

采用生物信息克隆的免疫算法

张波涛^{1,2}, 刘士荣^{1,2}, 吕 强²

(1. 华东理工大学 自动化研究所, 上海 200237; 2. 杭州电子科技大学 自动化学院, 浙江 杭州 310018)

摘要: 受克隆选择过程生物学原理的启发, 提出了一种采用生物信息克隆的免疫算法. 抗体克隆依赖于一个动态平衡的网络, 并与遗传因素相关. 为了解决传统克隆过程中信息不能充分利用的问题, 该进化算法将环境信息、抗体历史信息以及抗体遗传特征积累的影响引入人工免疫系统, 用这多种信息作为先验知识为克隆过程提供决策支持, 引导抗体系统的更新. 同时采用实数与二进制混合编码方式增加种群多样性, 提高收敛速度, 然后分析了该算法的收敛性. 仿真实验结果表明, 该克隆策略能较大的提高免疫克隆算法的优化能力; 与几种高级免疫克隆算法和进化算法相比, 该算法寻优精度高, 收敛速度快, 能有效的克服早熟现象, 并具有很好的高维优化能力.

关键词: 智能计算; 人工免疫系统; 克隆选择

中图分类号: TP18 **文献标识码:** A

Immune algorithm with biologic information clone

ZHANG Bo-tao^{1,2}, LIU Shi-rong^{1,2}, LÜ Qiang²

(1. Research Institute of Automation, East China University of Science and Technology, Shanghai 200237, China;

2. Institute of Automation, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China)

Abstract: The clonal selection process depending on a homeostasis network is closely linked with genetic factors. Inspired by this immunology principle, an immune algorithm with biologic information clone(IABIC) is proposed. To deal with the missing information in traditional artificial clonal selection process, this evolutionary algorithm introduces the environment information, history information of antibody and hereditary characteristics into the artificial immune system, which serves as the prior knowledge to provide decision support for clonal selection process and conduct the rebirth of antibody system. Meanwhile, hybrid coding was adopted to improve the diversity of the population and the convergence speed, and then, the global convergence of this algorithm was analyzed. Simulation results indicate that this clonal strategy makes significant improvement in optimizing the ability of immune clonal algorithm. Compared with other immune clonal optimization algorithms and evolution algorithms, it has higher precision, faster convergence, remarkable convergence character for higher dimensional optimization, and the ability for avoiding premature convergence.

Key words: intelligent computing; artificial immune system; clonal selection

1 引言(Introduction)

人工免疫系统是模仿高等生物免疫系统功能的一种智能方法, 鉴于其解决问题的潜力, 它已经成为人工智能领域继进化计算、神经网络、模糊逻辑之后的又一研究热点^[1]. 其成果涉及控制、故障诊断、优化计算等领域^[2~4]. 免疫克隆是免疫系统研究的重要分支, 它是由De Castro首次提出的^[5,6]. 因其卓越的信息处理能力, 现在已经得到越来越多的应用和关注^[7,8,10].

以往研究侧重于获得性免疫记忆, 克隆过程被处理为根据亲和度按比例复制和高概率变异, 免疫过程的大量信息被忽略, 从而导致演化的盲目性, 进而限制免疫克隆算法的寻优能力. 受此启发, 将环境信

息、历史信息以及抗体遗传特征积累引入人工免疫克隆系统, 以此模拟利用更充分的信息进行克隆选择的过程, 这里称其为采用生物信息克隆的免疫算法(immune algorithm with biologic information clone, IABIC). 将IABIC与几种不同的算法进行对比测试, 结果表明该算法性能良好, 并且可以解决各类复杂模型的优化问题.

2 采用生物信息克隆的免疫算法(Immune algorithm with biologic information clone)

免疫优化问题为寻找 $\mathbf{x}^* = \{x_1, x_2, x_3, \dots, x_c\}$, 使 $f(\mathbf{x}^*) = \min\{f(\mathbf{x}') : \mathbf{x}' \in I\}$. 其中: $I = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ 为抗体系统, I_k 为第 k 代抗体种群空间, c 为问题规模, n 为种群规模. 应用举例: 在机器

人路径规划中,寻找一组参数使得路径最短.

2.1 IABIC的克隆选择算子(Clonal selection operator of IABIC)

克隆选择算子(CSO)包括:克隆操作、免疫基因操作、克隆选择操作、克隆修正操作.克隆修正操作为IABIC的核心部分,其余3部分根据传统人工克隆理论进行构架^[6,7].演化过程表示为

$$I_k \rightarrow (OP^C(M) + W) \cup OP^{Ge}(OP^C(M)) \rightarrow \text{sel}(OP^{Ge}(OP^C(M))) + \text{init}(W) = I'_k, \quad (1)$$

$$I_{k+1} = \text{mod}(I'_k). \quad (2)$$

$OP^C(M)$ 为记忆集 M 复制, $OP^{Ge}(OP^C(M))$ 为抗体基因操作, $\text{sel}(OP^{Ge}(OP^C(M)))$ 为记忆集更新, $\text{init}(W)$ 为抗体消亡和更新, I'_k 为过渡抗体集, $\text{mod}(I'_k)$ 为过渡抗体系统更新.

第 i 个抗体的抗体-抗原亲和度 appe_i 定义为

$$\text{appe}_i = \alpha_1 * \frac{\sum_{j=1}^n f(\mathbf{x}_j)}{f(\mathbf{x}_i)}. \quad (3)$$

第 i 个抗体的克隆规模 cnum_i 定义为

$$\text{cnum}_i = \text{fix}(\alpha_2 * \text{appe}_i). \quad (4)$$

第 i 个抗体的变异量 m_{q_i} 定为

$$m_{q_i} = \alpha_3 \frac{1}{\text{appe}_i} \times v_1. \quad (5)$$

$f(\mathbf{x}_i)$ 是第 i 个抗体的函数值, $\text{fix}(\cdot)$ 为向下取整函数, α_1, α_2 为实数, v_1 为服从高斯分布的随机实数,且 $v_1 \in (0, 1)$.

2.2 生物信息对传统克隆的修正(Amendment to traditional CSO with biologic information)

抗体系统信息包括体液环境信息矩阵和抗体历史信息矩阵以及遗传信息积累矩阵.因为体液中的抗原抗体复合物会对克隆产生抑制作用,复合物的多少与亲和度有关,所以环境信息矩阵储存亲和度的相关信息.环境信息矩阵 E 主要存储两部分信息:

$$E = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad (6)$$

\mathbf{e}_1 储存全局最优抗体与当前抗体亲和度的关系, \mathbf{e}_2 储存历史最优抗体与当前抗体亲和度的关系.抗体历史信息矩阵 H 也包括两部分:

$$H = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}, \quad (7)$$

\mathbf{h}_1 存储全局最优抗体, \mathbf{h}_2 储存历史最优抗体.

IABIC中认为有利于清除有害抗原的信息会在

抗体系统中不断积累,其遗传信息积累表示为

$$\text{history}_i = \alpha_4 \times v_2 \times |\mathbf{x}_i - \sum_{j=1}^n \mathbf{x}_{\text{local}j}/n|. \quad (8)$$

\mathbf{x}_i 为当前抗体, $\mathbf{x}_{\text{local}j}$ 为第 j 个抗体的历史最优, α_4 为比例常数, ξ_2 为服从高斯分布的随机实数, $v_2 \in (0, 1)$.

环境信息更新1:

$$\mathbf{e}_1^{k+1} = \mathbf{e}_1^k + \mathbf{z}(|f(\mathbf{x}_i) - f(\mathbf{x}_{\text{global}})| + \alpha_5). \quad (9)$$

环境信息更新2:

$$\mathbf{e}_2^{k+1} = \mathbf{e}_2^k + \mathbf{z}(|f(\mathbf{x}_i) - f(\mathbf{x}_{\text{local}i})| + \alpha_5). \quad (10)$$

抗体信息更新:

$$\mathbf{h}_1 = \mathbf{x}_{\text{global}}, \quad (11)$$

$$\mathbf{h}_2 = \mathbf{x}_{\text{local}i}, \quad (12)$$

$\mathbf{x}_{\text{global}}$ 为全局最优抗体, $\mathbf{x}_{\text{local}i}$ 为第 i 个抗体的历史最优, α_5 为速度调节因子, \mathbf{z} 是 n 维元素为1的向量, \mathbf{e}_1^k 与 \mathbf{e}_2^k 为第 k 代环境信息矩阵的元素.

利用以上抗体信息对克隆后的抗体进行更新

$$T_1 = U \cdot (1/\mathbf{e}_1) + (1 + \xi \times U) \cdot \mathbf{h}_1, \quad (13)$$

$$T_2 = U \cdot (1/\mathbf{e}_2) + (1 + \xi \times U) \cdot \mathbf{h}_2, \quad (14)$$

$$T_3 = \text{history}_i, \quad (15)$$

ξ 为更新因子($\xi_{n+100} = \xi_n + \gamma$), γ 为控制因子,当 $n \leq 100$ 时取 $\xi = 0.7$, “ \cdot ”为矩阵元素乘, U 为服从高斯分布的随机矩阵.第 i 个抗体 \mathbf{x}_i 被修正为

$$\mathbf{x}_i = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3, \quad (16)$$

其中 $\beta_1, \beta_2, \beta_3$ 为修正权值.

3 IABIC算法的收敛性(Convergence of IABIC)

定义1 S 为有界闭区域,目标函数在 S 上连续.用IABIC求解连续优化问题定义为:从 $\forall \mathbf{x}_0 \in S$ 开始,搜索至少一个 $\mathbf{x}^* \in S$,满足 $f(\mathbf{x}^*) \leq f(\mathbf{x})$ (极小值问题), $\forall \mathbf{x} \in S$.若抗体系统中至少存在一个 $\mathbf{x}^* \in S$,则称种群处于状态 D_1 ;若抗体系统中不存在 $\mathbf{x}^* \in S$,则称种群处于状态 D_0 .

引理1(Borel-Cantelli) 设 A_1, A_2, \dots 是概率空间上的一个事件序列,令 $p_k = P\{A_k\}$.若

$$\sum_{k=1}^{\infty} p_k < \infty,$$

则

$$P\left\{\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right\} = 0. \quad (17)$$

若 $\sum_{k=1}^{\infty} p_k = \infty$, 且各 A_k 相互独立, 则

$$P\{\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\} = 1. \quad (18)$$

定理 1 IABIC 算法以概率 1 收敛.

证 根据定义 1, 记第 k 代抗体系统处于状态 D_0 的概率为

$$P_0(k) = P\{D_0^k\}, \quad (19)$$

处于状态 D_1 的概率为

$$P_1(k) = P\{D_1^k\}, \quad (20)$$

由免疫记忆单元的性质可知, $P\{D_0^{k+1}/D_1^k\} = 0$, 故

$$P_0(k+1) = P\{D_0^{k+1}/D_0^k\} \times P\{D_0^k\}. \quad (21)$$

设 $P'_0(k+1)$ 为经传统搜索后, 不存在 $\mathbf{x}^* \in S$ 的概率, $P'_1(k+1)$ 为 $\exists \mathbf{x}^* \in S$ 的概率, 由免疫记忆的性质可知

$$P'_1(k+1) \leq P_1(k+1) < 1.$$

由文献[7]的定理 1 证明可知 $P'_0(k+1) \in (0, 1)$. 由

$$\begin{aligned} P'_1(k+1) &= 1 - P'_0(k+1), \\ 0 < P'_1(k+1) &\leq P_1(k+1) < 1, \end{aligned}$$

故

$$0 < (1 - P_1(k+1)) = P_0(k+1) < 1. \quad (22)$$

由公式(21)可知, 从 k 代不存在最优解, $k+1$ 代仍

然不存在最优解的事件 Miss_k 的发生概率为

$$\begin{aligned} P(\text{Miss}_k) &= P\{D_0^{k+1}/D_0^k\} \times P\{D_0^k\} = \\ &P_0(k+1) \in (0, 1). \end{aligned} \quad (23)$$

令 $P(\text{Miss}_k)$ 的最大值为 $\max P(\text{Miss}_k) = t$, 则 $t \in (0, 1)$.

第 k 代不存在最优解的事件 Event_k 是由从 I_0 到 I_k , k 次迁移 $\text{Miss}_1, \text{Miss}_2, \dots, \text{Miss}_k$ 一系列组合事件构成. 其概率和

$$\begin{aligned} \sum_{k=1}^{\infty} P(\text{Event}_k) &\leq \sum_{k=1}^{\infty} t^k = \lim_{k \rightarrow \infty} \frac{t(1-t^k)}{1-t} = \\ &\frac{t}{1-t} < \infty. \end{aligned} \quad (24)$$

由引理 1 可知

$$P\{\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} [|f(\mathbf{x}^*(k)) - f^*| \geq 0]\} = 0, \quad (25)$$

因此

$$P\{\lim_{k \rightarrow \infty} f(\mathbf{x}^*(k)) = f^*\} = 1,$$

即 IABIC 算法以概率 1 收敛. 证毕.

4 仿真试验与性能分析(Simulation and performance analysis)

为评测 IABIC 的性能, 下面采用一些实例进行测试, 测试分两部分: 1) 给定评价次数下的性能比较. 2) 高维复杂函数测试. 实例从文献[9]中选择, 如表 1 所示.

表 1 测试函数

Table 1 Test function

函数序号	最优值	函数表达式	最优解	搜索空间
f_1	0	$f_1 = \sum_{i=1}^n x_i^2$	$0, 0, \dots, 0$	$-100 \leq x_i \leq 100$
f_2	0	$f_2 = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	$0, 0, \dots, 0$	$-10 \leq x_i \leq 10$
f_3	0	$f_3 = \sum_{i=1}^n (\sum_{j=1}^i x_j)^2$	$0, 0, \dots, 0$	$-100 \leq x_i \leq 100$
f_4	0	$f_4 = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$	$0, 0, \dots, 0$	$-30 \leq x_i \leq 30$
f_5	0	$f_5 = \sum_{i=1}^n ([x_i + 0.5])^2$	$0, 0, \dots, 0$	$-100 \leq x_i \leq 100$
f_6	0	$f_6 = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10)$	$0, 0, \dots, 0$	$-5.12 \leq x_i \leq 5.12$
f_7	0	$f_7 = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos 2\pi x_i) + 20 + \exp(1)$	$0, 0, \dots, 0$	$-32 \leq x_i \leq 32$
f_8	0	$f_8 = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(x_i/\sqrt{i}) + 1$	$0, 0, \dots, 0$	$-600 \leq x_i \leq 600$

4.1 给定评价次数下的算法性能比较

(Comparison with fix evaluation times of function)

本指标表征相同计算资源耗费调价下算法的性能. 为了与FCA, Opt-IMMAlg进行比较, 将IABIC与二者放于相同配置下: 种群规模25, 评价次数 1×10^5 , 变量维数为30, IABIC的克隆规模按

公式(4)计算, 各算法随机独立运行20次. 统计结果如表2所示, FCA, Opt-IMMAlg的数据来源于文献[8].

表2的统计结果表明, 就统计结果总体而言, 在给定评价次数的条件下, IABIC的收敛性能优于FCA, Opt-IMMAlg. 因为当前配置是FCA的最优配置, 如果IABIC采用自身的最佳种群数量, 效果会更好.

表2 IABIC与Opt-IMMAlg, FCA在给定评价次数下的算法性能比较

Table 2 Comparison of IABIC, Opt-IMMAlg and FCA with fix evaluation times

函数	成功率/%	IABIC		FCA		Opt-IMMAlg	
		均值	方差	均值	方差	均值	方差
f_1	100	0	0	0	0	0	0
f_2	100	0	0	0	0	0	0
f_3	100	0	0	4.53×10^{-7}	1.52×10^{-7}	0	0
f_4	0	6.0825×10^{-6}	4.4179×10^{-11}	2.74	2.49	16.29	13.96
f_5	100	0	0	0	0	0	0
f_6	80	1.7066×10^{-9}	1.9949×10^{-14}	0	0	0.596	4.178
f_7	0	4.4409×10^{-15}	0	1.56×10^{-7}	3.12×10^{-7}	0	0

4.2 高维函数的测试(Test of high dimensional function)

当测试函数维数变高, 求解会越来越复杂, 国内针对高维函数的测试很少. 该指标可以有效检验算法处理复杂问题的能力. IMCPA为文献[10]提出的一种针对高维函数的优化算法, nIMCPA是无记忆单元的免疫克隆算法, AEA是一种具有较好高维性能的进化算法. IABIC的种群规模为5, 克隆规模按公式(4)计算. 为与文献[10]的统计试验一致, 本文采取与之相同的指标, 统计结果如表格3所示. 其中, 对应算法下的数据为达到精度 ε 所需的评价次数均值. IMCPA, nIMCPA, AEA的数据均来源于文献[10].

由表3的统计结果可知, IABIC与IMCPA计算资源消耗之比在200~400维之间约为1/5~2/3, 到达1000维时约为1/5~2/3. 与其余两种算法相比, 则具有更明显的优势, 最佳情况下可达1/10. 由此可见, IABIC在超过200维时可以节省大量计算资源, 适用于处理复杂问题.

5 结论(Conclusions)

为模拟免疫克隆的生物学机理, 完善人工免疫克隆过程, 本文利用环境信息、历史信息以及抗体遗传特征积累对传统克隆后的抗体进行修正, 有效的提高了免疫克隆算法的寻优能力, 并证明了

这种改进的有效性. 更重要的是该克隆策略为混沌、小生境、记忆单元等其他人工免疫克隆计算研究提供了良好的平台.

表3 对高维函数的测试结果

Table 3 High dimensional function test

维数	$f_6(\varepsilon = 0.1)$			
	IABIC	IMCPA	nIMCPA	AEA
100	2890	4988	7221	4798
200	3692	5747	13354	10370
400	6297	12563	15475	23588
1000	16015	24408	37250	46024
维数	$f_7(\varepsilon = 0.001)$			
	IABIC	IMCPA	nIMCPA	AEA
100	1703	5784	9759	22710
200	1934	9728	11305	43527
400	2873	13915	29477	78216
1000	3344	26787	33343	160940
维数	$f_8(\varepsilon = 0.0001)$			
	IABIC	IMCPA	nIMCPA	AEA
100	2232	6713	12166	17228
200	2681	8460	21577	36760
400	3518	15365	36515	61975
1000	6383	30906	75217	97600

参考文献(References):

- [1] DASGUPTA D, FORREST S. Artificial immune systems in industrial applications[C] // *Proceedings of the 2nd International Conference on Intelligent Processing and Manufacturing of Materials*. Hawaiiian, USA: IEEE, 1999: 257 – 267.
- [2] WHITBROOK A M, AICKELIN U, GARIBALDI J M. Idiotypic immune networks in mobile-robot control[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2007, 37(6): 1581 – 1598.
- [3] 李中华, 张雨浓, 谭洪舟, 等. 一类具有精英学习能力的增强型人工免疫网络优化算法[J]. *控制理论与应用*, 2009, 26(3): 283 – 290. (LI Zhonghua, ZHANG Yu'nong, TAN Hongzhou, et al. An enhanced artificial immune network with elitist-learning capability for optimization problems[J]. *Control Theory & Applications*, 2009, 26(3): 283 – 290.)
- [4] GONG M G, JIAO L C. Multiobjective immune algorithm with non-dominated neighbor-based selection[J]. *Evolutionary Computation*, 2008, 16(2): 225 – 255.
- [5] DE CASTRO L N, VON ZUBEN F J. An evolutionary immune network for data clustering[C] // *Proceedings of the 6th Brazilian Symposium on Neural Networks*. Rio de Janeiro, Brazil: IEEE, 2000: 84 – 89.
- [6] DE CASTRO L N, VON ZUBEN F J. Learning and optimization using the clonal selection principle[J]. *IEEE Transaction on Evolutionary Computation*, 2002, 6(3): 239 – 351.
- [7] 杜海峰, 公茂果, 刘若辰, 等. 自适应混沌克隆进化规划算法[J]. *中国科学E辑*, 2005, 35(8): 817 – 829. (DU Haifeng, GONG Maoguo, LIU Ruochen, et al. Adaptive chaos clonal evolutionary programming[J]. *Science in China Series E*, 2005, 35(8): 817 – 829.)
- [8] NITESH KHILWANI, ANOOP PRAKASH, RAVI SHANKAR, et al. Fast clonal algorithm[J]. *Engineering Applications of Artificial Intelligence*, 2008, 21(1): 106 – 128.
- [9] XIN Y. Evolutionary programming made faster[J]. *IEEE Transactions on Evolutionary Computation*, 1999, 3(2): 82 – 101.
- [10] 杜海峰, 公茂果, 焦李成, 等. 用于高维函数优化的免疫记忆克隆规划算法[J]. *自然科学进展*, 2004, 14(8): 925 – 933. (DU Haifeng, GONG Maoguo, JIAO Licheng, et al. Immune memory clonal programming algorithm for high-dimensional function numerical optimization[J]. *Progress in Natural Science*, 2004, 14(8): 925 – 933.)

作者简介:

张波涛 (1982—), 男, 博士研究生, 研究方向为智能机器人与智能系统、智能优化算法, E-mail: waveact@163.com;

刘士荣 (1952—), 男, 教授, 博士生导师, 研究方向为复杂系统建模、优化与控制、智能系统与智能机器人等, E-mail: liushirong@hdu.edu.cn;

吕强 (1977—), 男, 讲师, 博士, 研究方向为智能优化算法。