的, 是冗余的

例 4 　在表 1 所示不完备决策表中, 由例 2 的结果得出

$$\mathrm{dred}(A_T) = \mathrm{gred}(A_T) = A_T,$$
$$\mathrm{dcore}(A_T) = \mathrm{gcore}(A_T) = \phi$$

说明 A_T 中的属性对于一般决策规则以及确定性决策规则的获取都是相对必要的

4 3　信息系统两类约简的关系

信息系统的 G_S- 约简 (D_S- 约简) 是将所有一般(确定性) 决策规则简化的极小属性子集, 而 G- 约简 (D- 约简) 则是基于所有一般(确定性) 优化决策规则的获取

一般而言, 信息系统的 G_S- 约简 (D_S- 约简) 不是唯一的; 而信息系统的 G- 约简 (D- 约简) 则是唯一的, 其中的属性对于从信息系统中获取一般(确定性) 优化决策规则是必要的

由 G_S- 约简 (D_S- 约简) 的极小性知, 其中的每个属性必属于某个条件属性描述子的某个约简, 从而它必属于 G- 约简 (D- 约简). 可见, 每个 G_S- 约简 (D_S- 约简) 是 G- 约简 (D- 约简) 的子集, 但 G- 约简 (D- 约简) 中的属性不一定在某个 G_S- 约简 (D_S- 约简) 中, 如例 1 中的属性 b. 从优化决策规则获取的角度看, G_S- 约简 (D_S- 约简) 可能造成信息丢失 一般而言, G- 核是 G_S- 核的子集, D- 核是 D_S- 核的子集

利用本文方法讨论文献[9] 中表 2 关于小汽车的不完备信息系统, 除了得到[9] 中方法的 4 条优化决策规则外, 还可得到另外两条潜在的优化决策规则, 即

$$(P, \mathrm{low}) \wedge (M, \mathrm{low}) \wedge (X, \mathrm{high}) \Rightarrow (d, \mathrm{excellent}),$$
$$(P, \mathrm{high}) \wedge (M, \mathrm{high}) \wedge (X, \mathrm{low}) \Rightarrow (d, \mathrm{poor}).$$

注 5 　所谓潜在的优化决策规则, 是指在信息表中虽没有确定的对象支持它, 但可能有对象支持它, 且肯定没有任何对象与它相抵触 例如信息表中的数据显示, 只要对象的条件属性描述为 $(P, \mathrm{low}) \wedge (M, \mathrm{low}) \wedge (X, \mathrm{high})$, 则必可得到决策结论 $(d, \mathrm{excellent})$; 对象 x_2 可能支持此规则, 且没有其他任何对象与此规则相抵触 不完备信息系统中所蕴含的这种决策规则也是有意义的

5　结　语

利用描述子的约简及其区分函数来求一般决策规则的优化决策规则, 能够充分挖掘信息系统中的决策信息 通过描述子的简化来定义信息系统的 G_S- 约简和 D_S- 约简, 通过描述子的约简来定义信息系统的 D- 约简和 G- 约简, 可从不同角度对信息系统中各条件属性的重要性作出全面的评估 这些结论在完备信息系统和含有属性空值的不完备信息系统中是成立的, 可视为文献[3] 定义的相对约简、文献[4] 关于协调完备信息系统约简的推广, 以及文献[5] 关于优化决策规则获取方法的改进

参考文献(References)

[1] Pawlak Z Rough Sets [J]. *Int J of Computer and Information Sciences*, 1982, 11: 341-356.

[2] Pawlak Z *Rough Sets: Theoretical Aspects of Reasoning about Data* [M]. London: Kluwer Academic Publishers, 1991.

[3] Pawlak Z Rough Set Approach to Multi-attribute Decision Analysis [J]. *European J of Operational Research*, 1994, 72: 443-459.

[4] Zhang W X, Liang Y, Wu W Z *Information Systems and Knowledge Discovery* [M]. Beijing: Science Press, 2003: 22-47.

[5] Kryszkiewicz M. Rules in Incomplete Information Systems [J]. *Information Sciences*, 1999, 113: 271-292.

[6] Huang B, Zhou X Z Extension of Rough Set Model Based on Connection Degree under Incomplete Information Systems [J]. *J of System Engineering — Theory and Practice*, 2004, 24(1): 88-92.

[7] Kryszkiewicz M. Rough Set Approach to Incomplete Information Systems [J]. *Information Sciences*, 1998, 112: 39-49.

[8] Leung Y, Li D. Maximal Consistent Block Technique for Rule Acquisition in Incomplete Information Systems [J]. *Information Sciences*, 2003, 153: 85-106.

[9] Leung Y, Wu W Z, Zhang W X. Knowledge Acquisition in Incomplete Information Systems: A Rough Set Approach [J]. *European J of Operational Research*, 2004, 168: 164-180.

[10] Skowron A, Rauszer C. The Discernibility Matrices and Functions in Information Systems [A]. *Intelligent Decision Support: Handbook of Applications and Advances of Rough Sets Theory* [C]. Dordrecht: Kluwer Academic Publisher, 1992: 331-362.

[11] Skowron A. Extracting Laws from Decision Tables: A Rough Set Approach [J]. *Computational Intelligence*, 1995, 110: 371-388.