Prophet模型是将时间序列分解为三个部分效应，分别为趋势效应，周期效应（如季节周期性）和节假日或异常值。

Prophet模型构建



为趋势函数，为周期函数，为节假日或者异常事件带来的效应，为随机扰动项。

，趋势函数通常采用了logistic增长函数。其中，C是饱和值（承载能力）表示最大极限值？，k是增长率，m是偏置参数。其中C是可以设置的参数，也可以是动态的，另外可以考虑转折点，即前后存在k,m的差异。

当要预测的问题没有饱和增长的趋势时，分段连续增长率模型有时候非常有效。 是判断转折点，转折点为，k是增长率，是增长率的调整值，m是偏置参数，设置为来使得函数连续。

转折点的设定采用稀疏先验自动确定

周期性依靠傅里叶级数构造P为我们想要的时间序列的规则周期长度。N需要提前确定，对于年和周的季节性，我们发现N=10和N=3效果最好。a,b为待估参数

异常值通常采用虚拟变量的方法为过去和未来节假日的集合参数？

Arima模型  
首先预处理保证模型平稳性

为何要满足平稳性：伪回归检验不可靠，时序的一致性，基于经济学含义，基于随机变量的历史和现状来推测其未来，假设随机变量的历史和现状具有代表性或可延续性，随机变量的基本特性必须能在包括未来阶段的一个长时期里维持不变。

平稳性：ADF检验，看图也可（如果有趋势和周期性基本就不平稳）

趋势方法：一般而言，取对差分

ACF和PACF的任务

主要可以识别ar和ma阶数，但是往往更多用AIC，BIC信息准则识别，另外ACF图也可识别平稳性，ACF如果既不截尾又不拖尾，呈现一个单调的形式就说明存在趋势。

Arima和prophet的区别

|  |  |  |
| --- | --- | --- |
|  | Prophet | arima |
| 预处理 | 没必要平稳性  基本不需要预处理，缺失值和异常值的存在依然可以估计，但是异常值对趋势项会有影响，可以通过将异常值做缺失值处理解决 | 必须是平稳性数据 |
| 模型的理论基础 | 将时间序列分解三个效应，趋势、周期、节假日或异常点 | 认为历史对未来会有影响，包括模型能被解释的被解释变量和模型无法被解释的部分随机扰动项 |
| 线性和非线性 | 能够处理线性关系，也能处理非线性关系，但部分需要人为识别，比如趋势项的逻辑增长 | 只能捕捉线性关系 |
| 趋势增长的区别 | 可以通过先验检验自动识别转折点，即不同区间存在不同的增长水平，可能会导致过拟合或欠拟合 | 基本用一阶差分进行退势，增长水平都维持一个水平下的估计，基本难以出现过拟合问题 |
| 季节性的区别 | 通过傅里叶级数识别周期性，因此周期性的处理非常好,但是周期需要人为识别，但同样或容易造成过拟合或欠拟合 | 基本无法处理周期性问题，虽然有SARIMA模型，但处理简单，也就季节差分，导致需要人为识别周期，并且基本不会过拟合 |
| 异常值或者节假日 | 需要人为识别，工程量可能巨大，同时可能造成过拟合和欠拟合问题 | 影响很大且基本无法规避 |
| 模型参数估计 | 模型参数众多，三个项需要估计的参数可能都十分庞大，估计方法复杂 | 模型简单，模型参数较少，一般而言不会有太高阶数模型 |
| 模型的衍生 | 基本不知道是否有衍生的模型 | 衍生的模型众多，有ARIMAX,SARIMA,ARIMAGARCH，starma，和其他模型融合的很多 |
| 模型优点 | 由于存在效应分解，解释力度较高，另外有一定非线性处理能力，通过调参可能有一个比较好的结果 | 模型简单，估计方法容易，基本不用考虑过拟合和欠拟合问题 |
| 模型缺点 | 模型参数较多，有较多人为调参的步骤，这些可能对结果影响较大 | 有假设，模型被限制较多，很可能存在数据和方法不匹配的问题，只能处理线性关系，有一定局限性 |