

文章编号: 1001-0920(2003)04-0507-03

## 灰色模型在春运客流量预测应用中的优化

秦元庆<sup>1</sup>, 陈少鸿<sup>2</sup>, 孙德宝<sup>1</sup>, 郭和伟<sup>3</sup>

(1. 华中科技大学 控制科学与工程系, 湖北 武汉 430074; 2 羊城铁路总公司 广州分公司, 广东 广州 510100; 3 孝感职业技术学院 科技产业总公司, 湖北 孝感 432100)

**摘要:** 将火车站视为本征性灰色系统, 选用灰色预测方法对春运客流量作出预测, 并根据实际预测结果对基本模型进行优化。采用残差预测模型和新陈代谢模型等方法, 解决了预测精度和可信度的问题。实际数据表明预测结果是可信的。

**关键词:** 客流量; 灰色系统; 预测模型; 残差; 优化

**中图分类号:** TP14 **文献标识码:** A

## Optimization of gray model of the passenger flux forecast during the spring festival

QIN Yuan-qing<sup>1</sup>, CHEN Shao-hong<sup>2</sup>, SUN De-bao<sup>1</sup>, GUO Hewei<sup>3</sup>

(1. Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; 2 Guangzhou Subcompany, Yangcheng Railway Company, Guangzhou 510100, China; 3 Science and Technology Industry Company, Xiaogan Vocational Technology College, Xiaogan 432100, China)

**Abstract:** According to the analysis of the passenger flux data, the railway station is taken as an essential gray system. Therefore, the gray forecast method is used to predict the passenger flux during the spring festival. Several measures are used to optimize the basal forecast model, such as choosing the optimized parameter in the error-corrected model, using new information model etc. The actual data indicate that the forecasting result is reliable.

**Key words:** Passenger flux; Gray system; Forecast model; Remained error; Optimize

### 1 引言

由于南下民工和南方高校学生数量极大, 春节期间大部分民工和学生要返乡过节, 极易形成客流高峰, 这使得担任北上客流输送和南下客流接待任务的广州火车站每年春运都面临一次严峻考验。如果能根据往年的客流数据, 运用合理的预测方法找出其中的规律, 对下一年的春运客流量做出科学合理的推断, 就会给车站以充分的决策支持, 使他们能提前作好应付客流高峰的准备, 最大限度地减少旅客滞留人数, 提高服务质量。

本文选用灰色预测方法, 是考虑到火车站本身的特殊情况: 影响春运客流的因素太多, 无法一一列举分析, 而且所知的仅是往年的客流记录以及车站的输送情况, 各因素的作用机制根本无法用精确的数学模型来描述。如果把火车站作为一个系统来分析, 那么这个系统是本征性的灰色系统, 即只有部分信息是已知的系统<sup>[1]</sup>。由邓聚龙教授提出的灰色预测理论, 对解决灰色系统的趋势预测问题具有独特的优势: 它所需数据量小, 建模过程简单, 预测精度高, 现已在许多领域得到了广泛的应用, 尤其是在社

收稿日期: 2002-04-09; 修回日期: 2002-06-18。

作者简介: 秦元庆(1977—), 男, 山东临沂人, 硕士生, 从事模型预测、信号处理等研究; 孙德宝(1941—), 男, 湖北云梦人, 教授, 博士生导师, 从事人工智能、信号处理等研究。

会经济方面<sup>[2-4]</sup>。基于灰色模型 GM (1, 1) 的预测称为灰色预测。基本预测模型的建立参见文献[1]。

## 2 模型精度检验与优化

为了保证模型的精度和可信度, 一般情况下可采用两种方法来检验模型的精度, 即残差检验和后验差检验。如果原始模型不能达到精度要求, 则需对模型进行校正或优化。所谓残差检验, 就是把预测数据同实际数据相比较, 观察其绝对误差和相对误差是否满足预测要求。一般要求相对误差不大于10%, 原点(距离将来最近的点, 即已有数据中的最新数据点) 误差不大于2%。后验差检验以小误差概率  $P$  和后验差比值  $C$  作为检验的尺度, 按照  $P$  和  $C$  的大小, 可将预测精度分为好、合格、勉强、不合格4类。评价标准参见文献[1]。

现有 1994 ~ 2001 年的春运客流数据, 要对 2002 年的客流量进行预测。以直通线春节前第 15 天客流量为例, 建立 GM (1, 1) 模型, 可得如下预测计算公式

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\mu}{a}\right)e^{-ak} + \frac{\mu}{a} = 107.2923e^{0.0283k} - 103.7122 \quad (1)$$

$k = 0, 1, \dots, 7$

根据式(1)得到的模型计算数据如表1所示。

表1 式(1)模型计算数据

年度	模型生成数据	实际数据	误差	误差百分数
	万人	万人	万人	%
1995	3 082.5	4 002.9	0 920.4	22.9937
1996	3 171.0	2 836.7	- 0 334.3	- 11.7863
1997	3 262.1	2 861.0	- 0 401.1	- 14.0212
1998	3 355.9	2 814.6	- 0 541.3	- 19.2307
1999	3 452.3	3 170.1	- 0 282.2	- 8.9013
2000	3 551.5	3 814.6	0 263.1	6.8982
2001	3 653.5	4 039.7	0 386.2	9.5602

由表1可见误差较大, 难以令人满意。根据后验差检验原理, 计算得  $C = 0.9296$ ,  $P = 0.3750$ 。可知预测结果的精度等级为不合格, 因此必须进行模型优化。GM (1, 1) 模型优化, 就是对于给定的数据序列, 寻求精度最高的 GM (1, 1) 模型待识别参数  $a = (a, u)^T$ 。

根据实际数据情况, 本文采取以下优化措施:

### 2.1 优化模型

由于该模型的参数是用最小二乘法求得的, 符合最小二乘的性质<sup>[5]</sup>。如果把原始数据序列的每一个数据都加上一个常数  $a_0$ , 构成新的数据序列  $y^{(0)}$ ,

即

$$y^{(0)} = (x^{(0)}(1) + a_0, \dots, x^{(0)}(n) + a_0) = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)) \quad (2)$$

利用  $y^{(0)}$  建模所得待识别参数  $a^*$  及  $u^*$  必定受  $a_0$  的影响, 从而影响到模型的计算值和精度。利用后验差检验和原点误差检验等方法, 通过模拟计算出模型精度综合最优的  $a_0$ , 最后取预测精度最高的  $a_0$  进行建模预测。但序列  $x^{(0)}$  加  $a_0$  之后, 应满足新数列为正或负数列。设取常数  $a_0$  时模型精度最优, 按照 GM (1, 1) 建模, 可得模型计算公式

$$\hat{x}^{(0)}(k) = \left(y^{(0)}(1) - \frac{\mu^*}{a^*}\right) \times (1 - e^{a^*})e^{-a^*(k-1)} - a_0 \quad (3)$$

式(3)即为优化模型, 它是提高精度的一种有效方法。

### 2.2 残差校正模型

残差校正模型就是对预测误差建立 GM (1, 1) 模型, 对基本预测模型进行校正。定义残差为  $q^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$ , 则有残差序列  $q^{(0)} = (q^{(0)}(1), q^{(0)}(2), \dots, q^{(0)}(n))$ 。对  $q^{(0)}$  建立 GM (1, 1) 模型, 并把所得模型加到主模型中对其进行校正。设主模型和校正模型的优化参数分别为  $a_0$  和  $a_1$ , 则残差校正综合模型为

$$\hat{x}^{(0)}(k) = \left(y^{(0)} - \frac{\mu^*}{a^*}\right) (1 - e^{a^*})e^{-a^*(k-1)} - (a_0 + a_1) + \delta(k - i_0) \times \left(q^{(0)}(1) - \frac{\mu^*}{a^*}\right) (1 - e^{a^*})e^{-a^*(k-i_0-1)} \quad (4)$$

式中

$$\delta(k - i_0) = \begin{cases} 1, & k > i_0 \\ 0, & k = i_0 \end{cases} \quad (5)$$

$i_0 = n - n$ ,  $n$  表示实际数据序列的个数,  $n$  表示用于建立残差模型的残差个数。

当  $k = n + j$  时, 代入式(4)计算, 可得综合模型的第  $j$  个预测值。

### 2.3 新陈代谢模型

对于本征性灰色系统或广义能量系统而言, 随着时间的推移, 未来的一些扰动和因素等将相继进入系统并造成影响。对 GM (1, 1) 模型有预测意义的数据仅是数据  $x^{(0)}(n)$  以后的一两个数据, 越往未来发展, GM (1, 1) 计算的预测数据的预测意义越小, 其可信度越低<sup>[1]</sup>。为使模型能够多次重复使用, 并充

分利用已有的信息, 本文使用一种新陈代谢模型。该模型的基本思想是越接近现在的数据, 对未来的影响越大。也就是说, 随着时间的推移, 能够不断得到新的数据, 而那些老数据的信息意义将随之降低。这一思想无论从理论上还是在实际中都是合理的。

车站每年都可得到新的春运客流数据, 这些新的数据包含着一些新的扰动和因素, 体现了未来的发展趋势。需要把这些数据不断加入到模型中, 同时去掉一些最老的数据, 这样既可获得最能体现未来发展趋势的灰参数数据, 又能维持模型计算量的不变。例如: 要对 2002 年的春运客流量进行预测, 使用的是 1994 ~ 2001 年的客流数据, 当得到 2002 年实际的客流数据后, 便可去掉 1994 年的数据, 利用 1995 ~ 2002 年的数据对 2003 年的春运客流量进行预测。

### 3 预测结果及分析

根据 1994 ~ 2001 年 8 年的历史数据来建立模型, 预测 2002 年的客流量。也就是以 8 个数据为原始数据来建模, 预测第 9 个数据, 并用前 8 个数据的实际值和预测值来检验模型的精度和可信度。实际预测结果表明, 优化效果是明显的, 模型的精度为优。图 1 和图 2 给出了 2002 年预测结果与 2001 年实际客流数据的同比图, 其中: 实线为 2001 年实际数据, 虚线为 2002 年预测数据。可以看出, 二者的大体趋势是完全一致的。

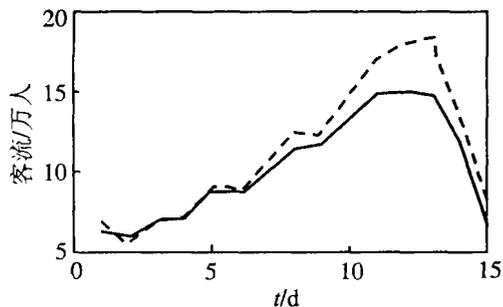


图 1 春节节前 15 天发送旅客同比图

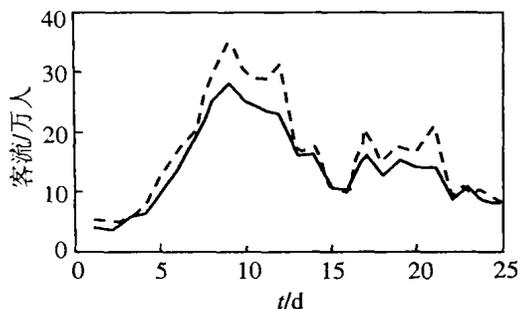


图 2 春节节后 25 天发送旅客同比图

现对预测结果分析如下:

1) 由往年的客流数据曲线可以看出, 客流总体表现出一定的规律: 节前的客流在距春节前十几天时开始缓慢爬升, 到节前五六天时达到峰点, 并且这几年的峰点有后移的趋势, 而在节前的两三天内迅速回落, 春节到来时客流高潮便已结束。节后的客流量会出现两三次高峰: 初四、初五时会出现第 1 个高峰, 这几天的民工人数会出现激增现象, 必须做好接待工作, 及时疏散到站旅客; 元宵节后会出现第 2 个高峰, 但已不如第 1 个高峰猛烈, 这种现象也是符合民工的回返规律的。

2) 本预测方案是完全数据驱动和定量的, 因此有一定的局限性。灰色预测模型是按广义能量系统的增长规律来预测的, 也就是按正常情况(包括经济发展正常, 天气正常, 不出现影响铁路运输的灾害性事件等)进行预测。因此, 车站不能完全依赖于预测数据, 在做出决策时最好能增加一定的定性分析, 以弥补完全数据驱动的不足。例如: 如果春运期间天气晴好, 汽运市场繁荣, 则乘火车的短途旅客将会减少, 铁路的运输压力会有所缓解, 反之则会加重。

### 4 结 语

本文采用灰色预测理论, 根据广州火车站 1994 ~ 2001 年的春运客流数据, 对 2002 年春运期间的客流量作出预测, 并把全部预测过程编写成一个可执行的应用程序, 使用户能够重复利用, 实现客流的滚动预测。本研究成果可对车站的春运客流接待准备工作起到较好的决策支持作用, 也会对预测技术在经济社会领域的应用起到一定的推动作用。

### 参考文献 (References):

- [1] 邓聚龙. 灰色预测与决策[M]. 武汉: 华中理工大学出版社, 1986
- [2] 王好芳, 董增川. 区域经济-人口系统需水量的灰色预测方法研究[J]. 西北水资源与水工程, 2001, 12(2): 53-54 (Wang Haofang, Dong Zengchuan. Water demand on system of regional economy and population[J]. Northwest Water Resources & Water Eng, 2001, 12(2): 53-54)
- [3] 赵茜, 李九红, 王瑞骏. 径流式水电站年发电量的灰色预测[J]. 陕西水力发电, 2001, 17(3): 1-4 (Zhao Qian, Li Jiuhong, Wang Ruijun. Grey prediction model for annual energy output of a hydropower station without storage[J]. Shanxi Water Power, 2001, 17(3): 1-4)

(下转第 512 页)

的有向边的个数

$$T^k = e(-g_m + x)(-g_m + y)(g_m + z) = e[g_m(-xz + xy - yz) + xyz] \quad (5)$$

可见,凡是不包含  $x$  或  $y$  的项均为负,否则为正。

由式(4)便可彻底消除此类冗余项。

## 2.2 第2类冗余项

第2类冗余项的情况较为复杂,现对照图1进行分析:

情况1:如果在图的分解过程中,使得  $\{v_i, v_j\}$  为一个广义顶点  $v_{i,j}$ ,  $\{v_k, v_m\}$  为另一个广义顶点  $v_{k,m}$ , 4条有向边都存在,这时连通广义顶点  $v_{i,j}$  和  $v_{k,m}$  的有向边权之和为

$$g_m - g_m + g_m - g_m = 0$$

这样就自然地消去了一些冗余项。

情况2:如果  $v_m, v_{i,j}$  和  $v_k$  分别为广义顶点,不论  $v_k$  还是  $v_m$  同广义参考顶点相关联或通过一条路径相连通(此路径是指不通过所讨论的4条有向边的任何一条),都有  $g_m - g_m = 0$ ,即消去了冗余项。如果  $v_{i,j}$  通过一条路径同广义参考顶点相连通,按照定理1删去射出边,也不存在冗余项。

情况3:  $v_{k,m}$  在一个广义顶点,  $v_i$  和  $v_j$  在不同的广义顶点,仍有  $g_m - g_m = 0$ ,即消去了一些冗余项。

情况4:无论  $v_i, v_j, v_k$  和  $v_m$  怎样组合和怎样与广义参考顶点相联,每棵有向树都只有一条有向边存在(指图3中的4条有向边)。这时产生的冗余项是无法消除的,因为此时还无法判断它是否为冗余项,只有全部树生成后才有可能将其消去。这是本文方法的一个不足之处,但这些冗余项是非常少的,即使留在全部树中,也不至于使表现形式显得臃肿。要想消去它们,利用有向边权的符号,可以缩小寻找范围。

从上面分析可以看出,该方法消去冗余项主要依靠两种手段:一是式(4),二是利用广义边权的性质消去权相等、符号相反的边。

从方便程度考虑,利用广义边权中各边权的正负号消去冗余项最为简便,其次是利用式(4)。因此应尽量使权相等而符号相反的边在同一广义边之中,即通过短路广义无向边(不含有向边的广义边),使权相等而符号相反的边成为并联边。于是,先短路和开路广义无向边,当图中的广义无向边全部处理完毕后,再短路和开路图中的有向边,按照这样的次序处理图中的所有广义边,可使生成的广义树冗余项较少。

## 3 结语

所谓冗余项的问题是指:当生成一个边集合或一组边集合时,如果不能确认是否为冗余项,而一定要到全部边集合都生成之后,才能逐项比较并加以消除的情况。好的算法是在生成一个边集合或一组边集合时,就能确认其冗余项并将其消除,或在生成一个边集合的一个子集时,就发现冗余项的问题。本文提出的两种消去冗余项的方法,都是在生成边集合或边集合的子集时就将冗余项消去,因此这些方法是先进而有效的。

## 参考文献(References):

- [1] 全茂达,朱英辉.符号网络函数与不定导纳矩阵[M].北京:高等教育出版社,1983:43-52
- [2] Shieu S D, Chan S P. Topological formulation of symbolic network functions and sensitivity analysis of acyive networks[J]. *IEEE Trans on Circuit and Systems*, 1974, 21(1): 39-45
- [3] Zhou Lixing, Chen Yan, Li Weiguo. The systematically generating method of symbolic network functions in cohere fom [A]. *Proc of ICEE [C]*. Kyongju, 1998: 21-25
- [4] 李卫国.一种生成有向树的拓扑方法——广义树法[J].长沙水电师院学报,1990,5(1):130-136  
(Li Weiguo. A new topological theory of generating direct tree tems: Generalized tree method [J]. *J of Changsha Normal University of Water Resources and Electric Power*, 1990, 5(1): 130-136)

(上接第509页)

- [4] 李洪娟,吕渭济,张建军.城市道路交叉口高峰小时流量的灰色预测[J].辽宁工程技术大学学报,2000,19(6):656-658  
(Li Hongjuan, Lu Weiji, Zhang Jianjun. Application of grey prediction model in urban cross peak hour flow [J]. *J Liaoning Tech Univ*, 2000, 19(6): 656-658)

- [5] 李希灿.动态平差灰色预测优化模型[J].测绘工程,1999,8(1):34-40  
(Li Xican. Grey prediction optimal model in dynamic adjustment [J]. *Eng Surveying and Mapping*, 1999, 8(1): 34-40)