

引用格式:赵耀,付皖皖,陈冉,等.基于角度分段线性近似和改进密度峰值聚类的户变关系识别[J].电力科学与技术学报,2025,40(1):113-125.

Citation: ZHAO Yao, FU Wanwan, CHEN Ran, et al. Meter-to-transformer relationship identification based on APLR and ICFSFDP[J]. Journal of Electric Power Science and Technology, 2025, 40(1): 113-125.

基于角度分段线性近似和改进密度峰值聚类的户变关系识别

赵耀¹, 付皖皖¹, 陈冉², 张涛¹

(1. 上海电力大学电气工程学院, 上海 200090; 2. 国网上海市电力公司电力科学研究院, 上海 200437)

摘要:为解决因排查效率低、数据更新不及时等因素导致低压配电网户变关系连接形式与实际不符的问题,提出一种基于角度分段线性近似(angle piecewise linear representation, APLR)和改进密度峰值聚类(improved clustering by fast search find of density peaks, ICFSFDP)相结合的户变关系识别方法。首先,根据电压曲线中相邻线段的角度变化量提取曲线的转折点,利用APLR对曲线进行自适应降维重构;随后,使用ICFSFDP算法对降维数据组展开聚类分析,在决策图中由拟合函数与坐标轴围成面积的最小值得到最优类簇数目,进而得到聚类和非聚类中心用户;最后,使用动态时间弯曲(dynamic time warping, DTW)距离计算聚类和非聚类中心用户之间的距离相似度,进而得到户变关系。将所提方法应用于模拟和真实数据中,均可证实所提方法的有效性。算例分析结果表明:该方法能够对时间间隔不同、不等维的序列进行分析,且不需要人为设定聚类算法的参数,户变关系识别准确率高。

关键词:配电网;户变关系;改进分段线性近似;动态时间弯曲距离;最小面积法;密度峰值聚类

DOI: 10.19781/j.issn.1673-9140.2025.01.012 中图分类号: TM743 文章编号: 1673-9140(2025)01-0113-13

Meter-to-transformer relationship identification based on APLR and ICFSFDP

ZHAO Yao¹, FU Wanwan¹, CHEN Ran², ZHANG Tao¹

(1. College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China; 2. Electric Power Research Institute, State Grid Shanghai Municipal Electric Power Company, Shanghai 200437, China)

Abstract: Factors like low troubleshooting efficiency and untimely data updates make meter-to-transformer wiring relationships in low-voltage distribution networks deviate from the actual situation. To address this issue, a meter-to-transformer relationship identification method based on the combination of angle piecewise linear representation (APLR) and improved clustering by fast search find of density peaks (ICFSFDP) is proposed. Initially, inflection points in the voltage curve are extracted by analyzing the angle variations between neighboring segments, and the curve undergoes adaptive dimensionality reduction and reconstruction using APLR. Then, the ICFSFDP method is deployed to cluster the data sets after dimensionality reduction, and the optimal number of clusters is determined by identifying the minimum area enclosed by the fitted function and the coordinate axis within the decision graph. This allows the identification of central clustered and non-clustered consumers. Finally, the dynamic time warping (DTW) distance is utilized to measure the distance similarity between the central clustered and non-clustered consumers, obtaining meter-to-transformer relationships. The application of this method on both simulated and real data has validated its effectiveness. Results from the analytical cases indicate that this approach can analyze sequences with varied time intervals and dimensions without the need for manually setting clustering algorithm parameters, delivering a high accuracy in identifying meter-to-transformer relationships.

Key words: distribution network; meter-to-transformer relationship; improved piecewise linear representation; dynamic time warping distance; minimum area method; clustering by fast search find of density peaks

收稿日期: 2024-03-01; 修回日期: 2024-04-17

基金项目: 国家自然科学基金(52377111); 国家电网有限公司科技项目(B3094023000D); 上海市自然科学基金(21ZR1425400)

通信作者: 赵耀(1987—), 男, 博士, 副教授, 主要从事低压配电网拓扑辨识、动态感知等研究; E-mail: nihaozhaoyao@163.com

户变关系是指用户电表与变压器/台区的连接关系,是低压拓扑关系研究中的核心内容,也是低压配电网研究的基础。但是在线路改造过程中,用户随意接线导致户变关系连接复杂,拓扑的变化信息上传不及时、漏报误报等问题普遍存在^[1-5],且缺少有效的校验方案,亟须研究准确高效的户变关系识别方法^[6-8]。

目前,低压拓扑识别方法主要分为2类:物理通信法^[9-15]和数据分析法^[16-24]。基于物理通信的方法主要通过电表的高速电力线通信(high-speed power line communication, HPLC)技术^[9-12]实现户变关系识别。但是基于HPLC技术的识别方法有3个缺点:①发出的信号抗干扰能力差且容易互扰;②对采集器、集中器等硬件要求较高;③识别准确率受限于传感器和智能仪表等装置的测量精度。若台区用户电压的波动很大时,则此方法的识别率较低^[13],且需要投入一定的人力,效率较低。而数据分析法是根据智能电表得到的用户电压、电流等数据,通常采用相关性分析、大数据及邻接矩阵分析等方法识别中低压拓扑结构。数据分析法与物理通信法相比,不受限于设备的精度,准确率较高。配电变压器与其连接的用户电表在电气量上的相关性是拓扑识别的一个关键参数,文献[16-18]分别采用Tanimoto相似度、T型灰色关联度和Pearson分析法对电压进行分析;文献[19]建立线性回归方程求解拓扑与线路参数;文献[20]提出一种多粒度聚类 and 多元特征统计的低压配电网拓扑辨识和检测方法。

上述列举的数据分析法大多研究的是中低压线路之间的连接关系,而对于用户与变压器之间的户变关系鲜有研究。近几年来从变压器和用户电气量相似性的角度出发,通过聚类算法可以得到更准确的户变连接关系。文献[21]通过自组织映射算法和K-均值算法结合形成两阶段的聚类算法对户变关系进行识别;文献[22]根据自适应分段聚合近似方法提取电压特征,随后使用改进的密度聚类算法得到户变关系;文献[23]采用t-随机近邻嵌入方法对归一化后的时序电压进行降维,再利用基于层次的聚类方法实现单相用户辨识。然而,随着时间的推移,用户随意接线次数增加,导致采集到的数据差异大,使用传统的聚类算法效果不好;同时对于不同的台区数据,聚类算法每次计算都需要人为调整参数,不利于实际应用。因此,要求聚类算法不仅要能够发现任意形状的簇,还要有参数自适

应能力。

基于此,本文提出一种基于角度分段线性近似(angle piecewise linear representation, APLR)与改进密度峰值聚类(improved clustering by fast search find of density peaks, ICFSFDP)算法的低压配电网户变关系辨识方法。首先,该方法以角度值和变化段为依据寻找转折点,使用bottom-up全局算法实现对原始数据集的自适应重构,降低电压序列的维度;随后,对上述得到的降维曲线进行聚类分析,并提出在决策图中使用最小面积法得到聚类数目和聚类中心;最后,使用动态时间弯曲(dynamic time warping, DTW)距离计算聚类和非聚类中心用户之间的距离,进而得到户变连接关系。通过将角度和变化段引入分段线性近似中,本文所提方法可以克服原算法局部性的问题;采用ICFSFDP聚类算法可以避免传统聚类算法初值选择与聚类数人为选择的难题;通过最小面积法确定最优决策值,使得决策值选择有自适应能力,可以提高聚类结果的准确度。

1 角度分段线性近似

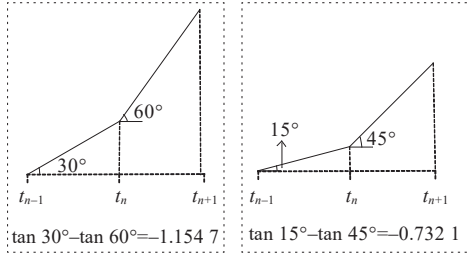
1.1 算法流程

为了保留电压数据的主要变化特征,减少冗余数据给计算带来的复杂度,采用时间序列分段线性近似(piecewise linear representation, PLR)算法对数据进行特征提取以降低维度、提高效率^[25]。

传统PLR算法是通过判断相邻采样点值的变化量是否大于所设定的阈值,然后把点依次连接即完成曲线的降维重构^[26]方法,以此在原曲线中找到“分段点”,但该分段点的提取方法受阈值和扰动的影响较大,且仅反映局部特征,缺乏全局性。基于斜率提取边缘点的时间序列分段线性表示方法则根据表征曲线变化程度的斜率变化量寻找分段点^[27],但是当角度变化量相同时,对应的斜率变化量可能是不同的,如图1所示(t_n 表示第 n 个时间点, t_{n-1} 、 t_{n+1} 同理),当起始角度、角度变化量均为 30° 时,斜率变化量为 -1.1547 ;当起始角度为 15° 、角度变化量为 30° 时,斜率变化量为 -0.7321 ,相同的角度变化量斜率的变化值不同。

为了解决上述分段点的局部性和相同变化角度斜率变化量不同的问题,根据APLR算法,通过引入角度变化量代替斜率变化量来寻找分段点。该方法考虑全局性对相邻分段点进行组合再与阈值比较形成变化段,算法流程如图2所示,其中数学

符号含义如下: φ 为电压曲线数据的角度变化量阈值, σ 为极值变化段阈值, β_k 为 u_k 点的前后角度变化值, Δ 为2个转折点 f_p 之间的差值, F 为电压转折点集合, D_k 为降维得到的分段点集合。



(a) 起始 30° ,变化量 30° (b) 起始 15° ,变化量 30°

图1 变化角度相同不同结构斜率变化量不同
Figure 1 Different slope variations for different structures with same angle change

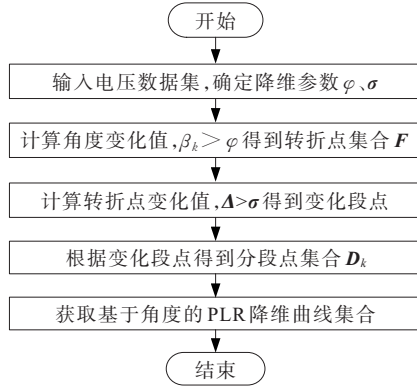


图2 APLR算法流程

Figure 2 Flow chart of APLR algorithm

1.2 算法实现步骤

本文所提方法的输入是智能电表获取的电压序列,形成 j 维 i 列的数据矩阵:

$$C = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,i} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,i} \\ \vdots & \vdots & \ddots & \vdots \\ u_{j,1} & u_{j,2} & \cdots & u_{j,i} \end{bmatrix} \quad (1)$$

式中, i 为某用户日电表采集的电压数据总个数; j 为台区用户总数; $u_{j,i}$ 为第 j 个用户第 i 个采样点的电压数据。

第 k 个用户的时间序列为 $C_k = [c_{k1}, c_{k2}, \dots, c_{ki}]$,经过本文所提APLR算法降维后 C_k 变化为 e 维的向量 $D_k = [d_{k1}, d_{k2}, \dots, d_{ke}]$ 。

1) 根据角度变化值得到转折点。

首先计算电压曲线中2点之间的角度值:

$$\begin{cases} \alpha_{k1} = \arctan(u_{j,k} - u_{j,k-1})/\Delta t \\ \alpha_{k2} = \arctan(u_{j,k+1} - u_{j,k})/\Delta t \end{cases} \quad (2)$$

式中, $u_{j,k}$ 为第 j 个用户第 k 个采样点的电压数据值。

弧度制 α_k 转化为角度制 $\hat{\alpha}_k$,即

$$\hat{\alpha}_{ki} = \alpha_{ki} (180/\pi), i=1, 2, \quad (3)$$

计算角度变化量, β_k 表示 $u_{j,k}$ 点的角度变化值:

$$\beta_k = \begin{cases} |\hat{\alpha}_{k1} - \hat{\alpha}_{k2}|, & \hat{\alpha}_{k1} \cdot \hat{\alpha}_{k2} > 0 \\ |\hat{\alpha}_{k1}| + |\hat{\alpha}_{k2}|, & \hat{\alpha}_{k1} \cdot \hat{\alpha}_{k2} < 0 \end{cases} \quad (4)$$

判断 β_k 与所设定的角度阈值 φ 的关系,若大于 φ 则为电压转折点 F_i ,否则不是转折点。最终得到电压转折点集合 $F = [F_1, F_2, \dots, F_n]$ 。

2) 根据转折点得到变化段。

将 F 中的相邻元素做差,得到转折变化量集合 $f = [f_1, f_2, \dots, f_p, \dots, f_{n-1}]$ 。子集 f_p 是步骤1)得到的相邻转折点 F_{p+1} 与 F_p 的差值。步骤1)充分考虑了局部性,但未兼顾全局性,因此,可以引入极值段的概念解决这一问题。当 f_p 为极小值时,取变化段最小值一端的值为变化段点;当 f_p 为极大值时,取变化段最大值一端的值为变化段点。与极值变化段阈值 σ 进行比较得到极值段为

$$\Delta = |f_p - f_{p+1}| > \sigma \text{ 或 } \Delta = |f_p - f_{p-1}| > \sigma \quad (5)$$

3) 根据变化段得到分段点序列。

$$d_{ki} = \begin{cases} \max(c_{k,i}, c_{k,i+1}), & f_p \text{ 为极大值} \\ \min(c_{k,i}, c_{k,i+1}), & f_p \text{ 为极小值} \end{cases} \quad (6)$$

式中, $c_{k,i}$ 为式(1)中第 k 个用户第 i 点的时间序列,当 $i=1$ 时表示采样点时刻为00:05:00。 d_{ki} 形成分段点集合 $D_k = [d_{k1}, d_{k2}, \dots, d_{ke}]$ 。

通过直接使用角度变化值代替斜率值,保留了角度变化明显的分段矩阵,降低了曲线平稳时的数据维度和识别算法对用户间电压特征异同性的挖掘难度。此外,算法难度较低,在使用角度变化量反映局部特征后,又提出使用转折点的变化段来反映全局特征,阈值确定也较为灵活,对实际工程中智能电表采集的频率和台区数量等参数的适应性较高。

1.3 降维效果分析

基于重要点的PLR算法^[28]和APLR算法对96个采样点的电压曲线进行分析,原曲线在0~40和40~80采样时间段的电压值分别呈现凸形和凹形。若采取重要点的PLR算法,曲线维度降至29维,得到降维曲线 M_1 ,如图3(a)所示,可以看出,该算法将波形低谷时期的最低点电压采纳为特征点,导致重构不准确,且含有大量的扰动点,无法准确反映原始曲线的特性。APLR算法对日电压曲线进行降维

的效果如图 3(b)所示,维度降低到 33,大部分的波动点、极值点都可以清晰展示,能更好地反映曲线的全局性。

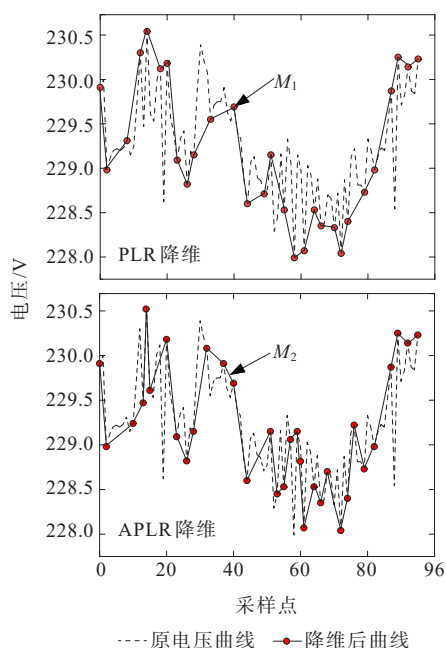


图 3 2 种算法降维效果

Figure 3 Dimensionality reduction performance of two algorithms

相似度与距离在本质上是一个物理量,即分析相似度时可采用距离公式^[25](推导过程参见文献[29]附录 A)来计算,即

$$d^2(X, Y) = 2n - 2 \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x^2 \sigma_y^2} \quad (7)$$

式中, $d^2(X, Y)$ 为样本序列 X, Y 之间的距离平方值; μ_x 和 μ_y , σ_x^2 和 σ_y^2 分别为样本序列 X, Y 的均值和方差。

将式(7)代入样本差和距离平方之间以及 Pearson 相似度与方差均值之间的关系公式中,得到距离的平方与相似度 P 之间的关系为 $d^2(X, Y) = 2n - 2P$, 因此,本文使用 DTW 距离表征曲线之间的相似度。

将原始曲线 M 分别与降维后的曲线 M_1, M_2 进行拟合度分析, $d(M_1, M) = 8.741 18$, $d(M_2, M) = 8.683 26$, APLR 算法降维后所得曲线与原曲线拟合度更高,更能反映电压曲线的变化特性。在时间方面, M_1 与 M 运算时间为 5.274 ms, 而 M_2 与 M 曲线运算时间为 4.817 ms, 减少了 0.457 ms。

从式(1)电压曲线集合中随机取出一个电压曲线序列 N , 分别使用重要点 PLR 和本文所提 APLR 算法得到降维后的 N_1, N_2 曲线。分别计算 $M_1 N_1,$

$M_2 N_2, MN$ 曲线之间的 DTW 距离, 即可表征相似度关系, 计算结果如表 1 所示。

表 1 不同方法降维后相似度和运算时间对比

Table 1 Similarity and time comparison of different methods after dimensionality reduction

算法	曲线序列	DTW 值	运算时间/ms
未使用	MN	6.117 65	15.795 8
PLR	$M_1 N_1$	3.924 19	4.023 5
APLR	$M_2 N_2$	3.773 04	2.124 7

由表 1 可知, 不同算法得到的降维曲线可以显著降低 DTW 值。APLR 算法得到的降维曲线比原始曲线在 DTW 距离和时间上分别缩短了 1.6、7.4 倍; 且 APLR 相比于 PLR 算法, 降维后与原始曲线的 DTW 指标和运算时间上分别减少了 3.9%、47.193%。

综上, 本文所提 APLR 算法通过提取曲线的角度变化局部特征和极大、极小值的总体变化特征, 对原始电压曲线进行降维, 在保证相似度和运算时间上都提高的前提下能够尽可能地把曲线维度降到最低。该方法所得曲线拟合度较高、运算时间短, 适用于用户电压曲线的聚类计算。

2 基于 APLR 和 DTW 距离的 ICFSFDP 算法

2.1 面向户变关系的基于密度峰值的聚类算法

由于不同台区的用户电压序列会形成不同的集群, 对所有用户在配变台区层级进行聚类得到聚类中心, 而基于密度峰值的聚类 (clustering by fast search find of density peaks, CFSFDP) 算法可以自动确定聚类中心, 能够发现任意形状的簇, 所以 CFSFDP 算法与户变关系识别问题契合度很高。对于文 1.2 中得到的降维后曲线数据集 $D_k = [d_{k1}, d_{k2}, \dots, d_{ke}]$, 绘制其决策图得到聚类中心和聚类数目。引入局部密度 ρ_i 、数据点与高密度点之间的最小距离 δ_i 。

1) 局部密度。

基于截断核的局部密度为

$$\rho_i = \sum_{i \neq j} \chi(d_{i,j} - d_c) \quad (8)$$

式中, χ 为 0, 1 的逻辑判断函数。

基于高斯核的局部密度计算公式为

$$\rho_i = \sum_{j \neq i} \exp(-d_{i,j}^2/d_c^2) \quad (9)$$

式中, $d_{i,j}$ 为 i,j 之间的欧氏距离; d_c 为截断距离。本文中 d_c 是 $d_{i,j}$ 按照升序排列后的前 1%~2% 的距离^[30]。

在局部密度选择上,由于截断核计算时容易形成相同的局部密度值,导致无法进行下一步的计算和排序工作,因此本文采用高斯核函数进行局部密度的计算。

2) 最小距离

设 $\{a_i\}_{i=1}^s$ 是上述局部密度降序排列的序下标值,满足 $\rho_{a_1} \leq \rho_{a_2} \leq \dots \leq \rho_{a_s}$, 最小距离对应的数据点被称为数据点 i 的最近邻点。计算如下:

$$\delta_i = \begin{cases} \min_{j < i, a_j} (d_{a_i, a_j}), & i \leq s \\ \max_{j \geq 2} (\delta_{a_j}), & i = s \end{cases} \quad (10)$$

δ_i 是样本数据点 i 和高于 i 点密度且距离最小的样本点之间的距离,当数据点 j 是最高密度点时, δ_i 是数据集中与数据点 j 距离最大的数据点与其之间的距离^[31]。对于数据集中的每一个数据点都有相应的 2 个数据与之相对应形成一个数据点 (ρ_i, δ_i) 。把所有的 $\{(\rho_i, \delta_i) | i=1, 2, \dots, N\}$ 在坐标轴上以 ρ 为横轴、 δ 为纵轴画出其图像,该图即为决策图,根据决策图可以判断并选择出聚类中心。对 ρ_i, δ_i 分别进行最大、最小归一化处理:

$$\rho'_i = \frac{\rho_i - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \quad (11)$$

$$\delta'_i = \frac{\delta_{a_i} - \delta_{\min}}{\delta_{\max} - \delta_{\min}} \quad (12)$$

由式(11)、(12)得到的具体数值绘制决策图,但其结果不能清晰地得到聚类中心的信息,这里综合考虑后将两者的乘积作为指标^[32],即

$$\gamma'_i = \rho'_i \cdot \delta'_i \quad (13)$$

其中, γ'_i 为归一化后的决策值, γ'_i 越大表示作为聚类中心的可能性越大,只需要对其进行降序排列,在决策图中选出较大的点作为决策中心。确定聚类中心后给其分配标签,将其余数据点分配到距离最近且密度最大的标签中,即完成聚类。

spiral 数据集广泛应用于聚类算法的性能测试中,其包含 3 个类簇,呈现螺旋形状,故称为“螺旋数据集”;每个簇的数据都呈现流形分布,各个样本之间没有交叉。spiral 数据集原始状态如图 4(a) 所示; ρ_i, δ_i 间的关系如图 4(b) 所示,此时截断距离 $d_c = 1.74929$;将 ρ_i, δ_i 乘积作为变量绘制决策图,如图

4(c) 所示,可知决策图不能直观表示分类结果;按照降序排列的聚类结果如图 4(d) 所示,可以清晰地表示出数据被聚成了 3 类。

相同地,使用带噪空间基于密度的聚类(density-based spatial clustering of application with noise, DBSCAN)算法与通过点排序识别聚类结构的密度聚类(ordering points to identify the clustering structure, OPTICS)算法对 spiral 数据集进行分类,如图 5 所示, r_{eps} 表示密度聚类的搜索半径。

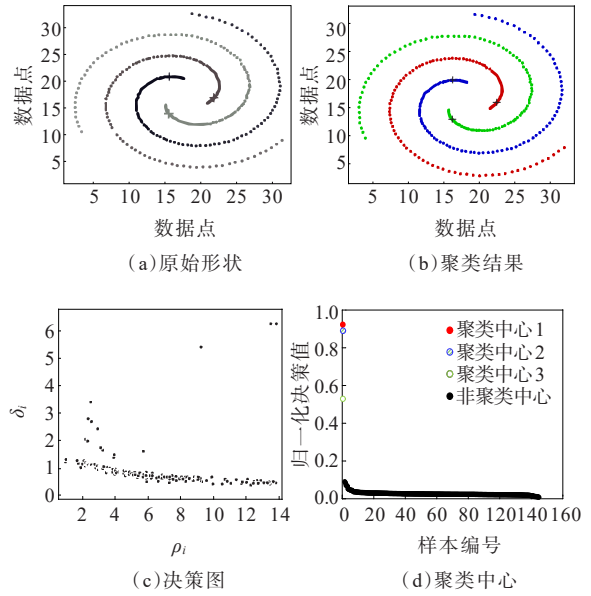


图 4 基于 CFSFDP 算法的 spiral 聚类

Figure 4 Spiral clustering based on CFSFDP algorithm

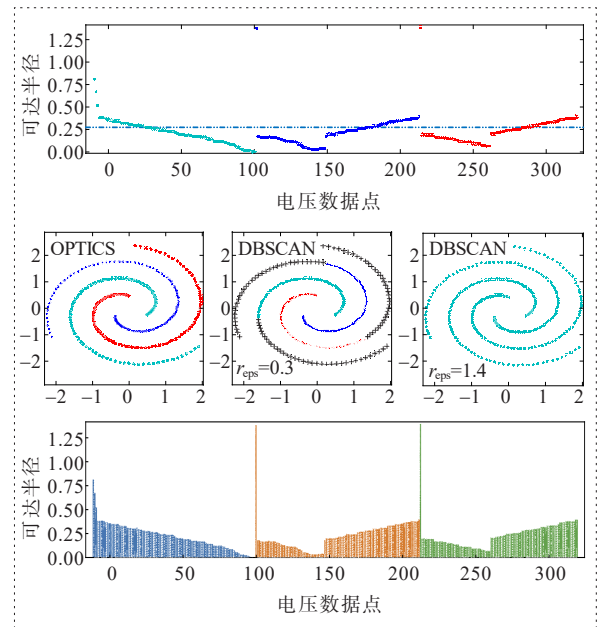


图 5 DBSCAN、OPTICS 在 spiral 数据集上的聚类

Figure 5 DBSCAN and OPTICS clustering on spiral dataset

由图5可知,DBSCAN、OPTICS把数据均分成了3类,但是DBSCAN受参数(r_{eps} 、 ϵ)的影响较大(ϵ 表示最小点数),若参数选择不合理则聚类效果较差,如 $r_{\text{eps}}=0.3$ 时数据聚类不充分, $r_{\text{eps}}=1.4$ 时数据聚类出现错误,即算法把数据分成了一类。OPTICS不能直接给出聚类结果,并且还需使用DBSCAN算法来进行分类,步骤较繁琐;相比于这2种算法,CFSFDP算法在参数设置、聚类效果方面有更加突出的优势,因此本文选择CFSFDP算法进行户变关系聚类分析。

2.2 CFSFDP算法改进

CFSFDP算法的第1步是计算欧式距离,这就要求输入的数据必须是等维矩阵,而PLR方法得到的降维曲线维度是不统一的,因此,本文采用DTW距离代替欧式距离使得输入矩阵可以不等维。此外,聚类中心的选择是根据决策图的分布情况人为确定的,无合理的标准,存在主观因素的影响,因此本文提出最小面积法寻找最优类簇中心。

1) 使用DTW距离进行局部密度的计算。

不同于欧氏距离需要严格根据相同时间点对应的序列值进行距离计算,DTW算法可以调整时间序列不同点元素的关系获得一条最优的弯曲路径,使得沿此路径累加的距离最少,用来衡量曲线整体之间的相似度。

对于任意给定的2个时间序列 m 、 n ,构建欧氏距离矩阵 $d_{i,j}$ 。弯曲路径是由距离矩阵中每一组相邻元素组成的集合,记作 $q=\{q_1, q_2, \dots, q_s, q_k\}$, k 为路径的总个数,如图6所示(DTW距离为0.048 677)。

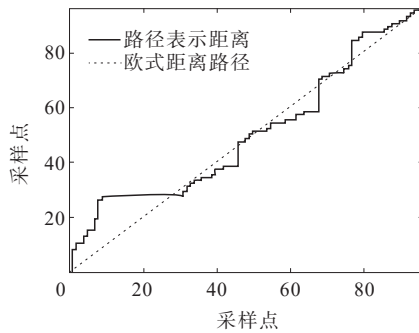


图6 DTW距离的动态规划示意

Figure 6 Dynamic programming of DTW distance

上述路径有很多条,DTW算法的目的是找到一条最优的路径,使得序列之间的累积距离最小。序列 A 、 B 之间的DTW距离定义为

$$d_{\text{DTW}(A,B)} = \min \sum_{s=1}^k D(q_s) \quad (14)$$

式中, $D(q_s)$ 为弯曲路径的距离。

用动态规划法构造矩阵 G 求解 $d_{\text{DTW}(A,B)}$,即

$$G_{i,j} = d_{i,j} + \min(G_{i-1,j-1}, G_{i,j-1}, G_{i-1,j}) \quad (15)$$

式中, $i \in [1, n]$, $j \in [1, m]$; G 为距离累积矩阵; $G_{i,j}$ 为 G 中第 i 行、第 j 列元素。 $G_{0,1}=0$ 、 $G_{i,0}=G_{0,j}=+\infty$,故序列 A 、 B 之间的DTW距离为 $d_{\text{DTW}(A,B)}=G_{n,m}$ 。

2) 面积最小法寻找 k_0 。

由于聚类中心的决策值 γ' 远大于其他数据点的 γ' 值,所以当 γ' 从大到小排序时,绘制在以横轴为排序序号点、纵轴为 γ' 的图像中会出现明显的“分层”现象,基于此,本文通过2条虚线对 γ' 进行分段线性拟合,寻找拟合直线与坐标轴围成的面积最小值 S'_{min} ,此时分段点之前的 γ' 即为最优类簇中心,如图7所示, S_1 、 S_2 表示阴影部分的面积。

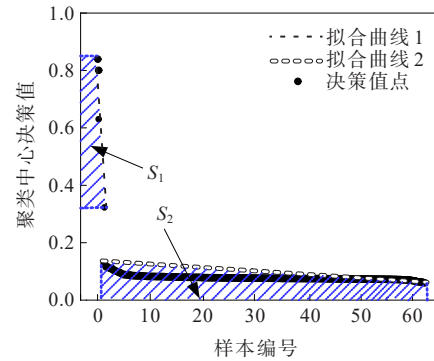


图7 决策值线性拟合与坐标轴围成的面积

Figure 7 Area enclosed by linear fit of decision values and coordinate axes

簇中心自动选择方法的具体步骤如下:

1) 将决策值 γ' 按照降序排列,以 n 个点 (i, γ_i) 的值作为数据集 L , $L=\{L_1, L_2, \dots, L_t, \dots, L_n\}$,且 L_1 对应最大 γ' , L_n 是数据集最后一个值且对应最小 γ' ;

2) 设置隔离点初始值 $t=2$,分别过 L_1 、 L_t 和 L_t 、 L_n 作直线,这2条直线需尽可能多地接近决策图上的点,此时图中的点分成了2层;

3) 计算拟合直线与坐标轴包含的面积,计算公式分别为

$$S_1 = (\gamma_1 - \gamma_t) \cdot (1 + t) / 2 \quad (16)$$

$$S_2 = (\gamma_n + \gamma_t) \cdot (n - t) / 2 \quad (17)$$

$$S' = S_1 + S_2 \quad (18)$$

4) 令 $t=t+1$,当 $t>n$ 时,转到步骤5),否则转到步骤2);

5) 找到最小 S'_{\min} , t 即为最优聚类中心数 k_p , 前 t 个样本所代表的 γ 即为聚类中心 ξ , 即聚类中心集合为 $\xi = \{\xi_1, \xi_2, \dots, \xi_t\}$ 。

2.3 户变关系识别整体框架

首先, 本文基于 APLR 算法对原始数据集进行降维处理; 然后, 使用 ICFSFDP 算法确定数据集的聚类数目和聚类中心; 最后, 计算聚类中心曲线和其余曲线之间的 DTW 距离, 根据 K 最近邻 (K-nearest neighbor, KNN) 算法将其余曲线归入所在类别内。本文总体技术路线如图 8 所示, 算法具体步骤如下。

1) 对配电变压器及用户数据的预处理, 包括台账信息及数据的规范化处理。对于少量电压数据

缺失的采样集合 C_k , 运用牛顿插值法处理缺失部分的采样数据。充分利用可信的档案信息, 基于地理信息对相邻的用户进行初步归纳合并, 合并以后的数据用于后续户变关系识别部分的分析。对于无法获取低压拓扑台账的信息, 其识别过程没有基于地理信息的用户初步合并的步骤。

2) 提取数据特征。对数据集 C 进行 APLR 处理, 得到降维曲线集合 $D_k = [d_{k1}^T, d_{k2}^T, \dots, d_{ke}^T]$ 。

3) 曲线聚类得到聚类中心。对数据集进行 ICFSFDP 聚类, 得到聚类个数 R 和聚类中心 $\xi = \{\xi_1, \xi_2, \dots, \xi_R\}$, 基于 KNN 算法把非聚类中心曲线归入与其相似度最大的类别中, 即完成户变关系识别。

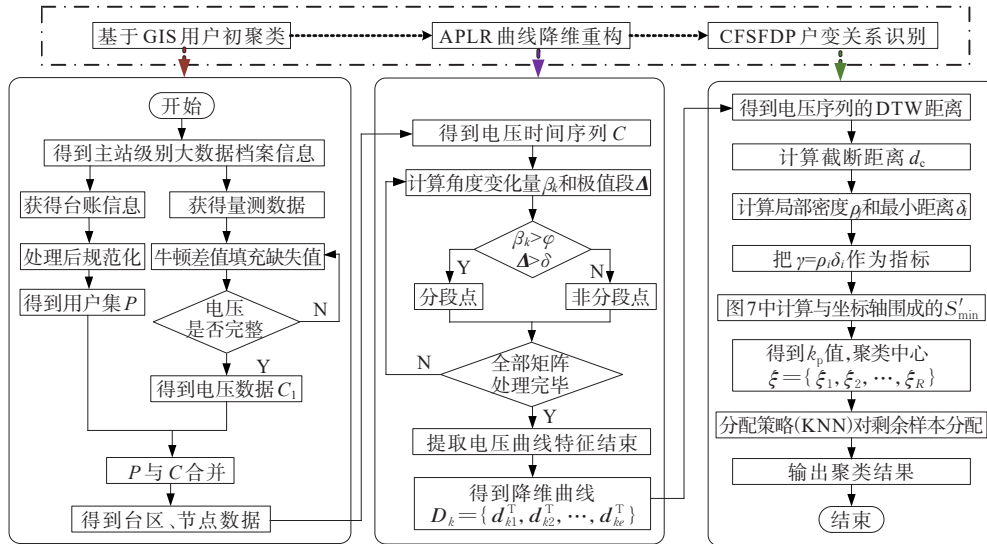


图 8 总体技术路线

Figure 8 Overall technical roadmap

3 算例分析

3.1 模拟数据识别结果

3.1.1 低压台区拓扑结构

首先使用 pandapower 建立台区的仿真模型, 假定各用户的功率在 $[1, 3]$ kW 内随机变化, 功率因数在 $[0.90, 0.95]$ 内随机变化, 由仿真模型可获得 24 h 的台变和用户节点的电压数据。然后根据上述条件建立低压拓扑结构, 如图 9 所示, 该拓扑结构由 T_1 、 T_2 和 T_3 共 3 个台区组成, T_1 台区中第 2 至第 8 节点、 T_2 台区中第 2 至第 7 节点、 T_3 台区中第 2 至第 8 节点均属于用户的电压节点; 算例分析中采样周期分别设置为 5、15、30 min 和 1、2 h。

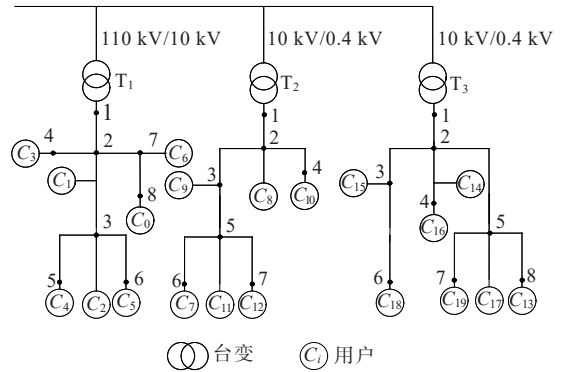


图 9 配电台区拓扑模型

Figure 9 Topology model of distribution substation area

3.1.2 用户聚类分析

为了验证本文所提聚类算法的正确性, 与主流的聚类算法 DBSCAN 作为对照组进行比较。

APLR算法中一共有角度变化值 φ 和变化段阈值 σ 这2个参数, φ 是为了寻找角度变化明显的点, φ 过小时会使分段点数量激增; φ 过大时时间序列中适合分段的点会被忽略,本文取 $\varphi=20^\circ$; σ 被用于过滤噪声,本文取 $\sigma=0.1$ 。当用户样本平衡时,台区 T_1 、 T_2 、 T_3 分别有7、6、7个用户,根据式(18)确定本文算法的聚类个数。

T_1 、 T_2 台区的节点电压波形如图10所示, T_{2-A} 和 T_{2-B} 分别表示台区2中的A、B相用户,可知用户电压曲线 T_{2-A} 和 T_{2-B} 整体上变化趋势相似,电压值重叠部分多。但是 T_1 、 T_2 台区内用户的变化趋势和电压相差很大。

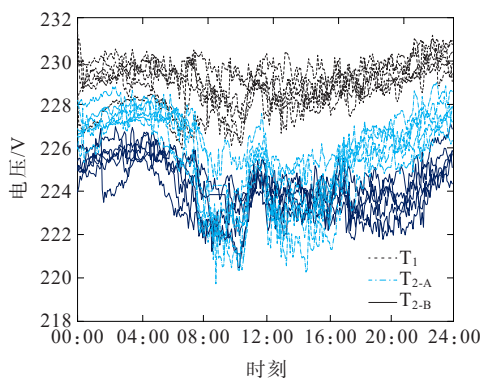


图10 3个低压台区用户电压序列

Figure 10 Household voltage sequences in three low-voltage substation areas

本文所提方法的户变关系聚类决策如图11所示,此时截断距离 $d_c=0.006\ 09$,该方法将原始数据集 D 映射到由局部密度 ρ_i 和最小距离 δ_i 组成的二维特征空间中,代表聚类中心的样本点应该同时具有较大的 ρ_i 和 δ_i ,通过特征空间决策图的可视化可以实现基于中心的聚类划分。由图11可知,样本点5、11、14是远离集群点的,代表 ρ_i 和 δ_i 的值都很大,很有可能是聚类中心。

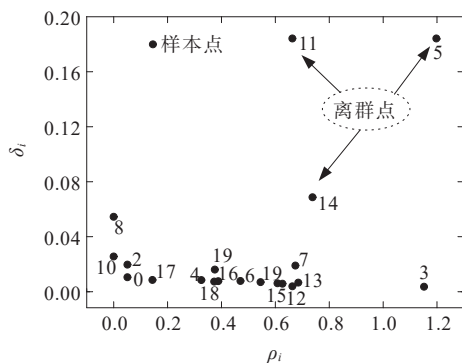


图11 ICFSFDP的决策图

Figure 11 Decision graph based on ICFSFDP

根据文2.2中ICFSFDP算法,对不同的分层点 l 对应的 S_1 、 S_2 进行计算并求和,当面积和最小时,得到最优聚类中心 $\xi=\{\xi_1, \xi_2, \dots, \xi_l\}$,如表2所示,可知 $S'_{\min}=0.528\ 05$,对应的最优聚类中心 ξ 为 $\{1, 2, 3\}$,不同分层点对应的面积如图12所示。为了在决策图中更加直观地看出聚类和非聚类中心,绘制样本编号与决策值 γ 图像,如图13所示,可以根据最小面积得到聚类中心,20个用户被分成了3类聚类中心,分别是用户5、11和14,且分别属于台区 T_1 、 T_2 和 T_3 ,可知本文所提方法存在3个较大的 γ ,能识别出3个聚类中心。

表2 分层点和聚类中心的候选值

Table 2 Candidates for hierarchical points and cluster centers

分层点	S_1	S_2	S'_{\min}	聚类中心 ξ
2	0.000 00	1.344 00	1.344 00	1, 2
3	0.147 59	0.380 46	0.528 05	1, 2, 3
4	0.374 99	0.153 52	0.528 51	1, 2, 3, 4
5	0.519 46	0.075 36	0.594 82	1, 2, 3, 4, 5
6	0.643 65	0.058 60	0.702 25	1, 2, 3, 4, 5, 6
7	0.756 24	0.056 24	0.812 48	1, 2, 3, 4, 5, 6, 7
8	0.865 65	0.052 97	0.918 62	1, 2, 3, 4, 5, 6, 7, 8
9	0.975 57	0.053 72	1.029 29	1, 2, 3, 4, 5, 6, 7, 8, 9

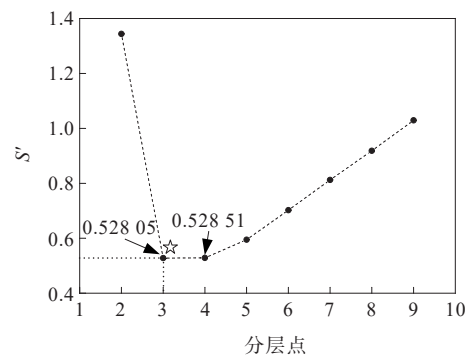


图12 不同分层点对应的面积

Figure 12 Areas for different hierarchical points

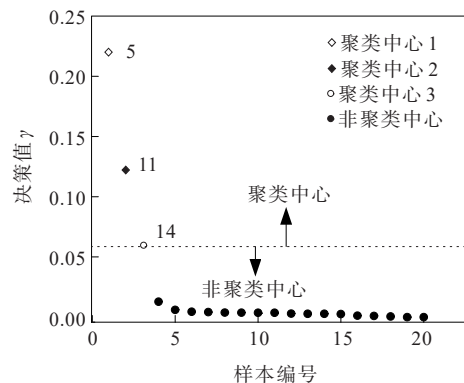


图13 ICFSFDP聚类中心决策

Figure 13 Cluster center decision based on ICFSFDP

对聚类中心用户 5、11、14 和非聚类中心用户的电压序列之间计算 DTW 距离。采样周期分别为 0.25、0.5、1、2 h, 20 个节点电压序列之间的 DTW 距离(具体参见文献[29]附录 B 表 1、2), 可知 DTW 距离能够对聚类结果中的中心和非中心用户的电压序列相似性进行准确的分析, 并且不受电压采样周期的影响。聚类结果如图 14 所示, 最终识别结果与图 13 的拓扑结构一致, 验证了本文所提方法的有效性。

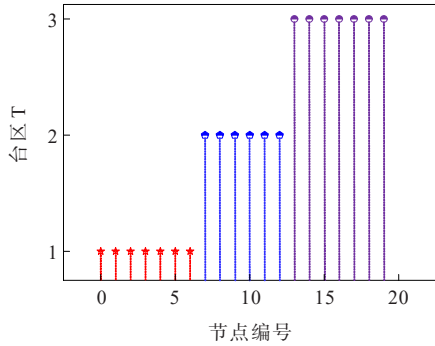


图 14 ICFSFDP 聚类结果

Figure 14 ICFSFDP clustering results

3.1.3 采样周期、噪声比例对算法灵敏度的影响

工程实际中电表的采样周期和误差的不同使得电压数据在维度和噪声程度上存在差异。在 15 min、30 min、1 h、2 h 这 4 种采样周期下, 本文分析噪声比例设置为 10%、50% 时的识别结果, 其中, 10% 的噪声比例表示在数据中有 10% 的用户加入 1% 的高斯噪声, 50% 的噪声比例同理。不同采样周期以及噪声比例下户变关系的识别结果如表 3 所示, 其中, N_{true} 、 N_{false} 分别表示户-变关系识别正确和错误的用户个数, R 表示由最小面积法得到的最优类簇数。

表 2 不同参数设置下的户变关系识别结果

Table 2 Identification of meter-to-transformer relationship under different parameter settings

采样周期	噪声比例					
	10%			50%		
	R	(N_{true}, N_{false})	准确率/%	R	(N_{true}, N_{false})	准确率/%
15 min	3	(20, 0)	100	3	(20, 0)	100
30 min	3	(20, 0)	100	3	(19, 1)	95
1 h	3	(19, 1)	95	2	—	—
2 h	3	(18, 2)	90	2	—	—

由表 3 可知, 当噪声比例维持在较低水平(10%)时, 采样周期对准确率的不利影响随着周期变大而逐渐增加, 采样周期在 30 min 之内的聚类结果是完全正确的; 当噪声比例较大(50%)时, 采样

周期越大识别效果越差; 当数据含有极高比例噪声且采样周期大于 1 h 时, 算法不能得到正确的聚类数目, 导致户变识别结果出现错误。

3.1.4 算法性能检验

在经过最长公共子序列方法提取电压特征后, 将本文所提方法与使用 DBSCAN、OPTICS 算法识别户变关系的性能进行对比, 如图 15 所示, 识别率分为正确率和错误率两部分; 整体台区中不同算法识别结果如表 3 所示, 可以看到, DBSCAN、OPTICS 在 3 个台区的识别准确率分别为 85%、90%, 低于本文所提方法的识别准确率。

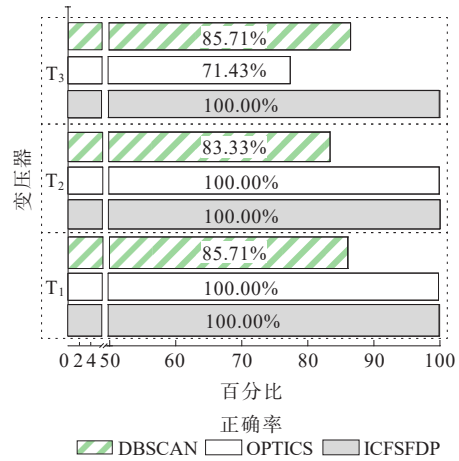


图 15 具体台区中不同算法识别正确率

Figure 15 Identification accuracy of different algorithms in specific substation areas

表 3 整体台区中不同算法识别结果

Table 3 Identification results of different algorithms in overall substation area

算法	采样周期/ (min·次 ⁻¹)	参数设置	待识别用 户数	准确率/ %
DBSCAN	5	$r_{eps}=0.8, \epsilon=7$	20	85
OPTICS	5	$r=0.5, \rho=4$	20	90
ICFSFDP	5	$\varphi=20^\circ, \sigma=0.1$	20	100

表 3 中 r 、 ρ 分别表示 OPTICS 的半径、密度。与 DBSCAN、OPTICS 算法相比, 本文所提方法在配电系统中具有更高的准确性。这是因为基于 APLR 曲线重构方法考虑了电压数据集的局部和全局特性, 不同配电系统电压簇的密度存在显著的差异, 因此本文所提方法能够满足户变关系的识别。DBSCAN、OPTICS 算法对输入参数敏感, 导致结果差异大; 通过提取电压特征, 本文所提方法提前知道聚类数并对数据进行密度峰值聚类, 具有较强的数据适应性和较好的聚类效果。

为了进一步验证 ICFSFDP 算法的有效性, 分别在样本数量平衡和不平衡的情况下进行户变关

系的聚类识别。当用户样本数量平衡时,由图 9 可知,台区 T_1 、 T_2 、 T_3 内分别有 7、6、7 个用户。当用户样本数量不平衡时,电网正常运行中负序与正序分量的均方根之比超过 2%,此时台区 T_1 、 T_2 、 T_3 内分别有 3、2、3 个用户。根据式(19)确定 CFSFDP 和 ICFSFDP 的聚类个数,DBSCAN、OPTICS 的聚类个数可以自发设定。样本数量平衡和不平衡下户变关系识别结果如图 16 所示,当样本数量平衡时,CFSFDP 和 ICFSFDP 算法均能正确得到 3 个决策中心,即数据中含有 3 个台区;当样本数量不平衡时,CFSFDP 算法只能得到 2 个决策中心,即 2 个台区,而 ICFSFDP 算法仍然可以得到 3 个决策中心即可以正确判断出含有 3 个台区。所以当用户数量不平衡时,直接聚类可能导致同一台区不同相的用户干扰户变关系的识别结果。

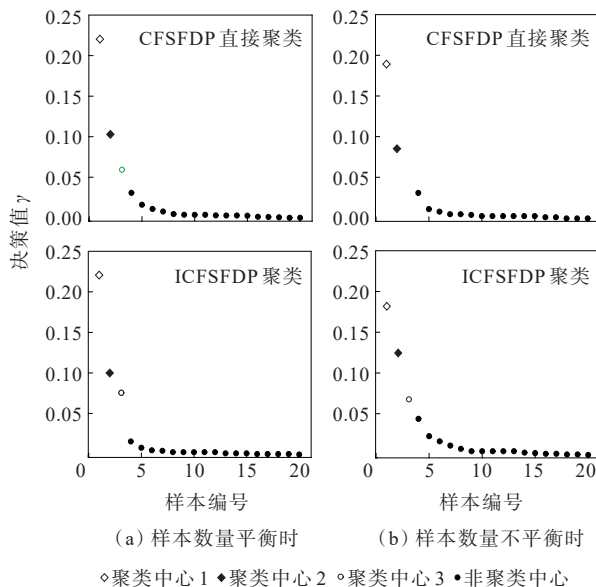


图 16 样本数量平衡和不平衡下户变关系识别结果

Figure 16 Identification results of meter-to-transformer relationships with balanced and imbalanced samples

使用 4 种方法进行识别的聚类结果如表 4 所示,CFSFDP、ICFSFDP 算法均能在数据平衡时得到正确的台区数量,内部所含样本数量均正确,但是 DBSCAN、OPTICS 结果出现错误。DBSCAN 在样本数量不平衡时错误地将 3 个台区的数据分成了 2 类;OPTICS 分类个数正确但是内部包含的数据个数出现错误;ICFSFDP 能在数据不平衡时正确识别出 3 类,并且每个类别中所含个数也均正确。可见,当样本数量不平衡度较大时,原始数据冗余将会直接导致聚类出现错误,通过利用 APLR 算法处理数据能有效提取数据的特征,根据密度峰值聚类发现聚类中心的方法能明显提升聚类精度。

表 4 户变关系聚类结果

Table 4 Clustering results of meter-to-transformer relationship

聚类 算法	样本数量平衡时			样本数量不平衡时		
	T_1	T_2	T_3	T_1	T_2	T_3
ICFSFDP	7	6	7	3	2	3
CFSFDP	7	6	7	3	4	2
DBSCAN	5	5	7	4	0	4
OPTICS	6	5	7	2	3	5

3.2 真实量测数据识别结果

在江苏省某地区选取 5 个低压配电网的变压器及用户电压值开展试点研究,典型农村和城市户变结构的地理位置如图 17 所示,#1~#5 表示 5 个台区。

选取 2023 年 1 月 30 日的电压数据(参见文献[29]附录表 C),采样时间间隔为 5 min。由于台区变压器是三相数据,故使用三相电压的平均值作为 5 个台区的台变电压平均值。绘制 5 个台区的三相电压平均值,如图 18 所示,可以看出, T_1 、 T_2 台区电压曲线比较接近, T_3 、 T_4 、 T_5 台区电压曲线比较接近。由图 17 也能看出, T_1 、 T_2 台区相邻, T_3 、 T_4 、 T_5 台区相邻。本文采用 APLR 算法得到降维曲线,使用 ICFSFDP 算法得到聚类和非聚类中心,最后计算聚类与非聚类中心间的 DTW 距离,根据此距离对台区和变压器进行分析,得到户变关系。



(a) 典型农村配电网



(b) 典型城市配电网

图 17 某地区低压配电网实际户变关系卫星图

Figure 17 Satellite image of actual meter-to-transformer relationship of low-voltage distribution network in a certain area

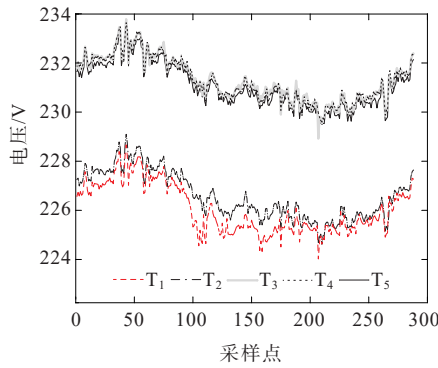


图 18 5 个台区变压器的电压平均值

Figure 18 Average voltage of five transformers in substation area

本文采用配电变压器低压侧出口处日冻结 288 点的电压数据绘制曲线,进行所提方法的验证,结果如表 5 所示,5 个试点的低压配电网中户变识别准确率均在 93% 以上,未能达到 100%。原因是,实际生产过程中一方面受限于电表时移存在误差,DTW 计算没有考虑时间延迟的影响,另一方面受电表间的计量误差是否符合标准的影响,使得采集的数据和理想数据之间存在差异,导致算法的准确率没达到理想状态。对比 DBSCAN、OPTICS 和 ICFSFDP 算法,结果如表 6 所示。对比分析可知,原始数据的冗余度高会使聚类效果下降,通过 APLR 和 ICFSFDP 算法的使用,提高了聚类的精度。

表 5 本文算法结果

Table 5 Results of proposed algorithm

台区	待识别用户数	算法识别准确个数	准确率/%
#1	9	9	100.00
#2	9	9	100.00
#3	108	101	93.52
#4	63	59	93.65
#5	40	38	95.00

表 6 算法结果对比

Table 6 Comparison of algorithm results

台区基本信息		准确率/%		
台区	待识别用户数	DBSCAN	OPTICS	ICFSFDP
T ₁	9	100.00	100.00	100.00
T ₂	9	100.00	100.00	100.00
T ₃	108	90.74	91.67	93.52
T ₄	63	90.47	92.06	93.65
T ₅	40	90.00	92.50	95.00

4 结语

为准确提取用户电压曲线的特征点,并提高现有户变关系聚类算法的聚类质量,本文提出基于 APLR 与 DTW 距离相似度衡量的 ICFSFDP 算法。

首先,基于 APLR 提取电压曲线的特征进行自适应降维;然后,使用基于 DTW 距离相似度衡量的 ICFSFDP 算法进行聚类;最后,根据改进的最小面积方法得到聚类分界值 k_p ,以准确找出聚类中心。本文所提方法在模拟数据和真实数据中进行了验证,并与其他算法进行了对比测试,得到如下结论:

1) 基于 APLR 算法,根据相邻线段的角度变化值作为选择转折点的依据,能解决选用极值点的过程中转折点提取不充分的缺点,又能避免直接使用斜率带来的非线性问题。

2) 采用 DTW 作为距离相似度衡量方式合理。此方法可以比较准确地衡量电压曲线间的动态特性以及不等维时间序列间的相似度,适用于 APLR 降维后的不等维数据组用于密度峰值聚类算法的分析中。

3) 提出一种密度峰值聚类簇中心自动选择的方法,直线拟合后使直线与坐标轴之间围成面积最小值的直线分界点作为聚类分界值 k_p ,避免了决策值人为选取的弊端。

参考文献:

[1] 马尚,卫志农,黄蔓云,等.基于智能电表数据的低压配电网拓扑与线路参数联合辨识[J].电力系统自动化,2024,48(2): 60-70.
 MA Shang, WEI Zhinong, HUANG Manyun, et al. Joint identification of topology and line parameters for low-voltage distribution network based on smart meter data[J]. Automation of Electric Power Systems,2024,48 (2): 60-70.

[2] 俞林刚,李铭,伍栋文,等.配电网拓扑参数未知场景下中压用户窃电检测方法[J].电网与清洁能源,2023,39(9):91-100.
 YU Lingang, LI Ming, WU Dongwen, et al. An electricity theft detection method of medium voltage users in the case of unknown for distribution network topology parameters[J]. Power System and Clean Energy, 2023, 39(9):91-100.

[3] 周宁,方治,蔺晨晖,等.基于正则化拉格朗日乘子法的配电网网络拓扑辨识[J].供用电,2024,41(6):12-20.
 ZHOU Ning, FANG Zhi, LIN Chenhui, et al. Topology identification of distribution network based on regularized Lagrangian multiplier[J]. Distribution & Utilization, 2024, 41(6): 12-20.

[4] 张国驹,裴玮,杨鹏,等.中压配电网柔性互联设备的电路拓扑与控制技术综述[J].电力系统自动化,2023,47(6): 18-29.
 ZHANG Guoju, PEI Wei, YANG Peng, et al. Review on circuit topology and control technology of flexible

- interconnection devices for medium-voltage distribution network[J]. Automation of Electric Power Systems, 2023, 47(6): 18-29.
- [5] 黄代雄,汪志军,袁俑斌,等.考虑多源协同的主动配电网故障恢复策略[J].高压电器,2024,60(2):210-215.
HUANG Daixiong, WANG Zhijun, YUAN Yongbin, et al. Fault recovery strategy considering multi-source collaborative active distribution network[J]. High Voltage Apparatus, 2024, 60(2): 210-215.
- [6] 张津铭,赵健,唐维溢,等.动态拓扑下基于终端数据挖掘的配电网异常节点检测[J].智慧电力,2024,52(7):1-9.
ZHANG Jinming, ZHAO Jian, TANG Weiyi, et al. Abnormal node detection in distribution network based on terminal data mining in dynamic topology[J]. Smart Power, 2024, 52(7): 1-9.
- [7] 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.
- [8] 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, LYU 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.
- [9] 王艳,王阳,赵洪山,等.中压配电网电力线载波通信信道建模[J].电力系统保护与控制,2021,49(8):50-57.
WANG Yan, WANG Yang, ZHAO Hongshan, et al. Channel modeling of power line communication in a medium voltage distribution network[J]. Power System Protection and Control, 2021, 49(8): 50-57.
- [10] 朱晓荣,司羽.考虑物理-信息-交通网耦合的配电网多时段动态供电恢复策略[J].电工技术学报,2023,38(12):3306-3320.
ZHU Xiaorong, SI Yu. Multi-period dynamic power supply restoration strategy considering physical-cyber-traffic network coupling[J]. Transactions of China Electrotechnical Society, 2023, 38(12): 3306-3320.
- [11] 邓淑斌,李嵘,梁志飞,等.考虑碳交易和绿证交易的配电网优化运行策略[J].电力建设,2023,44(10):149-156.
DENG Shubin, LI Rong, LIANG Zhifei, et al. The optimal operation strategy of distribution network considering carbon trading and green certificate trading mechanisms [J]. Electric Power Construction, 2023, 44(10): 149-156.
- [12] XU C, LEI Y, ZOU Y. A method of low voltage topology identification[C]//IEEE Conference on Telecommunications, Optics and Computer Science, Shenyang, China, 2020.
- [13] 唐冬来,倪平波,张捷,等.基于离散弗雷歇距离的户变关系识别方法[J].电力系统自动化,2021,45(6):223-230.
- TANG Donglai, NI Pingbo, ZHANG Jie, et al. Identification method for relationship between household transformers based on discrete frechet distance[J]. Automation of Electric Power Systems, 2021, 45(6): 223-230.
- [14] 焦昊,殷岩岩,吴晨,等.基于安全强化学习的主动配电网有功-无功协调优化调度[J].中国电力,2024,57(3):43-50.
JIAO Hao, YIN Yanyan, WU Chen, et al. Coordinated optimization of active and reactive power of active distribution network based on safety reinforcement learning[J]. Electric Power, 2024, 57(3): 43-50.
- [15] 荆志朋,柴林杰,胡诗尧.基于改进LSTM-VAE的配电网异常负荷检测方法研究[J].电测与仪表,2024,61(9):71-76.
JING Zhipeng, CHAI Linjie, HU Shiyao. Research on abnormal load detection method for distribution network based on improved LSTM-VAE[J]. Electrical Measurement & Instrumentation, 2024, 61(9): 71-76.
- [16] 吴龙腾,徐春华,邱泽坚,等.考虑双侧量测误差的配电网拓扑识别及参数联合估计方法[J].电力系统保护与控制,2023,51(16):149-157.
WU Longteng, XU Chunhua, QIU Zejian, et al. Topology identification and joint parameter estimation of a distribution network considering bilateral measurement errors[J]. Power System Protection and Control, 2023, 51(16): 149-157.
- [17] 陈沛东,曹华珍,何璇,等.中压配电网近邻交互式分布式拓扑辨识算法[J].电力工程技术,2023,42(2):139-146.
CHEN Peidong, CAO Huazhen, HE Xuan, et al. Distributed topology identification algorithm of medium-voltage distribution network based on neighboring interaction[J]. Electric Power Engineering Technology, 2023, 42(2): 139-146.
- [18] LUAN W, PENG J, MARAS M, et al. Smart meter data analytics for distribution network connectivity verification[J]. IEEE Transactions on Smart Grid, 2015, 6(4): 1964-1971.
- [19] 戴雪婷,李辉,何柏娜,等.配电网拓扑分析与线路参数辨识研究[J].电力科学与技术学报,2023,38(4):46-56.
DAI Xueting, LI Hui, HE Bona, et al. Research on topology analysis and line parameter identification of distribution network[J]. Journal of Electric Power Science and Technology, 2023, 38(4): 46-56.
- [20] 郭上华,王钢.基于多粒度聚类和多元特征统计的低压配电网拓扑识别与监测[J].电力自动化设备,2023,43(6):86-93.
GUO Shanghua, WANG Gang. Research on topology analysis and line parameter identification of distribution network[J]. Electric Power Automation Equipment, 2023, 43(6): 86-93.

- [21] 宋玮琼,郭帅.基于电压时序数据的配电台区户变关系智能识别[J].电力系统及其自动化学报,2022,34(1):48-55.
SONG Weiqiong, GUO Shuai. Intelligent recognition of user-transformer relationship in distribution station area based on time series voltage data[J]. Proceedings of the CSU-EPSA, 2022, 34(1): 48-55.
- [22] 崔雪原,刘晟源.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. Consumer-transformer relationship and phase identification based on APAA and improved DBSCAN algorithm[J]. Power System Technology, 2021, 45(8): 3034-3043.
- [23] 连子宽,姚力.基于t-SNE降维和BIRCH聚类的单相用户相位及表箱辨识[J].电力系统自动化,2020,44(8):176-184.
LIAN Zikuan, YAO Li. Phase and meter box identification for single-phase users based on t-SNE dimension reduction and BIRCH clustering[J]. Automation of Electric Power Systems, 2020, 44(8): 176-184.
- [24] 杜婉琳,王玲,罗威,等.基于深度强化学习的有源配电网电压分层控制策略[J].发电技术,2024,45(4):734-743.
DU Wanlin, WANG Ling, LUO Wei, et al. Voltage hierarchical control strategy of active distribution network based on deep reinforcement learning[J]. Power Generation Technology, 2024, 45(4): 734-743.
- [25] 刘洋,许立雄,王剑.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):4093-4104.
LIU Yang, XU Lixiong, WANG Jian. A high performance extraction method for massive user load typical characteristics considering data class imbalance[J]. Proceedings of the CSEE, 2019, 39(14): 4093-4104.
- [26] 宋军英,崔益伟,李欣然,等.改进分段线性表示与动态时间弯曲相结合的负荷曲线聚类方法[J].电力系统自动化,2021,45(2):89-96.
SONG Junying, CUI Yiwei, LI Xinran, et al. Load curve clustering method combining improved piecewise linear representation and dynamic time warping[J]. Automation of Electric Power Systems, 2021, 45(2): 89-96.
- [27] 詹艳艳,徐荣聪,陈晓云,等.基于斜率提取边缘点的时间序列分段线性表示方法[J].计算机科学,2006,25(11):139-142.
ZHAN Yanyan, XU Rongcong, CHEN Xiaoyun, et al. Time series piecewise linear representation based on slope extract edge point[J]. Computer Science, 2006, 25(11): 139-142.
- [28] PRAT K B, FINK E. Search for patterns in compressed time series[J]. International Journal of Image and Graphics, 2002, 2(1): 89-106.
- [29] 付皖皖. 论文附录 [EB/OL]. https://pan.baidu.com/s/1wNMimra8NsSgCRdgV_PXCg?pwd=lg2s (提取码:lg2s), 2025-01-18.
FU Wanwan. Appendix to the paper [EB/OL]. https://pan.baidu.com/s/1wNMimra8NsSgCRdgV_PXCg?pwd=lg2s (extraction code: lg2s), 2025-01-18.
- [30] 杨震,王红军,周宇.一种截断距离和聚类中心自适应的聚类算法[J].数据分析与知识发现,2018,2(3):39-48.
YANG Zhen, WANG Hongjun, ZHOU Yu. A clustering algorithm with adaptive cut-off distance and cluster centers [J]. Data Analysis and Knowledge Discovery, 2018, 2(3): 39-48.
- [31] 陈俊艺,丁坚勇,田世明,等.基于改进快速密度峰值算法的电力负荷曲线聚类分析[J].电力系统保护与控制,2018,46(20):85-93.
CHEN Junyi, DING Jianyong, TIAN Shiming, et al. An improved density peaks clustering algorithm for power load profiles clustering analysis [J]. Power System Protection and Control, 2018, 46(20): 85-93.
- [32] 王俊,肖辉,王家奇,等.基于SNN-密度峰值聚类算法的商业用户典型负荷模式提取[J].电力学报,2023,38(1):64-72.
WANG Jun, XIAO Hui, WANG Jiaqi, et al. Extracting Typical load patterns of commercial users based on SNN-density peak clustering algorithm [J]. Journal of Electric Power, 2023, 38(1): 64-72.