# 基于改进生成对抗网络的虚假数据注入攻击检测方法

夏云舒[1]，王勇[1]，周林[1]，樊汝森[2]
(1. 上海电力大学计算机科学与技术学院，上海市 200120;
2. 国网上海电力公司青浦供电公司，上海市 201799)

**摘 要**：随着新型能源互联网的发展，大规模的传感量测系统为基于数据驱动的虚假数据注入攻击检测方法提供了数据支持，然而攻击样本数据不平衡问题会影响此类方法的性能。提出了基于改进生成对抗网络(generative adversarial network，GAN)和极端随机树的数据重平衡攻击检测模型。首先，为了生成高质量数据，设计GAN的结构使其训练稳定；其次，使用Copula函数构建电力系统状态量之间的空间关联性以适应分布式能源的接入；然后，对改进的GAN进行对抗训练得到重平衡的数据集，采用极端随机树分类器实现攻击检测。此外，设计基于多种分类器的数据有效性指标评估生成数据的质量。通过对比实验对所提方法进行验证，结果表明该方法能生成高质量的量测数据，可以有效解决数据不平衡问题，攻击检测率达98.95%。

**关键词**：虚假数据注入攻击；生成对抗网络；极端随机树；不平衡数据；机器学习；攻击检测

# False Data Injection Attack Detection Method Based on Improved Generative Adversarial Network

XIA Yunshu[1], WANG Yong[1], ZHOU Lin[1], FAN Rusen[2]
(1. School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200120, China;
2. State Grid Qingpu Electric Power Supply Company, Shanghai 201799, China)

**ABSTRACT**: With the development of new-type energy internet, large-scale sensing measurement systems provide data support for data-driven detection of false data injection attack. However, the problem of unbalanced attack data will affect the performance of such methods. Therefore, a data rebalance attack detection model based on improved generative adversarial network (GAN) and extremely randomized tree is proposed. Firstly, the GAN structure is designed to make the training procedure stable enough to generate high-quality data. Secondly, the Copula function is used to construct the spatial correlation between the power system states to adapt to the integration of the distributed energy resources. Then, a rebalanced dataset is obtained through the adversarial training of the improved GAN, and the extremely randomized tree classifier is used to detect the attack. In addition, the data validity index based on multiple classifiers is designed to evaluate the quality of the generated data. The effect of the proposed method is verified by comparative experiments. Results show that the method can generate high-quality measurement data, solve the problem of data imbalance, and the attack detection rate is 98.95%.

This work is supported by the National Natural Science Foundation of China(No. 61772327), Natural Science Foundation of Shanghai(No. 20ZR1455900), National Engineering Laboratory for Big Data Collaborative Security Technology(No. QAX-201803), Science and Technology Commission of Shanghai Municipality (No. 18511105700) and Science and Technology Commission of Shanghai Municipality(No. 19DZ2252800).

**KEYWORDS**: false data injection attack; generative adversarial network; extremely randomized tree; imbalanced dataset; machine learning; attack detection

**中图分类号**: TM 76 **文献标志码**: A **文章编号**: 1000-7229(2022)03-0058-08
DOI: 10.12204/j.issn.1000-7229.2022.03.007

**基金项目**：国家自然科学基金资助项目(61772327)；上海市自然科学基金资助项目(20ZR1455900)；大数据协同安全国家工程实验室项目(QAX-201803)；上海市科委科技创新行动计划(18511105700)；上海市科委电力人工智能工程技术研究中心项目(19DZ2252800)

## 0　引　言

随着智能电网建设的不断推进，传统电力系统与信息控制设备及通信传感网络深度融合，形成电力信息物理系统（cyber physical system，CPS）。新型能源互联网含有高比例的分布式新能源，是一个大型的电力 CPS，能更有效地发挥信息融合带来的优势，但也更容易遭受网络攻击[1]。网络攻击不仅会破坏信息系统的正常功能，还可能传导至物理系统，威胁电力系统的安全运行[2]。虚假数据注入攻击（false data injection attack，FDIA）是一种破坏电网信息完整性的网络攻击，它通过篡改电网量测数据，引起电网误动或拒动，是对电力系统威胁程度较高的攻击方式之一[3]。因此，研究如何提高 FDIA 检测率对于能源互联网安全运行有重要意义。

传统的 FDIA 检测方法主要基于状态估计。文献[4]使用自适应卡尔曼滤波对系统内部状态和噪声作出估计；文献[5]在掌握电网局部信息的情况下，针对单节点与多节点攻击场景提出一种基于非线性状态估计的模型；文献[6]提出了一种基于节点时间相关性的短期状态预测方法，通过计算实际得到的量测量与预测得到的量测量的一致性判断是否发生攻击。

随着能源互联网的建设与发展，量测数据的规模日益增长，传统 FDIA 检测方法逐渐难以应对。近年来，基于人工智能的 FDIA 检测方法被提出，如支持向量回归[7]、卷积神经网络[8]、长短期记忆（long short-term memory，LSTM）网络[9]等。这类方法不需要预先获取电力系统的模型参数信息，有强大的计算能力，能够快速、大规模地检测攻击。然而，基于人工智能的 FDIA 检测方法存在严重的数据不平衡问题。由于 FDIA 发生的频率低，目前在真实电网中还没有捕获 FDIA 的实例[10]，直接在不平衡的数据集上训练得到的算法性能较差[11]，很可能造成误判。

目前解决数据不平衡问题的方法主要基于算法层面和数据层面[12]。前者对传统分类算法进行改进以提高算法对少数类样本的识别能力，如集成学习法、代价敏感法；后者通过数据欠采样、过采样等技术调整样本数据的分布。少类样本合成过采样技术（synthetic minority oversampling technique，SMOTE）通过线性插值在 2 个少数类样本间合成新的样本[13]，是一种经典的数据过采样方法。但在面临不同类型的不平衡数据（如大数据、流数据、高维数据、数值型标签数据）时 SMOTE 方法还存在一些缺陷[14]。

生成对抗网络（generative adversarial network，GAN）能够学习复杂数据的概率分布并生成人工样本[15-16]，已被应用于生成电网中不同类型的数据。文献[17]使用 WGAN（Wasserstein GAN）生成电网量测数据，解决由于数据敏感，研究者难以获取真实可信数据的问题；文献[18]通过训练 GAN 生成 FDIA 攻击数据，达到在电力市场中获得经济利益的目的；文献[19]使用 cGAN（conditional GAN）构建能够逃过电网不良数据检测机制的 FDIA；文献[20]训练 GAN 学习电网正常运行场景下的量测数据分布，以恢复 FDIA 下电力 CPS 数据的完整性。

鉴于以上分析，若能训练 GAN 生成高质量的正常量测数据与 FDIA 攻击数据，对于解决电力 CPS 缺少真实数据、数据不平衡导致攻击检测率低等问题具有多重意义。为此，本文首先考虑 GAN 训练不稳定、模式崩塌等问题对生成数据质量的影响，设计结构更稳定的 CTGAN（conditional tabular GAN）；其次，考虑接入分布式能源后能源互联网各量测数据间的相互影响，使用 Copula 函数构建电力系统状态变量间的空间相关性；然后，使用改进的 GAN 对 FDIA 数据过采样，提出基于极端随机树（extremely randomized trees，ET）的 FDIA 检测模型，解决数据不平衡问题。此外，构建数据有效性指标，基于多个分类器的性能评估生成数据所包含的有效信息。最后，通过对比实验对所提方法进行验证。

## 1　相关技术原理

### 1.1　原始生成对抗网络

原始生成对抗网络由两个互相博弈的神经网络组成，分别为生成器（generator，G）与判别器（discriminator，D）。生成器负责生成新数据，判别器负责判断生成数据的好坏。在这个零和博弈的过程中，判别器旨在分辨真实数据与生成数据，生成器旨在生成足够真实的数据，使判别器无法准确分辨数据的真假，这 2 个网络同时训练，直到达到纳什平衡。GAN 的目标函数如式（1）所示：

$$\min_{G(z)}\max_{D(\boldsymbol{x})}E_{\boldsymbol{x}\sim P_r}\left[\log D(\boldsymbol{x})\right]+E_{\tilde{\boldsymbol{x}}\sim P_g}\left[\log(1-D(\tilde{\boldsymbol{x}}))\right] \tag{1}$$

式中：$E(\cdot)$ 为数学期望；$\boldsymbol{x}$ 表示真实样本；$P_r$ 为真实样本的概率分布；$D(\boldsymbol{x})$ 是判断原始样本为真实样本的概率；$\tilde{\boldsymbol{x}}$ 表示生成器生成的样本，$\tilde{\boldsymbol{x}}=G(\boldsymbol{z})$，$G(\cdot)$ 为生成器生成样本函数，$\boldsymbol{z}$ 为随机噪声；$P_g$ 表示 $\tilde{\boldsymbol{x}}$ 的概率分布；$D(\tilde{\boldsymbol{x}})$ 表示生成样本被判别为真实样本的概率。

### 1.2　改进的生成对抗网络

原始 GAN 在设计之初主要用于生成图像样本，

图像的像素值近似服从高斯分布。然而,许多表格类数据( tabular data) 不服从高斯分布且存在多模态,直接使用原始 GAN 会遇到梯度消失、模式崩塌、不收敛等问题。为了增强原始 GAN 学习表格类数据的能力,并捕捉数据间的相关性,CopulaGAN 将 CTGAN 与 Copula 函数结合[21-22],使用高斯 Copula 函数学习数据的概率分布,描述随机变量间的关联,并改进了原始 GAN 的网络结构和学习步骤。

Copula 函数可用于描述随机变量间的非线性相关性,近年来受到广泛关注[23]。设电网量测数据中 $n$ 维随机变量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$,其中 $x_k(k=1,2,\cdots,n)$ 的边缘分布函数为 $F(x_k)$,令 $u_k=F(u_k)$,故 $u_k$ 为服从 $[0,1]$ 间均匀分布的随机变量,则联合概率分布函数 $H(\boldsymbol{x})$ 与 Copula 分布函数 $C(\boldsymbol{u})$ ( $\boldsymbol{u}=(u_1,u_2,\cdots,u_n)$ ) 之间的关系如式( 2) 所示:

$$H(\boldsymbol{x})=H(x_1,x_2,\cdots,x_n)=C(u_1,u_2,\cdots,u_n)=C(\boldsymbol{u}) \tag{2}$$

对式( 2) 求导可得到对应的联合概率密度函数,如式( 3) 所示:

$$f(x_1,x_2,\cdots,x_n)=c(u_1,u_2,\cdots,u_n)\prod_{k=1}^{n}f(x_k)=c(\boldsymbol{u})\prod_{k=1}^{n}f(x_k) \tag{3}$$

式中:$f(x_1,x_2,\cdots,x_n)$ 为联合概率密度分布;$c(\boldsymbol{u})$ 为 $n$ 维 Copula 密度函数,表示相关性结构;$f(x_k)$ 为 $x_k$ 的边缘概率密度函数。

### 1.3 极端随机树算法

随机森林( random forest, RF) 是一种基于 Bagging 理论的集成学习算法,它不容易陷入过拟合,对噪声和异常值有较好的容忍性,对高维数据分类问题有良好的可扩展性和并行性[24]。

极端随机树是在随机森林的基础上改进得到的,具有更强的随机性[25]。在构成决策树时,它使用所有的训练样本,保证了训练样本的利用率;在划分节点时,它对分裂阈值设置进一步的随机,保证每颗决策树间的结构差异,减少过拟合。因此,使用极端随机树算法构建攻击检测分类器能够提高少数类样本的利用率,提升模型的泛化能力。

## 2 基于 CopulaGAN 的 FDIA 检测模型

基于 CopulaGAN 的 FDIA 检测框架如图 1 所示,由 CopulaGAN 模型训练、FDIA 攻击检测、模型评估 3 部分组成。首先,通过 CopulaGAN 生成器与判别器的对抗训练,得到能够同时生成正常量测数据与攻击数据的数据增强模型;然后,使用该模型对 FDIA 数据过采样,得到平衡的攻击检测数据集,并使用极端随机树分类器进行攻击检测;最后,使用数据增强模型生成相同数目的正常量测数据与攻击数据,得到一个平衡的数据集。以该数据集为训练集,原始数据集为测试集构建多个分类器,利用分类器在测试集上的性能指标评估 CopulaGAN 模型生成数据的有效性。

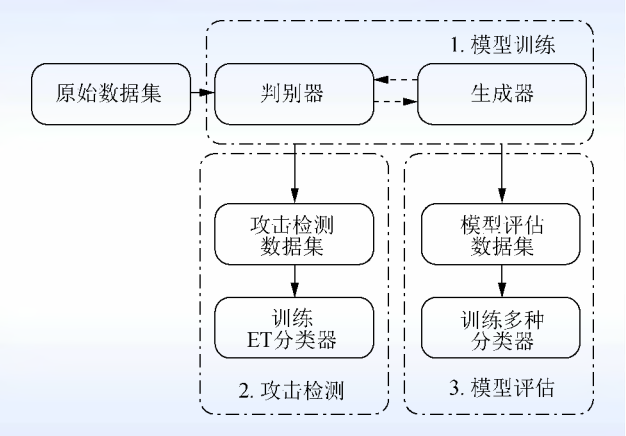


**图 1 基于 CopulaGAN 的 FDIA 检测框架**
**Fig. 1 Structure of FDIA detection based on CopulaGAN method**

### 2.1 CopulaGAN 模型训练

**步骤 1**: 数据预处理。能源互联网作为一个整体,其中各量测值之间相互影响。采用单一 GAN 难以采集不同数据样本间的联系,生成的训练数据与实际电网数据具有较大差别。因此,通过对原始样本中的随机变量进行概率积分变换,使变换后的样本在服从高斯分布的基础上,仍保持数据间的相关性,根据变换后的样本生成接近真实电网的数据,使训练过程更加精确,具体步骤如下:

1) 将原始电力 CPS 量测数据集分为训练集和测试集。

2) 使用高斯 Copula 函数学习数据的概率分布,描述训练集中的 $n$ 维随机变量间的关联关系并转换数据。

3) 对转换后的数据进行归一化。对于离散值,使用独热编码处理;对于连续值,具体转换策略如下:

( 1) 使用变分高斯混合模型估计随机变量的模态个数,拟合得到高斯混合分布;

( 2) 计算数据在每个模态中的概率,得出概率密度函数;

( 3) 由给定的概率密度函数采样得到模态,并用此模态对数据进行归一化处理。

**步骤 2**: GAN 结构设计。首先,为了确保训练过程稳定、收敛快速,引入 WGAN-GP( Wasserstein GAN with gradient penalty) [26] 中梯度惩罚的概念,把 Lipchitz 限制作为一个正则项加到 Wasserstein 损失

上，如式(4) 所示：

$$L = E_{\tilde{x}\sim P_g}[D(\tilde{x})] - E_{x\sim P_r}[D(x)] + \lambda E_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (4)$$

式中：$\lambda \in [0,\infty)$ 为惩罚权重；$\hat{x}$ 表示真实样本和生成样本的线性插值；$E_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$ 表示梯度惩罚项。

其次，为了捕获数据之间所有可能的关联关系，使用全连接网络。对于生成器，使用批标准化(batch normalization) 和 ReLU 激活函数，标量值由 tanh 函数激活，离散值由 softmax 函数激活；判别器中，在每个隐藏层上使用 leaky ReLU 函数和 dropout 方法。此外，使用 Adam 优化器，设置生成器与判别器的学习率衰减率，衰减率为 $10^{-6}$。并采用打包生成对抗网络(packing GAN，PacGAN) [27] 的方法，在将样本传递给判别器之前，将同一类别的 $n$ 个样本(本文选取 $n=10$) 打包，使得判别器能够同时看见多个样本，从一定程度上防止模式崩塌。

**步骤 3**：调整超参数。本文使用基于高斯 Copula 过程的贝叶斯优化方法[28] 寻找 GAN 模型的最优超参数，设计相似性分数 $A$ 为优化目标。贝叶斯优化在选择参数时考虑了选择的方向问题，可以缩短寻优时间，减少寻优过程的盲目性，基于 Copula 过程的贝叶斯优化方法把边缘参数变量变换为均匀分布参数变量，无需考虑参数的边缘分布，简化参数寻优过程。具体步骤如下：

1) 对判别器和生成器进行交替对抗训练。

2) 每训练得到一个模型，就生成一个包含相同数目正常运行数据和 FDIA 数据的数据集，并对该数据集进行反归一化处理。

3) 使用 K-S 检验(Kolmogorov-Smirnov test) 和 KL 散度(Kullback-Leibler divergence，KLD) 计算数据集与训练集之间数据的相似性，得到相似性分数 $A$。

4) 以相似性分数 $A$ 为目标，寻找模型的超参数。GAN 生成的数据与原数据越接近，$A$ 越接近于 1，以得分最高时获得的超参数作为 CopulaGAN 模型的最优超参数。

## 2.2　FDIA 攻击检测

CopulaGAN 模型捕捉样本间的关联性，生成大量攻击样本，使极端随机树分类器在提升少数类样本训练精度的同时选取更全面的特征来寻找全局最优的分裂属性，增强分类效果。具体步骤如下：

**步骤 1**：数据过采样。使用 CopulaGAN 模型对训练集的 FDIA 数据过采样，得到平衡的训练集，用于攻击检测分类器的训练。

**步骤 2**：构建极端随机树分类器。

1) 基于 CART 决策树算法生成基分类器，随机有放回地从攻击检测数据集中抽取所有样本，作为基分类器的训练集。

2) 随机地从训练集所有特征中选取 $m$ 个特征，作为待选择特征库。以基尼指数或信息增益熵选择最优属性进行分裂，且分裂过程不剪枝，对分裂产生的子集进行进一步分裂直到生成一颗决策树。

3) 重复 2)，得到由多颗决策树集成的极端随机树。

4) 使用极端随机树识别测试集的量测数据是否被篡改。

**步骤 3**：分类效果评价。使用混淆矩阵呈现分类器的预测结果，二分类算法检测 FDIA 得到的混淆矩阵如表 1 所示。

**表 1　二分类混淆矩阵**

**Table 1　Confusion matrix for binary classification**

| 真实值 | 预测值 | |
|---|---|---|
| | 1(FDIA) | 0(正常运行) |
| 1(FDIA) | $T_P$ | $F_P$ |
| 0(正常运行) | $F_P$ | $T_P$ |

注：$T_P$ 是真实值为 FDIA，预测值也为 FDIA 样本的数量；$F_N$ 是真实值为 FDIA 而预测值为正常运行样本的数量；$F_P$ 是真实值为正常运行样本而预测值为 FDIA 样本的数量；$T_N$ 是真实值和预测值都为正常运行样本的数量。

算法性能的评估指标可通过混淆矩阵计算，如准确率($\eta_{Accuracy}$)、查准率($\eta_{Precision}$)、查全率($\eta_{Recall}$) 以及查准率与查全率的调和平均值 F1 值($\eta_{F1}$)：

$$\eta_{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (5)$$

$$\eta_{Precision} = \frac{T_P}{T_P + F_P} \quad (6)$$

$$\eta_{Recall} = \frac{T_P}{T_P + F_N} \quad (7)$$

$$\eta_{F1} = \frac{2 \times \eta_{Precision} \times \eta_{Recall}}{\eta_{Precision} + \eta_{Recall}} \quad (8)$$

在检测 FDIA 时，相比于将正常运行样本预测为 FDIA 样本的误判情况，将 FDIA 样本预测为正常运行样本的漏检情况会导致更加严重的后果。因此，在预测结果准确率较高的情况下，算法的查全率越高，检测效果越好。

## 2.3　数据增强模型评估

CopulaGAN 模型能够生成大量的正常运行量测数据和 FDIA 数据，通过调整模型的超参数可以确保模型生成的数据与原始数据相似，并以相似性分数 $A$ 表示。然而，生成的数据不仅需要与原始数据有较高的相似性，还需要提供有效信息，使得分类

器充分学习样本的特征，提高分类器的性能。

因此，本文设计数据有效性指标评估 CopulaGAN 模型生成数据的有效性，生成数据包含越多的有效信息，分类器的性能越好，具体步骤如下：

1) 假设训练集中正常运行量测数据的样本数为 $q$，使用 CopulaGAN 模型分别生成 $q$ 个正常运行数据样本和 $q$ 个 FDIA 数据样本作为模型评估的训练集。

2) 使用多种经典机器学习算法构建多个攻击检测分类器，在训练集上训练，在原始测试集上测试。

3) 以分类器的准确率、F1 值为指标评估 CopulaGAN 模型生成数据所包含的有效信息。

## 3 算例分析

### 3.1 数据集描述

本文使用的数据集来源于密西西比州立大学和美国橡树岭国家实验室[29]，包含 6 类不同程度 FDIA 攻击场景(场景 0 至 5)、1 类电力系统正常运行场景(场景 6)。攻击者通过改变参数值(如电流、电压、序列分量等)模拟有效故障，使操作员误判电力系统的运行情况并误操作。算例选取这 7 类场景作为 FDIA 检测数据集，每条样本中包含 4 个同步相量量测单元(phasor measurement unit，PMU)量测得到的三相电压幅值、电压相角、电流幅值、电流相角。为保持数据分布的一致性、减少过拟合，按照 6∶2∶2 的比例使用分层采样法将数据集划分为训练集、验证集和测试集。表 2 记录了训练集中各类场景的具体情况。实验在 Python3.8 环境下完成。

**表 2 训练集中各类场景的具体描述**
**Table 2 Description of the training set**

| 场景 | 场景描述 | 原始样本数 | 需过采样样本数 |
|---|---|---|---|
| 0 | 线路 1 故障率为[10%,20%) | 1 138 | 2 386 |
| 1 | 线路 1 故障率为[20%,80%) | 1 006 | 2 518 |
| 2 | 线路 1 故障率为[80%,90%] | 1 015 | 2 509 |
| 3 | 线路 2 故障率为[10%,20%) | 1 402 | 2 122 |
| 4 | 线路 2 故障率为[20%,80%) | 1 534 | 1 990 |
| 5 | 线路 2 故障率为[80%,90%] | 1 570 | 1 964 |
| 6 | 线路 1、2 均正常运行 | 3 524 | |

### 3.2 数据增强模型训练与评估

使用基于高斯 Copula 过程的贝叶斯优化方法寻找 GAN 模型的最优超参数，以相似性分数 $A$ 为优化目标，训练 50 轮，超参数优化结果如表 3 所示。

**表 3 模型超参数优化结果**
**Table 3 Hyperparameters of the model**

| 超参数 | 生成器学习率 | 判别器学习率 | 判别器步数 |
|---|---|---|---|
| 默认超参数 | 0.000 20 | 0.000 20 | 1 |
| 优化超参数 | 0.000 91 | 0.000 23 | 5 |

使用 GAN 模型分别生成 7 类场景的数据各 3 524 条，得到用于模型评估的平衡数据集，该数据集的相似性分数 $A$ 如表 4 所示。

**表 4 生成数据的相似性分数**
**Table 4 Similarity score of synthetic data**

| 方法 | $A$ | | | |
|---|---|---|---|---|
| | 调参前 | | 调参后 | |
| | 正常运行 | FDIA | 正常运行 | FDIA |
| CTGAN | 0.772 7 | 0.799 0 | 0.867 8 | 0.893 0 |
| CopulaGAN | 0.770 6 | 0.804 7 | 0.852 7 | 0.882 0 |

由表 4 可见，调参后，模型生成的正常运行数据得分均在 0.85 以上，FDIA 数据得分均在 0.88 以上。由此可见，GAN 可以作为一种数据增强方法，生成大量与原始数据相似的量测数据。

为了评估 GAN 数据增强模型，分别基于 ET、RF、XGBoost 集成学习算法在模型评估数据集上训练多个分类器，根据分类器在原始测试集上的准确率、F1 值指标评估合成数据的有效性，结果如图 2 所示。

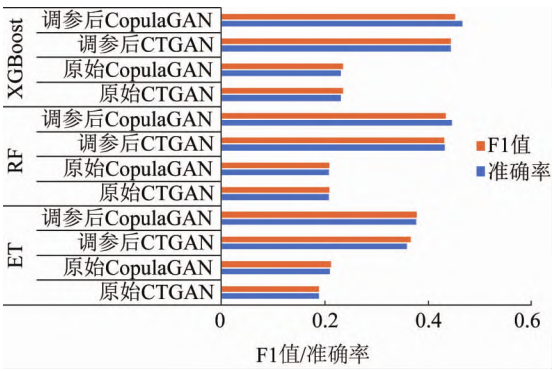


**图 2 生成数据的有效性评估**
**Fig. 2 Effectiveness evaluation of synthetic data**

由图 2 可见，调参后的 CopulaGAN 提高了每个分类器的性能，这说明 GAN 模型生成的数据能够在分类器训练时提供有效信息，且 CopulaGAN 对算法性能的提升效果略高于 CTGAN。

### 3.3 FDIA 检测结果对比

在原始数据集上使用 ET 算法构建攻击检测分类器，与 Adaboost、$K$ 邻近算法($k$-nearest neighbors，KNN)、RF、XGBoost 进行对比，算法的准确率如表 5 所示。

表 5　各种算法在原始数据集上的准确率
Table 5　Accuracy of algorithms on original dataset

| 算法 | Adaboost | KNN | RF | XGBoost | ET |
|---|---|---|---|---|---|
| 准确率 | 0.504 3 | 0.753 3 | 0.915 5 | 0.911 4 | 0.934 3 |

由表5 可见，ET 算法的准确率达到93%，远高于其他机器学习算法，这是因为在训练的过程中 ET 算法充分使用了所有的训练样本，且在划分节点时保证了每颗决策树间的结构差异，提高了分类性能。

对 FDIA 攻击数据进行过采样得到平衡的训练集后，将编号为 0 至 5 的 6 类不同程度的 FDIA 重新编号为 1，编号为 6 的正常运行数据重新编号为 0，如图 3 所示，得到 FDIA 检测的混淆矩阵。

分别使用调参后的 GAN 模型、随机过采样法（random over-sampling，ROS）、SMOTE 方法对原始数据集中的 FDIA 攻击数据进行过采样得到平衡的训练集，使用 ET 算法构建攻击检测分类器，得到的准确率、查全率如表 6 所示。

由表 6 可见，相对于其他数据过采样方法，CopulaGAN 模型提高了 ET 算法的准确率、查全率，减少了 FDIA 漏检的次数。本文提出的基于 CopulaGAN-ET 的检测方法对 FDIA 的检测率达到 98.95%。

表 6　各数据过采样方法下 ET 算法的性能
Table 6　Performance of ET algorithm with diffrent oversampling methods

| 过采样方法 | 准确率 | 查全率 |
|---|---|---|
| ROS | 0.984 6 | 0.986 2 |
| SMOTE | 0.986 5 | 0.987 9 |
| CTGAN | 0.987 6 | 0.989 0 |
| CopulaGAN | 0.987 6 | 0.989 5 |

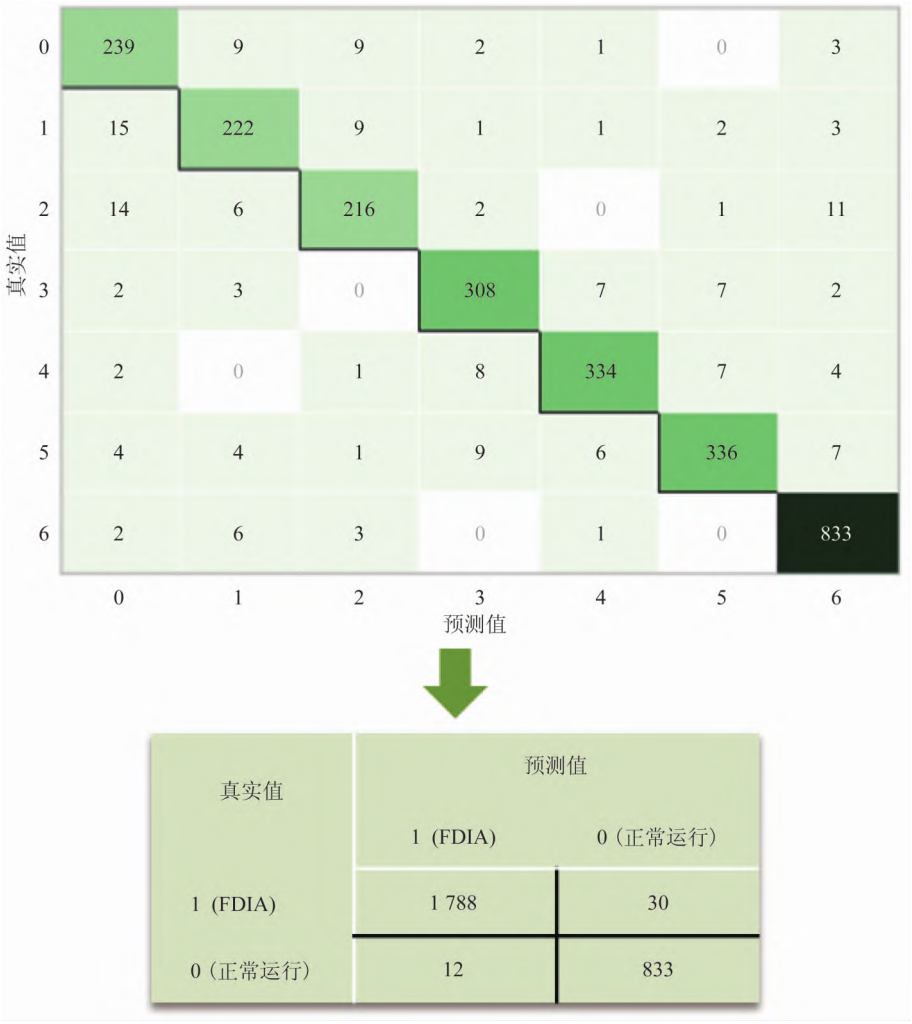


图 3　FDIA 检测混淆矩阵
Fig. 3　Confusion matrix of FDIA detection

## 4 结 论

针对新型能源互联网中 FDIA 攻击检测的数据不平衡问题，本文提出一种基于改进生成对抗网络和极端随机树算法的攻击检测方法。在电力系统攻击数据集上进行实验，得到以下结论：

1) CopulaGAN 模型能够合成质量较高的电力 CPS 正常量测数据与攻击数据，解决了现实中电力量测数据不足的问题。

2) 相较于 KNN、Adaboost、随机森林等经典机器学习算法，ET 算法能够减少攻击检测分类器误判攻击的情况，有效提高 FDIA 检测的准确率。

3) 相较于随机过采样、SMOTE 方法，本文数据增强方法能够提高 FDIA 检测率，减少漏检事件的发生。

4) 使用 GAN 时通常不需要定义规则或约束，便于推广应用于合成电力 CPS 中各种类型的数据。

未来，除了进一步提升 GAN 生成数据的精度，还可以研究 GAN 在新型能源互联网的数据隐私保护、数据压缩中的应用。

## 5 参考文献:


[1] 杨杉，谭博，郭静波. 基于双马尔科夫链的新型能源互联网虚假数据注入攻击检测[J]. 电力自动化设备，2021，41(2)：131-137.
YANG Shan，TAN Bo，GUO Jingbo. Detection of false data injection attack for new-type energy Internet based on double Markov chains [J]. Electric Power Automation Equipment，2021，41(2)：131-137.

[2] 汤奕，陈倩，李梦雅，等. 电力信息物理融合系统环境中的网络攻击研究综述[J]. 电力系统自动化，2016，40(17)：59-69.
TANG Yi，CHEN Qian，LI Mengya，et al. Overview on cyber-attacks against cyber physical power system[J]. Automation of Electric Power Systems，2016，40(17)：59-69.

[3] 王琦，邰伟，汤奕，等. 面向电力信息物理系统的虚假数据注入攻击研究综述[J]. 自动化学报，2019，45(1)：72-83.
WANG Qi，TAI Wei，TANG Yi，et al. A review on false data injection attack toward cyber-physical power system [J]. Acta Automatica Sinica，2019，45(1)：72-83.

[4] 罗小元，潘雪扬，王新宇，等. 基于自适应 Kalman 滤波的智能电网假数据注入攻击检测[J/OL]. 自动化学报，2020 (2020-07-13) [2021-08-12]. https://kns.cnki.net/kcms/detail/detail/11.2109.TP.20200710.1441.001.html.
LUO Xiaoyuan，PAN Xueyang，WANG XinYu，et al. Detection of false data injection attack in smart grid via adaptive kalman filtering [J/OL]. Acta Automatica Sinica，2020 (2020-07-13) [2021-08-12]. https://kns.cnki.net/kcms/detail/detail/11.2109.TP.20200710.1441.001.html.

[5] 赵丽莉，刘忠喜，孙国强，等. 基于非线性状态估计的虚假数据注入攻击代价分析[J]. 电力系统保护与控制，2019，47(19)：38-45.
ZHAO Lili，LIU Zhongxi，SUN Guoqiang，et al. Cost analysis of the false data injection attack based on nonlinear state estimation [J]. Power System Protection and Control，2019，47(19)：38-45.

[6] ZHAO J B，ZHANG G X，LA SCALA M，et al. Short-term state forecasting-aided method for detection of smart grid general false data injection attacks [J]. IEEE Transactions on Smart Grid，2017，8(4)：1580-1590.

[7] KHARE G，MOHAPATRA A，SINGH S N. A real-time approach for detection and correction of false data in PMU measurements [J]. Electric Power Systems Research，2021，191：106866.

[8] 李元诚，曾婧. 基于改进卷积神经网络的电网假数据注入攻击检测方法[J]. 电力系统自动化，2019，43(20)：97-104.
LI Yuancheng，ZENG Jing. Detection method of false data injection attack on power grid based on improved convolutional neural network [J]. Automation of Electric Power Systems，2019，43(20)：97-104.

[9] 陈刘东，刘念. 面向互动需求响应的虚假数据注入攻击及其检测方法[J]. 电力系统自动化，2021，45(3)：15-23.
CHEN Liudong，LIU Nian. False data injection attack and its detection method for interactive demand response [J]. Automation of Electric Power Systems，2021，45(3)：15-23.

[10] WANG J Y，SUN Z W，BAO B，et al. Malicious synchrophasor detection based on highly imbalanced historical operational data [J]. CSEE Journal of Power and Energy Systems，2019，5(1)：11-20.

[11] 李艳霞，柴毅，胡友强，等. 不平衡数据分类方法综述[J]. 控制与决策，2019，34(4)：673-688.
LI Yanxia，CHAI Yi，HU Youqiang，et al. Review of imbalanced data classification methods [J]. Control and Decision，2019，34(4)：673-688.

[12] 陈杰，张浩天，汤奕. 基于改进生成式对抗网络的电网异常数据辨识方法[J]. 电力建设，2021，42(5)：9-15.
CHEN Jie，ZHANG Haotian，TANG Yi. An abnormal data identification method based on improved generative adversarial network [J]. Electric Power Construction，2021，42(5)：9-15.

[13] 张阳，张涛，陈锦，等. 基于 SMOTE 和机器学习的网络入侵检测[J]. 北京理工大学学报，2019，39(12)：1258-1262.
ZHANG Yang，ZHANG Tao，CHEN Jin，et al. Research on network intrusion detection based on SMOTE algorithm and machine learning [J]. Transactions of Beijing Institute of Technology，2019，39(12)：1258-1262.

[14] 石洪波，陈雨文，陈鑫. SMOTE 过采样及其改进算法研究综述[J]. 智能系统学报，2019，14(6)：1073-1083.
SHI Hongbo，CHEN Yuwen，CHEN Xin. Summary of research on SMOTE oversampling and its improved algorithms [J]. CAAI Transactions on Intelligent Systems，2019，14(6)：1073-1083.

[15] GOODFELLOW I，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial networks [J]. Communications of the ACM，2020，63(11)：139-144.

[16] TANAKA F H K D S，ARANHA C. Data augmentation using GANs [EB/OL]. [2021-08-12]. https://arxiv.org/abs/1904.09135

[17] ZHENG X T，WANG B，XIE L. Synthetic dynamic PMU data generation: A generative adversarial network approach [C]//2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA). College Station，TX，USA: IEEE，2019：1-6.

[18] AHMADIAN S, MALKI H, HAN Z. Cyber attacks on smart energy grids using generative adverserial networks [C]//2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Anaheim, CA, USA: IEEE, 2018: 942-946.

[19] MOHAMMADPOURFARD M, GHANAATPISHE F, MOHAMMADI M, et al. Generation of false data injection attacks using conditional generative adversarial networks [C]//2020 IEEE PES Innovative Smart Grid Technologies Europe. Hague, Netherlands: IEEE, 2020: 41-45.

[20] LI Y C, WANG Y Y, HU S Y. Online generative adversary network based measurement recovery in false data injection attacks: A cyber-physical approach [J]. IEEE Transactions on Industrial Informatics, 2020, 16(3): 2031-2043.

[21] XU L, SKOULARIDOU M, CUESTA-INFANTE A, et al. Modeling tabular data using conditional GAN [DB/OL]. (2019-01-01) [2021-08-12]. https://arxiv.org/abs/1907.00503v2.

[22] PATKI N, WEDGE R, VEERAMACHANENI K. The synthetic data vault [C]//2016 IEEE International Conference on Data Science and Advanced Analytics. IEEE, 2016: 399-410.

[23] 赵渊, 刘庆尧, 邝俊威, 等. 电网可靠性评估中相关性变量的非参数 R 藤 Copula 模型 [J]. 中国电机工程学报, 2020, 40(3): 803-812.
ZHAO Yuan, LIU Qingyao, KUANG Junwei, et al. A nonparametric regular vine copula model for multidimensional dependent variables in power system reliability assessment [J]. Proceedings of the CSEE, 2020, 40(3): 803-812.

[24] 王奕森, 夏树涛. 集成学习之随机森林算法综述 [J]. 信息通信技术, 2018, 12(1): 49-55.
WANG Yisen, XIA Shutao. A survey of random forests algorithms [J]. Information and Communications Technologies, 2018, 12(1): 49-55.

[25] CAMANA ACOSTA M R, AHMED S, GARCIA C E, et al. Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks [J]. IEEE Access, 2020, 8: 19921-19933.

[26] GULRAJANI I, ALMED F, ARJOVSKY M, et al. Improved training of wasserstein gans [DB/OL]. (2017-03-31) [2021-0812]. https://arxiv.org/abs/1704.00028.

[27] LIN Z N, KHETAN A, FANTI G, et al. PacGAN: The power of two samples in generative adversarial networks [J]. IEEE Journal on Selected Areas in Information Theory, 2020, 1(1): 324-335.

[28] SMITH M J, SALA C, KANTER J M, et al. The machine learning bazaar: Harnessing the ML ecosystem for effective system development [C]//SIGMOD '20: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 2020: 785-800.

[29] BEAVER J M, BORGES-HINK R C, BUCKNER M A. An evaluation of machine learning methods to detect malicious SCADA communications [C]//2013 12th International Conference on Machine Learning and Applications. IEEE, 2013: 54-59.

**收稿日期**: 2021-09-12

**作者简介**:

夏云舒(1997),女,硕士研究生,通信作者,从事电力信息物理系统安全方面的研究工作,E-mail: carriexia@163.com;

王勇(1974),男,博士,教授,从事网络信息安全方面的研究工作,E-mail: wy616@126.com;

周林(1968),男,硕士,副教授,研究方向为计算机网络与应用,E-mail: lin.zhou@shiep.edu.cn;

樊汝森(1989),男,硕士,研究方向为输变电工程建设管理,Email: fanrusen107@163.com。

(编辑　张小飞)