

基于非侵入式电力负荷监测与分解的 电力分析

泰迪智能科技（武汉）有限公司

目录

contents

- ① 研究背景与目标
- ② 分析方法与过程
- ③ 数据预处理与特征工程
- ④ 模型搭建与训练
- ⑤ 结果分析及应用

第一部分

研究背景与目标

- 研究背景
- 数据说明
- 概念说明
- 分析目标

研究背景

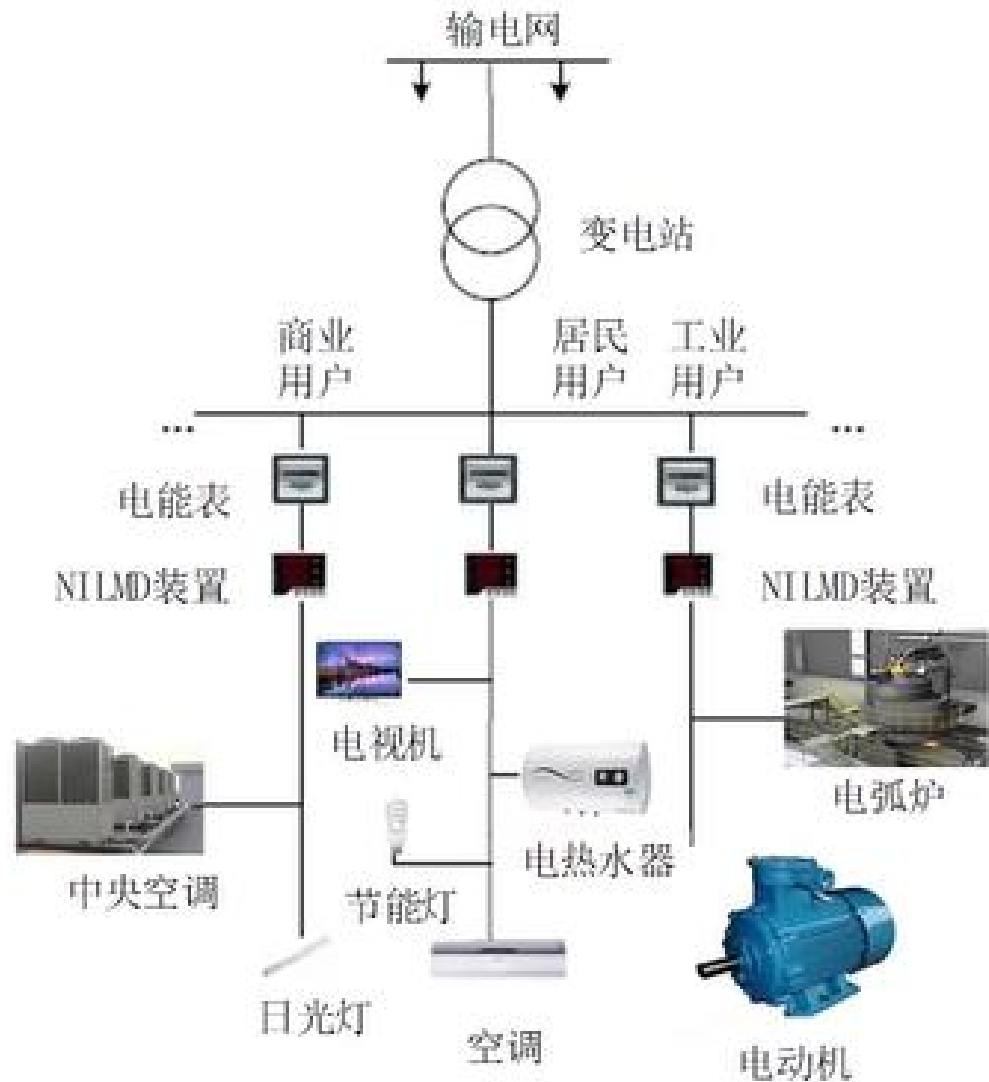
- 传统的电能能耗监测主要借助于电能表，在入户线上的电能表可以获取用户的总能耗数据，而电力分项计量可以对连接到入户线后的建筑物内各个用电设备所消耗的电能进行独立计量。基于电力分项计量的一系列技术，是将电器识别作为物联网的重要研究方向，能够为电力公司和用户带来很多便利，在生产和生活中有非常实际的意义。
- 电力分项计量技术主要分为两种：一种是侵入式电力负荷监测（Intrusive residential Load Monitoring, ILM），是为用户的每一个用电设备安装一个带有数字通信功能的传感器，通过网络采集各设备的用电信息；另一种是非侵入式电力负荷监测与分解（Non-Intrusive Load Monitoring and Decomposition, NILMD），是在用户的电能入口处安装一个传感器，通过采集和分析用户的用电总功率或总电流来监测每个或每类用电设备的功率及工作状态。
- 基于NILMD技术的用电分析计量具有简单、经济、可靠和易于迅速推广应用等优势，更加适用于居民用户。



研究背景与目标

研究背景

- 非侵入式电力负荷监测与分解结构如图所示；
- NILMD装置测量得到的是整个线路上的电压和电流数据，它们可以看作是各个用电设备的电压和电流数据的叠加；
- NILMD的核心是如何从采集到的整条线路的电压和电流数据中“分解”出每个用电设备独立的用电数据。



NILMD示意图

研究背景

- 如同人类的声纹和指纹等具有唯一性的生物特征可以用来实现个体识别一样，用电设备的负荷印记可以识别不同种类和型号的用电设备中相对稳定且较为显著的特征，如在运行过程中产生的电压、电流和谐波等时序数据中的特征。
- 而根据用电设备运行的过程，又可将数据分为暂态数据和稳态数据两大类，其中暂态数据主要指设备启动、设备停止、设备模式切换时的状态数据，稳态数据主要指设备稳定运行时的状态数据。
- NILMD系统的目标是根据不同类型用电设备独特的负荷印记，从一个能源网关设备记录的数据中检测出接入该能源网关设备的电路中各种用电设备的开关等操作，并对其用电量进行分项计量。



研究目标

1. 根据附件 1 提供的单一态数据，分析并给出各用电设备的运行特征，并估计各用电设备的实时用电量；
2. 根据附件 2 中的单一态数据，设计自动识别单一设备的数学模型和计算方法，并估计这一用电设备的实时用电量。



数据说明

- ▶ 本案例研究的用电设备共11种，其类型及工作参数如表所示。
- ▶ 本案例数据分为训练数据和测试数据两部分，训练数据包含了表所示的11种设备分别的用电数据，测试数据为表所示的11种设备中某两种设备的用电数据。
- ▶ 训练数据中的每一种设备都包含4张表，分别为设备数据、周波数据、谐波数据和操作记录。
- ▶ 测试数据中的每一种设备都包含3张表，分别为设备数据、周波数据和谐波数据。

序号	设备ID	设备类型	工作参数
1	YD1	落地风扇	220V, 60W
2	YD2	微波炉	220V, 输入: 1150W, 输出: 700W
3	YD3	热水壶	220V, 1800W
4	YD4	笔记本电脑	20V, 3.25A/4.5A
5	YD5	白炽灯	22V, 40W
6	YD6	节能灯	220V, 5W
7	YD7	激光打印机	220~240V, 50~60Hz, 4.6A
8	YD8	饮水机	220V, 制热: 430W, 制冷: 70W, 总: 500W
9	YD9	挂式空调	220V, 2600W
10	YD10	电吹风	220V, 50Hz, 1400W
11	YD11	液晶电视	220V, 50Hz, 150W

数据说明

设备数据表（sheet1）结构如表所示。



序号	特征	备注
1	time	年月日时分秒
2	IC	电流，单位：0.001A
3	UC	电压，单位：0.1V
4	PC	有功功率，单位：0.0001kW
5	QC	无功功率，单位：0.0001kVar
6	PFC	功率因数，单位：%
7	P	总有功功率，单位：0.0001kW
8	Q	总无功功率：单位：0.0001kVar
9	PF	总功率因数，单位：%

数据说明

- 周波数据表（sheet2）结构如表所示。
- 周波表示交流电完成一次完整变化的过程(即一个正弦波形)。因为我国交流电供电的标准频率为50Hz，所以NILMD装置在其中一个周期内（0.02秒）可采集128个时间点上的数据。



序号	特征	备注
1	time	年月日时分秒
2	ICi	i范围为001~128，电流一个周波的第i个采样点（XXX.XXX）
3	UCj	j范围为001~128，电压一个周波的第j个采样点（XXX.XXX）

数据说明

- 谐波数据表（sheet3）结构如表所示。



- 当供电线路中的正弦波电压施加在非线性电路上时，电流就变成非正弦波，非正弦电流在电网阻抗上产生压降，会使电压波形也变为非正弦波。非正弦波可用傅立叶级数分解，其中频率与工频相同的分量称为基波，频率大于基波的分量称为谐波。在电力行业中，谐波是指工频频率的整数倍的交流电。因为我国电网规定工频频率是50Hz，所以基波频率是50Hz，这样5次谐波电压（电流）的频率就是250Hz。

序号	特征	备注
1	time	年月日时分秒
2	ICi	i为02~51，i次电流谐波，表示谐波的含有率（XX.XX%）
3	UCj	j为02~51，j次电压谐波，表示谐波的含有率（XX.XX%）

数据说明

➤ 操作记录数据表（sheet4）结构如表所示。

设备数据

周波数据

谐波数据

操作记录

序号	时间	设备	工作状态	操作
1	2018/1/25 14:55:44	YD1	关闭	
2	2018/1/25 14:56:43	YD1	1档	启动
3	2018/1/25 14:59:38	YD1	2档	切换
4	2018/1/25 15:03:31	YD1	3档	切换
5	2018/1/25 15:06:59	YD1	关闭	关闭

»» 概念说明

- 有功功率（ P ）是保持用电设备正常运行所需的电功率，也就是将电能转换为其他形式能量（机械能、光能、热能）的电功率。如5.5kW的电动机是将5.5kW的电能转换为机械能，带动水泵抽水或脱粒机脱粒。而各种照明设备是将电能转换为光能，供人们生活和工作照明。
- 无功功率（ Q ）是用于电路内电场与磁场的交换，并用于在电气设备中建立和维持磁场的电功率。无功功率不对外做功，而是电场能与磁场能之间的相互转换。凡是有电磁线圈的电气设备要建立磁场，就要消耗无功功率。如40W的日光灯除了需要超过40W的有功功率（镇流器也需消耗一部分有功功率）来发光外，还需要80var左右的无功功率，用于供镇流器的线圈建立交变磁场。由于无功功率不对外做功，所以被称之为“无功”。

»» 概念说明

- 视在功率(S)等于电压有效值与电流有效值乘积，表示电源的输出能力。
- 功率因素为有功功率与视在功率的比值，由电压与电流之间的相位差角 φ 决定。
- 视在功率计算公式为 $S = UI$ ，单位为 VA ；有功功率计算公式为 $P = UI\cos\varphi = S\cos\varphi$ ，单位为 W ；无功功率计算公式为 $Q = UI\sin\varphi$ ，单位为 var 。其中， U 为电压， I 为电流， $\cos\varphi$ 为功率因素。



» 概念说明

- 根据用户用电设备工作状态的不同，可将用电设备分为以下3种类型。
 - 启/停二状态设备（ON/OFF）。这类用电设备只有运行和停机两种用电状态，如白炽灯、电热水壶等。
 - 有限多状态设备。这类用电设备通常具有有限个分立的工作状态，与之相对应的用电功率间是离散的，不同的功率水平即标志着不同的工作状态，如洗衣机、微波炉、电磁炉等。
 - 连续变电状态设备。这类用电设备的稳态区段功率无恒定均值，而是在一个范围内连续变动，如变频空调、电动缝纫机等。

»» 分析目标

- 本次分析的目标是根据NILMD装置测得的整个电路上的电力数据，利用k近邻模型，实现从整条线路中“分解”出每个用电设备的独立用电数据。

主要包括以下4个步骤。

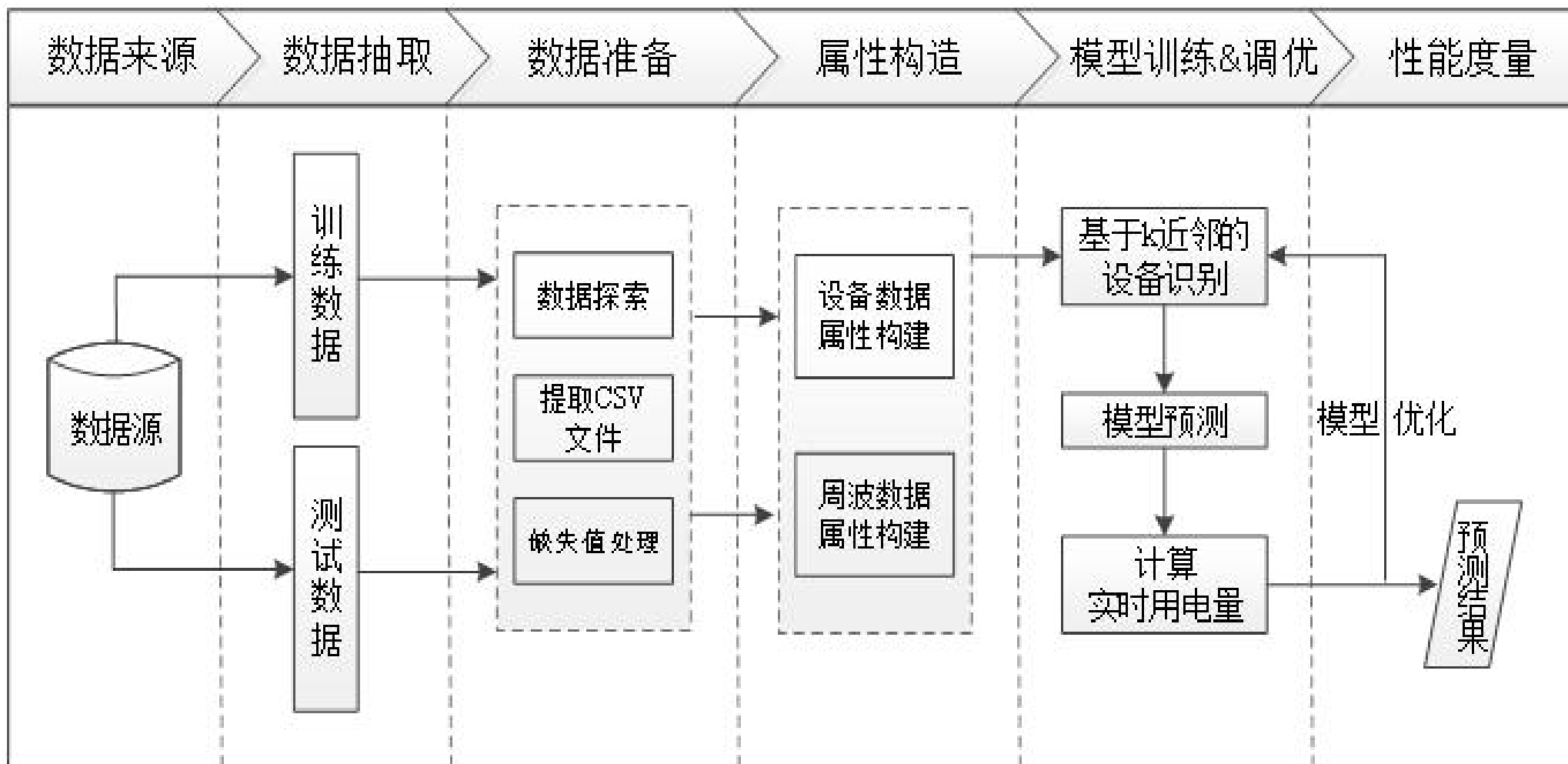
1. 抽取11个设备的电力分项计量的数据。
2. 对抽取的数据进行数据探索、缺失值处理和特征构造等操作。
3. 使用k近邻算法进行设备识别。
4. 计算实时用电量。

第二部分

分析方法与过程

- 分析过程图
- K近邻算法简介

分析过程图



»» K近邻算法简介

➤ K近邻算法简介：

K近邻算法（K-Nearest Neighbors，简称KNN）是一种基本且广泛使用的监督学习算法，主要用于分类和回归问题。它的核心思想是通过测量不同特征点之间的距离来进行预测和分类。

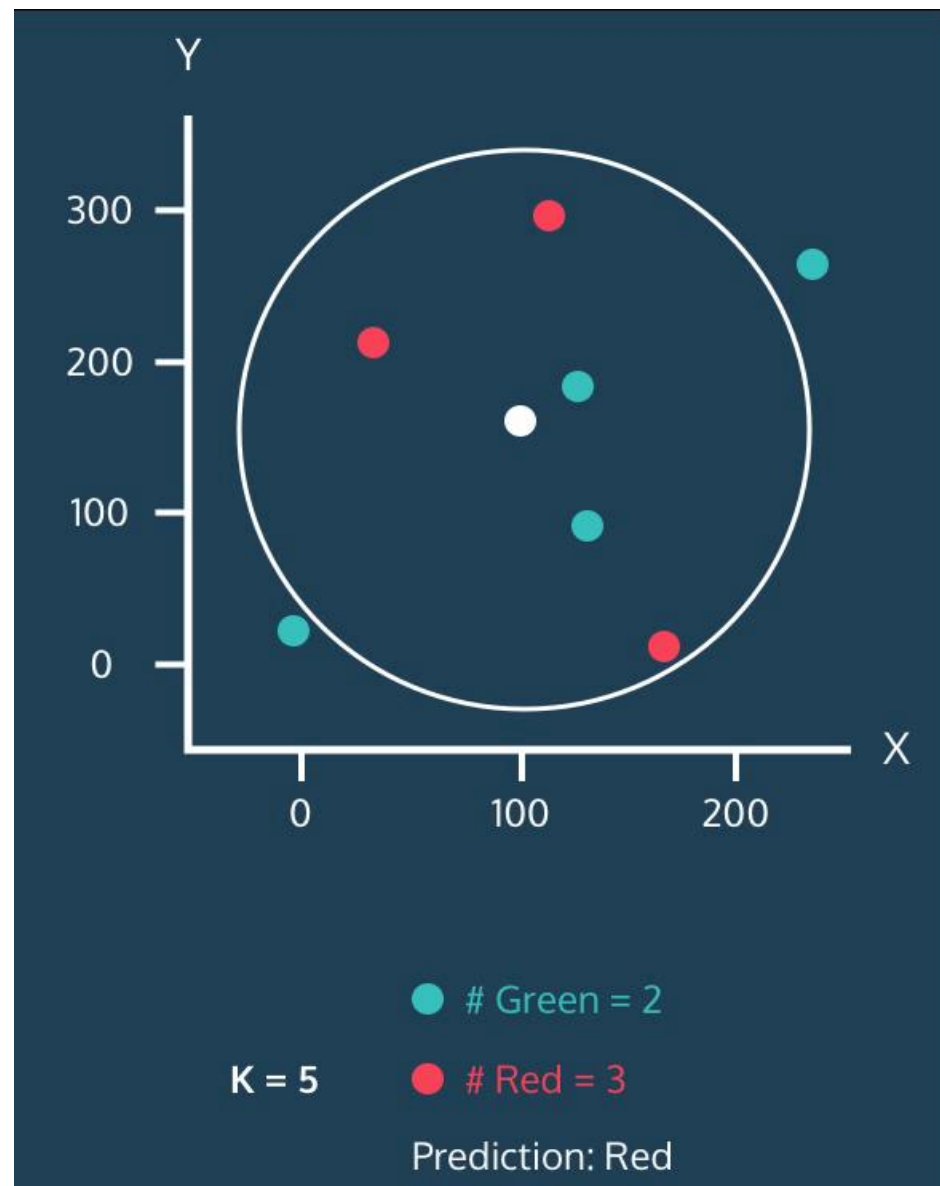
➤ 基本原理：

- 距离度量：首先确定一个距离度量方法，常用的有欧氏距离、曼哈顿距离和切比雪夫距离等。这个距离用来衡量样本点之间的相似度。
- 最近邻：对于一个给定的测试样本，计算它与训练集中所有样本的距离，并找出距离最近的K个样本点，这些点被称为最近邻。
- 投票机制（分类）：在分类问题中，K个最近邻的类别会被用来决策。通常采用投票机制，即哪一类的票数最多，测试样本就被分为哪一类。
- 平均机制（回归）：在回归问题中，K个最近邻的数值会被用来预测测试样本的数值。一般采用平均值或中位数作为预测结果。

» K近邻算法简介

➤ 工作流程

1. 确定K值：选择一个正整数K，它代表最近邻的数量；
2. 计算距离：对于每一个待分类的数据点，计算它与训练集中所有点之间的距离，距离的计算方式可以自己定义；
3. 找到最近的K个邻居：根据计算出的距离，找出距离最近的K个训练样本；
4. 进行投票：这K个邻居中最常见的类别（对于分类问题）或平均值（对于回归问题）将被用来预测新的数据点。



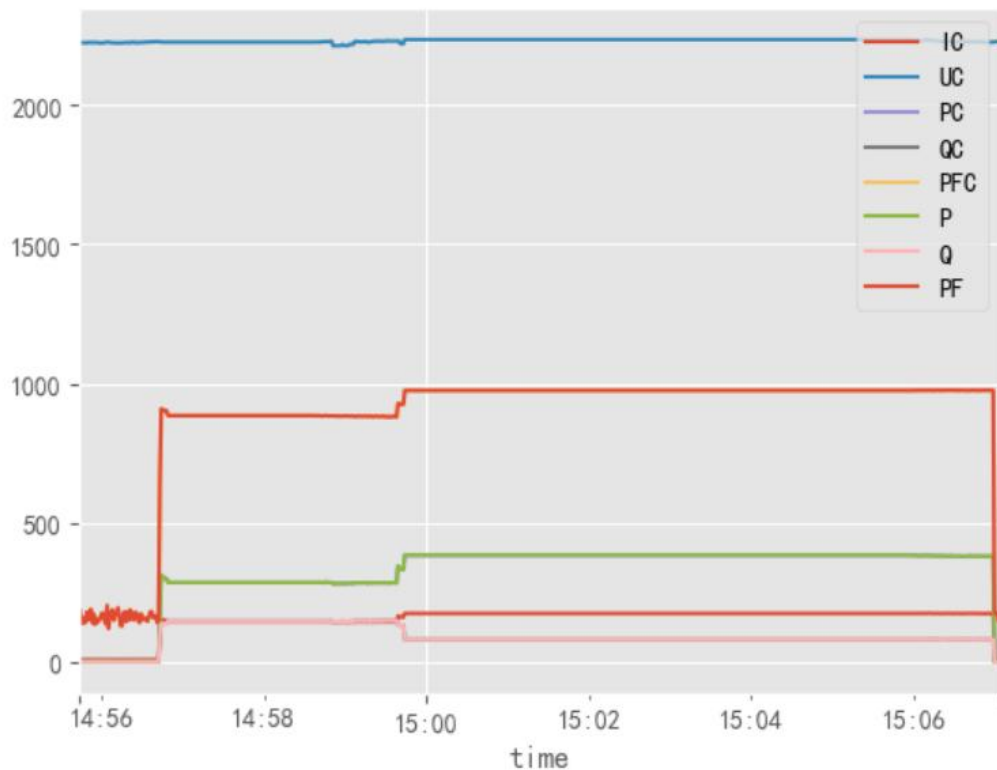
第三部分

数据预处理与特征工程

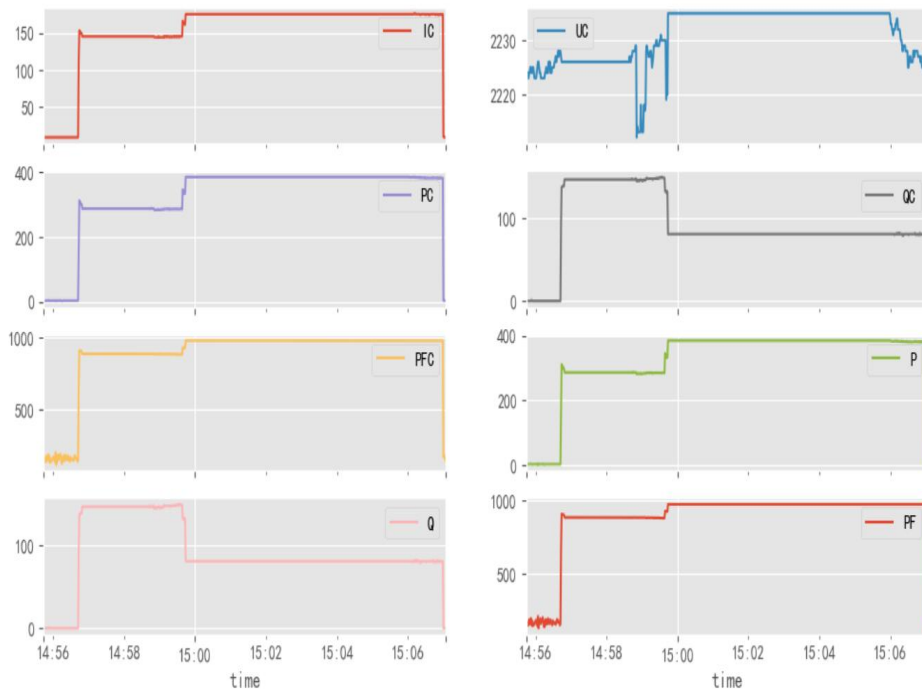
- 数据初探索
- 数据处理
- 特征工程

数据初探索（设备数据）

- 在获取数据后，由于数据表较多，每个表的特征也较多，所以需要我们对数据进行数据探索分析。在数据探索过程中主要根据原始数据特点，对每个设备的不同特征对应的数据进行可视化，并根据得到的折线图对数据特征进行分析。



设备数据YD1可视化（一张图）

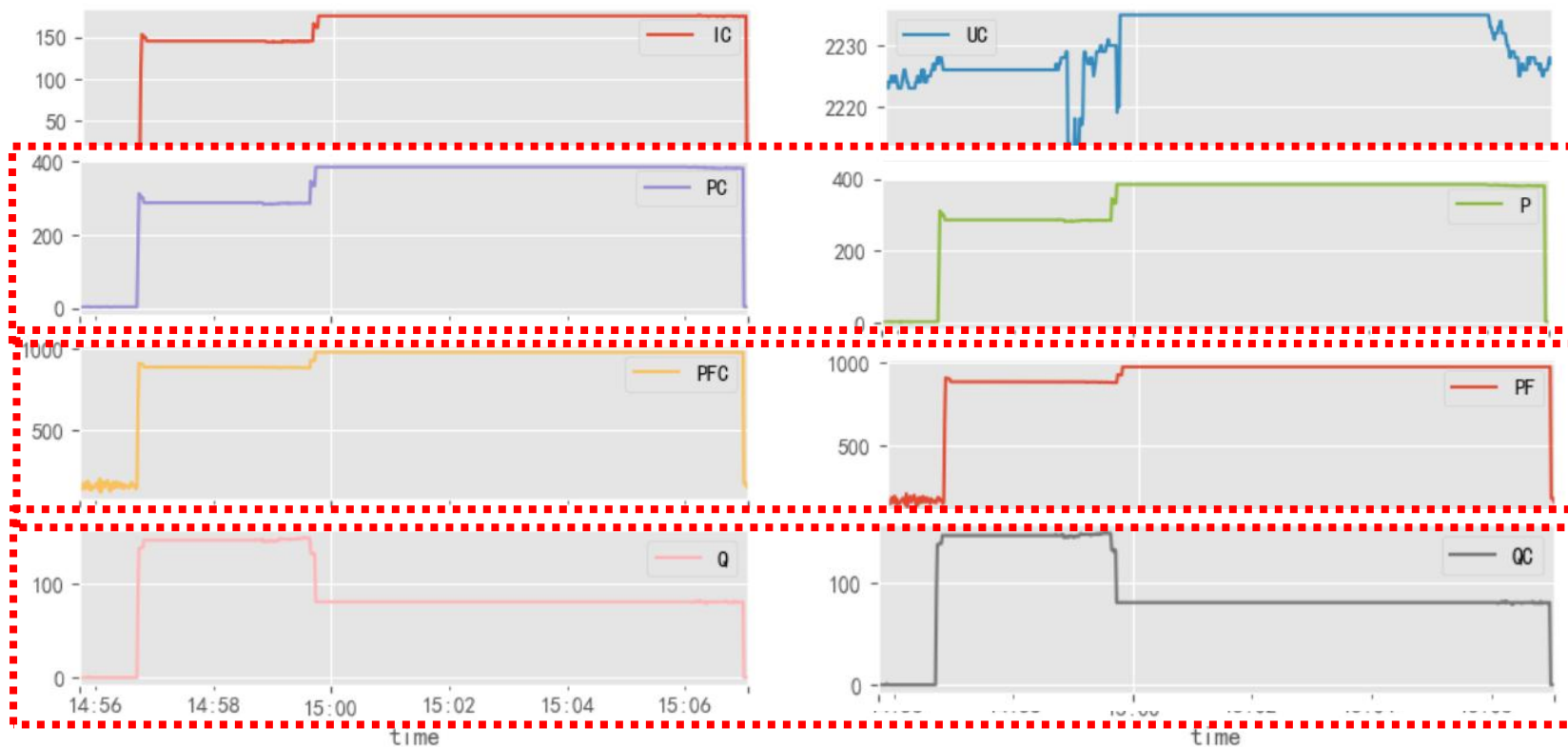


设备数据YD1可视化（分散）



结论1：设备数据初探索

- 同一种设备无功功率和总无功功率、有功功率和总有功功率、功率因素和总功率因素的折线图基本重合，说明对不同设备的区分度差别很小。



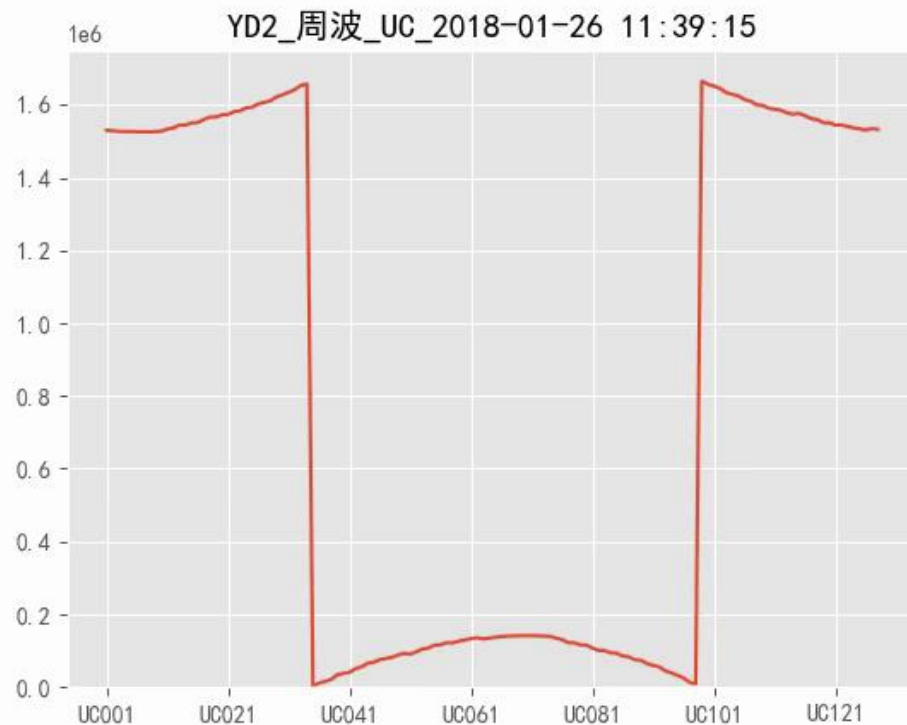
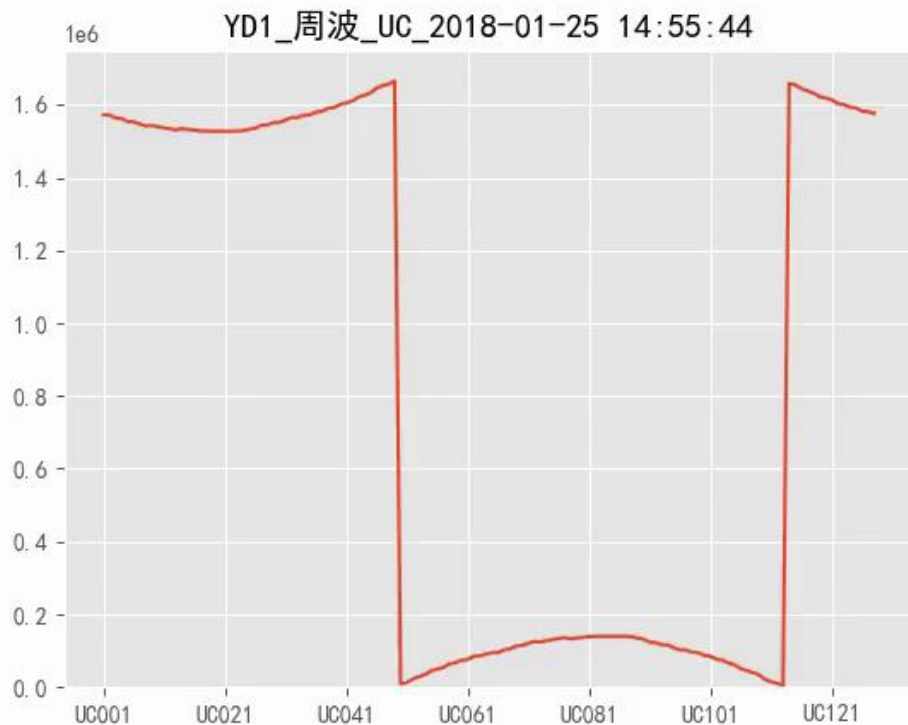
设备数据YD1各属性随时间变化对比

*可视化展示同一电器内、不同电器之间各变量之间存在的显著联系。（此处仅展示YD1）



结论2：周波数据电压变化对比

- 周波数据中电压均呈周期性，无明显变化，故后续不适用使用此特征作为模型输入。

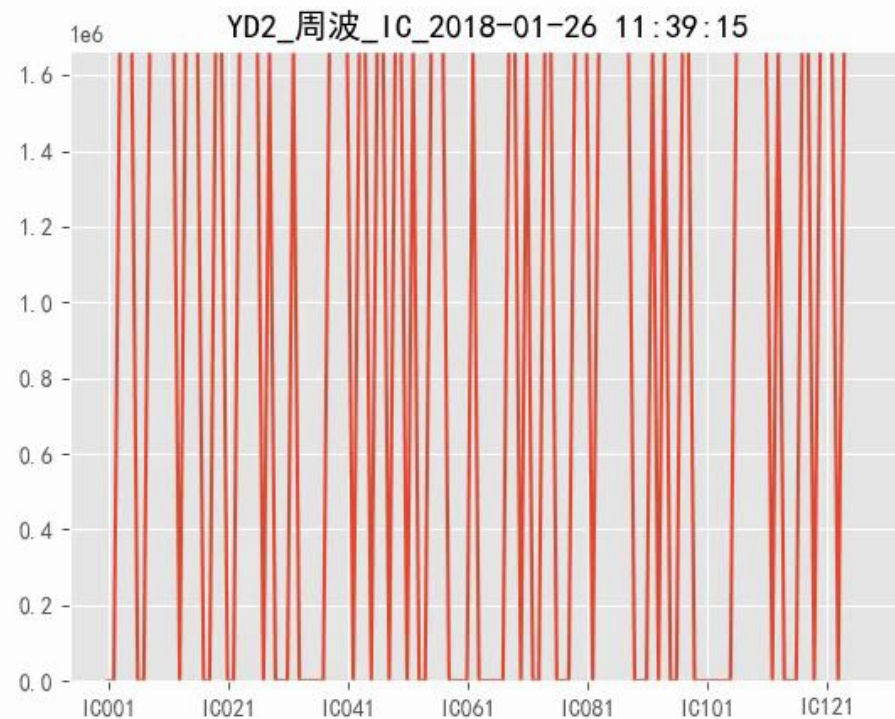
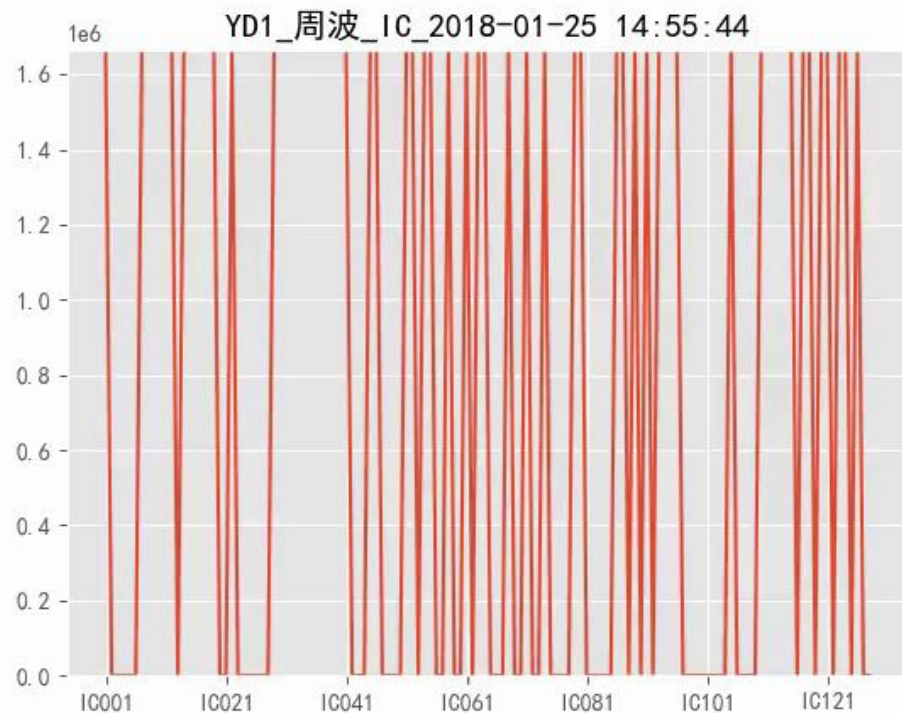


*YD1到YD11电压变化都一致，这里仅展示YD1和YD2



» 结论3：周波数据电流变化对比

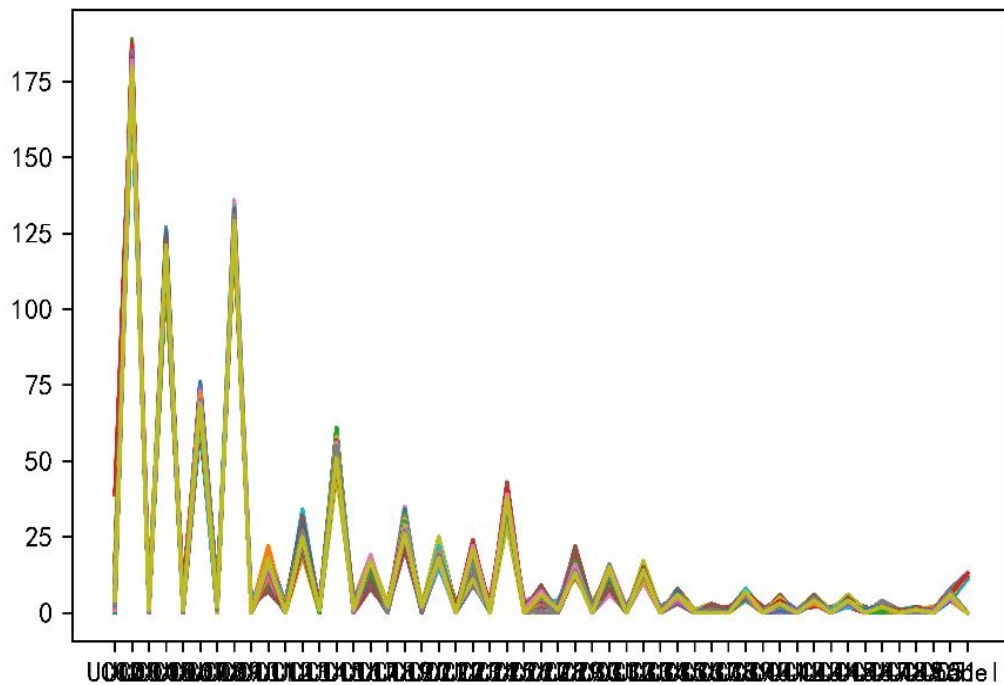
- 不同设备电流变化是存在明显变化的，且各设备电流变化不一样；故后续适用此特征作为模型输入。



*YD1到YD11电流变化都不一致，这里仅展示YD1和YD2

结论4：谐波数据电流电压对比

- 与周波数据类似，周波数据中电压呈周期性，无明显变化，故后续不适用使用此特征作为模型输入；
- 不同设备电流变化是存在变化的，且各设备电流变化不一样。



11种电器谐波电压变化



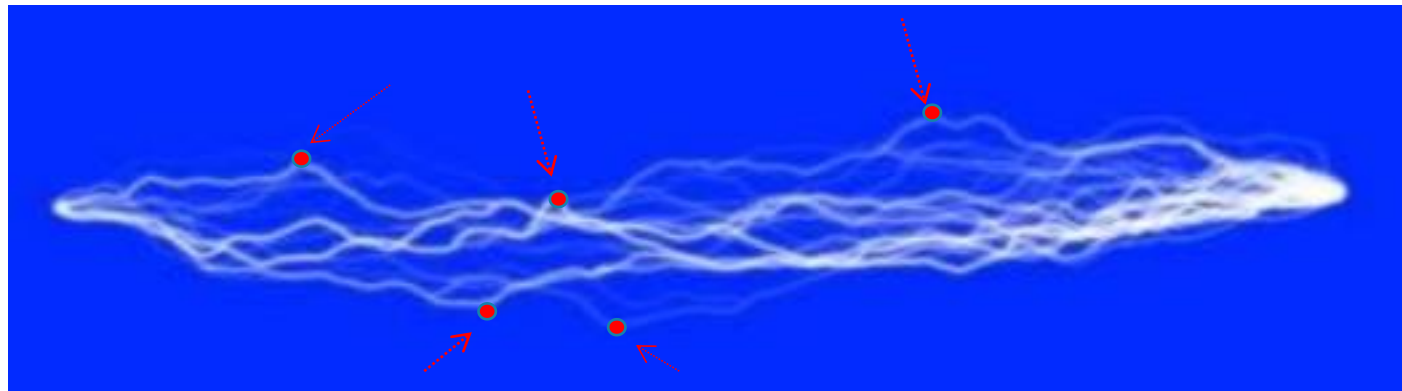
11种电器谐波电流变化

» 数据处理

- **提取数据：**在进行缺失值处理之前，需要将训练数据中所有设备数据中的设备数据表、周波数据表、谐波数据表和操作记录表，以及测试数据中所有设备数据中的设备数据表、周波数据表和谐波数据表都提取出来，作为独立的数据文件。
- **缺失值处理：**通过数据探索，发现数据中部分“time”特征存在缺失值，需要对这部分缺失值进行处理。由于每份数据中“time”特征的缺失时间段长不同，所以需要进行不同的处理。对于每个设备数据中具有较大缺失时间段的数据进行删除处理，对于具有较小缺失时间段的数据使用前一个值进行插补。
- **重复值处理：**由于缺失值处理后，每个设备数据都由一张表变为了多张表，所以需要将相同类型的数据表合并到一张表当中，如将所有设备的设备数据表合并到一张表当中。同时，因为缺失值处理的其中一种方式是使用前一个值进行插补，所以产生了相同的记录，需要对重复出现的记录进行处理。

特征工程

- **进行特征提取的重要性：**虽然在数据准备过程中对特征进行了初步处理，但是引入的特征太多，而且这些特征之间存在着信息的重复。故为了保留重要的特征，建立精确、简单的模型，需要对原始特征进一步筛选。
- 例如：在本案例中，不同电器的电流随时间变化各不相同，我们就可以提取电流的**波峰**和**波谷**作为判别的特征。



不同电器电流随时间变化

特征工程（设备数据）

【设备数据】特征提取：

在“数据探索”过程中发现，同一种设备无功功率和总无功功率、有功功率和总有功功率、功率因数和总功率因素的折线图基本重合，故上述特征二选一即可；故本案例在设备数据中选择**电流、电压、无功功率、有功功率、和功率**作为设备数据的特征构建判别特征库。结果如下图：

	IC	UC	PC	QC	PFC
time					
2018-01-25 14:55:44	9.0	2222.0	3.0	0.0	184.0
2018-01-25 14:55:45	9.0	2223.0	3.0	0.0	150.0
2018-01-25 14:55:46	9.0	2224.0	2.0	0.0	139.0
2018-01-25 14:55:47	9.0	2224.0	3.0	0.0	160.0
2018-01-25 14:55:48	9.0	2223.0	3.0	0.0	142.0
...



» 特征工程（周波数据）

- 在“数据探索”过程中发现，周波数据中的电流随着时间的变化有较大的起伏，不同设备的周波数据中的电流绘制出来的折线图的起伏不尽相同，具有明显的差异，故本案例选择波峰和波谷作为周波数据的特征构建判别特征库。
- 由于原始的周波数据中并未存在电流的波峰和波谷两个特征，所以需要进行特征构建。同时可以增加波峰数量和波谷数量来增加数据特征。运行结果如下图：

	波峰	波谷	波峰_num	波谷_num
2018-01-25 14:55:44	1666597.296875	70.28125	64.0	64.0
2018-01-25 14:55:45	1666580.951613	74.469697	62.0	66.0
2018-01-25 14:55:46	1666578.078947	70.730769	76.0	52.0
2018-01-25 14:55:47	1666587.533333	85.176471	60.0	68.0
2018-01-25 14:55:48	1666588.013333	90.603774	75.0	53.0
...



»» 特征工程（谐波数据）

- 在“数据探索”过程中发现，谐波数据中的电流随着时间的变化有较大的起伏，不同设备的谐波数据具有明显的差异，电流采样点折线图变化较为明显，但采样点特征过多，此处便采用基本统计量（**最大值、均值、标准差、偏度、峰度、中值、0.9分位数**）作为输入特征；
- 谐波数据中电压采样点特征无明显变化，故不采用。

	最大值	均值	标准差	峰度	偏度	中位数	0.9分位数
time							
2018-01-25 14:55:44	9671.0	4819.22	2557.989072	-0.936076	0.207642	4488.5	8098.6
2018-01-25 14:55:45	9060.0	4495.16	2169.792517	-0.413243	0.241491	4175.0	7577.9
2018-01-25 14:55:46	9937.0	5026.18	2762.524102	-1.314672	0.231930	4490.5	8974.0
2018-01-25 14:55:47	9553.0	5491.84	2679.053567	-0.995933	-0.283063	5693.5	8858.8
2018-01-25 14:55:48	9433.0	4088.14	1998.769908	0.032606	0.378618	3880.0	6723.9
...



» 特征工程（合并数据）

- 由于设备数据、周波数据和谐波数据各式一个dataframe数据框，此时我们需要把三者的特征进行合并。

代码：data_feature_fe = pd.concat([feature1,feature21, feature31_new],axis=1)

运行结果：

[17]:	IC	UC	PC	QC	PFC	波峰	波谷	波峰_num	波谷_num	最大值	均值	标准差	峰度	偏度	中位数	0.9分位数
2018-01-25 14:55:44	9.0	2222.0	3.0	0.0	184.0	1666597.296875	70.28125	64.0	64.0	9671.0	4819.22	2557.989072	-0.936076	0.207642	4488.5	8098.6
2018-01-25 14:55:45	9.0	2223.0	3.0	0.0	150.0	1666580.951613	74.469697	62.0	66.0	9060.0	4495.16	2169.792517	-0.413243	0.241491	4175.0	7577.9
...
2018-01-26 11:32:58	53.0	2254.0	4.0	116.0	35.0	1666160.047619	501.923077	63.0	65.0	1383.0	325.80	254.095916	6.127384	2.153046	267.0	539.7

4337 rows × 16 columns

第四部分

模型的搭建与训练

- 模型搭建
- 模型训练



» 模型搭建

➤ step1:数据标准化

由于不同特征 x_i 之间的数据量级相差过大，特征选择可能会受到特征尺度的影响。一些基于统计量的特征选择方法可能会偏向于那些尺度较大的特征，从而忽略了尺度较小但可能更重要的特征。标准化公式如下：

$$x_{ij} = \frac{x_{ij} - \frac{1}{N} \sum_{j=1}^N x_{ij}}{\sqrt{\frac{1}{N-1} \sum_{j=1}^N (x_{ij} - u)^2}} \quad (i = 1, \dots, 16)$$

➤ step2:使用KNN算法

在判别设备种类时，选择k近邻模型进行判别，邻居的个数 $N=5$ ，利用特征选择建立的特征库训练模型，然后利用训练好的模型对设备1和设备2进行判别。



模型训练

➤ 训练代码如下：

```
1 # 数据标准化
2 from sklearn.preprocessing import StandardScaler
3 scale0 = StandardScaler()
4 scale0.fit(x)
5 x0_ss = scale0.transform(x)
6 # 模型构建
7 from sklearn.neighbors import KNeighborsClassifier
8 model_knn = KNeighborsClassifier(n_neighbors=5)
9 model_knn.fit(x0_ss, y)
10 model_knn.score(x0_ss, y)
11 # 保存模型
12 from joblib import dump
13 dump(scale0, 'scale_ss')
14 dump(model_knn, 'model_knn')
```

➤ 在训练完模型之后保存模型是为了方便下一步做预测。

➤ 精确度对比：

原始125个特征	提取16个特征
0.6442	0.9186

➤ 结论：相较于原始的125个特征，提取的16个特征对于模型的分类精度提升帮助更大。

第五部分

结果分析与应用

- 测试数据预处理
- 模型预测
- 模型评估
- 分析与意义

测试数据预处理

- 此项目的模型预测与一般的有一点区别，之前的模型预测都是在一个数据集上进行训练集和测试集的划分，但是这个项目的训练和测试是完全分开的，因此我们在预测之前需要将测试数据的格式修改为训练集的样子。首先我们需要进行与之前类似的预处理，然后进行一样的特征提取。

```
# 预处理
def fix_na(data):
    tmp2 = data.copy()
    # 1 时间格式转换
    tmp2['time'] = pd.to_datetime(data['time'])
    # 2 时间作为索引
    tmp2 = tmp2.set_index('time')
    # 3 重采样
    tmp3 = tmp2.resample('S').max()
    # 4 缺失值处理
    ind = tmp3.iloc[:,0].isna()
    ind2 = pd.Series(ind[ind].index, index= ind[ind].index).diff().dt.seconds
    ind3 = ind2.loc[ind2 == 1].index
    # 删除
    tmp4 = tmp3.drop(ind3)
    tmp_res = tmp4.fillna(method='ffill')
    return tmp_res
```

- 处理后我们得到的数据集大小为 (526,16)

模型预测

- 根据分析目标，需要计算实时用电量。实时用电量计算的是瞬时的用电器的电流、电压和时间的乘积，如式所示。

$$W = \frac{100}{3600} \times P, P = UI \quad \text{其中} W \text{为实时用电量，单位是} 0.001kWh。P \text{为功率，单位为} W。$$

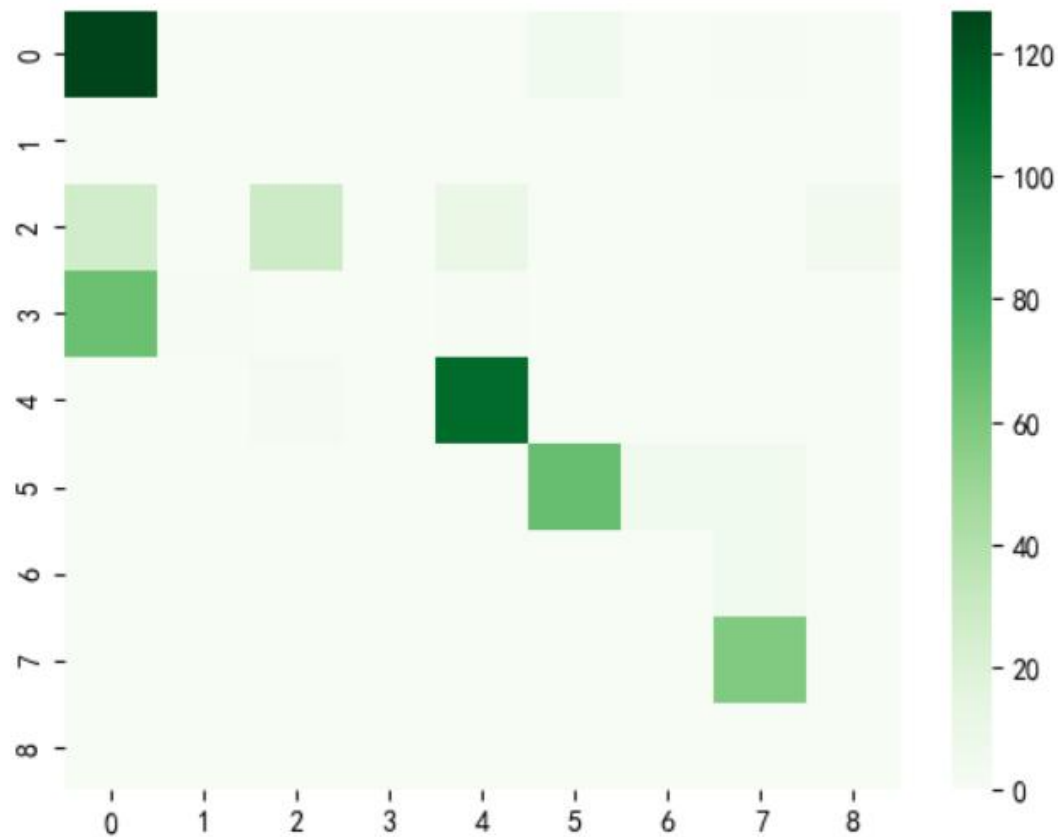
- 预测不同时段用电量和电器状态见右表

[13]:	用电量	label
2018-01-21 18:49:14	0.638889	0.0
2018-01-21 18:49:15	0.611111	0.0
2018-01-21 18:49:16	0.666667	0.0
2018-01-21 18:49:17	0.638889	0.0
2018-01-21 18:49:18	0.638889	0.0
...
2018-01-16 16:04:45	0.055556	0.0
2018-01-16 16:04:46	0.083333	0.0
2018-01-16 16:04:47	0.111111	0.0
2018-01-16 16:04:48	0.083333	0.0
2018-01-16 16:04:49	0.083333	0.0

1859 rows × 2 columns

模型评估（热力图）

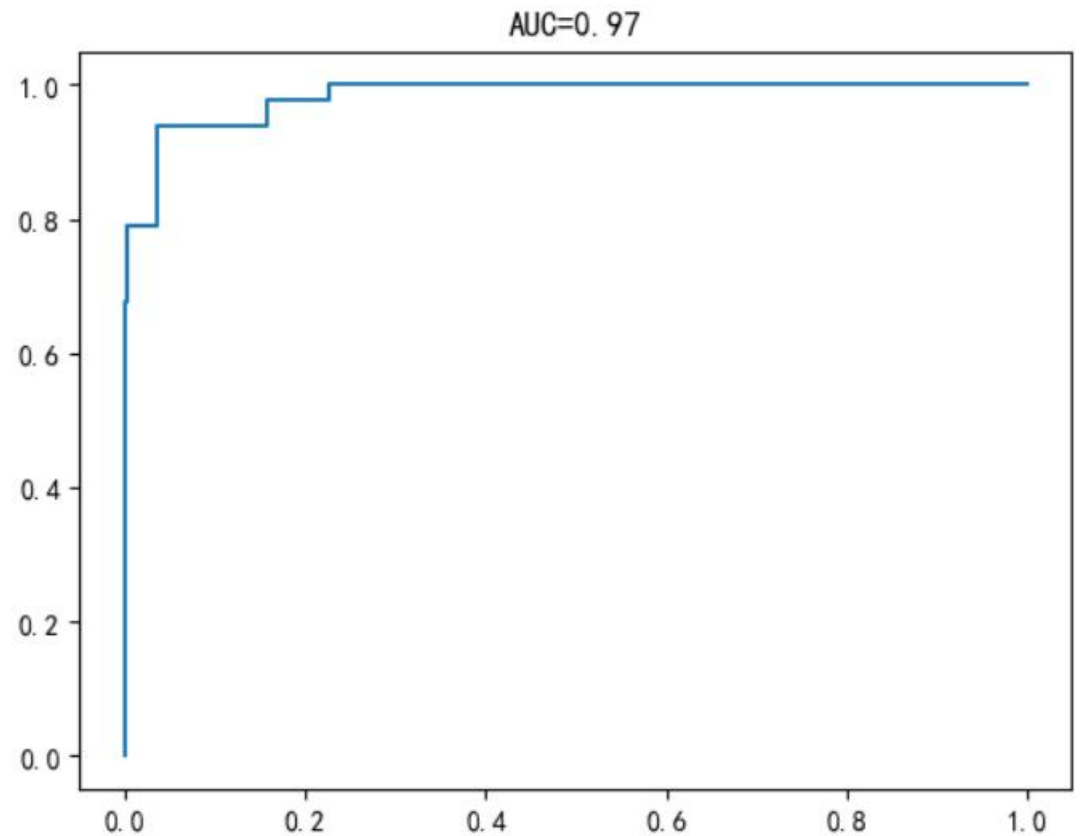
- 我们所建立的模型对测试数据预测的怎么样呢？我们就需要对此次预测进行评估。
- 理解这个热力图可以帮助我们评估模型的性能：
 - 对角线上的数值（TP和TN）越大越好，表示模型正确预测的正例和负例数量多。
 - 非对角线上的数值（FP和FN）越小越好，表示模型错误预测的正例和负例数量少。
- 通过观察热力图，可以快速识别模型在哪些类别上表现良好，以及哪些类别上可能需要改进。例如，如果热力图的左上角（TP）颜色很深，而右下角（FN）颜色较浅，这表明模型在正例分类上表现很好，但在负例分类上可能存在问题。



» 模型评估（ROC曲线）

➤ 根据代码建立的模型，对测试数据进行预测，模型的ROC曲线图如下；

- ROC曲线图，横轴FPR表示假正率（False Positive Rate），纵轴TPR表示真正率（True Positive Rate）ROC曲线通过展示在不同阈值下模型的分类性能，帮助我们理解模型在不同决策阈值下的表现。AUC值则提供了一个单一的性能度量，它告诉我们模型在所有分类阈值下的平均表现。
- ROC曲线尽可能地靠近左上角，这意味着在所有假正率水平下，模型都能保持较高的真正率。
 - AUC值为0.97接近于1，表示模型的分类性能越好。（AUC值为0.5时，表示模型没有区分能力，等同于随机猜测）。



»» 总结与分析

- 该项目通过在用户电能入口处安装传感器的NILMD技术，利用k近邻模型实现分解出每个用电设备的独立用电数据，并估计实时用电量的目标。
- 该项目首先对数据进行初步探索，包括可视化分析和特征分析、处理缺失值和重复值，提取不同sheet表格的关键特征，如电流波峰波谷、谐波数据的基本统计量等。
- 然后使用KNN算法建立模型，并通过特征选择优化模型。训练结果显示，提取的特征相较于原始特征能显著提高模型的分类精度。
- 最后通过热力图和ROC曲线评估模型性能，结果显示模型具有较高的分类准确性。
- 得到结论：该项目能够精确监测和分析居民用电情况，有助于节能减排和电力资源的合理分配。

»» 实际意义

- 该项目的实际意义主要体现在以下几个方面：
 - 节能和环保：通过对用电设备的精确监测和分析，用户可以了解各个设备的能耗情况，从而采取节能措施，减少能源浪费，对环保也有积极影响。
 - 电力管理优化：电力公司可以利用NILMD技术更好地管理电力分配和电网负载，提高电网运行效率和可靠性。
 - 用户用电行为分析：用户可以通过用电数据分析自己的用电行为，调整用电习惯，实现更经济的电费支出。
 - 智能电网建设：该项目是智能电网建设的一部分，有助于推动智能电网技术的发展和應用。

综上所述，基于非侵入式电力负荷监测与分解的电力分析项目具有重要的实际应用价值，能够在多个层面促进能源的高效使用和管理。

谢谢