

基于密度聚类的低压台区归属关系及相位识别方法

闫东辉[✉]

(海南电网有限责任公司, 海南 三亚 572000)

摘要: [目的] 供电部门记录的正确的拓扑信息有助于工作人员监测电网信息, 分析故障, 优化电网运行以满足低压配电台区精益化、智能化管理的需要。目前, 各式新型用电设备及用户的加入使低压配电网结构呈现出持续变化的特征, 线路维护成本被大大提高。[方法] 为此, 提出基于密度聚类的低压台区归属关系识别方法。首先, 提取智能电表有效电压数据生成高维时序电压矩阵; 其次, 采用 t 分布随机近邻嵌入方法 (t -distributed Stochastic Neighbor Embedding, t -SNE) 对高维时序电压数据进行特征提取与降维; 然后, 应用基于数据密度的噪声应用空间聚类方法 (Density-Based Spatial Clustering of Applications with Noise, DBSCAN) 对降维后的数据进行聚类分析, 实现低压用户台区归属信息的识别; 最后, 对海南省三亚市某台区实际数据进行分析, 并将所提方法与其他主流的拓扑识别算法进行对比。[结果] 分析结果表明所提方法能够达到 95% 以上的台区识别准确率, 高于目前其他主流的拓扑信息识别方法。[结论] 文章中的方法在解决此类问题上具有有效性与优势性, 可以为实际工程应用提供参考, 为低压台区拓扑信息识别领域提供不一样的研究思路。

关键词: 低压台区; 电压数据信息; t 分布随机近邻嵌入方法; 基于数据密度的噪声应用空间聚类方法; 台区归属关系识别; 相位识别

中图分类号: TM7; TM727

文献标志码: A

文章编号: 2095-8676(2023)05-0149-08

开放科学(资源服务)二维码:



Identification of Low-Voltage Distribution Network Attribution Relationship and Phase Information Based on Density Clustering

YAN Donghui[✉]

(Hainan Power Grid Co., Ltd., Sanya 572000, Hainan, China)

Abstract: [Introduction] The correct topology information recorded by the power supply department can help the staff monitor the power grid information, analyze the faults, and optimize the operation of the power grid to meet the needs of lean and intelligent management of low-voltage distribution networks. At present, the addition of various new types of electricity-using equipment and users has caused the low-voltage distribution network structure to show a continuous change in characteristics, and the line maintenance cost is greatly increased. [Method] Therefore, the identification method of low-voltage distribution network attribution relationship based on density clustering was proposed. First, the effective voltage data collected by smart meters were extracted to generate a high-dimensional time-series voltage matrix. Then, the t -distributed Stochastic Neighbor Embedding algorithm (t -SNE) and Density-Based Spatial Clustering of Applications with Noise algorithm (DBSCAN) were applied to cluster the voltage data to achieve identification of low-voltage distribution network attribution relationship. Finally, the actual data of a low-voltage distribution network in Sanya City, Hainan Province were analyzed, and the proposed method is compared with other mainstream topology identification methods. [Result] The analysis results show that the proposed method can achieve more than 95% of identification accuracy, which is higher than other mainstream topology identification methods. [Conclusion] The proposed method is effective and advantageous in solving such problems, and can provide reference for practical engineering applications and offer a different research idea in the field of topology identification of low-voltage distribution network.

Key words: low-voltage distribution network; voltage data information; t -SNE; DBSCAN; identification of low-voltage distribution

network attribution relationship; phase identification

2095-8676 © 2023 Energy China GEDI. Publishing services by Energy Observer Magazine Co., Ltd. on behalf of Energy China GEDI. This is an open access article under the CC BY-NC license (<https://creativecommons.org/licenses/by-nc/4.0/>).

0 引言

根据电压等级,配电台区可分为高压配电台区、中压配电台区和低压配电台区。文章研究的低压配电台区是指 10 kV/400 V 公共配电变压器下的居民区。低压配电台区网络主要是一个径向网络,从配电变压器的低压侧出口开始,到每个用户的接入点结束。供电部门记录的正确的拓扑信息有助于工作人员监测电网信息,分析故障,优化电网运行,满足低压配电台区精益化、智能化管理的需要^[1]。

目前,各式新型用电设备及用户的加入使低压配电网络结构呈现出持续变化的特征。因此,开展低压台区拓扑关系识别研究具有重要的工程应用价值。

低压台区配电网拓扑识别类型可归纳为以下 4 种:(1)识别低压台区与 10 kV 馈线的从属关系,该层关系识别错误率较低;(2)识别低压用户与低压台区的从属关系,部分用户易被误识别到邻近台区;(3)识别低压用户与电表表箱的从属关系,部分用户易被误识别到邻近表箱;(4)识别低压用户的相位信息。

目前,主流的拓扑结构识别方法可分为人工检测法、信号标记法和数据分析法^[2-3]。传统的人工检测法和信号标记法需要投入大量的人力成本和设备维护成本,经济效益低^[4-6]。数据分析方法是基于先进的计量基础设施所收集的大量电气测量数据,如电压和电流^[7],通过分析每个电气量的空间和时间特征,得出每个节点的潜在连接关系。因此,与人工检测法和信号标记法相比,数据分析法更加高效、便捷和环保,是目前低压配电网拓扑识别问题的一个重要研究方向。因此,国内外大量专家学者都相继在拓扑识别领域展开了数据分析法的研究。

文献 [8] 和文献 [9] 识别配电网 10 kV 母线之间的连接关系,即不同联络线上隔离开关闭合或断开时所形成的不同拓扑结构。文献 [10] 提出了识别电能表与低压配电网一级分支线(即低压配电网首端出线)的归属关系的方法。基于 T 型灰色关联度和 K-最近邻算法,结合供电部门记录的原拓扑结构图,

文献 [11] 提出了识别低压台区和 10 kV 的从属关系以及识别低压用户与台区的从属关系的方法。文献 [12] 通过对智能电表采集的电气量数据的读取分析,提出了基于主成分分析及其图论解释来生成稳定状态的网络拓扑。文献 [13] 使用智能电表数据,根据每小时电压曲线的相关因素和相对幅度水平,检测变压器附近错误连接的用户。文献 [14] 和文献 [15] 利用高质量电表电量数据,分别基于分层聚类算法以及半监督约束聚类算法解决单一台区的用户相位信息识别问题。

总体而言,现有的拓扑识别方法大多只能识别低压配电网的一种拓扑关系,很少有方法能识别一个完整拓扑的两种以上的连接信息^[16]。同时,目前方法在应用到大规模低压台区上时,准确率仍然无法达到实际工程上的要求^[9]。

在此背景下,针对当前拓扑识别研究中缺少低压用户归属台区误挂问题的识别方案以及多台区混合用户数据下用户相序识别方法的研究,文章提出了基于密度聚类的低压台区归属关系及相位识别方法。文章首先在阐述了对用户电压数据进行预处理的方法;接着详述了文章所用的 t 分布随机近邻嵌入降维方法(t -distributed Stochastic Neighbor Embedding, t -SNE)与基于数据密度的噪声应用空间聚类方法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)的算法原理;最后,通过对实际数据进行算法应用分析,验证所提方法的有效性与实用性。

1 高维时序电压数据预处理

文章算法需要用到的有效数据信息主要有用户测量点编号或智能电表编号、采集数据对应的时间信息与采集到的用户电压信息。因此,需要从供电公司的用户用电数据文件中对该类有效数据进行有效提取并生成电压数据矩阵 U ,如式(1)所示。

$$U = \begin{bmatrix} U_{1,t_1} & U_{1,t_2} & \cdots & U_{1,t_q} \\ U_{2,t_1} & U_{2,t_2} & \cdots & U_{2,t_q} \\ \vdots & \vdots & \ddots & \vdots \\ U_{p,t_1} & U_{p,t_2} & \cdots & U_{p,t_q} \end{bmatrix} \quad (1)$$

式中:

U_{j,t_k} ——电表 j 在 t_k 时刻测量的电压幅值;

p ——每一台区用户数量或节点数相加的总和;

q ——在一定时间段内的采样点数量, 作为每个用户或节点的原始完整特征维度。

目前智能电表采集的数据精度还较低, 采集的电压数据样本标准不一, 给电压数据分析工作造成了不小的困难, 甚至会直接影响后期降维聚类的精准度。因此, 文章选用 Z-Score 方法对 U 进行归一化处理, 方法表达式如下:

$$U'_{t_j} = \frac{U_{t_j} - \mu(U_{t_j})}{\sigma(U_{t_j})}, j = 1, 2, \dots, n \quad (2)$$

式中:

U'_{t_j} ——电压经标准化处理后在 t_j 采样时刻下所有台区用户节点电压列向量;

U_{t_j} ——在 t_j 采样时刻所有台区用户节点量测电压值列向量;

$\mu(U_{t_j})$ ——在 t_j 采样时刻下所有台区用户节点的电压均值;

$\sigma(U_{t_j})$ ——在 t_j 采样时刻下所有台区用户节点的电压标准差;

n ——表计采集的用户节点电压采样时刻总数。

2 基于 t-SNE 的高维时序电压数据特征降维算法

未经处理的高维时序电压数据中包含有异常干扰信息, 会在台区辨识与相位辨识过程中引入误差, 影响准确率。而降维可以对高维数据内部的本质结构进行有效提取, 减少低可信度信息造成的误差, 提高辨识精度。同时, 数据维度的降低更便于计算机处理和可视化操作, 缓解了高维空间数据的维度灾难, 有效提高了辨识效率。

2.1 主流降维算法背景概述

在数据挖掘领域, 目前已经存在着大量成熟的

数据降维算法, 可以分为线性降维算法与非线性降维算法两大类, 主要区别在于降维算法使用的映射函数的不同。主成分分析法 (Principal Component Analysis, PCA) 作为经典线性降维算法的代表, 其基于数据线性变换原理将高维数据线性映射到低维坐标中^[17], 具有计算简便高效、不受标签限制的优点, 但无法通过修改算法输入参数等方法对算法的降维处理过程进行可变性的人为干预, 对数据泛化能力差, 不适于处理拓扑复杂用户量大的高维时序电压数据; 局部线性嵌入方法 (Locally Linear Embedding, LLE) 作为流形学习的代表, 其基于高维数据点与点之间的线性组合重构出保留点与点之间局部关系的低维数据^[18], 优点是能够学习高维数据的局部线性低维流形, 有利于数据点间关系辨识, 算法输入参数少, 调参简便, 缺点是降维数据受异常数据点影响较大, 数据可视化效果差, 不适于处理数据采集异常率较高的用户电压数据, 而 t-SNE^[19] 作为目前最好的非线性概率降维算法之一, 受异常数据点影响较小, 对高维数据特征保留完好, 数据可视化效果好, 适用于对低压配电网中的高维的复杂用户电压数据的降维处理。因此, 选择 t-SNE 算法作为文章所用的数据降维算法。

2.2 t-SNE 降维算法原理

本小节将重点阐述 t-SNE 算法的原理步骤。首先, 计算联合概率以表示各数据点之间的关联性, 分别计算条件概率 $p_{j|i}$ 与 $p_{i|j}$, 如式(3)、式(4)所示。

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)} \quad (3)$$

$$p_{i|j} = \frac{\exp(-\|\mathbf{x}_j - \mathbf{x}_i\|^2 / 2\sigma_j^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_j - \mathbf{x}_k\|^2 / 2\sigma_j^2)} \quad (4)$$

式中:

$\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k$ ——任意 3 个输入向量;

σ_i, σ_j ——以 \mathbf{x}_i 和 \mathbf{x}_j 为高斯分布中心的模型方差。

再利用式(5)计算高维数据的联合概率分布 p_{ij} :

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \quad (5)$$

式中:

n ——高维输入数据集中高维向量的个数。

接着利用式(6)计算高维数据的联合概率分布 q_{ij} :

$$q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)} \quad (6)$$

式中:

y_i, y_j, y_k ——经过初步降维得到的 3 个输出向量, 此时算法还未收敛。

而满足 $p_{ij} = q_{ij}$ 这一条件是高低维数据局部特征保留完整、变化差异小的充要条件, 因此利用该条件作为算法的目标函数构建条件, 利用式(7)来构建目标函数优化 p_{ij} 和 q_{ij} 之间的 KL 散度:

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j) (1 + \|y_i - y_j\|^2)^{-1} \quad (7)$$

最后筛选出 KL 散度最小的特征集合 $Y^T = \{y_1, y_2, \dots, y_n\}$ 。算法 3.1 提供了 t -SNE 算法的伪代码。

算法 3.1: t 分布随机近邻嵌入方法

输入: 高维数据集 $X = \{x_1, x_2, \dots, x_n\}$; 困惑度 Perp ; 学习率 η ; 迭代次数 T ; 动量 $\alpha(t)$

输出: 低维数据集 $Y^T = \{y_1, y_2, \dots, y_n\}$

步骤:

- 1) 利用式(3)、式(4)计算在给定 Perp 下的条件概率 p_{ji} 与 p_{ij} ;
- 2) 令联合概率 $p_{ij} = \frac{p_{ji} + p_{ij}}{2n}$;
- 3) 用正态分布 $N(0, 10^{-4})$ 随机初始化目标低维数据集 $Y^0 = \{y_1, y_2, \dots, y_n\}$;
- 4) 迭代求解, 从 $t=1$ 到 T , 执行步骤 5) 至 7);
- 5) 利用式(6)计算低维空间下的联合概率 q_{ij} ;
- 6) 利用梯度下降法式(7)计算梯度 $\frac{\partial C}{\partial y_i}$;
- 7) 对目标低维数据集进行更新 $Y^t = Y^{t-1} + \eta \frac{\partial C}{\partial Y} + \alpha(t)(Y^{t-1} - Y^{t-2})$;
- 8) 结束。

3 基于 DBSCAN 的低维电压特征数据密度聚类算法

聚类分析指的是把一个输入数据集按照某种特定的标准度量划分成不同的簇, 且在此标准下簇内数据样本差异性低, 不同簇之间的数据样本差异性高, 最终实现类别划分的功能。对于完成特征降维的低维电压特征数据集而言, 其内部特征差异呈现

同类小、异类大的特点。因此, 选用无监督学习中的聚类算法可以完成对用户数据的分类任务。

3.1 主流聚类算法背景概述

目前常用的数据聚类算法主要可以分为划分子聚类、密度聚类、层次化聚类等。 K 均值聚类方法(k -means)是经典划分子聚类算法之一, 聚类依据为数据间的欧式距离大小, 其优点在于调参简单, 计算简便, 但聚类精度低, 对异常数据较为敏感, 无法处理非凸性数据的聚类问题, 因此不适用于处理用户拓扑种类多样的低维电压特征数据; 在层次化聚类方法中, 以平衡迭代规约和聚类方法(Balanced Iterative Reducing and Clustering Using Hierarchies, BIRCH)^[20] 为例, 其聚类依据为建立分类树结构按照树的节点一层一层聚类, 聚类速度快, 但主要缺点在于无法处理非凸性数据的聚类问题, 且调参复杂, 对同一种情况下的数据集, 输入不同的算法参数会有较大的结果差异, 加之此聚类算法没有科学的机器调参方法进行聚类效果判据分析, 需要进行人工调参, 最佳聚类参数需要通过枚举和人工分析求出, 工程应用效率低, 不适用于处理工程应用中的大量台区用户数据; 而在密度聚类算法中, 应用广泛的 DBSCAN^[21], 其聚类依据为根据已降维的数据密度进行聚类, 适于对可视化效果好、密度均匀的凸性或非凸性数据进行聚类分析, 与前述中的 t -SNE 降维算法相容性高, 同时有着科学的机器调参方法进行聚类效果判据分析, 不需要人工分析最佳的聚类参数, 有着较高的工程应用价值, 能够精细地处理复杂的数据量大的低维电压特征数据。因此, 选择 DBSCAN 算法作为文章所用的数据聚类算法。

3.2 DBSCAN 聚类算法原理

DBSCAN 算法是最有效的密度聚类算法之一。算法将某一对象的 ε 邻域内含大于等于 MinPts 数量的数据点的对象称为核心对象, 将某一核心对象与其的 ε 邻域内的对象间的关系称为密度可达关系。

DBSCAN 首先会遍历找到所有的核心对象, 然后随机选择一个作为种子, 找到与其有着密度可达关系的所有样本点集合作为一个聚类簇, 循环执行上述步骤, 直至所有核心对象都完成归类。最后, 算法将少数游离在簇外的异常样本点标记为噪音点, 它们不属于任何一个聚类簇, 并输出不同数据点的

类别结果。由于 DBSCAN 算法可以轻松识别和过滤掉离群值,因此在密度较大的数据集中表现良好,且不需要预先指定聚类的数量。这使得它成为机器学习领域中非常有用的聚类算法之一。算法 3.2 提供了 DBSCAN 算法的伪代码。

算法3.2: 基于数据密度的噪声应用空间聚类方法

输入: 数据集 $D = \{p_1, p_2, \dots, p_n\}$; 描述邻域关系参数: 邻域半径 ϵ ; 邻域样本对象密度阈值 MinPts

输出: 聚类簇划分 $C = \{C_1, C_2, \dots, C_k\}$

步骤:

- 1) 标记 D 内所有对象为未访问对象;
- 2) 随机选择一个未访问对象 p_i ;
- 3) 标记 p_i 为已访问对象;
- 4) 如果 p_i 的 ϵ 邻域至少有 MinPts 个对象, 循环执行步骤 5) 至 12), 否则跳转执行步骤 13);
- 5) 创建一个新簇 C_j , 并把 p_i 添加到簇 C_j 中;
- 6) 设 N 是 p_i 的 ϵ 邻域内的样本集合, 遍历 N 中的每个点 p' ;
- 7) 如果 p' 是未访问对象, 循环执行步骤 8) 至 11), 否则遍历下一个对象;
- 8) 标记 p' 为已访问对象;
- 9) 如果 p' 的 ϵ 邻域至少有 MinPts 个对象, 循环执行步骤 10) 至 11), 否则跳转执行步骤 7);
- 10) 把这些数据对象添加到 N 中;
- 11) 如果 p' 还不是任何簇的成员, 就把 p' 添加到簇 C_j 中;
- 12) 保存簇 C_j ;
- 13) 标记 p_i 为噪音点;
- 14) 搜寻是否还有未访问的对象, 若有:
- 15) 跳转执行步骤 2);
- 16) 结束。

4 算例分析

本章通过对中国海南省三亚市下属 3 个台区实际数据进行分析, 证明在实际工程应用中文章所提方法的有效性。3 个实际台区的基础参数如表 1 所示。

表 1 实际台区基础参数

Tab. 1 Basic parameters of the actual low-voltage distribution transformer networks

台区	用户数	A相表数量	B相表数量	C相表数量	数据采集完整率/%
1	220	80	74	66	100.0
2	216	72	68	76	94.6
3	233	81	75	77	93.1

各台区数据集由所有用户智能电表电压采集数据组成。其中, 数据采集长度为 3 d, 频率为每分钟一次, 总共 4320 个时刻点。

本章首先对台区类别以及各台区下用户相位进行识别, 接着将文章所提方法与其他方法进行比较, 证明文章所提方法在实际工程领域的有效性。

首先, 利用低压配电网智能电表采集的节点电压数据生成时序电压矩阵 U 。电压矩阵大小为 669×4320 , 669 为 3 个台区的用户测量点的总和, 电压矩阵行向量表示同一智能电表记录的某一用户的电压序列, 4320 个列向量表示相同时刻下各智能电表测得的电压大小。

接着, 对电压矩阵 U 进行归一化处理, 得到标准化时序电压矩阵 U' 。

然后, 采用 t -SNE 降维算法对该标准化时序电压特征矩阵 U' 做降维处理。降维结束后, 得到二维的电压特征矩阵 U_{2d} 。

将 U_{2d} 作为 DBSCAN 聚类算法的输入矩阵, 采用枚举迭代的方法对距离阈值参数 eps 与样本数阈值参数 min_samples 进行机器调参, 得出噪声点最小、轮廓系数最大的参数组合, 作为 DBSCAN 聚类算法对该组输入数据矩阵的最佳输入参数。聚类完成后, 输出用户的拓扑分类结果, 用户拓扑归属关系识别完成。整体方案流程图如图 1 所示。

经过详细统计与计算, 最终得出低压用户与台区归属关系的识别准确率为 100%, 不同台区用户相位识别平均准确率为 99.1%。聚类识别效果图如图 2 所示。

为了证明文章所提方法与当前其他主流辨识算法相比具有更高的准确性与普适性, 文章从线性降维算法中选用基于距离映射降维的 PCA 降维法, 从流形降维算法中选用基于数据点与点之间流形关系降维的 LLE 降维法, 与文章所提的基于概率映射降维的 t -SNE 降维法进行对照; 从划分式聚类算法中选用基于欧氏距离聚类的 k -means 聚类法, 从层次化聚类算法中选用基于建立分类树结构按照树节点层层聚类的 BIRCH 聚类法, 与文章所提的基于数据密度的 DBSCAN 密度聚类法进行对照, 最终辨识结果准确率如表 2 所示。

分析表 2 数据可知, 对于实际台区中的用户时序电压数据, 在采用相同的降维算法的情况下,

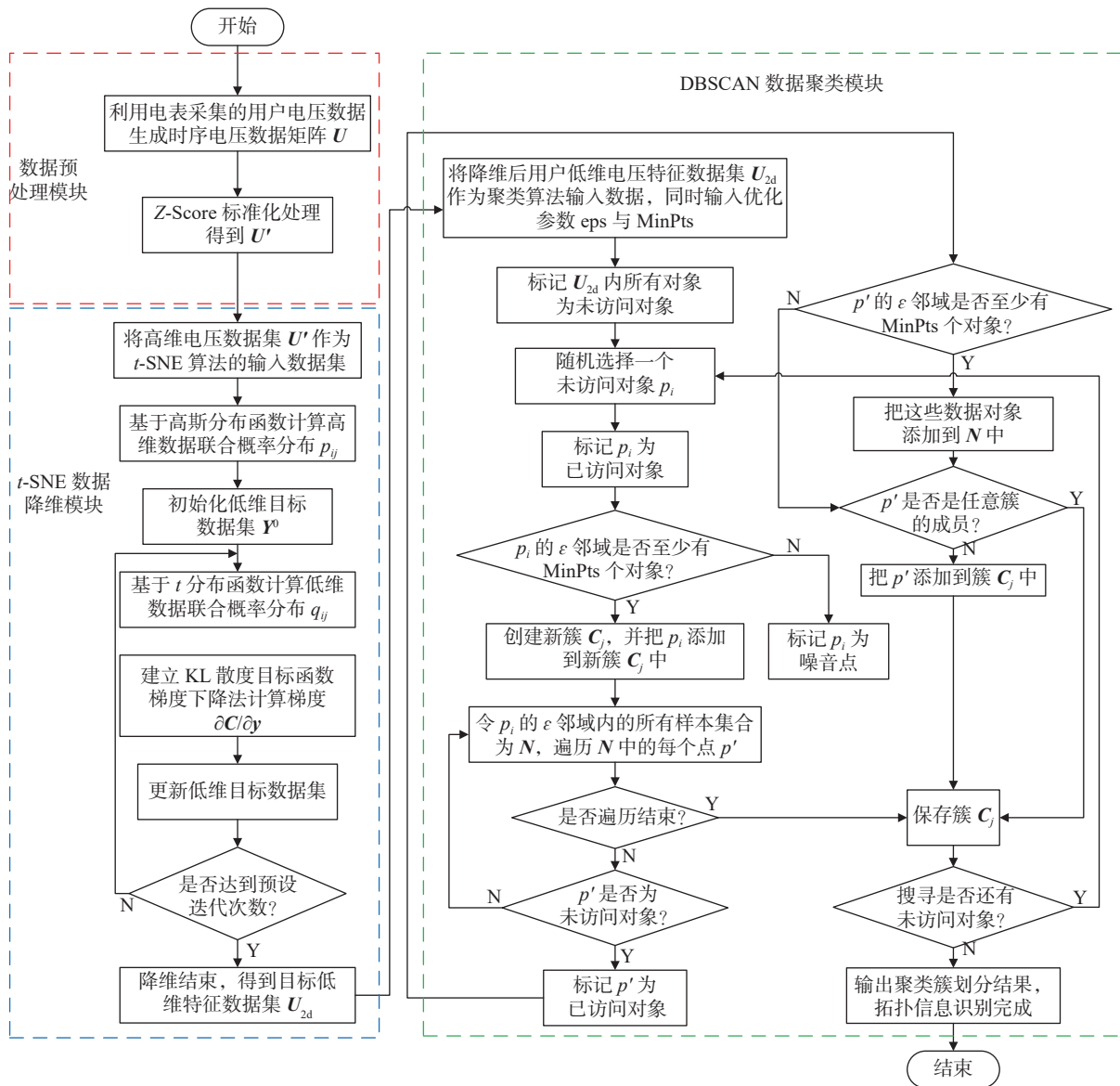


图 1 低压台区归属关系及相位识别方案流程图

Fig. 1 Flowchart of the identification of low-voltage distribution network attribution relationship and phase information

DBSCAN 算法相较于 BIRCH 算法与 k -means 算法在聚类数据上有着更高的台区归属关系识别准确率与相位识别准确率; 在采用相同的聚类算法的情况下, t -SNE 算法相较于 LLE 算法与 PCA 算法有着更高的识别准确率。

其主要原因在于由智能电表采集的用户电压数据体量大、拓扑信息复杂数据簇分支多、精度较低噪点多, 从算法原理的角度来看, PCA 算法泛化能力差, LLE 算法与 k -means 算法易受数据噪点影响, BIRCH 算法仅擅长处理树形数据, 在处理该类数据时的效果不如泛化性强、数据可视化效果好的 t -

SNE 算法与擅长处理密度均匀数据的 DBSCAN 算法。

因此, 分别结合在两方面表现最优的算法, 即 t -SNE+DBSCAN 的组合, 能够解决实际情况下含噪点用户数据的台区归属辨识问题以及多台区下用户相序辨识问题, 证明了文章所提方法的有效性与优势性。

5 结论

对文章的研究内容总结如下:

1) 提出了一种基于密度聚类的低压台区归属关系及相位识别新方法, 该方法改进了已有的数据预

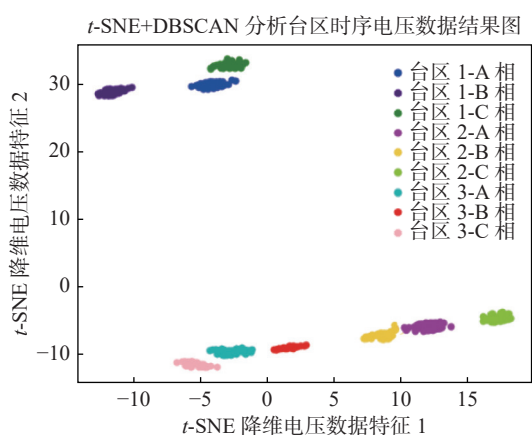


图2 低压台区归属关系及相位识别聚类效果图

Fig. 2 Clustering effect diagram of low-voltage distribution network attribution relationship and phase information

表2 不同方法下台区识别准确率和不同台区用户相位识别准确率

Tab. 2 Identification accuracy of low-voltage distribution network attribution relationship and phase information under different methods

识别算法	台区归属识别准确率/%			用户相位识别准确率/%		
	台区1	台区2	台区3	A相	B相	C相
<i>t</i> -SNE+DBSCAN	100.0	100.0	100.0	100.0	99.4	98.9
<i>t</i> -SNE+BIRCH	100.0	100.0	100.0	100.0	93.6	94.4
<i>t</i> -SNE+ <i>k</i> -means	100.0	100.0	100.0	100.0	91.2	86.9
LLE+DBSCAN	99.0	96.5	98.8	76.3	78.6	75.6
LLE+BIRCH	96.5	95.7	91.5	73.1	69.5	68.0
LLE+ <i>k</i> -means	93.6	94.0	90.6	72.5	63.1	62.0
PCA+DBSCAN	96.5	94.1	97.6	70.8	69.0	69.2
PCA+BIRCH	96.2	93.7	91.2	69.6	68.7	64.0
PCA+ <i>k</i> -means	92.0	91.7	92.1	63.1	59.5	61.8

处理方法、降维算法和聚类算法,将3种方法有机地结合在一起用于拓扑信息识别问题上,为低压台区拓扑信息的识别提供了不一样的研究思路。

2) 经过严谨的实验分析和比较,结果表明,所提出的方法能够很好地利用智能电表采集的含噪声的实际工程电压数据,在识别台区归属信息与相位信息问题上能够达到95%以上的准确率,并高于现有的其他主流识别算法,证明了文章所提方法在解决此类问题上具有有效性与优势性。

但文章所提方法仍具有一定的局限性,体现在基于概率降维的*t*-SNE降维方法无法保留原始高维

数据的流形信息,即降维后的数据点图无法保留分支与分支之间和用户与用户之间的电气距离信息,无法对低压台区的拓扑信息进行更深入准确的识别。下一阶段将探索更有效的深度学习方法,针对低压台区用户分支的电气距离识别问题作进一步的研究。

参考文献:

- [1] 章晋龙, 史述青, 杨海森, 等. 国内外直流配电网技术发展研究综述 [J]. *南方能源建设*, 2016, 3(增刊1): 93-98. DOI: 10.16516/j.gedi.issn2095-8676.2016.S1.020.
- ZHANG J L, SHI S Q, YANG H S, et al. Overview of development of DC power distribution network [J]. *Southern energy construction*, 2016, 3(Suppl. 1): 93-98. DOI: 10.16516/j.gedi.issn2095-8676.2016.S1.020.
- [2] 李宁, 郭泽林, 袁铁江, 等. 基于电力量测信息的低压台区电力拓扑生成方法 [J]. *分布式能源*, 2020, 5(5): 48-55. DOI: 10.16513/j.2096-2185.DE.2005006.
- LI N, GUO Z L, YUAN T J, et al. Power topology generation method in low-voltage transformer area based on electric power metrical information [J]. *Distributed energy*, 2020, 5(5): 48-55. DOI: 10.16513/j.2096-2185.DE.2005006.
- [3] 李健, 马彬, 张植华, 等. 基于网格的城市配电网优化规划方法研究 [J]. *南方能源建设*, 2015, 2(3): 38-42. DOI: 10.16516/j.gedi.issn2095-8676.2015.03.007.
- LI J, MA B, ZHANG Z H, et al. Research on a grid-based optimal planning method for urban distribution system [J]. *Southern energy construction*, 2015, 2(3): 38-42. DOI: 10.16516/j.gedi.issn2095-8676.2015.03.007.
- [4] 李遥, 许烈, 李永东. 模块化多电平矩阵变换器开关次数最优的拓扑连接策略 [J]. *电工技术学报*, 2018, 33(2): 342-352. DOI: 10.19595/j.cnki.1000-6753.tces.161106.
- LI Y, XU L, LI Y D. Topology connecting strategy of the modular multilevel matrix converter for optimal switching times [J]. *Transactions of China electrotechnical society*, 2018, 33(2): 342-352. DOI: 10.19595/j.cnki.1000-6753.tces.161106.
- [5] 郑杰, 赵敬凤, 王海江, 等. 一种基于载波通信的电力相位识别方法 [J]. *电测与仪表*, 2015, 52(增刊1): 77-79. DOI: 10.3969/j.issn.1001-1390.2015.z1.018.
- ZHENG J, ZHAO J F, WANG H J, et al. A method to recognizing phase of power based on carrier communication [J]. *Electrical measurement & instrumentation*, 2015, 52(Suppl. 1): 77-79. DOI: 10.3969/j.issn.1001-1390.2015.z1.018.
- [6] 高磊, 杨毅, 卜强生, 等. 基于MAC地址匹配的智能变电站网络设备拓扑动态识别 [J]. *电力系统自动化*, 2014, 38(20): 81-84. DOI: 10.7500/AEPS20140207003.
- GAO L, YANG Y, BU Q S, et al. Dynamical topology identification of smart substation network equipment based on MAC address matching [J]. *Automation of electric power*

- systems, 2014, 38(20): 81-84. DOI: [10.7500/AEPS20140207003](https://doi.org/10.7500/AEPS20140207003).
- [7] 佟为明, 梁建权, 李中伟, 等. 基于数据可恢复的AMI无线传感器网络安全数据聚合技术 [J]. *电力自动化设备*, 2017, 37(5): 211-216, 223. DOI: [10.16081/j.issn.1006-6047.2017.05.032](https://doi.org/10.16081/j.issn.1006-6047.2017.05.032).
TONG W M, LIANG J Q, LI Z W, et al. Security data aggregation based on data recoverability for wireless sensor network of AMI [J]. *Electric power automation equipment*, 2017, 37(5): 211-216, 223. DOI: [10.16081/j.issn.1006-6047.2017.05.032](https://doi.org/10.16081/j.issn.1006-6047.2017.05.032).
- [8] 李昂, 刘亚东, 孙鹏, 等. 基于 μ PMU相位数据的配电网拓扑识别方法 [J]. *智慧电力*, 2017, 45(11): 31-36. DOI: [10.3969/j.issn.1673-7598.2017.11.006](https://doi.org/10.3969/j.issn.1673-7598.2017.11.006).
LI A, LIU Y D, SUN P, et al. Distribution network topology identification method based on phase data by μ PMU [J]. *Smart power*, 2017, 45(11): 31-36. DOI: [10.3969/j.issn.1673-7598.2017.11.006](https://doi.org/10.3969/j.issn.1673-7598.2017.11.006).
- [9] 裴宇婷, 秦超, 余贻鑫. 基于LightGBM和DNN的智能配电网在线拓扑辨识 [J]. *天津大学学报(自然科学与工程技术版)*, 2020, 53(9): 939-950. DOI: [10.11784/tdxbz201907064](https://doi.org/10.11784/tdxbz201907064).
PEI Y T, QIN C, YU Y X. Online topology identification for smart distribution grids based on LightGBM and deep neural networks [J]. *Journal of Tianjin university (science and technology)*, 2020, 53(9): 939-950. DOI: [10.11784/tdxbz201907064](https://doi.org/10.11784/tdxbz201907064).
- [10] 唐捷, 蔡永智, 周来, 等. 基于数据驱动的低压配电网线户关系识别方法 [J]. *电力系统自动化*, 2020, 44(11): 127-134. DOI: [10.7500/AEPS20191021007](https://doi.org/10.7500/AEPS20191021007).
TANG J, CAI Y Z, ZHOU L, et al. Data-driven based identification method of feeder-consumer connectivity in low-voltage distribution network [J]. *Automation of electric power systems*, 2020, 44(11): 127-134. DOI: [10.7500/AEPS20191021007](https://doi.org/10.7500/AEPS20191021007).
- [11] 陈招安, 黄纯, 张志丹, 等. 基于T型灰色关联度和KNN算法的低压配电网台区拓扑识别方法 [J]. *电力系统保护与控制*, 2021, 49(1): 163-169. DOI: [10.19783/j.cnki.pspc.200260](https://doi.org/10.19783/j.cnki.pspc.200260).
CHEN Z A, HUANG C, ZHANG Z D, et al. Topology identification method of a low voltage distribution network based on T-type grey correlation degree and KNN algorithm [J]. *Power system protection and control*, 2021, 49(1): 163-169. DOI: [10.19783/j.cnki.pspc.200260](https://doi.org/10.19783/j.cnki.pspc.200260).
- [12] 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*, 2018, 9(5): 5113-5122. DOI: [10.1109/tsg.2017.2680542](https://doi.org/10.1109/tsg.2017.2680542).
- [13] LUAN W P, PENG J S, MARAS M, et al. Distribution network topology error correction using smart meter data analytics [C]// Proceedings of 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, Canada, July 21-25, 2013. New York: IEEE, 2013: 1-5. DOI: [10.1109/pesmg.2013.6672786](https://doi.org/10.1109/pesmg.2013.6672786).
- [14] 唐捷, 蔡永智, 李其霖, 等. 基于三相表特征约束聚类的低压台区用户相序识别方法 [J]. *电力系统自动化*, 2022, 46(8): 60-67. DOI: [10.7500/AEPS20210616007](https://doi.org/10.7500/AEPS20210616007).
TANG J, CAI Y Z, LI Q L, et al. Phase sequence identification method for users in low-voltage distribution-station area based on feature constraint clustering of three-phase meter [J]. *Automation of electric power systems*, 2022, 46(8): 60-67. DOI: [10.7500/AEPS20210616007](https://doi.org/10.7500/AEPS20210616007).
- [15] 连子宽, 姚力, 刘晟源, 等. 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. DOI: [10.7500/AEPS20190621001](https://doi.org/10.7500/AEPS20190621001).
LIAN Z K, YAO L, LIU S Y, et al. 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. DOI: [10.7500/AEPS20190621001](https://doi.org/10.7500/AEPS20190621001).
- [16] 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. DOI: [10.7500/AEPS20170922003](https://doi.org/10.7500/AEPS20170922003).
LIU Y X, ZHANG N, KANG C Q. A review on data-driven analysis and optimization of power grid [J]. *Automation of electric power systems*, 2018, 42(6): 157-167. DOI: [10.7500/AEPS20170922003](https://doi.org/10.7500/AEPS20170922003).
- [17] ABDI H, WILLIAMS L J. Principal component analysis [J]. *WIREs computational statistics*, 2010, 2(4): 433-459. DOI: [10.1002/wics.101](https://doi.org/10.1002/wics.101).
- [18] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, 290(5500): 2323-2326. DOI: [10.2307/3081722](https://doi.org/10.2307/3081722).
- [19] VAN DER MAATEN L, HINTON G. Visualizing data using t -SNE [J]. *Journal of machine learning research*, 2008, 9(86): 2579-2605.
- [20] ZHANG T, RAMAKRISHNAN R, LIVNY M. BIRCH: an efficient data clustering method for very large databases [J]. *ACM SIGMOD record*, 1996, 25(2): 103-114. DOI: [10.1145/235968.233324](https://doi.org/10.1145/235968.233324).
- [21] BIRANT D, KUT A. ST-DBSCAN: an algorithm for clustering spatial-temporal data [J]. *Data & knowledge engineering*, 2007, 60(1): 208-221. DOI: [10.1016/j.datak.2006.01.013](https://doi.org/10.1016/j.datak.2006.01.013).

作者简介:



闫东辉

闫东辉(通信作者)

1960-, 男, 海南电网有限责任公司基建部副总经理, 东北电力学院电气工程及其自动化专业学士, 主要从事电力生产、管理, 配电网规划、设计和运行分析研究工作(e-mail) 978045113@qq.com。

(编辑 叶筠英)