优势关系下不协调信息系统的正域约简*)

陈 娟[1] 王国胤[1] 胡 军[1,2]
(重庆邮电大学计算机科学与技术研究所 重庆 400065)[1]
(西安电子科技大学电子工程学院 西安 710071)[2]

摘 要 传统粗糙集对信息系统的处理是基于等价关系的,对于含有数值型的信息系统首先必须进行离散化,因此等价关系不利于处理连续值,也不能反映现实数据之间存在的序关系。本文基于优势关系在不协调信息系统中引入正域约简的概念,提出了优势关系下基于属性重要性的正域约简算法,为获取可信度较高的规则的循环正域约简算法。

关键词 不协调信息系统,正域约简,优势关系

Positive Domain Reduction Based on Dominance Relation in Inconsistent System

CHEN Juan[1] WANG Guo-Yin[1] HU Jun[1,2]
(Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065)[1]
(School of Electronic Engineering, XiDian University, Xi'an 710071)[2]

Abstract Research for information systems is based on equivalence relation in the classical rough set theory. Numerical System must be discretized. So, equivalence relation is not good at dealing with continuous attribute values, and couldn't reflect the order relations among real data. In this paper, we introduce a new concept of reduction named Positive Domain Reduction based on dominance relation, propose an algorithm for positive domain calculation based on condition attribute significance, and an algorithm for positive domain reduction to extract rules with high credibility.

Keywords Inconsistent information system, Positive domain reduction, Dominance relation

粗糙集理论[1]是近年来发展起来的一种软计算工具。知识约简是数据挖掘的一个重要课题,也是粗糙集理论的核心问题之一。众所周知,知识库中描述知识的属性并不是同等重要的,甚至其中有些是冗余的。所谓知识约简,就是在保持知识库分类能力不变的前提下,删除其中不必要的属性。通过知识约简,可以使知识表示简化,又不丢失基本信息,使人们能够更深入地理解知识并进行决策。

目前,许多学者对知识约简做了深入的研究,并取得了很多成果[2~4]。我们知道,在 Pawlak 近似空间下的信息系统,每个条件属性集和目标属性集决定了一个二元不可区分关系,即等价关系。要定义论域上的等价关系,属性值必须是离散的。但是,在实际问题中属性值更多的是连续的,其中部分具有偏序关系。传统粗糙集方法这类信息系统时先将其离散化[5,6],导致信息丢失,所以建立基于优势关系的信息系统有助于处理连续属性和偏序关系的问题。有不少学者对这一问题进行了大量的研究[7,8],在对基于优势关系的信息系统知识约简的研究已有相当的成绩[9]。本文在此基础上对基于优势关系的信息系统知识约简这一问题进行了更深入的探讨研究,给出了优势关系下不协调信息系统的正域约简的概念,提出了正域约简的算法,通过实验证明该算法在处理连续属性和偏序关系的问题是有效的。

1 基于优势关系的不协调信息系统

定义 1[5] 称一个四元组(U,R,V,f)为一个信息系统,其中,

U 是有限对象集,$U=\{x_1,x_2,\cdots,x_n\}$;

$R=C\cup D$ 是属性集合,子集 C 和 D 分别称为条件属性集和决策属性集;

V 是属性值的集合,$V=\cup V_r$,其中 $r\in R$,V_r 是属性 r 的有限值域;

$f:U\times R\to V$ 是一个信息函数,它指定 U 中每个对象 x 的属性值。

本文研究的信息系统是既有条件属性又有决策属性的一种信息系统。文[9]中建立的基于优势关系的信息系统是针对整个信息系统,即条件属性集与决策属性集都基于优势关系。本文讨论的信息系统只有条件属性集是基于优势关系,而决策属性是基于等价关系。

定义 2 设(U,R,V,f)为信息系统,对于 $B\subseteq C$,令 $R_B^{\preccurlyeq}=\{(x_i,x_j)\in U\times U: f_l(x_i)\preccurlyeq f_l(x_j),\forall a_l\in B\}$,$R_D=\{(x_i,x_j)\in U\times U: g_m(x_i)=g_m(x_j),\forall d_m\in D\}$。

其中 $R_B^{\preccurlyeq}$ 称为信息系统条件属性的优势关系,R_D 称为信息系统决策属性的等价关系。此时信息系统称为是基于优势关系的信息系统。记:

$$[x_i]_B^{\preccurlyeq}=\{x_j\in U:(x_i,x_j)\in R_B^{\preccurlyeq}\}=\{x_j\in U: f_l(x_i)\preccurlyeq f_l(x_j),\forall a_l\in B\}$$

$$[x_i]_D=\{x_j\in U:(x_i,x_j)\in R_D\}=\{x_j\in U: g_m(x_i)=g_m(x_j),\forall d_m\in D\}。$$

优势关系具有如下性质:

1)$R_C^{\preccurlyeq}$ 是自反的和传递的,未必是对称的。

*)本文获新世纪优秀人才支持计划(NCET)、重庆市自然科学基金(No. 2005BA2003)、重庆邮电大学自然科学基金(A2006-56)资助。陈 娟 硕士研究生,主要研究领域为智能信息系统;王国胤 博士,教授,博导,主要研究领域包括 Rough 集理论、神经网络、机器学习、数据挖掘等;胡 军 博士研究生,讲师,主要研究领域包括 Rough 集理论、粒计算等。

2)若 $B\subseteq C$,则 $R_C^{\preccurlyeq}\subseteq R_B^{\preccurlyeq}$。

3)若 $B\subseteq C$,则$[x_i]_C^{\preccurlyeq}\subseteq[x_i]_B^{\preccurlyeq}$。

4)若 $x_j\in[x_i]_C^{\preccurlyeq}$,则$[x_j]_C^{\preccurlyeq}\subseteq[x_i]_C^{\preccurlyeq}$,且$[x_i]_C^{\preccurlyeq}=\bigcup\{[x_j]_C^{\preccurlyeq}|x_j\in[x_i]_C^{\preccurlyeq}\}$成立。

5)$\bigcup\{[x_i]_C^{\preccurlyeq}|x\in U\}$是 U 的覆盖。

定义 3[9] 设 $S=(U,R,V,f)$为基于优势关系的目标信息系统,若 $R_C^{\preccurlyeq}\subseteq R_D$,则称该信息系统在优势关系下是协调的。否则,称该信息系统在优势关系下是不协调的。

本文研究的信息系统都是不协调信息系统。

定义 4[9] 对于任意的 $X\subseteq U$,定义 X 关于优势关系 $R_C^{\preccurlyeq}$ 的上、下近似为:

$\overline{R_C^{\preccurlyeq}}(X)=\{x_i\in U:[x_i]_C^{\preccurlyeq}\cap X\neq\varnothing\}$,

$\underline{R_C^{\preccurlyeq}}(X)=\{x_i\in U:[x_i]_C^{\preccurlyeq}\subseteq X\}$。

上下近似与 Pawlak 近似空间的性质相似。

2 基于优势关系的不协调信息系统的知识约简

2.1 知识约简

文[9]给出了下面几种基于优势关系的不协调信息系统知识约简的定义:

定义 5 设$(U,A\cup D,F,G)$是决策表,$B\subseteq A$,记:

$U/R_B^{\preccurlyeq}=\{[x_i]_B^{\preccurlyeq}|x_i\in U\}$,

$U/R_D=\{D_1,D_2,\cdots,D_r\}$,

$$\mu_B(x)=\left(\frac{|D_1\cap[x]_B^{\preccurlyeq}|}{|U|},\frac{|D_2\cap[x]_B^{\preccurlyeq}|}{|U|},\cdots,\frac{|D_r\cap[x]_B^{\preccurlyeq}|}{|U|}\right),$$

$$\gamma_B(x)=\max\left(\frac{|D_1\cap[x]_B^{\preccurlyeq}|}{|U|},\frac{|D_2\cap[x]_B^{\preccurlyeq}|}{|U|},\cdots,\frac{|D_r\cap[x]_B^{\preccurlyeq}|}{|U|}\right),$$

$\sigma_B(x)=\{D_j|D_j\cap[x]_B^{\preccurlyeq}\neq\varnothing,x\in U\}$。

称 $\mu_B(x)$为论域 U 上关于属性子集 B 的分布函数,称 $\gamma_B(x)$为论域 U 上关于属性子集 B 的最大分布函数,称 $\sigma_B(x)$为论域 U 上关于属性子集 B 的分配函数。

1) 若 $\forall x\in U,\mu_B(x)=\mu_C(x)$,则称 B 是分布协调集。若 B 是分布协调集,但 B 的任何真子集都不是分布协调集,则称 B 是分布约简。

2) 若 $\forall x\in U,\gamma_B(x)=\gamma_C(x)$,则称 B 是最大分布协调集。若 B 是最大分布协调集,但 B 的任何真子集都不是最大分布协调集,则称 B 是最大分布约简。

3) 若 $\forall x\in U,\sigma_B(x)=\sigma_C(x)$,则称 B 是分配协调集。若 B 是分配协调集,但 B 的任何真子集都不是分配协调集,则称 B 是分配约简。

分布协调集是保持对象在每个决策类的隶属程度不变的属性集,若 B 是(U,R,V,f)的分布约简,则由 B 产生的规则与由 C 产生的规则具有相同的可信度。

分配协调集是保持所有对象的可能决策类不变。

然而以上属性约简方法判断属性集可约简条件是严格的,大部分数据集很难达到约简的目的。在此基础上,本文提出一种新的知识约简方法,即基于优势关系的正域约简。

定义 6 条件属性集分类相对于决策属性集分类的正域为:

$$POS_C(D)=\bigcup_{X\in U/D}\underline{R_C^{\preccurlyeq}}(X)。$$

正域是可以完全确定地划分到某一决策类的所有元素集合。

对于 $B\subseteq C$,记 $POS_B(D)=\bigcup_{X\in U/D}\underline{R_B^{\preccurlyeq}}(X)$。若 $POS_B(D)=POS_C(D)$,则 B 为正域协调集。若 B 是正域协调集,但 B 的任何真子集都不是正域协调集,则称 B 是正域约简。

分配协调集是保持所有对象的可能决策类不变,而正域协调集实际上只保持原来具有确定决策为一个的不变即可,所以若 B 是分配协调集,则必是正域协调集。同理,若 B 是分布协调集,则必是正域协调集。反之,不一定成立。因此,可有如下定理:

定理 1 设(U,R,V,f)为基于优势关系的目标信息系统,则分配协调集必为正域协调集。

定理 2 设(U,R,V,f)为基于优势关系的目标信息系统,则分布协调集必为正域协调集。

2.2 正域约简的属性重要性算法及规则获取

通过在基于优势关系的信息系统中引入正域约简的概念,本文提出了基于属性重要性的正域约简算法,得到信息系统的正域约简。

定义 7 属性子集 $B'(B'\subset B\subseteq C)$的重要性表示为

$$SGF(B',B,D)=\frac{|POS_B(D)|-|POS_{B\backslash B'}(D)|}{|U|}。$$

由定义 7 可以知道,当属性子集 B'的重要性 $SGF(B',B,D)=0$ 时,该属性子集对于属性集 B 而言是不重要的,可以约简。否则,属性子集 B'相对于 B 而言是重要的,不可以约简。

若 $SGF(B',B,D)=0$,则属性子集 $B-B'$是正域协调集;若 $\forall b\in B-B'$,$SGF(\{b\},B-B',D)\neq0$,则属性子集 $B-B'$是正域约简。

例 1 表 1 给出了一个基于优势关系的目标信息系统。

表 1 一个目标信息系统

$U\times R$	a	b	c	d
x_1	2	3	5	2
x_2	5	1.4	4	3
x_3	6	2.6	6	3
x_4	4	0.8	3	2
x_5	1	1	1	1
x_6	3	2	2	1

由表 1 我们可以得到:

$[x_1]_C^{\preccurlyeq}=\{x_1\}$,　$[x_2]_C^{\preccurlyeq}=\{x_2,x_3\}$,

$[x_3]_C^{\preccurlyeq}=\{x_3\}$,　$[x_4]_C^{\preccurlyeq}=\{x_2,x_3,x_4\}$,

$[x_5]_C^{\preccurlyeq}=\{x_1,x_2,x_3,x_5,x_6\}$,　$[x_6]_C^{\preccurlyeq}=\{x_3,x_6\}$,

$D_1=\{x_1,x_4\}$,$D_2=\{x_2,x_3\}$,$D_3=\{x_5,x_6\}$。

根据定义 6,$POS_C(D)=\{x_1,x_2,x_3\}$。

$\{[x_1]_C^{\preccurlyeq},[x_2]_C^{\preccurlyeq},[x_3]_C^{\preccurlyeq},[x_4]_C^{\preccurlyeq},[x_5]_C^{\preccurlyeq},[x_6]_C^{\preccurlyeq}\}$是 U 的覆盖,$\{D_1,D_2,D_3\}$是 U 的划分。经过计算,只有属性子集$\{c\}$的重要性为 0,$\{a,b\}$是正域协调集,且$\{a\}$和$\{b\}$都不是正域协调集,所以$\{a,b\}$是正域约简。约简后得到表 2。

因为信息系统的不协调性,有 $POS_C(D)\subset U$,对于 $\forall x\in POS_C(D)$,$\forall y\in U$ 都与 x 不矛盾,故由 x 获取的规则是可靠的;对于 $\forall y\in(U-POS_C(D))$,$\exists x\in POS_C(D)$,y 与 x 矛盾,故由 y 获取的规则是不可靠的。如表 2 中,非正域元组 x_4 与正域元组 x_2、x_3 矛盾,所以由 x_4 获取的规则不可取。

表 2 约简后的目标信息系统

$U\times R$	a	b	d
x_1	2	3	2
x_2	5	1.4	3
x_3	6	2.6	3
x_4	4	0.8	2
x_5	1	1	1
x_6	3	2	1

通常规则的可靠性可以用可信度来衡量。

定义 8 规则的可信度定义为 $\gamma=\frac{|X\cap Y|}{|X|}$，其中，$X=\{x|x\in U\wedge C_x\}$，$Y=\{y|y\in U\wedge D_y\}$。

正域中的元组获取规则的 $\gamma=1$。

从表 2 可得到如下规则：

$R_1:(a\geqslant 2)\wedge(b\geqslant 3)\rightarrow(d=2)$

$R_2:(a\geqslant 5)\wedge(b\geqslant 1.4)\rightarrow(d=3)$

$R_3:(a\geqslant 6)\wedge(b\geqslant 2.6)\rightarrow(d=3)$

R_2 与 R_3 是传递的，R_3 包含于 R_2 中，这是因为 $[x_2]_C^{\leqslant}\subset[x_3]_C^{\leqslant}\subseteq D_2$，易知规则 R_3 是多余的。通过正域约简得到的规则虽然有高可信度，但由于属性值的序关系，正域中元组的属性值往往较大，某些规则的适用性偏低。规则的适用性可以用覆盖率来衡量。

定义 9 规则的覆盖率定义为 $\beta=|X|/|U|$，其中 $X=\{x|x\in U\wedge C_x\}$。

R_1，R_2 的覆盖率分别是 1/6，2/6，且是对正域的覆盖，无法表达非正域的信息。改进的方法是去掉已获取规则的元组，对剩下的元组再次求取正域并约简，获取规则，直到所有元组都能成为正域中的元组为止。这是一个将不协调信息系统协调化的过程。

表 2 经过上述操作后得到表 3。

表 3 新信息系统

$U\times R$	a	b	c	d
x_4	4	0.8	3	2
x_5	1	1	1	1
x_6	3	2	2	1

$[x_4]_A^{\leqslant}=\{x_4\}$，$[x_5]_A^{\leqslant}=\{x_5,x_6\}$，$[x_6]_A^{\leqslant}=\{x_6\}$，$D_1'=\{x_4\}$，$D_2'=\{x_5,x_6\}$，

$POS_C(D)'=\{x_4,x_5,x_6\}$，

$[x_6]_C^{\leqslant}\subset[x_5]_C^{\leqslant}\subseteq D_2'$。

经计算，$\{b,c\}$是其正域约简。设置一个参数 α 来区别每次获取的规则，每循环一次 α 加 1。于是每次获取规则的可信度 $\gamma=1$，在规则后加上参数对 (α,β)，这样得到的全部规则为：

$R'_1:(a\geqslant 2)\wedge(b\geqslant 3)\rightarrow(d=2)|(\alpha=1,\beta=1/6)$

$R'_2:(a\geqslant 5)\wedge(b\geqslant 1.4)\rightarrow(d=3)|(\alpha=1,\beta=2/6)$

$R_4:(b\geqslant 0.8)\wedge(c\geqslant 3)\rightarrow(d=2)|(\alpha=2,\beta=1/6)$

$R_5:(b\geqslant 1)\wedge(c\geqslant 1)\rightarrow(d=1)|(\alpha=2,\beta=2/6)$

获取规则的流程图如图 1 所示。

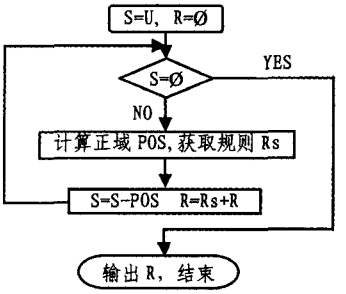

图 1 流程图

算法结束时，所有元组最终都会被纳入正域，这样就会使得到的规则有较强的泛化性及高识别率。样本识别时从 $\alpha=1$ 的规则开始，若在此 α 值的所有规则中能找到样本条件属性满足的规则，则不需再找 α 值更大的规则；若在同一 α 值规则中有多条决策属性值不同的规则均能使样本条件属性满足，则选择 β 最大的规则。

3 实验分析

文[9]在介绍分布约简与分配约简的同时，提出了实现分配约简的矩阵算法，但因为分配约简是个较严格的约简，在实际问题中较难达到约简的目的。本文旨在说明优势关系处理连续属性问题是有意义的，用正域约简方法对信息系统进行约简并获取规则是可行的，对优势关系下的不协调信息系统的深入研究是必要的。

我们在 VC6 上实现了优势关系下目标信息系统正域约简算法（以下简称 PDRIS），并选用 UCI 机器学习数据库（ftp://ftp.ics.uci.edu/pub/machine-learning-databases/）中的 4 个数据集（Ecoli，Glass，Iris，Pima-diabetes）进行实验，这 4 个数据集的条件属性都是连续属性，属性值是偏序关系；决策属性表示分类，都是离散属性，正是本文的研究对象。

本算法对每个数据集进行 5 次实验，随机取 50%作为训练集，其余样本作为测试集，对 5 次实验求平均值作为实验结果。实验对比对象是文[13]中采用的基于信息熵的离散化算法将连续属性离散化，然后进行属性约简而得到推理规则，最后采用获得的规则对测试集进行测试（以下简称 A1）。结果如表 4 所示。

表 4 实验对比结果

数据集	N#	M#	PDRIS				A1			
			识别率(%)	正确识别率(%)	误识率(%)	拒识率(%)	识别率(%)	正确识别率(%)	误识率(%)	拒识率(%)
Ecoli	336	8	95.6	73.9	21.7	4.4	93.5	70.3	23.2	6.5
Glass	214	7	91.2	63.4	27.8	8.8	94.4	64.4	30.0	5.6
Iris	150	3	99.7	96.3	3.4	0.3	100.0	96.0	4.0	0.0
Pima-diabetes	768	2	94.2	70.9	23.3	5.8	92.0	63.1	28.9	8.0

注：N#—元组数；M#—决策分类数；算法 A1 的数据来源于文[13]。

（下转第 227 页）

表 2 基于粒度计算的覆盖算法与领域覆盖算法以及 SVM 算法的实验结果比较

	学习样本数	测试样本数	覆盖数	正确识别个数	拒识个数	正确识别率(%)	拒识率(%)
覆盖算法	2521	278	27	262	0	94.24	0
基于粒度计算的覆盖算法(较细粒度)	2521	278	24	273	0	98.20	0
基于粒度计算的覆盖算法(较粗粒度)	24	278	14	223	50	80.22%	17.99
SVM 算法	2521	278	/	257	/	92.45%	/

继续将覆盖的圆心作为粗粒度下的样本进行学习,虽然正确识别率有所下降,拒识率有所上升,但这是粒度变粗的必然结果,论域变粗的过程中会丢失一些信息,但是其覆盖数大大减少,计算复杂性大大降低。是否将粒度继续变粗,则要根据实际需要进行。

结论 本文提出粒度计算与覆盖算法想结合的思想,并将其应用于实际数据中,得到理想的实验结果。指出在分析研究某一论域时,可以适当将论域粗化,在粗粒度世界抓住事物主要特征进行讨论,并且对某些在一个粗粒度世界无法识别或识别很弱的特征可以通过不同侧面对其进行分析,甚至可以进一步将讨论的粒度世界粗化,从而全面了解原事物,最终获得论域完整的分类结果。我们目前将不同粒度世界看作同样重要,测试样本时,只是进行一般的投票。事实上,人类在考虑问题时,对问题的不同粒度的分析是有主次之分的,因此,在计算机进行粒度计算时亦可考虑给不同粒度世界赋上权值,对由于不同粒度世界而得到的不同结果可以采取加权的形式处理。这种加权的思想,也正是概率逻辑神经网络的思想。此项工作正在进一步研究中。

参 考 文 献

1 Vapnik V N. Statistical Learning Theory [M]. INC. New York: John Wiley & Sons, 1998
2 边肇祺,张学工. 模式识别[M]. 北京:清华大学出版社,2001
3 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data[M]. Kluwer Academic Publishers, Dordrecht, 1991
4 张燕平. 机器学习中特征提取的新方法—重复覆盖算法[J]. 安徽大学学报,2002(3)
5 Zadeh L A. A Theory of Approximate Reasoning [M]. New York: Machine Intelligence, 1979. 149～194
6 Zadeh L A. Fuzzy Logic = Computing With Words [J]. IEEE Transactions on Fuzzy Systems, 1996(4):103～111
7 Zadeh L A. Towards Theory of Tuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic [J]. Fuzzy Sets and Syatema, 1997(19):111～127
8 Zadeh L A. Announcement of GrC, 1997. http://www.cs.uregina.ca/～yyao/GrC/
9 Pawlak Z. Rough Sets Theoretical Aspects of Reasoning about Data [J]. Dordrecht, Boston, London: Kluwer Academic Publishers, 1991
10 张铃,张钹. 问题求解的理论及应用[M]. 北京:清华大学出版社,1990
11 Zhang Ling. Granular Computing based on Rough Sets, Quotient Space Theory, and Belief Functions [J]. Lecture Notes in Computer Science, 2003, 2871:152～159
12 张铃,张钹. 多层前向网络的交叉覆盖设计算法[J]. 软件学报,1999,10(7):737～742

(上接第 218 页)

由对比结果可以看出,基于优势关系的信息系统正域约简在保持数据原有信息和基础上,可以有效地进行规则获取,对于处理具有序关系的数据集,两种算法的样本识别率大致相同,总体上 PDRIS 算法对未知样本的误识率比算法 A1 低,正确识别率基本比算法 A1 高,这说明用优势关系处理序信息系统比先将信息系统离散化后用等价关系来处理能得到更好的结果。

另外,基于优势关系对序信息系统获取的规则更符合人类对现实世界的认识。如 Iris(鸢尾属植物)标准数据集,每个元组包含 Iris 花的 4 个属性(萼片的长、宽,花瓣的长、宽),共有 3 种不同的花(Iris Setosa, Iris Versicolour, Iris Virginica)。该数据集获取的其中三条规则为:

$b \geqslant 3.5 \wedge c \geqslant 1.3 \wedge d \geqslant 0.2 \rightarrow d=0(\alpha=1,\beta=10/75)$,

$b \geqslant 2.7 \wedge c \geqslant 4.9 \wedge d \geqslant 1.8 \rightarrow e=2(\alpha=1,\beta=22/75)$,

$a \geqslant 5.0 \wedge b \geqslant 2.3 \wedge c \geqslant 3.3 \rightarrow d=1(\alpha=2,\beta=21/75)$。

即当萼片宽大于等于 3.5cm,花瓣长大于等于 1.3cm,花瓣宽大于等于 0.2cm 时该花是 Iris Setosa,且在该测试集中满足这种特性的元组有 10 个,另两条规则的含义类同。

结论 传统的粗糙集规则获取方法在处理连续属性时都先对连续属性进行离散化等处理,导致数据信息部分丢失。要想从基于优势关系的不协调信息系统中获取简洁的知识就必须对系统进行知识约简。本文提出的基于优势关系的不协调信息系统正域约简并获取规则的方法有效解决了连续属性和偏序关系问题,有一定的实用价值。然而基于属性重要性的正域约简算法在执行效率上有待改善,寻找更高效的属性约简方法是我们接下来的研究内容。

参 考 文 献

1 Pawlak Z. Rough Sets. International Journal of Computer and Information Sciences, 1982, 11:314～356
2 Beynon M. Reducts within the Variable Precision Rough Set Model: a further Investigation. European Journal of Operational research, 2001, 134:592～605
3 Pawlak Z. Rough Sets—Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Publishers
4 Wu W Z, Zhang M, Li H Z, Mi J S. Knowledge Reduction in Random Information Systems via Dempster-Shafer Theory of Evidence. Information Sciences, 2005, 174(3-4): 143～164
5 王国胤. Rough 集理论与知识获取. 西安:西安交通大学出版社, 2001
6 Stefanowski J. Handling Continuous Attributes in Discovery of Strong Decision Rules. RSCTC'98, 1998. 394～401
7 Sai Y, Yao Y Y, Zhang N. Data Analysis and Mining in Order Information Table. 2001 IEEE Int. Conf. on Data Mining. IEEE Computer Society Press, 2001. 497～504
8 Shao M W, Zhang W X. Dominance Relation and Rules in an Incomplete Ordered Information System. International Journal of Intelligent Systems, 2005, 20:13～27
9 张文修, 姚一豫, 梁怡. 粗糙集与概念格. 西安:西安交通大学出版社, 2006
10 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简. 计算机学报, 2003, 26(1):12～18
11 徐伟华, 张文修. 基于优势关系下的协调近似空间. 计算机科学, 2005, 32(9):164～166
12 邵明文, 张红英. 序信息系统上的优势关系与规则获取. 工程数学学报,2005, 22(4):697～702
13 谢宏, 程浩忠, 牛东晓. 基于信息熵的粗糙集连续属性离散化算法. 计算机学报, 2005, 28(9):1570～1574