

基于 LSTM 的风力系统上网功率预测

未央-能动 02 徐智昱 2020012991

摘要

在双碳目标下，根据历史数据对风电机未来一段时间内的运行状况结合其它设施进行调峰具有重大意义。本研究采用实际获得的风电机组一个月的运行数据与当地一个月的风力数据，进行未来 1 小时每 15 分钟一个点的功率预测。风电机的功率与当地风速有着紧密关系，而风速与风功率也有明显的自回归现象，因此考虑能够考虑这种自回归与相关性的模型进行建模。循环神经网络 (RNN) 就是这样一种方法，本研究采用了长短期记忆网络 (LSTM) 对风功率的预测进行了研究，并通过调整参数试图找到较为适合的模型。同时，也考虑利用传统的时间序列分析 (TSA) 的方法对此数据进行考量，比较经典方法与现代方法之间效果的差异。

关键词： TSA, RNN, LSTM

1 引言

1.1 数据描述与预处理

获得的风电数据来自江苏的风电场，包含风速与风功率两组变量，在 2015/10/1 0:00 到 2015/10/31 23:59 被采集，每 30 秒采集一次数据，内有一系列缺失数据。根据国家标准，需要对未来 1 个小时的 4 个，每个间隔 15 分钟的点进行预测，因此将数据集每 15 个点进行合并，可以借此作出散点图与时序图。

由时序图可以发现其存在较多的时间缺失区间，因此在实际建立模型时选择使用分段的区间进行模型的训练，放弃具有缺失值的部分。

同时可以作出的风功率与风速关系的图像，并同时用 Kernel Regression 对数据符合的形式作初步的拟合，结果如下图所示：

由图可见，二者关系可能近似符合类似于 Logistic Function 的形式；同时可能经过恰当的变换后，可以找到用于描述的线性函数描述其变化。

在时间序列中，常用的检测离群值 (Outlier) 的方法是认为每个数据独立同分布地服从于正态分布，由正态分布的置信区间判断有无明显的离群值。利用这个方法可以作出此数据的离群值：

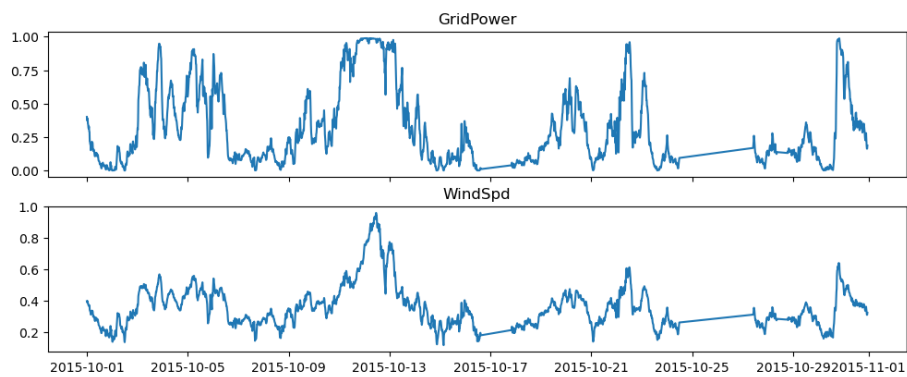


图 1: 原始数据时序图

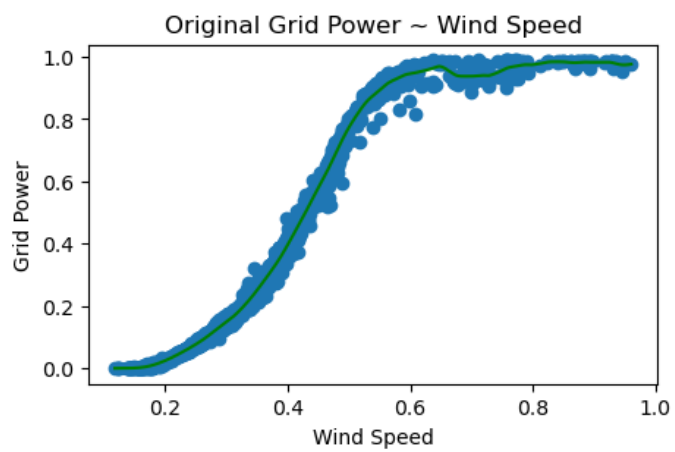


图 2: 风功率风速关系图

由图像可以发现，数据的特性良好，存在极少的离群值，因此不需要对数据的离群值进行处

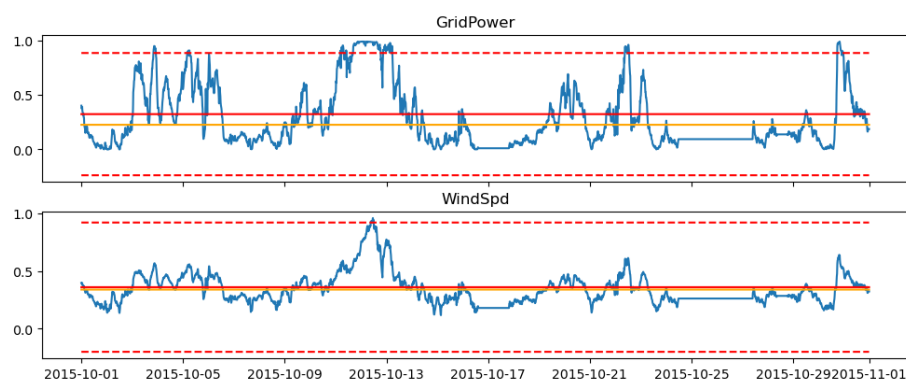


图 3: 离群值检测

理。

由图 1 可见，数据实际上存在较多的噪音，因此考虑使用滑动平均 (Moving Average) 的方法，降低数据的噪音。下图是以 20 个数据点为窗口的滑动平均的结果：

滑动平均后数据明显变得光滑，但也失去了一部分的信息，将对滑动平均前后分别建立模型

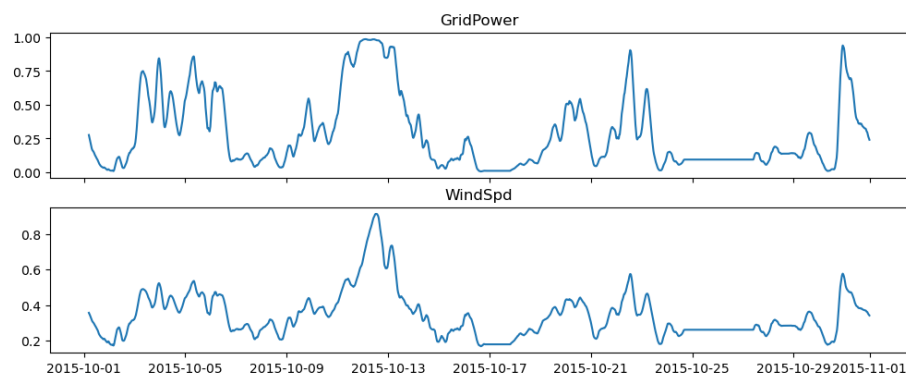


图 4: 滑动平均数据

进行研究来探究实际中应该使用何种模型进行训练。

1.2 方法概述

1.2.1 LSTM

在实际数据中,经常会遇到序列数据,数据之间是存在时空相关性的情况,如本研究中遇到的风功率与风速数据。循环神经网络 (RNN) 可以很好地处理序列信息,通过引入状态变量存储过去的信息和当前的输入,从而确定当前的输出。

RNN 的本质是一种自回归模型,认为当前值具有某种意义上的马尔可夫性质,其概率分布只与过去一段时间的数据点相关,因此通过观察一个非常短的历史就能够预测未来的情况如何。由于在实际中,时间较为久远的信息对当前的影响一般会随时间减小,因此经常可以忽略不计,马尔可夫模型是能够在各种场景中大量使用的。在本研究中也采用马尔可夫模型对序列进行预测。

在 RNN 模型中,利用迭代的方法对模型进行优化,在实际使用中使用 batch 的方法,采用 Stochastic Gradient Descent 利用梯度下降的方法对模型参数进行优化,从而得到在 loss function 意义下具有优越性的模型。为了提高模型的鲁棒性,又提出了 LSTM 模型。

由于隐变量模型存在着长期信息保存和短期输入缺失的问题,长短期记忆网络 (LSTM) 引入了计算机的逻辑门,其中一个门用来输出,另一哥们用来决定何时读入,并增加控制何时对单元进行重置的门。三个门都利用线性运算与隐藏层相关联。通过三个门的状态不同,决定 LSTM 在不同时刻的计算状态,从而实现对过去数据的记忆与以往的过程,达到比较良好的效果。

由于 LSTM 天然地具有良好的处理时空相关性的能力,因此本文采用 LSTM 进行部分的研究。

1.3 评价指标

本文采用 MSE 和 Qualification Rate 两个指标对不同模型得到的结果进行评价。二者定义分别如下。

$$MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$Q = \frac{1}{n} \sum_{i=1}^n B_i \times 100\% \quad (2)$$

$$B_i = \begin{cases} 1, & 1 - |y_i - \hat{y}_i| \geq 0.85 \\ 0, & 1 - |y_i - \hat{y}_i| < 0.85 \end{cases} \quad (3)$$

其中 y_i 为序列中第 i 项的真值, \hat{y}_i 为序列中第 i 项的预测值。MSE 可以用来表征数据的预测程度的好坏, MSE 越小代表预测值与真值在均方误差意义下更加接近。Q 则为观测预测值符合国家标准的数量, 更直观地观测预测方法的好坏。