

模拟自然树生长的竞争算法及在曲线拟合中的应用

郭改文, 黄卡玛

(四川大学电子信息学院, 四川成都 610065)

摘要: 利用自然树生长过程中生长、凋落矛盾统一的原理, 建立了自然树生长的竞争模型, 提出了模拟自然树生长的竞争算法. 为验证算法的合理性和有效性, 将其应用于复杂曲线的拟合, 与标准遗传算法进行对比, 该算法具有运行速度快、内存占用率低、拟合精确度高. 与经典的最小二乘法进行对比, 该算法内存占有率低且具有抗噪特性. 该算法为优化设计和计算提供了一种新的思路.

关键词: 自然树; 生长; 竞争; 优化; 曲线拟合

中图分类号: TN15; TP202. 7 **文献标识码:** A **文章编号:** 0372 2112 (2008) 09 1839 05

A Competition Algorithm of Simulated Growing Natural Tree and Its Application in Curve Fitting

GUO Garwen, HUANG Karma

(School of Electronics & Information Engineering, Sichuan University, Chengdu, Sichuan 610065, China)

Abstract: According to the principle of unity of opposite on natural tree growing and wilting, a competition model is developed. Then, a competition algorithm of simulated growing natural tree is presented. To verify the algorithm, it is applied to the curves fitting. Compared with the standard genetic algorithm, it runs faster with less memory and has higher precision; compared with the classical least square method, it requires less memory with noiseproof feature. This algorithm may provide a novel solution to the design and calculation of optimization.

Key words: natural tree; growth; competition; optimization; curve fitting

1 引言

最优化算法的研究长期以来是应用数学、计算机科学、管理科学等学科研究的重要课题之一. 自20世纪60年代以来, 以美国密西根大学教授 Holland 创立的遗传算法(简称 GA) 为代表^[1] 的模拟生物进化算法, 使现代最优化算法在近几十年里得到了蓬勃的发展. GA 与传统的解析法相比, 其优点是: 对目标函数和约束条件没有特殊要求; 有更多的机会求得全局最优解; 具有通用性且便于并行处理. 然而 GA 存在着明显的固有缺点: 局部搜索能力差; 收敛速度慢且存在未成熟现象; 采用全局随机搜索技术, 可以避免陷入局部极小值, 但是将花费大量的运算时间. 因而近年来遗传算法的应用研究主要是结合其他算法对其进行改进^[2]. 1983年, Kirkpatrick 等人利用固体物质退火过程中的热平衡问题与随机寻优问题的相似性提出模拟退火算法来达到全局最优, 其具有较好的局部搜索能力, 但存在着解的质量与求解时间长之间的矛盾且对参数的依赖性很强^[3].

1992年, Colomi 提出模拟蚂蚁集体寻径的仿生类蚁群算法, 其采用分布式并行机制, 具有较强的鲁棒性, 但搜索时间长、易陷入局部最优解^[4]. 自1997年以来, 人工免疫系统成为人工智能领域的研究热点, 但由于人们对生物免疫机理还不十分深入, 有关免疫算法的研究局限于利用免疫机理改进其他算法, 主要集中在遗传算法上^[5].

鉴于上述算法存在着不同程度的不足之处, 我们考虑到相对生物进化的漫长过程, 自然树的生长时间很短, 其生长过程本身又是生长、凋落矛盾统一的过程. 本文提出一种完全不同于模拟生物进化过程的新思想——模拟自然树生长的竞争算法(下文简称 GCA). 其基本思想是: 自然树生长的动力来源于获得充足的阳光进行光合作用; 其枝条的生长和凋落都受到阳光和营养的主要调控. 自然树通过枝条的生长和凋落, 其最后的形体必然是与生长环境相适应的最优结果.

国内外有限的关于树的建模研究工作, 主要是基于如 A. Lindenmayer 和 P. Prusinkiewicz 等人的模拟植物生

长的分枝模型,用于分形领域及计算机图形学.文献[6]提出模拟植物生长算法求解整数规划问题,是基于植物向光性的概率生长模型.本文提出的GCA,完全基于自然界中树的生长过程,体现了自然树生长过程中的辩证关系,其基本思想与已有的研究不同.

GCA机理比较简单,容易被人们理解和接受.本文首先阐述了自然树生长的生物学原理,抓住主要矛盾,将自然树的复杂生长过程抽象以建立自然树生长的竞争模型,并通过数学定义描述自然树生长的竞争算法.最后以两条有代表性的单调升函数、复杂振荡函数曲线拟合的结果说明了该算法的合理性和有效性,并将结果与通用性较强的标准GA^[7]和经典的最小二乘法(简称LS)^[8]进行对比,结果表明该算法在曲线拟合中显示出其优越性.

2 自然树生长的生物学原理

从生物学原理看,自然树生长的内在动力是要获得充足的阳光进行光合作用,延续种群的发展.自然树的生长受光照、温度、降水、气体成分、土壤物理化学性质、季节、气候、昆虫、海拔高度及其他动植物等环境因素的影响^[9],其过程是非常复杂的.本文抓住主要矛盾,将自然树的复杂生长过程抽象出来,只考虑树干和树枝的生长与凋落,而不考虑树的根和叶的生长,也不考虑树的生殖过程.树枝的生长和凋落由阳光和营养因子共同确定:树在生长过程中为了获得足够的阳光进行光和作用,树枝必须向阳光迅速生长,这就使得树体顶端的枝条具有天然的竞争优势,优先生长,遮挡先前生长的枝叶,被遮挡的枝叶因得不到足够的阳光渐渐凋落;但同时,树枝的生长也需要从土壤中获取营养物质,离根部越远的树枝因得不到充足的土壤营养物质而减缓生长,甚至可能凋落.因此,树的生长过程是适应环境的最优化过程,树的最终形态是适应环境的最佳结果.

3 模拟自然树生长的竞争模型与算法

从系统动力学^[10]的观点看,最优化过程相当于在问题的可行解空间内从一个初态向着吸引源(最优解)不断演化的过程.这里我们提出模拟自然树的生长,根据自然树生长的内在动力并考虑阳光和营养因素的作用,建立枝条生长和凋落的动力机制,使树枝在可行域中向最优状态不断生长和演化,最终获得问题的最优解.

3.1 模拟自然树生长的竞争模型

种子从根基点发芽成为树干 a_0 , a_0 上又产生树枝,然后在树干及树枝上产生分枝 $a_i(i=1,2,\dots)$.因此,树的生长过程是一个反复迭代的过程.新枝条的发芽点可以是旧枝条上的任意点; $\lambda \in (0,1)$ 表示枝条向光生长

的速率.由于树的生长要获取最大阳光进行光合作用,我们定义整棵树的阳光适应度如下:

定义1 阳光适应度

$$\eta = \exp(-c_0|\phi - f|) \tag{1}$$

式中, ϕ 为虚拟的阳光函数,它与解决的问题有关,对于最优化问题而言,它就是吸引源; f 为与树的形态有关的函数,对于最优化而言,它就是需要优化的对象; c_0 为调整阳光适应度的正常数.显然,分枝的产生和枝条的生长应该使 $f \rightarrow \phi$,即 $\eta \rightarrow 1$.

树的生长需要的营养物质是从土壤输送到每一根枝条的,由于受到重力的作用,营养物质的输送距离是有限制的.因此,我们定义每一根枝条的营养因子为枝条顶端距根部总长度的归一化值,即:

定义2 枝条 a_i 的营养因子

$$\alpha_i = \exp\left(-\sum_{r=0}^m |l_r|\right) \tag{2}$$

式中, m 表示营养从根部传送到枝条 a_i 的顶端所经过的总枝条数目; $|l_r|$ 为第 r 根枝条的长度.

在营养充足的条件下,树枝的生长还受到遮挡因子的影响,而遮挡因子与树枝向光竞争生长过程中阳光适应度的变化有关,描述如下:

定义3 枝条 a_i 的遮挡因子

$$\beta_i = \eta - \eta_{a_i} \tag{3}$$

式中, η_{a_i} 表示枝条 a_i 向光竞争生长结束时的阳光适应度; η 表示当前树体上所有枝条向光竞争生长结束时的阳光适应度.特别地,新生枝条的遮挡因子为0.而每一年中枝条 a_i 的向光竞争生长过程可用式(4)描述为:

$$l_i^t = l_i^{t-1} + Dw^t(\lambda)^t \tag{4}$$

式中, $t=1,2,\dots,T$, T 为每年树枝向光探测生长的次数; w^t 为生长速率权重,由初值 w^{int} 、终值 w^{end} 及探测的次数确定,即: $w^t = (w^{\text{int}} - w^{\text{end}})(T - t)/T + w^{\text{end}}$; D 初值为1,在探测生长过程中由式(5)决定:

$$D = \begin{cases} l_i/|l_i|, & \eta_{a_i} > \eta_{a_i^{t-1}} \\ -l_i/|l_i|, & \eta_{a_i} \leq \eta_{a_i^{t-1}} \end{cases} \tag{5}$$

在本次探测生长的过程中,只要 $\eta_{a_i} > \eta_{a_i^{t-1}}$, a_i 就生长一个增量,否则反方向探测.显然,向光竞争生长始终使整棵树的阳光适应度增大.

考虑到树枝的生长还与营养和被其他枝条遮挡有关,在探测生长结束后,如果满足式(6)中的凋落条件,则树枝 a_i 进行凋落:

$$l_i = \begin{cases} 0, & a_i < \varepsilon_1 \text{ 或 } \beta_i > \varepsilon_2 \\ l_i^t, & \text{其他} \end{cases} \tag{6}$$

式中, ε_1 、 ε_2 为(0,1)之间的正常数,他们分别表示枝条营养的最低阈值和遮挡的最大阈值.虽然枝条的凋落可

能使当年整棵树的阳光适应度降低,但从整个生长过程看,这有利于整棵树向阳光适应度最好的方向生长.

3.2 模拟自然树生长的竞争算法的流程

基于模拟自然树生长的竞争模型, GCA 在每一年(代)的生长过程中分为以下四步进行,依次是:分枝产生、向光竞争生长、计算枝条营养因子和遮挡因子、有条件的凋落. 设 k 为当前代, K 为最大生长代数, E 为设定的最佳阳光适应度, 分枝概率为 p , 其流程如图 1 所示.

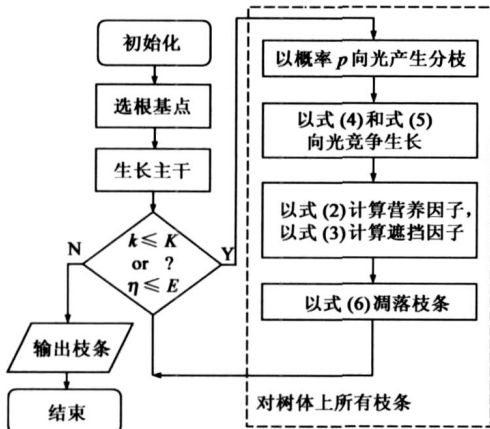


图 1 GCA 流程框图

为了验证 GCA 的合理性和有效性, 本文首先将其应用于解决曲线拟合问题.

4 GCA 在曲线拟合中的应用

在科学实验及统计分析的研究工作中, 我们经常需要从测量得到的数据出发, 寻找一个合理的函数能从整体上反映数据的基本变化趋势. 假设给定一组数据对 $(x_j, y_j) (j=1, 2, \dots, N, y_j = \phi(x_j))$, 要求在函数类 F 中寻找一个“最好”的函数 $f(x)$, 满足下式:

$$\min \sum_{j=1}^N |\phi(x_j) - f(x_j)| \quad (7)$$

我们知道, 在完备函数空间中 $f(x)$ 可以用一组基函数展开, 即:

$$f(x) = \sum_{i=0}^M g_i f_i(x) \quad (8)$$

其中 g_i 是待定的系数, $f_i(x)$ 是基函数. 显然, 上述问题可以转化为式(9), 即求出系数 g_i , 使得 η 取得最大值:

$$\eta = \frac{1}{N} \sum_{j=1}^N \exp \left[-c_0 \left| \phi(x_j) - \sum_{i=0}^M g_i f_i(x_j) \right| \right] \quad (9)$$

这里, 系数 g_i 的求解就可以采用上述的模拟自然树生长的竞争算法 GCA. 本文以两条典型曲线的拟合来说明.

以下问题的定义域为 $[0, 1]$, 且 $N = 100$, 我们在 Intel CPU 2.26GHz、内存 1G Bytes 的机器上编程实现. 本文还将结果与标准 GA 和经典 LS 的拟合结果进行对比.

并比较了三者的耗时、占用内存情况及拟合精度(精度^a为无噪声拟合, 精度^b为加入 5% 的高斯噪声拟合). 这里的标准 GA 程序采用广泛使用的由 Carroll 编写的 GA170. f; 经典 LS 程序使用文献[8]的 D9r4. f90; GCA 的参数 p 、 ϵ_1 、 ϵ_2 是根据经验值选取的.

例 1 单调升函数 $y = x \sin x \cos x + 3$

三种算法取相同的基函数 $f_i(x) = x^i$, 即采用多项式拟合.

首先利用本文算法进行拟合, 相关的参数取值为: $c_0 = 1, K = 100, T = 5, \lambda = 0.6, p = 0.15, \epsilon_1 = 0.01, \epsilon_2 = 0.9, w^m = 0.9, w^{nd} = 0.4$. 我们随机运行 50 次取平均值, 平均拟合精确度为 98.89%, 平均枝条数目 $M = 15$, 即有 15 个系数. 对标准 GA, 同样取 15 个系数进行拟合, 相关的参数取值为: 最大代数 = 200, 每代个体数 = 50, 交叉率 = 0.5, 变异率 = 0.02. 三种算法曲线拟合结果对比如图 2 所示, 运行时间、占用内存和拟合精度对比如表 1 所示.

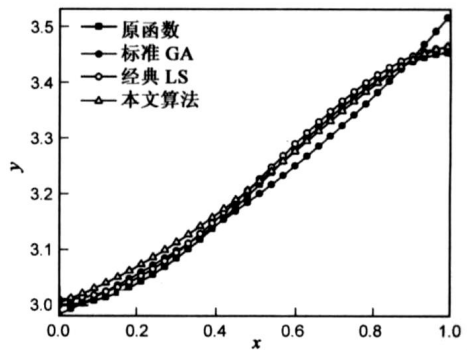


图 2 单调升函数三种算法拟合效果对比

表 1 单调升函数三种算法耗时和占用内存对比

算法名称	时间 (s)	内存 (Bytes)	精度 ^a	精度 ^b
标准 GA	3.20	12000	98.15%	95.11%
经典 LS	0.28	36480	99.99%	95.91%
本文算法	1.50	2800	99.12%	98.16%

从图 2 可以看出: 本例中, 相对于原函数, 三种算法拟合的结果都能从整体上反映数据的基本变化趋势. 但在区间 $[0.8, 1]$ 范围内, 标准 GA 拟合的曲线走势明显与原函数曲线不符, 而本文算法拟合的曲线与原函数曲线和经典 LS 拟合的曲线相吻合. 从表 1 可以看出: 本文算法相对于标准 GA 拟合精度高, 运算时间短, 占用内存少; 本文算法相对于经典 LS, 虽然在运算时间和精确度上稍有逊色, 但占有内存少. 从表 1 我们还可以看出: 在加入 5% 的高斯噪声干扰时, 三种算法的拟合精度都有所下降, 但本文算法相对其他两种算法下降程度较小, 在拟合精度上甚至超过了经典 LS.

例 2 复杂振荡函数 $y = x \sin 10\pi x + 2$

三种算法取相同的基函数 $f_i(x) = \cos(i \arccos x)$.

即采用切比雪夫正交基函数拟合。

首先利用本文算法进行拟合, 相关的参数取值为: $c_0=1, K=100, T=3, \lambda=0.36, p=0.15, \varepsilon_1=0.01, \varepsilon_2=0.9, w^{\text{int}}=0.9, w^{\text{end}}=0.4$ 。我们随机运行 50 次取平均值, 平均拟合精确度为 97.47%, 平均枝条数目 $M=39$, 即有 39 个系数。对标准 GA, 同样取 39 个系数进行拟合, 相关的参数取值为: 最大代数=200, 每代个体数=50, 交叉率=0.5, 变异率=0.02。三种算法曲线拟合结果对比如图 3 所示, 运行时间、占用内存和拟合精度对比如表 2 所示。

表 2 复杂振荡函数三种算法耗时和占用内存对比

算法名称	时间(s)	内存(Byte)	精度 ^a	精度 ^b
标准 GA	13.73	31200	81.27%	81.06%
经典 LS	0.69	117120	99.98%	96.34%
本文算法	7.82	7784	97.44%	96.78%

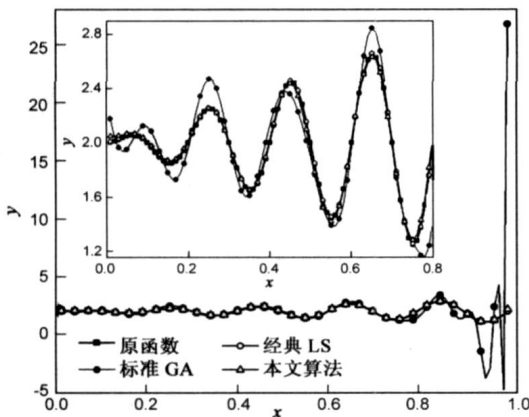


图 3 复杂振荡函数三种算法拟合效果对比

本例是一条比较复杂的振荡函数曲线。从图 3 和表 2, 我们可以看出: 与标准 GA 相比, 本文算法除了具有运算时间短, 占用内存少的优点之外, 平均拟合精度仍保持在 97% 以上, 拟合效果仍然比较好, 基本与原函数曲线和经典 LS 拟合的曲线相吻合。而标准 GA 在区间 $[0, 0.8]$ 范围内, 只是拟合形状大概相似, 但在区间 $[0, 0.1]$ 和 $[0.8, 1]$ 范围内拟合形状相差很大, 可以说, 标准 GA 没有从整体上反映数据的基本变化趋势。从表 2 我们还可以看出: 在加入 5% 的高斯噪声干扰时, 三种算法的拟合精度都有所下降, 但本文算法相对其他两种算法下降程度较小, 在拟合精度上也超过了经典 LS。

综合上述两例函数曲线拟合的结果, 我们可以看出: 与标准 GA 相比, 本文算法拟合的精确度高, 拟合的效果好, 运行耗时短, 占用内存少, 且对复杂振荡函数曲线仍然保持较好的拟合效果。同时, 本文算法与经典 LS 相比: 避免了解大规模病态方程组; 不需人为的确定拟合曲线的次数; 当加入高斯噪声干扰时, 拟合精度高于经典 LS。

5 结论

本文提出一种模拟自然树生长的竞争优化算法 GCA, 并将其应用在曲线拟合中。通过与标准 GA 和经典 LS 对比两条有代表性的函数曲线拟合的结果, 显示出该算法在解决大规模复杂优化问题时的优越性, 也充分说明了该算法具有深入研究的价值。分析其主要原因: 一方面, GA 模拟生命的进化过程, 个体基因编码的需要, 开始就必须确定问题域的维数, 使问题复杂化; 另一方面, GA 采用全局随机搜索策略, 局部搜索能力差, 使其收敛速度慢, 运算时间长。而本文算法模拟自然树的生长过程, 开始不需要确定问题域的维数, 枝条(问题域)依据阳光适应度(目标函数值)动态的分枝和生长, 并根据营养因子和遮挡因子有条件的凋落, 使问题的解决由简单到复杂; 且采用竞争生长策略, 使其局部搜索能力强, 收敛速度快, 运算时间短。

作为一种全新的优化算法还需要做更深入的研究工作, 本文的曲线拟合只是其应用的一个方面。由于树的结构很容易拓展到线天线, 我们下一步的研究工作是将其应用于天线的优化设计和计算, 然后推广应用到解决其他复杂工程实践问题。我们可以展望本算法的开发应用将会有助于解决大量的管理学、运筹学、工程实践、科学研究中的大规模最优化问题。

参考文献:

- [1] Holland J H. Adaptation in Nature and Artificial System [M]. Ann Arbor: The University of Michigan Press, 1975.
- [2] 张晓伟, 刘三阳. 一种新的区间遗传算法[J]. 电子学报, 2007, 35(8): 1567-1571.
ZHANG Xiao wei, LIU Sar yang. A novel interval genetic algorithm[J]. Acta Electronica Sinica, 2007, 35(8): 1567-1571. (in Chinese)
- [3] Alfonzetti S, Diletto E, Salerno N. Simulated annealing with restarts for the optimization of electromagnetic devices [J]. IEEE Trans Magnetics, 2006, 42(4): 1115-1118.
- [4] HO S L, YANG S Y, NI G Z, et al. A modified ant colony optimization algorithm modeled on Tabu Search methods [J]. IEEE Trans Magnetics, 2006, 42(4): 1195-1198.
- [5] 焦李成, 杜海峰. 人工免疫系统进展与展望[J]. 电子学报, 2003, 31(10): 1540-1548.
JIAO Li cheng, DU Hai feng. Development and prospect of the artificial immune system [J]. Acta Electronica Sinica, 2003, 31(10): 1540-1548. (in Chinese)
- [6] LI Tong. A global optimization bionics algorithm for solving integer programming——plant growth simulation algorithm [A]. Proceedings of 2004 International Conference on Management Science and Engineering [C]. Harbin: Harbin Institute of Techn

nology Press, 2004. 328– 386.

- [7] Vose M D. The Simple Genetic Algorithm: Foundations and Theory[M]. Massachusetts: The MIT Press, 1999.
- [8] 何光渝, 高永利. Visual Fortran —— 常用数值算法集[M]. 北京: 科学出版社, 2002. 384– 405.

- [9] 李江风, 袁玉江等. 树木年轮水文学研究与应用[M]. 北京: 科学出版社, 2000. 25– 47.
- [10] 陈忠, 盛毅华. 现代系统科学[M]. 上海: 上海科学技术文献出版社, 2005. 303– 430.

作者简介:



郭改文 女, 1971 年出生于河南南阳, 现为河南教育学院信息技术系副教授, 四川大学电子信息学院在读博士研究生. 主要研究方向为: 智能计算、并行计算等.

E mail: guogaowen710825@163.com



黄卡玛 男, 1964 年出生于重庆, 教授、博士生导师, 1991 年电子科技大学获博士学位, 2001 年美国 Clenslon 大学博士后. 现为中国电子学会理事、会士, IEEE 高级会员. 近年来主要从事电磁场理论、微波化学等方面的研究工作.

E mail: kmhuang@scu.edu.cn