

文章编号: 1005—0523(2007)01—0123—06

一种免疫算法的改进

曾 毅

(华东交通大学 基础科学学院, 江西 南昌 330013)

摘要:针对一种免疫算法在求解多模态函数优化问题的不足,利用基于相似结构的小生境技术和近优淘汰算子对免疫算法加以改进.通过仿真实验,本文分析了交叉算子、变异算子及近优淘汰算子对算法的影响.仿真结果表明改进后的算法的性能有明显的提高.

关 键 词:免疫算法;近优淘汰算子;小生境技术;多模态最优化

中图分类号:TP18

文献标识码:A

0 引言

随着人类基因工程技术的发展,人们越来越注重生物系统诸多特性在工程中的应用.受生物免疫系统的启发,而发展起来的人工免疫系统及应用已成为崭新的研究领域正受到计算智能学者的广泛关注^[1],它是继神经网络及进化计算之后的又一研究热点.基于免疫机理的算法主要有两种途径:一种是基于免疫应答的免疫算法(IAs: Immune Algorithms);另一种是借鉴免疫系统中的接种育苗或抗体的促进和抑制机制构建增强群体多样性的算子,并结合遗传算法提出的免疫遗传算法(IGAs: Immune Genetic Algorithms).文章通过对一种免疫遗传算法(IGAs)分析,指出该算法的不足,利用基于相似结构的小生境技术及近优淘汰算子,对算法加以改进,仿真表明改进后的算法改善了群体的多样性,算法的性能明显优于原算法.

1 免疫算法的描述及分析

免疫算法(IAs)是由 T. Fukuda 等人(1998)所提出并发展起来的一类求解多模态优化问题的算法,

它模拟生物免疫系统的识别多样性与多样性免疫细胞的产生与维持机制.在免疫算法中,生物个体对应着抗体,以定长二进制串表示;个体适应性度量对应着抗体与抗原(即问题与约束)之间的亲和力;选择取作免疫选择;在鼓励高亲和力抗体的同时,抑制高浓度的抗体;繁殖对应作免疫细胞的分化与增值.为了便于对免疫算法的分析,我们先介绍几个相关的基本概念:

1) 信息熵:设等位基因 $\Gamma = \{k_1, k_2, \dots, k_s\}$, $X = \{X_1, X_2, \dots, X_n\} \in H_L^N$ 为一个 N 种群,每个个体 $X_i = a_{i1} a_{i2} \dots a_{il}$, $a_{ij} \in \Gamma$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, L$).

定义 $H(X) = \frac{1}{L} \sum_{j=1}^L H_j(X)$.

其中 $H_j(X) = -\sum_{i=1}^s p_{ij} \log p_{ij}$, p_{ij} 为在种群 X 中 N 个个体的第 j 位取值 k_i 的频率,称为 $H_j(X)$ 为 X 的第 j 位的信息熵(information entropy), $H(X)$ 为 X 的信息熵.信息熵 X 中诸个体多样性的一种度量.

2) 相似度:抗体 X_i 和 X_j 的相似度定义为

$$\alpha v(X_i, X_j) = \frac{1}{1 + H(X_i, X_j)},$$

其中 $H(X_i, X_j)$ 表示种群 $X = \{X_i, X_j\} \in H_L^2$ 的信息

收稿日期: 2006—12—30

基金项目: 华东交通大学校级科研课题《免疫算法参数的研究》

作者简介: 曾毅(1965—),男,江西吉安人,副教授,硕士.

熵.

3) 抗体浓度: 对种群 $X = \{X_1, X_2, \dots, X_n\} \in H_L^2$ 中的任何抗体 $X_i (i=1, 2, \dots, N)$ 的度定义为 $C_{X_i} = \frac{1}{N} \sum_{j=1}^N \zeta_{ij}$, 其 $\zeta_{ij} = \begin{cases} 1, & \text{若 } (x_i, x_j) \geq Tac^0, \\ 0, & \text{其它,} \end{cases}$ 其中 Tac^0 为取定的阈值(本文 $Tac^0=0.85$).

以求解以下多模态函数优化问题为例, 依上述基本概念可将免疫算法^[2]描述为:

问题: $\max f(b), b \in R^n$,

其中 $b = (b_1, b_2, \dots, b_n)$, $v_i \leq b_i \leq u_i, 0 \leq f(b) < +\infty$, 且 $f(b) \neq \text{常数}$.

step 1(初始化) 随机生产 N_1 体抗体 $X_1(0), X_2(0), \dots, X_{N_1}(0)$, 并从免疫记忆库中提取 N_2 个抗全 $X_{N_1+1}(0), X_{N_1+2}(0), \dots, X_N(0)$ (其中 $N = N_1 + N_2$). 如果记忆库中无抗体, 可随机选取.

令初始种群为: $X(0) = (X_1(0), X_2(0), \dots, X_N(0))$; 设定停机准则, 并置 $t = 0$.

step 2(种群进化)

1) 评价 对 $X(t)$ 中的个抗体 $X_i(t) (i=1, 2, \dots, N)$ 进行评价, 即计算:

(1) $X_i(t)$ 的浓度 $C_{X_i(t)}$;

(2) $X_i(t)$ 与抗原的亲合力 $J(X_i(t))$;

(3) $X_i(t)$ 的期望繁殖概率 $P_{X_i(t)} = \frac{J(X_i(t))/C_{X_i(t)}}{\sum_{j=1}^N J(X_j(t))/C_{X_j(t)}}$

(注: 本文取目标函数值作为 $J(X_i(t))$)

2) 免疫选择 以 $\{R_{X_i(t)}\}$ 为概率分布在 $X(t)$ 中选择 N_1 对抗体作母本; 同时选择 N_2 个抗体存入记忆库.

3) 细胞分化与增殖 对 $X(t)$ 中选择的 N_1 对母本执行杂交与变异生成 N_1 个新的抗体 $X_1(t+1), X_2(t+1), \dots, X_{N_1}(t+1)$ 在记忆库中取出 N_2 个抗体;

$X_{N_1+1}(t+1), X_{N_1+2}(t+1), \dots, X_N(t+1)$ 一并生成新一代抗体种群;

$x(t+1) = (X_1(t+1), X_2(t+1), \dots, X_N(t+1))$.

step 3 终止检验 如果所设终止条件满足, 则输出 $X(t+1)$ 中最高亲和力的本作为最优解, 终止计算, 否则置 $t = t+1$ 并转 step 2.

step 3 终止检验 如果所设终止条件满足, 则输出中最高亲和力的抗体作为最优解, 终止计算, 否则置并转 step 2.

通过对多模态函数优化问题进行模拟, 表明免疫算法存在两个方面的缺陷, 其一, 对多模态函数求解时, 不能很好地找出全部极值点; 其二, 当搜索到极值点附近时, 搜索过程停滞不前或进展缓慢. 这是由于:

(1) 采用免疫选择策略时, 虽然群体多样性有较好地改善, 但对于期望繁殖概率较大地个体而言, 当其期望繁殖概率高于种群的平均期望繁殖概率时, 则其在种群中的数量会急剧增加, 以至于支配整个群体, 此现象一旦出现, 便会造成冗余现象, 交叉操作难于产生新的个体. 仅靠偶尔的随机变异, 不可能有效快速地跳出此状态, 从而使免疫算法难于收敛到全部极值点.

(2) 父代个体进行杂交和变异操作以产生新的个体, 增大搜索能力, 由于这种进化缺乏有效的控制机制, 交叉和变异操作产生的子代个体是随机的, 即可能优于父代个体, 也可能比父代个体更差, 甚至是淘汰的个体, 从而降低了在优良解附近进行有效搜索局部收敛速度.

(3) 对于多模态函数优化问题而言, 由于要搜索到给定自变量空间的多个极值点, 为了不降低任何一个极值点的搜索概率, 所以相对于单峰函数, 则必须按极值点的个数成倍增加群体规模, 这势必造成运算量过大. 加之, 大部分问题在优化之前极值点的个数未知, 这也给选定初始群体规模造成了困难.

(4) 以亲和力高或期望繁殖概率作为选取记忆库个体的条件, 这有可能将有的极值点对应的个体排除在记忆库之外, 从而也造成免疫算法难于收敛到全部极值点.

2 算法的改进

2.1 基于相似结构的小生境技术

Goldberg^[3]指出, 在进化算法几种流行的选择机制中, $(u+\lambda)$ 选择机制具有最强的选择压, $(u+\lambda)$ 选择允许 u 个父代个体和 λ 个子代个体共同竞争, 确定性地选择 u 个高适应值的个体进入新的种群. 仿真实验表明, 当在整个种群进行随机配对的交叉操作时, 用 $(u+\lambda)$ 选择能产生最快的局部收敛速度. 因此, 在应用 $(u+\lambda)$ 选择提供强的选择压的同时, 追求个体的多样性并改善交叉、变异操作的进化效果成为改善算法的主要目标. 具体的做法为: 父代个体交叉操作后, 立即应用选择机制, 确定性地选择 2 个优良个体进入下一代, 变异操作按概率进行, 仅

对种群中的最佳个体变异时,应用选择机制,以保证全局的收敛性,而对其它个体仅做随机变异,不作选择.

2.2 近优淘汰算子

为了提高搜索出所有极值点的概率,避免出现抗体的冗余现象.引入近优淘汰算子,并在此基础上产生记忆库中的个体.

近优淘汰算子的实现:

step 1 设抗体群为 $X(t)=(X_1(t), X_2(t), \dots, X_N(t))$ (1)

计算亲和力 $J(X_i(t))$ ($i=1, 2, \dots, N$)并按亲和力从大到小对抗体群中的个体进行排序,不妨仍记为 (1)

step 2 计算抗体 $X_1(t)$ 与 $X_j(t)$ ($j=1, 2, \dots, N$)之间的距离

$d(X_1(t), X_j(t)) = \|X_1(t) - X_j(t)\|$, 并取阈值 $Tacl, Tac1'$.

记小生境子群 $M_{p1} = \{X_j | \|X_1(t) - X_j(t)\| < Tac1, J(X_1(t)) - J(X_j(t)) < Tac1'\}$, 称 $X'(t) = X(t) - M_{p1}$ 为剩余子群(注:保持个体的相对排序不变). (注本文 $Tacl=0.15$)对小生境子群 M_{p1} 实施以下操作:

首先,保留个体 $X_1(t)$,

取 $M'_{p1} = \{X_j | \|X_1(t) - X_j^*(t)\| < Tac2; J(X_1(t)) - J(X_j^*(t)) < Tac2', X_j^* \in M_{p1}\}$.

其中 $Tac2 \ll Tac1, Tac2' \ll Tac1'$.

其次,将 M'_{p1} 中的个体 $X_j^*(t)$ 删除,取 $M_{p1} = M_{p1} - M'_{p1}$, 对 M_{p1} 实施以上两步,直至每个个体都实施了以上操作对每次保留的个体,依 $|M_{p1}|(|M_{p1}|$ 指最初的小生境子群远见模)的值的大小按一定的比例从大到小选取到记忆库,同时随机生成新的个体补充删除的个体.

step 3 令 $X(t) = X'(t)$, 重复 step 2, 直至剩余子群为 $\emptyset$.

显然,淘汰处芭保持群体的多样性,同时又避免出现个体的冗余现象,提高搜索出全部极值点的概率.

3 新的免疫算法及算例分析

3.1 新的免疫算法的描述

step 1 初始化(同原免疫算法)

step 2 种如进化

1) 评价 (同原免疫算法)

2) 免疫选择 以 $\{R_{X_i(t)}\}$ 为概率分布在 $X(t)$ 中选择 N_1 对抗体作母本.

3) 细胞分化与增殖 对 $X(t)$ 中选择的 N_1 对母本执行 $(2+2)$ 杂交与变异操作,生成 N_1 上新的抗体 $X_1(t+1), X_2(t+1), \dots, X_{N_1}(t+1)$; 在从记忆库中取出 N_2 个抗体 $X_{N_1+1}(t+1), X_{N_1+2}(t+1), \dots, X_N(t+1)$ 一并生成新一代抗体种群:

$X(t+1) = (X_1(t+1), X_2(t+1), \dots, X_N(t+1))$.

4) 对抗体种群 $X(t+1)$ 进行近优淘汰操作,并生成新记忆库.

step 3 终止检验 如果所设终止条件满足,则输出记忆库中的抗体作为最优解,终止计算,否则,置 $t:=t+1$ 并转 step 2.

3.2 算例分析

为分析对此改进前后免疫算法的性能,对下面典型的多峰函数进行测试.

$$f(x) = \exp[-(x-0.1)^2] \sin^6(5\pi x^{\frac{3}{4}}), x \in [0, 1]$$

为分析对比改进前后免疫算法的性能,对下面典型的多峰函数进行测试.

$$f(x) = \exp[-(x-0.1)^2] \sin^6(5\pi x^{\frac{3}{4}}), x \in [0, 1]$$

3.2.1 杂交算子对算法的影响

取参量: 群体规模 $N=100$, 染色体的串长 $L=15$, 迭代次数 $T=40$, 变异概率 $P_m=0.2$, 而杂交概率 $P_c=0.3, P_c=0.6, P_c=0.9$ 进行比较. 图 1 中的实线是多峰值函数的曲线, “o” 是保留于记忆库中的个体. 图 1 左边的三个图是改进前的实验结果, 右边的三个图是该进后的实验结果. 从左图的三个结果来看, 较大 P_c 的容易破坏群体的中已形成的优良模式, 使搜索具有太大的随机性, 而较小 P_c 的使发现新个体(特别是优良新个体)的速度太慢. 比较理想的方式是非一致地使用交叉概率: 在免疫算法的前期使用交大的 P_c , 在免疫算法的后期使用较小的 P_c , 甚至不使用交叉概率. 从右图三个结果来看, 近优淘汰算子使用使算法具有较好的稳定性, 在一定程度上可以弥补交叉概率对算法的影响.

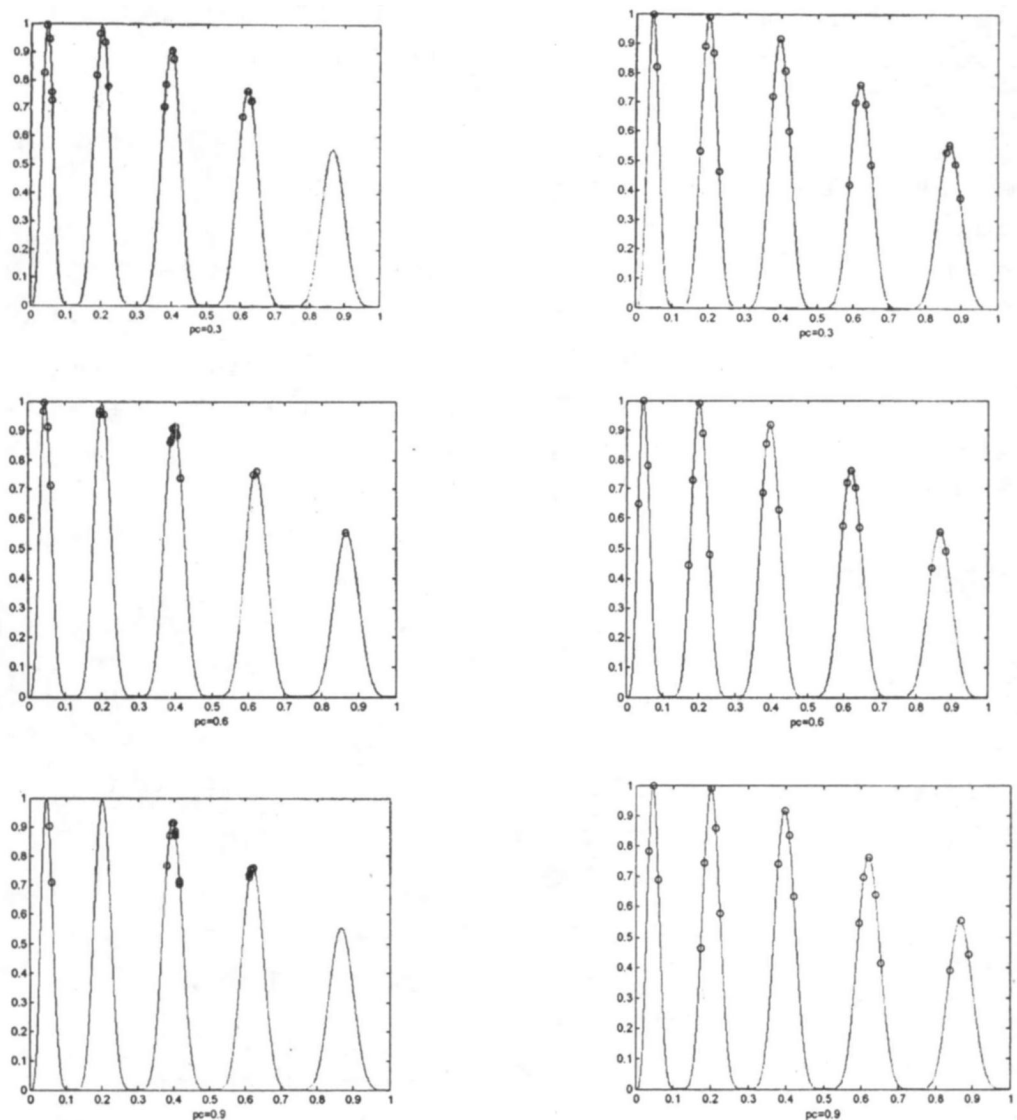
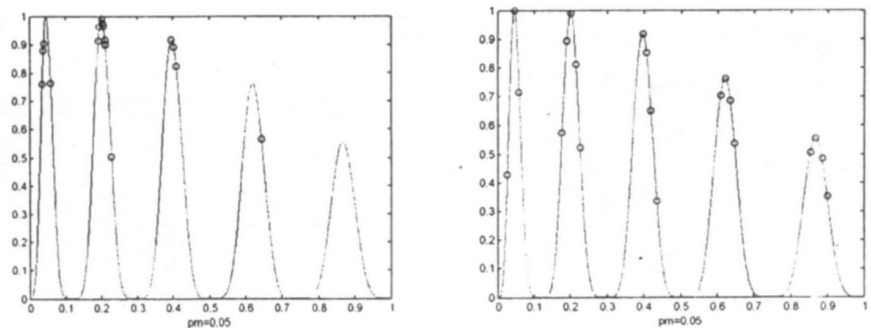


图 1

3.2.2 变异算子对算法的影响

取参量：群体规模，染色体的串长，迭代次数交叉概率，而变异概率分别取进行比较. 图 2 中的实线是多峰值函数的曲线，“o”是保留于记忆库中的个体. 图 2 的左边的三个图是改进前的实验结果，右边的三个图是该进后的实验结果. 对个体变异操作，实

际上是子代基因按小概率扰动产生的变化，而从左边的三个图来看，变异概率的变化，不足以改善“近亲繁殖”的现象，以至于无法寻找找到较多的极值点. 而从右边的三个图来看，合理的使用近优淘汰操作对于改善群体的多样性有明显的效果. 改善了算法的性能，100%地寻找找到多峰值函数的各个极值点.



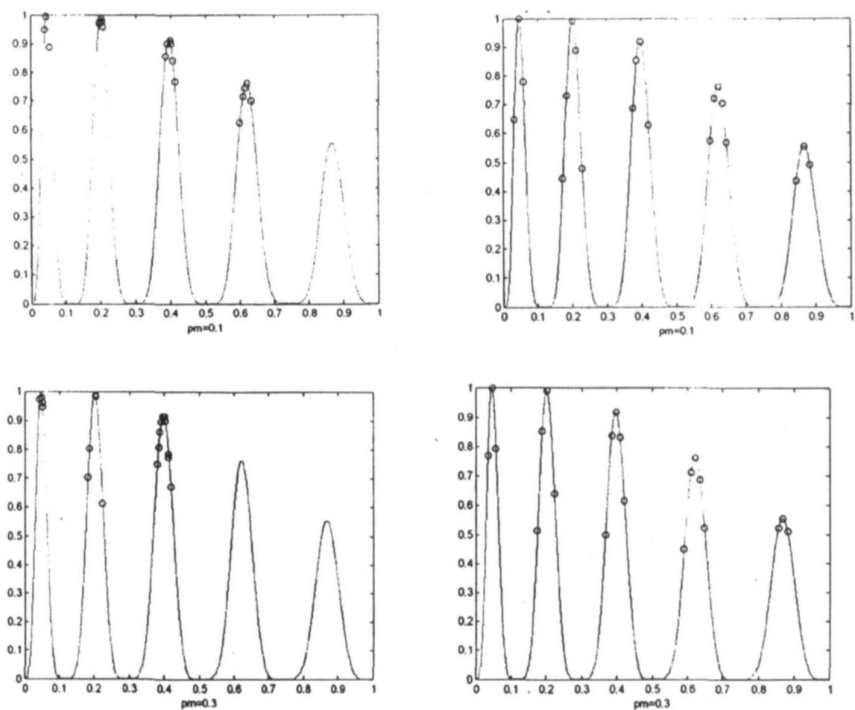


图 2

3.2.3 淘汰算子对算法的影响

取参量：群体规模 $N=100$ ，染色体的串长 $L=15$ ，迭代次数 $T=40$ ，杂交概率 $P_c=0.6$ ，变异概率 $P_v=0.1$ ，由于目标函数为连续函数，在实施近优淘汰操作时没必要考虑条件 $J(X_1(t)) - J(X_j(t)) < Tac1'$ 及 $J(X_1(t)) - J(X_j(t)) < Tac2'$ ，淘汰操作的

阈值 $tac2$ 分别取 $Tac2=0.005$ ， $Tac2=0.01$ ， $Tac2=0.08$ 进行比较。图 3 中的实线是多峰值函数的曲线，“o”是保留于记忆库中的个体。从实验结果来看，根据问题选取合适的阈值 $tac2$ ，可以明显地改善群体的多样性，从而在求解多峰值函数优化问题时，改进后的免疫算法在性能上明显地优于免疫算法。

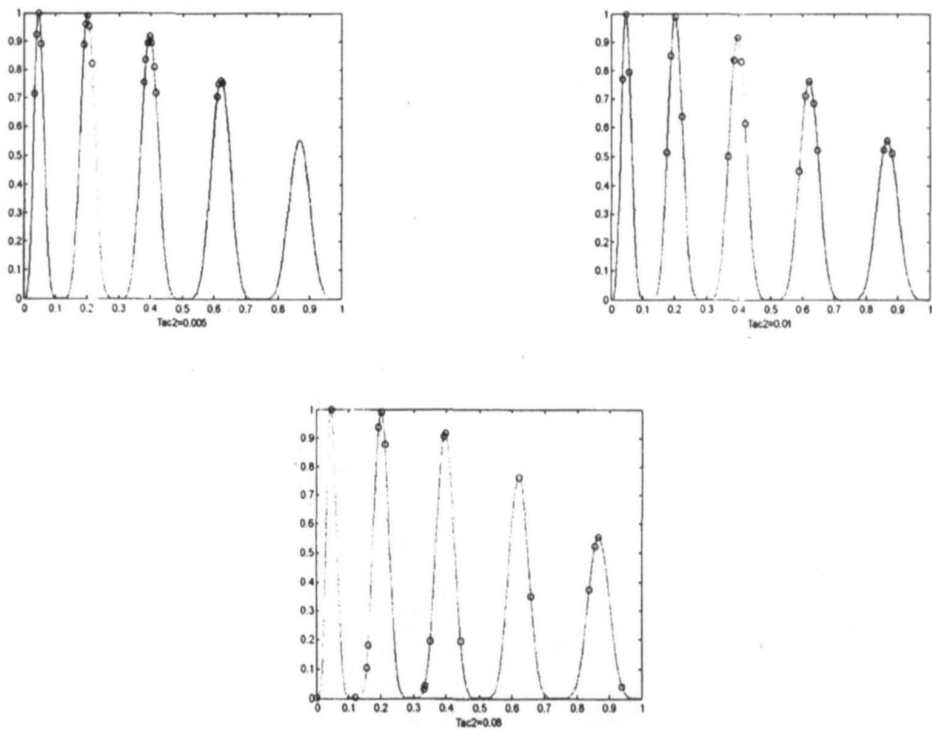


图 3

4 结束语

本文针对免疫算法不足的基础上,利用基于相似结构的小生境技术及近优淘汰算子,对原算法加以改进,明显地改善群体的多样性,使改进后的算法的性能有较为明显的改善.并提出了确定记忆库个体的方法.同时,发现在使用信息熵作为评判个体近似程度也有许多不足之处,将另文加以讨论.

参考文献:

[1]de CASTROLN,Von ZUBEN F J. learning and optimization using the clonal selection principle [J]. IEEE trans on Evolution-

ary computation ,special Issue on Artificial Immune systems, 2002. 6. (3);239—251.

[2]徐宗本. 计算智能(第一册)模拟进化计算[M]. 北京:高等教育出版社,2004.

[3]D·E·Goldberg and k·Deb, A Comparative Analysis of selection Schemes used in Genetic Algorithms. In Genetic Algorithms. In Rawins, 1991;69—91.

[4]周明,孙树栋. 遗传算法及应用[M]. 北京:国防工业出版社,2000.

[5]张著洪,黄席樾. 一种新的免疫算法及其在多模态函数优化中的应用[J]. 控制理论与应用,2004,21(2);17—21.

[6]葛红,毛宗源. 免疫算法的实现[J]. 计算机工程,2003,29(5);62—63.

An Improved Immune Algorithm

ZENG Yi

(School of Natural Science, East China Jiaotong University, Nanchang 330013, China)

Abstract:To the drawback that an immune algorithms solve the multi-model optimizing problem, the immune algorithms is improved by using a niche technology of similar structure and eliminating near best result operator. Through the simulation test, the paper analyses the impact of crossover operator, mutation operator and eliminating near best result operator on the algorithms. Simulation results show that the property of the improved algorithms is improved greatly.

Key words:immune algorithms; eliminating near best result operator; niche technology; multi-model optimization