序目标信息系统中分布约简的矩阵算法*

张晓燕[1], 徐伟华[1],张文修[2]

(1. 重庆理工大学 数理学院,重庆 400054;2. 西安交通大学 理学院,西安 710049)

摘 要:在序信息系统中引入了优势矩阵和目标分布矩阵的概念,建立了此类信息系统中分布约简的矩阵算法,同时通过实例分析验证了该算法的有效性,其优点是对数据复杂的信息表可相对容易地给出所有的分配约简,为序信息系统分配约简提供了一种便捷操作的属性约简方法.

关 键 词:粗糙集;信息系统;分布约简;矩阵计算

中图分类号:TP18 **文献标识码**:A **文章编号**:1674 -8425(2010)03 -0056 -06

Matrix Computation for Distribution Reduction in Ordered Information Systems

ZHANG Xiao-yan[1], XU Wei-hua[1], ZHANG Wen-xiu[2]

(1. School of Mathematics and Physics, Chongqing Univtrsity of Technology, Chongqing 400054, China;
2. School of Sciences, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: This paper studies the problem of matrix computation for distribution reduction in ordered information systems. The dominance matrix and decision distribution matrix are introduced in ordered information systems. Furthermore, the algorithm of distribution reduction is obtained, from which we can provide new approach to attributes reductions in inconsistent ordered information systems. Finally, an example illustrates the validity of this method and shows that the method is applicable to a complex information system.

Key words: rough set; information system; distribution reduction; matrix computation

粗糙集理论[1]是近年来发展起来的一种处理不精确性、不确定性和模糊知识的软计算工具,已被成功地应用于人工智能、数据挖掘、模式识别与智能信息处理等领域[2-5],并越来越引起国际学术界的关注。经典粗糙集是以完备信息系统为研究对象,以等价关系(满足自反性、对称性、传递性)为基础,通过等价关系对论域分成互不相交的等价类,划分越细,知识越丰富,信息越充分。

属性约简是粗糙集理论的核心问题之一. 在实际的知识库中描述知识的属性并不是同等重要的,甚至其中有些属性是冗余的。所谓属性约简,就是在保持知识库分类能力不变的条件下,删除其中不相关或不重要的属性. 通过属性约简去掉不必要的属性,可以既使知识表示简化,又不丢

* **收稿日期**:2009 -09 -09
项目基金:重庆市教委科学技术研究项目(KJ090612)
作者简介:张晓燕(1979—),女,山西怀仁人,硕士,讲师,主要从事人工智能的数学基础、动力系统等方面研究.

失基本信息。目前，许多学者通过不同的方法从不同的角度对属性约简作了深入的研究，并取得了很多成果[6-11]。然而，这些研究主要是在等价关系下的信息系统进行的，而在实际问题中有许多信息系统由于各种原因(如噪声、信息缺损等)是基于优势关系的，而且是不协调的。要想从这种复杂的基于优势关系的不协调信息系统中获取简洁的不确定性命题，就必须对系统进行属性约简,因而对优势关系下的不协调目标信息系统属性约简的研究是非常有意义的[12-15]。为此，本文中对这一问题进行了探讨,在基于优势关系下的信息系统中引入了优势矩阵和目标分布矩阵的概念，并进一步建立了优势关系下信息系统分布约简的矩阵算法,同时通过实例分析验证了该算法的有效性,并说明了其优点是对数据复杂的信息表也可相对容易地给出所有的分配约简。该方法提供了在优势关系下信息系统分配约简的便捷操作方法。

1 序目标信息系统及其分布约简

目标信息系统是既有条件属性又有目标属性(决策属性)的一种特殊信息系统。目标信息系统主要是研究条件属性和目标属性之间的关系问题。为了方便理解，下面先给出一些基本概念。

定义 1[7] 称一个五元组 $I=(U,A,F,D,G)$ 为一个目标信息系统，其中:(U,A,F) 是信息系统;A 称为条件属性集;D 称为目标属性集,即: U 是有限对象集，$U=\{x_1,x_2,\cdots,x_n\}$;A 是有限条件属性集，$A=\{a_1,a_2,\cdots,a_p\}$;D 是有限目标属性集，$D=\{d_1,d_2,\cdots,d_q\}$; F 是 U 与 A 的关系集，$F=\{f_k:U\to V_k,k\leqslant p\}$，$V_k$ 是 a_k 的有限值域;G 是 U 与 D 的关系集，$G=\{g_{k'}:U\to V_k{}',k'\leqslant q\}$，$V_k{}'$是 $d_{k'}$ 的有限值域。

我们知道,在 Pawlak 近似空间意义下的信息系统，对每个属性集和目标属性集就决定了一个二元不可区分关系,即等价关系。然而，在实际生活中有许多系统并不是基于等价关系的，有不少是基于优势关系的，即对每个属性值域和目标属性值域有一个偏序关系，如一个班级成员的各科成绩情况等问题。这时就需要建立基于优势关系下的信息系统。

定义 2[7] 设 $I=(U,A,F,D,G)$ 为目标信息系统，对于 $B\subseteq A$，令

$R_B^{\leqslant}=\{(x_i,x_j)\in U\times U:f_l(x_i)\leqslant f_l(x_j),\forall a_l\in B\}$

$R_D^{\leqslant}=\{(x_i,x_j)\in U\times U:g_m(x_i)\leqslant g_m(x_j),\forall d_m\in D\}$

其中 $R_B^{\leqslant},R_D^{\leqslant}$ 称为目标信息系统的优势关系，此时该目标信息统称为是序目标信息系统。

记:

$[x_i]_B^{\leqslant}=\{x_j\in U:(x_i,x_j)\in R_B^{\leqslant}\}=\{x_j\in U:f_l(x_i)\leqslant f_l(x_j),\forall a_l\in B\}$

$[x_i]_D^{\leqslant}=\{x_j\in U:(x_i,x_j)\in R_D^{\leqslant}\}=\{x_j\in U:g_m(x_i)\leqslant g_m(x_j),\forall d_m\in D\}$

易见优势关系有下面性质:

命题 1[7]

1) $R_B^{\leqslant}$ 是自反的和传递的，未必是对称的，因而一般不再是等价关系。

2) 当 $B_1\subseteq B_2\subseteq A$ 时,有 $R_A^{\leqslant}\subseteq R_{B_2}^{\leqslant}\subseteq R_{B_1}^{\leqslant}$。

3) 当 $B_1\subseteq B_2\subseteq A$ 时,有$[x_i]_{B_2}^{\leqslant}\subseteq[x_i]_{B_1}^{\leqslant}$。

4) 当 $x_j\in[x_i]_B^{\leqslant}$ 时,有$[x_j]_B^{\leqslant}\subseteq[x_i]_B^{\leqslant}$。

对于任意 $X\subseteq U$，定义 X 关于优势关系下 $R_B^{\leqslant}$ 的下近似和上近似分别为:

$\underline{R_B^{\leqslant}}(X)=\{x_i\in U:[x_i]_B^{\leqslant}\subseteq X\}$

$\overline{R_B^{\leqslant}}(X)=\{x_i\in U:[x_i]_B^{\leqslant}\cap X\neq\varnothing\}$

优势关系下的上、下近似也满足类似于 Pawlak 近似空间中的许多性质,详细请参考文献[7]。

为了叙述方便,本文中在没有特别说明时信息系统都是指基于优势关系下的信息系统。

定义 3[7] 设 $I=(U,A,F,D,G)$ 为序目标信息系统,若 $R_A^{\leqslant}\subseteq R_D^{\leqslant}$,则称该序目标信息系统是协调的，否则若 $R_A^{\leqslant}\varnothing R_D^{\leqslant}$,则称该系统是不协调的。

文献[15]中在基于优势关系的信息系统中引入了分布约简的概念,并给出了分布约简的判定定理和辨识矩阵。下面简单列出有关概念。

设 $I=(U,A,F,D,G)$ 为序目标信息系统，$R_B^{\leqslant}$，$R_D^{\leqslant}$ 分别为属性集 A 和目标属性集 D 生成的 U 上的优势关系,对于 $B\subseteq A$,记

$$\mu_B(x)=\left(\frac{|D_1\cap[x]_B^{\leqslant}|}{|U|},\frac{|D_2\cap[x]_B^{\leqslant}|}{|U|},\cdots,\frac{|D_r\cap[x]_B^{\leqslant}|}{|U|}\right)$$

其中$[x]_B^{\leqslant}=\{y\in U:(x,y)\in R_B^{\leqslant}\}$,称 $\mu_B(x)$ 为论

域 U 上的关于属性子集 B 的分布函数。

定义 4[15] 设 $\alpha=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$ 和 $\beta=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$ 为 2 个 n 维 $n\times 1$ 向量,有下面定义:

若 $a_i=b_i(i=1,\cdots,n)$,称向量 α 等于向量 β,记做 $\alpha=\beta$。

若 $a_i\leqslant b_i(i=1,\cdots,n)$,称向量 α 小于等于向量 β,记做 $\alpha\leqslant\beta$;否则若存在某个 $i_0,i_0\in\{1,2,\cdots,n\}$,使得 $a_{i_0}>b_{i_0}$,称向量 α 不小于等于向量 β,记做 $\alpha\nleqslant\beta$。

如 $(1,2,3)\nleqslant(1,1,4)$,且 $(1,1,4)\nleqslant(1,2,3)$。

显然,由以上结论可立即得到下面命题:

命题 2[15]

1) 当 $B\subseteq A$ 时,对任意的 $x\in U$ 有 $\mu_A(x)\leqslant\mu_B(x)$。

2) 对 $\forall x,y\in U$,当 $[y]_B^{\leqslant}\subseteq[x]_B^{\leqslant}$ 时,有 $\mu_B(y)\leqslant\mu_B(x)$。

定义 5[15] 设 $I=(U,A,F,D,G)$ 为目标信息系统。若 $\forall x\in U$,有 $\mu_B(x)=\mu_A(x)$,则称 B 是分布协调集,如 B 的任何真子集不是分布协调集,则称 B 为分布协调约简。

例 1[7] 表 1 给出了一个序目标信息系统。于是,按照优势关系的定义有:

$[x_1]_A^{\leqslant}=\{x_1,x_2,x_5,x_6\}$,$[x_2]_A^{\leqslant}=\{x_2,x_5,x_6\}$,

$[x_3]_A^{\leqslant}=\{x_2,x_3,x_4,x_5,x_6\}$,$[x_4]_A^{\leqslant}=\{x_4,x_6\}$,

$[x_5]_A^{\leqslant}=\{x_5\}$,$[x_6]_A^{\leqslant}=\{x_6\}$,

$[x_1]_d^{\leqslant}=[x_5]_d^{\leqslant}=\{x_1,x_5\}$,

$[x_2]_d^{\leqslant}=[x_4]_d^{\leqslant}=\{x_1,x_2,x_4,x_5\}$,

$[x_3]_d^{\leqslant}=[x_6]_d^{\leqslant}=\{x_1,x_2,x_3,x_4,x_5,x_6\}$

表 1 某序目标信息系统

U	a_1	a_2	a_3	d
x_1	1	2	1	3
x_2	3	2	2	2
x_3	1	1	2	1
x_4	2	1	3	2
x_5	3	3	2	3
x_6	3	2	3	1

显然 $R_A^{\leqslant}\varnothing R_d^{\leqslant}$,因此该序目标信息系是不协调的。

若在例 1 的信息系统中记:

$D_1=[x_1]_d^{\leqslant}=[x_5]_d^{\leqslant}$,$D_2=[x_2]_d^{\leqslant}=[x_4]_d^{\leqslant}$,

$D_3=[x_3]_d^{\leqslant}=[x_6]_d^{\leqslant}$

则有

$\mu_A(x_1)=\left(\frac{1}{3},\frac{1}{2},\frac{2}{3}\right)$;$\mu_A(x_2)=\left(\frac{1}{6},\frac{1}{3},\frac{1}{2}\right)$;

$\mu_A(x_3)=\left(\frac{1}{6},\frac{1}{2},\frac{5}{6}\right)$;$\mu_A(x_4)=\left(0,\frac{1}{6},\frac{1}{3}\right)$;

$\mu_A(x_5)=\left(\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)$;$\mu_A(x_6)=\left(0,0,\frac{1}{6}\right)$。

在文献[15]中通过利用辨识矩阵得到例 1 只有一个分布约简 $\{a_2,a_3\}$。

本文中未经特别说明信息系统均指不协调目标信息系统。

2 优势矩阵与目标分布矩阵

定义 6 设 $I=(U,A,F,D,G)$ 为序目标信息系统,其中 $U=\{x_1,x_2,\cdots,x_n\}$。若记

$$M_B=(m_{ij})_{n\times n},m_{ij}=\begin{cases}1,x_j\in[x_i]_B^{\leqslant}\\0,\text{其他}\end{cases}$$

则称矩阵 M_B 为条件属性子集 $B\subseteq A$ 的优势矩阵;若记 $|B|=l$,也称 M_B 是信息系统 $I=(U,A,F,D,G)$ 的 l 级优势矩阵。

下面给出 2 个条件属性子集优势矩阵交的定义。

定义 7 设 $I=(U,A,F,D,G)$ 为目标信息系统,且 $B,C\subseteq A$,优势矩阵分别为 M_B 和 M_C。定义它们的交如下:

$$M_B\cap M_C=(m_{ij})_{n\times n}\cap(m_{ij}')_{n\times n}=(min\{m_{ij},m_{ij}'\})_{n\times n}$$

由上面定义可以直接得到:

命题 3 优势矩阵有下面性质:

1) $m_{ii}=1$。

2) 若 $B,C\subseteq A$,则 $M_{B\cup C}=M_B\cap M_C$。

定义 8 设 $I=(U,A,F,D,G)$ 为目标信息系统,其中 $U=\{x_1,x_2,\cdots,x_n\}$。若记:

$$M_D=(r_{ij})_{n\times n},r_{ij}=\begin{cases}1,\mu_A(x_j)\leqslant\mu_A(x_i)\\0,\text{其他}\end{cases}$$

则称矩阵 M_D 为信息系统 I 的目标分布矩阵。

从定义 6 和 7 可以看出条件属性子集优势矩阵确定了信息系统中所有对象的优势关系,而目标分布矩阵则描述了所有对象之间确定不同目标的区分关系。

例 2 计算表 1 给出的信息系统的几个条件

属性子集的优势矩阵和目标分布矩阵。

$$M_A=\begin{bmatrix}1&1&0&0&1&1\\0&1&0&0&1&1\\0&1&1&1&1&1\\0&0&0&1&0&1\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}$$

$$M_D=M_{\{d\}}=\begin{bmatrix}1&1&0&1&1&1\\0&1&0&1&1&1\\0&1&1&1&1&1\\0&0&0&1&0&1\\0&0&0&0&1&1\\0&0&0&0&0&1\end{bmatrix}$$

定义 9 设有矩阵 $M_A=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm T}$ 和 $M_B=(\beta_1,\beta_2,\cdots,\beta_n)^{\mathrm T}$,其中 α_i 和 $\beta_i(i=1,\cdots,n)$ 都是 n 维 $n\times1$ 向量。若 $\alpha_i\leqslant\beta_i(i=1,\cdots,n)$,则称 M_A 小于 M_B,记做 $M_A\leqslant M_B$。

显然优势矩阵有下面性质:

命题 4 设 $I=(U,A,F,D,G)$ 为目标信息系统,且 $B\subseteq A$,优势矩阵分别为 M_B 和 M_A,则有 $M_A\leqslant M_B$。

3 分布矩阵的矩阵算法

设 $I=(U,A,F,D,G)$ 为目标信息系统,且 $B\subseteq A$,优势矩阵分别为 M_B 和 M_A,目标分布矩阵是 M_D。下面首先给出本文中算法的理论根据。

定理 1 信息系统 $I=(U,A,F,D,G)$ 中,若 $B\subseteq A$,则 B 是分不协调集充要条件是 $M_B\leqslant M_D$。

证明 “$\Rightarrow$”需证明:若对 $\forall x\in U$ 有 $\mu_B(x)=\mu_A(x)$,则 $M_B\leqslant M_D$ 成立。由定义 8 可知只需证当 $m_{ij}=1$ 时有 $r_{ij}=1$。事实上有

$m_{ij}=1\Rightarrow x_j\in[x_i]_B^{\leqslant}\Rightarrow\mu_B(x_j)\leqslant\mu_B(x_i)\Rightarrow\mu_A(x_j)\leqslant\mu_A(x_i)\Rightarrow r_{ij}=1$

“$\Leftarrow$”:对任意的 $x_i\in U$,由于 $B\subseteq A$,由命题 2 有 $\mu_A(x_i)\leqslant\mu_B(x_i)$。

下面只需证 $\mu_B(x_i)\leqslant\mu_A(x_i)$。

显然当 $\mu_B(x_i)=0$ 时 $\mu_B(x_i)\leqslant\mu_A(x_i)$ 成立。

当 $\mu_B(x_i)\neq0$ 时,则存在某个 $D_k(k\in\{1,2,\cdots,r\})$,使得对任意的 $x_i\in U$ 有 $\dfrac{|D_k\cap[x_i]_B^{\leqslant}|}{|U|}\neq0$,故此时有 $|D_k\cap[x_i]_B^{\leqslant}|\neq0$,即 $D_k\cap[x_i]_B^{\leqslant}\neq\Phi$。不妨设 $x_j\in D_k\cap[x_i]_B^{\leqslant}$,则 $x_j\in D_k$ 且 $x_j\in[x_i]_B^{\leqslant}$。由 $x_j\in[x_i]_B^{\leqslant}$ 便得 $m_{ij}=1$。又 $M_B\leqslant M_D$,于是可知 $r_{ij}=1$,即有 $\mu_A(x_j)\leqslant\mu_A(x_i)$。

另外又由于 $x_j\in[x_j]_A^{\leqslant}$,则 $x_j\in D_k\cap[x_j]_A^{\leqslant}$,故有 $|D_k\cap[x_i]_B^{\leqslant}|\leqslant|D_k\cap[x_j]_A^{\leqslant}|$,即有 $\mu_B(x_i)\leqslant\mu_A(x_j)$。

因此 $\mu_B(x_i)\leqslant\mu_A(x_i)$ 成立。

推论 1 信息系统 $I=(U,A,F,D,G)$ 中,若 $B\subseteq A$,则 B 是 I 的分布约简充要条件是 $M_B\leqslant M_D$,且 B 的任意真子集 B 都不满足 $M_{B'}\leqslant M_D$。

设 $I=(U,A,F,D,G)$ 为目标信息系统,其中 $U=\{x_1,x_2,\cdots,x_n\}$。若条件属性子集 B 所确定的优势矩阵记为 $M_B=(\beta_1,\beta_2,\cdots,\beta_n)^{\mathrm T}$,$I$ 的目标分布矩阵记为 $M_D=(\gamma_1,\gamma_2,\cdots,\gamma_n)^{\mathrm T}$,其中 $\beta_i,\gamma_i(i=1,\cdots,n)$ 分别是指 M_B 和 M_D 的第 i 行元素做成的 $n\times1$ 列向量,上标 T 表示矩阵的转置,则由以上定理的思想便可以获得一个序信息系统分布约简的矩阵算法,具体算法描述如下。

算法 1 优势关系下的信息系统分布约简的矩阵算法。

输入 一个优势关系下信息系统 $I=(U,A,F,D,G)$ 的信息表,其中:$U=\{x_1,x_2,\cdots,x_n\}$;$A=\{a_1,a_2,\cdots,a_p\}$。

输出 信息系统 $I=(U,A,F,D,G)$ 的所有分布约简。

Step1 将信息表中完全相同的行进行合并,并记录频数。

Step2 计算 $I=(U,A,F,D,G)$ 的目标分布矩阵 $M_D=(\gamma_1,\gamma_2,\cdots,\gamma_n)^{\mathrm T}$。

Step3 对每个 $a_l\in A,(1\leqslant l\leqslant p)$ 计算其一级优势矩阵 $M_{a_l}=M_{a_l}{}^{(1)}=(\tau_1{}^{(1)},\tau_2{}^{(1)},\cdots,\tau_n{}^{(1)})^{\mathrm T}$,For $i=1$ to n,若 $0\neq\tau_i{}^{(1)}\leqslant\gamma_i$,则令 $\tau_i{}^{(1)}=0$,并将得到的新矩阵记做 $FM_{\{a_l\}}^{(1)}$,进入下一步。

Step4 将上一步得到的

$FM_{\{a_l\}}^{(1)}=(\tau_1^{(1)},\tau^{(1)},_2\cdots,\tau_n^{(1)})^{\mathrm T},a_l\in A,(1\leqslant l\leqslant p)$

称为一级分布矩阵,若 $FM_{\{a_l\}}{}^{(1)}=0$,则输出一个一级分布约简 $\{a_l\}$,否则进入下一步。

Step5 由 Step3 得到的所有非 0 一级分布矩阵求交得到所有的二级等价优势矩阵:

$M_{\{a_la_s\}}^{(2)},(M_{\{a_la_s\}}^{(2)}\neq M_{\{a_l\}}^{(1)},M_{\{a_la_s\}}^{(2)}\neq M_{\{a_s\}}^{(1)})$

对所有的二级分布矩阵用 Step3 的方法求出所有的二级分布约简。

Step6 重复 Step5 可得到三级以上的分布约简，直到 0 矩阵结束。

可以容易分析出述算法的时间复杂度为 $O(|U|^2 2^{|A|})$。

4 实例分析

本节中将利用算法 1 来说明该算法的有效性。仍然考虑例 1 给出的优势关系下的目标信息系统。

由于在表 1 中没有完全相同的行，故 Sptep1 省略。目标分布矩阵已在例 2 中给出。下面分别计算每个 $a_l \in A, 1 \leqslant l \leqslant 3$ 的优势矩阵：

$$M_{\{a_1\}}=\begin{bmatrix}1&1&1&1&1&1\\0&1&0&0&1&1\\1&1&1&1&1&1\\0&1&0&1&1&1\\0&1&0&0&1&1\\0&1&0&0&1&1\end{bmatrix}$$

$$M_{\{a_2\}}=\begin{bmatrix}1&1&0&0&1&1\\1&1&0&0&1&1\\1&1&1&1&1&1\\1&1&1&1&1&1\\0&0&0&0&1&0\\1&1&0&0&1&1\end{bmatrix}$$

$$M_{\{a_3\}}=\begin{bmatrix}1&1&1&1&1&1\\0&1&1&1&1&1\\0&1&1&1&1&1\\0&0&0&1&0&1\\0&1&1&1&1&1\\0&0&0&1&0&1\end{bmatrix}$$

$$M_D=M_{\{d\}}=\begin{bmatrix}1&1&0&1&1&1\\0&1&0&1&1&1\\0&1&1&1&1&1\\0&0&0&1&0&1\\0&0&0&0&1&1\\0&0&0&0&0&1\end{bmatrix}$$

分别将优势矩阵 $M_{\{a_1\}}, M_{\{a_2\}}, M_{\{a_3\}}$ 与目标分布矩阵 $M_{\{d\}}$ 的行进行对比，发现没有一级分布约简，于是一级分布矩阵为

$$FM^{(1)}_{\{a_1\}}=\begin{bmatrix}1&1&1&1&1&1\\0&0&0&0&0&0\\1&1&1&1&1&1\\0&1&0&1&1&1\\0&1&0&0&1&1\\0&1&0&0&1&1\end{bmatrix}$$

$$FM^{(1)}_{\{a_2\}}=\begin{bmatrix}0&0&0&0&0&0\\1&1&0&0&0&0\\1&1&1&1&1&1\\1&1&1&1&1&1\\0&0&0&0&0&0\\1&1&0&0&1&1\end{bmatrix}$$

$$FM^{(1)}_{\{a_3\}}=\begin{bmatrix}1&1&1&1&1&1\\0&1&1&1&1&1\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&1&1&1&1&1\\0&0&0&1&0&1\end{bmatrix}$$

下面构造二级优势矩阵：

$$M^{(2)}_{\{a_1,a_2\}}=FM^{(1)}_{\{a_1\}}\cap FM^{(1)}_{\{a_2\}}=\begin{bmatrix}0&0&0&0&0&0\\0&0&0&0&0&0\\1&1&1&1&1&1\\0&1&0&1&1&1\\0&0&0&0&0&0\\0&1&0&0&1&1\end{bmatrix}$$

$$M^{(2)}_{\{a_1,a_3\}}=FM^{(1)}_{\{a_1\}}\cap FM^{(1)}_{\{a_3\}}=\begin{bmatrix}1&1&1&1&1&1\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&1&0&0&1&1\\0&0&0&0&0&1\end{bmatrix}$$

$$M^{(2)}_{\{a_2,a_3\}}=FM^{(1)}_{\{a_2\}}\cap FM^{(1)}_{\{a_3\}}=\begin{bmatrix}0&0&0&0&0&0\\0&1&0&0&1&1\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&1\end{bmatrix}$$

将 $M^{(2)}_{\{a_1,a_2\}}$ 与 $M_{\{d\}}$ 进行对比，发现 $M^{(2)}_{\{a_1,a_2\}}$ 的第 3,4,6 行向量并没有 $M_{\{d\}}$ 中相应的行向量小，故 $\{a_1,a_2\}$ 不是该系统的二级分布约简；同理可知 $\{a_1,a_3\}$ 也不是该系统的二级分布约简。然而发现 $M^{(2)}_{\{a_2,a_3\}}$ 的所有行都比 $M_{\{d\}}$ 的小，所以可知该系

统的所有二级分布约简为$\{a_2,a_3\}$。

令

$$FM^{(2)}_{\{a_2,a_3\}}=\begin{bmatrix}0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\end{bmatrix}$$

于是算法结束。

因此由算法 1 可知例 1 的信息系统的所有分布约简是$\{a_2,a_3\}$,这与文献[15]中的结果是一致的，而且从上面分析中可以看出,即使是对象集与属性集等数据都非常复杂的庞大信息表该算法也是容易进行操作的，只需要按照算法 1 编写程序计算,这样就方便了许多。

5 结束语

如要想从复杂的基于优势关系的不协调信息系统中获取简洁的不确定性命题,就必须对系统进行知识约简,因此对于优势关系下的不协调目标信息系统的知识约简的研究是非常有意义的。本文中在基于优势关系下的信息系统中引入了优势矩阵和目标分布矩阵的概念，并进一步建立了文献[15]中提出的优势关系下信息系统分布约简的矩阵算法,同时通过实例分析验证了该算法的有效性,并说明了其优点是对数据复杂的信息表也可相对容易地给出所有的分布约简。该方法提供了在优势关系下信息系统知识分布约简的便捷操作方法，而且对一般的经典等价关系下的信息系统也是完全适用的。

参考文献:

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

[2] Pawlak Z. Rough Sets[J]. Communication of the ACM, 1995,38(1): 89 -95.

[3] 王国胤. Rough 集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001.

[4] 苗夺谦,王珏. 基于粗糙集的多变量决策树构造方法[J]. 软件学报, 1997, 8(6): 425 -431.

[5] 米据生,吴伟志,张文修. 不协调目标信息系统知识约简的比较研究[J]. 模糊系统与数学, 2003,17(3):54 -60.

[6] 王珏,苗夺谦,周育健. 关于 Rough Set 理论与应用的综述[J]. 模式识别与人工智能,1996,9(4): 337 -344.

[7] 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.

[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.

[9] Kryszkiewicz M. Comparative Studies of Alternative of Knowledge Reduction in Inconsistent Systems [J]. Intelligent Systems, 2001,16(1): 105 -120.

[10] Grecos M B, Slowinski R. Rough Approximation of Preference Relation by Dominance Relations [J]. European Journal of Operational Research, 1999,117: 63 -68.

[11] Shao M W, Zhang W X. Dominance relation and rules in an incomplete ordered information system[J]. International Journal of Intelligent Systems, 2005(20): 13 -27.

[12] Xu W H,Zhang W X. Methods for knowledge reduction in inconsistent ordered information systems [J]. Journal of Applied Mathematics & Computing,2008,26(1/2):313 -323.

[13] 徐伟华,张文修. 基于优势关系信息系统分配约简的矩阵算法[J]. 计算机工程,2007, 33(14): 4 -7.

[14] 徐伟华,张文修. 基于优势关系下不协调目标信息系统的知识约简[J]. 计算机科学,2006,33(2):182 -184.

[15] 徐伟华,张文修. 基于优势关系下不协调目标信息系统的分布约简[J]. 模糊系统与数学,2007,21(4): 124 -131.

(责任编辑 刘 舸)