一种基于概率粗糙集的属性约简加速算法

刘 芳[1] 李天瑞[2]
(内江师范学院数学与信息科学学院 内江 641101)[1] (西南交通大学信息科学与技术学院 成都 611756)[2]

摘 要 介绍了基于概率粗糙集模型的启发式属性约简算法,提出了概率粗糙集模型中的概率近似精度和改进概率近似精度的增量更新机制,通过比较概率近似精度的更新值得到属性核,然后通过比较改进概率近似精度的值逐步得到概率粗糙集中的属性约简。最后提出了一种概率粗糙集模型中属性核与属性约简的加速求解算法,并举例说明了所提算法的有效性和可行性。

关键词 概率粗糙集,属性约简,增量学习,更新近似集
中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.011

Accelerated Attribute Reduction Algorithm Based on Probabilistic Rough Sets

LIU Fang[1] LI Tian-rui[2]
(Department of Mathematics and Information Science, Neijiang Normal University, Neijiang 641101, China)[1]
(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China)[2]

Abstract A heuristic attribute reduction algorithm based on probabilistic rough sets was introduced. Incremental approaches for computing the probabilistic approximation accuracy and the modified probabilistic approximation accuracy in probabilistic rough sets were presented. The attribute core is obtained by comparing the updated values of the probabilistic approximation accuracy. Then, the attribute reduction of probabilistic rough sets is gradually obtained by comparing the updated values of the modified probabilistic approximation accuracy. Finally, a fast algorithm for calculating the attribute core and attribute reduction based on probabilistic rough sets is developed. And the effectiveness and feasibility of the proposed accelerated algorithm for attribute reduction are validated by illustrative examples.

Keywords Probabilistic rough sets, Attribute reduction, Incremental learning, Updating approximations

1 引言

粗糙集理论是一种有效处理模糊、不确定问题的数学工具,已被成功应用到许多研究领域,如机器学习、数据挖掘、知识发现和智能数据分析等[1-4]。属性约简是粗糙集理论中最重要的研究内容之一,通过属性约简可以提高分类学习算法的性能、简化数据描述和避免过拟合[5-8]。

经典粗糙集模型假设的分类必须是完全正确或确定的,因此,它不能有效地处理有噪声的数据和挖掘其潜在的有用知识。为了解决这种问题,众多研究学者通过引入概率阈值提出了很多扩展概率粗糙集模型,如0.5概率粗糙集模型、变精度粗糙集模型和决策粗糙集模型等[9-11]。由于在概率粗糙集模型中属性集的变化对分类区域的影响不再具有单调性,因此经典粗糙集中的属性约简方法在概率粗糙集模型中不再适用,众多研究学者对这一问题展开了深入研究。Yao提出了一种在决策粗糙集模型中进行属性约简的方法[12]。Chen等基于容差矩阵和属性核的最小约简的概念,提出了变精度粗糙集的属性核思想,并给出了基于属性核的启发式约简以求解最小约简[13]。Jia等以决策成本最小化为目标函数,分别提出了启发式算法、遗传算法和模拟退火算法来对决策粗糙集模型中的属性约简问题进行优化[14]。Jia等还设计了一种自适应学习算法来得到决策粗糙集中的成本函数和阈值,并提出了一种用粒子群优化方法来计算属性约简的方法[15]。Wang等给出了在概率粗糙集中进行启发式属性约简时的单调不确定性度量方法[16],具有较高性能。然而,当单个属性删除或增加时,该方法需要频繁地计算概率近似精度和改进概率近似精度的更新值,从而逐步得到属性核和最小属性约简。这种重复性的计算导致算法的时间复杂度和空间复杂度都较高,而且可能导致错误率增加,最终使得算法的效率低下。

增量学习方法是一种在当前的样本训练中充分利用历史的训练结果,从而显著地减少了后续训练时间的技术,可大幅提高分类的效率。开发基于增量学习技术的高效知识获取方法,已成为粗糙集理论与方法研究中的一个热点问题。例如,Chan提出了Pawlak粗糙集模型中边界集的定义,并利用边界集的更新计算实现了Pawlak粗糙近似集的增量式计算方法[17];Li提出了在特征关系粗糙集模型下利用边界集增量式计算近似集的方法[18];Chen提出了一种当属性值粗化和细化时近似集的增量更新方法[19];Luo提出了在集值有序决策

到稿日期:2015-12-30 返修日期:2015-05-04 本文受国家自然科学基金项目(61175047)资助。
刘 芳(1984—),女,硕士生,讲师,主要研究方向为粗糙集与粒计算等,E-mail:lytb@163.com;李天瑞(1969—),男,教授,博士生导师,主要研究方向为数据挖掘与知识发现、粗糙集与粒计算等,E-mail:trli@swjtu.edu.cn(通信作者)。

信息系统中当属性泛化时近似集的动态更新方法[21]；Zhang提出了复合粗糙集模型中当对象发生变化时近似集的增量式更新方法[22]。

本文针对文献[16]中对概率粗糙集模型中属性约简的计算方法进行了改进，提出了概率粗糙集模型中的概率近似精度和改进概率近似精度的增量式计算方法，进一步给出了概率粗糙集模型中属性核与属性约简的一种快速求解算法。最后，通过实例分析进一步验证了本文所提方法的有效性和可行性。

2 概率粗糙集相关知识

本节简要介绍概率粗糙集模型的基本概念及其属性约简方法[9,11,16,23]。

$IS=(U,A,V,f)$是一个信息系统，$U=\{x_1,x_2,\cdots,x_n\}$表示非空有限的对象集合，A为非空有限的属性集，$V=\bigcup_{a\in A}V_a$，其中V_a是a的值域，映射$f:U\times A\to V$是一个信息函数。$IS=(U,A,V,f)$可以简写为$IS=(U,A)$。

定义1[23] 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq A$和$X\subseteq U$，定义$U$的任意子集$X$的概率上、下近似集分别为：

$$\overline{apr}_R^{(\alpha,\beta)}(X)=\{x\in U\mid p(X\mid[x]_R)>\beta\}$$

$$\underline{apr}_R^{(\alpha,\beta)}(X)=\{x\in U\mid p(X\mid[x]_R)\geqslant\alpha\}$$

概率粗糙集模型中的正域、边界域和负域分别为：

$$POS_R^{(\alpha,\beta)}(X)=\{x\in U\mid p(X\mid[x]_R)\geqslant\alpha\}$$

$$BND_R^{(\alpha,\beta)}(X)=\{x\in U\mid\beta<p(X\mid[x]_R)<\alpha\}$$

$$NEG_R^{(\alpha,\beta)}(X)=\{x\in U\mid p(X\mid[x]_R)\leqslant\beta\}$$

其中，$p(X\mid[x]_R)=\dfrac{|[x]_R\cap X|}{|[x]_R|}$。

定义2[16] 给定一个决策信息系统$DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$C$为条件属性集，$D$为决策属性集，$R\subseteq C$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$为$U$中划分的决策类集合，$D$关于$C$的概率上、下近似集分别定义为：

$$\overline{apr}_R^{(\alpha,\beta)}(U/D)=\overline{apr}_R^{(\alpha,\beta)}(Y_1)\cup\overline{apr}_R^{(\alpha,\beta)}(Y_2)\cup\cdots\cup\overline{apr}_R^{(\alpha,\beta)}(Y_M)$$

$$\underline{apr}_R^{(\alpha,\beta)}(U/D)=\underline{apr}_R^{(\alpha,\beta)}(Y_1)\cup\underline{apr}_R^{(\alpha,\beta)}(Y_2)\cup\cdots\cup\underline{apr}_R^{(\alpha,\beta)}(Y_M)$$

定义3[16] 给定一个决策信息系统$DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$为$U$中划分的决策类集合，定义$U/D$关于$R$的概率近似精度为：

$$\alpha_R^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{apr}_C^{(\alpha,\beta)}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{apr}_R(\overline{apr}_C^{(\alpha,\beta)}(Y_j))|}$$

其中，$j\in\{1,2,\cdots,M\}$。

定义4[16] 设$\pi=\{X_1,X_2,\cdots,X_K\}$是$U$的一个划分，$m:2^U\to\mathcal{R}$是子集的粒度的度量，则在划分$\pi$下的预期粒度定义为：

$$EG_m(\pi)=E_{p_\pi}(m(\cdot))=\sum_{i=1}^{K}m(X_i)p(X_i)$$

其中，$P_\pi=(p(X_1),p(X_2),\cdots,p(X_K))=\left(\dfrac{|X_1|}{|U|},\dfrac{|X_2|}{|U|},\cdots,\dfrac{|X_K|}{|U|}\right)$是在划分$\pi$下的概率分布，$E_{p_\pi}(\cdot)$是相对于分布$P_\pi$的数学期望。

定义5[16] 给定一个决策信息系统$DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$为$U$中划分的决策类集合，$U/D$关于$R$的改进的概率近似精度定义为：

$$\gamma_R^{(\alpha,\beta)}(U/D,m(\cdot))=EG_m(\Pi_1)-(1-\alpha_R^{(\alpha,\beta)}(U/D))EG_m(U/R)$$

其中，$\Pi_1=\{U\}$。

定义6[16] 给定一个决策信息系统$DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，$R$是$DS$在概率粗糙集模型下的属性约简，当且仅当下列式子成立：

(1) $\alpha_R^{(\alpha,\beta)}(U/D)=\alpha_C^{(\alpha,\beta)}(U/D)$；

(2) 对$\forall a\in R$，$\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D)\neq\alpha_C^{(\alpha,\beta)}(U/D)$。

定义7[16] 给定一个决策信息系统$DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$DS$在概率粗糙集模型下的属性核定义为：

$$CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)=\{c\in C\mid\alpha_{C-\{c\}}^{(\alpha,\beta)}(U/D)\neq\alpha_C^{(\alpha,\beta)}(U/D)\}$$

3 概率粗糙集的约简计算方法

在Wang等[16]提出的基于概率粗糙集的启发式属性约简算法中，首先需要计算属性核，并通过在属性核的基础上逐步添加属性计算出一个属性约简，然后在得到的一个属性约简中去掉冗余的属性进而求得最小属性约简。在此过程中会频繁计算单个属性c在已知属性集中删除后的概率近似精度$\alpha_{C-\{c\}}^{(\alpha,\beta)}(U/D)$、单个属性$a$在已知属性集中添加后的改进概率近似精度$\gamma_{R\cup\{a\}}^{(\alpha,\beta)}(U/D,m(\cdot))$。这将导致该算法的效率低下，为解决这一问题，本文在文献[17-22]的基础上，提出了概率粗糙集的概率近似精度和改进概率近似精度的增量式计算方法，并给出了该模型中基于边界集的属性约简加速算法。

3.1 理论分析

定义8[17] 给定一个信息系统$IS=(U,A)$，如果在Pawlak粗糙集模型中的边界域为$BND_R(X)$，下边界集为$\underline{\Delta}_R(X)$，上边界集为$\overline{\Delta}_R(X)$，则边界域和上下边界集的关系可表示为：

$$\begin{aligned}&BND_R(X)=\overline{\Delta}_R(X)\cup\underline{\Delta}_R(X)\\&\underline{\Delta}_R(X)=X-\underline{R}(X)\\&\overline{\Delta}_R(X)=\overline{R}(X)-X\end{aligned}\tag{1}$$

定义8提出了在等价关系Pawlak粗糙集模型中的边界集的概念，将边界域$BND_R(X)$分成了下边界集$\underline{\Delta}_R(X)$和上边界集$\overline{\Delta}_R(X)$两个部分。

定理1[17] 给定一个信息系统$IS=(U,A)$，$X\subseteq U$，$P\subseteq A$，属性$a\notin P$。当新增属性a添加到属性集P时，Pawlak粗糙集中的下近似集更新如下：

$$\underline{apr}_{P\cup\{a\}}(X)=\underline{apr}_P(X)\cup\underline{apr}_{\{a\}}(X)\cup Y$$

其中，$Y=\{x\in\underline{\Delta}_P(X)\cap\underline{\Delta}_{\{a\}}(X)\mid\bigcap_{b\in P\cup\{a\}}[x]_b\subseteq X\}$。

引理1 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$P\subseteq A$，属性$a\notin P$。$\underline{apr}_A^{(\alpha,\beta)}(X)$是概率粗糙集中关于属性集合$A$的下近似集，当新增属性$a$添加到属性集$P$时，下近似集更新如下：

$$\underline{apr}_{P\cup\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cup\underline{apr}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))\cup Y'$$

其中，$Y'=\{x\in\underline{\Delta}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cap\underline{\Delta}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))\mid\bigcap_{b\in P\cup\{a\}}[x]_b\subseteq\underline{apr}_A^{(\alpha,\beta)}(X)\}$。

证明：因为在概率粗糙集模型中关于属性集合A的下近似集$\underline{apr}_A^{(\alpha,\beta)}(X)$是一个对象集合，并且$\underline{apr}_A^{(\alpha,\beta)}(X)\subseteq U$，对于

下近似集$\underline{apr}_{P\cup\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$中的任意对象$x$，如果$x\notin\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cup\underline{apr}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$，那么$x$一定在$Y'$中。这是因为如果$x\notin\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cup\underline{apr}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$，当且仅当$x$在$\underline{\Delta}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cap\underline{\Delta}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$中；而且如果$x\in\underline{apr}_{P\cup\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$，当且仅当$\bigcap\limits_{b\in P\cup\{a\}}[x]_b\subseteq\underline{apr}_A^{(\alpha,\beta)}(X)$成立。因此$x$一定在$Y'$中。即可得到$\underline{apr}_{P\cup\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cup\underline{apr}_{\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))\cup Y'$。证毕。

定理 2[17] 给定一个信息系统$IS=(U,A)$，$X\subseteq U$，$P\subseteq A$，属性$a\notin P$。当新增属性a添加到属性集P时，Pawlak粗糙集中的上近似集更新如下：

$$\overline{apr}_{P\cup\{a\}}(X)=\overline{apr}_P(X)\cap\overline{apr}_{\{a\}}(X)-Z$$

其中，$Z=\{xin\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(X)\mid\bigcap\limits_{b\in P\cup\{a\}}[x]_b\subseteq\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(X)\}$。

引理 2 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$P\subseteq A$，属性$a\notin P$。$\overline{apr}_A^{(\alpha,\beta)}(X)$是概率粗糙集中关于属性集合$A$的上近似集，当新增属性$a$添加到属性集$P$时，上近似更新如下：

$$\overline{apr}_{P\cup\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))=\overline{apr}_P(\overline{apr}_A^{(\alpha,\beta)}(X))\cap\overline{apr}_{\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))-Z'$$

其中，$Z'=\{xin\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))\mid\bigcap\limits_{b\in P\cup\{a\}}[x]_b\subseteq\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))\}$。

证明：令$x\in\overline{apr}_{P\cup\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$，并且$x\notin X$，根据定义8中上边界集的定义可知$x$在$\overline{\Delta}_{P\cup\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$中，这表示$x$在$\overline{\Delta}_P(\overline{apr}_A^{(\alpha,\beta)}(X))$中，而且$\bigcap\limits_{b\in P\cup\{a\}}[x]_b\cap\overline{apr}_A^{(\alpha,\beta)}(X)\neq\varnothing$。因为$(\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X)))\cap\overline{apr}_A^{(\alpha,\beta)}(X)=\varnothing$，然而$\bigcap\limits_{b\in P\cup\{a\}}[x]_b$不是$\bigcap\limits_{b\in P\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$的子集，所以$x$不在$Z'$中。因此可得到$\overline{apr}_{P\cup\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))=\overline{apr}_P(\overline{apr}_A^{(\alpha,\beta)}(X))\cap\overline{apr}_{\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))-Z'$。证毕。

定理 3[17] 给定一个信息系统$IS=(U,A)$，$X\subseteq U$，$P\subseteq A$，属性$a\in P$。当从属性集P中删除属性a时，Pawlak粗糙集中的下近似集更新如下：

$$\underline{apr}_{P-\{a\}}(X)=\underline{apr}_P(X)-\underline{\Delta}_{P-\{a\}}(X)$$

其中，$\underline{\Delta}_{P-\{a\}}(X)=\{xin\bigcap\limits_{b\in P-\{a\}}\underline{\Delta}_{\{b\}}(X)\mid\bigcap\limits_{b\in P-\{a\}}[x]_b\not\subset X\}$。

引理 3 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$P\subseteq A$，属性$a\in P$。$\underline{apr}_A^{(\alpha,\beta)}(X)$是概率粗糙集中关于属性集合$A$的下近似集，当从属性集$P$中删除属性$a$时，下近似集更新如下：

$$\underline{apr}_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))-\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$$

其中，$\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\{xin\bigcap\limits_{b\in P-\{a\}}\underline{\Delta}_{\{b\}}(\underline{apr}_A^{(\alpha,\beta)}(X))\mid\bigcap\limits_{b\in P-\{a\}}[x]_b\not\subset\underline{apr}_A^{(\alpha,\beta)}(X)\}$。

证明：一般而言，$\underline{apr}_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))\subseteq\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))$，$\underline{\Delta}'_P(\underline{apr}_A^{(\alpha,\beta)}(X))\subseteq\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$。属性$a$从属性集$P$中删除前后的边界集的变化可表示为$\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))-\underline{\Delta}'_P(\underline{apr}_A^{(\alpha,\beta)}(X))=\{xinU\mid x\in\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$ and $x\notin\underline{\Delta}'_P(\underline{apr}_A^{(\alpha,\beta)}(X))\}$。所以，当从属性集$P$中删除属性$a$时，下近似集更新为$\underline{apr}_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))-(\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))-\underline{\Delta}'_P(\underline{apr}_A^{(\alpha,\beta)}(X)))$。又因为$\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\cap\underline{\Delta}'_P(\underline{apr}_A^{(\alpha,\beta)}(X))=\varnothing$，所以$\underline{apr}_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))=\underline{apr}_P(\underline{apr}_A^{(\alpha,\beta)}(X))-\underline{\Delta}'_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$。证毕。

定理 4[17] 给定一个信息系统$IS=(U,A)$，$X\subseteq U$，$P\subseteq A$，属性$a\in P$。当从属性集P中删除属性a时，Pawlak粗糙集中的上近似集更新如下：

$$\overline{apr}_{P-\{a\}}(X)=(X)\cup\overline{\Delta}_P(X)\cup W$$

其中，$W=\{xin\bigcap\limits_{b\in P-\{a\}}\overline{\Delta}_{\{b\}}(X)\mid\bigcap\limits_{b\in P-\{a\}}[x]_b\not\subset\bigcap\limits_{b\in P-\{a\}}\overline{\Delta}_{\{b\}}(X)\}$。

引理 4 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$P\subseteq A$，属性$a\in P$。$\overline{apr}_A^{(\alpha,\beta)}(X)$是概率粗糙集中关于属性集合$A$的上近似集，当从属性集$P$中删除属性$a$时上近似集更新如下：

$$\overline{apr}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))=\overline{apr}_A^{(\alpha,\beta)}(X)\cup\overline{\Delta}_P(\overline{apr}_A^{(\alpha,\beta)}(X))\cup W'$$

其中，$W'=\{xin\bigcap\limits_{b\in P-\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))\mid\bigcap\limits_{b\in P-\{a\}}[x]_b\not\subset\bigcap\limits_{b\in P-\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))\}$。

证明：令$x\in\overline{apr}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$，$x\notin X$，根据定义8可知$x$在$\overline{\Delta}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$中，通常$\overline{\Delta}_P(\underline{apr}_A^{(\alpha,\beta)}(X))\subseteq\overline{\Delta}_{P-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))$，因此，如果$x\in\overline{apr}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$，$x\notin X$，而且$x\notin\overline{\Delta}_P(\underline{apr}_A^{(\alpha,\beta)}(X))$，那么$x$一定在$W'$中。这是因为如果$\bigcap\limits_{b\in P-\{a\}}[x]_b\subseteq\bigcap\limits_{b\in P-\{a\}}\overline{\Delta}_{\{b\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$，当且仅当$x\in\overline{apr}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))$成立。所以$\overline{apr}_{P-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))=\overline{apr}_A^{(\alpha,\beta)}(X)\cup\overline{\Delta}_P(\overline{apr}_A^{(\alpha,\beta)}(X))\cup W'$。证毕。

定理1和定理2说明了在Pawlak粗糙集中，当新增属性a添加到属性集P时，利用上、下边界集对粗糙集的上、下近似集做增量式更新。引理1和引理2将定理1和定理2的结果扩展到了概率粗糙集模型中，说明了在概率粗糙集中，当新增属性a添加到属性集P时，利用上、下边界集更新粗糙集的上、下近似集的方法。定理3和定理4说明了在Pawlak粗糙集中，当从属性集P中删除属性a时，利用上、下边界集对粗糙集的上、下近似集做增量式更新。引理3和引理4将定理3和定理4的结果扩展到了概率粗糙集模型中，说明了在概率粗糙集中，当从属性集P中删除属性a时，利用上、下边界集更新粗糙集的上、下近似集的方法。

根据引理1—引理4和定义3可以得到定理5—定理8。

定理 5 给定一个信息系统$IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$R\subseteq A$，属性$a\notin R$，当新增属性$a$添加到属性集$R$中时，概率粗糙集中的概率近似精度更新如下：

$$\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_R(\underline{P}(X))\cup\underline{apr}_{\{a\}}(\underline{P}(X))\cup Y'|}{|\overline{apr}_R(\overline{P}(X))\cap\overline{apr}_{\{a\}}(\overline{P}(X))-Z'|}$$

其中，$\underline{P}(X)=\underline{apr}_A^{(\alpha,\beta)}(X)$，$\overline{P}(X)=\overline{apr}_A^{(\alpha,\beta)}(X)$，$Y'=\{xin\,\underline{\Delta}_R(\underline{P}(X))\cap\underline{\Delta}_{\{a\}}(\underline{P}(X))\mid\bigcap\limits_{b\in R\cup\{a\}}[x]_b\subseteq\underline{P}(X)\}$，$Z'=\{xin\bigcap\limits_{b\in R\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(X))\mid\bigcap\limits_{b\in R\cup\{a\}}[x]_b\subseteq\bigcap\limits_{b\in R\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(X))\}$。

证明：因为$\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_{R\cup\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))|}{|\overline{apr}_{R\cup\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))|}$，令$\underline{P}(X)=\underline{apr}_A^{(\alpha,\beta)}(X)$，$\overline{P}(X)=\overline{apr}_A^{(\alpha,\beta)}(X)$，则$\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_{R\cup\{a\}}(\underline{P}(X))|}{|\overline{apr}_{R\cup\{a\}}(\overline{P}(X))|}$，根据引理1得到$\underline{apr}_{R\cup\{a\}}(\underline{P}(X))=\underline{apr}_R(\underline{P}(X))\cup\underline{apr}_{\{a\}}(\underline{P}(X))\cup Y'$。根据引理2得到$\overline{apr}_{R\cup\{a\}}(\overline{P}$

$(X))=\overline{apr_R}(\overline{P}(X))\cap\overline{apr_{\{a\}}}(\overline{P}(X))-Z'$，所以 $\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr_R}(\underline{P}(X))\cup\underline{apr_{\{a\}}}(\underline{P}(X))\cup Y'|}{|\overline{apr_R}(\overline{P}(X))\cap\overline{apr_{\{a\}}}(\overline{P}(X))-Z'|}$。证毕。

定理 6 给定一个信息系统 $IS=(U,A)$，$0\leqslant\beta<\alpha\leqslant1$，$X\subseteq U$，$R\subseteq A$，属性 $a\in R$，当从属性集 R 中删除属性 a 后，概率粗糙集中的概率近似精度更新如下：

$$\alpha_{R-\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr_R}(\underline{P}(X))-\underline{\Delta}'_{R-\{a\}}(\underline{P}(X))|}{|\overline{P}(X)\cup\overline{\Delta}_R(\overline{P}(X))\cup W'|}$$

其中，$\underline{P}(X)=\underline{apr}_A^{(\alpha,\beta)}(X)$，$\overline{P}(X)=\overline{apr}_A^{(\alpha,\beta)}(X)$，$\underline{\Delta}'_{R-\{a\}}(\underline{P}(X))=\{xin\bigcap_{b\in R-\{a\}}\underline{\Delta}_{\{b\}}(\underline{P}(X))\mid\bigcap_{b\in R-\{a\}}[x]_b\not\subset\underline{P}(X)\}$，$W'=\{xin\bigcap_{b\in R-\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(X))\mid\bigcap_{b\in R-\{a\}}[x]_b\not\subset\bigcap_{b\in R-\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(X))\}$。

证明：因为 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_{R-\{a\}}(\underline{apr}_A^{(\alpha,\beta)}(X))|}{|\overline{apr}_{R-\{a\}}(\overline{apr}_A^{(\alpha,\beta)}(X))|}$，令 $\underline{P}(X)=\underline{apr}_A^{(\alpha,\beta)}(X)$，$\overline{P}(X)=\overline{apr}_A^{(\alpha,\beta)}(X)$，则 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_{R-\{a\}}(\underline{P}(X))|}{|\overline{apr}_{R-\{a\}}(\overline{P}(X))|}$，根据引理 3 得到 $\underline{apr}_{R-\{a\}}(\underline{P}(X))=\underline{apr}_R(\underline{P}(X))-\underline{\Delta}'_{R-\{a\}}(\underline{P}(X))$。根据引理 4 得到 $\overline{apr}_{R-\{a\}}(\overline{P}(X))=\overline{P}(X)\cup\overline{\Delta}_R(\overline{P}(X))\cup W'$，所以 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(X)=\frac{|\underline{apr}_R(\underline{P}(X))-\underline{\Delta}'_{R-\{a\}}(\underline{P}(X))|}{|\overline{P}(X)\cup\overline{\Delta}_R(\overline{P}(X))\cup W'|}$。证毕。

定理 7 给定一个决策信息系统 $DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，属性 $a\notin R$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$ 为 U 中划分的决策类集合，当新增属性 a 添加到属性集 R 中时，概率粗糙集中的概率近似精度更新如下：

$$\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{P}(Y_j))\cup\underline{apr}_{\{a\}}(\underline{P}(Y_j))\cup Y''|}{\sum_{Y_j\in U/D}|\overline{apr}_R(\overline{P}(Y_j))\cap\overline{apr}_{\{a\}}(\overline{P}(Y_j))-Z''|}$$

其中，$\underline{P}(Y_j)=\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\overline{P}(Y_j)=\overline{apr}_C^{(\alpha,\beta)}(Y_j)$，$Y''=\{xin\underline{\Delta}_R(\underline{P}(Y_j))\cap\underline{\Delta}_{\{a\}}(\underline{P}(Y_j))\mid\bigcap_{b\in R\cup\{a\}}[x]_b\subseteq\underline{P}(Y_j)\}$，$Z''=\{xin\bigcap_{b\in R\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(Y_j))\mid\bigcap_{b\in R\cup\{a\}}[x]_b\subseteq\bigcap_{b\in R\cup\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(Y_j))\}$，$j=\{1,2,\cdots,M\}$。

证明：根据定义 3 可知 $\alpha_R^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{apr}_C^{(\alpha,\beta)}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{apr}_R(\overline{apr}_C^{(\alpha,\beta)}(Y_j))|}$，当属性 a 添加到属性集 R 中时，$\underline{apr}_C^{(\alpha,\beta)}(Y_j)$ 和 $\overline{apr}_C^{(\alpha,\beta)}(Y_j)$ 的值是不会发生变化的，只有 $\underline{apr}_R(\underline{apr}_C^{(\alpha,\beta)}(Y_j))$ 和 $\overline{apr}_R(\overline{apr}_C^{(\alpha,\beta)}(Y_j))$ 的值会随着 R 的变化而变化，所以 $\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_{R\cup\{a\}}(\underline{apr}_C^{(\alpha,\beta)}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{apr}_{R\cup\{a\}}(\overline{apr}_C^{(\alpha,\beta)}(Y_j))|}$，令 $\underline{P}(Y_j)=\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\overline{P}(Y_j)=\overline{apr}_C^{(\alpha,\beta)}(Y_j)$，根据引理 1 可知 $\underline{apr}_{R\cup\{a\}}(\underline{P}(Y_j))=\underline{apr}_R(\underline{P}(Y_j))\cup\underline{apr}_{\{a\}}(\underline{P}(Y_j))\cup Y''$。根据引理 2 可知 $\overline{apr}_{R\cup\{a\}}(\overline{P}(Y_j))=\overline{apr}_R(\overline{P}(Y_j))\cap\overline{apr}_{\{a\}}(\overline{P}(Y_j))-Z''$，所以 $\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{P}(Y_j))\cup\underline{apr}_{\{a\}}(\underline{P}(Y_j))\cup Y''|}{\sum_{Y_j\in U/D}|\overline{apr}_R(\overline{P}(Y_j))\cap\overline{apr}_{\{a\}}(\overline{P}(Y_j))-Z''|}$。证毕。

定理 8 给定一个决策信息系统 $DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，属性 $a\in R$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$ 为 U 中划分的决策类集合，当从属性集 R 中删除属性 a 时，概率粗糙集中的概率近似精度更新如下：

$$\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{P}(Y_j))-\underline{\Delta}''_{R-\{a\}}(\underline{P}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{P}(Y_j)\cup\overline{\Delta}_R(\overline{P}(Y_j))\cup W''|}$$

其中，$\underline{P}(Y_j)=\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\overline{P}(Y_j)=\overline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\underline{\Delta}''_{R-\{a\}}(\underline{P}(Y_j))=\{xin\bigcap_{b\in R-\{a\}}\underline{\Delta}_{\{b\}}(\underline{P}(Y_j))\mid\bigcap_{b\in R-\{a\}}[x]_b\not\subset\underline{P}(Y_j)\}$，$W''=\{xin\bigcap_{b\in R-\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(Y_j))\mid\bigcap_{b\in R-\{a\}}[x]_b\not\subset\bigcap_{b\in R-\{a\}}\overline{\Delta}_{\{b\}}(\overline{P}(Y_j))\}$，$j=\{1,2,\cdots,M\}$。

证明：根据定义 3 可知 $\alpha_R^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{apr}_C^{(\alpha,\beta)}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{apr}_R(\overline{apr}_C^{(\alpha,\beta)}(Y_j))|}$，当属性 a 从属性集 R 中删除时，$\underline{apr}_C^{(\alpha,\beta)}(Y_j)$ 和 $\overline{apr}_C^{(\alpha,\beta)}(Y_j)$ 的值是不会发生变化的，只有 $\underline{apr}_R(\underline{apr}_C^{(\alpha,\beta)}(Y_j))$ 和 $\overline{apr}_R(\overline{apr}_C^{(\alpha,\beta)}(Y_j))$ 的值会随着 R 的变化而变化，所以 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_{R-\{a\}}(\underline{apr}_C^{(\alpha,\beta)}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{apr}_{R-\{a\}}(\overline{apr}_C^{(\alpha,\beta)}(Y_j))|}$，令 $\underline{P}(Y_j)=\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\overline{P}(Y_j)=\overline{apr}_C^{(\alpha,\beta)}(Y_j)$，根据引理 3 可知 $\underline{apr}_{R-\{a\}}(\underline{P}(Y_j))=\underline{apr}_R(\underline{P}(Y_j))-\underline{\Delta}''_{R-\{a\}}(\underline{P}(Y_j))$。根据引理 4 可知 $\overline{apr}_{R-\{a\}}(\overline{P}(Y_j))=\overline{P}(Y_j)\cup\overline{\Delta}_R(\overline{P}(Y_j))\cup W''$。所以 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D)=\frac{\sum_{Y_j\in U/D}|\underline{apr}_R(\underline{P}(Y_j))-\underline{\Delta}''_{R-\{a\}}(\underline{P}(Y_j))|}{\sum_{Y_j\in U/D}|\overline{P}(Y_j)\cup\overline{\Delta}_R(\overline{P}(Y_j))\cup W''|}$。证毕。

定理 5 和定理 6 说明了在信息系统中，当已知属性集 R 中添加和删除属性 a 时概率粗糙集中的概率近似精度的增量式更新方法。定理 7 和定理 8 说明了在决策信息系统中，当已知属性集 R 中添加和删除属性 a 时概率粗糙集中的概率近似精度的增量式更新方法。

根据定义 5 和定理 5－定理 8 容易得到定理 9 和定理 10。

定理 9 给定一个决策信息系统 $DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，属性 $a\notin R$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$ 为 U 中划分的决策类集合，当新增属性 a 添加到属性集 R 中时，概率粗糙集中改进的概率近似精度更新如下：$\gamma_{R\cup\{a\}}^{(\alpha,\beta)}(U/D,m(\cdot))=EG_m(\Pi_1)-(1-\alpha_{R\cup\{a\}}^{(\alpha,\beta)}(U/D))EG_m(U/R\cup\{a\})$，其中，$\Pi_1=\{U\}$。

定理 10 给定一个决策信息系统 $DS=(U,C\cup D)$，$0\leqslant\beta<\alpha\leqslant1$，$R\subseteq C$，属性 $a\in R$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$ 为 U 中划分的决策类集合，当从属性集 R 中删除属性 a 时，概率粗糙集中改进的概率近似精度更新如下：$\gamma_{R-\{a\}}^{(\alpha,\beta)}(U/D,m(\cdot))=EG_m(\Pi_1)-(1-\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D))EG_m(U/R-\{a\})$，其中，$\Pi_1=\{U\}$。

定理 9 和定理 10 说明了在决策信息系统中，当已知属性集 R 中添加和删除属性 a 时概率粗糙集中的改进概率近似精度的增量式更新方法。

由引理 1－引理 4、定理 7－定理 10 可以看出概率粗糙集模型中概率近似精度和改进概率近似精度的求解可通过边界集的计算实现增量式的更新求解，进而避免重复计算。基于此，本文给出了概率粗糙集模型中属性核的增量求解算法。

3.2 算法描述

算法 1 概率粗糙集的属性核计算算法

输入：一个决策表 $DS=(U,C\cup D)$，阈值 (α,β)，$C=\{a_1,a_2,\cdots,a_n\}$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$

输出：概率粗糙集的属性核 $CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)$

Step 1　计算 U 在 $\forall a_i \in C$ 上的划分 U/a_i，其中 $i=\{1,2,\cdots,n\}$；

Step 2　计算 $U/C=\bigcap_{i=1}^{n} U/a_i$；

Step 3　计算概率粗糙集模型中的关于条件属性 C 的上、下近似集 $\overline{apr}_C^{(\alpha,\beta)}(Y_j)$、$\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，$\overline{P}(Y_j)=\overline{apr}_C^{(\alpha,\beta)}(Y_j)$，并令 $\underline{P}(Y_j)=\underline{apr}_C^{(\alpha,\beta)}(Y_j)$，其中 $j\in\{1,2,\cdots,M\}$；

Step 4　计算 Pawlak 粗糙集模型中的关于条件 C 的上、下近似集 $\overline{apr}_C(\overline{P}(Y_j))$、$\underline{apr}_C(\underline{P}(Y_j))$，其中 $j\in\{1,2,\cdots,M\}$，然后根据定义 3 计算出 $\alpha_C^{(\alpha,\beta)}(U/D)$；

Step 5　计算 $\alpha_{C-\{a_i\}}^{(\alpha,\beta)}(U/D)$，其中 $i=\{1,2,\cdots,n\}$；

5.1 计算 Pawlak 粗糙集模型下的关于单个属性的上、下近似集 $\overline{apr}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{\{a_i\}}(\underline{P}(Y_j))$，其中 $i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$；

5.2 根据式(1)计算关于单个属性的上、下边界集 $\overline{\Delta}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_j))$，其中 $i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$；

5.3 根据定理 4 和定理 3 计算关于删除一个属性后的属性集的上、下近似集 $\overline{apr}_{C-\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{C-\{a_i\}}(\overline{P}(Y_j))$，其中 $i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$；

5.4 根据定理 8 计算 $\alpha_{C-\{a_i\}}^{(\alpha,\beta)}(U/D)$，其中 $i\in\{1,2,\cdots,n\}$；

Step 6　令 $CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)=\emptyset$，根据 Step 5 的计算结果进行判断，如果 $\alpha_{C-\{a_i\}}^{(\alpha,\beta)}(U/D)\neq\alpha_C^{(\alpha,\beta)}(U/D)$，那么 $CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)=CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)\cup\{a_i\}$，其中 $i\in\{1,2,\cdots,n\}$；

Step 7　$CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)$ 就是所求属性核。

算法 1 描述了一种概率粗糙集模型中的属性核的计算方法，首先计算出对象集在各个属性下的划分及概率粗糙集模型中的关于条件属性 C 的上及下近似集；然后计算从条件属性 C 中删除属性 a_i 时的概率近似精度 $\alpha_{C-\{a_i\}}^{(\alpha,\beta)}(U/D)$；最后判断该概率近似精度，如果 $\alpha_{C-\{a_i\}}^{(\alpha,\beta)}(U/D)\neq\alpha_C^{(\alpha,\beta)}(U/D)$，则 a_i 属于属性核。

在算法 1 的基础上根据对改进概率近似精度的增量式计算，进一步提出了概率粗糙集模型中的属性约简加速算法。

算法 2　概率粗糙集的属性约简算法

输入：

(1)一个决策表 $S=(U,C\cup D)$，阈值 (α,β)，$C=\{a_1,a_2,\cdots,a_n\}$，$U/D=\{Y_1,Y_2,\cdots,Y_M\}$

(2)在算法 1 中已经计算得到的概率粗糙集模型中的关于条件属性 C 的上、下近似集 $\overline{P}(Y_j)$、$\underline{P}(Y_j)$；U 在 $\forall a_i\in C$ 上的划分 U/a_i；Pawlak 粗糙集模型下的关于单个属性的上、下近似集 $\overline{apr}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{\{a_i\}}(\underline{P}(Y_j))$；关于单个属性的上、下边界集 $\overline{\Delta}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_j))$，其中 $i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$

(3)属性核 $CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)$

输出：概率粗糙集的最小属性约简 R

Step 1　令 $R=CORE_{\alpha^{(\alpha,\beta)}(U/D)}(C)$，$CA=C-R$，根据定义 3 计算出 $\alpha_R^{(\alpha,\beta)}(U/D)$；

Step 2　判断 $\alpha_R^{(\alpha,\beta)}(U/D)$ 与 $\alpha_C^{(\alpha,\beta)}(U/D)$ 是否相等，如果不相等，转到 Step 3；如果相等，则得到一个属性约简 R，转到 Step 5；

Step 3　计算 $\gamma_{RU\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$，其中 $a_i\in CA$；

3.1 利用算法 1 中已经计算得到的中间结果：$\overline{P}(Y_j)$、$\underline{P}(Y_j)$、U/a_i、$\overline{apr}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{\{a_i\}}(\underline{P}(Y_j))$、$\overline{\Delta}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_j))$，根据定理 2 和定理 1 计算关于添加一个属性后的属性集的上、下近似集 $\overline{apr}_{RU\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{RU\{a_i\}}(\underline{P}(Y_j))$，其中 $a_i\in CA$，$i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$；

3.2 根据定理 7 计算 $\alpha_{RU\{a_i\}}^{(\alpha,\beta)}(U/D)$，其中 $a_i\in CA$；

3.3 根据定义 4 计算 $EG_m(U)$ 和 $EG_m(U/R\cup\{a_i\})$，其中 $a_i\in CA$；

3.4 根据定理 9 计算 $\gamma_{RU\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$，其中 $a_i\in CA$；

Step 4　根据 Step 3 的计算结果找出使得 $\gamma_{RU\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$ 的值为最大的 a_i，令 $R=R\cup\{a_i\}$、$CA=CA-\{a_i\}$，然后转到 Step 2；

Step 5　删除属性约简 R 中冗余的属性，得到最小属性约简；

5.1 令 $CD=R$，利用算法 1 中已经计算得到的中间结果：$\overline{apr}_{\{a_i\}}(\overline{P}(Y_j))$、$\underline{apr}_{\{a_i\}}(\underline{P}(Y_j))$，计算 $\gamma_{\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$，其中 $a_i\in CD$，$i\in\{1,2,\cdots,n\}$，$j\in\{1,2,\cdots,M\}$；

5.2 对集合 CD 中的对象按照 $\gamma_{\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$ 的大小升序排序；

5.3 令 $CD=CD-\{a\}$，其中 a 为排序后的集合 CD 中的第一个对象。判断如果 $\alpha_{R-\{a\}}^{(\alpha,\beta)}(U/D)$ 和 $\alpha_C^{(\alpha,\beta)}(U/D)$ 相等，就把 a 从约简 R 中去掉，即 $R=R-\{a\}$；如果不相等，则循环执行本步骤直到 $CD\neq\emptyset$ 为止。

Step 6　R 即为所求的最小属性约简。

算法 2 描述了一种概率粗糙集模型中的属性约简加速算法，利用算法 1 的计算结果计算出改进概率近似精度 $\gamma_{RU\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$，找出使得改进概率近似精度最大的属性 a_i 并将其添加到属性集 R 中，直到 $\alpha_R^{(\alpha,\beta)}(U/D)=\alpha_C^{(\alpha,\beta)}(U/D)$ 为止，此时 R 就是一个属性约简。最后利用算法 1 的计算结果计算出 R 中各个属性的改进概率近似精度 $\gamma_{\{a_i\}}^{(\alpha,\beta)}(U/D,m(\cdot))$，按属性的重要度对属性进行排序，并按这个顺序去掉属性约简 R 中冗余的属性，从而得到最小属性约简。

4　算例

给定一个决策信息系统 $DS=(U,C\cup D)$ 如表 1 所列，$U=\{x_1,x_2,\cdots,x_{10}\}$，$C=\{a_1,a_2,a_3,a_4,a_5,a_6\}$。设 $\alpha=0.75$，$\beta=0.60$。

表 1　一个决策信息系统

Car	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	0	0	0	1	1	1	1
x_2	0	0	0	1	1	1	1
x_3	0	0	0	1	1	1	0
x_4	0	1	0	0	0	0	0
x_5	1	1	1	1	0	1	1
x_6	1	1	1	0	1	0	1
x_7	1	1	1	1	0	0	1
x_8	0	0	1	0	0	0	0
x_9	0	0	1	0	0	0	0
x_{10}	0	0	1	0	0	1	1

根据表 1 计算条件属性等价类和决策属性等价类为：

$U/a_1=\{\{x_1,x_2,x_3,x_4,x_8,x_9,x_{10}\},\{x_5,x_6,x_7\}\}$

$U/a_2=\{\{x_1,x_2,x_3,x_8,x_9,x_{10}\},\{x_4,x_5,x_6,x_7\}\}$

$U/a_3=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6,x_7,x_8,x_9,x_{10}\}\}$

$U/a_4=\{\{x_1,x_2,x_3,x_5,x_7\},\{x_4,x_6,x_8,x_9,x_{10}\}\}$

$U/a_5=\{\{x_1,x_2,x_3,x_6\},\{x_4,x_5,x_7,x_8,x_9,x_{10}\}\}$

$U/a_6=\{\{x_1,x_2,x_3,x_5,x_{10}\},\{x_4,x_6,x_7,x_8,x_9\}\}$

$U/C=\{\{x_1,x_2,x_3\},\{x_4\},\{x_5\},\{x_6\},\{x_7\},\{x_8,x_9\},\{x_{10}\}\}$

$U/d=\{Y_1,Y_2\}=\{\{x_1,x_2,x_5,x_6,x_7,x_{10}\},\{x_3,x_4,x_8,x_9\}\}$

4.1　计算属性核

概率粗糙集模型中关于条件属性 C 的上、下近似集为：

$\underline{apr}_C^{(0.75,0.60)}(Y_1)=\{x_5,x_6,x_7,x_{10}\}$

$\underline{apr}_C^{(0.75,0.60)}(Y_2)=\{x_4,x_8,x_9\}$

$\overline{apr}_C^{(0.75,0.60)}(Y_1)=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$

$\overline{apr}_C^{(0.75,0.60)}(Y_2)=\{x_4,x_8,x_9\}$

令$\underline{P}(Y_1)=\underline{apr}_C^{(0.75,0.60)}(Y_1)$，$\underline{P}(Y_2)=\underline{apr}_C^{(0.75,0.60)}(Y_2)$，$\overline{P}(Y_1)=\overline{apr}_C^{(0.75,0.60)}(Y_1)$，$\overline{P}(Y_2)=\overline{apr}_C^{(0.75,0.60)}(Y_2)$。

Pawlak 粗糙集中关于条件属性 C 的上、下近似集为：

$\overline{apr}_C(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$

$\overline{apr}_C(\overline{P}(Y_2))=\{x_4,x_8,x_9\}$

$\underline{apr}_C(\underline{P}(Y_1))=\{x_5,x_6,x_7,x_{10}\}$

$\underline{apr}_C(\underline{P}(Y_2))=\{x_4,x_8,x_9\}$

根据定义 3 计算得到 $\alpha_C^{(0.75,0.60)}(U/d)=\frac{7}{10}$。

Pawlak 粗糙集中关于单个属性的上、下近似集为：

$\underline{apr}_{\{a_1\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7\}$；$\underline{apr}_{\{a_i\}}(\underline{P}(Y_1))=\emptyset$，其中 $i=\{2,3,\cdots,6\}$；$\underline{apr}_{\{a_i\}}(\overline{P}(Y_2))=\emptyset$，$i\in\{1,2,\cdots,6\}$；

$\overline{apr}_{\{a_i\}}(\overline{P}(Y_1))=U$，其中 $i=\{1,2,\cdots,6\}$；

$\overline{apr}_{\{a_1\}}(\overline{P}(Y_2))=\{x_1,x_2,x_3,x_4,x_8,x_9,x_{10}\}$；

$\overline{apr}_{\{a_2\}}(\overline{P}(Y_2))=\overline{apr}_{\{a_3\}}(\overline{P}(Y_2))=U$；

$\overline{apr}_{\{a_4\}}(\overline{P}(Y_2))=\{x_4,x_6,x_8,x_9,x_{10}\}$；

$\overline{apr}_{\{a_5\}}(\overline{P}(Y_2))=\{x_4,x_5,x_7,x_8,x_9,x_{10}\}$；

$\overline{apr}_{\{a_6\}}(\overline{P}(Y_2))=\{x_4,x_6,x_7,x_8,x_9\}$。

根据式(1)得到关于单个属性的上、下边界集为：

$\underline{\Delta}_{\{a_1\}}(\underline{P}(Y_1))=\underline{P}(Y_1)-\underline{apr}_{\{a_1\}}(\underline{P}(Y_1))=\{x_{10}\}$；

$\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7,x_{10}\}$，其中 $i\in\{2,3,\cdots,6\}$；

$\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_2))=\{x_4,x_8,x_9\}$，其中 $i\in\{1,2,\cdots,6\}$；

$\overline{\Delta}_{\{a_i\}}(\overline{P}(Y_1))=\overline{apr}_{\{a_i\}}(\overline{P}(Y_1))-\overline{P}(Y_1)=\{x_4,x_8,x_9\}$，其中 $i\in\{1,2,\cdots,6\}$；

$\overline{\Delta}_{\{a_1\}}(\overline{P}(Y_2))=\{x_1,x_2,x_3,x_{10}\}$；

$\overline{\Delta}_{\{a_2\}}(\overline{P}(Y_2))=\overline{\Delta}_{\{a_3\}}(\overline{P}(Y_2))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$；

$\overline{\Delta}_{\{a_4\}}(\overline{P}(Y_2))=\{x_6,x_{10}\}$；

$\overline{\Delta}_{\{a_5\}}(\overline{P}(Y_2))=\{x_5,x_7,x_{10}\}$；

$\overline{\Delta}_{\{a_6\}}(\overline{P}(Y_2))=\{x_6,x_7\}$。

根据引理 3 得到关于删除一个属性后的属性集的下近似集为：

$\underline{apr}_{C-\{a_1\}}(\underline{P}(Y_1))=\underline{apr}_C(\underline{P}(Y_1))-\underline{\Delta}_{C-\{a_1\}}(\underline{P}(Y_1))$，其中 $\underline{\Delta}_{C-\{a_1\}}(\underline{P}(Y_1))=\{x\,in\bigcap_{b\in C-\{a_1\}}\underline{\Delta}_{\{b\}}(\underline{P}(Y_1))\,|\bigcap_{b\in C-\{a_1\}}[x]_b\not\subset\underline{P}(Y_1)\}$。因为 $\bigcap_{i=2}^{6}\underline{\Delta}_{\{a_i\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7,x_{10}\}$，而且 $\bigcap_{i=2}^{6}[x_5]_{a_i}=\{x_5\}\subset\underline{P}(Y_1)$、$\bigcap_{i=2}^{6}[x_6]_{a_i}=\{x_6\}\subset\underline{P}(Y_1)$、$\bigcap_{i=2}^{6}[x_7]_{a_i}=\{x_7\}\subset\underline{P}(Y_1)$、$\bigcap_{i=2}^{6}[x_{10}]_{a_i}=\{x_{10}\}\subset\underline{P}(Y_1)$，所以对象 x_5，x_6，x_7 和 x_{10} 都不在 $\underline{\Delta}_{C-\{a_1\}}(\underline{P}(Y_1))$ 中，即 $\underline{\Delta}_{C-\{a_1\}}(\underline{P}(Y_1))=\emptyset$。所以 $\underline{apr}_{C-\{a_1\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7,x_{10}\}-\emptyset=\{x_5,x_6,x_7,x_{10}\}$。

同理，可以计算出 $\underline{apr}_{C-\{a_6\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7\}$；$\underline{apr}_{C-\{a_i\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7,x_{10}\}$，其中 $i\in\{2,3,4,5\}$。$\underline{apr}_{C-\{a_i\}}(\underline{P}(Y_2))=\{x_4,x_8,x_9\}$，其中 $i\in\{1,2,\cdots,5\}$；$\underline{apr}_{C-\{a_6\}}(\underline{P}(Y_2))=\{x_4\}$。

根据引理 4 得到关于删除一个属性后的属性集的上近似集为：

$\overline{apr}_{C-\{a_1\}}(\overline{P}(Y_1))=\overline{P}(Y_1)\cup\overline{\Delta}_C(\overline{P}(Y_1))\cup Z'$，其中 $\overline{\Delta}_C(\overline{P}(Y_1))=\overline{apr}_C(\overline{P}(Y_1))-\overline{P}(Y_1)=\emptyset$；$Z'=\{xin\bigcap_{b\in C-\{a_1\}}\overline{\Delta}_{\{b\}}(\underline{P}(Y_1))\,|\bigcap_{b\in C-\{a_1\}}[x]_b\not\subset\bigcap_{b\in C-\{a_1\}}\overline{\Delta}_{\{b\}}(\underline{P}(Y_1))\}$。因为 $\bigcap_{i=2}^{6}\overline{\Delta}_{\{a_i\}}(\underline{P}(Y_1))=\{x_8,x_9\}$，而且 $\bigcap_{i=2}^{6}[x_8]_{a_i}=\bigcap_{i=2}^{6}[x_9]_{a_i}=\{x_8,x_9\}\subset\bigcap_{i=2}^{6}\overline{\Delta}_{\{a_i\}}(\underline{P}(Y_1))$，所以对象 x_8 和 x_9 都不在 Z' 中，即 $Z'=\emptyset$。所以 $\overline{apr}_{C-\{a_1\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$。

同理，可以计算出 $\overline{apr}_{C-\{a_i\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$，其中 $i\in\{2,3,4,5\}$；$\overline{apr}_{C-\{a_6\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_8,x_9,x_{10}\}$。

$\overline{apr}_{C-\{a_i\}}(\overline{P}(Y_2))=\{x_4,x_8,x_9\}$，其中 $i\in\{1,2,\cdots,5\}$；$\overline{apr}_{C-\{a_6\}}(\overline{P}(Y_2))=\{x_4,x_8,x_9,x_{10}\}$。

根据定理 8 得到 $\alpha_{C-\{a_i\}}^{(0.75,0.60)}(U/d)=\frac{7}{10}$，其中 $i\in\{1,2,\cdots,5\}$；$\alpha_{C-\{a_6\}}^{(0.75,0.60)}(U/d)=\frac{4}{13}$。

根据算法 1 中属性核的判断方法可知决策系统的属性核为$\{a_6\}$。

4.2 计算最小属性约简

令 $R=\{a_6\}$，$CA=\{a_1,a_2,a_3,a_4,a_5\}$，根据定义 3 得到 $\alpha_{\{a_6\}}^{(0.75,0.60)}(U/d)=0\neq\alpha_C^{(0.75,0.60)}(U/d)$。

根据定理 9 得到改进的概率近似精度为：$\gamma_{R\cup\{a_1\}}^{(0.75,0.60)}(U/d,m(\cdot))=EG_m(U)-(1-\alpha_{R\cup\{a_1\}}^{(0.75,0.60)}(U/d))EG_m(U/R\cup\{a_1\})$，其中 $\alpha_{R\cup\{a_1\}}^{(0.75,0.60)}(U/d)=\frac{\sum_{Y_j\in U/d}|\underline{apr}_{R\cup\{a_1\}}(\underline{P}(Y_j))|}{\sum_{Y_j\in U/d}|\overline{apr}_{R\cup\{a_1\}}(\overline{P}(Y_j))|}$。因为根据引理 1 可知 $\underline{apr}_{R\cup\{a_1\}}(\underline{P}(Y_1))=\underline{apr}_{\{a_6\}}(\underline{P}(Y_1))\cup\underline{apr}_{\{a_1\}}(\underline{P}(Y_1))\cup Y$，其中 $Y=\{xin\,\underline{\Delta}_{\{a_6\}}(\underline{P}(Y_1))\cap\underline{\Delta}_{\{a_1\}}(\underline{P}(Y_1))\,|\bigcap_{b\in\{a_1,a_6\}}[x]_b\subseteq\underline{P}(Y_1)\}=\emptyset$，所以 $\underline{apr}_{R\cup\{a_1\}}(\underline{P}(Y_1))=\{x_5,x_6,x_7\}$。同理，可以得到 $\underline{apr}_{R\cup\{a_2\}}(\underline{P}(Y_1))=\{x_5\}$，$\underline{apr}_{R\cup\{a_3\}}(\underline{P}(Y_1))=\{x_5,x_{10}\}$，$\underline{apr}_{R\cup\{a_4\}}(\underline{P}(Y_1))=\{x_7,x_{10}\}$，$\underline{apr}_{R\cup\{a_5\}}(\underline{P}(Y_1))=\{x_5,x_6,x_{10}\}$；$\underline{apr}_{R\cup\{a_1\}}(\underline{P}(Y_2))=\{x_4,x_8,x_9\}$，$\overline{apr}_{R\cup\{a_2\}}(\underline{P}(Y_2))=\{x_8,x_9\}$，$\underline{apr}_{R\cup\{a_3\}}(\underline{P}(Y_2))=\{x_4\}$，$\underline{apr}_{R\cup\{a_4\}}(\overline{P}(Y_2))=\emptyset$，$\underline{apr}_{R\cup\{a_5\}}(\underline{P}(Y_2))=\emptyset$。

根据引理 2 可以得到：

$\overline{apr}_{R\cup\{a_1\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_{10}\}$；

$\overline{apr}_{R\cup\{a_2\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_{10}\}$；

$\overline{apr}_{R\cup\{a_3\}}(\overline{P}(Y_1))=\{x_1,x_2,x_3,x_5,x_6,x_7,x_8,x_9,x_{10}\}$；

$\overline{apr}_{R\cup\{a_4\}}(\overline{P}(Y_1))=\overline{apr}_{R\cup\{a_5\}}(\overline{P}(Y_1))=U$；

$\overline{apr}_{R\cup\{a_1\}}(\overline{P}(Y_2))=\{x_4,x_8,x_9\}$；

$\overline{apr}_{R\cup\{a_2\}}(\overline{P}(Y_2))=\overline{apr}_{R\cup\{a_3\}}(\overline{P}(Y_2))=\{x_4,x_6,x_7,x_8,x_9\}$；

$\overline{apr}_{R\cup\{a_4\}}(\overline{P}(Y_2))=\{x_4,x_6,x_8,x_9\}$；

$\overline{apr}_{R\cup\{a_5\}}(\overline{P}(Y_2))=\{x_4,x_7,x_8,x_9\}$。

所以 $\alpha_{RU\{a_1\}}^{(0.75,0.60)}(U/d)=\frac{3}{5}$，$\alpha_{RU\{a_2\}}^{(0.75,0.60)}(U/d)=\frac{3}{13}$，$\alpha_{RU\{a_3\}}^{(0.75,0.60)}(U/d)=\frac{3}{14}$，$\alpha_{RU\{a_4\}}^{(0.75,0.60)}(U/d)=\frac{1}{7}$，$\alpha_{RU\{a_5\}}^{(0.75,0.60)}(U/d)=\frac{3}{14}$。

根据定义4计算得到 $EG_m(U)=1$，$EG_m(U/R\cup\{a_1\})=\frac{4^2+3^2+1^2+2^2}{100}=\frac{3}{10}$，$EG_m(U/R\cup\{a_2\})=EG_m(U/R\cup\{a_3\})=EG_m(U/R\cup\{a_5\})=\frac{3}{10}$，$EG_m(U/R\cup\{a_4\})=\frac{17}{50}$。

然后可以计算得到：

$$\gamma_{RU\{a_1\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{22}{25}=0.88$$

$$\gamma_{RU\{a_2\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{10}{13}=0.77$$

$$\gamma_{RU\{a_3\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{107}{140}=0.76$$

$$\gamma_{RU\{a_4\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{124}{175}=0.71$$

$$\gamma_{RU\{a_5\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{107}{140}=0.76$$

其中，$\gamma_{RU\{a_1\}}^{(0.75,0.60)}(U/d,m(\cdot))$的值最大，所以令 $R=\{a_6\}\cup\{a_1\}$，$CA=\{a_2,a_3,a_4,a_5\}$。

因为 $\alpha_{\{a_1,a_6\}}^{(0.75,0.60)}(U/d)=\frac{3}{5}\neq\alpha_C^{(0.75,0.60)}(U/d)$，所以根据上述计算改进概率近似精度的方法循环计算得到：

$\alpha_{RU\{a_2\}}^{(0.75,0.60)}(U/d)=\frac{3}{5}$；

$\alpha_{RU\{a_i\}}^{(0.75,0.60)}(U/d)=\frac{7}{10}$，其中 $i=\{3,4,5\}$；

$EG_m(U/R\cup\{a_2\})=\frac{13}{50}$，$EG_m(U/R\cup\{a_3\})=\frac{1}{5}$，$EG_m(U/R\cup\{a_4\})=EG_m(U/R\cup\{a_5\})=\frac{11}{50}$。

然后可以计算得到：

$$\gamma_{RU\{a_2\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{112}{125}=0.90$$

$$\gamma_{RU\{a_3\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{47}{50}=0.94$$

$$\gamma_{RU\{a_4\}}^{(0.75,0.60)}(U/d,m(\cdot))=\gamma_{RU\{a_5\}}^{(0.75,0.60)}(U/d,m(\cdot))=\frac{467}{500}=0.93$$

其中，对象 a_3 的改进概率近似精度值最大，将 a_3 添加到属性集 R 中形成新的属性约简。根据以上计算结果可知 $\alpha_{RU\{a_3\}}^{(0.75,0.60)}(U/d)=\frac{7}{10}=\alpha_C^{(0.75,0.60)}(U/d)$，所以由算法2和算法3得到的决策表1的属性约简为 $R=\{a_1,a_3,a_6\}$。

5 实验测试与分析

为了验证提出的基于边界集的概率粗糙集的属性约简算法比非增量式的概率粗糙集的属性约简方法在时间效率上更高，从UCI数据集下载了2个数据集(分别为Flags和Soybean-small)，数据集描述见表2。分别用两种方法对2个数据集进行了测试，并对所消耗的时间进行了比较，实验测试的软硬件环境：CPU Intel Celeron 单核 2.6GHz，内存 2.0GB，Win 7 操作系统，Visual Studio 2010 的 C++开发平台。

表2 数据集描述

数据集	属性个数	对象个数	本文方法(ms)	非增量方法(ms)
Flags	29	194	2363	4568
Soybean-small	35	47	141	746
本文算例	6	10	39	82

从表2可以看出本文提出的基于边界集的概率粗糙集的属性约简算法比非增量式的概率粗糙集的属性约简方法在时间效率上更高。

结束语 本文基于增量学习方法提出了一种基于概率粗糙集的属性约简加速算法，该方法通过借鉴 Pawlak 粗糙集中增量式更新上、下近似集的方法来计算概率粗糙集中当单个属性增加或减少时的上、下近似集。进一步得到概率近似精度和改进概率近似精度，通过比较概率近似精度的值得到属性核，然后在属性核的基础上通过比较改进概率近似精度的值逐步实现概率粗糙集模型中的属性约简快速求解。通过理论分析和实例表明，本文提出的概率粗糙集的属性约简算法是可行的。下一步工作将探讨概率粗糙集模型下对象集变化时的增量式约简算法。

参考文献

[1] Herawan T, Deris M M, Abawajy J H. A rough set approach for selecting clustering attribute[J]. Knowledge-Based Systems, 2010, 23(3): 220-231

[2] Herbert J P, Yao J T. Criteria for choosing a rough set model [J]. Computers & Mathematics with Applications, 2009, 57(6): 908-918

[3] Lin T Y, Syau Y R. Unifying variable precision and classical rough sets: granular approach[C]// Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam. Springer, 2013: 365-373

[4] Shen Q, Jensen R. Rough sets, their extensions and applications [J]. International Journal of Automation & Computing, 2007, 4 (3): 217-228

[5] Hedar A R, Wang J, Fukushima M. Tabu search for attribute reduction in rough set theory[J]. Soft Computing, 2008, 12(9): 909-918

[6] Miao D Q, Zhao Y, Li H X, et al. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model [J]. Information Sciences, 2009, 179(24): 4140-4150

[7] Parthalain N M, Shen Q, Jensen R. A distance measure approach to exploring the rough set boundary region for attribute reduction[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(3): 305-317

[8] Thangavela K, Pethalakshmi A. Dimensionality reduction based on rough set theory: A review[J]. Applied Soft Computing, 2009, 9(1): 1-12

[9] Pawlak Z, Wong S K M, Ziarko W. Rough sets: probabilistic versus deterministic approach[J]. International Journal of Man-Machine Studies, 1988, 29(1): 81-95

[10] Yao Y Y, Wong S K M, Lingras P. A decision-theoretic rough set model[M]. Ras Z W, Zemankova M, Emrich M L. eds., Methodologies for Intelligent Systems 5. 1990

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

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

[13] Chen H, Yang J A, Zhuang Z Q. The core of attributes and minimal attributes reduction in variable precision rough set[J]. Chinese Journal of Computers, 2012, 35(5): 1011-1017

[14] Jia X Y, Liao W H, Tang Z M, et al. Minimum cost attribute reduction in decision-theoretic rough set models[J]. Information Sciences, 2013, 219(10): 151-167

[15] Jia X Y, Tang Z M, Liao W H, et al. On an optimization representation of decision-theoretic rough set model[J]. International Journal of Approximate Reasoning, 2014, 55(1): 156-166

[16] Wang G Y, Ma X A, Yu H. Monotonic uncertainty measures for attribute reduction inprobabilistic rough set model[J]. International Journal of Approximate Reasoning, 2015, 59: 41-67

[17] Chan C C. A rough set approach to attribute generalization in data mining[J]. Journal of Information Sciences, 1998, 107: 169-176

[18] Li T R. A rough sets based characteristic relation approach for dynamic attribute generalization in data mining[J]. Knowledge-Based Systems, 2007, 20: 485-494

[19] Chen H M, Li T R, Qiao S J, et al. A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values[J]. International Journal of Intelligent Systems, 2010, 25(10): 1005-1026

[20] Liu D, Li T R, Ruan D, et al. An incremental approach for inducing knowledge from dynamic information systems[J]. Fundamenta Informaticae, 2009, 94(2): 245-260

[21] Luo C, Li T R, Zhang J B. Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization[J]. Information Sciences, 2014, 257(2): 210-228

[22] Zhang J B, Li T R, Chen H M. Composite rough sets for dynamic data mining[J]. Information Sciences, 2014, 257: 81-100

[23] Yao Y Y. Two semantic issues in a probabilistic rough set model [J]. Fundamenta Informaticae, 2011, 108(3): 249-265

(上接第 45 页)

[3] Qi Chao, Chen Hong-chang, Yu Yan. Micro-blog information diffusion effect based on behavior analysis[J]. Journal of Computer Applications, 2014, 34(8): 2404-2408(in Chinese)
齐超,陈鸿昶,于岩. 基于行为分析的微博信息传播效果[J]. 计算机应用,2014,34(8):2404-2408

[4] Yi Lan-li. Research on Statistical Characteristic Analysis and Modeling for Behavior in Microblog Community Based on Human Dynamics[D]. Beijing: Beijing University of Posts and Telecommunications, 2012(in Chinese)
易兰丽. 基于人类动力学的微博用户行为统计特征分析与建模研究[D]. 北京:北京邮电大学,2012

[5] Zhou Dong-hao, Han Wen-bao. DiffRank: A Novel Algorithm for Information Diffusion Detection in Social Networks[J]. Chinese Journal of Computers, 2014, 37(4): 884-893(in Chinese)
周东浩,韩文报. DiffRank:一种新型社会网络信息传播检测算法[J]. 计算机学报,2014,37(4):884-893

[6] Zhang Yan-chao, Liu Yun, Zhang Hai-feng, et al. The Research of Information Dissemination Model on Online Social Network [J]. Acta Physica Sinica, 2011, 60(5): 66-72(in Chinese)
张彦超,刘云,张海峰,等. 基于在线社交网络的信息传播模型[J]. 物理学报,2011,60(5):66-72

[7] Chen Qian-guo, Zhang Zi-li. Dynamics Behavior and Immune Control Strategies of SIRS Model with Immunization on Scale-free Complex Networks[J]. Computer Science, 2013, 40(6): 211-214(in Chinese)
陈乾国,张自力. 无标度网络上带人工免疫的 SIRS 模型动力学行为及其免疫控制策略[J]. 计算机科学,2013,40(6):211-214

[8] Yang Zi-long, Huang Shu-guang, Wang Zhen, et al. Study on Micro Blog Reposting Model Based on Characteristics of Information Obsolescence[J]. Computer Science, 2014, 41(12): 82-85 (in Chinese)
杨子龙,黄曙光,王珍,等. 基于信息老化特征的微博传播模型[J]. 计算机科学,2014,41(12):82-85

[9] Gu Yi-ran, Xia Ling-ling. The Propagation and Inhibition of Rumors in Online Social Network[J]. Acta Physica Sinica, 2012, 61(23): 238701(in Chinese)
顾亦然,夏玲玲. 在线社交网络中谣言的传播与抑制[J]. 物理学报,2012,61(23):238701

[10] Wang Hui, Han Jiang-hong, Deng Lin, et al. Dynamics of Rumor Spreading in Mobile Social Networks[J]. Acta Physica Sinica, 2013, 62(11): 110505(in Chinese)
王辉,韩江洪,邓林,等. 基于移动社交网络的谣言传播动力学研究[J]. 物理学报,2013,62(11):110505

[11] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in twitter: the million follower fallacy [C]// Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media(ICWSM 2012). Menlo Park: AAAI Press, 2010: 10-17

[12] Crandall D, Cosley D, Huttenlocher D, et al. Feedback effects between similarity and social influence in online communities[C]// Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA, ACM, 2008: 160-168

[13] Singla P, Richardson M. Yes, there is a correlation: From social networks to personal behavior on the Web[C]// Proceeding of the 17th International World Wide Web Conference. Beijing, China, 2008: 665-664

[14] Mao Jia-xin, Liu Yi-qun, Zhang Min, et al. Social Influence Analysis for Micro-Blog User Based on User Behavior[J]. Chinese Journal of Computers, 2014, 37(4): 791-795(in Chinese)
毛佳昕,刘奕群,张敏,等. 基于用户行为的微博用户社会影响力分析[J]. 计算机学报,2014,37(4):791-795

[15] Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media? [C]// Proceddings of the 19th international conference on World Wide Web. ACM, 2010: 591-600

[16] Newman M. Network: an introduction [M]. New York: Oxford University Press, 2009: 449

[17] Li He-yuan, Yu Xiao-ming, Liu Yue, et al. Research on Detecting Spammer in Micro-blogs[J]. Journal of Chinese Information Processing, 2014, 28(3): 62-67(in Chinese)
李赫元,俞晓明,刘悦,等. 中文微博客的垃圾用户检测[J]. 中文信息学报,2014,28(3):62-67