

# 连续值属性决策表中的可变精度粗糙集模型及属性约简

冯 林<sup>1,2</sup> 李天瑞<sup>2</sup> 余志强<sup>3</sup>

(四川师范大学计算机科学学院 成都 610101)<sup>1</sup> (西南交通大学信息科学与技术学院 成都 610031)<sup>2</sup>  
(成都飞机工业(集团)有限责任公司 成都 610092)<sup>3</sup>

**摘 要** 属性约简是粗糙集理论研究的一个核心问题。为了有效地处理决策表中连续值属性约简,提出了连续值属性决策表中的可变精度粗糙集模型以及基于此模型的连续值属性约简算法。仿真实验结果表明,该算法可以对连续值属性进行约简,而且比经典粗糙集相关方法在处理连续值属性约简方面更有效。

**关键词** 粗糙集,属性约简,可变精度粗糙集,属性重要性  
**中图法分类号** TP301 **文献标识码** A

## Attributes Reduction Based on the Variable Precision Rough Set in Decision Tables Containing Continuous-valued Attributes

FENG Lin<sup>1,2</sup> LI Tian-rui<sup>2</sup> YU Zhi-qiang<sup>3</sup>

(College of Computer Science, Sichuan Normal University, Chengdu 610101, China)<sup>1</sup>  
(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)<sup>2</sup>  
(Chengdu Aircraft Industrial(Group) Co., Ltd, Chengdu 610092, China)<sup>3</sup>

**Abstract** Attribute reduction is one of the key problems of the rough set theory. In order to effectively use the rough set theory to deal with the problem of attribute reduction in Decision Tables containing Continuous-Valued Attributes (DTCVA) directly, a new variable precision rough set model and a heuristic algorithm for attributes reduction in DTCVA were developed. Simulation results show that the proposed approach is effective for reduction of continuous-valued attributes, and more efficient than the classical rough set approaches in processing attribute reduction in decision information systems containing continuous-valued attributes.

**Keywords** Rough set, Attribute reduction, Variable precision rough set, Attributes significance

粗糙集(Rough Set, RS)理论是波兰科学家 Pawlak 提出的一种处理不确定和含糊信息的重要数学工具。自这一理论提出以来,越来越多的研究人员对它进行研究并不断发展完善,从理论上建立了粗糙集理论的数学模型,提出了用粗糙集理论构建诸如机器学习、智能控制、不确定数据处理、故障诊断等问题的众多算法模型<sup>[1-3]</sup>。

现实世界中存在很多连续值决策表,比如在医院诊疗系统中,有身高、体温、血压等连续值数据。用经典的粗糙集理论对这类性质的决策表进行属性约简和知识获取时,首先需要对连续值数据进行离散化处理。但是离散化的过程可能导致信息丢失,因而影响知识获取的结果。针对这一问题,一些学者已经进行了有益的研究。如肖迪等人<sup>[4]</sup>提出了广义邻域关系下的实域粗糙集扩展模型,利用广义邻域关系对论域划分相容类,构成集合的上、下近似,从而避免了离散化过程;刘文军等人<sup>[5]</sup>引入  $t$  相似类的概念,定义了相似划分算法和广义决策表,并给出了面向连续值信息系统的决策规则提取算法。该方法也可以避免离散化的步骤,但是在该方法中参数  $t$

的选择标准和优化还是一个待解决的问题;Pal 等人<sup>[6]</sup>利用模糊集理论把连续值属性模糊化,然后基于模糊粗糙集理论提出了专家系统中案例推理、检索方法,但模糊隶属函数的确定需要专家经验。

本文把 Ziarko 的可变精度粗糙集模型 VPRS<sup>[7]</sup>推广到连续值属性决策表 DTCVA(Decision Tables containing Continuous-Valued Attributes)中,提出了 DTCVA 中的可变精度粗糙集模型 VPRSCA(Variable Precision Rough Set in decision tables containing Continuous-valued Attributes);基于 VPRSCA,提出了 DTCVA 中的属性约简方法 CVAR(Continuous-Valued Attributes Reduction)。与经典粗糙集相关方法在处理连续值属性约简方面进行了对比实验,结果表明了本文方法的有效性。

### 1 基本定义

粗糙集理论对知识进行表达和处理的基本工具是决策表(也称决策信息系统)。它可以表示为  $S=(U, R, V, f)$ 。其中

到稿日期:2009-10-28 返修日期:2010-01-15 本文受国家自然科学基金(No. 60873108),四川省教育厅科研基金(No. 09ZC079),四川师范大学重点研究课题资助。

冯 林(1972-),男,博士,副教授,主要研究方向为粗糙集理论、数据挖掘等, E-mail: scfengyc@126.com;李天瑞(1969-),男,博士后,教授,博士生导师,主要研究方向为粗糙集理论、粒计算、数学模型等;余志强(1963-),男,研究员级高级工程师,主要研究方向为智能控制。

$U$  是有限非空对象集合, 也称论域;  $R=C \cup D$  是有限非空属性集合,  $C \cap D = \emptyset, D \neq \emptyset$ , 子集  $C$  和  $D$  分别称为条件属性集和决策属性集;  $V = \bigcup_{r \in R} V_r$  是属性值的集合,  $V_r$  表示属性  $r$  的值域;  $f: U \times R \rightarrow V$  是一个信息函数, 它指定  $U$  中的每一个对象  $x$  的属性值。

本文中,  $\forall c \in C$ , 如果  $c$  取连续实数值, 则  $S$  称为连续值属性决策表 DTCVA。

在 Pawlak 粗糙集模型中, 各对象之间用不分明关系(等价关系)来刻画。针对 DTCVA, 由于属性为连续值, 我们用相似关系来描述两个对象之间的不分明程度。

**定义 1** 给定 DTCVA, 对于任意  $B \subseteq C$  和  $a_i \in B$ , 定义  $B$  上的一个相似关系  $R(B)$  为

$$R(B)(x, y) = \frac{1}{1 + \sum_{i=1}^{|B|} |f(x, a_i) - f(y, a_i)|} \quad (1)$$

式中,  $a_i \in B$ 。  $R(B)(x, y)$  的值域为  $(0, 1]$ , 满足自反性、对称性。

本文实验部分, 在使用定义 1 计算各对象的相似程度时, 所有连续值属性用最大-最小值方法被标准化到  $[0, 1]$ , 以减少因各属性量纲不一致对结果的影响。

记  $\theta_B(x)$  表示对象  $x$  在属性子集  $B$  上满足式(1)的相似类;  $\theta_B(x)(z)$  表示  $x$  与对象  $z$  的相似程度。借鉴模糊集中模糊蕴涵的思想, 对象  $x$  在知识空间(属性集)  $B$  上包含于  $X(X \subseteq U)$  的 Rough 包含函数  $I_B(x, X)$  可定义为

$$I_B(x, X) = \inf_{y \in U} (1 - \theta_B(x)(y) + \theta_B(x)(y) \times X(y)) \quad (2)$$

式中,  $X(y)$  表示对象  $y$  隶属于集合  $X$  的程度。由于  $X \subseteq U$ , 因此式(2)可重新定义为

$$I_B(x, X) = \inf_{y \notin X} (1 - \theta_B(x)(y)) \quad (3)$$

于是, DTCVA 中的  $\beta$  下近似集  $B^\beta_-(X)$  和  $\beta$  上近似集  $B^{-\beta}(X)$  可定义为

$$\begin{cases} B^\beta_-(X) = \bigcup \{x | x \in U \wedge I_B(x, X) \geq 1 - \beta\} \\ B^{-\beta}(X) = \bigcup \{x | x \in U \wedge I_B(x, X) \geq \beta\} \end{cases} \quad (4)$$

称  $(B^\beta_-(X), B^{-\beta}(X))$  为连续值信息系统中的可变精度粗糙集模型 VPRSCA。其中,  $\beta$  是依赖于数据中噪音程度的、取值在  $[0, 0.5]$  上的数。

**定义 2** 给定 DTCVA  $(U, C \cup D, V, f)$ , 对任意  $B(B \subseteq C)$ , 相对于决策属性  $D$  的  $\beta$  正域  $POS^\beta_B(D)$  定义为

$$POS^\beta_B(D) = \bigcup_{x \in U/D} B^\beta_-(X) \quad (5)$$

**定义 3** 给定 DTCVA  $(U, C \cup D, V, f)$ , 对任意的  $B(B \subseteq C)$ ,  $B$  对  $D$  的  $\beta$  近似分类质量  $\gamma^\beta_B(D)$  定义为

$$\gamma^\beta_B(D) = \frac{|\bigcup_{x \in U/D} B^\beta_-(X)|}{|U|} \quad (6)$$

式(6)反映了应用知识  $B$  对论域  $U$  进行分类时, 在噪声程度为  $\beta$  的情况下, 能够确定分类于决策  $D$  的对象在  $U$  中所占的比例。

**定理 1** 设 DTCVA  $(U, C \cup D, V, f)$ , 对任意的  $M \subseteq N \subseteq C$  和  $X \subseteq U$ , VPRSCA 模型具有以下性质:

- 1) 对任意  $x(x \in U)$ , 则  $I_M(x, X) \leq I_N(x, X)$ ;
- 2)  $M^\beta_-(X) \subseteq N^\beta_-(X)$ ;
- 3)  $POS^\beta_M(D) \subseteq POS^\beta_N(D)$ ;
- 4)  $\gamma^\beta_M(D) \leq \gamma^\beta_N(D)$ 。

证明: 1): 由式(2), 对任意的  $x(x \in U)$ ,

$$I_B(x, X) = \inf_{y \notin X} (1 - \theta_B(x)(y)) \quad (7)$$

当  $M \subseteq N \subseteq C$  时, 由式(1), 有  $\theta_M(x)(y) \geq \theta_N(x)(y)$  成立, 于是由式(7),  $I_M(x, X) \leq I_N(x, X)$  成立。

2)-4): 结论易证, 在此不赘述。

定理 1 保证了 VPRSCA 模型中正域与属性集之间的单调关系, 这些性质在本文第 3 节属性约简方法中起了重要的作用。

## 2 连续值属性决策表中的属性重要性度量

针对连续值属性的特点, 我们基于马氏距离衡量各连续值属性对决策类的分离程度, 以此度量连续值属性的重要性。

设 DTCVA  $(U, C \cup D, V, f)$ , 设 DTCVA 有  $l$  个决策类, 进一步, 设  $D$  把 DTCVA 分成  $l$  个决策表  $S_i = (U_i, C \cup \{d_i\}, V, f_i), U = U_1 \cup U_2 \cup \dots \cup U_l, i = 1, 2, \dots, l$ , 设  $\{x_{i1}, x_{i2}, \dots, x_{i|U_i|}\}$  是  $U_i$  的对象集, 在属性子集  $B = \{a_1, a_2, \dots, a_m\}$  下, 定义 DTCVA 的协方差矩阵  $W$  为

$$W = \frac{1}{l} \sum_{i=1}^l \frac{1}{|U_i|} \sum_{k=1}^{|U_i|} [f_i(x_{ik}, B) - \mu_i] [f_i(x_{ik}, B) - \mu_i]^T \quad (8)$$

式中,  $f_i(x_{ik}, B)$  是一个信息函数值向量, 它指出了  $S_i$  中的对象  $x_{ik}$  在  $B$  上的取值, 即

$$f_i(x_{ik}, B) = [f_i(x_{ik}, a_1), f_i(x_{ik}, a_2), \dots, f_i(x_{ik}, a_m)]^T \quad (9)$$

$\mu_i$  是  $S_i$  上各对象分别在条件属性  $a_1, a_2, \dots, a_m$  上的均值向量, 即

$$\mu_i = \left[ \frac{1}{|U_i|} \sum_{k=1}^{|U_i|} f_i(x_{ik}, a_1), \frac{1}{|U_i|} \sum_{k=1}^{|U_i|} f_i(x_{ik}, a_2), \dots, \frac{1}{|U_i|} \sum_{k=1}^{|U_i|} f_i(x_{ik}, a_m) \right]^T \quad (10)$$

$T$  表示矩阵的转置。

根据协方差矩阵  $W$ , 可定义条件属性集  $B$  上决策类  $d_i$  与  $d_j$  之间的马氏距离  $\Phi_{ij}^B$  为

$$(\Phi_{ij}^B)^2 = [\mu_i - \mu_j]^T W^{-1} [\mu_i - \mu_j] \quad (11)$$

式中,  $W^{-1}$  是  $W$  的逆矩阵。设在决策属性  $D$  上,  $\Phi_{ij}^B$  的均值为  $m_B$ 。于是, 在决策表 DTCVA 上, 当  $l=2$  时, 属性集  $B$  对决策  $D$  的“散度”为

$$\delta_B(D) = \Phi_{ij}^B \quad (12)$$

当  $l > 2$  时, 属性集  $B$  对决策  $D$  的“散度”为

$$\delta_B(D) = \sqrt{\sum_{i=1}^l \sum_{j=i+1}^l (\Phi_{ij}^B - m_B)^2} \quad (13)$$

“散度”指出了在属性子集  $B$  上各决策类之间的分散程度。

**定义 4** 给定 DTCVA  $(U, C \cup D, V, f)$ , 属性子集  $B(B \subseteq C)$  在属性集  $C$  中相对于决策  $D$  的重要性  $SIG(B, C, D)$  定义为

$$SIG(B, C, D) = \delta_C(D) - \delta_{C \setminus B}(D) \quad (14)$$

式(14)指出了在属性全集  $C$  中, 去掉属性子集  $B$  后对各决策类之间“可分离性”的影响。

## 3 DTCVA 中的属性约简

**定义 5** 给定 DTCVA  $(U, C \cup D, V, f)$ , 条件属性  $C$  关于决策  $D$  的  $\beta$  约简定义为  $C$  的一个属性子集 REDU, 且满足

- 1)  $\gamma_C^\beta(D) = \gamma_{REDU}^\beta(D)$ ;
- 2) 若从 REDU 中再去掉任何一个属性, 1) 将不成立。

定义 5 保证了  $C$  的一个最小子集 REDU 与  $C$  的近似分

类质量相等。

下面给出属性约简算法 CVAR。算法从所有的条件属性 C 开始,以第 3 节定义的属性重要性为启发式信息,逐次去掉重要性小且分类质量较差的属性,从而得到属性约简结果。

**算法 1 启发式属性约简算法 CVAR**

输入: DTCVA=(U,CUD,V,f)及 $\beta$ ;

输出: DTCVA 的一个  $\beta$  约简 REDU。

Step1 按式(6)计算 DTCVA 中决策 D 相对条件属性集 C 的  $\beta$  近似分类质量  $\gamma_C^\beta(D)$ ;

Step2 按式(14)计算每个条件属性的重要性程度,并按重要性程度升序排列所有的条件属性,设排序后的结果为  $C = \{a_1, a_2, \dots, a_n\}$ ;

Step3 令 REDU=C;

Step4 依顺序对 C 中的每个属性  $a_i$  重复以下操作:

Step4.1 计算决策 D 相对于约简 REDU 在删掉  $a_i$  后的  $\beta$  近似分类质量  $\gamma_{REDU-\{a_i\}}^\beta(D)$ ;

Step4.2 如果  $\gamma_C^\beta(D) = \gamma_{REDU-\{a_i\}}^\beta(D)$  成立,则属性  $a_i$  可约, REDU=REDU- $\{a_i\}$ ; 否则,属性  $a_i$  不能被约简, REDU 保持不变。

Step5 输出 REDU,则 REDU 是 DTCVA 中的一个约简,算法结束。

**4 实验测试**

为了验证前述 VPRSCA 模型及 CVAR 算法的效果,我们在 UCI 机器学习数据库上取 4 个连续值属性数据集进行实验。数据集的基本情况如表 1 所列。实验分为 3 个部分,第 1 个部分是 CVAR 算法的实验结果与分析;第 2 个部分是 VPRSCA 模型与 VPRS 模型<sup>[7]</sup>约简结果的对比;第 3 个部分是 VPRSCA 模型与粗糙集 RS 约简结果的对比。

以下实验 1—3,我们从整个数据集 allDataSet 中随机抽取 50%的数据作为训练数据集 TrDataSet,用 allDataSet 作为

测试数据集。为了验证属性约简效果,本文给出了利用 SVM 分类器的识别准确率来验证属性约简效果的方法。在实验 1、实验 2、实验 3 中支持向量机的参数设置为 SVM Type: C\_SVC, Kernel Function: RBF, Multiclass Method: one-against-one。对表 3 中的各数据集分别进行实验,每组数据实验 3 次,总体分类准确率取 3 次实验结果的平均值。

表 1 实验数据集的基本特性

数据集序号	数据集	样本数	连续值属性数	决策类别数
1	Glass	214	9	6
2	Wine	178	13	3
3	Pima	768	8	2
4	Taylor	2048	10	2

**实验 1 CVAR 算法的实验结果及分析**

通过本文 CVAR 算法建立数据集上的属性约简,用支持向量机 SVM 对属性约简数据集进行分类,对其分类结果与属性全集在支持向量机 SVM 上的分类结果进行比较。实验步骤如下:

Step1 对训练集 TrDataSet,用 CVAR 算法生成约简数据集 rDataSet,用支持向量机 SVM 对 rDataSet 数据集进行训练,用 allDataSet 进行测试,输出识别率;

Step2 对训练集 TrDataSet,用支持向量机 SVM 对 TrDataSet 数据集进行训练,用 allDataSet 进行测试,输出识别率。

本实验 CVAR 算法参数的具体设置为  $\beta=0.25$  和  $\beta=0.4$ 。

实验 1 结果如表 2 所列。从表 2 可以看出,在给定的  $\beta$  域值下,Wine 与 Taylor 约简后的数据集比属性全集(未约简数据集)的测试准确率高,而 Pima 与 Glass 约简后的数据集比属性全集(未约简数据集)的测试准确率几乎一致,但条件属性的个数已经减少。实验结果说明,数据集的冗余属性不能增加对数据集的分类率,另一方面也说明了 CVAR 方法的有效性。

表 2 支持向量机 SVM 识别结果

数据集	条件属性集 C	实验次数	rDataSet+SVM						TrDataSet+SVM	
			$\beta=0.25$			$\beta=0.4$			测试准确率(%)	平均(%)
			约简	测试准确率(%)	平均(%)	约简	测试准确率(%)	平均(%)		
Pima	{a1, a2, a3, a4, a5, a6, a7, a8}	1	{a1, a2, a5, a6, a8}	82.6		{a1, a2, a6, a8}	82.7		82.6	
		2	{a1, a2, a5, a6, a8}	82.6	82.6	{a1, a2, a5, a6, a8}	82.6	82.6	82.6	82.6
		3	{a1, a2, a5, a6, a8}	82.6		{a1, a2, a5, a6, a8}	82.6		82.6	
Wine	{a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11, a12, a13}	1	{a2, a4, a11, a12, a13}	73.0		{a2, a4, a5, a7, a11, a12}	82.6		70.2	
		2	{a2, a5, a12, a13}	71.3	76.2	{a2, a4, a5, a7, a11, a12}	81.5	82.2	70.2	70.2
		3	{a1, a2, a5, a7, a10, a12}	84.3		{a2, a5, a7, a10, a12}	82.6		70.2	
Taylor	{a1, a2, a3, a4, a5, a6, a7, a8, a9, a10}	1	{a9}	99.8		{a9}	99.8		76.9	
		2	{a8}	99.5	99.7	{a8}	99.5	99.7	76.7	76.8
		3	{a9}	99.8		{a9}	99.8		76.9	
Glass	{a1, a2, a3, a4, a5, a6, a7, a8, a9}	1	{a2, a3, a4, a5, a6, a7, a8}	79.0		{a2, a3, a4, a5, a6, a7}	79.4		70.9	
		2	{a2, a3, a4, a5, a6, a7, a8}	79.0	79.1	{a2, a3, a4, a5, a6, a7, a9}	80.4	79.6	80.4	79.6
		3	{a2, a3, a4, a5, a6, a7, a8, a9}	79.4		{a2, a3, a4, a5, a6, a7, a9}	79.0		79.4	

**实验 2 VPRSCA 模型约简与 VPRS 模型约简的对比及分析**

在粗糙集理论各种扩展模型中,VPRS 模型已经取得了较成功的应用。实验 2 的目的是把本文的 VPRSCA 模型与 VPRS 模型在连续值决策表中的属性约简结果进行对比。实验步骤如下:

Step1 对训练集 TrDataSet,用 CVAR 算法生成随  $\beta$  变化的约简数据集 rDataSet,用 SVM 对 rDataSet 数据集进行训

练,用 allDataSet 进行测试,输出识别率;

Step2 对训练集 TrDataSet,用 ROSE 测试平台<sup>[8]</sup>中可变精度粗糙集模型生成随  $\beta$  变化的约简数据集 rsDataSet,用 SVM 对 rsDataSet 数据集进行训练,用 allDataSet 进行测试,输出识别率。实验中采用 ROSE 系统的 Local 离散化方法。属性约简采用 Manual Search 方法,如果训练集有多个约简,则任意选取一个属性个数少的约简。

实验 2 的结果如图 1、图 2 及表 3 所示,其中“\*”表示用

VPRS 模型得到的 SVM 识别结果。表 3 记录了 VPRSCA 模型中各数据集在不同噪音阈值上的 SVM 识别结果。图 1、图 2 反映了基于 VPRSCA 模型的属性约简与基于 VPRS 模型的属性约简测试准确率的变化趋势。从实验结果可以看出,当噪音值  $\beta$  较小时,Wine 数据集使用 VPRS 模型约简效果较好,Pima 与 Taylor 数据集使用 VPRS 模型与 VPRSCA 模型几乎一致;但当噪音值  $\beta$  较大时,基于 VPRSCA 模型的约简效果更好;另一方面,Glass 数据集在整个噪音域值  $\beta$  上,使用 VPRSCA 模型约简效果好,这说明本文建立的 VPRSCA 模型具有比 VPRS 模型更强的噪音处理能力。

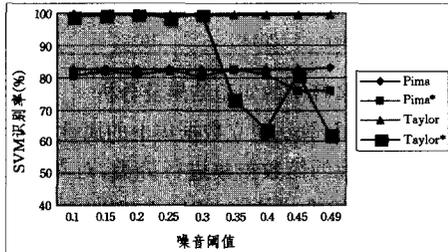


图 1 Pima, Taylor 数据集在 VPRSCA 及 VPRS 模型上约简的识别结果

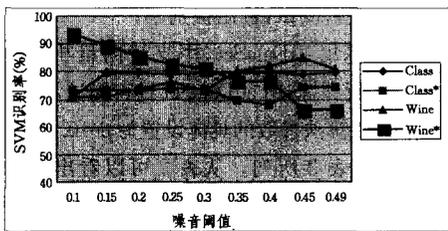


图 2 Glass, Wine 数据集在 VPRSCA 及 VPRS 模型上约简的识别结果

表 3 VPRSCA 模型约简的测试结果

Threshold $\beta$	0.1	0.15	0.2	0.25	0.30	0.35	0.4	0.45	0.49
Pima	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.8
Wine	71.1	71.0	73.6	76.2	74.3	80.3	82.2	84.6	80.9
Glass	70.9	79.7	79.7	79.1	79.6	79.6	79.6	79.3	79.6
Taylor	99.8	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7

实验 3 VPRSCA 模型约简与经典粗糙集模型 RS 约简的对比及分析

由于 RS 属性约简已有多种方法,实验 3 的目的是将 CVAR 方法与多种属性约简方法进行对比。实验步骤如下:

Step1 对表 3 中各训练集 TrDataSet,采用信息熵的离散化方法<sup>[9]</sup>对连续值属性做离散化处理。

Step2 用多种约简方法进行约简,生成约简数据集 rs-DataSet,用 SVM 对 rsDataSet 数据集进行训练,用 allDataSet 进行测试,输出识别率。

实验采用重庆邮电大学计算机与科学技术研究所开发的 RIDAS 测试平台来生成<sup>[10]</sup>。本文中 RS 模型下各种约简方法均来自文献<sup>[9]</sup>。

实验 3 的结果如表 4 所列。表 4 反映了各数据集在 RS

模型下各约简集上的 SVM 识别率。结合实验 2 中表 3 所列的结果,可以看出,只有 Wine 数据集在基于信息熵的约简、基于分明矩阵约简上的测试准确率比 CVAR 方法稍高,其它各数据集的测试准确率均低于 CVAR 约简方法上的测试准确率。这也说明了本文建立的 VPRSCA 模型的有效性以及对噪音的处理能力。

表 4 RS 模型下不同约简方法<sup>[9]</sup>的 SVM 测试结果

Dataset	基于信息熵约简	一般属性约简	特征选择约简	MIBARK 属性约简	基于分明矩阵约简
Pima	72.4	81.0	82.3	69.7	80.0
Wine	87.1	76.4	75.3	66.9	87.6
Glass	53.3	64.5	44.4	45.3	63.6
Taylor	79.6	99.5	99.5	99.5	79.6

结束语 本文提出了连续值属性决策表中的可变精度粗糙集模型 VPRSCA 和连续值属性决策表中属性约简方法 CVAR。该方法不需要事先对连续值属性做离散化处理,从而有效减少了连续值属性离散化的信息损失。UCI 数据集上的实验结果表明,与经典粗糙集模型及 VPRS 模型相比,该方法能够较好地适应连续值属性的决策表中的属性约简,并具有较好的性能。基于 VPRSCA 的高效约简算法是我们下一步研究工作的一项内容。

## 参考文献

- [1] 张雪英,刘凤玉, Krause J. 粗糙集分类算法的近似决策规则和规则匹配方法[J]. 计算机科学, 2005, 32(6): 129-132
- [2] 薄洪光,刘晓冰,马跃,等. 基于粗糙集的钢铁行业工艺知识发现方法[J]. 计算机集成制造系统, 2009, 15(1): 135-141
- [3] 徐非非,苗夺谦,魏莱,等. 基于互信息的模糊粗糙集属性约简[J]. 电子与信息学报, 2008, 30(6): 1372-1375
- [4] 肖迪,胡寿松. 实域粗糙集理论与属性约简[J]. 自动化学报, 2007, 33(3): 253-258
- [5] 刘文军,古云东,李洪兴. 一种基于 t 相似分类的连续值域决策表的决策算法[J]. 模式识别与人工智能, 2005, 18(6): 652-656
- [6] Pal S K. Case Generation Using Rough Sets with Fuzzy Representation[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(3): 292-230
- [7] Ziarko W. Variable Precision Rough Sets Model[J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59
- [8] Prędko, Słowiński R, Stefanowski J, et al. A Data Analysis and Decision Support System for Multiple Attributes Classification Problems Based on Rough Sets with Indiscernibility Relation [DB/OL]. <http://idss.cs.put.poznan.pl/site/rose.html>, 2009-01-06
- [9] 王国胤. 粗糙集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001
- [10] Wang G Y, Zheng Z, Zhang Y. RIDAS-A Rough Set Based Intelligent Data Analysis System[C]//Proceedings of ICMLC 2002. Beijing: IEEE Press, 2002: 646-649

(上接第 193 页)

- [2] Bauer C, King G. Hibernate in action[M]. Manning Publication Co., 2005
- [3] 孙卫琴. 精通 Hibernate Java 对象持久化技术详解[M]. 北京: 电子工业出版社, 2005
- [4] Hibernate Reference[EB/OL]. <http://www.hibernate.org/hib>

does/reference/htm I/. 2004-03-29

- [5] 孙卫民,曹正凤. Hibernate 对 Struts 框架的扩展研究[J]. 计算机工程与设计, 2008, 29(04): 858-861
- [6] 赵广利. 基于 NHibernate 的数据持久化方案[J]. 计算机工程, 2009, 35(20): 53-55
- [7] 顾春华,贾欣歌,顾兢. 一种动态对象/关系映射框架及其实现[J]. 华东理工大学学报: 自然科学版, 2009, 30(6): 882-885