

简化*E.Coli*觅食优化算法及其在非线性模型 参数辨识中的应用

刘益剑^{1,2}, 方彦军¹, 孙 冀², 张建明³

(1. 武汉大学 自动化系, 湖北 武汉 430072; 2. 南京师范大学 检测技术与自动化装置研究所, 江苏 南京 210042;
3. 浙江大学 先进控制研究所 工业控制技术国家重点实验室, 浙江 杭州 310027)

摘要: 针对非线性系统模型参数辨识问题, 提出了一种简化*E.Coli*觅食优化算法, 该算法是基于大肠杆菌(*E.Coli*)的化学趋向性行为, 包含翻转算子和游动算子, 同时还采用个体历史最佳位置和群体最佳位置对*E.Coli*群体进行更新(优值跟踪算子). 以重油热解三集总非线性模型参数估计验证了简化*E.Coli*觅食优化算法的效果, 实验结果表明了简化*E.Coli*觅食优化算法的有效性, 为非线性系统模型参数估计提供了一种新方法.

关键词: 简化*E.Coli*; 觅食算法; 参数辨识

中图分类号: TP18 **文献标识码:** A

Simplified *E.Coli* foraging optimization algorithm and its application to parameter identification of nonlinear system model

LIU Yi-jian^{1,2}, FANG Yan-jun¹, SUN Ji², ZHANG Jian-ming³

(1. Department of Automation, Wuhan University, Wuhan Hubei 430072, China;
2. Institute of Measurement Technology and Automatic Device, Nanjing Normal University, Nanjing Jiangsu 210042, China;
3. National Laboratory of Industrial Control Technology, Institute of Advanced Process Control Zhejiang University, Hangzhou Zhejiang 310027, China)

Abstract: A simplified *E.Coli* foraging optimization algorithm based on the chemotactic behavior of *E.Coli* is presented in this paper for the parameter estimation problem of nonlinear system model (NSM). The simplified *E.Coli* algorithm includes a tumbling operator, a swimming operator, the optimal position of individual *E.Coli* and the location of all *E.Coli* swarm are adopted to update the locations of swarm (tracing operator of optimal values). The effectiveness of the proposed simplified *E.Coli* foraging optimization algorithm is demonstrated by simulation experiments on the parameter estimation of the NSM for heavy oil thermal cracking. The results show that the simplified *E.Coli* model is valid and provides an attractive method to the estimation of parameters of NSM.

Key words: simplified *E.Coli*; foraging algorithm; parameter optimization

1 引言(Introduction)

在自动控制领域中, 系统模型的建立通常是控制方案设计的前提条件之一, 但是由于被控过程或者对象的复杂特性, 使得系统模型尤其是非线性系统的模型建立变得困难, 这也促使众多学者进行了非线性系统辨识的有益研究. 传统的系统辨识方法包括参数辨识方法有最小二乘法(least squares methods)、极大似然法(maximum likelihood methods), 并且在许多参数辨识问题中得到了广泛应用^[1~3]. 然而由于上述辨识方法都是基于梯度法原理的, 所以如果目标函数不可导或辨识参数具有非线性,

则LS和ML的应用就会受到局限. 而进化算法是人们模拟自然界中生物的优化现象, 抽象出来的解决优化问题的数学方法, 随着多年不断的研究, 已经有了很多成果, 包括研究比较早的遗传算法(genetic algorithm)、蚁群算法(ant algorithm)、鱼群算法(fish school algorithm), 以及近年来研究比较热的微粒群算法(particle swarm optimization algorithm)、免疫算法(immune algorithm)等. 由于系统辨识问题可以划归为优化问题, 因此上述的进化算法也可以用来解决系统辨识^[4,5].

在研究宏观生物智能优化行为的同时, 学者们也

将视角转向了微生物. 在研究*E.Coli*(大肠杆菌)的化学趋向性的现象基础上, 提出的*E.Coli*觅食优化算法是一种新型的进化算法^[6], 也是基于群体的概念, 可以用来处理优化问题. 本文在文献[6]研究的基础上, 将*E.Coli*觅食优化算法进行简化, 在保留实现化学趋向性行为的翻转、游动等算子的基础上, 引入*E.Coli*群体中的历史最佳位置, 作为*E.Coli*的更新算子, 加强*E.Coli*个体之间的信息交流, 提出了一种新的简化*E.Coli*觅食优化算法, 并将此优化算法应用于非线性系统的参数辨识.

2 问题描述(Description of problem)

通常工业过程对象模型特性部分已知, 因此模型建立问题就转化为灰箱建模问题: 根据物理规律或者经验公式, 假设系统模型类已知, 利用输入输出数据, 在一定的准则下确定模型公式的参数, 图1描述了这种过程.



图1 灰箱模型示意

Fig. 1 Illustration of grey box model

图中: $u(t)$, $y(t)$ 表示系统输入、输出数据; $f(\theta, t)$ 表示系统模型公式, $\theta = [\theta_1, \theta_2, \dots, \theta_D]$ 表示要辨识得到的参数; $\xi(t)$ 表示干扰噪声.

参数 θ 辨识需要满足一定的准则, 文中采用如下的函数作为辨识准则, 同时也作为简化*E.Coli*觅食优化算法中*E.Coli*个体位置好坏的适应度评价函数:

$$f_{\text{fitness}} = \sum_{i=1}^T [y(t) - \tilde{y}(t)]^2. \quad (1)$$

式(1)中: $y(t)$ 表示辨识模型输出, $\tilde{y}(t)$ 表示实际输出.

3 简化*E.Coli*觅食优化算法(Simplified *E.Coli* foraging optimization algorithm)

3.1 *E.Coli*觅食过程(Foraging process of *E.Coli*)

大肠杆菌(*E.Coli*)是寄生在动物体内的微生物, 通常其个体大小约为直径 $1\mu\text{m}$, 长 $2\mu\text{m}$. 在*E.Coli*生命过程中, 表现出一种称之为“化学趋向性”的行为, 即*E.Coli*总是向着富含有机营养的区域游动, 而同时远离有害物质的区域.

假设在觅食时间 T 内*E.Coli*获得了 E 大小的能量, 那么*E.Coli*的化学趋向性就可以用 E/T 表示, 即觅食效率. *E.Coli*觅食与优化问题可以转换为极大化 E/T , 即在整个觅食区域中单位时间内获得尽可能多的能量的优化过程.

另外, *E.Coli*也往往基于群体进行觅食. 在*E.Coli*群体中, 个体在保持自身觅食行为的同时, 也与其他*E.Coli*个体进行信息的传递, 同时*E.Coli*个体之间也会因为争抢食物引发竞争. 除此之外, *E.Coli*觅食过程中还存在分裂、消亡或者迁移等复杂行为.

3.2 简化*E.Coli*觅食优化算子(operators of simplified *E.Coli* foraging optimization algorithm)

化学趋向性是*E.Coli*觅食过程中获得高的觅食效率的重要环节, 简化*E.Coli*觅食优化算法就是模拟了这个优化行为, 而忽略了其他觅食行为. *E.Coli*的化学趋向性是不断游动和改变方向的行为, 简化*E.Coli*觅食优化算法采用3个算子.

① 游动算子(Swim).

*E.Coli*在化学趋向性中位置的改变称为游动. 通过游动算子的作用使得*E.Coli*在觅食优化空间中进行位置的更新. 在游动过程中设定游动的步数为 N_s , 决定了*E.Coli*个体在优化解方向上游动的次数.

② 翻转算子(Tumble).

当评价到当前位置得到的适应度函数比上一步差时, 就要进行游动方向的改变, 在简化*E.Coli*模型中称之为翻转算子. 通过翻转算子实际上是保证简化*E.Coli*觅食优化算法具有随机性, 从而能够脱离局部最小值, 保证了简化*E.Coli*觅食优化算法具有全局最优性能.

③ 优值跟踪算子(Tracing of optimal value).

为了加速优化过程, 采用*E.Coli*群体中的历史最佳信息来更新*E.Coli*群体的位置, 具体介绍见下节简化*E.Coli*模型原理中的介绍.

为了下面描述方便, “化学趋向性步”表示紧跟着上一次的翻转算子或者紧跟着上一次游动算子的翻转.

3.3 简化*E.Coli*觅食优化原理(Principle of simplified *E.Coli* foraging optimization algorithm)

假设 p 维的优化问题, θ 表示要寻优的参数, 对应为觅食区域中*E.Coli*的位置, θ 的所有可行解就构成了优化问题的论域, 也即对应于*E.Coli*觅食环境空间.

S 为*E.Coli*群体大小, $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T \in \mathbb{R}^p$ 表示第 i 个*E.Coli*个体在觅食优化空间中的位置, 其中 $i = 1, 2, \dots, S$, $eval_i = f(X_i)$ 表示评价第 i 个*E.Coli*个体位置($\theta_i \in \mathbb{R}^p$)好坏的适应度函数, 通常就是所要优化问题的函数值. P_{best_i} 和 $P_i^{\text{pbest}} = (p_{i1}^{\text{pbest}}, p_{i2}^{\text{pbest}}, \dots, p_{ip}^{\text{pbest}})^T$ 分别表示第 i 个*E.Coli*个体经历过的历史最佳适应度函数值及其对应的位置; G_{best} 和 $G^{\text{gbest}} = (g_1^{\text{gbest}}, g_2^{\text{gbest}}, \dots, g_p^{\text{gbest}})^T$ 分别表示*E.Coli*群体中*E.Coli*经历过的历史最佳适应

度函数值及其对应的位置.

假设 N_c 为*E.Coli*的生命周期步数, 在简化*E.Coli*觅食优化算法中设定为化学趋向性步数, 在第 $n(n \leq N_c)$ 次化学趋向性步中, 各个*E.Coli*的位置先由算式(2)进行更新:

$$X_i^{n+1} = X_i^n + C(i)\phi(n), \quad (2)$$

$$X_i^{n+1} = X_i^n + w_1 r_1^n (P_i^{pbest} - X_i^n) + w_2 r_2^n (G^{gbest} - X_i^n). \quad (3)$$

式中: C_i 表示在游动方向 $\phi(n)$ 下游动算子的步长, $\phi(n)$ 为分布在 $[-1,1]$ 之间的随机数, 由游动算子产生. 优化值跟踪算子定义为: 如果在位置 X_i^{n+1} 时的适应度函数值 $eval_i^{n+1} = f(X_i^{n+1})$ 好于在位置 X_i^n 时的适应度函数值 $eval_i^n$, 则保持 $\phi(n)$ 的大小值, 在游动算子作用下, 以式(2)进行位置更新, 直到达到最大游动步数 N_s ; 否则, 利用式(3)进行位置的更新, 式中 w_1, w_2 为加权系数, r_1^n, r_2^n 为分布在区间 $[0,1]$ 的随机数.

3.4 简化*E.Coli*觅食优化算法流程(Procedure of simplified *E.Coli* foraging algorithm)

用于优化问题的简化*E.Coli*模型实现流程描述如下:

步骤 1 进行相关参数的初始化, 包括优化参数 θ 的论域范围, 化学趋向性步数 N_c , 游动算子的步数 N_s 和步长 $C(i)$, 种群数 S , *E.Coli*初始位置 $X_i(i = 1, 2, \dots, S)$, 加权系数 w_1, w_2 以及算法提前结束的条件等.

步骤 2 对于每个*E.Coli*个体, 用式(2)进行位置的更新, 评价其适应度函数 $eval_i^{n+1}$.

步骤 3 如果 $eval_i^{n+1} < eval_i^n$, 设置游动计数器 $Counter$ 为0, 保持游动方向 $\phi(n)$ 不变, 进行游动算子, 直到最大游动步数 N_s 或者不满足 $eval_i^{n+1} < eval_i^n$ 的条件. 同时更新 P_{best_i}, P_i^{pbest} , 以及 G^{gbest} 和 G^{gbest} .

步骤 4 如果 $eval_i^{n+1} > eval_i^n$, 则利用式(3)进行*E.Coli*位置的更新.

步骤 5 满足提前结束条件, 退出计算; 否则, 继续流程.

步骤 6 转到步骤 2, 直到群体中最后一个*E.Coli*个体.

步骤 7 转入下一次化学趋向性步.

4 仿真实验及结果(Simulation experiment and results)

4.1 试验对象描述(Description of experimental object)

为了验证本文提出的简化*E.Coli*觅食优化算法非线性系统建模效果, 根据文献[7]中提供的重油热

解三集总模型及实验数据进行模型参数估计. 实验数据可参见文献[7].

根据反应动力学特性, 重油热解三集总模型^[7]可表示为

$$x_L = \frac{K_{LP0}e^{-E_{LP}/T}}{n_L} [1 - (1-x)^{n_L}] + \frac{K_{WP0}K_{WLP0}e^{-(E_{WP}+E_{WLP})/T}}{n_w - K_{WLP0}e^{-E_{WLP}/T}} \cdot \left\{ \frac{1}{K_{WLP0}e^{-E_{WLP}/T}} (1 - (1-x)^{K_{WLP0}e^{-E_{WLP}/T}}) + \frac{1}{n_w} [(1-x)^{n_w} - 1] \right\}. \quad (4)$$

其中: x 和 T (反应温度)是自变量, x_L 为因变量.

利用简化*E.Coli*觅食优化算法进行模型参数估计的目的就是要确定式(4)中的模型系数: $K_{LP0}, K_{WP0}, K_{WLP0}, E_{LP}, E_{WP}, E_{WLP}, n_L, n_w$, 使得 x_L 与实验测得的 $\tilde{x}_L$ 具有较高的拟合度. 本文利用文献[7]中的任意20组实验数据作为样本, 进行参数估计. 简化*E.Coli*觅食优化算法的评价函数采用式(1)离散化的形式, 重新表述为

$$fitness = \sum_{i=1}^n [x_L(i) - \tilde{x}_L(t)]^2. \quad (5)$$

式中 n 为用于模型参数估计的样本个数.

4.2 试验过程(Experimental process)

1) 确定辨识参数的论域大小.

根据经验, 大致确定待估计参数的取值范围: $K_{LP0} \in (0, 10), K_{WP0} \in (0, 10), K_{WLP0} \in (0, 10), E_{LP} \in (800, 1500), E_{WP} \in (1500, 4000), E_{WLP} \in (1500, 4500), n_L \in (0, 5), n_w \in (0, 5)$.

2) 确定简化*E.Coli*觅食优化算法参数: $S = 20, N_c = 500, N_s = 4, w_1 = 0.2, w_2 = 0.2$, 初始游动步长 ϕ_n 设为0.01. 因本文针对系统模型的8个参数进行估计, 故取 $p = 8$.

3) 编程实现非线性参数简化*E.Coli*觅食优化NSM参数估计算法.

4.3 试验结果与分析(Results and analysis of experiment)

表1给出了本文方法与文献[7]中SGA, EGA的3种方法模型参数估计试验结果.

利用表1中3种方法得到的估计值作为模型参数, 并对从文献[7]的试验数据中随机选取的20组数据(建模样本), 进行模型拟合度的比较. 比较结果为: 采用SGA算法辨识得到的模型剩余标准差为0.0858, 采用EGA算法得到的模型剩余标准差为0.1089, 采

用简化*E.Coli*觅食优化算法得到的模型剩余标准差为0.0753. 3种算法的模型拟合曲线如图2所示. 从模型拟合曲线及模型剩余标准差的比较结果可以看出, 本文方法的拟合精度高于SGA和EGA算法. 可见本文方法是有效的、可行的.

表1 模型参数估计结果

Table 1 Estimated results of model parameters

参数	方法		
	SGA	EGA	Simplified <i>E.Coli</i>
K_{LP0}	4.680	4.032	4.161
K_{WP0}	5.155	5.058	2.925
K_{WLP0}	4.197	4.228	3.117
E_{LP}	1257	1166	1124
E_{WP}	1850	3676	2916
E_{WLP}	3776	2382	3198
n_L	1.191	1.357	1.614
n_w	1.488	1.570	1.721

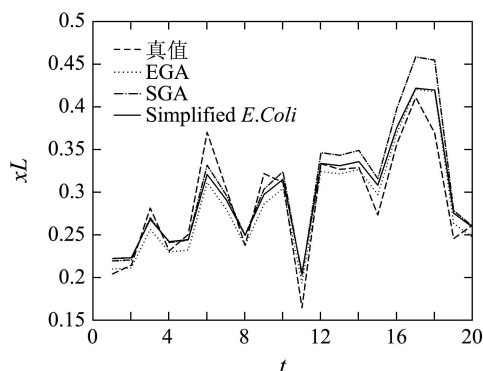


图2 3种算法的模型拟合曲线比较

Fig. 2 Regressive compared curves by three methods

从文献[7]的测试数据中再次随机抽取20组数据作为测试样本, 分别采用上述方法得到的模型参数来验证SGA, PGA和本文的简化*E.Coli*觅食优化算法得到的模型的拟合程度. 进行3次这样的测试试验, 并计算平均模型剩余标准差, 得到的结果列于表2. 从表2可以看出, 对于测试样本简化*E.Coli*觅食优化算法得到的非线性模型拟合程度也优于SGA和EGA.

表2 测试样本计算结果

Table 2 Calculated results of measured sample data

	辨识方法		
	SGA	EGA	Simplified <i>E.Coli</i>
平均剩余标准差	0.1052	0.1396	0.0769

5 结论(Conclusion)

本文提出了一种新的简化*E.Coli*觅食优化算法, 并将其用于非线性系统模型(NSM)参数辨识.

简化*E.Coli*觅食优化模型结构简单, 实现容易, 游动算子能够保证了算法在某一优化方向的持续优化, 翻转算子和对群体最优值的跟踪算子保证了简化*E.Coli*觅食优化算法的随机性, 这些算子的共同作用, 保证了简化*E.Coli*觅食优化算法的随机性能和寻优能力. 对重油热解三集总非线性模型进行了参数估计的实验结果表明: 利用简化*E.Coli*觅食优化算法进行模型参数估计是可行和有效的, 从而为解决非线性系统模型参数估计问题提供了新方法.

参考文献(References):

- [1] 尹力, 刘强, 王惠文. 偏最小二乘相关算法在系统建模中的两类典型应用[J]. 系统仿真学报, 2003, 15(1): 135 – 137, 145.
(YIN Li, LIU Qiang, WANG Huiwen. Two typical applications of the related partial least squares algorithms in the field of system modeling[J]. *J of System Simulation*, 2003, 2(5): 135 – 137, 145.)
- [2] 邓自立, 马建为, 杜洪越. ARMA 模型参数估计的两段最小二乘法[J]. 科学技术与工程, 2002, 2(5): 3 – 5.
(DENG Zili, MA Jianwei, DU Hongyue. Two-stage least squares algorithm of parameter estimation for ARMA model[J]. *Science Technology and Engineering*, 2002, 2(5): 3 – 5.)
- [3] ASO M, SAIKAWA T, HATTORI T. Mobile station location estimation using the maximum likelihood method in sector cell systems[C] // *Proc of Vehicular Technology Conf*. New York: IEEE Press, 2002, 2: 1192 – 1196.
- [4] 潘峰, 冯东梅, 韩如成. 基于遗传算法的时变时滞参数辨识[J]. 仪器仪表学报, 2004, 25(4): 910 – 912, 918.
(PAN feng, FEN Dongmei, HAN Rucheng. On-line identification of time-varying time-delay based on genetic algorithm for large time-delay systems[J]. *Chinese J of Scientific Instrument*, 2004, 25(4): 910 – 912, 918.)
- [5] 柯晶, 钱积新. 应用粒子群优化的非线性系统辨识[J]. 电路与系统学报, 2003, 8(4): 12 – 15, 7.
(KE Jin, QIAN Jixin. Nonlinear system identification using particle swarm optimization[J]. *J of Circuits and Systems*, 2003, 8(4): 12 – 15, 7.)
- [6] PASSINO K M. Biomimicry of bacterial foraging for distributed optimization and control[J]. *IEEE Control Systems Magazine*, 2002, 22(3): 52 – 67.)
- [7] 宋晓峰, 陈德钊, 胡上序, 等. 基于进化策略的遗传算法对重油热解模型参数的估计[J]. 高校化学工程学报, 2003, 17(4): 411 – 417.
(SONG Xiaofeng, CHEN Dezhaoh, HU Shangxu, et al. Eugenic evolution strategy genetic algorithms for estimating parameters of heavy oil thermal cracking model[J]. *J of Chemical Engineering of Chinese Universities*, 2003, 17(4): 411 – 417.)

作者简介:

刘益剑 (1978—), 男, 南京师范大学讲师, 武汉大学博士研究生, 目前研究方向为智能优化算法及嵌入式系统设计, E-mail: liuyujian_2002@163.com;

方彦军 (1957—), 男, 教授, 博士生导师, 主要研究方向为检测与控制方面的教学与研究工作, E-mail: yjfang@163.net;

孙冀 (1963—), 男, 副教授, 主要研究方向为分布参数控制理论与应用, E-mail: sg@mail.nsgk.net;

张建明 (1968—), 男, 副研究员, 主要研究方向为先进控制理论与应用、智能计算, E-mail: jmzhang@iipc.zju.edu.cn.