

一种改进的遗传算法求解旅行商问题

刘荷花¹, 崔超², 陈晶³

(1. 太原大学 计算机系, 山西, 太原 030009; 2. 齐齐哈尔大学 应用技术学院, 黑龙江, 齐齐哈尔 161005;
3. 齐齐哈尔市第一中学, 黑龙江, 齐齐哈尔 161005)

摘要: 针对在解决旅行商问题时标准遗传算法效率不高, 很容易陷入局部最优解的问题, 提出了一种改进的遗传算法. 根据种群个体的多样性和分布情况, 提出了判定遗传算法截止代数的方法. 研究表明, 通过加入了初始化信息, 改进交叉算子, 可提高遗传算法的精确性和收敛性.

关键词: 遗传算法; 旅行商 (TSP); 截止代数; 交叉算子

中图分类号: TP 399 **文献标志码:** A **文章编号:** 1001-0645(2013)04-0390-04

An Improved Genetic Algorithm for Solving Travel Salesman Problem

LIU He-hua¹, CUI Chao², CHEN Jing³

(1. Department of Computer, Taiyuan University, Taiyuan, Shanxi 030009, China;
2. School of Applied Science and Technology, Qiqihar University, Qiqihar, Heilongjiang 161005, China;
3. No. 1 Middle School of Qiqihar, Qiqihar, Heilongjiang 161005, China)

Abstract: Standard genetic algorithm in solving the traveling salesman problem (TSP) is not efficient since it is easy to fall into local optimal solution. To improve the efficiency of genetic algorithm, this paper presents an improved genetic algorithm. First, according to the diversity of individuals and the population distribution, the method to determine the cut-off algebraic of genetic algorithm is proposed. Second, by adding initialization information and improving cross-operator, the accuracy and convergence of the genetic algorithm could be improved.

Key words: genetic algorithm; travelling salesman problem(TSP); end algebra; crossover operator

遗传算法是借鉴生物界自然选择和进化机制发展起来的全局的概率搜索算法, 遗传算法最早由美国的 Holland 教授提出, 起源于对自然和人工自适应系统的研究^[1]. De Jong 基于遗传算法的思想在计算机上进行了大量的纯数值函数优化计算试验^[2]. 在一系列研究工作的基础上, Goldberg 进行归纳总结, 形成了遗传算法的基本框架^[3-4].

在解决旅行商问题时标准遗传算法效率不高, 很容易陷入局部最优解. 作者提出了一种改进的遗传算法. 根据种群个体的多样性和分布情况, 提出了判定遗传算法截止代数的方法.

1 旅行商问题

旅行商问题 (traveling salesman problem, TSP) 可以描述为: 已知 n 个城市之间的距离, 现有一个推销员从某一个城市出发, 必须访遍这 n 个城市, 并且每个城市只能访问 1 次, 最后又必须返回到出发城市, 如何安排他对这些城市的访问次序, 可使其旅行路线的总长度最短.

用图论的术语来说, 假设有一个图 $g=(v, e)$, 其中 v 为顶点集, e 为边集, 设 $d=d_{ij}$ 为由顶点 i 和顶点 j 之间的距离所组成的距离矩阵, 旅行商问题

就是求出一条通过所有顶点且每个顶点只通过1次的具有最短距离的回路. 若对于 n 个城市 $v = \{v_1, v_2, \dots, v_n\}$ 的一个访问顺序为 $t = (t_1, t_2, \dots, t_i, \dots, t_n)$, 其中 $t_i \in v, (i=1, 2, \dots, n)$, 且记 $t_{n+1} = t_1$, 则旅行商问题的数学模型为

$$\min I = \sum_{i=1}^n \text{od}(t(i), t(i+1)), \quad (i=1, 2, \dots, n) \quad (1)$$

TSP问题是十分重要的组合优化问题,它在计算机科学、运筹学及工程等领域都有着广泛的应用. 从理论角度看,不可能设计出理想的能求出最优解的精确算法. 现阶段出现的解决TSP问题的各种优化算法,都是通过牺牲解的精确性来换取较少的耗时. 其他一些启发式的搜索算法则依赖于特定的问题域,缺乏通用性. 相比较而言,遗传算法是一种通用性很好的全局搜索算法.

2 遗传算法

遗传算法GA(genetic algorithm)是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种通过模拟自然进化过程搜索最优解的方法,该算法具有自组织、自适应、自学习和群体进化功能,有很强的解决问题的能力,在许多领域都得到了应用.

2.1 基本运算过程

遗传算法的基本运算过程如下.

① 初始化:设置进化代数计数器 $t=0$,设置最大进化代数 T ,随机生成 M 个个体作为初始群体 $P(0)$.

② 个体评价:计算群体 $P(t)$ 中各个个体的适应度.

③ 选择运算:将选择算子作用于群体. 选择的目的是将优化的个体直接遗传到下一代或通过配对交叉产生新的个体再遗传到下一代. 选择操作建立在群体中个体的适应度评估基础上.

④ 交叉运算:将交叉算子作用于群体. 交叉是指将两个父代个体的部分结构加以替换重组而生成新个体的操作. 遗传算法中起核心作用的就是交叉算子.

⑤ 变异运算:将变异算子作用于群体. 即是对群体中的个体串的某些基因座上的基因值作变动.

群体 $P(t)$ 经过选择、交叉、变异运算之后得到下一代群体 $P(t+1)$.

⑥ 终止条件判断:若 $t=T$,则以进化过程中所得到的具有最大适应度个体作为最优解输出,终止

计算.

2.2 理论基础

遗传算法的理论基础是模式定理,它是Holland为了解释基于二进制编码的标准遗传算法(SGA)而建立的.

遗传算法的群体进化过程,可以视为通过遗传算子不断寻找较好模式的过程. 选择算子对模式的作用在于适应问题解的较好模式被选择的概率大,它在群体中的个体采样数量也相应较大. 好模式上的重要基因或者有效的基因就可以遗传下来. 交叉算子和变异算子在破坏一种模式的同时,产生了一种新的模式,这是群体进化的动力.

定义1 模式:字符串集 $\{0,1\}$ 的字符串,基于三值字符集 $\{0,1,*\}$,能描述某些结构相似性. 其中, $*$ 可以是0或1.

定义2 模式不同,长度相同,模式样本(所匹配串)的个数不一定相同. 引入概念模式阶,来反映这种差异.

模式阶记作 $O(H)$,在模式 H 中,负责确定位置的个数.

定义3 针对模式阶不能反映模式所有性质的特点,引入模式定义距的概念. 定义距为:把模式 H 中取值为0或1的第一个确定位置,与最后一个确定位置之间的距离,记为 $\xi(H)$.

设 $P(t)$ 为第 t 代规模是 n 的群体,在 $P(t)$ 中模式 H 的生存数量为 $m(H,t)$,适应度为 $f(n)$,阶为 $O(H)$,矩为 $\xi(H)$,整个群体的平均适应度为 \bar{f} ,交叉概率和变异概率分别为 P_c 和 P_m ,染色体串的长度为 L .

如果只考虑选择算子对模式 H 生存数量的影响,则有

$$m(H,t) = m(H,0) \times (1 + P_c)^t. \quad (2)$$

综合考虑选择、交叉和变异算子的作用,模式 H 在下一代中的存活量可以表示为

$$m(H,t+1) \geq m(H,t) \frac{f(H)}{\bar{f}} \times \left[1 - P_c \frac{\xi(H)}{L-1} - P_m O(H) \right]. \quad (3)$$

可以看到,在遗传算子选择、交叉、变异的作用下,低阶、低矩和适应度超过群体平均适应度的模式的生存数量,随着迭代次数的增加以指数级增长.

3 改进的遗传算法解决旅行商问题

标准遗传算法的执行效率不高,而且容易在局部最优解出收敛.为了提高搜索的速度和效率,更好地解决 TSP 问题,作者对标准遗传算法进行了改进.

3.1 遗传算法截止代数的判定

3.1.1 信息熵

熵概念来自热力学,它反映了系统的微观混乱程度.

定义 4 已知单符号离散无记忆信源的数学模型为

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{P}(\mathbf{X}) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots & x_n \\ P(x_1) & P(x_2) & \cdots & P(x_i) & \cdots & P(x_n) \end{bmatrix},$$

其中, $0 \leq P(x_i) \leq 1 (i = 0, 1, 2, \dots, n)$ 且 $\sum_i P(x_i) = 1$.

信源各个离散消息的自信息量的数学期望(即概率加权的统计平均值)为信源的平均信息量,一般称为信源的信息熵,也叫信源熵和香农熵,有时称为无条件熵或熵函数,简称为熵,记为 $H(x)$,且

$$H(X) = -C \sum_i P(x_i) \lg P(x_i). \quad (4)$$

式中: C 为正整数,一般取 $C=1$.

在遗传算法的搜索过程中,每一代种群都可以看作是一个单符号离散无记忆信源.根据式(4)可计算出当前代的信息熵,信息熵一旦确定则相应的信源随之确定^[5-6].

只有在变异算子的作用下,遗传算法最终收敛到初始种群中的最有个体,不管初始种群给出何种分布,在变异算子的重复作用下,其极限分布都是均匀的.种群信息熵小时,遗传算法不具备进化能力,算法很难找到更优解.由于变异算子的存在,使种群的信息熵为 0 具有随机性.

3.1.2 遗传算法截止代数的判定

遗传算法的截止代数在信息熵达到某一极小值时可以判断.如果种群方差大、信息熵大,说明种群中个体可能分散在搜索空间的各个角落,这适合于初始种群的情况.如果种群方差大、信息熵小,说明种群中的个体分散在较大的空间中,但个体的多样性弱.种群方差小、信息熵大,说明种群个体集中在较小空间中^[7],但个体的多样性强.经过多代进化,种群收敛,方差和信息熵变小.遗传算法的截止代

数在方差和信息熵小于某一阈值时,可以判断.

3.2 TSP 问题的求解算子

3.2.1 编码方法

现阶段用 GA 解决 TSP 问题时,常见的编码方法如下:

① 顺序表示:每个城市是一个遗传因子,依旅行顺序,把所有要旅行的城市排列成表,旅行每经过一个城市,就去掉这个城市,所有城市处理完后,将表示每个城市的遗传因子连接起来,旅程就形成一条用染色体编码的基因表示.

② 近邻表示:将路径表示为 n 个城市的一个排列,且在第 i 位城市为 j 当且仅当路径中从 i 所到达的下一个城市为 j .如,路径:1-2-4-3-8-5-9-6-7 用排列(2, 4, 8, 3, 9, 7, 1, 5, 6)来表示^[7].如果用这种方法表示,则每一条路径都唯一对应一个近邻表示,然而,任一近邻排列却不一定都对应于合法路径,这样就会产生小圈现象,出现非法路径,加大操作难度.

③ 矩阵表示:用一个 $n \times n$ 二进制矩阵表示一条路径.这种方法最大的缺点是:如果城市规模增加,则导致存储量大幅增加.

④ 路径表示:这种方法符合逻辑,能简单、自然地与表示旅程的基因编码相对应.如初始基因固定,这种编码方式具有唯一性,否则,不具有唯一性.这种方法最符合逻辑,经过每一个节点,而且不产生子回路^[8].本文主要采用路径表示法编码.

3.2.2 改进交叉算子

传统 GA 解决 TSP 问题的交叉算子的设计有部分匹配交叉(PM)、顺序交叉法(OX)、类顺序交叉法(OX)、循环交叉(CX)、单性交叉(asexual crossover)等.在运用遗传交叉算子时,会出现 2 个缺点:基因重复和父代的优秀基因子代不能保留、编码条件需改动子代.

遗传交叉的主要目的是子代尽可能继承父代的优秀基因^[9].为使交叉后的子代保留父代的优秀基因,可用贪心交叉算子加以改动.

贪心交叉算子的基本思想是:首先选择父代的第一个城市,然后在双方父代中对比剩下的城市,选择距离较近的城市,继续城市旅行.如果该城市已在旅行中出现,则选择另一父代的城市,如果两城市都出现过,则随机生成未选择过的城市作为下一个旅行城市.贪心交叉算子充分利用染色体的局部信息指导遗传进化搜索.

循环贪心交叉(GX)可以用以下形式描述:

假设有 n 个城市 $1, 2, \dots, n$; 染色体编码采用路径表示法, 现有待交叉的双亲为

$$\begin{cases} x_1 = (r_{11}, r_{12}, \dots, r_{1n}), \\ x_2 = (r_{21}, r_{22}, \dots, r_{2n}). \end{cases}$$

依据贪婪算法, 将以上染色体视作一个环, 交叉的步骤设计如下.

步骤 1 定当前城市为 r_1 , 将 r_1 加入子代.

步骤 2 比较与 r_1 相邻的两个父个体, 选择距离最近的城市扩展子个体.

步骤 3 被扩展城市如存在子个体中, 扩展城市可定为距离次近的城市. 由此推演, 所有相邻城市如存在子个体中, 则可把一个没有被扩展的城市, 随机加入子个体中.

步骤 4 不断重复步骤 2、步骤 3, 子代则可完整生成.

3.2.3 设计变异算子

TSP 求解时, 具有局部搜索能力的任何算子都是它的变异算子. 与杂交算子相比, 设计变异算子更为灵活.

逆转变异: 在串中随机选择两点, 在将这两点内的子串按反序插入到原来的位置中.

插入变异: 从串中随机选择 1 个码, 将此码插入随机选择的插入点中间.

对换变异: 随机选择串中的两点, 交换其值.

点位变异: 变异仅以一定的概率(通常很小)对串的某些位做值得变异.

4 结 论

对遗传算法提出了改进方案, 加入初始化启发信息, 改进贪心交叉算子. 根据种群的多样性, 以及个体在种群中的分布, 找出了一种判定遗传算法的截止代数的方法. 通过比较发现: 在解决 TSP 问题的领域内, 遗传算法具有较强的寻优能力; 如果在种群初始化时引入启发信息, 则能够提高问题的求解精度和收敛性; 贪心交叉算子改进后, 能较好保留父代的优秀基因, 保证子代的可行性, 提高遗传算法的交叉操作次数和寻优能力.

参考文献:

[1] 邓娟, 陈莘萌. 一种基于最大相似性的 TSP 问题求解算法[J]. 计算机工程, 2006, 9(16): 127 - 129.

Deng Juan, Chen Shenmeng. Greatest similarity based

on TSP problem solving algorithm [J]. Computer Engineering, 2006, 9(16): 127 - 129. (in Chinese)

[2] 高经纬, 张煦, 李峰, 等. 求解 TSP 问题的遗传算法实现[J]. 计算机时代, 2006(3): 35 - 37.

Gao Jingwei, Zhang Xu, Li Feng, et al. Solve the TSP problem using a genetic algorithm [J]. The Computer Age, 2006(3): 35 - 37. (in Chinese)

[3] 谢胜利, 张燕姑, 李广. 基于遗传算法的旅游商问题求解[J]. 温州师范学院学报, 2005(6): 45 - 47.

Xie Shengli, Zhang Yangu, Li Guang. Tourism business based on genetic algorithm problem solving [J]. Wenzhou Normal College, 2005(6): 45 - 47. (in Chinese)

[4] 张春霞, 王蕊. 基于遗传算法求解 TSP 问题的算法设计[J]. 安阳工学院学报, 2008(6): 77 - 79.

Zhang Chunxia, Wang Rui. Genetic algorithm based on TSP problem algorithm design [J]. Anyang Institute of Technology, 2008(6): 77 - 79. (in Chinese)

[5] 穆艳玲, 李学武, 高润泉. 遗传算法解 TSP 问题的并行实现[J]. 北京联合大学学报, 2008(4): 99 - 101.

Mu Yanling, Li Xuewu, Gao Runquan. Genetic algorithm for parallel implementation of TSP problem [J]. Beijing Union University, 2008(4): 99 - 101. (in Chinese)

[6] 杨华芬, 魏延. 一种求解 TSP 问题的改进遗传算法[J]. 重庆工学院学报: 自然科学版, 2009(7): 113 - 115.

Yang Huafen, Wei Yan. Problem solving TSP improved genetic algorithm [J]. Chongqing Institute of Technology: Natural Science, 2009(7): 113 - 115. (in Chinese)

[7] 熊伟清, 郭举良, 魏平. 一种快速求解 TSP 问题的遗传算法[J]. 微电子学与计算机, 2008(2): 135 - 136.

Xiong Weiqing, Guo Juliang, Wei Ping. A fast algorithm for solving TSP problem [J]. Microelectronics and Computer, 2008(2): 135 - 136. (in Chinese)

[8] 敖友云, 迟洪钦. 基于遗传算法求解 TSP 问题的一种算法[J]. 计算机与数字工程, 2008(5): 75 - 77.

Ao Youyun, Chi Hongqin. An algorithm based on genetic algorithm for solving TSP problem [J]. Computer and Digital Engineering, 2008(5): 75 - 77. (in Chinese)

[9] 封全喜, 刘诚. 物流配送车辆路径问题的并行遗传算法研究[J]. 铁道科学与工程学报, 2006(6): 95 - 96.

Feng Quanxi, Liu Cheng. Vehicle routing problem based on genetic algorithm [J]. Railway Science and Engineering, 2006(6): 95 - 96. (in Chinese)

(责任编辑: 刘芳)