

交互式遗传算法的噪声及降噪策略

周 勇^{1,2}, 巩敦卫²

(1. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221008;

2. 中国矿业大学 信息与电气工程学院, 江苏 徐州 221008)

摘要: 在交互式遗传算法中, 人对进化个体的评价含有噪声. 如何降低噪声对评价的不利影响, 至今没有有效的方法. 这严重制约了交互式遗传算法在复杂优化问题中的广泛应用. 本文首先针对交互式遗传算法中人评价个体适应值的不确定性和漂移性, 分析交互式遗传算法噪声的来源, 定义认知评价度和疲劳评价度, 给出交互式遗传算法的3阶段噪声模型; 然后, 给出基于个体海明距离的认知评价度和疲劳评价度刻画以及基于适应值可信度的降噪策略; 最后, 通过在服装设计中的应用实例验证噪声模型的正确性和降噪策略的有效性.

关键词: 遗传算法; 交互; 噪声; 降噪策略

中图分类号: TP18 **文献标识码:** A

Noises in interactive genetic algorithms and strategies for its reduction

ZHOU Yong^{1,2}, GONG Dun-wei²

(1. School of Computer Science and Technology, China University of Mining and Technology,
Xuzhou Jiangsu 221008, China;

2. School of Information and Electrical Engineering, China University of Mining and Technology,
Xuzhou Jiangsu 221008, China)

Abstract: Noises exist in human's subjective evaluation on individuals in interactive genetic algorithms. There is no effective way to alleviate the influence of noises on evaluation, which restricts broad applications of interactive genetic algorithms to complicated optimization problems. In this paper, sources of noises in interactive genetic algorithms are analyzed in relation with uncertainties and fluctuation in human's subjective evaluation on individuals; the degree of evaluating cognition and the degree of evaluating fatigue are defined; and a model for noises with 3 phases in interactive genetic algorithms is proposed. Then, descriptions of the degree of evaluating cognition and the degree of evaluating fatigue based on individuals' Hamming distance, and a strategy for reducing noises based on fitness reliabilities are presented. Finally, the efficiencies of the model for noises and the strategy of reducing noises are validated through the application examples in fashion design.

Key words: genetic algorithms; interaction; noise; strategy of reducing noises

1 引言(Introduction)

交互式遗传算法是20世纪90年代中期由日本学者Takagi提出的一种基于人的主观评价得到进化个体适应值的进化计算方法^[1]. 该方法将传统进化计算与人的智能评价有机结合, 是一种解决隐式性能指标优化问题的有效方法, 已经广泛应用于艺术设计、动画设计、音乐创作、语音处理、虚拟现实、知识获取与数据挖掘、控制与机器人等多领域^[2,3]. 由于交互式遗传算法中的个体适应值由人主观评价所得, 人的疲劳会引起个体适应值评价的波动, 从而形

成了个体适应值的噪声. 因此, 研究交互式遗传算法中的噪声问题和降噪策略对解决由人的疲劳和人的认知快慢程度引起的进化过程的波动是十分有意义的.

已有研究表明, 一般遗传算法中噪声的来源主要有3个方面: 一是决策变量; 二是个体适应值的量测; 三是优化问题本身^[4~6]. 噪声对整个进化过程的影响主要表现在: 进化过程的学习效率低; 无法正确保留已学习的信息; 决策变量空间的开发和搜索能力受限种群适应值不能随进化代数增加而增长^[7,8].

一般在线优化遗传算法中的噪声补偿策略主要有以下3种^[8,9]:一是增大种群规模;二是增加采样次数,然后取多次采样的平均值;三是改进遗传算子。因此,一般遗传算法的噪声补偿策略主要是通过重复采样技术提高个体适应值的精度,但交互式遗传算法中人是宝贵的资源,且人的疲劳问题直接影响交互式遗传算法的优化结果,重复对同一进化个体进行评价必然造成人的疲劳,影响进化的效果。所以,一般遗传算法中的噪声补偿策略不适用于交互式遗传算法,必须根据交互式遗传算法自身的特点和噪声的分布形式研究其特点的降噪策略。

齐岩等指出用户在不同的时间对个体的评价价值会不同^[10],蒋姗姗等指出用户评价具有一定偏好^[11],Gong等给出了用户偏好的漂移模型^[12],但以上文献都没有给出交互式遗传算法的噪声模型,也没有分析这种噪声对个体适应值评价的影响。针对这些问题,本文给出交互式遗传算法的噪声模型和降噪策略,并通过服装设计实例验证该模型的正确性和降噪策略的有效性。

2 交互式遗传算法的噪声模型(Models for noises in interactive genetic algorithms)

本文考虑如下一般优化问题:

$$\begin{cases} \max f(x), \\ x \in S \subseteq \mathbb{R}^n, \\ F(x) = f(x) + \delta(t). \end{cases} \quad (1)$$

式中: x 是 D 维决策变量, S 是其取值范围, $f(x)$ 是被优化的性能指标,且不能用显式函数表示, $\delta(t)$ 是噪声。在不引起混乱的情况下,仍记进化个体和搜索空间分别为 x 和 S ,人在第 t 代赋予 x 的适应值为 $F(x)$ 。

2.1 交互式遗传算法的噪声来源(Sources of noises in interactive genetic algorithms)

交互式遗传算法的噪声主要来源于两个方面:其一是由人对评价对象认知不充分导致的进化个体适应值漂移;其二是由人的疲劳导致的进化个体适应值波动。

进化个体适应值噪声强度随人对评价对象认识的程度而变化。当人评价一定数量的个体后,对评价对象已经认识得比较全面,形成了稳定的评价标准。此时,人给出的适应值仅含有很小的随机噪声,但基本可以真实反映待评价个体的优劣程度。随着种群进化代数的增加,人产生了疲劳,从而影响个体评价的客观性。

由上述分析可以看出,交互式遗传算法的噪声具有阶段性、非结构性以及空间依赖性。

2.2 认知评价度(Degree of cognition)

由人认识事物的常识可知,人对评价对象的熟悉程度随着已评价进化个体数量的增多而逐渐加深;

当人评价的进化个体数量达到某一阈值时,已完全熟悉该评价对象。这说明,评价进化个体的数量达到该阈值前,人对评价对象的熟悉程度应正比于其评价的个体数量。

定义 1 人的认知评价度是一个数值,当人对某对象评价的个体数量大于该值时,则人已经完全熟悉了评价对象。记人的认知评价度为 N_c 。

2.3 疲劳评价度(Degree of fatigue)

假设人对每一进化个体的评价时间都相同。在种群进化后期,当评价的进化个体数量达到某一阈值时,人产生了疲劳。此后,人给出的适应值存在某种程度的波动,也不能客观反映个体优劣。而且,人的疲劳度随已评价进化个体数量的增多而加深,这使得适应值波动的强度也不断变。

定义 2 人的疲劳评价度是一个数值,当人对某对象评价的个体数量大于该值时,已经开始疲劳。记人的评价疲劳度为 N_f 。

注 1 对于非常复杂的评价对象,人一时很难对其熟悉甚至当人已经产生了疲劳,还没有完全熟悉评价对象,此种情况较复杂,不在本文的考虑范围之内。所以本文中 $N_f > N_c$ 。

2.4 交互式遗传算法的噪声模型(Models for noises in interactive genetic algorithms)

根据以上分析和模型定义,设种群规模为 N ,若人当前评价的进化个体为第 t 代的第 $i+1$ 个体,记 N_e 为人已评价的进化个体数量,则有 $N_e = (t-1) \cdot N + i$,则交互式遗传算法的噪声 $\delta(t)$ 可表示为如下关于 N_e 的函数:

$$\delta(N_e) = \begin{cases} k_1 \cdot (N_c - N_e) \cdot e^{-N_c}, & N_e < N_c, \\ \varepsilon \cdot N(0, 1), & N_c \leq N_e \leq N_f, \\ k_2 \cdot e^{(N_e - N_f)} \cdot N(0, 1), & N_e \geq N_f. \end{cases} \quad (2)$$

其中: k_1, k_2 为相应阶段的调节系数, ε 为 $(0, 1)$ 之间的小数,表示噪声强度的临界值, $N(0, 1)$ 为标准正态分布噪声。如图1所示,交互式遗传算法的噪声具体可以分为认知阶段、中间阶段以及疲劳阶段3个阶段。

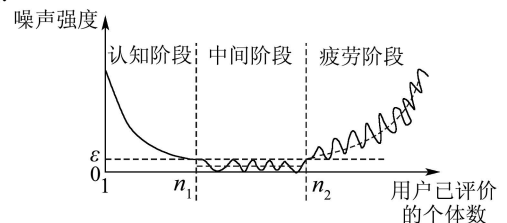


图1 噪声强度曲线

Fig. 1 Curve of noise intensity

各个阶段的出现与评价对象的复杂程度和评价者对评价对象的先验知识有关。但是,考虑问题的一

般性, 图1给出了含有3阶段的噪声模型. 图中的 ε 为噪声强度的临界值, 当个体适应值的噪声强度小于 ε 时, 则认为系统处于中间阶段.

3 交互式遗传算法的降噪策略(Strategies of reducing noises in interactive genetic algorithms)

3.1 算法思想(Methodology of algorithms)

根据交互式遗传算法进化个体适应值噪声的特点, 在不同阶段采用不同策略, 调整人赋予的进化个体适应值, 将调整后的适应值参与遗传操作, 以最大限度地降低适应值噪声对种群进化的影响.

3.2 适应值的可信度(Degree of reliabilities of fitness)

为了降低适应值噪声对选择操作的影响, 在2.4节噪声模型的基础上, 本文对每一阶段个体的适应值评价过程都确定一个可信度函数. 适应值的可信度值随着进化过程而不断变化, 则可信度函数 $R(N_e)$ 可表示为

$$R(N_e) = \begin{cases} 1 - k_1 \cdot (N_c - N_e) \cdot e^{(-N_e)}, & N_e < N_c, \\ 1, & N_c \leq N_e \leq N_f, \\ k_2 \cdot e^{-(N_e - N_f)} \cdot N(0, 1), & N_e > N_f. \end{cases} \quad (3)$$

3.3 基于个体海明距离的认知评价度和疲劳评价度刻画(Descriptions of degree of cognition and degree of fatigue based on individuals' Hamming distance)

对于相同或相似的个体, 如果人评价给出的适应值相差很小或基本相同, 则可以认为人评价的适应值可信度较高或者噪声很小, 此时人的评价过程处于中间阶段; 如果人评价给出的适应值相差很大, 则可以认为人评价的适应值可信度较低或者噪声很大, 此时人的评价过程处于认知阶段或者疲劳阶段.

假设个体采用二进制编码, 个体的编码长度为 ℓ , 则第 t_1 代的第 i 个体 $x_i(t_1)$ 和第 t_2 代的第 j 个体 $x_j(t_2)$ 的海明距离为

$$H(x_i(t_1), x_j(t_2)) = \sum_{k=1}^{\ell} |x_{ik}(t_1) - x_{jk}(t_2)|. \quad (4)$$

其中 $x_{ik}(t_1)$ 表示 $x_i(t_1)$ 的第 k 个基因. 人赋予这两个个体的适应值之差为

$$\Delta x_j(t_2) = |F(x_i(t_1)) - F(x_j(t_2))|. \quad (5)$$

定义3 认知偏差是由于人不熟悉进化个体或者已经疲劳导致的评价进化个体产生的偏差, 数学表示为

$$\delta(x_i(t_1), x_j(t_2)) = \left| \frac{1}{\ell} H(x_i(t_1), x_j(t_2)) - \frac{1}{F_{\max}} \Delta F(x_i(t_1), x_k(t_2)) \right|. \quad (6)$$

其中 $F_{\max}$ 为进化个体适应值的上限.

考虑评价个体 $x_m(t)$, 其前面相邻的 K 个个体依次为 $x_1, x_2, \dots, x_k$, 则 $x_m(t)$ 与 $x_1, x_2, \dots, x_k$ 的认知偏差分别为

$$\delta(x_m(t), x_k) = \left| \frac{1}{\ell} H(x_m(t), x_k) - \frac{1}{F_{\max}} \Delta F(x_m(t), x_k) \right|, \quad k = 1, 2, \dots, K. \quad (7)$$

其中 K 为认知偏差的跟踪深度 $K \in \{1, 2, \dots, N\}$.

记 δ_e 为认知偏差阈值, 若 $\forall k \in \{1, 2, \dots, K\}$, 均满足 $\delta(x_m(t), x_k) \leq \delta_e$, 则认知过程进入中间阶段, 认知评价度为 $N_c = (t-1)N + m$; 若 $\forall k \in \{1, 2, \dots, K\}$, 均满足 $\delta(x_m(t), x_k) \geq \delta_e$, 则认知过程可能处于认知阶段或者疲劳阶段. 由于认知过程3个阶段的时间顺序性, 只有当系统已经确定处于中间阶段以后, 若 $\forall k \in \{1, 2, \dots, K\}$, 均满足 $\delta(x_m(t), x_k) \geq \delta_e$, $N_f = (t-1)N + m$, 否则系统处于认知阶段. 在评价了一代个体后, 若 $\forall k \in \{1, 2, \dots, K\}$, 均不满足 $\delta(x_m(t), x_k) \geq \delta_e$, 为了种群继续进化, 则近似取 $N_c = (t+1)N$.

3.4 交互式遗传算法的降噪策略(Strategies of reducing noises in interactive genetic algorithms)

在刻画认知评价度和疲劳评价度的基础上, 可以调整人赋予的进化个体适应值, 即以适应值可信度的折扣接受人赋予的进化个体适应值. 因此, 可得如下进化个体适应值调整公式:

$$F'(x_i(t)) = R((t-1)N + i - 1) \cdot F(x_i(t)). \quad (8)$$

其中 $F'(x_i(t))$ 为进化个体 $x_i(t)$ 修正后的适应值. 进化个体适应值的修正在人评价完每代种群后统一进行.

3.5 算法步骤(Steps of algorithms)

算法步骤如下:

步骤1 确定 k_1 , k_2 和 K 以及遗传操作参数, 生成初始种群;

步骤2 解码, 用户评价进化个体;

步骤3 依式(6)计算评价偏差;

步骤4 依第3.3节方法确定 N_c 和 N_f ;

步骤5 依式(8)调整进化个体适应值;

步骤6 判断算法是否满足停机准则, 若是, 输出最优个体, 算法结束; 否则, 转步骤7;

步骤7 遗传操作, 生成下一代种群, 转步骤2.

4 实例(Application)

本文以设计只有上衣和裙子的二维女装进化设计系统作为应用实例. 上衣和裙子的款式各32套, 分别可以被染成16种颜色, 因此, 搜索空间大小

为 $2^5 \times 2^5 \times 2^4 \times 2^4$. 个体采用二进制编码, 共有18位, 如图2所示, 其中前5位表示上衣的代码, 第6位到第10位表示裙子的代码, 第11位到第14位表示上衣颜色的代码, 第15位到第18位表示裙子颜色的代码. 颜色编码与相应的颜色取值如表1所示. 上衣和裙子的款式以bmp图片格式存放在统一的文件夹中并按照一定的顺序将图片排列、命名, 名称分别是从0到31的整数, 对应于相应部位款式二进制代码的十进制值. 例如图2中上衣代码10110对应于编号为22的上衣款式, 裙子的代码00100对应于编号为4的裙子图片; 上衣颜色代码0110对应于编号为6的黄色, 裙子颜色的代码1001对应于编号为9的亮蓝色.

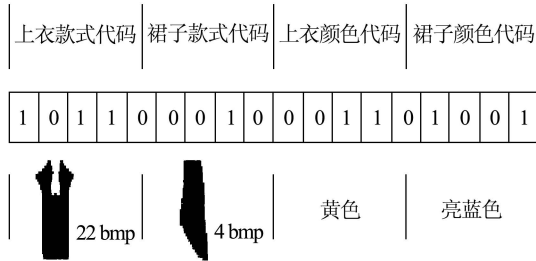


图2 个体编码
Fig. 2 Individual's coding

表1 颜色编码与取值
Table 1 Color's coding and its value

编码	值	颜色	编码	值	颜色
0000	0	黑色	1000	8	灰色
0001	1	蓝色	1001	9	亮蓝色
0010	2	绿色	1010	10	亮绿色
0011	3	青色	1011	11	亮青色
0100	4	红色	1100	12	亮红色
0101	5	洋红色	1101	13	亮洋红色
0110	6	黄色	1110	14	亮黄色
0111	7	白色	1111	15	亮白色

本文算例的种群规模为8, 个体适应值采用百分制打分, 采用适应值比例选择、多点交叉和单点变异, 交叉概率为1, 变异概率为0.01, 进化截止代数20, 采取精英保留策略. 取认知评价阈值 $\delta_e = 0.05$, 此时能够较恰当地反映人的认知过程, 可以达到跟踪人的认知过程的目的. 由于人对服装的评价过程较简单, 所以认知偏差的跟踪深度取 $K = 1, k_1 = k_2 = 0.5$.

为了保证实验结果的客观性, 随机选择10名男生和10名女生作为评价者, 进行设计, 10名男生分别为 $M_1, M_2, \dots, M_{10}$, 10名女生分别为 $O_1, O_2, \dots, O_{10}$, 其中编号为 M_1 到 M_5 的5名男生和编号为 O_1 到 O_5 的5名女生的服装设计采用含降噪策略的交互式遗传算法, 另外5名男生和5名女生的

服装设计采用一般的交互式遗传算法. 不含降噪策略的交互式服装设计的进化代数和算法的收敛性列于表2. 含降噪策略的交互式服装设计的认知评价度、疲劳评价度、进化代数和算法的收敛性列于表3. 表4列出了含降噪策略和不含降噪策略的男生和女生的平均进化代数.

表2 不含降噪策略的交互式服装设计的运行结果
Table 2 The results of interactive fashion design without strategies of reducing noises

评价者	进化代数	算法的收敛性
M_6	20	不收敛
M_7	19	收敛
M_8	20	不收敛
M_9	20	不收敛
M_{10}	20	不收敛
O_6	15	收敛
O_7	12	收敛
O_8	10	收敛
O_9	11	收敛
O_{10}	12	收敛

表3 含降噪策略的交互式服装设计的运行结果
Table 3 The results of interactive fashion design with strategies of reducing noises

评价者	认知评价度 n_1	疲劳评价度 n_1	进化代数	算法的收敛性
M_1	12	78	17	收敛
M_2	10	97	16	收敛
M_3	8	80	19	收敛
M_4	16	79	18	收敛
M_5	14	68	16	收敛
O_1	1	无	10	收敛
O_2	2	96	13	收敛
O_3	1	无	11	收敛
O_4	1	无	10	收敛
O_5	1	无	11	收敛

表4 含降噪策略和不含降噪策略的男生和女生的平均进化代数

Table 4 The average evolution generations of boys and girls with and without strategies of reducing noises

含降噪策略的服装设计	男生平均进化代数	17.2
	女生平均进化代数	11.0
不含降噪策略的服装设计	男生平均进化代数	不收敛
	女生平均进化代数	12.0

由表2可知, 在不含降噪策略的实验组中, 女生的服装设计全部收敛, 男生的服装设计, 在截止代数内只有1名男生都找到了自己喜欢的服装. 而在含降噪策略的对照组中, 所有的男生和女生都可以设计出

自己喜欢的服装. 由此可知, 含降噪策略的交互式服装设计可以改进算法的收敛性, 修正用户在设计过程中的偏差, 有利于用户在搜索空间中找到自己喜欢的服装样本. 通过表4可以进一步看出含降噪策略的交互式遗传算法中男生服装设计的平均进化代数为17.2, 并且算法全部收敛, 可见含降噪策略的交互式遗传算法可以有效地提高进化效率, 加快搜索速度.

但是, 由表2~表4可知是否含有降噪策略对女生的服装设计过程影响不是很明显, 女生含降噪策略的服装设计与不含降噪策略相比进化代数只提高了1代. 这说明女生对服装设计过程本身就比较感兴趣, 且设计的又是女裙装, 所以女生可以很快进入设计状态, 设计出自己喜欢的服装. 由表3可以看出, 5名女生中有4名跨过了设计的认知阶段直接进入了设计的中间阶段, 只有1名女生进入了疲劳阶段, 且其疲劳评价度明显较男生高. 相比较而言, 男生进入设计状态的时间稍微长一些, 一般都是评价了十几个个体才能熟悉设计对象, 进入设计的中间阶段, 且5名男生都经过了设计的疲劳阶段, 其平均疲劳评价度为80.4.

由上可知, 交互式遗传算法中的噪声大小不仅与设计对象的复杂程度有关, 而且设计者对设计对象的熟悉程度也直接影响设计的进程. 因此, 本文中的噪声模型引入认知评价度和疲劳评价度刻画设计过程是比较客观的, 能够恰当反映交互式遗传算法中设计者对设计进化过程的影响, 服装设计的实例验证了本文算法的有效性.

5 结论(Conclusions)

交互式遗传算法的适应值噪声不利于种群进化, 且用于传统遗传算法的降噪策略在这里不再适用. 本文根据种群的不同进化阶段, 对人给出的进化个体适应值赋予不同可信度, 并用其调整人给出的进化个体适应值, 调整后的适应值参与后续遗传操作, 从而达到降低适应值噪声的目的. 在服装进化设计系统中的应用及对比结果验证了方法的可行性和有效性.

为了简化讨论, 本文没有考虑疲劳阶段和认知阶段的交叠问题, 即在评价对象较复杂时, 有可能还没有进入中间阶段, 用户已经感觉疲劳了, 这种情况较复杂, 是需要进一步研究的问题. 另外, 交互式遗传算法中的噪声对进化模式影响的机理, 及对进化效

率影响的定量分析也是进一步研究的内容.

参考文献(References):

- [1] TAKAGI H. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation[C]//*Proceedings of the IEEE*. 2001, 89(9): 1275 – 1296.
- [2] CHO S B. Towards creative evolutionary systems with interactive genetic algorithm [J]. *Applied Intelligence*, 2002, 16(2): 129 – 138.
- [3] TOKUMARU M, MURANAKA N, IMANISHI S. Virtual stylist project: examination of adapting clothing search system to user's subjectivity with interactive genetic algorithms[C]//*Proceedings of the IEEE Congress on Evolutionary Computation*. Chicago: IEEE Service Center, 2003: 1036 – 1043.
- [4] BRANKE J, SCHMIDT C. Sequential sampling in noisy environments[M]//*Lecture Notes in Computer Science*. Berlin: Springer, 2004: 202 – 211.
- [5] SANO Y, KITA H. Optimization of noisy fitness functions by means of genetic algorithms using history of search with test of estimation[C]//*Proceedings of IEEE Congress on Evolutionary Computation (CEC2002)*. New York: Morgan Kaufmann, 2002: 360 – 365.
- [6] BEYER H G. Actuator noise in recombinant evolution strategies on general quadratic fitness models[C]//*Proceedings of Genetic and Evolutionary Computation Conf(GECCO2004)*. Seattle, WA, USA: Springer, 2004: 654 – 665.
- [7] JIN Yaochu, BRANKE Jürgen. Evolutionary optimization in uncertain environments-A survey[J]. *IEEE Transactions on Evolutionary Computation*, 2005, 9(3): 303 – 317.
- [8] DI PIETRO A, WHILE L. Applying evolutionary algorithms to problems with noisy, time-consuming fitness functions[C]//*Proceedings of the 2004 Congress on Evolutionary Computation*. Portland, OR, USA: [s.n.], 2004: 1254 – 1261.
- [9] CANTU-PAZ E. Adaptive sampling for noisy problems[C]//*Proceedings of Genetic and Evolutionary Computation Conf(GECCO'2004)*. Seattle, WA, USA: Springer, 2004: 947 – 958.
- [10] 齐岩, 卢德唐. 交互式遗传算法在基于内容的图像检索中的应用[J]. *中国图象图形学报*, 2004, 9(1): 46 – 55.
(QI Yan, LU Detang. The Application of IGA in the CBIR[J]. *Journal of Image and Graphics*, 2004, 9(1): 46 – 55.)
- [11] 蒋姗姗, 曹先彬, 王煦法. 基于IGA的用户Agent模型与设计[J]. *模式识别与人工智能*, 2004, 17(2): 244 – 249.
(JIANG Shanshan, CAO Xianbin, WANG Xufa. Users' agent model and design using IGA[J]. *Pattern Recognition and Artificial Intelligence*, 2004, 17(2): 244 – 249.)
- [12] GONG D, ZHOU Y, Li T. Cooperative interactive genetic algorithm based on user's preference[J]. *International Journal of Information Technology*, 2005, 11(10): 1 – 10.

作者简介:

周 勇 (1974—), 男, 江苏省徐州人, 讲师, 博士生, 从事进化计算的研究, E-mail: zhoyongchina@126.com;

巩敦卫 (1970—), 男, 博士, 教授, 博士生导师, 从事进化计算与智能控制的研究.