

# 融合粗糙集与小生境免疫优化的属性约简方法

徐雪松<sup>1,2</sup>, 陈荣元<sup>1</sup>

(1. 湖南商学院计算机与信息工程学院, 湖南长沙 410205; 2. 国防科学技术大学信息系统与管理学院, 湖南长沙 410073)

**摘要:** 在粗糙集核属性的基础上, 融合小生境免疫优化提出一种决策属性约简方法. 将核属性参数作为抗体编码的先验信息, 通过疫苗自适应提取算法对抗体群接种疫苗, 提高抗体群多样性及稳定性. 为降低属性约简的计算复杂度, 引入属性集合的分类近似标准作为免疫优化的亲和度, 采用小生境免疫共享机制动态调整抗体群的亲和力, 提高算法局部搜索能力. 通过免疫记忆算子操作促使优良个体的保存, 在保证收敛速度的同时具有较强的全局和局部寻优能力. 通过滚动轴承故障诊断及 UCI 数据集的属性约简实验, 显示本算法在属性约简精度和效率方面具有较好效果.

**关键词:** 粗糙集; 核属性; 小生境; 免疫共享; 属性约简

**中图分类号:** TP182

**文献标识码:** A

**文章编号:** 0372-2112 (2014)08-1545-06

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2014.08.103

## Attribute Decision Reduction Method Based on Hybrid Rough Sets and Niche Immune Optimization

XU Xue-song<sup>1,2</sup>, CHEN Rong-yuan<sup>1</sup>

(1. College of Computer and Information Engineering, Hunan University of Commerce, Changsha, Hunan 410205, China;

2. Information System and Management College, National University of Defense Technology, Changsha, Hunan 410073, China)

**Abstract:** On the basis of analyzing the rough set theory and immune computational theory, a hybrid attribute reduction algorithm is proposed in this paper. In order to enhance the diversity and stability of the antibodies group, the attribute kernel parameter is taken into antibodies coding. Then the population is vaccinated stochastically through a bacterin extraction algorithm. The approximate quality is taken as the affinity function objection, which is in order to reduce computational complexity of this algorithm. A niche immune sharing mechanism is introduced in the optimization process, which can dynamically adjust the affinity of antibodies and improve the local searching ability of the population. Through all these operators to prompt the convergence speed and kept the balance of global and local optimization. The experiments of the rolling bearing fault diagnosis and UCI data sets reduction have illustrated that the algorithm has outstanding advantages in precision and efficiency.

**Key words:** rough set; immune sharing; clone selection; niche; attribute reduction

## 1 引言

粗糙集 (Rough Set) 理论是 Pawlak Z<sup>[1]</sup> 教授于 1982 年提出的一种能够定量分析并处理不精确、不完整信息的数学工具. 基于粗糙集的研究主要集中在属性约简、规则获取及智能算法等方面, 而属性约简中求出所有约简与求最小约简是典型 NP 难问题, 其计算复杂度会随着属性个数的增加而呈指数增加<sup>[2]</sup>. 为此大多数研究学者退而求其次, 提出各种算法来求解次优约简. 文献[3~5]从信息论角度阐述了粗糙集属性约简方法并取得

一定成效. 另有学者在属性约简过程中引入启发式信息, 采用数学规划及相对约简方法来减低约简的复杂度, 提高约简效率<sup>[6~9]</sup>. 近年来, 文献[10]提出了基于免疫克隆优化的粗糙集约简; 文献[11]提出基于邻域粒化和粗糙逼近的数值属性约简; 文献[12~15]结合量子计算、蛙跳算法、克隆优化及样例选取等提出了一系列基于进化算法的属性约简方法. 虽然进化算法具有较好局部寻优能力, 但为求全局最优解则需要付出较大的代价, 容易忽视问题的特征信息对求解问题时的辅助作用, 在求解一些规模较大的属性约简问题时就不灵活

了.本文在粗糙集核属性的基础上,设计了自适应疫苗提取算法对抗体接种疫苗,并融合小生境免疫优化提出一种决策属性约简方法.通过该方法对决策表中的条件属性进行简化,实验仿真表明其具有较好的效果.

## 2 粗糙集理论及核属性

决策表作为一种知识表达系统在粗糙集理论中有着重要的作用.为便于叙述,我们先对信息系统的属性约简和核属性进行简要描述.

**定义 1** 设  $S = \langle U, A, V, f \rangle$  为一知识表达系统,其中,  $U$  是所有对象的集合,称为论域;  $A = C \cup D$  是属性的集合,子集  $C$  和  $D$  分别称为条件属性集合和决策属性集合;  $V = \bigcup_{r \in R} V_r$  是属性值的集合,  $V_r$  表示属性  $r \in R$  值域,  $f: U \times A \rightarrow V$  是一个信息函数,它指定  $U$  中的每一个对象  $x$  属性值.

**定义 2** 给定  $S = (U, A)$ ,  $U$  是所有对象的集合,若  $B \subseteq A$  且  $B \neq \emptyset$ ,则  $B$  中所有等价关系的交集  $\cap B$  也是一个等价关系,称为  $B$  上的不可分辨关系,记为  $\text{IND}(B)$ .

**定义 3** 给定决策表信息系统  $S = \langle U, A, V, f \rangle$ ,对于每个子集  $X \subseteq U$  和不可分辨关系  $B \subseteq A$ ,  $[x]_B$  表示  $U$  中所有与  $x$  的不可分辨关系  $\text{IND}(B)$ ,  $X$  的上近似集  $B^+(X)$  和下近似集  $B_-(X)$  分别定义为:

$$\begin{aligned} B^+(X) &= \{x \mid (x \in U \wedge [x]_B \cap X \neq \emptyset)\} \\ B_-(X) &= \{x \mid (x \in U \wedge [x]_B \subseteq X)\} \end{aligned} \quad (1)$$

集合  $BN_B(X) = B^+(X) - B_-(X)$  称为  $X$  的  $B$  边界域.

粗糙集理论定义了相对约简与核属性,而核属性是描述对象条件属性不可缺少的信息,并且为无法约简的属性.为此,有学者提出可辨识矩阵方式来获取决策表系统核属性<sup>[16]</sup>,文献<sup>[17]</sup>进一步深入的分析了决策表核属性的计算方法,为决策属性约简提供了可行思路.

## 3 属性分类近似质量及优化目标

本文采用分类近似质量来衡量属性的分类能力.

**定义 4** 设  $X \subseteq U$  为论域上的一个子集.条件属性集为  $C$ ,决策属性集为  $D$ .  $X$  的下近似  $B_-(X)$ ,  $|U|$  表示集合中元素个数,  $Y_i \subseteq X$ ,  $Y_i$  为决策属性的类别,则信息系统的分类近似质量为:

$$f_B(U, D) = \sum_{i=1}^n \frac{|B_-(Y_i)|}{|U|} \quad (2)$$

公式(2)中分类近似质量表示应用条件属性集合能确切进行分类的比例,当对目标进行优化时只寻找那些分类能力出众且属性个数较少的子集.设定优化的亲和函数为公式(3):

$$F(B) = f_B(U, D) + \left[ \frac{N-m}{N} \right] \quad (3)$$

其中  $N$  为所有属性的个数,  $m$  为所选择抗体中所含属性的个数.这样优化求解条件属性集中所有属性值可以对数据库的一次扫描中同时求出,计算量得以简化.

## 4 属性信息编码及小生境免疫优化

### 4.1 抗体编码及接种疫苗

假设有 12 个条件属性  $\{z_1, z_2, \dots, z_{12}\}$  的决策表的核属性为  $\{z_3, z_6, z_8\}$ ,那么我们在初始群体选择时对其编码为“\* \* 1 \* \* 1 \* 1 \* \* \* \*”,并将其携带的先验信息引入抗体编码.令  $A = a_1, a_2 \dots a_l$  是二进制的抗体编码,在抗体  $A$  中,  $a_i$  为抗体基因并分为  $n$  段,每段长为  $l_i$ .接种疫苗过程中采用自适应疫苗抽取算法,通过对前  $k-1$  代保留下来的最优个体群  $Ab(k-1)$  和当前代的最优个体群  $Ab(k)$  进行分析,分别提取两代中最优个体基因位上共同有效信息用作为疫苗.按照先验知识强制性修改抗体群部分基因,使所得个体具有更高的亲和度.算法实现如算法 1 所示:

#### 算法 1 疫苗自适应提取及接种算法

```

Begin
While(condition = true)
统计父代群体,确定  $k-1$  代及  $k$  代最佳个体
 $Ab_{best}^k = \text{statistics}(Ab_i^k \mid i = 1, \dots, n)$ ;
 $Ab_{best}^{k-1} = \text{statistics}(Ab_i^{k-1} \mid i = 1, \dots, n)$ 
分解最佳个体,提取免疫基因信息:
 $H = \{h_j = Ab_{best,j}^k = Ab_{best,j}^{k-1} \mid j = 1, 2, \dots, l\}$ ;
Gauss 变异:  $Ab_i^k = \text{Mutation}(Ab_i^k)$ ;
for  $i = 1$  to  $n$ 
if  $\{p\} = \text{True}$ 
 $J = \text{random}(l)$ ;
接种疫苗:  $Ab_{H,i}^k = \text{Vaccine}(Ab_i^k, Ab_i^{k-1}, h_j)$ ;
免疫检测: if  $Ab_{H,i}^k < Ab_i^{k-1}$ , then  $Ab_i^k = Ab_i^{k-1}$ ;
else  $Ab_i^k = Ab_{H,i}^k$ ;
end;
end;
end;
```

其中  $Ab_{H,i}^k$  为对第  $k$  代第  $i$  个抗体  $Ab_i^k$  接种疫苗后所得抗体;  $p$  为接种疫苗的概率;  $Ab_{best,j}^k$  表示第  $k$  代优秀抗体的基因位信息;  $\text{Vaccine}(Ab_i^k, Ab_i^{k-1}, h_j)$  表示按模式  $h_j$  修改个体  $Ab_i^k$  上基因的接种疫苗操作;  $J$  表示随机生成 1 到  $l$  间的一个任意正整数;  $n$  和  $l$  分别为群体规模和个体基因长度.接种疫苗加速了优良模式的繁殖,修复了交叉、变异破坏的优良模式.通过种群与具有先验信息的疫苗相互作用、协同进化,保证了种群的多样性及收敛速度.

## 4.2 小生境免疫共享机制及免疫算子操作

本文通过小生境技术将种群中的优良个体保存于免疫记忆池中,通过共享选择策略来调整各个个体的适应度,以维护群体的多样性及实现信息共享.

### (1) 小生境免疫共享机制

我们首先定义共享函数,当个体之间的关系比较密切的时候,共享函数值较大(接近于 1),反之则较小(接近于 0).

**定义 5** 对于群体  $A(b) = \{Ab_1, Ab_2, \dots, Ab_n\}$ , 令  $H(d_{ij})$  表示个体  $Ab_i$  和  $Ab_j$  间的共享函数,其表达式为式(4)所示.

$$H(d_{ij}) = \begin{cases} 1 - \left(\frac{d_{ij}}{r}\right)^a, & \text{当 } d_{ij} < r \\ 0, & \text{当 } d_{ij} \geq r \end{cases} \quad (4)$$

其中,  $r$  为小生境半径,  $d_{ij}$  为个体间距离测度,  $d_{ij}$  为欧式距离.  $a$  为共享函数的调整参数. 一般  $a$  取 2.

**定义 6** 对于群体  $A(b) = \{Ab_1, Ab_2, \dots, Ab_n\}$ , 个体  $Ab_i$  在群体中的亲和度为:

$$S_i = \sum_{j=1}^n H(d_{ij}), i = 1, 2, \dots, n \quad (5)$$

**定义 7** 根据公式(3), 基于个体的共享适应度值的调整方法为:

$$F(Ab_i) = \frac{F(Ab_i)}{S_i}, i = 1, 2, \dots, n \quad (6)$$

可见,共享函数是一种特殊的非线性适应度值标度变换,其依据是群体中个体间的相似性. 该机制限制了群体内特殊抗体的无限制增长,其中共享半径  $r$  取值影响算法的搜索性能.

### (2) 免疫记忆算子

设抗体群  $A(b) = \{Ab_1, Ab_2, \dots, Ab_n\}$ , 其亲和力函数如公式(3)表示为  $F(Ab)$ , 由欧几里德距离计算抗体间距离测度. 根据  $d_{uv} < r$ , 确定局部小生境优良种群  $A_S$  并初始化记忆种群  $A_M$ .

$$A_M = \{Ab_{m1}, Ab_{m2}, \dots, Ab_{mm}\} \quad (7)$$

如果  $A_S$  中包含种群  $A$  中亲和力最高的抗体, 则保留其中一个最高亲和力的抗体, 并按  $F^*(Ab_i) = \frac{F(Ab_i)}{S_i}$ ,  $i = 1, 2, \dots, n$  更新其余抗体亲和力. 按亲和力大小进行各抗体排序, 选择前  $M$  个抗体更新记忆抗体集  $A_M$ .

### (3) 免疫检测算子

免疫检测算子对接种了疫苗的个体进行检测, 若其亲和力不如父代, 则将该个体被父代中所对应的个体所替代. 如果子代亲和力优于父代, 则子代将代替父代进入下一代种群. 免疫选择对算法的收敛性将起到决定性作用, 加强接种算子的积极作用, 消除其负面影响, 具有较强的鲁棒性.

## 5 算法执行过程

算法 2 协同进化在两个群体上, 当某个子种群发现最优解时, 将该最优解选入小生境集合. 当个体与任意一个小生境核的距离小于小生境半径  $r$  时, 将其适应度缩减为一个较小值, 各个子种群按新的亲和力进行进化. 令抗体群规模为  $N$ , 记忆库规模  $N_0$ , 克隆抗体规模为  $N_c$ , 迭代数为  $g_{\max}$ , 则抗体群及记忆种群的时间复杂度分别为  $O(N)$  和  $O(rN + N_0)$ , 克隆操作时间复杂度为  $O(N_c)$ , 抗体群及记忆库更新操作时间复杂度为  $O(NN_c^2)$  及  $O(NN_0^2)$ , 记忆单元学习时间复杂度最差为  $O(N_0^2)$ . 因此算法 2 的总时间复杂度最差为:

$$\begin{aligned} & O(N) + O(rN + N_0) + g_{\max} [ (O(r(N + N_0)^2 + O(N_n) \\ & \quad + O(NN_n^2) + O(NN_0^2) + O(N_n^2)) ] \\ & = O(N + rN + N_0 + g_{\max}(r(N + N + N_0)^2 + N_n \\ & \quad + NN_n^2 + NN_0^2 + N_n^2)) \\ & = O(g_{\max}(r(N + N_0)^2 + (N + 1)N_n^2 + NN_0^2)) \end{aligned}$$

### 算法 2 决策属性约简算法

输入: 一个决策表  $S = (U, A, V, f)$ ,  $A = C \cup D$ ,  $A$  为属性集合,  $C, D$  分别为条件属性和决策属性

输出: 决策表的一个属性约简  $R$ .

(1) **初始化**: 初始规模为  $N$  的抗体群  $A(b)$ , 引入属性核  $\text{Core}(P)$  对初始抗体群编码; 给定克隆规模  $n_c$ , 变异概率  $p_m$ , 记忆规模  $n_m$ , 小生境半径  $r$ . 从  $A(b)$  中选取  $M$  个抗体初始化记忆库  $A(b)_M$ ;

(2) **接种疫苗**: 根据 4.1 节计算父代种群最优抗体, 通过本文算法 1 对抗体群  $A(b)$  随机接种疫苗;

(3) **亲和力计算**: 通过共享策略调整抗体群及记忆种群个体的亲和力. 当任意小生境核的距离小于小生境半径  $r$  时, 将其亲和力缩减为一个较小值;

(4) **算子操作**: 根据克隆规模  $n_c$  及变异概率  $p_m$  对抗体实施克隆、变异等操作;

(5) **抗体补充**: 随机产生规模为  $N_r$  的抗体群, 根据当代提取疫苗概率性对种群  $N_r$  接种疫苗作为补充抗体;

(6) **更新记忆库**: 用记忆库  $A(b)_M$  中  $N_s$  个亲和力最高的个体代替抗体群中亲和力低的个体, 并对记忆库  $A(b)_M$  进行更新操作: 按照 4.2 节中所述的方法对  $k+1$  代抗体群操作, 得到下一代记忆单元  $Ab_M(k+1)$ ;

(7) **输出**: 当亲和力不再提高或者迭代结束, 则终止计算, 输出记忆库集合  $M$  即为优化的属性约简  $R$  集合.

## 6 试验仿真

**实验 1** 机械滚动轴承故障诊断中, 故障决策信息的获取是一重要环节. 设部分数据集论域  $U = \{1, 2, \dots, 16\}$ , 条件属性集  $C = \{s_1, s_2, \dots, s_{12}\}$ ; 决策属性  $D = \{0, 1\}$ , 它包括低频/高频故障及正常两种状态. 每组样本

利用频谱特征提取了 12 个条件属性和 1 个决策属性,其中条件属性  $S_1, S_2, \dots, S_{12}$  分别表示测点振动数据在频谱区间(  $0 - 0.2, 0.2 - 0.4, 0.4 - 0.6, 0.6 - 0.8, 0.8 - 1.0, 1, 2, 3, 4, 5$  倍频和大于 5 倍频)中的最大幅值. 由于表征频率分量上幅值大小的条件属性是连续变化, 因此采用了模糊 C 均值聚类方法进行离散化处理, 将每个属性离散成高、中、低 3 种状态, 如表 2 所示. 算法在 CPU 主频 2.0G, 内存 2G 的系统运行环境中, 参数初始值选择如表 1 所示. 运行 30 次取平均值得实验结果如表 3.

表 1 算法 2 运行参数

参数名	参数值	参数名	参数值
共享函数调整 $a$	2	共享半径 $r$	0.8 ~ 2.0
克隆规模 $n_c$	20	补充抗体群 $N_r$	10
变异概率 $p_m$	0.1	迭代次数 $g$	100
抗体规模 $N$	100	接种概率 $p$	0.3
记忆规模 $n_m$	0.5		

表 2 故障信息离散数据表

U	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	D
1	高	高	低	低	中	低	中	高	中	低	低	高	0
2	高	中	中	中	高	低	中	中	中	高	中	低	0
3	低	高	高	中	中	中	高	低	中	高	中	低	0
4	高	高	高	低	低	高	低	低	高	高	中	高	0
5	高	中	高	低	中	低	高	高	高	低	高	高	1
6	高	高	中	中	中	高	低	低	中	中	高	低	0
7	高	高	中	低	高	低	高	高	低	中	中	低	0
8	中	中	高	高	高	高	高	高	低	高	低	中	0
9	高	高	中	低	高	高	中	低	低	中	高	高	0
10	高	低	中	低	低	高	低	高	低	高	低	低	0
11	高	中	低	高	低	高	高	低	中	中	高	中	0
12	高	低	中	中	低	中	高	低	高	高	高	中	0
13	高	高	低	中	低	中	中	中	中	中	中	高	0
14	高	高	中	高	高	中	高	低	低	高	高	中	1
15	高	高	中	高	中	中	中	中	中	高	高	低	1
16	高	高	中	中	高	低	高	高	高	中	高	中	1

表 5 决策表基本信息及约简比较

决策表	U	C	Core	最小约简数目	文献[19]B算法			文献[19]C算法			本文算法 2		
					约简结果	是否含冗余属性	执行时间/ms	约简结果	是否含冗余属性	执行时间/ms	约简结果	是否含冗余属性	执行时间/ms
Patient	90	8	8	8	8	否	1.58	8	否	0.73	8	否	1.22
onks-1	432	6	3	3	4	是	2.97	3	否	1.32	3	否	1.93
Vote	435	16	7	9	10	是	24.05	9	否	18.12	9	否	16.34
Tic-Tac-Toe	958	9	0	8	8	否	17.90	8	否	23.57	8	否	15.02
Led17	2000	22	14	18	22	是	452.13	18	否	226.86	18	否	187.6
Poker	25010	10	5	7	7	否	902.27	7	否	760.57	7	否	522.45

表 3 属性约简计算结果

迭代次数	最优个体	个体适应度
1	110000111000	0.6321
3	100001110100	0.6521
5	000001111100	0.7344
9	110011000000	0.8542
13	110010000000	0.8741

当属性约简过程中搜索的属性数目庞大, 寻找最小约简是一个十分费时的过程. 由表 3 可以看到本文算法 2 在第 13 代即寻找到最小约简  $\{S_1 S_2 S_5\}$ , 种群快速收敛并保持稳定, 具有较快的全局收敛速度. 将该方法与 Rosetta 软件 (<http://www.lcb.uu.se/tools/rosetta/>) 进行属性约简完备性比较, 两种方法计算属性最小约简结果一致(如表 4 所示). 但本文算法 2 约简更为全面, 具有较好的完备性, 其所得条件属性更为直观而简单.

表 4 属性约简完备性比较

本文算法 2 属性项	Rosetta 属性项
S1 S2 S7 S8 S9	S1 S2 S5 S6 S7 S8 S9 S10 S11 S12
S1 S6 S7 S8 S10	S1 S2 S5 S6 S7 S8 S9 S10
S2 S6 S7 S8 S9	S2 S6 S7 S8 S9
S1 S2 S5 S6	S6 S7 S8 S9 S10
S1 S2 S5	S1 S2 S5

实验 2 为验证算法 2 对复杂决策表约简的有效性, 选择 UCI 数据库中 6 个决策表为实验数据(数据来源: <http://kdd.ics.uci.edu/summary.data/>), 其基本信息如表 5. 为统一运行环境, 实验选择 Petium4 2.8G, 内存 512M 的微机系统, 编程工具采用 VC++ 6.0, 算法参数选择如表 1, 并与文献[19]所介绍的 B、C 两种算法进行典型值比较, 运行 10 次平均结果如表 5 所示.

从表 5 数据分析可以看到,本文算法 2 和文献[19]的算法 C 执行效率普遍高于文献[19]算法 B,并且随着数据集的增大,这种优势越明显.在对本文算法 2 和文献[19]算法 C 单独比较中可以发现:当约简集与核属性集属性数目比值较小时,算法 2 性能基本与算法 C 一致,其原因主要是算法 2 的在抗体群和记忆抗体群协同优化过程中,接种疫苗及更新记忆库过程所耗费的计算量对优化对象规模较小情况时所占比偏高.例如,数据集 Patient 和 Monks-1 的约简集和核属性集是相同的,算法 C 当求出核属性集以后就等于约简算法完成了,但本文算法 2 却增加了规则集的寻优过程,增加了算法开销.

在论域和属性集较大的情况下,本文算法 2 的优越性要高于文献[19]B、C 算法.主要是在处理规模较大的属性约简过程中,本文通过核属性参数对优化过程进行先验信息的引导,算法的编码和适应度函数简单,并通过免疫记忆种群实现精英个体的保留,在提高全局搜索能力同时也提高了抗体群的多样性,同时保持了该算法快速收敛特性.特别的,在数据集 Tic-Tac-Toe 约简过程中,算 C 花费的时间却大于算法 B,这是因为该数据集中核属性数目为 0,约简集中属性均为非核属性,算法 C 的最后阶段需要对约简集中所有属性进行反向消除检查,这必然要花费一些时间,但本文算法 2 在执行算法第 2 步操作时,由于  $Core(P)$  为 0,将不再产生疫苗库进行接种疫苗.从分析表中属性约简结果可以看出,算法 C 和本文算法 2 的约简结果是完备的,并且获得的属性约简是最小约简,而算法 B 却不一定能保证约简的完备性.

## 6 结论

本文从数据挖掘过程对规则提取的要求出发,提出了一种融合免疫优化的粗糙集决策属性约简方法.为提高约简完备性和效率,文中设计了自适应疫苗提取算法对抗体接种疫苗,将核属性先验信息对抗体编码;引入小生境共享函数,动态调整抗体群的亲和度,对属性集合的分类近似标准进行优化.通过免疫记忆操作保持优良抗体,提高全局搜索能力同时保持该算法快速收敛特性.通过试验仿真分析表明,该方法能够有效地对规模较大的决策表进行约简,在获得最简约简的同时保持了较好的约简完备性.

## 参考文献

- [1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11: 341 – 356.
- [2] Kryszykiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1): 39 – 49.
- [3] Peters J F, Skowron A. A rough sets approach to knowledge discovery[J]. International Journal of Intelligent Systems, 2002, 17(2): 109 – 112.
- [4] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报, 1999, 10(2): 113 – 116.  
Miao Duoqian, Wang Jue. An information representation of the concepts and operations in rough set theory[J]. Journal of Software, 1999, 10(2): 113 – 116. (in Chinese)
- [5] 马建敏, 张文修, 宋朝晖. 基于信息量的序信息系统的属性约简[J]. 系统工程理论与实践, 2010, 30(9): 1679 – 1683.  
Ma Jianmin, Zhang Wenxiu, Zhu Zhaohui. Information quantity-based attribute reduction in ordered information systems[J]. System Engineering-Theory & Practice, 2010, 30(9): 1679 – 1683. (in Chinese)
- [6] Mi J S. Approaches to knowledge reduction based on variable precision rough set model[J]. Information Sciences, 2004, 159(3): 255 – 272.
- [7] 代建华, 李元香. 粗集中属性约简的一种启发式遗传算法[J]. 西安交通大学学报, 2002, 36(2): 1287 – 1290.  
Dai Jianhua, Li Yuanxiang. Heuristic genetic algorithm for reduction of attributes in rough set theory[J]. Journal of Xi'an Jiaotong University, 2002, 36(2): 1287 – 1290. (in Chinese)
- [8] 张腾飞, 肖健梅, 王锡淮. 粗糙集理论中属性相对约简算法[J]. 电子学报, 2005, 33(11): 2080 – 2083.  
Zhang Tengfei, Xiao Jianmei, Wang Xihuai. Algorithms of attribute relative reduction in rough set theory[J]. Acta Electronica Sinica, 2005, 33(11): 2080 – 2083. (in Chinese)
- [9] 苗夺谦, 周杰, 张楠. 基于代数方程组的属性约简研究[J]. 电子学报, 2010, 38(5): 1021 – 2027.  
Miao Duo-qian, Zhou Jie, Zhang Nan. Research of attribute reduction based on algebraic equations[J]. Acta Electronica Sinica, 2010, 38(5): 1021 – 2027. (in Chinese)
- [10] 梁霖, 徐光华. 基于克隆选择的粗糙集属性约简方法[J]. 西安交通大学学报, 2005, 39(11): 1231 – 1235.  
Liang Lin, Xu Guanghua. Reduction of rough set attribute based on immune clone selection[J]. Journal of Xi'an Jiaotong University, 2005, 39(11): 1231 – 1235. (in Chinese)
- [11] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报, 2008, 19(3): 640 – 649.  
Hu Qinghua, Yu Daren, Xie Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation[J]. Journal of Software, 2008, 19(3): 640 – 649. (in Chinese)
- [12] Ke Liangjun, Feng Zuren, Ren Zhigang. An efficient ant colony optimization approach to attribute reduction in rough set theory[J]. Pattern Recognition Letters, 2008, 29: 1351 – 1357.
- [13] Jiao Licheng, Li Yangyang. Quantum inspired immune clonal algorithm for global numerical optimization[J]. IEEE Transac-

- tions on System, Man and Cybernetics, Part B, 2008, 38(5): 1234 – 1253.
- [14] 丁卫平, 王建东, 管致锦. 基于量子蛙跳协同进化的粗糙属性快速约简[J]. 电子学报, 2011, 39(11): 2597 – 2603.  
Ding Weiping, Wang Jiandong, Guan Zhijin. Efficient rough attribute reduction based on quantum frog-leaping co-evolution [J]. Acta Electronica Sinica, 2011, 39(11): 2597 – 2603. (in Chinese)
- [15] 王熙照, 王婷婷, 翟俊海. 基于样例选取的属性约简算法[J]. 计算机研究与发展, 2012, 49(11): 2305 – 2310.  
Wang Xizhao, Wang Tingting, Zhai Junhai. An attribute reduction algorithm based on instance selection [J]. Journal of Computer Research and Development, 2012, 49(11): 2305 – 2310. (in Chinese)
- [16] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086 – 1088.  
Ye Dongyi, Chen Zhaojiong. A new discernibility matrix and the computation of a core [J]. Acta Electronica Sinica, 2002, 30(7): 1086 – 1088. (in Chinese)
- [17] 王国胤. 决策表核属性的计算方法[J]. 计算机学报, 2003, 26(5): 612 – 615.  
Wang Guoyin. Calculation methods for core attributes of decision table [J]. Journal of Computer, 2003, 26(5): 612 – 615. (in Chinese)
- [18] Yang T. Li Q. C. Reduction about approximation spaces of covering generalized rough sets [J]. International Journal of Approximate Reasoning. 2010, 51(3): 335 – 345.

- [19] 葛浩, 李龙澍, 杨传健. 基于冲突域的高效属性约简算法[J]. 计算机学报, 2012, 35(2): 342 – 250.  
Ge Hao, Li Longshu, Yang Chuanjian. An efficient attribute reduction algorithm based on conflict region [J]. Journal of Computer, 2012, 35(2): 342 – 250. (in Chinese)

#### 作者简介



徐雪松 男, 1978 年生于湖南郴州, 2009 年湖南大学控制科学与工程专业博士毕业, 现为国防科学技术大学国家重点实验室博士后, IEEE Member, 计算机学会会员, 自动化学会会员. 主要研究方向为复杂系统优化, 数据挖掘和人工智能.

E-mail: daniel613@126.com



陈荣元 (通信作者) 男, 1976 年生于江苏兴化, 2010 年武汉大学测绘与遥感专业博士毕业, 计算机学会会员. 目前主要研究方向为图像处理、数据挖掘和人工智能.

E-mail: chenrong yuan@126.com