Computer Science and Application 计算机科学与应用, 2023, 13(12), 2474-2499
Published Online December 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.1312247

Hans汉斯

多种数据补全策略对商超客流量预测影响研究

关家豪[1]，屈俊峰[2]

[1]河北地质大学信息工程学院，河北 石家庄
[2]湖北文理学院计算机工程学院，湖北 襄阳

收稿日期：2023年11月26日；录用日期：2023年12月21日；发布日期：2023年12月29日

摘 要

数据的准确性对传统线下零售商超的客流量预测有着较为显著的影响。通过对某商超2021年的客流量分析发现，受ERP系统不稳定的影响，传统线下零售商超的客流量曾出现了明显的波动，在2021年3月初到2021年7月末，客流量数据出现了断崖式下降。因此在进行客流量预测时，应当针对数据不稳定性对数据进行一定的特征工程处理。为了探究不同特征工程对客流量预测的影响，我们提出了采用删除异常值、Arima模型填充异常值、Mean-Value填充异常值等多种特征工程处理异常值的方式；使用SArima模型、Holt-Winters模型。在多轮实验中，实验训练集将包含2021年3月初到2021年7月末。并且时间序列模型的参数将以数据为准，选择最优参数进行训练，最终实验生成的模型都将使用2021年8月1日至2021年8月31日的数据做预测验证。结果显示，当不使用任何数据补充的特征工程策略时，Holt-Winters模型较好，而使用了数据补全策略后，SArima模型的预测能力得到了显著提升，并且使用Arima模型进行的数据填充，在一定程度上可以提高SArima模型的预测能力。

关键词

客流量预测，时间序列，ARIMA模型，Holt-Winters模型，特征工程

Research on the Influence of Multiple Data Completion Strategies on the Forecast of Supermarket Passenger Flow

Jiahao Guan[1], Junfeng Qu[2]

[1]School of Information Engineering, Hebei GEO University, Shijiazhuang Hebei
[2]School of Computer Engineering, Hubei University of Arts and Sciences, Xiangyang Hubei

Received: Nov. 26th, 2023; accepted: Dec. 21st, 2023; published: Dec. 29th, 2023

文章引用: 关家豪, 屈俊峰. 多种数据补全策略对商超客流量预测影响研究[J]. 计算机科学与应用, 2023, 13(12): 2474-2499. DOI: 10.12677/csa.2023.1312247

Abstract

The accuracy of data has a significant impact on the customer flow prediction of traditional offline retailers. Through the analysis of the passenger flow of a supermarket in 2021, it is found that due to the instability of the ERP system, the passenger flow of traditional offline retailers had a significant fluctuation, and from the beginning of March 2021 to the end of July 2021, the passenger flow data showed a cliff decline. Therefore, when the passenger flow forecast is carried out, it is necessary to carry out some feature engineering processing to the data in view of the instability of the data. In order to explore the influence of different feature engineering on passenger flow prediction, we propose several feature engineering methods to deal with outliers, such as deleting outliers, filling outliers with Arima model, and filling outliers with Mean-Value. Use the SArima model, Holt-Winters model. In multiple rounds of experiments, the experimental training set will cover the period from early March 2021 to the end of July 2021. In addition, the parameters of the time series model will be based on the data, and the optimal parameters will be selected for training. The final models generated by the experiment will use the data from August 1, 2021 to August 31, 2021 for prediction verification. The results show that the Holt-Winters model is better when no data-supplemented feature engineering strategy is used, but the prediction ability of the SArima model is significantly improved after the data completion strategy is used, and the data filling with the Arima model can improve the prediction ability of the SArima model to a certain extent.

Keywords

Passenger Flow Forecast, Time Series, ARIMA Model, Holt-Winters Model, Feature Engineering

Copyright © 2023 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

由于传统线下零售商超的数字化、信息化能力不足，因此 ERP 系统中的存在大量的沉睡数据。如何科学合理的刺激市场经济，满足市场需求，增强经济增长的内生动力，为市场主体的发展营造更好的环境，是企业的重要战略。为了应对多变的市场变化，制定合理的零售策略，我们将通过对客流量进行预测分析[1]，并充分考虑老旧 ERP 系统不稳定因素的影响，除了采用常用的均值、去除空值等方式来对数据的异常值进行预处理外[2]，我们还使用 Arima 时间预测模型对实验数据进行异常值补全，最后利用、SArima 模型、Holt-Winters 模型进行客流量的预测。本文旨在研究多种数据补全策略对商超客流量预测的影响。Arima 模型时常用的时间预测模型，可以用作短期数据补全。SArima 模型和 Holt-Winters 模型是时间序列中常用的预测模型[3] [4]。对客流量的预测具有一定的应用研究价值。本研究提出了基于多种数据补全策略特征工程的 SArima 模型和 Holt-Winters 模型，并利用 2021 年 2 月至 2021 年 4 月某商超线下实际客流量进行了模型训练和预测，来预测多种数据补全策略对商超客流量的影响[5]。同时我们还进行了对比研究与误差分析，以评估模型的准确性和实用性。

总体而言，本文的研究为预测商超购物客流量提供了有效的理论依据与可行的方法。通过使用多种数据补全策略对数据进行预处理；引入不同的模型，使我们能够更加精准的预测商超客流量，为企业的下一步规划，提供了强有力的理论依据，让数据发声，用数据说话。该研究对于零售行业的销售预测、

运营策略、销售营销都具有一定的理论和实践意义。

2. 文献综述

客流量预测是零售企业管理的重要内容[6]，它可以帮助提高顾客消费转化效率，优化资源配置等[7]。客流量预测的方法有很多，其中一种常用的方法是基于时间序列的分析[8]。时间序列是指按照时间顺序排列的数据序列，它反映了数据随时间变化的规律和趋势。时间序列分析是利用历史数据来建立数学模型，从而对未来数据进行预测的技术。

近年来，很多学者对此进行了研究，主要分为以下三种预测的方法：

1) 基于统计学方法的客流量预测[9] [10]

时间序列分析的方法有很多，其中比较经典和常用的是自回归移动平均模型(ARMA) [11]和自回归综合移动平均模型(ARIMA) [12]。这两种模型都是基于线性回归的思想，将当前时刻的数据表示为过去时刻数据和误差项的线性组合。ARMA 模型由两部分组成，分别是自回归部分(AR)和移动平均部分(MA)。AR 部分表示当前时刻的数据与过去时刻数据之间的关系，MA 部分表示当前时刻的数据与过去时刻误差项之间的关系。ARMA 模型适用于平稳时间序列，即时间序列的均值、方差和自相关性不随时间变化的序列。

然而，在实际应用中，很多时间序列并不是平稳的，例如客流量数据可能受到季节性、趋势性等因素的影响[13]。为了处理非平稳时间序列，可以对原始数据进行差分处理，即用相邻两个时刻的数据之差代替原始数据，从而消除非平稳性[14]。这样得到的新序列称为差分时间序列，它可以用 ARMA 模型进行建模和预测。这种在 ARMA 模型基础上引入差分处理的方法称为自回归综合移动平均模型(ARIMA)。

何雪晴(2019)使用网络搜索数据和降噪处理进行的旅游客流量预测研究中，引入了经验模态分解方法进行降噪处理，使用经验模态分解方法对每个搜索指标数据分别进行了分解[15]，一定程度上避免了不同搜索指标之间噪声的干扰，提升了预测的精度。

段然(2017)对某铁路局全部列车旅客乘车数据进行清洗和规约，将数据划分为节假日和非节假日两种类型，对具有周期性的非节假日数据选用在传统 ARIMA 模型中加入周期性的 SARIMA 模型进行预测，而对节假日数据则选用波动系数模型进行预测[16]。最后，通过计算预测值与真实值的相对误差来评价模型的预测准确度，最终验证 SARIMA 模型与波动系数模型相结合的方法对铁路客流量的预测较为准确。

2) 基于机器学习的客流量预测[17]

方昇越(2022)基于 XGBoost 模型，对短时客流量进行预测[18]，过对地铁客流内外部因素的特征分析，判断客流在时间和空间上的分布特点，确定相邻时段客流、预测时段、工作日和雨雪天气四个特征变量，最终确定短时客流量 XGBoost 模型，可以有效提升客流量预测的精度[19]。

鄢仕林(2022)通过对机场航线客流量分析，构建了 GM(1,1)模型对序列的趋势性进行预测，并使用支持向量机模型对模型预测值的残差进行修正补偿[20]，通过实证分析发现该模型预测精度较高，可以为航线新开航线客流量预测提供决策支持。

马毅(2022)提出的一种改进的 PAPSO-BP 预测算法模型，对短期客流量情况进行了预测，改进粒子群算法。针对传统粒子群算法容易陷入局部最优值这一缺陷，对其做出改进：将惯性权重设置为随粒子适应度值自适应变化，且在位置更新公式中加入扰动因子[21]，仿真结果证明该改进的 PAPSO 方法有效提高了算法精确度。

冒志恒(2022)使用神经网络模型，结合历史数据对短时客流量进行预测。考虑到反向传播(Back Propagation, BP)神经网络的不足，提出改进 BP 神经网络模型预测精度的方法，构建了两种组合预测模型，即 GA-BP 模型和 MEEMD-BP 模型[22]，最终结果表明，组合预测模型在提升短时客流量预测的准确性

上都有较好的效果，但使用 MEEMD 算法改进数据质量的 MEEMD-BP 组合模型对 BP 神经网络模型的预测精度提升更大。

蔡纯(2019)针对游乐园行业、管理者和游客三方面主体进行了细致的需求分析；然后，以此为依据，设计了基于多源数据的游乐园运营分析指标体系，旨在从多元视角，整合乐园游客、设备、销售等相关的实时运营数据，分析挖掘游乐园多源数据背后的关联与规律。使用传统回归预测模型和三种数据挖掘预测模型——SVR 模型、决策树回归模型和 BP 神经网络模型进行比较[23]，分析了适用于游乐园客流量预测的有效模型。

成翔(2018)针对商铺客流量的特点，分析了影响商铺客流量的环境因素和产品因素，研究了不同因素导致的影响，为取舍预测数据建立了理论基础，再对数据进行学习从而提出最适当的数据分类的方法，并在理论上分析了不同的商铺客流量预测的方法[24]，证明了基于 BP 神经网络的商铺客流量预测是可行的。

3) 结合多种方法的客流量预测[25]

邓雨菲(2022)针对旅游客流量的真实数据，基于 ARIMA-ATT-LSTM 模型对景区客流量进行精准预测[26]，为高峰期旅客爆满、景区拥堵、环境破坏、安全隐患、淡季旅游资源闲置等问题提供了理论支持。

常昊(2022)采用 LSTM 神经网络对地铁短时客流量进行预测，通过建立 ARIMA 短时客流预测模型实现周期性客流的波动特征提取，采用 Light GBM 预测算法实现对客流峰值的泛化，建立 LSTM 预测模型，在此基础上，提出 LGB-LSTM 并联模型，通过对客流空间和时间数据特征提取[27]，解决预测精度和非线性最优拟合问题。

殷志敏(2022)使用杭州市 AFC 系统的所有地铁站刷卡数据信息，对数据进行处理后建立短期客流量预测模型，从站点和时间两个方面对客流量数据特征进行分析，对工作日和周末分别建立客流量 LightGBM 预测模型，将 LightGBM 模型和 LSTM 模型进行加权融合[28]，最终预测精度得到提升。

吕高帆(2019)使用 K-means 聚类算法对原有历史数据进行数据聚类处理，结合 CART 机器学习算法及统计学分析方法的优势，提出一种建设性较强，且适用性突出的无人超市客流量组合预测模型[29]，为管理者提供比较及时的客流分布，而且可以更好地为决策者提供可参考的决策信息。

张璐(2022)基于最小二乘支持向量机风电功率预测模型求解短期风功率预测问题，采用了一种基于密度的聚类算法以目标风功率属性特征密度作为分群指标模型对训练样本进行聚类与划分，对风电机组的异常数据进行识别并排除[30]，提高模型预测结果的准确性。

3. 模型方法

3.1. Arima

Arima (Autoregressive Integrated Moving Average model)模型，即差分整合移动平均自回归模型，又称整合移动平均自回归模型或整合滑动平均自回归模型[31]，是时间序列预测分析方法之一，通过发现数据之间的相互关联性，以时间为轴，进行线性预测。Arima 模型主要由三部分构成：AR、I、MA。

AR 是“自回归”(Autoregressive)模型，因为其变量只包含一个值，即它本身，所以称其为自回归。AR(p)模型的含义为当前的时间点所对应是值等于过去 p 个时间点所对应的值的回归，其中 p 为阶数，公式为：

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t + \mu .$$

I 是集成(Integrated)的含义，是对数据进行了关于时间的差分，因为时间序列分析对数据的平稳性具有一定的要求，因此在进行模型训练前，应通过一定的数学方法将现有数据序列转换为平稳的数据序列，一般采用的方法是进行差分，公式为：

$$\Delta X_t = X_t - X_{t-1} = X_t - LX_t = (1-L)X_t$$

用 t 时刻的数据值减 $t-1$ 时刻的数据值，此时得到的新的数据序列称之为 1 阶差分序列。按照此规律，再次进行相同操作将得到 2 阶差分序列，公式为：

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = (1-L)X_t - (1-L)X_{t-1} = (1-L)^2 X_t$$

一般的时间序列算法最多进行 2 阶差分，若数据依然不具备平稳特征，此时可以认为数据缺乏平稳性，不具备进行时间预测的前提，此时应当更换数据集，或是探究造成数据不平稳的缘由。

由于数据具有一定的周期性，所以引入了一种特殊的季节性差分 S，即将原本的数据按照一定的周期拆分，让 t 时刻的数据值减 $t-T$ 时刻的数据值，此时得到的便是带有季节性的差分序列。

MA 是移动平均(Moving Average)模型或滑动平均模型，MA(q)是当前时刻的数据值等于过去 q 个历史预测误差值的回归，其中 q 为阶数，公式为：

$$X_t = \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_q\varepsilon_{t-q} + \varepsilon_t + \mu$$

将上述的 AR(p)和 MA(q)两个模型进行结合，就得到了 Arima 模型，公式为：

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \beta_1\varepsilon_{t-1} + \cdots + \beta_q\varepsilon_{t-q} + \varepsilon_t + \mu .$$

由于时间序列算法在进行训练时，要求数据必须是平稳的。因此在数据准备阶段，我们应当对数据的平稳性进行检验，如果数据不平稳，那么应当进行差分，至多 2 轮后，若数据平稳，继续进行实验，其中差分阶数 d 是 Arima 模型中参数 d 的值。通过 AIC_BIC 检验，确定 p,q 的值，最后利用 Arima(p,d,q)模型进行预测。Arima 模型实验流程图 1 如下：

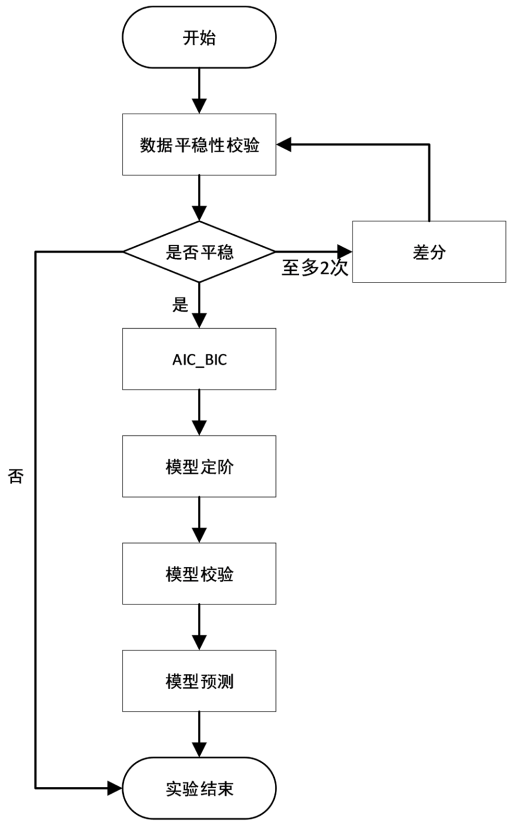

Figure 1. Flowchart of Arima model experiment
图 1. Arima 模型实验流程图

3.2. Holt-Winters

Holt-Winters 模型，是在 Holt 模型的基础上引入了 Winters 季节项，是一种较为高效的预测模型。与其他时间序列模型只能使用平稳序列数据不同，它对有线性趋势和周期波动的非平稳序列也同样适用。

Holt-Winters 模型主要分为如下两类：

1) Holt-Winters 加法模型又称为加性季节模型。因此我们应当有 v_t 用于描述时间序列的线性趋势也就是趋势因子 u_t，设两个平滑参数 α，β，公式如下：

$$u_t = \alpha \times (x_t - s_{t-T}) + (1-\alpha) \times (u_{t-1} + v_{t-1})$$

$$v_t = \beta \times (u_t - u_{t-1}) + (1-\beta) \times v_{t-1}$$

同时还应当有对季节 T 进行描述的季节因子 s_t，设一个平滑参数 γ，公式如下：

$$s_t = \gamma \times (x_t - u_t) + (1-\gamma) \times s_{t-T}$$

我们假定时间序列 x_t 的趋势因子 u_t 与季节因子 s_t 是相加关系，公式如下所示：

$$x_t = u_t + s_t$$

三个平滑参数 α、β、γ，其取值都在 0 到 1 之间，是模型预报值与实测预测值之间的平衡权重。参数 α、β、γ 越大，表示时间序列 x_t 的非平稳性越强，模型的可预报时间越短。如果能用较小的参数 α、β、γ 与历史数据吻合上，则可预报时间较长。

2) Holt-Winters 乘法模型。

Holt-Winters 乘法模型又称为乘性季节模型，与加法模型类似，我们设三个平滑参数 α、β、γ，其取值都在 0 到 1 之间，因此将得到如下公式：

$$u_t = \alpha \times (x_t \div s_{t-T}) + (1-\alpha) \times (u_{t-1} + v_{t-1})$$

$$v_t = \beta \times (u_t - u_{t-1}) + (1-\beta) \times v_{t-1}$$

$$s_t = \gamma \times (x_t \div u_t) + (1-\gamma) \times s_{t-T}$$

可以看到用于描述时间序列的线性趋势的 v_t 不变，u_t 和 s_t 中的减法运算变成除法运算。与加法模型不同，在乘法模型中，我们将假设 u_t 和 s_t 相乘后的结果是 x_t

$$x_t = u_t \times s_t$$

在预测是将三个平滑参数 α、β、γ 的值都设置为 0，此时根据公式计算预测值。因此乘法模型是非线性模型，能够更好的处理随趋势成分变化的季节波动的振幅变化，所以它比加法模型更依赖一个好的初始值。

Holt-Winters 模型的具体实验流程如图 2。

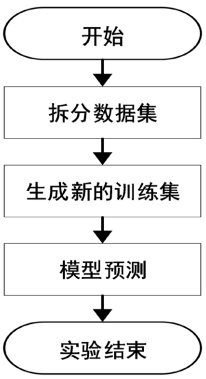

Figure 2. Flow chart of Holt-Winters model experiment
图 2. Holt-Winters 模型实验流程图

3.3. Mean-Value

Mean-Value 方法，称为均值填充方法，是常用的异常值填充方法，可以分为两类均值填充方法。

全局均值填充，每次填充时，都计算将当前全部数据的平均值，将本次计算的平均值用作填充异常值。并在下次计算时，将本次结果也用作计算集的一部分。具体公式如下：

$$x_t = \frac{\sum_{i=1}^{n} x_i}{n}$$

区域均值填充，每次填充时，将采用当前值前后 m 个区域内的数字作为数据集，计算平均值，当遇到连续异常值时，应当跳过异常值向后取数，进行填充，具体公式如下：

$$x_t = \frac{\sum_{i=t-m}^{t+m} x_i}{2m}$$

全局均值填充可以更好的对平稳数据填充，但较容易收到极值的影响，而导致填补值出现异常。

区域均值填充可以对异常值数据的前后进行记录，不易收到极值影响，较为稳定，因此在本实验中，我们将使用区域均值填充作为均值填充算法。具体实验步骤如下图 3。

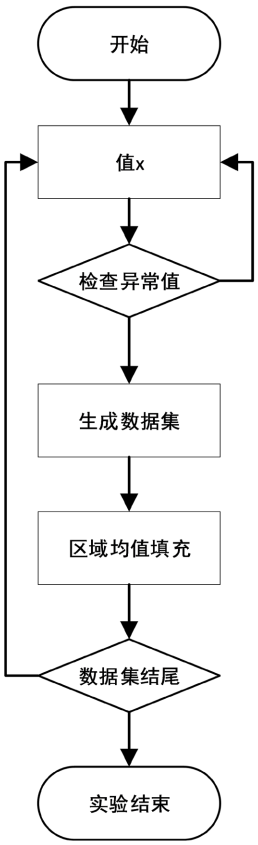

Figure 3. Flowchart of Mean-Value model experiment
图 3. Mean-Value 模型实验流程图

4. 实证分析

从某零售企业获取某卖场的每日成交数据，用于作为客流量预测的数据集，分别对数据集的异常值

进行删除空值、Mean-Value 填充、Arima 填充等方法进行处理，然后创建基于多种数据补全策略的 Arima 模型、SArima 模型和 Holt-Winters 模型进行预测，采用不同的评价指标对模型进行对比，来分析多种数据补全策略对零售商超卖场客流量预测的影响。

4.1. 数据准备

客流量数据来自于某头部零售商超卖场成交数据，由于数据量较为庞大，且因为 ERP 系统不稳定，因此数据中存在大量的异常值，因此在使用这些数据进行训练模型前，应当对数据进行预处理。常用的数据预处理方式包括数学方法、统计学方法等。例如在处理异常值时，应当分为如下状况：

重复值：当发现重复值时，应当删除多余重复值，并只保留一条有效数据。

缺省值：当数据中出现缺省值时，应先判断数据类型，如果是有规律数据，应当按照客观规律填入；例如，4 月 5 日，[缺省值]，4 月 7 日，则此时的缺省值应当填入 4 月 6 日。若是数据为统计类数据，应当按照数据规则，填入平均数、中位数或以其他数学规则填入，如果无法找到对应规则填入，应当删除这条无效数据。

异常值：当数据中发现数据远超出于其他数据时，即数据极大或数据极小，又或者是数据不符合常识，例如，顾客生日为 1900 年 1 月 1 日；这时应当按照缺省值的补全策略来完成数据的修补。

使用上述规则对数据进行清洗后，可以基本保证数据的准确性与有效性。

在数据存入数据仓库后，我们就可以依据自己的实验，来抽取所需要的数据，使用数据导出工具，框选好时间日期后，选择相对应的卖场，对全部成交数据进行汇总，得出了某卖场的真实客流量数据。

本文中选取了某头部零售商超卖场真实客流量数据，由于数据在进入数据仓库时只进行了简单的清洗，去除了无效值，因此我们对异常值的处理将按照实验规划，采用删除空值、Mean-Value 填充、Arima 填充方法进行处理。同时为了保证实验的严谨性、准确性、可复用性，实验的测试集将只进行数据清洗，不进行任何数据补全，且数据完备率大于 95%，并保证测试集完全相同。所以我们将以时间升序的方式，对数据进行抽取。将时间作为索引项，可以绘制出如下的折线图，为了构建疫情因素影响的模型，我们将使用 2021 年 3 月 1 日至 2021 年 8 月 31 日的数据作为本次实验的数据集。其中 2021 年 3 月 1 日至 2021 年 7 月 31 日的数据作为训练集，2021 年 8 月 1 日至 2021 年 8 月 31 日的数据集作为测试集。生成的基于多种数据补全策略不同时间序列图像如下图 4、图 5、图 6、图 7 所示。通过对模型的训练最终实现预测，并对比分析各个模型的精准度。

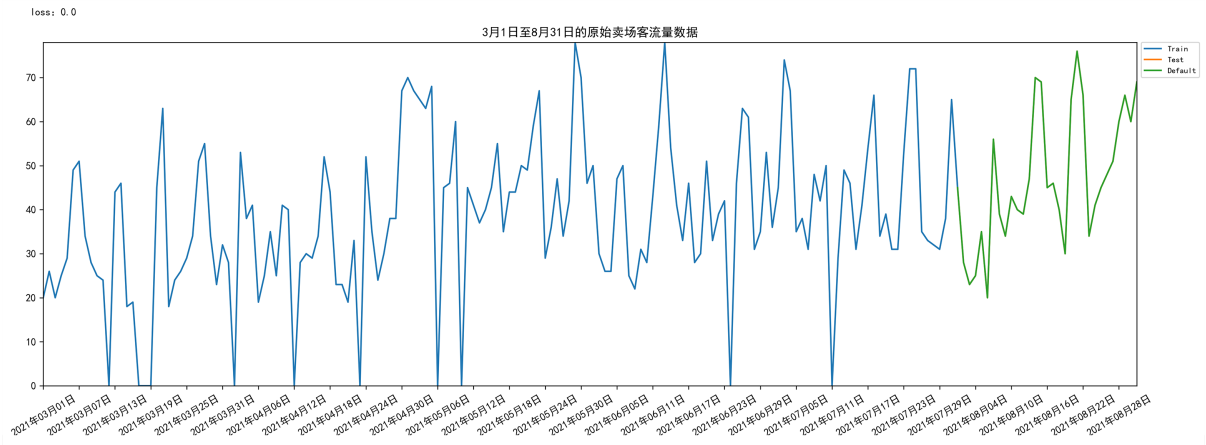

Figure 4. Statistical diagram of original store passenger flow data
图 4. 原始卖场客流量数据统计图

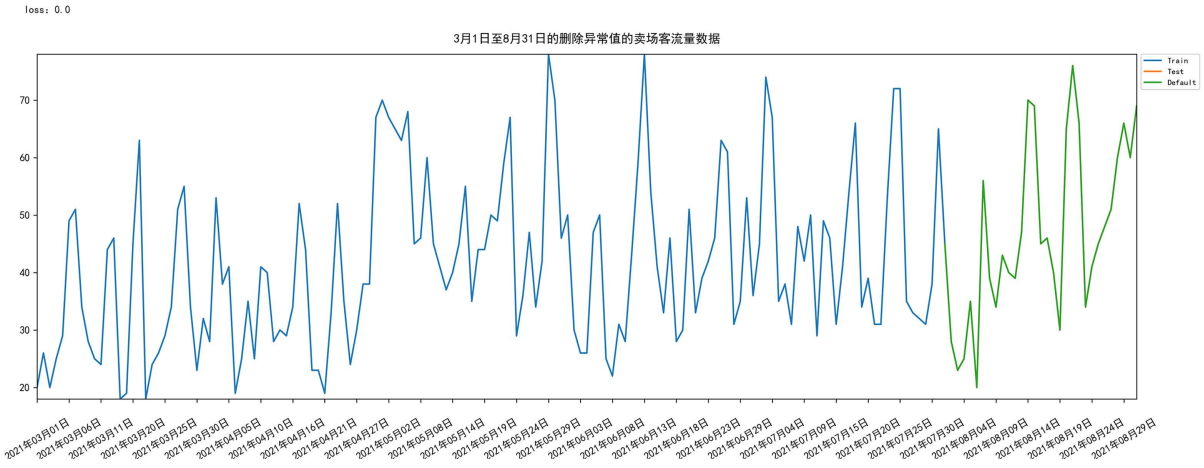

Figure 5. Statistical diagram of store traffic data with outliers deleted
图 5. 删除异常值的卖场客流量数据统计图

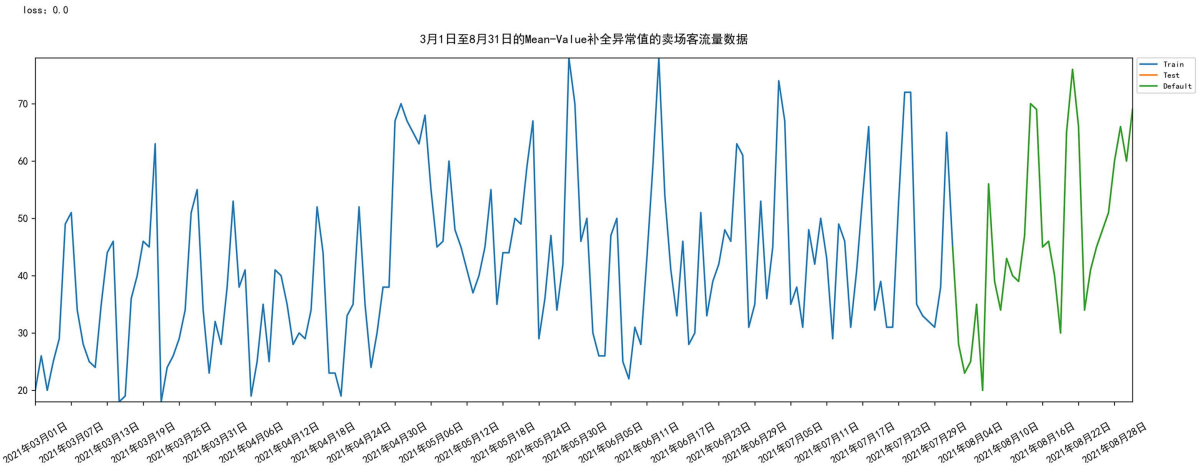

Figure 6. Statistical diagram of customer traffic data of Mean-Value complete outlier
图 6. Mean-Value 补全异常值的卖场客流量数据统计图

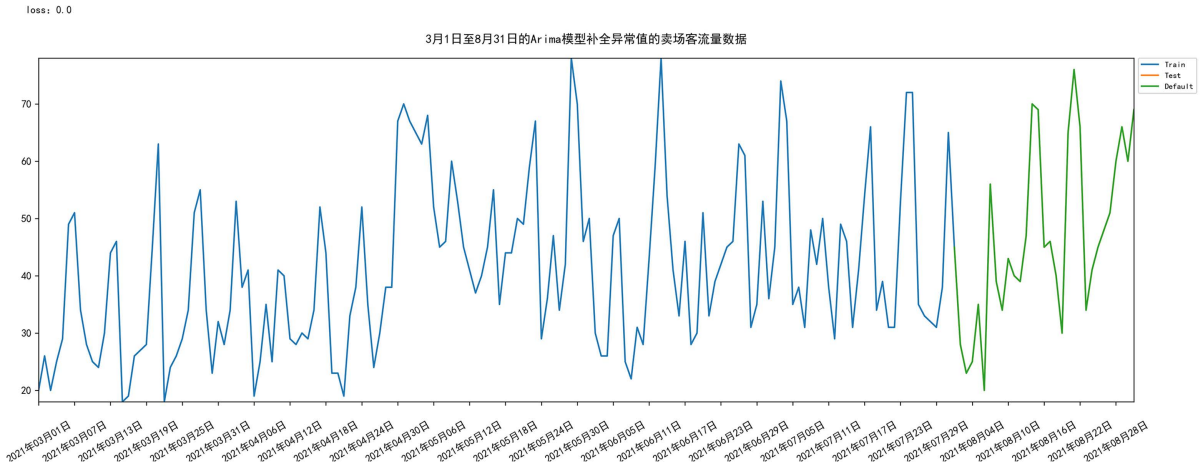

Figure 7. Statistical diagram of shopping mall traffic data with Arima model completing outliers
图 7. Arima 模型补全异常值的卖场客流量数据统计图

4.2. Arima 模型预测

使用 Arima 模型时，必须保证数据序列的平稳性，所以在进行模型训练前，应当首先判断数据序列是否具备平稳性，绘制图像的差分图，如图 8 所示。

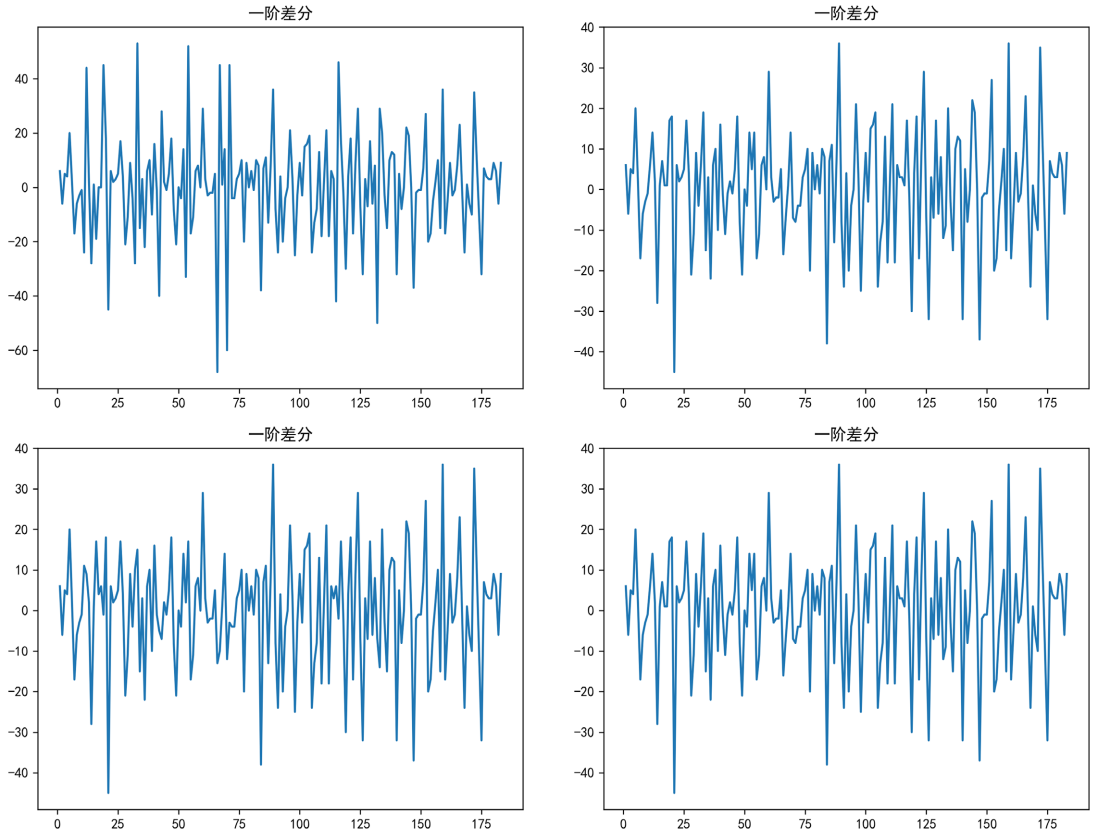

Figure 8. First-order difference graphs of four data completion methods
图 8. 四种数据补全方法的一阶差分图

通过观察分析可知，四份数据均在 1 阶时稳定，此时显著水平 p 值、ADF 的检验结果、显著性水平分别如下表 1 所示：

Table 1. 0-order data ADF verification
表 1. 0 阶数据 ADF 校验

数据补全方式	项目	T 统计量	p 值
无任何操作	ADF 检验统计量	−7.167698e+00	2.857762e−10
	1%显著性水平	−3.469413e+00	
	5%显著性水平	−2.878696e+00	
	10%显著性水平	−2.575917e+00	
删除异常值	ADF 检验统计量	−2.924997	3.819871e−10
	1%显著性水平	−3.470616e+00	
	5%显著性水平	−2.879221e+00	
	10%显著性水平	−2.576197e+00	

Continued

Mean-Value 填充	ADF 检验统计量	−6.320255e+00	3.078648e−08
	1%显著性水平	−3.469413e+00	
	5%显著性水平	−2.878696e+00	
	10%显著性水平	−2.575917e+00	
Arima 填充	ADF 检验统计量	−6.122570e+00	8.783295e−08
	1%显著性水平	−3.469181e+00	
	5%显著性水平	−2.878595e+00	
	10%显著性水平	−2.575863e+00	

4.3. SArima 模型预测

建立 SArima 模型，在上一步中，我们已经确定数据在 1 阶时，具有稳定性，因此 SArima 模型的三个参数(p,d,q)中，可以认定差分阶数 $d = 1$。接下来应当确定 p 和 q 两个参数，我们有多种方法来定阶，常用的方法为绘制 ACF 图和 PACF 图，通过观察 ACF 图与 PACF 图可以大概确定模型的 p，q 值。但是由于这种方法所生成的 p，q 值具有较强的主观性与经验性，所以可以使用 AIC_BIC 热力图的方式，来确认生成的模型参数。

1) 不进行任何数据补全的 SArima 模型

对数据进行预处理后生成的 AIC_BIC 热力图如下图 9 所示。

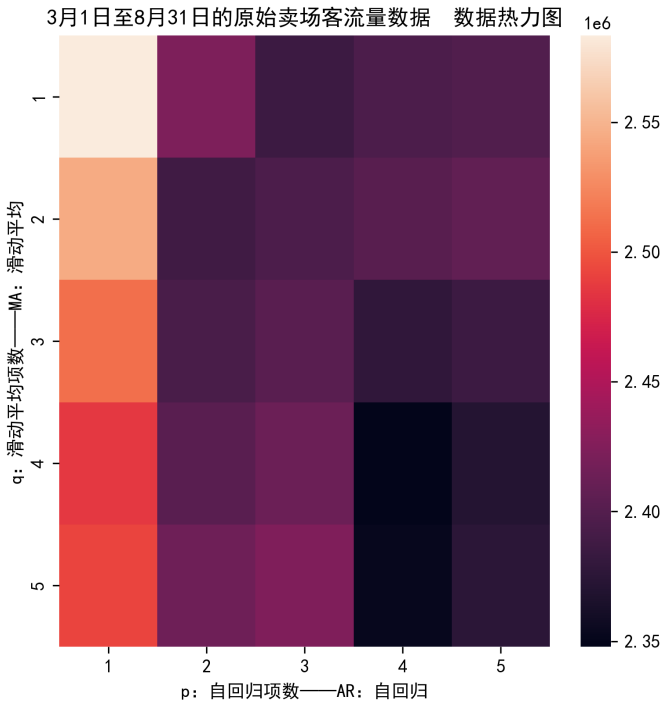

Figure 9. Raw data AIC_BIC heat map without data completion
图 9. 不进行数据补全的原始数据 AIC_BIC 热力图

可以看到，图 9 中深色区域的部分为(p,q)的值分别为(4,4)，在使用网格搜索，对 P、D 进行调优，利用 SArima 模型对客流量进行预测，绘制折线图，如下图 10 所示。

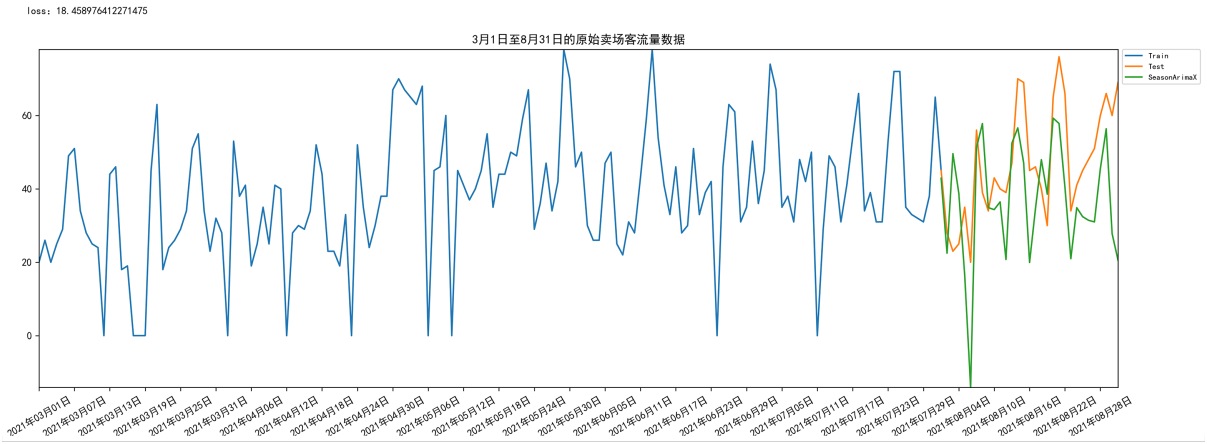

Figure 10. Prediction effect of SArima model
图 10. SArima 模型预测效果图

将预测值与真实值进行求差计算，制作统计表，如下表 2 所示。

Table 2. Error table of SArima model prediction results
表 2. SArima 模型预测结果误差表

日期	真实值	预测值	误差
2021-08-01	45	42.66667	2.333326
2021-08-02	28	22.17409	5.825912
2021-08-03	23	49.14716	−26.1472
2021-08-04	25	38.05967	−13.0597
2021-08-05	35	16.12297	18.87703
2021-08-06	20	−13.9759	33.97587
2021-08-07	56	51.27939	4.72061
2021-08-08	39	57.51167	−18.5117
2021-08-09	34	34.59544	−0.59544
2021-08-10	43	33.84833	9.151675
2021-08-11	40	35.71872	4.281281
2021-08-12	39	20.41649	18.58351
2021-08-13	47	52.64206	−5.64206
2021-08-14	70	56.66623	13.33377
2021-08-15	69	46.76441	22.23559
2021-08-16	45	19.62592	25.37408
2021-08-17	46	34.07842	11.92158
2021-08-18	40	47.19308	−7.19308
2021-08-19	30	38.26693	−8.26693
2021-08-20	65	59.52833	5.47167
2021-08-21	76	57.87853	18.12147
2021-08-22	66	40.09952	25.90048
2021-08-23	34	20.62774	13.37226
2021-08-24	41	34.32539	6.674606
2021-08-25	45	31.64328	13.35672

Continued

2021-08-26	48	31.14862	16.85138
2021-08-27	51	31.08126	19.91874
2021-08-28	60	45.2413	14.7587
2021-08-29	66	55.57972	10.42028
2021-08-30	60	26.70845	33.29155
2021-08-31	69	19.68455	49.31545

2) 删除异常值的 SArima 模型

对数据进行预处理后生成的 AIC_BIC 热力图如下图 11 所示。

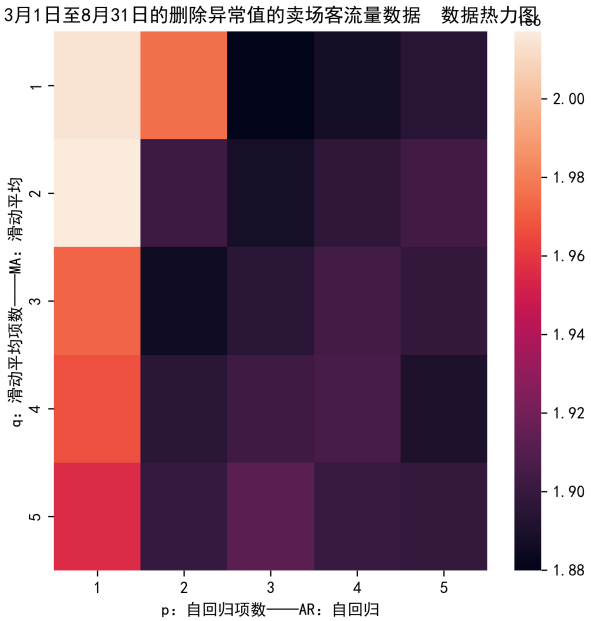

Figure 11. AIC_BIC heat map of data with deleted outliers
图 11. 删除异常值的数据的 AIC_BIC 热力图

利用 Arima 模型对 2021 年的客流量进行预测，绘制折线图，如下图 12 所示。

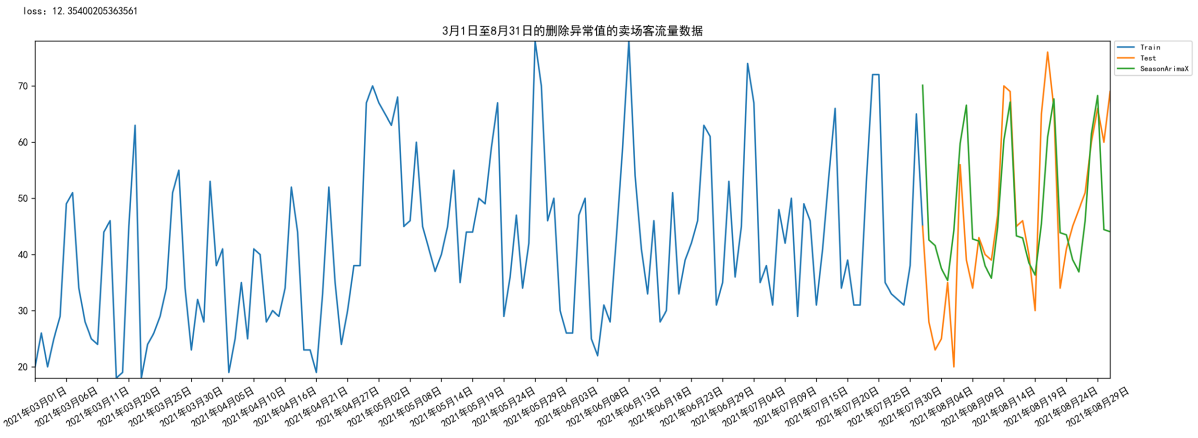

Figure 12. Prediction effect of SArima model
图 12. SArima 模型预测效果图

将预测值与真实值进行求差计算，制作统计表，如下表 3 所示。

Table 3. Error table of SArima model prediction results
表 3. SArima 模型预测结果误差表

日期	真实值	预测值	误差
2021-08-01	45	70.12217	−25.1222
2021-08-02	28	42.5839	−14.5839
2021-08-03	23	41.57066	−18.5707
2021-08-04	25	37.50728	−12.5073
2021-08-05	35	35.38552	−0.38552
2021-08-06	20	44.34471	−24.3447
2021-08-07	56	59.7644	−3.7644
2021-08-08	39	66.57298	−27.573
2021-08-09	34	42.75884	−8.75884
2021-08-10	43	42.37621	0.623785
2021-08-11	40	37.94153	2.058475
2021-08-12	39	35.79196	3.208038
2021-08-13	47	44.87478	2.125218
2021-08-14	70	60.33572	9.664283
2021-08-15	69	67.13576	1.86424
2021-08-16	45	43.31808	1.681924
2021-08-17	46	42.93901	3.06099
2021-08-18	40	38.50631	1.493685
2021-08-19	30	36.35671	−6.35671
2021-08-20	65	45.43935	19.56065
2021-08-21	76	60.90037	15.09963
2021-08-22	66	67.70049	−1.70049
2021-08-23	34	43.88282	−9.88282
2021-08-24	41	43.50375	−2.50375
2021-08-25	45	39.07105	5.928948
2021-08-26	48	36.92145	11.07855
2021-08-27	51	46.00409	4.995909
2021-08-28	60	61.46511	−1.46511
2021-08-29	66	68.26523	−2.26523
2021-08-30	60	44.44756	15.55244
2021-08-31	69	44.06849	24.93151

3) 使用 Mean-Value 填补异常值的 SArima 模型

对数据进行预处理后生成的 AIC_BIC 热力图如下图 13 所示。

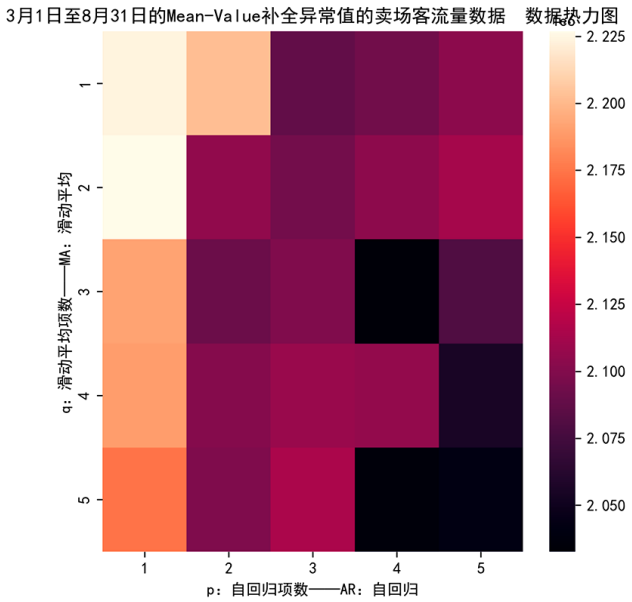

Figure 13. AIC_BIC heat map with Mean-Value completion of outliers
图 13. 使用 Mean-Value 补全异常值的 AIC_BIC 热力图

利用 Arima 模型对 2021 年的客流量进行预测，绘制折线图，如下图 14 所示。

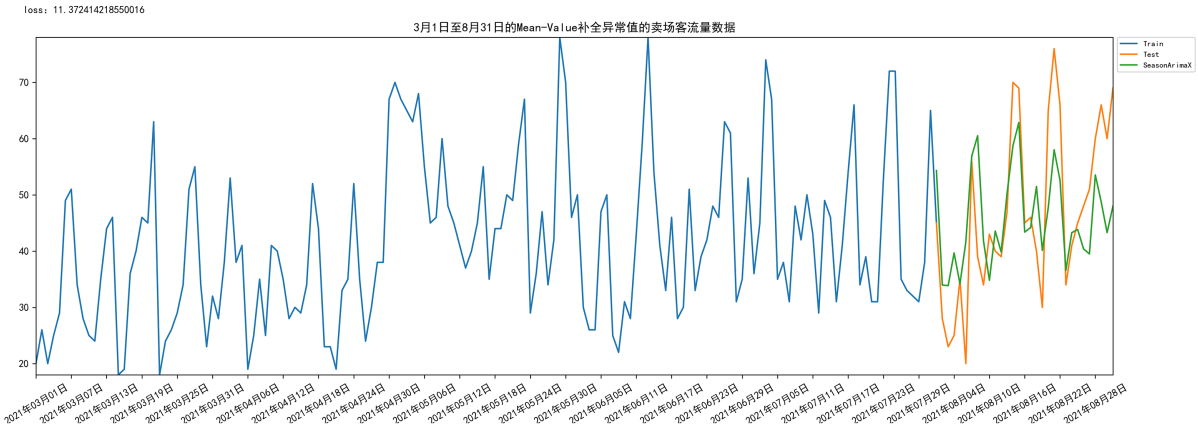

Figure 14. Prediction effect of SArima model
图 14. SArima 模型预测效果图

将预测值与真实值进行求差计算，制作统计表，如下表 4 所示。

Table 4. Error table of SArima model prediction results
表 4. SArima 模型预测结果误差表

日期	真实值	预测值	误差
2021-08-01	45	52.35714	−7.35714
2021-08-02	28	32.24044	−4.24044
2021-08-03	23	30.63942	−7.63942
2021-08-04	25	37.18575	−12.1858
2021-08-05	35	33.91311	1.086891

Continued

2021-08-06	20	43.76604	−23.766
2021-08-07	56	58.41126	−2.41126
2021-08-08	39	58.44797	−19.448
2021-08-09	34	41.05194	−7.05194
2021-08-10	43	34.40468	8.595325
2021-08-11	40	42.37975	−2.37975
2021-08-12	39	41.38634	−2.38634
2021-08-13	47	51.71612	−4.71612
2021-08-14	70	60.64925	9.350748
2021-08-15	69	63.63395	5.366051
2021-08-16	45	43.91264	1.087364
2021-08-17	46	43.21232	2.787677
2021-08-18	40	49.60273	−9.60273
2021-08-19	30	39.07505	−9.07505
2021-08-20	65	47.83876	17.16124
2021-08-21	76	58.86741	17.13259
2021-08-22	66	52.90643	13.09357
2021-08-23	34	35.62791	−1.62791
2021-08-24	41	40.42461	0.575395
2021-08-25	45	40.98411	4.015891
2021-08-26	48	39.35929	8.640707
2021-08-27	51	39.38862	11.61138
2021-08-28	60	55.34035	4.659651
2021-08-29	66	47.08214	18.91786
2021-08-30	60	42.72818	17.27182
2021-08-31	69	47.85268	21.14732

4) 使用 Arima 填补异常值的的 SArima 模型

对数据进行预处理后生成的 AIC_BIC 热力图如下图 15 所示。

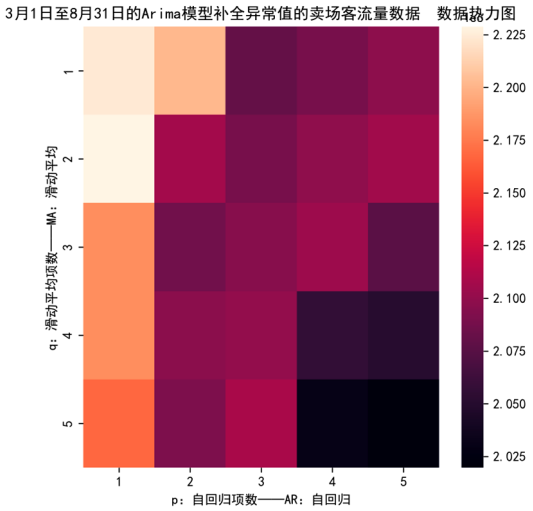

Figure 15. AIC_BIC heat map using Arima model to complete outliers

图 15. 使用 Arima 模型补全异常值的 AIC_BIC 热力图

利用 Arima 模型对 2021 年的客流量进行预测，绘制折线图，如下图 16 所示。

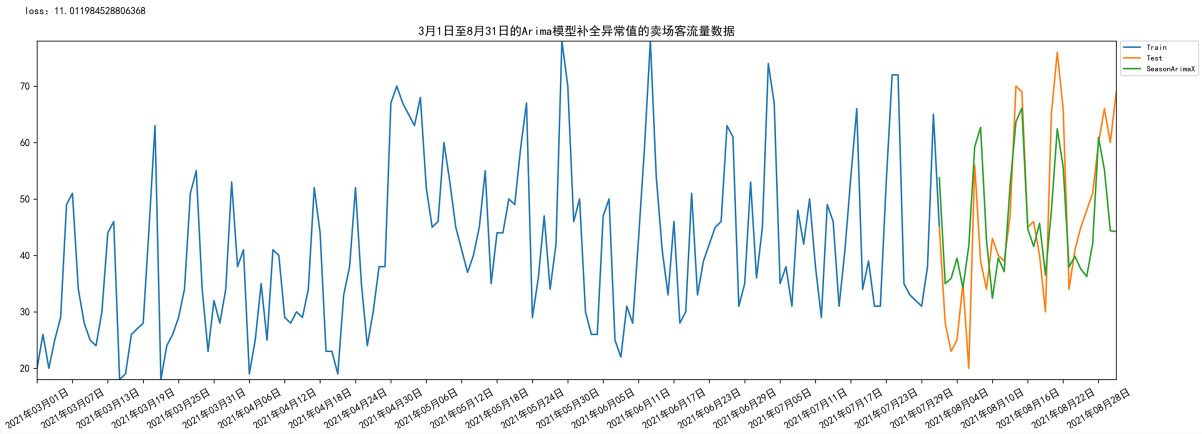

Figure 16. Prediction effect of SArima model
图 16. SArima 模型预测效果图

将预测值与真实值进行求差计算，制作统计表，如下表 5 所示。

Table 5. Error table of SArima model prediction results
表 5. SArima 模型预测结果误差表

日期	真实值	预测值	误差
2021-08-01	45	53.68029	−8.68029
2021-08-02	28	35.00859	−7.00859
2021-08-03	23	36.00158	−13.0016
2021-08-04	25	39.59415	−14.5942
2021-08-05	35	34.27874	0.721265
2021-08-06	20	41.67846	−21.6785
2021-08-07	56	59.1781	−3.1781
2021-08-08	39	62.61238	−23.6124
2021-08-09	34	42.69568	−8.69568
2021-08-10	43	32.50631	10.49369
2021-08-11	40	39.61825	0.381747
2021-08-12	39	37.14491	1.855089
2021-08-13	47	52.31952	−5.31952
2021-08-14	70	63.64477	6.355228
2021-08-15	69	66.05399	2.946007
2021-08-16	45	44.77296	0.227035
2021-08-17	46	41.68511	4.314888
2021-08-18	40	45.74378	−5.74378
2021-08-19	30	36.49591	−6.49591
2021-08-20	65	47.79212	17.20789
2021-08-21	76	62.35169	13.64831
2021-08-22	66	55.63867	10.36133

Continued

2021-08-23	34	38.02437	−4.02437
2021-08-24	41	39.96603	1.033975
2021-08-25	45	37.61004	7.389956
2021-08-26	48	36.23311	11.76689
2021-08-27	51	41.99275	9.007252
2021-08-28	60	60.83719	−0.83719
2021-08-29	66	55.13589	10.86411
2021-08-30	60	44.46564	15.53436
2021-08-31	69	44.35498	24.64502

4.4. Holt-Winters 预测

采用网格调参的方式，找到 Holt-Winters 模型在数据集上的最优解，得到的结果如下所示。

1) 不进行任何数据补全的 Holt-Winters 模型

Holt-Winters 模型生成的预测图如下图 17 所示：

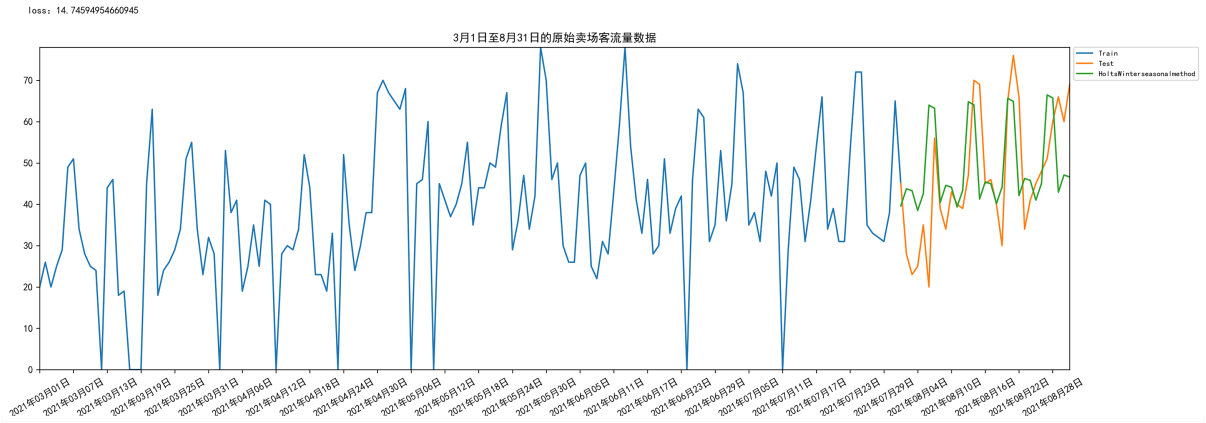

Figure 17. Prediction effect of Holt-Winters model
图 17. Holt-Winters 模型预测效果图

Holt-Winters 模型预测的结果与真实值之间的误差如下表 6 所示：

Table 6. Error table of prediction results of Holt-Winters model
表 6. Holt-Winters 模型预测结果误差表

日期	真实值	预测值	误差
2021/8/1	45	39.55097	5.449026362
2021/8/2	28	43.69106	−15.691065
2021/8/3	23	43.23972	−20.2397249
2021/8/4	25	38.45884	−13.4588445
2021/8/5	35	42.47564	−7.47563684
2021/8/6	20	63.91756	−43.9175592
2021/8/7	56	63.10523	−7.10522991
2021/8/8	39	40.37129	−1.3712868

Continued

2021/8/9	34	44.51138	−10.5113781
2021/8/10	43	44.06004	−1.0600381
2021/8/11	40	39.27916	0.720842369
2021/8/12	39	43.29595	−4.29595001
2021/8/13	47	64.73787	−17.7378724
2021/8/14	70	63.92554	6.074456928
2021/8/15	69	41.1916	27.80840004
2021/8/16	45	45.33169	−0.33169129
2021/8/17	46	44.88035	1.11964874
2021/8/18	40	40.09947	−0.09947079
2021/8/19	30	44.11626	−14.1162632
2021/8/20	65	65.55819	−0.55818554
2021/8/21	76	64.74586	11.25414376
2021/8/22	66	42.01191	23.98808687
2021/8/23	34	46.152	−12.1520044
2021/8/24	41	45.70066	−4.70066442
2021/8/25	45	40.91978	4.080216042
2021/8/26	48	44.93658	3.063423667
2021/8/27	51	66.3785	−15.3784987
2021/8/28	60	65.56617	−5.5661694
2021/8/29	66	42.83223	23.16777371
2021/8/30	60	46.97232	13.02768239
2021/8/31	69	46.52098	22.47902241

2) 删除异常值的 Holt-Winters 模型

Holt-Winters 模型生成的预测图如下图 18 所示：

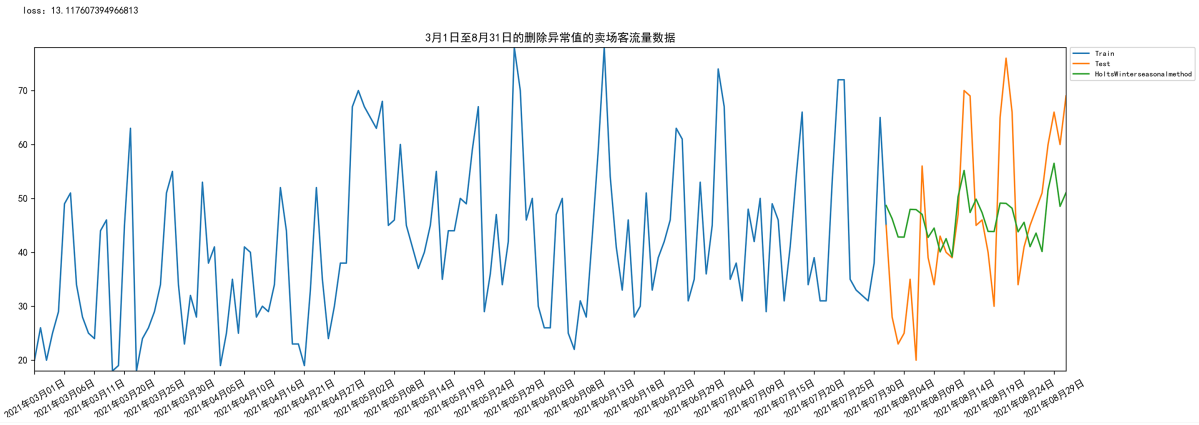

Figure 18. Prediction effect of Holt-Winters model
图 18. Holt-Winters 模型预测效果图

Holt-Winters 模型预测的结果与真实值之间的误差如下表 7 所示：

Table 7. Error table of prediction results of Holt-Winters model
表 7. Holt-Winters 模型预测结果误差表

日期	真实值	预测值	误差
2021/8/1	45	48.68307	−3.68307436
2021/8/2	28	46.23639	−18.2363851
2021/8/3	23	42.84135	−19.841351
2021/8/4	25	42.81987	−17.8198661
2021/8/5	35	47.96467	−12.9646671
2021/8/6	20	47.92323	−27.9232339
2021/8/7	56	47.03899	8.961005427
2021/8/8	39	42.7743	−3.77429891
2021/8/9	34	44.49407	−10.4940733
2021/8/10	43	40.08347	2.91653391
2021/8/11	40	42.53184	−2.53184313
2021/8/12	39	39.2056	−0.20559828
2021/8/13	47	50.35212	−3.35211989
2021/8/14	70	55.17793	14.82207279
2021/8/15	69	47.39051	21.60948579
2021/8/16	45	49.85733	−4.85733181
2021/8/17	46	47.35163	−1.35162729
2021/8/18	40	43.8747	−3.87470353
2021/8/19	30	43.8527	−13.8527003
2021/8/20	65	49.1216	15.87840373
2021/8/21	76	49.07916	26.92083639
2021/8/22	66	48.1736	17.82640391
2021/8/23	34	43.80603	−9.80603407
2021/8/24	41	45.56729	−4.56729015
2021/8/25	45	41.0503	3.949702844
2021/8/26	48	43.55773	4.442269857
2021/8/27	51	40.15125	10.84874522
2021/8/28	60	51.56664	8.43336454
2021/8/29	66	56.50884	9.491156591
2021/8/30	60	48.53359	11.46640546
2021/8/31	69	51.05991	17.94008713

3) 使用 Mean-Value 填补异常值的 Holt-Winters 模型

Holt-Winters 模型生成的预测图如下图 19 所示：

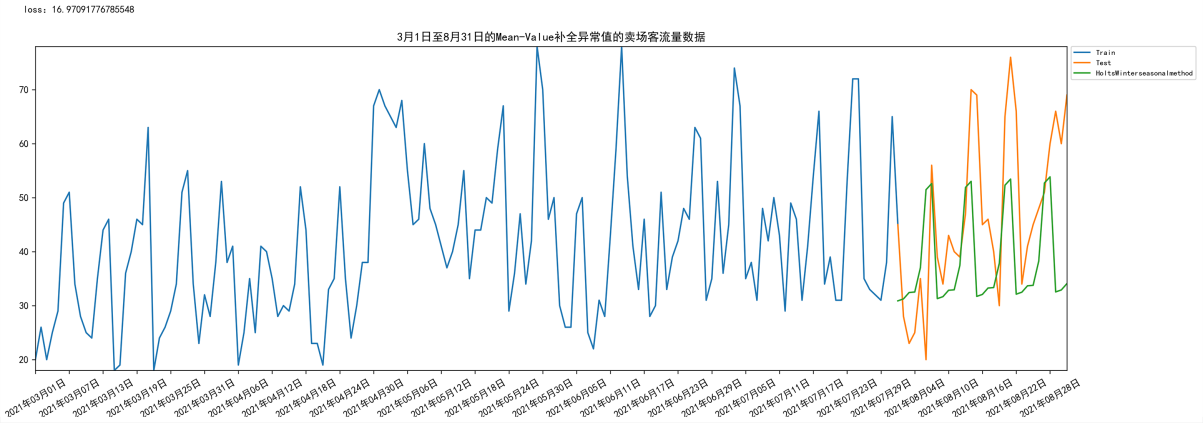

Figure 19. Prediction effect of Holt-Winters model
图 19. Holt-Winters 模型预测效果图

Holt-Winters 模型预测的结果与真实值之间的误差如下表 8 所示：

Table 8. Error table of prediction results of Holt-Winters model
表 8. Holt-Winters 模型预测结果误差表

日期	真实值	预测值	误差
2021/8/1	45	30.88416	14.11584353
2021/8/2	28	31.2464	−3.24640196
2021/8/3	23	32.42923	−9.42923145
2021/8/4	25	32.51837	−7.51836817
2021/8/5	35	37.01588	−2.01588382
2021/8/6	20	51.46988	−31.4698799
2021/8/7	56	52.60782	3.392180474
2021/8/8	39	31.2984	7.701595188
2021/8/9	34	31.66065	2.339349699
2021/8/10	43	32.84348	10.15652021
2021/8/11	40	32.93262	7.067383484
2021/8/12	39	37.43013	1.569867834
2021/8/13	47	51.88413	−4.88412824
2021/8/14	70	53.02207	16.97793213
2021/8/15	69	31.71265	37.28734685
2021/8/16	45	32.0749	12.92510136
2021/8/17	46	33.25773	12.74227187
2021/8/18	40	33.34686	6.653135142
2021/8/19	30	37.84438	−7.84438051
2021/8/20	65	52.29838	12.70162342
2021/8/21	76	53.43632	22.56368379
2021/8/22	66	32.1269	33.8730985
2021/8/23	34	32.48915	1.510853014

Continued

2021/8/24	41	33.67198	7.328023527
2021/8/25	45	33.76111	11.2388868
2021/8/26	48	38.25863	9.741371148
2021/8/27	51	52.71262	−1.71262493
2021/8/28	60	53.85056	6.149435446
2021/8/29	66	32.54115	33.45885016
2021/8/30	60	32.9034	27.09660467
2021/8/31	69	34.08622	34.91377518

4) 使用 Arima 填补异常值的 Holt-Winters 模型

Holt-Winters 模型生成的预测图如下图 20 所示：

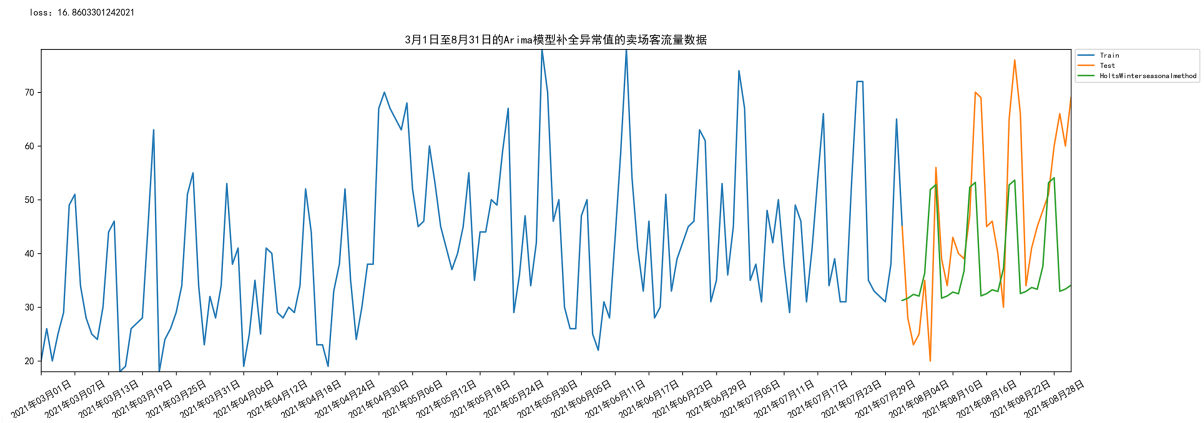

Figure 20. Prediction effect of Holt-Winters model
图 20. Holt-Winters 模型预测效果图

Holt-Winters 模型预测的结果与真实值之间的误差如下表 9 所示：

Table 9. Error table of prediction results of Holt-Winters model
表 9. Holt-Winters 模型预测结果误差表

日期	真实值	预测值	误差
2021/8/1	45	31.24465	13.7553534
2021/8/2	28	31.65328	−3.65328383
2021/8/3	23	32.38079	−9.38079107
2021/8/4	25	32.06355	−7.06354776
2021/8/5	35	36.3355	−1.33550098
2021/8/6	20	51.88061	−31.8806104
2021/8/7	56	52.79045	3.209547311
2021/8/8	39	31.67102	7.328977526
2021/8/9	34	32.07966	1.920340297
2021/8/10	43	32.80717	10.19283305
2021/8/11	40	32.48992	7.510076364

Continued

2021/8/12	39	36.76188	2.238123148
2021/8/13	47	52.30699	−5.30698632
2021/8/14	70	53.21683	16.78317144
2021/8/15	69	32.0974	36.90260165
2021/8/16	45	32.50604	12.49396442
2021/8/17	46	33.23354	12.76645718
2021/8/18	40	32.9163	7.08370049
2021/8/19	30	37.18825	−7.18825273
2021/8/20	65	52.73336	12.26663781
2021/8/21	76	53.6432	22.35679556
2021/8/22	66	32.52377	33.47622578
2021/8/23	34	32.93241	1.067588549
2021/8/24	41	33.65992	7.340081302
2021/8/25	45	33.34268	11.65732462
2021/8/26	48	37.61463	10.3853714
2021/8/27	51	53.15974	−2.15973806
2021/8/28	60	54.06958	5.930419688
2021/8/29	66	32.95015	33.0498499
2021/8/30	60	33.35879	26.64121267
2021/8/31	69	34.08629	34.91370543

4.5. 对比分析

对基于多种数据补全策略的 SArima 模型、Holt-Winters 模型训练效果进行对比分析，使用常用的误差损失函数来判断模型的效果。

1) 均方差误差 MSR 值越小，预测值与真实值的误差越小。

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

2) 均方根误差 RMSE 值越小，预测值与真实值的误差越小。

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

3) 平均绝对误差 MAE 值越小，预测值与真实值的误差越小。

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

4) 平均绝对百分比误差 MAPE 值越小，预测值与真实值的误差越小。

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

公式中 n 代表样本个数，$\hat{Y}_i$ 为预测值，Y_i 为真实值。

在实验结束中，我们将分别对应多种数据补全策略情况下分别生成 Holt-Winters 模型、SArima 模型，并使用 2021 年 8 月 1 日至 2021 年 8 月 31 日的数据集作为测试集，使用 MSE、RMSE、MAE 和 MAPE 值进行统计，制表 10 如下。

Table 10. Results of evaluation indexes of the three models
表 10. 三种模型评价指标结果表

模型名称 补充策略	SArima 模型				Holt-Winters 模型			
	MSE	RMSE	MAE	MAPE	MSE	RMSE	MAE	MAPE
不进行任何数据补全	340.7338	18.4589	15.2209	0.3658	217.4430	14.7459	10.9188	0.2891
删除异常值	323.3502	17.9819	14.2314	0.3411	**172.0705**	**13.1175**	**10.7949**	**0.2730**
使用 Mean-Value 填补异常值	129.3318	11.3724	9.2566	0.2225	288.0120	16.9709	12.9562	0.2810
使用 Arima 填补异常值	**121.2638**	**11.0119**	**8.7760**	**0.2178**	284.2707	16.8603	12.8767	0.2796

根据表中数据，我们可以发现，在进行了数据补全后，SArima 模型效果在各个指标上均优于 Holt-Winters 模型，而未进行数据补全时，Holt-Winters 模型具有较低的损失，其原因是 SArima 模型不单单考虑了数据之间的线性关系，还计算了 N 天前的数据，因此如果数据中存在较多的干扰异常值时，SArima 模型效果会下降，即便删除了这些异常值，SArima 模型在拟合时，由于周期性改变，因此也无法继续进行正确的预测。在使用了 Mean_Value 与 Arima 模型进行异常值填充后，SArima 模型的预测效果得到了大幅度的提升，Holt-Winters 模型在进行填充后反而错误率提高，原因时因为 Holt-Winters 模型只考虑普通线性关系，当预测数量越多，误差也就越大。对比两种异常值填充的方法，我们发现使用 Arima 模型进行的数据填充在 SArima 模型和 Holt-Winters 模型中，都有一定程度的提升。

总体而言，在进行时间序列预测时，我们应在数据清洗完成后，对异常值进行处理，可以极大的提升模型预测效果。并且在进行异常值填充时可以考虑使用 Arima 模型作为填充策略，在一定程度上，可以增强模型的预测效果。

5. 结语

本文研究了多种数据补全策略对商超客流量预测的影响，分别使用不进行数据补全、删除异常值、基于 Mean-value、基于 Arima 模型的数据补全的 SArima、Holt-Winters 模型建立了客流量预测模型。通过多个评价指标的对比，表明若数据只进行简单补全操作时，Holt-Winters 模型的效果较好，而使用复杂的补全方式后 SArima 模型的效果较好，并有了较大的提升，优化效果达到了 Holt-Winters 模型的两倍以上；并且对比两次实验，发现 SArima 模型会受到不同补全策略的干扰，但是，当使用一定的补全策略后，将达到一个较好预期的结果。所以得到结论，不同的补全策略将极大影响 Sarima 模型的稳定性，且使用 Arima 模型进行数据补全将提升模型的准确程度。

由于现有的模型都只是考虑了特征之间的线性关系，并没有考虑到其他非线性特征对模型预测的影响，例如当天的天气、是否为节假日、是否有促销活动等。未来可以通过特征工程的方法，将这些特征添加到预测中去，或使用大语言模型，先对各项特征进行向量化，将向量特征添加到预测中去，来提高模型预测的精度。

总之，本研究为客流量预测的数据补全提供了一些参考，但仍需要进一步改进和完善，以更好地应

对实际应用需求。

基金项目

2020 年襄阳市农业领域重点科技创新计划项目“基于多源异构大数据平台的精准农业智能化服务系统的研发与应用”(项目编号：2020ABA002240)。

参考文献

[1] 赵雷. GA-SVM 研究及在航空客流量预测中的应用[D]: [硕士学位论文]. 昆明: 云南大学, 2012.

[2] Siami-Namini, S., Tavakoli, N. and Namin, A.S. (2018) A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*), Orlando, 17-20 December 2018, 1394-1401. https://doi.org/10.1109/ICMLA.2018.00227

[3] 洪嘉灏. 基于 Holt-Winters 模型的股价趋势预测研究[D]: [硕士学位论文]. 广州: 暨南大学, 2017.

[4] 王莹, 韩宝明, 张琦, 李得伟. 基于 SARIMA 模型的北京地铁进站客流量预测[J]. 交通运输系统工程与信息, 2015, 15(6): 205-211. https://doi.org/10.16097/j.cnki.1009-6744.2015.06.031

[5] 邢金城, 孙家丽, 凌继红, 等. 天津市西青区某大型超市客流量等的调查研究[C]//中国建筑学会暖通空调分会, 中国制冷学会空调热泵专业委员会. 全国暖通空调制冷 2008 年学术年会资料集. 重庆: 暖通空调, 2008: 106.

[6] 赵加宁, 武丽霞, 王昭俊, 方修睦. 大型超市客流量的调查与分析[J]. 暖通空调, 2004(6): 53-56.

[7] 王旭天. 基于 BP 神经网络的我国汽车销量预测分析[D]: [硕士学位论文]. 上海: 东华大学, 2016.

[8] 姜春雷, 张树清, 张策, 李华朋, 丁小辉. 基于 SARIMA-BP 神经网络组合方法的 MODIS 叶面积指数时间序列建模与预测[J]. 光谱学与光谱分析, 2017, 37(1): 189-193.

[9] 毛远宏, 孙琛琛, 徐鲁豫, 刘曦, 柴波, 贺鹏超. 基于深度学习的时间序列预测方法综述[J/OL]. 微电子学与计算机, 2023(4): 8-17. https://doi.org/10.19304/J.ISSN1000-7180.2022.0725

[10] 李海林, 张丽萍. 时间序列数据挖掘中的聚类研究综述[J]. 电子科技大学学报, 2022, 51(3): 416-424.

[11] 孙湘海, 刘潭秋. 基于神经网络和 SARIMA 组合模型的短期交通流预测[J]. 交通运输系统工程与信息, 2008(5): 32-37. https://doi.org/10.16097/j.cnki.1009-6744.2008.05.004

[12] 赵喜仓, 周作杰. 基于 SARIMA 模型的我国季度 GDP 时间序列分析与预测[J]. 统计与决策, 2010(22): 18-20. https://doi.org/10.13546/j.cnki.tjyjc.2010.22.023

[13] Siami-Namini, S. and Namin, A.S. (2018) Forecasting Economics and Financial Time Series: ARIMA vs. LSTM.

[14] Fattah, J., Ezzine, L., Aman, Z., *et al*. (2018) Forecasting of Demand Using ARIMA Model. *International Journal of Engineering Business Management*, **10**, 3-7. https://doi.org/10.1177/1847979018808673

[15] 何雪晴. 基于网络搜索数据和降噪处理的旅游客流量预测研究[D]: [硕士学位论文]. 大连: 东北财经大学, 2019. https://doi.org/10.27006/d.cnki.gdbcu.2019.000084

[16] 段然. 基于 SARIMA 模型的铁路站点客流量预测[D]: [硕士学位论文]. 柳州: 广西科技大学, 2017. https://doi.org/10.27759/d.cnki.ggxgx.2017.000166

[17] 苏耘. 基于深度学习的时间序列分类方法综述[J]. 电子技术与软件工程, 2022(14): 259-262.

[18] 邱敦国, 杨红雨. 一种基于双周期时间序列的短时交通流预测算法[J]. 四川大学学报(工程科学版), 2013, 5(5): 64-68. https://doi.org/10.15961/j.jsuese.2013.05.001

[19] 方昇越. 基于 XGBoost 的地铁短时客流量预测研究[D]: [硕士学位论文]. 大连: 大连海事大学, 2022. https://doi.org/10.26989/d.cnki.gdlhu.2022.001171

[20] 鄢仕林. 中小机场航线客流量预测问题研究[D]: [硕士学位论文]. 德阳: 中国民用航空飞行学院, 2022. https://doi.org/10.27722/d.cnki.gzgmh.2022.000152

[21] 马毅. 基于改进粒子群算法的 BP 神经网络研究及旅游客流量预测[D]: [硕士学位论文]. 荆州: 长江大学, 2022. https://doi.org/10.26981/d.cnki.gjhsc.2022.000780

[22] 冒志恒. 基于改进神经网络的交通短时客流量预测研究[D]: [硕士学位论文]. 镇江: 江苏大学, 2021. https://doi.org/10.27170/d.cnki.gjsuu.2021.000425

[23] 蔡纯. 基于多源数据的游乐园运营分析研究[D]: [硕士学位论文]. 成都: 西南财经大学, 2019. https://doi.org/10.27412/d.cnki.gxncu.2019.001934

[24] 成翔. 基于机器学习的商铺客流量预测[D]: [硕士学位论文]. 成都: 电子科技大学, 2018.

[25] 熊笑. 基于梯度上升决策回归树的网约出租车需求动态预测[D]: [硕士学位论文]. 武汉: 华中科技大学, 2017.

[26] 邓雨菲. ARIMA-ATT-LSTM 在旅游客流量预测中的应用研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2022. https://doi.org/10.26991/d.cnki.gdllu.2022.000310

[27] 常昊. 基于 LSTM 神经网络的地铁短时客流量预测研究[D]: [硕士学位论文]. 西安: 西京学院, 2022. https://doi.org/10.27831/d.cnki.gxjxy.2022.000059

[28] 殷志敏. 基于 LightGBM 和 LSTM 模型的地铁客流量短期预测[D]: [硕士学位论文]. 武汉: 华中科技大学, 2020. https://doi.org/10.27157/d.cnki.ghzku.2020.001134

[29] 吕高帆. 超市客流预测模型的研究与应用[D]: [硕士学位论文]. 北京: 北京工业大学, 2019. https://doi.org/10.26935/d.cnki.gbjgu.2019.000381

[30] 张璐. 基于 GA-LSSVM 与 ARIMA 组合的短期风功率预测[D]: [硕士学位论文]. 包头: 内蒙古科技大学, 2022. https://doi.org/10.27724/d.cnki.gnmgk.2022.000809

[31] 龙会典, 严广乐. 基于 SARIMA、GM(1,1)和 BP 神经网络集成模型的 GDP 时间序列预测研究[J]. 数理统计与管理, 2013, 32(5): 814-822. https://doi.org/10.13860/j.cnki.sltj.2013.05.004