第 24 卷第 5 期
2007 年 10 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 24 No. 5
Oct. 2007

文章编号：1000−8152(2007)05−0766−05

属性约简矩阵特征结构及分层约简快速算法

徐 宁[1,2]，章 云[1]，孙海卫[3]，熊红艳[1]

(1. 广东工业大学 自动化学院, 广东 广州 510090;

2. 上海应用技术学院 计算机科学与信息工程系, 上海 200233; 3. 澳门大学 科技学院, 澳门)

摘要: 已有的属性约简算法往往只能提供一个可行解, 并且不能保证是最小约简解. 因此, 详细分析属性约简特征并获得最小约简解具有重要意义. 本文针对信息系统数据集提出属性约简矩阵, 通过矩阵的结构特征分析得到属性的约简特征, 因此采用矩阵代数的方法求解属性约简问题, 并得到分层约简快速算法. 经实例运算可见, 矩阵代数的处理和算法大大降低了最小约简解获取的计算量.

关键词: Rough集; 属性约简; 约简矩阵; 特征结构; 分层约简; 快速算法

中图分类号: TP18 TP391　　　**文献标识码**: A

Structural features of attribute reduction matrix and layer fast algorithm

XU Ning[1,2], ZHANG Yun[1], SUN Hai-wei[3], XIONG Hong-yan[1]

(1. Automation Faculty, Guangdong University of Technology, Guangzhou Guangdong 510090, China;

2. Department of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 200233, China;

3. Faculty of Science and Technology, University of Macau, Macau, China)

Abstract: Current algorithms provide only feasible attribute reductions, they are not necessarily the minimum ones. Thus, it is significant to analyze the characteristics of attribute reduction and get the minimum one from all reductions. Firstly, an attribute reduction matrix is defined based on the information system. The attribute reduction can be characterized by the structural features of the matrix. Then, a layer fast algorithm by matrix algebraic theory is proposed to solve the problem. Finally, numerical results are given to show the advantage of our method by highly reducing the computational cost.

Key words: rough sets; attribute reduction; reduction matrix; structural features; layer reduction; fast algorithm

1 引言(Introduction)

Rough集(又译粗糙集)于20世纪80年代初提出[1], 十年后受到世界范围的关注[2~4]. Rough集研究的核心是数据集的属性约简, 提出了基于数据分类的约简理论. 该理论现已广泛的应用于数据分析、数据预处理的属性约简中, 成为属性约简的重要数学工具. 关于高维数据集的属性约简, 因属性组合将出现爆炸性增长, 运用广度优先的方法来获得最小约简解, 在理论上已被证明是NP-hard问题[5], 因此多采用深度优先的方法. 深度优先通常可获得一个可行解, 但这个解不一定是完备解[6], 而且这个解与最小解的关系, 从理论上和计算上现都缺乏有效的证明或说明[7].

在Rough集约简理论下, 一个数据集的约简解不唯一, 而且最小约简解也不唯一. 不同的约简解含有不同的知识, 有必要对获得数据集的所有最小约简解进行讨论, 文献[8]介绍了广度优先算法的研究对NP问题的获解研究以及计算理论的发展具有着重要的意义.

本文从一定规模数据集出发, 寻求有效算法的设计, 在广度上进行属性约简研究: 首先提出源于数据集的属性约简矩阵; 依据此矩阵, 将数据集的属性约简问题抽象成为整数规划问题；继而看到属性约简矩阵的结构特征与数据集中属性组合的密切关系, 由此采用矩阵代数的方法, 得到了一个快速且可获取全最小约简集的分层约简算法.

收稿日期: 2004−10−26; 收修改稿日期: 2006−10−20.
基金项目: 广东省自然科学基金资助项目(04009480,06301299).

2 问题描述 (Problem statement)

2.1 几个概念(Basic concepts)

依据Rough集理论, 定义一个数据集上的信息系统 $S=\{U,A,V,f\}$, 其中论域 $U=\{u_1,u_2,\cdots,u_n\}$为对象集合; $A=C\bigcup D$为属性集合, 其中$C=\{c_1,c_2,\cdots,c_m\}$为条件属性集合, $D=\{d_1,d_2,\cdots,d_q\}$为决策属性集合; $f: U\times A\rightarrow V$是映射关系; $V=V_c\bigcup V_d$是属性取值值域, 其中$V_c(i,j)=f_c(u_i,c_j)\in V_c$, $V_d(i,k)=f_d(u_i,d_k)\in V_d$.

设$\mathbb{R}$为论域 U上所有等价关系组成的集合. $R\in\mathbb{R}$, $[u]_R$ 表示包含元素u的R等价类. U/R表示R的所有等价类构成的集合, 是 U的一个划分.

定义 1 $X\subseteq U, R\in\mathbb{R}$, 设:

$$R_-(X)=\bigcup\{Y\in U/R\mid Y\subseteq X\}=\{u\in U\mid [u]_R\subseteq X\},$$

$$R^-(X)=\bigcup\{Y\in U/R\mid Y\bigcap X\neq\emptyset\}=\{u\in U\mid [u]_R\bigcap X\neq\emptyset\}.$$

称$R_-(X)$为X的R下近似集, $R^-(X)$为X的R上近似集.

定义 2 $POS_R(U,X)=R_-(X)$称为X的R正域; $NEG_R(U,X)=U-R^-(X)$称为X的R负域.

设$P\subseteq\mathbb{R}$是$\mathbb{R}$上的一族等价关系, 其交集也是一个等价关系, 记为$\mathrm{ind}(P)=\bigcap P$.

"对象在第j个条件属性下取值相同"是一个等价关系, 以符号c_j 表示. $U/\mathrm{ind}(D)=\{Y_1,Y_2,\cdots,Y_k\}$是以决策属性作出的一个划分. 设$\overline{C}\subseteq C$, 记

$$POS_{\overline{C}}(U,D)=\bigcup_{i=1}^{k}POS_{\mathrm{ind}\overline{C}}(U,Y_i).$$

定义 3 $\overline{C}\subseteq C$, 如果$POS_{C-\overline{C}}(U,D)=POS_C(U,D)$, 则称$\overline{C}$是$C$中不必要的(可约简的); 否则称$\overline{C}$是$C$中必要的.

问题 1 信息系统$S=\{U,A,V,f\}$的属性约简问题可描述为求$\overline{C}$, 使得

$$J=\max|\overline{C}|,$$

$$POS_{C-\overline{C}}(U,D)=POS_C(U,D).$$

2.2 正域的性质(Properties of positive region)

文献[9]对正域的性质进行了讨论, 得到的研究结果之一是: 记$U^\#=POS_C(U,D)$, 则对任何$P\subseteq C$, 有$POS_P(U,D)=POS_P(U^\#,D)$.

因此可知, 信息系统S的属性约简可分别在相容记录和不相容记录中进行.

定义 4 设 $\pi_1=\{A_1,A_2,\cdots,A_n\}$, $\pi_2=\{B_1,B_2,\cdots,B_m\}$, 是集合$U$的两个划分, 如果$\pi_1$中每个 A_i 都是π_2中某个B_j的子集, 则称划分π_1是划分π_2的一个细分(refinement), 表示为: $\pi_1\preccurlyeq\pi_2$.

引理 1 如果信息系统S的论域 $U=U^\#=POS_C(U,D)$, $R\subseteq C$则:

$$POS_R(U,D)=POS_C(U,D)\Leftrightarrow\pi_R\preccurlyeq\pi_D. \quad (1)$$

证 设

$$\pi_R=U/\mathrm{ind}(R)=\{X_1,X_2,\cdots,X_r\},$$

$$\pi_D=U/\mathrm{ind}(D)=\{Y_1,Y_2,\cdots,Y_d\},$$

根据正域和划分的定义, 式(1)可等效为:

$$\forall u\in U,\exists\, i,j,\ \text{s.t.}[u]_R=X_i\subseteq Y_j,\Leftrightarrow$$
$$\forall i,\exists j,\ \text{s.t.}\ X_i\subseteq Y_j,$$
$$i=1,2,\cdots,r;j=1,2,\cdots,d. \quad (2)$$

1) 如果式(2)右边成立, 则 $\forall u\in U$, $\exists\, i,u\in X_i\subseteq Y_j$, 即 $u\in POS_R(U,D)$.

2) 如果式(2)左边成立, 则 $\forall i$ 和 $\forall u_1,u_2\in X_i$, $u_1\neq u_2$, $\exists j_1,j_2$, $[u_1]_R\in Y_{j_1}$, $[u_2]_R\in Y_{j_2}$. 显然$[u_1]_R=[u_2]_R\neq\emptyset$, 如果$j_1\neq j_2$, 则$Y_{j_1}\bigcap Y_{j_2}\neq\emptyset$, 与划分的定义矛盾. 所以, $j_1=j_2=j$, 即 $X_i\subseteq Y_j$.

综合上面两部分, 式(1)成立. 证毕.

由此得到:

$$POS_R(U^\#,D)=POS_R(U,D)=POS_C(U,D)=POS_C(U^\#,D).$$

则属性约简可表述为:

问题 2 信息系统$S=\{U,C\bigcup D,V,f\}$ 的属性约简问题是: 求最小的$R\subseteq C$, 且新的信息系统$\overline{S}=\{U^\#,\overline{A}=R\bigcup D,V,f\}$不出现不相容情况.

2.3 属性约简矩阵(Reduction matrix)

从问题2的表述知, 属性约简问题的实质是确保条件属性值对对象的描述可区分[10]. 为此, 下面给出基于区分的属性约简矩阵定义.

定义 5 对2个对象 u_{i1} 和 u_{i2}, $i_1\neq i_2$, 且$V_d(i_1,k)\neq V_d(i_2,k)$, 由其第$j$个属性的区分值得其区分向量的第$j$个分量为:

$$a((i_1,i_2),j)=\begin{cases}1, V_c(i_1,j)\neq V_c(i_2,j),\\ 0, V_c(i_1,j)=V_c(i_2,j),\end{cases}$$
$$j=1,2,\cdots,m.$$

定义 6 信息系统$S=\{U,A=C\bigcup D,V,f\}$的属性约简矩阵A由下列决策属性值不同的两两对象: $(1,2),(1,3),\cdots,(1,n),(2,3),(2,4)\cdots,(2,n,),$ $(3,4),(3,5),\cdots,(n-1,n)$的条件属性区分向量组成. 属性约简矩阵$A$的维数为$N\times m$, m为C中的属性数, N是在两两对象比较后产生的区分行向量数, 最大值$N\leqslant(n-1)^2$.

问题 3 在以上定义下可获得信息系统S的属性约简矩阵A, 则属性约简问题可再描述为:

$$J=\min\sum_{j=1}^{m}x_j, \tag{3a}$$

$$Ax\geqslant I, \tag{3b}$$

$$x=[x_1,x_2,\cdots,x_m]^{\mathrm T}, I=[1,1,\cdots,1]^{\mathrm T},$$

$$x_j=0 \text{ or } 1, j=1,2,\cdots,m. \tag{3c}$$

在以上定义下的属性约简矩阵A一定不存在全0行, 问题1定有解, 关键在如何找到全部解.

解由 x 表示, 其中 $x_j=0$ 表示属性 c_j 被约简, $x_j=1$则对应的属性组成约简集.

对问题3的另一个直观描述是: 在A中划掉最多的列(也划掉相应的 x_j), 且保证剩下的矩阵不出现全 0 行. 下面按照这样的思路来讨论问题的求解.

3 属性约简矩阵的结构特征 (Structural features of attribute reduction matrix)

3.1 对角规范型(Diagonal normal matrix)

引理 2 设a_{i1}, a_{i2}是属性约简矩阵A的两个行向量, 且$a_{i2}-a_{i1}\geqslant 0$, 将矩阵A中的a_{i2}行划掉得到矩阵A', 则: $Ax\geqslant I$ 与$A'x\geqslant I'$ 有相同的解(x满足式(3c)).

证 不失一般性, 设$A=\begin{bmatrix}a_{i1}\\a_{i2}\\\overline{A}\end{bmatrix}$, 那么$A'=\begin{bmatrix}a_{i1}\\\overline{A}\end{bmatrix}$. 因此有:

$$Ax\geqslant I\Leftrightarrow\begin{bmatrix}a_{i1}\\a_{i2}\end{bmatrix}x\geqslant\begin{bmatrix}1\\1\end{bmatrix}, \overline{A}x\geqslant\overline{I}, \tag{4a}$$

$$A'x\geqslant I'\Leftrightarrow a_{i1}x\geqslant 1,\ \overline{A}x\geqslant\overline{I}. \tag{4b}$$

1) 易知, $Ax\geqslant I$的解一定是 $A'x\geqslant I'$的解.

2) 如果x是$A'x\geqslant I'$的解, 从式(4)并考虑到$a_{i2}-a_{i1}\geqslant 0$, 所以x也一定是$Ax\geqslant I$的解. 证毕.

从该引理知, 如果属性约简矩阵中某一行的"1"完全覆盖了另一行, 则可以将该行划掉或删除掉.

定义 7 满足引理2条件的约简矩阵变换称为行吸收变换.

不难看出, 对属性约简矩阵A实施行或列调换, 也不会影响问题的求解(若进行了列调换, 将x 中相应的元进行调换即可).

定理 1 对属性约简矩阵A中所有满足引理2条件的行实施行吸收变换, 再经过适当的行或列调换, 可将矩阵A变换为如下对角规范型:

$$A^*=\begin{bmatrix}J & \\ & A_0\end{bmatrix}_{N^*\times m},\ J=\mathrm{diag}\{J_1,J_2,\cdots,J_r\}, \tag{5a}$$

$$J_k=\mathrm{diag}\{I_k^{\mathrm T},I_k^{\mathrm T},\cdots,I_k^{\mathrm T}\}_{N_k\times kN_k},$$

$$I_k=[\overbrace{1\ \cdots\ 1}^{k}]^{\mathrm T},\ k=1,2,\cdots,r, \tag{5b}$$

$$A_0=[\ ]_{N_0\times m_0},\ N_k\geqslant 0,$$

$$N_0+\sum_{k=1}^{r}N_k=N^*\leqslant N, m_0+\sum_{k=1}^{r}kN_k=m. \tag{5c}$$

证 1) 求矩阵A每一行的和, 按和由小到大重新排列.

2) 扫描矩阵A每一行, 对满足引理2条件的行实施行吸收变换.

3) 对实施行吸收变换后的矩阵进行适当的列调换, 将会得到式(5)的对角规范型. 因为, 如果有行和为"1"的行, 则经过所有可能的行吸收变换后, 为"1"的列将只有这一个1, 再经适当列调换, 便可得到J_1. 同样的分析, 可得到J_k. 对于不满足上述特征的行和列归入到矩阵A_0. 证毕.

定理1给出的对角规范型, 把条件属性集合进行了一个新的划分:

$C_k^*=\{J_k$所对应的属性$\}, k=1,2,\cdots,r$;

$C_0^*=\{A_0$所对应的属性$\}$.

定义 8 C_k^*中的属性称为k阶核属性, C_0^* 中的属性称为耦合属性.

不难看出, 对矩阵A^*的属性约简求解可划为对J和A_0的求解, 前者的求解是容易的, 问题3的求解将主要针对A_0的求解. 为此, 可假定属性约简矩阵A只有耦合部分.

3.2 上三角规范型(Up triangular normal type)

对属性约简矩阵A实施一定的行或列调换, 可变换为一个上三角规范型. 变换算法UTT:

步骤 1 $k=1,\ A_k=A$.

步骤 2 求矩阵A_k各列的和, 并按和的大小由小到大进行重排, 将列和最大的列全部调到最左边.

步骤 3 将A_k首列有"1"的行调到前部并将与首列相同的列调到一起.

步骤 4 将A_k中与首列相同的列调到一起(所得到的非零部分为A_{kk}.

步骤 5 $k=k+1$, 将首行有"1"的列和与首行相同的行划掉, 得到余下矩阵A_k, 若没有剩下的列则停止, 否则回到步骤2.

不难证明下面的定理.

定理 2 经过UTT算法, 将划掉行列对应恢复, 属性约简矩阵$A(N\times m)$变成如下形式:

$$A=\begin{bmatrix} A_{11}\,A_{12}\,A_{13}\cdots\,A_{1l}\\ A_{22}\,A_{23}\cdots\,A_{2l}\\ A_{33}\cdots\,A_{3l}\\ \ddots\ \vdots\\ A_{ll}\end{bmatrix}, \tag{6a}$$

其中:

$$A_{kk}=[a_{ij}]_{n_k\times m_k}, a_{ij}=1,$$
$$i=1,\cdots,n_k;\ j=1,\cdots,m_k;\ k=1,\cdots,l, \tag{6b}$$
$$m=\sum_{k=1}^{l}m_k,\ N=\sum_{k=1}^{l}n_k. \tag{6c}$$

同样的道理, 对属性约简矩阵A实施一定的行或列调换, 可变换为一个下三角规范型, 变换算法DTT(down triangular normal type), 该算法与UTT对偶, 在此不赘述.

4　问题的求解 (Problem solution)

记$A(j_1,j_2,\cdots,j_k)$ 为条件属性子集对应的属性约简矩阵, $\forall(j_1,j_2,\cdots,j_k)\subseteq\mathbb{C}$, 共有$\mathbb{C}_m^k$ 种组合. 若不考虑约简矩阵的结构特征, 按最小组合式穷举, 其算法(MINENUM)为:

步骤 1 $k=1$;

步骤 2 对$\forall(j_1,j_2,\cdots,j_k)\subseteq\mathbb{C}$, 判断$A(j_1,j_2,\cdots,j_k)$有无全0行;

步骤 3 若无全0行, 则计算停止, 得到全部最小解且$J_{\min}=k$; 否则, $k=k+1$, 重执行步骤2,3.

在前面的讨论知道, 经过适当的变换可得到式(6)的规范型, 这样利用结构特征大大减少了计算工作量.

记$A(\overleftarrow{j_1},j_2,\cdots,j_k)$ 为$A(j_1,j_2,\cdots,j_k)$ 中去掉第j_1 列和该列有“1”的行剩下的矩阵. 不失一般性, 可设$A(j_1,j_2,\cdots,j_k)$ 有如下形式:

$$A(j_1,j_2,\cdots,j_k)=\begin{bmatrix} I''\ A''\\ \ \ A'\end{bmatrix},$$
$$A'=A(\overleftarrow{j_1},j_2,\cdots,j_k). \tag{7}$$

因此, 可得引理.

引理 3　记$x=[1\ \ x']^{\mathrm{T}}$, 则

$$A(j_1,j_2,\cdots,j_k)x\geqslant I\Leftrightarrow A(\overleftarrow{j_1},j_2,\cdots,j_k)x'\geqslant I'.$$

问题3的可行解域: 若属性约简矩阵A的首行有k_1个“1”, 假定为$(j_1^1,j_2^1,\cdots j_k^1)$, 则对应的这$k_1$个属性至少要保留一个, 这会有$k_1$种可能; 假定保留$j_1^1$列, 那么在去掉第$j_1^1$列和该列有“1”的行剩下的矩阵中的首行若有$k_2$ 个“1”, 假定为$(j_1^2,j_2^2,\cdots,j_k^2)$, 则对应的这$k_2$个属性至少要保留一个, 这会有$k_2$ 种可能; 照此类推, 经过有限次后将不再剩下矩阵. 根据引理3知, 在此过程中保留下来的属性就是问题3的一个可行解. 这个求解过程是一个不断分支加层的树状结构.

综上, 可得到分层约简算法(La UTT / La DTT):

步骤 1 以UTT(或DTT)算法得到三角规范型(参见式(6)), 显然$J_{\min}\leqslant l$. 取$k_{\min}=l$. 记初始约简矩阵为A_1.

步骤 2 在A_1首行找到有“1”的列, 记下对应的属性序号; 将该列有“1”的行全部划掉, 得到A_2; 以此类推, 直到A_k为全“1”(此时找到一个可行解; 如果 $k<k_{\min}$, 则 $k_{\min}\Leftarrow k$)或者判断 $k=k_{\min}$?(表明再继续下去已无意义).

步骤 3 在A_{k-1}首行找到另一个有”1”的列, 记下对应的属性序号; 将该列有“1”的行全部划掉, 得到新的A_k. 判断A_k是否为全“1”, 若满足, 则为一个可行解. 重复步骤3, 直到扫描完A_{k-1}首行中所有有“1”的列.

步骤 4 继续回溯, 在A_{k-2}首行找到另一个有“1”的列, 记下对应的属性序号; 将该列有“1”的行全部划掉, 得到新的A_{k-1}. 对新的A_{k-1}采用步骤2和步骤3. 重复步骤4, 直到扫描完A_{k-2}首行中所有有“1”的列. 以此类推, 一直回溯到A_1.

该算法充分利用约简矩阵的结构特征(“1”的位置), 因此可以在远少于组合数的计算量下得到全最小约简集. 另外, 在算法的运行过程中, $k_{\min}$在自适应减少, 相应也减少了计算量. 若不采用三角规范型, 初始 $k_{\min}=m$.

5　实例计算 (Examples)

采用VB6.0平台和Access数据库, 在普通个人计算机(x86 Family 6 Model 86, 128 MRAM 133 MHz, W2000OS)上, 对若干实例数据库依据算法运算, 获得有益的研究结果.

例 1　CTR汽车数据库[4], 如表1. 该数据库有9个条件属性, 1个决策属性, 21个汽车样本数据, 修改数据$V_c(12,e)=1$改为$V_c(12,e)=0$. 经算法, 存在两个4个属性的最小解: a,d,e,i 和a,b,d,i. 采用启发式方法仅获得第1个解.

例 2　数据库源于文献[11], 为HSV病人医疗检测数据, 共有11个条件属性, 40个记录, 经离散化处理获得分类分析数据表. 最小解$J_{\min}=5$, 共有7个最小属性约简解.

例 3　数据源同例2, 有122个记录, 经离散化处理获得分类分析数据表. 算法解得最小解$J_{\min}=5$, 仅1个最小属性约简解.

例 4　数据源于文献[12], 数据库共有100个记录对象, 20个条件属性. 最小解$J_{\min}=5$, 共有12个最小属性约简解.

例 5　数据源同上例, 数据库共有568个记录对

象, 20个条件属性. 最小解$J_{\min} = 8$, 共有13个最小属性约简解.

各例不同算法下的计算结果和时间(生成A后)的比较如表2所示. 可见, 根据属性约简矩阵结构特征得到的算法, 不但可获取所有最小解, 而且计算量大大减少.

表 1 CTR汽车数据库

Table 1 Dataset of CTR cars

U	a	b	c	d	e	f	g	h	i	D
1	0	1	1	1	1	1	1	1	0	1
2	0	1	0	1	1	0	1	0	0	1
3	0	1	0	1	1	1	1	0	0	1
4	0	0	1	1	1	1	1	0	1	2
5	0	1	0	1	1	0	0	0	0	1
6	0	1	0	0	1	0	0	1	2	0
7	0	1	0	1	1	0	1	0	2	0
8	1	0	0	0	0	1	2	0	1	2
9	0	0	0	0	0	1	2	0	0	1
10	0	0	0	0	0	1	0	1	0	1
11	1	0	0	1	0	1	2	0	1	2
12	1	0	0	1	0	0	0	0	0	2
13	0	0	0	0	1	0	0	0	0	1
14	1	0	1	1	0	1	1	0	0	2
15	1	0	0	0	0	0	2	0	0	2
16	0	0	1	1	1	0	1	0	0	1
17	0	1	0	1	1	0	1	1	0	1
18	0	0	0	1	1	0	1	1	0	1
19	1	0	0	1	0	1	0	0	0	2
20	0	0	0	1	0	1	0	0	0	2
21	0	0	0	0	0	1	0	0	0	1

表 2 计算时间的对比

Table 2 Comparison of compution time s

实例	MINENUM	La UTT	La DTT	计算结果
1	0.012	0.0008	0.0008	$J_{\min}=4$, 2个
2	0.078	0.0039	0.0022	$J_{\min}=5, 7$个
3	0.126	0.0031	0.0052	$J_{\min}=5, 1$个
4	45.096	5.3281	4.625	$J_{\min}=5, 12$个
5	520.362	18.609	13.109	$J_{\min}=8, 13$个

6 结束语 (Conclusion)

属性约简是知识发现、机器学习、人工智能等领域研究的重要专题. 从数据集在进行属性约简时的结构特征出发, 在算法上寻找解决途径, 对获得最小约简解有重要意义. 本文提出基于信息系统的属性约简矩阵, 通过矩阵的结构特征反映属性约简的结构特征, 获得了分层约简算法, 相对于直接由全组合获取全最小约简集的广度搜索算法, 大幅降低了计算量, 是广度研究属性约简问题的重要开拓. 关于该算法的复杂性有进一步的研究.

参考文献(References):

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

[2] PAWLAK Z. Rough set theory and application to data analysis[J]. *J of Cybernetics and Systems: An International Journal*, 1998, 29: 661 – 356.

[3] 曾黄麟. 粗集理论与应用[M]. 重庆: 重庆大学出版社, 1996.
(ZENG Huanglin. *Rough Sets Theory and Applications*[M]. Chongqin: Chongqin University Publishing, 1996.)

[4] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
(ZHANG Wenxiu, WU Weizhi, LIANG Jiye, et al. *Rough Sets Theory and Methods*[M]. Beijing: Science Press, 2001.)

[5] WONG S K M, ZIARKO W. On optional decision rules in decision tables[J]. *J of Bulletin of Polish Academy of Science*, 1985, 33: 693 – 696.

[6] WANG J, MIAO D Q. Analysis on attribute reduction strategies of rough set[J]. *J Computer Science & Technology*, 1998, 113(2): 189 – 193.

[7] 陈志平, 徐宗本. 计算机数学-计算复杂性理论与NPC、NP难问题的求解[M]. 北京: 科学出版社, 2001.
(CHEN Zhiping, XU Zongben. *Computer Mathematics-Theories of Computational Complexity and the Solution of NPC & NP-hard Problems*[M]. Beijing: Science Press, 2001.)

[8] 周波涛. 可满足性问题(SAT)的快速算法的研究[D]. 广东, 广州: 华南理工大学, 2000.
(ZHOU Botao. *High-speed algorithm study on the satisfiability problem (SAT)*[D]. Guangzhou Guangdong: South China University of Technology, 2000.)

[9] 叶东毅, 陈昭炯. 粗糙集正域的若干性质[J]. 福州大学学报(自然科学版), 2002, 30(5): 521 – 523.
(YE Dongyi, CHEN Zhaojiong. Some properties of positive regions in Rough Set[J]. *Journal of Fuzhou University (Natural Science*, 2002, 30(5): 521 – 523.)

[10] SKOWRON A, ORLOWSKI M W. The discernibility matrices and functions in information systems[C]// *Int Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*. USA: Kluwer Academic Publishers, 1992: 331 – 362

[11] SLOWINSKI K. Rough classification of HSV patients[C]// *Int Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*. USA: Kluwer Academic Publishers, 1992: 77 – 93.

[12] GUO J Y. *Rough set-based approach to data mining*[D]. USA: Case Western Reserve University, 2003.

作者简介:

徐 宁 (1956—), 女, 副教授, 博士, 目前研究方向数据挖掘、Rough集、智能信息处理, E-mail: xun@sit.edu.cn;

章 云 (1963—), 男, 教授, 博士生导师, 博士, 目前研究方向嵌入式技术、智能信息处理等, E-mail: yz@gdut.edu.cn;

孙海卫 (1962—), 男, 助理教授, 目前研究方向科学计算等, E-mail: hsun@umac.mo;

熊红艳 (1975—), 女, 讲师, 硕士, 目前研究方向智能信息处理智能控制等.