DOI:10.12158/j.2096-3203.2024.02.014

融合无监督和有监督学习的虚假数据注入攻击检测

黄冬梅[1]，王一帆[2]，胡安铎[1]，周游[3]，时帅[2]，胡伟[4]
（1. 上海电力大学电子与信息工程学院，上海 201306；2. 上海电力大学电气工程学院，
上海 200090；3. 国网江苏省电力有限公司苏州供电分公司，江苏 苏州 215004；
4. 上海电力大学经济与管理学院，上海 201399）

摘 要：虚假数据注入攻击（false data injection attack，FDIA）是智能电网安全与稳定运行面临的严重威胁。文中针对 FDIA 检测中存在的有标签数据稀少、正常和攻击样本极不平衡的问题，提出了融合无监督和有监督学习的 FDIA 检测算法。首先引入对比学习捕获少量攻击数据特征，生成新的攻击样本实现数据扩充；然后利用多种无监督检测算法对海量的无标签样本进行特征自学习，解决有标签样本稀缺的问题；最后将无监督算法提取的特征与历史特征集进行融合，在新的特征空间上构建有监督 XGBoost 分类器进行识别，输出正常或异常的检测结果。在 IEEE 30 节点系统上的算例分析表明，与其他 FDIA 检测算法相比，文中方法增强了 FDIA 检测模型在有标签样本稀少和数据不平衡情况下的稳定性，提升了 FDIA 的识别精度并降低了误报率。

关键词：虚假数据注入攻击（FDIA）；有监督学习；无监督学习；对比学习；数据扩充；特征融合
中图分类号：TM712 **文献标志码**：A **文章编号**：2096-3203（2024）02-0134-08

0 引言

近年来，电网与信息通信网络不断融合，成为信息空间与物理空间深度耦合的信息物理系统（cyber-physical system，CPS）[1-4]。信息与物理两侧的协同互动提升了 CPS 运行能力，同时也可能引入新的安全风险。虚假数据注入攻击（false data injection attack，FDIA）是针对电力系统状态估计的网络攻击手段，通过篡改数据采集和监控（supervisory control and data acquisition，SCADA）系统中的量测数据，破坏信息的完整性和正确性，导致状态估计出现错误，引起误操作[5-6]。因此，研究高效的 FDIA 检测算法对保障 CPS 安全稳定运行具有重要意义。

基于机器学习的 FDIA 检测可以分为有监督学习检测和无监督学习检测。有监督学习检测利用训练集的少量标签样本，学习正常和异常数据的特征[7-9]，从而判断测试集数据是否异常。文献[10]提出神经网络和随机森林这 2 种有监督算法相结合的模型，采用神经网络算法从原始量测数据中提取特征集，再输入随机森林分类器进行检测，但在正常数据与虚假数据极不平衡时检测效果较差，易造成误分类。文献[11]提出采用重采样技术来解决有监督算法中的数据不平衡问题，提高异常检测能力，但模型训练依赖于有标签数据，当有标签数据稀少时，模型的分类精度无法满足检测需求。

收稿日期：2023-09-06；修回日期：2023-11-09
基金项目：国家社科基金资助项目（19BGL003）

无监督学习检测利用 FDIA 数据的几何特性，挖掘隐含的正常和异常特征，如 K 均值聚类[12]、局部离群因子[13]等。文献[14]提出基于深度信念网络（deep belief network，DBN）的 FDIA 检测模型，利用 DBN 对原始数据逐层进行特征提取，然后构建分类模型，由于缺乏有标签数据的引导，计算复杂度高，检测速度慢。文献[15]提出一种基于无监督学习的检测算法，当训练集和测试集相似性高且数据丰富、有代表性时该算法有效，然而实际电力系统中发生 FDIA 的频率低，攻击数据与正常数据极不平衡，若直接在不平衡的数据集训练，通常检测性能表现不佳。

鉴于无监督学习和有监督学习在 FDIA 检测中存在的问题，文中将图像领域的对比学习[16]引入 FDIA 检测领域，建立一种融合无监督和有监督学习的 FDIA 检测方法。首先引入对比学习对原始量测数据集中的少量攻击数据进行扩充，利用多种无监督检测算法从海量无标签样本中提取重要特征，生成攻击评分函数（attack scoring function，ASF）；然后将 ASF 特征与原始特征集融合，生成一个新的特征空间，提高样本特征的聚合性；最后输入到有监督分类器 XGBoost 进行检测。实验结果表明，在有标签数据稀少和数据集极不平衡时，所提方法仍能有效实现 FDIA 检测，且分类精度和泛化性能优于其他方法。

1 FDIA 攻击原理

电力系统控制中心通过 SCADA 系统获取的测

量值来估计系统的实时状态,然而由于 CPS 是复杂的非线性系统,高度非线性化的交流状态估计无法保证收敛到全局最优解。为简化问题,通常将交流模型在操作点附近泰勒展开,得到近似为线性的直流状态估计模型[17],可以表示为:

$$\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} \tag{1}$$

式中:$\boldsymbol{Z}$ 为 SCADA 的量测值,包括线路潮流和节点注入功率;$\boldsymbol{H}$ 为量测雅可比矩阵;$\boldsymbol{x}$ 为状态变量,表示总线的电压相角;$\boldsymbol{e}$ 为服从零均值高斯分布的测量噪声。在传统的加权最小二乘估计中,其目标函数为:

$$J(\boldsymbol{x}) = \operatorname{argmin}(\boldsymbol{Z} - \boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{Z} - \boldsymbol{H}\boldsymbol{x}) \tag{2}$$

式中:$\boldsymbol{R}$ 为权重矩阵。对式(2)求解可得到直流状态估计模型的系统状态变量为:

$$\tilde{\boldsymbol{x}} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Z} \tag{3}$$

检测电力系统中的不良数据,可以采用坏数据检测(bad data detection,BDD)机制。基于测量值和估计值之间的残差 $\boldsymbol{r}$ 进行统计检验,如式(4)所示。

$$\| \boldsymbol{r} \| = \| \boldsymbol{Z} - \boldsymbol{H}\tilde{\boldsymbol{x}} \| \tag{4}$$

若 $\| \boldsymbol{r} \| < \upsilon$,$\upsilon$ 为判断阈值,则测量数据中至少存在一个不良数据。由于 $\boldsymbol{r}^2 \sim \chi^2$,阈值 υ 可以根据显著性水平 α 来确定[18]。

FDIA 中攻击者设计操作,通过篡改测量值实现隐藏攻击,欺骗电网中的状态估计的 BDD 模块,以免触发警报。设 $\boldsymbol{Z}_{\mathrm{bad}} = \boldsymbol{Z} + \boldsymbol{a}$ 为攻击后的测量值,$\boldsymbol{a}$ 为注入系统的非零攻击向量。攻击后的残差见式(5)。

$$\begin{aligned} \| \boldsymbol{r}_{\mathrm{bad}} \| &= \| \boldsymbol{Z}_{\mathrm{bad}} - \boldsymbol{H}\tilde{\boldsymbol{x}}_{\mathrm{bad}} \| = \\ &\| \boldsymbol{Z} + \boldsymbol{a} - \boldsymbol{H}(\tilde{\boldsymbol{x}} + \boldsymbol{c}) \| = \\ &\| \boldsymbol{Z} - \boldsymbol{H}\tilde{\boldsymbol{x}} + (\boldsymbol{a} - \boldsymbol{H}\boldsymbol{c}) \| \end{aligned} \tag{5}$$

式中:$\tilde{\boldsymbol{x}}_{\mathrm{bad}}$ 为攻击后估计状态变量;$\boldsymbol{c}$ 为攻击前后估计状态变量的偏差向量。

如果注入的攻击向量满足式(6),则 FDIA 不会改变系统残差,可以避过电网 BDD 模块的残差检测,实现 FDIA。

$$\boldsymbol{a} = \boldsymbol{H}\boldsymbol{c} \tag{6}$$

2 FDIA 检测模型

融合无监督学习和有监督学习的 FDIA 检测模型主要分为 3 个阶段。第一阶段利用对比学习,扩充攻击样本的数量,提升模型在不平衡数据集上的表现;第二阶段采用多种无监督算法对海量的无标签数据进行特征提取,解决数据集中有标签数据稀少的问题;第三阶段将提取的特征与历史特征集融合,形成新的特征空间,提升数据特征的聚合性,然后使用有监督分类器 XGBoost 检测 FDIA,同时对增广的特征空间进行剪枝,控制计算复杂度,提升检测效率。FDIA 检测流程如图 1 所示,其中分类为异常的数据将触发报警模块,实现 FDIA 检测,分类为正常则不报警。

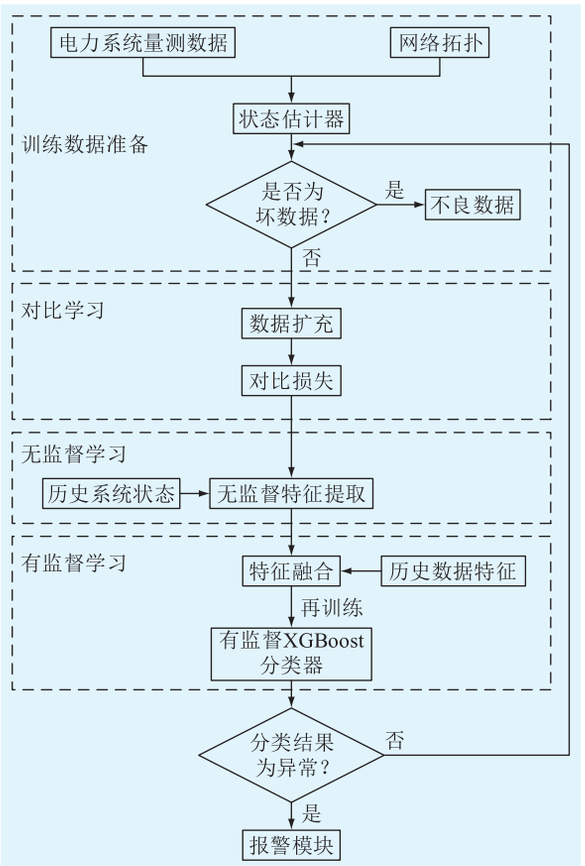


图 1 FDIA 检测算法流程

Fig.1 Flow chart of FDIA detection algorithm

2.1 基于对比学习的攻击数据扩充

实际电网中发生 FDIA 的概率较低且攻击样本数量较少,因此受攻击样本远少于正常样本。为了增加 FDIA 事件的样本数量,采用对比学习扩充攻击样本的数量,解决数据不平衡引起的 FDIA 高误报率和训练效率低的问题。对比学习的核心思想是将正常样本和攻击样本在特征空间进行对比,学习样本的特征表示,使得新增样本与攻击样本的特征表示尽可能接近,而与正常样本的特征表示尽可能不同[19],模型如图 2 所示。

设原始攻击数据集由样本 $Y=\{\boldsymbol{v}_i\}$ 组成,每个样本为 d 维向量。从攻击样本 $\boldsymbol{v}_i$ 中随机选取其中的 m 维进行构造切分,生成的 2 个样本[20]记为:

$$\varphi(\boldsymbol{v}_i) = \{a_{i,j}, b_{i,j}\} \quad 1 \leqslant j \leqslant d - m + 1 \tag{7}$$

其中:

$$\begin{cases} a_{i,j} = \{\boldsymbol{v}_{i,j}, \boldsymbol{v}_{i,j+1}, \cdots, \boldsymbol{v}_{i,j+m-1}\} \\ b_{i,j} = \{\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \cdots, \boldsymbol{v}_{i,j-1}\} \cup \{\boldsymbol{v}_{i,j+m}, \cdots, \boldsymbol{v}_{i,d}\} \end{cases} \tag{8}$$

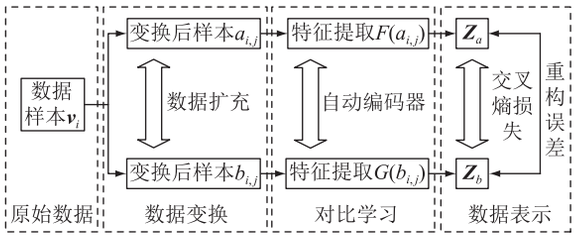


图 2 对比学习流程

Fig.2 Flow chart of comparative learning

将 2 个样本分别输入特征提取器，得到编码后的特征 $F(a_{i,j})$、$G(b_{i,j})$，其中 F、G 为自动编码器。采用对比框架构造攻击样本 $\boldsymbol{Z}_a$、$\boldsymbol{Z}_b$[21]，最大限度提高 $F(a_{i,j})$ 与原始攻击样本 $G(b_{i,j})$ 的相似性，同时最小化正常样本 $G(b_{i,j'})$ 的相似性，其中 $j \neq j'$。

为了使生成的新样本接近原始攻击样本，通过计算生成的新样本与原始攻击样本中心之间的重构误差，剔除重构误差较大的新样本，让生成的新样本均趋向攻击样本中心，如图 3 所示。

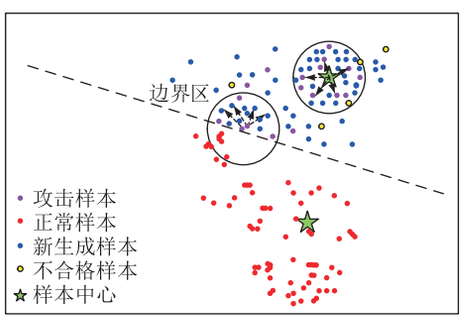


图 3 对比学习扩充攻击样本

Fig.3 Comparative learning expands attack samples

利用自动编码器 F 和 G 的重构误差 $e_{\text{f\_loss}}$、$e_{\text{g\_loss}}$ 组成建立对比模型的损失函数，可以表示为：

$$e_{\text{loss}} = e_{\text{f\_loss}} + e_{\text{g\_loss}} \tag{9}$$

模型的损失函数 e_{loss} 采用交叉熵函数 l：

$$l(F,G,j) = -\ln \frac{\exp(F(a_{i,j})G(b_{i,j})/\tau)}{\sum_{j'=1}^{m}\exp(F(a_{i,j})G(b_{i,j'})/\tau)} \tag{10}$$

式中：τ 为温度系数。

2.2 无监督特征提取

无监督特征提取模块中采用 ASF 提取数据特征。ASF 对于给定的具有 p 个特征的 n 维数据 $\boldsymbol{X} \in \mathbf{R}^{n\times p}$，建立一个映射 $\boldsymbol{\Phi}(\cdot)$：$\boldsymbol{X} \in \mathbf{R}^n$，为矩阵 $\boldsymbol{X}$ 的每一行分配一个实值输出。

不同的无监督算法对应的 ASF 不同，文中采用多种无监督学习算法从样本数据中提取特征，将输出的多个 ASF 值并入原始特征空间，实现样本的特征空间扩展。

将无监督算法生成的 k 维 ASF 组合起来，得到矩阵：

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \cdots \quad \boldsymbol{\Phi}_k] \tag{11}$$

在数据集 $\boldsymbol{X}$ 上应用 $\boldsymbol{\Phi}(\cdot)$，得出 ASF 矩阵 $\boldsymbol{\Phi}(\boldsymbol{X})$ 为：

$$\boldsymbol{\Phi}(\boldsymbol{X}) = [\boldsymbol{\Phi}_1^{\mathrm{T}}(\boldsymbol{X}) \quad \boldsymbol{\Phi}_2^{\mathrm{T}}(\boldsymbol{X}) \quad \cdots \quad \boldsymbol{\Phi}_k^{\mathrm{T}}(\boldsymbol{X})] \in \mathbf{R}^{n\times k} \tag{12}$$

多个无监督算法作为特征变换的评分函数，形成异构的基函数，可以捕获特定数据集中异常值的不同特征，保持多样性和准确性之间的平衡，从而提高模型的泛化能力[22]。

2.3 特征融合

结合 2.2 节 ASF 提取的特征和原始特征，融合形成新的特征空间，提高样本特征的聚合性。采用 K 近邻（K-nearest neighbor，KNN）、支持向量机（support vector machine，SVM）、局部异常因子（local outlier factor，LOF）、局部异常概率（local outlier probabilities，LOOP）、孤立森林（isolation forest，iForest）这 5 个基于距离和密度的无监督学习算法提取 ASF 特征。不同无监督算法对样本进行特征提取后输出不同类别特征向量，拼接处理后得到最终特征向量。

设 $\boldsymbol{L}$ 为原始量测值构成的特征空间，记为：

$$\boldsymbol{L} = [(\boldsymbol{x}_1,\boldsymbol{y}_1) \quad (\boldsymbol{x}_2,\boldsymbol{y}_2) \quad \cdots \quad (\boldsymbol{x}_n,\boldsymbol{y}_n)] \in \mathbf{R}^{n\times p} \tag{13}$$

式中：$(\boldsymbol{x}_n,\boldsymbol{y}_n)$ 为第 n 个特征对。结合提取的 k 维新特征，可以构造出一个新的特征数据集。

$$l = p + k \tag{14}$$

最终形成融合后的特征空间：

$$\boldsymbol{F}_{\text{new}} = [\boldsymbol{X} \quad \boldsymbol{\Phi}(\boldsymbol{X})] \in \mathbf{R}^{n\times l} \tag{15}$$

5 个分类器训练结束后，每个分类器都会得到长度为 k 维的特征向量。最终将基分类器结果拼接后形成 l 维的特征。特征融合框架如图 4 所示。

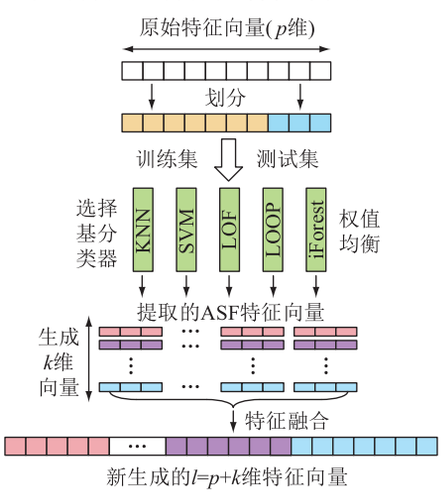


图 4 特征融合框架

Fig.4 Framework of feature fusion

2.4 有监督学习检测

有监督学习检测是将异常攻击视为二分类或多分类问题,用详细标记的正常、异常标签样本训练模型,提取正常、异常样本之间更具区分性的特征[23]。

在特征融合后构成的新特征向量上应用有监督分类器 XGBoost 生成最终输出,具体的预测模型为:

$$\tilde{y}_l = \sum_{k=1}^{K} f_k(x_i) \quad f_k \in N \tag{16}$$

式中:$\tilde{y}_l$ 为预测值;K 为树的数目;f_k 为与第 k 棵树的结构和叶子权重值有关的函数;x_i 为输入的第 i 个样本;N 为决策树组成的函数空间。

第 t 次迭代时 XGBoost 的目标函数为:

$$O^{(t)} = \sum_{i=1}^{n} l(y_i, \tilde{y}_l^{(t)}) + \sum_{k=1}^{K} \Omega(f_k) \tag{17}$$

式中:y_i 为第 i 个样本的实际攻击类别;$\tilde{y}_l^{(t)}$ 为第 t 次迭代的预测攻击类别;$l(y_i, \tilde{y}_l^{(t)})$ 为损失函数,表示预测攻击类别与实际攻击类别的差异;$\Omega(f_k)$ 为正则项。

将目标函数泰勒二阶展开,使用贪心枚举法找到梯度提升树,得到最优 XGBoost 模型[24]。

3 算例分析

3.1 实验数据与仿真设置

仿真系统选用 IEEE 30 节点系统作为测试环境,收集纽约独立运营商 2020 年 2 月—6 月的负荷数据,其网络拓扑、节点数据、支路参数等均从 MATPOWER中获得,并通过潮流计算生成正常量测数据[25]。攻击向量按照文献[26]中的单点注入攻击方法进行设计,确保 FDIA 能够绕过 BDD 系统。

实验中使用 5 个基于距离和密度的估计值评分函数作为特征转换,分别为 KNN、SVM、LOF、LOOP、iForest。考虑到 ASF 提取特征的准确性依赖于领域参数,对于不同的无监督算法评分函数,设置不同领域参数。文中定义 KNN、LOF 的领域参数为{1,5,10,15,…,100};在 SVM 中使用了交叉验证来优化参数并实现自动寻优,领域参数设置为{0.01,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.99};由于 LOOP 算法在大数据集上的计算复杂度较高,因此定义领域参数为较小的范围{1,3,5,10};对于 iForest,为了达到算法的检测性能,需要采取较大的领域参数{10,20,50,70,100,150,200,500}。值得注意的是,选取不同值会产生不同的 ASF。

随机选择 9 000 条量测样本作为实验数据,将数据集中正常样本标记为 0,攻击样本标记为 1。通过去除均值和缩放到单位方差来对样本进行归一化,按照 6∶4随机抽取样本制作训练集和测试集。

3.2 评估指标

文中采用精确率、召回率和接受者操作特征(receiver operating characteristic,ROC)作为 FDIA 检测评价指标,验证所提 FDIA 检测方法的有效性和可行性。

精确率 I_{pre} 计算如下:

$$I_{\text{pre}} = \frac{\rho_{\text{TP}}}{\rho_{\text{TP}} + \rho_{\text{FP}}} \tag{18}$$

式中:ρ_{TP} 为分类器预测为攻击实际也是攻击的数量;ρ_{FP} 为分类器预测为攻击而实际是正常的数量。I_{pre} 越高,误检率越低,分类器性能越好。

召回率 I_{rec} 计算如下:

$$I_{\text{rec}} = \frac{\rho_{\text{TP}}}{\rho_{\text{TP}} + \rho_{\text{FN}}} \tag{19}$$

式中:ρ_{FN} 为分类器预测为正常,实际是攻击的数量。I_{rec} 越高则分类器性能越好。

根据混淆矩阵可以计算分类器的真阳性率 I_{tpr} 和假阳性率 I_{fpr},分别可以反映分类器的检出率和误检率,定义为:

$$I_{\text{tpr}} = I_{\text{rec}} = \frac{\rho_{\text{TP}}}{\rho_{\text{TP}} + \rho_{\text{FN}}} \tag{20}$$

$$I_{\text{fpr}} = \frac{\rho_{\text{FP}}}{\rho_{\text{FP}} + \rho_{\text{TN}}} \tag{21}$$

式中:ρ_{TN} 为分类器预测为正常,实际也是正常的数量;I_{tpr} 为正确检测出的攻击数据个数占攻击数据总数的比例,即检出率;I_{fpr} 为分类器预测为攻击而实际是正常占所有正常的数量的比例,即误检率。

3.3 仿真结果分析

3.3.1 对比学习扩充攻击样本效果评估

为了衡量对比学习对算法迭代收敛的影响,选取量测数据集中攻击样本分别为 100、200、300、500、1 000 组进行模型训练,收敛效果见图 5。图 5 中,加入对比学习前,数据不平衡程度较大,损失达到 13.07%;加入对比学习后,增加新的攻击检测数据集,降低了数据的不平衡度,交叉熵损失降至 1.10%。可以看出,加入对比学习对攻击数据集预处理后,正常样本和攻击样本的不平衡度会降低,模型可以进一步收敛,同时增强模型的分类效果。

3.3.2 有标签样本稀少处理方法比较

为了衡量量测数据集中的有标签样本个数对检测的影响,使用文中模型与参考文献[27]的多种有监督与无监督单分类器、无监督集成(Ens_Un)模

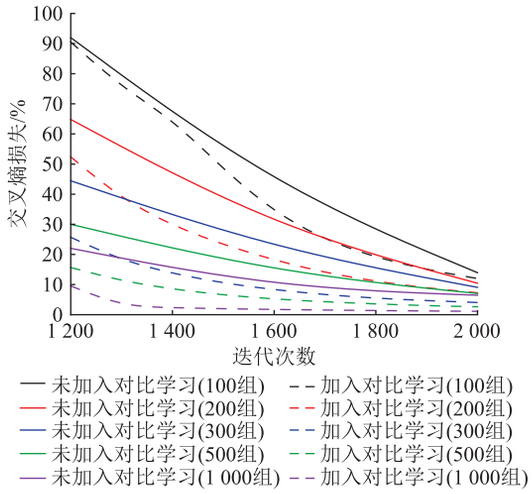


图 5　加入对比学习前后的收敛效果

Fig.5　Convergence effect before and after contrastive learning

型、有监督集成(Ens_Su)模型进行对比实验,其中数据集包含 95%的无标签数据样本和 5%的有标签数据样本。图 6 为检测的精确率和召回率结果,其中 LR 为逻辑回归分类器,DT 为决策树分类器。可以看出,使用 Ens_Un 模型的检测器性能略优于 Ens_Su模型,与文献[27]的实验结果一致。但是,文献[27]中表现性能较好的 Ens_Un 模型检测精确率和召回率仅为 70%左右,检测精度并不高,这是因为单分类器在集成中采用多数投票的方法进行选择,对 FDIA 的识别整体性能偏低,难以保护电网的数据安全。文中模型检测召回率为 90.49%、精确率为 94.25%,明显优于其他模型,主要原因是文中模型对于海量的无标签数据,采用无监督算法提取 ASF 特征,与历史特征集进行融合,增加了数据特征的多样性和完整性,有利于检测出电网遭受攻击。

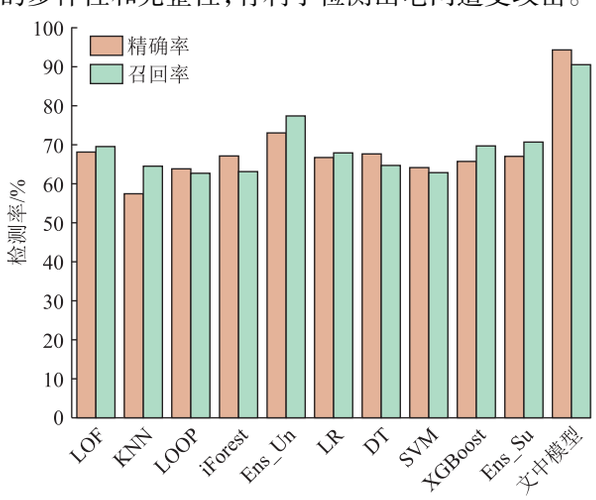


图 6　各检测器检测指标对比

Fig.6　Comparison of the detection indicators of detectors

ROC 曲线以假阳性率为横轴、真阳性率为纵轴,能够描述检出率与误检率之间的相对关系,如图 7 所示。ROC 曲线下的面积(area under curve,AUC)是衡量检测器性能的重要指标,如果检测器性能优越,AUC 值会逼近于 1。各检测器的 AUC 值见表 1,可以看出对比实验的检测模型 AUC 值低于 75%,而文中所提出的检测方法,在低误检率的情况下,能保证较高的检测率,AUC 值达到 97.2%,说明文中模型可以准确检测攻击,分类性能较好,可以有效保护电网安全。

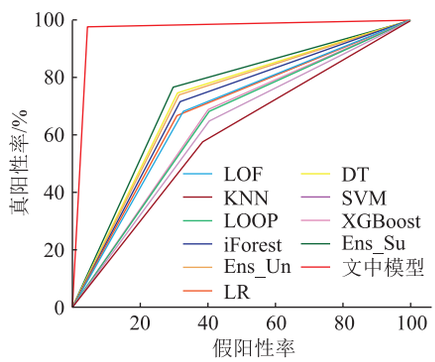


图 7　各检测器的 ROC 曲线对比

Fig.7　Comparison of ROC curves of detectors

表 1　各检测器的 AUC 值

Table 1　AUC values of detectors

检测器	AUC 值	检测器	AUC 值
LOF	0.692	LR	0.701
KNN	0.574	DT	0.727
LOOP	0.641	SVM	0.668
iForest	0.718	XGBoost	0.693
Ens_Un	0.725	Ens_Su	0.734
文中模型	0.972		

3.3.3　不同比例正负样本检测效果

由于在实际电力系统中,发生 FDIA 的频率较低,攻击类的实例远少于正常类的实例,即数据集极不平衡。为了验证文中模型在极不平衡数据集下对 FDIA 识别的适用性,选取了处理不平衡数据集的算法,如 SMOTE 过采样技术、卷积神经网络[28]模型和半监督自动编码器模型[29],与文中模型进行对比。其中 SMOTE 过采样设置惩罚函数为 0.1,当测试集精确率基本不变时完成训练。卷积神经网络模型设置 5 层卷积层、5 层最大池化层和 3 层全连接层,激励函数选择 Sigmoid 函数,学习率取0.01,采用随机梯度下降优化器,为了避免过拟合,交叉熵损失函数加入正则化项。自动编码器参数设置为一层编码器和一层解码器,使用交叉熵作为损失函数,采用 Adam 优化器。各检测方案在不同正负

样本比例下的精确率见表 2。一方面,随着正负样本比例不平衡性的增大,各检测方案的检测精度都在下降,这是因为正负样本比例增大,算法的识别率降低。另一方面,对比发现即使在正负样本比例为 50∶1的极度不平衡条件下,文中模型的检测精确率仍能达到 90%。

表 2 不同正负样本比例下的检测精确率

Table 2 Precision of detection under different ratios of positive and negative samples %

检测方案	8∶1	16∶1	19∶1	40∶1	50∶1
SMOTE 过采样	69.87	63.17	58.76	58.14	57.39
卷积神经网络	82.01	80.51	72.97	66.87	59.20
自动编码器	85.42	81.60	79.66	74.31	70.13
文中模型	96.10	94.77	94.25	91.85	91.09

为进一步验证文中模型对不平衡数据的检测能力,图 8 展示了过采样、卷积网络和文中模型 3 种算法在正负样本比例为 50∶1时的 F1 值箱型图。

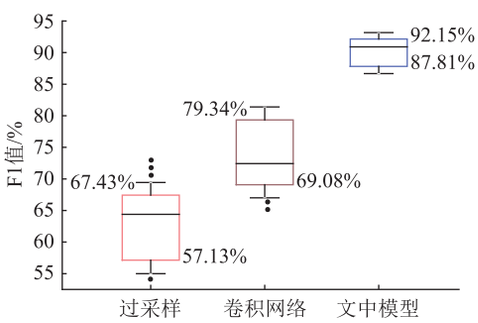
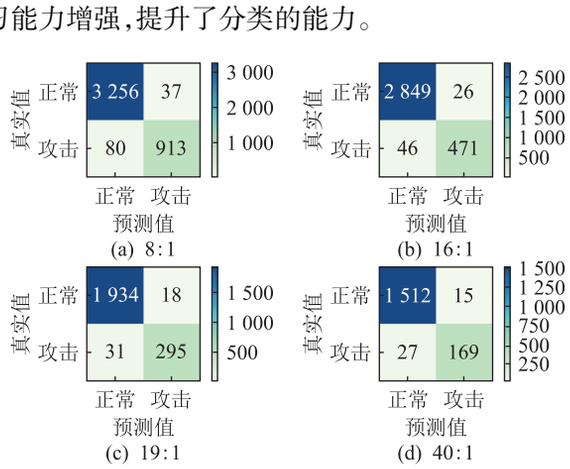


图 8 正负样本比例为 50∶1的 F1 值箱型图

Fig.8 F1 value box diagram with 50∶1 ratio of positive and negative samples

由图 8 可以看出文中模型检测结果的 F1 值为 90.65%,明显优于其他算法,且实验结果分布更集中,表明模型训练时可以学习到更多的受攻击类别的特征,模型分类能力更好。图 9 为文中模型的 FDIA 检测混淆矩阵,可以看出,经过对比学习进行数据平衡和无监督学习改进的特征融合后,模型学习能力增强,提升了分类的能力。



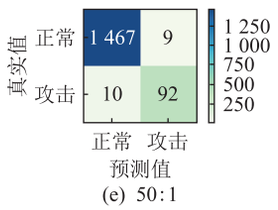


图 9 文中模型的 FDIA 检测混淆矩阵

Fig.9 FDIA detection confusion matrix of model in this paper

4 结论

针对电力系统中实际 FDIA 有标签样本稀少、正常和攻击样本不均衡导致常规 FDIA 检测模型检测精度差,泛化能力弱的问题,文中提出了融合无监督和有监督学习的 FDIA 检测算法。研究结果表明:

(1) 在训练阶段引入对比学习生成高质量的攻击数据,进行数据扩充,能够降低正常样本和攻击样本的不平衡度。采用对比学习扩充攻击样本后收敛性能明显提升,增加了模型对 FDIA 识别的可靠性。

(2) 采用无监督学习进行特征挖掘归纳,使用各种无监督算法特征提取生成 ASF,使得模型参数适应有标签样本稀少的特征表达,在精确率、召回率及 ROC 指标方面均优于传统机器学习模型,解决了有标签样本稀缺的问题。

(3) 结合基于特征融合的优势,将生成的 ASF 新特征和历史数据特征进行融合,提高了样本特征的聚合性。进一步地,通过实验结果验证了文中方法有较高的检测精度,并且检测效率较高,具有更好的泛化能力。

鉴于实际电力网络量测数据的完整性与准确性偏差,未来可以研究考虑数据清洗、补全等预处理的 FDIA 检测方法。

参考文献:

[1] LI J,SUN C W,SU Q Y. Analysis of cascading failures of power cyber-physical systems considering false data injection attacks [J]. Global Energy Interconnection,2021,4(2):204-213.

[2] 秦博雅,刘东. 电网信息物理系统分析与控制的研究进展与展望[J]. 中国电机工程学报,2020,40(18):5816-5827.
QIN Boya,LIU Dong. Research progresses and prospects on analysis and control of cyber-physical system for power grid[J]. Proceedings of the CSEE,2020,40(18):5816-5827.

[3] 李妍莎,蔡晔,曹一家,等. 面向联合检修的电力信息物理系统输电线路脆弱相关性辨识[J]. 电力系统保护与控制,2022,50(24):120-128.
LI Yansha,CAI Ye,CAO Yijia,et al. Vulnerable correlation identification of a transmission line in the power cyber physical

system for federated maintenance[J]. Power System Protection and Control,2022,50(24):120-128.

[4] 胡起凡,郝丽丽,陈从霜. 智能配电网信息物理系统故障协调恢复策略[J]. 电力电容器与无功补偿,2023,44(1):103-112.
HU Qifan,HAO Lili,CHEN Congshuang. Coordination recovery strategy of cyber physical system fault of smart distribution network[J]. Power Capacitor & Reactive Power Compensation,2023,44(1):103-112.

[5] 张鹏,熊雅琴,蹇洁. 基于多目标双层规划的智能电网虚假数据注入攻击研究[J]. 运筹与管理,2023,32(1):22-26.
ZHANG Peng,XIONG Yaqin,JIAN Jie. Research on false data injection attack of smart grid based on multi-objective bi-level programming[J]. Operations Research and Management Science,2023,32(1):22-26.

[6] 伊娜,徐建军,陈月,等. 电力 CPS 多阶段低代价虚假数据注入攻击方法[J]. 浙江电力,2023,42(11):39-47.
YI Na,XU Jianjun,CHEN Yue,et al. A multi-stage low-cost false data injection attack method for power CPS[J]. Zhejiang Electric Power,2023,42(11):39-47.

[7] JENA P K,GHOSH S,KOLEY E,et al. An ensemble classifier based scheme for detection of false data attacks aiming at disruption of electricity market operation[J]. Journal of Network and Systems Management,2021,29(4):1-26.

[8] 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.

[9] HU P F,GAO W G,LI Y F,et al. Detection of false data injection attacks in smart grids based on expectation maximization [J]. Sensors,2023,23(3):1683.

[10] 黄崇鑫,洪明磊,伏帅,等. 考虑虚假数据注入攻击的有源配电网分布式状态估计[J]. 电力工程技术,2022,41(3):22-31.
HUANG Chongxin,HONG Minglei,FU Shuai,et al. Distributed state estimation of active distribution network considering false data injection attack[J]. Electric Power Engineering Technology,2022,41(3):22-31.

[11] KUMAR A,SAXENA N,JUNG S,et al. Improving detection of false data injection attacks using machine learning with feature selection and oversampling[J]. Energies,2021,15(1):212.

[12] YANG Y,HAO J A,ZHAO J G,et al. Computer user behavior anomaly detection based on K-means algorithm[J]. Security and Communication Networks,2022,2022:1-8.

[13] 田亚静. 安全 PMU 配置下基于 OPTICS-LOF 的虚假数据注入攻击检测与定位[D]. 秦皇岛:燕山大学,2020.
TIAN Yajing. Detection and location of false data injection attacks based on OPTICS-LOF in secure PMU configuration[D]. Qinhuangdao:Yanshan University,2020.

[14] 胡聪,洪德华,张翠翠,等. 一种基于特征映射与深度学习的虚假数据注入检测方法[J]. 现代电力,2023,40(1):125-132.
HU Cong,HONG Dehua,ZHANG Cuicui,et al. A method to detect false data injection based on feature mapping and deep learning[J]. Modern Electric Power,2023,40(1):125-132.

[15] 谢娟英,丁丽娟,王明钊. 基于谱聚类的无监督特征选择算法[J]. 软件学报,2020,31(4):1009-1024.
XIE Juanying,DING Lijuan,WANG Mingzhao. Spectral clustering based unsupervised feature selection algorithms [J]. Journal of Software,2020,31(4):1009-1024.

[16] XIAO S,BAI T,CUI X C,et al. A graph-based contrastive learning framework for medicare insurance fraud detection[J]. Frontiers of Computer Science,2023,17(2):1-3.

[17] RADHOUSH S,VANNOY T,LIYANAGE K,et al. Distribution system state estimation and false data injection attack detection with a multi-output deep neural network[J]. Energies,2023,16(5):2288.

[18] 谢云云,严欣腾,桑梓,等. 面向交直流混联系统的虚假数据注入攻击方法[J]. 电力工程技术,2022,41(1):165-172.
XIE Yunyun,YAN Xinteng,SANG Zi,et al. False data injection attack method against AC-DC hybrid systems[J]. Electric Power Engineering Technology,2022,41(1):165-172.

[19] NING H Y,QUAN D,ZHANG X R,et al. Unsupervised outlier detection using memory and contrastive learning [J]. IEEE Transactions on Image Processing,2022,31:6440-6454.

[20] SHENKAR T,WOLF L. Anomaly detection for tabular data with internal contrastive learning[C]//International Conference on Learning Representations. 2021.

[21] XIAO S,BAI T,CUI X C,et al. A graph-based contrastive learning framework for medicare insurance fraud detection[J]. Frontiers of Computer Science,2023,17(2):1-3.

[22] DASH C S K,BEHERA A K,DEHURI S,et al. An outliers detection and elimination framework in classification task of data mining[J]. Decision Analytics Journal,2023,6:100164.

[23] 杨玉泽,刘文霞,李承泽,等. 面向电力 SCADA 系统的 FDIA 检测方法综述[J/OL]. 中国电机工程学报:1-22[2023-03-31]. http://kns.cnki.net/kcms/detail/11.2107.tm.20221101.1539.003.html.
YANG Yuze,LIU Wenxia,LI Chengze,et al. Overview of FDIA detection methods for power SCADA systems [J/OL]. Proceedings of the CESS:1-22[2023-03-31]. http://kns.cnki.net/kcms/detail/11.2107.tm.20221101.1539.003.html.

[24] PAVITHRA J,SELVAKUMARASAMY S. An adaptive-feature centric XGBoost ensemble classifier model for improved malware detection and classification[J]. Journal of Cyber Security,2022,4(3):135-151.

[25] ZIMMERMAN R D,MURILLO-SÁNCHEZ C E,THOMAS R J. Matpower:steady-state operations,planning,and analysis tools for power systems research and education[J]. IEEE Transactions on Power Systems,2011,26(1):12-19.

[26] 陆孝锋,李鹏,高莲,等. 基于 DKPCA 的电力信息系统虚假数据注入攻击检测方法[J]. 电子测量技术,2022,45(2):91-97.

LU Xiaofeng,LI Peng,GAO Lian,et al. False data injection attack detection method based on dynamic kernel principal component analysis for power information system[J]. Electronic Measurement Technology,2022,45(2):91-97.

[27] ASHRAFUZZAMAN M,DAS S,CHAKHCHOUKH Y,et al. Detecting stealthy false data injection attacks in the smart grid using ensemble-based machine learning[J]. Computers & Security,2020,97:101994.

[28] 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.

[29] CHEN L A,GU S L,WANG Y,et al. Stacked autoencoder framework of false data injection attack detection in smart grid [J]. Mathematical Problems in Engineering,2021(30):1-8.

作者简介:


黄冬梅

黄冬梅(1964),女,硕士,教授,博士生导师,研究方向为海洋与电力时空信息技术(E-mail:dmhuang_dl@163.com);

王一帆(1998),男,硕士在读,研究方向为虚假数据注入攻击检测;

胡安铎(1983),男,博士,讲师,研究方向为电力负荷预测、虚假数据注入攻击检测。

Detection method of false data injection attack based on unsupervised and supervised learning

HUANG Dongmei[1], WANG Yifan[2], HU Anduo[1], ZHOU You[3], SHI Shuai[2], HU Wei[4]

(1. College of Electronics and Information Engineering,Shanghai University of Electric Power,Shanghai 201306,China;
2. College of Electric Power Engineering,Shanghai University of Electric Power,Shanghai 200090,China;
3. State Grid Suzhou Power Supply Company of Jiangsu Electric Power Co.,Ltd.,Suzhou 215004,China;
4. College of Economics and Management,Shanghai University of Electric Power,Shanghai 201399,China)

Abstract:False data injection attack (FDIA) is a serious threat to the security and stable operation of smart grids. In this paper,a FDIA detection algorithm that combines unsupervised and supervised learning is proposed,solving the problems of scarce labeled data and extremely imbalanced normal and attack samples. Firstly,contrastive learning is introduced to capture the features of a small amount of attack data,and it generates new attack samples to achieve data augmentation. Then,various unsupervised detection algorithms are used to perform feature self-learning on a large number of unlabeled samples,addressing the problem of scarce labeled samples. Finally,the features extracted by the unsupervised algorithm are fused with the historical feature set,and a supervised XGBoost classifier is constructed to identify and output the detection results. The results on the IEEE 30-node system show that the proposed method can enhance the stability of the FDIA detection model under scarce labeled samples and imbalanced data,compared with other FDIA detection algorithms. The proposed method can improve recognition accuracy and reduce false alarm rate.

Keywords:false data injection attack(FDIA);supervised learning;unsupervised learning;contrastive learning;data expansion;feature enhancement

(编辑 方晶)