

基于数据变换的优化 GM (1,1) 模型

汤旻安^{1,2}, 李 滢¹

(1. 兰州交通大学自动化与电气工程学院, 甘肃 兰州 730070)

(2. 兰州理工大学机电工程学院, 甘肃 兰州 730050)

摘 要: 本文研究了提高灰色 GM (1,1) 模型预测精度的问题. 利用复合函数变换对原始数据序列经过一定处理的基础上同时优化模型的背景值和初始值的方法, 获得了比改进单个模型条件更高预测精度的 GM (1,1) 模型, 推广了灰色预测模型的适用范围.

关键词: 灰色理论; GM (1,1) 模型; 数据变换; 初始条件; 背景值

MR(2010) 主题分类号: 49N30

中图分类号: O175.7

文献标识码: A

文章编号: 0255-7797(2015)04-0957-06

1 引言

灰色模型作为灰色系统理论的重要内容之一, 在多个领域得到了广泛的应用. 尤其是 GM (1,1) 模型, 以其所需建模样本少, 计算简单方便等特点, 因而比其他传统预测方法更具优越性. 为提高模型的拟合和预测精度, 许多学者从提高原始数据序列的光滑度, 改进模型的参数估计方法, 优化背景值和初始条件等方面做了大量广泛的研究.

对改进灰色预测模型的研究, 为提高建模序列的光滑度文献 [1-9] 分别提出了对数函数变换、正弦函数变换、余弦函数变换、幂函数变换、正切函数变换、正切 - 指数函数变换、幂函数 - 指数函数变换、负指数函数变换等, 这些变换均能在一定程度上提高原始序列的光滑度, 并在实际应用中取得了比较满意的结果. 由 GM (1,1) 模型的时间响应式可以看出影响模型模拟和预测精度的因素取决于初始值的选取和参数 a, b 的值, 而参数的值又依赖于背景值, 因此, 初始值的选取和背景值的构造是否合理直接影响到模型的模拟和预测精度. 传统的 GM (1,1) 模型以历史数据中的第一个数据作为初始条件, 这就等于拟合曲线必须经过历史数据中的第一个数据点, 但这并不一定符合事实. 文献 [10-17] 从改进背景值和初始条件方面研究了提高模型预测精度的方法. 然而, 单方面改进对模型精度的贡献有限, 同时, 在多方面优化可使模型精度有更大的提高. 本文在前人研究的基础上以文献 [9] 提出的函数 $f(x) = c \ln x + d$ 变换技术和文献 [14] 提出的将 $x^{(1)}$ 的第 n 个分量作为灰色微分模型的初始条件并将原背景值 $z^{(1)}(k)$ 优化为 $z^{(1)}(k) = \frac{x^{(1)}(k) - x^{(1)}(k-1)}{\ln x^{(1)}(k) - \ln x^{(1)}(k-1)}$ 的方法相结合对 GM (1,1) 模型进行了改进, 从而使所建模型精度进一步提高.

2 传统灰色预测模型的建立

*收稿日期: 2013-10-17

接收日期: 2014-09-03

基金项目: 国家自然科学基金资助 (61263004); 甘肃省科技支撑计划项目基金资助 (090GKCA009; 1304GKCA023); 兰州市科技攻关计划项目基金资助 (2013-4-18).

作者简介: 汤旻安 (1973-), 男, 陕西勉县, 副教授, 主要研究方向: 交通信息工程及控制, 智能交通.

设有非负原始数据序列

$$x^{(0)} : x^{(0)}_{(k)} = \{x^{(0)}_{(1)}, x^{(0)}_{(2)}, x^{(0)}_{(3)}, \dots, x^{(0)}_{(n)}\}, \quad (2.1)$$

其中, $x^{(0)}_{(k)} \geq 0, k = 1, 2, \dots, n$; $x^{(1)}$ 为 $x^{(0)}$ 的一次累加生成数列

$$x^{(1)}_{(k)} = \{x^{(1)}_{(1)}, x^{(1)}_{(2)}, x^{(1)}_{(3)}, \dots, x^{(1)}_{(n)}\}, \quad (2.2)$$

其中, $x^{(1)}_{(k)} = \sum_{i=1}^k x^{(0)}_{(i)}, k = 1, 2, \dots, n$; $z^{(1)}$ 为 $x^{(1)}$ 的紧邻均值生成序列

$$z^{(1)}_{(k)} = \{z^{(1)}_{(2)}, z^{(1)}_{(3)}, \dots, z^{(1)}_{(n)}\}, \quad (2.3)$$

其中 $z^{(1)}_{(k)} = \frac{1}{2}(x^{(1)}_{(k)} + x^{(1)}_{(k-1)}), k = 2, 3, \dots, n$.

建立关于 $x^{(1)}$ 的一阶白化微分方程: $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$, 并用最小二乘法求解参数 a 和 b .

$$\hat{\alpha} = (B^T B)^{-1} B^T Y_N = [a, b]^T, \quad (2.4)$$

式中 a, b 为待识别参数. a 为灰系数, b 为灰作用量, 分别反映了原始数据序列的增长速度和数据变化关系.

$$B = \begin{bmatrix} -z^{(1)}_{(2)} & 1 \\ -z^{(1)}_{(3)} & 1 \\ \dots & \dots \\ -z^{(1)}_{(n)} & 1 \end{bmatrix}, \quad Y_N = \begin{bmatrix} -x^{(0)}_{(2)} \\ -x^{(0)}_{(3)} \\ \dots \\ -x^{(0)}_{(n)} \end{bmatrix}. \quad (2.5)$$

a, b 确定后求解微分方程可得到 GM (1,1) 模型:

$$\hat{x}^{(1)}_{(k+1)} = \left(x^{(0)}_{(1)} - \frac{b}{a}\right) e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n. \quad (2.6)$$

其还原模型为

$$\hat{x}^{(0)}_{(k+1)} = (1 - e^a) \left(x^{(0)}_{(1)} - \frac{b}{a}\right) e^{-ak} = \hat{x}^{(1)}_{(k+1)} - \hat{x}^{(1)}_{(k)}, k = 1, 2, \dots, n. \quad (2.7)$$

3 优化 GM (1,1) 模型

3.1 原始数据序列的数据变换

对原始数列经过某种变换可以减小光滑比, 但不适当的变换虽可以满足光滑性要求却不一定获得较高的精度. 文献 [9] 在保证构造的数据变换满足减小光滑比、调节级比压缩、保持序列凹凸性不变、还原误差不增大等条件下, 提出非负 $f(x) = c \ln x + d$ 函数变换, 其中 $c \geq \max\{x_{(k)}, k = 1, 2, \dots, n\}, x_{(k)} > e$, 并证明利用该函数变换对原始数据序列进行处理可以保持在原有数据序列基本规律不变的情况下, 提高数据序列的光滑度.

3.2 背景值和初始值的优化

对背景值的优化文献 [10] 提出用 $z^{(1)}(k) = \frac{x^{(1)}(k) - x^{(1)}(k-1)}{\ln x^{(1)}(k) - \ln x^{(1)}(k-1)}$ 代替 $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1))$ 从而减小由原背景值产生的误差, 可在一定程度上提高建模精度. 而初始值也是影响 GM (1,1) 模型模拟与预测精度的另一个重要因素. 文献 [13] 基于灰色系统理论的新信息优先原理和最少信息原理, 将原 GM 模型的初始条件由 $x^{(0)}(1)$ 改为以 $x^{(1)}$ 的第 n 个分量 $x^{(1)}(n)$ 作为灰色模型的初始条件, 从而使得原始序列的信息通过 $x^{(1)}(n)$ 都可以得到充分反映, 因而可以克服 GM 建模与 $x^{(0)}(1)$ 无关的弊端. 由于背景值优化和初始值的改进均能使得模型的精度提高, 并且二者相互独立. 因此, 可先对背景值进行优化再将原模型 $\hat{x}_{(k+1)}^{(0)} = (1 - e^a)(x^{(0)}(1) - \frac{b}{a})e^{-ak}$ 中的 $x^{(0)}(1)$ 换成 $x^{(1)}(n)$ 即得到新的时间响应序列 $\hat{x}_{(k+1)}^{(0)} = (1 - e^a)(x^{(1)}(n) - \frac{b}{a})e^{-ak}$. 文献 [14] 研究了采用初始条件与背景值同时优化的方法, 得到改进后的新模型既适用于低增长指数序列建模, 也适用于高增长指数序列建模, 且拟合精度都很高.

3.3 优化后 GM (1,1) 模型的建模过程

新的 GM (1,1) 模型的建模过程如下:

(1) 设有非负原始数据序列 $y^{(0)}: y^{(0)}(k) = \{y^{(0)}(1), y^{(0)}(2), y^{(0)}(3), \dots, y^{(0)}(n)\}, k = 1, 2, \dots, n$. 对序列 $y^{(0)}$ 做一次累加得数据序列 $y^{(1)}: y^{(1)}(k) = \{y^{(1)}(1), y^{(1)}(2), y^{(1)}(3), \dots, y^{(1)}(n)\}, k = 1, 2, \dots, n$.

(2) 对原始数列做光滑性和准指数规律检验, 建模可行性分析. 光滑性检验公式: $\rho(k) = \frac{y^{(0)}(k)}{y^{(1)}(k-1)}, \rho(k) \in (0, 0.5)$. 准指数规律检验公式: $\delta^{(1)}(k) = \frac{y^{(1)}(k)}{y^{(1)}(k-1)}, \delta^{(1)}(k) \in (1, 1.5)$.

(3) 对原始数列做函数 $f(x) = c \ln x + d$ 变换, 得到新数列

$$x^{(0)}: x^{(0)}(k) = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\}, k = 1, 2, \dots, n.$$

(4) 对数列做一次累加生成数列 $x^{(1)}: x^{(1)}(k) = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)\}$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$.

(5) 对 $x^{(1)}$ 做紧邻生成序列 $z^{(1)}: z^{(1)}(k) = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$, 其中 $z^{(1)}(k) = \frac{x^{(1)}(k) - x^{(1)}(k-1)}{\ln x^{(1)}(k) - \ln x^{(1)}(k-1)}, k = 2, 3, \dots, n$.

(6) 由一阶累加生成序列 $x^{(1)}$ 建立 GM (1,1) 模型对应的微分方程: $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ 并用最小二乘法求解参数 a 和 b : $\hat{\alpha} = (B^T B)^{-1} B^T Y_N = [a, b]^T$, 其中

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad Y_N = \begin{bmatrix} -x^{(0)}(2) \\ -x^{(0)}(3) \\ \dots \\ -x^{(0)}(n) \end{bmatrix}.$$

(7) 令初始值为 $x^{(1)}(n)$, 解上述微分方程得到时间响应序列

$$\hat{x}_{(k+1)}^{(1)} = \left(x^{(1)}(n) - \frac{b}{a} \right) e^{-a(k-n)} + \frac{b}{a}, k = 1, 2, \dots, n.$$

(8) 其还原值为 $\hat{x}_{(k+1)}^{(0)} = (1 - e^a) \left(x^{(1)}(n) - \frac{b}{a} \right) e^{-a(k-n)} = \hat{x}_{(k+1)}^{(1)} - \hat{x}_{(k)}^{(1)}, k = 1, 2, \dots, n$.

(9) 由逆函数变换 $\hat{y}_{(k)}^{(0)} = e^{\frac{\hat{x}_{(k)}^{(0)} - d}{c}}$ 得到 $\hat{y}^{(0)}, \hat{y}_{(k)}^{(0)} = (\hat{y}_{(1)}^{(0)}, \hat{y}_{(2)}^{(0)}, \dots, \hat{y}_{(n)}^{(0)})$, $k = 1, 2, \dots, n$.

(10) 用均方差比 C 和小误差概率 p 进行模型精度等级检验. $C = \frac{S_1}{S_2}$, 其中 $S_1^2 = \frac{1}{n} \sum_{k=1}^n (x_{(k)}^{(0)} - \bar{x})^2$, $S_2^2 = \frac{1}{n} \sum_{k=1}^n (\varepsilon_{(k)} - \bar{\varepsilon})^2$, $p = P(|\varepsilon_{(k)} - \bar{\varepsilon}| < 0.6745 S_1)$, 均方差比值越小越好而小误差概率越大越好. 模型精度等级检验表如表 1 所示.

表 1: 模型精度等级检验表

精度等级	均方差比值 C	小误差概率 p
一级 (合格)	$C < 0.35$	$p > 0.95$
二级 (良好)	$0.35 \leq C < 0.5$	$0.8 \leq p < 0.95$
三级 (勉强)	$0.5 \leq C < 0.65$	$0.7 \leq p < 0.8$
四级 (不合格)	$C < 0.8$	$p > 0.6$

4 优化后的 GM (1,1) 在轴承套磨损中的预测

以某轴承套随时间的磨损为例进行预测, 该轴承套已经过磨合期, 且每半月对其做一次磨损测量记录, 得到磨损量数据如表 2 所示.

表 2: 轴承套磨损量

序号	1	2	3	4	5	6	7	8
磨损量 (mm)	19.6	19.9	20.2	20.6	21.0	21.5	22.0	22.7

首先对原始数据序列进行光滑性和准指数规律检验. 当 $k > 3$ 时, 光滑比分别为 0.3451、0.2615、0.2122、0.1792、0.1568, 均在 (0,0.5) 内, 且光滑比递减, 满足准光滑性条件. 当 $k > 3$ 时, 准指数规律比分别为 1.3451、1.2675、1.2122、1.1792、1.1568, 均在 (1,1.5) 内, 因此原始序列的一次累加生成数列具有准指数规律, 可对原始数列建立 GM (1,1) 模型.

4.1 传统 GM (1,1) 模型

建立传统 GM (1,1) 预测模型为 $\hat{x}_{(k+1)}^{(1)} = 142.0473e^{0.0535k} - 134.5028$, $k = 0, 1, \dots, 7$.

4.2 函数变换 GM (1,1) 模型

根据文献 [9] 函数变换技术 $f(x_{(k)}) = c \ln x_{(k)} + d$, 建立 GM (1,1) 模型, 得到变换后的模型:

$$\hat{x}_{(k+1)}^{(1)} = 142.0473e^{0.0535k} - 134.5028, k = 0, 1, \dots, 7.$$

其中, $c = \max\{x_{(k)}^{(0)}\} = 22.7, d = -60$.

4.3 函数变换与初始值和背景值同时优化的 GM (1,1) 模型

本文运用函数变换技术先将原始数据序列进行变换, 后再与背景值和初始值同时改进相结合的方法对传统 GM (1,1) 模型进行优化得到优化后的预测模型:

$$\hat{x}_{(k)}^{(1)} = 206.6667e^{0.0536(k-8)} - 134.5299, k = 1, 2, \dots, 8.$$

4.4 检验模型精度

由上述三种方法得出的结果进行对比分析如表 3 所示.

表 3: 三种模型的模拟值及相对误差比较

序号	测量值	原始模型 模拟值	相对误差 %	文 [9] 方法 模拟值	相对误差 %	本文方法 模拟值	相对误差 %
1	19.6						
2	19.9	19.7653	0.6767	19.8275	0.3643	19.8389	0.3070
3	20.2	20.2029	0.0144	20.2058	0.0287	20.2185	0.0916
4	20.6	20.7202	0.5837	20.6128	0.0621	20.6270	0.1311
5	21.0	21.0375	0.1784	21.0500	0.2381	21.0668	0.0206
6	21.5	21.6448	0.6736	21.5227	0.1056	21.5419	0.1949
7	22.0	21.9825	0.0795	22.0341	0.1550	22.0541	0.2459
8	22.7	22.5407	0.7013	22.5837	0.5141	22.6079	0.4057
平均相对误差 %			0.4154		0.2097		0.1995

注 表中平均相对误差为百分数的绝对值.

对比表 3 中的数据可以看出, 本文提出的方法所得的平均相对误差较原始模型和文献 [9] 提出的方法得到的误差小且该新模型预测精度达到 99.8% 以上. 新模型的均方误差比 $C = 0.126871642 < 0.35$, 小误差概率 $p = 1 > 0.95$. 根据表 1 可知模型精度属于一级. 由此可见, 本文提出的方法具有较高的拟合和预测精度, 并具有一定的实用价值.

4.5 三种方法预测值与实际值对比图

三种方法所得预测值与实际值比较如图 1 所示:

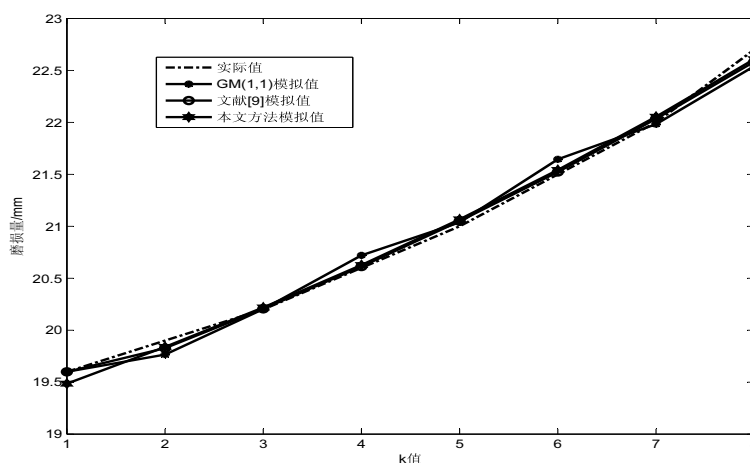


图 1: 三种 GM (1,1) 方法预测值与实际值比较

5 结语

本文提出在对原始数据序列进行函数变换, 提高原始数据序列光滑度的基础上, 对模型背景值和初始值同时进行优化, 得到改进后的 GM (1,1) 模型. 最后应用实例将改进后的模型

与改进单一条件模型相比较, 验证了改进后的新模型具有比改进单一条件模型误差小, 精度高的优点, 从而说明了该模型的实用性和有效性.

参 考 文 献

- [1] 陈涛捷. 灰色预测模型的一种拓广 [J]. 系统工程, 1990, 8(4): 50–52.
- [2] 曹昶, 樊重俊, 胡兆龙. 基于正弦函数变换的灰色预测模型研究及其应用 [J]. 数学杂志, 2013, 33(4): 697–701.
- [3] 郑锋, 魏勇. 提高灰建模数据光滑度的一种新方法 [J]. 统计与决策, 2013, 9: 37–38.
- [4] 李群. 灰色预测模型的进一步拓广 [J]. 系统工程理论与实践, 1993, 13(1): 64–66.
- [5] 李翠凤, 戴文战. 基于函数 $\cot x$ 变换的灰色建模方法 [J]. 系统工程, 2005, 23(3): 110–114.
- [6] 关叶青, 刘思峰. 基于函数 $\cot x^a$ 变换的灰色 GM (1,1) 建模方法 [J]. 系统工程, 2008, 26(9): 89–93.
- [7] 陈洁, 许长新. 灰色预测模型的改进 [J]. 辽宁师范大学学报 (自然科学版), 2005, 28(3): 262–264.
- [8] 何斌, 蒙清. 灰色预测模型拓广方法研究 [J]. 系统工程理论与实践, 2009, 22(9): 138–141.
- [9] 崔立志, 刘思峰. 基于数据变换技术的灰色预测模型 [J]. 系统工程, 2010, 28(5): 104–107.
- [10] 罗党, 刘思峰, 党耀国. 灰色模型 GM (1,1) 优化 [J]. 中国工程科学, 2003, 5(8): 50–53.
- [11] 廖飞. 对背景值优化的新 GM (1,1) 模型 [J]. 数学的实践与认识, 2009, 39(18): 107–113.
- [12] 骆公志, 崔杰, 谢乃明. 灰色 GM (1,1) 模型新的改进方法 [J]. 统计与决策, 2008, (22): 11–13.
- [13] 党耀国, 刘思峰, 刘斌. 以 $x^{(1)}_{(n)}$ 为初始条件的 GM 模型 [J]. 中国管理科学, 2005, 13(1): 132–134.
- [14] 张怡, 魏勇, 熊常伟. 灰色模型 GM (1,1) 的一种新优化方法 [J]. 系统工程理论与实践, 2007, 27(4): 141–146.
- [15] 李俊峰, 戴文战. GM (1,1) 改进模型的研究及在上海市发电量建模中的应用 [J]. 系统工程理论与实践, 2005, 3: 140–144.
- [16] 尹方平. 改进的以 $x^{(1)}_{(n)}$ 为初始条件的 GM 模型 [J]. 统计与决策, 2001, 3: 164–165.
- [17] 张彬, 西桂权. 基于背景值和边值修正的 GM (1,1) 模型优化 [J]. 系统工程理论与实践, 2013, 33(3): 682–688.

OPTIMIZED GM (1,1) MODEL BASED ON DATA TRANSFORMATION TECHNOLOGY

TANG Min-an ^{1,2}, LI Ying ¹

(1. School of Automation and Elec. Engin., Lanzhou Jiaotong University, Lanzhou 730070, China)

(2. School of Mechanical and Elec. Engin., Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: In this paper, we study the accuracy of grey GM (1,1) forecasting model improvement. Using the composite function transformation to deal with the original data sequence, and optimizing the background value and initial value of the model, we obtain a grey forecasting model which has a higher prediction accuracy than single condition improvement GM (1,1) model. The study extends the scope of GM (1,1) forecasting model.

Keywords: grey theory; GM (1,1) model; data transformation; initial condition; background value

2010 MR Subject Classification: 49N30