

# Arima 模型数据分析

by 赵嘉欣

## 一、 Arima 模型简述

ARIMA 模型全称为自回归积分滑动平均模型，主要用于时间序列预测。其主要思想为，将序列转化为平稳序列后，因变量的值由其滞后值与随机误差的现值和滞后值表示。

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

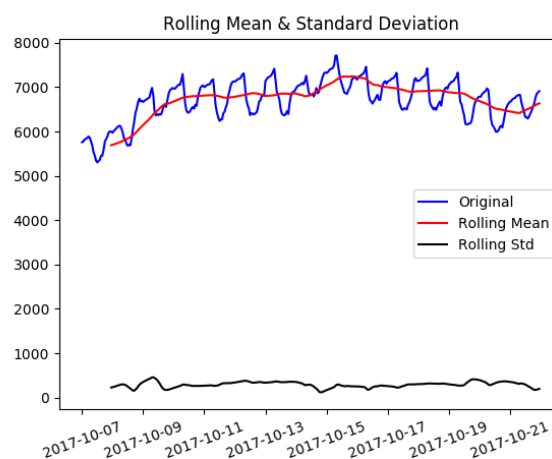
其中，L 为滞后算子， $\phi$ 、 $\theta$  为待求系数。p 为自回归阶数、q 为滑动平均阶数、d 为差分阶数。

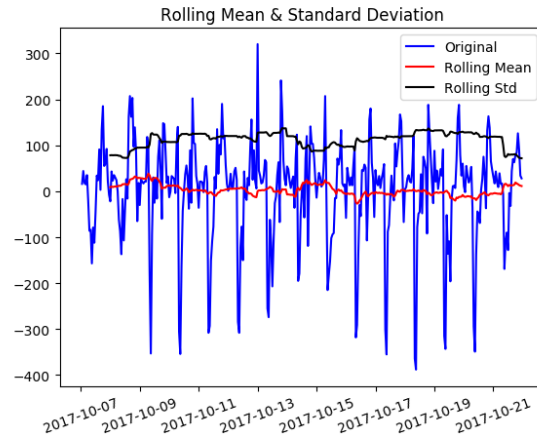
## 二、 拟合过程

下面的分析是，对于#840 号基站网络为中心的 9 个网格的每小时驻留人数的时序序列进行的。

### 1. 检测序列是否平稳

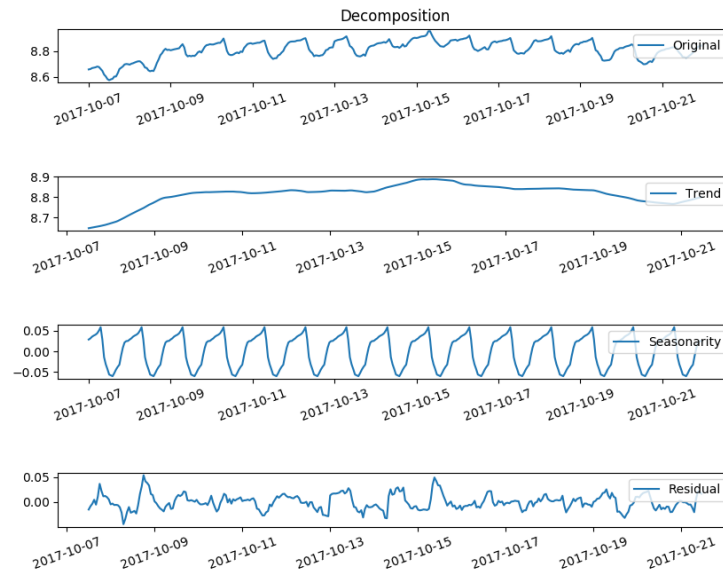
取两周的数据，观察序列的滑动平均与滑动方差（窗口长度为 12），可以看到均值在较大范围内变化。进一步对序列差分，均值在很小范围内波动，基本可以视为平稳序列。



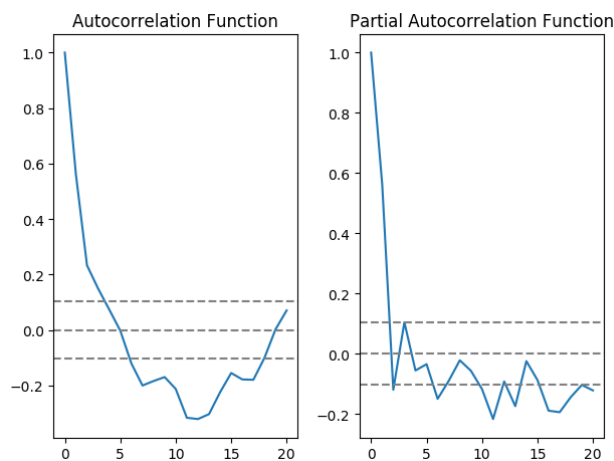


## 2. 对序列进行周期性和趋势性分解

可以看到数据有长期的变化趋势（Trend）和很明显的周期性（Seasonarity）变动，残差项接近白噪声。



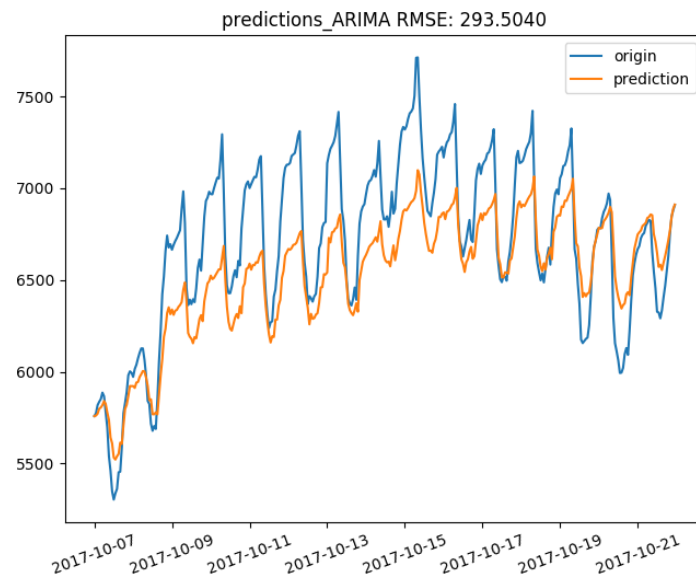
## 3. 通过自相关和偏相关函数确定模型阶数



自相关函数第一次降低到误差线以下可以确定 MA 阶数  $q=4$ 。偏相关函数第一次降低到误差线以下可以确定 AR 阶数  $p=2$ 。

### 三、 结果

最后得到模型并对数据进行预测，与原始序列对比。预测结果趋势一定程度与真值相近，但是峰值上存在较大误差。



结论是，利用 Arima 模型对本项目的数据进行建模可以预测趋势但是存在一定误差。可能的原因是序列的相关性是跳跃的，即某时刻的数据可能与 1 小时前、2 两小时前、24 小时前、48 小时前、一周前的数据等相关性更高，而不是与之前的数据随时间间隔增大相关性逐渐降低。