

光伏发电预测第二次报告

张翼鹏

December 16, 2019

目录

1	误差	2
1.1	计算方式	2
2	数据	2
2.1	数据缺失（测试集）	2
2.2	数据异常	2
3	特征	3
3.1	含义	3
3.2	关系	3
4	处理	3
5	进展	4

1 误差

1.1 计算方式

- 1、一天24个小时，每15分钟记录一次数据。一天之内如果没有数据缺失，应为96个数据。在这96个数据中，按照公式，计算实际功率大于 $0.03 * P$ 的点的平均值，得到当日平均绝对偏差。
- 2、计算一个月内每一天的当日平均绝对偏差的平均值，得到当月绝对平均偏差。
- 3、对于每个电站，计算其每个月的当月绝对平均偏差的平均值，得到电站绝对平均偏差。
- 4、计算4个电站的电站绝对平均偏差的平均值，得到最终的误差。

2 数据

2.1 数据缺失（测试集）

电站1：5月至8月，无数据缺失。

电站2：5月至9月，无数据缺失。

电站4：8月1日至10月8日；

缺少9月19日、9月26日、9月29日数据；

9月27日只有一条数据。

电站6：5月1日至10月8日；

缺少5月30-31日、9月30日、10月1-7日数据；

5月29日、6月12日、9月27日、9月29日只有一条数据；

6月10日只到14:00，8月15日只到10:00。

2.2 数据异常

[训练集]

电站2：2017-2-17实际功率为0；

2017-6-25 19:45至2017-6-29 23:45实际功率为0；

2017-12-27 9:15至2017-12-29 9:30实际功率为0.6；

2018-3-7 7:15至2018-3-7 14:45实际功率为0;

2018-4-10 17:00至2018-4-12 8:00实际功率为4.4。

电站4: 2017-8-4 13:45至2017-8-6 8:30实际功率为29.667;

2017-10-17 14:15至2017-10-17 19:00实际功率为0;

2017-12-25至2017-12-29实际功率为0。

电站6: 2017-5-24 17:45至2017-5-28 6:15实际功率为0。

[测试集]

电站2: 2018-5-13实际功率为0;

2018-6-5 17:00至2018-6-7 12:15实际功率为4.2。

电站6: 2018-8-18至2018-8-19实际功率为0。

3 特征

3.1 含义

自变量中辐照度和所有气象数据均为预测值; 实际功率和实发辐照度均为实测值。

3.2 关系

经验证, 实际功率与实发辐照度间具有强线性关系, 所以实际情况下(且辐照度预测较准确), 线性模型是最理想的。

所以我们一定程度上没有必要去建立过于复杂的模型, 重点在于数据预处理和模型融合。

4 处理

去掉夜间数据(或去掉实发辐照度/实际功率为0的数据)。

去掉训练集异常数据。

加入时间变量。

先以实发辐照度为标签, 再以实际功率为标签。

5 进展

电站1: 0.1205 (svr, gamma=0.004)

电站2: 0.1514 (岭回归, 正则化参数330)

电站4: 0.1408 (多项式岭回归, 次数4, 正则化参数400)

电站6: 0.12605554 (线性回归)