

DOI: 10.19650/j.cnki.cjsi.J2311874

# 基于时序关联特性的错误接线漏电用户定位方法\*

周凯欣<sup>1</sup>,冯萧飞<sup>1</sup>,苏盛<sup>1,2</sup>,李彬<sup>1</sup>

(1.长沙理工大学电气与信息工程学院 长沙 410014; 2.电网防灾减灾全国重点实验室 长沙 410114)

**摘要:**针对用户零线、地线错误接线故障引起的台区剩余电流超标、剩余电流动作保护器(RCD)投运率低和漏电事故频发的问题,提出基于时序关联特性的错误接线漏电用户定位方法。接入用户的负荷电流与台区剩余电流存在因果关系,正常用户的影响有限,而异常用户负荷电流主导台区剩余电流变化。首先,运用 Apriori 算法挖掘出错误接线时台区剩余电流和故障用户负荷电流呈强关联特性;然后,进一步构造台区剩余电流与各用户负荷电流的自适应 Lasso 回归模型,筛选出不同故障场景下的可疑用户变量;再结合可疑用户的标准化回归系数绝对值大小,可快速识别与台区剩余电流大幅异动有强关联特性的错误接线漏电用户;最后,基于真型配电网实验室数据验证了所提方法的有效性。

**关键词:** 剩余电流;错误接线;漏电定位;Apriori 算法;自适应 Lasso 回归

中图分类号: TN912 TH89 文献标识码: A 国家标准学科分类代码: 510.40

## User location method of Erroneous wiring and leakage electricity based on timing correlation characteristics

Zhou Kaixin<sup>1</sup>, Feng Xiaofei<sup>1</sup>, Su Sheng<sup>1,2</sup>, Li Bin<sup>1</sup>

(1. School of Electrical & Information Engineering, Changsha University of Science and Technology, Changsha 410014, China;

2. National Key Laboratory of Power Grid Disaster Prevention and Mitigation, Changsha 410014, China)

**Abstract:** Aiming at the problems of excessive residual current in the station, low operation rate of residual current device (RCD) and frequent leakage accidents caused by erroneous wiring faults of user neutral and ground wires, an erroneous wiring leakage user location method based on timing correlation characteristics is proposed. There is a causal relationship between the load current of connected users and the residual current of station. It's noted that the influence of normal users is limited, while the load current of abnormal users dominates the change of residual current in the station. First, the Apriori algorithm is used to infer the strong correlation between the residual current in the station area and the load current of faulty when there are erroneous connections; then, the adaptive Lasso regression model is constructed to characterize the relation between residual current of the station area and the load current of each user, correspondingly the suspicious user variables in different fault scenarios can be screened out; By utilizing the additional absolute value of standardized regression coefficient of suspicious users, the erroneous wiring leakage users with strong correlation characteristics with obviously abnormal changes in the residual current in the station can be quickly identified. Finally, the effectiveness of proposed method is verified with the laboratory data of realistic distribution network.

**Keywords:** residual current; erroneous wiring; leakage localization; apriori algorithm; adaptive lasso regression

## 0 引言

用电多发,易造成漏电故障,是造成电气火灾和人身触电事故的重要肇因<sup>[1]</sup>。为保障用电安全,一般要求在台区低压侧、分支箱和用户入户处安装分级剩余电流保护(residual current devices, RCD),通过各级 RCD 的整定配合来实现选择性保护<sup>[2]</sup>。

低压配网线路和用户设备情况复杂,不规范接线的

目前,国内外 RCD 普遍采用幅值比较法进行保护控制,根据主回路中的剩余电流是否达到整定值来判断是否跳闸<sup>[3]</sup>。但运行中除线路与电器设备的浪涌电流和线路对地电容不对称等非故障因素干扰外,重过载、三相不平衡和谐波成分较高等情况也会造成阈值设置较小的用户三级 RCD 频繁误动<sup>[4-5]</sup>。因此,部分低压用户未正确配置和使用三级 RCD,发生故障时容易造成分支箱二级乃至台区一级 RCD 越级动作。由于低压系统接线复杂且用户并不一定规范配置和使用三级 RCD,低压漏电故障往往很难及时定位排查。一、二级 RCD 频繁跳闸时,为保障用户供电往往不得不退出 RCDs。对福建某地的调研表明,当地台区一级 RCD 安装率为 95.52%,但因频繁误动无法投运的比例高达 72.85%<sup>[6]</sup>,极大地增加了人身触电的风险。

近年来,国内外针对 RCD 投运困难的问题,以避免 RCD 误动、拒动为目标,从工作机制出发深入分析了绝缘故障漏电电弧的频谱特征,围绕利用该特征来识别触电类型等方面开展了大量研究。文献[7-10]针对如何准确快速识别触电故障类型的问题,分别提取分析了生物触电剩余电流的时频奇异谱特征、固有模态能量特征和时域频域特征,为提高 RCD 的灵敏性、降低人身触电伤害后果提供了理论支撑。从 RCD 保护控制算法的改进设计层面,文献[11-13]分别引入负荷电流和三相不平衡等信息,来屏蔽台区固有剩余电流等干扰因素的影响,能满足多级 RCD 的灵敏度要求,提高系统可靠性。

工程实际中,除绝缘故障外,用户零线、地线错误接线也是造成 RCD 频繁动作、无法投运的常见原因。错误接线用户用电时其负荷电流将全部转化为台区剩余电流,会引起台区剩余电流大幅异动且远超 RCD 所设阈值。用户错误接线隐蔽性强且不影响用电,很难及时准确定位,而持续通流的负荷电流还会烧损地线接头、造成接地电阻异常升高,使与该地线相连所有用户的电器导电外壳带电,显著放大触电伤亡事故的风险,当前国内外对如何定位该类故障尚缺乏必要的研究。

本文针对零线、地线错误接线用户漏电定位难题,利用台区剩余电流和用户负荷电流计量数据,深入分析两者的波动特性,并通过 Apriori 算法进行关联规则挖掘,在此基础上采用自适应 Lasso 回归建立最优模型,准确刻画用户负荷电流和台区剩余电流之间的函数关系,解决电力用户数据集存在的多重共线性和模型过拟合的问题。通过定位错误接线漏电用户,提高台区一级、二级 RCDs 的投运率。

### 1 错误接线危害机制与机理分析

#### 1.1 故障危害机制

低压居民供电按接地和保护方式的不同主要可分为 TT 系统、TN-S 系统和 TN-C-S 系统,3 种供电系统的分级 RCD 配置示意如图 1。

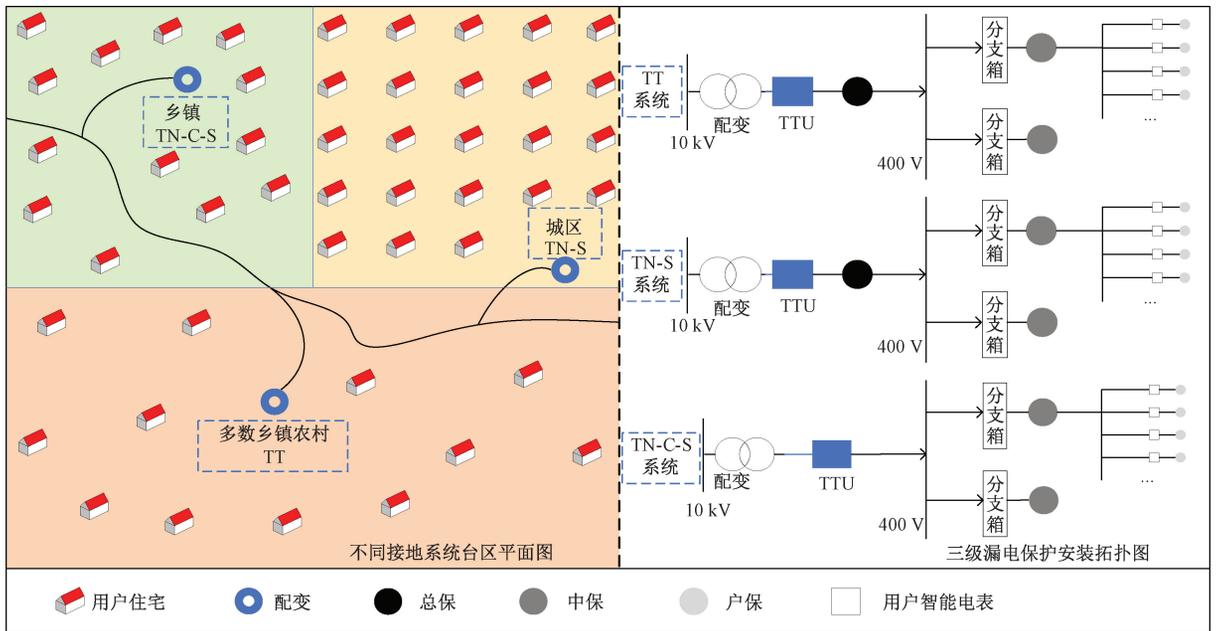


图 1 不同接地系统分级 RCD 安装拓扑图

Fig. 1 Installation topology diagram of hierarchical RCDs in various earth systems

1) 接地系统为三相四线制供电,配变电源中性点直接接地,电器外壳独立直接接地,多用于农村居民。其

中,一户/多户的保护接地(protective earth, PE)可共用同一 PE。

2) TN-S 接地系统为三相五线制供电,配变电源中性点直接接地,从接地点引出 N 线和 PE 线至用电设备处,并要求 N 线和 PE 线严格分开,多用于城市用户。

3) TN-C-S 接地系统为部分三相五线制供电,配变电源中性点直接接地,在电源首端采用三相四线制,到用电负荷附近将 N 线分开形成单独的 N 线和 PE 线。因为 N 线在电源首端与 PE 线相连,不能装设一级 RCD,只能配置二级和三级 RCD。

在上述 3 种接地系统中,错误接线用户的负荷电流随其生活习惯有规律地波动变化,用户用电低谷期大功率电器多处于待机或关闭状态,此时其接入的台区如无其他漏电故障则台区剩余电流会在 300 mA 以下;错误接线用户用电开启大功率电器时,分支线和台区剩余电流将会远超 300 mA。若错误接线用户未投用三级 RCD,将导致一级和二级 RCD 频繁动作。

从防护台区触电伤亡事故的完整流程来看,保障分级 RCD 正常投运、RCD 正确动作隔离故障和发生漏电故障时接地保护线将电器外壳电压控制在安全水平,是防护触电伤亡事故、保障用电安全的三道防线。以 TT 系统为例,零线、地线错误接线故障危害机制示意图如图 2 所示。由于故障用户的负荷电流  $I_d$  经保护地线流入大地,其流经途径为“相线—大地—变压器中性点”,保护地线长期通流较大负荷电流会烧损接头、造成接地电阻上升,使得接入该地线所有用户的电器外壳带上危险电压  $U_{touch}$ 。由于电器外壳带电电源来自保护地线,即便受影响用户正确配置了三级 RCD 也起不到保护作用,将穿透破坏触电保护的三道防线,任一用户错误接线均可能导致接入同一保护地线的全部用户电器外壳带电,显著放大触电伤亡事故风险。

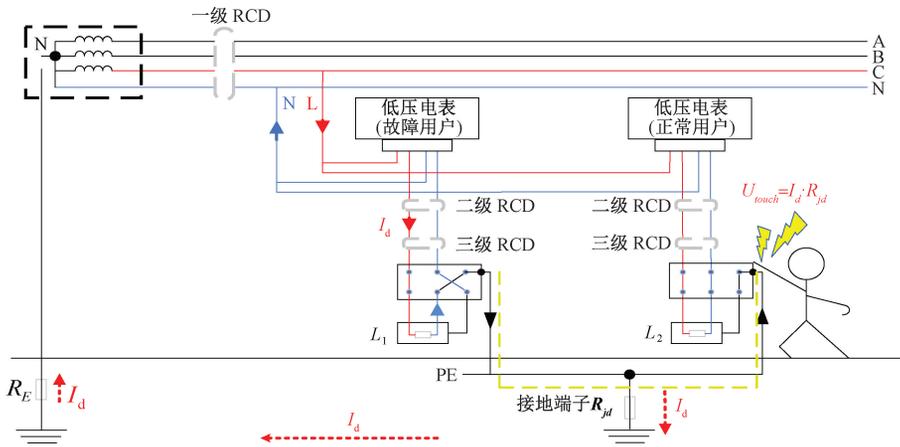


图 2 零线、地线错误接线用户危害机制示意图

Fig. 2 Installation topology diagram of three-level residual current operated protective device of each system

1.2 故障机理分析

一级 RCD 采集装置的工作原理示意图如图 3 所示,其根据传感器采集的相线和零线电流总和测量剩余电流,即:

$$i_{re} = i_A + i_B + i_C + i_N \quad (1)$$

式中:  $i_A$ 、 $i_B$ 、 $i_C$  分别为 A、B、C 三相主干线路上的电流数值;  $i_N$  为零线上的电流数值。

当用户接线正确时,用户负荷电流将经零线流回,由基尔霍夫定律可得:

$$-i_N = i_a + i_b + i_c \quad (2)$$

由式(1)、(2)可得:

$$i_{re} = i_A + i_B + i_C - (i_a + i_b + i_c) \quad (3)$$

用户零线、地线错误接线时,异常用户的用电回路都将由正常的“一相一零”变成“一相一地”。相线经用户负载接地,电流经保护地线并通过大地与变压器接地装

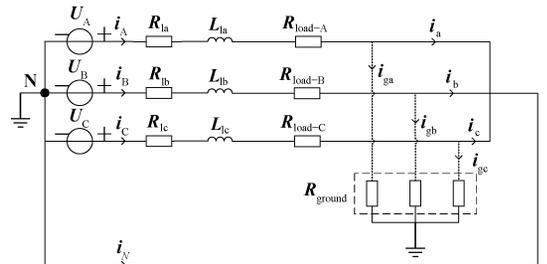


图 3 剩余电流采集装置的工作原理

Fig. 3 Schematic of the working principle of RCD

置构成回路,此时通过基尔霍夫定律得:

$$i_A = i_a + i_{ga} \quad (4)$$

$$i_B = i_b + i_{gb} \quad (5)$$

$$i_C = i_c + i_{gc} \quad (6)$$

式中:  $i_a$ 、 $i_b$ 、 $i_c$  为流经三相上所接用户负载电流;  $i_{ga}$ 、 $i_{gb}$ 、 $i_{gc}$  为流经三相所接保护地线的电流。

由式(3)~(6)可得:

$$i_{re} = i_{ga} + i_{gb} + i_{gc} \quad (7)$$

由式(7)可知,零线、地线错误接线用户的负荷电流不再经零线流回,全部转化为台区剩余电流,引起台区剩余电流超标,使得一级、二级 RCD 无法投运。

## 2 基于 Apriori 算法的时序关联特性分析

### 2.1 台区剩余电流变化趋势分析

在台区和分支线进行剩余电流监测,可在台区计量数据基础上极大丰富可用数据,通过领域知识与多源监测数据的融合,实现错误接线用户定位。结合存在零线、地线错误接线的台区剩余电流及异常用户的负荷电流数据,对正常时和出现零线、地线错误接线用户时台区剩余电流进行波动分析。图4所示为在有、无错误接线用户前后两天按15 min 间隔采集的台区剩余电流曲线和该用户的负荷电流曲线,竖直虚线左边代表台区正常,其中小框子图为自然剩余电流波动的局部放大图,竖线右边代表有错误接线用户接入。

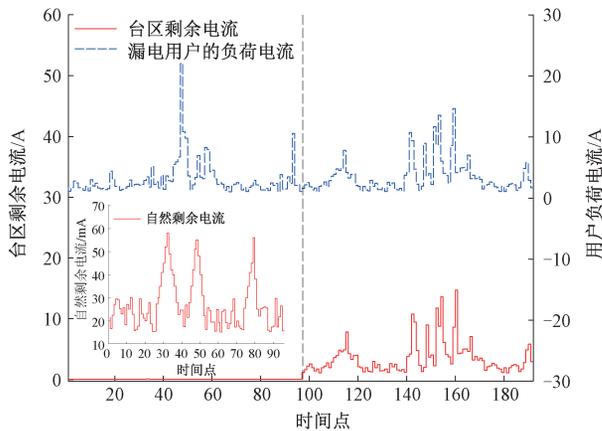


图4 台区剩余电流故障前后对比

Fig. 4 Comparison of residual current before and after the fault in station area

由图4可知,正常无漏电故障台区也存在微小剩余电流,但数值不会超过一、二级 RCD 动作值、造成跳闸。台区存在零线、地线错误接线用户时,台区剩余电流会随着故障用户的用电情况大幅度波动,远超 RCD 动作值,造成一级、二级 RCD 频繁跳闸。

将台区中单个用户、两个同相用户和两个异相用户错误接线故障时的台区剩余电流和异常用户负荷电流如图5~7所示。分析波动曲线可知,当台区存在零线、地线错误接线用户时,异常用户负荷电流与台区剩余电流的幅值变化波形拟合程度高,峰值点所对应的时间计量点也极为相似,两者的波动特性一致。以下引入 Apriori 算法挖掘它们之间的相互依存性和关联特性。

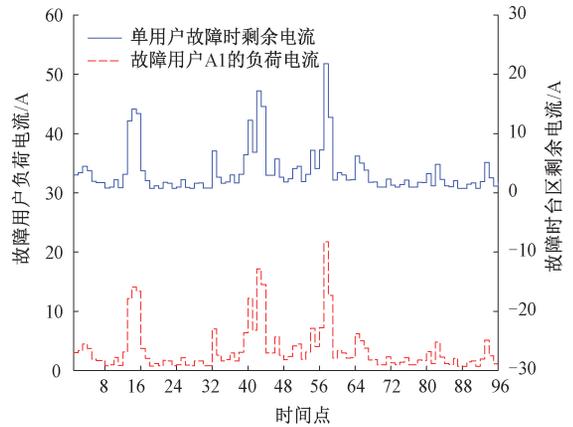


图5 单用户故障时其负荷电流与台区剩余电流  
Fig. 5 A1 Load current and residual current of substation area when a single user fails

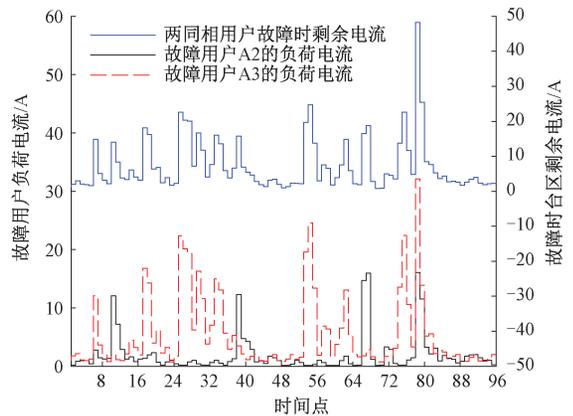


图6 两同相用户故障时其负荷电流与台区剩余电流  
Fig. 6 A2 Load current and residual current of substation area when two in-phase users fail

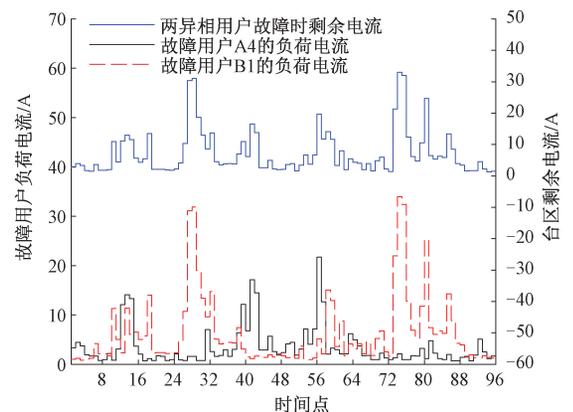


图7 两异相用户故障时其负荷电流与台区剩余电流  
Fig. 7 Load current and residual current of substation area when two out-of-phase users fail

### 2.2 基于 Apriori 算法的剩余电流关联分析

Apriori 算法是关联规则挖掘中经典的数据分析算法,其基本思路是扫描整个数据集  $C_1$  确定其中各项的最小支持度  $S_{\min}$  和最小置信度  $C_{\min}$ ,基于阈值对候选集  $L_h$  进行剪枝,提取不同项集长度候选集  $L_h$  中的频繁项集  $C_h$ ,并结合支持度、置信度、提升度以及确信度来进一步挖掘  $C_h$  中的强关联规则<sup>[14]</sup>。算法迭代过程如图 8 所示,其中  $A \sim D$  为关联规则的项集。

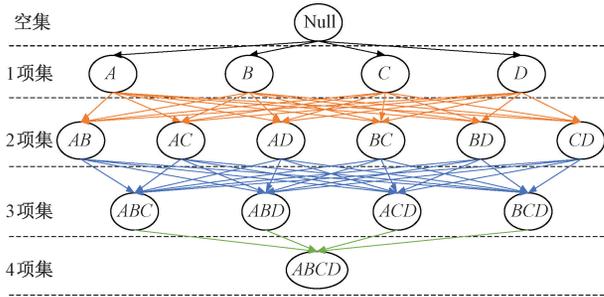


图 8 Apriori 算法基本迭代示意图

Fig. 8 Schematic diagram of the basic iteration of the Apriori algorithm

其中,两子项集之间的关联规则  $R$  可表示为:

$$R: A \Rightarrow B \quad (8)$$

式中: $A$  代表关联规则  $R$  的条件项集 LHS,  $B$  代表关联规则  $R$  的结果项集 RHS。支持度 Support 可衡量各项集在数据集中的频繁共现程度,规则  $R$  的支持度计算公式如下:

$$Support(A \Rightarrow B) = \frac{Count(A \cup B)}{N} \quad (9)$$

式中: $Count(A \cup B)$  为数据集中  $A$  与  $B$  同时出现在一条事务记录中的次数; $N$  为数据集中事务的总数。

置信度 Confidence 可衡量关联规则的可信程度,用来判断规则是否具有实际意义,其计算公式如下:

$$Confidence(A \Rightarrow B) = \frac{Count(A \cup B)}{Count(A)} \quad (10)$$

式中: $Count(A)$  为数据集中涵盖  $A$  的事务数。

提升度 Lift 可衡量结论项集是否依赖于前提项集,反映了  $A$  的出现对  $B$  出现的影响程度,为关联规则中的相关性度量指标,其计算公式如下:

$$Lift(A \Rightarrow B) = \frac{Confidence(A \Rightarrow B)}{Support(B)} \quad (11)$$

若提升度  $>1$ ,则  $A$  和  $B$  呈正相关;若提升度  $=1$ ,则  $A$  和  $B$  相互独立,不存在相关性;若提升度  $<1$ ,则  $A$  和  $B$  呈负相关。

确信度 (Conviction) 可衡量关联规则的确信程度,反映事务  $A$  出现而事务  $B$  不出现的概率,其计算公式

如下:

$$Conviction(A \Rightarrow B) = \frac{1 - Support(B)}{1 - Confidence(A \Rightarrow B)} \quad (12)$$

若确信度  $>1$ ,表示  $B$  越独立于  $A$ ,即关联规则是更加可信的;若确信度  $=1$ ,表示  $B$  与  $A$  的出现与否并没有影响;若确信度  $<1$ ,表示  $B$  与  $A$  是相关联的,但相关性可能是负向的。

### 3 自适应 Lasso 回归分析模型

通过 Apriori 算法分析可以分析用户对台区剩余电流变化的因果关系,找到具有强关联规则的嫌疑用户。但台区接入大量用户、特别是存在多用户错误接线时,可能仅根据强弱关联规则难以精准定位异常用户,以下采用合适的回归模型进行定量分析,准确衡量接入用户对台区剩余电流异动的贡献程度。

#### 3.1 回归分析

回归分析是用来定量分析两个及两个以上的变量相关性的统计方法,常用于预测分析时间序列模型。由理论值和实际漏电台区数值可知,错误接线用户对台区剩余电流贡献较大,反之正常用户的较小。据此,可将台区剩余电流作为结果变量,用户负荷电流作为解释变量,构造线性回归模型,利用回归系数的绝对值大小来表征用户的贡献程度。设基于漏电台区背景的经典多元线性回归模型为:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_m X_{m,t} + \varepsilon_t \quad (13)$$

式中: $Y_t$  对应  $t$  时刻台区剩余电流计量值; $X_{m,t}$  对应台区第  $m$  个用户在  $t$  时刻的负荷电流计量值; $\beta_0$  是台区在  $t$  时刻的计量扰动量; $\beta_1, \beta_2, \dots, \beta_p$  是各用户在  $t$  时刻的异常嫌疑系数; $\varepsilon_t$  为误差项。为了满足线性回归的一致性,对于误差项  $\varepsilon_t$  引入两个重要假定:误差项的期望值等于 0 以及误差项与解释变量不相关。

为了保证回归模型的有效性和稳定性,除引入误差项假设外,还需考虑实际台区对回归模型的影响。低压台区接入用户数较多,而且用电行为习惯较为相似,用户负荷之间的相关性可能使回归模型难以估计准确,造成多重共线性。此外,由于目前所公开的可用电力数据集较为有限,将很有可能会出现当前分析得到的参数估计模型仅仅适用于当前数据集的过拟合现象,此时模型的泛化能力和鲁棒性严重劣化。

文献[15]采用逐步回归法剔除变量来获得有效变量集,进而解决模型多重共线性问题,但 Serang 等<sup>[16]</sup>指出逐步回归方法违背了回归分析的假设前提,在研究者的理论假设并不明确的背景下,采用多重检验修正的验证性方法(如逐步回归)从理论上讲是错误的,在剔除变量之后纳入的变量在这种情况下常常会因为与之前的变量存在相关而被削弱影响,并在一定程度上加剧模型

过拟合问题。文献[17]采用岭回归分析放弃最小二乘法的无偏性,以损失部分信息、降低精度为代价来消除多重共线性,并能够缓解过拟合问题,但难以将冗余预测变量的系数压缩为0,且容易过度压缩较重要的回归系数,不能很好地贴合于存在错误接线用户的台区实际情况。

### 3.2 自适应 Lasso 回归

在利用回归分析构造时间序列分析模型以及寻找各变量之间的因果关系时,需要一个标准来判别求解参数  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  是否为最佳。最小二乘法 (ordinary least square, OLS) 是数据分析中常用的模型系数估计方法,通过最小化损失函数来估计回归模型中的参数。OLS 的系数估计为:

$$\tilde{\beta}_o(n) = \operatorname{argmin} \sum_{i=1}^n (y_i - \beta_0 - \sum_{k=1}^m \beta_k X_{k,i})^2 \quad (14)$$

OLS 的核心是对当前数据集进行无偏估计,易造成模型过拟合现象,在共线性较高和解释变量较多的电力用户数据中更为严重,将难以在台区中准确识别电表零线、地线错误接线用户。

为了弥补 OLS 分析的缺陷,可采用正规化来减少模型估计参数数量,目前有两种主流正规化方法,  $L_1$  正规化 (Lasso 回归) 和  $L_2$  正规化 (岭回归)。加入  $L_1$  正规项和  $L_2$  正规项后的参数估计为<sup>[18]</sup>:

$$\tilde{\beta}_{en}(n) = \tilde{\beta}_o(n) + \lambda_1 [\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2] \quad (15)$$

式中:  $\lambda_1$  为调整参数,表征模型对标准化回归系数的惩罚强度;  $\alpha$  为超参数,当  $\alpha = 0$  时,模型为岭回归,当  $\alpha = 1$  时,模型为 Lasso 回归;  $\|\beta\|_1$  为  $L_1$  正规项,其值为  $\sum_i |\beta_i|$ ;  $\|\beta\|_2^2$  为  $L_2$  正规项,其值为  $\sqrt{\sum_i \beta_i^2}$ 。

$L_1$  正规化函数和  $L_2$  正规化函数的二维平面示意图如图9所示, loss 函数的最优解出现在图中阴影区域和等值线的交点处。当惩罚力度不断增大时 loss 函数等值线将不断向坐标轴扩散,  $L_1$  正规化的交点多出现在坐标轴上,即大多数变量的标准化回归系数将为0,而  $L_2$  正规化的交点多出现在象限内,即大多数变量的标准化回归系数非0<sup>[18-19]</sup>。

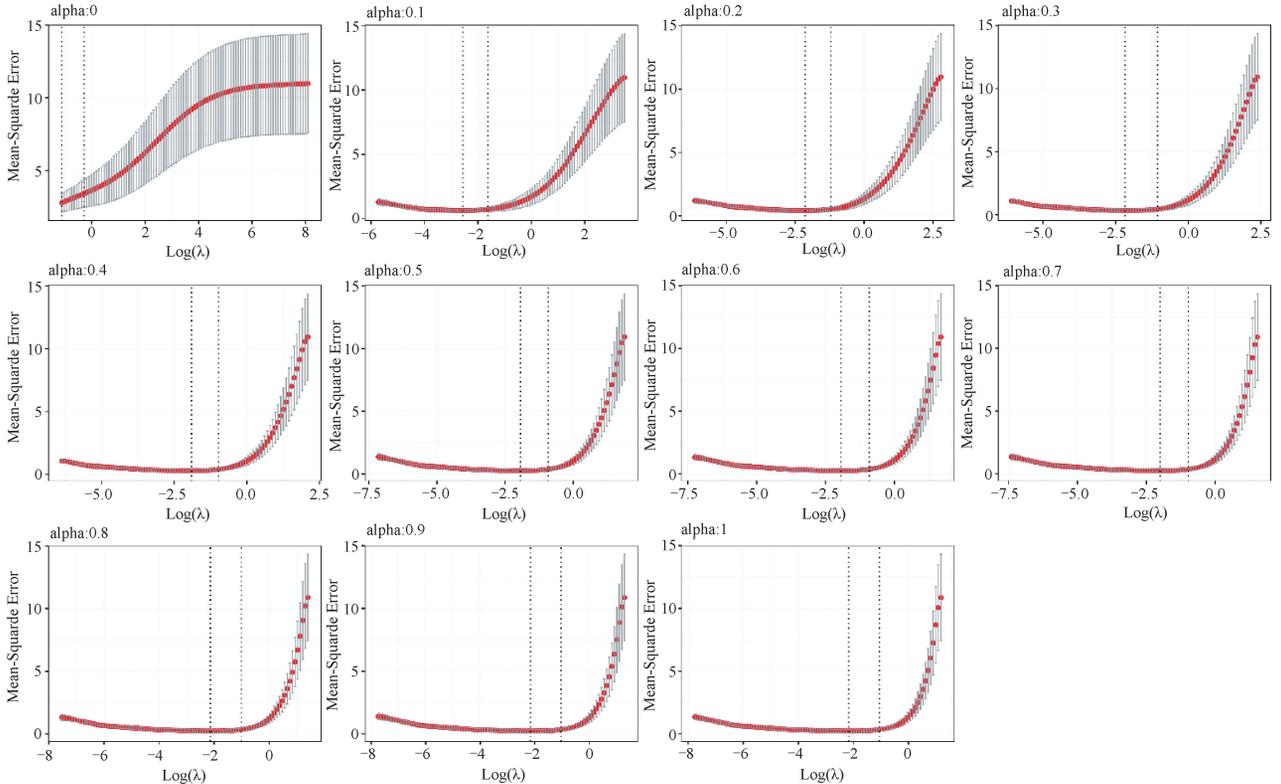


图9 单用户故障时不同模型的交叉验证结果

Fig. 9 Cross-validation results of different models when a single user fails

在  $L_1$  正规化的基础上,自适应 Lasso 回归对不同解释变量应用不同的惩罚力度,得到更为精简模型以更好的推广到其他数据集,减小算力要求的同时还具有好的模型概化和泛化能力<sup>[20]</sup>。自适应 Lasso 回归的参数估

计为:

$$\tilde{\beta}_{adj} = \tilde{\beta}_o(n) + \lambda_2 \sum_{k=1}^p \hat{\omega}_k^{(n)} |\beta_k| \quad (16)$$

式中:  $\lambda_2$  为调整参数,取值范围为  $(0, +\infty]$ ,用于控制

回归系数压缩程度,数值越大就表明惩罚力度越强;  
 $\hat{\omega}_k^{(n)} = |\hat{\beta}_k^{(n)}|^{-r_1}$ ,表示系数  $\beta$  的权重。

### 3.3 时序关联特性的错误接线漏电用户定位流程

以台区剩余电流与接入用户计量数据为例,定位错误接线用户的整体流程如下。

1) 对剩余电流及接入用户计量数据进行预处理;处理后的数据整段剩余电流超标时段用户负荷电流数据定义为时间序列  $X$ ,剩余电流数据定义为  $Y$ 。

2) 通过 Apriori 算法进行关联特性分析,挖掘用户与剩余电流之间的关联规则;若异常用户负荷电流与剩余电流存在强关联规则,表明错误接线用户是造成剩余电流  $Y$  的因,两者存在较强的关联特性,反之则不存在较强的关联特性。

3) 进一步对用户负荷电流与剩余电流构建自适应 Lasso 回归模型,以定量分析接入用户对台区剩余电流的贡献程度。利用回归分析得出的异常嫌疑系数,将数值明显偏大的用户识别为与台区剩余电流超标有强因果关联特性的错误接线嫌疑用户。

## 4 算例分析

为验证方法有效性,在用电异常实验平台搭建由 80 个用户构成的低压台区,模拟用户错误接线漏电故障场景,以 15 min 为间隔采集台区剩余电流数据以及用户负荷电流数据。

### 4.1 剩余电流关联分析算例

考虑到 Apriori 算法通过迭代生成候选项集并剪枝来找频繁项集的过程中,将随机选定关联规则的前后件,可能产生大量不符合预期结果或不具有实际意义的关联规则。为了避免以上情况,研究随机选取 1 个正常用户和 1 个异常用户作为 1 项集,进行剩余电流与用户之间的关联规则挖掘,结果如表 1 所示。

表 1 关联规则挖掘结果

Table 1 Association rule mining results

序号	关联规则	支持度	置信度	提升度	确信度
1	(正常用户)→(剩余电流)	0.697 9	0.697 9	1.00	1.35
2	(剩余电流)→(正常用户)	0.697 9	1.000 0	1.00	$\infty$
3	(异常用户)→(剩余电流)	0.687 5	1.000 0	1.43	$\infty$
4	(剩余电流)→(异常用户)	0.687 5	0.985 1	1.43	20.94

由表 1 可知,第 1 条和第 2 条关联规则的提升度均等于 1.00,表示正常用户与剩余电流之间不存在关联特性;确信度均大于 1.00,表明关联规则是可靠可信的。第 3 条和第 4 条关联规则结果中,约 68.75% 的事务中包含条件项集,并且这些事务中几乎包含全部的结果项集。

同时,提升度大于 1.00,确信度大于 1.00,表示“剩余电流”和“异常用户”之间存在具有实际意义的强关联规则。

### 4.2 错误接线漏电用户定位算例

为验证方法的工程适用性,随机抹除 500 个计量点来模拟数据缺失的情况,并在单用户故障时、两同相用户故障时和两异相用户故障时的 3 个场景下开展定位。将台区异常当天的各用户负荷电流数据定义为解释变量  $X_1 \sim X_{80}$ ,将台区剩余电流定义为结果变量  $Y$ ,建立自适应 Lasso 回归模型来定位故障用户。

#### 1) 单个用户错误接线

随机选取台区用户 2 使其发生零线、地线错误接线故障,并记录在此条件下持续 1 d 的 15 min 间隔计量数据。为了验证在该背景下自适应 Lasso 回归模型为最佳,对每个  $\alpha$  值进行十折交叉验证,过程示意如图 10 所示,计算结果如图 11 所示。其中当  $\alpha = 1$  时模型所对应的均方误差最小,即自适应 Lasso 回归模型在此背景下最为合适。

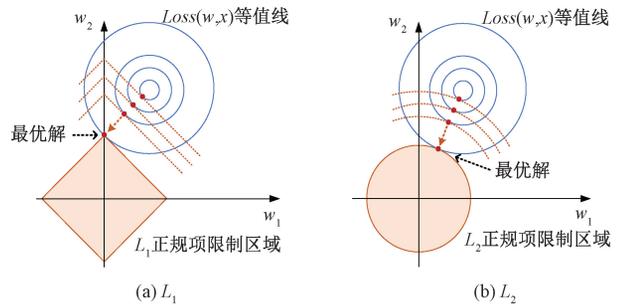


图 10  $L_1, L_2$  正则化效果对比

Fig. 10 Comparisons chart of  $L_1$  and  $L_2$  regularization effects

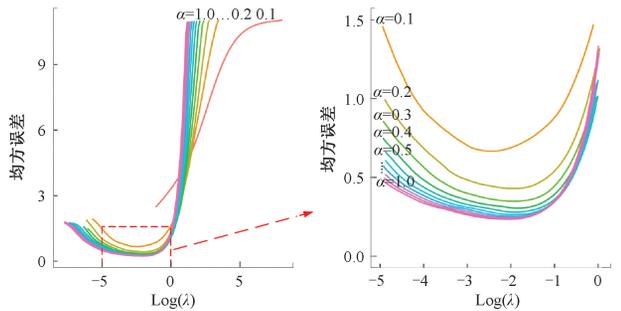


图 11 单用户故障时各  $\alpha$  值的比较结果

Fig. 11 The comparison results of each  $\alpha$  values when a single user fails

挑选  $\alpha = 1$  的十折交叉验证结果进行分析,如图 12 所示。比对均方误差 (mean-squared error, MSE) 大小,可确定最佳调整参数  $\lambda$  值为 0.114 3。

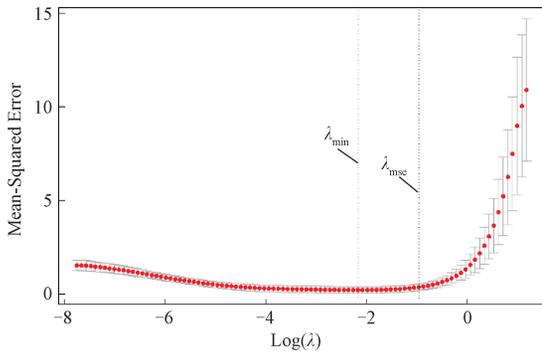


图 12 单用户故障时交叉验证结果

Fig. 12 Cross-validation when a single user fails

图 13 所示为单用户错误接线时  $\log(\lambda)$  与各解释变量的标准化回归系数的关系,图中错误接线用户 2 的嫌疑系数变化曲线用  $X_2$  标注,其他未标注曲线为正常用户。随着  $\log(\lambda)$  的不断增大,即惩罚力度的不断增大,标准化回归系数最终会被压缩至 0,当调整参数  $\lambda$  值为 0.114 3 时,即  $\log(\lambda) = -2.168 6$  时,存在 4 个标准化回归系数非 0 的变量,输出结果如表 2 所示。

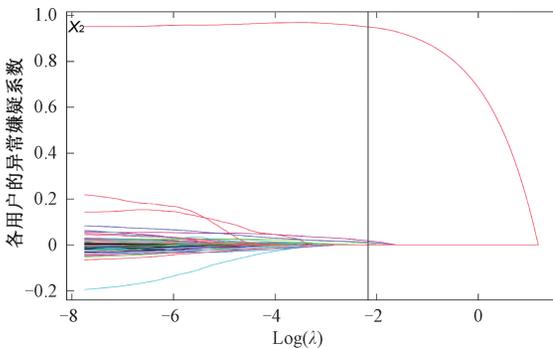


图 13 单用户故障时自适应 Lasso 分析

Fig. 13 Adaptive Lasso analysis diagram when a single user fails

表 2 单用户故障时各用户的异常嫌疑系数

Table 2 Abnormal suspect coefficients when a single user fails

Variable	Coe	Variable	Coe
用户 2	<b>0.950 4</b>	用户 57	0.005 4
用户 12	0.017 6	用户 79	0.010 8
常量	0.063 6	其他用户	*

注: \* 表示其回归系数被压缩至 0

由表 2 可知,解释变量  $X_2$ 、 $X_{12}$ 、 $X_{57}$  和  $X_{79}$  被保留了下来,即自适应 Lasso 回归避免了在压缩不重要变量系数的同时对重要变量(用户 2)的压缩。根据筛选后各变量的异常嫌疑系数可知,用户 2 对台区剩余电流异常的贡献最大,可初步判断用户 2 存在错误接线故障,该结果与

设定一致。为了保证模型的有效性和可靠性,对该情况下的自适应 Lasso 回归模型进行拟合优度检验,如表 3,其中  $R^2$  的取值范围为 0~1,越接近 1 表示拟合程度越好。调整后  $R^2$  为 0.983 2,估计后的标准误差为 0.917 2,表明模型拟合效果好,可解释大部分的因变量变化。

表 3 单用户故障时的拟合优度检验

Table 3 Goodness of fit test when a single user fails

样本量	$R^2$	调整后 $R^2$	估计的标准误差
96	0.983 9	0.983 2	0.789 4

2) 两同相用户错误接线

随机选取台区 A 相用户 18 和 A 相用户 10 发生故障,同样记录在此条件下持续 1 d 的 15 min 间隔计量数据。由图 14~16 可知,该场景下自适应 Lasso 回归的最佳调整参数  $\lambda$  值为 0.285 6。

图 17 呈现了两同相用户故障时  $\log(\lambda)$  与各解释变量的标准化回归系数的关系,其中错误接线用户 10 和用户 18 的嫌疑系数变化曲线用  $X_{10}$  和  $X_{18}$  进行标注,其他未标注曲线为正常用户。当调整参数  $\lambda$  值为 0.285 6 时,即  $\log(\lambda) = -1.253 2$  时,有 4 个变量的标准化回归系数不为 0,输出结果表 4 所示,可知,解释变量  $X_{10}$ 、 $X_{18}$ 、 $X_{22}$  和  $X_{38}$  被保留了下来。但其中解释变量  $X_{10}$  和  $X_{18}$  的异常嫌疑系数远大于其他变量,可初步判定用户 10 和用户 18 为可疑故障用户,该结果与设定一致。该故障情况下自适应 Lasso 模型的拟合优度检验结果如表 5,可知,调整后  $R^2 = 0.917 2$ ,估计后的标准误差为 0.853 0。

表 4 两同相用户故障时各用户的异常嫌疑系数

Table 4 Abnormal suspect coefficient when two in-phase users fail

Variable	Coe	Variable	Coe
用户 10	<b>0.823 0</b>	用户 22	0.012 9
用户 18	<b>0.901 2</b>	用户 38	0.006 4
常量	1.388 6	其他用户	*

注: \* 表示其回归系数被压缩至 0

表 5 两同相用户故障时的拟合优度检验

Table 5 Goodness of fit test when two in-phase users fail

样本量	$R^2$	调整后 $R^2$	估计的标准误差
96	0.920 7	0.917 2	0.853 0

3) 两异相用户错误接线

随机选取台区 A 相用户 26 和 B 相用户 33 发生故障,记录在此条件下持续 1 d 的 15 min 间隔计量数据。由图 18~20 可知,该场景下自适应 Lasso 回归的最佳调整参数  $\lambda$  值为 0.833 3。

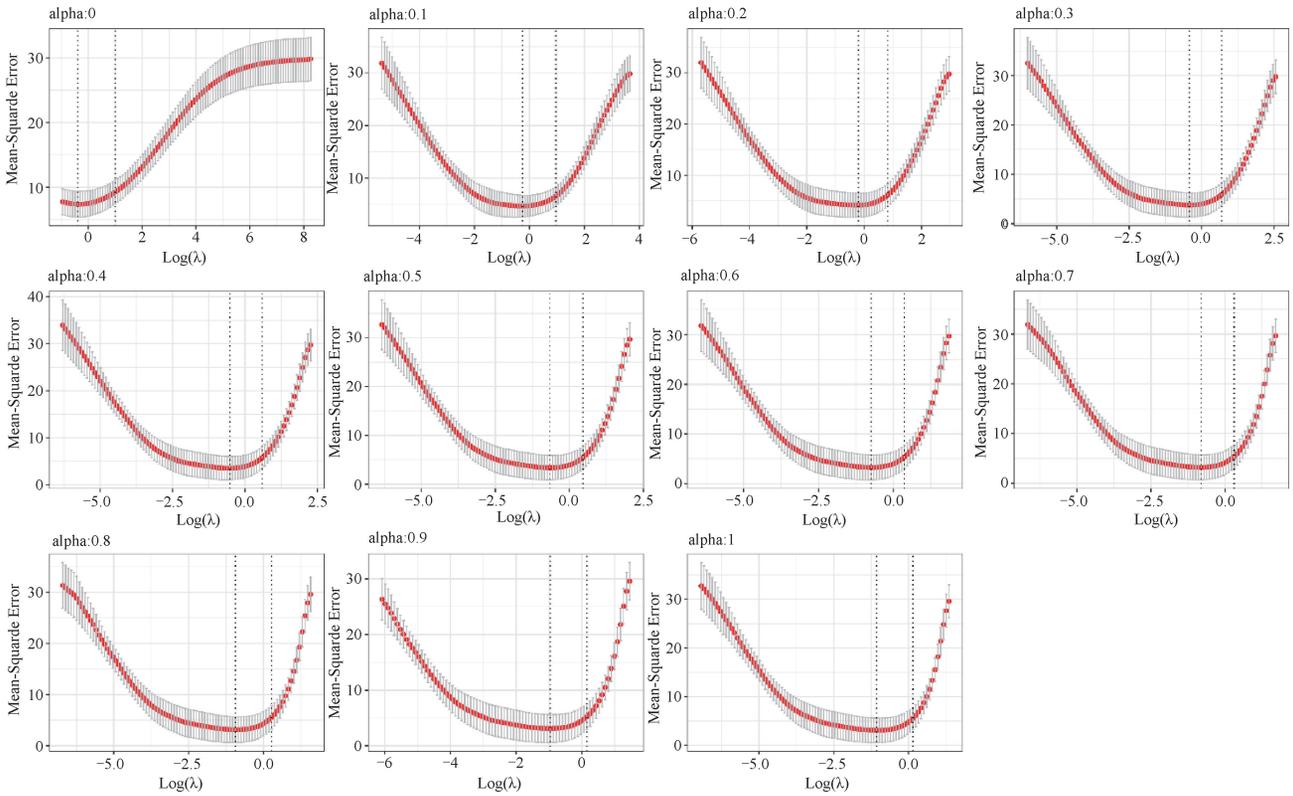


图 14 两同相用户故障时不同模型的交叉验证结果

Fig. 14 Cross-validation results of different models when two in-phase users fail

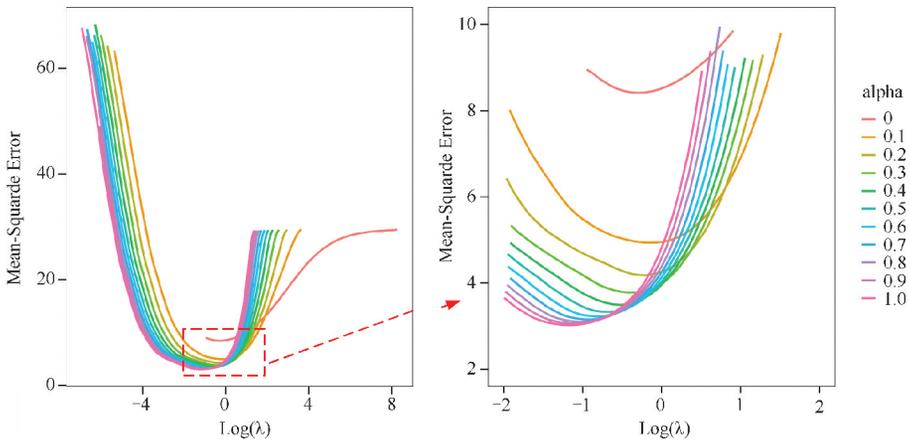


图 15 两同相用户故障时各  $\alpha$  值的比较结果

Fig. 15 The comparison results of each  $\alpha$  value when two in-phase users fail

图 21 呈现了两个异相用户故障时  $\log(\lambda)$  与各解释变量的标准化回归系数的关系,其中错误接线用户 26 和用户 33 的嫌疑系数变化曲线用  $X_{26}$  和  $X_{33}$  进行标注,其他未标注曲线为正常用户。当调整参数  $\lambda$  值为 0.833 3 时,即  $\log(\lambda) = -0.182 3$  时,共有 5 个解释变量被保留。输出结果如表 6 所示,可知,在这 5 个变量中,解释变量  $X_{26}$  和  $X_{33}$  的异常嫌疑系数远大于其他变量,可初

步判断用户 26 和用户 33 为故障用户,结果与设定一致。该故障情况下自适应 Lasso 模型的拟合优度检验结果如表 7,可知,调整后  $R^2$  为 0.905 8,估计后的标准误差为 1.219 4。

### 4.3 对比试验

合适的回归模型能够基于台区接入用户的负荷电流计量数据和台区剩余电流计量数据之间的相关性来估算

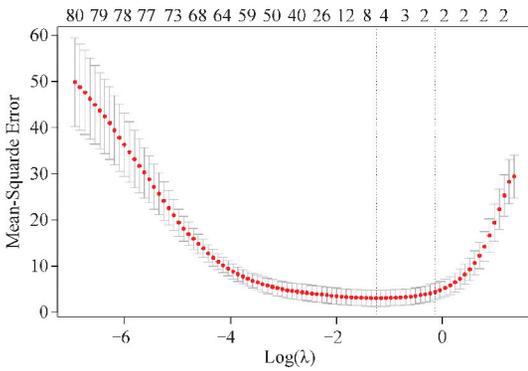


图 16 两同相用户故障时交叉验证结果

Fig. 16 Cross-validation results when two in-phase users fail

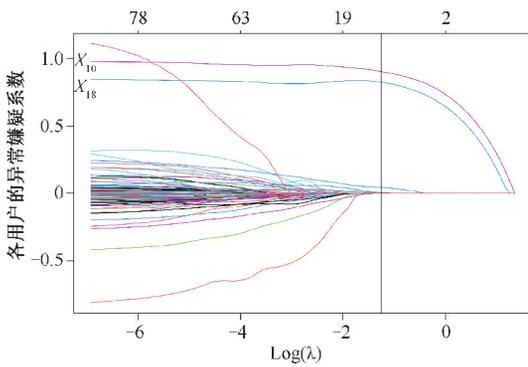


图 17 两同相用户故障时自适应 Lasso 分析图

Fig. 17 Adaptive Lasso analysis diagram when two in-phase users fail

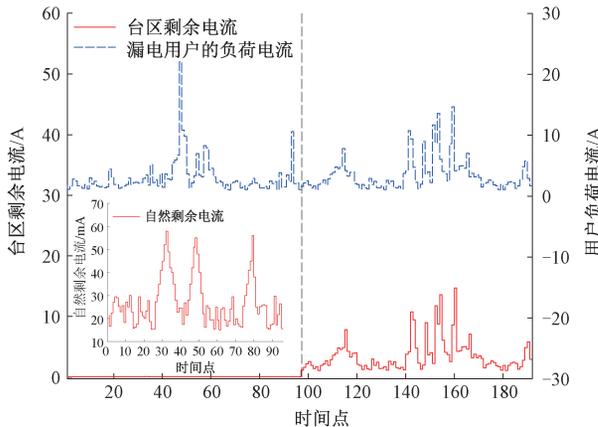


图 18 两异相用户故障时不同模型的交叉验证结果

Fig. 18 Cross-validation results of different models when two out-of-phase users fail

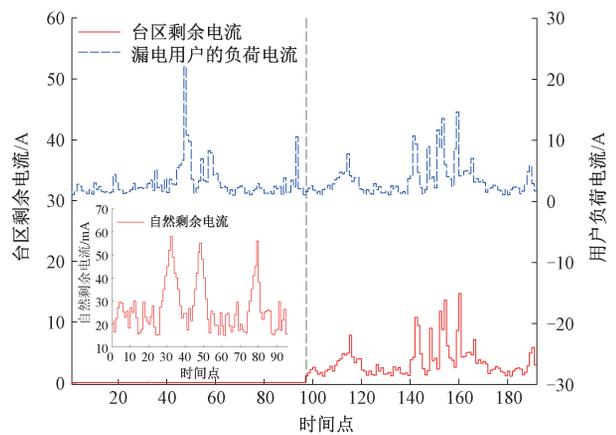


图 19 两异相用户故障时各 α 值的比较结果

Fig. 19 The comparison results of each α value when two out-of-phase users fail

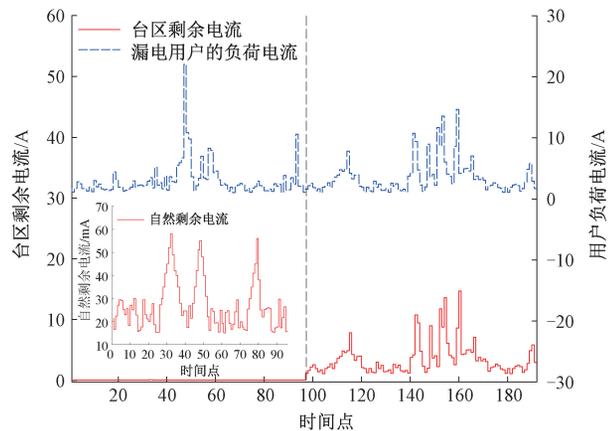


图 20 两异相用户故障时交叉验证结果

Fig. 20 Cross-validation results when two out-of-phase users fail

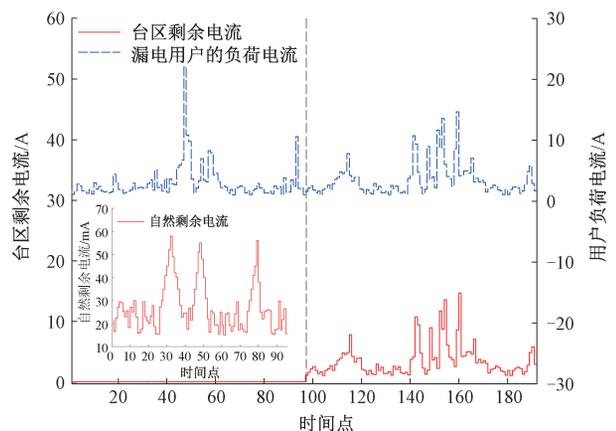


图 21 两异相用户故障时自适应 Lasso 分析图

Fig. 21 Adaptive Lasso analysis diagram when two out-of-phase users fail

每个用户的异常系数,进而定位故障用户。当仅有单相用户故障或者故障用户同相时,由于相关性非常明确,大多数回归模型都能较好的识别。但当故障用户异相时,相关性开始变得模糊,将容易造成误判,且这一情况

表 6 两异相用户故障时各用户的异常嫌疑系数

Table 6 Abnormal suspect coefficient when two out-of-phase users fail

Variable	Coe	Variable	Coe
用户 2	0.056 0	用户 14	0.061 7
用户 26	<b>0.463 5</b>	用户 33	<b>0.634 0</b>
用户 36	0.013 5	其他用户	*
常量	1.319 4	-	-

注: \* 表示其回归系数被压缩至 0

表 7 两异相用户故障时的拟合优度检验

Table 7 Goodness of fit test when two out-of-phase users fail

样本量	$R^2$	调整后 $R^2$	估计的标准误差
96	0.910 3	0.905 8	1.219 4

在数据缺失的背景下更为严重。为了突出本文所提错误接线用户识别方法具有更好的适用性,以下采取前节两异相用户故障的数据和文献[15]所提逐步回归法进行对比试验,实验结果如表 8 所示。

表 8 两异相用户故障时的逐步回归系数

Table 8 Stepwise regression coefficients when two out-of-phase users fail

变量	未标准化系数		标准化系数	$t$	显著性
	B	标准误差	$\beta$		
常量	0.044	0.844	-	0.052	0.958
$X_{33}$	0.862	0.049	<b>0.802</b>	17.425	0
$X_{26}$	0.755	0.077	<b>0.467</b>	9.836	0
$X_{14}$	0.107	0.031	0.156	3.438	0
$X_2$	0.604	0.106	<u>0.315</u>	5.715	0.004
$X_{48}$	-0.251	0.113	-0.103	-2.215	0.030
$X_{13}$	-0.230	0.081	-0.145	-2.847	0.006
$X_{42}$	-0.169	0.055	-0.148	-3.054	0.003
$X_{46}$	-0.143	0.041	-0.163	-3.467	0.000
$X_{69}$	0.273	0.132	0.092	2.069	0.042
$X_{36}$	0.245	0.067	0.172	3.649	0
$X_{73}$	-0.139	0.036	<u>-0.221</u>	-3.865	0
$X_{30}$	0.211	0.077	0.130	2.731	0.008
$X_3$	0.501	0.175	0.146	2.865	0.005
$X_{41}$	-0.349	0.172	-0.097	-2.028	0.046

通过逐步回归分析,得到由解释变量  $X_{33}$ 、 $X_{26}$ 、 $X_{14}$ 、 $X_2$ 、 $X_{48}$ 、 $X_{13}$ 、 $X_{42}$ 、 $X_{46}$ 、 $X_{69}$ 、 $X_{36}$ 、 $X_{73}$ 、 $X_{30}$ 、 $X_3$  和  $X_{41}$  构成的回归模型。由各解释变量的标准化回归系数绝对值大小可

知,共有 12 个变量处于 0.1 数量级,且用户 2、用户 73 和用户 26 较为接近。工程实践中,运维人员无法预判故障用户个数,此时易误判用户 2 和用户 73 故障。此外通过逐步回归建立的回归模型包含冗余变量较多,不仅加大了算力要求,且易造成误判,无法适用于接入用户数量多的中大型台区。

本文采用自适应 Lasso 回归方法筛选最优用户变量构建最优线性回归模型,不仅精简可靠,而且能够解决电力数据集的过拟合问题、多重共线性问题和大数据量等问题,在工程实践中较为有效。

## 5 结 论

针对由用户零线、地线错误接线故障所引起漏电故障定位困难、影响台区一级、二级 RCD 投运的问题,融合台区剩余电流和用户负荷电流数据,提出了基于时序关联特性的错误接线漏电用户定位方法。

1) 本文打破绝缘故障引起台区漏电故障的思维定势,指出引起台区剩余电流严重超标的另一种常见故障类型,并对该类故障进行了危害分析和故障机理分析。

2) 利用 Apriori 算法进行关联分析,得到在发生错误接线时台区剩余电流将与异常用户负荷电流之间呈强关联耦合性特点。在此基础上,建立了自适应 Lasso 回归模型识别出接线异常用户。所提方法对数据要求较低,可在残缺数据的条件下能有效识别。

3) 与主流回归方法相比,本文所用的自适应 Lasso 回归具有模型精简、数据要求较低且贴合漏电台区实际等特点,可以精准筛选出对台区剩余电流波动贡献大的异常用户,有效避免漏判和误判。

4) 基于低压台区用电异常实验平台构建错误接线仿真模型,验证了该方法的可行性和有效性,对于单用户、两同相用户和两异相用户电表零线异常故障的 3 种场景都进行了有效识别,识别结果与预设一致。

实际系统漏电故障种类繁多,本文主要针对由用户零线、地线错误接线所引起的非绝缘漏电故障,为提高台区 RCD 投运率探索了一条新思路,对于其他漏电情形还处于研究过程中。如何拓展引入增量测量数据来定位排查漏电故障,是未来低压用电安全的重要研究方向。

## 参考文献

[1] 鲁文帅,尤睿,周扬,等. 基于单片 TMR 磁传感器的非侵入电流监测微系统[J]. 仪器仪表学报, 2020, 41(6): 1-9.

LU W SH, YOU R, ZHOU Y, et al. Non-invasive current monitoring microsystem based on a single TMR sensor [J]. Chinese Journal of Scientific Instrument, 2020, 41(6): 1-9.

- [ 2 ] 鲍光海, 江润. 基于磁通不对称分布的串联电弧故障检测研究[J]. 仪器仪表学报, 2019, 40(3): 54-61.  
BAO G H, JIANG R. Research on series arc fault detection based on asymmetrical distribution of magnetic flux[J]. Chinese Journal of Scientific Instrument, 2019, 40(3): 54-61.
- [ 3 ] 李天友, 郭峰. 低压配电的触电保护技术及其发展[J]. 供用电, 2019, 36(12): 2-8.  
LI T Y, GUO F. Electric shock protection technology and its development of low-voltage distribution [ J ]. Distribution & Utilization, 2019, 36(12): 2-8.
- [ 4 ] 李奎, 解晨雨, 牛峰, 等. 考虑非线性特性的剩余电流互感器建模及其输出调理电路参数设计[J]. 中国电机工程学报, 2022 42(10): 3815-3826.  
LI K, XIE CH Y, NIU F, et al. Modeling of residual current transformer considering nonlinear characteristics and design of output conditioning circuit parameters[J]. Proceedings of the CSEE, 2022 42(10): 3815-3826.
- [ 5 ] 杨帆, 宿磊, 沈煜, 等. 基于交叉自编码网络的故障漏电流分离方法[J]. 电子测量与仪器学报, 2021, 35(11): 185-193.  
YANG F, SU L, SHEN Y, et al. Fault leakage current separation method based on cross auto encoder network[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(11): 185-193.
- [ 6 ] 黄超艺, 陈宏, 王晨. 低压配电网接地方式及三级剩余电流保护应用实践[J]. 供用电, 2019, 36(12): 29-34.  
HUANG CH Y, CHEN H, WANG CH. Application practice of grounding mode and three-level residual current protection in LV distribution network [ J ]. Distribution & Utilization, 2019, 36(12): 29-34.
- [ 7 ] 王金丽, 刘永梅, 杜松怀, 等. 基于剩余电流固有模态能量特征的生物触电故障诊断模型[J]. 农业工程学报, 2016, 32(21): 202-208.  
WANG J L, LIU Y M, DU S H, et al. Fault diagnosis model for biological electric shock based on residual current intrinsic mode function energy features [ J ]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(21): 202-208.
- [ 8 ] 韩晓慧, 杜松怀, 李振, 等. 基于泄漏电流时频奇异谱和模糊聚类的触电故障诊断[J]. 农业工程学报, 2018, 34(4): 217-222.  
HAN X H, DU S H, LI ZH, et al. Diagnosis of electric shock fault based on time-frequency singular value spectrum of leakage current and fuzzy clustering [ J ]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(4): 217-222.
- [ 9 ] LIU Y M, DU S H, SHENG W X. Classification and identification of electric shock current for safety operation in power distribution network [ J ]. IET Cyber-Physical Systems: Theory & Applications, 2020 5(2): 145-152.
- [ 10 ] 赵怀军, 杨文强, 杜松峰, 等. 多参量融合的路灯触电事故检测方法研究[J]. 电子测量与仪器学报, 2022, 36(3): 70-78.  
ZHAO H J, YANG W Q, DU S F, et al. A current separation leakage protection method using double mutations for TN-C-S systems[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 ( 3 ): 70-78.
- [ 11 ] 梁栋, 王玮, 孙中玉, 等. TN-C-S 系统双突变电流分离漏电保护方法[J]. 电力系统保护与控制, 2022, 50(15): 168-177.  
LIANG D, WANG W, SUN ZH Y, et al. TN-C-S system double mutation current separation leakage protection method[J]. Power System Protection and Control, 2022, 50(15): 168-177.
- [ 12 ] 孙荣可, 单玉凯, 薛永端, 等. 基于负荷电流闭锁的低压配电网漏电保护[J]. 电力系统及其自动化学报, 2023, 35(3): 21-28.  
SUN R K, SHAN Y K, XUE Y D, et al. Leakage protection for low-voltage distribution network based on load current locking[J]. Proceedings of the CSU-EPSA, 2023, 35 ( 3 ): 21-28.
- [ 13 ] 翟国亮, 梁栋, 王玮, 等. 基于剩余电流和不平衡电流突变向量比的 TN-C-S 系统漏电故障检测方法[J/OL]. 中国电机工程学报: 1-12 [ 2023-09-31 ]. <https://doi.org/10.13334/j.0258-8013.pcsee.222522>.  
ZHAI G L, LIANG D, WANG W, et al. Leakage fault detection method for TN-C-S system based on residual current and unbalanced current mutation vector ratio[J/OL]. Proceedings of the CSEE; 1-12 [ 2023-09-31 ]. <https://doi.org/10.13334/j.0258-8013.pcsee.222522>.
- [ 14 ] 张延旭, 胡春潮, 黄曙, 等. 基于 Apriori 算法的二次设备缺陷数据挖掘与分析方法[J]. 电力系统自动化, 2017, 41(19): 147-151+163.  
ZHANG Y X, HU CH CH, HUANG SH, et al. Apriori algorithm based data mining and analysis method for secondary device defects [ J ]. Automation of Electric Power Systems, 2017, 41 ( 19 ): 147-151,163.
- [ 15 ] 赖志强, 李彬, 李文松, 等. 基于台区剩余电流监测

的接线错误用户漏电故障定位[J/OL]. 中国电机工程学报;1-12[2023-09-31]. <https://doi.org/10.13334/j.0258-8013.pcsee.221924>.

LAI ZH Q, LI B, LI W S, et al. Leakage fault location of users with abnormal wiring based on residual current monitoring in serving zone of distribution transformer[J/OL]. Proceedings of the CSEE;1-12[2023-09-31]. <https://doi.org/10.13334/j.0258-8013.pcsee.221924>.

- [16] SERANG S, JACOBUCCI R, BRIMHALL K C, et al. Exploratory mediation analysis via regularization [J]. Structural Equation Modeling: A Multidisciplinary Journal, 2017, 24(5): 733-744.
- [17] 杨楠. 岭回归分析在解决多重共线性问题中的独特作用[J]. 统计与决策, 2004(3): 14-15.
- YANG N. The unique role of ridge regression analysis in solving multicollinearity problems [J]. Statistics & Decision, 2004(3): 14-15.
- [18] MAZILU S, IRIA J. L1 vs. L2 regularization in text classification when learning from labeled features [C]. 10th International Conference on Machine Learning and Applications and Workshops, IEEE, 2011: 166-171.
- [19] 魏新园, 钱牧云, 赵洋洋, 等. 基于正则化的机床热误差自适应稳健建模算法[J]. 仪器仪表学报, 2022, 43(5): 77-85.
- WEI X Y, QIAN M Y, ZHAO Y Y, et al. Adaptive robust modeling method for machine tool thermal error based on regularization[J]. Chinese Journal of Scientific Instrument, 2022, 43(5): 77-85.
- [20] 张沥今, 魏夏琰, 陆嘉琦, 等. Lasso 回归: 从解释到预测[J]. 心理科学进展, 2020, 28(10): 1777-1791.
- ZHANG L J, WEI X Y, LU J Q, et al. Lasso regression: From explanation to prediction[J]. Advances in Psychological Science, 2020, 28(10): 1777-1791.

## 作者简介



周凯欣, 2021 年于湖南科技大学获得学士学位, 现为长沙理工大学硕士研究生, 主要研究方向为低压漏电故障定位。

E-mail: 2568039648@qq.com

**Zhou Kaixin** received his B. Sc. degree from Hunan University of Science and Technology in 2021. Now he is a M. Sc. candidate at Changsha University of Science and Technology. His main research interests include low-voltage leakage fault location.



苏盛, 1998 年于武汉水利电力大学获得学士学位, 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 B. Sc. degree from Wuhan University of Water Conservancy and Electric Power in 1998, M. Sc. degree from Wuhan University in 2002, and Ph. D. degree from Huazhong University of Science and Technology in 2009. Now he is 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.



李彬 (通信作者), 2016 年于湖南城市大学获得学士学位, 2019 年于长沙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和电力气象灾害分析。

E-mail: libin2021666@163.com

**Li Bin** (Corresponding author) received his B. Sc. degree from Hunan City University in 2016, M. Sc. degree from Changsha University of Science and Technology in 2019. Now he is a Ph. D. candidate at Changsha University of Science and Technology. His main research interests include power distribution big data analysis and power meteorological disaster analysis.