

文章编号: 1000-5013(2007) 01-0027-03

TSP 的等价解及其对免疫遗传算法的干扰

杨四海

(华侨大学 信息科学与工程学院, 福建 泉州 362021)

摘要: 以群论的观点考察货郎担问题(TSP) 的等价解. 指出在利用免疫遗传算法求解 TSP 问题时, 采用平均信息量的方法计算亲和度是不合理的, 并且传统的交叉算子存在缺陷. 依据群论的观点, 提出一种新方法计算亲和度, 并提出两种新的交叉算子.

关键词: 免疫遗传算法; 遗传算法; 亲和度; 交叉算子; 货郎担问题

中图分类号: TP 183; O 152 文献标识码: A

TSP 问题是一个典型的非确定的多项式(Nondeterministic Polynomial, NP) 难题, 常被用来评判各种算法的性能^[13]. 利用免疫遗传算法求解 TSP 问题, 其改进之处在于种群多样性的设置, 而抗体之间亲和度的计算是计算种群浓度和多样性的基础. 文[4] 在计算抗体间亲和度时, 采用了计算平均信息量的方法, 它没有考虑 TSP 问题的等价解结构, 因而不合理的. 本文研究如何改进免疫遗传算法的操作策略, 以排除 TSP 问题等价解的干扰.

1 亲和度的计算

在求解 TSP 问题时, 计算两个抗体 u 和 v 之间亲和度的常用方法是利用信息熵的概念计算, 有

$$A_{u,v} = 1/(1 + H(2)).$$

式中, $A_{u,v}$ 为抗体之间亲和度, $H(2)$ 为抗体 u 和 v 的平均信息量. 设有 N 个抗体, 抗体长度为 M , 采用符号集大小为 S , 则

$$H(N) = - \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^S p_{ij} \cdot \log p_{ij}.$$

上式中, p_{ij} 为第 i 个符号出现在基因座 j 上的概率, 定义 p_{ij} 等于在基因座 j 上出现第 i 个符号的总个数与 N 的比值. 利用平均信息量来表征群体的多样性(或反过来表征其近似程度) 是一种统计的观点, 在解空间连续分布时, 可以很好地反映个体之间的差异. 但对于 TSP 问题, 城市编号在某个位置出现并不重要, 重要的是该城市与哪些城市相连接. 因此, 用等位基因出现的概率及平均信息量来表征解的近似程度是不合理的. 例如, 对于 5 个城市的 TSP 问题, 以下 3 组解完全一致. 即 $\{1\ 3\ 2\ 4\ 5\}$, $\{5\ 4\ 2\ 3\ 1\}$ 和 $\{3\ 2\ 4\ 5\ 1\}$ 的亲度应为 1. 然而, 利用平均信息量来计算, 结果却完全不同.

TSP 问题所求解是一个无向连通图, 亦可拓扑变换为正 N 边形的顶点, 其等价解可通过对正 N 边形施加旋转和反射操作得到. 本文仅考虑对称 TSP, 对非对称 TSP 问题, 只需考虑旋转操作. 因此, 可用群论的方法加以处理. 实际上, 求解 N 个城市 TSP 问题所得解的等价解, 可由正 N 边形的二面体群刻画, 所有不同的等价解类构成了 TSP 问题所有可能解的 1 个划分. 设定城市数目为 N , 有 2 个主要结论. (1) 对任一 TSP 问题的解, 在解空间中其等价解的个数为 $2N$. (2) 在解空间中, 所有不同解集的个数为 $N! / 2N = (N - 1)! / 2$. 利用免疫遗传算法计算抗体多样性时, 应在此基础上计算平均信息量.

在计算 2 个抗体间的亲和度时, 同样应考虑旋转和反射操作的影响. 例如对 $\{1\ 5\ 2\ 3\ 4\}$, $\{5\ 4\ 3\ 2\ 1\}$ 2 个解, 尽管它们是不同解, 但都保留了 2-3-4 和 5-1 的连接, 具有一定的亲和度. 旋转操作的影响使 3

收稿日期: 2006-05-11

作者简介: 杨四海(1975-), 男, 讲师, 硕士, 主要从事遗传算法与神经网络的研究. E-mail: ysh_12@163.com.

基金项目: 福建省自然科学基金资助项目(A0540005)

在 2 个解中处于不同位置, 而反射操作的影响使第 1 个解中的 2-3-4 变成了第 2 个解中的 4-3-2, 计算亲和度时, 应能对此加以识别. 考虑到以上情况(对个体 u, v , 城市数目为 N), 构造求解亲和度的算法有 5 个步骤. (1) 查找个体 u 中第 $i=1$ 位置城市编号 j 及其后面城市编号 m . (2) 在 v 中查找城市编号 j 所在位置及其后面城市编号 n , 若 m, n 相等, 亲和度置 1. (3) 取个体 u 第 $i+1$ 位置, 重复步骤(2), 直到 N 为止. (4) 逆序排列 v , 执行前 3 步操作, 累加得到亲和度之和 S . (5) 计算 S/N 作为个体亲和度的值. 超出个体边界时首尾循环补位. 如对于个体 132457689 和 324519867, 两者均保留了 3-2-4-5 及 7-6-8-9-1 的连接, 采用此算法计算, 可得

$1 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 7 \rightarrow 6 \rightarrow 8 \rightarrow 9$	对应位置的亲和度为	$1 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 7 \rightarrow 6 \rightarrow 8 \rightarrow 9$
$3 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 9 \rightarrow 8 \rightarrow 6 \rightarrow 7$		0 1 1 1 0 0 0 0 0
$7 \rightarrow 6 \rightarrow 8 \rightarrow 9 \rightarrow 1 \rightarrow 5 \rightarrow 4 \rightarrow 2 \rightarrow 3$	逆序排列后的亲和度为	0 0 0 0 0 1 1 1 1

两者之和为 7, 于是个体间亲和度数值为 $7/9$. 这和两个个体间相同的局部连接数是一致的.

2 遗传算子

由模式定理可知, 遗传算子起作用的根本点在于优势模式有更多的几率在子代中繁衍. 如果局部的优势模式得不到保持, 遗传算法将退化为随机搜索算法. 对于 TSP 问题, 局部的优势模式可以被认为相邻城市的连接, 故交叉及变异操作应针对城市间的连接. 文[57]介绍了目前常用的几种遗传算子, 从群论的角度看, 这些算子对父个体结构的破坏较大, 故而本文针对 TSP 问题的具体特点使用下述算子.

(1) 选择算子. 为确保收敛, 本文取父个体中较优者及子代中较优者作为下一代的初始个体. (2) 交叉算子. 目前较好的交叉算子^[7](图 1(a)). 取个体 u 的部分位元作为新个体的前一部分, 从个体 v 中找出

剩余位元填入新个体中. 本文采用下述两种算子(设城市数目为 N). (1) 第 1 种算子, 如图 1(b) 所示. (a) 随机选取个体 u 介于 1 到 N 之间的任意 m 个城市编号(即随机选取 m 个位置). (b) 查找个体 v 中此 m 个城市所在位置, 并与个体 u 交换与此 m 个城市相邻的城市编号(个体 u 和 v 中城市编号 3, 7 位置的相邻城市间交换见图 1(b)). (c) 调整个体 v 使之有效解. 本文在实现时每次交换一个城市, 循环执行 m 次. (2) 第 2 种算子, 如图 1(c) 所示. (a) 随机选取个体 u 的部分编码段作为新个体的前一部分(如 53124).

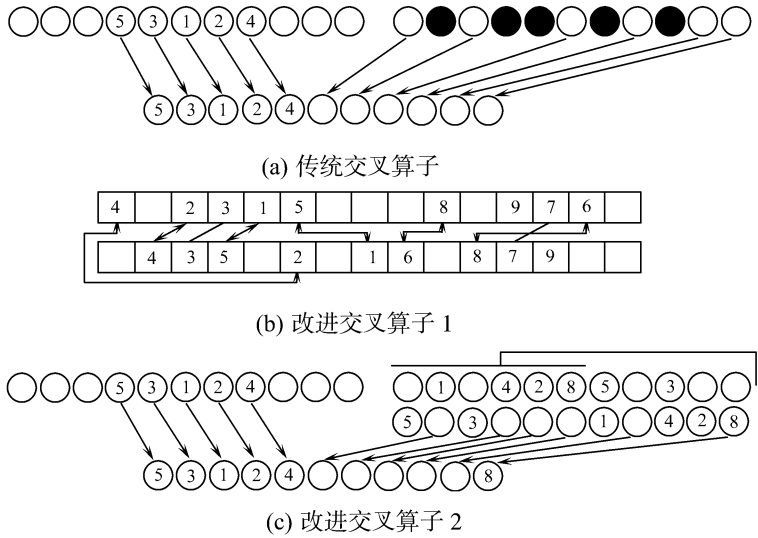


图 1 交叉算子

Fig. 1 Crossover operator

市编号 5, 并判断个体 v 中此城市编号 5 前一位的城市编号 8 是否在个体 u 所选出的编码段中编号 5 的后面. (c) 若 u 选出编码段中编号 8 没在编号 5 的后面(图 1(c)), 则循环移位个体 v 中城市编号 5 至位首, 从个体 v 中找出与 u 中选出的编码段不同的剩余位元填入新个体中. (d) 若 u 选出编码段中编号 8 在编号 5 的后面, 则个体 v 全部编码逆序排列, 然后将城市编号 5 移至个体 v 的位首, 并找出与 u 中选出的编码段不同的剩余位元填入新个体中. 算子 2 基于 u 中城市 5 的部分连接得以保持, 所以希望能在交叉时保持 v 的部分连接. 传统方法的缺陷在于两个父个体即使是同一类解, 子代仍然会不同, 这是不合理的, 也不利于优势个体的保留. 产生这种缺陷的实质在于, 传统方法没有考虑旋转和反射的影响, 而新算子 1 和 2 则考虑了这两种影响. 如对父个体 543219876 和 987654321(两者相同), 若取前者的 24 位作为子代的首部, 传统算子的结果为 432987651 和父代不同, 而新算子的结果则和父代相同.

变异算子采取如下方式^[6]: 随机选定两城市, 其间城市连接全部逆序排列. 对 TSP 问题而言, 这和遗传算法的变异操作是一致的, 其优势在于保持了两城市与其间城市的连接.

3 仿真结果

采用文[8]中数据(取前 3 位), 选取中国的直辖市、省会城市(包括港、澳、台), 计算最短路径(采用交叉算子 1(图 1(b)) 称方法 1, 交叉算子 2(图 1(c)) 称方法 2; 传统交叉算子(图 1(a)) 称方法 3). 两次实验的参数设定, 如表 1 所示.

(1) 种群规模 200, 最大世代数 300, 交叉概率 0.5, 变异概率 0.1/34, 试验次数 5. (2) 种群规模 200, 最大世代数 1 000, 交叉概率 0.5, 变异概率 0.1/34, 试验次数 1. 表中, 使用最佳个体更新次数 N_{br} 和 100 代以后最佳个体的更新次数 N_{100} 作为算子性能的两个评价指标, P_{mb} 为平均最优路径, P_b 为最优路径, G_{mb} 为平均最优代数, P_L 为最终路径, G_b 为最优代数. 仿真结果表明, 相对于传统算法, 新算子具有更好的性能. 这不仅体现在最终结果上, 也体现在最佳个体的更新次数上. 方法 1, 2 的最佳个体更新次数远大于方法 3, 这说明新算子具有更高的局部调整能力, 而这种局部调整能力的获得以较好保持个体结构为基础.

表 1 试验结果
Tab. 1 Test result

方法	300 代 5 次					1 000 代 1 次	
	P_{mb}	P_b	G_{mb}	N_{br}	N_{100}	P_L	G_b
方法 1	17.028	16.47	296.2	52.4	25.0	15.80	709
方法 2	17.104	16.58	298.0	50.4	25.8	15.68	855
方法 3	18.808	17.75	284.2	33.8	12.8	16.03	789

4 结束语

针对 TSP 存在的等价解问题, 提出了新的交叉算子及计算亲和度的方法. 试验结果表明, 该方法提高了免疫遗传算法求解 TSP 问题的整体性能, 对于有等价解存在的其他问题, 具有参考价值.

参考文献:

[1] 马良. 旅行推销员问题的算法综述[J]. 数学的实践与认识, 2000, 30(2): 156-165.
[2] 王熙法, 张显俊, 曹先彬, 等. 一种基于免疫原理的遗传算法[J]. 小型微型计算机系统, 1999, 20(2): 117-120.
[3] 邹鹏, 周智, 陈国良, 等. 求解 TSP 问题的多极归约算法[J]. 软件学报, 2003, 14(1): 35-42.
[4] 莫宏伟. 人工免疫系统原理与应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 2003: 1-30.
[5] 刘勇, 康立山, 陈毓屏. 非数值并行算法——遗传算法[M]. 北京: 科学出版社, 1995: 165-170.
[6] 陈国良, 王熙法, 庄镇泉, 等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996: 81-97.
[7] 冯文镛, 杨灿军, 陈鹰. ACR 原型系统的全局路径规划遗传算法研究[J]. 控制理论与应用, 2002, 19(2): 282-286.
[8] 康立山, 谢云, 尤矢勇, 等. 非数值并行算法——模拟退火算法[M]. 北京: 科学出版社, 1994: 150-151.

Equivalent Solutions of TSP and Its Interference
on Immune- Genetic Algorithm
YANG Si-hai

(College of Information Science and Engineering, Huaqiao University, 362021, Quanzhou, China)

Abstract: In this paper the equivalent solutions of TSP was analyzed by using group theory. We point out that to solve TSP problem by using immune genetic algorithm, it is unreasonable to compute genetic affinity using entropy, although the classic crossover operators exist some defects. Base on the group theory, a new method to calculate affinity and two novel crossover operators was proposed.

Keywords: immune genetic algorithm; genetic algorithm; affinity; crossover operator; traveling salesman problem

(责任编辑: 黄仲一)