

第六届“泰迪杯” 数据挖掘挑战赛

优秀作品

作品名称：基于非侵入式负荷检测与分解的电力数据挖掘

荣获奖项：特等奖

作品单位：西南石油大学

作品成员：李红燕 瞿丹 林江

指导老师：丁显峰

基于非侵入式负荷检测与分解的电力数据挖掘

摘要

近年来因传统的电力负荷监测方法成本高、效率低，在应用上受到很大制约，而非侵入式负荷检测系统仅在电网的用户入口处安装一个传感器，通过采集和分析用户用电总功率或总电流来监测每个或每类用电设备的用电功率和工作状态，从而了解用户家中每个或每类用电设备的耗电情况和用电规律，对于指导用户科学合理用电和电力公司准确预测电力负荷、科学制定电网调度方案、提高电力系统稳定性和可靠性有着重要意义。

在本文中，利用 `matlab` 软件对所给出的相关数据通过小波降噪，清除异常点等预处理。

问题一，分析预处理的数据，提取单一设备的暂态特征，有功功率与无功功率、电流谐波、电压-电流轨迹（V-I trajectory）这四类特征，并给出了各负荷特性特征值的计算提取方法。采用公式计算各个单一设备的实时用电量，计算结果详见附件 `energy1.xlsx`。

问题二，建立模型对任意单一设备进行识别。通过问题一提取的四类负荷特征建立基于相似度的负荷识别模型，首先定义了特征相似性来表示任意两个设备在某个特征下的相似程度，通过熵值法确定所得各个特征的相似度权系数；其次将特征相似度加权求和确定总的特征相似度，选择相似度最高的设备与未知设备匹配；最后采用模型计算未知设备的实时用电量，计算结果详见附件 `energy2.xlsx`。

问题三，采用基于事件检测的非侵入式用电负荷分解方法对多负荷进行状态识别。首先对可能发生事件的时间点进行检测，本文以有功功率的改变值作为事件检测的依据并给出了算法的流程和具体处理方法；其次建立 0-1 规划模型进行多负荷的运行状态识别，提出了一种基于连续化的 0-1 二次规划的非侵入式负荷识别算法，确定出各用电设备的状态、操作与操作时间，并通过附件 3 已知操作记录的设备组验证了上述方法的有效性；最后建立了多负荷下的实时用电量估计模型，用以计算设备组的各设备的实时用电量，计算结果详见附件 `energy3.xlsx`。

问题四，采用基于决策树的非侵入式用电负荷识别方法。本文利用决策树算法结合用电设备数据库，实现了在不同组合负荷运行时的采样数据的负荷识别。首先将提取出的待识别负荷的特征参数与数据库中已知负荷的特征参数进行比较匹配，选择最近似的作为识别结果确定设备组的设备构成；其次采用问题三的 0-1 规划模型确定各个设备的状态、操作及操作时间，计算结果详见附件 `operation4.xlsx`；最后采用多负荷下的实时用电量估计模型计算设备组的各设备的实时用电量，计算结果详见附件 `energy4.xlsx`。

关键词：非侵入式；负荷识别；负荷分解；事件检测；0-1 规划；决策树

Abstract

Traditional monitoring methods have been severely limited in application due to cost and efficiency. In recent years, non-intrusive load detection systems only need to install a sensor at the total incoming line of a single-family home, through the residential power entrance. The total electricity consumption at the site is decomposed to obtain the electricity consumption and working status of the specific load, which not only can reduce the monitoring cost, but also can be used to guide the user's electricity consumption behavior, which in turn facilitates the electricity users to save electricity and achieve energy conservation and consumption reduction.

In this paper, we use MATLAB to carry out wavelet noise reduction processing on the relevant data and remove the abnormality point and other preconditioning steps.

The first problem is to analyze the preprocessed data, extract the transient characteristics of a single device, the four characteristics of active power and reactive power, current harmonic, voltage current track (V-I trajectory), and give the calculation and extraction method of characteristic eigenvalues of each load. The formula is used to calculate the real-time power consumption of each single device. The results are detailed in Annex 1.

The second problem, we establish a model to identify any single device. The similarity degree based load identification model is established by the four types of load characteristics extracted by the first problem. Firstly, the feature similarity is defined to represent the similarity degree of any two equipment under a certain feature. The entropy method is used to determine the similarity weight coefficient of each characteristic. Finally, the feature similarity is weighted to determine the total characteristics. Choose to match the highest similarity device with the unknown device. Finally, a formula is applied to calculate the real-time power consumption of the unknown equipment. The results are detailed in Annex 2.

The third problem is state recognition based on event detection and non-intrusive power load decomposition. First, the time point of the event can be detected. In this paper, the change value of active power is used as the basis of the event detection, and the process and the specific processing method of the algorithm are given. Then the 0-1 programming model is established for multi load running status recognition, and a non-intrusive load identification algorithm based on continuous 0-1 two times programming is proposed. The device group of the known operation records of Annex 3 validates the effectiveness of the above method. Finally, the real-time electricity consumption estimation model under multi load is established, and the real-time power consumption of each device is calculated. The calculation results are detailed in Annex 3.

The fourth problem, using the decision tree based non-intrusive power load identification method, proposes to use the decision tree algorithm and the database to realize the load identification from the sampling data of different combined load running. The characteristic parameters of the load identified by the load identification are compared with the characteristic parameters of the known load in the database, and the nearest similar identification results are selected to determine the equipment composition of the equipment group, and then the state of each device is determined by the 0-1 programming model of problem three. Finally, the real-time power consumption estimation model based on multi load is applied to calculate the real-time power consumption of each device in the device group. The results are detailed in Annex 4.

Key words: non-intrusive; load identification; load decomposition; event detection; decision

目录

1. 研究背景与意义.....	4
2. 变量说明.....	4
3 问题分析.....	5
4.问题一	6
4.1 数据预处理.....	6
4.1.1 降噪处理.....	6
4.1.2 数据变换.....	7
4.2 负荷特征分析.....	9
4.2.1 暂态特征.....	9
4.2.2 稳态特征.....	11
5.问题二	17
5.1 相似度与权系数.....	17
5.2 模型建立.....	18
5.3 模型求解.....	19
6.问题三	20
6.1 事件检测算法.....	20
6.2 模型建立.....	22
6.3 模型求解.....	24
7 问题四	25
7.1 负荷识别决策树算法.....	25
7.2 模型建立.....	27
7.3 模型求解.....	27
8. 结论	28
参考文献.....	28

1 研究背景与意义

与常规的连接在入户线上的电能表只能获得总能耗数据不同,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。分项计量技术目前主要分为两种:一种在总负荷内部为每个用电设备配备带有数字通信功能的传感器,通过通讯网络采集各用电设备的用电信息,这种方式称为侵入式电力负荷监测;另一种仅在电网的用户入口处安装一个传感器,通过采集和分析用户用电总功率或总电流来监测每个或每类用电设备的用电功率和工作状态,从而了解用户家中每个或每类用电设备的耗电情况和用电规律,这种方式称作非侵入式电力负荷监测与分解,这一技术量测得到的是整个线路上的电压、电流数据,它们可以看作是各个用电设备的电压、电流数据的叠加,它最重要的是如何从采集到的整条线路的电压、电流数据中“分解”出每个用电设备独立的用电数据,这种技术具有简单、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。

电力分项计量的一系列技术,是将电器识别作为物联网的一个重要研究方向,从物理层面去真正查看、操控物联网设备,达到绝对可信的要求,这是未来物联网的目标。同时,电力分项计量对于电力公司准确预测电力负荷、科学制定电网调度方案、提高电力系统稳定性和可靠性有着重要意义;对用户而言可以帮助用户了解用电设备的使用情况、提高用户的节能意识、促进科学合理用电。不仅如此,分项计量还可以针对能耗数据来实现设备老化、故障预警的功能,在我们的生产、生活中有非常实际的意义。

2 变量说明

符号	意义
$i(t)$	均值
i_{rms}	均方根
Δt	暂态过渡时间
K_p	冲击功率倍数
itc	电流跨度
$area$	轨迹面积
$absarea$	绝对面积
D	瞬时电阻标准差
S	域特征相似性
$P_{e,(i,j)}(t)$	设备 i 在时刻 t 处于状态 j 时的有功功率的估计值
$P_{r(i,j)}$	设备 i 处于状态 j 时的有功功率参考值 (均值)

3 问题归纳分析

根据题目要求及所提供数据，对所提的四个问题简要归纳如下：本题主要在已知设备 1 至设备 11 的各个采样点的设备数据（包括有功功率，无功功率，操作记录等数据），以及一个周期内的周波数据，谐波数据的基础上，对数据进行处理和分析，找出识别不同设备的负荷印记，分析已提取的负荷印记并用以识别各个单一负荷以及多负荷识别。

本文主要针对不同的用电设备从以下角度进行分析：

第一，设备是启/停二状态设备还是有限多状态设备。

此问题中，涉及设备状态的档位划分，即将档位用数学语言表示。本文根据其电流、有功功率、无功功率等数据的复合结果，按照从小到大的顺序将有限多状态和启/停状态设备的各个设备状态进行划分并标号。设备 YD1-YD11 的状态划分如下：

表 1 档位划分

	0 档位	1 档位	2 档位	3 档位	4 档位	5 档位
YD1	关闭	1 档	2 档	3 档		
YD2	关闭	低火	中低火	中火	中高火	高火
YD3	关闭	打开				
YD4	关闭	睡眠	重启	打开		
YD5	关闭	打开				
YD6	关闭	打开				
YD7	关闭	结束	扫描	打开	打印	复印
YD8	关闭	保温	制冷	加热	加热、制冷	
YD9	关闭	制冷	辅热	除湿		
YD10	关闭	1 档冷风	2 档冷风	1 档热风	2 档热风	
YD11	关闭	打开				

第二，设备是纯电阻设备还是非纯电阻用电设备。

纯电阻电器，电流只有热效应，其 $V-I$ 轨迹曲线是类似于直线的一个曲线（即曲线面积接近于 0）；而非纯电阻电器，电流除了产生热效应，还产生其它效应，将电能转化为其它能（化学能、机械能等），它的 $V-I$ 轨迹围成了类似椭圆的曲线。

在问题一中，本文除了提取各个设备的各个档位的 $V-I$ 轨迹特征，还提取了 4 种暂态特征和 4 种稳态特征的负荷印记特征，比较全面的含盖了用电设备的运行特征。

在问题二中，在问题一提取的负荷特征的基础上，用待识别的单一设备和设备库的数据进行识别匹配。即考虑的多种因素转化相应的评价标准，利用对应的负荷特征进行相似度匹配，负荷印记区分出不同设备，并量化其定义特征，从而识别出待识别设备。

第三，设备的工作有功功率和无功率的关系。

对于问题三，首先对设备的有功功率和无功率变化进行了分析，设定合理的功率改变阈值，采用事件检测算法对事件发生时间点检测。在此基础上根据有功、无功功率的均值与方差的负荷特征，建立了 0-1 二次规划的优化模型。该模型中，以各个设备的各个档位的关闭为 0，开启为 1，以优化其误差 ϵ 为目标，在事件检测算法的基础上，很好的识别出用电设备组的状态、操作与其对应的操作时间。

第四，设备工作时产生的谐波特征。

在问题四中，首先利用负荷识别决策树算法，识别出各个用电设备，其次利用上一章中

单个负荷的特征参数建立完成负荷识别所需的数据库,包括平均功率数据库和谐波幅值数据库,最后建立 0-1 二次规划模型,确定出用电设备组的转态、操作及操作时间。

4 问题一分析及求解

对于问题一的解答,主要采取以下几个步骤,如图 1 所示:



图 1 问题一流程图

4.1 数据预处理

4.1.1 降噪处理

由于各个设备的采样数据点较多,数据波动较大,数据中含有突变点,突变点在求平均值的过程中会影响到结果的准确性,因此需要对数据中的突变点和异常点进行清理。文采用小波分解 W_k 值的方法检测并区分异常点和突变点^[1]。在数学上,利用 Lipschitz 指数(用 ν 表示)表述函数的光滑程度:函数越光滑 ν 越大。由于阶跃函数的 ν 值为 0,因此斜坡函数较阶跃函数更光滑连续,其 ν 值为 1。Mallat 等人在 1992 年建立 Lipschitz 指数与小波系数的关系,并以此提出小波变换模极大值原理,其中小波变换模极大值与 Lipschitz 指数关系如下:

$\nu > 0$ 时,小波系数随小波尺度的增大而增大;

$\nu = 0$ 时,小波系数与尺度无关。

利用此关系可以检测并区分异常值和突变点,具体算法如下:

步骤 1. 在两个小波尺度下对拟合残差 e_t , $t=1,2,\dots$ 进行在线小波分解;

步骤 2. 计算两尺度下小波分解系数的模,并计算差值得到 E_k ;

步骤 3. 异常点、突变点检测:

- 1) 步骤 1 中未出现模极大值点处,并且步骤 2 中 E_k 没有突变,说明此处 W_k 值曲线始终维持在零附近,没有发生变化,说明此处既没有异常点也没有突变点;
- 2) 步骤 1 中出现模极大值,而步骤 2 中 E_k 没有模极大值点,说明此处两尺度下小波系数相同,应为异常点所在处;

- 3) 步骤1和步骤2中均存在模极大值点,说明两尺度下小波系数不同,应为突变点所在处。通过上述异常值检验方法检验了 YD1-YD11 的有功功率数据,图2是某设备的某段时间的异常点检测结果。

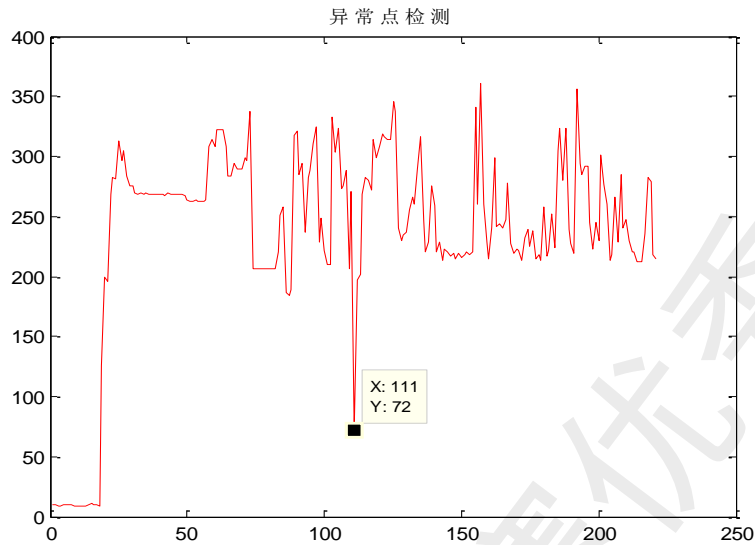


图2 异常点检测图

4.1.2 数据变换

我们通过小波降噪的原理来进行数据变换。在信号检测中经常存在噪声污染,为了更加精准的研究信号传输变化,往往会采用降噪的方法进行去除噪声。常见的降噪方法有:希尔伯特变换降噪法、小波变换降噪方法、傅里叶分析降噪处理法等。由于小波分析是一种对信号降噪最普遍方法,并且在信号低频部分降噪处理中具有良好的频率分辨率,因此,本文采用小波包降噪法去除信号中复高斯白噪声。

小波降噪是利用噪声在时域和频域内的差别,实现有效的信噪分离,从而获得较为理想的降噪效果。

设信号 $X(t)$ 被噪声污染后为 $S(n)$ 其基本噪声模型可以表示为:

$$S(t) = X(t) + \sigma e(t) \quad (1)$$

其中, $e(t)$ 为噪, σ 为噪声强度^[2]。

利用小波降噪对信号降噪的过程分为:(1)信号进行小波分解;(2)信号进行小波重构^[2]。由于一般噪声信号包含具有较高频率的细节中,因此利用门限阈值等形式对所分解的小波系数进行处理。于是小波降噪分为如下几个步骤:

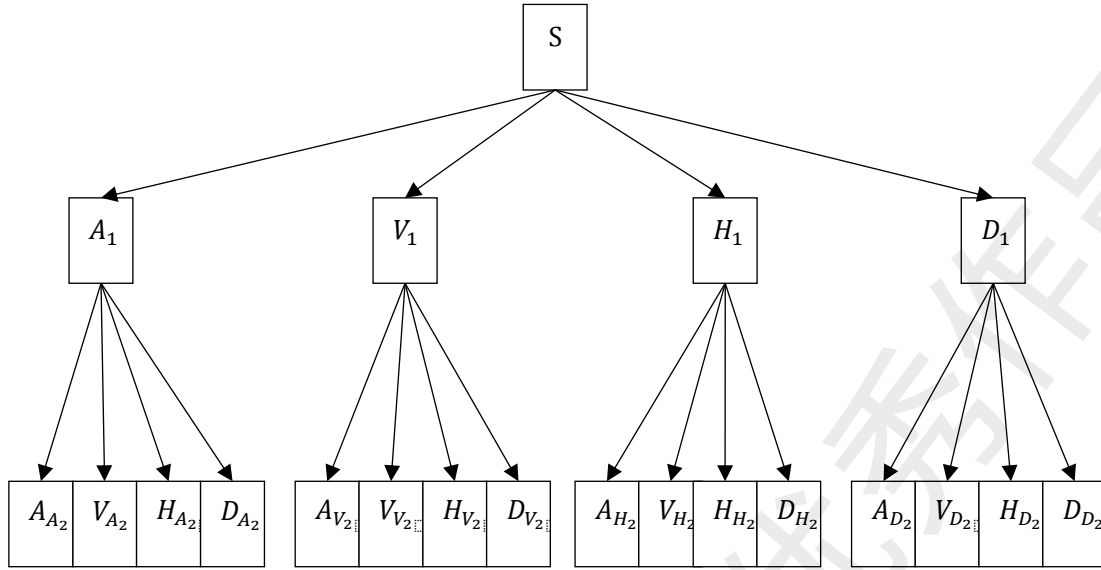


图3 小波包降噪的3层分解示意图

1、小波系数分解

根据多尺度分析方法，设信号 $S(t) \in V_0$ 可以将其投影到 V_1 空间和 W_1 空间，分解成 V_1 空间的概貌和 W_1 空间的细节部分。再将 V_1 空间的信号投影继续分解，可以分成 V_2 空间和 W_2 空间。如此继续，最终可以将信号 $S(t)$ 分解投影到 $W_1, W_2, \dots, W_j, \dots$ 空间中去，分解模型示意图如下所示：

其中信号的多尺度数学分解过程如下：

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}} \phi(2^{-j}t - k) \quad (2)$$

上面两式分别表示在不同尺度 j 下，随着 k 的位移，张成不同的尺度空间和小波空间 $\{W_j\}_{j \in \mathbb{Z}}$ ，则信号 $S(t)$ 可分解成：

$$S(t) = \sum_k c_{j,k} \phi_{j,k}(t) + \sum_k d_{j,k} \psi_{j,k}(t) \quad (3)$$

其中， $c_{j,k}$ 和 $d_{j,k}$ 分别为尺度空间的尺度系数和小波系数。

2、小波阈值处理

小波分析用于降噪的核心步骤就是对信号经过小波分解后的系数确定阈值。在小波变换中，对各层系数降噪所需的阈值一般根据原始信号的信噪比来选取，即通过波各层分解系数的标准差来求取。在得到信号噪声强度 σ 后，可以根据信号噪声强度确定各层的阈值，即：

$$T = \sqrt{2 \lg(n)} * \sigma \quad (4)$$

其中 n 为信号长度。

在求得阈值以后，有两种在信号作用阈值的方法：（1）令所有绝对值小于阈值的信号点的值为0，即硬阈值为 $d_1(t)$ ；（2）软阈值 $d_2(t)$ 计算公式分别为：

$$d_1(t) = \begin{cases} t, & |t| > T \\ 0, & |t| \leq T \end{cases} \quad (5)$$

$$d_2(t) = \begin{cases} t-T, t > T \\ t+T, t < -T \\ 0, \text{其他} \end{cases} \quad (6)$$

3、小波信号重构

信号重构过程与信号分解过程相似，只是方向相反，在对小波系数进行阈值处理后，再根据信号测试结果，将信号分解恢复降噪后的信号。

综上，经过小波降噪后，获取出降噪后的数据，然后通过 Matlab 软件绘制出降噪后的波形图。基于篇幅，此处以 YD_1 的周波数据的一个采样周期为例，给出其降噪后的信号。

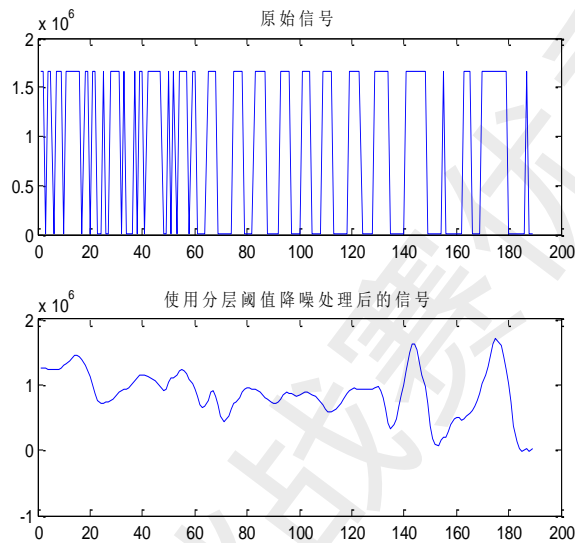


图 4 小波降噪图

4.2 负荷特征分析

负荷特征是指电器处于运行状态时某个物理量所表现出来的特性，是负荷识别的重要依据，本文通过对各个用电设备的数据分析，提取了两类负荷特征——暂态特征和稳态特征；这里的稳态特征分析认为负荷运行是稳定状态，而暂态分析考虑到过渡过程。通过对特征的分析，将各个负荷特征转化为可量化的负荷特征，这些特征可以有效地区分各个用电设备，从而达到区分各个设备的目的。

4.2.1 暂态特征

暂态特征是指电器工作状态发生转变时所表现出来的特征，研究发现主要负荷的暂态特征是比较明显的，它不像稳态特征那样具有重叠性，并且较之稳态特征能为负荷识别提供更多的信息用于负荷识别。暂态特征的不足在于为了获取暂态数据需要较高的采样频率，这就要求数据采集装置需要对系统进行持续不间断的监测，才不会遗漏负荷投入/切除运行时稍纵即逝的暂态过程。

暂态特征主要包括暂态过渡时间、电流均值及均方根、冲击电流功率倍数、瞬时电压、瞬时电流、瞬时功率等。其提取方法相对而言较为复杂，常用方法有频谱分析、小波分析等频域及小波域分析方法。负荷的暂态特征是比较明显的，并且负荷的暂态特征不像稳态特征

那样具有重叠性，然而唯一的局限性就是为了获取暂态数据需要较高的采样频率。下面就对电流均值及均方根、暂态过渡时间和冲击电流功率倍数 4 种非侵入式负荷监测中的暂态特征的实现方法与负荷特点进行分析：

均值与均方根

1、均值（直流分量），通常对应于某段时间电压或电流的均值，而在正式的定义中，计算信号均值需要在某段时间上对信号波形进行积分。信号 $i(t)$ 的时间均值表达式如下：

$$\langle i(t) \rangle = \frac{1}{T} \int_0^T x(t) dt \quad (7)$$

其中，T是积分时间。

2、均方根，表征信号相对均值的起伏和波动。使用信号 $i(t)$ 的均方根来表征交流波形的电压，该值定义为：

$$i_{rms} = \sqrt{\frac{1}{T} \int_0^T i(t) dt} \quad (8)$$

显然，对于不同的信号波形，均方值和均方根给出量化这些时变电信号的合理测量。均方根实际上是对信号平方的平均值求平方根；通过对信号的均值和均方根的正式定义，相应的电流的均值和均方根是类似的定义。

研究表明，电阻性负载开启后直接进入稳态，没有暂态过程，而非阻性电器启动时存在电磁暂态过程，一些电器的暂态过程伴随有 3-9 倍冲击电流(功率)的出现，具有非常强的辨识度。暂态开关特性可用暂态过渡时间和冲击倍数两个物理量来表示^[3]。

1、暂态过渡时间

设暂态过程开始时间为 t_{ion} ，暂态过程结束时间为 t_{toff} ，则暂态过渡时间 Δt 可通过下式计算：

$$\Delta t = t_{toff} - t_{ion} \quad (9)$$

2、冲击电流(功率)倍数

设电器投入前的稳态电流为 I_{s1} 暂态开关过程中的电流最大值为 I_{peak} ，电器投入后的稳态电流为 I_{s2} ，则冲击电流倍数 K_I 的计算公式为：

$$K_I = \frac{I_{peak} - I_{s1}}{I_{s2} - I_{s1}} \quad (10)$$

同理，冲击功率倍数 K_p 的计算公式：

$$K_p = \frac{P_{peak} - P_{s1}}{P_{s2} - P_{s1}} \quad (11)$$

其中， P_{peak} 为暂态开关过程中的功率最大值， P_{s1} 为电器投入前的稳态平均功率， P_{s2} 为电器投入后的稳态平均功率。根据上面的介绍以及附件 1 中提供的单一态数据，我们得到如下的数据表格：

表 2 设备暂态特征值

	电流均值	电流均方根值	暂态过度时间	冲击电流倍数
YD1	110.021	71.860	4	1.003
YD 2	3229.075	2635.514	3	3.376
YD 3	4362.044	3874.353	1	1.014
YD 4	213.690	74.269	4	1.612
YD 5	98.903	87.131	1	1.005
YD 6	25.685	16.077	2	1.030
YD 7	435.202	1079.552	6	7.127
YD 8	1242.107	863.184	2	1.005
YD 9	100.004	54.712	7	1.551
YD 10	2153.312	2129.346	3	1.037
YD 11	289.769	239.148	6	1.011

从上表可以看出，不同的设备由于结构与工艺的差别，用电负荷从开启到稳定状态的过渡阶段的变化形式各不相同；纯电阻型负荷直接从开启进入到稳态，而其它负荷在启动时会含有脉冲电流，启动的时间、脉冲大小各不相同，不同用电负荷的开关暂态各不相同，因此暂态特征可以用作区分用电设备。

4.2.2 稳态特征

稳态特征是指电器运行状态稳定时所表现出的特征，换句话说，稳态特征是两种稳定运行状态某些特性分析差异的结果。被用于负荷识别的稳态特征参数包括有功功率、无功功率、电流谐波、功率因数、V-I 轨迹等。稳态特征易于采集和检测，应用较多，本文研究将用到 V-I 轨迹、功率特征和谐波矩阵。

1、V-I 轨迹

香港大学的 Lam 等于 2007 年提出了二维的电压-电流 (V-I) 轨迹法。他们利用稳态时一个周期的电压电流数据来绘制 V-I 轨迹，进而根据轨迹的形状特征分辨出不同的负荷。V-I 轨迹法主要采纳的形状特征包括曲率、斜率、总面积、左右分区面积、不对称性、相交点、轨迹循环方向以及中间分区的峰值等。为避免不同负荷的电压、电流幅度差异对 V-I 轨迹大小的影响，需要在比较形状特征前对两个参数做归一化处理。利用附件 1 提供的周波数据，以归一化后的电压为横坐标，归一化后的电流为纵坐标绘制部分设备的 V-I 曲线，如图所示：

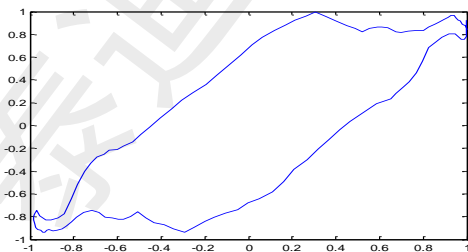


图 5 YD1 (奥克斯风扇)

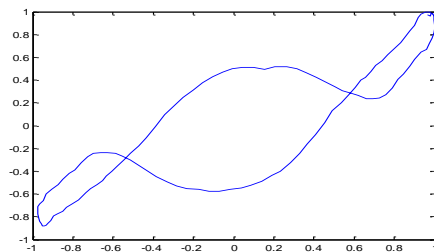


图 6 YD2 (美的微波炉)

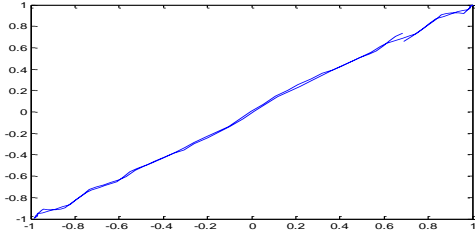


图 7 YD3 (九阳热水壶)

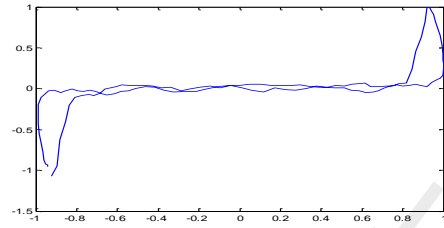


图 8 YD6 (节能灯)

从上图可以看出，对于电阻型负荷，如电热水壶，V-I 是一条直线，而对于微波炉这类含有谐波含量较高的负荷，V-I 至少含有一个交点，且它们轨迹相差十分明显，因此结合 V-I 轨迹作为负荷识别分解的特征，能够提高负荷分解精度。本文利用电流跨度、轨迹面积、绝对面积、瞬时电阻标准差四种特征来与预设负荷特征的负荷稳态特征匹配。以下是其主要步骤^[4]：

1) 获取电流跨度

记所述电流跨度为 itc ，定义为：

$$itc = \max(I) - \min(I) \quad (12)$$

其中， I 为电流序列，即 V-I 轨迹曲线的纵坐标； $\max(I)$ 代表所述电流序列的最大值， $\min(I)$ 代表所述电流序列的最小值。

2) 根据所述 V-I 轨迹曲线生成标么化的 V-I 轨迹曲线

获取电压序列 V_m 标么化值 V'_m ，定义为：

$$V'_m = \frac{V_m}{\max(V)} \quad (13)$$

其中， $\max(V)$ 为电压序列的最大值及电压序列 V ， $m \in [1, NT + ip]$ ， NT 为一个周期的采样点个数， ip 为预设插值点数；

获取电流序列 I_m 标么化值 I'_m ，定义为：

$$I'_m = \frac{I_m}{\max(I)} \quad (14)$$

其中， $\max(I)$ 为电流序列的最大值及电流序列 I ；记电压最大值点为 $v \max(V'_{v \max}, I'_{v \max})$ ，记电压最小值点为 $v \min(V'_{v \min}, I'_{v \min})$ ；根据所述电压最大值点 $v \max$ 及所述电压最小值点 $v \min$ 将 V-I 轨迹曲线分为 A, B 两个部分：

$$A = \{(V'_{K1}, I'_{K1}) | K1 \in [1, v \min]\} \quad (15)$$

$$B = \{(V'_{K2}, I'_{K2}) | K2 \in [v \min + 1, NT + ip]\} \quad (16)$$

3) 获取所述标么化的 V-I 轨迹曲线的轨迹面积

记所述轨迹面积为 $area$ ，定义为：

$$area = \sum_{m=1}^{NT+ip-1} \frac{1}{2} (V'_{m+1} - V'_m)(I'_{m+1} - I'_m) \quad (17)$$

4) 获取所述标么化的 V-I 轨迹曲线的绝对面积

记所述绝对面积为 $absarea$ 定义为：

$$absarea = \sum_{m=1}^{vmin-1} \frac{1}{2} |V'_{m+1} - V'_m| (|I'_{m1} - I'_m| + |I'_{m2} - I'_{m+1}|) \quad (18)$$

其中， $m1$ 满足 $\min(V'_{m1} - V'_m)$ ， $m1 \in [vmin+1, NT+ip]$ ， $m2$ 满足 $\min(V'_{m2} - V'_m)$ ， $m2 \in [vmin+1, NT+ip]$

5) 获取瞬时电阻标准差^[5]

记所述瞬时电阻标准差为 D ，定义为：

$$D = \sqrt{\frac{\sum_{n=1}^{NT} (R_n - \bar{R})^2}{NT}} \quad (19)$$

其中， R_n 为第 n 采样点的瞬时电阻， $R_n = \frac{V'_n}{I'_n}$ ， V'_n 为第 n 采样点表么化后的电压值， I'_n 为第 n 采样点表么化后的电流值， $m \in [1, NT+ip]$ ， NT 为一个周期采样点个数， ip 为预设插值点数； \bar{R} 为 R_n 的平均值。

首先根据功率的大小，将设备的工作状态分为若干档位，功率越大，档位越高。设备 1 至设备 11 最多出现了五种工作状态，故将设备工作状态划分为一、二、三、四、五档，从各个运行状态中随机选取 1 秒绘制 V-I 轨迹。通过以上的步骤以及附件 1 提供的单一态数据，得到如下的数据（缺省行代表设备无该档位）：

表 3 设备 1 档位的 V-I 轨迹特征

V-I 轨迹特征	电流跨度	面积	绝对面积	瞬时电阻标准差
YD1	3.253	0.0723	0.085	4.410
YD2	77.582	0.012	0.072	59.810
YD 3	95.075	0.077	0.077	2.120
YD 4	11.086	0.014	0.022	30.925
YD 5	4.100	0.072	0.079	3.874
YD 6	2.336	0.017	0.060	170.892
YD 7	16.185	0.016	0.024	38.097
YD 8	31.091	0.069	0.084	0.191
YD 9	0.881	0.062	0.110	6.843
YD 10	36.897	0.027	0.051	12.263
YD 11	11.780	0.059	0.062	9.119

表 4 设备 2 档位的 V-I 轨迹特征

V-I 轨迹特征	电流跨度	面积	绝对面积	瞬时电阻标准差
YD 1	2.9098	0.0711	0.091	11.723
YD 2	4.4406	0.060	0.067	59.402
YD 4	13.551	0.012	0.018	141.87
YD 7	6.026	0.017	0.032	7.239
YD 8	31.814	0.069	0.084	0.215
YD 9	3.5010	0.030	0.082	42.053
YD 10	73.01	0.076	0.077	0.191

表 5 设备 3 档位的 V-I 轨迹特征

V-I 轨迹特征	电流跨度	面积	绝对面积	瞬时电阻标准差
----------	------	----	------	---------

YD 1	3.725	0.070	0.078	5.411
YD 2	4.461	0.062	0.068	5.100
YD 4	4.936	0.013	0.026	12.027
YD 7	1.565	0.020	0.092	2.348
YD 8	31.562	0.070	0.085	0.400
YD 9	3.914	0.039	0.078	7.411
YD 10	1.023	0.013	0.079	2.355

表 6 设备 4 档位的 V-I 轨迹特征

V-I 轨迹特征	电流跨度	面积	绝对面积	瞬时电阻标准差
YD 2	184.994	0.011	0.038	475.7641
YD 7	51.627	0.043	0.044	6.578
YD 8	16.397	0.021	0.031	815.209
YD 10	0.977	0.009	0.087	2.386

表 7 设备 5 档位的 V-I 轨迹特征

V-I 轨迹特征	电流跨度	面积	绝对面积	瞬时电阻标准差
YD 2	72.398	0.017	0.073	68.673
YD 7	57.096	0.066	0.068	0.231

2、功率特征

有功功率是负荷运行期间消耗的总功率,如果负荷为纯电阻则电流和电压波形将总是同相位,则不存在无功分量。对于纯电抗负荷相移 90 度,则将不存在有功功率。但是,由于存在电感元件和电容元件,因此总存在电流和电压波形之间的相移,这就会产生或消耗无功功率。有功功率及无功功率的计算如下所示^[6]:

$$P = \sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} U_k I_k \cos(\phi_k)$$

$$Q = \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} U_k I_k \sin(\phi_k)$$

其中, U 为用电负荷的运行时的电压有效值, I 为用电负荷的运行时的电流的有效值, ϕ 为用电负荷的运行时的功率因数角, k 为谐波次数。

我们将各个设备的有功功率和无功率和时间的图像绘制在同一坐标轴上,得到各个设备的有功功率与无功功率的对比图,选取 5 个设备的对比图:

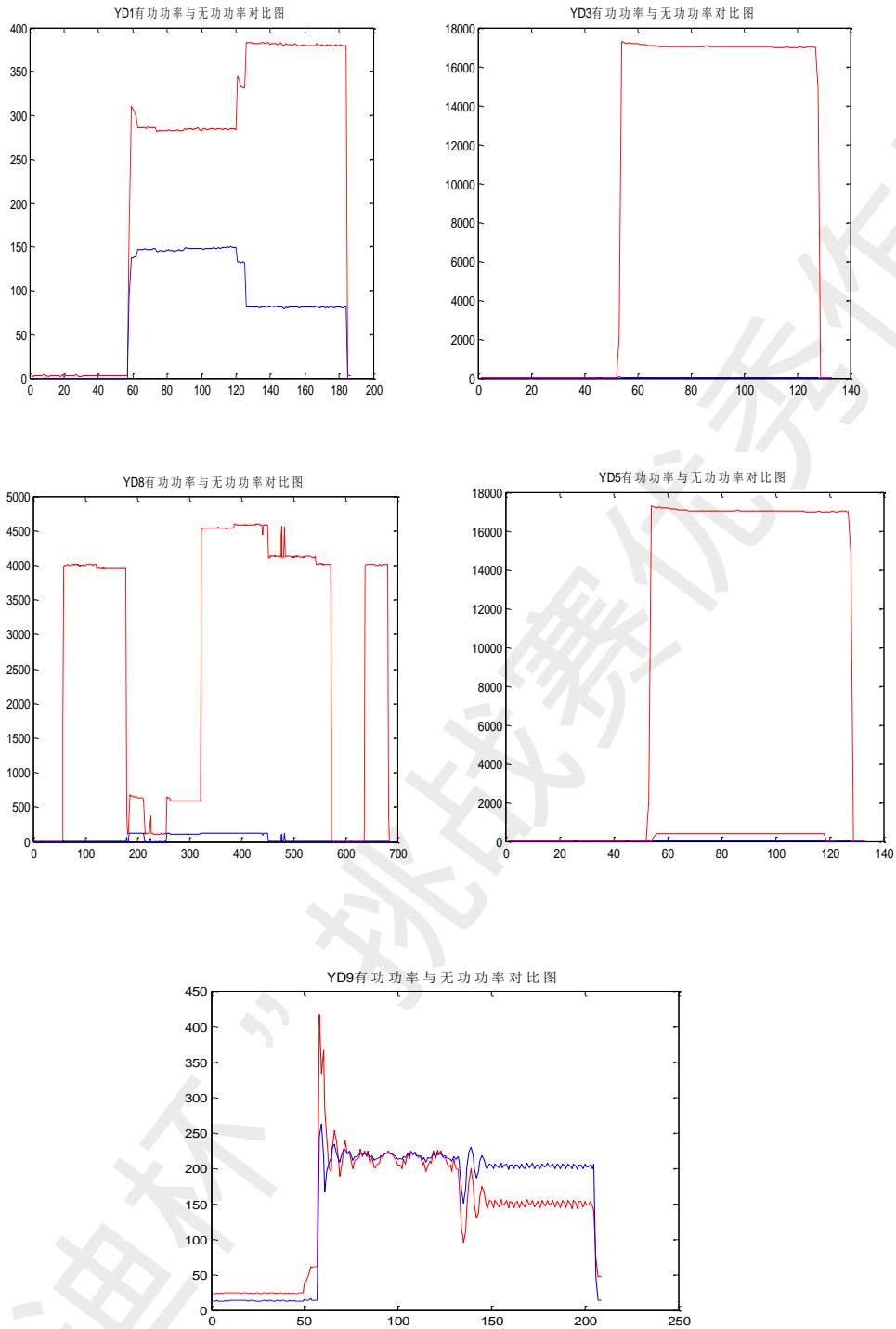


图9 有功功率与无功功率对比图

从图中可以看出，YD1、YD3、YD5 和 YD8 这四个设备中，有功功率大于无功功率，而 YD9 设备的有功功率不是恒大于无功功率的，其中有一段采样期间有功功率小于无功功率，因此 YD9 设备在有功功率和无功功率对比中明显不同于其它设备。

3、谐波矩阵

由于不同的用电器的电路设计不同，产生的谐波也是不同的。因此谐波数据中包含有不同用电器独特的特征。通过傅里叶变换或小波变换可提取负荷电压或电流的谐波，进而实现对负荷的辨识。需注意的是，大多数负荷为近似双向对称元件，因而产生的偶次谐波幅度较

小，而奇次谐波的幅度较大。低次谐波含有的信息量大，因此本文选取了 2 次至 11 次谐波数据研究，计算各个设备的各次谐波含有率幅值，得到如下表格，每一行的数据是每个设备的 k 次谐波含有率的幅值 ($k = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$)。

表 8 设备 k 次谐波含有率幅值

幅值	2 次	3 次	4 次	5 次	6 次	7 次	8 次	9 次	10 次	11 次	二范数
YD1	99.68	99.71	96.9	98.28	94.51	89.62	98.51	98.19	95.49	96.21	305.958
YD2	99.2	99.44	95.8	96.9	99.32	98.6	98.04	97.98	97.89	99.33	310.714
YD3	99.9	94.5	96.48	99.01	93.5	97.43	97.54	93.74	97.31	96.84	305.622
YD4	21.06	96.19	19.6	88.75	18.47	82.88	17.43	76.13	15.01	62.38	188.104
YD5	97.15	99.85	97.84	99.45	95.64	99.68	99.54	96.9	93.93	94.94	308.364
YD6	97.83	92.98	97.89	94.48	96.08	96.67	84.73	94.65	98.6	97.28	301.039
YD7	79.93	94.24	36.77	83.92	19.74	78.01	19.49	73.88	13.26	58.86	199.043
YD8	99.51	98.09	98.93	92.82	98.65	97.63	92.07	98.63	99.54	98.93	308.366
YD9	17.75	53.85	34.85	62.77	25.88	31.83	32.73	37.85	42.27	46.98	128.647
YD10	66.59	38.27	38.37	21.88	14.4	20.07	8.67	22.51	10.96	13.5	96.665
YD11	6.14	31.41	6.05	10.67	4.64	15.31	4.53	17.42	4.01	8.97	43.025

不同负荷产生的谐波不同，通过负荷的电压或电流的谐波，进而实现对负荷的辨识。阻性负荷，如白炽灯、热水壶，产生的谐波较少；非阻性负荷，如电磁炉，电风扇产生的谐波丰富。题目中给出了各次谐波含有率数据，可以看到 YD1, YD2, YD3, YD5, YD6, YD8 的二、三次谐波含有率达到了 90% 以上，但 YD9, YD10, YD11 的谐波含量明显低于 90%，可将这些负荷区分开来。

为描述各个设备在不同工作状态下的谐波含有率变化趋势，本文计算了各设备不同工作状态下的电流谐波含有率方差。其中缺省值表示该设备不存在该档位，比如设备 1 不能切换成 4 档和 5 档，表格如下：

表 9 设备个工作状态下谐波含有率方差

方差	关闭	档位 1	档位 2	档位 3	档位 4	档位 5
YD1	763.2436	1.21656	0.97821	0.24295		
YD2	702.8478	768.2751	0.582456	0.58845	90.05154	42.24794
YD3	777.4941	703.4177				
YD4	17.23165	555.0503	480.2903	474.9479		
YD5	582.9355	0.088918				
YD6	583.3648	220.6958				
YD7	8.813617	480.7053	473.4962	75.56688	32.17358	1.453375
YD8	166.9674	290.4106	0.128187	175.1147	274.4583	
YD9	36.81396	38.92979	51.2206	38.28987		
YD10	16.59171	107.8209	0.063964	12.29556	8.93058	
YD11	9.557222	5.628287				

从上面表格可以看出，在关闭状态下，YD1、YD2、YD3、YD5、YD6 的谐波含有率方差较其它设备更大，即谐波含有率的变化更大，而对于一个设备来说，例如 YD4，首先在关闭状态下谐波含有率方差较小，然后切换成 1 档时谐波含有率迅速增加，最后又随着档位切换的越高，谐波含有率方差慢慢减少，谐波含有率几乎保持恒定，因此谐波含有率方差可

作为识别依据。

5 问题二分析及求解

5.1 相似度与权系数

要自动识别某一未知的单一设备，可通过分析负荷印记的特征相似性。域特征相似性 S 定义为：

$$S = \frac{1}{\left(\frac{Y_i - Y_x}{Y_i}\right)^2} \quad (20)$$

其中， Y_x 表示未知设备的特征向量； Y_i 表示设备 i 的特征向量。 S 的值越大，表示未知设备 X 与已知设备 I 的相似程度越高。

将本文提取的负荷印记相似度分为四类计算，其中 $s_{ix_i} = \frac{1}{\left(\frac{Z_i - Z_x}{Z_i}\right)^2}$ 表示设备 YD_i 和待测设备 YD_x 的暂态特征相似度。

同理， $s_{iv_i} = \frac{1}{\left(\frac{V_i - V_x}{V_i}\right)^2}$ 和 $s_{ix_x} = \frac{1}{\left(\frac{X_i - X_x}{X_i}\right)^2}$ 表示设备 YD_i 和待测设备 YD_x 的 V-I 轨迹特征和谐波特

征的相似度， H_{ix} 表示有功功率无功功率对比相似度。 H_{ix} 定义为两个设备的有功功率与无功功率对比图的图像相似度，主要用于对于两幅图像之间内容的相似程度进行打分，根据分数的高低来判断图像内容的相近程度。最后采用加权的方式计算总的相似度，权重的确定采用熵值法。在信息论中，熵是对不确定性的一种度量。信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也越大。根据熵的特性，我们可以通过计算熵值来判断一个事件的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，指标的离散程度越大，该指标对综合评价的影响越大。熵表示从一组不确定事物中提供信息量的多少。在多指标决策问题中，某项指标的变异程度越大，信息熵越小，该指标提供的信息量就越大，那么在方案评价中所取得的作用就越大，该指标的权重也就越大；反之，某指标的变异程度越小，信息熵越大，该指标所提供的信息量越小，那么该指标的权重也就越小。根据各指标值的变异程度，利用信息熵计算各指标的权重。

熵值法确定所得相似度的权系数步骤如下：

1、数据的非负数据化处理：

由于熵值法计算采用的是各个方案某一指标占同一指标值总和的比值，因此不存在量纲的影响，不需要进行标准化处理，若数据中有负数，就需要对数据进行非负化处理。此外，为了避免求熵值时对数的无意义，需要进行数据平移：

$$\text{对于越大越好的指标: } r_{ij} = \frac{x_{ij} - x_i^{\min}}{x_i^{\max} - x_i^{\min}} + 1 \quad (21)$$

$$\text{对于越小越好的指标: } r_{ij} = \frac{x_i^{\max} - x_{ij}}{x_i^{\max} - x_i^{\min}} + 1 \quad (22)$$

为了方便起见，仍记非负化处理后的数据为 r_{ij}

2、由 $R=(r_{ij})_{m \times n}$ 计算第 i 项指标下第 j 个方案占该指标的比重 f_{ij}

$$f_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}}, i=1,2,\dots,m; j=1,2,\dots,n \quad (23)$$

3、第 i 个评价指标 f_i 输出的熵

$$H_i = -K \sum_{j=1}^n f_{ij} \ln f_{ij}, j=1,2,\dots,n \quad (24)$$

4、各目标的熵权系数

$$w_i = \frac{1-H_i}{m - \sum_{i=1}^m H_i} \quad i=1 \quad (25)$$

通过上述熵值法，得到各类特征相似度的权重， $w=(0.151,0.342,0.375,0.132)$ 。

5.2 模型建立

基于相似度的负荷识别模型是将各类特征相似度加权求和得到了负荷特征相似度的一个总相似度，具体的模型如下：

$$S_{ix_i} = \frac{w_1}{\left(\frac{Z_i - Z_x}{Z_i}\right)^2} + \frac{w_2}{\left(\frac{V_i - V_x}{V_i}\right)^2} + \frac{w_3}{\left(\frac{X_i - X_x}{X_i}\right)^2} + w_4 H_{ix}^2 \quad (26)$$

其中， Z_i ——表示设备 YDi 的暂态特征向量

Z_x ——表示待测设备 YDx 的暂态特征向量

V_i ——表示设备 YDi 的 V-I 轨迹特征向量

V_x ——表示待测设备 YDx 的 V-I 轨迹特征向量

X_i ——表示设备 YDi 的谐波特征向量

X_x ——表示待测设备 YDx 的谐波特征向量

H_{ix} ——表示待测设备与已知设备的有功功率无功功率对比相似度

通过上述 V-I 轨迹和谐波矩阵中介绍的方法，得到未知设备 X_1 和 X_2 的特征匹配数据如下：

表 10 待测设备 1 档位的 V-I 轨迹特征

	电流跨度	面积	绝对面积	瞬时电阻标准差
X1	1.642	0.022	0.054	64.249
X2	1.277	0.049	0.079	13.390

表 11 待测设备 2 档位的 V-I 轨迹特征

	电流跨度	面积	绝对面积	瞬时电阻标准差
X1	31.4444	0.066	0.083	0.168
X2	5.108	0.061	0.073	16.166

表 12 待测设备 3 档位的 V-I 轨迹特征

	电流跨度	面积	绝对面积	瞬时电阻标准差
X1				
X2	3.314	0.032630025	0.08610039	12.973325

表 13 设备 k 次谐波含有率幅值

幅值	2	3	4	5	6	7	8	9	10	11	二范数
X1	0.98	0.99	0.97	0.98	0.88	0.95	0.98	0.88	0.99	0.94	3.01947
X2	15.57	54.12	21.21	49.32	15.3	34.93	17.81	29.42	15.72	29.94	99.16741

表 14 设备个工作状态下谐波含有率方差

方差	未工作	1	2	3	4	5
X1	0.047832	0.02007	7.51E-06			
X2	0.047832	0.080229	0.046945	0.037575		

5.3 模型求解

通过建立的模型与相关的数据，得到未知设备 X_1 和 X_2 与设备 1 至设备 11 的相似度的计算结果如下：

表 15 待测设备 1 的相似度结果

相似度	暂态特征相似度	V-I 轨迹特征相似度	有功功率无功功率对比相似度	谐波相似度	总相似度
设备 1—YD1	0.321	0.302	0.945	1.000708	0.63868
设备 1—YD2	1.228	1.134	0.675	1.000417	0.98893
设备 1—YD3	1.147	0.920	0.961	1.000407	0.99420
设备 1—YD4	0.434	0.973	0.782	1.000839	0.81413
设备 1—YD5	0.321	0.305	0.956	1.0011	0.64316
设备 1—YD6	0.148	1.265	0.943	1.000979	0.89269
设备 1—YD7	1.320	1.176	0.910	1.001123	1.09019
设备 1—YD8	3.001	0.664	0.898	1.002281	1.26949
设备 1—YD9	0.297	0.346	0.928	1.010204	0.64403
设备 1—YD10	1.361	0.786	0.926	1.006951	0.98783
设备 1—YD11	0.604	0.515	0.955	1.034707	0.76915

表 16 待测设备 2 的相似度结果

相似度	暂态特征相似度	V-I 轨迹特征相似度	有功功率无功功率对比相似度	谐波相似度	总相似度
设备 2—YD1	2.488	0.772	0.9585	1.01977	1.22088
设备 2—YD2	1.013	1.047	0.7076	1.011319	0.931428
设备 2—YD3	1.009	1.003	0.9296	1.010978	0.983971
设备 2—YD4	1.386	1.278	0.8192	1.021115	1.110883
设备 2—YD5	2.021	0.753	0.9456	1.031931	1.120526
设备 2—YD6	0.609	1.041	0.9594	1.027882	0.927835
设备 2—YD7	1.038	1.197	0.9267	1.029494	1.050958

设备 2—YD8	1.039	0.976	0.9022	1.073814	0.986183
设备 2—YD9	4.198	1.025	0.9735	1.510558	1.741483
设备 2—YD10	1.018	1.044	0.9365	1.156208	1.029209
设备 2—YD11	1.187	1.146	0.9492	0.881266	1.042491

从表中可以看出：未知设备 X_1 与设备 8 的相似度最高，即未知设备 X_1 为设备 8；未知设备 X_2 与设备 9 的相似度最高，即未知设备 X_2 为设备 9。

6 问题三分析及求解

负荷分解是指对电力负荷中不同类型用电设备功率消耗比例的实时辨识。由于非侵入式用电负荷分解的数据含有大量的冗余与噪音特征，并且电力入口处采集到的数据可能会受到其他因素的影响，如电压异常和测量误差、存在多种具有相似能耗特征负荷、负荷同时切换开关状态以及某些不常使用的负荷的开启等等，对于实时负荷分解来说就变得比较困难。因此，为了解决这些问题，本文采用基于事件检测的非侵入式用电负荷分解方法^[7]。

6.1 事件检测算法

负荷特征提取和事件检测是相辅相成的，本文以有功功率 P 的改变值 ΔP 作为事件检测的判据，方便直观。根据赛题中表 1 提供的用电设备及工作参数，设置合理的功率改变值的阈值。除了照明类电器有功功率较小，其他电器大多都在 50W 以上，因此功率改变阈值设定为 50W。但是，附件三中的设备组 3-5 中的设备组合各不相同，若统一将功率改变阈值设定为 50W 会降低事件检测的准确度，因此这里根据不同的设备组合对附件三中的无操作记录的组采用不同的功率改变阈值。设定阈值的方法是，由于小功率用电器的功率改变值较小，如果设定的阈值太大就会容易忽略掉小功率用电设备的状态切换，造成事件检测的遗漏，所阈值应该以小功率设备的功率改变值为基准设定。但设备组中同时含有最小功率的设备（YD6）和较大功率设备，此时阈值设定太大会造成大功率设备掩盖小功率设备，阈值设定太小会造成检测到的事件数倍增，因此阈值设定必须既考虑设备组所含设备的功率大小。本文采用如下的方法确定功率阈值：以时间为横轴，功率值为纵轴画出时间—功率图，通过观察图中曲线，同时统计用电设备功率值所占百分比，确定功率阈值。

表 17 功率阈值设定表

功率阈值	P_1	P_2
设备组 4	3000	4000
设备组 5	5000	16000
设备组 6	3000	5000

通过上述方法确定好各个设备组的事件检测算法功率阈值后，进行事件检测，算法的流程如下：

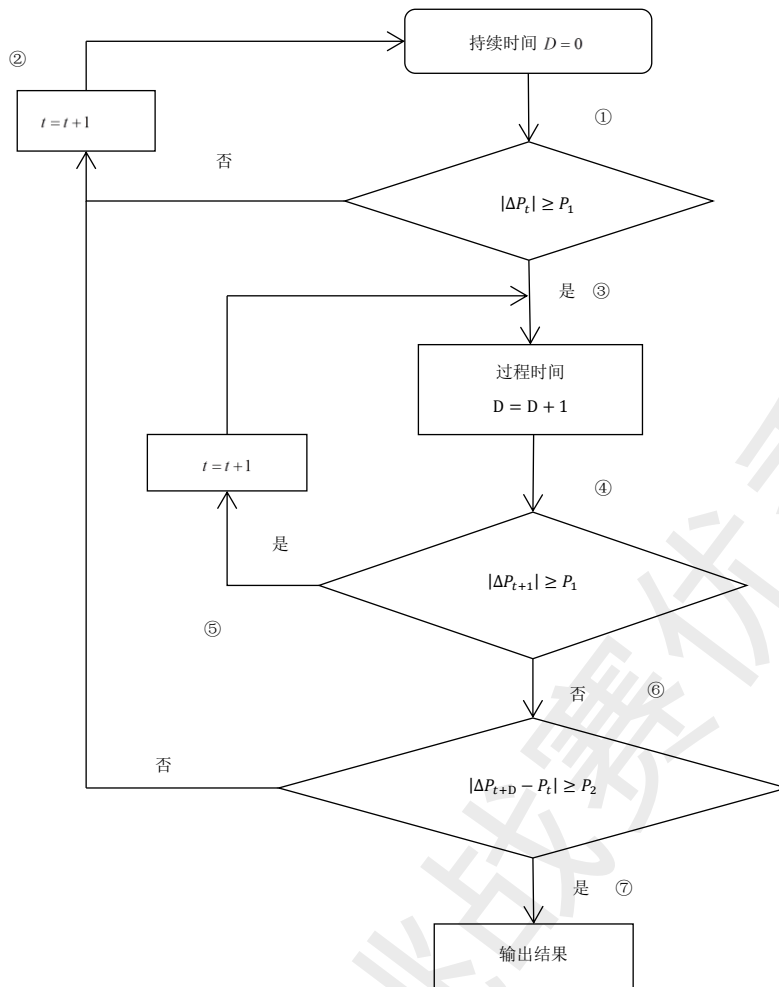


图 10 检测算法流程图

具体的检测算法流程如图 10 所示^[3]：

图中 ΔP_t 代表 t 时刻有功功率与 $t-1$ 时刻有功功率之差的绝对值， D 表示事件过程持续时间，其初始值是 0 秒。

① 计算当前时刻有功功率与前一时刻之差 ΔP_t ，并判断其绝对值是否大于或等于阈值 P_1 ，若不是则转入步骤②，否则进入步骤③。

② 读入下一时刻的数据，继续执行步骤①。

③ 事件持续时间 D 在其原有基础上增加 1 秒，继续执行步骤④。

④ 读入下一时刻的数据，计算得到 $\Delta P_{t+1} = P_{t+1} - P_t$ ，并计算其绝对值是否大于或等于 P_1 ，若是则转入步骤⑤，否则进入步骤⑥。

⑤ 读入下一时刻的数据，继续执行步骤③。

⑥ 根据事件持续时间 D 可得到事件结束时刻为 $t+D$ ，计算事件发生前后有功功率的变化值 $P_{t+D} - P_t$ ，如果该结果绝对值大于或等于阈值 P_2 ，则转入步骤⑦，否则认为不是一次事件，转入步骤②继续执行。

⑦ 输出结果：根据 $P_{t+D} - P_t$ 的正负情况，可判定这次事件是功率上升沿事件或者功率下降沿事件。如果结果为正，说明系统有功功率增加，判定为上升沿事件，一般是由电器投入运行或者状态发生改变导致；反之，如果结果为负，说明系统有功功率减少，判定为下降沿事件，一般是由电器退出运行护着状态发生改变导致。 t 时刻认为是事件发生的起始时刻， $t+D$ 时刻认为是事件的结束时刻。

以设备组 5 为例，利用事件检测算法找到电器运行状态发生改变的时刻，如图 11 所示：

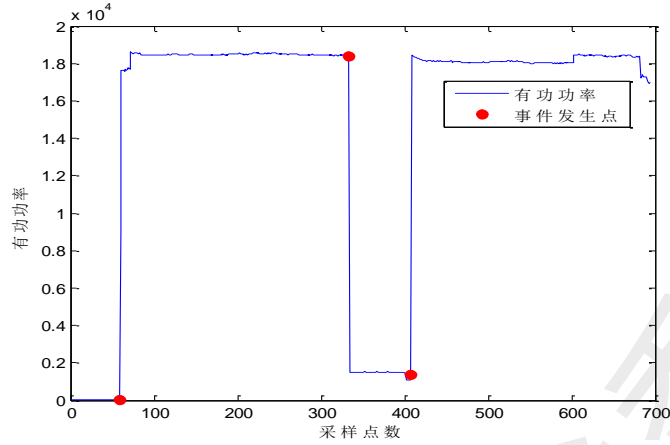


图 11 设备组五的事件检测图

通过设置合理的功率改变阈值，事件检测算法能很好地识别有功功率改变值较大的负荷事件，对分析各个用电设备的运行状态有很大的帮助。因此本文采用该事件检测算法对设备组的运行状态进行分段后，采用 0-1 规划模型进行负荷分解。

6.2 模型建立

通过问题一提取负荷特征计算所有设备的负荷特性矩阵，其负荷特性矩阵如下式所示：

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1N} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{M1} & \psi_{M1} & \cdots & \psi_{MN} \end{bmatrix} \quad (27)$$

N 为设备数目，对于多状态工作模式电器将每一状态工作模式看做一个电器处理，即 N 将大于电器设备的实际数目。 M 为识别算法中采用的负荷特性种类数，这里我们取 $M=2$ ，即有功功率与无功功率两种负荷特性。

$$\psi_{j,i} = [f_1, f_2, \cdots, f_n]^T \quad (28)$$

上式中的 $\psi_{j,i}$ 为设备 i 的负荷特性 j 的存储列向量， n 为对应的负荷特性 j 的数据长度。

实测数据提取出的待识别的特征向量 Y ：

$$Y' = [y'_1, y'_2, \cdots, y'_M]^T \quad (29)$$

$$y'_j = [f_1, f_2, \cdots, f_n]^T \quad (30)$$

y'_j 为实测数据提取出的负荷特性 j 的存储列向量。

所求状态向量：

$$\vec{X} = [x_1, x_2, x_3, \cdots, x_N]^T \quad (31)$$

\vec{X} 为负荷的状态向量（0 表示关断状态，1 表示开状态）

通过上述负荷监测算法可以确定一设备状态发生的改变的情况下，即状态向量 \bar{X} 已知，则从负荷特性数据库中可得知此设备的负荷特性向量 Y ：

$$Y = [y_1, y_2, \dots, y_M]^T = \psi \bar{X} \quad (32)$$

则 Y' 与 Y 的关系为：

$$Y' = Y + \varepsilon = \psi \bar{X} + \varepsilon \quad (33)$$

其中 ε 为各数值点之间的误差，其表达式如下式所示。

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M]^T$$

$$\varepsilon_j = [\varepsilon_{j,1}, \varepsilon_{j,2}, \dots, \varepsilon_{j,m}]^T$$

因此，在暂态事件检测算法检测到有事件发生后，相应的负荷特性提取算法提取出待识别的特征向量 Y' ，依据已建好的负荷特性数据库来求解状态向量 \bar{X} ，使得误差 ε 最小。问题描述如下：

目标函数用矩阵表示：

$$Y' = \psi \bar{X} + \varepsilon \quad (34)$$

约束条件：

$$\sum_{i=1}^N x_i = 1 \quad x_i = \{0, 1\} \quad (35)$$

其中， Y' 为实测数据提取出的待识别的特征向量， \bar{X} 为负荷的状态向量； ψ 负荷特性值矩阵。

由于 Y' 是冗余的测量，因此不可能直接根据上述方程组进行求解（如果不考虑误差，方程式 (34) 无解，因为方程的个数多于未知数的个数），但可以找到与方程组 (34) 最接近的解。可以使得上述 $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M]^T$ 方差最小来确定参数。将冗余的方程组转化下面的求极小值问题。

$$\min J = \varepsilon^T \varepsilon \quad (36)$$

$$\min J = (Y' - \psi \bar{X})^T (Y' - \psi \bar{X}) \quad (37)$$

$$\min J = Y'^T Y' - 2Y'^T \psi \bar{X} + \frac{1}{2} (\bar{X}^T \psi^T 2\psi \bar{X}) \quad (38)$$

因此上述问题转化为 0-1 二次规划问题，其数学模型如式 (39) 所示。

$$\begin{cases} \min J = Y'^T Y' - 2Y'^T \psi \bar{X} + \frac{1}{2} (\bar{X}^T \psi^T 2\psi \bar{X}) \\ \sum_{i=1}^N x_i = 1 \\ x_i = \{0, 1\} \end{cases} \quad (39)$$

由线性代数相关知识我们可以证明矩阵 $\psi^T 2\psi$ 为正定（或半正定）矩阵，则目标函数为严格凸函数（或凸函数）且可行域为凸集，因此上述规划问题属于凸规划问题。由非线性

规划问题中的凸规划可以知道上述模型的最优解即为全局最优解。

由于约束条件是0-1规划问题所以求解上述规划问题只能用离散方法。离散算法是从设计变量的离散特性出发直接求解整数规划。传统的离散方法大多属于组合算法，如隐枚举法、穷举法等。这类算法虽然能准确地找到问题的全局最优解，但随着问题规模增大，其计算量剧增。另一种是离散的启发式算法，如遗传算法。这类算法的主要缺点是不能很好地处理约束，且容易出现过早收敛问题。而连续化方法则不存在这样的问题，因此将上述问题转化为连续化方法进行求解。

式(39)中的约束条件 $x_i \in \{0,1\}$ 为离散型，将其进行连续化处理，找到其连续性约束的等价条件代换，本文采取的连续性等价代换的约束条件如式(40)所示。

$$\sum_{i=1}^N a_i (x_i - x_i^2) = 0 \quad (40)$$

连续化方法求解等价模型如式(41)所示。

$$\begin{cases} \min J = Y^T Y' - 2Y^T \psi \bar{X} + \frac{1}{2} (\bar{X}^T \psi^T 2\psi \bar{X}) \\ \sum_{i=1}^N x_i = 1 \\ \sum_{i=1}^N a_i (x_i - x_i^2) = 0 \end{cases} \quad (41)$$

6.3 模型求解

通过lingo软件编程求解上述0-1规划模型，求解结果详见附件3。求得各个时间点的设备组各个设备后，用如下公式估计每个设备的实时功率：

$$P_{e,(i,j)}(t) = \frac{P_{r(i,j)} P(t)}{\sum_{i=I_1}^{I_N} \sum_{j=1}^{n(i)} a_{i,j}(t) \cdot P_{r(i,j)}} \quad (42)$$

其中： $P(t)$ —— t 时刻设备组的有功功率的值；

$a_{i,j}$ ——0-1变量，取1表示设备*i*处于第*j*个状态；

$n(i)$ ——设备*i*可切换的状态总数；

$I = \{I_1, I_2, \dots, I_N\}$ ——表示设备组的设备ID集合；

N ——表示设备组所含设备个数；

$P_{e,(i,j)}(t)$ ——表示用电设备*i*在时刻*t*处于状态*j*时的有功功率的估计值；

$P_{r(i,j)}$ ——表示用电设备*i*处于状态*j*时的有功功率参考值（均值）；

负荷分解算法通常是将提取的未知负荷的特征参数与数据库中已知负荷的特征参数做比较，找出与提取的负荷特征参数最接近的已知负荷作为识别结果，算法的实现离不开已知负荷数据库的建立，这里负荷数据库的建立主要是功率特征数据库和无功功率特征数据库，本文所用的功率特征参数是指电器运行状态功率的平均值。因此需要计算电器在不同状态运行时的功率平均值。通过事件检测算法可以将各个设备组的设备数据分成若干采样段，将每一采样段的每个采样点处的有功功率与设备库的平均功率进行匹配，即求解目标函数，可得

该设备在该采样点所属的运行状态，汇总该采样时间段的求解状态即可。最后求解上述多负荷实时用电量估计模型即可得设备组中各个设备的实时用电量，最终计算结果详见附件 3。

7 问题四分析及求解

7.1 负荷识别决策树算法

决策树 (Decision Tree) 本是数据挖掘领域的一种分类算法，是一种类似于树形的决策支持工具，表现对象值与属性间的映射关系。树中每个节点表示某个对象，而每个分叉路径则代表的某个可能的属性值，而每个叶节点则对应从根节点到该叶节点所经历的路径所表示的对象的值。决策树是依托于策略抉择而建立起来的分类规则。决策树分类方法采用自顶向下的递归方式，在决策树的内部节点进行属性值的比较并根据不同的属性值判断从该节点向下的分支，在决策树的叶节点得到结论^[8]。

决策树应用在负荷识别中，可认为每个负荷都自成一类，相当于决策树中的叶节点，通过决策树进行分类直到每个类别只含有唯一结果即叶节点纯净为止。决策树中的属性即负荷特征参数，是判断不同向下分支的依据。决策树算法的分治思想是将要解决的问题分解为若干规模更小且易于解决的子问题，先将这些子问题逐个解决，最后由于子问题的解合并构成原问题的解。在划分和解决子问题的过程中，就可充分利用不同负荷特征的优势。如果设备是一个近似纯电阻的负荷，对它进行识别最有效的负荷特征是 V-I 轨迹，构建决策树算法的过程中可以先判定其是电阻类负荷，从而只需要与数据库中同是电阻类的负荷进行相似性比较，从而省去不必要的比较。在识别算法将提取的特征参数与数据库进行比较的过程中，若出现与多种已知负荷都类似的情况是无法准确识别未知负荷的，这时可通过其他负荷特征实现对未知负荷的进一步识别，虽然前面的负荷特征不足以得到最终正确的识别结果，但能缩小后续进行特征参数相似性比较的范围：只需对相似性足够接近的若干负荷用另一种特征参数识别。从上述分析中可以看出基于决策树的识别算法计算量相对较小，能在一定程度上避免使用低效率的负荷特征进行识别。

类似于负荷分解算法，负荷识别算法同样是将提取的未知负荷的特征参数与数据库中已知负荷的特征参数做比较，找出与提取的负荷特征参数最接近的已知负荷作为识别结果，因此也需要负荷数据库的建立，而负荷数据库的建立是在单个负荷特征参数提取的基础上。本次负荷数据库采用两种：第一种数据库仍为设备不同状态下的平均功率组成的数据库，第二种为谐波含有率幅值数据库。

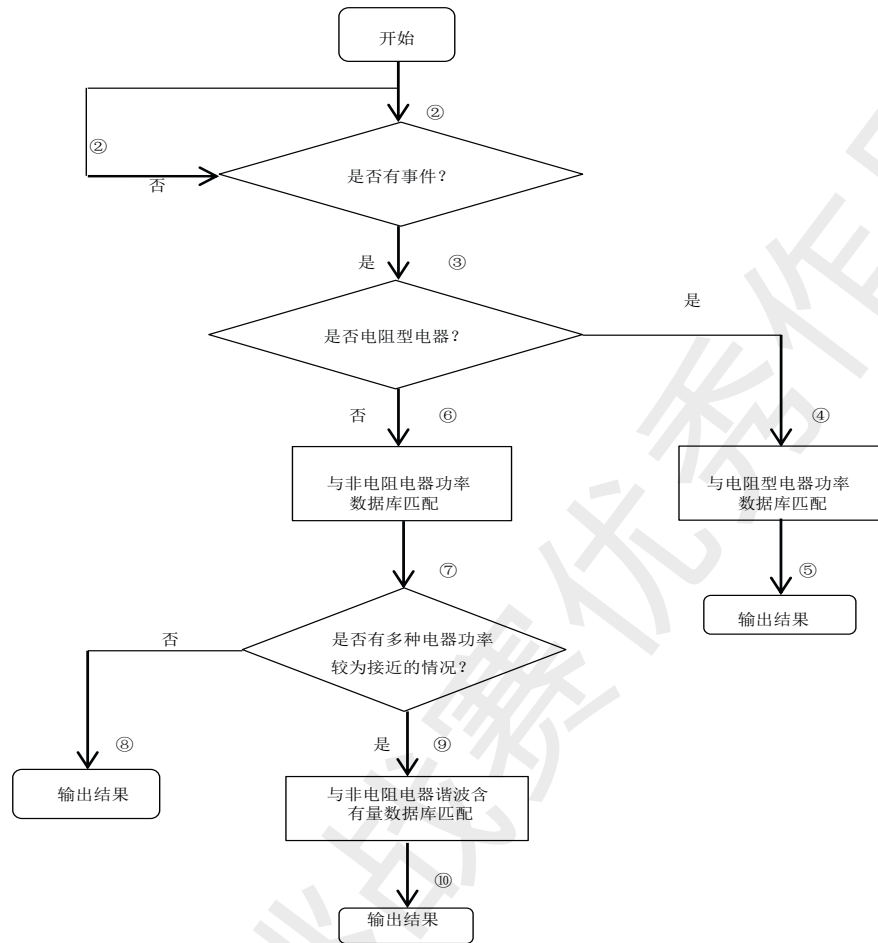


图 12 负荷识别决策树算法流程图

图 12 是基于决策树的负荷识别算法流程示意图，具体流程介绍如下：

- ①由算法判断是否有负荷投切事件发生，若没有则进入步骤②，否则执行③。
- ②读取下一时刻的数据，继续执行步骤①。
- ③判断事件发生的负荷是否为电阻型电器，若是电阻型电器则执行流程④，否则执行⑤。
- ④与电阻型电器功率数据库进行比较，因为事件负荷是电阻型的，只比较有功功率可。
- ⑤输出有功功率最为接近的电器作为识别结果。
- ⑥与非电阻型电器功率数据库进行比较。
- ⑦判断步骤⑥中是否有多种电器相似度很接近，若没有则执行⑧，否则执行⑨。
- ⑧输出⑥中复功率最为接近的电器作为识别结果。
- ⑨提取引发事件负荷的电压-电流轨迹，将得到的谐波含有率与数据库进行比较。
- ⑩输出⑨中最接近的电器作为识别结果。

步骤④⑥⑨中的匹配根据是欧氏距离，将检测到的事件负荷特征参数分别视为空间中的点，同样地将数据库中的特征参数也视为欧氏空间中的点，设点 $x = (x_1, \dots, x_n)$ 和 $y = (y_1, \dots, y_n)$ 分别代表提取到的特征参数和数据库中的特征参数，用式(43)表示其近似程度，其值越小说明近似程度越高。

$$d(x, y) = \frac{\sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{\sqrt{x_1^2 + \dots + x_n^2}} \quad (43)$$

步骤⑦中衡量匹配结果是否存在较为接近是指步骤⑥中按式(43)进行比较, 如果得到最小的若干结果相差在 ε 以内, 则认为比较接近, 随后进入步骤⑨中用谐波含有率进一步识别。

7.2 模型建立

通过上述负荷识别决策树算法确定附件四中的三个设备组的设备组成后采用问题三的模式进行状态识别:

$$\begin{cases} \min J = Y^T Y' - 2Y^T \psi \bar{X} + \frac{1}{2} (\bar{X}^T \psi^T 2\psi \bar{X}) \\ \sum_{i=1}^N x_i = 1 \\ \sum_{i=1}^N (x_i - x_i^2) = 0 \end{cases}$$

其中: N ——设备数目
 ψ ——负荷特性值矩阵。
 Y ——实测数据提取出的待识别的特征向量;
 \bar{X} ——负荷的状态向量 (0 表示关断状态, 1 表示开状态)

7.3 模型求解

通过 lingo 软件编程求解上述 0-1 规划模型, 求解结果详见附件 4。求得各个时间点的设备组中的各个设备后, 用如下公式估计每个设备的实时功率:

$$P_{e,(i,j)}(t) = \frac{P_{r,(i,j)} P(t)}{\sum_{i=I_1}^{I_N} \sum_{j=1}^{n(i)} a_{i,j}(t) \cdot P_{r,(i,j)}} \quad (44)$$

其中: $P(t)$ —— t 时刻设备组的有功功率的值;
 $a_{i,j}$ ——0-1 变量, 取 1 表示设备 i 处于第 j 个状态;
 $n(i)$ ——设备 i 可切换的状态总数;
 $I = \{I_1, I_2, \dots, I_N\}$ ——表示设备组的设备 ID 集合;
 N ——表示设备组所含设备个数;
 $P_{e,(i,j)}(t)$ ——表示用电设备 i 在时刻 t 处于状态 j 时的有功功率的估计值;
 $P_{r,(i,j)}$ ——表示用电设备 i 处于状态 j 时的有功功率参考值 (均值);

求解上述多负荷实时用电量估计模型即可得设备组中各个设备的实时用电量, 计算结果详见附件 4。

8 结论

非侵入式电力负荷监测与分解 (NILMD) 技术在优化电力系统运行、规划和管理, 指导用户优化用电以节省电费和电量, 加速能效技术革新和诱发能效市场变革, 以及促进全社会把提高生态文明意识付诸实践等方面均有重要意义, 而且同侵入式电力负荷监测 (ILM) 相比, 具有易操作、低成本 (投资回收期短)、高可靠、数据完整性好和易于迅速推广等优势, 因此总体上具有广阔的发展前景和无可取代的工程意义。

本文通过分析单一设备数据提取了电流均值, 电流均方根值, V-I 轨迹, 谐波含有率幅值等负荷印记, 将各个负荷印记转化为可量化的指标。通过衡量各类负荷印记的相似度, 确定暂态特征相似度、V-I 轨迹特征相似度、谐波特征的相似度和有功功率无功功率对比相似度的加权组合建立模型用以识别未知设备和状态。针对多负荷状态识别问题, 采用事件检测算法对可能发生事件的时间点进行搜索, 在此基础上通过求解 0-1 二次规划模型进行多负荷的运行状态识别。针对负荷组成未知的状态识别问题, 采用基于决策树的非侵入式用电负荷识别方法, 提出利用决策树算法和数据库, 实现从不同组合负荷运行时的采样数据中, 进行负荷识别, 然后采用 0-1 规划模型确定各个设备的状态。

从分析结果可以看出, 本文采用的分析方法及所建立的模型能够确定各用电设备的状态、操作及操作时间, 并进一步计算实时用电量。

参考文献

- [1] 苏卫星, 朱云龙, 刘芳, 胡琨元. 时间序列异常点及突变点的检测算法[J]. 计算机研究与发展, 2014(4):781-788.
- [2] 冯毅, 王香华. 小波变换降噪处理及其 Matlab 实现[J]. 数据采集与处理, 2006(21):37-39.
- [3] 姚程. 非侵入式用户内部回路低压故障监测系统的研究设计[D]. 北京:华北电力大学, 2017.
- [4] 华栋, 陈小闽, 汪隆君. 基于 V-I 轨迹的非侵入式负荷分解方法及系统:中国, 107302217[P].2017-10-27.
- [5] Hassan T, Javed F, Arshad N. An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring[J]. IEEE Transactions on Smart Grid, 2013, 5(2):870-878.
- [6] 姜波. 基于深度学习的非侵入式住宅用电负荷分解方法[D]. 安徽:合肥工业大学, 2017.
- [7] 王志超. 住宅用电负荷的非侵入式监测方法研究[D]. 重庆:重庆大学, 2015.
- [8] 张棣, 曹健. 面向大数据分析的决策树算法[J]. 计算机科学, 2016, 43(s1).