doi: 10.3969/j.issn.1674-8425(z).2013.01.021

序信息系统中变精度粗糙集属性约简的 Matlab 实现

徐伟华[1,2],张先韬[1],王巧荣[1]

(1. 重庆理工大学 数学与统计学院,重庆 400054; 2. 西安交通大学 管理学院,西安 710049)

摘 要:根据序信息系统中变精度粗糙集的理论知识,应用 Matlab 程序实现序信息系统中变精度粗糙集属性约简的计算。当序信息系统变精度粗糙集考虑单个优势决策时,约简类型不受系统协调性的影响,因此进行了考虑单个优势决策的约简计算研究,并通过算例验证了程序的正确性和可行性。

关 键 词:序信息系统;变精度;属性约简;单个优势决策

中图分类号:TP18 **文献标识码**:A **文章编号**:1674-8425(2013)01-0107-09

Experimental Computing on Attribute Reduction by Matlab in Dominance-Based Variable Precision Rough Set

XU Wei-hua[1,2], ZHANG Xian-tao[1], WANG Qiao-rong[1]

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

Abstract: According to the theory on Dominance-based variable precision rough set, we study and program Matlab codes to compute the attribute reductions and realize the variable precision rough set in ordered information systems. While a single dominance decision is considered in an ordered information system, the forms of reductions are not affected by the consistence of the system. And the attribute reductions with respect to a single decision are considered in this paper. Furthermore, cases are employed to illustrate the corrective of the program.

Key words: ordered information system; variable precision; attribute reduction; single dominance decision

粗糙集理论作为一种有效的软计算工具,在多个领域和学科中得到了卓有成效的应用,为工程应用和科学研究创造了很大的价值[1-15]。经典粗糙集理论最初由波兰数学家 Pawlak[7] 于 1982 年提出,是一种基于等价关系的处理不确定性的数学工具。粗糙集理论处理的主要是带有不确定

收稿日期:2012-08-21

作者简介:徐伟华(1979—),男,山西大同人,博士,副教授,主要从事粗糙集、模糊集、人工智能的数学基础等方面研究。

性的分类问题，能利用该理论的知识近似刻画不确定的概念，从而发掘有用信息，简化知识表达，为规则提取提供简洁的数据和知识。

经典粗糙集理论对关系要求苛刻，仅限于等价关系，而且对于概念的近似也仅限于支持和可能支持，这使得经典粗糙集的可拓性受到限制[16]。为使粗糙集理论应用更为广泛，学者们从多个角度对其进行了拓展研究[4]，提出和研究了变精度粗糙集[5-6]、优势关系粗糙集[1-3,11-13]、相容关系粗糙集、程度粗糙集、模糊粗糙集、概率粗糙集等[16]，适应了多种分类问题。

优势关系粗糙集和变精度粗糙集是2类非常重要的粗糙集模型，在实际应用中能发现重要的有用的信息，有很大的实际应用价值[16]。变精度粗糙集是一类带有一定容错能力的粗糙集模型，能克服经典粗糙集模型的局限性，发现较大程度上正确的信息，可实现多数支持，忽略正确程度较小的信息，从而实现信息筛选[5-6,16]。本文在优势关系变精度粗糙集(DB-VPRS)及属性约简理论的基础上进行了 Matlab 实现。首先介绍序信息系统中的变精度粗糙集及属性约简理论[14]，然后通过 Matlab 设计代码实现优势关系变精度粗糙集及约简的获取。在 Matlab 平台上进行程序编制，调试正确并成功完成序信息系统中变精度粗糙集及其属性约简的计算，为序信息系统知识发现的理论和应用研究提供参考。

1 序信息系统中变精度粗糙集理论

粗糙集理论用信息系统来表示对象和对象具有的信息，一个信息系统是一张数据表，包含对象、属性和对象在属性下的取值。

定义 1[16] 称一个四元组 $I=(U,A,V,f)$ 为信息系统，其中：$U=\{x_1,x_2,\cdots,x_n\}$ 为论域，x_i 为对象；$A=\{a_1,a_2,\cdots,a_m\}$ 为属性集，a_i 为对象的属性；$V=\bigcup_{a\in A}V_a$ 为属性值域，V_a 为属性 $a\in A$ 的值域；$f:U\times A\rightarrow V$ 为信息函数，$\forall x\in U, f(x,a)\in V_a$。

若属性集 $A=AT\cup DT$，且 $AT\cap DT=\varnothing$，AT 为条件属性集，DT 为决策属性集，则称信息系统 $I=(U,AT\cup DT,V,f)$ 为目标信息系统或决策表。

在序信息系统中，由属性或属性集可以对应得到优势关系。信息系统中优势关系定义见定义2。

定义 2[3,14,16] 设 $I=(U,AT\cup DT,V,f)$ 为目标信息系统，该信息系统上的2个二元关系定义如下：

$$R_{AT}^{\leqslant}=\{(x,y)\mid f(x,a)\leqslant f(y,a),\ \forall a\in AT\}$$

$$R_{DT}^{\leqslant}=\{(x,y)\mid f(x,a)\leqslant f(y,a),\ \forall a\in DT\}$$

$R_{AT}^{\leqslant}$ 和 $R_{DT}^{\leqslant}$ 称为目标信息系统 I 上的优势关系。

定义 3[3,14,16] 设 $I=(U,AT\cup DT,V,f)$ 为目标信息系统，在 I 中根据优势关系对对象分类得到如下形式的对象类：

$$[x]_{AT}^{\leqslant}=\{y\in U\mid f(x,a)\leqslant f(y,a),\ \forall a\in AT\}$$

$$[x]_{DT}^{\leqslant}=\{y\in U\mid f(x,a)\leqslant f(y,a),\ \forall a\in DT\}$$

$[x]_{AT}^{\leqslant}$ 和 $[x]_{AT}^{\leqslant}$ 分别称为对象 x 关于 AT 和 DT 的优势类。在 I 中根据优势关系对对象分类还可得到如下形式的对象类：

$$[x]_{AT}^{\geqslant}=\{y\in U\mid f(y,a)\leqslant f(x,a),\ \forall a\in AT\}$$

$$[x]_{DT}^{\geqslant}=\{y\in U\mid f(y,a)\leqslant f(x,a),\ \forall a\in DT\}$$

$[x]_{AT}^{\geqslant}$ 和 $[x]_{AT}^{\geqslant}$ 分别称为对象 x 关于 AT 和 DT 的劣势类。依优势关系分类的信息系统称为序信息系统。

变精度粗糙集最初多数包含关系定义，经过研究和统一，现在主要使用包含度来定义，但其实质是一样的，因此本文不再给出包含度的公理化定义。在普通集上，包含度应该特殊化。康托集的包含度定义应该加强，以适应康托集的性质。根据公理化定义，重新定义的康托集的包含度见定义4。

定义 4[14] 设 U 为论域，$P(U)$ 为 U 的幂集，D 称为 $P(U)$ 上的包含度，若 D 对任意的 $E,F,H\in P(U)$ 满足如下性质：① $0\leqslant D(E/F)\leqslant 1$；② $E\subseteq F\Rightarrow D(F/E)=1$；③ $E\subseteq F\subseteq H\Rightarrow D(E/H)\leqslant D(E/F)$；④ $E\subseteq F\Rightarrow D(E/H)\leqslant D(F/H)$；⑤ $D(F/E)+D(\sim F/E)=1$；⑥ $E\cap F=\varnothing\Rightarrow D(E/F)=D(F/E)=0$。

该定义使康托集上的包含度满足康托集性质，对于变精度粗糙集研究和应用也大有好处，能

更具体地进行研究。

定理 1[16] 设 U 为论域,$P(U)$ 为 U 的幂集,$\forall E,F\in P(U)$,则 D 是 $P(U)$ 上的包含度,

$$D(E/F)=\frac{|E\cap F|}{|F|}$$

根据上面的分类和包含度,序信息系统中的变精度粗糙集的近似集可以定义为:

定义 5[14] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统,$B\subseteq AT,\beta\in(0.5,1]$,任意 $X\subseteq U$ 关于 B 的 β-优势下近似和 β-优势上近似定义为:

$$\underline{B_\beta^\leqslant}(X)=\{x\in U|D(X/[x]_B^\leqslant)\geqslant\beta\}$$

$$\overline{B_\beta^\leqslant}(X)=\{x\in U|D(X/[x]_B^\leqslant)>1-\beta\}$$

若$\underline{B_\beta^\leqslant}(X)=\overline{B_\beta^\leqslant}(X)$,称 X 关于 $R_B^\leqslant$ 为β-可定义集;若$\underline{B_\beta^\leqslant}(X)\neq\overline{B_\beta^\leqslant}(X)$,称 X 关于 $R_B^\leqslant$ 为β-粗糙集。该粗糙集模型称为优势变精度粗糙集模型。

定义 6[14] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统,$B\subseteq AT,\beta\in(0.5,1]$,任意 $X\subseteq U$ 关于 B 的 β-劣势下近似和 β-劣势上近似也可定义为:

$$\underline{B_\beta^\geqslant}(X)=\{x\in U|D(X/[x]_B^\geqslant)\geqslant\beta\}$$

$$\overline{B_\beta^\geqslant}(X)=\{x\in U|D(X/[x]_B^\geqslant)>1-\beta\}$$

若$\underline{B_\beta^\geqslant}(X)=\overline{B_\beta^\geqslant}(X)$,称 X 关于 $R_B^\leqslant$ 为β-可定义集;若$\underline{B_\beta^\geqslant}(X)\neq\overline{B_\beta^\geqslant}(X)$,称 X 关于 $R_B^\leqslant$ 为β-粗糙集。该粗糙集模型称为劣势变精度粗糙集模型。

对于优势关系变精度粗糙集,概念的粗糙性可用类似 Pawlak 粗糙集的形式来度量,其形式为

$$\rho(X,B_\beta^\leqslant)=\frac{|\underline{B_\beta^\leqslant}(D_j)|}{|\overline{B_\beta^\leqslant}(D_j)|}$$

本文计算代码的编写按照上面定义分别实现,代码流程仅按照定义 5 的形式给出,下面介绍的内容仅以定义 5 形式给出,定义 6 形式的流程和本节剩余内容的表述在定义 6 形式下均类似可得,不再详述。

记 $U_{AT}^\leqslant=\{[x]_{AT}^\leqslant|x\in U\}$ 为优势关系 $R_{AT}^\leqslant$导出的全体分类,$U_{DT}^\leqslant=\{D_1,D_2,\cdots,D_r\}$ 为优势关系 $R_{DT}^\leqslant$导出的全体分类,本文不再一一表述。推广的序信息系统中变精度粗糙集模型依赖度的定义介于 0~1,表述如下:

定义 7[14] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统,$B\subseteq AT,\beta\in(0.5,1]$。$B$ 关于决策 DT 的 β-依赖度定义为:

$$\gamma_{B_\beta^\leqslant}(DT)=\frac{1}{r}\cdot\sum_{D_j\in U_{DT}^\leqslant}\frac{|\underline{B_\beta^\leqslant}(D_j)|}{|U|}$$

由 β-依赖度可定义属性子集 B 的 β-重要度:

$$\sigma(B_\beta^\leqslant)=\gamma_{AT_\beta^\leqslant}(DT)-\gamma_{(AT-B)_\beta^\leqslant}(DT)$$

若 $\sigma(B_\beta^\leqslant)=0$,则称 B 为 β-不重要的;若 $\sigma(B_\beta^\leqslant)>0$,则称 B 为 β-重要的。特别,当 $B=\{a\}$时,$\sigma(\{a\}_\beta^\leqslant)$表示属性 a 的重要度。

在序信息系统的变精度粗糙集模型中研究属性约简时,信息系统的协调性是必须考虑的,下面给出序信息系统协调性的定义。

定义 8[14,16] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统。若 $U_{AT}^\leqslant\leqslant U_{DT}^\leqslant$,即 $\forall x\in U\Rightarrow[x]_{DT}^\leqslant$,则称 I 为协调序信息系统;否则,称 I 为不协调序信息系统,即 $\exists x_0\in U$, s. t. $[x_0]_{AT}^\leqslant\not\subset[x_0]_{DT}^\leqslant$。

在协调信息系统中,变精度粗糙集就是经典粗糙集,但在优势关系下,优势关系变精度粗糙集却不是优势关系粗糙集,这是因为在优势关系下的分类可能不构成对论域的划分而只构成覆盖。在协调信息系统中,分析可知由上面依赖度定义的约简和由区分矩阵定义的约简是不同类型的约简,因此,序信息系统中使用变精度粗糙集。信息系统的协调性对于考虑多种类型的约简是没有影响的。根据 Pawlak 粗糙集,依赖度也可以称为下近似质量,对应的上近似形式称为上近似质量。在序信息系统中需要考虑多种类型的约简。下面给出近似分布约简和近似质量约简的定义。

定义 9[14] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统,$B\subseteq AT,\beta\in(0.5,1]$。$I$ 关于 B 的β-优势下近似分布函数和上近似分布函数分别定义为:

$$f(\underline{B_\beta^\leqslant})=(\underline{B_\beta^\leqslant}(D_1),\underline{B_\beta^\leqslant}(D_2),\cdots,\underline{B_\beta^\leqslant}(D_r))$$

$$f(\overline{B_\beta^\leqslant})=(\overline{B_\beta^\leqslant}(D_1),\overline{B_\beta^\leqslant}(D_2),\cdots,\overline{B_\beta^\leqslant}(D_r))$$

1) 若$f(\underline{B_\beta^\leqslant})=f(\underline{AT_\beta^\leqslant})$,则称 B 为β-优势保下近似分布集,记为 $\beta_{\leqslant,dis}-set$。若还满足 B 为 $\beta_{\leqslant,dis}-set$,但 B 的任意真子集不是 $\beta_{\leqslant,dis}-set$,则

称 B 为 I 的 β - 优势保下近似分布约简，记为 $\beta_{\leqslant,dis}-reduct$。

2）若 $f(\overline{B_\beta^{\leqslant}})=f(\overline{AT_\beta^{\leqslant}})$，则称 B 为 β - 优势保上近似分布集，记为 $\beta^{\leqslant,dis}-set$。若还满足 B 为 $\beta^{\leqslant,dis}-set$，但 B 的任意真子集不是 $\beta^{\leqslant,dis}-set$，则称 B 为 I 的 β - 优势保上近似分布约简，记为 $\beta^{\leqslant,dis}-reduct$。

3）若 B 为 $\beta_{\leqslant,dis}-reduct$ 且为 $\beta^{\leqslant,dis}-reduct$，则称 B 为 β - 优势保近似分布约简，记为 $\beta(\leqslant,dis)-reduct$。

定义 10[14] 设 $I=(U,AT\cup DT,V,f)$ 为序信息系统，$B\subseteq AT,\beta\in(0.5,1]$。$I$ 关于 B 的 β - 优势下近似质量函数和上近似质量函数分别定义为：

$$g(\underline{B}_\beta^{\leqslant})=\frac{1}{r}\cdot\frac{\sum_{k=1}^{r}|\underline{B}_\beta^{\leqslant}(D_k)|}{|U|}$$

$$g(\overline{B}_\beta^{\leqslant})=\frac{1}{r}\cdot\frac{\sum_{k=1}^{r}|\overline{B}_\beta^{\leqslant}(D_k)|}{|U|}$$

1）若 $g(\underline{B}_\beta^{\leqslant})=g(\underline{AT}_\beta^{\leqslant})$，则称 B 为 β - 优势保下近似质量集，记为 $\beta_{\leqslant,qua}-set$。若还满足 B 为 $\beta_{\leqslant,qua}-set$，但 B 的任意真子集不是 $\beta_{\leqslant,qua}-set$，则称 B 为 I 的 β - 优势保下近似质量约简，记为 $\beta_{\leqslant,qua}-reduct$。

2）若 $g(\overline{B}_\beta^{\leqslant})=g(\overline{AT_\beta^{\leqslant}})$，则称 B 为 β - 优势保上近似质量集，记为 $\beta^{\leqslant,qua}-set$。若还满足 B 为 $\beta^{\leqslant,qua}-set$，但 B 的任意真子集不是 $\beta^{\leqslant,qua}-set$，则称 B 为 I 的 β - 优势保上近似质量约简，记为 $\beta^{\leqslant,qua}-reduct$。

3）若 B 为 $\beta_{\leqslant,qua}-reduct$ 且为 $\beta^{\leqslant,qua}-reduct$，则称 B 为 β - 优势保近似质量约简，记为 $\beta(\leqslant,qua)-reduct$。

关于序信息系统中变精度粗糙集的属性约简还可以进行深入研究。随着对应变精度粗糙集和优势关系粗糙集的属性约简的研究，相关工作已深入展开，并在其他研究中体现，本文不作详细介绍和探讨。

2 优势关系变精度粗糙集 Matlab 实现

以本文介绍的内容为理论依据，按照定义设计算法流程，逐步实现各个计算过程，最终得到信息系统对应类型的约简。

按照约简定义，可以实现 β - 优势近似质量约简的计算。计算的目标是在 $\beta\in(0.5,1]$ 范围内，寻找对应类型的全部 β - 优势约简。根据约简定义，需要实现 β - 优势下、上近似质量约简和 β - 优势下、上近似分布约简的计算。设计的计算流程见图 1。

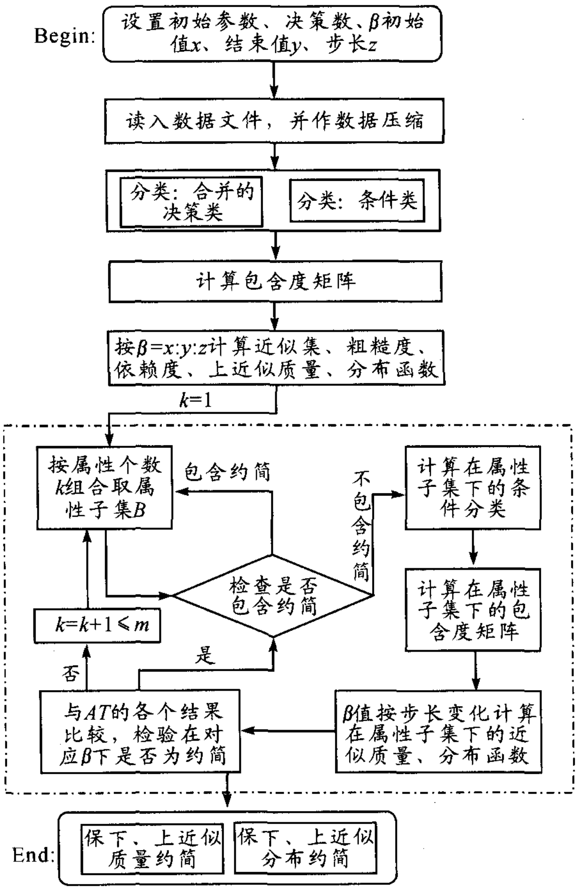

图 1 计算流程

根据图 1，按照本文介绍的定义编写 Matlab 代码。

设 $I=(U,AT\cup DT,V,f)$ 是一个压缩的信息系统，$U=\{x_1,x_2,\cdots,x_n\}$ 为论域，$AT=\{a_1,a_2,\cdots,a_m\}$ 为条件属性集，$DT=\{d_1,d_2,\cdots,d_p\}$ 为决策属性集，$U_{AT}^{\leqslant}=\{[x_i]_{AT}^{\leqslant}|i\leqslant n\}$ 为依条件属性的优势分类（dominating），$U_{DT}^{\leqslant}=\{D_1,D_2,\cdots,D_r\}$ 为合并的优势决策类，$U1_{DT}^{\leqslant}=\{[x_i]_{DT}^{\leqslant}|i\leqslant n\}$ 为未合并的优势决策类。按照本文介绍的近似集、依赖度、近似质

量和约简定义分别实现各个量的计算，最后输出了对应形式的约简。

1）初始化。设定计算的初始参数：精度β的起始值x、步长y、终止值z，指定数据存储文件（data.txt）以便读入数据，确定决策属性数目（dn）。

2）数据读取和压缩。采用文本文档（data.txt）存储数据，用 load 函数载入数据并赋给矩阵存储，通过逐行比较将相同行进行压缩。

3）编写各个功能函数。① 依条件分类：按优势类定义计算分类。② 依决策分类并压缩：类似地依决策属性获取决策类，并逐个比较删除重复决策类。③ 计算包含度：根据定理 1 计算各个决策类对条件分类的包含程度，并用矩阵进行记录，称为包含度矩阵。④ 计算上、下近似：根据定义 4，对包含度矩阵中的各个数值与β进行比较，若大于等于β对应记录为 1，否则记录为 0，得到的矩阵称为下近似特征矩阵。下近似特征矩阵中为 1 的位置标记了对应决策类下近似中的元素，对应的包含度矩阵中若各个数值大于$1-\beta$则记录为 1，否则记录为 0，得到的矩阵称为上近似特征矩阵。上近似特征矩阵中为 1 的位置标记了对应决策类上近似中的元素，根据上、下近似特征矩阵得到各个近似集。⑤ 计算粗糙度、近似质量、近似分布：根据计算得到的各个量，按照定义形式进行计算，各个计算结果存储为元胞形式，供下一步计算使用。⑥ 求约简：按照属性个数由少到多取组合，得到属性子集，在各个精度下检验是否为β-下（上）近似分布集、β-下（上）近似质量集，将检验结果通过矩阵记录，矩阵的行记录属性子集，矩阵的列记录精度，将记录矩阵各列比较，保留的极小子集即为约简。⑦ 输出约简：根据记录矩阵和各列对应的精度β的值，分别输出各个精度下的约简。

简要的过程设计代码如下，基本思想是对于每一个属性子集按照各个精度β分别检验并记录，最后将所有属性子集记录关于每一个β做比较，得出约简。

L_1 [Initialize] Set precision $\beta = x:y:z, x = 0.55, y = 0.05, z = 1$; Data file: data.txt; Decision number: dn = 1.

L_2 [Read data and compress them] OIS = load('data.txt') is the original information system. IS stands for the information system with objects being compressed;

L_3 [Classify] Classify by dominating conditions and dominating decisions in IS;

L_4 [Inclusion matrix (incmatrix)]

$j=1:r$ $i=1:n$ $A=[x_i]^{\leqslant}_{AT}$; $B=D_j$; $b\{i,j\}$ = intersect(A,B);

incmatrix(i,j) = length(b{i,j})/|U|);

% Calculate inclusion degree

L_5 [$(\beta = x:y:z)$: Approximations, approximation characteristic matrix, ρ, f, g in IS]

Calculate approximations *lapprD* (or *uapprD*) and approximation characteristic matrix *lapprmatrix* (or *uapprmatrix*) in IS;

beta = x:y:z [l,n] = size(incmatrix);

j = 1:n lapprD{j} = [ ]; uapprD{j} = [ ];

i = 1:l if dgincmatrix(i,j) > = beta

lappr = [i]; lapprmatrix(i,j) = 1;

lapprD{j} = union(lapprD{j}, lappr);

else lapprmatrix(i,j) = 0;

end

if incmatrix(i,j) > 1 - beta

uappr = [i]; uapprmatrix(i,j) = 1; uapprD{j} = union(uapprD{j}, uappr);

else uapprmatrix(i,j) = 0;

end

ρ can be calculated by approximations;

f can be represented by the matrix lapprmatrix and uapprmatrix;

g can be calculated by the sum of all elements in *lapprmatrix* and *uapprmatrix*.

L_6 [All reductions]

fl = 0; fu = 0; du = 0; dl = 0; ls = 0;

$k=1:$m subISs = nchooksek(1:1:m,k) % Select information system with k conditions ($B \subseteq AT$), results are arranged in a cell subISs;

[~ ,v] = size(subISs) ;

j = 1 : v ly = 0; ls = ls + 1; a = subISs{j} ;

Classify by conditions in a;

Calculate inclusion matrix sincmatrix in a (By step L_4);

β = x : y : z ly = ly + 1;

Calculate approximations, approximation matrix *slapprmatrix* and *suapprmatrix*, ρ, f in a (By step L_5);

Compare *slapprmatrix*, *suapprmatrix*, f in a with that in *IS* and record the subsets in corresponding matrices as follows;

if $f(\underline{B_\beta^{\leqslant}}) = f(\underline{AT_\beta^{\leqslant}})$

remarksetl{ls,ly} = B;

else remarksetl{ls,ly} = [ ]; end

if $f(\overline{B_\beta^{\leqslant}}) = f(\overline{AT_\beta^{\leqslant}})$

remarksetu{ls,ly} = B;

else remarksetu{ls,ly} = [ ]; end

if *slapprmatrix* = *lapprmatrix*

remarksetdl{ls,ly} = B;

else remarksetdl{ls,ly} = [ ]; end

if *suapprmatrix* = *uapprmatrix*

remarksetdu{ls,ly} = B;

else remarksetdu{ls,ly} = [ ]; end

Sort all the subsets in the corresponding recording matrix;

if remarkset * {i,j} ≠ ∅ && remarkset * {i, j}(remarkset * {k,j} && i≠k

remarkset * {k,j} = [ ];

end

L_7 [Output reductions]

[u,v] = size(remarkset *); beta = x - y;

j = 1 : vbeta = beta + y;

i = 1 : ured * = remarkset * {i,j} ;

if isempty(lred) = = 0

red *

end

fprintf('- - - - * reduction: beta = % g. ', beta);

where * in L_5, L_6 stands for the form of reduction (lower or upper approximation quality, lower or upper distribution).

3 实例计算

算例 1 首先使用有 6 个对象、3 个条件属性、1 个决策属性的协调序信息系统(表 1)来进行计算。

表 1 算例 1 的协调序信息系统

(U,A)	α_1	α_2	α_3	α_4
x_1	1	2	1	1
x_2	3	2	2	1
x_3	1	1	2	1
x_4	2	1	3	1
x_5	3	3	2	2
x_6	3	2	3	3

用 Matlab 程序计算上面信息系统的约简,得到在各个精度 β 下的对应约简(表 2)。

表 2 算例 1 约简情况

Forms of reduction	Reductions	β(Step length = 0.05)
$\beta_{\leqslant,qua}$ - *reduct*	$\{a_1,a_3\}$, $\{a_2,a_3\}$ $\{a_2\}$ $\{a_3\}$	[0.55,0.65] 0.8 [0.85,1]
$\beta^{\leqslant,qua}$ - *reduct*	$\{a_1\}$ $\{a_2\}$, $\{a_3\}$	0.7,0.75 [0.55,0.65]
$\beta \leqslant, qua$ - *reduct*	$\{a_2\}$ $\{a_3\}$	0.7,0.75,[0.85,1] [0.7,0.8]
$\beta_{\leqslant,dis}$ - *reduct*	$\{a_1,a_3\}$, $\{a_2,a_3\}$ $\{a_3\}$	[0.55,0.65] [0.85,1]
$\beta^{\leqslant,dis}$ - *reduct*	$\{a_2\}$, $\{a_3\}$	[0.55,0.65]
$\beta \leqslant, dis$ - *reduct*	$\{a_2\}$ $\{a_3\}$	[0.7,1] [0.7,0.8]

通过人工验证可知,表 2 约简计算正确。

下面给出一个较大规模信息系统计算约简,以验证程序的可行性。

算例 2 有 60 个对象、9 个条件属性、1 个决策属性的协调序信息系统见表 3。

表 3 算例 2 的协调序信息系统

(U,A)	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_1	27	8	3.34	11.4	81.5	42.6	11.7	59	922
x_2	23	11	3.14	11.0	78.8	50.7	14.4	57	998
x_3	29	10	3.21	9.8	81.6	39.4	12.4	54	962
x_4	45	7	3.41	11.1	77.5	50.2	20.6	56	982
x_5	35	8	3.44	9.6	84.6	43.7	14.3	55	1 071
x_6	45	8	3.45	10.2	66.8	43.1	25.5	54	1 030
x_7	30	11	3.23	12.1	83.9	49.2	11.3	56	935
x_8	30	9	3.29	10.6	86.0	40.4	10.5	56	900
x_9	24	9	3.31	10.5	83.2	42.5	12.6	61	1 002
x_{10}	27	10	3.36	10.7	79.3	41.0	13.2	59	912
x_{11}	42	8	3.39	9.6	69.2	41.3	24.2	56	1 018
x_{12}	26	9	3.2	10.9	83.4	44.9	10.7	58	1 025
x_{13}	34	9	3.21	10.2	77.0	45.7	15.1	57	970
x_{14}	28	9	3.29	11.1	86.3	44.6	11.4	60	986
x_{15}	31	8	3.26	11.9	78.4	49.6	14.9	58	959
x_{16}	46	7	3.22	11.8	79.9	51.2	16.1	54	860
x_{17}	30	8	3.35	11.4	81.9	44.0	12.0	58	936
x_{18}	30	8	3.15	12.2	84.2	53.1	12.7	38	872
x_{19}	27	7	3.44	10.8	87.0	43.5	13.6	59	959
x_{20}	24	7	3.53	10.8	79.5	33.8	12.4	61	941
x_{21}	45	7	3.22	11.4	80.7	48.1	18.5	53	892
x_{22}	24	9	3.37	10.9	82.8	45.2	12.3	61	871
x_{23}	40	6	3.45	10.4	71.8	41.4	19.5	53	971
x_{24}	27	9	3.25	11.5	87.1	51.6	9.5	56	887
x_{25}	55	6	3.35	11.4	79.7	46.9	17.9	59	953
x_{26}	29	9	3.23	11.4	78.6	46.6	13.2	60	969
x_{27}	31	9	3.10	12.0	78.3	48.6	13.9	55	920
x_{28}	32	10	3.38	9.5	79.2	43.7	12.0	54	844
x_{29}	53	9	2.99	12.1	90.6	48.9	12.3	47	862
x_{30}	35	8	3.37	9.9	77.4	42.6	17.7	57	989
x_{31}	42	7	3.49	10.4	72.5	43.3	26.4	59	1 006
x_{32}	67	10	2.98	11.5	88.6	47.3	22.4	60	861
x_{33}	20	9	3.26	11.1	85.4	44.0	9.4	64	929
x_{34}	12	9	3.28	12.1	83.1	51.9	9.8	58	858
x_{35}	40	8	3.32	10.1	70.3	46.1	24.1	56	961
x_{36}	30	10	3.16	11.3	83.2	45.3	12.2	58	923
x_{37}	54	7	3.36	9.7	72.8	45.5	24.2	62	1 113
x_{38}	33	10	3.03	10.7	83.5	48.7	12.4	58	995
x_{39}	32	9	3.32	10.5	87.5	45.3	13.2	54	1 015
x_{40}	29	10	3.32	10.6	77.6	45.5	13.8	56	991
x_{41}	38	11	2.99	12.0	81.5	50.3	13.5	73	894

续表

(U,A)	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_{42}	29	11	3.19	10.1	79.5	38.8	15.7	56	939
x_{43}	33	11	3.08	9.6	79.9	38.6	14.1	54	946
x_{44}	39	8	3.32	11.0	79.9	49.5	17.5	53	1 026
x_{45}	25	1	13.21	11.1	82.5	46.4	10.8	60	874
x_{46}	32	9	3.23	9.7	76.8	45.1	15.3	57	954
x_{47}	55	7	3.11	12.1	88.9	51.0	14.0	61	840
x_{48}	48	9	2.92	12.2	87.7	51.2	12.0	71	912
x_{49}	49	7	3.36	12.2	90.7	51.9	9.7	71	791
x_{50}	40	10	3.02	12.2	82.5	54.3	10.1	72	899
x_{51}	28	11	3.21	11.1	82.6	41.9	12.3	56	904
x_{52}	24	10	3.34	11.4	78.0	50.5	11.1	61	951
x_{53}	26	9	3.22	10.7	81.3	43.9	13.6	59	972
x_{54}	23	11	3.28	10.3	73.8	47.4	13.5	60	912
x_{55}	37	6	3.25	12.3	89.5	59.7	10.3	52	968
x_{56}	32	7	3.27	12.1	81.0	51.6	13.2	54	824
x_{57}	33	8	3.39	11.3	82.2	47.3	10.9	56	1004
x_{58}	24	12	3.25	11.1	79.8	44.8	14.0	56	896
x_{59}	33	10	3.22	9.0	76.2	42.2	14.5	54	912
x_{60}	28	9	3.48	10.7	79.8	37.5	13.0	58	954

观察上面信息系统,可以看到决策属性和条件属性的取值都是实数,且属性值较多,以等价类进行计算将得不到有效结果,可考虑使用优势关系进行分类。用 Matlab 程序计算表 3 的约简,得到在各个精度 β 下各形式的约简(表4)。

经分析和验证,上述几种约简定义和计算在单决策序信息系统中与协调性没有关系,优势关系变精度粗糙集中考虑上述约简不需要关注序信息系统的协调性。序信息系统情形与等价关系下的粗糙集是不一样的,序信息系统比较复杂,单个决策和多个决策可能差别很大,单决策下协调性对于上述几种约简没有影响,但是多决策下可能会产生影响。对于多优势决策同时考虑的情形比单个决策复杂得多,因此本文只取优势决策属性 $dn=1$,对于多个优势决策属性的情形不做算法的验证。决策数 $dn>1$ 的情形可用本文程序辅助继续研究。需要说明的是,笔者尝试以等价决策类来进行约简计算,并得到正确的结果。考虑等价决策时可以设置决策属性数量 $dn>1$ 进行约简计算,本文对这种情形不再列举实例验证结果。

表4 算例2约简情况

Forms of reduction	Reductions	β(Step length =0.05)
$\beta_{\leqslant,qua}-reduct$	$\{a_3,a_4,a_6,a_7,a_8\}$	0.55
	$\{a_1,a_2,a_3,a_5,a_6,a_7,a_8\}$	0.55,0.7
	$\{a_2,a_3,a_4,a_5,a_6,a_7a,8\}$	[0.6,1]
$\beta^{\leqslant,qua}-reduct$	$\{a_1,a_2,a_7,a_8\}$	0.55
	$\{a_1,a_3,a_4,a_5,a_8\}$	0.6
	$\{a_1,a_2,a_3,a_4,a_5,a_7,a_8\}$	[0.65,1]
	$\{a_1,a_2,a_6\}$	0.7
$\beta\leqslant,qua-reduct$	$\{a_1,a_2,a_3,a_5,a_6,a_7,a_8\}$	0.6,0.65,[0.75,1]
$\beta_{\leqslant,dis}-reduct$	$\{a_2,a_3,a_4,a_5,a_6,a_7,a_8\}$	[0.55,1]
$\beta^{\leqslant,dis}-reduct$	$\{a_1,a_2,a_3,a_4,a_5,a_7,a_8\}$	[0.55,1]
$\beta\leqslant,dis-reduct$	$\{a_1,a_2,a_3,a_5,a_6,a_7,a_8\}$	[0.55,1]

4 结束语

本文叙述了优势关系变精度粗糙集的基本理论，对序信息系统中基于变精度粗糙集方法的属性约简进行了 Matlab 程序实现。数据实例计算验证表明，程序正确无误。在序信息系统中引入变精度方法，使应用粗糙集理论处理信息系统中的知识和信息时更加合理，能有效地处理由噪声引起的数据不准确问题，这种容错性使序信息系统中的粗糙集理论和方法更加完善，增强了粗糙集理论在序信息系统中的可拓性。序信息系统中的变精度粗糙集就是优势关系变精度粗糙集。本文程序实现有助于促进序信息系统中粗糙集理论的发展，为相关理论、方法和应用的研究提供一种可行的计算工具支撑。

参考文献：

[1] Greco S, Matarazzo B, Słowinski R, et al. Variable consistency model of dominance-based rough sets approach. Rough Sets and Current Trends in Computing[M]. Berlin: Springer Berlin/Heidelberg, 2005: 170 - 181.

[2] Greco S, Matarazzo B, Slowinski R. Rough set theory for multicriteria decision analysis[J]. Europe Journal of Operation Research, 2001, 129: 11 - 47.

[3] Greco S, Matarazzo B, Slowinski R. Rough approximation by dominance relations[J]. International Journal of Intelligent Systems, 2002, 17(2): 153 - 171.

[4] Komorowski J, Polkowski L, Skowron A. Rough sets: A tutorial [M]. New York: Springer-Verlag New York, Inc. Secaucus, 1999.

[5] Masahiro I, Yukihiro Y, Yoshifum K. Variable-precision dominance-based rough set approach and attribute reduction[J]. International Journal of Approximate Reasoning, 2009, 50(8): 1199 - 1214.

[6] Mi J S, Wu W Z, Zhang W X. Approaches to knowledge reduction based on variable precision rough set model [J]. Information Sciences, 2004, 159(3/4): 255 - 272.

[7] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11: 341 - 356.

[8] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177: 3 - 27.

[9] Pawlak Z, Skowron A. Rough sets: some extensions[J]. Information Sciences, 2007, 177: 28 - 40.

[10] Pawlak Z, Skowron A. Rough sets and Boolean reasoning [J]. Information Sciences, 2007, 177(1): 41 - 73.

[11] Xu W H, Yang H Z, Zhang W X. Uncertainty measures of roughness of knowledge and rough sets in ordered information systems[J]. Lecture Notes in Computer Science, 2007, 4682: 759 - 769.

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

[13] Xu W H, Zhang X Y, Zhong J M, et al. Attribute reduction in ordered information systems based on evidence theory[J]. Knowledge and Information Systems, 2010 (29): 169 - 184.

[14] 张晓燕，徐伟华，张文修. 序目标信息系统中分布约简的矩阵算法[J]. 重庆理工大学学报：自然科学版，2010, 24(3): 56 - 61.

[15] Yoshifumi K, Masahiro I. A unified approach to reducts in dominance-based rough set approach[J]. Soft Computing-A Fusion of Foundations, Methodologies and Applications, 2010, 14(5): 507 - 515.

[16] Zhang W X, Leuang Y, Wu W Z. Information system and knowledge discovery[M]. Beijing: Science Press, 2003.

（责任编辑　刘　舸）