文章编号: 1001-0920 (2011) 06-0867-06

不一致不完备决策系统中属性约简的比较研究

蒙祖强, 黄柏雄

(广西大学 计算机与电子信息学院，南宁 530004)

摘　要: 相对其他决策系统, 不一致不完备决策系统对于复杂数据具有更强的数据建模和表示能力, 这决定了其在实际应用中的重要作用. 鉴于此, 在已有研究成果的基础上深入探讨了不一致不完备决策系统中约简的定义方法, 提出了不一致不完备决策系统中的 5 种约简概念, 并找出了它们之间的关联, 为进一步研究此类系统中知识约简的理论和方法奠定了基础.

关键词: 决策系统；不一致；不完备；约简

中图分类号: TP18　　**文献标识码:** A

Comparative study of attribute reduction in inconsistent incomplete decision system

MENG Zu-qiang, *HUANG Bo-xiong*

(College of Computer and Electronics Information，Guangxi University，Nanning 530004，China．Correspondent: MENG Zu-qiang，E-mail：zqmeng@126.com)

Abstract: Inconsistent incomplete decision system(IIDS) has strong superiority in modeling and expressing complex data compared with other decision systems, which shows its important value. Therefore, based on existed research results, this paper investigates the definition method of reduction in IIDS. Five kinds of reduction in IIDS, such as possible reduct, are presented and the relationships between them are also found out. This work can form the foundation for further studying related theory and method of reduction in IIDS.

Key words: decision system；inconsistent；incomplete；reduction

1 引　言

决策系统从完备与否的角度可分为完备决策系统和不完备决策系统; 从一致与否的角度可分为一致决策系统和不一致决策系统. 据此, 决策系统可分为一致完备决策系统、一致不完备决策系统、不一致完备决策系统和不一致不完备决策系统 4 种类型.

目前, 对于一致完备决策系统的约简问题, 利用传统的 Rough 集理论在约简效率方面可以得到较好的解决. 该理论中等价关系下的数据样本呈线序关系, 具有良好的数学性质[1], 因此可以利用排序技术来求正域, 并构造高效的约简算法[2], 后又经进一步改进, 使得约简效率已十分接近线性复杂度[3-4].

对于一致不完备决策系统, 由于系统中的数据样本一般不存在线序关系 (而是一种格结构), 不能用排序技术来构造高效算法, 即传统 Rough 集理论对此“力不从心”. 通过粒化方法对基于条件属性的信息系统进行分解, 形成一系列的相交粒. 在同一粒中, 数据样本呈线性关系, 引入快速排序技术实现该粒相对于决策类的划分, 通过各粒的划分计算基于条件属性的信息系统对于决策类的正域, 进而基于正域计算决策系统的约简, 并由此提出基于相容关系的快速属性约简算法[5], 这在一定程度上解决了此类系统的高效约简问题.

对于不一致完备决策系统, 条件属性下的相容粒度划分对于决策类形成多个交叉子集, 这样, 传统约简定义所依赖的“系统分类能力不变”的意义发生了变化, 即从不同的角度看, 对“系统分类能力不变”有着不同的理解, 导致出现了多种约简概念[6]. 张文修等人[7-8]对此进行了深入的研究, 进一步丰富了约简的概念, 给出了基于分辨矩阵的各种约简的计算方法, 但其效率有待于进一步提高.

收稿日期: 2010-03-15；**修回日期:** 2010-05-11.

基金项目: 国家自然科学基金项目(61063032)；广西教育厅科学基金项目(201012MS010).

作者简介: 蒙祖强(1974−), 男, 教授, 博士后, 从事粒度计算、知识发现等研究；黄柏雄(1971−), 男, 讲师, 从事数据挖掘等研究.

对于不一致不完备决策系统, 由于目前的复杂数据在形式上高维、海量、异构, 在内容上不完整、不确定、无序、歧义, 使得对这些数据进行结构化转换后所形成的表示模型通常是不一致不完备决策系统, 即不一致不完备决策系统更适合于对现今复杂数据的建模和表示, 更具有实际应用价值. 但针对不一致不完备决策系统约简问题的研究报道还较为少见, 相应的约简理论和方法非常欠缺. 为此, 本文拟对此类系统中的约简概念进行探讨, 揭示各类约简之间的关联, 为进一步研究约简理论和方法奠定基础.

2 不一致不完备决策系统

一个信息系统 (IS) 通常可以表示为四元组 $(U, A, \{V_a\}, f_a)_{a\in A}$. 其中: U 为非空有限对象集; A 为非空有限的属性集; V_a 为属性 a 的值域; f_a 为 U 到 V_a 的函数, 表示为 $f_a : U \to V_a$, 称为 IS 的信息函数. 若存在 $x \in U$ 且 $a \in A$, 使得 $f_a(x)$ 的值为一个丢失值 (空值、未知值或不确定值[9]), 则此信息系统称为不完备信息系统, 否则称为完备信息系统. 丢失值用"$*$"表示, 这样, 不完备信息系统即可表示为

$$(U, A, \{V_a\}, f_a)_{a\in A},\ * \in \bigcup_{a\in A} V_a.$$

若 V_a 和 f_a 是已知的, 则 $(U, A, \{V_a\}, f_a)_{a\in A}$ 也可以简写为 (U, A). 对于任意 $B \subseteq A$, 子集 B 决定如下二元关系:

$$\mathrm{TR}(B) =\{(x, y)|f_a(x) = f_a(y) \text{ or } f_a(x) = * \text{ or } f_a(y) = *, \text{for } \forall a \in B,\ \text{and } x, y \in U\}. \quad (1)$$

容易证明, TR(B)是自反、对称的, 因此是一个相容关系. 令 $S_B(x)$ 表示在相容关系 B 下所有与对象 x 不可分辨的对象的最大集, 即

$$S_B(x) = \{y|(x, y) \in \mathrm{TR}(B),\ y \in U\}. \quad (2)$$

令 $X \subseteq U$ 且 $B \subseteq A$, 则 X 的下近似定义为

$$\underline{\mathrm{B}}\mathrm{X} = \{x \in U|S_B(x) \subseteq X\}. \quad (3)$$

下近似也称为正域, 可以表示为 $\mathrm{POS}_B(X)$, 即 $\mathrm{POS}_B(X) = \underline{\mathrm{B}}\mathrm{X}$. 子集 X 的上近似定义为

$$\overline{\mathrm{B}}\mathrm{X} = \{x \in U|S_B(x) \bigcap X \neq \varnothing\}. \quad (4)$$

$\underline{\mathrm{B}}\mathrm{X}$ 中的对象肯定属于 X, $\overline{\mathrm{B}}\mathrm{X}$ 中的对象可能属于 X.

不完备决策系统是一种特殊的不完备信息系统, 它可以看作是对信息系统 (U, A) 中的属性集 A 增加一些属性而形成. 这样, 一个不完备决策系统也可以表示为一个四元组, 即

$$\big(U, C \bigcup D, \{P(V_d)\}, \partial_B\big)_{d\in D, B\subseteq C}. \quad (5)$$

其中: U 和 C 的意义同 IS, D 为非空有限的属性集, 称为决策属性集, 且 $C \bigcap D = \varnothing$; $P(V_d)$ 为 V_d 的幂集, V_d 为属性 $d \in D$ 的值域, 但 $* \notin V_d$; ∂_B 为函数 ∂_B: $U \to P(V_d), B \subseteq C$, 定义为 $\partial_B(x) = \{f_d(y)|y \in S_B(x)\}$, $f_d(y)$ 是不完备信息系统 $\big(U, A = C \bigcup D, \{V_d\}, f_a\big)_{a\in A}$ 的信息函数.

如果对于任意 $x \in U$, 均有 $|\partial_C(x)| = 1$, 则称不完备决策系统是一致的, 否则称其是不一致的[10], 相应的系统分别称为一致不完备决策系和不一致不完备决策系 (IIDS). 这样, 一个 IIDS 可以形式化表示为如下四元组:

$$\big(U, C \bigcup D, \{P(V_d)\}, \partial_B\big)_{d\in D, B\subseteq C},$$
$$\exists x_0 \in U,\ |\partial_C(x_0)| > 1. \quad (6)$$

在相关条件已明确的情况下, IIDS$(U, C \bigcup D, \{P(V_d)\}, \partial_B)_{d\in D, B\subseteq C}$ 也可以简写为 $(U, C \bigcup D)$, 即 IIDS $= (U, C \bigcup D)$.

根据以上描述, 在一个不完备、不一致决策系统中, 至少存在一个对象 $x_0 \in U$ 以及两个不同的决策类 D' 和 $D''(D' \neq D'')$, 使得 $S_C(x_0) \bigcap D' \neq \varnothing$ 且 $S_C(x_0) \bigcap D'' \neq \varnothing$.

3 IIDS 中约简的定义

为了表述方便, 如果不特别说明, 则下文中提到的"决策系统"指的是 IIDS, 并约定对于 IIDS $= (U, C \bigcup \{d\})$, $V_d = \{v_1, v_2, \cdots, v_r\}$, $X_{v_i} = \{x \in U|f_d(x) = v_i\}$, 有 $U/\{d\} = \{X_{v_1}, X_{v_2}, \cdots, X_{v_r}\}$. 下面将不一致完备系统中的约简概念[6-8]拓展到 IIDS 中, 并作适当的补充, 从而定义 IIDS 中约简的概念.

定义 1 在 IIDS $= (U, C \bigcup \{d\})$ 中, 若 $B \subseteq C$, 则称 B 为 IIDS 的一个可能约简, 当且仅当对于任意 $x \in U$, $S_B(x) \subseteq \overline{\mathrm{C}}\mathrm{X}_{\mathrm{v}}$, $v = f_d(x)$ 和 $B' \subset B$, 总存在 $x_0 \in U$, 使得 $S_{B'}(x_0) \nsubseteq \overline{\mathrm{C}}\mathrm{X}_{\mathrm{v_0}}$, $v_0 = f_d(x_0)$.

可能约简是具有如下特性的最小属性集: 能够将 x 区分于那些不属于 X_v 上近似中的对象, 其中 $v = f_d(x)$.

定义 2 称 $B \subseteq C$ 为 IIDS 的一个属性约简, 当且仅当对于 $x \in U$, 有 $S_B(x) = S_C(x)$, 且对于任意 $B' \subset B$, 总存在 $x_0 \in U$, 使得 $S_{B'}(x_0) \neq S_C(x_0)$.

定义 2 要求对于任意 $x \in U$, $S_B(x)$ 均不能改变 (总等于 $S_C(x)$). 属性约简是具有如下特性的最小属性集: 能将 x 区别于那些其邻域不等于 x 领域的对象.

定理 1 令 IIDS $= (U, C \bigcup \{d\})$, 有:

1) 令 $B', B'' \subseteq C$, 对于 $\forall x \in U$, 均有 $S_{B'}(x) = S_{B''}(x)$, 当且仅当 $\mathrm{TR}(B') = \mathrm{TR}(B'')$;

2) $\mathrm{TR}(B) = \bigcap_{a\in B} TR(\{a\}), B \subseteq C$;

3) 若 $B' \subseteq B''$, 则对于 $\forall x \in U$, 有 $S_{B'}(x) \supseteq S_{B''}(x)$.

证明 1) 充分性是显然的, 仅需证明必要性. 对

于 $\forall(y,z) \in \mathrm{TR}(B')$, 有 $z \in S_{B'}(y)$. 根据条件 $S_{B'}(y) = S_{B''}(y)$, 有 $z \in S_{B'}(y) = S_{B''}(y)$, 即 $z \in S_{B''}(y)$. 这意味着 $(y,z) \in \mathrm{TR}(B'')$, $\mathrm{TR}(B') \subseteq \mathrm{TR}(B'')$, 类似可证, $\mathrm{TR}(B') \supseteq \mathrm{TR}(B'')$, $\mathrm{TR}(B') = \mathrm{TR}(B'')$.

2) 证明过程见文献 [10].

3) 假设 $B' = \{a_1, a_2, \cdots, a_p\}$, 由于 $B' \subseteq B''$, 可假设 $B'' = \{a_1, a_2, \cdots, a_p, a_{p+1}, a_{p+2}, \cdots, a_q\}$. 根据 2), 有

$$
\begin{aligned}
&\mathrm{TR}(B'') = \bigcap_{a \in B''} \mathrm{TR}(\{a\}) = \\
&\mathrm{TR}(\{a_1\}) \bigcap \mathrm{TR}(\{a_2\}) \bigcap \cdots \bigcap \mathrm{TR}(\{a_p\}) \bigcap \\
&\mathrm{TR}(\{a_{p+1}\}) \bigcap \mathrm{TR}(\{a_{p+2}\}) \bigcap \cdots \bigcap \mathrm{TR}(\{a_q\}) \subseteq \\
&\mathrm{TR}(\{a_1\}) \bigcap \mathrm{TR}(\{a_2\}) \bigcap \cdots \bigcap \mathrm{TR}(\{a_p\}) = \\
&\bigcap_{a \in B'} \mathrm{TR}(\{a\}) = \mathrm{TR}(B'),
\end{aligned}
$$

即 $\mathrm{TR}(B'') \subseteq \mathrm{TR}(B')$. 因此, 对于 $\forall y \in S_{B''}(x)$, 有 $(x, y) \in \mathrm{TR}(B'') \subseteq \mathrm{TR}(B')$. 这说明 $y \in S_{B'}(x)$, 从而有 $S_{B''}(x) \subseteq S_{B'}(x)$. □

定义 3　在 $\mathrm{IIDS}=(U, C \bigcup \{d\})$ 中, $U/\{d\}=\{X_{v_1}, X_{v_2}, \cdots, X_{v_r}\}$, 令

$$
\begin{aligned}
&\boldsymbol{\mu}_B(x) = (\mu^B_{v_1}(x), \mu^B_{v_2}(x), \cdots, \mu^B_{v_r}(x)), \\
&\mu^B_{v_i}(x) = \left|X_{v_i} \bigcap S_B(x)\right| / |S_B(x)|, \\
&\qquad B \subseteq C,\ i = 1, 2, \cdots, r.
\end{aligned}
$$

则称 B 为 IIDS 的一个分布约简, 当且仅当对于 $\forall x \in U$, 有 $\boldsymbol{\mu}_B(x) = \boldsymbol{\mu}_C(x)$, 且对于 $\forall B' \subset B$, 总存在 $x_0 \in U$, 使得 $\boldsymbol{\mu}_{B'}(x_0) \neq \boldsymbol{\mu}_C(x_0)$.

显然, $\mu_{v_i}(x)$ 是一种隶属度函数, 其值表示对象 x 隶属于决策类 $X_{v_i} (i = 1, 2, \cdots, r)$ 的程度. $\boldsymbol{\mu}_B(x) = (\mu_{v_1}(x), \mu_{v_2}(x), \cdots, \mu_{v_r}(x))$ 是 $U/\{d\}$ 上的一个概率分布函数, 其中 $\mu_{v_1}(x) + \mu_{v_2}(x) + \cdots + \mu_{v_r}(x) = 1$.

定义 4　令 $\theta_B(x) = \{X_v \| X_v \bigcap S_B(x)|/|S_B(x)| > 0, X_v \in U/\{d\}\}$, 则称 $B \subseteq C$ 为 IIDS 的一个分配约简, 当且仅当对于任意 $x \in U$, 有 $\theta_B(x) = \theta_C(x)$, 且对于任意 $B' \subset B$, 总存在 $x_0 \in U$, 使得 $\theta_{B'}(x_0) \neq \theta_C(x_0)$.

定理 2　对于任意 $x \in U$, 存在 $\theta_B(x) = \theta_C(x)$, 当且仅当 $\partial_B(x) = \partial_C(x)$.

证明　首先证明 $\partial_B(x) = \partial_C(x)$. 由 $B \subseteq C$, 易知 $\partial_B(x) \supseteq \partial_C(x)$, 以下只需证明 $\partial_B(x) \subseteq \partial_C(x)$. 对于任意 $v \in \partial_B(x)$, 总存在 $y \in S_B(x)$, 使得 $v = f_d(y)$, 这意味着 $y \in X_v$. 因此, $y \in X_v \bigcap S_B(x) \neq \varnothing$, 从而有 $|X_v \bigcap S_B(x)|/|S_B(x)|>0$, 即 $X_v \in \theta_B(x)$. 因为 $\theta_B(x) = \theta_C(x)$, 所以 $X_v \in \theta_C(x)$, 且 $|X_v \bigcap S_C(x)|/|S_C(x)| > 0$, 这表明 $X_v \bigcap S_C(x) \neq \varnothing$. 假设 $y' \in X_v \bigcap S_C(x)$, 则有 $y' \in S_C(x)$ 且 $f_d(y') = v$, 这说明 $v \in \{f_d(y)|y \in S_C(x)\} = \partial_C(x)$. 进而有 $\partial_B(x) \subseteq \partial_C(x)$.

下面证明 $\theta_B(x) = \theta_C(x)$. 由于 $B \subseteq C$, 显然有 $\theta_B(x) \supseteq \theta_C(x)$, 以下只需证明 $\theta_B(x) \subseteq \theta_C(x)$. 对于任意 $X_v \in \theta_B(x)$, 有 $|X_v \bigcap S_B(x)|/|S_B(x)| > 0$, 这意味着 $X_v \bigcap S_B(x) \neq \varnothing$. 这样, 可以假设 $z \in X_v \bigcap S_B(x)$, 则有 $v = f_d(z) \in \{f_d(y)|y \in S_B(x)\} = \partial_B(x) = \partial_C(x)$, 即 $v \in \partial_C(x)$. 这说明, 存在 $y_0 \in S_C(x)$, 使得 $f_d(y_0) = v$, 即 $y_0 \in X_v$. 因此, $y_0 \in X_v \bigcap S_C(x) \neq \varnothing$, 说明 $|X_v \bigcap S_C(x)|/|S_C(x)| > 0$, 即有 $X_v \in \theta_C(x)$, 进而有 $\theta_B(x) \subseteq \theta_C(x)$. □

由定理 2 可以进一步推出下列结论.

定理 2'　在 IIDS 中, 分配约简和广义决策约简是等价的.

根据定理 2', 分配约简亦可定义如下.

定义 4'　称 $B \subseteq C$ 为 IIDS 的一个分配约简, 当且仅当对于任意 $x \in U, \partial_B(x) = \partial_C(x)$ 和 $B' \subset B$, 总存在 $x_0 \in U$, 使得 $\partial_{B'}(x_0) \neq \partial_C(x_0)$.

定义 5　令

$$
\begin{aligned}
\gamma_B(x) = &\Big\{X_v \Big| \left|X_v \bigcap S_B(x)\right| / |S_B(x)| = \\
&\max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_B(x)\right| / |S_B(x)|\right)\Big\},
\end{aligned}
$$

则称 $B \subseteq C$ 为 IIDS 的一个最大分布约简, 当且仅当对于任意 $x \in U$, 有 $\gamma_B(x) = \gamma_C(x)$, 且对于任意 $B' \subset B$, 总存在 $x_0 \in U$, 使得 $\gamma_{B'}(x_0) \neq \gamma_C(x_0)$.

4　IIDS 中属性约简之间的关系

定理 3　假设 PR 是 IIDS 的所有可能约简的集合, AR 是分配约简的集合, 则 $\mathrm{PR} \subseteq \mathrm{AR}$.

证明　对于任意 $B \in \mathrm{PR}$ 和 $v \in \partial_B(x) (x \in U)$, 总存在 $y_1 \in S_B(x)$, 使得 $v = f_d(y_1)$. 由于 $B \subseteq C$ 是 IIDS 的一个可能分布约简, 由定义 1 可知, $S_B(x) \subseteq \overline{\mathrm{CX}}_v$. 于是有 $y_1 \in \overline{\mathrm{CX}}_v$, 这说明 $S_C(x) \bigcap X_v \neq \varnothing$. 假设 $y_2 \in S_C(x) \bigcap X_v$, 则 $y_2 \in S_C(x)$ 且 $y_2 \in X_v$. 这表明 $v = f_d(y_2) \in \{f_d(y)|y \in S_C(x)\} = \partial_C(x)$, $\partial_B(x) \subseteq \partial_C(x)$. 由于 $B \subseteq C$, $\partial_B(x) \supseteq \partial_C(x)$, 从而 $\partial_B(x) = \partial_C(x)$.

根据定义 1, 对于任意 $B' \subset B$, 必存在 $x_0 \in U$, 使得 $S_{B'}(x_0) \not\subseteq \overline{\mathrm{CX}}_{v_0}$, 其中 $v_0 = f_d(x_0)$. 这说明存在 $x_1 \in S_{B'}(x_0)$, 使得 $x_1 \notin \overline{\mathrm{CX}}_{v_0}$. 一方面, 由 $x_1 \in S_{B'}(x_0)$ 有 $x_0 \in S_{B'}(x_1)$, 从而 $v_0 = f_d(x_0) \in f_d(S_{B'}(x_1)) = \partial_{B'}(x_1)$, 即 $v_0 \in \partial_{B'}(x_1)$; 另一方面, 由 $x_1 \notin \overline{\mathrm{CX}}_{v_0}$ 有 $S_C(x_1) \bigcap X_{v_0} = \varnothing$, 这说明 $v_0 \notin f_d(S_C(x_1)) = \partial_C(x_1)$. 由于 $v_0 \in \partial_{B'}(x_1)$ 且 $v_0 \notin \partial_C(x_1)$, $\partial_{B'}(x_1) \neq \partial_C(x_1)$, 即存在 $v_0 = f_d(x_0)$ 和 x_1, 使得 $\partial_{B'}(x_1) \neq \partial_C(x_1)$. 根据定义 4', B 是 IIDS 的一个分配约简, 即 B

$\in$ AR, 因此 PR $\subseteq$ AR. □

定理 4 令 $B \subseteq C$ 为 IIDS 的一个属性约简, 则对于任意 $x \in U$, 有:

1) $S_B(x) \subseteq \overline{\mathrm{CX}}_v$, 其中 $v = f_d(x)$;

2) $S_B(x) = S_C(x)$;

3) $\boldsymbol{\mu}_B(x) = \boldsymbol{\mu}_C(x)$;

4) $\theta_B(x) = \theta_C(x), \partial_B(x) = \partial_C(x)$;

5) $\gamma_B(x) = \gamma_C(x)$.

证明 仅需证明 3), 其他证明与前文类似. 由定义 2 可知, 对于任意 $x \in U$, 有 $S_B(x) = S_C(x)$, 因此有

$$\boldsymbol{\mu}_B(x) = (\mu^B_{v_1}(x), \mu^B_{v_2}(x), \cdots, \mu^B_{v_r}(x)),$$
$$\mu^B_{v_i}(x) = |X_{v_i} \bigcap S_B(x)|/|S_B(x)| = |X_{v_i} \bigcap S_C(x)|/|S_C(x)| = \mu^C_{v_i}(x),$$
$$i = 1, 2, \cdots, r.$$

这表明 $\boldsymbol{\mu}_C(x) = \boldsymbol{\mu}_B(x)$. □

定理 4 说明, 任意属性约简均是可能约简、分布约简、分配约简和最大分布约简的一个超集, 即每一个属性约简至少包含一个可能约简、分布约简、分配约简和最大分布约简, 反之不成立. 但属性约简不一定是可能约简、分布约简、分配约简或最大分布约简, 它只是包含后者而已.

定理 5 假设 $B \subseteq C$ 是 IIDS 的一个分布约简, 则存在 IIDS 的一个最大分布约简 B', 使得 $B \supseteq B'$.

证明 如果 $\gamma_B(x) = \gamma_C(x)$, 则容易验证必存在 IIDS 的一个最大分布约简 B', 使得 $B \supseteq B'$. 因此, 只需证明 $\gamma_B(x) = \gamma_C(x)$.

因为 B 是一个分布约简, $\boldsymbol{\mu}_B(x) = \boldsymbol{\mu}_C(x)$, 即

$$(\mu^B_{v_1}(x), \mu^B_{v_2}(x), \cdots, \mu^B_{v_r}(x)) = (\mu^C_{v_1}(x), \mu^C_{v_2}(x), \cdots, \mu^C_{v_r}(x)).$$

所以有

$$\max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_B(x)\right|/|S_B(x)|\right) = \max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_C(x)\right|/|S_C(x)|\right).$$

对于任意 $X_v \in \gamma_B(x)$, 必存在 $X_{v_i} = X_v (i \in \{1, 2, \cdots, r\})$, 使得

$$\left|X_{v_i} \bigcap S_B(x)\right|/|S_B(x)| = \max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_B(x)\right|/|S_B(x)|\right) = \max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_C(x)\right|/|S_C(x)|\right).$$

由此可以得到

$$|X_{v_i} \bigcap S_C(x)|/|S_C(x)| = \mu^C_{v_i}(x) = \mu^B_{v_i}(x) = \left|X_{v_i} \bigcap S_B(x)\right|/|S_B(x)| = \max_{v' \in U/\{d\}} \left(\left|X_{v'} \bigcap S_C(x)\right|/|S_C(x)|\right). \tag{7}$$

这表明 $X_v = X_{v_i} \in \gamma_C(x)$, 从而 $X_v \in \gamma_C(x)$, 于是有 $\gamma_B(x) \subseteq \gamma_C(x)$. 类似可证 $\gamma_B(x) \supseteq \gamma_C(x)$, 从而有 $\gamma_B(x) = \gamma_C(x)$. □

定理 5 说明, 每一个分布约简至少包含一个最大分布约简, 反之不成立.

定理 6 设 $B \subseteq C$ 是 IIDS 的一个分布约简, 则必存在 IIDS 的一个分配约简 B', 使得 $B \supseteq B'$.

证明 类似于定理 5 的证明, 只需证明 $\theta_B(x) = \theta_C(x)$.

对于任意 $X_{v_i} \in \theta_B(x)$, 由分配约简的定义可知, $|X_{v'} \bigcap S_B(x)|/|S_B(x)| > 0$. 由于 B 是一个分布约简, 有 $\boldsymbol{\mu}_B(x) = \boldsymbol{\mu}_C(x)$. 于是有

$$\left|X_{v'} \bigcap S_C(x)\right|/|S_C(x)| = \mu^C_{v_i}(x) = \mu^B_{v_i}(x) = \left|X_{v'} \bigcap S_B(x)\right|/|S_B(x)| > 0.$$

这说明 $X_{v_i} \in \theta_C(x)$, 从而有 $\theta_B(x) \supseteq \theta_C(x)$. 用类似方法可以证明 $\theta_B(x) \subseteq \theta_C(x)$, 进而有 $\theta_B(x) = \theta_C(x)$. □

由定理 6 可以发现, 分配约简类似于最大分布约简, 它们都是相关分布约简的子集, 但实际上最大分布约简与分配约简之间不存在包含关系.

综上所述, 定理 1∼ 定理 6 说明了各个约简之间的关系, 可以用图 1 表示. 图 1 表明: 1) 分配约简与广义决策约简是等价的; 2) 所有其他约简必为某一个属性约简的子集; 3) 任意一个最大分布约简和分配约简必为相应分布约简的子集; 4) 所有可能约简的集合是所有分配约简集合的一个子集, 即可能约简必定是分配约简. 图 1 中各箭头符号的意义如下: A⇔B 表示 A 和 B 等价; A⇒B 表示 {A}⊆{B}; B → A 表示 A⊇B.

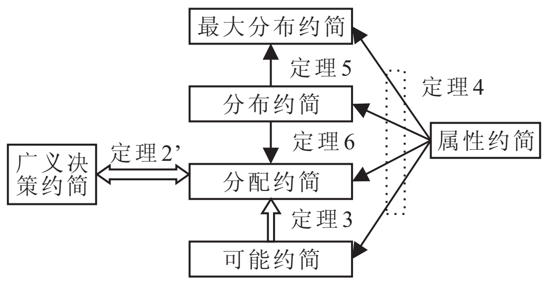

图 1 各种约简之间的关系

5 实例分析

考虑不完备决策系统 $(U, C \bigcup \{d\})$, 如表 1 所示. 其中: $U = \{1, 2, \cdots, 6\}$; $C = \{P, M, S, X\}$ = {Price, Mileage, Size, Max-speed}; $\{d\}$={Acceleration}. 分别计算 $S_C(x)$ 和 $\partial_C(x)$, 结果如表 2 所示. 由表 2 可见, $|\partial_C(4)| = |\partial_C(5)| = |\partial_C(6)| = 2 > 1$, 因此该系统是一个 IIDS.

表 1　不一致不完备决策系统

U	P	M	S	M	d
1	high	high	full	Low	good
2	low	*	full	Low	good
3	*	*	compact	high	poor
4	low	*	full	high	good
5	*	*	full	high	excel
6	low	high	full	*	good

表 2　$S_C(x)$ 和 $\partial_C(x)$ 的值

x	$S_C(x)$	$\partial_C(x)$
1	$\{1\}$	$\{\text{good}\}$
2	$\{2,6\}$	$\{\text{good}\}$
3	$\{3\}$	$\{\text{poor}\}$
4	$\{4,5,6\}$	$\{\text{good},\text{excel}\}$
5	$\{4,5,6\}$	$\{\text{good},\text{excel}\}$
6	$\{2,4,5,6\}$	$\{\text{good},\text{excel}\}$

5.1　计算属性约简

对于表 1 所示的 IIDS $=(U,C\bigcup\{d\})$, 分别计算 $S_{\{P\}}(x)$, $S_{\{S\}}(x)$, $S_{\{X\}}(x)$, $S_{\{P,S\}}(x)$, $S_{\{P,X\}}(x)$, $S_{\{S,X\}}(x)$ 和 $S_{\{P,S,X\}}(x)$. 可以发现, 对于任意 $x\in U$, 有 $S_{\{P,S,X\}}(x)=S_C(x)$, 而对于所有 $B\subset\{P,S,X\}$, 均存在 $x_0\in U$, 使得 $S_B(x)\neq S_C(x)$, 这说明 $\{P,S,X\}$ 是 IIDS 的属性约简.

5.2　计算可能约简

经计算可得 $X_{\text{good}}=\{1,2,4,6\}$, $X_{\text{poor}}=\{3\}$, $X_{\text{excel}}=\{5\}$. 由表 2 可知, $\overline{\text{CX}}_{\text{good}}=\{1,2,4,5,6\}$, $\overline{\text{CX}}_{\text{poor}}=\{3\}$, $\overline{\text{CX}}_{\text{excel}}=\{4,5,6\}$. $S_{\{S,X\}}(x)$, $S_{\{S\}}(x)$ 和 $S_{\{X\}}(x)$ 的计算结果如表 3 所示. 由表 3 可见, 对于所有 $x\in U$, 有 $S_{\{S,X\}}(x)\subseteq\overline{\text{CX}}_{f_d(x)}$. 但 $S_{\{S\}}(5)$ 不属于 $\overline{\text{CX}}_{f_d(5)}$, $S_{\{X\}}(y)$ 不属于 $\overline{\text{CX}}_{f_d(y)}(y=3,4,5,6)$. 由定义 1 可知, $\{S,X\}$ 是此 IIDS 的可能约简.

表 3　$S_{\{S,X\}}(x)$, $S_{\{S\}}(x)$ 和 $S_{\{X\}}(x)$ 的值

x	$S_{\{S,X\}}(x)$	$S_{\{S\}}(x)$	$S_{\{X\}}(x)$	$\overline{\text{CX}}_{f_d(x)}$
1	$\{1,2,6\}$	$\{1,2,4,5,6\}$	$\{1,2,6\}$	$\{1,2,4,5,6\}$
2	$\{1,2,6\}$	$\{1,2,4,5,6\}$	$\{1,2,6\}$	$\{1,2,4,5,6\}$
3	$\{3\}$	$\{3\}$	$\{3,4,5,6\}$	$\{3\}$
4	$\{4,5,6\}$	$\{1,2,4,5,6\}$	$\{3,4,5,6\}$	$\{1,2,4,5,6\}$
5	$\{4,5,6\}$	$\{1,2,4,5,6\}$	$\{3,4,5,6\}$	$\{4,5,6\}$
6	$\{1,2,4,5,6\}$	$\{1,2,4,5,6\}$	$\{1,2,3,4,5,6\}$	$\{1,2,4,5,6\}$

经过类似计算可以发现, 其他属性子集 (如 $\{P,X\}$) 均不是此 IIDS 的可能约简.

5.3　计算分布约简

容易观察到, $U/\{d\}=\{X_{\text{good}},X_{\text{poor}},X_{\text{excel}}\}$. 其中: $X_{\text{good}}=\{1,2,4,6\}$, $X_{\text{poor}}=\{3\}$, $X_{\text{excel}}=\{5\}$. 利用表 2 可以计算所有的 $\boldsymbol{\mu}_C(x)$, 进而可以计算 $\theta_C(x)$ 和 $\gamma_C(x)$, 结果如表 4 所示. 对于任意 $x\in U$, 有

$\boldsymbol{\mu}_{\{S,X\}}(x)=\boldsymbol{\mu}_C(x)$;

$\boldsymbol{\mu}_{\{S\}}(x)=(4/5,0,1/5)$, $x=1,2,4,5,6$;

$\boldsymbol{\mu}_{\{S\}}(3)=(0,1,0)$; $\boldsymbol{\mu}_{\{X\}}(1)=\boldsymbol{\mu}_{\{X\}}(2)=(1,0,0)$;

$\boldsymbol{\mu}_{\{X\}}(3)=\boldsymbol{\mu}_{\{X\}}(4)=\boldsymbol{\mu}_{\{X\}}(5)=(2/4,1/4,1/4)$;

$\boldsymbol{\mu}_{\{X\}}(6)=(4/6,1/6,1/6)$.

表 4　$\boldsymbol{\mu}_C(\boldsymbol{x}),\boldsymbol{\partial}_C(\boldsymbol{x}),\boldsymbol{\theta}_C(\boldsymbol{x})$ 和 $\boldsymbol{\gamma}_C(\boldsymbol{x})$ 的值

x	$\boldsymbol{\mu}_C(x)$	$\partial_C(x)$	$\theta_C(x)$	$\gamma_C(x)$
1	$(1,0,0)$	good	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
2	$(1,0,0)$	good	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
3	$(0,1,0)$	poor	$\{X_{\text{poor}}\}$	$\{X_{\text{poor}}\}$
4	$(2/3,0,1/3)$	good,excel	$\{X_{\text{good}},X_{\text{excel}}\}$	$\{X_{\text{good}}\}$
5	$(2/3,0,1/3)$	good,excel	$\{X_{\text{good}},X_{\text{excel}}\}$	$\{X_{\text{good}}\}$
6	$(3/4,0,1/4)$	good,excel	$\{X_{\text{good}},X_{\text{excel}}\}$	$\{X_{\text{good}}\}$

与表 4 中的数据比较并根据定义 3 可以发现, $\{S,X\}$ 是 IIDS 的分布约简.

5.4　计算分配约简

对于任意 $x\in U$, 分别计算 $S_{\{P\}}(x)$, $S_{\{S\}}(x)$, $S_{\{X\}}(x)$, $S_{\{P,S\}}(x)$, $S_{\{P,X\}}(x)$ 和 $S_{\{S,X\}}(x)$, 并利用表 2 可以求出所有的 $\theta_B(x)$. 由 $\theta_B(x)$ 的值并根据定义 4 可知, $\{S,X\}$ 是分配约简, 类似地可以计算出其他属性子集均不是分配约简.

5.5　计算最大分布约简

利用已有的 $S_B(x)$ 计算 $\gamma_B(x)$, 结果如表 5 所示. 由表 5 并根据定义 5 可知, $\{S\}$ 是最大分布约简.

表 5　$\boldsymbol{\gamma}_B(\boldsymbol{x})$ 的值 $(\boldsymbol{B}\subseteq\boldsymbol{C})$

x	$\gamma_C(x)$	$\gamma_{\{S,X\}}(x)$	$\gamma_{\{S\}}(x)$	$\gamma_{\{X\}}(x)$
1	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
2	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
3	$\{X_{\text{poor}}\}$	$\{X_{\text{poor}}\}$	$\{X_{\text{poor}}\}$	$\{X_{\text{good}}\}$
4	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
5	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$
6	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$	$\{X_{\text{good}}\}$

5.6　计算广义决策约简

类似地, 利用 $S_B(x)$ 计算 $\partial_B(x)$, 由 $\partial_B(x)$ 的值并根据文献 [6] 可知, $\{S,X\}$ 是广义决策约简.

由上述计算结果可以发现, 属性约简 $\{P,S,X\}$ 是所有其他约简的超集, 分配约简 $\{S,X\}$ 与广义决策约简 $\{S,X\}$ 是相等的, 最大分布约简 $\{S\}$ 和分配约简 $\{S,X\}$ 均包含于分布约简 $\{S,X\}$, 可能约简集合 $\{S,X\}$ 确实是分配约简集合 $\{\{S,X\}\}$ 的子集. 以上结果与图 1 所示的结论一致, 这表明本文结论是有效的.

6　结　　论

IIDS 对于复杂数据具有较强的数据建模和表示能力, 使其在当今复杂的数据环境下具有重要的实

际应用价值. 本文将不一致完备决策系统中的约简概念拓展到IIDS中, 并作适当的补充, 从而成功定义了IIDS中约简的概念, 并找到它们之间的关联. 结论如下: 1) 分配约简与广义决策约简是等价的; 2) 所有其他约简必为某一个属性约简的子集; 3) 任意一个最大分布约简和分配约简必为相应分布约简的子集; 4) 可能约简必定是分配约简.

下一步的工作将充分利用这些约简之间的关系研究IIDS中约简的相关理论, 提出面向海量数据的各种约简的高效计算方法, 并将其应用于解决实际问题.

参考文献(References)

[1] 蒙祖强, 史忠植. 粒度世界拓扑结构的理论研究[J]. 控制与决策, 2007, 22(9): 1017-1021.
(Meng Z Q, Shi Z Z. Theoretic research on topological structure of granular worlds[J]. Control and Decision, 2007, 22(9): 1017-1021.)

[2] 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.)

[3] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为 $\max(O(|C||U|), O(|C|^2|U/C|))$ 的快速属性约简算法[J]. 计算机学报, 2006, 29(3): 391-399.
(Xu Z Y, Liu Z P, Yang B R, et al. A quick attribute reduction algorithm with complexity of $\max(O(|C||U|), O(|C|^2|U/C|))$[J]. Chinese J of Computers, 2006, 29(3): 391-399.)

[4] 胡峰, 王国胤. 属性序下的快速约简算法[J]. 计算机学报, 2007, 30(8): 1429-1435.
(Hu F, Wang G Y. Quick reduction algorithm based on attribute order[J]. Chinese J of Computers, 2007, 30(8): 1429-1435.)

[5] Meng Z Q, Shi Z Z. A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets[J]. Information Sciences, 2009, 179: 2774-2793.

[6] Kryszkiewicz M. Comparative study of alternative types of knowledge reduction in inconsistent systems[J]. Int J of Intelligent Systems, 2001, 16: 105-120.

[7] Zhang W X, Mi J S, Wu W Z. Approaches to knowledge reduction in inconsistent information systems[J]. Int J of Intelligent Systems, 2003, 18: 989-1000.

[8] 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.
(Zhang W X, Mi J S, Wu W Z. Knowledge reduction in inconsistent information systems[J]. Chinese J of Computers, 2003, 26(1): 12-18.)

[9] Wu W Z. Attribute reduction based on evidence theory in incomplete decision systems[J]. Information Sciences, 2008, 178: 220-255.

[10] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112: 39-49.

下 期 要 目