

# 基于极小析取范式的属性约简算法

索中英<sup>1</sup>, 程嗣怡<sup>2</sup>, 苟新禹<sup>2</sup>

(1.空军工程大学理学院,陕西西安,710051;2.空军工程大学航空航天工程学院,陕西西安,710038)

**摘要** 属性约简是粗糙集理论的核心问题之一,针对求取决策表所有决策约简集的 NP 问题,化繁为简将问题转化为对象动态增加下的决策约简求取问题。在深入分析了可辨识矩阵中可辨识集的特点及相互关系的基础上,优化改进决策辨识矩阵:①两对象之间不作逆向比较;②将决策可辨识矩阵列简化为属性等价类;③正域等价类作为决策可辨识矩阵的行,分情况给出了新决策表求取所有决策约简集的极小析取范式属性约简方法。该方法统一解决了相容和不相容决策表所有决策约简集的求取问题,最后通过实例分析验证了算法的可行性与有效性,为决策表的属性约简提供了一条高效的途径。

**关键词** 粗糙集;属性约简;决策约简集;可辨识矩阵;极小析取范式

**DOI** 10.3969/j.issn.1009-3516.2012.06.014

**中图分类号** TK401 **文献标志码** A **文章编号** 1009-3516(2012)06-0066-05

属性约简是粗糙集理论的重要应用,也是其核心问题之一<sup>[1]</sup>。知识库中的属性并不是同等重要的,某些属性冗余。属性约简就是在保持分类能力不变的情况下,删除信息系统中不相关或不重要的属性。Skowron 教授等人在文献[2]中提出的基于可辨识矩阵(亦称差别矩阵或分明矩阵)的分辨函数方法是计算决策表全部属性约简的一个有代表性的方法,不少关于属性约简及其拓展的研究工作都是以该方法(特别是其中的 Skowron 可辨识矩阵)的思想为基础的<sup>[3-7]</sup>。然而文献[8~9]指出:当决策表不相容时,常用的计算全部属性约简(代数意义下)的基于 Skowron 可辨识矩阵的分辨函数方法会产生错误的结果,并给出了一个改进的 Skowron 可辨识矩阵,解决了不相容的情况下,决策表的属性约简求取问题。

对一个决策表求取所有的决策约简集是一个 NP 难题。在实际应用中,人们已提出若干以启发式为基础的属性约简算法<sup>[5]</sup>,但这往往只能求出某种次优的属性约简。实际上,当决策表中包含对象相对较少时,可以通过构造决策表的可辨识矩阵和分辨函数,利用布尔函数得到全部决策约简集。鉴于此,本文在改进的 Skowron 可辨识矩阵的基础上,借助极小析取范式,统一解决了相容和不相容决策表的全部决策约简集求取问题,提供了一条较为高效的求取所有决策约简集的途径。

## 1 相关概念及理论

### 1.1 决策辨识集与辨识矩阵

决策表是一类特殊而重要的知识表达系统,在粗糙集理论中起着重要的作用<sup>[10]</sup>。决策表可以根据知识表达系统定义如下:

称  $S = (U, A, F, d)$  是一个决策表, $U$  为对象集,即  $U = \{x_1, x_2, \dots, x_n\}$ , $U$  中的每一个元素  $x_i (i \leq n)$  称为一个对象。 $A$  为属性集,即  $A = \{a_1, a_2, \dots, a_m\}$ , $A$  中的每个元素  $a_l (l \leq m)$  称为一个属性。 $F$  为  $U$  与  $A$  之间

收稿日期:2011-11-04

基金项目:陕西省自然科学基金资助项目(2012JQ8019);陕西省电子信息系统系统集成重点实验室基金资助项目(201113Y01)

作者简介:索中英(1981—),女,山西忻州人,讲师,博士,主要从事智能决策与信息系统研究。  
E-mail:csy\_316@163.com

的关系集,即  $F = \{f_l : U \rightarrow V_l (l \leq m)\}$ , 其中  $V_l$  为  $a_l (l \leq m)$  的值域。  $d : U \rightarrow V_d, V_d$  取有限值。

记:  $R_A = \{(x_i, x_j) \mid f_l(x_i) = f_l(x_j) (a \in A)\}$ ,  $R_d = \{(x_i, x_j) \mid d(x_i) = d(x_j)\}$ , 若  $R_A \subseteq R_d$ , 则称  $S = (U, A, F, d)$  为相容决策表。若  $R_A \not\subseteq R_d$ , 则  $S$  为不相容决策表。设  $B \subset A$ , 若  $Pos_B(d) = Pos_A(d)$ , 且不存在  $B' \subset B$  使得  $Pos_{B'}(d) = Pos_A(d)$  成立, 则称  $B$  为  $A$  的决策约简集。其中  $Pos_P(Q)$  表示  $Q$  的  $P$  正区域, 是  $U$  中所有根据分类  $U/R_P$  可以准确分到关系  $Q$  的等价类  $U/R_Q$  中的对象集合。

记:  $[x_i]_A = \{x_j \mid (x_i, x_j) \in R_A\}$ ,  $[x_i]_d = \{x_j \mid (x_i, x_j) \in R_d\}$ ,  $U/R_A = \{[x_i]_A \mid x_i \in U\}$ ,  $U/R_d = \{[x_i]_d \mid x_i \in U\}$ ,  $D_d([x_i]_A, [x_j]_A) = \begin{cases} \{a \in A \mid f_l(x_i) \neq f_l(x_j)\} & , m_{ij} \\ \emptyset & , \text{其它} \end{cases}$ , 其中  $m_{ij}$  表示  $[d(x_i) \neq d(x_j)] \wedge [\min(|d([x_i]_A)|, |d([x_j]_A)|) = 1]$ , 称  $D_d([x_i]_A, [x_j]_A)$  为  $[x_i]_A$  与  $[x_j]_A$  的决策辨识集, 称  $D_d = (D_d([x_i]_A, [x_j]_A) \mid [x_i]_A, [x_j]_A \in U/R_A)$  为决策表的决策可辨识矩阵。

### 1.2 决策约简集与布尔代数的关系描述

将一个布尔函数称为决策分辨函数, 用  $\Delta$  表示。对每个属性  $a \in A$ , 指定一个布尔变量  $a$ 。

若  $D_d([x_i]_A, [x_j]_A) = \{a, a, \dots, a_k\} \neq \emptyset$ , 则指定一个布尔函数  $a \vee a \vee \dots \vee a_k$ , 用  $\sum D_d([x_i]_A, [x_j]_A)$  来表示; 若  $D_d([x_i]_A, [x_j]_A) = \emptyset$ , 则指定布尔常量 1。布尔决策分辨函数  $\Delta$  可定义为:  $\Delta = \prod_{[x_i]_A, [x_j]_A \in U/R_A} \sum D_d([x_i]_A, [x_j]_A)$ 。作为决策分辨函数的布尔函数具有以下性质<sup>[11]</sup>:

- 1) 布尔表达式  $E$  是一个范式, 如果  $E$  仅仅由布尔变量和常量通过析取与合取运算来表达;
- 2)  $E$  是一个合取范式, 如果  $E$  是由一些析取式组成的合取所构成的范式;
- 3)  $E$  是一个析取范式, 如果  $E$  是由一些合取式组成的析取所构成的范式;
- 4)  $E$  是一个极小析取范式, 如果  $E$  是一个析取范式且包含最小数目的合取式。

决策分辨函数  $\Delta$  的极小析取范式中的所有合取式的析取是属性集  $A$  的所有决策约简集。

## 2 基于极小析取范式的属性约简

为有效提高约简的求取效率, 首先对决策可辨识矩阵作出优化改进: ① 两对象之间不作逆向比较; ② 将决策可辨识矩阵简化为属性等价类; ③ 正域等价类作为决策可辨识矩阵的行。

**定理 1** 设决策表  $S = (U, A, F, d)$  共有  $n-1$  个对象, 假设收集到一与  $S$  原有对象不矛盾的新对象  $x_n$ , 则有  $\Delta_n = \Delta_{n-1} \wedge \Delta_n$ 。其中:  $\Delta_n = \prod_{\substack{[x_i]_A, [x_j]_A \in U/R_A \\ i, j = 1, 2, \dots, n}} \sum D_d([x_i]_A, [x_j]_A) \Delta_k = \prod_{[x_i]_A \in U/R_A} \sum D_d([x_i]_A, [x_k]_A)$ 。

由于析取和合取运算满足交换律、结合律以及分配律, 因此利用数学归纳法容易证明, 本文不再赘述。

**定理 2** 设决策表  $S^*$  和  $S^\#$  的决策分辨函数分别为  $\Delta_{S^*}$  和  $\Delta_{S^\#}$ , 且决策表  $S^*$  和  $S^\#$  之间不存在矛盾对象时, 由其全部对象所组成的新决策表  $S$  的决策辨识函数  $\Delta_S = \Delta_{S^*} \wedge \Delta_{S^\#}$ 。由于析取和合取运算满足交换律、结合律以及分配律, 定理容易得证。

**定理 1** 和 **定理 2** 分别解决了对象逐个增加和批量增加情况下的决策辨识函数求取问题。

基于极小析取范式求取全部决策约简集的流程见图 1。步骤如下:

步骤 1 给出已有决策表  $S$ ;

步骤 2 计算并保留  $S$  的极小析取范式  $\Delta_S$ ,  $[x_i]_A$  与  $[x_j]_A$  之间的决策辨识集  $a_{ij}$  (其中  $i > j, [x_i]_A \subset [x_i]_d, [x_j]_A \subset [x_j]_d$ ), 以及  $[x_i]_A$  与  $[x_j]_A$  之间的决策辨识集  $b_{ij}$  (其中  $[x_i]_A \subset [x_i]_d$  且  $[x_j]_A \not\subset [x_j]_d$ );

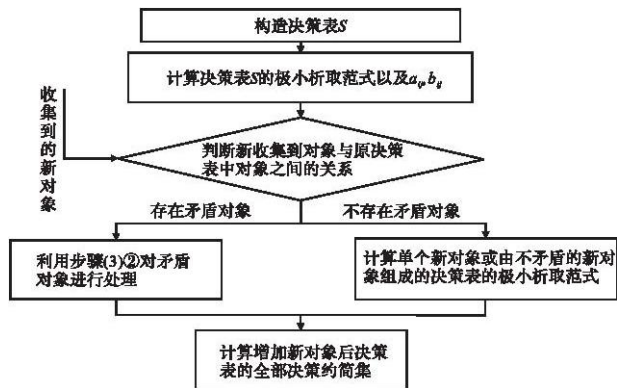


图 1 基于极小析取范式的决策属性约简求取流程

Fig. 1 The process of attribute reduction based on minimal disjunctive normal form

步骤 3 假设收集到单个对象  $x_k$  :

1) 当  $x_k$  与  $S$  中对象重复时,决策约简集不变;

2) 当  $x_k$  与  $S$  中对象  $x_l$  矛盾时,删去  $[x_l]_A$  相关的决策辨识集,保留  $a_{ij} (i, j \neq l)$  和  $b_{ij} (i \neq l)$ ; 计算  $[x_i]_A$  与  $[x_k]_A$  之间的决策辨识集  $b_{ik}$  并保留(其中  $[x_i]_A \subset [x_k]_A$ ), 收集到对象  $x_k$  后的决策表  $S^\#$  的极小析取范式  $\Delta_s^\# = a_{ij} \wedge b_{ij} \wedge b_{ik}$ 。

3) 当  $x_k$  与  $S$  中对象既不重复也不矛盾,即收集到一新对象时,计算  $\Delta_k = \prod_{[x_i] \in U/R_A} \sum D_d([x_i]_A, [x_k]_A)$ ,

则增加新对象  $x_k$  后新决策表  $S^*$  的极小析取范式为:  $\Delta_s^* = \Delta_s \wedge \Delta_k$ , 替代  $\Delta_s$ , 计算决策辨识集  $a_{ij}, b_{ij}$  并增加记录。

步骤 4 当收集到多个数据时,首先分析是否与  $S$

中对象矛盾,如果存在矛盾对象,则应用步骤 3 中的 2) 方法进行单独处理,如果不存在矛盾对象,则求取由新对象构成的新决策表的极小析取范式,再利用定理 2 计算原有决策表增加新的对象后的极小析取范式。

步骤 5 计算极小析取范式中的所有合取式的析取,得到相应决策表的全部决策约简集。

表 1 决策表

Tab.1 Decision Table

	$a$	$\alpha$	$\alpha$	$a$	$d$
$x_1$	1	0	1	1	1
$x_2$	0	0	1	0	0
$x_3$	1	1	0	0	1
$x_4$	1	1	1	1	1
$x_5$	1	0	1	0	0
$x_6$	1	0	1	0	1
$x_7$	1	0	1	0	0
$x_8$	0	0	1	0	0
$x_9$	1	1	1	1	1
$x_{10}$	0	0	1	0	0
$x_{11}$	0	1	1	0	1
$x_{12}$	1	1	0	0	0

### 3 实例分析

#### 3.1 例 1

给定一决策表  $S$  见表 1, 其中  $U = \{x_1, x_2, \dots, x_{10}\}, A = \{a, \alpha, \alpha, a\}, d$  为决策属性。 $x_{11}$  和  $x_{12}$  为决策表依次增加的对象。

则有:  $U/\{d\} = \{\{x_1, x_3, x_4, x_6, x_9\}, \{x_2, x_5, x_7, x_8, x_{10}\}\}, U/A = \{\{x_1\}, \{x_2, x_8, x_{10}\}, \{x_3\}, \{x_4, x_9\}, \{x_5, x_6, x_7\}\}$ 。

决策表  $S$  优化后的可辨识矩阵见表 2 所示。其中正域等价类:  $[x_i]_{Pos_A(d)} = \{x_j \mid (x_i, x_j) \in R_A \wedge (x_i, x_j) \in R_d\}$

表 2 决策表的可辨识矩阵

Tab.2 Discernibility matrix of decision table

	$a_{ij}$				$b_{ij}$	
	$[x_1]_A$	$[x_2]_A$	$[x_3]_A$	$[x_4]_A$	$[x_5]_A$	$[x_{12}]_A$
$[x_1]_{Pos_A(d)}$	$\emptyset$				$\{a\}$	$\{\alpha, \alpha, a\}$
$[x_2]_{Pos_A(d)}$	$\{a, a\}$	$\emptyset$			$\{a\}$	$\{a, \alpha, \alpha\}$
$[x_3]_{Pos_A(d)}$	$\emptyset$	$\{a, \alpha, \alpha\}$	$\emptyset$		$\{\alpha, \alpha\}$	
$[x_4]_{Pos_A(d)}$	$\emptyset$	$\{a, \alpha, a\}$	$\emptyset$	$\emptyset$	$\{\alpha, a\}$	$\{\alpha, a\}$
$[x_{11}]_{Pos_A(d)}$	$\emptyset$	$\{\alpha\}$	$\emptyset$	$\emptyset$	$\{\alpha, \alpha\}$	$\{a, \alpha\}$

决策表  $S$  的极小析取范式  $\Delta_s = a \wedge a \wedge (\alpha \vee \alpha)$ , 针对该差别矩阵只需保留  $a_{21}, a_{22}, a_{23}$  以及  $b_{ij} (i=1, 2, 3, 4, j=5)$ 。当收集到  $x_{11}$  时, 发现对于原决策表,  $x_{11}$  是一全新对象, 因此首先计算  $\Delta_{11} = \alpha \wedge (a \vee \alpha) = \alpha$ , 则新决策表  $S^*$  的极小析取范式为:

$$\Delta_s^* = \Delta_s \wedge \Delta_{11} = (a \wedge a \wedge (\alpha \vee \alpha)) \wedge \alpha = a \wedge \alpha \wedge a$$

将其替代  $\Delta_s$ , 并增加记录  $a_{1.2}, b_{11.5}$ 。于是决策表  $S^*$  的决策约简集为  $\{a, \alpha, a\}$ 。

当继续增加对象  $x_{12}$  后, 发现新对象  $x_{12}$  与  $x_3$  矛盾, 首先删除记录  $a_{22}, b_{35}$ , 其次增加  $b_{i.12} (i=1, 2, 4, 11)$ , 则增加对象  $x_{12}$  后决策表  $S^\#$  的极小析取范式:

$$\Delta_s^\# = [(a \vee a) \wedge \alpha] \wedge [a \wedge a \wedge (\alpha \vee \alpha) \wedge (a \vee \alpha)] \wedge [(\alpha \vee \alpha \vee a) \wedge (a \vee \alpha \vee \alpha) \wedge (\alpha \vee a) \wedge (a \vee \alpha)] = a \wedge \alpha \wedge a$$

即决策表  $S^\#$  的决策约简集为  $\{a, a_1, a_2\}$ 。

### 3.2 例2

当2个决策表之间不存在相互矛盾的对象时,对于新决策表,可以利用定理2求取其全部决策约简集。

假设存在决策表  $S^*$  由表1决策表中的对象  $\{x_1, x_2, x_4, x_5, x_6, x_7\}$  组成,而后收集到对象  $\{x_3, x_8, x_9, x_{10}, x_{11}, x_{12}\}$ ,通过分析发现由  $\{x_1, x_2, x_4, x_5, x_6, x_7\}$  构成的决策表  $S^*$  和由  $\{x_3, x_8, x_9, x_{10}, x_{11}, x_{12}\}$  构成的决策表  $S^\#$  之间不存在相互矛盾的对象。并且由于  $\Delta_{S^*} = a \wedge a_1, \Delta_{S^\#} = a_2 \wedge (a_3 \vee a_4) \wedge (a_5 \vee a_6)$ ,所以依据定理2  $\Delta_S = \Delta_{S^*} \wedge \Delta_{S^\#} = [a \wedge a_1] \wedge [a_2 \wedge (a_3 \vee a_4) \wedge (a_5 \vee a_6)] = a \wedge a_1 \wedge a_2$ ,所得结果与示例1结果一致。

## 4 结语

本文提出的属性约简方法,以改进的 Skowron 差别矩阵以及布尔代数的特点为基础,化繁为简将求取决策表所有决策约简的 NP 问题转化为决策表对象动态增加的问题,针对所增加对象与原决策表对象重复、矛盾(决策表不相容)、既不重复也不矛盾(决策表相容)3种情况,深入剖析决策辨识集的特点及其相互关系,构造了新决策表求取所有决策约简集的方法,此外给出了对象批量增加情况下的决策约简集求取途径。

对于一个信息系统而言,逐步收集到的数据应以全新对象为主,本文所提出的基于极小析取范式的属性约简算法能高效的处理此类情况,此外即使收集到的是一矛盾对象,考虑到决策可辨识矩阵中非空辨识集的稀疏性,亦可以较为方便的得到决策表的所有决策约简集。

### 参考文献(References):

- [1] Pawlak Z. Rough sets[J]. International journal of information and computer sciences, 1982, 11:341-356.
- [2] Skowron A, Rauszer C. The discernibility matrices and functions in information systems intelligent decision support: handbook of applications and advance of the rough sets theory[M]. [s. n.]: Kluwer academic publishers, 1992:331-362.
- [3] GUAN J W, BELLD A. Matrix computation for information system[J]. Information sciences, 2001, 131: 129-256.
- [4] Wang J. Reduction algorithms based on discernibility matrix: the ordered attributes method[J]. J comput sci technol, 2001, 16(6):489.
- [5] 陶志, 许宝栋, 汪定伟. 一种基于分明矩阵的启发式知识约简方法[J]. 系统工程与电子技术, 2005, 27(4):734-736.  
TAO Zhi, XU Baodong, WANG Dingwei. Approach to heuristic knowledge reduction based on discernibility matrix[J]. Systems engineering and electronics, 2005, 27(4):734-736. (in Chinese)
- [6] Yao Y Y, Zhao Y. Discernibility matrix simplification for constructing attribute reducts [J]. Information sciences, 2009, 179(5):867-882.
- [7] 秦奕青, 杨炳儒, 徐章艳. 基于 Skowron 分明矩阵的有效属性约简算法[J]. 北京科技大学学报, 2009, 31(3):398-404.  
Qin Yiqing, YANG Bingru, XU Zhangyan. Efficient algorithm of attribute reduction based on Skowron's discernibility matrix[J]. Journal of Beijing university of science and technology, 2009, 31(3):398-404. (in Chinese)
- [8] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7):1086-1088.  
YE Dongyi, CHEN Zhaojiong. A new discernibility matrix and the computation of a core[J]. Acta electronica sinica, 2002, 30(7):1086-1088. (in Chinese)
- [9] 叶东毅, 陈昭炯. 不相容决策表全部属性约简计算的一个改进方法[J]. 小型微型计算机系统, 2006, 27(10):1909-1913.  
YE Dongyi, CHEN Zhaojiong. Improved method for computing all attribute reducts of an inconsistent decision table[J]. Mini-micro system, 2006, 27(10):1909-1913. (in Chinese)
- [10] 张文修, 仇国芳. 基于粗糙集的不确定决策[M]. 北京:清华大学出版社, 2006.  
ZHANG Wenxiu, QIU Guofang. Uncertain decision making based on rough sets[M]. Beijing: Tsinghua university press, 2006. (in Chinese)
- [11] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京:科学出版社, 2001.  
ZHANG Wenxiu, WU Weizhi, LIANG Jiye, et al. Rough sets theory and methods[M]. Beijing: Science press, 2001. (in Chinese)

# The Algorithm of Attribute Reduction Based on Minimal Disjunctive Normal Form

SUO Zhong-ying<sup>1</sup>, CHENG Si-yi<sup>2</sup>, GOU Xin-yu<sup>2</sup>

(1. Science College, Air Force Engineering University, Xi'an 710051, China; 2. Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

**Abstract:** Attribute reduction is a key problem in theoretical research of rough set. Aiming at the NP-hard problem of acquiring all decision reduction in decision table, the paper simplifies the problem to decision reduction computation in incremental object. The arithmetic of attribute reduction based on minimal disjunctive normal form is respectively offered on the basis of analyzing the characteristic and correlation of discernibility set in discernibility matrix, in which the decision discernibility matrix is optimized by ① omitting repeated contrast between the objects; ② transforming the row of the decision discernibility matrix to the equivalent set; ③ treating the equivalent set of the positive field as the other row of the decision discernibility matrix. The method offers a tool to uniformly acquire all decision reduction of consistent and inconsistent decision table. The example analysis shows this algorithm's feasibility and validity, which offers an effective way for attribute reduction.

**Key words:** rough set; attribute reduction; decision reduction; discernibility matrix; minimal disjunctive normal form

## 本刊相关链接文献:

- [1]史超,程咏梅.基于改进证据理论的群决策专家意见集结方法[J].空军工程大学学报:自然科学版,2011,12(5):85-90.
- [2]袁修久,高生强,杨宇.分级粗糙集和分级知识约简[J].空军工程大学学报:自然科学版,2009,10(2):91-94.
- [3]周炜,雷英杰.Pawlak粗糙集的若干重要性质[J].空军工程大学学报:自然科学版,2008,9(3):76-78.
- [4]袁修久,何华灿.优势关系下模糊目标信息系统约简的辨识矩阵[J].空军工程大学学报:自然科学版,2006,7(2):81-84.
- [5]赵惠文,朱根标,张凤鸣.基于粗集和模糊集的启发式规则约简算法研究[J].空军工程大学学报:自然科学版,2005,6(5):88-90.
- [6]安芹力,李安平.不协调决策表的属性约简模型及规则提取[J].空军工程大学学报:自然科学版,2005,6(3):88-91.
- [7]袁修久,杨合俊,张小水.广义决策约简同相对约简的关系[J].空军工程大学学报:自然科学版,2005,6(1):44-47.