第21卷第4期 2007年8月 | 模糊系统与数学 Fuzzy Systems and Mathematics | Vol. 21, No. 4 Aug., 2007

文章编号:1001-7402(2007)04-0124-08

基于优势关系下不协调目标信息系统的分布约简*

徐伟华,张文修

(西安交通大学 理学院 信息与系统科学研究所,陕西 西安 710049)

摘 要:在基于优势关系下的不协调目标信息系统中引入了分布约简和最大分布约简的概念,并讨论了二者之间的关系,而且得到了分布和最大分布约简的判定定理以及辨识矩阵,建立了不协调目标信息系统的分布和最大分布约简的具体方法,同时通过实例验证了该方法的有效性,从而进一步丰富了粗糙集理论。

关键词:粗糙集;信息系统;分布约简;最大分布约简;辨识矩阵

中图分类号:TP18 **文献标识码**:A

1 引言

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,它已被成功的应用于人工智能、数据挖掘、模式识别与智能信息处理等领域[2-5],并越来越引起了国际学术界的关注。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。

知识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。所谓知识约简,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。通过知识约简去掉不必要的属性,可以使知识表示简化,又不丢失基本信息。目前,许多学者通过不同的方法从不同的角度对知识约简做了深入的研究,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,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,就必须对系统进行知识约简。因而,对于优势关系下的不协调目标信息系统知识约简的研究是非常有意义的[12,13]。为此,本文对这一问题进行了探讨研究,在基于优势关系下的不协调目标信息系统中引入了分布约简和最大分布约简的概念,并讨论了二者之间的关系,而且得到了分布和最大分布约简的判定定理以及辨识矩阵,建立了不协调目标信息系统的分布和最大分布约简的具体方法,同时通过实例验证了该方法的有效性,从而进一步丰富了粗糙集理论。

* 收稿日期:2005-03-26
基金项目:国家"九七三"计划资助项目(2002CB31200)
作者简介:徐伟华(1979-),男,西安交通大学博士研究生,研究方向:粗糙集理论与应用;张文修(1940-),男,西安交通大学博士生导师,教授,研究方向:模糊集,粗糙集,人工智能的数学理论。

2 基于优势关系的信息系统

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 2.1[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 2.2[7] 设 $I=(U,A,F,D,G)$ 为目标信息系统，对于 $B\subseteq A$，令

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

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

$R_B^{\leqslant},R_D^{\leqslant}$ 称为目标信息系统的优势关系，此时该目标信息统称为是基于优势关系下的目标信息系统。

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 2.1[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：$[x_i]_A^{\leqslant}\subseteq[x_i]_{B_2}^{\leqslant}\subseteq[x_i]_{B_1}^{\leqslant}$.

(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

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

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

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[7]。

为了叙述方便，下文我们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 2.3[7] 设 $I=(U,A,F,D,G)$ 为基于优势关系的目标信息系统，若 $R_A^{\leqslant}\subseteq R_D^{\leqslant}$，则称该基于优势关系的目标信息系统是协调的，否则若 $R_A^{\leqslant}\not\subseteq R_D^{\leqslant}$ 称该系统是不协调的。

例 2.1[7] 表1给出了一个基于优势关系的目标信息系统。

表1

$U\times(A\cup D)$	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

于是，按照优势关系的定义有：

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

显然，$R_A^{\leqslant}\not\subseteq R_d^{\leqslant}$. 因此该目标信息系统在优势关系下是不协调的。

3 不协调目标信息系统的分布约简

由于优势关系不再是等价关系，不能形成对象集上的划分，而是一个覆盖。因此，对于优势关系下的信息系统中不能采取 Pawlak 近似空间下的信息系统中的方法定义分布函数和最大分布函数。下面我们给出在优势关系下的信息系统的分布函数和最大分布函数的定义方式。

设(U,A,F,D,G)为目标信息系统，$R_B^{\leqslant}$,$R_D^{\leqslant}$ 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 $B\subseteq A$，$x\in U$，记

$$U/R_B^{\leqslant} = \{[x_i]_B^{\leqslant}:x_i \in U\}$$
$$U/R_D^{\leqslant} = \{D_1,D_2,\cdots,D_r\}$$
$$\mu_B(x) = (\frac{|D_1 \cap [x]_B^{\leqslant}|}{|U|},\frac{|D_2 \cap [x]_B^{\leqslant}|}{|U|},\cdots,\frac{|D_r \cap [x]_B^{\leqslant}|}{|U|})$$
$$\gamma_B(x) = \max\{\frac{|D_1 \cap [x]_B^{\leqslant}|}{|U|},\frac{|D_2 \cap [x]_B^{\leqslant}|}{|U|},\cdots,\frac{|D_r \cap [x]_B^{\leqslant}|}{|U|}\}$$

其中$[x]_B^{\leqslant}=\{y\in U:(x,y)\in R_B^{\leqslant}\}$，我们称 $\mu_B(x)$为论域U 上的关于属性子集 B 的分布函数，$\gamma_B(x)$为论域 U 上的关于属性子集 B 的最大分布函数。

定义 3.1 设 $\alpha=(a_1,a_2,\cdots,a_n)^T$ 和 $\beta=(b_1,b_2,\cdots,b_n)^T$ 为两个 n 维 $n\times1$ 向量，若 $a_i=b_i(i=1,2,\cdots,n)$称向量 α 等于向量 β,记做 $\alpha=\beta$；若 $a_i\leqslant b_i(i=1,2,\cdots,n)$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 $\alpha\not\leqslant\beta$.

如：$(1,2,3)\not\leqslant(1,1,4)$且$(1,1,4)\not\leqslant(1,2,3)$。

显然由以上可立即得到下面命题。

命题 3.1　(1) 当 $B\subseteq A$ 时，对任意的 $x\in U$ 有 $\mu_A(x)\leqslant\mu_B(x)$。

(2) 当 $B\subseteq A$ 时，对任意的 $x\in U$ 有 $\gamma_A(x)\leqslant\gamma_B(x)$。

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

(4) $\forall\ x,y\in U$，当$[x]_A^{\leqslant}\subseteq[x]_B^{\leqslant}$ 时，有 $\gamma_B(y)\leqslant\gamma_B(x)$。

定义 3.2　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，且 B 的任何真子集不是分布协调集，则称 B 为分布协调约简。

定义 3.3　设 $I=(U,A,F,D,G)$ 为目标信息系统。若$\forall\ x\in U$，有 $\gamma_B(x)=\gamma_A(x)$，则称 B 是最大分布协调集，且 B 的任何真子集不是最大分布协调集，则称 B 为最大分布协调约简。

例 3.1　考虑例 2.1 给出的不协调目标信息系统。

若在该信息系统中，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：

$$\mu_A(x_1)=(\frac{1}{3},\frac{1}{2},\frac{2}{3})$$

$$\mu_A(x_2)=(\frac{1}{6},\frac{1}{3},\frac{1}{2})$$

$$\mu_A(x_3)=(\frac{1}{6},\frac{1}{2},\frac{5}{6})$$

$$\mu_A(x_4)=(0,\frac{1}{6},\frac{1}{3})$$

$$\mu_A(x_5)=(\frac{1}{6},\frac{1}{6},\frac{1}{6})$$

$$\mu_A(x_6)=(0,0,\frac{1}{6})$$

$$\gamma_A(x_1)=\frac{2}{3}$$

$$\gamma_A(x_2)=\frac{1}{2}$$

$$\gamma_A(x_3)=\frac{5}{6}$$

$$\gamma_A(x_4)=\frac{1}{3}$$

$$\gamma_A(x_5)=\frac{1}{6}$$

$$\gamma_A(x_6)=\frac{1}{6}$$

若取 $B=\{a_2,a_3\}$时，容易验证$\forall\ x\in U$ 有：$[x]_A^{\leqslant}=[x]_B^{\leqslant}$，因此有 $\mu_B(x)=\mu_A(x)$和 $\gamma_B(x)=\gamma_A(x)$。故$\{a_2,a_3\}$是个分布协调集，也是个最大分布协调集。进一步可以计算$\{a_2\}$，$\{a_3\}$均不是分布协调集和最大分布协调集，因此 $B=\{a_2,a_3\}$是个分布约简，也是个最大分布约简。

容易验证 $B'=\{a_1,a_3\}$和$\{a_1,a_2\}$都不是分布协调集，也不是最大分布协调集，因此该目标信息

系统只有一个分布约简，也只有一个最大分布约简，即$\{a_2,a_3\}$。

下面我们具体给出不协调目标信息系统的分布约简与最大分布约简的关系以及判定定理。

定理 3.1 设(U,A,F,D,G)为目标信息系统，则分布协调集一定是最大分布协调集。

证明 由定义直接可得。

推论 3.1 设(U,A,F,D,G)为目标信息系统，则分布协约简一定是最大分布约简。

定理 3.2 设(U,A,F,D,G)为目标信息系统，$B\subseteq A$，则 B 是分布协调集当且仅当$\forall\ x,y\in U$，当 $\mu_A(y)\not\leqslant\mu_A(x)$时有$[y]_B^{\leqslant}\not\subseteq[x]_B^{\leqslant}$.

证明 "$\Rightarrow$"反证。

假设当 $\mu_A(y)\not\leqslant\mu_A(x)$时$[y]_B^{\leqslant}\not\subseteq[x]_B^{\leqslant}$ 不成立。故此时有$[y]_B^{\leqslant}\subseteq[x]_B^{\leqslant}$，由命题 2(3)知 $\mu_B(y)\leqslant\mu_B(x)$，而 B 又是分布协调集，于是有 $\mu_A(x)=\mu_B(x)$，$\mu_A(y)=\mu_B(y)$。故有 $\mu_A(y)\leqslant\mu_A(x)$，矛盾。

因此当 $\mu_A(y)\not\leqslant\mu_A(x)$时有$[y]_B^{\leqslant}\not\subseteq[x]_B^{\leqslant}$ 成立。

"$\Leftarrow$"要证 B 是分布协调集须证明对任意的 $x\in U$ 有 $\mu_A(x)=\mu_B(x)$。而由命题 2(1)知，只需 $\mu_B(x)\leqslant\mu_A(x)$证即可。

当 $\mu_B(x)=$时显然成立。下面证明 $\mu_B(x)\neq0$ 时同样成立。

由条件知，$\forall\ x,y\in U$，若 $\mu_A(y)\not\leqslant\mu_A(x)$时有$[y]_B^{\leqslant}\not\subseteq[x]_B^{\leqslant}$ 成立。因此$\forall\ x,y\in U$，若$[y]_B^{\leqslant}\subseteq[x]_B^{\leqslant}$ 成立，则 $\mu_A(y)\leqslant\mu_A(x)$成立。

另外，当$\frac{|D_i\cap[x]_B^{\leqslant}|}{|U|}\neq0$ 时，便有$|D_i\cap[x]_B^{\leqslant}|\neq0$。不妨设 $y_i\in D_i\cap[x]_B^{\leqslant}$，则 $y_i\in D_i$ 且 $y_i\in[x]_B^{\leqslant}$，故由命题 1(4)知$[y_i]_B\subseteq[x]_B$ 成立。因此可得 $\mu_A(y_i)\leqslant\mu_A(x)$。又 $y_i\in[y_i]_A^{\leqslant}$，所以 $y_i\in D_i\cap[y_i]_A$，即有$|D_i\cap[x]_B^{\leqslant}|\leqslant|D_i\cap[y_i]_B^{\leqslant}|$。因此 $\mu_B(x)\leqslant\mu_A(y_i)$，故 $\mu_B(x)\leqslant\mu_A(x)$成立。

用同样的方法我们可以得到最大分布协调集的充要条件。

定理 3.3 设(U,A,F,D,G)为目标信息系统，$B\subseteq A$，则 B 是最大分布协调集当且仅当$\forall\ x,y\in U$，当 $\gamma_A(y)>\gamma_A(x)$时有$[y]_B^{\leqslant}\not\subseteq[x]_B^{\leqslant}$。

4 不协调目标信息系统的分布约简方法

第 2 节中的定理给出了不协调目标信息系统的分布协调集和最大分布协调集的等价刻画，这是判断属性子集是否协调的理论所在，由此我们可进一步得出分布约简和最大分布约简的方法。下面先给出辨识属性矩阵的概念。

定义 4.1 设(U,A,F,D,G)为不协调目标信息系统，记

$$D_\mu^*=\{(x_i,x_j):\mu_A(x_i)\not\leqslant\mu_A(x_j)\}$$

$$D_\gamma^*=\{(x_i,x_j):\gamma_A(x_i)>\gamma_A(x_j)\}$$

用 $f_{a_k}(x)$表示属性关于 a_k 对象 x 的取值。定义

$$D_\mu(x_i,x_j)=\begin{cases}\{a_k\in A,f_{a_k}(x_i)>f_{a_k}(x_j)\}, & (x_i,x_j)\in D_\mu^*\\ \varnothing, & (x_i,x_j)\notin D_\mu^*\end{cases}$$

$$D_\gamma(x_i,x_j)=\begin{cases}\{a_k\in A,f_{a_k}(x_i)>f_{a_k}(x_j)\}, & (x_i,x_j)\in D_\gamma^*\\ \varnothing, & (x_i,x_j)\notin D_\gamma^*\end{cases}$$

称 $D_\mu(x_i,x_j)$与 $D_\gamma(x_i,x_j)$分别为 x_i 与 x_j 的分布与最大分布可辨识属性集。矩阵 $M_\mu=(D_\mu(x_i,x_j),x_i,x_j\in U)$与 $M_\gamma=(D_\gamma(x_i,x_j),x_i,x_j\in U)$分别称为该目标信息系统的分布辨识矩阵和最大分布辨识矩阵。

定理 4.1 设(U,A,F,D,G)为不协调目标信息系统，$B\subseteq A$，则：B 是分布协调集当且仅当 B

$\cap D_\mu(x_i,x_j)\neq\varnothing$，$\forall\ (x,y)\in D_\mu^*$.

证明　"⇒"$\forall\ (x,y)\in D_\mu^*$，则有：$\mu_A(x)\npreceq\mu_A(y)$，而 B 又是分布协调集，故由定理 2 有$[y]_B^{\leqslant}\nsubseteq[x]_B^{\leqslant}$. 因此$[x]_B^{\leqslant}$与$[y]_B^{\leqslant}$的关系有三种情况：(1)$[x]_B^{\leqslant}\subset[y]_B^{\leqslant}$，(2)$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}=\varnothing$，(3)$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}\subset[x]_B^{\leqslant}$且$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}\subset[y]_B^{\leqslant}$. 下面证明在这三种情况下均有 $B\cap D_\mu(x_i,x_j)\neq\varnothing$成立。

(1)若$[x]_B^{\leqslant}\subset[y]_B^{\leqslant}$时，则至少存在一个 $z\in[y]_B^{\leqslant}$，但 $z\notin[x]_B^{\leqslant}$. 由 $z\notin[x]_B^{\leqslant}$可知至少存在一个 $a_k\in B$，使得 $f_{a_k}(x)>f_{a_k}(z)$，而 $z\in[y]_B^{\leqslant}$，则 $f_{a_k}(y)\leqslant f_{a_k}(z)$。于是有：$f_{a_k}(x)>f_{a_k}(y)$，因此 $a_k\in D_\mu(x_i,x_j)$，即有 $B\cap D_\mu(x_i,x_j)\neq\varnothing$.

(2)若$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}=\varnothing$时。事实上，此时必然至少存在一个 $a_k\in B$，使得 $f_{a_k}(x)>f_{a_k}(y)$，即 $B\cap D_\mu(x_i,x_j)\neq\varnothing$. 否则，若$\forall\ a_l\in B$都有 $f_{a_k}(x)\leqslant f_{a_k}(y)$，则 $y\in[x]_B^{\leqslant}$，这与$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}=\varnothing$矛盾。

(3)若$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}\subset[x]_B^{\leqslant}$且$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}\subset[y]_B^{\leqslant}$时，证明同(1)。因为此时也至少存在一个 $z\in[y]_B^{\leqslant}$，但 $z\notin[x]_B^{\leqslant}$.

因此，若 B 是分布协调集，则$\forall\ (x,y)\in D_\mu^*$有 $B\cap D_\mu(x,y)\neq\varnothing$.

"⇐"若$\forall\ (x,y)\in D_\mu^*$有 $B\cap D_\mu(x,y)\neq\varnothing$，则存在一个 $a_k\in B$ 使得 $a_k\in D_\mu(x,y)$，故有 $f_{a_k}(x)>f_{a_k}(y)$，所以 $y\notin[x]_B^{\leqslant}$，而又 $y\in[y]_B^{\leqslant}$. 因此有$[x]_B^{\leqslant}\cap[y]_B^{\leqslant}\neq[y]_B^{\leqslant}$，即$[y]_B^{\leqslant}\nsubseteq[x]_B^{\leqslant}$. 而另一方面因为$(x,y)\in D_\mu^*$，则 $\mu_A(x)\npreceq\mu_A(y)$。故当 $\mu_A(x)\npreceq\mu_A(y)$成立时，有$[y]_B^{\leqslant}\nsubseteq[x]_B^{\leqslant}$成立。由定理 2 知 B 是分布协调集。

定理 4.2　设(U,A,F,D,G)为不协调目标信息系统，$B\subseteq A$，则：B 是最大分布协调集当且仅当 $B\cap D_\gamma(x_i,x_j)\neq\varnothing$，$\forall\ (x,y)\in D_\gamma^*$.

证明　同定理 4.1。

定义 4.2　设(U,A,F,D,G)为不协调目标信息系统，M_μ 和 M_γ 分别为其分布和最大分布辨识矩阵，若记

$$F_\mu=\wedge\ \{\vee\ \{a_k:a_k\in D_\mu(x_i,x_j)\},x_i,x_j\in U\}=\wedge\ \{\vee\ \{a_k:a_k\in D_\mu(x_i,x_j)\},x_i,x_j\in D_\mu^*\}$$

$$F_\gamma=\wedge\ \{\vee\ \{a_k:a_k\in D_\gamma(x_i,x_j)\},x_i,x_j\in U\}=\wedge\ \{\vee\ \{a_k:a_k\in D_\gamma(x_i,x_j)\},x_i,x_j\in D_\gamma^*\}$$

称 F_μ 和 F_γ 分别为该信息系统的分布和最大分布辨识公式。

定理 4.3　设(U,A,F,D,G)为不协调目标信息系统。分布辨识公式 F_μ 的极小析取范式为 $F_\mu=\bigvee\limits_{k=1}^{p}(\bigwedge\limits_{s=1}^{q_k}a_s)$，若记 $B_\mu^k=\{a_s,s=1,2,\cdots,q_k\}$，则$\{B_\mu^k,k=1,2,\cdots,p\}$是所有分布约简形式的集合。

证明　对任意的$(x_i,x_j)\in D_\mu^*$，由极小析取范式的定义知 $B_\mu^k\cap D_\mu(x_i,x_j)\neq\varnothing$，再由定理 4 知 B_μ^k 是分布协调集。同时，$F_\mu=\bigvee\limits_{k=1}^{p}(B_\mu^k)$在 B_μ^k 中去掉一个元素形成 $B_\mu^{k\prime}$，则必然存在某个$(x_i,x_j)\in D_\mu^*$ 使得 $B_\mu^{k\prime}\cap D_\mu(x_i,x_j)=\varnothing$，故 $B_\mu^{k\prime}$不是分布协调集，从而 B_μ^k 是分布约简。而分布辨识公式中包含了所有的 $D_\mu(x_i,x_j)$，因此不存在其他分布约简。

定理 4.4　设(U,A,F,D,G)为不协调目标信息系统。最大分布辨识公式 F_γ 的极小析取范式为 $F_\gamma=\bigvee\limits_{k=1}^{p}(\bigwedge\limits_{s=1}^{q_k}a_s')$，若 $B_\gamma^k=\{a_s',s=1,2,\cdots,q_k\}$记，则$\{B_\gamma^k,k=1,2,\cdots,p\}$是所有最大分布约简形式的集合。

证明　同定理 4.3。

例 4.1　对于表例 1 给出的不协调目标信息系统中的分布函数和最大分布函数在例 2 中已经得到，而且可以计算：

$$D_\mu^*=\{(x_1,x_2),(x_1,x_3),(x_1,x_4),(x_1,x_5),(x_1,x_6),(x_2,x_4),(x_2,x_5),(x_2,x_6),(x_3,x_1),$$

$$(x_3,x_2),(x_3,x_4),(x_3,x_5),(x_3,x_6),(x_4,x_5),(x_4,x_6),(x_5,x_4),(x_5,x_6)\}$$

$$D_\gamma^* = \{(x_1,x_2),(x_1,x_4),(x_1,x_5),(x_1,x_6),(x_2,x_4),(x_2,x_5),(x_2,x_6),$$
$$(x_3,x_1),(x_3,x_2),(x_3,x_4),(x_3,x_5),(x_3,x_6),(x_4,x_5),(x_4,x_6)\}$$

于是知该信息系统的分布辨识矩阵和最大分布辨识矩阵如表 2、表 3 所示。

表 2

M_μ	x_1	x_2	x_3	x_4	x_5	x_6
x_1	$\varnothing$	$\varnothing$	a_2	a_2	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	a_1,a_2	$\varnothing$	$\varnothing$
x_3	a_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	a_3	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	a_1,a_2	$\varnothing$	$\varnothing$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

表 3

M_μ	x_1	x_2	x_3	x_4	x_5	x_6
x_1	$\varnothing$	$\varnothing$	$\varnothing$	a_2	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	a_1,a_2	$\varnothing$	$\varnothing$
x_3	a_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	a_3	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

故可得：

$$F_\mu = F_\gamma = a_2 \wedge a_3 \wedge (a_1 \vee a_2) = a_2 \wedge a_3$$

因此$\{a_2,a_3\}$是该不协调目标信息系统的所有分布约简而且也是最大分布约简。这与例 2 的结果是一致的。

5 结论

我们知道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。本文在基于优势关系下的不协调目标信息系统中引入了分布约简和最大分布约简的概念，并讨论了二者之间的关系，而且得到了分布和最大分布约简的判定定理以及辨识矩阵，建立了不协调目标信息系统的分布和最大分布约简的具体方法，同时通过实例验证了该方法的有效性。

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] 徐伟华，张文修. 基于优势关系下协调近似空间[J]. 计算机科学, 2005, 32(9): 164～165.
[13] 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.

Distribution Reduction in Inconsistent Information Systems Based on Dominance Relations

XU Wei-hua, ZHANG Wen-xiu
(Institute of Information and System Sciences, Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Knowledge reduction is one of the most important problems in rough set theory. However, most of information systems are based on dominance relations because of various factors. To acquire brief decision rules from inconsistent systems based on dominance relations, knowledge reductions are needed. The main aim of this paper is to study the problem. The concepts of distribution reduction and maximum distribution reduction are introduced in information systems based dominance relations. And the relationships among them are discussed. The judgment theorem and discernable matrix are obtained, from which we can provide the approach to those reductions in inconsistent systems based on dominance relations. Finally, an example illustrates the validity of this method.

Key words: Rough Set; Inconsistent System; Distribution Reduction; Maximum Distribution Reduction; Discernable Matrix