doi:10.3969/j.issn.1672-1292.2016.03.006

基于ARIMA模型的城市公共自行车需求量短期预测方法研究

林燕平,窦万峰

(南京师范大学计算机科学与技术学院,江苏 南京 210023)

[摘要] 预测在城市公共自行车的研究中占重要地位,对站点未来需求量进行分析和预测,可为管理者提前分配自行车和用户合理制定出行方案提供依据.本文采用自回归求积移动平均(ARIMA)模型,对公共自行车高峰时段的需求量时间序列进行拟合和预测,并与基线法(Baseline)预测误差比较,结果显示对于不同站点类型的预测,此模型的预测值与实际值的平均相对误差均低于Baseline预测方法.ARIMA模型的预测精度相对较高,且预测结果可信,可为城市公共自行车管理和使用提供预测的理论与方法.

[关键词] 公共自行车,ARIMA模型,需求量,短期预测

[中图分类号]U491.1 [文献标志码]A [文章编号]1672-1292(2016)03-0036-05

Research on Short-Term Prediction Method of Demand Number in Urban Public Bicycle Based on the ARIMA Model

Lin Yanping,Dou Wanfeng

(School of Computer Science and Technology,Nanjing Normal University,Nanjing 210023,China)

Abstract:Prediction occupies an important position in study of urban public bicycle. Analyzing and predicting the demand numbers at every station in future can provide a basis,which managers allocate bicycles and the users make travel plan in advance. It is necessary to use the Autoregressive Integrated Moving Average(ARIMA)model,which models the demand number time series of public bicycle during peak hours of the week. Comparing with prediction error of the Baseline method,the results show that the average relative error of the value of the prediction and the actual are both lower than the Baseline prediction method for different stations. The prediction precision of the ARIMA model is relatively high,and the prediction result is credible. It provides theory and method of the prediction for management and use of the urban public bicycle.

Key words:public bicycle,ARIMA model,demand number,short-term prediction

随着快速城市化和机动化进程的推进,城市的布局和交通规划不够完善,不同交通之间的衔接不够紧密,城市居民面临“最后一公里”的尴尬问题.公共自行车的出现[1],不仅解决了此类问题,还有效地缓解了交通压力.在公共自行车运营系统中,由于交通潮汐性现象,早晚高峰时段常会出现双向流量严重不平衡的现象[2],部分租赁点单一方向租车需求量大,而另一方向停车需求量大,即在高峰期时段,一些站点呈现空位状态,用户无法借车;一些站点呈现满位状态,用户无法还车.

这推动了学者相继对公共自行车进行研究.公共自行车需求量预测是城市公共自行车研究的一部分[3].目前,对公共自行车预测的研究大多数采用调研法,数据量少且不连续;而有些使用数据挖掘[4-5]、抓取网站数据等方法来获取研究数据[6],研究预测的时间都较长,对于使用者来说,实际意义不大.关于预测的方法有很多,其中常用的有线性回归模型、非线性回归模型及时间序列模型等.由于公共自行车数据是自相关非平稳的时间序列,而ARIMA模型可有效地处理自相关的非平稳数据,因此,本文使用ARIMA

收稿日期:2016-05-16.

基金项目:国家自然科学基金(41171298).

通讯联系人:窦万峰,教授,研究方向:公共自行车项目.E-mail:douwanfeng@njnu.edu.cn

模型对站点自行车的需求量进行短期预测,便于使用者了解站点未来多长时间将有多少辆车被借或被还,根据这些信息来决定是否租借或归还到其他站点;使管理人员准确把握站点自行车的租赁情况,并对站点实施有效的车辆分配 .

1 数据描述

本文采用的数据来源于杭州市公共自行车站点后台运营的真实数据,为保证数据的有效性及消除异常样本对预测的干扰,对原始样本做了如下处理:在样本中剔除噪音数据,此种运营状况异常,若将此类数据作为样本,会大大降低预测精度 . 经过筛选,本文用于公共自行车站点需求量预测的数据样本有 160 个 .

从筛选的数据中可以掌握人们出行的规律性和时间模式 . 本文定义两个时间类型:工作日和周末/节假日 . 图 1 所示为两种工作日类型的需求量比较,工作日使用模式清晰地呈现出早高峰、晚高峰、平峰 3 个时段,其使用模式不同于周末,且需求量远高于周末,这说明城市居民在工作日对自行车的使用最为频繁,且可推断出使用人群以上班族为主;此外,每个站点有着不同的使用模式,人们的出行模式是由时间、地点等众多因素决定的 . 图 2 所示为两种不同站点类型的使用模式,居民区站点在工作日的高峰时段不同于商业区站点,这些使用模式特点与人们的出行时间和活动地点相关 .

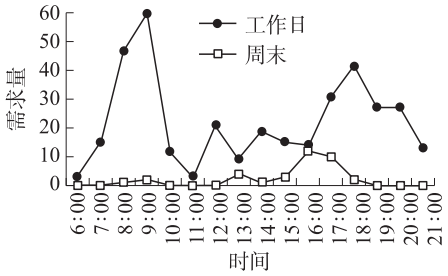

图 1 不同日期类型的需求量对比图

Fig.1 Demand comparison of different date types

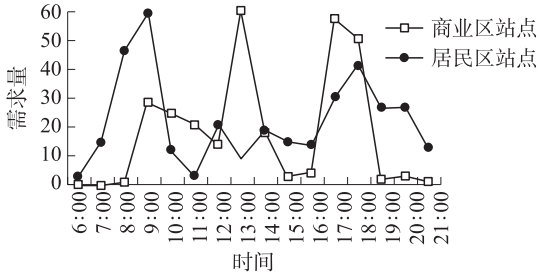

图 2 不同站点类型的需求量对比图

Fig.2 Demand comparison of different station types

图 3 为 2013 年 3 月份某站点所有工作日 8:00~10:00 高峰时段平均需求量变化图 . 由图 3 可知,此站点高峰时段持续时间较长,且不同工作日需求量变化呈现相似的使用模式 . 因此,本文选择工作日的高峰时段数据作为实验数据 .

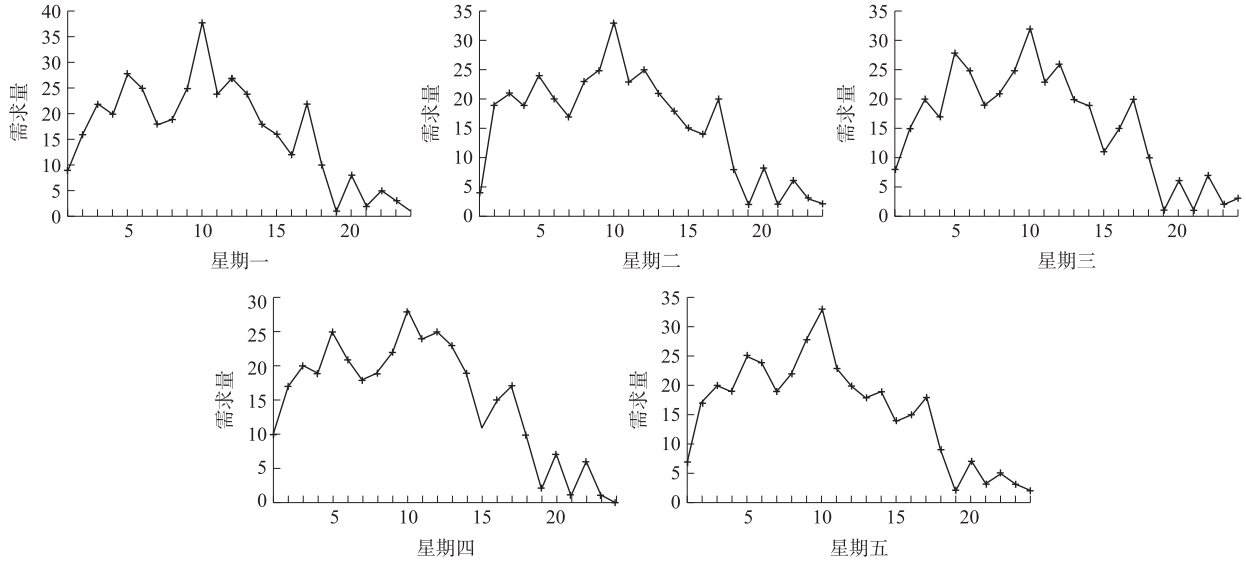

图 3 不同工作日的平均需求量对比图

Fig.3 Average demand comparison of different weekday types

2 ARIMA 模型介绍

对未来尚未观察的量进行预测或预报,是建立序列模型的主要原因之一 . 当模型作为简单的确定性

趋势加上零均值白噪声时,预测即等同于外推趋势.而当模型包含自相关关系时,便会充分利用其相关性,以得到更好的预测.ARIMA模型展示了该特性,并研究了有关预测的计算和性质.ARIMA模型是一种精确度较高的时间序列预测方法[7],它将预测对象随时间变化形成的序列,看作是一个随机序列,通过建立相应的数学模型,更本质地认识这些时间序列的内在结构和复杂性,从而达到在最小方差意义下的最佳预测.

ARIMA模型运用一阶差分的平稳过程,将非平稳的时间序列转化为平稳的时间序列,然后对滞后项不断迭代,回归因变量的滞后值和随机误差项的现值和滞后值.若一个时间序列$\{y_t\}$的d次差分$W_t=\nabla^d y_t$是一个平稳的ARIMA过程,则称$\{y_t\}$为自回归积分滑动平均求和模型.由于本文预测的是一组时间序列,而非一个值,需对其一般形式进行转换变形如式(1)所示:

$$Y_{m+1}(t)=\alpha_{11}Y_{11}(t-5)+\cdots+\alpha_{mp}Y_{mp}(t-p*5)+\beta_{11}\mu_{11}(t-5)+\cdots+\beta_{mp}\mu_{mp}(t-p*5)+N(t)\,, \tag{1}$$

式中,$Y_{m+1}(t)$表示提前m天预测站点在$[t-p*5,t]$时段的租借车数,此模型是由m天的1至p个步长的观测值和随机项线性组合而成.对于ARIMA(p,d,q)模型,d为差分次数,通常取d=1或最多为2,差分阶数d的确定是实验的第一步。$\nabla^d y_t=W_t$是平稳ARIMA(p,q)的过程,标准的假设是平稳模型有一个零均值,即实际上研究的是相对常数均值的偏离值,差分和对数变换是实现平稳性的有效方法.

3　ARIMA模型预测的实证分析

本文所需研究的是准确地预测公共自行车每个站点在工作日的需求量.由于自行车使用者的忍耐度是有限的,当租赁站点无车时,更长时间的等待是不现实的,因而选择提前5 min短期预测,而不是更长时间.本节运用变形的ARIMA模型分析已经获取的高峰时段需求量,统计20天工作日早高峰期使用数据,将早高峰按每5 min分为8个时间段,总共得到160个数据,通过做相关分析、模型定阶、参数估计、模型拟合,预测未来某一天的高峰时段每隔5 min的自行车需求量.

3.1　模型识别

原时间序列为非平稳的公共自行车需求量,须对其进行一阶差分处理,来消除其显著趋势.一阶差分的时间序列如图4所示,可以看出时间序列显著平稳.为证实该结论,本文对其做ADF检验,其显著性水平在0.05以下,拒绝存在单位根的原假设,说明一阶差分的时间序列是平稳的.

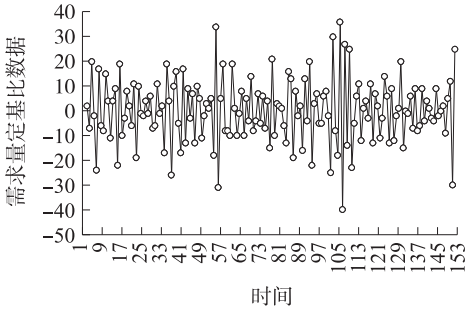

图4　一阶差分时序图

Fig.4　Time sequence diagram of first difference

在平稳性分析中,本文得到了差分的阶数d=1,经观察自相关和偏自相关函数图可知:自相关系数与偏自相关系数均是拖尾的.当模型ARIMA的阶数p=7,q=1时,AIC和BIC信息准则最小,且由各个模型系数的t检验统计量的值可知,其系数都显著不为零,此模型的参数估计结果如表1所示.

表1　参数估计结果表

Table 1　Parameter estimation results

变量	系数	标准误差	t-统计	概率
C	23.857 310	1.556 666	15.325 900	0.000 0
AR(1)	0.778 030	0.133 126	5.844 291	0.000 0
AR(7)	0.138 393	0.070 720	1.956 906	0.052 2
MA(1)	−0.733 505	0.147 269	−4.980 721	0.000 0

参数估计后,对拟合模型的适应性进行检验,实质是对模型残差序列进行白噪声检验.若残差序列不是白噪声,说明仍有一些重要数据信息未被提取,应重新定阶、设定模型.检验结果表明残差序列的样本自相关函数与偏自相关函数均可控制在95%的置信区间之内,因此,残差序列为白噪声过程.此外,根据ARIMA模型参数估计,ARIMA(7,1,1)模型如式(2)所示:

$$Y_2(t) = 0.778\,030Y_{11}(t-5) + 0.138\,393Y_{17}(t-35) + (-0.733\,505)\mu_{11}(t-5) + 23.857\,31. \quad (2)$$

误差项方差的估计值为 $\hat{\sigma}_a = 1.557\,713$. 模型的拟合效果如图 5 所示. 由图 5 可知，ARIMA 模型有明显较好的拟合效果，残差服从正态分布，说明此模型可以短期预测.

3.2 预测结果分析

根据上述分析结果，采用 ARIMA 模型对杭州市公共自行车站点的需求量进行短期预测，结果如图 6 所示，给出了预测结果图，曲线为实际的城市公共自行车需求量，黑色点为预测值，只有黑点的部分为预测区间，由图可以看出城市公共自行车的需求量存在明显的周期性. 构建 ARIMA 模型的关键在于时间序列的平稳性，通过差分来提高模型参数的平稳性. 但对于任意的非平稳模型，预测误差的方差会随前置时间的增加而无限增大，因为非平稳序列遥远的未来相当不确定.

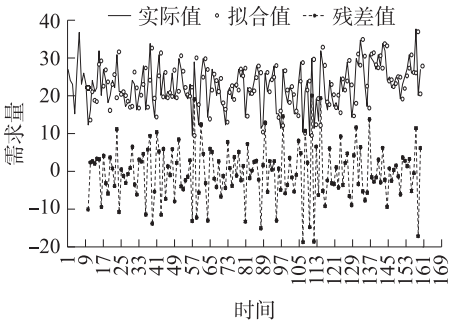

图 5 拟合模型图

Fig.5 Fitting model figure

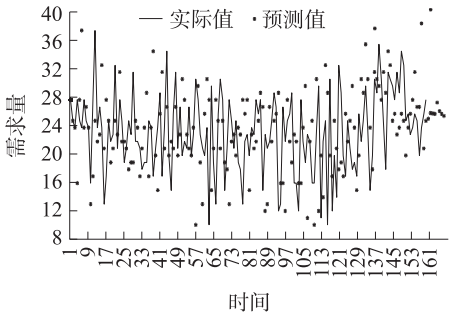

图 6 需求量预测结果图

Fig.6 Demand prediction results

为检验本文预测方法的性能，选用两个站点类型数据（商业中心类型和居民区类型），并分别与 Baseline 预测方法进行对比分析. 通过公共自行车实际的需求量测试数据集验证两种预测方法的相对误差及平均相对误差，如表 2 所示.

表 2 需求量预测相对误差

Table 2 Demand prediction relative error

时间序列		商业区站点相对误差		居民区站点相对误差	
序列	时间	Baseline	ARIMA	Baseline	ARIMA
161	8:20—8:25	0.086 957	0.099 565	0.130 435	0.012 609
162	8:25—8:30	0.000 000	0.087 500	0.333 333	0.045 833
163	8:30—8:35	0.160 000	0.041 600	0.160 000	0.058 800
164	8:35—8:40	0.346 154	0.000 385	0.038 461	0.008 462
165	8:40—8:45	0.200 000	0.079 667	0.333 333	0.046 333
166	8:45—8:50	0.400 000	0.055 200	0.120 000	0.015 200
167	8:50—8:55	0.259 259	0.037 778	0.222 222	0.036 296
168	8:55—9:00	0.423 077	0.011 538	0.038 461	0.065 385
平均相对误差		0.234 431 000	0.051 654 092	0.172 030 844	0.036 114 727

通过对两种方法预测相对误差分析，发现 ARIMA 预测两种类型的平均相对误差均小于 Baseline 法，这表明 ARIMA 模型的预测精度高于 Baseline 预测方法，且对两种类型的站点均适用. 因此选用 ARIMA 模型是合理的，可以做出相对精确度较高的预测. 此外，还发现两种方法居民区平均相对误差低于商业区，这是因为居民区用户的出行具有周期性，而商业区人们对于自行车的使用随机性强，影响了预测的精度.

4 结语

本文通过对后台记录的时间序列数据分析研究，以站点自行车高峰时段的需求量为样本数据，建立 ARIMA 模型对公共自行车时间序列进行分析和预测. 和 Baseline 预测方法比较显示，对于不同类型站点，ARIMA 模型平均相对误差均低于 Baseline 法，说明此模型的预测精度较好，具有一定的可信度，可为

城市公共自行车的管理者合理安排站点自行车提供数据支持.同时,当站点自行车无空位时,用户可以通过手机客户端查看目的站点5 min内将有多少辆车被还,也方便了城市居民的出行活动.由于公共自行车在运营过程中受诸多因素的影响,如天气、节假日等因素[8],未来的研究需将这些因素考虑在内,使得预测更具有实际指导意义.

[参考文献](References)

[1] 姚遥,周扬军.杭州市公共自行车系统规划[J].城市交通,2009,7(4):30-38.
YAO Y,ZHOU Y J. Bike-sharing planning system in Hangzhou[J]. Urban transport of China,2009,7(4):30-38.(in Chinese)

[2] 曹静,宫建,杨孝宽.解决北京市潮汐性交通拥堵的措施研究[J].武汉理工大学学报(交通科学与工程版),2009,33(6):1 116-1 119.
CAO J,GONG J,YANG X K. Analysis of resolving reversible traffic congestion in Beijing[J]. Journal of Wuhan university of technology(transportation science & engineering edition),2009,33(6):1 116-1 119.(in Chinese)

[3] 徐叶冉子,沈瑾.基于圆分布法和时间序列模型的公共自行车需求量分析[J].工业工程,2014,17(2):54-63.
XU Y R Z,SHEN J. Demand analysis of public bicycle system based on circular distribution method and time series model[J]. Industrial engineering journal,2014,17(2):54-63.(in Chinese)

[4] OLIVER O B,JAMES C,MICHAEL B. Mining bicycle sharing data for generating insights into sustainable transport systems[J]. Journal of transport geography,2014,34:262-273.

[5] CÔME E,OUKHELLOU L. Model-based count series clustering for bike sharing system usage mining,a case study with the V'lib's system of Paris[J]. ACM transactions on intelligent systems and technology,2014,5(3):1-21.

[6] ADVAIT S,NEAL L,CECILIA M. Comparing cities' cycling patterns using online shared bicycle maps[J]. Transportation,2015,42:541-559.

[7] BOX G E P,JENKINS G M. Time series analysis:forecasting and control[M]. San Francisco:Holden Day,1976.

[8] JONATHAN C,TIEBEI L,DAVID R,et al. Spatio-temporal patterns of a public bicycle sharing program:the effect of weather and calendar events[J]. Journal of transport geography,2014,41:292-305.

[责任编辑:严海琳]