《能源互联网导论》课程大作业

**挑战性任务报告**

题目：电力负荷大数据分析

组 号：第八组

组 员 一：朱继涛 2021011455

组 员 二：李德瑜 2021011459

组 员 三：吴晨聪 2022010311

2024年6月25日

目 录

[1 任务1：基于负荷历史数据的用户用能模式分析 3](#_Toc8883)

[1.1 聚类模型和流程分析 3](#_Toc335)

[1.2 数据预处理 3](#_Toc17655)

[1.2.1 降噪 4](#_Toc28027)

[1.2.2 降采样 4](#_Toc9128)

[1.2.3 归一化 5](#_Toc4963)

[1.2.4 相似度度量 6](#_Toc26905)

[1.3 提取典型日 6](#_Toc9779)

[1.3.1 肘部法则 6](#_Toc28001)

[1.3.2 轮廓系数 7](#_Toc12503)

[1.3.3 聚类选择 7](#_Toc7716)

[1.3.4 典型日曲线 8](#_Toc26404)

[1.4 用户聚类 9](#_Toc29085)

[1.4.1 聚类选择 9](#_Toc6460)

[1.4.2 聚类结果 10](#_Toc9965)

[1.4.3 聚类分析 11](#_Toc8985)

[1.4.4 聚类统计 12](#_Toc26117)

[2 任务2：用户行为与需求响应 12](#_Toc16738)

[2.1 聚类准备 12](#_Toc9114)

[2.1.1 数据集 12](#_Toc15862)

[2.1.2 灵活性潜力 13](#_Toc4615)

[2.2 关于电价影响的聚类 13](#_Toc14806)

[2.3 用户参与需求响应的灵活性潜力分析 14](#_Toc28177)

[2.3.1 减小负荷 14](#_Toc18260)

[2.3.2 增加负荷 15](#_Toc12001)

[3 拓展任务1：数据价值评估方案设计 15](#_Toc24234)

[3.1 用户数据价值评估策略 16](#_Toc6833)

[3.2 数据交易商业模式 16](#_Toc18540)

[参考文献 17](#_Toc3580)

[附录A 任务分工记录表格 17](#_Toc13428)

# 任务1：基于负荷历史数据的用户用能模式分析

基于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。

## 聚类模型和流程分析

聚类模型总体框架如图1所示。

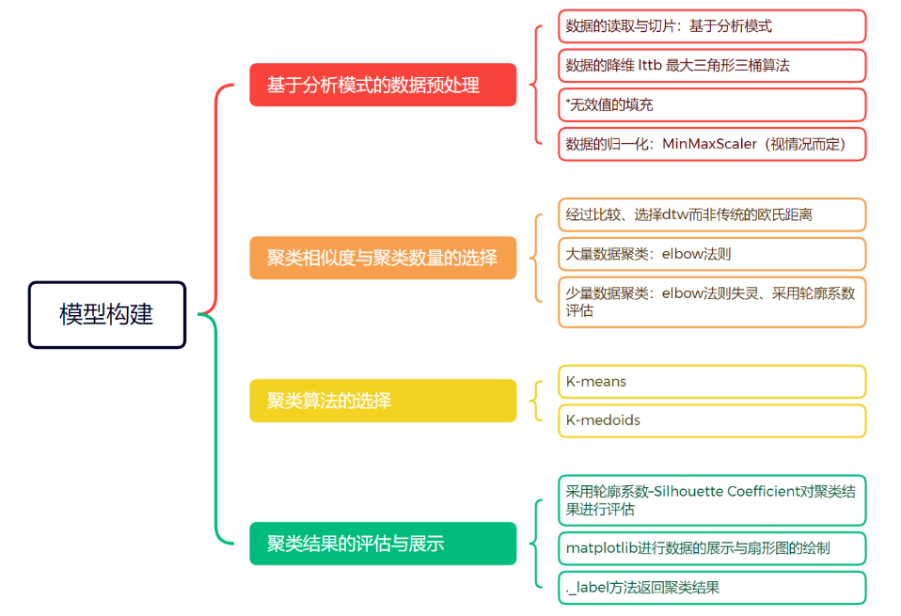


图1聚类模型总体框架

聚类模型包括数据预处理、提取典型日、聚类分析、评估聚类效果。

## 数据预处理

任务1中采用的数据集来源于MIT发布的2016年全年114单户公寓的用电数据，单位为千瓦，采样时间间隔为1分钟。每年在一个家庭采集超过50万个负荷数据。图1为apartment 1一天的原始用电数据图。

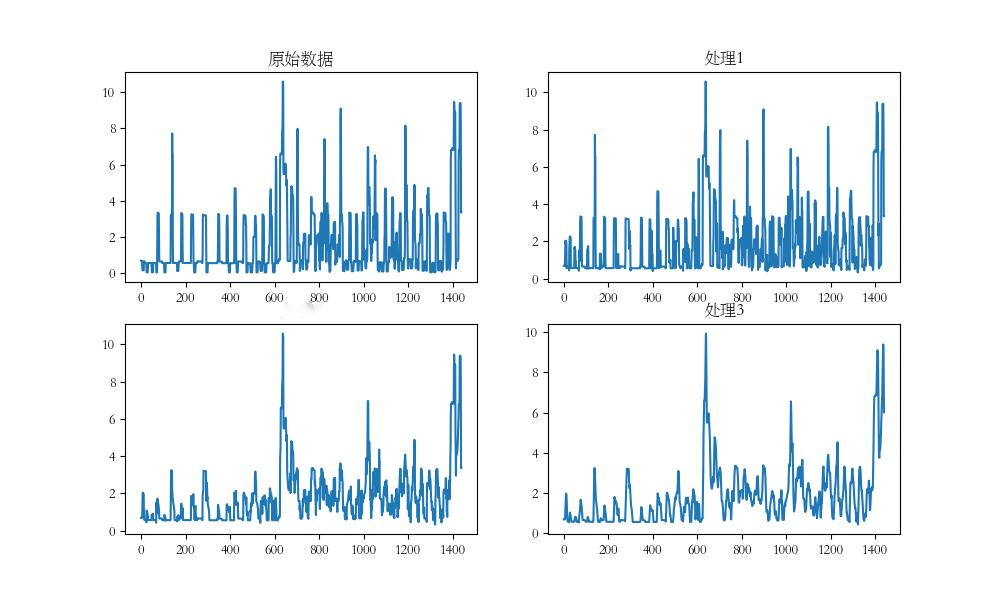


图2 apartment 1单日用电曲线

从图中可以看出，数据波动非常明显，用户的用电模式被各种干扰所掩盖。为了提取具有实际价值的用户信息，需要对数据进行预处理。

### 降噪

在查阅相关文献后，公寓1一天的原始用电数据进行了以下的降噪操作：

处理1缺失数据剔除。对于负荷明显较小的数值，进行剔除处理。然后，采用当月同一个星期日同一时刻的平均值进行填充。这样可以有效减少由于缺失数据对整体分析的影响，同时确保数据的连续性和一致性。这种方法的优点在于利用历史数据的相似性来填补缺失值，保持了数据的统计特性。

处理2异常数据补全。在数据采集过程中，可能会出现错误记录，这些异常数据通常表现为负荷值的突然升高或降低。我们通过前后负荷变化率来判定异常数据，并进行补全。异常数据的处理是为了确保数据的准确性，避免由于异常点影响整体趋势分析。采用变化率判定方法可以有效识别并修正突变值，保证数据的合理性。

处理3动平均法处理。采用平滑公式处理随机噪声。动平均法通过对数据进行平滑处理，减少了随机波动的影响，使得数据更加平稳和易于分析。这种方法能够有效地提取出数据中的主要趋势和模式，提高数据的质量。

apartment1一天的原始用电数据降噪处理后的效果图如图2所示。

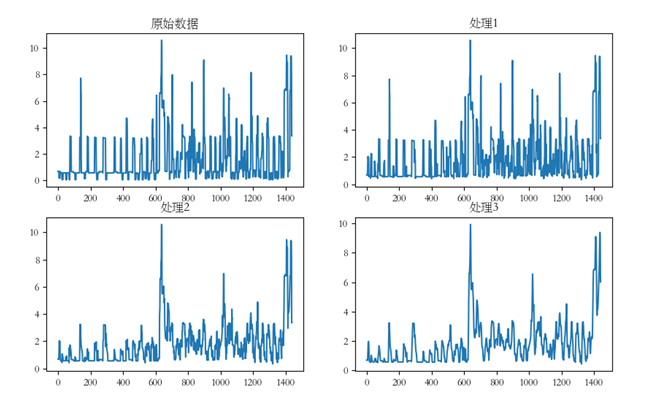


图3 降噪对比图

### 降采样

庞大的数据量会大大增加运算时间和复杂度，而且用户感知环境变化、用电习惯影响等因素带来用电变化的时间尺度较长。因此，采样过密不会增加有效信息，会带来冗余。为了减少计算时间和复杂度，需要对数据进行降采样处理。

直接间隔选取数据的方法断然不可取，这存在很大的盲目性，容易丢失序列中的重要特征。为尽可能保留数据序列中的特征，采用最大三角形三桶算法（Largest-Triangle-Three-Buckets, LTTB）进行降采样。

最大三角形三桶算法，其核心思想是通过选择能够最大化保留数据特征的点，从而减少数据量的同时保留序列中的主要信息。具体步骤如下：首先，将原始数据分成若干个等间隔的桶（buckets）；其次，在每个桶中选择一个点，使得该点与其相邻点构成的三角形面积最大，保证所选点在视觉上具有代表性；最后，将这些点组成新的降采样后的时间序列。LTTB算法能够有效减少数据点数，同时保持数据的主要特征和趋势，非常适合大数据量的时间序列数据的降采样处理。通过这种方法，可以大大降低计算复杂度，提高数据处理效率。

apartment1一天用电数据在降采样前后的对比图如图3所示。

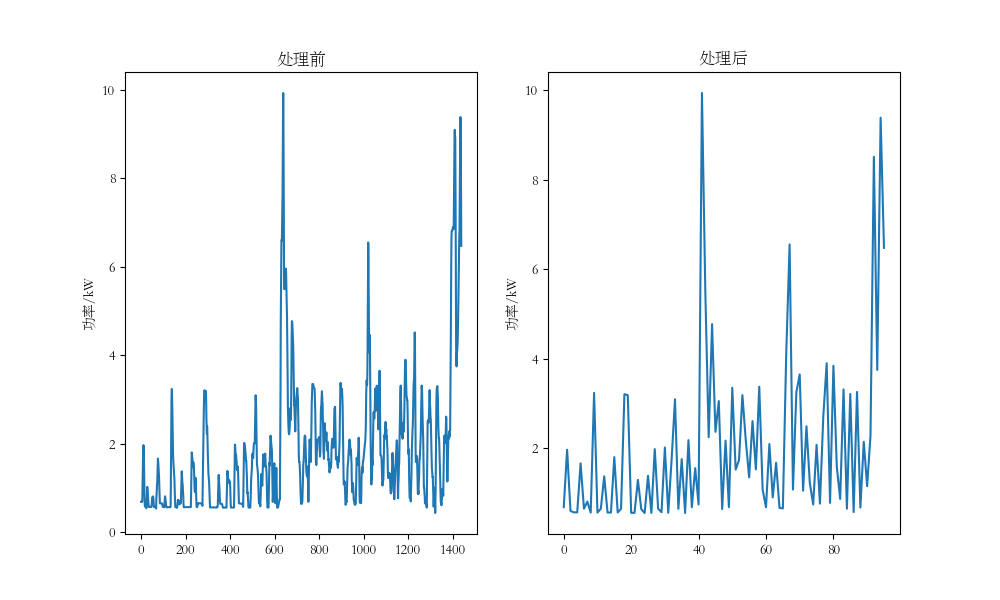


图4 降采样对比图

可以看出，经过LTTB算法处理后，抽样间隔由1分钟降至15分钟，采样点由一天1440个点降至一天96个点，数据量显著减少，数据的主要特征得到了有效保留，为后续的聚类分析提供了更加具有代表性的数据基础。

### 归一化

在使用一整年数据分析用户的用电模式时，不可避免地会出现季节性的用电变化。为了避免不同日期负荷数量级对聚类效果的影响，我们通过归一化将数据压缩在[0, 1]的范围内。本模型采用最大值、最小值归一化的方法，即：

这种方法有效地将数据标准化，使得不同日期的负荷数据在同一个尺度上进行比较，从而提高聚类分析的准确性和可比性。

然而，在某些情况下，例如分析同一家庭四季的用电模式差异时，功率差异本身可能是重要的分析因素。在这种情况下，可以选择不进行归一化处理，以保留原始数据中的功率差异信息，从而更准确地反映季节性变化对用电模式的影响。

### 相似度度量

在进行聚类分析时，可以依据样本间相似性的高低来判断是否将其归为同一类。

动态时间规整（Dynamic Time Warping, DTW）是一种通过调节时间序列采样点的对应关系，寻找最佳匹配路径，从而度量两个长度不同时间序列相似度的方法。具体实现步骤包括：首先计算时间序列A和时间序列B中各元素间的距离，然后从一个对角出发，到达另一个对角，累加经过的距离，最终找到距离最短的路径。DTW不仅可以有效衡量长度不同序列的相似性，还能够处理时间尺度上的不一致性。

除了需要衡量样本之间的距离外，一些聚类算法（如层次聚类）还需要衡量聚类间的距离。DTW的优势在于其能够灵活适应时间序列的非线性变化，从而在不同聚类算法中广泛应用，提升聚类结果的准确性和可靠性。因此，DTW在处理时间序列数据时，尤其是在序列长度不一致的情况下，显示出显著的优越性。

## 提取典型日

采用肘部法则（Elbow Method）和轮廓系数(Silhouette coefficient，SC)评估聚类效果，进而选择合适的聚类数目和聚类算法。

### 肘部法则

肘部法则的核心指标是误差平方和（Sum of the Squared Errors, SSE）。其核心思想是随着聚类数的增大，样本划分会更加精细，每个簇的聚合程度会逐渐提高，从而使SSE逐渐减少。当超过合适的聚类数时，再增加，SSE的下降程度会骤减并趋于平缓。通过绘制SSE与聚类数的关系图，当图中SSE明显下降的“肘部”出现时，即为最佳聚类数。具体计算SSE的方法如下：

其中为聚类数，为第𝑖个簇，𝑥为簇中的样本点，为簇的质心，表示样本点与质心之间的欧几里得距离。

### 轮廓系数

选择聚类的内部指标来评价聚类效果的好坏至关重要。遵循“簇内相似度越高，簇间相似度越低，聚类效果越好”的原则，选取轮廓系数SC进行分析。

轮廓系数的计算公式如下：

其中，为样本总数，是样本到其他簇样本的最小平均距离，是样本到同一簇其他样本的平均距离。SC值的范围在[-1, 1]之间，SC值越大，表示聚类效果越好。具体来说，SC接近1表示样本点正确聚类，SC接近0表示样本点在聚类边界上，SC接近-1表示样本点被错误聚类。

### 聚类选择

为了获得较好的聚类效果，本任务选取了三种聚类算法：Kmeans算法、Kmedoids算法和DTW-Kmedoids算法。每种算法在理论上都有其特定的应用场景，如K-means算法适用于一般的数值数据聚类；Kmedoids算法在处理带有噪声和异常值的数据时更稳健；DTW-Kmedoids算法专门用于时间序列数据的聚类，可以更好地处理时间序列中的变形和错位。

一張含有 文字, 螢幕擷取畫面, 行, 繪圖 的圖片

自動產生的描述

图5 apartment 1肘部曲线

一張含有 行, 螢幕擷取畫面, 圖表, 繪圖 的圖片

自動產生的描述

图6 apartment 1轮廓系数图

通常情况下，随着聚类数目的增大，SSE会逐渐减少，但当聚类数目增大至一定程度时，SSE的变化趋于平缓。因此选取聚类数为2，初步设想将公寓1的一天用电量分为周末和工作日两类。图7为对apartment 1进行按天用电量的聚类数为2的聚类分析。

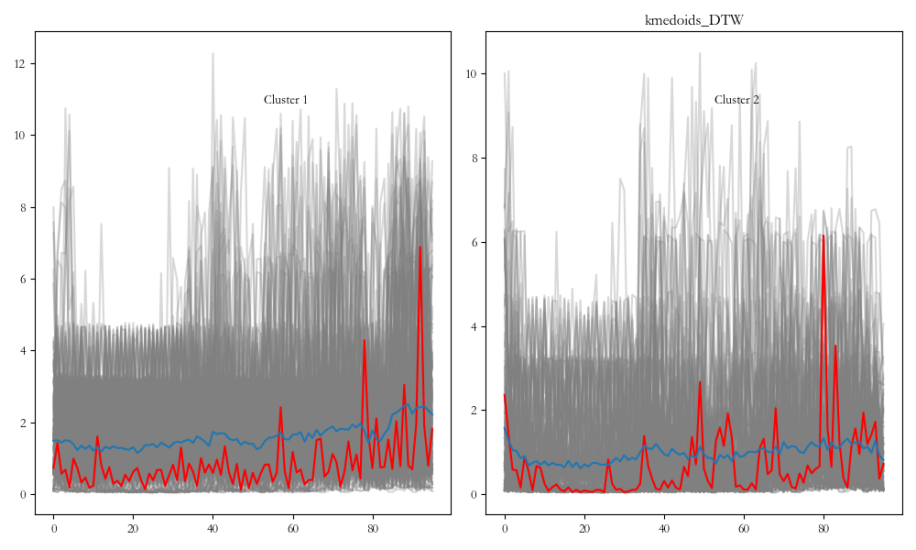


图7 apartment 1聚类结果

可以看出，这两个类之间的差异并不显著，可能是由于同一用户的用电模式本身较为固定，难以划分聚类。不如直接将每户家庭的日用电量作为一个类，根据DTW最小选择聚类中心作为典型用电模式。

### 典型日曲线

下图为apartment1在2016年的聚类结果，图中的红色曲线是典型日的用电曲线。

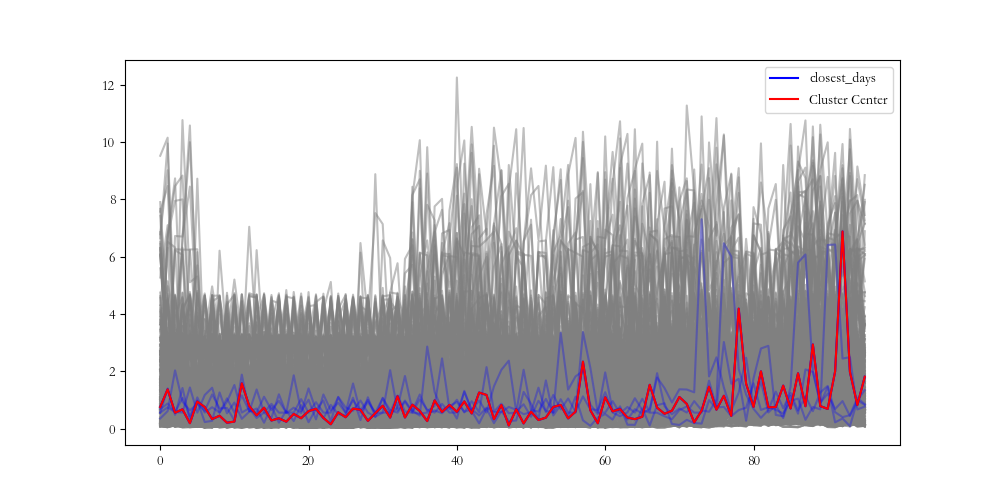


图8 apartment1选取的典型日曲线

通过该曲线，我们可以观察到该家庭的用电习惯。该家庭的用电高峰集中在傍晚，推测家庭中可能存在放学后回家打游戏的学生。

此外，图中还绘制了与典型日相似度最高的几日用电曲线，即蓝色曲线。可以看出，尽管用电模式相似，但是用电高峰会“漂移”，这表明用户用电行为存在一定的不确定性，也给聚类带来一定的困难。

## 用户聚类

* + 1. 聚类选择

首先找出114个用户的典型日用电数据，再绘制肘部曲线和轮廓系数图，选择适合的聚类数和算法进行聚类分析。

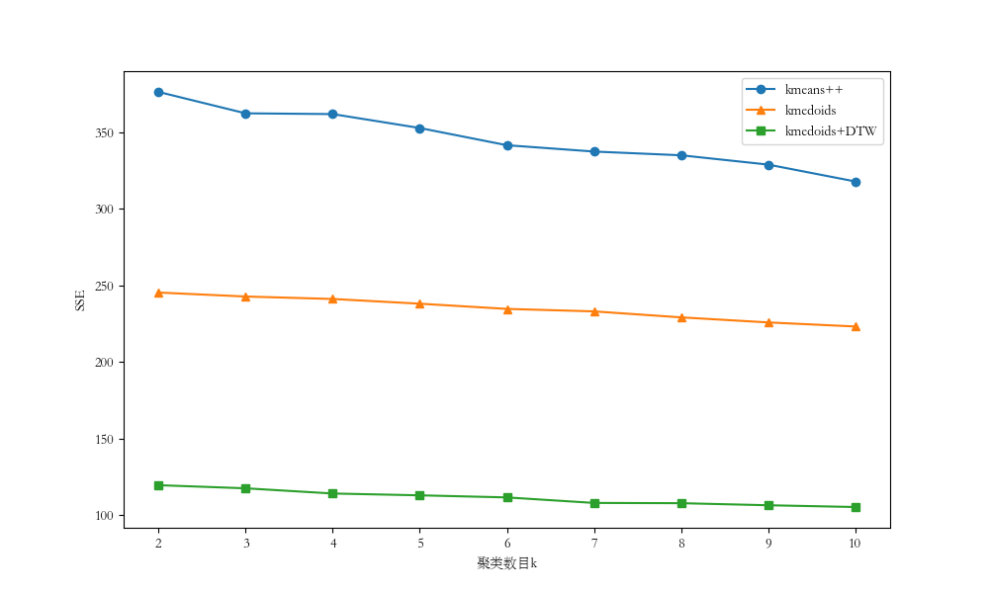


图9 114个用户的肘部曲线 一張含有 圖表, 行, 繪圖, 斜率、斜坡 的圖片

自動產生的描述

图10 114个用户的轮廓系数图

最终选取聚类数为5，聚类算法为kmedoids+DTW。

* + 1. 聚类结果

图11为聚类结果。

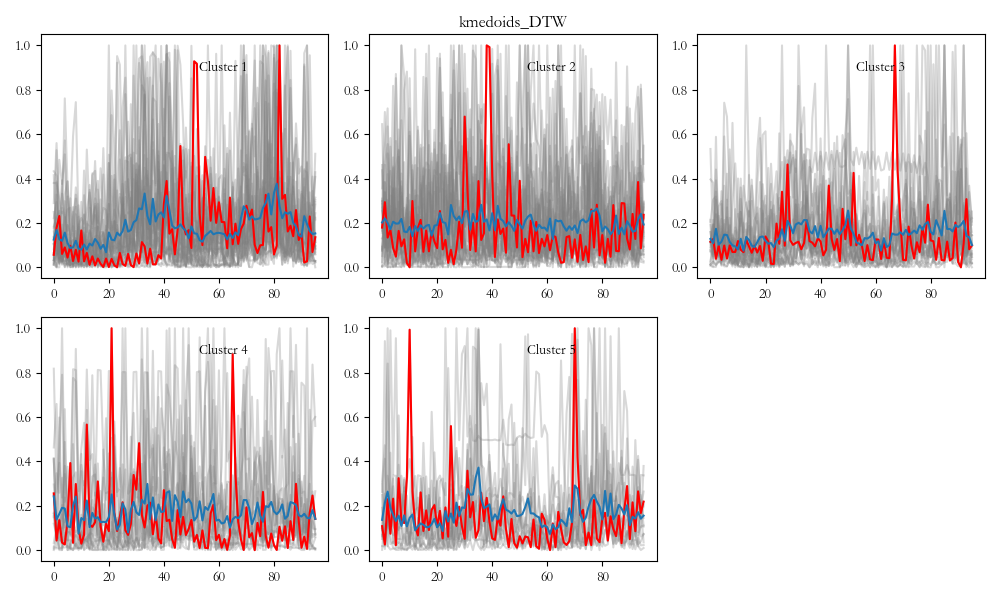


图11 聚类结果

* + 1. 聚类分析

对图11进行初步分析，首先确定了各类家庭的用电高峰期。如表1所示。

表1 聚类分析

|  |  |  |
| --- | --- | --- |
| 家庭类型 | 用电高峰期1 | 用电高峰期2 |
| 幸福小家 | 10:30-13:00 | 20:00-21:00 |
| 老人小家 | 07:30-11:00 |  |
| 自由小家 | 16:15-17:30 |  |
| 轮班小家 | 03:30-07:30 | 16:15-17:30 |
| 夜猫小家 | 01:30-02:30 | 17:30-19:00 |

第一类家庭的用电高峰期集中在上午10:30到下午1:00，以及晚上的8:00到9:00。基于这一观察，我们将其定义为“幸福小家”。这一命名的原因在于，上午10:30到下午1:00用电量大的时段，可能是家庭成员，尤其是父母与孩子分别去工作和上学，而母亲独自在家进行家务或准备午餐。晚上8:00到9:00的高峰时段，则可能是家庭成员下班放学后，一同在家中休闲娱乐，如观看电视。

类似地，我们对其他四类家庭的用电高峰期进行了分析和命名。老人小家在07:30至11:00的用电高峰，反映了老年人在早晨活动频繁的特点。自由小家虽然有明显的高峰期在16:15至17:30，但由于家中有自由工作者，其生活方式的波动性以致用电的波动。轮班小家则因工作时间的不固定，其高峰期在凌晨和下午，反映了轮班工作者的作息规律。夜猫小家在深夜和晚间的高峰用电则显示了这类家庭作息时间较晚的特点。

通过这种方式，可以更直观地理解不同用户群体的生活习惯和用电模式。

### 聚类统计

图12为各家庭类型占比。

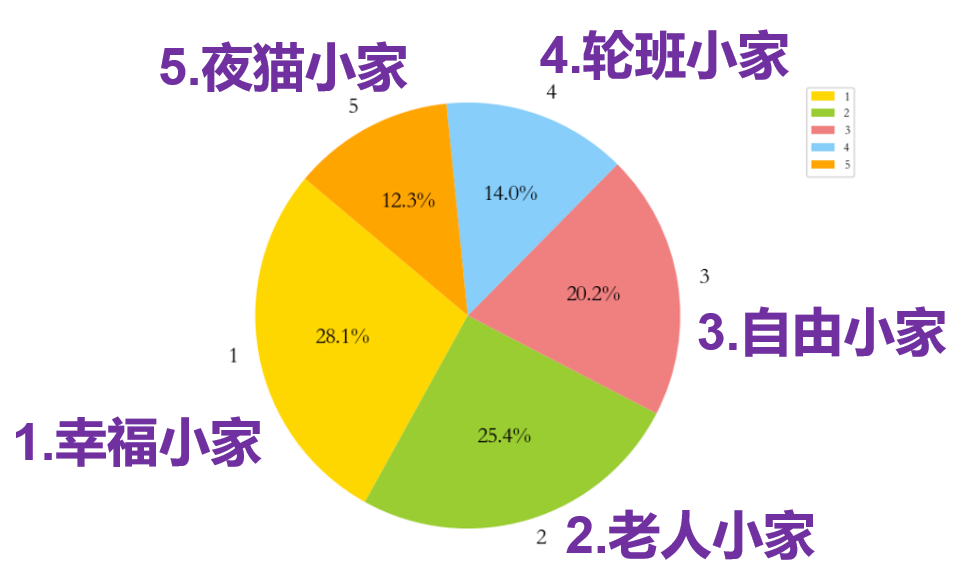


图12 家庭类型占比图

进一步对五类家庭在整个小区中的占比进行了统计，结果显示其分布符合现实情况。其中，幸福小家占比最多，占据了小区的大部分比例；其次是老人小家，反映了小区中老年居民的普遍存在；自由小家和轮班小家则占据了较小的比例；而夜猫小家占比最小，表明这种生活方式较为少见。

# 任务2：用户行为与需求响应

自选利益主体（如负荷聚合商、售电商等），自行设计激励条件，评估用户参与需求响应的灵活性潜力，并根据任务1中刻画的用户行为特征给出对应分析。

## 聚类准备

* + 1. 数据集

本任务中选用的数据集来源于2011年11月至2014年2月期间，参与英国电力网络主导的低碳伦敦项目的5,567户伦敦家庭的用电量数据。该数据集每半小时采样一次，总计包含约1.67亿行数据。数据集包含两组客户：

第一组包括约1100名用户，这些用户在2013年期间受到动态电价的影响。动态电价分为三档：高档（High）为67.2便士/千瓦时（p/(kW·h)），低档（Low）为3.99便士/千瓦时，和标准档（Normal）为11.76便士/千瓦时。

第二组包括其余约4500名用户，其电价在整个期间保持稳定。

在研究电价对用户用电模式的影响时，我们单独提取了第一组用户中有较为完整的2013年用电数据的用户。经过筛选，共有978名用户参与聚类分析。聚类过程中，将电价变化时的用电量与前后各一天同时刻用电的平均值进行对比，分别计算电价为高档（High）和低档（Low）时的用电变化。

聚类时，将电价变化时的用电量与前后各三天同时刻用电的平均值做对比，分别计算电价为High和Low时的用电变化。

* + 1. 灵活性潜力

灵活性潜力定义为：

图13为MAC005921的灵活性潜力。

一張含有 文字, 繪圖, 圖表, 行 的圖片

自動產生的描述

图13 灵活性潜力（以MAC005921为例）

* 1. 关于电价影响的聚类

图14为电价对用电模式的影响聚类。以电价为Low的电量变化为横坐标，电价为High时的电量变化为纵坐标，绘制成所有用户的散点图。最终结果分布在四个象限，将同一象限的用户作为一个类，划分出四类用户。

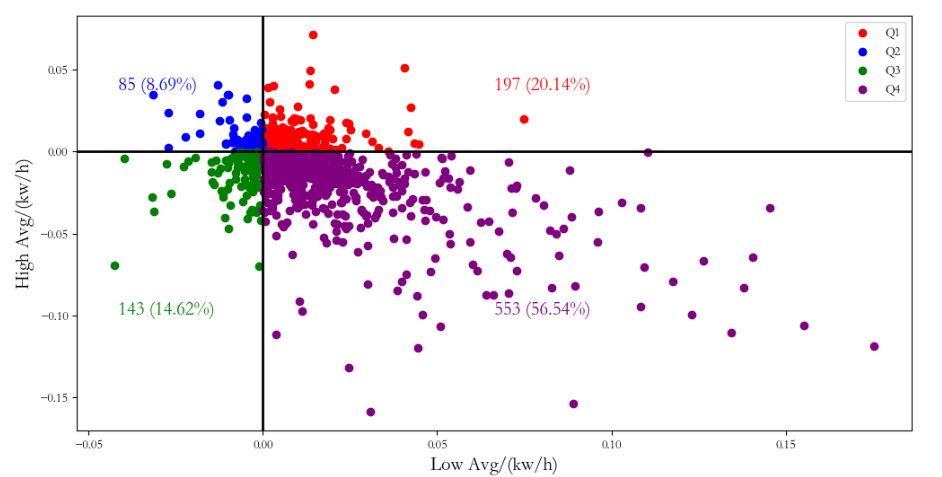


图14电价对用电模式的影响聚类

可以发现，数据点集中在（0，0）附近，表明对于大部分用户而言，电价变化对其用电行为的影响较小，即对电价变化的敏感度较低。

此外，大部分用户分布在第四象限，当电价升高时，其用电量减少；而当电价降低时，用电量增加。这一现象符合一般的预期。这为电价弹性和用户响应度的分析提供了一定依据，并对电力定价策略的制定和需求响应计划的优化具有参考价值。

## 用户参与需求响应的灵活性潜力分析

在本任务中，选择的利益主体为电网公司。当负荷需要调节时，例如在夏季和冬季用电高峰期，为了保证供电的稳定性，需要采取措施降低负荷。此外，当需要消纳新能源时，可能需要增加负荷以避免弃风、弃光等问题的发生。

通过动态电价策略，电网公司可以有效引导用户调整用电行为，平衡供需关系。

* + 1. 减小负荷

选择通过增加电价的方式来降低负荷，激励对象为分布在第三、四象限的用户。

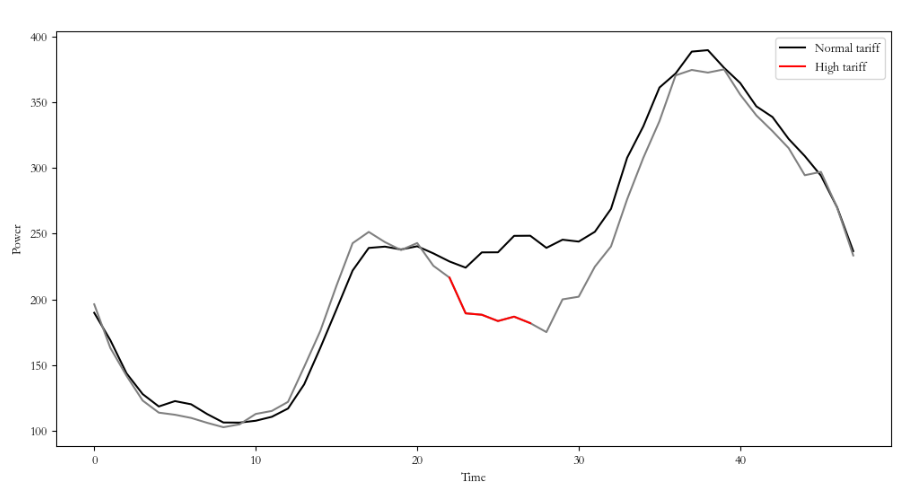


图15 电价增加时的负荷变化

可以看出，当电价从Normal变为High时，若针对第三、四象限的用户进行激励，978户家庭的总用电量在对应时刻将下降近50 kW·h。

由于数据集中的电价并不是连续变化的，因此无法准确得出各电价增加对应的负荷变化。假设电价增加与负荷减小呈线性关系，可以认为电价增加对负荷的影响可线性表示为0.9018 kW·h/p。

### 增加负荷

选择通过减小电价的方式来增加负荷，激励对象为分布在第一、四象限的用户。图16为电价降低时的负荷变化。

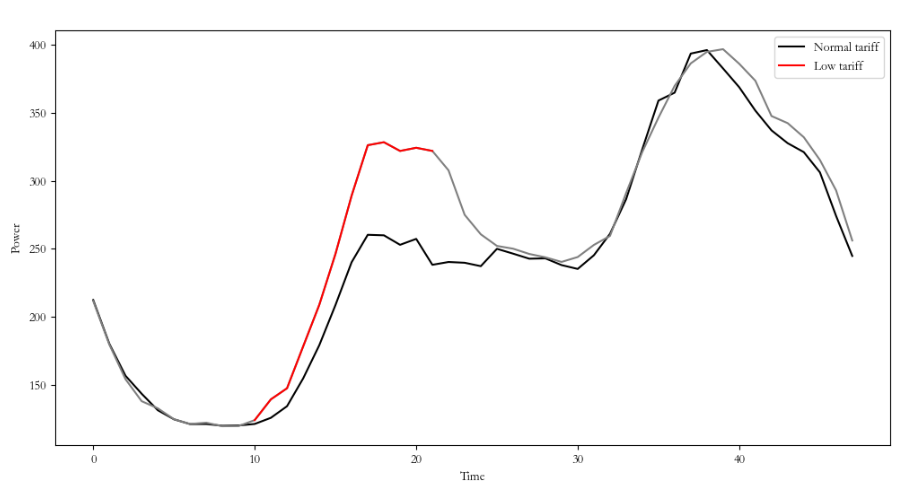


图16 电价降低时的负荷变化

可以看出，当电价从Normal变为Low时，若针对第一、四象限的用户进行激励，家庭的总用电量在对应时刻将增加近80 kW·h。如果假设电价减少与负荷增加呈线性关系，可以认为电价减少对负荷的影响可线性表示为10.30 kW·h/p。与前述电价增加的影响相比，二者相差一个数量级，这也表明居民负荷中存在部分硬性需求，难以通过增加电价来显著降低，但可以通过降低电价来刺激用电量的增加。

此外，还发现一个有趣的现象，即在电价回调后的一段时间内，负荷依然处于需求响应的状态，用户的用电恢复存在一定的滞后性。这意味着可以通过较短时间的电价激励来实现更长时间的负荷调整，从而有效节约调节成本。

# 拓展任务1：数据价值评估方案设计

结合已有数据集，设计用户数据价值评估方法及交易商业模式，需考虑获取成本、用户模型效益、盈利空间等方面对数据进行定价并设计一种考虑安全性和市场价值的数据交易模式。

## 用户数据价值评估策略

在分析一个用户数据价值评估策略是否合理时，可以从以下设计原则入手：

1、综合性原则：在评价用户数据价值时，需要全面考虑数据的获取成本、用户模型效益、盈利空间等多个维度，确保评估结果的客观性和准确性。这其中数据的获取成本包括数据采集、存储和处理成本，而用户模型效益评估涉及用户需求预测、用户满意度提升以及用户价值挖掘，盈利空间评估也包含多方面，包括市场潜力分析、竞争对手分析、盈利模式设计。

2、动态性原则：随着电力市场环境和用户需求的变化，数据价值评估方法应当具有动态调整的能力，以适应不断变化的市场需求。

3、可操作性原则：一个合理的评估策略对于电力企业而言应当具有较强的可操作性，以便根据实际情况进行数据定价。

后两条原则在设计具体数据定价策略时有重要作用，从这两条原则出发可以提出三种不同的定价策略：基于成本加成的定价策略是以数据获取成本为基础，加上一定的利润空间，以此来确定数据价格；基于市场需求的定价策略则是根据市场对数据的需求程度，调整数据价格，从而实现供需平衡；基于用户价值的定价策略是根据用户模型效益和盈利空间，对数据价格进行动态调整。

## 数据交易商业模式

合格的数据交易商业模式必须兼顾以下三点：安全性、市场价值以及绿色。

在安全性方面，可以采用区块链技术实现去中心化的数据交易，保证数据的完整性和不可篡改性，同时通过加密技术保护数据隐私。

在市场价值方面，可以设立数据交易所或者相关平台，提供透明的市场价格信息，促进供需双方的匹配，同时引入第三方评估机构，对数据的质量和价值进行公正评价。

同时为了实现绿色，在数据交易的过程中，应当鼓励和支持使用可再生能源和节能技术，减少数据处理和传输过程中的碳足迹。

参考文献

1. 陈苏豫，顾亦然，张腾飞．基于DLT-Kmedoids算法的用电负荷聚类分析[J]．计算机技术与发展，2024，34(04)：205-211．
2. 白雅玲，周亚同，刘君．基于深度卷积嵌入聚类的日负荷曲线聚类分析[J]．电网技术，2022，46(06)：2104-2113．
3. 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.

附录A 任务分工记录表格

|  |  |
| --- | --- |
| 组员 | 工作量占比（%） |
| 朱继涛 | 33.33 |
| 吴晨聪 | 33.33 |
| 李德瑜 | 33.33 |