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 $CR - C^{\prime}R$ 的充要条件，并证明所定义的属性约简与文献[4,7]中所述的等价。

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.isnn.1002-137X.2014.10.052

1 引言

属性约简是知识发现的一个重要研究课题[1-4]。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，可以使信息系统中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、胚约简、相似关系下的约简[4-8]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。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 $C_R$ 与 $C_r$ 的性质，得到它们与不可辨识属性集族 $\Omega$ 的关系，同时给出协调信息系统属性集的闭算子方法。最后证明在一定条件下本文所定义的约简与文献[4,7]中所述的等价。

2 基本概念

设 $(U,A,F,\{d\},\{g_d\})$ 为目标信息系统[4]。若 $B \subseteq A$，则 $R_B=(\{\{(x_1,x_2)\in U\times U| f_1(x_1)=f_1(x_2)\}\ \forall \ x_1 \in B\}) \subseteq R_A$。若 $R_B 

$ R_B 

定义 $1$ 设 $(U,A,F,\{d\},\{g_d\})$ 为目标信息系统。若 $R_B \subseteq R$, 则称目标信息系统 $B$ 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 $R_B \subseteq R_u$，且对于任意 $b \in B$，有 $R_{\bar{b}} \subseteq R_u$, 则称 $B$ 是信息系统的一个约简。

定义 $2$ 设 $A$ 是非空集合，$R$ 是 $A$ 的幂集 $\mathcal{B}(A)$ 上的等价关系，若 $R$ 满足条件：$\forall X,Y,Z \in \mathcal{B}(A), (X,Y) \in R \Rightarrow (X \cup Y, Z \cup Y) \in R$, 则称 $R$ 是 $\mathcal{B}(A)$ 上的一致关系。

3 闭算子与闭族集的性质

设 $(U,A,F,\{d\},\{g_d\})$ 是协调目标信息系统，定义 $R=((B,C) \in \mathcal{B}(A) \times \mathcal{B}(A) | R_B \subseteq R_c \ \text{且} \ \text{R}_c \subseteq R_d)$ 或 $R_b$
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. 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 $B(A)$ 上的一个划分: 

$B(A)/R = \{ [B]_{R} \mid B \subseteq A \}$

其中，$[B]_{R} = \{ x \in A \mid (x, c) \in R \}$. 同样地，由 $r$ 得到一个划分: 

$A/r = \{ [B]_{r} \mid B \subseteq A \}$

其中，$[B]_{r} = \{ x \in A \mid (x, c) \in R \}$. 由 $C$ 得到 $B(A)$ 上的闭算子: 

$C(R) : B(A) \rightarrow B(A)$

下同。再证 $C(r)$. 即对任意 $B \in \Omega$, 有 $C(R)(B) = B$. 由于 $C(R)$ 为闭算子，因此 $C(R)(B) \subseteq B$. 进一步证明 $C(R)(B) = B$. 因为 $B \in \Omega$，若 $B = \emptyset$，则 $C(R)(\emptyset) = \emptyset$. 那么 $C(R)(B) \subseteq B$.

若 $B \neq \emptyset$，则存在 $x_1, x_2$ 使 $B = \{ (a_1 \in A \mid f_1(x_1) = f_1(x_2) \}$. 且 $g_1(x_1) = g_1(x_2)$，显然 $R_1 \subseteq R_2$. 通过 $C(R)(B) = C(R)((a_1 \in A \mid f_1(x_1) = f_1(x_2)))$. 对于 $x_1, x_2 \in B$, 使得 $g_1(x_1) = g_1(x_2)$，于是 $a \in B$, 即 $C(R)(B) = U [B]_{R} \subseteq B$. 从而 $C(R)(B) \subseteq B$.

再证 $C(r)$. 即对任意 $B \in \Omega$, 有 $C(r)(B) = B$. 由于 $C(r)$ 为闭算子，因此 $C(r)(B) \subseteq B$. 由此证明 $C(r)(B) = B$. 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 $x_1, x_2$ 使 $B = \{ (a_1 \in A \mid f_1(x_1) = f_1(x_2) \}$. 且 $g_1(x_1) = g_1(x_2)$，显然 $R_1 \subseteq R_2$. 通过 $C(r)(B) = C(r)((a_1 \in A \mid f_1(x_1) = f_1(x_2)))$. 对于 $x_1, x_2 \in B$, 使得 $g_1(x_1) = g_1(x_2)$，于是 $a \in B$, 即 $C(r)(B) = U [B]_{R} \subseteq B$. 从而 $C(r)(B) \subseteq B$.

再证 $C(r)$. 即对任意 $B \in \Omega$, 有 $C(r)(B) = B$. 由于 $C(r)$ 为闭算子，因此 $C(r)(B) \subseteq B$. 由此证明 $C(r)(B) = B$. 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 $x_1, x_2$ 使 $B = \{ (a_1 \in A \mid f_1(x_1) = f_1(x_2) \}$. 且 $g_1(x_1) = g_1(x_2)$，显然 $R_1 \subseteq R_2$. 通过 $C(r)(B) = C(r)((a_1 \in A \mid f_1(x_1) = f_1(x_2)))$. 对于 $x_1, x_2 \in B$, 使得 $g_1(x_1) = g_1(x_2)$，于是 $a \in B$, 即 $C(r)(B) = U [B]_{R} \subseteq B$. 从而 $C(r)(B) \subseteq B$.

下同。再证 $C(r)$. 即对任意 $B \in \Omega$, 有 $C(r)(B) = B$. 由于 $C(r)$ 为闭算子，因此 $C(r)(B) \subseteq B$. 由此证明 $C(r)(B) = B$. 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 $x_1, x_2$ 使 $B = \{ (a_1 \in A \mid f_1(x_1) = f_1(x_2) \}$. 且 $g_1(x_1) = g_1(x_2)$，显然 $R_1 \subseteq R_2$. 通过 $C(r)(B) = C(r)((a_1 \in A \mid f_1(x_1) = f_1(x_2)))$. 对于 $x_1, x_2 \in B$, 使得 $g_1(x_1) = g_1(x_2)$，于是 $a \in B$, 即 $C(r)(B) = U [B]_{R} \subseteq B$. 从而 $C(r)(B) \subseteq B$.}

再证 $C(r)$. 即对任意 $B \in \Omega$, 有 $C(r)(B) = B$. 由于 $C(r)$ 为闭算子，因此 $C(r)(B) \subseteq B$. 由此证明 $C(r)(B) = B$. 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 $x_1, x_2$ 使 $B = \{ (a_1 \in A \mid f_1(x_1) = f_1(x_2) \}$. 且 $g_1(x_1) = g_1(x_2)$，显然 $R_1 \subseteq R_2$. 通过 $C(r)(B) = C(r)((a_1 \in A \mid f_1(x_1) = f_1(x_2)))$. 对于 $x_1, x_2 \in B$, 使得 $g_1(x_1) = g_1(x_2)$，于是 $a \in B$, 即 $C(r)(B) = U [B]_{R} \subseteq B$. 从而 $C(r)(B) \subseteq B$.
由此可得 $\Omega = (\{a_1\}, \{a_2\})$。

$C_k = (\{a_1\}, \{a_2\})_A$。

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 $\Omega \subseteq C_r \subseteq C$, 且 $\Omega \cup \{A\} = C_k$.

4 属性约简的闭算子方法

一般来讲，与 $c_r$ 不同。例如，例 1 所给的目标信息系统中，$C_k = (\{a_1\}, \{a_2\}, \{a_3\}) = B(A)$，显然 $C_k \neq C_r$。

下面给出这两个闭算子等价的充要条件。

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

证明：需证明如下逆否命题：

定理 7 的条件成立，则有 $(B, C) \notin R$。

必要性 若有 $B \neq C$，则存在 $B = \{E_i\} \subseteq \Omega$, $C = \{E_i\} \subseteq \Omega$, 当 $B \neq C$ 时，有 $(B, C) \notin R$。

充分性 若有 $B \neq C$，则存在 $B = \{E_i\} \subseteq \Omega$, $C = \{E_i\} \subseteq \Omega$, 当 $B \neq C$ 时，有 $(B, C) \notin R$。

由定理 7 的条件可得 $C = C_r(B) = \{B\}$，有 $B = \{E_i\} \subseteq \Omega$, $C = \{E_i\} \subseteq \Omega$, $R_B \subseteq R_c$，则 $C = A$, $R_c = R_c$, 且 $R_q \subseteq R_c$。若 $R_B \subseteq R_c$，则有 $(B, C) \in R$，且 $(B, C) \in R$。若 $D \subseteq \Omega$, $R_c \subseteq R_c$, 则 $R_c \subseteq R_c$，且 $(B, C) \in R$，且 $(B, C) \in R$。

综上可得，$(B, C) \in R$。

充分性 若存在 $a, \beta \subseteq A/r$, 且 $a \neq \beta$，则有 $B = \{E_i\} \subseteq \Omega$, $C = \{E_i\} \subseteq \Omega$，有 $(B, C) \in R$。若 $R_B \subseteq R_c$，则 $C = A$, $R_c = R_c$, 且 $R_q \subseteq R_c$。若 $R_B \subseteq R_c$，则 $R_c = R_c$, 且 $(B, C) \in R$，且 $(B, C) \in R$。

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

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

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

(1) $(\Omega, \{E_i\})$, 此时 $B = C = \Omega$。

(2) $(\Omega, \{E_i\})$, 此时 $B = C = \{E_i\}$。

(3) $(\Omega, A/r)$, 此时 $B = C = A$。

(4) $(\{E_i\}, \{E_i\})$, 此时 $B = C = \{E_i\}$。

(5) $(\{E_i\}, A/r)$, 此时 $B = E_i, C = A$。

(6) $(\{E_i\}, A/r)$, 此时 $B = E_i, C = A$。
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。可以尝试使用其它西矩阵作为参与人的策略，以观察结果。

(2)本文只考虑了两人两策略的最简单博弈模型,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


索引语: 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聚类上定义了两个闭算子，讨论了由它们所生成的两个闭算子的性质，提出了一种新的方法。给出了 $C_0 = C$, 若 $B$ 使 $C_0 = C$。从而在一定条件下简化了信息系统的知识发现。