第 32 卷 第 22 期 计 算 机 工 程 2006 年 11 月
Vol.32 No.22 Computer Engineering November 2006
·博士论文· 文章编号：1000—3428(2006)22—0026—02 文献标识码：A 中图分类号：TP311

集值信息系统的知识约简与属性特征

宋笑雪[1,2]，李鸿儒[1]，张文修[1]
(1. 西安交通大学理学院信息与系统科学研究所，西安 710049；2. 咸阳师范学院计算机系，咸阳 712000)

摘 要：定义了集值信息系统中的一种新的关系，讨论了在这种关系下集值信息系统的知识约简问题，给出了集值信息系统属性约简的判定定理和辨识矩阵，得到了计算约简的具体方法；讨论了集值信息系统中 3 种不同类型的属性特征及每一种属性的判定定理。
关键词：集值信息系统；知识约简；可辨识矩阵；属性特征

Knowledge Reduction and Attributes Characteristics in Set-valued Information System

SONG Xiaoxue[1,2], LI Hongru[1], ZHANG Wenxiu[1]
(1. Institute of Information and System Science, Faculty of Science, Xian Jiaotong University, Xi'an 710049;
2. Department of Computer, Xianyang Normal College, Xianyang 712000)

【Abstract】 A new relation is defined in set-valued information system, based on the relation the problem of knowledge reduction in set-valued information system is discussed, the judgment theorems and the discernibility matrices of knowledge reduction are then given in set-valued information system, from which a new approach to knowledge reductions can be provided. The characteristics of different type of attributes in set-valued information system are studied, and the judgment theorems about them are obtained.
【Key words】 Set-valued information system; Knowledge reduction; Discernibility matrices; Attributes characteristics

糙集理论是由波兰数学家Z.Pawlak于 1982 年提出的用于数据分析的理论[1]。由于粗糙集理论能够分析处理不精确、不完备信息，因此作为一种具有极大潜力和有效的知识获取工具受到了广大研究者的关注。粗糙集理论已被成功地应用在机器学习与知识发现、数据挖掘、决策支持与分析、模式识别等领域[4~8]。

在实际情况中，由于问题的复杂性，要保证每个对象的所有属性值的完整性往往是非常困难的。在不确定信息和缺省信息，即不完备信息的情况下，需要研究集值信息系统[4]。

知识约简是粗糙集理论的核心问题。众所周知，知识库中描述知识的属性并不是同等重要的，甚至其中某些属性是冗余的。所谓知识约简就是在保持知识库分类能力不变条件下，删除其中不相关或不重要的属性[2,3]。本文定义了集值信息系统中的一种新的关系，讨论了在这种关系下集值信息系统的知识约简问题，给出了集值信息系统属性约简的判定定理和辨识矩阵，从而得到了集值信息系统知识约简的具体操作方法。

另外，由于每一种属性在集值信息系统的约简中所起的作用不同[6]，因此本文还讨论了在集值信息系统的约简中起不同作用的属性，按照属性在约简中的作用将属性分为 3 类，并讨论了 3 种不同类型的属性特征及每种属性的判定定理。

1 集值信息系统

定义 1 称 (U,A,F) 是集值信息系统，若 $U=\{x_1,...,x_n\}$ 为对象集，每个 $x_i\,(i\le n)$ 称为一个对象；$A=\{a_1,...,a_m\}$ 为属性集，每个 $a_j\,(j\le m)$ 称为一个属性；$F=\{f_l:l\le m\}$ 为对象属性值映射，其中 $f_l:U\to \mathrm{P}_0(V_l)(l\le m)$，$V_l$ 是属性 a_l 的值域，$\mathrm{P}_0(V_l)$ 表示 V_l 的非空子集全体。

设 (U,A,F) 是一个集值信息系统，任意属性子集 $B\subseteq A$，定义二元关系 $R_B^*=\{(x,y)\in U\times U:f_l(x)\subseteq f_l(y)\ (\forall a_l\in B)\}$，并记

$$[x]_{B^*}=\{y\in U:(x,y)\in R_B^*\}=\{y\in U:f_l(x)\subseteq f_l(y)\ (\forall a_l\in B)\}$$

容易证明以下性质成立：

(1) R_B^* 是自反和传递的，未必是对称的，因此一般不是等价关系。

(2)当 $B_1\subseteq B_2\subseteq A$ 时，$R_{B_1}^*\supseteq R_{B_2}^*\supseteq R_A^*$。

(3)当 $B_1\subseteq B_2\subseteq A$ 时，$[x]_{B_1^*}\supseteq[x]_{B_2^*}\supseteq[x]_{A^*}$。

(4) $\{J=\{[x]_{B^*}:x\in U\}$ 是 U 的一个覆盖。

(5)当 $y\in[x]_{B^*}$ 时，$[y]_{B^*}\subseteq[x]_{B^*}$。

定义 2 设 (U,A,F) 是集值信息系统，对于任意 $X\subseteq U$，记

$$\underline{R_B^*}(X)=\{x\in U:[x]_{B^*}\subseteq X\},$$

$$\overline{R_B^*}(X)=\{x\in U:[x]_{B^*}\cap X\neq\phi\},$$

则 $\underline{R_B^*}(X)$ 和 $\overline{R_B^*}(X)$ 分别称为 X 在关系 R_B^* 下关于属性集 B 的下近似和上近似。

设 (U,A,F) 是一个集值信息系统，若 $B\subseteq A$，则对于任意 $X,Y\subseteq U$，以下性质成立：

(1) $\underline{R_B^*}(X)\subseteq X\subseteq\overline{R_B^*}(X)$；

基金项目：国家“973”计划基金资助项目(2002CB312200)；咸阳师范学院专项科研基金资助项目(04XSYK225)
作者简介：宋笑雪(1967－)，女，副教授、博士生，主研方向：粗糙集，模糊集，人工智能；李鸿儒，副教授、博士生；张文修，教授、博导
收稿日期：2005-12-01 **E-mail：**songxiaoxue@stu.xjtu.edu.cn

(2) $\underline{R_B^*}(X)=\sim\overline{R_B^*}(\sim X)$，$\overline{R_B^*}(X)=\sim\underline{R_B^*}(\sim X)$；

(3) $\underline{R_B^*}(\phi)=\phi=\overline{R_B^*}(\phi)$，$\underline{R_B^*}(U)=U=\overline{R_B^*}(U)$；

(4) $\underline{R_B^*}(X\cap Y)=\underline{R_B^*}(X)\cap\underline{R_B^*}(Y)$，$\overline{R_B^*}(X\cup Y)=\overline{R_B^*}(X)\cup\overline{R_B^*}(Y)$；

(5) $\underline{R_B^*}(X\cup Y)\supseteq\underline{R_B^*}(X)\cap\underline{R_B^*}(Y)$，$\overline{R_B^*}(X\cap Y)\subseteq\overline{R_B^*}(X)\cap\overline{R_B^*}(Y)$；

(6) $X\subseteq Y\Rightarrow\underline{R_B^*}(X)\subseteq\underline{R_B^*}(Y)$，$\overline{R_B^*}(X)\subseteq\overline{R_B^*}(Y)$；

(7) $\underline{R_B^*}(X)=\underline{R_B^*}(\underline{R_B^*}(X))$，$\overline{R_B^*}(X)=\overline{R_B^*}(\overline{R_B^*}(X))$；

证明：直接由定义即可得证。

例 1　集值信息系统(表 1)

表 1 集值信息系统

	a1	a2	a3	a4
x1	{1}	{1}	{2}	{2}
x2	{1}	{2}	{1}	{2}
x3	{1,2}	{1,2}	{2}	{1}
x4	{1}	{1,2}	{1}	{1}
x5	{1,2}	{1,2}	{1}	{1}
x6	{2}	{1}	{1}	{1,2}

由定义可以得到

$[x_1]_{A^*}=\{x_1\}$，$[x_2]_{A^*}=\{x_2\}$，$[x_3]_{A^*}=\{x_3\}$，

$[x_4]_{A^*}=\{x_4,x_5\}$，$[x_5]_{A^*}=\{x_5\}$，$[x_6]_{A^*}=\{x_6\}$。

取 $X=\{x_2,x_3,x_4\}$，则 $\underline{R_A^*}(X)=\{x_2,x_3\}$，$\overline{R_A^*}(X)=\{x_2,x_3,x_4\}$。

2 集值信息系统的知识约简

定义 3　称 $B\subseteq A$ 是集值信息系统 (U,A,F) 的协调集，若满足 $R_B^*=R_A^*$。若进一步对任意 $b\in B$，$R_{B-\{b\}}^*\neq R_A^*$，称 B 是集值信息系统 (U,A,F) 的约简。

记 $D(x_i,x_j)=\{a_l\in A: f_l(x_i)\subsetneq f_l(x_j)\}\quad(x_i,x_j\in U)$，称 $D(x_i,x_j)$ 为集值信息系统 (U,A,F) 在关系 R_A^* 下的辨识属性集，$D=\{D(x_i,x_j)\ ;(x_i,x_j\in U)\}$为集值信息系统 (U,A,F) 在关系 R_A^* 下的辨识矩阵。记 $D_0=\{D(x_i,x_j)\neq\phi;\ (x_i,x_j\in U)\}$。

定理 1　设 $D(x_i,x_j)$ 为集值信息系统 (U,A,F) 在关系 R_A^* 下的辨识属性集，$B\subseteq A$，则下列命题等价：

(1) B 是协调集；

(2) $\forall D(x_i,x_j)\neq\phi$，有 $B\cap D(x_i,x_j)\neq\phi$；

(3) $\forall B'\subseteq A$，若 $B'\cap B=\phi$，则 $B'\notin D_0$；

证明：(1) $\Leftrightarrow$ (2)：

B 是协调集 $\Leftrightarrow R_B^*\subseteq R_A^*$

$\Leftrightarrow (x_i,x_j)\notin R_A^*$ 时，有 $(x_i,x_j)\notin R_B^*$；

$\Leftrightarrow D(x_i,x_j)\neq\phi$ 时，有 $(x_i,x_j)\notin R_B^*$，即 $\exists a\in B$，$a\in D(x_i,x_j)$；

$\Leftrightarrow D(x_i,x_j)\neq\phi$ 时，$B\cap D(x_i,x_j)\neq\phi$；

(3)由(1)、(2)即得。

定理 1 说明了利用辨识矩阵可以计算集值信息系统的属性约简。

例 2　表 1 集值信息系统在关系 R_A^* 下的辨识矩阵

$$D=\begin{pmatrix}
\phi & \{a2,a3\} & \{a4\} & \{a3,a4\} & \{a3,a4\} & \{a1,a3\}\\
\{a2,a3\} & \phi & \{a3,a4\} & \{a4\} & \{a4\} & \{a1,a2\}\\
\{a1,a2,a4\} & \{a1,a2,a3,a4\} & \phi & \{a1,a3\} & \{a3\} & \{a1,a2,a3\}\\
\{a2,a3,a4\} & \{a2,a4\} & \{a3\} & \phi & \phi & \{a1,a2\}\\
\{a1,a2,a3,a4\} & \{a1,a2,a4\} & \{a3\} & \{a1\} & \phi & \{a1,a2\}\\
\{a1,a3,a4\} & \{a1,a2,a4\} & \{a3,a4\} & \{a1,a4\} & \{a4\} & \phi
\end{pmatrix}$$

取 $B=\{a1,a3,a4\}$，则 B 满足定理 1，于是 B 是协调集，并且 $B_1=\{a1,a3\}$，$B_2=\{a1,a4\}$，$B_3=\{a3,a4\}$ 均不满足定理 1，从而 B 是约简集。

3 集值信息系统的属性特征

设 $B=\{B_k:k\le l\}$ 是集值信息系统 (U,A,F) 的约简集全体，记 $C=\bigcap\limits_{k\le l}B_k$，$K=\bigcup\limits_{k\le l}B_k-C$，$I=U-(K\cup C)$，称 C 为 (U,A,F) 的核心属性集，K 为 (U,A,F) 的相对必要属性集，I 为 (U,A,F) 的不必要属性集。

定理 2　设 (U,A,F) 是一个集值信息系统，则以下命题等价：

(1) a 是核心属性；

(2)存在 $x_i,x_j\in U$，使 $D(x_i,x_j)=\{a\}$；

(3) $R_{A-\{a\}}^*\subsetneq R_A^*$；

证明：(1) $\Rightarrow$ (2)　若(2)不成立，即 $a\in D(x_i,x_j)$，$\left|D(x_i,x_j)\right|\ge 2$，记 $B=\cup\{(D(x_i,x_j)-\{a\}):[x_i]_A^*\cap[x_j]_A^*=\phi\}$，则对任意 $[x_i]_A^*\cap[x_j]_A^*=\phi$，有 $B\cap D(x_i,x_j)\neq\phi$。由定理 2 知，B 是协调集。从而存在约简 $C\subseteq B$，且 $a\notin C$，这与 a 是核心属性矛盾。

(2) $\Rightarrow$ (3)　由(2)知，存在 x_i,x_j，$f_a(x_i)\subsetneq f_a(x_j)$，且 $f_b(x_i)\subseteq f_b(x_j)$ $(b\neq a)$，所以 $(x_i,x_j)\in R_{A-\{a\}}^*$，$(x_i,x_j)\notin R_A^*$。从而 $R_{A-\{a\}}^*\subsetneq\Rightarrow$。

(3) $\Rightarrow$ (1)　若 a 不是核心属性，则存在属性约简 B，使 $a\notin B$，于是 $B\subseteq A-\{a\}$，从而 $R_{A-\{a\}}^*\subseteq R_B^*\subseteq R_A^*$，与(3)矛盾，则证。

定理 3　设 (U,A,F) 是一个集值信息系统，则 a 是不必要属性当且仅当 $R^*(a)\subseteq R_A^*$，其中 $R_{A-\{a\}}^*=\cup\{R_{B-\{a\}}^*:R_B^*\subseteq R_A^*,B\subseteq A\}$。

证明：若 a 是不必要属性，则 a 不存在于任何约简之中。于是 $\forall R_B^*\subseteq R_A^*(B\subseteq A)$，有 $R_{B-\{a\}}^*\subseteq R_A^*$，否则，若 $R_{B-\{a\}}^*\subsetneq R_A^*$，$\forall B'\subseteq B-\{a\}$，有 $R_{B'}^*\subsetneq R_A^*$，从而 B 是约简，且 $a\in B$，与 a 是不必要属性矛盾。

若 $R^*(a)\subseteq R_a^*$，则 $\forall B\subseteq A$，$R_B^*\subseteq R_A^*$，有 $R_{B-\{a\}}^*\subseteq R_a^*\subseteq R_a^*\cup R_A^*$，于是 $R_{B-\{a\}}^*\cap R_a^{*c}\subseteq R_A^*$，所以 $R_{B-\{a\}}^*=R_{B-\{a\}}^*\cap(R_A^*\cup R_a^{*c})=R_B^*\cup(R_{B-\{a\}}^*\cap R_a^{*c})\subseteq R_A^*$，即 a 不存在于任何约简之中，所以 a 是不必要属性。

定理 4　设 (U,A,F) 是一个集值信息系统，则：

(1) $a\in C$ 当且仅当 $R_{A-\{a\}}^*\subsetneq R_A^*$；

(2) $a\in K$ 当且仅当 $R_{A-\{a\}}^*\subseteq R_A^*$，且 $R^*(a)\subsetneq R_A^*$；

(3) $a\in I$ 当且仅当 $R^*(a)\subseteq R_A^*$。

证明：由定理 2、定理 3 即得。

以上结论说明了在集值信息系统中每个属性的类别和作用，并获得了每一种属性的判定定理，由此得到了另一种求约简的方法。

(1)判断属性 $a\in A$ 的类型：1)如果 $R_{A-\{a\}}^*\subsetneq R_A^*$，则 a 为核心属性；2)否则计算 $R^*(a)$，如果 $R^*(a)\subsetneq R_A^*$，则 a 为相对必要属性；3)如果 $R^*(a)\subseteq R_a^*$，则 a 为绝对不必要属性。

(2)寻找集值信息系统的约简集：1)根据上述方法判断属性类型；2)如果没有相对必要属性，则约简唯一，即核心 K；3)如果有相对必要属性，则从相对必要属性中取一个属性，

(下转第 36 页)