

带有分级思想的自适应遗传算法

黄 利^{1,2} 丁立新¹ 杜伟伟³

(武汉大学软件工程国家重点实验室 武汉 430072)¹

(武汉理工大学理学院数学系 武汉 430070)² (中国北方自动控制技术研究所 太原 030006)³

摘 要 为了平衡演化算法的搜索效果和效率,在自适应遗传算法中引入分级思想,即根据种群中个体适应值的相似性对其进行分级,使得优秀个体和较差个体充分发挥各自的职能。然而,过早收敛也是遗传算法亟待解决的问题之一,因此对遗传算法中的选择操作进行改进,定义了一种新的选择机制:一方面是在选择过程中引入一个新的参数——成活率,以有效地避免算法过早收敛;另一方面结合模拟退火中的参数——温度,通过变异杂交生成虚拟种群,以有效扩大搜索空间,保持种群多样性。实验结果表明,利用新算法处理 TSP 问题能够产生高质量的解,并能有效解决过早收敛问题。

关键词 分级思想,遗传算法,自适应,选择机制

中图分类号 TP301.6 **文献标识码** A

Self-adaptive Genetic Algorithm with Classification

HUANG Li^{1,2} DING Li-xin¹ DU Wei-wei³

(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072, China)¹

(Department of Mathematics, School of Science, Wuhan University of Technology, Wuhan 430070, China)²

(North Automatic Control Technology Institute, Taiyuan 030006, China)³

Abstract Aiming at the balance of search results and search speed of evolutionary algorithm, we proposed a search strategy to classify the individuals by the similarity of their fitness. This differentiated respective function of individuals in search process. Nevertheless, premature convergence was one of GA-difficulties. So, an improved selection mechanism in GA was used to deal with the mentioned drawback. On the one hand, a new parameter named success ratio which is higher setting causes higher selection pressure. It could keep the algorithm from premature convergence. On the other hand, another parameter T the same as the simulated annealing algorithm's temperature was recommended. While mutation and crossover happened, a virtual population could be generated according to this parameter for enlarging the search space and keeping the diversity of generation. Finally, experimental results on benchmark problems of TSP show that the new method is capable of producing highest quality solutions and preventing premature convergence efficiently.

Keywords Classification, Genetic algorithm, Self-adaptive, Selection mechanism

1 引言

遗传算法是 20 世纪六七十年代由美国 Michigan 大学的 Holland 教授及其学生和同事发展起来的一种全局搜索优化算法,它具备自适应、自组织和自学习的特征,但所涉及到的个体大小及表示方法、遗传操作类型及概率设置等诸多因素在遗传算法的全部运行过程中都是静态的。为使遗传算法具有更强的自适应性,人们提出了自适应遗传算法,它考虑的是如何使遗传算法在求解某一问题的不同阶段自动地调整某些因素,如遗传操作类型及操作概率等,以达到“探测”和“开发”之间的最优平衡^[1]。在 1994—2006 年期间,与自适应算法相关的文献,2006 年的篇数是 1994 年的 5 倍之多,可见算法的自适应性越来越被人们所关注。从自适应遗传算法的内容上

来看,对遗传算法参数——种群规模 N 、交叉概率、变异概率的自适应设置是研究的热点。Arabas 提出对个体赋予寿命来调整 N ,保持种群多样性^[2]。Strinivas 提出了杂交概率和变异概率随父串适应度自适应变化^[3]。

本文在自适应遗传算法中引入了分级思想。分级思想是利用适应值对个体进行分级的搜索策略,通过对个体的分级,可以区分个体在搜索过程中的职能:优秀的个体进行局部极小值的开采,其它的个体进行搜索空间的探索,以发现新的局部极小值,从而达到搜索效果和效率的平衡^[4]。

将自适应的选择机制引入到遗传算法中是为了解决算法的过早收敛问题。所提及的自适应性主要体现在以下两点:一是引入新参数——成活率,用来控制父代个体在子代中成功存活下来的个数;二是对父代执行变异杂交操作,生成与温

到稿日期:2009-07-01 返修日期:2009-09-10 本文受高等学校博士点基金(20070486081),湖北省杰出青年基金(2005ABB017)资助。

黄 利(1981—),女,博士,讲师,主要研究方向为计算机软件与理论等,E-mail:l_huang@126.com;丁立新(1967—),男,教授,主要研究方向为计算机软件与理论等;杜伟伟(1981—),男,工程师。

度有关的虚拟种群,其中温度思想来源于模拟退火算法,采用几何级数递减的方式。这种选择机制不但可以有效地防止算法过早收敛,而且在一定程度上扩大了搜索空间,确保了种群多样性。

本文第2节详细描述了带有分级思想的自适应遗传算法;第3节用TSP问题来检验算法的有效性,并对结果进行了分析;最后是对全文的总结和展望。

2 带有分级思想的自适应遗传算法

2.1 分级思想^[4]

优化问题求解过程一般有两个特点:一是采用局部搜索算法,效率好、效果差;二是采用全局搜索算法,效果好、效率差。如果能够设计一种算法来融合局部搜索和全局搜索两方面的本质属性,就有可能实现算法效率与效果的平衡。

因此本文采取了种群分级的策略,根据适应值的相似性对个体进行分级。

分级的原则是把适应值相似的个体分配在同一级,它们的相似性根据适应值之间的距离来度量。

分级的目的是让优秀的个体进行局部极小值的搜索,让较差的个体进行搜索空间的探索,以发现新的局部极小值。产生新个体的规则是由它在搜索解空间的任务所确定,也就是说它与个体所在的级别相关联,即个体在父代中的级别次序确定产生新个体的搜索范围和它们所生成的个体数目。具体是在父代中级别较高的个体的搜索范围较小,生成的个体数目较少;父代中级别较低的个体的搜索范围较大,它生成的个体数目较多。

2.1.1 距离

定义1 设 $V=(v_1, v_2, \dots, v_n)$ 为图 $G(E, V)$ 的顶点集, E 为完全图的边集,任取 $e_i \in E, E_1 \subseteq E$, 记 $d(e_i, E_1) = \begin{cases} 1, & e_i \in E_1 \\ 0, & \text{others} \end{cases}$, 则称 $d(e_i, E_1)$ 为边 e_i 到边集 E_1 的距离。

定义2 设 $E_2=(e_1, e_2, \dots, e_m), E_1, E_2 \subseteq E$, 则称 $d(E_1, E_2) = \sum_{i=1}^m d(e_i, E_2)$ 为边集 E_1 到 E_2 的距离。

定义3 设 $E_1, E_2 \subseteq E$, 则称 $D(E_1, E_2) = d(E_1, E_2) + d(E_2, E_1)$ 为 E_1, E_2 之间的距离。

2.1.2 相似性

定义4 对于TSP问题的任意两个个体 X, Y , 定义 $S_{XY} = n - D_{XY}$, 其中 n 为TSP问题的城市数目(也即个体的长度), 则称 S_{XY} 为个体 X 和个体 Y 的相似度。

显然, S_{XY} 反映了个体之间相似性的大小。随着个体之间距离 D_{XY} 的减小, S_{XY} 增大, 表明个体之间的相似性增加。

2.2 自适应的选择机制

对种群的选择操作采用一种新的自适应的选择机制,即通过一个新的变量成活率^[6](sr , 指的是下一代要存活的个体占种群中总个体的比值)来控制父代个体在子代中存活个数。

因此,子代中的一部分个体由父代直接复制得来,具备较好的性质。这些个体通过该子种群的个数 M 与成活率相乘得到,即 $M * sr$; 而另一部分是先对父代个体实行变异和杂交操作,生成虚拟种群,然后在虚拟种群中进行选取,选择的个数为 $M * (1 - sr)$ 。虚拟种群的种群大小 V 由该子种群的个数与模拟退火中的温度 T 相乘得到:

$$V = M * T$$

同传统的遗传算法比较,这种处理方式会产生更多的个体,无疑扩大了解的搜索空间,并保持了种群多样性。随着演化的进行,逐渐降温,即 T 值逐渐减小,虚拟种群大小逐渐减小,搜索空间随之减小。在演化的过程中,温度使用几何级数的递减方式,即

$$T = T * \alpha$$

通过上述的选择方式分析,可以看到在这种新的选择机制下,选择出来的个体数是与 T 有关系的,而 T 又是在不断变化的(在一次循环中 T 不断减小;在子种群合并后, T 又要重新设置为一个较大的值,以便在新的一轮循环中降温),因此种群的选择压力是可变的。这样,就可以控制种群的多样性,从而避免陷入过早收敛。

2.3 算法框架

图1是本文算法的流程图。

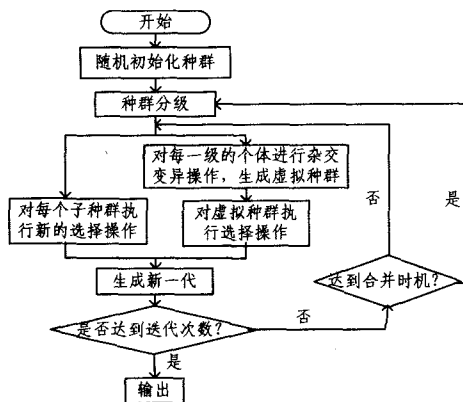


图1 算法流程图

Procedure 带有分级思想的自适应遗传算法

```

随机初始化种群;
计算初始种群的适应值;
while (未达到最大进化代数)
{
    t=0; //t 表示进化代数
    if (达到合并的时机)
    {
        k=k-1; //k 表示分级的级数
        重新设置温度 T;
    }
    计算第 t+1 代种群 P(t+1)=calculate(P(t), T, k, N);
    //N 表示种群大小
    徐徐降温; //采用几何级数的递减方式降温
    t=t+1;
}
function calculate(P(t), T, k, N) //生成新一代种群
for(j=1; j<=k; j++)
{
    for(i=1; i<=N; i++)
    {
        a=max f(Xi) - min f(Xi);
        b=a/k;
        if (min f(Xi) + (j-1) * b <= f(Xi) <= min f(Xi) + j * b)
        {
            将第 t 代的个体 Xi 归到第 j 级, 得到第 j 级的子种群;
            对每一级的个体计数;
        }
    }
}
    
```

计算第 j 级子种群的适应值;
 计算每一级虚拟种群的大小;

根据自适应选择机制生成第 $t+1$ 代的一部分个体;
 对第 j 级子种群的个体执行变异和杂交,生成虚拟种群;
 计算虚拟种群的适应值;
 在虚拟种群中选择第 $t+1$ 代的另一部分个体;

在整个演化过程中,种群的个体按一定的变异概率、杂交概率进行变异和杂交。变异操作可以采用贪心对换变异^[7],而杂交操作就可以选择基于路径的贪心杂交法^[7]。

对于算法框架中提到的合并时机,这里采用监测子种群的个体适应值变化的方式。当各个子种群的适应值在一定时间内不再变化时,就对子种群进行合并。即当算法陷入局部最优时,停止对子种群的演化。然后将各个子种群按照一定的原则合并,级数相应地减 1,此时温度需要重新设置为一个较大的值,以便在对新一代种群演化过程中再次徐徐降温。最后,对新一代种群划分后的各个子种群再次独立地进行演化,直到种群进化到一定代数 G 时,算法终止。

3 实验与分析

在 VC++ 6.0 平台下编程实现了本算法。选取 3 个 TSP 典型问题 oliver30, berlin52 和 kroA100 进行了仿真实验,检验了本算法性能。

3.1 参数设置

选择基于路径表示的编码方式,最大演化代数 $G=2000$,温度 $T=2$,冷却率 $\alpha=0.99$,成活率 $sr=0.2$,变异概率 $p_m=0.2$,杂交概率 $p_c=0.9$ 。

3.2 实验结果

按照上述取定的参数值,对上述 3 个 TSP 问题分别仿真 10 次,并将算法找到的最好路径结果用下述巡回图表示,如图 2,图 3 和图 4 所示, length 为路径距离总和。

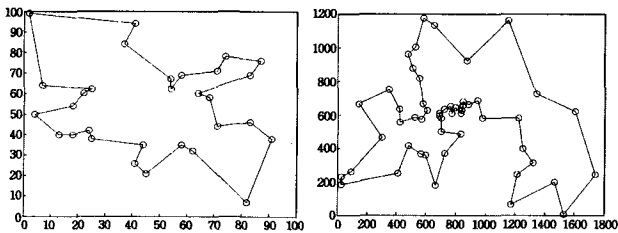


图 2 oliver30 的巡回路线
(length = 423.741)

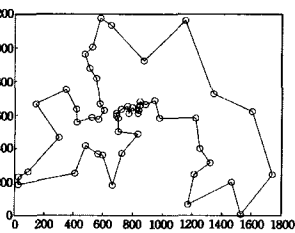


图 3 berlin52 的巡回路线
(length = 7544)

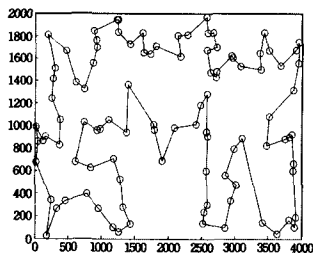


图 4 kroA100 的巡回路线 (length = 21285)

图 5 描绘的是 kroA100TSP 问题第 1 级子种群的收敛曲线。从图中可以清晰看出,第 1 级子种群能够较快地收敛到近似最优解。

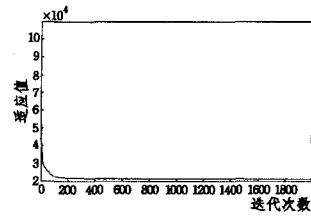


图 5 kroA100 第 1 级子种群收敛曲线

为了检验算法的收敛性能,对实验结果进行相对误差分析,相对误差计算公式如下:

$$relativeError = \left(\frac{result}{optimal} - 1 \right) * 100\%$$

式中, $result$ 是算法计算结果; $optimal$ 是当前公布最优结果。

利用上述 3 个 TSP 问题独立运行 10 次得到的实验数据计算各自相对误差,统计结果如表 1 所列。

表 1 收敛性能分析

问题	种群大小	子种群数	迭代次数	相对误差(%)		
				最小值	最大值	平均值
oliver30	40	1	2000	0.22	1.52	0.97
oliver30	40	4	2000	0	0.22	0.05
berlin52	200	1	2000	0.02	3.33	1.21
berlin52	200	4	2000	0.02	0.02	0.02
kroA100	400	1	2000	0.01	0.77	0.36
kroA100	400	4	2000	0.01	0.34	0.19

表 1 对分级前和分级后 ($k=4$) 的 3 个 TSP 问题在保持相同种群大小及迭代次数的情况下运行所产生的相对误差进行了比较,可看出分级后的相对误差较小。因此,可以肯定分级对算法搜索最优解确实有一定的帮助。

本文算法仿真得到的实验结果也优于文献[4]中算法的结果。从图 2—图 4 可以看出,得到的最短路径图基本上没有交叉的路径,可见算法的有效性,并且结果都十分接近最优值。从统计出的相对误差值可以看出,算法可以有效避免过早收敛,具有较好的稳定性和收敛性。如果扩大初始种群,增加迭代次数,算法的收敛效果应该会更好。

结束语 为了调节遗传算法效率和效果的不平衡性,以及缓解算法容易发生过早收敛的缺点,本文提出带有分级思想的自适应遗传算法。算法采用以下方法来解决上述问题。

(1)对遗传算法种群中的个体按适应值的相似性来分级,其目的是更好地搜索到局部最优值。此分级方式同种群无规律的任意划分方式^[5]比较,具备一定的优越性,它可以适当减少演化时间,从而提高算法运行的效率;与种群不进行任何划分方式比较,它能找到更优的解。

(2)对遗传算法的选择操作采用一种新的自适应选择机制,通过产生虚拟种群大小的温度 T 来自适应调整选择压力,保持种群的多样性。这在很大程度上避免了过早收敛的发生。

实验结果显示,本算法在平衡演化算法的搜索效果和效率以及避免算法过早收敛方面都有显著的效果。但是,由于采用较少的实例来验证算法的有效性,因此今后还需要从理论上更加详细地证明本算法的收敛性和演化代数等。

参考文献

[1] 巩敦卫,潘凤萍. 自适应遗传算法理论与应用[M]. 徐州:中国矿业大学出版社,2003:1-7

(下转第 250 页)

3.2 色差评价

CIELAB 为色度空间,其色差公式无法对光谱的转换精度进行评价,尤其无法给出同色异谱情况下的光谱比较^[14]。但其在色差评价过程中考虑到人眼视觉感知系统,因此可通过它对光谱转换后合成的颜色精度进行度量。实验选择 CIELAB 的标准色差公式 ΔE_{cb} ,其定义如下:

$$\Delta E_{cb} = \sqrt{((\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2)} \quad (7)$$

实验对 IT8.7/3 光谱数据、young girl 图、fruits and flowers 图的 ΔE_{cb} 统计结果如表 4 所列。通常当 ΔE_{cb} 小于 3 时,人眼几乎感知不到颜色差异;当 ΔE_{cb} 介于 3 和 6 之间时,人眼能够感知到色差,但色差可接受;当 ΔE_{cb} 大于 6 时,颜色严重失真,色差无法接受^[16]。从表 4 可以看出,3 幅图的平均 ΔE_{cb} 均小于 3,其中 IT8.7/3 和 young girl 图的平均 ΔE_{cb} 不到 1。空间转换引起的色差,young girl 图中几乎无法感知到,fruits and flowers 图中仅有极小部分的色差可被感知。

表 4 多光谱测试图色差统计

测试图	平均 ΔE_{ab}	最大 ΔE_{ab}	最小 ΔE_{ab}
IT8.7/3	0.1089	0.8693	0.0013
图 2 (a)	0.1231	0.3276	0.0040
图 2 (b)	1.2733	6.9524	0.0491

结束语 为解决光谱色彩管理中多光谱图像的输出问题,本文提出一种面向打印的中间空间构造方法。该方法在光谱空间与设备颜色空间之间引入中间空间作为空间变换桥梁,有效解决了光谱空间的高维度给光谱色彩管理带来的问题。针对多光谱图像的打印输出,在建立中间空间时,采用打印机的光谱特征化样本作为基础样本,通过主成分分析法构造低维空间,使得空间变换成本低,光谱及色度精度高。低维中间空间能用于查找表的建立,可实现与打印机颜色空间的转换。在构造中间空间时,需要对光谱空间进行降维,降维方法对光谱转换精度的影响很大,采用更适合的降维方法是需要进一步研究的方向。另外,中间空间的维度直接影响着光谱转换的精度,维数确定的标准也需要深入研究。

参考文献

[1] Zolliker P, Simon K. Retaining Local Image Information in Gamut Mapping Algorithms[J]. IEEE Trans on Image Processing, 2007, 16(3):664-672

[2] Chin E M, Garcia E K, Gupta M R. Color Management of Printers by Regression over Enclosing Neighborhoods[C]// IEEE Int. Conf. on Image Processing. San Antonio, TX, USA, 2007, 2: 161-164

(上接第 167 页)

[2] Arabas J, Michalewicz Z. GAVAPS — A Genetic Algorithm with Varying Population Size[C]// Proceeding of 1st Conference On Evolutionary Computation, 1994:73-78

[3] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in GAs[J]. IEEE Trans. on SMC, 1994, 24(4): 678-691

[4] 黄樟灿. 演化计算的搜索策略研究[D]. 武汉: 武汉大学, 2004: 48-53

[5] Affenzeller M. A New Approach to Evolutionary Computation;

[3] Zhao Y, Berns S R, Taplin A L, et al. An investigation of multi-spectral imaging for the mapping of pigments in paintings[C]// Proc. SPIE. 2008, 6810:1-9

[4] Bochko V, Tsumura N, Miyake Y. A spectral color imaging system for estimating spectral reflectance of paint[J]. Journal of Imaging Science and Technology, 2007, 51(1):70-78

[5] Munzenmayer C, Paulus D, Wittenberg T. A spectral color correction framework for medical applications[J]. IEEE Trans on Biomedical Engineering, 2006, 53(2):254-265

[6] Tsumura N, Miyake Y, Imai H F. Medical vision; measurement of skin absolute spectral-reflectance-image and the application to component analysis[C]// Proc. Third Int. Conf. on Multispectral Color Science MCS'01. 2001:25-28

[7] Hardeberg J Y. Recent Advances in Acquisition and Reproduction of Multispectral Images[C]// 14th European Signal Processing Conference. Florence, 2006

[8] Bochko V, Tsumura N, Miyake Y. Spectral Color Imaging System for Estimating Spectral Reflectance of Paint[J]. The Journal of Imaging Science and Technology, 2007, 51(1):70-78

[9] Nieves J L, Valero E M, Nascimento S M C, et al. Multispectral Synthesis of Daylight Using a Commercial Digital CCD Camera [J]. The Journal of Optical Society of America, 2005, 44(27): 5696-5703

[10] Cheung V, Westland S, Li C, et al. Characterization of Trichromatic Color Cameras by Using a New Multispectral Imaging Technique[J]. The Journal of Optical Society of America A, 2005, 22(7):1084-7529

[11] Urban P, Rosen R M, Berns S R. Accelerating spectral-based color separation within the Neugebauer subspace[J]. Journal of Electronic Imaging, 2007, 16(4):1-11

[12] Berns S R, Taplin A L, Liang Z T. Spectral color reproduction with six color output[P]. United States Patent:6698860. 2004

[13] 丁二锐, 曾平, 罗雪梅, 等. 非均匀三维查找表的快速查找技术[J]. 计算机科学, 2007, 34(2):258-259, 276

[14] Imai F H, Rosen M R, Berns R S. Comparative Study of Metrics for Spectral Match Quality[C] //The First European Conference on Colour Graphics, Imaging, and Vision. Poitiers, France, 2002:492-496

[15] 汤顺青. 色度学[M]. 北京:北京理工大学出版社, 1990:102-113

[16] Conde J, Haneishi H, Yamaguchi M, et al. CIE-XYZ Fitting by Multispectral Images and Mean Square Error Minimization with a Linear Interpolation Function[J]. REVISTA MEXICANA DE FISICA, 2004, 50(6):601-607

Segregative Genetic Algorithms (SEGA) [C] // Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence. Lecture Notes of Computer Science. 2001:594-601

[6] Affenzeller M, Wagner S. SASEGASA: An Evolutionary Algorithm for Retarding Premature Convergence by Self-adaptive Selection Pressure Steering[C]// IWANN 2003. Lecture Notes of Computer Science. 2003

[7] 林志毅. 改进的遗传算法求解 TSP 问题[D]. 武汉: 武汉理工大学, 2006:17-29