计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190500162

基于局部可调节多粒度粗糙集的属性约简

侯成军[1] 米据生[1] 梁美社[1,2]
1 河北师范大学数学与信息科学学院 石家庄 050024
2 石家庄职业技术学院科技发展与校企合作部 石家庄 050081
(13933134223@163.com)

摘 要 经典的多粒度粗糙集模型采用多个等价关系(多粒度结构)来逼近目标集。根据乐观和悲观策略,常见的多粒度粗糙集分为两种类型:乐观多粒度粗糙集和悲观多粒度粗糙集。然而,这两个模型缺乏实用性,一个过于严格,另一个过于宽松。此外,多粒度粗糙集模型由于在逼近一个概念时需要遍历所有的对象,因此非常耗时。为了弥补这一缺点,进而扩大多粒度粗糙集模型的使用范围,首先在不完备信息系统中引入了可调节多粒度粗糙集模型,随后定义了局部可调节多粒度粗糙集模型。其次,证明了局部可调节多粒度粗糙集和可调节多粒度粗糙集具有相同的上下近似。通过定义下近似协调集、下近似约简、下近似质量、下近似质量约简、内外重要度等概念,提出了一种基于局部可调节多粒度粗糙集的属性约简方法。在此基础上,构造了基于粒度重要性的属性约简的启发式算法。最后,通过实例说明了该方法的有效性。实验结果表明,局部可调节多粒度粗糙集模型能够准确处理不完备信息系统的数据,降低了算法的复杂度。

关键词:多粒度;粗糙集;不完备信息系统;近似质量;属性约简
中图法分类号 TP18

Attribute Reduction Based on Local Adjustable Multi-granulation Rough Set

HOU Cheng-jun[1], MI Ju-sheng[1] and LIANG Mei-she[1,2]
1 College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China
2 Department of Scientific Development and Technology and School-Business Cooperation, Shijiazhuang University of Applied Technology, Shijiazhuang 050081, China

Abstract In classical multi-granulation rough set models, multiple equivalent relations (multiple granular structures) are used to approximate a target set. According to optimistic and pessimistic strategies, there are two types of common multi-granulation called optimistic multi-granulation and pessimistic multi-granulation respectively. The two combination rules seem to lack of practicability since one is too restrictive and the other too relaxed. In addition, multi-granulation rough set model is highly time-consuming because it is necessary to scan all the objects when approximating a concept. To overcome this disadvantage and enlarge the using range of multi-granulation rough set model, this paper firstly introduced the adjustable multi-granulation rough set model in incomplete information system and defined the local adjustable multi-granulation rough set model. Secondly, this paper proved that local adjustable multi-granulation rough set and adjustable multi-granulation rough set have the same upper and lower approximations. By defining the concepts of lower approximation cosistent set, lower approximation reduction, lower approximation quality, lower approximation quality reduction, and importance of internal and external, a local adjustable multi-granulation rough set model for attribute reduction was proposed. Furthermore, a heuristic algorithm of attribute reduction was constructed based on granular significance. Finally, the effectiveness of the method was illustrated through examples. The experimental results show that local adjustable size rough set model can accurately process the data of incomplete information system, and it can reduce the complexity of the algorithm.

Keywords Multi-granulation, Rough set, Incomplete information system, Approximate quality, Attribute reduction

经典的粗糙集理论[1-2]是由 Pawlak 于 1982 年提出的,它已经成为了一种处理不确定性数据的有效工具。在使用粗糙集模型处理数据时,不需要数据的任何先验知识[3],因此该方法在很多方面得到了应用,如特征选择[4-5]、规则提取[6-7]、不

到稿日期:2019-05-30 返修日期:2019-07-21 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61573127);河北省自然科学基金(A2018210120);河北师范大学研究生创新项目基金(CXZZSS2017046)
This work was supported by the National Natural Science Foundation of China (61573127), Natural Science Foundation of Hebei province (A2018210120), Graduate Innovation Project Foundation of Hebei Normal University(CXZZSS2017046).
通信作者:米据生(mijsh@263.net)

确定性推理[8-9]、粒计算[10]、模式识别[11-12]、分类[13]和知识发现[14-16]等。

经典的粗糙集理论主要是利用等价关系来处理单个粒空间上的目标近似逼近理论。考虑到属性之间的关系可能是相互独立的，Qian 等[17]提出了多粒度粗糙集的概念，Xu 等[18]提出了可调节多粒度粗糙集模型。此后，多粒度粗糙集的相关研究受到了国内外学者的广泛关注。

目前，众多学者对粗糙集属性约简已经有了成熟、系统的探究，并取得了大量的研究成果，这是本文研究的理论基础。在分析已有属性约简算法时发现，基于全局的可调节多粒度粗糙集属性约简的计算复杂度较高，适应性较差，而基于局部的可调节多粒度粗糙集属性约简的研究相对较少。因此，本文在不完备信息系统中，利用可调节多粒度粗糙集模型进行局部属性约简，并定义了下近似协调集、下近似约简、下近似质量、下近似质量约简，以及内部重要度和外部重要度的概念，最后通过实例验证了该方法的有效性。

1 预备知识

本节主要回顾粗糙集、多粒度粗糙集的相关概念。

1.1 粗糙集

信息系统 $S=(U,A,\{V_a:a\in A\},\{f_a:a\in A\})$（简记 $S=(U,A)$），其中 U 为非空有限论域（对象集），A 为非空有限属性集合，V_a 为对象关于属性 a 的值域，$f_a:U\rightarrow V_a$ 是 $U\times A$ 上的一个信息函数。$X\subseteq U$，则 X 在 U 上的补集记为 $\sim X$：$\sim X=\{x\in U:x\notin X\}$。

当信息系统中某些属性的取值是缺省或者未知时，称该系统为不完备信息系统，仍然记为 $S=(U,A)$。用“$*$”表示缺省值或未知值，即 $f_a(x)=*$ 表示对象 x 在属性 a 上的取值是未知的。

设 $S=(U,A)$ 为一个不完备信息系统，对于非空子集 $B\subseteq A$，定义 $R_B=\{(x,y)\in U\times U:\forall a\in B$ 或 $f_a(x)=*$ 或 $f_a(y)=*\}$，则 R_B 为 U 关于属性集合 B 的相似关系（满足自反性和对称性）。称 $[x]_B=\{y\in U:(x,y)\in R_B\}$ 为对象 x 关于 R_B 的相似类。

定义 1[1-2] 设 $S=(U,A)$ 为一个不完备信息系统，$B\subseteq A$，$X\subseteq U$，X 关于 R_B 的下上近似分别为：

$$\underline{R_B}(X)=\{x\in U:[x]_B\subseteq X\}$$

$$\overline{R_B}(X)=\{x\in U:[x]_B\cap X\neq\emptyset\}$$

若 $\underline{R_B}(X)=\overline{R_B}(X)$，则称 X 是可定义的，否则称 X 是粗糙的。称序对 $(\underline{R_B}(X),\overline{R_B}(X))$ 为对象集合 X 关于 B 的粗糙集。

1.2 多粒度粗糙集

从粒计算的角度看，Pawlak 粗糙集是在单个粒度空间上对目标进行的近似刻画。现实中，考虑属性间可能具有独立性这一特征，Qian 等[17]提出了多粒度模型，下面简单介绍主要的多粒度粗糙集模型。

定义 2[17] 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,a_2,\cdots,a_m\}$，$[x]_{a_i}$ 是由 a_i 产生的相似类，$X\subseteq U$，X 的多粒度乐观下上近似分别为：$\underline{\sum_{i=1}^{m}R_i^O}(X)=\{x\in U:[x]_{a_1}\subseteq X\vee[x]_{a_2}\subseteq X\vee\cdots\vee[x]_{a_m}\subseteq X\}$，$\overline{\sum_{i=1}^{m}R_i^O}(X)=\sim\underline{\sum_{i=1}^{m}R_i^O}(\sim X)$。序对 $(\underline{\sum_{i=1}^{m}R_i^O}(X),\overline{\sum_{i=1}^{m}R_i^O}(X))$ 称为 X 的乐观多粒度粗糙集。

定义 3[3] 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,a_2,\cdots,a_m\}$，$[x]_{a_i}$ 是由 a_i 产生的相似类，$X\subseteq U$，X 的多粒度悲观下上近似分别为：$\underline{\sum_{i=1}^{m}R_i^P}(X)=\{x\in U:[x]_{a_1}\subseteq X\wedge[x]_{a_2}\subseteq X\wedge\cdots\wedge[x]_{a_m}\subseteq X\}$ $\overline{\sum_{i=1}^{m}R_i^P}(X):=\sim\underline{\sum_{i=1}^{m}R_i^P}(\sim X)$。序对 $(\underline{\sum_{i=1}^{m}R_i^P}(X),\overline{\sum_{i=1}^{m}R_i^P}(X))$ 称为 X 的悲观多粒度粗糙集。

针对经典粗糙集在处理分类问题时过于严格及其容错性不强等问题，许多专家根据实际需要提出了程度粗糙集和变精度粗糙集模型，并将其应用于多粒度模型中。利用类似方法可以得到悲观情况下多粒度粗糙集的相应定义。

程度多粒度粗糙集和变精度多粒度粗糙集在一定程度上放宽了边界域的严格定义，有利于不完备和噪声数据的处理。但是，与经典的多粒度粗糙集类似，上述多粒度模型中乐观多粒度下近似的定义过于宽松，即 m 个粒度中只要有一个粒度满足 k 程度或 β 精度即可；而悲观下近似的定义过于严格，即 m 个粒度都要满足 k 程度或 β 精度。基于此，Xu[18]给出了一种可调节多粒度粗糙集模型。

2 可调节多粒度粗糙集

实际生活中，我们会遇到类似投票的例子，即只要候选人得到的票数达到一定比例，该候选人即获胜。从粒的角度分析，每个投票者都是一个单独的粒。基于此，在 m 个粒结构下，只要存在一定数目的知识粒满足 $[x]_{a_i}\subseteq X$，则认为 x 属于 X 的下近似。下面给出可调节多粒度粗糙集（简记为 AMGRS）的定义及相关性质。

定义 4[18] 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,a_2,\cdots,a_m\}$，$[x]_{a_i}$ 是由 a_i 产生的相似类，$X\subseteq U$，其在 a_i 下的特征函数定义为：

$$S_X^i(x)=\begin{cases}1, & [x]_{a_i}\subseteq X(i\leqslant m)\\ 0, & 其他\end{cases}$$

定义 5 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,a_2,\cdots,a_m\}$，$[x]_{a_i}$ 是由 a_i 产生的相似类，$X\subseteq U$，令 $0<\frac{1}{m}\leqslant\alpha\leqslant 1$，$X$ 的 α 可调节多粒度粗糙集的下上近似可分别定义为：

$$\underline{\sum_{i=1}^{m}R_i^\alpha}(X)=\{x\in U:\sum_{i=1}^{m}S_X^i(x)/m\geqslant\alpha\}$$

$$\overline{\sum_{i=1}^{m}R_i^\alpha}(X)=\sim\underline{\sum_{i=1}^{m}R_i^\alpha}(\sim X)$$

定理 1 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,a_2,\cdots,a_m\}$，$[x]_{a_i}$ 是由 a_i 产生的相似类，$X\subseteq U$，令 $0<\frac{1}{m}\leqslant\alpha\leqslant 1$，$X$ 的 α 可调节多粒度粗糙集的上近似可表示为：$\overline{\sum_{i=1}^{m}R_i^\alpha}(X)=\{x\in U:\sum_{i=1}^{m}S_{\sim X}^i(x)/m<\alpha\}$。

证明过程详见文献[18]。

定理 2 设 $S=(U,A)$ 为一个不完备信息系统，$A=\{a_1,

$a_2,\cdots,a_m\}$,$[x]_{a_i}$ 是由 a_i 产生的相似类,$X\subseteq U$,则以下结论成立:

(1)若 $\alpha=\dfrac{1}{m}$,则 $\underline{\sum_{i=1}^{m}R_i^{1/m}}(X)=\underline{\sum_{i=1}^{m}R_i^{O}}(X)$,$\overline{\sum_{i=1}^{m}R_i^{O}}(X)=\underline{\sum_{i=1}^{m}R_i^{O}}(X)$;

(2)若 $\alpha=1$,则 $\underline{\sum_{i=1}^{m}R_i^{1}}(X)=\underline{\sum_{i=1}^{m}R_i^{P}}(X)$,$\overline{\sum_{i=1}^{m}R_i^{1}}(X)=\underline{\sum_{i=1}^{m}R_i^{P}}(X)$。

证明过程详见文献[18]。

同理,当 $\alpha=1$ 时,也可以得到:$\underline{\sum_{i=1}^{m}R_i^{1}}(X)=\underline{\sum_{i=1}^{m}R_i^{P}}(X)$,$\overline{\sum_{i=1}^{m}R_i^{1}}(X)=\underline{\sum_{i=1}^{m}R_i^{P}}(X)$。

由定理2可以看出,在不完备信息系统中,乐观多粒度粗糙集和悲观多粒度粗糙集均是可调节多粒度粗糙集的特殊形式。

定理3 设 $S=(U,A)$ 为一个不完备信息系统,$A=\{a_1,a_2,\cdots,a_m\}$,$[x]_{a_i}$ 是由 a_i 产生的相似类,$X\subseteq U$,令 $1/m\leqslant\alpha\leqslant 1$,则有:$\underline{\sum_{i=1}^{m}R_i^{P}}(X)\subseteq\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\subseteq\underline{\sum_{i=1}^{m}R_i^{O}}(X)\subseteq X$,$\overline{\sum_{i=1}^{m}R_i^{O}}(X)\subseteq\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\subseteq\overline{\sum_{i=1}^{m}R_i^{P}}(X)$。

证明:由定理2可直接证明。

仿照经典多粒度粗糙集的性质,下面给出可调节多粒度粗糙集的一些性质。

性质1 设 $S=(U,A)$ 为一个不完备信息系统,$A=\{a_1,a_2,\cdots,a_m\}$,$[x]_{a_i}$ 是由 a_i 产生的相似类,$X\subseteq U$,令 $0\leqslant\alpha\leqslant 1$,则有:

(1)$\underline{\sum_{i=1}^{m}R_i^{\alpha}}(U)=\overline{\sum_{i=1}^{m}R_i^{\alpha}}(U)=U$;

(2)$\underline{\sum_{i=1}^{m}R_i^{\alpha}}(\varnothing)=\overline{\sum_{i=1}^{m}R_i^{\alpha}}(\varnothing)=\varnothing$;

(3)若 $0\leqslant\alpha_1\leqslant\alpha_2\leqslant 1$,$\underline{\sum_{i=1}^{m}R_i^{\alpha_2}}(X)=\underline{\sum_{i=1}^{m}R_i^{\alpha_1}}(X)$,$\overline{\sum_{i=1}^{m}R_i^{\alpha_1}}(X)\subseteq\overline{\sum_{i=1}^{m}R_i^{\alpha_2}}(X)$;

(4)$\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X\cap Y)\subseteq\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\cap\underline{\sum_{i=1}^{m}R_i^{\alpha}}(Y)$;

(5)$\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X\cup Y)\supseteq\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\cup\underline{\sum_{i=1}^{m}R_i^{\alpha}}(Y)$;

(6)$\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X\cap Y)\subseteq\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\cap\overline{\sum_{i=1}^{m}R_i^{\alpha}}(Y)$;$\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X\cup Y)\supseteq\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X)\cup\overline{\sum_{i=1}^{m}R_i^{\alpha}}(Y)$。

证明:可根据定理1和定理2,仿照经典多粒度粗糙集模型的证明方法进行简单证明。

推论1 设 $S=(U,A)$ 为一个不完备信息系统,$A=\{a_1,a_2,\cdots,a_m\}$,$[x]_{a_i}$ 是由 a_i 产生的相似类,$X_1\subseteq X_2\subseteq U$,令 $1/m\leqslant\alpha\leqslant 1$,则有:$\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X_1)\subseteq\underline{\sum_{i=1}^{m}R_i^{\alpha}}(X_2)$;$\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X_1)\subseteq\overline{\sum_{i=1}^{m}R_i^{\alpha}}(X_2)$。

证明:由性质1可简单证明。

例1 设 $S=(U,A)$ 为一个不完备信息系统,$A=\{a_1,a_2,a_3,a_4\}$,如表1所列,令 $X=\{x_1,x_2,x_3,x_5,x_6,x_8\}$。

表1 不完备信息表

Table 1 Incomplete information table

U	a_1	a_2	a_3	a_4
x_1	1	1	3	1
x_2	2	1	3	1
x_3	2	*	4	2
x_4	3	3	3	3
x_5	3	4	4	3
x_6	4	3	4	*
x_7	5	4	4	4
x_8	1	2	3	2

根据定义4,计算对象关于 X 在 a_i 下的特征函数,如表2所列。

表2 基于 X 的特征函数

Table 2 Eigenfunction according to X

U	S_X^1	S_X^2	S_X^3	S_X^4
x_1	1	1	0	1
x_2	1	1	0	1
x_3	1	0	0	1
x_4	0	0	0	0
x_5	0	0	0	0
x_6	1	0	0	0
x_7	0	0	0	0
x_8	1	1	0	1

根据定义5,我们可计算 X 的可调节多粒度下近似:$\underline{\sum_{i=1}^{4}R_i^{0.5}}(X)=\{x_1,x_2,x_3,x_8\}$,$\underline{\sum_{i=1}^{4}R_i^{0.75}}(X)=\{x_1,x_2,x_8\}$。类似地,我们计算 X 的可调节多粒度上近似:$\overline{\sum_{i=1}^{4}R_i^{0.5}}(X)=\overline{\sum_{i=1}^{4}R_i^{0.75}}(X)=U$。

3 局部 AMGRS 的属性约简

3.1 局部 AMGRS 的下上近似

本节主要讨论在不完备决策信息系统中,获取局部可调节多粒度下近似属性约简的方法。利用类似的方法,我们也可以获取局部可调节多粒度上近似约简。首先给出局部AMGRS下上近似的定义。

在不完备信息系统 $S=(U,A)$ 中加入决策属性 d,$d\notin A$,则构成的新信息系统 $S=(U,A\cup\{d\})$,称为一个不完备决策信息系统。设 $[x]_d$ 是由 $\{d\}$ 产生的决策类,U 在决策属性 d 上的划分记为 $U/R_d=\{D_1,D_2,\cdots,D_K\}$。

定义6 设 $S=(U,A\cup\{d\})$ 为不完备决策信息系统,对于 $\forall D_k\in U/R_d$,令 $1/m\leqslant\alpha\leqslant 1$,$D_k$ 关于 R_d 的局部 AMGRS 下上近似分别为:

$$\underline{\sum_{i=1}^{m}LR_i^{\alpha}}(D_k)=\{x\in D_k:\sum_{i=1}^{m}S_X^i(x)/m\geqslant\alpha\}\quad\overline{\sum_{i=1}^{m}LR_i^{\alpha}}(D_k)=\sim\underline{\sum_{i=1}^{m}LR_i^{\alpha}}(\sim D_k)$$

定理4 设 $S=(U,A\cup\{d\})$ 为不完备决策信息系统,对于 $\forall D_k\in U/R_d$,令 $1/m\leqslant\alpha\leqslant 1$,则有:

(1)$\underline{\sum_{i=1}^{m}LR_i^{\alpha}}(D_k)=\underline{\sum_{i=1}^{m}R_i^{\alpha}}(D_k)$

(2)$\overline{\sum_{i=1}^{m}LR_i^{\alpha}}(D_k)=\overline{\sum_{i=1}^{m}R_i^{\alpha}}(D_k)$

证明：对于 $\forall x\in \underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)$，$\forall D_k\subset U/R$

$\Leftrightarrow x\in\{y\in D_k:\sum_{i=1}^{m}S_{D_k}^{i}(y)/m\geqslant\alpha\}$

$y\in D_k\subset U\Rightarrow x\in\{y\in U:\sum_{i=1}^{m}S_{D_k}^{i}(y)/m\geqslant\alpha\}$

$\Rightarrow x\in\underline{\sum_{i=1}^{m}}R_i^{\alpha}(D_k)$

因此，$\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)\subseteq\underline{\sum_{i=1}^{m}}R_i^{\alpha}(D_k)$。反之，对于 $\forall x\in\underline{\sum_{i=1}^{m}}R_i^{\alpha}(D_k)$，由定理 3 可知 $x\in D_k$，因此

$x\in\{y\in U:\sum_{i=1}^{m}S_{D_k}^{i}(y)/m\geqslant\alpha\}$

$\Rightarrow x\in\{y\in D_k:\sum_{i=1}^{m}S_{D_k}^{i}(y)/m\geqslant\alpha]$

$\Rightarrow x\in\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)$

故 $\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)\supseteq\underline{\sum_{i=1}^{m}}R_i^{\alpha}(D_k)$。

综上所述，$\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)=\underline{\sum_{i=1}^{m}}R_i^{\alpha}(D_k)$。

利用类似的方法也可以证明(2)。

定理 4 说明，局部 AMGRS 的下上近似与全局的 AMGRS 下上近似的结果是一样的。

3.2 局部 AMGDRS 的局部下近似约简

设 $S=(U,A\cup\{d\})$ 为一个不完备决策信息系统，$[x]_d$ 是由 $\{d\}$ 产生的决策类，U 在决策属性 d 上的划分记为 $U/R_d=\{D_1,D_2,\cdots,D_K\}$。

定理 5 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$B\subset A$，$\forall\alpha\in(0,1]$，$\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)\subseteq\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)$，$\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)=\left\{x\in D_k:\frac{\sum_{a_i\in B}S_{D_k}^{i}(D_k)}{m}\geqslant\alpha\right\}$。

定义 7 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$B\subset A$，如果 $\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)=\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)$，则称 B 是一个局部下近似协调集。更进一步地说，如果 B 是一个局部下近似协调集，并且没有任何一个集合 $C\subseteq B$ 是一个局部下近似协调集，则 B 是一个局部下近似约简。

定义 8 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$B\subseteq A$，局部下近似质量可定义为：$\sigma_{\alpha}(B,k)=\frac{|\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)|}{|D_k|}$。

定义 9 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$B\subseteq A$，$\sigma_{\alpha}(B,k)=\sigma_{\alpha}(A,k)$，则 B 是一个下近似质量协调集。更进一步地说，如果 B 是一个局部下近似质量协调集，并且没有任何一个真子集 $C\subset B$ 是一个局部下近似质量协调集，则 B 是一个局部下近似质量约简。

定理 6 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$B\subseteq A$，B 是一个局部下近似约简，当且仅当 B 是一个局部下近似质量约简。

证明：(必要性) 若 B 是一个局部下近似约简，则：$\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)=\sum_{a_i\in B}LR_i^{\alpha}(D_k)$

$$\sigma_{\alpha}(A,k)=\frac{|\underline{\sum_{i=1}^{m}}LR_i^{k}(D_k)|}{|D_k|}$$

$$\sigma_{\alpha}(B,k)=\frac{|\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)|}{|D_k|}$$

因此，$\sigma_{\alpha}(A,k)=\sigma_{\alpha}(B,k)$。

(充分性) 由于 $\sigma_{\alpha}(A,k)=\sigma_{\alpha}(B,k)$，因此 $\frac{|\underline{\sum_{i=1}^{m}}LR_i^{k}(D_k)|}{|D_k|}=\frac{|\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)|}{|D_k|}$，故 $|\underline{\sum_{i=1}^{m}}LR_i^{k}(D_k)|=|\underline{\sum_{a_i\in B}}LR_i^{\alpha}(D_k)|$。由定理 5 可知，$\sum_{a_i\in B}LR_i^{\alpha}(D_k)\subset\underline{\sum_{i=1}^{m}}LR_i^{k}(D_k)$，因此 $\underline{\sum_{i=1}^{m}}LR_i^{\alpha}(D_k)=\sum_{a_i\in B}LR_i^{\alpha}(D_k)$。

由于不同的粒度结构在 D_k 中的作用各不相同，下面定义重要度来探讨各个粒度的重要性。

定义 10 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$A=\{a_1,a_2,\cdots,a_m\}$，$B\subseteq A$，$\forall a_i\in A\backslash B$，则 a_i 关于 A 的内部重要度可定义为：

$Sig_{\text{in}}(A,a_i,k)=\sigma_{\alpha}(A,k)-\sigma_{\alpha}(A\backslash\{a_i\},k)$

定义 11 $S=\{U,A\cup\{d\}\}$，D_k 是第 k 个决策类，$A=\{a_1,a_2,\cdots,a_m\}$，$B\subset A$，$\forall a_i\in A\backslash B$，则 a_i 关于 B 的外部重要度可定义为：

$Sig_{\text{out}}(A,a_i,k)=\sigma_{\alpha}(A\cup\{a_i\},k)-\sigma_{\alpha}(A,k)$

若 $Sig_{\text{in}}(A,a_i,k)\neq0$，则说明第 i 个属性是必不可少的。所有必不可少的属性构成的集合为核心属性集合，记为 $Core(A)$。若 $Sig_{\text{out}}(A,a_i,k)\neq0$，说明第 i 个属性是相对必不可少的；若 $Sig_{\text{out}}(A,a_i,k)=0$，说明第 i 个属性是完全不必要属性。

下面给出一个启发式算法来计算局部下近似约简。

算法 1 局部可调节多粒度粗糙集属性约简

输入：多粒度信息系统 $S=\{U,A\cup\{d\}\}$，其中 $A=\{a_1,a_2,\cdots,a_m\}$，D_k 是第 k 个决策类，信息水平 α

输出：局部可调节多粒度粗糙集属性约简 red_k

Step1 初始化，令 $red_k=\varnothing$，$Core(A)=\varnothing$；

Step2 根据定义 9，计算 a_i 关于 A 的内部重要度：$sig_{in}(A,a_i)$；

Step3 若 $sig_{in}(A,a_i)\neq0$，则 $Core(A)=Core(A)\cup\{a_i\}$；

Step4 令 $red_k=Core(A)$，重新排列 $A\text{-}red_k=\{a_1',a_2',\cdots,a'_{|A\text{-}red_k|}\}$，其中 $sig_{out}(A,a_i')\geqslant sig_{out}(A,a'_{i+1})$；

Step5 令 $i=1:|A\text{-}red_k|$，如果 $\sigma_{\alpha}(red_k,k)\neq\sigma_{\alpha}(A,k)$，则 $red_k=red_k\cup\{a_i'\}$，$i\leftarrow i+1$；

Step6 $S=red_k-Core(A)$ $=\{b_1,b_2,\cdots,b_{|red_k-Core(A)|}\}$ $i=1:|red_k-Core(A)|$，若 $sig_{in}(red_k,b_i)=0$，则 $red_k=red_k-\{b_i\}$，$i\leftarrow i+1$；

Step7 输出 red_k。

算法 1 中，第 1 步为初始化；第 2－3 步根据定义 9 计算核心属性集合，即将所有内部重要度不为 0 的属性选出来组成核心粒度集 $Core(A)$，其计算复杂为 $O(|D_k||A||U|)$；第 4－5 步求解下近似质量协调集合，如果 $\sigma_{\alpha}(red_k,k)\neq\sigma_{\alpha}(A,k)$，则计算粒度 $a_i\in A\backslash red_k$ 关于 $Core(A)$ 的外部重要度，将各属性按照外部重要度从大到小排列，并顺次添加属性到核心属性集合中，直至满足 $\sigma_{\alpha}(red_k,k)=\sigma_{\alpha}(A,k)$，其计算复杂度为 $O(|D_k||A|^2|U|)$；第 6 步去掉冗余属性，由于核心属性是

不能删除的，因此只计算属性 $b_i \in red_k - Core(A)$ 关于 A 的内部重要性，若 $sig_{in}(A,b_i)=0$，则去掉 b_i，直到对于任意 $b_i \in red_k\text{-}Core(A)$ 均有 $sig_{in}(A,b_i)\neq 0$，其计算复杂度为 $O(|D_k||A||U|)$；第 7 步输出局部可调节多粒度粗糙集属性约简 red_k。综上，该算法的计算复杂度之和为 $O(|D_k||A|(1+|A|)|U|)$。

4 实例分析

设信息系统 $S=\{U,A\cup\{d\}\}$，如表 3 所列，$A=\{a_1,a_2,a_3,a_4,a_5\}$，$D_1=\{x_1,x_2,x_3,x_5,x_6,x_8\}$，$D_2=\{x_4,x_7\}$。

表 3 不完备信息表

Table 3 Incomplete information table

U	a_1	a_2	a_3	a_4	a_5	d
x_1	1	1	2	1	1	1
x_2	2	1	3	1	2	1
x_3	2	*	4	2	3	1
x_4	3	3	3	3	2	0
x_5	3	4	4	3	3	1
x_6	4	3	1	*	5	1
x_7	5	4	1	4	5	0
x_8	1	2	2	2	1	1

根据算法 1 进行模拟计算，取 $\alpha=0.5$，则 $\underline{\sum_{i=1}^{m} LR_i^{0.5}(D_1)}=\{x_1,x_2,x_3,x_8\}$，由 Step2 可计算 A 中各属性关于 A 的内部重要度：$\sigma_{0.5}(A,1)=4/6$，$\sigma_{0.5}(A\backslash\{a_1\},1)=3/6$。因此，$Sig_{\text{in}}(A,a_1,1)=\sigma_{0.5}(A,1)-\sigma_{0.5}(A\backslash\{a_1\},1)=1/6$。同理可得：$Sig_{\text{in}}(A,a_2,1)=1/6$，$Sig_{\text{in}}(A,a_3,1)=0$，$Sig_{\text{in}}(A,a_4,1)=1/6$，$Sig_{\text{in}}(A,a_5,1)=0$。由 Step3 得到核心属性集：$Core(A)=\{a_1,a_2,a_4\}$。由 Step4 计算 a_3 和 a_5 关于 red_1 的外部重要度：$Sig_{\text{out}}(red_1,a_3,1)=\sigma_{0.5}(red_1\cup\{a_3\},1)-\sigma_{0.5}(red_1,1)=1/6$，$Sig_{\text{out}}(red_1,a_5,1)=\sigma_{0.5}(red_1\cup\{a_5\},1)-\sigma_{0.5}(red_1,1)=1/6$。由于 $\sigma_{0.5}(A,1)\neq\sigma_{0.5}(red_1,1)$，因此由 Step5 和 Step6 把 a_3 加入到 red_1 中，此时 $\sigma_{0.5}(A,1)=\sigma_{0.5}(red_1,1)$，则输出约简结果为 $red_1=\{a_1,a_2,a_3,a_4\}$。

结束语 本文在不完备信息系统中利用相似关系建立可调节多粒度模型，定义下近似质量以及内、外属性重要度等概念，设置信息水平参数，给出了局部属性约简的启发式算法，并举例说明了该算法的优点。通过实例分析可知：局部属性约简在删除冗余属性方面具有明显优势，这对我们研究决策问题具有较为重要的作用。

参 考 文 献

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

[2] PAWLAK Z. Rough sets：some extensions [J]. Information Sciences，2007，177(1)：3-27.

[3] PAWLAK Z. Rough sets：Theoretical aspects of reasoning about data[M]. Kluwer Academic Publishers，Dordrecht，1991.

[4] QIAN Y H，LIANG J Y，PEDRYCZ W. Positive approximation：an accelerator for attribute reduction in rough set theory[J]. Artificial Intellegence. 2010，174 (9)：597-618.

[5] MA X A，WANG G Y，YU H，et al. Decision region distribution preservation reduction in decision-theoretic rough set model[J]. Information Sciences，2014，278：614-640.

[6] YAO Y Y，ZHAO Y. Attribute reduction in decision-theoretic rough set models[J]. Information Sciences，2008，178(17)：3356-3373.

[7] DU Y，HU Q H，ZHU P F，et al. Rule learning for classification based on neighborhood covering reduction [J]. Information Sciences，2011，181 (24)：5457-5467.

[8] LI J H，MEI C L，LV Y J. Incomplete decision contexts：Approximate concept construction，rule acquisition and knowledge reduction[J]. International Journal of Approximate Reasoning，2013，54(1)：149-165.

[9] POLKOWSKI L，SKOWRON A. Rough mereology：A new paradigm for approximate reasoning[J]. International Journal of Approximate Reasoning，1996，15(4)：333-365.

[10] SHE Y H，HE X L. On the structure of the multigranulation rough set model[J]. Knowledge-Based System，2012，36：81-92.

[11] QIAN Y H，LIANG J Y，WU W Z，et al. Information Granularity in Fuzzy Binary GrC Model[J]. IEEE Transactions on Fuzzy Systems，2011，19(2)：253-264.

[12] SWINIARSKI R，SKOWRON A. Rough set methods in feature selection and recognition[J]. Pattern Recognition Letters，2003，24(6)：833-849.

[13] WEI J M，WANG S Q，YUAN X J. Ensemble Rough Hypercuboid Approach for Classifying Cancers[J]. IEEE Transactions on Knowledge and Data Engineering，2010，22(3)：381-391.

[14] LIU Z G，YU H，WANG G Y. An automatic method to determine the number of clusters using decision-theoretic rough set [J]. International Journal of Approximate Reasoning，2014，55(1)：101-115.

[15] CHEN H M，LI T R，LUO C，et al. A decision-theoretic rough set approach for dynamic data mining[J]. IEEE Transactions on Fuzzy Systems，2015，23(6)：1-14.

[16] LIU D，LI T R，LIANG D C. Incorporating logistic regression to decision-theoretic rough sets for classifications[J]. International Journal of Approximate Reasoning，2014，55(1)：197-210.

[17] QIAN Y H，LIANG J Y，YAO Y Y，et al. A multigranulation rough set[J]. Information Sciences，2010，180：949-970.

[18] XU W H，SUN W X，ZHANG X Y，et al. Multiple granulation rough set approach to ordered information systems[J]. International Journal of General Systems，2012，41(5)：475-501.

HOU Cheng-jun，postgraduate. His research interests include rough set and granular computing.

MI Ju-sheng，Ph.D，professor. His research interests include granular computing and approximate reasoning.