

电力负荷大数据分析

——第三小组期末作业

杨焜棋 电 15 2021011488

胥职体 电 15 2021011489

蔺国欣 电 15 2021011480

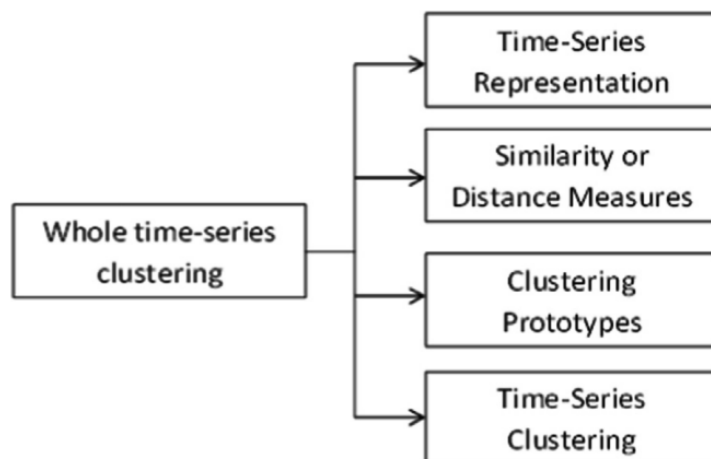
2023 年春

1 基本任务一：基于负荷历史数据的用户用能模式分析

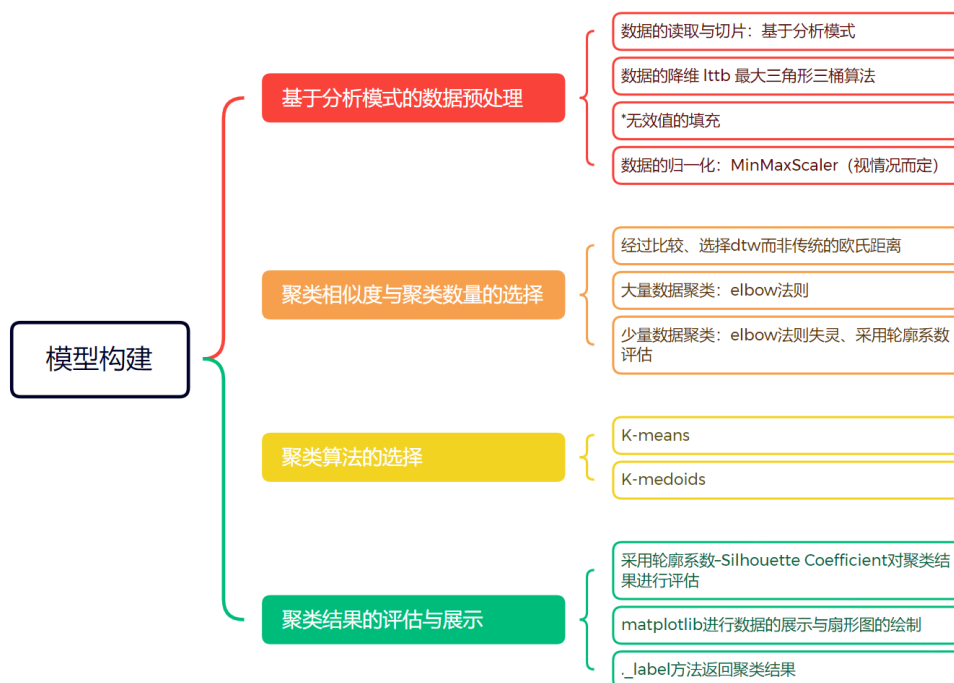
题目回顾：基于数据集 1 中的家庭智能电表与天气数据，统计分析不同家庭的用能行为模式差异，对不同的用户负荷进行聚类分析，尝试从中挖掘具有解释性的结论。

1.1 聚类模型的构建

1.1.1 总体框架



基于 Saeed 等人在关于时间序列聚类的综述中提及的模型组成（如上图），本组时间序列聚类模型的总体如下：

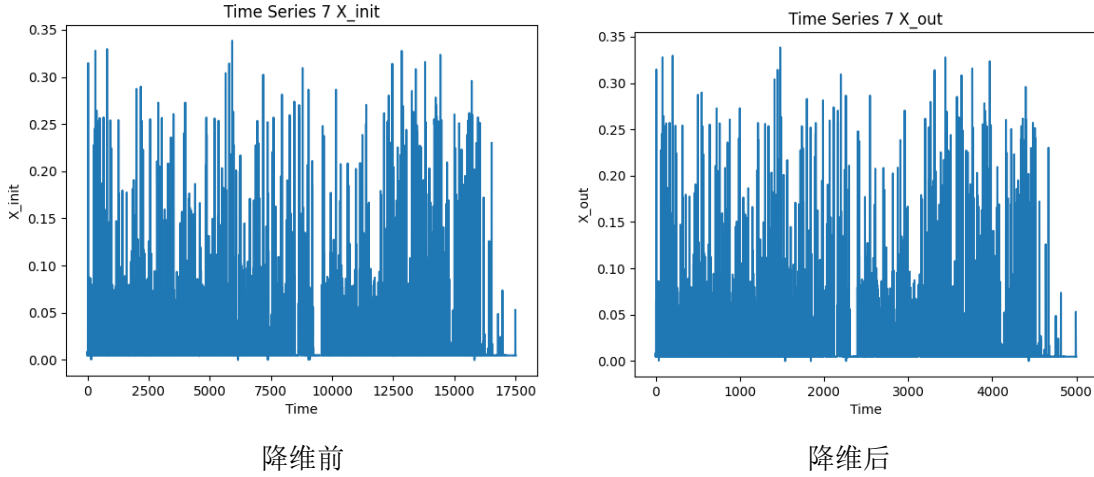


1.1.2 数据降维与归一化

从.csv 文件中读取的数据被转化为 numpy 数组分析，具体的读取与切片方式会在第二

部分介绍。

由于一年的时间序列数量过于庞大，这会极大增加运算压力，因此找到一个能保证相似度的降维方法十分重要。在查找文献之后，本模型中使用了 lttb（最大三角形三桶算法）来对时间序列进行降维，结果如下。可见降维前后的时间序列基本一致。



数据的归一化采用 sklearn 库中的 MinMaxScale 处理。

1.1.3 相似度的选择

不同于传统对于单个数据点的聚类，常用的欧几里得度量不适合时间序列。如果两个时间序列高度相关，但其中一个被移动了哪怕一个时间步长，欧几里得度量就会错误地测量它们的距离。在时间序列聚类中，更常使用的方法是 dynamic time warping (DTW, 动态时间规整)。其原理如下：

DTW between x and y is formulated as the following optimization problem:

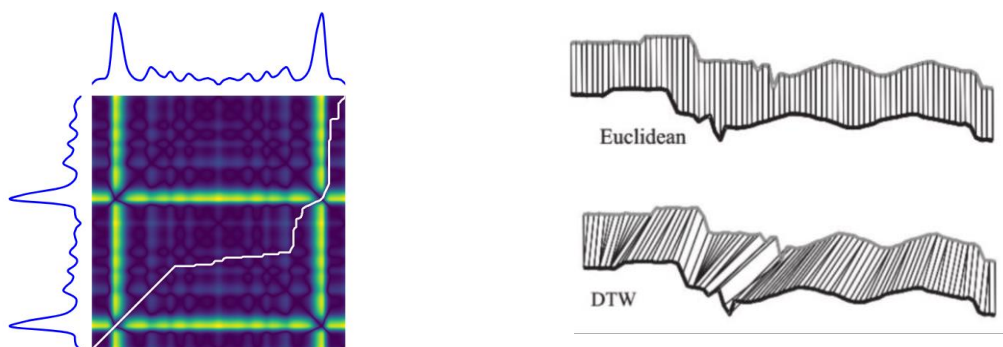
$$DTW(x, y) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} d(x_i, y_j)^2}$$

where $\pi = [\pi_0, \dots, \pi_K]$ is a path that satisfies the following properties:

- it is a list of index pairs $\pi_k = (i_k, j_k)$ with $0 \leq i_k < n$ and $0 \leq j_k < m$
- $\pi_0 = (0, 0)$ and $\pi_K = (n-1, m-1)$
- for all $k > 0$, $\pi_k = (i_k, j_k)$ is related to $\pi_{k-1} = (i_{k-1}, j_{k-1})$ as follows:
 - $i_{k-1} \leq i_k \leq i_{k-1} + 1$
 - $j_{k-1} \leq j_k \leq j_{k-1} + 1$

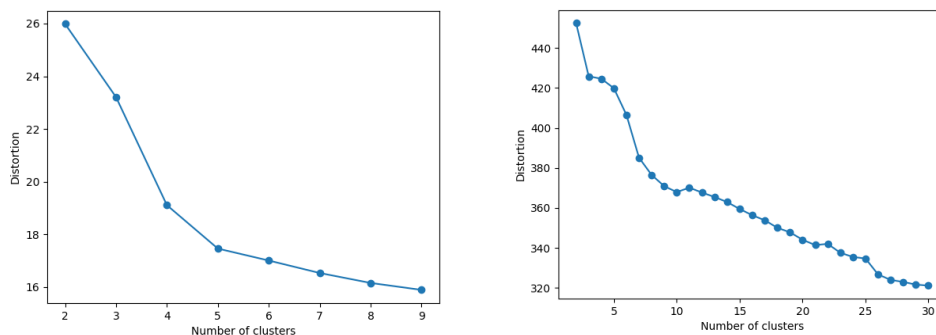
简而言之，它将两组时间序列张成一个二维的矩阵之后，试图找出一个路径（对应关系），使得总体得到的距离最小（见下左图）。

与欧式距离不同，DTW 寻找的“相关点”并不是严格一一映射的，而是通过错位映射使得整体的距离最小（见下右图）。

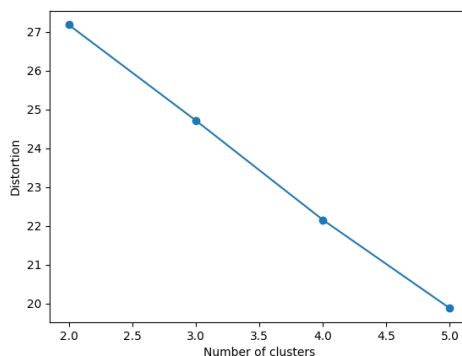


1.1.4 聚类数目的选择

聚类数目的选择采用“肘部法则” (Elbow Method)，将每个簇的质点与簇内样本点的平方距离误差和称为畸变程度 (distortions)，那么，对于一个簇，它的畸变程度越低，代表簇内成员越紧密，畸变程度越高，代表簇内结构越松散。畸变程度会随着类别的增加而降低，但对于有一定区分度的数据，在达到某个临界点时畸变程度会得到极大改善，之后缓慢下降，这个临界点就可以考虑为聚类性能较好的点。Distortion 下降的情况举例如下，比如在对用户 B 的 2015 年所有家用电器进行聚类分析时（下左图），分成 5 类比较合适。对用户 B 的 2016 年总用电情况进行按日切片并聚类分析时（下右图），分成 10 类比较合适。



当然，如果时间序列总量较少，则可能遇到肘部法则失灵的情况，如下图：



此时采用轮廓系数 (Silhouette Coefficient)，它是类的密集与分散程度的评价指标。

可以在试验的聚类数目中选取轮廓系数最大的那个。轮廓系数也被用于对于聚类结果的评估，其公式如下：

$$S = (b - a) / \max(a, b)$$

1.1.5 聚类算法选取

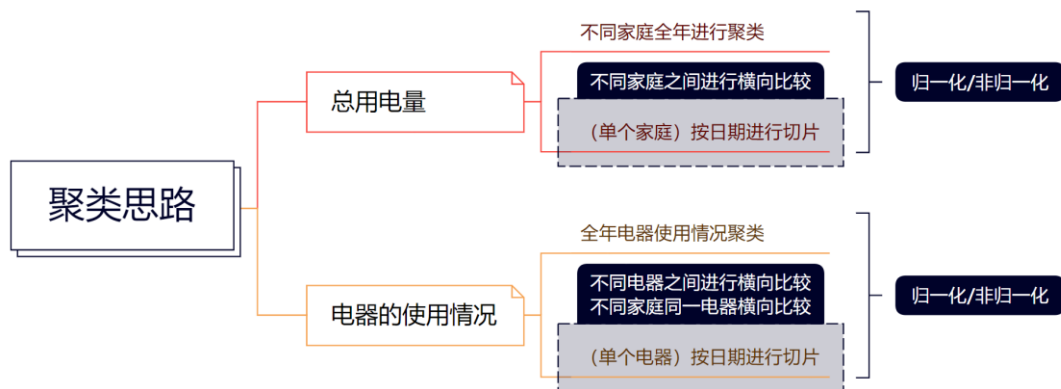
参考文献综述中给出的结论（如下两图），选取 K-means 和 K-Medoids 两种算法进行聚类，分别使用 sklearn 和 tslearn 两个库。

A partitioning clustering method makes k groups from n unlabelled objects in the way that each group contains at least one object. One of the most used algorithms of partitioning clustering is **k-Means** [169] where each cluster has a prototype which is the mean value of its objects. The main idea behind k-Means clustering is the minimization of the

Prototype in k-Means process is defined as mean vector of objects in a cluster. However, when it comes to time-series clustering, it is a challenging issue and is not trivial [149]. Another member of partitioning family is **k-Medoids (PAM)** algorithm [150], where the prototype of each cluster is one of the nearest objects to the centre of the cluster. Moreover

1.2 数据处理及结果分析

1.2.1 整体思路



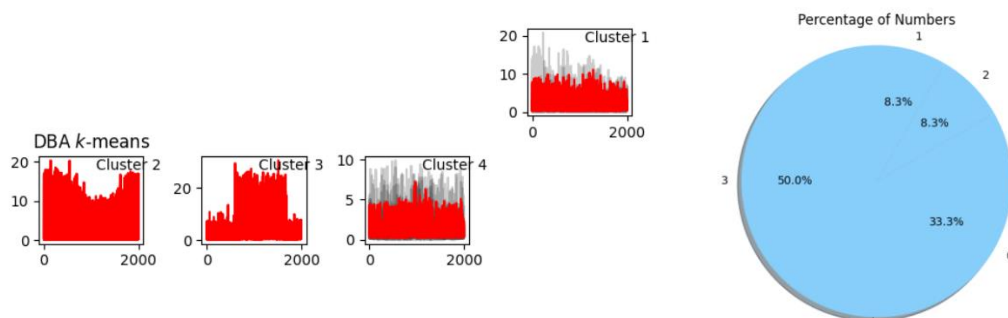
1.2.2 归一化/非归一化

在模型的设计中加入了归一化的函数，但是并不是在每次聚类过程中都要调用，是否对数据进行归一化处理，要看运营商对于数据的需求。在选做任务一中，我们选择的利益主体是负荷聚合商。对于这一主体来说，可以通过归一化的数据评估不同用户的负荷弹性和响应时间，通过将不同用户的功率数据标准化到相同的尺度上，可以消除因功率范围不同而导致的差异。而通过非归一化的数据可以评估用户能够提供的响应容量，即用户能够降低或增加的用电量范围，因为此时强调的是功率水平的绝对值。

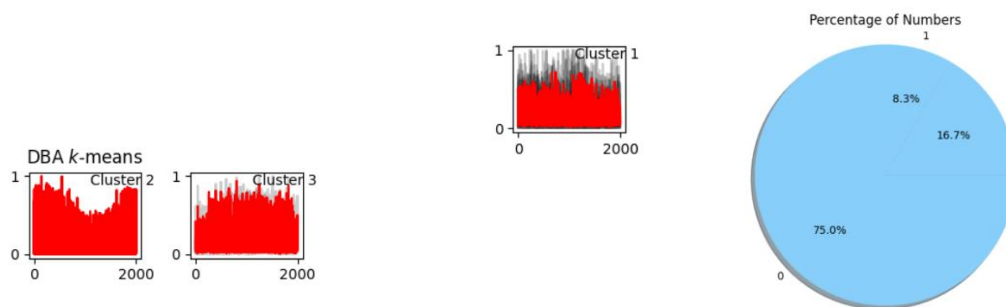
1.2.3 所有用户 2014-2016 全年总用电量进行聚类分析

分别将归一化的和非归一化的数据聚类并分析。在非归一化的结果中可以看到，大多数家庭都属于第一类和第四类，全年用电情况分布较均匀且平均功率在 5-10kW，而 HomeG 被单独分到第二类，特点为夏季用电出现明显波谷，而全年平均用电在 15-20kW，因此这户家庭

很有可能提供更大的负荷弹性容量。；类似的，对于 HomeF2016 年的情况而言，其冬天用电水平趋于正常居民用电水平，但是在其他季节则表现出明显的高额用电。

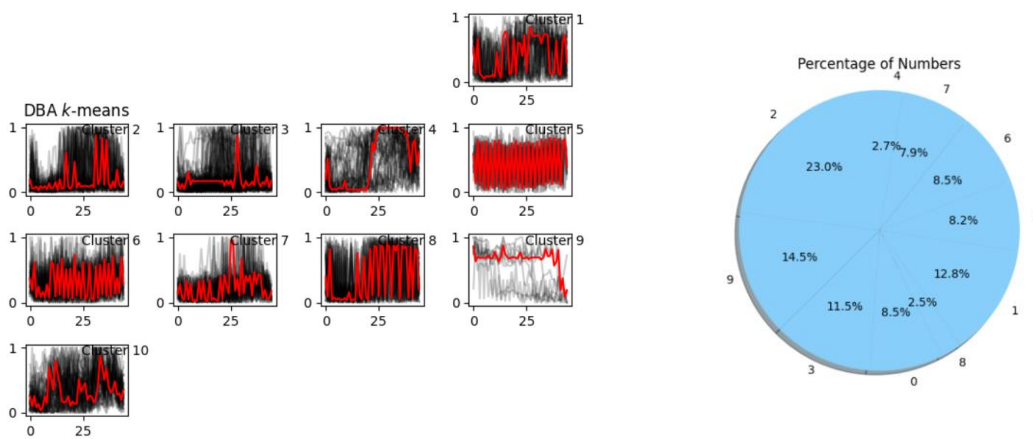


再看归一化之后聚类的结果，其中，家庭 B/C/D/E 在 2014-2016 年都属于第一类，可以看出有一定波动幅度，这说明这些家庭的弹性负荷能力较好。HomeG 仍被单独分到第二类，特点为夏季用电出现明显波谷；而 HomeF2015、2016 两年的情况被分至第三类，表现出的波动性较差。虽然 HomeF 为高额用电用户，但是有可能并不能提供较大的负荷弹性。

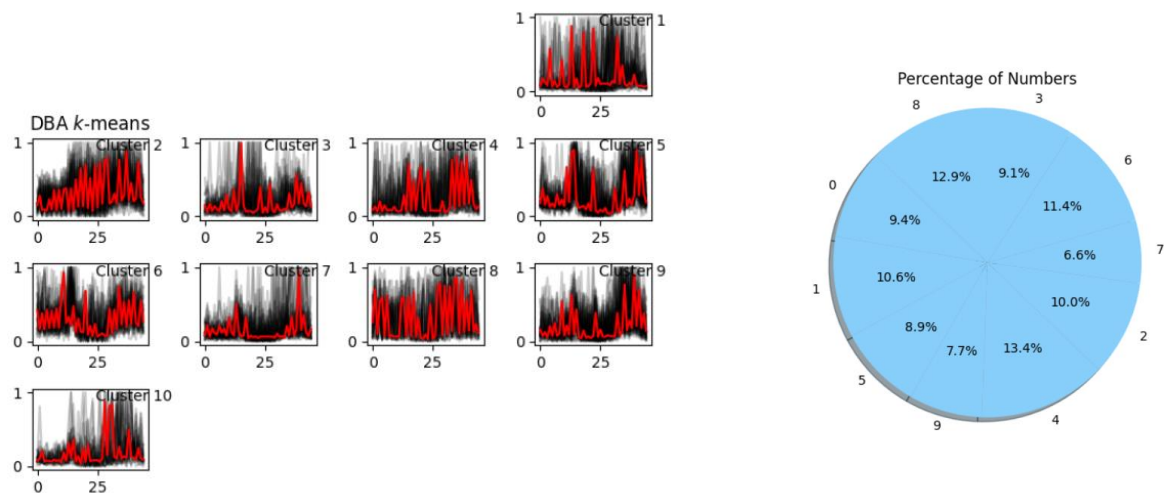


1.2.4 单个家庭总功率按日期进行切片并聚类之后横向比较

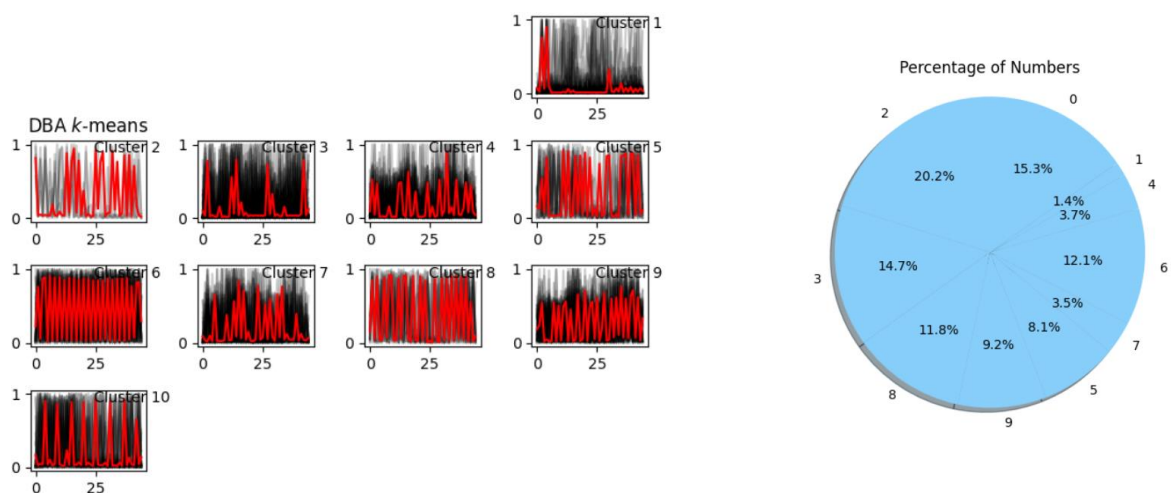
上述全年总用电量的分析只能大体上描述出用户的负荷弹性程度，想要对用户进行更细致的画像，这里对单个家庭按日期进行切片并聚类之后再进行横向比较（均为 2016 年）。



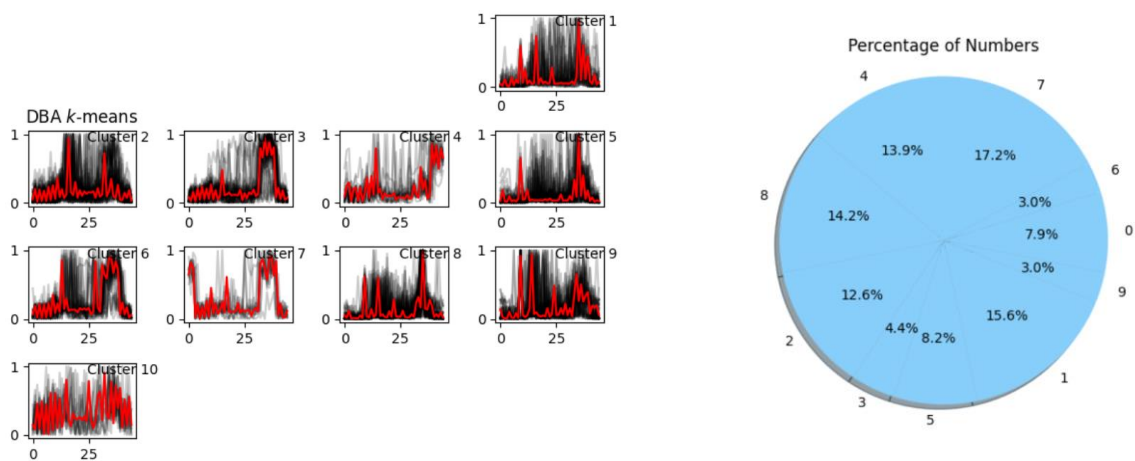
HomeB 按日切片后聚类结果（归一化）



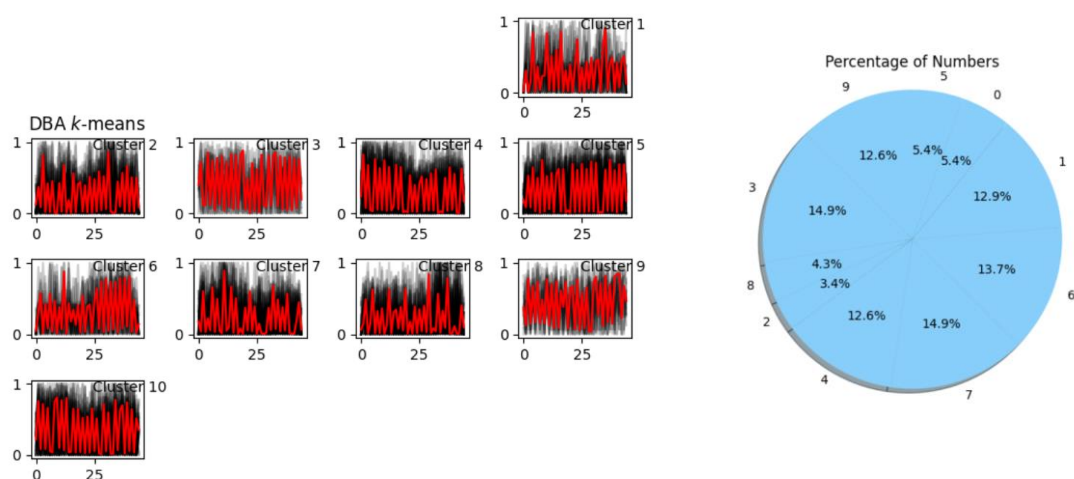
HomeC 按日切片后聚类结果（归一化）



HomeD 按日切片后聚类结果（归一化）

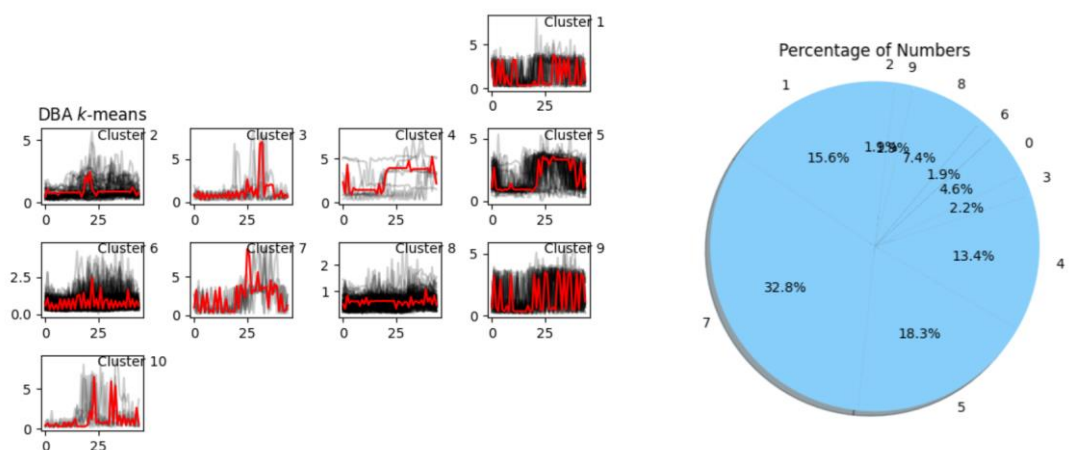


HomeF 按日切片后聚类结果（归一化）

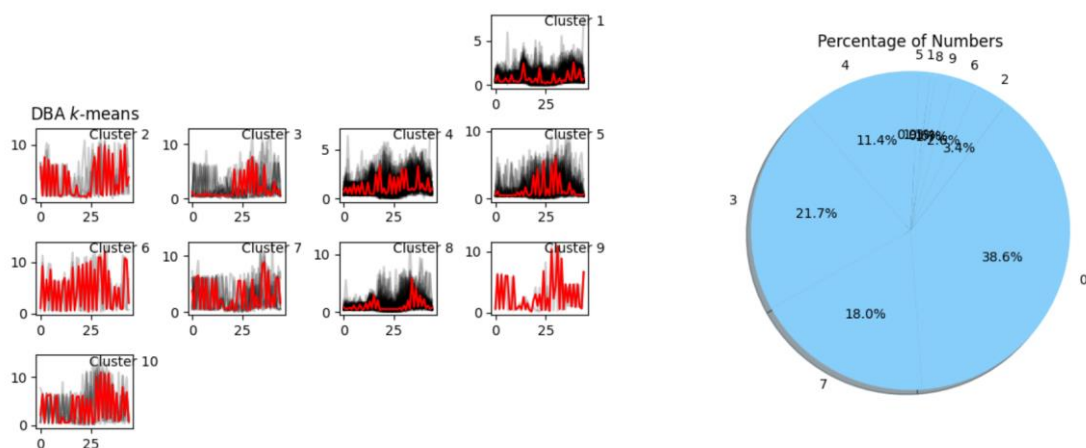


HomeG 按日切片后聚类结果（归一化）

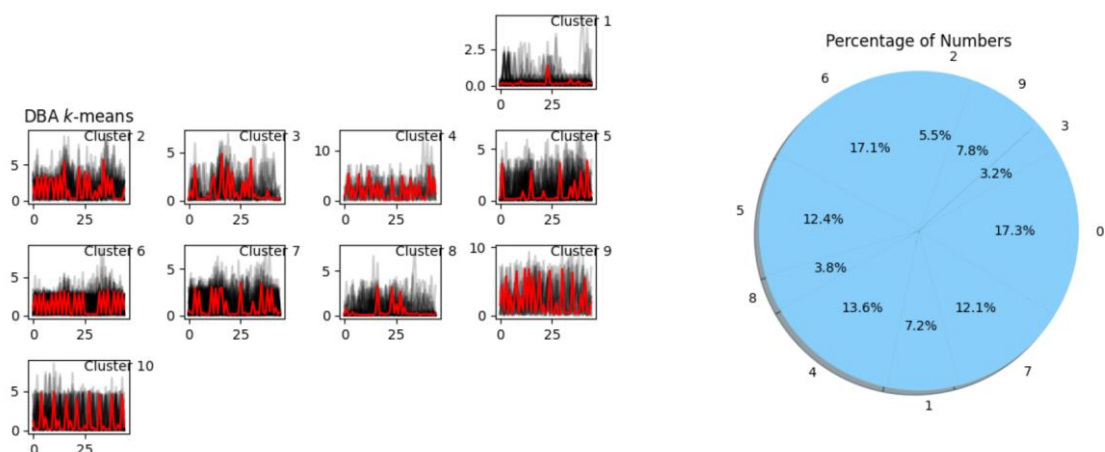
可以看出，HomeD、HomeG 在一年中的大部分时间，用电负荷情况呈现明显的波动性；而 HomeB、HomeC 则大多数时候呈现出用电的规律性，波动性用电的情况很少；而 HomeF 则几乎没有出现用电波动，每一天的峰谷都呈现出很强的规律性。



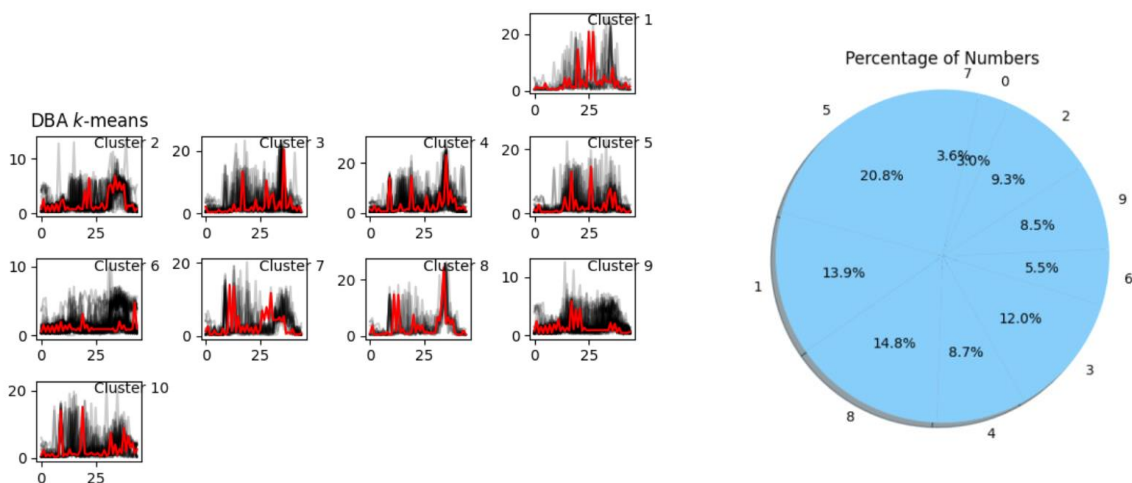
HomeB 按日切片后聚类结果（非归一化）



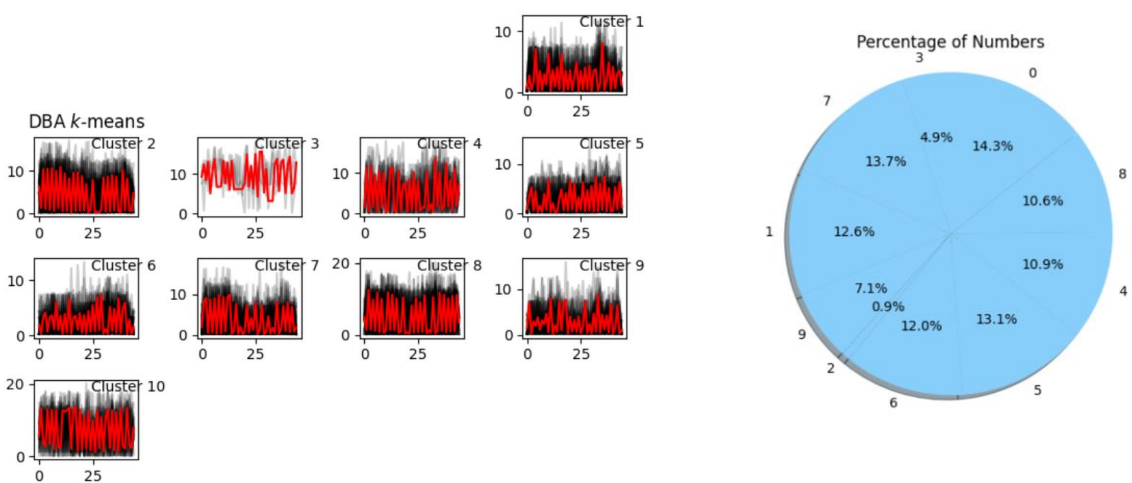
HomeC 按日切片后聚类结果（非归一化）



HomeD 按日切片后聚类结果（非归一化）



HomeF 按日切片后聚类结果（非归一化）



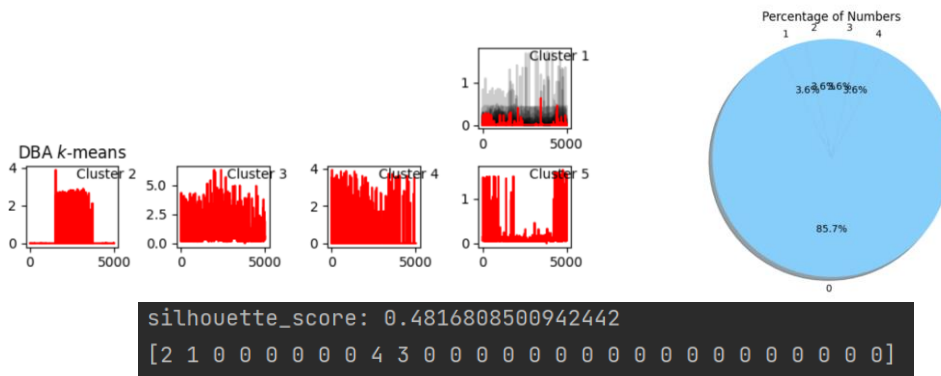
HomeG 按日切片后聚类结果（非归一化）

HomeB 大多数情况下用电水平较低；而从非归一化的数据结果来看，HomeC 一年中有超

过 50% 的日子用电波动性较强，且用电水平在 5kW-10kW 之间，故与 B 相比能提供更高的弹性；HomeD、HomeG 在一年中的大部分时间，用电负荷情况仍然呈现很明显的波动性，都具有良好的负荷弹性，同时从绝对功率值来看，HomeG 可以提供的响应容量更大。

1.2.5 全年电器使用情况

家庭与负荷聚合商合作的直接手段是控制某些家用电器的通断，某些设备或过程可能具有较高的弹性，可以在不影响正常运行的情况下进行调整，例如空调温度调节、照明控制等。因此对某些家用电器的使用情况进行画像也可以帮助负荷聚合商评估一户家庭可能的负荷弹性。这里以 2015 年用户 B 为例，对该用户所有家用电器进行聚类分析。这里对电器进行非归一化处理更好，这是由于不同电器使用功率差异较大，这样可以区分出各个用电器能够提供的响应容量。



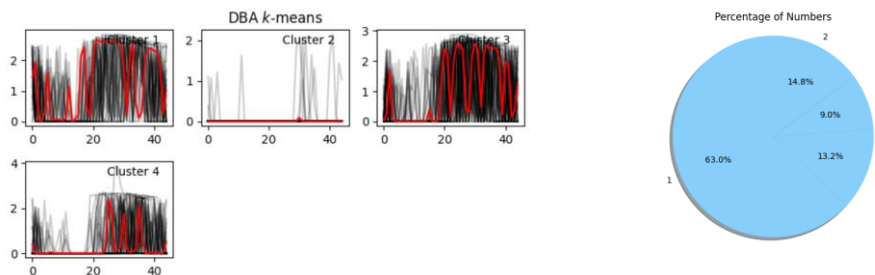
以上的 list 表示表格中对应的某一个电器被分到了哪一类。

可见大部分家用电器最多可以提供 1kW 左右的容量，且使用时间并不固定，这里面包括了各个房间的灯具、洗衣机等；而可提供响应容量功率最多，在图中表现地最有弹性的为空调（2 表示被聚到了第三类）。

1.2.6 单个家用电器按日期进行切片并聚类之后横向比较

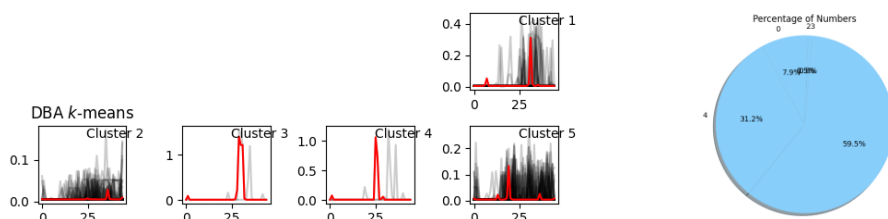
根据 1.2.5 中设备控制方法和时段特性，可以对负荷进行如下的划分：对功率可变且用电时段分布较为固定的设备，认为其为可调型设备，如空调、智能照明设备等；对功率不可变，即仅为开/关状态的用电设备中，日用电时间波动较大，用电时段分布较为分散的设备，则认为其为可停型设备，如常规照明设备、电视等；对功率不可变，日用电时间基本恒定，用电时段分布较为分散的设备，则认为其为可转移型设备，如热水器、微波炉、电磁炉等；对于功率不可变，且用电时间稳定，分布时段广泛均匀的设备，策略化的可调节空间较小的设备，认为其是不可控型设备，典型如部分用户的冰箱、公共区域照明设备等。而用户的用电弹性来源，也可以认为是依照以上介绍的顺序依次递减的。

根据以上结论，这里以空调和微波炉为例进行分析。

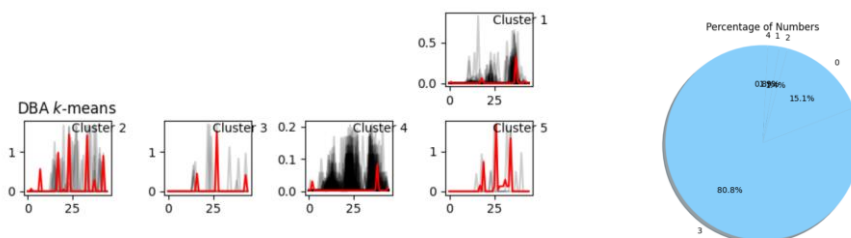


用户 B 的空调 2015 年使用情况进行按日切片（非归一化）

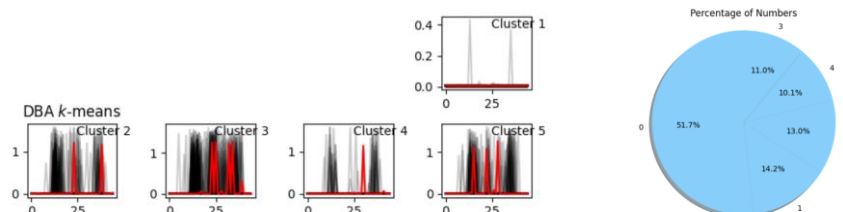
可见空调在使用期间呈现出了很好的波动性，而且可以提供的容量在 2-4kW，是比较优秀的用于参与负载聚合商合作的用电器。



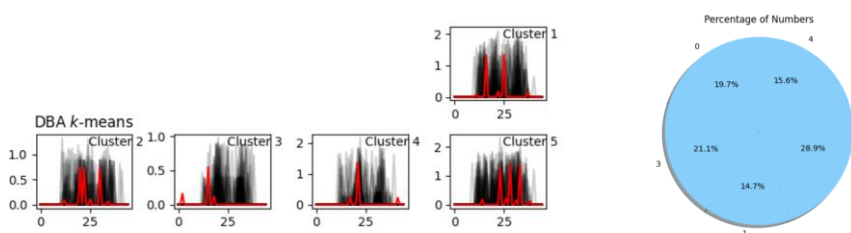
用户 B 的微波炉 2015 年使用情况进行按日切片（非归一化）



用户 C 的微波炉 2015 年使用情况进行按日切片（非归一化）



用户 D 的微波炉 2016 年使用情况进行按日切片（非归一化）



用户 F 的微波炉 2016 年使用情况进行按日切片（非归一化）

横向比较以上用户的微波炉使用情况，可以发现微波炉的使用具有以下特征：每次使用的功率需求不同，但在一次使用过程中功率需求稳定，可以看作“可调档的电器”；使用通常是短暂的且为间歇性的；通常表现为二到三个使用高峰。而四户家庭对于微波炉使用的情況不同主要体现在每次使用对于微波炉功率的需求上。比如 F 集中在 2kW 左右，而 B 则集中在 0.2kW 左右，差别较大，但是在使用高峰的分布上并没有明显的差别。对于这类设备可以通过施加激励鼓励用户转移使用的时间，比如稍微错峰 1-2 小时等，即可实现与负载聚合商之间的互动。

1.2.7 其他需要说明的问题

首先是天气对于功率的影响，在任务二作预测的过程中可以得出天气中对于总功率预测影响最大的是气温。除此之外，通过相关性分析等手段，也可以了解天气中不同因素对于某一家庭某一电器的使用的影响。

除此之外，在数据处理过程中，除了在报告中展示的按天切片，模型里还同样设计了按月切片的功能，该功能可以用来探究不同用户在不同季节的用电情况，进而为负荷聚合商提供建议。

2 基本任务二：多源数据融合的日前负荷预测与分析

题目回顾：基于数据集 2 中的公寓用电总负荷历史数据进行负荷预测；在此基础上，对训练得到的负荷预测模型进行分析，分析不同预测输入（如输入历史数据的时间周期）、不同预测对象（单个公寓或群体）对模型预测结果的影响，并给出解释。

加入数据集 2 中的温度、湿度等天气历史数据，再次进行公寓负荷预测；在此基础上，对训练得到的日前负荷点预测模型进行分析，具体分析外来数据的加入对模型预测结果的影响，识别其中的重要影响因素，并给出解释。

2.1 基本原理和模型

采用深度学习的方法，选择 LSTM（长短期记忆）模型。LSTM 在时序数据上有着优异的表现，因为 LSTM 在 t 时间时会将 t 时间之前的隐节点作为当前时间点的输入，LSTM 具有图示的结构。

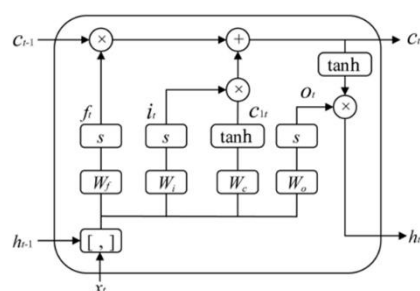
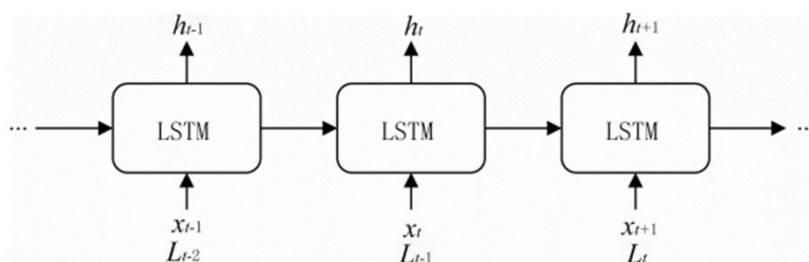


图1 LSTM网络单元结构图



这样之前时间点的信息也用于进行当前时间点的计算，而传统模型的输出只取决于当前时间点的输入，LSTM 在时间序列预测问题上有更好的表现。

2.2 数据的处理

2.2.1 缺失值填充

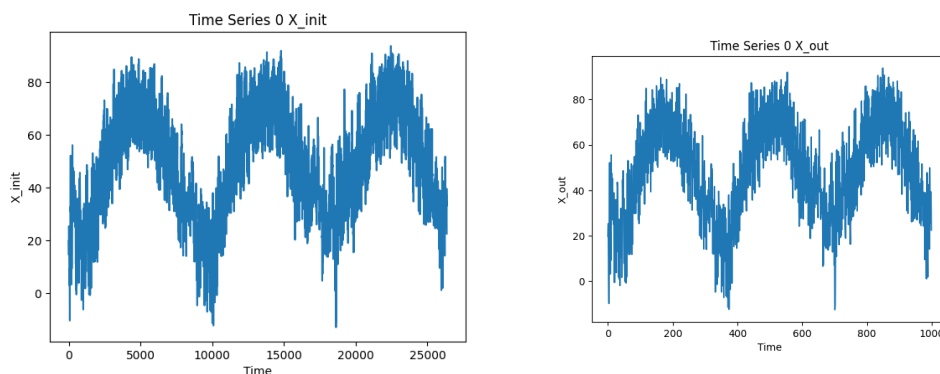
数据存在一些缺失，用插值法和后向填充对缺失值进行填充。

2.2.2 调整时间粒度

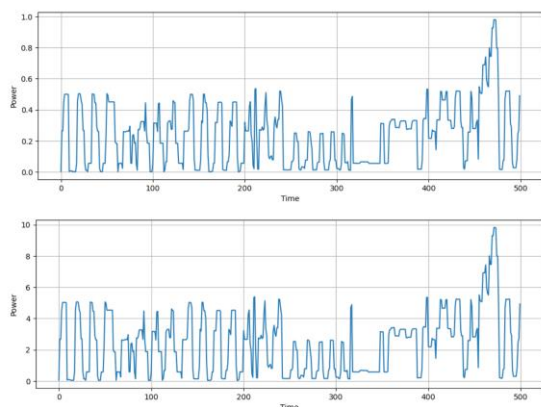
功率数据的时间粒度为一分钟一取样居多，有一些时间段取样稀疏，天气数据则是一小

时一取样，无法对齐。因此均按两点间数据值线性变化的方式在两个数据点之间插入新的数据点，扩充至每分钟一取样。训练时选择不同的数据量的大小和密度进行训练，如一分钟一取样的三天的数据，和一小时一取样的半年的数据进行训练。寻找更合适的取样方式。

除此之外，还有对数据进行保持曲线形状的压缩，具体压缩方法同任务一



如图所示，25000 个数据点压缩到 1000 个数据点保持了形状大致不变。但在一日的变化中有所失真。尽管如此可以得到大致的变化走势，当压缩倍数较小时一天的基本特征也能被保留，如下图为压缩倍数约为 3 的情况，将一天从 1440 个数据点压缩至 500 个：



通过这种方法可以任意调整时间粒度，进而在训练数据总量不变的情况下调整训练时间的范围。

2.2.3 构造历史功率

训练时，对作为标签的某个时间点，采用不同的周期取出 100 个该时间点间隔该周期整数倍的以前的功率（历史功率），且两点间时间跨度大于预测时间，将其作为该点进行训练的输入。

预测时，按同样的规则输入得到预测结果，由于某一预测点和它作为输入的历史功率相距时间跨度大于预测时间，故能保证输入数据中不含预测时间段的功率数据，避免未来数据泄露的问题。

2.2.4 其他处理

将部分字符串型的天气数据用独热编码转化为整型数据。

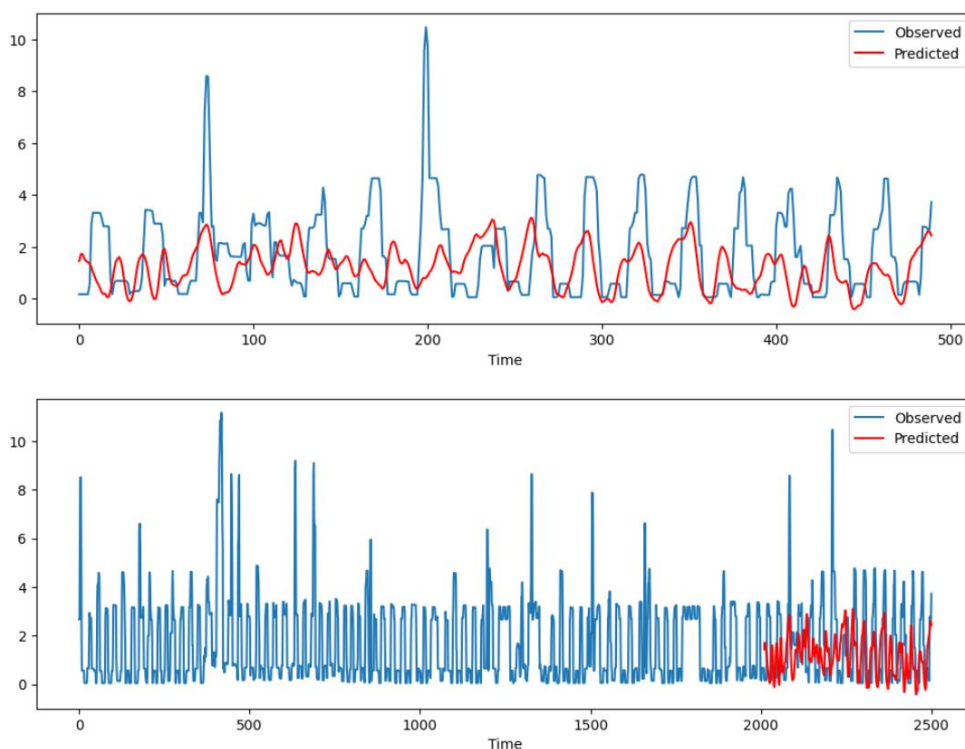
将数据归一化，转化为在 0 到 1 之间的浮点型数据。

数据 data 被分为训练集 train_data 和测试集 test_data。再分别将其构造为输入数据 train_X, test_X 和标签数据 train_Y, test_Y。对于数据 data，数据个数/时间为一个维度，变量类型（如某一公寓功率，该区域总功率，间隔第 N 个周期的历史功率，温度等各类天气数据）为另一个维度，对于构造出的输入数据 X，增加了一个 look_back 维度，长度为 LSTM 模型中计算当前节点时利用了多少个之前节点。标签数据为待预测量（某一公寓功率或者总功率）。

2.3 负荷预测和结果分析

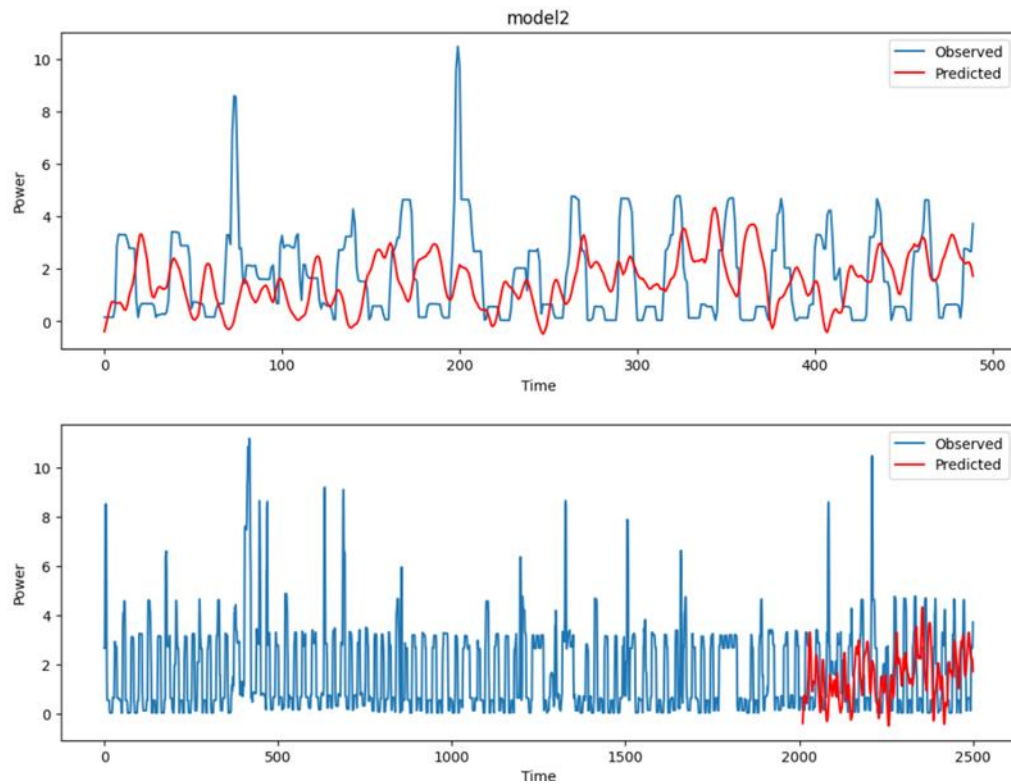
2.3.1 基于历史数据的负荷预测

(1) 对 apartment1 一日的负荷预测，为了得到更大范围的历史数据，将时间粒度调整为三分钟一取样，一天约为 500 个数据点，由于是单变量时间序列预测，输入数据来自构造的历史功率，按间隔一天的周期构造历史功率（即当前时刻 n 天前的功率， $n=1, 2, 3, \dots$ ），日前负荷预测如下图所示

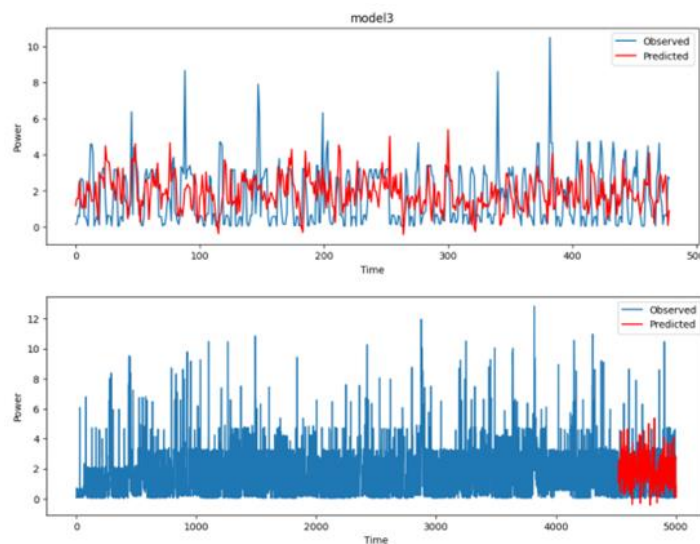


按间隔一周构造的历史功率（即当前时刻 n 周前的功率， $n=1, 2, 3, \dots$ ），日前负荷预测

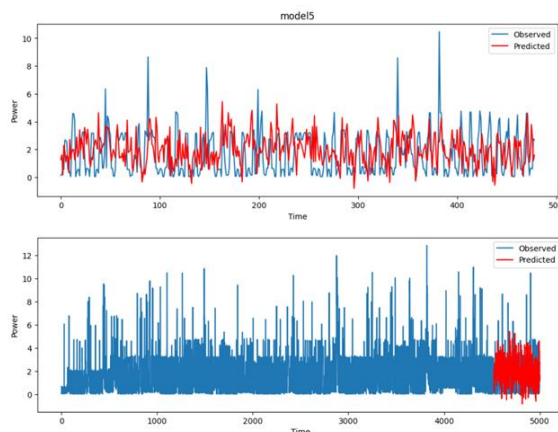
如下图所示。根据 MSE 计算的验证集损失 validation_loss 差别不明显, 可知历史功率的间隔不是对预测精度显著影响的因素。



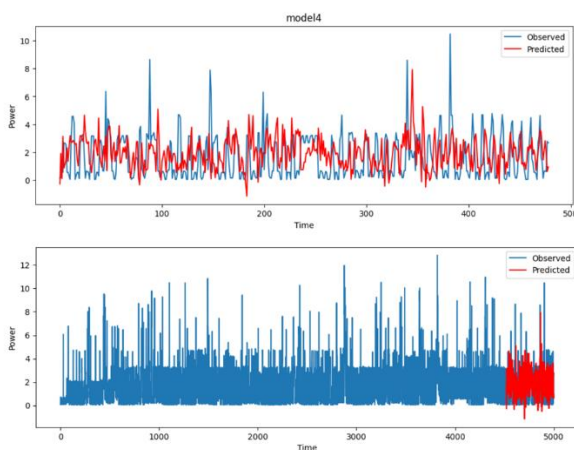
进一步调整时间粒度和训练时间范围, 然后比较了不同模型参数对训练结果影响, 如下图所示。



改变迭代次数等模型参数, 对训练精度的影响不明显:

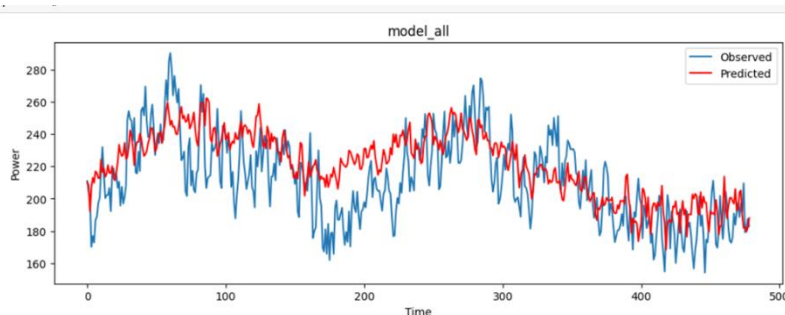


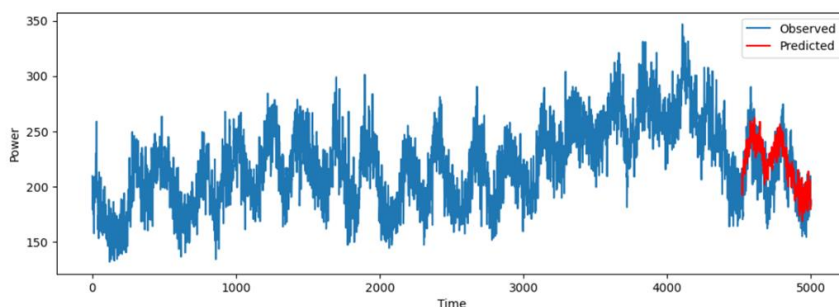
look_back 长度适当增加可以提高模型的预测效果：module4 在验证集上的 validation_loss (0.1116) 小于 module3 和 module5。增加 look_back 相当于增加了对单个标签进行拟合时的总输入量，参与计算的历史功率更多了，因此效果更好。如下图所示



```
226/226 [=====] - 1s 4ms/step - loss: 0.0351 - val_loss: 0.1127
Epoch 26/30
226/226 [=====] - 1s 3ms/step - loss: 0.0343 - val_loss: 0.1098
Epoch 27/30
226/226 [=====] - 1s 3ms/step - loss: 0.0325 - val_loss: 0.1080
Epoch 28/30
226/226 [=====] - 1s 3ms/step - loss: 0.0311 - val_loss: 0.1091
Epoch 29/30
226/226 [=====] - 1s 3ms/step - loss: 0.0306 - val_loss: 0.1097
Epoch 30/30
226/226 [=====] - 1s 3ms/step - loss: 0.0299 - val_loss: 0.1116
{'batch_size': 20, 'epochs': 30, 'optimizer': 'adam'}
```

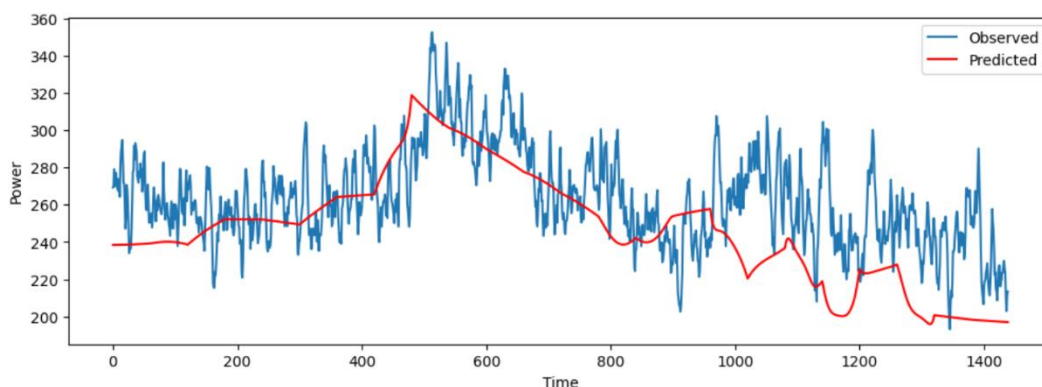
(2) 对于该区域总负荷进行预测：





总负荷相比于单个公寓的符合波动较小,预测曲线也能大致反映走向,更具有参考价值。从实际负荷曲线即可得出,区域总负荷变化相对单个公寓负荷变化更平缓,因此对于模型训练和负荷预测也更友好。

改变时间粒度为一分钟一个数据点,并且将训练数据从 5000 个增加至 500000 个,对于总负荷的预测曲线有较为显著的变化,可以看到预测曲线波动显著变小,但能较好地反映某一时段的平均功率,根据 validation_loss 得到的预测精度也较好。



```
Epoch 8/10
225/225 [=====] - 1s 3ms/step - loss: 0.0132 - val_loss: 0.0655
Epoch 9/10
225/225 [=====] - 1s 3ms/step - loss: 0.0131 - val_loss: 0.0601
Epoch 10/10
225/225 [=====] - 1s 3ms/step - loss: 0.0130 - val_loss: 0.0500
{'batch_size': 20, 'epochs': 10, 'optimizer': 'adam'}
```

因此适当增加训练数据改变输入时间的范围对预测有正面影响,在预测曲线较为波动(变化率大)时,波峰和波谷形状与实际负荷曲线相似,但存在一定错位,因此对于某个时刻的功率误差较大,而当预测曲线变化率变小,变化更平缓时,预测的负荷可以代表某一时段的平均负荷,更有参考意义。

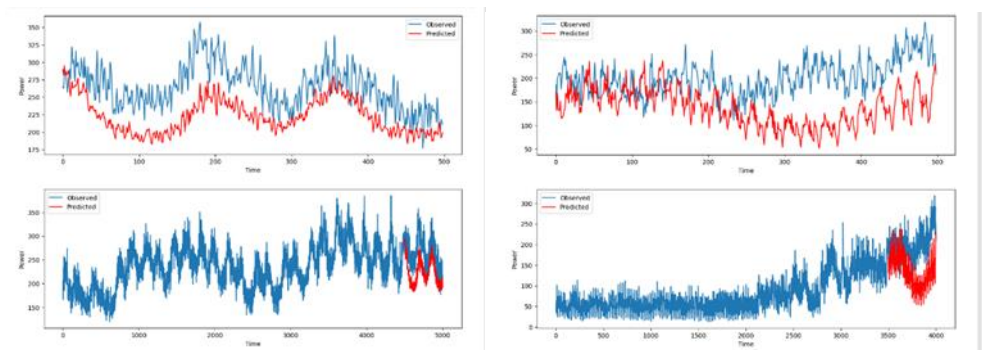
2.3.2 基于天气数据的负荷预测

(1) 天气数据对预测的影响

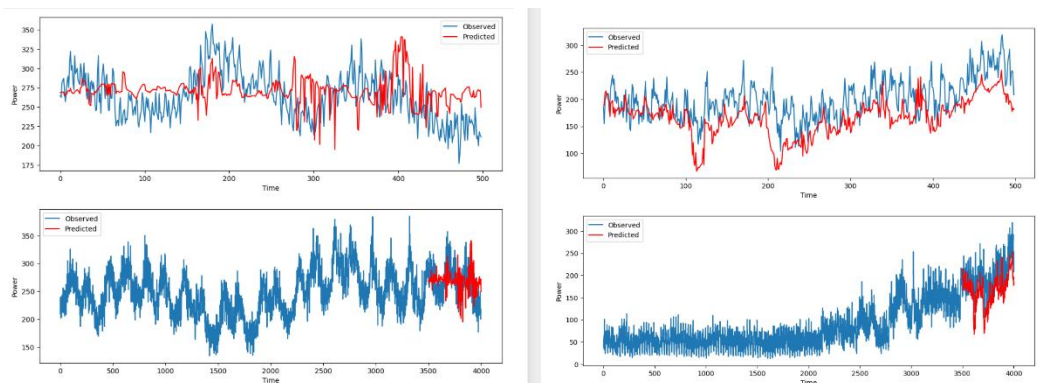
根据上一环节的结论,负荷预测对较长时间周期和较平缓的曲线预测效果更好,因此更适合预测较长时间周期的大致变化,故设置预测时间为一周,同时预测对象设置为区域总负

荷，使得曲线更平缓。

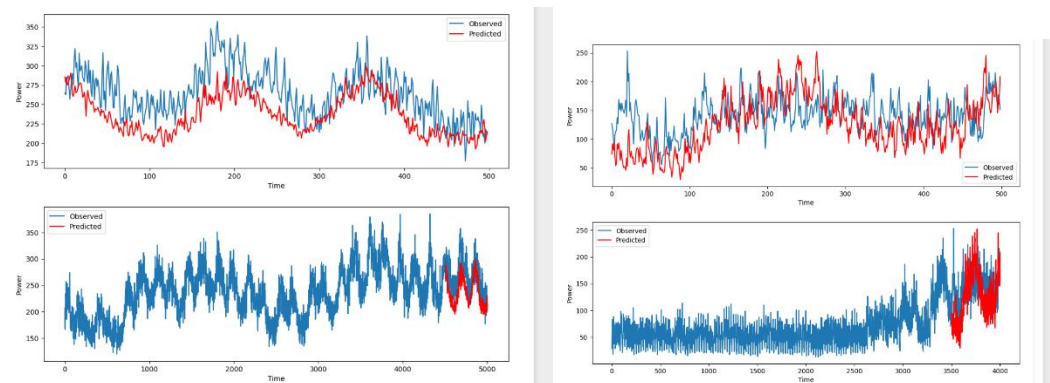
为进行对比，仍给出仅根据历史功率进行负荷预测的结果（方法一）：



单独使用天气数据进行负荷预测的结果如下图所示（方法二）：



将天气数据和历史功率数据进行结合预测，如下图所示：

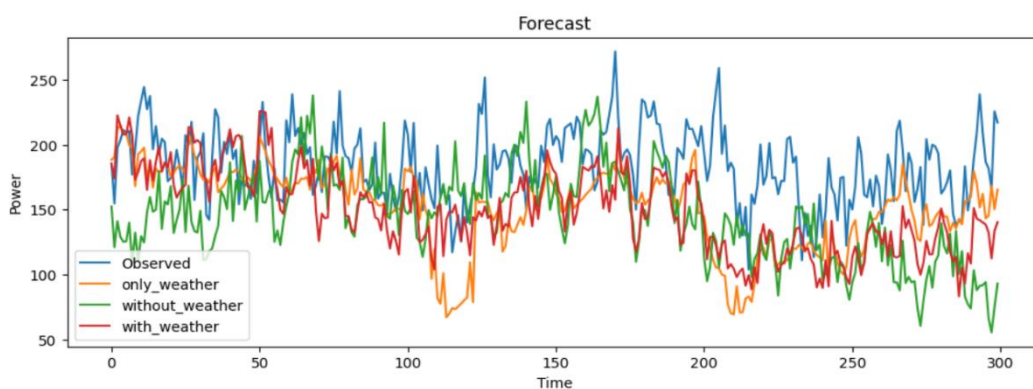


观察曲线可知，两种方法均有一定效果，仅利用天气数据进行负荷预测也得到了较好的结果，预测曲线的大致走势曲线（忽略高频的波动）变化不明显，相对于方法一波动更小，更加贴近实际符合曲线，但对于高频的波动更不敏感。因此利用天气数据进行预测也适合对一段时间的平均功率提高参考。

将两种方法结合使用得到了效果更好的曲线，相比方法一在验证集上的损失更小，说明天气数据能够提高预测精度，相比方法二的损失差别不大，但大致走势曲线变化更明显，在

需要知道大致走势的场合中更值得采用。

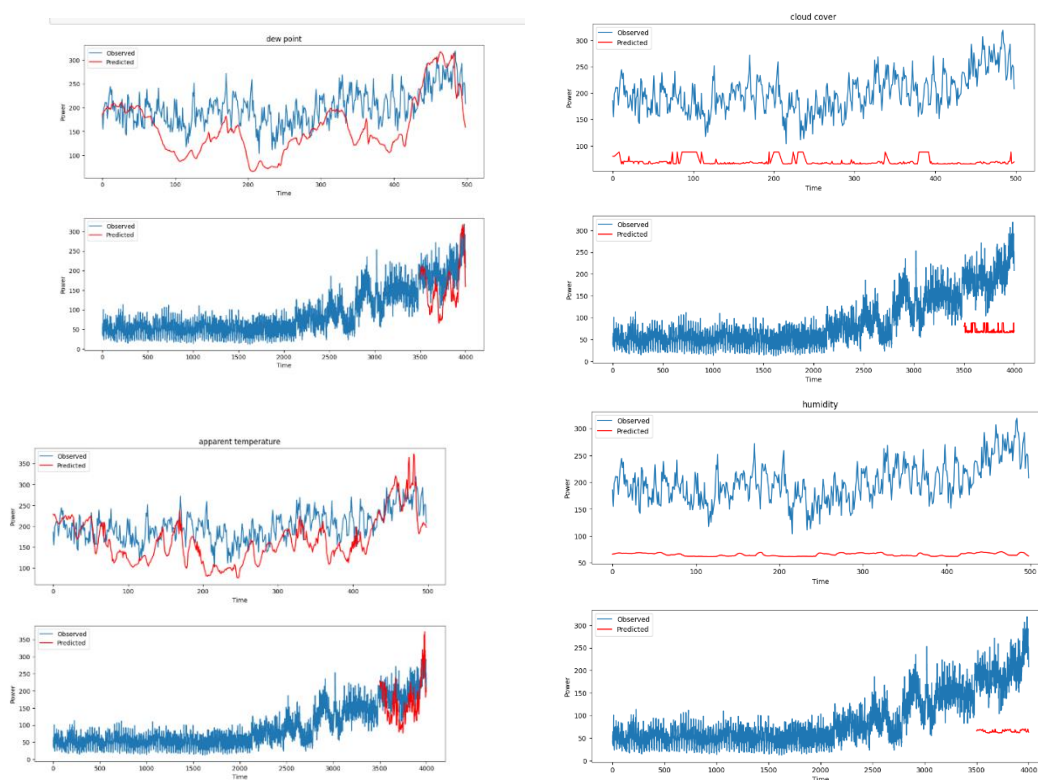
将模型下载下来，在不同的测试数据上使用，结果如下：

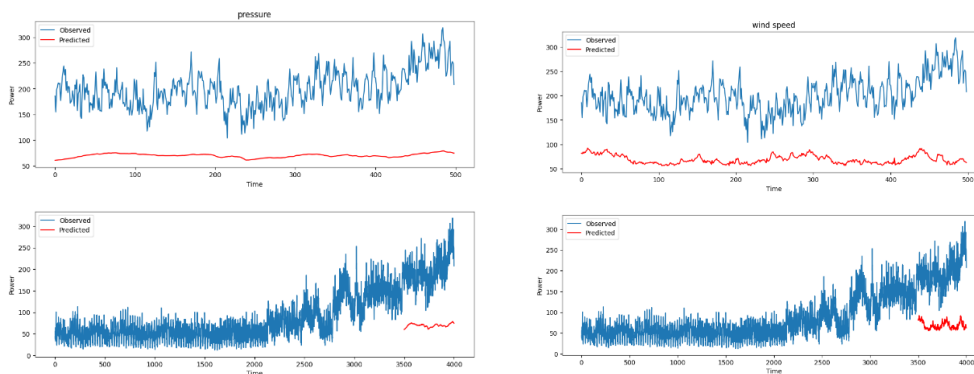


仍是将两种方法结合的预测效果最好，图中红线为两种方法结合。

(2) 识别天气数据中的关键影响因素

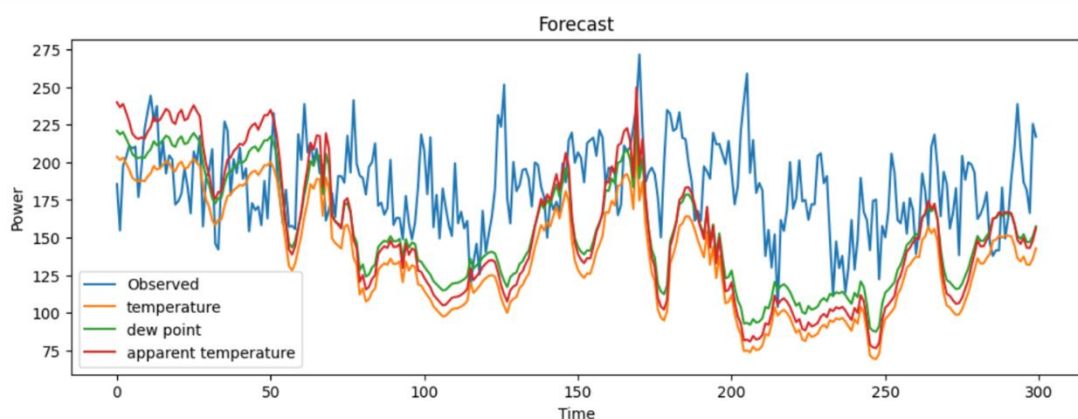
由于与方法一结合的情况，只能考虑每一个天气数据对预测的增益，而“预测精度的提升”这种变化不明显，因此将各种天气数据单独作为输入进行预测。给出部分不同的天气数据训练出的模型的预测结果如下。





大部分天气数据无法得到有效的曲线。

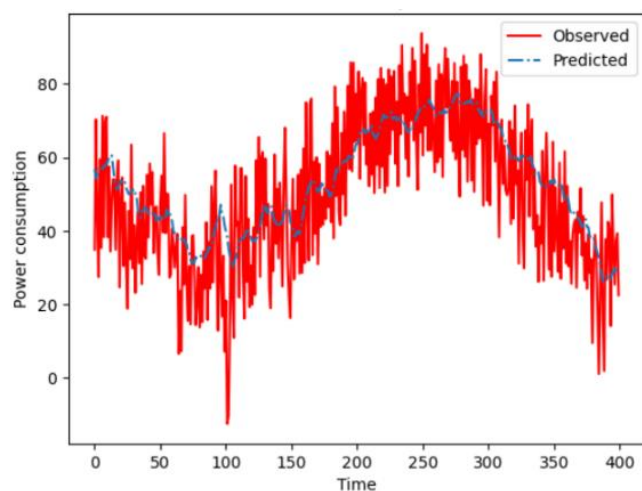
能有较好拟合效果的数据只有三种，如下图所示：



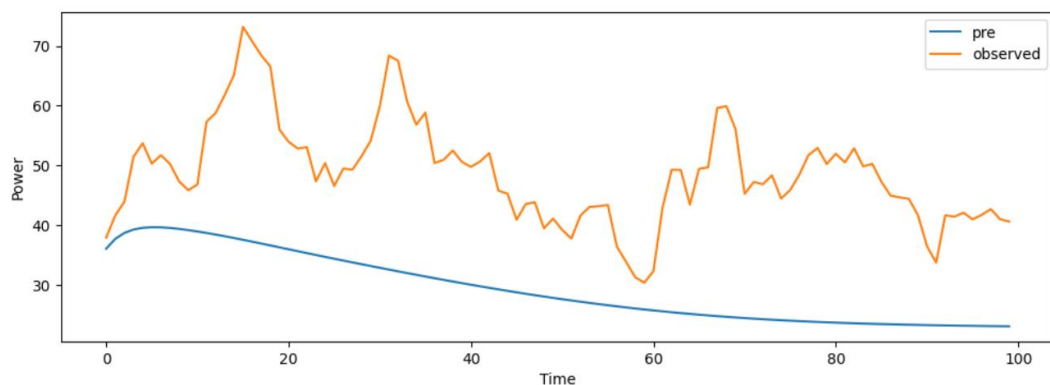
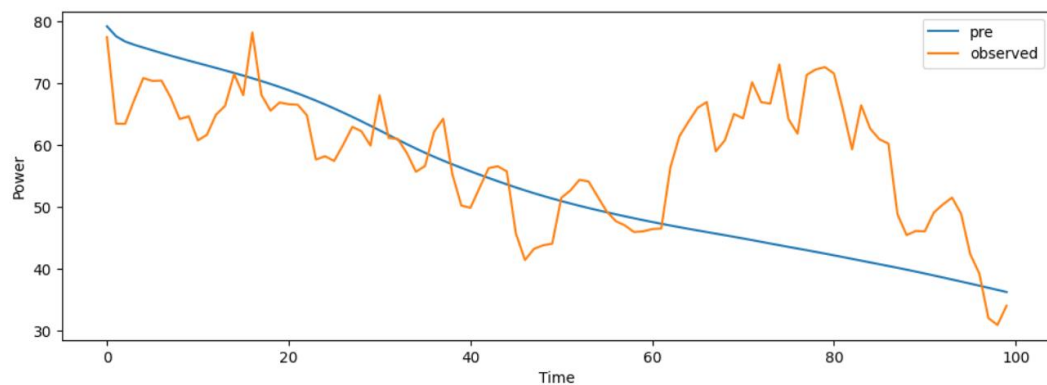
因此气温、露点和表观温度是对预测有显著影响的天气数据。露点是指饱和蒸气压随温度降低，降低至空气中的水可以凝结时的温度，因此和气温一样同属于温度类数据，由此可以得出结论：温度类数据对公寓负荷的影响最为显著。推测原因是有许多家用电器的功率和温度正相关，如需要保持温度的冰箱和空调等。此外空调和电风扇之类的电器使用时间本身也受温度影响。

2.4 其他

由于对模型参数和数据的要求不熟悉（如训练的数据量可能过小），导致上述预测结果均不是十分精确。在进行上述预测时，由于天气数据按一小时一取样，因此采用的将天气和负荷数据的时间对齐，只取了整小时的功率数据，导致总的训练数据较少。而且由于在短时间内，功率变化剧烈，无法较好地反映一段时间的平均功率（刚好在功率为 0 点或者峰值）。因此，通过数据处理中介绍的方法，将天气数据按两个数据点之间的数据满足线性变化的要求插入 59 个数据，扩充至一分钟一取样的数据集，增大了训练的时间范围，此时进行压缩也能较好地保留一段时间（如一个小时）之内的功率特征。此外再进行其他有益于负荷预测的参数调整，最后得到的较好的预测结果如下图所示：



为充分利用 LSTM 的特性,应该在预测当前时间点数据时输入前几个(长度为 look_back)数据进行计算。但直接输入测试集会导致未来数据泄露的问题。因此在上述过程中采取了构造历史功率的办法。但是想到有一种逐步预测的方法是在不输入未来数据也不另外构造历史功率的情况下进行预测,即写一个循环,输入数据 X 为测试集前的一段长度为 look_back 的数据,得到对下一时刻的预测数据,将这个数据添加至输入数据 X 的尾部,同去掉首位数据构成新的 X,以此为输入进行下一个数据的预测。预测结果如下:



可以看到只在预测第一个数据时较为准确，之后的预测几乎完全无效，推测是以预测数据作为输入导致误差逐渐累积。但不确定是方法固有的局限性还是由于对模型训练（参数和数据规模的选择等）不熟练导致的问题，有可能是训练的模型有误（之前的模型找不到了，此时采用的新训练的模型），希望日后能对此加以探讨。

3 选做任务一：用户需求响应方案设计

题目回顾：近年来，智能电表的推广为更为精细化的用户行为分析提供了支撑。在基本任务 1 研究的基础上，讨论智能电表的应用价值，并分析不同类型用户负荷灵活性与响应能力；自选利益主体（如负荷聚合商、售电商等），自定指标评估用户的需求侧响应能力，设计具体的需求侧响应方案并说明有效性。

3.1 负荷聚合商角度分析不同类型用户负荷灵活性与响应能力

3.1.1 传统用户分类与负荷聚合商

大型工业用户、商业用户对电力供应的要求相对灵活，是可中断负荷的重要来源，同时也对电价变化十分敏感，价格弹性较高，是电力系统很好的柔性负荷；居民用电则规模较小、可供改变的容量也很小，对电价变化不敏感，我国对居民用电仅仅利用分时电价进行有限的调节。以浙江省为例，在早期的电力现货市场中，只允许 110kV 及以上的用户直接参与，这些用户称为 I 类用户，特征为体量大，对价格信号的响应比较灵敏；110kV 以下用户称为 II 类用户，他们受到电网公司的管理，整体的负荷预测值作为市场出清的边界条件。通俗来讲，在日前电力市场的交易中，发电侧先列出自己的发电计划，然后减去 II 类用户整体的负荷预测值，剩下的发电量由 I 类用户进行竞价。这种模式下系统的运行较稳定，但是 II 类用户无法响应现货市场价格，需求侧弹性并没有被释放。

II 类用户的特点为基数大、分散性较强，但单个 II 类用户体量较小、弹性水平低。这时候，负荷聚合商就应运而生。负荷聚合商可以和一个小区甚至更多的人签订协议，获得他们家一部分用电设备的控制权，相应地，会支付一定的报酬或者提供更优惠的电价；也可以和小作坊、小店铺等 II 类用户签协议，获得调节他们的生产活动的部分权力。在日前现货市场，电力出现供需不平衡时，负荷聚合商可以向电网自荐，利用手中对很多用户用电设备的控制权帮电网平衡功率。与电网达成协议后，可以给所有用户下发通知，要求他们在某些时间段减少使用大功率电器，比如空调的温度调高一点，或者在此期间不要给电动汽车充电等等。

3.1.2 智能电表的价值

一方面，智能电表可以记录用户的用电习惯并具体到家用电器，方便负荷聚合商通过用户画像来决定是否与该用户进行合作，以及合作的具体方式。比如在必做任务中单个家庭总功率按日期进行切片并聚类之后横向比较，就刻画了每个用户的用电行为，可以通过其中的峰值负荷（即在某个时间段内功率需求显著增加）、波动性（观察功率是否频繁变化或波动

较大), 以及平时用电的平均功率大小来判断该用户是否有响应弹性。可以将用户按照响应弹性和响应容量两个指标进行分类。

而具体到某一个电器, 则可以根据必做任务中提到的四类电器分别分析, 来评估用户的响应弹性以及可以提供的响应容量。通过对各个家庭各个电器用电情况的聚类分析, 可以得知哪些用户家中具有柔性负荷, 并将他们作为潜在客户, 提高推销的效率。

另一方面, 智能电表可以解决每个用户在响应时的参与度, 进而解决利益分配的问题。智能电表不仅具有传统机械式电表电能计量的功能, 还可以实现分时计价、实时数据采集和传输。通过智能电表传输的数据, 利用计算机程序, 我可以很轻松地判断出用户的参与程度, 同时防止上述钻空子等情形出现。

3.2 需求侧响应方案设计

3.2.1 需求响应内涵

需求响应指电力用户根据价格信号或激励机制做出响应, 改变固有习惯用电模式的行为。根据价格做出的响应简称为 PBDR, 根据激励做出的响应简称为 IBDR。简单来说, PBDR 是电网不对用户进行任何强制措施, 通过调整电价来引导用户在高峰负荷时段减少用电或在低负荷时段增加用电, 从而平衡供需关系, 优化电力系统的运行效率和稳定性。我国最近实施的峰谷电价即为 PBDR 应用的实例。而 IBDR 则是电网对帮助削峰填谷的用户进行经济补偿或价格优惠, 其形式多种多样, 既包括可中断负荷用户与电网签订协议让渡一部分用电设备的控制权, 还包括需求侧竞价等手段。

3.2.2 评估用户侧响应能力的指标

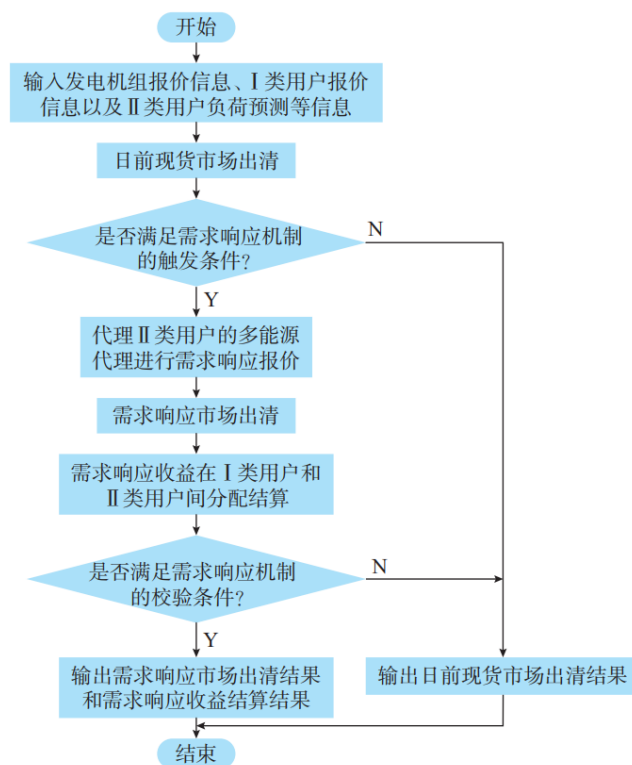
1. 可调节的弹性负荷功率。我会更加青睐拥有很多空调、购买了电动汽车且配备了户用充电桩的用户, 而不愿意与只使用弹性较低电器的用户合作。
2. 参与的积极性。这个指标在合作后才能被具体化。当负荷聚合商下发通知后, 迅速做出响应的用户便是优质的用户, 应该考虑适当给予更多的优惠。
3. 响应持续时间。用户可以在多长的时间内持续按需求做出调整。

3.2.3 需求响应方案的具体机制

需求响应的设计方案详见下一页中的流程图, 这里介绍如下: 首先发电企业和 I 类用户列出各自的发电、用电计划, 在 II 类用户的从整体负荷预测值作为边界条件的前提下进行交易。如果供需大致平衡, 那么需求响应机制便不需要派上用场; 如果出现了严重的供求矛盾, 导致电价出现极端时, 需求响应机制便会触发, 负荷聚合商便开始报价, 来协助电网实

现供需平衡。需求响应市场出清后，若还是不能达到供需平衡，此时只能调整发电侧或Ⅰ类用户的生产计划。

在时间顺序上，需求响应市场位于初期现货市场下的日前市场与实时市场之间，本质上是提供必要的需求侧调节资源来应对电网供需紧张时出现的日前极端电价，其不影响初期现货市场中的日前市场与实时市场的原有运行规则。



4 选做任务二：用户需求响应方案设计

题目回顾：实际生产运行中，各类数据可能把握在不同主体手中而存在信息壁垒。在基本任务 2 和 3 研究的基础上，自选场景和主体（如数据中心、售电商等），评估上述负荷预测相关附加数据在不同场景下可能带来的经济价值，并讨论不同主体未来可能怎样共享这些数据。

4.1 负荷预测相关附加数据在不同场景下可能带来的经济价值

4.1.1 主体：数据中心

数据中心，拥有产业链上下游历史数据，具有数据交易资质，且日常运行耗能大。

数据中心可能拥有大量的服务器和设备，对能源的需求量较大。使用负荷预测模型，结合附加数据如温度、湿度等天气数据，可以预测未来的能源需求，从而优化能源采购和管理。

数据中心可以通过共享其负荷预测数据和附加数据，与售电商等能源供应商建立合作关系。售电商可以根据这些数据更准确地进行电力市场交易和能源供应计划，从而提高市场运营效率和降低成本。

4.1.2 主体：售电商

售电商可以利用负荷预测模型和附加数据，对未来的市场需求进行预测，从而优化定价策略。准确的负荷预测有助于减少供需不平衡和供电不足的风险。

售电商可以与数据中心等大能耗用户建立数据共享合作，获取其负荷预测数据和附加数据，以提高自身预测精度和市场响应能力。同时，售电商可以为数据中心等主体提供增值服务，如能源成本优化、能效改进等。

4.1.3 对于不同主体对附加数据的利用的具体考虑

(1) 利用实时的天气数据参与负荷预测能够明显地提高精度，因此作为实时功率存在信息壁垒的主体，可以采用相应的模型，并且实时测量天气数据作为输入，得到对当下时间的功率的预测值，这对于该类主体有经济价值和参考意义，约等于向数据提供方购买数据的经济价值（当然精度有限），或者是用于优化自己的能源供应计划得到的经济价值。

(2) 对负荷预测有需求的主体，还可以利用天气预报的数据进行未来一段时间的功率预测。

(3) 可以采用不构造历史功率的 LSTM 模型（如 2.4 所述），即用当前时间点之前的功率进行预测，虽然只能预测很少的数据，但仍有一定参考意义，对于对较短时间内的负荷变化有更精确的需求的主体可以利用此类数据。

4.2 不同主体未来可能怎样共享这些数据

数据共享的实现方式有：

- (1) 通过互联网平台进行数据共享
- (2) 通过数据中心提供 API 服务，对所需数据进行访问
- (3) 达成数据共享协议，实现数据共享
- (4) 由政府部门或行业巨擘牵头发起数据合作项目

数据共享的具体实现重点有以下几点：

- (1) 隐私保护：在数据共享过程中，需要确保对数据的隐私和安全进行保护，遵循相关的数据隐私法律法规，并采取安全措施，如数据脱敏、加密等。
- (2) 数据标准化：为了实现数据的互操作性和共享效益，需要制定统一的数据标准和接口规范，确保不同主体间的数据可以互相理解和使用。
- (3) 激励机制：为鼓励数据共享，可以设计合适的激励机制，如经济激励、合作共赢等方式，使各主体愿意分享自己的数据，并获得相应的经济或其他回报。

参考文献

- [1] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, Teh Ying Wah. Time-series clustering – A decade review [J] Information Systems, 2015, 53: 16-38.
- [2]Steinarsson, S. (2013). Downsampling Time Series for Visual Representation. 21-25.
- [3]裴赢鑫. 面向居民用户的自动需求响应系统应用研究[D]. 华北电力大学(北京), 2017.
- [4]尹逊虎, 丁一, 惠红勋, 包铭磊, 徐立中, 唐学用, 桑茂盛. 初期现货市场下考虑用户响应行为的需求响应机制设计[J]. 电力系统自动化, 2021, 45(23): 94-103.
- [5]田世明, 王蓓蓓, 张晶. 智能电网条件下的需求响应关键技术[J]. 中国电机工程学报, 2014, 34(22): 3576-3589.
- [6]王蓓蓓. 面向智能电网的用户需求响应特性和能力研究综述[J]. 中国电机工程学报, 2014, 34(22): 3654-3663.
- [7]汤奕, 鲁针针, 宁佳, 林扬宇. 基于电力需求响应的智能家电管理控制方案[J]. 电力系统自动化, 2014, 38(09): 93-99.