

文章编号:1006-3080(2009)02-0298-04

一种 DEA-PSO 混合算法及其在丙烯腈收率软测量中的应用

林 晨， 俞金寿

(华东理工大学自动化研究所, 上海 200237)

摘要:提出了一种 DEA 与 PSO 相结合的混合算法,即用 DEA 算法对 PSO 中适应值较差的粒子群进行重组和优化。将此混合算法与 PSO 算法同时用于一些常见测试函数的优化问题,通过对比表明:与 PSO 算法相比,DEA-PSO 混合算法的优化效果更佳。用 DEA-PSO 混合算法训练神经网络,并将其用于丙烯腈收率软测量建模,结果显示了该混合算法在丙烯腈软测量建模中的可行性与有效性。

关键词:DNA 进化算法(DEA); PSO; 丙烯腈; 软测量

中图分类号:TP18

文献标志码:A

A Hybrid Algorithm of DEA and PSO and Its Application to Soft Sensing of Acrylonitrile Yield

LIN Chen, YU Jin-shou

(Institute of Automation, East China University of Science and Technology, Shanghai 200237, China)

Abstract: A hybrid algorithm of DEA and PSO is proposed, in which the DEA is utilized to improve the bad sub-swarms of PSO. By resolving the optimization problems of several widely used test functions, it is showed that the proposed algorithm has better optimization performance than the standard PSO. Finally, the hybrid algorithm is employed to train the artificial neural network that is applied to soft-sensing of acrylonitrile yield. The results show that the hybrid algorithm is feasible and effective in soft-sensing of acrylonitrile yield.

Key words: DNA evolutional algorithm(DEA); PSO; acrylonitrile; soft sensing

粒子群优化算法(PSO)是基于对鸟群社会行为模拟的演化算法。PSO 算法由于概念简单、实现容易,短短几年时间便获得了很快的发展。但是对于一些复杂的问题,仍存在一些缺陷,如在搜索后期,因为粒子趋于同一化,失去了多样性而最终陷于局部最优解等^[1-2]。本文引入了采用单亲生物繁殖机理的 DEA 算法与 PSO 算法结合,提出了一种新型的 DEA-PSO 混合算法。实验结果表明该混合算法能够有效地解决这一问题。

1 粒子群优化算法

如前所述,PSO 通过模拟鸟群捕食的行为来寻找全局最优解。当求解优化问题时,问题的解就对应于搜索空间中一只鸟的位置,即“粒子”,每个粒子都有自己的位置和速度。各个粒子记忆、追随当前的最优粒子,并根据它自身的“经验”和相邻粒子群的最佳“经验”在解空间中向更好的位置“飞行”。

收稿日期:2008-04-14

作者简介:林晨(1984-),男,福建人,硕士生,主要研究方向:优化算法,DNA 计算。

通讯联系人:俞金寿,E-mail:jshyu@ecust.edu.cn

假设搜索区域为 N 维空间, 则第 i 个粒子的位置可以表示为 $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$, 速度可以表示为 $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$, 粒子根据下列方程更新本身的速度和位置^[2]:

$$v_{id} = \chi[\omega v_{id} + c_1 r_1(p_{id} - x_{id}) + c_2 r_2(p_{gd} - x_{id})] \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

2 DEA 算法

2.1 基本定义^[3]

定义 1 个体编码(DNA 单链)和它的互补链及它们的逆转(反演)所构成的集合, 称为该个体的分裂子群。在二进制编码中, 0 和 1 互补。但生物 DNA 分子中, 两条互补链方向相反。为不产生歧义, 把 DNA 单链编码的互补链区分为同(向)互补链和异(向)互补链。例如, 编码 111010 的同向互补链是 000101, 而异向互补链是 101000。

定义 2 DNA 单链经过不多于一次变异所能迁移的局部范围(或构成的点集), 称为该链的变异邻域。例如, 一条定长为 L 的二进制编码有 L 个变异数包含于它的变异邻域。它们中每一个只有一位的值与原编码的位值不同。

定义 3 定长编码的全体 DNA 单链的集合构成一个进化空间, 记做 D 。 D 中全体子集构成的集合称做 D 上的一个集类(Aggregates), 记为 D^* 。

2.2 基本操作算子

基本操作算子有 3 种: 分裂、变异和选择。

分裂算子: 如果 F 是一个 $D \rightarrow D^*$ 上的映射, 且对于 $\forall v \in D$, 有

$F(v) = \{v, -v, \bar{v}, -\bar{v}\} \equiv [v \ -v \ \bar{v} \ -\bar{v}] \in D^*$, 则称 F 是 $D \rightarrow D^*$ 上的分裂算子。这里是 $\{v, -v, \bar{v}, -\bar{v}\}$ 的 v 分裂集, $-v, \bar{v}, -\bar{v}$ 分别表示串 v 的倒置、同互补和异互补。

变异算子: 如果 M 是 $D \rightarrow D^*$ 上的一个映射, 且对于 $\forall v \in D$, 有 $M(v) = \{v_1, v_2, \dots, v_L, v\} \equiv [m_1 \ m_2 \ \dots \ m_L \ m] \otimes v = [v_1 \ v_2 \ \dots \ v_L \ v]$, 则称 M 为 $D \rightarrow D^*$ 上的变异算子。这里 $\{v_1, v_2, \dots, v_L, v\}$ 是 v 的变异邻域, m_i 是 $D \rightarrow D^*$ 上的位变异映射($i=1, 2, \dots, L$), 它使得 v 上的第 i 位发生变异。变异算子 M (位变异算子集)的作用是使基因映射成它的变异近邻(集合)。

选择算子: 选择算子有 2 个。一个是从分裂集中选取一个 DNA‘表形’, 另一是从变异邻域中选取一个 DNA‘变型’。前者称做水平(分裂)选择, 后者

称做垂直(变异)选择。

3 DEA-PSO 混合算法

3.1 基本原理

通过对 PSO 算法的分析可以发现, 如果相邻粒子群的历史最优适应值在一定时间内未发生变化, 则在粒子群很接近历史最优值时, 更新主要由第一部分决定, 因此粒子群表现出强烈的“趋同性”, 易陷入局部最优解。而 DEA 算法的独特算子使得它拥有较好的全局和局部搜索能力, 因此将这两种算法进行结合, 从 PSO 算法中选取适应值较差的那部分粒子交给 DEA 算法进行分裂、变异和选择, 然后将处理后的粒子返回至 PSO 算法的粒子群中, 这样既保持了 PSO 算法快速收敛的特性, 又丰富了粒子的多样性, 扩大了粒子的搜索范围, 使得 PSO 算法比较容易跳出局部极小, 从而能够比较好地收敛到最优值。

3.2 算法流程

DEA-PSO 混合算法的基本流程:

(1) 初始化 DEA-PSO 的参数;

(2) 对每个粒子的状态进行初始化, 计算对应的适应值, 并保存每个粒子个体最优位置、分群的最优位置与全群的最优位置, 及其对应的适应值;

(3) 对粒子按照适应值的优劣进行排序, 选出适应值较差的那部分粒子, 进行编码转换, 将粒子由十进制转换为二进制或者 DNA 编码, 以方便进行 DEA 算法计算;

(4) 对编码转换后的粒子进行 DEA 操作, 其中变异算子和分裂算子可以分别单独使用, 也可以结合起来并行或者串行使用;

(5) 将 DEA 算法优化后的粒子进行解码操作, 将粒子由二进制或者 DNA 编码转换为十进制, 并用这些粒子替代原先的粒子。更新个体最优位置、全局最优位置以及全局最优适应值;

(6) 如果停止条件未满足, 转步骤(3); 否则停止迭代, 输出全局最优解。

4 测试函数优化结果分析

4.1 测试函数

本文使用 3 个测试函数。

(1) F1 函数

$$\max f(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + x_2 \sin(20\pi x_2)$$

$$-3.0 \leqslant x_1 \leqslant 12.1, 4.1 \leqslant x_2 \leqslant 5.8$$

这个函数只在(11.625 5, 5.725)处收敛至唯一的全局最大值, 全局最大值为 38.850 3。这是一个很难的多峰值优化函数。

(2) F2 函数(考虑二元的情况)

$$\max f(x_1, x_2) = \frac{\sin(\sqrt{(x_1-50)^2 + (x_2-50)^2} + e)}{(\sqrt{(x_1-50)^2 + (x_2-50)^2} + e)} + 1,$$

$$-100 \leqslant x_1, x_2 \leqslant 100$$

该函数也是多峰值优化函数, 只在全局最小值点(50, 50)处取得最大值 1.151 1。

(3) Schaffer F6 函数

$$\max f(x_1, x_2) = 0.5 - \frac{(\sin \sqrt{x_1^2 + x_2^2})^2 - 0.5}{(1 + 0.001(x_1^2 + x_2^2))^2},$$

$$-10 \leqslant x_1, x_2 \leqslant 10$$

该函数也是很难的多峰值优化函数, 它有无数

个极大值点, 其中只有在(0, 0)处取得全局最大值, 最大值为 1。

4.2 测试结果与分析

本文同时使用 DEA-PSO 混合算法和 PSO 对上述测试函数进行优化。最大迭代次数均为 2 000 代。惯性权值都是按照余弦规律由 1.8 衰减到 0.02。加速系数 c_1 和 c_2 均取 2。粒子的个数均为 30 个。DEA-PSO 混合算法的编码方式采用 DNA 编码, 每个 DNA 的编码长度为 16 位。

由于 DEA-PSO 混合算法和 PSO 都是随机搜索算法, 单凭一两次搜索的结果难以说明搜索的有效性。本文从概率角度来对比算法的有效性, 优化时分别用各算法对每一测试函数进行 100 次独立测试, 并进行统计汇总。测试结果的统计汇总对比见表 1。

表 1 测试函数优化结果

Table 1 Optimization results of test functions

Test functions	DEA-PSO			PSO		
	Success rate/ (%)	Optimal fitness	Mean fitness	Success rate/ (%)	Optimal fitness	Mean fitness
F1	56	38.850 3	38.820 6	1	38.850 3	38.733 9
F2	32	1.151 1	1.138 9	6	1.151 1	1.130 1
Schaffer F6	87	1	0.999 34	68	1	0.997 04

由表 1 可知, 在算法参数等同的情况下, DEA-PSO 混合算法寻优时获得全局最优解的概率明显高于 PSO 算法。在测试过程中还发现, PSO 算法容易陷入局部极值点, 而且一旦陷入就很难跳出局部最优。特别明显的是在 F1 函数中, PSO 算法极其容易陷入局部极大值 38.732 8, 而 DEA-PSO 混合算法则比较好地收敛到全局最大值 38.850 3。以上数据表明由于加入了 DEA 算法, 相比于 PSO 算法, DEA-PSO 混合算法的性能更加优秀。

5 DEA-PSO 混合算法在丙烯腈收率软测量中的应用

在丙烯腈的生产过程中, 丙烯腈收率是一个关键指标^[4], 及时、准确地获取产品中丙烯腈收率是对丙烯腈生产过程进行先进控制的关键因素。然而, 由于在线分析仪表价格昂贵、滞后大, 最终将导致控制质量的性能下降, 难以满足生产要求, 因此采用软测量技术对丙烯腈收率进行预测具有重要的实际意义。

5.1 用 DEA-PSO 混合算法构造神经网络应用于丙烯腈收率软测量

使用 DEA-PSO 混合算法训练 BP 神经网络, 构建基于 DEA-PSO 混合算法的神经网络结构作为丙烯腈收率的软测量模型。根据工艺机理和操作经验, 选择反应压力、中段温度、纯丙烯腈量、空气比、氨比、触媒密度、反应线速为软测量建模的辅助量, 共 7 个输入、1 个中间层, 中间层节点数为 15, 丙烯腈的收率为模型的输出, 构造出结构为 7-15-1 的神经网络作为软测量模型, 神经网络的传递函数均为 S 函数, 该模型的目标函数为^[5]:

$$\min E = \sum (t - y)^2, \text{ 其中 } t \text{ 为收率实际值, } y \text{ 为收率预测值。}$$

5.2 两种 BP 神经网络模型的比较

将采集到的 328 组数据根据 3σ 准则剔除异常数据, 进行平滑、归一化处理后, 将得到的 317 组数据分为两份, 其中的 267 组数据用于神经网络训练。使用 DEA-PSO 算法进行训练后, 得出的训练样本均差为 0.407 7, 绝对误差的平均值为 0.194 9。没参加训练的 50 组数据作为检验样本对所建模型进

行检验,检验样本的均方差为 0.239 9,绝对误差的平均值为 0.173 4。所有的样本数据中,有 74% 的样本数据的误差绝对值小于 0.2。这些数据表明:所建模型的预测值与实际化验值的拟合程度比较好,模型的泛化性能也比较好,该模型能较好地预测丙烯腈收率(AN)。实际值与预测值的对比结果如图 1 所示。

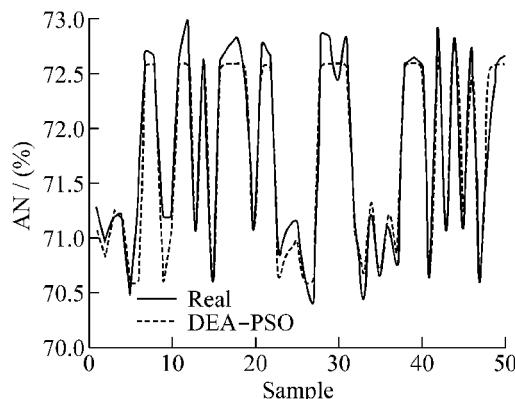


图 1 DEA-PSO BPNN 的实际值与预测值对比

Fig. 1 Predicted value of DEA-PSO BPNN model and real value

采用同样的数据,使用 PSO 训练神经网络,构建出粒子群神经网络(PSO BPNN)的丙烯腈收率软测量模型,模型的训练与检验结果也还可以,但是不如 DEA-PSO BPNN 模型得到的结果好。它的检验样本的均方差与绝对误差的平均值都比 DEA-PSO BPNN 高,所有样本数据中也只有 60% 的样本误差小于 0.2。这表明 DEA-PSO BPNN 模型相比较 PSO BPNN 的测量精度有了一定的提高。具体对比见表 2。

表 2 DEA-PSO BPNN 模型与 PSOBPNN 模型比较

Table 2 Comparison of neural network models based on DEA-PSO and PSO

Model	Training sample		Test sample	
	MSE	MAE	MSE	MAE
DEA-PSO BPNN	0.407 7	0.194 9	0.239 9	0.378 3
PSO BPNN	0.439 8	0.244 3	0.173 4	0.224 9

6 结束语

本文提出的 DEA-PSO 混合算法能够比较好地克服 PSO 算法容易陷入局部极值的缺点,通过测试函数的测试可发现,DEA-PSO 混合算法在优化速度和优化精度方面都有所提高。并且,基于 DEA-PSO 混合算法建立的神经网络软测量模型也有较好的表现。DEA-PSO 混合算法为 PSO 算法的改进提供了一条新的思路与方法。

参考文献:

- [1] 陈国初,俞金寿.微粒群优化算法[J].信息与控制,2005,34(3):318-324.
- [2] 柯晶,钱积新,乔谊正.一种改进粒子群优化算法[J].电路与系统学报,2003,8(5):87-91.
- [3] 余文,李人厚.一种有效的双向进化算法[J].小型微型计算机系统,2003,24(3):527-530.
- [4] 聂大仕,张强,陈章茂.丙烯腈的研究与应用进展[J].化学工业与工程技术,2005,26(2):35-36.
- [5] 田景文,高美娟.人工神经网络算法研究及应用[M].北京:北京理工大学出版社,2006.