Vol.42 No.7 Jul. 2022

# 基于模体挖掘与调和函数半监督学习的 非侵入式负荷监测

周勇军<sup>1,2</sup>,吴元香<sup>3</sup>,董智华<sup>1</sup>,胡誉蓉<sup>2</sup>,肖先勇<sup>2</sup>,张 妹<sup>2</sup>

(1. 国网西藏电力有限公司拉萨供电公司,西藏 拉萨 850010;2. 四川大学 电气工程学院,四川 成都 610065;3. 国网西藏电力有限公司,西藏 拉萨 850000)

摘要:针对现有非侵入式负荷监测(NILM)方法成本高昂等问题,提出一种基于模体挖掘与调和函数半监督 学习的NILM方法。基于低频采样数据,根据从监测数据得到的功率阶跃量,利用时间序列分析法和模体挖 掘法划分设备的运行窗;在设备运行窗中,根据设备特性与统计方法定义设备开启最大值到稳定运行点的斜 率、设备稳定运行时的波动幅度2个新的特征量;构建设备运行窗的特征向量,并利用基于调和函数的半监 督学习算法对运行窗中的设备类型进行识别。基于参考能量分解数据集,分别从事件匹配和设备识别的角 度将模体挖掘和基于调和函数的半监督学习算法与其他NILM方法进行对比,验证了所提方法的准确性和可 推广性。

关键词:非侵入式负荷监测;时间序列;模体挖掘;调和函数;半监督学习 中图分类号:TM 73 文献标志码:A

DOI:10.16081/j.epae.202202013

## 0 引言

在用户与电网的双向互动服务模式<sup>[1]</sup>中,对用 户侧设备用电情况的分类计量是一项基础工作<sup>[2]</sup>, 其中非侵入式负荷监测 NILM (Non-Intrusive Load Monitoring)是所采用的一种典型方法,该方法是指 利用用户供电总线处的数据分析出详细的设备使用 状态等信息。这些信息可以帮助用户通过更换节能 电器、调整电器参数、错峰用电等方式来减少电费。 用户也可将这些信息共享给电力企业,以获取更多 的增值服务。通过 NILM 方法,电力企业能更加详 细地掌握用户的用能信息,科学地制定区域电力系 统的发展规划与供电计划<sup>[3]</sup>。

现有关于NILM方法的研究,大多需依靠对单 台设备的单独测量来建立负荷特征库进行模式识 别,但特征库是侵入式收集的,会影响用户正常的生 产生活,因此需要一种可以非侵入式收集、提取和存 档任意用户每台设备唯一特征向量的收集方法<sup>[4]</sup>。 随着用户设备类型与数量的增加,具有相近运行特 征的设备也逐渐增多,当这类设备交叉投切时,准确 分析设备运行时间极为困难。文献[5]通过大量的 算例验证了基于设备运行窗的NILM方法准确率更

# 收稿日期:2021-10-31;修回日期:2021-12-20 在线出版日期:2022-02-22

基金项目:国家自然科学基金资助项目(52007126);国家电网 公司西藏电力有限公司科技项目(52311020009X)

Project supported by the National Natural Science Foundation of China (52007126) and the Science and Technology Project of State Grid Corporation of Tibet Electric Power Co.,Ltd.(52311020009X) 高,但其没有对设备窗的选定进行说明。文献[4]根 据历史标记事件数据、相位、设备可能的运行时间段 等信息以及聚类算法确定设备的运行窗,在一定程 度上提升了分析精度,但辅助信息过多且实时性不 足。文献[6-7]分别通过二分图模型与迭代运算进 行事件匹配,但均未考虑多状态和相近功率的设备。

负荷特征量是用来区分设备类型或运行状态的 重要标签,主要包括设备运行时的稳态特征和状态 转换时的暂态特征。其中,稳态特征包括功率需求、 谐波特性、电压-电流轨迹、电压和电流的波形参数 等。暂态特征包括瞬态功率、谐波和电压波形参数 等<sup>[8]</sup>。但波形参数与谐波分析都需要较高的采样频 率,对配置的测量设备要求较高,这也会导致存储与 分析成本的增加,因此,有必要提取分析更为实用的 负荷特征量,以满足NILM方法推广应用的需求。

最经典的负荷识别方法是以单台设备的特征向 量为对象进行模板匹配<sup>[9]</sup>,并根据与期望值的偏差 情况进行判定。随着对准确度需求的提升,基于判 别函数的分类法被用于负荷识别中,包括基于线性 函数的支持向量机<sup>[10]</sup>、基于非线性函数的神经网 络<sup>[11]</sup>以及基于概率函数的贝叶斯<sup>[12]</sup>分类法。但这 些监督类方法对标记样本的数量需求较高,使得相 应的 NILM 方法在应用中面临成本高昂、实施繁琐 以及居民舒适度低等问题<sup>[13]</sup>。如何降低该过程的工 作量,是 NILM 方法广泛应用面临的关键难题。近 年来,模糊聚类<sup>[14]</sup>与隐马尔可夫模型<sup>[15]</sup>等不对样本 进行标记的无监督类方法也被应用到 NILM 方法领 域,但这类方法存在分类准确度较低且分析结果需 要人工标识等新问题。 为此,本文从NILM方法的实用性角度出发,提 出一种基于模体挖掘与调和函数半监督学习的 NILM方法。首先,基于1Hz低频采样的功率数据, 利用设备运行的基本约束与事件发生的先后逻辑, 确定唯一匹配的设备运行窗并建立设备特征库,对 不能唯一确定的匹配事件,以模体挖掘法分析所 有的可能性;然后,根据不同类型设备开启与稳定运 行时的差异,进一步提取设备运行窗中的特征向量, 并利用对标记样本需求少的基于调和函数的半监督 学习算法对设备类型和运行状态进行识别;最后,基 于参考能量分解数据集 REDD(Reference Energy Disaggregation Data set),分别从事件匹配和设备识 别的角度与其他 NILM方法进行对比,验证本文所 提方法的优越性。

# 1 基于模体挖掘与调和函数半监督学习的 NILM方法架构

本文所提基于模体挖掘与调和函数半监督学习的NILM方法是根据设备投切时在用户总功率曲线上产生的阶跃量,利用完整的设备运行基本约束、事件发生的先后逻辑与模体挖掘,对所有可能的确定设备的运行过程进行划分,并在设备运行窗中利用不同类型的设备在开启瞬态与稳定运行时的波动差异等特征量对各类型设备进行识别的方法。该方法的具体实现流程如图1所示。其中,基于滑动窗的双边累积和事件检测方法<sup>[16]</sup>检测设备投切与状态变化的时刻 $t_i$ 与对应的功率阶跃量 $\Delta P_i$ ,判断功率阶跃量的阈值 $\eta$ =40 W。



图1 本文所提方法基本流程

Fig.1 Basic process of proposed method

# 2 基于时间序列与模体挖掘的设备运行窗划分

从功率曲线的角度,用户设备主要有3种典型 运行过程:单状态设备在开关时具有相近的功率阶 跃量;多状态设备在运行过程中会出现多次不同状 态间功率的需求变换,如洗衣机、烤箱等;通常状态 变化的设备在开关时的功率变化量不同,在运行过 程中设备的功率需求不断变化,如变频空调。

## 2.1 基于时间序列的设备运行窗识别

设备运行窗是指设备从不工作的初始状态到被 开启运行,再到最后被关闭重新回到初始状态的数 据段。根据定义可知设备运行的基本约束为:设备 运行窗的有功功率变化量总和为0;设备在任何状 态都不可能以负有功功率需求运行;设备在关闭前 一定被打开,这被称为事件发生的时间逻辑,即当检 测到一个( $\Delta P_i < 0, t_i$ )的事件时,一定有某一功率阶 跃量 $\Delta P_x > 0$ 且 $\Delta P_x \approx |\Delta P_i|$ (单状态设备)或 $\Delta P_x > 0$ (多 状态设备)发生在 $t_i$ 时刻之前。基本约束对应的单 状态和多状态设备的匹配条件分别如式(1)和式(2) 所示。

$$\left| \Delta P_x + \Delta P_i \right| < \varepsilon \quad t_x < t_i \tag{1}$$

$$\left|\Delta P_{x1} + \Delta P_{x2} + \dots + \Delta P_{i}\right| < \varepsilon \quad t_{x1} < t_{x2} < \dots < t_{i} \quad (2)$$

式中:*ε*为判断匹配的阈值,其值约为10W。

设备运行窗的确定步骤如下。

1)当检测出用户总功率曲线上的所有阶跃量后,在时间轴上形成一系列以阶跃量 $\Delta P_i$ 为符号的事件。

2)读取起始有功功率P<sub>0</sub>,将未匹配的功率阶跃 事件暂时记录于式(3)所示待匹配状态P<sub>w</sub>中,待匹 配状态功率总和始终大于0,且分析数据段的第一 个阶跃量应为正,若为负,则说明某设备的开启事件 不在分析数据段内,不能进行运行窗匹配。

 $P_{w} = P_{0} + \Delta P_{1} + \Delta P_{2} + \dots + \Delta P_{i-1} > 0$  (3)

3)当检测到 $\Delta P_i < 0$ 时,设置匹配标识n=0,依次 假设设备类型为单状态、两状态、三状态等,将  $\Delta P_i$ 如图 2所示向前匹配,判断是否满足基本约束式(1) 或式(2)。每匹配到1次,n值加1。当 $\Delta P_i$ 与此前的 阶跃量全部匹配后,若n=1,对于单状态设备,则说 明在此期间没有发生功率需求相近的设备同时投入 运行的情况,转至步骤6),而对于多状态设备,则需 要进一步分析,转至步骤6),而对于多状态设备,则需 零进一步分析,转至步骤4);若 $n\neq1$ ,则发生了功率 需求相近设备相继开启或者已经开启的某单状态设 备与某多状态设备的状态改变量相近的情况,需等 待下一个  $\Delta P_i < 0$ 的事件,转至步骤3)。



# 图2 各状态设备匹配

Fig.2 Matching of each state device

4)当功率匹配满足多状态设备约束式(2)时,有可能出现如图3所示的3种情况,图中数据单位均为W。情况1下不同状态对应的阶跃量绝对值不同,



Fig.3 Condition of state window when cumulative sum is zero

在当前阶段可以唯一确定单个多状态设备类型,转 至步骤6)。而情况2和3下除了满足式(2)外,还存 在多个单状态设备以及某多状态设备与某状态变化 量相近的单状态设备重叠的可能性,设备所属类型 不能唯一确定,需进行进一步分析,因此将情况2和 3暂时记录于待匹配事件序列式(3)中。

5)检测 P<sub>w</sub>是否小于0:若是,则说明发生了某长时间运行设备关闭的事件,且该设备开启事件发生的时间早于分析数据段对应的时间,此时应调整式(3)中P<sub>0</sub>与阶跃量的关系;否则,返回步骤3)。

6)根据2.3节步骤输出匹配的设备运行窗。

根据上述步骤可以唯一确定一部分设备运行 窗,然后将已确定的设备运行窗从总线数据上移除, 以便于分析其他设备运行窗。

#### 2.2 基于模体挖掘的设备运行窗划分

模体挖掘<sup>[17]</sup>是根据数据挖掘的思想,在时间序 列中寻找重复出现的相似片段。数据挖掘以序列符 号化的形式进行,即在分析运行窗前将监测数据处 理为不同时刻下的功率阶跃量信息。一组能够满足 设备运行窗基本约束的功率阶跃量为一个模体,模 体挖掘即是在一组符号化信息中挖掘出重复出现的 相似片段模体,因此,确定每类设备的第1个模体是 很关键的。

1)在所研究的待匹配事件序列式(3)中,按照时 间顺序对ΔP<sub>i</sub>>0的事件依次以单状态设备或多状态 设备为假设,以设备运行窗基本约束为判定依据,确 定第1个模体。

2)每个模体一一向后对应,挖掘监测数据中所 有可能的设备运行窗。

3)执行上述步骤,直到分析数据的最后一个阶 跃量。

# 2.3 运行窗数据提取

基于上述分析,监测到的所有设备状态都被划 分,一部分是确定的设备运行窗,另一部分是可能的 设备运行窗。确定的设备运行窗将作为已标记样 本,用于辨识可能的设备运行窗是否属于某类设备, 因此需要从分析结果中进一步提取运行窗数据。

以单状态设备为例,当事件匹配后,设备运行窗 中的数据提取过程如下。

1)检查在时间轴上 $\Delta P_i = \Delta P_x$ 事件之间是否有 其他设备事件,若没有,则认为该设备在运行时无其 他设备交叉重叠开启,该设备运行窗的数据窗口取  $(t_x, t_i)$ 之间的数据与 $t_x$ 前一段平稳区段平均值 $P_{av}$ 的 差值,如图4所示,图中, $A_{an}$ 和 $A_{aff}$ 分别为设备A开和 关的标识。



Fig.4 Schematic diagram of data window processing under Case 1

2)若在时间轴上 $\Delta P_x = \Delta P_i$ 事件之间有其他设 备事件,则除了取差值,还需要将其他设备事件从该 数据窗口中移除。考虑到设备重叠运行时大幅度波 动的设备会覆盖小幅度波动的设备,影响本文所提 的设备运行时稳定波动幅度特征值,因此在数据移 除时,若 $\Delta P_x = \Delta P_i$ 之间是 $\Delta P_y > 0$ 的事件,则将 $\Delta P_y$ 到 $\Delta P_i$ 之间的数据替换为 $\Delta P_x$ 到 $\Delta P_y$ 之间的数据,反 之,则将 $\Delta P_x$ 到 $\Delta P_y$ 之间的数据替换为 $\Delta P_y$ 到 $\Delta P_i$ 之 间的数据,如图5所示,图中, $B_{ex}$ 和 $B_{ex}$ 分别为设备 B





开和关的标识。

# 3 负荷特征量的分析与选取

设备在运行过程中受到主要元器件、设备功能 以及人为因素等多方面的影响,将会产生不同的运 行特征。在对1Hz频率采样的功率数据进行分析时,不同类型设备的表现特征如表1所示。为表征 不同类型设备在开启和运行状态下瞬态和波动情况的差异,本文提出2个新的负荷特征量进行负荷 识别。

## 表1 不同类型设备的表现特征

Table 1	Performance	of	different	types	of	equipments
---------	-------------	----	-----------	-------	----	------------

に友						
以甘	设备开启	设备运行				
电阻类	无瞬态	无波动				
电机驱动类	启动时间大于3s,最长启动时间为25s,最大功率与平稳运行功率比值 最大值和最小值分别为9.6和1,最大功率点与平稳运行点相隔3~20s	负载的增减会造成有功功率需求的波动				
电子馈电类	启动时间为0~2s,最大功率与平稳运行功率比值一般为1.3~1.5, 最大功率点与平稳运行点相隔0~1s	一些存在波动,一些无波动				
受人为活动影响类	一些存在较长的瞬态,一些无瞬态	稳定运行时,功率需求变化受人为活动影响较大				

#### 3.1 设备开启最大值到稳定运行点的斜率

图6给出了2种典型的设备开启、稳定运行直到 设备关闭的功率曲线,设备1为电阻类或受人为活 动影响类设备,设备2为电机驱动类或电子馈电类 设备。图中:*P<sub>max</sub>*为设备开启时的有功功率最大值; *P<sub>s</sub>*为特征量,表示设备开启到开始稳定运行点的斜 率,该变量是考虑不同类型设备在开启时的有功功 率需求最大值、稳定运行值与在过渡时间段不同而 设定的;*P<sub>st</sub>*为设备开始稳定运行点的功率;*P<sub>e</sub>*为设 备最终稳定运行点的功率。



图6 2种典型的设备运行窗



为计算  $P_s$ ,对图 6(b)运行窗中的数据进行  $\Delta t$ = 1 s的功率变化量分析,如图 7所示,图中最大的功率 变化量  $\Delta P_{max}$ 可直接定位, $\Delta P_{max}$ 后第 1 个为负值的变 化量对应的有功功率值为  $P_{max}$ ,而最小的功率变化 量  $\Delta P_{min}$ 向前对应的功率值为  $P_s$ 。







由图6(b)可以看出, P<sub>max</sub>到P<sub>st</sub>并不是连续下滑的过程, 而是存在间歇性的功率值回升现象, 因此根据累加和的思想判断P<sub>st</sub>的位置。具体计算步骤

如下。

1)设置阈值 $\beta$ =3,读取 $P_{max}$ 后的阶跃量,置i=1。

2)判断  $\Delta P_i$ 是否为正:若是,则转至步骤3);否则,  $\Diamond_{i=i+1}$ ,继续向前判断  $\Delta P_i$ 的正负。

3)判断  $\Delta P_i + \Delta P_{i+1} + \Delta P_{i+2}$  是否大于 $\beta$ :若是,则 认为已经进入稳定运行区段,输出 $P_{st}$ 对应的时刻 点;否则,认为仍处于过渡过程,令i=i+1,返回 步骤2)。

图 6(b)的设备运行窗连续3个变化量相加的结 果如附录 A 图 A1 所示。当分别确定 $P_{max}$ 和 $P_{st}$ 对应 的时刻点 $t_1, t_2$ 后,根据式(4)计算 $P_{sc}$ 。

$$P_{s} = \left(P_{max} - P_{st}\right) / \left(t_{2} - t_{1}\right)$$

$$\tag{4}$$

#### 3.2 设备稳定运行时的波动幅度

不同类型的设备在*P*<sub>a</sub>到*P*<sub>e</sub>的功率变化量如附 录A图A2所示。由图可以看出,不同设备在稳态运 行时波动幅度不同。设备稳定运行时功率变化量 (取自设备样本a—d)的频数分布直方图如图8所示, 可以看出,数据近似服从正态分布,因此取90%的 置信水平将每个运行窗稳态运行时的波动上下限*P*<sub>f</sub> 作为设备的特征量。

#### 3.3 特征向量的构建

每台设备运行窗对应一个特征向量  $T = [P_{\text{max}} P_{\text{f}} P_{\text{d}}]$ ,其中, $P_{\text{d}}$ 为 $P_{\text{st}}$ 到 $P_{\text{s}}$ 之间设备稳定运行时的 功率需求。

## 4 基于调和函数半监督学习的设备类型识别

本文以基于调和函数的半监督学习算法进行运行窗所属设备类型的辨识。该算法的主要思想是在 学习训练过程中综合利用已标记和未标记样本,从 未标记样本中提取有效的信息来提高学习的性能<sup>[18]</sup>。该算法在所构建的样本图*G*上建立连续预测 函数的分布模型,充分考虑样例的分类概率,解决了 传统离散预测的图割法对相邻类样本误判的问题<sup>[19]</sup>。此外,该算法的基本思想在后续流形正则化







方法的开发中被沿用,是准确性与实用性兼备的一 类半监督学习算法。

将全部样本集X(包括已标记样本集 $X_v$ 和未标 记样本集 $X_v$ )构建在包含N个点和E条边的无向图 G=(N,E)中,图G中每个点表示一个样本。X共包 含N个样本,其中 $X_v$ 的样本数为 $l,X_v$ 的样本数为u, 且N=l+u。X中样本 $x_i$ 的特征向量为T,当 $x_i$ 为已标 记样本时,其标记值为 $y_i$ ,所有已标记样本的标记值 构成集合Y。图G中每条边的权重为 $w_{ij}$ ,表示样本 $x_i$ 和 $x_j$ 之间的相似程度,所有权重构成矩阵W。相似 程度一般是通过距离进行判定的,本文用带宽为 $\sigma$ 的高斯核函数表示样本特征量间的距离。

对于已标记样本,调和函数值与标记值对应;对 于未标记样本,调和函数值满足权重平均原则。设 调和函数为*f*(*x*),则:

$$\begin{cases}
X_{U} = \{x_{i} \in X | i = l + 1, l + 2, \dots, l + u\} \\
X_{V} = \{(x_{i}, y_{i}) \in X \times Y | i = 1, 2, \dots, l\} \\
X = \{x_{1}, x_{2}, \dots, x_{l+u}\} \\
Y = \{y_{V}, y_{2}, \dots, y_{l}\}
\end{cases}$$
(5)

$$d(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right)$$
(6)

$$f(x) = \begin{cases} y_i & i = 1, 2, \cdots, l\\ \sum_{j=1}^{l+u} w_{ij} f(x_i) / \sum_{j=1}^{l+u} w_{ij} & i = l+1, l+2, \cdots, l+u \end{cases}$$
(7)

设计使目标函数值最小的标记预测函数f(x), 对于调和函数而言,它是优化问题式(8)的解。

$$\min_{f(x)\in\mathbb{R}} \left[ \infty \sum_{i=1}^{l} (y_i - f(x_i))^2 + \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} w_{ij} (f(x_j) - f(x_i))^2 \right]$$
(8)

式中: $f(x) \in \mathbf{R}$ 表示调和函数值属于实数域,具有闭 合解。式(8)中第1项为带无穷大加权的损失函数, 用于保证已标记样本的值在计算过程中保持不变,第 2项为正则项 $\Omega(f)$ ,用于保证预测函数对样本数据 点标记值的平滑性,使得邻近样本的预测标记相似。

设对角矩阵 $D = (D_{ij})_{N\times N}$ ,其中,*i≠j*时, $D_{ij} = 0$ ,*i=j* 时, $D_{ij} = \sum w_{ij}$ 。对应图G的拉普拉斯矩阵L = D - W, 令 $f = [f(x_1) f(x_2) \cdots f(x_{l+u})]^{\mathsf{T}}$ 。在图G中,数据 点被划分为2个独立的分别标记为正和负的数据 组,而分割这2个独立数据组的边可以表示为正则 化函数 $\Omega(f) = f^{\mathsf{T}}Lf$ 。

拉普拉斯矩阵L可以划分为4个部分:

$$L = \begin{bmatrix} L_{\rm v} & L_{\rm vu} \\ L_{\rm uv} & L_{\rm u} \end{bmatrix}$$
(9)

式中:L为(l+u)×(l+u)阶矩阵; $L_v$ 、 $L_{uv}$ 、 $L_v$ 、 $L_v$ 、 $J_v$ 、 $L_v$ 为拉 普拉斯矩阵L的4个部分, $L_v$ 为l>l阶矩阵,与已标 记样本部分对应, $L_v$ 为u×u阶矩阵,与未标记样本部 分对应, $L_{uv}$ 为u×l阶矩阵, $L_{vv}$ 为l×u阶矩阵,且 $L_{uv}$ =  $L_{vvo}$ 

类似地, *f* 也可按照已标记和未标记样本分为2 个部分:

$$f = [f(x_1) f(x_2) \cdots f(x_{l+u})]^{\mathrm{T}} = \begin{bmatrix} f_{\mathrm{V}} \\ f_{\mathrm{U}} \end{bmatrix}$$
(10)

式中: $f_v$ 为l维列向量,与已标记样本部分对应; $f_v$ 为u维列向量,与未标记样本部分对应。

正则化函数存在式(11)对应的关系。

 $f^{\mathrm{T}}Lf = f_{\mathrm{V}}^{\mathrm{T}}L_{\mathrm{V}}f_{\mathrm{V}} + 2f_{\mathrm{V}}^{\mathrm{T}}L_{\mathrm{VU}}f_{\mathrm{U}} + f_{\mathrm{U}}^{\mathrm{T}}L_{\mathrm{U}}f_{\mathrm{U}} \qquad (11)$ 

由于正则化函数主要是针对标签中的 $f_{v}$ 进行 求解,对于式(8)中的优化问题也是求极值解,因此 将式(11)对 $f_{v}$ 求偏导并令结果为0,可得:

$$f_{\rm U} = -L_{\rm U}^{-1}L_{\rm UV}f_{\rm V} \tag{12}$$

综上,基于调和函数半监督学习的设备识别步 骤为:

1)所有样本的特征向量形成邻接矩阵₩;

2) 定义已标记样本的调和函数值;

3) 计算拉普拉斯矩阵 *L*, 根据式(12) 计算未标记样本的调和函数值, 以此区分设备类型。

#### 5 算例验证

本节使用麻省理工学院(MIT)发布的从真实 家庭中收集的数据集 REDD<sup>[20]</sup>,验证工作主要从 REDD\_house1和 REDD\_house3数据展开,其中有功 功率的采样频率为1 Hz。为了测试模体挖掘和基于 调和函数的半监督学习算法的性能,从2个方面进 行验证:保持所提方法中其他部分不变,将所提方法 中模体挖掘分别替换为文献[6-7]中事件匹配法进 行对比;保持所提方法中其他部分不变,将所提方法 中基于调和函数的半监督学习算法分别替换为监督 类和无监督类算法进行对比。

# 5.1 模体挖掘验证

负荷识别过程中,样本处理过程为:

1)对每台设备均以5%的已标记样本作为间隔 进行7组测试,其中,每组中正负样本比例和已标记 样本比例均相同:

2)为了保证测试结果的准确性,每组测试反复 随机选择已标记样本计算10次:

3)将10次测试结果的平均值作为最终的结果。

本节利用式(13)中综合评价指标识别准确度 $F_1$ 对分类精确度进行评价。F1的值越高,说明分析越 准确。

$$F_1 = 2pr/(p+r) \tag{13}$$

$$p = T_{\rm p} / \left( T_{\rm p} + F_{\rm p} \right) \tag{14}$$

$$r = T_{\rm p} / \left( T_{\rm p} + F_{\rm n} \right) \tag{15}$$

式中:p为预测为本类的样本中正确的比例;r为本类 样本中预测正确的比例;T。为样本类别预测正确的 个数;F,为将其他类样本预测为本类样本的个数;T, 为将本类的样本预测为其他类样本的个数。

实验结果如图9以及附录A图A3、图A4所示。 在图9的REDD\_house1中,烤箱、微波炉、热水器和 熨斗运行时的功率需求非常接近,文献[6]方法对这 些设备仅进行随机匹配,非常容易出现错误,文献 [7]方法在时间上进行就近匹配,仅部分情况能够匹 配正确,而本文模体挖掘在匹配时考虑了所有的可 能性,采用本文所提方法得到的结果远大于其他方 法。在图 A3 的 REDD house1 中,台灯是多状态设 备,在重叠设备序列中难以进行匹配,并且所有的特



REDD house1

征样本在训练时集中在同一个图中,采用文献[6-7] 方法时无效特征向量太多,会进一步增大识别难度, 导致其他类别设备的识别准确度也略低于本文所提 方法。同样地,在图A4的REDD\_house3中,电视机 与冰箱以及微波炉与热水器运行时的功率需求非 常接近,且卧室灯和厨房灯是多状态设备,因此文 献[6-7]方法对这几类设备的识别准确度也远低于 本文所提方法。

#### 5.2 基于调和函数的半监督学习算法验证

本节将负荷识别的基于调和函数的半监督学习 算法分别与监督类算法中的反向传播 BP(Back Propagation)神经网络和k最邻近kNN(k-Nearest Neighbor)法<sup>[15]</sup>以及无监督类算法中的基于信息熵的模糊 聚类算法<sup>[14]</sup>进行对比。

1)与监督类算法的对比。

监督类算法与半监督类算法均需要部分已标记 样本,样本处理及实验过程与5.1节相同。对比结果 如图10和附录A图A5、图A6所示。可以看出:随着 标记率的提高,3种算法的识别准确度均逐渐升高; 在样本比例较低时,基于调和函数的半监督学习算 法的识别准确度最高,BP神经网络和kNN的识别准 确度较低;在标记率为35%及以上时,3种算法对部 分设备的识别准确度较接近:在标记率为40%及以 下时,基于调和函数的半监督学习算法对部分设备 的识别准确度始终高于另外2种算法,这说明本文 所提方法的准确度与可推广度更高。



Fig.10 Comparison results of three algorithms in REDD house1

# 2)与无监督类算法的对比。

无监督类算法理论上不依赖于已标记样本,但 是在负荷分类结束后仍需要利用部分信息判断所划 分的负荷属于哪一类,因此本节将少量的已标记样 本用于识别,并与同样数量样本识别的半监督类算 法进行对比。图11和附录A图A7给出了在10%的 已标记样本下2种算法的对比结果,图中设备编号 与名称的对应关系见附录A表A1。可以看出,对大 部分设备而言,基于调和函数的半监督学习算法在 较少已标记样本的情况下比基于信息熵的模糊聚类 算法的准确度高。



图 11 在 REDD house1 中 2 种算法的对比结果

Fig.11 Comparison results of two algorithms in REDD\_house1

# 6 结论

减少对人工成本的依赖,是NILM方法推广应 用的关键。本文从实用性角度出发,基于低频采样 数据,提出一种基于模体挖掘与调和函数半监督学 习的NILM方法。利用设备运行的基本约束与事件 发生的时间逻辑,确定唯一匹配的设备运行窗并建 立设备特征库,对不能唯一确定的匹配事件,以模体 挖掘法分析所有的可能性。根据不同类型设备开启 过程与稳定运行时的差异,采用累积和频数分布等 统计方法提取出能反映设备特性的特征量,并通过 构建设备运行窗的特征向量,利用基于调和函数的 半监督学习算法对设备类型进行识别。从事件匹配 和设备识别的角度,将本文所提方法与其他 NILM 方法进行对比,算例分析结果表明:相比于事件匹配 法,本文模体挖掘能够对功率相近的设备和多状态 设备进行准确识别;在负荷识别方面,在准确度需求 相同时,基于调和函数的半监督学习算法比有监督 类算法所需样本标记率低,在低标记率时比无监督 和有监督类算法的准确度高。因此,算例测试证明 了本文所提方法具有更高的准确度和可推广度。

附录见本刊网络版(http://www.epae.cn)。

#### 参考文献:

- [1] 王守相,葛磊蛟,王凯. 智能配电系统的内涵及其关键技术
   [J]. 电力自动化设备,2016,36(6):1-6.
   WANG Shouxiang,GE Leijiao,WANG Kai. Main contents and key technologies of smart distribution system[J]. Electric Power Automation Equipment,2016,36(6):1-6.
- [2] 牟魁翌,杨洪耕.基于 PLA-GDTW 支持向量机的非侵入式负荷监测方法[J].电网技术,2019,43(11):4185-4193.
   MOU Kuiyi, YANG Honggeng. Non-intrusive load identification method based on PLA-GDTW support vector machine[J]. Power System Technology,2019,43(11):4185-4193.
- [3]梁甜甜,高赐威,王蓓蓓.智能电网下电力需求侧管理应用 [J].电力自动化设备,2012,32(5):81-85. LIANG Tiantian,GAO Ciwei,WANG Beibei. Applications of de-

mand side management in smart grid[J]. Electric Power Automation Equipment, 2012, 32(5):81-85.

- [4] DONG M, MEIRA P C M, XU W, et al. Non-intrusive signature extraction for major residential loads [J]. IEEE Transactions on Smart Grid, 2013, 4(3):1421-1430.
- [5] DONG M, MEIRA P C M, XU W, et al. An event window based load monitoring technique for smart meters[J]. IEEE Transactions on Smart Grid, 2012, 3(2):787-796.
- [6] XIAO Y, HU Y, HE H J, et al. Non-intrusive load identification method based on improved KM algorithm[J]. IEEE Access, 2019, 7: 151368-151377.
- [7] XU Z G, CHEN W, WANG Q F. A new non-intrusive load monitoring algorithm based on event matching[J]. IEEE Access, 2019,7:55966-55973.
- [8] 程祥,李林芝,吴浩,等. 非侵入式负荷监测与分解研究综述
   [J]. 电网技术,2016,40(10):3108-3117.
   CHENG Xiang,LI Linzhi,WU Hao,et al. A survey of the research on non-intrusive load monitoring and disaggregation
   [J]. Power System Technology,2016,40(10):3108-3117.
- [9] 宋旭帆,周明,涂京,等. 基于k-NN结合核Fisher判别的非侵入 式负荷监测方法[J]. 电力系统自动化,2018,42(6):73-80.
   SONG Xufan, ZHOU Ming, TU Jing, et al. Non-intrusive load monitoring method based on k-NN and kernel Fisher discriminant[J]. Automation of Electric Power Systems, 2018, 42(6): 73-80.
- [10] 涂京,周明,宋旭帆,等.基于监督学习的非侵入式负荷监测算法比较[J].电力自动化设备,2018,38(12):128-134.
  TU Jing,ZHOU Ming,SONG Xufan, et al. Comparison of supervised learning-based non-intrusive load monitoring algorithms[J]. Electric Power Automation Equipment,2018,38(12): 128-134.
- [11] 张玉天,邓春宇,刘沅昆,等. 基于卷积神经网络的非侵入负荷 辨识算法[J]. 电网技术,2020,44(6):2038-2044.
  ZHANG Yutian, DENG Chunyu, LIU Yuankun, et al. Non-intrusive load identification algorithm based on convolution neural network[J]. Power System Technology, 2020, 44(6):2038-2044.
- [12] MARCHIORI A, HAKKARINEN D, HAN Q, et al. Circuit-level load monitoring for household energy management[J]. IEEE Pervasive Computing, 2011, 10(1):40-48.
- [13] LIU Q, KAMOTO K M, LIU X D, et al. Low-complexity nonintrusive load monitoring using unsupervised learning and generalized appliance models[J]. IEEE Transactions on Consumer Electronics, 2019, 65(1):28-37.
- [14] 孙毅,崔灿,陆俊,等. 基于差量特征提取与模糊聚类的非侵入 式负荷监测方法[J]. 电力系统自动化,2017,41(4):86-91.
   SUN Yi, CUI Can, LU Jun, et al. Non-intrusive load monitoring method based on delta feature extraction and fuzzy clustering[J]. Automation of Electric Power Systems,2017,41(4): 86-91.
- [15] AGYEMAN K, HAN S, HAN S. Real-time recognition nonintrusive electrical appliance monitoring algorithm for a residential building energy management system[J]. Energies, 2015, 8(9):9029-9048.
- [16] 牛卢璐,贾宏杰. 一种适用于非侵入式负荷监测的暂态事件检测算法[J]. 电力系统自动化,2011,35(9):30-35.
   NIU Lulu,JIA Hongjie. Transient event detection algorithm for non-intrusive load monitoring[J]. Automation of Electric Power Systems,2011,35(9):30-35.
- [17] LAXMAN S, TANKASALI V, WHITE R W. Stream prediction using a generative model based on frequent episodes in event sequences [C] // Proceeding of the 14th ACM SIGKDD

International Conference on Knowledge Discovery and Data Mining-KDD 08. New York, USA: ACM Press, 2008:453-461.

- [18] ZHU X, LAFFERTY J, CS L, et al. Semi-supervised learning using gaussian fields and harmonic functions [C]//20th International Conference on Machine Learning. Washington DC, USA:IEEE,2003:912-919.
- [19] 刘建伟,刘媛,罗雄麟.半监督学习方法[J]. 计算机学报, 2015,38(8):1592-1617.
  LIU Jianwei, LIU Yuan, LUO Xionglin. Semi-supervised learning methods[J]. Chinese Journal of Computers, 2015,38(8): 1592-1617
- [20] KOLTER J Z, JOHNSON M J. REDD: a public data set for energy disaggregation research [C]//Workshop on Data Mining Applications in Sustainability. San Diego, USA: ACM Press, 2011:59-62.

作者简介:



周勇军(1975—),男,博士研究生,主要 研究方向为电力系统负荷分析、电力调度管 理及控制(**E-mail**:zyj.tepc@163.com);

胡誉蓉(1997—),女,硕士研究生,研 究方向为负荷特性分析与辨识、谐波分析与 治理(E-mail:2385612083@qq.com);

肖先勇(1968—),男,教授,博士研究生 导师,博士,主要研究方向为电能质量与智能

电网(E-mail:xiaoxianyong@163.com);

张 妹(1988—), 女, 助理研究员, 通信作者, 主要研究 方向为配电网保护与定位、负荷特性分析与建模(E-mail: ZS20061621@163.com)。

(编辑 王锦秀)

# Non-intrusive load monitoring based on motif mining and harmonic function based semi-supervised learning

ZHOU Yongjun<sup>1,2</sup>, WU Yuanxiang<sup>3</sup>, DONG Zhihua<sup>1</sup>, HU Yurong<sup>2</sup>, XIAO Xianyong<sup>2</sup>, ZHANG Shu<sup>2</sup>

(1. Lhasa Power Supply Company of State Grid Tibet Electric Power Co., Ltd., Lhasa 850010, China;

2. College of Electrical Engineering, Sichuan University, Chengdu 610065, China;

3. State Grid Tibet Electric Power Co., Ltd., Lhasa 850000, China)

Abstract: Aiming at the problems such as high cost of existing NILM(Non-Intrusive Load Monitoring) methods, a NILM method based on motif mining and harmonic function based semi-supervised learning is proposed. Based on low-frequency sampling data, according to the power step quantity obtained from the monitoring data, the time series analysis method and motif mining method are used to partition the equipment operation window, two new characteristic quantities of slope from the maximum value when equipment opens to the stable operation point, and the fluctuation amplitude during stable operation of the equipment are defined according to the characteristics of the equipment and statistical methods. The feature vector of equipment operation window is constructed, and the harmonic function based semi-supervised learning algorithm is used to identify the equipment type in the operation window. Based on the reference energy disaggregation data set, the motif mining and harmonic function based semi-supervised learning algorithm are compared with other NILM methods respectively from the perspectives of event matching and equipment identification, and the accuracy and generalization of the proposed method are verified. **Key words**:non-intrusive load monitoring;time series;motif mining;harmonic function;semi-supervised learning

附录 A:



图 A13个连续变化量的累加和







Fig.A2 Fluctuation condition during stable operation of different equipments









Fig.A4 Comparison results of three methods in REDD\_house3





Fig.A5 Comparison results of three methods in REDD\_house1









Fig.A7 Comparison results of two methods in REDD\_house3

数据集 REDD_house1		数据集 REDD_house3		
设备编号	设备名称	设备编号	设备名称	
1	烤箱	14	电视机	
2	空调	15	冰箱	
3	冰箱	16	吸尘器	
4	洗碗机	17	烤炉	
5	台灯	18	卧室灯	
6	洗衣机	19	洗衣机	
7	微波炉	20	烘干机	
8	热水器	21	微波炉	
9	打印机	22	地下室灯	
10	熨斗	23	厨房灯	
11	卧室灯	24	热水器	
12	办公室灯	25	洗碗机	
13	烘干机			

# 表 A1 设备编号与设备名称对应关系

Table A1 Relationship between equipment label and equipment name