

DOI: 10.19783/j.cnki.pspc.231575

基于改进FCM和最小互信息算法的户变关系辨识方法

李彦伦¹, 窦晓波¹, 卜强生², 徐晓春³, 吕朋蓬²

(1. 东南大学电气工程学院, 江苏 南京 210096; 2. 国网江苏省电力有限公司电力科学研究院, 江苏 南京 211103;
3. 国网江苏省电力有限公司淮安供电分公司, 江苏 淮安 223002)

摘要: 随着低压分布式光伏的高比例接入, 传统的户变关系辨识方法利用电压与功率特征进行判断, 受光伏出力影响较大。针对分布式光伏大规模接入下低压配电网的电气特征, 提出了一种基于数据驱动的低压配电网户变关系辨识方法。首先, 基于改进模糊C均值(fuzzy C-means, FCM)聚类算法, 利用电压相关性对用户进行初步聚类划分。其次, 在初步聚类的基础上, 利用最小互信息算法排除光伏出力的影响, 建立配变与用户的连接关系识别模型并用回归分析的思想求解, 实现对户变关系的精确辨识。最后, 通过对某地实际数据进行算例分析, 验证了基于改进FCM和最小互信息算法的户变关系辨识方法在大规模光伏接入场景下的有效性。实验表明, 该方法相比传统的电压和功率特征判别方法具有更高的准确率。

关键词: 低压配电网; 户变关系辨识; 拓扑识别; 电压相关性; 互信息

Identification method of transformer-customer relationship based on an improved FCM algorithm and minimum mutual information

LI Yanlun¹, DOU Xiaobo¹, BU Qiangsheng², XU Xiaochun³, LÜ Pengpeng²

(1. College of Electrical Engineering, Southeast University, Nanjing 210096, China; 2. Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211103, China; 3. Huai'an Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd., Huaian 223002, China)

Abstract: With the high proportion of low-voltage distributed PV access, traditional transformer-customer relationship identification methods based on voltage and power characteristics are largely affected by photovoltaic input. Given the electrical characteristics of the low-voltage distribution network when there is large-scale distributed PV access, a data-driven identification method for transformer-customer relationship in the low-voltage distribution network is proposed. First, using voltage correlation, customers are preliminarily divided based on an improved fuzzy C-means (FCM) clustering algorithm. Then, based on the initial clustering, a minimum mutual information algorithm is used to eliminate the influence of photovoltaic output, and an identification model of the connection relationship between distribution transformers and customers is established. Regression analysis is used to solve the problem, so as to realize accurate identification. Finally, the effectiveness of the identification method of transformer-customer relationship based on improved FCM algorithm and minimum mutual information is verified by analyzing the actual data of a certain place. Experiments show that this method has higher accuracy than the traditional discrimination method using voltage and power characteristics.

This work is supported by the Science and Technology Project of the Headquarters of State Grid Corporation of China (No. 5108-202218280A-2-367-XG).

Key words: low-voltage distribution network; transformer-customer relationship identification; topology identification; voltage correlation; mutual information

0 引言

户变关系是指终端用户的电表与配电台区变压器

器之间的从属连接关系^[1-2], 准确的户变关系是故障定位、线损治理、线路改造等业务开展的重要保证^[3]。低压台区作为配电网的末端系统, 直接面向用户供电^[4-7], 是保证供电质量的重要环节。近年来, 随着低压台区用户数量的急剧增长, 台区户变关系变化频繁, 对台区精细化管理水平提出了更高要求^[8]。但由于排查效率低、信息更新不及时等问题,

基金项目: 国网有限公司总部科技项目资助(5108-202218280A-2-367-XG); 国家自然科学基金项目资助(52077036)

户变关系的档案信息可能存在错误, 影响台区管理及故障研判等业务的展开^[9-10]。

户变关系识别方法主要分为注入信号法与数据分析法^[11-12]两类。注入信号法通过在网络中安装专门的设备注入谐波或无功功率等信号, 在相邻节点处读取信号的特征以实现户变关系辨识^[13-14], 但需要安装额外的设备数据, 使用场景存在局限性且成本较高。数据分析法通过智能电表采集的电压、功率数据分析用户间的相关性, 实现户变关系辨识。文献[15]提出一种状态量测的电网拓扑结构反向识别方法; 文献[16]通过分析节点间的电气距离, 提出了虚拟阻抗的概念实现拓扑关系异常设备的辨识; 文献[17]提出了基于最大似然估计框架的变量误差模型, 以机器学习的方式进行户变关系辨识。文献[18]通过节点相关性分析和基于拓扑建模的马尔可夫随机场研究系统拓扑, 并引入了伪似然方法和 L2 正则化理论提高计算效率。文献[19]基于有功功率序列的时频变换进行特征提取, 利用回归分析的方法辨识分支节点的层次关系和区段信息。文献[20]基于自适应分段聚合近似(adaptive piecewise aggregate approximation, APAA)和改进 DBSCAN(density-based spatial clustering of application with noise, DBSCAN)算法提取电压曲线特征, 避免了曲线相似性难以区分的问题; 文献[21]结合地理信息系统(geographic information system, GIS)和电压波形相关性分析估计次级电路拓扑。文献[22]综合地理信息、电压与功率特征, 基于关联卷积与功率平衡原理实现户变关系识别。文献[23]针对量测数据采集周期长、数据不平衡的问题, 基于两层堆叠的图卷积网络提出了一种两阶段的新型配电网拓扑辨识方法。

目前常用的数据分析法利用电压或功率量测值分析各个节点的特征, 从而实现对户变关系的辨识。但是, 随着分布式光伏的装机规模和范围逐渐扩大, 高比例分布式光伏的接入给户变关系识别带来了新的问题^[24]。光伏的接入使配电网的电气特性复杂化, 产生电压波动、配网有源化、功率倒送等问题, 从而导致功率、电压等电气特征量数据区分度下降^[25], 传统的基于电压和功率特征的方法辨识难度增加。针对上述问题, 文献[26]基于定量关联规则挖掘(quantitative association rule mining, QARM), 利用光伏节点与用电节点间的电压相似性, 实现了对二次配电网的动态感知, 但该方法仅对光伏节点位置进行准确判断, 无法判断非光伏节点的连接关系。文献[27]基于双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)算法建立光伏权重矩阵后利用关联卷积算法进行户变关系辨识, 无

法完全消除光伏对特征量产生的影响, 且该方法需要的实时光伏量测数据不易获得。

综合上述情况, 传统的辨识方法一般未考虑光伏接入的影响, 准确率和经济性较低。现有的考虑分布式光伏接入的方法在使用场景和数据获取方面均存在不足, 本文针对高比例分布式光伏接入下, 户变关系不明确的问题, 对低压台区户变关系进行了辨识。首先, 基于模糊 C 均值(fuzzy C-means, FCM)算法, 利用电压相关性对用户进行初步聚类。接着, 基于最小互信息排除光伏出力的影响, 并基于功率平衡原理建立多元线性回归方程, 通过回归分析精确辨识户变关系。最后, 通过某地实际数据进行算例分析, 验证了本文所提方法的有效性。

1 含高渗透率分布式光伏的低压台区户变关系辨识框架

1.1 低压台区拓扑结构

图 1 展示了电力系统中典型低压台区配电网的拓扑结构。低压台区的配网拓扑以放射型与树干型结构为主, 由“配变-分支箱-表箱-用户”的连接关系组成。在这种结构下, 潮流存在确定的流动方向, 物理条件上满足拓扑辨识的要求。节点的电压变化量主要与配变低压母线侧电压的变化量、上游线路电气距离及传输功率的变化量有关。同一台区用户之间的电气距离越短, 其电压时序曲线的相似程度越高。

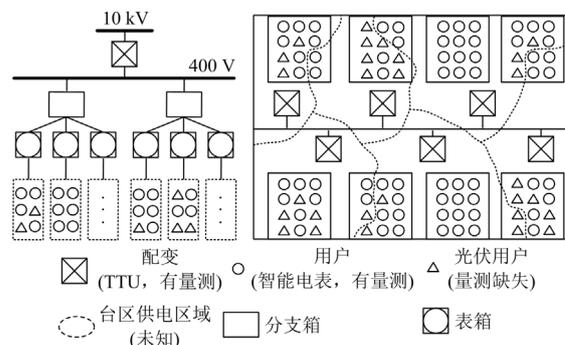


图 1 低压台区拓扑结构示意图

Fig. 1 Schematic diagram of structural topology in low voltage distribution network

1.2 户变关系辨识问题描述

目前, 低压台区户变关系异常主要包括:

1) 由于用户私自改接、维修时线路调整、人工录入出错等因素, 系统中记录的户变关系与实际情况不一致;

2) 由于低压配网改造频繁, 扩建后户变关系未

能及时更新，导致用户的台区所属未知^[28]。

因此，户变关系异常无法被及时感知。一方面初始户变关系可能发生错误，另一方面，即使初始户变关系正确，后期变化等因素也会导致户变关系出现错误，所以需要进行后期排查和校验。

此外，当前的户变关系辨识存在以下问题：

1) 由于城市中低压台区用户较为密集，变压器的供电范围往往存在重叠，同一栋小区居民楼往往会有 2 台或以上的变压器进行供电，导致电气距离较近的用户并不一定属于同一台区；

2) 低压台区分布式光伏渗透率逐步提高，但考虑到成本、政策、隐私等影响因素，对于用户的光伏出力以不监控或弱监控为主，无法实时感知光伏出力^[29]。

1.3 户变关系辨识流程

基于上述分析，本文提出含高比例分布式光伏的低压台区户变关系辨识模型。本文的辨识方法依赖于用户侧和配变侧的历史功率和电压数据，这些数据可通过配电网营销系统获取，识别流程如图 2 所示，共分为 3 个阶段。阶段一：基于用户侧智能

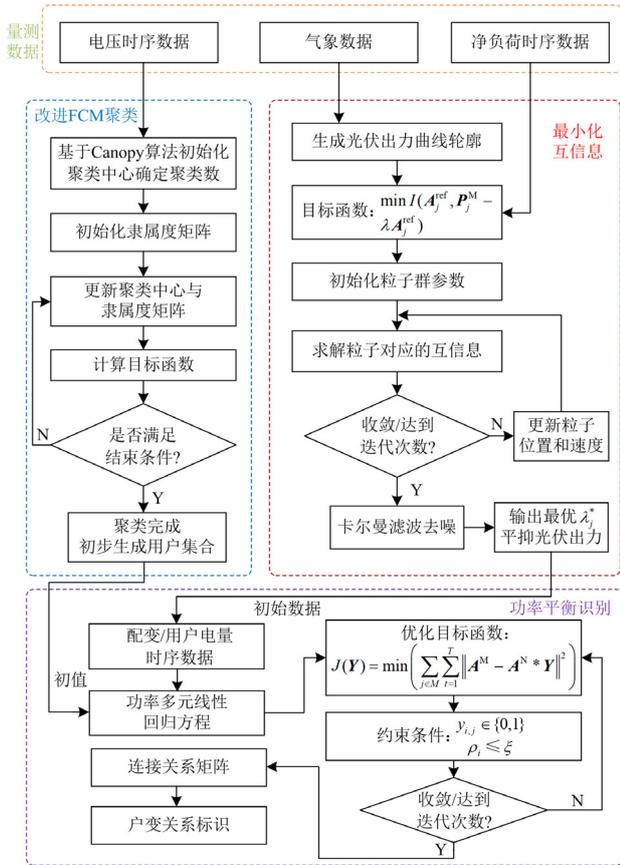


图 2 户变关系辨识流程

Fig. 2 Identification process of transformer-customer relationship

电表的电压量测数据分析电压曲线特征，利用电压相关性实现用户的初聚类；阶段二：基于配变侧和用户侧的净负荷量测数据，结合气象数据或少量光伏量测数据，使用最小互信息算法削减光伏出力的影响；阶段三：将上述两步生成的用户集合作为迭代初值，光伏出力分解后的负荷数据作为初始数据，基于功率平衡原理建立多元回归方程并求解，实现户变关系的辨识。

2 基于改进 FCM 算法的电压初聚类

通常情况下，从属于同一配变的用户具有更紧密的电气连接和更近的电气距离，而电压相关性可以反映用户之间的电气距离。因此，引入改进 FCM 聚类算法来分析用户间的相关程度，对台区用户进行初步聚类。

2.1 传统 FCM 算法

传统 FCM 聚类算法是在经典 K-means 聚类算法基础上提出的改进算法。相较于传统算法采用的硬聚类方式，FCM 聚类算法结合模糊理论，实现柔性的模糊聚类。

FCM 算法主要通过构造拉格朗日函数，迭代求取各样本到聚类中心的全局加权距离平方和最小值以获得最优聚类中心。

针对一个有 m 台配变和 n 个用户的区域，对于第 j 个用户，有电压量测序列 $\mathbf{x}_j = [v_{j,1}, v_{j,2}, \dots, v_{j,T}]$ ，全部用户的电压量测数据可以用数据集 $\mathbf{D} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ 来表示。其中， T 表示每个用户采集的时间断面数，即特征维度， $v_{j,T}$ 表示用户 j 在 T 时刻的电压量测。

若标记聚类中心矩阵为 \mathbf{V} ，对应的隶属度矩阵为 \mathbf{U} ，FCM 算法的目标函数可表示为

$$J_w(\mathbf{U}, \mathbf{V}) = \sum_{j=1}^n \sum_{k=1}^c u_{kj}^w d_{kj} \quad (1)$$

式中： w 为模糊化指数； u_{kj} 表示第 j 个用户样本对第 k 个聚类的隶属度； d_{kj} 表示第 j 个用户样本到第 k 个聚类中心的距离； n 为用户数； c 表示聚类数。

采用拉格朗日乘法求出式(1)的极值，隶属度和聚类中心的迭代公式分别为

$$u_{kj} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{kj}}{d_{hj}} \right)^{\frac{2}{w-1}}} \quad (2)$$

$$\mathbf{c}_k = \frac{\sum_{j=1}^n u_{kj}^w}{\sum_{j=1}^n \sum_{k=1}^c u_{kj}^w} \mathbf{x}_j \quad (3)$$

式中: c_k 代表第 k 个聚类中心; d_{hj} 表示第 j 个用户样本到第 h 个聚类中心的距离。

2.2 基于密度 Canopy 的改进 FCM 算法

传统 FCM 算法的初始聚类中心是随机生成的。如果初始聚类中心选择不合理, 算法容易陷入局部最优解^[30]。因此, 提出一种基于密度的 Canopy 算法自动确定聚类数和初始聚类中心, 以此提高算法的自适应性和稳定性。

通过计算数据集 D 中密度值最大的序列, 确定聚类中心 c_k 的不相似度在截断距离阈值之内的序列, 并在数据集 D 中删去该序列。

$$MeanDist(D) = \frac{2}{n(n-1)} \sum_{l=j+1}^n \sum_{l=j+1}^n d_{lj} \quad (4)$$

$$\begin{cases} \rho_j = \sum_{j=1}^n f[d_{lj} - MeanDist(D)] \\ f(x) = \begin{cases} 1, x < 0 \\ 0, x \geq 0 \end{cases} \end{cases} \quad (5)$$

式中: $MeanDist(D)$ 表示平均不相似度; d_{lj} 表示序列 x_i 和 x_j 间的距离; ρ_j 为序列 x_j 的数据密度。

计算数据集 D 中除 c_k 外剩余序列的聚类中心权重 w_j , 选出聚类中心权重值最大的序列, 得到下一个聚类中心, 然后确定不相似度在截断距离之内的序列, 并在数据集 D 中删去该序列, 如式(6)~式(8)所示。

$$a_j = \frac{2}{\rho_j(\rho_j - 1)} \sum_{l=j+1}^{\rho_j} \sum_{l=j+1}^{\rho_j} d_{lj} \quad (6)$$

$$s_j = \begin{cases} \min(d_{lj}), \exists j, \rho_l > \rho_j \\ \quad \quad \quad x_l: \rho_l > \rho_j \\ \max(d_{lj}), \forall j, \rho_l < \rho_j \\ \quad \quad \quad x_l \in D \end{cases} \quad (7)$$

$$w_j = \rho_j \cdot \frac{1}{a_j} \cdot s_j \quad (8)$$

式中: a_j 表示平均不相似度, 其由与序列 x_j 不相似度在 $MeanDist(D)$ 范围内所有序列计算得到; s_j 表示该序列与其他聚类间的不相似度; ρ_l 表示序列 x_l 的数据密度。

迭代计算式(4)~式(8), 重复上述步骤直至数据集 D 为空, 得到聚类中心集合并重新计算初始聚类中心。计算各序列与聚类内其他序列的不相似度, 选取此聚类中不相似度最小的序列, 确定初始聚类中心。

在 Canopy 算法确定聚类数和初始聚类中心后, 使用 FCM 算法进行聚类分析。在利用电压作初步分类的基础上, 接下来将考虑功率数据的特征, 进

一步对户变关系进行精确辨识。

3 基于最小互信息的户变关系精确识别

在聚类算法的基础上, 针对高比例光伏接入对量测数据的影响, 首先基于最小互信息算法提取光伏出力特征, 分解用户负荷和光伏出力。接着, 利用功率平衡原理建立平衡方程, 并求解多元线性回归问题, 得到以 0-1 变量表示的用户与变压器连接关系矩阵 Y , 实现户变关系识别。

3.1 基于最小互信息的光伏特征提取

首先基于最小互信息算法进行光伏特征提取, 对混有光伏的原始净负荷数据进行预处理。

标准互信息(normalized mutual information, NMI) 是一种用于衡量两个随机变量共同含有信息量的指标^[31-32]。对于离散变量, 通过式(9)~式(11)计算标准互信息。

$$H(A) = -\sum_{a \in \mathcal{X}} p(a) \log_2 p(a) \quad (9)$$

$$H(A, B) = -\sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{V}} p(a, b) \log_2 p(a, b) \quad (10)$$

$$I(A, B) = \frac{H(A) + H(B) - H(A, B)}{\sqrt{H(A)H(B)}} \quad (11)$$

式中: A 和 B 表示两个随机变量; a 、 b 表示两个随机变量; $H(A)$ 和 $H(B)$ 分别表示 A 和 B 的信息熵; \mathcal{X} 和 \mathcal{V} 分别表示 a 和 b 的取值范围; $p(a)$ 和 $p(b)$ 分别表示 A 和 B 的概率分布; $H(A, B)$ 表示 A 、 B 之间的联合熵; $p(a, b)$ 表示 A 、 B 之间的联合概率分布; $I(A, B)$ 表示 A 、 B 之间的标准互信息。

光伏出力特征的提取可视为一个最小互信息问题。由于用户侧和配变侧的光伏出力都是未知的, 需要对光伏用户和配变的净负荷数据进行处理, 对于某一用户 j , 台区内的净负荷数据表示为

$$P_j^M = A_j^L + A_j^{PV} \quad (12)$$

$$A_j^{\text{ref}} = \frac{1}{\lambda_j^*} A_j^{PV} \quad (13)$$

式中: P_j^M 表示用户 j 在 T 个时间断面上的净负荷矩阵; A_j^L 表示非光伏的负荷矩阵; A_j^{PV} 表示光伏出力矩阵; A_j^{ref} 表示参考光伏出力; λ_j^* 表示实际光伏出力与参考光伏出力之间的比例, 即可实现光伏出力特征的提取。

设 λ 为待求解的光伏倍率, 在 λ 非最优解的情况下, 式(13)中求解出的负荷分量就会混合一定比例的光伏分量。 λ 越趋近最优解, 负荷信息中包含的光伏信息就越少, 分解效果就越好, 即两者的互信息最小。以 λ 作为未知量, 设置评估分解效果的

目标函数为

$$I(\mathbf{A}_j^{\text{pv}}, \mathbf{A}_j^{\text{l}}) \approx \min I(\mathbf{A}_j^{\text{ref}}, \mathbf{P}_j^{\text{M}} - \lambda \mathbf{A}_j^{\text{ref}}) \quad (14)$$

考虑到实际物理情况, 对该优化问题添加约束, 增加对变量 λ 的约束, 表示为

$$0 \leq \lambda \leq S^{\text{T}} / \max(\mathbf{A}_j^{\text{ref}}) \quad (15)$$

式中, S^{T} 表示变压器容量, 显然台区所装光伏容量不会超过变压器容量, 因此该约束条件表示光伏倍率不会超过变压器容量 S^{T} 与参考光伏曲线最大值之间的比值。

考虑到噪声的影响, 仅采用最小互信息的方式估计光伏出力存在误差。因此, 采用卡尔曼滤波算法对光伏倍率进行滤波去噪。卡尔曼滤波分为预测与校正两个过程, 预测过程如式(16)、式(17)所示。

$$\lambda_t = \lambda_{t-1} \quad (16)$$

$$W_t = W_{t-1} + Q \quad (17)$$

式中: λ_{t-1} 为 $t-1$ 时刻经过卡尔曼滤波的光伏倍率; λ_t 为根据 $t-1$ 时刻的数据得到的 t 时刻的光伏倍率预测值; W_t 表示根据校正后的卡尔曼增益更新预测的误差; Q 为卡尔曼模型的预测误差。

误差校正过程如式(18)—式(20)所示。

$$K_t = \frac{W_t}{W_t + R} \quad (18)$$

$$\lambda_t = \lambda_t + K_t(\lambda - \lambda_t) \quad (19)$$

$$W_t = (1 - K_t)W_t \quad (20)$$

式中: R 表示基于互信息进行分解得到的光伏倍率的噪声方差, 其量级与 Q 相同; K_t 表示 R 与预测误差进行加权后获得的 t 时刻的卡尔曼增益; λ 为 t 时刻由互信息算法得到的光伏倍率, 根据该测量值与预测值得到 t 时刻获得的光伏倍率滤波值 λ_t 。

在长时间尺度上, 实际台区内的光伏存在投切与新装, 这会导致光伏容量变化, 也就是光伏倍率改变, 而以上建立的卡尔曼静止模型追踪这种阶跃式的突变存在较大延时, 卡尔曼控制方程本身没有能够表征光伏容量变化的输入。由上述最小互信息方法计算得到的光伏倍率可能会产生突变, 该突变可能会被滤波算法当作噪声进行平滑, 所以引入光伏倍率动态变量 a , 则 t 时刻光伏实时出力表示为

$$\mathbf{A}_j^{\text{pv}}(t) = [a\lambda + (1-a)\lambda_t] \mathbf{A}_j^{\text{ref}}(t) \quad (21)$$

针对每个用户节点, 均可使用上述方法求得各个节点最优的 λ_j^* 。经过卡尔曼滤波后, 进一步基于式(21)与式(12)求解得到各个节点的光伏出力 \mathbf{A}_j^{pv} 及负荷 \mathbf{A}_j^{l} , 并将其作为下一步识别的数据基础。

3.2 基于多元线性回归的户变关系识别

在由改进FCM算法给出求解算法的初值, 并由最小互信息算法处理负荷数据后, 基于多元线性回归进行户变关系精确识别。

这一步骤的基础为功率守恒原理, 即排除量测误差和功率损耗的影响后, 在任意一个时刻, 用户负荷之和与配变负荷相等。基于该理论建立多元线性回归方程组, 利用优化算法求解该线性回归问题。

采集某低压台区 T 个时间断面的净负荷数据, 并经过互信息分解光伏出力后得到配变侧与用户侧的负荷数据, 如式(22)、式(23)所示。

$$\mathbf{A}^{\text{M}} = [\mathbf{A}_{t,i}^{\text{M}}]_{T \times m} \quad (22)$$

$$\mathbf{A}^{\text{N}} = [\mathbf{A}_{t,j}^{\text{N}}]_{T \times n} \quad (23)$$

式中: \mathbf{A}^{M} 、 \mathbf{A}^{N} 分别为配变侧与用户侧的负荷数据; $\mathbf{A}_{t,i}^{\text{M}}$ 表示配变 i 在 t 时刻的负荷; $\mathbf{A}_{t,j}^{\text{N}}$ 表示用户 j 在 t 时刻的负荷。

在任意时刻 t , 对于配变 i , 有式(24)所示的功率平衡关系。

$$\mathbf{A}_{t,i}^{\text{M}} = \sum_{j \in \Omega_i} \mathbf{A}_{t,j}^{\text{N}} + \varepsilon_{t,i} \quad (24)$$

式中: $\varepsilon_{t,i}$ 表示功率损耗与量测误差; Ω_i 表示从属于配变 i 的所有用户集合。

为描述该配网下 m 个配变与 n 个用户的连接关系, 设定连接关系矩阵 $\mathbf{Y}_{m \times n} = [y_{i,j}]_{m \times n}$, 其中 $y_{i,j}$ 为矩阵 \mathbf{Y} 中的元素。 \mathbf{Y} 是由 0、1 组成的二元矩阵, 用于表示用户 j 与配变 i 是否存在从属关系, 1 代表存在, 0 代表不存在。

对连接关系矩阵的求解可以视为一个多元线性回归问题, 将式(24)转化为线性回归的表达形式, 列出多元线性回归方程组如式(25)所示。

$$\mathbf{A}^{\text{N}} * \mathbf{Y} = \mathbf{A}^{\text{M}} - \varepsilon_{t,i} \quad (25)$$

式中, “*” 代表卷积运算。

为了保证问题的可解性与精确性, 使用优化求解的思路解决这个线性回归问题。构建优化目标函数 $J(\mathbf{Y})$ 如式(26)所示。

$$J(\mathbf{Y}) = \min \left(\sum_{j \in M} \sum_{t=1}^T \left\| \mathbf{A}^{\text{M}} - \mathbf{A}^{\text{N}} * \mathbf{Y} \right\|^2 \right) \quad (26)$$

式中, M 表示用户集合。

由于配变和用户数量较多, 为提高优化问题的可解性和计算效率, 使用式(27)将二元变量 $y_{i,j}$ 松弛为区间(0,1)上的连续变量 a_{ij} 。

$$a_{ij} = \frac{\exp(y_{i,j})}{\sum_{i=1}^n \exp(y_{i,j})} \quad (27)$$

在模型优化结果中, 变量越趋近于 1, 则表明更有可能存在从属关系。总体上, 若各变量趋近于 0 或 1, 说明有较好的优化效果。反之优化结果越趋近于中间值, 优化效果就越差。

结合实际情况, 增加一定约束条件, 在正常情况下, 配变的供电曲线与其所属用户的总耗电曲线有类似的波形。设定约束条件为

$$\begin{cases} \rho_i = \frac{\text{cov}(A_i^M, A_i^N * Y_i^M)}{\sigma(A_i^M)\sigma(A_i^N * Y_i^M)} \\ \rho_i \leq \xi \end{cases} \quad (28)$$

式中: ρ_i 代表配变 i 所对应的相似系数, ρ_i 的取值小于 1, 其值越趋近于 1, 相似度越高; A_i^M 表示配变 i 的量测功率序列; $A_i^N * Y_i^M$ 表示配变 i 下属用户功率之和; ξ 为相似度阈值。

综合式(26)—式(28), 得到优化目标函数及约束条件表示为

$$\begin{cases} \min \sum_{j \in M} \sum_{l=1}^T \|A_j^M - A_j^N * Y\|^2 \\ \text{s.t. } y_{i,j} \in \{0,1\} \\ \rho_i \leq \xi \end{cases} \quad (29)$$

将前置步骤中改进 FCM 算法的结果作为初值代入式(29), 使用 Gurobi 求解器对该优化问题进行求解, 当误差小于设定值或迭代次数达到上限时, 算法结束, 解出符合要求的连接关系矩阵 Y 。

4 算例分析

4.1 验证方法设计

本文采用淮安某地区数据, 包含 4 台配变和 424 户终端用户, 以 15 min 为间隔的智能电表实际测量数据为例, 其中 1 号配变下属 106 户, 2 号配变下属 108 户, 3 号配变下属 94 户, 4 号配变下属 116 户, 其中安装户用光伏的用户数量共 283 户。量测包括 2020 年 1 月 1 日到 2020 年 1 月 31 日用户侧与配变侧实时量测负荷、实时量测电压、天气数据、户用光伏出力、初始户变关系数据。标准数据集记录的户变关系如表 1 所示。

表 1 标准数据集信息

配变	初始台账用户数	用户编号	户用光伏安装数量
配变 1	106	NO.1—106	74
配变 2	108	NO.107—214	53
配变 3	94	NO.215—308	89
配变 4	116	NO.309—424	67
总计	424	—	283

4.2 基于电压聚类的效果分析

本节仅使用聚类算法初步判别用户与配变的连接关系。一方面验证改进 FCM 聚类算法的有效性, 另一方面验证光伏出力对聚类算法效果的影响。

首先基于改进 FCM 聚类算法辨识户变关系。将聚类识别结果与标准数据集进行对比, 若从从属关系相同, 则认为识别结果正确。将本文的改进 FCM 聚类算法与其他聚类算法进行对比, 验证本文算法的有效性。为判断户变识别结果的准确性, 比较识别结果与标准数据集间的偏差。除了准确率之外, 本文采用 F1 分数这一指标作为户变关系辨识效果的评价标准, 结果如表 2 所示。

表 2 聚类方法识别结果统计

算法	是否指定聚类数	准确率/%	F1 分数
K-means	是	51.4	0.679
标准 FCM	是	54.2	0.703
改进 GMM	是	65.1	0.789
改进 FCM	否	67.9	0.809

表 2 的识别结果表明, 由于初值选取的随机性, 标准 K-means 算法和标准 FCM 算法的效果并不稳定。本文的改进算法有效规避了随机性的问题, 基于密度 Canopy 的改进在确定聚类数的同时确定了初始聚类中心和隶属度矩阵, 在不指定聚类数目的条件下实现更快的收敛。

由表 2 可知, 单纯基于电压聚类的方法辨识准确率最高仅为 67.9%, 无法满足实际需求。准确率较低的原因在于分布式光伏大规模接入使各个节点的电压特征趋于相似, 节点特征不够显著, 因此, 电压聚类算法仅能作为辅助验证步骤。

4.3 基于最小互信息的光伏特征提取分析

本节验证采用最小互信息理论的光伏特征提取算法。算法输入为一台光伏表的数据、配变总表数据, 将算法分解出的光伏出力与实际光伏出力进行对比。动态变量 $a=0.2$, 卡尔曼滤波参数 $Q=1 \times 10^{-4}$, $R=0.1$ 。光伏出力分解效果见图 3。

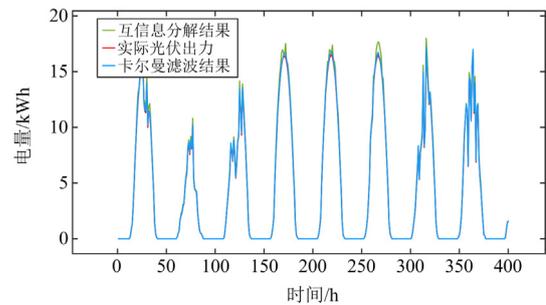


图 3 光伏出力分解效果

Fig. 3 Decomposition effect of photovoltaic output

由图 3 可以看出, 单纯以最小互信息为标准进行光伏特征提取, 其效果受噪声影响较大。而经过卡尔曼滤波之后的光伏出力结果与实际光伏出力基本重合, 满足进一步识别的需求。

4.4 本文方法的识别结果分析

本节综合改进 FCM 聚类及最小互信息算法, 通过多元线性回归分析, 对台区户变关系进行精确辨识。对比辨识结果与标准数据集的差异, 分析有无 FCM 聚类对辨识结果的影响, 分析有无最小互信息分解对辨识结果的影响。阈值参数 $\zeta=0.9$ 。辨识结果如图 4 所示。

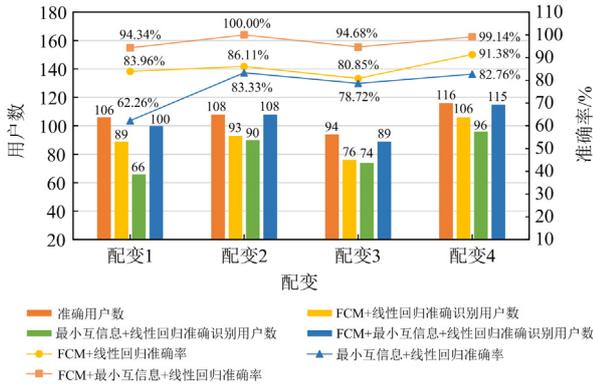


图 4 户变关系识别结果

Fig. 4 Identification results of transformer-customer relationship

将本文方法的识别结果用连接关系矩阵表示, 对应的热力图如图 5 所示。图中, 横轴为用户, 纵轴为配变, 两者对应的区域代表连接关系, 线条的颜色深浅程度表示配变与用户间的连接关系强弱, 连接关系越强, 颜色越深, 邻接矩阵中的元素取值越趋近于 1, 该用户就更有可能属于该配变。无 FCM 聚类及无最小互信息分解辨识方法所对应的热力图分别如图 6 和图 7 所示。

由图 4—图 7 可知, 本文方法可以将 424 户用户准确匹配至对应的配变, 正确判别出大部分用户与台区之间的关系, 识别准确率达 97.1%。然而,

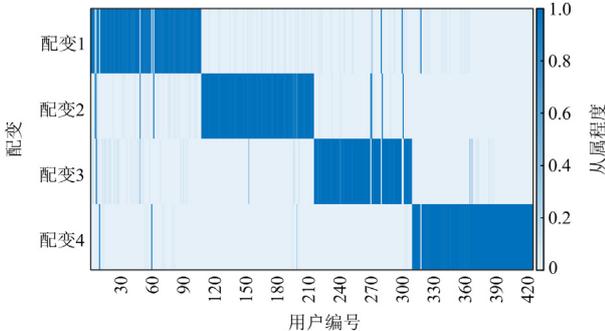


图 5 本文方法辨识结果热力图

Fig. 5 Hotspots map of identification results of this paper

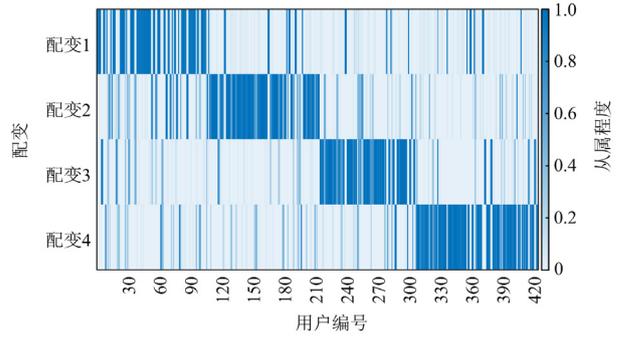


图 6 无 FCM 聚类结果热力图

Fig. 6 Hotspots map of identification results without FCM algorithm

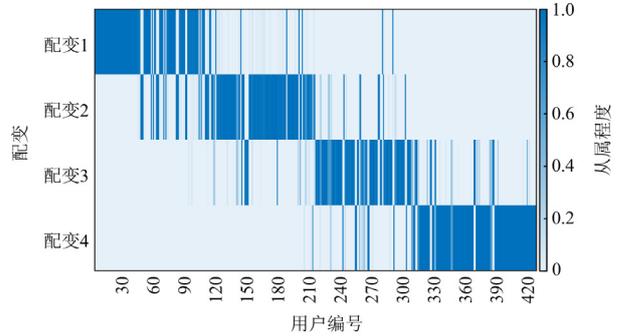


图 7 无最小互信息分析辨识结果热力图

Fig. 7 Hotspots map of identification results without minimum mutual information

如图 4 所示, 在未经过最小互信息算法分解光伏出力, 直接采用净负荷代入计算的情况下, 由于光伏出力对电压和功率特征的影响, 辨识准确率仅为 85.8%。未使用改进 FCM 算法初聚类筛选用户的方法无法准确辨识户变关系, 准确率仅为 76.9%。准确率较低的原因可能是用户数量较大, 仅采用功率数据建立的线性回归模型易产生维度爆炸, 无法找到最优解。

4.5 不同算法识别结果对比

在本文场景中使用 GIS+电压波形相关性分析方法^[21]及地址聚类+GMM 聚类+关联卷积方法^[22]进行户变关系辨识, 并与本文采用的算法进行对比, 结果如表 3 所示。对比算法对应的热力图如图 8 和图 9 所示。

表 3 各户变关系识别方法对比

Table 3 Comparison of different transformer-customer identification methods

方法	准确率/%	F1 分数
GIS+电压波形相关性	83.7	0.911
地址聚类+GMM 聚类+关联卷积	89.6	0.945
本文方法	97.1	0.982

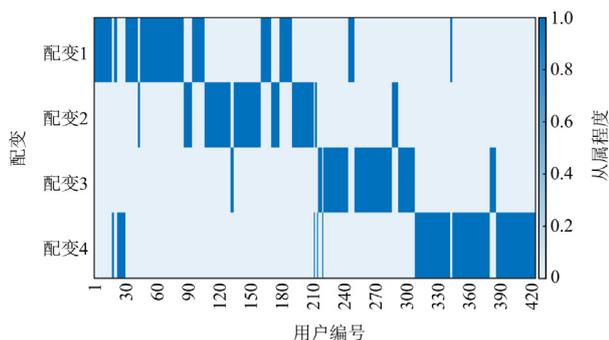


图 8 GIS+电压波形相关性辨识结果热力图

Fig. 8 Hotspots map of identification results using GIS+voltage waveform correlation

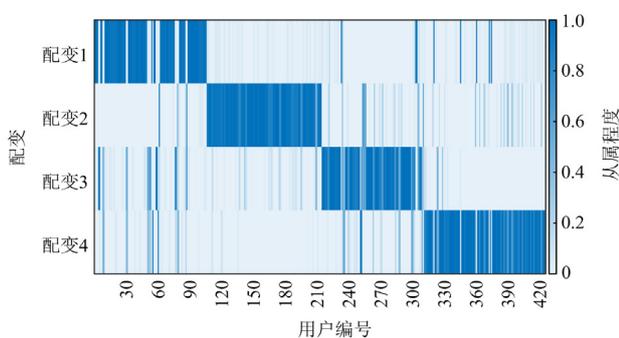


图 9 地址聚类+GMM 聚类+关联卷积方法辨识结果热力图

Fig. 9 Hotspots map of identification results using address clustering+GMM clustering+correlation convolution

分析表 3 数据, 对比图 5 与图 8、图 9 可知, 基于 GIS+电压波形相关性的识别方法的准确率仅为 83.7%, 而基于地址聚类+GMM 聚类+关联卷积的准确率可以达到 89.6%。准确率较低的原因包括: (1) 未对光伏出力进行处理, 电压和功率特征不明显; (2) 地址信息相近的用户可能实际上从属于不同配变。这些问题在本文方法中得以规避, 本文的改进 FCM 聚类和最小互信息算法, 能够获得 97.1% 的准确率, 可以有效地应用于工程实践中。

5 结论

本文针对实际工程中台区户变关系不明确的问题, 考虑高渗透率分布式光伏对量测数据的影响, 结合传统的数据驱动方法, 提出一种低压台区户变关系辨识方法。该方法基于实际工程原理, 能够较好地挖掘用户电压与电量特征, 其识别结果拥有较高的准确度。本文的工作主要包括以下内容:

1) 提出基于密度 Canopy 的改进 FCM 算法, 利用电压相关性进行初聚类, 为后续精确计算提供了初值, 降低了计算量并提高了精确性。

2) 提出了最小化互信息方法, 平抑分布式光伏对量测数据的干扰, 有效解决了用户分布式光伏不断增加的场景下户变关系的识别问题。

3) 综合考虑电压与功率量测数据特征, 精确辨识户变关系。有效利用了多维度数据并综合聚类算法和线性回归算法的优势, 避免了以电气距离为单一判据带来的误差。

本文所提辨识方法仍存在一定的局限性和改进空间: 本文辨识的用户均为单相用户, 没有考虑到三相用户的影响, 三相不平衡电网中不对称负荷的渗透率较高, 可能对本文方法造成一定影响, 未来还需进一步研究。

参考文献

- [1] 张博, 唐巍, 蔡永翔, 等. 面向高比例户用光伏消纳的储能系统与通信网络协同规划[J]. 电网技术, 2018, 42(10): 3161-3169.
ZHANG Bo, TANG Wei, CAI Yongxiang, et al. Collaborative configuration of energy storage systems and communication networks for accommodation of high-penetration residential PVs[J]. Power System Technology, 2018, 42(10): 3161-3169.
- [2] 杨虎臣, 王晓东, 史志鹏, 等. 基于电压相似性评估的低压配电网户变关系和相位识别技术研究[J]. 电测与仪表, 2023, 60(2): 91-96.
YANG Huchen, WANG Xiaodong, SHI Zhipeng, et al. Research on household variation relation and phase recognition technology of low voltage distribution network based on voltage similarity assessment[J]. Electrical Measurement & Instrumentation, 2023, 60(2): 91-96.
- [3] ZHAO G, CHU J, DENG L, et al. Research on line-transformer-user topological anomaly recognition model based on multi-source data mining[C]// 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), June 4-7, 2020, Chengdu, China: 192-196.
- [4] 陈斌, 李昀宸, 马仲坤, 等. 低压配电网台区多端口能量路由器优化配置研究[J]. 供用电, 2023, 40(2): 32-38.
CHEN Bin, LI Yunchen, MA Zhongkun, et al. Research on optimal configuration of multi-port energy router in low voltage distribution network[J]. Distribution & Utilization, 2023, 40(2): 32-38.
- [5] 仇继扬, 李涵, 王泽忠, 等. 含分布式光伏的低压台区三相线损理论计算方法研究及其应用[J]. 供用电, 2023, 40(11): 69-75.
QIU Jiyang, LI Han, WANG Zezhong, et al. Research and application on three-phase line loss improvement algorithm for low voltage distribution network with distributed photovoltaic and its application[J]. Distribution & Utilization,

- 2023, 40(11): 69-75.
- [6] 刘浩, 赵伟, 温克欢, 等. 低压配电台区模块化智能融合终端构建方案与实现[J]. 电测与仪表, 2022, 59(1): 168-175.
LIU Hao, ZHAO Wei, WEN Kehuan, et al. Construction scheme and realization of modularized intelligent fusion terminal in low-voltage distribution station areas[J]. Electrical Measurement & Instrumentation, 2022, 59(1): 168-175.
- [7] 吴善, 郝思鹏, 杨李星, 等. 基于 CIM 的智能配电台区信息模型及应用[J]. 电测与仪表, 2018, 55(10): 46-51, 89.
WU Shan, HAO Sipeng, YANG Lixing, et al. Information model and application of intelligent distribution area based on CIM[J]. Electrical Measurement & Instrumentation, 2018, 55(10): 46-51, 89.
- [8] FANG X, YANG Q, YAN W. Power generation maximization of distributed photovoltaic systems using dynamic topology reconfiguration[J]. Protection and Control of Modern Power Systems, 2022, 7(3): 508-522.
- [9] MOTEPE S, TWALA B, STOPFORTH R. Determining South African distribution power system big data integrity using fuzzy logic: power measurements data application[C] // 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), November 30-December 1, 2017, Bloemfontein, South Africa: 139-143.
- [10] 赵腾, 张焰, 张东霞. 智能配电网大数据应用技术与前景分析[J]. 电网技术, 2014, 38(12): 3305-3312.
ZHAO Teng, ZHANG Yan, ZHANG Dongxia. Application technology of big data in smart distribution grid and its prospect analysis[J]. Power System Technology, 2014, 38(12): 3305-3312.
- [11] 刘羽霄, 张宁, 康重庆. 数据驱动的电力网络分析与优化研究综述[J]. 电力系统自动化, 2018, 42(6): 157-167.
LIU Yuxiao, ZHANG Ning, KANG Chongqing. A review on data-driven analysis and optimization of power grid[J]. Automation of Electric Power Systems, 2018, 42(6): 157-167.
- [12] 周苏荃, 柳焯. 新息图法拓扑错误辨识[J]. 电力系统自动化, 2000, 24(4): 23-27.
ZHOU Suquan, LIU Zhuo. An innovation graph approach to topology error identification[J]. Automation of Electric Power Systems, 2000, 24(4): 23-27.
- [13] 王日宁, 武一, 魏浩铭, 等. 基于智能终端特征信号的配电网台区拓扑识别方法[J]. 电力系统保护与控制, 2021, 49(6): 83-89.
WANG Rining, WU Yi, WEI Haoming, et al. Topology identification method for a distribution network area based on the characteristic signal of a smart terminal unit[J]. Power System Protection and Control, 2021, 49(6): 83-89.
- [14] 唐冬来, 李强, 陈永东, 等. 基于电压时空聚合曲线的乡镇配电台区户变关系识别方法[J]. 电力系统自动化, 2023, 47(19): 86-96.
TANG Donglai, LI Qiang, CHEN Yongdong. Identification method for relationship between residential electricity meters and distribution transformers in township distribution station area based on voltage spatial-temporal aggregation curves[J]. Automation of Electric Power Systems, 2023, 47(19): 86-96.
- [15] 刘道伟, 章锐, 李柏青, 等. 基于状态量测的电网拓扑结构反向识别方法研究[J]. 中国电机工程学报, 2019, 39(24): 7107-7117, 7484.
LIU Daowei, ZHANG Rui, LI Baiqing, et al. Research on reverse recognition method of power system topological structure based on state measurement[J]. Proceedings of the CSEE, 2019, 39(24): 7107-7117, 7484.
- [16] 殷涵, 宣羿, 黄怡, 等. 基于虚拟阻抗的低压配电网拓扑识别方法[J]. 电力系统保护与控制, 2024, 52(3): 83-93.
YIN Han, XUAN Yi, HUANG Yi, et al. Virtual impedance-based low-voltage distribution network topology identification method[J]. Power System Protection and Control, 2024, 52(3): 83-93.
- [17] YU J, WENG Y, RAJAGOPAL R. PaToPa: a data-driven parameter and topology joint estimation framework in distribution grids[J]. IEEE Transactions on Power Systems, 2018, 33(4): 4335-4347.
- [18] ZHAO J, LI L, XU Z, et al. Full-scale distribution system topology identification using Markov random field[J]. IEEE Transactions on Smart Grid, 2020, 11(6): 4714-4726.
- [19] 张巍, 徐进, 刘家鹏. 基于分支有功功率的低压有源配电网拓扑辨识方法[J]. 电力系统自动化, 2022, 46(15): 85-95.
ZHANG Wei, XU Jin, LIU Jiapeng. Topology identification method for low-voltage active distribution network based on branch active power[J]. Automation of Electric Power Systems, 2022, 46(15): 85-95.
- [20] 崔雪原, 刘晟源, 金伟超, 等. 基于 APAA 和改进 DBSCAN 算法的户变关系及相位识别方法[J]. 电网技术, 2021, 45(8): 3034-3043.
CUI Xueyuan, LIU Shengyuan, JIN Weichao, et al. Consumer-transformer relationship and phase identification based on APAA and improved DBSCAN algorithm[J]. Power System Technology, 2021, 45(8): 3034-3043.
- [21] 黄旭波, 黄磊, 艾丹, 等. 基于 GIS 和电压波形相关性分析的低压配电网拓扑识别方法[J]. 哈尔滨理工大学学报, 2022, 46(15): 85-95.

- HUANG Xubo, HUANG Lei, AI Dan, et al. Topology identification method of low-voltage distribution network based on GIS and voltage waveform correlation analysis[J]. Journal of Harbin University of Science and Technology, 2022, 46(15): 85-95.
- [22] 徐明昕, 赵健, 王小宇, 等. 基于电压聚类 and 关联卷积的配电网户变关系识别方法[J]. 电力系统保护与控制, 2022, 50(4): 92-102.
- XU Mingxin, ZHAO Jian, WANG Xiaoyu, et al. Transformer-customer identification method for a low-voltage distribution network based on voltage clustering and incidence convolution[J]. Power System Protection and Control, 2022, 50(4): 92-102.
- [23] 陈碧云, 吕怡博, 梁志坚, 等. 考虑数据失衡的新型配电网两阶段拓扑辨识[J]. 电力系统保护与控制, 2023, 51(21): 57-65.
- CHEN Biyun, LÜ Yibo, LIANG Zhijian, et al. Two-stage topology identification of a new-type distribution network considering data imbalance[J]. Power System Protection and Control, 2023, 51(21): 57-65.
- [24] 许晓艳, 黄越辉, 刘纯, 等. 分布式光伏发电对配电网电压的影响及电压越限的解决方案[J]. 电网技术, 2010, 34(10): 140-146.
- XU Xiaoyan, HUANG Yuehui, LIU Chun, et al. Influence of distributed photovoltaic generation on voltage in distribution network and solution of voltage beyond limits[J]. Power System Technology, 2010, 34(10): 140-146.
- [25] WU S, HU W, DONG Y. A method for identifying connectivity relationship in low-voltage area based on voltage big data[C] // 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), May 21-24, 2019, Chengdu, China: 3624-3628.
- [26] SALEEM B, YANG W, GONZALES F M. Association rule mining for localizing solar power in different distribution grid feeders[J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2589-2600.
- [27] 朱音洁, 赵健, 宣羿, 等. 高比例户用光伏接入下低压配电网户变关系识别[J/OL]. 电网技术: 1-13 [2023-07-19]. DOI:10.13335/j.1000-3673.pst.2022.2371.
- ZHU Yinjie, ZHAO Jian, XUAN Yi, et al. Transformer-customer relationship identification of low-voltage distribution network with high proportion of household PV system[J/OL]. Power System Technology: 1-13 [2023-07-19]. DOI:10.13335/j.1000-3673.pst.2022.2371.
- [28] PAPPU S J, BHATT N, PASUMARTHY R, et al. Identifying topology of low voltage distribution networks based on smart meter data[J]. IEEE Transactions on Smart Grid, 2017, 9(5): 5113-5122.
- [29] 刘岩, 李文文, 周丽霞, 等. 基于高斯混合模型的光伏发电出力中高比例异常数据检测方法研究[J]. 电测与仪表, 2021, 58(9): 14-21.
- LIU Yan, LI Wenwen, ZHOU Lixia, et al. Research of high proportion of outliers detection method in photovoltaic power output data based on Gaussian mixture model[J]. Electrical Measurement & Instrumentation, 2021, 58(9): 14-21.
- [30] 孔祥玉, 胡启安, 董旭柱, 等. 引入改进模糊 C 均值聚类的负荷数据辨识及修复方法[J]. 电力系统自动化, 2017, 41(9): 90-95.
- KONG Xiangyu, HU Qi'an, DONG Xuzhu, et al. Load data identification and correction method with improved fuzzy C-means clustering algorithm[J]. Automation of Electric Power Systems, 2017, 41(9): 90-95.
- [31] 周治宇, 陈豪. 盲信号分离技术研究及算法综述[J]. 计算机科学, 2009, 36(10): 16-20, 31.
- ZHOU Zhiyu, CHEN Hao. Research and survey on algorithms of blind signal separation technology[J]. Computer Science, 2009, 36(10): 16-20, 31.
- [32] WILDEBOER R R, SCHOOT B C, WIJKSTRA H, et al. Blind source separation for clutter and noise suppression in ultrasound imaging: review for different applications[J]. IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control, 2020, 67(8): 1497-1512.

收稿日期: 2023-12-12; 修回日期: 2024-02-28

作者简介:

李彦伦(2000—), 男, 博士研究生, 研究方向为电力系统稳定与控制; E-mail: 18952731328@qq.com

窦晓波(1979—), 男, 通信作者, 教授, 博士生导师, 研究方向为分布式电源高渗透配电网、分布式发电与微电网、分布式电源(储能)变流器优化控制。E-mail: dxh_2001@sina.com

(编辑 许威)