

# 基于 python 的风功率可视化及预测

小组成员：

杨思颖 (2020012981)

胡清茹 (2020012996)

**摘要：**随着风力发电在能源结构中具有越来越大的重要性，风功率预测也逐渐成为了研究的热点话题，较为准确的风功率预测能够为电场调峰和储能进行良好的指导作用，促进风电的使用与发展。本文针对给定的风电场数据，首先对风电场数据进行探索性分析，然后分别尝试使用深度学习和统计学的方法对风功率进行超短期预测。本文重点探索深度学习方法中不同的神经网络对于风电预测的效果，从而更深刻地理解到大数据与人工智能在能源与动力领域的重要作用。

**关键词：**风功率预测 深度学习神经网络

## 目录

0 引言 .....	2
1 探索性数据分析及预处理 .....	4
1.1 数据统计特性的描述和分析 .....	5
1.1.1 位置估计及变异性估计 .....	5
1.2 数据清洗和预处理 .....	7
1.2.1 缺失值处理 .....	7
1.2.2 数据平滑 .....	10
1.3 预处理后的数据可视化 .....	12
1.3.1 数据随时间的变化 .....	12
1.3.2 相关性探索 .....	13
1.4 特征提取 .....	14
1.4.1 统计量特征提取 .....	14
1.4.2 时间序列的去趋势特征 .....	15
1.4.3 数据的自相关和偏相关性 .....	15

1.4.4 数据的傅里叶去噪声 .....	15
1.5 探索性分析总结 .....	17
2 深度学习方法预测风功率 .....	17
2.1 人工神经网络 ANN .....	17
2.1.1 MLP 前馈神经网络 .....	17
2.1.2 BP 算法用于 MLP .....	20
2.2 循环神经网络 RNN .....	22
2.2.1 RNN .....	22
2.2.2 LSTM 长短期记忆 .....	25
2.2.3 GRU 门控循环神经网络 .....	29
2.2.4 LSTM 和 GRU 的比较 .....	32
2.3 卷积神经网络 CNN .....	34
2.3.1 一维 CNN .....	34
2.3.2 TCN 时序卷积网络 .....	37
3 统计学方法预测风功率 .....	41
3.1 ARMA 模型原理 .....	42
3.2 ARMA 预测 .....	43
4 总结 .....	45
5 参考文献 .....	45
6 课程收获与建议 .....	错误！未定义书签。
合作分工 .....	错误！未定义书签。

## 0 引言

随着世界经济的快速发展，人类对能源的需求与日俱增。化石能源是目前全球消耗的最主要能源，但因其不可再生性和人类的大量开采，正在逐步走向枯竭。截至 2020 年底，我国石油储量仅占全球的 1.5%，储采比仅有 18.2 年，煤炭和天然气的储采比均小于 50 年，

化石能源储量并不乐观。同时，化石能源的消耗也导致了二氧化碳的过量排放，进一步加剧温室效应，导致气候异常、生态系统紊乱、疾病多发等危害人类生存的严峻后果。碳减排也成为了各国共同努力的目标。我国提出碳达峰碳中和的“双碳”目标，在十四五规划中，明确大力发展战略性新兴产业，并要求大力发展战略性新兴产业，2030年风电和太阳能发电总装机容量需达1200GW以上。

风电作为技术较为成熟并且安全可靠的清洁能源，是替代传统化石能源的重点之一。其全生命周期碳排放不到火力发电的1%，且用地小、发电功率密度相对较高、设备安装灵活、发电时间持续较长，这些优点使得风电易于替代传统火电并和其他能源形成互补。但是，风电却有一大不足，由于风电完全受风力影响，因此具有明显的波动性和间歇性，导致风力发电有很强的不确定性，这对于风电并网造成了很大的困难，严重时可能会危及电网安全。<sup>1</sup>因此，如果能够成功对风功率进行较为精确的预测，电力部门就可以提前根据预测结果安排调度计划，这也就减少了风电本身不确定性带来的负面影响。因此，风电预测成为了研究的重点。

为统一预测标准，国家市场监管局和国家标准管理委员会发布了《调度侧风电或光伏功率预测系统技术要求》的国家标准，自2022年5月1日起实施，规定了调度侧风电或光伏预测系统在数据、软件、硬件和性能指标方面的技术要求。<sup>2</sup>在预测的时间尺度方面，分为超短期功率预测（预测风电场未来15min~4h的有功功率）、短期功率预测（预测零时起到未来72h的有功功率）、中期功率预测（预测零时起到未来240h的有功功率），时间分辨率皆为15min。预测性能指标为（以准确率为标准，准确率定义为1-均方根误差rmse）：超短期大于等于87%，短期大于等于83%，中期大于等于70%。

预测准确率依赖于科学的预测理论方法和合理的预测模型，主要分为模型驱动方法和物理方法。物理方法主要是根据物理信息分析风速特征，但是这类方法缺乏对历史数据的利用，精度较差。模型驱动方法可以分为四类<sup>3</sup>：第一类是概率统计模型，代表是时间序列模型；第二类是机器学习模型，比如支持向量机、随机森林、提升回归树等；第三类是深度学习模型，这种模型能够基于神经网络构造非线性函数，拥有强大的模型拟合能力，包括人工神经

<sup>1</sup> 纪代颖.基于风电的不确定性对电力系统的影响分析[J].电子测试, 2022,36(10): 99-101. DOI: 10.16520/j.cnki.1000-8519.2022.10.003.

<sup>2</sup> GB/T 40607-2021, 调度侧风电或光伏功率预测系统技术要求[S].

<sup>3</sup> 冉靖,张智刚,梁志峰,鲁宗相,陈敏.风电场风速和发电功率预测方法综述[J].数理统计与管理, 2020,39(06): 1045-1059. DOI: 10.13860/j.cnki.sltj.20200930-002.

网络（ANN），循环神经网络（RNN），卷积神经网络（CNN），长短期记忆网络（LSTM），门限循环网络（GRU）等等。其中部分模型很适合对时间序列建模，有着不错的效果。第四类是模型分解方法，考虑并区分了序列数据的线性和非线性性，分别建立模型，并将两部分的预测结果组合起来。另外，也可以综合以上四种方法使用组合预测，也就是将我们的序列数据分解为多个信号，分别采取适合的方法建模，然后再用各个预测值重构整体序列的预测值。

在本论文中，针对给定的风电场数据，笔者将首先对风电场数据进行探索性分析，然后将使用多种模型和多种方法对风功率进行预测，并比较效果，得出结论。

## 1 探索性数据分析及预处理

本节将介绍笔者对于拿到的数据进行初步描述、分析、可视化的过程，以及对于数据做清洗和预处理的过程、结果等。

拿到的数据分为三列，分别是 DateAcqTime，代表获取风电数据的时间，是从 2015 年 10 月 1 日 0:00 至 2015 年 10 月 31 日 23:59 的数据，其中有缺失，将在之后介绍；GridPower，代表风功率；wind\_Spd，代表风速。其中风功率和风速由于风电场保密要求，拿到的数据已做归一化处理。这一方面意味着在做数据预处理时，不用再次归一化输入模型，另一方面意味着我们无法将数据反归一化，因而算出来的准确率是基于归一化数据的，由于数据本身较小，故 rmse 的值也会较小，因此本文在之后的预测模型中，多注重对各个模型之间的比对，比较各个模型和不同方法间的 rmse 的相对大小，从而得出有趣的结论。原始数据格式如下图所示：

	DateAcqTime	GridPower	wind_Spd
1			
2	2015/10/1 0:00	0.38065	0.39627
3	2015/10/1 0:00	0.36943	0.39592
4	2015/10/1 0:01	0.38529	0.39538
5	2015/10/1 0:01	0.38892	0.39579
6	2015/10/1 0:02	0.4122	0.39627
7	2015/10/1 0:02	0.42277	0.39871
8	2015/10/1 0:03	0.40035	0.39998
9	2015/10/1 0:03	0.41485	0.40036
10	2015/10/1 0:04	0.41225	0.40208
11	2015/10/1 0:04	0.4024	0.4027
12	2015/10/1 0:05	0.39804	0.40327
13	2015/10/1 0:05	0.37576	0.40026
14	2015/10/1 0:06	0.39463	0.39892
15	2015/10/1 0:06	0.41636	0.39866
16	2015/10/1 0:07	0.41607	0.39913
17	2015/10/1 0:07	0.41616	0.40154
18	2015/10/1 0:08	0.40367	0.40072
19	2015/10/1 0:08	0.40117	0.40095
20	2015/10/1 0:09	0.39506	0.40008

图 1. 原始数据格式展示

## 1.1 数据统计特性的描述和分析

### 1.1.1 位置估计及变异性估计

首先考察数据最基本的位置估计，包括均值、结尾均值和中位数，结果如下表所示：

	均值	截尾均值	中位数
GridPower	0.327315	0.290137	0.229580
WindSpeed	0.362716	0.347014	0.342385

表 1. 位置估计

接着考察变异性估计，观察数值之间的差异性，计算标准差、中位数绝对偏差 MAD、四分位距 IQR、极差 Range、峰度、偏度，总结如下表所示：

	标准差	MAD	IQR	极差	峰度	偏度
GridPower	0.287181	0.235181	0.410835	0.999715	-0.307591	0.923892
WindSpeed	0.137817	0.121047	0.165935	0.877830	2.859494	1.392202

表 2. 变异性估计

### 1.1.2 数据分布可视化

接下来，绘制几种能够描述数据分布的图形。首先绘制箱型图，可以看出，风速的离群值较多，而 GridPower 的数据分布范围较广。

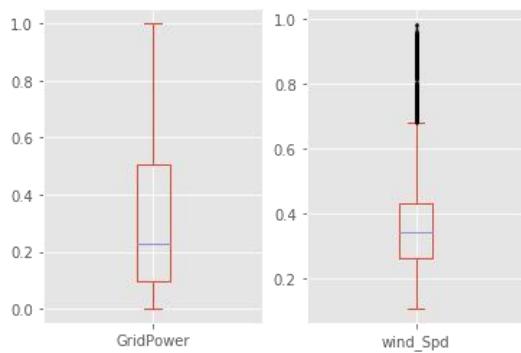


图 2. 箱型图

接着绘制小提琴图，可以明显地观察到，风功率在 0.1 左右有明显的峰值。

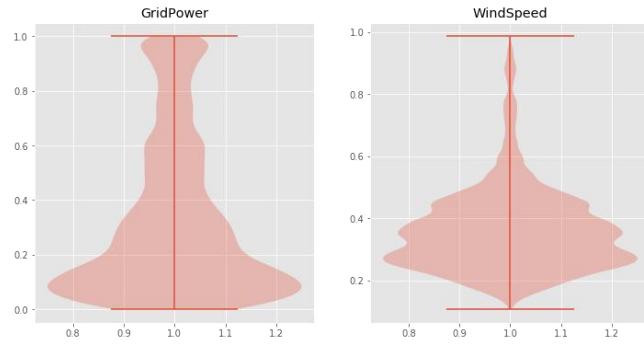


图 3. 小提琴图

接着分别绘制每天的小提琴图，可以看出天与天之间的数据分布存在较大的区别。

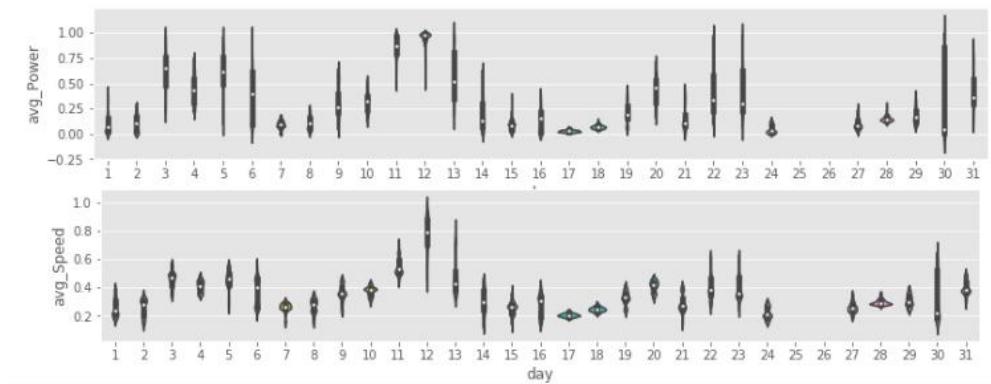


图 4. 小提琴图（按日期绘制）

然后绘制带有概率密度分布的频率直方图，可以发现风速基本呈现一个有点偏移的正态分布，而风功率的分布就不太均匀了，最左侧有一个峰，而在右侧尾部又呈现出一个小峰。

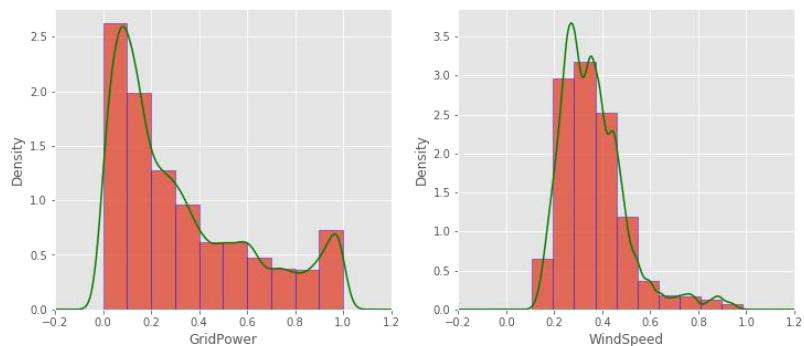


图 5. 频率直方图

## 1.2 数据清洗和预处理

首先对日期进行预处理，将日期格式拆分为年月日时分秒分别列入表格之中，并附加一列 HH:MM:SS。这可以更加便于我们定位哪些时间的数据是缺失的。

### 1.2.1 缺失值处理

下一步我们需要定位缺失值并且预留为 NAN 空值。首先展示原数据的缺失情况，如下图所示：

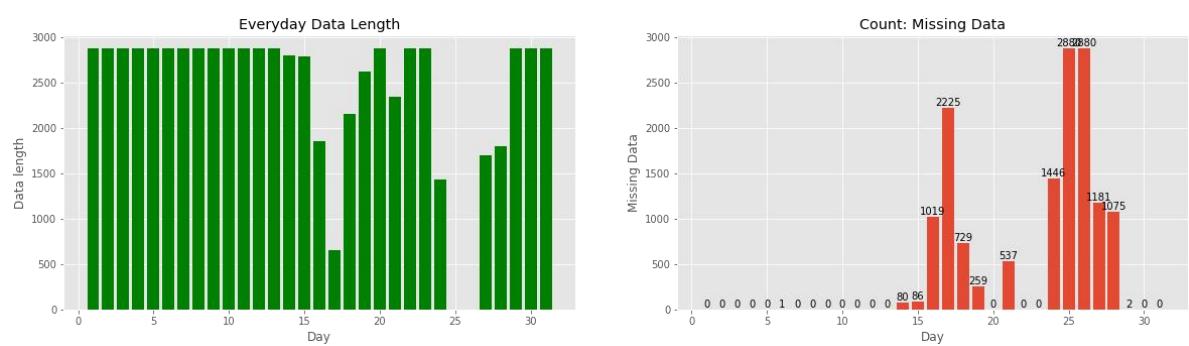


图 6. 缺失值按日期的分布

可以看到，存在数据缺失的日期为：6、14-19、21、24-29，共计 14 天。其中 6 日和 29 日仅仅缺失个别数据，25 日和 26 日的数据全部缺失，其余几日的缺失有的缺少了一半左右，有的缺少了几十个到几百个。因此必须对缺失值进行处理，否则会影响模型的输入。

首先为缺失的位置留出空行，便于填充，接着我们使用两种方法分别尝试并观察对缺失值填充的效果，采取的验证填充效果的方式是：在完整的日期中自行随机产生缺失点，然后采用方法填充，并与删除前的完整数据做比对，由此观察填补效果。最终决定到底是使用哪种填充方式。

第一，尝试 Forward Fill 前向填充，这种方法即是用前面的一个数据来填充缺失空位。考虑到我们的风电场数据中每分钟有 2 次采集点，因此认为紧连的两个数据间的差异不会很大，因此这种填充方法应当误差不会太大。

选取10/1(数据完整), 构造缺失点, 利用ForwardFill填充并与真实值比较的结果

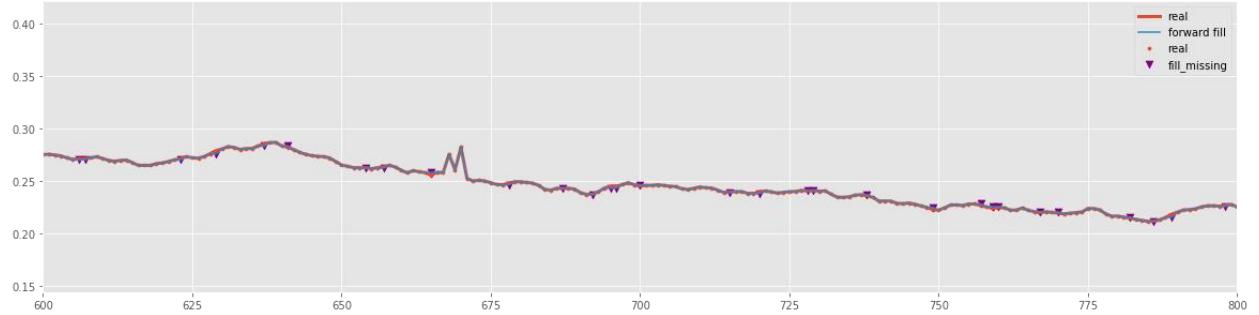


图 7. 前向填充填补缺失值的效果

于是用前向填充法, 对 31 天所有的缺失值进行填充, 效果如下图所示。

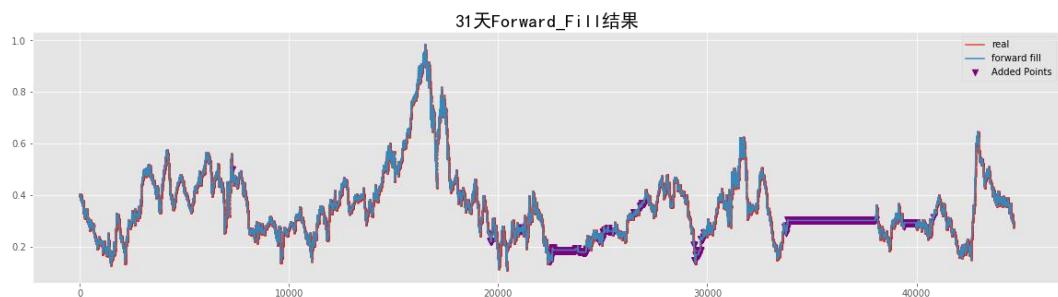


图 8. 前向填充 31 天缺失值的效果

可以看出, 对于离散的数据点缺失, 前向填充的方法还是较为精确的, 但是对于连续的缺失, 这种方法就会出现一条横线的状况, 现在观察 10 月 17 日的填充效果。

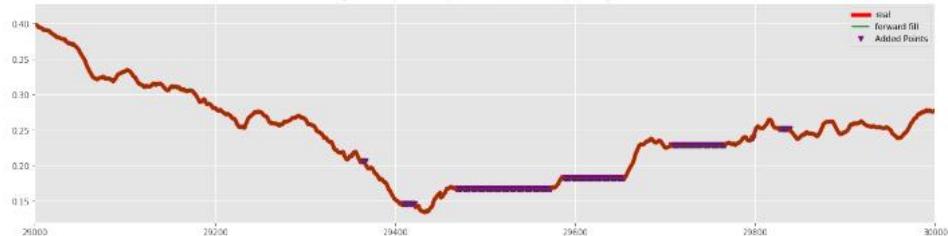


图 9. 前向填充 10/17 日缺失值效果

可以看到中间连续缺失的地方在填补后成为了一条直线, 放大观察之后横线的情况看起来更为明显:



图 10. 前向填充连续缺失值效果放大

因此得到结论, 对于大段连续的缺失值, 这种方法并不适合于填充。

第二，尝试另一种方法，使用 Moving Average 移动平均法。核心思想是取出缺失值发生之前的一段滚动时间内的值，计算其平均值或中位数来填补缺失。在有些场景下，这种方法会比前向填充效果更好，例如数据的噪声很大，对于单个数据点有很大的波动，但用移动平均的方法就可以弱化这些噪声。使用这种方法的填充效果如下图所示。

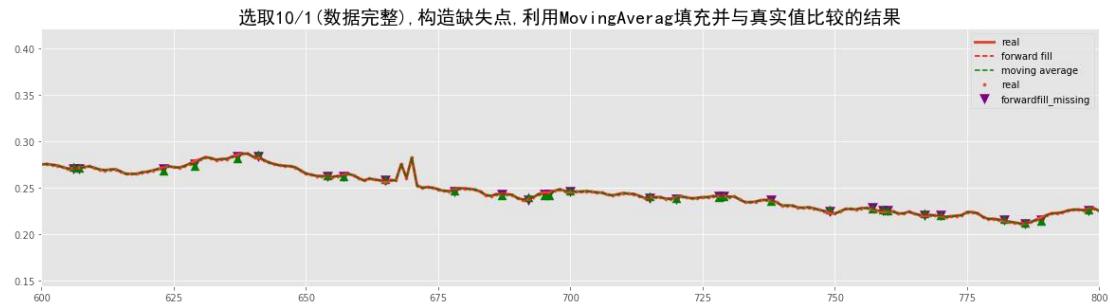


图 11. 移动平均填充缺失值效果

由此，对于这两种方法，我们认为他们对零散的缺失值的填充效果都比较好，但是对于大段的缺失，填充效果都不太好。因此我们决定使用如下的缺失值填补方法：

当连续缺失数据少于 60 条时，采用移动平均法填充缺失值；当连续缺失的数据大于 60 条时，令这一段数据空着不填充，未来也不输入模型。

以分钟为标准衡量缺失值，原来的缺失分钟数为 7235，按这种方法填充后，缺失分钟数减少为 6919，且其中还包括 25 和 26 日完全缺失的两天。画图比较填充前后的效果如下图所示：

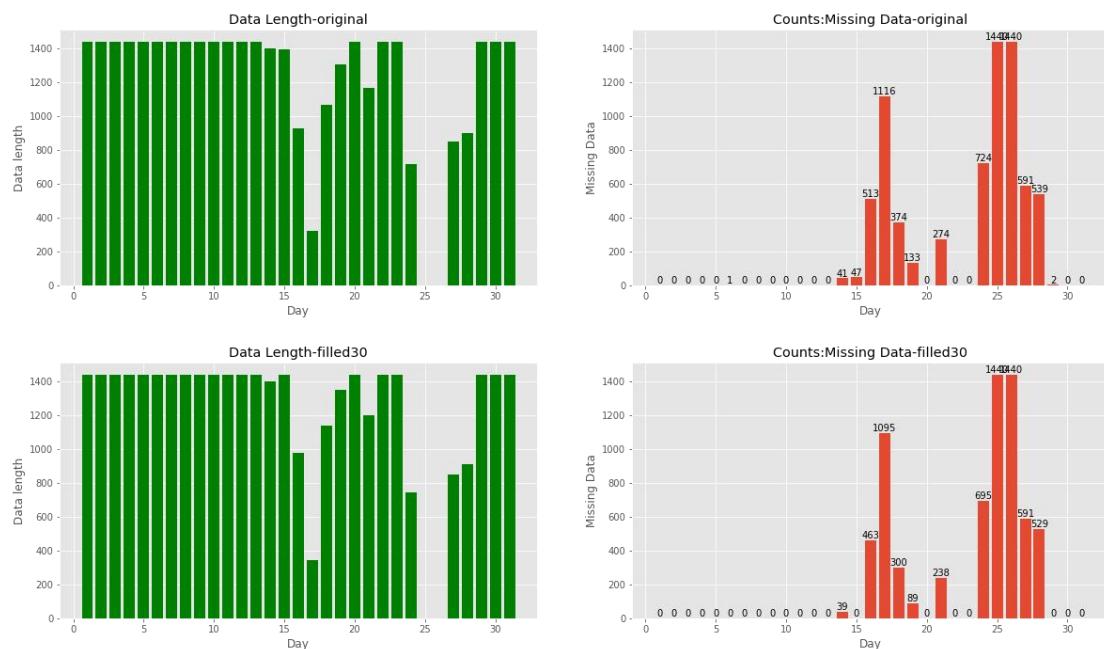


图 12. 填充缺失值前后缺失的分钟数对比

## 1.2.2 数据平滑

首先进行平稳性检验。使用 ADF Test，得到如下结果， $p$  值  $> 0.05$ ，故序列不平稳。

```
*****  
ADF Statistic of speed: -2.156373243644567; p-value: 0.22245854612670973  
Critic Values:  
1%, -3.430699593622523  
Critic Values:  
5%, -2.8616945075066655  
Critic Values:  
10%, -2.5668522402260887  
*****  
ADF Statistic of Power: -2.44921142239032; p-value: 0.1283256264396535  
Critic Values:  
1%, -3.4307002302198053  
Critic Values:  
5%, -2.861694788842625  
Critic Values:  
10%, -2.566852389976601
```

图 13. ADF Test 平稳性检验结果

继续使用 KPSS Test，得到结果如下图所示， $p$  值  $< 0.05$ ，故序列不平稳。

```
*****  
KPSS Statistic of Speed: 6.5581799605; p-value: 0.010  
Critic Values:  
10%, 0.347  
Critic Values:  
5%, 0.463  
Critic Values:  
2.5%, 0.574  
Critic Values:  
1%, 0.739  
*****  
KPSS Statistic of Power: 5.0830161736; p-value: 0.010  
Critic Values:  
10%, 0.347  
Critic Values:  
5%, 0.463  
Critic Values:  
2.5%, 0.574  
Critic Values:  
1%, 0.739
```

图 14. KPSS Test 平稳性检验结果

接着进行白噪声检验，使用 Ljung–Box Test，得到结果如下表所示， $p < 0.05$ ，不是白噪声。

	lb_stat	lb_pvalue
GridPower	18697.207	0.0
WindSpeed	18717.570	0.0

表 3. Ljung–Box 白噪声检验

由此我们验证了数据是不平稳的，并且不是白噪声。下一步对数据进行平滑。

第一使用移动平均法平滑，得到如下结果：

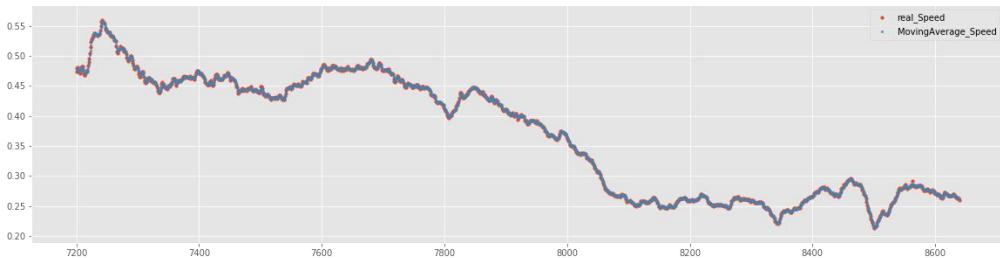


图 15. 移动平均法平滑效果

第二使用 Loess 平滑，结果如下图所示，放大每天的效果观察也可以发现效果较好。

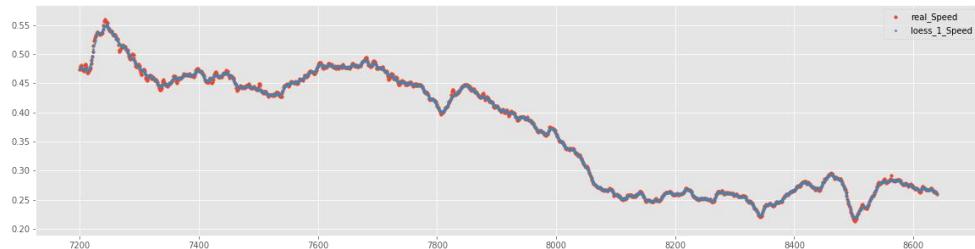


图 16. Loess 平滑效果

最终对于平滑数据的需求，我们采用 Loess 平滑，对于每天分别平滑，然后拼接成完整的一个月。风速和风功率的 Loess 平滑后效果如下图所示。

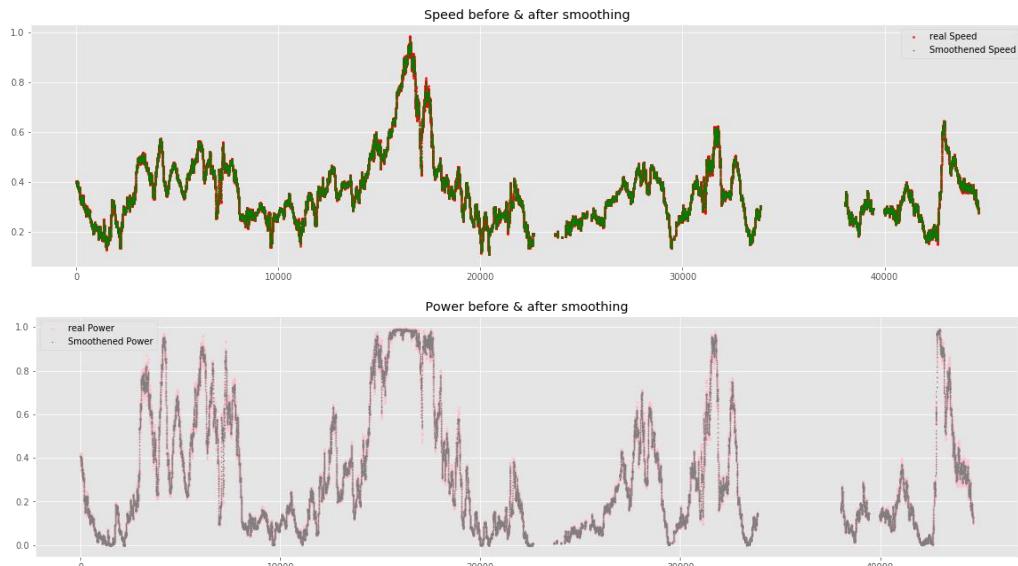


图 17. 对全月进行 Loess 平滑的效果

需要注意的是，在之后输入模型的时候，由于担心平滑后的数据会缺失一些信息，因此我们仍旧使用的是平滑前的数据，但在未来继续探索时可能会尝试使用平滑后的数据。

## 1.3 预处理后的数据可视化

在初步的预处理和缺失值填充之后，我们的数据相较于最初已经完整许多了，因此可以在此基础上更多地进行数据可视化。

### 1.3.1 数据随时间的变化

风功率和风速在 31 天中随时间的变化如下图所示：

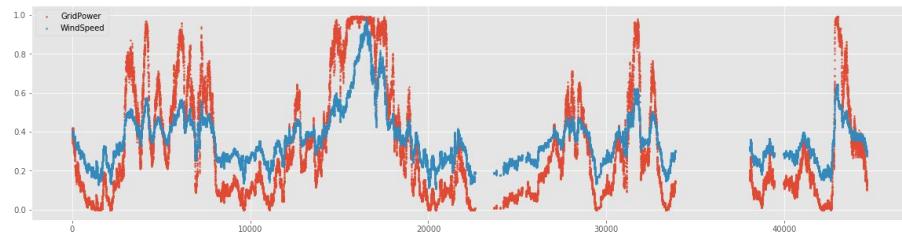


图 18. 风功率和风速随时间变化（全月）

按每天分别画图，如下图所示。可以发现二者的变化趋势是非常非常相近的，有紧密的联系，从直观上也很好理解，因为风速是直接决定风功率的，并且其他决定风功率的因素非常少，基本完全由风速决定。

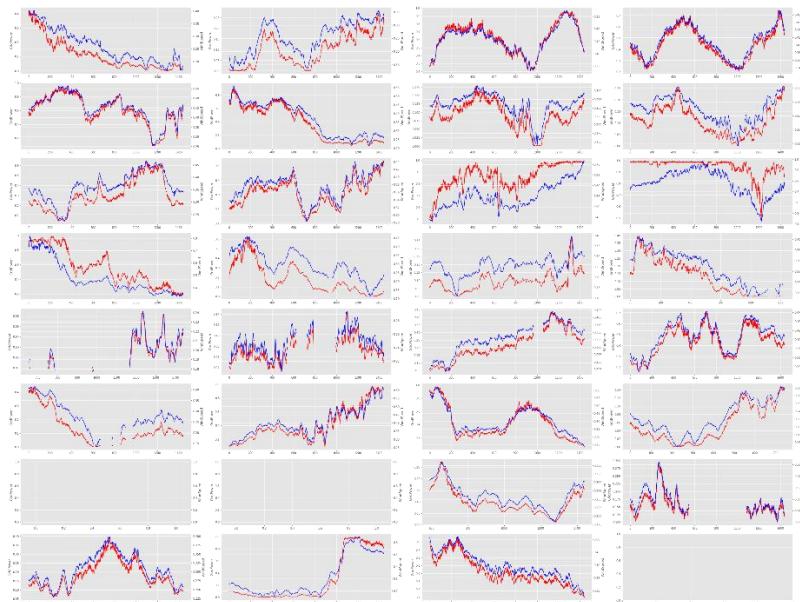


图 19. 风功率和风速随时间变化（按日期）

从趋势相近，则说明二者的相关度很高，由此进行相关性探索。

### 1.3.2 相关性探索

对风速和风功率进行相关性探索，首先绘制散点图。可以看出二者拟合的很好，有非常大的相关性，在中间端甚至呈现比较标准的线性，在风速为 0.6-0.8 的范围内，二者的分布稍微有点散，但是在前段和后段可以说是拟合的非常好。因此这也启示我们，是否可以先预测风速，然后根据这个拟合情况由此预测风功率，这是一种可行的预测方法。

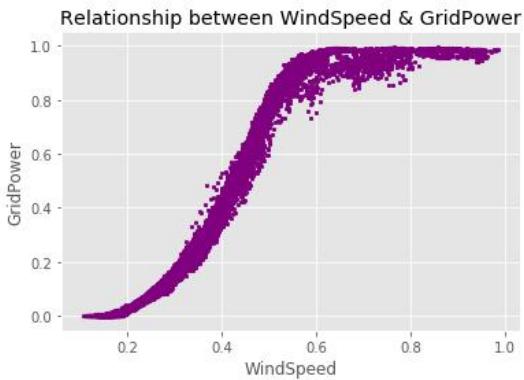


图 20. 风速风功率相关性

绘制其他带有更多信息的相关性图如下图所示。

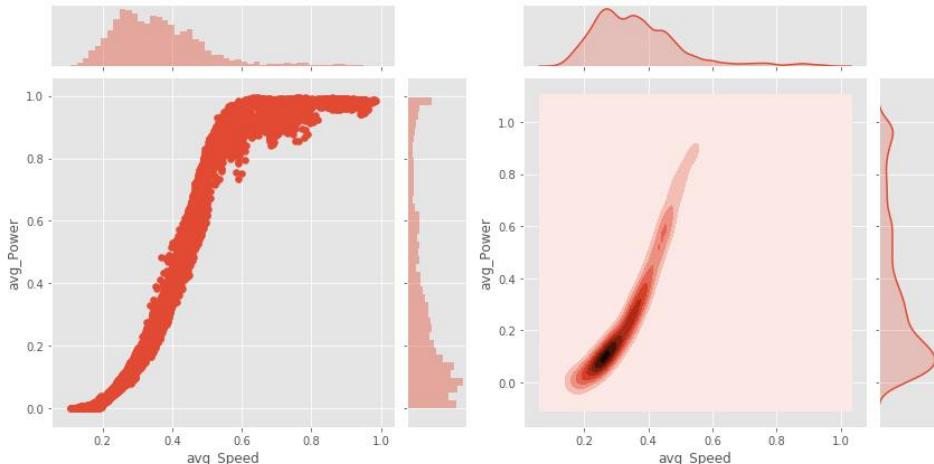


图 21. 风速风功率相关性进一步探索

接着绘制相关性热力图，几个参数包括日期、一天中的时间、风速、风功率，得到结果如下图所示。可以看出，日期和风速、风功率无关，时间和风速风功率略相关，风速和风功率二者紧密相关。

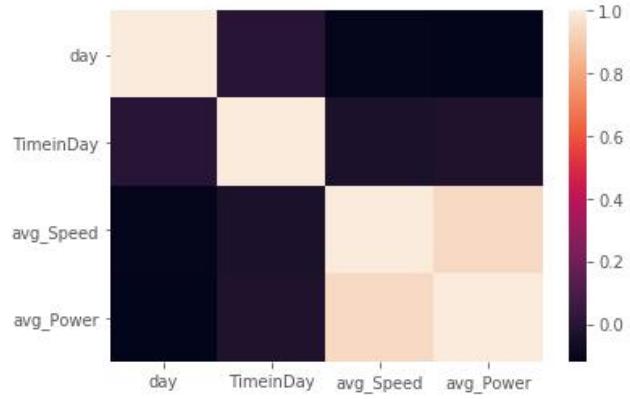


图 22. 相关性热力图

## 1.4 特征提取

### 1.4.1 统计量特征提取

为了从数据中提取更多信息，便于作为特征未来输入模型，首先提取典型的统计量特征。对风速和风功率分别提取以下统计量：相比于前 1min 的变化百分比、相比于前 5min 的变化百分比、该时间点的前 5min 的数据的均值、前 5min 的标准差、前 5min 的最大值、前 5min 的最小值。部分统计量特征随时间的变化如下图所示：

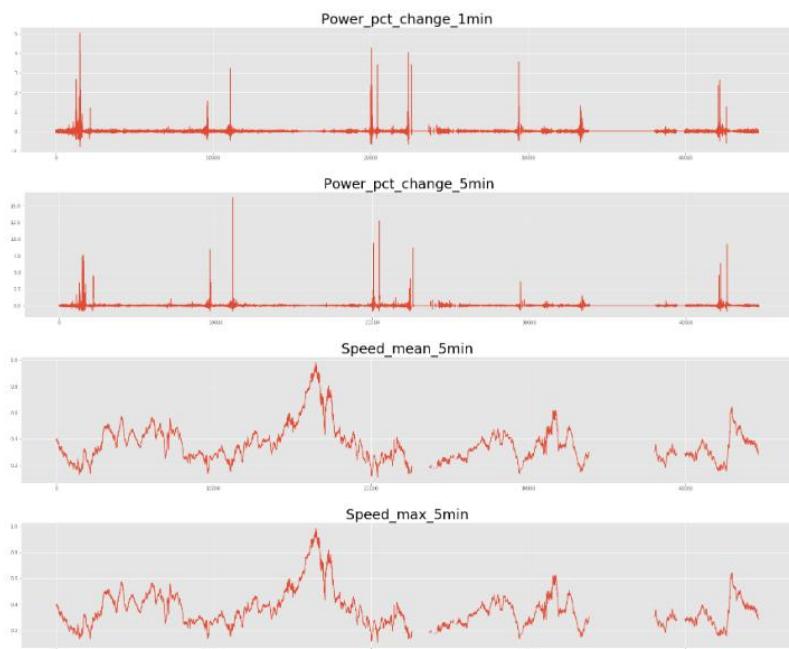


图 23. 统计量特征随时间的变化

### 1.4.2 时间序列的去趋势特征

选取前 13 天完整的数据，分别将风速和风功率分解为趋势（Trend）、季节性（Seasonality，我们将 periods 取为 1 天）、随机误差（Residual），并提取出数据。经过分解后，各部分随时间的变化如下图所示。可以看出，是可以拆分成明显的趋势的，尤其是季节性，我们是以天为季节的，从图上可以看出明显的周期性。

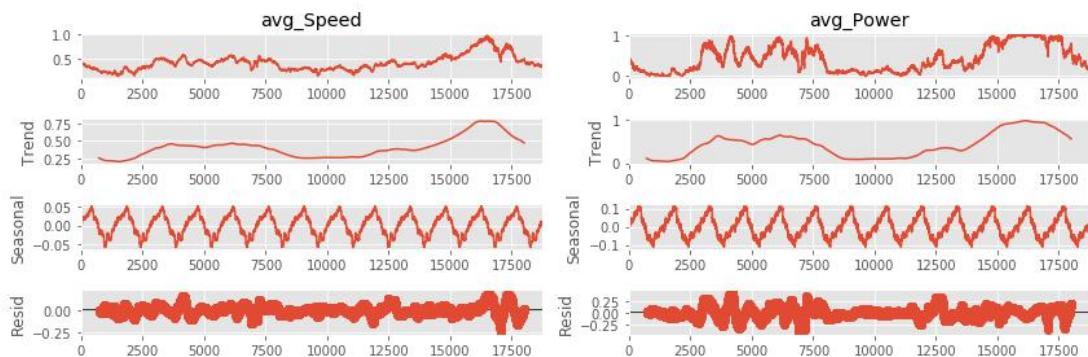


图 24. 时间序列的去趋势特征

### 1.4.3 数据的自相关和偏相关性

绘制数据的自相关和偏相关性如下图所示。

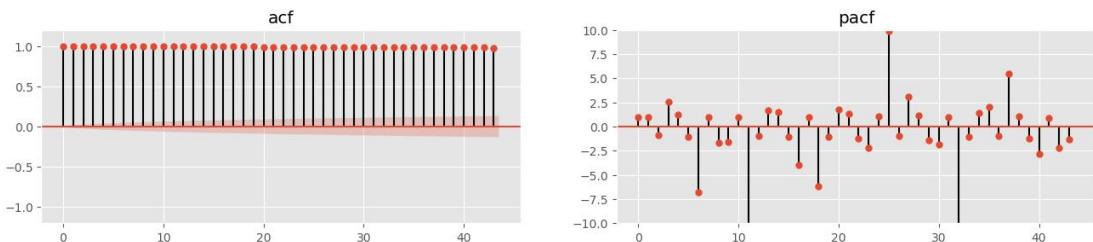


图 25. 数据的自相关和偏相关性

### 1.4.4 数据的傅里叶去噪声

对风速和风功率分别进行傅里叶去噪声，效果如下图所示。

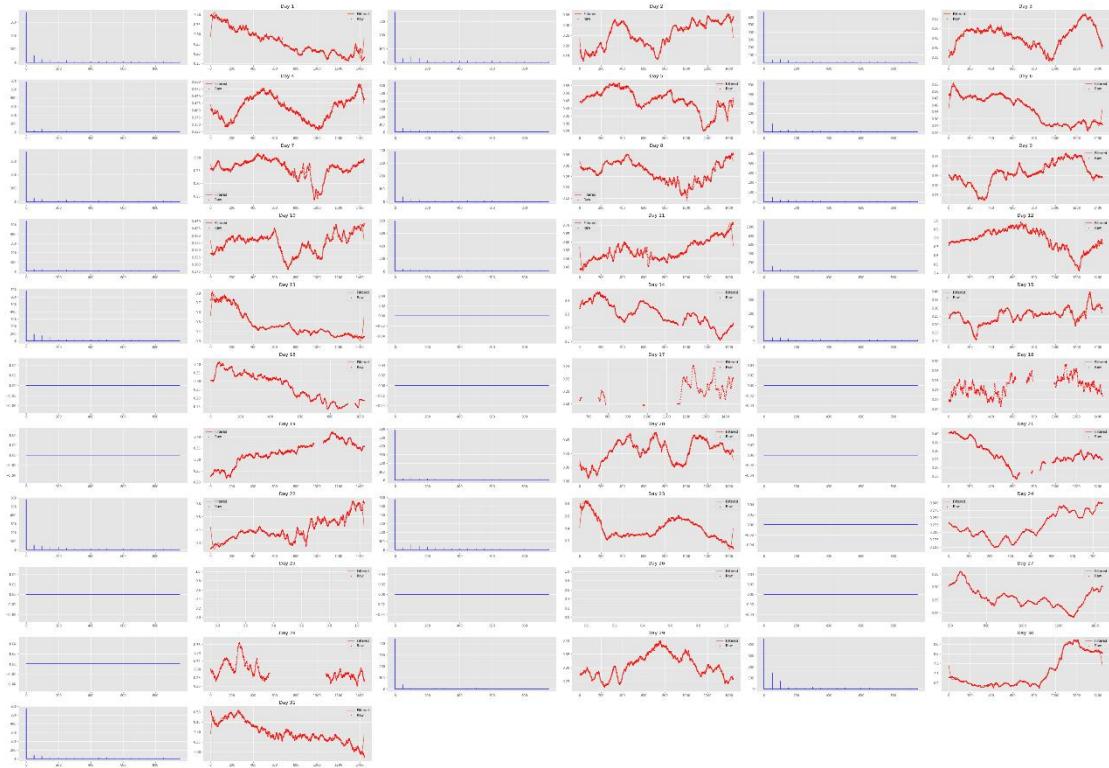


图 26. 风速的傅里叶去噪声结果

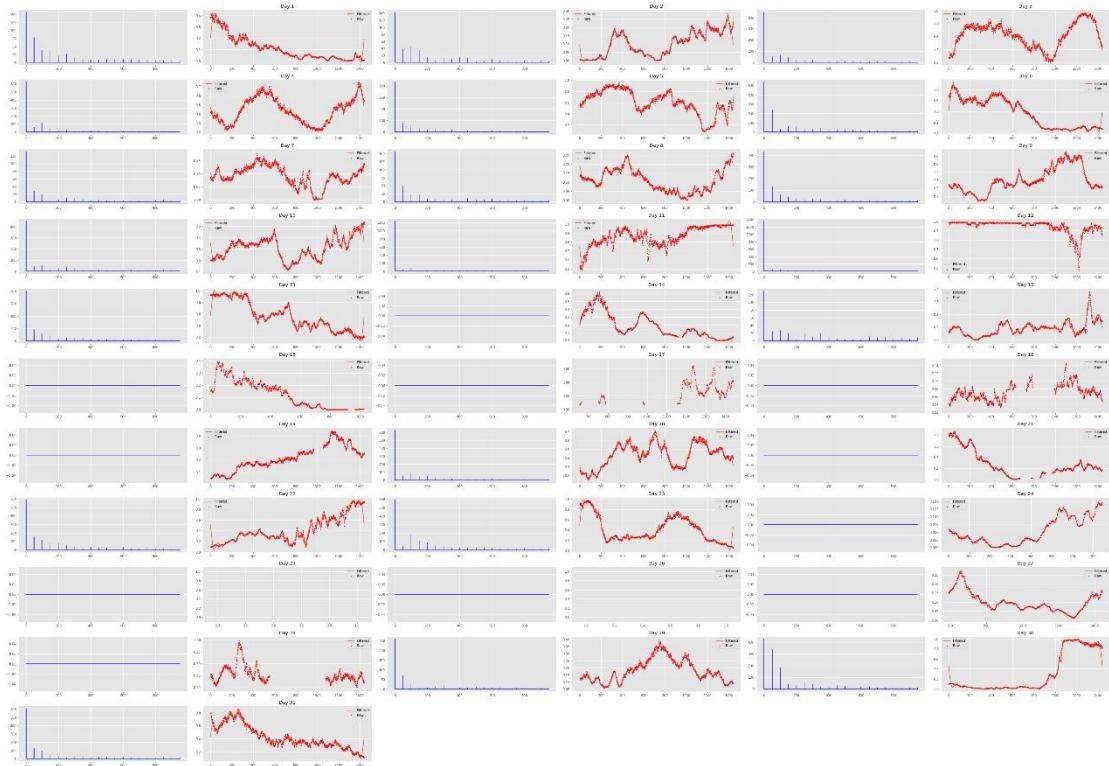


图 27. 风功率的傅里叶去噪声

## 1.5 探索性分析总结

本节我们对数据做了初步分析，观察了其统计特性，包括位置估计和变异性估计；接着对数据分布进行了可视化；然后对数据做了清洗和预处理，填补了缺失值，对数据做了平滑等；之后对预处理后的数据进行了可视化，展示随时间变化、相关性等；接着对数据做了特征提取，提取了统计量特征，观察了其变化趋势，并且提取了时间序列的去趋势特征、自相关性和偏相关性、傅里叶分析等。在进行这些工作的同时，我们保存了好几版数据，便于之后输入模型的时候选取最合适的数据。

## 2 深度学习方法预测风功率

首先介绍深度学习方向的方法。在深度学习中，有三种重要类型的神经网络，它们构成了深度学习中大多数预训练模型的基础，这三种类型分别为：人工神经网络（ANN）、循环神经网络（RNN）和卷积神经网络（CNN）。

神经网络是对生物神经元的模拟和简化，生物神经元由树突、细胞体、轴突等部分组成。树突是细胞体的输入端，其接受四周的神经冲动；轴突是细胞体的输出端，其发挥传递神经冲动给其他神经元的作用，生物神经元具有兴奋和抑制两种状态，当接受的刺激高于一定阈值时，则会进入兴奋状态并将神经冲动由轴突传出，反之则没有神经冲动。神经网络的结构也是以此为基础搭建的。接下来将介绍不同的神经网络模型进行超短期风功率预测的结果。

### 2.1 人工神经网络 ANN

#### 2.1.1 MLP 前馈神经网络

人工神经网络（ANN）中最典型基础的就是 MLP（多层感知器），也叫前馈神经网络，是最早发明的人工神经网络之一。MLP 就是由生物神经元而产生的。最典型的 MLP 包括三层：输入层、隐层和输出层，MLP 神经网络不同层之间是全连接的（上一层的任何一个神经元与下一层的所有神经元都有连接），其最大的特点就是信息只进行单向传输，没有循环或回路。神经网络有三个重要要素：权重（weights）、偏置（bias）和激活函数（activation）。其

中权重表示神经元之间的连接强度，表示可能性大小；偏置能够保证输出值不被随便激活；激活函数其非线性映射作用，可将神经元的输出幅度限制在一定范围内。最常见的是 ReLU 和 Sigmoid 激活函数。

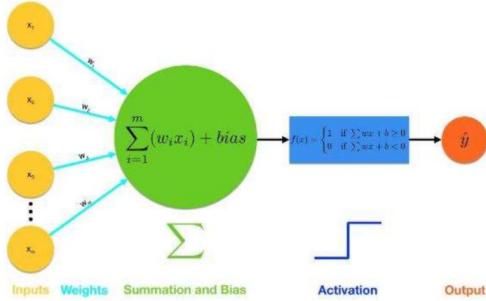


图 28. MLP 基本结构

首先尝试搭建了四层的 MLP，每层的神经元个数依次为 500, 500, 250, 250，进行训练。使用前 40 个时刻的风功率预测后一个时刻的风功率，由于预测时间很短，故准确率非常高，准确率高达 98.2%。预测结果和实际的比较见下图所示。

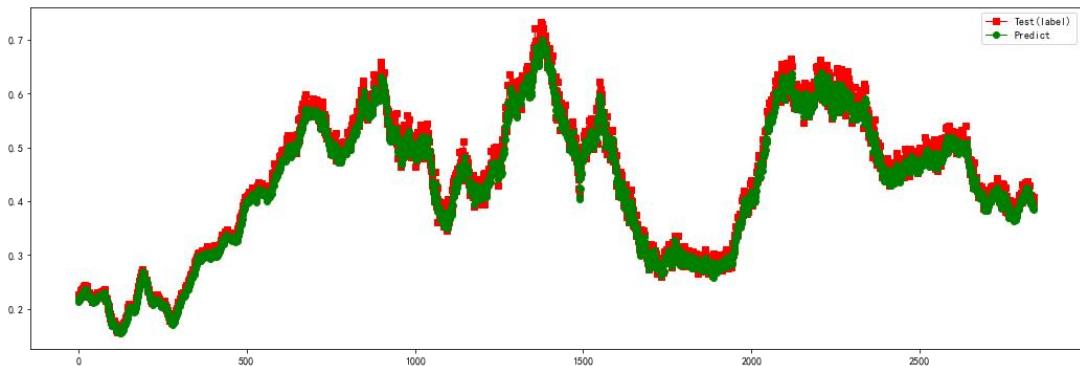


图 29. MLP 预测后一时刻结果

接下来进一步尝试，仍然只使用风功率数据，使用前 60min 预测后 30min 的风功率。得到的结果如下图所示。可以看出能够预测出大致趋势，且预测前面的时刻时的准确率较高，预测越到后面，精确率越低，整体精确率大概在 96% 左右。这也引导我们思考，或许可以将模型预测出来的结果再输入模型，接着预测下一时刻的结果，形成一个滚动的预测，这样精度应该会有所提升，但不排除预测出来的数据的误差导致误差一步步加大，在之后的模型和方法中我们将会尝试这种方法。

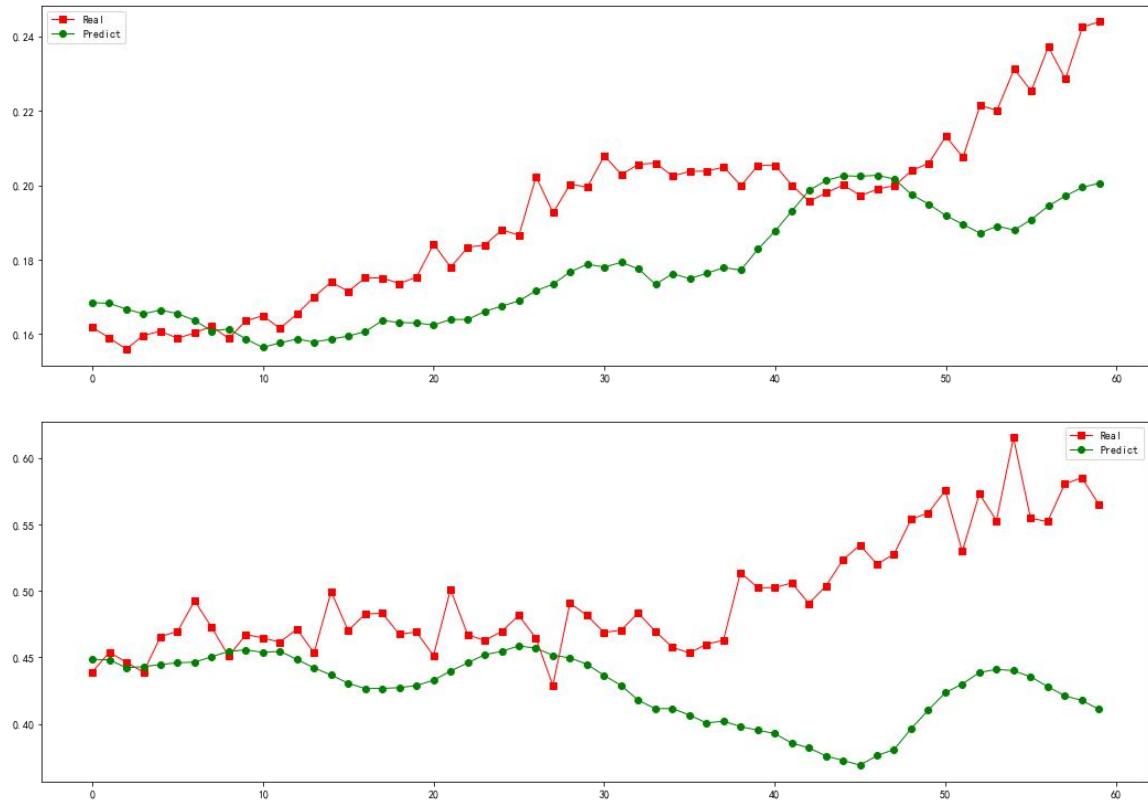


图 30. MLP 预测后 30min 的结果举例

接下来，加入风速的数据，仍然是使用前 60min 的数据预测后 30min 的数据，发现加入风速数据之后，预测的效果看似有所改善，尤其是在后半部分的偏差上，稍微减小了一点，但仍然是在预测的后时段的效果仍然不是很理想。另外，注意到有些真实数据的波动看似是在预测中被放大了，体现在图中就发现部分预测样例做出来的预测的波动比真实值要大。

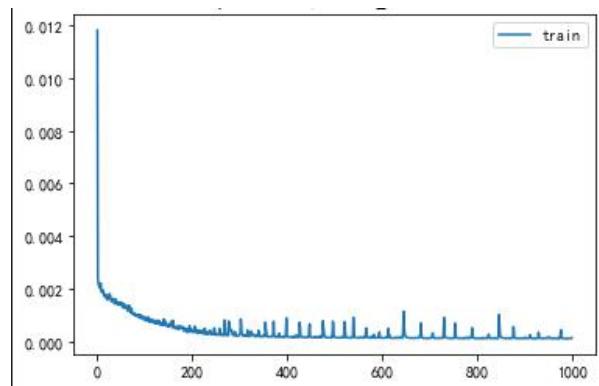


图 31. 训练 loss 随迭代次数增多的变化

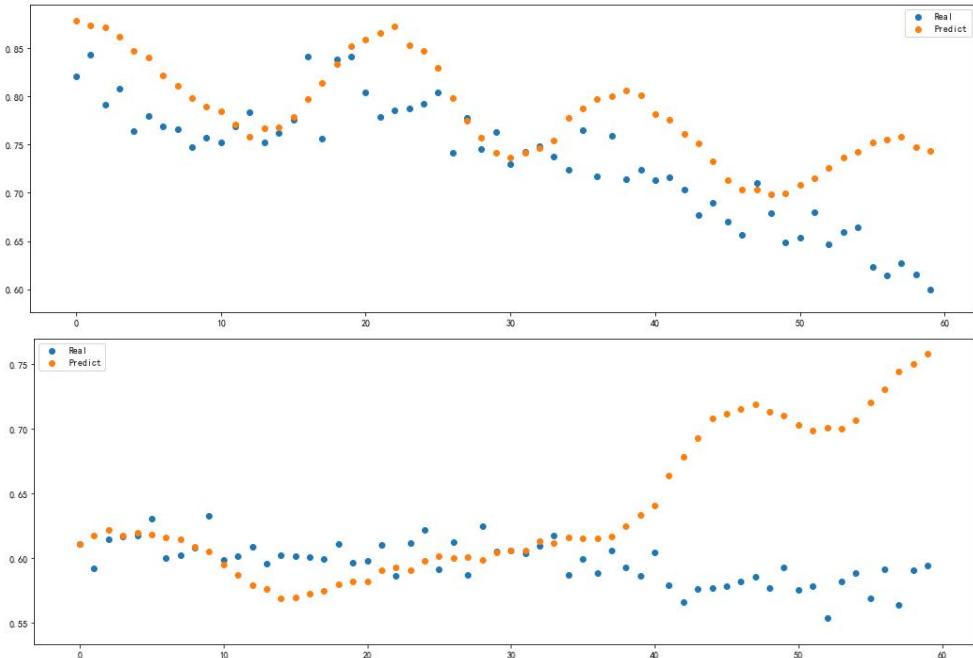


图 32. 加入风速数据后 MLP 预测 30min 后的结果举例

### 2.1.2 BP 算法用于 MLP

接着使用 BP 算法对于 MLP 进行训练。之前所用的 MLP 都是从输入层到输出层的前向传递的过程，成为前向传播。而 BP 算法的核心思想是使用梯度下降来搜索可能的权向量的假设空间，以找到最佳的拟合样例的权向量。具体而言，即利用损失函数，每次向损失函数负梯度方向移动，直到损失函数取得最小值。也就是说，反向传播算法，是根据损失函数，求出损失函数关于每一层的权值及偏置项的偏导数，也称为梯度，用该值更新初始的权值和偏置项，一直更新到损失函数取得最小值或是设置的迭代次数完成为止。以此来计算神经网络中的最佳的参数。

使用 BP 算法用于 MLP，但是感觉效果不是很好，可能是因为这个手动编写的算法存在一定的问题。

下图是训练集预测出来的结果：可以看出对于训练的 rmse 已经很好了，达到了 rmse 为 6% 的结果。

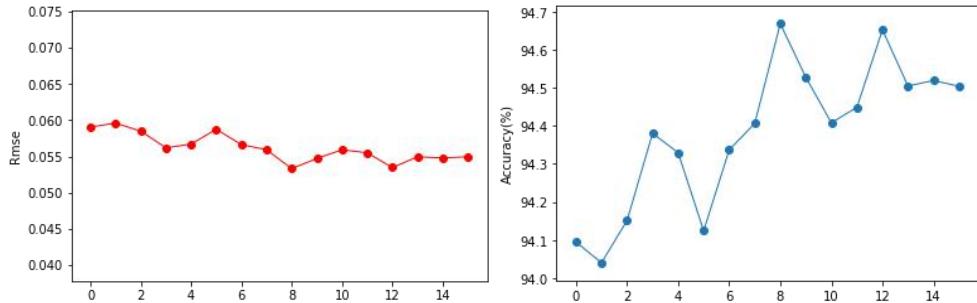


图 33. BP 算法训练集的 rmse 和准确率

但是测试集的结果还是不太好，考虑可能是过拟合了，但是最终也没能成功调出来一个好的结果。测试集的结果如下图所示，随时间增加，准确率大幅下降，最终的准确率只有 55% 左右，虽然一开始的准确率也不高，只有 80% 左右，平均的准确率仅有 61%。这个结果不太好，但还是呈现在这里。

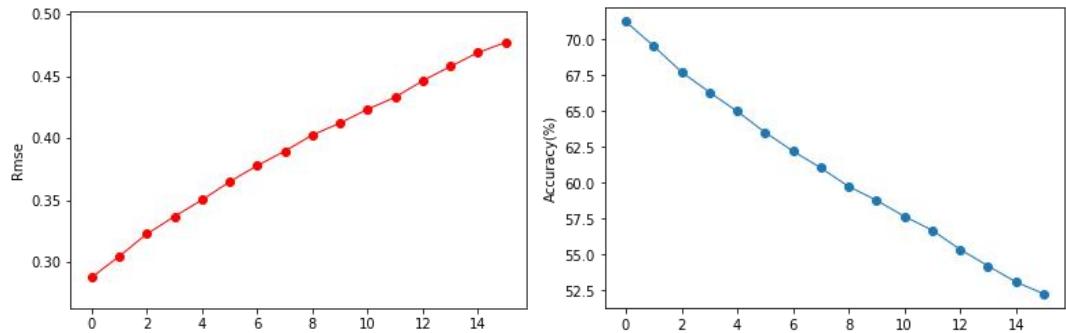


图 34. BP 算法测试集的 rmse 和准确率

这是训练集中的预测结果，预测至少是有一定的波动的。

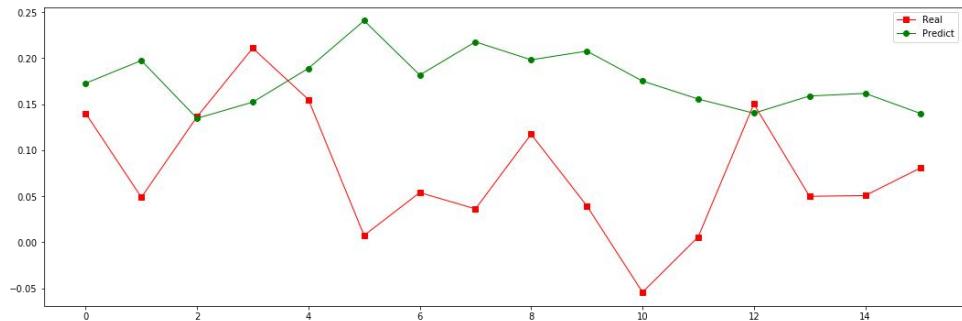


图 35. BP 算法训练集的预测结果

而这是测试集中的结果，发现预测结果几乎不波动，效果不好。

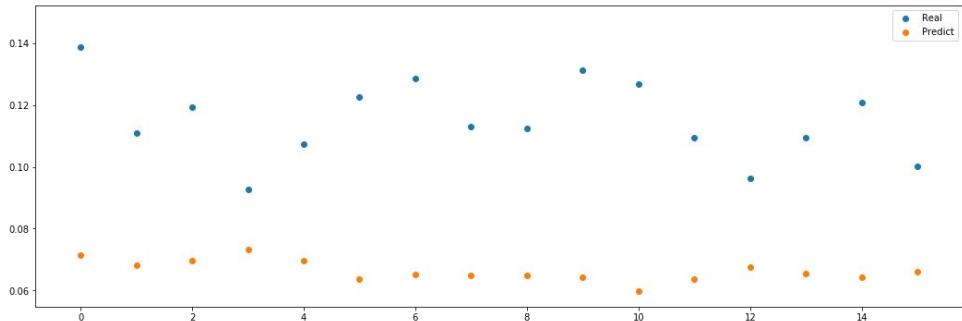


图 36. BP 算法测试集的预测结果

## 2.2 循环神经网络 RNN

### 2.2.1 RNN

对于单一的 MLP 来说，它只能单独地去处理一个个输入，前一个输入和后一个输入没有关系。但是我们的数据是序列信息的，前面的输入和后面的输入是有关联的，因此循环神经网络就因此诞生了，其本质是在 MLP 的基础上加入了循环，使得历史的序列信息被输入模型。最基础的 SimpleRNN 的结构如下图所示：

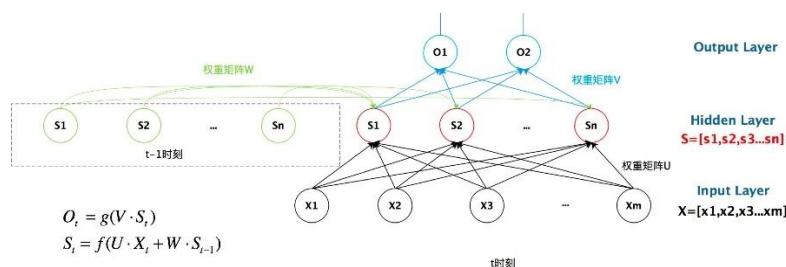


图 37. RNN 结构

首先输入向量  $x$ ,  $s$  为隐藏层的值,  $U$  是输入层到隐藏层的权重矩阵,  $o$  也是一个向量, 它表示输出层的值;  $V$  是隐藏层到输出层的权重矩阵。 $W$  是比较特殊的权重矩阵, 因为循环神经网络 RNN 是需要输入历史的信息的, 则其隐藏层的值  $s$  不仅仅取决于本次输入的  $x$ , 而且还取决于上一次隐藏层的值  $s$ , 权重矩阵就是隐藏层上一次的值作为这一次输入的权重, 按时间线展开可以表现为下图:

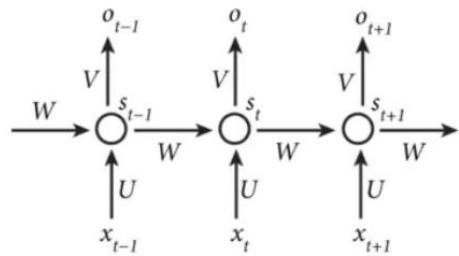


图 38. RNN 按时间展开表示

RNN 在  $t$  时刻接受输入  $x_t$  之后，隐藏层的值是  $s_t$ ，输出值是  $o_t$ ，其中  $s_t$  不仅取决于本时刻的输入  $x_t$ ，还取决于上一时刻的  $s_{t-1}$ ，公式可以写为：

$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$

### 1) 使用 RNN 预测未来 30min (60 个时间点)

首先进行非常短期的预测，观察模型效果。预测目标为预测未来的 60 个时间点，也即 30 分钟。输入模型的数据为风速和风功率。搭建循环神经网络，不断调整参数，观察训练结果。训练集和验证集的 loss 函数随着迭代次数的增多，不断减小，如下图所示。

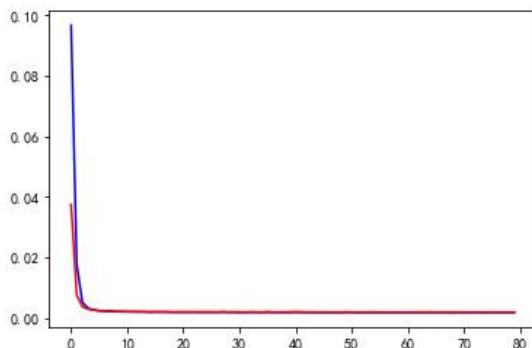


图 39. RNN 训练 loss 函数与迭代次数关系图

接着，使用未被输入模型的数据验证模型的效果，计算损失函数 rmse 和准确率，得到结果如下图所示。可以发现，随着预测的时间越来越长，预测的效果也就越来越差，但是整体的准确率还是非常不错的，在第 60 个时间点时，rmse 误差也只有 3.5% 左右，准确率仍然高达 96.5%，效果是不错的。

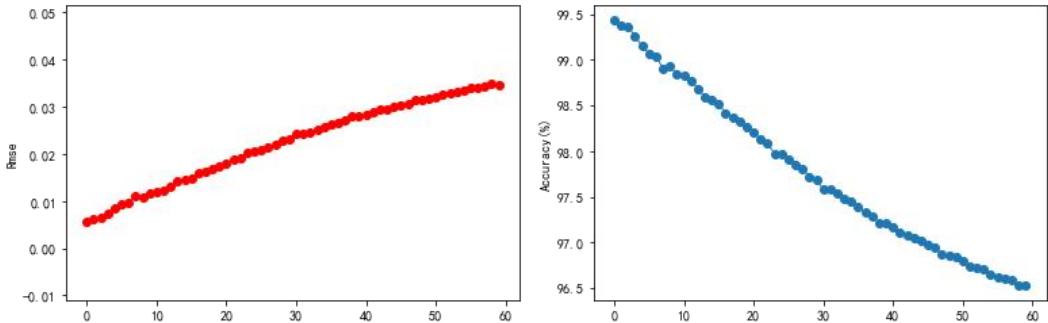


图 40. 误差和准确率随预测时间的变化

选取一个验证效果进行展示，也可以看到随着时间增长，预测和真实值的差距越来越大了。

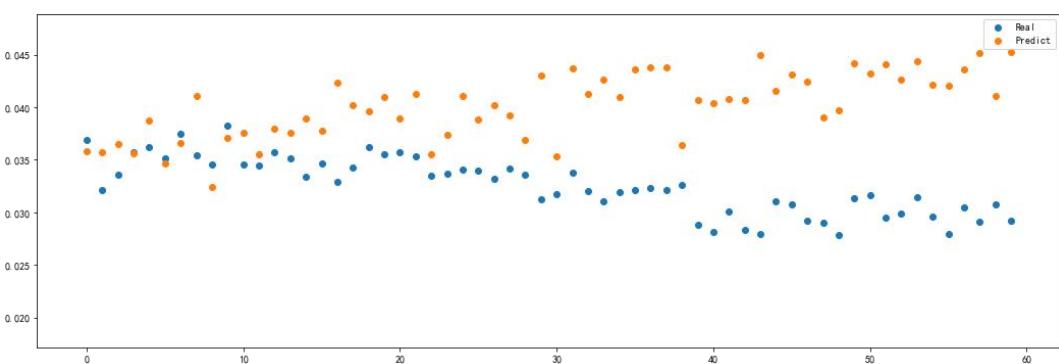


图 41. RNN 预测后 60 个时间点效果

## 2) 使用 RNN 预测未来 4 小时

接着，尝试使用 RNN 预测未来 4 个小时的风功率效果。按照时间分辨率为 15min 进行预测，也即预测未来 16 个点的风功率，每 15min 是一个点。经过参数的调节，训练过程中的训练集和验证集的损失函数随迭代而不断减小。

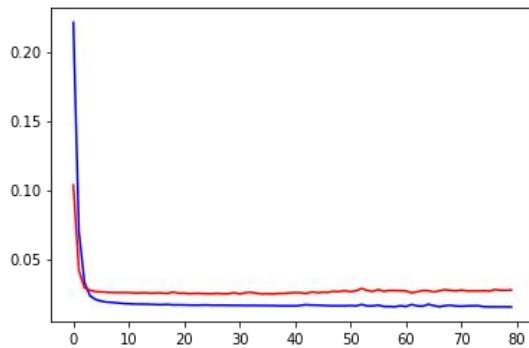


图 42. 损失函数与迭代次数关系图

使用没有输入模型的独立数据来做测试，可以发现和之前预测后 60 个时刻的规律一致，随着预测时间的增长，误差不断增大，准确率随之下降。可以看到，最后一个时刻的准确率也

达到了 80% 左右，最开始一个小时的准确率都在 90% 以上。整体来看，这 4h 内的平均准确率达到了 87%。

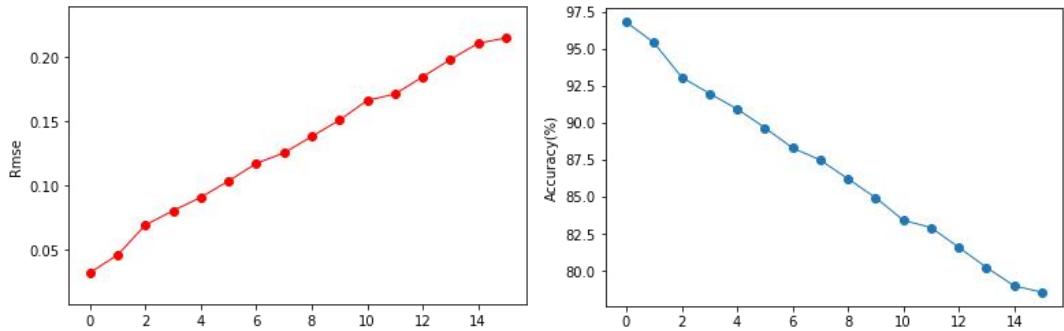


图 43. 误差和准确率随预测时间的变化

选取一个例子，绘制真实风功率和预测值于一张图中，结果如下图所示：

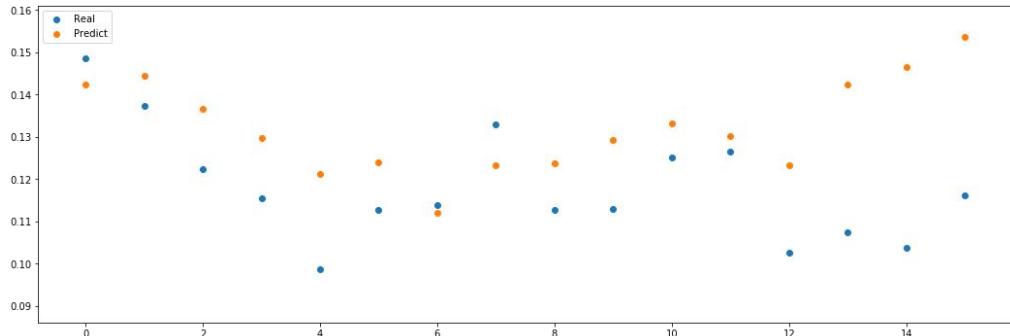


图 44. 真实风功率和预测值比较（预测未来 4h）

## 2. 2. 2 LSTM 长短期记忆

长短期记忆 (Long short-term memory, LSTM) 是一种特殊的循环神经网络 (RNN)，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。简单来说，就是相比普通的 RNN，LSTM 能够在更长的序列中有更好的表现。

### 1) LSTM 网络工作原理

LSTM 将信息存储在一个个记忆细胞中，不同隐藏层的记忆细胞之间通过少量线性交互形成一条传送带实现信息的流动；同时引入“遗忘门”、“记忆门”和“输出门”，用来新增或删除记忆细胞中的信息，控制信息的流动。下图表示 LSTM 的一个最小单元， $x_t$  表示  $t$  时刻的输入， $h_t$  表示  $t$  时刻的输出， $C_t$  表示  $t$  时刻的单元状态。

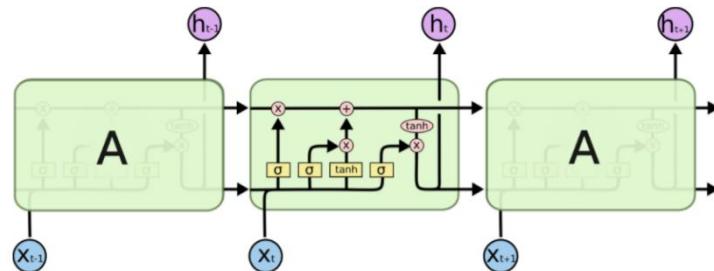


图 45. LSTM 神经元结构

首先确定从节点中丢弃哪些信息：“遗忘门”Sigmoid 层查看  $h_{t-1}$  和  $x_t$ ，并为单元状态  $C_{t-1}$  输出 0-1 之间的数字，1 代表“完全保持这个状态”，0 代表“完全摆脱这个状态”。

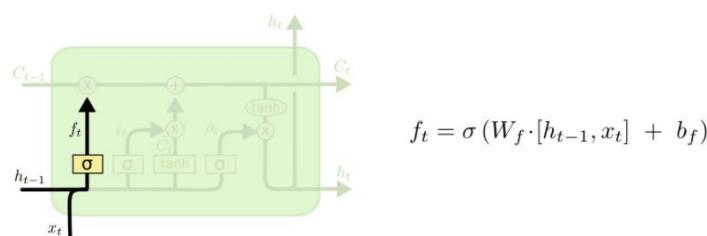


图 46. LSTM “遗忘门” 结构

然后确定在单元节点状态中存储哪些新信息：首先，称为“输入门”的 Sigmoid 层决定我们将更新哪些值。接下来，tanh 层创建可以添加到状态的新候选值  $\tilde{C}_t$  的向量。

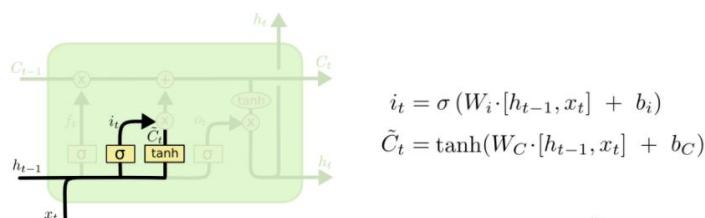


图 47. LSTM “输入门” 结构

然后确定新的单元状态：将旧状态  $C_{t-1}$  乘以  $f_t$  进行遗忘，同时根据一定的缩放比例添加  $\tilde{C}_t$ 。

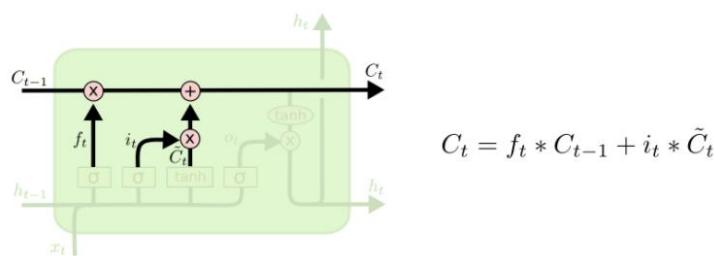


图 48. LSTM “输出门” 计算结构

最后决定我们要输出的内容：首先，运行一个 Sigmoid 层决定我们要输出的单元状态的哪些部分；然后，将单元状态置于 tanh 并将其乘以 sigmoid 门的输出结果。

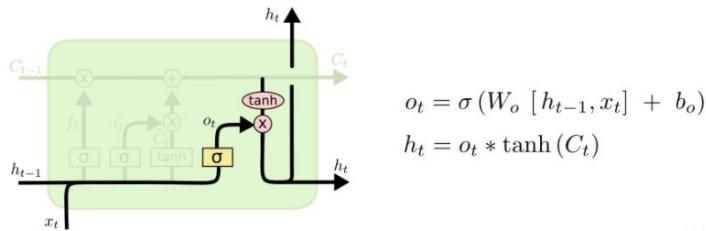


图 49. LSTM “输出门” 结构

## 2) 使用 LSTM 预测未来 4h 的功率

搭建一维 LSTM 神经网络，选取 10 月 1 日-6 日的数据作为训练集（17280 个 sample），10 月 7 日-10 日的数据作为验证集（11520 个 sample），10 月 10 日-13 日的数据作为测试集（8640 个 sample）。LSTM 的输入为前 200 个点（时间跨度约为 1.7h）的风速与功率数据，输出为预测的后 4h 的功率（时间分辨率为 15min，输出维度为 16）。

LSTM 最为重要的参数是 LSTM 层的神经元的数目 (Units)，选取不同的神经元数目 (10, 20, 30, 40, 50) 对上述训练集与验证集分批进行训练，比较预测结果。

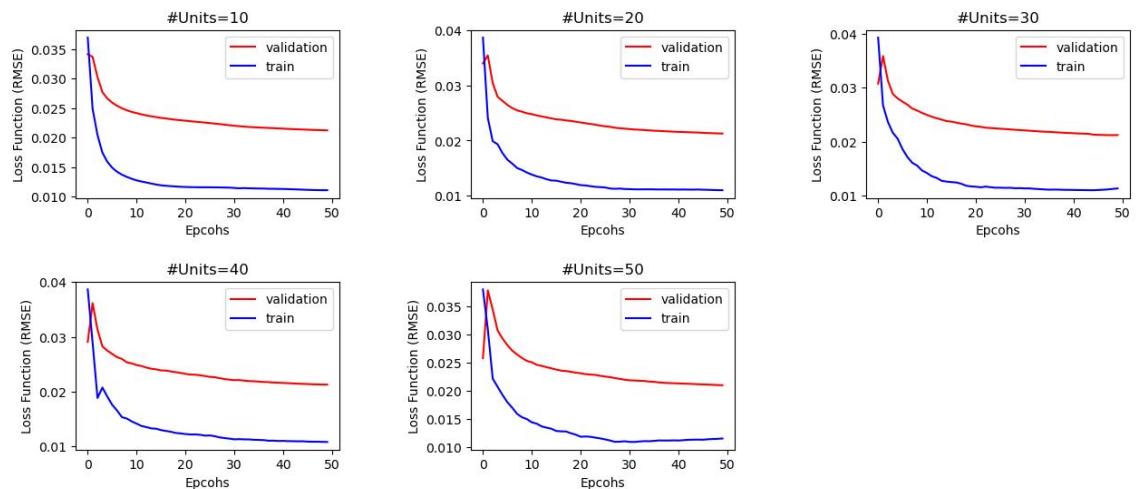


图 50. 不同神经元数目下 LSTM 训练过程中损失函数的变化

训练的批次大小为 32，一共训练 50 轮，训练过程中的损失函数变化如图所示。从图中可以看到在训练过程中训练集和验证集的损失函数经过几轮的训练之后都在不断地下降最后趋于平稳，并且神经元数目越少损失函数下降的越快（因为模型的参数更少了）。此外，由于训练的数据量较大并且训练的轮数足够多，发现在不同的神经元数目下 LSTM 都能够在训练集和验证集上达到较好的预测效果。

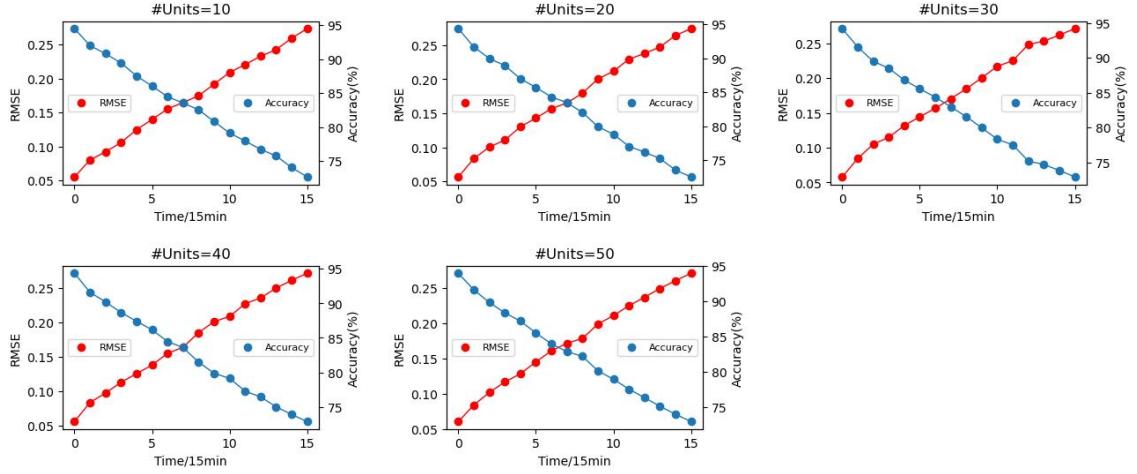


图 51. LSTM 误差和准确率随预测时间的变化

对预测的时间中的每一个时刻点计算 RMSE 和准确率，不同神经元数目的结构如图。所示，可以看到随着预测时间的不断增加，LSTM 预测的准确率也在不断地下降，这可能是因为预测的时间越长，其功率就和前 200 个点的关联越来越小，加之随机因素的影响，使得 LSTM 很难从前 200 个点中记忆和学习得到准确的预测值。此外，不同数目的神经元的 RMSE 和准确率的变化与范围基本一致，且平均准确率均在 82% 左右，这可能也是因为训练的数据量较大并且训练的轮数足够多，不同的神经元数目下 LSTM 都具有了足够的泛化能力可以对测试集给出较为准确的预测。

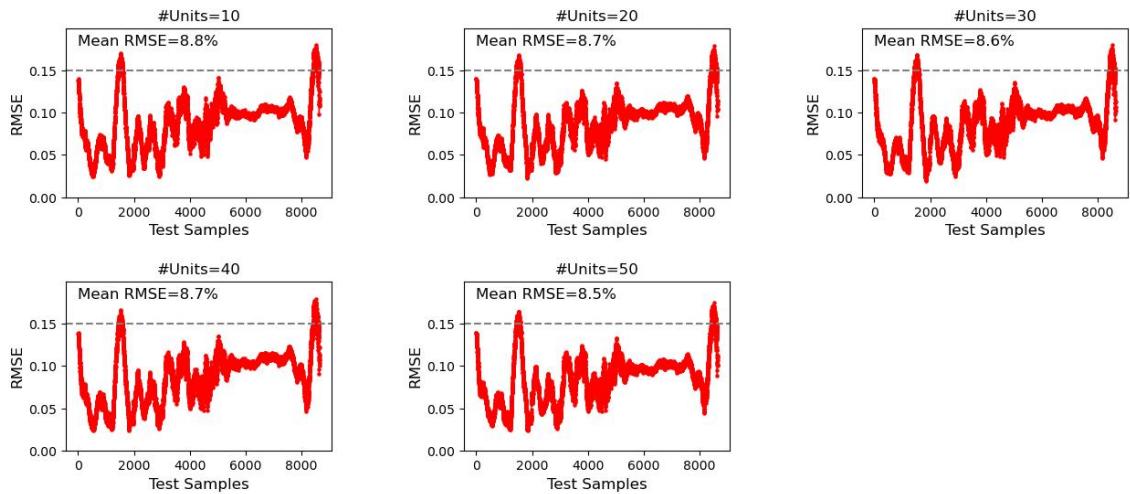


图 52. LSTM 误差 RMSE 随测试集中样本顺序的变化

对测试集中的每一个样本计算 RMSE，不同神经元数目的结果如图。所示。从图中可以看到不同神经元数目下 LSTM 对于相同样本的预测准确率基本一致，绝大多数的样本的 RMSE 都低于 15%，平均 RMSE 均在 8.5% 左右。

随机选取测试集中的一个样本，绘制真实风功率和预测值于一张图中，不同神经元的 LSTM 结果如下图所示。可以看到不同神经元数目的 LSTM 对于同一个样本的预测的趋势都是

和真实值相一致的，也说明了在该风功率预测问题中，LSTM 的神经元数目对于预测结果的影响并不是非常的大。

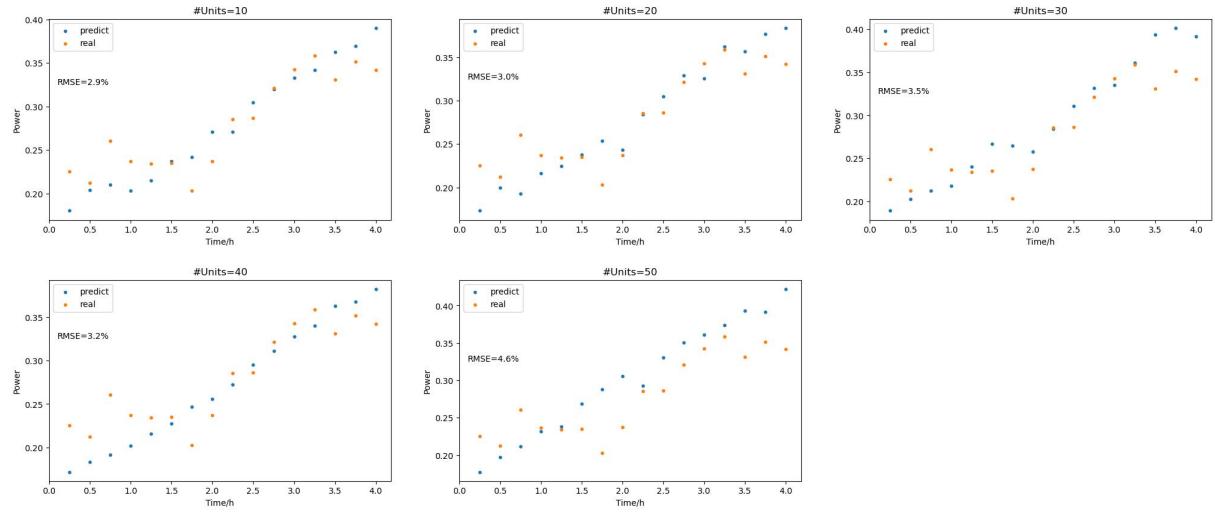


图 53. 真实风功率与预测值比较（未来 4h）

## 2.2.3 GRU 门控循环神经网络

门控循环神经网络 (Gate Recurrent Unit) 也是循环神经网络 RNN 的一种，和 LSTM 一样也是为了解决长期记忆和反向传播中的梯度等问题而提出来的。GRU 在大部分情况下能够达到和 LSTM 一样的效果，并且相比之下更容易进行训练，能够提高训练效率。

GRU 输入输出的结构与普通的 RNN 相似，其中的内部思想与 LSTM 相似，与 LSTM 相比 GRU 内部少了一个“门控”，参数比 LSTM 少，但是却也基本上能够达到与 LSTM 相当的功能。

### 1) GRU 网络工作原理

GRU 的输入输出结构和普通的 RNN 是一样的，有一个当前的输入  $x_t$ ，和上一个节点传递下来的状态  $h_{t-1}$ ，这个隐状态包含了之前节点的相关信息。结合  $x_t$  和  $h_{t-1}$ ，GRU 会得到当前隐藏节点的输出  $y_t$  和传递给下一个节点的隐状态  $h_t$ 。

GRU 神经元的内部结构如下图所示，通过上一个传输下来的状态  $h_{t-1}$  和当前节点的输入  $x_t$  来获取“重置门”与“更新门”的状态  $r_t$  和  $z_t$ ，通过“重置门”决定如何将新的输入信息与前面的记忆相结合，通过“更新门”定义前面记忆保存到当前时间步的量，从而使得 GRU 在长期的预测中也能表现的很好。

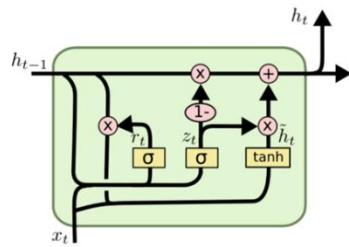


图 54. GRU 神经元的基本结构

## 2) 使用 GRU 预测未来 4h

搭建一维 GRU 神经网络，选取 10 月 1 日-6 日的数据作为训练集（17280 个 sample），10 月 7 日-10 日的数据作为验证集（11520 个 sample），10 月 10 日-13 日的数据作为测试集（8640 个 sample）。GRU 的输入为前 200 个点（时间跨度约为 1.7h）的风速与功率数据，输出为预测的后 4h 的功率（时间分辨率为 15min，输出维度为 16）。

GRU 最为重要的参数也是 GRU 层的神经元的数目 (Units)，选取不同的神经元数目 (10, 20, 30, 40, 50) 对上述训练集与验证集分批进行训练，比较预测结果。

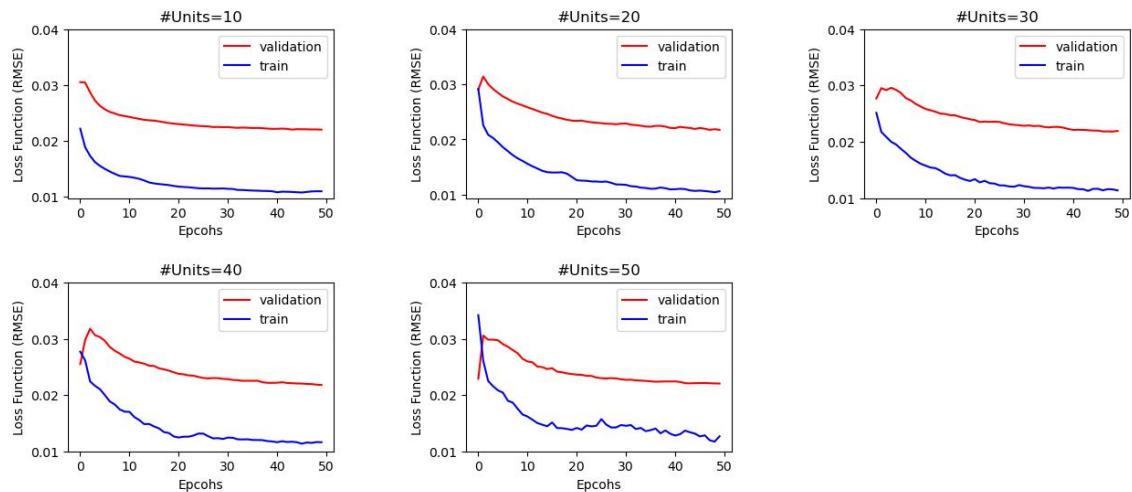


图 55. 不同神经元数目下 GRU 训练过程中损失函数的变化

训练的批次大小为 32，一共训练 50 轮，训练过程中的损失函数变化如图。所示。从图中可以看到在训练过程中训练集和验证集的损失函数经过几轮的训练之后都在不断地下降最后趋于平稳。此外，由于训练的数据量较大并且训练的轮数足够多，发现在不同的神经元数目下 GRU 都能够在训练集和验证集上达到较好的预测效果。但是当神经元数目越多的时候，需要训练的参数也就越多，GRU 训练过程中的损失函数也出现了一些不稳定的波动。由此可以看出，神经元的数目不是越多越好，而是要根据深度学习的具体问题和数据集的特点来综合考量。

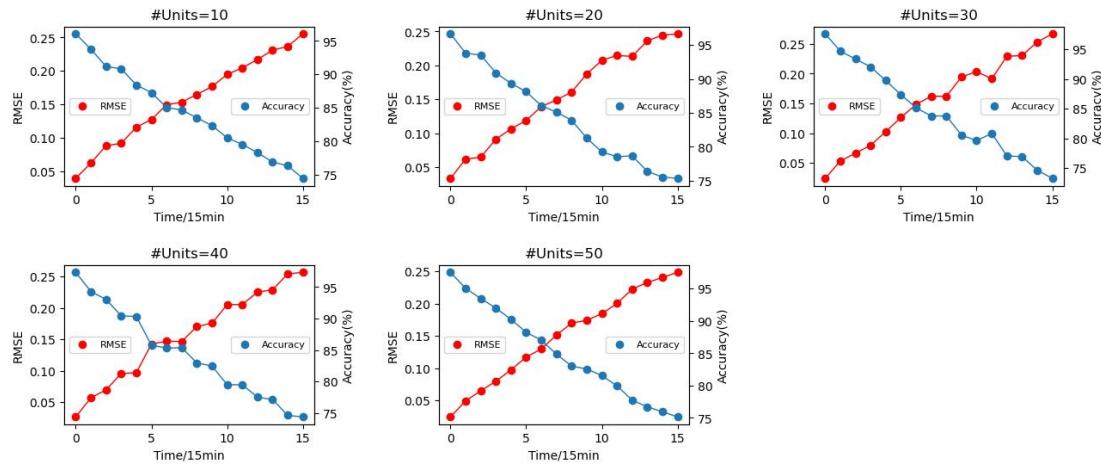


图 56. GRU 误差和准确率随预测时间的变化

对预测的时间中的每一个时刻点计算 RMSE 和准确率，不同神经元数目的结构如图。所示，可以看到随着预测时间的不断增加，GRU 预测的准确率也在不断地下降，这与 LSTM 表现一致，是由于时间序列预测的特点决定的。此外，不同数目的神经元的 RMSE 和准确率的变化与范围基本一致，且平均准确率均在 84% 左右，这可能也是因为训练的数据量较大并且训练的轮数足够多，不同的神经元数目下 GRU 都具有了足够的泛化能力可以对测试集给出较为准确的预测。

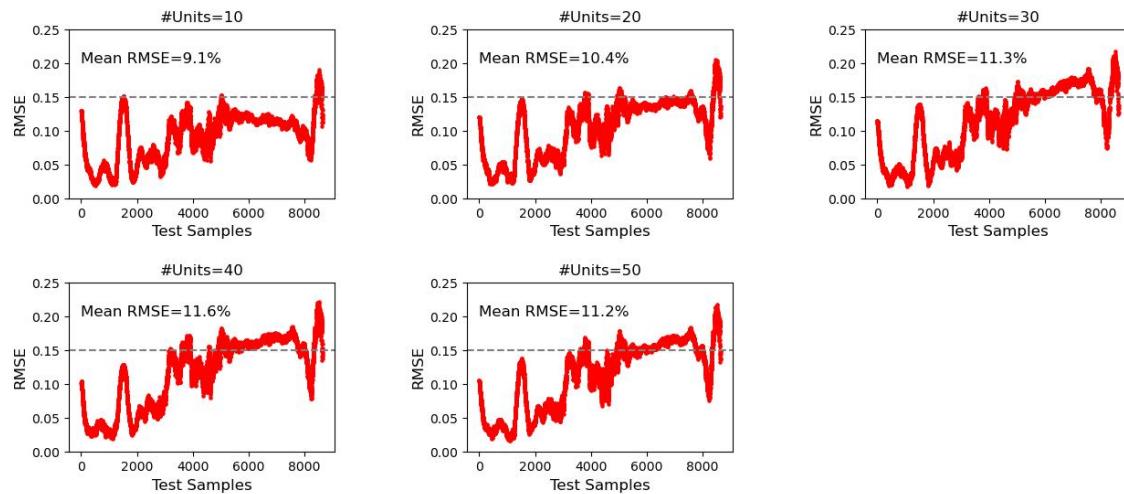


图 57. GRU 预测误差随测试集中样本顺序的变化

对测试集中的每一个样本计算 RMSE，不同神经元数目的结果如图。所示。从图中可以看到当 GRU 的神经元数目为 10 或者 20 的时候，绝大多数测试集中样本的 RMSE 都是小于 15% 的，但是当 GRU 的神经元数目为 30、40 和 50 的时候，有一小半的样本的 RMSE 都大于了 15%，在这里也可以看到神经元的数目不是越多越好。关于测试集的 RMSE 其他的一些特征将在下一部分 LSTM 和 GRU 的对比之中详细分析。

随机选取测试集中的一个样本，绘制真实风功率和预测值于一张图中，不同神经元的 GRU 结果如下图所示。可以看到不同神经元数目的 GRU 对于同一个样本的预测的趋势都是和真实值相一致的，但是神经元数目较大的模型与真实值的偏离也较大，与上面测试集 RMSE 体现出来的神经元数目越多预测结果越差的现象一致。由此可见，GRU 模型神经元的数目对于预测结果的影响还是比较显著的，要根据深度学习问题和数据集的特征选择最为合适的参数。

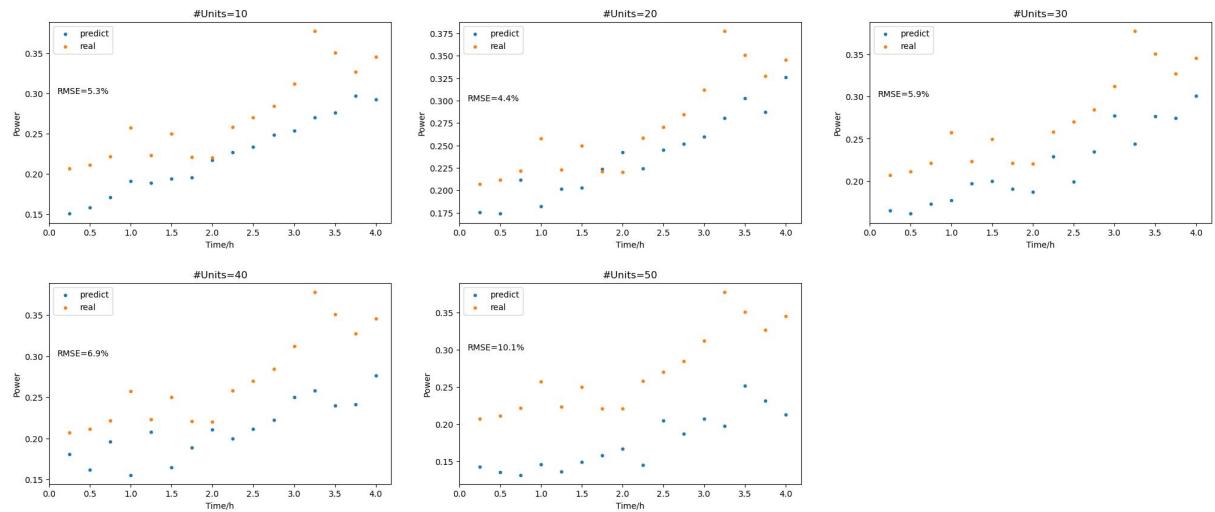


图 58. 真实风功率与预测值比较（未来 4h）

## 2.2.4 LSTM 和 GRU 的比较

对比 LSTM 和 GRU 循环神经网络的训练过程中的损失函数的变化（下图左为 LSTM，右为 GRU），我们可以发现当神经元数目增多的时候，LSTM 的损失函数的变化一直比较稳定，但是 GRU 损失函数在神经元数目为 40 的时候就开始发生挺大的波动。

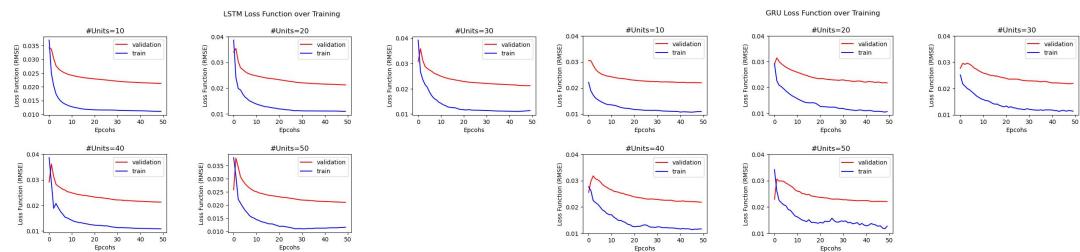


图 59. 训练过程中 LSTM 与 GRU 的损失函数变化比较

测试集是选取的接在训练集和验证集之后的 10 月 11 日-13 日的数据，测试集的样本顺序其实是和真实的时间顺序一一对应的。对比 LSTM 和 GRU 在测试集上的顺序排布的 RMSE 可以看到：在时间跨度为 3 天的测试集上，不同神经元数目下 LSTM 的预测效果都是几乎一

致比较稳定的；但是不同神经元数目下 GRU 的预测效果，当神经元数目比较少，并且测试集的样本比较靠前（接近训练集和验证集）的时候，预测的结果比较准确。GRU 在神经元数目和时间跨度上都没有 LSTM 稳定，最为有意思的一点是当测试集的样本比较靠后也就是在时间上远离训练集与验证集的时候，GRU 某些模型的预测效果就会开始下降，说明 GRU 虽然较 LSTM 训练更为高效，但是其在抽取长期特征和进行长期预测的效果上还是不如老牌的 LSTM 稳定和优秀。

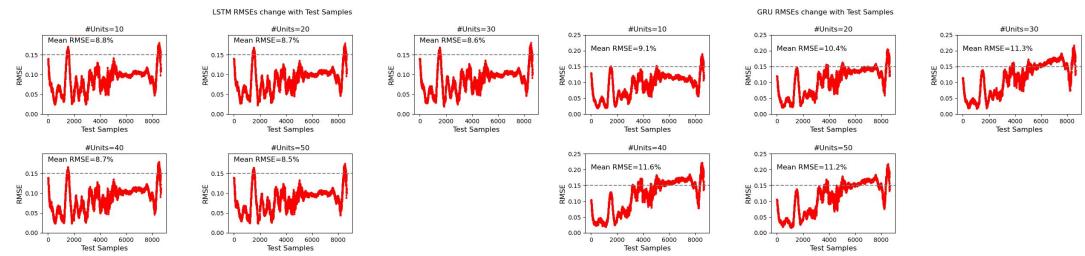


图 60. LSTM 与 GRU 在测试集上的误差随样本顺序变化的比较

绘制出测试集样本预测结果的 RMSE 的箱型图如下图所示（左为 LSTM，右为 GRU），从图中可以很清楚地看到：LSTM 的预测结果的 RMSEs 范围在 2%-18%，其上边缘大致在 15%左右，大于 15%的都被认定为是异常值，其中位数略小于 10%，误差小精度高；而 GRU 的预测结果的 RMSEs 范围在 2%-22%，其上四分位数几乎就在 15%的位置，误差虽然分布比较均匀但和 LSTM 相比均较高。

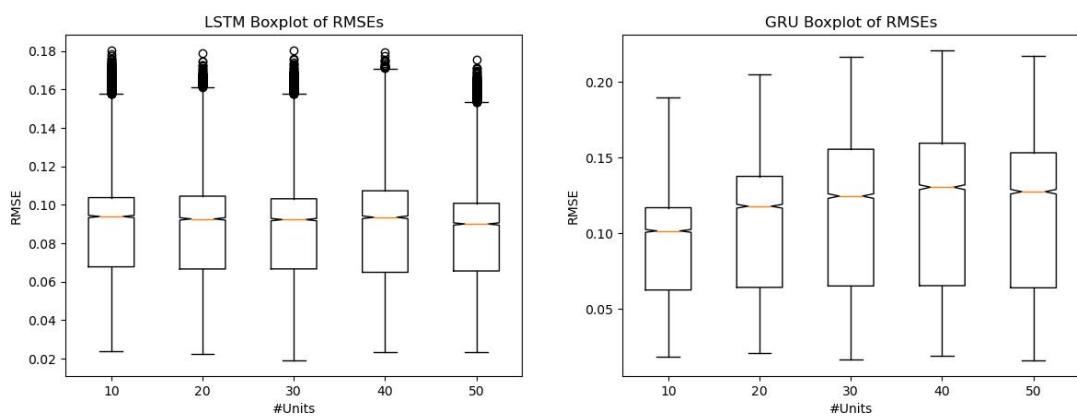


图 61. LSTM 和 GRU 在测试集上的 RMSE 的箱型图

由此可以得出，当需要进行长期的较高精度的预测（比如风功率预测）的时候，如果有足够的算力和时间还是选择 LSTM 来进行预测结果可能会更好。

## 2.3 卷积神经网络 CNN

### 2.3.1 一维 CNN

我们知道，CNN 在计算机视觉领域是非常有用的，应用在图像上的 CNN 卷积核是 2 维的，而 1 维的 CNN 在时间序列的预测上也可以起到一定作用，因为 CNN 一般包含卷积层和池化层，卷积计算互相关，所以可以挖掘时序的相关性，池化层可以进行降维，用于提取特征。

#### 1) 1 维 CNN 的基本工作原理

二维卷积神经网络（2D CNN）在图像识别上的研究有很多，并且广泛运用。但一维卷积神经网络（1D CNN）被研究的相比而言较少，有部分学者在研究 1DCNN 在自然语言处理（NLP）上的应用。但是目前很少找到文章能够解释如何构造 1DCNN 来解决机器学习问题。但是其实 CNN 可以很好地识别出数据中的简单模式，然后使用这些简单模式在更高级的层中生成更复杂的模式。当我们想要从整体数据中较短的片段中获得特征，并且这个特征在这个数据片段中的位置不具有高相关性时，1D CNN 是有一定效果的，它可以应用于传感器数据的时间序列分析，也可以很好地用于分析具有固定长度周期的信号数据。

一个 CNN 主要由 5 层构成，数据输入层、卷积计算层、ReLU 激励层、池化层、全连接层。卷积计算层也是 CNN 得名卷积神经网络，与其他神经网络区分的原因。在卷积层中，使用卷积核进行局部感知，这个卷积核是一个小的权重矩阵，能够覆盖输入数据，对应位置加权相乘，其和作为结果，卷积核在数据数据上滑动，并且形成新的矩阵，最终经过 Relu 激励层和池化层经全连接输出结果。

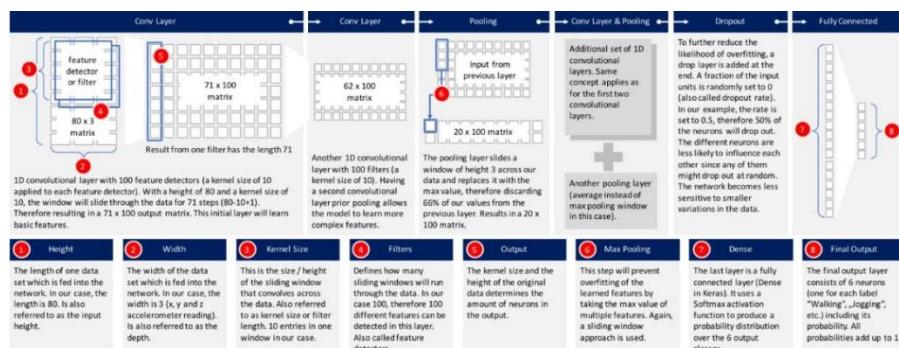


图 62. 卷积神经网络的原理

无论是 1D 还是 2D 还是 3D，卷积神经网络都有相同的特点和处理方法，其区别是输入数据的维数以及“滤波器”如何在数据之间滑动。下图展示了一维 CNN 的处理过程。

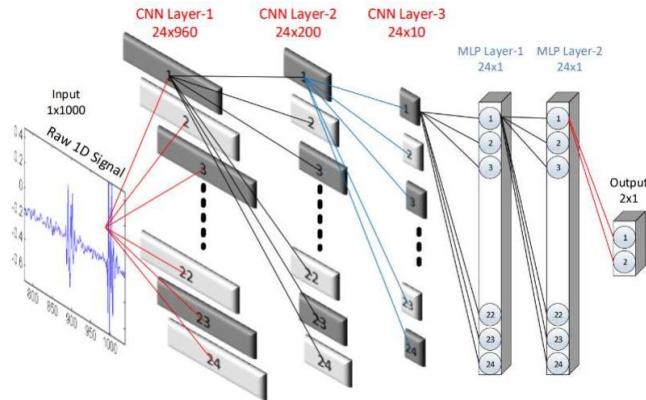


图 63. 一维 CNN 原理

## 2) 1 维 CNN 预测

搭建一维 CNN，输入风功率数据，使用前 30 个时刻，预测下一个时刻的风功率，搭建一个卷积层 (`filters=64, kernel_size=2`)，接着再叠加一个 (`kernel_size=3`) 的卷积层，然后加入一维池化层 (`pool_size=2`)，叠加 Dense 层 (`unit=50`)，最终是输出层，中间插入了一层 Dropout (0.3) 的层避免过拟合。迭代 1000 次之后，训练的损失函数已经很低了，如下图所示。

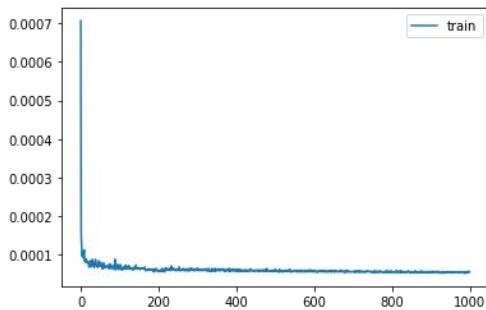


图 64. CNN 损失函数

对于 CNN，发现很有意思的是在调节参数的过程中，`train_loss` 一般都能够保持稳步降低，但是 `val_loss` 却变化非常大，一点参数的修改都会使得 `val_loss` 曲线出现很大的变化。猜测这是因为 CNN 非常容易过拟合和欠拟合所导致的结果。当我添加过多的 Dropout 以及训练数据不足时，就会出现类似以下左图的结果。`Val_loss` 大幅震荡，并且先下降后上升。优化参数减少欠拟合后，又会逐渐出现下右图的情况。终经过参数调节，使得过拟合和欠拟合都减少了不少，但仍有不少震荡，最终我确定的模型如下中图所示。

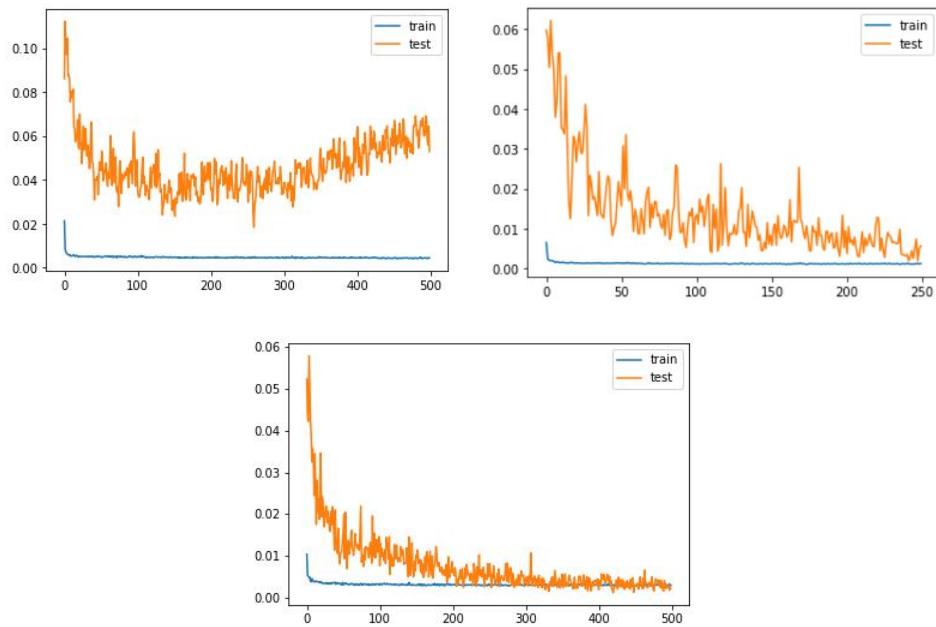


图 65. CNN val\_loss 变化

Rmse 和精确率如下图所示，震荡很明显，模型并不是很好。

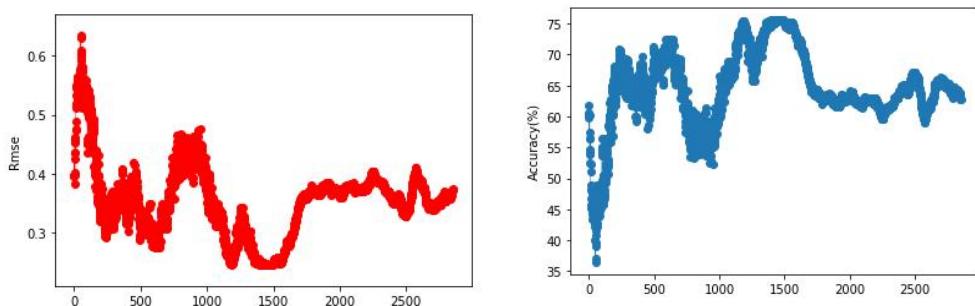


图 66.CNN 误差及精确率

预测结果并不是很理想，可能是因为参数调节以及训练集输入的数据量还不是非常理想，  
预测结果如下图所示。

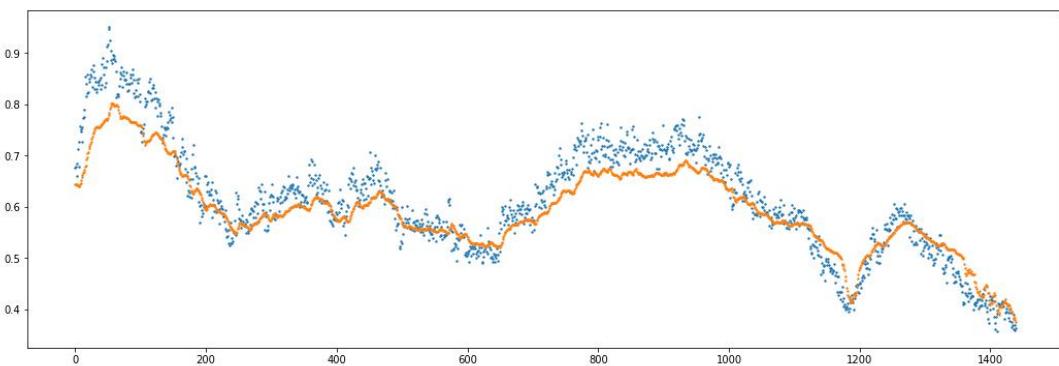


图 67. CNN 预测结果

## 2.3.2 TCN 时序卷积网络

时序卷积网络（Temporal Convolutional Network）是在 2018 年新提出的一个模型，可以用于时序数据处理。卷积网络已被证明在提取结构化数据中的高级特征方面具有不错的效果。而时间卷积网络 TCN 则是一种利用因果卷积和空洞卷积的神经网络模型，它可以适应时序数据的时序性并可以提供视野域用于时序建模。

### 1) TCN 网络工作原理

首先，TCN 是一种因果卷积网络，所谓因果卷积网络是指上一层  $t$  时刻的值只能依赖于下面那层  $t$  时刻以及  $t$  时刻之前的值。与传统的神经网络不同，TCN 不能看见未来的数据，只有前面的因才有后面的果。

其次，TCN 采用膨胀卷积通过跳过给定步骤的大小，然后在一个比它的大小更大的区域上应用滤波器。通过增加采样间隔的大小，使得网络在较少的层数下就可记住较多以前的信息，一般来说，越高的层使用的 Dilation 越大。这保证了历史记录中每个输入的接收域，并通过使用深度网络获得一个非常大的接收域作为有效的历史记录。下图是因果膨胀卷积的图示。

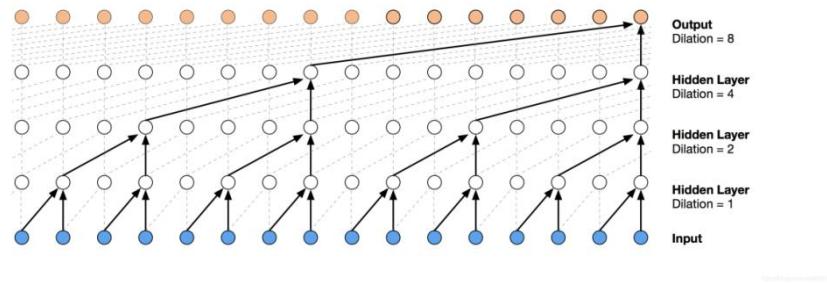


图 68. TCN 膨胀卷积示意图

最后，由于 TCN 的接收域大小取决于网络深度  $n$ 、滤波器大小  $k$  和空洞扩张因子  $d$ ，因此使 TCN 变深变大是获得足够大的接收域的关键。从经验上看，使网络变深变窄，即叠加大量的层，选择较细尺度的过滤器尺寸，是一种有效的架构。残差连接已被证明是非常有效的训练深度网络。在残差网络中，整个网络采用跳跃连接，以加快训练过程，避免深度模型的梯度消失问题。

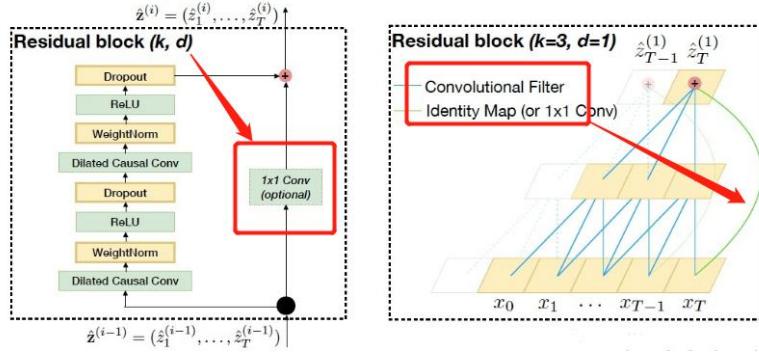


图 69. TCN 残差连接示意图

## 2) 使用 TCN 预测未来 4h 的风功率

搭建一维 TCN 神经网络，选取 10 月 1 日-6 日的数据作为训练集（17280 个 sample），10 月 7 日-10 日的数据作为验证集（11520 个 sample），10 月 10 日-13 日的数据作为测试集（8640 个 sample）。GRU 的输入为前 200 个点（时间跨度约为 1.7h）的风速与功率数据，输出为预测的后 4h 的功率（时间分辨率为 15min，输出维度为 16）。

TCN 神经网络的主要参数是过滤器（filter）的数目和卷积核的大小（kernel\_size），尝试对于不同的过滤器数目（10, 15, 20）和不同的卷积核大小（4, 6, 8）采用均方误差（mse）作为损失函数进行训练，对比不同神经网络参数下的训练结果。

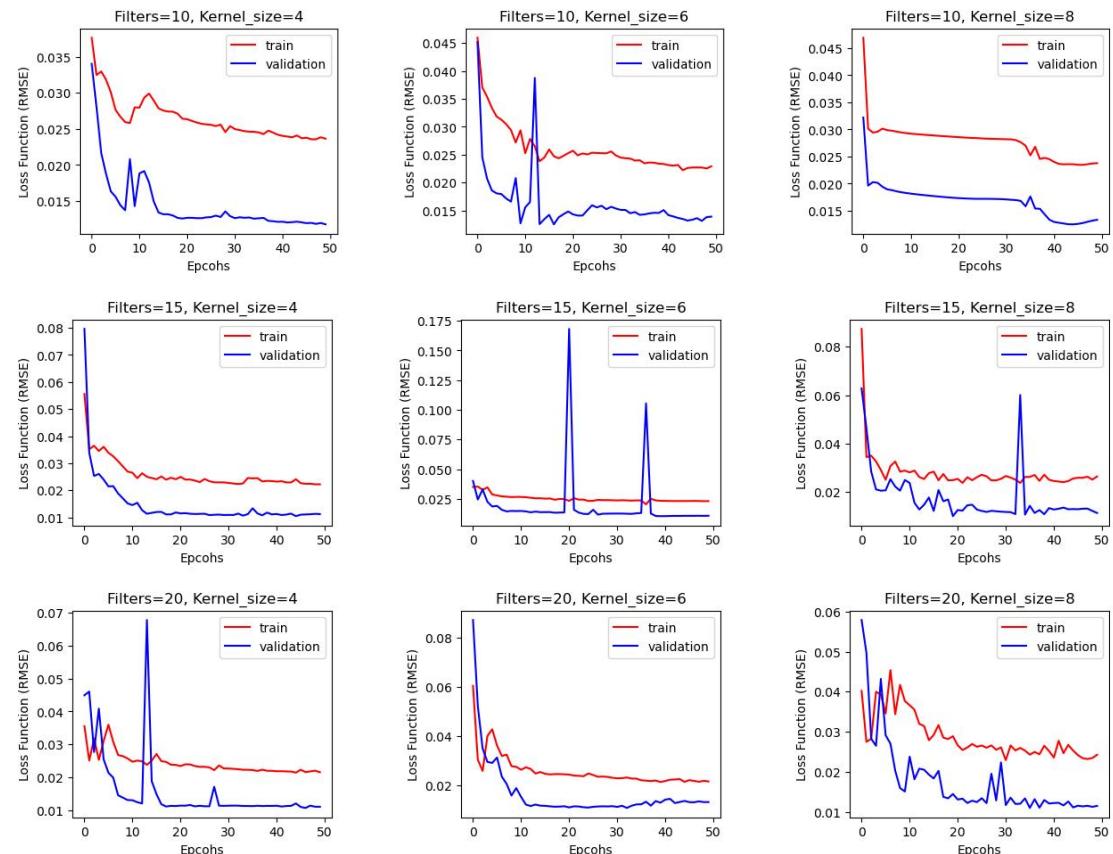


图 70. TCN 训练过程中损失函数的变化

训练的批次大小为 32，一共训练 50 轮，训练过程中的损失函数变化如上图。从图中可以看到除了不同过滤器数目和不同大小卷积核下损失函数随训练周期的变化非常不一致，而且验证集有可能出现异常的峰，表示验证集在该次训练的模型上误差很大。

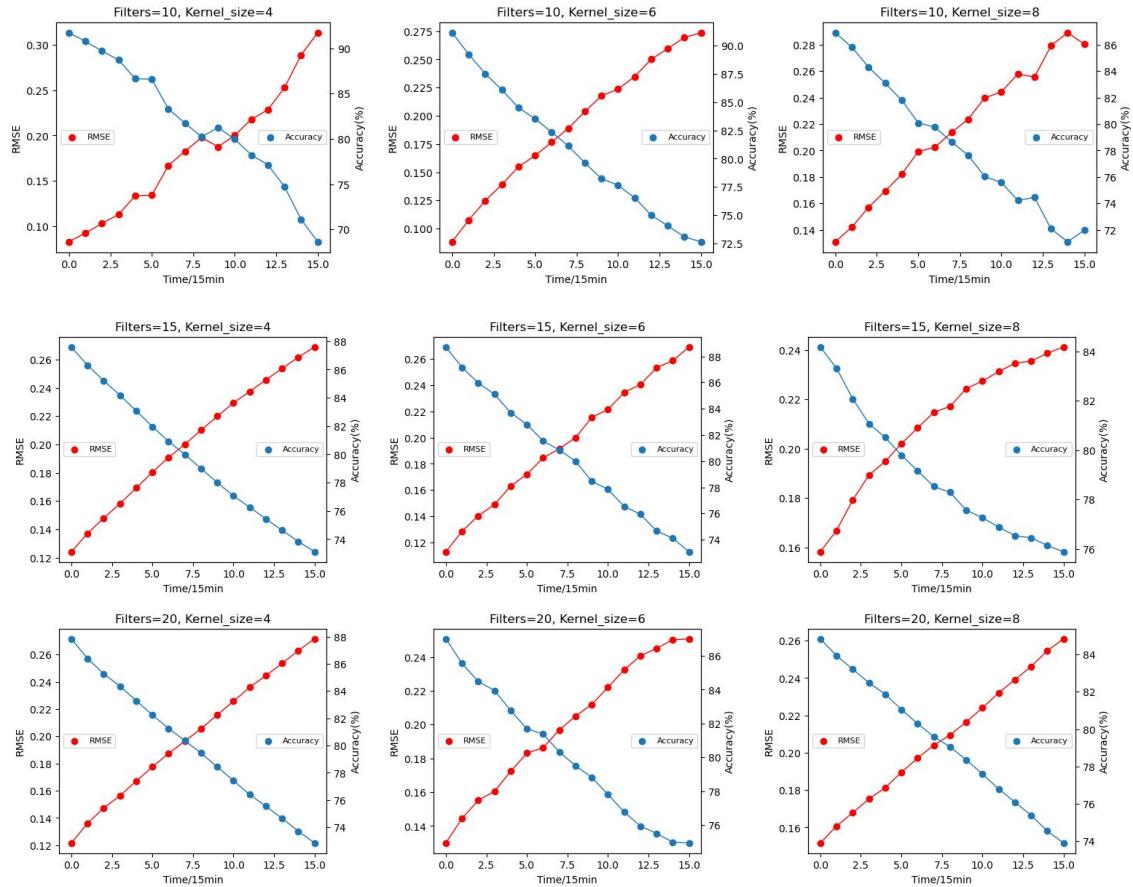
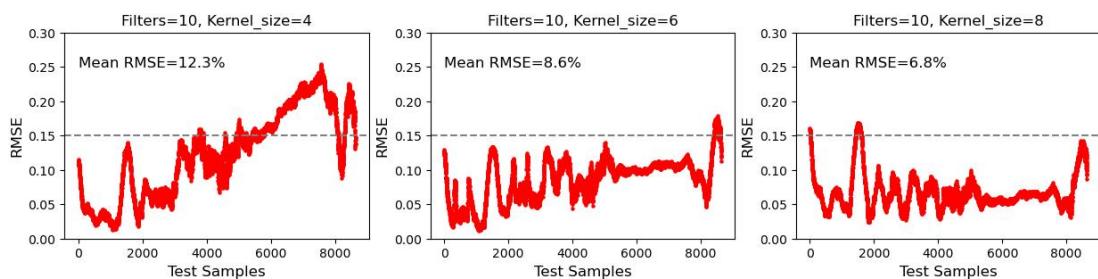


图 71.TCN 误差与准确率随时间的变化

对预测的时间中的每一个时刻点计算 RMSE 和准确率，不同神经元数目的结构如上图。所示，可以看到随着预测时间的不断增加，TCN 预测的准确率也在不断地下降，这是由于时间序列预测的特点决定的。此外，不同参数的 TCN 和准确率的变化与范围基本一致，且平均准确率均在 80% 左右，这可能也是因为训练的数据量较大并且训练的轮数足够多，不同的神经元数目下 GRU 都具有了足够的泛化能力可以对测试集给出较为准确的预测。



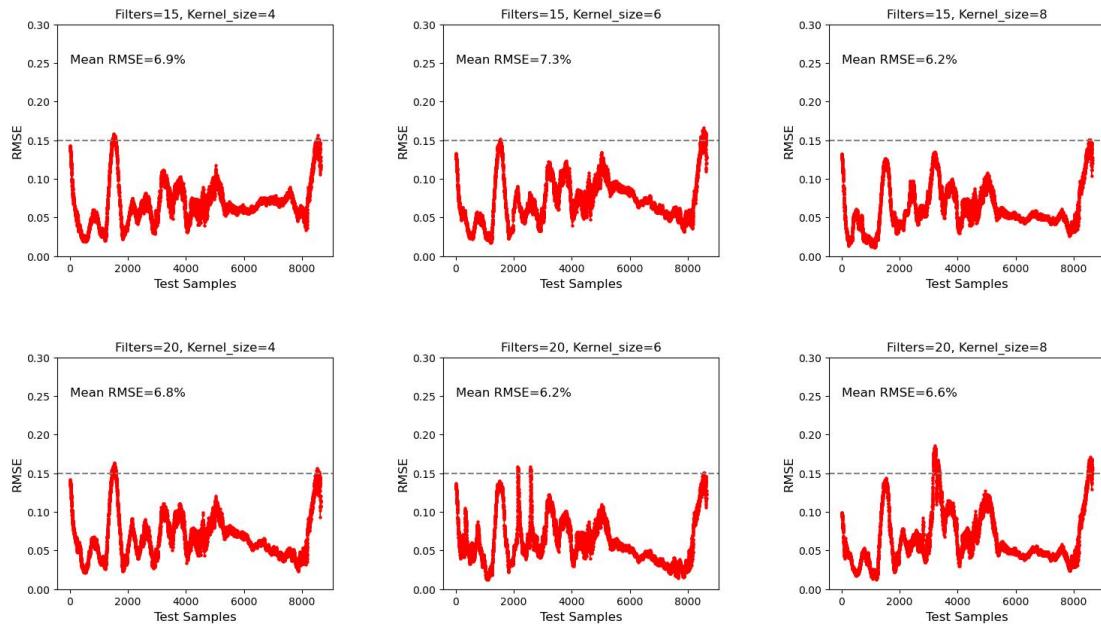


图 72.TCN 误差 RMSE 随测试集顺序样本的变化

从 TCN 误差 RMSE 随测试集顺序样本的变化可以发现，除了个别（过滤器数目为 10，卷积核大小为 4）的模型预测的误差较大外，其余的大部分预测的误差 RMSE 都低于 15%，甚至比 LSTM 的结果还要好。

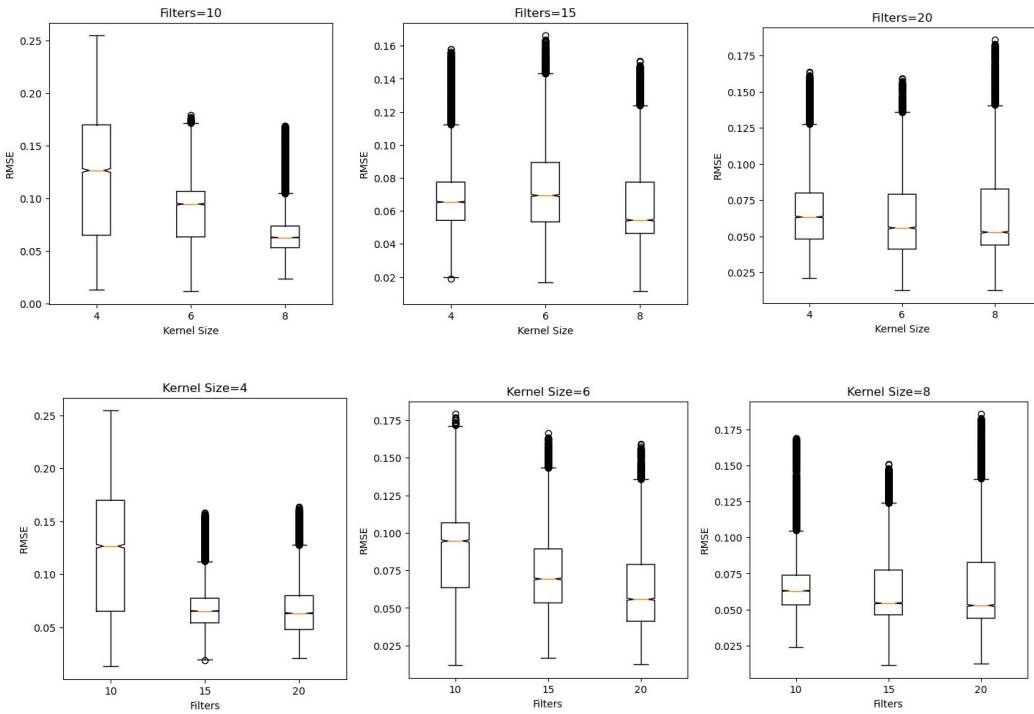


图 73. 不同 TCN 模型的测试集误差分布

分别固定过滤器数目和卷积核大小看测试集的误差分布。滤波器数目为 10 卷积核大小为 4 的 TCN 模型预测能力最差，其他的模型预测能力相当。

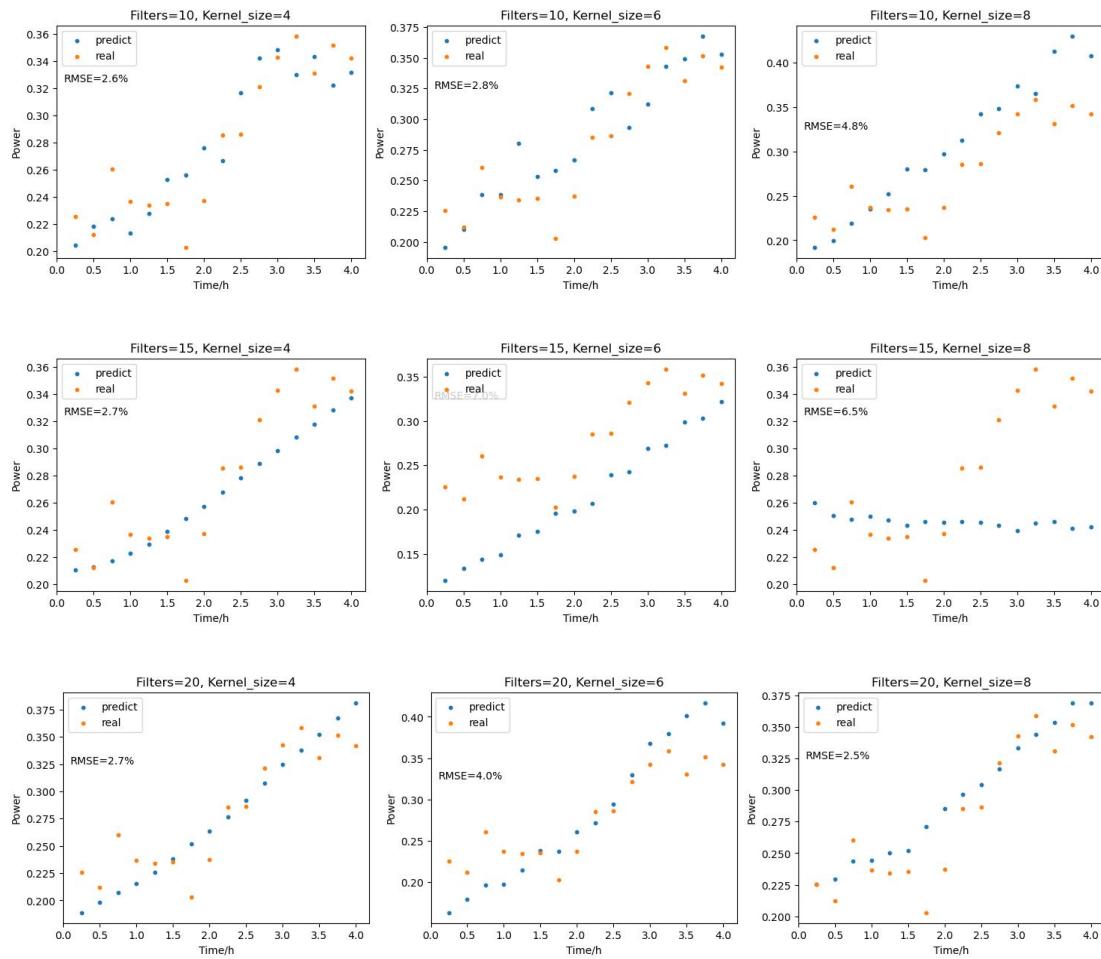


图 74.TCN 的预测风功率与真实风功率

随机选取测试集中的一个样本，绘制真实风功率和预测值于一张图中，发现不同参数的 TCN 模型对于同一个样本的预测除了个别的模型基本一致，并且预测的的 RMSE 误差均小于 15%。综合来看 TCN 的预测效果是不错的。

### 3 统计学方法预测风功率

当然除了深度学习的神经网络能够预测风功率之外，传统的统计学方法也能够根据时间序列自身的统计特征来预测其走向，但是统计学的方法一般只能进行短期的预测，且精度一般没有机器学习和深度学习的方法高。自回归移动平均模型（ARMA 模型），它是目前最常用的拟合平稳时间序列的统计学模型。ARMA 模型是 AR 模型和 MA 模型的组合，都属于统计学范畴下的多元线性回归模型。

### 3.1 ARMA 模型原理

#### 1) AR 模型

具有如下结构的模型称为 p 阶自回归模型，简记为 AR(p)：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t$$

在 t 时刻随机变量  $x_t$  的取值是前 p 个  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  多元线性回归，误差项是当前的随机干扰  $\epsilon_t$ ，为零均值白噪声序列。平稳 AR (p) 模型的性质如下图所示。

统计量	性质	统计量	性质
均值	常数均值	自相关系数 (ACF)	拖尾
方差	常数方差	偏自相关系数 (PACF)	p 阶截尾

图 75. 平稳 AR 模型的性质

#### 2) MA 模型

具有如下结构的模型称为 q 阶自回归模型，简记为 MA (q)：

$$x_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

在 t 时刻随机变量  $X_t$  的取值是前 q 个随机扰动  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}$  的多元线性回归函数， $\mu$  是序列  $\{X_t\}$  的均值，误差项是当前的随机干扰  $\epsilon_t$ ，为零均值白噪声序列。平稳 MA (q) 模型的性质如下图所示。

统计量	性质	统计量	性质
均值	常数均值	自相关系数 (ACF)	$q$ 阶截尾
方差	常数方差	偏自相关系数 (PACF)	拖尾

图 76. 平稳 MA 性质的性质

#### 3) ARMA 模型

具有如下结构的模型称为自回归移动平均模型，简记为 ARMA (p, q)：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

在 t 时刻随机变量  $X_t$  的取值是前 p 个  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  和前 q 个随机扰动  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}$  的多元线性回归函数，误差项是当前的随机干扰  $\epsilon_t$ ，为零均值白噪声序列。认为  $x_t$  主要是受过去 p 期的序列值和过去 q 期的误差项的共同影响。平稳 ARMA (p, q) 模型的性质如下图所示。

统计量	性质	统计量	性质
均值	常数均值	自相关系数 (ACF)	拖尾
方差	常数方差	偏自相关系数 (PACF)	拖尾

图 77. 平稳 MARA 模型的性质

## 3.2 ARMA 预测

我们尝试使用 ARMA 对于给定的数据集进行预测，将 10 月 1 日-10 日的风功率数据作为训练集，11 日-13 日的风功率数据作为测试集。

首先对风功率序列进行 ADF 平稳定性检验如下图所示，p 值显著地大于 0.05，所以原始数据是非平稳序列。

	value
Test Statistic Value	-2.525867
p-value	0.109307
Lags Used	25
Number of Observations Used	37414
Critical Value(1%)	-3.430525
Critical Value(5%)	-2.861617
Critical Value(10%)	-2.566811

图 78. 风功率序列 ADF 检验

所以对原始风功率数据进行一次差分，再次进行 ADF 平稳定性检验如下图所示，p 值显著地小于 0.05，所以一阶差分之后的风功率是平稳序列。

	value
Test Statistic Value	-36.006611
p-value	0.0
Lags Used	24
Number of Observations Used	37414
Critical Value(1%)	-3.430525
Critical Value(5%)	-2.861617
Critical Value(10%)	-2.566811

图 79. 风功率一次差分序列 ADF 检验

接着对一阶差分之后的风功率数据进行白噪声检验，其 LB 和 PB 的 p 值都显著地小于 0.05，所以一阶差分之后的风功率数据不是白噪声，可以使用 ARMA 模型拟合。一阶差分之后的风功率数据的自相关图和偏相关图如下所示，根据自相关图和偏相关图判定 ARMA 模型的 q=7, p=4。

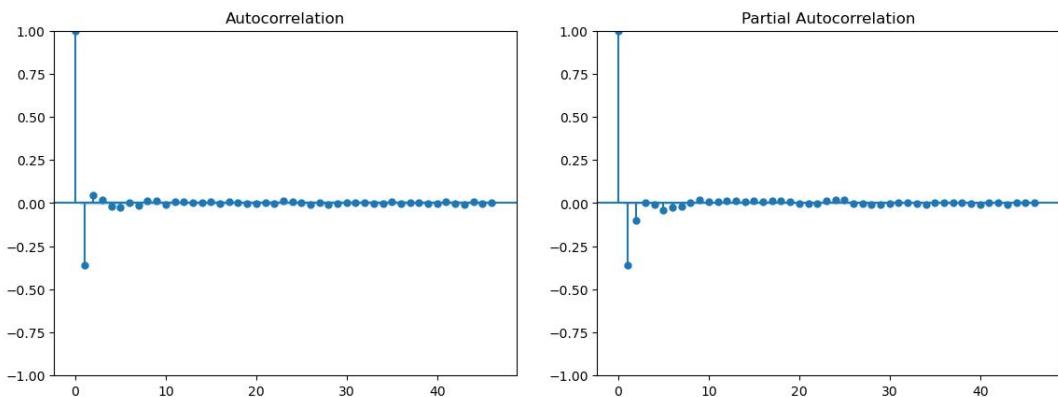


图 80. 风功率一阶差分序列的自相关和偏相关图

使用 ARMA 模型拟合的结果如图所示：

SARIMAX Results						
Dep. Variable:	power	No. Observations:	28800			
Model:	ARIMA(4, 1, 7)	Log Likelihood	89130.312			
Date:	Wed, 11 Jan 2023	AIC	-178236.625			
Time:	23:27:34	BIC	-178137.407			
Sample:	10-01-2015 - 10-10-2015	HQIC	-178204.728			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.2465	5.150	0.048	0.962	-9.847	10.340
ar.L2	0.0669	3.773	0.018	0.986	-7.329	7.463
ar.L3	0.0563	1.523	0.037	0.971	-2.928	3.041
ar.L4	-0.2071	1.280	-0.162	0.871	-2.716	2.302
ma.L1	-0.5642	5.150	-0.110	0.913	-10.658	9.529
ma.L2	0.0165	5.325	0.003	0.998	-10.420	10.453
ma.L3	-0.0284	2.222	-0.013	0.990	-4.383	4.327
ma.L4	0.1712	1.632	0.105	0.916	-3.028	3.370
ma.L5	-0.0897	0.672	-0.133	0.894	-1.407	1.228
ma.L6	0.0011	0.082	0.014	0.989	-0.160	0.162
ma.L7	0.0119	0.086	0.138	0.890	-0.157	0.181
sigma2	0.0001	5.44e-07	220.514	0.000	0.000	0.000
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	27777.15			
Prob(Q):	0.94	Prob(JB):	0.00			
Heteroskedasticity (H):	0.42	Skew:	0.10			
Prob(H) (two-sided):	0.00	Kurtosis:	7.81			

图 81. ARMA 模型拟合结果

使用该模型预测训练集紧接着的 30s (60 个数据点) 的结果如下图所示，可以看到预测的风功率几乎是一条直线，预测效果并不理想。

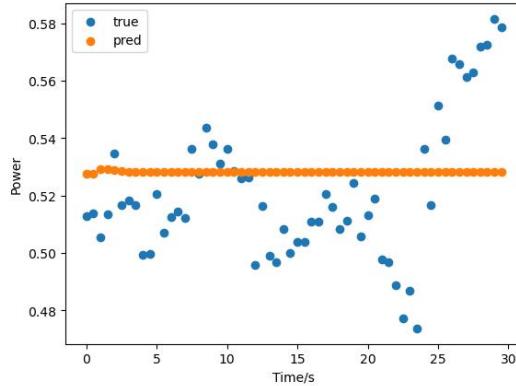


图 82.ARMA 预测的风功率和实际风功率

## 4 总结

本论文中，我们首先介绍了风功率预测的研究背景，接着对于拿到的数据进行了探索性数据分析和预处理，描述了数据的统计特性，接着对数据做了清洗和预处理，然后对数据进行了多种可视化的呈现，并对数据做了特征提取，为未来的模型训练做准备。然后，我们尝试使用多种办法对风功率做预测，使用了 ANN 人工神经网络，包括 MLP 前馈神经网络、BP 算法等；接着使用了 RNN 循环神经网络，包括基础的 RNN 及 LSTM、GRU 等；然后使用了 CNN 卷积神经网络。我们还是用了统计学方法进行实验，使用了 ARMA 等方法。在这过程中，我们对各种模型的原理和效果做了解释和比较，阐释了不同的预测效果，并且其中有不少得到了较好的结果。在这个过程中，我们不仅加深了对本数据的了解，做了相当多收获颇丰的工作，也对 data science 有了更深刻的理解和体会。

## 5 参考文献

- [1] 纪代颖. 基于风电的不确定性对电力系统的影响分析 [J]. 电子测试, 2022, 36(10):99-101. DOI:10.16520/j.cnki.1000-8519.2022.10.003.
- [2] GB/T 40607-2021, 调度侧风电或光伏功率预测系统技术要求 [S].
- [3] 冉靖, 张智刚, 梁志峰, 鲁宗相, 陈敏. 风电场风速和发电功率预测方法综述 [J]. 数理统计与管理, 2020, 39(06):1045-1059. DOI:10.13860/j.cnki.sltj.20200930-002.
- [4] 强硕, 马进. 风电功率预测方法研究综述 [C]//2017 中国自动化大会 (CAC2017) 暨国际智能制造创新大会 (CIMIC2017) 论文集. [出版者不详], 2017:987-992.

- [5] 钱政, 裴岩, 曹利宵, 王婧怡, 荆博. 风电功率预测方法综述 [J]. 高电压技术, 2016, 42(04): 1047-1060. DOI: 10.13336/j.1003-6520.hve.20160405021.
- [6] 崔颢, 马平. 基于优化 BP 神经网络和非参数估计的风功率区间预测 [J]. 电子设计工程, 2022, 30(13): 6-10. DOI: 10.14022/j.issn1674-6236.2022.13.002.
- [7] 周晓玲, 银磊, 张瑞光, 刘贞贞, 覃明文, 汤蔚. 风功率数据清洗方法、清洗装置及预测方法 [P]. 广东省: CN114969017B, 2022-11-11.
- [8] 邓艾东, 丁雪, 邓敏强, 史曜炜, 刘洋. 一种基于改进 LSTM-CNN 的风功率短期预测方法 [P]. 江苏省: CN115115090A, 2022-09-27.
- [9] 陈海鹏, 李赫, 阚天洋, 赵畅, 张忠, 于海薇. 考虑风电时序特性的深度小波-时序卷积网络的超短期风功率预测 [J/OL]. 电网技术: 1-14[2023-01-12]. DOI: 10.13335/j.1000-3673.pst.2022.1019.
- [10] 李静茹, 姚方. 引入注意力机制的 CNN 和 LSTM 复合风电预测模型 [J]. 电气自动化, 2022, 44(06): 4-6.
- [11] 张晋华, 程鹏, 黄慧, 李茂茗. 基于长短期记忆网络的风电功率短期预测方法 [P]. 河南省: CN109214566B, 2021-02-26.
- [12] 朱乔木, 李弘毅, 王子琪, 陈金富, 王博. 基于长短期记忆网络的风电场发电功率超短期预测 [J]. 电网技术, 2017, 41(12): 3797-3802. DOI: 10.13335/j.1000-3673.pst.2017.1657.
- [13] 易善军, 王汉军, 向勇, 田长翼, 高大禹, 陈志奎. 基于集成多尺度 LSTM 的短时风功率预测 [J]. 重庆大学学报, 2021, 44(07): 75-81.
- [14] 赵攀, 戴义平, 夏俊荣, 盛迎新. 卡尔曼滤波修正的风电场短期功率预测模型 [J]. 西安交通大学学报, 2011, 45(05): 47-51.
- [15] 梁涛, 董玉兰. 基于改进的网格法和 BA-BP-ARMA 模型的风功率预测 [J]. 水电能源科学, 2018, 36(07): 211-214.
- [16] 梁涛, 杨改文, 姜文, 李永强. 基于 ARMA 误差修正的 LM-BP 模型的风功率预测 [J]. 高技术通讯, 2019, 29(05): 487-493.
- [17] 刘帅, 朱永利, 张科, 高佳程. 基于误差修正 ARMA-GARCH 模型的短期风电功率预测 [J]. 太阳能学报, 2020, 41(10): 268-275. DOI: 10.19912/j.0254-0096.2020.10.036.
- [18] 张群, 唐振浩, 王恭, 杨阳, 童瑶. 基于长短时记忆网络的超短期风功率预测模型 [J]. 太阳能学报, 2021, 42(10): 275-281. DOI: 10.19912/j.0254-0096.tynxb.2019-1193.
- [19] 刘军涛, 欧阳孜. 基于 ARMA 模型的短期风功率预测研究 [J]. 西北水电, 2017(06): 100-104.

- [20]田波,朴在林,郭丹,王慧.基于 ARMA-GARCH 模型的超短期风功率预测研究[J].电测与仪表,2016,53(17):12-17.
- [21]王玉. 基于深度学习的风电场短期风速预测组合模型[D]. 中国科学技术大学,2021.DOI:10.27517/d.cnki.gzkju.2021.000735.
- [22]季梦凡. 风速和风电场功率组合预测及其不确定性分析[D]. 华北电力大学(北京),2018.
- [23]杨江平. 基于神经网络组合预测的风电场风速及发电功率短期预测[D]. 重庆大学,2012.
- [24]张国强,张伯明. 基于组合预测的风电场风速及风电机功率预测[J]. 电力系统自动化,2009,33(18):92-95+109.
- [25]周松林,茆美琴,苏建徽. 基于主成分分析与人工神经网络的风电功率预测[J]. 电网技术,2011,35(09):128-132.DOI:10.13335/j.1000-3673.pst.2011.09.004.
- [26]于安兴. 风电场短期风电功率预测研究[D]. 华东理工大学,2013.
- [27]王丽婕,冬雷,廖晓钟,高阳. 基于小波分析的风电场短期发电功率预测[J]. 中国电机工程学报,2009,29(28):30-33.
- [28]陈垣毅. 风能及光伏发电功率短期预测方法研究[D]. 浙江大学,2013.
- [29]张丽丽. 基于 K-means 聚类—层次分析的风电场短期功率预测方法[D]. 郑州大学,2019.
- [30]马晓博. 基于小波变换和 BP 神经网络的短期风电功率预测[J]. 电力科学与技术学报,2015,30(02):92-97.
- [31]师洪涛,杨静玲,丁茂生,王金梅. 基于小波—BP 神经网络的短期风电功率预测方法[J]. 电力系统自动化,2011,35(16):44-48.
- [32]陈聪聪,王维庆. 基于小波神经网络法的短期风电功率预测方法研究[J]. 工业控制计算机,2010,23(10):47-48.
- [33]叶林,刘鹏. 基于经验模态分解和支持向量机的短期风电功率组合预测模型[J]. 中国电机工程学报,2011,31(31):102-108.DOI:10.13334/j.0258-8013.pcsee.2011.31.014.
- [34]薛禹胜,郁琛,赵俊华,Kang LI,Xueqin LIU,Qiuwei WU,Guangya YANG. 关于短期及超短期风电功率预测的评述[J]. 电力系统自动化,2015,39(06):141-151.
- [35]杨永恒. 风电场风速和发电功率预测研究[J]. 居舍,2018(12):188.
- [36]黄慧,齐岩. 基于机组动态特性的风电功率预测不确定性量化方法[P]. 河南省:CN113313139A,2021-08-27.
- [37]黄慧,贾嵘,师小雨,王颂凯. 考虑机组动态特性的超短期风电功率预测及不确定性量化分析[J]. 电力系统保护与控制,2021,49(08):109-117.DOI:10.19783/j.cnki.pspc.200750.