优势-等价关系下序贯三支决策的属性约简

李 艳[1,2] 张 丽[1] 王雪静[1] 陈俊芬[1]
(河北大学数学与信息科学学院河北省机器学习与计算智能重点实验室 河北 保定 071002)[1]
(北京师范大学珠海分校应用数学学院 广东 珠海 519087)[2]

摘 要 序贯三支决策方法是一种能够表示问题中的多重层次粒度,并将多粒度结合起来解决不确定决策问题的有效途径。优势-等价关系粗糙集则是针对条件属性具有偏好关系的分类问题,提取有序信息,对目标概念进行近似,从而形成决策知识。利用传统的优势关系粗糙集方法进行知识约简和提取的效率低下,而目前大部分序贯三支决策方法则局限在符号值属性的信息系统中,对连续值和有序值不能进行有效处理,造成一定程度的信息丢失。因此,将序贯三支决策的思想应用于优势关系粗糙集模型中,定义了一种新的基于序贯三支决策的属性约简及相应的属性重要度,对具有偏好值属性的信息系统进行更加高效的处理,通过多粒度的表示和关系的研究,加速了知识约简过程。选取了多组UCI数据进行实验,结果表明所提出的基于优势关系的序贯三支决策方法能够在保证约简质量的基础上明显降低时间耗费。

关键词 粗糙集,优势关系,决策理论粗糙集,序贯三支决策,属性约简
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.02.037

Attribute Reduction for Sequential Three-way Decisions Under Dominance-Equivalence Relations

LI Yan[1,2] ZHANG Li[1] WANG Xue-jing[1] CHEN Jun-fen[1]
(Key Lab of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding, Hebei 071002, China)[1]
(School of Applied Mathematics, Beijing Normal University, Zhuhai, Zhuhai, Guangdong 519087, China)[2]

Abstract Sequential three-way decision is an effective way to solve problems under multiple levels granularity. Dominance-equivalence relation based rough set approach can be used to handle classification problems for conditional attributes with preference ordered, extract related information, approximate target concepts and finally form the decision-making knowledge. The traditional dominance relation-based rough sets model is very time consuming for knowledge reduction and extraction, however, most of current sequential three-way decision models are limited to information systems of symbolic attributes, which can not process continuous and ordinal values effectively, and will cause a certain degree loss of information. Therefore, this paper applied the idea of sequential three-way decisions to the dominance relation-based rough sets models, defined a new attribute reduction method based on sequential three-way decisions and the corresponding attribute importance measure, and then accelerated the processing of information systems with ordinal attributes. Finally, the efficiency of knowledge reduction is improved through multiple granularity representations and relationships. Several UCI data sets are selected for experiments. The results show that the proposed sequential three-decision method based on dominance relations can reduce the time consumption noticeably and guarantee the quality of the attribute reduction.

Keywords Rough set, Dominance relation, Decision theory rough set, Sequential three-way decisions, Attribute reduction

1 引言

粗糙集理论是近年来发展起来的一种处理不精确、不确定及模糊信息的软件计算工具。传统的粗糙集模型只能处理名义型符号值数据,对解决实际问题具有很大的局限性。为了处理带有偏好值的数据,Greco等[1]最早基于优势关系定义了优势类和上下联合的上近似和下近似,其能够有效地利用条件属性和决策中的有序信息,并提取相应的决策知识。

到稿日期:2018-02-08 返修日期:2018-05-12 本文受国家自然科学基金(61473111),河北省自然科学基金(F2018201096,F2016201161),河北大学自然科学研究计划项目(799207217069),北京师范大学珠海分校教师科研能力促进计划资助。

李 艳(1976-),女,博士,教授,CCF会员,主要研究方向为机器学习、Rough集理论、计算智能,E-mail:ly@hbu.cn(通信作者);张 丽(1990-),女,硕士生,主要研究方向为粒计算与知识发现;王雪静(1994-),女,硕士生,主要研究方向为粒计算与知识发现;陈俊芬(1976-),女,博士,副教授,主要研究方向为计算智能、图像处理。

王国胤等[2]研究了条件属性具有偏好值而决策为分类的多准则分类问题,提出了正域约简的概念;李艳等进一步针对多准则分类问题,考虑不协调目标信息系统,在条件属性中引入优势关系,在决策属性上引入等价关系,并定义了相应的近似集、属性重要度和约简[3];后又于 2017 年提出了一种基于优势-等价关系的快速正域约简方法[4]。在决策问题的研究中,研究者基于多软件概念,在软集理论中将对象包含在决策类的价值集中,用于发现基于偏好关系的对象之间的关系[5]。另一方面,为了更好地处理不确定的信息,姚一豫等[6-8]提出了决策粗糙集理论模型,此类模型通过应用贝叶斯决策理论确定了两个阈值,而后再对论域进行分类。基于人类解决问题的思维特点,Yao 等[9-11]最先提出三支决策理论,用粗糙集理论中的概率正域、负域和边界域来体现接受、拒绝和延迟决定 3 种不同的决策。该理论模拟了人类思维,适用于解决不确定决策问题。近年来,三支决策被广泛应用于实际应用领域中,如投资管理[12]、聚类分析[13-14]、人脸识别[15]、属性约简[16-17]和其他问题[18-19]。姚一豫等[20]于 2011 年又提出了序贯三支决策,实现了一种渐进式计算的粒度计算思想,利用粗细不同的粒度表示与多粒度之间的关系,能够有效提高知识约简和提取的效率。在此基础上,Qian 等[21]提出了动态粒化下的序贯三支决策的属性约简方法,并将序贯三支决策的思想应用于决策粗糙集模型,避免了决策修正带来的大量计算耗费。目前的序贯三支决策方法主要是处理不带偏好关系的符号值数据,为了扩展序贯三支决策的应用,将其用于处理连续值和带有偏好关系的符号值数据,我们把序贯三支决策的思想引入到优势-等价关系的粗糙集模型中。首先给出优势-等价关系下的概率近似、概率正域、负域和边界域的定义;然后进一步建立相应的属性约简方法。实验说明,该方法在保证了属性约简效果的同时,降低了计算耗费。

本文第 2 节分别回顾决策粗糙集模型和基于优势-等价关系的粗糙集模型的概念;第 3 节给出基于优势-等价关系的序贯三支决策的概念;第 4 节给出优势关系下的序贯三支决策的属性约简、相应的属性重要度以及相应的属性约简的算法;第 5 节给出实验结果以说明算法的有效性;最后总结全文并提出对未来的展望。

2　基本知识

定义 1(决策表)　决策表定义为四元组 $S=(U,A,V,I)$,其中 U 是对象的有限非空集合,即论域 $U=\{x_1,x_2,\cdots,x_n\}$;$A=C\cup D$ 是非空属性集,其中 $C=\{c_1,c_2,\cdots,c_m\}$ 是条件属性集,D 是决策属性集;V 是所有属性值的集合;$I:U\times A\to V$ 是 U 中的对象 x 到 V 的映射。为了简单起见,假设 $D=\{d\}$,其中 d 是有 k 个不同决策值的决策属性,且 $V_d=\{1,2,\cdots,k\}$。通过考虑原始决策属性的笛卡尔积,一个多重决策属性可以转化成一个单个决策属性。

定义 2(等价关系和等价类,以及基于等价类的划分)　设 $B\subseteq A$,B 的一个等价关系定义为 $IND(B)=\{(x,y)\in U\times U\mid \forall a\in B, I_a(x)=I_a(y)\}$,其中 $I_a(x)$ 表示对象 x 在属性 a 下的属性值。由不可分辨关系 $IND(B)$ 诱导的划分由 $U/IND(B)$ 表示。为了简便,用 U/B 来表示 $U/IND(B)$。对象 x 的等价类定义为 $[x]_{IND(B)}=\{y\in U\mid (x,y)\in IND(B)\}=\{y\in U\mid \forall b\in B, I_b(x)=I_b(y)\}$。类似地,用 $[x]_B$ 来代替 $[x]_{IND(B)}$。

基于上述定义,决策属性集 D 对论域 U 的划分为 $U/D=\{D_1,D_2,\cdots,D_k\}$,条件属性 B 对论域的划分表示为 $U/B=\{B_1,B_2,\cdots,B_r\}$。

定义 3(优势关系、优势类)　给定 $P\subseteq C$ 且 $x\in U$,优势关系定义为 $R_P^{\preccurlyeq}=\{(x_i,x_j)\in U\times U\mid I_a(x_i)\leqslant I_a(x_j),\forall a\in P\}$。论域中一个对象 x 的优势类定义为:$[x]_P^{\preccurlyeq}=R_P^{+}(x)=\{y\in U\mid xR_P^{\preccurlyeq}y\}=\{y\in U\mid I_a(y)\geqslant I_a(x),\forall a\in P\}$,其中,$R_P^{+}$ 是优于 x 的集合,称为 x 的 P-优势集[1]。

条件属性集 C 决定的优势类为 $U/R_C^{\preccurlyeq}=\{[x_1]_C^{\preccurlyeq},[x_2]_C^{\preccurlyeq},\cdots,[x_n]_C^{\preccurlyeq}\}$。

定义 4(优势-等价关系)　对于条件属性集 C 的任意子集 B(即 $\forall B\subseteq C$)和决策属性集 D,定义:

$$R_B^{\preccurlyeq}=\{(x_i,x_j)\in U\times U\mid I_l(x_i)\leqslant I_l(x_j),\forall a_l\in B\}$$

$$R_D=\{(x_i,x_j)\in U\times U\mid I_l(x_i)=I_l(x_j),\forall a_l\in B\}$$

定义 5[3](基于优势-等价关系的近似集)　对于 $P\subseteq C$,在优势-等价关系下,所有确定属于决策类 D_i 的对象的集合构成 D_i 的 P-下近似,记作 $\underline{P}(D_i)$;所有可能属于 D_i 的对象集构成 D_i 的 P-上近似,记作 $\overline{P}(D_i)$,即:

$$\underline{P}(D_i)=\{x\in U:R_P^{+}(x)\subseteq D_i\}$$

$$\overline{P}(D_i)=\{x\in U:R_P^{+}(x)\cap D_i\neq\varnothing\},\text{对于 } t=1,2,\cdots,k$$

定义 6[7-8](决策粗糙集中的近似集、正域、边界域和负域)　设 $B\subseteq C$,考虑决策类 D_i 关于划分 U/B 的近似集,由贝叶斯决策理论确定两个参数 α 和 β,概率下近似和上近似有如下定义:

$$\underline{P}_{U/B}^{(\alpha,\beta)}(D_i)=\{x\in U\mid p(D_i\mid[x]_B)\geqslant\alpha\}=\bigcup_{p(D_i\mid[x]_B)\geqslant\alpha}[x]_B$$

$$\overline{P}_{U/B}^{(\alpha,\beta)}(D_i)=\{x\in U\mid p(D_i\mid[x]_B)>\beta\}=\bigcup_{p(D_i\mid[x]_B)>\beta}[x]_B$$

其中,$p(D_i\mid[x]_B)$ 是等价关系 $[x]_B$ 中的对象 x 属于 D_i 的条件概率,即 $p(D_i\mid[x]_B)=\dfrac{|[x]_B\cap D_i|}{|[x]_B|}$。

对应以上的近似集,则正域、边界域和负域的定义如下:

$$POS_{U/B}^{(\alpha,\beta)}(U/B)=\bigcup_{1\leqslant i\leqslant k}\underline{P}_{U/B}^{(\alpha,\beta)}(D_i)$$

$$BND_{U/B}^{(\alpha,\beta)}(U/B)=\bigcup_{1\leqslant i\leqslant k}(\overline{P}_{U/B}^{(\alpha,\beta)}(D_i)-\underline{P}_{U/B}^{(\alpha,\beta)}(D_i))$$

$$NEG_{U/B}^{(\alpha,\beta)}(U/B)=U-POS_{U/B}^{(\alpha,\beta)}(U/B)\cup BND_{U/B}^{(\alpha,\beta)}(U/B)$$

经典的粗糙集模型可以看成是概率粗糙集模型的特殊情况,即当 $\alpha=1,\beta=0$ 时的模型。

3　基于优势-等价关系的序贯三支决策

本节结合传统的序贯三支决策方法和基于优势关系粗糙集的模型提出:优势-等价关系下的概率近似集,优势-等价关系下的概率正域、边界域和负域,优势-等价关系下的动态粒序及目标概念的近似集,优势-等价关系下的 l 级动态粒序,目标概念的近似集及优势-等价关系下的 l 级动态粒序及目标概念的正域、边界域和负域的概念,建立基于优势-等价关系的决策粗糙集理论和方法,进一步引入序贯三支决策的思想,进而提出相应的约简算法。

定义 7(优势-等价关系下的概率近似集)　类似于决策

理论粗糙集，在优势关系粗糙集中引入一对阈值 α 和 β（$0\leqslant\beta<\alpha\leqslant1$），$\alpha$ 和 β 是由贝叶斯决策理论程序确定的，因此，给定一对阈值 α 和 β（$0\leqslant\beta<\alpha\leqslant1$），对 $Q\subseteq C$ 决策类 D_i 关于 $[x]_Q^{\leqslant}$ 的上近似和下近似可分别定义为：

$$\overline{P}_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D_i)=\{x\in U\mid p(D_i\mid[x]_Q^{\leqslant})>\beta\}=\bigcup_{p(D_i\mid[x]_Q^{\leqslant})>\beta}[x]_Q^{\leqslant}$$

$$\underline{P}_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D_i)=\{x\in U\mid p(D_i\mid[x]_Q^{\leqslant})\geqslant\alpha\}=\bigcup_{p(D_i\mid[x]_Q^{\leqslant})\geqslant\alpha}[x]_Q^{\leqslant}\tag{1}$$

其中，$p(D_i\mid[x]_Q^{\leqslant})$ 表示决策类 D_i 中对象属于优势类 $[x]_Q^{\leqslant}$ 的概率，即 $p(D_i\mid[x]_Q^{\leqslant})=\dfrac{|D_i\cap[x]_Q^{\leqslant}|}{|[x]_Q^{\leqslant}|}$。

定义 8（优势-等价关系下的概率正域、边界域和负域）对应于定义 7 的正域、边界域和负域可分别定义为：

$$POS_{R_Q^{\leqslant}}^{(\alpha,\beta)}(U/D)=\bigcup_{1\leqslant i\leqslant k}\underline{P}_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D_i)$$

$$BND_{R_Q^{\leqslant}}^{(\alpha,\beta)}(U/D)=\bigcup_{1\leqslant i\leqslant k}(\overline{P}_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D_i)-\underline{P}_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D_i))$$

$$NEG_{R_Q^{\leqslant}}^{(\alpha,\beta)}(U/D)=U-POS_{R_Q^{\leqslant}}^{(\alpha,\beta)}(U/D)\cup BND_{R_Q^{\leqslant}}^{(\alpha,\beta)}(U/D)\tag{2}$$

显然，当 $\alpha=1,\beta=0$ 时，该模型是优势关系粗糙集，属于基于优势关系的决策理论粗糙集的特殊情况。

定理 1 给定决策表 $S=(U,A,V,I)$，$Q\subseteq C$，Q 是 C 的真子集。$U/D=\{D_1,D_2,\cdots,D_k\}$，对于 $x\in U$，定义 $m(D\mid[x]_Q^{\leqslant})=\max\limits_{D_i\in U/D}\dfrac{|[x]_Q^{\leqslant}\cap D_i|}{|[x]_Q^{\leqslant}|}$，则有：

$$x\in POS_{R_Q^{\leqslant}}^{(\alpha,\beta)}(D)\Rightarrow\begin{cases}x\in POS_{R_C^{\leqslant}}^{(\alpha,\beta)}(D),\ m([x]_C^{\leqslant},D)\geqslant\alpha\\ x\notin POS_{R_C^{\leqslant}}^{(\alpha,\beta)}(D),\ m([x]_C^{\leqslant},D)<\alpha\end{cases}$$

证明：假设 $x\in POS_{R_Q^{\leqslant}}^{\alpha}(D)$，则满足 $m([x]_Q^{\leqslant},D)\geqslant\alpha$，$Q\subset C$，有 $Q\neq C$，则对于 $\forall D_i\in U/D$，有：

$$\frac{|[x]_C^{\leqslant}\cap D_i|}{|[x]_C^{\leqslant}|}\geqslant\frac{|[x]_Q^{\leqslant}\cap D_i|}{|[x]_Q^{\leqslant}|}\text{或}$$

$$\frac{|[x]_C^{\leqslant}\cap D_i|}{|[x]_C^{\leqslant}|}<\frac{|[x]_Q^{\leqslant}\cap D_i|}{|[x]_Q^{\leqslant}|}$$

因此有两种结果：

（1）$m([x]_C^{\leqslant},D)\geqslant\alpha$；

（2）$m([x]_C^{\leqslant},D)<\alpha$。

对于（1），存在 $D_i\in U/D$ 满足 $\dfrac{|[x]_C^{\leqslant}\cap D_i|}{|[x]_C^{\leqslant}|}\geqslant\alpha$，于是有 $p(D_i\mid[x]_C)\geqslant\alpha$，从而 $x\in POS_{R_C^{\leqslant}}^{(\alpha,\beta)}(D)=\bigcup\limits_{D_i\in U/D}POS_{R_C^{\leqslant}}^{(\alpha,\beta)}(D_i)$；对于（2），有 $\forall D_i\in U/D$，$\dfrac{|[x]_C^{\leqslant}\cap D_i|}{|[x]_C^{\leqslant}|}<\alpha$ 成立，因此 $x\notin POS_{R_C^{\leqslant}}^{(\alpha,\beta)}(D)$。

由定理 1 可知，优势-等价关系下的正域不一定会随着属性的增加而增加，因此，在加入新的属性时，可能需要决策更正。

基于以上定义，进一步引入序贯三支决策的思想，通过不同的粒度表示粗糙集模型的 3 个域。级别较低的粒度较粗，级别较高的粒度较细，较高级别需要较大的 α 阈值和较小的 β 阈值，因此，阈值满足下列关系：

$$0\leqslant\beta_m\leqslant\beta_{m-1}\leqslant\cdots\leqslant\beta_1<\alpha_1\leqslant\alpha_2\leqslant\cdots\leqslant\alpha_m\leqslant1$$

优势关系下第 l 级的 (α_l,β_l)-下近似和 (α_l,β_l)-上近似如定义 9 所示。

定义 9（优势-等价关系下动态粒序及目标概念的近似集）对于决策表 S，令 $Q_1\subseteq Q_2\subseteq\cdots\subseteq Q_m$ 是条件属性集的序列，给定 $P_m^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_m^{\leqslant}\}$，$R_t^{\leqslant}$ 是由 $B_t(t=1,2,\cdots,m)$ 诱导的优势关系。给定一个序 $P_l^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_l^{\leqslant}\}(1\leqslant l\leqslant m)$ 和一个动态阈值参数序列 $(\alpha,\beta)_l=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_l,\beta_l)\}$，在当前造粒序 $P_l^{\leqslant}$ 下，(α_l,β_l) 的下近似 $\underline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)$ 和 (α_l,β_l) 的上近似 $\overline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)$ 可定义为：

$$\begin{aligned}&\underline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)=\{x\mid p(D_i^l\mid[x]_{R_l}^{\leqslant})\geqslant\alpha_l,x\in U_i\}\\&\overline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)=\{x\mid p(D_i^l\mid[x]_{R_l}^{\leqslant})>\beta_l,x\in U_i\}\end{aligned}\tag{3}$$

其中，$U_1=U$，$U_{l+1}=\overline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)-\underline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)$，论域逐渐缩减，$D_i^l$ 表示在划分 U_l/D 中包含 x 的等价类。

定义 10（优势-等价关系下 l 级动态粒序及目标概念的近似集） 对于决策表 S，令 $Q_1\subseteq Q_2\subseteq\cdots\subseteq Q_m$ 是条件属性集的序列，给定 $P_m^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_m^{\leqslant}\}$，$R_t^{\leqslant}$ 是由 $B_t(t=1,2,\cdots,m)$ 诱导的优势关系。给定一个序 $P_l^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_l^{\leqslant}\}(1\leqslant l\leqslant m)$ 和一个动态阈值参数序 $(\alpha,\beta)_l=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_l,\beta_l)\}$，在当前造粒序 $P_l^{\leqslant}$ 下，关于 D_i 的 $P_l^{(\alpha,\beta)_l}$ 下近似 $\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)$ 和 $P_l^{\{\alpha,\beta)_l}$ 上近似 $\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)$ 的定义如下：

$$\begin{aligned}\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)&=\bigcup_{1\leqslant t\leqslant l}\underline{P}_{R_t^{\leqslant}}^{U_t(\alpha_t,\beta_t)}(D_i^t)\\&=\bigcup_{1\leqslant t\leqslant l}\{x\mid p(D_i^t\mid[x]_{R_t}^{\leqslant})\geqslant\alpha_t,x\in U_t\}\\\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)&=\bigcup_{1\leqslant t\leqslant l}\overline{P}_{R_t^{\leqslant}}^{U_t(\alpha_t,\beta_t)}(D_i^t)\\&=\bigcup_{1\leqslant t\leqslant l}\{x\mid p(D_i^t\mid[x]_{R_t}^{\leqslant})>\beta_t,x\in U_t\}\end{aligned}\tag{4}$$

其中，$U_1=U$，$U_{l+1}=\overline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)-\underline{P}_{R_l^{\leqslant}}^{U_l(\alpha_l,\beta_l)}(D_i^l)$，论域逐渐缩减，$D_i^l$ 表示在划分 U_l/D 中包含 x 的等价类。

定义 11（优势-等价关系下 l 级动态粒序及目标概念的正域、边界域和负域） 对于决策表 S，令 $Q_1\subseteq Q_2\subseteq\cdots\subseteq Q_m$ 是条件属性集的序列，给定 $P_m^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_m^{\leqslant}\}$，$R_t^{\leqslant}$ 是由 $B_t(t=1,2,\cdots,m)$ 诱导的优势关系。给定一个序 $P_l^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_l^{\leqslant}\}(1\leqslant l\leqslant m)$ 和一个动态阈值参数序列 $(\alpha,\beta)_l=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_l,\beta_l)\}$，在当前造粒序 $P_l^{\leqslant}$ 下，一个决策划分 $U/D=\{D_1,D_2,\cdots,D_k\}$，在粒化序列 P_l 下 U/D 的近似集为：

$$\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)=\{\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_1),\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_2),\cdots,\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_k)\},$$

$$\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)=\{\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_1),\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_2),\cdots,\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_k)\}$$

根据以上定义，基于优势关系的关于目标决策划分 $U/D=\{D_1,D_2,\cdots,D_k\}$ 的概率正域、边界域和负域可定义为：

$$\begin{aligned}&POS_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)=\bigcup_{1\leqslant i\leqslant k}\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)\\&BND_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)=\bigcup_{1\leqslant i\leqslant k}(\overline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i)-\underline{P}_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D_i))\\&NEG_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)=U-POS_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)\cup BND_{P_l^{\leqslant}}^{(\alpha,\beta)_l}(D)\end{aligned}\tag{5}$$

根据以上定义，下面来说明优势关系下的关于目标决策划分 U/D 的概率正域的单调性。

定理 2 给定一个决策表 S，令 $Q_1\subseteq Q_2\subseteq\cdots\subseteq Q_m$ 是条件属性集的序列，给定 $P_m^{\leqslant}=\{R_1^{\leqslant},R_2^{\leqslant},\cdots,R_m^{\leqslant}\}$，$R_t^{\leqslant}$ 是由 $B_t(t=1,2,\cdots,m)$ 诱导的优势关系。给定一个序 $P_l^{\leqslant}=\{R_1^{\leqslant},$

$R_2^{\preccurlyeq},\cdots,R_l^{\preccurlyeq}\}(1\leqslant l\leqslant m)$,在现有的粒序 P_l 下,给定一个动态阈值参数序列$(\alpha,\beta)_l=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2)\cdots,(\alpha_n,\beta_n)\}$和一个决策划分 $U/D=\{D_1,D_2,\cdots,D_k\}$,对于任意粒序 $P_l^{\preccurlyeq}$,有 $POS_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)\subseteq POS_{P_2^{\preccurlyeq}}^{(\alpha,\beta)_2}(D)\subseteq\cdots\subseteq PSO_{P_l^{\preccurlyeq}}^{(\alpha,\beta)_l}(D)$。

证明:对于任意的两个动态粒化序 P_{j+1} 和 P_j,其中 $P_{j+1}=P_j\cup\{R_{j+1}\}$, $POS_{P_{j+1}^{\preccurlyeq}}^{(\alpha,\beta)_{j+1}}(D)=POS_{P_j^{\preccurlyeq}}^{(\alpha,\beta)_j}(D)\cup\bigcup\limits_{1\leqslant i\leqslant k}\underline{P}_{R_{j+1}^{\preccurlyeq}}^{U_{j+1}^{(\alpha_{j+1},\beta_{j+1})}}(D_i^{j+1})$。因此,$POS_{P_j^{\preccurlyeq}}^{(\alpha,\beta)_j}(D)\subseteq POS_{P_{j+1}^{\preccurlyeq}}^{(\alpha,\beta)_{j+1}}(D)$ 成立,进而 $POS_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)\subseteq POS_{P_2^{\preccurlyeq}}^{(\alpha,\beta)_2}(D)\subseteq\cdots\subseteq POS_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)$ 成立。

上述定理说明正域随着属性的增加而增大,因此能够有效避免决策修订。

算例　考虑一个简单决策表,如表 1 所列。其中,论域 $U=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}\}$,条件属性 $C=\{a_1,a_2,a_3,a_4,a_5,a_6,a_7\}$,决策属性 $D=\{d\}$。给定序 $P_3^{\preccurlyeq}=\{R_1^{\preccurlyeq},R_2^{\preccurlyeq},R_3^{\preccurlyeq}\}$,其中,$R_1^{\preccurlyeq},R_2^{\preccurlyeq},R_3^{\preccurlyeq}$ 分别是由$\{a_2,a_7\}$,$\{a_2,a_7,a_3,a_6\}$,$\{a_2,a_7,a_3,a_6,a_5\}$诱导的偏序关系。通过贝叶斯决策理论确定的阈值$(\alpha,\beta)_3=\{(0.85,0.55),(0.95,0.5),(1,0.45)\}$。

表 1　算例中的决策表

Table 1　Decision table in example

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
x_1	2	1	1	1	2	3	1	2
x_2	1	2	2	1	2	1	3	3
x_3	1	2	3	1	1	1	2	2
x_4	1	2	3	1	2	3	3	3
x_5	2	1	1	2	2	1	3	2
x_6	1	3	1	1	2	2	3	2
x_7	2	3	3	3	1	3	2	2
x_8	2	3	1	1	1	2	1	2
x_9	3	2	3	1	2	3	1	3
x_{10}	3	2	3	3	3	1	1	1

令 $U_1=U$,$U/D=\{\{x_{10}\},\{x_1,x_3,x_5,x_6,x_7,x_8\},\{x_2,x_4,x_9\}\}$。

(1)第一层的优势类

$U_1/R_1^{\preccurlyeq}=\{\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}\},\{x_2,x_4,x_6\},\{x_2,x_3,x_4,x_6,x_7\},\{x_2,x_4,x_6\},\{x_2,x_4,x_5,x_6\},\{x_6\},\{x_6,x_7\},\{x_6,x_7,x_8\},\{x_2,x_3,x_4,x_6,x_7,x_8,x_9,x_{10}\},\{x_2,x_3,x_4,x_6,x_7,x_8,x_9,x_{10}\}\}$

$POS_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)=\bigcup\limits_{1\leqslant i\leqslant k}\underline{P}_{R_1^{\preccurlyeq}}^{(\alpha_1,\beta_1)}(D_i)$

$=\underline{P}_{R_1^{\preccurlyeq}}^{(\alpha_1,\beta_1)}(D_1^1)\cup\underline{P}_{R_1^{\preccurlyeq}}^{(\alpha_1,\beta_1)}(D_2^1)\cup\underline{P}_{R_1^{\preccurlyeq}}^{(\alpha_1,\beta_1)}(D_3^1)$

$=\{x_6,x_7,x_8\}$

(2)更新减少的论域

$U_2=\overline{P}_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)-\underline{P}_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)=\{x_1,x_2,x_3,x_4\}$

$U_2/D=\{\{x_1,x_3\},\{x_2,x_4\}\}$

$U_2/R_2^{\preccurlyeq}=\{\{x_1,x_4\},\{x_2,x_4\},\{x_3,x_4\},\{x_4\}\}$

$POS_{P_2^{\preccurlyeq}}^{(\alpha,\beta)_2}(D)=\{x_2,x_4,x_6,x_7,x_8\}$

(3)进一步更新减少的论域

$U_3=\overline{P}_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)-\underline{P}_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)=\{x_1,x_3\}$

$U_2/D=\{\{x_1,x_3\}\}$

$U_3/R_3^{\preccurlyeq}=\{\{x_1\},\{x_3\}\}$

$POS_{P_3^{\preccurlyeq}}^{(\alpha,\beta)_3}(D)=\{x_1,x_2,x_3,x_4,x_6,x_7,x_8\}$

可以看出,由于每层的论域都是上一层的边界域而不是所有论域,随着属性的增加,论域逐渐减少,所要计算的范围即论域逐渐减小,这会显著降低计算正域所需要的时间和存储耗费。

定理 3　给定一个决策表 S,令 $Q_1\subseteq Q_2\subseteq\cdots\subseteq Q_m$ 是条件属性集的序列,给定 $P_m^{\preccurlyeq}=\{R_1^{\preccurlyeq},R_2^{\preccurlyeq},\cdots,R_m^{\preccurlyeq}\}$,$R_t^{\preccurlyeq}$ 是由 $B_t(t=1,2,\cdots,m)$ 诱导的优势关系。给定一个序 $P_l^{\preccurlyeq}=\{R_1^{\preccurlyeq},R_2^{\preccurlyeq},\cdots,R_l^{\preccurlyeq}\}(1\leqslant l\leqslant m)$,在现有的粒序 $P_l^{\preccurlyeq}$ 下,给定一个动态阈值参数序列$(\alpha,\beta)_l=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_l,\beta_l)\}$和一个决策划分 $U/D=\{D_1,D_2,\cdots,D_k\}$。存在最小整数 $t\in\{1,2,\cdots,m\}$使得 $POS_{P_t^{\preccurlyeq}}^{(\alpha,\beta)_t}(D)=POS_{P_{t+1}^{\preccurlyeq}}^{(\alpha,\beta)_{t+1}}(D)=\cdots=POS_{P_m^{\preccurlyeq}}^{(\alpha,\beta)_m}(D)$成立。

证明:根据定理 2,有 $POS_{P_1^{\preccurlyeq}}^{(\alpha,\beta)_1}(D)\subseteq POS_{P_2^{\preccurlyeq}}^{(\alpha,\beta)_2}(D)\subseteq\cdots\subseteq POS_{P_l^{\preccurlyeq}}^{(\alpha,\beta)_l}(D)$,显然,必然存在 $t\in\{1,2,\cdots,m\}$使得 $POS_{P_t^{\preccurlyeq}}^{(\alpha,\beta)_t}(D)=POS_{P_m^{\preccurlyeq}}^{(\alpha,\beta)_m}(D)$,因此,有 $POS_{P_t^{\preccurlyeq}}^{(\alpha,\beta)_t}(D)=POS_{P_{t+1}^{\preccurlyeq}}^{(\alpha,\beta)_{t+1}}(D)=\cdots=POS_{P_m^{\preccurlyeq}}^{(\alpha,\beta)_m}(D)$。

由定理 3 可以看出,在定义了序贯三支决策下的优势-等价关系的概率正域后,正域随着属性的增加而增大,当属性集增加到最大时,概率正域也是最大的。

4　优势关系下的序贯三支决策的属性约简

本节在优势-等价关系中引入序贯三支决策模型,定义了基于优势-等价关系的几种不同约简,进而给出相应约简算法的步骤。

定义 12(优势关系下的正域及正域约简)[3]　设决策表 $S=(U,A,V,I)$,$POS_C(D)=\bigcup\limits_{1\leqslant i\leqslant k}\underline{P_C^{\preccurlyeq}}(D_i)$是决策属性 D 关于条件属性 C 的正域。对于条件集 C 的子集 $B\subseteq C$,如果 $POS_B(D)=POS_C(D)$,则称集合 B 是优势关系下的正域协调集。若 B 是正域协调集且 B 的任意真子集都不是优势关系下的正域协调集,则称 B 是优势关系下的正域约简。

类似于优势关系下的正域及正域约简,下面给出优势-等价关系下的概率正域约简的定义。

定义 13(优势-等价关系下的概率正域约简)　设决策表 $S=(U,A,V,I)$,$POS_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}(D)=\bigcup\limits_{1\leqslant i\leqslant k}\underline{P}_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}(D_i)$是决策属性 D 关于条件属性集 C 的概率正域。对于 $Q\subseteq C$,如果 $POS_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(D)=POS_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}(D)$,则称 Q 是优势关系下的概率正域协调集。如果 Q 是优势关系下的概率正域协调集,且 Q 的任意真子集都不是概率正域协调集,则称 Q 是优势关系下的概率正域约简。

类似于优势关系下的相容约简[3],下面给出优势关系下的概率相容约简的定义。

定义 14(优势关系下的概率相容约简)　设决策表 $S=(U,A,V,I)$,对于 $Q\subseteq C$,$Q\neq\varnothing$,若 $COM_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}=COM_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}$,则称 Q 是优势关系下的概率相容协调集。若 Q 是优势关系下的概率相容协调集,且其任意真子集都不是优势关系下的概率相容协调集,称 Q 是优势关系下的概率相容约简。其中,

$COM_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}=\{x_i\in U\mid p([x_i]_D\mid[x_i]_C^{\preccurlyeq})\geqslant\alpha\}$，$p([x_i]_D\mid[x_i]_C^{\preccurlyeq})=\frac{|[x_i]_D\cap[x_i]_C^{\preccurlyeq}|}{|[x_i]_C^{\preccurlyeq}|}$。

徐伟华等[22]已经给出了优势关系下的分布约简与最大分布约简，相应地，我们给出优势关系下的概率分布约简与概率最大分布约简的定义。

定义 15（优势关系下的概率分布约简与概率最大分布约简） 设决策表 $S=(U,V,A,I)$，对于 $Q\subseteq C$，$R_Q^{\preccurlyeq}$ 是基于条件属性集 Q 的偏序关系，R_D 是基于决策属性的等价关系，定义：

$$\eta_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)=\{x_i\mid x_i\in U\wedge\frac{|[x_i]_Q^{\preccurlyeq}\cap[x_i]_D|}{|[x_i]_Q^{\preccurlyeq}|}\geqslant\alpha\}$$

$$\mu_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)=\{\frac{|[x_1]_Q^{\preccurlyeq}\cap[x_1]_D|}{|[x_1]_Q^{\preccurlyeq}|},\frac{|[x_2]_Q^{\preccurlyeq}\cap[x_2]_D|}{|[x_2]_Q^{\preccurlyeq}|},\cdots,\frac{|[x_r]_Q^{\preccurlyeq}\cap[x_r]_D|}{|[x_r]_Q^{\preccurlyeq}|}\}$$

其中，$U/R_Q^{\preccurlyeq}=\{[x_i]_Q^{\preccurlyeq}\mid x_i\in U\}$，$U/D=\{[x_i]_D\mid x_i\in U\}$。

如果 $\mu_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)=\mu_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}(S)$，则称 Q 是 S 的概率分布协调集；如果 Q 是 S 的概率分布协调集，且 Q 的任意真子集都不是概率分布协调集，则称 Q 为优势关系下的概率分布约简。

如果 $\eta_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)=\eta_{R_C^{\preccurlyeq}}^{(\alpha,\beta)}(S)$，则称 Q 是概率最大分布协调集；如果 Q 是概率最大分布协调集，且 Q 的任意真子集都不是概率最大分布协调集，则称 Q 是优势关系下的概率最大分布约简。其中，(1)$\mu_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)$是决策表 S 在优势关系下的概率分布函数；(2)$\eta_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}(S)$是决策表 S 在优势关系下的概率最大分布函数。

依赖函数用于刻画一个属性子集对于给定决策的依赖度，在等价关系下，钱宇华等[19]给出了一个粒度序 P 关于决策属性 D 的依赖函数。类似地，我们定义优势关系下粒度 $P_l^{\preccurlyeq}$ 序关于决策 D 的概率依赖函数。

定义 16（概率依赖函数） 令决策表 $S=(U,A,V,I)$，则粒度序 $P_l^{\preccurlyeq}$ 关于决策属性 D 的概率依赖函数定义为：

$$\gamma_{P_m^{\preccurlyeq}}^{(\alpha,\beta)}(D)=\frac{|POS_{P_m^{\preccurlyeq}}^{U}(D)|}{|U|}\tag{6}$$

其中，$POS_{P^{\preccurlyeq}}^{U_l}(D)$是决策 D 关于 U_l 上粒度序 $P_l^{\preccurlyeq}$ 的正域，$0\leqslant\gamma_{P_l^{\preccurlyeq}}^{(\alpha,\beta)}(D)\leqslant1$。

定义 17 对于决策表 $S=(U,A,V,I)$，其中 $A=C\cup D$，$Q\subseteq C$，$\forall a\in Q$，则 a 在 Q 的内重要度定义为：

$$SIG_{\text{inner}}^{(\alpha,\beta)}(a,Q,D)=\gamma_{R_Q^{\preccurlyeq}}(D)-\gamma_{R_{Q-\{a\}}^{\preccurlyeq}}(D)\tag{7}$$

其中，$\gamma_{R_Q^{\preccurlyeq}}(D)=\frac{|POS_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}|}{|U|}$。

定义 18 对于决策表 $S=(U,A,V,I)$，其中 $A=C\cup D$，$Q\subseteq C$，$\forall a\in C-Q$，则 a 在 Q 的外重要度定义为：

$$SIG_{\text{outer}}^{(\alpha,\beta)}(a,Q,D)=\gamma_{R_{Q\cup\{a\}}^{\preccurlyeq}}(D)-\gamma_{R_Q^{\preccurlyeq}}(D)\tag{8}$$

其中，$\gamma_{R_Q^{\preccurlyeq}}(D)=\frac{|POS_{R_Q^{\preccurlyeq}}^{(\alpha,\beta)}|}{|U|}$。

算法 1 优势-等价关系下的序贯三支决策属性约简算法

输入：决策表 S；$(\alpha,\beta)=\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_l,\beta_l)\}$

输出：优势-等价关系下的概率正域约简 $\text{Red}^{(\alpha,\beta)}$

Step1 $\text{Red}^{(\alpha,\beta)}=\emptyset$，$U_1=U$，$t=1$，$Q_1=\emptyset$，$P^{\preccurlyeq}=\emptyset$

Step2 如果 $U_t=\emptyset$，或 $t\geqslant|C|$，则转 Step7

Step3 $a\in C-Q_t$

{计算 $SIG_{\text{outer}}^{(\alpha_t,\beta_t)}(a,Q_t,D)$，如果 $SIG_{\text{outer}}^{(\alpha_t,\beta_t)}(c,Q_t,D)=\max\{SIG_{\text{outer}}^{(\alpha_t,\beta_t)}(a,Q_t,D),a\in C-Q_t\}$

$\text{Red}^{(\alpha,\beta)}=\text{Red}^{(\alpha,\beta)}\cup\{c\}$

}

Step4 $Q_t=Q_t\cup\{c\}$，$P^{\preccurlyeq}=P^{\preccurlyeq}\cup\{R_{Q_t}^{\preccurlyeq}\}$

Step5 $U_{t+1}=\overline{P}_{P_{Q_t}^{\preccurlyeq}}^{(\alpha,\beta)_t}(D)-\underline{P}_{P_{Q_t}^{\preccurlyeq}}^{(\alpha,\beta)_t}(D)$

Step6 $Q_{t+1}=Q_t$，$t+1$，转 Step2

Step7 输出 $\text{Red}^{(\alpha,\beta)}$

该算法的主要思想是：在由属性 Q_1 决定的最粗的造粒空间中，删除属于概率正域和负域的那些对象，并获得新的论域 U_2，再在通过添加条件属性决定的较粗粒度空间中，对于更新的论域 U_2，在动态粒序 $P_2^{\preccurlyeq}$ 下，删除属于概率正域和负域的对象，再次更新论域。重复此过程，直到已更新的论域变成空集或者没有条件属性可以增加。其中，Step3 是根据定义 18 计算的属性重要度由高到低依次添加条件属性集中的属性到约简集中。

5 实验结果与分析

在一台 PC 机上进行实验，配置为 Windows7，3.20 GHz CPU 和内存 8GB，编程软件为 Matlab7.1。对传统的优势-等价关系下的正域约简和序贯三支决策下的正域约简两种方法进行了比较，优势-等价关系下的序贯三支决策在添加新的条件属性时不再对所有的对象进行处理，而是对上一级留下的边界域进行处理，因此明显节省了计算时间。表 2 给出了实验所需要的数据[23]。为方便起见，后文中用 $\text{data}_1-\text{data}_9$ 来表示这些数据集。

表 2 实验数据

Table 2 Data sets

数据集	样例个数	属性个数
Gotracktracks (data_1)	163	8
Bupa (data_2)	345	7
Ecoli (data_3)	336	8
AutoMPG (data_4)	393	8
Backuplarge (data_5)	212	28
IndianLiverPatientDatasetILPD (data_6)	583	11
Pima (data_7)	768	9
Statlog (data_8)	690	15
Yeast (data_9)	1484	9

在此，选取了不同大小的数据集，样例个数从 163 到 1484，属性个数从 7 到 28。本实验中更倾向于选择属性个数较多的数据集，以便构造动态粒序，体现实验的效果。其中，虽然有几个数据的大小比较相近，如 data_3 和 data_4，但是它们的属性类型不同，data_3 为数值型数据，data_4 为混合型数据。

表 3 列出了在 9 组数据集上的优势-等价关系下的传统粗糙集方法和序贯三支决策的计算时间、所得属性约简和缩减比（所提方法的运行时间相对于传统方法减少的比率）。由本文所提算法可以看出，该算法是根据属性的重要度将候选的条件属性依次添加进约简中，直至在动态粒序下的边界域为空集。这就说明所得到的属性约简保持了原决策表的辨识能力，也就是保留了决策表中最重要的信息。对于原决策表的属性集所能辨别的样例对，约简后的决策表也可以辨别。在这种情况下，我们认为约简属性集越小越好。

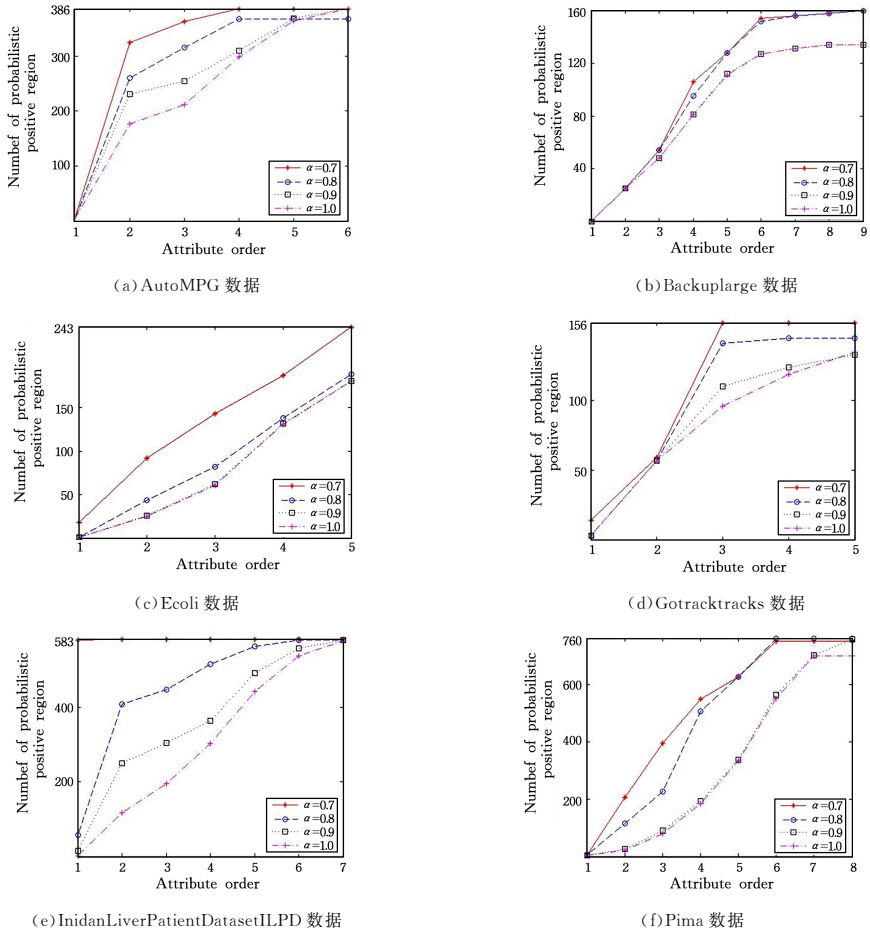
表 3　约简时间结果比较

Table 3　Comparison of reduction time

数据集	原始属性	传统方法		序贯三支决策方法		缩比率/%
		约简后的属性	时间/s	约简后的属性	时间/s	
$data_1$	8	6	0.95651	6	0.5195	45.68
$data_2$	7	6	3.4400	6	3.4400	24.41
$data_3$	8	6	3.9998	5	3.3126	17.18
$data_4$	8	6	4.8783	6	2.3447	51.93
$data_5$	28	16	5.9673	13	3.9455	33.88
$data_6$	11	10	14.4648	9	7.4958	48.17
$data_7$	9	8	21.7792	7	15.5978	28.18
$data_8$	15	13	28.6488	9	10.4751	63.43
$data_9$	9	8	172.1694	8	158.7173	7.81

由表 3 可以得知:优势-等价关系下的序贯三支决策比传统的优势-等价关系粗糙集的计算时间短,而且其中一些数据约简后的属性比传统方法约简后的属性还要少,例如 $data_1$,$data_3$,$data_4$,$data_5$ 和 $data_6$ 中。上述结果中,参数取值分别为:$\alpha=0.8$,$\beta=0$。

下面从中取出 6 组数据来研究阈值对正域的影响,如图 1 所示,横轴表示添加的属性个数,纵轴表示正域中所含样例的个数。其中,设置 $\alpha=0.7,0.8,0.9,1$。显然,概率正域的大小随着属性的增加而均匀增加。如图 1 所示,概率正域($\alpha=0.7$)中对象的数目比其他($\alpha=0.8,0.9$)的数目多得多。概率正域($\alpha=1$)中对象的数量最少。通常,α 越大,概率正域的大小越小,当 α 变小时,其迅速达到概率正域的最大值。



(a) AutoMPG 数据　(b) Backuplarge 数据

(c) Ecoli 数据　(d) Gotracktracks 数据

(e) InidanLiverPatientDatasetILPD 数据　(f) Pima 数据

图 1　概率阈值对正域的影响

Fig. 1　Effect of probability threshold on positive region

结束语　为处理带有偏好值的信息系统并提高约简效率,定义了优势-等价关系下的概率粗糙集模型和序贯三支决策模型,给出了新的序贯三支决策的概率正域;建立了优势关系下的序贯三支决策的属性约简、相应的属性重要度以及相应的属性约简的算法,并在 UCI 数据集中验证了方法的有效性。

在以后的工作中,将进一步研究成本结构和粒度级别,并研究不同的动态粒序下基于优势关系的属性约简方法。今后的工作中还考虑研究动态信息系统中的相关属性约简方法。

参 考 文 献

[1] GRECO S, MATARAZZO B, SLOWINSKI R. Rough approximation by dominance relations[J]. International Journal of Intelligent Systems, 2002, 17(2): 153-171.

[2] CHEN J, WANG G Y, HU J. Positive domain reduction based on dominance relation in inconsistent system[J]. Computer Science, 2008, 35(3): 216-218. (in Chinese)
陈娟,王国胤,胡军. 优势关系下不协调信息系统的正域约简[J]. 计算机科学, 2008, 35(3): 216-218.

[3] LI Y, SUN N X, ZHAO J, et al. Reductions based on dominance-

equivalence relations and rule extraction methods [J]. Computer Science, 2011, 38(11): 220-224. (in Chinese)
李艳, 孙娜欣, 赵津, 等. 基于优势-等价关系的几种约简及规则抽取方法[J]. 计算机科学, 2011, 38(11): 220-224.

[4] JIN Y, LI Y, HE Q. A fast positive-region reduction method based on dominance-equivalence relations [C] // International Conference on Machine Learning and Cybernetics. IEEE, 2017: 152-157.

[5] AWANG M I, ROSE A N M, AWEANG M K, et al. Multiple criteria preference relation by dominance relations in soft set theory[M]. Berlin: Springer International Publishing, 2016: 475-484.

[6] YAO Y Y. Decision-theoretic rough set models[C] // International Conference on Rough Sets and Knowledge Technology. Springer, Berlin, Heidelberg, 2007: 1-12.

[7] YAO Y Y, WONG S K M. A decision theoretic framework for approximating concepts[J]. International Journal of Man-Machine Studies, 1992, 37(6): 793-809.

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

[9] YAO Y Y. An outline of a theory of three-way decisions [C] // International Conference on Rough Sets and Current Trends in Computing. Springer, Berlin: Heidelberg, 2012: 1-17.

[10] 贾修一, 商林, 周献中, 等. 三支决策理论与应用[M]. 南京: 南京大学出版社, 2012.

[11] YAO Y. Rough sets and three-way decisions[M]. Berlin: Springer International Publishing, 2015.

[12] LIU D, YAO Y Y, LI T R. Three-way investment decisions with decision-theoretic rough sets[J]. International Journal of Computational Intelligence Systems, 2011, 4(1): 66-74.

[13] LINGRAS P, CHEN M, MIAO D. Rough cluster quality index based on decision theory[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(7): 1014-1026.

[14] YU H, LIU Z, WANG G. An automatic method to determine the number of clusters using decision-theoretic rough set[M]. Amsterdam: Elsevier Science Inc. 2014.

[15] LI H, ZHANG L, HUANG B, et al. Sequential three-way decision and granulation for cost-sensitive face recognition [J]. Knowledge-Based Systems, 2016, 91(C): 241-251.

[16] JIA X, SHANG L, ZHOU B, et al. Generalized attribute reduct in rough set theory [J]. Knowledge-Based Systems, 2016, 91(C): 204-218.

[17] MENG Z, SHI Z. On quick attribute reduction in decision-theoretic rough set models[J]. Information Sciences, 2016, 330: 226-244.

[18] LI J, HUANG C, QI J, et al. Three-way cognitive concept learning via multi-granularity[J]. Information Sciences, 2017, 378(1): 244-263.

[19] LI J, MEI C, XU W, et al. Concept learning via granular computing: a cognitive viewpoint [J]. Information Sciences, 2015, 298(1): 447-467.

[20] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data[M]. Boston: Kluwer Academic Publishers, 1991.

[21] QIAN J, DANG C, YUE X, et al. Attribute reduction for sequential three-way decisions under dynamic granulation[J]. International Journal of Approximate Reasoning, 2017, 85: 196-216.

[22] XU W H, ZHANG X X, ZHANG W X. Upper approximation reduction in inconsistent target information system based on dominance relations [J]. Computer Engineering, 2009, 35(18): 191-193. (in Chinese)
徐伟华, 张晓燕, 张文修. 优势关系下不协调目标信息系统的上近似约简[J]. 计算机工程, 2009, 35(18): 191-193.

[23] BACHE K, LICHMA M. UCI Machine Learning Repository [OL]. http://archive.ics.uci.edu/ml.