

DOI: 10.12158/j.2096-3203.2025.01.004

基于层次近邻传播聚类的用户低电压越限模式挖掘方法

沈枢昊¹, 钟庆¹, 许中², 王钢¹, 李海锋¹, 汪隆君¹

(1. 华南理工大学电力学院, 广东 广州 510640; 2. 广州供电局有限公司, 广东 广州 510600)

摘要:开展用户低电压越限模式挖掘工作可以为用户低电压问题的治理提供指导。针对目前台区低电压用户电压复杂多变、低电压越限模式未知的问题,文中提出基于层次近邻传播(hierarchical affinity propagation, HAP)聚类的用户低电压越限模式挖掘方法。首先,通过HAP聚类算法对大规模低电压用户电压数据集进行聚类分析,获得若干聚类簇。然后,将不同的聚类簇视作不同的低电压越限模式,并从越限时长和越限电压幅值两方面定义低电压越限模式的4项基本特征指标,通过计算各聚类簇的基本特征指标,反映其所对应低电压越限模式的特征。最后,将该方法运用到某地区低电压用户的电压数据集中,有效挖掘出该地区低电压用户的4种低电压越限模式,从而根据不同低电压越限模式的特征,有针对性地开展低电压用户的监管、分析工作,并制定用户低电压问题治理的优先级。

关键词:低电压用户;层次近邻传播(HAP)聚类;低电压越限模式;越限时长;越限电压幅值;治理优先级

中图分类号:TM714

文献标志码:A

文章编号:2096-3203(2025)01-0030-09

0 引言

随着经济发展和城镇化加速,配电台区数量激增,台区低电压问题愈发严重^[1-4]。低电压问题不仅会影响用户的正常用电,还会增加电网的功率损耗,严重影响配电网的安全稳定运行^[5-6]。开展配电台区的电压监测工作是治理低电压问题的有效手段。我国电网部门在配电台区首端以及用户侧安装了大量的监测设备,并开展监测工作。随着监测数据的指数级增长,电网数据呈现明显的“大数据”特征^[7-9]。长期监测工作积累了丰富的电压时间序列数据,监测数据中有价值的时间序列特征信息未能得到充分挖掘利用^[10],迫切需要高效且合适的特征分析方法,实现监测数据增值应用^[11-12]。

电力系统中包含多种类型的时间序列数据,如负荷数据、电能质量数据等。针对电力行业中时间序列数据的特征提取及模式识别技术,国内外学者开展了大量研究工作。在电网实际运行中,由于缺乏监测数据时间序列模式特征的先验知识,尤其在供电环境复杂多变的情况下,有监督的分析方法并不适用于现有监测数据^[13-15],因此现有研究主要采用聚类这种无监督的分析方法。针对电网中的电能质量数据,文献^[16]通过量子聚类算法从稳态电能质量监测数据中挖掘出具有显著差异的指标越

限模式,并对各越限模式开展严重度评估工作。针对电网中的负荷数据,文献^[17-19]在对负荷数据进行一定预处理的基础上,分别采用不同的聚类方法,开展负荷曲线的模式识别研究工作。配电台区电压监测数据的研究与应用主要聚焦于分析用户与台变的电压时序数据,以解决配电台区的拓扑及用户相位识别问题^[20-21]。

不同的低电压产生原因可能使得用户电压具有不同的低电压越限模式。导致用户低电压问题的常见原因有供电半径过长、三相负荷分配不平衡、配变运行档位不合理、供电线径过细等^[22]。近年来,随着电网规模持续扩大,如何避免低电压治理中的低效投资并确定治理优先级,成为电力企业关注的重点之一^[23-24]。开展用户的低电压越限模式识别工作有助于电网公司掌握低电压用户治理的优先级,从而有针对性地开展低电压用户的监管、分析和差异化治理工作。

一方面,低电压用户来自不同配电台区,而不同配电台区具有不同的拓扑,因此低电压用户电压数据本身就具备一定的层次结构;另一方面,低电压用户数据庞大、变化模式复杂且缺乏先验知识。根据低电压用户电压数据的特点,文中提出基于层次近邻传播(hierarchical affinity propagation, HAP)聚类的用户低电压越限模式挖掘方法。首先,采用HAP聚类算法对大规模低电压用户电压数据集进行聚类分析,获得若干聚类簇;然后,将不同的聚类簇视作不同的低电压越限模式,从越限时长和越限电压幅值两方面定义4种基本特征指标,分别计算

收稿日期:2024-06-11;修回日期:2024-08-25

基金项目:国家自然科学基金资助项目(52077082);广东省自然科学基金资助项目(2021A1515012087)

各聚类簇的基本特征指标,从而获得其所对应低电压越限模式的特征;最后,利用文中方法实现对某市低电压用户的低电压越限模式挖掘工作。

1 用户低电压越限模式挖掘方法

1.1 用户电压监测数据的 HAP 聚类分析

1.1.1 动态时间弯曲距离的定义

动态时间弯曲(dynamic time warping, DTW)距离是一种常见的度量时间序列相似度的方法。相较于欧氏距离,DTW 距离能够衡量不同时间序列间的整体相似性,对噪声和局部波动的鲁棒性更强^[25-28]。

用户电压数据受用电习惯影响,其电压曲线的峰谷时间并不完全一致,但仍呈现一定的周期性。相比于传统的距离度量,DTW 距离允许时间序列在时间轴上存在偏移或拉伸,因此文中利用 DTW 距离度量不同用户电压时间序列数据之间的相似程度,从而构造相似度矩阵。

假设用户 i 的电压数据 $\mathbf{U}_i = [u_{i,1} \ u_{i,2} \ \dots \ u_{i,l}]^T$, $u_{i,l}$ 为用户 i 第 l 个采样点的电压数据;用户 j 的电压数据 $\mathbf{U}_j = [u_{j,1} \ u_{j,2} \ \dots \ u_{j,h}]^T$, $u_{j,h}$ 为用户 j 第 h 个采样点的电压数据。通过 2 个用户的电压监测数据可以构建一个 $l \times h$ 的距离矩阵 \mathbf{D} , 矩阵中元素 $d_{p,q}$ 为:

$$d_{p,q} = \sqrt{(u_{i,p} - u_{j,q})^2} \quad 1 \leq p \leq l; 1 \leq q \leq h \quad (1)$$

DTW 算法的目的在于找出一条最优对齐路径 $W = \{w_1, w_2, \dots, w_s\}$, 其中路径上的第 e 个元素定义为 w_e , $e = 1, 2, \dots, s$ 。若 $w_e = d_{p,q}$, 则表示用户 i 电压数据 \mathbf{U}_i 中 $u_{i,p}$ 与用户 j 电压数据 \mathbf{U}_j 中的 $u_{j,q}$ 对应。在该最优对齐路径下,用户电压数据 \mathbf{U}_i 和 \mathbf{U}_j 间的累计距离最小,该累计距离即对应两者的 DTW 距离。

计算用户 i 、 j 的电压数据 \mathbf{U}_i 和 \mathbf{U}_j 的 DTW 距离,其目标函数 f_{DTW} 可表示为:

$$f_{\text{DTW}}(\mathbf{U}_i, \mathbf{U}_j) = \min \left(\sum_{e=1}^s w_e \right) \quad (2)$$

为找出最优对齐路径,通过动态规划方法构造一个 $l \times h$ 的累积距离矩阵 $\boldsymbol{\gamma}$, 其元素计算公式为:

$$\gamma_{p,q} = d_{p,q} + \min(\gamma_{p-1,q-1}, \gamma_{p-1,q}, \gamma_{p,q-1}) \quad 1 \leq p \leq l; 1 \leq q \leq h \quad (3)$$

式中: $\gamma_{0,0} = 0$; $\gamma_{1,0} = \gamma_{0,1} = +\infty$ 。

累积距离矩阵 $\boldsymbol{\gamma}$ 的最后一个元素 $\gamma_{l,h}$ 即为计算得到的 DTW 距离。

$$f_{\text{DTW}}(\mathbf{U}_i, \mathbf{U}_j) = \gamma_{l,h} \quad (4)$$

1.1.2 低电压用户电压监测数据的 HAP 聚类

HAP 聚类算法基于近邻传播(affinity propagation, AP)聚类算法进行改进,不仅无须提前指定聚类个数,还弥补了 AP 聚类算法在大规模数据集中表现不佳的缺点。此外,该聚类算法对数据的分布形式没有严格要求,可以处理不同类型的数据。

HAP 聚类的基本思想为将大规模数据集划分为多个子集,各子集经过 AP 聚类获得各自的聚类中心,然后通过对所有子集的聚类中心进行 AP 聚类获得全局的聚类中心,通过计算每条用户数据与全局聚类中心的距离选择该数据所属的聚类中心。

具体聚类分析过程如下:

(1) 指定每个子集包含的用户个数为 m , 并按照顺序对用户电压数据集 U 进行采样,如果采样最后剩余用户数量小于子集用户数量 m , 则剩余用户数据独自构成一个子集。最后,用户电压数据集 U 被划分成 r 个子集 X_1, X_2, \dots, X_r 。子集划分过程如图 1 所示,其中 N 为用户总数量。

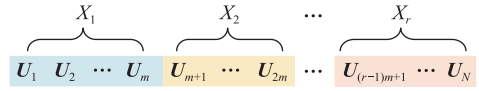


图 1 子集划分过程

Fig.1 Subset partitioning process

(2) 对于第 x 个子集 X_x ($x = 1, 2, \dots, r$), 先通过 DTW 算法计算子集内各用户电压数据之间的距离,进而构造该子集的相似度矩阵 \mathbf{S}_x , 如式(5)所示。

$$\mathbf{S}_x = \begin{bmatrix} s_{1,1} & \dots & s_{1,j} & \dots & s_{1,m} \\ \vdots & & \vdots & & \vdots \\ s_{i,1} & \dots & s_{i,j} & \dots & s_{i,m} \\ \vdots & & \vdots & & \vdots \\ s_{m,1} & \dots & s_{m,j} & \dots & s_{m,m} \end{bmatrix} \quad (5)$$

式中: $s_{i,j}$ 为数据 \mathbf{U}_i 和数据 \mathbf{U}_j 之间的相似度。

对于相似度矩阵 \mathbf{S}_x 的非对角线元素 $s_{i,j}$, 其计算公式为:

$$s_{i,j} = -f_{\text{DTW}}(\mathbf{U}_i, \mathbf{U}_j) \quad i \neq j \quad (6)$$

对于相似度矩阵 \mathbf{S}_x 对角线上的元素 $s_{i,i}$, 其计算公式为:

$$s_{i,i} = \min(\mathbf{S}_x) \frac{m \sum_{j=1, j \neq i}^m s_{i,j}}{\sum_{i=1}^m \sum_{j=1, j \neq i}^m s_{i,j}} \quad (7)$$

(3) 各子集基于所得相似度矩阵,计算可靠度矩阵 \mathbf{R} 和有效度矩阵 \mathbf{A} , 并对两者进行迭代更新,从而完成 AP 聚类^[29-31]。在聚类过程中,可靠度表示

一个样本数据作为聚类中心的可能性,有效度表示一个样本数据作为聚类成员的可能性。用户数据的可靠度和有效度之和越大,其越有可能成为聚类中心。可靠度矩阵 \mathbf{R} 中的元素 $r_{i,j}$ 表示数据 U_i 向数据 U_j 传递数值信息,数据 U_j 作为数据 U_i 的聚类中心的可能性,其中第 t 次迭代时的值为 $r_{i,j,t}$,其计算公式为:

$$r_{i,j,t} = s_{i,j} - \max_{k \neq j} (a_{i,k,t-1} + s_{i,k}) \quad (8)$$

式中: $a_{i,k,t-1}$ 为第 $t-1$ 次迭代时数据 U_i 和数据 U_k 之间的有效度,有效度的迭代初值为 0; $s_{i,k}$ 为数据 U_i 和数据 U_k 之间的相似度。

有效度矩阵 \mathbf{A} 中元素 $a_{i,j}$ 第 t 次迭代时的值为 $a_{i,j,t}$,其计算公式为:

$$a_{i,j,t} = \begin{cases} \min \left(0, r_{j,j,t} + \sum_{k \in (i,j)} \max(0, r_{k,j,t}) \right) & i \neq j \\ \sum_{k \neq j} \max(0, r_{k,j,t}) & i = j \end{cases} \quad (9)$$

式中: $r_{j,j,t}$ 、 $r_{k,j,t}$ 分别为第 t 次迭代时可靠度矩阵中元素 $r_{j,j}$ 、 $r_{k,j}$ 的数值。

在迭代求解过程中,引入阻尼系数以避免因存在噪声数据而发生振荡的问题,可靠度和有效度实际等于当前迭代值和上一次迭代值的加权平均。可靠度和有效度的更新迭代公式为:

$$\begin{cases} r_{i,j,t+1} = (1 - \lambda)r_{i,j,t} + \lambda r_{i,j,t} \\ a_{i,j,t+1} = (1 - \lambda)a_{i,j,t} + \lambda a_{i,j,t} \end{cases} \quad (10)$$

式中: λ 为阻尼系数,为 0~1 之间的实数,通常 λ 取 0.9^[32]。

若 $r_{i,j}$ 与 $a_{i,j}$ 越大,则用户 j 的电压数据成为聚类中心的可能性越大,且用户 i 与用户 j 同属一类的可能性也越大。通过计算矩阵 \mathbf{R} 与矩阵 \mathbf{A} 对应元素之和,找到各用户所在列的最大值对应索引,该索引对应数据即为该用户所属的聚类中心。

当循环迭代满足停止条件,即超过预设的迭代次数或者聚类中心不再发生变化时,迭代停止,输出第 x 个子集的聚类中心 c_x 。

(4) 将各个子集经过 AP 聚类后所得的聚类中心 c_1, c_2, \dots, c_r 组成一个新的数据集,对该数据集进行 AP 聚类,从而获得全部数据集的聚类中心 C_1, C_2, \dots, C_K , K 为全局聚类簇个数。然后,每个用户电压数据根据各自与全局聚类中心的 DTW 距离寻找对应的聚类中心,从而实现 HAP 聚类的全过程。为更好地反映聚类效果,文中采用 Davies-Bouldin 指标 (Davies-Bouldin index, DBI) I_{DBI} 对聚类结果进行评价^[18]。DBI 的计算如式(11)所示。

$$I_{DBI} = \frac{1}{K} \sum_{z=1}^K \max_{z \neq y} \left(\frac{d(X_z) + d(X_y)}{d(C_z, C_y)} \right) \quad (11)$$

式中: $d(X_z)$ 、 $d(X_y)$ 分别为第 z 类、第 y 类时间序列曲线与该类聚类中心曲线的平均距离; $d(C_z, C_y)$ 为第 z 类聚类中心与第 y 类聚类中心曲线间的距离。DBI 越小,同类之间距离越小,不同类之间距离越大,聚类效果越好。

(5) 由于划分的子集大小直接决定了最终的聚类结果,指定不同子集所含用户个数 m 并进行 HAP 聚类,计算其各自聚类结果的 DBI;通过对比所得聚类结果的 DBI,选出聚类效果最佳的子集所含用户个数,将其对应的聚类结果作为低电压用户电压数据的聚类结果,以进行下一步的模式分析工作。

1.2 低电压越限模式的特征指标

经过 HAP 聚类分析可以获取低电压用户电压数据的多个聚类簇,将不同的聚类簇视作不同的低电压越限模式。文中定义了 4 种基本的电压越限特征,从越限时长和越限电压幅值两方面对不同越限模式的越限程度进行分析,量化不同低电压越限模式的特征。

(1) 平均越限时长占比 A_K 。

文中定义用户电压数据低于标准规定阈值的时长占其采样总时长的比值为用户的越限时长占比,其计算公式为:

$$Y_i = \sum_{\alpha=1}^l \frac{g(u_{i,\alpha})}{l} \quad (12)$$

式中: Y_i 为用户 i 的越限时长占比; $g(u_{i,\alpha})$ 为反映用户 i 第 α 个采样点电压数据 $u_{i,\alpha}$ 是否低于 0.9 的布尔变量,其计算公式如式(13)所示。

$$g(u_{i,\alpha}) = \begin{cases} 1 & u_{i,\alpha} < 0.9 \\ 0 & u_{i,\alpha} \geq 0.9 \end{cases} \quad (13)$$

通过对各聚类簇中用户的越限时长占比取平均值,获得各聚类簇的平均越限时长占比。平均越限时长占比可以客观反映各聚类簇中的用户在一定时间范围内出现低电压问题的时间长短。

$$A_K = \frac{1}{n_K} \sum_{i \in C_K} Y_i \quad (14)$$

式中: n_K 为属于第 K 个聚类簇的用户数量。

(2) 平均最大单日越限时长占比 E_K 。

文中定义用户单日电压数据低于标准规定阈值的时长占单日总时长的比值为用户的单日越限时长占比,其计算公式为:

$$B_{i,d} = \sum_{\alpha=24(d-1)+1}^{24d} \frac{g(u_{i,\alpha})}{24} \quad d = 1, 2, \dots, 7 \quad (15)$$

式中: $B_{i,d}$ 为用户 i 第 d 天的单日越限时长占比。

对于用户 i 而言,在采样时间范围内对每天的单日越限时长占比取最大值可以得到其最大单日越限时长占比 T_i ,如式(16)所示。

$$T_i = \max(B_{i,d}) \quad d = 1, 2, \dots, 7 \quad (16)$$

因此聚类簇的平均最大单日越限时长占比 E_k 可以通过计算聚类簇中各用户的最大单日越限时长占比的平均值获得。平均最大单日越限时长占比可以客观反映各聚类簇中的用户单日出出现低电压问题的时间长短,其计算公式为:

$$E_k = \frac{1}{n_k} \sum_{i \in C_k} T_i \quad (17)$$

(3) 最低电压平均值 H_k 。

用户电压最低值对应其电压越限最严重时的电压值。计算聚类簇中各用户在采样时间范围内最低电压的平均值,可以客观反映该聚类簇中用户在一定时间范围内电压幅值越限的最严重程度,其计算公式为:

$$H_k = \frac{1}{n_k} \sum_{i \in C_k} \min(U_i) \quad (18)$$

(4) 平均越限电压 O_k 。

当某个时刻电压小于标准规定阈值时,文中定义该时刻电压为越限电压,因此用户越限电压 $f(U_i)$ 的计算公式为:

$$f(U_i) = \frac{\sum_{\alpha=1}^l g(u_{i,\alpha}) u_{i,\alpha}}{\sum_{\alpha=1}^l g(u_{i,\alpha})} \quad (19)$$

通过计算各聚类簇中用户越限电压的平均值获得各聚类簇的平均越限电压 O_k ,该指标可以客观反映各聚类簇中用户电压幅值每次越限的程度,其计算公式为:

$$O_k = \frac{1}{n_k} \sum_{i \in C_k} f(U_i) \quad (20)$$

以上4种越限特征多角度地刻画了电压越限的严重程度。 A_k 和 E_k 越大表示用户电压越限时间越长, H_k 和 O_k 越小表示用户的越限电压幅值越小,电压越限程度越大。

文中通过计算各聚类簇的上述特征指标,获得其各自对应低电压越限模式的特征。

1.3 方法流程

文中所提基于 HAP 聚类的用户低电压越限模式挖掘方法的流程如图2所示。

首先,以用户为单位获取电压监测数据,构成数据集 $U = \{U_1, U_2, \dots, U_N\}$,其中用户 i 的电压时间序列数据记为 $U_i = [u_{i,1} \dots u_{i,\alpha} \dots u_{i,l}]^T$ 。然后,对低电压用户的电压数据集 U 采用 HAP 聚类算法进

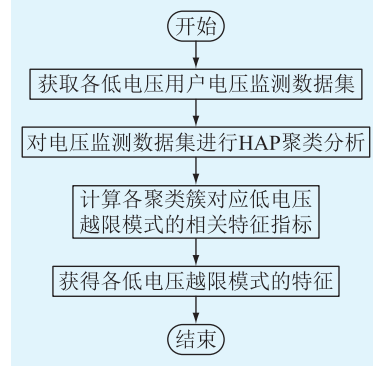


图2 用户低电压越限模式挖掘流程

行聚类,获得若干聚类簇,不同聚类簇对应着不同低电压越限模式。最后,通过计算各聚类簇的4种基本特征指标,获得其各自对应的低电压越限模式特征。

2 算例分析

文中采用的电压数据集来自某市1050个低电压用户2023年第三季度的电压监测数据,数据采样间隔为1h,每个用户的电压数据采样数为168个,电压数据均已标幺化。根据GB/T 12325—2008中对电网供电电压的规定,220V单相供电电压偏差为标称电压的+7%、-10%^[33],因此文中设定用户电压低于标称电压的90%即为出现低电压问题。

2.1 用户低电压越限模式挖掘

以数据集中的某周电压监测数据为例,在选取的时间范围内,用户的过电压问题严重,其电压数据如图3所示,不同时间的低电压用户电压数据之间差异巨大、复杂多变,难以直接分辨其低电压越限模式,因此采用文中所提方法挖掘该周的用户低电压越限模式。

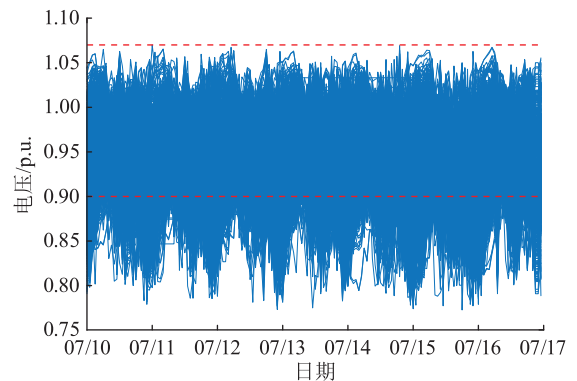


图3 某市低电压用户电压时间序列

Fig.3 Time series of low voltage users in a city

首先,通过 HAP 算法对低电压用户数据进行聚类分析。将1050组低电压用户数据构成的数据集

划分成多个子集。由于子集所含数据量超过 500 后单个子集的 AP 聚类效率会降低,文中子集大小不超过 500^[34]。为选出合适的子集大小,在大量聚类分析的基础上,计算不同子集对应的聚类评价指标,选出最优的子集大小,其所得的 HAP 聚类结果即为下一步特征分析的对象。不同子集大小对应的 DBI 值如图 4 所示。

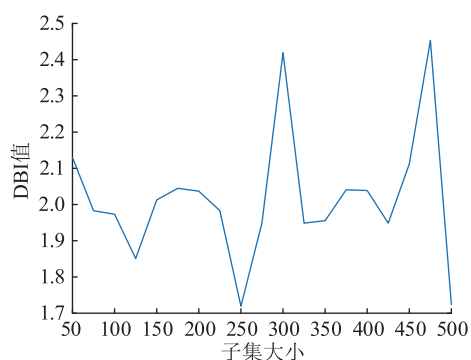


图 4 不同子集大小下的 DBI 值

Fig.4 DBI values for different subset sizes

根据图 4 中的结果,文中设置每个子集有 250 个用户数据。1 050 个低电压用户电压数据经过 HAP 聚类后共得到 4 个聚类簇,对应着 4 种不同的低电压越限模式。各低电压越限模式对应数量如表 1 所示。

表 1 各低电压越限模式对应数量

Table 1 The corresponding number of each low-voltage over-limit pattern

越限模式	聚类簇编号	数量
模式一	a	196
模式二	b	188
模式三	c	328
模式四	d	338

分别计算 4 种低电压越限模式所对应的特征指标,计算结果如图 5 所示。结合各低电压越限模式对应聚类簇的电压曲线特点与特征指标计算结果,对各低电压越限模式进行具体分析。

4 种模式的电压曲线如图 6—图 9 所示。

(1) 模式一。由图 6 可知,模式一下电压在一周内频繁越限,越限时间占比在所有的聚类簇中最大,用户平均越限时长占比接近 0.4,甚至存在某一天长时间处于低电压的情况,同时越限程度在所有越限模式中最大。该模式下的用户低电压问题严重,需要及时治理。

(2) 模式二。由图 7 可知,模式二下的用户电压在绝大多数时间里保持在标准规定范围内,每天电压偶有越限,但存在某天电压越限时长突然增加

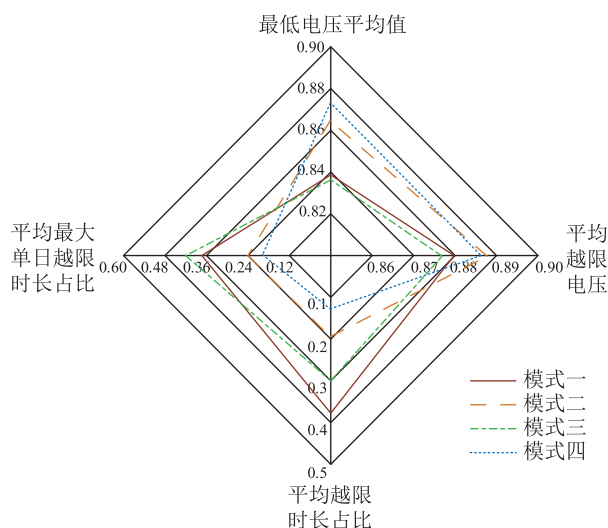


图 5 各低电压越限模式特征指标计算结果雷达图

Fig.5 Radar chart of calculation results for characteristic indexes of low-voltage over-limit patterns

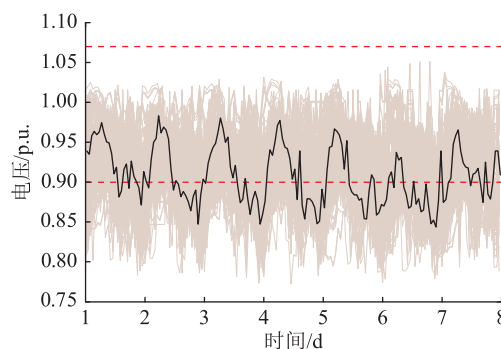


图 6 聚类簇 a

Fig.6 Cluster a

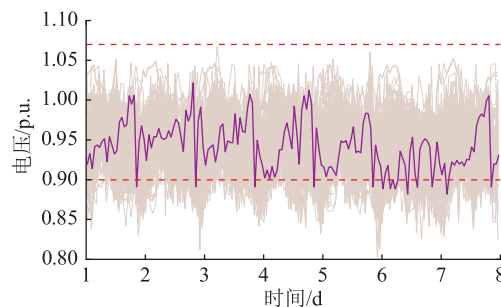


图 7 聚类簇 b

Fig.7 Cluster b

的情况。该模式下的越限程度较轻,越限电压幅值仅小于模式四。

(3) 模式三。由图 8 可知,模式三下的用户电压在一周内多次越限,越限行为呈现一定的周期性,且电压越限发生在每天相对固定的时间段内。越限时间段与用户的用电习惯有很大关系。该模式下的越限严重程度仅次于模式一。

(4) 模式四。由图 9 可知,模式四下的用户电

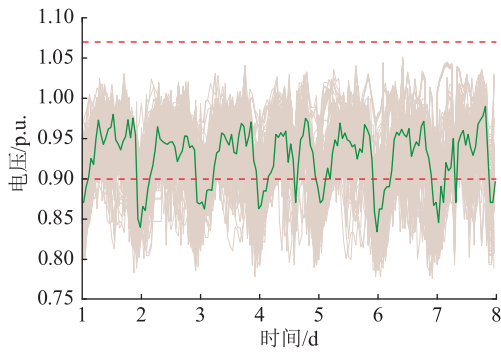


图8 聚类簇 c
Fig.8 Cluster c

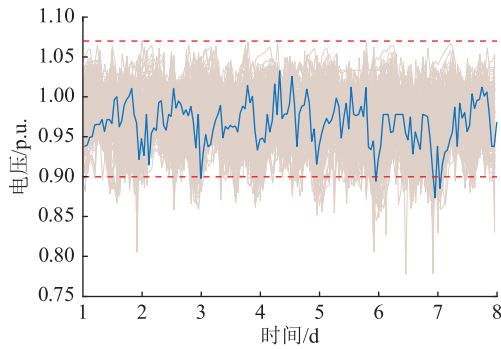


图9 聚类簇 d
Fig.9 Cluster d

压在一周的绝大多数时间里保持在标准规定范围内,电压越限现象偶有发生,越限时长占比在4种模式中最小,且越限程度较轻。该模式下电压越限存在一定的偶然性。

因此,针对这4种不同的越限模式,可以根据不同的越限特点,结合用户投诉次数、受影响用户数量等用户侧需求,对低电压问题治理进行评估,并制定低电压问题治理的优先级。对于低电压问题持续时间长、电压幅值较低的模式一与模式三,其低电压问题较严重,须及时开展低电压成因辨识工作并采取相应措施进行治理,因此这2种模式具有较高的治理优先级。对于模式二与模式四,其用户发生低电压问题存在一定的偶然性,应重点做好监测预警和日常巡查工作。

2.2 文中方法的适应性性与有效性分析

为验证文中所用聚类算法的适应性,任意选取电压数据集中其他时间段的电压监测数据进行聚类分析,结果如图10所示。可见,将文中方法运用到数据集的其他时间段仍然可以获得不错的聚类结果,且各聚类簇对应低电压越限模式也与之前分析结果类似。通过计算各低电压越限模式对应的特征指标,并结合实际情况,可以合理制定低电压用户的治理优先级。

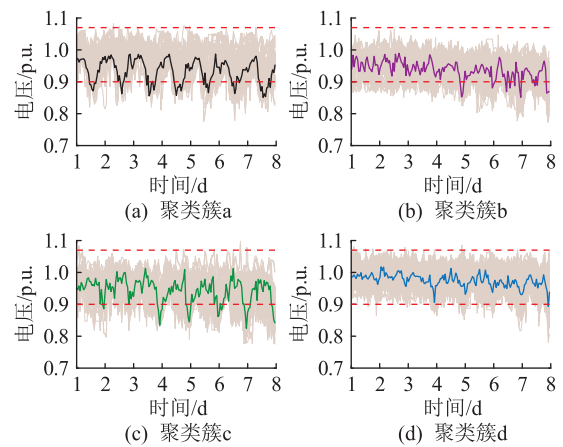


图10 聚类结果

Fig.10 clustering results

为验证文中所用聚类算法的有效性,将该算法与传统的AP聚类算法、层次聚类算法以及常用的K-means聚类算法进行比较,其中层次聚类以及K-means聚类的聚类簇个数取DBI最优下的值,各聚类算法对应的DBI如表2所示。相比于文中算法,传统的AP聚类虽然在DBI上占优,但其聚类簇个数过多会导致聚类结果过于碎片化,失去实际价值。而文中算法相比于层次聚类和K-means聚类算法,不仅不需要事先指定聚类簇数量,在聚类结果上也更优。

表2 不同聚类算法的对比

Table 2 Comparison among different clustering algorithms

聚类算法	所得聚类簇个数	DBI值	是否需要事先指定聚类簇个数
HAP 聚类算法	4	1.71	否
AP 聚类算法	41	1.41	否
层次聚类算法	4	1.84	是
K-means 聚类算法	6	2.64	是

3 结语

开展用户低电压越限模式挖掘工作能够科学指导用户的低电压问题治理。根据低电压用户电压监测数据的特点,文中提出基于HAP聚类算法的用户低电压越限模式挖掘方法,并将该方法运用到某地区的实际算例中。经过聚类分析,一共获得该地区的4种聚类簇,其对应4种低电压越限模式。文中还定义了平均越限时长占比、平均最大单日越限时长占比、最低电压平均值和平均越限电压4种低电压越限模式的特征指标以量化越限模式特征,通过计算各越限模式对应的特征指标,获得各模式的越限特征。根据不同低电压越限模式的特点,相关部

门可以采取针对性的措施开展低电压用户监管、分析工作,并制定用户低电压问题治理的优先级。

参考文献:

- [1] 赖振宏,易皓,黄华,等. 用于低电压问题治理的柔性交直流切换配电系统设计[J]. 电力系统自动化,2023,47(6):123-132.
LAI Zhenhong, YI Hao, HUANG Hua, et al. Design of flexible AC/DC switching distribution system for solving low-voltage problems[J]. Automation of Electric Power Systems, 2023, 47(6):123-132.
- [2] 宋强. 基于大数据分析的配电网低电压变化态势预测研究[J]. 自动化技术与应用,2024,43(3):96-98,103.
SONG Qiang. Research on low voltage change trend prediction of distribution network based on big data analysis[J]. Techniques of Automation and Applications, 2024, 43(3):96-98, 103.
- [3] 周杨珺,张斌,黄伟翔,等. 基于LSTM-BP组合模型的配电网台区低电压预测[J]. 电力科学与技术学报,2023,38(5):177-186.
ZHOU Yangjun, ZHANG Bin, HUANG Weixiang, et al. A low voltage prediction based on LSTM-BP combined model for distribution station area[J]. Journal of Electric Power Science and Technology, 2023, 38(5):177-186.
- [4] 徐峰亮,王发义,赵建涛,等. 采用有载调容调压变和MPC技术的配电网台区低电压治理[J]. 电力建设,2023,44(10):117-126.
XU Fengliang, WANG Fayi, ZHAO Jiantao, et al. Low-voltage management in distribution platform area by on-load capacity and voltage regulating transformer and model predictive control technology[J]. Electric Power Construction, 2023, 44(10):117-126.
- [5] 范李平,朱庆,李黄强,等. 基于多任务辅助学习的配网低电压成因分析[J]. 国外电子测量技术,2024,43(8):125-133.
FAN Liping, ZHU Qing, LI Huangqiang, et al. Analysis on the causes of low voltage in distribution network based on multi-task assisted learning[J]. Foreign Electronic Measurement Technology, 2024, 43(8):125-133.
- [6] 唐会祥,陈文涛,闵子慧,等. 考虑三相负荷不平衡的10kV配电网电压损耗计算模型[J]. 电气自动化,2023,45(6):24-26,30.
TANG Huixiang, CHEN Wentao, MIN Zihui, et al. Calculation model of voltage loss in 10 kV distribution network considering three-phase load unbalance[J]. Electrical Automation, 2023, 45(6):24-26, 30.
- [7] 胡维昊,曹迪,黄琦,等. 深度强化学习在配电网优化运行中的应用[J]. 电力系统自动化,2023,47(14):174-191.
HU Weihao, CAO Di, HUANG Qi, et al. Application of deep reinforcement learning in optimal operation of distribution network[J]. Automation of Electric Power Systems, 2023, 47(14):174-191.
- [8] 葛磊蛟,刘航旭,孙永辉,等. 智能配电网多元电力用户群体特性精准感知技术综述[J]. 电力系统自动化,2023,47(20):174-191.
GE Leijiao, LIU Hangxu, SUN Yonghui, et al. Review on accurate awareness technology for characteristics of diversified power user groups in smart distribution network[J]. Automation of Electric Power Systems, 2023, 47(20):174-191.
- [9] 吴旻昊,王建功,朱英刚,等. 基于t-SNE降维与聚类的主动配电网运行方式在线识别[J]. 电力建设,2023,44(8):52-60.
WU Minhao, WANG Jiangong, ZHU Yinggang, et al. Online identification of active distribution network operation mode based on t-SNE dimensionality reduction and clustering[J]. Electric Power Construction, 2023, 44(8):52-60.
- [10] SI C, XU S L, WAN C, et al. Electric load clustering in smart grid: methodologies, applications, and future trends[J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(2):237-252.
- [11] 张慧波,王守相,赵倩宇,等. 考虑数据不均衡的居民用户负荷曲线分类方法[J]. 电力工程技术,2022,41(3):186-193.
ZHANG Huibo, WANG Shouxiang, ZHAO Qianyu, et al. Residential user load curve classification method considering data imbalance[J]. Electric Power Engineering Technology, 2022, 41(3):186-193.
- [12] CHEN J, YIN Z B, CHENG X M, et al. Big data analysis based identification method of low-voltage substation area[C]//2021 2nd International Conference on Big Data and Informatization Education (ICBDIE). Hangzhou, China. IEEE, 2021:169-172.
- [13] 赵日晓,闫冬,周翔,等. 人工智能支撑新型电力系统能源供给及消纳[J]. 全球能源互联网,2023,6(2):186-195.
ZHAO Rixiao, YAN Dong, ZHOU Xiang, et al. Artificial intelligence supports energy supply and consumption in new power system[J]. Journal of Global Energy Interconnection, 2023, 6(2):186-195.
- [14] 袁翔宇,张蓬鹤,熊素琴,等. 一种基于组合算法的异常用电模式辨识方法[J]. 电测与仪表,2023,60(6):160-166.
YUAN Xiangyu, ZHANG Penghe, XIONG Suqin, et al. Identification of abnormal power consumption mode based on combination algorithm[J]. Electrical Measurement & Instrumentation, 2023, 60(6):160-166.
- [15] 李琪林,严平,宿欣宇,等. 基于深度孪生自回归网络的无监督异常用电检测[J]. 计算机应用研究,2023,40(12):3717-3722,3727.
LI Qilin, YAN Ping, SU Xinyu, et al. Unsupervised abnormal power consumption detection via deep siamese autoregressive network[J]. Application Research of Computers, 2023, 40(12):3717-3722, 3727.
- [16] 钟庆,梁家豪,熊能,等. 基于量子聚类的稳态电能质量综合评估方法[J]. 电力系统及其自动化学报,2023,35(4):85-92.
ZHONG Qing, LIANG Jiahao, XIONG Neng, et al. Comprehensive evaluation method for steady-state power quality based on quantum clustering[J]. Proceedings of the CSU-EPSA, 2023,

- 35(4):85-92.
- [17] 白雅玲,周亚同,刘君. 基于深度卷积嵌入聚类的日负荷曲线聚类分析[J]. 电网技术,2022,46(6):2104-2113.
BAI Yaling, ZHOU Yatong, LIU Jun. Clustering analysis of daily load curve based on deep convolution embedding clustering[J]. Power System Technology, 2022, 46(6):2104-2113.
- [18] 孙园,李秋雨,黄冬梅,等. 基于粒计算和双尺度相似性的负荷曲线集成聚类算法[J]. 电力建设, 2022, 43(5):117-126.
SUN Yuan, LI Qiuyu, HUANG Dongmei, et al. Clustering ensemble model based on granular computing and dual-scale similarity[J]. Electric Power Construction, 2022, 43(5):117-126.
- [19] 孙毅,毛烨华,李泽坤,等. 面向电力大数据的用户负荷特性和可调节潜力综合聚类方法[J]. 中国电机工程学报, 2021, 41(18):6259-6270.
SUN Yi, MAO Yehua, LI Zekun, et al. A comprehensive clustering method of user load characteristics and adjustable potential based on power big data[J]. Proceedings of the CSEE, 2021, 41(18):6259-6270.
- [20] 周赣,茅欢,冯燕钧,等. 基于多特征符号聚合近似和层次聚类的户变关系识别方法[J]. 电力系统自动化, 2024, 48(3):133-141.
ZHOU Gan, MAO Huan, FENG Yanjun, et al. Identification method for user-transformer relationship based on multi-feature symbolic aggregate approximation and hierarchical clustering[J]. Automation of Electric Power Systems, 2024, 48(3):133-141.
- [21] 廖志伟,谢讯恺,郑广显,等. 基于电压曲线形态特征聚类的低压台区相位辨识[J]. 电力系统自动化, 2023, 47(7):142-149.
LIAO Zhiwei, XIE Xunkai, ZHENG Guangyu, et al. Phase identification of low-voltage distribution station area based on morphological characteristic clustering of voltage curves[J]. Automation of Electric Power Systems, 2023, 47(7):142-149.
- [22] 黄嘉楠,向小民,邢彧. 基于低电压治理的配电网优化策略研究[J]. 电力电容器与无功补偿, 2020, 41(1):200-206.
HUANG Jianan, XIANG Xiaomin, XING Yu. Study on optimization strategy of distribution network based on low voltage treatment[J]. Power Capacitor & Reactive Power Compensation, 2020, 41(1):200-206.
- [23] 林跃欢,黄彦璐,田兵,等. 配电网低电压治理方案动态评价方法研究[J]. 电测与仪表, 2023, 60(6):74-80.
LIN Yuehuan, HUANG Yanlu, TIAN Bing, et al. Research on dynamic evaluation method of low-voltage governance scheme in distribution network[J]. Electrical Measurement & Instrumentation, 2023, 60(6):74-80.
- [24] 贺春光,檀晓林,周兴华,等. 基于博弈论组合赋权的智能配电网项目投资效益评价[J]. 电力科学与技术学报, 2022, 37(1):161-167.
HE Chunguang, TAN Xiaolin, ZHOU Xinghua, et al. Investment benefit evaluation of intelligent distribution network project based on game theory combination weighting[J]. Journal of Electric Power Science and Technology, 2022, 37(1):161-167.
- [25] 黄子蒙,余娟,向明旭,等. 基于改进动态时间弯曲的PMU频率异常检测及类型识别[J]. 电力系统自动化, 2022, 46(24):104-112.
HUANG Zimeng, YU Juan, XIANG Mingxu, et al. Frequency anomaly detection and type identification of PMU based on improved dynamic time warping[J]. Automation of Electric Power Systems, 2022, 46(24):104-112.
- [26] 李克明,江亚群,黄世付,等. 基于DTW距离和聚类分析的配电台区低压拓扑结构辨识方法[J]. 电力系统保护与控制, 2021, 49(14):29-36.
LI Keming, JIANG Yaqun, HUANG Shifu, et al. Topology identification method of a low-voltage distribution station area based on DTW distance and cluster analysis[J]. Power System Protection and Control, 2021, 49(14):29-36.
- [27] 于惟坤,朱若源,陈旭,等. 基于双尺度相似性和改进DBSCAN算法的低压配电台区相序识别方法[J]. 电力建设, 2024, 45(9):74-88.
YU Weikun, ZHU Ruoyuan, CHEN Xu, et al. Phase sequence identification method for low-voltage distribution stations area based on dual-scale similarity and improved DBSCAN algorithm[J]. Electric Power Construction, 2024, 45(9):74-88.
- [28] 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.
- [29] 王晓霞,俞敏,霍泽健,等. 基于近邻传播聚类与LSTNet的分布式光伏电站群短期功率预测[J]. 电力系统自动化, 2023, 47(6):133-141.
WANG Xiaoxia, YU Min, HUO Zejian, et al. Short-term power forecasting of distributed photovoltaic station clusters based on affinity propagation clustering and long short-term time-series network[J]. Automation of Electric Power Systems, 2023, 47(6):133-141.
- [30] 赵庆明. 基于近邻传播算法和随机森林回归模型的台区线损率计算方法[J]. 电力系统及其自动化学报, 2020, 32(9):94-98.
ZHAO Qingming. Calculation method for line loss rate in transformer district based on affinity propagation algorithm and random forest regression model[J]. Proceedings of the CSU-EPSA, 2020, 32(9):94-98.
- [31] 陈泓光,刘俊勇,李林果,等. 基于AP聚类算法的充电站/光伏电站一体化规划方法[J]. 可再生能源, 2024, 42(10):1371-1380.
CHEN Xuanguang, LIU Junyong, LI Linguo, et al. Integrated planning method of charging station/PV integrated station based on AP clustering algorithm[J]. Renewable Energy Resources, 2024, 42(10):1371-1380.

- [32] 梁家豪. 基于时序特征的稳态电能质量评估与预警方法研究[D]. 广州:华南理工大学,2023.
LIANG Jiahao. Research on steady-state power quality assessment and early warning method based on time series characteristics[D]. Guangzhou;South China University of Technology, 2023.
- [33] 国家质量监督检验检疫总局,中国国家标准化管理委员会. 电能质量 供电电压偏差:GB/T 12325—2008[S]. 北京:中国标准出版社,2009.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Power quality-deviation of supply voltage;GB/T 12325-2008[S]. Beijing;

Standards Press of China,2009.

- [34] LI Y, GUO C H, SUN L L. Fast clustering by affinity propagation based on density peaks [J]. IEEE Access, 2020, 8: 138884-138897.

作者简介:



沈枢昊

沈枢昊(2000),男,硕士在读,研究方向为电能质量分析与控制(E-mail: 1494795180@qq.com);

钟庆(1978),男,博士,教授,研究方向为电能质量分析与控制;

许中(1986),男,硕士,高级工程师,从事电能质量分析与电力技术监督管理工作。

A method for mining over-limit patterns of low voltage in users based on hierarchical affinity propagation clustering

SHEN Shuhao¹, ZHONG Qing¹, XU Zhong², WANG Gang¹, LI Haifeng¹, WANG Longjun¹

(1. School of Electric Power Engineering, South China University of Technology, Guangzhou 510640, China;

2. Guangzhou Power Supply Bureau Co., Ltd., Guangzhou 510600, China)

Abstract: Mining different over-limit patterns of the low voltage is very important to guide the management of low voltage issues in users. Due to the complexity and the ever-changing nature of voltage, the over-limit patterns of low voltage are always inherently unknown in users. A pattern mining method for low voltage in users based on hierarchical affinity propagation clustering (HAP) is proposed in this paper. Firstly, large-scale voltage data is clustered into several clusters using the HAP clustering algorithm, and these clusters are regarded as the different over-limit patterns of low voltage. Then, four indices are defined from two aspects of the duration and amplitude to characterize the features of the clusters. The features of the over-limit patterns are then derived by calculating the indices for each cluster. Finally, the proposed method is applied to a real dataset, effectively mining four over-limit patterns of low voltage. The characteristics of different patterns provide the important information for the supervision and analysis of low voltage issues in users, and the priorities of low voltage problems management in users can be well leveled.

Keywords: low voltage users; hierarchical affinity propagation (HAP) clustering; over-limit patterns of low voltage; over-limit duration; over-limit voltage amplitude; management priority

(编辑 陆海霞)