

文章编号:1671-6833(2002)02-0009-04

基于遗传算法的最小生成树的参数优化研究

周荣敏, 雷延峰

(郑州大学环境与水利学院, 河南 郑州, 450002)

摘要:应用遗传算法求最小生成树问题是一种可行而有效的新方法. 为了提高算法的进化效率和收敛性, 通过数值模拟试验, 对算法中主要控制参数的优化组合模式进行研究, 获得了主要参数的优化组合规律: 群体规模取10~30; 换位率在0.5~0.9之间选取; 选择率可采用自适应方式确定, 进化初期采用较大值, 随着进化过程的前进可逐渐减小; 编码长度不应过长等.

关键词:遗传算法; 最小生成树算法; 参数优化

中图分类号: O 157.6 **文献标识码:** A

1 基于遗传算法的最小生成树算法

最小生成树算法在工程实际中应用比较广泛, 许多工程实际问题, 如局域网布局、通讯网络设计、电路分析设计、最短线路连接、给排水系统的优化布局等, 常可转化为求最小生成树或最小支撑树(简称最小树)问题. 图论中已有一些经典的最小树算法, 如Kruskal 算法、Prim 算法、Sollon 算法等^[1,2], 但是这些算法一般仅能得到一个最小生成树. 对许多实际问题而言, 不同形式的最小生成树代表不同的设计方案, 有时权值次小的生成树, 也可能是实际问题的一个较好的设计方案, 如果任选其中的一个, 将会丢弃或忽略更优的方案. 为了能够在较短时间内, 以较高的概率获得一组最小树或次最小树, 作为进行实际问题优化方案评价和选择的依据, 作者提出了一种基于遗传算法的最小生成树算法^[3](Minimum Spanning Tree Algorithms based on Genetic Algorithms, 简称MST-GA), 并通过实例应用研究证明了应用该算法求最小树问题是可行和有效的.

应用遗传算法求解最小生成树问题可按照如下的基本步骤进行: ①将待求解的问题转化为一个求网络图的最小生成树问题; ②采用二进制编码串表示网络图的生成树; ③定义最小生成树问题的适应度函数; ④以单亲换位算子和单亲逆转算子作为主要的遗传运算, 对遗传群体实施遗传操作; ⑤在遗传进化过程中综合采用代间竞争和群体单一化策略、平等选择和优先选择相结合的

混合选择机制、单亲换位算子和逆转算子的随机执行策略等进化控制策略, 以提高遗传算法的进化效率和收敛性, 使进化过程不断向理想的进化方向前进; ⑥从最终的遗传群体中选取若干个最优体(即最小生成树)作为所研究问题的优化方案, 结合其它因素进行综合评价和决策.

2 MST-GA 的参数优化配置

MST-GA 所涉及参数较多, 如最大遗传代数、群体规模、选择率、换位率和编码长度等. 不同的参数组合模式将直接影响算法的性能和进化效率. 为了提高算法的效率和实用性, 应通过数值模拟试验进行参数优化配置研究, 寻找各参数之间的组合规律, 为实际应用提供合理的参数选择依据.

以图1所示的网络图为例, 进行参数优化配置研究. 图中共有10个节点, 23条边, 括号内的数值为各边的权值. 计算机系统的配置为Pentium III, 64 M 内存; 中文Windows 98 操作系统; 所有程序采用Matlab 5.2 编写.

2.1 最大遗传代数

遗传算法主要通过遗传群体的反复迭代搜索全局最优解. 一般采用最大遗传代数作为遗传进化的终止条件, 即当进化过程达到预定的最大遗传代数时终止算法, 防止陷入无穷死循环, 浪费计算时间. 最大遗传代数可根据算法的进化速度和群体收敛性确定, 通常从一个较小的最大遗传代数开始运行, 通过观察遗传进化的速度和进化结

收稿日期: 2002-01-15; 修订日期: 2002-02-20

作者简介: 周荣敏(1971-), 女, 河南省新野县人, 郑州大学讲师, 博士, 主要从事遗传算法及其应用研究.

果的收敛程度,逐渐增大进化代数直至获得一组满意的稳定解.一般可取最大遗传代数为100~1000,甚至更大.另外,可在程序中设置适当的中断控制条件,当发现进化过程已进入成熟收敛状态时,提前中止算法,避免浪费计算时间.

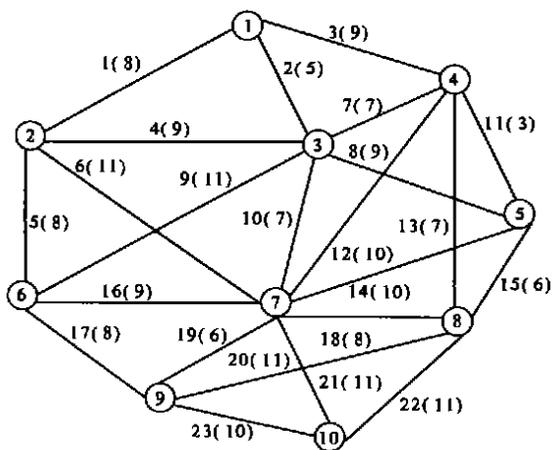


图1 网络图F

Fig.1 Graph F

2.2 群体规模

群体规模是指每一代遗传群体中个体的数目.群体规模影响遗传算法的最终结果和执行效率.群体规模较小时,遗传算法的优化性能一般不会太好,容易引起未成熟早收敛的现象;群体规模较大时,可产生足够多的遗传个体,使遗传算法从多个可行解点出发搜索最优解,减少遗传算法陷入局部最优解的机率,但是,较大的群体规模将导致计算复杂程度增加,影响算法的计算效率.

在保持最大遗传代数 $Maxgen = 100$ 、换位率 $P_t = 0.5$ 和选择率 $P_s = 0.5$ 的情况下,分别以不同的群体规模执行一次遗传进化过程.研究表明:在最大遗传代数、换位率和选择率相同的情况下,程序运行时间随着群体规模的增大而急剧增长,如图2所示.因此,在遗传进化过程中,群体规模不宜过大.在实际应用中,应根据问题的性质,选择合适的群体规模.对于一般的最小生成树问题,一次获得5~10个最优方案就可以满足方案评价的需要.应用MST-GA求解最小生成树问题时,建议将群体规模控制在10~30之间,避免因群体规模太大而造成程序运行时间过长,影响算法的计算效率.

2.3 选择率和换位率

选择率和换位率是影响MST-GA性能的主要参数,这两个参数均在算法的初始化阶段确定.选择率(P_s)是用来控制遗传进化过程产生每一代

子代群体时执行平等选择机制和优先选择机制的概率;换位率(P_t)是用来控制在遗传进化过程中产生每一个子代个体执行换位算子和逆转算子的概率.

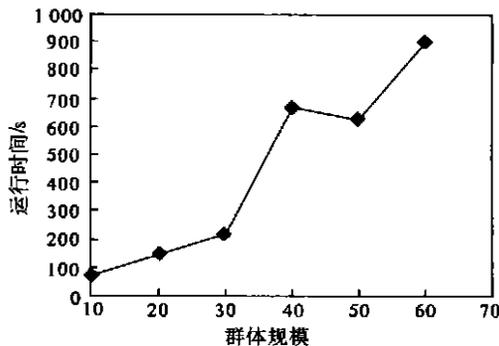


图2 群体规模与运行时间的关系

Fig.2 The relation between population size and run time

在遗传进化过程中,产生每一代子代群体所采用的选择机制由一个0~1之间的随机数 R_s 决定.如果 $R_s \geq P_s$,按优先选择机制选择产生子代的母体,否则按平等选择机制选择产生子代的母体.对所选择的母体执行的遗传运算由一个0~1之间的随机数 R_t 决定.如果 $R_t \geq P_t$,对所选择的母体执行单亲换位算子产生新的子代个体,否则执行单亲逆转算子产生新的子代个体.

选择率和换位率对遗传算法性能的影响可以从计算时间和群体平均适应度的变化情况两个方面评价.在保持群体规模和最大遗传代数不变的情况下,以不同的选择率和换位率组合模式执行遗传进化计算.

从程序运行时间的变化趋势可以看出,当换位率一定时,选择率的变化使算法运行速度不稳定,程序运行时间在一定的范围内波动.随着换位率的逐渐增大,程序运行速度加快,运行时间显著减小,如图3所示.

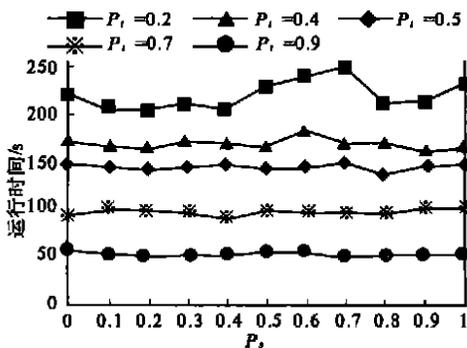


图3 不同 P_s 和 P_t 组合情况下程序运行时间的变化

Fig.3 The program run time under different combinations of P_s and P_t

从群体平均适应度的变化情况可以看出,当换位率一定,选择率变化时,随着遗传进化过程的进行,群体平均适应度都能不断改善提高,并且逐渐收敛到一个特定的平均适应度范围,如图4所示.相比较而言,选择率较小时,群体的整体进化性能较好;选择率较大时,群体的初期进化速度较快.这主要是因为,在进化过程中,当选择率较小时,即优先选择机制占优势时,可以使每一代群体中适应度较高、性能优良的个体,有更多的机会被选择为母体,有更高的概率产生后代个体,并将其优良特性遗传到下一代,保证群体整体性能向好的方向发展.当选择率较大时,即平等选择机制占优势时,在进化的初期阶段可以从尽可能多的点出发进行优化搜索,搜索的空间范围较大,使群体的平均适应度提高较快.当整个遗传群体进化到一定程度时,群体平均适应度基本稳定,不同进化群体中个体性能的差异较小,选择机制对进化性能的影响相对减弱,进化速度趋于平稳.

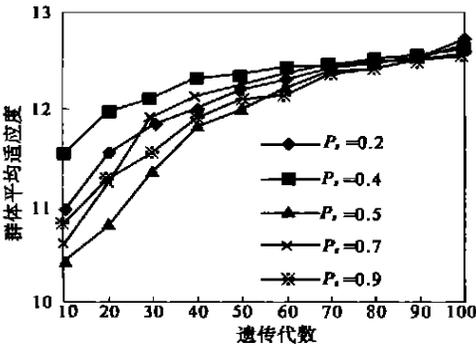


图4 群体平均适应度的变化 ($P_t = 0.5$)

Fig. 4 The change of population average fitness ($P_t = 0.5$)

由图5可以看出,在选择率一定的情况下,较大的换位率使遗传群体呈现出较好的收敛性能.这主要是因为,在子代群体的生成过程中,当换位率较大时,子代个体的产生主要通过对其亲体的某个基因段进行逆转算子产生后代个体.逆转算子不仅执行速度快,而且能更好地继承亲体中高性能的基因成分,从而使进化效率较高.当换位率较小时,换位算子在遗传操作中占绝对优势,对亲体基因链执行频繁的基因位对换,容易破坏亲体中性能较好的基因段,无法保证子代个体具有比亲体更优良的特性,导致整个进化过程非常缓慢.因此,在遗传进化过程中,采用较大的换位率对提高算法运行速度非常有效.

根据选择率和换位率对遗传算法进化性能的影响,在进行实际问题的求解时,采用较小的选择

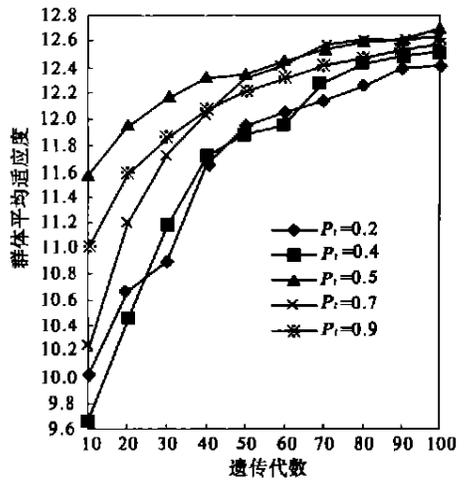


图5 群体平均适应度的变化 ($P_s = 0.3$)

Fig. 5 The change of population average fitness ($P_s = 0.3$)

率和较大的换位率是比较合理的参数组合模式.推荐采用的换位率取值范围为0.5~0.9.选择率的取值可以采用自适应方式确定,使其在进化初期阶段具有较大的选择率,如在0.6~0.9之间变化,充分利用平等选择机制的优势,扩大初期的搜索范围.随着进化过程的前进,可自适应地减小选择率,如在0.2~0.5之间变化,充分利用优先选择机制的优势,使性能优良的个体具有更多的机会产生其子代个体,改善群体的进化性能.

2.4 编码长度

在MST-GA中,个体编码长度由网络图中候选边的数目 m 决定.对一个有 n 个节点的最小树的问题,基因链中只有 $n-1$ 个有效基因位,其对应于网络图中的 $n-1$ 条边.在长度为 m 的二进制编码串中, $n-1$ 个基因位为1的可能组合方案数目为

$$C_m^{n-1} = \frac{m!}{(m-n+1)! (n-1)!}$$

由图论原理可知,过 n 个顶点的树的数目为 n^{n-2} 个;有 n 个顶点的完全连接图的边数为 $n \times (n-1) / 2$ 个,即网络图中候选边数目 m 最大为 $n \times (n-1) / 2$ 个.当过 n 个顶点的树的数目一定时,候选边数 m 越大,可能的组合方案数目就越大,有效可行解(生成树)所占的概率就越低,影响算法的搜索效率.

例如,算例中的网络图共有10个节点,23条候选边,即 $n=10, m=23$.过10个顶点的树的数目为 10^8 .若用枚举法从 10^8 个树中寻找权最小的树,计算量很大.当 m 取最大可能边数目,即 $m=10 \times 9 / 2 = 45$ 时,个体编码串的所有可能组合方案

数目为: $C(45, 9) = 45! / 36! / 9! = 886, 163, 135$. 当 $m=23$ 时, 个体编码串的所有可能组合方案数为: $C(23, 9) = 23! / 14! / 9! = 817, 190$, 这时所要搜索的解空间由 886163135 个可能组合方案缩小为 817190 个可能组合方案, 仅为最大解空间的 0.9%. 因此, 在确定初始网络图时, 必须根据问题的实际情况, 尽可能排除性能较差的连线(边), 减少候选边数目, 增加组合方案中可行方案的比例, 从而减小问题规模和需要搜索的解空间, 提高算法效率.

3 结束语

通过大量的数值模拟试验, 对基于遗传算法的最小生成树算法的参数优化配置进行了深入细

致的研究, 获得了有关群体规模、最大遗传代数、换位率、选择率、编码长度等参数的组合规律和合理的取值范围, 为实际应用提供了可靠的参考依据. 在实际应用中可以根据所求解问题性质的不同, 选择适当的参数组合模式, 以提高算法的计算效率、收敛性能和实用性.

参考文献:

- [1] 谢金星, 邢文训. 网络优化[M]. 北京: 清华大学出版社, 2000.
- [2] 卢开澄, 卢华明. 图论及其应用[M]. 第二版. 北京: 清华大学出版社, 1995.
- [3] 周荣敏, 买文宁, 雷延峰. 基于遗传算法的最小生成树算法[J]. 郑州工业大学学报, 2002, 23(1): 45-48.

Study on the Parameters Optimization of Minimum Spanning Tree Algorithms Based on Genetic Algorithms

ZHOU Rong-min, LEI Yan-feng

(College of Environmental & Hydraulic Engineering, Zhengzhou University, Zhengzhou 450002, China)

Abstract: It's a new feasible and efficient method to solve the minimum spanning tree problems by genetic algorithms. In order to improve the evolutionary efficiency and convergence of the algorithms this paper makes a thorough research to the combination model of the main controlling parameters of the algorithms by numerical simulation tests. According to the results of the research, the rules of optimal parameter combinations are acquired. The population size can be chosen between 10 and 30, the transpose probability (P_t) can be chosen between 0.5 and 0.9, the select probability (P_s) can be choosed by self adapting method. P_s can be larger at the beginning of the process of evolutionary, and then becomes smaller as the evolutionary process goes ahead, and the coding length shouldn't be too long. Those rules provide a valuable reference for the application of the algorithms.

Key words: genetic algorithms; minimum spanning tree; parameter optimization