

# 最小相关性最大依赖度属性约简

翟俊海 万丽艳 王熙照

(河北大学数学与计算机学院 河北省机器学习与计算智能重点实验室 保定 071002)

**摘要** 在经典粗糙集中,基于重要度的决策表属性约简算法只考虑了决策属性与条件属性之间的依赖度,没有考虑约简中条件属性之间的相关性,由此求出的约简中可能依然包含冗余属性。针对这一问题,提出了一种改进算法,它利用最小相关性和最大依赖度准则求决策表属性约简。与基于重要度的决策表属性约简算法相比,本算法求出的约简包含的属性个数少、冗余小。实验结果显示,本算法优于基于重要度的决策表属性约简算法。

**关键词** 粗糙集,决策表,属性约简,最小相关性,最大依赖度

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.12.031

## Attribute Reduction with Principle of Minimum Correlation and Maximum Dependency

ZHAI Jun-hai WAN Li-yan WANG Xi-zhao

(Hebei Province Key Lab of Machine Learning and Computational Intelligence, College of Mathematics and Computer Science, Hebei University, Baoding 071002, China)

**Abstract** In the classical rough set, the reduction algorithm based on significance for decision table only considers the dependency of decision attribute and condition attribute, and does not consider the correlation between the condition attributes in reduct. The reduct calculated with this kind of algorithm may include redundant attributes. In order to deal with this problem, an improved algorithm was proposed in this paper, which calculates the reduct with the principle of minimum correlation and maximum dependency. Compared with the reduction algorithm based on significance for decision table, less attributes are remained in the reducts calculated with the proposed algorithm, and the redundancy of the reduct is smaller. The experimental results show that the proposed algorithm outperforms the reduction algorithm based on significance for decision table.

**Keywords** Rough sets, Decision table, Attribute reduct, Minimum correlation, Maximum dependency

## 1 引言

粗糙集<sup>[1]</sup>作为一种新的处理不精确、不确定和不完备数据的数学工具,已被广泛应用于模式识别、数据挖掘和决策分析等领域<sup>[2]</sup>。粗糙集的主要应用是属性约简,本文在决策表框架下讨论属性约简。决策表属性约简算法大致可分为 3 类:基于差别矩阵的属性约简算法<sup>[3]</sup>、基于信息熵的属性约简算法<sup>[4]</sup>和基于重要度的属性约简算法<sup>[1]</sup>。基于差别矩阵的属性约简算法是 1992 年由波兰科学家 A. Skowron 提出的,他创造性地定义了决策表的差别矩阵和差别函数,利用差别矩阵来表示知识,利用差别函数来求决策表属性约简。差别矩阵  $C_{n \times n}$  中的元素  $c_{ij}$  表示能区分论域中对象(样例)  $x_i$  和  $x_j$  的属性子集,其中  $n$  表示对象(样例)的个数,差别矩阵中所有单属性子集的并集构成了决策表的核。差别函数是由  $c_{ij}$  中的元素求析取,不同的  $c_{ij}$  之间求合取得的关于属性的函数,它是一个析取合取范式,将其等价变换为合取析取范式,每一个析取子式就是一个属性约简。基于差别矩阵的属性约简算

法是一种非启发式算法,计算时间复杂度为  $O(n^2d)$ ,空间复杂度为  $O(n^2+d)$ ,其中  $d$  为属性的个数。该算法的优点是能求出决策表的所有约简,缺点是差别矩阵的计算复杂度高。基于信息熵的属性约简算法是一种启发式算法,它利用属性的信息熵度量属性的重要性,将信息熵小的属性逐一加入到候选属性约简中,直到满足预定义的停止条件。基于信息熵的属性约简算法的计算时间复杂度为  $O(nd)$ ,空间复杂度为  $O(n+d)$ ,该算法的优点是算法易于理解,也易于实现,缺点是它只能求出决策表的一个约简,而且不一定是最优约简。基于重要度的属性约简算法也是一种启发式算法,它利用依赖度作为属性重要度的度量,将重要度大的属性逐一加入到候选属性约简中,直到满足预定义的停止条件。基于重要度的属性约简算法的计算时间复杂度为  $O(nd)$ ,空间复杂度为  $O(n+d)$ 。与信息熵属性约简算法类似,该算法易于理解,容易实现,但它也只能求出决策表的一个约简,而且也不一定是最优约简。

其他的属性约简算法,包括针对经典粗糙集扩展模型的

到稿日期:2014-01-15 返修日期:2014-04-13 本文受国家自然科学基金项目(71371063,61170040),河北省自然科学基金项目(F2013201220,F2013201110),河北省高等学校科学技术研究重点项目(ZD20131028)资助。

翟俊海(1964—),男,博士,教授,主要研究方向为数据挖掘与模式识别,E-mail:mczjh@126.com;万丽艳(1986—),女,硕士生,主要研究方向为机器学习;王熙照(1963—),男,教授,博士生导师,主要研究方向为机器学习与计算智能。

属性约简算法,都是在这3种算法的基础上发展而来的。针对粗糙集属性约简等价类的计算问题, Qian等提出了属性约简的加速算法<sup>[5]</sup>, Chen等提出了样本对属性约简算法<sup>[6]</sup>, 这两个算法都极大地提高了计算属性约简的效率。针对离散和连续混合属性数据库, Hu等提出了一种模糊粗糙集属性约简算法<sup>[7]</sup>。E. Tsang等将经典粗糙集属性约简的概念扩展到模糊粗糙集模型, 针对模糊粗糙集模型, 提出了基于差别矩阵的属性约简算法<sup>[8]</sup>。受矩阵消元法的启发, Yao等提出了简化的差别矩阵属性约简算法<sup>[9]</sup>。Wang等将属性序关系引入到差别矩阵中, 也提出简化的差别矩阵属性约简算法<sup>[10]</sup>。Qian等将信息熵扩展为组合熵, 提出了基于组合熵的属性约简算法<sup>[11]</sup>。王钰等通过重新定义信息熵, 建立了相容粗糙集模型下的知识与信息熵的关系, 进而提出了针对相容粗糙集的基于信息熵的属性约简算法<sup>[12]</sup>。N. Parthala和 R. Jensen针对相容粗糙集模型和模糊粗糙集模型, 分别提出了基于重要度的属性约简算法<sup>[13,14]</sup>, 它们可直接处理连续值数据库, 而不需要离散化的过程。Yao等将 Bayes 决策理论引入粗糙集模型, 提出了决策粗糙集模型及相应的属性约简算法<sup>[15]</sup>。这些约简算法均没有考虑约简中条件属性之间的相关性, 这样求出的约简中可能依然包含冗余属性。针对这一问题, 本文提出了一种改进算法, 它利用最小相关性和最大依赖度准则求决策表属性约简。

## 2 基础知识

本节给出将要用到的粗糙集的基础知识及信息熵和互信息的基础知识。

### 2.1 粗糙集

本文在分类的框架下给出粗糙集的基础知识<sup>[2]</sup>。

**定义1** 四元组  $DT=(U, A \cup D, V, f)$  称为决策信息系统, 简称决策表。其中  $U=\{x_1, x_2, \dots, x_n\}$ , 为对象的非空有限集合, 称为论域;  $A=\{a_1, a_2, \dots, a_m\}$ , 是描述对象的属性集合;  $D=\{d\}$ , 是单决策属性(若  $U$  中的样例分为  $p$  类, 则  $d \in \{1, 2, \dots, p\}$ );  $V$  为属性值域,  $f: U \times A \rightarrow V$  为信息函数。

**定义2** 对于任意条件属性子集  $B \subseteq A$ , 定义  $B$  上的不可分辨关系(等价关系)为:

$$IND(B)=\{(x, y) \in U \times U \mid \forall b \in B, b(x)=b(y)\} \quad (1)$$

其中,  $b(x)$  和  $b(y)$  分别是对象  $x$  和  $y$  在属性  $b$  上的属性值。

**定义3** 令  $X \subseteq U$  是论域上的一个对象子集,  $B \subseteq A$  是条件属性子集, 定义  $X$  在条件属性子集  $B$  上的下近似为:

$$\underline{BX}=\{x \mid [x]_B \subseteq X\} \quad (2)$$

**定义4** 令  $X \subseteq U$  是论域上的一个对象子集,  $B \subseteq A$  是条件属性子集, 定义  $X$  在条件属性子集  $B$  上的上近似为:

$$\overline{BX}=\{x \mid [x]_B \cap X \neq \emptyset\} \quad (3)$$

**定义5** 令  $X \subseteq U$  是论域上的一个对象子集,  $B \subseteq A$  是条件属性子集, 定义  $X$  在条件属性子集  $B$  上的边界域为:

$$BN(X)=\overline{BX}-\underline{BX} \quad (4)$$

**定义6** 决策属性  $D$  划分论域  $U$  为  $U/D$ , 定义条件属性子集  $B$  关于决策属性  $D$  的正域为:

$$POS_B(D)=\bigcup_{x \in U/D} \underline{BX} \quad (5)$$

**定义7** 给定属性子集  $R \subseteq A$ , 对于任意属性  $a \in R$ , 称属性  $a$  在决策系统  $DT$  中对于  $R$  是不必要的, 如果  $POS_R(D)=$

$POS_{(R-\{a\})}(D)$ ; 否则称属性  $a$  对  $R$  是必要的。如果对于任意的属性  $a \in R$ ,  $a$  对  $R$  都是必要的, 则称  $R$  是独立的。

**定义8** 给定属性子集  $R \subseteq A$ , 称  $R$  是决策表  $DT$  的约简, 如果  $R$  满足:

$$(1) R \text{ 是独立的}; (2) POS_R(D)=POS_A(D).$$

记决策表所有约简的集合为  $RED_D(A)$ 。

**定义9** 对于任意属性  $a \in A$ , 决策属性  $D$  依赖于条件属性  $a$  的依赖度定义为:

$$\gamma_a(D)=\frac{|POS_a(D)|}{|U|} \quad (6)$$

若用依赖度作为条件属性重要性的度量, 则经典的基于重要度的属性约简算法可描述如下:

输入: 决策系统  $DT=(U, A \cup D, V, f)$

输出: 一个约简  $R \in RED_D(A)$

步骤1:  $R=\emptyset$ ;

步骤2: 重复如下步骤, 直到  $POS_A(D)=POS_R(D)$

步骤2.1: 对于所有的  $a \in A$ , 根据式(6)计算每个条件属性  $a$  的依赖度;

步骤2.2: 选出依赖度最大的属性:

$$a'=\operatorname{argmax}_{a \in A-R} \{\gamma_a(D)\};$$

步骤2.3:  $R=R \cup \{a'\}$ ;

步骤3: 输出  $R \in RED_D(A)$ 。

### 2.2 互信息

本节简要介绍信息熵和互信息的基本概念<sup>[16]</sup>。

熵可以用来度量一个随机变量的不确定性。设  $X$  是一个离散随机变量, 其值域为  $V_x$ , 它的概率分布函数是  $p(x)=P(X=x), x \in V_x$ 。信息熵的定义为:

$$H(X)=-\sum_{x \in V_x} p(x) \log_2 p(x) \quad (7)$$

如果  $X$  和  $Y$  是两个离散随机变量, 它们的联合概率分布为  $p(x, y)$ , 那么  $X$  和  $Y$  的联合熵定义为:

$$H(X, Y)=-\sum_{x \in V_x} \sum_{y \in V_y} p(x, y) \log_2 p(x, y) \quad (8)$$

给定随机变量  $X$  的情况下, 随机变量  $Y$  的不确定性可用如下的条件熵来度量:

$$\begin{aligned} H(Y|X) &= -\sum_{x \in V_x} p(x) H(Y|X=x) \\ &= -\sum_{x \in V_x} p(x) \sum_{y \in V_y} p(y|x) \log_2 p(y|x) \\ &= -\sum_{x \in V_x} \sum_{y \in V_y} p(x, y) \log_2 p(y|x) \end{aligned} \quad (9)$$

联合熵和条件熵的关系为:

$$H(X, Y)=H(X)+H(Y|X) \quad (10)$$

两个随机变量  $X$  和  $Y$  的相关程度可用互信息度量, 互信息的定义为:

$$I(X, Y)=-\sum_{x \in V_x} \sum_{y \in V_y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (11)$$

如果  $X$  和  $Y$  的互信息的值大(小), 那么  $X$  和  $Y$  的相关性大(小), 当知道其中一个变量后, 增加另一个变量对信息量的增加几乎没有影响。如果  $X$  和  $Y$  的互信息值为 0, 那么  $X$  和  $Y$  是完全不相关的。

互信息和熵及条件熵有如下关系:

$$I(X, Y)=H(X)-H(X|Y) \quad (12)$$

$$I(X, Y)=H(Y)-H(Y|X) \quad (13)$$

$$I(X, Y)=H(X)+H(Y)-H(X, Y) \quad (14)$$

对于连续型随机变量, 熵和互信息的公式中加号相应变

为积分的形式,  $p(x)$  和  $p(y)$  分别是  $X$  和  $Y$  的概率密度函数 (pdf), 而  $p(x, y)$  是  $X$  和  $Y$  的联合概率密度函数。

$$H(X) = - \int p(x) \log_2 p(x) dx \quad (15)$$

$$I(X, Y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (16)$$

对于给定的离散值决策表, 离散值条件属性的概率分布可用其频率分布来逼近, 对于任意的离散值条件属性  $X$  和  $Y$ , 它们的熵及其互信息可直接用给定的决策表计算得到。但对于给定的连续值决策表, 若其概率分布是未知的, 则需要用所给的决策表来估计连续值条件属性的概率密度函数。在没有任何先验知识的情况下, 估计连续值条件属性的概率密度函数是很困难的, 本文只研究离散值情况。

### 3 改进的属性约简算法

经典粗糙集中, 基于重要度的属性约简算法仅仅选出属性重要度大或依赖度大的条件属性加入到约简中。没有考虑约简中条件属性相互之间的相关性, 仅仅选择重要度大的条件属性来进行有效属性约简是不够的, 如果两个条件属性相关性很大, 那么在已选择其中一个条件属性的情况下, 再选择另一个条件属性显然是不必要的。因此, 用经典的基于属性重要度约简算法求得的约简依然含有冗余属性。我们希望选出来的条件属性相对于决策属性的依赖度大, 而条件属性之间的相关性小。条件属性之间的相关性用互信息度量, 基于这一思想, 本文提出了基于最小相关性和最大依赖度准则的属性约简算法。

算法从  $R = \emptyset$  开始, 每次循环对所有待选条件属性分别计算它们的依赖度, 选出依赖度值最大的那个条件属性, 再计算它与  $R$  中所有已选属性的互信息的平均值, 如果这个平均值不大于预先定义的阈值  $\lambda$ , 那么把它加入到约简  $R$  中, 否则不加入, 并从候选条件属性中删除此属性。算法直到在整个条件属性集上的正域等于在约简条件属性上的正域为止, 算法描述如下。

输入: 决策系统  $DT = (U, A \cup D, V, f)$ , 预定义阈值  $\lambda$

输出: 一个约简  $R \in RED_D(A)$

步骤 1:  $R = \emptyset; R' = A - R$

步骤 2: 重复如下步骤, 直到  $POS_{R'}(D) = POS_R(D)$

步骤 2.1: 对于  $R'$  中所有的条件属性  $a' \in R'$ , 利用式(6)计算其依赖度, 并选出依赖度最大的属性, 仍记为  $a'$ ;

步骤 2.2: 对于所有已选条件属性  $a \in R$ , 利用式(11)计算互信息  $I(a, a')$ ;

步骤 2.3: 如果  $\frac{1}{|R|} \sum_{a \in R} I(a, a') \geq \lambda$ , 则  $R' = R' - \{a'\}$ , 转到步骤 2; 否则  $R = R \cup \{a'\}, R' = R' - \{a'\}$ , 转到步骤 2;

步骤 3: 输出约简  $R$ 。

### 4 实验结果及统计分析

为了进一步验证本文算法的有效性, 我们在 8 个 UCI 数据集<sup>[17]</sup>上进行了实验, 8 个 UCI 数据集的基本信息如表 1 所列, 包括: 4 个离散值数据库 Breast Cancer (DB1), Lymphography (DB2), Soybean (Small) (DB3) 和 Congressional Voting Records (DB4); 1 个混合值数据库 Zoo (DB5); 3 个连续值数据库 Semeion Handwritten Digit (DB6), Connectionist Bench

(Sonar, Mines vs. Rocks) (DB7) 和 Wine (DB8)。对于连续值属性, 我们采用基于信息熵的离散化方法<sup>[2]</sup>进行离散化。实验采用十折交叉验证的方法, 比较本文算法和基于属性重要度算法求出的约简包含的属性个数及分类精度, 采用 C4.5 算法<sup>[18]</sup>求分类精度, 设置相关性阈值  $\lambda = 0.65$ 。实验环境是 PC 机, 双核 1.86G CPU, 2G 内存, Windows XP 操作系统, Matlab 7.1 实验平台, 实验结果如表 2 所列。

表 1 实验所用数据库的基本信息

DB	样例个数	属性个数	类别个数
DB1	286	9	2
DB2	148	18	4
DB3	47	35	4
DB4	435	16	2
DB5	101	17	7
DB6	1593	256	10
DB7	208	60	2
DB8	178	13	3

表 2 实验结果

DB	本文算法		基于属性重要度算法	
	属性个数	测试精度	属性个数	测试精度
DB1	4	0.9246	4	0.9181
DB2	6	0.7320	8	0.7280
DB3	2	0.9750	5	0.8750
DB4	9	0.9345	13	0.8462
DB5	6	0.9163	7	0.9132
DB6	20	0.6213	21	0.6057
DB7	15	0.7521	16	0.7464
DB8	9	0.8797	10	0.8783

从表 2 可以看出, 利用本文算法求出的约简包含的属性个数均少于或等于利用基于属性重要度的算法求出的约简包含的属性个数。其原因是本文算法不仅考虑了决策属性与条件属性之间的依赖关系, 还考虑了条件属性之间的相关性。因此, 利用本文算法求出的约简包含的属性个数更少, 但从实验结果发现, 利用本文算法求出的约简不一定是利用基于属性重要度算法求出的约简的子集。利用两种算法得到的约简来简化原数据集, 然后用 C4.5 算法从简化后的数据集中训练决策树, 并计算训练出的决策树的测试精度, 结果如表 2 所示。从表 2 也可以看出, 利用本文算法训练出的决策树的测试精度均高于利用基于属性重要度的算法训练出的决策树的测试精度, 其原因是利用本文算法得到的约简包含的冗余信息更少。实验结果证明了本文算法优于基于属性重要度的属性约简算法。

**结束语** 粗糙集的核心应用是属性约简, 基于属性重要度的约简算法仅考虑了决策属性与条件属性之间的依赖关系, 而没有考虑条件属性之间的相关性计算出的约简依然可能包括冗余属性。针对这一问题, 本文提出了最小相关性最大依赖度属性约简算法, 通过实验得出结论: 与基于属性重要度的约简算法相比, 利用本文算法求出的属性约简包含的属性个数更少, 但利用本文算法求出的约简不一定是利用基于属性重要度算法求出的约简的子集。实验结果表明本文提出的算法在求出的约简包含的属性个数及测试精度两方面均优于其他相关算法。

本文在处理连续值数据库时, 利用了基于信息熵的离散化方法, 离散化会丢失原数据集的一些有用信息。我们下一

(下转第 154 页)

后的算法容易过早收敛,造成算法在执行初期便无法再产生更优的可行解,以致于算法对解空间探索不充分,给搜索结果和质量带来很大的损失。针对该问题,本文提出一种带有摄动的导向性蚁群算法 PGACO,并设计了一种新的摄动方法。这种方法能够充分利用导向性局部搜索中惩罚因子在寻优过程积累的导向信息以及蚁群算法的正反馈特性指导搜索过程,当算法存在过早收敛的趋势时,通过在启发式信息上加入惩罚因子的影响来优化解构建的过程,使当前解移动到相邻的解空间中,以搜索更多的可能解。实验结果表明,与最大最小蚁群算法以及文献[3]中加入导向性局部搜索的蚁群算法相比,本文算法改善了过早收敛的问题,有效提升了算法性能,使其具有更强的全局搜索能力、更高的解质量,同时有更快的执行速度,其总体求解时间优于加入导向性局部搜索的蚁群算法,时效性较好。

### 参 考 文 献

[1] Dorigo M. Optimization, Learning and Nature Algorithms[D]. Milan: Dipartimento di Elettronica, Politecnico di Milano, 1992

[2] Dorigo M, Birattari M, Stützle T. Ant colony optimization[J]. Computational Intelligence Magazine, IEEE, 2006, 1(4): 28-39

[3] Hani Y, Amodeo L, Yalaoui F, et al. Hybrid optimization method for the facility layout problem[M]// Swarm Intelligence: Focus on Ant and Particle Swarm Optimization, 2007: 331-342

[4] Zhao N, Wu Z, Zhao Y, et al. Ant colony optimization algorithm with mutation mechanism and its applications[J]. Expert Systems with Applications, 2010, 37(7): 4805-4810

[5] Marinakis Y, Marinaki M, Doumpos M, et al. A hybrid ACO-GRASP algorithm for clustering analysis[J]. Annals of Opera-

tions Research, 2011, 188(1): 343-358

[6] Tairan N, Zhang Q. Population-based guided local search: Some preliminary experimental results[C]// 2010 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2010: 1-5

[7] Vansteenwegen P, Souffriau W, Berghe G V, et al. A guided local search metaheuristic for the team orienteering problem[J]. European Journal of Operational Research, 2009, 196(1): 118-127

[8] Hani Y, Amodeo L, Yalaoui F, et al. Ant colony optimization for solving an industrial layout problem[J]. European Journal of Operational Research. 2007, 183: 633-642

[9] Chehade H, Yalaoui F, Amodeo L, et al. Ant colony optimization for assembly lines design problem[C]// Proceedings of the 8th International FLINS'08 Conference on Computational Intelligence in Decision and Control. Madrid, 2008: 1135-1140

[10] 邢立宁, 陈英武. 基于混合蚁群优化的卫星地面站系统任务调度方法[J]. 自动化学报, 2008, 34(4): 414-418

[11] Voudouris C, Tsang E P K, Alsheddy A. Guided local search [M]. Springer US, 2010

[12] Stützle T, López-Ibáñez M, Pellegrini P, et al. Parameter adaptation in ant colony optimization [M]. Autonomous Search. Springer Berlin Heidelberg, 2012: 191-215

[13] Tsutsui S, Fujimoto N. Parallel ant colony optimization algorithm on a multi-core processor [M]. Swarm Intelligence. Springer Berlin Heidelberg, 2010: 488-495

[14] Dan Z, Hongyan H, Yu H. The Optimal Selection of the Parameters for the Ant Colony Algorithm with Small-Perturbation[C]// 2010 International Conference on Computing, Control and Industrial Engineering (CCIE). IEEE, 2010. 2: 16-19

(上接第 150 页)

步的工作是考虑直接从原数据集中利用非参数方法(如核密度估计方法)估计连续值属性的概率密度函数,然后计算条件属性之间的互信息,以度量条件属性之间的相关性,并研究本文算法在离散情况下得到的结论是否依然成立。

### 参 考 文 献

[1] Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356

[2] 苗夺谦, 李道国. 粗糙集理论、算法与应用 [M]. 北京: 清华大学出版社, 2008

[3] Skowron A, Rauszer C. The discernibility matrices and functions in information systems [C]// Slowinski R, ed. Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory. Kluwer, Dordrecht, 1992

[4] 苗夺谦, 王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论 [J]. 模式识别与人工智能, 1998, 11(1): 34-40

[5] Qian Y, Liang J, Pedrycz W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory [J]. Artificial Intelligence, 2010, 174(9/10): 597-618

[6] Chen D, Zhao S, Zhang L, et al. Sample Pair Selection for Attribute Reduction with Rough Set [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(11): 2080-2093

[7] Hu Q, Xie Z, Yu D. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation [J]. Pattern Recognition, 2007, 40(12): 3509-3521

[8] Tsang E, Chen D, Yeung D. Attributes reduction using fuzzy rough sets [J]. IEEE Transactions on fuzzy system, 2008, 16

(5): 1130-1141

[9] Yao Y Y, Zhao Y. Discernibility matrix simplification for constructing attribute reducts [J]. Information Sciences, 2009, 179(5): 867-882

[10] Wang J, Wang J. Reduction algorithms based on discernibility matrix; the ordered attributes method [J]. Journal of Computer Science and Technology, 2001, 16(6): 489-504

[11] Qian Y, Liang J. Combination Entropy & Combination Granulation in Rough Set Theory [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2008, 16(2): 179-193

[12] 王钰, 刘三阳, 刘振华. 相容 RS 理论中的属性约简算法 [J]. 计算机科学, 2003, 30(2): 161-162

[13] Parthala N, Shen Q. Exploring the boundary region of tolerance rough sets for feature selection [J]. Pattern Recognition, 2009, 42(5): 655-667

[14] Jensen R, Shen Q. New Approaches to Fuzzy-Rough Feature Selection [J]. IEEE Transactions on Fuzzy Systems, 2009, 17(4): 824-838

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

[16] Cover T M, Thomas J A. Elements of information theory (Second Edition) [M]. Hoboken: Wiley-Interscience, 2006

[17] Blake C L, Merz C J. UCI Repository of Machine Learning Databases[OL]. [2011]. <http://www.ics.uci.edu/~mllearn/ML-Repository.html>

[18] Mitchell T M. Machine Learning [M]. 北京: 机械工业出版社, 2003: 55-73