

DOI: 10.19650/j.cnki.cjsi.J2514760

# 基于互信息图与邻接特征的非高损线路窃电检测方法

周文晴<sup>1</sup>, 杨茂涛<sup>2,3</sup>, 苏 盛<sup>1</sup>, 赵 斌<sup>1</sup>, 李 彬<sup>1</sup>

(1. 长沙理工大学电气与信息工程学院 长沙 410114; 2. 国网湖南省电力有限公司供电服务中心(计量中心) 长沙 410007; 3. 智能电气量测与应用技术湖南省重点实验室 长沙 410004)

**摘要:**现有窃电检测方法多聚焦于日线损率异常突出的高损线路,而非高损线路中的潜在短时、间歇性窃电行为易被负荷波动、计量误差与运行方式变化掩盖,难以有效辨识。为此,提出了一种基于互信息差值动态图与邻接特征的非高损线路窃电检测方法。首先,在小时尺度下采用滑动窗口计算用户用电量与线损电量的互信息,构建互信息差值动态图序列;并引入逐窗口自适应阈值以刻画用户用电量与线损之间随时间演化的非线性依赖关系,从而避免固定阈值导致的图结构不稳定。其次,设计融合图自编码器与动态权重平衡机制的特征提取方法;基于跨窗口的邻接关系重构任务,捕捉用户在群体网络中连接模式的时序演变,提取反映用户与群体依赖关系稳健性的邻接特征;同时采用动态权重加权损失,抑制正负样本不平衡引起的训练偏置,防止图重构学习发生退化。最后,采用主成分分析对嵌入特征降维优选,并结合 K-Means++ 聚类实现无监督异常识别。仿真算例与真实数据实验表明,所提方法能够在非高损场景下有效识别窃电用户,验证了方法的可行性与有效性。该方法从群体关系演化视角刻画用户用电行为与线损之间的动态依赖关系,将窃电检测问题转化为图异常检测问题,为中压线路窃电检测提供了一种有效且具工程实用性的技术路径。

**关键词:** 非高损线路;窃电检测;数据驱动;互信息;图自编码器

**中图分类号:** TN93 TH89 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Electricity theft detection method for non-high-loss feeders based on mutual information graphs and adjacency features

Zhou Wenqing<sup>1</sup>, Yang Maotao<sup>2,3</sup>, Su Sheng<sup>1</sup>, Zhao Bin<sup>1</sup>, Li Bin<sup>1</sup>

(1. School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China;  
2. State Grid Hunan Electric Power Company Limited Power Service Center (Metrology and Test Center), Changsha 410007, China;  
3. Hunan Province Key Laboratory of Intelligent Electrical Measurement and Application Technology, Changsha 410004, China)

**Abstract:** Existing electricity theft detection methods primarily focus on high-loss feeders, where anomalies in the daily average line-loss rate are pronounced. However, potential short-term and intermittent electricity theft on non-high-loss feeders is easily masked by load fluctuations, metering errors, and changes in operating conditions, making such theft difficult to detect effectively. To address this issue, this paper proposes a method for electricity theft detection on non-high-loss feeders based on a mutual-information-difference dynamic graph sequence and adjacency features. First, at an hourly resolution, a sliding window is employed to calculate the mutual information between users' electricity consumption and line-loss electricity, thereby constructing a mutual-information-difference dynamic graph sequence. A window-wise adaptive threshold is introduced to capture the nonlinear dependency between users' electricity usage and line loss as it evolves over time, avoiding graph-structure instability caused by a fixed threshold. Second, a feature extraction method combining a graph autoencoder and a dynamic weight balancing mechanism is developed. Based on a cross-window adjacency reconstruction task, it captures the temporal evolution of connection patterns within the population network and extracts adjacency features that reflect the robustness of users' dependency relationships with the group. Meanwhile, a dynamically weighted loss function is applied to mitigate training bias caused by the imbalance between positive and negative samples and to prevent degeneration in graph reconstruction learning. Finally, principal component analysis is used to reduce and select the embedded features, and K-Means++

clustering is applied for unsupervised anomaly identification. Simulation studies and experiments on real-world data demonstrate that the proposed method can effectively detect electricity-theft users in non-high-loss scenarios, verifying its feasibility and effectiveness. From the perspective of group relationship evolution, this method characterizes the dynamic dependencies between users' electricity consumption and line loss, transforms electricity theft detection into a graph anomaly detection problem, and provides an effective and practically applicable technical solution for electricity theft detection on medium-voltage feeders.

**Keywords:** non-high-loss feeders; electricity theft detection; data-driven methods; mutual information; graph auto-encoder

## 0 引 言

中压配电线路中,用户窃电行为是导致配电损耗的主要因素之一<sup>[1]</sup>。此类行为不仅直接侵害供电企业的经济利益,还会破坏电网的安全稳定运行,并可能诱发火灾、触电乃至大范围停电事故<sup>[2-3]</sup>。智能电表在配电网中的普及推动了计量方式的智能化,但与此同时,窃电手段也呈现出多样化与隐蔽化趋势,使得窃电检测的难度不断加大<sup>[4-5]</sup>。窃电行为已成为全球电力行业亟待解决的重大问题。相关研究表明,全球138个国家因窃电造成的年度经济损失高达1 012亿美元<sup>[6]</sup>,凸显了该问题的普遍性和严重性。因此,研究并应用有效的窃电检测技术,对于提升电能利用效率、保障电力企业利益以及维护公共用电安全,具有重要的现实意义。

传统的窃电检测主要依赖人工巡检与经验统计,其检测效率低、覆盖面有限,难以满足当前配电系统的需求。随着高级计量基础设施的逐步普及,大规模用电数据的获取为数据驱动的窃电检测方法提供了可能<sup>[7]</sup>。近年来,数据驱动的窃电检测技术在国内外均进行了广泛研究,主要可分为基于机器学习、深度学习以及状态估计的方法<sup>[8-10]</sup>。基于机器学习的方法认为正常用户的用电行为具有相似性和规律性,可以通过用户负荷特征或统计指标划分正常与异常群体。文献[11]通过融合多种异质学习器来构建集成模型,从而更好地发挥不同学习器的互补优势。然而,基于机器学习的方法实现相对简单,但往往依赖于特征工程,难以自适应捕捉中压用户负荷与线损之间复杂的非线性动态关系。基于深度学习的方法通过构建神经网络模型,挖掘用户用电数据中的潜在模式。文献[12]结合卷积自编码器与长短期记忆网络构造窃电检测模型,通过二维转换和自编码器提取电力数据的空间特征,融合时空特征实现窃电用户定位。基于深度学习的方法具备更强的特征学习能力,但其训练过程往往需要大量历史数据和高质量的标注样本,且可解释性较为不足。基于状态估计的方法将系统的运行状态与电气量之间的数学关系建立起来,并通过状态估计算法对当前时刻的量测值进行一致性检验。文献[13]基于三相状态估计器的数据分析技术,提出了一种利用加权最小二乘与坏数据分析框架,并结合测量残

差相关性指标的检测与定位方法,用于识别配电网中的窃电用户。基于状态的方法物理意义明确,然而往往依赖高精度的量测装置和准确的系统参数,在实际应用场景中受到一定限制。近年来,图表示学习的快速发展也进一步推动了图异常检测研究,相关方法通常在图结构与节点表示空间中识别偏离模式,实现对异常节点、异常边或异常子图的检测<sup>[14]</sup>。文献[15]通过将时间序列特征转化为嵌入向量并构建图结构数据,以提高复杂异常模式下的检测性能。然而,现有无监督图异常检测方法在建模上仍多聚焦于节点属性与固定的相关性构图,侧重于识别节点属性显著偏离或结构异常的节点。同时,静态图难以刻画关系结构随时间演化的动态过程,从而对非高损线路中短时、间歇性窃电引起的时序关系偏离不够敏感。

现有方法虽在不同层面取得一定成效,但大多数研究仍集中于低压住宅用户场景,对于酒店、工业厂房等中压线路下的专变用户关注不足<sup>[16]</sup>。一方面,中压线路用户的用电特性复杂,工商业用户电量常随生产计划呈现显著波动<sup>[4]</sup>,难以直接从负荷曲线中提取稳健特征;基于深度学习的有监督方法虽具备自动特征学习能力,但依赖大量标注样本且可解释性不足;而基于状态估计的方法虽然物理意义明确,却往往依赖高精度量测装置和准确的系统参数,实际应用受限<sup>[10]</sup>。另一方面,工程实践中普遍以日均线损率为主要判据,选取配电损耗显著高于阈值的线路作为检测对象,而对非高损线路中潜在的窃电行为缺乏有效的识别方法<sup>[17]</sup>。在非高损线路中,窃电行为更可能以短时、分时段、间歇性方式发生,其对日均线损的边际贡献有限且统计特征不稳定,容易被工商业负荷波动及日尺度统计的平滑效应所掩盖,从而使以日均线损阈值为入口的工程筛查难以有效覆盖该类风险。与此同时,稽查资源往往优先投向高损线路,使得非高损场景更难获取充足且高质量的窃电标注样本,进一步限制了依赖大量标注数据的有监督模型在该场景的适用性与可迁移性。

针对这些问题,为克服现有静态图建模与数值特征分析在非高损线路窃电识别中的局限,本文从群体关系演化视角切入,在互信息依赖关系分析的基础上,结合群体关系建模与图表示学习技术,以小时尺度用电信息为基础,提出一种面向非高损中压线路的窃电检测方法。

本文的主要贡献包括:

1) 提出基于互信息差值的动态图建模框架。通过滑动窗口计算用户用电量与线路总线损电量的互信息,并采用逐窗口自适应阈值构建互信息差值时序图,实现随时间窗口自适应的关系构建,减少对固定阈值与经验参数的依赖。从群体关系演化视角刻画用户用电行为与线损之间的动态依赖关系,将窃电检测问题转化为图异常检测问题。

2) 设计基于图自编码器的邻接特征提取方法。通过跨窗口的邻接关系重构任务,捕捉用户在群体网络中连接模式的时序演变规律。该机制能够有效提取反映用户与群体依赖关系稳健性的深层特征,利用正常用户与窃电用户在图结构中邻接关系稳定性的差异实现异常行为的表征学习。

3) 提出融合样本平衡机制的无监督异常识别框架。针对图中正负边样本严重不平衡问题,在图自编码器训练中引入动态权重调节与损失加权策略,防止模型在图重构任务中发生退化。在此基础上,结合主成分分析和K-Means++聚类算法,实现窃电用户的无监督识别。

仿真与实测数据结果表明,所提方法在无需标注数据的情况下即可准确识别非高损线路中的窃电用户,为中压配电网窃电检测提供了一种有效且具工程可实施性的技术路径。

## 1 非高损线路窃电行为机理与关联性分析

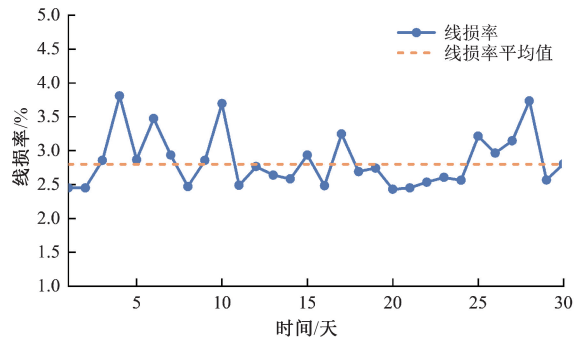
### 1.1 非高损线路窃电问题分析

中压配电线路是电能输送的重要环节,其运行状态直接关系到电网的经济效益与安全可靠性。在工程实践中,线路日均线损率常被作为判别依据<sup>[16]</sup>,一般假设阈值为7%,该数值来源于当前欧盟平均线损率7%<sup>[18]</sup>。当某线路的日均线损率超过这一预设阈值时,即被认定为异常高损线路,并进一步对其中用户进行排查。

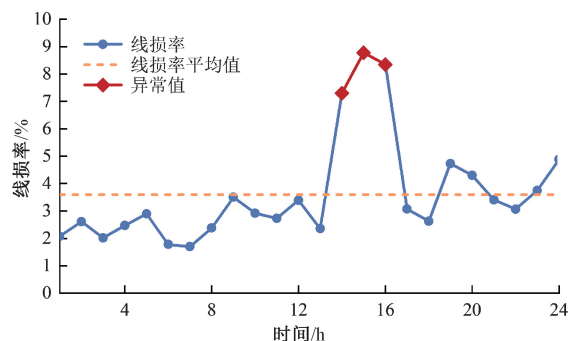
然而,这种基于日均线损率的分析方法存在局限性,容易忽略非高损线路中潜在的窃电行为,导致隐蔽性较高的窃电用户难以有效识别。以中国某地区的一条10 kV线路线损数据为例<sup>[19]</sup>。如图1(a)所示,在日尺度上,该线路的日线损率在2%~4%波动,处于正常范围,未能触发基于日均线损率的异常检测阈值。然而,实际情况表明,在第4个采集日发生了窃电行为。如图1(b)所示,在小时尺度下,部分时段的线损率骤增至8.7%,显著偏离正常范围。由于局部时段的剧烈波动在日均化过程中被平滑,窃电行为未被有效识别。

这一现象表明,即使线路整体日线损率表现为非高损,仍可能存在隐蔽的窃电问题。此类线路在大部分时间保持正常,仅在局部时段发生异常用电,造成短时线损

率剧烈波动。因此,仅依赖日线损率的高损线路判别方法存在明显局限性。要有效识别非高损线路中的窃电行为,必须结合更细粒度的时间尺度分析,并引入滑动窗口机制,从而捕捉日均化掩盖的短时异常特征。



(a) 日尺度线损率曲线  
(a) Daily-scale line loss rate curve



(b) 小时尺度线损率曲线  
(b) Hourly-scale line loss rate curve

图1 10 kV中压线路线损率曲线

Fig. 1 Line-loss rate curve of the 10 kV medium-voltage feeder

### 1.2 典型窃电模式分析

窃电用户通常通过绕表法、欠流法、移相法或网络攻击等方式篡改计量数据,以减少记录的用电量并获取非法电力。文献<sup>[20]</sup>归纳了当前常见的8种窃电类型,如图2所示。

前7种类型包括固定比例窃电、削峰窃电、阈值窃电、动态比例窃电、基于均值的随机比例窃电、分时段置零窃电和均值替代窃电方法。这些窃电方式通过干扰或篡改电能计量系统,人为降低用户的用电量,导致线路的非技术性损失增加,进而造成电力企业的经济损失和安全问题。相比之下,第8种窃电类型具有更高的隐蔽性。这种方式利用分时电价机制,通过篡改高峰时段的计量数据来减少计量电量。这种窃电策略使得窃电行为在日尺度用电数据中难以显现异常,增加了检测难度。

因此,有必要采用更细粒度的时序分析,并结合用户群体关系网络挖掘潜在异常。通过将单用户负荷特征分析拓展为群体依赖关系建模,有助于在非高损线路场景下实现更有效的窃电检测。

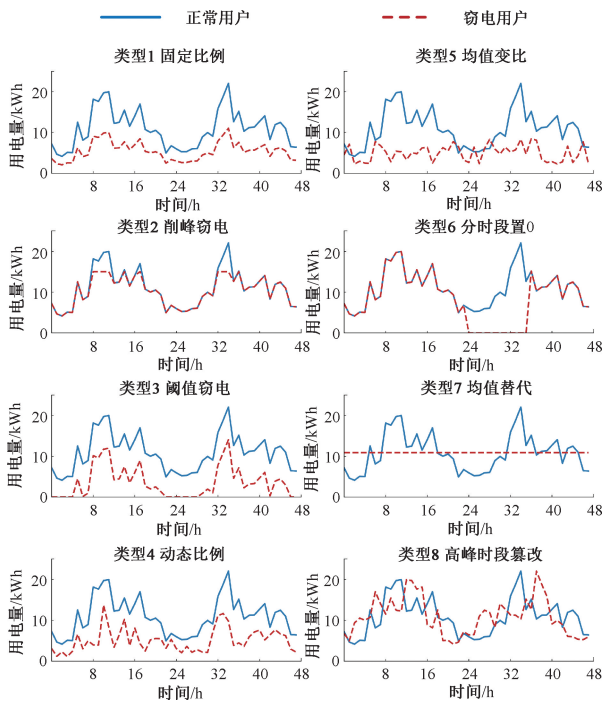


图2 8种窃电类型的示意图

Fig. 2 Schematic diagram of eight electricity-theft types

### 1.3 基于互信息的用户负荷与线损关系分析

配电网中,中压线路的损耗来源主要包括技术损耗(technical loss, TL)与非技术损耗(non-technical loss, NTL)。在线变户表关系正确的条件下,线路线损率 $L$ (单位为%)可用式(1)计算<sup>[21]</sup>,即:

$$L = \frac{I - S}{I} = \frac{\Delta W}{I} = \frac{TL + NTL}{I} \quad (1)$$

式中: $I$ 为线路输入电量; $S$ 为线路售电量; $\Delta W$ 表示线损电量,包含线路自身损耗导致的TL和主要由窃电所导致的NTL。当窃电发生时,输入电量超过售电量,导致NTL和 $\Delta W$ 增加,从而增加整体线损。

为深入分析用户用电行为与线损之间的依赖关系,本文引入互信息方法进行分析<sup>[22]</sup>。互信息是衡量随机变量间统计依赖程度的重要指标,能够同时捕捉线性和非线性相关性。相较于传统相关系数,互信息不受线性关系限制,因此更适合刻画用户负荷与线损之间的内在联系。本文采用滑动窗口方法,计算了某10 kV中压线路F16下各用户用电量与该路线损电量之间的互信息。数据集包含该线路的总线损电量以及线路下34个工商业专变用户的用电量,时间范围为2023年3月1日至2023年5月31日,共计2 208个连续小时的数据<sup>[19]</sup>。

图3展示了互信息分布的箱线图。箱体中间的横线表示互信息值的中位数。箱体的顶部和底部是互信息值的上四分位数和下四分位数,箱体外的圆点表示离群值。

由图可见,用户的箱体高度普遍较小,四分位距集中。互信息的中值、下四分位数、上四分位数分别分布在 $[1.79, 2.02]$ ,  $[1.56, 1.86]$ ,  $[2.06, 2.23]$ 。表明正常用户的互信息值波动性较小且分布模式相对一致,且其负荷与线损电量之间存在稳定的依赖关系。该群体层面的稳定依赖关系为后续互信息差值图的构建提供了依据。

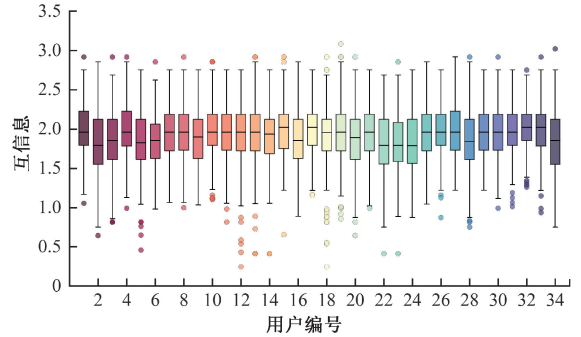


图3 互信息分布箱线图

Fig. 3 Boxplot of mutual information distribution

然而,仅依靠互信息分析进行用电行为特征刻画存在一定局限性。在实际配电网中,当线路数量众多且接入用户规模庞大时,仅凭箱线图或互信息统计特征难以直观识别异常模式。此外,窃电行为导致的局部互信息下降与正常用户因数据噪声或负荷波动引起的短时异常难以区分,容易造成误判。为此,本文进一步引入动态图构建方法。将用户视为节点,根据用户间互信息强度的相似程度构建边连接,从群体网络层面刻画用电行为与线损的依赖模式。相比单用户用电特征分析,图结构不仅保留了个体特征信息,还能通过群体关系约束凸显异常用户的相对偏离程度,从而更有效地识别潜在的窃电行为。

## 2 窃电检测模型及原理

### 2.1 互信息差值图构建

图是一种由节点和边组成的数据结构,相较于其他数据类型,能够更有效地表示对象之间的复杂关联关系<sup>[23]</sup>。与传统的序列化或向量化表示不同,图不仅保留了节点的个体属性,还通过边的连接体现节点之间的相互作用,从而在结构化层面揭示潜在的模式与规律。近年来,图结构已广泛应用于网络安全、金融交易监测和社交网络分析等领域,在复杂关系建模与异常检测中展现出独特优势。

在窃电检测问题中,单个用户的负荷曲线往往具有较强的随机性与波动性,仅依靠个体时序特征难以全面刻画其行为模式<sup>[24]</sup>。将多个用户置于统一的图结构框架下,可以通过节点表示用户自身特性,通过边表

示用户之间的相似性或差异性,从而在群体层面揭示协同行为与异常模式。互信息是一个随机变量中包含的关于另一个随机变量的信息量,是变量间相互依赖性的量度。1.3节已对用户负荷与线路损耗的关系进行了分析。基于此,本文将用户集合表示为图的节点,并根据用户之间互信息强度的相似性建立边。对于每个窗口 $t$ ,对每个用户计算用电量 $X$ 和线损电量 $W$ 之间的互信息<sup>[25]</sup>,即:

$$I(X;W) = \sum_{x,w} p(x,w) \ln \frac{p(x,w)}{p(x)p(w)} \quad (2)$$

式中: $I(X;W)$ 表示变量 $X$ 和 $W$ 之间的互信息; $P(x)$ 和 $p(w)$ 分别表示变量 $X$ 和 $W$ 的边际密度函数; $p(x,w)$ 表示变量 $X$ 和 $W$ 的联合概率密度函数。互信息还可以表示为熵的组合形式,即:

$$I(X;W) = H(X) + H(W) - H(X,W) \quad (3)$$

其中, $H(X)$ 为变量 $X$ 的信息熵,表示 $X$ 的不确定性; $H(W)$ 为变量 $W$ 的信息熵; $H(X,W)$ 为 $X$ 和 $W$ 的联合熵,表示两变量共同的不确定性。

采用长度为 $h$ ,步长为 $s$ 的滑动窗口对时间序列进行分段。在第 $t$ 个窗口内,得到每个用户的归一化后的互信息(normalized mutual information, NMI)。归一化处理有助于降低不同用户负荷幅值与波动尺度差异带来的影响,使不同用户的依赖强度具备可比性。群体建模中,将同一时间窗口内用户负荷对线损响应的依赖强度差异作为行为一致性的度量。当用户之间满足互信息差值较小,表示二者对线损响应的依赖强度相近,其依赖模式在群体层面表现一致;反之,当差值增大,表明其依赖强度存在更明显差异,行为一致性降低。基于此,以互信息差值描述用户间依赖相似度,并据此确定窗口 $t$ 的邻接关系。将所有用户表示为节点集合 $V = \{v_1, v_2, \dots, v_n\}$ 。若用户 $i$ 与用户 $j$ 的NMI值接近,说明其对线损的依赖模式相似,则在两者之间建立边,即:

$$A_{ij}^{(t)} = \begin{cases} 1, & |NMI_i^{(t)} - NMI_j^{(t)}| < \theta^{(t)} \\ 0, & \text{其他} \end{cases} \quad (4)$$

式中: $\theta^{(t)}$ 为逐窗口自适应阈值,取第 $t$ 窗口内所有用户间NMI差值分布的 $p$ 分位数,表达式为:

$$\theta^{(t)} = Q_p(\{|NMI_i^{(t)} - NMI_j^{(t)}|; i < j\}) \quad (5)$$

式中: $Q_p(\cdot)$ 表示分位数函数。该方法能够根据窗口内数据的统计特性动态调整阈值,避免固定阈值带来的过度稀疏或过度密集问题。 $p$ 采用第三四分位数。最终得到邻接矩阵 $A^{(t)}$ ,从而在每个窗口内形成一张互信息差值图 $G(t) = (V, A^{(t)})$ 。

## 2.2 图自编码器特征提取模块

图自编码器(graph autoencoder, GAE)是一类基于神经网络的图表示学习模型<sup>[26]</sup>。该方法通常由图卷积编

码器与内积解码器构成,通过对图结构的重建来学习节点的低维表示。凭借结构简洁、训练高效等优势,GAE在人脸识别、生物医学信息处理等领域得到了广泛应用<sup>[27]</sup>。

GAE特征提取模块示意图如图4所示。每个时间窗口 $t$ 均生成一张互信息差值图 $G(t)$ 。

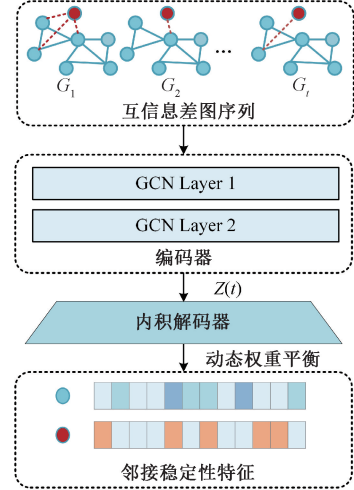


图4 GAE特征提取模块示意图

Fig. 4 Diagram of the GAE feature extraction module

码器部分采用两层图卷积网络(graph convolutional network, GCN),通过卷积操作聚合邻居节点的信息,得到节点嵌入表示 $Z^{(t)}$ ,即:

$$\begin{cases} Z^{(t)} = \tilde{A}(t) H^{(t)} \alpha_2 \\ H^{(t)} = \sigma(\tilde{A}(t) X^{(t)} \alpha_1) \end{cases} \quad (6)$$

式中: $X^{(t)}$ 为第 $t$ 窗口的节点输入特征; $\tilde{A}(t)$ 为第 $t$ 窗口的对称归一化邻接; $H^{(t)}$ 为第1层图卷积后的中间表示; $\alpha_1$ 和 $\alpha_2$ 为两层图卷积的可训练权重矩阵; $\sigma(\cdot)$ 为逐元素非线性激活。解码器通过内积重建邻接关系,即:

$$\hat{A}_{ij}^{(t)} = \text{sigmoid}(Z_i^{(t)} \cdot Z_j^{(t)T}) \quad (7)$$

式中: $\hat{A}_{ij}^{(t)}$ 表示节点 $i$ 与节点 $j$ 预测的邻接关系。

由于不同时间窗内差值分布存在差异,构建得到的图在各窗口的边密度及正负边比例不同,从而使窗口内正负边样本呈现类别不平衡。若直接采用未加权的二元交叉熵进行重构学习,模型的梯度易被数量占优的一类样本主导,从而产生重构偏置并降低对少数类结构信息的学习能力,因此需要引入样本平衡机制。常见的样本平衡策略还包括随机过采样、欠采样和负采样<sup>[28]</sup>等。采样方法会改变窗口内边分布并引入额外随机性,可能影响图结构重构任务的一致性。相比之下,按窗口自适应的类频率加权无需额外超参数,在不改变数据分布的前提下实现梯度平衡,适配滑动窗口构建的动态图训练场景。

为缓解边级样本不平衡导致的优化偏置, 在每个窗口内采用类频率平衡的加权二元交叉熵。其核心目标是在窗口  $t$  内使正负样本对损失的总贡献相同。该权重随窗口内正负边比例自适应变化, 从而避免未加权训练下多数类主导导致的退化, 并提升跨窗口稀疏度变化下的训练稳定性。损失函数在标准二元交叉熵的基础上引入动态权重平衡机制<sup>[29]</sup>, 即:

$$L = \sum_{i=1}^T \sum_{(i,j)} [-\alpha_i^+ A_{ij}^{(t)} \log \hat{A}_{ij}^{(t)} - \alpha_i^- (1 - A_{ij}^{(t)}) \log (1 - \hat{A}_{ij}^{(t)})] \quad (8)$$

式中:  $A_{ij}^{(t)}$  表示真实的邻接关系。权重  $\alpha_i^+$  和  $\alpha_i^-$  根据第  $t$  窗口中正负边比例自适应计算, 即:

$$\begin{cases} \alpha_i^+ = \frac{N_i}{2P_i} \\ \alpha_i^- = \frac{N_i}{2N_i^-} \end{cases} \quad (9)$$

式中:  $P_i$  和  $N_i^-$  分别为正例和负例边数, 总样本数为  $N_i$ 。不同时间窗样本量差异较大时, 损失按窗口样本量加权汇总, 避免某个稀疏或稠密窗口主导优化方向。编码器参数在全部时间窗之间共享, 以提取跨时段一致的结构表达。为提升收敛效率与鲁棒性, 训练过程采用学习率自适应调度结合早停机制。训练完成后, 跨窗口提取每个用户节点的邻接特征。

### 2.3 异常检测模块

在经过 GAE 的处理后, 每个用户均得到一组邻接特征, 这些特征能够反映其在群体结构中的行为模式。但由于原始特征维度较高, 部分维度之间存在较强相关性, 直接使用可能带来冗余与噪声, 降低聚类的判别效果。因此, 本文在异常检测阶段引入了主成分分析 (principal component analysis, PCA) 与 K-Means++ 结合的方法, 对特征进行优选和聚类判别。

K-Means 聚类作为经典的无监督学习算法, 其核心机制是基于相似性度量将数据样本分配至  $k$  个不同的

簇, 并用质心向量表征各簇的中心位置。针对传统 K-Means 初始聚类中心敏感的缺陷, 本文采用了改进的 K-Means++ 算法。该算法采用概率选取策略确定起始聚类中心, 通过最大化初始质心间的空间间距, 有效提升了算法的初始化质量, 降低了收敛至次优解的可能性。相较于随机初始化方案, K-Means++ 在处理复杂数据分布时展现出更优的群体划分能力和结果稳定性<sup>[30]</sup>。

聚类完成后, 考虑到实际电力场景中窃电用户占比低, 将样本量最小的簇定义为异常簇, 其成员标记为异常用户。在 PCA 子空间里, 异常用户个体通常远离主体群体, 表现为孤立点或小规模团簇。其跨窗连接行向量往往呈现结构相似度波动更大、邻居集合更替更频繁或连接强度突变等特征。采用 PCA 结合 K-Means++ 的方式能够在不牺牲主要判别信息的前提下完成去相关与抑噪, 并通过基于距离的高质量初始化配合多次重启获得稳定的原型聚类结果, 从而实现可疑窃电用户稳健的无监督识别。

### 2.4 总体框架

为实现非高损线路场景下窃电用户的异常检测, 本文提出了一种基于互信息图与邻接特征的无监督方法。所提方法从群体关系演化的视角出发, 将用户作为群体网络中的节点, 利用互信息差值图刻画小时尺度上用户用电量与线损电量之间的依赖模式, 并通过图表示学习提取跨时间窗口的邻接特征, 以实现异常用户的识别定位。整体框架流程如图 5 所示。主要的检测流程为:

1) 数据预处理。收集小时尺度的用户用电量序列  $X(t)$  与线损电量序列  $W(t)$ 。针对实际采集中可能存在的缺失记录、量纲差异与噪声干扰, 对原始数据进行清洗与归一化处理, 以保证时序完整性、数值尺度一致性等, 得到规整的输入序列。

2) 互信息差值图构建。采用滑动窗口切分, 在每个窗口内计算各用户用电量与线损电量的互信息。将互信息值相近的用户视为具有相似的线损关联模式, 在图中

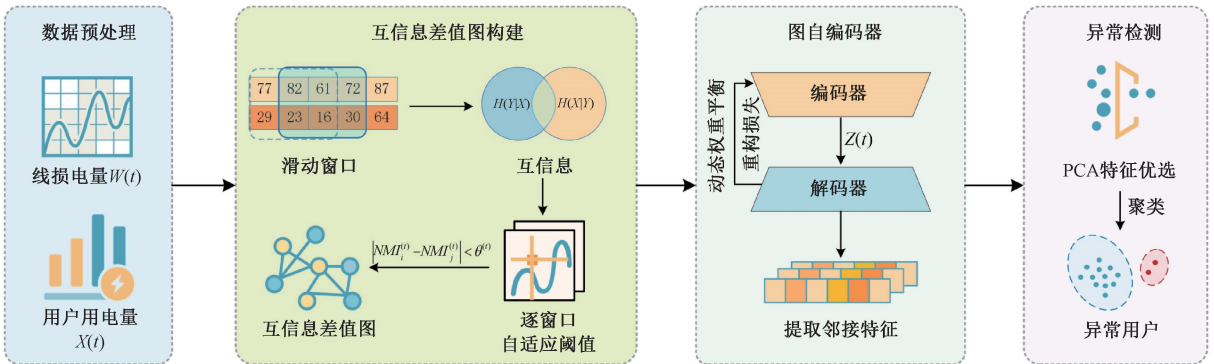


图 5 窃电检测方法的整体框架流程

Fig. 5 Overall framework and workflow of the electricity theft detection method

建立边连接。采用逐窗自适应阈值,避免固定阈值带来的图结构不稳定问题,构建出反映群体依赖关系的互信息差值图序列。

3) GAE 邻接特征提取。采用两层参数共享的 GAE,在编码端获取节点的局部与全局嵌入表示,在解码端以内积方式重建邻接关系。训练过程中引入动态权重平衡机制,对正负边样本赋予自适应权重,以缓解边类不均衡对模型收敛的影响。基于跨窗口的邻接重建结果,进一步提取能够刻画节点关系一致性的邻接特征。

4) 异常检测。对提取的特征采用 PCA 进行特征优选,以去除冗余维度并突出主要判别信息。在此基础上,利用 K-Means++ 聚类实现无监督分类,将聚类结果中孤立点或小规模类簇识别为潜在的窃电用户。模型的参数设置如表 1 所示。

表 1 参数设置  
Table 1 Settings of parameters

模块	参数名称	数值/说明
互信息差值图	滑动窗口长度 $h$	12
	滑动步长 $s$	6
	分位数参数 $p$	第三四分位数
GAE 特征提取	隐藏层	32
	嵌入层	16
	激活函数	ReLU
	优化器	Adam
	学习率	0.01
异常检测	聚类簇数 $K$	2
	K-Means 最大迭代次数	300

## 3 算例分析

### 3.1 仿真实验

为了验证该方法的有效性,首先在某地区正常的中压线路上进行了仿真实验。数据集包含 16 条中压线路 F1~F16,以小时采样频率采集,时间范围为 2023 年 3 月 1 日至 2023 年 5 月 31 日,共计 2 028 个连续小时的数据<sup>[19]</sup>。选择线路 F16 作为本节仿真实验对象。线路 F16 涵盖 34 个工商业用户的用电量记录及相应的线损电量<sup>[19]</sup>。为了模拟真实的窃电场景,从中随机选择 3 个用户,并按 1.2 节所示的策略生成窃电样本。实验中,抽取用户 U16、U29 和 U26 设定为窃电类型 2 的窃电用户,窃电时段占比为 15%、20%、25%。其余用户保持正常。窃电模拟完成后,线路的平均日线损率低于 7%,整体仍处于非高损线路区间。

用户 U1~U34 的小时用电量定义为时间序列  $X_1 \sim X_{34}$ ,相应的线损电量定义为时间序列  $W$ 。本文采用滑动窗口方法,在各时间窗口内计算  $X_i$  与  $W$  的互信息,并据此得到完整的互信息序列。如图 6 所示,正常用户的箱体高度普遍较小,四分位距集中,上四分位数和下四分位数分别分布在  $[2.06, 2.23]$  和  $[1.56, 1.86]$ ,波动性有限。而用户 U16、U26、U29 的箱体明显拉长,离群点变多。上下四分位数分布在  $[2.01, 2.12]$ ,  $[1.21, 1.41]$ 。结果表明,正常用户的互信息值波动范围小,用电量与线损电量之间的依赖关系稳定。而窃电用户的互信息值波动范围更大,用电量与线损电量之间的依赖关系减弱且时序稳定性较差。由 1.3 节分析可知,窃电行为导致用户的实际用电量与计量电量不一致,从而引入额外的 NTL,打破了原有的物理约束关系。窃电用户的部分负荷虽然未被计入计量系统,但仍然对线损产生贡献。导致用户负荷与线损之间的统计依赖关系被削弱,从而互信息值下降。

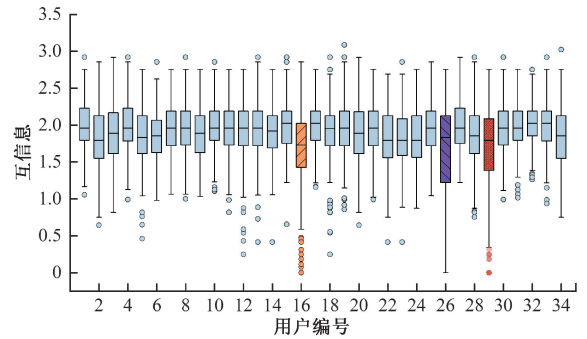


图 6 互信息分布箱线图

Fig. 6 Boxplot of mutual information distribution

基于互信息结果,采用逐窗口分位数法设定阈值,构建一组随时间演化的互信息差值图  $G(t)$ ,用户作为节点  $V = \{v_1, v_2, \dots, v_{34}\}$ 。利用 GAE 模型对图序列进行嵌入学习,提取节点的低维表示,并通过邻接稳定性指标捕捉异常模式。通过 PCA 对嵌入特征进行降维,并采用 K-Means++ 聚类方法对用户进行划分,从而识别潜在异常用户。聚类后的 t 分布随机邻域嵌入 (t-distributed stochastic neighbor embedding, t-SNE) 结果如图 7 所示,其中用户 U16、U26 和 U29 被识别为异常类簇并标记为窃电用户。仿真结果表明,该方法能够准确识别非高损线路中的窃电用户,验证了所提方法的有效性。

### 3.2 消融实验

为进一步评估所提方法的性能,本文基于 1.2 节所示的窃电方法,在 3.1 节数据集 16 条线路下的用户中随机抽取部分用户并注入窃电行为,窃电时段占比在 15%~35%,构建混合攻击模式下的窃电样本,进而生成

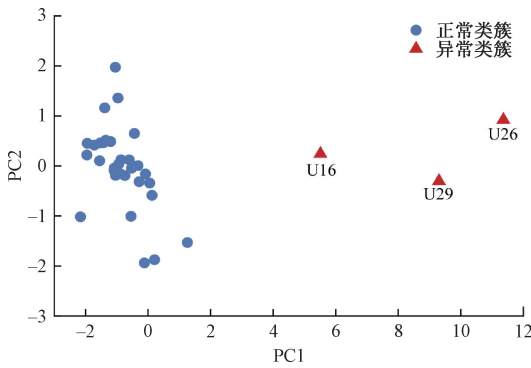


图 7 K-Means++ 聚类结果

Fig. 7 Clustering results of K-Means++

32 个不同的数据集进行实验测试。所有数据集的平均日线损率均控制在 7% 以下, 仍属于非高损线路范围。

消融实验包括 4 种模型。模型 A 采用固定阈值代替自适应分位数阈值策略。模型 B 移除动态平衡机制, 使所有边在损失函数中的权重相同。模型 C 取消 PCA 步骤, 直接对高维嵌入进行聚类。模型 D 采用随机欠采样策略代替动态权重平衡机制, 使正负边样本数量平衡后进行重构学习, 其余设置保持不变。检测性能采用准确率 (accuracy, ACC), 查准率 (precision, PRE), 检出率 (true positive rate, TPR), F1 分数 (F1-score) 和误报率 (false positive rate, FPR) 进行评估<sup>[31]</sup>。

最终结果的平均值如图 8 所示。模型 A 的 F1 分数从 0.933 3 降至 0.772 7, 下降了 17.2%, PRE 从 0.901 5 降至 0.723 4, 下降了 19.8%, FPR 从 0.028 8 上升至 0.086 5。这表明固定阈值无法适应不同时间窗口的数据分布差异, 导致图构建不稳定, 大量正常用户被误判为异常。模型 B 的 F1 分数降至 0.710 1, 下降了 23.9%, TPR 从 0.967 5 降至 0.796 7, 下降了 17.7%, 检出率明显下降。PRE 也降至 0.640 5, FPR 上升至 0.122。结果表明图中正负边样本分布不平衡, 在缺乏权重调节的情况下, 模型对少数类窃电用户的学习能力明显不足。模型 C 的 F1 分数从 0.933 3 降至 0.797 0, 下降了 14.6%, PRE 降至 0.729 7, 下降了 19.1%, TPR 降至 0.878, 下降了 9.3%, FPR 上升至 0.088 7。这表明直接进行聚类会受到高维噪声的干扰, PCA 降维优选能够有效提升聚类的稳定性和判别能力。模型 D 的 F1 分数降至 0.819 9, 下降了 12.2%, PRE 降至 0.775 4, 下降了 14.0%。随机欠采样虽能通过平衡正负边数量缓解不平衡, 但会丢弃部分边样本并改变窗口内邻接分布, 造成结构信息损失, 从而削弱邻接重构与异常表征学习效果。相比之下, 动态权重平衡机制在保留全部样本信息的前提下按窗口自适应调整权重, 实现更稳定的训练与更优的检测性能。综

上所述, 消融实验充分验证了自适应阈值策略、动态权重平衡机制和 PCA 降维优选的有效性。

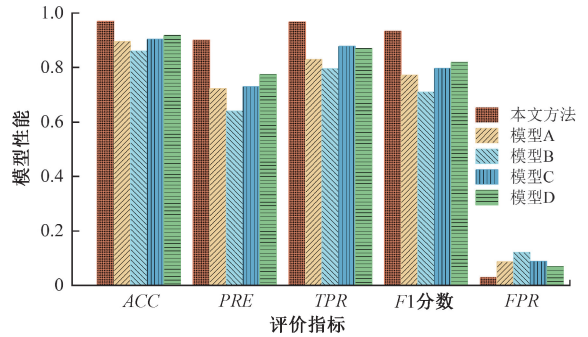


图 8 性能指标结果

Fig. 8 Results of performance indicators

为进一步分析所提方法对不同窃电类型的检测能力, 本文在 32 个数据集上统计了所提方法对各窃电类型的 TPR、PRE 和 F1 分数, 结果如图 9 所示。

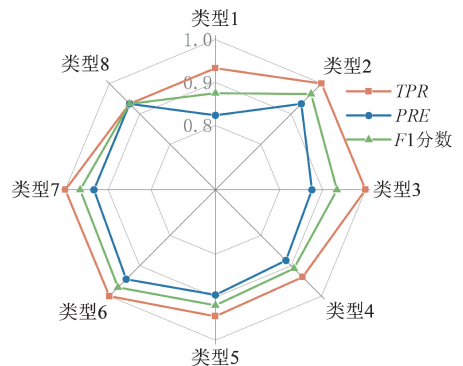


图 9 不同窃电类型的检测效果

Fig. 9 Detection effectiveness of different types of electricity theft

由图 9 可见, 所提方法对 8 种窃电类型均表现出良好的检测性能。具体而言, 类型 2、3、6、7 的检测效果最好, 表明这些窃电行为对用户负荷与线损的依赖关系破坏较为明显, 在互信息图中表现为显著的邻接模式异常, 因此能够被完全识别。类型 5 的 TPR 为 94.44%, 类型 4 的 TPR 为 93.75%, 仍保持在较高水平。类型 1 和类型 8 的 TPR 均为 93.33%, 其中类型 1 的 F1 分数相对较低, 为 87.5%, 这是因为固定比例窃电在窃电比例较小时, 对线损的影响较为微弱, 导致互信息变化不明显, 部分样本被漏检。类型 8 的 TPR 为 93.33%, PRE 达到 93.33%, F1 分数为 93.33%, 验证了基于互信息图与邻接特征的方法能够有效识别隐蔽性窃电行为。这种窃电方式虽然在日均用电量层面难以显现异常, 但由于其破坏了用户用电量与线损之间在小时尺度上的稳定依赖关系, 导致互信息值波动增大且邻接模式不稳定, 从而被成功检出。

### 3.3 对比实验

为了全面评估所提出方法的性能和优势,本节以第3.2节使用的数据集作为检测对象。检测性能采用ACC,PRE,TPR,F1分数和FPR进行评估。对比方法包括模型1基于线损用电量相关性的检测方法<sup>[19]</sup>、模型2基于负荷曲线特征的方法<sup>[32]</sup>和模型3基于深度学习的方法<sup>[33]</sup>进行比较。具体而言,模型1基于格兰杰因果检验,利用用户用电量与线路损耗的时序关系来识别导致线损升高的窃电用户<sup>[7]</sup>。模型2基于用户负荷数据特征,利用局部离群因子度量用户与群体的偏离程度,从而识别出明显异常的窃电行为<sup>[32]</sup>,模型3基于卷积神经网络(convolutional neural network,CNN)与长短期记忆网络(long short-term memory,LSTM),通过CNN提取负荷曲线的局部模式特征,结合LSTM捕捉时序依赖关系,并通过聚类识别异常用户<sup>[33]</sup>。

对比实验结果如图10所示。可以看到,本文方法在核心指标上均优于对比方法。准确率达到0.9704,显著高于模型1的0.9007、模型2的0.8885和模型3的0.9077,表明整体判别能力更强。本文方法的PRE为0.9015,远高于模型1的0.7292、模型2的0.7218和模型3的0.7273,这意味着正常用户被误判的概率显著降低。模型1虽能捕捉时序因果关系,但缺少群体关系建模,PRE仅为0.7292,FPR高达0.0865。模型2基于局部离群因子,仅依赖单用户统计特征,在非高损场景下对隐蔽窃电行为的识别能力不足,PRE为0.7218,FPR达到0.0820,误报率较高。模型3采用CNN-LSTM提取负荷曲线的时空特征,虽能有效学习时序模式,但缺少群体关系约束,PRE为0.7273,且FPR高达0.0931,为所有方法中最高。相比之下,本文方法通过互信息图构建群体网络,在捕捉个体时序特征的同时引入了用户间的依赖关系,能够从群体视角识别邻接稳定性异常,FPR仅为0.0288,显著低于对比方法。综上所述,所提方法通过融合互信息图建模与邻接特征提取,能够在非高损线路场景下有效识别窃电用户,在保持较高准确率和查准率的同时显著降低了误报率。

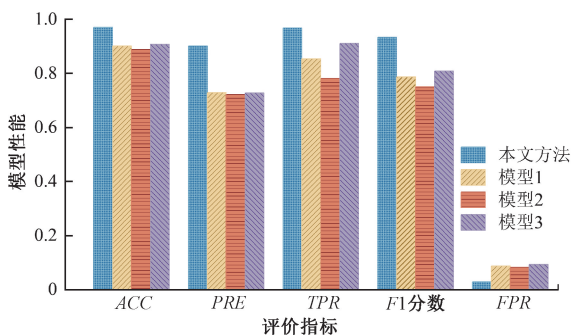


图10 对比模型性能指标结果

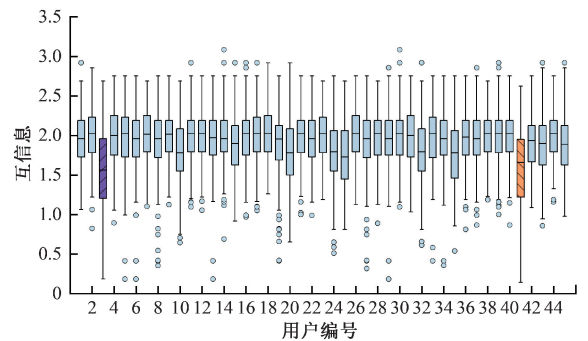
Fig. 10 Performance indicator results of comparative models

### 3.4 真实线路案例

为了进一步验证所提方法的有效性,本文基于真实10 kV中压线路数据开展评估。数据以1 h为采样间隔,共包含3条10 kV中压线路<sup>[19]</sup>。线路A覆盖2024年3月1日至5月31日连续2208 h,包含该线路下45个工商业用户的用电量及线损电量数据,日平均线损率约为3.5%;线路B覆盖2025年6月1日至8月31日连续2208 h,包含该线路下39个工商业用户的用电量及线损电量数据。该线路的日平均线损率约为3.4%;线路C覆盖2025年3月1日至6月1日连续2232 h,包含该线路下42个工商业用户的用电量及线损电量数据。该线路的日平均线损率约为4.4%。在工程实践中这些线路通常被视为正常的非高损线路,并非窃电排查的重点。因小时尺度下线损率存在间歇性升高现象,可采用本文所提方法对这些线路进行检测。

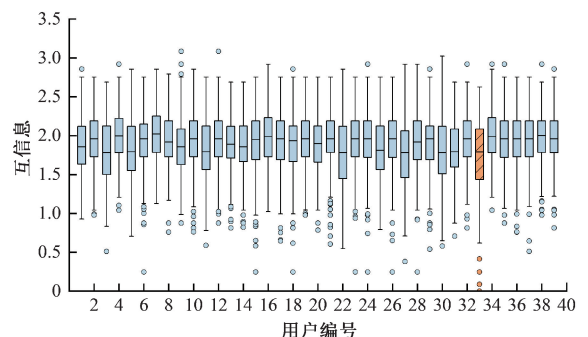
对于每条线路,用户 $U_i$ 的每小时用电量数据定义为时间序列 $X_i$ ,相应的线损电量被定义为时间序列 $W$ 。滑动窗口后在每个窗口内计算 $X_i$ 与 $W$ 的归一化互信息,构建互信息差值图序列 $G(t)$ 。将图序列输入GAE,学习得到每个用户的低维嵌入;随后用PCA保留主要成分,并以K-Means++在嵌入空间完成聚类以标注可疑用户。箱线图 and 聚类的t-SNE可视化如图11和12所示。

如图11(a)所示,线路A中U3和U41的箱体位置低于其他用户。上下四分位数分别为 $[1.93, 1.95]$ , $[1.21, 1.22]$ ,而其他用户的上下四分位数则分布在



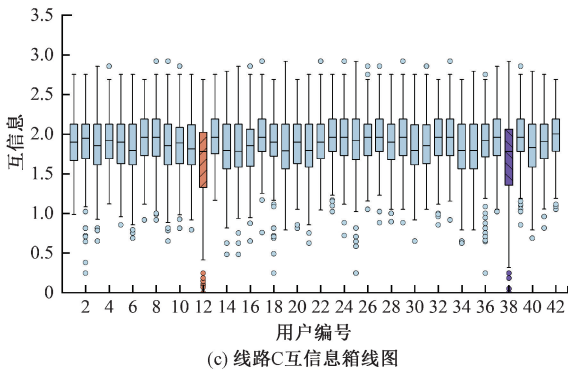
(a) 线路A互信息箱线图

(a) Boxplot of mutual information for line A



(b) 线路B互信息箱线图

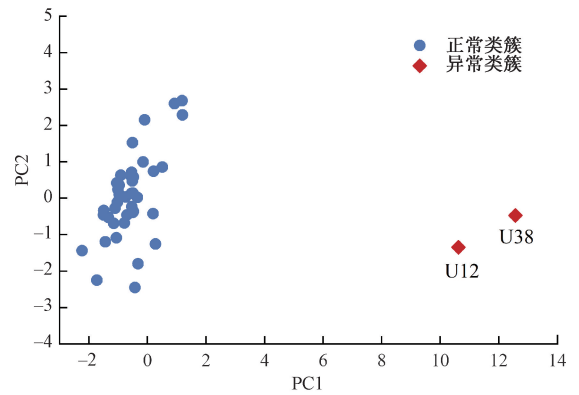
(b) Boxplot of mutual information for line B



(c) 线路C互信息箱线图  
(c) Boxplot of mutual information for line C

图 11 互信息分布箱线图

Fig. 11 Boxplot of mutual information distribution



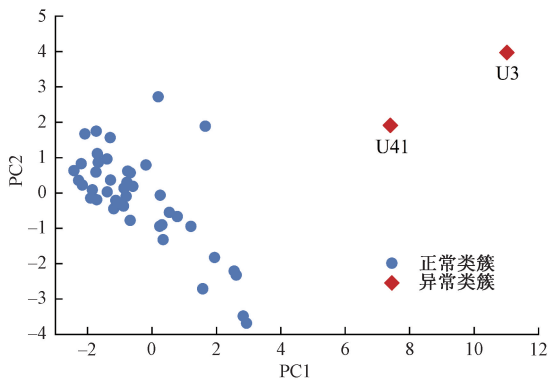
(c) 线路C聚类图  
(c) Clustering result for line C

图 12 K-Means++ 聚类结果

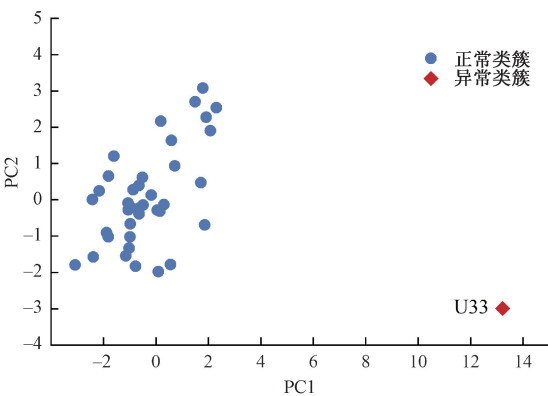
Fig. 12 Clustering results of K-Means++

[2.06, 2.25]、[1.46, 1.79]。如图 12(a)所示,从聚类结果可以观察到,模型将用户划分为两个类簇,异常类簇在低维空间与主体类簇呈明显分离。异常类簇中用户 U3 和 U41 被标记为异常用户。经现场稽查人员确认,用户 U3 和 U41 使用欠流法窃电,致使电表的电流输入回路只通过部分电流,从而导致电量少计。

线路 B 的互信息箱线图如图 11(b)所示, U33 表现为低值离群点增多,反映其与线损响应之间的依赖强度发生弱化或波动。同时, U33 的箱体中位数与四分位距



(a) 线路A聚类图  
(a) Clustering result for line A



(b) 线路B聚类图  
(b) Clustering result for line B

整体仍与某些用户存在较大重叠。如 1.3 节所述,在该线路上仅依赖互信息的难以形成稳定、直观的异常判别依据。因此,有必要进一步基于互信息差值进行图建模,并结合邻接稳定性特征学习增强对弱异常的识别能力。模型聚类结果如图 12(b)所示, U33 被清晰划分为远离主体类簇的异常类簇。经现场稽查人员核实,用户 U33 通过绕越计量装置达到电量少计的目的。实验验证了所提方法对弱偏离异常的增强识别能力。

线路 C 的互信息箱线图如图 11(c)所示, U12 和 U38 相较于其他用户表现出箱体更长以及低值离群点增多的特点,其上下四分位数分别为 [2.02, 2.07] 和 [1.32, 1.35]。如图 12(c)所示,模型将 42 名用户聚类为两个独立的簇。其中,用户 U12 和 U38 构成的异常类簇在特征空间中与主体类簇边界清晰,被识别为异常用户。经现场稽查人员确认,用户 U12 和 U38 在互感器内嵌遥控装置窃电,使计量链路受扰,从而造成电能表少计电量。

## 4 结 论

本研究针对非高损线路窃电检测场景,提出了一种基于互信息图与邻接特征的检测方法。本研究主要工作包括:

1) 进行了非高损线路窃电行为机理分析,指出现有工程实践中以日均线损率作为判据易忽略间歇型窃电行为。通过互信息分析验证了正常用户小时级用电量与线损电量之间依赖关系的稳定性,为后续基于互信息的图构建提供了依据。

2) 提出了基于滑动窗口的互信息差值图构建方法,将用户用电量与线损电量的依赖关系映射为动态图结构,从整体网络视角刻画用户之间的动态一致性与差异性。

3) 设计了融合 GAE 与动态权重平衡机制的特征提取方法,有效捕捉窃电用户在邻接稳定性上的异常表现,并结合 PCA 与 K-Means++ 聚类,在真实数据集上实现了无监督的窃电识别。

本文方法突破了传统以单用户负荷特征为主的检测思路,从群体网络的角度出发,捕捉用户节点之间的邻接特征,在非高损线路场景下实现了窃电用户识别。未来的研究可进一步考虑低压台区的适应性、多源数据的融合,以提升在实际电网中的应用价值。

## 参考文献

[1] 马健,滕召胜,邱伟,等. 基于 IGA-BP 神经网络的智能电能计量设备退化趋势研究[J]. 仪器仪表学报, 2021, 42(11): 123-131.

MA J, TENG ZH SH, QIU W, et al. Research on the degradation trend of smart energy metering equipment based on IGA-BP neural network[J]. Chinese Journal of Scientific Instrument, 2021, 42(11): 123-131.

[2] CUI X Y, LIU SH Y, LIN ZH ZH, et al. Two-step electricity theft detection strategy considering economic return based on convolutional autoencoder and improved regression algorithm [J]. IEEE Transactions on Power Systems, 2022, 37(3): 2346-2359.

[3] 陈静,郑垂锭,李桂敏,等. 考虑行业关联度的工业用户用电异常识别研究[J]. 仪器仪表学报, 2023, 44(4): 72-81.

CHEN J, ZHENG CH D, LI G M, et al. Research on power consumption anomaly identification of industrial users considering industry relevance[J]. Chinese Journal of Scientific Instrument, 2023, 44(4): 72-81.

[4] 陈敏,张逸,邹阳,等. 基于稳健回归和卷积神经网络的中压窃电类型检测方法[J]. 电网技术, 2024, 48(11): 4729-4738.

CHEN M, ZHANG Y, ZOU Y, et al. A medium-voltage stealing type detection method based on robust regression and convolutional neural network [J]. Power System Technology, 2024, 48(11): 4729-4738.

[5] 陈磊,苏华锋,苏盛,等. 低压台区用户表前分支线漏电故障定位方法[J]. 仪器仪表学报, 2025, 46(6): 276-289.

CHEN L, SU H F, SU SH, et al. Leakage fault location method for branch lines before user meters in low voltage distribution system [J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 276-289.

[6] 黄彦钦,余浩,尹钧毅,等. 电力物联网数据传输方案:现状与基于 5G 技术的展望[J]. 电工技术学报, 2021, 36(17): 3581-3593.

HUANG Y Q, YU H, YIN J Y, et al. Data transmission

schemes of power internet of things: Present and outlook based on 5G technology [J]. Transactions of China Electrotechnical Society, 2021, 36(17): 3581-3593.

- [7] ZHENG Y, CHEN F, YANG H M, et al. Edge computing based electricity-theft detection of low-voltage users [J]. Frontiers in Energy Research, 2022, 10: 892541.
- [8] 陈静,王铭海,刘煜寒,等. 基于 Transformer 与单分类支持向量机的窃电时间识别方法[J]. 电网技术, 2025, 49(5): 2109-2118.
- CHEN J, WANG M H, LIU Y H, et al. Electricity theft time identification method based on Transformer and one-class support vector machine[J]. Power System Technology, 2025, 49(5): 2109-2118.
- [9] 刘康,刘鑫,张蓬鹤,等. 基于负荷尖峰特征 LSTM 自编码器的窃电识别方法[J]. 电力系统自动化, 2023, 47(2): 96-104.
- LIU K, LIU X, ZHANG P H, et al. Identification method of electricity theft based on long short-term memory autoencoder with load peak features[J]. Automation of Electric Power Systems, 2023, 47(2): 96-104.
- [10] 刘永光,谭煌,李志鹏. 一种基于电能表误差和窃电分析的线损分层定位方法[J]. 电测与仪表, 2022, 59(9): 188-194.
- LIU Y G, TAN H, LI ZH P. A hierarchical line loss location method based on the analysis of electricity meter error and electricity stealing[J]. Electrical Measurement and Instrumentation, 2022, 59(9): 188-194.
- [11] 游文霞,李清清,杨楠,等. 基于多异学习器融合 Stacking 集成学习的窃电检测[J]. 电力系统自动化, 2022, 46(24): 178-186.
- YOU W X, LI Q Q, YANG N, et al. Electricity theft detection based on multiple different learner fusion by Stacking ensemble learning [J]. Automation of Electric Power Systems, 2022, 46(24): 178-186.
- [12] 董立红,肖纯朗,叶鸥,等. 一种基于 CAEs-LSTM 融合模型的窃电检测方法[J]. 电力系统保护与控制, 2022, 50(21): 118-127.
- DONG L H, XIAO CH L, YE OU, et al. Electricity theft detection method based on a CAEs-LSTM fusion model[J]. Power System Protection and Control, 2022, 50(21): 118-127.
- [13] RAGGI L M R, TRINDADE F C L, CUNHA V C, et al. Non-technical loss identification by using data analytics and customer smart meters [J]. IEEE Transactions on Power Delivery, 2020, 35(6): 2700-2710.
- [14] 陈波冯,李靖东,卢兴见,等. 基于深度学习的图异

- 常检测技术综述[J]. 计算机研究与发展, 2021, 58(7): 1436-1455.
- CHEN B F, LI J D, LU X J, et al. Survey of deep learning based graph anomaly detection methods[J]. Journal of Computer Research and Development, 2021, 58(7): 1436-1455.
- [15] 严盛辉, 陈志德. 基于图自编码器的无监督多变量时间序列异常检测[J]. 计算机系统应用, 2023, 32(5): 308-315.
- YAN SH H, CHEN ZH D. GAE-based unsupervised anomaly detection of multivariable time series[J]. Computer Systems and Applications, 2023, 32(5): 308-315.
- [16] 孔祥玉, 马玉莹, 赵鑫, 等. 基于多阶段递推数据分析的低压台区窃电检测方法[J]. 中国电机工程学报, 2024, 44(15): 5921-5934.
- KONG X Y, MA Y Y, ZHAO X, et al. Detection method of electric theft in low voltage station area based on multi-stage recursive data analysis[J]. Proceedings of the CSEE, 2024, 44(15): 5921-5934.
- [17] LIAO W L, YANG D CH, GE L J, et al. Electricity theft detection in integrated energy systems considering multi-energy loads[J]. International Journal of Electrical Power and Energy Systems, 2025, 164: 110428.
- [18] YORUKOGLU S, NASIBOV F, MUNGAN M, et al. The effect of the types of network topologies on nontechnical losses in secondary electricity distribution systems[J]. IEEE Transactions on Industry Applications, 2016, 52(5): 3631-3643.
- [19] 邹念, 张颖, 苏盛, 等. 基于小时尺度周期特征自编码器的用户窃电识别方法[J]. 电网技术, 2023, 47(6): 2558-2567.
- ZOU N, ZHANG Y, SU SH, et al. User electric theft detection method based on hour scale periodic feature LSTM-transformer[J]. Power System Technology, 2023, 47(6): 2558-2567.
- [20] WANG X Y, XIE H P, TANG L F, et al. Decentralized privacy-preserving electricity theft detection for distribution system operators[J]. IEEE Transactions on Smart Grid, 2024, 15(2): 2179-2190.
- [21] 刘康, 李彬, 薛阳, 等. 基于传递熵密度聚类的用户窃电识别方法[J]. 中国电机工程学报, 2022, 42(20): 7535-7546.
- LIU K, LI B, XUE Y, et al. User electric theft detection method based on transfer entropy density clustering[J]. Proceedings of the CSEE, 2022, 42(20): 7535-7546.
- [22] 郑楷迪, 缪希仁, 林雨润, 等. 基于最大互信息系数的低压光伏台区用户相序辨识方法[J]. 电力自动化设备, 2024, 44(12): 108-114.
- ZHENG K D, MIAO X R, LIN Y R, et al. Phase sequence identification method for users in low-voltage transform district with photovoltaic based on maximal information coefficient[J]. Electric Power Automation Equipment, 2024, 44(12): 108-114.
- [23] 纪泽源, 于潇颖, 付文兴. 基于图神经网络特征点匹配的视觉 SLAM 算法[J]. 仪器仪表学报, 2024, 45(9): 34-43.
- JI Z Y, YU X Y, FU W X. Visual SLAM algorithm based on graph neural network feature point matching[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 34-43.
- [24] ZHU Y Y, ZHANG Y, LIU L B, et al. Hybrid-order representation learning for electricity theft detection[J]. IEEE Transactions on Industrial Informatics, 2023, 19(2): 1248-1259.
- [25] 刘万里, 刘扬, 张学亮. 融合互信息和多特征约束的激光雷达与相机外参标定方法[J]. 仪器仪表学报, 2025, 46(1): 29-41.
- LIU W L, LIU Y, ZHANG X L. Extrinsic calibration of LiDAR and camera through mutual information integrated multi-feature constraints[J]. Chinese Journal of Scientific Instrument, 2025, 46(1): 29-41.
- [26] 盛志文, 王镜毓, 石东源. 基于图自编码器的输电网单线图布局自适应生成方法[J]. 中国电机工程学报, 2022, 42(6): 2133-2145.
- SHENG ZH W, WANG J Y, SHI D Y. An adaptive generation method for transmission network one-line diagram layouts based on graph auto-encoder[J]. Proceedings of the CSEE, 2022, 42(6): 2133-2145.
- [27] 陈洁, 刘斌斌, 赵姝, 等. 基于模块度感知图自编码器的重叠社区发现模型[J]. 清华大学学报(自然科学版), 2024, 64(8): 1319-1329.
- CHEN J, LIU B B, ZHAO SH, et al. Overlapping community detection model based on modularity-aware graph auto-encoder[J]. Journal of Tsinghua University (Science and Technology), 2024, 64(8): 1319-1329.
- [28] 李川, 伍依凡, 杨帅. 不平衡分布的数据驱动故障诊断的研究进展[J]. 仪器仪表学报, 2023, 44(8): 181-197.
- LI CH, WU Y F, YANG SH. Research progress of fault diagnostics driven by imbalanced data distribution[J]. Chinese Journal of Scientific Instrument, 2023, 44(8): 181-197.
- [29] ZHANG L, WU M L, CAI X CH, et al. Joint analysis of acoustic scenes and sound events in multitask learning based on Cross \_ MMOE model and class-balanced

- loss[J]. *IEEE Sensors Journal*, 2024, 24(11): 18082-18089.
- [30] 闵永智, 郝大宇, 王果, 等. 基于深度自适应 K-means++ 算法的电抗器声纹聚类方法[J]. *电力系统保护与控制*, 2025, 53(8): 1-13.  
MIN Y ZH, HAO D Y, WANG G, et al. Reactor voiceprint clustering method based on deep adaptive K-means++ algorithm [J]. *Power System Protection and Control*, 2025, 53(8): 1-13.
- [31] 刘钊瑞, 高云鹏, 郭建波, 等. 基于深度自编码器高斯混合模型的窃电行为检测[J]. *电力系统保护与控制*, 2022, 50(18): 92-102.  
LIU ZH R, GAO Y P, GUO J B, et al. Abnormal detection of electricity theft using a deep autoencoder Gaussian mixture model [J]. *Power System Protection and Control*, 2022, 50(18): 92-102.
- [32] PENG Y L, YANG Y L, XU Y J, et al. Electricity theft detection in AMI based on clustering and local outlier factor[J]. *IEEE Access*, 2021, 9: 107250-107259.
- [33] ASIF M, NAZEER O, JAVAID N, et al. Data augmentation using BiWGAN, feature extraction and classification by hybrid 2DCNN and BiLSTM to detect non-technical losses in smart grids [J]. *IEEE Access*, 2022, 10: 27467-27483.

## 作者简介



**周文晴**, 2019年于武汉理工大学获得硕士学位, 现为长沙理工大学博士研究生, 主要研究方向为配用电大数据分析 with 异常用电检测研究。

E-mail: wenqingzhou2019@163.com

**Zhou Wenqing** received her M. Sc. degree from the Wuhan University of Technology in 2019. She is currently pursuing her Ph. D. degree at the Changsha University of Science and Technology. Her research interests include big data analysis for power distribution and utilization, and abnormal electricity consumption detection.



**杨茂涛**, 2001年于武汉水利电力大学获得学士学位, 2004年于华南理工大学获得硕士学位, 现为国网湖南省电力有限公司供电服务中心(计量中心)副总工程师, 主要研究方向为电力计量、智能量测、用电信息采集等。

E-mail: 496189@qq.com

**Yang Maotao** received his B.Sc. degree from Wuhan University of Hydraulic and Electric Engineering in 2001,

received his M. Sc. degree from South China University of Technology in 2004. He is currently the deputy chief engineer of the Power Supply Service Center (Metrology Center) of State Grid Hunan Electric Power Co., Ltd. His main research interests include power metering, intelligent measurement, and electricity consumption information collection.



**苏盛**, 2002年于武汉大学获得硕士学位, 2009年于华中科技大学获博士学位, 现为长沙理工大学电气与信息工程学院教授, 主要研究方向为基于配用电大数据的用电异常检测。

E-mail: susheng@163.com

**Su Sheng** received his M. Sc. degree from Wuhan University in 2002, his Ph. D. degree from Huazhong University of Science and Technology in 2009. He is currently a professor at Changsha University of Science and Technology. His main research interests include power anomaly detection based on big data of power distribution and utilization.



**赵斌**, 2005年于华北电力大学获得硕士学位, 2012年于华北电力大学获得博士学位, 现为长沙理工大学教授, 研究方向为新能源发电与电能存储、电站设备状态监测与故障诊断、多能互补与微电网。

E-mail: zhaobin@csust.edu.cn

**Zhao Bin** received his M. Sc. and Ph. D. degrees both from North China Electric Power University in 2005 and 2012, respectively. He is currently a professor at Changsha University of Science and Technology. His main research interests include renewable energy generation and storage, condition monitoring and fault diagnosis of power station equipment, and multi-energy complementary systems and microgrids.



**李彬**(通信作者), 2019年于长沙理工大学获得硕士学位, 2025年于长沙理工大学获得博士学位, 现为长沙理工大学电气与信息工程学院讲师, 主要研究方向为配用电大数据分析 and 电力气象灾害分析。

E-mail: 009295@csust.edu.cn

**Li Bin** (Corresponding author) received his M. Sc. degree from Changsha University of Science and Technology in 2019, his Ph. D. degree from Changsha University of Science and Technology in 2025. He is currently a lecturer in the School of Electrical and Information Engineering at Changsha University of Science and Technology. His main research interests include big data analysis of power distribution and utilization and power meteorological disaster analysis.