第 31 卷 第 6 期
Vol. 31 No. 6

控 制 与 决 策
Control and Decision

2016 年 6 月
Jun. 2016

文章编号: 1001-0920 (2016) 06-1027-05 **DOI:** 10.13195/j.kzyjc.2015.0684

基于置信优势关系粗糙集的近似集动态更新方法

苟光磊[1,2,3], 王国胤[2]

(1. 西南交通大学 信息科学与技术学院，成都 610031；2. 中国科学院重庆绿色智能技术研究院 大数据挖掘及应用中心，重庆 400714；3. 重庆理工大学 计算机科学与工程学院，重庆 400054)

摘　要: 置信优势关系粗糙集是处理不完备有序信息的重要模型, 上、下近似集的计算是核心内容之一. 在实际应用中, 属性集通常会发生变化. 根据属性集的增加或减少, 首先讨论置信优势类及劣势类变化情况, 随之给出上、下近似集增量式的变化规律, 提出相应的近似集动态更新方法. 通过 Matlab 在 UCI 数据集上的实验结果表明, 与非增量式方法相比, 所提出的置信优势关系粗糙集下的上、下近似集的增量式更新方法可行、高效.

关键词: 置信优势关系；增量更新；近似集；粗糙集

中图分类号: TP391 **文献标志码:** A

Incremental updating approximations in confidential dominance relation based rough set

GOU Guang-lei[1,2,3], *WANG Guo-yin*[2]

(1. School of Information Science and Technology，Southwest Jiaotong University，Chengdu 610031，China；2. Big Data Mining and Applications Center，Chongqing Institute of Green and Intelligent Technology of Chinese Academy of Science，Chongqing 400714，China；3. School of Computer Science and Engineering，Chongqing University of Technology，Chongqing 400054，China．Correspondent：WANG Guo-yin，E-mail：Wanggy@cqupt.edu.cn)

Abstract: Confidential dominance relation based rough set is a model of incomplete ordered information processing, computation of approximations of which is a core issue. In real-life applications, the attribute set is dynamically changed. According to the variation of the attribute set, confidential dominance and dominated class are firstly calculated. Then the principles of incremental updating approximations are discussed when some attributes are added or deleted. Furthermore, incremental approaches and algorithms in the confidential dominance relation based on rough set are proposed. Finally, the experiments on UCI datasets developed on Matlab are designed to evaluate the performance of the proposed incremental updating method and non-incremental updating method. The results show that the proposed algorithms are effective and feasible under the variation of the attribute set.

Keywords: confidential dominance relation；incremental updating；approximations；rough sets

0 引　言

粗糙集理论[1]用于处理不精确、不确定问题, 将知识视为一种分类的能力, 概念用论域中的子集表示. 实际应用中, 由于数据采集技术的限制、传输故障及一些人为因素等原因造成的数据缺损和丢失现象时常发生, 人们常常面对信息不确定、不完全以及具有决策偏好信息的决策问题. 在这种背景下, 置信优势关系粗糙集[2]被提出, 用于处理不完备有序信息系统下的决策分析.

概念近似集的计算是粗糙集及其扩展理论的核心问题之一. 传统的概念近似集的计算方法大都是静态方法, 假定属性集和对象集都是固定不变的, 当信息系统发生变化后, 从头开始计算近似集, 时间复杂度高, 消耗大. 考虑到信息系统在现实环境下是动态变化的, 高效地计算出粗糙集概念上、下近似集的更新是热点研究问题之一[3]. Chan[4]提出了在对象集保持不变而发生单个属性的增加或删除时, 近似集的增量式更新方法及规则提取方法; 刘少辉等[5]提出了

收稿日期: 2015-05-31；**修回日期:** 2015-08-01.

基金项目: 国家科技重大项目(2014ZX07104-006)；中国科学院百人计划项目(Y21Z110A10)；国家自然科学基金项目(61073146, 61173184).

作者简介: 苟光磊(1980−), 男, 讲师, 博士生, 从事粗糙集、粒计算的研究；王国胤(1970−), 男, 教授, 博士生导师, 从事粗糙集、智能信息处理等研究.

一种新的计算正域的增量式更新方法; Zheng 等[6]提出了基于规则树的增量式知识获取 RRIA 算法, 能够快速地获取新知识; 王磊等[3]提出了对象集变化时近似集动态维护的矩阵方法. 针对不完备信息系统, Li 等[7]提出了特性关系粗糙集模型的概念上、下近似集的增量式更新方法; 杨习贝等[8]探讨了在容差关系下增减属性导致上、下近似集更新的方法; Li 等[9]提出了在属性集变化下, 优势关系粗糙集的上、下近似集的增量式更新方法; 石为人等[10]给出了一种基于优势区分矩阵的增量求核算法, 通过修改矩阵的某一行或某一列来增量得到决策表的核; Chen 等[11]给出了不完备有序信息系统中属性值粗化和细化的一种定义, 研究了属性值发生粗化和细化时概念近似集的增量式更新方法.

综上, 概念集上、下近似集的增量式更新方法取得了较多的研究成果, 但在不完备有序信息系统下, 属性集的增加或删除, 对于上、下近似集的变化和增量式更新方法, 还未有涉及. 本文以置信优势关系粗糙集为基础, 当属性集变化后, 考察置信优势类和置信劣势类的变化, 给出上、下近似集的更新情况, 提出置信优势关系粗糙集在属性变化下的动态更新方法. 仿真实验结果表明, 本文提出的动态更新方法能够有效求解上、下近似集, 并提高时间效率.

1 基础理论

定义 1[2] (不完备有序决策系统) 设有一个决策系统 $\mathrm{DS}=(U,A,V,f)$. 其中: U 是论域, 即非空的对象集合; A 是属性集合, $A=C\bigcup D$, C 和 D 分别表示条件属性集和决策属性集; V 是属性值域, 具有偏好; $f:U\times A\to V$ 是信息函数, $f=\{f(x_i,a)|f(x_i,a):x_i\to v_a, a\in C, x_i\in U, 1\leqslant i\leqslant |U|\}$, $f(x_i,a)=v_a$ 表示对象 x_i 在属性 a 上的取值. 如果所有的属性值都已知, 则称为完备有序决策系统; 如果存在缺失值, 则称为不完备有序决策系统 (IODS), 缺失值用 “$*$” 表示.

设 $a\in A$ 具有偏好有序关系, 记为 $\succeq_a$. $x\succeq_a y$, $x,y\in U$ 表示在属性 a 上对象 x 至少和 y 一样好. 增序偏好下, $x\succeq_a y$ 表示 $f(x,a)\geqslant f(y,a)$; 降序偏好下, 则表示 $f(x,a)\leqslant f(y,a)$. 本文为了便于讨论, 在不做特别说明的情况下, 将采用增序偏好.

定义 2[2] (置信优势关系) 设不完备有序决策系统 $\mathrm{IODS}=(U,A,V,f)$, $x,y\in U$, $P\subseteq C$, $B_P=\{b|b\in P\bigwedge f(x,b)\neq *\}$. 关于 P 的置信优势关系定义为

$$\begin{aligned}&\mathrm{CDR}(P)=\\&\{(x,y)\in U^2|(|B_P(x)\bigcap B_P(y)|/|B_P(x)|=1)\bigwedge\\&\forall_{q\in P}((f(x,q)=*\bigwedge f(y,q)=*)\bigvee\\&(f(x,q)=*\bigwedge f(y,q)\neq *))\bigvee(f(y,q)\succeq f(x,q))\},\end{aligned}$$

用 $y\,D_P^{\mathrm{CDR}}\,x$ 表示“y 置信优势于 x”.

定义 3[2] $\mathrm{Cl}_t^{\geqslant}$ 和 $\mathrm{Cl}_t^{\leqslant}$ 的上、下近似定义如下:

$$\begin{aligned}&\underline{P}(\mathrm{Cl}_t^{\geqslant})=\{x\in U|D_p^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}\},\\&\overline{P}(\mathrm{Cl}_t^{\geqslant})=\bigcup_{x\in \mathrm{Cl}_t^{\geqslant}} D_P^{\mathrm{CDR}+}(x)=\\&\qquad\{x|D_P^{\mathrm{CDR}+}(x)\bigcap \mathrm{Cl}_t^{\geqslant}\neq \varPhi\},\\&\underline{P}(\mathrm{Cl}_t^{\leqslant})=\{x\in U|D_P^{\mathrm{CDR}-}(x)\subseteq \mathrm{Cl}_t^{\leqslant}\},\\&\overline{P}(\mathrm{Cl}_t^{\leqslant})=\bigcup_{x\in \mathrm{Cl}_t^{\leqslant}} D_P^{\mathrm{CDR}-}(x)=\\&\qquad\{x|D_P^{\mathrm{CDR}-}(x)\bigcap \mathrm{Cl}_t^{\leqslant}\neq \varPhi\}.\end{aligned}$$

通过上、下近似可得边界域, 表示如下:

$$\mathrm{Bn}_p(\mathrm{Cl}_t^{\geqslant})=\overline{P}(\mathrm{Cl}_t^{\geqslant})-\underline{P}(\mathrm{Cl}_t^{\geqslant}),$$
$$\mathrm{Bn}_p(\mathrm{Cl}_t^{\leqslant})=\overline{P}(\mathrm{Cl}_t^{\leqslant})-\underline{P}(\mathrm{Cl}_t^{\leqslant}).$$

推论 1 对于 $\mathrm{IODS}=(U,A,V,f)$, $P\subseteq C$, 有:

1) $\underline{P}(\mathrm{Cl}_t^{\geqslant})=U-\overline{P}(\mathrm{Cl}_{t-1}^{\leqslant})$,
$\underline{P}(\mathrm{Cl}_t^{\leqslant})=U-\overline{P}(\mathrm{Cl}_{t-1}^{\geqslant})$;

2) $\mathrm{Bn}_P(\mathrm{Cl}_t^{\geqslant})=\mathrm{Bn}_P(\mathrm{Cl}_{t-1}^{\leqslant})$.

证明 1) 对于任意的 $x\in U$, 存在两种可能: $D_P^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}$; $D_P^{\mathrm{CDR}+}(x)\not\subseteq \mathrm{Cl}_t^{\geqslant}$. $D_P^{\mathrm{CDR}+}(x)\not\subseteq \mathrm{Cl}_t^{\geqslant}$ 意味着 $D_P^{\mathrm{CDR}+}(x)\bigcap(U-\mathrm{Cl}_t^{\geqslant})\neq\varnothing$. 因为 $U-\mathrm{Cl}_t^{\geqslant}=\mathrm{Cl}_{t-1}^{\leqslant}$, 有 $D_P^{\mathrm{CDR}+}(x)\bigcap \mathrm{Cl}_{t-1}^{\leqslant}\neq\varnothing$, 满足 $\overline{P}(\mathrm{Cl}_{t-1}^{\leqslant})$ 的定义. 因此 $\underline{P}(\mathrm{Cl}_t^{\geqslant})=U-\overline{P}(\mathrm{Cl}_{t-1}^{\leqslant})$ 得证.

同理可得 $\underline{P}(\mathrm{Cl}_t^{\leqslant})=U-\overline{P}(\mathrm{Cl}_{t-1}^{\geqslant})$.

2)
$$\begin{aligned}&\mathrm{Bn}_P(\mathrm{Cl}_t^{\geqslant})=\\&\overline{P}(\mathrm{Cl}_t^{\geqslant})-\underline{P}(\mathrm{Cl}_t^{\geqslant})=\\&(U-\underline{P}(\mathrm{Cl}_{t-1}^{\leqslant}))-(U-\overline{P}(\mathrm{Cl}_{t-1}^{\leqslant}))=\\&\mathrm{Bn}_P(\mathrm{Cl}_{t-1}^{\leqslant}).\end{aligned}$$
□

2 基于置信优势关系粗糙集的近似域更新方法

本节将讨论在属性集变化下, 置信优势关系下的近似域更新方法.

定义 4 设 $\mathrm{IODS}=(U,A,V,f)$, 对于任意的 $x\in U$, 有

$$l_P(x)=\min\{D_P^{\mathrm{CDR}+}(x)\bigcap \mathrm{Cl}_t\neq\varnothing\},$$
$$u_P(x)=\max\{D_P^{\mathrm{CDR}-}(x)\bigcap \mathrm{Cl}_t\neq\varnothing\}.$$

定理 1 设 $\mathrm{IODS}=(U,A,V,f)$, $P\subset C$, $Q\subseteq C-P$. 属性集 P 中增加 Q 时, 置信优势关系的变化满足:

1) $D_{P\cup Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)$;

2) $D_{P\cup Q}^{\mathrm{CDR}-}(x)=D_P^{\mathrm{CDR}-}(x)\bigcap D_Q^{\mathrm{CDR}-}(x)$.

证明 1) 假定 $y\in D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)$, 即 $y\in D_P^{\mathrm{CDR}+}(x)$ 且 $y\in D_Q^{\mathrm{CDR}+}(x)$. 这说明 (x,y) 在属性集 P 和属性 Q 上均满足置信优势关系, $(x,y)\subseteq$

$\mathrm{CDR}(P)$, 且 $(x,y)\subseteq \mathrm{CDR}(Q)$, 故 $y\in D_{P\cup Q}^{\mathrm{CDR}+}(x)$. 因此 $D_{P\cup Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)$ 成立.

2) 同理可证. □

定理 2　设 $P\subseteq C$, $Q\subseteq P$. 属性集 P 中去掉 Q 时, 置信优势关系的变化满足:

1) $D_{P-Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcup A_Q^{\geqslant}(x)$;

2) $D_{P-Q}^{\mathrm{CDR}-}(x)=D_P^{\mathrm{CDR}-}(x)\bigcup A_Q^{\leqslant}(x)$.

其中

$$A_Q^{\geqslant}(x)=$$
$$\{y\in U-D_P^{\mathrm{CDR}+}(x)|\sim D_P^{\mathrm{CDR}+}(x)\bigwedge D_{P-Q}^{\mathrm{CDR}+}(x)\},$$
$$A_Q^{\leqslant}(x)=$$
$$\{y\in U-D_P^{\mathrm{CDR}-}(x)|\sim D_P^{\mathrm{CDR}-}(x)\bigwedge D_{P-Q}^{\mathrm{CDR}-}(x)\}.$$

符号"∼"表示取非.

证明　1) 采用反证法. 若 $y\in U-D_P^{\mathrm{CDR}+}(x)$, 且 $y\notin A_Q^{\geqslant}(x)$, 即 $y\notin(\sim D_P^{\mathrm{CDR}+}(x)\bigwedge D_{P-Q}^{\mathrm{CDR}+}(x))$, 则 $y\in D_P^{\mathrm{CDR}+}(x)$, 或者 $y\in\sim D_{P-Q}^{\mathrm{CDR}+}(x)$. ① 当 $y\in D_P^{\mathrm{CDR}+}(x)$ 时, 因为 $Q\subseteq P$, 所以与 $y\in U-D_P^{\mathrm{CDR}+}(x)$ 矛盾. ② 当 $y\in\sim D_{P-Q}^{\mathrm{CDR}+}(x)$, 即 $(x,y)\subseteq\sim D_{P-Q}^{\mathrm{CDR}+}(x)$ 时, 显然等式 $D_{P-Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcup A_Q^{\geqslant}(x)$ 左边均有 $(x,y)\subseteq D_{P-Q}^{\mathrm{CDR}+}(x)$, 而等式右边存在 $(x,y)\subseteq\sim D_{P-Q}^{\mathrm{CDR}+}(x)$, 故等号不成立, 矛盾.

2) 同理可证. □

定理 3　属性集 P 中添加 Q 时, 上、下近似变化为:

1) $\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=\underline{P}(\mathrm{Cl}_t^{\geqslant})\bigcup T(\mathrm{Cl}_t^{\geqslant})$,

$$T(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x|l_{P\cup Q}(D_{P\cup Q}^{\mathrm{CDR}+}(x))\geqslant \mathrm{Cl}_t: x\in \mathrm{Bn}_P(\mathrm{Cl}_t^{\geqslant})\};$$

2) $\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=\overline{P}(\mathrm{Cl}_t^{\geqslant})-K(\mathrm{Cl}_t^{\geqslant})$,

$$K(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x|u_{P\cup Q}(D_{P\cup Q}^{\mathrm{CDR}+}(x))<\mathrm{Cl}_t: x\in \overline{P}(\mathrm{Cl}_t^{\geqslant})\}.$$

证明　1) 因为

$$\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=\{x\in U: D_{P\cup Q}^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}\},$$

又有

$$D_{P\cup Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x),$$

所以有

$$\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x\in U: D_P^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}\}\bigcup$$
$$\{x\in U-\underline{P}(\mathrm{Cl}_t^{\geqslant}): D_{P\cup Q}^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}\}.$$

已知

$$x\in U-\underline{P}(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x\in(\mathrm{Bn}_P(\mathrm{Cl}_t^{\geqslant})\bigcup(U-\overline{P}(\mathrm{Cl}_t^{\geqslant})))\}.$$

由于 $\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})\subseteq\overline{P}(\mathrm{Cl}_t^{\geqslant})$, 有 $U-\overline{P}(\mathrm{Cl}_t^{\geqslant})\subseteq U-\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})$. 若 $x\in U-\overline{P}(\mathrm{Cl}_t^{\geqslant})$, 则必有 $x\in U-\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})$, 那么 $x\notin\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})$. 设

$$T(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x|l_{P\cup Q}(D_{P\cup Q}^{\mathrm{CDR}+}(x))\geqslant \mathrm{Cl}_t: x\in \mathrm{Bn}_P(\mathrm{Cl}_t^{\geqslant})\},$$

则 1) 得证.

2) $\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=$

$$\bigcup\nolimits_{x\in \mathrm{Cl}_t^{\geqslant}} D_{P\cup Q}^{\mathrm{CDR}+}(x)=$$
$$\bigcup\nolimits_{x\in \mathrm{Cl}_t^{\geqslant}}(D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)),$$

而

$$D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)-(\sim D_{P\cup Q}^{\mathrm{CDR}+}(x)).$$

因此

$$\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=$$
$$\overline{P}(\mathrm{Cl}_t^{\geqslant})-\{u_{P\cup Q}(D_{P\cup Q}^{\mathrm{CDR}+}(x))<\mathrm{Cl}_t: x\in\overline{P}(\mathrm{Cl}_t^{\geqslant})\},$$

设

$$K(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x|u_{P\cup Q}(D_{P\cup Q}^{\mathrm{CDR}+}(x))<\mathrm{Cl}_t: x\in\overline{P}(\mathrm{Cl}_t^{\geqslant})\},$$

则 2) 得证. □

定理 4　属性集 P 中去掉 Q 时, 上、下近似变化为:

1) $\underline{P-Q}(\mathrm{Cl}_t^{\geqslant})=\underline{P}(\mathrm{Cl}_t^{\geqslant})-W(\mathrm{Cl}_t^{\geqslant})$,

$$W(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x|l_{P-Q}(A_Q^{\geqslant}(x))<\mathrm{Cl}_t, x\in\underline{P}(\mathrm{Cl}_t^{\geqslant})\};$$

2) $\overline{P-Q}(\mathrm{Cl}_t^{\geqslant})=\overline{P}(\mathrm{Cl}_t^{\geqslant})\bigcup R(\mathrm{Cl}_t^{\geqslant})$,

$$R(\mathrm{Cl}_t^{\geqslant})=\{A_Q^{\geqslant}(x): x\in\overline{P}(\mathrm{Cl}_t^{\geqslant})\}.$$

证明

$$\underline{P-Q}(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x\in U: D_{P-Q}^{\mathrm{CDR}+}(x)\subseteq \mathrm{Cl}_t^{\geqslant}\}=$$
$$\underline{P}(\mathrm{Cl}_t^{\geqslant})-\{x\in\underline{P}(\mathrm{Cl}_t^{\geqslant}): D_{P-Q}^{\mathrm{CDR}+}(x)\nsubseteq \mathrm{Cl}_t^{\geqslant}\},$$

设

$$W(\mathrm{Cl}_t^{\geqslant})=$$
$$\{x\in\underline{P}(\mathrm{Cl}_t^{\geqslant}): D_{P-Q}^{\mathrm{CDR}+}(x)\nsubseteq \mathrm{Cl}_t^{\geqslant}\}=$$
$$\{x|l_{P-Q}(A_Q^{\geqslant}(x))<\mathrm{Cl}_t, x\in\underline{P}(\mathrm{Cl}_t^{\geqslant})\},$$

则 1) 得证.

2) $\overline{P-Q}(\mathrm{Cl}_t^{\geqslant})=$

$$\bigcup\nolimits_{x\in \mathrm{Cl}_t^{\geqslant}} D_{P-Q}^{\mathrm{CDR}+}(x)=$$
$$\overline{P}(\mathrm{Cl}_t^{\geqslant})\bigcup\{\mathrm{A}_Q^{\geqslant}(x): x\in\overline{P}(\mathrm{Cl}_t^{\geqslant})\},$$

则 2) 得证. □

综上, 根据定理 1 和定理 3, 给出增加属性集合近似集动态更新方法 (见算法 1); 根据定理 2 和定理 4, 给出删除属性集合近似集动态更新方法 (见算法 2).

算法 1　增加属性集合动态更新近似集.

输 入: $D_P^{\mathrm{CDR}+}(x)$, $D_P^{\mathrm{CDR}-}(x)$, $D_Q^{\mathrm{CDR}+}(x)$, $D_Q^{\mathrm{CDR}-}(x)$, $\underline{P}(\mathrm{Cl}_t^{\geqslant})$, $\overline{P}(\mathrm{Cl}_t^{\geqslant})$, Q;

输出: $\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})$, $\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})$.

1. for $i=1:|U|$
2. $D_{P\cup Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcap D_Q^{\mathrm{CDR}+}(x)$
3. $D_{P\cup Q}^{\mathrm{CDR}-}(x)=D_P^{\mathrm{CDR}-}(x)\bigcap D_Q^{\mathrm{CDR}-}(x)$
4. end for
5. for $i=1:|n|$
6. 计算 $T(\mathrm{Cl}_t^{\geqslant})$ 和 $K(\mathrm{Cl}_t^{\geqslant})$
7. $\underline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=\underline{P}(\mathrm{Cl}_t^{\geqslant})\bigcup T(\mathrm{Cl}_t^{\geqslant})$
8. $\overline{P\bigcup Q}(\mathrm{Cl}_t^{\geqslant})=\overline{P}(\mathrm{Cl}_t^{\geqslant})-K(\mathrm{Cl}_t^{\geqslant})$
9. end for

算法 2 删除属性集合动态更新近似集.

输入: $\underline{P}(\mathrm{Cl}_t^{\geqslant}),\overline{P}(\mathrm{Cl}_t^{\geqslant}),Q$;

输出: $\underline{P-Q}(\mathrm{Cl}_t^{\geqslant}),\overline{P-Q}(\mathrm{Cl}_t^{\geqslant})$.

1. for $i=1:|U|$
2. 计算 $A_Q^{\geqslant}(x)$
3. $D_{P-Q}^{\mathrm{CDR}+}(x)=D_P^{\mathrm{CDR}+}(x)\bigcup A_Q^{\geqslant}(x)$
4. end for
5. for $i=1:|n|$
6. 计算 $W(\mathrm{Cl}_t^{\geqslant})$ 和 $R(\mathrm{Cl}_t^{\geqslant})$
7. $\underline{P-Q}(\mathrm{Cl}_t^{\geqslant})=\underline{P}(\mathrm{Cl}_t^{\geqslant})-W(\mathrm{Cl}_t^{\geqslant})$
8. $\overline{P-Q}(\mathrm{Cl}_t^{\geqslant})=\overline{P}(\mathrm{Cl}_t^{\geqslant})\bigcup R(\mathrm{Cl}_t^{\geqslant})$
9. end for

例 1 设不完备有序决策系统 $\mathrm{IODS}=(U,C\bigcup D)$, $C=\{a_1,a_2,a_3\}$, 见表 1.

表 1 不完备决策表

	a_1	a_2	a_3	D
x_1	4	4	3	4
x_2	3	2	3	3
x_3	4	*	2	3
x_4	2	2	2	2
x_5	2	1	2	2
x_6	3	1	2	1
x_7	*	2	2	2
x_8	4	1	2	2
x_9	3	*	2	3
x_{10}	4	3	3	3

假设 $P=\{a_1,a_3\}$, $Q=\{a_2\}$. 显然, $C=P\bigcup Q$, $P=C-Q$. 限于篇幅, 以 x_1, x_4 为例, 可计算得

$$A_Q^{\geqslant}(x_1)=\{x_{10}\},$$

$$A_Q^{\geqslant}(x_4)=\{x_3,x_5,x_6,x_8,x_9,x_{10}\}.$$

根据定理 2 有

$$D_P^{\mathrm{CDR}+}(x_1)=D_C^{\mathrm{CDR}+}(x_1)\bigcup A_Q^{\geqslant}(x_1)=\{x_1,x_{10}\},$$

$$D_P^{\mathrm{CDR}+}(x_4)=D_C^{\mathrm{CDR}+}(x_4)\bigcup A_Q^{\geqslant}(x_4)=U-\{x_7\}.$$

$P=C-Q$ 的下近似域变化, 计算得

$$W(\mathrm{Cl}_4^{\geqslant})=$$

$$\{x|l_{C-Q}(A_Q^{\geqslant}(x))<\mathrm{Cl}_4,x\in\underline{P}(\mathrm{Cl}_4^{\geqslant})\}=\{x_1\},$$

$$W(\mathrm{Cl}_3^{\geqslant})=\varnothing,$$

$$W(\mathrm{Cl}_2^{\geqslant})=\{x_4,x_7\},$$

$$W(\mathrm{Cl}_1^{\geqslant})=\varnothing.$$

根据定理 4 有

$$\underline{P}(\mathrm{Cl}_4^{\geqslant})=\underline{C-Q}(\mathrm{Cl}_4^{\geqslant})=\varnothing,$$

$$\underline{P}(\mathrm{Cl}_3^{\geqslant})=\{x_1,x_2,x_{10}\},$$

$$\underline{P}(\mathrm{Cl}_2^{\geqslant})=\{x_1,x_2,x_3,x_8,x_{10}\},$$

$$\underline{P}(\mathrm{Cl}_1^{\geqslant})=U.$$

$R(\mathrm{Cl}_t^{\geqslant})$、$T(\mathrm{Cl}_t^{\geqslant})$ 和 $K(\mathrm{Cl}_3^{\geqslant})$ 可同理计算, 这里不再赘述. 根据定理 3 和定理 4, 可获得相应的近似集.

3 仿真实验

本节通过在 UCI[12] 的 5 个数据集和本文例 1 (见表 2) 进行实验, 以验证上述方法的正确性和有效性. 实验环境为 Lenovo X220 i, 配置为 Intel i3-2370 2.4 GHz, 内存 4GB, Windows 7 (32bit). 使用 Matlab R2012b 编程实现.

表 2 数据集描述

ID	数据集名称	$\lvert U\rvert$	$\lvert C\rvert$
1	Hepatitis	155	19
2	Ionosphere	351	34
3	breast-cancer	699	10
4	mammographic-masses	961	5
5	Statlog (Landsat Satellite)	6 435	36
6	本文例 1	10	3

首先随机删除 1 个、$|C|/4$、$|C|/2$ 个条件属性, 通过增量式方法 (算法 2) 得到删除属性后的近似集; 然后重新添加这些被删除的属性, 通过算法 1 重新获得原来数据集的近似集. 通过求解近似集的运行时间对比增量式动态更新方法与非增量式方法的性能.

非增量式方法和算法 2 在随机删除 1 个、$|C|/4$ 个、$|C|/2$ 个条件属性情况下的运行时间对比见表 3.

表 3 算法 2 与非增量式方法的时间对比

ID	非增量式	删除 1 个	删除 $\lvert C\rvert/4$	删除 $\lvert C\rvert/2$
1	0.117	0.126	0.142	0.146
2	0.612	0.531	0.515	0.505
3	1.922	1.837	1.644	1.913
4	3.718	2.630	2.630	2.677
5	93.176	66.26	62.303	59.592
6	0.120 2	0.009 6	0.009 6	0.011 2

从表 3 可以看出: 在样本数量很小时, 算法 2 的时间效率并不一定会优于非增量式方法, 如数据集 Hepatitis; 随着样本数量的增加, 动态更新方法下获取置信优势类的时间效率明显优于非增量式方法; 随着删除属性个数的增加, 算法 2 动态更新近似集的运行时间则会有所降低, 在属性个数较多的数据集更显著, 如 Statlog (Landsat Satellite)、Ionosphere.

再将这些随机删除的属性添加回去, 考察算法 1 与非增量式求解近似集的性能比较, 实验结果见表 4.

表 4　算法 1 与非增量式方法的时间对比

ID	非增量式	增加 1 个	增加 $\lvert C\rvert/4$	增加 $\lvert C\rvert/2$
1	0.117	0.070	0.081	0.085
2	0.612	0.153	0.212	0.350
3	1.922	0.430	1.377	1.485
4	3.718	0.769	0.769	2.535
5	93.176	8.153	22.963	22.194
6	0.1202	0.0099	0.010	0.0103

从表 4 可以看出, 算法 1 增加属性后近似集的动态更新方法比非增量式方法时间效率高. 随着在原属性集合上增加的属性个数的增多, 如增加 1 个属性与增加 $|C|/4$ 个属性相比, 算法 1 的运行时间会有所增加.

综合上面的实验结果可知, 通常情况下, 与非增量式方法相比, 动态更新方法在属性集增加 (算法 1) 和属性集减少 (算法 2) 时求解近似集, 不需要重新计算整个论域, 时间效率更高. 随着增加属性个数的增加, 动态更新方法的运行时间也会随之增加, 而随着删除属性个数的增加, 运行时间则会有所减少.

4　结　论

在监测、医疗领域, 经常面对不完备有序决策, 以及属性集 (监/检测指标) 的改变问题. 本文探讨了不完备有序信息系统随着属性集的变化, 上、下近似集的变化规律, 提出了基于置信优势关系粗糙集的近似集动态更新方法, 与非增量式计算近似集方法相比, 时间效率更高. 进一步的研究工作将利用近似集的动态更新方法, 实现不完备信息系统增量式属性约简, 从而提高约简效率.

参考文献(References)

[1] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177(1): 3-27.

[2] 苟光磊, 王国胤, 利节, 等. 基于置信优势关系的粗糙集近似模型[J]. 控制与决策, 2014, 29(7): 1325-1329.
(Gou G L, Wang G Y, Li J, et al. Confidential dominance relation based rough approximation model[J]. Control and Decision, 2014, 29(7): 1325-1329.)

[3] 王磊, 李天瑞, 刘清, 等. 对象集变化时近似集动态维护的矩阵方法[J]. 计算机研究与发展, 2013, 50(9): 1992-2004.
(Wang L, Li T R, Liu Q, et al. A matrix-based approach for maintenance of approximation under the variation of object set[J]. J of Computer Research and Development, 2013, 50(9): 1992-2004.)

[4] Chan C C. A rough set approach to attribute generalization in data mining[J]. Information Science, 1998, 107(97): 169-176.

[5] 刘少辉, 盛秋戬, 史忠植. 一种新的快速计算正区域的方法[J]. 计算机研究与发展, 2003, 40(5): 637-642.
(Liu S H, Sheng Q J, Shi Z Z. A new method for fast computing positive region[J]. J of Computer Research and Development, 2003, 40(5): 637-642.)

[6] Zheng Zheng, Wang Guoyin. RRIA: A rough set and rule tree-based incremental knowledge acquisition algorithm[J]. Fundamenta Informaticae, 2004, 59(3): 299-313.

[7] Li Tianrui, Ruan Da, Gerret W, et al. A rough set based characteristic relation approach for dynamic attribute generalization in data mining[J]. Knowledge-based Systems, 2007, 20(5): 485-494.

[8] 杨习贝, 吴陈, 傅凡. 不完备信息系统中属性增减下粗集近似概念的更新[J]. 江苏科技大学学报: 自然科学版, 2005, 19(6): 65-69.
(Yang X B, Wu C, Fu F. Updating approximate concept of rough set resulting from inserting or deleting attributes in incomplete information system[J]. J of Jiangsu University of Science and Technology: Natural Science Ed, 2005, 19(6): 65-69.)

[9] Li Shaoyong, Li Tianrui, Liu Dun. Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set[J]. Knowledge-Based Systems, 2013(40): 17-26.

[10] 石为人, 李伟漳, 贾修一. 基于优势关系区别矩阵的一种增量求核方法[J]. 计算机应用研究, 2008, 25(7): 2050-2052.
(Shi W R, Li W W, Jia X Y. Improvement of dominance discernibility matrix and incremental computation of core[J]. Application Research of Computers, 2008, 25(7): 2050-2052.)

[11] Chen H M, Li T R, Ruan D. Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining[J]. Knowledge-Based Systems, 2012(31), 140-161.

[12] Lichman M. UCI machine learning repository[OL]. (2013-01-15)[2013-03-11]. http://archive.ics.uci.edu/ml.

(责任编辑：齐　霁)