

ARIMA

2022-05-04

1. 基本概念
2. ARIMA简介
3. ARIMA建模
 - 3.1. ARIMA模型分解
 - 3.2. ARIMA模型参数
 - 3.3. 模型参数定阶
 - 3.4. AR模型
 - 3.5. MA模型
 - 3.6. AR与MA模型对比
 - 3.7. 建模过程
4. 参考文档

1. 基本概念

1. 拖尾：指的是并不存在某一阶突然跳变到0而是逐渐衰减为0。
2. 截尾：指的是从某阶开始均为（接近）0的性质，例如，5阶结尾，即可以用前5个数可以预测下一个数。
3. 自相关（ACF）：衡量的是序列自身在不同时刻随机变量的相关性。
4. 偏自相关（PACF）：剔除了两时刻之间其他随机变量的干扰，是更加纯粹的相关。
5. 严平稳：数据的分布不随时间的改变而改变。
6. 弱平稳：期望与相关系数（依赖性）不变，未来时刻的值依赖过去的信息。

2. ARIMA简介

ARIMA (Auto Regressive Integrated Moving Average) 全称为差分自回归异动平均模型，该模型基于历史时序数据对未来数据进行预测。ARIMA可以对任何存在一定规律的非季节性时间序列进行预测，如果时间序列存在季节性，则需要使用SARIMA (Seasonal ARIMA) 进行预测。

3. ARIMA建模

3.1. ARIMA模型分解

ARIMA模型可以分解为：AR + I + MA

1. AR：自回归模型，以过去的观察值来做线性组合。

2. I：差分，非平稳数据可以通过差分得到平稳性，但是，过度差分会使时序数据失去自相关性，从而失去使用AR模型的条件。
3. MA：滑动平均模型，以过去的残差项也就是白噪声来做线性组合，MA的出发点是通过组合残差项来观察残差的振动。

3.2. ARIMA模型参数

ARIMA模型有三个参数：p、d、q

1. p：AR模型的阶数，表示y的当前值和前p个历史值有关。
2. d：使序列平稳的最小差分阶数，一般是1阶。
3. q：MA模型的阶数。

3.3. 模型参数定阶

1. d：差分阶数
 - 如果时序数据本身是平稳的，则不需要差分，此时d=0。
 - 如果时序数据不平稳，需要结合时序数据的acf图。如果acf表现为10阶或以上的拖尾，则需要进一步差分；如果acf表现为1阶截尾，则可能差分过度。最好的差分阶数是使acf先拖尾几阶，然后截尾。
 - 如果在2个阶数的acf表现类似，则选择标准差较小的序列。
2. p：AR模型参数
 - 通过pacf图确定，因为AR各项的系数就代表了各项自变量x对因变量y的偏自相关性。
3. q：MA模型参数
 - 通过acf图来设定，因为MA是预测误差，预测误差是自回归预测和真实值之间的偏差。定阶过程类似AR阶数的设定过程。

3.4. AR模型

模型作用：描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测。

数据要求：必须满足平稳性的要求，且自相关系数大于0.5。

p阶自回归过程的公式如下：

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t$$

(1)

其中， y_t 是当前值
 μ 是常数项
 p 是阶数
 γ_i 是自相关系数
 ϵ_t 是误差

3.5. MA模型

模型作用：关注的是自回归模型中误差项的累积，能够有效地消除预测中的随机波动。

q阶移动平均模型的公式如下：

$$y_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2)$$

3.6. AR与MA模型对比

AR模型：自相关系数是拖尾，偏自相关系数是截尾

MA模型：自相关系数是截尾，偏自相关系数是拖尾

自相关系数截尾性和偏自相关系数的拖尾性是MA模型重要的识别依据。例如，当我们画出自相关图和偏自相关图，发现自相关图（ACF图）的自相关系数在某个步长后等于0（截尾），偏自相关图（PACF图）的自相关系数接近但不等于0（拖尾），那么我们就可以选择MA模型来拟合这组时间序列数据。

3.7. 建模过程

使用ARIMA(p, d, q)建模的步骤：

1. 白噪声检查：使用LB检查时序数据是否是白噪声。
2. 平稳性检查：使用ADF检查时序数据是否是平稳的，如果不平稳，则使用差分进行处理。
3. AR(p)模型定阶：PACF。
4. MA(q)模型定阶：ACF。
5. ARIMA拟合。
6. 判断残差是否为白噪声，如果是，则说明预测比较好。否则，需要调整阶数，继续预测并判断。
7. Metric：AIC、BIC，越小越好。

4. 参考文档

1. [pmdarima.ima.auto_ima — pmdarima 1.8.5 documentation \(alkaline-ml.com\)](https://pmdarima.readthedocs.io/en/1.8.5/documentation.html)
2. [5 移动平均模型 | 金融时间序列分析讲义 \(pku.edu.cn\)](http://www.pku.edu.cn/~gao/teaching/financial_time_series_analysis/)
3. [用python做时间序列预测九：ARIMA模型简介 - 云+社区 - 腾讯云 \(tencent.com\)](https://cloud.tencent.com/developer/article/1541111)
4. [【python算法+代码案例】时间序列ARIMA模型及预测/机器学习哔哩哔哩bilibili](https://www.bilibili.com/video/BV18t4y1j7hY/)
5. [【实战代码】如何使用ARIMA模型预测世界肺炎确诊人数？【时序数据预测】哔哩哔哩bilibili](https://www.bilibili.com/video/BV18t4y1j7hY/)
6. [KeshengZhang/Covid-19-ARIMA-Time-Series-Prediction: 如何使用ARIMA模型预测世界肺炎确诊人数？【时序数据预测】 \(github.com\)](https://github.com/KeshengZhang/Covid-19-ARIMA-Time-Series-Prediction)
7. [时间序列预测之 AUTO-ARIMA - iUpoint - 博客园 \(cnblogs.com\)](http://www.cnblogs.com/iupoint/)
8. [Python时间序列分析指南! \(qq.com\)](http://www.qq.com/)