

课程项目：风电功率预测

张平文，文再文课题组

December 20, 2017

1 功率预测原理介绍

风电功率预测是指以风电场历史功率、历史风速、地形地貌、数值天气预报、风力发电机组运行状态等数据建立的风电场输出功率预测模型。风电功率预测按时间尺度可以分为长期（数年）、中期（一周）、短期（0-72h）、超短期（0-4h），按预测范围可以分为单台风机与风力机群。

虽然有众多分类，但总体而言，风电功率预测模型可能利用的数据有历史大气数据、历史风速、历史功率、地理条件、历史风电场机组情况，输入的参数可以包括实时大气数据、天气预报结果、风电场机组情况（预测时间风电机组实际运行数量与最大功率），输出为风电场的功率。观察这些量，我们可以发现因为对于某一给定风电场，地理条件与风电场机组情况是长期不变的，所以预测可以转化为两个核心问题：

- 构建风速预测模型；
- 构建功率关于风速的函数。

对于风速预测问题，主要有两种思路，一种是用物理方法模拟大气运动及地理环境，这种思路不需要大量历史数据，但相关方程求解困难，计算量大，另一种是用神经网络等统计方法从历史数据中预测，这种办法计算速度快，但需要大量历史数据。

对于功率函数问题，目前主要的思路有两种：一是使用物理模型，但这需要大量背景知识；另一种是仿效第一个问题的第二种思路利用历史数据，目前是主流的办法。

2 风速预测描述

2.1 目标 and 需求

每天下午3点提供从次日0点开始向后24小时的功率预报结果。假设由风速转化为功率可在30分钟之内完成，则需每天下午2:30之前提供风速预测结果。

2.2 方法

风速预测方法主要有基于物理模型的数值天气预报(NWP)和基于已有风速数据的统计方法和机器学习方法。基于NWP的风速功率预测方法已较成熟，这是人们比较认可用于较长时间风速预测的一种较为准确方法。

目前我们采用数值天气预报和统计分析相结合的方法进行风速预测。数值天气预报采用Weather Research and Forecasting 软件（简称WRF）进行。具体步骤为：首先下载NCEP（美国国家环境预报中心）提供的全球数值预报结果，再将其输入WRF进行中尺度预报，得到风机所在位置的预测风速。但是这个风速往往不太准确，需要风机提供的历史风速测量数据对数值预报结果进行校正。

2.3 我们的问题

根据WRF给出的风机所在位置的预测风速，和历史测量风速，给出风机所在位置的更准确的预测风速。

3 风电场功率预测描述

3.1 数学问题

对给定一个区域的风场，我们的任务是在每天的 15:00 之前给出对第二天的风电功率的预测。我们可以使用的是未来几天的 NWP (天气预报) 的数据以及目前已有的风场的测量数据。

此外，对于压强和温度，只有预测数据，缺乏实地测量数据。但依然在某些模型下可以当作输入进行训练。

值得注意的是，在实际功率的数据中，由于机械或者网络原因，会存在风机中的数据出现异常情况（多为向量中的一段全0），大部分情况下，都会有风机的状态异常的记录。但某些情况下，也存在风机状态异常但没有任何异常状态标志的情况。例如， $VR_t > 0, T_2 \leq t < T_3$ ，但实际功率 $PR_t = 0, T_2 \leq t < T_3$ ，并且没有任何标志表明此段 $T_2 - T_3$ 的功率数据几率有误。此种情况若风速较低，可能是风速不足切入风速，导致没有足够动力进行发电。但当风速较高是，这种情况显然是数据异常，我们在进行预测的时候要注意这种情况的识别。

关于风速和功率的关系，我们可以近似认为风机的输出功率和当时的风速率有着一一对应的函数关系，称为功率曲线。利用功率曲线可以得到较好的结果，但仍会有部分误差。且我们已知不同型号的风机之间的功率曲线不相同，但相同型号的风机的功率曲线基本相同。

4 数据集概述

该数据集由南通某风场的历史实时数据和通过数值天气预报系统计算得出的预测数据组成及部分风机参数数据，总共包含了2016年4月1日到2016年7月31日的6台风机的数据。下面给出各类数据的具体描述。

4.1 历史数据

4.1.1 原始数据

原始历史数据存放在 windpowerdata/HisRawData/ 文件夹下，按时间分文件存储，如 201605023.mat 存储了2016年5月23日的原始历史数据。

每个mat文件通过MATLAB读入后，包含三个字段：speed、power和status，分别记录了风机的速度、功率及工作状态。例如speed,为风机当天的风速数据，读入后为一个两列的向量，第一列为时间点，以s为单位，取值在1-86400之间；第二列为该时间点对应的数据。

其中status表示风机的运行状态，如果是11则表示风机工作正常，其余数字表示风机工作不正常。而且在下一个时间点出现前，风机运行状态保持不变。比如说，status有3行，分别是

seconds	status
12000	11
40000	21
60000	11

这意味着12000s至39999s风机工作正常，40000s至59999s风机工作不正常。

4.2 预测数据

预测数据存放在 windpowerdata/PredictData/ 文件夹下，按时间分文件夹存放，如2016年5月23日的数据存放在201605/23文件夹下。

每天的预测数据包括风速、温度、气压、风向等多种预测数据，其中speed_out_ncl.txt为风速的预测数据，NTspeed_Corrected.txt为初步线性矫正后的数据（建议采用），P_out_ncl.txt为气压的预测数据，T_out_ncl.txt为温度的预测数据，u_out_ncl.txt, v_out_ncl.txt和w_out_ncl.txt分别为三个维度上的风速数据；而TimeSeries.txt为每次预测覆盖的时间点，文件中使用的是零时区的时间，为前一天18:00到三天后的16:00，换算到东八区为当天02:00到三天后的24:00，除此之外我们使用的都是东八区的时间，包括之前的测量数据。时间间隔是每15分钟一个时间点，共377个时间点。每个预测数据 txt 文件中，均存储有一个长为377的向量，第 k 行的数据代表第 k 个时间点对应的预测数据。

4.3 风机参数数据

风机参数数据存放在windpowerdata/Parameter/文件夹下，包括文件P_install_S_full_NT.mat,分别记录了满载功率及最大发电功率。

P_install_S_full_NT.mat记载了各台风机的满载功率，通过MATLAB读入后包含两个变量:S_full和P_install_all。S_full 等于14，代表风速达到或超过14m/s 时，该风场所有风机均达到满载状态； P_install_all为满载功率。

5 问题描述

5.1 数据清理

我们得到测量数据不能直接使用，原因主要有两点：

- 测量数据时间分布是不均匀的，可能每隔几秒会有一个测试点，也有可能间隔几分钟，甚至有时候连续几小时缺失数据；
- 测量的功率数据未必是有效的，由于风机可能不在正常工作状态。

在进行预测问题前，大家需要识别异常的功率数据点，并对时间分布不均匀的问题做恰当的处理。

5.2 风速预测的问题

5.2.1 可利用的数据

- 截止至预测当天12:00之前的历史风速测量值.
(以12月21日为例，历史风速测量值包括12月21日12:00前的历史风速测量值.)
- 由WRF（数值天气预报）得到的结果，包括原始预测风速、原始预测温度和原始预测压强等。利用数值天气预报的方法，每一天我们会给出预测当天02:00开始的之后94h的预测结果，每15min一个预测点。
(以12月21日为例，WRF得到的结果为12月21日02:00之后94h的预测结果，每15min一个预测点。原始预测风速、原始预测温度和原始预测压强等.)
- 实际操作中发现，原始预测风速与真实风速有较强的相关性，但预测精度仍不是很理想。

5.2.2 预测量

- 次日00:00开始的之后24h的预测风速，每15min一个预测点。
(以12月21日为例，预测量为12月22日00:00开始的之后24h的预测风速，每15min一个预测点.)

5.3 功率预测的问题

5.3.1 可利用的数据

- 风速预测中可利用的所有数据。
- 截止至预测当天12:00之前的历史功率测量值。

5.3.2 预测量

- 次日00:00开始的之后24h的预测功率，每15min一个预测点。

5.4 异常风速和功率预测的问题

我们发现 WRF (数值天气预报) 给出风速预测有时会有很大的误差。如果用该数据进行功率预测, 那么会导致很大的误差。然而对风速预测的修正并不是一件容易的事, 为了解决该问题, 我们可以考虑一个更简单的问题: 对 WRF (数值天气预报) 给出异常风速预测进行检测和预报。

首先我们需要对异常风速进行定义, 一个简单的定义就是预测风速和实际风速的差大于 $X m/s$ 时, 就认为该预测风速是异常的其中 X 可以自行选取, 比如取为 $3m/s$ 。

还有一种方式是检测功率预测异常。首先利用直接利用预测风速进行功率预测, 如果得到的预测功率与实际功率的差大于某一定值时, 比如说大于满发功率的 30%, 就认为该功率预测异常。

有了异常风速和异常功率的定义, 我们就可以预测之后是否会有异常风速和异常功率, 在此基础上, 对出现异常风速和异常功率的时间点, 我们可以用其他方式做预测已达到更高的预测精度。

6 风电预测模型的评估

6.1 预测功率评估

对功率预测的评估我们使用的是风场总功率, 以下是功率预测模型的评估方式: $P(t+k)$ 表示在 $t+k$ 时刻的整个风场的实际风电功率; $P(t+k|t)$ 则表示在 t 时刻做出的对 $t+k$ 时刻整个风场的风电功率的预测值; $P_{install}$ 则表示风机的装机功率, n 表示时间点数目。在风电预测模型评估中, 主要有两种误差测度, 即规范平均误差 (NMAE) 和均方误差 (RMSE)。其定义分别如下:

NMAE:

$$\frac{1}{P_{install}} \left(\frac{\sum_{k=1}^n |P_{pred}(t+k|t) - P_{real}(t+k)|}{n} \right)$$

RMSE:

$$\frac{1}{P_{install}} \left(\frac{\sum_{k=1}^n (P_{pred}(t+k|t) - P_{real}(t+k))^2}{n} \right)^{1/2}$$

根据中国国家能源局的官方文件, 我国所采纳的误差测度为 RMSE。更具体的情况如下:

- 风场需要在当天 15:00 前提供第二天的每 15min 的平均风场总功率预测值以便进行评估;
- 测量并记录第二天的每 15min 的实际平均风场总功率值; 如果 15min 内无观测点则记为 0, 该时间点标记为无效;
- 计算该天的 RMSE 的值作为准确性的评估。无效的时间点不纳入误差的计算。

我们的目标就是使得上述的风场总功率的 RMSE 尽量小, 我们目前只给出一台风机的数据, 因此只需计算一台风机的 RMSE。

在这个课程项目中, 请大家给出数据集中最后一个月 (7月) 的所有预测功率, 并计算 RMSE。可以使用的训练数据参见 5.3。

6.2 预测风速评估

对风速预测的评估, 我们是对单台风机进行评估, 以下是风速预测模型的评估方式: $S(t+k)$ 表示在 $t+k$ 时刻的实际风电功率; $S(t+k|t)$ 则表示在 t 时刻做出的对 $t+k$ 时刻风电功率的预测值; $S_{install}$ 则表示风机的满发时的风速, n 表示时间点数目。在风电预测模型评估中, 对单台风机的规范平均误差 (NMAE) 和均方误差 (RMSE), 其定义分别如下:

NMAE:

$$\frac{1}{S_{install}} \left(\frac{\sum_{k=1}^n |S_{pred}(t+k|t) - S_{real}(t+k)|}{n} \right)$$

RMSE:

$$\frac{1}{S_{install}} \left(\frac{\sum_{k=1}^n (S_{pred}(t+k|t) - S_{real}(t+k))^2}{n} \right)^{1/2}$$

更具体的情况如下:

- 风场需要在当天 15:00 前提供第二天的每15min 的每台风机的平均风速预测值以便进行评估；
- 测量并记录第二天的每15min 的实际风机的平均风速，如果15min内无观测点则记为0，该时间点标记为无效；
- 计算该天误差（无效的时间点不纳入误差的计算）。

在这个课程项目中，请大家给出数据集中最后一个月（7月）的所有预测风速，并计算 RMSE。可以使用的训练数据参见 5.2.