

文章编号: 1000-8152(2004)06-0989-04

## 一类自适应范围 DNA 软计算模型

黄自元<sup>1,3</sup>, 师黎<sup>1,2</sup>, 费敏锐<sup>1</sup>, 胡传平<sup>1,3</sup>, 王鹤寿<sup>3</sup>

(1. 上海大学 机电工程与自动化学院, 上海 200072; 2. 郑州大学 信息工程学院, 河南 郑州 450052;  
3. 公安部 上海消防研究所, 上海 200032)

**摘要:** 提出了自适应范围 DNA(ARDNA)软计算模型框架,并着重论述了设计变量统计特性自适应调整策略和自适应范围 DNA 解码方法及其简化计算公式. ARDNA 克服了传统 DNA 计算中设计变量编码长度和计算精度之间的矛盾,通过有限长的 DNA 编码实现预定的计算精度. 函数优化应用实例说明 ARDNA 对设计变量的取值范围无需先验知识,且具有较强的全局收敛能力.

**关键词:** 自适应范围 DNA 软计算; 编码; 解码

**中图分类号:** TP301 **文献标识码:** A

## Adaptive range deoxyribonucleic acid soft computing model

HUANG Zi-yuan<sup>1,3</sup>, SHI Li<sup>1,2</sup>, FEI Min-rui<sup>1</sup>, HU Chuan-ping<sup>1,3</sup>, WANG He-shou<sup>3</sup>

(1. School of Mechatronical Engineering & Automation, Shanghai University, Shanghai 200072, China;  
2. School of Information Engineering, Zhengzhou University, Zhengzhou Henan 450052, China;  
3. Shanghai Fire Research Institute, The Ministry of Public Security Shanghai 200032, China)

**Abstract:** A framework of an adaptive range deoxyribonucleic acid (ARDNA) soft computing model was proposed. ARDNA distributed design candidates according to the normal distribution of design variables in the present population. The adaptive adjusting strategy for statistic characteristics of design variables, and ARDNA decoding method and its approximately realized technology were given. ARDNA resolved the contradiction between the length of DNA string and computing precision. Finally, it was applied to minimization of the multi-modal function. The simulation results demonstrate that it does not need the prior knowledge for the range of design variables and it has the strong capability of searching globally.

**Key words:** adaptive range DNA soft computing; encoding; decoding

### 1 引言(Introduction)

脱氧核糖核酸(DNA)是生命遗传信息的载体.通过转录作用将遗传信息转录到信使 RNA(mRNA),构成由 3 个连续碱基组成的密码子序列.然后,通过翻译作用,将 mRNA 携带的遗传信息翻译成含有特定氨基酸序列的蛋白质.由于不同的蛋白质表现出不同的性状,从而使得整个世界丰富多彩.生物技术研究结果表明<sup>[1]</sup>:可以采用不同的酶对 DNA 片断进行切除、粘贴和聚合等操作,产生人们所期望的 DNA 串,与电子计算机对“0”,“1”序列所进行的按位加、按位与、按位或等基本操作实现复杂的算术和逻辑运算极其类似.人们不禁要问:基于 DNA 汤(种群)的试管系统是否具有电子计算机强大的计算能力呢?

1994 年, L. Adleman 首次用实验的方法显示了 DNA 用于计算的可能性.在《Science》杂志上发表的文章中<sup>[2]</sup>,他详细介绍了用 DNA 计算解决图论中 Hamiltonian 七节点路径问题的实验方法和结果.这不仅表明了采用 DNA 进行特定目的计算的可行性,而且也说明其可作为具有强大计算能力和潜在并行性的通用计算硬件的可能性.该文的发表,引起了许多学者尤其是计算机科学家的广泛关注.目前研究已涉及许多方面,如 DNA 计算的能力、模型和算法等.最近,许多学者将 DNA 计算与遗传算法、神经网络、模糊系统和混沌系统等软计算方法相结合,逐步形成一个比较活跃的 DNA 生物软计算研究领域.

在人工 DNA 计算模型与进化算法的融合中,例

如 DNA 遗传算法(DNA-GA)、DNA 进化算法(DNA-EA)等,一般预先设定设计变量的取值范围[*min*, *max*],并认为其服从[*min*, *max*]上的均匀分布,从而得到从 DNA 编码串到设计变量的解码关系式.该解码方法存在两个问题:其一,设计变量的最优解未必落在预先设定的取值范围内,而为确保这一点,势必扩大取值范围,导致算法效率下降;其二,取值范围确定的情况下,为提高算法精度,势必增长设计变量的编码串,使得算法计算复杂度增加.本文基于文献[3]的思想,提出自适应范围DNA(简称 ARDNA)解码方法,给出了自适应范围 DNA 计算模型框架.假设设计变量服从正态分布  $N(\mu, \sigma^2)$ ,其中均值  $\mu$  和标准差  $\sigma$  随种群进化过程不断调整,在进化策略的引导下,算法逐步搜索到设计变量最优解所在的区域.与此同时,由正态分布的特征可知,算法精度仅仅与设计变量的 DNA 编码长度和标准差  $\sigma$  有关,从而可以通过有限长的 DNA 编码实现预定的计算精度.由此可见,自适应范围 DNA 软计算不仅解决了传统 DNA 软计算预先设定设计变量取值范围所带来的问题,而且增强了算法的全局搜索能力.本文给出的函数优化应用实例充分说明了这一点.

**2 ARDNA 软计算模型 (ARDNA soft computing model)**

ARDNA 软计算方法,是一种基于 DNA 汤(种群)和生物进化机制的随机搜索算法,其设计变量服从均值和方差随进化过程变化的正态分布,不必预先设定其取值范围,且算法引导种群逐步向优化区域搜索,确保其全局收敛能力.

**2.1 编码(Coding)**

基于生物 DNA 结构,可以建立 DNA 计算模型,用来解决具体的优化和工程设计问题.将设计变量编码成由 A, T, G 和 C 组成的一个固定(或可变)长度 DNA 串,每个设计变量按照其计算精度对应该串的一部分,其中每个组成单元为由 A, T, G 和 C 表示的三联体密码子.

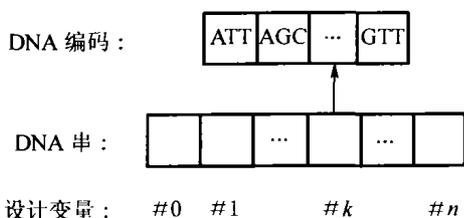


图 1 DNA 编码  
Fig. 1 DNA Coding

**2.2 解码(Decoding)**

首先,将设计变量所对应的 DNA 串中每一三联体密码子按照表 1 翻译成 0 ~ 20 之间的整数.设变量  $X_k$  被编码成  $sl$  个三联体密码子,第  $i$  个翻译成整数  $b_i$ .然后,将该整数序列映射到一个[0,1]的数  $r$ ,满足  $r = \frac{1}{21^{sl} - 1} \sum_{i=1}^{sl} b_i \cdot 21^{i-1}$ . 本文将设计变量取值范围自适应机制引入 DNA 软计算,认为  $X_k$  服从正态分布  $N(\mu, \sigma^2)$ ,从而  $X_k$  被解码为  $X_k = \mu + \sigma \cdot P_r$ . 这里  $P_r$  为标准正态分布  $N(0,1)$  中概率  $r$  所对应的分位数,亦即

$$r = \int_{-\infty}^{P_r} \rho(z) dz. \tag{1}$$

其中  $\rho(z)$  为标准正态分布随机变量的密度函数.在遗传优化机制的引导下,种群逐步进化,种群中的优势群体也逐步趋于最优个体集,优势群体的均值  $\mu$  逐步趋于最优解,而方差  $\sigma^2$  也逐渐缩小.那么,ARDNA 通过种群中优势群体的统计特性不断调整设计变量的统计特性,从而引导算法搜索到最优区域.由正态分布的特性可知,变量取值落在  $[\mu - \sqrt{6}\sigma, \mu + \sqrt{6}\sigma]$  内的概率为 0.999 以上,因此在当前种群统计特性情况下,算法计算精度为  $\frac{2\sqrt{6}\sigma}{21^{sl}}$ ,则适当的编码长度可以实现较高的计算精度.

表 1 DNA 译码表  
Table 1 DNA decoding table

碱基 1	碱基 2				碱基 3
	T	C	A	G	
T	Phe(1)	Ser(3)	Tyr(4)	Cys(5)	T
	Phe(1)	Ser(3)	Tyr(4)	Cys(5)	C
	Leu(2)	Ser(3)	Stop(0)	Stop(0)	A
	Leu(2)	Ser(3)	Stop(0)	Trp(6)	G
C	Leu(2)	Pro(7)	His(8)	Arg(10)	T
	Leu(2)	Pro(7)	His(8)	Arg(10)	C
	Leu(2)	Pro(7)	Gln(9)	Arg(10)	A
	Leu(2)	Pro(7)	Gln(9)	Arg(10)	G
A	Ile(11)	Thr(13)	Asn(14)	Ser(3)	T
	Ile(11)	Thr(13)	Asn(14)	Ser(3)	C
	Met(12)	Thr(13)	Lys(15)	Arg(10)	A
	Met(12)	Thr(13)	Lys(15)	Arg(10)	G
G	Val(16)	Ala(17)	Asp(18)	Gly(20)	T
	Val(16)	Ala(17)	Asp(18)	Gly(20)	C
	Val(16)	Ala(17)	Glu(19)	Gly(20)	A
	Val(16)	Ala(17)	Glu(19)	Gly(20)	G

对于每个  $r$ ,如果都通过解积分方程(1)求得  $P_r$ ,显然计算量太大.现在给出基于标准正态分布表

计算  $P_r$  的简化公式. 记标准正态分布表为  $(r_i, P_{r_i})$ ,  $P_{r_i}$  为标准正态分布  $N(0,1)$  中概率  $r_i$  所对应的分位数, 且  $0.5 \leq r_i \leq 1$ . 若  $r_0 (0.5 \leq r_0 \leq 1)$  在标准正态分布表中, 则可以通过查表得到相应的分位数  $P_{r_0}$ . 否则, 必存在两个与  $r_0$  最接近的, 不妨记为  $r_1$  和  $r_2$ , 且  $r_1 < r_0 < r_2$ .  $P_{r_0}$  可以通过下面的式(2) 或式(3) 近似求得, 其证明过程略. 当  $0 \leq r_0 < 0.5$  时, 其分位数为  $(1 - r_0)$  分位数的相反数.

$$P_{r_0} = P_{r_1} + \frac{(P_{r_2} - P_{r_1}) \cdot (k_1 \cdot \rho(P_{r_1}) - \rho(P_{r_2}))}{k_1 \cdot \rho(P_{r_1}) - \rho(P_{r_2})}, \quad (2)$$

$$P_{r_0} = P_{r_1} + \frac{(1 - k_2) \cdot (P_{r_2} - P_{r_1}) \cdot \rho(P_{r_1})}{\rho(P_{r_1}) - \rho(P_{r_2})}. \quad (3)$$

其中

$$k_1 = \sqrt{\frac{S}{S + r_2 - r_0}}, \quad k_2 = \sqrt{\frac{S + r_2 - r_0}{S + r_2 - r_1}},$$

$$\text{而 } S = \frac{1}{2} \rho^2(P_{r_2}) \cdot \frac{P_{r_2} - P_{r_1}}{\rho(P_{r_1}) - \rho(P_{r_2})}.$$

### 2.3 ARDNA 算法结构 (Structure of ARDNA algorithm)

A) 初始化 DNA 汤 (种群). 根据实际问题, 以及各个设计变量的计算精度确定其 DNA 片断长度, 并按上述编码方法编码所有设计变量, 构成 DNA 串, 并以某种方式 (如完全随机和借助先验知识) 产生由  $N$  个 DNA 串初始化 DAN 汤 (种群). 另外, 初始化每一设计变量的分布参数  $\mu$  和  $\sigma_2$ .

B) 评价. 按照上述解码方法将每个设计变量解码, 并以某评价函数计算种群中每条 DNA 串的适应度.

C) 遗传操作. 对 DNA 汤 (种群) 施以选择、交叉、变异和倒位等遗传操作, 具体操作方法参见文献[4].

D)  $\mu, \sigma^2$  自适应调整. 为了避免设计变量与当前种群的统计特性不一致, 种群进化  $M$  代, 调整一次  $\mu$  和  $\sigma$ . 基于种群进化速度和先验知识, 本文给出如下  $\mu$  和  $\sigma$  自适应调整策略, 以增强算法鲁棒性和智能度.

$$\mu_{\text{new}} = \mu_{\text{present}} + \frac{\bar{f}_{\text{present}} \cdot (\mu_{\text{sampling}} - \mu_{\text{present}})}{\bar{f}_{\text{present}} + \bar{f}_{\text{old}}}, \quad (4)$$

$$\sigma = \sigma_{\text{present}} + 2 \cdot \frac{(\bar{f}_{\text{old}} + \bar{f}_{\text{present}}) \cdot (\sigma_{\text{sampling}} - \sigma_{\text{present}})}{\bar{f}_{\text{sampling}} + \bar{f}_{\text{present}}}, \quad (5)$$

$$\sigma_{\text{new}} = \begin{cases} \sigma, & \text{if } \sigma > \sigma_\epsilon, \\ \sigma_\epsilon, & \text{else.} \end{cases} \quad (6)$$

式中  $\bar{f}_{\text{old}}, \bar{f}_{\text{sampling}}, \bar{f}_{\text{present}}$  分别为上一次调整  $\mu$  和  $\sigma$  时, 对当前种群采样和当前种群的平均适应度,  $\sigma_\epsilon$  为预先设定的最小的标准差. 引入  $\sigma_\epsilon$  以防止搜索区域变得太小, 避免出现早熟现象.

ARDNA 算法流程图见图 2.

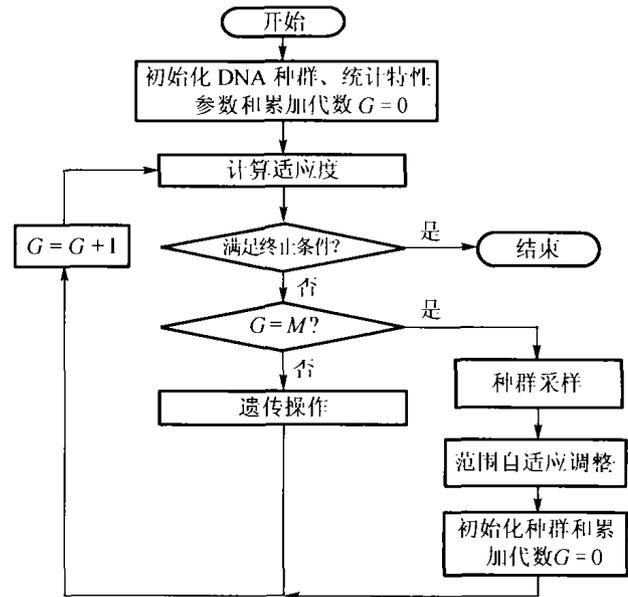


图 2 ARDNA 算法流程图  
Fig. 2 Flow chart of ARDNA

### 3 应用实例 (Application example)

为了说明 ARDNA 的有效性, 本文进行了数字仿真, 将其应用到多模态函数 (multi-modal function, 简称 MMF) 优化问题中. 最优化问题为

$$\min_{x_i \in [-100, 100]} MMF = \sum_{i=1}^n (x_i^2 + 5(1 - \cos(x_i \cdot \pi))).$$

上述为  $n$  维的多模态函数优化问题, 在此仅解决一维情况, 其函数曲线见图 3.

由图 3 可知, 一维多模态函数有多个局部最优解和唯一的全局最优解  $x = 0$ . ARDNA 算法主要运行参数为: 设计变量初始均值和方差分别为  $\mu = 30$  和  $\sigma^2 = 25$ , 每  $M = 50$  代自适应调整一次设计变量的均值和方差, 最多调整 5 次, 设计变量编码长度为 12 位字符串, 交叉概率和变异概率均为 0.3, 种群采样率为 0.6, 采用杰出者 (当前种群中最好个体) 保留策略. 为便于比较, 给出 DNA 软计算的优化结果, 其设计变量初始均值  $\mu$  和方差  $\sigma^2$  以及其它运行参数与 ARDNA 方法相同, 唯一不同的是运行过程中没有自适应调整设计变量的均值和方差. 到目前为止搜索到的最好解 (记为 Elitist) 的进化结果见图 4,

图5给出了最好解的多模态函数值的进化结果.从图4中可以看出,由于预先假定设计变量服从的正态分布的中心距离最优解太远,使得DNA软计算方法搜索的仅为局部最优解5.05.而ARDNA可以搜索到全局最优解,其原因就在于设计变量的均值和方差随种群进化过程自适应调整,使得算法逐步搜索到全局最优解所在区域,经过250代进化,可得近似全局最优解为0.000047(理论上全局最优解为0,故误差小于0.005%).若进化代数增加,则可进一步逼近全局最优解.

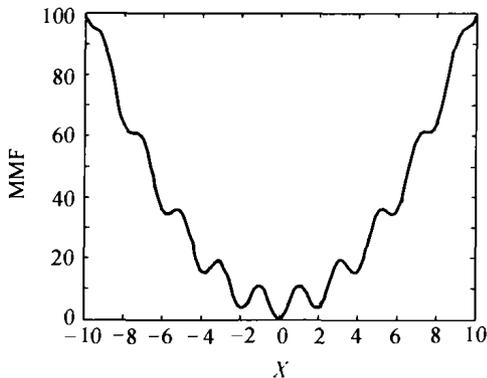


图3 一维多模态函数曲线  
Fig. 3 Curve of multi-modal function

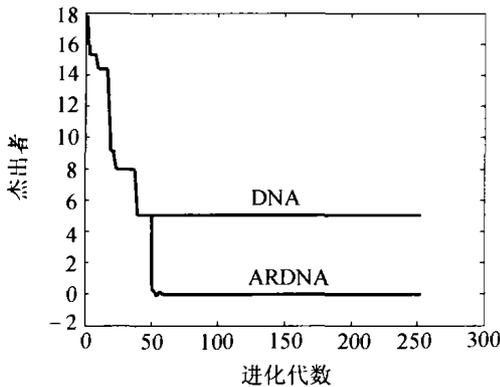


图4 最好解进化结果  
Fig. 4 Evolution results of elitist

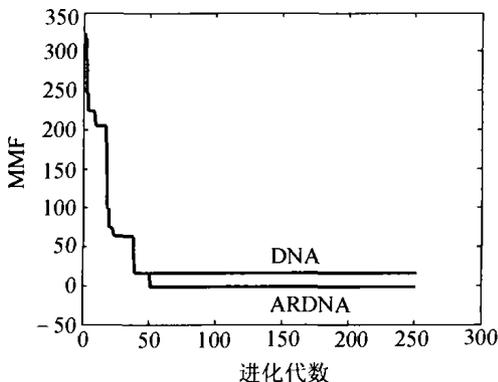


图5 多模态函数值进化结果  
Fig. 5 Evolution results of value of MMF

## 4 结论(Conclusion)

本文将设计变量的取值范围自适应技术引入到DNA软计算,给出ARDNA计算模型框架.该模型的优势在于:其一是对设计变量的取值范围无需先验知识,随着种群进化过程算法逐步逼近最优解所在的区域;其二是解决了DNA编码长度与算法精度之间的矛盾,有限的编码长度实现较高的计算精度;其三是其具有较强的全局搜索能力.由此可见,ARDNA具有很强的通用性和适应能力,其有望应用于流线型机翼形状优化设计、最优控制、非线性建模等领域.

## 参考文献(References):

- [1] GARZON M H, DEATON R J. Biomolecular computing and programming [J]. *IEEE Trans on Evolutionary Computation*, 1999, 3(3): 236 - 250.
- [2] ADLEMAN L M. Molecular computation of solutions to combinatorial problems [J]. *Science*, 1994, 266(5187): 1021 - 1023.
- [3] OYAMA Akira, OBAYASHI Shigeru, NAKHASHI Kazuhiro. Real-coded adaptive range genetic algorithm and its application to aerodynamic design [J]. *Int J of Japan Society of Mechanical Engineers (JSME)*, Series A, 2000, 43(2): 124 - 129.
- [4] 丁永生, 任立红, 邵世煌. DNA计算与软计算[J]. 系统仿真学报, 2001, 13(增刊): 198 - 201.  
(DING Yongsheng, REN Lihong, SHAO Shihuang. DNA computing and soft computing [J]. *J of System Simulation*, 2001, 13(Suppl.): 198 - 201.

## 作者简介:

黄自元 (1973—),男,博士,公安部上海消防研究所助理研究员,毕业于上海大学机电工程与自动化学院控制理论与控制工程专业,在国内外重要会议和核心期刊上已发表论文12篇,目前主要研究兴趣是模糊逻辑、进化算法、神经网络和人工DNA等软计算技术及其在函数优化、自动控制 and 火灾识别与预测等领域中的应用, E-mail: hzy666@21cn.com;

师黎 (1964—),女,上海大学在职博士生,郑州大学电气学院副教授,负责承担和完成省部级等项目10余项,发表论著20余篇,主要研究兴趣是智能、自适应控制理论,目前主要进行模糊神经网络自适应容错控制方面的研究;

费敏锐 (1961—),男,博士,教授,博士生导师,负责承担和完成国家和省部级等项目20余项,已发表论著80余篇,目前主要研究兴趣是智能、适应控制理论及学习方法,分散、现场控制系统及其相关技术和应用、虚拟现实及计算机仿真技术, E-mail: mrfei@staff.shu.edu.cn;

胡传平 (1960—),男,博士,公安部上海消防研究所研究员,毕业于同济大学系统工程专业,发表论著50余篇,主要从事消防科学研究及公共安全管理, E-mail: cphu@china.fire.com.cn;

王鹤寿 (1967—),男,博士,公安部上海消防研究所副研究员,毕业于上海大学机械设计及理论专业,发表论著20余篇,主要从事消防科学及工程应用研究, E-mail: wahesh538@vip.sina.com.