

## 基于字符串编码克隆选择的建模 算法及可靠性预测

张伟<sup>1</sup> 胡昌华<sup>2</sup> 焦李成<sup>1</sup> 尚荣华<sup>1</sup>

**摘要** 提出了一种基于字符串编码克隆选择的建模算法,其核心是通过字符串编码的克隆选择策略在初等函数空间上进化、选择初等函数和函数间的运算关系,构成预测模型,并优化模型参数,实现预测模型形式的自动生成和参数的自适应确定.字符串编码设计使随机生成函数表达式成为可能;克隆选择策略实现了字符串编码抗体的进化训练;对基于字符串编码的克隆选择操作算子设计,增加了种群的多样性,使得不同抗体协同进化,有效避免了进化“早熟”现象.用该算法建立某武器系统可靠度预测模型,实验结果表明:该模型具有较高的精度,为可靠性预测、故障预报与维修奠定了基础.

**关键词** 字符串编码,克隆选择,建模算法,可靠性预测  
**中图分类号** TP277

### Modeling Algorithm Based on String-coding Clone Selection and Reliability Forecasting

ZHANG Wei<sup>1</sup> HU Chang-Hua<sup>2</sup> JIAO Li-Cheng<sup>1</sup>  
SHANG Rong-Hua<sup>1</sup>

**Abstract** Modeling algorithm based on string-coding clone selection is put forward. In the algorithm, evolution and choice of function structure, arithmetic relation and parameters are completed in elementary function space with string-coding clone selection, therefore, automatic production of forecasting model structures and adaptive determination of forecasting model parameters are realized. Design of string-coding makes it possible to randomly produce an expression in elementary function space; clone selection strategy achieves evolutionary training of string-coding antibodies. Moreover, by design of clone operator to string-coding antibodies, different antibodies evolve coordinately, diversity of antibodies restrains “precocity” in evolution. Reliability forecasting model of the weapon system is built using the algorithm. Results of experiments show that the forecasting model is of high accuracy, and lays a foundation of reliability forecasting, fault forecasting and maintenance for the control system.

**Key words** String-coding, clone selection, algorithm of modeling, reliability forecasting

基于实验数据建立被测对象变化规律的预测模型,关键是恰当地选择合适的模型和模型参数,如果模型结构和参数选择不当,模型逼近实际系统的能力就会受到影响<sup>[1-2]</sup>.潘正军提出一种通过遗传算法学习、确定模型结构和参数的思

路<sup>[3]</sup>,但标准遗传算法<sup>[4-5]</sup>在选择、交叉、变异环节上存在不足:首先,选择环节主要依靠适应度调节,因而过早地剥夺了某些存在优化潜力、而适应度暂时不高的个体进化的机会;其次,经过选择后保留的个体,多样性已经受损,交叉操作进一步将优势个体的部分特征传递给其他个体;再次,很小的变异概率使变异处于背景算子的地位,种群很难摆脱局部寻优的困境.综上,容易由多样性不足引发“早熟”.另外,作者用二叉树表示函数,为避免模型过于复杂,二叉树的深度是有限的,即二叉树的深度超过某一数值,函数表示及运用遗传算法针对二叉树结点的交叉、变异等操作将复杂到难以实现的地步.为此,本文提出一种基于字符串编码克隆选择的建模算法,其基本思想是:1)设计了一种抗体字符串编码,基于这种编码方式,能够随机而自动地生成基本初等函数、运算关系、参数及初等函数空间的函数表达式;2)在建模算法中,设计了特有的抗体-抗体亲合度函数、变异策略、选择概率的多样性系数,目的是实现对这种字符串编码抗体的克隆选择操作,从而用该字符串编码克隆选择策略对抗体群进行数据样本的拟合性训练、进化学习;3)经一代代进化,最终从未代进化抗体群中选择反映研究对象变化规律的最优抗体,作为预测模型.上述算法实现了字符串编码抗体在克隆选择进化训练中的建模,有利于数学模型结构、参数的优化搜索,有利于抗体群多样性的改善,有利于不同抗体间的自适应协同进化.

基于字符串编码克隆选择的建模算法不同于传统意义上的二进制、实数编码的克隆选择优化<sup>[6-8]</sup>,后者侧重于寻找函数的全局最优数值解;而前者则实现了初等函数空间上函数模型的自适应确定、进化及其参数的自适应调节.

本文以某武器系统可靠性时间序列实验数据为研究对象,建立了可靠度预测模型.实验表明,基于该建模算法建立的预测模型具有良好的精度.

## 1 基于字符串编码克隆选择的建模算法

### 1.1 抗体字符串编码设计

字符串编码克隆选择建模算法中的抗体采用字符串编码,即每个抗体对应一个字符串,一个字符串就是一个待优化的初等函数表达式.为产生完整、合乎数学逻辑、便于克隆选择操作的函数表达式,建立函数算子集 $\{\sin, \log, \exp, \dots\}$ 、自变量集 $\{x_1, x_2, \dots, x_m\}$ 、运算符集 $\{+, -, *, /, ()\}$ 、参数集 $\{\text{实数}\}$ .

函数算子集用于产生初始抗体的基本初等函数算子,也用于在抗体变异时产生新基本初等函数算子替代旧基本初等函数算子.为实现函数表达式在进化过程中的递归调用,函数算子集中元素宜选用定义域为 $(-\infty, +\infty)$ 的基本初等函数算子,以保证定义域、值域的封闭性.

自变量集用于产生相应维的自变量.

运算符集用于产生数学运算符、优先级运算符,规定运算操作性质,也用于在抗体变异时变更运算符,它出现在抗体的系数与函数、函数与函数、函数与常数之间,使生成的字符串-函数表达式合乎数学逻辑.

参数集用于依照概率在初始抗体中产生系数、常数或产生幂函数算子中的指数,也用于在抗体变异时修改相应参数,从而使字符串表示的函数式更完整.

不同抗体中必备的部分包括:基本初等函数算子、不同维自变量、指数(基本初等函数为幂函数时产生)和运算符.有些抗体中除必备的部分外还会依据概率产生系数和常数.抗

收稿日期 2006-12-18 收修改稿日期 2007-07-03  
Received December 18, 2006; in revised form July 3, 2007  
国家自然科学基金(60736026),教育部“新世纪优秀人才支持计划”,国防研究项目(203020301)资助  
Supported by National Natural Science Foundation of China (60736026), Program for New Century Excellent Talents in University, National Defence Research Program (203020301)  
1. 西安电子科技大学智能信息处理研究所 西安 710071 2. 第二炮兵工程学院 302 室 西安 710025  
1. Institute of Intelligent Information Processing, Xidian University, Xi'an 710071 2. 302 Unit, the Second Artillery Engineering Institute, Xi'an 710025  
DOI: 10.3724/SP.J.1004.2008.00105

体的组成部分通过链接方式生成完整的抗体函数表达式。

为使函数表达式便于克隆选择操作, 定义了抗体有效性准则, 如: 每维变量在函数表达式中只出现一次, 系数排列在函数项之前, 常数项排列在函数表达式末尾等。

## 1.2 基于字符串编码克隆选择的建模算法

**Step 1.** 按字符串编码策略和抗体有效性准则随机生成  $n$  个初始抗体, 并组成初始抗体群; 建立抗体库, 放入初始抗体群。

**Step 2.** 除对抗体群建立专门记录各抗体形式的抗体库外, 还建立了函数类型库、函数次序库、参数数值库、符号类型库、符号次序库, 以记录、管理进化各代各抗体中不同维变量的函数类型及位置、系数/指数/常数数值及位置、运算符类型及位置。

**Step 3.** 设  $t$  为训练集样本数目,  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$  为输入样本集, 其中:  $\mathbf{x}_j = (x_1, x_2, \dots, x_m), j \in t, m$  为输入样本维数;  $\{y_1, y_2, \dots, y_t\}$  为输出样本集;  $A_r$  为抗体,  $r \in n$ ; 构造抗体-抗原亲合度函数  $f(A_r)$

$$f(A_r) = \frac{1}{1 + E_r}, \quad r = 1, 2, \dots, n \quad (1)$$

$0 < f(A_r) < 1, E_r$  为训练集样本均方根相对误差

$$E_r = \sqrt{\frac{1}{t} \sum_{j=1}^t \left( \frac{P_{r,j} - y_j}{y_j} \right)^2}, \quad j \in t, r \in n \quad (2)$$

$P_{r,j}$  为第  $j$  个样本在抗体  $A_r$  函数表达式中的预测值;  $y_j$  是第  $j$  个样本实际输出值。

**Step 4.** 构造抗体-抗体亲合度函数  $\theta_r$ . 抗体间的亲合度体现了两抗体的相似程度. 抗体间亲合度愈大, 抗体间差异愈大, 抗体间抑制作用愈小; 反之, 抗体间亲合度愈小, 抗体间相似度愈高, 抗体间抑制作用愈大. 依据上述准则, 在字符串编码方式下, 构造了其特有的抗体-抗体亲合度函数  $\theta_r$ , 该函数通过比较两抗体中相应维变量的函数类型、两抗体中相应自变量函数有/无系数及取值、两抗体中有/无常数项及取值、两抗体中运算符/运算优先级种类及位置, 以相异得分的方式对  $\theta_r$  赋值. 分数值愈高, 表示抗体间相异性愈强, 即

$$\theta_r = \min(\psi_{r,s} + \phi_{r,s} + \varphi_{r,s} + \xi_{r,s}) \quad (3)$$

$s \neq r, s = 1, 2, \dots, n$

其中,

$\psi_{r,s}$  为两抗体  $A_r, A_s$  中相应维变量函数算子相异得分函数;

$\phi_{r,s}$  为两抗体  $A_r, A_s$  中相应维变量的系数相异得分函数;

$\varphi_{r,s}$  为抗体中运算符及运算优先级相异得分函数;

$\xi_{r,s}$  为两抗体  $A_r, A_s$  中常数项相异得分函数。

**Step 5.** 克隆操作. 设克隆规模系数为  $D$ , 对于  $A_r$ , 其克隆倍数  $q_r$  是抗体-抗体亲合度  $\theta_r$  和抗体-抗原亲合度  $f(A_r)$  的函数。

$$q_r = \text{int} \left( D \frac{f(A_r)}{\sum_{r=1}^n f(A_r)} \theta_r \right), \quad r \in n \quad (4)$$

$\text{int}(\cdot)$  为取整函数,  $q_r$  愈大, 经高频变异后抗体的形式愈多, 意味着它所得到的搜索方向愈多, 因而它进化的机会愈多, 由 (4) 知: 经一代代训练、搜索、进化,  $\theta_r, f(A_r)$  值均变大的抗体, 其克隆规模也会自适应地变大; 而  $\theta_r$  或  $f(A_r)$  单项变大的抗体, 其克隆规模也会有所增长. 这样在兼顾种群多样性的同时, 又使具有不同特性的抗体得以协同进化。

**Step 6.** 对抗体库、函数类型库、函数次序库、参数数值库、符号类型库、符号次序库及时添加克隆抗体信息。

**Step 7.** 高频变异操作. 对新增克隆抗体进行变异操作, 变异策略主要包括:

1) 从函数算子集中产生新算子替换变异抗体中被随机抽到的某维变量的旧算子; 重构变异抗体。

2) 交换变异抗体中被随机抽到的不同维变量各自的算子; 重构变异抗体。

3) 变更变异抗体中被随机抽到的某个运算符; 重构变异抗体。

4) 对变异抗体中被随机抽到的某维变量进行系数增、减、变值操作; 重构变异抗体。

5) 对变异抗体中的常数项进行增、减、变值操作; 重构变异抗体。

6) 对变异抗体中被随机抽到的某维变量为幂函数算子的指数进行变值操作; 重构变异抗体。

**Step 8.** 修改变异后的克隆抗体在抗体库、函数类型库、函数次序库、参数数值库、符号类型库中的信息。

**Step 9.** 克隆选择操作. 设由抗体  $A_r (r = 1, 2, \dots, n)$  克隆、变异后的抗体群  $B_{r,d}, r \in n, d = 1, 2, \dots, q_r$  中抗体-抗原亲合度最大的抗体为  $B_{r,g}, r \in n, g \in q_r$ , 其抗体-抗原亲合度为  $f(B_{r,g})$ , 与  $f(A_r)$  比较、选择, 进而产生新一代抗体, 其中第  $r$  个抗体  $A'_r$  的选择概率为式 (5)。

式 (5) 中, 抗体群多样性系数  $\alpha > 0$ , 多样性愈好, 其值愈大. 该算法中  $\alpha$  取值设计为本代抗体-抗体亲合度  $\theta_r (r = 1, 2, \dots, n)$  的均值, 这样克隆选择概率能够实现自适应调节. 克隆选择概率使优势抗体直接进入新一代, 同时这种 Gauss 型选择概率也给某些非优势抗体得以保留和继续进化的机会, 进一步增加了抗体群的多样性。

$$P(A'_r = B_{r,g}) = \begin{cases} 1 & f(A_r) < f(B_{r,g}) \\ \exp\left(-\frac{f(A_r) - f(B_{r,g})}{\alpha}\right) & f(A_r) \geq f(B_{r,g}) \text{ 且 } A_r \text{ 非本代最优抗体} \\ 0 & f(A_r) \geq f(B_{r,g}) \text{ 且 } A_r \text{ 是本代最优抗体} \end{cases}$$

$$P(A'_r = A_r) = \begin{cases} 0 & f(A_r) < f(B_{r,g}) \\ 1 - \exp\left(-\frac{f(A_r) - f(B_{r,g})}{\alpha}\right) & f(A_r) \geq f(B_{r,g}) \text{ 且 } A_r \text{ 非本代最优抗体} \\ 1 & f(A_r) \geq f(B_{r,g}) \text{ 且 } A_r \text{ 是本代最优抗体} \end{cases} \quad (5)$$

**Step 10.** 用新一代抗体更新抗体库、函数类型库、函数次序库、参数数值库、符号类型库、符号次序库。

**Step 11.** 重复 Steps 3~10, 直至满足终止条件, 停止进化, 保留最优抗体。

注. Step 11的终止条件采用设定进化代数 and 停止准则相结合的方式, 即若满足停止准则则停止进化, 跳出预先设置的进化循环; 若不满足停止准则则继续进化循环训练, 直至设定的进化代数. 因此, 进化代数与克隆规模系数的设定应兼顾训练的实际效果, 处理具体问题时可据实验结果调节. 停止准则定义为

$$|1 - f(A_r^*)| < \varepsilon \quad (6)$$

其中, 1 是训练样本集的抗体 - 抗原亲合度期望值, 它由实际训练样本集的输出值按照式 (1)、(2) 计算得到;  $f(A_r^*)$  是当前进化中最优抗体的抗体 - 抗原亲合度;  $\varepsilon$  是根据实际问题的精度要求确定的误差值。

## 2 某武器系统可靠性预测模型建模实验

以某武器系统可靠性实测数据为研究对象, 基于这些数

据建立可靠度变化规律的预测模型, 根据经验知识和评价指标, 将所得的 75 个可靠度时间序列数据构成 6 维 69 组输入样本和 1 维 69 组输出样本, 用其中 50 组作为训练集, 19 组作为测试集. 用训练集训练字符串编码克隆选择的抗体群, 得到的末代最优抗体作为该系统可靠度预测模型, 测试集用于检验预测模型的泛化能力. 训练前设定: 初始抗体群规模  $n$  取 50, 克隆规模系数  $D$  分别取 10、20, 抗体进化代数  $G$  分别取 5、10、20. 在上述参数不同组合设置下, 分别随机地进行 30 次预测模型的训练、泛化实验, 相继得到训练、泛化结果, 主要指标均值如表 1 所示. 图 1、2 是预测模型训练和泛化的预测、误差曲线, 该预测模型形如式 (7). 用最小二乘 AR 算法、RBF 核的 SVM 回归算法建立该例的预测模型进行对比实验, 实验结果如表 2(见下页) 所示. 上述实验中误差均用式 (2) 计算。

由基于字符串编码克隆选择建模算法得到的预测模型为

$$f(\mathbf{x}) = x_3^{0.86605} + \frac{x_1^{0.60291} \ln x_6}{\sin x_5} - \frac{\ln x_4}{x_2^{0.59787}} \quad (7)$$

表 1 字符串编码克隆选择的可靠度建模实验主要指标数值

Table 1 Main parameter values of reliability modeling experiments based on string-coding clone selection

克隆规模系数	代数	最优抗体 - 抗原亲合度均值	平均训练时间 (s)	抗体克隆平均值 (个)	训练集误差均值	泛化集误差均值
D=10	G=5	0.907475	55.45	12.227333	0.002577	0.0151
	G=20	0.970178	6.31067E+002	39.370667	8.836667E-004	0.015953
	G=50	0.977232	2.47790E+003	47.767333	6.829333E-004	0.010496
D=20	G=5	0.937859	1.16365E+002	36.434667	0.0018	0.017999
	G=20	0.975922	1.36082E+003	90.631333	7.119667E-004	0.01233
	G=50	0.982328	4.73772E+003	105.514	5.185E-004	0.008539

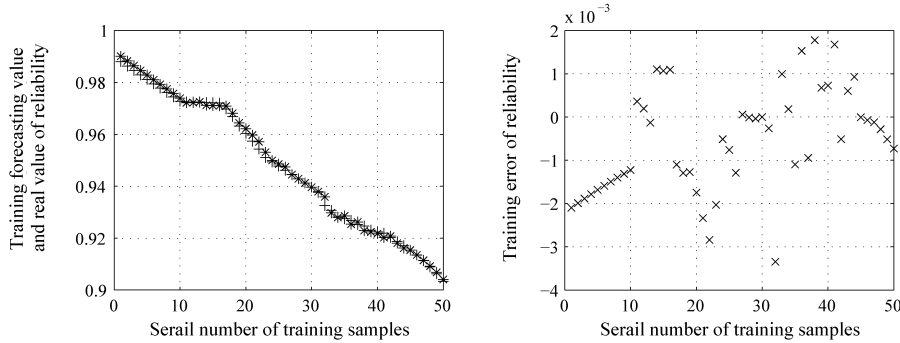


图 1 字符串编码克隆选择的可靠度模型训练预测、误差曲线

Fig. 1 Training forecasting and error graph of reliability model based on string-coding clone selection

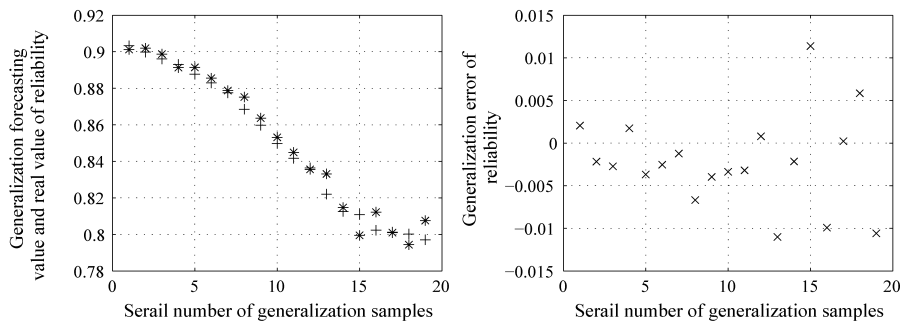


图 2 字符串编码克隆选择的可靠度模型泛化预测、误差曲线

Fig. 2 Generalization forecasting and error graph of reliability model based on string-coding clone selection

表 2 最小二乘 AR 可靠度模型与 SVM 可靠度模型误差  
Table 2 Errors of LSAR and SVM reliability model

模型	训练误差	泛化误差
最小二乘 AR 模型	0.002972	0.026045
SVM 模型 (核为 RBF)	0.026838	0.045546

### 3 实验结果分析

1) 从表 1 可以看出: 当克隆规模系数一定时, 随着进化代数增加, 抗体得以优化, 训练误差、泛化误差呈下降趋势; 伴随着抗体 - 抗原亲合度的增大, 抗体平均克隆数递增, 训练的时间开销增大。

2) 采用字符串编码, 为在初等函数空间上随机生成函数表达式 (抗体) 提供了条件, 待训练、优化函数族 (克隆抗体群) 在建模实验中规模最多时达到 5000 余个, 因而扩大了预测模型的候选范围, 有利于模型结构、参数进化; 字符串编码的克隆选择算法实现了初等函数空间上函数族的训练和优化, 建模和优化同时完成; 所建模型针对性强, 模型预测精度高。

3) 在该建模算法中, 构造了字符串编码抗体 - 抗体亲合度函数, 实现了抗体间相似度量, 从而保护了抗体群的多样性; 克隆计算分别赋予抗体 - 抗原亲合度与抗体 - 抗体亲合度不同组合的数值以不同的克隆倍数, 使具有不同特性的抗体协同进化; 高频变异算子设计给予克隆抗体群以更多的搜索方向, 兼顾了局部与全局寻优, 增加了训练中进化的机会; 克隆选择概率使优势抗体得以保留, 同时 Gauss 型选择概率也给了非优势抗体被保留的机会。

4) 用字符串编码克隆选择的建模算法建立可靠度预测模型; 与最小二乘 AR 可靠度模型、核为 RBF 的 SVM 可靠度模型相比, 具有更高的精度, 这是因为最小二乘 AR 模型本质上是线性模型, 在训练集上有一定的精度, 在泛化集上当样本有所“波动”时, 模型拟合精度变差; SVM 回归模型以结构风险最小化为优化准则, 辅以映射能力强的核函数, 更适于解决复杂的非线性建模问题。

### 4 结论

本文设计了一种字符串编码, 基于这种编码方式, 实现了初等函数空间函数表达式的自动生成; 提出了基于这种字符串编码的克隆选择建模算法, 实现了在初等函数空间上进化、选择初等函数和函数间的运算关系, 构成预测模型, 并优化其模型参数, 从而完成了预测模型形式的自动生成和参数的自适应确定。克隆选择策略中抗体 - 抗体亲合度函数的构造和高频变异算子的设计等有效避免了进化过程中的“早熟”现象。运用该建模算法建立某武器系统可靠度预测模型的实验表明, 预测模型具有较高的精度。该建模方法为可靠性预测、故障预报与预防维修提供了一种实用的建模方法。

### References

- 1 Chapelle O, Vapnik V, Bousque O, Mukherjee S. Choosing multiple parameters for support vector machines. *Machine Learning*, 2002, **46**(1): 131–159

- 2 Wong M L, Leung K S. An efficient data mining method for learning Bayesian networks using an evolutionary algorithm-based hybrid approach. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(4): 378–404
- 3 Pan Zheng-Jun, Kang Li-Shan, Chen Yu-Ping. *Evolutionary Computation*. Beijing: Tsinghua Press, 1998. 111–123 (潘正军, 康立山, 陈毓屏. 演化计算. 北京: 清华大学出版社, 1998. 111–123)
- 4 Gallagher J C, Vignraham S, Kramer G. A family of compact genetic algorithm for intrinsic evolvable hardware. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(2): 111–126
- 5 Zhong Wei-Cai, Liu Jing, Jiao Li-Cheng. Optimal approximation of linear systems by multi-agent genetic algorithm. *Acta Automatica Sinica*, 2004, **30**(6): 933–938 (钟伟才, 刘静, 焦李成. 多智能体遗传算法用于线性系统逼近. 自动化学报, 2004, **30**(6): 933–938)
- 6 Kim J, Bentley P J. Towards an artificial immune system for network intrusion detection: an investigation of dynamic clone selection. In: Proceedings of Congress on Evolutionary Computation. Hawaii, USA: IEEE, 2002. 1015–1020
- 7 Du H F, Gong M G, Jiao L C. A novel algorithm immune system for high-dimensional function numerical optimization. *Progress in Natural Science*, 2005, **15**(5): 463–471
- 8 Gong Mao-Guo, Du Hai-Feng, Jiao Li-Cheng. Approximation of linear systems based on artificial immunity response. *Science in China Series E: Information Sciences*, 2005, **35**(12): 1288–1303 (公茂果, 杜海峰, 焦李成. 基于人工免疫响应的线性系统逼近. 中国科学 E 辑: 信息科学, 2005, **35**(12): 1288–1303)

张 伟 第二炮兵工程学院指挥自动化系副教授, 西安电子科技大学智能信息处理研究所博士研究生。主要研究方向为人工智能、故障诊断。本文通信作者。E-mail: zhang1966wei@163.com  
(ZHANG Wei Associate professor in Department of Command Automation at Second Artillery Engineering Institute, Ph.D. candidate at Xidian University. Her research interest covers artificial intelligence and fault diagnosis. Corresponding author of this paper.)

胡昌华 第二炮兵工程学院自动控制系教授。主要研究方向为故障诊断、可靠性工程。E-mail: hch6603@263.net  
(HU Chang-Hua Professor in Department of Automation at Second Artillery Engineering Institute. His research interest covers fault diagnosis and reliability engineering.)

焦李成 西安电子科技大学智能信息处理研究所教授。主要研究方向为进化计算、智能信息处理。E-mail: lchjiao@mail.xidian.edu.cn  
(JIAO Li-Cheng Professor at Institute of Intelligent Information Processing, Xidian University. His research interest covers evolutionary computation and intelligent information processing.)

尚荣华 西安电子科技大学智能信息处理研究所博士研究生。主要研究方向为进化计算。E-mail: rhshang@mail.xidian.edu.cn  
(SHANG Rong-Hua Ph.D. candidate at Institute of Intelligent Information Processing, Xidian University. Her main research interest is evolutionary computation.)