

面向旅行商问题的一种改进遗传算法

丁华福, 刘晓路, 唐远新, 石福斌

(哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080)

摘要:针对传统遗传算法解决组合优化领域的典型难题——旅行商问题时存在的缺陷与不足,对传统遗传算法加以改进。引入了种群熵和基因座多样性来测度种群的多样性,并利用求得的测度值修改了个体的适应度,达到了预期目的。还提出启发式交叉算子、二分局部搜索、自适应遗传参数,使该算法在提高搜索效率与解质量方面取得了综合平衡。在算法的仿真和测试中,改进后的算法明显优于传统的遗传算法。这表明,该算法具有良好的可行性和实用性。

关键词:遗传算法;种群多样性;启发式交叉算子;二分局部搜索;自适应遗传参数

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2011)06-0051-04

An Improved Genetic Algorithm for Traveling Salesman Problem

DING Hua-fu, LIU Xiao-lu, TANG Yuan-xin, SHI Fu-bin

(Sch. of Computer Sci. and Tech., Harbin Univ. of Sci. and Tech., Harbin 150080, China)

Abstract: By analyzing the deficiency of traditional genetic algorithm in solving the traveling salesman problem, one representative problem of the combination optimization, the algorithm structure of traditional genetic algorithm was improved. Entropy and the locus of population diversity are introduced to measure the population diversity in this paper, and the measure values obtained are used to modify the fitness of the individual to achieve the desired purpose. Also introduces self-adjust adapt value and proposing heuristic crossover operation, second detach local searching and self-adapt genetic parameter, the algorithm achieved a balance between quality and efficiency. According to the analysis and test, the improved genetic algorithm can get the better result than the traditional genetic algorithm. This shows that the method has better feasibility and practicability.

Key words: genetic algorithm; population variation; heuristic crossover operation; second detach local search; self-adapt genetic parameter

0 引言

当给定一组 N 个城市和它们两两之间的直达距离时,旅行商问题(Traveling Salesman Problem, TSP)是被描述为寻找一条闭合的路线,这条路线刚好经过每个城市一次且总的距离最短^[1]。TSP 问题是一个 NP 完全问题,属于经典的组合优化问题,其可能的搜索路径随着城市数目的增加,呈指数增长。所以较为实际的方法是尽快地发现其近似最优解,想要精确求出最优解是极为困难的^[2]。遗传算法简单易用、鲁棒性强,具有的全局收敛性和隐并行性,能够轻易地获得问题的全局最优解,是一种全局优化搜索算法,问题越复杂,它相对于其他算法的优越性就越明显,故十分适合解决这类问题。但传统的遗传算法往往由于种群多样性的减小而使算法过早地进入局部最优,即出现了过早收敛现象^[3]。经过实验发现,维持遗传算法在进化过程中的种群多样性可以很好地解决这个问题。因

而,有必要针对具体问题,对经典遗传算法加以改进。

马欣等提出了单亲进化遗传算法(PEGA),该算法通过利用父体所提供的有效边的信息,保留最小边的来优化个体,提高算法的收敛速度^[4]。吴志远等提出对算法的参数 P_c, P_m , 群体规模 N 进行自适应调整以维护群体的多样性,防止算法的过早收敛^[5]。Rudolph G 为了保证算法的收敛性,提出了保持群体中最好的个体不丢失的方法,即用精英选择策略^[6]。

文中基于适值自适应调整方法和遗传参数自适应调整方法增强种群的多样性,种群多样性的增加必然导致收敛速度的降低,因此文中又引入了启发式交叉算子和二分局部搜索技术来提高种群的收敛速度,并取得了不错的效果。最后为了证明算法的有效性和先进性,文中利用计算机仿真了国际标准数据库中的经典算例,仿真结果表明新算法在收敛速度和最优解的质量上均优于简单遗传算法和单亲进化遗传算法。

1 遗传算法的改进

1.1 基于熵测度的种群多样性增强方法

典型的种群的多样性测试方法有海明距离法,但

收稿日期:2010-11-01;修回日期:2011-02-10

基金项目:黑龙江省教育科研项目(11531049)

作者简介:丁华福(1962-),男,哈尔滨人,硕士研究生导师,研究员,研究方向为数据挖掘、数据库。

是这个方法有一定的局限性与不足之处^[7]。因此文中采用熵测度并结合基因座多样性来计算种群的多样性,并利用该多样性测度值动态调整了适应度函数,从而通过保留较小适应度的个体达到了增加种群多样性的目的。

种群熵的定义为:对于第 m 代种群,该种群有 M 个子集: $S_1^m, S_2^m, \dots, S_M^m$, 每个子集的个体数为: $|S_1^m|, |S_2^m|, \dots, |S_M^m|$, 设 $a, b \in \{1, 2, \dots, M\}, S_a^m \cap S_b^m = \emptyset, \bigcup_{b=1}^M S_b^m = A^m$, 则 A^m 为第 m 代种群的集合,那么第 m 代种群的熵就定义为:

$$E_m = - \sum_{i=1}^M a_i \log(a_i) \quad (1)$$

其中 $p_i = |S_i| / N$, N 为种群中的个体数目。

基因座的多样性^[8]表现了种群中不同的个体在同一个基因位上的差异程度。设 P_{ij}^k 表示个体 x_i 和个体 x_j 在第 k 个基因位上的差异,那么基因座多样性定义为:

$$P_{ij}^k = \begin{cases} 1, & g_{ik} \neq g_{jk} \\ 0, & \text{其他} \end{cases} \quad (2)$$

$$P^k(m) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N P_{ij}^k \quad (3)$$

其中 N 为种群个体数目, g_{ik} 和 g_{jk} 是个体在 k 个基因位上的基因。

文中定义在复制操作中,个体 x_i 的适应度为 $f(x_i)$,当被复制到下一代的时候适应度被调整为:

$$f'(x_i) = e^{uf(x_i)} \quad (4)$$

其中 $u = \partial(E_m/E_{\max} + \sum_{i=1}^{i=L} P^i(m)/L * N + \beta) (0 < \partial, \beta < 1)$, E_{\max} 表示种群中可能得到的最大的熵, L 为 x_i 个体的染色体长度,即基因个数, N 为种群中的个体数。通过将适应度调整后,就可以在种群多样性减小的情况下,增大选择操作的选择范围,从而有利于保留适应度较小的个体,也就增大了种群的多样性。

1.2 一种新的交叉算子

对于求解 TSP 问题,常见的杂交算子有顺序杂交算子、部分匹配杂交算子、循环杂交算子等^[9]。但是这些杂交算子有一个共同的缺点,就是都没有考虑 TSP 问题中边的信息并加以利用,因此文中采取了如下的启发式交叉算子通过利用边的信息来提高算法的性能。

文中算法的个体编码采用实数编码方式,将路径经过的每一个城市用一个实数来表示,从而生成一个数字串来表示一条巡回路径,例如:编码串 (x_1, x_2, \dots, x_n) 表示从城市 x_1 出发,依次经过城市 x_2, \dots, x_n ,最后回到出发城市 x_1 。对于群体中的任意两个父代个体 $X = (x_1, x_2, \dots, x_n), Y = (y_1, y_2, \dots, y_n)$,将 X 和 Y 看成是

两个首尾相连的环,然后通过文中提出的启发式杂交算子产生两个子代个体 child1、child2 的过程如图 1 所示。

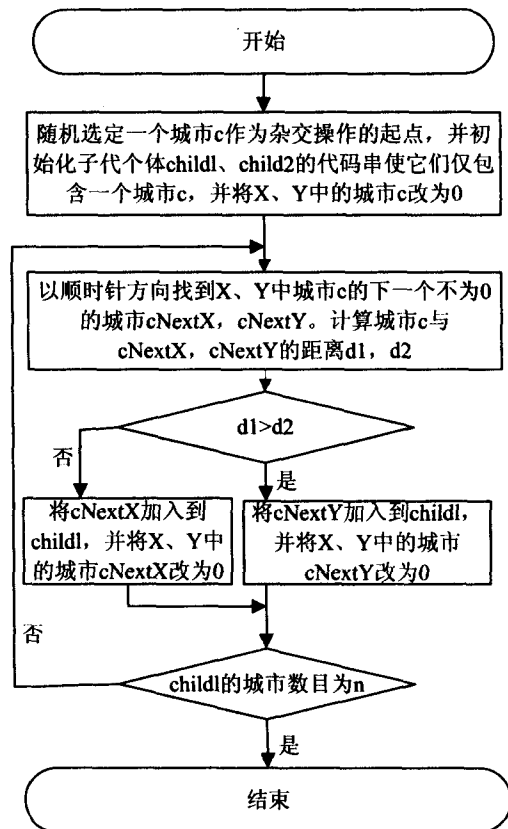


图 1 启发式杂交操作过程

同理,将相关操作改为逆时针查找 c 的下一个城市就可生成另一个子代个体 child2,从而产生了两个新个体。

1.3 动态变化的遗传参数

近年来的研究表明,控制参数对遗传算法的性能影响很大。在简单遗传算法中,交叉概率 P_c 、变异概率 P_m 等控制参数在算法执行前就确定了。用传统遗传算法不变的控制参数来使遗传进化得到控制就可能产生“早熟”现象,使得算法的搜索效率降低^[10]。文中为了克服固定交叉变异概率的这个缺点,引入了动态自适应技术来动态的调整 P_c 和 P_m 。具体做法为:当种群趋于收敛时,减小交叉概率、增大变异概率以保持种群多样性;当种群多样性比较大时,增大交叉概率、减小变异概率来增加算法的收敛速度。基于文中提出的评价种群多样性程度的指标 E ,提出如下的自适应调整遗传算法控制参数的新策略,使得交叉概率 P_c 、变异概率 P_m 在进化过程中随着种群多样性 E 的变化而变化,数学描述如下:

$$P_c = \frac{1}{1 + \exp(-k_1 * E)}$$

$$P_m = k_2 \left[1 - \frac{1}{1 + \exp(-k_1 * E)} \right] \quad (5)$$

式中 $k_1, k_2 > 0$, 由于 $E > 0$, 利用该式就可以使交叉概率和变异概率根据种群多样性的具体情况来动态地调整自己的大小, 从而避免算法陷入局部解并提高遗传算法的搜索速度。

1.4 局部搜索技术的改进

为了平衡由种群多样性增加带来的性能下降问题, 文中引入了二分搜索技术^[11]来提高算法的收敛速度。实验表明对种群中的任一条路径, 进行二分操作可有效提高种群的搜索速度。二分局部搜索算法的实现原理如图2所示, 对于一条巡回路径, 从中选择两条不相邻的边, 将这条巡回路径分成两段路径, 通过重组这两段路径, 生成另一条巡回路径。计算新得到的两条路径的长度, 如果新路径长度小于旧的路径, 则用新路径替换旧的路径, 否则不做改变。经过这样的操作以后就使已得路径得到了进一步的优化, 也就达到了提高局部搜索性能的期望。在图2中, 虚线表示在虚线两端的城市之间不存在城市或存在一个或多个城市。如果 $L_{mn} + L_{pq} > L_{mq} + L_{np}$ (L 表示边的长度), 则用新得到的路径代替旧的路径。具体操作是用边 mq 和 np 取代边 mn 与 pq , 并将城市 n 与 q 之间的城市逆序。

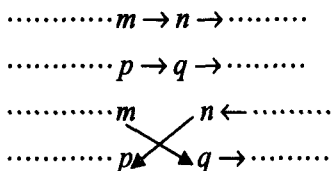


图2 二分优化操作

2 文中改进算法的实现

基于以上提出的对遗传算法的改进方法, 结合经典遗传算法的基本操作, 文中提出的改进遗传算法实现步骤如下:

步骤1: 根据编码机制, 对种群进行初始化;

步骤2: 基于排列编码的基因型多样性计算种群多样性 Δ ;

步骤3: 计算种群中每个染色体的适应度, 根据公式(2)重新计算每个染色体的适应度。再按照适应度大小进行排序;

步骤4: 采用轮盘赌进行选择算子^[12]操作;

步骤5: 根据公式(3)计算 P_c , 并以启发式交叉算子进行杂交操作;

步骤6: 根据公式(3)计算 P_m , 以概率 P_m 进行变异操作;

步骤7: 随机取一定量染色体, 对其进行二分局部搜索操作;

步骤8: 经过上述操作后对新得到的染色体, 计算其适应度值并将得到的染色体都放到种群中。当染色体的个体超过容量时, 若假定染色体的容量一定, 就将适应度小的染色体从池中删去;

步骤9: 判定整个种群是否进化完 M 代遗传操作, 如果满足条件, 则继续向下执行, 否则的话, 跳回到步骤2;

步骤10: 根据进化的结果统计信息, 输出求解结果。

3 实验结果

为了表明算法的有效性, 用文中提出的改进遗传算法(IGA)分别和简单遗传算法(SGA)进行比较。在主频为 2.40GHz 的 PC 机器上, 采用了 Matlab 2010 软件仿真了国际标准数据库 TSPLIB 中的部分算例, 其中算法参数为: 群体规模: $N = 100$; 最大迭代代数: $M = 200$ 。表1是各算法得到的结果。

表1 SGA 和 IGA 的实验结果比较

Problem	Known Best	Algorithm	Best	Ave	Ave Gen
Eil51	426	SGA	493	540	89
		IGA	428	431	56
St70	675	SGA	823	920	89
		IGA	679	685	57
Gr21	2707	SGA	2865	3492.5	58
		IGA	2707	2707	40
Koral100	21282	SGA	21746	221220	80
		IGA	21285	21353	65
burma14	30	SGA	30	31.18	63
		IGA	30	29	97

在实验过程中发现, 用改进的遗传算法求解经典的 TSP 问题实例时, 最优解更接近于现有的最好解, 由于种群多样性的增加, 在有的实例实验中, 文中的改进算法甚至超过了现有的最好解, 尽管需要比较多的进化代数。由于文中引入了多种遗传操作, 使得算法在个体变分散时, 仍然能够较快速地找到最终解, 其中 Eil51 和 Gr21 的进化过程分别如图3和图4所示。

由图3和图4的实验结果可以看出改进遗传算法

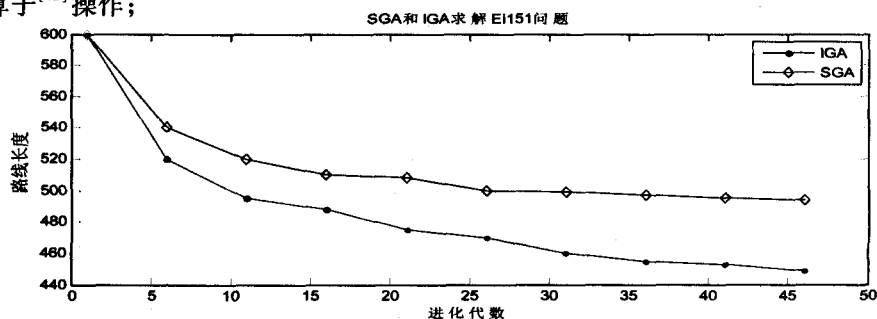


图3 SGA 和 IGA 求解 Eil51 实例的进化图

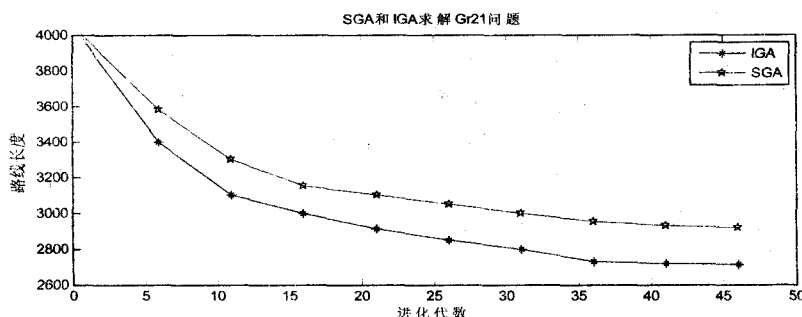


图 4 SGA 和 IGA 求解 Gr21 实例的进化图

无论在全局搜索能力还是收敛速度均好于简单遗传算法。

4 结束语

文中基于自适应调整方法和遗传参数自适应调整方法增强了种群的多样性,引入了启发式交叉算子和二分局部搜索技术来提高种群的收敛速度,克服了简单遗传算法和单亲进化遗传算法种群多样性不佳,易陷入“早熟”现象的缺点,最后文中利用计算机仿真了国际标准数据库中的经典算例,证明了改进算法的有效性。

今后的工作是需要继续探索新的方法来进一步提高算法的性能,并经过实验来验证种群多样性的提高方法。

参考文献:

- [1] 冯春松,王军宇,周松盛,等. TSP 问题的一种改进遗传算法[J]. 武汉理工大学学报,2006(4):116-117.
- [2] 宋海州. 启发式交叉求解 TSP 问题的混合遗传算法[J]. 系统工程学报,2006(12):625-629.

(上接第 50 页)

- and Applications[M]. 2nd ed. New York: Wiley,2005:97-99.
- [2] 赵亚萍. 基于 Visual C#. NET 的空间缓冲区分析开发[J]. 计算机技术与发展,2009,19(12):29-30.
- [3] 宋关福,钟耳顺,王尔琪. WebGIS——基于 Internet 的地理信息系统[J]. 中国图像图形学报,1998,3(3):251-254.
- [4] 袁伟,洪政,魏冬梅. 基于 B/S 模式的 WebGIS 设计与实现[J]. 计算机技术与发展,2008,18(8):8-9.
- [5] 刘立辉. 电子地图应用系统数据分离框架设计[J]. 计算机技术与发展,2010,20(8):203-204.
- [6] 李晓军. GIS 空间分析方法研究[D]. 杭州:浙江大学,2007.
- [7] 胡圣武,王宏涛. 空间关系的研究进展[J]. 测绘科学,2007,32(1):153-156.
- [8] Azevedo L G, Zimbrão G, de Souza J M. Approximate Query Processing in Spatial Databases Using Raster Signatures[C]//

- [3] Bakhquya M, Gaber J. An Immune Inspired - based Optimization Algorithm: Application to the Traveling Salesman Problem[J]. Advanced Modeling and Optimization, 2007, 9: 89-95.
- [4] 马欣,朱双东,杨斐. 旅行商问题的一种改进遗传算法[J]. 计算机仿真,2003,20(4):36-37.
- [5] 吴志远,邵惠鹤,吴新余. 一种新的自适应遗传算法及其在多峰值函数优化中的应用[J]. 控制理论和应用,1999,16(1):127-129.
- [6] Rudolph G. Convergence Analysis of Canonical Genetic Algorithms[J]. IEEE Transactions on Neural Networks, 1994, 5(1):96-101.
- [7] 韩生廉,周文愚,何华. 遗传算法进化状态的测度及非随机性遗传操作方法[J]. 信息与控制,2002(3):223-226.
- [8] 曹道友,程家兴. 基于改进的选择算子和交叉算子的遗传算法[J]. 计算机技术与发展,2010,20(2):45-47.
- [9] Tao Z. TSP Problem Solution Based on Improved Genetic Algorithm[C]//Proceedings of the 2008 Fourth International Conference on Natural Computation. Washington, DC: IEEE Computer Society, 2008:686-690.
- [10] 李航,寇纪淞,李敏强. 参数搜索空间规模对遗传算法稳定性的影响[J]. 系统工程学报,2007(4):162-169.
- [11] Burke K, Kendall G S. A survey and analysis of diversity measures in genetic programming[C]//Proceedings of the Genetic and Evolutionary Computation Conference. [s. l.]: [s. n.], 2002:716-723.
- [12] 汪松泉,程家兴. 遗传算法和模拟退火算法求解 TSP 的性能分析[J]. 计算机技术与发展,2009,19(11):98-100.

- VIII Brazilian Symposium on GeoInformatics. Campos do Jordão, Brazil; [s. n.], 2006:3-17.
- [9] Brinkhoff T, Kriegel H P, Schneider R, et al. Multi-step Processing of Spatial Joins[C]//Proceedings of the 1994 ACM SIGMOD Conference. Minneapolis, USA; [s. n.], 1994.
- [10] Foley J D, van Dam A, Feiner S K, et al. Computer Graphics: Principles and Practice in C[M]. 2nd ed. [s. l.]: Addison-Wesley Professional, 1996:67-72.
- [11] Shemanarev M. Anti-Grain Geometry [EB/OL]. 2002. <http://www.antigrain.com/doc/index.html>.
- [12] Bresenham J E. Algorithms for computer control of a digital plotter[J]. IBM Systems Journal, 1965, 4(1):25-30.
- [13] 董慧,程振林,方金云,等. 一种基于栅格方法的空间谓词判断方法及其系统[C]//2010 中国地理信息产业论坛暨协会年会. 出版地不详:中国地理信息系统协会, 2010.