

文章编号: 1006 2475( 2007) 05 0038 03

# 基于选择性集成的整数编码遗传算法及 TSP问题求解

蒋望东<sup>1</sup>, 林士敏<sup>2</sup>

( 1. 湖南财经高等专科学校信息 管理系, 湖南 长沙 410205 2 广西师范大学计算机科学系, 广西 桂林 541004)

**摘要:** 基于遗传算法的求解 TSP( Traveling Salesman Problem) 研究是近几年的研究热点。设计高效的遗传算法求解, 有重要的理论意义和实用价值。本文考察了基于整数编码的遗传算法的选择算子、交叉算子、变异算子, 运用选择性集成的思想, 将几种算子集成, 随进化的进程对交叉概率和变异概率做自适应调整, 用 Matlab 编写遗传算法程序, 求解中国 31 城市 TSP 问题, 获得了优于目前同类工作的结果。

**关键词:** 整数编码; 遗传算法; 旅行商问题; Matlab 应用; 选择性集成

**中图分类号:** TP311

**文献标识码:** A

## Integer Coding Genetic Algorithm Based on Selective Ensemble and Traveling Salesman Problem Solving

JIANG Wang dong<sup>1</sup>, LN Shim in<sup>2</sup>

( 1. Department of Information Management Hunan Financial &amp; Economic College Changsha 410205 China

2 Computer Science Department Guangxi Normal University Guilin 541004 China)

**Abstract** The research on traveling salesman problem using genetic algorithms is a hotspot in recent years. It is very important in theories and applications to design high quality genetic algorithms programs. This paper investigates select operators, cross operators, mutation operators based on integer coding, and applies them on solving traveling salesman problem with ensemble above operators based on selective ensemble concept. Making cross probability and mutation probability adaptive to evolutional process in programs writing by Matlab, experiment about TSP of Chinese 31 cities shows that this algorithm gets better result than current similar works.

**Key words** integer coding; genetic algorithm; TSP; Matlab applications; selective ensemble

## 0 引 言

旅行商问题 ( Traveling Salesman Problem-TSP) 是一个经典的 NP 完全问题, 只能求近似最优解。近几年的研究热点是基于遗传算法的求解研究。在遗传算法研究中, TSP 问题被用来检验评价不同的遗传选择和操作机制的性能; 在实际应用中, 求解 TSP 问题的有效算法可以用来解决一批同类的组合优化问题, 比如用来求解贝叶斯网络结构学习 K2 算法的节点排序问题。因此, 设计高效的求解 TSP 问题的遗传算法, 有重要的理论意义和实用价值。

TSP 问题的表述很简单: 寻找一条遍历  $n$  个城市的最短路径, 或者说搜索到自然数集  $X = \{1, 2, 3, \dots, n\}$  ( $X$  中元素表示对  $n$  个城市的编号) 的一个全排列  $\pi(X) = \{v_1, v_2, \dots, v_n\}$ , 使

$$T_d = \sum_{i=1}^{n-1} d(v_i, v_{i+1}) + d(v_1, v_n)$$

取最小值, 式中的  $d(v_i, v_{i+1})$  表示城市  $v_i$  到城市  $v_{i+1}$  的距离。

## 1 基于选择性集成的整数编码遗传算法设计

遗传算法的创始人美国著名学者、密西根大学教

收稿日期: 2006 06 08

基金项目: 国家自然科学基金资助项目 ( 60473115)

作者简介: 蒋望东 ( 1971- ), 男, 湖南永州人, 湖南财经高等专科学校信息系讲师, 硕士, 研究方向: 人工智能, 机器学习; 林士敏 ( 1941- ), 男, 广西贵港人, 广西师范大学计算机科学系教授, 硕士生导师, 研究方向: 机器学习, 数据采掘。

授 John H. Holland认为,可以用一组编码来模拟一组计算机程序,并且定义了一个衡量每个“程序”的度量:“适应值”。Holland模拟自然选择机制对这组“程序”进行“进化”,直到最终得到一个正确的“程序”为止。编码方式有:二进制编码、十进制编码和符号编码等方法。整数编码与符号编码一般用于与顺序有关的组合优化方面的问题。根据TSP的特点,本文的工作采用整数编码。染色体长度与TSP路线结点个数相同,染色体的每个基因的编码即为TSP路线结点的编号。因此,每条染色体由1到n的一个全排列组成。

### 1.1 染色体的选择策略

染色体的选择方法一般有:

(1)适应度比例法,又称轮盘赌选择法。即根据适应度函数计算各染色体的适应度值,求出所有染色体的适应度值的总和,计算每个染色体的选择概率,计算每个染色体的累积概率 $q_i$ ,产生随机数 $r$ ,若 $r < q_i$ ,则选择第一个染色体。重复选取,直到选择的染色体总数等于初始染色体总数。显然,在这种染色体选择方法中,适应值越大的染色体即结果越好的染色体被选择的概率也越大,但这种方法容易引起过早收敛。完全依赖概率随机地进行寻优操作虽然可以避免陷入局部极小,但受寻优条件的限制,一般只能得到全局范围内的次优解,很难得到最优解。

(2)最佳个体保留法。即把群体中适应度最高的个体不经过配对交叉直接复制到下一代中,但这种方法的全局搜索能力差,只适用于只有一个峰值的搜索空间,对于具有多个峰值的搜索空间不适用。

(3)竞争选择法。上述两种选择方法各有所长,也各有缺点,差异性较大,依据选择性集成思想<sup>[9]</sup>,表现好的个体学习器越精确,差异越大,集成后可以获得的结果越好。因此,本文先采用适应度比例法进行选择,交叉后产生下一代,再利用最佳个体保留法将上一代的最佳个体直接保存下来,然后从新群体中淘汰一个适应度最差的个体。这种方法集成了上述两种方法的优点并克服了它们的缺点,称之为竞争选择法。

### 1.2 染色体的交叉操作

在采用整数编码的染色体中,交叉操作会形成染色体中的非法基因,即重复基因。所以实现染色体交叉要将重复的基因清除。只使用一种交叉方法容易引起过早收敛,即“早熟”。依据选择性集成思想<sup>[9]</sup>等概率使用以下两种差异性较大的交叉方法,扩大遗传算法的搜索范围,避免过早收敛。

(1)尾部交叉法。预先对染色体设定交叉点位置,只对该交叉点以后的基因片断进行交叉。比较两个染色体的交叉部分,取其中不同的基因按原来的先后顺序分别放入两个数组 $P[]$ 和 $q[]$ 中。在交叉操作前,交叉部分的基因若与数组 $P[]$ 中的基因相同,则先换成 $q[]$ 中对应的基因,反之亦然。交叉

部分不同于数组 $P[]$ 或 $q[]$ 中的基因则无需预先交换。最后进行交叉操作。

例如,有两个染色体为 $\{3\ 2\ 4\ 5\ 9\ 8\ 7\ 6\ 1\}$ , $\{9\ 1\ 3\ 2\ 6\ 5\ 4\ 7\ 8\}$ 。选择第五个位置为交叉点,数组 $P[] = \{6\ 1\}$ , $q[] = \{5\ 4\}$ 。在基因片断 $\{8\ 7\ 6\ 1\}$ 中,将6换为5,1换为4其余不变;在基因片断 $\{5\ 4\ 7\ 8\}$ 中,将5换为6,4换为1。最后将变换后的基因片断进行交叉。交叉后的染色体变为 $\{3\ 2\ 4\ 5\ 9\ 6\ 1\ 7\ 8\}$ 和 $\{9\ 1\ 3\ 2\ 6\ 8\ 7\ 5\ 4\}$ 。

(2)区域交叉法。随机在染色体中选择一个交叉区域,例如 $\{3\ 2\ 4\ 5\ 9\ 8\ 7\ 6\ 1\}$ , $\{9\ 1\ 3\ 2\ 6\ 5\ 4\ 7\ 8\}$ 。将第二条染色体的交叉区域加在第一条染色体的前面,第一条染色体的交叉区域加在第二条染色体的前面,得到: $\{3\ 2\ 6\ 5\ 3\ 2\ 4\ 5\ 9\ 8\ 7\ 6\ 1\}$ , $\{4\ 5\ 9\ 8\ 9\ 1\ 3\ 2\ 6\ 5\ 4\ 7\ 8\}$ ;在交叉区域后依次删除与交叉区域相同的基因,得到最后的两条子染色体为: $\{3\ 2\ 6\ 5\ 4\ 9\ 8\ 7\ 1\}$ , $\{4\ 5\ 9\ 8\ 1\ 3\ 2\ 6\ 7\}$ 。

### 1.3 染色体的变异操作

与交叉方法一样,如果只使用一种变异方法,同样可能会引起“早熟”。为了避免过早收敛,依据选择性集成思想选择以下三种个性好且差异性较大的变异方法,等概率使用以扩大搜索范围。

(1)邻居交换变异。产生一个随机数,将该数对应的基因和其后的基因交换;若该数对应的基因是染色体中的最后一个基因,则将该基因与染色体的第一个基因交换。

如:  $\text{rand\_num} = 3$

染色体:  $3\ 2\ 1\ 4\ 5$       交换结果:  $3\ 2\ 4\ 1\ 5$

(2)区间反转变异。先产生两个不同的随机数,用它们作为边界确定变异区间;然后对个体在变异区间内的基因进行反转。如:

染色体:  $3\ 2\ 1\ 4\ 5$       产生的两个随机数为1和4

变异结果:  $4\ 1\ 2\ 3\ 5$

(3)两点交换变异。先产生两个不同的随机数,确定两个交换点;然后对个体在此两点的基因进行交换。如:

染色体:  $3\ 2\ 1\ 4\ 5$       产生的两个随机数为1和4

变异结果:  $4\ 2\ 1\ 3\ 5$

在标准遗传算法中,变异概率是个常数,经过多次迭代后,种群中的个体适应度容易趋向一致,出现“近亲繁殖”,对群体的进化会产生不利的影响。本文采用自适应变异概率,变异概率随着群体的进化而不断调整。在进化的初期,变异概率较大,随着迭代次数的增加,变异概率逐步减少。交叉概率和变异概率随进化的进程做自适应调整,种群中个体性能提高时交叉概率提高,反之则变异概率增加。

### 1.4 测度函数与适应度的计算

显然,每条染色体对应的路线的距离  $T_d =$

$\sum_{i=1}^{n-1} d(v_i, v_{i+1}) + d(v_n, v_1)$  是其天然的测度函数,因

此,我们就用这个距离函数来计算每个个体对应的适应度。先求所有个体对应的距离值总和,再求每个个体对应的距离值与距离值总和之比,即为每个个体的适应度。若用于其他问题求解,只需要改变测度函数与适应度计算公式。

## 2 基于选择性集成的整数编码遗传算法求解 TSP 问题

### 2.1 实验设计

实验是在 Intel P4 2.0G /256M DDR /Windows 2000 professional /64M AGP 环境下进行的。程序设计工具是 Matlab 6.5 这不同于目前采用 C 语言的流行做法<sup>[1]</sup>。遗传算法诸多算子(如选择、交叉、变异等),都是针对所谓染色体进行的,而染色体实质上是一个向量,可将其看成一个  $1 \times n$  的矩阵,因此这些算子实质上是进行矩阵运算。Matlab 的基本数据单元就是一个维数不加限制的矩阵,用 Matlab 编写遗传算法程序,比用 C 语言等其它高级语言要简单、灵活、快捷,程序篇幅也缩小许多。由于无需用户考虑有关矩阵运算采用何种算法等低层问题,也不必深入了解相应算法的具体细节,因而对用户算法语言方面的要求十分宽松。

为了便于比较,选用解决中国 31 个省会城市和直辖市的 TSP 问题进行实验。首先用 T 记录各城市的相对坐标:

$T = [1304 \ 2312 \ 3639 \ 1315 \ 4177 \ 2244 \ 3712 \ 1399 \ 3488 \ 1535 \ 3326 \ 1556 \ 3238 \ 1229 \ 4196 \ 1044 \ 4312 \ 790 \ 4386 \ 570 \ 3007 \ 1970 \ 2562 \ 1756 \ 2788 \ 1491 \ 2381 \ 1676 \ 1332 \ 695 \ 3715 \ 1678 \ 3918 \ 2179 \ 4061 \ 2370 \ 3780 \ 2212 \ 3676 \ 2578 \ 4029 \ 2838 \ 4263 \ 2931 \ 3429 \ 1908 \ 3507 \ 2376 \ 3394 \ 2643 \ 3439 \ 3201 \ 2935 \ 3240 \ 3140 \ 3550 \ 2545 \ 2357 \ 2778 \ 2826 \ 2370 \ 2975]$

实验步骤如下:

- (1)初始化,随机生成第一代若干个  $1 \sim 31$  的整数编码种群,种群规模取 200
- (2)根据每个个体对应的结点顺序,一一求相邻两点的距离及第一点与最后一点的距离,其和为该个体对应的总距离。
- (3)将本代各个体对应的总距离转换为相对应的适应度值。
- (4)循环对上一代种群依据其对应的适应度进行选择、交叉、变异。初始的交叉概率取 0.8 初始的变异概率取 0.05 循环代数取 2000 代。选择算子选用竞争选择算子,两种交叉和三种变异算子均等概率使用。每循环一次,交叉概率增加  $1/10000$  变异概率减少  $1/10000$  中止条件是直到 2000 代或 200 代内最短距离值无变化。
- (5)输出最短距离对应的 TSP 图。

### 2.2 实验结果分析

目前中国 31 城市 TSP 问题的最好解是通过复杂的改进的模拟退火算法得来的<sup>[3~4]</sup>,如图 1 所示。本算法运行结果如图 4 所示,它比图 1、图 2 所示结果更好,且程序编制相对简单。实验数据如表 1 所示。由于在程序中综合实现了多种选择、交叉和变异算子,使得算法全局寻优能力显著加强,本算法的结果优于用遗传算法求解 TSP 问题的同类工作的结果<sup>[3~8]</sup>,而且程序运行结果稳定,说明本算法的有效性和鲁棒性。

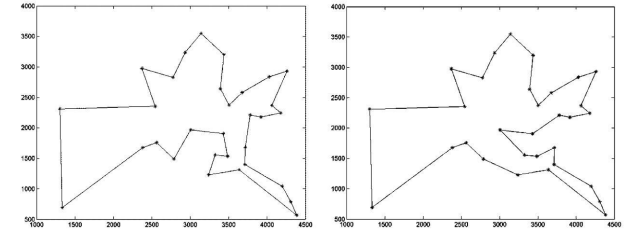


图 1 中国 31 城市的 TSP 文献[3]、[4]结果 图 2 中国 31 城市的 TSP 文献[4]、[7]结果

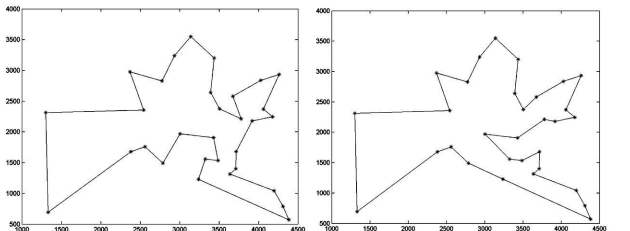


图 3 中国 31 城市的 TSP 文献[5]结果 图 4 中国 31 城市的 TSP 本文算法结果

表 1 结果比较

算 法	文献[3]、[4]	文献[4]、[7]	文献[5]	文献[6]	文献[8]	本文算法
求出最短距离	1598.395	1593.485	1583.777	16379.635	16174.792	15882.544

## 3 结束语

通过将基于整数编码的遗传算法引入到 TSP 问题求解中,运用选择性集成思想,采用多种算子集成混合编程,交叉概率和变异概率随进化的进程做自适应调整,得到优于用遗传算法求解 TSP 问题的同类工作的结果,该算法有较好的有效性和鲁棒性。但对于大规模的 TSP 问题不能盲目地靠扩大种群规模或扩大遗传代数,而应该寻找更好的进化策略,我们将进一步研究。

### 参考文献:

- [1] Katayama K, Sakamoto H, Narihisa H. The efficiency of hybrid mutation genetic algorithm for the travelling salesman problem[J]. Mathematical and Computer Modelling 2000 31(10): 197-203
- [2] Lin Yung-Chien, Hwang Kao-Shing, Wang Feng-Sheng. A mixed encoding scheme of evolutionary algorithms to solve mixed integer nonlinear programming (下转第 43 页)

时,如表 1所示在其它位置用不同的颜色表示是否可调课。拖动完成后,检查是否冲突,是则提示,否则处理排课,并回显调课结果。

(4)排课效率的优化。为了提高排课效率,我们采取了以下措施:

①把所有数据一次性读到内存中处理,从而避免了不停地访问数据库,提高了数据的访问速度,减少了数据存取时间。在数据库中,对所有的字符信息用整数编码;在内存中,直接对编码进行处理,从而提高了处理效率。

②在动态规划阶段,对每个等价类按优先级进行处理,从而减少了回溯的次数。

③在模拟退火阶段,对下次变换课程的选择按一定规则选取,提高了产生优化解的概率。

(5)排课质量的优化。为了自动排课能完成,我们采取了以下措施:

①在排课前,先确保信息的输入完整。用约束条件对基本的数据进行检查,比如一个班级只有 25个课时可以排课,但教学计划排了 28节,这样就无可行解,则要求用户修改基本数据。

②使用了缓冲区技术,在排课时,遇到难排的课表时,先把部分课排在缓冲区中,初始解产生后,再把缓冲区的课调出。

③排课时,数据处理在内存中完成,存储空间动态分配,因此排课时不受班级数、教师数、学科数、上课天数和节数的限制,与全部用数据库来实现的软件相比,增加了使用的灵活性。

3.3 课表的输出

课表的输出提供了可由用户选择的两种方式,一是通过 MFC 的文档 视图类提供的打印框架,实现了软件的打印功能;二是采用 ADO 技术,利用 Word 提供的编程接口,把数据写入 Word 文档,用户在 Word

中编辑、排版、打印。

4 排课效率和质量

根据对 65个客户的回访,效果如表 2所示。

表 2 软件使用情况信息反馈统计表

自动排课时间	1分钟内	1-5分钟	5分钟以上
	51	12	2
自动排课满意度	78.46%	18.46%	3.07%
	非常满意	满意	不满意
	30	34	1
	46.15%	52.30%	1.53%

从表 2可以看出,超过一半的客户自动排课能在 1分钟之内完成,排课效率较高。有 98.45%的客户对本系统满意。

参考文献:

[ 1 ] Siu Cheung Kong, Lan For Kwok. A conceptual model of knowledge based time tabling system [ J ]. Knowledge Based System, 1999, 12( 3): 84-93.

[ 2 ] Masood A. Abdri, Donald L. Davis, Donna F. Davis, John Holling Sworth. A multi objective courses cheduling model combining faculty preferences for courses and times [ J ]. Computer Ops. Res., 1998, 25( 4): 303-316.

[ 3 ] 顾运筠. 遗传算法应用于排课问题中的教师安排最优化[ J ]. 计算机应用与软件, 2006, 23(6): 65~67.

[ 4 ] 程学先, 祝苏薇. 一种基于动态规划的 课程调度算法的研究与实现[ J ]. 武汉理工大学学报( 交通科学与工程版 ), 2006, 30( 3): 485~488.

[ 5 ] 刘继清, 陈传波. 模拟退火算法在排课中的应用[ J ]. 武汉船舶职业技术学院学报, 2003, 9( 3): 22~24.

[ 6 ] 胡顺仁, 邓毅, 王铮. 基于高校排课系统中的图论问题研究[ J ]. 计算机工程与应用, 2002, 4: 221~256.

( 上接第 40 页 ) problem[ J ]. Computers and Mathematics with Applications, 2004, 47( 8-9): 1295-1307.

[ 3 ] 康立山, 谢云, 尤矢勇. 模拟退火算法 [ M ]. 北京: 科学出版社, 1994, 150~151.

[ 4 ] 孙承意, 余雪丽, 王皖贞. 遗传算法求解 TSP 的进化策略 [ J ]. 太原重型机械学院学报, 1996, 17( 2): 128~132.

[ 5 ] 储理才. 基于 MATLAB 的遗传算法程序设计及 TSP 问题求解 [ J ]. 集美大学学报( 自然科学版 ), 2004, 6( 1): 14~19.

[ 6 ] 高尚. 基于 MATLAB 遗传算法优化工具箱的优化计算 [ J ]. 微型电脑应用, 2002, 18( 2): 38~40.

[ 7 ] 朱享荣, 刘伟铭, 宋丹. 改进遗传算法求解 TSP 问题 [ J ]. 株洲工学院学报, 2004, 18( 2): 38~40.

[ 8 ] 傅玉芳. 基于遗传算法求解 TSP 问题的一种新方法 [ J ]. 计算机应用研究, 2004, 21( 9): 45~49.

[ 9 ] 周志华. 选择性集成 ( Selective Ensemble ) [ A ]. 第九届中国机器学习会议( 上海复旦大学 ) [ C ]. 2004.