

基于优势关系下的协调近似空间(续)*)

徐伟华¹ 张晓燕² 苏雅娟¹ 张文修¹

(西安交通大学理学院信息与系统科学研究所 西安 710049)¹

(广东海洋大学理学院数学与信息科学系 湛江 524088)²

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和最大分布协调近似空间的概念，并证明了在优势关系下不协调目标信息系统也可以转化为一个分布以及最大分布协调近似空间，这更加方便了基于优势关系下不协调目标信息系统的研究，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，协调近似空间

Consistent Approximation Spaces Based on Dominance Relations (II)

XU Wei-Hua¹ ZHANG Xiao-Yan² SU Ya-Juan¹ ZHANG Wen-Xiu¹

(Institute of Information and System Sciences, Faculty of Science, Xi'an Jiaotong University, Xi'an 710049)¹

(School of Sciences, Guangdong Ocean University, Zhanjiang 524088)²

Abstract In real-world, most of information systems are not only inconsistent, but also based on dominance relations because of various factors. Moreover, it is extremely important to turn an inconsistent system into a consistent one. The main aim of this paper is to discuss the problem. The concepts of approximation space of distribution consistent are proposed based on dominance relation in the paper, and it is shown that the inconsistent system can be turned into a distribution consistent approximation space, and any knowledge isn't lost. Using this approach some problems can be simplified.

Keywords Rough set, Information system, Dominance relation, Approximation space of distribution consistent

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。

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，许多学者已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，并取得了可喜的成果^[6~10]。

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^[11~13]。在文[11]中给出了协调近似空间的定义，并证明了在优势关系下不协调目标信息系统也可以转化为一个协调近

似空间。另外，文[13]作者引入了优势关系下分布协调集与最大分布协调集的概念，并讨论了它们的有关特征，得到了比较好的结果。为此，本文对这一问题进行了探讨研究，进一步给出优势关系下协调目标信息系统的另外一种非常有意义的协调近似空间的转化方法。在基于优势关系下的不协调目标信息系统中引入了分布和最大分布协调近似空间的概念，并证明了在优势关系下不协调目标信息系统也可以转化为一个分布以及最大分布协调近似空间，这更加方便了基于优势关系下不协调目标信息系统的研究，从而进一步丰富了粗糙集理论。

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，它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， $U=\{x_1, x_2, \dots, x_n\}$ ；

A 是有限条件属性集， $A=\{a_1, a_2, \dots, a_p\}$ ；

D 是有限目标属性集， $D=\{d_1, d_2, \dots, d_q\}$ ；

F 是 U 与 A 的关系集， $F=\{f_k : U \rightarrow V_k, k \leq p\}$ ， V_k 是 a_k 的有限值域；

G 是 U 与 D 的关系集， $G=\{g_k : U \rightarrow V_k, k' \leq q\}$ ， V_k 是

*) 国家“九七三”计划项目资助（项目号：2002CB31200）。徐伟华 博士研究生，研究方向：粗糙集理论与应用；张晓燕 硕士，研究方向：粗糙集、动力系统；苏雅娟 硕士研究生，研究方向：粗糙集；张文修 博导，教授，研究方向：模糊集、粗糙集、人工智能的数学理论。

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^[7] 设 $I = (U, A, F, D, G)$ 为目標信息系统, 对于 $B \subseteq A$, 令

$$R_B^{\leq} = \{(x_i, x_j) \in U \times U : f_i(x_i) \leq f_i(x_j), \forall a_i \in B\},$$

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

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

记: $[x_i]_B^{\leq} = \{x_j \in U : (x_i, x_j) \in R_B^{\leq}\} = \{x_j \in U : f_i(x_i) \leq f_i(x_j), \forall a_i \in B\}$,

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

易见, 优势关系有下面性质:

命题 1^[7] (1) R_B^{\leq} 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 $R_A^{\leq} \subseteq R_{B_2}^{\leq} \subseteq R_{B_1}^{\leq}$ 。

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

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

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

$$R_B^{\leq}(X) = \{x_i \in U : [x_i]_B^{\leq} \subseteq X\}, \bar{R}_B^{\leq}(X) = \{x_i \in U : [x_i]_B^{\leq} \cap X \neq \emptyset\}.$$

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[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 3^[7] 设 $I = (U, A, F, D, G)$ 为基于优势关系的目標信息系统, 若 $R_A^{\leq} \subseteq R_D^{\leq}$, 则称该基于优势关系的目標信息系统是协调的。否则, 若 $R_A^{\leq} \not\subseteq R_D^{\leq}$, 称该系统是不协调的。

例 1^[7] 表 1 给出了一个基于优势关系的目標信息系统。

表 1 一个基于优势关系的目標信息系统

$U \times (A \cup D)$	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

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

$$[x_4]_A^{\leq} = \{x_4, x_6\}, [x_5]_A^{\leq} = \{x_5\}, [x_6]_A^{\leq} = \{x_6\},$$

$$[x_1]_D^{\leq} = [x_5]_D^{\leq} = \{x_1, x_5\}, [x_2]_D^{\leq} = [x_4]_D^{\leq} = \{x_1, x_2, x_4, x_5\}, [x_3]_D^{\leq} = [x_6]_D^{\leq} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

显然, $R_A^{\leq} \not\subseteq R_D^{\leq}$, 因此该目標信息系统在优势关系下是不协调的。

3 基于优势关系下的分布协调近似空间

在文[11]中给出了协调近似空间的定义, 并证明了在优

势关系下不协调目標信息系统也可以转化为一个协调近似空间。这里, 我们将进一步给出优势关系下协调目標信息系统另外一种非常有意义的协调近似空间的转化方法。首先将文[11]中的有关概念列出。

定义 4^[11] 设 $S^{\leq} = (U, A, F, D, G)$ 为目標信息系统, 若存在 $U \times U$ 上的二元关系 R_A^{\leq} , 使得有 $R_A^{\leq} \subseteq R_D^{\leq}$, 称 S^{\leq} 为基于优势关系下的协调近似空间。

定理 1^[11] 设 $S^{\leq} = (U, A, F)$ 为非目標信息系统, 则 S^{\leq} 是基于优势关系下的协调近似空间。

定理 2^[11] 设 $S^{\leq} = (U, A, F, D, G)$ 为协调目標信息系统, 则 S^{\leq} 是基于优势关系下的协调近似空间。

定义 5^[12] 设 $S^{\leq} = (U, A, F, D, G)$ 为目標信息系统, 若 $\forall x \in U$, 有 $\sigma_B^{\leq}(x) = \sigma_A^{\leq}(x)$, 则称 B 是分配协调集, 且 B 的任何真子集不是分配协调集, 则称 B 为分配协调约简。

定义 6^[11] 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目標信息系统(在优势关系下)。二元关系

$$R^{\leq} = \{(x_i, x_j) \in U \times U : \sigma_A^{\leq}(x_i) \supseteq \sigma_A^{\leq}(x_j)\}$$

称为不协调目標信息系统的分配协调优势关系。

定理 3^[11] 不协调目標信息系统 $S^{\leq} = (U, A, F, D, G)$ 一定是协调近似空间。

定理 4^[11] 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目標信息系统, R^{\leq} 为其分配协调优势关系, $B \subseteq A$, 则有 $R_B^{\leq} \subseteq R^{\leq} \Leftrightarrow \sigma_B^{\leq}(x) = \sigma_B^{\leq}(x)$, 对 $\forall x \in U$ 。

通过上述定理 3、4, 我们可以看出, 利用分配协调关系 R^{\leq} 使不协调目標信息系统转化为了分配协调近似空间, 而且并没有改变原来信息系统的信息, 还保持了与文[12]中的分配协调集的一致性。

另外, 文[13]中作者讨论了优势关系下的分布协调集与最大分布协调集, 那么它们是否也能够做成一个优势关系, 使原来的不协调信息空间为一个协调的呢? 为了方便讨论这一问题, 我们先将上面的有关定理做如下修改。

定义 6' 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目標信息系统(在优势关系下)。二元关系

$$R^{\leq} = \{(x_i, x_j) \in U \times U : \sigma_A^{\leq}(x_i) \supseteq \sigma_A^{\leq}(x_j)\}$$

称为不协调目標信息系统的分配协调优势关系。

定理 3' 不协调目標信息系统 $S^{\leq} = (U, A, F, D, G)$ 在分配协调优势关系 R^{\leq} 下一定是协调近似空间, 并将之称为分配协调近似空间。

下面将文[13]中的分布和最大分布协调集的有关定义列出。

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) 为目標信息系统, R_B^{\leq}, R_D^{\leq} 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

$$U/R_B^{\leq} = \{[x_i]_B^{\leq} : x_i \in U\},$$

$$U/R_D^{\leq} = \{D_1, D_2, \dots, D_r\},$$

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

$$\gamma_B^{\leq}(x) = \max \left\{ \frac{|D_1 \cap [x]_B^{\leq}|}{|U|}, \frac{|D_2 \cap [x]_B^{\leq}|}{|U|}, \dots, \frac{|D_r \cap [x]_B^{\leq}|}{|U|} \right\},$$

$$\frac{|D_r \cap [x]_B^{\leq}|}{|U|} \}$$

其中 $[x]_B^{\leq} = \{y \in U : (x, y) \in R_B^{\leq}\}$ 。文[13]中称 $\mu_B^{\leq}(x)$ 为论域 U 上的关于属性子集 B 的分布函数, $\gamma_B^{\leq}(x)$ 为论域 U 上的关于属性子集 B 的最大分布函数。

定义 7^[13] 设 $\alpha = (a_1, a_2, \dots, a_n)^T$ 和 $\beta = (b_1, b_2, \dots, b_n)^T$ 为两个 n 维 $n \times 1$ 向量, 若 $a_i = b_i$ ($i = 1, \dots, n$) 称向量 α 等于向量 β , 记作 $\alpha = \beta$; 若 $a_i \leq b_i$ ($i = 1, \dots, n$) 称向量 α 小于等于向量 β , 记作 $\alpha \leq \beta$; 否则若存在某个 i_0 , ($i_0 \in \{1, 2, \dots, n\}$), 使得 $a_{i_0} > b_{i_0}$, 称向量 α 不小于等于向量 β , 记作 $\alpha \not\leq \beta$ 。

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

命题 2^[13] (1) 当 $B \subseteq A$ 时, 对任意的 $x \in U$, 有 $\mu_A^{\leq}(x) \leq \mu_B^{\leq}(x)$ 。

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

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

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

定义 8^[13] 设 $I = (U, A, F, D, G)$ 为目標信息系统。若 $\forall x \in U$, 有 $\mu_B^{\leq}(x) = \mu_A^{\leq}(x)$, 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 9^[13] 设 $I = (U, A, F, D, G)$ 为目標信息系统。若 $\forall x \in U$, 有 $\gamma_B^{\leq}(x) = \gamma_A^{\leq}(x)$, 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。

我们在此基础上可以定义分布协调优势关系如下:

定义 10 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目标信息系统(在优势关系下)。二元关系

$$R_{\mu}^{\leq} = \{(x_i, x_j) \in U \times U : \mu_A^{\leq}(x_j) \leq \mu_A^{\leq}(x_i)\}$$

$$R_{\gamma}^{\leq} = \{(x_i, x_j) \in U \times U : \gamma_A^{\leq}(x_j) \leq \gamma_A^{\leq}(x_i)\}$$

分别称为不协调目标信息系统的分布和最大分布协调优势关系。

定理 5 不协调目标信息系统 $S^{\leq} = (U, A, F, D, G)$ 在分布协调优势关系 R_{μ}^{\leq} 下一定是协调近似空间, 并将之称为分布协调近似空间。

证明: 只需证明可以找到 R_0^{\leq} , 使得有 $R_A^{\leq} \subseteq R_0^{\leq}$ 成立。

事实上, 只需取 $R_{\mu}^{\leq} = R_0^{\leq}$ 即可。即须证: 对于 $(x_i, x_j) \in U \times U$, 若 $(x_i, x_j) \in R_A^{\leq}$, 则 $(x_i, x_j) \in R_{\mu}^{\leq}$ 。

由于 $(x_i, x_j) \in R_A^{\leq}$, 即 $x_j \in [x_i]_A^{\leq}$, 由命题 1(4) 可知 $[x_i]_A^{\leq} \supseteq [x_j]_A^{\leq}$, 又由命题 2(3) 有 $\mu_A^{\leq}(x_j) \leq \mu_A^{\leq}(x_i)$, 因此有 $(x_i, x_j) \in R_{\mu}^{\leq}$ 成立。

用同样的方法可以得到:

定理 6 不协调目标信息系统 $S^{\leq} = (U, A, F, D, G)$ 在最大分布协调优势关系 R_{γ}^{\leq} 下一定是协调近似空间, 并将之称为最大分布协调近似空间。

定理 7 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目标信息系统, R_{γ}^{\leq} 为其分布协调优势关系, $B \subseteq A$, 则有 $R_B^{\leq} \subseteq R_{\mu}^{\leq} \Leftrightarrow \mu_A^{\leq}(x) = \mu_B^{\leq}(x)$, 对 $\forall x \in U$ 。

证明: “ \Rightarrow ”由命题 2(1) 知, 只需证 $\mu_B^{\leq}(x) \leq \mu_A^{\leq}(x)$ 即可。

当 $\mu_B(x) = 0$ 时显然成立。下面我们证明 $\mu_B(x) \neq 0$ 时同样成立。

由条件知, 对 $\forall x, y \in U$, 若 $(x, y) \in R_B^{\leq}$, 则 $(x, y) \in R_{\mu}^{\leq}$ 。于是, 若 $[y]_B^{\leq} \subseteq [x]_B^{\leq}$ 成立, 则 $\mu_B^{\leq}(y) \leq \mu_B^{\leq}(x)$ 成立。

另外, 当 $\frac{|D_i \cap [x]_B^{\leq}|}{|U|} \neq 0$ 时, 便有 $|D_i \cap [x]_B^{\leq}| \neq 0$ 。不妨设 $y_i \in D_i \cap [x]_B^{\leq}$, 则 $y_i \in D_i$, 且 $y_i \in [x]_B^{\leq}$, 故由命题 1(4) 知 $[y_i]_B \subseteq [x]_B$ 成立。因此可得 $\mu_A^{\leq}(y_i) \leq \mu_A^{\leq}(x)$ 。又 $y_i \in [y_i]_B^{\leq}$, 所以 $y_i \in D_i \cap [y_i]_A$, 即有 $|D_i \cap [x]_B^{\leq}| \leq |D_i \cap [y_i]_B^{\leq}|$ 。因此 $\mu_B^{\leq}(x) \leq \mu_B^{\leq}(y_i)$, 故 $\mu_B^{\leq}(x) \leq \mu_A^{\leq}(x)$ 成立。

“ \Leftarrow ”若 $(x_i, x_j) \in R_B^{\leq}$, 则有 $[x_j]_B^{\leq} \subseteq [x_i]_B^{\leq}$ 。故由命题 2(3), 有 $\mu_B^{\leq}(x_j) \leq \mu_B^{\leq}(x_i)$ 。又由已知 $\mu_A^{\leq}(x) = \mu_B^{\leq}(x)$, 因此可得 $\mu_A^{\leq}(x_j) \leq \mu_A^{\leq}(x_i)$, 即 $(x_i, x_j) \in R_{\mu}^{\leq}$ 。

定理 8 设 $S^{\leq} = (U, A, F, D, G)$ 为不协调目标信息系统, R_{γ}^{\leq} 为其最大分布协调优势关系, $B \subseteq A$, 则有 $R_B^{\leq} \subseteq R_{\gamma}^{\leq} \Leftrightarrow \gamma_A^{\leq}(x) = \gamma_B^{\leq}(x)$, 对 $\forall x \in U$ 。

证明: 同定理 7。

由上述定理 5, 6, 7, 8 我们可以看出, 利用分布和最大分布协调关系 $R_{\mu}^{\leq}, R_{\gamma}^{\leq}$, 使不协调目标信息系统转化为了分布以及最大分布协调近似空间, 而且并没有改变原来系统的信息量, 还保持了与文[13]中的分布以及最大分布协调集的一致性。这样, 大大方便了我们的研究, 丰富了粗糙集理论。

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重要的。本文对这一问题进行了探讨研究, 给出优势关系下协调目标信息系统的另外一种非常有意义的协调近似空间的转化方法。在基于优势关系下的不协调目标信息系统中引入了分布和最大分布协调近似空间的概念, 并证明了在优势关系下不协调目标信息系统也可以转化为一个分布以及最大分布协调近似空间, 这更加方便了基于优势关系下不协调目标信息系统的研究, 从而进一步丰富了粗糙集理论。

参 考 文 献

- 1 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data. Boston: Kluwer Academic Publishers, 1991
- 2 Pawlak Z. Rough Sets. Communication of the ACM, 1995, 38(1): 89~95
- 3 王国胤. Rough 集理论与知识获取. 西安: 西安交通大学出版社, 2001
- 4 吴伟志, 张文修, 徐宗本. 粗糙模糊集的构造与公理化方法[J]. 计算机学报, 2004, (2): 197~203
- 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. 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. 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. 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. 计算机学报, 2003, 26(1): 12~18
- 9 Kryszkiewicz M. Comparative Studies of Alternative of Knowledge Reduction in Inconsistent Systems. Intelligent Systems, 2001, 16(1): 105~120
- 10 Grcos M B, Slowinski R. Rough Approximation of Preference Relation by Dominance Relations. European Journal of Operational Research, 1999, 117: 63~68
- 11 徐伟华, 张文修. 基于优势关系下协调近似空间. 计算机科学, 2005, 32(9): 164~165
- 12 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. 计算机科学, 2006, 33(2): 182~184
- 13 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. 模糊系统与数学(已录用)