

# 基于改进TimeGAN数据增强的用户窃电识别研究

吴佐平,王宏岩,张千福,谢青

(北京中电普华信息技术有限公司,北京100085)

**摘要:**用户窃电是电网电能非技术损失的主要原因,对电力系统造成了巨大的经济损失和资源浪费。相较于大量用户正常用电样本,窃电用户属于少数类样本,而传统窃电分类方法在样本稀疏或失衡情况下表现不佳。由此,提出一种基于改进时间序列生成对抗网络(TimeGAN)数据增强的用户窃电分类方法,使用TimeGAN对原始小样本窃电数据进行增强,生成与原始数据分布相似的增广样本,考虑到增广样本存在噪声或不可信等问题,利用马氏距离评估增广样本的质量,完成不可信样本剔除。通过卷积神经网络(CNN)对数据增强后的用电负荷数据进行特征提取,采用长短时记忆网络(LSTM)提取特征量的时序相关性并完成特征分类,进一步,利用麻雀搜索算法(SSA)对CNN-LSTM网络进行参数优化,提高模型检测精度。实验结果表明,所提方法可有效解决用户窃电行为识别中样本不平衡的二分类问题。

**关键词:**TimeGAN模型;马氏距离;麻雀搜索算法(SSA);窃电识别

**中图分类号:**TP39;F426.61   **文献标识码:**A   **DOI:**10.19457/j.1001-2095.dqcd26311

## Research on User Electricity Theft Recognition Based on Improved TimeGAN Data Enhancement

WU Zuoping, WANG Hongyan, ZHANG Qianfu, XIE Qing

(Beijing China-Power Information Technology Co.,Ltd., Beijing 100085, China)

**Abstract:** Electricity theft by users is the main cause of non-technical loss of electric energy in power grids, which causes huge economic losses and resource wastage to the power system. Compared with a large number of users' normal electricity samples, electricity theft users belong to a minority class of samples, and the traditional electricity theft classification methods perform poorly in the case of sparse or imbalanced samples. As a result, a user electricity theft classification method based on the data enhancement of improved time series generative adversarial network (TimeGAN) was proposed. TimeGAN was used to enhance the original small-sample electricity theft data, generating the augmented samples similar to the distribution of the original data, and considering that the augmented samples are noisy or untrustworthy, the quality of augmented samples was evaluated using the Mahalanobis distance to complete the untrustworthy sample rejection. Convolutional neural network (CNN) was used to extract features from the augmented electricity load data, and long-short time memory network (LSTM) was used to extract the temporal correlation of the feature quantities and complete the feature classification, and furthermore, the sparrow search algorithm (SSA) was used to optimize the parameters of the CNN-LSTM network, so as to improve the accuracy of the model detection. The experimental results show that the proposed method can effectively solve the binary classification problem of sample imbalance in the identification of user's electricity theft behavior.

**Key words:** TimeGAN model; Marginal distance; sparrow search algorithm(SSA); electricity theft identification

随着电网规模的不断扩大,电网的电能损耗也逐渐成为一个关注的焦点。电网的电能损耗主要包括两类,技术损耗(电力传输中的固有损失)和非技术损耗(窃电、记录异常等)。其中,窃电行为给电网带来了巨大的经济损失,因此,准

确地判断出用户窃电行为,对维护电力市场的正常运转和降低电能损耗提高电网运行效率有着重要意义<sup>[1]</sup>。

传统的窃电行为检测主要是依靠人工排查,但该方法效率低下,很难精准地发现窃电用户。

基金项目:国网信息通信产业集团基金(SGJTPH00YXJS2310260)

作者简介:吴佐平(1980—),男,本科,工程师,主要研究方向为电力信息化系统集成技术等,Email:wuzuoping11@163.com

随着机器学习和大数据技术的发展,国内外的研究者通过机器学习算法对用户用电信息进行分析,开展异常用电检测研究。吴蕊等<sup>[2]</sup>利用改进 K-means 算法进行电力数据异常检测。耿俊成等<sup>[3]</sup>提出了一种基于局部离群点检测的用户窃电识别方法,可以输出用户窃电的嫌疑度排序,从而提高了反窃电的效率。随着深度学习方法在分类任务上取得优异性能,不少研究者开始采用深度学习模型进行用户异常诊断。万磊等<sup>[4]</sup>提出一种基于置信规则推理(belief rule-based, BRB)和长短时记忆网络(LSTM)的用户窃电行为诊断方法。邹念等<sup>[5]</sup>提出一种基于小时尺度周期特征自编码器方法,用于非高损线路窃电行为识别。文献[6]给出了一种基于双神经网络自学习的智能电网自适应窃电检测方法。

然而在实际工况条件下,电网中窃电用户的数量远少于正常用户,造成用电样本不平衡<sup>[7]</sup>。直接处理会导致训练的模型倾向于给样本数量较多的类别更多的权重,使少数类样本产生较差的预测精度<sup>[8]</sup>。王欢等<sup>[9]</sup>提出了一种基于生成对抗网络(GAN)的诊断模型,将数据输入到 IW-GAN 中进行样本扩充,显著提升了小样本情况下模型的整体性能。杨艺宁等<sup>[10]</sup>利用 GAN 在少量有窃电标签数据场景下通过扩充标签样本数据集,提升了模型检测准确度和泛化能力。由此表明,通过生成与原始数据相似的高质量用户侧数据,来平衡数据达到窃电行为的准确检测具有重要意义。

有鉴于此,本研究首次应用改进的基于时间序列的生成对抗网络(TimeGAN)进行窃电小样本数据增强,其能捕捉时间序列动力学特征以生成高质量虚拟样本,且引入马氏距离评估筛选增广样本,确保样本质量。提出的 SSA-CNN-LSTM 模型融合各组件优势,卷积神经网络(CNN)捕捉空间特征、LSTM 提取时序相关性,同时利用麻雀搜索算法(SSA)优化模型参数,增强全局与局部搜索能力。经实验对比,所提方法在多项评价指标上显著优于传统方法,有效解决样本不平衡问题,为窃电检测提供更准确可靠手段。

## 1 基于 TimeGAN 的窃电少数类样本生成

### 1.1 TimeGAN 网络模型原理

GAN 在计算机视觉领域取得了很大成功。

由于实际工作中电网窃电用电行为数据较少,使得模型在现有样本不平衡分类下很容易过拟合或欠拟合。由此,可利用 GAN 网络生成更多的少类样本数据。

在现有 GAN 用于生成序列数据的研究中,很少考虑时间动力学特征。由此,文献[11]给出了 TimeGAN,该方法结合了无监督学习和有监督学习,兼顾灵活性和训练可控性。同时,将 GAN 和自编码器结合进行训练,可以避免时间序列数据增强中存在的自回归问题,图 1 展示了 TimeGAN 的基本结构。将生成对抗网络在嵌入空间中完成对生成器和判别器的训练,这样对高维数据时序相关性进行拟合。

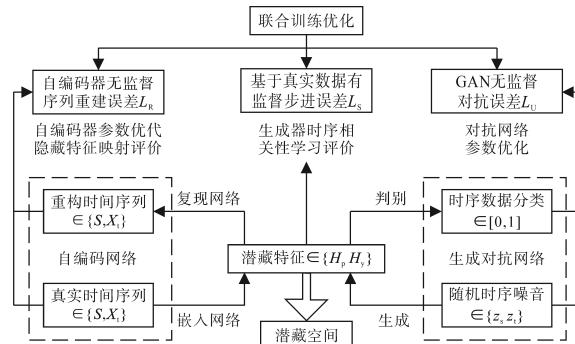


图 1 TimeGAN 模型结构图

Fig.1 TimeGAN model structure

实际上,将窃电一维时间序列转换成二维矩阵,采用覆盖所有列的相同滑动时间窗,将数据转化为三维时间序列。将转换后的三维时间序列在自编码中进行数据重构,嵌入函数可以表示为

$$\begin{cases} h_s = e_s(S) \\ h_t = e_x(h_s, h_{t-1}, x_t) \end{cases} \quad (1)$$

式中: $S$  为静态数据; $X$  为基于时间变化的动态序列; $e$  为变量所对应嵌入函数; $h_s$  为静态数据潜在空间; $h_t$  为动态序列的潜在空间。

而复现函数可表示为

$$\begin{cases} \tilde{S} = r_s(h_s) \\ \tilde{x}_t = r_x(h_t) \end{cases} \quad (2)$$

式中: $r$  为复现函数; $\tilde{S}, \tilde{x}_t$  为复现解码后输入数据。由此,生成对抗网络模型可以由下式表示:

$$\begin{cases} \hat{h}_s = g_s(z_s) \\ \hat{h}_t = g_x(\hat{h}_s, \hat{h}_{t-1}, z_t) \end{cases} \quad (3)$$

$$\begin{cases} \tilde{y}_s = d_s(\tilde{h}_s) \\ \tilde{y}_t = d_x(\tilde{u}_t, \vec{u}_t) \end{cases} \quad (4)$$

式中: $g$  为生成函数; $d$  为判别函数; $z$  为噪声; $\hat{h}_s, \hat{h}_t$

为生成数据;  $\tilde{y}_s, \tilde{y}_t$  为判别结果。

设计中 TimeGAN 采用联合训练方法, 依靠编码器损失  $L_R$ 、非监督损失  $L_U$  和监督损失  $L_S$  对生成对抗网络进行训练。 $L_R$  反映了自编码器对输入数据内在特征的掌握程度, 计算公式如下:

$$L_R = E_{S, X_i, T-P} [\|S - \tilde{S}\|_2 + \sum_t \|x_t - \tilde{x}_t\|_2] \quad (5)$$

$L_U$  反映生成器生成的数据能够逼近经过自编码器编码后真实数据的程度。公式如下:

$$\begin{aligned} L_U &= E_{S, X_i, T-P} [\log y_s + \sum_t \log y_t] + \\ &E_{S, X_i, T-P} [\log (1 - \tilde{y}_s) + \sum_t \log (1 - \tilde{y}_t)] \end{aligned} \quad (6)$$

$L_S$  反映生成器和判别器的竞争对抗互动, 表示为

$$L_S = E_{S, X_i, T-P} [\sum_t \|h_t - g_x(h_s, h_{t-1}, z_t)\|_2] \quad (7)$$

## 1.2 增广样本评估

TimeGAN 生成的增广样本与真实窃电样本分布可能存在一定差异, 且生成样本存在噪声或不可信等问题。未经筛选的增广样本会对模型训练产生负面影响, 影响故障子空间的故障表征。考虑到用电数据存在非线性、多时间维度, 以及数据特征之间的相关性, 选择马氏距离来对增广窃电样本进行筛选。其中, 生成样本  $x_i^*$  与原始样本集合  $X = (x_1, x_2, \dots, x_j)$  之间的距离为

$$Dis(x_i^*, X) = [(x_i^* - \bar{x})^T (U)^{-1} (x_i^* - \bar{x})]^{0.5} \quad (8)$$

式中:  $\bar{x}$  为生成样本平均值;  $U$  为协方差矩阵。

设经过 TimeGAN 生成的  $m$  个增广窃电样本为  $x_i (i = 1, 2, \dots, m)$ , 分别计算各个原始样本之间距离的最大值  $\max Dis(x_i, X)$ , 如果  $Dis(x_i^*, X)$  大于  $\max Dis(x_i, X)$ , 则将生成样本保留, 否则舍弃。

## 2 特征提取与分类

本文提出的特征提取与分类模型主要包括 SSA, CNN 和 LSTM。SSA 主要用于优化数据及特征提取分类模型的参数, CNN 捕捉时间序列的空间特征, 为 LSTM 提供丰富特征信息。

### 2.1 SSA 种群优化方法

SSA 模拟麻雀觅食和反捕食行为, 具有较强的全局搜索能力和局部开发能力<sup>[12]</sup>。SSA 中, 适应度越大则表示当前位置越可能发现食物并可以躲避天敌, 而较低的适应度则可能表示麻雀需要前往新的位置继续搜索食物, 算法原理步骤如下:

$$X = DB + rand(0, 1) \times (AB - DB) \quad (9)$$

式中:  $DB, AB$  分别为麻雀位置的最大、最小限制。

搜索过程如下:

$$X_i^{t+1} = \begin{cases} X_i^t \pm \alpha |X_i^t - X_k^t| & R_2 < S_t \\ X_i^* + Q \times L & R_2 \geq S_t \end{cases} \quad (10)$$

式中:  $X_i^t$  为第  $i$  个麻雀在第  $t$  时刻的位置;  $X_k^t$  为第  $k$  个麻雀在第  $t$  时刻的位置;  $X_i^*$  为当前全局最优位置;  $R_2$  为预警值,  $R_2 \in [0, 1]$ ;  $S_t$  为安全值,  $S_t \in [0, 1]$ ;  $\alpha, Q$  为常数;  $L$  为一个服从标准正态分布的随机数。

当  $R_2 < S_t$  时, 表示当前位置处于安全区域, 探索者可以进行广泛的搜捕工作; 当  $R_2 \geq S_t$  时, 表示种群中已经有一些麻雀发现了天敌, 并迅速向其他同类发布警报信息, 此时其他觅食的麻雀会到其他安全的位置进行觅食。每轮位置信息更新的时候都会有一批新的加入者加入搜索过程。加入者位置更新描述公式如下:

$$X_{ij}^{t+1} = X_i^t + \beta (X_{best}^t - X_i^t) + \varepsilon (X_k^t - X_i^t) \quad (11)$$

式中:  $X_{ij}^{t+1}$  为加入者在第  $t+1$  时刻的位置;  $X_{best}^t$  为当前全局最优位置;  $\beta$  为一个步长控制参数;  $\varepsilon$  为一个随机数。

### 2.2 CNN-LSTM 特征提取与分类

电网数据具有很强的时间依赖性, 所提 SSA-CNN-LSTM 结构能够在复杂的电网数据中进行有效的搜索, 找到最优或接近最优的解决方案。模型还包含 2 个卷积层和 2 个池化层的 CNN, CNN 输出作为特征量输入到 LSTM 网络中对时序特征进行提取。处理后的特征量输入到全连接层, 最后通过 Softmax 分类器输出分类结果, 结构如图 2 所示。

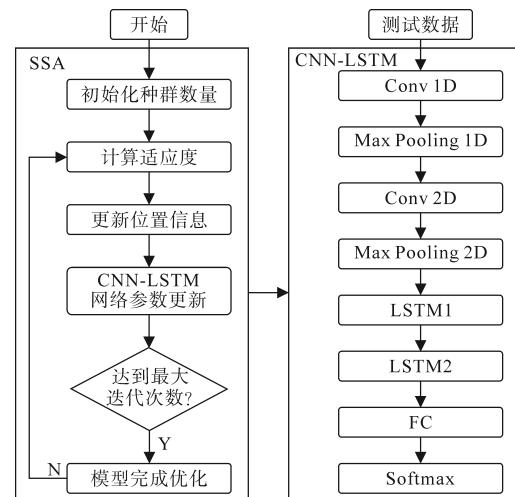


图 2 SSA-CNN-LSTM 模型结构

Fig.2 SSA-CNN-LSTM model structure

CNN 主要通过考虑前后数据输入的关联性, 赋予先前数据一定权重, 使网络对前面内容具有一定记忆能力。而 LSTM 主要解决了 CNN 在长时间中容易产生“梯度消失”或“梯度爆炸”的问题。LSTM 设置了 3 个门: 遗忘门、输入门和输出门。首先, 通过遗忘门  $f_t$  来决定去除掉时间序列中重要度低的信息,  $\sigma$  为激活函数,  $X_t$  为当前输入,  $H_{t-1}$  为上一时刻的输出。接着, 通过输入门决定进入细胞状态的信息量, 细胞状态的更新状态为  $C_t$ , 通过遗忘过去数据和对当前数据进行增强来更新细胞状态, 从而得到当前输入信息。最后, 可以得到该时间段的预测值, 计算出隐藏点的输出  $H_t$ 。

### 3 窃电识别流程

#### 3.1 识别流程

本文将 TimeGAN 与 SSA-CNN-LSTM 相结合, 实现不平衡二分类窃电用户识别。模型流程如图 3 所示。

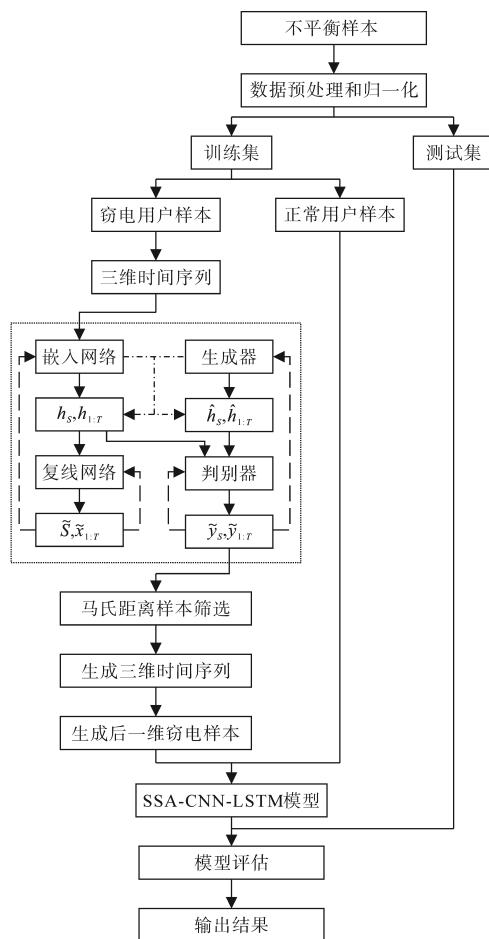


图 3 TimeGAN-CNN-LSTM 窃电行为识别结构

Fig.3 TimeGAN-CNN-LSTM electricity theft recognition structure

1) 归一化预处理后数据, 并划分训练集和测试集。归一化的公式为

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (12)$$

式中:  $X$  为原始数据;  $X_n$  为归一化后的数据;  $X_{\max}$  为原始数据的最大值;  $X_{\min}$  为原始数据的最小值。

2) 窃电用户数据少数类样本增广。将少数类训练样本数据由一维转化为三维时间序列, 输入到 TimeGAN 模型中, 学习更多的少数类, 利用马氏距离评估增广样本的质量, 完成不可信样本剔除, 达到总体样本平衡。

3) 用合成数据训练 SSA-CNN-LSTM 模型。利用 CNN 进行空间维度特征提取, 采用 LSTM 提取特征量的时序相关性, 利用 SSA 完成模型参数优化, 模型的损失函数为 Softmax。

#### 3.2 评价指标

用户窃电行为识别本质上是二分类问题, 分为正常用电数据和窃电数据。对于二分类常用的评价指标是精确率和召回率。对于测试集结果预测,  $TP$  表示将样本中正类别预测为正类的数目;  $FN$  表示将样本中正类别预测为负类的数目;  $FP$  表示将样本中负类别预测为正类的数目;  $TN$  表示将样本中负类别预测为负类的数目。在实验中将窃电用户类定义为正类, 将正常用户类定义为负类。精确率和召回率的定义分别为

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

由于两类样本差异较大, 从精确率表现去判断分类器分类性能, 会忽略掉少数类的预测结果。由此使用  $AUC$ (area under the curve) 值对分类性能进行定量判断。

## 4 实验与分析

### 4.1 数据预处理

本文采用国家电网公司发布的真实用电量数据<sup>[13]</sup>, 该数据集包括 1 035 天内(2014 年 1 月 1 日至 2016 年 10 月 31 日)42 372 个用电客户的用电数据量, 其中包含了 3 615 个窃电用户, 异常占比占总数的 9%, 少数类样本(窃电用户)和大类样本(正常用户)数量差异较大。图 4 分别展示了正常用户和窃电用户 1 个月的用电曲线, 其中正常用户不同周中每天用电趋势相近, 而窃电用户

不同周中每天用电趋势波动较大。实际分析验证中由于数据包含很多错误数据和缺失值,利用文献[13]中的预处理方法,利用插值方法来恢复缺失值,并对极端异常错误值进行剔除。由于神经网络对不同数据的敏感性,选择MAX-MIN标度法对预处理后数据进行归一化。

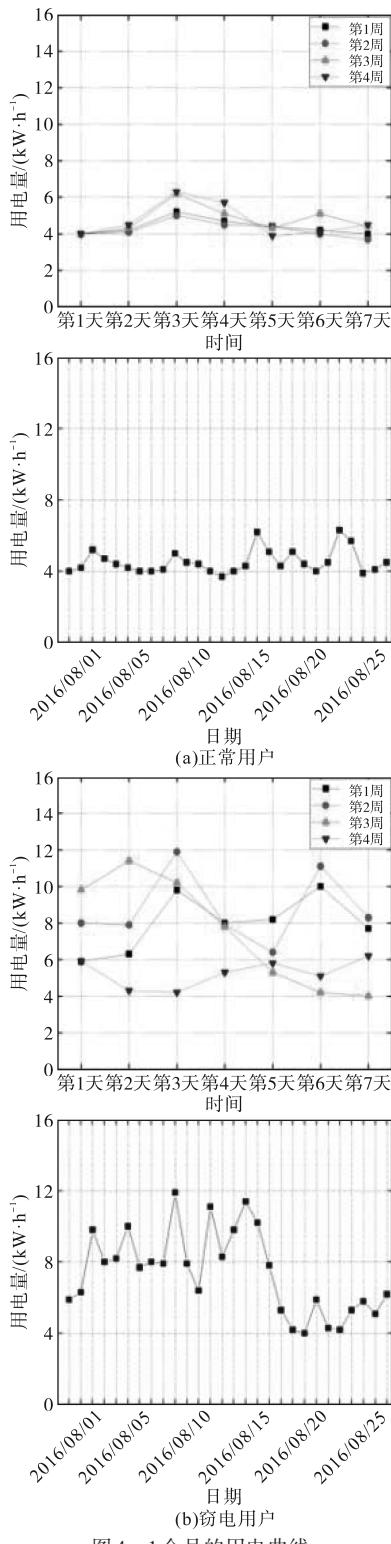


Fig.4 Electricity consumption curve for 1 month

## 4.2 参数设置

该实验中使用TensorFlow深度学习模型框架,TensorFlow版本为2.5,实验环境为CPU Intel Core i5-8300H, GPU NIVIDA GTX1060,内存为16 GB,操作系统为Windows 10 (64 bit)。通过实验观察模型在不同迭代次数下生成样本的质量以及与原始样本分布的拟合程度。在初步实验中,发现随着迭代次数增加,生成样本的稳定性和与原始样本的相似性逐渐提高。当迭代次数达到一定值后,继续增加迭代次数对模型性能提升不明显,且可能导致过拟合。因此,综合考虑模型性能和计算成本,最终设置GAN的迭代次数为6 000,其他设置如表1所示。

表1 TimeGAN模型参数表

Tab.1 TimeGAN model parameter table

AutoEncoder	参数	GAN模型	参数
采样步长	16	生成器噪声层数	32
特征维数	3	GAN隐藏层	20
隐藏层	1	激活函数	ReLU
学习率	5e-4	判别器真实损失	1

将经过扩充后的数据输入到SSA-CNN-LSTM模型中进行分类训练,参数如表2所示。为增强模型的泛化能力和抗干扰能力,引入批归一化并对第一层卷积核进行0.5的Dropout处理。实验中将总样本的70%作为训练集,剩余30%作为测试集。

表2 SSA-CNN-LSTM网络参数设置

Tab.2 SSA-CNN-LSTM network parameter settings

参数	取值	参数	取值
学习率	0.001	池化层取值	1×2
迭代次数	2 000	LSTM隐藏层节点	32
批尺寸	64	LSTM步长	10
卷积核1尺寸	1×64	全连接层	64
卷积核2尺寸	1×3	softmax层	4
卷积核1深度	16	优化算法	Adam
卷积核2深度	32	损失函数	交叉熵

## 4.3 检测结果

为验证本文算法的有效性,将提出的模型与现有经典少数类样本生成模型进行对比。分别利用自适应样本<sup>[14]</sup>、小类过采样<sup>[15]</sup>、随机欠采样<sup>[16]</sup>、GAN<sup>[10]</sup>和TimeGAN<sup>[11]</sup>及本文方法(TimeGAN-马氏距离筛选)与SSA-CNN-LSTM结合进行对比,精确率、召回率、AUC和准确率结果如表3所示。表3表明,TimeGAN数据生成方法取得了最好的效果,对比自适应样本生成方法准确率

提升了 2.8%, AUC 值提升了 12.2%; 对比小类过采样方法准确率提升了 1.9%, AUC 值提升了 10.9%; 对比随机欠采样方法准确率提升了 3.1%, AUC 值提升了 12.9%; 对比 GAN 方法准确率提升了 1.2%, AUC 值提升了 4.9%。

表 3 基于 SSA-CNN-LSTM 不同样本生成方法对比

Tab.3 Comparison of different sample generation methods based on SSA-CNN-LSTM

样本生成方法	精确率	召回率	AUC	准确率
自适应样本	0.615	0.863	0.671	0.934
小类过采样	0.632	0.872	0.683	0.942
随机欠采样	0.613	0.861	0.667	0.931
GAN	0.658	0.914	0.718	0.949
TimeGAN	0.672	0.932	0.742	0.951
本文方法	0.702	0.962	0.753	0.960

将本文提出模型与 TimeGAN-SSA-CNN, TimeGAN-SSA-LSTM 和 TimeGAN-CNN-LSTM 分别进行了对比来验证特征提取和分类的有效性。表 4 显示了不同模型在精确率、召回率、AUC 和准确率指标中的表现性能。表 4 表明, 本文提出的分类方法取得了最好的分类效果, 对比直接用 SSA-CNN 准确率提升了 1.3%, AUC 值提升了 4.7%; 对比 SSA-LSTM 方法准确率提升了 0.8%, AUC 值提升了 3.5%。其中, 含 SSA 模块在 4 个指标的表现优于 CNN-LSTM, 这说明 SSA 特征有助于判断窃电行为, 应在以后研究中充分考虑。

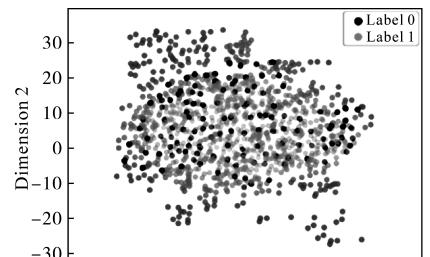
表 4 基于 TimeGAN 样本生成不同分类算法对比

Tab.4 Comparison of different classification algorithms based on TimeGAN sample generation

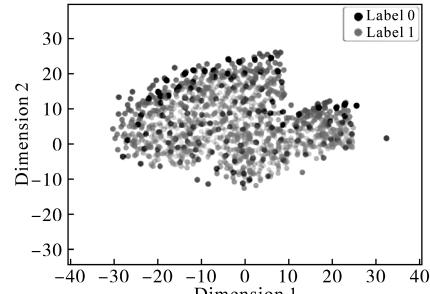
分类方法	精确率	召回率	AUC	准确率
SSA-CNN	0.675	0.928	0.719	0.948
SSA-LSTM	0.682	0.932	0.728	0.952
CNN-LSTM	0.671	0.922	0.681	0.922
SSA-CNN-LSTM	0.702	0.962	0.753	0.960

选取不同迭代次数(3 000 次和 6 000 次)通过 t-SNE 进行数据降维分析, 来验证所提少数类样本生成方法的有效性, 可视化的结果如图 5 所示。图中 Label0 深色点为窃电用户真实数据, Label1 浅色点群为通过 TimeGAN 生成的少数类样本数据。随着模型迭代次数的增加, 生成数据与真实样本的相似性增加, 外围拟合的边界逐渐清晰。当迭代次数为 6 000 时, 内部数据分散均匀, 尽管有部分生成点离群, 但整体生成数据较为稳定, 后续可以将 t-SNE 得到结果对 TimeGAN

模型进行实时反馈, 对模型进行进一步调优。



(a) 3 000 次迭代



(b) 6 000 次迭代

图 5 窃电用户两类数据可视化 t-SNE 图

Fig.5 Visualization of two types of data for electricity theft users t-SNE plot

## 5 结论

本文针对电网用户异常窃电行为检测, 提出了一种基于改进 TimeGAN 和 SSA-CNN-LSTM 的混合模型。其中, 利用 TimeGAN 网络对窃电用户数据进行生成, 生成时间序列数据, 并利用马氏距离对增广样本的质量进行评估, 剔除不可信样本。考虑到用户用电量行为的时序相关性, 采用 CNN-LSTM 对用电行为特征进行提取和分类, 利用 SSA 方法对模型参数进行优化来提升模型分类性能。通过实验对比证明了本文所提方法能够对电网中用户窃电行为进行精准识别。未来通过软硬件相结合, 利用深度学习技术融合智能硬件技术, 可以更好实现窃电行为实时监测。

## 参考文献

- [1] 梁键, 何江海, 杨海涛. 基于聚类算法的感应式机械电能表防窃电监测方法[J]. 自动化与仪表, 2023, 38(10):81-85.  
LIANG Jian, HE Jianghai, YANG Haitao. Monitoring method of induction mechanical watt hour meter against electricity stealing based on clustering algorithm[J]. Automation & Instrumentation, 2023, 38(10):81-85.
- [2] 吴蕊, 张安勤, 田秀霞, 等. 基于改进 K-means 的电力数据异常检测算法[J]. 华东师范大学学报(自然科学版), 2020(4): 79-87.

- WU Rui, ZHANG Anqin, TIAN Xiuxia, et al. Anomaly detection algorithm based on improved K-means for electric power data[J]. Journal of East China Normal University (Natural Science), 2020(4):79–87.
- [3] 耿俊成,张小斐,周庆捷,等.基于局部离群点检测的低压台区用户窃电识别[J].电网与清洁能源,2019,35(11):30–36.  
GENG Juncheng, ZHANG Xiaofei, ZHOU Qingjie, et al. A low voltage electricity theft identification method based on improved LOF[J]. Power System and Clean Energy, 2019, 35 (11):30–36.
- [4] 万磊,陈成,黄文杰,等.基于BRB和LSTM网络的电力大数据用电异常检测方法[J].电力建设,2021,42(8):38–45.  
WAN Lei, CHEN Cheng, HUANG Wenjie, et al. Power abnormality detection method based on power big data applying BRB and LSTM network[J]. Electric Power Construction, 2021, 42 (8):38–45.
- [5] 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 Nan, ZHANG Ying, SU Sheng, et al. User electric theft detection method based on hour scale periodic feature LSTM-transformer[J]. Power System Technology, 2023, 47(6) : 2558–2567.
- [6] ALTHOBAITI A, ROTSOS C, MARNERIDES A K. Adaptive energy theft detection in smart grids using self-learning with dual neural network[J]. IEEE Transactions on Industrial Informatics, 2024, 20(2):2776–2786.
- [7] 韩建富,肖春,宋小兵,等.基于GA-BP神经网络的能源互联网窃电行为识别方法[J].电气传动,2022,52(14):38–44.  
HAN Jianfu, XIAO Chun, SONG Xiaobing, et al. An identification method of electricity theft on energy internet based on GA-BP neural network[J]. Electric Drive, 2022, 52(14):38–44.
- [8] 肖进胜,王澍瑞,吴原硕,等.时空关键区域增强的小样本异常行为识别[J].计算机学报,2025,48(1):68–81.  
XIAO Jinsheng, WANG Shurui, WU Yuanxu, et al. Anomalous action recognition with spatio-temporal key region enhancement and few-shot learning[J]. Chinese Journal of Computers, 2025, 48(1):68–81.
- [9] 王欢,王昕,张峰,等.基于改进生成对抗网络的变压器声纹故障诊断[J].智慧电力,2024,52(4):24–31.  
WANG Huan, WANG Xin, ZHANG Feng, et al. Transformer voiceprint fault diagnosis based on improved generative adversarial network[J]. Smart Power, 2024, 52(4):24–31.
- [10] 杨艺宁,张蓬鹤,夏睿,等.基于CT-GAN的半监督学习窃电检测方法研究[J].湖南大学学报(自然科学版),2024,51 (6):211–222.  
YANG Yining, ZHANG Penghe, XIA Rui, et al. Research on semi-supervised learning detection method of electricity theft based on CT-GAN[J]. Journal of Hunan University (Natural Sciences), 2024, 51(6):211–222.
- [11] 张丽娜,陈会娟,余昭旭.基于TimeGAN-CNN-LSTM模型的河流水质预测研究[J].自动化仪表,2022,43(8):11–15.  
ZHANG Lina, CHEN Huijuan, YU Zhaoxu. Research on river water quality prediction based on TimeGAN-CNN-LSTM model [J]. Process Automation Instrumentation, 2022, 43(8):11–15.
- [12] 王毅,郑宏志,黄欣,等.基于多阶段调度框架的麻雀搜索优化算法[J].电子学报,2024,52(9):3086–3096.  
WANG Yi, ZHENG Hongzhi, HUANG Xin, et al. Sparrow search optimization algorithm based on multi-stage scheduling framework[J]. Acta Electronica Sinica, 2024, 52 (9) : 3086–3096.
- [13] ZHENG Z, YANG Y, NIU X, et al. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids[J]. IEEE Transactions on Industrial Informatics, 2018, 14 (4):1606–1615.
- [14] 王骏,梁东,高迪,等.基于改进随机森林算法的低压用户窃电行为自适应监测方法[J].微型电脑应用,2024,40(6): 202–205, 210.  
WANG Jun, LIANG Dong, GAO Di, et al. An adaptive monitoring method based on improved random forest algorithm for power stealing behavior of low-voltage users[J]. Microcomputer Applications, 2024, 40(6):202–205, 210.
- [15] 谷铮,陈学斌,张宏扬,等.基于凝聚式层次聚类的微调筛选过采样方法[J].计算机应用,2025,45(7):2138–2144.  
GU Zheng, CHEN Xuebin, ZHANG Hongyang, et al. A fine-tuned screening oversampling method based on cohesive hierarchical clustering[J]. Journal of Computer Applications, 2025, 45 (7):2138–2144.
- [16] APPIAH S Y, AKOWUAH E K, IKPO V C, et al. Extremely randomised trees machine learning model for electricity theft detection[J]. Machine Learning with Applications, 2023, 12: 100458.

收稿日期:2024-11-12

修改稿日期:2024-12-19