

# 一种改进型遗传算法及其在规则提取中的应用

倪世宏<sup>1</sup>, 刘敏智<sup>2</sup>, 夏岩<sup>1</sup>, 苏晨<sup>1</sup>

(1. 空军工程大学 工程学院, 陕西 西安 710038; 2. 93356 部队, 辽宁 锦州 121000)

**摘要:**传统的遗传算法在处理复杂的优化问题时容易早熟收敛,陷入局部最优解。为此将免疫原理引入遗传算法,提出了一种新的亲和度定义策略——正弦型亲和度,该策略在对适应度调整时,前期有近似线性的抑制作用,后期则变得平缓。据此设计了一种改进型遗传算法,以提高遗传算法的全局寻优和局部搜索能力。实验结果表明,改进型遗传算法在处理高维多峰函数的收敛速度和收敛精度方面均优于基本遗传算法。以发动机稳定工作状态为例,应用改进型遗传算法实现了飞行状态分类规则的自动获取。测试结果表明,只要训练样本选取得当,获取的规则简洁、有效。

**关键词:**遗传算法;免疫机制;专家系统;规则提取

**中图分类号:** TP301.6    **文献标识码:** A    **文章编号:** 1009-3516(2008)06-0033-05

众所周知,遗传算法(Genetic Algorithm, GA)<sup>[1]</sup>是一种借鉴生物界自然选择和自然遗传机制的导向随机搜索算法,它具有算法简单、通用性强、鲁棒性好、适于并行处理等特点,被广泛应用于机器学习、神经网络和模式识别等各个方面<sup>[2-4]</sup>。但由于传统遗传算法的基本思想是在编码后对编码空间进行搜索,基本不利用问题空间的知识,而仅基于适应度函数值评价个体并指导搜索,容易造成搜索效率低、收敛速度慢,有时甚至无法收敛到全局最优解<sup>[5]</sup>。

免疫系统是一种高度进化的生物系统,它针对外来抗原的输入产生识别并消灭抗原的抗体,具有很强的学习、记忆和识别能力。1986年, Farmer 和 Packard 首先提出免疫系统的人工智能模型<sup>[6]</sup>。此后,免疫系统逐渐被应用于解决机器故障诊断、配电网规划、入侵检测、参数优化、关联规则挖掘等多种工程实际问题。

近年来,将免疫系统与遗传系统结合构成免疫遗传算法一直受到人们的青睐,将二者结合时,若将所处理问题的目的函数和约束条件对应为抗原,将问题的解对应于抗体,则可将问题空间与 GA 空间融为一体,充分利用遗传算法的全局搜索能力及免疫算法的局部优化能力,提高搜索效率。

## 1 改进型遗传算法

### 1.1 抗体多样性衡量

免疫机制中,抗体多样性的衡量建立在对抗体群相似度衡量的基础之上,而抗体群的相似度又是通过抗体群的信息熵得到的,信息熵描述了抗体的不规则度(多样性)。设有  $N$  个抗体,每个抗体长度为  $M$ ,采用的符号集大小为  $S$ (二进制编码中,  $S=2$ ),则抗体基因座  $v$  的信息熵  $H_v(N)$  可定义<sup>[7]</sup>为  $H_v(N) = -\sum_{u=1}^S p_{uv} \log_2 p_{uv}$ ,  $p_{uv}$  为第  $u$  个符号出现在第  $v$  个基因座上的概率,且可定义为  $p_{uv} = n_{uv}/N$ ,其中  $n_{uv}$  为在基因座  $v$  上出现第  $u$  个符号的总个数( $u=1,2,\dots,S; v=1,2,\dots,M$ )。由此可得平均熵  $H(N) = \frac{1}{M} \sum_{v=1}^M H_v(N)$ 。

\* 收稿日期:2008-04-22

基金项目:教育部重大项目培育基金:城市智能化综合交通与运输安全

作者简介:倪世宏(1963-),男,江苏南京人,教授,博士生导师,主要从事飞机状态监控与地面数据处理研究。

E-mail: nsh649@vip.163.com

2个抗体  $a_i$  和  $a_j$  之间的相似程度称作相似度,记作  $S_{i,j}$ ,定义为  $S_{i,j} = 1/(1 + H(2))$ 。

$S_{i,j} \in [0,1]$ ,  $S_{i,j}$  越大,表示两个抗体越相似。由此,抗体群的相似度即为  $S(N) = 1/(1 + H(N))$ 。

$S(N)$  代表整个种群的相似程度,  $S(N) \in [0,1]$ ,  $S(N)$  越大,抗体群的相似程度越高,多样性程度越低,反之亦然。

## 1.2 免疫调节算子

免疫机制中的自我调节是基于浓度机制进行的。抗体浓度是群体中与某抗体相似的抗体所占的比重,设与抗体  $a_i$  的相似度大于  $\lambda$  的抗体数量为  $n_i$ ,则该抗体浓度  $C_i$  为

$$C_i = \sum_{j=1}^N n_j / N \quad (1) \quad n_j = \begin{cases} 1, & S_{i,j}/N \geq \lambda \\ 0, & S_{i,j}/N < \lambda \end{cases}, i, j \in \{1, 2, \dots, N\} \quad (2)$$

基本遗传算法(SGA)中的遗传操作是,适应度值越大,个体被选择的概率就越大。这样,在算法的后期,适应度高的个体占据了群体的大部分,个体的多样性不足。免疫机制基于亲和度评价抗体优劣,目前对亲和度的定义通常有以下两种策略<sup>[8-9]</sup>:

$$\text{ST1.} \quad A_i = F_i / (1 + \alpha \ln(1 + C_i)) \quad (3) \quad \text{ST2.} \quad A_i = F_i \exp(-kC_i) \quad (4)$$

式中  $F_i$  为抗体  $a_i$  的适应度( $\alpha > 0, k > 0$ )。

亲和度是抗体的适应度与浓度的综合评价。上述两种策略的共同特点是:浓度一定时,适应度越大则亲和度越大,抗体被选择的概率也越大;适应度一定时,浓度越高则亲和度越小,抗体被选择的概率越小。这体现了免疫系统的自我调节功能,既可以保留高适应度的抗体又可以抑制浓度过高的抗体,形成一种多样性保持策略。

借鉴上述思想,这里提出一种新的亲和度定义策略 STN。

$$\text{STN.} \quad A_i = F_i [1 - q \sin(\pi C_i / 2)], (0 < q < 1) \quad (5)$$

## 1.3 免疫记忆算子

免疫系统的记忆效应体现在免疫再次应答中,即免疫初次应答保留的优秀抗体可以加快对再次入侵抗原的识别和消灭,免疫应答体现出更迅速、更有效的现象。在算法中,具体操作为:对于初次应答,全部初始抗体随机产生;对于再次应答,部分初始抗体由记忆单元获取,其余抗体随机产生<sup>[10]</sup>。

## 1.4 算法测试

选取以下几个函数进行算法性能的测试: $f_1$  为 3 维 Sphere 函数,  $f_2$  为 2 维 Rosenbrock 函数,  $f_3$  为 2 维 Rastrigin 函数,  $f_4$  为 10 维 Schwefel 函数。

### 1.4.1 策略仿真比较

以 Rastrigin 为测试函数,设定种群规模  $n = 40$ ,位串长度  $L = 20$ ,交叉概率  $p_c = 0.7$ ,变异概率  $p_m = 0.018$ 。对于 ST1 策略,选取  $\alpha = 7$ ,  $S_0 = 0.5$ ;对于 ST2 策略,选取  $k = 0.8$ ,  $S_0 = 0.5$ 。

免疫调节的本质是通过浓度对适应度的调整,实现多样性的保持。ST1 策略和 ST2 策略通过合适的参数选取,对典型的多峰函数有较好的优化效果。本文提出的 STN 策略基于正弦规律对亲和度进行定义,同样具备 ST1 策略和 ST2 策略的特点,3 种策略对适应度的调整情况见图 1,其中纵坐标为适应度调整比例 ( $A_i/F_i$ ),横坐标为个体浓度  $C_i$ 。

在对适应度的前期调整中,STN 策略与 ST2 策略类似,对适应度有近似线性的抑制作用,但 STN 策略的这种抑制作用在适应度调整后变得平缓,因为对于优良个体,对其抑制作用也应是有限度的,这种特点与 ST1 策略类似。因此,STN 策略在 ST1 策略与 ST2 策略之间较好地寻找到了平衡点。分别使用 3 种策略对 Rastrigin 函数进行优化,独立运行 50 次后取平均值,结果见图 2。

STN 策略与 ST1 策略的优化效果相当,且收敛速度优于 ST2 策略,因此 STN 是一种较好的亲和度定义策略。

### 1.4.2 算法对比实验

将使用 STN 策略优化的遗传算法(Sine based Immune Genetic Algorithm, SIGA)与 SGA 算法进行对比,算法控制参数设置及相应的实验结果见表 1、表 2、图 3 - 图 6 所示。

表 1 算法控制参数设置

Tab. 1 The setting of control parameter of algorithm

函数	种群大小	位串长度	交叉概率	变异概率	浓度阈值	最大代数	运算次数	控制精度
$f_1$	40	20	0.7	0.012	0.3	1 000	30	$10^{-7}$
$f_2$	50	40	0.7	0.009	0.1	1 000	50	$10^{-1}$
$f_3$	50	40	0.7	0.009	0.1	1 000	50	$10^{-10}$
$f_4$	40	20	0.7	0.004	0.3	6 000	20	$2 \times 10^{-6}$

表 2 测试结果对比

Tab. 2 Comparison of testing results

函数	平均收敛代数		平均相对误差		成功次数	
	SGA	SIGA	SGA	SIGA	SGA	SIGA
$f_1$	109.1	107.9	$e-8.715\ 0$	$e-8.958\ 4$	30	30
$f_2$	227.2	53.4	$e-1.916\ 1$	$e-2.437\ 5$	49	50
$f_3$	190.1	167.2	$1.989\ 9e-2$	$3.129\ 7e-11$	49	50
$f_4$	2 501.1	2 389.2	$9.894e-3$	$1.413\ 6e-3$	13	19

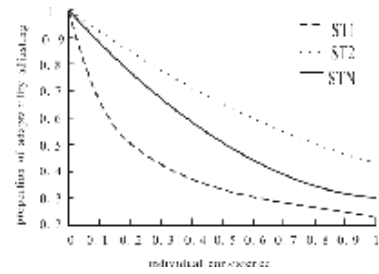


图 1 3 种策略对适应度调整比较

Fig. 1 Comparison of adaptability affected by three strategy

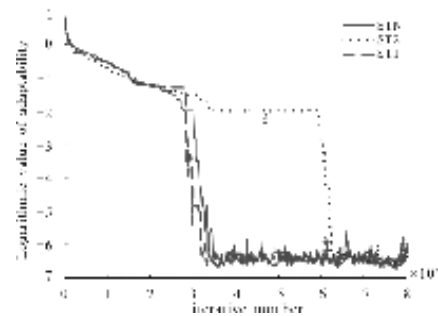


图 2 3 种策略优化性能比较

Fig. 2 Comparison of optimizing performance of three strategy

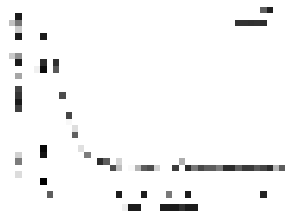


图 3 Sphere 函数对比测试  
Fig. 3 Testing comparison of Sphere function

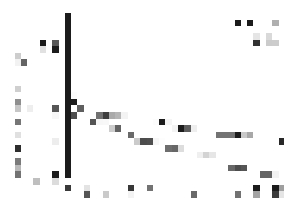


图 4 Rosenbrock 函数对比测试  
Fig. 4 Testing comparison of Rosenbrock function



图 5 Rastrigin 函数对比测试  
Fig. 5 Testing comparison of Rastrigin function



图 6 Schwefel 函数对比测试  
Fig. 6 Testing comparison of Schwefel function

由图 3 - 图 6 及表 2 可知, SIGA 算法在收敛精度和收敛速度方面均优于 SGA 算法。其一, 对于简单函数  $f_1$ , SIGA 算法与 SGA 算法的优化效果几乎相当; 对于病态函数  $f_2$ 、多峰函数  $f_3$  及高维多峰函数  $f_4$ , SIGA 算法的优化效果则明显好于 SGA 算法。其二, SIGA 算法可以有效地跳出局部极值, 例如对函数  $f_3$ , 在 50 次独立运行中, SGA 算法有落入局部极值的情况, 而 SIGA 算法没有; 对函数  $f_4$ , 在 20 次独立运行中, SGA 算法有 7 次落入局部极值的情况, 而 SIGA 算法只有 1 次。需要说明的是, 伴随免疫机制的加入, SIGA 算法的运行时间会多于 SGA 算法。

## 2 改进型遗传算法在规则提取中的应用

### 2.1 特征参数提取

与飞行状态密切相关的关键参数有以下 10 个: 发动机低压转子转速  $N_1$ 、发动机高压转子转速  $N_2$ 、发动机油门杆位置 PYD、发动机涡轮后燃气温度  $T_4$ 、倾斜角 KPEH、俯仰角 TAHG、垂直过载  $N_y$ 、迎角 ATAKA 及发动机喘振信号 PIOM 和武器发射信号 BK, 其中前 8 个参数为模拟量, 后 2 个参数为开关量。从飞行状态的数据特征量库中取 14 个特征量作为参数属性, 分别记作  $C_1, C_2, \dots, C_{14}$ , 其中  $C_1 - C_{12}$  分别表示结束时刻与起始时刻的发动机低压转子转速差、高压转子转速差、油门杆位置差、涡轮后燃气温度差、过程中的最大和最小倾斜角、最大和最小俯仰角、最大和最小垂直过载、最大和最小迎角,  $C_{13}$  和  $C_{14}$  取值为若连续 2 个采样点  $PIOM = 1$ , 则  $C_{13} = 0$ , 若连续 2 个采样点  $BK = 1$ , 则  $C_{14} = 0$ 。

### 2.2 算法设计

前期准备工作需要获得某特定飞行状态的正例样本  $\{T_1, T_2, \dots, T_n\}$  和反例样本  $\{F_1, F_2, \dots, F_m\}$ 。对于正例样本,采用统计的方法获得上述前 12 个模拟量参数属性的极限值,即  $\{\text{Min}(C_i), \text{Max}(C_i)\}, (i = 1, 2, \dots, 12)$ ,对于后 2 个开关量参数属性,统计其相应的采样值(0 或 1),这样便得到 26 个属性。

### 2.2.1 抗体编码设计

采用二进制编码。将上述 26 个属性按照任意给定的顺序排列,例如:  $\{\text{Min}(C_1), \text{Max}(C_1), \text{Min}(C_2), \text{Max}(C_2), \dots, \text{Min}(C_{12}), \text{Max}(C_{12}), C_{13}, C_{14}\}$ 。个体长度为 26 位,该位为 1 表示相应属性被选中,该位为 0 表示相应属性未被选中,例如个体“110000000000000000000000”表示的规则即为  $\{C_1 > \text{Min}(C_1) \text{ 且 } C_1 < \text{Max}(C_1)\}$ 。

### 2.2.2 适应度函数设计

优良的分类产生式规则应较多地排斥反例且尽量简洁,适应度函数设计为  $F(x_1, x_2) = (1 - w)x_1 - wx_2$ 。其中,  $x_1$  为规则能够排斥的反例个数,  $x_2$  为规则中属性的个数,也就是抗体个体中基因位为 1 的个数,  $w$  为调整前后两项的权值,取为 0.1。

### 2.2.3 算法流程

- 1) 输入样本及设定参数:准备正例、反例样本,由正例样本计算属性值,设定种群规模  $N$ 、交叉率  $p_c$  和变异率  $p_m$ ;
- 2) 抗体群初始化及编码:根据应答的情况,初次应答则初始抗体全部随机产生,再次应答则初始抗体一部分随机产生,另一部分通过记忆单元产生,并按照二进制方式进行编码;
- 3) 抗体群遗传操作:计算各抗体的适应度  $F_i$ ,使用选择、交叉和变异算子产生新抗体;
- 4) 种群多样性衡量:计算抗体群相似度  $S$ ,对给定阈值  $S_0$ ,若  $S \geq S_0$ ,则转步骤 5),否则转步骤 7);
- 5) 免疫调节操作:随机产生  $R$  个新抗体(此时种群规模为  $N + R$ ),计算抗体浓度  $C_i$ ,并根据  $F_i$  和  $C_i$  计算抗体亲和度  $A_i$ ;
- 6) 群体更新:选取亲和度较高的  $N$  个抗体组成新抗体群,对更新后的群体进行遗传操作;
- 7) 判断终止条件:若达到截止代数,则按比例选取亲和度较高的部分抗体存入记忆单元,然后终止,否则转步骤 6);
- 8) 分析评价:获得产生式规则并由领域专家进行评价。

## 2.3 实验分析

以发动机稳定工作状态为例,验证方法的有效性。选取发动机稳定工作状态正例样本 30 个,反例样本 56 个,位串长度  $L = 26, N = 40, p_c = 0.7, p_m = 0.01, S_0 = 0.5$ ,最大代数 100,算法独立运行 20 次。图 7 是规则获取的进化过程曲线。

最终获得最优个体的适应度值  $F = 50.1$ ,个体的表示字符串为“01000000000000000000000011”,表示的分类产生式规则为  $\{C_1 < 2.166 \text{ 且 } C_{13} = 1 \text{ 且 } C_{14} = 1\}$ 。其物理含义为:若发动机低压转子转速差小于 2.166,同时无发动机喘振及武器发射信号,则在此时间段内发动机处于稳定工作状态。同时获得另外一个最优个体的适应度值  $F = 50.1$ ,个体的表示字符串为“00000100000000000000000011”,表示的分类产生式规则为  $\{C_3 < 3.278 \text{ 且 } C_{13} = 1 \text{ 且 } C_{14} = 1\}$ 。其物理含义为:若发动机油门杆位置差小于 3.278,同时无发动机喘振及武器发射信号,则在此时间段内发动机处于稳定工作状态。

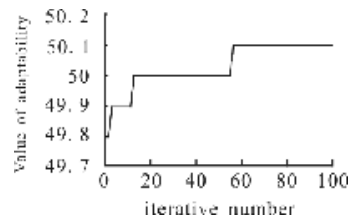


图 7 规则获取的进化过程曲线  
Fig.7 The evolution process of attaining rule on engine steady state

## 3 结论

将免疫机制融入遗传算法可有效改进遗传算法处理高维多峰函数的收敛速度和收敛精度。将关键飞行参数特征量符号化后,应用本文提出的改进型遗传算法,可自动获取飞行状态分类规则。对发动机稳定工作状态的规则提取结果表明,该方法得到的规则简洁、正确。

## 参考文献:

- [ 1 ] Holland J H. Adaptation in Natural and Artificial Systems [M]. Ann Arbor, Michigan, USA: University of Michigan, 1975.
- [ 2 ] Whitley D, Starkweather T, Shaner D. The Traveling Salesman and Sequence Scheduling: Quality Solutions Using Genetic Edge Recombination [M] // Davis L. Handbook of Genetic Algorithm. New York: Van Nostrand Reinhold, 1991: 350 - 372.
- [ 3 ] Holland J H. Genetic Algorithms and Classifier Systems: Foundations and Future Directions [C] // Proceedings of the Second International Conference on Genetic Algorithm. New Jersey: Lawrence Erlbaum Associates Publishers, 1989: 82 - 89.
- [ 4 ] Hegde S U. Efficacy of Parallel Genetic Algorithms [D]. Charlottesville, VA: University of Virginia, 1988.
- [ 5 ] 王小平, 曹立明. 遗传算法——理论、应用与软件实现 [M]. 西安: 西安交通大学出版社, 2002.  
WANG Xiaoping, CAO Liming. Genetic Algorithms Theory, Application and Implement in Software [M]. Xi'an: Xi'an Jiaotong University Publishing Company, 2002. (in Chinese)
- [ 6 ] Farmer J D, Packard N H. The Immune System, Adaptation, and Machine Learning [J]. Physica, 1986, 22(2): 187 - 204.
- [ 7 ] 王熙法, 张显俊, 曹先彬, 等. 一种基于免疫原理的遗传算法 [J]. 小型微型计算机系统, 1999, 20(2): 38 - 41.  
WANG Xifa, ZHANG Xianjun, CAO Xianbin, et al. An Improved Genetic Algorithm Based on Immune Principle [J]. Mini - Micro Systems, 1999, 20(2): 38 - 41. (in Chinese)
- [ 8 ] 郑德玲, 梁瑞鑫, 付冬梅, 等. 人工免疫系统及人工免疫遗传算法在优化中的应用 [J]. 北京科技大学学报, 2003, 25(3): 284 - 287.  
ZHENG Deling, LIANG Ruixin, FU Dongmei, et al. Application of Artificial Immune System and Artificial Immune Genetic Algorithm to Optimization [J]. Journal of Beijing University of Science and Technology, 2003, 25(3): 284 - 287. (in Chinese)
- [ 9 ] 陈丽安, 张培铭. 免疫遗传算法在 MATLAB 环境中的实现 [J]. 福州大学学报: 自然科学版, 2004, 32(5): 554 - 559.  
CHEN Lian, ZHANG Peiming. Realization of Immune Genetic Algorithm in MATLAB [J]. Journal of Fuzhou University: Natural Science Edition, 2004, 32(5): 554 - 559. (in Chinese)
- [ 10 ] 朱红霞, 沈炯, 王培红, 等. 基于免疫遗传算法的模糊优化控制及其仿真 [J]. 东南大学学报: 自然科学版, 2005, 35(1): 64 - 68.  
ZHU Hongxia, SHEN Jiong, WANG Peihong, et al. Fuzzy Optimization Control Based on Immune Genetic Algorithm and Its Simulating Study [J]. Journal of Southeast University: Natural Science Edition, 2005, 35(1): 64 - 68. (in Chinese)

(编辑: 姚树峰, 徐敏)

## An Improved Genetic Algorithm and Its Application in Rules Extraction

NI Shi - hong<sup>1</sup>, LIU Min - zhi<sup>2</sup>, XIA Yan<sup>1</sup>, SU Chen<sup>1</sup>

(1. Engineering Institute, Air Force Engineering University, Xi'an 710038, China; 2. Unit 93356, Jingzhou 121000, Liaoning, China)

**Abstract:** The traditional SGA has the characters of converging early and obtaining easily local best result in the case of processing complicated optimizing problem. So the immune principle is brought in SGA and a new affinity definition strategy (affinity based on sine function) is put forward. This strategy restrains adaptability in the manner of approximate line prophase and flatness anaphase, and a sine based immune genetic algorithm (SIGA) is designed to improve its global and local searching abilities. The experiment results demonstrate that the convergent precision and speed of SIGA is better than those of SGA. Taking the engine stabilization for example, by applying the SIGA, picking up the aviation state classification rules is successfully realized. The testing results indicate that the rule acquired is simple and effective if the training sample is selected properly.

**Key words:** genetic algorithm; immune mechanism; expert system; rules extraction