基于优势关系下的协调近似空间

徐伟华　张文修

（西安交通大学理学院信息系与系统科学研究所　西安 710049）

摘　要　在基于优势关系下的信息系统中引入了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协调近似空间，这进一步方便了基于优势关系下不协调目标信息系统的研究，丰富了粗糙集理论。

关键词　粗糙集，信息系统，优势关系，协调近似空间

Consistent Approximation Spaces 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)

Abstract　In practices, most of information systems are not only inconsistent, but also based on dominance relations because of various factors. It is every difficult to acquire brief decision rules from inconsistent systems based on dominance relations. Furthermore, how to turn an inconsistent system into a consistent one is extremely meaningful. The main objective of this paper is to study the problem. The concept of consistent approximation space based on dominance relation is proposed in the paper, and it is shown that the inconsistent system can be turned into a consistent approximation space based on dominance, and any knowledge isn't lost. Using this approaches some problems can be simplified.

Keywords　Rough set, Information system, Dominance relation, Consistent approximation space

1　引言

粗糙集理论[1]是一种处理不精确、不确定和模糊知识的软计算工具，它已被成功地应用于人工智能、数据挖掘、模式识别与之能信息处理等领域[2~5]，并越来越引起了国际学术界的关注。众所周知，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，这正是人们所期望的。

目前，许多学者对知识约简做了深入的研究，并取得了很多成果[6~11]，但是这些研究主要是在等价关系下的信息系统中进行的。然而，在实际问题中有许多信息系统是基于优势关系的，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，就存在一定的难度。因而，如何将优势关系下的不协调目标信息系统转化为一个协调的信息系统是非常有意义的。为此，本文对这一问题进行了探讨研究，给出了在优势关系下的协调近似空间，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，而且并没有改变原来信息系统的信息，还保持了与文[9]中的分配协调集的一致性，这样大大方便了我们的研究。

2　基于优势关系的目标信息系统

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和术语。

定义 1　称一个五元组 $S=(U,A,F,D,G)$ 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，$F=\{f_k: U\rightarrow V_k, k\leqslant p\}$，$V_k$ 是 a_k 的有限值域；

G 是 U 与 D 的关系集，$G=\{g_{k'}: U\rightarrow v'_k, k'\leqslant q\}$，$v'_k$ 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问题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　设 $S=(U, A, F, D, G)$ 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，用 $S^{\leqslant}$ 表示。

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　(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：$[x_i]_A^{\leqslant}\subseteq[x_i]_{B_2}^{\leqslant}\subseteq[x_i]_{B_1}^{\leqslant}$。

徐伟华　博士研究生，主要研究人工智能、粗糙集；张文修　博导，教授，主要研究模糊集、粗糙集、人工智能。

(4)当 $x_j \in [x_i]_B^{\geqslant}$ 时有：$[x_j]_B^{\geqslant} \subseteq [x_i]_B^{\geqslant}$。

定义 3 设 $S=(U,A,F,D,G)$ 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，或该目标信息系统在优势关系下是协调的，否则是不协调的。

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

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

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，详情请参考文[7]。

3 基于优势关系下的协调近似空间

这一部分中，我们给出了基于优势关系下协调近似空间的定义，并说明了它的作用。

定义 4 设 $S^{\leqslant}=(U,A,F,D,G)$ 为目标信息系统，若存在 $U \times U$ 上的二元关系 $R_0^{\leqslant}$，使得有 $R_A^{\leqslant} \subseteq R_0^{\leqslant}$，称 $S^{\leqslant}$ 为基于优势关系下的协调近似空间。

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

证明：若 $R_0^{\leqslant}=R_A^{\leqslant}$，由定义立即可得结论成立。 □

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

证明：显然只需取 $R_0^{\leqslant}=R_D^{\leqslant}$，由定义 3 知 $R_A^{\leqslant} \subseteq R_0^{\leqslant}$ 成立，故满足定义 4。 □

由以上可知，基于优势关系下的一般信息系统和协调目标信息系统都是基于优势关系下的协调近似空间，那么对于基于优势关系下的不协调目标信息系统是否也能够找到一个二元关系 $R_0^{\leqslant}$ 使之成为一个协调近似空间呢？

在文[9]中，作者给出了基于优势关系下不协调目标信息系统的分配协调集，这是否和我们这里的协调近似空间一致呢？下面我们给出说明。

设 $S^{\leqslant}=(U,A,F,D,G)$ 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 $B \subseteq A$，记

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

$U/R_D^{\leqslant}=\{D_1, D_2, \cdots, D_r\}$，

$\sigma_B^{\leqslant}(x)=\{D_j : D_j \cap [x]_B^{\leqslant} \neq \varnothing, x \in U\}$，

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

显然由以上可立即得到

命题 2 (1) $\overline{R_B^{\leqslant}}(D_j)=\bigcup\{[x]_B^{\leqslant} : D_j \in \sigma_B^{\leqslant}(x)\}$。

(2)当 $B \subseteq A$ 时，有 $\sigma_B^{\leqslant}(x) \supseteq \sigma_A^{\leqslant}(x)$，$\forall x \in U$。

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

文[9]中关于分配协调集的定义如下：

定义 5 设 $S^{\leqslant}=(U,A,F,D,G)$ 为目标信息系统，若 $\forall x \in U$，有 $\sigma_B^{\leqslant}(x)=\sigma_A^{\leqslant}(x)$，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。

下面，我们给出

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

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

称为不协调目标信息系统的分配协调优势关系。

定理 3 不协调目标信息系统 $S^{\leqslant}=(U,A,F,D,G)$ 一定是协调近似空间。

证明：只需证明可以找到 $R_0^{\leqslant}$，使得有 $R_A^{\leqslant} \subseteq R_0^{\leqslant}$ 成立。事实上只需取定义 6 中的 $R^{\leqslant'}=R_0^{\leqslant}$ 即可。即只需证明：对于 $(x_i,x_j) \in U \times U$，若 $(x_i,x_j) \in R_A^{\leqslant}$ 则 $(x_i,x_j) \in R^{\leqslant'}$。

由于 $(x_i,x_j) \in R_A^{\leqslant}$，即 $x_j \in [x_i]_A^{\leqslant}$，由命题 1(4)可知 $[x_i]_A^{\leqslant} \supseteq [x_j]_A^{\leqslant}$，若 $D_k \in \sigma_A^{\leqslant}(x_j)$，则 $D_k \cap [x_j]_A^{\leqslant} \neq \varnothing$，故有 $D_k \cap [x_i]_A^{\leqslant} \neq \varnothing$，于是 $D_k \in \sigma_A^{\leqslant}(x_i)$，因此有 $\sigma_A^{\leqslant}(x_i) \supseteq \sigma_A^{\leqslant}(x_j)$，即 $(x_i,x_j) \in R^{\leqslant'}$ 成立。

定理 4 设 $S^{\leqslant}=(U,A,F,D,G)$ 为不协调目标信息系统，$R^{\leqslant'}$ 为其分配协调优势关系，$B \subseteq A$，则有：$R_B^{\leqslant} \subseteq R^{\leqslant'} \Leftrightarrow \sigma_A^{\leqslant}(x)=\sigma_B^{\leqslant}(x)$，对 $\forall x \in U$。

证明："⇒"设 $\forall x_i \in U$，若 $D_k \in \sigma_B^{\leqslant}(x_i)$ 则 $[x_i]_B^{\leqslant} \cap D_k \neq \varnothing$。不妨设 $x_j \in [x_i]_B^{\leqslant} \cap D_k$，则 $x_j \in D_k$ 且 $x_j \in [x_i]_B^{\leqslant}$，故 $(x_i,x_j) \in R_B^{\leqslant} \subseteq R^{\leqslant'}$。因此有：$\sigma_A^{\leqslant}(x_i) \supseteq \sigma_A^{\leqslant}(x_j)$。又因为 $x_j \in [x_j]_A^{\leqslant}$，于是 $x_j \in [x_j]_A^{\leqslant} \cap D_k$，即 $[x_j]_A^{\leqslant} \cap D_k \neq \varnothing$，故 $D_k \in \sigma_A^{\leqslant}(x_j)$。而又上面得到了 $\sigma_A^{\leqslant}(x_i) \supseteq \sigma_A^{\leqslant}(x_j)$，所以有 $D_k \in \sigma_A^{\leqslant}(x_i)$，于是可得 $\sigma_B^{\leqslant}(x_i) \subseteq \sigma_A^{\leqslant}(x_i)$ 成立。因此对 $\forall x \in U$ 有 $\sigma_B^{\leqslant}(x) \subseteq \sigma_A^{\leqslant}(x)$。

另一方面，由命题 2(2)知对 $\forall x \in U$ 有 $\sigma_A^{\leqslant}(x) \subseteq \sigma_B^{\leqslant}(x)$，因此对 $\forall x \in U$ 有 $\sigma_A^{\leqslant}(x)=\sigma_B^{\leqslant}(x)$。

"⇐"若 $(x_i,x_j) \in R_B^{\leqslant}$，则有 $[x_i]_B^{\leqslant} \supseteq [x_j]_B^{\leqslant}$。另设 $D_k \in \sigma_A^{\leqslant}(x_j)$，又对 $\forall x \in U$ 有 $\sigma_A^{\leqslant}(x)=\sigma_B^{\leqslant}(x)$，于是有 $D_k \in \sigma_B^{\leqslant}(x_j)$，即 $D_k \cap [x_j]_B^{\leqslant} \neq \varnothing$，故 $D_k \cap [x_i]_B^{\leqslant} \neq \varnothing$，所以 $D_k \in \sigma_B^{\leqslant}(x_i)=\sigma_A^{\leqslant}(x_i)$，因此可得 $\sigma_A^{\leqslant}(x_i) \supseteq \sigma_A^{\leqslant}(x_j)$，即 $(x_i,x_j) \in R^{\leqslant'}$。 □

通过上述定理，我们可以看出利用分配协调关系 $R^{\leqslant'}$ 使不协调目标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[9]中的分配协调集的一致性，这样大大方便了我们的研究。

结论 我们知道，在实际问题中有许多信息系统是基于优势关系的，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，就存在一定的难度。因而，如何将优势关系下的不协调目标信息系统转化为一个协调的信息系统是非常有意义的。为此，本文对这一问题进行了探讨研究，给出了在优势关系下的协调近似空间，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，而且并没有改变原来信息系统的信息，还保持了与文[9]中的分配协调集的一致性，这样大大方便了我们的研究。然而，等价意义下的分布约简在优势关系下并不一定成立，如何找出基于优势关系下不协调目标信息系统的分布约简，并可转化为一个基于又是关系下的协调近似空间，这将是我们接下来的研究内容。

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. 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 苗夺谦，胡贵容. 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

5 王珏，王任，苗夺谦，等. 基于 Rough Set 理论的"数据浓缩". 计算机学报，1998，21(5)：393～400

6 张文修，吴伟志，梁吉业，李德玉. 粗糙集理论与方法. 北京：科学出版社，2001

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 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(已投稿)

10 Kryszkiewicz M. Comparative Studies of Alternative of Knowledge Reductin in Inconsiostent Systems. Intelligent Systems, 2001, 16(1): 105～120

11 Grecos, Matarazzo B, Slowinski R. Rough Approximation of Preference Relation by Dominance relations. European Journal of Operational Research, 1999, 117: 63～68