文章编号:1674-2974(2024)06-0211-12 DOI:10.16339/j.cnki.hdxbzkb.2024241

基于CT-GAN的半监督学习窃电检测方法研究

杨艺宁[1],张蓬鹤[1],夏睿[2],高云鹏[2†],王飞[3],朗珍白桑[3]
(1. 中国电力科学研究院有限公司, 北京 100192;
2. 湖南大学 电气与信息工程学院,湖南 长沙 410082;
3. 国网西藏电力有限公司, 西藏 拉萨 850000)

摘 要:针对电网公司获取有标签数据成本高、难度大,而获取的无标签数据难以训练有效窃电检测模型的问题,提出了在少量有窃电标签数据场景下基于联合训练生成对抗网络(Co-training Generative Adversarial Networks, CT-GAN)的半监督窃电检测方法.首先,探究了生成对抗网络及半监督生成对抗网络的原理与结构.其次,提出了采用Wasserstein距离取代JS(Jensen-Shannon)散度和KL(Kullback-Leibler)散度距离以解决生成对抗网络因梯度消失和模式崩溃原因导致的模型训练不稳定和生成数据质量低的问题,并构建了多判别器联合训练模型,避免了单个判别器分布误差高的问题,同时增强了GAN生成标签样本数据的能力,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.最后,采用爱尔兰电网数据集验证了该方法的准确性和有效性.

关键词:窃电检测;生成对抗网络;半监督学习;Wasserstein距离;判别器
中图分类号:TM715 **文献标志码**:A

Research on Semi-supervised Learning Detection Method of Electricity Theft Based on CT-GAN

YANG Yining[1], ZHANG Penghe[1], XIA Rui[2], GAO Yunpeng[2†], WANG Fei[3], Langzhenbaisang[3]
(1. China Electric Power Research Institute Co., Ltd., Beijing 100192, China;
2. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China;
3. State Grid Xizang Electric Power Co., Ltd., Lasa 850000, China)

Abstract: Aiming at the high cost and difficulty of obtaining labeled data for power grid companies, and the difficulty of training an effective electricity theft detection model with unlabeled data, this paper proposes a method based on CT-GAN (Co-training Generative Adversarial Networks) semi-supervised electricity theft detection method. Firstly, the principles and structures of generative adversarial networks and semi-supervised generative adversarial networks are explored. Secondly, it is proposed to replace the JS (Jensen-Shannon) divergence and KL (Kullback-Leibler) divergence distance with the Wasserstein distance to solve the problem of unstable model

* **收稿日期**:2023-09-11
基金项目:中国电力科学研究院研究开发项目(JL8422-003),Research and Development Project of China Electric Power Research Institute (JL8422-003)
作者简介:杨艺宁(1996—),女,山东菏泽人,中国电力科学研究院有限公司工程师
† 通信联系人,E-mail:gaoyp@hnu.edu.cn

training and low quality of generated data caused by the gradient disappearance and mode collapse of the generative confrontation network problem, and built a multi-discriminator Co-training model to avoid the problem of high distribution error of a single discriminator. At the same time, it enhanced the ability of GAN to generate label sample data. By expanding the label sample data set, the model detection accuracy and generalization ability were improved. Finally, the accuracy and effectiveness of the method are verified using the Irish power grid dataset.

Key words: electricity theft detection; generative adversarial network; semi-supervised learning; Wasserstein distance; discriminator

窃电指以蓄意绕开电费缴纳为目的,通过非法手段来最小化用电量的行为,不仅会极大损害电网公司的经济利益、严重扰乱电力市场的正常秩序,而且危及电力设备的安全,给电能稳定供给带来巨大挑战[1].据国家电力部门不完全统计,我国因窃电造成的年均亏损约为两百亿元人民币[2].近年来,随着虚拟货币狂潮席卷世界,因“挖矿”导致的窃电现象日益猖獗,2019年安徽和江苏警方先后破获两起“比特币挖矿”窃电案件[3-4],累计窃电价值分别为10万元和2 000万元.而在世界范围内,2014年因偷电造成的经济损失约为893亿美元[5],2017年增至960亿美元[6].图1所示为部分窃电较为严重国家的窃电数据统计饼状图,由图1可知,发展中国家因窃电导致的经济损失远超发达国家[7].因此对于窃电用户准确检测,对于维护电力市场的正常运转、保障供电企业的经济利益具有重要的现实意义和社会价值.

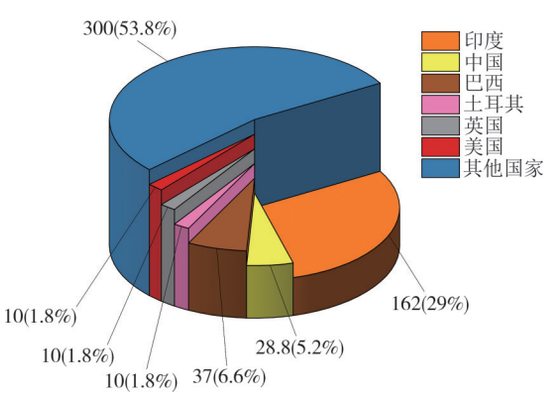


图1 窃电数据统计饼状图(单位:亿美元)

Fig.1 Statistical pie chart of electricity theft data (unit: 10^8 dollar)

传统窃电检测主要依赖于人工排查,该方式不仅效率低下,且难以精准定位到台区下的窃电用户[8].同时,物理手段仍然无法应对网络攻击等高级手段对电表实施的人为干预.因此,近年来众多研究者提出多种窃电检测方法自动识别疑似窃电用户.

检测窃电方法主要分为三类:面向网络方法、面向网络混合方法以及面向数据方法[9-10].面向网络方法主要由状态估计方法[11-12]和传感器网络方法组成[13],通过分析电网中物理规则或利用传感器数据进行窃电检测.然而,面向网络方法因强烈依赖于电网的拓扑信息[14],且通常需特定的传感器,部署成本较高,难以在电网公司全面展开.

相比之下,面向数据方法仅利用与用户用电相关数据(如电量、用电类型、电压电流等),该类方法具有开发成本小、可开发程度高的优点[15],随着电网公司大面积全面推广高级量测体系(Advanced Metering Infrastructure, AMI)[16],用电数据的利用率和分辨率得到极大提升[17],有力地推动了面向数据窃电检测技术的发展.

目前,基于AMI的面向数据窃电检测方法研究主要分为无监督学习方法和有监督学习方法.无监督学习方法无需标签信息,设窃电用户与正常用户的用电习惯上存在显著差异,通过密度、距离、近邻度、相似性等方法筛选出与多数用户相异的用户为疑似窃电用户[18-19].文献[20]提出基于模糊聚类的方法,利用高斯核函数改进局部离群因子对用户窃电行为进行检测,但检测模型存在解耦问题.文献[21]为解决模型解耦问题,提出将估计网络和高斯混合模型进行联合优化以避免模型解耦.文献[22]构建台区线损和与用电量之间的向量自回归模型,进一步提高检测的识别精度.然而,无监督学习检测模型对参数比较敏感,难以训练出最佳模型.

有监督学习方法需包含标签数据,通常比无监督学习方法具有更好的检测性能[23].文献[24-25]分

别采用支持向量机和决策树对窃电用户进行识别,文献[26-27]为解决单分类算法难以全面学习数据的不同特征,分别采用AdaBoost和Bagging的集成学习方法进一步提高检测精度.但上述有监督学习未进行特征工程提取用电数据的时间序列特征,为此,文献[28]利用卷积神经网络提取用电负荷特征,随机森林作为分类器;文献[29]利用卷积神经网络提取用电高维特征,LSTM用于提取用电负荷间的实际关联特征.文献[30]提出了宽度和深度CNN模型,宽度部分利用线性模型强记忆能力挖掘提取特征间的相关性,深度部分从二维的用电数据中提取周期性特征,再将两部分模型进行联合训练得出最终分类结果.文献[31]提出在CNN基础上引入注意力机制,优化CNN的特征提取过程,以提高检测精度并提升效率.

在实际情况中,电网公司所面临的主要挑战是获取有标签的数据成本高、难度大,而实际采集到的数据包含大量未标记的数据难以训练有效窃电检测模型.为此,本文提出了在少量有窃电标签数据场景下基于联合训练生成对抗网络(Co-Training Generative Adversarial Networks,CT-GAN)的半监督窃电检测方法.首先探究半监督生成对抗网络的原理与结构;其次提出采用最优传输距离(Wasserstein Distance)取代JS散度(Jenson's Shannon,JS)和KL散度(Kullback-Leibler Divergence,KL)距离,并构建多判别器联合训练模型;最后,采用爱尔兰电网数据集进行实验验证.

1　数据预处理与数据准备

1.1　数据预处理

在进行用电负荷数据采集时,由于各种因素如软硬件故障或特殊事件等,数据中可能存在错误值或缺失值,会影响用电记录的连续性,因此需对原始数据集进行预处理.本文选择3σ定律对错误值进行修复[30],计算式为:

$$f(x_i)=\begin{cases}\dfrac{x_{i-1}+x_{i+1}}{2}, & 若x_i>3\cdot\sigma(\boldsymbol{X}_i)且x_{i-1},x_{i+1}\neq \mathrm{NaN}\\ x_i, & 其他情况\end{cases} \tag{1}$$

式中:x_i表示客户一天的用电量值,$\sigma(\boldsymbol{X}_i)$表示向量X_i的标准差,NaN表示x_i不是一个数字符号.

此外,对于数据集中的缺失值,采用如下计算方式进行恢复:

$$f(x_i)=\begin{cases}\mathrm{mean}(\boldsymbol{X}_i), & 若x_i\in \mathrm{NaN}\\ x_i, & 其他情况\end{cases} \tag{2}$$

式中:mean($\boldsymbol{X}_i$)表示向量$\boldsymbol{X}_i$的平均值.

电力负荷包括基荷和变动负荷,对负荷数据进行标准化处理,有利于模型训练和收敛,本文采用min-max标准化方法可去除基荷并突出变动负荷走势,同时避免数量级相差较大的影响,也有助于降低异常天气、错误数据以及季节性波动的影响,其计算式为:

$$x_{i,j}^k=\frac{x_{i,j}^k-x_{i\min}^k}{x_{i\max}^k-x_{i\min}^k} \tag{3}$$

式中:$x_{i\min}^k$为第i个用户在第k天用电负荷的最小值,$x_{i\max}^k$为第i个用户在第k天用电负荷的最大值.

1.2　数据准备

本文通过为爱尔兰ISET公开数据集[32]验证提出算法的有效性和准确性,该数据集均为正常用户数据集(即不包含窃电用户).因此对窃电检测模型进行训练,需人为通过公式针对正常数据集进行篡改以模拟窃电行为,来构造出包含窃电用户与正常用户的数据集.表1为通过6种篡改公式针对ISET正常数据集篡改以模拟窃电行为[33],其中涉及随机数生成的部分,使用均匀分布的随机数.

表1　6种篡改模式

Tab.1　Six types of malicious samples

攻击类型	攻击方式
类型1	$\tilde{x}_t=\alpha x_t,\ 0.2<\alpha<0.8$
类型2	$\tilde{x}_t=\alpha_t x_t,\ 0.2<\alpha_t<0.8$
类型3	$\tilde{x}_t=\beta x_t,\ \beta=\begin{cases}1, & 若t_1<t<t_2\\ 0, & 其他情况\end{cases}$
类型4	$\tilde{x}_t=\alpha_t\bar{X},\ 0.2<\alpha_t<0.8$
类型5	$\tilde{x}_t=\bar{X}$
类型6	$\tilde{x}_t=x_{48-t}$

注:$\tilde{x}_t$为窃电后计量电量;x_t为正常用电量;$\bar{X}$为用电量均值.

ISET数据集中用户的用电量记录频率为每半小时记录一次,因此可以使用向量$X=[x_1,x_2,\cdots,x_{48}]$来表示某用户一天内的用电量时间序列.其中,类型1表示一天内所有时刻采集的用电量乘以时不变随机因子(0.2~0.8);类型2表示用电量乘以时变随机因子α_t;类型3表示将t_1-t_2以外的用电量置0,t_1-t_2是随机选取的一天内某个大于6 h的时间段;类型5表示

将一天内所用电量篡改为平均值的形式;类型4是将类型5中的用电量平均值乘以时变随机因子α_t;类型6表示将一天内的用电量颠倒,以达到将用电量高峰转移到低电价时段的目的.如图2所示为正常日常用电量模式与六种篡改公式生成的用电曲线图.

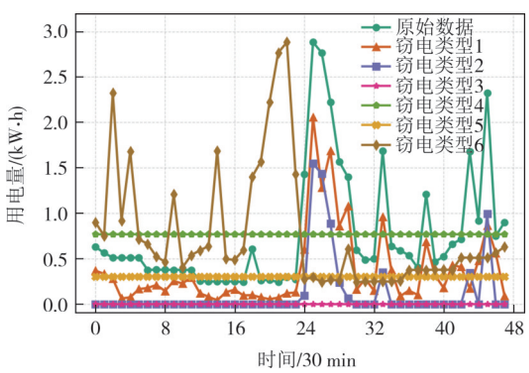


图2 6种篡改公式产生的用电曲线图

Fig.2 The electricity curve generated by 6 tampering formulas

2 基于CT-GAN的窃电检测方法

2.1 半监督生成对抗网络

Wang等学者提出了一种基于半监督学习的生成对抗网络(Semi-supervised Learning with Generative Adversarial Networks, SGAN)[34],该方法以GAN为基础结构搭建半监督学习模型,并利用少量的有标签样本和大量的无标签样本来训练判别器D,网络的输出结果为K+1维含类别信息的分类标签,如[Class_1,Class_2,Class_3,…,Class_K,Fake],SGAN网络结构如图3所示.

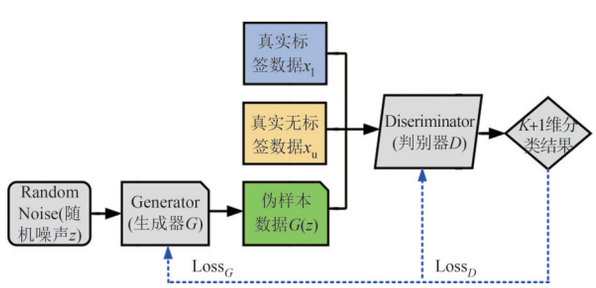


图3 SGAN网络结构

Fig.3 SGAN network structure

SGAN中判别器D的训练数据包括生成器G生成的伪样本$G(z)$、真实有标签样本x_l和真实无标签样本x_u组成的K维样本集合,判别器的最后一层采用softmax非线性函数作为分类器.该分类器的输出是一个K+1维的向量$[l_1,l_2,\cdots,l_{K+1}]$,该向量的前K维表示对应类别的置信概率,而第K+1维表示判定样本为"伪"的置信概率.

尽管SGAN采用基于GAN的半监督学习模型,并在某些数据集下取得较好的分类准确性,但仍存在较为严重的训练不稳定问题.主要原因是SGAN在模型训练中,如果单个判别器的分布误差明显,这意味着判别器的预测与实际样本类别之间存在明显的差异,即判别器无法准确地预测样本所属的类别概率分布,会出现梯度消失现象,进而导致判别器网络不收敛.为解决半监督生成对抗网络训练不稳定的弊端,同时进一步提高窃电行为检测分类的准确性,本文提出一种基于联合训练生成对抗网络(CT-GAN)的半监督窃电检测方法,该方法采用如图4所示的网络结构.

由图4可知,判别器D_1和判别器D_2均加入了网络的训练之中,即进行联合训练,该网络可在有效增强网络训练的稳定性同时提高窃电检测的准确性.D_1、D_2判别器共用同一生成器G,且它们的网络结构和初始参数设置完全相同.值得注意的是,为保证CT-GAN网络在训练过程中D_1、D_2判别器处于动态平衡变化,需将真实标签数据x_l和真实无标签数据x_u顺序随机调整再分别输入D_1、D_2判别器中.为消除前述半监督学习的生成对抗网络中单个判别器分布误差较大带来的影响,CT-GAN在训练过程中取D_1、D_2判别器损失的平均值计算其网络的损失函数,而网络的最终输出为K+1维分类结果,即$[l_1,l_2,\cdots,l_{K+1}]$.为进一步优化网络结构,提高网络收敛速度,在图4所示的CT-GAN网络结构中,针对D_1、D_2判别器,采用相同的置信度阈值策略:若生成样本数据$G(z)$的置信度高于该阈值,将其视为高质量生成样本,并将其标记为"伪标签".随后,将这些伪标签的生成样本数据重新融合到初始的真实标签数据集中,以丰富数据集的多样性和扩大其规模.这个策略旨在进一步提高生成模型的性能,确保生成的样本质量较高,同时扩展可用于训练的数据集规模.

在CT-GAN网络中,生成器G的主要作用是生成具有高度逼真性质的合成数据,以模糊判别器D对待真实数据和生成数据的能力,从而实现数据分布的伪装,增加判别器D的难度,使其难以区分输入数据的真实性,即$D[G(z)]$无限接近于1,判别器D每次输出的概率为1/2,而此时$V(D,G)$取得极小值,生成器损失函数定义为:

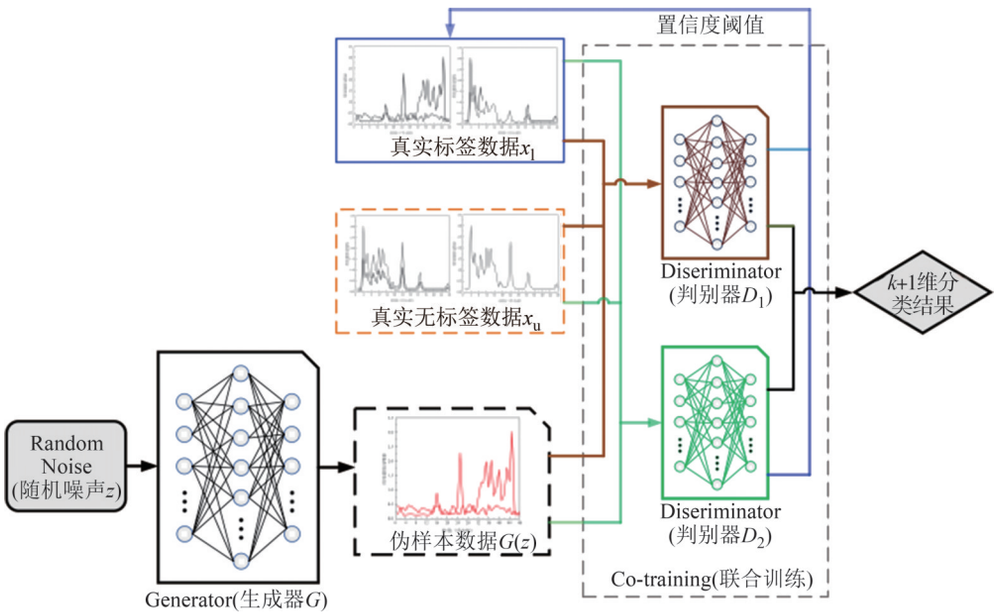


图 4　CT-GAN 网络结构

Fig.4　CT-GAN network structure

$$\min E_{z\sim P_z}\left(\sum_{j=1}^{2}\log\left[1-D_j(G(z))\right]\right) \tag{4}$$

式中:$G(z)$是生成器生成的数据,其目的是使得这些数据在统计上尽量与真实数据分布相符合;$D(\cdot)$是判别器判断给定输入样本数据来自真实分布的概率[$0\leqslant D(\cdot)\leqslant 1$],当$D(x)$越接近1时,$D[G(z)]$越接近0,此时$V(D,G)$取得极大值.

对于模型的生成器损失函数,本节采用特征匹配的方法对其损失进行约束,以使生成器生成的数据分布尽可能逼近真实数据分布,其特征匹配的损失定义为:

$$\left\| E_{x\sim P_{\text{data}}(x)}\sum_{j=1}^{2}f_j(x)-E_{z\sim P_z}\sum_{j=1}^{2}f_j\left[G(z)\right]\right\|_2^2 \tag{5}$$

式中:$f_j(\cdot)$代表判别器D_j在其神经网络中,接近全连接层的一层输出的特征值.结合公式(4)和公式(5)可得出CT-GAN生成器G的总损失为:

$$\min \text{Loss}_G=\min E_{z\sim P_z}\left(\sum_{j=1}^{2}\log\left\{1-D_j\left[G(z)\right]\right\}\right)+$$

$$\left\| E_{x\sim P_{\text{data}}(x)}\sum_{j=1}^{2}f_j(x)-E_{z\sim P_z}\sum_{j=1}^{2}f_j\left[G(z)\right]\right\|_2^2 \tag{6}$$

本文判别器的损失值由有监督损失和无监督损失两部分组成,旨在综合考虑有监督学习和无监督学习的优势,以更好地优化判别器的性能.首先介绍有监督损失函数,该方式需将包含标签信息的数据作为输入,采用交叉熵(Cross Entropy, CE)作为损失函数的定义:

$$\max \text{Loss}_{\text{sup}\,D}=$$

$$E_{x,y\sim P_{\text{data}}(x,y)}\sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i\log D_j(x_i)+(1-y_i)\log\left[1-D_j(x_i)\right]\right\} \tag{7}$$

式中:y_i为输入样本的第i维标签;$D_j(x_i)$为判别器D_j将输入样本标签判别为第i维的概率.

CT-GAN模型需将D_1、D_2判别器联合训练的情况考虑其中以计算无监督损失,故其函数表达式如式(8)所示:

$$\max \text{Loss}_{\text{sup}\,D}=E_{x\sim P_{\text{data}}(x)}\left\{\log\left(1-P_{\text{model}}\left[y\middle|(x,y)\right]\right)\right\}+$$

$$E_{x\sim P_g}\left\{\log\left(1-P_{\text{model}}\left[y=y<K+1\middle|x\right]\right)\right\}=$$

$$E_{x\sim P_{\text{data}}(x)}\sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i'\log D_j(x_i)+(1-y_i')\log\left[1-D_j(x_i)\right]\right\}+$$

$$E_{z\sim P_z}\left(\sum_{j=1}^{2}\log\left\{1-D_j\left[G(z)\right]\right\}\right) \tag{8}$$

式中:y_i'为判别器前一次迭代中无标签样本第i类别预测置信度;$D_j(x_i)$为判别器D_j预测有标签样本的标签为第i类别的置信度.

联立公式(7)和公式(8)可得,CT-GAN判别器D的总损失函数表达式如式(9)所示:

$$max\text{Loss}_D = max\text{Loss}_{\sup D} + max\text{Loss}_{\text{un}\sup D} =$$
$$E_{x,y\sim P_{\text{data}}(x,y)} + \sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i \log D_j(x_i) + (1-y_i)\log\left[1-D_j(x_i)\right]\right\} +$$
$$E_{x\sim P_{\text{data}}(x)}\sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i' \log D_j(x_i) + (1-y_i')\log\left[1-D_j(x_i)\right]\right\} +$$
$$E_{z\sim P_z}\left(\sum_{j=1}^{2}\log\left\{1-D_j\left[G(z)\right]\right\}\right) \tag{9}$$

为计算CT-GAN整体总损失函数,可将生成其总损失即式(6)和判别器总损失即式(9)联立,其表达式为:

$$\min_G \max_D V(D,G) = E_{x,y\sim P_{\text{data}}(x,y)} +$$
$$\sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i \log D_j(x_i) + (1-y_i)\log\left[1-D_j(x_i)\right]\right\} +$$
$$E_{x\sim P_{\text{data}}(x)}\sum_{i=1}^{K}\sum_{j=1}^{2}\left\{y_i' \log D_j(x_i) + (1-y_i')\log\left[1-D_j(x_i)\right]\right\} +$$
$$E_{z\sim P_z}\left(\sum_{j=1}^{2}\log\left\{1-D_j\left[G(z)\right]\right\}\right)$$
$$\left\| E_{x\sim P_{\text{data}}(x)}\sum_{j=1}^{2} f_j(x) - E_{z\sim P_z}\sum_{j=1}^{2} f_j\left[G(z)\right]\right\|_2^2 \tag{10}$$

2.2 半监督生成对抗网络窃电检测模型

CT-GAN模型不仅可解决半监督生成对抗网络训练不稳定的弊端,同时能利用GAN生成标签样本数据的能力,对于电网公司实际存在的获取标签数据难以及获取成本极高的难题提出一种有效的解决方式.而CT-GAN模型能否有效实现的关键是其网络联合训练(Co-training)的合理性.对于判别器D,其主要目的是判断输入数据的真伪性而无需判别其类别属性;对于生成器G生成的伪数据,对其赋予"伪标签"的伪数据已重新作为扩充数据添加到初始真实标签数据集中,因此,在CT-GAN模型中判别器D_1、D_2无需考虑伪数据的输入,只需考虑标签数据及无标签数据的输入.

CT-GAN模型的联合训练示意图如图5所示.为保证CT-GAN网络在训练过程中D_1、D_2判别器处于动态平衡,在进行联合训练前,首先对真实标签数据x_l和真实无标签数据x_u进行随机顺序调整,然后将它们分别输入判别器D_1和D_2中.

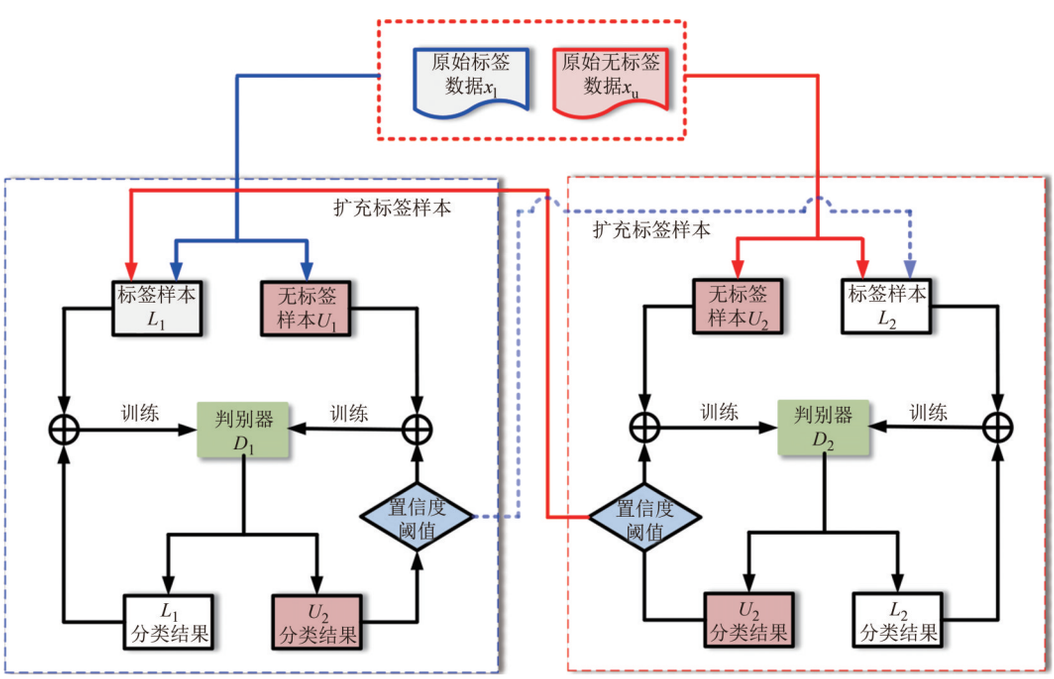


图5 联合训练示意图

Fig.5 Schematic diagram of co-training

在联合训练网络中,判别器D_1的训练过程包括三个关键操作,如下所示:

1)使用标签样本L_1来训练判别器D_1,以获取L_1分类属性,再根据公式(7)计算判别器的监督损失来训练判别器D_1.

2)通过判别器D_1对无标签样本数据U_1进行分类预测.再使用最近一次迭代获得的U_1分类标签与当前U_1的预测标签一同用于计算公式(8)所示的无监督损失,以便进一步优化U_1的标签属性.

3)在判别器D_1的设置中,引入标签样本数据L_2

的扩展.这一扩展包括在D_1侧引入一个置信度阈值,用于评估无标签样本数据U_1的分类预测结果,若其大于设置的置信度阈值,即将该数据贴上“伪标签”并将其放置于对应的标签样本数据L_2中继续用于训练,进而可有效地扩展训练数据集,加快网络收敛速度.

判别器D_2的训练过程与D_1基本相似,如果判别器预测某个样本属于某个类别的概率高于其他所有类别的概率之和,则将此样本分配给该类别标签.若$P_{\max}$为无标签样本的预测类别的概率,则

$$P_{\max} \geqslant \sum_{i=1}^{K} P_i - P_{\max} \tag{11}$$

式中:P_i为某分类类别的概率,K为类别数.由此可得,置信度阈值设置为0.5最为合理.

在CT-GAN模型中,通过D_1、D_2判别器的网络联合训练能消除生成对抗网络中单个判别器分布误差大的缺陷,进而解决半监督生成对抗网络训练不稳定的弊端.同时利用GAN生成标签样本数据的优势,同样可利用无标签数据在训练过程中对于有标签数据集的扩充,以此在丰富标签数据集后,对网络收敛起到有效的加速作用.

对于CT-GAN网络结构而言,设计合理、高效的生成器网络和判别器网络结构对数据生成质量、数据分类预测结果起到至关重要的作用.本文选用生成数据质量好、网络收敛速度快的U-net全卷积神经网络结构作为生成器网络,如图6所示为U-net结构示意图.

该结构由8个残差块组成,可分为编码(Encoder)和解码(Decoder)两部分.编码部分主要由4次下采样和卷积操作组成;解码部分主要由4次上采样和卷积操作组成.上采样和下采样过程的具体模型参数设置如表2所示.

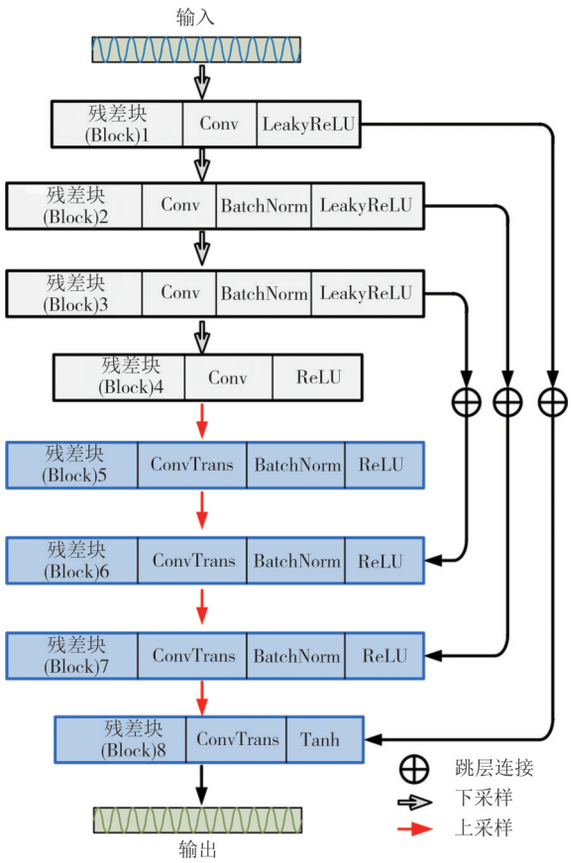


图6　U-net结构示意图

Fig.6　Schematic diagram of U-net structure

首先,U-net网络结构的输入尺寸为1×48×48的用电数据矩阵,该数据经过4个残差块的操作后网络输出为一系列尺寸为3×3的特征矩阵,其中残差块1的下采样操作包括1次卷积(Convolution,简称Conv)操作和1次LeakyReLU操作;残差块2和残差块3的下采样操作分别包括1次卷积操作、1次批归一化操作和1次LeakyReLU操作;残差块4的下采样

表2　U-net结构参数设置

Tab.2　U-net structure parameter setting

残差块序号	输入尺寸	卷积核	卷积步长stride	LeakyReLU参数	动量	输出尺寸
Block1	1×48×48	3×3	2	0.2	—	16×24×24
Block2	16×24×24	3×3	2	0.2	0.9	32×12×12
Block3	32×12×12	3×3	2	0.2	0.9	64×6×6
Block4	64×6×6	3×3	2	—	—	64×3×3
Block5	64×3×3	3×3	2	—	0.9	64×6×6
Block6	64×6×6	3×3	2	—	0.9	32×12×12
Block7	32×12×12	3×3	2	—	0.9	16×24×24
Block8	16×24×24	3×3	2	—	—	1×48×48

操作包括1次卷积操作和1次ReLU操作.

其次,经过上述4个残差块的下采样操作后,对获得的特征矩阵采用上采样操作以实现特征融合,其中在上采样过程中,U-net网络结构在残差块6、7和8的上采样操作前选用3次跳层连接操作,而跳层连接操作指的是:将下采样过程中产生的特征数据与此次上采样输入相同尺寸的特征数据根据通道拼接组成新的特征数据,再对其执行反卷积操作,跳层连接操作可有效地将高维输入数据的特征更快地融合至输出数据里.其中残差块5、6及7的上采样操作包括1次反卷积(简称ConvTrans)操作、1次批归一化操作和1次ReLU操作.残差块8的上采样操作包括1次反卷积操作及Tanh激活函数操作.经过上述上采样操作,U-net网络输出与输入相同尺寸的生成矩阵,即1×48×48的生成用电数据矩阵.

CT-GAN结构的判别器网络选用如图7所示的神经网络结构.将输入为1×48×48的数据进行残差块1的操作,包括1次3×3卷积核的卷积操作(Conv1)、1次LeakyReLU激活函数操作及1次Dropout操作(在神经网络中添加Dropout操作,其目的是在每次训练前随机丢掉一些网络连接,以防止模型过拟合,增加模型的泛化能力);分别进行残差块2、3、4的操作,分别包括1次3×3卷积核的卷积操作(Conv1)、1次LeakyReLU激活函数操作、1次Dropout操作及1次批归一化(BatchNorm)操作.

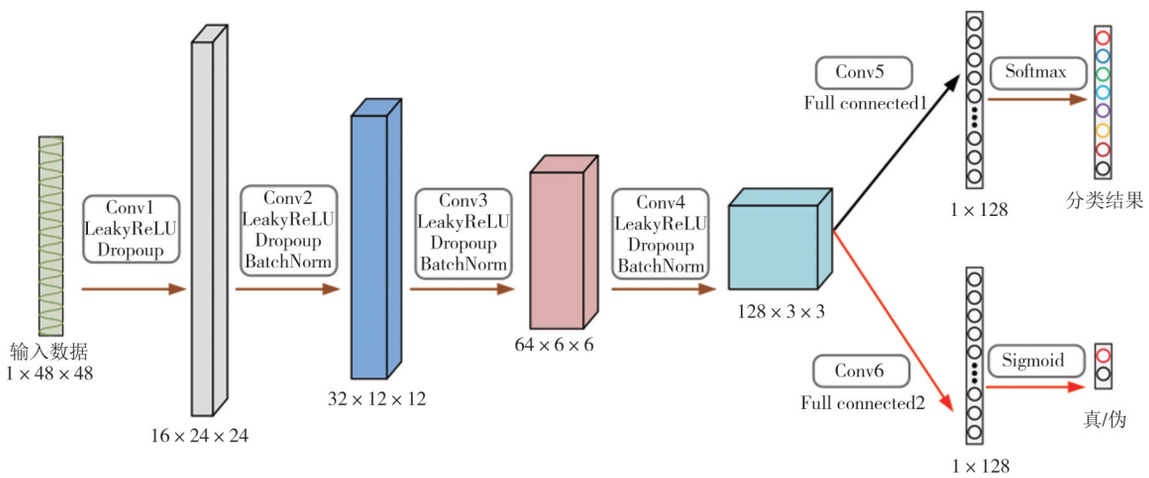


图7 判别器网络结构示意图

Fig.7 Schematic diagram of discriminator network structure

最后,将残差块4得到的128×3×3的数据进行ReShape操作,其中通过第1个全连接层(Full connected1)128×3×3卷积核的卷积操作得到1×128的数据;通过第2个全连接层(Full connected2)128×3×3卷积核的卷积操作得到1×128的数据;进一步,将利用两个全连接层提取的特征数据分别通过Softmax操作和Sigmoid操作进行处理,从而得到分类结果和真伪判别结果.具体而言,判别器网络结构参数已在表3中列出.

表3 判别器网络结构参数设置

Tab.3 Discriminator network structure parameter setting

参数名称	参数值
卷积核大小	3×3
卷积步长stride	2
Padding(填充)	1
全连接层神经元个数	128
学习率	0.1

3 实验验证与分析

为验证本文所提出的算法的有效性和准确性,选用一台配置为64位、6核心、十二线程的Intel Core i7-8750H CPU@2.20 GHz的计算平台作为实验平台,并采用了深度学习框架PyTorch作为实验的基础工具,实验数据为前述的爱尔兰ISET公开数据集.窃电行为检测可视为二元分类任务,算法对用户进行分类识别后,还需要进行准确性评估以验证检测方法的有效性.通常采用如表4所示的混淆矩阵作为依据.

根据表4的混淆矩阵可得准确率(Accuracy,ACC)=(TP+TN)/(TP+TN+FP+FN)和检出率TPR=TP/(TP+FN)及FPR=FP/(TN+FP).误检率FPR描述正常用户被误判为窃电用户占所有正常用户的比

例;TPR 描述检测出真实窃电用户占总窃电用户的比例.F1 值为 TPR 和 FPR 的加权平均,表示为 F1=2TP/(2TP+FN+FP).受试者操作特性(Receiver Operating Characteristic, ROC)曲线分别以 FPR 和 TPR 为横轴和纵轴,显示检测器 TPR 和 FPR 增长速率的相对关系.AUC(Area Under ROC Curve)为 ROC 曲线下的面积值,其以数值定量评估检测器性能,AUC 数值越高,检测器有越大的概率将窃电用户准确识别同时有越小的概率将正常用户误判为窃电.

表 4　混淆矩阵

Tab.4　Confusion matrix

用户	检测为窃电用户	检测为正常用户
实际窃电用户	TP (True Positive)	FN (False Negative)
实际正常用户	FP (False Positive)	TN (True Negative)

本次实验中选择该数据集中从 2009 年至 2010 年共 528 天以半小时为采样间隔 5 000 个电力用户的用电量数据组成所需的数据集,将每个数据分为 11 个 48 天的样本(每一天有 48 个半小时采样点),即得到 11×5 000=55 000 个正常样本.本实验将数据集以 6:2:2 的比例拆分为训练集、验证集和测试集,考虑到电网实际情况下窃电用户的比例占总用户比例较小,从数据集中随机选择 5% 的比例,通过前述篡改公式针对 ISET 正常数据集篡改以模拟窃电行为.

半监督学习方法为利用少量标签数据训练初始学习器,再对未知类别数据进行检验分类,因此需对数据集进行相应预处理,本次实验中,按照一定的比例删除部分标签数据的标签信息,以此构成一定的无标签数据.针对预处理好的数据集,分别选用含标签比例为 0.25%、2.5%、6.25%、12.5%、25% 和 50% 分别进行实验,以研究含不同标签数据比例的数据集在本文所提出的 CT-GAN 窃电检测方法检测性能.如图 8 所示为含标签比例为 0.25%、2.5%、6.25%、12.5%、25% 和 50% 时 CT-GAN 窃电检测模型生成器损失变化曲线图.

从图 8 可知,随着标签数据比例的不断增加,生成器模型的损失趋于稳定所需的迭代次数也不断增加,当含标签比例为 0.25% 时,损失趋于稳定所需的迭代次数最少,而当含标签比例为 50% 时,损失趋于稳定所需的迭代次数最多.对于含标签比例的数据,CT-GAN 窃电检测模型的损失值随着迭代次数的增加逐渐减少,并收敛至相对较低的水平.

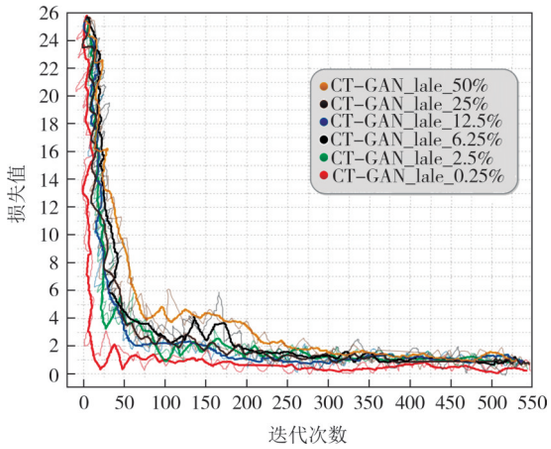


图 8　CT-GAN 窃电检测模型生成器损失变化曲线

Fig.8　CT-GAN electricity theft detection model generator loss change curve

如图 9 所示为含标签比例为 0.25%、2.5%、6.25%、12.5%、25% 和 50% 时 CT-GAN 窃电检测模型判别器损失变化曲线图,随着标签数据比例从 0.25% 增加至 50% 时,判别器模型的损失趋于稳定所需的迭代次数不断增加,当含标签比例为 0.25% 时,损失趋于稳定所需的迭代次数最少,这是由于标签数据的数量越少,整个数据集所包含的有效信息量越少,进而判别器可有效学习的信息相应减少,从而使得模型的收敛速度加快.

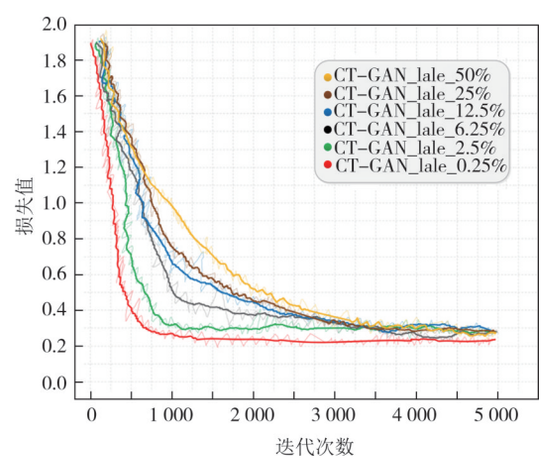


图 9　CT-GAN 窃电检测模型判别器损失变化曲线

Fig.9　CT-GAN electricity theft detection model discriminator loss change curve

为进一步验证本文所提出方法的准确性和有效性,对通过上述预处理后的数据集进行验证,其中对比了现存 6 种检测效果较好的半监督学习方法,包括自训练(Self-Training, ST)[35]、协同训练(Co-Training, CT)[36]、基于图的方法(Graph-Based Methods, GBM)[37]、生

成式模型(Generative Models, GM)[38]、半监督支持向量机(S3VM)[39]和半监督生成对抗网络(SGAN)[34].如表5所示在设置60%数据为训练数据集,余下40%数据为验证数据集,且进行20次重复试验下,在不同的标签比例数据样本的各半监督学习方法分类的平均精度,即ACC值.

从表5的分类平均精度结果可知,本文提出的CT-GAN窃电检测模型的分类准确性在不同的标签比例数据下,相比于其他方法更高;随着样本数据数量不断增加,该方法的分类精度均有不同程度的提升,如在标签数据比例仅为0.25%时,CT-GAN窃电检测模型的分类精度值可达56.87%,相比于半监督生成对抗网络提升了17.91%,提升效果明显.从以上的分类结果可以分析出,CT-GAN窃电检测模型可有效提升当标签数据偏少的情况下模型的检测精度和分类准确性,相当程度上解决了传统生成对抗网络在小样本情况下的模型过拟合问题的弊端.

表5 不同标签比例数据样本的各半监督学习方法分类精度

Tab.5 ACC of each semi-supervised learning method for data samples with different label ratios %

比例/%	ST	CT	GBM	GM	S3VM	SGAN	CT-GAN
0.25	25.37	24.46	24.31	25.55	24.89	38.96	56.87
2.5	39.86	37.17	36.15	41.47	38.12	61.02	70.25
6.25	49.35	47.27	45.11	52.66	48.33	68.18	81.42
12.5	58.96	56.37	54.42	60.17	57.58	72.32	87.15
25	65.43	62.08	64.76	71.62	63.43	78.47	89.13
50	78.12	76.11	77.75	82.86	77.07	84.47	91.29

为更有效地评估模型的训练和分类效果,采用AUC值作为重要的参考依据.本文利用标签数据比例为50%的数据进行比较试验,设置60%数据为训练数据集,余下40%数据为验证数据集,如图10所示为各个方法的ROC曲线.

如图10所示,本文提出的基于CT-GAN窃电行为检测方法AUC值高达93.3%,均优于其他方法.本文利用Wasserstein距离解决传统生成对抗网络梯度消失和模式崩溃的问题,提升了模型的训练稳定性,提高了生成数据的质量;而采用的基于半监督学习理论的联合训练网络,通过两个判别器联合训练,不仅解决了半监督生成对抗网络训练不稳定的弊端,同时通过增强GAN生成标签样本方面的能力达到高效扩充标签样本数据集的目的,从而成功提升了半监督分类任务的精确度,并有效增强了模型的泛化性能.因此,本文提出方法相比于现有的几种半监督机器学习方法,在模型训练效率和窃电行为识别准确率方面均表现出显著的优势.

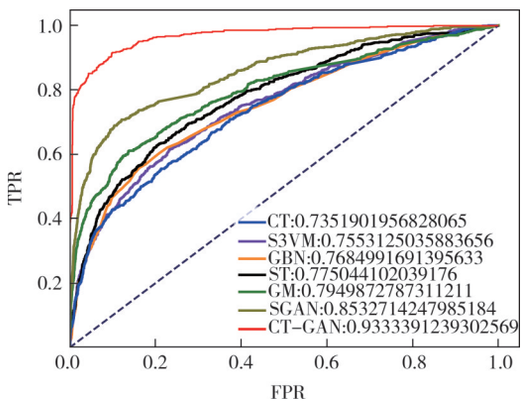


图10 不同方法的ROC曲线

Fig.10 ROC curves of different methods

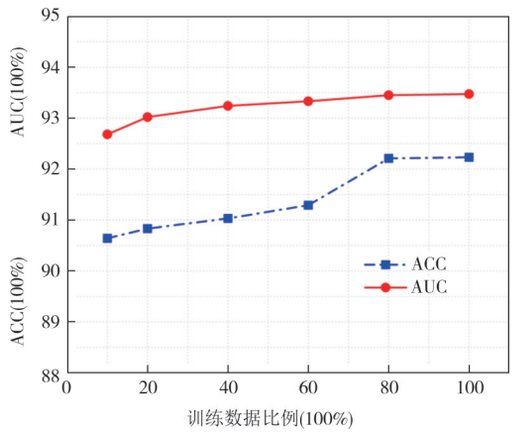
本文通过实验比较不同训练数据比例下模型的性能表现,如表6和图11(a)所示,当使用的训练样本比例降低时,本文所提出模型的性能略有下降,而当训练样本的比例高于40%时,模型的性能几乎保持不变,且在比例为80%时性能较好,该模型的其他评估指标显示出类似趋势.结果表明,当达到一定数量的训练样本,即原始数据集的40%时,模型性能已能达到最佳值.

表6 不同比例训练样本的模型性能

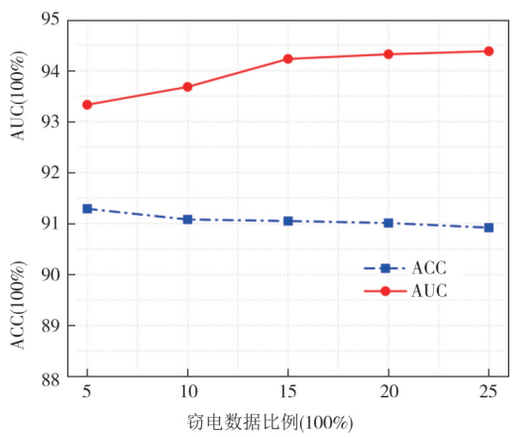
Tab.6 Model performance at different proportions of training samples %

比例	评价指标				
	AUC	ACC	FPR	TPR	F_1
10	92.68	90.64	3.79	89.04	86.69
20	93.02	90.83	3.86	89.96	88.78
40	93.24	91.03	2.65	90.42	89.55
60	93.33	91.29	2.97	90.96	89.64
80	93.45	92.21	3.15	91.22	90.31
100	93.47	92.23	2.36	90.36	90.66

为分析不平衡程度对模型性能的影响,本文设置不平衡度(窃电数据比例)分别为5%、10%、15%、20%和25%.实验结果如表7所示,ACC和AUC的可视化结果如图11(b)所示,可看出,当窃电样本比例在5%以上时,尽管样本不平衡程度增加,但AUC值基本不变;当比例从25%降至5%时,FPR、TPR及F1均下降约2%,而ACC略有上升,但各指标变化较小.综上所述,随着窃电比例降低,模型的整体性能变化较小,因此,模型具有较好的鲁棒性.



(a)训练数据



(b)窃电数据

图 11　不同比例数据下模型的性能图

Fig.11　Performance graph of the model under different scale data

表 7　不同比例样本不平衡数据的模型性能

Tab.7　Model performance on imbalanced data with different ratio samples　　%

比例	评价指标				
	AUC	ACC	FPR	TPR	F_1
5	93.33	91.29	2.97	90.96	89.64
10	93.68	91.08	3.29	91.63	90.39
15	94.23	91.05	4.36	91.98	91.03
20	94.32	91.01	4.78	92.54	91.69
25	94.38	90.92	4.69	92.89	92.42

4　结 论

本文提出一种基于联合训练生成对抗网络CT-GAN的半监督窃电检测方法,主要适用于解决电网公司难以获取疑似窃电用户标签数据的现实困难.本文首先介绍了生成对抗网络及半监督生成对抗网络的原理和结构,并阐述它们存在的缺陷;其次针对上述问题,提出采用Wasserstein距离取代JS散度和KL散度以解决生成对抗网络因梯度消失和模式崩溃导致的模型训练不稳定和生成数据质量低的问题,随后提出利用两个判别器联合训练,消除单个判别器分布误差高的问题,提高了半监督生成对抗网络训练的稳定性,达到高效扩充训练数据集的目的,从而提高了模型的窃电检测精度和泛化能力.最后通过不同标签比例数据样本对各半监督学习方法进行比较分析,进一步验证了本文所提CT-GAN半监督窃电检测方法的有效性和准确性,为电网公司提供准确有效的检测手段,以发现当前智能电网中的异常用电行为并稽查潜在的窃电用户,旨在降低企业运营成本,同时保证电网的可靠安全运行,为此提供了重要的技术支持.

参考文献

[1] 卿柏元,陈珏羽,李金瑾,等. 基于CNN-LG模型的窃电行为检测方法研究[J]. 湖南大学学报(自然科学版),2022,49(8):138-148.
QING B Y, CHEN J Y, LI J J, et al. Research on detection method of electricity theft behavior based on CNN-LG model[J]. Journal of Hunan University (Natural Sciences), 2022, 49(8): 138-148.(in Chinese)

[2] 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 S, SU S, XUE Y, 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.(in Chinese)

[3] 新华社. 江苏破获特大盗电"挖"比特币案件案值近2000万元. https://www. 163. com/tech/article/EJSLMB6B00097U7R. html. 2019-07-12.
Xinhua News Agency. Jiangsu cracked a big stolen "digging" Bitcoin case in the case value of nearly 20 million yuan[EB/OL]. Https://www. 163. com/tech/article/ejslMB6B00097u7R. html. 2019-07-12.(in Chinese).

[4] 李伟,谭冯忍. 一起"比特币挖矿机"窃电案的启示[J]. 农村电工,2020,28(6):18.
LI W, TAN F R. Enlightenment from a case of "Bitcoin digger" stealing electricity[J]. Rural Electician, 2020, 28(6): 18.(in Chinese)

[5] JINDAL A, DUA A, KAUR K, et al. Decision tree and SVM-based data analytics for theft detection in smart grid[J]. IEEE Transactions on Industrial Informatics, 2016, 12(3): 1005-1016.

[6] CHAUHAN A A. Non-technical losses in power system and monitoring of electricity theft over low-tension poles[C]//2015 Second International Conference on Advances in Computing and Communication Engineering. Dehradun, India. IEEE, 2015: 280-284.

[7] GAO Y P, FOGGO B, YU N .A physically inspired data-driven model for electricity theft detection with smart meter data[J].IEEE Transactions on Industrial Informatics, 2019, 15(9): 5076-5088.

[8] HUANG S C, LO Y L, LU C N. Non-technical loss detection using state estimation and analysis of variance[J]. IEEE Transactions on Power Systems, 2013, 28(3): 2959-2966.

[9] MESSINIS G M, HATZIARGYRIOU N D. Review of non-technical loss detection methods[J]. Electric Power Systems Research, 2018, 158: 250-266.

[10] 王方雨,刘文颖,陈鑫鑫,等. 基于"秩和"近似相等特性的同期线损异常数据辨识方法[J]. 电工技术学报,2020,35(22):4771-4783.
WANG F Y, LIU W Y, CHEN X X, et al. Abnormal data identification of synchronous line loss based on the approximate equality of rank sum[J]. Transactions of China Electrotechnical Society,2020,35(22):4771-4783.(in Chinese)
[11] SALINAS S A, LI P. Privacy-preserving energy theft detection in microgrids: a state estimation approach[J]. IEEE Transactions on Power Systems,2016,31(2):883-894.
[12] 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.
[13] GUO Y H, TEN C W, JIRUTITIJAROEN P. Online data validation for distribution operations against cybertampering[J]. IEEE Transactions on Power Systems,2014,29(2):550-560.
[14] 陈厚合,丛前,姜涛,等. 多能协同的配电网供电恢复策略[J]. 电工技术学报,2022,37(3):610-622.
CHEN H H, CONG Q, JIANG T, et al. Distribution systems restoration with multi-energy synergy[J]. Transactions of China Electrotechnical Society,2022,37(3):610-622.(in Chinese)
[15] 王毅,张宁,康重庆,等. 电力用户行为模型:基本概念与研究框架[J]. 电工技术学报,2019,34(10):2056-2068.
WANG Y, ZHANG N, KANG C Q, et al. Electrical consumer behavior model: basic concept and research framework [J]. Transactions of China Electrotechnical Society, 2019, 34(10): 2056-2068.(in Chinese)
[16] RAZAVI R, GHARIPOUR A, FLEURY M, et al. A practical feature-engineering framework for electricity theft detection in smart grids[J]. Applied Energy,2019,238:481-494.
[17] 郭庆来,王博弘,田年丰,等. 能源互联网数据交易:架构与关键技术[J]. 电工技术学报,2020,35(11):2285-2295.
GUO Q L, WANG B H, TIAN N F, et al. Data transactions in energy internet: architecture and key technologies [J]. Transactions of China Electrotechnical Society, 2020, 35(11): 2285-2295.(in Chinese)
[18] BISWAS P P, CAI H Y, ZHOU B, et al. Electricity theft pinpointing through correlation analysis of master and individual meter readings[J]. IEEE Transactions on Smart Grid, 2020, 11(4):3031-3042.
[19] ZHENG K D, CHEN Q X, WANG Y, et al. A novel combined data-driven approach for electricity theft detection[J]. IEEE Transactions on Industrial Informatics,2019,15(3):1809-1819.
[20] 孙毅,李世豪,崔灿,等. 基于高斯核函数改进的电力用户用电数据离群点检测方法[J]. 电网技术,2018,42(5):1595-1606.
SUN Y, LI S H, CUI C, et al. Improved outlier detection method of power consumer data based on Gaussian kernel function[J]. Power System Technology,2018,42(5):1595-1606.(in Chinese)
[21] 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 Z R, GAO Y P, GUO J B, 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.(in Chinese)
[22] 殷涛,薛阳,杨艺宁,等. 基于向量自回归模型的高损线路窃电检测[J]. 中国电机工程学报,2022,42(3):1015-1023.
YIN T, XUE Y, YANG Y N, et al. Electricity theft detection of high-loss line with vector autoregression[J]. Proceedings of the CSEE,2022,42(3):1015-1023.(in Chinese)
[23] XIA X F, XIAO Y, LIANG W, et al. Detection methods in smart meters for electricity thefts: a survey[J]. Proceedings of the IEEE, 2022, 110(2): 273-319.
[24] PEREIRA D R, PAZOTI M A, PEREIRA L A M, et al. Social-spider optimization-based support vector machines applied for energy theft detection[J]. Computers & Electrical Engineering, 2016,49:25-38.
[25] JINDAL A, DUA A, KAUR K, et al. Decision tree and SVM-based data analytics for theft detection in smart grid[J]. IEEE Transactions on Industrial Informatics,2016,12(3):1005-1016.
[26] 游文霞,申坤,杨楠,等. 基于AdaBoost集成学习的窃电检测研究[J]. 电力系统保护与控制,2020,48(19):151-159.
YOU W X, SHEN K, YANG N, et al. Research on electricity theft detection based on AdaBoost ensemble learning [J]. Power System Protection and Control, 2020, 48(19): 151-159.(in Chinese)
[27] 游文霞,申坤,杨楠,等. 基于Bagging异质集成学习的窃电检测[J]. 电力系统自动化,2021,45(2):105-113.
YOU W X, SHEN K, YANG N, et al. Electricity theft detection based on Bagging heterogeneous ensemble learning [J]. Automation of Electric Power Systems,2021,45(2):105-113.(in Chinese)
[28] LI S, HAN Y H, YAO X, et al. Electricity theft detection in power grids with deep learning and random forests [J]. Journal of Electrical and Computer Engineering,2019,2019:4136874.
[29] BUZAU M M, TEJEDOR-AGUILERA J, CRUZ-ROMERO P, et al. Hybrid deep neural networks for detection of non-technical losses in electricity smart meters [J]. IEEE Transactions on Power Systems,2020,35(2):1254-1263.
[30] ZHENG Z B, YANG Y T, NIU X D, 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.
[31] 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 R, GAO Y P, ZHU Y Q, 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.(in Chinese)
[32] ISSDA. Data from the Commission for Energy Regulation (CER)—smart metering project[EB/OL].2016.
[33] VIEGAS J L, ESTEVES P R, MELÍCIO R, et al. Solutions for detection of non-technical losses in the electricity grid: a review[J]. Renewable and Sustainable Energy Reviews,2017,80:1256-1268.
[34] WANG S W, WANG Q Y, JIANG Z W, et al. A weak coupling of semi-supervised learning with generative adversarial networks for malware classification[C]//2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy. IEEE, 2021: 3775-3782.
[35] NARTEY O T, YANG G W, WU J Z, et al. Semi-supervised learning for fine-grained classification with self-training [J]. IEEE Access,2019,8:2109-2121.
[36] VALE K M O, GORGÔNIO A C, DA LUZ E GORGÔNIO F, et al. An efficient approach to select instances in self-training and co-training semi-supervised methods[J]. IEEE Access, 2021, 10: 7254-7276.
[37] MA J L, WANG X L, XIAO B. An image segmentation method based on simple linear iterative clustering and graph-based semi-supervised learning[C]//2015 International Conference on Orange Technologies (ICOT). Hong Kong, China: IEEE, 2015:10-13.
[38] BOND-TAYLOR S, LEACH A, LONG Y, et al. Deep generative modelling: a comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11):7327-7347.
[39] PAN D W, NIE L Q, KANG W X, et al. UAV anomaly detection using active learning and improved S3VM model [C]//2020 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD). Xi'an, China. IEEE,2020:253-258.