

# 一种基于生物免疫遗传学的新优化方法

罗小平, 韦 巍

(浙江大学电气工程学院, 浙江杭州 310027)

**摘 要:** 本文将免疫遗传学的基本思想引入到优化设计中. 模拟生物体的实际免疫行为, 设计出了融合应答、免疫记忆、基因重组、新陈代谢、浓度控制、隔离小生境技术和混沌思想的实用化的免疫遗传算法. 然后利用此算法对多峰值函数的寻优过程进行实验研究, 证明了本文的方法能够改善简单遗传算法的优化能力, 所得结果令人满意.

**关键词:** 免疫遗传算法; 优化

**中图分类号:** TN911 **文献标识码:** A **文章编号:** 0372-2112 (2003) 01-0059-04

## A New Optimization Method on Immunogenetics

LUO Xiao-ping, WEI Wei

(College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China)

**Abstract:** In this paper, idea on basic immunogenetics is introduced to optimization. By simulating the behavior of biological immune system, we design the immunogenetic algorithm. Experiments have been carried out on the function optimization, the results show that the method can improve the optimization capability of Simple Genetic Algorithm (SGA) successfully and so the satisfaction is proved.

**Key words:** immunogenetic algorithm; optimization

### 1 引言

简单遗传算法 (Simple Genetic Algorithm, 简记为 SGA) 是由 Michigan 大学的 Holland 教授等创立的, 主要理论基础是生物进化论和群体遗传学. SGA 的基本特征是: 利用群体进化——即在求解过程中, 通过使种群不断优化, 从而找到满意解或最优解. 对于许多常规方法难以有效解决的非线性优化问题, SGA 往往能够奏效, 因此在许多工程问题的解决中, SGA 获得了广泛的应用.

SGA 在理论上借鉴生物进化理论以及遗传学机理, 已经形成了一套较为完善的算法体系, 然而在实际使用中, 还有许多问题有待进一步研究探讨. 例如, 对于单调函数或单峰值函数, 在初始时很快向最优值逼近, 但是在最优值附近收敛较慢; 而对于多峰值函数的优化问题, 它往往出现“早熟”, 即收敛于局部极值. 究其原因, 主要是通常使用的 SGA 的选择策略多采用个体繁殖机会同其适应值成正比例的方法, 这样就很容易导致超级个体问题和多个相似数字串问题<sup>[1]</sup>. 交叉算子的设计一般都采用随机交叉的方式, 由两个个体的交叉产生两个新个体, 其结果是父代与子代间很相似, 这也会导致如上的问题. 因此, 有必要研究如何改进 SGA, 采用合适的算法加快多峰值函数的寻优速度和质量.

另一方面, 目前许多有关智能系统的研究都是围绕人脑智能及其学习机制进行的. 这些拟人化方法都忽略了与人脑

行为方式并不明显相关的另一类智能系统——免疫系统. 生物体的免疫系统具备很高的智能级 (high level of intelligence), 但它与人脑的行为方式的确没有明显联系. 由于实际的生物体免疫系统具备免疫记忆、记忆开发 (长期和短期)、混沌识别、自适应网络调节等许多优良功能, 因此, 我们可以考虑模拟其实际行为规律, 设计出相应的数学算法来解决实际问题.

在免疫系统中产生的抗体是用于对付和消除外来抗原的. 如果我们进行类比——对于一个优化问题而言, 抗原对应问题的目标函数而抗体对应问题的最优解, 那么免疫系统的特点对于改进和提高遗传算法的能力就具有重要的启迪作用. 文[7]中模拟部分免疫特点 (如自我免疫等) 对 SGA 进行改进而设计出一种免疫遗传算法, 该算法的优化能力相对 SGA 而言有一定的改进. 本文对如何有效模拟免疫重组、免疫记忆、混沌增殖等免疫行为而设计出免疫遗传算法以更好地进行多峰值函数的寻优进行探讨.

### 2 免疫遗传学思想及免疫遗传算法设计

#### 2.1 免疫遗传学基本思想

在生物医学科学领域, 免疫遗传学 (Immunogenetics) ——作为免疫学 (Immunology) 和遗传学 (Genetics) 这两门学科的边缘学科, 丰富和发展了现代遗传学理论. 免疫遗传的研究说明了免疫物质不仅受遗传基因的控制, 而且由此发生的免疫

功能(免疫应答)也同样受到遗传调控。

当传染媒介侵入生物系统后,免疫系统的工作就是中和或者清除异物.免疫系统模型如图 1 所示.免疫系统的特点详述如下:

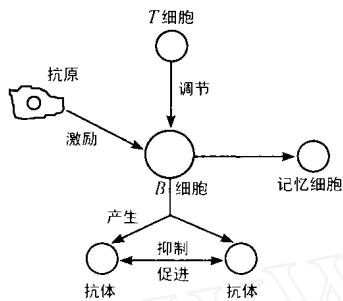


图 1 免疫系统模型

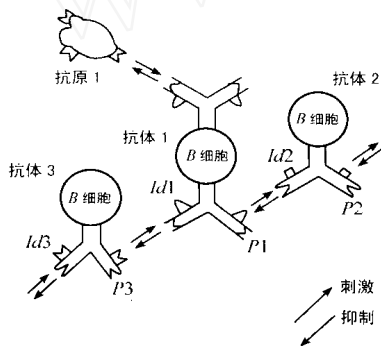


图 2 免疫网络示意图

(1) 基因重组:当外来抗原 (antigen) 侵入生物体时,免疫系统先对不同的抗原进行识别,针对不同的抗原生成对应的抗体 (antibody) 进行中和或破坏.抗体的产生是通过 DNA 分子上特定段的任意重组进行的<sup>[2]</sup>.

(2) 网络作用:当一个抗体传递给生物体中的 B 细胞时,一次免疫响应就启动了.如果与抗原匹配,B 细胞将被激活. B 细胞的激励水平不仅依赖于其与抗原匹配的程度,而且依赖于它同机体免疫系统中其它 B 细胞的亲和力.根据 Jerne 提出的免疫网络理论(见图 2),计算抗体激励水平及浓度的动态方程如下<sup>[3]</sup>:

$$\frac{dA_i(t)}{dt} = \left( \frac{\sum_{j=1}^N m_{ij} a_j(t)}{N} - \frac{\sum_{k=1}^N m_{-i-k} a_k(t)}{N} + g_i - k_i \right) a_i(t) \quad (1)$$

$$a_i(t+1) = \frac{1}{1 + \exp(0.5 - A_i(t+1))} \quad (2)$$

其中,  $A_i(t)$  表示  $t$  时刻第  $i$  个个体的激励水平,  $a_i(t)$  表示  $t$  时刻第  $i$  个个体的浓度,  $m_{ij}$  表示抗体  $i, j$  间的亲和系数,  $m_{-i-k}$  表示抗体  $i, k$  间的排斥系数,  $a_i$  分别表示抗体  $i$  对于其它抗体和抗原的交互作用率,  $k_i$  表示抗体  $i$  的自然死亡率,  $g_i$  表示抗体  $i$  与抗原间的匹配率,  $N$  为抗体数目.

按照式(1)、(2)的计算结果,如果激励水平超过阈值, B

细胞将生成自身的许多克隆 (clones),而这个行为又将启动一个变异机制——在抗体分子的编码基因 (genes) 中产生变异.反之,如果激励水平低于阈值, B 细胞不会自我复制,并且到时候死亡.这样,在免疫理论中很重要的一点需要引起我们注意的是:传递一个抗原给能够约束它的 B 细胞不仅引起 B 细胞对抗原的分析,而且会导致许多新的 B 细胞的生成.所有的这些 B 细胞可以依次分析抗原并进一步产生 B 细胞——这是新的 B 细胞生成方式之一.

(3) 混沌增殖:对于有关细胞生长方面的研究,文[4]中针对白细胞生成控制模型进行了讨论,表明在确定性的时间系统中存在着内在混沌的可能性;同时,第三军医大学的徐启旺对绿脓杆菌的生长作了初步的探讨,实验结果表明:细菌的生长不是同步分裂而是呈现混乱无序的混沌态增殖,因此其本质是一种波形状态的生命活动并伴随以非线性动态的混沌增殖<sup>[4]</sup>.混沌理论中,通过 logistic 方程:  $x_{n+1}(t) = f(\mu, x_n) = \mu x_n (1 - x_n)$  ( $n=0, 1, 2, \dots$ ) 生成的混沌序列是遗传学中用以描述昆虫数目世代变化规律的方程,可以用于模拟免疫细胞的增殖方式.混沌序列较通常的 RANDOM 函数相比,具有更好的随机性、快速性——这是由混沌的本身特色决定,因此可以更好保持多样性,并能够有效解决全局搜索能力和局部搜索能力的矛盾.

(4) 免疫变异:高等生物尤其是哺乳动物,在进化进程中对多种多样的病原体或原源性蛋白甚至人工合成的抗原都能够产生相对应的抗体.实际上一个淋巴细胞只产生一种抗体,而生物体内的确存在成千上万种抗体,这种适应性功能的发展程度很让人惊讶.基因是抗体的基本单体,对于淋巴系统能够产生如此多种的抗体分子,体细胞突变学说 (somatic mutation theory, Cohn, 1968) 认为数量如此巨大的基因贮存在种质 (即生殖细胞染色体) 中是不可能的,也是不经济的.该学说认为种质传下来的只是少量 V 基因 (一种基本基因),而 V 基因的多样性主要是由于个体发育期间体细胞突变造成的.这一说法也得到了 Leder 等人的支持.这些变异主要反映在高可变区 (hv 区域, 抗体分子的一部分) 上,主要由它决定了抗体的多样性.这已为许多观察结果所支持<sup>[10]</sup>.所以,我们可以看到,变异机制在抗体的生成及多样性的维持方面起到了重要的作用.

(5) 免疫选择:由于免疫过程中许多的变异行为会破坏抗体对抗原的亲合力——即使假定变异机制是在完全随机的情况下,因此免疫系统需要通过有选择地增加高亲和力的抗体数目来解决这个问题<sup>[2]</sup>.

(6) 免疫记忆:在抗原消失后数月甚至数年,抗体形成细胞仍然具有免疫记忆.因此以后抗原侵入时生物体根据自身免疫网络可以很快生成对应的抗体来中和抗原<sup>[10]</sup>.

(7) 免疫代谢:由于机体本身的新陈代谢,每天有 5% 的低激励值 B 细胞死亡,而代之以从骨髓中生成全新 B 细胞.只有那些与网络中已存在 B 细胞具有较强亲和力的新 B 细胞才会加入到免疫系统中,否则将死亡<sup>[8]</sup>.

(8) 浓度控制:在免疫研究中我们发现免疫过程完成后机体内抗体浓度不会过高,否则,正常体细胞会受到损害.这是

免疫行为的另一个重要特点<sup>[10]</sup>。

(9) 隔离小生境:研究发现,生物在进化过程中,之所以具有形成许多物种的能力,主要是因为生物群具有分化小种群的能力,小种群沿着不同的方向进化,彼此间差距越来越大,渐渐由同一物种分化为许多不同的种。其中隔离起着非常重要的作用——它防止了被隔离种群间的杂交,因而防止了其间的基因交流。隔离对于种群的分化是很有帮助的。因此,为了进一步保证群体中个体的多样性,我们在算法的设计中引入小生境(niche)技术<sup>[11]</sup>。

如上所述,我们可以看到:免疫行为的许多特点能够使免疫细胞的多样性较好地得到维持,因而我们可以考虑模拟生物体的实际免疫行为设计出一种新的优化算法——免疫遗传算法(Immunogenetic Algorithm),简记为 IGA,它较 SGA 而言将更加有利于寻优。

## 2.2 免疫遗传算法的设计

根据 2.1 节中讨论的内容,我们可以进行免疫遗传算法的设计。根据免疫行为的特点,我们设计了如下几个核心模拟算法:

(1) 小生境:模拟隔离小生境特点,引入小生境技术<sup>[1]</sup>,将演化种群分为若干小子群以实现隔离。这包含在初始抗体生成中。

(2) 重组:模拟基因重组,在子群内将各基因位统一编码,然后进行基因的随机重组。这里不使用 SGA 的交叉算子,因而可以有效地避免父代与子代间的相似问题。

(3) 免疫变异:以类似 SGA 的方式和较大的概率进行变异操作。

(4) 抗体产生:模拟免疫网络作用、免疫选择和浓度控制,我们进行以下工作。差分计算式(1):令

$$\frac{dA_i(t)}{dt} = \frac{A_i I(n+1) T - A_i(n) T}{T} = A_i(n+1) - A_i(n)$$

计算整理得

$$A_i(n+1) = A_i(n) + \left( \frac{\sum_{j=1}^N a_{ij}(n)}{N} + g_i - k_i \right) a_i(n) \quad (3)$$

$$a_i(n) = \frac{1}{1 + \exp(0.5 - A_i(n))} \quad (4)$$

其中  $a_{ij}$  表示抗体  $i, j$  间的综合激励系数。

根据式(3)、(4)的迭代,计算出相应抗体的浓度  $a_i(n)$ ,再按照 SGA 的方法计算出第  $i$  个个体的适应值与平均适应值之比  $SGAslct\_prop(i)$ ,计算  $\min(a_i(n), SGAslct\_prop_i(n))$ ;再考虑机体中抗体浓度不能够过大的原则,进行浓度控制。算法如下:

if  $\min(a_i(n), SGAslct\_prop_i(n)) >$  规定浓度 then

复制概率  $copy\_prob_i(n) =$  规定浓度

else  $copy\_prob_i(n) = \min(a_i(n), SGAslct\_prop_i(n))$

以此实现模拟免疫系统选择性地增加高亲和力的抗体数目的行为以及免疫过程完成后机体内抗体浓度不会过高的现象。

(5) 免疫记忆:各子群依次计算完毕后,计算总群体的最

大适应值。如果本次计算得到的最大适应值大于免疫网络中已存在抗体所能提供的最大适应值,则将对应的个体作为新抗体加入抗体记忆表中,本次最大适应值加入抗体适应值表;否则,启动免疫记忆,根据抗体表中所记忆的历史抗体,利用 logistic 方程寻找出新的高适应值个体作为更好的免疫抗体,混沌初始值分为历史抗体及抗体的变异个体两种。此方法模拟如下免疫行为:当激励水平超过阈值时,机体 B 细胞将生成自身的许多克隆(clones),同时启动变异机制——在抗体分子的编码基因(genes)中产生变异。

(6) 免疫代谢:模拟免疫行为中细胞的混沌增殖现象,找出子群内 5% 的低激励水平个体并去除;利用 logistic 方程从免疫记忆表中生成新的高适应值个体加入种群中。

免疫遗传算法具体流程如下:

步骤 1 初始抗体生成( $n=1$ ) (含小生境隔离);

步骤 2 对于第  $m$  个子群进行以下的操作:

(1) 免疫重组; (2) 免疫变异; (3) 抗体生成; (4) 免疫记忆; (5) 免疫代谢; (6) 群体更新

步骤 3 各子群体依次计算完毕后,如果  $n >$  指定演化代数,则结束计算,否则  $n = n + 1$ ,转到第 2 步。

IGA 算法分析:(1) 在多峰值函数的寻优中,IGA 的局部搜索能力和全局搜索能力是一对矛盾。而我们希望遗传算法具有自动在解空间随机探索新点的能力,同时能够在某一区域内很快收敛最优解。显然,IGA 很难满足我们的要求。因为 IGA 的选择算子和交叉算子的设计方法往往会导致父代与子代间很相似,因而很难脱离局部极值。其主要赖以摆脱局部极值的变异算子的变异概率若取得太大,固体会增强全局搜索能力,但是又会导致局部搜索能力降低;为解决此问题,本文在 IGA 的设计中,引入了隔离小生境和混沌的思想。隔离小生境技术将遗传的竞争过程分为子群体间的竞争和子群体内个体间的竞争两级——前者体现为全局搜索能力,后者体现为局部搜索能力,这样通过两级竞争,就有效地解决了局部搜索能力和全局搜索能力的矛盾。(2) 隔离机制的引入使子群体的进化既同整个群体的进化密切相关,又有相对独立性,这有利于种群个体多样性的维持;同时,子群体的进化可以并行处理,这将大大提高优化的速度。(3) 为了解决由于 SGA 的选择和交叉算子设计方式而导致父子代间的相似问题,本文模拟免疫行为中抗体浓度的控制和基因的重组而设计了浓度控制策略和重组算子以进一步维持个体的多样性。(4) 本文提出了模拟免疫记忆和新陈代谢而将混沌思想引入到 SGA 中的方法。利用混沌对初值的敏感性以及随机性、规律性、遍历性,算法可以有效地跳出局部收敛点以更快地收敛于最大值,从而能够比较有效地解决通常遇到的“早熟”问题。

## 3 实验研究

本文以多峰值函数  $f(x, y) = (-1)^x (x^2 + 2y^2 - 0.3\cos(3x) - 0.4\cos(4y)) + 4$  为例,用 IGA 进行寻优比较研究。该函数有很多个局部极大值,其中最大值为 4.7,此时  $x, y$  均取值为 0。为了更好的体现本文算法的优化能力,我们还对本文的 IGA 与文[7]的算法(记为“[7]”)进行优化比较。

实验中,自变量  $x, y$  取值范围是  $[-1, 1]$ , 种群大小为 50, 个体编码长度均为 20 位. 算法 [7] 的交叉概率取 0.4, 随机取 1 位变异. IGA 中随机取 2 位变异. 实际计算 20 代. 定义  $x, y$  为  $x$  的  $y$  次方, 则我们可以定义个体间综合激励系数  $ij$  的计算公式为:

- (1) 将个体  $i, j$  按位异或, 得到新个体  $k$ ;
- (2) 计算出个体  $k$  直接对应的十进制整数  $I(k)$ ;
- (3)  $ij = \frac{I(k)}{2^{\text{编码位数}-1}}$

实验结果比较如下(以下各图纵坐标表示最大适应值, 图 3 中横坐标表示演化代数, 图 4 中横坐标表示统计次数).

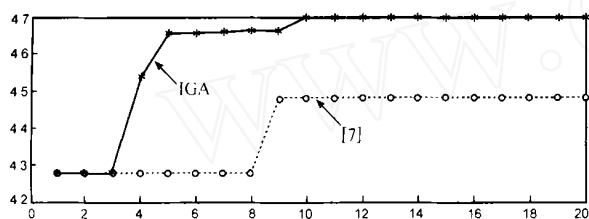


图 3 IGA 与 GA 寻优实验结果比较示意图

从表 1 的数据比较

中我们可以看到, 本文 IGA 的寻优结果比 [7] 好. 考虑到算法中使用

分类	最大值	$x$	$y$
IGA	4.6994	-0.0005	0.0041
[7]	4.4849	0.1263	-0.0178

了随机操作, 仅仅由一次实验得到的结果是不能充分说明问题的, 因此, 我们再用 IGA 和 [7] 分别进行统计比较. 本文中进行 50 次统计实验, 比较结果如表 2.

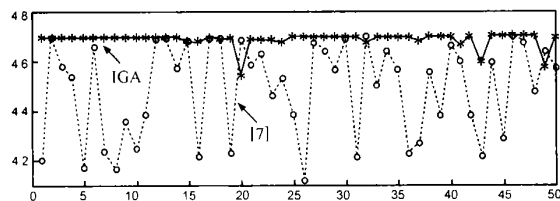


图 4 寻优统计结果比较示意图

表 2 寻优统计结果数据比较表

分类	平均值	标准差	统计实验得到的最大值	最大值大于 4.69 的次数统计
IGA	4.6885	0.0319	4.7000	39
[7]	4.5095	0.1512	4.6998	7

从实验的得到的各图形和表格数据的比较中, 我们可以看到: 在多峰值函数的优化过程中, 本文的 IGA 具有良好的搜索能力.

## 4 结论

通过实验比较, 我们可以看到 IGA 具有更强的寻优性能,

结果令人满意. 由于 IGA 具备能够有效地解决局部搜索能力和全局搜索能力的矛盾、有效维持种群多样性、较好地防止早熟等优点, 从而对今后的实际规划和设计工作能够起到良好的促进和帮助作用. 但是由于实际上抗体是一个具有变区 (VAQ) 和常区 (C 区) 的二官能分子 (bifunctional molecules)<sup>[2]</sup>, 而本文的算法主要根据 V 区的规律而设计的, 因此如何将 C、V 两个区的特点综合考虑设计出更好的算法是值得以后继续深入研究的.

## 参考文献:

- [1] 林焰等. 隔离小生境遗传算法研究[J]. 系统工程学报, 2000, 15 (1): 86 - 91.
- [2] Quagliarella et al. Genetic algorithms in engineering and computer science [M]. New York: John Wiley & Son, 1997.
- [3] Yasuhiko Dote. Soft Computing (Immune Networks) In Artificial Intelligence [A]. SMC 98 Proceeding, 1998 IEEE International Conference on Systems, Man, and Cybernetics [C]. 1998, 1 - 7.
- [4] 吴祥兴等. 混沌学导论[M]. 上海: 上海科学技术文献出版社, 1996.
- [5] 林飞卿等. 细胞免疫学研究进展[M]. 北京: 人民卫生出版社, 1981.
- [6] Dong-Wook Lee, et al. Artificial immune network-based cooperative control in collective autonomous mobile robots [A]. Proceedings, 6th IEEE International Workshop on Robot and Human Communication [C]. RO-MAN 97 Sendai, 1997, 58 - 63.
- [7] Wang Lei, et al. The immune genetic algorithm and its converge [A]. Signal Processing Proceedings ICSP 98, 1998 Fourth International Conference on [C]. 1998, 1347 - 1350.
- [8] John E. Hunt, et al. An adaptive, distributed learning system based on immune system [A]. 1995 IEEE International Conference on Systems, Man and Cybernetics, Intelligent Systems for the 21st Century [C]. 1996, 2494 - 2499.
- [9] 王宁等. 基于嵌入混沌序列的遗传算法[J]. 系统工程理论与实践, 1999, 19(11): 1 - 7.
- [10] 林剑. 免疫遗传学基础[M]. 北京: 北京大学出版社, 1982.
- [11] Isao Tazawa et al. An evolutionary optimization based on the immune system and its application to the VLSI floor-plan design problem [J]. Electrical Engineering in Japan, 1998, 124(4): 27 - 36.

## 作者简介:

罗小平 男, 1974 年出生于四川成都, 博士研究生, 主要研究方向为进化计算、智能控制. email: luo\_xiaoping@sina.com

韦 巍 男, 1964 年出生于浙江东阳, 教授, 博导, 主要研究方向为智能控制理论及应用、机器人控制.