

基于灰色预测模型的航线客流量预测^①

刘 夏¹, 李苑辉¹, 陈 磊¹, 陈焕东², 陈明锐³

¹(三亚航空旅游职业学院, 三亚 572000)

²(海南师范大学 教务处, 海口 571158)

³(海南大学 信息科学技术学院, 海口 570228)

摘 要: 准确的预测航线的客流量, 对于航空企业制定航线销售政策有着重要的作用. 针对民航客流量预测具有诸多不确定性和数据不足的特点, 在选取某航空公司近六年三亚-北京往返航线数据的基础上, 提出以回归分析、灰色预测方法对该航线2016年的客流量进行分析预测. 经实证分析, 结果表明: 灰色预测方法对航线客流量的预测的精确度较高, 对航空公司预估客流量和制订销售政策有直接的指导意义.

关键词: 航线数据; 回归分析; 灰色预测模型; 客流量

引用格式: 刘夏, 李苑辉, 陈磊, 陈焕东, 陈明锐. 基于灰色预测模型的航线客流量预测. 计算机系统应用, 2017, 26(7): 221-226. <http://www.c-s-a.org.cn/1003-3254/5880.html>

Prediction for Air Route Passenger Flow Based on Grey Prediction Model

LIU Xia¹, LI Yuan-Hui¹, CHEN Lei¹, CHEN Huan-Dong², CHEN Ming-Rui³

¹(Sanya Aviation Tourism College, Sanya 572000, China)

²(Academic Affairs Office, Hainan Normal University, Haikou 571158, China)

³(College of Information Science Technology, Hainan University, Haikou 570228, China)

Abstract: The accurate prediction on airline passenger flow plays an important role in sales policy for aviation enterprises. Based on the data of Sanya-Beijing route of a particular airline in 6 years, this paper uses the regression analysis, gray forecast method to analyze the passenger flow in 2016. The result shows the Grey forecasting method of airline passenger flow forecasting is more accurate than others, which sets a significantly guiding model for airlines sales policies.

Key words: air route data; regression analysis; grey prediction model; passenger flow

1 引言

准确的预测某一航线的客流量, 对航空企业有着至关重要的作用. 以“客流量预测”为关键字在CNKI进行检索, 该研究从1983年开始, 在2014年达到研究高峰, 总共有175篇文章. “客流量预测”的跨学科研究也发展迅猛, 已深入到计算机科学与技术、管理科学与工程、数学、系统科学、交通运输工程、等多个学科, 并衍生出多个交叉学科主题. 在计算机科学与技术方面, 包括神经网络、支持向量机、支持向量回归、遗

传算法、粒子群算法、遗传神经网络等方法; 在管理科学与工程方面, 包括组合预测、组合模型、多元回归模型、时间序列模型等方法; 在数学方面, 包括时间序列、季节调整、ARMA模型、数学模型、参数估计、排队论等方法; 在系统科学方面, 包括组合预测模型、灰色预测、灰色系统、灰色模型等方法.

在灰色预测方面, 张兆宁、郭爽(2007)基于GM(1, N)模型群, 使用了灰色区间预测模型, 对首都机场未来的流量区间进行了预测^[1]. 杨名桂、杨晓霞(2010)依据

^① 基金项目: 2014年三亚市院地合作科技项目(2014YD52)

收稿时间: 2016-11-17; 收到修改稿时间: 2017-01-04

灰色预测对重庆市近十年入境客流量的进行分析,利用近4年的数据构建了基于GM(1, 1)模型,预测了未来5年的客流量^[2]. 在支持向量机方面, 演克武、朱金福(2010)将支持向量机的回归模型与BPANN、线性回归算法做了对比, 得出支持向量机回归模型误差较小, 预测精度较高^[3]. 在组合预测方面, 路川、胡欣杰(2010)把神经网络与支持向量机进行组合, 实现了有效预测^[4]. 关静(2013)将灰色预测与支持向量机结合, 提高了民航旅游吞吐量的预测精度^[5]. 屈拓(2014)把灰色模型和BP神经网络进行了组合, 提高了机场旅客吞吐量预测精度^[6]. 任崇岭、曹成铨、李静等人(2011)基于Matlab编程将小波神经网络用于客流量短时的预测, 预测精度高, 误差小^[7]. 刘夏等人基于机场历史月度客流量数据, 采用组合加权的方法对机场未来两年的客流量进行了预测, 准确度高^[8]. Sun Y, Zhang G, Yin H基于非参数回归理论, 通过比较与卡尔曼滤波模型和支持向量机回归模型, 结果表明, 非参数回归模型的优点移植能力强, 测量精度高, 可以准确预测客流转移不同的间隔^[9]. Biao, LENG, Jiabei, 等人提出了一种新颖的面向metro-net方法, 首先学习和获得各种OD对的出现概率, 基于实时数据起源, 基于目标基准计算, 然后以几种metro-net来对客流进行预测^[10]. 上述文献虽然在不同层面预测了客流量, 但因民航的特殊性, 未能获取到某一航线的准确数据诸如: 旅客数、客座率、平均票价、平均折扣款等关键字段进行研究分析, 也缺乏对客流量与票价之间关系的研究.

2 数据预处理及描述性分析

2.1 数据的选取

本文选取2005年1月到2015年12月三亚和北京往返共12016个航班的旅客数, 客座率, 完成比率, 平均票

价, 平均折扣款和促销费用等指标作为本文的研究对象. 数据来源于某航空公司市场销售部门.

2.2 数据预处理

对收集到的数据进行整理, 按照月度对各个航班的旅客数, 促销费用和平均折扣率进行计算得到最终的月份数据. 其中每月的每个航线的旅客数:

$$y_i = \sum_{k=1}^n y'_{ik}, i = 1, 2, k = 1, 2 \dots n \quad (1)$$

其中1表示北京-三亚的航线, 2表示三亚到北京的航线, k 表示每月的航班数, y'_{ik} 表示每个航班的旅客数. 促销费用的计算公式为:

$$pc_i = \sum_{k=1}^n pc'_{ik} \quad (2)$$

pc'_{ik} 表示每月每个航班的促销费用. 平均折扣率的计算公式为:

$$ad_i = \frac{\sum_{k=1}^n p'_{ik} y'_{ik}}{\sum_{k=1}^n p_{ik} y'_{ik}} \quad (3)$$

其中 p'_{ik} 是指第*i*个航线第*k*个航班的折扣机票价格, p_{ik} 是指第*i*个航线第*k*个航班的全价机票. 根据上述的计算公式可以得到每月每条航线的旅客数, 促销费用和平均折扣款的指标数据.

2.3 描述性分析

对北京-三亚和三亚-北京航线进行描述性分析得到的结果如表1所示.

根据表1可得北京-三亚和三亚-北京的航线在样本内每月的平均发送旅客数量分别为15408人和15235人, 平均机票价格为1568.37元和1536.46元, 平均客座率为84%和83.48%.

表1 两条航线各个指标在样本内的均值

均值	旅客数	促销费用(元)	机票价格	折扣率	客座率	机票收入(元)
北京-三亚	15408	377727.9	1568.37	0.669624	0.8400	24729979
三亚-北京	15235	350264.6	1536.46	0.656763	0.8348	23402584

3 基于回归分析

回归分析是处理变量与变量之间关系的一种统计分析方法, 应用非常的普遍. 它的主要目的是研究被解释变量与解释变量之间的数量关系. 根据选取的变量以旅客数(Y)作为被解释变量, 以促销费用(PC), 平均折扣率(AD)和航线(AC)作为解释变量, 建立一个多元回

归模型, 其模型的表达式为:

$$Y_t = \beta_0 + \beta_1 PC_t + \beta_2 AD_t + \beta_3 AC_t + \varepsilon_t \quad (4)$$

其中 ε_t 是随机误差项. 其中AC是虚拟变量 $AC = \begin{cases} 1, B-S \\ 0, S-B \end{cases}$, 用R3.2.3软件进行回归得到的结果如表2所示.

表2 回归结果

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.70E+04	2.65E+03	10.196	<2e-16***
PC	3.25E-02	2.27E-03	14.321	<2e-16***
AD	-3.60E+04	4.25E+03	-8.471	2.99e-14***
AC	2.55E+02	7.44E+02	0.343	0.732
Adjusted R-squared:	0.5892			
F-statistic:	69.36 on 3 and 140 DF			
p-value:	<2.2e-16			

根据表2所示回归结果,可以得到总体回归方程的 P 值为零,小于0.05的显著水平,说明总体回归方程是显著地.得到的回归方程为:

$$PN = 27000 + 0.0325PC - 36000AD + 255AC \quad (5)$$

根据上述的回归结果可以看出,在0.05的显著水平下,促销费用和平均折扣率对旅客人数具有显著性的影响.其回归系数 $\beta_1=0.0325$ 表示在其他变量不变的条件下,促销费用每增加1元,航班旅客数平均增加0.0325人.回归系数 $\beta_2=-36000$ 表示航线的机票价格每增加一折的折扣,航班旅客人数平均增加36000人.航线的概率 P 值为0.732大于0.05的显著水平,说明航线对旅客的数量没有显著性的影响.即北京-三亚和三亚-北京的旅客人数没有显著性的差异.另一方面,可以看出旅客数量与促销费用呈正相关关系,即促销力度越大旅客数越多.与平均折扣率呈负相关关系,即平均折扣力度越大,旅客数越多.

4 基于灰色模型的预测方法

4.1 灰色预测模型简介

邓聚龙提出的灰色系统理论,以“部分信息已知,部分信息未知”的“小样本、贫信息”不确定性系统为研究对象,通过对“部分”已知信息的生成、开发,提取有价值的信息,实现对系统运行行为、演化规律的正确描述和有效监控^[11].鉴于航班上的乘客数量符合灰色系统的特点,使用灰色预测模型来预测某一航班的乘客数量则有着较强的针对性.

4.2 模型的建立

首先,设时间序列 $X^{(0)}$ 有 n 个观察值, $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$,通过累加生成新序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$,则GM(1,1)模型相应的微分方程为:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \quad (6)$$

这个方程称为白化方程,也称影子方程.其中: δ 称为发展灰数, μ 称为内生控制灰数.

其次,设 \hat{a} 为待估参数向量, $\hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix}$,可用最小二乘法求解.解得:

$$\hat{a} = (B^T B)^{-1} B^T Y_n \quad (7)$$

其中:

$$B = \begin{bmatrix} -\frac{1}{2}[X^{(1)}(1) + X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2) + X^{(1)}(3)] & 1 \\ \dots & \dots \\ -\frac{1}{2}[X^{(1)}(n-1) + X^{(1)}(n)] & 1 \end{bmatrix}, \quad (8)$$

$$Y_n = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \dots \\ X^{(0)}(n) \end{bmatrix}$$

求解微分方程,即可得预测模型:

$$\hat{X}^{(1)}(k+1) = \left[X^{(0)}(1) - \frac{\mu}{a} \right] e^{-ak} + \frac{\mu}{a} \quad (9)$$

$$k = 0, 1, 2, \dots, n$$

接着对上式进行累减还原,即可得:

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) \quad (10)$$

得原始数列的 $X^{(0)}$ 灰色预测模型为:

$$\hat{X}^{(0)}(K+1) = (1 - e^a) \left[X^{(0)}(1) - \frac{\mu}{\alpha} \right] e^{-ak}, k = 0, 1, 2, \dots, n \quad (11)$$

4.3 模型检验

灰色预测模型检验主要有残差检验,关联度检验和后验差检验.

(1) 残差检验:计算原始序列和原始序列的灰色预测序列之间的绝对误差和相对误差.

绝对误差:

$$\Delta^{(0)}(i) = X^{(0)}(i) - \hat{X}^{(0)}(i); (i = 1, 2, \dots, n) \quad (12)$$

相对误差:

$$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(i)} \times 100\%; (k = 1, 2, \dots, n) \quad (13)$$

相对误差越小,模型精度越高.

(2) 关联度检验. 根据前面所属关联度计算方法, 计算出 $X^{(0)}$ 与 $\hat{X}^{(0)}$ 的关联度, 关联度大于0.6便可以.

(3) 后验差检验. 首先计算原始序列 $X^{(0)}$ 的标准差:

$$S_1 = \sqrt{\frac{\sum_{i=1}^n [X^{(0)}(i) - \bar{X}^{(0)}]^2}{n-1}} \quad (14)$$

然后计算绝对误差序列的标准差:

$$S_2 = \sqrt{\frac{\sum_{i=1}^n [\Delta^{(0)}(i) - \bar{\Delta}^{(0)}]^2}{n-1}} \quad (15)$$

再计算方差比 $c = \frac{S_2}{S_1}$, 最后计算小误差概率 $p = \{|\Delta^{(0)} - \bar{\Delta}^{(0)}| < 0.6745 \cdot S_1\}$. 根据下面的预测精度等级划分表确定模型的精度.

表3 预测精度等级划分表

小误差概率P值	方差比C值	预测精度等级
>0.95	<0.35	好
>0.80	<0.5	合格
>0.70	<0.65	勉强合格
≤0.70	≥0.65	不合格

若残差检验、关联度检验、后验差检验都通过, 则可以用所建模型进行预测.

4.4 数据处理

分别选取北京到三亚和三亚到北京这两条航线2010年到2015年的每月的旅客数作为样本, 利用灰色模型进行拟合, 最后对样本外进行预测. 做出两条航线的时序图, 得到如图1的结果.

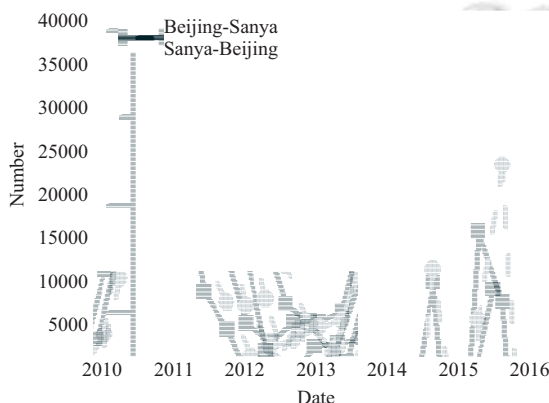


图1 航线时序图

根据上述的时序图可以看出两条航线的的旅客数具有明显的季节趋势, 因此, 需要对两条航线的旅客数

进行季节调整. 以消除明显的季节趋势. 季节调整公式为:

$$\bar{y}_j = \frac{\sum_{i=1}^m y_{ij}}{m}, \text{ 其中 } y_{ij} \text{ 表示第 } i \text{ 年 } j \text{ 期, } j = 12 \quad (16)$$

$$\bar{y} = \frac{\sum_{i=1}^m \sum_{j=1}^{12} y_{ij}}{12m} \quad (17)$$

进一步计算得到的季节指数为:

$$S_j = \frac{\bar{y}_j}{\bar{y}} \quad (18)$$

根据2010到2015年两条航线一共72个月的每月旅客数量, 利用上述公式计算得到的每月的季节指数, 如表4所示.

表4 每月季节指数

月份	1月	2月	3月	4月	5月	6月
三亚-北京	1.4005	1.5809	1.2888	0.9422	0.8542	0.6064
北京-三亚	1.5035	1.4627	1.2075	0.9146	0.7053	0.5807
月份	7月	8月	9月	10月	11月	12月
三亚-北京	0.6037	0.6011	0.647	0.91	1.1921	1.3731
北京-三亚	0.6143	0.5804	0.6729	0.9715	1.3158	1.4709

4.5 模型建立

计算出调整后的两个航线的旅客人数, 根据上述模型的原理, 利用R软件, 根据北京-三亚和三亚-北京2010年1月至2015年12月调整后的旅客数量建立模型.

首先, 令 $X_1^{(0)} = \{11425.63, 10990.77, \dots, 22191.78\}$ 和 $X_2^{(0)} = \{11311.72, 11576.28, \dots, 21290.56\}$ 对原始列进行累加生成新数列: $X_1^{(1)} = \{11425.63, 22416.40, \dots, 1109400.00\}$ 和 $X_2^{(1)} = \{11311.72, 22888.00, \dots, 1096893.00\}$; 然后解微分方程 $\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu$, 利用最小二乘法求参数向量 $\hat{\alpha} = (B^T B)^{-1} B^T Y_n$. 估计的结果为 $\alpha_1 = -0.0085$, $\mu_1 = 11187.61$ 和 $\alpha_2 = -0.0845$, $\mu_2 = 11063.48$ 得到北京-三亚旅客的预测模型为:

$$\hat{X}_1^{(1)}(k+1) = 1327615.042e^{-0.0085k} - 1316189.412, \quad (19)$$

$$k = 0, 1, 2, \dots, n$$

和三亚-北京的旅客数的预测模型为:

$$\hat{X}_2^{(1)}(k+1) = 1321374.479e^{-0.0845k} - 1309287.574, \quad (20)$$

$$k = 0, 1, 2, \dots, n$$

运用R软件运算上述模型预测北京-三亚和三亚-

北京的旅客数量,得到的样本内预测值与真实值的比较结果如图2、图3所示。

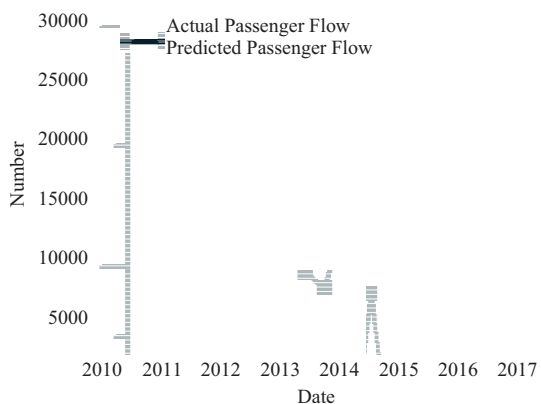


图2 北京-三亚航线实际值与预测值的时序图

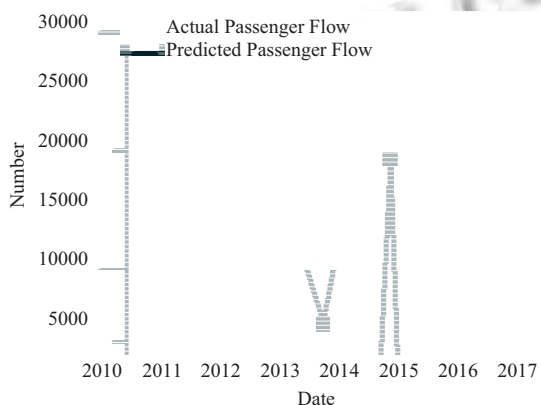


图3 三亚-北京航线实际值与预测值的时序图

4.6 模型检验

对上述模型进行检验,通过残差检验和后验检验来判别模型的可靠程度,得到的检验结果如表5所示。

表5 模型的检验参数

指标	北京-三亚	三亚-北京
平均相对误差/%	12.52	13.13
后验比值	0.4655603	0.5141915
精度等级	合格	勉强合格

根据表5的检验结果可以得到两个航线的模型的后验比值都在0.5左右,模型也是基本合格的.可以用来预测样本外两个航线的年客流量。

根据图2、图3可以得出上述建立的灰色模型大体拟合了两条航线的大体走势。

4.7 模型预测

利用上述模型对2016年两个航线的客流量进行预

测,得到的预测值为 y_i^* ,最终的预测值 $yf = y_i^* * S_j$,得到的结果如表6、图4所示。

表6 样本外旅客数预测

日期	北京-三亚		
	预测值	季节指数	预测旅客数
2016/1	20653	1.50	31051
2016/2	20828	1.46	30465
2016/3	21005	1.21	25364
2016/4	21184	0.91	19374
2016/5	21363	0.71	15067
2016/6	21545	0.58	12512
2016/7	21728	0.61	13347
2016/8	21912	0.58	12718
2016/9	22098	0.67	14870
2016/10	22286	0.97	21652
2016/11	22475	1.32	29572
2016/12	22666	1.47	33338

日期	三亚-北京		
	预测值	季节指数	预测旅客数
2016/1	20653	1.40	28925
2016/2	20828	1.58	32928
2016/3	21005	1.29	27071
2016/4	21184	0.94	19959
2016/5	21363	0.85	18249
2016/6	21545	0.61	13064
2016/7	21728	0.60	13116
2016/8	21912	0.60	13171
2016/9	22098	0.65	14298
2016/10	22286	0.91	20280
2016/11	22475	1.19	26793
2016/12	22666	1.37	31124

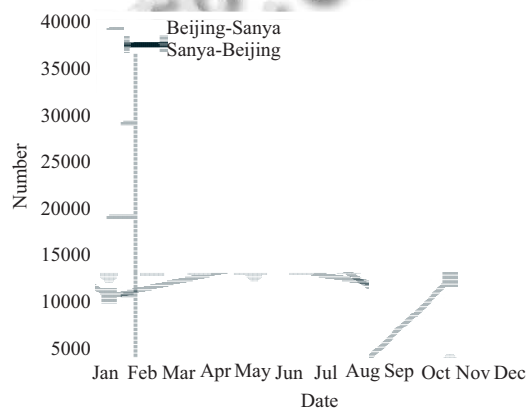


图4 2016年的样本外旅客数预测图

5 结语

由于航班客流量受天气影响、航空管制等诸多不确定因素与灰色预测模型的模糊性与不确定性相一致、因此使用历史数据对未来的客流进行预测是具有可行性的.要提高预测的精度,在未来的预测中需要进

一步考虑本文中由于季节调整带来的预测误差,可以采用分月份建立灰色预测模型进行预测,从而可以避免季节因素对预测精度的影响,以提高调整后的拟合优度值 R^2 ,使其预测更接近真实值。

从预测的结果可以看出,未来一年不管是北京-三亚还是三亚-北京的航线的旅客数都会继续增加的,随着人数的增加,在2016年年底两个航线的人数有望突破三万人。在高铁快速发展对民航的迅猛冲击下,北京-三亚和三亚-北京航班的平均客座率为84.00%和83.48%,这显然得益于三亚旅游业的快速发展。在未来,这条往返航线的乘客人数将持续增加,航空公司更应该通盘考虑,适当增加运力、调整机票出价格或投放宽体执飞,增加营业收入。

参考文献

- 1 张兆宁, 郭爽. 首都机场飞行流量的灰色区间预测. 中国民航大学学报, 2007, 25(6): 1-4.
- 2 杨名桂, 杨晓霞. 基于灰色预测模型的重庆市入境旅游客流量预测. 西南师范大学学报(自然科学版), 2010, 35(3): 259-263.
- 3 滨克武, 朱金福. 基于支持向量机回归算法的航空公司客流量预测研究. 企业经济, 2010, (3): 88-90.
- 4 路川, 胡欣杰. 区域航空市场航线客流量预测研究. 计算机技术与发展, 2010, 20(4): 84-88, 92.
- 5 关静. 基于灰色支持向量机的民航旅客吞吐量预测. 大连交通大学学报, 2013, 34(3): 41-43.
- 6 屈拓. 组合模型在机场旅客吞吐量预测中的应用. 计算机仿真, 2012, 29(4): 108-111.
- 7 任崇岭, 曹成铉, 李静, 等. 基于小波神经网络的短时客流量预测研究. 科学技术与工程, 2011, 11(21): 5099-5103, 5110. [doi: 10.3969/j.issn.1671-1815.2011.21.031]
- 8 刘夏, 陈磊, 李苑辉, 等. 基于组合方法的三亚机场客流量预测. 计算机系统应用, 2016, 25(8): 23-28.
- 9 Sun YJ, Zhang GH, Yin HH. Passenger flow prediction of subway transfer stations based on nonparametric regression model. Discrete Dynamics in Nature & Society, 2014: 397154.
- 10 Leng B, Zeng JB, Xiong Z, *et al.* Probability tree based passenger flow prediction and its application to the Beijing subway system. Frontiers of Computer Science, 2013, 7(2): 195-203. [doi: 10.1007/s11704-013-2057-y]
- 11 邓聚龙. 灰理论基础. 武汉: 华中科技大学出版社, 2002: 361-369.