

基于空穴序列的灰色预测模型及其应用

教培, 牟龙华

(同济大学电子与信息工程学院, 上海 201804)

摘要: 在实际预测中, 原始数据往往呈现空穴序列, 无法直接应用传统的基于等间距序列数据的灰色模型对其进行预测。提出直接利用有限的、不连续的观测点数据建立 GM(1, 1) 模型, 并利用遗传算法确定模型参数。最后, 以某地区的电力系统负荷为算例, 分别采用该文直接基于原始空穴序列建模和传统的内插数据后再建模, 并将两种模型的预测结果进行了分析比较。结果表明, 与传统内插 GM(1, 1) 模型相比较, 基于空穴序列的 GM(1, 1) 模型的模型精度和预测精度均较高。

关键词: GM(1, 1) 模型; 空穴序列; 遗传算法; 非线性规划; 灰色预测

Grey forecasting model based on vacant data series and application

AO Pei, MU Long-hua

(College of Electrons and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: The original data series are often the vacant data series in actual forecasting. But traditional grey model based on equal interval data series can not be used to predict directly. In this paper, limited and interval data are directly used to build GM(1,1). Genetic algorithm is applied to determine parameter of GM(1,1). At last, power system load of certain region is treated as an example. Models based on vacant data and traditional interpolated data are built respectively. Then, forecasting results of two models are analyzed and compared. It is demonstrated that the accuracy and reliability of GM(1,1) based on vacant data series is higher than traditional interpolated GM(1,1).

Key words: GM(1,1) model; vacant data series; genetic algorithm; nonlinear programming; grey forecasting

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2009)24-0090-04

0 引言

近年来, 基于等时距序列 (即序列中没有空穴) 的 GM(1, 1) 模型, 在预测领域中得到了较为广泛的应用。而在一些实际问题中, 由于各种主客观因素的影响, 原始数据有可能出现不连续的空穴序列。当空穴较多时, 传统的采用均值生成、级比生成或差值生成等局部生成的方式填补空穴后建立的模型, 模型的精度与可信度较低, 预测结果的精度及可靠性不高。本文提出直接利用有限的、不连续的观测点数据建立 GM(1, 1) 模型, 并利用遗传算法确定模型参数。最后, 以某地区电力系统负荷数据为算例, 将本文提出的模型与传统内插补缺 GM(1, 1) 模型的模型精度和预测精度进行了分析比较。结果表明, 该模型克服了数据缺失对负荷预测造成的不利影响, 提高了预测精度与预测可信度。

1 灰色预测模型

本文讨论最常用的灰色预测模型——GM(1,

1) 模型^[1], 它是基于连续等间距时间数列建立的只包含单变量的一阶微分方程, 具体模型如下:

设有变量为 $x^{(0)}$ 的原始数据序列

$$x^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)]$$

用 1-AGO 生成一阶累加生成序列

$$x^{(1)} = [x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)]$$

其中:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (1)$$

由于序列 $x^{(1)}$ 具有指数增长规律, 而一阶微分方程的解恰是指数增长形式的解, 因此, 构造下述一阶线性微分方程模型

$$\frac{dx^{(1)}}{dt} x^{(1)} + \hat{a}x^{(1)} = \hat{u} \quad (2)$$

其中: \hat{a} 为模型的发展系数, \hat{u} 为模型的协调系数。利用最小二乘法求解参数 \hat{a} 、 \hat{u} 有

$$\hat{A} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T Y_n \quad (3)$$

其中:

$$B = \begin{bmatrix} \frac{1}{2}[x^{(0)}(1)+x^{(0)}(2)] & 1 \\ \frac{1}{2}[x^{(0)}(2)+x^{(0)}(3)] & 1 \\ \vdots & \vdots \\ \frac{1}{2}[x^{(0)}(n-1)+x^{(0)}(n)] & 1 \end{bmatrix}, Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

$x^{(1)}$ 的灰色预测模型为

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{\hat{u}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \quad (4)$$

$(k = 0, 1, 2, \dots, n-1)$

为判别模型优劣, 可使用后验差检验法^[3].

$$\text{令 } \varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k) \quad (k = 1, 2, \dots, n)$$

$$\bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^n \varepsilon(k), \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$$

$$S_1^2 = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{x}]^2$$

$$S_2^2 = \frac{1}{n} \sum_{k=1}^n [\varepsilon^{(0)}(k) - \bar{\varepsilon}]^2$$

$$\text{若 } C = \frac{S_2}{S_1} < 0.35, P = P\{|\varepsilon(k) - \bar{\varepsilon}| < 0.6745S_1\} \geq 0.95$$

则认为负荷预测模型满足一级精度。

2 遗传算法

遗传算法^[4] GA (Genetic Algorithm) 模拟生命进化机制, 从任意一个初始种群出发, 通过随机选择、交叉和变异操作, 不断进行适应环境的繁殖、进化, 最后收敛到一群最适应环境的个体上, 求得问题的最优解。GA 具有广泛的可行解表示、采用内在启发式群体随机搜索、无需适应度函数数值以外的辅助信息、不易陷入局部最优解等特点。目前, 遗传算法作为一种全新的优化方法, 已经在约束规划^[5]、非线性规划^[6]、控制系统优化^[7]等方面得到广泛的应用。遗传算法的主要实现步骤如下:

(1) 编码: GA 在求解之前, 先将问题解空间的可行解表示成遗传空间的基因型串结构数据, 串结构数据的不同组合构成了不同的可行解;

(2) 生成初始群体: 随机产生 N 个初始串结构数据, 每个串结构数据称为一个个体, N 个个体构成一个群体, GA 以该群体作为初始迭代点;

(3) 适应性值评估检测: 根据实际标准计算个体的适应度, 评判个体的优劣, 即该个体所代表的可行解的优劣;

(4) 选择: 从当前群体中选择优良的个体, 使其作为父代以大概率繁殖后代;

(5) 交叉: 组合其父辈个体特性, 得到新一代个体;

(6) 变异: 在群体中随机选择一个个体, 以

一定概率随机地改变串结构数据中某个串的值。

本文首先将求解基于空穴序列 GM(1, 1) 模型参数的问题转化为求解非线性规划的最优参数问题, 然后利用遗传算法求解该非线性规划问题的最优参数。

3 基于空穴序列的灰色预测模型

第 1 部分所述的 GM(1, 1) 模型基本关系, 首先要求做一次累加生成 (1-AGO), 当数据存在空穴时, 这种形式的累加便无法实现, 这时可以假设: 连续等时距的原始数据是客观存在的, 只是由于某些原因使其中的一些数据缺失, 因而出现了不连续的空穴数列。进一步假设, 已经得到了与这些原始数据较为符合的 GM(1, 1) 模型曲线, 曲线的离散形式如式 (4) 所示, 这里设 $\hat{c} = x^{(0)}(1) - \frac{\hat{u}}{\hat{a}}$, 则式

(4) 可改写为

$$\hat{x}^{(1)}(k+1) = \hat{c}e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \quad (k = 0, 1, 2, \dots) \quad (5)$$

还原后原始数据估计值为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = \hat{c}e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} - [\hat{c}e^{-\hat{a}(k-1)} + \frac{\hat{u}}{\hat{a}}] = \hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}k} \quad (6)$$

$$(k = 0, 1, 2, \dots, n-1)$$

设有空穴数列, 初始时间序列值为 0, 时间序列 $T^{(0)} = \{0, t_2, t_3, \dots, t_m\}$, 则有:

$$\hat{x}^{(0)}(t_i) = \hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}t_i} \quad (7)$$

$$(t_i = t_2, t_3, \dots, t_m, m \text{ 为原始数列个数})$$

这里, 最终是要确定上式中参数 \hat{c} 、 \hat{a} 的值, 只要找到适当的 \hat{c} 、 \hat{a} 的值, 就可以建立相应的灰色系统模型。

由于期望得到的是比较理想的预测模型, 即要求估计值 $\hat{x}^{(0)}(t_i)$ 与原始值 $x^{(0)}(t_i)$ 越接近越好。因此, 令 $\hat{x}^{(0)}(t_i)$ 与原始值 $x^{(0)}(t_i)$ 的偏差平方和最小, 即

$$\min F(\hat{c}, \hat{a}) = \sum_{i=2}^m (\hat{x}^{(0)}(t_i) - x^{(0)}(t_i))^2 = \sum_{i=2}^m (\hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}t_i} - x^{(0)}(t_i))^2 \quad (8)$$

此外, 对于所有的原始数据, 还应满足近似方程组

$$\begin{cases} x^{(0)}(t_2) \approx \hat{x}(t_2) = \hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}t_2} \\ x^{(0)}(t_3) \approx \hat{x}(t_3) = \hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}t_3} \\ \vdots \\ x^{(0)}(t_m) \approx \hat{x}(t_m) = \hat{c}(1 - e^{-\hat{a}})e^{-\hat{a}t_m} \end{cases} \quad (9)$$

不难看出，问题最终转化为在等式 (9) 约束下，求解使目标函数式 (8) 最小化的参数 \hat{c} 、 \hat{a} 的非线性规划问题，即

$$\begin{aligned} \min F(\hat{c}, \hat{a}) &= \sum_{i=2}^m (\hat{x}^{(0)}(t_i) - x^{(0)}(t_i))^2 = \\ & \sum_{i=2}^m (\hat{c}(1 - e^{\hat{a}})e^{-\hat{a}t_i} - x^{(0)}(t_i))^2 \quad (10) \\ \text{s.t. } x^{(0)}(t_i) &= \hat{x}(t_i) = \hat{c}(1 - e^{\hat{a}})e^{-\hat{a}t_i} \\ & i = 2, 3, \dots, m \end{aligned}$$

运用惩罚策略^[8]，选择惩罚函数^[9]为

$$p(\hat{c}, \hat{a}) = \begin{cases} 0 & \text{若 } x \text{ 可行} \\ \sum_{i=2}^m r_i \hat{x}(t_i) = \sum_{i=2}^m r_i \hat{c}(1 - e^{\hat{a}})e^{-\hat{a}t_i} & \text{其他} \end{cases} \quad (11)$$

进而，将式 (10) 的约束问题转化为无约束问题，即

$$\begin{aligned} \min(F(\hat{c}, \hat{a}) + p(\hat{c}, \hat{a})) &= \\ \sum_{i=2}^m (\hat{x}^{(0)}(t_i) - x^{(0)}(t_i))^2 + p(x) &= \\ \sum_{i=2}^m (\hat{c}(1 - e^{\hat{a}})e^{-\hat{a}t_i} - x^{(0)}(t_i))^2 + p(x) \end{aligned} \quad (12)$$

最后，采用遗传算法求解最优化参数 \hat{c} 、 \hat{a} 。

4 算例分析

本文以某地区 1990~2005 年的电力系统负荷数据为算例，假设除 1991、1994、1998 和 2001 年数据外，其它年份数据缺失（如表 1），分别采用传统内插补缺等时距 GM(1, 1) 模型和本文基于空穴序列的 GM(1, 1) 模型对该地区 2002~2005 年的电力系统负荷进行预测。

表 1 某地区全社会用电量

Tab.1 Total social quantity of electricity in certain region

年份	1991	1994	1998	2001
全社会用电量 / (万 kWh)	1 428 794	2 236 449	3 201 954	44 391 860

首先，采用传统的内插法填补原始数据空穴，生成等时距序列，然后建立 GM(1,1) 模型，得 $\hat{c}1=15038167$ ， $\hat{a}1=-0.10079$ 。

然后，采用本文提出的直接利用空穴序列建立 GM(1, 1) 模型。在利用遗传算法求解 \hat{c} 、 \hat{a} 的过程中，初始种群个体数目选定为 40 个，根据精度要求， \hat{c} 、 \hat{a} 分别选用 23 位和 14 位格雷码进行编码，选用随机遍历抽样算法以 0.9 的概率选择适应度较高的个体到下一代群体，采用多点交叉方法以 0.7 的概率产生新个体，变异概率取为 0.001。经过 1 000

代进化后，得到 $\hat{c} = 14003125$ 、 $\hat{a} = -0.10469$ ，对所得灰色模型采用后验差检验法进行检验，得 $C=0.08 < 0.35$ ， $P=1 > 0.95$ ，模型精度等级为一级，能够较好地进行预测。

两种模型在 1991、1994、1998 和 2001 年电力负荷数据点的相对误差见表 2。从表 2 中不难看出，采用传统内插补缺等时距 GM(1, 1) 模型，平均相对误差为 -3.9482%，原点误差达到 -2.8766%；而采用本文基于空穴序列的 GM(1, 1) 模型，平均相对误差为 -0.8610%，原点误差仅为 0.2436%。

表 2 两种模型的精度比较

Tab.2 Comparison of accuracy between two models

年份 \ 模型	基于空穴序列 GM(1, 1) 模型相对误差 / (%)	传统内插补缺等时距 GM(1, 1) 模型相对误差 / (%)
1991	-8.163 9	-11.296 6
1994	5.400 5	1.868 8
1998	-0.437 1	-3.488 5
2001	0.243 6	-2.876 6

最后，分别应用传统模型和本文模型对 2002~2005 年电力负荷值进行预测，预测结果如表 3。可以看出，采用传统内插补缺等时距 GM(1, 1) 模型，平均相对误差为 -2.6675%，而采用本文基于空穴序列的 GM(1, 1) 模型，平均相对误差为 -0.3313%。

表 3 两种模型预测精度比较

Tab.3 Comparison of forecasting accuracy between two models

年份	实际值 / (万 kWh)	基于空穴序列的 GM(1, 1) 模型		传统内插补缺等时距 GM(1, 1) 模型	
		预测值 / (万 kWh)	相对误差 / (%)	预测值 / (万 kWh)	相对误差 / (%)
		2002	4 968 387	4 888 671	1.604 5
2003	5 476 290	5 428 216	0.877 9	5 558 833	-1.507 3
2004	5 984 194	6 027 308	-0.720 5	6 163 298	-2.993 0
2005	6 492 097	6 692 520	-3.087	6 833 492	-5.258 6

以上算例结果表明，本文基于空穴序列的 GM(1, 1) 模型的模型精度和预测结果优于传统内插补缺等时距 GM(1, 1) 模型。

5 结论

在原始数据空穴较多时，本文提出的直接利用有限的、不连续的观测点数据建立 GM(1, 1) 模型，克服了传统的采用均值生成、级比生成或差值生成等局部生成的方式填补空穴后建模方法模型精度差和预测结果可信度较低的问题。通过算例验证，基于空穴序列的 GM(1, 1) 模型的模型精度和预测精度均高于传统内插补缺等时距 GM(1, 1) 模型。

参考文献

- [1] 邓聚龙.灰色系统基本方法(第2版)[M].武汉:华中科技大学出版社,2005.
DENG Ju-long. The Primary Methods of Grey System Theory[M].Wuhan: Huazhong Univeristy of Science and Technology,2005.
- [2] WEI Li, HAN Zhu-hua, NIU Dong-xiao. Improved Genetic Algorithm-GM(1,1) for Power Load Forecasting Problem[A].in: Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies[C]. 2008.1147-1152.
- [3] ZHOU Jiang-wen. Mathematical Processing of Systematic Error[A]. in:Proceedings of New Exploration of Surveying Error Theory[C].Beijing: Seismological Press, 1999.
- [4] 王小平, 曹立明.遗传算法——理论、应用与软件实现[M].西安: 西安交通大学出版社, 2002.
WANG Xiao-ping, CAO Li-ming. Genetic Algorithm[M]. Xi'an: Xi'an Traffic University Press, 2002.
- [5] Homaifar A, Lai S H Y, Qi X. Constrained Optimization Via Genetic Algorithms[J]. Simulation, 1994, 62, 242-254.
- [6] Zbigniew Michalewicz, Martin Schmidt. TCG-2: A Test-case Generator for Non-linear Parameter Optimization Techniques[M]. Hcidelberg, Germany: Springer, 2003.193-212.
- [7] Hansen J V. Genetic Search Methods in Air Traffic Control[J]. Computers and Operations Research,2004,31((3):445-459.
- [8] 杨新武, 刘椿年.基于演化周期的惩罚策略[J].计算机工程与应用,2003,17.
YANG Xin-wu, LIU Chun-nian. The Penalty Strategy Based on Evolution Period[J]. Computer Engineering and Application, 2003, 17.
- [9] 胡旭晓, 潘晓弘, 何卫, 等.惩罚函数的构造及多模态平稳过渡策略[J].机械工程学报, 2008, 44(4).
HU Xu-xiao, PAN Xiao-hong, HE Wei, et al. Construction of Penalty Function and Multi-modal Smooth Transition Strategy[J]. Chinese Journal of Mechanical Engineering, 2008, 44(4).
- 收稿日期: 2008-12-16; 修回日期: 2009-03-11
作者简介:
教培(1979-), 女, 博士研究生, 助理研究员, 研究方向为智能测试理论与装置. E-mail:aopei16.student@sina.com
车龙华(1963-), 男, 教授, 博导, 研究方向为电力系统保护、智能电器和电能质量.
- (上接第89页 continued from page 89)
Technology Report of Research and Application on Substation Secondary System[Z].Henan Electric Power Corporation,XJGroup Corporation,2008.
- [11] 周江龙.电力操作闭锁方法及其分布式智能实现模式(硕士学位论文)[D].湘潭:湘潭大学,2008.
ZHOU Jiang-long. Research of On-line Interlocking and its Distriuted Intelligent Implementation Method,Thesis[D]. Xiangtan: Xiangtan University, 2008.
- [12] 窦晓波.与GOOSE联动的数字化变电站巡视系统[J].电力自动化设备,2008.
DOU Xiao-bo. Remote Superviwer System Accompany with GOOSE Information[J]. Electric Power Automation Equipment, 2008.
- [13] GOOSE 实时通信的分析与实现[J].电力系统保护与控制,2009,38.
Analysis and Realization of GOOSE Real-time Communication[J]. Power System Protection and Control ,2009,38.
- [14] 基于IEC 61850的数字化继电保护GOOSE功能测试[J].电力系统保护与控制,2008,36(7).
GOOSE Test of Digital Relays Based on IEC 61850[J]. Power System Protection and Control, 2008,36(7).
- [15] 数字化(智能)变电站调研情况总结[Z].国家电网公司内部资料,2009.
Summary report of Investigation and Study on Digital (Smart) Substation[Z]. State Grid,2009.
- 收稿日期: 2009-09-15; 修回日期: 2009-11-24
作者简介:
郑新才(1962-), 男, 高级工程师, 从事继电保护研究工作; E-mail:zhengxincai22@126.com
周鑫(1979-), 男, 助工, 从事继电保护检修安装调试工作;
王素华(1971-), 女, 继电保护技师, 从事继电保护工作.