

基于闭算子的属性约简

刘 静 米据生

(河北师范大学数学与信息科学学院 石家庄 050016)

摘 要 对于协调的信息系统,定义了其条件属性集的幂集上的两个闭算子 $C(R)$ 与 $C(r)$,讨论了相应闭集族的性质,并证明了它们与不可辨识属性集族之间的关系。提出属性约简的一种新方法,给出 $C_r = C_R$ 的充要条件,并证明所定义的属性约简与文献[4,7]中约简的等价性。

关键词 属性约简,闭算子,粗糙集,等价关系,信息系统

中图法分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.10.052

Attribute Reduction Based on Closure Operators

LIU Jing MI Ju-sheng

(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050016, China)

Abstract For a consistent information system, the definitions of two closure operators $C(R)$ and $C(r)$ on the power set of conditional attribute set were first defined respectively, and then the properties of two closed set families C_r and C_R were also discussed. The relationships among C_r , C_R and the set of all discernibility attribute sets Ω were examined, from which we provided a simple method to attribute reduction in consistent decision table defined in reference [4]. Meanwhile, the sufficient and necessary condition of $C_r = C_R$ was proved. Finally, we proved that under condition $C_r = C_R$, the proposed method is equivalent to those in references [4] and [7].

Keywords Attribute reduction, Closure operator, Rough set, Equivalence relation, Information system

1 引言

属性约简是知识发现的一个重要研究课题^[1-3]。所谓属性约简,就是在保持知识库分类能力不变的条件下,删除其中不相关或不重要的属性,由剩余属性确定的知识与全体属性确定的知识是相同的。通过约简,可以使信息系统的知识表示更加简洁,使隐藏的知识更加清晰。因此,国内外许多学者对多种信息系统的约简进行了研究,如近似约简、分布约简、分配约简、最大分布约简、 β 约简、相似关系下的约简^[4-6]等等。这些约简都能通过可辨识属性矩阵与布尔表达式求得。这些新的思想为信息系统的知识发现提供了很多有益的思路。从另一个角度来讲,几乎所有的约简都是通过研究信息系统的对象集及其幂集上的等价关系来获得的。1992年,Novotny提出了依赖空间的概念^[7,8],在属性集而不是对象集上研究等价关系与等价类,使得属性约简在数学意义上更加清晰。文献[9,10]利用其思想研究了信息系统上的属性约简问题。

本文遵循此思路,在协调信息系统属性集的幂集上定义两个闭算子 $C(R)$ 与 $C(r)$,通过研究相应闭集族 C_r 、 C_R 的性质,得到它们与不可辨识属性集族 Ω 的关系,同时给出协调信息系统属性约简的闭算子方法。最后证明在一定条件下本文所定义的约简与文献[4,7]中的约简是等价的。该方法把对信息系统的对象集的幂集的研究转移到条件属性集上来,

在一定条件下简化了信息系统的知识发现。

2 基本概念

设 $(U, A, F, \{d\}, \{g_d\})$ 为目标信息系统^[4], $B \subseteq A$, 则 $R_B = \{(x_i, x_j) \in U \times U \mid f_i(x_i) = f_i(x_j), \forall a_i \in B\}$, $R_d = \{(x_i, x_j) \in U \times U \mid g_d(x_i) = g_d(x_j)\}$ 是 U 上的两个等价关系,从而产生 U 上的两个分划:

$$U/R_B = \{[x_i]_B \mid x_i \in U\}$$

$$U/R_d = \{[x_i]_d \mid x_i \in U\}$$

定义 1^[4] 设 $(U, A, F, \{d\}, \{g_d\})$ 为目标信息系统。若 $R_A \subseteq R_d$, 则称目标信息系统是协调的, 否则称为不协调的。

对于协调目标信息系统,若存在 $B \subseteq A$, 使得 $R_B \subseteq R_d$, 且对于任意 $b \in B$, 有 $R_{B-\{b\}} \not\subseteq R_d$, 则称 B 是信息系统的约简。

定义 2^[7] 设 A 是非空集合, R 是 A 的幂集 $B(A)$ 上的等价关系, 如果 R 满足条件: $\forall X, Y, X_1, Y_1 \in B(A), (X, X_1) \in R, (Y, Y_1) \in R \Rightarrow (X \cup Y, X_1 \cup Y_1) \in R$, 则称 R 是 $B(A)$ 上的一致关系。

3 闭算子与闭集族的性质

设 $(U, A, F, \{d\}, \{g_d\})$ 是协调目标信息系统。定义

$$R = \{(B, C) \in B(A) \times B(A) \mid "R_B \subseteq R_d \text{ 且 } R_C \subseteq R_d" \text{ 或 } "R_B$$

到稿日期:2013-12-17 返修日期:2014-03-16 本文受国家自然科学基金(61170107),高等学校博士学科点专项科研基金(20101303110004)资助。

刘 静(1981—),女,讲师,主要研究方向为粗糙集、概念格与近似推理,E-mail:lj9068@163.com;米据生(1966—),男,教授,博士生导师,主要研究方向为粗糙集与概念格、数据挖掘与近似推理等。

$$=R_C \not\subseteq R_d''\}$$

$$r=\{(b,c)\in A\times A|R_b=R_c\}$$

则有如下结果:

定理 1 r 是 A 上的等价关系, R 是 $B(A)$ 上的一致关系。

定义 3^[7] 设 A 是非空集合, 称 $C:B(A)\rightarrow B(A)$ 为闭算子, 若 C 满足以下条件:

- (1) $\forall B\in B(A), B\subseteq C(B)$;
- (2) $\forall B_1, B_2\in B(A), B_1\subseteq B_2\Rightarrow C(B_1)\subseteq C(B_2)$;
- (3) $\forall B\in B(A), C(C(B))=C(B)$ 。

由 R 得到 $B(A)$ 上的一个分划:

$$B(A)/R=\{[B]_R|B\subseteq A\}$$

其中, $[B]_R=\{C\subseteq A|(B,C)\in R\}$ 。

同样地, 由 r 得到一个分划:

$$A/r=\{[b]_r|b\in A\}$$

其中, $[b]_r=\{c\in A|R_b=R_c\}$ 。

定理 2 定义算子 $C(R):B(A)\rightarrow B(A)$:

$$\forall B\subseteq A, C(R)(B)=\begin{cases} \bigcup[B]_R, & R_B\not\subseteq R_d \\ A, & R_B\subseteq R_d \end{cases}$$

则 $C(R)$ 为 $B(A)$ 上的闭算子。

同样地, 算子 $C(r):B(A)\rightarrow B(A)$ 也是 $B(A)$ 上的闭算子。其中: $C(r)(B)=\bigcup\{E|E\cap B\neq\emptyset, E\in A/r\}, \forall B\subseteq A$ 。

定理 3 设 $\forall B\neq A$, 则 $C(R)(B)=B\iff \forall C\subseteq A$, 若 $R_C=R_B\not\subseteq R_d$, 则 $C\subseteq B$ 。

证明: 由以上定义可证。

记 $C_R=\{B|C(R)(B)=B\}$

$$C_r=\{B|C(r)(B)=B\}$$

则 C_R 与 C_r 分别是由闭算子 $C(R)$ 与 $C(r)$ 产生的闭集族, 且以下结论成立。

定理 4 C_R, C_r 与 A/r 有以下性质:

- (1) $\emptyset\in C_r, A\in C_r$;
- (2) $\emptyset\in C_R, A\in C_R$;
- (3) $E\in A/r, \forall b\in E$, 有 $\{b\}\in [E]_R$;
- (4) $E\in A/r, \forall E_1\subseteq E, E_1\in [E]_R$;
- (5) $A/r\subseteq C_r$;
- (6) $C_r=\{\bigcup E|E\in A/r\}$;
- (7) $E\in A/r, E\cap B\neq\emptyset$, 有 $B\cup E\in [B]_R$;
- (8) 记 $A=\{a_1, \dots, a_m\}$, 若 $A/r=\{E_1, \dots, E_m\}$, 即 $E_i=\{a_i\}$, 则有 $C_r=B(A)$;

(9) 若 $E_1\in A/r, E_2\in A/r, A=E_1\cup E_2$, 且 $R_{E_1}\not\subseteq R_d, R_{E_2}\not\subseteq R_d$, 则 $C(R)(E_1)=E_1, C(R)(E_2)=E_2$, 且此时 $C_R=C_r$ 。

证明: (1)–(8)略。

(9) 由定理 3 知, 要证明 $C(R)(E_1)=E_1$, 只需证明满足 $R_C=R_{E_1}\not\subseteq R_d$ 的任意属性集 $C\subseteq A$ 都是 E_1 的子集, 即 $C\subseteq E_1$ 。

假设存在 $C_1, R_{C_1}=R_{E_1}\not\subseteq R_d$, 但 $C_1\not\subseteq E_1$, 则 $C_1-E_1\neq\emptyset$ 。又 $A=E_1\cup E_2$, 所以 $C_1-E_1\subseteq E_2$, 从而 $R_{C_1-E_1}=R_{E_2}$ 。故 $R_{C_1\cup E_1}=R_{E_1\cup E_2}=R_A$ 。又 $R_{C_1}=R_{E_1}\not\subseteq R_d$, 所以 $R_{C_1\cup E_1}=R_{E_1}$ 。从而 $R_{E_1}=R_A\subseteq R_d$, 矛盾。所以满足 $R_C=R_{E_1}\not\subseteq R_d$ 的任意属性集 $C\subseteq A$ 都是 E_1 的子集, 从而 $C(R)(E_1)=E_1$, 同理可证 $C(R)(E_2)=E_2$ 。

下面证明 $C_R=C_r$, 我们设法证明 $C_R=C_r=\{\emptyset, E_1, E_2, A\}$ 。先证 $C_R=\{\emptyset, E_1, E_2, A\}$ 。由以上证明过程知 $E_1, E_2\in$

C_R 。再由性质(2)得 $A, \emptyset\in C_R$ 。以下证明 C_R 中除此 4 个元素之外不含其他元素, 即除此 4 个集合之外, A 的任何真子集均不属于 C_R 。显然, 满足 $C\subseteq E_1\subseteq A$ 与满足 $C\subseteq E_2\subseteq A$ 的 A 的两类真子集均不属于 C_R 。

下面只需说明由 E_1 中的部分元素和 E_2 中的部分元素组成的 A 的任意真子集 C 也不属于 C_R 。不妨设 $C=\{a, b\}$, 其中 $a\in E_1, b\in E_2$ 。由 $R_{\{a\}}=R_{E_1}$ 和 $R_{\{b\}}=R_{E_2}$ 得 $R_{\{a,b\}}=R_{E_1\cup E_2}$, 即 $R_C=R_A$ 。从而 $C(R)(C)=A\neq C$, 即 $C\notin C_R$, 所以 $C_R=\{\emptyset, E_1, E_2, A\}$ 。类似地, 可以证明 $C_r=\{\emptyset, E_1, E_2, A\}$ 。

设 $(U, A, F, \{d\}, \{g_d\})$ 是协调的目标信息系统, 定义其不可辨识属性矩阵 $\Omega=(C_{ij})_{n\times n}, i, j\leq n$ 。其中 $C_{ij}=\begin{cases} \emptyset, & g_d(x_i)=g_d(x_j) \\ \{a_l|f_l(x_i)=f_l(x_j)\}, & g_d(x_i)\neq g_d(x_j) \end{cases}$ 为 x_i 与 x_j 的不可辨识属性集。为方便起见, 我们记不可辨识属性集合的全体为 $\Omega=\{C_{ij}|1\leq i, j\leq n\}$ 。则有以下性质:

定理 5 $\Omega\subseteq C_R\subseteq C_r$ 。

证明: 先证 $\Omega\subseteq C_R$, 即对任意 $B\in\Omega$, 有 $C(R)(B)=B$ 。

由于 $C(R)$ 为闭算子, 因此 $C(R)(B)\supseteq B$, 下面证明 $C(R)(B)\subseteq B$ 。

因为 $B\in\Omega$, 若 $B=\emptyset$, 则 $C(R)(\emptyset)=\emptyset$ 。显然 $C(R)(B)\subseteq B$ 。若 $B\neq\emptyset$, 则存在 x_i, x_j 使 $B=\{a_l\in A|f_l(x_i)=f_l(x_j)\}$, 且 $g_d(x_i)\neq g_d(x_j)$ 。显然 $R_B\not\subseteq R_d$, 所以 $C(R)(B)=C(R)(\{a_l|f_l(x_i)=f_l(x_j)\})$ 。由 $\forall C\in [B]_R$ 有 $R_B=R_C\not\subseteq R_d$ 。由 $(x_i, x_j)\in R_B$ 得 $(x_i, x_j)\in R_C$ 。因此 $\forall a\in C$, 有 $f_a(x_i)=f_a(x_j)$ 。于是 $a\in B$, 即 $C\subseteq B$ 。这样 $\bigcup[B]_R\subseteq B$ 。从而 $C(R)(B)\subseteq B$ 。

再证 $C_R\subseteq C_r$ 。即对于 $\forall B\in C_R$, 有 $B\in C_r$ 成立。

若 $B\in C_R$, 则 $C(R)(B)=B$ 。

(1) 若 $B=A$, 则 $B\in C_r$ 。

(2) 若 $B\neq A, B\subseteq A$, 则 $C(R)(B)=B\neq A$ 。由闭算子 $C(R)$ 的定义知 $R_B\not\subseteq R_d$ 。则 $C(R)(B)=\bigcup[B]_R=B$ 。所以由 $\forall C\in [B]_R$, 有 $C\subseteq B$, 且 $R_B=R_C\not\subseteq R_d$ 。 $\forall E\in A/r$, 当 $E\cap B\neq\emptyset$ 时, 由定理 4 的性质(7)可知, $E\cup B\in [B]_R$, 所以 $B\subseteq E\cup B\subseteq B$, 从而 $E\subseteq B$ 。因此 $C(r)(B)=\bigcup\{E\in A/r|E\cap B\neq\emptyset\}=\bigcup\{E\in A/r|E\subseteq B\}\subseteq B$ 。又 $\forall b\in B, \exists E\in A/r$, 使得 $b\in E$, 故 $E\cap B\neq\emptyset$ 。这样, $b\in E\subseteq C(r)(B)$, 从而 $B\subseteq C(r)(B)$ 。因此 $B=C(r)(B)$, 即 $B\in C_r$ 。综上可知 $C_R\subseteq C_r$ 。

定理 6 $\Omega\cup\{A\}=C_R$ 。

例 1 表 1 是一个协调的目标信息系统 $(U, A, F, \{d\}, \{g_d\})$ 。

表 1

	a_1	a_2	a_3	d
x_1	1	2	3	1
x_2	2	1	1	1
x_3	1	3	2	2
x_4	3	3	3	1
x_5	2	3	2	2
x_6	3	1	2	2

该目标信息系统的不可辨识属性集由表 2 给出。

表 2

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	$\emptyset$					
x_2	$\emptyset$	$\emptyset$				
x_3	$\{a_1\}$	$\emptyset$	$\emptyset$			
x_4	$\emptyset$	$\emptyset$	$\{a_2\}$	$\emptyset$		
x_5	$\emptyset$	$\{a_1\}$	$\emptyset$	$\{a_2\}$	$\emptyset$	
x_6	$\emptyset$	$\{a_2\}$	$\emptyset$	$\{a_1\}$	$\emptyset$	$\emptyset$

由此可得 $\Omega = \{\emptyset, \{a_1\}, \{a_2\}\}$ 。

$C_R = \{\emptyset, \{a_1\}, \{a_2\}, A\}$ 。

另外,易知 $A/r = \{\{a_1\}, \{a_2\}\}$ 。因此由定理 4 的性质 (8) 可得 $C_r = B(A)$ 。

于是得到 $\Omega \subseteq C_R \subseteq C_r$, 且 $\Omega \cup \{A\} = C_R$ 。

4 属性约简的闭算子方法

一般来讲, C_R 与 C_r 不同。例如, 例 1 所给的目标信息系统中, $C_R = \{\emptyset, \{a_1\}, \{a_2\}, A\}$, $C_r = B(A)$, 显然 $C_R \neq C_r$ 。

下面给出这两个闭族族相等的充要条件。

定理 7 $C_R = C_r \Leftrightarrow$ 对于任意的 $\alpha, \beta \subseteq A/r$ 且 $\alpha \neq \beta$, 记 $B = \bigcup_{E_i \in \alpha} E_i, C = \bigcup_{E_i \in \beta} E_i$, 当 $B \neq C$ 时, 有 $(B, C) \notin R$ 。

证明: 需证明如下逆否命题:

$C_R \neq C_r \Leftrightarrow$ 存在 $\alpha, \beta \subseteq A/r$, 且 $\alpha \neq \beta$, 记 $B = \bigcup_{E_i \in \alpha} E_i, C = \bigcup_{E_i \in \beta} E_i$, 有 $B \neq C$, 但 $(B, C) \in R$ 。

必要性 若 $C_R \neq C_r$, 由定理 5 可知, 存在 $B \in C_r$, 但 $B \notin C_R$, 即 $C(R)(B) \neq B$, $C(R)(B) \neq B$ 。由于 $B \in C_r$, 根据定理 4 的性质 (6), B 可写成 $B = \bigcup_{E_i \in \alpha} E_i, \alpha \subseteq A/r$ 。记 $C = C(R)(B)$, 则由于 $C(R)$ 是闭算子, 因此 $C(R)(C) = C$ 。所以 $C \in C_R \subseteq C_r$, 从而 $C = \bigcup_{E_i \in \beta} E_i, \beta \subseteq A/r$ 。显然 $\alpha \neq \beta, B \neq C$ 。但 $C = C(R)(B)$ 。

由 $C(R)$ 的定义可得 $C = C(R)(B) = \begin{cases} \bigcup [B]_R, & R_B \not\subseteq R_d \\ A, & R_B \subseteq R_d \end{cases}$, 若 $R_B \subseteq R_d$, 则 $C = A, R_C = R_A \subseteq R_d$ 。若 $R_B \not\subseteq R_d$, 则 $C = \bigcup [B]_R$ 。又 $\forall D \in [B]_R, R_D = R_B \not\subseteq R_d$ 。所以 $R_C = R_{\bigcup [B]_R} = \bigcap_{D \in [B]_R} R_D = R_B \not\subseteq R_d$ 。

综上可知, $(B, C) \in R$ 。

充分性 若存在 $\alpha, \beta \subseteq A/r$, 且 $\alpha \neq \beta$ 。

记 $B = \bigcup_{E_i \in \alpha} E_i, C = \bigcup_{E_i \in \beta} E_i$, 且当 $B \neq C$ 时, 有 $(B, C) \in R$, 则 $B, C \in C_r$ 。若 $R_B = R_C \not\subseteq R_d$, 则 $R_{B \cup C} = R_B = R_C \not\subseteq R_d$ 。若 $R_B \subseteq R_d$ 且 $R_C \subseteq R_d$, 则 $R_{B \cup C} = R_B \cap R_C \subseteq R_d$ 。故 $(B \cup C, B) \in R, (B \cup C, C) \in R$, 所以 $B \subseteq B \cup C \subseteq C(R)(B), C \subseteq B \cup C \subseteq C(R)(C)$, 也就是 $B, C \in C_R$, 因此 $C_R = C_r$ 。

例 2 考察表 3 所列的协调目标信息系统 $(U, A, F, \{d\}, \{g_d\})$ 。

表 3

	a_1	a_2	a_3	d
x_1	1	0	1	1
x_2	0	1	1	2
x_3	0	1	0	1
x_4	0	1	1	2
x_5	0	1	0	1
x_6	0	1	0	1

容易计算 $A/r = \{\{a_1, a_2\}, \{a_3\}\}$ 。记 $E_1 = \{a_1, a_2\}, E_2 = \{a_3\}$ 。则 A/r 的子集共有 4 个, 分别是 $\emptyset, \{E_1\}, \{E_2\}, A/r$ 。满足定理 7 的 (α, β) 组合共有 6 种:

- (1) $\{\emptyset, \{E_1\}\}$ 。此时 $B = \emptyset, C = E_1$ 。
- (2) $\{\emptyset, \{E_2\}\}$ 。此时 $B = \emptyset, C = E_2$ 。
- (3) $\{\emptyset, A/r\}$ 。此时 $B = \emptyset, C = A$ 。
- (4) $\{\{E_1\}, \{E_2\}\}$ 。此时 $B = E_1, C = E_2$ 。
- (5) $\{\{E_1\}, A/r\}$ 。此时 $B = E_1, C = A$ 。
- (6) $\{\{E_2\}, A/r\}$ 。此时 $B = E_2, C = A$ 。

通过计算, 以上 6 种情形都有 $(B, C) \notin R$ 。

定理 7 的条件成立, 故在本例中 $C_R = C_r$ 。

对定理 7 进行分析易知, 若 $C_R = C_r$, 则可以对条件属性集 A 进行约简。具体方法是: 在每个 $E_i \in A/r$ 中任选一个元素, 这些元素组成的集合就是 A 的一个约简。

定理 8 当 $C_R = C_r$ 时, 记 $A/r = \{E_1, \dots, E_n\}$ 。则 B 为 A 的约简 $\Leftrightarrow B = \{e_1, \dots, e_n\}$, 其中 $e_i \in E_i (i=1, \dots, n)$ 。

证明: (必要性) 若 B 为 A 的约简, 则 $R_B \subseteq R_d$ 。所以 $(A, B) \in R$ 。假设 $B \neq \{e_1, \dots, e_n\}$ 。不妨设 B 不含 E_n 中的元素, 并记 $B = \{e_1, \dots, e_{n-1}\}$ 。则 $R_B = R_{\bigcup_{i=1}^{n-1} E_i} = \bigcap_{i=1}^{n-1} R_{E_i} = \bigcap_{i=1}^{n-1} R_{\bigcup_{j=1}^{n-1} E_j}$ 。因为 $R_B \subseteq R_d$, 所以 $(B, \bigcup_{i=1}^{n-1} E_i) \in R$ 。因此 $(A, \bigcup_{i=1}^{n-1} E_i) \in R$ 。但 $A = \bigcup_{i=1}^n E_i \neq \bigcup_{i=1}^{n-1} E_i$, 由定理 7 得 $(A, \bigcup_{i=1}^{n-1} E_i) \notin R$, 矛盾。所以 $B = \{e_1, \dots, e_n\}$ 。

(充分性) 若 $B = \{e_1, \dots, e_n\}$, 则 $R_B = R_{\bigcup_{i=1}^n E_i} = \bigcap_{i=1}^n R_{E_i} = R_{\bigcup_{i=1}^n E_i} = R_A \subseteq R_d$ 。假设 B 不是 A 的约简, 则存在 $B_1 \subseteq B$ 且 $B_1 \neq B$, 使 B_1 为 A 的约简。所以 $R_{B_1} \subseteq R_d$, 且 $(B_1, A) \in R$ 。令 $B_1 = \{e_{i_1}, \dots, e_{i_m}\}$, 则 $R_{B_1} = R_{\bigcup_{j=1}^m E_{i_j}} = \bigcap_{j=1}^m R_{E_{i_j}} = \bigcap_{j=1}^m R_{\bigcup_{i=1}^m E_{i_j}}$, 所以 $(B_1, \bigcup_{j=1}^m E_{i_j}) \in R$, 因此 $(\bigcup_{j=1}^m E_{i_j}, A) \in R$ 。但 $\bigcup_{j=1}^m E_{i_j} \neq \bigcup_{i=1}^n E_i = A$, 由定理 7 得 $(\bigcup_{j=1}^m E_{i_j}, A) \notin R$, 矛盾。所以 B 为 A 的约简。

下面给出协调目标信息系统 $(U, A, F, \{d\}, \{g_d\})$ 的另一种属性约简方法。

定义 4 设 $(U, A, F, \{d\}, \{g_d\})$ 为协调的目标信息系统, $B \subseteq A$ 。 $R = \{(B, C) \in B(A) \times B(A) \mid "R_B \subseteq R_d \text{ 且 } R_C \subseteq R_d" \text{ 或 } "R_B = R_C \not\subseteq R_d"\}$ 。

称 B 为 A 的约简当且仅当 B 是 $M = \{D_1 \subseteq A, D_1 \cap D \neq \emptyset, \forall D \in \overline{C_R}\}$ 的极小元。其中 $\overline{C_R} = \{D^c \mid D \in C_R, D \neq A\}$ 。

例 3 (续例 2) 在表 3 所示的协调目标信息系统中, 由例 2 得 $C_R = C_r = \{\emptyset, \{a_1, a_2\}, \{a_3\}, A\}$, $A/r = \{\{a_1, a_2\}, \{a_3\}\}$ 。记 $E_1 = \{a_1, a_2\}, E_2 = \{a_3\}$, 则由定理 8 知, 该目标信息系统的约简为 $\{a_1, a_3\}, \{a_2, a_3\}$ 。

文献[7]给出了属性约简的另一种形式:

定义 5^[7] 设 $(U, A, F, \{d\}, \{g_d\})$ 为协调目标信息系统。 $R = \{(B, C) \in B(A) \times B(A) \mid "R_B \subseteq R_d \text{ 且 } R_C \subseteq R_d" \text{ 或 } "R_B = R_C \not\subseteq R_d"\}$ 。则 B 是 A 的约简当且仅当 B 是 A 所在的 R 等价类 $B \in B(A)/R$ 的极小元。

例 4 (续例 3) 应用定义 4 和定义 5 分别对表 3 所示的目标信息系统进行约简。

由例 3 得 $C_R = C_r = \{\emptyset, \{a_1, a_2\}, \{a_3\}, A\}$, $A/r = \{\{a_1, a_2\}, \{a_3\}\}$ 。

应用定义 4。由于 $\overline{C_R} = \{\{a_1, a_2\}, \{a_3\}, A\}$, $M = \{A, \{a_1, a_3\}, \{a_2, a_3\}\}$, 因此 M 的极小元为 $\{a_1, a_3\}, \{a_2, a_3\}$, 所以 A 的约简为 $\{a_1, a_3\}, \{a_2, a_3\}$ 。

应用定义 5。因为 $B(A)$ 关于 R 的等价类为 $\{\emptyset\}, \{\{a_1\}, \{a_2\}, \{a_1, a_2\}\}, \{\{a_3\}\}, \{A, \{a_1, a_3\}, \{a_2, a_3\}\}$, 而 A 所在的 R 等价类 $\{A, \{a_1, a_3\}, \{a_2, a_3\}\}$ 的极小元为 $\{a_1, a_3\}, \{a_2, a_3\}$,

(下转第 265 页)

西矩阵都可以作为参与人的策略,本文为了简化运算,将参与人的策略做了限定。策略矩阵不同,博弈终止态也会不同,从而影响收益和最后的聚类精度。所以可以尝试使用其它西矩阵作为参与人的策略,以观察结果。

(3)本文只考虑了两人两策略的最简博弈模型,后续研究可以考虑多人同时参与博弈的情形^[13],如3个邻居同时参与。这样,在量子情形下,每个人的收益都将受到多个邻居的影响,博弈的收益矩阵需要重新设计,博弈终止态也将发生变化,从而得到不同的聚类结果。

参考文献

- [1] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(1):48-61
- [2] 焦宝聪,陈兰平,方海光. 博弈论-思想方法及应用[M]. 北京:中国人民大学出版社,2013:118-121
- [3] Meyer D A. Quantum strategies[J]. Physical Review Letters, 1998,82(5):1052-1055
- [4] 李强. 动点聚类算法及其量子化研究[D]. 杭州:浙江大学电气工程学院,2009

- [5] 王龙,王靖,武斌. 量子博弈:新方法与新策略[J]. 智能系统学报,2008,3(4):294-304
- [6] Eisert J, Wilkens M, Lewenstein M. Quantum games and quantum strategies[J]. Physical Review Letters A, 1999, 83(15): 3077-3080
- [7] 李威,赵红敏,林家逖. 量子博弈论及其应用[J]. 大学物理, 2003,22(12):3-8
- [8] 钱国红. 量子算法及其在数据挖掘中的应用[D]. 杭州:浙江工业大学,2012
- [9] Guo Hong, Zhang Ju-heng, Koehler G J. A survey of quantum games[J]. Decision Support Systems, 2008,46(1):318-332
- [10] Hauert C, Doebeli M. Spatial structure often inhibits the evolution of cooperation in snowdrift game[J]. Nature, 2004, 428: 643-646
- [11] 钱国红,黄德才,陆亿红. 广义加权 Minkovski 距离及量子遗传聚类算法[J]. 计算机科学,2013,40(5):224-228
- [12] 赵恒,杨万海. 模糊 K-Modes 聚类精确度分析[J]. 计算机工程, 2003,29(12):27-28
- [13] 周杰,李筠,马雷. 任意多人的量子博弈[J]. 量子电子学报, 2006,23(2):173-177

(上接第251页)

故A的约简为 $\{a_1, a_3\}$ 与 $\{a_2, a_3\}$ 。

由定义4和定义5求得的约简与例3的结论完全一样。

一般地,我们有以下定理。

定理9 当 $C_R=C_r$ 时,由定义1、定义4和定义5所定义的3种属性约简等价。

证明:我们只需证明下面3个条件等价:

(1) $B=\{e_1, \dots, e_n\}, e_i \in E_i, E_i \in A/r(i=1, \dots, n)$ 。

(2) B 是 A 所在的 R 等价类的极小元。

(3) B 是 $M=\{D_1 \subseteq A, D_1 \cap D \neq \emptyset, \forall D \in \overline{C_R}\}$ 的极小元。

其中 $\overline{C_R}=\{D^c | D \in C_R, D \neq A\}$ 。

(1) $\Rightarrow$ (2):若 $B=\{e_1, \dots, e_n\}$,则 $R_B=R_{\bigcup_{i=1}^n E_i}=\bigcap_{i=1}^n R_{E_i}=R_{\bigcup_{i=1}^n E_i}=R_A \subseteq R_d$ 。所以 $(A, B) \in R$ 。如果 B 不是 A 所在的 R 等价类 $B \in B(A)/R$ 的极小元,则存在 $B_1 \subseteq B$ 且 $B_1 \neq B$,使 B_1 为 B 的极小元。不妨设 $B_1=\{e_{i_1}, \dots, e_{i_m}\}$ 。则 $R_{B_1}=R_{\bigcup_{j=1}^m E_{ij}}=\bigcap_{j=1}^m R_{E_{ij}}=R_{\bigcup_{j=1}^m E_{ij}}$ 。又 $R_A=R_{B_1} \subseteq R_d$,所以 $(\bigcup_{j=1}^m E_{ij}, A) \in R$ 。

但 $\bigcup_{j=1}^m E_{ij} \neq \bigcup_{i=1}^n E_i=A$,这与定理7矛盾。故 B 是 A 所在的 R 等价类 $B \in B(A)/R$ 的极小元。

(2) $\Rightarrow$ (1):若 B 是 A 所在的 R 等价类 $B \in B(A)/R$ 的极小元, $B=\{e_1, \dots, e_n\}$,则 $R_{B_1}=R_{\bigcup_{i=1}^n E_i}=R_A$ 。所以 $(B_1, A) \in R$,从而 $B \subseteq B_1$ 。假设 $B \neq B_1$,则 B 为 B_1 的真子集。不妨设 $B=\{e_{i_1}, \dots, e_{i_m}\}$,则 $R_B=R_{\bigcup_{j=1}^m E_{ij}}=R_{\bigcup_{j=1}^m E_{ij}}$ 。所以 $(\bigcup_{j=1}^m E_{ij}, B) \in R$ 。又 $\bigcup_{j=1}^m E_{ij} \neq \bigcup_{i=1}^n E_i$,由定理7知 $(\bigcup_{j=1}^m E_{ij}, A) \notin R$,所以 $(B, A) \notin R$ 。这与已知条件矛盾,因此 $B=\overline{B_1}=\{e_1, \dots, e_n\}$ 。

(1) $\Leftrightarrow$ (3):当 $C_R=C_r$ 时, $\overline{C_R}=\overline{C_r}=\{D^c | D \in C_r, D \neq A\}$,而 $C_r=\{\bigcup_i E_i | E_i \in A/r\}$ 。于是 $B=\{e_1, \dots, e_n\} \Leftrightarrow B$ 是 $\{D_1 \subseteq A | D_1 \cap E_i \neq \emptyset\}$ 的极小元 $\Leftrightarrow B$ 是 $\{D_1 \subseteq A | D_1 \cap D \neq \emptyset, \forall D \in C_r, D \neq \emptyset\}$ 的极小元 $\Leftrightarrow B$ 是 $M=\{D_1 \subseteq A | D_1 \cap E_i \neq \emptyset\}$ 的极小元。

结束语 本文另辟新径,对于协调信息系统,在其条件属性集而非对象集的幂集上定义了两个闭算子,讨论了由它们所生成的两个闭集族的性质,提出了属性约简的一种新方法。给出了 $C_r=C_R$ 的充要条件,并在此条件下证明了本文所提出的新方法已有结论的等价性。利用这些结果,可以把对信息系统的对象集的幂集的研究转移到条件属性集上来,从而在一定条件下简化了信息系统的知识发现。

参考文献

- [1] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法. 北京:科学出版社,2001
- [2] 王珏,王任,苗夺谦,等. 基于 Rough set 理论的“数据浓缩”[J]. 计算机学报,1998,21(5):393-400
- [3] 王国胤. 粗糙集理论与知识获取[M]. 西安:西安交通大学出版社,2001
- [4] 张文修,梁怡,吴伟志. 信息系统与知识发现[M]. 北京:科学出版社,2003
- [5] Kryzkievich M. Comparative study of alternative types of knowledge reduction in inconsistent systems[J]. International Journal of Intelligent Systems, 2001,16:105-120
- [6] 张文修,米据生,吴伟志. 不协调目标信息系统的知识约简[J]. 计算机学报,2003,26(1):12-18
- [7] Novotny M. Dependence spaces of information systems[C]// Ortowska E, et al. Incomplete informations: Rough set analysis. Physica-Verlag, 1998:193-246
- [8] Novotny M, Pawlak Z. On a problem concerning dependence spaces[J]. Fundamenta Informaticae, 1992,16:275-287
- [9] Ling Wei, Zhang Wen-xiu. Attribute reduction based on equivalence relation defined on attribute set and its power set[C]// International workshop on monitoring, security and rescue techniques in multiagent systems MSRAS, 2004:317-326
- [10] Wei Ling, Li Hong-ru, Zhang Wen-xiu. Knowledge reduction based on the equivalence relations defined on attribute set and its power set[J]. Information Sciences, 2007,177:3178-3185