

研究简报

# 基于遗传算法的模糊树建模方法

代冀阳 张建刚 毛剑琴

(北京航空航天大学第七研究室 100083 北京)

(E-mail: dumao@public.bta.net.cn)

**关键词** 模糊树模型, 遗传算法, 矩阵编码, 非线性系统建模.

## FUZZY-TREE MODELING BASED ON GENETIC ALGORITHM

DAI Jiyang ZHANG Jiangan MAO Jianqin

(The Seventh Research Division (System & Control), Beijing University of Aeronautics  
& Astronautics, Beijing 100083)

**Key words** Fuzzy-tree model, genetic algorithm, matrix coding, nonlinear system modeling.

## 1 引言

近年来模糊逻辑理论在非线性系统辨识领域中得到了广泛的应用. 模糊逻辑建模方法的主要内容可分为两个层次, 一是模型结构的辨识, 另一则是模型参数的估计. 模糊模型的结构辨识问题是指如何划分输入空间. 模糊树模型(FT模型)是一种利用二叉树结构描述输入空间模糊划分的模糊建模方法, 其主要特点是建模精度高、计算量小<sup>[1]</sup>. 遗传算法(GA)是一种模拟生物进化过程的计算模型, 特别适于处理传统寻优方法难以解决的复杂非线性寻优问题<sup>[2, 3]</sup>. 本文提出的基于遗传算法的模糊树建模方法(FT/GA方法)就是以模糊树模型作为个体, 采用矩阵编码方式, 利用遗传算法在整个模型空间搜索最优模糊树. 该方法能有效地进化模糊树的结构, 得到一个精度较高而复杂度较低的次优模型.

## 2 模糊树模型

考虑多输入单输出非线性系统的建模问题

$$y = f(x), \quad (1)$$

其中  $x = [1, x_1, \dots, x_n]^T \in R^{n+1}$  为输入向量,  $y \in R$  为输出.

一个模糊树模型与一个二叉树  $T$  相对应. 在每个节点  $t \in T$ , 定义权向量  $\mathbf{c}_t = [c_t(0), \dots, c_t(n)]^T$  和模糊集  $N_t$ , 其隶属度函数为  $\mu_t(x): R^{n+1} \rightarrow [0, 1]$ . 这样, 可在每个节点上定义一个线性模型, 而每个节点的模糊集  $N_t$  则描述了一个模糊子空间  $\chi_t$ . 在节点  $t \in T, t \neq r(T)$ ,  $r(T)$  表示叶节点, 定义辅助隶属度函数

$$\hat{\mu}_t(x) = \frac{1}{1 + \exp[-\alpha_t(\mathbf{c}_{p(t)}^T \mathbf{x} - \theta_t)]}, \quad (2)$$

式中  $p(t)$  表示  $t$  的父节点. 则模糊集  $N_t$  的隶属度函数为  $\mu_t(x) = \mu_{p(t)}(x) \hat{\mu}_t(x)$ .

于是, 模糊树模型将输入空间划分为所有叶节点对应的模糊子空间, 其中每个模糊子空间内定义一条规则: If  $\mathbf{x}$  is  $N_t (t \in \tilde{T}, \tilde{T}$  表示  $T$  的叶节点集合), then  $\hat{y}_t = \mathbf{c}_t^T \mathbf{x}$ . 采用中心平均反模糊化方法, 则模糊树模型的输出  $\hat{y}_T$  为所有叶节点的输出的加权平均.

设给定的训练数据集为  $\{(x^i, y^i), x^i \in R^{n+1}, y^i \in R, i = 1, 2, \dots, L\}$ . 定义能量函数

$$e = \sum_{i=1}^L e^i = \sum_{i=1}^L \frac{1}{2} (y^i - \hat{y}_T^i)^2, \quad (3)$$

可采用混合训练算法, 通过最小化能量函数  $e$  来确定模糊树模型中的待定参数  $\{\mathbf{c}_t | t \in T\}$  及  $\{\alpha_t, \theta_t | t \in T, t \neq r(T)\}$ , 具体的算法详见文献[1]. 因此, 一个二叉树结构就对应着输入空间的一种划分, 通过一定的训练算法确定出各叶节点的参数, 即可得到一个模糊树模型. 本文采用如下步骤确定在树结构给定时的各叶节点参数:

- 1) 按文献[1]的方法初始化模糊树根节点,  $j=2$ ,  $j$  为模糊树的层数;
- 2) 搜索第  $j$  层上的所有节点, 将这些节点当作叶节点, 采用混合训练算法计算出它们相应的  $\theta_t, \alpha_t, \mathbf{c}_t$ ;
- 3) 判断这些节点是否为叶节点, 若不是, 则  $j=j+1$ , 转2); 若是, 则结束算法.

### 3 搜索最优模糊树模型

讨论利用遗传算法搜索最优模糊树的方法和步骤.

1) 编码. 对于树结构个体, 无法采用二进制编码. 本文提出一种有效的树结构编码方式——矩阵编码, 这种编码方式简单, 而且非常适合遗传算法的各个遗传操作. 用矩阵表示二叉树结构的过程如图1所示, 图中“0”表示“无”.

2) 群体初始化. 遗传算法中, 群体的规模  $N$  是一个很重要的设计参数, 但该参数的选取目前还没有统一的准则, 本文将群体规模取在10~20之间, 并且采用递归循环算法随机产生初始树个体, 为了避免过大的计算量, 树的深度  $d$  固定.

3) 模型评价. 模型的优劣可从精度和复杂度这两个方面来衡量, 模型的精度由能量函数  $e$  表示,  $e$  越小, 则精度越高; 模糊树模型的复杂度可由模糊树叶节点的个数  $m$  来反映,  $m$  越小, 则模型的复杂度就越低. 因此, 反映模型品质的树个体适应度可定义如下:

$$g(T) = W_e \times \frac{1}{e} + W_m \times \frac{1}{m}, \quad (4)$$

式中  $g(T)$  表示树个体  $T$  的适应度;  $W_e, W_m$  为权值.

4) 复制或选择. 本文采用最常用的适应度比例选择或赌轮选择方法, 选择概率取为1.

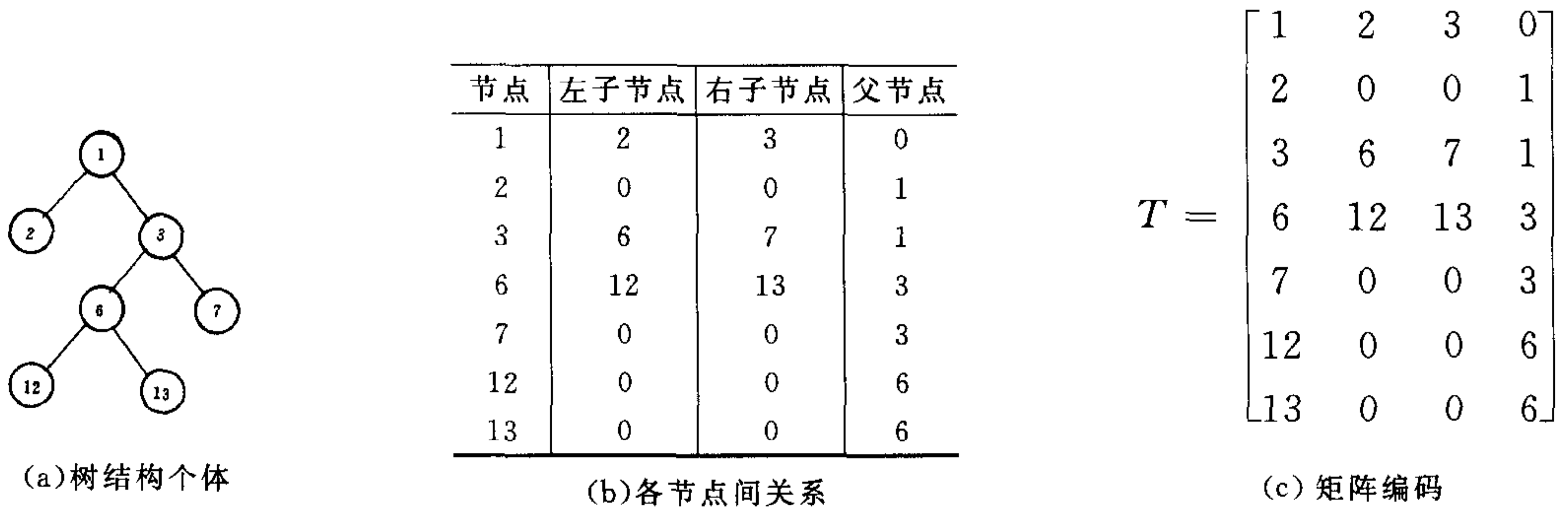


图1 二叉树结构的矩阵编码

5) 交叉. 两个个体的交叉由交叉概率  $p_c$  决定是否进行. 当进行交叉时, 在群体中随机选择两个树个体作为父代, 从这两个树个体上分别随机选取一个子树, 然后这两个个体相互交换各自子树的结构, 这样就产生了新一代. 两个个体交叉的过程如图2所示.

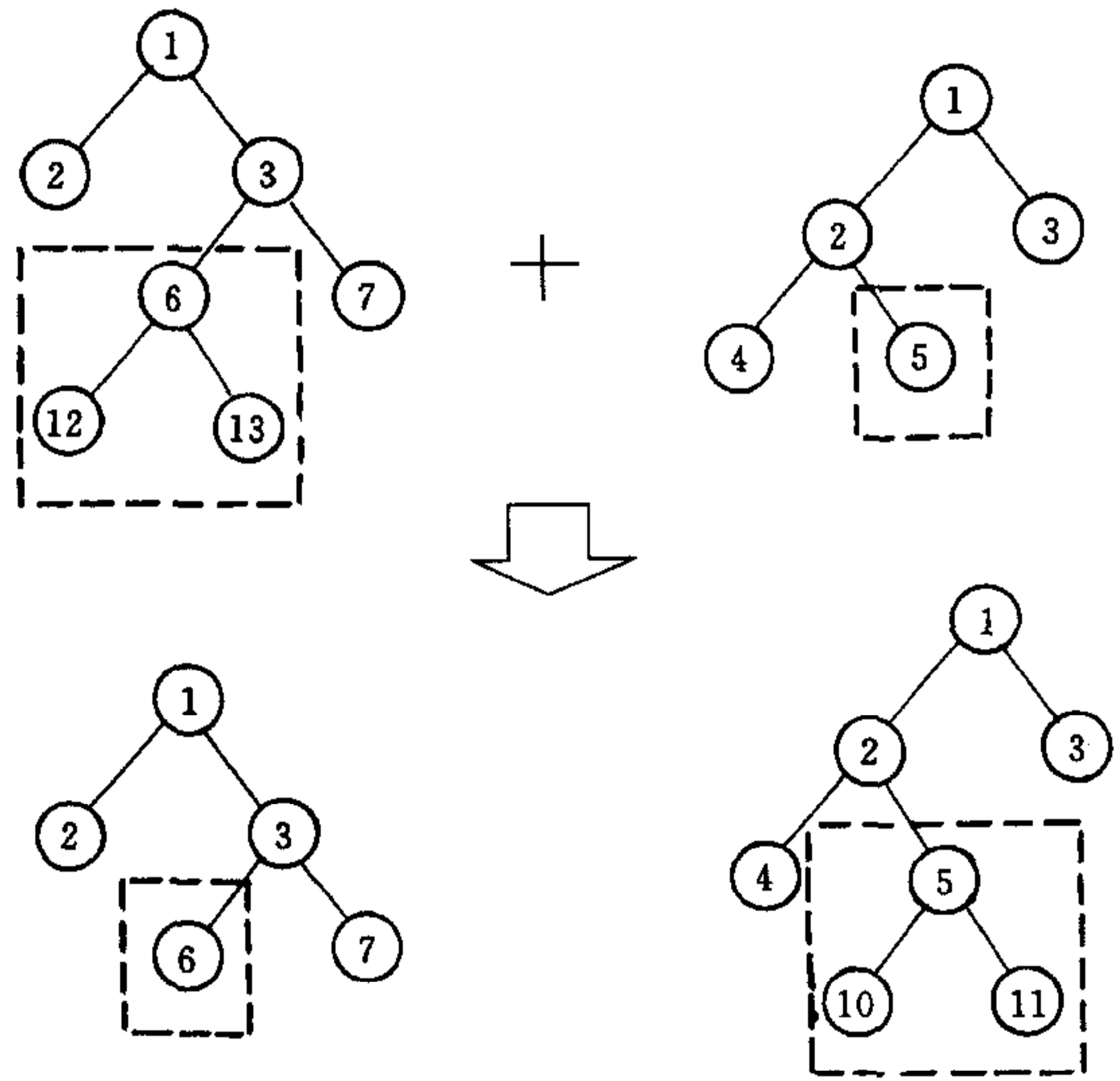


图2 树个体交叉

6) 变异. 树个体是否变异由变异概率  $p_m$  决定. 当个体变异时, 在树个体上随机选取一个子树, 然后用一个随机产生的树来代替被选的子树, 同样地, 这种替换也是结构上的替换.

7) 剪枝. 经过上述三种遗传操作后, 将会产生更深的树, 过深的树结构将会使计算量大大增加, 而且缺乏一般性. 因此, 本文对超过给定深度的树进行剪枝, 也即将树中超过给定深度的节点全部删去.

8) 结束条件. 当迭代次数超过给定代次或最优个体误差小于给定值时, 结束算法.

### 4 仿真算例

采用文献[4]中给出的例子进行仿真. 考虑如下两输入单输出非线性静态系统

$$y = x_1 e^{(-x_1^2 - x_2^2)}, \quad -2 \leq x_1, x_2 \leq 2,$$

其中  $x_1, x_2$  为输入量,  $y$  为输出量.

仿真实验中, 在输入空间内任意产生200对独立均匀分布的随机数  $\{(x_1(k), x_2(k)), k=1, 2, \dots, 200\}$ , 再加上它们的函数值  $y$  构成训练样本集. 本次实验所选用的遗传算法各参数为: 群体规模  $N=14$ , 树的深度  $d=6$ , 选择概率  $p_s=1$ , 交叉概率  $p_c=0.95$ , 变异概率  $p_m=0.01$ , 适应度中的权值  $W_e=1, W_m=10$ .

本次寻优搜索过程共进行50代, 并把每一代的最佳个体记录下来. 从整体变化趋势看每一代的最佳个体适应度及平均适应度随代次的增加而增大, 但并非在每一代都能产生

最佳个体. 在整个搜索过程中, 第38代出现了最大适应度值, 其值为  $g=75.4464$ , 此代之后, 每代的最佳个体之间的差别并不是很大, 因而作者把第38代的最佳个体作为最优模糊树模型. 该模型将输入空间划分为23个模糊子空间, 即采用了23条模糊规则, 其估计上述非线性函数的残差曲线如图4所示(图中  $e_s$  表示误差,  $k$  表示样本数), 最大误差为

$$\max_{1 \leq k \leq 200} \{ |y(k) - \hat{y}_T(k)| \} = 0.0375.$$

该例表明, 以树结构为个体, 采用矩阵编码方式, 遗传算法能在较短的代次内有效地搜索到一个次优模糊树模型. 对于同样的例子, 在树的深度同为6的情况下, FT 模型得到的规则数为32, 最大误差为0.0198; 而文献[4]得到的规则数为22, 但最大误差是0.0903. 因此, 同文献[4]的方法相比, 当所采用的规则数相差不多时, 本文提出的 FT/GA 方法的精度较高; 同 FT 模型建模方法相比, 本文的方法在保证模型具有足够精度的前提下, 优化了输入空间的划分, 降低了模型的复杂度.

## 5 结论

本文提出了一种基于遗传算法的模糊树建模方法. 为了能够有效地搜索最优模糊树模型, 在遗传算法中, 采用了树个体的矩阵编码方法, 此编码方法非常适合进行各种遗传操作, 易于算法实现. 模糊树模型通过遗传算法得以进化, 从而得到一个次优解, 该次优模型既具有较低的复杂度, 又能以较高的精度逼近非线性系统, 仿真结果验证了该建模方法的正确性和有效性. 对于如何用遗传算法同时优化模糊树的结构及各叶节点的参数, 这是今后需要进一步研究的课题.

## 参 考 文 献

- 1 张建刚. 复杂系统辨识——方法、软件及应用研究[博士论文]. 北京:北京航空航天大学, 1999
- 2 Goldberg D. Genetic Algorithms in Search, Optimization, and Machine Learning. Massachusetts: Addison-Wesley, 1989
- 3 陈国良, 王煦法, 庄镇泉, 王东生. 遗传算法及其应用. 北京:人民邮电出版社, 1996
- 4 Masahiro Tanaka, Ju Ye, Tetsuzo Tanino. Fuzzy modeling by genetic algorithm with tree-structured individuals. *International Journal of Systems Science*. 1996, 27(2):261~268

**代冀阳** 1966年生. 1991年毕业于南京航空航天大学, 获硕士学位. 目前在北京航空航天大学第七研究室控制理论与控制工程专业攻读博士学位. 主要研究方向为智能控制理论及应用、鲁棒控制、线性矩阵不等式及其应用等.

**张建刚** 1971年生. 1999年4月毕业于北京航空航天大学获博士学位, 研究方向为小波理论在系统辨识中的应用、系统智能化辨识与参数的智能化估计等.

**毛剑琴** 见本刊25卷2期.