

文章编号: 1007-2284(2009)06-0054-02

# 遗传程序在水位流量关系拟合中的应用

段春青<sup>1</sup>, 陈晓楠<sup>2</sup>, 韩黎明<sup>2</sup>

(1. 北京师范大学, 北京 100875; 2. 南水北调中线干线工程建设管理局, 北京 100053)

**摘要:** 针对传统水位流量关系曲线拟合过程中存在精度不高的问题, 应用遗传程序拟合水位流量关系。该方法能够通过演化计算自动寻找最优的模型结构, 比起传统统计方法具有较大的灵活性和智能性, 避免了预先假定具体数学表达式的不足。将基于遗传程序的水位流量关系模型同基于最小二乘法、遗传算法以及蚁群算法的模型进行详细比较, 实例表明, 基于遗传程序设计的水位流量关系模型适应性强, 拟合精度最高, 为水位流量关系拟合提供了新的有效方法。

**关键词:** 水位流量关系; 遗传程序; 遗传算法; 蚁群算法  
**中图分类号:** P337.3      **文献标识码:** A

水位和流量关系反映水位和流量的对应状态, 由于流量测验技术比较复杂、耗资比较昂贵, 难以连续进行, 而连续地阅读水位则容易办到, 因此通常将连续的水位资料, 通过水位流量关系推算、转换为连续的流量资料, 供水文计算或水文预报分析使用; 有时也因某种需要, 由流量通过水位流量关系反推水位, 如河道防汛水位, 闸坝电站下游水位等。所以, 水位流量关系具有重要的实用意义<sup>[1]</sup>。然而, 传统方法首先假定水位流量关系为某一非线性函数, 然后以计算值与实际值误差最小为原则, 利用最小二乘法、遗传算法或蚁群算法等求解计算模型参数<sup>[2-3]</sup>。这种方法可能产生由于结构选择不当而造成拟合精度不高的问题, 因为并不确定水位流量关系最符合何种具体函数。

遗传程序 (Genetic Programming, GP), 是在遗传算法的基础上加以延伸和扩展而形成的一种新的演化算法, 它依据自然选择的原理和生物进化论, 通过进化由树结构表示的计算机程序, 已成功地求解了许多领域的问题, 在建模、预测、控制、神经网络设计、人工生命等方面得到了成功应用<sup>[4-9]</sup>。这是一种自动化编程技术, 擅长于模型结构的自动搜索, 具有适应性强、精度高等优点。本文借助遗传程序建立水位流量函数模型, 并同传统方法进行拟合精度比较。

## 1 遗传程序

遗传程序的具体步骤如下:

**步骤 1 确定目标函数。** 设输入和输出样本系列分别为  $X$

$= \{(x_{11}, x_{12}, \dots, x_{1m}), (x_{21}, x_{22}, \dots, x_{2m}) \dots, (x_{n1}, x_{n2}, \dots, x_{nm})\}$ ,  $Y = \{y_1, y_2, \dots, y_n\}$ 。设计过程就是确定一个最佳函数表达式  $G(x_1, \dots, x_m)$ , 使误差极小化:

$$\min f = \sum_{k=1}^n |G(x_{k1}, \dots, x_{km}) - y_k| \quad (1)$$

式中:  $c$  为实常数。

**步骤 2 编码。** 确定终端集  $T$  和函数集  $F$ 。终端集  $T$  的元素定为变量  $x$  和常数  $c$ 。在本文中, 函数集  $F$  的元素定为算术操作  $\{+, -, \times, /\}$  和初等函数  $\{\sin, \cos, \arctan, \operatorname{arccot}\}$ 。对并集  $D = T \cup F$  进行统一编码, 表 1 给出了一种编码方案。若编码值为 0 则在区间  $[-1, 1]$  上随机选取一数; 若表中的编码值为 1, 则对应  $x_i (i=1, 2, \dots, m)$  中的某一个。在 GP 中称函数  $G(x_1, \dots, x_m)$  为式 (1) 问题的解, 并用二叉树的形式来表示。

表 1 遗传程序设计的一种编码方案

并集 $D$ 中的元素	编码值	并集 $D$ 中的元素	编码值
$c$	0	/	5
$x$	1	$\sin$	6
$+$	2	$\cos$	7
$-$	3	$\arctan$	8
$\times$	4	$\operatorname{arccot}$	9

**步骤 3 父代群体的初始化。** 设群体规模为  $N$ , 则共产生  $N$  棵树, 这些树的最大深度 (即层数) 应取较小的整数 (例如 4~6), 以便对由 GP 搜索到的最优函数表达式进行分析解释。产生初始群体  $N$  棵树时, 每棵树的根节点可在函数集  $F$  对应的编码值中随机选择, 中间结点可在并集  $D$  对应的编码值中随机选择, 叶节点可在终端集  $T$  对应的编码值中随机选择。

**步骤 4 父代群体的解码和适应度评价。** 把函数表达式  $G$

带入式(1), 可得各个体所对应的目标函数值  $f(i)$ ,  $i=1, 2, \dots, N$ , 把  $f(i)$  按从小到大排序, 排在最前面的几个个体为优秀个体。定义排序后第  $i$  个父代个体的适应度函数值  $F(i)$  为

$$F(i) = 1/[f(i) \times f(i) + 0.001] \tag{2}$$

步骤 5 对父代群体进行遗传操作。选择操作取比例选择方式, 则父代个体  $i$  的选择概率为

$$s(i) = F(i) / \sum_{i=1}^N F(i) \tag{3}$$

令  $p(0)=0$ ,  $p(i)=\sum_{k=1}^i s(k)$ ,  $i=1, 2, \dots, N$ 。生成一个  $[0, 1]$  区间的均匀随机数  $u$ , 若  $u$  在  $[p(i-1), p(i)]$  中, 则第  $i$  个个体被选中。为增强持续全局搜索的能力, 把最优秀的 5 个父代个体直接复制为子代个体。

在 GP 中交叉操作就是在两棵父代个体树上随机产生两个交叉点, 然后交换以交叉点为根节点的子树。

在 GP 中的变异操作就是在选得的个体中随机产生一个变异点, 然后以变异点为根节点, 将其以下的子树(包括变异点)按照步骤 3 的方式随机产生一棵子树来代替。

在产生每个子代个体中, GP 以相应得操作概率执行上述三种遗传操作之一: 设选择操作概率为  $p_s$ , 交叉操作概率为  $p_c$ , 则变异操作概率为  $p_m=1-p_s-p_c$ 。产生一个  $[0, 1]$  区间上的均匀随机数  $u$ , 若  $u \leq p_s$  则进行选择操作, 若  $p_s < u \leq (p_s + p_c)$  则进行交叉操作, 若  $u > (p_s + p_c)$  则进行变异操作。如此反复进行  $N-5$  次, 完成 1 次进化迭代。

步骤 6 记录最佳的个体, 并把子代群体作为新的父代群体, 的转入步骤 4, 如此反复演化, 直至进化迭代次数大于预设值, 或目标函数值达到预设值, 结束算法的运行。此时最佳个体为最终结果。

2 基于遗传程序设计的水位流量函数

传统拟合方法是假设水位流量关系为某一非线性函数, 其数学模型一般为

$$Q = aH^b \tag{4}$$

式中:  $Q$  为流量,  $\text{m}^3/\text{s}$ ;  $H$  为水位,  $\text{m}$ ;  $a$ ,  $b$  为待拟定的模型参数。

然后根据残差平方和最小准则、绝对残差绝对值和最小准

则或相对残差绝对值和最小准则, 借助最小二乘法、遗传算法或蚁群算法等优化方法计算模型参数。这种方法会由于模型结构选择不合适而得不到很好的拟合效果。本文使用遗传程序自动寻找最优的结构和参数。

设有  $N$  组实测水位流量数据  $(H_i, Q_i)$ ,  $i=1, 2, \dots, N$ , 按下述公式作标准化处理:

$$x_i = (H_i - H_{\min}) / (H_{\max} - H_{\min}) \quad i=1, 2, \dots, N \tag{5}$$

$$y_i = (Q_i - Q_{\min}) / (Q_{\max} - Q_{\min}) \quad i=1, 2, \dots, N \tag{6}$$

式中:  $x_i$ ,  $y_i$  分别为处理后的作为输入和输出的训练样本, 均为比值, 无量纲;  $Q_{\min}$ ,  $Q_{\max}$  为数据中流量的最小、最大值;  $Q_{\min}$ ,  $Q_{\max}$ ;  $H_{\min}$ ,  $H_{\max}$  为数据中水位的最小、最大值。

利用所得的训练样本, 通过遗传程序进行模型的自动寻优, 即可找出理想的水位流量函数。利用所得函数计算出的是标准化后的流量, 根据(6)转化回实际的即可。

用残差平方和、平均绝对误差以及平均相对误差来衡量, 可由下列各式计算:

$$Err_1 = \sum_{i=1}^N (\hat{Q}_i - Q_i)^2 \tag{7}$$

$$Err_2 = \frac{1}{N} \sum_{i=1}^N |\hat{Q}_i - Q_i| \tag{8}$$

$$Err_3 = \frac{1}{N} \sum_{i=1}^N |\hat{Q}_i - Q_i| / Q_i \tag{9}$$

式中:  $Err_1$ ,  $Err_2$ ,  $Err_3$  分别为残差平方和、平均绝对误差以及平均相对误差;  $\hat{Q}_i$  为利用模型计算出流量  $\text{m}^3/\text{s}$ 。误差越小, 拟合的精度越高。

3 算 例

某水文站, 共有 13 组水位流量的原始观测数据, 如表 2 所示, 假设水位流量关系符合(4)的形式, 采用最小二乘法, 结合残差平方和最小准则以及采用遗传算法、蚁群算法, 结合绝对残差绝对值和最小准则计算模型参数, 然后同利用遗传程序自动优化的模型结果进行比较, 如表 3 所示。遗传程序计算的流量见表 2 第三行。

表 2 水位流量关系拟合数据

水位/ m	15. 50	14. 90	14. 10	14. 55	12. 60	12. 47	12. 67	8. 30	11. 40	10. 30	10. 70	9. 48	7. 77
实际流量/( $\text{m}^3 \cdot \text{s}^{-1}$ )	596	561	542	574	435	433	448	204	372	309	331	258	182
计算流量/( $\text{m}^3 \cdot \text{s}^{-1}$ )	595	563	521	545	438	431	442	205	370	308	330	263	182

表 3 各种模型拟合结果比较

模型	残差平方和	平均绝对误差	平均相对误差/ %
遗传程序	1 373. 1	6. 32	1. 33
最小二乘法 <sup>[1]</sup>	2 136. 3	9. 85	2. 10
遗传算法 <sup>[3]</sup>	3 837. 4	8. 76	1. 81
蚁群算法 <sup>[2]</sup>	3 817. 0	8. 76	1. 81

最小二乘法得出的结果为

$$Q = 4. 306 9 \times H^{1. 827 2}$$

遗传算法得出的结果为

$$Q = 4. 363 8 \times H^{1. 822 0}$$

蚁群算法得出的结果为

$$Q = 4. 262 2 \times H^{1. 822 1}$$

从上面计算结果可以看出, 无论哪种评价指标, 遗传程序的拟合效果非常满意。

(下转第 58 页)

此,当预报数据为一确切的实数时,灰色动态规划方法就退化为普通的动态规划方法。表1给出了全供水期的实测流量,取与灰色动态规划方法同样的计算参数,计算结果见表4。

表4 常规动态规划计算结果表

时段	月份	初水位/ m	末水位/ m	发电流量/ ( $\text{m}^3 \cdot \text{s}^{-1}$ )	出力/ 万 kWh
6	11	760.00	760.00	308.0	24 002.65
5	12	760.00	758.81	201.2	15 672.55
4	1	758.81	755.82	204.0	15 638.01
3	2	755.82	748.66	255.3	18 769.50
2	3	748.66	732.00	356.3	23 598.51
1	4	732.00	720.00	358.9	20 745.00

对比普通动态规划方法与灰色动态方法计算的结果,库水位变化过程如图1。

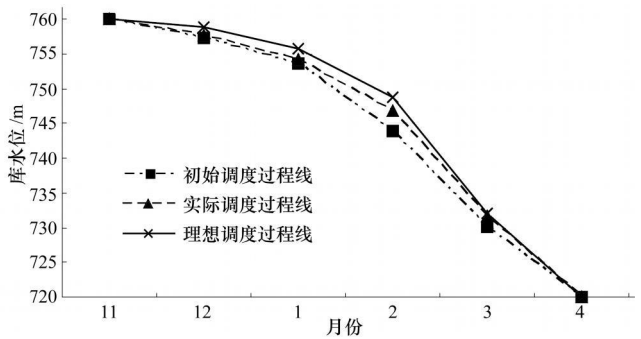


图1 乌江渡水库水位变化过程线

其中:理想调度过程线是预报数据为确切实数的水位变化过程线;初始调度过程线为全供水期初,根据灰色动态规划方法计算的水位变化过程线(表3中第一行数据);实际调度过程线是根据预报调度模型在调度期结束后整个供水期实际的水位变化过程线(表3中有下化线的数据)。实际调度过程线比初始调度过程线更接近理想调度线。从计算结果可知:当预报精度较高时,预报调度的结果同最优结果(确定型)比较接近,预报数据越精确,调度的结果越理想;当实际径流在预报灰数

范围内,实际出力也在预期出力范围内;当实际径流不在预报灰数范围内,实际出力也不在预期出力范围内,实际出力和实际流量在预报灰数的同一侧。

4 与其他动态规划方法的比较

灰色动态规划与其他方法的区别主要表现在径流描述方面。当径流的描述方式采用随机过程及优化调度中需要满足设计保证率时,一般用随机动态规划;模糊动态规划的模糊性主要体现在模糊约束、模糊目标和模糊决策,模糊动态规划的径流描述与普通动态规划方法一样分为确定型和随机型,无论是随机动态规划还是模糊随机动态规划都需要水库的流量频率曲线;灰色动态规划的径流描述则根据不同的径流预报结果、结合当年水文年的类型(如“平水年”、“枯水年”)确定的区间灰数。

与确定型动态规划方法相比,灰色动态规划方法考虑了径流预报的不确定性,更符合工程实际;与随机型动态规划相比,该方法不需要历年径流资料,还考虑了整个计算时段的预报值,充分利用了已知条件,而在随机模型需考虑径流所有可能的概率,扩大了径流可能的范围。

5 结 语

随着科技的发展,径流预报的精度比以前有了较大提高,如何充分利用预报值成了调度工作者的一个新问题。灰色动态规划方法结合径流预报,用区间灰数描述径流的不确定型,即考虑了随机因素,又充分利用预报的已知信息,更符合工程实际。

参考文献:

[1] 徐泽水,达庆利. 区间数的排序方法研究[J]. 系统工程, 2001, 19(11): 94—96.  
[2] 徐泽水,达庆利. 区间数排序的可能度法及其应用[J]. 系统工程学报, 2003, (2): 67—70.  
[3] Weng Wenping. Study on grey linear programming[J]. The Journal of Grey System, 1997.

(上接第55页)

4 结 语

根据遗传程序的基本原理,充分吸收演化计算理论的最新研究成果,建立了水位流量关系拟合模型。该模型避开预先假定具体函数式的局限,通过程序自动搜索模型结构,避免了由于结构不合理而造成拟合效果不够理想的缺陷。利用实际资料,将遗传程序模型与最小二乘法、遗传算法、蚁群算法模型详细地比较,结果表明基于遗传程序设计的水位流量拟合精度最高,而且该方法具有强大的自动搜索模型的能力,使用方便,适应性强,有推广应用的价值。

参考文献:

[1] 黄才安. 水位流量回归的优化研究[J]. 水利水电技术, 1995, 26(10): 2—5.  
[2] 詹世昌,徐 婕. 蚁群算法在水位流量关系拟合中的应用[J]. 杭州师范学院学报(自然科学版), 2005, 4(2): 109—113.  
[3] 杨晓华,陆桂华,郦建强. 自适应加速遗传算法及其在水位流量拟合关系中的应用[J]. 水文, 2002, 22(2): 14—18.  
[4] 唐丽珏,李 森,张 建. 混合 GP-GA 用于信息系统建模预测的研究[J]. 计算机工程与应用, 2004, (25): 44—48.  
[5] 刘健勤. 人工生命理论及其应用[M]. 北京: 冶金工业出版社, 1997.  
[6] 刘大有,卢奕南,王 飞,等. 遗传程序设计方法综述[J]. 计算机研究与发展, 2001, 38(2): 213—222.