

DOI: 10.19783/j.cnki.pspc.231147

# 基于 ISSA-SVC 的配电网高损台区窃电检测方法研究

赖健<sup>1</sup>, 许志浩<sup>1,2,3</sup>, 康兵<sup>1,2</sup>, 王宗耀<sup>1,2</sup>, 丁贵立<sup>1,2</sup>, 袁小翠<sup>1,2</sup>

(1. 南昌工程学院电气工程学院, 江西 南昌 330099; 2. 江西省高压大功率电力电子与电网智能量测工程研究中心, 江西 南昌 330099; 3. 江西博微新技术有限公司, 江西 南昌 330096)

**摘要:** 针对现有的基于机器学习的用户窃电行为检测方法检测效率和准确率不高等问题, 提出一种基于改进麻雀搜索算法(improved sparrow search algorithm, ISSA)优化支持向量分类机(support vector classification, SVC)参数的 ISSA-SVC 窃电检测模型。首先, 该模型通过分析台区每一天的线损率与窃电电量、窃电用户计量电量与窃电电量、窃电用户计量电量与线损电量、台区供电量与窃电电量、用户最近一天用电量和相邻几天用电量、具有相似特征用户用电量曲线的相关性提取用户窃电特征参量。其次, 利用动态时间规整(dynamic time warping, DTW)方法计算得到它们的相关系数。最后, 采用 ISSA 优化 SVC 惩罚参数  $C$  和核参数  $g$ , 并对台区内窃电用户进行检测。仿真算例与实际电网数据分析表明, 所提方法与传统的窃电检测方法相比, 具有更高的效率和准确率。

**关键词:** 机器学习; 窃电检测; 用户窃电特征参量; 相关系数; ISSA-SVC

## A detection method for electricity theft in a high loss station area of a distribution network based on ISSA-SVC

LAI Jian<sup>1</sup>, XU Zhihao<sup>1,2,3</sup>, KANG Bing<sup>1,2</sup>, WANG Zongyao<sup>1,2</sup>, DING Guili<sup>1,2</sup>, YUAN Xiaocui<sup>1,2</sup>

(1. School of Electrical Engineering, Nanchang Institute of Engineering, Nanchang 330099, China; 2. Jiangxi Engineering Research Center of High Electricity Electronics and Grid Smart Metering, Nanchang 330099, China; 3. Jiangxi Boeway New Technology Co., Ltd., Nanchang 330096, China)

**Abstract:** Existing machine learning based user electricity theft detection methods have insufficient detection efficiency and accuracy. Thus an ISSA-SVC model based on the improved sparrow search algorithm (ISSA) to optimize the parameters of support vector classification (SVC) is proposed. First, the model analyzes the correlation between line loss rate and electricity theft, metered electricity consumption and electricity theft, metered electricity consumption and line loss, electricity supply and electricity theft, electricity consumption on the most recent day and adjacent days, and electricity consumption curves of the users with similar characteristics on each day of the station area to extract the characteristics of the users' electricity theft. Secondly, it uses the dynamic time warping (DTW) algorithm to calculate their correlation coefficients. Finally, it uses ISSA to optimize the SVC penalty parameter  $C$  and kernel parameter  $g$ , and detect electricity theft users in the station area. Simulation examples and analysis of real grid data show that the proposed method has higher efficiency and accuracy than traditional power theft detection methods.

This work is supported by the National Natural Science Foundation of China (No. 62001202).

**Key words:** machine learning; electricity theft detection; characteristics of the users' electricity theft; correlation coefficients; ISSA-SVC

## 0 引言

窃电行为在发展中国家普遍存在<sup>[1]</sup>, 美国智能电网咨询服务公司 Northeast Group 在 2017 年发布

的一项研究报告<sup>[2]</sup>中指出, 50 个被调研的发展中国家每年因窃电行为带来的损失高达 647 亿美元; 同时, 窃电行为也给电网的安全稳定运行带来了极大的安全隐患<sup>[3]</sup>。配电网台区分布范围广、供电范围大、供电线路长, 窃电行为的检测通常依赖于分析台区的线损情况, 即供电量与售电量之间的差异。

基金项目: 国家自然科学基金项目资助(62001202)

这种方法需要辅助工作人员进行巡查,巡查过程耗时费力,尤其是对于偏远的农村地区而言,巡查异常困难,且窃电行为往往不容易被巡查人员及时发现<sup>[4]</sup>。智能电表的普及改变了传统人工抄表的方式,大量的用户用电量数据传入数据中台,存储在用电信息采集系统中。因此,充分利用用电数据检测用户窃电行为,对智能电网的发展具有重要意义。

随着大数据、机器学习和人工智能技术的快速发展,国内外许多专家、学者通过分析电力数据对用户窃电行为进行检测<sup>[5]</sup>。文献[6]提出利用用电信息采集系统中反窃电模块推送的用户异常用电告警信息、再辅助工作人员进行现场排查的用户窃电检测方法,这种方法在一定程度上减少了窃电用户企图通过买通工作人员进行窃电的概率,也缩小了工作人员巡查的范围,但是用电采集系统每天要接受和处理千亿级量测数据,容易出现数据上传处理不及时、用户异常用电告警信息推送错误等情况,导致不能及时安排工作人员进行现场排查用户窃电行为。文献[7]提出一种基于蚁狮算法优化支持向量机(ant lion optimizer-support vector machine, ALO-SVM)的用户窃电检测方法,该方法可以有效地检测电力用户的窃电行为,但仅通过用户自身用电量数据的变化提取窃电特征用于用户窃电行为检测误差较大。文献[8]提出一种基于深度自编码器高斯混合模型的窃电行为检测方法。该方法能较好地检测出窃电用户,但只验证了窃电比例对检测结果的影响,未考虑用户实际窃电方式的多样性。文献[9]结合卷积自编码器网络(convolutional auto-encoder, CAEs)和长短期记忆网络(long short term memory, LSTM)的优点,提出一种 CAEs-LSTM 模型来检测用户窃电行为,该模型具有自动提取特征、且融合的特征能较好地表征数据集的特点,在窃电检测的数据集上表现出很好的效果,但在实际窃电用户检测中的精度和稳定性还不够高,需要进一步提高。文献[10]提出一种基于通道注意力网络改进卷积神经网络模型的窃电行为检测方法,可准确检测实际电网情况下的窃电行为,但检测流程和建立窃电评价指标体系的过程相对复杂。

基于上述研究,本文提出一种基于改进麻雀搜索算法优化支持向量分类机(improved sparrow search algorithm-support vector classification, ISSA-SVC)的配电网高损台区窃电检测方法,数据集采用接入居民侧的单相智能电表和台区总表采集的电量数据,通过分析高损台区发生窃电行为每一天的线损率与窃电电量、窃电用户计量电量与窃电电量、窃电用户计量电量与线损电量、台区供电量与窃电电量、

用户最近一天用电量和相邻几天的用电量曲线、具有相似特征的用户用电量曲线的相关性提取用户窃电特征参量,再利用动态时间规整(dynamic time warping, DTW)方法计算它们的相关系数,最后利用 ISSA-SVC 模型进行窃电检测。

## 1 非技术线损分析与窃电特征参量提取

### 1.1 台区拓扑结构

在电力系统的输配电中,为了向众多用电用户提供电力供应,电力公司一般以供电台区为最小单元对其供电范围内的用电用户进行供电。在配电网中,台区是指由一台变压器供电的范围或区域,该范围以配电变压器为中心,并以其供电半径  $R$  为半径向四周扩散,处于该范围内的用户被称为潜在供电用户,并根据其所在的台区相别将其归入 3 个扇形区域。供电台区是电网供给用户的最终供电环节,连接着千家万户。在实际供电过程中,台区配变的供电范围会受到人口分布、地势、建筑物、道路布局、施工难度和经济效益等条件的限制,因此,其供电范围往往呈现出不规则的形态<sup>[11]</sup>。因此,台区拓扑结构是对供电网络进行分析和设计的基础,也是构建智能配电网的关键。配电网台区拓扑结构如图 1 所示,配电台区 10 kV 中压馈线共接了  $A$  台配电变压器  $T_1, T_2, \dots, T_A$ , 0.4 kV 侧  $A$  个台区分别接了  $z$  个用户。

### 1.2 台区非技术线损分析

台区向用户供电电能并不是保持平衡的,即台区供电量并不等于台区内所有用户总用电量之和,而是会大于台区内所有用户总用电量之和,因为在供电的过程中会有一部分电能损耗,该损耗可以分为技术线损和管理线损两部分。技术线损又称为理论线损,包括由导线、变压器绕组、电能表电流、电压线圈和铁芯、电容器介质等产生的电能损耗;管理线损又称为不明损耗,主要包括由用户违章用电、窃电、漏电和电能表计量误差等产生的电能损耗<sup>[12]</sup>。正常状态下城网的线损率不超过 8%,农网的线损率不超过 11%<sup>[13]</sup>。台区线损率计算公式为

$$L = \frac{G - Y}{G} = \frac{\Delta E}{G} = \frac{\Delta E_1 + \Delta E_2}{G} \quad (1)$$

式中:  $L$  为台区线损率;  $G$  和  $Y$  分别为台区供电量和售电量;  $\Delta E$  为台区总电能损耗;  $\Delta E_1$  为供电线路产生的理论电能损耗;  $\Delta E_2$  为窃电产生的管理电能损耗。

### 1.3 窃电特征提取

配电网台区线损异常主要由用电用户窃电所致,基于机器学习的台区居民用户窃电检测方法提

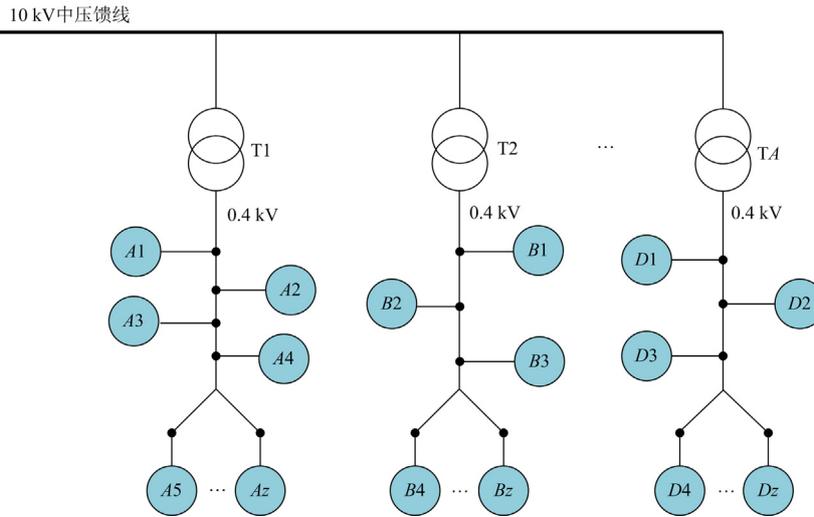


图1 台区拓扑结构

Fig. 1 Topological structure of the substation area

取的典型特征参量主要有峰时耗电量、谷电系数、日负荷率和日最大负荷、平段用电百分比、日平均负荷、日峰谷差、日峰谷差率<sup>[14]</sup>，但以上特征参量不易提取，且计算量较为复杂。因此，不少学者提出仅通过较容易获取的电量数据分析线损率与窃电电量、窃电用户计量电量与窃电电量、窃电用户计量电量与线损电量、台区供电量与窃电电量、用户最近一天计量电量和相邻几天的计量电量曲线、具有相似特征的用户计量电量曲线的相关性来判断用户是否存在窃电行为。文献[15]指出，在供电电量不变且用户窃电量与计量电量呈正相关关系的条件下，窃电用户窃电电量与线损率存在正相关性，窃电电量与台区电量也存在相关性。文献[16]指出，非零电量窃电用户的窃电量与电表计量电量大多也呈正相关，窃电用户计量电量与线损电量之间也存在潜在关联性。文献[17]分析当前和历史用电量的差异，认为用户最近一天用电量和相邻几天的用电量曲线也具有相关性，还对比了同一台区下具有相似特征的用户，认为他们之间的用电量曲线也具有相关性。

用户在进行窃电时最终会影响用户电量值发生突降、突升或较大的偏离正常范围值，台区线损率也会超出正常范围值，因此，本文据此从用户用电量和线损电量数据方面，分析台区线损率与窃电电量、窃电用户计量电量与窃电电量、窃电用户计量电量与线损电量、台区供电量与窃电电量、用户最近一天计量电量和相邻几天的计量电量曲线、具有相似特征的用户计量电量曲线的相关性，计算得到它们的相关系数作为窃电特征参量。常见的相关系数计算方法包括余弦相似度、皮尔逊相关系数、互

信息和 DTW 等<sup>[18-19]</sup>。

DTW 方法常用于比较不同长度的两个数组或时间序列之间的相似性或计算两者间的距离<sup>[20]</sup>，由于电量曲线、线损率曲线是由一系列与时间顺序相关的数据所组成，会存在数据缺失导致时间不匹配、数据长度不同的情况，因此本文采用衡量时间序列数据相似性的 DTW 方法来计算它们的相关系数。DTW 具体计算步骤如下。

1) 通过欧式距离公式分别计算两个对应窃电特征参量数据集各个点之间的距离，如式(2)所示。

$$d_{xy} = \sqrt{\sum_{n=1}^6 (x_n - y_n)^2} \quad (2)$$

式中： $n=1,2,\dots,6$  表示窃电特征参量数据集个数； $x_n$ 、 $y_n$  分别表示两个对应窃电特征参量数据集。

2) 已知两个对应窃电特征参量数据集点  $\phi_x(k)$  与点  $\phi_y(k)$  之间的对应关系为  $\phi(k) = (\phi_x(k), \phi_y(k))$ ， $\phi_x(k)$  取值范围为  $1,2,\dots,N$ ， $\phi_y(k)$  的取值范围为  $1,2,\dots,M$ ， $M$  和  $N$  分别表示对应两个序列的长度。接着求解两个序列的累积距离，其计算公式为

$$d_\phi(x_n, y_n) = \sum_{k=1}^P d_{xy}(\phi(k)) \quad (3)$$

式中， $d_\phi(x_n, y_n)$  为  $P$  个从  $x_n$  序列中点到  $y_n$  序列中点的累积距离。

3) 寻找一条从两个窃电特征参量数据集的累积距离矩阵左上角到右下角的路径，使得路径上的元素和最小，其计算公式为

$$d_a = \min d_\phi(x_n, y_n), a \in [1, 6] \quad (4)$$

式中： $a$  为提取的窃电特征参量个数； $d_a$  为各个窃

电特征参量数据集  $x_n$  和  $y_n$  的最小距离。

4) 定义各个窃电特征参量的相关系数  $\rho_a$ , 其计算公式为

$$\rho_a = \frac{d_a}{\sum_{a=1}^6 d_a}, a \in [1, 6] \quad (5)$$

5) 定义窃电阈值函数  $S$ , 如式(6)所示。

$$S = \sum_{a=1}^6 r_a \rho_a \quad (6)$$

式中,  $r_a$  为各窃电特征参量相关系数的权重。  $S \in (0, 0.8)$  认为用户正常用电,  $S \in [0.8, 1]$  认为用户为窃电用户; 相关系数权重取  $r_1 + r_2 + \dots + r_a = 1$ 。此外, 阈值和相关系数权重可以根据具体情况进行设定。

## 2 基于 ISSA-SVC 方法的窃电检测

### 2.1 ISSA 算法

麻雀搜索算法(sparrow search algorithm, SSA)是根据自然界中麻雀捕食和反捕食过程提出的一种群智能优化算法, 它具有较好的全局搜索能力和较快的收敛速度, 但随着迭代次数的增加, 算法前期全局搜索能力不足, 后期容易陷入局部最优<sup>[21]</sup>。针对以上问题, 本文引入曲线自适应权重向量  $w_1$ 、惯性动态权重向量  $w_2$  分别对 SSA 中发现者、加入者位置进行改进, 使麻雀能够跳出不断缩小的迭代搜索空间中的最优位置, 在更大的空间内进行更广泛的搜索, 防止算法陷入局部优化, 提高算法鲁棒性。改进麻雀搜索算法(improved sparrow search algorithm, ISSA)的步骤如下。

1) 设定麻雀种群规模、最大迭代次数、权值上边界、权值下边界、阈值上边界、阈值下边界、安全值、发现者比例、警戒者的比例, 由  $c$  只麻雀组成的种群  $X$  可表示为

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & & \vdots \\ x_{c,1} & x_{c,2} & \dots & x_{c,d} \end{bmatrix} \quad (7)$$

式中,  $d$  为待求解问题的维度。

2) 利用适应度函数计算种群中各个麻雀的适应度值, 并按照数值大小进行排序, 筛选当前种群内最优与最劣的个体, 适应度大的定义为当前最优麻雀, 所有麻雀的适应度值  $F_x$  表示为

$$F_x = \begin{bmatrix} f([x_{1,1} \ x_{1,2} \ \dots \ x_{1,d}]) \\ f([x_{2,1} \ x_{2,2} \ \dots \ x_{2,d}]) \\ \vdots \\ f([x_{c,1} \ x_{c,2} \ \dots \ x_{c,d}]) \end{bmatrix} \quad (8)$$

式中,  $f$  表示适应度值, 代表寻优目标函数。

3) 利用 ISSA, 在发现者更新方式中引入曲线自适应权重向量  $w_1$ 。  $w_1$  随着迭代次数的增加开始缓慢下降, 中间下降速度较快, 最后再缓慢下降, 改善了前期全局搜索能力和后期容易陷入局部最优的问题。曲线自适应权重  $w_1$  的计算公式为

$$w_1 = (\cos(\pi \cdot t / T_{\max}) + w_{\max})(w_{\max} + w_{\min}) / 2 + h \quad (9)$$

式中:  $t$  为当前迭代次数;  $T_{\max}$  为最大迭代次数;  $w_{\max}$  为权重最大值;  $w_{\min}$  为权重最小值;  $h$  为调整系数。

加入曲线自适应权重向量  $w_1$  后, 发现者位置更新公式为

$$X_{i,j}^{t+1} = \begin{cases} (1 - w_1) \cdot X_{i,j}^t \cdot \exp\left(-\frac{i}{\beta \cdot T_{\max}}\right) & \text{if } R_2 < S_T \\ X_{i,j}^t + Q \cdot L' & \text{if } R_2 \geq S_T \end{cases} \quad (10)$$

式中:  $j = 1, 2, \dots, d$ ;  $X_{i,j}^t$  表示第  $t$  次迭代时第  $i$  只麻雀在第  $j$  维中的位置信息;  $\beta$  是一个随机数,  $\beta \in (0, 1]$ ;  $R_2$  和  $S_T$  分别表示预警值和安全值,  $R_2 \in [0, 1]$ ,  $S_T \in [0.5, 1]$ ;  $Q$  是服从正态分布的随机数;  $L'$  表示一个  $1 \times d$  的矩阵。

4) 在加入者更新方式中引入惯性动态权重向量  $w_2$ ,  $w_2$  随着迭代次数  $t$  的增加缓慢递减, 防止因盲目跟随发现者而错过更好位置, 进而提高算法前期全局搜索能力, 后期进行更精确的局部搜索。  $w_2$  的表达式为

$$w_2 = w_{\max} + (w_{\max} - w_{\min}) \left[ \frac{2t}{T_{\max}} - \left( \frac{t}{T_{\max}} \right)^2 \right] \quad (11)$$

经过多次实验, 当  $w_{\max} = 0.8$ ,  $w_{\min} = 0.3$ ,  $h = 0.5$  时, 寻优结果最好。

加入惯性动态权重向量  $w_2$ , 改进后的加入者位置更新公式为

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{X_{\text{worst}}^t - X_{i,j}^t}{i^2}\right) & \text{if } i > c/2 \\ X_p^{t+1} + w_2 \cdot X_p^{t+1} & \text{otherwise} \end{cases} \quad (12)$$

式中:  $X_p^{t+1}$  表示当前迭代全局最优位置;  $X_{\text{worst}}$  表示当前迭代全局最差的位置。

本文所提 ISSA 算法的麻雀初始种群数目设置为 50, 迭代次数为 200, 经 ISSA 算法优化后支持向量分类机(support vector classification, SVC)的惩罚参数  $C$  为 0.1, 核参数  $g$  为 0.03, ISSA 算法迭代过程如图 2 所示。

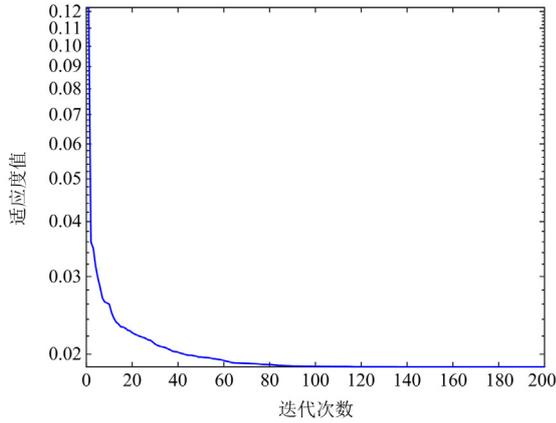


图 2 ISSA 迭代过程

Fig. 2 ISSA iterative process

### 2.2 SVC 算法

SVC 是支持向量机(support vector machine, SVM)的一种二分类形式, 它的基本思想是通过构造最优超平面来划分数据集标签<sup>[22]</sup>。最优超平面可以表示为

$$\begin{cases} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{u=1}^l \varepsilon_u \\ \text{s.t.} & y_u (\mathbf{w}^T x_u + \mathbf{b}) \geq 1 - \varepsilon_u, \varepsilon_u \geq 0, u = 1, 2, \dots, l \end{cases} \quad (13)$$

式中:  $\mathbf{w}$  为权重向量;  $\varepsilon_u$  为松弛变量;  $y_u$  为介于[-1, 1]之间的输出类别;  $x_u$  为窃电特征参量相关系数数据集第  $u$  个训练样本;  $\mathbf{b}$  为偏移向量;  $l$  为样本个数。

式(13)是一个含有不等式的优化问题, 本文选择拉格朗日乘数法进行求解。

计算得到的窃电特征参量相关系数数据集是非线性不可分的, 因此需要引入核函数。目前常用的 SVC 核函数有线性核函数、高斯径向基核函数、多项式核函数、Sigmoid 核函数、傅里叶核函数。本文利用高斯核函数将窃电特征参量相关系数数据集映射到无限维, 使得决策边界更为多样且只有一个参数, 实现非线性变换并分类。非线性 SVC 分类步骤如下。

1) 选择高斯核函数将窃电特征参量相关系数数据集映射到无限维, 实现非线性变换并分类。高斯核函数表达式为

$$k(x_u, x_v) = \exp(-g \|x_u - x_v\|^2), g > 0 \quad (14)$$

式中,  $x_v$  为窃电特征参量相关系数数据集第  $v$  个训练样本。

2) 求解高斯核函数对应的分类决策函数, 其计算公式为

$$f(x) = \text{sign} \left( \sum_{u,v=1}^l a_u^* y_u k(x_u, x_v) + \mathbf{b}^* \right) \quad (15)$$

式中:  $\text{sign}$  为符号函数;  $a_u^*$  为最优拉格朗日因子;  $\mathbf{b}^*$  为最优偏移向量。

综上所述, SVC 利用高斯核函数能够有效解决非线性分类问题, 通过 ISSA 算法对惩罚参数  $C$  和核参数  $g$  进行优化, 能够使 SVC 具有较好的分类效果, 本文将优化后的 SVC 算法用于窃电用户的识别, 提高检测精度。

### 2.3 窃电检测模型

配电网台区常见的窃电方式主要有 3 种: 1) 进行无表法窃电, 通过绕越电能表接线或者在供电线路上私拉电线达到窃电的目的; 2) 通过买通电力企业内部工作人员进行窃电; 3) 改变计量装置, 通过一些技术和非技术手段使计量装置少计或者不计电量数据, 如欠压法、欠流法、移相法、扩差法、强磁干扰窃电法<sup>[23]</sup>。无论以哪种方式进行窃电, 虽然会引起电压、电流、相位角的改变, 但最终目的都是使智能电表计量电量数据减少或者为 0, 体现在电量和台区线损率数据异常中。因此, 本文主要从电量数据入手, 考虑窃电用户窃电方式和台区内用户窃电比例不同会对模型检测结果产生影响, 将正常用电用户电量数据按表 1 中 5 种窃电模型进行更改<sup>[24]</sup>。

表 1 5 种窃电模型

Table 1 Five electricity theft models

| 窃电模型 | 描述  |
|------|---|
| 1    | $X_\lambda^\delta = \beta x_\lambda^\delta$ , 其中 $\beta \in (0.1, 0.8)$ , 且 $\beta$ 为随机数  |
| 2    | $X_\lambda^\delta = \theta x_\lambda^\delta$ , 其中 $\theta \in (0.1, 0.8)$ , 且 $\theta$ 为随机数   |
| 3    | $X_\lambda^\delta = 0$ , 全时段电量数据为 0   |
| 4    | $X_\lambda^\delta = f(\lambda) x_\lambda^\delta$ , 其中 $f(\lambda) = \begin{cases} \beta x_\lambda^\delta & \lambda \in (\lambda_1, \lambda_2) \\ 1 & \text{其他} \end{cases}$ |
| 5    | $X_\lambda^\delta = f(\lambda) x_\lambda^\delta$ , 其中 $f(\lambda) = \begin{cases} 0 & \lambda \in (\lambda_1, \lambda_2) \\ 1 & \text{其他} \end{cases}$                      |

表 1 中, 5 种窃电模型可以表示台区用户进行窃电的常见方式。  $X_\lambda^\delta$  代表用户  $\delta$  发生窃电后  $\lambda$  时刻的电量;  $x_\lambda^\delta$  代表用户  $\delta$  正常用电  $\lambda$  时刻的电量;  $\lambda_1$  和  $\lambda_2$  分别代表用户窃电起止时间。模型 1 认为窃电用户长期按照某一个固定的比例进行窃电, 多为欠电压、欠电流窃电。模型 2 认为窃电用户按照时变的比例进行窃电, 常见于通过远程遥控窃电模块篡改电能表的计量数据进行窃电。模型 3 认为窃电用户通过无表法直接断开电能表接线、绕越电能表、私接电线或者将电能表计量电量数据全时段置为 0

进行窃电。模型 4 认为窃电用户选择时段地将计量电量按时变比例进行减小进行窃电。模型 5 认为窃电用户通过遥控开关选择时段地将计量电量数据置为 0。

ISSA-SVC 窃电检测模型流程如图 3 所示, 具体步骤如下。

- 1) 初始化 ISSA 的参数, 确定 SVC 中  $C$ 、 $g$  参数的取值范围。
- 2) 窃电特征参量提取, 并将窃电特征参量划分成训练集、测试集, 输入训练集进行训练。
- 3) ISSA 算法对 SVC 中  $C$ 、 $g$  参数进行寻优。
- 4) 确定 ISSA 算法的适应度函数并计算种群中各个麻雀的适应度值, 并保留最优麻雀的适应度值和位置。
- 5) 确定是否达到终止条件, 若是, 则执行步骤 6); 否则继续更新  $C$ 、 $g$  参数, 转至步骤 4)。
- 6) 根据最优麻雀个体的位置得到最优参数( $C$ ,  $g$ )并赋给 SVC, 建立优化后的 SVC 分类模型。
- 7) 将测试样本输入模型进行分类。
- 8) 输出测试样本的预测分类结果并与实际值对比, 评估模型的准确率。

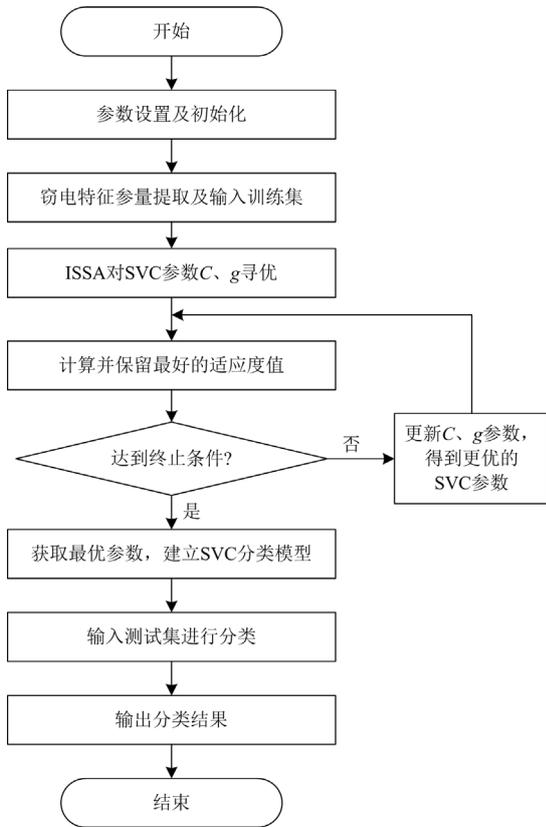


图 3 ISSA-SVC 窃电检测模型流程图

Fig. 3 Flowchart of electricity theft detection model

## 2.4 模型评价指标

窃电检测在本质上属于二分类问题, 即将所检测的用电用户分为正常用户和窃电用户两类<sup>[25]</sup>。本文采用混淆矩阵对模型分类结果进行评价。混淆矩阵是评估分类器可信度的一个较好的工具<sup>[26]</sup>, 混淆矩阵显示了分类器所有可能的分类结果, 其中行对应检测对象实际所属的类别, 列则表示分类器检测的类别。混淆矩阵如表 2 所示。

表 2 混淆矩阵

Table 2 Confusion matrix

| 用户     | 检测为窃电用户 | 检测为正常用户 |
|--------|---------|---------|
| 实际窃电用户 | $T_p$   | $F_N$   |
| 实际正常用户 | $F_p$   | $T_N$   |

表 2 中, 将用电用户按真实标签与检测标签分为真阳性( $T_p$ )、假阳性( $F_p$ )、真阴性( $T_N$ )与假阴性( $F_N$ )。 $T_p$  表示将窃电用户正确地检测出来;  $F_p$  表示将正常用户错误地检测为窃电用户;  $F_N$  表示将窃电用户错误地检测为正常用户;  $T_N$  表示将正常用户正确地检测出来。

在混淆矩阵的基础上可以推导出多个分类器的评价指标: 准确率  $A_{CC}$ 、查准率  $P_{RE}$ 、查全率  $R_{ccall}$  和 F1 分值  $F_1$ 。其中,  $A_{CC}$  为所检测样本中检测正确的比例;  $P_{RE}$  为所有检测为某类别的用户中实际为此类用户的占比;  $R_{ccall}$  为所有实际为某类别的用户中检测为此类用户的占比;  $F_1$  为与  $P_{RE}$  和  $R_{ccall}$  有关的综合指标。各指标计算式分别如式(16)~式(19)所示。

$$A_{CC} = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (16)$$

$$P_{RE} = \frac{T_p}{T_p + F_p} \quad (17)$$

$$R_{ccall} = \frac{T_p}{T_p + F_N} \quad (18)$$

$$F_1 = \frac{2P_{RE} \cdot R_{ccall}}{P_{RE} + R_{ccall}} \quad (19)$$

## 3 仿真实验与实例分析

文中选取 IEEE European Low Voltage Test Feeder 模型搭建仿真台区和实际发生窃电的一个农村配电网台区, 采用上述方法进行分析验证。仿真台区数据集是利用电力仿真软件 OpenDSS, 根据 IEEE European Low Voltage Test Feeder 模型, 通过设置电压等级、阻抗、用电用户、线路长度等参数得到, 农村配电网台区数据集是从电力用户用电采集系统导出。数据集包括台区供电量、线损电量数

据及用户侧电量数据,时间跨度为3个月,每30 min采集一次数据。为了分析窃电方式及用户窃电比例不同对本文方法辨识结果的影响,选取仿真台区数据集中正常用电用户,按照表1中5种窃电模型构造窃电比例为5%、20%、50%的窃电用户,并选其中用户1一天的电量曲线进行分析,如图4所示。

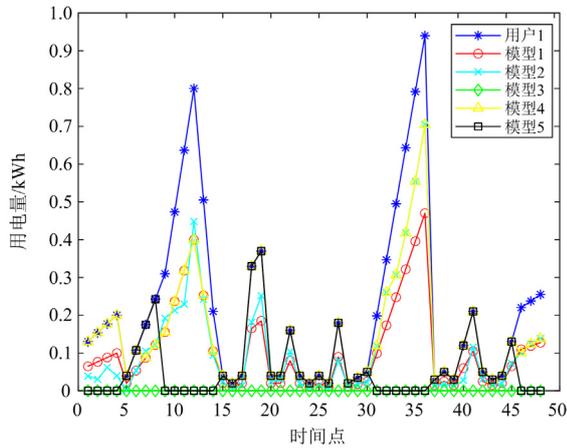


图4 5种窃电模型电量曲线

Fig. 4 Electricity curves of five types of electricity theft models

由图4可知,对用户1构造的5种窃电模型中,模型1为窃电用户长期将电能表计量电量减半进行窃电,模型2为窃电用户按照时变的比例进行窃电,模型3为窃电用户将电能表计量电量数据全时段置为0进行窃电,模型4为窃电用户将用电高峰期时段的计量电量减小进行窃电;模型5为窃电用户将用电高峰期时段的计量电量数据置为0。

本文将仿真数据集中窃电样本与正常样本组成数据集,并按照不同窃电模型和不同窃电比例的窃电用户将数据集分为训练集和测试集。分析不同窃电模型及不同窃电比例的窃电用户对模型检测结果的影响。经多次试验,综合检测结果对比如表3所示。

由表3可知,在仿真数据集中台区窃电比例为50%且窃电用户采取模型1和模型5方式窃电时准确率较高,文中所提方法的检测结果准确率分别为98.56%和95.23%,查准率分别为92.61%和90.72%,查全率分别为95.38%和92.56%,同时有较高的F1分值,表现出较好的检测性能。在5%和20%低窃电比例以及窃电方式为模型2、模型3和模型5下,基于本文方法检测窃电用户的检测性能较差。

表3 窃电检测结果对比

Table 3 Comparison of electricity theft detection results

| 窃电模型 | 窃电比例/% | 准确率/% | 查准率/% | 查全率/% | F1 分值 |
|------|--------|-------|-------|-------|-------|
| 模型 1 | 5      | 70.23 | 30.88 | 50.23 | 40.02 |
|      | 20     | 78.42 | 49.58 | 64.32 | 58.12 |
|      | 50     | 98.56 | 92.61 | 95.38 | 94.24 |
| 模型 2 | 5      | 62.36 | 28.81 | 46.42 | 34.23 |
|      | 20     | 65.31 | 42.68 | 57.65 | 51.69 |
|      | 50     | 87.23 | 83.82 | 88.32 | 85.13 |
| 模型 3 | 5      | 55.42 | 24.98 | 39.72 | 30.28 |
|      | 20     | 58.71 | 35.88 | 50.42 | 48.71 |
|      | 50     | 83.22 | 76.78 | 81.57 | 81.54 |
| 模型 4 | 5      | 50.12 | 24.47 | 32.12 | 26.42 |
|      | 20     | 57.30 | 30.97 | 48.85 | 43.73 |
|      | 50     | 80.12 | 77.90 | 79.65 | 78.75 |
| 模型 5 | 5      | 72.23 | 35.38 | 55.74 | 43.31 |
|      | 20     | 83.51 | 53.47 | 70.75 | 62.14 |
|      | 50     | 95.23 | 90.72 | 92.56 | 94.31 |

基于本文所提方法对窃电比例为50%的实际台区窃电用户进行检测,其中检测样本中有20个正常用电用户、15个窃电用户。窃电检测混淆矩阵如图5所示。混淆矩阵行代表预测值为某一个类别,列代表真实标签为某一个类别,0表示正常用电用户,1表示窃电用户。

由图5可知,文中所提方法正确识别出了测试集中20个正常用电用户、14个窃电用户、1个窃

电用户被错误识别成了正常用电用户,准确率高达97.14%。

为了进一步说明本文所提方法的优越性,对比了SVM、ALO-SVM、粒子群优化算法优化支持向量机(partical swarm optimization-support vector machine, PSO-SVM)等3种常见的窃电检测方法。经过多次试验,在高窃电比例台区下各窃电检测方法对比结果如表4所示。

由表 4 可知, 针对本文所构建的窃电用户, 基于 ISSA-SVC 的窃电检测方法比其他 3 种方法具有更高的检测准确率和精度。

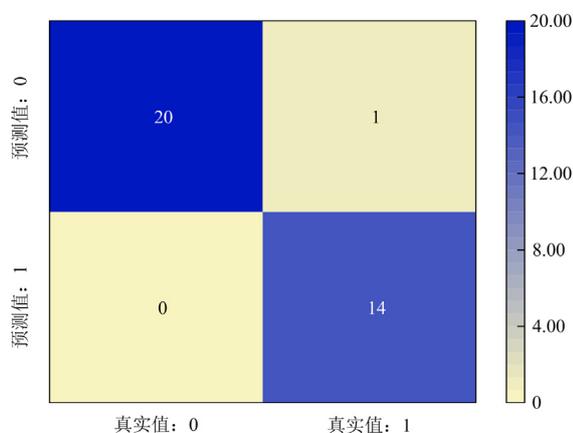


图 5 窃电检测混淆矩阵

Fig. 5 Confusion matrix for electricity theft detection

表 4 窃电检测方法对比

Table 4 Comparison of electricity theft detection methods

| 方法       | 准确率/% | 查准率/% | 查全率/% | F1 分值 |
|----------|-------|-------|-------|-------|
| SVM      | 86.23 | 80.88 | 82.23 | 84.02 |
| ALO-SVM  | 92.36 | 88.82 | 90.32 | 91.28 |
| PSO-SVM  | 94.10 | 90.36 | 92.63 | 93.52 |
| ISSA-SVC | 97.14 | 92.32 | 93.56 | 95.30 |

## 4 结论

本文基于 IEEE European Low Voltage Test Feeder 模型仿真数据集、实际配电网台区总表和居民用户单相智能电表 HPLC 模块高频次采集的数据集, 提出了一种基于 ISSA-SVC 的配电网高损台区窃电检测模型。针对台区不同窃电比例和不同窃电方式的窃电用户, 对台区总表和居民用户单相智能电表量测数据进行分析, 提取窃电特征参量, 计算相应窃电特征参量的相关系数; 针对 SVC 分类器分类效果不佳的问题, 利用 ISSA 优化了 SVC 算法中的惩罚参数  $C$  和核参数  $g$ , 提高了 SVC 分类准确性。仿真与实验结果表明: 相比于现有其他检测方法, 本文所提方法受窃电样本影响小, 具有更高的检测准确率和精度, 可准确识别窃电用户。但在窃电方式较复杂和台区内窃电比例较小的情况下检测准确率较低。下一步的工作重点是获取更多的用户实际窃电样本数据进行用户窃电检测, 不断优化方法适用性, 提高检测准确率。

## 参考文献

- [1] FAISAL J, EATZAZ A. Policy considerations for limiting electricity theft in the developing countries[J]. Energy Policy, 2019, 129: 452-458.
- [2] Northeast Group LLC. Emerging markets smart grid: outlook 2017[EB/OL]. [2017-02-22]. <https://www.reportbuyer.com/product/3866771>.
- [3] 陈文瑛, 陈雁, 邱林, 等. 应用大数据技术的反窃电分析[J]. 电子测量与仪器学报, 2016, 30(10): 1558-1567. CHEN Wenyong, CHEN Yan, QIU Lin, et al. Analysis of anti-stealing electric power based on big data technology[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(10): 1558-1567.
- [4] 陈启鑫, 郑可迪, 康重庆, 等. 异常用电的检测方法: 评述与展望[J]. 电力系统自动化, 2018, 42(17): 189-199. CHEN Qixin, ZHENG Kedi, KANG Chongqing, et al. Detection methods of abnormal electricity consumption behaviors: review and prospect[J]. Automation of Electric Power Systems, 2018, 42(17): 189-199.
- [5] 县国成, 王永攀, 高俊, 等. 基于 ELM-SVM 模型与电能计量大数据的窃电识别技术研究[J]. 智慧电力, 2022, 50(9): 82-89. XIAN Guocheng, WANG Yongpan, GAO Jun, et al. Power theft identification technology based on ELM-SVM model and big data of electric energy measurement[J]. Smart Power, 2022, 50(9): 82-89.
- [6] 程超, 张汉敬, 景志敏, 等. 基于离群点算法和用电信息采集系统的反窃电研究[J]. 电力系统保护与控制, 2015, 43(17): 69-74. CHENG Chao, ZHANG Hanjing, JING Zhimin, et al. Study on the anti-electricity stealing based on outlier algorithm and the electricity information acquisition system[J]. Power System Protection and Control, 2015, 43(17): 69-74.
- [7] 金晟, 苏盛, 薛阳, 等. 数据驱动窃电检测方法综述与低误报率研究展望[J]. 电力系统自动化, 2022, 46(1): 3-14. JIN Sheng, SU Sheng, XUE Yang, et al. Review on data-driven based electricity theft detection method and research prospect for low false positive rate[J]. Automation of Electric Power Systems, 2022, 46(1): 3-14.
- [8] 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 Zhaorui, GAO Yunpeng, GUO Jianbo, et al. Abnormal detection of electricity theft using a deep auto-encoder Gaussian mixture model[J]. Power System Protection and Control, 2022, 50(18): 92-102.
- [9] 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 Lihong, XIAO Chunlang, 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.
- [10] 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.
- [11] 李彦兆. 低压配电网拓扑模型构建及用户用电行为分析[D]. 太原: 太原理工大学, 2021.
- LI Yanzhao. Construction of low-voltage distribution network topology model and analysis of consumer power consumption behavior[D]. Taiyuan: Taiyuan University of Technology, 2021.
- [12] 林振智, 崔雪原, 金伟超, 等. 用户侧窃电检测关键技术[J]. 电力系统自动化, 2022, 46(5): 188-199.
- LIN Zhenzhi, CUI Xueyuan, JIN Weichao, et al. Key technologies of electricity theft detection at consumer side[J]. Automation of Electric Power Systems, 2022, 46(5): 188-199.
- [13] 张勤, 马玮, 付锦, 等. 中压配电网规划态线损计算方法[J]. 电气技术, 2018, 19(2): 42-45.
- ZHANG Qin, MA Wei, FU Jin, et al. Line loss calculation for middle voltage distribution network planning[J]. Electrical Engineering, 2018, 19(2): 42-45.
- [14] 曾虎. 电能计量数据聚类分析与窃电检测研究[D]. 昆明: 昆明理工大学, 2017.
- [15] 金晟, 苏盛, 曹一家, 等. 基于格兰杰归因分析的高损台区窃电检测[J]. 电力系统自动化, 2020, 44(23): 82-89.
- JIN Sheng, SU Sheng, CAO Yijia, et al. Electricity-theft detection for high-loss distribution area based on Granger causality analysis[J]. Automation of Electric Power Systems, 2020, 44(23): 82-89.
- [16] 郑应俊, 杨艺宁, 舒一飞, 等. 基于边缘计算的低压用户窃电检测[J]. 电力系统自动化, 2022, 46(11): 111-120.
- ZHENG Yingjun, YANG Yining, SHU Yifei, et al. Electricity theft detection for low-voltage users based on edge computing[J]. Automation of Electric Power Systems, 2022, 46(11): 111-120.
- [17] 吴迪. 基于曲线相似性分析的窃电用户判断[J]. 中国电力, 2017, 50(2): 181-184.
- WU Di. Electricity theft identification method based on curve similarity[J]. Electric Power, 2017, 50(2): 181-184.
- [18] 许阳, 任盛, 刘鑫, 等. 基于皮尔逊相关系数法的台区窃电精准分析[J]. 大众用电, 2021, 36(8): 14-15.
- XU Yang, REN Sheng, LIU Xin, et al. Accurate analysis of electricity stealing in station area based on Pearson correlation coefficient method[J]. Popular Utilization of Electricity, 2021, 36(8): 14-15.
- [19] 刘苏, 黄纯, 侯帅帅, 等. 基于 DDTW 距离与 DBSCAN 算法的户变关系识别方法[J]. 电力系统自动化, 2021, 45(18): 71-77.
- LIU Su, HUANG Chun, HOU Shuaishuai, et al. Identification method for household-transformer relationship based on derivative dynamic time warping distance and density-based spatial clustering of application with noise algorithm[J]. Automation of Electric Power Systems, 2021, 45(18): 71-77.
- [20] 杨洁, 康宁. 动态时间规整 DTW 算法的研究[J]. 科技与创新, 2016(4): 11-12.
- [21] 薛建凯. 一种新型的群智能优化技术的研究与应用: 麻雀搜索算法[D]. 上海: 东华大学, 2020.
- XUE Jiankai. Research and application of a novel swarm intelligence optimization technique: sparrow search algorithm[D]. Shanghai: Donghua University, 2020.
- [22] 张梦楠, 李红娇. 基于 DCNN 和 SVC 的窃电检测[J]. 计算机仿真, 2022, 39(6): 92-97, 429.
- ZHANG Mengnan, LI Hongjiao. Electricity theft detection based on deep convolutional neural network and support vector classification[J]. Computer Simulation, 2022, 39(6): 92-97, 429.
- [23] 冯郁. 用电监察中窃电与反窃电技术的研究[J]. 自动化应用, 2023, 64(8): 38-39, 45.
- FENG Yu. Research on power stealing and anti-power stealing technology in electricity monitoring[J]. Automation Application, 2023, 64(8): 38-39, 45.
- [24] 赵云, 肖勇, 曾勇刚, 等. 一种相关性与聚类自适应融合技术窃电检测方法[J]. 南方电网技术, 2021, 15(9): 69-74.
- ZHAO Yun, XIAO Yong, ZENG Yonggang, et al. An electricity theft detection method with correlation and clustering fusion technique[J]. Southern Power System Technology, 2021, 15(9): 69-74.
- [25] 李秋红. 基于改进粒子群算法的前馈神经网络识别用户窃电行为[J]. 电气技术, 2022, 23(11): 44-48.
- LI QiuHong. Feedforward neural network based on improved particle swarm optimization algorithm for identification of user electricity stealing behavior[J]. Electrical Engineering, 2022, 23(11): 44-48.
- [26] 李金瑾, 陈珏羽, 黄柯颖. 基于 CNN-LSTM 混合模型的多类别窃电行为检测[J]. 电力科学与技术学报, 2023, 38(1): 226-234.
- LI Jinjin, CHEN Jueyu, HUANG Keying. Multi-class electricity theft detection based on the CNN-LSTM hybrid model[J]. Journal of Electric Power Science and Technology, 2023, 38(1): 226-234.

收稿日期: 2023-09-04; 修回日期: 2023-11-29

作者简介:

赖健(1997—), 男, 硕士研究生, 研究方向为电力系统大数据应用分析; E-mail: 957651802@qq.com

许志浩(1988—), 男, 通信作者, 博士, 副教授, 研究方向为输配电网智能无损检测技术。E-mail: zhxuhi@whu.edu.cn

(编辑 周金梅)