

问题二的建模与求解

思路：

建立基于历史功率的光伏电站日前发电功率预测模型，进行发电功率预测，根据附件 1 中考核要求分析你所采用方法的准确性。

分析

只利用历史数据，对后续 7 天进行预测。

按照题目要求，第 2、5、8、11 个月最后一周数据作为测试集，其他数据为训练集，说明数据需要用尽可能多的年份，通过机器学习/深度学习算法，同时挖掘**年度周期信息**和**月份波动信息**。要求的时间分辨率为 15 分钟，这要求提供的数据集的颗粒度尽可能更小或者一致。最后根据附件 1 计算白昼时段的误差统计指标。

目前考虑使用基于滚动窗口 + LightGBM 模型来进行时间序列的建模，并显式地将上述提到的**年度周期信息**和**月份波动信息**加入网络，计算指标并进行对比。

问题解决：

以纬度 (Latitude): 51.8830° N, 经度 (Longitude): 4.4849° E 的位于荷兰鹿特丹市区南部的 Sevwilaweg 55 Rotterdam 的发电站为对象进行分析，这是一家 244.8kW 系统规模的发电站，数据集从 2013 年到 2023 年共 10 年时间。

为了确保能够将第 2、5、8、11 个月最后一周数据作为测试集，且保证数据的连续性，采取如下的方式进行预测：

● 表 1 数据集划分说明

训练集	测试集
2013 年到 2022 年的所有数据，加上 2 月最后一周之前的数据	2023 年 2 月最后一周
2013 年到 2022 年的所有数据，加上 5 月最后一周之前的数据	2023 年 5 月最后一周
2013 年到 2022 年的所有数据，加上 8 月最后一周之前的数据	2023 年 8 月最后一周
2013 年到 2022 年的所有数据，加上 11 月最后一周之前的数据	2023 年 11 月最后一周

得到如下四个数据集划分：

```
Month 2: Train = 355582, Test = 672
Month 5: Train = 364414, Test = 672
Month 8: Train = 373246, Test = 672
Month 11: Train = 381982, Test = 672
```

图 8 划分数据集尺寸

模型建立

一、模型概念

1. 目标：预测未来一周（672 个 15 分钟点）的光伏发电功率，只使用历史功率数据。

方法核心：

2. 利用 LightGBM 作为学习器建模非线性映射；
3. 显式引入周期性时间特征，如：小时、日、月等；
4. 使用滑动窗口机制生成训练样本；
5. 支持逐步滚动预测或多步预测。

二、输入输出形式：

1. 输入特征 X：

历史功率窗口（如前 96 个点 = 1 天）；

时间周期特征（hour, minute, dayofyear, weekday, month, is_daytime 等）；

2. 输出目标 y：

未来一个时间点的功率，可扩展为多个时间点（多输出或迭代滚动）

三、算法流程（逐点滚动预测）

步骤 1：准备数据

输入序列为 pandas Series: train_series 和 test_series。

设置开机容量 P_CAP 用于归一化和误差计算。

步骤 2：构造特征（滑动窗口）

对于每一个训练样本：提取前 N 个历史功率点作为特征；追加周期性时间特征，比如目标时刻的：

- 小时（hour）
- 分钟（minute）
- 星期（weekday）
- 一年中的第几天（dayofyear）
- 月份（month）

滑窗大小（如 96）和目标预测步长（如 1）可调。

步骤 3：训练 LightGBM 模型

构造特征矩阵 X_train 和标签 y_train；

用 LightGBM 的 LGBMRegressor 拟合；

支持 early stopping 或默认参数先调试。

步骤 4：滚动预测（逐点）

对于每个未来时间点：构造输入窗口（含滑动历史和时间特征）；

使用模型预测当前时间点；

将预测结果加入窗口，继续滚动；

保持自回归性：

步骤 5：评估与误差分析

未来每个点都用自己预测值作为下一轮输入。

使用真实测试集 test_series 比较模型输出；

计算如下指标：

1. RMSE、
2. MAE、

3. ME
4. Pearson 相关系数
5. 准确率（归一 RMSE）
6. 合格率（基于阈值 25% 的误差）

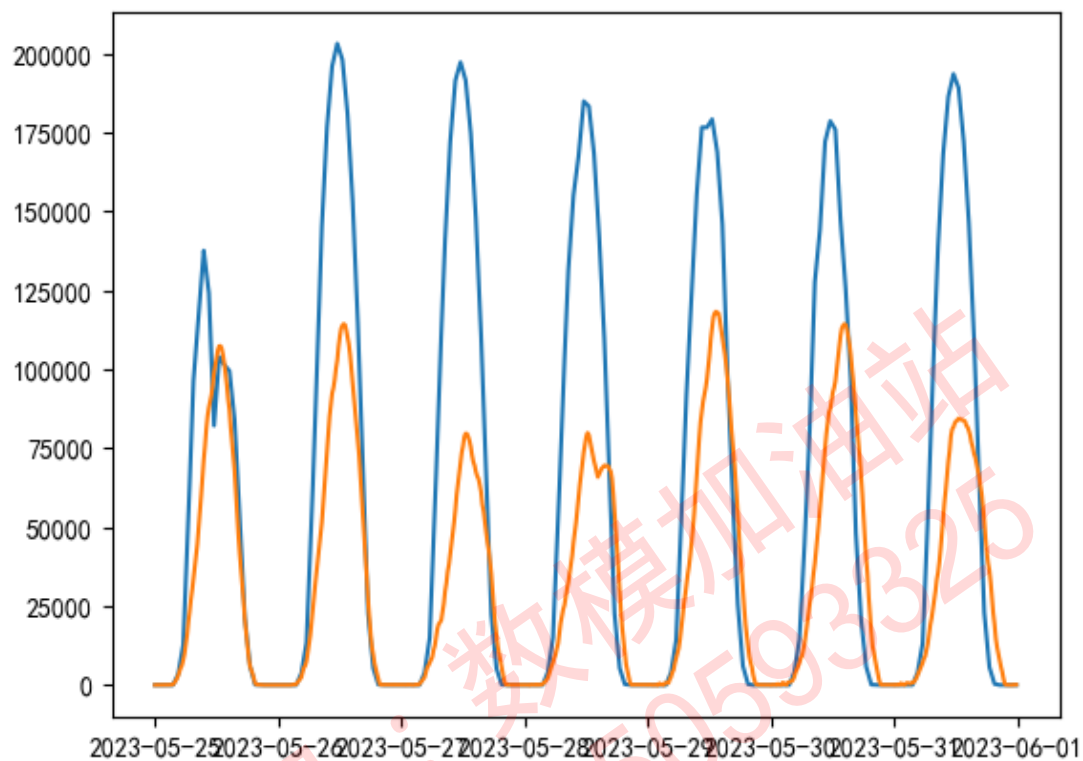


图9 5月最后一个周的数据对比（未加入周期特征）

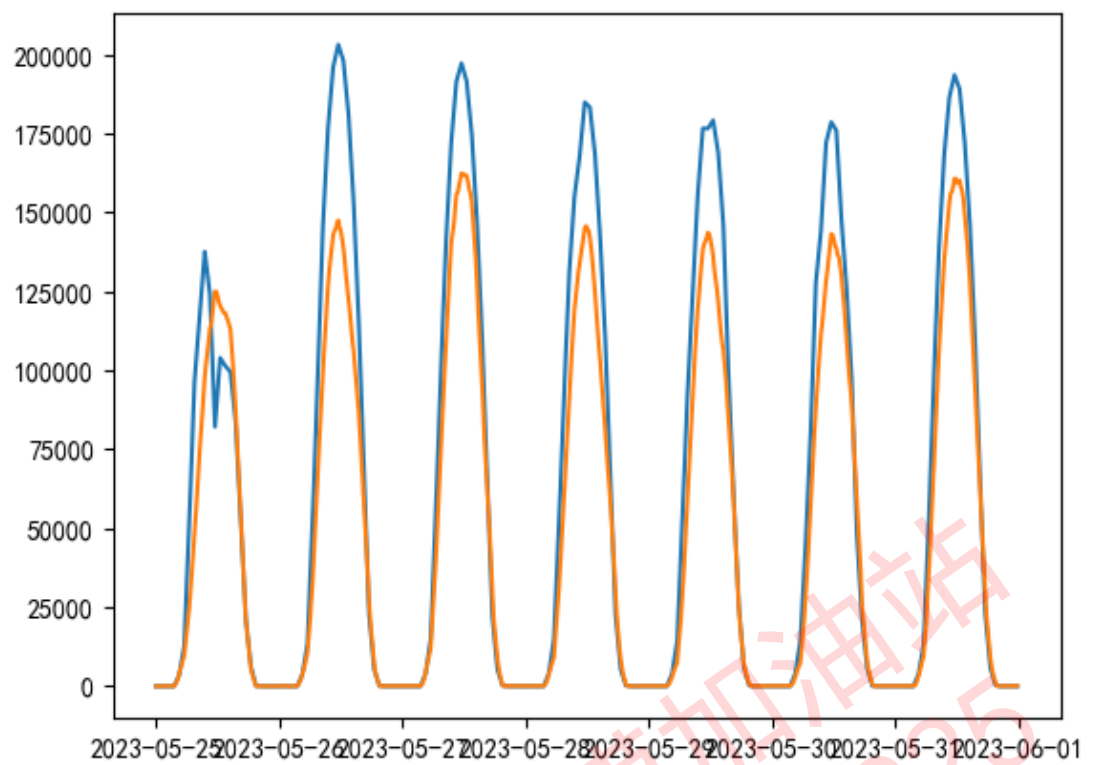


图 10 5 月最后一个周的数据对比（加入周期特征）

计算指标，得到

表 2 2 月最后一个周的数据对比

指标	数值
RMSE	30361.18
MAE	15043.12
ME	12647.61
相关系数	0.8424
准确率	87.60%
合格率	0.92

表 3 5 月最后一个周的数据对比

指标	数值
RMSE	21584.53
MAE	13671.78
ME	-12011.30
相关系数	0.9845
准确率	91.18%
合格率	1.0

表 4 8 月最后一个周的数据对比

指标	数值
----	----

RMSE	16053.93
MAE	8775.96
ME	1445.20
相关系数	0.9431
准确率	93.44
合格率	0.99

表 5 11 月最后一个周的数据对比

指标	数值
RMSE	31712.23
MAE	17485.10
ME	11543.81
相关系数	0.6822
准确率	87.05%
合格率	0.92

存在较多问题，该模型的学习特征只有时间序列的发电功率，基本只学到了趋势，但如果类似与 11 月这种，前几天可能是由于雨天导致发电量较少，仅仅通过发电功率趋势没有办法进行精准预测。

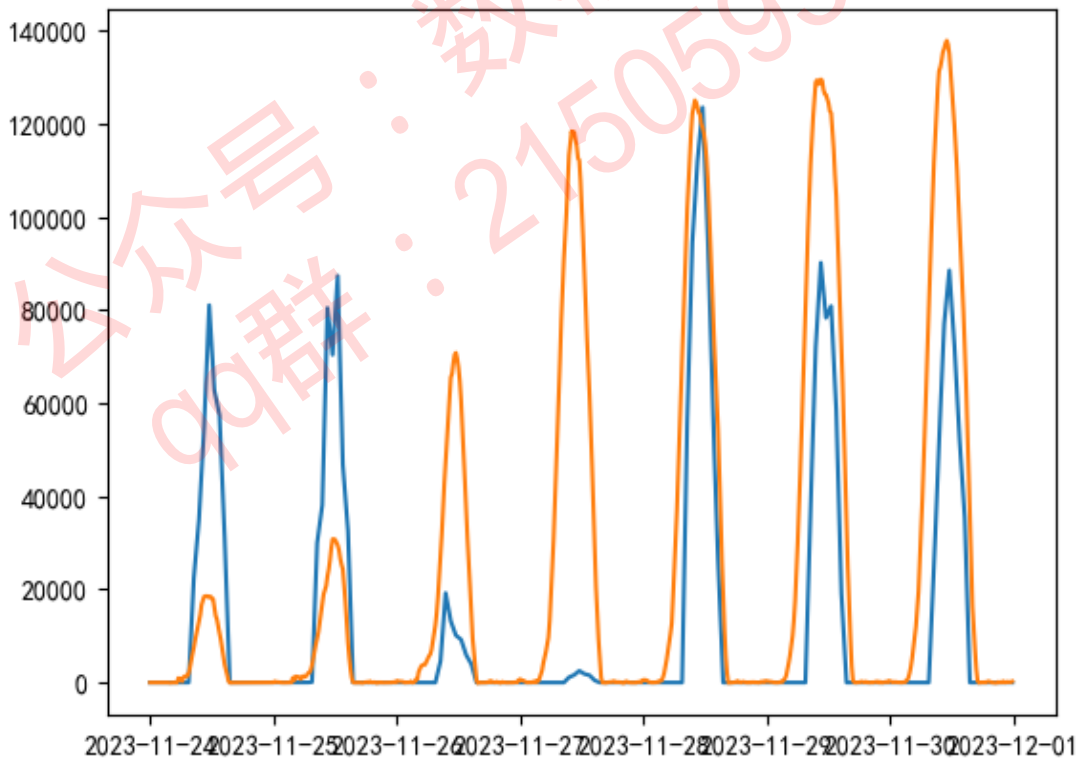


图 11 11 月最后一个周的数据对比（加入周期特征）

所以需要把天气信息也加入进来。