

基于改进遗传算法的可靠性网络优化设计

刘 东¹, 丁照宇²

(1. 成都理工大学 信息工程学院, 四川 成都 610059;

2. 成都理工大学 网络教育学院, 四川 成都 610059)

摘 要:在可靠性条件约束下,使网络成本最低是网络规划 NP-hard 问题。从遗传算法的基本原理出发并对其进行改进,分析带有可靠性约束条件的通信网设计中的网络优化问题,这一方法的最大优点是可将其推广到求解一般带有约束的网络优化问题。而且结果表明无论是解的精度还是运算速度遗传算法都优于分枝定界法及其它启发式算法。

关键词:通信网络;遗传算法;网络优化;可靠性约束

中图分类号:TP393.02

文献标识码:A

文章编号:1673-629X(2007)01-0063-02

Reliability Optimization for Network Design Based on Improved Genetic Algorithm

LIU Dong¹, DING Zhao-yu²

(1. College of Information Engineering, Chengdu University of Technology, Chengdu 610059, China;

2. College of Network Education, Chengdu University of Technology, Chengdu 610059, China)

Abstract: It is NP-hard problem to find a network topology for a set of nodes whose total link-cost is minimized, subjected to the condition of reliability. Basing on the theorem of genetic algorithm, this paper analyses the problem of design of network optimization with a constraint. This technique is very powerful because the same approach can be extended to solve different types of optimization problems with constraint. Results show that GA can produce better solutions than branch and bound in accuracy and computation time.

Key words: communication networks; genetic algorithm; network optimization; reliability constraint

0 引 言

通信网系统由若干个结点及结点之间的连接边构成,网络系统之间的连接分为直接连接和间接连接。现代通信网在运行的过程中,随着时间的推移,存在着网络的结点扩充及重组等问题。其数学模型是构造一个网络结构使其达到在全终端可靠性约束下使网络结构费用最小的问题。此类问题一般都属于 NP 问题。实际应用中人们主要考虑的是基于分枝定界法下求其问题的精确解^[1,2]。由于受到搜索空间的限制,分枝定界法仅限于结点和边较少的网络。

遗传算法(GA)模拟自然进化过程,是一种具有并行特征的搜索算法,利于对解空间进行搜索,加快了对解的搜索速度,便于推广到多结点的网网络优化设计中,是解决大规模网络优化问题的有效工具^[3]。

1 符号、定义与假设

(1) 符号与定义。

$G(N, L)$ ——结点集合为 N , 边集合为 L 的无向图(网络)。网络 G 的结点数为 $|N|$, 边数为 $|L|$; (i, j) ——结点为 i, j 的无向边; p, q ——分别表示边的可靠度与不可靠度,并且满足 $p + q = 1$; $R(G, p)$ ——网络 G 各边具有独立可靠度 p 的全终端可靠性; $e_{i,j}$ ——与结点 i, j 是否连结的状态函数。当边 (i, j) 存在时 $e_{i,j} = 1$; 否则 $e_{i,j} = 0$; e ——全体构成 $e_{i,j}$ 的集合, $e = \{e_{1,2}, e_{1,3}, \dots, e_{1,n}, e_{2,3}, e_{2,4}, \dots, e_{n-1,n}\}$; $f(e)$ —— e 所表示的网络的系统可靠性; p_0 ——给定可靠性下界(约束); $C_{i,j}$ 或 $C(e_{i,j})$ ——边 (i, j) 或边 $e_{i,j}$ 的费用; d_i ——结点 i 的度,且 $\sum_{i=1}^n d_i = D$; d ——度序列, $d = (d_1, d_2, \dots, d_n)$; $d(\alpha, j)$ ——存在数 d 使当 $i = 1, 2, \dots, j$ 时, $d_i = \alpha$; 其它 $d_i = \alpha + 1$ ($1 \leq j \leq n$); $r(l)$ —— l 边网的最大系统可靠性; $\bar{r}(l)$ —— $r(l)$ 的上界; SIZE——群体的大小(也称群体规模); LENGTH——染色体串长度; p_c ——杂交概率; d_m ——变异概率。

(2) 假设。

① 网络的边以一定的概率 p 工作(称为边不可靠),全部结点总是正常工作(称为点可靠);整个网络及每个边只有两个状态(正常工作或失效)。② 整个网络及每个边

收稿日期:2006-04-13

作者简介:刘 东(1973-),男,四川绵阳人,硕士,研究方向为网络计算技术;丁照宇,硕士生导师,教授,研究方向为数据库应用。

的工作状态具有统计独立性。③ 网络各边的费用给定而且是不变的。④ 网络不包含重边的情形。

(3) 问题。

求解可靠性约束下网络结构优化问题的抽象模型为：

目标函数： $Z = \min \sum_{i,j} C_{i,j} * e_{i,j}$ ；满足约束条件： $f(e) \geq p_0$ 。

2 网络优化问题的遗传算法

遗传算法包括的要素有：编码、初始群体的设定、适应度函数的设计、控制参数的设定、遗传操作设计等，这 5 个要素构成了遗传算法的核心内容^[4]。编码的目的是将优化问题的解空间转化为遗传算法易于操作的编码空间。一般编码方式模仿生物的染色体结构，进行二进制串（基因）位编码，使解空间的每一个元素（个体）成为一个基因型。与优化目标函数相关的适应度函数是一种评价函数，用于评价个体在整个群体中的生存能力。初始群体是 GA 的初始操作对象，它的生成一般是随机的，但为了使搜索更加有效，也可以根据经验选择初始群体。遗传操作则是 GA 模仿生物界中的遗传繁殖和自然选择机制，对群体中的个体的基因进行交换、重组或突变的操作过程，称为选择操作、杂交操作和变异操作等。参数包括群体规模 N 、迭代停止条件 G 、杂交概率 p_c 和变异概率 p_m 等。

自适应遗传算法流程如下：

步骤 1：编码 对染色体进行串型编码，上节中的变量 $X_{i,j}$ 为 0、1 变量，可直观地转化为遗传算法的编码。对于一个 N 结点的网络，其 G 图最多有 $N(N-1)/2$ 条边，所以染色体的长度可定长为 $N(N-1)/2$ 的二进制串。问题的解可用如下的上三角矩阵表示，称其为 G 的邻接矩阵 E 。

$$\begin{bmatrix} 0 & e_{1,2} & e_{1,3} & \cdots & e_{1,n} \\ & 0 & e_{2,3} & \cdots & e_{2,n} \\ & & \ddots & & \vdots \\ & & & 0 & e_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

图 1 问题解的邻接矩阵

步骤 2：初始化种群 选定种群规模 $N = P$ ，随机产生种群中每个个体。

步骤 3：计算种群中每个个体的适应度 依解的编码规则 and 问题的特性，所考虑的适应值函数应反映以下三方面的内容：a. $f(x)$ 应反映网络 G 的连通性 f_1 ；b. $f(x)$ 应满足可靠性 p_0 的约束 f_2 ；c. $f(x)$ 的解所对应的网络费用最低 f_3 。为此首先给定适应值的一个上界 m （取 $m = 1$ ），以上三部分的权值分别为 $0.5m, 0.3m, 0.2m$ 。因为染色体 x 对应的可达矩阵中“1”的个数 k 表示其连通程度，因此可令： $f_1 = 0.5mk$ ，根据 Jan Ronghong 的研究^[5]，对可靠性上界作如下计算：

$$r(n-1) = p^{n-1}; r(n) = p^n + np^{n-1}q;$$

$$r(n+1) = p^{n+1} + (n+1)p^nq + \frac{(n+1)^2 p^{n-1} q^2}{3}$$

$$r(n+2) = p^{n+2} + (n+2)p^{n+1}q + \frac{3(n+2)^2 p^n q^2}{8} + \frac{(n+1)^2 p^{n-1} q^3}{16}$$

$$\bar{r}(l) = H(d(\alpha, j)) \quad l \geq (n+3)$$

$$\alpha = \lfloor D/n \rfloor, j = n(\alpha+1) - D$$

$$H(d) \equiv 1 - \left[\sum_{i=1}^n q^{d_i} \prod_{k=1}^{m_i} (1 - q^{d_k-1}) \prod_{k=m_i+1}^{i-1} (1 - q^{d_k}) \right]$$

$m_i = \min(d_i, i-1)$ 。如果 $\bar{r}(l) \geq p_0$ ，并且 $r(l-1) < p_0$ ，取 l 为满足可靠性约束所必须达到的边的最大上界 l ，令 $f_2 = 0.3m(k/l)$ 。

若令染色体 $e = \{e_{1,2}, e_{1,3}, \dots, e_{1,n}, e_{2,3}, e_{2,4}, \dots, e_{n-1,n}\}$ ，其对应网络费用为 $\sum_{i,j} C_{i,j} * X_{i,j}$ ，而最大网络费用为 $\sum C_{i,j}$ 。因此可令： $f_3 = 0.2m(\sum_{i,j} C_{i,j} * X_{i,j} / \sum C_{i,j})$ 。最后可得问题的适应值函数： $f = f_1 + f_2 + f_3$ 。

步骤 4：遗传操作 包括选择、交叉和变异 3 个基本遗传算子。交叉算子和变异算子具有自适应功能。选择过程可采用 $(\mu + \lambda)$ 选择和转轮选择的混合选择方法。 $(\mu + \lambda)$ 选择是 Back 和 Hoffmeister 提出的，它从 μ 个双亲和 λ 个后代中选择 μ 个最好的染色体。如果没有 μ 个不同的染色体，种群中的空缺则用转轮选择来填满。

由于 p_c 和 p_m 对遗传算法的运行性能有较大的影响，故应防止 p_c 和 p_m 的值选择不当造成的进化过程过早收敛或收敛速度极慢的现象出现^[6]。自适应算子 p_c 和 p_m 的值的选取，对远离最优值的个体采用较大的 p_c 和 p_m ，以提高搜索速度；对接近最优值的个体采用较小的 p_c 和 p_m ，保证不收敛于局部最优解。为此， p_c 和 p_m 的选取就不能简单地随适应值 f 线性变化。这样，文中构造的自适应遗传算子为：

$$p_c = \begin{cases} \frac{p_{c1} + p_{c2}}{2} + \frac{p_{c1} - p_{c2}}{2} \cos \left[\frac{f - f_{avg}}{f_{min} - f_{avg}} \pi \right] & f \leq f_{avg} \\ p_{c1} & f > f_{avg} \end{cases}$$

$$p_m = \begin{cases} \frac{p_{m1} + p_{m2}}{2} + \frac{p_{m1} - p_{m2}}{2} \cos \left[\frac{f - f_{avg}}{f_{min} - f_{avg}} \pi \right] & f \leq f_{avg} \\ p_{m1} & f > f_{avg} \end{cases}$$

式中： $p_{c1} = 0.9$ ； $p_{c2} = 0.1$ ； $p_{m1} = 0.2$ ； $p_{m2} = 0.01$ 。

文中采用无退还按个体 p_c 选择群体及单点杂交方法，实现遗传算法中群体杂交的过程。变异是无退还按个体 p_m 选择群体并随机产生一个变异点，在变异点位置将字符由 0（或 1）变为 1（或 0）。

步骤 5：终止条件为适应值 50 代保持不变或达到最大允许代数 400。计算判断满足终止条件否，如满足，则停止；否则转回到步骤 3。

3 实例及结果

这里给出一个网络优化问题 P 的仿真实例。已知各结点间架设通讯线路的代价，要在八结点（编号 1~8）间规划网络拓扑结构，使网络建设代价最小。约束条件为：

（下转第 68 页）

分类器的识别很难大幅度提高,而文中提出的两层结构分类器设计可以很快达到满意的程度。

由图 4 可知,随着学习了更多的样本,单层分类器的领域训练时间一直是上升的趋势,而且上升都保持较高幅度;而两层结构分类器设计,当学习样本增长到一定程度后,反而出现了下降的趋势,在“5:1 比例”处,训练时间出现了下降,这是因为在两层结构分类器设计中,将和单层分类器一样的学习样本按照分层次的结构进行学习的结果。

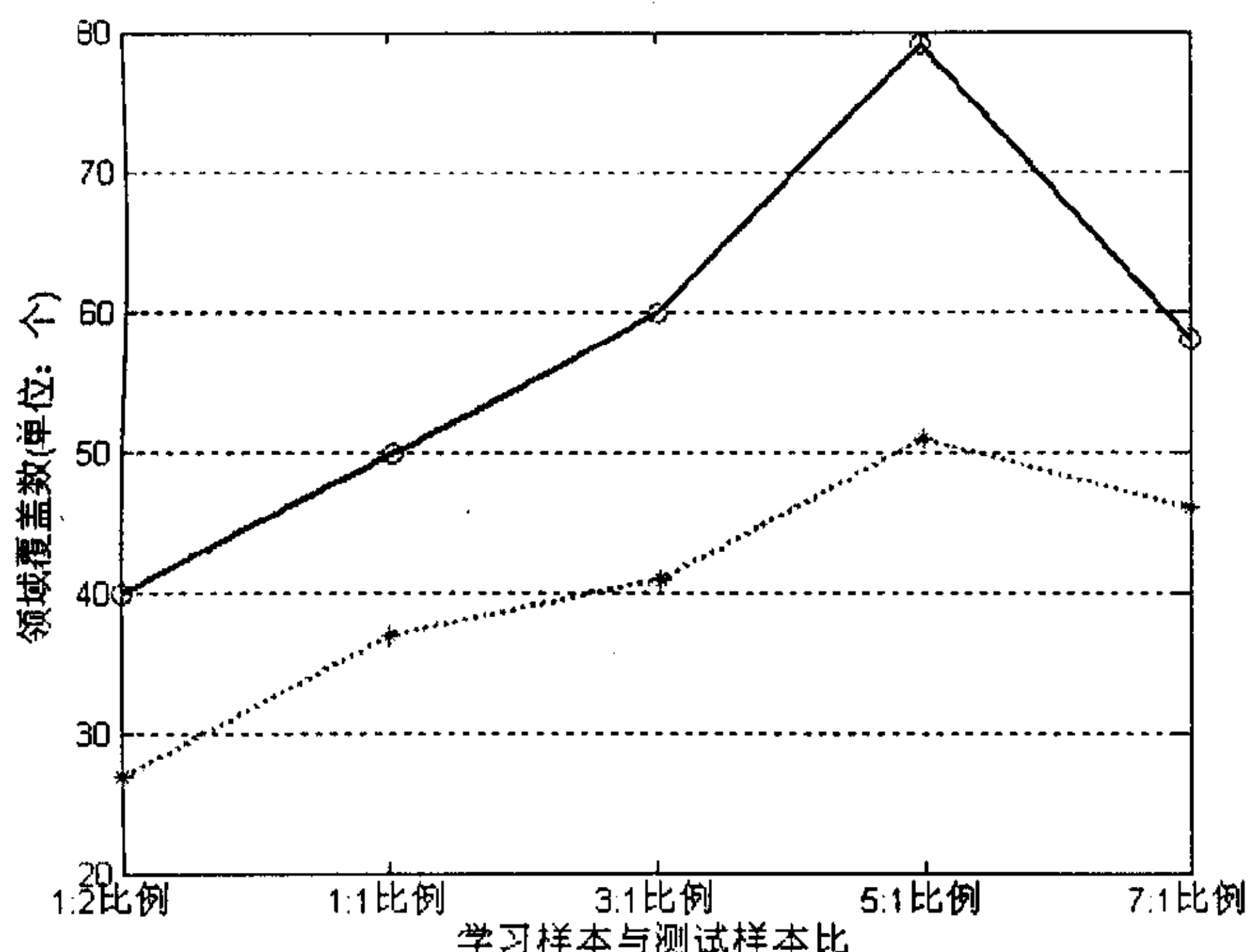


图 5 学习样本增加时的领域覆盖数变化图

由图 5 可知,当学习样本相对于测试样本比例不断增大时,两层结构分类器设计比单层分类器的领域覆盖数一

直要多,即分类器设计的复杂度要高一些。但是动态增长的幅度基本上两种方法趋于一致。

4 结 论

综合以上分析可知,当学习样本与测试样本比例大到一定数值时,与单层的覆盖算法构造的分类器相比,两层结构分类器在不明显增加分类器设计复杂度的情况下,使分类器的性能主要是识别正确率和分类器的训练学习时间分别得到了较大的优化,特别是最重要的识别正确率大幅度地提高到了令人满意的程度。

参考文献:

- [1] 边肇祺,张学工. 模式识别[M]. 第 2 版. 北京:清华大学出版社,1999:176-212.
- [2] 沈淑娟,姜建国,曹建春. 手写体字符识别的多特征多分类器设计[J]. 计算机工程与应用,2004(16):116-118.
- [3] Zhang Ling, Zhang Bo. A Geometrical Representation of McCulloch-Pitts Neural Model and Its Applications[J]. IEEE Trans. on Neural Networks, 1999, 10(4):925-929.
- [4] 张 铃,张 钺. 多层反馈神经网络的 FP 学习和综合算法[J]. 软件学报,1994,8(4):252-258.
- [5] 张 铃,张 钺. 多层前向网络的交叉覆盖设计算法[J]. 软件学报,1999,10(7):737-742.
- [6] 张 铃,张 钺. 神经网络的规划学习算法[J]. 计算机学报,1994,17(9):669-675.

(上接第 64 页)

$f(e) \geq 0.90$ 。表 1 为各结点间线路规划代价。经过计算,可得到最小代价为 588,网络拓扑结构如图 2 所示。

表 1 线路规划代价

	1	2	3	4	5	6	7	8
1	0	39	57	123	54	81	69	87
2		0	21	96	81	117	93	78
3			0	102	102	138	102	90
4				0	126	177	192	54
5					0	51	108	75
6						0	99	126
7							0	162
8								0

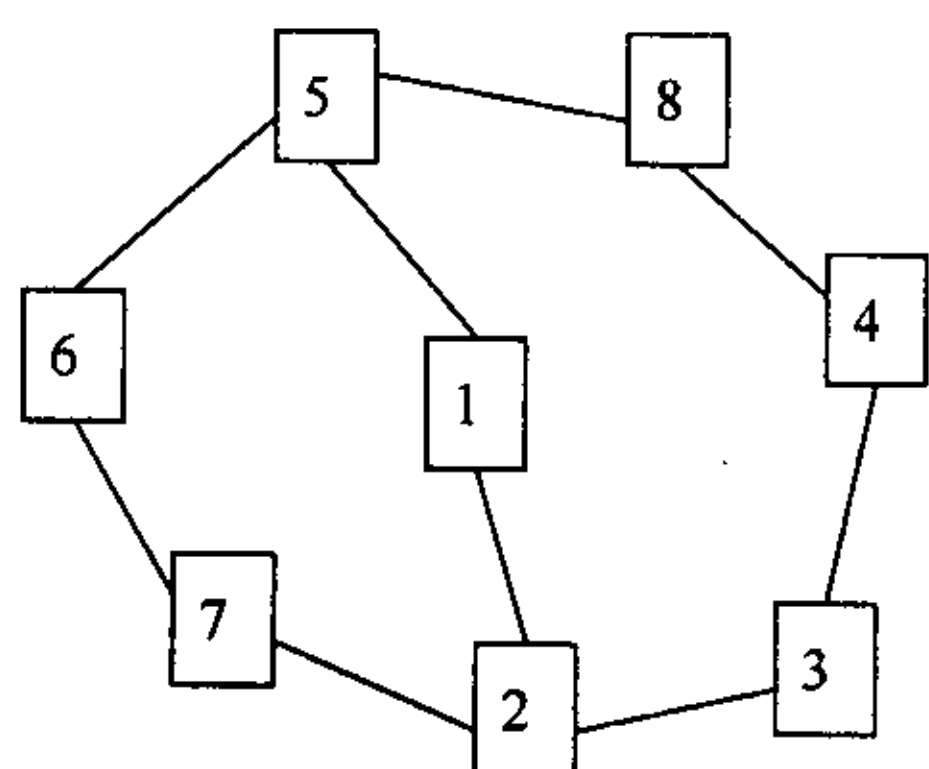


图 2 网络优化拓扑结构图

4 结 论

改进了的交叉算子和变异算子的自适应遗传算法能

够根据适应度的大小自适应地选取交叉算子和变异算子,从而可以避免问题解的“早熟”或收敛速度过慢现象出现。通过一个简单网络优化问题的仿真,验证了这种策略具有较好的效果。

参考文献:

- [1] Jan Ronghong, Huang Fungien, Cheng Shengtong. Topological optimization of a communication network subject to a reliability constraint[J]. IEEE Trans Reliab, 1993, 42:63-70.
- [2] Zhao Lianchang, Shao Fangming. Optimization of connecting two communication network subject to a reliability constraint [J]. Microelectron Reliab, 1997, 37(4):629-633.
- [3] Kumar A, Pathak R A, Gupta Y P. Genetic - algorithm - based reliability optimization for computer network expansion [J]. IEEE Trans Reliab, 1995, 44:63-72.
- [4] 阎平凡,张长水. 人工神经网络与模拟进化计算[M]. 北京:清华大学出版社,2000.
- [5] Jan Ronghong. Design of reliable networks[J]. Computers Ops, 1993, 20(1):25-34.
- [6] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutations in GAs[J]. IEEE Trans on SMC, 1994, 24(4): 656-666.