

文章编号: 2095-2163(2021)12-0165-06

中图分类号: U491.1

文献标志码: A

基于 ARIMA 模型的城市路网交通运行指数预测研究

董洁霜, 方天源, 周亦威

(上海理工大学 管理学院, 上海 200093)

摘要: 受新冠肺炎疫情的冲击,城市路网高峰行程延时指数(Delay Index)在特殊交通管制措施影响下急剧降低。本研究利用 ARIMA 模型对杭州市西湖区 2019 年 9 月 15 日至 2021 年 9 月 15 日的工作日路网高峰延时指数时间序列数据进行建模分析和预测,完成了样本内预测和样本外预测。结果发现样本内和样本外延时指数预测值的平均绝对百分比误差分别为 2.435、2.625 均小于 10,该数据集对应的 ARIMA(6,0,0)模型预测结果客观真实。此外,对新冠肺炎疫情爆发初期的时间段预测结果分析发现,ARIMA 模型在大样本数据下的预测效果较优,时间序列预测在受短时的外部环境冲击下韧性良好。

关键词: 时间序列预测; ARIMA 模型; 路网高峰行程延时指数(Delay Index); 新冠肺炎

Research of prediction on urban road network traffic operation index based on ARIMA model

DONG Jieshuang, FANG Tianyuan, ZHOU Yiwei

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] Affected by the COVID-19, the peak travel delay index in the urban road network decreases sharply under the influence of special traffic control measures. This research uses the ARIMA model to analyze and predict the peak delay exponential time series data of Hangzhou Xihu District from 15 September 2019 to 15 September 2021. The within-sample prediction and extent-sample prediction is completed. The average absolute percentage error of the within-sample and the extent-sample index is 2.435, 2.625, respectively where the corresponding ARIMA (6, 0, 0) model prediction is objective. Besides, the analysis of the time period prediction result in the initial period of the COVID-19 shows that the predicted effect of the ARIMA model under large sample data is excellent, and the time series prediction can stand short-term environmental impact.

[Key words] time series prediction; ARIMA model; peak travel delay index; COVID-19

0 引言

城市道路交通运行状态数据,是交通出行决策和拥堵治理的关键。路网高峰行程延时指数,简称“延时指数(Delay Index)”,作为城市道路路况重要指标之一,其根据交通部门提供的实时监控和各地图导航后台返回的大数据综合判断,为出行者提供导航路线参考。

现状研究主要针对拥堵延时指数的空间属性。袁浩^[1]等人研究人口、用地属性、交通基础设施密度等空间异质属性对拥堵延时指数的关系。冯海霞^[2]等人通过地理加权回归(geographically weighted regression, GWR)模型,定量分析拥堵延时指数与空气质量指数的强相关性。汪传雷^[3]等人利用熵值法研究不同车型对拥堵延时指数的影响程

度,提出分车型的治堵措施。胡成雨^[4]等人基于 Kruskal-Wallis 检验,对中国不同城市拥堵延时指数进行比较和排名,量化各城市的交通健康差异情况。相较统计理论模型,深度学习的方法对短时交通流数据的捕捉能力更好^[5],但统计学模型适用于长时间跨度的交通流状态指标分析预测。相关研究表明,城市道路拥堵延时指数在时间上有明显的周期性^[6]。其表现为,节假日的始末段拥堵延时指数较高,工作日的周一、周五拥堵延时指数高于该周的其他工作日。但实际上,在长跨度的时间序列数据中,受天气条件、临时交通管制政策、交通基础设施建设进度等多因素影响,并非周一、周五拥堵延时指数都比其余工作日高。如果在建模分析时默认存在时间序列周期性,会对区域交通运行状态指标研判工作有一定的干扰。

基金项目: 上海理工大学管理学院新进博士及引进人才科研启动费项目(BSQD202108);上海市高水平学科建设项目。

作者简介: 董洁霜(1973-),女,博士,副教授,主要研究方向:交通运输规划与管理;方天源(1998-),男,硕士研究生,主要研究方向:交通运输规划与管理;周亦威(1987-),男,博士,讲师,主要研究方向:交通出行分析、空间计量模型。

通讯作者: 周亦威 Email: zywadam@163.com

收稿日期: 2021-11-07

综上所述,研究大多关注空间效应对拥堵延时指数的影响,时间层面受到数据样本量的限制会适当回避。然而,与时间层面影响最大的是政策效应^[7]。例如:单双号限行、小汽车牌照申领指标限制、鼓励错峰出行等。随着时间推移,各个交通管控措施的依次(或叠加)实施,都会对拥堵延时指数产生影响。工作日的拥堵延时指数对市民的通勤出行更具有引导意义。工作日高峰时期拥堵延时指数的上升,主要是市民通勤导致的交通流集聚在部分道路,相较节假日,工作日的城市路网拥堵延时指数一般更高,这是工作日的出行时间更集中,方式单一且路径基本固定所导致。本研究通过将杭州西湖区 2019 年 9 月 15 日至 2021 年 9 月 15 日(共计 499 个数据样本)每日的高峰拥堵延时指数进行时间序列分析(剔除节假日),利用 ARIMA 模型进行样本内数据预测和样本外数据预测。同时考虑新冠肺炎疫情特殊时期的交通管控措施政策对拥堵延时指数的影响。

1 ARIMA 模型设计

ARIMA (Autoregressive Integrated Moving Average model)模型综合考虑了时间序列的趋势变化、周期性和随机干扰^[8-9]。其建模步骤依次为数据预处理、模型识别、模型诊断及检验、模型预测。

ARIMA(p, d, q) 模型原理如下:

$$\Phi(B) \nabla^d x_t = \Theta(B) \varepsilon_t \quad (1)$$

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, s \neq t \quad (2)$$

其中, $E(\varepsilon_s \varepsilon_t) = 0, s < t; x_t$ 表示时间序列数据, x_t 与 $x_{t-i} (i = 1, 2, \dots, p)$ 相关; ε_t 表示残差项, ε_t 与 $\varepsilon_{t-j} (j = 1, 2, \dots, q)$ 相关; B 表示延迟算子, 满足 $B^n x_t = x_{t-n}; p$ 表示自回归阶数; q 表示平均阶数; d 表示差分阶数; ∇ 表示差分算子; $\nabla^d = (1 - B)^d$ 。

式(1)中的自回归系数多项式的具体表达式如下:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3)$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (4)$$

残差 ε_t 是独立于 x_{t-i} 和 ε_{t-i} 的白噪声序列, 满足:

$$\varepsilon_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} - \phi_0 - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} - x_t \quad (5)$$

2 预测数据描述

此次研究的数据来源于高德智慧交通公共服务平台,数据类型为杭州市西湖区工作日的路网高峰行程延时指数(高峰时间为 7:00-9:00 和 16:30-

18:30),数据时间跨度为 2019 年 9 月 15 日至 2021 年 9 月 15 日,共计 499 个样本数据。另外,为测试 ARIMA 模型在样本外的预测能力,补充时间跨度为 2021 年 9 月 16 日至 2021 年 10 月 15 日的工作日路网高峰行程延时指数数据。

路网高峰行程延时指数的计算原理源自旅行时间指数。旅行时间指数 TTI (Travel Time Index)^[10] 是常用的城市拥堵程度的评价指标,其反映实际花费的行程时间与自由流花费行程时间的比值关系,与拥堵程度正相关,值越大表示交通运行状态越差。高德地图后台基于海量的轨迹数据,可以准确计算路网高峰行程延时指数,用直观的形式表达城市拥堵程度。

计算基本思想:定义一条路段 link 有两个连续的时间片,分别为 t_1, t_2 , 路段的长度为 S , 则 t_1 到 t_2 这段时间内路段的平均速度 v 为 $2S/(t_1+t_2)$ 。同一条路段在一个时间片内,旅行时间指数为自由流速度和实际速度的比值。当轨迹覆盖度较低时,根据路段的长度和路况的可信度进行过滤。集合 S 的定义为所有路段 link 的集合, $S = \{link_1, link_2, link_3, link_4, \dots, link_N\}$ 。集合中路段 link 的总数为 N, L_i 为路段 link 的长度, W_i 为路段 link 的权重, V_{free_i} 为路段 link 的自由流速度, V_i 为实时路况速度。故计算方式如式(6)所示。

$$TTI = \frac{\sum_{i=1}^N \frac{L_i}{V_i} \cdot W_i}{\sum_{i=1}^N \frac{L_i}{V_{free_i}} \cdot W_i}; \quad speed = \frac{\sum_{i=1}^N L_i \cdot W_i}{\sum_{i=1}^N \frac{L_i}{V_i} \cdot W_i} \quad (6)$$

以上是对路段延时指数的计算方式,由每个路段的计算结果得到区域拥堵延时指数,计算逻辑如图 1 所示。

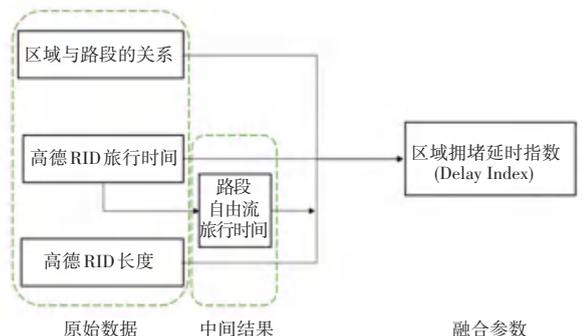


图 1 区域拥堵延时指数的计算逻辑

Fig. 1 Calculation logic of the regional congestion delay index

3 方法与模型

3.1 模型实现

数据样本工作日延时指数时间序列如图 2 所示。由于数据选取的是工作日的延时指数,从图中可见没有明显季节性特征。但在第 91 个时间序列点(2020 年 1 月 23 日)处延时指数骤减,而当天正是武汉市发布“封城”通告时间。从该日开始,城市公交、地铁、长途客运暂停运营,全国各级市也随即响应相关交通出行和管控政策,城市路网的延时指数急剧下降。直到 2020 年 3 月 2 日,随着复工复产的有序推进,杭州市西湖区的路网高峰行程延时指数自骤减后首次回到 1.2 以上。

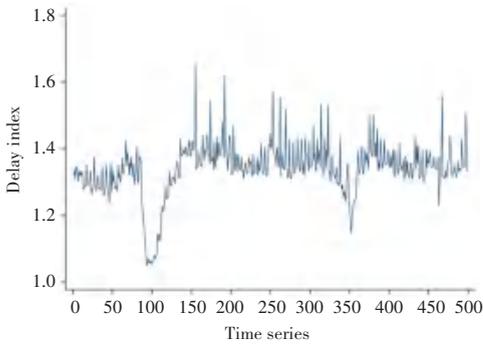


图 2 时间序列(2019/9/15~2021/9/15)工作日路网高峰行程延时指数

Fig. 2 Weekday road network peak delay index of time series from 2019/9/15 to 2021/9/15

平稳性描述的是时间序列的统计性质关于时间平移的不变性。研究时间序列是希望通过时间序列的历史数据规律,得到其未来的一些预测。在模型运行前,对该时间序列进行 ADF 单位根检验结果见表 1。结果显示, $P\text{-value} < 0.05$, 数据稳定, 拒绝原假设。从 ADF 单位根检验结果可以看出 Test statistic = -8.949 小于 1%、Critical value = -3.440。表明 99% 的置信区间下都满足数据平稳性, 无需进行差分处理, 即 $ARIMA(p, d, q)$ 中参数 $d = 0$ 。

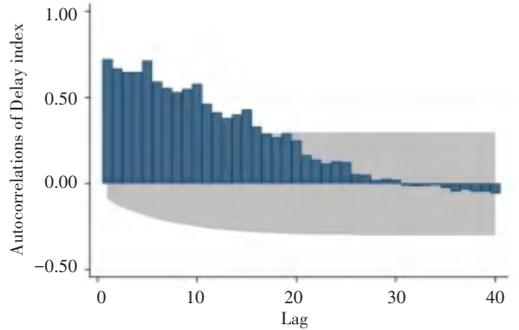
表 1 ADF 单位根检验

Tab. 1 ADF unit root inspection

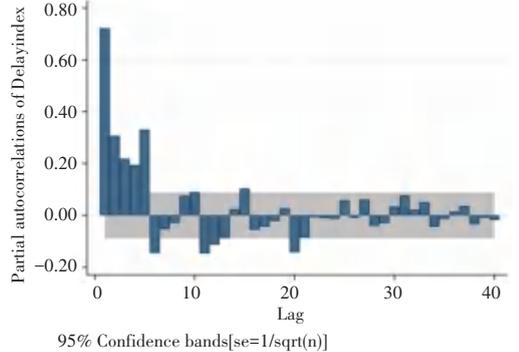
Test-Statistic	1% Critical Value	5% Critical Value	10% Critical Value
$Z(t)$	-8.949	-3.440	-2.870
MacKinnon approximate p-value for $Z(t) = 0.000 0$			

使用自相关(ACF)与偏自相关(PACF)图判断 $ARIMA(p, d, q)$ 模型中的 p, q 阶数, 必要时可以用 AIC (Akaike information criterion)、BIC (Bayes Information Criterion) 检验进行再确定。自相关函数

(ACF)是将有序的随机变量序列与其自身相比较,反映了同一序列在不同时序取值之间的相关性;偏自相关函数(PACF)是剔除了中间随机变量干扰,所得到的两个变量之间的相关性,之后计算两个变量之间的相关性。利用 Stata 软件对序列进行自相关(ACF)图和偏自相关(PACF)图的绘制,其结果如图 3 所示。



(a) 自相关



(b) 偏自相关

图 3 研究序列的自相关(ACF)及偏自相关(PACF)图

Fig. 3 Autocorrelation (ACF) diagram and partially related (PACF)

从图 3(a)中可以初步判断 ACF 图拖尾,故 $q = 0$ 。从图 3(b)可以确定 PACF 图截尾, p 的取值可以为 5、6、7、8。通过 AIC、BIC 检验值最小原则,进行 p 值的再判断,结果见表 2。 $ARIMA(6, 0, 0)$ 模型中的 AIC 值和 BIC 值最小,故 $p = 6, q = 0, d = 0$ 为 $ARIMA(p, d, q)$ 模型最优参数。

表 2 AIC、BIC 值检验结果

Tab. 2 Test results of AIC and BIC value

$ARIMA(p, d, q)$	AIC	BIC
$ARIMA(5, 0, 0)$	-1 614.707	-1 585.218
$ARIMA(6, 0, 0)$	-1 622.944	-1 589.243
$ARIMA(7, 0, 0)$	-1 622.289	-1 584.375
$ARIMA(8, 0, 0)$	-1 620.674	-1 578.548

拟合模型后,需要对残差序列检验,判断是否为

白噪声。数据集中生成了新的残差序列,对得到的残差序列进行 ADF 单位根检验,得到残差是平稳序列,见表 3。模型检验结果如图 4 所示。对残差进行 Ljung-Box 检验后,通过白噪声检验($Q = 63.3772, P = 0.0107$),从残差自相关(ACF)图来看,残差服从均值为 0 的正态分布,此模型残差大部分落入 95%可信区间,说明有效信息被充分提取,模型的有

效性拟合效果较好。模型检验结果如图 4 所示。

表 3 残差的平稳性检验

Tab. 3 Residual stability test

Test-Statistic	1% Critical	5% Critical	10% Critical
	Value	Value	Value
$Z(t)$	-22.385	-3.440	-2.870
MacKinnon approximate p-value for $Z(t) = 0.0000$			

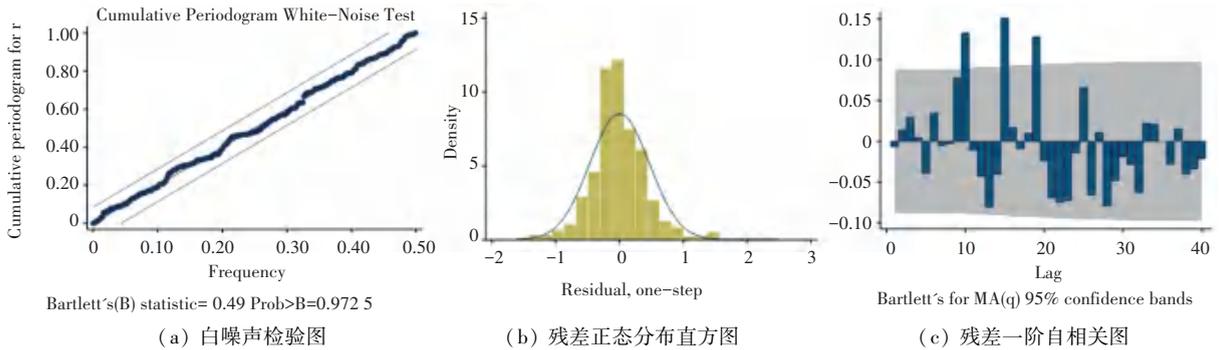


图 4 模型检验

Fig. 4 Model Validation

3.2 模型预测

使用 $ARIMA(6,0,0)$ 对模型进行样本内和样本外预测,结果如图 5 所示。整体来看,预测线与实际线贴合程度较好。对样本外的延时指数数据进行预测发现,预测线逐渐呈现收敛状态,若样本外数据容

量过大,会使预测数据趋于一个常量。样本内平均绝对百分比误差 MAPE (Mean Absolute Percentage Error) 为 $2.435 < 10$, 样本外平均绝对百分比误差 MAPE 为 $2.625 < 10$, 说明模型的预测精度较高。

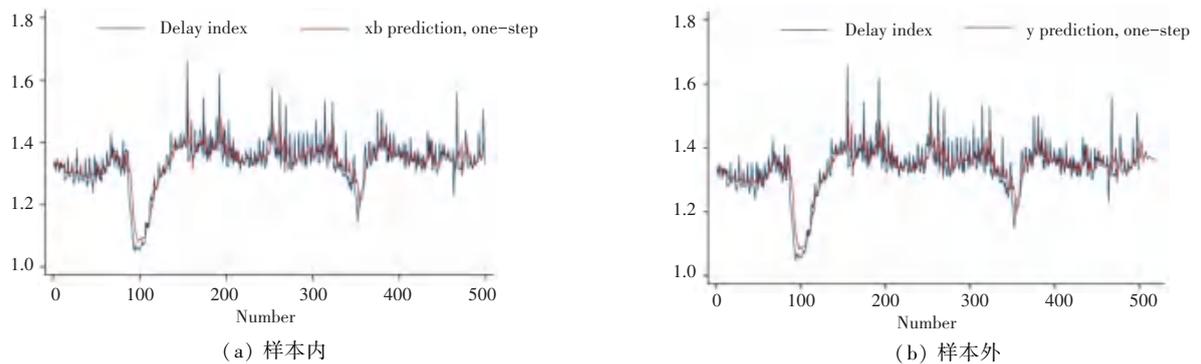


图 5 预测线及实际线

Fig. 5 Prediction line and actual line

4 实验结果及分析

将实验结果从 Stata 软件中导出,见表 4,可见各项统计检验值均显著。从表 5 可以发现,各阶滞后后的 z 值均显著。拥堵延时指数预测值与实际值的绝对误差小、预测结果客观真实。

为检验 ARIMA 模型在突变数据时间段内的预测韧性,考虑新冠肺炎疫情对延时指数的冲击,针对 2020 年 1 月 14 日~2 月 14 日延时指数实际值和预

测值进行分析对比,如图 6 所示。2020 年 1 月 23 日是武汉“封城”开始日,全国各地包括杭州也随即响应特殊交通管制措施。整体看,路网拥堵延时指数从原来的 1.4 左右下降到 1.2 以下。通过该特殊时间段工作日路网高峰行程延时指数的实际值与预测值对比,发现在 1 月 23 日后预测值与实际值有一段数据“落差”,但很快两者又逐渐贴合,可见 ARIMA 模型的预测韧性较好。

表 4 ARIMA(6,0,0) 模型统计检验值结果

Tab. 4 Statistical test value of ARIMA (6, 0, 0) model

ARIMA(6,0,0)	Coef.	Std. Err.	z	P > z	[95%Conf.Interval]	
L 1.	0.375 233	0.031 320	11.98 **	0.000	0.313 846 4	0.436 619
L 2.	0.125 641	0.048 661	2.58 **	0.010	0.030 266 6	0.221 016
L 3.	0.087 068	0.046 726	1.86 *	0.062	-0.004 513 7	0.178 65
L 4.	0.083 220	0.042 922	1.94 *	0.053	-0.000 906 7	0.167 346
L 5.	0.375 753	0.038 352	9.80 **	0.000	0.300 584 1	0.450 922
L 6.	-0.143 44	0.044 795	-3.20 **	0.001	-0.231 240 2	-0.055 65
_cons	1.340 049	0.025 787	51.97 **	0.000	1.289 507	1.390 591
/sigma	0.046 745	0.000 962	48.57	0.000	0.044 858 4	0.048 631

注:表 4 中 **、* 分别表示 $P < 0.05$ 、 $P < 0.1$

表 5 拥堵延时指数预测值与实际值

Tab. 5 The prediction value and actual value of congestion delay index

序列	实际值	预测值	误差/%	序列	实际值	预测值	误差/%
1	1.319 6	1.340 0	1.55	484	1.325 6	1.317 5	-0.61
2	1.338 0	1.325 4	-0.94	485	1.315 1	1.327 2	0.92
3	1.308 8	1.332 8	1.84	486	1.336 6	1.324 3	-0.92
4	1.347 6	1.321 7	-1.92	487	1.362 6	1.340 2	-1.64
5	1.342 2	1.333 8	-0.63	488	1.329 2	1.338 6	0.71
6	1.325 8	1.332 4	0.50	489	1.340 3	1.332 7	-0.56
7	1.303 1	1.335 2	2.46	490	1.335 0	1.333 2	-0.14
8	1.334 2	1.313 8	-1.53	491	1.344 5	1.341 4	-0.23
9	1.318 4	1.339 5	1.60	492	1.430 4	1.349 2	-5.68
10	1.331 2	1.326 5	-0.35	493	1.367 5	1.366 8	-0.05
11	1.319 0	1.324 8	0.44	494	1.349 5	1.363 3	1.02
12	1.329 3	1.316 8	-0.94	495	1.343 6	1.353 3	0.73
	...			496	1.351 4	1.354 9	0.26
83	1.300 2	1.311 0	0.83	497	1.506 6	1.381 2	-8.33
84	1.379 8	1.324 7	-3.99	498	1.437 1	1.402 4	-2.41
85	1.369 1	1.354 0	-1.11	499	1.331 4	1.398 3	5.02
86	1.368 6	1.374 6	0.44	样本外			
		
90	1.164 8	1.266 3	8.71	503	1.389 9	1.379 3	-0.76
91	1.127 7	1.243 4	10.26	504	1.357 2	1.360 1	0.22
92	1.091 1	1.176 9	7.87	505	1.373 3	1.403 7	2.21
93	1.060 7	1.160 4	9.40	506	1.375 3	1.366 9	-0.61
94	1.048 7	1.128 9	7.65	507	1.389 8	1.343 7	-3.32
95	1.069 8	1.112 9	4.03	508	1.374 1	1.343 4	-2.24
96	1.068 5	1.102 3	3.17	509	1.364 2	1.371 5	0.53
97	1.055 2	1.092 5	3.54	510	1.370 7	1.425 7	4.01
98	1.063 7	1.082 0	1.72	511	1.370 2	1.326 7	-3.17
99	1.058 7	1.085 0	2.49	512	1.373 8	1.398 1	1.77
100	1.054 7	1.092 6	3.59	513	1.366 8	1.502 6	9.93
101	1.062 2	1.086 6	2.30	514	1.363 7	1.353 1	-0.78
102	1.071 4	1.084 4	1.21	515	1.365 8	1.382 9	1.25
103	1.080 5	1.093 1	1.17	516	1.364 8	1.400 7	2.63
		...		517	1.365 2	1.430 6	4.79

(注:序列 91 对应 2020 年 1 月 23 日,为武汉“封城”开始日)

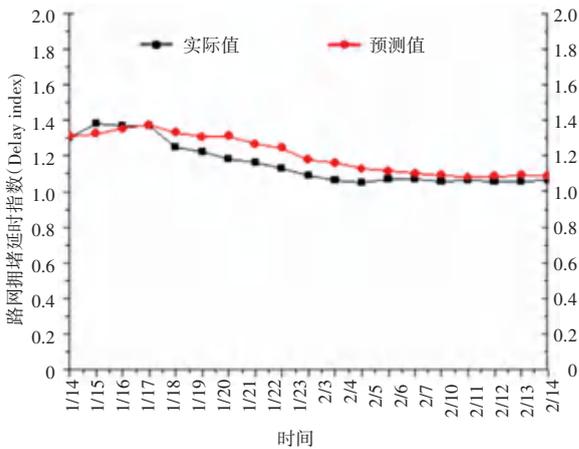


图6 新冠肺炎疫情发生前后的延时指数实际与预测值

Fig. 6 The actual and predicted value of the delay index before and after the emergence of COVID-19

5 结束语

本文利用杭州市西湖区 2019/9/15~2021/9/15 连续工作日的城市路网高峰行程延时指数数据样本,建立延时指数的 ARIMA 预测模型。同时考虑新冠肺炎疫情对城市道路交通运行状态的冲击,对样本内和样本外数据进行预测分析。对比结果误差小、精度高,证明 ARIMA 模型能够适应短时的数据突变,有一定的模型预测韧性,预测效果具有实用价值。未来研究将同时考虑长时间跨度下的空间因素变化对延时指数的影响。

研究过程到以下结论:

(1)新冠肺炎疫情特殊交通管控措施影响下,杭州市西湖区城市路网高峰行程延时指数显著下降,路网交通运行比往常工作日更畅通;

(2)ARIMA 模型能够捕捉短时的数据突变并预测时间序列趋势,对时间序列预测客观有效;

(3)城市路网高峰行程延时指数能够为城市交通管理者和出行者提供出行参考,其历史数据、现状实时数据以及预测数据都具有研究意义;

(4)时间序列数据中含政策效应因子,政策效应对预测序列的影响不仅在冲击前后时间段,其影响可能会持续一段时间。

参考文献

- [1] 袁浩,任杰,杨睿,等. 考虑空间异质性的城市建成环境对交通拥堵状态的影响[J/OL]. 武汉理工大学学报(交通科学与工程版);1-14[2021-11-03].
- [2] 冯海霞,王琦,杨立才,等. 拥堵环境下道路交通对城市空气质量的影响[J]. 山东大学学报(工学版),2021,51(1):128-134.
- [3] 汪传雷,吴雯宇. 绿色交通视角下车辆类型与交通拥堵关系研究——以合肥市为例[J]. 哈尔滨学院学报,2020,41(5):27-32.
- [4] 胡成雨,张焕明,朱家明. 基于多元 ANOVA 和主成分分析对我国城市交通健康的评价[J]. 佳木斯大学学报(自然科学版),2020,38(1):139-145.
- [5] 孟晓彤. 基于实时路况的兰州市城关区交通拥堵时空特征及对策探究[D]. 兰州大学,2018.
- [6] 李莎,林晖. 结合 MLR 和 ARIMA 模型的时空建模及预测[J]. 计算机工程与应用,2021,57(13):276-282.
- [7] 千银敏. 不同限行政策下郑州市区道路交通运行状况比较分析[D]. 郑州:郑州大学,2018.
- [8] LV Tiantian, WU Yanqin, ZHANG Le. A Traffic Interval Prediction Method Based on ARIMA[A]. Federation University Australia. Proceedings of the 5th International Conference on Machine Vision and Information Technology (CMVIT 2021)[C]. Federation University Australia; ,2021, 7:1-3.
- [9] ZHOU K, WANG W Y, HU T, et al. Comparison of Time Series Forecasting Based on Statistical ARIMA Model and LSTM with Attention Mechanism[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1631(1): 012141.
- [10] ERDELIC T, CARIC T, ERDELIC M, et al. Estimating congestion zones and travel time indexes based on the floating car data [J]. Computers Environment and Urban Systems, 2021, 87: 1-3.

(上接第 164 页)

- [8] Energy - Renewable Energy; New Renewable Energy Research from Institut Teknologi Sepuluh Nopember Described (Design and Implementation of Real-Time Monitoring System for Solar Power Plant in Surabaya, Indonesia)[J]. Energy Weekly News,2020.
- [9] 周俊,刘明,邱爱兵.基于 FPGA 的基 2-FFT 算法在谐波检测系统中的研究[J]. 电力电容器与无功补偿,2016,37(6):100-104.

- [10]王艳芬,王刚,张晓光,等. 数字信号处理原理及实现[M]. 北京:清华大学出版社,2017.
- [11]邱良丰,刘敬彪,于海滨. 基于 STM32 的全相位 FFT 相位差测量系统[J]. 电子器件,2010,33(3):357-361.
- [12]徐哲鑫,张艺林,林潇,等. 基于 OpenWrt 与 ZigBee 的智能家居路由器设计[J]. 计算机工程,2017,43(3):94-98,104.