DOI:10.3969/j.issn.1673-4785.201307012
网络出版地址:http://www.cnki.net/kcms/doi/CNKI:23-1538/TP.20131105.1201.003.html

α 优势关系下粗糙集模型的属性约简

韦碧鹏[1,2],吕跃进[2],李金海[3]

(1.柳州职业技术学院 公共基础部,广西 柳州 545006; 2. 广西大学 数学与信息科学学院,广西 南宁 530004; 3.昆明理工大学 理学院,云南 昆明 650500)

摘　要:不完备序信息系统中现有优势关系存在一些不足,提出了 α 优势关系的概念;然后在 α 优势关系的粗糙集模型上,构造了不完备序信息系统以及序决策系统的优势区分矩阵和优势决策区分矩阵。在此基础上设计出了不完备序信息系统以及序决策系统的属性约简算法。此外,对比分析表明了 α 优势关系既具备了现有优势关系的优点,克服了它们的缺点。实例分析验证了所提出方法的有效性。

关键词:粗糙集理论;不完备序信息系统;不完备序决策系统; α 优势关系;属性约简

中图分类号: TP18　**文献标志码**:A　**文章编号**:1673-4785(2014)02-0251-08

中文引用格式:韦碧鹏, 吕跃进,李金海. α 优势关系下粗糙集模型的属性约简[J]. 智能系统学报, 2014, 9(2): 251-258.

英文引用格式:WEI Bipeng, LÜ Yuejin, LI Jinhai. Attribute reduction based on the rough set model under α dominance relation [J]. CAAI Transactions on Intelligent Systems, 2014, 9(2): 251-258.

Attribute reduction based on the rough set model under α dominance relation

WEI Bipeng[1,2], LÜ Yuejin[2], LI Jinhai[3]

(1. Public Infrastructure Department, Liuzhou Vocational and Technical College, Liuzhou 545006, China; 2. School of Mathematics and Information Science, Guangxi University, Nanning 530004, China; 3. Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Through the analysis of the existing dominance relations in incomplete ordered information systems, the concept of 了 α dominance relation is presented. Based on the rough set model under the dominance relation, the notion of a dominance discernibility matrix of an incomplete ordered information system is proposed as well as that of a dominance decision discernibility matrix of an incomplete ordered decision system, and then the attribute reduction algorithms are developed in these two types of databases. Compared with the existing dominance relations, the proposed dominance relation not only has distinct advantages, but it also avoids the others' shortcomings. Finally, a real example was used to demonstrate the effectiveness of the presented algorithms.

Keywords: rough set theory; incomplete ordered information system; incomplete ordered decision system; α dominance relation; attribute reduction

波兰数学家 Pawlak 于 1982 年提出了粗糙集理论[1],它是一种处理模糊、不精确性以及不确定性的数学工具。近年来,由于它具有诸多优势,已经被成功地运用于数据挖掘、模式识别、数据处理、决策分析等领域[2-4]。然而,在现实生活中,由于噪声、测量数据的不完整性等因素,不完备信息系统依然广泛存在。而 Pawlak 提出的经典粗糙集并不适用不完备信息系统。这就有必要对它进行扩充以适用于处理不完备数据。目前,针对不完备信息系统缺失值的不同理解,对经典粗糙集的扩充研究有如下

收稿日期:2013-07-05.　**网络出版日期**:2013-11-05.
基金项目:国家自然科学基金资助项目(61305057, 71361002);广西自然科学基金资助项目(2013GXNSFAA019016);2014 年度广西高校科研一般项目资助项目(YB2014501).
通信作者:李金海. E-mail:jhlixjtu@163.com.

模型;Kryszkiewicz[5]提出了基于容差关系的粗糙集模型,把不完备信息系统中的缺失值看作是遗漏型的,即可以和任意的对象进行比较;Stefanowski 等[6]提出了基于非对称相似关系和量化容差关系的粗糙集模型,把不完备信息系统中的缺失值看作是缺席型的;为了克服扩展模型的不足,王国胤[7]又给出了基于限制容差关系的粗糙集模型。这方面的更多研究,可以查看文献[8-10]。

在现实世界中,很多的信息系统其属性值域可能具有偏序性。此时,经典粗糙集方法显得无能为力。为了能够处理具有偏序关系的信息系统,Greco 等[11]提出了基于优势关系的粗糙集模型,即运用优势关系代替等价关系。为了让优势关系下的粗糙集模型能够同时处理缺失值,Shao 等[12]把不完备序信息系统中的缺失值看作是遗漏型的,提出了基于优势关系的不完备序信息系统,并讨论了属性约简和规则提取。针对文献[12]提出的优势关系过于宽松,容易将实际不满足条件的对象误判为同一个优势类,胡明礼等[13]通过引入一个阈值得到了广义扩展优势关系的概念,并对规则提取进行了重新考虑。此外,针对优势系统中缺失值是缺席型的情况,杨习贝等[14]提出了相似优势关系的粗糙集模型,并对属性约简做了研究。然而这并没有弥补文献[12]提出的优势关系过于宽松的缺陷,并且改进后的模型还容易将一些本属于同一决策类的对象误判为不同决策类。这在一定程度上会影响文献[12]提出的模型决策分析的效果。因此,Luo 等[15]进一步又给出了限制优势粗糙集模型,既可以避免文献[12]提出的优势关系过于宽松的现象,又不容易将一些本属于同一决策类的对象误判为不同的决策类。但文献[15]提出的限制优势关系在某些情况下又显得过于严格,例如:当2个对象在各属性下的取值为 $x=(4,3,2,1)$, $y=(*,*,2,1)$ 时,且各属性值域的最大值均为4,那么根据文献[15]提出的限制优势关系,对象 x 是不优于对象 y 的;可现实生活中,对象 x 优于对象 y 的可能性是非常大的,这有可能导致数据集中的决策规则没有充分提取。

基于此,本文在分析不完备序信息系统中现有的几种扩展粗糙集模型的基础上,提出了基于 α 优势关系的粗糙集模型,它既吸收了其他扩展模型的优点,又能有效克服它们的局限性,更有利于处理现实中的数据集。此外,基于 α 优势关系的粗糙集模型,给出了不完备序信息系统和序决策系统的区分矩阵属性约简算法,并表明了其区分矩阵只能运用属性集的幂集进行构造,而不能简单地运用单属性集进行构造。

1　基本概念

四元组 $IS=(U,AT,V,f)$,其中 U 是非空有限的对象集合,AT 是非空有限的属性集合,AT 为条件属性集,$V=\bigcup_{a\in AT}V_a$,且 V_a 是属性 a 的值域,满足 $f_a:U\times a\rightarrow V_a$ ($\forall x\in U, a\in AT$),称四元组 IS 为一个信息系统。

若存在一个属性 $a\in AT$ 使得 V_a 为空值(记作:$f(x,a)=*$),则称该信息系统是不完备的信息系统,记作:IIS;否则称为完备的信息系统。

定义1[11]　设 $IS=(U,AT,V,f)$ 是一个完备信息系统,对于 $\forall A\subseteq AT$, $\forall x,y\in U$,有:

$$R_A^{\geqslant}=\{(x,y)\in U\times U\mid f(x,a)\geqslant f(y,a),\forall a\in A\}$$

则称 $R_A^{\geqslant}$ 为信息系统下的优势关系,满足这种关系的完备信息系统称之为序信息系统,记作:OIS。

记:$[x]_A^{\geqslant}=\{y\in U\mid (y,x)\in R_A^{\geqslant}\}=\{y\in U\mid f(y,a)\geqslant f(x,a),\forall a\in A\}$

为对象 x 在序信息系统中的优势类,且 $U/R_A^{\geqslant}=\{[x]_A^{\geqslant}\mid x\in U\}$。

定义2[12]　设 $IIS=(U,AT,V,f)$ 是一个不完备信息系统,对于 $A\subseteq AT$, $\forall x,y\in U$,有

$$R_A^{*\geqslant}=\{(x,y)\in U\times U\mid \forall a\in A, f(x,a)\geqslant f(y,a)\vee f(x,a)=*\vee f(y,a)=*\}$$

则称 $R_A^{*\geqslant}$ 是不完备信息系统上的一个优势关系,满足具有优势关系的不完备信息系统称之为不完备序信息系统,记作:$IOIS$。

通过 $R_A^{*\geqslant}$ 的定义,显然可以看出,在条件属性集 A 下,对象 x 优于对象 y。

记:$[y]_A^{*\geqslant}=\{x\in U\mid (x,y)\in R_A^{*\geqslant}\}$,则 $[y]_A^{*\geqslant}$ 描述的是:论域中的对象在条件属性 A 下,优于或者等于对象 y 的集合,简称为对象 y 的优势类。因此,$U/R_A^{*\geqslant}$ 中所有的优势类构成了论域 U 的覆盖,而不是论域 U 的划分,即 $\cup U/R_A^{*\geqslant}=U$。

定义3[12]　设 $IOIS=(U,AT,V,f)$ 是一个不完备序信息系统,$\forall X\subseteq U$, $A\subseteq AT$,对象集合 X 在优势关系 $R_A^{*\geqslant}$ 下关于属性集 $A\subseteq AT$ 的上下近似集为

$$\underline{R}_A^{*\geqslant}(X) = \{x \in U \mid [x]_A^{*\geqslant} \subseteq X\}$$

$$\overline{R}_A^{*\geqslant}(X) = \{x \in U \mid [x]_A^{*\geqslant} \cap X \neq \varnothing\}$$

通过分析定义 2,可以得出文献[12]在不完备序信息系统中提出的优势关系是以空值可以等于任意值为假设前提的,即认为空值可以优于任意值,同时又认为任意值可以优于空值,这显然不符合实际情况。例如:当 $x = (1,1,*)$, $y = (1,1,4)$, $z = (*,*,4)$ 时,其中“4”表示属性值下最大的取值,“1”为属性值下最小的取值。根据定义 2 进行分析可得:$(x,y) \in R_A^{*\geqslant}$, $(x,z) \in R_A^{*\geqslant}$, $(y,x) \in R_A^{*\geqslant}$, $(y,z) \in R_A^{*\geqslant}$, $(z,x) \in R_A^{*\geqslant}$, $(z,y) \in R_A^{*\geqslant}$ 成立。然而事实上,缺失值 $*$ 优于属性值 4 以及属性值 1 优于缺失值 $*$ 的可能性是非常之小,但是属性值 4 优于缺失值 $*$ 以及缺失值 $*$ 优于属性值 1 必然成立。因此,$(x,y) \in R_A^{*\geqslant}$, $(x,z) \in R_A^{*\geqslant}$, $(y,z) \in R_A^{*\geqslant}$ 成立的可能性是非常之小。这说明文献[12]提出的优势关系的划分粒度过大,容易把不属于优势类的对象归结到优势类中。针对文献[12]提出的不完备序信息系统粗糙集模型的缺陷,文献[15]提出了基于限制优势关系的粗糙集模型。

定义 4[15]　设 $IOIS = (U,AT,V,f)$ 是一个不完备序信息系统,对于 $A \subseteq AT$, $\forall x,y \in U$, 对象在属性集 A 下的限制优势关系为

$$R_A^{*L\geqslant} = \{(x,y) \in U^2 \mid \forall a \in A, f(x,a) \geqslant f(y,a) \vee (f(x,a) = \max V_a \wedge f(y,a) = *) \vee (f(x,a) = * \wedge f(y,a) = \min V_a)\} \cup I_U$$

其中 $\max V_a = \{v \in V_a \mid \forall v' \in V_a, v \geqslant v'\}$, $\min V_a = \{v \in V_a \mid, \forall v' \in V_a, v \leqslant v'\}$, I_U 是一个确定关系, $I_U = \{(x,x) \mid x \in U\}$。因此, $[y]_A^{*L\geqslant} = \{x \in U \mid (x,y) \in R_A^{*L\geqslant}\}$

定义 5[15]　设 $IOIS = (U,AT,V,f)$ 是一个不完备序信息系统, $\forall X \subseteq U$, $A \subseteq AT$, 对象集合 X 在优势关系 $R_A^{*L\geqslant}$ 下关于属性集 $A \subseteq AT$ 的上下近似集为

$$\underline{R}_A^{*L\geqslant}(X) = \{x \in U \mid [x]_A^{*L\geqslant} \subseteq X\}$$

$$\overline{R}_A^{*L\geqslant}(X) = \{x \in U \mid [x]_A^{*L\geqslant} \cap X \neq \varnothing\}$$

通过分析定义 4 提出的限制优势关系,可以得出文献[15]提出的限制优势关系过于严格,划分的粒度过细。例如:当 $x = (4,3,2,1)$, $y = (*,*,2,1)$ 时,其中“4”表示属性值下最大的取值,“1”为属性值下最小的取值。根据定义 4 分析可得:$(x,y) \notin R_A^{*L\geqslant}$。然而,事实上,$(x,y) \in R_A^{*L\geqslant}$ 的可能性是非常大。这说明文献[15]提出的限制优势关系划分的粒度过细,容易把一些本该属于优势类的对象误判为不是优势类的对象,这样容易导致提取决策规则的不完全性。基于此,在不完备序信息系统中,有必要提出一种相对于文献[12]提出的优势关系以及文献[15]提出的限制优势关系更加灵活的优势关系。

2　α 优势关系的粗糙集模型

设 $IOIS = (U,AT,V,f)$ 是一个不完备序信息系统,对于 $\forall a \subseteq AT$, $\forall x,y \in U$, 对象在属性 a 下的取值为 $V_a = \{a_1,a_2,\ldots,a_m\}$, 并且 $a_1 < a_2 < \ldots < a_m$, 根据 V_a 的排序,令 $a_1 = 1, a_2 = 2, \ldots, a_m = m$, 则对象在属性 a 下的取值转变为 $V_a' = \{1,2,\ldots,m\}$, 其中 $V_a'(x)$ 为对象 x 在 V_a' 中转化的值。

定义 6　设 $IOIS = (U,AT,V,f)$ 是一个不完备序信息系统,对于 $\forall a \subseteq AT$, $\forall x,y \in U$, 对象在属性 a 下的取值为 $V_a = \{a_1,a_2,\ldots,a_m\}$, 并且 $a_1 < a_2 < \ldots < a_m$, 则对象 x 在属性集 a 下优于 y 的概率为

$$R_a(x,y) = \begin{cases} 1, f(x,a) \geqslant f(y,a) \\ \dfrac{V_a'(x)}{m}, f(x,a) \neq * \wedge f(y,a) = * \\ \dfrac{m - V_a'(y) + 1}{m}, f(x,a) = * \wedge f(y,a) \neq * \\ \dfrac{1}{m}, f(x,a) = * \wedge f(y,a) = * \\ 0, f(x,a) < f(y,a) \end{cases}$$

从定义 6 中得知,在单属性下,2 个对象优于的程度在 0 和 1 之间。当对象 x 在属性 a 下取值完全小于对象 y 的取值时,用数值 0 来表示对象 x 劣于对象 y 的程度,即概率;当对象 x 在属性 a 下取值完全大于对象 y 的取值时,用数值 1 来表示对象 x 优于对象 y 的程度。然而,当对象 x 和对象 y 有一个缺失值时,且对象在属性下取值的多样性,使得很难确定 2 个对象的优于程度;根据概率的含义,即在 m 个数中,有 n 个数优于一个确定数值的概率为 n/m, 得出定义 6,2 个对象中有一个为缺失值时它们的优于程度。当 2 个对象都为缺失值时,没有根据可以判别它们之间优于程度,因此,运用 $1/m$ 概率来表示,意味着优于程度很小。

定义 7 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于 $A \subseteq AT$, $\forall x, y \in U$，则对象 x 在属性集 A 下优于 y 的概率为

$$R_A(x, y) = \prod_{a \in A} R_a(x, y)$$

通过定义 7，可以得出不完备序信息系统各个对象的优势类如下：

定义 8 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于 $A \subseteq AT$, $\forall x, y \in U$，有

$$R_A^{*\alpha\geqslant} = \{(x, y) \in U \times U \mid R_A(x, y) \geqslant \alpha\}$$

则称 $R_A^{*\alpha\geqslant}$ 是不完备序信息系统上的一个 α 优势关系，其中 $0 < \alpha \leqslant 1$。

记：$[y]_A^{*\alpha\geqslant} = \{x \in U \mid (x, y) \in R_A^{*\alpha\geqslant}\}$，则 $[y]_A^{*\alpha\geqslant}$ 描述的对象是：在条件属性集 A 下，以 α 的优势度优于对象 y 的最大对象的集合，简称为对象 y 的 α 优势类。由于 $U/R_A^{*\alpha\geqslant}$ 表示所有与优势关系族 $R_A^{*\alpha\geqslant}$ 相关的知识，记作 $U/R_A^{*\alpha\geqslant} = \{[x]_A^{*\alpha\geqslant} \mid x \in U\}$，则 $U/R_A^{*\alpha\geqslant}$ 中的任何一个元素都是属于 α 优势类。从而可以得知 $U/R_A^{*\alpha\geqslant}$ 中所有的 α 优势类构成了 U 中的覆盖，而不是 U 中的划分，即 $\cup U/R_A^{*\alpha\geqslant} = U$。

根据定义 8，结合定义 2 和定义 4 的概念，显然可以得出：当 $\alpha > 0$ 时，$R_A^{*\alpha\geqslant} = R_A^{*\geqslant}$；当 $\alpha = 1$ 时，$R_A^{*\alpha\geqslant} = R_A^{*L\geqslant}$。因此，在不完备序信息系统中，本文提出的 α 优势关系是文献[12]提出的优势关系和文献[15]提出的限制优势关系的一种扩展形式。

性质 1 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，则 α 优势关系满足如下性质：

1) $R_A^{*\alpha\geqslant}$ 满足自反性，不满足对称性和传递性；

2) 当 $B \subseteq A \subseteq AT$ 时，$R_{AT}^{*\alpha\geqslant} \subseteq R_A^{*\alpha\geqslant} \subseteq R_B^{*\alpha\geqslant}$；

3) 当 $B \subseteq A \subseteq AT$ 时，$\forall x \in U$，

$$[x]_{AT}^{*\alpha\geqslant} \subseteq [x]_A^{*\alpha\geqslant} \subseteq [x]_B^{*\alpha\geqslant}。$$

性质 2 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于 $A \subseteq AT$，$R_A^{*\geqslant}$、$R_A^{*L\geqslant}$、$R_A^{*\alpha\geqslant}$ 三者优势关系满足如下性质：

1) $R_A^{*\geqslant} \prec R_A^{*L\geqslant} \prec R_A^{*\alpha\geqslant}$，$\prec$ 表示优于；

2) $\forall x \in U$，$[x]_A^{*L\geqslant} \subseteq [x]_A^{*\alpha\geqslant} \subseteq [x]_A^{*\geqslant}$。

定义 9 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，$\forall X \subseteq U$，$A \subseteq AT$，对象集合 X 在优势关系 $R_A^{*\alpha\geqslant}$ 下关于属性集 $A \subseteq AT$ 的上下近似集为：

$$\underline{R}_A^{*\alpha\geqslant}(X) = \{x \in U \mid [x]_A^{*\alpha\geqslant} \subseteq X\} = \{x \in X \mid [x]_A^{*\alpha\geqslant} \subseteq X\}$$

$$\bar{R}_A^{*\alpha\geqslant}(X) = \{x \in U \mid [x]_A^{*\alpha\geqslant} \cap X \neq \varnothing\} = \cup\{[x]_A^{*\alpha\geqslant} \mid x \in X\}$$

可以看出，$\underline{R}_A^{*\alpha\geqslant}(X)$ 是确定属于 X 的一些对象集，然而 $\bar{R}_A^{*\alpha\geqslant}(X)$ 是可能属于 X 的一些对象集。

定理 1 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于属性集 $\forall A \subseteq AT$，对象集 $\forall X \subseteq U$，则可以得出 $\underline{R}_A^{*\geqslant}(X) \subseteq \underline{R}_A^{*\alpha\geqslant}(X) \subseteq \underline{R}_A^{*L\geqslant}(X)$，$\bar{R}_A^{*L\geqslant}(X) \subseteq \bar{R}_A^{*\alpha\geqslant}(X) \subseteq \bar{R}_A^{*\geqslant}(X)$。

证明 由性质 2 可得，$\forall A \subseteq AT, \forall x \in U$，$[x]_A^{*L\geqslant} \subseteq [x]_A^{*\alpha\geqslant} \subseteq [x]_A^{*\geqslant}$，结合三者关系上下近似的定义，$\forall X \subseteq U$，得出 $\underline{R}_A^{*\geqslant}(X) \subseteq \underline{R}_{A(X)}^{*\alpha\geqslant} \subseteq \underline{R}_A^{*L\geqslant}(X)$，$\bar{R}_A^{*L\geqslant}(X) \subseteq \bar{R}_A^{*\alpha\geqslant}(X) \subseteq \bar{R}_A^{*\geqslant}(X)$。证毕。

从定理 1 可知，当 $\alpha > 0$ 时，$\underline{R}_A^{*\alpha\geqslant}X = \underline{R}_A^{*\geqslant}X$，$\bar{R}_A^{*\alpha\geqslant}X = \bar{R}_A^{*\geqslant}X$；当 $\alpha = 1$ 时，$\underline{R}_A^{*\alpha\geqslant}X = \underline{R}_A^{*L\geqslant}X$，$\bar{R}_A^{*L\geqslant}X = \bar{R}_A^{*\alpha\geqslant}X$。从而又可以验证了基于 α 优势关系的粗糙集模型是文献[12]提出的优势关系粗糙集模型和文献[15]提出的限制优势关系粗糙集模型的一种扩展形式；因此，本文提出的 α 优势关系的粗糙集模型更具有灵活性，更加符合实际情况，更有利于去处理现实中存在的不完备序信息系统。

3 基于 α 优势关系粗糙集模型的属性约简

3.1 不完备序信息系统的属性约简

定义 10 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于属性集 $A \subseteq AT$，称属性集 A 是 $IOIS$ 的一个优势约简当且仅当 $R_A^{*\alpha\geqslant} = R_{AT}^{*\alpha\geqslant}$，且 $\forall B \subset A, R_B^{*\alpha\geqslant} \neq R_{AT}^{*\alpha\geqslant}$ 成立。

定义 10 表明了基于 α 优势关系的不完备序信息系统的属性约简保持的是对象的 α 优势类不变的最小属性组成的集合。基于此，下面给出其优势区分矩阵构造的方法：

定义 11 设 $IOIS = (U, AT, V, f)$ 是一个不完备序信息系统，对于 $\forall x, y \in U$，有

$$D_{AT}^{*\alpha\geqslant}(x, y) = \begin{cases} \{A \in AT \mid y \notin [x]_A^{*\alpha\geqslant}\}, y \notin [x]_{AT}^{*\alpha\geqslant} \\ AT, y \in [x]_{AT}^{*\alpha\geqslant} \end{cases}$$

称 $D_{AT}^{*\alpha\geqslant} = \{D_{AT}^{*\alpha\geqslant}(x, y) \mid \forall x, y \in U\}$ 为 $IOIS$ 关于 α 优势关系的优势区分矩阵，其中属性集 A 为属性集 AT 的幂集。

根据对以往知识的了解，区分矩阵的构造是可

以仅仅在单元素下进行。但是，在 α 优势关系的粗糙集模型中，这样的构造是不成立的，如

假设基于 α 优势关系的优势区分矩阵可以如下构造：

$$D_{AT}^{*\alpha\geqslant}(x,y)=\begin{cases}\{a\in AT\mid y\notin[x]_a^{*\alpha\geqslant}\},y\notin[x]_{AT}^{*\alpha\geqslant}\\ AT,y\in[x]_{AT}^{*\alpha\geqslant}\end{cases}$$

存在这样的不完备序信息系统：$U=\{x,y\}$，$AT=\{a,b,c\}$，$x=(3,*,2)$，$y=(*,2,1)$，其中在属性 a 下，最大的取值为 4；在属性 b 下，最大的取值为 3；根据定义 6 和定义 7，计算得出对象在条件属性集下的优势概率为：$R_{AT}(x,y)=3/4\times 2/3=1/2$。当 $\alpha=0.6$ 时，可以得出对象 x 不是对象 y 的 α 优势类，即 $x\notin[y]_{AT}^{*\alpha\geqslant}$；因此，运用单元素构成的优势区分矩阵对不完备序信息系统进行属性约简，可以得出其约简为空集，这显然不符合属性约简的概念。但是，当根据定义 11，即运用属性集的幂集构成的优势区分矩阵对该不完备序信息系统进行属性约简时，可以得出属性集 $\{a,b\}$ 为该系统的约简。显然，可以验证基于 α 优势关系粗糙集模型的优势区分矩阵应该是在属性集的幂集下进行的。

定理 2　设 $IOIS=(U,AT,V,f)$ 是一个不完备序信息系统，对于 $A\subseteq AT$，$D_{AT}^{*\alpha\geqslant}(x,y)$ 是基于 α 优势关系的优势区分矩阵，$\forall x,y\in U$，$A\cap D_{AT}^{*\alpha\geqslant}(x,y)\neq\varnothing(D_{AT}^{*\alpha\geqslant}(x,y)\neq\varnothing)$ 当且仅当 $R_A^{*\alpha\geqslant}=R_{AT}^{*\alpha\geqslant}$。

证明　（$\Rightarrow$）对于 $\forall x,y\in U,A\subseteq AT$，由性质 1 可得 $R_{AT}^{*\alpha\geqslant}\subseteq R_A^{*\alpha\geqslant}$。要证明 $R_A^{*\alpha\geqslant}=R_{AT}^{*\alpha\geqslant}$，因此只需要证明 $R_{AT}^{*\alpha\geqslant}\supseteq R_A^{*\alpha\geqslant}$ 即可。然而由于 $A\cap D_{AT}^{*\alpha\geqslant}(x,y)\neq\varnothing(D_{AT}^{*\alpha\geqslant}(x,y)\neq\varnothing)$，即存在 $a\in A,a\in D_{AT}^{*\alpha\geqslant}(x,y)$，可以得出 $y\notin[x]_a^{*\alpha\geqslant}$，即 $y\notin[x]_A^{*\alpha\geqslant}$，则 $y\notin[x]_{AT}^{*\alpha\geqslant}$；因为 y 是任意的，所以可以得出 $R_{AT}^{*\alpha\geqslant}\supseteq R_A^{*\alpha\geqslant}$，即 $R_A^{*\alpha\geqslant}=R_{AT}^{*\alpha\geqslant}$。

（$\Leftarrow$）由于 $R_A^{*\alpha\geqslant}=R_{AT}^{*\alpha\geqslant}$，可得 $\forall x\in U$，$[x]^{*\alpha\geqslant}{}_A=[x]_{AT}^{*\alpha\geqslant}$，则存在 $a\in A$，当 $y\notin[x]_{AT}^{*\alpha\geqslant}$ 时，$y\notin[x]_a^{*\alpha\geqslant}$，即 $A\cap D_{AT}^{*\alpha\geqslant}(x,y)\neq\varnothing$。证毕。

因此，令 $\Delta^{*\geqslant}=\bigwedge_{(x,y)\in U\times U}\bigvee D_{AT}^{*\alpha\geqslant}(x,y)$ 为 α 优势关系下不完备序信息系统的区分函数，$\Delta^{*\geqslant}(x)=\bigwedge_{y\in U}\bigvee D_{AT}^{*\alpha\geqslant}(x,y)$ 为对象 x 在 α 优势关系下不完备序信息系统的区分函数。

在 α 优势关系的不完备序信息系统中，由于区分函数 $\Delta^{*\geqslant}=\bigwedge_{(x,y)\in U\times U}\bigvee D_{AT}^{*\alpha\geqslant}(x,y)$ 中 $D_{AT}^{*\alpha\geqslant}(x,y)$ 是属性集的幂集组成的集合，因此，运用 $\wedge$ 作为连接符。例如：$D_{AT}^{*\alpha\geqslant}(x,y)$ 中区分这 2 个对象的属性集合为 $a\wedge b$，可以写成 ab。即在优势区分矩阵中，“，”表示“或”含义。

推论 1　设 $IOIS=(U,AT,V,f)$ 是一个不完备序信息系统，对于 $AT=\{a,b,c\}$，若区分函数 $D_{AT}^{*\alpha\geqslant}(x,y)$ 中的 2 个对象可以运用属性集 a、b、ab、abc 加以区分，则 ab、abc 可以省略。

推论 2　设 $IOIS=(U,AT,V,f)$ 是一个不完备序信息系统，对于 $AT=\{a,b,c\}$，若区分函数 $D_{AT}^{*\alpha\geqslant}(x,y)$ 中的 2 个对象仅仅能用属性集 ab 来区分，则属性集 ab 为这 2 个对象的区分属性集。

3.2　不完备序决策系统的属性约简

在 α 优势关系的不完备序信息系统中，若添加决策属性 d，则不完备序信息系统可以转变为 $IODS=(U,AT\cup\{d\},V,f)$ 情形，其中属性 d 也是一个准则，称之为决策属性，$d\notin AT$ 且 $*\notin V_d$，则称 $IODS$ 为不完备序决策系统。

记：$R_d^{*\alpha\geqslant}=\{(x,y)\mid f(x,d)\geqslant f(y,d)\}$，则称 $R_d^{*\alpha\geqslant}$ 为决策属性 d 的一个优势关系。因此，可以记作：$[y]_d^{*\alpha\geqslant}=\{x\in U\mid(x,y)\in R_d^{*\alpha\geqslant}\}$。

定义 12　设 $IODS=(U,AT\cup\{d\},V,f)$ 是一个不完备序决策系统，若 $R_{AT}^{*\alpha\geqslant}\subseteq R_d^{*\alpha\geqslant}$，则称 α 优势关系的不完备序决策系统是协调的；否则称为不协调的。

定义 13　设 $IODS=(U,AT\cup\{d\},V,f)$ 是一个不完备序决策系统，对于属性集 $A\subseteq AT$，称属性集 A 是 $IODS$ 的一个优势约简当且仅当 $R_A^{*\alpha\geqslant}\subseteq R_d^{*\alpha\geqslant}$，且 $\forall B\subset A,R_B^{*\alpha\geqslant}\not\subset R_d^{*\alpha\geqslant}$ 成立。

从定义 13 中可以得出，基于 α 优势关系的不完备序决策系统的属性约简保持 α 优势关系的协调性不变的最小子集。下面给出其优势决策区分矩阵构造的方法。

定义 14　设 $IODS=(U,AT\cup\{d\},V,f)$ 是一个不完备序决策系统，对于 $\forall x,y\in U$，有

$$D_d^{*\alpha\geqslant}(x,y)=\begin{cases}\{A\in AT\mid y\notin[x]_A^{*\alpha\geqslant}\},y\notin[x]_d^{*\alpha\geqslant}\\ AT,y\in[x]_d^{*\alpha\geqslant}\end{cases}$$

其中 $D_d^{*\alpha\geqslant}=\{D_d^{*\alpha\geqslant}(x,y)\mid\forall x,y\in U\}$ 称为 $IODS$ 中基于 α 优势关系的优势决策区分矩阵。

定理 3　设 $IODS=(U,AT\cup\{d\},V,f)$ 是一个

不完备序决策系统，对于 $A \subseteq AT$，$D_d^{*\alpha\geqslant}(x,y)$ 是基于 α 优势关系的优势决策区分矩阵，$\forall x,y \in U$，$A \cap D_d^{*\alpha\geqslant}(x,y) \neq \varnothing(D_d^{*\alpha\geqslant}(x,y) \neq \varnothing)$ 当且仅当 $R_A^{*\alpha\geqslant} \subseteq R_d^{*\alpha\geqslant}$。

证明 （$\Rightarrow$）对于 $\forall x \in U, A \subseteq AT$，根据性质1，可以得出 $[x]_{AT}^{*\alpha\geqslant} \subseteq [x]_A^{*\alpha\geqslant}$。又由于 $\forall x,y \in U, A \cap D_d^{*\alpha\geqslant}(x,y) \neq \varnothing(D_d^{*\alpha\geqslant}(x,y) \neq \varnothing)$，即 $y \notin [x]_d^{*\alpha\geqslant}, y \notin [x]_A^{*\alpha\geqslant}$，由于对象 y 的任意性，则可以得出 $[x]_A^{*\alpha\geqslant} \subseteq [x]_d^{*\alpha\geqslant}$，即 $R_A^{*\alpha\geqslant} \subseteq R_d^{*\alpha\geqslant}$。

（$\Leftarrow$）由于 $R_A^{*\alpha\geqslant} \subseteq R_d^{*\alpha\geqslant}$，则 $\forall x \in U$，有 $[x]_A^{*\alpha\geqslant} \subseteq [x]_d^{*\alpha\geqslant}$ 成立。即当 $y \notin [x]_d^{*\alpha\geqslant}$ 时，$y \notin [x]_A^{*\alpha\geqslant}$，因此 $A \cap D_d^{*\alpha\geqslant}(x,y) \neq \varnothing$。证毕。

因此，$\Delta_d^{*\geqslant} = \bigwedge_{(x,y)\in U\times U} \vee D_d^{*\alpha\geqslant}(x,y)$ 为 α 优势关系下不完备序决策系统的区分函数，$\Delta_d^{*\geqslant}(x) = \bigwedge_{y\in U} \vee D_d^{*\alpha\geqslant}(x,y)$ 为对象 x 在 α 优势关系下不完备序决策系统的区分函数。

4 实例分析

例 1 下面给出一个不完备序决策系统，其中 $U=\{1,2,3,4,5\}$，$AT=\{a,b,c\}$ 为条件属性，d 为决策属性，且对象在单个条件属性 a、b、c 下最大的取值分别为 4、3、3，最小值都是 1。运用表 1 给出的不完备序决策系统来分析文献[12]提出的优势关系、文献[15]提出的限制优势关系以及本文提出的 α 优势关系之间的性能；并计算 α 优势关系不完备序信息系统与决策系统的属性约简。

表 1 不完备序决策系统

Table 1 Incomplete ordered decision system

U	a	b	c	d
1	1	1	1	1
2	4	3	*	2
3	3	*	2	2
4	1	2	3	3
5	*	2	1	2

1）把表 1 中的决策属性去掉，不完备序决策系统转化为不完备序信息系统，分析三者优势关系之间的关系。

① 根据定义 2，可得文献[12]提出的优势关系下，各个对象的优势类为：$[1]_{AT}^{*\geqslant}=\{1,2,3,4,5\}$，$[2]_{AT}^{*\geqslant}=\{2\}$，$[3]_{AT}^{*\geqslant}=\{2,3\}$，$[4]_{AT}^{*\geqslant}=\{2,4\}$，$[5]_{AT}^{*\geqslant}=\{2,3,4,5\}$。

② 根据定义 4，可以得出文献[15]提出的限制优势关系下，各个对象的优势类为：$[1]_{AT}^{*L\geqslant}=\{1,2,3,4,5\}$，$[2]_{AT}^{*L\geqslant}=\{2\}$，$[3]_{AT}^{*L\geqslant}=\{3\}$，$[4]_{AT}^{*L\geqslant}=\{4\}$，$[5]_{AT}^{*L\geqslant}=\{2,5\}$。

③ 令 $\alpha=0.6$，根据定义 8，可得 α 优势关系下，各个对象的 α 优势类为：$[1]_{AT}^{*\alpha\geqslant}=\{1,2,3,4,5\}$，$[2]_{AT}^{*\alpha\geqslant}=\{2\}$，$[3]_{AT}^{*\alpha\geqslant}=\{2,3\}$，$[4]_{AT}^{*\alpha\geqslant}=\{4\}$，$[5]_{AT}^{*\alpha\geqslant}=\{2,5\}$。

根据定义得出三者优势关系下各个对象的优势类，在文献[12]提出的优势关系下，对象 4 的优势类为对象 2、4，然而在条件属性 c 下，由于对象 4 取得最大值，因此，对象 2 优于对象 4 的可能性是非常之小；另外，对象 5 的优势类为对象 2、3、4、5，然而在条件属性 a 下，对象 4 优于对象 5 的可能性也是非常之小。因此，可以说文献[12]提出的优势关系划分的粒度过大了。在文献[15]提出的限制优势关系下，对象 3 的优势类为对象 3，然而对象 2 优于对象 3 的可能性是非常之大。因此，说明了文献[15]提出的限制优势关系划分的粒度过细了。然而，在 α 优势关系下，设定确定的 α 取值，各个对象的优势类就不会出现上述的情况；可以看出，α 优势关系吸取了上述两者优势关系的优点，丢弃了两者的缺陷，更加符合实际情况，更有利于去处理现实生活中存在复杂的不完备序信息系统。

2）不完备序信息系统的属性约简

根据定义 11，结合 α 优势类，可以得出不完备序信息系统的优势区分矩阵如表 2。

表 2 优势区分矩阵

Table 2 Dominance discernibility matrix

U	1	2	3	4	5
1	a,b,c	a,b,c	a,b,c	b,c	a,b
2	a,b,c	a,b,c	a,b,c	c	a,b,c
3	a,b,c	a,b	a,b,c	c	ab
4	a,b,c	a,b	a	a,b,c	a
5	a,b,c	a,b,c	a,c	c	a,b,c

故 $\Delta^{*\geqslant}=(a \vee b \vee c) \wedge (a \vee b) \wedge a \wedge c \wedge (a \vee c) \wedge (a \wedge b) \wedge (b \vee c)=abc$，即该不完备序信息系统的属性约简为 abc。

3）不完备序决策系统的属性约简。

由于各个对象在决策属性下的优势类如下：

$[1]_d^{*\alpha\geqslant}=\{1,2,3,4,5\}$, $[2]_d^{*\alpha\geqslant}=\{2,3,4,5\}$, $[3]_d^{*\alpha\geqslant}=\{2,3,4,5\}$, $[4]_d^{*\alpha\geqslant}=\{4\}$, $[5]_d^{*\alpha\geqslant}=\{2,3,4,5\}$。

因此,得出 $R_{AT}^{*\alpha\geqslant}\subseteq R_d^{*\alpha\geqslant}$, 即该不完备序决策系统是协调的。

根据定义 14,可以得出不完备序决策系统的优势决策区分矩阵如表 3。

表 3　优势决策区分矩阵

Table 3　Dominance decision discernibility matrix

U	1	2	3	4	5
1	a,b,c	a,b,c	a,b,c	b,c	a,b
2	a,b,c	a,b,c	a,b,c	c	a,b,c
3	a,b,c	a,b,c	a,b,c	c	a,b,c
4	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
5	a,b,c	a,b,c	a,b,c	c	a,b,c

$\Delta_d^{*\geqslant}=(a\vee b\vee c)\wedge(b\vee c)\wedge c\wedge(a\vee b)=ac\vee bc$,即表 1 中不完备序决策系统的属性约简为 $\{a,c\}$ 或者 $\{b,c\}$ 。

5　结束语

由于现实中处理的数据很多是不完备的,且存在偏序性,因此研究处理这种复杂数据情况的粗糙集方法是很有实际意义的。本文通过对现有优势关系的分析后提出了 α 优势关系及其相应的粗糙集模型,以使得对不完备序信息系统的数据分析更加合理。此外,在基于 α 优势关系的粗糙集模型上,给出了不完备序信息系统的优势区分矩阵以及不完备序决策系统的优势决策区分矩阵,从而实现了属性约简,同时也表明了区分矩阵只能运用属性集的幂集进行构造,而不能运用单个属性集进行构造。

需要指出的是,在 α 优势关系的基础上,还可以进一步研究不协调不完备序决策系统的属性约简算法,这是本文的下一步工作任务。

参考文献:

[1] PAWLAK Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[2] 张文修,吴伟志,梁吉业,等.粗糙集理论与方法[M].北京:科学出版社, 2001: 14-19.

[3] 李金海,吕跃进.决策系统的快速属性约简算法[J].电子科技大学学报, 2007, 36(6): 1237-1240.
LI Jinhai, LÜ Yuejin. Quick attribute reduction algorithm on decision system[J]. Journal of University of Electronic Science and Technology of China, 2007, 36(6): 1237-1240.

[4] 覃丽珍,姚炳学,李金海.基于信息量的完备覆盖约简算法[J]. 计算机科学, 2012, 39(10): 235-239.
QIN Lizhen, YAO Bingxue, LI Jinhai. Complete algorithm for covering reduction based on information quantity [J]. Computer science, 2012, 39(10): 235-239.

[5] KRYSZKIEWICZ M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112: 39-49.

[6] STEFANOWSKI J, TSOUKIAS A. On the extension of rough sets under incomplete information[C]//Proceedings of New Directions in Rough Sets, Data Mining and Granular-Soft Computing. Berlin: Springer, 1999: 73-81.

[7] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39 (10): 1238-1243.
WANG Guoyin. Extension of rough set under incomplete information systems[J]. Journal of Computer Research and Development, 2002, 39 (10): 1238-1243.

[8] LIANG J Y, QU K S. Information measures of roughness of knowledge and rough sets in incomplete information systems [J]. Journal of System Science and System Engineering, 2001, 24(5): 544-547.

[9] GRZYMALA-BUSSE J W. Characteristic relations for incomplete data: a generalization of indiscernibility relation[C]// Proceedings of Rough Sets and Current Trends in Computing. Berlin: Springer, 2004: 244-253.

[10] 杨习贝, 杨静宇, 吴陈, 等. 不完备信息系统中的集对分析方法[J]. 计算机科学, 2007, 34(4): 171-174.
YANG Xibei, YANG Jingyu, WU Chen, et al. Set pair analysis in incomplete information systems[J]. Information Science, 2007, 34(4): 171-174.

[11] GRECO S, MATARAZZO B, SLOWINGSKI R. Rough sets theory for multicriteria decision analysis [J]. European Journal of Operational Research, 2001, 129(1): 1-47.

[12] 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.

[13] 胡明礼, 刘思峰. 基于广义扩展优势关系的粗糙决策分析方法[J]. 控制与决策, 2007, 22(12): 1347-1351.
HU Mingli, LIU Sifeng. Rough analysis method of multi-attribute decision making based on generalized extended dominance relation[J]. Control and Decision, 2007, 22 (12): 1347-1351.

[14] YANG X B, YANG J Y, WU C, et al. Dominance-based

rough set approach and knowledge reductions in incomplete ordered information system [J]. Information Sciences, 2008, 178: 1219-1234.

[15] LUO G Z, YANG X B. Limited dominance-based rough set model and knowledge reductions in incomplete decision system[J]. Journal of Information Science and Engineering, 2010, 26: 2199-2211.

作者简介：

韦碧鹏，男，1987 年生，硕士，主要研究方向为粗糙集的理论、粒计算，规则提取。近年来，发表学术论文 5 篇，其中 EI 检索 2 篇。

吕跃进，男，1958 年生，教授，中国工业与应用数学学会理事、数学模型专业委员会委员，中国复杂系统研究会理事，广西数学学会理事、数学模型专业委员会主任。主要研究方向为决策理论与方法，数据挖掘。2006 年获广西高校首届教学名师奖，主持 2 项国家、2 项广西自然科学基金项目，发表学术论文近 100 篇，主要论文被同行引用达 700 多篇次。

李金海，男，1984 年生，博士，《美国数学评论(MR)》评论员，《Journal of Computer Engineering and Informatics》编委，主要研究方向为粗糙集、概念格与粒计算。现主持国家自然科学青年基金 1 项，近几年，发表学术论文 10 余篇，其中被 SCI 检索 5 篇，EI 检索 2 篇。

第 3 届智能机械国际会议

3rd International Conference on Intelligent Machines (ICIM)

The 3rd ICIM'2014 aims to offer a meeting opportunity for researchers, students, and industry-related researchers from around the world to present recent results and discuss the latest advances and new trends in the theory and application of Intelligent Machines with both efficient and flexible behaviour.

he topics of interest include, but are not limited to:

computational intelligence;

intelligent hybrid systems;

documents analysis and recognition;

intelligent multimedia data analysis and processing;

cognition and robotic systems;

Submissions should be up to 6 pages, in double column IEEE proceeding format. Submitted papers will be double-blind reviewed by the Program Committee on the basis of novelty, relevance, and technical quality.

Conference content will be submitted for inclusion into IEEE Xplore as well as other Abstracting and Indexing (A&I) databases.

Website: http://icim.regim.org/