

DOI: 10.19783/j.cnki.pspc.230835

# 基于多维场景划分的台区线损率异常研判及 关联用户精准追踪方法

陈光宇<sup>1</sup>, 张盛杰<sup>1</sup>, 杨里<sup>2</sup>, 黄文灏<sup>3</sup>, 南钰<sup>4</sup>, 张仰飞<sup>1</sup>, 郝思鹏<sup>1</sup>

(1. 南京工程学院电力工程学院, 江苏 南京 211167; 2. 国网福建省电力有限公司, 福建 福州 361004; 3. 国网福建省电力有限公司三明供电公司, 福建 三明 365000; 4. 国网河南省电力公司开封供电公司, 河南 开封 475000)

**摘要:** 针对台区线损率异常辨识困难, 关联用户难以精准追踪的问题, 提出一种基于多维场景划分的台区线损率异常研判及关联用户精准追踪方法。首先, 提出了一种基于改进完备集合经验的模态分解(improved complete EEMD with adaptive noise, ICEEMDAN)算法的数据预处理策略, 通过分析台区历史线损率曲线的变化趋势, 构建台区线损率粗放场景集, 并在此基础上采用支持向量聚类算法进一步对场景集进行二次精细划分, 实现对台区多维场景集的建立。其次, 提出一种基于簇类个案数目的区间动态平移策略, 用于确定不同场景下线损率标准库的区间范围, 并采用区间重叠率策略对划分中的孤立区间进行合并, 实现对台区线损率标准库的完备划分。最后, 给出一种基于关联分析的线损异常追踪方法, 通过灰色关联分析法确定线损异常强相关因素, 并基于改进Adtributor算法定量分析各因素和台区用户间的内在关联度, 提高对台区线损率异常用户的追踪精度。算例采用台区真实数据进行仿真分析, 结果表明了所提方法的有效性和实用性。

**关键词:** 台区; 灰色关联; 多维场景集; 线损率标准库; 精准追踪

## Station line loss rate anomaly identification and accurate tracking method of associated users based on multi-dimensional scene division

CHEN Guangyu<sup>1</sup>, ZHANG Shengjie<sup>1</sup>, YANG Li<sup>2</sup>, HUANG Wenhao<sup>3</sup>, NAN Yu<sup>4</sup>, ZHANG Yangfei<sup>1</sup>, HAO Sipeng<sup>1</sup>

(1. School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China; 2. State Grid Fujian Electric Power Company, Fuzhou 361004, China; 3. Sanming Power Supply Company, State Grid Fujian Electric Power Co., Ltd., Sanming 365000, China; 4. Kaifeng Power Supply Company, State Grid Henan Electric Power Company, Kaifeng 475000, China)

**Abstract:** There is a difficulty in identifying station line loss rate anomalies and tracking associated users accurately. Thus a method of station line loss rate anomaly research and tracking associated users based on multi-dimensional scene division is proposed. First, a data preprocessing strategy based on the improved complete EEMD with adaptive noise (ICEEMDAN) algorithm is proposed. By analyzing the changing trend of the historical line loss rate curve of the station area, the extensive scene set of the station line loss rate is constructed. On this basis, the SVM clustering algorithm is used to make further quadratic partition to scene set, thus to establish multi-dimensional scene set. Secondly, an interval dynamic translation strategy based on the number of cluster cases is proposed to determine the interval range of the line loss rate standard library in different scenarios. Also an interval overlap rate strategy is used to merge the isolated intervals in the division, so as to realize the complete division of the line loss rate standard library in the station area. Finally, an abnormal line loss tracking method based on correlation analysis is given. This determines the strong correlation factors of abnormal line loss through grey correlation analysis, and quantitatively analyzes the internal correlation degree between each factor and station users based on an improved Adtributor algorithm to improve the tracking accuracy of abnormal line loss rate users. Cases are simulated and analyzed by using real data in station area. The simulation results show that the proposed method is effective and practical.

This work is supported by the National Natural Science Foundation of China (No. 52107098).

**Key words:** station; grey correlation; multi-dimensional scene set; line loss ratio standard library; accurate tracking

基金项目: 国家自然科学基金项目资助(52107098)

## 0 引言

长期以来,随着国家节能减排政策的贯彻实施,台区线损精益化管理逐渐成为当下的主流,节能降损工作也取得了显著成效<sup>[1-4]</sup>。目前,台区线损率异常、用户窃电等问题普遍存在,其排查方法仍主要依靠运维人员的经验进行判断,这不但极大地增加了运维人员的工作强度,而且还会出现部分异常难以发现导致电量损失。因此,如何建立完备的台区线损率异常智能研判体系并且以此为基础快速和准确地找到造成台区线损率异常的关联用户已变得刻不容缓。

目前,国内外研究人员对于线损率的研究主要围绕线损率的计算<sup>[5]</sup>和预测<sup>[6-7]</sup>展开,而对于台区线损率异常的辨识与追踪等问题却较少涉及。文献[8]提出了一种基于“秩和”近似相等特性的同期线损异常数据深度辨识方法,在传统的 $\alpha$ 阈值计算基础上采用改进 $\alpha \cap \beta$ 阈值计算方法,解决了对台区不同类型异常数据辨识困难的问题;文献[9]则针对如何有效辨识线损异常的问题,并结合低压台区的真实数据,提出一种基于二次K-means聚类的线损异常辨识方法,仿真结果表明该方法在很大程度上解决了传统线损异常辨识方法存在的片面性和局限性问题。尽管以上文献对台区线损率异常问题进行了分析并取得了较好的结果,但并没有进一步挖掘线损率与用户之间的内在关联,对造成台区线损率异常的用户缺乏进一步的研究和分析。

此外,部分学者在台区用户侧异常用电方面也展开了深入研究,并取得不少成果。文献[10]从考虑降低电力公司运营成本的角度出发,提出基于网络的局部离群因子算法的异常用电模式检测模型,通过检测异常度排序靠前的部分用户,完成异常用户的追踪,研究结果表明该方法能提高排查效率;与文献[10]不同的是,文献[11]从深度学习的角度出发,讨论了在用户异常用电模式检测领域深度学习算法如何更加有效实现的问题,采用长短时记忆网络(long short-term memory, LSTM)模型进行序列特征提取,并结合多层全连接网络实现对异常用电数据的检测;文献[12]则基于用户行为的关联度,提出了一种异常用户间相关性的概率模型,用一阶逻辑语法推导出了行为相关性发生的概率,实验表明该方法与传统异常检测算法相比精度优势明显。与文献[8-9]的侧重点不同,以上文献大多从用户侧的角度出发,重点研究了台区内用户用电模式的异常问题,但并没有涉及用户侧电量异常和台区线损率异常间的内在关联,有造成异常误判的风险。

文献[13]针对窃电用户的线损特点,提出了采

用欧式(euclidean distance, ED)、余弦和街区距离计算用户负荷曲线与异常馈线线损曲线的相似性,进而迅速缩小窃电用户的排查范围;文献[14]则从线损与用户电量关系的角度出发,采用皮尔逊相关系数(Pearson correlation coefficient, PCCs)算法求解其相关程度,实现了对异常电能表用户的准确追踪。以上方法分别从形状<sup>[13]</sup>和相关性<sup>[14]</sup>两个层面展开论述,探讨了线损率和用户负荷之间的某种函数关系,并取得了较好的效果,但针对部分台区负荷类型较为复杂的情况,仅从形状或相似性来追踪异常用户仍存在一定的片面性。

随着电网的日益复杂以及对台区线损率精细化管理要求的提高,一方面,由于配电台区众多,而不同台区下用户类型及分布情况各有差异,仅依靠传统“一刀切”式的台区线损率标准难以实现对台区线损率异常的精准判别;另一方面,对于导致台区线损率异常的关联用户的快速查找和追踪方法仍需改进,对于一些情况较为复杂的台区用户,气候、节假日等因素都会对其线损率产生影响。因此当前的追踪方法仍需要改进,其判断线损率的标准过于笼统,缺乏特殊性,需进一步研究和探讨。因此,针对上述问题,本文提出一种数据驱动下的台区线损率异常智能研判及关联用户精准追踪方法。首先,提出一种基于改进完备集合经验的模态分解(improved complete EEMD with adaptive noise, ICEEMDAN)策略,分析台区历史线损率曲线的大致变化趋势,并构建台区线损率的粗放场景集,在此基础上采用支持向量聚类(support vector clustering, SVC)算法进一步对场景集进行二次精细划分,进而完成台区多维场景集的建立;其次,给出一种基于簇类个案数目的区间动态平移策略,用于确定不同场景下线损率标准库的区间范围,采用区间重叠率策略对划分中的孤立区间进行合并,实现对台区线损率标准库的完备划分,并基于标准库完成对台区线损率异常的自动判别;最后,采用一种改进 Adtributor 算法定量分析线损率和用户间的内在关联度,实现对台区线损率异常关联用户的精准追踪。算例以某配电网公用变台区的真实用户电量和线损率数据进行仿真分析,结果表明,本文构建的标准库能够较好地给出台区多维场景下的线损率标准,实现对台区异常线损率的智能研判;而提出的异常追踪策略相比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 台区线损率异常智能研判框架

台区线损率异常智能研判总体框架如图 1 所

示, 包含 4 个部分, 具体如下。

第 1 部分: 基于历史数据挖掘的台区线损率多维场景划分策略。

首先, 针对原始采集的线损率曲线, 采用 ICEEMDAN 算法进行初步预处理, 去除曲线中的“毛刺”或“噪声”数据, 完成数据预处理; 其次, 挖掘线损率曲线整体变化趋势, 采用 K-means++ 方法对预处理后的线损率进行聚类分析, 初步构建较为粗放的场景; 最后, 采用 SVC 算法进行二次聚类, 实现对台区多维场景的精准划分。

第 2 部分: 基于台区多维场景的划分结果完成台区线损率标准库的构建。

首先, 提取各场景下的线损率数据并进行聚类分析, 初步生成台区线损率标准库; 其次, 提出一种基于簇类个案数目的区间动态平移法, 修正标准库区间的划分结果, 进一步获取标准库精确上下限; 最后, 给出一种区间重叠率算法对重叠率较高的区间进行合并处理, 实现多维场景下台区线损率标准库的构建。

第 3 部分: 基于台区线损率标准库实现台区线损率异常的智能研判。

以标准库为基础, 通过对比待研究的台区线损

率形态和标准形态, 分析线损率异常时间和线损率异常形态, 实现异常自动判定。

第 4 部分: 针对线损率异常台区, 实现台区异常关联用户的精准追踪。

首先, 研究异常台区线损率曲线的形态特征, 提取异常线损率及对应的台区用户; 其次, 引入灰色关联分析法和改进 Adtributor 算法, 求解并挖掘出最具有嫌疑的用户; 最后, 锁定异常关联用户, 实现精准追踪。

## 2 基于历史数据挖掘的台区线损率多维场景划分策略

目前台区线损率的异常判断大多仍依靠企业制定的统一标准或运维人员自身经验进行主观判断, 但随着配电网规模日趋庞大, 台区线损及负荷种类情况日益复杂, 传统的线损率标准和人工经验已难以满足电网发展的需求。

因此, 为了能准确把握台区线损率的真实情况, 首先需要对台区线损率海量历史数据进行挖掘, 分析出台区线损率各类典型场景, 并以此为基础构建多维场景下的台区线损率标准形态。图 2 给出了多维场景划分策略的主要流程。

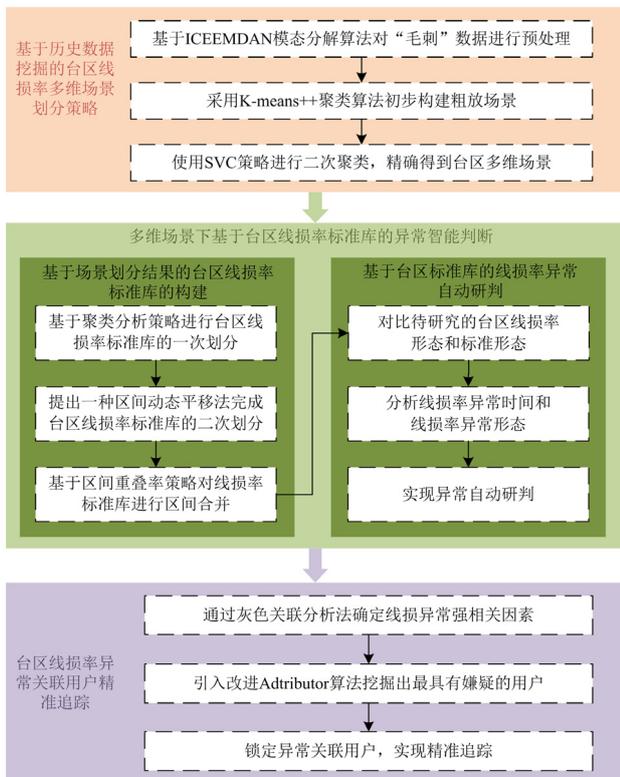


图 1 台区线损率异常智能研判框架

Fig. 1 Framework of intelligent research and judgment of abnormal line loss rate of substation area

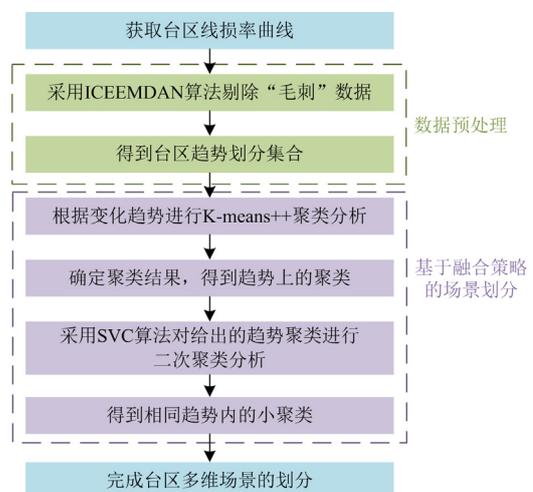


图 2 多维场景划分策略架构图

Fig. 2 Architecture of multi-dimensional scene partitioning strategy

### 2.1 数据预处理策略

数据的噪声会影响线损率曲线的态势分析, 因此考虑对原始线损率数据进行降噪处理。传统的处理方法主要有  $3-\sigma$  去噪法<sup>[15]</sup>、小波阈值降噪<sup>[16]</sup>、经验模态分解(empirical mode decomposition, EMD)<sup>[17]</sup>等, 这些方法在一定程度上能消除部分干扰, 但对分解过程中引起的残余噪声模态和模态混叠等问题

仍未能较好地解决。为此, 本文给出一种基于 ICEEMDAN 算法的台区线损率曲线预处理策略, 用于剔除线损率曲线中的干扰数据, 步骤如下:

首先, 提取台区原始线损率数据, 得到相应曲线, 采用 ICEEMDAN 算法分解生成  $k$  个本征模函数(intrinsic mode function, IMF);

其次, 针对分解得到的 IMF 分量, 引入归一化自相关函数(normalized autocorrelation function, NAF)作为信号相关性评价指标, 计算与原始台区线损率曲线的关联系数;

最后, 设置阈值分析, 筛选出高相关的 IMF 分量集合, 并叠加集合中的所有分量, 进而完成重构过程, 最终得到预处理后的台区线损率曲线。

总体流程如图 3 所示。

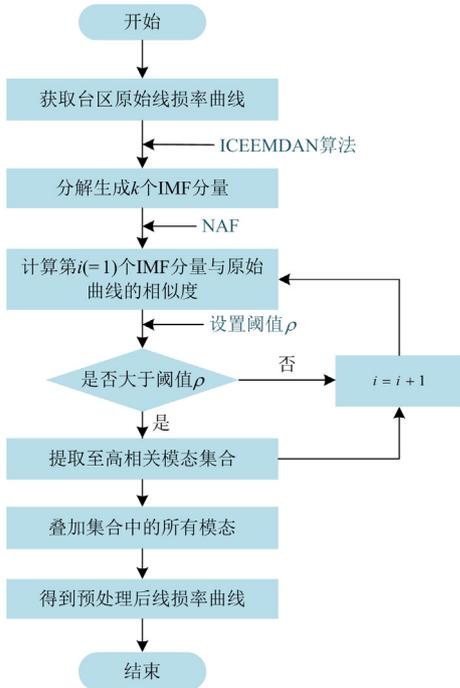


图 3 基于 ICEEMDAN 的数据预处理策略

Fig. 3 Data preprocessing strategy based on ICEEMDAN

### 2.1.1 ICEEMDAN 算法简介

ICEEMDAN 是 Colominas 等人提出的一种改进降噪算法<sup>[18]</sup>, 克服了传统 EMD 方法直接舍弃前几阶 IMF 分量的弊端, 能有效削弱 IMF 分量中的残余噪声, 改善传统降噪算法中伪模态和模态混叠问题, 不易使重构信号失真, 能够更好地对初始数据进行平滑处理, 为后续的研究打下坚实的基础。

假设存在一信号  $X(t)$ , 构造一个包含特殊白噪声  $E_1[\omega^{(i)}(t)]$  的加噪信号, 如式(1)所示。

$$X^{(i)}(t) = X(t) + \alpha_0 E_1[\omega^{(i)}(t)] \quad (1)$$

式中:  $X^{(i)}(t)$  为本文构造的加噪信号;  $\alpha_0$  为添加特

殊白噪声的标准差;  $i$  为填入次数;  $\omega^{(i)}(t)$  为添加的高斯白噪声。

$X^{(i)}(t)$  经 ICEEMDAN 生成 IMF 分量的具体流程见表 1。

表 1  $X^{(i)}(t)$  经 ICEEMDAN 生成 IMF 分量流程

Table 1 Progress of  $X^{(i)}(t)$  generating IMF component through ICEEMDAN

$X^{(i)}(t)$ 经 ICEEMDAN 生成 IMF 分量流程:	
Step 1	计算加噪信号 $X^{(i)}(t)$ 通过 EMD 生成的第一个残余分量。
	$r_1(t) = \langle M(X^{(i)}(t)) \rangle$
Step 2	求解第一阶 IMF 分量 $F_1(t)$ 。
	$F_1(t) = X(t) - r_1(t)$
Step 3	进行类比迭代, 计算第 $k$ 个残余分量 $r_k(t)$ 。
	$r_k(t) = \langle M(r_{k-1}(t) + \alpha_{k-1} E_k[\omega^{(i)}(t)]) \rangle$
	其中, $\alpha_{k-1}$ 为第 $k$ 次 EMD 添加特殊白噪声的标准差。
Step 4	求解第 $k$ 阶 IMF 分量 $F_k(t)$ 。
	$F_k(t) = r_{k-1}(t) - r_k(t)$

### 2.1.2 IMF 的选取原则

针对 ICEEMDAN 的 IMF 分量, 分析原信号  $X^{(i)}(t)$  与各阶 IMF 分量的相关性, 提取与  $X^{(i)}(t)$  相关程度较大的部分, 进而完成相关模态的重构, 提高特征提取的准确率。本文选用 NAF 作为信号相关性评价指标来筛选出包含原始信号特征的 IMF 分量。

NAF 反映了同一信号序列在不同时刻数值之间相关性的统计特性, 具体计算方法如式(2)<sup>[19]</sup>所示。

$$\rho(t_1, t_2) = \frac{\sum_{i=0}^k \sum_{t_1, t_2 \in N} R_{\text{IMF}_i}(t_1, t_2) R(t_1, t_2)}{\sqrt{\sum_{i=0}^k \sum_{t_1, t_2 \in N} R_{\text{IMF}_i}^2(t_1, t_2) \sum_{t_1, t_2 \in N} R^2(t_1, t_2)}} \quad (2)$$

$$R(t_1, t_2) = \frac{1}{N} \sum_{t_1, t_2 \in N} X^{(i)}(t_1) X^{(i)}(t_2)$$

式中:  $\rho(t_1, t_2)$  表示原信号  $X^{(i)}(t)$  与各阶 IMF 分量的相关系数;  $R_{\text{IMF}_i}(t_1, t_2)$  表示原信号  $X^{(i)}(t)$  的第  $i$  阶 IMF 分量;  $k$  表示 IMF 阶数;  $t_1$  和  $t_2$  表示随机时刻;  $R(t_1, t_2)$  表示原信号  $X^{(i)}(t)$  和各阶 IMF 分量的自相关函数;  $N$  表示信号集合。

### 2.2 基于融合策略的场景划分方法

采用传统的聚类策略进行划分时常会出现相近或交叠的数据无法划分的问题, 图 4(a)给出了采用 K-means++ 聚类<sup>[20]</sup>策略对某一线损率曲线聚类后的

结果。从图 4(a)中可以看出，每一维趋势聚类集合中的线损率变化幅度很小，数值相差不大，且整体呈“半月”型分布，导致聚类效果明显不佳，甚至存在聚类错误的情况。为了解决聚类精度不足的问题，本节进一步对给出的趋势聚类进行二次聚类分析，考虑到 SVC 算法可产生任意形状的簇边界<sup>[21]</sup>，能较为理想地完成相近或交叠数据的聚类，因此本节通过将 K-means++和 SVC 算法相融合实现台区场景的精细划分，具体步骤如下所述。

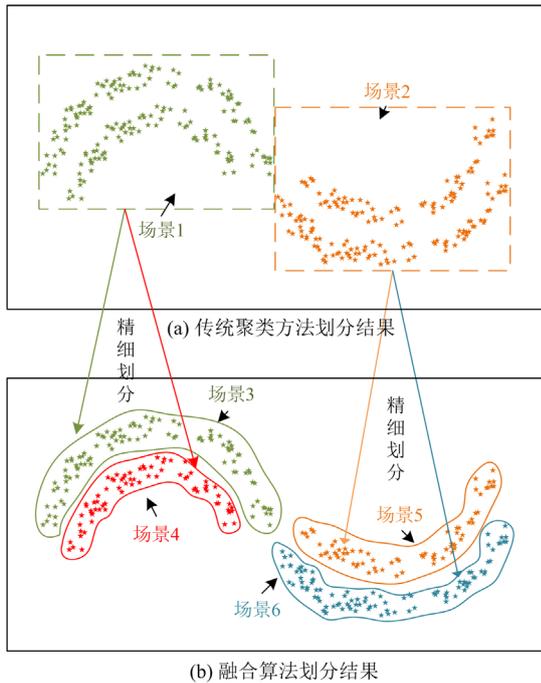


图 4 不同聚类策略的场景划分结果示意图  
Fig. 4 Schematic diagram of scene division results of different clustering strategies

步骤 1: 获取台区线损率曲线，采用 2.1 节所提的 ICEEMDAN 算法剔除曲线中的“毛刺”数据，得到台区趋势划分集合。

步骤 2: 根据变化趋势进行 K-means++聚类分析，确定历史数据的聚类结果，得到趋势上的聚类，如图 4(a)所示(场景 1 和场景 2)，图中横轴表示时间月份，纵轴表示线损率。

步骤 3: 在步骤 2 的基础上，采用 SVC 算法继续对给出的趋势聚类结果进行二次聚类，如图 5 所示，在二次聚类时，参与聚类的属性还包含台区售电量、供电量、容量等，同时还会考虑季节、特殊节假日等因素；将数据归一化处理，再结合一次聚类所划分出的场景，进行更加精细的聚类和场景划分，如图 4(b)所示(场景 3 至场景 6)。

步骤 4: 输出台区场景划分结果。

从图 4(a)和图 4(b)的比较中可以看出，通过传统聚类策略只能区分出两种场景，而融合算法策略可以划分出 4 种场景，结果更加精细。

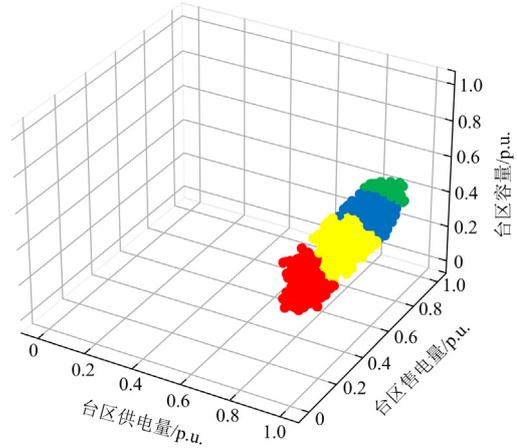


图 5 多属性台区二次聚类示意图  
Fig. 5 Schematic diagram of secondary clustering of multi-attribute substation area

### 3 基于台区线损率标准库的线损率异常研判策略

传统台区线损率划分标准采用简单的“一刀切”模式，但该方法忽略了不同台区内各类负荷特性的差异，难以适应日益复杂的台区线损率异常判别。为了能精准地辨识异常线损率，本节在台区多维场景划分的基础上，建立多维场景下的线损率标准库，进而实现台区线损率异常的智能研判，具体流程如图 6 所示。

#### 3.1 基于聚类分析的线损率标准库一次划分

根据场景划分结果将相同场景下的线损率存放在同一集合  $W_i$  中，研究不同场景下线损率的分布情况。传统划分方法将集合  $W_i$  中的线损率数据进行排序，取最小值作为区间下限，最大值作为区间上限，如图 7 中橙色线段限制的区域。

传统方法考虑了全部的集合数据，忽略了不同簇类中离群点的取舍问题，进而导致划分精度不足。因此，本节给出一种基于聚类分析的线损率标准库一次划分策略，具体步骤如下所述。

步骤 1: 针对集合  $W_i$  中的线损率数据进行挖掘分析，动态调整聚类数目，分别计算每一个聚类数对应的轮廓系数(silhouette coefficient, SC)<sup>[22]</sup>，进而确定最佳聚类数目  $K$ 。

步骤 2: 基于步骤 1 的计算结果，采用 K-means++进行聚类分析，并提取个案数目最多的簇类  $C_i$ (图 7 中的蓝色点群)。

步骤 3: 选取  $C_i$  中元素的最小值作为区间下限, 最大值作为区间上限, 最终得到一次划分后的线损率标准库区间(图 7 中蓝色线段限制的区域)。

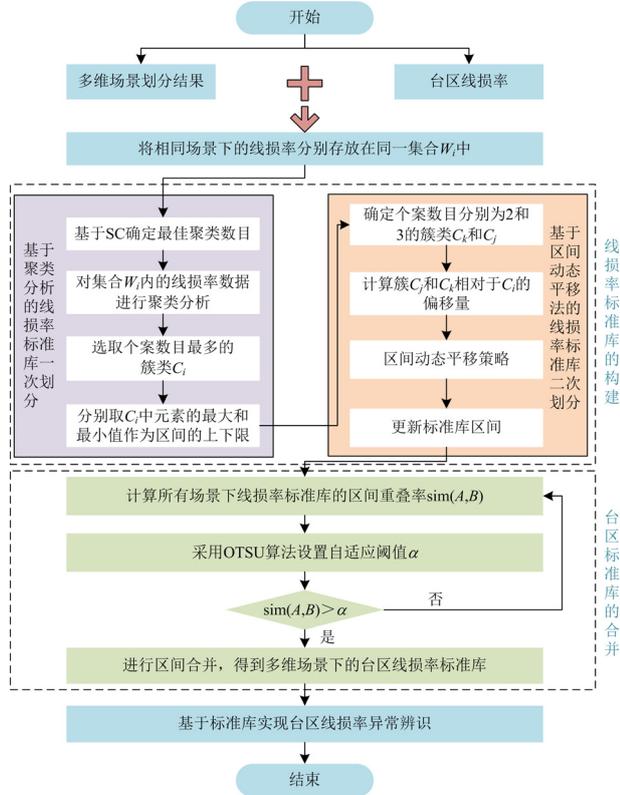


图 6 基于台区线损率标准库的异常研判流程图

Fig. 6 Flow chart of anomaly detection based on line loss rate standard library of the substation area

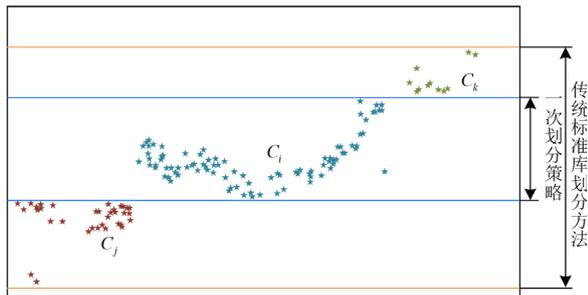


图 7 不同划分策略下的标准库区间示意图

Fig. 7 Schematic diagram of standard reservoir areas under different partitioning strategies

### 3.2 基于区间动态平移法的线损率标准库二次划分

3.1 节给出了一次划分策略, 但该策略主要依据簇类中的个案数目进行区间划分, 且仅考虑了个案数据最多的簇类, 忽略了个案数据较少的簇类样本, 难以保证划分精度。为解决上述问题, 本节提出一种基于簇类个案数目的区间动态平移法, 使一次划

分的标准库区间尽可能覆盖全部线损率数据, 且满足一定的准确度需求。

在采用区间动态平移方法计算时, 需分析个案数目最多的 3 个簇类, 为了便于说明, 规定簇类的空间分布顺序从上至下依次为  $\{C_k, C_i, C_j\}$ 。区间动态平移法的原理步骤如下所述。

步骤 1: 研究多维场景下线损率数据经 K-means++ 聚类生成的  $K$  个簇, 保留个案数目最多的簇  $C_i$ 。

步骤 2: 假定簇类  $C_k$  的个案数目小于簇类  $C_j$ 。选取个案数目分别为 2、3 的簇类  $C_k$  和  $C_j$ , 计算相对于  $C_i$  的偏移量, 计算如式(3)和式(4)所示。

$$\Delta C_j = L_{C_j} \frac{m_j}{n} \quad (3)$$

$$\Delta C_k = L_{C_k} \frac{m_k}{n} \quad (4)$$

式中:  $\Delta C_j$  和  $\Delta C_k$  分别表示簇类  $C_j$  和  $C_k$  相对于簇类  $C_i$  的偏移量;  $L_{C_j}$  和  $L_{C_k}$  分别为簇类  $C_j$  和  $C_k$  的区间长度;  $m_j$  和  $m_k$  分别为簇类  $C_j$  和  $C_k$  的个案数目;  $n$  为总样本容量。

步骤 3: 结合步骤 2 中的相对偏移量结果, 求解并更新标准库区间, 如式(5)所示。

$$[y_{\text{lower}}, y_{\text{upper}}] = [C_{i\text{min}} - \Delta C_j, C_{i\text{max}} + \Delta C_k] \quad (5)$$

式中:  $y_{\text{lower}}$  和  $y_{\text{upper}}$  分别表示二次划分后标准库区间的下限和上限;  $C_{i\text{min}}$  和  $C_{i\text{max}}$  分别表示簇  $C_i$  内元素最小值和最大值。

### 3.3 基于区间重叠率的台区标准库合并策略

由于部分台区场景划分维度较多, 难免出现场景重叠的现象。重叠现象是指两个或多个台区线损率标准库区间出现大量交集, 大量的重叠致使各标准库区间区分度不高, 增加系统研判时间, 影响工作效率。图 8 给出了不同场景间重叠的示意图, 其中,  $X$  轴代表场景,  $Y$  轴为线损率标准区间, 由图 8 即可看出各场景的区间重叠率。场景 1 和场景 3 的区间重叠率很高, 所以将两个场景进行合并。

从图 8 中可以看出, 场景 1 和场景 3 的重叠区间较大, 因此, 为了适当减少台区线损率的标准库数量, 提高运维人员的工作效率, 本节给出一种基于区间重叠率的线损率标准库的区间合并策略, 进一步对重叠场景进行缩减。构造两场景重叠率定义: 设  $A$ 、 $B$  是两个确定区间, 则  $A$ 、 $B$  的区间重叠率表示为

$$\text{sim}(A, B) = \frac{L(A \cap B)}{L(A) + L(B) - L(A \cap B)} \quad (6)$$

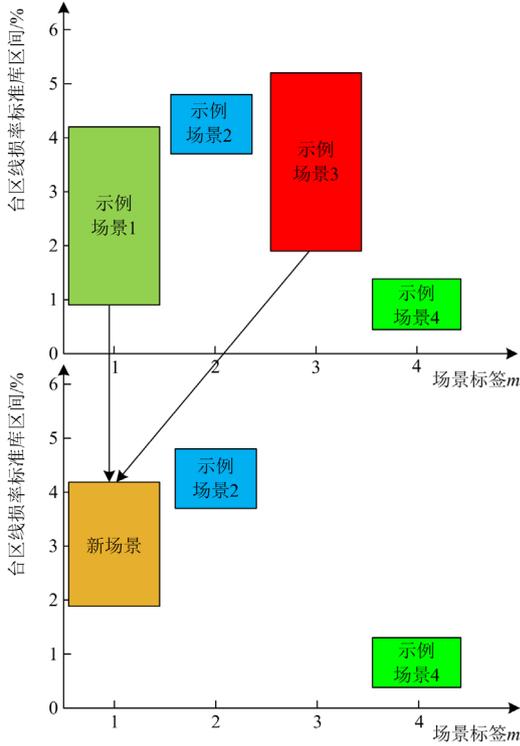


图 8 多维场景下部分线损率标准库的重叠示意图

Fig. 8 Overlapping schematic of partial line loss standard library in multi-dimensional scenarios

式中： $L$  表示相应区间的长度； $A \cap B$  表示  $A$ 、 $B$  的重叠区间。其中， $\text{sim}(A, B) \in [0, 1]$ 。

为保证标准库足够精简，本节采用 OTSU 算法<sup>[23]</sup>对区间重叠率设置一个自适应阈值  $\alpha$ ，通过  $\alpha$  分割求取的类间方差作为评价指标，具体如式(7)所示。

$$\delta^2 = w_1 w_2 (u_1 - u_2)^2 \quad (7)$$

式中： $\delta^2$  为类间方差； $w_1$  和  $w_2$  分别表示  $[0, \alpha)$  和  $[\alpha, 1]$  两部分所占的比例； $u_1$  和  $u_2$  分别为每一部分重叠率的平均值。

求出  $\delta^2$  最大时对应的阈值作为最佳阈值  $\alpha_{\text{best}}$ ，并将重叠率大于等于  $\alpha_{\text{best}}$  的标准库区间按交集原则进行合并，反之则不合并。

### 3.4 基于台区标准库的线损率异常自动研判策略

3.3 节给出了多维场景下台区线损率标准库的构建方法，本节将基于多维标准库对台区给定时间段内的线损率异常情况进行自动研判，具体流程如图 9 所示。

步骤 1：导入待研究的台区线损率序列  $\{P_i\}$ ；

步骤 2：获取序列  $\{P_i\}$  中的台区编号和时间日期，根据场景划分结果确定其对应的场景，并提取

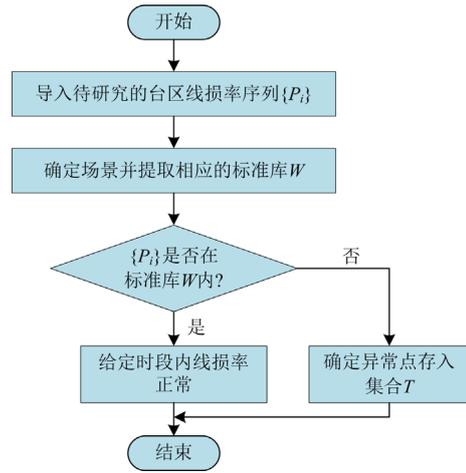


图 9 基于台区线损率标准库的异常自动研判流程图

Fig. 9 Automatic anomaly detection flow chart based on line loss rate standard library of the substation area

相应的标准库  $W$ ；

步骤 3：判断集合  $\{P_i\}$  是否在标准库  $W$  内，若在转入步骤 5，否则转入步骤 4；

步骤 4：给定时段内台区线损率异常，并将异常点存入集合  $T$ ；

步骤 5：给定时段内台区线损率正常，结束。

## 4 台区线损率异常关联用户精准追踪方法

以上分析了基于标准库的线损率异常辨识方法，但台区线损率的异常通常是由多种因素造成，传统基于距离<sup>[13]</sup>或相关系数法<sup>[14]</sup>的异常追踪策略在复杂场景下的辨识效果欠佳，在台区内用户较多的情况下，要想准确辨识出导致台区线损率异常的用户难度较大，在计算时间和准确度方面表现不稳定，部分场景下计算结果存在一定的误差，且得出的追踪结果所考虑的因素较为单一。针对以上问题，本文考虑引入灰色关联分析法和改进 Adtributor 算法，从多个角度对异常用户进行追踪。

### 4.1 灰色关联分析法

灰色关联分析法是灰色系统中的一个重要组成部分，其基本思想是根据数据曲线几何形状的相似程度来判断不同数据标签之间的联系是否紧密，如果两个因素的相关性越强，则灰色关联度越大，反之则越小。计算灰色关联度的关联系数和关联度表达式<sup>[24]</sup>为

$$\gamma(x_{0z}, x_{iz}) = \frac{\min_i \min_z |x_{0z} - x_{iz}| + \rho \max_i \max_z |x_{0z} - x_{iz}|}{|x_{0z} - x_{iz}| + \rho \max_i \max_z |x_{0z} - x_{iz}|} \quad (8)$$

$$\gamma_{0i} = \frac{1}{n} \sum_{z=1}^n \gamma(x_{0z}, x_{iz}) \quad (9)$$

式中:  $\gamma(x_{0z}, x_{iz})$  为第  $i$  个元素  $z$  时刻的关联系数;  $x_{0z}$  和  $x_{iz}$  为归一化处理后的值;  $x_{0z}$  为参考序列  $z$  时刻的数值;  $x_{iz}$  为第  $i$  个因素  $z$  时刻的数值;  $\rho$  为分辨系数, 一般取 0.5;  $\gamma_{0i}$  为第  $i$  个元素的关联度;  $n$  表示时刻总长度。

#### 4.2 Adtributor 算法

针对异常线损率, Adtributor 算法<sup>[25]</sup>可用于求解并挖掘出最具有嫌疑的用户。采用算法分析时, 会涉及到意外性值(surprise, S)和解释能力值(explanatory power, EP)两个参量的计算。

S 值的计算公式如式(10)所示。

$$S_{ij}(m) = 0.5 \left[ p \log\left(\frac{2p}{p+q}\right) + q \log\left(\frac{2q}{p+q}\right) \right] \quad (10)$$

式中:  $S_{ij}(m)$  表示当前序列  $i$  和参考序列  $j$  之间的意外性值;  $p$  表示预测概率;  $q$  表示真实概率;  $m$  表示异常指标。

$p$  和  $q$  的作用是衡量实际观测到的概率分布与预测的概率分布之间的差异程度, 通过比较  $p$  和  $q$  的差异性, 有助于判断台区用户线损是否异常。

其中, 预测概率  $p$  的计算公式如式(11)所示。

$$p_{ij}(m) = \frac{F_{ij}(m)}{F(m)} \quad (11)$$

真实概率  $q$  的计算公式如式(12)所示。

$$q_{ij}(m) = \frac{A_{ij}(m)}{A(m)} \quad (12)$$

EP 值的计算公式如式(13)所示。

$$E_{ij}(m) = \frac{A_{ij} - F_{ij}}{A(m) - F(m)} \quad (13)$$

式中:  $p_{ij}(m)$  表示当前序列  $i$  和参考序列  $j$  之间的预测概率;  $q_{ij}(m)$  表示当前序列  $i$  和参考序列  $j$  之间的真实概率;  $E_{ij}(m)$  表示当前序列  $i$  和参考序列  $j$  之间的解释能力值;  $A_{ij}$  表示故障真实值;  $F_{ij}$  表示正常预测值;  $A(m)$ 、 $F(m)$  分别为  $A_{ij}$ 、 $F_{ij}$  的总和。

#### 4.3 Adtributor 算法的改进

实际电网中, 由于用户负荷波动引起的小扰动现象较为常见, 而 Adtributor 算法对扰动因素的敏感度较强, 同时, 算法采用的 ARMA 时间序列模型<sup>[26]</sup>对非线性数据的处理能力较弱, 如果继续采用传统的 Adtributor 算法将导致式(13)中  $F_{ij}$  的求解效率不高, 预测精度难以保证, 严重影响异常用户追

踪的效率和准确度。

为解决传统算法存在的问题, 本文对 Adtributor 算法中的预测模型进行改进。通过采用 Lagrange 插值法<sup>[27]</sup>求取异常点的预估值, 在一定程度上能实现非线性函数的拟合, 并大幅减少计算量。图 10 给出了基于 Lagrange 插值的线损率异常预估示意图。

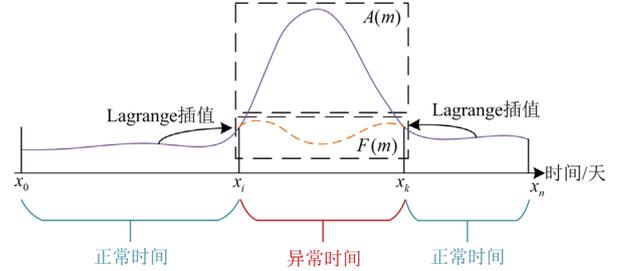


图 10 基于 Lagrange 插值的线损率异常预估示意图

Fig. 10 Schematic diagram of line loss rate anomaly prediction based on Lagrange interpolation

具体步骤如下所述。

步骤 1: 基于标准库甄别出线损率异常值(图中  $A(m)$  区域);

步骤 2: 提取正常时间段内的线损率数据, 采用 Lagrange 插值法求解异常点的预估值(图中  $F(m)$  区域);

设线损率曲线在区间  $[x_0, x_i] \cup [x_k, x_n]$  上的值  $f(x_1), \dots, f(x_n)$  已知, 则  $[x_{i+1}, x_k]$  中某一位置的预估值  $f(x_j)$  应满足式(14)。

$$P_n(x) = \sum_{s=0, s \neq [i, k]}^n [f(x_s) \prod_{j=0, j \neq s}^n \frac{x - x_j}{x_s - x_j}] \quad (14)$$

式中:  $P_n(x)$  表示插值多项式;  $s$  的取值范围代表正常时间段。

步骤 3: 结合步骤 1 和 2 中得到的  $A(m)$  和  $F(m)$  值, 根据式(13)计算改进后的 EP 值。

#### 4.4 线损率异常关联用户精准追踪

结合标准库分析台区线损率异常的时间段及所有相应用户的曲线形态, 比较台区线损率异常和用户电量异常间的差异, 研究各自异常曲线形态特征, 对不同异常形态间的内在关联进行识别和计算, 具体流程如图 11 所示。

步骤 1: 基于第 3 节生成的标准库进行台区异常筛选, 提取线损率异常值, 确定异常时间段。

步骤 2: 计算台区用户强相关因素 S 值, 判定是否大于阈值 0<sup>[25]</sup>。若 S 大于 0, 则构建可疑用户集, 同时将满足条件的用户存入集合中; 反之, 判为正常。

步骤 3: 计算可疑用户集中的电量 EP 值, 判定是否大于阈值 0<sup>[25]</sup>。若 EP 大于 0, 则保留在可疑用户集中; 反之, 判为正常。

步骤 4: 计算可疑用户(即 EP > 0)对应的 S 值总和, 通过降序排列, 追踪可疑用户。

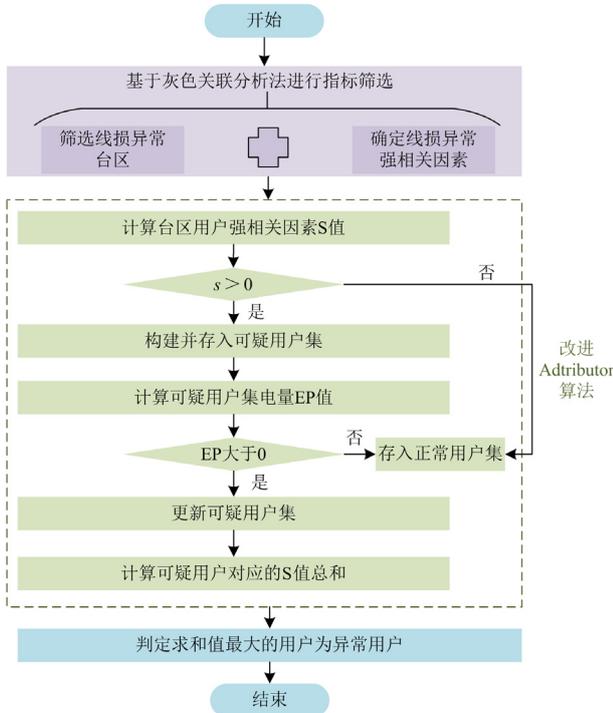


图 11 台区线损率异常关联用户精准追踪流程图

Fig. 11 Flow chart of accurate positioning associated with abnormal line loss rate user of the substation area

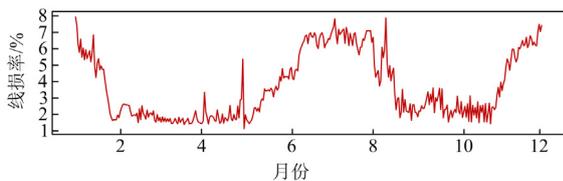
### 5 算例分析

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 630 kVA, 时间跨度为 1 年(2021 年 1 月 1 日至 12 月 31 日), 采样频率为 1 天, 原始线损率数据共 365 条, 负荷数据共 1000 条。

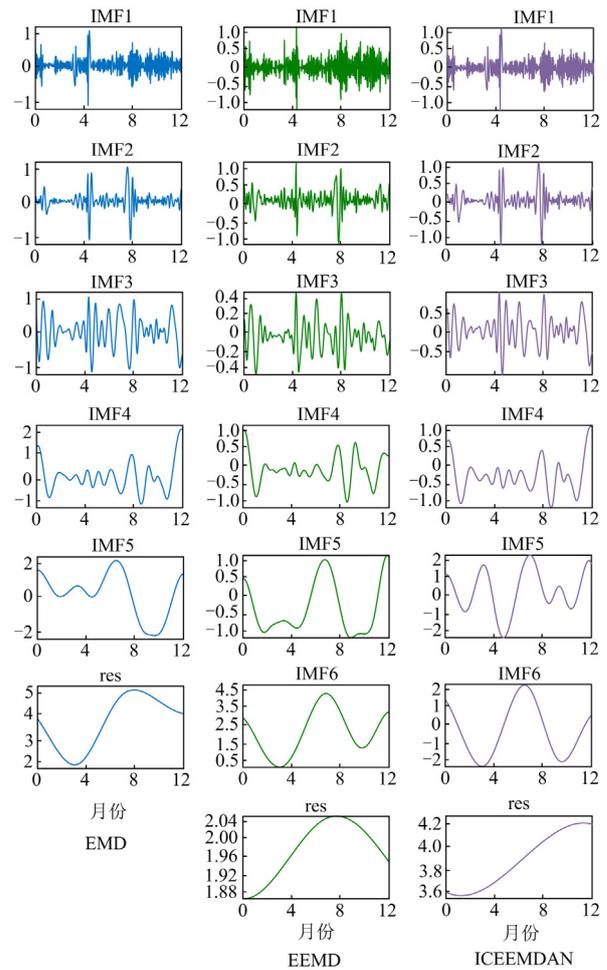
#### 5.1 多维场景的构建分析

为验证本文给出的 ICEEMDAN 算法的优越性, 分别采用 EMD<sup>[28]</sup>和 EEMD<sup>[29]</sup>进行对比分析, 如图 12 所示。

图 12(a)和图 12(b)分别给出了某台区一年的真实线损率曲线以及不同算法下线损率曲线的模态分



(a) 某台区一年的真实线损率曲线



(b) 不同算法下线损率曲线的模态分解效果图

图 12 EMD、EEMD 和 ICEEMDAN 算法对原始线损率曲线模态分解效果对比图

Fig. 12 Comparison of modal decomposition effect of EMD, EEMD and ICEEMDAN algorithms on original line loss rate curve

解效果图。从图 12(b)可以看出, EMD 和 EEMD 产生的模态混叠现象依然存在(如 IMF2 和 IMF3 分量)。ICEEMDAN 算法尽管局部极值依然存在短时间内的跳变, 但变化频率较低且波动幅值不大, 较好地实现了对模态混叠的抑制。

为了实现模态的重构, 需筛选出内涵模态分量 IMF 中的有效分量。以 ICEEMDAN 算法为例, 表 2 给出了原始线损率与各阶 IMF 分量的 NAF 计算结果。当 NAF 大于 0.3 时两者具有强相关性<sup>[30]</sup>。从表 2 中可以看出, IMF5 和 IMF6 的 NAF 均大于阈值 0.3, 因此保留这两阶 IMF 分量进行模态重构。

图 13 给出了 EMD、EEMD 及 ICEEMDAN 3 种算法进行分解重构后的线损率曲线。

表 2 各阶内涵模态分量的 NAF 计算结果

Table 2 The original line loss rate and the NAF calculation results of each IMF component

IMF	NAF	IMF	NAF
1	0.023	4	0.236
2	0.035	5	0.571
3	0.187	6	0.609

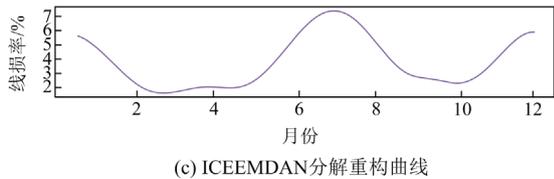
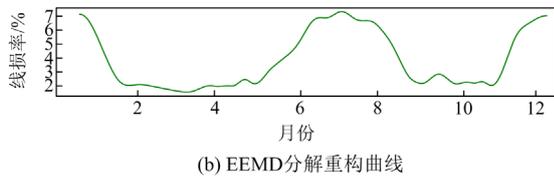
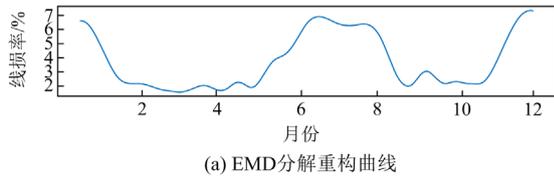


图 13 不同模态算法分解重构的线损率曲线对比

Fig. 13 Comparison of line loss curves for decomposition and reconstruction of different modal algorithms

从图 13 中可以看出, EMD 和 EEMD 重构曲线中仍保留着部分噪声, 而经 ICEEMDAN 算法重构的线损率曲线不仅保持了原有曲线的基本变化趋势, 而且整体平滑, 无明显的“毛刺”, 噪声滤除效果显著。

对图 13(c)采用 K-means++ 聚类分析, 可得到台区的粗放场景集, 如图 14(a)所示, 将台区划分为 A、B、C 三个场景。由于 K-means++ 仅考虑了趋势上的聚类, 处理后的曲线仍无法完整反映场景内部线损率数据的分布情况。因此, 进一步对粗放场景集内的线损率进行二次划分。

以场景 A 中 1 月 21 日至 4 月 30 日(连续时段)内线损率数据为二次划分对象, 图 14(b)给出了二次划分结果示意图, 从图 14(b)中可以看出, 数据集被划分为 3 个不同的聚类场景, 分别为场景 1: 1 月 21 日—1 月 26 日(绿色区域); 场景 2: 1 月 27 日—3 月 30 日及 4 月 20 日—4 月 30 日(蓝色区域); 场景 3: 3 月 31 日—4 月 19 日(红色区域)。通过对比日历信息, 场景 1 对应的时段为特殊节假日(春节),

而场景 3 也包含了特殊节假日(清明), 说明二次划分能考虑节假日场景等因素的影响。

同理, 场景 B 和 C 也可进一步精细划分, 此处不再赘述。

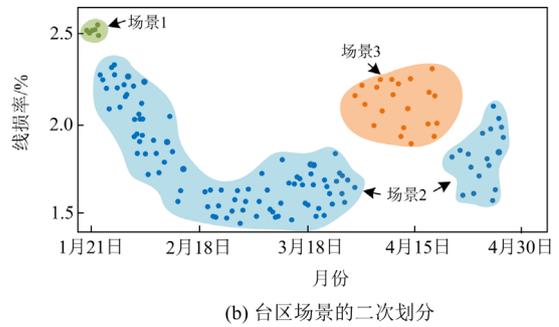
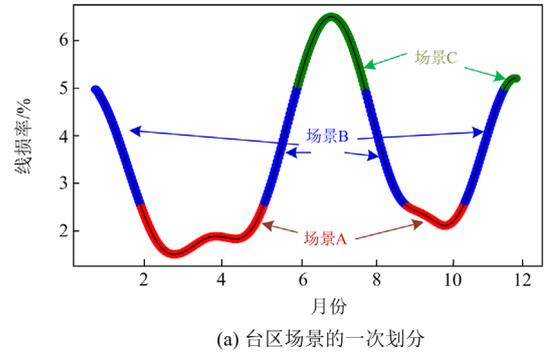


图 14 台区场景划分效果图

Fig. 14 Effect of scene division of the substation area

图 15 刻画了台区多维场景的划分结果。由图 15 可以明显看出, 台区共被划分为了 8 个不同的场景, 其中蓝色区域面积占比最大, 涵盖了场景 B 中所有的日期, 而场景 C 被细分为了 2 个小场景, 分别为场景 6 和 7; 面积较小的几块区域也都具有各自的独特性, 例如浅红区域(9 月 19 日—10 月 4 日)反映的是特殊节假日(国庆)。

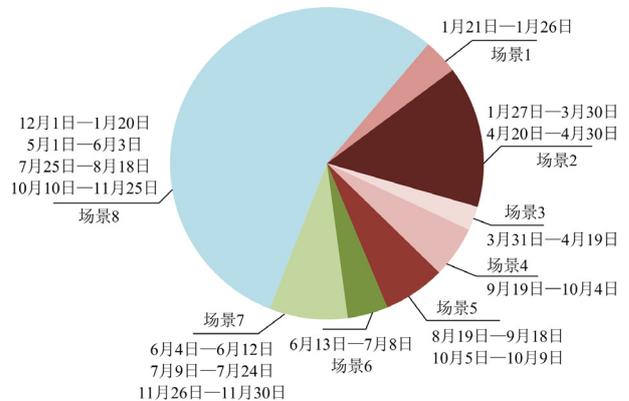


图 15 多维场景划分结果

Fig. 15 Multi-dimensional scene partitioning results

### 5.2 台区线损率标准库构建分析

以上给出了台区多维场景下的划分结果，在此基础上，本节以图 14(b)中的场景 2 为例，分析多维场景下台区线损率标准库的构建过程。

首先，基于 SC 确定最佳聚类数  $K$ ，图 16 为聚类数在[2,5]时 SC 的对比图，每块颜色区域代表数据集中每个样本本身的 SC，区域中白色柱状条表示所有样本 SC 的均值，红线为进行大小比较的基准线。从图 16 可以看出，当聚类数目为 4 时，每个样本的 SC 总和及均值皆为最大，因此，选定  $k$  为 4。

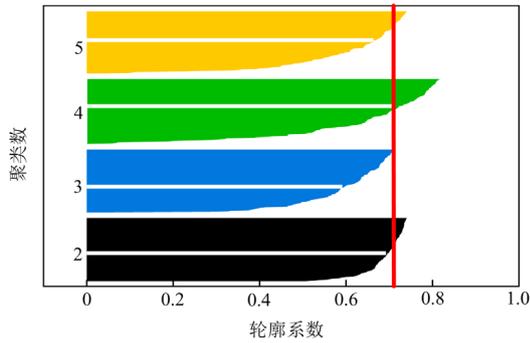


图 16 不同聚类数的轮廓系数对比图

Fig. 16 Silhouette coefficient comparison of different cluster numbers

基于最佳聚类数  $K$  对该场景下线损率数据进行聚类，图 17 给出了线损率分布的可视化情况。

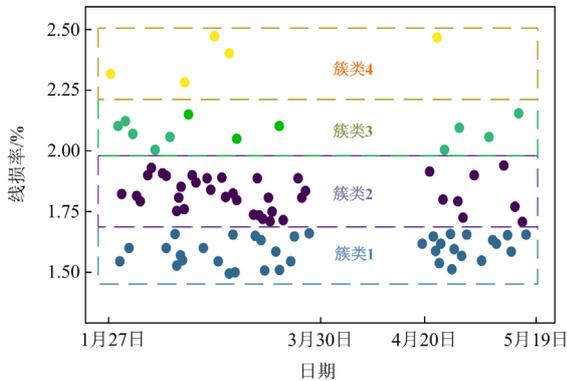


图 17 线损率聚类可视化

Fig. 17 Visualization of line loss rate clustering

根据区间动态平移策略，需保留簇类 2(即个案数目最多的簇类)中所有的线损率数据，得到区间 [1.68, 1.99]，依据式(13)和式(14)再分别计算簇类 1 和 3 的区间长度 ( $L_1 = 0.2, L_3 = 0.14$ )、簇类 1 相对簇类 2 下限的相对偏移量  $\Delta C_j$  ( $\Delta C_j = 0.0774$ ) 以及簇类 3 的相对偏移量  $\Delta C_k$  ( $\Delta C_k = 0.0108$ )，最终得到台区线损率标准库区间 [1.6026, 2.0008]。表 3 给出了图 15 中 8 个场景的标准库区间。

表 3 多维场景下台区线损率标准库

Table 3 Line loss standard library of the substation area under multidimensional scene

场景	标准库下限/%	标准库上限/%
1	2.4981	2.5195
2	1.6026	2.0008
3	2.1945	2.4674
4	2.1532	2.421
5	2.423	2.5068
6	5.0625	5.2969
7	5.3787	7.1306
8	2.6944	4.8396

图 18 进一步展示了不同场景下标准库区间的重叠率情况。从图 18 中可以看出，区间重叠率矩阵高度稀疏，说明了大部分场景之间不存在重合部分，难以被合并。其中，场景 3 和 4 的区间相似度为 0.721，而场景 5 求得的相似度均在 0.1 左右。

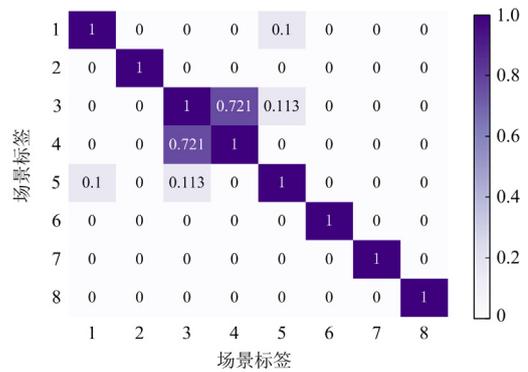


图 18 不同场景下标准库区间的重叠率

Fig. 18 Overlap rate between standard libraries in different scenarios

图 19 给出了类间方差与重叠率阈值的学习曲线。从图 19 中可以看出，当  $\alpha \in [0.148, 0.7]$  时，类间方差最大，本文选取最佳阈值为 0.5，因此可将场景 3 和场景 4 按照区间交集原则进行合并，得到新的标准库区间 [2.1945, 2.421]，而场景 5 的重叠度小于 0.5，则不合并。

图 20 给出了不同标准库建立方法的对比图。图 20(a)规定的线损率正常范围为 0~8%，因为按人工经验指定的标准，10 kV 供电的线损率应在 0~8%，因而正常范围为 0~8%；图 20(b)则考虑了线损率的最大和最小值，以上两种方法都无法准确地辨识出异常值；图 20(c)仅选取了个案数最多的簇类，常导致误判；而图 20(d)基于各簇类个案数的分布情况，构造了相对偏移量，在传统区间的基础上扩展了两块新区域，一定程度上增加了标准库的包容性。

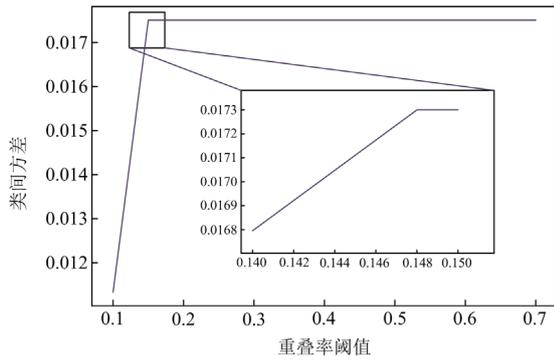


图 19 类间方差与重叠率阈值的学习曲线

Fig. 19 Learning curve of inter-class variance and overlap threshold

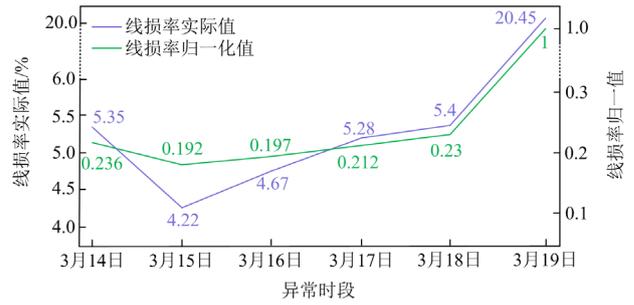


图 21 异常时段内的线损率分布情况

Fig. 21 Distribution of line loss rate in abnormal time

表 4 异常线损率与部分用户的 PCCs 和 ED 计算结果

Table 4 Abnormal line loss rate and PCCs and ED calculation results of some users

用户编号	传统关联算法		异常追踪结果	
	PCCs	ED	PCCs	ED
1	0.512	1.56		
2	0.463	2.626		
3	0.778	1.056		✓
4	0.862	1.692	✓	
5	0.796	1.365		○
6	0.69	1.27		

注：“✓”表示仅考虑 PCCs 或 ED 的异常追踪结果(即可得用户 4 或 3 为嫌疑用户)；“○”表示根据现场排查得出的真实异常用户(即可得用户 5 为嫌疑用户)。

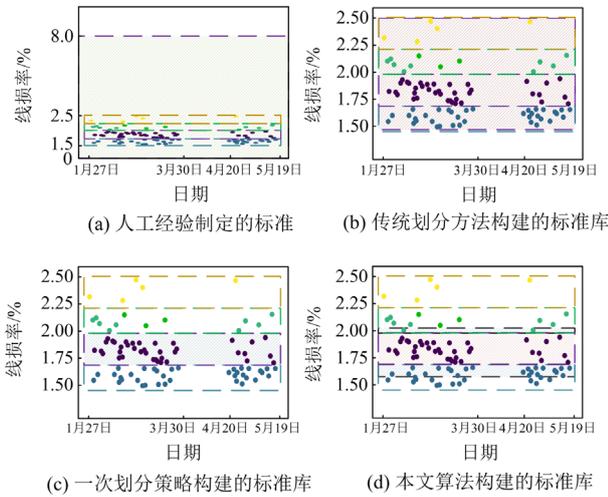


图 20 不同标准库建立方法对比图

Fig. 20 Comparison of different libraries building methods

### 5.3 台区线损率异常研判分析

图 21 给出了异常时段  $T_c$  内的线损率分布情况。从图 21 中可以看出,  $T_c$  分布在 3 月 14 日至 3 月 19 日; 在 3 月 15 日线损率最低为 4.22, 3 月 16 至 3 月 18 日逐步上升, 在 19 日发生突增的现象, 较 18 日增加了 278%左右。此外, 3 月 14 日至 18 日的线损率归一值变化趋势较为平稳, 而 19 日达到峰值 1, 因此, 3 月 19 日线损率存在异常的可能性最大, 需进一步分析异常线损率和台区用户的关联性。

表 4 给出了异常线损率与部分用户的皮尔逊相关系数 PCCs 和欧氏距离 ED 计算结果(限于篇幅, 只罗列了包含 PCCs 最大和 ED 最小的部分用户)。从表 4 中可以看出, 用户 4 的 PCCs 值最大, 为 0.862; 用户 3 的 ED 值最小, 为 1.056。由以上可知, 传统关联算法追踪异常用户的标准过于笼统, 缺乏特殊性, 导致结果与实际情况不符, 无法实现异常关联追踪。

为解决传统算法的误判问题, 首先采用灰色关联分析法确定线损异常强相关因素, 表 5 给出了本文 8 个场景各因素的灰色关联系数情况<sup>[24]</sup>。从表 5 中可以看出, 这 8 个场景售电量和供电量对线损率关联度较高, 场景 5—7 处于夏季, 因此平均温度对线损率的影响较大。为了减少计算量, 每个场景选择关联度排名靠前的因素进行后续的研究。

表 5 8 个场景各因素的灰色关联系数结果

Table 5 Results of grey correlation coefficients of factors in 8 scenarios

场景编号	售电量	供电量	平均温度	公变台数	公变总容量
1	<b>0.837</b>	<b>0.841</b>	0.513	<b>0.653</b>	0.641
2	<b>0.859</b>	<b>0.846</b>	0.544	0.608	<b>0.639</b>
3	<b>0.845</b>	<b>0.852</b>	0.551	0.642	<b>0.647</b>
4	<b>0.861</b>	<b>0.877</b>	0.598	0.615	0.622
5	<b>0.889</b>	<b>0.884</b>	<b>0.715</b>	0.677	0.669
6	<b>0.863</b>	<b>0.875</b>	<b>0.729</b>	0.650	0.653
7	<b>0.849</b>	<b>0.831</b>	<b>0.671</b>	0.604	0.602
8	<b>0.846</b>	<b>0.840</b>	0.505	0.636	<b>0.640</b>

对于各场景线损异常强相关因素,采用改进 Adtributor 进行分析,表 6 给出了用户  $i(i=1,2,\dots,6)$  S 值的计算结果。从表 6 中可以看出,用户 4 和 6 的 S 值均为 0,说明以上两个用户不存在异常用电嫌疑。同时,将剩余用户存入可疑用户集(包含用户 1、2、3 和 5)。

表 6 部分用户 S 值的计算结果

Table 6 Calculation results of S value of partial users' electric quantity

$T_c$	用户编号					
	1	2	3	4	5	6
3月14日	0	0	0	0	0	0
3月15日	0	<b>0.03</b>	0	0	0	0
3月16日	<b>0.02</b>	0	<b>0.01</b>	0	<b>0.01</b>	0
3月17日	<b>0.03</b>	0	<b>0.06</b>	0	<b>0.02</b>	0
3月18日	0	0	0	0	<b>0.33</b>	0
3月19日	0	0	0	0	<b>0.48</b>	0

表 7 给出了可疑用户集电量 EP 值的计算结果。从表 7 中可以看出,当 S 值大于 0(表 6 中粗体部分)时,相应的 EP 值才存在。进一步地,用户 2 的 EP 值为 0,说明不存在异常;用户 1 和 3 的 EP 总和位于区间(0,100%),说明以上两个用户存在嫌疑;而用户 5 含有两个 EP 大于 100%(即 130%和 250%),则说明用户 5 在 3 月 18 日和 19 日的电量存在异常的概率极大<sup>[25]</sup>。此时更新可疑用户集(包含用户 1、3 和 5)。

表 7 可疑用户集电量 EP 值的计算结果

Table 7 Calculation result of the EP value of the collection electric quantity of the suspect user

$T_c$	用户编号				
	1	2	3	4	5
3月14日	—	—	—	—	—
3月15日	—	0	—	—	—
3月16日	10%	—	0	—	0
3月17日	25%	—	60%	—	0
3月18日	—	—	—	—	130%
3月19日	—	—	—	—	250%

结合表 6 和表 7,计算可疑用户集中嫌疑电量(即 EP 大于 0)对应的 S 值总和,可得

$$\begin{cases} S_{\Sigma 1} = S_{EP(25\%)} + S_{EP(10\%)} = 0.03 + 0.02 = 0.05 \\ S_{\Sigma 3} = S_{EP(60\%)} = 0.06 \\ S_{\Sigma 5} = S_{EP(250\%)} + S_{EP(130\%)} = 0.48 + 0.33 = 0.81 \end{cases} \quad (15)$$

由式(15)可知,用户 5 的 S 值最大,且用户 1 和 3 较用户 5 相差甚多,因此,可判定用户 5 为异常用户,与表 4 中现场排查结果一致,进而验证了

改进 Adtributor 算法在实际工作中相较于传统关联算法的优越性。

通过以上分析可以看出,本文所提方法较好地完成了台区异常线损率的智能辨识,进而实现了异常关联用户的精准追踪。

### 5.4 实际应用分析

本节以 8 个公变台区的真实数据作为样本进行实际应用分析。采用软件 PyQt5 进行开发,部分功能结果如下。

图 22 给出了 8 个台区的多维场景的划分结果。图 23 展示了 8 个台区线损率异常的研判结果,从图 23 中可以看出,有且仅有 1 个台区存在异常(红色方框标记),其余均为正常台区。

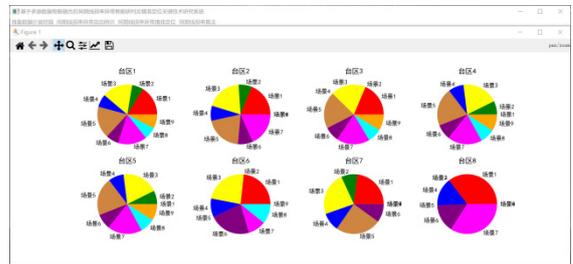


图 22 多维场景划分结果

Fig. 22 Results of multidimensional scene partition

1	2	3
台区编号	台区名称	台区线损是否异常
2 900907971	恒*小区变电站间隔配电变压器	正常
3 4193939	恒*东方#7箱变变压器	正常
4 354812	恒*东方	正常
5 900921893	华*家园2	正常
6 900921892	华*小区	正常
7 900922423	华*小区3	正常
8 715811	金*变压器	正常
9 4264248	角*公司变压器	异常

图 23 台区线损率异常判断结果

Fig. 23 Judgment result of abnormal line loss rate of substation area

图 24 给出了异常台区线损率关联用户的追踪结果。从图 24 中可以发现,编号为 5600097158 和 5600086740 的用户(红色方框标记)存在异常用电的可能性最大,通过现场排查发现编号为 5600097158 的用户日电量超过 300 kWh,电流超过 60 A,导致表计量不准确,和计算结果一致,进一步验证了本文所提方法在实际电网中应用的可行性和合理性。

为进一步验证本文策略在实际电网应用中的可行性,继续用山东某区域 978 个公变台区的真实数据作为样本进行实际应用分析。分别使用 4 种标准库来测算这 978 个台区中的异常台区数量,并用

PCCs、ED 和本文所述 S 值追踪到的异常用户数量, 结果如表 8 所示。

1 用户编号	2 用户名称	3 可疑用户电量的S值
2 5600097158	张"阳	0.5630658527652115
3 5600086740	洪"成	0.5568927369631227
4 5600086803	曾"盛	0.4982928818891852
5 7505947016	陈"	0.450488460288038
6 5600086813	李"花	0.44035287139505397

图 24 线损率异常关联用户的追踪结果

Fig. 24 Tracking results of abnormal line loss rate associated with user

表 8 山东某区域台区的筛选结果

Table 8 Screening results of a station in Shandong province

	人工经验	传统划分	一次划分	本文划分
PCCs	288	311	324	335
ED	269	320	333	339
S 值	314	334	345	355

从表 8 中可以看出, 通过本文划分策略得出的标准库能够研判出更多的线损异常用户。由此可得本文所提出的异常用户的追踪方法是领先可行的。

为了比较传统聚类策略和本文融合方法之间的优劣, 将间隙统计-轮廓系数(gap static algo-rithm-silhouette coefficient, GSA-SC)聚类算法作为对照, 在山东某地区选取 978 个台区作为样本, 将各方法研判结果对照现场排查结果, 得到线损研判准确率<sup>[31]</sup>。图 25 为传统聚类策略下 4 种标准库的线损研判准确率, 图 26 为本文融合策略下 4 种标准库的线损研判准确率; 在统一使用本文标准库生成方法的情况下, 通过计算, 采用传统聚类策略的线损研判平均准确率为 94.48%, 而采用本文融合方法的线

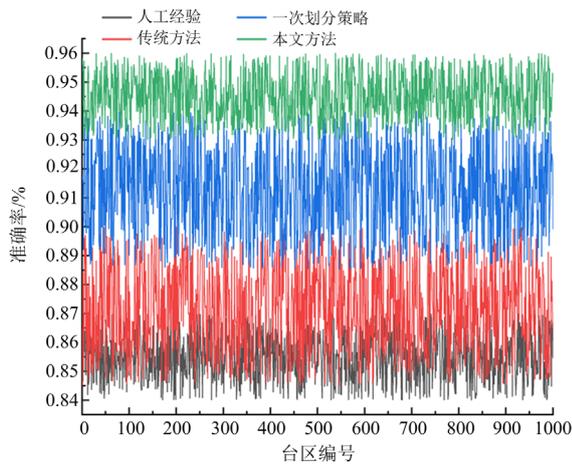


图 25 传统聚类策略下 4 种标准库的线损研判准确率

Fig. 25 Line loss estimation accuracy of four standard libraries under traditional clustering strategy

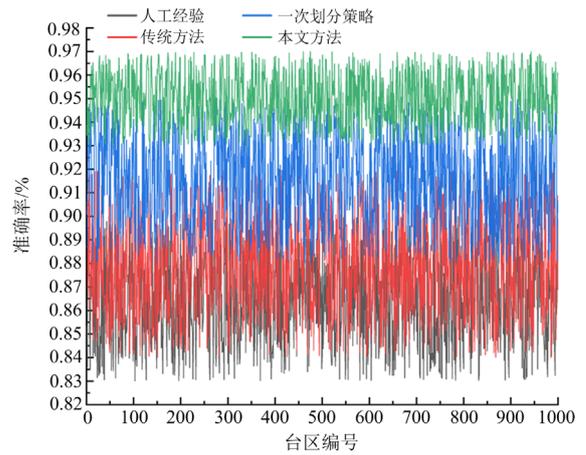


图 26 融合策略下 4 种标准库的线损研判准确率

Fig. 26 Line loss estimation accuracy of four standard libraries under fusion strategy

损研判平均准确率为 95.86%, 提高了 1.38%, 故体现本文融合方法精细划分场景的优越性; 同时, 从图 26 可以看出, 采用本文方法得出的标准库、线损研判准确率普遍高于其余 3 种方法, 故体现出采用本文区间动态平移策略得出标准库进行线损异常研判的可行性。

## 6 结论

本文提出了一种数据驱动下台区线损率异常智能研判及关联用户精准追踪方法, 结论如下。

1) 提出一种基于历史数据挖掘的台区线损率多维场景划分策略, 采用模态分解算法进行预处理, 并给出一种融合算法提高聚类精度, 实现了对台区线损率标准库的完备划分, 为后续线损率异常辨识提供了数据支撑。

2) 给出一种基于台区线损率标准库的线损率异常研判策略, 通过构建台区多维场景下的线损率标准形态, 进而对台区异常线损进行自动研判, 减轻了运维人员的工作强度, 提高了排查的精度和准确性。

3) 给出一种台区线损率异常关联用户精准追踪方法, 通过灰色关联分析法和改进 Adtributor 算法降低对异常关联用户的误判概率, 相比传统关联算法能更加准确地发现线损率异常的关联用户。

4) 构建了一种基于数据驱动的台区线损率异常智能研判及关联用户精准追踪框架, 该框架不仅对单个台区的线损率异常辨识和追踪较为准确, 而且适用于多台区场景。

本文对台区线损率异常辨识和关联用户追踪等问题进行了深入的研究, 然而, 随着排查精度的不

断提高, 用户异常产生的原因将更多地引起运维人员的关注, 因此, 用户异常的成因分析将是未来值得进一步深入探讨的问题。

### 参考文献

- [1] 林宝德, 杨铮宇. 基于多维特征的电网台区线损数据异常识别研究[J]. 电力系统保护与控制, 2022, 50(9): 172-178.  
LIN Baode, YANG Zhengyu. Anomaly recognition of line loss data in power grid stations based on multi-dimensional features[J]. Power System Protection and Control, 2022, 50(9): 173-178.
- [2] 夏睿, 高云鹏, 朱彦卿, 等. 基于 SE-CNN 模型的窃电检测方法研究[J]. 电力系统保护与控制, 2022, 50(20): 117-126.  
XIA Rui, GAO Yunpeng, ZHU Yanqing, et al. A detection method of electricity theft behavior based on an SE-CNN model[J]. Power System Protection and Control, 2022, 50(20): 117-126.
- [3] 张辰琪, 朱成昊, 李洋, 等. 基于随机模糊潮流的三相不平衡配电网节能降损方法[J]. 电力科学与技术学报, 2022, 37(6): 212-218.  
ZHANG Chenqi, ZHU Chenghao, LI Yang, et al. Research on energy saving and loss reduction method of three-phase unbalanced power grid based on stochastic fuzzy power flow[J]. Journal of Electric Power Science and Technology, 2022, 37(6): 212-218.
- [4] 李政麒, 蔡晔, 唐夏菲, 等. 面向降损场景的移动储能与网络重构协同优化策略[J]. 电力建设, 2023, 44(9): 137-148.  
LI Zhengqi, CAI Ye, TANG Xiafei, et al. Collaborative optimization strategy of mobile energy storage devices and distribution network reconfiguration for power loss reduction scenarios[J]. Electric Power Construction, 2023, 44(9): 137-148.
- [5] 周王峰, 李勇, 郭钊秀, 等. 基于 DAE-LSTM 神经网络的配电网日线损率预测[J]. 电力系统保护与控制, 2021, 49(17): 48-56.  
ZHOU Wangfeng, LI Yong, GUO Yixiu, et al. Daily line loss rate forecasting of a distribution network based on DAE-LSTM[J]. Power System Protection and Control, 2021, 49(17): 48-56.
- [6] 仇继扬, 李涵, 王泽忠, 等. 含分布式光伏的低压台区三相线损理论计算方法研究及其应用[J]. 供用电, 2023, 40(11): 69-75.  
QIU Jiyang, LI Han, WANG Zezhong, et al. Research and application on three-phase line loss improvement algorithm for low-voltage distribution network with distributed photovoltaic and its application[J]. Distribution & Utilization, 2023, 40(11): 69-75.
- [7] 李晋源, 保富, 胡凯, 等. 基于 Stacking 集成模型的台区线损率预测方法研究[J]. 电测与仪表, 2023, 60(1): 71-77.  
LI Jinyuan, BAO Fu, HU Kai, et al. Research on prediction method of line loss rate in station area based on Stacking ensemble model[J]. Electrical Measurement & Instrumentation, 2023, 60(1): 71-77.
- [8] 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.
- [9] ZHAO G, CHU J, DENG L, et al. Research on line-transformer-user topological anomaly recognition model based on multi-source data mining[C] // 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), 2020: 192-196.
- [10] 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.
- [11] 赵文清, 沈哲吉, 李刚. 基于深度学习的用户异常用电模式检测[J]. 电力自动化设备, 2018, 38(9): 34-38.  
ZHAO Wenqing, SHEN Zheji, LI Gang. Anomaly detection for power consumption pattern based on deep learning[J]. Electric Power Automation Equipment, 2018, 38(9): 34-38.
- [12] YANG X, SUN Y, LI W. Abnormal user detection based on the correlation probabilistic model[J]. Security and Communication Networks, 2020(43): 1-12.
- [13] VIEGAS J L, VIEIRA S M. Clustering-based novelty detection to uncover electricity theft[C] // 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 9-12, 2017, Naples, Italy: 779-1552.
- [14] CHAI Z L, ANG S P, SAIAM M A, et al. Demand-based photovoltaic orientation and power flow assessments of a 11 kV network[C] // 7th Brunei International Conference on Engineering and Technology 2018 (BICET 2018), 2018: 1-4.
- [15] 张宁, 朱永利, 高艳丰, 等. 基于变分模态分解和概率密度估计的变压器绕组变形在线检测方法[J]. 电网技术, 2016, 40(1): 297-302.

- ZHANG Ning, ZHU Yongli, GAO Yanfeng, et al. An on-line detection method of transformer winding deformation based on variational mode decomposition and probability density estimation[J]. *Power System Technology*, 2016, 40(1): 297-302.
- [16] 吴春华, 黄宵宵, 李智华, 等. 光伏系统直流微弱电弧信号检测研究[J]. *中国电机工程学报*, 2019, 39(20): 6025-6033, 6183.
- WU Chunhua, HUANG Xiaoxiao, LI Zhihua, et al. Research on DC weak arc signal detection in photovoltaic system[J]. *Proceedings of the CSEE*, 2019, 39(20): 6025-6033, 6183.
- [17] 刘云鹏, 许自强, 董王英, 等. 基于经验模态分解和长短期记忆神经网络的变压器油中溶解气体浓度预测方法[J]. *中国电机工程学报*, 2019, 39(13): 3998-4008.
- LIU Yunpeng, XU Ziqiang, DONG Wangying, et al. Concentration prediction of dissolved gases in transformer oil based on empirical mode decomposition and long short-term memory neural networks[J]. *Proceedings of the CSEE*, 2019, 39(13): 3998-4008.
- [18] SIBTAIN M, LI X, SALEEM S, et al. A multistage hybrid model ICEEMDAN-SE-VMD-RDPG for a multivariate solar irradiance forecasting[J]. *IEEE Access*, 2021, 9: 37334-37363.
- [19] HUANG Z, QIU J, KONG D, et al. A novel in-band OSNR measurement method based on normalized autocorrelation function[J]. *IEEE Photonics Journal*, 2018, 10(2): 1-8.
- [20] SINAGA K P, YANG M. Unsupervised K-means clustering algorithm[J]. *IEEE Access*, 2020, 8: 80716-80727.
- [21] LEE J, LEE D. An improved cluster labeling method for support vector clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3): 461-464.
- [22] 王帅, 杜欣慧, 姚宏民, 等. 面向含多种用户类型的负荷曲线聚类研究[J]. *电网技术*, 2018, 42(10): 3401-3412.
- WANG Shuai, DU Xinhui, YAO Hongmin, et al. Research on load curve clustering with multiple user types[J]. *Power System Technology*, 2018, 42(10): 3401-3412.
- [23] LOPEZ-RAMIREZ M, LEDESMA-CARRILLO L M, GARCIA-GUEVARA F M, et al. Automatic early broken-rotor-bar detection and classification using otsu segmentation[J]. *IEEE Access*, 2020, 8: 112624-112632.
- [24] 韩平平, 陈思远, 张楠, 等. 考虑分布式光伏的低压台区线损异常辨识方法[J]. *电力系统保护与控制*, 2023, 51(8): 140-148.
- HAN Pingping, CHEN Siyuan, ZHANG Nan, et al. Line loss anomaly identification method for low-voltage station area considering distributed PV[J]. *Power System Protection and Control*, 2019, 51(8): 140-148.
- [25] BHAGWAN R, KUMAR R, RAMJEE R, et al. Adtributor: revenue debugging in advertising systems[C]// *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI' 14)*, 2014: 43-55.
- [26] 张文韬, 王秀丽, 李言, 等. 大规模风电并网下多区域互联系统热电综合调度模型[J]. *电网技术*, 2018, 42(1): 154-161.
- ZHANG Wentao, WANG Xiuli, LI Yan, et al. An analysis model of multi-area interconnected power systems with large-scale wind power involved in comprehensive heating and power system scheduling[J]. *Power System Technology*, 2018, 42(1): 154-161.
- [27] 关治, 陆金甫. *数值方法*[M]. 北京: 清华大学出版社, 2006.
- [28] 陈浩文, 刘文霞, 李月乔. 基于奇异谱分析与神经网络的中期负荷预测[J]. *电网技术*, 2020, 44(4): 1333-1347.
- CHEN Haowen, LIU Wenxia, LI Yueqiao. Medium-term load forecast based on singular spectrum analysis and neural network[J]. *Power System Technology*, 2020, 44(4): 1333-1347.
- [29] 邓带雨, 李坚, 张真源, 等. 基于EEMD-GRU-MLR的短期电力负荷预测[J]. *电网技术*, 2020, 44(2): 593-602.
- DENG Daiyu, LI Jian, ZHANG Zhenyuan, et al. Short-term electric load forecasting based on EEMD-GRU-MLR[J]. *Power System Technology*, 2020, 44(2): 593-602.
- [30] 别锋锋, 都腾飞, 庞明军, 等. 基于ICEEMDAN-ELM的管道声信号识别方法研究[J]. *噪声与振动控制*, 2019, 39(4): 205-211.
- BIE Fengfeng, DU Tengfei, PANG Mingjun, et al. Research on the method of acoustic signals recognition for pipelines based on ICEEMDAN-ELM[J]. *Noise and Vibration Control*, 2019, 39(4): 205-211.
- [31] 徐迪, 陆煜铨, 肖勇, 等. 基于孤立森林算法的配电网线损异常判定[J]. *电力系统保护与控制*, 2021, 49(16): 12-18.
- XU Di, LU Yuxin, XIAO Yong, et al. Identification of abnormal line loss for a distribution power network based on an isolation forest algorithm[J]. *Power System Protection and Control*, 2021, 49(16): 12-18.

收稿日期: 2023-07-02; 修回日期: 2024-02-05

作者简介:

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

张盛杰(1999—), 男, 硕士研究生, 研究方向为配电网线损异常分析; E-mail: zhangsjphl@foxmail.com

杨里(1979—), 男, 硕士, 高级工程师, 研究方向为同期线损分析. E-mail: Yang\_li@fj.sgcc.com.cn

(编辑 魏小丽)