王旭,何宇,袁梦薇. 基于 Stacking 和孤立森林的虚假数据注入攻击防御策略[J]. 智能计算机与应用,2024,14(7):222-226.
DOI:10.20169/j.issn.2095-2163.240735

基于 Stacking 和孤立森林的虚假数据注入攻击防御策略

王 旭, 何 宇, 袁梦薇
(贵州大学 电气工程学院, 贵阳 550025)

摘 要: 针对电力信息物理系统受到虚假数据注入攻击后无法安全稳定运行的问题,本文提出了一种基于 Stacking 和孤立森林的两阶段数据清洗方法。首先,由多异质学习器组成的 Stacking 分类模型对实时量测数据样本进行异常检测,判断当前时刻量测样本中是否存在虚假数据;其次,虚假数据的量测样本与基于负荷预测和潮流计算生成的当前时刻伪量测数据作差,得到量测误差向量,将量测误差向量输入孤立森林异常检测模型中进行二次辨识,定位受攻击的量测位置,并由伪量测数据进行替换修正;最后,通过 IEEE-33 节点测试系统仿真实验验证本文所提方法的有效性。

关键词: 虚假数据注入; 数据清洗; Stacking; 孤立森林
中图分类号: TM769 **文献标志码**: A **文章编号**: 2095-2163(2024)07-0222-05

False data injection attacks defense strategy based on stacking and isolation forest

WANG Xu, HE Yu, YUAN Mengwei
(School of Electrical Engineering, Guizhou University, Guiyang 550025, China)

Abstract: To address the problem of the power cyber physical system cannot operate safely and stably after being attacked by false data, a two-stage data cleansing method that combines Stacking and Isolation Forest is proposed. First, a Stacking classification model composed of multiple heteroge-neous learners performs anomaly detection on real-time measurement data samples, and determines whether there are false data in the current measurement samples. Then, the measurement samples with false data will be subtracted from the pseudo-measurement data at the current moment generated based on load forecasting and power flow calculation to obtain the measurement error vector. The measurement error vector will be input into the Isolation Forest anomaly detection model for secondary identification, locating the attacked measurement position and replacing it with pseudo-measurement data for correction. Finally, simulation experiments on IEEE-33 node test system verify the effectiveness of the proposed method.

Key words: false data injection; data cleansing; Stacking; Isolation Forest

0 引 言

随着电力系统不断向现代化、智能化推进,现代电力系统已经逐步发展成为信息系统和物理系统高度耦合的电力信息物理系统(Cyber Physical Systems, CPS)[1-3]。电力 CPS 依靠高效的数据传输,能够有效地发挥信息物理高度融合的优势,但过于频繁的数据通信也导致其更容易受到网络攻击的威胁。2015 年,乌克兰多个地区电网由于遭受黑客的攻击导致大规模停电,是电力 CPS 遭受外部攻击导致的系统崩溃[4]。网络安全问题已经成为电力 CPS 首要问题。虚假数据注入攻击(False Data Injection Attack, FDIA)作为网络攻击中威胁性较高的攻击之一,具有极高的隐蔽性,能够绕开现有能量管理系统对状态信息检测机制,篡改电网量测数据,造成调控中心对当前电力 CPS 状态的错误估计,从而下达错误指令,影响电网的安全稳定运行[5-7]。文献[8]提出一种 FDIA 方法能够绕过传统不良数据检测机制,直接影响状态估计,但该方法需要攻击者得知系统的全部拓扑结构;文献[9]放宽了攻击条件,攻击者仅需知道攻击区域内的相角差即可构建躲过不良数据检测机制的 FDIA。因此,FDIA 的

基金项目: 黔科合支撑[2022]一般 014。
作者简介: 王 旭(1996-),男,硕士研究生,主要研究方向:电力大数据应用; 袁梦薇(1997-),女,硕士研究生,主要研究方向:机器学习在电力系统中的应用。
通讯作者: 何 宇(1978-),女,硕士,副教授,主要研究方向:电力系统规划,电力系统稳定与运行。Email:yhe7@ gzu.edu.cn
收稿日期: 2023-05-15

检测与识别已经成为保障电力 CPS 安全稳定运行的必要条件。现有的 FDIA 检测方法主要分为基于模型驱动的状态估计检测方法和基于数据驱动的机器学习检测方法。状态估计 FDIA 检测方法一般基于电网的运行状态，通过残差搜索进行 FDIA 检测与辨识[10]。文献[11]通过自适应卡尔曼滤波算法得到状态估计结果，并结合中心极限定理提出检测算法；文献[12]通过在最小二乘估计法中引入额外的扩展卡尔曼滤波算法，计算二者残差来判断系统是否遭受攻击；文献[13]通过脉冲神经网络构建了伪量测模型，结合非线性滤波算法计算估计值偏差，辨识 FDIA；文献[14]针对新型能源互联网运行状态多变的特征，生成 2 个马尔可夫链模型，通过对比二者状态估计精确度来检测 FDIA。随着电力系统不断朝着智能化发展，大数据逐渐在电力 CPS 中应用，基于模型驱动的检测方法已无法应对不断增多的电网数据量，无法满足不断增大的电力系统在线应用需求[15]。基于数据驱动的 FDIA 检测方法不需要物理建模，直接利用大量历史数据拟合攻击数据的特征来进行 FDIA 检测。文献[16]利用递归神经网络学习数据间的时间相关性，从而进行 FDIA 检测；文献[17]采用 CNN-GRU 神经网络提取数据的时空特征进行 FDIA 检测；文献[18]结合随机边删减策略和 Deepwalk 算法将系统节点映射为低维向量，并采用机器学习方法对其进行分类以进行 FDIA 检测；文献[19]利用生成式对抗网络对量测数据进行重建，通过重建误差来定位不良数据；文献[20]运用基于 Spark 的 K 均值聚类算法提取电力系统的日负荷特征曲线，通过与特征曲线的对比结果来辨识和处理不良数据，解决了传统 K 均值算法聚类速度慢的问题。数据驱动的 FDIA 检测方法在保持精度的同时速度较快，有良好的应用前景。

上述方法虽能够实现假数据检测，但仅局限于检测量测数据中是否存在 FDIA，防御效果不佳，为此本文提出一种基于 Stacking 和孤立森林的数据驱动假数据注入防御策略，通过 Stacking 方法检测量测样本中是否存在 FDIA；通过孤立森林模型定位量测样本中受到 FDIA 的数据，最后将受到 FDIA 的数据用伪量测数据对应替换，达到精准定位并修正脏数据的目的。

1　相关模型原理

1.1　电力系统状态估计原理

电力系统状态估计的本质是通过迭代计算非线性方程，最小化量测值与状态值之间的误差。对于一个给定的 m 个量测值，n 个状态变量的电力 CPS，$\boldsymbol{z} \in \mathbb{R}^m$ 为实时量测向量，一般包含的量测量为节点电压幅值、节点注入功率和支路功率，公式(1)：

$$\boldsymbol{z} = h(\boldsymbol{x}) + \boldsymbol{\nu} \tag{1}$$

其中，$\boldsymbol{x} \in \mathbb{R}^n$ 为系统状态变量；$h(\cdot)$ 为非线性量测函数；$\boldsymbol{\nu} \in \mathbb{R}^m$ 为量测误差向量。

一般加权最小二乘(WLS)估计算法以加权残差平方和最小为目标，即式(2)：

$$\min J = [\boldsymbol{z} - h(\boldsymbol{x})]^{\mathrm{T}} \boldsymbol{R}^{-1} [\boldsymbol{z} - h(\boldsymbol{x})] \tag{2}$$

其中，$\boldsymbol{R}$ 为 $m \times m$ 阶量测值角协方差矩阵。

状态变量的最终估计值，公式(3)：

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

其中，$\boldsymbol{H}$ 为雅克比矩阵。

1.2　虚假数据注入攻击原理

通过精心构建攻击向量，攻击者能够使被攻击的量测数据满足电力系统物理定律，逃过不良数据辨识。受攻击的量测向量表示为式(4)：

$$\boldsymbol{z}' = \boldsymbol{z} + \boldsymbol{a} \tag{4}$$

其中，$\boldsymbol{a}$ 为攻击向量。

当系统中不存在不良数据时，量测残差 r 应满足 $\| \boldsymbol{r} \|_2 < \delta$，$\delta$ 为设定的残差阈值。若攻击向量 $\boldsymbol{a}$ 满足条件(5)，即可成功避开不良数据检测机制，公式(5)：

$$\| \boldsymbol{a} - h(\hat{x} + \boldsymbol{c}) + h(\hat{\boldsymbol{x}}) \|_2 \leqslant \tau_1 - \| \boldsymbol{r} \|_2 \tag{5}$$

1.3　Stacking 集成学习原理

集成学习是机器学习方法中一种提升模型精度的常用方法，通常分为并行集成方法和串行集成方法。Stacking 是并行集成方法的一种，一般为两层结构，第一层中使用的学习器称为基学习器，第二层中使用的学习器称为元学习器。

Stacking 使用异质模型作为基学习器，通过不同的基学习器对原始数据集进行不同数据空间角度和不同数据结构角度的观测，提取出不同的特征并输出。将每一个基学习器提取出的特征组合，称之为元特征。为了防止过拟合现象，元学习器不直接观测原始数据集，而是观测元特征，并输出最终的预测结果。

2　虚假数据注入攻击防御策略

2.1　基于 Stacking 的虚假数据检测

Stacking 方法能够结合各种不同类型机器学习模型的优点，提高模型的精度和泛化能力，本文利用 Stacking 分类模型检测电网量测数据中是否存在 FDIA，具体检测方法如图 1 所示。

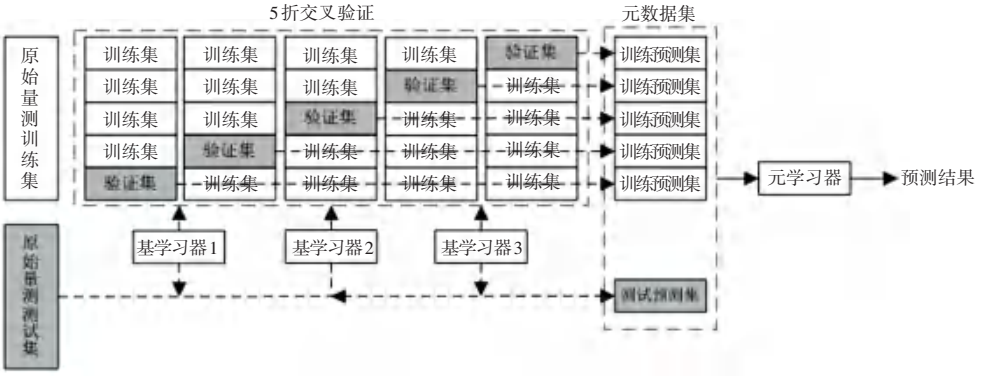

图 1 基于 Stacking 的 FDIA 检测方法

Fig. 1 Stacking based FDIA detection method

2.2 基于孤立森林的虚假数据定位与修正

2.2.1 基于 Stacking 的伪量测生成

当检测出量测数据中存在 FDIA 时,可使用伪量测数据对受到 FDIA 的数据进行对应替换。本文使用 Stacking 预测模型进行负荷预测,使用前 $t-4$ 个时刻的负荷值来预测 t 时刻的负荷值,并基于 t 时刻的负荷预测结果进行潮流计算,得到 t 时刻的伪量测数据。

2.2.2 基于孤立森林的虚假数据定位

孤立森林异常检测算法是一种基于统计计算的异常检测算法,其基于随机森林算法,通过计算每个样本点的异常分数来检测异常点,孤立森林的异常检测分为两个阶段,第一阶段即训练阶段使用训练集的样本构建隔离树,第二阶段即测试阶段将测试样本在隔离树中层层传递,获取最终的异常分数。

本文中孤立森林模型的输入数据集为真实量测与伪量测的误差,即式(6):

$$\boldsymbol{\zeta}_k = |\boldsymbol{m}_k - \hat{\boldsymbol{m}}_k| \tag{6}$$

其中,$\boldsymbol{\zeta}_k$、$\boldsymbol{m}_k$ 和 $\hat{\boldsymbol{m}}_k$ 分别为第 k 个时间断面的量测误差向量、实时量测向量和伪量测向量。

孤立森林模型通过学习量测误差向量中每个数据,计算其异常分数来定位虚假数据,避免了经验法设置阈值带来的高误检率和漏检率。

2.3 电网虚假数据注入攻击防御策略

在用大量历史量测数据训练所有模型后,即可组合模型进行虚假数据注入攻击防御工作,具体流程如图 2 所示。

3 仿真分析

为验证本文所提方法的有效性,本文基于 IEEE-33 节点系统进行仿真实验。首先通过直流潮流计算生成各个时间断面的潮流真实值,负荷数据来自美国纽约电力管理局统计的纽约地区 2021 年用电负荷,攻击向量参照文献[9]设置,确保其能够躲过状态估计中的不良数据检测机制,具体过程如下:

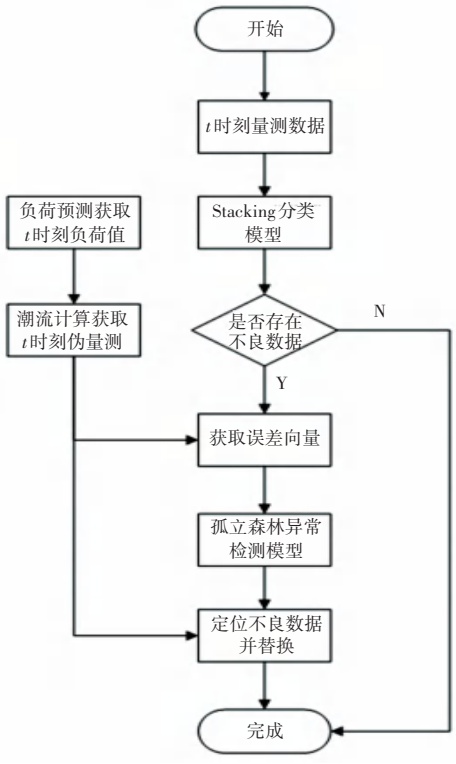

图 2 电网 FDIA 防御策略流程

Fig. 2 Power grid FDIA defense strategy process

首先,选定攻击区域,根据其当前时刻系统状态变量 $\boldsymbol{x} = [\theta, V]^{\mathrm{T}}$ 计算攻击向量 $[p, q, P, Q]^{\mathrm{T}}$,并确保其满足功率上下界约束;

其次,通过求解最优化问题得到状态变量增量 $\Delta\boldsymbol{x} = [\Delta\theta, \Delta V]^{\mathrm{T}}$;

最后,更新状态变量 $\boldsymbol{x}' = \boldsymbol{x} + \Delta\boldsymbol{x}$。

通过对比本文模型与其它对比模型在数据集上的评价指标验证本文方法的有效性。

3.1　评价指标

本文使用混淆矩阵(Confusion Matrix)及其衍生出来的准确率(*Accuracy*)、查准率(*Precision*)、查全率(*Recall*)、*F*1 分数(*F*1 - score)、*ROC* 曲线下面积(Area Under ROC Curve，*ROC - AUC*)对模型的性能进行评价，同时采用成对多样性度量指标 *Q* 统计值来衡量学习器之间的多样性。

3.2　基学习器和元学习器的选择

3.2.1　基学习器性能测试

在进行 Stacking 集成学习时，首先需根据“好而不同”原则选择模型的基学习器。本文中初步考虑基学习器为 Logistic 回归(LR)、K 最邻近算法(KNN)、支持向量机(SVM)、随机森林(RF)、极端随机树(ET)、极端梯度提升树(XGBoost，XGB)和光梯度提升机(LightGBM，LGB)。分别在数据集中进行测试，通过评价指标对比分析，充分考虑准确度和差异化要求，确定最终选取的初级学习器，各模型评价指标见表 1。

表 1　各模型评价指标

Table 1　Evaluation indicators of each model

模型	准确率	*F*1 分数	*ROC - AUC*
LR	0.620 2	0.593 5	0.620 3
KNN	0.845 7	0.817 6	0.845 8
SVM	0.575 3	0.286 0	0.575 5
RF	0.997 9	0.997 9	0.997 9
ET	0.999 2	0.999 2	0.999 2
XGB	0.976 4	0.975 9	0.976 4
LGB	0.994 9	0.994 8	0.994 9

由表 1 可知，7 个模型中 SVM 表现最差，其准确率与 *ROC* 曲线下面积均低于 0.6，*F*1 分数仅为 0.28，说明 SVM 模型稳定度低，泛化能力差，无法准确识别 FDIA。LR 在剩余模型中的表现最不佳，其 3 个评价指标均处于 0.6 左右。SVM 与 LR 的 *ROC* 曲线下面积接近 0.5，说明二者易将正常数据判断为虚假数据，同时也容易把虚假数据判断为正常数据，会极大地影响电网正常运行，首先考虑将 SVM 与 LR 从基学习器中排除。

差异化的学习器从训练集中观测到的特征也是差异化的，这会让元学习器有更大的改进空间，避免元学习器产生过拟合现象。剩余 5 个学习器的 *Q* 值对比如图 3 所示。

图 3 中，KNN 与其他学习器的 *Q* 值均最小。由于面向二分类问题，故效果较好的学习器之间的 *Q* 值一般较大，可以看到 RF、ET、XGB 和 LGB 之间的 *Q* 值均超过 0.9。考虑到 RF 和 ET 是基于装袋法集成决策树的改进算法，XGB 和 LGB 是基于提升法集成决策树的改进算法，基于学习器的算法多样性、分类性能以及训练时间综合评估，选择 ET 和 LGB 作为模型的基学习器。综上所述，本文选择的基学习器为 KNN、ET 和 LGB。

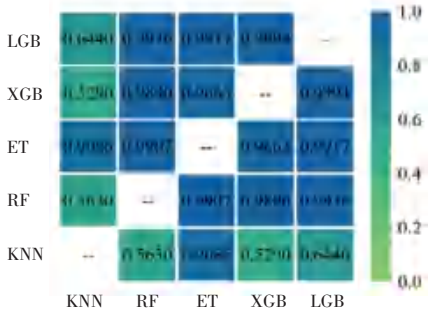

图 3　多样性度量指标 *Q* 值对比

Fig. 3　Comparison of diversity metric *Q* value

3.2.2　元学习器性能测试

在 Stacking 集成学习中，基学习器面对的是较高维的原始数据集，而元学习器则面对基学习器输出的较低维的元数据集，故一般考虑选择较简单的学习器作为元学习器，来拟合各基学习器的输出。本文选取了 LR 和决策树(DT)作为候选模型，通过对比二者性能确定元学习器。

分别将 LR 和 DT 作为元学习器，已选取的 3 个基学习器通过 Stacking 集成学习策略组合，在数据集上训练 10 次，其平均精度、*F*1 分数、*ROC - AUC* 值和训练时间结果见表 2。从表 2 中可见，经过 Stacking 集成后，LR 和 DT 的 3 个评价指标均高于 0.98，充分说明 Stacking 集成策略的有效性，而 LR 在评价指标和训练耗时上均优于 DT，故选取 LR 作为元学习器。

表 2　元学习器评价指标对比

Table 2　Evaluation indicators of each model

模型	准确率	*F*1 分数	*ROC - AUC*	耗时/s
LR	0.999 4	0.999 3	0.999 4	24.147 2
DT	0.985 6	0.985 7	0.985 7	28.396 3

3.3　伪量测数据生成

本文建立了一个用于负荷预测的 Stacking 回归模型，与 Stacking 分类模型不同，Stacking 回归模型中所采用的各种学习器均为经过实验验证的高精度回归学习器。通过综合对比分析各种基学习器与元学习器组合的预测精度，选取了 XGB、LGB 和梯度提升类别型特征(CatBoost)作为基学习器，LGB 作为元学习器来完成负荷预测任务。负荷预测值与真实值对比如图 4 所示。

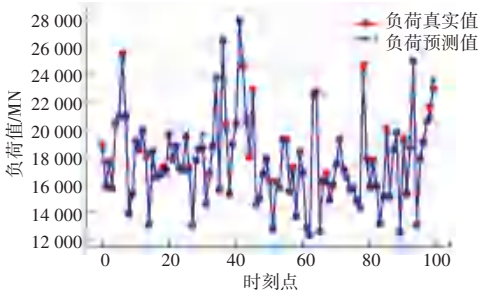

图 4 负荷预测值与真实值对比

Fig. 4 Comparison between predicted and actual load

图 4 中截取了 100 个截面对预测负荷与真实负荷作对比,可以看出经过 Stacking 集成学习的负荷预测模型很好地拟合了真实负荷走势曲线,仅在一些峰谷值上有一些误差。

将本时刻的预测负荷值输入仿真软件中进行潮流计算,得到本时刻的计算量测数据,作为本时刻的伪量测数据。

3.4 虚假数据注入攻击防御效果分析

量测数据经过 Stacking 模型的初级辨识后,输入孤立森林模型中进行二次辨识,完成虚假数据的精准定位,将伪量测数据与对应的虚假数据进行替换,完成数据清洗任务。

本文选取基于“3σ”准则的异常数据鉴别方法、基于中心极限定理的异常数据鉴别方法进行对比,验证本文所提方法的有效性,对比结果见表 3。

表 3 各方案系统运行成本对比

Table 3 Comparison of system operation costs by option

方法	准确率/%	误报率/%
3σ 准则法	69.50	0.54
中心极限定理	96.07	2.51
本文方法	99.40	0.32

根据表 3 可知,3σ 准则法的精度最低,这是由于除了虚假数据不满足 3σ 准则以外,电力系统运行时的偶然波动会导致数据产生较大的变动,这些数据虽然波动幅度大,但仍属于正常数据;中心极限定理辨识法虽然准确度在 96%以上,但误报率却远大于其余两个方法;本文所提方法表现最为优秀,辨识准确率在 99.4%以上的同时误报率均低于 0.4%。

4 结束语

本文提出一种基于 Stacking 和孤立森林的虚假数据注入攻击防御策略,利用历史量测数据选择合适的机器学习算法作为 Stacking 模型的基学习器和元学习器,结合孤立森林模型自适应辨识虚假数据,最后利用伪量测数据对应替换虚假数据完成数据清洗。在 IEEE-33 节点系统中进行了验证,结果表明本文所提策略准确率高、误报率低,具有一定的实用价值。

参考文献

[1] 陈清清, 苏盛, 畅广辉, 等. 电力信息物理系统内部威胁研究综述[J]. 南方电网技术, 2022,16(6): 1-13.

[2] 遥宝中, 李庚银, 王剑晓, 等. 计及监测与控制功能的电力信息物理系统关键输电线路辨识方法[J]. 中国电机工程学报, 2022,42(7): 2556-2566.

[3] 张艺伟, 刘文霞, 刘耕铭, 等. 考虑拓扑相关和双重耦合的电力信息物理系统建模与脆弱性分析[J]. 中国电机工程学报, 2021,41(16): 5486-5500.

[4] 郭庆来, 辛蜀骏, 王剑辉, 等. 由乌克兰停电事件看信息能源系统综合安全评估[J]. 电力系统自动化, 2016,40(5): 145-147.

[5] WANG J, HUI L C K, YIU S M, et al. A survey on cyber attacks against nonlinear state estimation in power systems of ubiquitous cities[J]. Pervasive and Mobile Computing, 2017,39: 52-64.

[6] CHAOJUN G, JIRUTITIJAROEN P, MOTANI M. Detecting false data injection attacks in ac state estimation[J]. IEEE Transactions on Smart Grid, 2015,6(5): 2476-2483.

[7] 王琦,郜伟,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.

[8] LIU Y, NING P, REITER M K. False data injection attacks against state estimation in electric power grids [J]. Acm Transactions on Information and System Security, 2011,14(1): 309-341.

[9] LIU X, LI Z. False data attacks against ac state estimation with incomplete network information[J]. IEEE Transactions on Smart Grid, 2017,8(5): 2239-2248.

[10] 于尔铿. 电力系统状态估计中不良数据的零残差辨识法[J]. 电网技术, 1981(2): 66-73.

[11] 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]. 自动化学报, 2022,48(12): 2960-2971.

[12] 王竞才,李琰,徐天奇. 基于扩展卡尔曼滤波的智能电网虚假数据检测[J]. 智慧电力,2022,50(3):50-56.

[13] 陈碧云,李弘斌,李滨. 伪量测建模与 AUKF 在配电网虚假数据注入攻击辨识中的应用[J]. 电网技术,2019,43(9):3226-3236.

[14] 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.

[15] 张东霞,苗新,刘丽平,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-12.

[16] 邱荣福, 谢型浪, 谢虎, 等. 基于递归神经网络的智能电网虚假数据检测[J]. 自动化技术与应用, 2021,40(1): 105-109.

[17] 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.

[18] 连祥龙,钱瞳,张银,等. 基于 DeepWalk 算法的电力系统错误数据注入网络攻击分类方法[J]. 电力自动化设备,2023,43(3): 166-171.

[19] 臧海祥,郭镜玮,黄蔓云,等. 基于改进 Wasserstein 生成式对抗网络的电力系统不良数据辨识[J]. 电力自动化设备,2022,42(9):50-56,110.

[20] 孟建良,刘德超. 一种基于 Spark 和聚类分析的辨识电力系统不良数据新方法[J]. 电力系统保护与控制,2016,44(3):85-91.