基于下近似分布粒度熵的变精度悲观多粒度粗糙集粒度约简

孟慧丽 马媛媛 徐久成
(河南师范大学计算机与信息工程学院 新乡 453007)
(河南省高校计算智能与数据挖掘工程技术研究中心 新乡 453007)
(智慧商务与物联网技术河南省工程实验室 新乡 453007)

摘 要 将下近似分布约简引入变精度悲观多粒度粗糙集,定义了变精度悲观多粒度粗糙集的下近似分布粒度熵,基于下近似分布粒度熵定义了变精度悲观多粒度粗糙集粒度的重要度,并设计了基于下近似分布粒度熵的悲观多粒度粗糙集启发式粒度约简算法,通过实例验证了算法的有效性。

关键词 下近似分布约简,下近似分布粒度熵,变精度悲观多粒度粗糙集,粒度约简
中图法分类号 TP182 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.018

Granularity Reduction of Variable Precision Pessimistic Multi-granulation Rough Set Based on Granularity Entropy of Lower Approximate Distribution

MENG Hui-li MA Yuan-yuan XU Jiu-cheng
(College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, China)
(Engineering Technology Research Center for Computing Intelligence & Data Mining of Henan Province, Xinxiang 453007, China)
(Engineering Lab of Intelligence Business & Internet of Things of Henan Province, Xinxiang 453007, China)

Abstract The lower approximate distribution reduction was introduced into the variable precision pessimistic multi-granulation rough set. The granularity entropy of the lower approximate distribution of the variable precision pessimistic multi-granulation rough set was defined. The importance of a granularity was also defined based on the granularity entropy of the lower approximate distribution, and a heuristic granularity reduction algorithm of variable precision pessimistic multi-granulation rough set was presented. The experimental results show the validity of the algorithm.

Keywords Lower approximate distribution reduction, Granularity entropy of lower approximate distribution, Variable precision pessimistic multi-granulation rough set, Granularity reduction

粗糙集理论[1]是波兰数学家Pawlak提出的能有效处理信息系统中不精确、不确定信息的分析方法。粒计算[2]是人工智能领域中一个新的快速发展的领域,强调从不同层次和粒度分析问题,多粒度是粒计算的一个核心概念。从粒计算的角度来看,经典粗糙集理论主要基于等价关系对论域进行划分形成一个粒度空间,在单粒度空间下对目标概念进行近似逼近。钱宇华等人分析了单一粒度空间下粗糙集的不足,认为应当从多个粒度空间对目标概念进行近似逼近,针对实际的决策问题,可获得更加合理、更加满意的求解[4]。文献[3-5]提出了多粒度粗糙集的概念,定义了悲观多粒度粗糙集和乐观多粒度粗糙集,多粒度粗糙集已成为粗糙集理论研究的一个新的热点,引起了许多学者的关注。文献[6-12]分别对多粒度粗糙集的模型扩展及粒度约简算法进行了研究。

Ziarko于1993年在经典粗糙集的基础上提出了变精度粗糙集模型[13],文献[10,11]将变精度粗糙集的思想引入了多粒度粗糙集,定义了变精度多粒度粗糙集模型,但没有给出变精度多粒度粗糙集粒度约简的方法。文献[12]主要研究了基于容差关系的不完备系统中变精度多粒度粗糙集的约简方法。目前对完备决策系统下变精度多粒度粗糙集的粒度约简的研究较少,对变精度多粒度粗糙集中的粒度进行约简,针对不同的精度,删除粒度集中不必要的粒度,有利于进一步从变精度多粒度粗糙集中提取简洁的决策规则。

本文将下近似分布约简[14]的概念引入变精度悲观多粒度粗糙集,并定义了变精度悲观多粒度粗糙集的下近似分布粒度熵,基于下近似分布粒度熵定义了变精度悲观多粒度粗糙集粒度的重要度,设计了基于下近似分布粒度熵的变精度悲观多粒度粗糙集启发式粒度约简算法,该算法的时间复杂度为$O(m^2|U|^2)$,为变精度多粒度粗糙集的应用提供了理论基础。

1 基本概念

1.1 粗糙集的基本概念

定义1[1] 四元组$S=(U,AT,V,f)$称为信息系统,其中U表示对象的非空有限集合,称为论域;AT表示属性的非空有限集合;V_a表示属性a的值域,V表示全部对象在各个属

到稿日期:2015-05-11 返修日期:2015-06-29 本文受国家自然科学基金项目(60873104,61370169),河南省科技攻关重点项目(112102210194),河南省教育厅自然科学研究项目(2011A520054)资助。
孟慧丽(1978-),女,硕士,讲师,主要研究方向为粗糙集理论、数据挖掘,E-mail:menghuili93@163.com;马媛媛(1982-),女,硕士,主要研究方向为粗糙集理论、数据挖掘;徐久成(1963-),男,博士,教授,主要研究方向为数据挖掘、粒计算理论、生物信息等。

性上的取值构成的集合；f 表示 $U\times AT\to V$ 的一个信息函数，$\forall a\in AT, x\in U, f(x,a)\in V_a$。

定义 2[1] 设 $S=(U,AT,V,f)$ 为信息系统，$\forall A\subseteq AT$，定义属性集 A 的不可区分关系 $IND(A)$ 为：$IND(A)=\{(x,y)\in U\times U\mid \forall a\in A, f(x,a)=f(y,a)\}$，$U/IND(A)$ 表示不可区分关系 $IND(A)$ 在 U 上导出的划分，简记为 U/A。对 $\forall x\in U, [x]_A=\{y\mid f(y,a)=f(x,a), \forall a\in A\}$ 称为 x 在属性集 A 下的等价类。

定义 3[1] 设 $S=(U,AT,V,f)$ 为信息系统，$\forall A\subseteq AT$，$X\subseteq U$，X 关于属性集 A 的下近似集和上近似集分别定义为：$\underline{A}(X)=\{x\in U: [x]_A\subseteq X\}$，$\overline{A}(X)=\{x\in U: [x]_A\cap X\neq\varnothing\}$。

定义 4[10] 设 $S=(U,AT,V,f)$ 是一个信息系统，$X\subseteq U$，定义

$$P([x]_A\mid X)=\frac{|[x]_A\cap X|}{|[x]_A|}$$

其中，$|\cdot|$ 表示集合的基数，$P([x]_A\mid X)$ 表示集合 $[x]_A$ 中的元素包含在集合 X 中的比例。

性质 1[10] $0\leqslant P([x]_A\mid X)\leqslant 1$。

定义 5[10] 设 $S=(U,AT,V,f)$ 为信息系统，$\forall A\subseteq AT$，$X\subseteq U$，对 $\beta\in(0,1]$，定义 X 关于 A 的可变精度下近似：

$\underline{A_\beta}(X)=\{x\in U: P([x]_A\mid X)\geqslant\beta\}$

性质 2[10] 设 $S=(U,AT,V,f)$ 为信息系统，$\forall A\subseteq AT$，$X\subseteq U$，对 $\beta1,\beta2\in(0,1]$，$\beta1\leqslant\beta2$，则 $\underline{A_{\beta2}}(X)\subseteq\underline{A_{\beta1}}(X)$。

1.2 变精度悲观多粒度粗糙集的基本概念

在多粒度粗糙集中，四元组 $S=(U,AT,V,f)$ 是一个完备信息系统，其中 $A_1,A_2,\cdots,A_m\subseteq AT$，每个属性集 A_i 称为一个粒度，可以对 U 基于等价关系 $IND(A_i)$ 划分得到一个粒度空间，$A=\{A_1,A_2,\cdots,A_m\}$ 称为一个粒度集。

定义 6[10] 设 $S=(U,AT,V,f)$ 是一个信息系统，其中 $A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$\forall X\subseteq U$，$\beta\in(0,1]$，$X$ 的变精度悲观多粒度下近似、上近似分别定义为：

$$\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(X)=\{x\in U: P([x]_{A_1}\mid X)\geqslant\beta\wedge P([x]_{A_2}\mid X)\geqslant\beta\wedge\cdots\wedge P([x]_{A_m}\mid X)\geqslant\beta\}$$

$$\overline{\sum_{i=1}^{m}A_{i\beta}^{P}}(X)=\sim\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(\sim X)$$

性质 3[10] 设 $S=(U,AT,V,f)$ 是一个信息系统，其中 $A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$\forall X\subseteq U$，则 $\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(X)=\underline{\bigcap_{i=1}^{m}A_{i\beta}}(X)$。

2 基于下近似分布粒度熵的变精度悲观多粒度粗糙集粒度约简

定义 7 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$D$ 为决策属性，$A=\{A_1,A_2,\cdots,A_m\}$，$\beta\in(0,1]$，$B\subseteq A$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，则变精度悲观多粒度粗糙集下近似分布定义为：

$$\delta_A^{\beta}(U,D)=\{\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_1),\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_2),\cdots,\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_n)\}$$

若 $\delta_B^{\beta}(U,D)=\delta_A^{\beta}(U,D)$，则称 B 为 A 的 β 变精度悲观多粒度下近似分布一致集。

定义 8 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$\beta\in(0,1]$，$B\subseteq A$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，若 B 为 A 的 β 变精度悲观多粒度下近似分布一致集，且对 $\forall B1\subseteq B$，都有 $\delta_{B1}^{\beta}(U,D)\neq\delta_A^{\beta}(U,D)$，则称 B 为 A 的 β 变精度悲观多粒度下近似分布约简。

定义 9 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta\in(0,1]$，$\forall A_i\in A$，如果 $\delta_{A\{A_i\}}^{\beta}(U,D)=\delta_A^{\beta}(U,D)$，则称 A_i 在粒度集 A 中是不必要的，否则称 A_i 在粒度集 A 中是必要的。

定理 1 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta\in(0,1]$，则变精度悲观多粒度粗糙集下近似分布 $\delta_A^{\beta}(U,D)=\{\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_1),\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_2),\cdots,\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_n)\}$。

证明：因为 $\delta_A^{\beta}(U,D)=\{\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_1),\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_2),\cdots,\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_n)\}$，由性质 3 可知对 $\forall Y_j\in U/D$，有 $\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_j)=\underline{\bigcap_{i=1}^{m}A_{i\beta}}(Y_j)=\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_j)$，则 $\delta_A^{\beta}(U,D)=\{\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_1),\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_2),\cdots,\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_n)\}$。

定义 10 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$\beta\in(0,1]$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，定义

$$G_\beta(A\mid D)=\frac{1}{|U|^2}\sum_{j=1}^{n}|\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_j)|^2$$

为变精度悲观多粒度粗糙集中粒度集 A 的 β 下近似分布粒度熵。这里 $|\cdot|$ 表示集合的基数。

定理 2 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta\in(0,1]$，$B\subseteq A$，则 $G_\beta(A\mid D)\leqslant G_\beta(B\mid D)$。

证明：因为 $B\subseteq A$，对 $\forall A_i\in A$，$\forall Y_j\in U/D$，有 $\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_j)\subseteq\underline{\bigcap_{A_i\in B}A_{i\beta}}(Y_j)$，则 $|\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_j)|\leqslant|\underline{\bigcap_{A_i\in B}A_{i\beta}}(Y_j)|$，从而 $G_\beta(A\mid D)=\frac{1}{|U|^2}\sum_{j=1}^{n}|\underline{\bigcap_{A_i\in A}A_{i\beta}}(Y_j)|^2\leqslant\frac{1}{|U|^2}\sum_{j=1}^{n}|\underline{\bigcap_{A_i\in B}A_{i\beta}}(Y_j)|^2=G_\beta(B\mid D)$。

定理 2 说明随着粒度集中粒度的增加，粒度集的下近似分布粒度熵逐渐变小，粒度集对对象决策分类的影响程度越高；粒度熵越大，粒度集对对象决策分类的影响程度越小。

定理 3 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta1,\beta2\in(0,1]$，$\beta1\leqslant\beta2$，则 $G_{\beta2}(A\mid D)\leqslant G_{\beta1}(A\mid D)$。

证明：因为 $\beta1\leqslant\beta2$，所以对 $\forall A_i\in A$，$\forall Y_j\in U/D$，由性质 2 有 $\underline{A_{i\beta2}}(Y_j)\subseteq\underline{A_{i\beta1}}(Y_j)$，从而 $\underline{\bigcap_{A_i\in A}A_{i\beta2}}(Y_j)\subseteq\underline{\bigcap_{A_i\in A}A_{i\beta1}}(Y_j)$，则 $|\underline{\bigcap_{A_i\in A}A_{i\beta2}}(Y_j)|\leqslant|\underline{\bigcap_{A_i\in A}A_{i\beta1}}(Y_j)|$，从而有 $G_{\beta2}(A\mid D)=\frac{1}{|U|^2}\sum_{j=1}^{n}|\underline{\bigcap_{A_i\in A}A_{i\beta2}}(Y_j)|^2\leqslant\frac{1}{|U|^2}\sum_{j=1}^{n}|\underline{\bigcap_{A_i\in A}A_{i\beta1}}(Y_j)|^2=G_{\beta1}(A\mid D)$。

定理 4 设 $S=(U,AT\cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m\subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta\in(0,1]$，$B\subseteq A$，则 $\delta_B^{\beta}(U,D)=\delta_A^{\beta}(U,D)$ 的充要条件是 $G_\beta(B\mid D)=G_\beta(A\mid D)$。

证明：对 $\forall Y_j\in U/D$，由于 $\underline{\sum_{i=1}^{m}A_{i\beta}^{P}}(Y_j)=\underline{\sum_{A_i\in A}A_{i\beta}^{P}}(Y_j)$，则当 $\delta_B^{\beta}(U,D)=\delta_A^{\beta}(U,D)$ 时，$\underline{\sum_{A_i\in A}A_{i\beta}^{P}}(Y_j)=\underline{\sum_{A_i\in B}A_{i\beta}^{P}}(Y_j)$，由

性质 3，对 $\forall Y_j \in U/D$，有 $\underline{\sum_{A_i \in A} A_{i\beta}^P}(Y_j) = \underline{\bigcap_{i=1}^{m} A_{i\beta}}(Y_j) = \underline{\bigcap_{A_i \in A} A_{i\beta}}(Y_j)$，从而 $\underline{\bigcap_{A_i \in B} A_{i\beta}}(Y_j) = \underline{\bigcap_{A_i \in A} A_{i\beta}}(Y_j)$，则有 $G_\beta(B|D) = G_\beta(A|D)$。反之，当 $B \subseteq A$ 时，对 $\forall Y_j \in U/D$，有 $\underline{\bigcap_{A_i \in A} A_{i\beta}}(Y_j) \subseteq \underline{\bigcap_{A_i \in B} A_{i\beta}}(Y_j)$，所以当 $G_\beta(B|D) = G_\beta(A|D)$ 时，必有 $\underline{\bigcap_{A_i \in A} A_{i\beta}}(Y_j) = \underline{\bigcap_{A_i \in B} A_{i\beta}}(Y_j)$，即 $\underline{\sum_{A_i \in A} A_{i\beta}^P}(Y_j) = \underline{\sum_{A_i \in B} A_{i\beta}^P}(Y_j)$，从而 $\delta_B{}^\beta(U,D) = \delta_A{}^\beta(U,D)$。

定义 11 设 $S=(U,AT \cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m \subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$\beta \in (0,1]$，粒度 A_i 在粒度集 A 中的重要度定义为：

$SGF(A_i,A)=G_\beta(A-\{A_i\}|D)-G_\beta(A|D)$

性质 4 粒度 $A_i \in A$ 在粒度集 A 中是必要的，当且仅当 $SGF(A_i,A)>0$。

证明：当 $SGF(A_i,A)>0$ 时，即 $G_\beta(A-\{A_i\}|D)-G_\beta(A|D)>0$，$G_\beta(A-\{A_i\}|D) \neq G_\beta(A|D)$，从而由定理 4 可知 $\delta_{A-\{A_i\}}{}^\beta(U,D) \neq \delta_A{}^\beta(U,D)$，所以 $A_i \in A$ 在粒度集 A 中是必要的。反之，若 $A_i \in A$ 在粒度集 A 中是必要的，则必有 $\delta_{A-\{A_i\}}{}^\beta(U,D) \neq \delta_A{}^\beta(U,D)$，从而 $G_\beta(A-\{A_i\}|D) \neq G_\beta(A|D)$，由定理 2 可知 $G_\beta(A|D) \leqslant G_\beta(A-\{A_i\}|D)$，则 $G_\beta(A-\{A_i\}|D)-G_\beta(A|D)>0$，即 $SGF(A_i,A)>0$。

定义 12 粒度集 A 的核定义为 $Core(A)=\{A_i \in A | SGF(A_i,A)>0\}$。

定理 5 设 $S=(U,AT \cup D,V,f)$ 是一个完备决策信息系统，$A_1,A_2,\cdots,A_m \subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$，$U/D=\{Y_1,Y_2,\cdots,Y_n\}$，$B \subseteq A$，若 $G_\beta(B|D)=G_\beta(A|D)$ 且对 $\forall A_i \in B$，$SGF(A_i,B)>0$，则 B 为 A 的一个变精度悲观下近似分布粒度约简。

证明：由定理 4 可知当 $G_\beta(B|D)=G_\beta(A|D)$ 时，有 $\delta_B{}^\beta(U,D)=\delta_A{}^\beta(U,D)$，即 B 为 A 的变精度悲观多粒度下近似分布一致集，而对 $\forall A_i \in B$，$SGF(A_i,B)>0$，即 $G_\beta(B-\{A_i\}|D)-G_\beta(B|D)>0$，$G_\beta(B-\{A_i\}|D) \neq G_\beta(B|D)$，从而 $\delta_{B(A_i)}{}^\beta(U,D) \neq \delta_B{}^\beta(U,D)$，因此 B 为 A 的变精度悲观多粒度下近似分布粒度约简。

3 基于下近似分布粒度熵的变精度悲观多粒度粗糙集粒度约简算法

本节基于粒度集的下近似分布粒度熵设计了变精度悲观多粒度粗糙集的下近似分布粒度约简算法，将粒度集中不影响对象决策分类的不重要的粒度约简掉，基本思路是首先计算粒度集的核心粒度，并逐个选择粒度重要度最大的粒度加入到核中，最终得到粒度集的一个约简。

算法 1 基于下近似分布粒度熵的变精度悲观多粒度粗糙集粒度约简算法

输入：决策信息系统 $S=(U,AT \cup D,V,f)$，$A_1,A_2,\cdots,A_m \subseteq AT$，$A=\{A_1,A_2,\cdots,A_m\}$

输出：决策信息系统的一个变精度悲观下近似分布粒度约简 C

Step1：对每一个 $A_i \in A$，计算 U/A_i 和 U/D。

Step2：对每一个基于单粒度的粒度空间 U/A_i 和 $\forall Y_j \in U/D$，求出 U/A_i 中各等价类包含在 Y_j 中的比例，即 $P([x]_{A_i}|Y_j)$，针对给定的 β，将 $P([x]_{A_i}|Y_j) \geqslant \beta$ 的 x 加入到 $\underline{A_{i\beta}}(Y_j)$，对 $\forall A_i \in A$，$\forall Y_j \in U/D$，计算出 $\underline{A_{i\beta}}(Y_j)$。

Step3：令 $B=\varnothing$，对 $\forall A_i \in A$，计算 $SGF(A_i,A)$，将使 $SGF(A_i,A)>0$ 的 A_i 增加到粒度集 B。

Step4：如果 $G_\beta(B|D)=G_\beta(A|D)$，则 $C=B$，转 Step5；否则令 $B_1=A-B$，对 $\forall A_i \in B_1$，计算 $G_\beta(B|D)-G_\beta(B \cup \{A_i\}|D)$，将使下近似分布粒度熵减少最多的粒度增加到属性集 B 中，$B_1=B_1-\{A_i\}$，转 Step4。

Step5：输出粒度约简 C，算法结束。

算法时间复杂度分析：计算各个粒度对论域划分的时间复杂度为 $O(m|U|^2)$，计算一个粒度划分下的各等价类包含在决策分类 Y_i 中的比例的时间复杂度为 $O(|U|^2)$，则计算不同粒度划分下的等价类包含在决策分类 Y_i 中的比例的时间复杂度为 $O(m|U|^2)$，知道各对象在不同粒度下的等价类包含在决策分类 Y_i 中的比例后，计算 $\underline{A_{i\beta}}(Y_i)$ 的时间复杂度为 $O(|U|)$，则计算 $\underline{\bigcap_{A_i \in A} A_{i\beta}}(Y_i)$ 时，需计算 m 个 $\underline{A_{i\beta}}(Y_i)$ 的交集，其时间复杂度为 $O(m|U|^2)$，则计算 $SGF(A_i,A)$、$G_\beta(B|D)$、$G_\beta(A|D)$ 的时间复杂度为 $O(m|U|^2)$，从而对 m 个粒度，计算 $SGF(A_i,A)$、$G_\beta(A-\{A_i\}|D)$ 的时间复杂度为 $O(m^2|U|^2)$，所以算法总的时间复杂度为 $O(m^2|U|^2)$。

4 实例分析

表 1 是一个决策信息系统，$A=\{A_1,A_2,A_3\}$ 表示 3 种独立的评价指标，$U=\{x_1,x_2,x_3,x_4,x_5,x_6\}$ 表示 6 个待评价的对象，D 表示决策属性。$U/D=\{Y_1,Y_2\}$，$Y_1=\{x_2,x_3,x_5,x_6\}$，$Y_2=\{x_1,x_4\}$。根据每个评价指标，对对象的评价分为 3 类{1,2,3}，分别表示{差，一般，好}，如表 1 所列。

表 1 决策信息系统

U	A_1	A_2	A_3	D
x_1	1	1	2	no
x_2	2	2	1	yes
x_3	1	3	2	yes
x_4	2	1	1	no
x_5	3	2	2	yes
x_6	3	2	3	yes

对象在不同评价指标下的等价类如表 2 所列。

表 2 不同评价指标下对象 x_i 的等价类

U	$[x_i]_{A1}$	$[x_i]_{A2}$	$[x_i]_{A3}$
x_1	x_1,x_3	x_1,x_4	x_1,x_3,x_5
x_2	x_2,x_4	x_2,x_5,x_6	x_2,x_4
x_3	x_1,x_3	x_3	x_1,x_3,x_5
x_4	x_2,x_4	x_1,x_4	x_2,x_4
x_5	x_5,x_6	x_2,x_5,x_6	x_1,x_3,x_5

当 $U/D=\{Y_1,Y_2\}$，$Y_1=\{x_2,x_3,x_5,x_6\}$，$Y_2=\{x_1,x_4\}$ 时，各对象在不同评价指标下的等价类包含在 Y_i 中的比例如表 3 所列。

表 3 各对象在不同评价指标下的等价类包含在 Y_i 中的比例

U	$P([x_i]_{A1}\|Y_1)$	$P([x_i]_{A2}\|Y_1)$	$P([x_i]_{A3}\|Y_1)$	$P([x_i]_{A1}\|Y_2)$	$P([x_i]_{A2}\|Y_2)$	$P([x_i]_{A3}\|Y_2)$
x_1	1/2	0	2/3	1/2	1	1/3
x_2	1/2	1	1/2	1/2	0	1/2
x_3	1/2	1	2/3	1/2	0	1/3
x_4	1/2	0	1/2	1/2	1	1/2
x_5	1	1	2/3	0	0	1/3
x_6	1	1	1	0	0	0

当 $\beta=1/2$ 时，$\delta_A{}^\beta(U,D)=\{\{x_2,x_3,x_5,x_6\},\{x_4\}\}$，根据算法 1 计算可得 $\underline{A_{1\beta}}(Y_1)=U$，$\underline{A_{2\beta}}(Y_1)=\{x_2,x_3,x_5,x_6\}$，$\underline{A_{3\beta}}(Y_1)=U$，$\underline{A_{1\beta}}(Y_2)=\{x_1,x_2,x_3,x_4\}$，$\underline{A_{2\beta}}(Y_2)=\{x_1,x_4\}$，$\underline{A_{3\beta}}(Y_2)=\{x_2,x_4\}$，由于 $SGF(A_2,A)>0$，$SGF(A_3,A)>0$，因此 $B=\{A_2,A_3\}$，此时 $\delta_B{}^\beta(U,D)=\{\{x_2,x_3,x_5,x_6\},\{x_4\}\}=$

（下转第 104 页）

for discovering clusters in large spatial databases with noise [C]// Proceedings of the 2th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1996,226-231

[6] Sun Hao-jun, Wang Sheng-rui, Jiang Qing-shan. FCM-based model selection algorithms for determining the number of clusters[J]. Pattern Recognition,2004,37(10):2027-2037

[7] Bai Liang, Liang Ji-ye, Dang Chuang-yin. An initialization method to simultaneously find initial cluster centers and the number of clusters for clustering categorical data[J]. Knowledge-Based Systems,2011,24(6):785-795

[8] Liang Ji-ye, Zhao Xing-wang, Li De-yu, et al. Determining the number of clusters using information entropy for mixed data [J]. Pattern Recognition,2012,45:2251-2265

[9] Tou J, Gonzales R. Pattern Recognition Principles[M]. MA: Addison-Wesley. Reading,1974

[10] Pal N R, Bezdek J C. On clustering validity for the fuzzy c-means model[J]. IEEE Transactions on Fuzzy Systems,1995,3(3):370-379

[11] Xiao Yu, Yu Jian. Semi-Supervised Clustering Based on Affinity Propagation Algorithm[J]. Journal of Software,2008,19(11):2803-2813(in Chinese)
肖宇,于剑. 基于近邻传播算法的半监督聚类[J]. 软件学报,2008,19(11):2803-2813

[12] Bilenko M, Basu S, Mooney R J. Integrating constraints and metric learning in semi-supervised clustering [C]// Russ G, Dale S, eds. Proc. of the 21st Int'l Conf. on Machine Learning (ICML 2004). Banff: ACM Press,2004:81-88

[13] Basu S, Banerjee A, Mooney R J. Semi-supervised clustering by seeding[C]// Claude S, Achim GH, eds. Proc. of 19th Int'l Conf. on Machine Learning (ICML 2002). Sydney: Morgan Kaufmann Publishers,2002:27-34

[14] Kamvar S D, Klein D, Manning C D. Spectral learning[C]// Proc. of the 18th Int'l Joint Conf. on Artificial Intelligence (IJCAI 2003). Acapulco, Mexico: Morgan Kaufmann Publishers, 2003:561-566

(上接第 85 页)

$\delta_A^\beta(U,D)$,$G_\beta(B|D)=G_\beta(A|D)$,所以 $B=\{A_2,A_3\}$ 为 $\beta=1/2$ 时粒度集 A 的一个粒度约简;同理可计算出当 $\beta=1$ 时,$\delta_A^\beta(U,D)=\{\{x_6\},\emptyset\}$,当 $B=\{A_3\}$ 时,$\delta_B^\beta(U,D)=\{\{x_6\},\emptyset\}=\delta_A^\beta(U,D)$,$G_\beta(B|D)=G_\beta(A|D)$,所以 $B=\{A_3\}$ 为 $\beta=1$ 时粒度集 A 的一个粒度约简。

实例表明,在计算变精度悲观多粒度粗糙集的粒度约简时,针对给定的精度 β,首先计算粒度集的核心粒度,然后以下近似分布粒度熵的变化作为启发式信息,逐个选择粒度重要度最大的粒度加入到核中,最终求得的粒度约简与原始多粒度空间在相同精度下具有同样的决策能力。在本例中,当 $\beta=1/2$ 时,$B=\{A_2,A_3\}$ 为粒度集 A 的一个约简,评价指标 A_1 可忽略,当 $\beta=1$ 时,$B=\{A_3\}$ 为粒度集 A 的一个约简,评价指标 A_1、A_2 可忽略。

结束语 本文针对变精度悲观多粒度粗糙集模型的粒度约简进行了研究,设计了变精度悲观多粒度粗糙集粒度约简算法,并通过实例验证了该算法的有效性,针对不同的精度,计算粒度集的约简,有利于进一步从变精度悲观多粒度粗糙集中提取更加简洁的决策规则,这为变精度多粒度粗糙集的应用提供了理论基础。

参 考 文 献

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

[2] Lin T Y. Granular computing on binary relations II: Rough set representations and belief functions [M] // Rough Sets and Knowledge Discovery. 1998:122-140

[3] Qian Yu-hua, Liang Ji-ye. Rough set method based on multi-granulations[C]// Proceeding of the Fifth IEEE International Conference on Cognitive Informatics. Beijing, China, July 2006: 297-304

[4] Qian Yu-hua, Liang Ji-ye, Yao Yi-yu, et al. MGRS: A multigranulation rough set[J]. Information Sciences,2010,180:949-970

[5] Qian Yu-hua, Liang Ji-ye, Wei Wei. Pessimistic rough decision [C]// Second International Workshop on Rough Sets Theory. 2010:440-449

[6] Yang Xi-bei, Dou Hui-li, Yang Jing-yu. Hybird Multigranulation Rough Sets Based on Equivalence Relations[J]. Computer Science,2012,30(11):165-169(in Chinese)
杨习贝,窦慧莉,杨静宇. 基于等价关系的混合多粒度粗糙集[J]. 计算机科学,2012,30(11):165-169

[7] Zhang Ming, Tang Zhen-min, Xu Wei-yan, et al. Variable Multi-granulation Rough Set Model[J]. Pattern Recognition and Artificial Intelligence,2012,25(4):709-720(in Chinese)
张明,唐振民,徐维艳,等. 可变多粒度粗糙集模型[J]. 模式识别与人工智能,2012,25(4):709-720

[8] Sang Yan-li, Qian Yu-hua. A Granular Space Reduction Approach to Pessimistic Multi-Granulation Rough Sets[J]. Pattern Recognition and Artificial Intelligence,2012, 25(3):361-366(in Chinese)
桑妍丽,钱宇华. 一种悲观多粒度粗糙集中的粒度约简算法[J]. 模式识别与人工智能,2012,25(3):361-366

[9] Liu Cai-hui. Covering-based Multigranulation Rough Set Model Based on Maximal Description of Elements[J]. Computer Science,2013,40(12):64-67(in Chinese)
刘财辉. 一种元素最大描述下的多粒度覆盖粗糙集模型[J]. 计算机科学,2013,40(12):64-67

[10] Qian Yu-hua, Zhang Hu, Sang Yan-li, et al. Multigranulation decision-theoretic rough sets[J]. International Journal of Approximate Reasoning,2014,55(1):225-237

[11] Dou Hui-li, Wu Chen, Yang Xi-bei, et al. Variable Precision Multigranulation Rough Sets[J]. Journal of Jiangsu University of Science and Technology,2012,26(1):65-69(in Chinese)
窦慧莉,吴陈,杨习贝,等. 可变精度多粒度粗糙集模型[J]. 江苏科技大学学报,2012,26(1):65-69

[12] Zhai Yong-jian, Zhang Hong. Reduction of Variable Precision Multi-granulation Rough Sets[J]. Journal of Jinling Institute of Technology,2013,29(4):1-8(in Chinese)
翟永健,张宏. 变精度多粒度粗糙集的约简研究[J]. 金陵科技学院学报,2013,29(4):1-8

[13] Ziarko W. Variable precision rough set model [J]. Journal of Computer and System Sciences,1993,46(1):39-59

[14] Zhang Wen-xiu, Mi Ju-sheng, Wu Wei-zhi. Knowledge Reductions in Inconsistent Information Systems[J]. Chinese Journal of Computers,2003,26(1):12-18(in Chinese)
张文修,米据生,吴伟志. 不协调目标信息系统的知识约简[J]. 计算机学报,2003,26(1):12-18