

# 光伏发电预测第三次报告

张翼鹏

December 19, 2019

## 目录

<b>1</b>	<b>误差</b>	<b>2</b>
1.1	误差的某些规律 . . . . .	2
<b>2</b>	<b>参数</b>	<b>2</b>
<b>3</b>	<b>特征</b>	<b>3</b>
3.1	交叉项 . . . . .	3
<b>4</b>	<b>模型</b>	<b>3</b>

## 1 误差

### 1.1 误差的某些规律

计算误差时，列出每一天的误差，发现不同回归方法在某些天内的误差普遍较大。以电站1为例，列出几种回归下误差大于20%的天的序号。

[线性回归]

4、6、8、9、16、25、32、35、37、53、66、71、75、79、81、109、112、116

[支持向量机回归]

4、8、16、35、37、53、66、75、79、81、83、109、116

[多项式回归]

4、8、11、25、32、35、37、53、66、79、81、109、116

[随机森林回归]

4、8、21、32、35、37、53、66、79、81、109、112、116

【重合天数】

4、8、35、37、53、66、79、81、109、116

观察这些天的数据，初步发现两个原因：

- 1、若干个连续数据点，其相同特征之间相差并不大，但实际功率相差较大，造成预测值出现较大误差。
- 2、一般情况下，辐照度在以上每一种回归模型中都可以说是权重最大的一个特征，但在某些数据点中，实发辐照度与实际功率显然不匹配（辐照度高时实际功率过低），可能是其他因素影响了实际功率的值。

目前未想到合适的方法利用或处理这一现象。

## 2 参数

以电站1为例，首先建立线性模型，参数及误差见下表第一行；再在上述基础上，手动调参，参数及误差见下表第二行。

表 1: 线性模型参数

参数	辐照度	风速	风向	温度	压强	湿度	截距	误差
默认	4.97	-0.416	-2.855e-4	-0.963	0.402	-0.274	4.67	0.1293
手动调参	4.945	0.12	-9.1e-4	0.06	0.34	0.31	4.55	0.1164

虽然这种方法和结果不具有推广性，但是给了一个启示：

sklearn里的回归方法，代价函数是其默认的函数（比如线性回归的代价函数为均方误差），如果改用题目给出的误差函数，可能会有更好的结果。

### 3 特征

#### 3.1 交叉项

将气象特征两两组合生成新的特征，再进行线性回归，误差结果如下：

表 2: 添加交叉项的误差情况

	电站1	电站2	电站4	电站6	平均
原始	0.1293	0.1517	0.1556	0.1261	0.1407
风速*风向	0.1303	0.1515	0.1556	0.1268	0.1410
风速*温度	0.1303	0.1503	0.1551	0.1262	0.1405
风速*压强	0.1296	0.1513	0.1551	0.1260	0.1405
风速*湿度	0.1293	0.1512	0.1555	0.1258	0.1405
风向*温度	0.1293	0.1520	0.1559	0.1260	0.1408
风向*压强	0.1294	0.1512	0.1557	0.1260	0.1406
风向*湿度	0.1289	0.1519	0.1550	0.1261	0.1405
温度*压强	0.1287	0.1530	0.1534	0.1258	0.1402
温度*湿度	0.1288	0.1510	0.1542	0.1271	0.1403
压强*湿度	0.1291	0.1505	0.1550	0.1258	0.1401
后三项	0.1290	0.1507	0.1527	0.1277	0.1400

### 4 模型

将数据集分割，分别建模。此处选择用训练集每个月的数据分别建模，每组数据建立四个线性回归模型，分别是普通、加入三项交叉项、只用辐照度、只用辐照度和三项交叉项。以电站1为例，误差结果如下：

表 3: 添加交叉项的误差情况

月份	普通	交叉	辐照	辐照交叉
2016-4	0.1240	0.1215	0.1204	0.1258
2016-5	0.1275	0.1281	0.1258	0.1263
2016-6	0.1225	0.1282	0.1212	0.1211
2016-7	0.1736	0.1659	0.1355	0.1349
2016-8	0.1422	0.1543	0.1211	0.1201
2016-9	0.1377	0.1726	0.1313	0.1325
2016-10	0.1471	0.1535	0.1333	0.1546
2016-11	0.2126	0.2873	0.2793	0.3009
2016-12	0.4146	0.7499	0.4145	0.4293
2017-1	0.2633	0.3222	0.3308	0.3189
2017-2	0.1640	0.1566	0.2133	0.1929
2017-3	0.1333	0.1345	0.1492	0.1407
2017-4	0.1220	0.1229	0.1198	0.1201
2017-5	0.1212	0.1225	0.1197	0.1210
2017-6	0.1242	0.1289	0.1227	0.1234
2017-7	0.1183	0.1304	0.1189	0.1185
2017-8	0.1306	0.1413	0.1263	0.1271
2017-10	0.2401	0.3634	0.2258	0.2421
2017-11	0.2897	0.3125	0.3355	0.3431
2017-12	0.3215	0.3708	0.3699	0.3624
2018-1	0.3121	0.3269	0.3612	0.3467
2018-2	0.1771	0.2292	0.2304	0.2051
2018-3	0.1497	0.1652	0.1572	0.1554
2018-4	0.1186	0.1182	0.1203	0.1196
全部	0.1293	0.1290	0.1312	0.1312

可以发现，表中有些误差很小，最小达到0.1182，或许说明如果数据准确，只用一个月的数据集足以让我们将误差降至12%以下。

但是表中有不少月份的误差非常大，经检验，发现依然出现了辐照度与实际功率不匹配情况：

电站1测试集中，辐照度（-1至1）与实际功率（0至10）在每一天几乎都是先增大再减小，趋势比较一致，线性关系明显；但在训练集上表那些误差较大的月份中，实际功率依然趋势正常，但辐照度却几乎没有大于0.2过，我们用这样的训练集线性回归很难得出想要的结果。（电站2也有类似现象，但电站2的实际功率变化更加不稳定，所以误差普遍更大，在0.15以上。）

排除记录错误或计算错误，我认为出现这种情况的原因有如下几点：

- 1、由除辐照度外的特征导致。由上表可观察到误差在10月至次年2月之间普遍较大，观察数据仅发现温度特征在这段时间与其他时间明显不同（即温度低）。猜测在一年之中温度较低时，有一种因素（或许是温度或许是其他）可以在辐照度较低的情况下依然使电站达到高的发电功率。
- 2、由当时电站内其他因素所造成，但我们无法得知是何种因素，也就无法通过建模弥补这方面的误差。