



DOI:10.12404/j.issn.1671-1815.2403113

引用格式:赵振宇,郭丽宣.基于特征优选与自适应三支密度峰值法的多元负荷聚类及用能行为刻画[J].科学技术与工程,2025,25(5):1944-1953.

Zhao Zhenyu, Guo Lixuan. Multi-load clustering and energy consumption behavior characterization based on feature selection and three-way adaptive density peak algorithm[J]. Science Technology and Engineering, 2025, 25(5): 1944-1953.

基于特征优选与自适应三支密度峰值法的 多元负荷聚类及用能行为刻画

赵振宇,郭丽宣

(华北电力大学经济与管理学院,北京 102206)

摘要 随着向新型能源体系的转型加速,亟待开展对多元负荷用户的复杂用能特性分析的深入研究。提出了一种综合考虑电、冷、热多元负荷耦合特性的用户用能特性标签库构建技术及用户画像方法。首先运用快速相关性滤波算法剔除高冗余低相关特征,并通过随机森林和递归式特征消除算法精选出具有强区分能力的用能特征。在聚类阶段,改进的自适应三支密度峰值聚类算法(three-way adaptive density peak clustering,3W-ADPC)通过结合自适应近邻搜索和三支聚类算法提升负荷聚类效果。实证结果表明,所提方法具备在计算效率和聚类精度上的双重优势,能够精准揭示多元负荷用户综合用能特性和深层次信息,证实所提方法在多元负荷用户行为研究中的实用价值。

关键词 负荷聚类;多元负荷;用能行为特性;特征优选;用户画像

中图分类号 TM714; 文献标志码 A

Multi-load Clustering and Energy Consumption Behavior Characterization Based on Feature Selection and Three-way Adaptive Density Peak Algorithm

ZHAO Zhen-yu, GUO Li-xuan

(Economic and Management College, North China Electric Power University, Beijing 102206, China)

[Abstract] With the acceleration of the transition to new energy systems, it is urgent to carry out in-depth research on the complex energy characteristics of multi-load users. A technology of constructing user energy characteristic label library and a user portrait method were proposed, which comprehensively considered the coupling characteristics of electric, cold and thermal multiple loads. Firstly, the high redundancy and low correlation features were eliminated by the fast correlation filtering algorithm, and the features with strong distinguishing ability were selected by the random forest and recursive feature elimination algorithm. In the clustering stage, the improved three-way adaptive density peak clustering (3W-ADPC) algorithm improved the load clustering effect by combining the adaptive neighbor search and the three-branch clustering algorithm. The empirical results show that the proposed method has dual advantages in computational efficiency and clustering accuracy, and can accurately reveal the comprehensive energy use characteristics and deep information of multi-load users, which confirms the practical value of the proposed method in the study of multi-load users' behavior.

[Keywords] load clustering; multiple loads; energy use behavior characteristics; feature selection; user portrait

近年来,随着新型电力系统及能源互联网建设进程的推进,能源系统内冷、热、电等多种能源供应形式间耦合加深,用户能源消费形态日益多元化和复杂化,给系统掌握用户用能模式、制定需求响应策略造成了困难^[1]。在此背景下,深入探索多元负荷用户的用能特性与行为模式成为当前电力系统研究的一项重要课题。此外,得益于智能量测设备和大数据技术的进步,使得电网各环节紧密互联,用户的多元用能数据可以被实时采集并有效存储。

基于这些数据能够深入挖掘用户用能特性与行为偏好,开展用户用能行为精细化建模和画像,有助于电力企业准确掌握用户用能规律、感知用户需求、实现用户细分、提供差异化服务,提升企业服务水平与市场竞争力^[2]。

用户画像技术利用数据分析手段提炼出用户特征标签,是建立在一系列真实数据之上的用户虚拟代表模型^[3-4]。近年来,随着机器学习技术的革新与负荷数据的累积,用户画像技术在能源电力领

收稿日期:2024-04-26 修订日期:2024-11-20

基金项目:北京市自然科学基金(8232013)

第一作者:赵振宇(1969—),男,满族,辽宁丹东人,博士,教授。研究方向:电力工程项目管理。E-mail:zhaozhenyuxm@263.net。

投稿网址:www.stae.com.cn

域的研究和应用逐渐兴起。现有研究与应用可分为两类,一类是架构型研究,主要针对用户画像技术框架进行研究,即基于分析对象特性提取用户属性标签,构建画像模型,侧重于分析标签生成^[5-6]、画像刻画方法^[7-8];另一类是用户画像分类模型研究^[9-11],即通过对用户负荷数据的分析和聚类算法的应用,识别不同类型的用户用能行为模式,进而构建多类别用户画像,如文献[9]提出了一种结合谱聚类与集成强基分类器的用户日负荷曲线分类算法,基于分类结果构建用户用电行为画像并完成用户画像分类;文献[10]考虑用户对时空电价的响应度建立需求响应特征指标,采用 K-means 划分用户类别,运用特征选择与搜索策略提取关键指标,刻画各类用户互动潜力画像。但现有研究多集中于单一的电力消耗行为分析,较少关注到综合能源用户如冷热负荷等多种能源形式的用能特性。部分文献虽对多元负荷行为有所分析,但其用能特征均由主观选取,未考虑特征选取的有效性和特征间存在的冗余度,且缺乏对多元负荷间耦合关系特性的深入分析^[12-14]。

用能特性标签的选取对于构建精确的用户用能画像至关重要,其目标在于筛选出既精简又能有效体现用户用能行为本质的关键特征子集^[15]。因此,需采取有效的评估指标与搜索策略来确保特征子集的可靠性。现有研究^[15-18]通常借助相关系数或互信息度量特征间冗余度,使用信息增益判断特征重要性,但这些方法各有局限,如难以捕捉复杂非线性关系、处理高维数据效率低、易受噪声影响导致过拟合等问题。因此,亟须寻求更高效且鲁棒性更强的特征选择方法以确保用户用能画像构建的可靠性和准确性。

针对上述研究现状,现构建综合考虑电、冷、热多元负荷耦合特性的用户用能特性标签库,并提出一种结合两阶段特征优选与改进的自适应三支密度峰值聚类算法(three-way adaptive density peak clustering, 3W-ADPC)的用户画像方法。首先,从用户用能数据中提取多维度用能行为特性,构建用户用能画像标签库。接着采用快速相关性滤波算法剔除冗余特性,在此基础上结合随机森林和交叉验证的递归式特征消除算法优选出能够最大化区分用户用能行为模式的特征集。针对传统聚类方法的不足,提出的 3W-ADPC 算法通过引入加权皮尔逊距离和自适应 k 近邻改进了参数的计算,并运用三支聚类解决样本误分类问题,提升聚类精准度和鲁棒性。最后,通过雷达图直观展示各类用户用能特性评价量化结果。利用真实城市负荷数据进行

实例验证,以证明该方法能有效揭示多元负荷用户的综合用能特性,有助于电力企业精细化管理、精准定制需求响应策略以及差异化服务,提高用户用能行为的理解和管理水平。

1 多元负荷用户用能特性标签提取

通过对国内外能源用户用能行为特性的深入调研和梳理,围绕电、冷、热负荷水平、行为规律、互动能力等方面,并综合考虑多类型负荷间耦合特性,提取反映多元负荷用户用能特点的画像标签属性。此外,考虑能源消耗数据具有显著的随时间变化规律,在日、月、年等不同时间尺度上选取相应的标签,构建用能行为画像标签库,从多元维度和多时间尺度细致刻画用户用能行为模式,建立精细完整的多元负荷用户综合用能行为画像。

1.1 用电负荷水平标签

用电负荷水平标签包括反映用户负荷消纳水平与全天用电特性的特征指标,如日均负荷、日负荷率等,具体指标如表 1 所示。

表 1 用电负荷水平标签

一级标签	二级标签	
	短期标签	中长期标签
用电负荷水平	工作日平均负荷	月平均负荷
	非工作日平均负荷	月负荷率
	日负荷率	年丰(枯)期平均月负荷
	日峰(谷)时负荷率	月峰(谷)期平均日负荷

1.2 用电行为规律标签

用电行为标签包括表述负荷波动程度的特征指标如日峰谷差率等,反映用电习惯的特征指标如日峰值负荷发生时间等,以及反映异常用电行为的特征指标如异常用电数量等,具体指标如表 2 所示。

表 2 用电行为规律标签

一级标签	二级标签	
	短期标签	中长期标签
用电行为规律	日峰谷差率	年、月峰谷差率
	日峰值负荷发生时间	年丰期负荷发生时期
	峰值负荷利用小时数	年丰期负荷利用月数
		月峰值负荷发生时期
		月峰值负荷利用天数

1.3 用电互动能力标签

用电互动能力标签包括反映用户需求响应潜力的特征指标如需求响应潜力熵、基准负荷偏差,反映用户调峰能力的特征指标如削峰潜力^[10,19],如表 3 所示。

表3 用电互动能力标签

Table 3 Electricity interaction capability labels		
一级标签	二级标签	
	短期标签	中长期标签
用电互动能力	需求响应潜力熵	平均(最大)需求响应潜力熵
	基准负荷偏差	平均基准负荷偏差
	削峰潜力	平均(最大)削峰潜力

1.4 用冷(热)特性标签

由于冷、热负荷均具有一定的延时性,相较于电负荷对需求响应不敏感,因此用冷(热)特性标签主要围绕负荷水平与行为规律两方面选取。参考用电负荷水平与用电行为规律标签并结合冷(热)能实际消耗情况,选取标签如表4所示。

表4 用冷(热)标签

Table 4 Cold (heat) energy consumption labels		
一级标签	二级标签	
	短期标签	中长期标签
用冷(热)特性	工作日平均负荷	月平均负荷
	非工作日平均负荷	月负荷率
	日负荷率	年丰(枯)期平均月负荷
	日峰(谷)时负荷率	月峰(谷)期平均日负荷
	日峰谷差率	年、月峰谷差率
	日峰值负荷发生时间	年丰期负荷发生时期
	最大负荷利用小时数	年丰期负荷利用月数
		月峰值负荷发生时期 月峰值负荷利用天数

1.5 多元负荷耦合特性标签

相比传统的电力系统,综合能源系统发展下用户侧冷、热、电等多种负荷间的联系日益密切,且不同类型用户的不同能源形式间的耦合关系存在差异,因此在刻画用户用能行为特性时,需要对多元负荷间的耦合关系与耦合强度的差异进行挖掘。统一信息系数法(uniform information coefficient, UIC)是一种新型特征过滤技术,相较于传统皮尔逊和斯皮尔曼相关系数,它能够同时捕获变量间的线性与非线性关联,相比最大信息系数法,UIC在相关性分析中有更好的抗噪声能力和更低的计算成本,更适用于大规模负荷数据集的分析场景^[20]。

对于特征向量 $\mathbf{X} = [x_1, x_2, \dots, x_n]$ 与 $\mathbf{Y} = [y_1, y_2, \dots, y_n]$, n 为序列长度,则 \mathbf{X} 与 \mathbf{Y} 之间互信息系数的度量公式为

$$I_{MI}(\mathbf{X}; \mathbf{Y}) = \sum_{x \in \mathbf{X}} \sum_{y \in \mathbf{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

式(1)中: $I_{MI}(\mathbf{X}; \mathbf{Y})$ 为 \mathbf{X} 与 \mathbf{Y} 间的互信息系数; $p(x, y)$ 为 \mathbf{X} 和 \mathbf{Y} 的联合概率密度函数; $p(x)$ 和 $p(y)$ 分别为 \mathbf{X} 和 \mathbf{Y} 的边缘概率密度函数。

将 \mathbf{X} 与 \mathbf{Y} 按式(2)统一均匀划分为若干段。

$$\begin{cases} l_x = \frac{x_{\max} - x_{\min}}{a}, & 2 \leq a \leq 1 + \frac{n^{0.6}}{2} \\ l_y = \frac{y_{\max} - y_{\min}}{b}, & 2 \leq b \leq 1 + \frac{n^{0.6}}{2} \end{cases} \quad (2)$$

式(2)中: l_x 与 l_y 分别为 \mathbf{X} 与 \mathbf{Y} 的分区单元长度; x_{\max} 与 x_{\min} 、 y_{\max} 与 y_{\min} 分别为特征 \mathbf{X} 与 \mathbf{Y} 取值范围的最大、最小值; a 与 b 为分段数指标,用于划分 \mathbf{X} 与 \mathbf{Y} 的区间; $n^{0.6}$ 为分区网格,通常将其设为数据量的0.6次方。基于以上分区方法,UIC计算模型如式(3)所示。

$$I_{UIC}(\mathbf{X}; \mathbf{Y}) = \frac{I_{MI}(\mathbf{X}; \mathbf{Y})}{\log_2[\min(a, b)]} \quad (3)$$

式(3)中: $I_{UIC}(\mathbf{X}; \mathbf{Y})$ 为 \mathbf{X} 与 \mathbf{Y} 间的统一信息系数; $\min(a, b)$ 为 a 与 b 中的最小值。

计算不同时间尺度下冷、热、电三类负荷的统一信息系数,作为多元负荷耦合特性标签,如表5所示。

表5 多元负荷耦合特性标签

Table 5 Labels for coupling characteristics between multiple loads

一级标签	二级标签	
	短期标签	中长期标签
多元负荷耦合特性	电-热负荷短期信息系数	电-热负荷长期信息系数
	电-冷负荷短期信息系数	电-冷负荷长期信息系数
	冷-热负荷短期信息系数	冷-热负荷长期信息系数

2 多元负荷用能特性两阶段优选策略

准确识别用户用能行为模式的关键在于选择具有较强区分力的特征集合,低冗余、强相关的特征在聚类中有着更重要的区分能力,此外,不同排列组合方式得到的不同特征子集对划分行为模式的区分能力也有所差异。提出一种两阶段特征优选策略:第一阶段采用快速相关性滤波算法(fast correlation-based filter, FCBF)衡量特征的冗余度与相关性,实现对原始特征全集的初筛;第二阶段基于随机森林(random forest, RF)与交叉验证的递归式特征消除(recursive feature elimination with cross-validation, RFECV)算法相结合完成对特征重要度的评估与进一步择优,最终得到表征多元负荷用户多维度用能行为特性的最优特征子集。

2.1 基于FCBF算法的特征集初筛

快速相关性滤波算法(FCBF)通过引入信息熵理论中的对称不确定性(symmetrical uncertainty, SU)衡量特征与类别之间和不同特征之间的相关性^[10]。考虑到特征变量的统计学意义,需先将特征值归一化和区间离散化得到概率分布,进而完成对

相关性度量指标的计算。依据信息熵理论,两个变量间的相关性可通过信息增益来衡量,即一个变量对另一变量不确定性降低的有效程度,数值上表示为引入该变量前后另一个变量的信息熵变化差值,信息增益越大,表明两个变量间冗余度越高。但信息增益存在偏向信息量更大的特征这一偏差, SU 基于这一偏差进行了补偿,且将度量结果规范化至 $[0, 1]$ 区间内, $SU(\mathbf{X}, \mathbf{Y})$ 取值越大表明两个变量间相关性越大,其表达式为

$$SU(\mathbf{X}, \mathbf{Y}) = \frac{2I(\mathbf{X}, \mathbf{Y})}{H(\mathbf{X}) + H(\mathbf{Y})} \quad (4)$$

式(4)中: $I(\mathbf{X}, \mathbf{Y})$ 为随机变量 \mathbf{X} 与 \mathbf{Y} 间的联合信息增益; $H(\mathbf{X})$ 、 $H(\mathbf{Y})$ 分别为随机变量 \mathbf{X} 与 \mathbf{Y} 的信息熵。信息增益与信息熵的计算参考文献[10]中公式。

设定共有 C 类用户,影响用户用能行为的特征集为 $\mathbf{F} = (f_{ij})_{N \times M}$, 则特征 f_i 与特征 f_j 间的对称不确定性 $SU(f_i, f_j)$ 和特征 f_i 与用户类别 c 间的对称不确定性 $SU(f_i, c)$ 的表达式为

$$\begin{cases} SU(f_i, f_j) = 2 \frac{I(f_i, f_j)}{H(f_i) + H(f_j)} \\ SU(f_i, c) = 2 \frac{I(f_i, c)}{H(f_i) + H(c)} \end{cases} \quad (5)$$

为确保特征与类别间关联度最大且特征之间冗余度最小,利用 FCBF 算法对特征全集进行如下筛选。

(1) 筛除弱相关的特征: 设置阈值 σ , 若 $SU(f_i, c) < \sigma$, 表明特征 f_i 与类别间相关性较弱, 则筛除特征 f_i 。

(2) 筛除高冗余的特征: 计算余下各特征间的冗余度, 若 $SU(f_i, f_j) > SU(f_i, c)$, 表明特征 f_j 与其他特征间的冗余度较高, 则筛除特征 f_j , 保留余下特征作为第一阶段筛选后的特征集。

2.2 基于 RF-RFECV 算法的特征集择优

RF-RFECV 算法是一种基于随机森林算法和交叉验证递归特征消除策略的特征优选算法, 其利用 RF 中决策树模型的优势对特征重要性进行度量, 并引入 RFECV 算法迭代训练得到最优特征子集。

RF 算法结合集成学习理论与随机子空间方法, 由一定数量的决策树构成, 能高效处理复杂、缺失值数据, 具有高鲁棒性。RF 在决策树训练时通过基尼指数(Gini)评估特征重要性, 进而选取每个节点的最优划分特征。现有一个样本集 D , 假设其可被划分为 C 个类别, 每个样本具有 F 个特征, 其 RF 模型由 I 棵决策树构成, 设第 i 棵决策树节点 q 处的样本点属于第 c 类的概率为 $P_{q,c}$, 则第 i 棵决策树节点

q 的基尼指数 $Gini_q^{(i)}$ 为

$$Gini_q^{(i)} = \sum_{c=1}^C P_{q,c}^{(i)} [1 - P_{q,c}^{(i)}] = 1 - \sum_{c=1}^C [P_{q,c}^{(i)}]^2 \quad (6)$$

在 RF 模型中, 若样本集经过某特征划分后基尼指数的减少量越大, 表明该特征对于提升特征集纯度的作用越大, 即该特征越重要。因此, 对于特征 f_j , 其重要性评分的数值可以通过计算经 f_j 划分前后样本集 Gini 指数的变化量来评估, 表达式为

$$\begin{aligned} VIM_{f_j} &= \sum_{i=1}^I VIM_{f_j}^{Gini(i)} = \sum_{i=1}^I \sum_{q \in Q} VIM_{f_j, q}^{Gini(i)} \\ &= \sum_{i=1}^I \sum_{q \in Q} [Gini_q^{(i)} - Gini_{q-left}^{(i)} - Gini_{q-right}^{(i)}] \end{aligned} \quad (7)$$

式(7)中: $VIM_{f_j}^{Gini(i)}$ 为特征 f_j 在第 i 棵决策树的重要性评分; $VIM_{f_j, q}^{Gini(i)}$ 为特征 f_j 在第 i 棵决策树节点 q 的重要性评分; Q 为第 i 棵决策树中包含特征 f_j 的节点集合; $Gini_{q-left}^{(i)}$ 、 $Gini_{q-right}^{(i)}$ 分别为节点 q 分枝后的左节点、右节点的基尼指数。最后将求得的重要性评分进行归一化处理。

基于 RF 模型得出的特征重要性评估结果, 采用 RFECV 方法^[21]对特征集进行迭代训练: 每轮迭代都记录当前特征集的重要性评分, 并消除特征重要性较低的特征形成新的特征集进行下一轮训练, 迭代直至特征集维度缩减至最小阈值; 为增强选择结果的鲁棒性与可靠性, 将所有的特征子集进行 10 折交叉验证, 将重要性评分最高的特征子集作为最优特征子集, 用于划分用户类别和构建用户画像。

3 改进的 3W-ADPC 多元负荷用户群体聚类方法

基于两阶段特征优选结果, 继续对多元负荷用户的典型用能行为模式进行聚类分析, 可以有效地将具有相似用能行为特性的用户归为一类, 从而了解用户实际能源需求, 帮助电力企业实现对用户的高效分类管理与统一调度, 提升整体运营管理效率和服务质量。为此, 提出了一种改进的自适应三支密度峰值聚类算法(three-way adaptive density peak clustering, 3W-ADPC); 针对传统密度峰值聚类(density peak clustering, DPC)算法中存在的主观性与错误传播等缺陷, 引入加权皮尔逊距离与自适应 k 近邻(adaptive k -nearest neighbor, A-kNN)对局部密度和截断距离进行优化, 并采用三支聚类(three-way, 3W)算法解决误分配问题。

3.1 密度峰值聚类算法

DPC 是一种基于密度和距离的聚类算法, 其选

取聚类中心是基于两点假设:类簇中心的局部密度高于邻居点的密度、类簇中心之间距离较远。对于每个数据点 x_i 都需计算两个参数值:该数据点的局部密度 ρ_i 和相对距离 δ_i 。

(1)局部密度 ρ_i 。局部密度有两种计算方法:针对大规模数据集使用截断核,小规模数据集使用高斯核。对于数据集 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$,截断核和高斯核的计算公式分别如下。

$$\rho_i = \sum_j \chi(d_{ij} - d_c), \chi(x) = \begin{cases} 1, & x < 0 \\ 0, & \text{其他} \end{cases} \quad (8)$$

$$\rho_i = \sum_{j \neq i} \exp \left[- \left(\frac{d_{ij}}{d_c} \right)^2 \right] \quad (9)$$

式中: d_{ij} 为 x_i 和 x_j 间的欧式距离; d_c 为全局截断距离且 $d_c > 0$,通常取所有样本点间距离前 2% 处的值。

(2)相对距离 δ_i 。定义每个样本点 x_i 的相对距离 δ_i 为

$$\delta_i = \begin{cases} \min_{j: \rho_j > \rho_i} (d_{ij}), & \rho_i < \max(\rho) \\ \max_j (d_{ij}), & \rho_i = \max(\rho) \end{cases} \quad (10)$$

由式(10)可知,若 x_i 不是数据集中局部密度最大的点,其相对距离就是与局部密度高于自身的点之间的最小距离,当 x_i 为局部密度最大的点时,则选取所有数据点中与之最远的点之间的距离。

DPC 以 ρ_i 为横轴、 δ_i 为纵轴绘制决策图,将出现在坐标系中右上部分的数据点(即同时具有较高的局部密度 ρ_i 和较大的相对距离 δ_i) 标识为类簇中心。对于非类簇中心的剩余数据点的分配,首先按局部密度降序排序,然后依次将数据点分配到局部密度高于自身且距离最近的数据点的所属类簇中。

传统的 DPC 算法存在以下缺陷。

(1)局部密度计算方法过于简单,忽略了数据的局部结构和分布特点,当数据集密度分布不均匀时无法准确识别稀疏簇的聚类中心。

(2)聚类结果对于截断距离参数的设置异常敏感,导致应用于类簇分布情况较复杂的能源用户的聚类效果不佳。

(3)数据点分配策略容错能力差,若某数据点分配错误,则后续分配会进一步放大误差,对聚类结果产生严重影响。

3.2 三支聚类算法

传统聚类方法使用单一集合表示聚类结果,即对于给定的数据集 $V = \{x_1, x_2, \dots, x_n\}$,其聚类结果表示为具有 k 个类簇的集合 $C = \{C_1, C_2, \dots, C_k\}$ 。三支聚类算法^[22]基于传统聚类概念进行模糊扩展,

将类簇划分为核心和边缘部分,通过一个区间集合来表示聚类结果,则有

$$C_i = [C_i, \bar{C}_i] \\ = [\text{POS}(C_i), \text{POS}(C_i) \cup \text{BND}(C_i)] \quad (11)$$

式(11)中: C_i 为 C_i 下限, $\underline{C}_i = \text{POS}(C_i)$; \bar{C}_i 为 C_i 上限, $\bar{C}_i = \text{POS}(C_i) \cup \text{BND}(C_i)$; 另设 $\text{NEG}(C_i) = V - \bar{C}_i$,由此将一个类簇划分为 3 个区域:正域 $\text{POS}(C_i)$ 、边界域 $\text{BND}(C_i)$ 和负域 $\text{NEG}(C_i)$; $\text{POS}(C_i)$ 为类簇核心域,其中的点一定属于当前簇; $\text{BND}(C_i)$ 为边缘域,其中的点分布在类簇的边缘区域,有可能属于其他类簇; $\text{NEG}(C_i)$ 中包含了所有一定不属于当前簇的点。规定这 3 个区域不相交且覆盖整个集合,各类簇中一定存在正域且不同类簇正域不重叠,所有类簇的上限需覆盖整个集合,如式(12)所示。

$$\begin{cases} \text{POS}(C_i) \cap \text{BND}(C_i) = \emptyset \\ \text{POS}(C_i) \cap \text{NEG}(C_i) = \emptyset \\ \text{BND}(C_i) \cap \text{NEG}(C_i) = \emptyset \\ \text{POS}(C_i) \cup \text{BND}(C_i) \cup \text{NEG}(C_i) = V \\ \text{POS}(C_i) \neq \emptyset \\ \text{POS}(C_i) \cap \text{POS}(C_j) = \emptyset, i \neq j \\ \bigcup_{i=1}^n (\text{POS}(C_i) \cup \text{BND}(C_i)) = V \end{cases} \quad (12)$$

3.3 改进的自适应三支密度峰值聚类算法

针对传统 DPC 的缺陷,提出了一种改进的自适应三支密度峰值聚类算法(3W-ADPC)。一方面,针对参数计算问题,引入加权皮尔逊距离和自适应 k 近邻改进密度度量方法,避免手动选择截断距离;另一方面,针对错误传播问题,引入三支聚类算法将边界点逐步归入正域,以此作为非中心点的分配策略。

3.3.1 改进局部密度

现有大多数 DPC 的改进思想是使用 k 近邻理论代替截断距离来估计局部密度,可以平衡稀疏区域与密集区域的密度,进而易于找到稀疏类簇的聚类中心,其中常见形式是由 k 近邻的平均距离来计算局部密度。但是原始 k 近邻理论仍存在需要手动设置参数 k 这一缺陷,且算法结果受 k 值影响较大,当 k 值越大时,目标数据点周围被计入的数据点越多,可能意外地计入离群点,影响密度测量结果真实性。鉴于此,引入自适应 k 近邻理论改进局部密度的计算,该算法能够根据样本局部稠密情况自适应地获得数据点的近邻数量 k ,从而计算样本的局部密度,具体步骤如下。

步骤 1 初始化每个数据点的逆近邻。对于数据集 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 中的任意数据点 x_i, x_j ,若

x_j 是 x_i 的 k 近邻点之一,则称 x_i 是样本点 x_j 的逆近邻,即

$$\text{RNN}_k(x_j) = \{x_i | x_i \in \text{KNN}_k(x_j)\} \quad (13)$$

式(13)中: $x_i \in \text{KNN}_k(x_j)$ 为数据点 x_i 是 x_j 的 k 近邻,即 x_i 为距离 x_j 最近的 k 个数据点之一。

步骤2 迭代获取每个数据点的自适应 k 值。设定从搜索指数 $r = 1$ 开始迭代,每次迭代搜索每个数据点的 KNN_r 和 RNN_r ,当逆近邻个数为 0 的数据点的数量连续三次迭代不变时,认为算法达到自然稳定状态,停止迭代,每个数据点在迭代停止时的逆近邻个数即为数据点的自适应 k 值。

步骤3 基于自适应 k 值搜索各数据点的自适应 k 邻域,并计算局部密度,计算式为

$$\rho_i = \exp \left[- \frac{\sum_{x_j \in \text{ANN}_k(x_i)} d(x_i, x_j)}{k} \right] \quad (14)$$

式(14)中: $\text{ANN}_k(x_i)$ 为数据点 x_i 以自适应 k 值为参数的自适应 k 邻域; $d(x_i, x_j)$ 为数据点 x_i 和数据点 x_j 间的距离,这里采用加权皮尔逊距离来计算^[23]。

为进一步识别局部信息,将 x_i 的局部密度除以其自适应 k 近邻中的最大局部密度得到相对局部密度,可以更好地适应差异化的数据分布,从而得到更合理的聚类中心。改进后的相对局部密度定义为

$$\rho'_i = \frac{\rho_i}{\max[\rho_j | x_j \in A - \text{kNN}(x_i)]} \quad (15)$$

3.3.2 改进的 3W-ADPC 算法流程

基于以上理论概念,所提 3W-ADPC 算法的具体流程如下。

(1)通过自适应 k 近邻搜索算法获得各样本点自适应 k 近邻域(A-kNN),计算样本点相对距离与相对局部密度,绘制决策图并选择类簇中心,将剩余的样本点分配到最近的类簇中心,实现样本点的初步标记。

(2)检测潜在边界样本点。边界样本应同时具有较低的局部密度和相对局部密度,采用局部密度的四分位数作为阈值检测潜在边界样本。

(3)基于潜在边界样本的 A-kNN 进一步将潜在边界样本归入正域或边界域:如果一个潜在边界样本的 A-kNN 都属于一个簇,则该样本被分配到该簇的正区域;如果 A-kNN 属于不同的簇,那么样本被分配到这些簇的边界区域。对于非潜在边界样本,则将其继续保留在原所属初始类簇的正域中。

4 实例分析

采用某城市区域内 707 个能源用户的电、冷、热负荷实际数据进行实例分析,数据集采集了 2022 年

1 月 1 日 00:00—2022 年 12 月 31 日 23:45 的负荷曲线数据,每 15min 采集一次,全年共计 35 040 个采样点,用户按类型被划分为 7 类并附带类别标识。实例分析分为 3 个步骤:首先是多元负荷用户用能行为特性分析,对提出的两阶段特征优选策略进行验证;其次,运用 3W-ADPC 算法进行多元负荷用户群体聚类分析,评估其聚类效果及算法性能;最后,基于上述行为特性和聚类结果,对不同用户群体进行多元负荷用能画像分析,形成各用户群体的多维度综合用能行为画像。

4.1 多元负荷用户用能特性分析

依据构建的多元负荷用户用能特性标签库,基于实例数据集对各标签进行量化计算,形成原始特征集。采用两阶段特征优选策略对原始特征集进行筛选与择优。参考传统特征选择方法,设定优选后的最优特征子集保留 30%~40% 的特征数量。

4.1.1 特征相关度、冗余度初筛

第一阶段根据 FCBF 原则,设定阈值 σ 为 0.5,计算标签与用户类别间相关度以及各标签间的冗余度,计算结果如图 1 和图 2 所示,由于篇幅限制只展示部分计算结果。

由图 1 中各特征标签与用户类别间相关度计算结果可知,电-月峰谷差率、电-最大需求响应潜力熵等标签与类别间的相关度较低,故予以剔除;由图 2 中各特征间冗余度计算结果可知,电-非工作日平均负荷、电-年丰期平均月负荷等特征全局相关度低、与其他特征间冗余度高,故依据 FCBF 原则予以剔除。综上所述,第一阶段特征选择过程筛选了原始特征集中约 1/3 的低相关、高冗余特征。

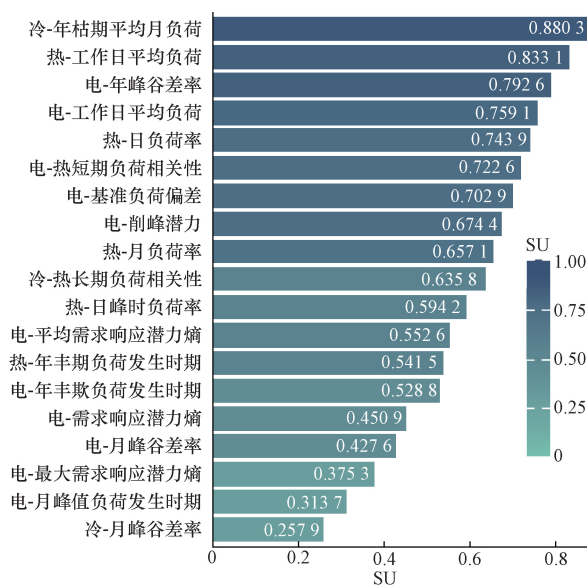


图 1 标签与用户类别间的相关度

Fig. 1 Correlation between labels and user categories

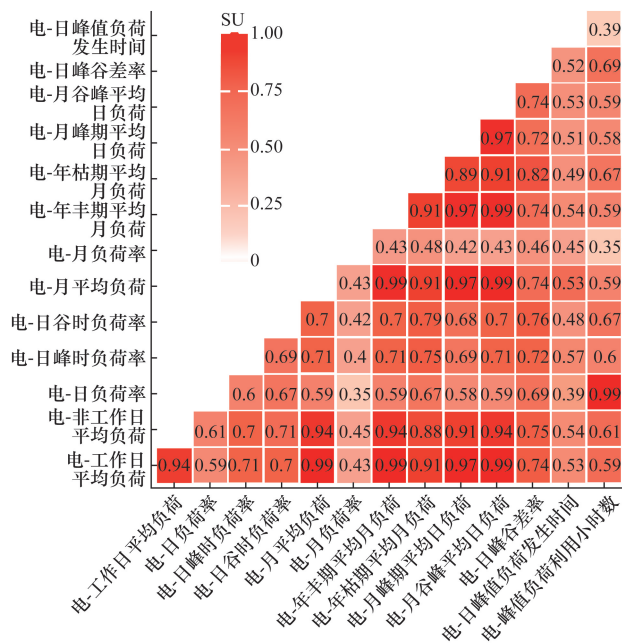


图2 标签间的冗余度

Fig. 2 Redundancy between labels

4.1.2 特征重要度评价分析

在第二阶段特征优选中,将第一阶段筛选后的特征集输入 RF 模型中进行特征重要性评分,并基于评分结果采用 RFECV 方法对特征集迭代训练并交叉验证,将重要性评分最高的特征子集作为最优特征子集。设定 RF 模型中决策树个数为 100,节点分裂过程中抽取的特征个数为总特征数的算术平方根。最终经过两阶段优选过程获得了 19 维的特征标签集合,这些特征涵盖了多维度 and 多个时间尺度下的用户用能特性,如表 6 所示。

表 6 两阶段特征优选策略最优特征子集

Table 6 The optimal feature subset of the two-stage feature optimization strategy

一级标签	二级标签	
	短期标签	中长期标签
用电负荷水平	工作日平均负荷 日负荷率	月负荷率
用电行为规律	—	年丰期负荷发生时期 年丰期负荷利用月数
用电互动行为	基准负荷偏差 削峰潜力	平均需求响应潜力熵
用冷特性	—	月负荷率 年枯期平均月负荷
用热特性	工作日平均负荷 日负荷率 日峰时负荷率	月负荷率 年丰期负荷发生时期
多元负荷耦合特性	电-热短期负荷相关性 电-冷短期负荷相关性 冷-热短期负荷相关性	冷-热长期负荷相关性

为验证所提两阶段特征优选策略的有效性,选取了 4 个特征集进行对比分析:原始特征集、基于最大相关最小冗余准则(maximal relevance and minimal redundancy, mRMR)^[15]选择的特征集、基于 RF 选择的特征集和经所提两阶段特征优选策略选择的特征集。以特征选择耗时和聚类准确率为评价标准,为准确体现不同特征集的分类性能,不受聚类方法本身干扰,均采用 DPC 进行聚类,设置截断距离为 2%,不同特征集的分类性能结果如表 7 所示。

实验结果显示,使用所构建的原始特征集进行用户分类的聚类准确率达到 88.97%,显著高于基于传统日负荷曲线的 79.07%,验证了通过特性指标替代功率向量进行聚类的可行性与实用性。3 种特征选择方法所得特征集的聚类准确率均优于传统负荷曲线及原始特征全集,体现了特征选择对于提高聚类效果和处理大型用能负荷数据集的重要性。

表 7 不同特征集分类性能对比

Table 7 Comparison of classification performance of different feature sets

特征集	特征集 维度	聚类 准确率/%	特征选择 过程耗时/s
典型日负荷曲线	96	79.07	—
原始特征集	76	88.97	—
mRMR 准则特征集	19	93.48	71.88
RF 特征集	29	94.91	21.45
两阶段特征优选策略特征集	19	94.06	42.65

进一步对比发现,在保持相同特征维度时,两阶段特征优选策略相比于 mRMR 不仅耗时更短,且聚类准确率略高,显示了该策略在计算效率上的优势;而相较于 RF 特征选择方法,本文策略能在更低维度特征集上达到理想的准确率表现,在减少冗余信息、提炼关键负荷特性方面表现较好。

综上,所提两阶段特征优选策略既能有效降低计算复杂度,又可精确刻画多元负荷用户综合用能行为,具备在特征选择效率和精度上的双重优势,能更好地适应大规模用能负荷数据的分析场景。

4.2 多元负荷用户群体聚类分析

基于两阶段特征优选策略得到的最优特征子集,采用 3W-ADPC 算法进行多元负荷用户群体划分,并与 K-means、DPC、仅引入自适应最近邻改进 DPC 的聚类效果进行比较。为全面评估聚类算法的性能,除聚类准确率和运行时间外,采用 3 个常用的聚类评价指标进行对比分析:戴维森-堡丁指数(Davies-Bouldin index, DBI)、邓恩指数(Dunn's Va-

lidity index, DVI) 和轮廓指数 (silhouette coefficient, SC), 聚类结果如表 8 所示。

由表 8 可知,三类 DPC 算法在多个评价指标上均较 *K*-means 有所提升,尤其在聚类准确率上,表明在大规模高维度数据集中 DPC 类算法表现更优,不过与此同时计算耗时增长,体现出性能与效率间的权衡。自适应最近邻 DPC 与 3W-ADPC 均展现出良好的簇内紧密度与簇间分离能力,对比来看,3W-ADPC 算法在多项评价指标上均较优,且聚类准确率较自适应最近邻 DPC 显著提升,尽管运算耗时有所增加,但这一增长幅度微乎其微,总的来看,3W-ADPC 在聚类效果、精度和时间效率方面均表现出其优越性。综上所述,3W-ADPC 方法能够有效地识别并精确划分不同用户群体的用能行为特性,实现基于负荷行为特性的多元用户群体分类。

表 8 不同算法的聚类效果对比

Table 8 Comparison of clustering effect of different algorithms

特征集	DVI	DBI	SC	聚类 准确率/%	运行 时间/s
<i>K</i> -means	0.34	0.57	0.57	83.45	4.22
DPC	0.21	0.51	0.59	94.06	6.87
自适应最近邻 DPC	0.27	0.50	0.59	95.33	13.08
3W-ADPC	0.32	0.51	0.60	98.59	13.60

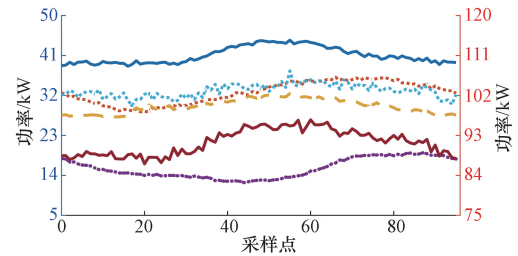
4.3 多元负荷用户综合用能画像分析

运用 3W-ADPC 算法对负荷进行聚类,并将各聚类中心作为各类用户的典型负荷,如图 3 所示,展示了各类用户中典型用户的电、冷、热三类日负荷曲线,可以看出每类用户具有独特的用能特点。

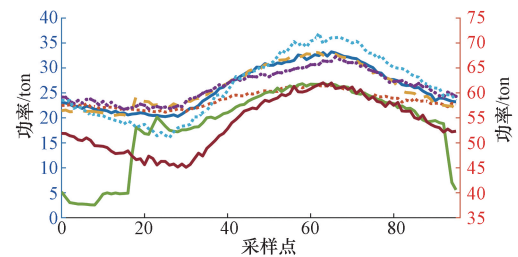
为清晰展示第 3 类和第 4 类用户较高的负荷水平,并同时呈现所有用户的负荷曲线特征,采用双纵轴图表绘制方式,横坐标是 96 个日负荷采集点。

基于优选特征与用户类簇划分结果,运用 CRITIC (criteria importance through intercriteria correlation) 法计算各用能特性标签的权重,采用标签打分制评估各类用户在各维度上的用能表现水平,并以雷达图为载体,直观展现不同类别多元负荷用户的综合用能行为画像,如图 4 所示,揭示了各类用户在不同时间尺度下的用能特性差异化分布。

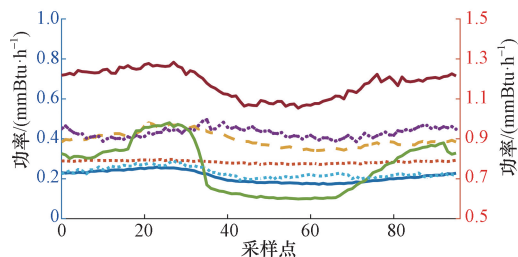
具体来看,第 1 类用户表现出较高的短期用电负荷,而长期负荷水平较低,反映其日常用电频繁且较均衡,整体较为平稳。用电行为规律分数不高,表明此类用户用电行为缺乏明显季节性规律,年初负荷集中,非高峰期负荷较低。用电互动短期评分较高,表明负荷难以准确预测且控制难度大,蕴含节能潜力,可通过智能定价和需求响应策略优化负荷管理,提升供电可靠性和用户满意度。在用



(a) 典型用户电负荷曲线



(b) 典型用户冷负荷曲线



(c) 典型用户热负荷曲线



第 3 类和第 4 类数据从右侧纵坐标读取;冷负荷 1 ton = 3.5 kW;
热负荷 1 mmBtu/h = 293.071 1 kW

图 3 各类典型用户的典型日负荷曲线

Fig. 3 Typical daily load curves for various typical users

冷和用热特性方面,该类用户冷负荷需求较强,可能是一年内有较长制冷期的场所,全年均维持一定水平的热能消耗,但热负荷并非核心需求。从多元负荷耦合角度看,尽管电-冷短期负荷关联度高,显示两者间有较强的同步变动关系,即电力消耗往往伴随着冷能输出,如空调等设备,但整体负荷间耦合性评分偏低。综上所述,第 1 类用户的用能画像具备典型的商业运营模式下多变的能源消耗特性,节能优化潜力大,如办公大楼、商业楼宇等场所。

类似的,对第 2~7 类用户的用能画像结果进行如上分析,可得出以下结论:第 2 类用户用电负荷适中且稳定,业务活动平稳、能源管理高效,如中型商业、轻工业等场所,全年冷热需求平衡,尤其是热能分配较平稳,冷热负荷长期耦合紧密,推测可能配备高效节能的智能冷暖系统;第 3 类用户全年负荷需求较强且有节能潜力,冷热负荷季节性同步变化,如大型工厂、数据中心等能源密集型场所,由于其高能耗属性,参与削峰管理挑战较大;第 4 类用户

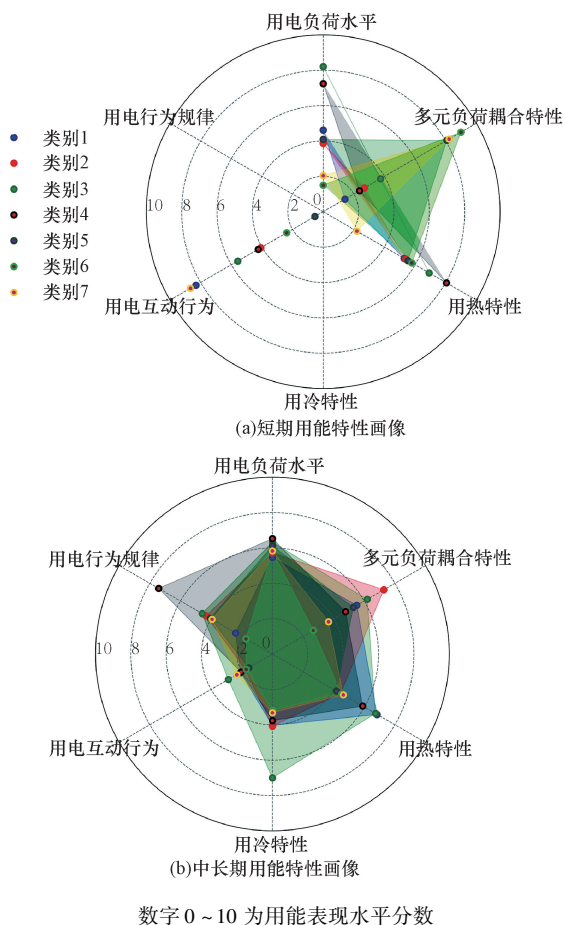


图4 不同类别多元负荷用户用能特性画像

Fig. 4 Portraits of different categories of multi-load user energy characteristics

全年冷热负荷需求大,如大型医疗设施、购物中心或酒店等场所,且具备削峰潜力,适合参与电网需求响应;第5类用户负荷稳定且能源管理良好,冷热负荷有季节性规律,如小型办公楼、学校等场所;第6类用户供暖需求稳定但主要依赖非电力供热系统,如燃气热水器或燃煤锅炉等,制冷需求有限,冷热负荷短期同步性强,如注重节能的小型商铺或住宅等;第7类用户与第6类相似,但其互动能力较强,削峰潜力大,制冷供热需求均较低,负荷间关联性明显,可能是受季节气候的影响。

根据多元负荷用户用能画像的结果,电力公司可以依据用户在不同用能行为特性方面的表现为其提供个性化的用能服务,例如对于第3类能源密集型生产用户,可通过合理实施分时电价政策,配合用户调整重要设备运行时间以降低尖峰时段的负荷压力,保障电力系统稳定性与经济性。此外,用户还可以根据自身综合用能画像与得分确定用能改进方向,如第1类用户在用电互动能力方面评分表现突出,可充分挖掘自身节能潜力,优化负荷管理,提升用能稳定性。

5 结论

针对当前用户用能行为研究中负荷类型单一、特征选择困难以及行为模式复杂的问题,构建了一个计及多元负荷耦合特性的多维用户用能行为标签体系,并在此基础上提出了一种基于两阶段特征优选策略与改进的3W-ADPC聚类算法的用户用能画像方法。实验表明,所提的两阶段特征优选策略有效揭示了多元负荷用户的多样化和精细化用能特征,显著提升了负荷分类准确性,且改进的3W-ADPC聚类方法在性能上优于传统模型,对于电力企业识别用户差异化需求、实现精准服务具有重要的实践指导意义。通过实证分析,成功构建了多元负荷用户综合用能画像,并完成了初步的特性分类与行为定性解析,证实了所提策略与算法在多元负荷用户行为研究中的实用价值。未来研究可进一步挖掘更多个性指标,优化定量评估模型,并探讨如何将用户画像技术用于精细化需求响应管理和个性化能源服务设计,以推动电力行业向智能化、个性化服务方向深入发展。

参考文献

- [1] 孔祥玉, 马玉莹, 艾芊, 等. 新型电力系统多元用户的用电特征建模与用电负荷预测综述[J]. 电力系统自动化, 2023, 47(13): 2-17.
Kong Xiangyu, Ma Yuying, Ai Qian, et al. Review on electricity consumption characteristic modeling and load forecasting for diverse users in new power system[J]. Automation of Electric Power Systems, 2023, 47(13): 2-17.
- [2] 刘诗语, 吴鸣, 李睿哲. 基于多维缩放和KICIC的电力负荷聚类[J]. 科学技术与工程, 2023, 23(3): 1096-1103.
Liu Shiyu, Wu Ming, Li Ruizhe. Power load curve clustering research based on multi-dimensional scaling and KICIC[J]. Science Technology and Engineering, 2023, 23(3): 1096-1103.
- [3] 洪于亮, 许宁, 秦焕美, 等. 基于用户画像的大型冬季户外活动客流需求特征挖掘——以哈尔滨和长春国际滑雪马拉松为例[J]. 科学技术与工程, 2023, 23(10): 4400-4407.
Hong Yuliang, Xu Ning, Qin Huanmei, et al. Characteristics mining of passenger flow based on user portrait in large winter outdoor events: take Harbin and Changchun International Skating Marathon as an example[J]. Science Technology and Engineering, 2023, 23(10): 4400-4407.
- [4] 余顺坤, 闫泓序, 杜诗悦, 等. 基于SC-RS的我国工业电力用户价值画像模型研究[J]. 中国管理科学, 2022, 30(3): 106-116.
Yu Shunkun, Yan Hongxu, Du Shiyue, et al. Research on the customer value portrait model of industrial power enterprise in China based on spectral clustering technology and rough set theory[J]. Chinese Journal of Management Science, 2022, 30(3): 106-116.
- [5] 刘文君, 董明, 徐元孚, 等. 电力设备运行状态大数据标签体系与关键技术[J]. 中国电力, 2022, 55(1): 126-132.
Liu Wenjun, Dong Ming, Xu Yuanfu, et al. Structure and key

- technologies of big data labeling system for power equipment operation status[J]. *Electric Power*, 2022, 55(1): 126-132.
- [6] 张智, 王韵楚, 林振智, 等. 基于套餐隐式评分与用户画像的电力套餐推荐方法[J]. *电力系统自动化*, 2023, 47(4): 91-101.
- Zhang Zhi, Wang Yunchu, Lin Zhenzhi, et al. Electricity plan recommendation method based on implicit score of electricity plan and user portrait[J]. *Automation of Electric Power Systems*, 2023, 47(4): 91-101.
- [7] 王永明, 陈宇星, 殷自力, 等. 基于大数据分析的电力用户行为画像构建方法研究[J]. *高压电器*, 2022, 58(10): 173-179, 187.
- Wang Yongming, Chen Yuxing, Yin Zili, et al. Research on construction method of power user behavior portrait based on big data analysis[J]. *High Voltage Apparatus*, 2022, 58(10): 173-179, 187.
- [8] 汪颖, 喻梦洁, 卢宏, 等. 基于最大互信息的干扰源类型识别及电能质量需求画像技术[J]. *电力系统自动化*, 2022, 46(9): 171-181.
- Wang Ying, Yu Mengjie, Lu Hong, et al. Interference source type identification and power quality demand portrait technology based on maximum mutual information[J]. *Automation of Electric Power Systems*, 2022, 46(9): 171-181.
- [9] 王磊, 刘洋, 李文峰, 等. 基于用电行为数字特征画像的电力用户两阶段分类方法[J]. *电力建设*, 2022, 43(2): 70-80.
- Wang Lei, Liu Yang, Li Wenfeng, et al. Two-stage power user classification method based on digital feature portraits of power consumption behavior[J]. *Electric Power Construction*, 2022, 43(2): 70-80.
- [10] 梁纪峰, 范辉, 李顺, 等. 计及响应度的电力用户互动潜力画像分析[J]. *科学技术与工程*, 2022, 22(15): 6130-6138.
- Liang Jifeng, Fan Hui, Li Shun, et al. Analysis of power user interaction potential portrait considering user response[J]. *Science Technology and Engineering*, 2022, 22(15): 6130-6138.
- [11] 范宇辉, 姜婷玉, 黄奇峰, 等. 基于画像的工业园区需求响应潜力评估[J]. *电力系统自动化*, 2024, 48(1): 41-49.
- Fan Yuhui, Jiang Tingyu, Huang Qifeng, et al. Portrait-based assessment on demand response potential of industrial parks[J]. *Automation of Electric Power Systems*, 2024, 48(1): 41-49.
- [12] 陈非凡, 高亚静, 梁海峰, 等. 基于地理分区的综合能源园区用能特性评价模型及供能分区方法[J]. *电力建设*, 2019, 40(6): 23-32.
- Chen Feifan, Gao Yajing, Liang Haifeng, et al. Energy consumption characteristic evaluation model and energy supply partition method of integrated energy park based on geographical partition[J]. *Electric Power Construction*, 2019, 40(6): 23-32.
- [13] 曹斯明, 张慧波, 曹凯, 等. 数据驱动的中大型能源用户多类型用能特性画像及评价方法[J]. *供用电*, 2022, 39(2): 71-79.
- Cao Siming, Zhang Huibo, Cao Kai, et al. Data-driven portrayal and evaluation method for multi-type energy consumption characteristics of large and medium-sized energy users[J]. *Distribution & Utilization*, 2022, 39(2): 71-79.
- [14] 朱辉, 闫腾飞, 丁一, 等. 基于改进的 AHP-CRITIC 综合权重计算法的多元用户用能行为特性刻画及评价方法[J]. *东北电力大学学报*, 2023, 43(5): 70-77.
- Zhu Hui, Yan Tengfei, Ding Yi, et al. Characterization and evaluation method for multi-user energy consumption behavior based on the improved AHP-CRITIC composite weighting approach[J]. *Journal of Northeast Electric Power University*, 2023, 43(5): 70-77.
- [15] 赵晋泉, 夏雪, 刘子文, 等. 电力用户用电特征选择与行为画像[J]. *电网技术*, 2020, 44(9): 3488-3496.
- Zhao Jinquan, Xia Xue, Liu Ziwen, et al. User electricity consumption feature selection and behavioral portrait[J]. *Power System Technology*, 2020, 44(9): 3488-3496.
- [16] 刘炼, 王强, 陈浩. 基于改进随机森林的电力系统暂态稳定评估[J]. *科学技术与工程*, 2022, 22(11): 4367-4374.
- Liu Lian, Wang Qiang, Chen Hao. Transient stability assessment of power system based on improved random forest[J]. *Science Technology and Engineering*, 2022, 22(11): 4367-4374.
- [17] 周晨璟, 邵振国, 陈飞雄, 等. 考虑多指标融合的电能质量扰动特征优选策略[J]. *电网技术*, 2023, 47(9): 3873-3885.
- Zhou Chenjing, Shao Zhenguo, Chen Feixiong, et al. Feature selections for power quality disturbance signals with multi-indicator fusion[J]. *Power System Technology*, 2023, 47(9): 3873-3885.
- [18] 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.
- [19] 张琳娟, 许长清, 王利利, 等. 基于多维度能耗分析的园区用户画像模型[J]. *可再生能源*, 2021, 39(8): 1078-1086.
- Zhang Linjuan, Xu Changqing, Wang Lili, et al. User profile model of park based on multi-dimensional energy consumption analysis[J]. *Renewable Energy Resources*, 2021, 39(8): 1078-1086.
- [20] 姜飞, 林政阳, 王文焯, 等. 考虑最小平均包络熵负荷分解的最优 Bagging 集成超短期多元负荷预测[J]. *中国电机工程学报*, 2024, 44(5): 1777-1789.
- Jiang Fei, Lin Zhengyang, Wang Wenye, et al. Optimal bagging ensemble ultra short term multi-energy load forecasting considering least average envelope entropy load decomposition[J]. *Proceedings of the CSEE*, 2024, 44(5): 1777-1789.
- [21] 魏寅孔, 李振华, 李振兴, 等. 特高压直流输电线路合成电场的天气影响因素及预测模型[J]. *科学技术与工程*, 2023, 23(9): 3769-3778.
- Wei Yinkong, Li Zhenhua, Li Zhenxing, et al. Weather factors and prediction model of synthetic electric field of ultra-high voltage direct-current transmission line[J]. *Science Technology and Engineering*, 2023, 23(9): 3769-3778.
- [22] Yu H, Chen L Y, Yao J T. A three-way density peak clustering method based on evidence theory[J]. *Knowledge-based Systems*, 2021, 211: 106532.
- [23] 夏飞, 张洁, 张浩, 等. 基于 BIC 准则和加权皮尔逊距离的居民负荷模式精细识别及预测[J]. *电子测量与仪器学报*, 2020, 34(11): 33-42.
- Xia Fei, Zhang Jie, Zhang Hao, et al. Fine recognition and prediction of resident load pattern based on BIC criterion and weighted Pearson distance[J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 34(11): 33-42.