文章编号：1000－8152(2005)05－0810－04

基于量子遗传算法的特征选择算法

张葛祥[1]，金炜东[1]，胡来招[2]
（1. 西南交通大学 电气工程学院，四川 成都 610031；2. 中国电子科技集团第29研究所，四川 成都 610036）

摘要：特征选择是模式识别和机器学习等领域中重要而困难的研究课题.提出一种最优特征子集评价准则和实现特征选择的一种新量子遗传算法(NQGA).NQGA采用量子门旋转角更新新方法和增强算法寻优能力及防止早熟收敛的移民和灾变策略.定性分析了NQGA的高效性.典型复杂函数测试和雷达辐射源信号特征选择的应用表明，NQGA寻优能力强、收敛速度快和能有效防止早熟现象.采用提出的准则函数和搜索策略实现特征选择，大大降低了特征维数，获得了更高的正确识别率.

关键词：遗传算法；特征选择；量子理论；量子遗传算法
中图分类号：TP18，TP301.6　　**文献标识码**：A

Feature selection algorithm based on quantum genetic algorithm

ZHANG Ge-xiang[1], JIN Wei-dong[1], HU Lai-zhao[2]
(1. School of Electrical Engineering, Southwest Jiaotong University, Chengdu Sichuan 610031, China;
2. China Electronics Technology Group Corporation No.29 Research Institute, Chengdu Sichuan 610036, China)

Abstract: Feature selection is always an important and difficult issue in pattern recognition and machine learning. This paper proposed a criterion function for selecting the optimal feature subset and a search strategy called novel quantum genetic algorithm (NQGA). NQGA adopted a novel update approach of rotation angles of quantum gates, and immigration and catastrophe operations to enhance search capability and to avoid premature convergence. Besides, high efficiency of NQGA was analyzed qualitatively. Testing results of typically complex functions and experimental results of feature selection in radar emitter signal recognition show that NQGA has good characteristics of strong search capability, rapid convergence and no premature convergence. The proposed feature selection algorithm reduces greatly the dimensions of original feature set and heightens accurate recognition rate of radar emitter signals.

Key words: genetic algorithm; feature selection; quantum theory; quantum genetic algorithm

1 引言(Introduction)

量子遗传算法(Quantum genetic algorithm, QGA)是量子计算理论和遗传算法原理相结合的产物，是一种具有勃勃生机和应用潜力的新遗传算法.QGA以量子计算原理[1,2]为基础，用量子位编码表示染色体，通过量子门作用实现进化搜索.QGA因具有种群规模小、寻优能力强、收敛速度快和计算时间短的特点而受到极大关注[3~7].

Narayanan等[8]提出了量子衍生遗传算法(QIGA)并成功求解了TSP问题，虽然QIGA仍属传统遗传算法(CGA)，但激发了量子计算原理与遗传算法结合的研究.Han等[3,4]采用量子位编码和量子门更新染色体，提出了遗传量子算法(GQA)和并行量子遗传算法(PQGA)，并成功求解了组合优化问题.李斌[5]采用量子交叉[8]和量子变异改进GQA[3]并实现频繁结构模式的发掘；杨俊安等[6]基于多量子位编码和量子门旋转角动态调整来改进GQA[3]并用于图像的盲源分离；李映等[7]用混合PQGA实现图像的边缘检测.这些结果[3~7]均表明GQA的性能大大优于CGA.可是，GQA的更新策略是在对问题最优解情况有所了解的前提下设计的，但在连续函数优化和实际应用中，最优解情况是未知的.

本文提出一种基于新量子遗传算法(NQGA)的特征选择算法.NQGA采用一种新的量子门旋转角更新策略，并引入移民和灾变策略来增强搜索能力和避免早熟收敛.复杂连续函数测试表明NQGA明

收稿日期：2003－06－24；收修改稿日期：2004－02－23.
基金项目：电子对抗技术预研基金项目(NEWL51435QT220401)；国家自然科学基金资助项目(69574026)；西南交通大学博士生创新基金资助项目(2003).

显优于 GQA.同时,本文提出一种新的最优特征子集评价准则,基于该准则及 NQGA,本文实现了雷达辐射源信号的特征选择.实验结果表明,采用选择出的 3 维特征获得了比原来 16 维特征更高的识别率,而且 NQGA 的性能也优于 GQA,表明了本文算法的有效性和实用性.

2 新量子遗传算法(Novel quantum genetic algorithm)

两态量子计算机存储的最小信息单元称为一个量子比特或量子位.量子位既可表示"0"和"1"两种状态,又可表示它们的任意线性叠加态,即

$$|\varphi\rangle = \alpha \cdot |0\rangle + \beta \cdot |1\rangle. \tag{1}$$

其中,$|0\rangle$ 和 $|1\rangle$ 分别表示自旋向下和自旋向上两种状态,α 和 β 是两个复常数,分别表示 $|0\rangle$ 和 $|1\rangle$ 的概率幅,且满足下面归一化条件:

$$|\alpha|^2 + |\beta|^2 = 1. \tag{2}$$

量子位的概率幅是指满足(1)和(2)式的一对数 α 和 β.量子位的相位是指满足(3)式的角度 $\zeta(\zeta \in [-\pi/2, \pi/2])$,即

$$\zeta = \arctan(\beta/\alpha). \tag{3}$$

数 α 和 β 的乘积用 d 表示,d 的正负值表示量子位相位 ζ 在平面坐标系中所处的象限,如果 d 为正,表示 ζ 处于一和三象限,否则处于二和四象限.

NQGA 的算法描述如下:

Ⅰ)初始化:确定种群大小 n 和量子位的数目 m,包含 n 个个体的种群 $P = \{p_1, p_2, \cdots, p_n\}$,其中 $p_j(j = 1,2,\cdots,n)$ 为种群中的第 j 个个体,其描述如下:

$$p_j = \left[\begin{array}{c|c|c|c} \alpha_{j1} & \alpha_{j2} & \cdots & \alpha_{jm} \\ \beta_{j1} & \beta_{j2} & \cdots & \beta_{jm} \end{array}\right]. \tag{4}$$

其中,所有的 $\alpha_{ji}, \beta_{ji}(i = 1,2,\cdots,m)$ 均取为 $1/\sqrt{2}$,表示在初始搜索时所有状态均以相同的概率进行叠加,进化代数初始值 g 设为 0;

Ⅱ)根据 P 中各个体的概率幅构造出量子叠加态的观测态 R,$R = \{a_1, a_2, \cdots, a_n\}$,其中 $a_j(j = 1,2,\cdots,n)$ 为每个个体的观测状态,且为一个长度为 m 的二进制串,即 $a_j = b_1 b_2 \cdots b_m$,其中 $b_k(k = 1,2,\cdots,m)$ 为一位二进制数"0"或"1".采用概率方式产生观测态,具体过程为:对于 P 中的每一个量子位的概率幅 $[\alpha_i, \beta_i]^{\mathrm{T}}(i = 1,2,\cdots,n \times m)$,随机产生 0 与 1 之间的一个数 r,若 $r < |\alpha_i|^2$,则相应的观测值 b 为"0",否则,相应的观测值 b 为"1".在NQGA算法中,由概率幅 P 构造观测态 R 的过程包含算法的解码过程,解码后便得到各优化参数的当前实际值;

Ⅲ)用适应度函数评价种群中的所有个体;

Ⅳ)保留最佳个体,并将其与保留的此代以前的最佳个体进行比较,保留两者中较优者,同时判断算法是否满足最大进化代数,若满足,则算法终止,否则,执行下一步;

Ⅴ)量子逻辑门选为量子旋转门 G,即

$$G = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{5}$$

其中,θ 为量子门的旋转角,$\theta = k \cdot h(\alpha,\beta)$,$k$ 是与算法收敛速度有关的系数,其值必须合理选取,k 值太大,算法易陷入局部极值而出现早熟现象;k 值太小,算法搜索的速度就很慢,甚至会处于停滞状态.若将 k 定义为变量更合理一些,但 k 又不能像在 CGA 中那样把自适应参数定义为最大适应值和平均适应值的函数,因为这会大大增加NQGA的计算时间,故将 k 定义为与进化代数和优化问题复杂性有关的变量,以便自适应地调整搜索网格的大小,本文 k 取为 $0.5e^{-t/\max t}$,t 为进化代数 g 除以种群发生灾变的代数 C_g 的余数,$\max t$ 为最大进化代数.$h(\alpha,\beta)$ 是搜索方向函数,其作用是使算法在解空间中朝着最优解的方向搜索,$h(\alpha,\beta)$ 的值如表 1 所示.于是,可计算出量子门的旋转角,并将其作用于种群中的所有个体的概率幅以更新 P,即

$$p_j^{t+1} = G(t) \cdot p_j^t. \tag{6}$$

其中,上标 t 为进化代数,$G(t)$ 为第 t 代的量子旋转门,p_j^t 和 p_j^{t+1} 分别为第 t 代和 $t+1$ 代的第 j 个个体的概率幅.

表 1　函数 $h(\alpha,\beta)$ 的查询表

Table 1　Look-up table of function $h(\alpha,\beta)$

$d_1 > 0$	$d_2 > 0$	$h(\alpha_i,\beta_i)$	
		$\lvert\zeta_1\rvert > \lvert\zeta_2\rvert$	$\lvert\zeta_1\rvert < \lvert\zeta_2\rvert$
True	True	+1	−1
True	False	−1	+1
False	True	−1	+1
False	False	+1	−1

注　在表 1 中,α_1,β_1 是最优解的概率幅,$d_1 = \alpha_1 \cdot \beta_1$,$\zeta_1 = \arctan(\beta_1/\alpha_1)$,$\alpha_2,\beta_2$ 是当前解的概率幅,$d_2 = \alpha_2 \cdot \beta_2$,$\zeta_2 = \arctan(\beta_2/\alpha_2)$.

Ⅵ)种群每进化 M_g 代进行一次移民操作,引入优质基因,使算法易于搜索到最优解,同时也可能使算法摆脱局部极值,具体操作为:采用种群初始化的方法重新产生一代种群,但其中每个 α 取 0 与 1 之间的随机值,相应的 β 为 $\pm\sqrt{1-|\alpha|^2}$,然后采用概率

方式将种群中最佳个体的某几位概率幅取代Ⅳ)中最佳个体的相应位置的概率幅;

Ⅶ) 保留的最佳个体连续 C_g 代都无变化,表明算法处于搜索停滞状态或陷入局部极值,需要实施灾变操作使其尽快摆脱进化迟钝状态或跳出局部极值,但灾变操作不能使种旋退化,故采用这样的灾变操作:采用Ⅵ)中的方法产生新种群并选出最佳个体来完善替代Ⅳ)中保留的最佳个体;

Ⅷ) 进化代数 g 增1,算法转至Ⅱ)继续执行,直到算法结束.

NQGA具有寻优能力强、收敛速度快、计算时间短和能有效防止早离收敛的特点,其原因在于:

i) 采用量子位编码方式,种群中的各个体在进化过程中始终携带着不同叠加态的信息,能有效保持种群的多样性和避免选择压力问题;

ii) 量子旋转门的旋转角自适应变化,旋转角先由大到小逐渐变化,开始旋转角变化大,便于在解空间中大范围搜索,随着进化代数的增加,旋转角逐渐减小,便于局部搜索,而每当发生种群灾变后,量子旋转门的旋转角又会增大,然后又逐渐减小,因此,此种更新方法简单有效,适用性大.

iii) 移民操作能将优良品质基因引入到种群中,使算法易于搜索到最优解,同时也可能使算法摆脱局部极值.灾变操作能使算法尽快摆脱进化迟钝状态或跳出局部极值.

3 性能测试(Performance test)

选取的典型复杂连续函数为:

1) 多峰函数:

$$f_1 = 10 + \frac{\sin(1/x)}{(x-0.16)^2+0.1},\ x \in [0.01,1].$$

函数在定义域内有无穷多个局部极值,全局极大值为 $f_1(0.1275) = 19.8949$.

2) 二维多模态函数:

$$f_2 = \cos(2\pi x_1)\cos(2\pi x_2)e^{-(x_1^2+x_2^2)/10}.$$

函数 f_2 是著名的多模态测试函数,其定义域为 $-1 \leqslant x_1, x_2 \leqslant 1$,函数在定义域内有13个起伏变化、函数值相近的模态峰点,最优点为 $f_2(0,0) = 1$.

测试中,NQGA和GQA的种群大小均为10,量子位数目为15,最大进化代数为1000,误差均为0.0001,NQGA的 M_g 和 C_g 分别取为40和100.表2给出了100次测试的统计结果.

表2 算法性能测试结果

Table 2 Results of performance tests of algorithms

函数	算法	平均代数	平均时间/s	成功率
f_1	GQA	177.17	3.5289	87.00%
	NQGA	139.00	3.2617	100.0%
f_2	GQA	358.21	13.2192	76.00%
	NQGA	249.07	10.7918	100.0%

由表2知,NQGA的平均搜索代数比GQA少,函数越复杂,差别越明显;从成功率看,NQGA明显优于GQA,对于 f_2,GQA的成功率仅为76%,而NQGA的成功率为100.0%,这说明算法中引入移民和灾变操作是有效的,使算法及时跳出了局部极值;NQGA的平均计算时间虽比GQA少,但与平均代数的差别相比并不太明显,这主要是由于在NQGA中引入了移民和灾变操作所致.

4 特征选择(Feature selection)

特征选择是一类典型的组合优化问题,除了搜索策略,最优特征子集的评价准则也是其关键问题.下面给出本文的评价准则函数.

将第 i 类雷达辐射源信号的类内聚集度 C_{ii} 定义为

$$C_{ii} = \max\left\{\left[\frac{1}{M_i^q}\sum_{k=1}^{M_i^q} \| x_{ik}^q - \mathrm{E}(X_i^q) \|^p\right]^{\frac{1}{p}}\right\}. \quad (7)$$

式中,$q = 1,2,\cdots,N$,N 是特征数,M_i^q 是第 i 类雷达辐射源信号第 q 种特征的样本数,x_{ik}^q 是第 i 类雷达辐射源信号第 q 种特征的第 k 个样本值,$X_i^q = [x_{i1}^q, x_{i2}^q, \cdots, x_{iM_i^q}^q]$,$\mathrm{E}(X_i^q)$ 是 X_i^q 的期望值,p 是大于1的整数.于是,可得到第 j 类雷达辐射源信号的类内聚集度 C_{jj}.

第 i 类和第 j 类雷达辐射源信号的距离可定义为 $D_{ij} = \min\{\| \mathrm{E}(X_i^q) - \mathrm{E}(X_j^q) \|\}$.这样,第 i 类与第 j 类雷达辐射源信号的类间分离度可定义为

$$S_{ij} = \frac{D_{ij}}{C_{ii} + C_{jj}}. \quad (8)$$

如果待识别的雷达辐射源信号共有 H 类,特征选择的准则函数定义为

$$f = \frac{2}{H(H-1)}\sum_{i=1}^{H-1}\sum_{j=i+1}^{H} S_{ij}. \quad (9)$$

显然,f 的值越大,选择出的特征子集就越好.

从10类雷达辐射源信号中提取出16种特征,分别用GQA和NQGA进行特征选择,种群大小均取10,量子位数均取16,适应度函数如式(9)所示,最大进化代数为1000,NQGA的 M_g 和 C_g 分别取为

40 和 100,量子观测态用 16 位二进制数表示,当某一位为"1"时,表示此种特征被选中,反之,当这一位为"0"时,表示此种特征未被选中.算法在 CPU 为 2GHz、内存为 512M 的 P-IV 计算机上实现.用 GQA 和 NQGA 运行 50 次后的各次适应度值分别如图 1 和图 2 所示.运行 50 次后的统计结果如表 3 所示,表中 AG 表示平均搜索代数,AT 表示平均计算时间(单位为 s),AF 表示平均适应值,MF 表示搜索到的最大适应值.

GQA 和 NQGA 搜索到的最优特征子集均由 3 维特征构成,最大适应值相同.图 1 和图 2 显示,NQGA 达到最大适应值的次数明显多于 GQA.在表 3中,NQGA 的平均适应值明显高于 GQA,但平均搜索代数略多于 GQA,相应的计算时间也大于 GQA,这一点也恰好说明 NQGA 中的移民和灾变操作起了作用.

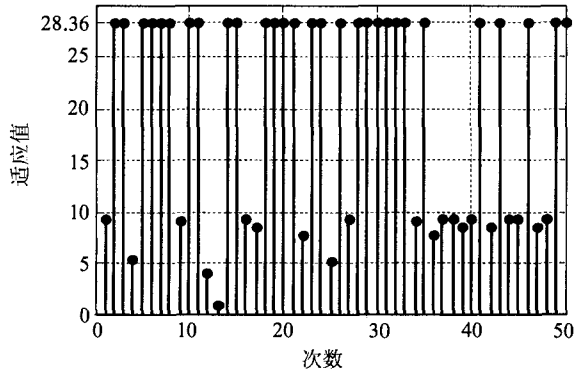

图 1　用 GQA 运行 50 次的适应值

Fig. 1　Fitness values of 50 runs using GQA

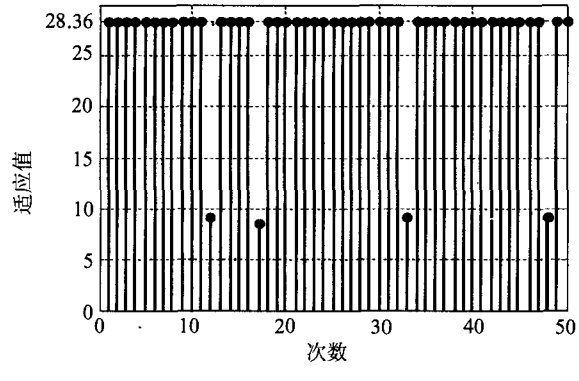

图 2　用 NQGA 运行 50 次的适应值

Fig. 2　Fitness values of 50 runs using NQGA

表 3　GQA 和 NQGA 的统计结果

Table 3　Statistical results of GQA and NQGA

算法	AG	AT	AF	MF
GQA	465.78	19.77	19.77	28.36
NQGA	486.84	26.08	26.81	28.36

本文还设计了神经网络(NN)分类器来验证选出的特征子集的性能,NN 分类器的结构为 3 - 15 - 10,隐层和输出层神经元函数分别为"tansig"和"logsig",误差为 0.05.为了对比,选择前的特征集也用来进行分类识别,其分类器的结构为 16 - 25 - 10.分类器的训练样本均取 50,测试样本均取 1000,选择前后的特征集的结构结果如表 4 所示,表中的数值为百分比数值.表 4 显示,选择前后的平均识别率分别为 96.21% 和 99.21%.由此可知,特征选择不仅省去了大量无用特征提取的开销,简化了分类器的设计,而且获得了比特征选择前更高的识别率.

表 4　选择前后特征集的识别率

Table 4　Recognition rates using different feature sets

类型	BPSK	QPSK	MPSK	LFM	NLFM
选择前	97.90	78.17	100	90.87	100
选择后	99.01	93.75	100	100	100
类型	CW	FD	FSK	IPFE	CSF
选择前	98.47	98.78	97.91	100	91.67
选择后	100	100	100	100	100

5　结束语(Concluding remarks)

最优特征子集的评价准则和搜索算法一直是特征选择算法研究的主要内容,本文给出了一种有效的评价准则和一种高效的搜索算法.通过典型复杂函数优化测试和雷达辐射源信号特征选择实例表明,本文提出算法的性能优于文献[3]中的算法.该算法大大降低了雷达辐射源信号识别中的特征维数,获得了比原特征集更高的正确识别率.

参考文献(References):

[1] NARAYANAN A. Quantum computing for beginners [C]// *Proc of the* 1999 *Congress on Evolutionary Computation*. Piscataway: IEEE Press, 1999: 2231 - 2238.

[2] GROVER L. Quantum computing [C]// *Proc of the* 12*th Int Conf on VLSI Design*. Piscataway: IEEE Pess, 1999: 548 - 553.

[3] HAN Kuk-Hyun, KIM Jong-hwan. Genetic quantum algorithm and its application to combinatorial optimization problems [C]// *Proc of the* 2000 *IEEE Conf on Evolutionary Computation*. Piscataway: IEEE Press, 2000: 1354 - 1360.

[4] HAN Kuk-Hyun, PARK Kui-Hong, LEE Chi-Lee, et al. Parallel quantum-inspied genetic algorithm for combinatorial optimization problms [C]// *Proc of IEEE Conf on Evolutionary Computation*. Piscataway: IEEE Press, 2001: 1442 - 1429.

(下转第 819 页)

Springer-Verlag,1985.

[6] SLOTINE J, LI W. *Applied Nonlinear Control* [M]. Englewood Cliffs, New Jersey: Prentice-Hall Inc, 1991.

作者简介:

宾　洋　(1976—),男,博士研究生,研究领域为鲁棒非线性系统控制,车辆 Stop and Go 巡航控制,E-mails: biny02@mails.tsinghua.edu.cn;

李克强　(1963—),男,工学博士,博士生导师,1992 年 ~ 1994 年在日本五十铃汽车公司车身技术中心作客座研究员,1994 年 ~ 1997 年为重庆大学汽车工程系教授,1997 年 ~ 1998 年在日本东京农工大学车辆动力学与控制研究室作访问学者,1998 年 ~ 2000 年在日本国立交通安全与公害研究所工作,2000 年至今,清华大学教师、STA Fellow,2003 年 ~ 2004 年在德国亚琛工业大学(RWTH-Aachen)IKA 作访问教授,研究兴趣为智能汽车与智能交通系统,混合动力电动汽车(HEV)整车控制系统,车辆噪声振动分析与控制;

连小珉　(1955—),男,工学博士,博士生导师,1982 年 ~ 1983 年在四川省交通科学研究所工作,1986 年至今,清华大学教师,1988 年 ~ 1990 年,日本五十铃汽车公司的研究人员,研究领域为汽车噪声与振动控制,GPS 汽车导航,智能交通系统,计算机辅助设计,计算机辅助测试技术.

(上接第 809 页)

参考文献(References):

[1] 李洪兴.变论域自适应模糊控制器[J].中国科学(E 辑),1999,29(1):32 - 42.
(LI Hongxin. Adaptive fuzzy controller based on variable universe [J]. *Science in China* (*Series E*), 1999, 29(1): 32 - 42.)

[2] 李洪兴.一类高精度模糊控制器的设计[J].控制理论与应用,1997,14(6):868 - 876.
(LI Hongxin. Design on a class of high-accuracy fuzzy controller [J]. *Control Theory & Applications*, 1997, 14(6): 868 - 876.)

[3] 李洪兴,苗志宏,王加银.四级倒立摆的变论域自适应模糊控制[J].中国科学(E 辑),2002,32(1):65 - 75.
(LI Hongxin, MIAO Zhihong, WANG Jiayin. Variable universe adaptive fuzzy control on the quadruple inverted pendulum [J]. *Science in China* (*Series E*). 1999, 29(1): 32 - 42.)

[4] 汪荣鑫.随机过程[M].西安:西安交通大学出版社.1993.
(WANG Rongxing. *Random Process* [M]. Xi'an: Xi'an Jiaotong University Press, 1993.)

[4] 李少远.模糊滑动模态控制系统的性质分析[J].控制理论与应用,2000,17(1):14 - 18.
(LI Shaoyuan. Analysis of property of fuzzy sliding mode control [J]. *Control Theory & Applications*, 2000, 17(1): 14 - 18.)

作者简介:

岳士弘　(1964—),男,博士,副教授,从事数据挖掘,数据融合,模糊控制和优化理论等工程应用研究,E-mail: shyue1999@tju.edu.cn;

张绍杰　(1972—),男,博士生,从事控制理论及其程应用研究;

李　平　(1954—),男,教授,博士生导师,浙江大学工业控制技术研究所所长,从事控制理论与应用的研究.

(上接第 813 页)

[5] 李斌.金融时间序列数据挖掘关键算法研究[D].合肥:中国科学技术大学,2001.
(LI Bin. *The main algorithm research on financial time series data mining* [D]. Hefei: University of Science and Technology of China, 2001.)

[6] YANG Junan, LI Bin, ZHUANG Zhenquan. Research of quatntum genetic algorithm and its application in blind source separation [J]. *J of Electronics*, 2003, 20(1): 62 - 68.

[7] 李映,焦李成.一种有效的基于并行量子进化算法的图像边缘检测方法[J].信号处理.2003,19(1):69 - 74.
(LI Ying, JIAO Licheng. An effective method of image edge detection based on parallel quantum evolutionary algorithm [J]. *Signal Processing*, 2003, 19(1): 69 - 74.)

[8] NARAYANAN A, MOORE M. Quantum-inspired genetic algorithm [C]// *Proc of IEEE Int Conf on Evolutionary Computation*. Piscataway: IEEE Press, 1996: 61 - 66.

作者简介:

张葛祥　(1974—),男,博士研究生,研究领域为进化计算、雷达辐射源信号处理、优化理论与优化控制、神经网络等,E-mail: dylan7237@sina.com;

金炜东　(1959—),男,博士,教授,博士生导师,研究领域为优化理论与优化控制、智能信息处理、系统仿真等;

胡来招　(1945—),男,博士,研究员,博士生导师,研究领域为信号处理、侦察接收机、无源定位等.