

[货物量预测]

基于ARIMA时间序列模型 解决分拣中心货物量预测问题

文/李锐天
青海大学能源与电气工程学院

得益于科技的蓬勃发展，人们可以使用各网络平台足不出户地进行网络购物，在网络购物逐渐占据主题的消费模式下，越来越多的货物需经过遥远的运输送达消费者手中。分拣中心作为物流环节上必不可少的环节承担着巨大压力，如若能对未来不同时间到达分拣中心的货物数量进行预测，必将能更快地处理货物中转环节，也能减轻管理者的管理负担。因此，提出一种预测模型，以此解决物流网络分拣中心的首要问题。

问题重述

互联网科技高速发展，如今人们更倾向于使用智能手机来进行网上购物，为使商品从厂家到消费者手中更方便准确与快捷，货物途中所经过的中转站极为重要，而分拣中心正是作为电商物流网络中的重要核心，其承担着货物运输枢纽与分类分拣等重要责任。它将各地中转过来的货物进行记录整理，随后再发往下一目的地，让物流更清晰地展现在消费者手中。合理地规划与排班工作能使得一个分拣中心的效率得到极大提高，在这之前就需对分拣中心的货物进行预测以便管理者预知下一时刻的货物量，做出更好的人员调动规划及充分发挥工作人员的工作效率。

由于货物受天气、节假日等各不确定因素影响，货物到达分拣中心的数目并不能体现出其具有简单的规律性，要想预测货物在具体时间到达分拣中心的具体数量是一大难点。基于题目所提供的数据训练出合适的模型，能基本符合要求对未来货物量进行大致准确的预测，同时为后续人员调动排班问题提供数据支撑。

问题说明

本文构建货量预测模型，对 57 个分拣中心从

2023-12-1到2023-12-30共计30天的货物进行预测。

模型假设

为简化研究问题，在理想情况下做出如下假设：

①假设各分拣中心均能正常运作；②假设货物均不存在退回分拣中心的情况；③假设货物均能及时运转，在下一时间节点不存在货物滞留；④假设大部分时间货物不存在极大增长或减少情况；⑤假设建模数据真实无误。

解决流程

求解流程如图1所示。

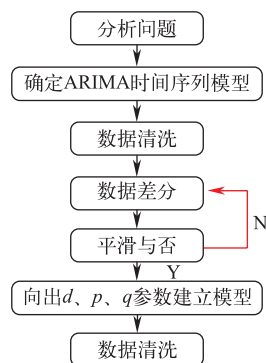


图1 求解流程图

问题要求对57个分拣中心进行未来货量预测，在训练模型中考虑到若将如此多的分拣中心数据输入训练，因每个分拣中心会受到不同误差因素影响，最后必将因误差因数过多而得到一个误差较大的模型，这显然不符合求解原则。所以此处选取SC1分拣中心以作代表，先进行数据清洗，随后采用ARIMA时间序列模型进行求解，最后将其引申至其他所有分拣中心求得结果。

数据清洗

1. 数据说明

数据包含的信息有“分拣中心”“日期”和“货量”三种。经过分析，总共包含了57个分拣中心从2023-8-1到2023-11-30共计122天的货量信息，且数据连续完整无缺失部分。

2. 异常值的检验及处理

本文采用箱线图的方法以识别数据中的异常值，即数据经过处理后绘制对应的箱线图，落在图中上下限以外的点即为异常值点，如图2所示。

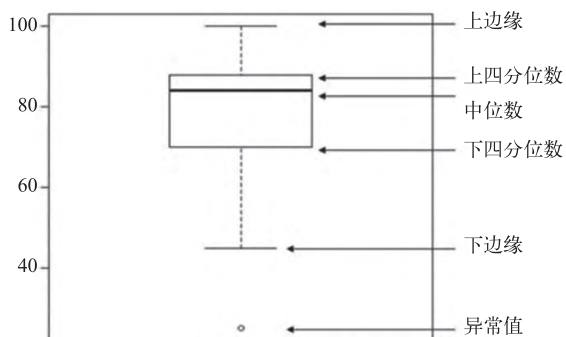


图2 箱线图原理图

将中抽样的SC1分拣中心数据绘制箱线图后，得到图3。

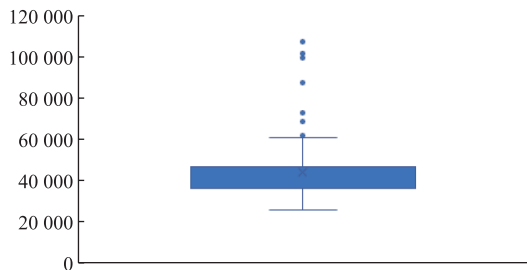


图3 原始数据箱线图

可以观测出图3中存在着超出上下限的值，这代表着SC1分拣中心记录的数据中存在异常值。观察

异常值的分布可以发现，这些异常点大都分布于11月11日这一节点附近，由此可以推测某几天货物量突然大幅增长超过正常值，是由于双十一购物潮所引起的。这种情况并非是数据有误的情况，但是为了减少后续数据计算的误差，此处选择合理删除此类异常值，同时保证数据的连续性，以防对后续预测产生影响。

建立ARIMA时间序列模型进行预测

1. ARIMA时间序列模型

ARIMA时间序列模型，全称自回归差分移动平均模型，是一种广泛应用的单变量时间序列数据预测方法。该模型包含三个关键部分：自回归（AR）、差分（I）和移动平均（MA）。

自回归（AR）部分主要体现当前数据与过去数据之间的关系，即使用时间序列数据的历史数据来预测未来值。差分（I）部分的处理可以使非平稳时间序列数据变得更平稳，以便收集其存在的时序规律。因为ARIMA模型要求时间序列数据必须是平稳的，如此才能在这基础上进行模型的建立。移动平均（MA）部分则考虑过去的预测误差对未来预测值的影响。

在建模过程中，先要进行平稳性检验，确保时间序列数据满足平稳性要求，在此期间根据使数据平稳而进行的差分运算的阶次来获取 d 的值。随后根据自相关系数和偏自相关系数确定回归系数 p 和滑动平均项数 q 。最后根据以上步骤求得的 d 、 p 、 q 拟合建立ARIMA模型。

2. ARIMA时间序列模型的建立

原始数据图如图4所示。

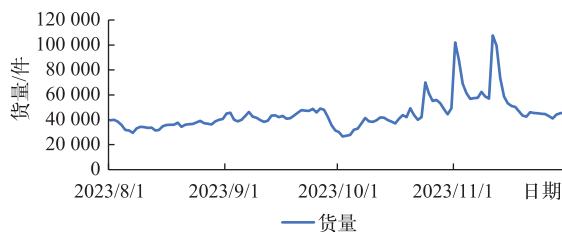


图4 原始数据图

由图4得知，SC1分拣中心原始数据存在着数据量大、数据多段跳跃和折线不平滑等缺点，这不符合上述提到的数据必须尽量平稳这一条件，因此不符合直接建立ARIMA模型的要求，需要对原始数

据作进一步处理。

经过简单的数据清洗后得到图5，对比图4明显能得出结论，经过初步的数据处理后得到的图像更平滑，没有明显的大幅上下跳跃折线出，又能反映原始数据的基本特征及趋势，但这依旧未能符合模型建立的要求，接下来应对数据进行差分处理以继续使数据平稳化。一阶差分序列图如图6所示。

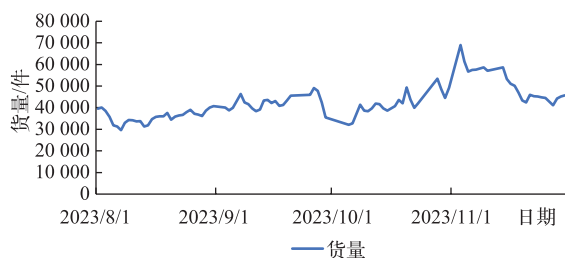


图5 初步处理数据图

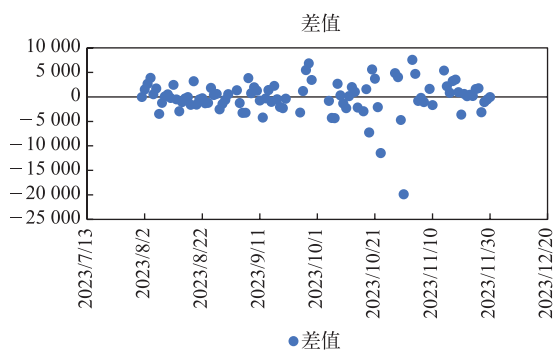


图6 一阶差分序列图

由图6分析其散点分布情况得出，数据经过一阶差分处理后已基本趋于稳定，因此直接可以确定模型中的 d 值取值为1，再用plot_acf与plot_pacf绘制其自相关图与偏自相关图，如图7所示。

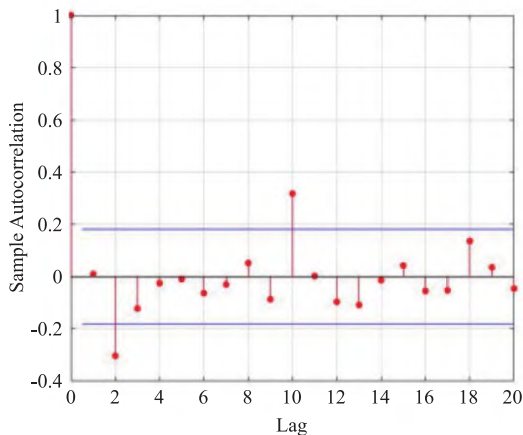


图7 ACF图

观察自相关图与偏自相关图可以发现，经过1阶差分后的数据基本都落在 $-0.2 \sim 0.2$ 的置信区间内，即可得回归系数 $p=1$ 和滑动平均项数 $q=1$ 。

用上文所求得的三个参数 d 、 p 、 q 建立ARIMA时间序列模型：

```
model= ARIMA(train_data, order=(1,1,1),  
freq=' CARGO' )
```

随后将历史数据导入预测模型即可求得SC1分拣中心在未来30天的货物量预测值。模型预测结果如图8所示。

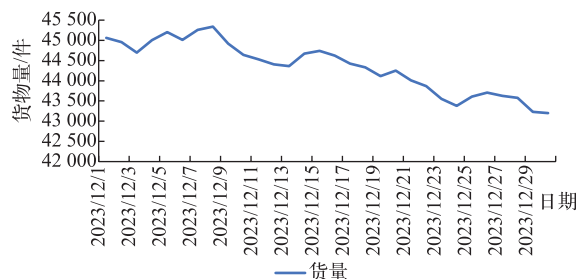


图8 模型预测结果

对模型预测的结果进行分析，可以发现从12/1到12/30未来30天内货物量的涨跌情况较为稳定，折线基本平滑，没有出现大幅度的跳跃且有一个基本稳定的趋势。可以得出结论，此模型预测的结果应该还是比较贴近真实值，数据结果不失真且稳定性高，预测模型有效。

结束语

本文对具体问题进行分析后决定采用ARIMA时间序列模型进行求解，ARIMA时间序列模型的组成较为简单，对复杂非线性预测问题的求解有广泛的应用。为了使用该模型，应先对数据进行平滑性处理，随后根据处理的数据分别求出模型中的 d 、 p 、 q 三个参数，以此建立合理的模型，即可求出题中一个分拣中心预测值。进而将上述方法扩展至其他所有分拣中心，即可分别求出所有的预测值。EA

（收稿日期：2024-08-14）

参考文献

- [1] 李玲玲，辛浩．时序模型ARIMA在数据分析中的应用[J]．福建电脑，2024，40(04)：25-29．