

DOI: 10.12158/j.2096-3203.2024.06.003

# 基于时间序列变密度处理的负荷曲线聚类分析

郭文熙<sup>1</sup>, 李知艺<sup>1</sup>, 尹建兵<sup>2</sup>, 陈琳<sup>2</sup>, 鞠平<sup>1,3</sup>

(1. 浙江大学电气工程学院, 浙江 杭州 310027; 2. 国网浙江省电力有限公司杭州供电公司, 浙江 杭州 310000; 3. 河海大学电气与动力工程学院, 江苏 南京 211100)

**摘要:** 负荷曲线聚类是分析用户负荷特性的基础, 能够从大量负荷数据中挖掘典型用电模式, 了解用户电力消费的特点, 对需求响应、电价设计、电网规划等应用具有重要意义。针对现有聚类方法对负荷时段特征考虑不足的问题, 为提升聚类精度和满足实际应用需求, 提出一种基于时间序列变密度处理的聚类方法。首先, 采用线性插值法增加峰、谷、爬坡等3个关键时段数据点的密度, 突出和放大其在聚类中的影响, 并基于自适应分段聚合近似 (adaptive piecewise aggregate approximation, APAA) 降维方法减小冗余数据密度。然后, 结合欧式距离和相关距离构建综合指标, 对负荷曲线开展  $k$ -medoids 聚类分析。最后, 利用 UCI 数据集的居民用户实测数据对所提方法进行验证。实验结果表明, 该方法能有效改善负荷聚类效果, 真实反映了居民用户的用电特性。

**关键词:** 变密度; 负荷聚类; 关键时段; 常规时段; 插值法; 自适应分段聚合近似 (APAA)

中图分类号: TM714

文献标志码: A

文章编号: 2096-3203(2024)06-0021-12

## 0 引言

随着智能电网的建设发展, 各种先进的智能测量设备投入安装, 能够实时快速地从用户处获取海量的负荷数据<sup>[1-3]</sup>。如何利用好这些数据, 挖掘和提取出有价值的信息, 已成为值得研究的重要问题<sup>[4-5]</sup>。近年来, 电力负荷不断增长<sup>[6-7]</sup>, 根据负荷特点对负荷进行归类并分析其特性, 可以把握用户用电规律, 对实施需求响应、电价制定、安排发电计划等都有指导作用<sup>[8-12]</sup>。同时, 随着源荷随机性和波动性逐渐增大, 负荷峰谷差也较大, 给电力系统的调度运行带来一系列困难。如: 随着光伏比例增加, 负荷表现为“鸭型曲线”, 波动加剧, 导致系统调峰压力变大<sup>[13-14]</sup>。因此, 更加迫切地需要负荷特性的准确认知和用电规律的精准掌握。

聚类方法是用电数据挖掘的基础<sup>[15]</sup>, 能够在无标签情况下对负荷进行划分归类<sup>[16]</sup>。目前, 关于负荷聚类的主要研究针对聚类方法及负荷归类的依据<sup>[17-18]</sup>。基于经典聚类算法的不足, 许多改进算法被提出, 如 G-mean 法<sup>[19]</sup>、自适应  $k$ -means 法<sup>[20]</sup>、改进跟随领导者 (modified follow the leader, MFTL) 算法等<sup>[21]</sup>。而根据负荷间相似性归类时, 一方面, 从负荷数据中提取合适的特征指标作为分类依据<sup>[22-28]</sup>, 提升聚类效率和精度; 另一方面, 在直接比

较负荷时间序列相似性时, 选取不同的度量方式替代传统欧式距离<sup>[29-31]</sup>, 保证聚类结果的准确、有效。文献[22]基于数据驱动思想, 使用长短时记忆网络 (long short-term memory, LSTM) 搭建自动编码器提取负荷特征; 文献[23-25]提取负荷频域的特征指标; 文献[26-27]选用日负荷曲线特性指标; 文献[28]为了解用户峰荷需求和用电变化, 根据负荷4个时段的特点定义特征指标。这些方法能够实现负荷数据的低维表达, 保证聚类效率和精度, 但采用的等时间分辨率数据序列难以充分体现负荷不同时间段的差异, 容易丢失极值、斜率等关键特征信息<sup>[32-33]</sup>。改进负荷相似性的度量方式, 主要是针对经典欧式距离判据仅能反映负荷各点数值差异的问题。文献[29]引入动态时间弯曲距离, 关注负荷形状差异; 文献[30]将差分余弦距离结合欧氏距离构成双尺度相似性判据; 文献[31]根据负荷曲线分段斜率方向的异同建立度量指标。因此, 目前研究比较注重负荷的形态变化特征, 有利于负荷的准确分类。

然而, 以上方法在对负荷时间序列提取特征归类的过程中, 以同等重要性看待各时间点数据, 没有根据聚类需求重点考虑负荷关键时段的特性。聚类分析时, 容易受到全局相似区域的影响, 使得不同类型负荷间缺少确切的分界依据, 难以区分。事实上, 考虑聚类目的和实际需求, 不同时段负荷特性认知和分析的压力以及重要程度不一样。峰、谷、爬坡时段的负荷特性在电量转移、电力电量平衡、削峰填谷等方面有重要意义, 对其重点关注, 将

收稿日期: 2024-03-12; 修回日期: 2024-06-09

基金项目: 国家电网有限公司科技项目“广义负荷柔性成分的分析、建模及应用验证方法研究”(5108-202218280A-2-445-XG)

其作为区分负荷差异的重要依据,有利于提升聚类精度,更好地服务于聚类目标和实际应用。同时,对于常规时段,其负荷曲线形状比较平缓、数据比较接近,且电网调度运行的压力较小,因此可以减弱这些非关键时段在聚类中的影响,帮助减轻计算负担和改善聚类效果。

对此,文中针对用户的日负荷曲线,提出一种基于时间序列变密度处理的聚类方法,准确有效地获取和掌握用户典型负荷特性,为电力系统调峰等决策提供依据。首先,考虑精确掌握负荷特性的聚类目的与需求,借鉴注意力机制的思想,提出一种增密度处理手段,重点关注负荷在峰、谷、爬坡3个关键时段的特点及差异,提升负荷聚类的准确性。其次,基于负荷不同时段的特征,提出一种基于时间序列变密度处理的负荷聚类方法,能够在不改变负荷实际功率、维持形态特征的情况下,实现对关键时段信息的强调和常规时段冗余信息的削减,有效改善聚类分析效果。

## 1 负荷时间序列的变密度处理

### 1.1 传统负荷聚类思路

负荷量测数据实质是等间隔的时间序列。假设每组负荷  $P$  共包含  $n$  个采样数据,则可表示为时间序列  $P = \{p_1, p_2, \dots, p_n\}$ 。

通常,对负荷  $P$  进行聚类的思路是:将负荷时间序列视为  $n$  维空间的点,计算各负荷间欧式距离,将相距较近的负荷划分为一类。但以此方式衡量相似性时,各时间点的数值差异累积得到总距离,难以刻画负荷形态变化等特征的相似性,且会同等地受到所有局部差异的影响,使得负荷间缺乏可用于区分的关键特征<sup>[32,34]</sup>。负荷曲线对比如图1所示,其中曲线  $L_1$ 、 $L_2$ 、 $L_3$  分别为不同用户的96点日负荷时间序列。

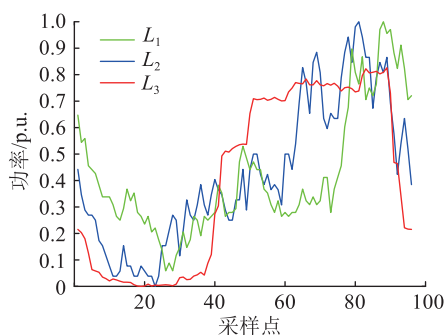


图1 负荷曲线对比

Fig.1 Comparison of load curves

将  $L_2$  作为参考负荷,得到负荷间欧式距离为  $D(L_1, L_2) = 2.1416$ ,  $D(L_2, L_3) = 1.8135$ 。若以此为

相似性判据,则负荷  $L_2$  与  $L_3$  的相似性更高,但图1中负荷  $L_1$ 、 $L_2$  的形态趋势和峰谷变化等都更加相似,因此在聚类时应当将负荷  $L_1$ 、 $L_2$  分入同一类,更符合实际情况。

因此,传统聚类思路对负荷各时间点的差异未作区分,没有充分地体现时段特征,得到的聚类结果也不够准确。图1中,划分到同一类的负荷为  $L_2$ 、 $L_3$ ,其峰、谷、爬坡时段变化趋势的差异较大,聚类时没能将这些关键时段特性相似的负荷挖掘出来,无法较好地贴合电网调峰等实际需求。

### 1.2 负荷重要时段的增密度处理

针对传统负荷聚类的不足,在比较相似性时常采用加权欧式距离作为替代,根据不同时间的重要程度为  $n$  维数据赋予权重,强化关键时段在聚类中的影响。

借鉴机器学习中注意力机制的原理,处理数据时将注意力集中于重要或感兴趣的信息上,有助于提高模型的性能<sup>[35]</sup>。日负荷曲线中,峰时段表示用电高峰,谷时段表示用电低谷,爬坡时段则表示用电快速增长。这些时段在维持供需平衡、实行电量转移等场景下具有重要价值,能更大程度地反映负荷特性及用户用电的相似性,因此将其作为“注意力”主要关注的时段,这些时段的负荷数据应当占有较高权重。设负荷各时间点的权重系数为  $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ,关键时段权重  $\omega_{ke}$  大于常规时段权重  $\omega_{else}$ ,对负荷各点赋予权重后可得加权负荷时间序列  $P_w = \{\omega_1 p_1, \omega_2 p_2, \dots, \omega_n p_n\}$ 。

然而,经过上述加权处理后,如图2(a)所示,不仅改变了负荷重要时段的实际功率值,还破坏了曲线原本的走势和变化特点,由此再开展聚类是不合理的。为克服直接加权的缺陷,文中对这些重要时段进行增密度处理,向负荷序列中补充数据点,增加重点区域负荷数据的密度,如图2(b)所示。相比于其他时段,峰、谷、爬坡段的功率幅值是正常的,但其中包含的数据点数更多,在聚类过程中自然会具有更大的影响,得到的聚类结果也能更多地体现关键时段的特性。因而能够在维持负荷曲线形态特征的同时,改善聚类效果。

假设要对负荷子时间段序列  $\{p_s, p_{s+1}, \dots, p_e\}$  做增密度处理,所得结果如式(1)所示。

$$P_{add} = \{p_s, p_{s+u}, p_{s+1}, \dots, p_{e-1}, p_{e-1+u}, p_e\} \quad (1)$$

式中:  $P_{add}$  为增密度后负荷子时间段序列;  $u = 1, 2, \dots, q$ , 其中  $q$  为每两点间补充的数据点数;  $s$  为负荷子时间段的起始时间点;  $e$  为终止时间点。负荷

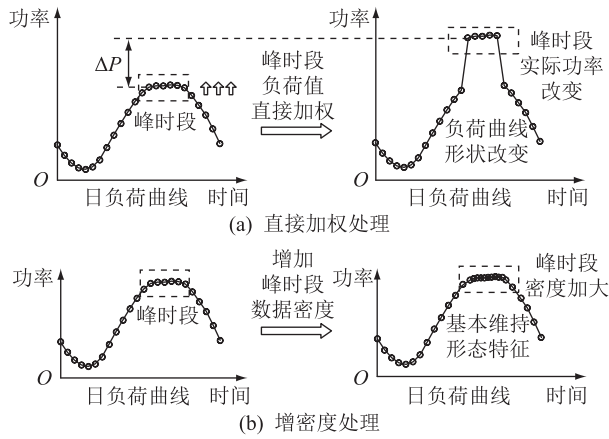


图2 负荷重要时段加权对比示意

Fig.2 Schematic diagram of increasing the data density of load in critical periods

新补充的数据点实质上属于时间序列的一部分,排列于已有的负荷数据之中,实现对关键时段数据密度的增加。

为保证这些数据点与原负荷曲线的变化趋势相符,采用线性内插法确定其对应的负荷值。线性内插法,是指基于已知的离散数据点,采用一次函数插值,以直线近似表示两点间的数据变化规律。假设离散数据点为 $(x_0, y_0)$ 、 $(x_1, y_1)$ ,由线性函数 $y = ax + b$ 可知,两点间内插公式为:

$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) \quad (2)$$

采用该方法,可以确保两侧插值节点误差为0,即不改变原负荷值。并且,以线性函数的近似在目标区域插入若干数据点,可以尽量维持负荷曲线的原变化趋势和形态特征,在保证负荷物理意义的情况下达到变相加权的效果。

如式(1)所示,假设 $u$ 为峰、谷、爬坡时段每个采样间隔插入的数据点, $q$ 为数据点数,依据式(2)在负荷数据间等距插值,则负荷插值点计算如下:

$$p_{t-u} = \frac{up_t + (q - u)p_{t+1}}{q} \quad u = 1, 2, \dots, q \quad (3)$$

式中: $p_t$ 为负荷在第 $t$ 个采样点的功率值; $p_{t+1}$ 为负荷在第 $t+1$ 个采样点的功率值; $p_{t-u}$ 为第 $t$ 和 $t+1$ 采样点之间插入的第 $u$ 个数据点。

### 1.3 负荷常规时段的减密度处理

增密度处理会导致数据点数进一步增加,计算量大,影响聚类精度,因而有必要削减负荷曲线中常规时段的冗余数据。由于负荷曲线本质上是具有形态意义的负荷序列<sup>[36]</sup>,采用自适应分段聚合近似(adaptive piecewise aggregate approximation, APAA)

降维方法,能减小负荷常规时段的数据密度,保留负荷的形态特征和波动特性<sup>[37]</sup>。

在分段聚合近似的基础上提出 APAA 法,根据以下 2 项指标判断负荷在各时段的波动和形态变化情况,从而确定减密度的时段。

(1) 爬坡事件数:在时间窗口 $T$ 内,负荷的最大值与最小值之差 $\Delta p$ 超过设定阈值 $\delta_p$ ,则判断该负荷曲线存在爬坡事件 $\varphi$ 。

$$\varphi = \begin{cases} 1 & \Delta p \geq \delta_p \\ 0 & \Delta p < \delta_p \end{cases} \quad (4)$$

统计所有用户负荷爬坡事件总数 $n_p$ ,即:

$$n_p = \sum_{i=1}^M \varphi_i \quad (5)$$

式中: $\varphi_i$ 为负荷 $i$ 在时间窗口 $T$ 的爬坡事件; $M$ 为负荷数。

(2) 边缘点数:计算负荷采样点 $t$ 前后斜率变化量,若超过设定阈值 $\delta_b$ ,则判断该点为边缘点。

$$\mu_t = \begin{cases} 1 & \Delta b_t \geq \delta_b \\ 0 & \Delta b_t < \delta_b \end{cases} \quad (6)$$

$$\Delta b_t = |(p_{t+1} - p_t) - (p_t - p_{t-1})| \quad (7)$$

式中: $\mu_t$ 为表征采样点 $t$ 是否为边缘点的量; $\Delta b_t$ 为负荷采样点 $t$ 左侧斜率到右侧斜率变化量的绝对值; $\delta_b$ 为设定的阈值。

统计所有用户负荷在 $T$ 内边缘点总数 $n_b$ ,即:

$$n_b = \sum_{i=1}^M \sum_{t=1}^T \mu_{i,t} \quad (8)$$

APAA 法具体实施如下:

(1) 输入预处理后的负荷数据集,设定爬坡事件和边缘点阈值 $\delta_p$ 和 $\delta_b$ ,时间窗宽度 $T$ 。

(2) 以每段时间宽度 $T$ ,对负荷分段。根据式(4)~式(8),统计所有负荷曲线此段的 $n_p$ 值和 $n_b$ 值。

(3) 设定阈值 $\sigma_p$ 和 $\sigma_b$ 。若 $n_p < \sigma_p$ 、 $n_b < \sigma_b$ ,以平均值替代这段序列;若 $n_p < \sigma_p$ 、 $n_b > \sigma_b$ ,以极大和极小值替代这段序列;其余情况,数据直接保留。

(4) 根据步骤(3)中方法,逐段对负荷数据进行处理,获得减密度负荷序列,并输出结果。

对于变化比较平缓的常规时段,采用减密度处理后,可以使用更少的数据点对该段负荷进行近似替代。由于曲线的波动较小,处理后能够保留负荷曲线的形态信息,不阻碍负荷间的区分归类,且不必要数据的削减有利于改善聚类效果。

### 1.4 负荷时间序列变密度处理的基本流程

增密度和减密度处理,实质上都是基于负荷时

间序列不同时段的特点对其进行处理:前者突显和强调关键区域,补充数据点,突出和放大峰、谷、爬坡段特性;后者削减平缓时段数据,以更精简的方式描述负荷。通过这种对各时段数据“扬长避短”的处理方式,能够得到考虑用户特性的变密度负荷时间序列,从而改善后续聚类的效果。

变密度处理后,负荷数据的长度、分布发生改变。但处理过程融合了时段特征,能够在不破坏形态特性的情况下,提高重要时段影响、削减常规时段冗余数据。因此,负荷自身特性得以筛选和提取,且强调了峰、谷、爬坡等关键时段的特性,有利于划分归类和改善聚类效果。

变密度处理的具体流程如图3所示。针对输入的日负荷序列,不同于传统方法中同等对待负荷各时间点数据、直接作为聚类输入的做法,首先需要重点关注的负荷时段,即峰、谷、爬坡时段;随后,采用APAA法寻找和削减负荷冗余数据点,得到减密度负荷序列;最后,为提高重要时段的影响和规避直接加权的不足,确定负荷减密度后关键时段对应的数据点,实行线性插值,得到变密度负荷时间序列,用于后续的聚类分析。

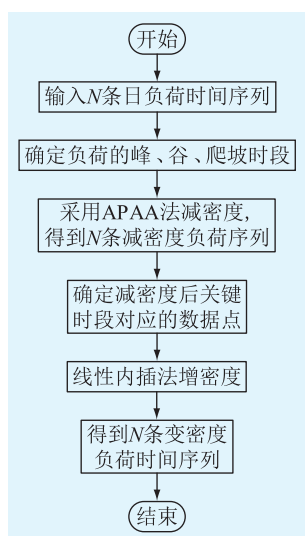


图3 变密度处理的基本流程

Fig.3 Basic flow chart of density-changing processing

## 2 基于变密度负荷时间序列的聚类分析

### 2.1 数据预处理

由于通信干扰、设备故障等原因,负荷数据集含有部分异常数据,需要进行数据清洁,将其剔除或修正。若某条负荷曲线的异常点数超过采样点数的10%,直接移除;否则,采用平滑修正公式<sup>[38]</sup>计算修正值替代异常值。

$$p_{i,t}^* = \frac{\sum_{a=1}^{a_1} p_{i,t-a} + \sum_{b=1}^{b_1} p_{i,t+b}}{a_1 + b_1} \quad (9)$$

式中: $p_{i,t}^*$ 为负荷*i*第*t*个采样点异常数据 $p_{i,t}$ 的修正值; $a_1$ 、 $b_1$ 分别为从第*t*点向前和向后所取的点数,通常设为4~6个点。

同时,各用户负荷的功率存在差异,为使其能在相同数值基准下比较相似性,进行归一化。

$$p'_{i,t} = \frac{p_{i,t} - p_{i,\min}}{p_{i,\max} - p_{i,\min}} \quad (10)$$

式中: $p'_{i,t}$ 为负荷*i*第*t*个采样点归一化后的数据; $p_{i,\max}$ 、 $p_{i,\min}$ 分别为负荷*i*的最大、最小负荷值。

### 2.2 负荷峰、谷、爬坡时段的确定

文中对负荷时间序列的重要时段进行增密度处理,首先须确定负荷的重要时段,即峰、谷、爬坡时段。由于负荷序列各数据点还代表着采样时间,增密度时也应保证对应时间的统一,即:根据时段特征,考虑负荷整体的用电规律,增密度时选取全体负荷都适用的、相同的时间段。

设输入*M*条日负荷曲线,采样点数为*n*,则共包含*n*-1个时间间隔。若将峰、谷及爬坡段的时间长度均选取为*T*,每段包含*v*个采样点,则每条曲线应包含*n*-*v*个时长为*T*的不同时段,可以从中选取最符合峰、谷、爬坡特征的时段。确定负荷关键时段时,对这*n*-*v*段逐一进行判别:对于某一选中的时段,判断*M*条负荷曲线在这段时间窗内属于用电高峰、低谷或快速增长;随后,统计这*n*-*v*个时段的高峰、低谷、快速增长情况对应的曲线数,包含曲线数量最多的时段分别作为全体负荷的峰、谷、爬坡段。确定某一时段是否属于高峰或低谷时,根据其相应特点,选取负荷水平同时高于或低于两侧负荷的时间段,如图4所示。

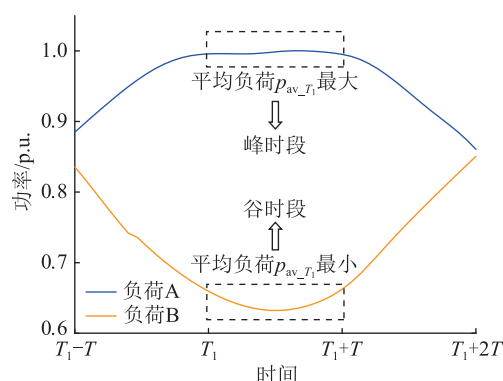


图4 确定负荷峰谷时段示意

Fig.4 Schematic diagram of determining the peak and trough periods

对于某条负荷曲线  $A$  或  $B$ , 若该负荷在子时间段  $\{p_{T_1}, p_{T_1+1}, \dots, p_{T_1+T}\}$  的平均负荷  $p_{av\_T_1}$  大于时间域  $(T_1 - T, T_1 + 2T)$  中其余任意连续时段的平均负荷, 则时段  $(T_1, T_1 + T)$  视为该负荷的一个高峰时段; 若小于任意时段平均负荷, 则视为低谷时段。

负荷快速增长, 选择上升幅值超过设定阈值的时间段。若某负荷在时段  $(T_1, T_1 + T)$  的上升量  $p_{add\_T_1}$  超过阈值  $\varepsilon$ , 即满足式(11), 则该时段视为快速增长时段。

$$p_{add\_T_1} = p_{T_1+T} - p_{T_1} \geq \varepsilon \quad (11)$$

依据以上判据, 通过式(12)~式(14)统计用电高峰在各时段对应的负荷曲线数  $M_{peak}$ 。

$$M_{peak} = \{M_1, M_2, \dots, M_{n-v}\} \quad (12)$$

$$M_h = \sum_{i=1}^M m_{i,h} \quad h = 1, 2, \dots, n - v \quad (13)$$

$$m_{i,h} = \begin{cases} 1 & \text{第 } i \text{ 条曲线的时段 } h \text{ 属于用电高峰} \\ 0 & \text{第 } i \text{ 条曲线的时段 } h \text{ 不属于用电高峰} \end{cases} \quad (14)$$

式中:  $M_h$  为时段  $h$  为用电高峰的负荷曲线数;  $m_{i,h}$  为判断第  $i$  条曲线的时段  $h$  是否属于用电高峰的变量;  $M$  为负荷曲线数。

同理, 计算可得用电低谷、快速增长在各时段对应的负荷曲线数  $M_{trough}$ 、 $M_{climb}$ 。根据统计结果  $M_{peak}$ 、 $M_{trough}$ 、 $M_{climb}$ , 分别取其最大值对应的时间段作为峰、谷、爬坡段。

值得指出的是, 尽管选用相同的时段不适用于所有负荷曲线, 但多数负荷在所选手段表现出峰、谷、爬坡的特点, 足以说明该时段比较关键, 应当在聚类时重点考虑, 即增密度处理, 用更多数据点描述其特征。

### 2.3 负荷相似性度量方式

为充分考虑负荷的数值分布和形态特征, 采取欧式距离和相关距离的综合判据<sup>[33]</sup>, 定义用于变密度时间序列的相似性指标, 计算负荷间差异。

将负荷的峰、谷、爬坡段统称为关键时段。设减密度处理后每组负荷的数据点数为  $m$ , 得到减密度后负荷序列  $P_r = \{p_1, p_2, \dots, p_m\}$ ; 关键时段共补充  $q$  个数据点, 增密度后补充的负荷序列  $P_c = \{p_{c1}, p_{c2}, \dots, p_{cq}\}$ 。

对于这些新插入的数据点, 与其余负荷数据同等看待。因为增加这些时段的数据密度, 不是为了获取更短采样间隔的负荷数据, 而是使关键时段数据在聚类划分时能具有更大的影响。

以负荷  $i, j$  为例, 定义基于变密度时间序列的欧式距离指标, 关注数值的分布特点。

$$d_{i,j} = \sqrt{\sum_{f=1}^m |p_{i,f} - p_{j,f}|^2 + \sum_{g=1}^q |p_{i,cg} - p_{j,cg}|^2} \quad (15)$$

式中:  $d_{i,j}$  为负荷  $i$  和  $j$  之间的欧式距离;  $p_{i,f}$  为负荷  $i$  减密度处理后第  $f$  个数据;  $p_{i,cg}$  为负荷  $i$  增密度处理后补充的第  $g$  个数据。

定义相关距离指标, 描述负荷的波动变化情况。其中, 负荷  $i$  与  $j$  之间的协方差为:

$$L_{i,j} = E((P_{ri} - \bar{P}_i)(P_{rj} - \bar{P}_j)) + E((P_{ci} - \bar{P}_i)(P_{cj} - \bar{P}_j)) \quad (16)$$

$$\bar{P}_i = \left( \sum_{f=1}^m p_{i,f} + \sum_{g=1}^q p_{i,cg} \right) / (m + q) \quad (17)$$

式中:  $L_{i,j}$  为负荷  $i$  和  $j$  之间的相关距离;  $E(\cdot)$  为期望;  $\bar{P}_i, \bar{P}_j$  分别为负荷  $i, j$  变密度处理后的均值;  $P_{ri}, P_{ci}$  分别为负荷  $i$  减密度、增密度处理后所得负荷序列。

负荷标准差为:

$$L_{i,i} = E((P_{ri} - \bar{P}_i)^2) + E((P_{ci} - \bar{P}_i)^2) \quad (18)$$

式中:  $L_{i,i}$  为负荷  $i$  的标准差。

基于以上 2 项指标, 定义相关距离指标如下:

$$r_{i,j} = \frac{L_{i,j}}{\sqrt{L_{i,i}L_{j,j}}} \quad (19)$$

$$z_{i,j} = 1 - r_{i,j} \quad (20)$$

式中:  $r_{i,j}$  为负荷  $i, j$  的相关系数,  $r_{i,j} \in [-1, 1]$ ;  $z_{i,j}$  为相关距离,  $z_{i,j} \in [0, 2]$ 。

基于已有研究对各距离指标归一化并加权作和的处理方式, 得到衡量变密度时间序列相似性的综合指标<sup>[33, 39]</sup>。

$$S_{i,j} = \alpha d_{i,j} + \beta z_{i,j} \eta \quad (21)$$

$$\eta = \max(d) / \max(z) \quad (22)$$

式中:  $S_{i,j}$  为负荷  $i, j$  之间的综合相似性指标;  $\alpha$  为欧式距离对应权重,  $\beta$  为相关距离对应权重,  $\alpha + \beta = 1$ ;  $\eta$  为最大值归一化的比例系数。

### 2.4 聚类算法的实施

输入预处理、变密度处理后的负荷时间序列, 基于式(21)中的相似性指标, 采用  $k$ -medoids 法<sup>[40]</sup> 进行聚类, 具体步骤如下。

(1) 聚类中心初始化: 设定聚类数  $K$ , 并在输入的负荷序列  $P$  中, 随机选取初始聚类中心  $c^{(k)}$ ,  $k = 1, 2, \dots, K$ 。

(2) 划分归类: 计算负荷与各聚类中心  $c^{(k)}$  的相似性  $S_{p,c}$ , 与最相似的聚类中心划分至同一类。

(3) 更新聚类中心: 计算各类中所有负荷与类内其余负荷相似性之和, 取最小者为新聚类中心。

(4) 迭代:若步骤(3)的聚类中心与步骤(2)不同,再次执行步骤(2)和(3);否则,迭代结束,输出聚类结果。

### 2.5 聚类有效性评价指标

如果所用负荷数据缺乏标签信息,需要建立合适的有效性指标对聚类具体效果进行评估,以此作为验证聚类有效性的依据。Davies-Bouldin 指标(Davies-Bouldin index, DBI)、误差平方和(sum of squared error, SSE)、Calinski-Harabasz 指标(Calinski-Harabasz index, CHI)等是典型的聚类评价依据,可以通过类间分散性及类内紧密性评估聚类的有效性<sup>[41]</sup>。文中将其选作聚类评价指标,定量分析聚类效果。

(1) DBI。基于变密度负荷时间序列的聚类对象,得到计算公式<sup>[42]</sup>如下:

$$I_{DBI} = \frac{1}{K} \sum_{k=1}^K R_k \quad (23)$$

$$R_k = \max_{k \neq j} \frac{S(Y_k) + S(Y_j)}{S(c_k, c_j)} \quad (24)$$

式中:  $I_{DBI}$  为聚类结果的 DBI;  $K$  为聚类数;  $S(Y_k)$  为第  $k$  个负荷类簇  $Y_k$  的类内平均相似性;  $S(c_k, c_j)$  为聚类中心  $k, j$  间相似性。  $I_{DBI}$  越小, 聚类有效性越高。

(2) SSE。用各负荷到所属聚类中心的欧式距离平方和表示, 计算公式如下:

$$I_{SSE} = \sum_{i=1}^K \sum_{P \in Y_i} S^2(c_i, P) \quad (25)$$

式中:  $I_{SSE}$  为聚类结果的 SSE 指标;  $S(c_i, P)$  为负荷与聚类中心间的相似性。  $I_{SSE}$  越小, 聚类有效性越高。随着聚类数  $K$  的增大,  $I_{SSE}$  会逐渐减小, 其拐点可作为确定聚类数的依据。

(3) CHI。通过类间分散性和类内紧密性评估, 计算公式如下:

$$I_{CHI} = \frac{\sum_{k=1}^K \sum_{i=1}^M w_{k,i} S^2(c_k, \bar{P}) / (K-1)}{\sum_{k=1}^K \sum_{i=1}^M w_{k,i} S^2(c_k, P_i) / (M-K)} \quad (26)$$

$$w_{k,i} = \begin{cases} 1 & P_i \in Y_k \\ 0 & P_i \notin Y_k \end{cases} \quad (27)$$

式中:  $I_{CHI}$  为聚类结果的 CHI;  $\bar{P}$  为所有负荷平均值;  $w_{k,i}$  为负荷  $i$  对第  $k$  类的从属关系。  $I_{CHI}$  越大, 聚类有效性越高。

## 3 算例分析

算例选用 UCI 数据集中的电力负荷数据集<sup>[43]</sup>。该数据集记录了葡萄牙 370 个家庭用户 2011 年—

2014 年的负荷数据, 采样间隔 15 min, 可以作为所需的居民负荷实测数据使用。从 2014 年的负荷数据中, 选取 10 d 用户日负荷数据, 共计 3 700 条日负荷曲线, 采用文中方法对其特性进行聚类分析。

### 3.1 负荷时间序列的变密度处理

#### 3.1.1 负荷减密度结果

取宽度  $T = 1$  h 的时间窗进行分段, 将负荷序列均分为 24 段, 每段 4 个数据点。设定 APAA 法阈值  $\delta_p = 0.18, \delta_b = 0.036; \sigma_p = 1\ 100, \sigma_b = 9\ 700$ 。

经过统计比较: 在第 15 段时间窗内负荷波动较缓, 取均值替代; 第 4 段时间窗内曲线波动剧烈但幅度小, 取极值替代。最终, APAA 法削减 53 个数据点, 将 96 维的日负荷时间序列降至 43 维。

#### 3.1.2 负荷增密度结果

为保证计算效率和避免维数灾, 负荷峰、谷及爬坡段的长度均选取为 1 h, 即在 4 个间隔内插入数据点。对负荷曲线峰、谷、爬坡段在一天中的分布情况进行统计, 取曲线数最多的时间段。最终, 将峰时段确定为 12:00—13:00, 谷时段确定为 05:00—06:00, 爬坡时段确定为 08:30—09:30。

将其与 APAA 法减密度处理后的负荷序列进行匹配: 峰时段对应减密度负荷序列的第 24、25、26 维数据值, 谷时段对应第 5、6 维数据值, 爬坡时段对应第 16、17、18、19、20 维数据值。

在选定的负荷数据间逐渐增加插值点数, 进行增密度处理, 后续聚类结果如表 1 所示。由表 1 可知, 当插值点数为 3 时, DBI 的增长达到拐点, 聚类效果最佳。因此, 增密度处理时, 在重要的峰、谷、爬坡时段数据间各补充 3 个数据点, 最终得到 64 维负荷序列, 用于后续聚类分析。

表 1 不同插值点数下的聚类效果  
Table 1 Clustering effect under different interpolation points

插值点数	DBI	插值点数	DBI
0	0.853 7	3	0.788 3
1	0.832 3	4	0.792 7
2	0.816 1	5	0.798 2

#### 3.1.3 负荷直接加权结果分析

基于 3.1.2 节确定的负荷数据点, 对 2 条总体趋势较相似的负荷曲线加权处理, 结果如图 5 所示。

由图 5 可见: 负荷曲线  $X_1, X_2$  加权前, 整体变化趋势比较相似; 峰、谷、爬坡时段各取 4 个数据点加权后, 由于功率值被改变, 整条曲线多出 3 个尖峰, 形态发生明显变化。虽然  $X_1, X_2$  峰时段仍相近, 但在谷、爬坡时段, 由于一条负荷曲线向上波动, 另一

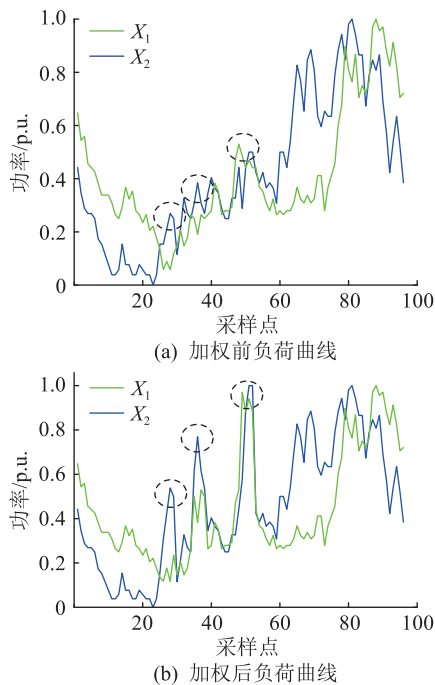


图5 直接加权得到的负荷曲线

Fig.5 The directly weighted load curves

条向下波动,加权后差距明显拉大,使得原本略有差异的负荷特性出现较大区别,直观展现了负荷直接加权的缺陷。

### 3.2 实际负荷曲线聚类分析

将聚类数  $K$  取值范围设定为 2~12 的整数。在不同聚类数  $K$  下,分别执行算法并计算有效性指标 DBI、SSE、CHI,结果如图 6 所示。根据图 6,聚类数  $K$  为 3 时 DBI 最小,CHI 最大,SSE 曲线也达到拐点,3 种有效性指标评估下的最优聚类数相同。因此,确定聚类数  $K = 3$ 。

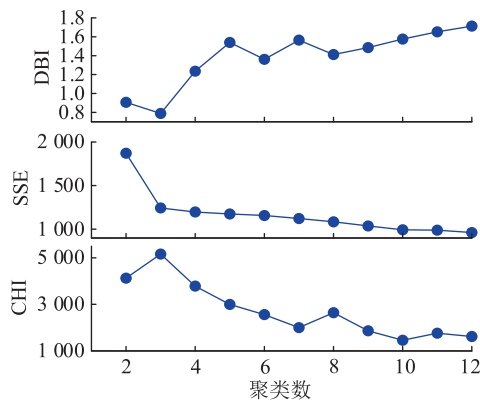


图6 不同聚类数下的聚类有效性指标

Fig.6 Cluster validity indexes with different cluster numbers

基于改进熵权法<sup>[39]</sup>,确定聚类中相似性衡量指标系数  $\alpha$ 、 $\beta$  分别为 0.352 8 和 0.647 2,对得到的负荷变密度时间序列开展聚类分析,结果如图 7 所示。

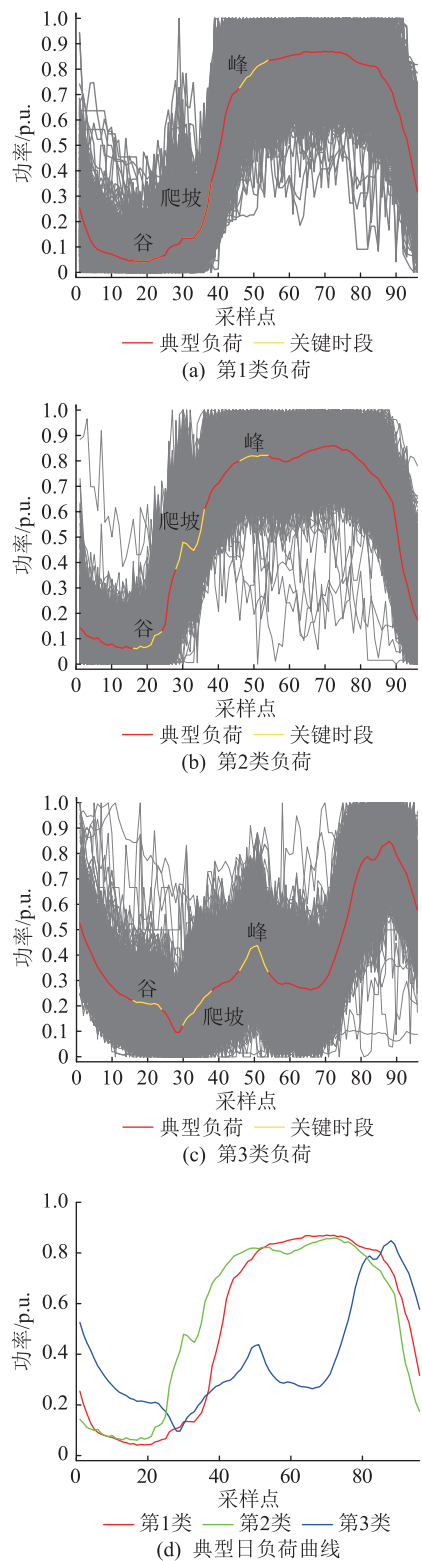


图7 负荷聚类结果

Fig.7 Load clustering results

经过预处理后,共 3 531 组居民负荷数据作为输入参与聚类,分为图 7(a)—图 7(c) 所示的 3 类负荷。对类内所有负荷取均值,获得对应的典型日负荷曲线如图 7(d) 所示。

根据典型负荷曲线的特点和变化趋势,可知各

类居民用户的负荷特性和用电特征,如表 2 所示。

表 2 各类居民用户的负荷特性  
Table 2 Load characteristics of each type of residential customers

类别	用户负荷曲线数	负荷特性
1	885	单峰,用电高峰大约在 11:00—22:00, 负荷水平很高,爬坡较快
2	1 643	单峰,用电高峰大约在 10:00—20:00, 负荷水平较高,爬坡较快,中途有降落
3	1 003	双峰,用电高峰分别在 13:00 和 22:00 左右, 13:00 附近负荷水平中等,爬坡较缓, 22:00 附近负荷水平较高,爬坡较快

由图 7 可知,峰、谷、爬坡段负荷的变化趋势等特性,是聚类时用于区分的重要依据,所得 3 类负荷反映着不同的用电规律。对于选择的时段:第 1 类负荷峰时段仍在上升,谷时段平缓、稳定,爬坡很快;第 2 类负荷峰时段平缓、稳定,谷时段开始有上升势头,爬坡快,但中途会出现下降;第 3 类负荷峰时段先上升后下降,谷时段开始有下降势头,爬坡缓。因此,所用方法较好地把握了负荷特点和关键时段特性,并可以依据这些特征实现负荷间的划分归类,聚类结果较为合理,验证了所提方法的聚类效果。并且,根据表 2 所示的不同类型负荷特性,可以帮助提供更好的电力服务,为用户参与电网调节、解决调峰问题等各项决策提供依据,满足聚类目的和实际需求。

相比于传统方法,在充分融合时段特征后,所用聚类方法能够准确地将峰、谷、爬坡时段相似的负荷划分到一起,且其余常规时段的相似性也较高,形态特征未被破坏。因此,数量众多的用户负荷可以用几类典型的负荷曲线表示特性,化繁为简、分门别类地获取和分析各自的特性。如图 7 和表 2 所示,各类负荷关键时段具有不同的变化特点,得到了有效细致的区分,并能够把握整体的变化趋势,满足细致认知峰、谷、爬坡时段的目标,解决负荷不同时段重要程度有差异的问题。

### 3.3 不同季节聚类结果比较

用户在不同季节的用电行为通常会表现出一定的差异性,尤其是峰、谷、爬坡这些重要时段的特性会发生改变,故采用文中方法对不同季节的负荷特性进行挖掘和比较,了解用户负荷特性的变化情况。随机选取冬季 1 月、夏季 8 月各 10 d 的用户日负荷数据,开展聚类分析,结果如图 8 所示。

图 8 表明,冬夏的用户负荷特性存在一定差异。尽管 3 种典型的日负荷曲线在总体趋势上比较相似,但许多时间点的负荷功率、变化情况仍表现出

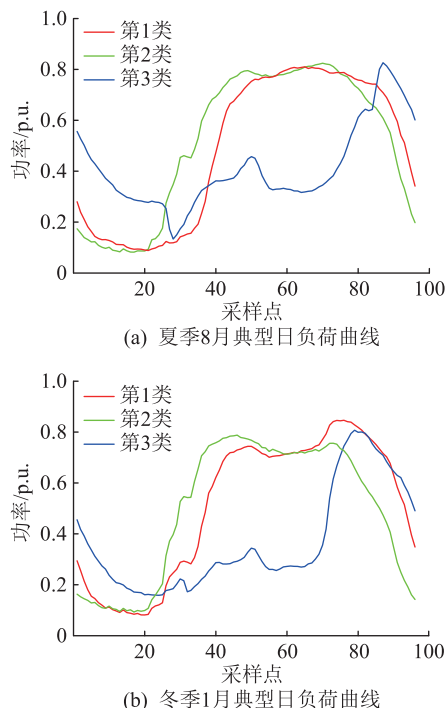


图 8 冬夏典型日负荷聚类结果

Fig.8 Load clustering results of summer and winter typical days

明显不同,因而有必要采用文中方法分析负荷在冬夏季节的不同特性,了解用电情况。

文中所提聚类方法,重点关注负荷在峰、谷、爬坡等 3 个重要时段的特点,能较好地把握负荷间差异。根据图 8 所示结果,冬夏负荷曲线的不同主要体现在上述 3 个重点时段:(1) 第 1、2 类在峰时段的功率变化,夏季峰负荷水平一直较高,相对平稳,冬季峰负荷相对波动更大;(2) 第 1 类爬坡段的功率变化,夏季负荷增长较快,冬季负荷在增长途中会出现短暂的降落;(3) 第 3 类在谷时段的功率变化,夏季谷负荷水平很低,冬季谷负荷相对更高,都体现着负荷特性的差异性。由此可见,在分析用户用电特征时,采用文中方法,可以较好地反映峰、谷、爬坡期的用户功率特性,细致把握不同场景下负荷特性的变化,从而针对性地指导应用。

### 3.4 聚类方法的对比分析

#### 3.4.1 增密度处理对负荷聚类质量的影响

为验证关键时段增密度处理的有效性,对低分辨率的负荷采样数据、增密度处理后的负荷数据以及在关键时段进行高分辨率采样的负荷数据分别进行聚类,以 DBI、SSE、CHI 3 项聚类评价指标对比分析聚类效果的好坏。由于所用数据集为 15 min 分辨率的居民负荷,为获取低分辨率数据,从中抽样得到 1 h 分辨率的 24 点负荷数据。

分别在 3 种场景下聚类:场景①,直接采用低分



分辨率的 1 h 数据,做  $k$ -means 聚类;场景②,基于 3.1.2 节确定的峰、谷、爬坡时段的对应时间,对 1 h 的负荷数据进行增密度处理,然后做  $k$ -means 聚类;场景③,在峰、谷、爬坡时段使用 15 min 分辨率的负荷数据,其余时段不变,然后做  $k$ -means 聚类。不同场景下的聚类有效性指标如表 3 所示。

表 3 不同场景下的聚类效果  
Table 3 Clustering effect under different scenarios

场景编号	DBI	SSE	CHI
①	1.277 9	1 080.1	3 663.9
②	1.035 2	975.7	4 039.9
③	1.014 4	880.1	4 315.3

由表 3 可知,关键时段更高的数据密度有助于提升负荷聚类质量。相比于使负荷各时间点差异在聚类中具有相同的影响,对关键时段特征的重视可以为划分归类提供更明确、更有针对性的依据,提高不同类型负荷的可区分性。因此,增密度处理和引入实测数据这 2 种场景下的聚类指标都更优,如 DBI 值分别提升了 18.99% 和 20.62%。虽然实测数据的聚类效果仅略优于增密度,但可以真实地提供关键时段的负荷信息,能更充分有效地反映这些时段的负荷特性,对于负荷聚类更有积极意义。

综上所述,增密度处理获得关键时段更高的数据密度有助于聚类。同时,实际应用中也可以此原则配置量测,以更高的采样频率获取峰、谷、爬坡段的负荷数据,如:针对常见的 15 min 分辨率量测,以 1 min 采样间隔获取部分时段的负荷数据,通过少量投入获取负荷关键信息,改善聚类效果。

#### 3.4.2 不同方法下的聚类结果及对比分析

基于聚类过程中采用的增密度线性内插、减密度 APAA 降维等手段,以及一些典型的聚类方法,使用 3.2 节所用负荷数据,对实施不同方法时的聚类效果进行对比分析,验证所用的变密度处理法是否能准确把握负荷关键特性,提升聚类的有效性。

根据负荷时间序列变密度处理过程,应用以下聚类方法对比,基于 3 种聚类评价指标进行定量分析。方法①,传统  $k$ -means 法;方法②,谱聚类;方法③,高斯混合聚类;方法④,增密度处理+综合指标  $k$ -medoids;方法⑤,增密度处理+欧氏距离  $k$ -medoids;方法⑥,减密度处理+综合指标  $k$ -medoids;方法⑦,经变密度处理,基于综合指标判据的  $k$ -medoids(文中方法)。

使用上述方法分别进行聚类分析后,得到表 4 所示结果。由表 4 可知,方法①、②、③所得结果的 DBI、SSE 值明显高于方法⑦,CHI 值明显低于方法

⑦。相比于前 3 种方法,文中方法的 DBI 指标分别提升了 37.92%、17.20%、28.66%,SSE 指标分别提升了 71.54%、35.47%、45.71%,CHI 指标分别提升了 38.25%、43.60%、33.25%。3 种指标的定量分析表明,文中所用方法考虑负荷的时段特征后,有效性明显优于典型的聚类方法。

表 4 不同方法的聚类效果

Table 4 Clustering effect applying different methods

方法编号	DBI	SSE	CHI
①	1.269 9	4 366.8	3 733.7
②	0.952 4	1 926.0	3 594.6
③	1.165 6	3 973.7	3 873.8
④	0.868 1	2 301.2	4 037.8
⑤	1.032 4	3 938.8	3 355.8
⑥	0.853 7	2 104.0	4 154.0
⑦	0.788 3	1 242.9	5 162.0

方法④、⑤所得结果的 DBI 显著低于方法①,SSE、CHI 相对比较接近,略优于方法①,再次说明了增密度处理能有效改善聚类质量。而方法④相比于方法⑤,以综合指标衡量负荷相似性,有效性指标进一步提升,DBI 从 1.032 4 减小至 0.868 1,SSE 从 3 938.8 减小至 2 301.2,CHI 从 3 355.8 增加至 4 037.8,说明了聚类时衡量负荷形态差异的重要性。

进一步地,方法⑥所得结果的 DBI 值、SSE 值低于方法①,CHI 值高于方法①,也证明了负荷降维减密度处理对于聚类结果的积极作用,能够提升聚类精度。相比于增密度和减密度处理后的负荷聚类,方法①没有充分体现时段特征,其有效性明显更低,说明了针对不同时段的处理对于负荷聚类的积极作用。

方法⑦,即文中方法,聚类评价指标均优于其他方法,证明了文中方法的优越性,能够充分地把握负荷关键特性,聚类时效果更佳。并且,方法⑦所得结果的指标优于方法④和方法⑥,以 DBI 值为例,分别提升了 9.19% 和 7.66%,聚类精度进一步提高。这说明负荷时间序列的变密度处理,对于峰、谷、爬坡时段的重要性、常规时段曲线形态特性的充分体现都具有积极意义,关注聚类需求的同时也省略了一些不必要的信息,相比于增密度或减密度处理的单独应用更具优势,也进一步验证了根据不同时段的特征处理负荷对于聚类的积极作用。

## 4 结论

针对负荷曲线的聚类分析问题,文中提出一种

基于时间序列变密度处理的聚类方法,对峰、谷、爬坡时段以及常规平缓时段负荷数据采用不同的处理方式,并结合  $k$ -medoids 法实现对负荷的聚类分析。得到以下结论:

(1) 相比于传统聚类方法,所用方法融合了负荷不同时段特征,其中关键时段的增密度处理和常规平缓时段的减密度处理都对负荷聚类具有积极作用,DBI、CHI、SSE 等聚类评价指标得到明显提升,有效地改善了聚类效果。

(2) 所得聚类结果能较好地反映用户峰、谷、爬坡段的特点,通过典型负荷曲线,可以获取各类负荷特性,且关键时段的特性更加精准。因而负荷能够得到合理分类,从而实行分组控制策略,更好地发挥其在需求响应等实际应用中的价值。

#### 参考文献:

- [1] 王保义,胡恒,张少敏. 差分隐私保护下面向海量用户的用电数据聚类分析[J]. 电力系统自动化,2018,42(2):121-127.  
WANG Baoyi,HU Heng,ZHANG Shaomin. Differential privacy protection based clustering analysis of electricity consumption data for massive consumers[J]. Automation of Electric Power Systems,2018,42(2):121-127.
- [2] TONG X,KANG C Q,XIA Q. Smart metering load data compression based on load feature identification[J]. IEEE Transactions on Smart Grid,2016,7(5):2414-2422.
- [3] 刘晨,白泰,王家驹,等. 用户用电信息采集系统的深化应用研究综述[J]. 电测与仪表,2022,59(2):1-8.  
LIU Chen,BAI Tai,WANG Jiaju,et al. Review on the deepening application of electricity consumption information acquisition system[J]. Electrical Measurement & Instrumentation,2022,59(2):1-8.
- [4] 严玉婷,薛冰,方力谦,等. 基于周期调整负荷成分指数的行业用电大数据价值挖掘[J]. 电力科学与技术学报,2022,37(6):181-189.  
YAN Yuting,XUE Bing,FANG Liqian,et al. Big data mining of industry power consumption based on component index about seasonal-adjusted load[J]. Journal of Electric Power Science and Technology,2022,37(6):181-189.
- [5] LIU J P,WANG Z,HU H,et al. Research on the optimization strategy of customers' electricity consumption based on big data[J]. Global Energy Interconnection,2023,6(3):273-284.
- [6] 张洁,夏飞,袁博,等. 基于特征优选策略的居民用电行为聚类方法[J]. 电力系统自动化,2022,46(6):153-159.  
ZHANG Jie,XIA Fei,YUAN Bo,et al. Clustering method for residential electricity consumption behavior based on feature optimization strategy[J]. Automation of Electric Power Systems,2022,46(6):153-159.
- [7] 解洋,梅飞,郑建勇,等. 基于 V-I 轨迹颜色编码的非侵入式负荷识别方法[J]. 电力系统自动化,2022,46(4):93-102.  
XIE Yang,MEI Fei,ZHENG Jianyong,et al. Non-intrusive load monitoring method based on V-I trajectory color coding[J]. Automation of Electric Power Systems,2022,46(4):93-102.
- [8] RYU S,CHOI H,LEE H,et al. Convolutional autoencoder based feature extraction and clustering for customer load analysis[J]. IEEE Transactions on Power Systems,2020,35(2):1048-1060.
- [9] 张铁峰,顾明迪. 电力用户负荷模式提取技术及应用综述[J]. 电网技术,2016,40(3):804-811.  
ZHANG Tiefeng,GU Mingdi. Overview of electricity customer load pattern extraction technology and its application[J]. Power System Technology,2016,40(3):804-811.
- [10] VAHEDI S,ZHAO L. Distributed auto-clustering for residential load profiling using AMI data from the U.S. high plains[J]. IEEE Transactions on Smart Grid,2023,14(6):4530-4541.
- [11] 李柏新,雷才嘉,方兵华,等. 基于日负荷指标及改进分布式 K-means 聚类的用户用电规律研究[J]. 电测与仪表,2023,60(10):104-111.  
LI Baixin,LEI Caijia,FANG Binghua,et al. Research on typical electricity consumption law based on daily load indicator and improved distributed K-means clustering[J]. Electrical Measurement & Instrumentation,2023,60(10):104-111.
- [12] 赵莎莎,朱雅魁,王悦. 基于大数据分析的综合能源系统负荷特性聚类分析[J]. 电测与仪表,2023,60(2):10-15,52.  
ZHAO Shasha,ZHU Yakui,WANG Yue. Cluster analysis of integrated energy system load characteristics based on big data analysis[J]. Electrical Measurement & Instrumentation,2023,60(2):10-15,52.
- [13] 陈皓勇,谭碧飞,伍亮,等. 分层集群的新型电力系统运行与控制[J]. 中国电机工程学报,2023,43(2):581-595.  
CHEN Haoyong,TAN Bifei,WU Liang,et al. Operation and control of the new power systems based on hierarchical clusters[J]. Proceedings of the CSEE,2023,43(2):581-595.
- [14] 叶伦,姚建刚,杨胜杰,等. 含高比例可再生能源电力系统的调峰成本量化与分摊模型[J]. 电力系统自动化,2022,46(15):20-28.  
YE Lun,YAO Jiangan,YANG Shengjie,et al. Quantification and allocation model of peak regulation cost for power system with high proportion of renewable energy[J]. Automation of Electric Power Systems,2022,46(15):20-28.
- [15] 杜向阳,熊小伏,王建,等. 基于负荷有序聚类的主动配电网储能规划方法[J]. 电力科学与技术学报,2023,38(6):187-197.  
DU Xiangyang,XIONG Xiaofu,WANG Jian,et al. Energy storage planning method of active distribution network based on load ordered clustering[J]. Journal of Electric Power Science and Technology,2023,38(6):187-197.
- [16] JAIN M,ALSKAIF T,DEV S. Validating clustering frameworks for electric load demand profiles[J]. IEEE Transactions on Industrial Informatics,2021,17(12):8057-8065.
- [17] 陈谦,陈嘉雯,王苏颖,等. 基于节点日负荷曲线的深度嵌入式聚类及其改进方法对比研究[J]. 电力科学与技术学报,2023,38(1):130-137.  
CHEN Qian,CHEN Jiawen,WANG Suying,et al. Comparative

- study on deep embedded clustering and its improved methods based on node daily load curve[J]. *Journal of Electric Power Science and Technology*, 2023, 38(1):130-137.
- [18] 徐毅, 吴鸣, 李广玮, 等. 基于多维缩放的日负荷曲线聚类分析[J]. *电测与仪表*, 2022, 59(10):80-86.
- XU Yi, WU Ming, LI Guangwei, et al. Clustering analysis of daily load curve based on multi-dimensional scaling[J]. *Electrical Measurement & Instrumentation*, 2022, 59(10):80-86.
- [19] METS K, DEPUYDT F, DEVELDER C. Two-stage load pattern clustering using fast wavelet transformation[J]. *IEEE Transactions on Smart Grid*, 2016, 7(5):2250-2259.
- [20] KWAC J, FLORA J, RAJAGOPAL R. Lifestyle segmentation based on energy consumption data[J]. *IEEE Transactions on Smart Grid*, 2018, 9(4):2409-2418.
- [21] CHICCO G, NAPOLI R, PIGLIONE F, et al. Load pattern-based classification of electricity customers[J]. *IEEE Transactions on Power Systems*, 2004, 19(2):1232-1239.
- [22] 庞传军, 余建明, 冯长有, 等. 基于 LSTM 自动编码器的电力负荷聚类建模及特性分析[J]. *电力系统自动化*, 2020, 44(23):57-63.
- PANG Chuanjun, YU Jianming, FENG Changyou, et al. Clustering modeling and characteristic analysis of power load based on long-short-term memory auto-encoder[J]. *Automation of Electric Power Systems*, 2020, 44(23):57-63.
- [23] ZHONG S Y, TAM K S. A frequency domain approach to characterize and analyze load profiles[J]. *IEEE Transactions on Power Systems*, 2012, 27(2):857-865.
- [24] ZHONG S Y, TAM K S. Hierarchical classification of load profiles based on their characteristic attributes in frequency domain[J]. *IEEE Transactions on Power Systems*, 2015, 30(5):2434-2441.
- [25] LI H, WANG Z, HONG T Z, et al. Characterizing patterns and variability of building electric load profiles in time and frequency domains[J]. *Applied Energy*, 2021, 291:116721.
- [26] 刘思, 李林芝, 吴浩, 等. 基于特性指标降维的日负荷曲线聚类分析[J]. *电网技术*, 2016, 40(3):797-803.
- LIU Si, LI Linzhi, WU Hao, et al. Cluster analysis of daily load curves using load pattern indexes to reduce dimensions[J]. *Power System Technology*, 2016, 40(3):797-803.
- [27] 宋军英, 何聪, 李欣然, 等. 基于特征指标降维及熵权法的日负荷曲线聚类方法[J]. *电力系统自动化*, 2019, 43(20):65-72.
- SONG Junying, HE Cong, LI Xinran, et al. Daily load curve clustering method based on feature index dimension reduction and entropy weight method[J]. *Automation of Electric Power Systems*, 2019, 43(20):65-72.
- [28] HABEN S, SINGLETON C, GRINDROD P. Analysis and clustering of residential customers energy behavioral demand using smart meter data[J]. *IEEE Transactions on Smart Grid*, 2016, 7(1):136-144.
- [29] TEERARATKUL T, O'NEILL D, LALL S. Shape-based approach to household electric load curve clustering and prediction[J]. *IEEE Transactions on Smart Grid*, 2018, 9(5):5196-5206.
- [30] 徐胜蓝, 司曹明哲, 万灿, 等. 考虑双尺度相似性的负荷曲线集成谱聚类算法[J]. *电力系统自动化*, 2020, 44(22):152-160.
- XU Shenglan, SI Caomingzhe, WAN Can, et al. Ensemble spectral clustering algorithm for load profiles considering dual-scale similarities[J]. *Automation of Electric Power Systems*, 2020, 44(22):152-160.
- [31] XIANG Y, HONG J H, YANG Z Y, et al. Slope-based shape cluster method for smart metering load profiles[J]. *IEEE Transactions on Smart Grid*, 2020, 11(2):1809-1811.
- [32] 李阳, 刘友波, 刘俊勇, 等. 基于形态距离的日负荷数据自适应稳健聚类算法[J]. *中国电机工程学报*, 2019, 39(12):3409-3419.
- LI Yang, LIU Youbo, LIU Junyong, et al. Self-adaptive and robust clustering algorithm for daily load profiles based on morphological distance[J]. *Proceedings of the CSEE*, 2019, 39(12):3409-3419.
- [33] 林顺富, 田二伟, 符杨, 等. 基于信息熵分段聚合近似和谱聚类的负荷分类方法[J]. *中国电机工程学报*, 2017, 37(8):2242-2252.
- LIN Shunfu, TIAN Erwei, FU Yang, et al. Power load classification method based on information entropy piecewise aggregate approximation and spectral clustering[J]. *Proceedings of the CSEE*, 2017, 37(8):2242-2252.
- [34] 刘洋, 刘洋, 许立雄, 等. 计及数据类别不平衡的海量用户负荷典型特征高性能提取方法[J]. *中国电机工程学报*, 2019, 39(14):4093-4103.
- LIU Yang, LIU Yang, XU Lixiong, et al. A high performance extraction method for massive user load typical characteristics considering data class imbalance[J]. *Proceedings of the CSEE*, 2019, 39(14):4093-4103.
- [35] 李晓旭, 刘忠源, 武继杰, 等. 小样本图像分类的注意力全关系网络[J]. *计算机学报*, 2023, 46(2):371-384.
- LI Xiaoxu, LIU Zhongyuan, WU Jijie, et al. Total relation network with attention for few-shot image classification[J]. *Chinese Journal of Computers*, 2023, 46(2):371-384.
- [36] 金伟超, 张旭, 刘晟源, 等. 基于剪枝策略和密度峰值聚类的行业典型负荷曲线辨识[J]. *电力系统自动化*, 2021, 45(4):20-28.
- JIN Weichao, ZHANG Xu, LIU Shengyuan, et al. Identification of typical industrial power load curves based on pruning strategy and density peak clustering[J]. *Automation of Electric Power Systems*, 2021, 45(04):20-28.
- [37] 王潇笛, 刘俊勇, 刘友波, 等. 采用自适应分段聚合近似的典型负荷曲线形态聚类算法[J]. *电力系统自动化*, 2019, 43(1):110-118.
- WANG Xiaodi, LIU Junyong, LIU Youbo, et al. Shape clustering algorithm of typical load curves based on adaptive piecewise aggregate approximation[J]. *Automation of Electric Power Systems*, 2019, 43(1):110-118.

[38] 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(6):2104-2113.  
BAI Yaling, ZHOU Yatong, LIU Jun. Clustering analysis of daily load curve based on deep convolution embedding clustering[J]. Power System Technology, 2022,46(6):2104-2113.

[39] 宋军英,崔益伟,李欣然,等. 基于欧氏动态时间弯曲距离与熵权法的负荷曲线聚类方法[J]. 电力系统自动化,2020,44(15):87-94.  
SONG Junying, CUI Yiwei, LI Xinran, et al. Load curve clustering method based on Euclidean dynamic time warping distance and entropy weight[J]. Automation of Electric Power Systems, 2020,44(15):87-94.

[40] ZHANG K Q, ZHU H, GUO S M. Dependency analysis and improved parameter estimation for dynamic composite load modeling[J]. IEEE Transactions on Power Systems, 2017, 32(4):3287-3297.

[41] CHICCO G. Overview and performance assessment of the clustering methods for electrical load pattern grouping[J]. Energy,

2012,42(1):68-80.

[42] WANG Y, JIA M S, GAO N, et al. Federated clustering for electricity consumption pattern extraction [J]. IEEE Transactions on Smart Grid, 2022, 13(3):2425-2439.

[43] TRINDADE, A. Electricity load diagrams 20112014[DB/OL]. UCI Machine Learning Repository. (2015) [2022-04-28]. <https://doi.org/10.24432/C58C86>.

作者简介:



郭文熙

郭文熙(2000),男,博士在读,研究方向为广义负荷的智能认知(E-mail:12210018@zju.edu.cn);

李知艺(1989),男,博士,研究员,博士生导师,研究方向为综合能源电力系统规划及运行、信息安全及隐私保护;

尹建兵(1975),男,硕士,高级工程师,从事电网规划工作。

## Clustering analysis of load curve based on time series density-changing processing

GUO Wenxi<sup>1</sup>, LI Zhiyi<sup>1</sup>, YIN Jianbing<sup>2</sup>, CHEN Lin<sup>2</sup>, JU Ping<sup>1,3</sup>

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China;

2. State Grid Hangzhou Power Supply Company of Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China;

3. College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China)

**Abstract:** Load curve clustering constitutes the foundational methodology for the analysis of customer load characteristics, enabling the extraction of typical power consumption patterns from a substantial volume of load data. Understanding these patterns is pivotal for applications including demand response, tariff design and power grid planning. In light of the inadequacies of existing clustering methodologies, specifically their insufficient consideration of the unique characteristics of load time periods, an advanced clustering technique based on time series density-changing processing is introduced in this study. The proposed method commences with the application of linear interpolation to augment the density of data points during critical periods, including peak, trough and ramp-up periods, thereby accentuating and amplifying their influence within the clustering framework. Concurrently, the technique incorporates the adaptive piecewise aggregate approximation (APPA) for dimensionality reduction to mitigate the density of superfluous data. Subsequently, a comprehensive index, formulated through the integration of Euclidean distance and correlation distance, is employed to conduct  $k$ -medoids clustering analysis of the load curves. The effectiveness of the proposed methodology is validated utilizing actual residential customer data from the UCI dataset. The results of these empirical investigations affirm that the method significantly enhances the efficacy of load clustering, thereby providing an authentic representation of the power consumption characteristics of residential customers.

**Keywords:** density-changing; load clustering; critical periods; ordinary periods; interpolation; adaptive piecewise aggregate approximation (APPA)

(编辑 方晶)