大创结题思路

最终呈现：ppt+论文+可视化预测界面

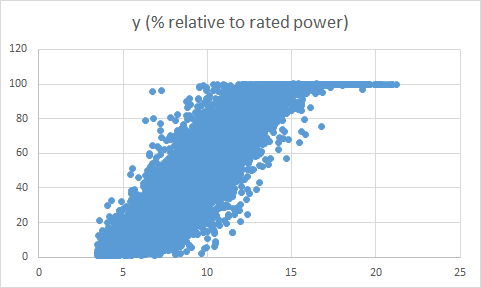
数据收集

数据集1：100万个数据点，包含风速，温度，功率等7维数据

缺点：时间连续性很差，无法填补缺失值，故只用作相关性分析

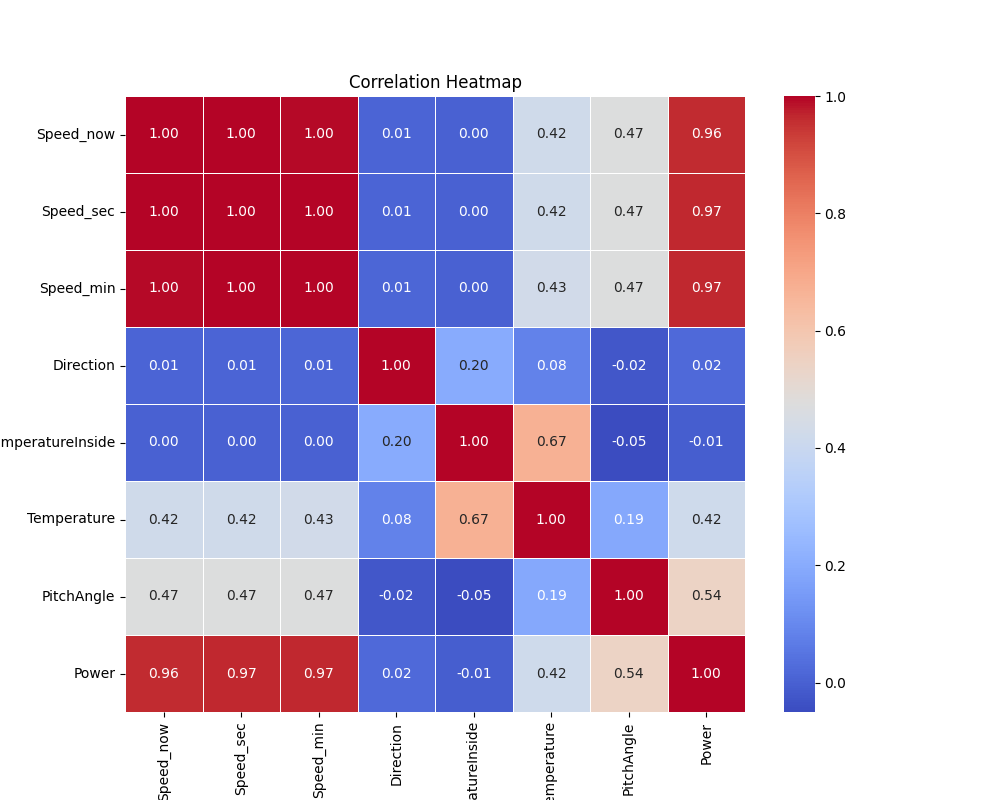
数据集2：4万个数据点，只包含风速和功率，时间连续性较好

数据处理（异常值：dbscan，缺失值：） 拟合风速功率曲线



根据风机规则和数据密度进行筛除

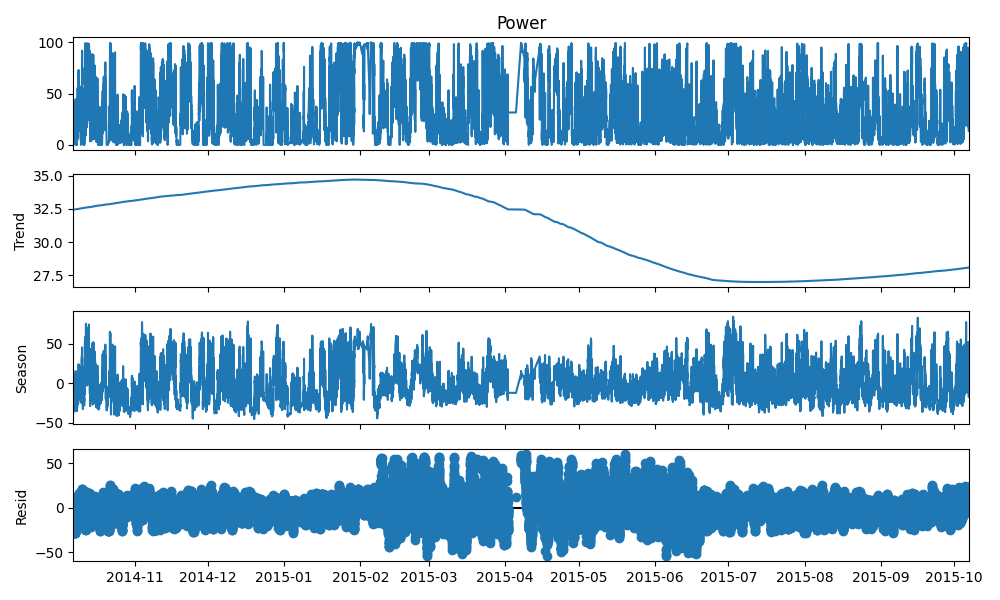
数据分析：相关性，热力图



发现功率主要与风速有强相关性，和浆距角有关，和风向，气温相关性较差

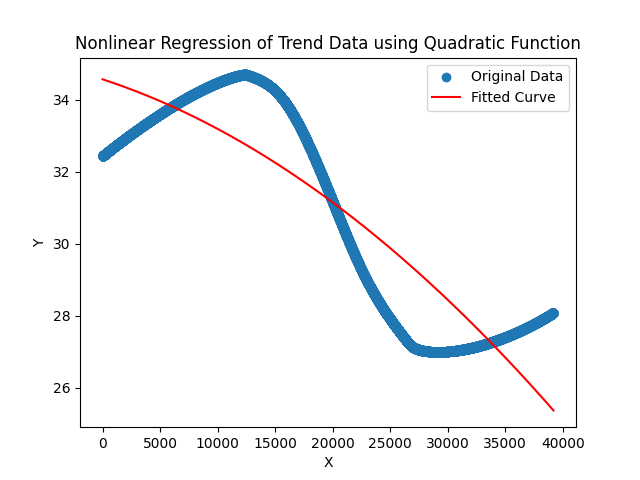
数据分解：STL

将功率数据分解为趋势，季节和残差三项，便于分析数据特征



模型训练：

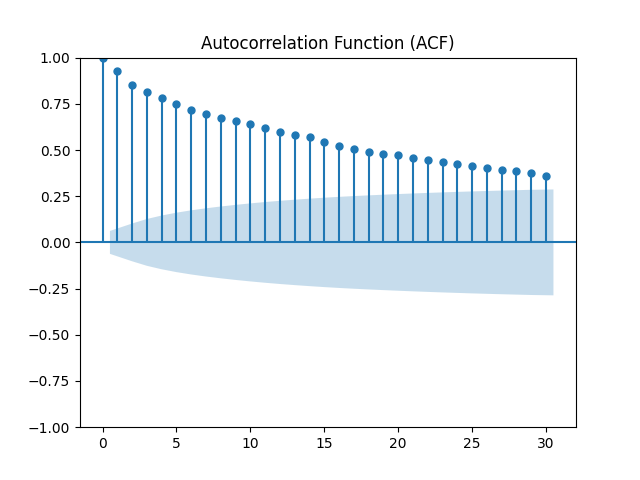
趋势项：较为简单，采用多元非线性回归

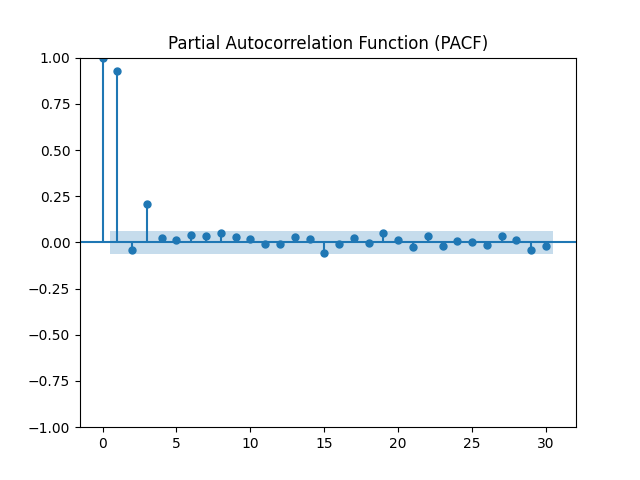


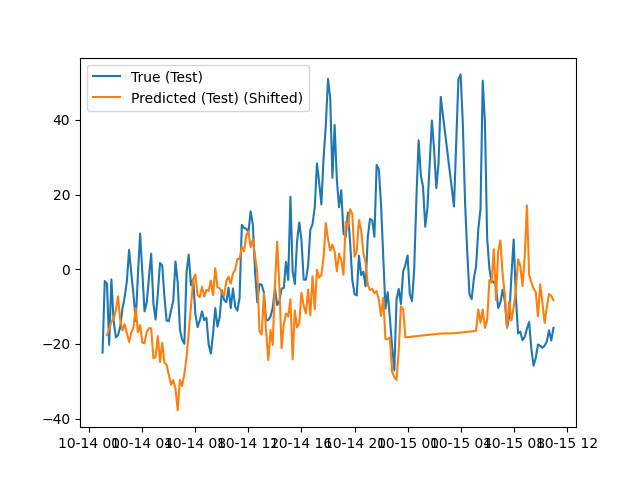
（二次函数拟合，效果并不好，需要选择更好的函数）

季节项：采用ARIMA捕捉季节变化趋势，使用滑动窗口逐步训练模型

根据ACF 和PACF图确定阶数

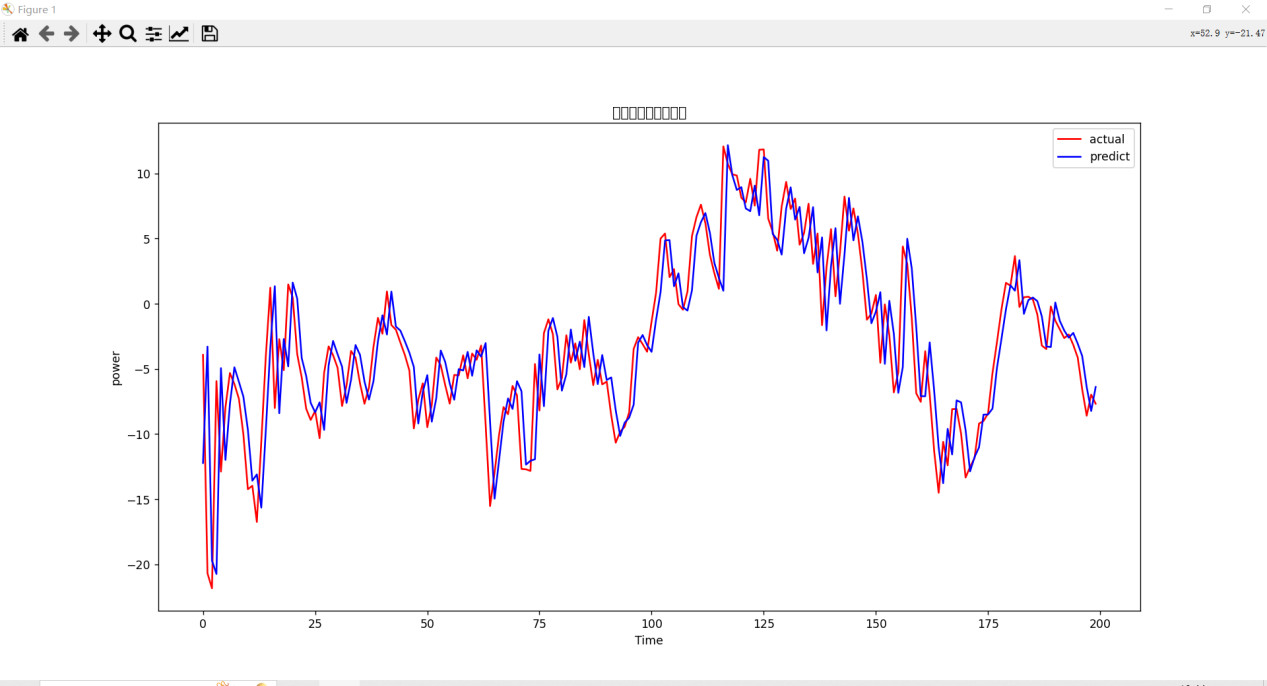




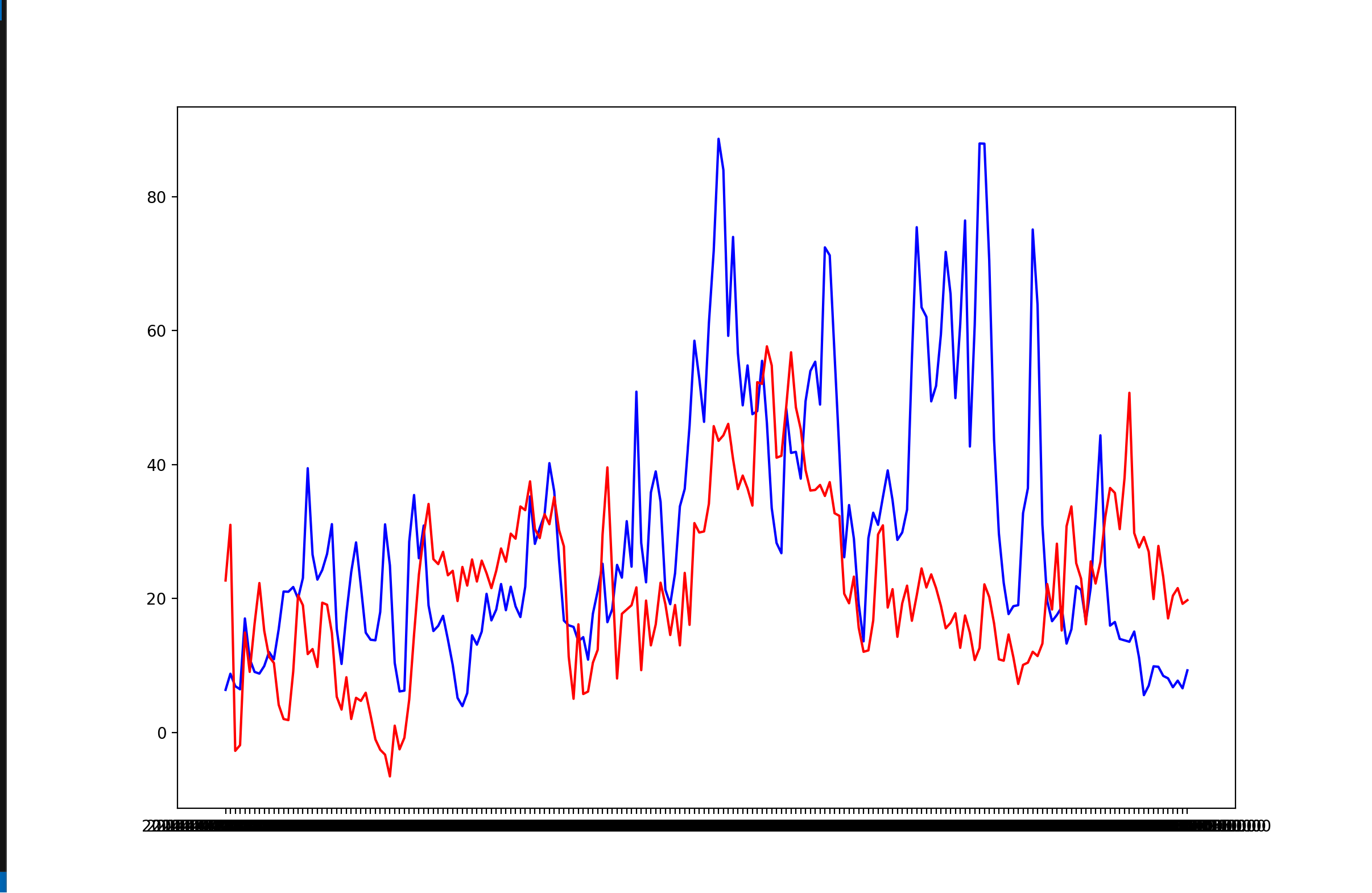


捕捉大概变化趋势，效果不佳是因为数据中有缺失值，得到残差继续输入神经网络(?)

残差：输入LSTM神经网络



模型预测：叠加



可视化：QT，直观显示未来的预测值