

文章编号:1000-582X(2008)02-0166-04

基于模糊聚类的改进遗传算法

唐 然, 龙腾锐, 龙向宇

(重庆大学 三峡库区生态环境教育部重点实验室, 重庆 400030)

摘 要:针对导致遗传算法早熟收敛的原因,提出一种基于模糊聚类的改进遗传算法(FMGA),给出了FMGA算法实施的详细步骤,并研究确定了算法控制参数的取值。最后,对FMGA进行了数值仿真,仿真结果表明,FMGA能有效避免早熟收敛,在较短时间内逼近全局最优解,运算结果较基本遗传算法的提高4个数量级,而且运算过程不存在震荡现象。

关键词:遗传算法;模糊聚类;早熟收敛

中图分类号:X192

文献标志码:A

Modified Genetic Algorithm Based on Fuzzy System

TANG Ran, LONG Teng-rui, LONG Xiang-yu

(Key Laboratory of Three Gorges Reservoir Region's Eco-nvironment, Ministry of Education, Chongqing University, Chongqing 400030, P. R. China)

Abstract: To solve premature convergence of Genetic Algorithm, Modified Genetic Algorithm based on Fuzzy System (FMGA) is presented, detailed steps for FMGA is developed, and the parameters of FMGA is determined. FMGA is applied to optimize Camel Function. The simulation results show that FMGA can well prevent premature convergence to get global optimal, and this algorithm is accurate and has good stability.

Key words: genetic algorithm; fuzzy system; premature convergence

遗传算法(Genetic Algorithm,简称GA)是一种模仿生物进化过程的随机并行自适应的全局优化方法。这种算法具有不要求目标函数可导及可行域可凸,以及不存在枚举法的维数灾问题的优点;并且,它充分利用搜索过程中积累的寻优信息来搜索解空间,降低了随机搜索的盲目性^[1-3]。然而,基本遗传算法(Canonical Genetic Algorithm,简称CGA)存在易早熟收敛的缺点^[1]。本文针对导致遗传算法过早收敛的主要原因,建立了一种基于模糊聚类的改进遗传算法(Modified Genetic Algorithm based on Fuzzy System,简称FMGA),并通过仿真试验证明了该算法的有效性。

1 基于模糊聚类的改进遗传算法(FMGA)

1.1 FMGA 设计思想

马氏链模型认为,导致遗传算法种群序列 $X(t)$ (t 为种群代数, $t=1, \dots, T$)过早收敛的原因是种群多样性过小,对此,可通过增大种群多样性或者增大种群满意值,来避免种群序列 $X(t)$ 过早收敛,且保证其能以概率1收敛到满意解。根据生物进化理论,不同种群间的交流能促进整个种群达到更高层次的平衡,即远亲繁殖能增强种群的多样性,从而实现种群进化到最优值。在遗传算法中,采用模糊聚类分析对种群个体进行分类,将不同类别的父代个体杂交,实现种群的远亲繁殖,从而保证种群序列收敛到满意解。另一方面,将父代中的最优个体保留到子代,能确保整

收稿日期:2007-11-12

基金项目:中国建设部荷兰赠款项目(MOC-NGGP-2003-3)

作者简介:唐然(1979-),女,重庆大学博士研究生,研究方向为水污染控制与规划。龙腾锐(联系人),男,重庆大学教授,博士生导师,(Tel)023-65126098;(E-mail)LongTR@126.com。

个群体中优秀个体的继续遗传和繁衍,并能增大种群满意值^[4]。对于遗传算法而言,引入启发式策略指导搜索过程是提高运算速率的有效手段;运算过程中,将优秀个体的取值范围作为下一步优化运算的可行域,能压缩搜索范围,提高运算速率。

基于上述思想,笔者采用模糊聚类分析法对父代个体分类,从选择、杂交及变异操作三方面对CGA予以改进:首先利用欧几里德距离描述种群个体间的相似关系,构造种群的相似矩阵;然后用模糊聚类分析将父代个体分类,将不同类的个体相杂交,保障种群的多样性;在算法中,将父代的优秀个体保留到子代,增大种群满意值,避免CGA的早熟。为进一步提高遗传算法的收敛速率,将两次“选择—杂交—变异”操作后得到的优秀个体的取值范围作为下一步优化运算的可行域。

1.2 FMGA 步骤

笔者从选择算子、杂交算子、变异算子以及解变量的变化区间4个方面对基本遗传算法予以改进,提出基于模糊聚类的改进遗传算法(FMGA),该算法共由8步组成。

不失一般性,设优化函数见式(1):

$$\begin{aligned} & \text{Min } f(x_1, \dots, x_m, \dots, x_M), \\ & \text{s. t. } a_m \leq x_m \leq b_m; \quad m = 1, \dots, M. \end{aligned} \quad (1)$$

步骤1 解变量初始变化区间的离散及其编码

1)采用 l 位二进制编码表示解变量 x_m ,其二进制符号串表示为: $bs(m, k)$ ($m = 1, \dots, M; k = 1, \dots, l$)。每个变量被视为一个遗传基因;则 M 维决策向量 $\mathbf{X} = (x_1, \dots, x_m, \dots, x_M)$ 可视为由 M 个遗传基因所组成的一个染色体。

2)将解变量 x_m 初始变化区间 $[a_m, b_m]$ 等分为 $(2^l - 1)$ 个子区间,每个子区间的长度为 $d_m = \frac{(b_m - a_m)}{(2^l - 1)}$,即

$$x_m = a_m + I_m \cdot d_m. \quad (2)$$

式(2)中, I_m —搜索步长,为小于 2^l 的任意十进制随机非负整数。

变量 x_m 的二进制和十进制表示的转换关系见式(3)

$$x_m = a_m + I_m \cdot d_m = a_m + \left[\sum_{k=1}^l bs(m, k) \cdot 2^{k-1} \right] \cdot d_m. \quad (3)$$

3) M 维决策向量 $\mathbf{X} = (x_1, \dots, x_m, \dots, x_M)$ 表示为 $L = M \cdot l$ 长的二进制符号串。

步骤2 初始父代群体 $\mathbf{X}(t)$ 的随机生成

1)设群体规模为 N ,则有:

$$\mathbf{X}(t) = (X_1, \dots, X_i, \dots, X_N)^T = \begin{bmatrix} \{bs(1,1,k)\} & \dots & \{bs(1,m,k)\} & \dots & \{bs(1,M,k)\} \\ \vdots & & \vdots & & \vdots \\ \{bs(i,1,k)\} & \dots & \{bs(i,m,k)\} & \dots & \{bs(i,M,k)\} \\ \vdots & & \vdots & & \vdots \\ \{bs(N,1,k)\} & \dots & \{bs(N,m,k)\} & \dots & \{bs(N,M,k)\} \end{bmatrix} =$$

$$\begin{bmatrix} \{BS(1,L)\} \\ \vdots \\ \{BS(i,L)\} \\ \vdots \\ \{BS(N,L)\} \end{bmatrix}; \quad (k = 1, \dots, l, L = 1, \dots, M \cdot l). \quad (4)$$

2)为得到 $\vec{\mathbf{X}}(t)$,对变量 x_m ,随机生成 N 个二进制符号串 $\{bs(i, m, k) | i = 1, \dots, N\}$ 。

步骤3 父代个体的排序

1)对每个染色体 $\{BS(i, L)\}$ ($i = 1, \dots, N$),将经由式(3)解码得到的个体 $(x_{i1}, \dots, x_{im}, \dots, x_{iM})$ 带入优化函数 $f(x_1, \dots, x_m, \dots, x_M)$,得到相应的函数值 f_i ,共得到 N 个 f_i 值。

2) f_i 值越小,表示该个体的适应度值越高,反之亦然。把 f_i 按从小到大的顺序排序,对应的决策向量 X_i 和染色体二进制数串 $\{BS(i, L) | i = 1, \dots, N; L = 1, \dots, M \cdot l\}$ 也随之排序。

3)将排序后位于最前面的 A 个个体定义为优秀个体。

步骤4 父代个体的选择

1)借鉴最优保存策略(elitist selection)思想,将 A 个优秀个体直接作为参与后续交叉和变异操作的父代个体。将适应度值好的父代个体保留到子代,可增大种群的满意值。

2)采用竞争选择法(tournament selection)确定其余的 $(N - A)$ 个参与后续交叉和变异运算的父代个体,由于个体被选中的概率不直接正比于其适应度值,从而改善了CGA的早熟收敛,tournament尺寸 $\text{Tour} \geq L^{[5]}$ 。

由1),2)得到第一组选择后的父代种群 $X'(t)$ 。

3)重复步骤4的2),得出另一组选择后的父代种群 $X''(t)$ 。

步骤5 父代个体的杂交

1)依据生物学理论中近亲繁殖易导致种群退化、远亲繁殖利于种群进化的思想,为保持样本的多样性,尽量将差异大的父代个体杂交。

对2组选择后的父代群体 $X'(t)$ 和 $X''(t)$ 的共 $2N$ 个父代个体 $\{X_i\}$ ($i = 1, \dots, 2N$),用欧几里德距离表示个体间的相似程度,记为:

$$r_{ij}(X_i, X_j) = \left[\sum_{m=1}^M (x_{im} - x_{jm})^2 \right]^{\frac{1}{2}}; \quad (i, j = 1, \dots, 2N), \quad (5)$$

用 $\{r_{ij}\}$ 构造种群 $\{X_i\}$ 的模糊相似矩阵 $\mathbf{R}_{2N \times 2N}$,根据直接聚类原则^[6],设定适宜的聚类水平 λ ,将 $\{X_i\}$ 分类,然后把不同类的 X_i 随机配对,由此来保持或增加种群的多样性。

2)杂交概率 $P_c = 1.0$,两点杂交,取杂交后的第一个个体为杂交结果,由此得到子代群体 $\mathbf{X}(t + 1)$ 。

步骤6 子代个体的变异

采用两点变异,变异的概率为 p_m ,且 $p_m = O\left(\sqrt{\frac{1}{N}}\right)^{[5]}$ 。

步骤 7 进化迭代

将得到的 N 个子代 $\mathbf{X}(t+1)$ 作为新的父代, 算法转入步骤 3, 进入下一次选择、杂交和变异的进化过程。将运行两次步骤 3 到步骤 7 操作的过程称为“提高收敛性的遗传算法”(FMGA1)。

步骤 8 加速收敛

1) 用经过一次 FMGA1 (即: 两次选择、杂交和变异操作) 后得到的 A 个优秀个体 $\{X_i\} (i=1, \dots, A)$ 的解变量 x_m 的取值区间 $[a'_m, b'_m]$ 替代其初始变化区间 $[a_m, b_m]$, 并以 $[a'_m, b'_m]$ 作为变量的新初始变化区间, 算法进入步骤 1, 并在新的初始变化区间上随机生成初始父代群体, 重新运行 FMGA1, 从而充分利用已积累的优化信息指导遗传算法在可行域空间的搜索, 达到降低无效搜索量和提高运算速率的目的。

2) 以上 8 步构成“基于模糊聚类的改进遗传算法”(FMGA)。

1.3 FMGA 控制参数的取值

FMGA 的控制参数有五个: 变量编码长度 l 、群体规模 N 、优秀个体数目 A 、聚类水平 λ 和变异概率 p_m 。

1) 变量编码长度 l

变量 x_m 二进制编码的长度 l 由实际问题解变量的取值范围和对解的精度要求共同确定^[7]。假设实际问题对解的精度要求为 ∂ , 变量的取值范围为 $[a, b]$, 则 l 为: $l = \ln 2^{\left(\frac{b-a}{\partial} + 1\right)}$ 。

2) 群体规模 N 和优秀个体数目 A

种群规模 N 决定初始种群的个体在解空间分布及其密度, 目前, N 多依靠经验确定。对一定规模的种群, 优秀个体数目 A 的设置必须合理。 A 若太小, 则优秀个体包围最优点的概率过小; A 若过大, 则优秀个体变量变化范围过大, 失去设置优秀个体达到降低无效搜索量和提高运算速率的目的。为确定 N 和 A 的适宜取值范围, 本文以 Camel 函数为测试函数, 取 $N=100, 200, 300, 400$ 和 500 , $A=5\%N, 10\%N, 20\%N, 30\%N, 40\%N$ 和 $50\%N$, 对每种 N, A 的组合, 分别运行 10 次 FMGA, 分析算法的运行情况。

① 群体规模 N

从寻优次数和运算结果的稳定性看, $N=200$ 的寻优效果优于 $N=100$, 当 $N \geq 300$ 时, 算法的稳定性显著增强且收敛速率明显提高; 从运算时间看, $N \leq 200$ 时的运算时间相近, 当 $N \geq 300$ 时, 运算时间稍有延长, 其具体表现为: $N=300$ 的运算时间比 $N \leq 200$ 延长约 2.5%, $N=400$ 的比 $N=300$ 的延长约 6%, $N=500$ 的比 $N=400$ 的延长约 30%。

② 优秀个体数目 A

在同一 N 下, A 值的变化对运算时间的影响不大。从寻优效果和算法稳定性看, 当 $N \leq 200$ 时, $A \geq 10\%N$ 时的运算效果明显优于 $A=5\%N$ 。当 $N \geq 300$ 时, 不同 A 值下的运算效果都较好。在同一 A 值下, 运算结果的准确度(用“十次运算结果的均值

- 函数理论值”表示) 随着 N 的增大而增大; 在同一 N 下, $A=5\%N \sim 30\%N$ 的准确度最好。

综上所述, 建议 FMGA 的群体规模 N 取为 300 及其以上, A 为 $5\%N \sim 30\%N$ 。为缩短运算时间, 在较大 N 下 A 可取小值, 在较小 N 下 A 可取大值。

3) 聚类水平 λ

聚类水平 λ 的取值首先要保证所分得各类的个体数目尽量相等; 其次, 分得的类别数宜为偶数; 最后, 尽量避免在达到算法终止条件前(例如一定的迭代数) 父代个体被归为一类。

由于种群中所有个体完全相同的状态的发生概率很低, 遗传算法收敛一般指种群进化到所有个体均相互接近的一种状态, 因此, 使父代个体不被归为一类的聚类水平 λ 是客观存在的。为便于后续的配对杂交操作, 宜将选择得到的个体分为两类。为使每类的个体数目尽量接近, 以 X_1 为基准分类, 根据直接聚类原则“在相似矩阵中, 若 $r_{ij} \geq \lambda$, 则 X_i 和 X_j 在 λ 水平上同类”, 确定相应的聚类水平 λ 。

① 对种群 $\{X_i\} (i=1, \dots, 2N)$ 的模糊相似矩阵 $R_{2N \times 2N}$, 取第一列的 $2N$ 个元素 $r_{i,1} (i=1, \dots, 2N)$, 将它们按从小到大的顺序排序。

$R_{2N \times 2N} =$

$$\begin{array}{c|cccc} & X_1 & X_2 & X_3 & \cdots & X_{2N} \\ \hline X_1 & r_{1,1}(1) & & & & \\ X_2 & r_{2,1} & r_{2,2}(1) & & & \\ X_3 & r_{3,1} & r_{3,2} & r_{3,3}(1) & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ X_{2N} & r_{2N,1} & r_{2N,2} & r_{2N,3} & \cdots & r_{2N,2N} \end{array} \quad \circ$$

② 将位于第 N 位的元素的值 r 定义为聚类水平 λ 。

4) 变异概率 p_m

FMGA 采用两点变异, 变异概率 $p_m = O\left(\sqrt{\frac{1}{N}}\right)$,

其中, $L = l \cdot M$ 。

2 仿真试验

本文利用 Delphi 语言在计算机上仿真, 将基于模糊聚类的改进遗传算法 (FMGA) 用于 Camel 函数, 并将它与基本遗传算法 (CGA) 的结果相比较, 验证 FMGA 的优越性。

Camel 函数的数学表达式为: $f(x, y) = (4 - 2.1 \cdot x^2 + \frac{x^4}{3}) \cdot x^2 + x \cdot y + (-4 + 4 \cdot y^2) \cdot y^2 (x, y \in [-100, 100])$ 。该函数有 6 个局部极小点, 其中的 $(-0.0898, 0.7126)$ 和 $(0.0898, -0.7126)$ 为全局最小点, 最小值为 $f(x, y) = -1.031628$ 。

为了便于比较两种算法 (即, FMGA 与 CGA), 两者的试验运行参数设置一致, 即种群规模取为 400, 变量编码长度为 10, 交叉概率为 1.0, 变异概率为 0.74, 具体见表 1。对 FMGA, 取优秀个体数目 $A=40$, 聚类水平 $\lambda=r$ 。

试验中,以执行一次“选择—杂交—变异”操作为一代,取演化代数 $T=100$ 为运算终止条件。每种算法运行 10 次,得到每种算法的运算结果,绘出运算中每代种群的最优值随进化代数 T 的演变情况。运算结果表明,CGA 容易陷入局部极小点,得到的最优值在点 -0.8913 或点 -1.0249 附近徘徊;相比而言,FMGA 能有效避免早熟收敛,每次运算结果在理论值附近稍有波动。表 2 列出了两种算法 10

次运行结果的平均值,FMGA 运算结果明显优于 CGA,前者较后者更逼近于理论值,而且运算结果的准确度更高。图 1 绘出了某次运算两种算法寻优过程,从图 1 可知,FMGA 除了较 CGA 能避免早熟收敛而收敛到理论值外,运算的稳定性也有显著改善。由此可以认为,基于模糊聚类的改进遗传算法(FMGA)较基本遗传算法(CGGA)更能克服早熟收敛获得最优值,且运算过程更为稳定。

表 1 FMGA 和 CGA 的参数设置

算法	种群规模	编码长度	优秀个体数	杂交	变异
CGA	400	10	\	不分类,单点杂交, $p_c = 1.0$	单点变异, $p_m = 0.74$
FMGA	400	10	40	分类,两点杂交, $p_c = 1.0$	两点变异, $p_m = 0.74$

表 2 两种算法运行效果比较

算法	$f(x,y)$		$ x $		$ y $	
	运算结果	理论值	运算结果	理论值	运算结果	理论值
CGA	-0.96538889 $\pm 2.3508104 \times 10^{-2}$	-1.031628	0.09767 $\pm 8.333 \times 10^{-5}$	0.0898	0.68433 $\pm 8.3333 \times 10^{-5}$	0.7126
FMGA	-1.0316232 $\pm 1.7297525 \times 10^{-6}$		0.08874 $\pm 3.166 \times 10^{-4}$		0.71283 $\pm 1.3652 \times 10^{-3}$	

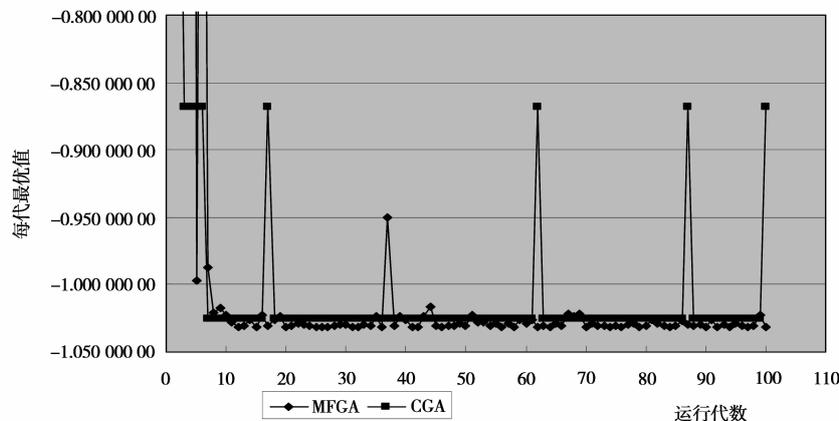


图 1 算法寻优曲线图

3 结 语

针对基本遗传算法易早熟收敛的缺点,笔者提出一种基于模糊聚类的改进遗传算法(FMGA),该方法首先采用欧几里德距离度量种群个体间的差异,然后用模糊聚类分析将父代个体分为两类,将不同类的个体相杂交;并且,将父代的部分优秀个体保留到子代,增大满意值;最后,将算法两次“选择—杂交—变异”操作后得到的优秀个体的取值范围作为下一步优化运算的可行域,以提高运算速率。为验证 FMGA 的优越性,笔者采用 Delphi 语言对 Camel 函数进行仿真,运算结果表明,FMGA 能很好的逼近全局最优值,能避免早熟收敛,运算结果的准确度高且运算过程稳定。

参考文献:

[1] KREINOVICH V, QUITANA C, FUENTES O. Genetic algo-

rithms; what fitness scaling is optimal [J]. Cybernetics and Systems, 1993, 24(1): 9-36.
 [2] LYER, SRIKANTH, SAXENA K, et al. Improved genetic algorithm for the permutation flowshop scheduling problem [J]. Computer and Operation Research, 2004, 31(4): 593-606.
 [3] 刘红, 韦穗. 遗传算法分析[J]. 计算机技术与发展, 2006, 16(10): 80-83.
 LIU HONG, WEI HUI. Analysis on genetic operators [J]. Computer technology and development, 2006, 16(10): 80-83.
 [2] GOLDBERG D E. Genetic algorithms in search, optimization, and machine learning[M]. New York: Addison-wesley Publishing Company INC, 1989.
 [3] 张文修, 梁怡编. 遗传算法的数学基础[M]. 西安: 西安交通大学出版社, 2003.
 [4] 杨伦标, 高英仪. 模糊数学原理及应用[M]. 广州: 华南理工大学出版社, 2003.
 [5] 王小平, 曹立明. 遗传算法—理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2005.

(编辑 张小强)