区间值决策系统的局部属性约简

尹继亮 张 楠 赵立威 陈曼如
(烟台大学数据科学与智能技术山东省高校重点实验室 山东 烟台 264005)
(烟台大学计算机与控制工程学院 山东 烟台 264005)

摘 要 区间值决策系统中已有的属性约简工作主要针对决策属性中所有的决策类。针对区间值决策系统中决策属性的某些特定类,引入了区间值决策系统局部约简的概念,提出了部分决策类约简的判定定理;利用差别矩阵方法研究局部约简的结构,并给出了基于差别矩阵的局部约简算法。通过局部约简的概念对区间值决策系统的全局约简结构进行进一步刻画,讨论了不协调区间值决策系统的局部约简和全局约简之间的关系。最后通过相关实验验证了所提算法的可行性和有效性。

关键词 区间值决策系统,特定类,局部约简,差别矩阵,全局约简

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.07.031

Local Attribute Reduction in Interval-valued Decision Systems

YIN Ji-liang ZHANG Nan ZHAO Li-wei CHEN Man-ru
(Key Lab for Data Science and Intelligence Technology of Shandong Higher Education Institutes,
Yantai University,Yantai,Shandong 264005,China)
(School of Computer and Control Engineering,Yantai University,Yantai,Shandong 264005,China)

Abstract The existing attribute reduction in interval-valued decision system is mainly relative to all decision classes. For some special classes of decision attributes in interval-valued decision system,the concept of local reduction and the judgment theorem of partial decision classes were introduced in this paper. Besides,the structure of local reduction was studied by using the method of discernibility matrix,and the local reduction algorithm based on discernibility matrix was given. The structure of the global reduction in interval-valued decision system was further depicted through the concept of the local reduction,and the relationship between the local reduction and global reduction was discussed. Finally, related experiments were carried out. The experimental results show the feasibility and effectiveness of the proposed algorithm.

Keywords Interval-valued decision system,Specific classes,Local reduction,Discernibility matrix,Global reduction

1 引言

属性约简[1-4]又称特征选择,是粗糙集理论[1-8]的重要研究内容之一,其目的是保留使信息系统中某种分类特征不变的最小属性子集,从而删除与信息系统不相关的属性。目前,由于区间值数据在实际应用中广泛存在,诸多学者利用启发式方法和差别矩阵方法对区间值系统展开了大量研究工作。Leung等[9]通过区间数间的错误分类率来计算对象间的相容度,并讨论了区间值信息系统中的规则获取;Yang等[10]在区间值信息系统中引入了α优势关系的概念,并构造出相应的上、下近似约简方法;文献[11]提出了基于正域和互信息的启发式算法,为电力数据分析提供了新的方法;2009年,张楠等[12]针对对象的分类结果存在冗余度大、错误分类率高等问题,在区间值信息系统中引入了α极大相容类的概念,并提出了广义决策保持属性约简;刘鹏惠等[13]根据区间数之间的相似程度定义了一种具有变精度的相似关系,进而得到求取属性约简的具体方法;Du等[14]在区间值序信息系统下引入了优势关系的概念,提出了相应的近似分布约简方法;2016年,张楠等[15]在不协调区间值决策系统中引入了确定性规则的概念,并构造了相应的差别矩阵方法;2017年,Dai等[16]在不完备区间值决策系统中引入了α-弱相似关系的概念,设计出了一种度量区间数之间相似关系的新方法。

上述约简方法针对决策属性的全部决策类,可以看成是一种全局约简。但在实际应用中,人们往往更加关注影响具

到稿日期:2018-03-17 返修日期:2018-05-03 本文受国家自然科学基金项目(61403329,61572418,61702439,61572419,61502410),山东省自然科学基金项目(ZR2018BA004,ZR2016FM42),烟台大学研究生科技创新基金项目(YDZD1807)资助。

尹继亮(1994-),男,硕士生,主要研究方向为粗糙集、数据挖掘与机器学习,E-mail:yinjiliangyt@126.com;**张 楠**(1979-),男,博士,硕士生导师,CCF会员,主要研究方向为粗糙集、认知信息学和人工智能,E-mail:zhangnan0851@163.com(通信作者);**赵立威**(1996-),男,主要研究方向为粗糙集、数据挖掘,E-mail:zlwazj1996@163.com。

体决策类的关键条件属性。例如，在医学诊断的决策过程中，如果决策值包括患病和不患病，则与患病密切相关的特征总是比其他的特征更容易引起注意，选择并分析这些关键特征具有重要意义。近几年，针对部分决策类的特征选择得到了广泛关注。特定类的概念于 1999 年被首次提出，Baggenstoss 等[17]采用特定类的思想设计了概念分类器；2010 年，Chen 等[18]在模糊决策系统中引进了特定类约简的概念，并设计了一种基于差别矩阵的局部正域约简方法；Yao 等[19]在经典粗糙集中提出了特定类的约简方法，并讨论了特定类约简和全局约简之间的关系；2017 年，鞠恒荣等[20]认为局部约简和全局约简之间存在一种内在的序贯关系，并采用特定类的思想设计了一个序贯三支分类器；Qian 等[21]认为经典粗糙集存在三大问题，即大数据中缺乏类标记、计算约简的效率低下和约简过拟合，并提出用局部粗糙集模型来解决以上问题，同时设计了相应的属性约简方法；Liu 等[22]提出了基于单个决策类的正域、分布和 β-分布约简概念，并设计了相应的差别矩阵方法。

基于上述研究，本文提出区间值决策系统的局部约简概念，为区间值决策系统提供一种针对部分决策类求取所有属性约简的新方法。本文第 2 节介绍不协调区间值决策系统的一些基本概念以及区间值决策系统中的正域约简方法；第 3 节提出不协调区间值决策系统的局部约简算法；第 4 节比较全局约简和针对特定的类局部约简的约简结果和效率；最后对本文进行总结。

2　基本知识

2.1　区间值决策系统的粗糙近似

相容关系粗糙集模型是对粗糙集模型的泛化，用于处理具有区间值条件属性的信息系统。下面首先给出其相关概念和性质。

给定区间值信息系统[12] $IS=(U,AT,V,f)$，其中，$U=\{x_1,x_2,\cdots,x_{|U|}\}$是有限对象集合；$AT=\{a_1,a_2,\cdots,a_{|AT|}\}$是有限属性集合；$V=\bigcup_{a_k\in AT}V_{a_k}$ 是全体属性的值域，V_{a_k} 是属性 $a_k\in AT$ 的值域；$f:U\times AT\rightarrow V$ 是一个映射函数，论域 U 中的每一个对象 x_i 在属性 a_k 上对应一个区间属性值。

假设属性集合 $AT=C\cup D$，$C=\{a_1,a_2,\cdots,a_{|C|}\}$是条件属性集合，$D=\{d\}$是决策属性集合；$V=V_C\cup V_D$，$V_C$ 为条件属性值集合，V_D 为决策属性值集合；$f:U\times C\rightarrow V_C$ 为条件属性映射函数，$f:U\times D\rightarrow V_D$ 为决策属性映射函数。则该区间值信息系统被称为区间值决策系统 $DS=(U,C\cup D,V,f)$。

定义 1[15]　设 $\eta_1=[l_i^k,u_i^k]$和 $\eta_2=[l_j^k,u_j^k]$为任意两个区间值，则区间值的相关运算如下：

$$\eta_1\cap\eta_2=\begin{cases}0, & (u_i^k<l_j^k)\vee(u_j^k<l_i^k)\\ [\max(l_i^k,l_j^k),\min(u_i^k,u_j^k)], & \text{otherwise}\end{cases}\tag{1}$$

$$\eta_1\cup\eta_2=[\min(l_i^k,l_j^k),\max(u_i^k,u_j^k)]\tag{2}$$

式(1)和式(2)分别表示两个区间数间的交运算和并运算。

定义 2　设 $DS=(U,C\cup D,V,f)$为区间值决策系统，$\forall x_i,x_j\in U$，$a_k\in C$，区间值 $\eta_1=[l_i^k,u_i^k]$和 $\eta_2=[l_j^k,u_j^k]$的 Jaccard 相似率[15] α_{ij}^k 为：

$$\alpha_{ij}^k=\frac{|[l_i^k,u_i^k]\cap[l_j^k,u_j^k]|}{|[l_i^k,u_i^k]\cup[l_j^k,u_j^k]|}\tag{3}$$

Jaccard 相似率用于度量区间数之间的相似性。

定义 3[13]　对于区间值决策系统 $DS=(U,C\cup D,V,f)$，$0\leqslant\alpha\leqslant1$，定义关于条件属性集 C 的 α-相容关系为：

$$T_C^\alpha=\{(x_i,x_j)\mid(x_i,x_j)\in U\times U,\alpha_{ij}^k>\alpha,a_k\in C\}\tag{4}$$

关于对象 x_i 在属性集 C 下的 α-相容类为：

$$S_C^\alpha(x_i)=\{x_j\mid x_j\in U,(x_i.x_j)\in T_C^\alpha\}\tag{5}$$

对于 $\forall x_i\in U$，关于属性集 C 的相容类集合为：

$$S_C^\alpha(U)=\{S_C^\alpha(x_1),S_C^\alpha(x_2),\cdots,S_C^\alpha(x_{|U|})\}\tag{6}$$

性质 1[12]　设区间值决策系统 $DS=(U,C\cup D,V,f)$，则 α-相容类 T_C^α 具有以下性质。

1）自反性：对于 $\forall x_i\in U$，有$(x_i,x_i)\in T_C^\alpha$。

2）对称性：对于 $\forall x_i,x_j\in U$，若$(x_i,x_j)\in T_C^\alpha$，则$(x_j,x_i)\in T_C^\alpha$。

3）非传递性：对于 $\forall x_i,x_j,x_l\in U$，如果$(x_i,x_l)\in T_C^\alpha$ 和$(x_l,x_j)\in T_C^\alpha$ 成立，则$(x_i,x_j)\in T_C^\alpha$ 不一定成立。

定义 4　给定区间值决策系统 $DS=(U,C\cup D,V,f)$，$S_C^\alpha(x_i)$为对象 x_i 的相容类，则定义集合 $X\subseteq U$ 关于 α-相容关系的上近似和下近似[12]分别为：

$$\overline{apr_C^\alpha}(X)=\{x_i\mid x_i\in U,S_C^\alpha(x_i)\cap X\neq\emptyset\}\tag{7}$$

$$\underline{apr_C^\alpha}(X)=\{x_i\mid x_i\in U,S_C^\alpha(x_i)\subseteq X\}\tag{8}$$

$U/D=\{D_1,D_2,\cdots,D_{|U/D|}\}$是决策属性 D 对 U 的划分，对于 $\forall x_i\in U$，关于决策属性 D 的上近似和下近似分别为：

$$\overline{apr_C^\alpha}(D)=\{x_i\mid D_j\in U/D,S_C^\alpha(x_i)\cap D_j\neq\emptyset\}\tag{9}$$

$$\underline{apr_C^\alpha}(D)=\{x_i\mid D_j\in U/D,S_C^\alpha(x_i)\subseteq D_j\}\tag{10}$$

$POS_C^\alpha(D)=\bigcup_{D_j\in U/D}\underline{apr_C^\alpha}(D_j)$为 α-相容关系的正域。

2.2　区间值决策系统的正域约简

文献[15]提出了不协调区间决策系统的正域约简。

定义 5　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$A\subseteq C$，若属性子集 A 满足：

1）$POS_A^\alpha(D)=POS_C^\alpha(D)$；

2）$\forall B\in A$，满足 $POS_B^\alpha(D)\neq POS_C^\alpha(D)$。

则称属性子集 A 为区间值决策系统的正域约简。

定义 6　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$U=\{x_1,x_2,\cdots,x_{|U|}\}$，则对于 $1\leqslant\forall i,j\leqslant|U|$：

$$M_D^\alpha(i,j)=\begin{cases}\{a_k\mid a_k\in C\wedge\alpha_{ij}^k<\alpha\}, & condition1\\ \emptyset, & \text{otherwise}\end{cases}\tag{11}$$

为对象 x_i 和 x_j 的正域差别属性集。其中，$condition1=x_i$ or $x_j\in POS_C^\alpha(D)\wedge d(x_i)\neq d(x_j)$，它表示两个对象中至少有一个在正域，并且两个对象的决策值不相同。

定义区间值决策系统的正域保持差别矩阵如下：

$$DM_D^\alpha=\bigwedge_{1\leqslant\forall i,j\leqslant|U|}M_D^\alpha(i,j)\tag{12}$$

定义 7　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$C=\{a_1,a_2,\cdots,a_{|C|}\}$，$M_D^\alpha(i,j)$为对象 x_i 和 x_j 的正域差别属性

集，则定义正域差别函数为：

$$\Delta_D^\alpha = \bigwedge_{1\leqslant \forall i,j\leqslant |U|} \vee M_D^\alpha(i,j) \tag{13}$$

定理 1 给定区间值决策系统 $DS=(U,C\cup D,V,f)$，$A\subseteq C$，若 A 是正域保持约简，当且仅当 $\wedge A$ 是差别函数 Δ_D^α 的极小析取范式中的一个合取项。

基于差别矩阵的正域约简算法(Positive regions Reduction Algorithm based on Discernibility Matrix，PRADM)的具体描述如算法1所示。

算法1 PRADM

输入：区间值决策系统DS，阈值α

输出：区间值决策系统的正域保持全部属性约简

Step 1 计算区间值决策系统DS的在阈值α下的相容类集合 $S_C^\alpha(U)$；

Step 2 根据每个对象对应的相容类，计算关于属性C的正域 $POS_C^\alpha(D)$；

Step 3 根据每个对象的正域，构造正域约简差别矩阵 DM_D^α；

Step 4 由差别矩阵 DM_D^α 计算正域约简可辨识函数 Δ_D^α；

Step 5 利用分配率和吸收率计算出所有正域保持约简。

例1 以表1所列的区间值决策系统为例，其中，$U=\{x_1,x_2,\cdots,x_8\}$，$C=\{a_1,a_2,a_3,a_4\}$，$D=\{d\}$。令 $\alpha=0.5$，则相容关系 $T_C^{0.5}$ 为：

$$T_C^{0.5}=\begin{bmatrix}1&0&0&0&0&0&0&0\\0&1&0&1&0&0&0&1\\0&0&1&0&0&0&0&0\\0&1&0&1&0&0&0&1\\0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&1&0&1&0&0&0&1\end{bmatrix}$$

根据相似率布尔矩阵，计算相容类集合：

$S_C^{0.5}(U)=\{S_C^{0.5}(x_1),S_C^{0.5}(x_2),\cdots,S_C^{0.5}(x_8)\}$

其中，$S_C^{0.5}(x_1)=\{x_1\}$，$S_C^{0.5}(x_2)=S_C^{0.5}(x_4)=S_C^{0.5}(x_8)=\{x_2,x_4,x_8\}$，$S_C^{0.5}(x_3)=\{x_3\}$，$S_C^{0.5}(x_5)=\{x_5\}$，$S_C^{0.5}(x_6)=\{x_6\}$，$S_C^{0.5}(x_7)=\{x_7\}$。

$U/D=\{\{x_1,x_5,x_6\},\{x_2,x_3,x_4\},\{x_7,x_8\}\}$ 为决策属性 D 对 U 的划分，关于决策属性 D 的正域为：

$POS_C^{0.5}(D)=\{x_1,x_3,x_5,x_6,x_7\}$

计算正域保持约简差别矩阵：

$DM_D^{0.5}=$

$$\begin{bmatrix}\emptyset & C & \{a_1,a_2\} & \{a_2,a_4\} & \emptyset & \emptyset & \{a_2\} & \{a_3,a_4\}\\ & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_1,a_3\} & C-\{a_2\} & \emptyset\\ & & \emptyset & \emptyset & \{a_3\} & \{a_1,a_4\} & \{a_1\} & C-\{a_2\}\\ & & & \emptyset & \{a_3,a_4\} & \{a_1\} & \{a_4\} & \emptyset\\ & & & & \emptyset & \emptyset & \{a_1,a_3\} & \{a_3,a_4\}\\ & & & & & \emptyset & \{a_1,a_4\} & \{a_3\}\\ & & & & & & \emptyset & \emptyset\\ & & & & & & & \emptyset\end{bmatrix}$$

计算正域约简差别函数：

$\Delta_D^{0.5}=a_1\wedge a_3\wedge a_4$

因此，正域保持约简为：$\{a_1,a_2,a_3,a_4\}$。

表1 不协调区间值决策系统

Table 1 Inconsistent interval-valued decision systems

U	a_1	a_2	a_3	a_4	d
x_1	[0.86,3.60]	[1.20,2.23]	[0.26,2.26]	[0.19,2.20]	1
x_2	[0.12,2.50]	[0.79,3.20]	[1.73,3.26]	[0.73,3.26]	2
x_3	[0.13,2.20]	[0.86,2.95]	[0.26,2.26]	[0.24,2.12]	2
x_4	[0.14,3.25]	[0.85,3.01]	[0.71,3.11]	[1.59,3.21]	2
x_5	[0.50,2.45]	[1.34,2.94]	[1.80,2.23]	[0.25,2.30]	1
x_6	[1.13,2.13]	[0.82,3.10]	[0.24,2.19]	[1.26,3.11]	1
x_7	[1.47,3.13]	[0.82,3.10]	[0.24,2.19]	[0.24,2.11]	3
x_8	[0.98,2.63]	[0.89,2.69]	[1.10,3.22]	[1.73,3.35]	3

3 区间值决策系统的局部约简

本节针对部分决策类进行属性约简，并讨论了全局约简和局部约简之间的关系。

定义 8 给定区间值决策系统 $DS=(U,C\cup D,V,f)$，$U/D=\{D_1,D_2,\cdots,D_{|U/D|}\}$ 是决策属性 D 对 U 的划分，设 $K\subseteq U/D$，即 K 是部分决策类的集合，$x_i\in U$，则定义关于 $1\leqslant|K|\leqslant|U/D|$ 的上近似和下近似分别为：

$$\overline{apr_C^\alpha}(K)=\{x_i\mid D_j\in K,S_C^\alpha(x_i)\cap D_j\neq\emptyset\} \tag{14}$$

$$\underline{apr_C^\alpha}(K)=\{x_i\mid D_j\in K,S_C^\alpha(x_i)\subseteq D_j\} \tag{15}$$

关于 K 相对于属性集 C 的正域为：

$$POS_C^\alpha(K)=\bigcup_{D_j\in K}\underline{apr_C^\alpha}(D_j) \tag{16}$$

当 $|K|=1$ 时，$POS_C^\alpha(K)$ 表示针对单个决策类确定性规则的集合；当 $|K|=|U/D|$ 时，$POS_C^\alpha(K)$ 表示决策属性 D 的确定性规则集合。

性质 2 如果 $K=\{D_{k_1},D_{k_2}\}$，$1\leqslant k_1,k_2\leqslant|U/D|$，即 K 包含两个决策类，则 $POS_C^\alpha(K)=POS_C^\alpha(D_{k_1})\cup POS_C^\alpha(D_{k_2})$ 成立。

证明：由 $K=\{D_{k_1},D_{k_2}\}$ 可得，$D_{k_1}\cap D_{k_2}=\emptyset$，因此 $POS_C^\alpha(D_{k_1})\cap POS_C^\alpha(D_{k_2})=\emptyset$ 成立，从而 $POS_C^\alpha(K)=POS_C^\alpha(D_{k_1})\cup POS_C^\alpha(D_{k_2})$。证毕。

定义 9 给定区间值决策系统 $DS=(U,C\cup D,V,f)$，若 $POS_C^\alpha(K)=POS_{C-\{a\}}^\alpha(K)$，则称属性 a 在 C 中相对于 K 是不必要的，否则称属性 a 在 C 中相对于 K 是必要的。若 C 中的每一个属性都是必要的，则称 C 相对于 K 是独立的。如果 $A\subseteq C$ 相对于 K 独立且 $POS_C^\alpha(K)=POS_A^\alpha(K)$ 成立，则 $A\subseteq C$ 是 C 相对于 K 的局部正域约简。C 中所有相对于 K 是必要的元素集合为关于 α-相容关系的局部核心，记为 $Core_K^\alpha(C)$。局部核心 $Core_K^\alpha(C)$ 是所有局部正域约简的交集，即 $Core_K^\alpha(C)=\bigcap Red_K^\alpha(C)$。

性质 3 如果 $K=\{D_{k_1},D_{k_2}\}$，即 K 包含两个决策类，则 $Core_K^\alpha(C)=Core_{D_{k_1}}^\alpha(C)\cup Core_{D_{k_2}}^\alpha(C)$ 成立。

证明：由性质2得知，$POS_C^\alpha(K)=POS_C^\alpha(D_{k_1})\cup POS_C^\alpha(D_{k_2})$，因此有 $a\in Core_K^\alpha(C)$ 当且仅当 $a\in Core_{D_{k_1}}^\alpha(C)$ 或者 $a\in Core_{D_{k_2}}^\alpha(C)$，从而有 $Core_K^\alpha(C)=Core_{D_{k_1}}^\alpha(C)\cup Core_{D_{k_2}}^\alpha(C)$ 成立。证毕。

当 $|K|>2$ 时，定理2的结论仍然成立，局部核心中的每一个属性对于相应的单个决策类都是必要的，因此人们在关

注部分决策类时，局部约简可以得到针对部分决策类的关键条件属性。

定义 10　如果 $K=\{D_{k_1},D_{k_2}\}$，即 K 包含两个决策类，且 $POS_C^{\alpha}(K)=POS_B^{\alpha}(K)$，则称 B 是局部正域一致集；若 B 是局部正域一致集且 B 的任意真子集都不是局部正域一致集，则称 B 是局部正域约简。

定理 2　在区间值决策系统 $DS=(U,C\cup D,V,f)$ 中，$U/D=\{D_1,D_2,\cdots,D_{|U/D|}\}$，$K\subseteq U/D$，若 B 是局部正域一致集，则其充分必要条件为：对于 $\forall D_j\in K$，有 $POS_C^{\alpha}(K)=POS_B^{\alpha}(K)$。

局部正域约简是保持每一个决策类下的正域不变的最小属性子集，由于正域得到的是确定性规则，因此局部正域约简是保持所有确定性规则不变的最小属性子集。

定理 3　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$A\subseteq C$，$K\subseteq U/D$，则 A 是局部正域一致集，当且仅当 x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)=d(x_j)$ 时，有 $S_A^{\alpha}(x_i)\neq S_A^{\alpha}(x_j)$。

证明：($\Rightarrow$) $K\subseteq U/D$，由 A 是局部正域一致集知 $POS_C^{\alpha}(K)=POS_A^{\alpha}(K)$ 成立，因此对于 $\forall x_i\in U$，有 $S_A^{\alpha}(x_i)=S_C^{\alpha}(x_i)$。假设当 x_i or $x_j\in POS_C^{\alpha}(D)$ 时，$S_A^{\alpha}(x_i)=S_A^{\alpha}(x_j)$ 成立，则必存在 x_i,x_j 使得 $d(x_i)=d(x_j)$ 成立，这与 $d(x_i)\neq d(x_j)$ 矛盾。因此，由 A 是局部正域一致集可推出：当 x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $S_A^{\alpha}(x_j)$。

($\Leftarrow$) 假设 A 不是一个局部正域一致集，即 $POS_C^{\alpha}(K)\neq POS_A^{\alpha}(K)$ 成立，则存在 $x\in U$，使得 $S_A^{\alpha}(x)\neq S_C^{\alpha}(x)$ 成立。当 x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $S_A^{\alpha}(x_i)\neq S_A^{\alpha}(x_j)$。因此，对于 $\forall x\in U$，有 $S_A^{\alpha}(x)=S_C^{\alpha}(x)$，这与存在 $x\in U$ 使 $S_A^{\alpha}(x)\neq S_C^{\alpha}(x)$ 成立相矛盾。因此，当 x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $S_A^{\alpha}(x_i)\neq S_A^{\alpha}(x_j)$，可推出 A 是局部正域一致集。证毕。

定义 11　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$K\subseteq U/D$，$U=\{x_1,x_2,\cdots,x_{|U|}\}$，则对于 $1\leqslant\forall i,j\leqslant|U|$：

$$M_K^{\alpha}(i,j)=\begin{cases}\{a_k\mid a_k\in C\wedge\alpha_{ij}^k<\alpha\}, & condition1\\ \emptyset, & \text{otherwise}\end{cases}\tag{17}$$

为对象 x_i 和 x_j 的局部正域差别属性集。其中，$condition1=x_i$ or $x_j\in POS_C^{\alpha}(K)\wedge d(x_i)\neq d(x_j)$。$DM_K^{\alpha}=\bigwedge_{1\leqslant\forall i,j\leqslant|U|}M_K^{\alpha}(i,j)$ 为区间值决策系统的局部正域差别矩阵。

性质 4　$Core_K^{\alpha}(C)=\{a\in C:M_K^{\alpha}(i,j)=\{a\}\}$，其中，$1\leqslant i,j\leqslant|U|$。

性质 4 表明，差别属性集个数为 1 时，该差别属性集中的属性为核属性。这是因为必存在两个对象能够唯一被该属性区分。

定理 4　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$A\subseteq C$，$K\subseteq U/D$，则 A 是局部正域一致集，当且仅当 $\forall x_i,x_j\in U$，x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $M_K^{\alpha}(i,j)\cap A\neq\emptyset$。

证明：($\Rightarrow$) 假设 $\exists x_i,x_j\in U$，x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，满足 $M_K^{\alpha}(i,j)\cap A\neq\emptyset$。从而 $\exists a_k\in A$，满足 $\alpha_{ij}^k\geqslant\alpha$，因此 A 不是局部正域一致集，这与 A 是局部正域一致集矛盾。因此，由 A 是局部正域一致集可推出当 $\forall x_i,x_j\in U$，x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $M_K^{\alpha}(i,j)\cap A\neq\emptyset$。

($\Leftarrow$) 假设 A 不是局部正域一致集，则 $\exists a_k\in A$，满足 $\alpha_{ij}^k\geqslant\alpha$，因此当 $\forall x_i,x_j\in U$，x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $M_K^{\alpha}(i,j)\cap A=\emptyset$ 成立，这与 $M_K^{\alpha}(i,j)\cap A\neq\emptyset$ 矛盾，所以当 $\forall x_i,x_j\in U$，x_i or $x_j\in POS_C^{\alpha}(D)$，$d(x_i)\neq d(x_j)$ 时，有 $M_K^{\alpha}(i,j)\cap A=\emptyset$，可推出 A 是局部正域一致集。证毕。

定义 12　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$C=\{a_1,a_2,\cdots,a_{|C|}\}$，$K\subseteq U/D$，$M_K^{\alpha}(i,j)$ 为对象 x_i 和 x_j 的正域差别属性集，则定义局部正域差别函数为：

$$\Delta_K^{\alpha}=\bigwedge_{1\leqslant\forall i,j\leqslant|U|}\vee M_K^{\alpha}(i,j)\tag{18}$$

定理 5　设区间值决策系统 $DS=(U,C\cup D,V,f)$，$A\subseteq C$，$K\subseteq U/D$，若 A 是局部正域约简，当且仅当 $\wedge A$ 是差别函数 Δ_K^{α} 的极小析取范式中的一个合取项。

证明：($\Rightarrow$) 假设 $\wedge A$ 是差别函数 Δ_K^{α} 的极小析取范式中的一个合取项，则存在 $M_K^{\alpha}(i,j)\cap A=\emptyset$，由定理 4 可知 A 是局部正域约简。

($\Leftarrow$) 根据定义 12，若 $\wedge A$ 中去掉一个属性，记为 $\wedge A'$，则 $\exists x_i,x_j\in U$，使得 $M_K^{\alpha}(i,j)\cap A'=\emptyset$，因此 A' 不是最大分布约简，从而 A 是其中一个局部正域约简。证毕。

基于差别矩阵的局部正域约简算法(Positive regions Local Reduction Algorithm based on Discernibility Matrix, PRLADM)的具体描述如算法 2 所示。

算法 2　PRLADM

输入：区间值决策系统 DS，阈值 α，部分决策类 $\mathrm{K}\subseteq\mathrm{U/D}$，$1\leqslant|\mathrm{K}|\leqslant|\mathrm{U/D}|$

输出：区间值决策系统的局部正域保持全部属性约简

Step 1　计算区间值决策系统 DS 在阈值 α 下的相容类集合 $\mathrm{S_C^{\alpha}(U)}$；

Step 2　根据每个对象对应的相容类，计算局部正域 $\mathrm{POS_C^{\alpha}(K)}$；

Step 3　根据每个对象的正域，构造局部正域约简差别矩阵 $\mathrm{DM_K^{\alpha}}$；

Step 4　由可辨识矩阵 $\mathrm{DM_K^{\alpha}}$ 计算局部正域约简差别函数 $\Delta_{\mathrm{K}}^{\alpha}$；

Step 5　利用分配率和吸收率计算出所有局部正域约简。

例 2　对于表 1 所列的区间值决策系统，令 $\alpha=0.5$，$U/D=\{\{x_1,x_5,x_6\},\{x_2,x_3,x_4\},\{x_7,x_8\}\}$ 为决策属性 D 对 U 的划分。由例 1 可知 DS 的相似率布尔矩阵和相容类。计算部分决策类相应的正域及约简：

1) 对于决策类 $D_1=\{x_1,x_5,x_6\}$，正域为 $POS_C^{0.5}(D_1)=\{x_1,x_5,x_6\}$；差别矩阵为：

$$DM_{D_1}^{0.5}=$$

$$\begin{bmatrix}\emptyset & C & \{a_1,a_2\} & \{a_2,a_4\} & \emptyset & \emptyset & \{a_2\} & \{a_3,a_4\}\\ & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_1,a_3\} & \emptyset & \emptyset\\ & & \emptyset & \emptyset & \{a_3\} & \{a_1,a_4\} & \emptyset & \emptyset\\ & & & \emptyset & \{a_3,a_4\} & \{a_1\} & \emptyset & \emptyset\\ & & & & \emptyset & \emptyset & \{a_1,a_3\} & \{a_3,a_4\}\\ & & & & & \emptyset & \{a_1,a_4\} & \{a_3\}\\ & & & & & & \emptyset & \emptyset\\ & & & & & & & \emptyset\end{bmatrix}$$

约简为 $\{a_1,a_2,a_3\}$。

2) 对于决策类 $D_2=\{x_2,x_3,x_4\}$，正域为 $POS_C^{0.5}(D_2)=\{x_3\}$；差别矩阵为：

$$DM_{D_2}^{0.5}=\begin{bmatrix} \varnothing & \varnothing & \{a_1,a_2\} & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ & & \varnothing & \varnothing & \{a_3\} & \{a_1,a_4\} & \{a_1\} & \{a_1,a_3,a_4\} \\ & & & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ & & & & \varnothing & \varnothing & \varnothing & \varnothing \\ & & & & & \varnothing & \varnothing & \varnothing \\ & & & & & & \varnothing & \varnothing \\ & & & & & & & \varnothing \end{bmatrix}$$

约简为$\{a_1,a_3\}$。

3)对于决策类$D_3=\{x_7,x_8\}$,正域为$POS_C^{0.5}(D_3)=\{x_7\}$;差别矩阵为:

$$DM_{D_3}^{0.5}=\begin{bmatrix} \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_1\} & \varnothing \\ & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_2\} & \varnothing \\ & & \varnothing & \varnothing & \varnothing & \varnothing & \{a_4\} & \varnothing \\ & & & \varnothing & \varnothing & \varnothing & \{a_1,a_3,a_4\} & \varnothing \\ & & & & \varnothing & \varnothing & \{a_1,a_3\} & \varnothing \\ & & & & & \varnothing & \{a_1,a_4\} & \varnothing \\ & & & & & & \varnothing & \varnothing \\ & & & & & & & \varnothing \end{bmatrix}$$

约简为$\{a_1,a_2,a_4\}$。

4)对于决策类$K=\{D_1,D_2\}=\{\{x_1,x_5,x_6\},\{x_2,x_3,x_4\}\}$,正域为$POS_C^{0.5}(K)=\{x_1,x_3,x_5,x_6\}$;差别矩阵为:

$$DM_K^{0.5}=$$

$$\begin{bmatrix} \varnothing & C & \{a_1,a_2\} & \{a_2,a_4\} & \varnothing & \varnothing & \{a_2\} & \{a_3,a_4\} \\ & \varnothing & \varnothing & \varnothing & \{a_3\} & \{a_1,a_3\} & \varnothing & \varnothing \\ & & \varnothing & \varnothing & \{a_3\} & \{a_1,a_4\} & \{a_1\} & \{a_1,a_3,a_4\} \\ & & & \varnothing & \{a_3,a_4\} & \{a_1\} & \varnothing & \varnothing \\ & & & & \varnothing & \varnothing & \{a_1,a_3\} & \{a_3,a_4\} \\ & & & & & \varnothing & \{a_1,a_4\} & \{a_3\} \\ & & & & & & \varnothing & \varnothing \\ & & & & & & & \varnothing \end{bmatrix}$$

约简为$\{a_1,a_2,a_3\}$。

$$POS_C^{0.5}(D)=POS_C^{0.5}(D_1)\cup POS_C^{0.5}(D_2)\cup POS_C^{0.5}(D_3)$$

决策属性D的正域是所有决策类对应正域的并集,从而验证了性质2的正确性。

4 实验分析

本节对本文提出的局部约简算法进行实验验证。实验包括两部分:1)比较全局约简与局部约简的约简结果;2)比较全局约简与局部约简的效率。采用UCI标准数据集1)进行实验。实验环境为:PC机,Windows7 64b操作系统;4 GB DDR3内存;Intel Core i5-3470 CPU;C++编程语言。

实验选取8组标准UCI数据集,对于缺失数据,用对应属性下占多数的属性值进行替换,对名词性数据进行{0,1}替换,对连续型数据采用等频分割[19]的方法,所有数据的预处理均在WEKA3.6中进行。数据集信息如表2所列,其中,$|U|$表示对象数,$|C|$表示条件属性个数,$|D|$表示决策属性对对象分类的个数。

1) http://archive.ics.uci.edu/ml/datasets.html

表2 UCI数据集信息

Table 2 Information of UCI datasets

编号	数据集	$\|U\|$	$\|C\|$	$\|D\|$
1	BLOGER	100	5	2
2	Fertility	100	9	2
3	Iris	150	4	3
4	Seeds	210	7	3
5	Glass	214	9	6
6	Ecoli	336	7	8
7	Auto MPG	398	7	3
8	Mammographic Mass	961	5	2

需要将单值数据转换为区间值数据。单值数据转换为区间值数据的方法在文献[23]中已有描述,现取$\lambda=2.4$。

4.1 约简结果对比

本节主要对全局约简与局部约简的约简结果进行对比。下面分别取α为0.5,0.6,0.7,0.8,共进行8组实验,实验结果如表3-表6所列,其中"-"表示正域为空集。实验结果表明,局部约简与全局约简结果不同,大多数情况下局部约简是全局约简的子集,其约简长度要短于全局约简长度。在实际应用中可以只提取与部分决策类相关的条件属性,这不仅可以提升约简效率,而且能降低成本,因此研究基于不同决策类下的局部约简是必要的。其中,编号为1,2,8的数据集,类1和类2分别表示决策值为"0"和"1";编号为3,4,7的数据集,类1、类2和类3分别表示决策值为"1""2"和"3";编号为5的数据集,类1、类2和类3分别表示决策值为"3""7"和"6";编号为6的数据集,类1、类2和类3分别表示决策值为"7""4"和"6"。

表3 约简结果对比(α=0.5)

Table 3 Comparison of reduction results(α=0.5)

数据集编号	全局约简	局部约简		
		类1	类2	类3
1	{1,2,4,5}	{2,4,5}	{1,2,4,5}	-
2	集合1	集合2	{1,2,3,5,6,7,9}	-
3	{1,2,3,4}	{3},{4}	{1,3,4}	{1,2,3,4}
4	{1,2,3,5,6,7}	集合3	集合4	{2,3,5,6}
5	C	集合5	集合6	集合7
6	C	{3,4,7}	集合8	集合9
7	{1,3,7}	{1,3,7}	-	{1,3}
8	C	C	{1}	-

集合1=集合2={1,3,4,5,6,7,8,9};集合3={{2,3,5,6,7}{1,3,4,5,6,7}};集合4={{1,3,6,7},{1,5,6,7}};集合5={{1,3,8},{1,4,5,6,7,8}};集合6={{1,2,3,8},{1,3,4,8},{1,3,5,8},{1,3,7,8},{1,6,7,8},{1,4,5,7,8}};集合7={{1,2,3,8},{2,3,5,8},{2,3,6,8},{3,6,8,9},{3,4,5,6,8},{3,4,6,7,8},{4,5,6,7,8}};集合8={{1,3,4,6},{1,3,4,7}};集合9={{2,3,5,7},{1,2,5,6,7}}。

表4 约简结果对比(α=0.6)

Table 4 Comparison of reduction results(α=0.6)

数据集编号	全局约简	局部约简		
		类1	类2	类3
1	C	C	C	-
2	C	C	C	-
3	{1,2,3,4}	{3},{4}	{1,3,4}	{1,2,3,4}
4	C	集合10	{1,2,3,6,7}	{2,3,5,6,7}
5	集合11	集合12	集合13	集合14
6	{1,2,3,5,6,7}	集合15	集合16	{2,3,5,6,7}
7	C	C	C	C
8	C	C	C	-

集合 10={{1,3,4,5,6,7},{2,3,4,5,6,7}};集合 11=集合 12={1,2,3,4,5,6,8,9};集合 13={{1,2,3,8},{1,3,4,8},{1,3,5,8},{1,3,6,8},{1,3,7,8},{1,2,6,8}{1,4,6,8},{1,6,7,8},{1,4,5,8},{1,5,7,8},{1,3,4,6},{1,3,5,6},{1,2,3,6,7},{1,3,6,7,9}};集合 14={{2,3,8},{3,4,6,8},{3,6,8,9},{2,3,4,6},{2,3,4,9},{4,6,7,8},{1,3,4,5,6},{3,4,5,6,7},{1,2,3,5,6},{1,3,5,6,9},{1,2,3,5,9}};集合 15={{3,4,7},{2,3,4,6},{3,4,5,6},{3,5,6,7}};集合 16={{1,3,4,6},{1,3,4,7},{1,3,5,6,7}}。

表 5　约简结果对比(α=0.7)

Table 5　Comparison of reduction results(α=0.7)

数据集编号	全局约简	局部约简		
		类 1	类 2	类 3
1	C	C	C	—
2	{3}	{3}	{3}	—
3	{1,3,4}	{3},{4}	{1,3,4}	{1,3,4}
4	{2,3,4,5,6,7}	集合 17	集合 18	{2,3,5,6,7}
5	集合 19	集合 20	集合 21	集合 27
6	{1,2,3,5,6,7}	集合 22	集合 23	{2,5,7}
7	C	C	C	C
8	C	C	C	—

集合 17={{1,3,4,6,7},{2,3,4,6,7},{3,4,5,6,7}};集 18={{1,3,4,6,7},{2,3,4,6,7},{3,4,5,6,7},{2,4,5,6,7}};集合 19={{1,3,8},{1,2,4,6,8},{1,4,5,6,8},{1,4,6,7,8},{1,2,4,5,8},{1,2,5,7,8}};集合 20={{1,2,3},{1,3,6},{1,3,7},{1,3,4,8},{1,3,5,8},{1,2,6,8},{1,4,6,8},{1,5,6,8},{1,6,7,8},{1,4,5,8},{1,5,7,8}};集合 21={{2,3,8},{3,6,8},{2,3,9},{1,3,4,8},{3,4,5,8},{3,4,7,8},{1,2,3,6},{2,3,4,6},{1,3,5,6},{3,5,6,7},{2,4,7,8},{4,6,7,8},{1,2,4,8},{2,5,7,8},{4,5,6,7},{1,4,5,6,8},{1,3,4,5,9},{3,4,5,7,9},{2,4,5,7,9}};集合 22={{1,5},{5,7},{3,4,7},{3,6,7},{3,5,6},{1,3,4,6},{2,3,4,6}};集合 23={{3,5},{3,4,7},{3,6,7},{1,3,4,6},{2,3,4,6}}。

表 6　约简结果对比(α=0.8)

Table 6　Comparison of reduction results(α=0.8)

数据集编号	全局约简	局部约简		
		类 1	类 2	类 3
1	C	{1,2,3,4}	{1,2,3,4,5}	—
2	{3}	{3}	{3}	—
3	{1,3,4},{2,3,4}	{3},{4},{1,2}	集合 24	集合 25
4	{3,4,6,7}	集合 26	集合 27	{3,4,6,7}
5	C	集合 28	集合 29	集合 30
6	{1,2,3,5,6,7}	集合 31	集合 32	集合 33
7	C	C	C	{7}
8	C	C	C	—

集合 24={{3,4},{1,2,4}};集合 25={{1,3,4},{2,3,4}};集合 26={{1,3,6,7},{2,3,6,7},{3,4,6,7}};集合 27={{5,6,7},{1,3,6,7},{2,3,6,7},{3,4,6,7},{1,4,6,7},{2,4,6,7}};集合 28={{1,3,8},{3,4,8},{1,2,4,8},{1,2,5,8},{1,2,6,8},{1,2,7,8},{1,4,6,8},{2,4,6,8},{4,6,7,8},{1,5,6,8},{2,4,7,8},{4,5,6,8,9}};集合 29={{1,4,8},{2,4,8},{3,4,8},{4,5,8},{4,6,8},{4,7,8},{1,2,3},{1,3,4},{1,3,5},{1,3,6},{1,3,7},{1,3,8},{1,2,4},{1,4,6},{1,4,7},{1,2,8},{1,6,8},{1,2,5},{1,2,7},{1,5,6},{1,6,7},{1,5,7,8},{3,4,5,7},{4,5,6,7},{3,4,5,9}};集合 30={{2,3,5},{2,3,8},{2,3,9},{4,5,6},{5,6,7},{2,5,7},{3,6,8},{1,2,8},{1,3,4,8},{3,4,5,8},{3,4,7,8},{1,4,6,8},{2,4,6,8},{4,6,7,8},{1,3,4,5},{3,4,5,9},{1,2,3,4},{1,2,3,6},{1,2,4,7},{1,2,6,7},{1,2,7,9},{2,3,4,6},{2,4,6,7},{2,4,7,8},{2,4,7,9},{1,3,5,6},{1,2,5,6},{1,5,6,8}};集 31={{1},{2,7},{5,7},{3,4,7},{3,6,7},{2,3,6},{3,5,6}};集合 32={{1,6},{2,6},{3,6},{4,6},{5,6},{6,7},{1,7},{2,7},{4,7},{5,7},{2,3},{2,4},{2,5},{1,3},{3,5}};集合 33={{1,7},{2,5,7},{5,6,7}}。

4.2　约简效率对比

本节主要对比了全局约简算法与局部约简算法在对象数量增加与条件属性数量增加时的约简效率,结果如图 1—图 8 所示。本节选取 Mammographic Mass 和 Auto MPG 两个数据集,选取 Mammographic Mass 数据集对比全局约简算法与局部约简算法的耗时随对象数量增加的变化情况;选取 Auto MPG 数据集对比全局约简算法与局部约简算法的约简耗时随着条件属性数量增加的变化情况。

图 1—图 4 为 λ 取 2.4,α 分别取 0.5,0.6,0.7,0.8 时,全局约简和两个局部约简的耗时随对象增加的变化情况。其中,横坐标表示 Mammographic Mass 数据集的对象数量,纵坐标表示运行时间(单位为 ms),类 1 表示决策值“0”,类 2 表示决策值“1”。

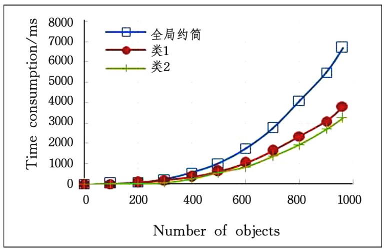

图 1　约简效率的对比(λ=2.4,α=0.5)

Fig. 1　Comparison of reduction efficiency(λ=2.4,α=0.5)

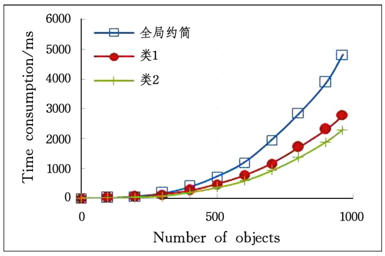

图 2　约简效率的对比(λ=2.4,α=0.6)

Fig. 2　Comparison of reduction efficiency(λ=2.4,α=0.6)

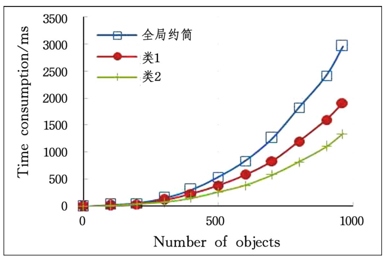

图 3　约简效率的对比(λ=2.4,α=0.7)

Fig. 3　Comparison of reduction efficiency(λ=2.4,α=0.7)

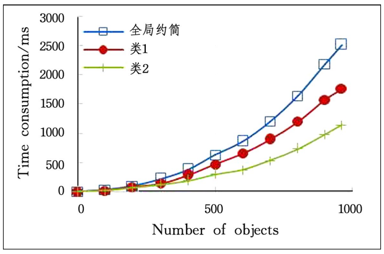

图 4 约简效率的对比(λ=2.4,α=0.8)

Fig. 4 Comparison of reduction efficiency(λ=2.4,α=0.8)

正方形点折线表示全局约简的运行时间随着对象增加的变化曲线,圆点折线和交叉点折线分别表示针对类 1 和类 2 的局部约简的耗时随着对象增加的变化曲线。实验结果表明,在对象数少于 600 时,由于算法之间的差别矩阵比较简单,全局约简与局部约简的运行时间几乎没有差距;但随着对象数量的增加,全局约简与局部约简的约简时间差异越来越明显。由于求取差别矩阵的时间复杂度为 $O(|C|\cdot|U|^2)$,此时局部约简的差别矩阵明显小于全局约简的差别矩阵,因此局部约简的运行速度比全局约简的速度快。

图 5—图 8 为 λ 取 3.5,α 分别取 0.5,0.6,0.7,0.8 时,各算法的耗时情况。其中,横坐标表示 Auto MPG 数据集的属性数量,纵坐标表示运行时间(单位为 ms),类 1 表示决策值为“1”,类 2 表示决策值为“2”,类 3 表示决策值为“3”。

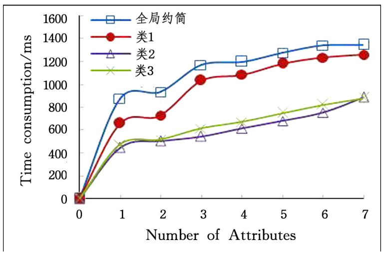

图 5 约简效率的对比(λ=3.5,α=0.5)

Fig. 5 Comparison of reduction efficiency(λ=3.5,α=0.5)

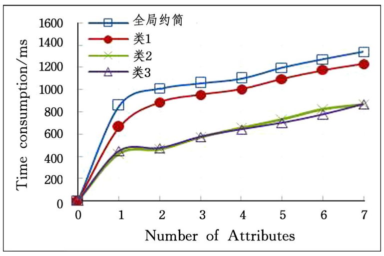

图 6 约简效率的对比(λ=3.5,α=0.6)

Fig. 6 Comparison of reduction efficiency(λ=3.5,α=0.6)

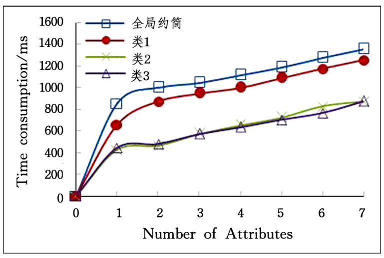

图 7 约简效率的对比(λ=3.5,α=0.7)

Fig. 7 Comparison of reduction efficiency(λ=3.5,α=0.7)

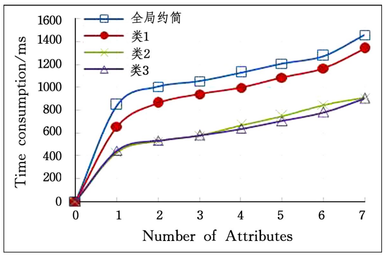

图 8 约简效率的对比(λ=3.5,α=0.8)

Fig. 8 Comparison of reduction efficiency(λ=3.5,α=0.8)

正方形点折线表示全局约简的运行时间随着属性增加的变化曲线,圆形点折线、交叉点折线和三角形点折线分别表示针对类 1、类 2 和类 3 的局部约简的耗时随着属性增加的变化曲线。实验结果表明,针对类 1 的局部约简和全局约简算法的约简耗时大致相同,而与类 2 和类 3 对应的局部约简耗时差距较大。这是由于针对类 1 的局部约简和全局约简算法对应的差别矩阵差距较小,而针对类 2 和类 3 的局部约简差别矩阵较复杂,因此局部约简的运行速度相对全局约简的速度更快。图 1—图 4 中的时间增速明显大于图 5—图 8 中折线的增速,这是由于求取差别矩阵的复杂度为 $O(|C|\cdot|U|^2)$,图 1—图 4 中的约简耗时随着对象的指数变化而增加,而图 5—图 8 中的约简耗时随着条件属性的线性变化而增加。

结束语 本文在相关研究工作的基础上,在区间值决策系统中引入了局部约简的概念,提出基于差别矩阵的局部正域约简,它保证了部分决策类的正域不变;同时讨论了局部约简算法和全局约简之间的关系。选取 8 组 UCI 标准数据集将局部约简和全局约简算法的约简结果和效率进行对比。实验结果表明,全局约简和局部约简在约简结果上没有明显联系,但局部约简的效率高于全局约简。由于本文提出的算法是基于差别矩阵的,其时间复杂度和空间复杂度较高,不利于在实际应用中推广,因此提出基于启发式的局部约简算法是未来的研究方向之一。

参 考 文 献

[1] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences,1982,11(5):341-356.

[2] PAWLAK Z. Rough sets:theoretical aspects of reasoning about data[M]. Boston:Kluwer Academic Publishers,1992.

[3] WANG G Y,YAO Y Y,YU H. A survey on rough set theory and applications[J]. Chinese Journal of Computers,2009,32(7):1229-1246.(in Chinese)
王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246.

[4] QIAN Y H,LIANG J Y,PEDRYCZ W,et al. Positive approximation:an accelerator for attribute reduction in rough set theory[J]. Artificial Intelligence,2010,174(9):597-618.

[5] CHEN H M,LIT R,RUAN D,et al. A rough-set based incrementalapproach for updating approximations under dynamic maintenance environments[J]. IEEE Transactions on Knowledge and

Data Engineering，2013，25(2)：274-284.

[6] HU Q H，YU D R，XIE Z X. Information-preserving hybrid data reduction based on fuzzy-rough techniques[J]. Pattern Recognition Letters，2006，27(5)：414-423.

[7] MIAO D Q，ZHAO Y，YAO Y Y，et al. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model[J]. Information Sciences，2009，179(24)：4140-4150.

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

[9] LEUNG Y，FISCHER M M，WU W Z，et al. A rough set approach for the discovery of classification rules in interval-valued information systems[J]. International Journal of Approximate Reasoning，2008，47(2)：233-246.

[10] YANG X B，QI Y，YU D J，et al. α-Dominance relation and rough sets in interval-valued information systems[J]. Information Sciences，2015，294(5)：334-347.

[11] XU F F，BI Z Q，LEI J S. Approximate reduction for the interval-valued decision table[M] // Rough Sets and Knowledge Technology. Springer International Publishing，2014：89-100.

[12] ZHANG N，MIAO D Q，YUE X D. Approach-es to knowledge reduction in interval-valued information systems[J]. Journal of Computer Research and Development，2010，47(8)：1362-1371. (in Chinese)
张楠，苗夺谦，岳晓冬. 区间值信息系统的知识约简[J]. 计算机研究与发展，2010，47(8)：1362-1371.

[13] LIU P H，CHEN Z C，QIN K Y. Decision attribute reduction of interval-valued informa-tion systems[J]. Computer Engineering and Applications，2009，45(28)：148-150. (in Chinese)
刘鹏惠，陈子春，秦克云. 区间值信息系统的决策属性约简[J]. 计算机工程应用，2009，45(28)：148-150.

[14] DU W S，HU B Q. Approximate distribution reducts in inconsistent interval-valued ordered decision tables[J]. Information Sciences，2014，271(7)：93-114.

[15] ZHANG N，XU X，TONG X R，et al. Knowledge reduction in inconsistent interval-valued decision systems[J]. Journal of Chinese Computer Systems，2017，38(7)：1585-1589. (in Chinese)
张楠，许鑫，童向荣，等. 不协调区间值决策系统的知识约简[J]. 小型微型计算机系统，2017，38(7)：1585-1589.

[16] DAI J H，WEI B，ZHANG X，et al. Uncertainty measurement for incomplete interval-valued information systems based on α-weak similarity[J]. Knowledge-Based Systems，2017，136(11)：159-171.

[17] BAGGENSTOSS P M. Class-specific feature sets in classification[J]. IEEE Transactions on Signal Processing，1999，47(12)：3428-3432.

[18] CHEN D G，ZHAO S Y. Local reduction of decision system with fuzzy rough sets[J]. Fuzzy Sets & Systems，2010，161(13)：1871-1883.

[19] YAO Y Y，ZHANG X Y. Class-specific attribute reducts in rough set theory[J]. Information Sciences，2017，418(38)：601-618.

[20] JU H R，LI H X，ZHOU X Z，et al. Spquential three-way classifier with local reduction[J]. Computer Science，2017，44(9)：34-39. (in Chinese)
鞠恒荣，李华雄，周献中，等. 基于 Local 约简的序贯三支分类器[J]. 计算机科学，2017，44(9)：34-39.

[21] QIAN Y H，LIANG X Y，LIANG J Y，et al. Local rough set：a solution to rough data analysis in big data[J]. International Journal of Approximate Reasoning，2018，97(6)：38-63.

[22] LIU G L，HUA Z，ZOU J Y. Local attribute reductions for decision tables[J]. Information Sciences，2017，422：204-217.

[23] ZHANG X，MEI C L，CHEN D G，et al. Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval-valued decision systems[J]. International Journal of Approximate Reasoning，2014，55(8)：1787-1804.

(上接第 142 页)

[12] CAO D，WANG X F，WANG F，et al. SA-IBE：A Secure and Accountable Identity-based Encryption Scheme[J]. Journal of Electronic & Information Technology，2011，33(12)：2922-2928. (in Chinese)
曹丹，王小峰，王飞，等. SA-IBE：一种安全可追责的基于身份加密方案[J]. 电子与信息学报，2011，33(12)：2922-2928.

[13] GOYAL V. Reducing trust in the PKG in identity based cryptosystem[C] // Advances in Crytology-CRYPTO 2007. Springer Berlin Heidelberg，2007：430-447.

[14] REN Y. Attribute-based Signature with Audita - biling in Standard Model[J]. Computer Science，2015，42(2)：142-146. (in Chinese)
任燕. 标准模型下可审计的基于属性的签名方案[J]. 计算机科学，2015，42(2)：142-146.

[15] LONG Y，XU X，CHEN K F. Two Identity Based Threshold Cryptosystem with Reduced Trust in PKG[J]. Journal of Computer and Development，2012，49(5)：932-938. (in Chinese)
龙宇，徐贤，陈克非. 两个降低 PKG 信任级的基于身份的门限密码体制[J]. 计算机研究与发展，2012，49(5)：932-938.

[16] FAN A W，YANG Z F，XIE L M，et al. Security analysis and improvement of strongly secure certificateless signature scheme[J]. Journal on Communications，2014，35(5)：118-123. (in Chinese)
樊爱宛，杨照峰，谢丽明，等. 强安全无证书签名方案的安全性分析和改进[J]. 通信学报，2014，35(5)：118-123.

[17] CAO X F，ZENG X W，KOU W D，et al. A novel anonymous authentication scheme over the insecure channel[J]. Journal of Xidian University，2007，34(6)：877-880，910. (in Chinese)
曹雪菲，曾兴雯，寇卫东，等. 一种新的不安全信道上的匿名认证方案[J]. 西安电子科技大学学报(自然科学版)，2007，34(6)：877-880，910.

[18] DAN B，MATT F. Identity-Based Encryptionfrom the Weil Pairing[J]. Lecture Notes in Computer Science，2001，2139(1)：213-229.

[19] NIST. Secure Hash Standard(SHS)[EB/OL]. http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf.