



清华大学  
Tsinghua University

# 风光储荷综合系统预测与模拟

电02 肖锦松 2023/06/07

THU



# 目录

## CONTENTS

- 0. Background
- 1. Power Generation
- 2. Energy Storage
- 3. Load
- 4. Conclusion

厚德載物  
自強不息  
Tsinghua

# 0. Background

我国可再生能源资源丰富，以风能、太阳能为主要能源的新型电力系统是我国未来能源电力的发展形态，但可再生能源存在间歇性、波动性的短板，而高比例的可再生能源渗透对电力系统安全稳定运行提出了新挑战。由分布式发电机(DG)和储能系统(ESS)组成的微电网(MG) 为负荷供电，这成为了一个研究的热点。

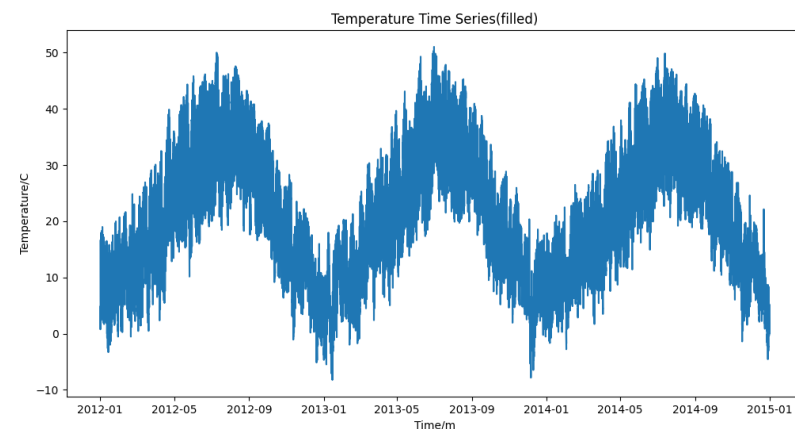
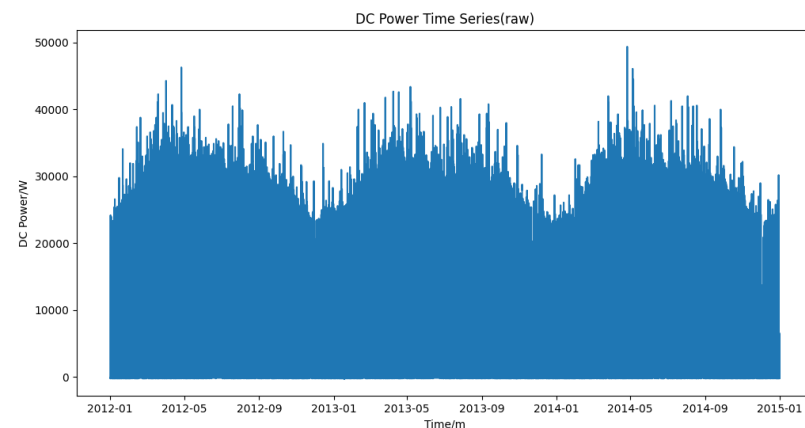
因此，本课程作业希望依次为出发点，争取解决一些风光储综合系统中与人工智能相关的问题。

- 发电侧：光伏出力中直流功率的预测，（风力发电直流功率的模拟）。
- 储能侧：电池单体的寿命预测，（电池储能系统的寿命预测）。
- 用户侧：用户用电量的预测，用户用电模式的聚类。

# 1.Power Generation Photovoltaic(PV) Prediction

数据处理:

- 数据来源于NREL（美国国家可再生能源实验室），描述了拉斯维加斯40kW光伏阵列三年左右的数据。
- 数据缺失值：数据集中存在一些缺失值，但分布较为零散，且量不是很大。最后采用了三次样条插值法将数据缺失值补全。
- 数据特征与目标：特征为环境温度、逆变器温度、模块温度、辐照度、相对湿度、风向、风速；目标为光伏阵列的直流输出功率。
- 预测方法：多元线性回归、LSTM。



# 1.Power Generation Photovoltaic(PV) Prediction

预测方法:

- 多元线性回归: 数据中给定七个特征, 尤其是环境温度和辐照度与预测目标即输出功率有着较强的**相关性**。因此可以考虑用多元线性回归的方法进行预测。
- LSTM: 适合于处理时间序列数据而光伏出力往往具有时序关系, 即过去的值对当前和未来的值可能有所影响。由于训练集数据量很大, 因此考虑采取一些防止过拟合的手段, 这里首先是对训练样本进行了**重采样** (15min->1h); 采用**Dropout的正则化方法**, 用于缓解神经网络的**过拟合问题**。在每个训练批次中, Dropout层按一定的比例随机选取丢弃输入数据, 使得模型不会过度依赖于某一些神经元。这里选取丢弃比例为0.4。

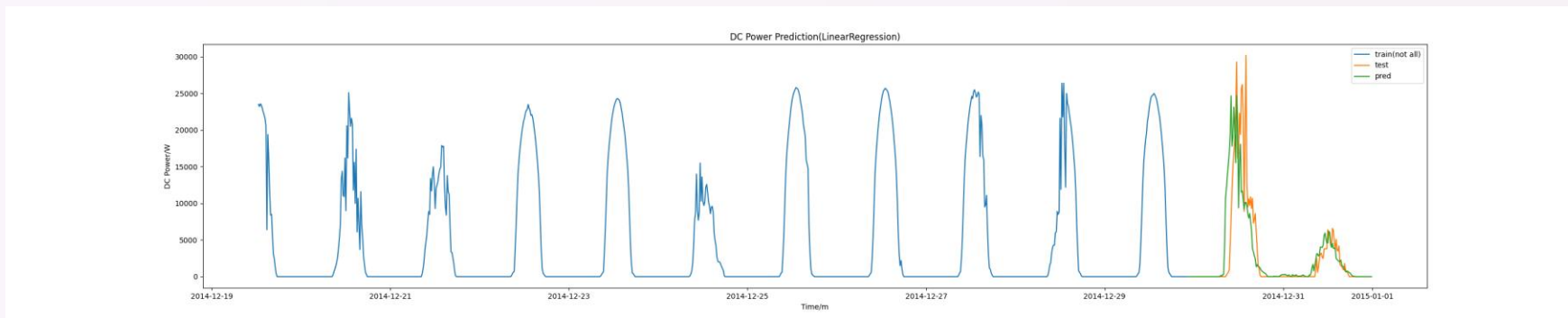
```
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

```
model = Sequential()
model.add(LSTM(64, input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dropout(0.4))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam')
history = model.fit(X_train, y_train, epochs=10, batch_size=10, validation_data=(X_test, y_test), verbose=2, shuffle=True)
```

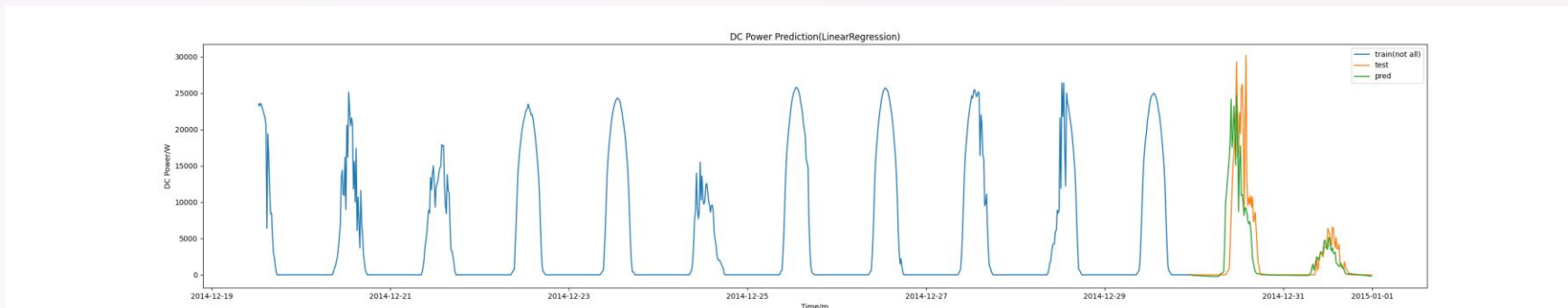
# 1. Power Generation Photovoltaic(PV) Prediction

预测结果: RMSE:3887 vs 3320

- 多元线性回归



- LSTM

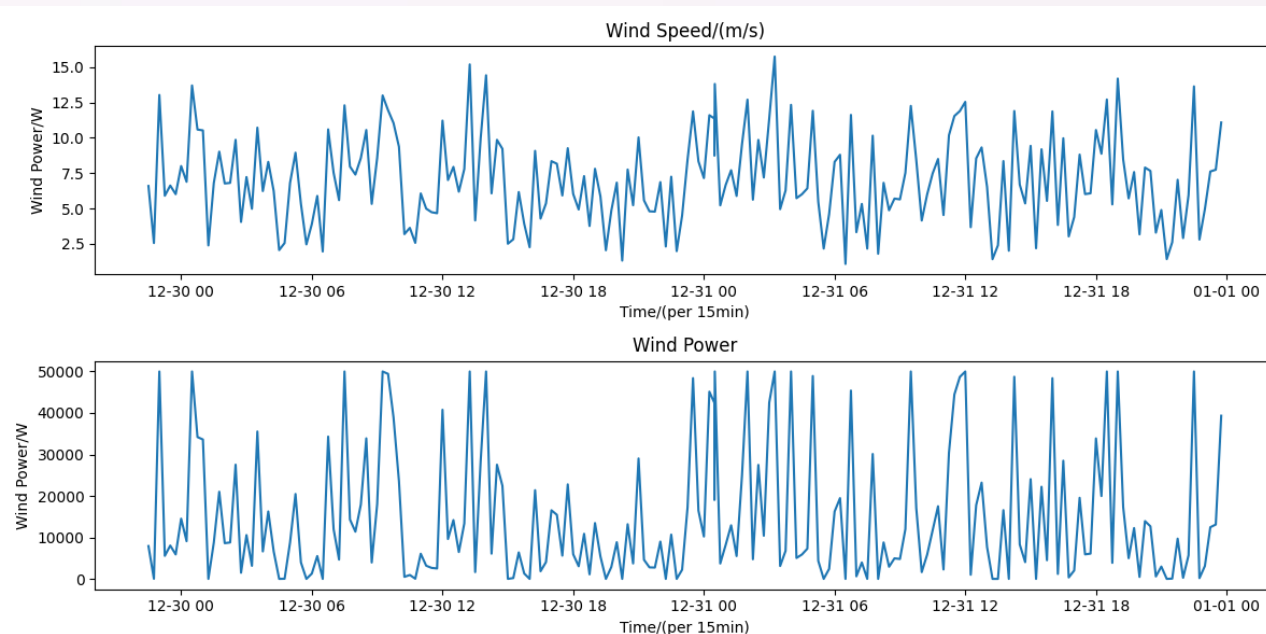




# 1.Power Generation Wind Power Simulation

由于仅有光伏出力无法满足功率需求，因此在光伏出力的基础上叠加风电，保证在光照条件差的情况下仍可以提供充足电量。以此为基础，模拟风-光-储混合的微电网系统。风机的输出功率与风速的关系可以用分段函数近似。

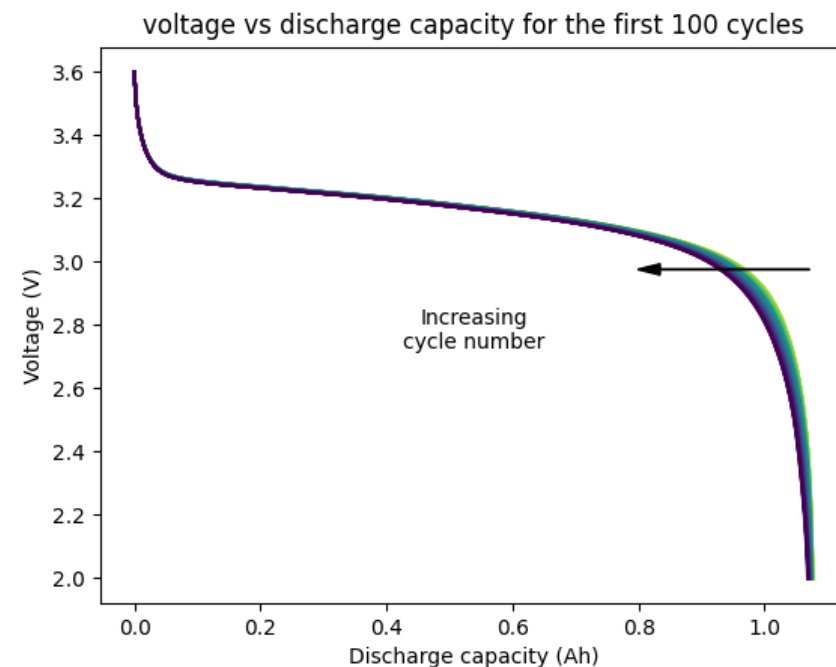
风速选择用**双参数Weibull模型**模拟，选择 $k=2.2$ 、 $\lambda=8$ ，利用上述关系获得风机出力情况。



## 2. Energy Storage battery lifetime prediction

数据处理:

- 数据集来源于Severson et al., 其中包括41 cells的training set, 43 cells的test set1, 40 cells的test set2。数据集中有前100次循环, 每次循环中有电压从3.6V下降到2.0V的1000个容量数据。目的是利用前100个循环的数据预测经对数转换的电池循环寿命。循环寿命被定义为容量低于额定容量的80%之前的循环次数。
- 可以发现, 随着循环次数的增加, 在相同的电压水平下, 容量逐渐递减; 在相同的容量水平下, 电压逐渐递减。





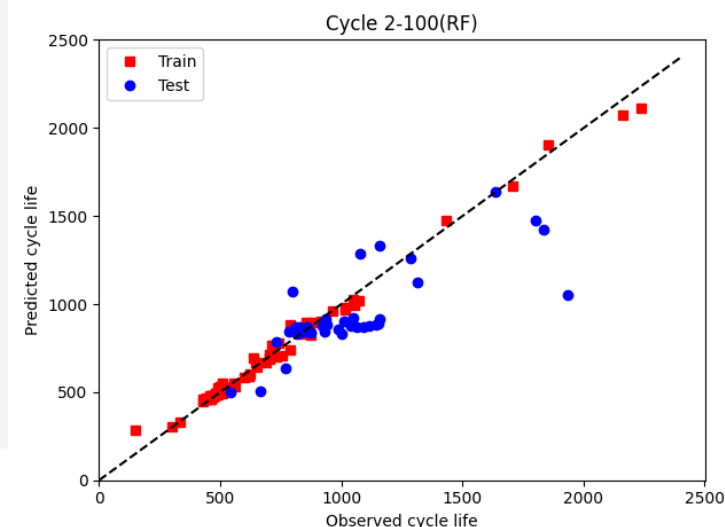
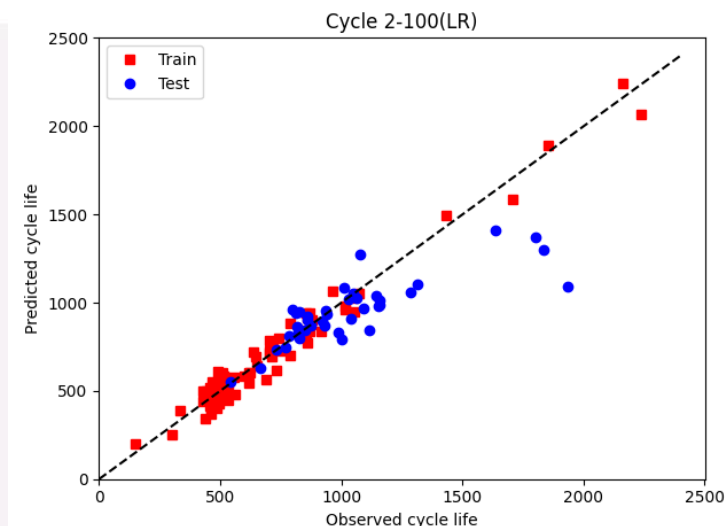
## 2. Energy Storage battery lifetime simulation

预测方法:

- 随机森林回归: 选择决策树的数量为100, 限制树的深度为300。再继续增加对预测精度的提高微乎其微。
- 同样地, 与多元线性回归进行对比

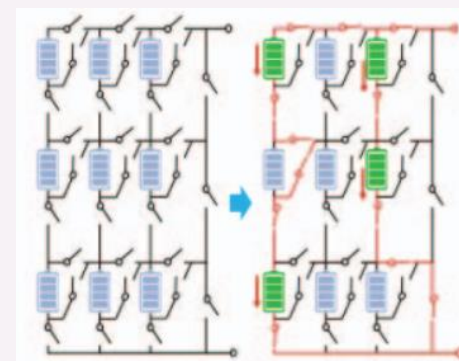
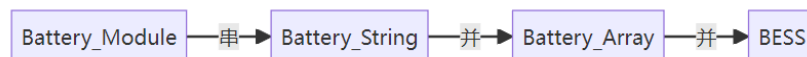
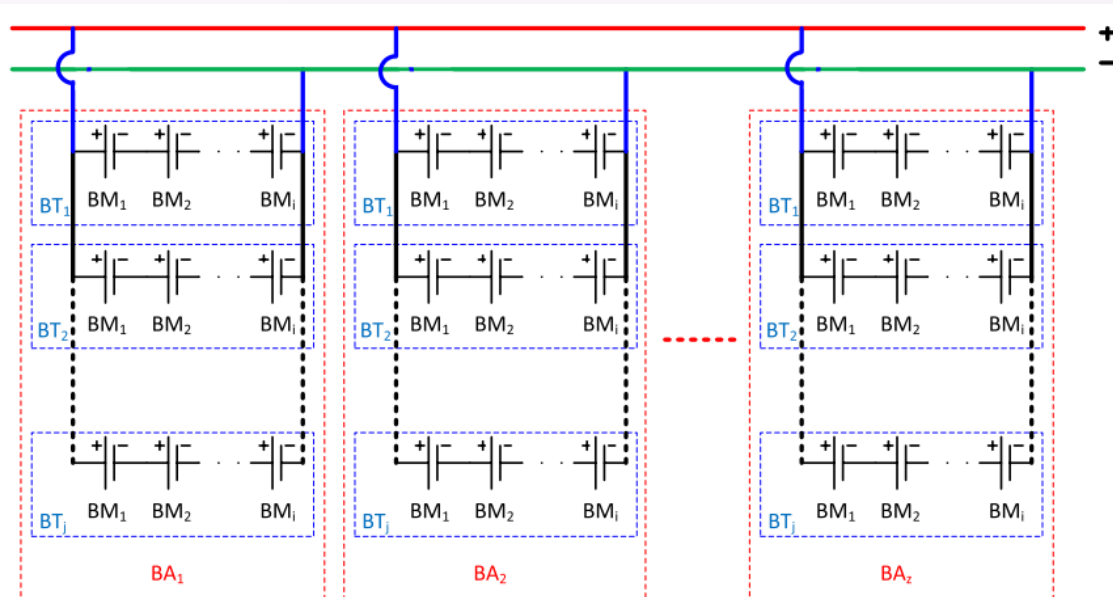
LR 均方误差: 42802.67632406935  
LR 均方根误差: 206.8880768049946  
LR 平均绝对误差: 127.93995893212113  
LR R\_square: 0.5390233147636393

RF 均方误差: 42453.82353061224  
RF 均方根误差: 206.04325645507603  
RF 平均绝对误差: 141.99214285714288  
RF R\_square: 0.5427803930160773



## 2. Energy Storage BESS lifetime calculation

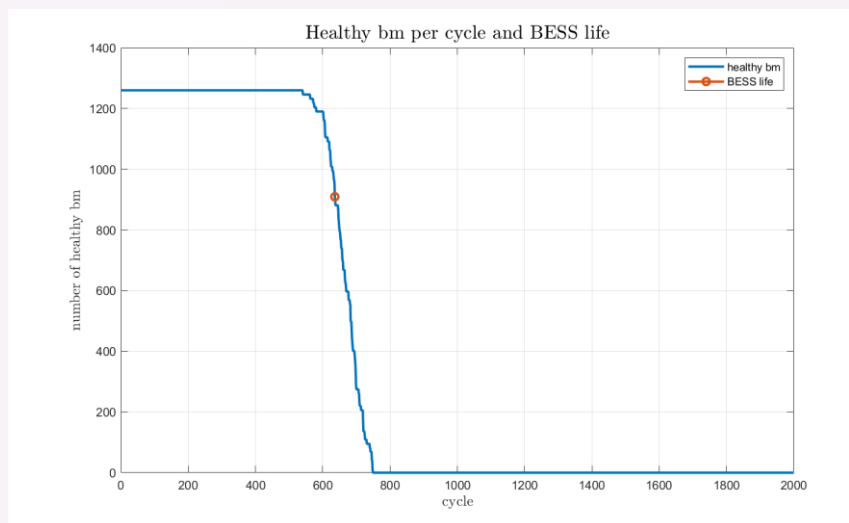
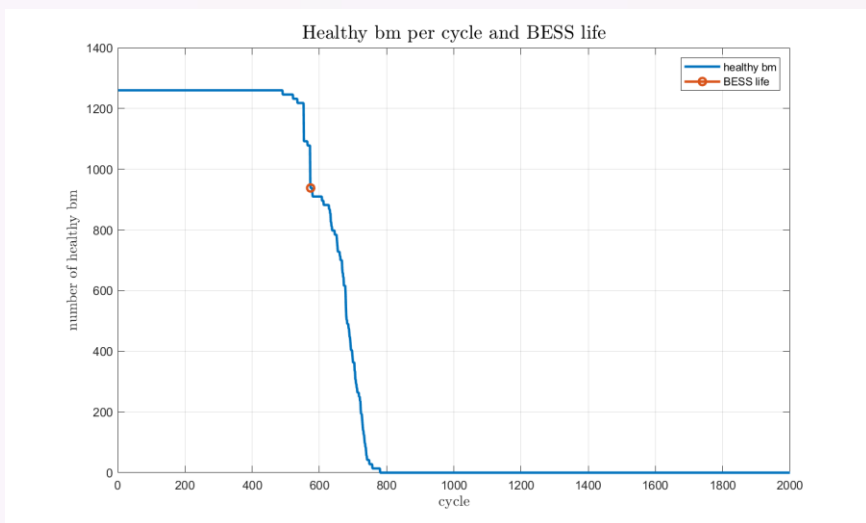
电池寿命预测的意义：对电池寿命用随机森林回归的方法进行预测，我们只需要得到电池前100个充放电循环中的电气数据即可利用训练好的模型进行寿命预测。我们将预测的寿命当作是当前电池的健康状态，即SOH(State of Health)新出厂电池为100%，完全报废为0%。此时相当于SOH变为可预测的，我们可以优先使用SOH高的电池，将所有电池的SOH趋于一致。这对分析电池储能系统的可靠性非常有帮助。



## 2. Energy Storage BESS lifetime calculation

蒙特卡洛模拟法：对BESS各个battery module进行全生命周期的抽样模拟，以电池充放电循环次数作为其寿命，在每个充放电循环下根据上一个充放电循环故障的电池数量对剩余电池的故障率进行重新评估，计算出故障概率。以此作为该充放电循环下的预期故障概率，继续进行这次抽样。

左：battery string里面的一个module故障后，整个battery string也会随之故障。  
右：可重构电池组，每串配用1个冗余电池，并优先使用预期寿命短的。



## 3. Load

## load prediction

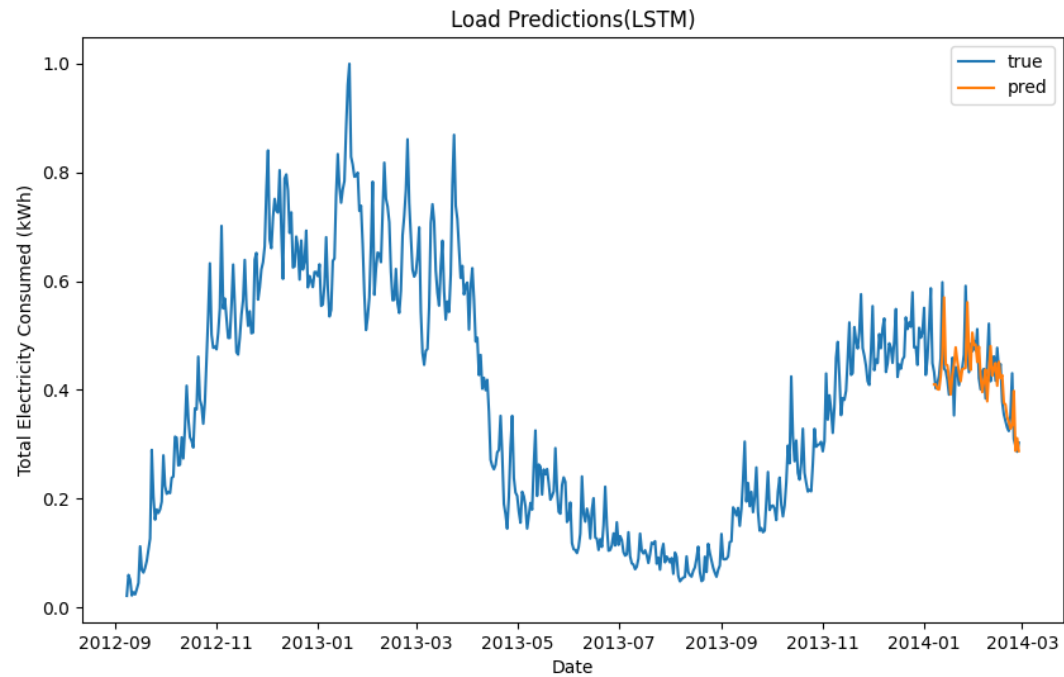
数据处理：

- 数据来源于Kaggle (Smart meters in London) : daily\_dataset: 统计了每天电能的使用情况，包括总量、平均值、中位数、最大值、最小值、标准差等等；acorn\_details: ACORN指的居民区的一些分类，数据中包含了居民的个人信息；household\_info: 包含了家庭的所有信息；
- 数据缺失值：缺失是因为某些用户没安装智能电表，所以在此之前的数据全为空。选择一个超过一半人安装智能电表的时间节点开始，**缺失用户则全部删除。**
- 数据特征与目标：特征为每天用户电量的最大值、最小值、标准差等等，目标为每天用户用电的总量。
- 预测方法：LSTM。

### 3. Load

### load prediction

预测结果:



### 3. Load

### load clustering

数据中有一个表格记录了每个用户的ACORN。ACORN作为一种英国伦敦的人口统计工具，将家庭、邮政编码和社区划分为 6 个类别、18 个组别和 62 个类型。

Affluent Achievers、Rising Prosperity、Comfortable Communities、Financially Stretched、Urban Adversity、Not Private Household。

电力水平或用电量经常能够反应一个国家的工业化水平和经济水平。因此可能能够解释不同经济水平的用户的消费模式，验证ACORN与用电模式之间的相关性。

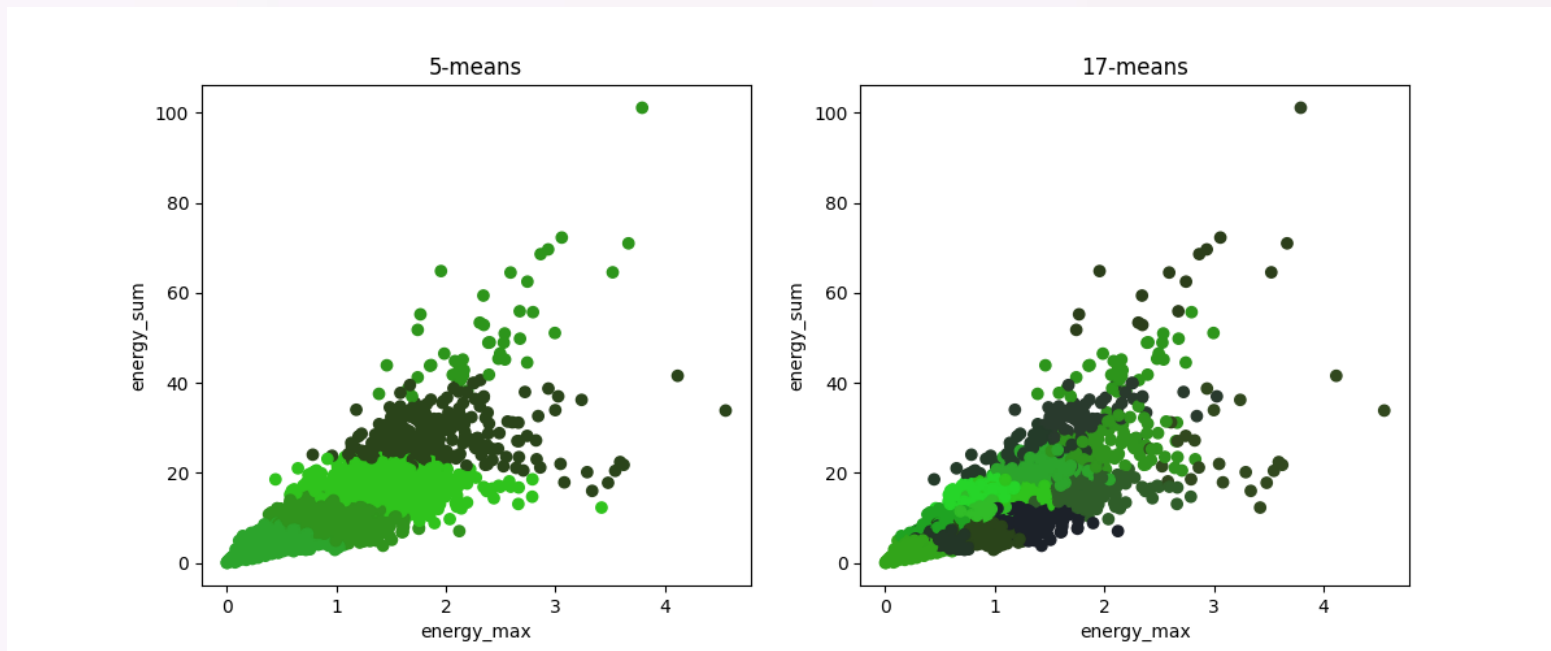
撇除ACORN不典型的第6类居民：not private households，则ACORN对居民的分类变为5个类别(categories)，17个组别(groups)，我们首先以5-means and 17-means来对居民用电模式进行聚类。



### 3.Load

### load clustering

聚类结果如下：似乎效果都不是特别好。

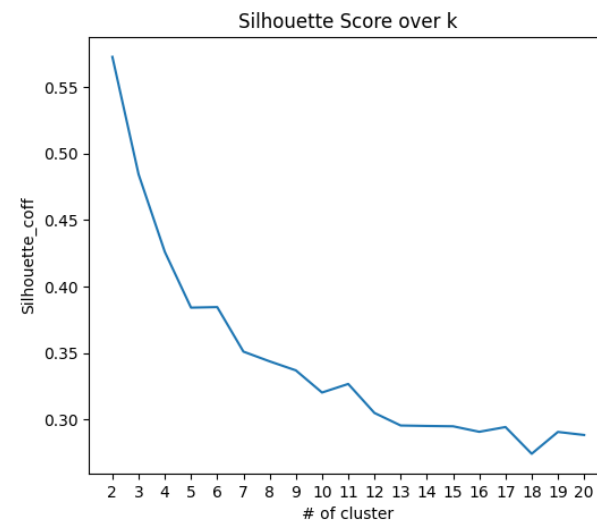
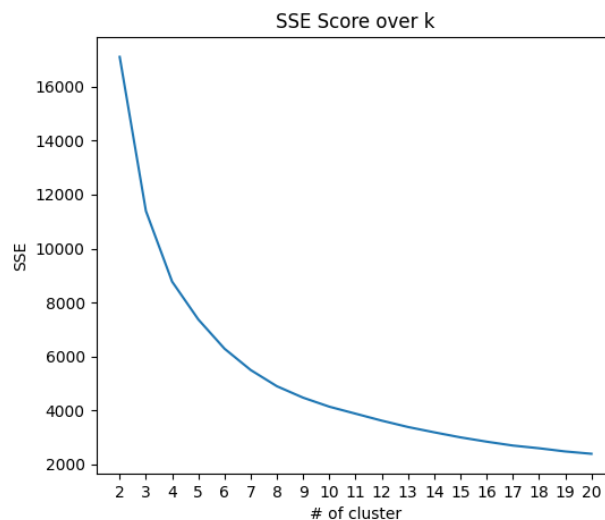


### 3. Load

### load clustering

试图选择出**最佳簇数k**，参考How to Determine the Right Number of Clusters。采用第一种方法#1: Within-Cluster Sum of Squares (WSS)，遗憾的是，这些数据似乎并没有出现比较明显的所谓的“拐点”（左图）；采用第二种方法#2 Average Silhouette Score（右图），可以发现轮廓系数最高的k<10依次为2、3、4、6、5.....，用电量能反映社会的贫富2个阶层的差距，后续可能是6个阶层分数更高一些；对于k>10，轮廓系数得分比较高的为11、10、12、13.....。

	k	SSE	Silhouette_coff
0	2	17112.724316	0.572939
1	3	11394.392881	0.484500
2	4	8772.158071	0.426185
4	6	6282.147377	0.384569
3	5	7367.648754	0.384159
5	7	5490.297224	0.351035
6	8	4888.765951	0.343715
7	9	4464.929330	0.336922
9	11	3871.048213	0.326695
8	10	4130.060755	0.320229
10	12	3612.492229	0.304858
11	13	3376.759538	0.295415
12	14	3179.619099	0.295067
13	15	2992.945159	0.294814
15	17	2688.676733	0.294242
14	16	2832.419110	0.290665
17	19	2469.031787	0.290563
18	20	2382.710786	0.288272



### 3. Load

### load clustering

虽然试图利用K-means聚类来验证伦敦居民的ACORN类型划分似乎不太可行，但本工作仍具有意义。居民类型划分并不是及时更新的，我们或许可以通过居民的用电模式聚类来试图寻找新的划分，可能有原先的Affluent Achievers以及Rising Prosperity因为投资失败破产，或是原先的Urban Adversity获取了大量财富，那么原先的分类无法反应他们当前的类型，而用电量能够反映当下比较真实居民类型划分。或许可以将其作为新一轮的人口普查、人口分类的参考资料。

## 4. Conclusion

- 《人工智能导论》这一门课是本专业知识以外很好的一个补充，也为我今后的学习增加了一些有用的工具。也谢谢助教在课后对我问题的解答。
- 虽然做了发电侧、储能侧、负荷侧一些比较初步的预测工作，但是由于使用的数据的来源都不同，难以进行进一步的分析。比如发电和负荷并不匹配、对储能的充放电具体过程并没有做出模型等等。

**THANK  
YOU**

