

遗传算法求解精度与种群大小的函数关系

李 刚^{1,2}, 薛惠锋¹, 邢书宝³

(1. 西安工业学院 经管学院, 陕西 西安 710032;

2. 西北工业大学 自动化学院, 陕西 西安 710072;

3. 西安科技大学 管理学院, 陕西 西安 710054)

摘 要:遗传算法在多方面具有运用,过早收敛是其主要的局限性,扩大种群可有效避免过早收敛。使用 Matlab 遗传算法工具箱收集了大量求解精度随种群的增大而提高的数据,通过曲线拟合,高精度地得到了遗传算法求解精度与种群大小的函数关系。

关键词:遗传算法;求解精度;种群;函数关系;曲线拟合

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2006)07-0096-03

Function of Genetic Algorithms Solution Precision to Population Size

LI Gang^{1,2}, XUE Hui-feng¹, XING Shu-bao³

(1. School of Economics & Management, Xi'an Institute of Technology, Xi'an 710032, China;

2. College of Automation, Northwestern Polytechnic University, Xi'an 710072, China;

3. School of Management, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: The genetic algorithms may be utilized in many aspects, premature convergence is its main limitation, population expanding can effectively avoid premature convergence. Use Matlab genetic algorithms toolbox to collect large numbers of data, through curvefitting, accurately obtain the function of genetic algorithms solution precision to the population size.

Key words: genetic algorithms; solution precision; population; function; curvefitting

0 引言

遗传算法(Genetic Algorithms, GA)由美国学者 Holland 于 1975 年首先提出,它是模拟达尔文的遗传选择和自然淘汰的生物进化论的计算模型^[1]。经过 30 年的发展,遗传算法的应用愈来愈普及。遗传算法的主要优点是简单、通用、鲁棒性强,适合于并行分布处理。

遗传算法作为一种优化方法,存在自身的局限性,过早收敛是其主要的缺点之一。扩大种群是避免过早收敛直接有效的办法,即“撒大网捕小鱼”,但种群的大小与求解精度间存在何种函数关系?少有报道,值得人们研究。

1 遗传算法基本理论

遗传算法是一类随机优化算法,模拟了自然选择和遗传中发生的复制、交叉和变异等现象,从任一初始种群(Population)出发,通过随机选择、交叉和变异操作,产生

一群更适应环境的个体,使群体进化到搜索空间中越来越好的区域,这样一代一代地不断繁衍进化,最后收敛到一群最适应环境的个体(Individual),求得问题的最优解。

1.1 遗传算法的运算过程

遗传算法的主要运算过程如下:

(1) 编码:从表现型到基因型的映射称为编码。遗传算法在进行搜索之前先将解空间的解数据表示成遗传空间的基因型串结构数据。

(2) 初始群体的生成:随机产生 N 个初始结构数据,每个串结构数据称为一个个体, N 个个体构成了一个初始种群 $P(0)$, N 即为种群 P 的大小。遗传算法以这 N 个串结构作为初始点开始迭代。

(3) 设置停止种群进化的条件:进化代数、时间限制、适应度限制、代数停滞限制、适应度停滞限制等。

(4) 适应度值评价检测:适应度函数表明个体的优劣性,计算群体 $P(t)$ 中各个个体的适应度。

(5) 选择:将选择算子作用于群体。

(6) 交叉:将交叉算子作用于群体。

(7) 变异:将变异算子作用于群体。

(8) 群体 $P(t)$ 经过选择、交叉、变异运算后得到下

收稿日期:2005-11-07

基金项目:霍英东教育基金资助项目(81044)

作者简介:李 刚(1971-),男,湖北襄樊人,博士研究生,讲师,研究方向为数据挖掘、信息化系统工程;薛惠锋,教授,博士生导师,兼任西安理工大学教授,研究方向为管理系统工程、信息化工程等。

一代群体 $P(t+1)$ 。

(9) 终止条件判断:未达到停止种群进化的条件,转到(5);否则,以得到的具有最大适应度的个体作为最优解输出,终止运算。

1.2 遗传算法的基本操作

遗传算法有 3 个基本操作^[2]:选择 (Selection)、交叉 (Crossover) 和变异 (Mutation)。

(1) 选择:选择的目的是为了从当前群体中选出优良的个体,使它们有机会作为父代为下一代繁殖子孙。选择运算体现了达尔文的适者生存原则。

(2) 交叉:交叉操作是遗传算法中最主要的遗传操作。通过交叉操作可以得到新一代个体,新个体组合了父辈个体的特性。将群体内的各个个体随机搭配成对,对每一个个体,以某个概率(称为交叉概率, Crossover Rate) 交换它们之间的部分染色体。交叉体现了信息交换的思想。

(3) 变异:对群体中的每一个个体,以某一概率(称为变异概率, Mutation Rate) 改变某一个或某一些基因座上的基因值为其他的等位基因。同生物界一样,遗传算法中变异发生的概率很低。变异为新个体的产生提供了机会。

2 实验环境

采用 Matlab7.0 遗传算法工具箱。

2.1 参数设置

种群类型 PopulationType: 'doubleVector'
精英个数 EliteCount: 2
交叉概率 CrossoverFraction: 0.8000
迁移方向 MigrationDirection: 'forward'
迁移间隔 MigrationInterval: 20
迁移概率 MigrationFraction: 0.2000
进化代数 Generations: Inf
时间限制 TimeLimit: Inf
适应度限制 FitnessLimit: - Inf
代数停滞限制 StallGenLimit: 50
适应度停滞限制 StallTimeLimit: Inf
初始种群 InitialPopulation: []
初始得分 InitialScores: []
绘图间隔 PlotInterval: 1
创建函数 CreationFcn: @gacreationuniform
适应度尺度函数 FitnessScalingFcn: @fitscalingrank
选择函数 SelectionFcn: @selectionstochunif
交叉函数 CrossoverFcn: @crossover scattered
变异函数 MutationFcn: @mutationgaussian
混合函数 HybridFcn: []
显示级别 Display: 'final'
绘图函数 PlotFcns: []
输出函数 OutputFcns: []

向量化 Vectorized: 'off':

种群初始范围 PopInitRange: 按测试函数要求设置

种群大小 PopulationSize: 5 - 10000

其它参数取默认值

2.2 重要参数解释

种群类型 - 'doubleVector': 种群中个体为双精度类型

精英个数 - 2: 具有最佳适应度的个体遗传到下一代的个体数为 2

进化代数 - Inf: 不限制进化代数

时间限制 - Inf: 不限制运算时间

适应度限制 - Inf: 不限制适应度精度

代数停滞限制 - 50: 若进化 50 代, 适应度无改进, 则终止运算

创建函数 - @gacreationuniform: 创建具有均匀分布的随机初始种群

适应度尺度函数 - @fitscalingrank: 根据个体适应度值排列顺序

选择函数 - @selectionstochunif: 随机均匀分布

交叉函数 - @crossover scattered: 创建一个二进制向量, 某位为 1, 基因从第一个父辈来; 为 0, 从第二个父辈来

变异函数 - @mutationgaussian: 把一高斯分布, 具有均值 0 的随机数加到父向量的每一项

种群初始范围: 指定变量求解的范围

种群大小: 指定种群中个体的个数, 文中在其它参数相同的情况下, 逐渐增大种群, 观察求解的精度

3 数据准备

3.1 测试函数

由于数学函数优化问题不需要专门领域知识, 且能较好地反映算法本身的实际效能, 所以常用于 GA 测试^[3], 文中采用 Shubert 函数用于测试。

Shubert 函数:

$$f(x, y) = \left\{ \sum_{i=1}^5 i \cos[(i+1)x + i] \right\} \left\{ \sum_{i=1}^5 i \cos[(i+1)y + i] \right\} + 0.5[(x + 1.42513)^2 + (y + 0.80032)^2]$$

此函数有 760 个局部极小点, 其中有一个 $(-1.42513, 0.80032)$ 为全局最小点。最小值为: -186.7309 。自变量的取值范围(PopInitRange) 为: $-10 < x, y < 10$ 。此函数极易陷入局部极小值 -186.34 ^[4]。

3.2 实验数据

实验参数如 2.1 所示设置, 实验设计为: 在同样种群值 p 下, 重复运行 100 次, 获取 100 个函数的最优值 y_i (求全局最小), 取其算术平均值作为此种群大小下函数求解结果 y , 即:

$$y = \sum_{i=0}^{100} y_i / 100$$

达优水平 S 定义为:

$$S = (y \times 100 / y_0) \%$$

y_0 为函数的理论最优解。

测试数据如表 1 所示:

表 1 Shubert 函数求解精度 S - 种群大小 P

P	10	20	30	40	50	60	70	80
S	53.07	71.11	81.45	81.42	87.31	89.14	91.39	91.88
P	90	100	110	120	130	140	150	160
S	93.72	92.95	94.67	95.17	96.33	95.68	97.22	97.05
P	170	180	190	200	210	220	230	240
S	99.59	99.66	99.70	99.62	99.75	99.76	99.74	99.78
P	250	500	1000	1500	2000	2500	5000	10000
S	98.45	99.52	99.75	99.82	99.89	99.88		

4 数据拟合及结论

根据表 1 数据画出求解精度 - 种群大小曲线如图 1 所示。

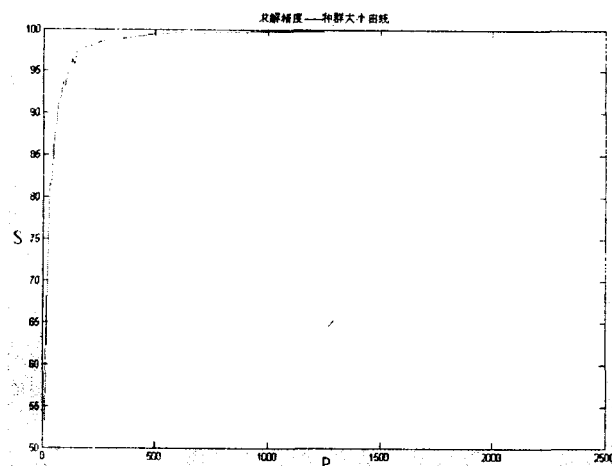


图 1 求解精度 - 种群大小曲线

由曲线的形态可以看出:GA 的求解精度 S 起始时随种群大小 P 的增大而迅速提高,在 $P = 500$ 时, S 值分别达到了 99.52, P 在 500 以上继续增大,但 S 值提高很慢。

观察可得:曲线形态符合 $f(x) = a * \exp(b * x) + c * \exp(d * x)$ 类型函数,用此函数对图 1 中曲线进行拟合,参数如表 2 所示。

表 2 拟合参数表

参数值				拟合评价
a	b	c	d	$R - \text{square}$
98.01	9.84E - 06	-43.75	-0.02467	0.9851

注: $R - \text{square} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
值愈接近于 1,拟合程度愈高^[5]。

拟合如图 2 所示。

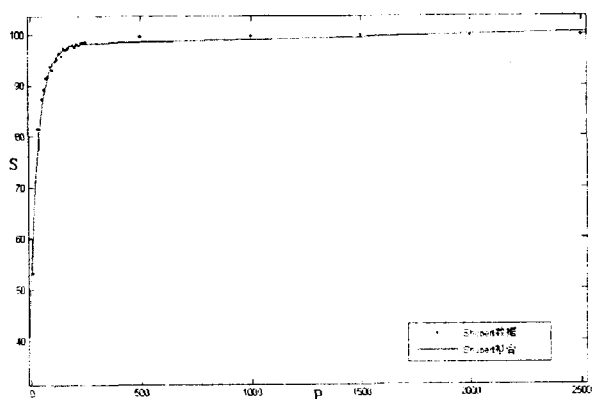


图 2 求解精度 - 种群大小拟合图

从表 2 可以看出:求解精度 - 种群大小曲线拟合的 $R - \text{square}$ 值很高(接近于 1),可以说明求解精度 - 种群大小的函数关系为:

$$f(x) = a * \exp(b * x) + c * \exp(d * x)$$

其中 a, b 为正值, a 值较大,接近于 100, b 值很小; c, d 为负值。从而保证种群在 10~100 时,求解精度上升很快,种群在 100 以上,求解精度上升较慢。

5 结束语

过早收敛是遗传算法的主要局限性。扩大种群可有效避免过早收敛,即“撒大网捕小鱼”,但网的大小与捕获精度之间的函数关系人们研究不多。

文中使用 Matlab7.0 遗传算法工具箱,收集 Shubert 函数的遗传算法求解精度随种群增大而变化的实验数据,以高精度拟合出它们的函数关系,揭示出求解精度 - 种群大小的函数关系的规律,对于优化算法参数具有较强的指导价值。

参考文献:

- [1] 周明,孙树栋. 遗传算法原理及应用[M]. 北京:国防工业出版社,1999.
- [2] 雷英杰,张善文,李续武,等. 遗传算法工具箱及应用[M]. 西安:西安电子科技大学出版社,2005.
- [3] 王晓平,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,2002.
- [4] 陈国梁,王煦法,庄镇泉,等. 遗传算法及应用[M]. 北京:人民邮电出版社,1996.
- [5] 傅鹂. 数学实验[M]. 北京:科学出版社,2001.

(上接第 95 页)

参考文献:

- [1] 焦李成. 神经网络系统理论[M]. 西安:西安电子科技大学出版社,1996.
- [2] 王永骥,涂健. 神经网络控制[M]. 北京:机械工业出版社,1998.
- [3] 陈详光,裴旭东. 人工神经网络技术及应用[M]. 北京:中国

电力出版社,2003.

- [4] Buckley J J. Stability and fuzzy controller[J]. Fuzzy Sets and Systems, 1996, 77: 167-173.
- [5] Wu Y, Shi H B. Fuzzy rules acquisition and parameters evolution based on fuzzy neural networks[A]. UCNN'99[C]. Washington D. C., USA: [s. n.], 1999.