第7卷第2期 空 军 工 程 大 学 学 报(自然科学版) Vol.7 No.2
2006年4月 JOURNAL OF AIR FORCE ENGINEERING UNIVERSITY (NATURAL SCIENCE EDITION) Apr. 2006

优势关系下模糊目标信息系统约简的辨识矩阵

袁修久[1,2]， 何华灿[1]
(1. 西北工业大学 计算机学院，陕西 西安 710072;2. 空军工程大学 理学院，陕西 西安 710051)

摘 要：约简是知识获取的重要方法之一，基于等价关系的粗糙集约简理论的研究已比较深入，而优势关系下约简理论的研究还比较少。定义了模糊目标信息系统在优势关系下的5种属性约简，并且给出了它们的判定定理和可辨识矩阵。证明了辨识矩阵对应的辨识公式给出的解就是所求约简的全体.最后通过一个例子说明如何用辨识矩阵算法求属性约简。

关键词：粗糙集；属性约简；优势关系；辨识矩阵

中图分类号：TP18 **文献标识码**：A **文章编号**：1009-3516(2006)02-0081-04

粗糙集理论是Pawlak提出的用于处理不完备和不精确知识的数学工具[1]。属性约简是粗糙集理论的重要内容之一，它已经成为数据挖掘的重要方法。通过约简，可以去掉冗余的知识，获得最简的知识。寻求属性约简的算法是约简理论的基本问题，在计算约简的各种算法中，有相当一部分算法是基于辨识矩阵的。不少文献都讨论了辨识矩阵。对一个基于辨识矩阵的算法，最核心的问题是给出辨识矩阵的正确形式。由于实际的需要不同，可以定义不同的属性约简，对于绝大多数属性约简的辨识矩阵的形式并不是显而易见的，因此有必要对具体的约简给出辨识矩阵的具体形式。对于模糊目标信息系统的约简已有很多的研究[2,3]，但是优势关系[4,5]下的模糊目标信息系统的约简还未见研究。在本文我们首先利用模糊粗糙集[6]的概念定义出模糊目标信息系统在优势关系下的5种属性约简，然后给出它们的判定定理和辨识矩阵，通过例子说明了优势关系下$L^{(1)}$型约简的辨识矩阵算法。

1 基本概念

定义1[2] 设U为非空有限对象集合，称为论域；A为非空有限的条件属性集，对于任意条件属性$a\in A$，有$a:U\rightarrow V$，其中V_a是对应属性a的值域；D是决策属性集合，对于任意$d\in D$有$d:U\rightarrow[0,1]$，即d是论域Y的模糊子集，称$(U,A\cup D)$为模糊目标信息系统。

在本文假定V_a是一个全序集，即对任意$x,y\in U$，$a(x)>a(y)$，$a(x)<a(y)$和$a(x)=a(y)$中必有一个成立，其中$a(x)$表示属性a在x处的值。

定义2[5] 设$B\subseteq A$，记

$$R_B^{\geqslant}=\{(x,y)\mid a(x)\leqslant a(y),a\in B\},R_B^{\leqslant}=\{(x,y)\mid a(x)\geqslant a(y),a\in B\}$$

称$R_B^{\geqslant}(R_B^{\leqslant})$为模糊目标信息系统$(U,A\cup\{D\})$的条件属性集$B$确定的优势关系。

记
$$[x]_B^{\geqslant}=\{y\mid (x,y)\in R_B^{\geqslant}\},[x]_B^{\leqslant}=\{y\mid (x,y)\in R_B^{\leqslant}\}$$

$$\begin{cases}\underline{R}_B^{\geqslant}(d)(x)=Min\{d(y)\mid y\in[x]_B^{\geqslant}\}\\ \overline{R}_B^{\geqslant}(d)(x)=Min\{d(y)\mid y\in[x]_B^{\geqslant}\end{cases}$$

$\underline{R}_B^{\geqslant}(d)(x)$理解为在条件属性集$B$所确定的知识下$x$肯定属于模糊集$d$的程度；$\overline{R}_B^{\geqslant}(d)(x)$理解为在

收稿日期：2005-10-08
基金项目：国家自然科学基金资助项目(60273087)；北京市自然科学基金资助项目(4032009)
作者简介：袁修久(1966-)，男，陕西旬阳人，副教授，博士(后)，主要从事数据挖掘研究；
何华灿(1938-)，男，重庆奉节人，教授，博士生导师，主要从事人工智能与泛逻辑研究.

条件属性集 B 所确定的知识下 x 可能属于模糊集 d 的程度。

性质 1 若$[x]_B^{\geqslant} \subseteq [y]_B^{\geqslant}$,则$\underline{R}_B^{\geqslant}(d)(x) \geqslant \underline{R}_B^{\geqslant}(d)(y)$,$\overline{R}_B^{\geqslant}(d)(x) \leqslant \overline{R}_B^{\geqslant}(d)(y)$。

记

$L_B^{(1)}(x) = (\underline{R}_B^{\geqslant}(d_1)(x), \underline{R}_B^{\geqslant}(d_2)(x), \cdots, \underline{R}_B^{\geqslant}(d_r)(x))$,$L_B^{(2)}(x) = (\overline{R}_B^{\geqslant}(d_1)(x), \cdots, \overline{R}_B^{\geqslant}(d_r)(x)$;$L_{B\alpha}^{(3)}(x) = \{d_j | \underline{R}_B^{\geqslant}(d_1)(x) \geqslant \alpha\}$,$L_{B\alpha}^{(4)}(x) = \{d_j | \overline{R}_B^{\geqslant}(d_1)(x) \geqslant \alpha\}$,$\alpha \geqslant 0$。

2 优势关系下模糊目标信息系统的约简

定义 3 设 $B \subseteq A$ 非空,若任给 $x \in U$,有 $L_B^{(i)}(x) = L_A^{(i)}(x), i=1,2, (L_{B\alpha}^{(i)} = L_{A\alpha}^{(i)}, i=3,4)$,则称 B 为 $L^{(i)}, i=1,2$ 型协调集($L_\alpha^{(i)}, i=3,4$ 型协调集)。设 B 为 $L^{(i)}, i=1,2$ 型协调集($L_\alpha^{(i)}, i=3,4$ 型协调集),其任何非空真子集都不是 $L^{(i)}, i=1,2$ 型协调集($L_\alpha^{(i)}, i=3,4$ 型协调集),则称 B 为 $L^{(i)}, i=1,2$ 型约简($L_\alpha^{(i)}, i=3,4$ 型约简)。

定义 3 对应于信息系统的上近似约简,下近似约简以及广义决策约简。设 B 是 $L^{(i)}$型约简,则从模糊目标信息系统中获取的知识形式为

$$[x]_B^{\geqslant} \to L_B^{(i)}(x)$$

在信息系统中基于正域约简获取的知识是确定的,在模糊目标信息系统中,讨论通过约简获得确定的知识是没有意义的,但我们可以在如下的意义下理解“确定的”, 即在一定的水平下,所得规则的后件中结果是唯一的或只有较少可能的结果,也就是说$|L_{B\alpha}^{(4)}(x)| = 1$ 或$|L_{B\alpha}^{(4)}(x)|$为较小数。其中$|\cdot|$ 表示集合的基。

记 $\mathrm{POS}_B^n(d) = \{x | |L_{B\alpha}^{(4)}(x)| \leqslant n\}$, 其中 n 为较小正整数。

定义 4 设 $B \subseteq A$,若 $x \in \mathrm{POS}_A^n(d)$,有 $L_{B\alpha}^{(4)}(x) = L_{A\alpha}^{(4)}(x)$,则称 B 为优势关系下近似协调集,若 B 是优势关系下近似协调集,且 B 的任何真子集不是优势关系下近似协调集,则 B 称为优势关系下近似约简。

性质 2 $\mathrm{POS}_B^n(d) \subsetneqq \mathrm{POS}_A^n(d)$ 。

因 $L^{(i)}, i=1,2$ 型约简和 $L_\alpha^{(i)}, i=3,4$ 型约简的判定定理的表达形式和证明过程类似,我们把它们写成一个定理。

定理 1 1) B 是优势关系下 $L^{(1)}$型协调集($L_\alpha^{(3)}$型协调集)当且仅当对任意的 $x,y \in U$,若 $L_A^{(1)}(x) \geqslant L_A^{(1)}(y)$ 不成立($L_{A\alpha}^{(3)}(x) \geqslant L_{A\alpha}^{(3)}(y)$ 不成立)时,存在 $a \in B$,使得 $a(x) < a(y)$ 。

2) B 是优势关系下 $L^{(2)}$型协调集($L_\alpha^{(4)}$型协调集)当且仅当对任意的 $x,y \in U$,若 $L_A^{(2)}(x) \geqslant L_A^{(2)}(y)$不成立($L_{A\alpha}^{(4)}(x) \geqslant L_{A\alpha}^{(4)}(y)$ 不成立)时,存在 $a \in B$,使得 $a(x) > a(y)$。

注:对于向量 $g1 = (a_1, a_2, \cdots, a_n)$, $g2 = (b_1, b_2, \cdots, b_n)$,$g1 \geqslant g2$ 系指 $a_i \geqslant b_i, i=1,2,\cdots,n$;$g1 \leqslant g2$ 系指 $a_i \leqslant b_i, i=1,2,\cdots,n$。

证明 仅就优势关系下 $L^{(1)}$型协调集进行证明。设 B 为优势关系下 $L^{(1)}$型协调集,则对于任意的 $x,y \in U$,$L_B^{(1)}(x) = L_A^{(1)}(x)$,$L_B^{(1)}(y) = L_A^{(1)}(y)$。若 $L_A^{(1)}(x) \geqslant L_A^{(1)}(y)$ 不成立时, 不存在 $a \in B$ 使得 $a(x) < a(y)$,则对于任意的$a \in B$,有$a(x) \geqslant a(y)$,从而$[x]_B^{\geqslant} \subseteq [y]_B^{\geqslant}$,由性质1,$L_B^{(1)}(x) \geqslant L_B^{(1)}(y)$,即$L_A^{(1)}(x) \geqslant L_A^{(1)}(y)$,这同假设矛盾。

反过来,假若 B 不是优势关系下 $L^{(1)}$型约简,则必有 $x' \in U$,使得 $L_B^{(1)}(x') \neq L_A^{(1)}(x')$,又因为$[x']_B^{\geqslant} \cup_{y \in [x']_B^{\geqslant}} [y]_A^{\geqslant}$, 必有 $y_0 \in [x']_B^{\geqslant}$和 $d \in D$ 使得$\underline{R}_A(d)(y_0) < \underline{R}_A(d)(x')$,即 $L_A^{(1)}(y_0) \geqslant L_A^{(1)}(x')$ 不成立,根据假设存在 $a \in B$,使得 $a(y_0) < a(x')$,故 $y_0 \notin [x']_B^{\geqslant}$,矛盾。其他情况类似可以证明。

定理 2 B 是优势关系下近似协调集的充分必要条件是对任意的 x,y,若 $x \in \mathrm{POS}_A^n(d)$,$y \notin \mathrm{POS}_A^n(d)$,或 $x,y \in \mathrm{POS}_A^n(d)$ 且 $L_{A\alpha}^{(4)}(x) \supseteq L_{A\alpha}^{(4)}(y)$ 不成立,则存在 $a \in B$ 使得 $a(x) > a(y)$ 。

证明 设 B 是优势关系下的近似协调集,则

1) $x \in \mathrm{POS}_A^n(d)$,$y \notin \mathrm{POS}_A^n(d)$, 假如对所有的 $a \in B$ 有 $a(x) \leqslant a(y)$,则$[x]_B^{\geqslant} \supseteq [y]_B^{\geqslant}$, 故 $L_{A\alpha}^{(4)}(x) \supseteq L_{A\alpha}^{(4)}(y)$, 因为 $L_{B\alpha}^{(4)}(x)| \leqslant n$, 从而$|L_{B\alpha}^{(4)}(y)| \leqslant n$, 即 $y \in \mathrm{POS}_B^n(d)$,从而 $Y \in \mathrm{POS}_A^n(d)$, 这同 $y \notin \mathrm{POS}_A^n(d)$ 相矛盾。

2) 若 $x,y \in \mathrm{POS}_A^n(d)$ 且 $L_{A\alpha}^{(4)}(x) \supseteq L_{A\alpha}^{(4)}(y)$ 不成立。假如对所有的 $a \in B$ 有,$a(x) \leqslant a(y)$,则 $[x]_B^{\geqslant} \supseteq [y]_B^{\geqslant}$,故 $L_{B\alpha}^{(4)}(x) \supseteq L_{B\alpha}^{(4)}(y)$,由于 $L_{B\alpha}^{(4)}(x) = L_{A\alpha}^{(4)}(x)$, $L_{B\alpha}^{(4)}(y) = L_{AB\alpha}^{(4)}(y)$,从而 $L_{A\alpha}^{(4)}(x) \supseteq L_{A\alpha}^{(4)}(y)$,这同假设

矛盾。

反过来,假若 B 不是优势关系下的近似协调集,则必有 $x \in \mathrm{POS}_A^n(d)$ 使得 $L_{B\alpha}^{(4)}(x) \neq L_{A\alpha}^{(4)}(x)$,因为 $[x]_B^{\geqslant} = \cup_{y\in[x]_B^{\geqslant}}[y]_A^{\geqslant}$,故必存在 $y_0 \in [x]_B^{\geqslant}$ 使得 $L_{A\alpha}^{(4)}(y_0) \subseteq L_{A\alpha}^{(4)}(x)$ 不成立。若 $y_0 \notin \mathrm{POS}_B^n(d)$,则根据假设 $a(x) > a(y_0)$,故 $y_0 \notin [x]_B^{\geqslant}$ 矛盾;若 $y_0 \in \mathrm{POS}_B^n(d)$,则也有 $a(x) > a(y_0)$,故 $y_0 \notin [x]_B^{\geqslant}$ 矛盾。

定义 5　对于任意的 $x,y \in U$,记

$$c_i(x,y) = \begin{cases} \{a \mid a(x) < a(y)\}, W^i; \\ A \qquad\qquad \text{否则} \end{cases} \quad i = 1,2,3,4$$

设 $C^i = (c_i(x,y))$,分别称 C^i 为 $L^{(i)}$,$i=1,2$ 型约简, $L_\alpha^{(i)}$,$i=3,4$ 型约简的辨识矩阵,其中 W^i ,$i=1,2$ 为“ $L_A^{(i)}(x) \geqslant L_A^{(i)}(y)$,$i=1,2$ 不成立”; W^i ,$i=3,4$ 为“ $L_A^{(i)}(x) \leqslant L_A^{(i)}(y)$,$i=3,4$ 不成立”。

定义 6　对于任意的 $x,y \in U$,记

$$c_5(x,y) = \begin{cases} \{a \mid a(x) < a(y)\}, & x \in \mathrm{POS}_A^n(d), y \notin \mathrm{POS}_B^n(d) \text{或} \\ & x,y \in \mathrm{POS}_A^n(d), \text{但} L_{A\alpha}^{(4)}(x) \supseteq L_{A\alpha}^{(4)}(y) \text{不成立}; \\ A & \text{否则} \end{cases}$$

称 $C_5 = (c_5(x,y))$ 为优势关系下近似约简的辨识矩阵。

定理 3　B 是优势关系下 $L^{(i)}$,$i=1,2$ 或 $L_\alpha^{(i)}$,$i=3,4$ 型约简和近似约简的充分必要条件是对任意的 $x,y \in U$,分别有 $c_i(x,y) \cap B \neq \varnothing$,$i=1,2,3,4,5$。

定义 7　分别称 $M_i = \wedge\{\vee\{c_i(x,y)\}\}$,$i=1,2,3,4,5$ 为模糊目标信息系统的优势关系下 $L^{(1)}$, $L^{(2)}$ 型,$L_\alpha^{(3)}$ 型, $L_\alpha^{(4)}$ 型约简和近似约简的辨识公式。

定理 4　设辨识公式 M_i 的极小析取范式为

$$M_i = \vee_{k=1}^{p_i}(\wedge_{s=1}^{q_{ik}} a_{iks}), i=1,2,3,4,5。$$

记 $B_{ik} = \{a_{iks} \mid s=1,2,\cdots,q_{ik}\}$,则 B_{ik} ,$k=1,2,\cdots,p_i$,$i=1,2,3,4,5$ 分别是优势关系下 $L^{(1)}$,$L^{(2)}$,$L_\alpha^{(3)}$, $L_\alpha^{(4)}$ 型约简和近似约简的全体。

依据定义 5,定义 6, 定理 4 可以给出定义 3 和定义 4 所给出的 5 种优势关系下模糊目标信息系统的约简的辨识矩阵算法。

例 1　模糊目标信息系统如表 1,写出 $L^{(1)}$ 型约简的辨识矩阵并求其约简。

表 1　模糊目标信息系统

U	a_1	a_2	a_3	a_4	d_1	d_2	d_3
x_1	1	1	1	3	0.1	0.6	0.5
x_2	1	2	0	1	0.5	0.8	0.1
x_3	2	0	1	2	0.3	0.4	0.5
x_4	0	1	0	1	0.6	0.3	0.2
x_5	0	0	2	2	0.4	0.5	0.1
x_6	1	1	1	1	0.2	0.2	0.6

易算得　$L_A^{(1)}(x_1) = (0.1, 0.6, 0.5)$;$L_A^{(1)}(x_2) = (0.5, 0.8, 0.1)$;$L_A^{(1)}(x_3) = (0.3, 0.4, 0.5)$;

$L_A^{(1)}(x_4) = (0.1, 0.2, 0.1)$;$L_A^{(1)}(x_5) = (0.1, 0.5, 0.1)$;$L_A^{(1)}(x_6) = (0.1, 0.2, 0.5)$ 。

由此可以得到 $L^{(1)}$ 型约简的辨识矩阵如表 2 所示。

表 2　优势关系下的 $L^{(i)}$ 型约简的辨识矩阵

	$L_A^{(1)}(x_1)$	$L_A^{(1)}(x_2)$	$L_A^{(1)}(x_3)$	$L_A^{(1)}(x_4)$	$L_A^{(1)}(x_5)$	$L_A^{(1)}(x_6)$
$L_A^{(1)}(x_1)$	A	a_2	a_1	A	A	A
$L_A^{(1)}(x_2)$	a_3, a_4	A	a_1, a_3, a_4	A	A	a_3
$L_A^{(1)}(x_3)$	a_2, a_3	a_2	A	a_3	a_2	a_1, a_4
$L_A^{(1)}(x_4)$	a_1, a_3, a_4	a_1, a_2	a_1, a_3, a_4	A	a_3, a_4	a_1, a_3
$L_A^{(1)}(x_5)$	a_1, a_2, a_4	a_1, a_2	a_1	A	A	a_1, a_2
$L_A^{(1)}(x_6)$	a_3, a_4	a_2	a_1, a_4	A	a_3, a_4	A

$M = a_2 \wedge a_1 \wedge \{a_3 \vee a_4\} \wedge \{a_1 \vee a_3 \vee a_4\} \wedge a_3 \wedge \{a_2 \vee a_3\} \wedge \{a_1 \vee a_4\} \wedge \{a_1 \vee a_2\} \wedge \{a_1 \vee a_3\} \wedge \{a_1 \vee a_2 \vee a_4\} = a_1 \wedge a_2 \wedge a_3$

故所求 $L^{(1)}$ 型约简为 $\{a_1, a_2, a_3\}$ 。

3 结论

本文给出了模糊目标信息系统在优势关系下的 $L^{(i)}, i=1,2, L_\alpha^{(i)}, i=3,4$ 型约简和近似约简的可辨识矩阵,由此可以给出相应约简的约简算法。

参考文献:

[1] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data[M]. Boston: Kluwer Academic Publishers, 1991.

[2] 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.

[3] 袁修久, 张文修. 模糊目标信息系统的属性约简[J]. 系统工程理论与实践, 2004, 24(5):116 - 120.

[4] Greco S, Matarazzo B, Slowinski R. Rough Sets Theory for Multicriteria Decisionanalysis[J]. European Journal of Operational Reasearch, 2002, 129:1 - 47.

[5] Greco S, Matarazzo B, Slowinski R. Rough approximation by domaince relations[J]. International Journal of Inteligent Systems, 2002, 17:153 - 171.

[6] Banerjee M, Pal S k. Roughness of a Fuzzy Set[J]. Information Sciences, 1996, 93:235 - 246.

[7] 安芹力,李安平. 不协调决策表的属性约简模型及其规则提取[J]. 空军工程大学学报(自然科学版)2005,6(3):88 - 91.

(编辑:田新华)

Discernibility Matrix with Respect to Reduction Based on Dominance Relation in Fuzzy Decision Information System

YUAN Xiu - jiu[1,2], HE Hua - can[1]

(1. College of Computer Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China; 2. The Science Institute, Air Force Engineering University, Xi'an, Shaanxi 710051, China)

Abstract: Knowledge reduction based on rough set theory is one of the important methods of knowledge - acquisition. Comparatively speaking, the attribute reduction based on equivalence relation has been investigated in depth, however, there is less study of attribute reduction based on dominance relation, so, in the paper five kinds of knowledge reduction are defined. The judgment theorems and discernibility matrixes with respect to these reductions are established, from which the algorithms of discermibility matrix for finding these reductions are obtained. Finally, an example is given.

Key words: rough set; attribute reduction; dominance relation; discernibility matrixes