

基于正交试验的遗传算法参数优化

王 雷, 蔡劲草, 李 明

(安徽工程大学机械与汽车工程学院, 安徽 芜湖 241000)

[摘要] 基本遗传算法求解优化问题的过程中存在着收敛缓慢、早熟现象以及求解的质量不高等问题。为了解决上述存在的问题, 提高遗传算法的求解质量, 提出使用正交试验法优化遗传算法中的主要参数, 即: 种群规模 N 、交叉概率 p_c 和变异概率 p_m 。通过使用正交试验法确定遗传参数, 大大提高了算法的收敛性和求解质量。仿真结果也表明采用正交试验法设计参数的科学性和有效性。

[关键词] 遗传算法, 参数优化, 正交试验

[中图分类号] TP18 **[文献标志码]** A **[文章编号]** 1672-1292(2016)02-0081-05

Parameter Optimization of Genetic Algorithm Based on Orthogonal Experiment

Wang Lei, Cai Jingcao, Li Ming

(School of Mechanical and Automotive Engineering, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: There exist slow convergence, premature problem, and the lower quality of the solution by using traditional genetic algorithm (GA) to deal with optimization problem. In order to solve these above-mentioned disadvantages and improve the solution quality, an orthogonal design method is proposed to optimize the main parameters of GA, namely population size N , crossover probability p_c and mutation probability p_m . As a result, the GA's evolutionary speed, global convergence and the solution quality can be improved. The simulation results indicate that this method is scientific and effective for dealing with parameter optimization problem.

Key words: genetic algorithm, parameter optimization, orthogonal experiment

遗传算法(genetic algorithm, GA)是在自然环境中模拟生物的“优胜劣汰”而形成的进化的智能化的优化方法, 往往能在搜索空间高度复杂的问题上取得较好的效果。其基本原理最早由 Holland 提出^[1]。近几十年来, 随着对理论和应用领域的不断深入研究和拓广, 遗传算法发展迅速, 并以其较优的性能引起了人们更多的关注, 已成功地应用于路径规划^[2-3]、作业车间调度^[4-7]、旅行商^[8-9]等领域。

尽管 GA 已成功应用于许多优化问题中, 但其本身也或多或少存在一些缺陷。比如, 算法基本模型中参数正确选择与否将对优化结果产生很大的影响。如何优化算法自身的参数使其性能较好, 从根本上来说就是一个较为复杂的研究难题。由于庞大的参数空间及各参数之间的关联性, 还没有最优参数确定的基本方法。经验法是目前较常用的优化方法, 但是经验法不能够保证针对每一类优化问题参数的选择都是最合适的。算法参数选择的理论性研究和实验分析论证非常必要。

作为一种试验的优化设计技术, 正交试验法的优点是能够以相当少的试验次数、较短的试验时间和较低的试验费用得到满意的试验结果。本文试图将正交试验法引入 GA 中, 对 GA 中关键参数——种群规模 N 、交叉概率 p_c 和变异概率 p_m 进行分析和研究。仿真结果表明了采用该方法优化参数的科学性和有效性, 也为该算法在其他研究方面选择参数时提供了一定的参考依据。

收稿日期: 2016-01-10.

基金项目: 国家自然科学基金(51305001)、安徽省高校优秀青年人才支持计划重点项目(gxyqZD2016125)、先进数控与伺服驱动安徽省重点实验室开放课题(xjsk003)。

通讯联系人: 王雷, 博士, 副教授, 研究方向: 智能优化算法及其在制造系统中的应用. E-mail: wangdalei2000@126.com

1 遗传算法与函数优化

遗传算法是从一组随机产生的被称为“种群”的初始解开始搜索过程,从而得到最优解的一种智能优化算法.遗传算法包括复制、交叉、变异3种操作算子.复制表明“优胜劣汰,适者生存”的自然进化选择思想;交叉算子能够使子代个体得以传承父代个体的优点;变异算子对扩大种群的多样性方面起着至关重要的作用.

1.1 函数优化

函数优化问题通常可归结为求极小值问题(对于极大值可以通过对函数加上负号,将其转化为最小值问题).一般问题的目标优化问题可以描述如下:

$$\min f(x), x \in S, \quad (1)$$

其中, S 为搜索空间, $f(x)$ 为目标函数.

1.2 函数优化问题中的遗传算法

1.2.1 编码

遗传算法编码方式有多种形式,如浮点数编码、二进制编码、整数编码、符号编码、矩阵编码等.对于实值函数编码,二进制编码是最常用的编码方法,其实际就是对搜索空间均匀划分,所求解的精度受到编码长度的限制.本文采用二进制编码.

1.2.2 适应度函数

遗传算法每进化一代后都需要对种群中的个体进行适应度评价,以便进行优胜劣汰,完成该评估的工作即称为适应度函数.一般以最小化函数为适应度函数,本文选择式(1)作为适应度的评价函数,即:

$$g = f(x). \quad (2)$$

1.2.3 复制

选择的作用是使适应度大的个体将以更大的概率被进化到下一代,这也是一般搜索算法与遗传算法的主要区别.已有许多如何选择最好个体的方法,如轮盘选择、波尔兹曼选择、锦标赛选择、排队选择等等.本文选用轮盘赌法.

1.2.4 交叉

当通过复制产生新的种群后,需要进行交叉操作.目前交叉操作方法主要有单点交叉、均匀交叉、多点交叉等.本文采用单点交叉,即:交叉使用两个个体,在一个随机选择的位置上将每一个体的染色体切割成前半段和后半段两个部分,然后将两个个体的前半部分或后半部分进行交换产生出两个新的子代个体.这两个个体各自继承了每个双亲的一些基因.

通常,交叉操作获得执行的概率相对变异操作要大得多,这是因为交叉操作能保持较好的模式,能够加强局部搜索能力,而变异操作是一个随机操作,用以在种群中引入多样性,扩展算法的全局搜索能力.交叉概率 p_c 一般设定在区间 $[0.6, 1]$ 之间.

1.2.5 变异

变异的主要作用是进一步扩大种群的搜索空间及多样性.为了使遗传算法能够保持较好的优化性能,变异概率 p_m 应设置在一个合适的范围内.变异操作改变了原有的染色体上的基因,对提高种群多样性具有明显的促进作用,但也有可能破坏原有已找到的优秀解的结构.若 p_m 过小,算法容易早熟;若 p_m 过大,可能会导致算法目前所处的搜索情况又回到原来较差的状态.因此,应将种群的变异限制在一定范围内,通常将 p_m 设定在区间 $[0.005, 0.1]$ 之间.图1给出了遗传算法求解优化函数问题的一般流程.

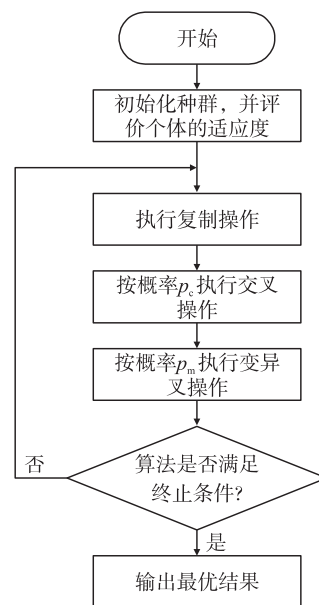


图1 遗传算法的基本流程

Fig.1 Basic flowchart of GA

2 基于正交试验的遗传算法参数优化

2.1 优化实例

求函数 $f(x) = 128x - x^2$ 的最大值, $x \in [0, 127]$. 已知当 $x=64.0$ 时, 其最大值为 4 096.

2.2 参数优化过程

按前述步骤通过遗传算法进行优化计算, 其中取进化代数 $T=100$, 只是在确定种群规模 N 、交叉概率 p_c 及变异概率 p_m 时, 需通过正交实验法来确定较好的遗传参数, 以便为类似函数的优化问题提供参考作用.

(1) 首先, 确定因素水平, 如表 1 所示.

(2) 其次, 选择 $L_{25}(5^6)$ 正交表, 并以优化值与已知最大值之间的误差 δ 和进化代数 T_0 为优化指标进行试验, 得到的试验结果如表 2 所示.

表 1 因素水平表

Table 1 Factors and their levels

水平	因数		
	N	p_c	p_m
1	40	0.5	0.01
2	60	0.6	0.05
3	80	0.7	0.1
4	100	0.8	0.2
5	120	0.9	0.3

表 2 GA 参数的整定实验

Table 2 Experiments of GA parameters' changes

试验号	P	p_c	p_m	误差 δ	进化代数 T
1	1(20)	1(0.5)	1(0.01)	0.000 019	51
2	1(20)	2(0.6)	2(0.05)	0.000 006	51
3	1(20)	3(0.7)	3(0.10)	0	51
4	1(20)	4(0.8)	4(0.20)	0.000 019	51
5	1(20)	5(0.9)	5(0.30)	0.000 010	51
6	2(40)	1(0.5)	2(0.05)	0.000 500	51
7	2(40)	2(0.6)	3(0.10)	0	32
8	2(40)	3(0.7)	4(0.20)	0.000 173	51
9	2(40)	4(0.8)	5(0.30)	0.000 700	51
10	2(40)	5(0.9)	1(0.01)	0.000 002	48
11	3(60)	1(0.5)	3(0.10)	0.000 203	51
12	3(60)	2(0.6)	4(0.20)	0.000 016	50
13	3(60)	3(0.7)	5(0.30)	0	27
14	3(60)	4(0.8)	1(0.01)	0	27
15	3(60)	5(0.9)	2(0.05)	0.000 075	51
16	4(80)	1(0.5)	4(0.20)	0	41
17	4(80)	2(0.6)	5(0.30)	0.000 006	51
18	4(80)	3(0.7)	1(0.01)	0	42
19	4(80)	4(0.8)	2(0.05)	0.000 053	50
20	4(80)	5(0.9)	3(0.10)	0.000 495	51
21	5(100)	1(0.5)	5(0.30)	0.001 901	51
22	5(100)	2(0.6)	1(0.01)	0.000 011	51
23	5(100)	3(0.7)	2(0.05)	0	34
24	5(100)	4(0.8)	3(0.10)	0	40
25	5(100)	5(0.9)	4(0.20)	0	39

(3) 对表 2 数据进行极差分析, 确定三因素水平对两个指标影响的主次关系, 并确定最优因素水平. 试验结果分析如表 3 所示. 表中, K_i 为水平号为 i 的指标值之和, R 为极差.

表 3 试验结果分析
Table 3 Analysis of experiment result

指标	N	p_c	p_m	指标	N	p_c	p_m
K_1	0.000 054	0.002 623	0.000 032	K_1	255	245	219
K_2	0.001 375	0.000 039	0.000 634	K_2	233	235	237
K_3	0.000 294	0.000 173	0.000 698	K_3	206	205	225
误差 δ	K_4	0.000 554	0.000 772	进化代数 K_4	235	219	232
	K_5	0.001 912	0.000 582	T_0	K_5	215	240
			0.002 617				231
极差 R	0.001 858	0.002 584	0.002 585	极差 R	49	40	18
因素主次	N	p_c	p_m	因素主次	p_m	p_c	N
优选方案 1	$N=20$	$p_c=0.6$	$p_m=0.01$	优选方案 2	$N=60$	$p_c=0.7$	$p_m=0.01$

(4)对最优解进行仿真.

从表 3 的分析结果知,对误差 δ 的影响因素主次分别是种群规模 N 、交叉概率 p_c 和变异概率 p_m ,且优选方案 1 为 $N=20$ 、 $p_c=0.6$ 和 $p_m=0.01$. 而对进化代数 T_0 的影响因素主次分别是变异概率 p_m 、交叉概率 p_c 和种群规模 N ,且优选方案 2 为 $N=60$ 、 $p_c=0.7$ 和 $p_m=0.01$. 尽管该方案在已经做过的 25 次试验中没有出现,但通过对这两个优选方案进行各自试验,结果表明这两个优选方案比其他方案的结果相对都要好,结果也表明采用该方法设计参数的科学性和有效性. 对两种优化方案分别试验 10 次取平均值,得到的误差 δ 、平均误差 δ_0 、进化代数 T 以及平均进化代数 T_0 如表 4、表 5 所示. 具体的两种方案比较结果见表 6.

表 4 优选方案 1 运行 10 次仿真结果

试验号	误差 δ	进化代数 T	平均误差 δ_0	平均进化代数 T_0
1	0.007 204	51	0.001 477	50
2	0	49		
3	0.002 328	51		
4	0.000 047	51		
5	0	48		
6	0	49		
7	0.000 075	51		
8	0	49		
9	0.000 006	51		
10	0.005 11	51		

通过表 4 和表 5 可看出,优选方案 1 尽管也有 4 次得到问题的最优解,但其平均误差以及平均进化代数要比优选方案 2 差得多,且优选方案 2 在随机仿真 10 次中,8 次得到问题的最优解. 由表 6 可以看出,在平均进化代数方面,优选方案 2 比方案 1 改进了 44%,而在平均误差方面改进了 99.6%.

因此,在综合比较两种优化方案后可知,当 $N=60$ 、 $p_c=0.7$ 和 $p_m=0.01$ 时,各项指标相对较优,能够满足优化的精度高和收敛速度快的要求. 因此,在本函数优化中的最优指标为 $N=60$ 、 $p_c=0.7$ 和 $p_m=0.01$.

3 结论

正交试验法利用正交表安排试验,使实验次数大大减少,效率较高,在这一点上值得遗传算法借鉴.

表 5 优选方案 2 运行 10 次仿真结果

试验号	误差 δ	进化代数 T	平均误差 δ_0	平均进化代数 T_0
1	0	28	0.000 006	28
2	0	37		
3	0	24		
4	0	28		
5	0	28		
6	0.000 050	27		
7	0	28		
8	0	24		
9	0	24		
10	0.000 010	28		

表 6 优选方案 2 与优选方案 1 比较结果
Table 6 Comparison results between scheme 1 and scheme 2

方案	指标	
	平均误差 δ_0	平均进化代数 T_0
优选方案 1	0.001 477	50
优选方案 2	0.000 006	28
改善程度(%)	99.6	44

备注: 改善程度 = $\frac{|\text{方案2} - \text{方案1}|}{\text{方案1}} \times 100\%$

通过使用该方法确定的遗传参数,大大提高了求解函数优化问题的收敛性和寻优精度。仿真结果也表明采用正交试验法设计参数的科学性和有效性。但针对不同的组合优化问题,遗传算法基本参数的选择会略有差别。

[参考文献](References)

- [1] HOLLAND J H. Adaptation in nature and artificial systems[M]. Ann Arbor: University of Michigan Press, 1975.
- [2] TUNCER A, YILDIRIM M. Dynamic path planning of mobile robots with improved genetic algorithm[J]. Computers & electrical engineering, 2012, 38(6): 1 564–1 572.
- [3] RAJA R, DUTTA A, VENKATESH K S. New potential field method for rough terrain path planning using genetic algorithm for a 6-wheel rover[J]. Robotics & autonomous systems, 2015, 72(C): 295–306.
- [4] IMEN D, KINZA N M, ASSIA L. A new genetic algorithm for flexible job-shop scheduling problems[J]. Journal of mechanical science and technology, 2015, 29(3): 1 273–1 281.
- [5] ISHIKAWA S, KUBOTA R, HORIO K. Effective hierarchical optimization by a hierarchical multi-space competitive genetic algorithm for the flexible job-shop scheduling problem[J]. Expert systems with applications, 2015, 42(24): 9 434–9 440.
- [6] KURDI M. An effective new island model genetic algorithm for job shop scheduling problem[J]. Computers & operations research, 2016, 67: 132–142.
- [7] CHEN J C, WU C C, CHEN C W, et al. Flexible job shop scheduling with parallel machines using Genetic Algorithm and Grouping Genetic Algorithm[J]. Expert systems with applications, 2012, 39(11): 10 016–10 021.
- [8] PAUL P V, MOGANARANGAN N, KUMAR S S, et al. Performance analyses over population seeding techniques of the permutation-coded genetic algorithm: an empirical study based on traveling salesman problems[J]. Applied soft computing, 2015, 32(C): 383–402.
- [9] 杨玉, 李慧, 戴红伟. 改进量子交叉遗传算法在 TSP 问题中的应用[J]. 南京师范大学学报(工程技术版), 2012, 12(9): 43–48.
YANG Y, LI H, DAI H W. Improved quantum crossover based GA and its application to traveling salesman problem[J]. Journal of Nanjing normal university(engineering and technology edition), 2012, 12(9): 43–48. (in Chinese)

[责任编辑: 严海琳]