

# 基于 ARMA 模型共享单车需求投放预测

张信<sup>1</sup> 王杨<sup>2</sup>

1 湖北工程学院新技术学院信息工程系, 湖北孝感, 432000;

2 枣阳市南城中心小学, 湖北省襄阳市, 441200;

**摘要:** 城市交通问题日益严重, 公共交通不具备很高的可达性, 在这种背景下, 共享单车系统孕育而生。本文通过鲸社区网站现有的共享单车系统运营的数据, 建立时间序列模型。将最优估算的方法以及改进算法运用于时间序列预测领域, 分析共享单车的运行特征, 并估计时间序列模型参数, 提高时间序列的预测准确性, 以便相关抉择者更好的管控。

**关键词:** 共享单车; 时间序列分析; ARMA 模型

**DOI:** 10.69979/3041-0673.24.11.003

## 引言

随着社会的进步, 人们的生活水准得到了明显的改善, 居民对汽车的需求多元化和对汽车品质的追求使得机动车的数量不断上升。交通拥堵问题日益严重, 为了杜绝此情况, 国家开始倡导绿色出行。现如今, 在绿色出行方式中, 地铁、公交车是最受欢迎的出行方式, 但无论是地铁还是公交都无法彻底解决最终的交通问题。在乘地铁和公共汽车旅行之后, 人们仍必须走路前往目的地。

发展共享单车既是解决这一问题的有效途径, 还能换乘带来便利, 解决节点交通问题; 与公交、地铁相比发展共享单车交通成本更低, 可以有效防止道路资源的浪费, 缓解停车压力。共享单车既能节能环保, 又能休闲和锻炼, 还能提升城市形象。

对于市民来说, 他们可以分享建设共享单车交通系统的成果, 不需要担心停车、修理和偷车等问题, 即可享受舒适的生活。一套有效的共享单车分配制度, 可以让所有的共享单车在不同的时间和空间上均匀分布。以满足乘客在早晚高峰时段的共享单车借还需求, 这对提高公交服务品质具有一定的现实意义<sup>[1-4]</sup>。

预测是一种科学方法, 在客观和历史资料发展的过程中, 人们可以利用有关的信息, 并利用科技知识工具与工具的相互联系, 在改变之前对客观事物的发展状况或趋势进行分析和推理。时间序列预测是根据某物的时间序列变化所反映的发展过程、趋势和方向, 通过分析时间序列来预测下一阶段或几年的变化。

由此可见, 时序预测对人类的日常生活有着深远的影响。时间序列预测在当前的社会生活中具有着重要的应用价值, 也是计算机应用领域研究的重点<sup>[5]</sup>。

本文以共享单车系统为研究对象, 着重阐述了共享单车在城市发展中的作用, 并对其参考价值进行了分析。

优化共享单车规划可以更好地将其与其他公共交通连接起来, 促进公共交通出行, 缓解早晚高峰拥挤问题。

## 1 理论模型

线性预测模型包括: 自回归模型、移动平均模型和自回归移动平均模型。ARMA 模型的基本思想是: 根据预测对象的时间而得到的一系列数据, 可以被视为一套随机序列, 由某个数学模型近似描述, 这个模型就能从过去和现在的时间序列值中预测出将来的值。现代的统计学和经济的测量模式对公司的未来有一定的帮助。

### 1.1 ARMA 模型的定义

ARMA 模型的定义为:

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(\varepsilon_t x_s) = 0, \forall s < t \end{cases} \quad 2-1$$

### 1.2 ARMA 模型的平稳性及可逆性

#### 1. 平稳条件与可逆条件

平稳条件:  $|\lambda_i| < 1 (i=1, 2, 3, \cdots, p)$ , 也就是  $\Phi(B) = 0$  根都在单位圆外

可逆条件:  $|\lambda_i| < 1 (i=1, 2, 3, \cdots, q)$ , 也就是  $\Theta(B) = 0$  根都在单位圆外

#### 2. 传递形式与逆转形式

传递形式:

$$x_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j} \quad 2-2$$

其中  $\{G_0, G_1, G_2, \cdots\}$  为 Green 函数。

可逆形式:

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)x_t = \sum_{j=0}^{\infty} I_j x_{t-j} \quad 2-3$$

其中  $\{I_1, I_2, \dots\}$  为可逆函数。

### 1.3 ARMA(p, q) 模型的统计性质

1. 均值:

$$Ex_t = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} \quad 2-4$$

2. 自协方差函数

$$\gamma_k = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} G_i G_{i+k} \quad 2-5$$

3. 自相关函数:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2} \quad 2-6$$

通过绘制分析自相关图 (ACF 图) 和偏自相关图 (PACF 图) 的特征再确定合适的 p、q 值, 总结出如表 2-1 的规律。

表 2-1

适用模型	ACF	PACF
AR(p)	拖尾	p 阶截尾
MA(q)	q 阶截尾	拖尾
ARMA(p,q)	拖尾	拖尾

### 1.4 ARIMA 模型的原理和结构

ARIMA 模型记作 ARIMA(p, d, q), p 为自回归项数; q 为滑动平均项数, d 为使序列平稳的差分次数。

AR 模型和 MA 模型可以组合成 ARIMA 模型, ARIMA 模型是对非平稳序列建立的。非平稳时间序列要先经过差分转化为平稳序列后建立 ARMA 模型。

## 2 实证分析

### 2.1 数据来源及描述

本次数据来源于和鲸社区, 采用 csv 文件。数据共 10886 行, 12 列。接下来用表 3.1 解释本次数据集中的各项数据的含义:

表 3.1

英文表示	含义
datetime	时间, 具体到年月日时
season	有 1,2,3,4 四个属性分别代表春夏秋冬
holiday	0 非节假日, 1 节假日
workingday	0 代表周末, 1 代表工作日
weather	1 晴天、2 雾天、3 雪天、4 雨天
temp	温度
atemp	体感温度
humidity	湿度
windspeed	风速
count	总租车数目
casual	非注册用户个数
registered	注册用户个数

### 2.2 平稳性检验和白噪声检验

对数据进行平稳性检验, 检验结果见表 3.1。

表 3.1: 平稳性检验

ADF 检验表							
变量	差分阶数	t	P	AIC	1%	5%	10%
count	0	-2.002	0.285	7147.389	-3.445	-2.868	-2.57
	1	-11.752	0.000***	7134.661	-3.445	-2.868	-2.57
	2	-11.684	0.000***	7152.502	-3.445	-2.868	-2.57

\*\*\*、\*\*、\* 分别代表 1%、5%、10% 的显著性水平。

表 3.2: 白噪声检验

自相关性					
序列: count					
延迟	自相关性	标准误差 a	值	自由度	显著性 b
1	-.346	.047	54.905	1	.000
2	-.088	.052	58.462	2	.000
3	-.014	.053	58.556	3	.000
4	-.120	.053	65.158	4	.000
5	.037	.053	65.774	5	.000
6	.041	.053	66.568	6	.000

根据表 3.1 检验的结果,可知该序列一阶差分平稳;  
由表 3.2 中检验结果,可知该序列为平稳非白噪声序列。

## 2.3 拟合 ARIMA 模型

1 阶差分后序列的自相关图和偏自相关图如下。

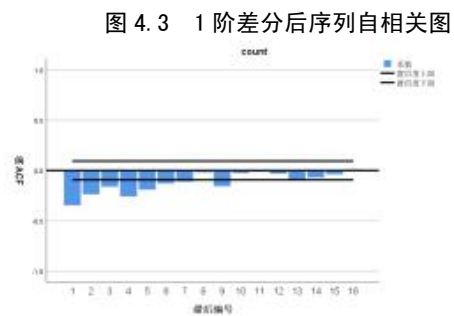
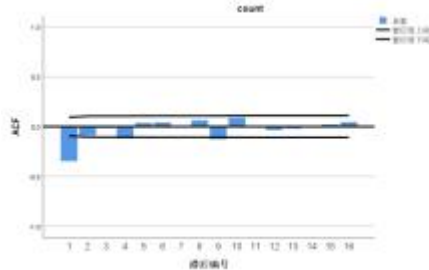


图 3.1: 1 阶差分后序列自相关图及偏自相关图

1 阶差分后序列的自相关图呈现 1 阶拖尾特征, 偏自相关图呈现 7 阶拖尾特征。使用 ARIMA(0, 1, 1)、ARIMA(7, 1, 0)、ARIMA(7, 1, 1)模型拟合原序列。

表 3.3: 模型检验对比

模型	模型参数检验	模型残差白噪声检验	模型是否可用
ARIMA(0,1,1)	通过检验	不通过检验	否
ARIMA(7,1,0)	通过检验	通过检验	是
ARIMA(7,1,1)	不通过检验		否

根据表 3.4 模型检验对比情况, 可知 ARIMA(7, 1, 0)模型通过检验, 最终选择该模型进行预测。

## 2.4 模型预测

用 ARIMA(7, 1, 0)对未来 10 天进行预测, 预测数据如表 3.5, 模型拟合预测图如 3.2。

表 3.4: 预测数据

模型	预测_count_模型_1	LCL_count_模型_1	UCL_count_模型_1
457	5299	3633	6964
458	5128	3330	6925
459	4952	3106	6797
460	4942	3066	6818
461	4920	3033	6807
462	5033	3111	6955
463	5139	3158	7121
464	5117	3069	7164
465	5091	2956	7226
466	5047	2856	7239

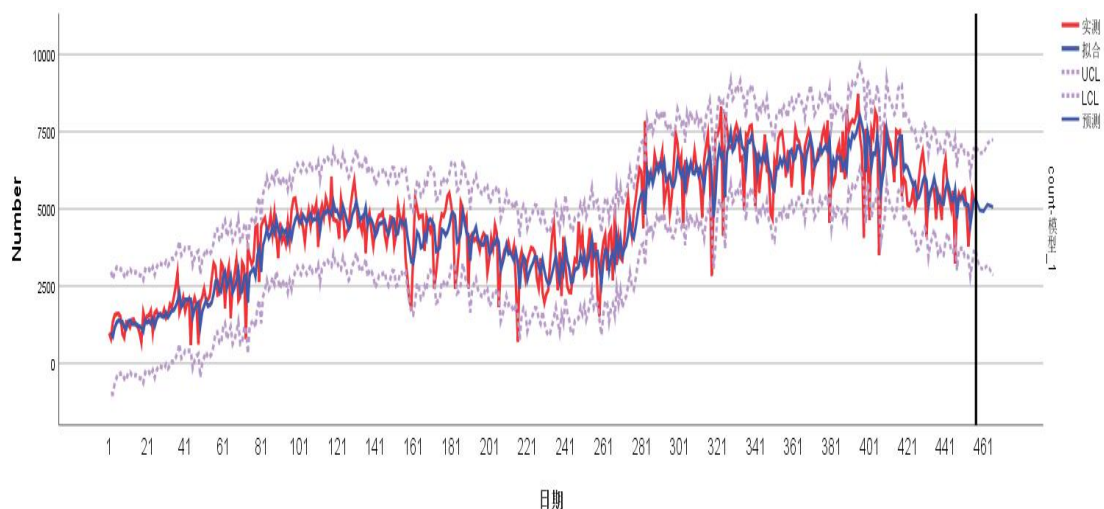


图 3.2: 模型预测图

### 3 结论与展望

#### 3.1 本文总结

本文从共享单车序列出发检验序列,通过差分使序列变成平稳序列,得到三个待选择 ARIMA 模型;分别对三个模型进行拟合查看模型参数确定模型有效性,根据三个模型的残差检验情况,选出最优模型,最后利用最优模型进行预测。本文的主要研究工作如下:

首先对数据进行处理,了解各个因素对共享单车使用量的影响,得出共享单车的使用群体大多为注册用户。

时间序列的预处理,通过介绍平稳性检验的方法,为判断时间序列为平稳或者非平稳以及如何验证该序列平稳性奠定了基础。给出 ARIMA 模型及 ARIMA 模型理论,方便预测。

对序列进行预测,首先通过时序图和 ADF 检验判断差分后的时间序列平稳性,再对差分平稳序列进行白噪声检验。其次对平稳非白噪声的差分后序列进行 ARMA 模型拟合,得到差分后的自相关图和偏自相关图,然后分别判断自相关图和偏自相关图有几阶截尾特征,对拟合的 ARIMA 模型进行参数显著性检验,参数有效再进行残差白噪声检验,确定模型并对未来进行预测。

#### 3.2 未来展望

本文研究采用相关历史数据,共享单车的使用量同时受多种因素的影响,怎样运用数据挖掘技术,从历史资料中获得更多有用的资讯与规则,支持对公共自行车需求的预测,有必要作更深入的探讨。

对 ARIMA 模型进行持续优化,可以改进模型的精确性;通过对季节、天气等多种因素的综合训练,可以提高模型的预测精度。在时间序列预测中引入群体智力的概念,能够改善预测的准确性。

#### 参考文献

- [1]刘臻.城市公共自行车运营中的多车场车辆调配优化研究[D].北京交通大学,2014.
- [2]张建国.城市公共自行车车辆调配问题研究[D].西南交通大学,2013.
- [3]罗赞.城市公共自行车选择行为研究[D].长安大学,2013.
- [4]谭玉龙.基于马尔可夫链模型的公共自行车站点供需研究[D].西南交通大学,2015.
- [5]向书坚.20 世纪 90 年代时间序列预测领域主要研究动态[J].中南财经大学学报,2001,125(2).