

DOI:10.12158/j.2096-3203.2022.04.009

基于相关性度量算法的台区线损异常判断及精准定位

陈光宇¹, 徐嘉杰¹, 卢兆军², 袁飞³, 张仰飞¹, 郝思鹏¹

(1. 南京工程学院电力工程学院, 江苏 南京 211167; 2. 国网山东省电力公司, 山东 济南 250001; 3. 国网山东省电力公司泰安供电公司, 山东 泰安 271000)

摘要:针对台区发生线损异常时关联用户辨识困难的实际问题,提出一种基于相关性度量算法的台区线损异常判断及精准定位方法。首先,通过间隙统计-轮廓系数融合算法确定数据集的最佳聚类数,并在此基础上采用二分K-means++构建台区线损标准库;其次,基于标准库完成台区线损异常辨识,确定异常时间段;再次,计算异常时间段内各用户电量和线损的斯皮尔曼相关性系数(SCC)和欧式-离散弗雷歇距离(E-DFD),并基于SCC和E-DFD构造综合评判指标分析用户关联性;最后,采用逼近理想解排序法(TOPSIS)对综合评判指标值进行排序,实现异常关联用户的精准定位。算例采用某台区真实现场数据进行分析,结果表明文中所提方法在聚类有效性、计算时间以及辨识准确度等方面具有较好的性能。

关键词:台区线损;异常判断;精准定位;线损标准库;综合评判指标;逼近理想解排序法(TOPSIS)

中图分类号:TM714

文献标志码:A

文章编号:2096-3203(2022)04-0067-08

0 引言

线损是反映电网规划设计、生产技术和运营管理水平的一个综合性技术经济指标^[1-3]。目前台区线损异常等问题普遍存在,其排查仍依靠运维人员的经验,这不但极大地增加了工作量,而且还会出现部分异常难以发现的情况,导致电量损失^[4-6]。因此,如何有效利用各系统间数据、深化数据分析、准确锁定异常用户成为了亟待解决的问题^[7-8]。

目前已有部分学者在台区线损异常研判方面进行了深入研究,并取得不少成果。文献[9-11]分别引入无监督学习算法和神经网络模型,解决了台区线损异常数据辨识困难的问题。文献[12-15]从数据挖掘的角度出发,通过分析用户用电行为实现对异常用户的检测。以上文献重点研究了线损和用户侧的异常问题,但没有涉及用户侧电量异常和台区线损异常间的内在关联,有造成异常误判的风险。文献[16]引入皮尔逊积距相关系数用以计算用户电量与对应配电房线损的曲线相似度,进而实现了违约用电异常用户的精准定位。文献[17]着重分析了线损电量与用户电量的线性关系,提出一种采用皮尔逊相关系数算法查找引起电量损失的异常电能表的方法,避免了定性分析的盲目性和不确定性。以上方法分别从形状^[16]和数值相似性^[17]层面展开,探讨了线损和用户负荷之间的某种函数关系,并取得了较好的效果,但仅从形状或相似性来定位用户仍存在片面性,尤其是应用于类

型复杂多样的台区时。

针对台区线损异常辨识困难,关联用户排查耗时且难以精准定位的问题,文中提出一种基于相关性度量算法的台区线损异常判断及关联用户精准定位方法。基于间隙统计-轮廓系数(gap static algorithm-silhouette coefficient, GSA-SC)融合算法确定数据集的最佳聚类数,采用二分K-means++完成台区线损标准库的构建,实现标准库的完备划分,并给出一种基于斯皮尔曼相关性系数(Spearman correlation coefficient, SCC)和欧式-离散弗雷歇距离(Euclidean-discrete Fréchet distance, E-DFD)的综合评判指标构建策略,最终实现异常用户的精准定位。算例采用某台区真实现场数据进行仿真分析,结果表明,文中构建的标准库能够较好地给出台区的线损标准,实现对台区异常线损率的智能研判。而提出的异常定位策略相比于传统关联算法能更加准确地发现可疑用户,降低误判概率。

1 基于台区线损标准库的异常判断策略

当前大多数台区线损的异常判断依靠企业制定的统一标准或运维人员自身经验进行衡量,但配电网规模日趋庞大,台区线损情况日益复杂,传统的线损标准和人工经验已难以应对电网发展的需求。因此,为实现线损异常的自动判别,提出一种基于台区线损标准库的异常判断策略,具体如图1所示。

1.1 GSA-SC的求解

SC是常用的聚类算法评价指标^[18],针对任意

收稿日期:2022-02-15;修回日期:2022-05-11

基金项目:江苏省自然科学基金资助项目(BK20181021)

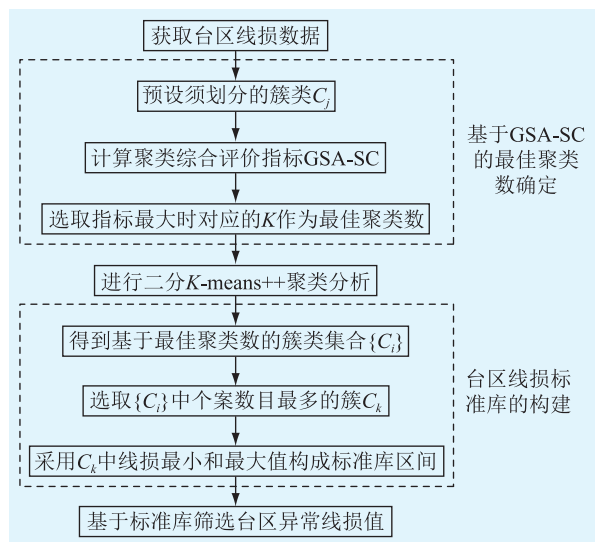


图1 基于台区线损标准库的异常判断策略

Fig.1 Anomaly judgment strategy based on line loss standard library

簇类 C_j , 单个样本的 SC 计算为:

$$s = \frac{b - a}{\max(a, b)} \quad (1)$$

其中:

$$a = \frac{1}{n} \sum_{x_i \in C_j} (x_i - c_j)^2 \quad (2)$$

$$b = \frac{1}{m} \sum_{x_i \in C_j, x_k \in C_l} (x_i - x_k)^2 \quad (3)$$

式中: s 为单个样本的 SC; a 为样本 x_i 与其自身所在类 C_j 中所有其他点的平均距离; b 为样本 x_i 与其距离最近的类 C_l 中所有点的平均距离; c_j 为簇类 C_j 的质心; m, n 分别为类 C_j 和 C_l 中的样本个数。

然而, SC 在凸型簇类上的表现会虚高, 因此在 SC 的基础上引入 GSA 算法, GSA 能估计数据集最佳聚类个数, 且在聚类数量设定值较低的场景下效果明显^[19]。

定义 K 个簇的聚类离散度为:

$$W(K) = \sum_{j=1}^K \sum_{x_i \in C_j} (x_i - c_j)^2 \quad (4)$$

GSA 算法的主要思路为: 对于每次设定的聚类数 K , 计算所对应聚类离散度的自然对数, 并与阈值进行对比, 进而确定最佳聚类数。经过自然对数处理后的数据态势更趋于线性化, 使聚类离散度与阈值的间隙值更方便确定, 间隙值定义如下:

$$A(K) = E \ln W_r(K) - \ln W(K) \quad (5)$$

式中: $A(K)$ 为间隙值; $W_r(K)$ 为选取的参考数据集的聚类离散度; E 为参考数据集的数学期望。

随着 K 值的不断增加, 当首次出现某个较大的 $A(K)$, 且变化趋势趋于稳定时, 认为此时的 K 为最

佳聚类数。

然而, 单纯采用 GSA 或 SC 方法进行最佳聚类数的确定, 可能会出现误判, 导致聚类效果不明显。因此, 为提高聚类结果的准确度, 引入 GSA-SC 融合算法确定最佳聚类数, 结合式(1)和式(5), 计算 $A(K)$ 与 s 的平均值, 构造新的聚类评价指标 G 。

$$G = \frac{A(K) + s}{2} \quad (6)$$

1.2 线损长期异常台区的剔除策略

为了实现台区线损异常的自动辨识, 首先须找到台区线损的正常线损范围, 故考虑通过聚类算法挖掘出台区的线损标准形态, 以此作为线损异常判别的标准。然而, 对于线损长期异常且稳定的台区, 难以构建较为准确的标准库, 会对线损的异常辨识产生影响, 可能造成误判, 因此须剔除线损长期异常的台区。

设原始线损序列为 $Y(t) = \{y(1), y(2), \dots, y(N)\}$, 对比分析线损数据和传统人工制定的标准区间(经验值为 $0\% \sim 8\%$ ^[20]), 计算异常线损的占比 α , 并设置阈值分析, 进而确定台区特征, 判断是否进行剔除。其中, α 的表现形式如下:

$$\alpha = \frac{k}{N} \quad (7)$$

式中: k 为线损序列 $Y(t)$ 中超出传统标准区间的线损个数; N 为原始采集的线损总个数。

当 $\alpha > \beta$ ($\beta = 0.5$) 时, 判定为长期异常台区; 反之, 则进行下一步研究。

1.3 台区线损标准库的生成策略

剔除线损长期异常台区后, 采用聚类算法对线损异常特征较为明显的台区进行分析, 完成线损标准库的构建。台区线损标准库的生成步骤如下:

(1) 剔除线损长期异常的台区。

(2) 基于最佳聚类数(为便于说明, 此处假设最佳聚类数为 3), 采用二分 K -means++^[21] 对台区线损进行聚类分析, 划分出 A、B、C 三大簇类, 如图 2 所示。

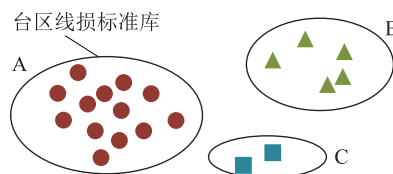


图2 基于二分 K-means++的台区线损聚类示意

Fig.2 Schematic diagram of line loss clustering in station area based on dichotomous K-means++

(3) 选取个案数目最多的簇类(簇类 A)作为台区线损标准库, 而簇类 B 和簇类 C 中的线损数据

自动归为异常点。

(4) 分析台区线损标准库(簇类 A)中的元素分布,选取线损最小值 y_{\min} 作为标准库区间的下限,最大值 y_{\max} 作为标准库区间的上限,进而实现台区线损标准库区间的完备划分。

1.4 基于标准库的线损异常自动判断

基于标准库对台区给定时间段内的线损异常情况进行智能辨识,如图 3 所示。

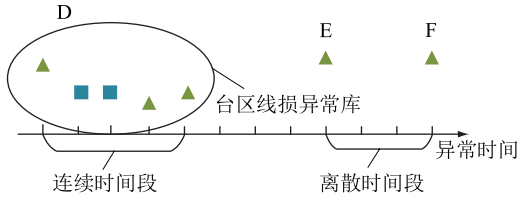


图 3 基于标准库的线损异常自动判断示意

Fig.3 Schematic diagram of automatic judgment of abnormal line loss based on the standard library

由图 2 可知,簇类 B 和簇类 C 中的线损均不在标准库内,故初步判定为异常。然而,依据运维人员的排查经验和专家建议,一般来说,异常线损对应的时间段要求满足连续性,即由若干个连续的日期组成。从图 3 可以看出,区域 D 中的异常线损分布在连续时间段内,而 E 和 F 均是离散点,因此,选取区域 D 中的线损构成台区的线损异常库。

2 台区线损异常关联用户精准定位方法

前文分析了基于台区线损标准库的异常判断策略,但台区线损的异常通常是由台区内用户电量异常造成,而台区内用户众多,准确辨识出导致台区线损率异常的用户的难度较大。传统基于距离^[16]或相关系数方法^[17]的异常定位策略仅考虑从形状或相似度来定位用户,在计算时间和准确度方面表现不稳定。因此文中考虑采用 SCC 和 E-DFD 构造一种新的综合评判指标来分析用户关联性,通过引入逼近理想解排序法(technique for order preference by similarity to an ideal solution, TOPSIS)对综合评判指标值进行排序,挖掘出最可能发生异常的用户,实现对异常关联用户的精准定位。

2.1 基于 SCC 和 E-DFD 的综合评价指标构建

2.1.1 斯皮尔曼相关性系数(SCC)

由于文中变量(线损和用户电量)的量纲不同,若使用协方差或皮尔逊相关系数计算其相似性,则须消除量纲后再分析,而 SCC 算法本身就是一个消除量纲的过程^[22],因此,采用 SCC 计算并分析台区线损和异常用户电量的关联程度更合适。

确定每个变量在数据序列上所属的等级,并计

算 2 组数据的等级差,SCC 的具体计算为:

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^T d_i^2}{M(M^2 - 1)} \quad (8)$$

式中: M 为样本量; d_i 为数据 x 和 y 在异常时间点 i 中用户电量和台区线损的等级差; T 为异常时间段。 r_{xy} 的绝对值越大表明相关性越强。

2.1.2 欧式-离散弗雷歇距离(E-DFD)

DFD 适用于路径空间的相似性度量,考虑相似度的同时还考虑了路径空间距离的因素,对于二维平面上 2 条曲线的相似度判断有良好的适用性。

由于用户电量和线损是等长时间序列且满足离散分布,因此文中在 DFD 的求解过程中,结合欧式距离判断异常库中各用户电量曲线与线损曲线的相似度,得到 E-DFD 距离。设用户电量序列分布为 $X(t) = \{x(1), x(2), \dots, x(N)\}$,线损序列分布为 $Y(t) = \{y(1), y(2), \dots, y(N)\}$,则此时对应的点对序列分布如下:

$$\sigma(X, Y) = \{(x(1), y(1)), (x(2), y(2)), \dots, (x(N), y(N))\} \quad (9)$$

定义 $\sigma(X, Y)$ 中序列对之间的长度 $\|L\|$ 为各序列对中欧式距离 d 最大的值:

$$\|L\| = \max_{s=1,2,\dots,N} d(x(s), y(s)) \quad (10)$$

此时, E-DFD 的表达式如下:

$$F(X, Y) = \min\{\|L\|\} \quad (11)$$

2.1.3 综合评价指标的构造

SCC 和 E-DFD 分别从数值相似性和形状相似性 2 个角度来描述用户和台区线损之间的关联,但 2 个指标各有利弊,因此文中考虑构建新的指标集合 P , 包含 SCC 和 E-DFD, 表示形式如下:

$$P = \{Q(r_{SCC}, F_{E-DFD})\} \quad (12)$$

式中: $Q(r_{SCC}, F_{E-DFD})$ 为关于 SCC 和 E-DFD 的函数。

SCC 取值范围为 $-1 \sim 1$, 其值越大, 相似程度越高, 而 E-DFD 越大相似程度越低, 因此 SCC 和 E-DFD 呈现反向状态, 不利于计算, 作同向化处理, 具体如下:

$$P = Q(|r_{xy}|, e^{-F(X, Y)}) \quad (13)$$

针对 SCC 求绝对值, 对 E-DFD 求 $e^{-F(X, Y)}$, 使其变换形式后的值在 $0 \sim 1$ 内, 同时也满足 E-DFD 越大, 相似程度越低的的关系。

2.2 基于综合评价指标的关联异常用户定位

TOPSIS 是一种常用的组内综合评价方法, 常用于解决多属性决策问题, 其结果能精确地反映各评价方案之间的差距^[23]。因此, 采用 TOPSIS 融合

SCC 和 E-DFD 确定台区异常用户,具体流程如图 4 所示。

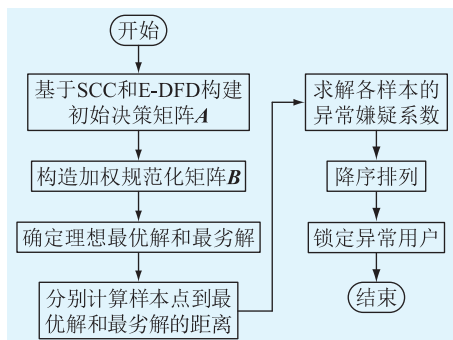


图 4 基于 TOPSIS 的关联性联合定位流程
Fig.4 Flow chart of the TOPSIS-based associated joint positioning

具体步骤如下:

(1) 根据式(8)和式(11),构建初始决策矩阵

$A = (A_{ij})_{N \times 2}$ 为:

$$A = \begin{bmatrix} r_1 & F_1 \\ r_2 & F_2 \\ \vdots & \vdots \\ r_N & F_N \end{bmatrix} \quad (14)$$

式中: r_i 为 SCC 参量; F_i 为 E-DFD 参量。

(2) 分别对指标属性 SCC 和 E-DFD 进行同向化和规范化,生成新的规范化决策矩阵 $B = (B_{ij})_{N \times 2}$ 。

$$B = \begin{bmatrix} k_1 r_1 & k_2 e^{-F_1} \\ k_1 r_2 & k_2 e^{-F_2} \\ \vdots & \vdots \\ k_1 r_N & k_2 e^{-F_N} \end{bmatrix} \quad (15)$$

其中:

$$k_j = \begin{cases} \sqrt{\sum_{i=1}^N A_{ij}^2} & j = 1 \\ \sqrt{\sum_{i=1}^N (e^{-A_{ij}})^2} & j = 2 \end{cases} \quad (16)$$

(3) 确定理想最优解 s^+ 和最劣解 s^- 。

$$s^+ = [\max\{B_{i1} | 1 \leq i \leq N\} \quad \max\{B_{i2} | 1 \leq i \leq N\}] \quad (17)$$

$$s^- = [\min\{B_{i1} | 1 \leq i \leq N\} \quad \min\{B_{i2} | 1 \leq i \leq N\}] \quad (18)$$

(4) 求解规范化决策矩阵 B 中行向量与最优、最劣解的 L^2 范数(即欧式距离) d_i^+ 和 d_i^- 。

$$d_i^+ = \|s^+ - B_i\|_2 \quad (19)$$

$$d_i^- = \|s^- - B_i\|_2 \quad (20)$$

(5) 计算台区第 i 个用户基于 L^2 范数的综合评

价指标 S_i ,并根据 S_i 大小进行排序,输出各用户样本 TOPSIS 评价结果。

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (21)$$

(6) 根据台区异常判断的综合评判指标 S_i 值的最终大小,确定用户异常嫌疑系数,从而精准定位异常用户。

3 台区线损异常判断及关联用户精准定位框架

综上所述,在文中场景下,对造成线损异常的台区以及台区下的所有用户进行深入研究和挖掘,基于数据挖掘理论,进一步实现台区线损异常情况的智能辨识和精准定位,流程如图 5 所示。

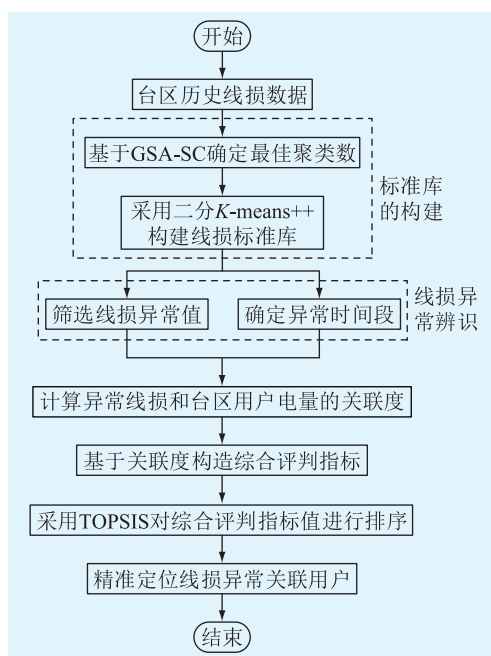


图 5 台区线损异常判断及关联用户精准定位框架

Fig.5 A framework for judging abnormal line loss in station area and precise positioning of associated users

(1) 基于聚类分析的线损标准库生成策略。针对台区的历史线损数据进行聚类分析,基于 GSA-SC 算法确定最佳聚类数目,采用二分 K -means++完成线损的聚类,然后研究各簇类的个案数目,进一步归纳和构建台区线损标准库。

(2) 基于标准库的台区线损异常辨识。以台区线损标准库为基础,对历史数据进行挖掘,比较台区线损和标准库之间的差异,确定线损异常的时间段以及线损异常数据。

(3) 台区线损异常关联用户的精准定位。首先,分析异常线损和台区用户电量间的关联性,采用 SCC 和 E-DFD 方法定量计算异常线损和台区用

户间的关联度。其次,依据相似度计算结果,构造综合评判指标分析用户关联性,并采用 TOPSIS 对综合评判指标值进行降序排列。最后,辅助台区用户电量和台区线损之间的内在关联,协同判断台区线损异常的关联用户,实现精准定位。

4 算例验证与分析

算例数据采用某配电网公用变台区真实用户电量和线损率。其中,台区容量为 $400 \text{ MV}\cdot\text{A}$,时间跨度为半年(2020年1月至7月),采样频率为 1 d,原始线损率数据共 183 条,负荷数据共 535 条。

4.1 基于台区线损标准库的异常自动辨识分析

对台区历史线损数据进行归一化处理。图 6 为台区历史线损实际值和归一化分布,可以看出,线损主要分布在 $[0, 2.5]$ 内,但存在 1 个孤立点(实际值为 20.45)。若采用传统聚类算法进行分析,由于聚类数为超参数,须人为输入,导致样本的最佳聚类数目不易确定,具有较强的随机性,因此考虑采用 GSA-SC 融合算法确定线损集的最佳聚类数。

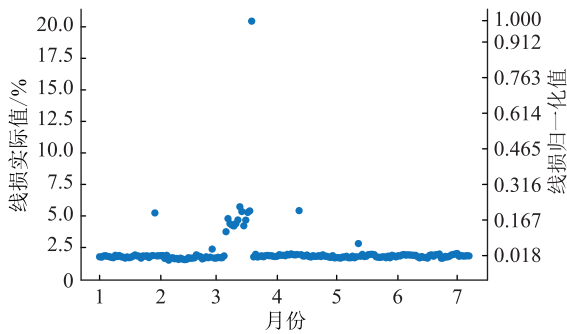


图 6 台区历史线损分布示意

Fig.6 Schematic diagram of historical line loss distribution in station area

为验证文中所提 GSA-SC 算法的优越性,分别采用 GSA 和 SC 进行对比分析,结果如表 1 所示,其中 $A(K)$, s , G 分别为 GSA, SC, GSA-SC 算法的聚类评价指标。

表 1 3 种不同算法的聚类指标结果对比

Table 1 Comparison of clustering index results of three different algorithms

K	$A(K)$	s	G
2	0.241 0	0.420 0	0.330 5
3	0.302 0	0.430 0	0.366 0
4	0.305 0	0.350 0	0.327 5
5	0.304 0	0.312 0	0.308 0
6	0.311 0	0.290 0	0.300 0

由表 1 可见,若仅采用 GSA 方法判断,当 $K \geq 3$ 时, $A(K)$ 的变化趋势不明显,各数值非常相近,导致

最佳聚类数不易确定;若仅采用 SC 算法分析,当 K 取 2 和 3 时,其对应的 s 值接近,最佳聚类数也不易确定。而采用文中所提的 GSA-SC 算法时,容易确定 K 为 3 时, G 值最大,且无相邻点值与之相似。因此,选取最佳聚类数为 3。

基于最佳聚类数对台区线损样本进行聚类分析,聚类结果如图 7 所示。可知,紫色部分的聚类结果为聚类的主体,故判别为标准库,而橙色和绿色部分均不在标准库内,可判为异常。

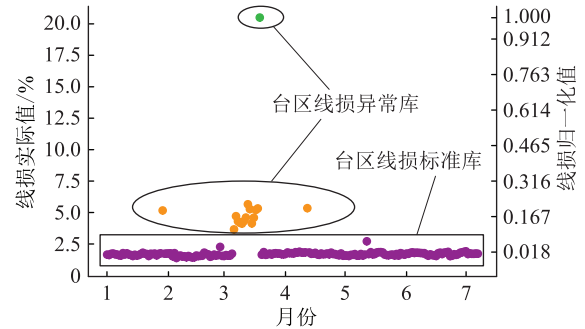


图 7 台区线损聚类结果

Fig.7 Area line loss clustering result

结合 IBM SPSS Statistics 软件实现数据聚类,结果如下:簇类 1、簇类 2、簇类 3 的聚类中心分别为 0.167, 0.018, 1; 个案数目分别为 15, 172, 1。可以看出,当簇类数设置为最佳簇类数 3 时,得到的最终聚类中心和每个聚类中的个案数目与图 7 中完全吻合。个案数目最多的簇类为簇类 2,共有 172 个样本数据,簇类 2 的聚类中心为 0.018,计算该簇中样本到聚类中心的最大上限距离 r_{upper} 和最大下限距离 r_{low} ,构成该台区的历史线损标准库,不难求出此时 r_{upper} 为 0.095, r_{low} 为 0.003,而标准库是由归一化之前系统采集的真实值组成的,故其应为 $[0.6, 3]$ 。

台区线损异常库的生成规则与标准库完全一致,由图 7 可知,该台区的历史线损异常库由两部分组成:一部分是以 0.167 为聚类中心,个案数目为 15 的簇类;另一部分是以 1 为聚类中心,个案数目为 1 的簇类。鉴于文中规定的 T 必须是连续的时间区间,故忽略线损波动小且周期为 1~2 d 的时间段,故异常时间可初步判断为 3 月 6 日—3 月 19 日。

图 8 为传统人工制定标准与文中构建的台区线损标准库对比示意。从图中可以看出,传统人工制定的正常台区线损范围应在 $0\% \sim 8\%$,相比于文中构建的标准库而言,另外囊括了一些线损数据,但根据现场排查发现其中若干个点也存在异常嫌疑。这也从侧面说明了文中所提的台区线损标准库构建策略满足工程实际应用的要求,在节约运维成本的同时提高了异常辨识的准确性。

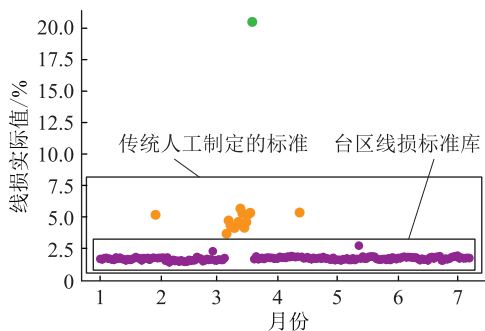


图8 传统人工制定与文中的台区线损正常区间对比

Fig.8 The normal interval of line loss comparison between the traditional manual establishment and the method proposed in this paper

4.2 线损异常关联用户精准定位分析

提取异常库中的线损数据,根据式(8)和式(11),分别计算异常时间段 T 内各用户电量和对应线损的 SCC 和 E-DFD (SCC 作绝对值化处理)。选取 SCC 大于阈值 $0.6^{[24]}$ 的用户并按用电量大小降序排列,计算结果见表 2。

表2 SCC 和 E-DFD 计算结果

Table 2 SCC and DFD calculation results

用户编号	r_{xy}	$F(X, Y)$
a	0.767 5	1.183 6
b	0.761 7	1.099 7
c	0.605 2	1.483 4
d	0.619 8	2.321 7
e	0.662 9	2.604 4

基于上述的计算结果,求得规范化决策矩阵为:

$$B = \begin{bmatrix} 0.524 & 0.587 \\ 0.411 & 0.639 \\ 0.413 & 0.435 \\ 0.423 & 0.188 \\ 0.453 & 0.142 \end{bmatrix} \quad (22)$$

由于经正向化和标准化处理的规范化决策矩阵中的指标均是极大型数据,故可取出每一列的最大值构成理想最优向量,每一列的最小值构成理想最劣向量,即:

$$s^+ = [0.524 \ 0.639] \quad (23)$$

$$s^- = [0.411 \ 0.142] \quad (24)$$

采用 TOPSIS 求得表 2 中 5 个存在异常嫌疑的用户与理想最优、劣向量的 L^2 范数 d_i^+, d_i^- 与台区异常判断的综合评判指标 S_i 值,具体结果如表 3 所示。图 9 为异常时间段内编号 a 和 b 的用户电量与台区线损的变化走势。从图中可以看出,由于 SCC 给出用户 a 和 b 的值大小相近,故这 2 家用户的电量曲线和线损曲线的趋势相似,因此,可初步断定以上 2

个用户均存在异常嫌疑。

表3 基于初始化决策矩阵求得的 d^+ 和 d^- 及 S_i

Table 3 d^+, d^- and S_i obtained based on the initial decision matrix

用户编号	r_{xy}	$e^{-F(X, Y)}$	d^+	d^-	S_i	排序
a	0.767 5	0.306 1	0.051 4	0.460 1	0.899 4	1
b	0.761 7	0.332 9	0.113 3	0.497 4	0.614 4	2
c	0.605 2	0.226 8	0.231 9	0.293 6	0.558 6	3
d	0.619 8	0.098 1	0.462 1	0.048 0	0.094 1	4
e	0.662 9	0.073 9	0.502 5	0.041 8	0.076 8	5

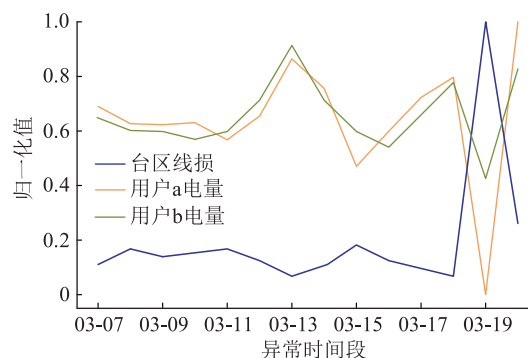


图9 异常时间段台区线损与嫌疑用户电量变化

Fig.9 Line loss and suspected user power change in abnormal time

进一步分析表 3 可知,若仅考虑采用 E-DFD 进行异常辨识,则用户 b 被判别为异常用户,但与实际情况不符。若采用文中方法可得到用户 a 的综合评价指标值最大,为 0.899 41,且与排序第 2 的用户指标相差较多,故可确定该用户异常用电嫌疑最大。对台区异常用户进行现场排查校验,结果表明文中所提方法较为准确。

5 结论

文中提出一种基于相关性度量算法的台区线损异常判断及精准定位方法,主要结论如下:

(1) 提出一种基于 GSA-SC 的二分 K -means++ 聚类算法用于构建台区线损标准库,该算法能够改善传统聚类算法中最佳聚类数不易确定的问题,提高聚类收敛性和效率。

(2) 构建一种新的台区线损标准库,相比传统人工制定的台区线损标准更为精细,其能更好地描述台区特征,实现对台区线损异常的准确辨识。

(3) 给出一种基于 SCC 和 E-DFD 的综合评判指标,能较为精确地分析台区线损异常和用户的内在关联性,实现台区线损异常关联用户的精准定位,降低误判概率。

文中对台区线损异常智能研判进行了研究,随

着配电网规模日趋庞大,台区线损情况日益复杂,同一台区的线损也会因台区内用户的用电特征和规律产生周期性变化,这种波动受到气候、节假日等多种因素的影响,因此如何准确把握台区线损的真实情况有待于进一步研究。

本文得到国网山东省电力公司科技项目(5206-091900C7)资助,谨此致谢!

参考文献:

- [1] 李滨,刘铸峰,黄柳军,等. 县级电网线损管理综合对标评价[J]. 电网技术,2016,40(12):3871-3880.
LI Bin, LIU Zhufeng, HUANG Liujun, et al. Comprehensive evaluation benchmark study on line loss management of county power grids [J]. Power System Technology, 2016, 40(12): 3871-3880.
- [2] 唐冬来,刘友波,熊智临,等. 基于时空关联矩阵的配电台区反窃电预警方法[J]. 电力系统自动化,2020,44(19):168-176.
TANG Donglai, LIU Youbo, XIONG Zhilin, et al. Early warning method of electricity anti-theft in distribution station area based on spatiotemporal correlation matrix[J]. Automation of Electric Power Systems, 2020, 44(19): 168-176.
- [3] 李亚,刘丽平,李柏青,等. 基于改进 K-Means 聚类 and BP 神经网络的台区线损率计算方法[J]. 中国电机工程学报,2016,36(17):4543-4552.
LI Ya, LIU Liping, LI Baiqing, et al. Calculation of line loss rate in transformer district based on improved K-Means clustering algorithm and BP neural network[J]. Proceedings of the CSEE, 2016, 36(17): 4543-4552.
- [4] 徐焕增,孔政敏,王帅,等. 基于动态线损及 FMRLS 算法的智能电表误差在线评估模型[J]. 中国电机工程学报,2021,41(24):8349-8358.
XU Huanzeng, KONG Zhengmin, WANG Shuai, et al. Online error evaluation model of smart meter based on dynamic line loss and FMRLS algorithm[J]. Proceedings of the CSEE, 2021, 41(24): 8349-8358.
- [5] 殷涛,薛阳,杨艺宁,等. 基于向量自回归模型的高损线路窃电检测[J]. 中国电机工程学报,2022,42(3):1015-1024.
YIN Tao, XUE Yang, YANG Yining, et al. Electricity theft detection of high-loss line with vector autoregression[J]. Proceedings of the CSEE, 2022, 42(3): 1015-1024.
- [6] 王方雨,刘文颖,李潇,等. 考虑线损灵敏度一致性的外网静态等值模型[J]. 电网技术,2020,44(6):2295-2305.
WANG Fangyu, LIU Wenying, LI Xiao, et al. Static equivalent model of external network considering consistency of line loss sensitivity[J]. Power System Technology, 2020, 44(6): 2295-2305.
- [7] 钟小强,陈杰,蒋敏敏,等. 基于深度学习的台区线损分析方法[J]. 电网技术,2020,44(2):769-774.
ZHONG Xiaoqiang, CHEN Jie, JIANG Minmin, et al. A line loss analysis method based on deep learning technique for trans-
- former district [J]. Power System Technology, 2020, 44(2): 769-774.
- [8] 王若丞. 配电网三相不平衡对线损增加率及电压偏移的影响[J]. 电力工程技术,2017,36(4):131-136.
WANG Ruocheng. Influence of distribution network three-phase unbalance on line loss increase rate and voltage offset[J]. Electric Power Engineering Technology, 2017, 36(4): 131-136.
- [9] 王方雨,刘文颖,陈鑫鑫,等. 基于“秩和”近似相等特性的同期线损异常数据辨识方法[J]. 电工技术学报,2020,35(22):4771-4783.
WANG Fangyu, LIU Wenying, CHEN Xinxin, et al. Abnormal data identification of synchronous line loss based on the approximate equality of rank sum[J]. Transactions of China Electrotechnical Society, 2020, 35(22): 4771-4783.
- [10] 代杰杰,宋辉,杨祎,等. 基于栈式降噪自编码器的输变电设备状态数据清洗方法[J]. 电力系统自动化,2017,41(12):224-230.
DAI Jiejie, SONG Hui, YANG Yi, et al. Cleaning method for status data of power transmission and transformation equipment based on stacked denoising autoencoders[J]. Automation of Electric Power Systems, 2017, 41(12): 224-230.
- [11] ZHAO G A, CHU J X, 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). Chengdu, China. IEEE, 2020:192-196.
- [12] 游文霞,申坤,杨楠,等. 基于 Bagging 异质集成学习的窃电检测[J]. 电力系统自动化,2021,45(2):105-113.
YOU Wenxia, SHEN Kun, YANG Nan, et al. Electricity theft detection based on bagging heterogeneous ensemble learning [J]. Automation of Electric Power Systems, 2021, 45(2): 105-113.
- [13] 游文霞,申坤,杨楠,等. 基于 AdaBoost 集成学习的窃电检测研究[J]. 电力系统保护与控制,2020,48(19):151-159.
YOU Wenxia, SHEN Kun, YANG Nan, et al. Research on electricity theft detection based on AdaBoost ensemble learning [J]. Power System Protection and Control, 2020, 48(19): 151-159.
- [14] 许刚,谈元鹏,戴腾辉. 稀疏随机森林下的用电侧异常行为模式检测[J]. 电网技术,2017,41(6):1964-1973.
XU Gang, TAN Yuanpeng, DAI Tenghui. Sparse random forest based abnormal behavior pattern detection of electric power user side [J]. Power System Technology, 2017, 41(6): 1964-1973.
- [15] 庄池杰,张斌,胡军,等. 基于无监督学习的电力用户异常用电模式检测[J]. 中国电机工程学报,2016,36(2):379-387.
ZHUANG Chijie, ZHANG Bin, HU Jun, et al. Anomaly detection for power consumption patterns based on unsupervised learning [J]. Proceedings of the CSEE, 2016, 36(2): 379-387.
- [16] 陈仕军,王长江. 基于曲线相似度和关联分析的窃电智能识别与预警[J]. 浙江电力,2017,36(12):33-36,62.

- CHEN Shijun, WANG Changjiang. Intelligent identification and early warning against power theft based on curve similarity and correlation analysis [J]. Zhejiang Electric Power, 2017, 36(12):33-36,62.
- [17] 王涓, 吴旭鸣, 王爱凤. 应用皮尔逊相关系数算法查找异常电能表用户[J]. 电力需求侧管理, 2014, 16(2):52-54.
WANG Juan, WU Xuming, WANG Aifeng. The application of Pearson correlation coefficient algorithm in searching for the users with abnormal watt-hour meters[J]. Power Demand Side Management, 2014, 16(2):52-54.
- [18] LI R M, HUANG Y F, WANG J. Long-term traffic volume prediction based on K-means Gaussian interval type-2 fuzzy sets [J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(6):1344-1351.
- [19] 吴军基, 杨伟, 葛成, 等. 基于 GSA 的肘形判据用于电力系统不良数据辨识[J]. 中国电机工程学报, 2006, 26(22):23-28.
WU Junji, YANG Wei, GE Cheng, et al. Application of GSA-based elbow judgment on bad-data detection of power system [J]. Proceedings of the CSEE, 2006, 26(22):23-28.
- [20] 国网浙江省电力公司. 台区线损异常治理一本通[M]. 北京: 中国电力出版社, 2016: 12-16.
State Grid Zhejiang Electric Power Company. Control of abnormal line loss in station area [M]. Beijing: China Electric Power Press, 2016: 12-16.
- [21] KŁOPOTEK M A. An aposterioral clusterability criterion for K-means++ and simplicity of clustering[J]. SN Computer Science, 2020, 1(2):1-38.
- [22] 贾科, 杨哲, 魏超, 等. 基于斯皮尔曼等级相关系数的新能源送出线路纵联保护[J]. 电力系统自动化, 2020, 44(15):103-111.
JIA Ke, YANG Zhe, WEI Chao, et al. Pilot protection based on Spearman rank correlation coefficient for transmission line connected to renewable energy source[J]. Automation of Electric Power Systems, 2020, 44(15):103-111.
- [23] 范洁, 杨世海, 武文广, 等. 基于组合赋权的非介入式终端辨识能力灰色 TOPSIS 评价[J]. 电力工程技术, 2020, 39(6):110-116, 131.
FAN Jie, YANG Shihai, WU Wenguang, et al. Grey TOPSIS evaluation of non intrusive terminal identification ability based on combination weighting [J]. Electric Power Engineering Technology, 2020, 39(6):110-116, 131.
- [24] BAKHTIAR A, SULIANTORO H, NINGSI R H, et al. Relationship of quality management system standards to industrial property rights in Indonesia using Spearman correlation analysis method[J]. IOP Conference Series: Earth and Environmental Science, 2021, 623(1):012092.

作者简介:



陈光宇

陈光宇(1980),男,博士,副教授,研究方向为电力系统运行与控制(E-mail: cgyhhu@163.com);

徐嘉杰(1997),男,硕士在读,研究方向为电力系统运行与控制;

卢兆军(1975),男,硕士,高级工程师,从事电网企业经营计划管理工作。

Judgment and precise location of abnormal line loss in station area based on correlation measurement algorithm

CHEN Guangyu¹, XU Jiajie¹, LU Zhaojun², YUAN Fei³, ZHANG Yangfei¹, HAO Sipeng¹

(1. School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China;

2. State Grid Shandong Electric Power Company, Jinan 250001, China;

3. State Grid Shandong Electric Power Company, Taian Power Supply Company, Taian 271000, China)

Abstract: Aiming at the practical problem of the difficulty in identifying associated users when abnormal line loss occurs in the station area, a method for judging and accurately locating the line loss abnormality in the station area based on the correlation measurement algorithm is proposed. Firstly, the optimal clustering number of the data set is determined by the gap statistics-contour coefficient fusion algorithm, and on this basis, the dichotomous K-means++ is used to construct the station area line loss standard library. Secondly, the station area line loss anomaly identification is completed based on the standard library and then the abnormal time is determined. The Spearman correlation coefficient (SCC) and Euclidean-discrete Fréchet distance (E-DFD) of each user's power and line loss during the abnormal time is calculated. And based on SCC and E-DFD, a comprehensive evaluation index to analyze user relevance is established. Finally, the technique for order preference by similarity to an ideal solution (TOPSIS) is used to sort the comprehensive evaluation index values to achieve precise positioning of abnormally associated users. The calculation example uses real field data in a certain area to analyze, and the results show that the method proposed in this paper has better performance in clustering effectiveness, calculation time, and identification accuracy.

Keywords: line loss in station area; abnormal judgment; precise positioning; line loss standard library; comprehensive evaluation indicators; technique for order preference by similarity to an ideal solution (TOPSIS)

(编辑 方晶)