

二进制可辨矩阵的变换及高效属性约简算法的构造^{*}

The Binary Discernibility Matrix's Transformation and High Efficiency Attributes Reduction Algorithm's Conformation

支天云¹ 苗夺谦²

(山西大学计算机科学系 太原市030006)¹(山西大学数学系 太原市030006)²

Abstract This paper dose a lot of mathematics transformation to Binary Discernibility Matrix, it is called BDM too. Through these transformations, we may both decrease a lot of memory space of BDM and contribute to make high efficiency attributes reduction algorithms. Based on BDM, we also give three algorithms which are in turn used to reduce BDM, attribute's reduction and judge reduction. Finally, we give three examples to illuminate the feasibility of three algorithms and analyze the complexity of algorithms.

Keywords Rough sets, Binary discernibility matrix, Attribute reduction

1 引言

由波兰理工大学教授 Z. Pawlak 等一批科学家提出的粗糙集(Rough Set)理论^[1]目前正受到国内外学者越来越广泛的重视。它从一种全新的视觉审视知识,认为知识与分类能力相关,两者成正比关系,即:拥有知识的多少通过分辨能力的高低来评判。基于此,许多有关知识的问题可以形式化之后通过计算机来处理。粗糙集理论自产生之日起,即有两个发展方向,一个是粗糙集理论的数学结构的研究,另一个是粗糙集理论在各个具体领域中的应用,如:通过粗糙集理论进行的属性约简等。我们一直认为,这两个发展方向应该有机地结合,即数学结构上的研究应该便于应用,转化为一些具体的算法,为应用服务。基于这种考虑,我们对二进制可辨矩阵进行了各种数学变换,并在此基础上构造了高效的属性约简算法,而这些数学变换又相应于属性约简的过程,即变换本身并不减少知识量和分类能力。

本文基于二进制可辨矩阵(Binary Discernibility Matrix, BDM),构造了高效的属性约简算法。首先,介绍了二进制可辨矩阵的表示、分析及其约简变换;然后是基于二进制可辨矩阵^[2]的属性约简算法,分别包括二进制可辨矩阵的化简、属性约简和约简判定三个算法;最后是算法复杂度分析和结束语。

2 二进制可辨矩阵^[2]

2.1 二进制可辨矩阵的表示

设决策表为: $T = (U, C, D, V, f)$, 其中 $U = \{u_1, u_2, \dots, u_n\}$, $C = \{c_1, c_2, \dots, c_m\}$, $D = \{d\}$, 则决策表 T 相应的二进制可辨矩阵 M_T 构造如下:

矩阵的每一列对应一个条件属性,共有 m 列,每一行对应一对论域中的对象 (u_p, u_q) , 且 $d(u_p) \neq d(u_q)$, 即这一对对象属于不同的决策类,至多有 $n(n-1)/2$ 行。设矩阵中一元素 $m_{((p,q),i)}$ 所在行对应对象对 (u_p, u_q) , 所在列对应条件属性 c_i , 则

$$m_{((p,q),i)} = \begin{cases} 1 & \text{如果 } c_i(u_p) \neq c_i(u_q) \\ 0 & \text{否则} \end{cases}$$

这样得到的一个矩阵,我们称之为相应于决策表 $T = (U, C, D, V, f)$ 的二进制可辨矩阵。

设 $I = (U, A, V, f)$ 是一信息系统,则相应于 I 的二进制可辨矩阵 M_I 可类似得到。由于信息系统是条件属性集与决策属性集相同的决策表,故它的行应对应在不可分辨关系 $IND(A)$ ^[1]下可分辨的对象对 (u_p, u_q) , 它的列对应属性集 A 中的属性 a_i 。设 $M_I = (m_{((p,q),i)})$, 则

$$m_{((p,q),i)} = \begin{cases} 1 & \text{如果 } a_i(u_p) \neq a_i(u_q) \\ 0 & \text{否则} \end{cases}$$

与 M_T 不同的是 M_I 中不存在全为0的行。

2.2 二进制可辨矩阵的分析

二进制可辨矩阵 M_T 或 M_I 直接描述了每个属性对论域中对象的分辨情况,所以可以直接反映决策表 T 或信息系统 I 中所蕴涵的知识。在二进制可辨矩阵 M_T 中,某个元素为1或0表示所在行的属于不同决策类的对象 u_p, u_q 在条件属性 c_i 下可分辨与不可分辨,因此,若二进制可辨矩阵中有全为0的行,则说明相应的决策表 $T = (U, C, D, V, f)$ 是不协调的,否则,决策表 T 是协调的。在信息系统 $I = (U, A, V, f)$ 中,由于我们列出的行是在不可分辨关系 $IND(A)$ ^[1]下可分辨的对象对 (u_p, u_q) , 故在信息系统 $I = (U, A, V, f)$ 相应的二进制可辨矩阵 M_I 中不存在全为0的行。在 M_I 或 M_T 中,若某一行只有一个元素为1,其余元素均为0,则这个元素1对应的列属性 a_i 或 c_i 一定属于核或相对核。因此,有下述命题:

命题1 若二进制可辨矩阵中某一行只有一个元素为1,其余元素均为0,则元素1所在列对应某个属性,所有这样的属性构成信息系统的核或决策表的相对核。若没有这样的行,则核或相对核为空。

由于约简是保持分类能力的最小属性子集,所以有如下命题:

命题2 在决策表 $T = (U, C, D, V, f)$ 中, $B \subseteq C$ 是 C 的一个 D -约简,其充要条件为:(1)在由 B 中所有属性对应的各列

^{*} 本文得到国家自然科学基金项目(NO. 69805004)、山西省青年科技基金(NO. 20001001)和模式识别国家重点实验室的资助。支天云 讲师,在职硕士研究生,主要研究方向为 Rough Set 理论及其应用、自然语言理解。苗夺谦 博士,教授,硕士生导师,主要研究方向为 Rough Set 理论与应用、KDD、Data Mining、人工智能和模式识别。

所构成的 M_T 的子阵中,不全为0的行数等于 M_T 中不全为0的行数;(2) $\forall B' \subset B, B'$ 不满足条件(1)。

与决策表的情况相仿,关于信息系统,有如下命题:

命题3 在信息系统 $I=(U, A, V, f)$ 中, $B \subseteq A$ 是 A 的一个约简,其充要条件为:(1)在由 B 中所有属性对应的各列所构成的 M_I 的子阵中,不存在全为0的行;(2) $\forall B' \subset B, B'$ 不满足条件(1)。

根据二进制可辨矩阵,我们很容易求出核或相对核,但如何求得约简呢?特别是如何求得最小约简呢?首先遇到的问题是二进制可辨矩阵最多有 $n(n-1)/2$ 行,在信息系统中,若某一行的元素全为1,说明相应的两个对象 u_p, u_q 在任何一个属性下都可分辨,此时,这一行可去掉而不影响约简。依据不影响约简,二进制可辨矩阵 M_I 也可作如下变换:

- (1)相同的行或列只出现一次。
- (2)行、列可以任意重新排序。
- (3)对于某两列,如 a_i 列与 a_j 列,若 $a_i + a_j = a_i$ (“+”表示逻辑加),则 a_i 列可以去掉。
- (4)对于某两行,如 (u_p, u_q) 行与 $(u_{p'}, u_{q'})$ 行,若 $(u_p, u_q) + (u_{p'}, u_{q'}) = (u_{p'}, u_{q'})$,则 (u_p, u_q) 行可以去掉。

对于决策表 $T=(U, C, D, V, f)$,二进制可辨矩阵 M_T 也可作与 M_I 相仿的变换而不影响约简,只不过在对二进制可辨矩阵 M_T 进行变换时,若 M_T 中有全为0的行,应首先把该行去掉。

2.3 二进制可辨矩阵的约简变换

根据以上的分析,二进制可辨矩阵的约简变换如下定义:

定义1 二进制可辨矩阵 M_T 或 M_I 的约简变换包括以下几种形式:

- (1)二进制可辨矩阵 M_T 中若有全为0的行,应首先去掉这样的行。
- (2)在变换过程中,可以随时将二进制可辨矩阵中出现的全为1的行去掉。
- (3)将二进制可辨矩阵中的行、列重新排序。
- (4)相同的行或列只出现一次。
- (5)对于某两列,如 a_i 列与 a_j 列,若 $a_i + a_j = a_i$ (“+”表示逻辑加),则 a_i 列可以去掉。
- (6)对于某两行,如 (u_p, u_q) 行与 $(u_{p'}, u_{q'})$ 行,若 $(u_p, u_q) + (u_{p'}, u_{q'}) = (u_{p'}, u_{q'})$,则 (u_p, u_q) 行可以去掉。

不难发现,二进制可辨矩阵 M_T 或 M_I ,经过以上变换之后,规模将会大大减小,并且列对应的属性将更接近于甚至等于属性(相对)约简。下面通过一个例子加以说明。

例1 考虑下面表1所表示的决策表。

表1 一个决策表

U \ A	a	b	c	d	e
1	1	2	0	1	1
2	1	2	0	1	1
3	2	0	0	1	0
4	0	0	1	2	1
5	2	1	0	2	1
6	0	0	1	2	2
7	2	0	0	1	0
8	0	1	2	2	1
9	2	1	0	2	2
10	2	0	0	1	0

其中 $C=\{a, b, c, d\}$ 是条件属性集, $D=\{e\}$ 是决策属性集。其二进制可辨矩阵 M_T 如表2所示。

表2 一个二进制可辨矩阵

(u_p, u_q)	a	b	c	d	(3,4)	1	0	1	1	(5,9)	0	0	0	0
(1,3)	1	1	0	0	(3,5)	0	1	0	1	(5,10)	0	1	0	1
(1,6)	1	1	1	1	(3,6)	1	0	1	1	(6,7)	1	0	1	1
(1,7)	1	1	0	0	(3,8)	1	1	1	1	(6,8)	0	1	1	0
(1,9)	1	1	0	1	(3,9)	0	1	0	1	(6,10)	1	0	1	1
(1,10)	1	1	0	0	(4,6)	0	0	0	0	(7,8)	1	1	1	1
(2,3)	1	1	0	0	(4,7)	1	0	1	1	(7,9)	0	1	0	1
(2,6)	1	1	1	1	(4,9)	1	1	1	0	(8,9)	1	0	1	0
(2,7)	1	1	0	0	(4,10)	1	0	1	1	(8,10)	1	1	1	1
(2,9)	1	1	0	1	(5,6)	1	1	1	0	(9,10)	0	1	0	1
(2,10)	1	1	0	0	(5,7)	0	1	0	1					

去掉全为0或1的行,且相同的行、列只出现一次,得到如下矩阵1:

$$\text{矩阵1} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

将一行加到另一行,若和仍等于这一行,将这一行去掉,结果为矩阵2:

$$\text{矩阵2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

可以发现,新矩阵中不存在只有一个元素为1,其余元素均为0的行,故相对核为 Φ 。一些列的逻辑和为 $(1, 1, 1, 1)^T$ ($(\dots)^T$ 表示转置),而它的任何子列都不满足这一条件,当然,这样的列对应的属性应为相对约简,在这个例子中相对约简为: $\{a, b\}, \{b, c\}, \{b, d\}, \{a, c, d\}$ 。

对于二进制可辨矩阵 M_I 或 M_T ,经过约简变换之后将得到规模大为减小的矩阵,为讨论方便,经过约简之后的矩阵记为 M'_I 或 M'_T ,则有如下命题:

命题4 在决策表 $T=(U, C, D, V, f)$ 中, $B \subseteq C$ 是 C 的一个 D 约简,其充要条件为:(1)在矩阵 M'_T 中,由 B 中各个属性所在列的逻辑和为 $(1, 1, \dots, 1)^T$; (2) $\forall B' \subset B, B'$ 不满足条件(1)。

命题5 在信息系统 $I=(U, A, V, f)$ 中, $B \subseteq A$ 是 A 的一个约简,其充要条件为:(1)在矩阵 M'_I 中,由 B 中各个属性所在列的逻辑和为 $(1, 1, \dots, 1)^T$; (2) $\forall B' \subset B, B'$ 不满足条件(1)。

值得注意的是,二进制可辨矩阵经过重复约简变换可以化为如下形式:

$$\begin{bmatrix} I & O \\ O & M \end{bmatrix}$$

其中 I 是单位矩阵, O 是元素全为0的矩阵, $M=(m_{ij})$, I 由简化的二进制可辨矩阵中只有一个元素为1,其余元素均为0的行经过重新排序构成,由于列经过消去变换,故它的列属性未必构成核或相对核,但一定包含核或相对核,并且它们一定属于某个约简。若 $M=\Phi$,则 I 中各个列属性构成了一个(相对)约简。若 $M \neq \Phi$,则首先把单位矩阵中元素1所在列对应的属性归入约简,并消去单位矩阵 I ,然后找出 M 中含元素1最少

的行,如果某行只有一个含1元素,则该元素1所在列对应的属性一定包含在属性的约简中,否则,就从这些行中元素1对应的列中找出元素1个数最多的列 c_{choice} ,并从 M 中消去所选列中元素1对应的各行及列 c_{choice} ,剩余部分再构成一个新的二进制可辨矩阵 M ,并把 c_{choice} 归入约简,重复上述步骤,直至 M 为空,依次选出的各列属性就构成了一个约简或相对约简。

3 基于二进制可辨矩阵的约简算法

在二进制可辨矩阵 M_i 或 M_T 中,对于那些只有一个元素为1其余元素均为0的行,元素1所在列的属性是必要的、不可缺少的,当然应当包含在(相对)约简中,而那些所含1的个数最多的列对应的属性虽未必必要,但具有很强的分辨能力,因此,在形成(相对)约简,尤其是最小(相对)约简的过程中具有重要地位。

基于上述考虑,下面给出几个在知识约简过程中具有重要作用的算法:

3.1 算法1(二进制可辨矩阵的化简算法:BDMR)

1. 二进制可辨矩阵 M_i 或 $M_T \rightarrow M$
2. 消去 M 中全为0的行,将得到 $p \times q$ 的新二进制可辨矩阵赋给 M ,置 $1 \rightarrow i$
3. 对于 $\forall k \neq i$,若 $(r_{i1}, r_{i2}, \dots, r_{iq}) + (r_{k1}, r_{k2}, \dots, r_{kq}) = (r_{k1}, r_{k2}, \dots, r_{kq})$ 则从 M 中消去第 k 行,并置 $p-1 \rightarrow p$,其中 $(r_{i1}, r_{i2}, \dots, r_{iq})$ 和 $(r_{k1}, r_{k2}, \dots, r_{kq})$ 分别为 M 中第 i 行和第 k 行,且 $k, i \leq p$
4. $i+1 \rightarrow i$, if $i \leq p$ then go 3 else go 6
5. $1 \rightarrow j$
6. 对于 $\forall l \neq j$, $(c_{1l}, c_{2l}, \dots, c_{pl}) + (c_{1l}, c_{2l}, \dots, c_{pl}) = (c_{1l}, c_{2l}, \dots, c_{pl})$ 则消去 M 的第 l 列,并置 $q-1 \rightarrow q$,其中 $(c_{1l}, c_{2l}, \dots, c_{pl})$ 和 $(c_{1l}, c_{2l}, \dots, c_{pl})$ 分别为 M 中第 j 列和第 l 列, $j, l \leq q$
7. $j+1 \rightarrow j$, if $j \leq q$ then go 6 else go 8
8. 将得到新的二进制可辨矩阵赋给 M

通过算法1,可以将二进制可辨矩阵的规模大大减少而不减少其中所蕴含的知识量,如(相对)核、(相对)约简等。下面通过一个例子说明该算法。

例2 考虑下面的一个信息系统,如表3。

表3 一个信息系统

U	a	b	c	d	e
1	1	0	2	1	1
2	2	1	0	1	0
3	2	1	2	0	2
4	1	2	2	1	1
5	1	2	0	0	2

首先构造生成二进制可辨矩阵 M_i ,如表4:

表4 一个二进制可辨矩阵

$(u_p, u_q) \setminus A$	a	b	c	d	e
(1,2)	1	1	1	0	1
(1,3)	1	1	0	1	1
(1,4)	0	1	0	0	0
(1,5)	0	1	1	1	1
(2,3)	0	0	1	1	1
(2,4)	1	1	1	0	1
(2,5)	1	1	0	1	1
(3,4)	1	1	0	1	1
(3,5)	1	1	1	0	0
(4,5)	0	0	1	1	1

将每一行与其它行进行逻辑加,若等于其它行,则消去其它

行,将得到的新矩阵3赋于 M ;

将每一列与其它列进行逻辑加,若等于这一列,则消去其它列,将得到的新矩阵4赋于 M ;

再用算法1对矩阵4进行化简,简化的二进制可辨矩阵恰为单位矩阵5,故它的一个约简为: $\{b, c\}$ 。

$$\text{矩阵3} \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \text{矩阵4} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{矩阵5} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3.2 算法2(属性约简算法:AR)

1. 调用算法1得到简化的 $p \times q$ 二进制可辨矩阵 M 。
2. 置矩阵 $M \rightarrow A$ 。用 Reduction 表示属性(相对)约简,初始值为 Φ 。
3. 将 A 的各行和各列进行横向和纵向相加,结果分别放入 $\text{Row}[i]$ 和 $\text{Col}[j]$ 中,其中 $i \leq p, j \leq q$
4. 求 $\text{Row}[i]$ 中的最小值: $\text{Row}_{\min} = \min\{\text{Row}[1], \text{Row}[2], \dots, \text{Row}[l]\}$, l 为矩阵 A 的行数
 - (1) if $\text{Row}_{\min} = 1$
 - 1) 求出所有使得 $\text{Row}_{\min} = 1$ 的各行 $\{r_1, r_2, \dots, r_w\}$ 及其元素1所对应的各列 $\{c_1, c_2, \dots, c_w\}$
 - 2) $\{c_1, c_2, \dots, c_w\} \rightarrow \text{Reduction}$
 - 3) 从 A 中消去各行 $\{r_1, r_2, \dots, r_w\}$ 及其元素1所对应的各列 $\{c_1, c_2, \dots, c_w\}$,将得到的新矩阵再赋给 A
 - 4) if $A \neq \emptyset$ then go 3 else go 5
 - (2) if $\text{Row}_{\min} > 1$
 - 1) $h = \text{Row}_{\min}$
 - 2) 求出所有使得 $\text{Row}_{\min} = h$ 的各行 $\{r_1, r_2, \dots, r_w\}$ 及其元素1所对应的各列 $L = \{c_{11}, c_{12}, \dots, c_{1h}, c_{21}, c_{22}, \dots, c_{2h}, \dots, c_{w1}, c_{w2}, \dots, c_{wh}\}$,在 L 中找出各列元素1的累积个数与 $\text{Col}[j]$ 中相对应的值
 - 3) 求出取得最大值 $\text{MAX}\{L\}$ 的列号 $\{c_{h2}, c_{h2}, \dots, c_{h2}\}$,若其中只有一列,则即为 c_{choice} ,否则,从 $\{c_{h2}, c_{h2}, \dots, c_{h2}\}$ 中任意取一个作为 c_{choice}
 - 4) $\text{Reduction} \cup \{c_{choice}\} \rightarrow \text{Reduction}$
 - 5) 从 A 中消去列 c_{choice} 中所有元素1所对应的行及其列 c_{choice} 后,得到的新矩阵再赋给 A
 - 6) if $A \neq \emptyset$ then go 3 else go 5
5. 输出一个属性(相对)约简 Reduction

在如上给出的 AR 算法中,总是首先将那些只有一个元素为1,其余元素均为0的行中元素1所在的列属性归入约简,然后将包含1的个数最多的列对应的属性归入约简。这样做,也许可以保证“收敛”的速度更快,并且可望用此方法得到最小属性(相对)约简。不难发现,AR 算法是完备的,这个算法可以用到例1中求得一个属性相对约简。

3.3 算法3(属性(相对)约简判定的完备算法:RJ)

1. 置二进制可辨矩阵 M_i 或 $M_T \rightarrow A$,判定 $\text{Reduction} = \{c_{r1}, c_{r2}, \dots, c_{rk}\}$ 是否为一个属性(相对)约简。
2. 消去列 $\{c_{r1}, c_{r2}, \dots, c_{rk}\}$ 及其这些列上元素1所在的各行,得到的新矩阵赋给 A ,若 $A \neq \Phi$,则输出“Reduction 不是属性(相对)约简”,否则,转第3步。

3. $1 \rightarrow i$

4. $\text{Reduction} - \{c_n\} \rightarrow \text{Reduction}$,消去 Reduction 中所有属性对应的各列及其这些列上元素1所在的各行,得到的新矩

(下转封四)

(上接第 142 页)

阵赋给 A, If $A = \Phi$ then Reduction 不是属性(相对)约简, 结束 else go 5.

5. $i+1 \rightarrow i$, if $i \leq k$ then go 4 else go 6

6. 如果一直进行下去, 得到的矩阵 A 都是非空的, 则输出“Reduction 是属性(相对)约简”。

可以证明, 该算法是完备的, 因此, 我们称它为属性约简判定的完备算法。

例 3 考虑例 2 中 $\{a, b\}, \{b, d\}$ 是否为一个属性约简?

1. $M_1 \rightarrow A$, Reduction = $\{a, b\}$

2. 消去列 a, b 及其这两列上元素 1 所在的各行, 得到的新矩阵赋给 A 为:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

因为 $A \neq \Phi$, 按照 RJ 算法, 得出 Reduction = $\{a, b\}$ 不是属性约简。若 Reduction = $\{b, d\}$, 则重做第 2 步, 得出 $A = \Phi$ 。

3. 从 Reduction 中消去 b, 得到 Reduction = $\{d\}$, 消去 d 列及其该列上元素 1 所在的各行, 得到矩阵 6。

4. 从 Reduction 中消去 d, 得到 Reduction = $\{b\}$, 消去 b 列及其该列上元素 1 所在的各行, 得矩阵 7。

$$\text{矩阵 6} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{矩阵 7} \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

由于得出的两个新矩阵均非空, 所以可以判定 Reduction = $\{b, d\}$ 是一个属性约简。

4 算法复杂度分析

设决策系统(信息系统) T 中有 M 个条件属性, N 个对象, 在最坏情况下, 构造二进制可辨矩阵需要比较 $MN(N-1)/2$ 次, 复杂度为 $O_1(MN^2)$; 算法 1(即 BDMR 算法)的复杂度为 $O_2(M^2+N^4)$; 算法 2(即 AR 算法)和算法 3(即 RJ 算法)的复杂度主要是由 T 的条件属性组合引起的, 在最坏情况下, 每次考虑的属性数依次为 M, M-1, ..., 1, 故总次数为 $M+(M-1)+\dots+1=M(M+1)/2$, 即算法 2 和算法 3 的复杂度分别为 $O_3(M^2)$ 和 $O_4(M^2)$ 。于是本文属性约简的复杂度为

$\text{MAX}\{O_1(MN^2), O_2(M^2+N^4), O_3(M^2), O_4(M^2)\} = O(M^2+N^4)$ 。因此, 在最坏情况下, 本文算法能够在 $O(M^2+N^4)$ 时间复杂度内找到满意的约简。

上述分析中, 我们考虑的是最坏情况, 实际上, 二进制可辨矩阵经过算法 1 的变换之后, 规模将会大大减少, 并且列对应的属性将更接近于甚至等于属性(相对)约简, 算法是高效率的。

结束语 本文对二进制可辨矩阵在保持属性约简后其分类能力不变的前提下进行了各种变换, 而这些变换又相应于属性的约简过程, 通过这些变换, 一方面可以大大减少 BDM 的存储空间, 另一方面有助于构造高效的属性约简算法, 基于 BDM 的概念, 又提出了 BDMR 算法、AR 算法和 RJ 算法, 并举例说明三个算法的可行性, 对于 AR 算法, 我们可望得到一个最小属性(相对)约简。粗糙集理论是全新的理论, 而属性约简与约简判定在这一领域中占有重要地位, 如何给出求得最小约简的完备算法, 仍然是一个未解决的问题, 本文并不是这一问题的终极解答, 若能对这一问题有所启发, 则是我们希望所至。

另外, 二进制可辨矩阵中的元素均为 1 或 0, 即二进制, 我们可以利用这一特点对二进制可辨矩阵的元素进行各种布尔运算, 并利用计算机的并行位操作对本文的属性约简算法作进一步的改进, 有关工作正在进行中。

参考文献

- 1 Pawlak Z. Rough sets-theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publishers, 1991
- 2 Felix R, Ushio T. Rough Sets-based Machine Learning Using a Binary Discernibility Matrix. IPMM'99 published, 1999. 299~305
- 3 苗夺谦, 胡桂荣. 知识约简的一种启发式算法. 计算机研究与发展, 1999, 36(6): 681~684
- 4 王珏, 苗夺谦, 等. 关于 Rough Set 理论与应用的综述. 模式识别与人工智能, 1996, 9(4): 337~344
- 5 王珏, 王任, 苗夺谦, 等. 基于 Rough Set 理论的“数据浓缩”[J]. 计算机学报, 1998, 21(5): 393~400

计算机科学

(1974 年 2 月创刊)

第 29 卷第 2 期 (月刊)

2002 年 2 月 25 日出版

中国标准刊号: ISSN 1002-137X
CN50-1075/TP

定价: 18.00 元 国外定价: 5 美元

邮发代号: 78-68

发行范围: 国内外公开

主管单位: 国家科学技术部

主办单位: 国家科技部西南信息中心

编辑出版: 《计算机科学》杂志社

重庆市渝中区胜利路 132 号 邮政编码: 400013

电话: (023) 63500828 E-mail: jsjxx@swic.ac.cn

主 编: 朱宗元

印刷者: 重庆科情印务有限公司

总发行处: 重庆市邮政局

订购处: 全国各地邮政局

国外总发行: 中国国际图书贸易总公司 (北京 399 信箱)

国外代号: 6210-MO