Modeling and Simulation 建模与仿真, 2024, 13(2), 1546-1555
Published Online March 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.132146

Hans汉斯

考虑样本不平衡的窃电检测模型

戴　宇，张　巍

上海理工大学机械工程学院，上海

收稿日期：2024年1月27日；录用日期：2024年3月20日；发布日期：2024年3月27日

摘　要

随着全社会用电量持续增加，窃电问题日益严重，给电力企业造成巨大的经济损失，并影响电网的运行。与此同时，窃电样本不足导致基于大数据的窃电检测方法受到限制，本文针对现实情况中窃电案例收集困难、数量稀少问题，提出了一种考虑样本不平衡的窃电检测模型。首先，通过生成对抗网络(GAN)生成器与判别器的对抗训练，学习窃电数据的时序相关性，制造出与窃电样本近似的样本，使得窃电数据集中正负样本趋于平衡。然后结合卷积神经网络、长短期记忆递归神经网络和注意力机制(CNN-LSTM-Attention)对用户进行窃电检测，将经过样本不平衡处理后的用户用电信息经过CNN进行特征提取，通过LSTM捕捉数据的时序变化信息，使用Attention对LSTM的输出赋予权重，强化有利于窃电检测的特征数据，弱化无关数据。算例分析表明，本文提出的方法能有效避免样本不平衡问题，更好地检测出用户窃电行为。

关键词

窃电检测，生成对抗网络，卷积神经网络，长短期记忆递归神经网络，注意力机制

Electricity Theft Detection Model Considering Sample Imbalance

Yu Dai, Wei Zhang

School of Mechanical Engineering, University of Shanghai for Science & Technology, Shanghai

Received: Jan. 27th, 2024; accepted: Mar. 20th, 2024; published: Mar. 27th, 2024

Abstract

With the continuous increase of electricity consumption in the whole society, the problem of power theft is becoming more and more serious, which causes huge economic losses to electric power enterprises and affects the operation of power grids. At the same time, the lack of power theft samples leads to the limitation of big data-based power theft detection methods. In this paper, we

文章引用: 戴宇, 张巍. 考虑样本不平衡的窃电检测模型[J]. 建模与仿真, 2024, 13(2): 1546-1555.
DOI: 10.12677/mos.2024.132146

propose a power theft detection model considering sample imbalance to address the problem of difficulty in collecting and scarcity of power theft cases in the real situation. Firstly, the temporal correlation of electricity theft data is learned through the adversarial training of generator and discriminator of Generative Adversarial Network (GAN), which creates samples that are close to the electricity theft samples, so that the positive and negative samples in the electricity theft dataset tend to be balanced. Then combine the convolutional neural network, long and short-term memory recurrent neural network and attention mechanism (CNN-LSTM-Attention) to detect power theft to the user, after the sample imbalance processing of the user's electricity consumption information through the CNN for feature extraction, through the LSTM to capture the temporal change of the data information, the use of Attention on the output of the LSTM to give weight, and strengthen the features that are favorable to the detection of power theft. Strengthen the feature data conducive to power theft detection and weaken the irrelevant data. Case analysis shows that the method proposed in this paper can effectively avoid the sample imbalance problem and better detect the user's power theft behavior.

Keywords

Electricity Theft Detection, Generative Adversarial Network, Convolutional Neural Networks, Long Short Term Memory, Attention Model

1. 引言

随着经济的不断发展，社会用电量持续攀升，伴随而来的问题是日益加剧的用户窃电现象。这一趋势对电力企业构成了严重的经济威胁，根据调查，全球各国家都因窃电问题导致了较大经济损失，如美国曾因窃电造成的经济损失超过 60 亿美元[1]，中国约 5.6 亿美元[2]，英国约 2.34 亿美元[3]，因窃电造成的异常数据还会影响调度决策从而直接威胁到整个电网的正常运行[4]。

近些年来高精度智能电表提供的详细用户用电信息结合大数据分析技术成为了有效的窃电检测方法。文献[5]利用先进计量基础设施(Advanced Metrology Infrastructure, AMI)收集到的用电数据通过主成分分析(Principal Component Analysis, PCA)进行特征提取并使用随机森林(Random Forest)进行窃电检测。然而由于现实世界中窃电样本的不足，导致的样本不平衡问题一直阻碍着基于大数据的窃电检测方法发展，若窃电样本仅占数据集的 10%，即使模型判定大部分样本为正常用户模型也能获得 90%的准确率。对此，文献[6]采用合成少数类过采样技术(Synthetic Minority Oversampling Technique, SMOTE)对窃电样本进行过采样操作，然而由于其是根据线性插值的原理，所生成的数据与用电数据有所差异。文献[7]采用自适应合成抽样(Adaptive Synthetic Sampling, ADASYN)平衡窃电数据集，同样不能很好模拟真实窃电样本。

针对大数据检测窃电行为中窃电样本稀少问题，本文提出了一种考虑样本不平衡的窃电检测模型，首先使用生成对抗网络(Generative Adversarial Network, GAN)中生成器和判别器的对抗训练模拟生成窃电样本，使得窃电数据集中正负样本平衡，然后基于卷积神经网络(Convolutional Neural Networks, CNN) [8]、长短期记忆递归神经网络(Long Short Term Memory, LSTM) [9]和注意力机制相结合(CNN-LSTM-Attention)进行窃电检测。该模型首先使用 GAN 中的生成器生成随机窃电样本，判别器负责区分真实和生成窃电样本，两者通过不断对抗训练得到与真实窃电样本相似的模拟窃电样本，由此解决样本不平衡

问题。再使用 CNN 对用户用电信息进行初步特征提取，通过带有记忆功能的 LSTM 模型提取时序变化信息，最后使用注意力机制赋予 LSTM 输出层权重，从而进一步降低模型的检测误差，提高模型的检测精度。为电力企业提供高效、精确的窃电检测模型。

2. 用户窃电行为分析

用户窃电行为的最终目的是为了少交、漏交电费，其使用的方法大多表现为用电量的减少或利用相关政策少交电费。根据电力企业统计的用户窃电行为，窃电用户使用的主要窃电手法主要可分为六类[10]，如表 1 所示。第一类和第二类用户主要通过用电高峰期电价和低谷期电价的差异，在不改变总用电量的情况下对用电曲线进行移峰，减小用电高峰期时的耗电量。$M_1(\cdot)$ 表示该天用电记录取其平均值，可以减少高峰期用电量；$M_2(\cdot)$ 表示将改天用电记录倒序记录。

Table 1. Main types of electricity theft
表 1. 主要窃电方式

序号	窃电方式	参数说明
1	$M_1(x_t)=\text{mean}(x_t)$	/
2	$M_2(x_t)=x_{48-t}$	/
3	$M_3(x_t)=f(t)\cdot x_t$	$f(t)=\begin{cases}0 & t_{start}<t<t_{end}\\ 1 & \text{otherwise}\end{cases}$
4	$M_4(x_t)=\max(x_t-\gamma,0)$	$0<\gamma<\max(x_t)$
5	$M_5(x_t)=\alpha_t\cdot x_t$	$\alpha_t=\text{random}(0.1,0.8)$
6	$M_6(x_t)=\begin{cases}x_t & x_t\le\beta\\ \beta & x_t\ge\beta\end{cases}$	$\beta=\text{random}(0,\max(x_t))$

第三类和第四类窃电用户通过篡改特定时间内的用电数据以达到窃电目的。第三类用户一般通过利用控制开关在用电高峰期或者被发现风险较低期间间歇性地将用电量置为 0，第四类用户通常将用电量减少一个固定的值以减少总用电量。$M_3(\cdot)$ 表示该天用电量在(t_{start},t_{end})内取 0，可以模拟一段时间内用电量置 0 的情况；$M_4(\cdot)$ 表示每一时刻用电量减去 γ，若小于 0 则取 0 可以模拟用电量不间断置 0 的情况。

第五类和第六类用户是将用电量减少到一定程度，其中第五类用户可能通过单相/两相分流或更换互感器将用电记录量随机按比例减少，第六类用户通过设定一个阈值，超过该阈值的用电量则记录为阈值。$M_5(\cdot)$ 表示该天的用电量乘以(0.1, 0.8)内的随机数，用以模拟按照相同比例削减用电量；$M_6(\cdot)$ 表示将超过阈值的用电量设为阈值，用以模拟随机削减用电量。

用户的用电行为会受天气、节日等因素的影响，但整体而言，在一定的地理区域内，同一类型的用户往往表现出相似的用电模式，因此，通过大数据分析，便可以找出其中具有显著不同用电行为的窃电用户，帮助电力企业进行稽查。

3. 研究方法及检测模型的构建

3.1. 基于 GAN 的窃电样本扩充方法

GAN 是一种由生成器和判别器组成的对抗网络，可用于学习用户用电数据中复杂的时序相关特性，目前已在图像生成、一维震动信号生成等领域使用，能够很好的解决相关领域样本获取困难问题。其生成器(Generator, G)的主要作用是能够根据原始输入的特点，将随机噪声映射到真实样本的分布上，制造

出于其含有同样特性的样本，并使得判别器(Discriminator, D)无法判别其与真实样本的真伪，GAN 模型结构图可参见图 1。

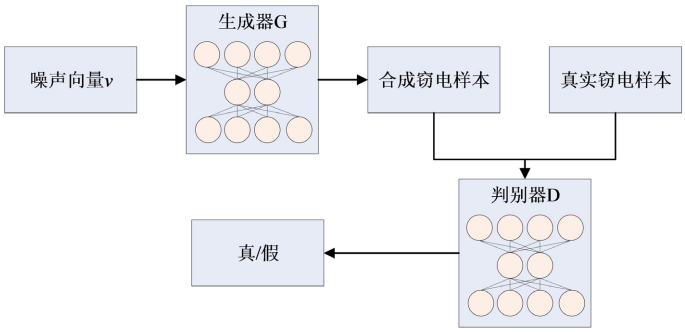

Figure 1. Diagram of the structure of the GAN model
图 1. GAN 模型结构图

假设一组满足高斯分布 P_v 的向量 v，生成器 G 接受输入 v，通过生成器 G 中的神经网络将噪声映射到样本分布上，合成初期窃电样本，尽量使判别器 D 无法判别样本是生成的还是真实样本，其目标函数可表示为：

$$L_G = \min - E_{v\sim P_v}\left[D\left(G\left(v\right)\right)\right] \tag{1}$$

式中，$G\left(v\right)$ 表示由生成器生成的窃电样本；$D\left(G\left(v\right)\right)$ 表示生成的样本被判别器判别为实际样本的概率，当生成的样本越接近真实样本时，$D\left(G\left(v\right)\right)$ 就越接近 1，此时目标函数 L_G 能取得最小值。

而判别器 D 的认识则是尽可能识别出生成样本，假设真实窃电样本 x_{data} 中符合某种分布关系 p_w，判别器 D 要尽量识别生成样本和窃电样本，其函数可以定义为：

$$L_D = \max E_{x_{data}\sim p_w}\left[D\left(x_{data}\right)\right] - E_{z\sim p_z}\left[D\left(G\left(v\right)\right)\right] \tag{2}$$

式中，$D\left(x_{data}\right)$ 表示真实样本被判别器 D 判为真实样本的概率。

GAN 整体的训练过程是一个对抗博弈的过程，其中生成器需要尽量学习窃电样本中的关键信息来生成类似的样本来迷惑判别器，而判别器需要学习真实样本与窃电样本之间的差别来区分两者，其总体的目标函数可以定义为：

$$\min_G \max_D E_{x_{data}\sim p_w}\left[D\left(x_{data}\right)\right] - E_{v\sim P_v}\left[D\left(G\left(v\right)\right)\right] \tag{3}$$

一般来讲，在训练 GAN 网络时，一般会固定一个网络来训练另一个网络进行往复训练，如更新生成器 G 时，判别器 D 是固定的，而当更新判别器 D 时，生成器 G 是固定的，以此最终 GAN 网络能够生成出与真实窃电样本极为相似的样本。基于 GAN 的样本平衡流程图如图 2 所示。生成器通过输入的随机噪声生成初期窃电样本，并与真实样本一起输入判别器。判别器根据真实样本与生成样本之间的差异，获得损失函数。生成器通过判别器的反馈来更新优化自身的参数，以减小生成样本与真实样本之间的差距，并使判别器无法辨别，当判别器无法辨别时，判别器更新优化自身参数，以提高对生成样本与真实样本的判别能力，两者不断更新优化参数，达到最终的迭代次数后输出生成样本。

3.2. 基于 CNN-LSTM-Attention 的窃电检测模型

3.2.1. 卷积神经网络

对用户的用电信息进行特征提取是判定窃电行为的关键一步。CNN 具有稀疏连接和全局共享的特点，

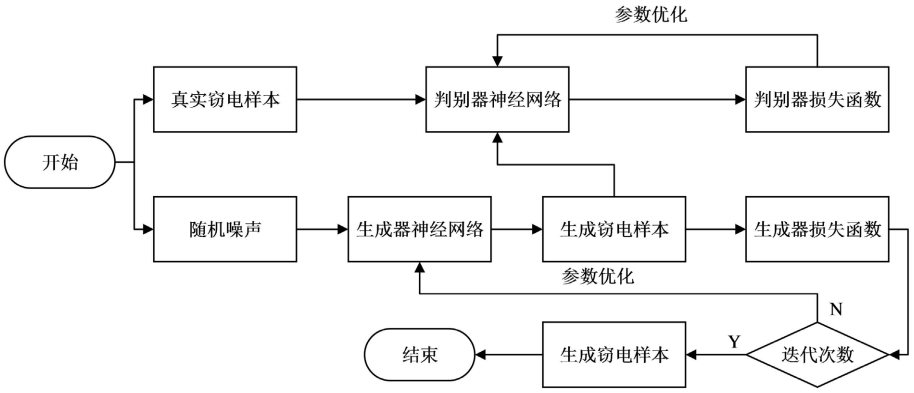

Figure 2. Flowchart of GAN model training
图 2. GAN 模型训练流程图

其主要由卷积层、池化层、全连接层和输出层组成，这个结构使其能够有效地从输入的用户用电数据中提取关键特征。具体来说，CNN 模型的卷积层通过对输入数据进行卷积操作，可以捕捉到数据中的局部特征，卷积过程的表达式如下式所示：

$$x_j^l = f\left(\sum_{i\in M_j} x_j^{l-1} * k_{ij}^{l-1} + b_j^l\right) \tag{4}$$

式中，*表示卷积操作；x_j^l 为 l 个卷积层的 j 个输出数据；k_{ij}^l 为该卷积层的卷积核；b_j^l 为该卷积层的偏置；$f(\cdot)$ 为 ReLU 函数，$f(x)=\max(0,x)$。

而池化层则通过降采样的方式减少数据的维度，减少模型运算时间并保留最显著的特征，一般选用最大池化作为池化方式。全连接层负责整合这些局部和全局的特征信息，如下式所示：

$$y = wx + b \tag{5}$$

式中，x 为全连接层的输入；w 为权值矩阵；b 为偏置。

3.2.2. 长短期记忆递归神经网络

通过 CNN 特征提取后的特征仍具有时序数据的特点，LSTM 通过遗忘门、输入门和输出门来对数据中的关键信息进行长期记忆，更能关注非线性数据中时序变化的信息。LSTM 模型结构图可参见图 3。

LSTM 中遗忘门负责选择性遗忘部分元素，避免过多记忆影响神经网络对输入数据的处理，如下式所示：

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{6}$$

式中，x_t 为 t 时刻输入向量；f_t 和 h_t 为 t 时遗忘门的激活值和隐藏层状态值；W_f 和 U_f 为输入与隐藏层的权重；b_f 为偏置；$\sigma(\cdot)$ 为 Sigmoid 函数，$\sigma(x)=\dfrac{1}{1+\mathrm{e}^{-x}}$。

输入门用来控制当前时刻所输入的数据，其中包括使用 Sigmoid 函数控制信息以及 tanh 层生成更新向量，将这两部分结合起来对一个细胞进行更新，如下式所示：

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{7}$$

$$m_t = \tanh\left(W_m x_t + U_c h_{t-1} + b_m\right) \tag{8}$$

式中，$\tanh(\cdot)$ 为双曲正切函数，$\tanh(x)=\dfrac{e^x - e^{-x}}{e^x + e^{-x}}$。

根据式(6) (7) (8)即可得到更新后的神经元状态，如下式所示：

$$C_t = f_t \cdot C_{t-1} + i_t \cdot m_t \tag{9}$$

式中，C_t 为隐藏状态的值；$\cdot$ 为向量的点积运算。

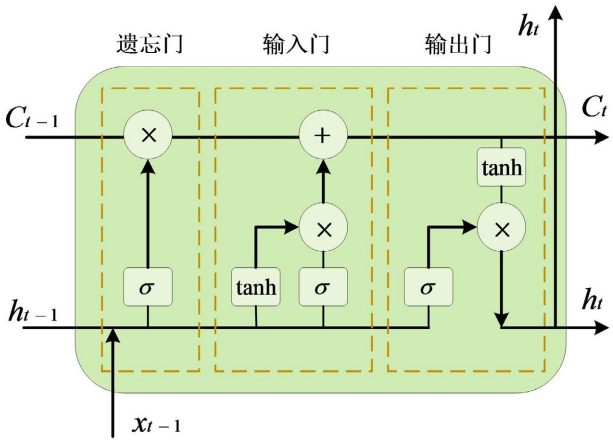

Figure 3. Diagram of the structure of the LSTM model
图 3. LSTM 模型结构图

输出门则通过 Sigmoid 和 tanh 函数来确定最终输出的值，其式如下所示：

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{10}$$

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{11}$$

3.2.3. 注意力机制

由于 CNN 和 LSTM 所提取的特征可能出现冗余的情况[11]，会干扰模型训练时的收敛性甚至精度，为了提升模型的工作效率以及检测准确度，本文引入注意力机制(Attention Module, AM)将有限的计算资源分配到更重要的特征信息上。

AM 首先通过得分函数计算出各项输入的权重值，得到其注意力分布，然后通过加权平均求出其单个输出值的信息。其得分函数如下所示：

$$p\left(x_i, q\right) = x_i^T W q \tag{12}$$

式中，x_i 为输入值；q 为神经网络的查询向量；W 为神经网络的参数。

取得分函数的 Softmax 值，即可得到对应的权重值，如下式所示：

$$k_i = s\left(p\left(x_i, q\right)\right) \tag{13}$$

$$k = \sum_{i=1}^{N} k_i x_i \tag{14}$$

式中，k_i 为对应的权重值；k 为加权平均值；$s(\cdot)$ 为 Softmax 函数，$s(x) = \dfrac{e^{x_i}}{\sum_i e^{x_i}}$。

3.3. 考虑样本不平衡的窃电检测流程

本文针对窃电检测中由于现实中窃电样本稀少导致窃电检测结果不够理想，采用 GAN 模拟生成窃电

样本，使得窃电数据集中正负样本趋于平衡，使模型能够充分学习窃电数据特征，而做出更加精准的判断。并基于 CNN-LSTM-Attention 进行窃电检测，图 4 为其检测流程图。

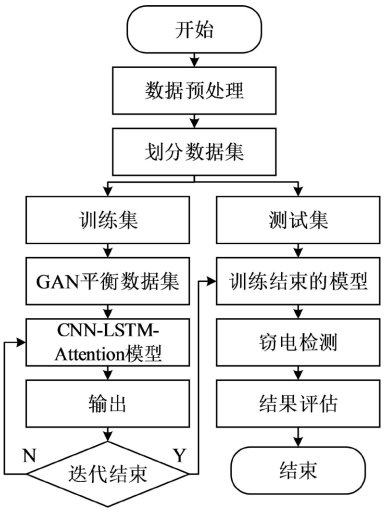

Figure 4. Diagram of the structure of the CAM model
图 4. 窃电检测流程图

首先，将经过预处理的数据按照一定比例划分为训练集和测试集。随后，将经过预处理的用户用电训练集数据输入到 GAN 模型中，通过博弈对抗生成一定数量的模拟窃电样本，使得训练集中正负样本数量平衡，使模型不会一味关注正常用电数据，同样也给予窃电数据一定的重视程度，能够更好地发现窃电数据中的特征。再通过 CNN-LSTM-Attention 模型经 CNN 的特征提取、LSTM 时间序列相关性分析，以及 AM 的赋值。进一步提高了模型对关键信息的关注度，从而提升了检测的准确性。最终，输出层输出模型的预测结果。当模型的迭代训练完成时，将测试集的数据送入已训练完成的模型中，进行窃电用户的检测。验证其泛化能力和实用性。

4. 算例结果与分析

4.1. 数据集

在算例实验中，本文采用了广泛应用与窃电检测的爱尔兰智能电表数据，该数据集中收集了爱尔兰 6000 多个住宅和商业用户连续一年多的用电记录，其时间分辨率为 30 min 记录一个数据[12]，符合大数据分析的特点。在剔除不良数据和缺失数据后，以天为单位选择 10,000 天的用电记录，并根据第二章所分析的窃电行为将 10,000 个样本中随机 1200 个样本生成 6 中窃电模式各 200 条样本，正负样本比例为 10:1.2，是严重的样本不平衡数据集，其典型用户如图 5 所示。

由于窃电检测在多数情况下只需要判定用户窃电与否，再交由电力企业工作人员进行线下核实，所以窃电检测事实上可以看作是一个二分类问题。本文采用窃电检测准确率 ACC 和 F1-Score 作为评价指标，F1-Score 是一个适用于二分类问题的评价指标，其能客观表现出模型的性能，尤其是在正负样本不平衡的情况下，而这恰恰符合窃电行为检测这一窃电样本较少的情景，其式如下所示：

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$F1 = 2\frac{precision \cdot recall}{precision + recall} \tag{16}$$

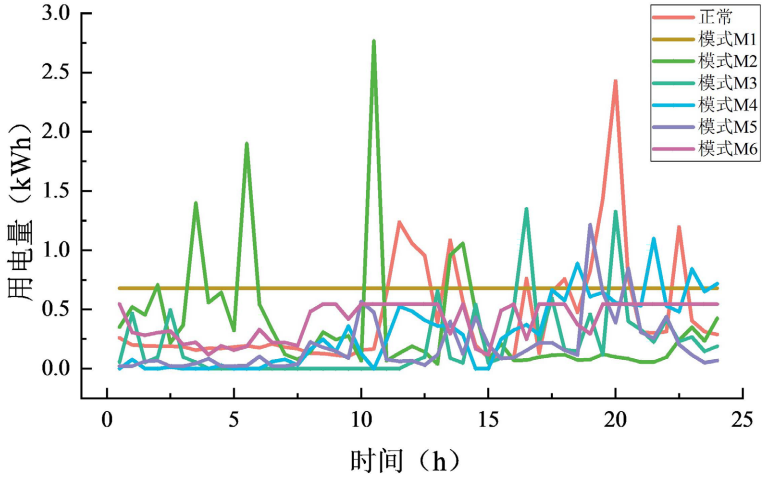

Figure 5. Various types of typical users of electricity consumption curve diagram
图 5. 各类典型用户用电曲线图

式中，*TP* 表示将正常用户判定为正常用户的数量；*TN* 表示将窃电用户判定为窃电用户的数量；*FP* 表示将窃电用户判定为正常用户的数量；*FN* 表示将正常用户判定为窃电用户的数量；*precision* 为精准度，表示被模型判定为正常用户中真实正常用户的比例，$precision = \frac{TP}{TP + FP}$；*recall* 为召回率，表示被判定为正常用户的数量占真实正常用户的比例，$recall = \frac{TP}{TP + FN}$。

4.2. 算例分析

算例采用的硬件环境为 inter(R) Xeon(R) W-2245 的 CPU，32 GB 内存，RTX4000 8 + 16 GB 的 GPU，使用 python 语言进行代码编写。其中各模型结构如表 2 所示。

Table 2. Table of the structure of each model
表 2. 各模型结构表

模型		结构
GAN	生成器 G	Input-Linear-ReLU- Linear-ReLU- Linear-Sigmoid
	判别器 D	Input-Linear-ReLU- Linear-ReLU-Flatten-Output
CNN-LSTM-Attention		Conv1D-MaxPooling1D-Conv1D-MaxPooling1D-Dropout-LSTM-Attention -Dense- Dense

随机选取 60%数据作为训练集用以生成窃电样本和训练检测模型，20%作为测试集用以评价检测模型优劣，剩余 20%数据作为验证集用于调检测整模型参数。每一轮训练后使用测试集测试模型故障识别准确率并通过反向传播优化模型参数。GAN 模型的迭代次数设置为 300 次，检测模型的学习率设置为 0.01，迭代次数设置为 100 次。首先使用 GAN 对训练集中的窃电样本进行生成以保证训练集正负样本平衡，GAN 能够近似学习到原始窃电数据中细微的时序信息，在训练初期可能还会离窃电曲线有所差异，随着迭代次数的增加，其模拟出的曲线会逐渐接近于真实窃电曲线，图 6 展示了其训练 50 次、100 次和 300 次与真实窃电样本的对照图。由图可以看出当迭代 50 次时，GAN 还不能很好把握生成样本与窃电样本之间的差异，迭代 100 次时，GAN 生成的窃电样本已经和窃电样本有所相似，当迭代到 300 次时，GAN

所生成的窃电样本已经与真实的窃电样本基本一致，但又加入了一定的随机性，可以有效扩充训练集中窃电样本，使得正负样本平衡，由样本平衡的训练集所训练出的检测网络才能够得到更好的效果。

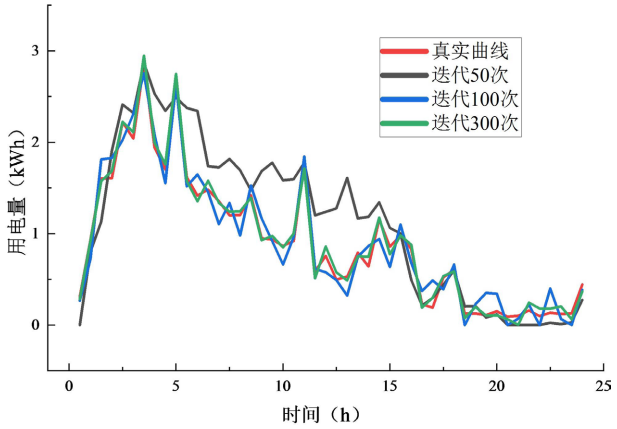

Figure 6. Comparison of GAN-generated electricity theft samples with real electricity theft samples
图 6. GAN 生成窃电样本与真实窃电样本对照图

使用经过样本平衡处理后的训练集训练 CNN-LSTM-Attention 窃电检测模型，将验证集送入模型验证其检测效果。为有效评估本文所提出解决样本不平衡方法的有效性，将其与经 SMOTE、ADASYN 算法进行样本平衡以及未进行样本平衡的数据集进行对比。其检测结果如图 7 所示，由图 7 可以看出，经过 GAN 样本平衡后的数据集用以进行窃电检测能够得到更好的检测效果，其在 ACC 上能够达到 96.6%，F1-Score 达到 98.01%。相较于 SMOTE 和 ADASYN 算法，GAN 能够获得更好的检测结果，说明 GAN 能够很好的学习窃电样本中的时序分布，并模拟生成高度相似的窃电样本供检测模型学习，使得模型能够更加关注到窃电样本，即能更好的检测出窃电用户。相比未经处理数据集的 91.8%的 ACC 和 95.2 的 F1-Score 分别提高了 4.8%和 2.81%，说明在实际窃电检测过程中是十分有必要考虑样本不平衡问题的，其正负样本的不平衡会严重干扰到基于大数据的窃电检测结果，而 GAN 作为样本平衡的方法，能够更好地解决这一问题。文章所采用的 GAN 样本平衡方法在窃电检测中表现出色，为提高检测模型的性能提供了有力的支持。

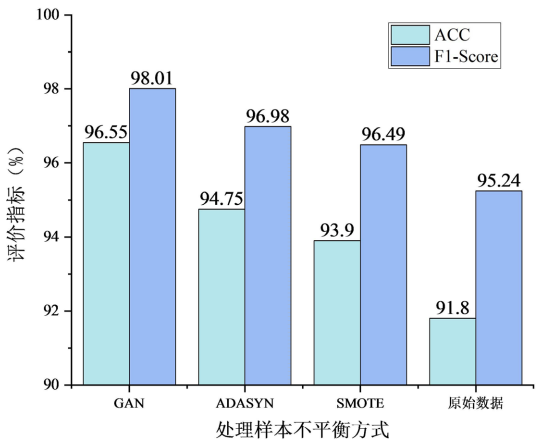

Figure 7. ACC for each model and F1-Score
图 7. 各模型的 ACC 和 F1-Score 值

5. 总结

文章考虑现实中窃电样本不平衡问题，采用 GAN 进行样本平衡再使用 CNN-LSTM-Attention 进行窃电检测，GAN 在训练 300 次后能够根据用户历史用电信息，生成高度相似的窃电数据，平衡数据集中正负样本。经过样本平衡后的数据集所训练出的 CNN-LSTM-Attention 具有较高的窃电检测能力，检测准确率达到 96.55%，较未经样本平衡的数据集准确率提高了 4.75%。所检测出的可疑用户能够方便后续相关部门的稽查。该项研究为电力企业基于大数据稽查窃电用户提供了一定的参考价值。

参考文献

[1] McDaniel, P. and McLaughlin, S. (2009) Security and Privacy Challenges in the Smart Grid. *IEEE Security & Privacy*, **7**, 75-77. https://doi.org/10.1109/MSP.2009.76

[2] Ismail, M., Shaaban, M.F., Naidu, M., *et al*. (2020) Deep Learning Detection of Electricity Theft Cyber-Attacks in Renewable Distributed Generation. *IEEE Trans Smart Grid*, **11**, 3428-3437. https://doi.org/10.1109/TSG.2020.2973681

[3] Jokar, P., Arianpoo, N. and Leung, V. (2017) Electricity Theft Detection in AMI Using Customers' Consumption Patterns. *IEEE Trans Smart Grid*, **7**, 216-226. https://doi.org/10.1109/TSG.2015.2425222

[4] 赵海波. 电力行业大数据研究综述[J]. 电工电能新技术, 2020, 39(12): 62-72.

[5] 刘文浩, 冯玥, 姜东良. 基于 AMI 数据驱动的窃电用户识别研究[J]. 制造业自动化, 2022, 44(11): 5-8.

[6] Figueroa, G., Chen, Y., Avila, N.F., *et al*. (2017) Improved Practices in Machine Learning Algorithms for NTL Detection with Imbalanced Data. 2017 *IEEE Power & Energy Society General Meeting*, Chicago, 16-20 July 2017, 1-5. https://doi.org/10.1109/PESGM.2017.8273852

[7] Aldegheishem, A., Anwar, M., Javaid, N., *et al.* (2021) Towards Sustainable Energy Efficiency with Intelligent Electricity Theft Detection in Smart Grids Emphasising Enhanced Neural Networks. *IEEE Access*, **9**, 25036-25061. https://doi.org/10.1109/ACCESS.2021.3056566

[8] 黄朝凯, 吴丹妍, 郑惠哲, 等. 基于 CNN 的电力数据分析模型研究[J]. 自动化仪表, 2023, 44(10): 65-69+74.

[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.

[10] Punmiya, R. and Choe, S. (2019) Energy Theft Detection Using Gradient Boosting Theft Detector with Feature Engineering-Based Preprocessing. *IEEE Transactions on Smart Grid*, **10**, 2326-2329. https://doi.org/10.1109/TSG.2019.2892595

[11] Han, K., Wang, Y., Tian, Q., *et al*. (2020) Ghostnet: More Features from Cheap Operations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 1580-1589. https://doi.org/10.1109/CVPR42600.2020.00165

[12] Commission for Energy Regulation (CER) (2012) CER Smart Metering Project-Electricity Customer Behaviour Trial, 2009-2010 [Dataset]. Irish Social Science Data Archive. https://www.ucd.ie/issda/data/commissionforenergyregulationcer/