

文章编号:1007-6492(2002)01-0045-04

# 基于遗传算法的最小生成树算法

周荣敏, 买文宁, 雷延峰

(郑州大学环境与水利学院, 河南 郑州 450002)

**摘要:**以图论和遗传算法为基础,提出了一种求最小生成树的改进遗传算法.该算法采用二进制编码表示最小树问题,用深度优先搜索算法进行图的连通性判断,并设计出相应的适应度函数、单亲换位算子和单亲逆转算子以及四种控制性进化策略,以提高算法执行速度和进化效率.与 Kruskal 算法相比,该算法能在一次遗传进化过程中获得一批最小生成树,适合于解决不同类型的最小树问题.

**关键词:**遗传算法;最小生成树;进化策略;网络优化

**中图分类号:**O 157.6

**文献标识码:**A

树是图论的重要概念之一,也是图论中结构简单、用途广泛的一种连通图.许多工程问题,如通讯网络设计、最短路连接,以及电力、煤气和排水系统设计等,通常都可以转化为求最小生成树问题.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<sup>[1,2]</sup>,但这些算法均属于贪心算法,一般只能得到一个最小生成树,而在实际应用中通常需要同时找出一组最小或次最小树作为方案评价和选择的依据.

遗传算法<sup>[3]</sup>是一种模拟生物进化过程的新兴全局优化算法,已在组合优化、模式识别、神经网络、经济预测等领域广泛应用.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,本文提出一种基于遗传算法的最小生成树算法,并通过实例应用证明算法的可行性和有效性.

## 1 遗传算法

遗传算法的基本思想来源于遗传进化,主要是借助于生物进化机制与遗传学原理,按照自然选择和适者生存的原则,利用简单的编码技术和繁殖机制,模拟自然界生物群体优胜劣汰的进化过程,实现对复杂问题的求解.应用遗传算法进行问题求解时,首先将待研究问题的解用某种编码方式(如二进制等)表示成染色体,并随机生成初始群体,然后以群体中个体的适应度为依据,对个体执行选择、交叉和变异三种基本遗传算子,使群体中的个体(问题的解)不断进化并逐渐逼近问题

的最优解.

在标准遗传算法中,选择算子的作用是将群体中优良个体(较优解)直接遗传到下一代,或通过交叉和变异遗传到下一代.目前,常用的选择算子有适应度比例法、最佳个体保留法、期望值法等.不同方法对算法的性能影响不同,具体应用时可根据问题求解特点,采用单一或组合方法.

交叉算子按一定概率随机从亲代群体中选择两个个体,随机将两个亲代个体的部分结构相互交换,生成两个新的子代个体.交叉算子可以采用单点交叉、两点交叉、均匀交叉、多点交叉等多种方式.交叉算子产生的两个子代个体,都包含两个亲代个体的遗传物质,但与亲代个体不同.因此,交叉算子能提高遗传算法的搜索能力,在适当选择策略下,通过交叉算子可实现向最优解的收敛.

变异算子以一个很小的随机概率改变个体字符串上的某些位.对于二进制编码,就是将某一位从 0 变为 1 或由 1 变为 0.变异算子具有增加和恢复群体多样性的功能,使遗传算法能进行非线性局部搜索.

## 2 应用遗传算法求最小生成树的基本原理

在图论中,一个无回路(圈)的连通子图称为树.设  $T=(N, E_T)$  是  $|N| \geq 3$  的一个图,则下列关于树的 6 个定义是等价的:①  $T$  连通且无回路;②  $T$  有  $|N|-1$  条边且无回路;③  $T$  连通且有  $|N|-1$  条边;④  $T$  连通且每条边都是割边;⑤  $T$  的任两

收稿日期:2001-10-11;修订日期:2001-12-25

作者简介:周荣敏(1971-),女,河南省新野县人,郑州大学讲师,博士,主要从事遗传算法最优化理论及应用研究.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

点间都有唯一的路相连;⑥ $T$  无回路,但在任一对不相邻的点之间加连一条边则构成唯一的回路.

设有一无向图  $G=(N, E, W)$ , 其中  $W=\sum_{e \in E} w_e$  为权函数,若树  $T=(N, E_T, W_T)$  包含了图  $G$  的所有顶点,则树  $T$  为  $G$  的一个生成树或支撑树,其中  $W_T=\sum_{e \in E_T} w_e$  为树  $T$  的权.图  $G$  的生成树不唯一,权最小的生成树称为  $G$  的最小生成树.

图论中经典的最小树算法,一般情况下仅得到相同且唯一的最小生成树.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,可能是一个较好设计方案,如果任选其一,可能会丢弃或忽略更优的方案.因此,在实际应用中希望能同时获得若干个权最小或次小的生成树作为待选方案.对于一个完全图  $G$ ,过  $n$  个顶点的树的数目为  $n^{n-2}$ .若用枚举法从  $n^{n-2}$  个树中寻找最适合实际问题的最小生成树,计算量非常大,有必要寻找更有效的算法,在较短的时间内,以较高概率确定出一批最小树和次最小树.

考虑如下一个寻找最小树的过程:根据图论中树的性质可知,一棵过  $n$  个顶点的树必有  $n-1$  条边,且是连通的.对于一个有  $ND$  个顶点、 $NP$  条边的图  $G$  来说,从  $NP$  条边中随机选取  $ND-1$  条边,则可构成一个子图.利用深度优先搜索算法(DFS)对该子图进行一次搜索,如果能搜索到  $ND$  个顶点,则这  $ND-1$  条边所构成的连接子图是一个连通图.过  $ND$  个顶点且有  $ND-1$  条边的连通图一定是该图的一棵生成树,即所选择的  $ND-1$  条边构成一个生成树.如果能采用一定的优化技术,搜索并评价尽可能多的生成树,就可以找出一组权值最小或次最小的生成树.

研究表明,遗传算法对组合优化问题的求解非常有效,能够在较短时间内,以较高的概率获得一批最优或次最优方案.上述寻找最小生成树的过程实际上也是一个典型的组合优化问题.根据遗传算法的基本原理,结合最小生成树的特点,可按照以下过程寻找最小生成树:首先选择一种合适的编码方式表示最小树问题,然后直接以树的权值最小为优化目标函数,从一组随机产生的初始解出发,以遗传算法控制优化搜索过程,通过不断搜索并评价可行的生成树,逐渐进化到一组权值最小的生成树,实现最小树的优化求解.

### 3 算法设计

#### 3.1 编码设计

设某一无向图中共有  $ND$  个顶点、 $NP$  条边,对图中的节点和边进行编号.以图中的所有边为编码变量,各个编码变量的取值为 0 或 1,则用一个长度为  $NP$  的二进制字符串可表示该图的所有子图.当字符串中某位上的字符值为 1 时,表示它所对应的一条边是构成子图的边;当字符值为 0 时,表示它所对应的边不是构成子图的边.例如,图 1(a) 所示的无向图中有 5 个节点,8 条边,图 1(b) 为该图一个生成树,由边 1,3,5 和 7 组成,则该生成树所对应的二进制编码为:  $\{10101010\}$ .

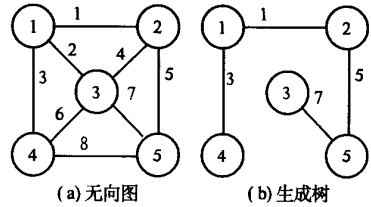


图 1 无向图和生成树

Fig. 1 Undirected graph and spanning tree

#### 3.2 适应度函数设计

遗传算法在进化搜索过程中只要求目标函数是非负的最大值形式,而对目标函数的定义域、连续性、可微性等没有限制.在具体应用中可根据问题的性质,采取一定的转化策略,将优化问题的目标函数映射成求非负的最大值形式,通常将转换后的目标函数称为“适应度函数”.

设某一无向图中各边的权值为  $w_i(i=1, \dots, NP)$ ,对于求权最小的生成树问题,可定义其适应度函数为

$$F=\begin{cases} F_0-\sum_{i=1}^{NP}w_iZ_i & \{Z_i \in \{0,1\} \mid i=1,\dots,NP\} \text{ 是生成树;} \\ 0 & (\text{其它}), \end{cases}$$

式中,  $F_0=(ND-1)\max(w_1, w_2, \dots, w_{NP})$  是一个较大的正常数,保证个体的适应度  $F$  总为非负;  $w_i$  为第  $i$  条边的权值;  $Z_i$  为个体编码中第  $i$  位上的字符值(对应于图中的第  $i$  条边),为 0 或 1.

在遗传进化过程中,必须对每一个个体所代表的子图进行一次深度优先搜索,若能搜索到所有的顶点,则说明该个体是图的一个生成树,可所定义的适应度函数表达式计算该个体的适应度值;否则令该个体的适应度值为零,使其在进化过程中的生存能力最低,逐渐从群体中淘汰出去.

#### 3.3 遗传算子设计

根据生成树的特点可知,在长度为  $NP$  的字符串中,必须有  $ND-1$  个字符位上的字符值为 1,才有可能是一个生成树,有小于或大于  $ND-1$  个字符值为 1 的个体必定不是一个生成树.为了减

少进化过程中不可行方案的产生,必须控制所产生的每个个体只有  $ND-1$  个字符值为 1,使其具备成为可行解的必要条件,然后通过连通性检验确定是否为一个生成树.

标准遗传算法主要通过交叉算子和变异算子产生新的子代个体.对于最小生成树问题,交叉算子和变异算子容易破坏子代成为可行解的特性,即具有  $ND-1$  个字符值为 1,所产生的新个体往往不能构成有效的生成树,降低了遗传算法搜索最小树的能力.为提高遗传算法搜索最小树的效率,结合生成树的图论特性,本算法放弃了标准遗传算法的交叉算子和变异算子,设计出新的单亲换位算子和逆转算子.

单亲换位算子对所选择个体的基因链上的任意一对基因进行交换,且所执行的基因交换次数以及被交换的基因位都是随机产生的.

个体 A  $10101010 \xrightarrow{\text{一次随机交换}} 11101000$  个体 A'

单亲换位算子能使任何一个母体通过有限次的基因换位生成一个新个体.通过这种基因重组方式可以从一组群体出发,以较高的概率搜索到解空间的各个可行解,能有效避免遗传过程中无效个体的产生.

单亲逆转算子则是将母体基因链上的任意一段基因进行逆转,一次性地完成从一个母体突变为一个新子体.与换位算子相比,逆转算子的执行速度较快,有助于将母体中未发生突变的有效基因段,直接遗传到子体中.

个体 A  $10101010 \xrightarrow{\text{基因段逆转}} 10010110$  个体 A'

改进后的遗传算法具有一个突出特点:在子代群体的生成过程中,每个子体只有一个母体,通过对母体执行换位算子或逆转算子,产生出具有不同性状的新个体.单亲换位算子和逆转算子不仅可以保证子代个体具有成为可行解的基本特性,而且可以提高对解空间的搜索能力.

3.4 进化策略设计

为使遗传进化过程能够不断向理想的优化方向前进,本算法综合采用如下控制性进化策略<sup>[4]</sup>.

3.4.1 代间竞争与群体单一化策略

为保证群体性能的优越性,采用代间竞争机制,即在保持群体规模不变的情况下,亲代和子代一起进行生存竞争,选择最优的若干个体构成新一代群体.同时,为了能在进化过程结束时得到一个具有多样性的有效群体,防止所有个体陷入同一极值而停止进化,算法中实施了群体单一

化策略,即不允许群体中有相同个体出现,保证群体中个体的唯一性和多样性.代间竞争机制和单一化策略的有机结合,既能保持群体多样性,又能保证群体性能不断提高,最终得到一个最优群体.

3.4.2 优先和相等相结合的混合选择机制

优先选择机制是按照优胜劣汰的原则,根据个体适应度的高低选择执行遗传运算的母体.适应度高的个体有更多的机会产生后代,优良特性得以繁衍;适应度低的个体产生后代的机会相对较低,并逐渐被淘汰.平等选择机制是指亲代群体中每个个体都有相同的机会产生自己的后代.在遗传算法的初始化阶段,首先确定一个选择机制的执行概率即选择率  $P_s$ ,每一代所采用的选择机制由随机数  $R_s$  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.这种混合选择机制,有利于扩大搜索空间,提高单亲遗传算法的全局寻优能力.

3.4.3 单亲换位算子和逆转算子随机执行策略

在遗传进化过程中单亲换位算子和逆转算子以一定的随机概率执行.首先定义一个单亲换位算子的执行概率即换位率  $P_t$ ,用伪随机数  $R_t$  决定所要执行的算子.如果  $R_t \geq P_t$ ,执行单亲换位算子,否则执行逆转算子.换位率的大小根据实际问题来调整.

3.4.4 进化终止策略

遗传算法是一种反复迭代的搜索方法,它通过多代进化逐渐地接近最优解,但不一定是最优解.因此,必须事先给定进化过程的终止条件.算法中采用最大遗传代数作为进化终止条件,即在程序开始运行时,预先指定最大遗传代数,防止陷入无穷循环.在进化过程中,通过观察和比较群体平均适应度的变化,判断进化状态.一旦群体平均适应度没有变化或变化很小,说明进化过程收敛.

4 实例应用

现有一给水系统的管网初步连接方案,共有 10 个给水节点,23 条可能的连接边,拟求一组管网总长度最短树状管网布置形式.首先将管网初步连接方案概化为图 2 所示的管网初步连接图,各边的编号及长度如图 2 所示.以管网初步连接图中各边的长度为权值,采用改进的遗传算法求最小树.控制参数设置:群体规模为 10,最大遗传代数为 400 代,以不同的选择率和换位率组合模

式进行优化,可获得一组权最小或次最小的生成树如表 1 所示.数值实验表明:当选择率和换位率的组合模式在  $[0,0.5]$  之间变化时,程序运行时间为  $60\sim 265\text{s}$ ,共搜索和评价了  $20\times 400=8000$  个方案,平均每个方案的计算时间为  $0.0075\sim 0.033\text{s}$ .作为对比,采用图论中的 Kruskal 算法只能得到一个权为 60 的最小生成树,且与表 2 中的方案 2 相同,程序运行时间为  $0.05\text{s}$ .

计算机配置为 Pentium II, 64M RAM; 中文 Windows 98 操作系统;所有程序用 Matlab 5.2 编写.

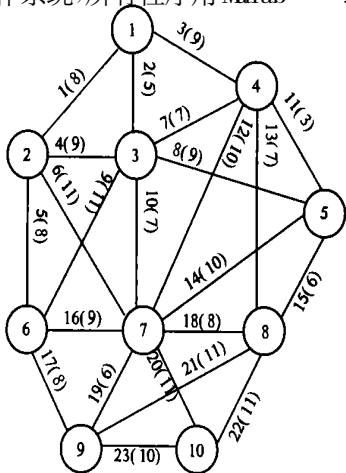


图 2 管网初步连接图

Fig. 2 Original connecting graph of pipe network

5 结束语

基于遗传算法的最小生成树算法可用于不同类型的最小树问题研究,能够在较短的时间内以较高的概率获得一批最小树或次最小树,为进行

实际工程的方案评价和决策提供更多的选择依据.算法所需数据量少,操作简单,且收敛性和适应性较强,可根据不同类型的最小树问题,构建不同的目标函数,寻找满足实际工程需要的最优解.在实际应用中,可以和一些经典的最小树算法配合使用,弥补贪心算法只能求出唯一最小树的不足,扩大方案评价和选择的范围.

表 1 用遗传算法求得的最小生成树

Tab. 1 Mini numspanning trees found by genetic algorithms

方案	权	组成最小生成树的边									
1	60	1	2	7	10	11	15	17	19	23	
2	60	1	2	5	7	10	11	15	19	23	
3	60	2	5	7	10	11	15	17	19	23	
4	61	2	4	7	10	11	15	17	19	23	
5	61	1	2	5	7	10	11	15	19	22	
6	61	2	5	10	11	15	17	18	19	23	
7	61	1	2	7	10	11	15	16	19	23	
8	61	1	2	7	10	11	15	17	19	22	
9	61	1	2	5	7	11	15	17	19	23	
10	61	2	5	7	11	15	17	18	19	23	

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 ] 陈国良.遗传算法及应用[ M ].北京:人民邮电出版社,1996.

[ 4 ] 周荣敏.遗传算法与人工神经网络优化理论及其在压力管网最优化中的应用研究[ D ].杨凌:西北农林科技大学,2000.

Mini num Spanning Tree Algorithms Based on Genetic Algorithms

ZHOU Rong-min, MAI Wen-ming, LEI Yan-feng

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

**Abstract :** Based on the graphic theory and genetic algorithm ,an i mproved genetic algorithm is introduced to search the mini num spanning trees . This algorithm uses binary code to represent the problem of mini num spanning trees and uses the depth first searching method to determine the connectivity of the graph .The corresponding fitness func - tion , single parent transposition operator , single parent reverse operators and four kinds of controlling evolutionary strategies are designed to i mprove its speed and efficiency .In comparison with kruskal algorithm ,it can acquire a set of mini num spanning trees during one genetic evolutionary process and is applicable to solving different kinds of mini num spanning tree problems .

**Key words :** genetic algorithms ; mini num spanning tree ; evolutionary strategy ; network opti mization