

基于 Log-ARIMA 的道路运行车速预测模型*

向红艳, 徐 韬, 何素贞

(重庆交通大学 交通运输学院, 重庆 400074)

摘要:道路运行车速预测是交通预测的难点,运行车速随交通条件的变化而变化,为提高道路运行车速预测精度,构建了自回归移动平均(ARIMA)时间序列模型,并结合实例对重庆市江北区红黄路早高峰小客车平均运行车速进行了预测。结果表明:相较于传统的线性回归、多项式拟合、指数拟合和模糊线性回归预测模型,ARIMA 预测模型的平均绝对误差分别下降了 19.1%,50%,6.5%和 3.7%;另外,将原始序列取自自然对数后再建立 ARIMA 的 Log-ARIMA 模型可进一步提高预测精度,预测绝对误差为 5.21,与普通 ARIMA 模型比较,平均相对误差下降了 29.9%。

关键词:ARIMA;道路运行车速;预测;效果比较

中图分类号:U491.1+4

文献标志码:A

文章编号:1672-6693(2016)03-0121-05

道路运行车速预测是交通预测中的重点和难点,道路车速受路段、时间、环境的影响较大,道路车速预测是利用数学模型,通过历史数据对未来一定时间的道路平均车速进行预测和分析,是制定正确的交通控制和交通管理措施的重要前提,也是目前广泛开展的智能交通的关键技术,因此,对道路车速预测进行研究具有重要的现实意义。

道路车速预测方法较多,常见的有线性回归、多项式拟合、指数拟合、神经网络模型和交通仿真等。解少博等人^[1]通过采集二级公路上 30 个样本路段的平曲线半径、纵坡度等线形数据和实测车速,利用模糊线性回归方法建立了小轿车第 85 百分位运行车速区间预测模型;王栋等人^[2]采用两阶段预测模式,建立了基于 BP 人工神经网络的山区高速公路直线段车速预测模型,并对模型进行了测试;解少博等人^[3]分别应用线性回归、多项式回归和 BP 神经网络和模糊神经网络建立了小轿车第 85 百分位车速模型,并对 4 种模型预测精度进行了对比分析。由于道路车速受到复杂交通条件的影响,随机性强,非线性特征突出,一般模型难以准确描述道路运行车速的变化,进一步加大了预测难度。传统预测模型如线性回归、多项式回归在预测非线性问题时精度较低,BP 神经网络模型和模糊神经网络模型建模相对复杂。

针对上述缺陷,文中应用自回归移动平均(ARIMA)时间序列模型进行车速预测分析,该模型能识别速度序列的变化趋势,在此基础上进行模型参数估计和预测,进而选择最合理的预测模型,因此,它的适用范围较广^[4-5]。本文运用重庆市江北区红黄路小客车平均车速数据建立了 ARIMA 模型,并提出对原始数据取自自然对数后建立新的 ARIMA 模型,比较了简单 ARIMA 和对数-ARIMA 的预测效果。

1 ARIMA 模型

自回归移动平均(ARIMA)时间序列模型主要解决两个问题:一是时间序列的平稳性、随机性和周期性;二是在对时间序列分析的基础上,选择适当的模型进行预测。基本模型有:自回归 AR(p)模型、移动平均 MA(q)模型和自回归移动平均 ARIMA(p, d, q)模型^[6]。ARIMA 模型可表示为:

$$\varphi_p(1-B)^d X_t = \theta_q(B)\mu_t, \quad (1)$$

其中, X_t 为原序列; μ_t 为白噪声序列,是一组均值为 0,方差为 σ^2 的随机变量序列; B 为滞后算子, $BX_t = X_{t-1}$; φ_p 为自回归算子, $\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)$, p 为模型的自回归阶数; θ_q 为移动平均算子, $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, q 为移动平均阶数; θ_0 为参数, $\theta_0 = v(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$, v 为平均数。

* 收稿日期:2015-07-08 修回日期:2015-12-25 网络出版时间:2016-04-30 9:42

作者简介:向红艳,女,副教授,博士,研究方向为交通运输规划与管理,E-mail:xiang-@126.com

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.n.20160430.0942.020.html

ARIMA 建模与预测步骤如下:

1) 对数据进行平稳化处理。如果数据序列是非平稳的,如存在一定的增长或下降趋势,则对数据进行差分处理,差分的次数为 d 。

2) 模型识别。对处理后时间序列的自相关函数(ACF)和偏相关函数(PACF)进行分析,初步确定自回归阶数(p)和移动平均阶数(q),得到初始模型 ARIMA(p, d, q)。

3) 参数估计和模型诊断。求出初始模型的各个系数,并对模型中的系数进行显著性检验,对模型进行白噪声检验。

4) 利用具有合适参数的模型进行预测和分析。

2 建模应用

2.1 序列平稳性判断

本文收集了重庆市江北区红黄路早高峰 7:00~8:00 的小客车平均车速,为方便建模,将每一分钟设置为一个时间序列,用 1~45 号序列进行建模,46~60 号序列进行模型拟合分析,见图 1 所示。

从图 1 时间序列相关图可知,随着序列的增加,速度先增后减,序列 16 达到峰值,序列 41 达到最低值,之后呈上升趋势,波动较大,类似于随机游走过程,属于非平稳过程,需进行差分处理。

为了确定差分次数,对原始序列进行 T 检验。假设 $H_0: \delta=0$, 序列为非平稳序列;备择假设 $H_1: \delta=1$, 序列为平稳序列。则原始序列 T 结果如下:

检验结果如表 1 所示,在上述各种显著性水平下,均接受 H_0 ,即该序列为非平稳过程。

将原序列进行一阶差分处理,再次进行 T 检验,结果见表 2 所示。

检验结果如表 2 所示,在上述各种显著性水平下,均接受 H_1 ,即该序列为平稳过程, $d=1$ 。

2.2 模型识别及参数估计

对于 ARIMA 模型,若自相关函数在滞后数为 p 后截尾和偏相关函数在滞后数为 q 后截尾,则阶数分别为 p 和 q ^[7]。

原始序列一阶差分后的偏相关函数和自相关函数见图 2 所示,原始序列一阶差分后的自相关函数从第 4 阶开始显著为 0,偏自相关函数从第 3 阶开始显著为 0,初步确定 $q=3, d=1, p=2$,初步确定的模型有:

ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(1,1,3), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA(2,1,2) 和 ARIMA(2,1,3),对 ARIMA(2,1,2) 进行参数检验,结果见表 3 所示。

ARIMA(2,1,2) 系数检验结果如表 3 所示,相伴概率均小于 0.05,故该模型通过系数检验,但仍需进行白噪声检验。

ARIMA(2,1,2) 模型白噪声检验结果如图 4 所示,相伴概率均大于 0.05,故通过白噪声检验,对其余模型做模型参数检验,最终确定 ARIMA(2,1,2) 为最优模型,ARIMA(2,1,2) 拟合函数为:

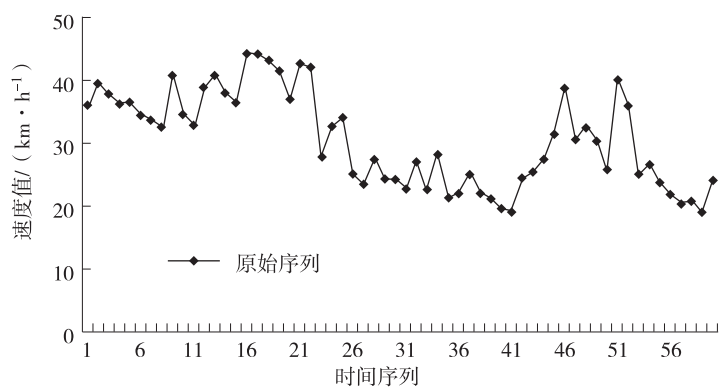


图 1 红黄路早高峰小客车平均车速时间序列

Fig. 1 Honghuang road morning peak passenger car average speed time series

表 1 原始序列 T 检验表

Tab. 1 T test table of the original sequence

显著性水平	T -检验临界值	T -实际值	检验结果
0.01	-3.588 509	-0.968 417	接受 H_0
0.05	-2.929 734	-0.968 417	接受 H_0
0.1	-2.603 064	-0.968 417	接受 H_0

表 2 原始序列 1 阶差分后 T 检验表

Tab. 2 Table of original sequence of one order difference in T test

显著性水平	T -检验临界值	T -实际值	检验结果
0.01	-3.588 509	-8.696 927	接受 H_1
0.05	-2.929 734	-8.696 927	接受 H_1
0.1	-2.603 064	-8.696 927	接受 H_1

$$(1-B)X_t =$$

$$\frac{1-0.977\ 162\ B+0.932\ 421\ B^2}{1-0.333\ 444\ B+0.508\ 197\ B^2}\mu_t, \quad (2)$$

其中 B 为滞后算子, X_t 为原始序列, μ_t 为白噪声序列。

2.3 Log-ARIMA 模型

由原始序列相关图(图 1)可知,该序列波动较大,为了使序列更加平稳,将原序列取自然对数(图 3)后再建立 ARIMA 模型,建模过程与上述过程一致,最终得到 ARIMA(2,1,2)模型,该模型通过了 ARIMA 模型系数显著性检验和白噪声检验,原序列取对数后的 ARIMA(2,1,2)模型拟合函数为:

$$(1-B)X_t =$$

$$\frac{1-0.944\ 541\ B+0.927\ 311\ B^2}{1-0.377\ 573\ B+0.513\ 878\ B^2}\mu_t, \quad (3)$$

其中 B 为滞后算子, X_t 为原始序列, μ_t 为白噪声序列。模型白噪声检验结果如图 4 所示。

2.4 各类模型预测精度分析

通过分析重庆市江北区红黄路早高峰 7:00~8:00 的小客车平均车速,建立了普通 ARIMA 和对数 ARIMA 模型,用 1~45 号序列进行建模,46~60 号序列进行模型预测和分析,为了更好地比较各类模型的预测精度,将原始数据进行线性回归、多项式拟合、指数拟合和模糊线性回归,拟合结果见图 5 所示。

从图 5 可知,原始序列呈现出了上下波动,线性拟合、多项式拟合和指数拟合均无法较好地拟合原始序列,由于道路车速受到周围环境的显著影响,变化较快,有固定趋势变化的线性拟合、多项式拟合及指数拟合显然无法较好地反应实际过程中道路车速的变化。ARIMA 时间序列模型通过自回归和移动平均,将历史数据实时反馈到模型中,故较好地拟合了原始序列,从图 5 可直观看出,ARIMA 模型随着序列变化而呈现出一定的波动,且将原始序列经过取自然对数处理后,拟合程度更高,可靠度更大,表明 Log-ARIMA 模型能较好地反应实际过程中的道路车速变化,各类模型预测精度如表 4 所示。

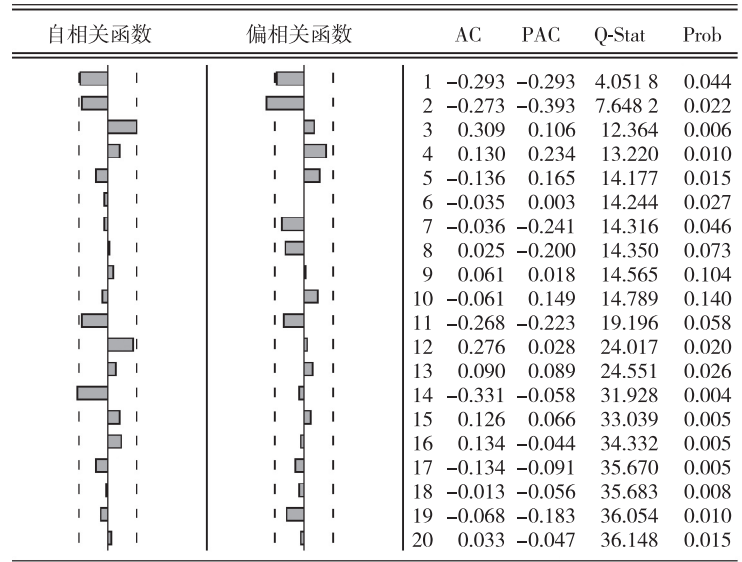


图 2 原始序列一阶差分的自相关和偏相关图

Fig. 2 Autocorrelation and partial correlation graph of original sequence in one order difference

表 3 ARIMA(2,1,2)模型参数检验表

Tab.3 ARIMA(2, 1, 2) model parameter test table

变量	系数	T 检验值	相伴概率
Ar(1)	0.333 444	2.309 438	0.026 4
Ar(2)	-0.508 197	-3.524 155	0.001 1
Ma(1)	-0.977 162	-18.238 93	0.000 0
Ma(2)	0.932 421	24.524 67	0.000 0

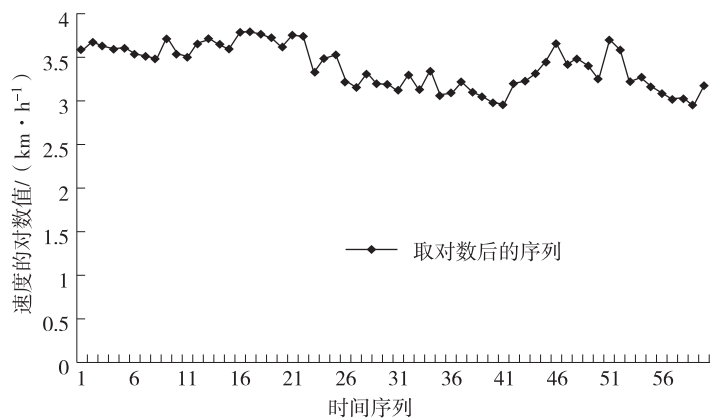


图 3 原始序列取自然对数后的时间序列

Fig. 3 Time series of the original sequence of the natural logarithm

表 4 各类预测模型效果比较

Tab.4 Comparison of various predictive model

误差类型	ARIMA	Log-ARIMA	线性	多项式	指数	模糊线性
平均绝对误差	7.2	5.2	8.9	14.4	7.7	5.4
平均相对误差	26.7%	18.7%	29.9%	52.2%	25.1%	19.1

3 结语

本文对重庆市江北区红黄路早高峰小客车车速进行了调查研究,通过分析原始序列的相关性,建立了道路运行车速的 ARIMA 时间序列模型,将原始序列进行对数处理后,建立了优化的 ARIMA 模型,从而进行道路运行车速预测。

通过原始序列建立的 ARIMA 模型预测绝对误差为 7.21,预测相对误差为 26.7%,相比与传统线性回归、多项式拟合、指数拟合和模糊线性回归,ARIMA 预测模型的平均绝对误差分别下降了 19.1%,50.0%,6.5%,3.7%;将原始序列取对数再构建的 ARIMA 模型进一步提高了预测精度,相较于普通 ARIMA 模型,Log-ARIMA 模型平均绝对误差下降了 29.9%,优化后的 Log-ARIMA 模型预测绝对误差为 5.21,相对误差仅 18.7%,可靠度较高,能较好地对道路车速进行拟合预测。

参考文献:

[1] 解少博,蒋晓君,魏朗.基于模糊线性回归的小轿车运行车速区间预测[J].公路交通科技,2014,31(8):126-130.
Xie S B, Jiang X J, Wei L. Prediction of driving speed interval of passenger car based on fuzzy linear regression[J]. Journal of Highway and Transportation Research and Development, 2014, 31(8):126-130.

[2] 王栋,邓北川,仇建华,等.山区高速公路直线段车速预测方法[J].交通科学与工程,2014,30(3):81-86.
Wang D, Deng B C, Qiu J H, et al. The method of the straight section speed prediction on the mountainous freeway[J]. Journal of Transport Science and Engineering, 2014, 30(3):81-86.

[3] 解少博,阿比旦,魏朗.公路运行车速预测模型对比分析

[J]. 长安大学学报:自然科学版,2013,33(5):81-85.
Xie S B, A B D, Wei L. Comparative analysis of prediction model for vehicle operating speed[J]. Journal of Chang'an University: Natural Science Edition, 2013, 33(5):81-85.

[4] Box G E P, Jenkins G M. Time series analysis, forecasting and control[M]. San Francisco: Holden-day, 1970.

[5] 徐国祥.统计预测和决策[M].上海:上海财经大学出版社,1998.
Xu G X. Statistical forecasting and decision making[M]. Shanghai: Shanghai University of Finance and Economics Press, 1998.

[6] 张杰,刘小明,贺玉龙,等. ARIMA 模型在交通事故预测中的应用[J].北京工业大学学报,2007,33(12):1295-1299.

自相关函数	偏相关函数	AC	PAC	Q-Stat	Prob	
		1	0.081	0.081	0.295 1	
		2	-0.063	-0.070	0.475 7	
		3	0.173	0.186	1.898 7	
		4	0.099	0.064	2.373 8	
		5	-0.036	-0.026	2.439 4	0.118
		6	0.031	0.018	2.488 6	0.288
		7	-0.109	-0.155	3.120 4	0.373
		8	-0.097	-0.071	3.634 9	0.458
		9	-0.057	-0.070	3.814 7	0.576
		10	-0.054	-0.019	3.981 7	0.679
		11	-0.147	-0.102	5.268 5	0.627
		12	0.148	0.209	6.613 9	0.579
		13	0.033	0.014	6.682 6	0.670
		14	-0.247	-0.213	10.694	0.382
		15	0.123	0.138	11.728	0.384
		16	0.132	0.006	12.968	0.371
		17	-0.134	-0.089	14.290	0.354
		18	-0.108	-0.142	15.188	0.365
		19	-0.155	-0.222	17.114	0.312
		20	-0.088	-0.027	17.770	0.338

图 4 模型白噪声检验结果图

Fig. 4 Model white noise test results

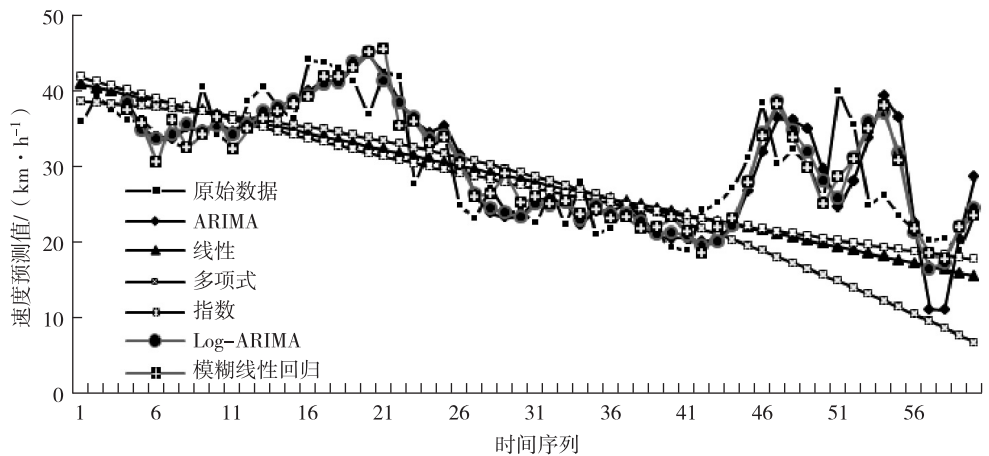


图 5 各类预测模型拟合曲线图

Fig. 5 Kinds of prediction model fitting curve

- Zhang J, Liu X M, He Y L, et al. Application of ARIMA model in forecasting traffic accidents[J]. Journal of Beijing Polytechnic University, 2007, 33(12): 1295-1299.
- [7] Bowerman B L, O'Connell R T. Forecasting and time series: an applied approach[M]. 3rd edition. 北京: 机械工业出版社, 2003.
- [8] Yang M, Liu Y, You Z. The reliability of travel time forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(1): 162-171.
- [9] Ahmed M S, Cook A R. Analysis of freeway traffic time series data by using box-jenkins techniques[J]. Transportation Research Record, 1979, 722: 1-9.
- [10] Myung J, Kim D K, Kho S Y, et al. Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system[J]. Transportation Research Record, 2011, 2256: 51-59.
- [11] Khosravi A, Mazloumi E, Nahavandi S, et al. Prediction intervals to account for uncertainties in travel time prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 537-547.

Log-ARIMA-Based Road Operating Speed Prediction Model

XIANG Hongyan, XU Tao, HE Suzhen

(School of Transportation, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: Prediction of operating speed is a difficult thing of road traffic prediction, operating speed change with traffic conditions, in order to improve the prediction accuracy of road operating speed, this paper constructs the Auto-regressive Integrated Moving Average Model (ARIMA), and forecasts car average operating speed in the morning rush with examples of Hong-huang Road of Chongqing Jiangbei District. The result shows that: compared with the traditional prediction model such as Linear regression, Polynomial fitting and Exponential fitting and Fuzzy Linear Regression, the average absolute error of ARIMA prediction model were decreased by 19.1%, 50%, 6.5% and 3.7%; in addition, using log-ARIMA model which the original sequence has been natural logarithm can deeply improve the prediction accuracy, and the average prediction absolute error is 5.21, the average relative error is decreased by 29.9% compared with ordinary ARIMA model.

Key words: ARIMA; road operating speed; prediction; performance comparison

(责任编辑 游中胜)