

文章编号: 1001-0920(2005)06-0717-04

自适应多模态免疫进化算法的研究与实现

杨孔雨^{1,2}, 王秀峰¹

(1. 南开大学 信息技术与科学学院, 天津 300071; 2 山东财政学院 计算机信息工程学院, 山东 济南 250014)

摘要: 基于免疫系统的动力学模型, 根据一类抗体可结合多个抗原表位并逐步达到亲和度成熟的机理, 研究并实现了一种多模态免疫进化算法(M IEA)。算法的主要算子包括正选择、记忆细胞产生、超变异和抗体相似性抑制。对不同的多峰值函数进行的仿真实验证明, 算法能够找到多模态问题的全部最优解或尽可能多的局部最优解。通过与同类算法进行比较和计算复杂性分析表明, 该算法不仅计算量小、具有更好的搜索性能, 而且无需任何先验知识, 可实现真正的自适应搜索。

关键词: 多模态优化; 免疫算法; 亲和度成熟; 超变异

中图分类号: TP18; TP302 **文献标识码:** A

Research and implement of adaptive multimodal immune evolution algorithm

YANG Kong-yu^{1,2}, WANG Xiu-feng¹

(1. College of Information Technology and Science, Nankai University, Tianjin 300071, China; 2 School of Computer Information Engineering, Shandong University of Finance, Ji nan 250014, China Correspondent: YANG Kong-yu, Email: yangkongyu@tsinghua.org.cn)

Abstract: Based on immune dynamic model, a multimodal immune evolution algorithm (M IEA) is designed. The algorithm is inspired by the mechanism in which a kind of antibody can identify multi-epitope of an antigen and gradually accomplish affinity maturation in the evolutionary process. The main operators of the algorithm include positive selection, memory cells producing, hypermutation and similar antibodies suppression. The simulation results show that it can find out all optimal solutions and local optimal solutions as many as possible and can realize adaptive searching without any transcendental presumptions.

Key words: multimodal optimization; immune algorithm; affinity maturation; hypermutation

1 引言

现实世界中绝大多数实际优化问题都属于多模态组合优化, 即存在多个最优解, 或者存在一个全局最优解与若干个局部最优解, 类似数学中的多峰函数。由于传统搜索方法一次只能非确定地搜索一个极值点, 对这类问题基本无效。遗传算法(GA)具有概率选择机制和全局搜索的特点, 在解决多模态优化问题中具有一定的可能性, 但搜索过程中适应度小的极值点不断地被淘汰, 因此很难同时搜索出多个峰值, 而往往仅能收敛到一个模态。为此, 人们提

出了许多改进的遗传算法, 试图解决多模态函数的优化问题。如文献[1]提出了基于适应度共享(SH)机制的小生境遗传算法; [2]提出一种新的聚类算法用于多峰值函数的优化, 并与SH算法进行了比较, 效果有所提高; [3]中介绍了多模态优化的分区处理策略, 但仅适用于较为简单且规则的问题(因为一般问题模态区域的数目很难事先确定); [4]提出了用于多模态搜索的平衡空间遗传算法, 较好地克服了现行方法的一些问题, 但该算法对不连续函数仍难以应用。

收稿日期: 2004-06-28; 修回日期: 2004-12-06

基金项目: 山东省教育厅科技计划项目(J02F06, J04A12)

作者简介: 杨孔雨(1967—), 男, 山东巨野人, 教授, 博士生, 从事免疫计算智能、网络信息安全等研究; 王秀峰(1942—), 男, 山东宁津人, 教授, 博士生导师, 从事计算智能、复杂系统建模等研究。

面对GA的种种问题,人们希望开辟新的思路,试图以新的生物学发现作为构建新算法的模型基础。随着免疫计算的兴起,近年来人们提出了许多免疫进化算法。如文献[5]提出了基于免疫记忆和免疫网络调节的集成免疫进化算法并证明了其收敛特性;文献[6]以信息熵作为描述工具,提出了解决多模态函数极值的并行进化免疫算法(IA)。

本文通过对免疫进化机制深入系统的分析,结合免疫系统的动力学模型,基于免疫细胞在自我进化中的亲和度成熟机理,提出一种全新的多模态免疫进化算法(MIEA):通过智能模拟高亲合度抗体的正选择、记忆细胞产生、免疫细胞超变异和抗体相似性抑制等进化机制,可最终找出多模态问题的所有最优解,或一个最优解和尽可能多的局部优化解。

2 免疫进化机理及其动力学模型

2.1 免疫进化机理分析

免疫系统^[7]是一个分布式、自组织和具有动态平衡能力的自适应复杂系统。其主要运行机制有:1)免疫响应:随着对高亲和度抗体(Ab)的免疫正选择和抗原(Ag)驱动的免疫细胞超变异的循环过程,抗体对抗原的亲和度不断增升直至达到亲和度成熟;2)免疫记忆:免疫系统通过学习抗原产生优秀抗体,并对有优良特性的抗体给予奖励(Baldwin效应),利用克隆选择机制产生记忆细胞;3)免疫调节:免疫系统内部各免疫细胞之间形成一个相互作用的动态平衡网络。当有外界抗原入侵时,通过免疫调节,达到新的免疫平衡;在无抗原入侵时,抗体间的相互促进与抑制作用可维持适当数量的必要抗体,以维持免疫平衡。抗体一旦产生,将分组进入淋巴系统实施进化,直至达到亲合度成熟,即抗体与抗原达到最佳匹配。一种抗原有多个不同的抗原表位,而抗体只有一种感受体,因此一个抗原能够和多个抗体结合。

2.2 免疫进化的动力学模型

Famer^[8]将免疫进化过程描述为免疫系统动态方程模型。根据Famer模型^[8]和Baldwin效应的数学机理,可设计评价抗体的4个要素:抗体之间的刺激度(Sim),抗体之间的抑制作用(Sup),抗体与抗原的亲和度(Aff)以及对优秀抗体的奖励(Praise)。这4个因素对于抗体评价(Fitness)的影响可表示为

$$\text{Fitness} = \text{Aff} + \text{Sim} - \text{Sup} + \text{Praise}, \quad (1)$$

其中加减表示各项对评价的作用方向。

新的研究结果还表明^[9]:在T细胞免疫响应期间,高亲和度的B细胞只出现在早期的免疫响应中(结合抗原),这说明高亲和度的B细胞比低亲和度的B细胞发生后期超变异的机会要少。B细胞受体

(BCR)的初始亲和度与其变异概率成反比,这说明了BCR亲和度能够直接调节变异率;而且高亲和度的B细胞未能经历与低亲和度B细胞相同轮回的超变异机会。这样就形成了以免疫正选择、记忆细胞产生、不同概率的细胞超变异、相似性抑制为流程的抗体循环进化过程。

3 多模态免疫进化算法设计

3.1 算法流程

算法主要涉及3方面内容:抗原与抗体的编码形式,抗原与抗体以及抗体与抗体之间的相互作用和整个系统的构造。通常抗原对应待解决的问题,抗体对应问题的解,抗原与抗体的相互作用机制和整个系统的构造将取决于问题本身的特点。在解多模态问题时,可将待求解问题(或其目标函数)定义为抗原,目标函数的多个峰值视为不同的抗原表位,免疫响应中产生的抗体视为问题的解;则抗体与抗原之间的匹配程度(即亲和度)可直接用抗体(解)所对应的目标函数值来评价。若采用二进制对多模态函数的自变量进行编码,则每个自变量为一个二进制编码串,自变量个数即抗体编码的维数,记为 d 。所有自变量的编码串连在一起构成一个抗体,抗体的变异、编码及解码方法与基本遗传算法相同。

整个进化过程完全遵循Famer的动力学模型,具体步骤如下:

1) 首先随机产生规模为 M 的初始抗体群,可分为 N 组以实现并行(或单组)进化。对各组抗体计算亲和度(目标函数求值),进行抗体评价,亲合度越大则评价越高。

2) 在每组抗体群中,依次选择 n_0 个亲和度高且编码不同的抗体(正选择)加入记忆细胞库,记忆细胞可直接进入下一代种群。该过程对应Baldwin效应,体现对优秀抗体的奖励。

3) 若记忆细胞满足结束条件,则算法结束;否则将每组抗体根据亲和度大小划分为高、中、低3个部分(数量可按1:2:1的比例设置)进行超变异操作,使其中高亲和度抗体部分有较小的变异率,低亲和度抗体部分有较大的变异率,而中亲和度抗体部分的变异率介于它们之间。该过程与高亲和度的B细胞不能与低亲和度B细胞有相同机会的变异机理一致,并且对应式(1)中的抗体刺激作用(Sim)。

4) 在记忆细胞库中根据抗体编码的相似性程度进行免疫调节,即进行抗体相似性抑制,抑制阈值设为 δ 。具体抑制方法是:若任意两抗体之间的欧氏距离小于等于 $\delta(D \leq \delta)$,则删除其中亲和度较低的抗体。该过程对应式(1)中的抗体抑制(Sup)和Famer模型中劣等抗体的死亡。其中两抗体 x_i 与 x_j

之间的欧氏距离可按式(2)计算(d 为抗体维数):

$$D = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (2)$$

重复进行上述过程, 直到满足算法循环的结束条件(如达到了预先定义的进化代数 T 或最大亲合度不再变化)为止。最后记忆细胞库中剩余的抗体就是所求问题的多个最优解。

3.2 MIEA 算法描述

Step 1: 在多模态函数的定义域内随机产生 M 个初始抗体。

Step 2: 将 M 个抗体划分为 N 个抗体组, 每组有 $n = M/N$ 个抗体, 每组抗体分别进行抗体评价, 即计算亲和度。

Step 3: 从每组中正选择出 n_0 个亲和度最高且编码各异的抗体, 克隆复制到记忆细胞库, 记忆细胞库规模为 m 。

Step 4: 判断结束条件。满足则输出记忆细胞, 解码后即所求问题的多个优化解, 其亲和度值即是目标函数的多个极值, 其中最大者也是全局最优值, 算法结束; 否则继续。

Step 5: 每一抗体组分别进入免疫系统, 根据抗体的亲和度值划分为高 (P_h), 中 (P_m), 低 (P_l) 3 个子抗体群进行超变异操作, 变异率应满足 $P_h < P_m < P_l$ 。

Step 6: 变异后的抗体与记忆细胞一起进行相似性抑制, 抑制阈值为 δ , 按规模 M 产生新一代抗体种群, 转 Step 2。

4 仿真实验与性能分析

4.1 实例研究与仿真实验

选择用于优化算法测试的两个经典多模态函数:

数:

$$f_1(x) = \sin^6(5\pi(x^{3/4} - 0.05)), \quad x \in [0, 1]; \quad (3)$$

$$f_2(x) = \exp\left[-2\ln 2 \cdot \left(\frac{x - 0.1}{0.8}\right)^2\right] \cdot \sin^6(5\pi x), \quad x \in [0, 1]; \quad (4)$$

函数 f_1 在定义域范围内有 5 个非均匀分布的相等峰值。文献[2]中列出了使用函数 f_1 对其聚类算法与 SH 算法、K 均值聚类方法的适应值共享遗传算法(KM-EAN)进行测试的比较结果, 实验中各算法均运行 20 次, 结果取 20 次的平均值。这里对函数 f_1 也重复测试 20 次, 结果取平均值, 实验条件和算法参数见表 1。实验中将初始抗体群等分为 3 个抗体组 ($N = 3$), 3 组变异率均相同(也可取不同)。算法未使用交叉算子, 峰值误差同其他算法, 抑制阈值 $\delta = 0.1$, 其余条件及实验结果一起列入表 1。

函数 f_2 为峰值不等的典型函数, 在定义域范围内有 5 个大小不等的峰。类似地, 文献[6]应用 IA 算法对函数 f_2 进行了仿真实验, 并与基本遗传算法(SGA)作了对比, 实验条件、参数设置及实验结果见表 2。这里使用本文算法及 SH 算法同样对函数 f_2 进行了仿真测试, 重复实验 20 次, 其余条件和实验结果如表 2 所示。每次算法均能搜索到函数 f_2 的全部峰值。

4.2 实验结果与性能分析

评判多模态优化算法的有效性主要从两个方面考虑: 一是考察能够搜索到多模态函数的峰值个数, 搜索到的峰值个数越多越好; 二是看计算量。计算量用算法评价个体的累计次数来估计, 对一个个体循环一次, 计为评价一次。从上述实验结果可以看出:

表 1 使用函数 f_1 对不同算法的测试条件及结果

算法	交叉率 P_c	变异率 P_m	峰值误差 e 抑制阈值 δ	个体编码长度 L /bits	群体规模 M	迭代次数 T	实验结果(搜索到的峰值个数)	算法计算量
文献[2]算法	1.0	0.01	$e < 0.02$	30	60	50	4.6	3 000
KM-EAN	1.0	0.01	$e < 0.02$	30	60	50	4.6	3 000
SHA	1.0	0.01	$e < 0.02$	30	60	50	4.8	3 000
本文算法(MIEA)	无	$P_m = \{P_h, P_m, P_l\}$ $P_m = \{0.04, 0.25, 0.4\}$	$e < 0.02$ $\delta = 0.1$	12	60 $N = 3$	30	5	1 800

表 2 使用函数 f_2 对不同算法的测试条件及结果

算法	交叉率 P_c	变异率 P_m	峰值误差 e 抑制阈值 δ	个体编码长度 L /bits	群体规模 M	迭代次数 T	实验结果(搜索到的峰值个数)	算法计算量
SGA	1.0	0.01	$e < 0.02$	16	100	200	1	20 000
IA	1.0	0.01	$e < 0.02$	16	100	200	5	20 000
SHA	0.88	0.003	$e < 0.02$	14	100	100	5	10 000
本文算法(MIEA)	无	$P_m = \{P_h, P_m, P_l\}$ $P_m = \{0.05, 0.35, 0.5\}$	$e < 0.02$ $\delta = 0.1$	12	60 $N = 3$	80	5	4 800

本文算法与相关的算法比较,不仅能够搜索到测试函数的全部或较多的峰值,而且因为个体编码较短,没有交叉操作,所以算法整体计算量大大减少.若考虑算法的并行因素,则进化速度将更快

本文算法的主要操作算子包括正选择、记忆抗体产生、细胞超变异及相似性抑制.其中正选择算子的作用在于保证免疫系统中抗体进化进程的方向性,由于采用确定性正向选择,不需要对目标函数进行变换以使函数值为正值.与其他搜索算法一样,变异算子属于全局搜索算子,其作用在于不断改变子空间,从而扩大搜索范围,原则上可扩大到整个搜索空间.算法中的变异率大小由抗体的亲和力高低来调节,这样,使得高亲和度的抗体能够在小范围内继续搜索以提高亲和力,而低亲和度和中等程度的抗体能够扩大搜索范围.这既符合免疫进化过程中免疫细胞的亲和力成熟机理,也具有定向诱发变异的特征,同时也体现了负反馈机理,即超变异后,若提高或降低了抗体的亲和力,则在下一轮变异时,亲和力将反作用于变异率,使低亲和度的抗体有较大的变异率,从而使进化过程具有稳定性.相似性抑制是要在记忆抗体群中消除低亲和力抗体,这与免疫模型中低亲和力抗体的消亡机理相一致.

算法中的主要计算过程包括正选择、超变异和相似性抑制.每一代中算法从每组 $n = M/N$ 个抗体中选择出 n_0 个高亲和力且编码不同的抗体所需时间为 $O(n)$,超变异部分的时间为 $O(nL)$,抑制运算部分的时间为 $O(m^2)$ (m 是每一代中记忆细胞的总数目,每一代的 m 值可能都在变化,但这种变化仅仅是线性增加),因此算法总的时间复杂性为 $O[(1+L)n+m^2]$.

5 结 论

本文提出了一种用于求解多模态问题的免疫进化算法 MEA.仿真实验和性能分析表明:该算法操作算子简单,易于实现,并可体现并行计算的特征;它不仅能搜索到多模态函数的所有全局最优解和尽可能多的局部最优解,具有很好的搜索性能,而且比同类算法计算量大大减少,且无须任何先验知识,从而实现了真正的自适应搜索.

参考文献(References)

- [1] Goldberg D E, Richardson J. Genetic algorithms with sharing for multimodal function optimization [A]. *Proc of the 2nd Int Conf on Genetic Algorithms* [C]. NJ: Lawrence Erlbaum, 1987: 41-49.
- [2] 于歆杰,王赞基.一种新的聚类方法及其在多峰优化中的应用[J]. *清华大学学报*, 2001, 41(4/5): 159-162.
(Yu X J, Wang Z J. A new clustering method and its applications on multimodal optimization [J]. *J of Tsinghua University*, 2001, 41(4/5): 159-162.)
- [3] 徐宗本,张讲社,郑亚林. *计算智能中的仿生学:理论与算法* [M]. 北京: 科学出版社, 2003: 36-42.
(Xu Z B, Zhang J S, Zheng Y L. *Bionics in computational intelligence: Theories and algorithms* [M]. Beijing: Science Press, 2003: 36-42.)
- [4] 刘洪杰,王秀峰.多峰搜索的自适应遗传算法[J]. *控制理论与应用*, 2004, 21(2): 302-304, 310.
(Liu H J, Wang X F. Adaptive genetic algorithm for multi-peak searching [J]. *Control Theory and Applications*, 2004, 21(2): 302-304, 310.)
- [5] 杨孔雨,王秀峰.一种集成免疫进化算法及其收敛性研究[J]. *计算机工程与应用*, 2003, 39(31): 33-35, 118.
(Yang K Y, Wang X F. Research of an integrated immune evolutionary algorithm and its convergence property [J]. *Computer Engineering and Applications*, 2003, 39(31): 33-35, 118.)
- [6] Fukuda T, Mori K, Tsukiyama M. Parallel search for multimodal function optimization with diversity and learning of immune algorithm [A]. *Artificial Immune Systems and Their Applications* [C]. Berlin: Springer-Verlag, 1999: 210-220.
- [7] Lydyard P M, Whelan A, Fanger M W. *Instant notes in immunology* [M]. Beijing: Science Press, 2001: 1-40.
- [8] Farmer J D, Packard N H, Perelson A S. The immune system adaptation and machine learning [J]. *Physica*, 1986, 22(D): 187-204.
- [9] Tien-A n Yang Shih, Eric M effre, Mario Roederer, et al. Role of BCR affinity in T cell-dependent antibody responses in vivo [J]. *Nature Immunology*, 2002, 3(6): 570-575.