

基于图注意力与多尺度并行融合卷积的虚假数据注入攻击定位检测

席磊^{1,2}, 陈采玉¹, 陈洪军¹, 李宗泽¹

(1. 三峡大学电气与新能源学院, 宜昌 443002;

2. 梯级水电站运行与控制湖北省重点实验室(三峡大学), 宜昌 443002)

摘要: 虚假数据注入攻击严重威胁电力信息物理系统的安全, 而传统攻击检测方法由于没有考虑量测数据间的拓扑并且特征提取能力差, 无法精确识别攻击并定位受攻击节点。因此, 该文提出一种基于图注意力与多尺度并行融合卷积模型的虚假数据注入攻击定位检测方法。该方法通过图注意力网络动态捕捉量测数据间的拓扑关系以提升检测方法的定位检测性能; 采用结合注意力特征融合模块增强的并行卷积神经网络提取数据的多尺度特征进一步提高检测方法的学习能力和泛化能力, 以实现高精度的定位检测。通过在 IEEE-14 节点测试系统和 IEEE-57 节点测试系统中进行评估研究, 与现有的定位检测方法相比, 该文所提方法具有更优的 F1 值, 分别高达 98.40%、95.29%。因此, 该方法能够更好地对虚假数据注入攻击进行定位检测。

关键词: 虚假数据注入攻击; 电力信息物理系统; 图注意力网络; 并行卷积; 特征融合

Localization Detection for False Data Injection Attacks Based on Graph Attention and Multi-scale Parallel Fusion Convolution

XI Lei^{1,2}, CHEN Caiyu¹, CHEN Hongjun¹, LI Zongze¹

(1. College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China;

2. Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China)

Abstract: False data injection attacks pose a significant threat to the security of cyber-physical power system. Traditional attack detection methods fail to accurately identify and localize attacked nodes due to their lack of consideration for the topological relationships among measurement data and their inadequate feature extraction capabilities. Therefore, this paper introduces a novel detection method for locating false data injection attacks, based on a graph attention network and a multi-scale parallel fusion convolutional model. This method dynamically captures the topological relationships among measurement data through the graph attention network, enhancing the localization performance of the detection technique. It utilizes a parallel convolutional neural network, augmented with an attention feature fusion module, to extract multi-scale features, thereby improving the learning and generalization capabilities of the detection method to achieve high-precision localization. Evaluation studies conducted on the IEEE-14 and IEEE-57 bus test systems demonstrate that the proposed method precedes existing localization techniques, achieving F1 scores of 98.40% and 95.29%, respectively. Consequently, this method provides a more effective solution for the localization detection of false data injection attack.

Key words: false data injection attack; cyber-physical power system; graph attention network; parallel convolution; feature fusion

0 引言

随着“双碳”目标的推进, 互联网技术和能源技术的融合程度不断加深, 传统电力系统正逐步转

变为信息物理高度耦合的电力信息物理系统(cyber-physical power system, CPPS)^[1-3]。CPPS 通过将先进的传感器、量测设备和通信设备集成到电力系统中, 实现了系统的智能化运行, 但这种集成也使其更容易受到各种网络攻击。虚假数据注入攻击(false data inject attack, FDIA)^[4-6]作为主要网络攻击

基金资助项目: 国家自然科学基金(52277108)。

Project supported by National Natural Science Foundation of China (52277108).

之一, 具有极高的隐蔽性和破坏性。FDIA 可以利用所掌握的电力系统参数和网络结构^[7], 恶意篡改测量仪表的量测值, 而不会被传统的不良数据检测机制(bad data detection, BDD)^[8]发现, 从而造成错误的状态估计, 严重影响 CPPS 的安全和稳定。因此, 设计一种有效的 FDIA 定位检测方法对电网安全运行具有重要意义。

目前, 针对 FDIA 检测方法^[9]的研究主要分为两类: 基于模型驱动^[10-12]检测方法和数据驱动^[13-15]检测方法。前者需要根据电力系统模型和参数, 利用连续时间段内的数据构建检测方法。但随着电力系统规模不断扩大、新能源并入和电力电子设备大量部署, 过于依赖模型参数的模型驱动方法的构建将愈发困难。而数据驱动方法不仅避免了繁琐的系统模型构建, 且检测方法适应性强, 能充分利用各种量测数据。文献[13]通过 K-means 过采样技术平衡量测数据, 并采用改进的级联机器学习模型, 增强了小样本环境下的 FDIA 检测性能。文献[14]利用图边缘条件卷积网络, 有效捕捉了量测数据之间的空间关联性以进行检测。此外, 考虑到量测数据具有一定的时间相关性, 文献[15]采用长短期记忆网络和门控循环单元提取数据的时序关系, 以识别潜在的攻击。然而, 上述基于数据驱动的检测方法虽然能够实现 FDIA 的检测, 但不能实现对 FDIA 的精确定位, 导致电网运营人员无法采用隔离被攻击的节点并重新进行调度等预防措施减少损失。

文献[16]为电力系统的每个节点都设计一个基于极限学习机的检测模型以进行攻击定位。文献[17]使用多层卷积神经网络捕获电网量测量的空间相关性并结合 BDD 来精确定位 FDIA。文献[18]则通过利用电力系统的拓扑特征来构造递归图神经网络以有效地对攻击进行检测和定位。

然而, 上述方法在面对电力系统结构变化和复杂性时难以有效利用节点之间的相关性, 也不能适应电力系统拓扑的动态变化。当系统的拓扑结构发生变化时, 现有检测方法的定位准确性和泛化能力均会受到严重影响。因此, 本文拟采用能够适应系统拓扑动态变化的图注意力网络(graph attention network, GAT)^[19]作为 FDIA 检测方法。GAT 在公共交通领域被首次提出, 与传统利用电网邻接矩阵进行静态拓扑关系建模的方法不同, 其利用注意力机制来自适应学习相邻节点的权值进而提取量测数据间的空间相关性, 且该方法成功捕捉了历史交通

数据的拓扑相关性构建出动态交通图。为此, 本文以期 GAT 可从电力系统量测数据中自动学习节点间的拓扑关系, 进而实现精确的 FDIA 定位检测。

而笔者长期探索发现, GAT 由于灵活的注意力机制会引入过多的参数, 从而导致其学习能力差及泛化能力一般, 因此 GAT 在大型电网中进行 FDIA 定位检测时难以提供准确和可靠的结果。为了增强检测方法的学习能力及泛化能力, 引入并行卷积神经网络(parallel convolutional neural network, PCNN)^[20]作为 GAT 的补充。PCNN 可以通过多尺度卷积广泛地提取数据中的全局和局部特征, 增强检测方法的学习能力和泛化性能。但提取的多尺度特征缺乏有效的整合, 使得检测方法难以充分利用特征之间的潜在联系。文献[21]引入注意力特征融合模块(attentional feature fusion, AFF)^[22]来融合深层与浅层特征, 优化了复杂背景下铁路受电弓图像的语义分割性能。因此, 本文使用 AFF 来动态融合 PCNN 提取的多尺度特征^[23], 进而得到的多尺度并行融合卷积网络(multi-scale parallel fusion convolution network, MPFCNN), 能够有效捕获并融合多尺度特征, 提升了检测方法的特征表达能力。

因此, 本文提出一种基于图注意力网络与多尺度并行融合卷积网络(GAT-MPFCNN)的 FDIA 定位检测方法。在利用 GAT 动态提取量测数据拓扑关系的基础上, 通过 MPFCNN 对数据进行多尺度特征提取和动态融合, 以提高检测方法的泛化能力和学习能力, 进而提升定位检测性能。本文在 IEEE-14 和 57 节点测试系统中进行仿真实验。实验结果验证了所提模型的有效性。

1 FDIA 相关问题描述

1.1 状态估计和不良数据检测机制

状态估计(state estimation, SE)是能量管理系统的重要组成部分^[24], 其能通过量测数据和电力系统模型进行电力系统状态估计, 从而得到可信的系统状态。而 SE 后得到的系统状态值通常会被用于经济调度、安全约束的最优潮流等方面^[25]。在交流系统中, 状态估计的量测方程如下所示:

$$z = h(x) + e \quad (1)$$

式中: $z \in \mathbf{R}^m$ 为量测向量, m 为量测量数目; $x = [\theta^T, v^T]^T \in \mathbf{R}^n$ 为系统的状态向量, θ 为节点电压相角, v 为节点电压幅值; $n=2N$, N 为网络中节点的总数目; $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 表示状态量与量测量之间的

非线性映射关系; $\mathbf{e} = [e_1, e_2, \dots, e_m] \in \mathbf{R}^m$ 为服从高斯分布的随机测量误差, 即 $\mathbf{e} \sim \mathcal{N}(0, \sigma^2)$, σ 为量测量的标准差。

状态估计通常采用加权最小二乘法求解最优的状态估计值 $\hat{\mathbf{x}}$, 求解方程如下所示:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} [\mathbf{z} - h(\mathbf{x})]^T \mathbf{W} [\mathbf{z} - h(\mathbf{x})] \quad (2)$$

式中: \mathbf{W} 为权重对角矩阵, 其对角元素 $W_i = \sigma_i^{-2}$ 为对应量测量 i 的权重系数。

由于量测与传输过程中存在随机干扰、偶然故障, 量测数据中会不可避免地出现不良数据。这些不良数据不仅会影响状态估计的收敛性能和估计精度, 而且可能会使其不收敛。因此需要使用 BDD 将不良数据剔除。而基于残差的 BDD 检测机制定义残差 $\mathbf{r} = \mathbf{z} - h(\hat{\mathbf{x}})$, 并使用 \mathbf{r} 的 L2 范数和预定义的阈值 τ 进行比较, 如果 $\|\mathbf{z} - h(\hat{\mathbf{x}})\|_2 > \tau$, 则存在不良量测数据, 否则, 则认为数据正常。

1.2 虚假数据注入攻击原理

攻击者通过攻击量测单元、通信网络来向 CPPS 注入虚假数据, 并躲过 BDD 从而造成错误的状态估计, 使得控制中心做出错误决策, 并污染云端数据库。针对 CPPS 的 FDIA 攻击示意图如图 1 所示。

假设 FDIA 攻击向量 $\mathbf{a} \in \mathbf{R}^{m \times 1}$, 攻击后的虚假量测向量变为 $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$ 。根据虚假量测向量进行状态估计得到的系统状态值则变为: $\hat{\mathbf{x}}_a = \hat{\mathbf{x}} + \mathbf{c}$ 。其中, $\mathbf{c} \in \mathbf{R}^{m \times 1}$ 表示攻击后系统状态变量的偏差值。则攻击向量 \mathbf{a} 的构建如下所示:

$$\mathbf{a} = h(\hat{\mathbf{x}} + \mathbf{c}) - h(\hat{\mathbf{x}}) \quad (3)$$

则注入攻击后的系统残差为:

$$\begin{aligned} \|\mathbf{r}_a\|_2 &= \|\mathbf{z}_a - h(\hat{\mathbf{x}}_a)\|_2 = \|\mathbf{z} + \mathbf{a} - h(\hat{\mathbf{x}} + \mathbf{c})\|_2 = \\ &= \|\mathbf{z} + h(\hat{\mathbf{x}} + \mathbf{c}) - h(\hat{\mathbf{x}}) - h(\hat{\mathbf{x}} + \mathbf{c})\|_2 = \|\mathbf{r}\|_2 \end{aligned} \quad (4)$$

由式(4)可知, 由于注入攻击前后系统的残差不会发生改变, 传统的基于残差的恶意数据检测方法已经失效, BDD 不再可靠。

2 GAT-MPFCNN

由于 CPPS 的底层物理架构是由发电机、变压器等电气元件经过线路相互连接构成, 故与底层物理架构高度耦合的信息层面量测数据具有特定的空间分布特征。因此, 利用 GAT 提取量测数据的拓扑关系, 可以充分利用量测数据的空间相关性。同时, 为提高检测方法的学习能力和泛化能力, 本文设计了一种多尺度并行融合卷积网络 MPFCNN 来对量

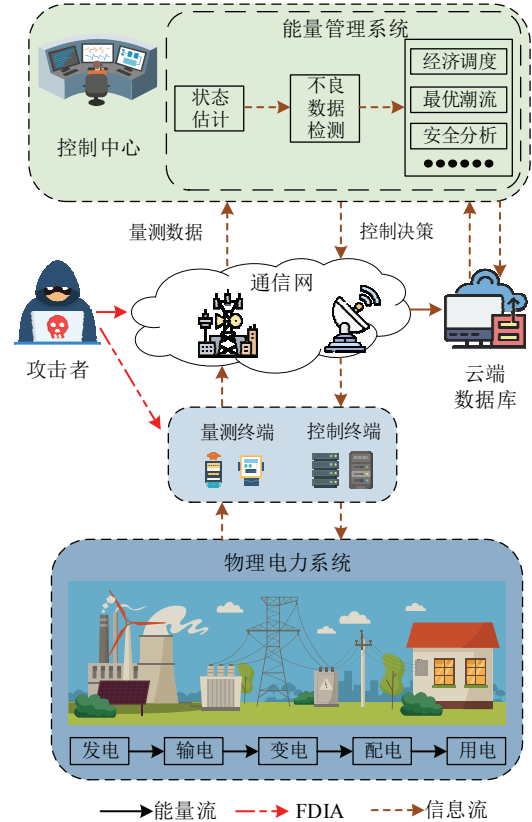


图 1 注入 FDIA 的 CPPS 示意图

Fig.1 Schematic diagram of CPPS injected into FDIA

测数据的多尺度特征进行特征提取及动态融合。

2.1 GAT

GAT 作为一种新颖的图神经网络, 其能利用自注意力机制动态地学习节点之间的权重, 使模型能够根据节点特征和其邻居的特征自适应地调整信息的聚合方式, 从而有效捕捉拓扑结构中的节点关系。通过这种方式, GAT 能够在不同的数据上, 自动学习到节点之间的重要性^[26], 进而提升模型性能。

若输入模型的数据 x_i 有 D 维, 则构造的图有 D 个节点, 即 $\{h_1, h_2, \dots, h_k\}$ 。相邻节点 j 对节点 i 的注意力 $e(h_i, h_j)$ 和归一化后的注意力 α_{ij} 计算式如下:

$$e(h_i, h_j) = \text{LeakyReLU}(\mathbf{u}^T [\mathbf{Q}h_i \parallel \mathbf{Q}h_j]) \quad (5)$$

$$\alpha_{ij} = \sigma_j(e(h_i, h_j)) = \frac{\exp(e(h_i, h_j))}{\sum_{j \in \Omega_i} \exp(e(h_i, h_j))} \quad (6)$$

式中: \parallel 表示节点特征拼接; $\mathbf{u} \in \mathbf{R}^{2 \times F}$ 与 $\mathbf{Q} \in \mathbf{R}^{F \times F}$ 为可学习的权重向量; LeakyReLU 为非线性激活函数; $j, j' \in \Omega_i$, Ω_i 为与节点 i 相连接的邻域节点集合; σ 为 sigmoid 激活函数。各节点的输出为:

$$h_{i, \text{out}} = \sigma \left(\sum_{j \in \Omega_i} \alpha_{ij} \mathbf{Q}h_j \right) \quad (7)$$

式中: $h_{i,\text{out}}$ 为节点 i 的输出, 与输入 h_i 维数相同; 为使 GAT 学习过程更稳定, 本文采用多层图注意力层来实现对量测数据空间特征的有效提取, 并将来自不同层的信息进行平均来整合, 最后将 D 维特征输出。而本文所使用的注意力层数为 3, 经过多层图注意力层后得到输出的特征为 Z_3 。GAT 结构示意图如图 2 所示。

2.2 MPFCNN

MPFCNN 主要由 PCNN 和 AFF 两部分组成。

其中, PCNN 利用并行卷积充分提取量测数据中的多尺度特征; 在此基础上, AFF 将 PCNN 提取的多尺度特征进行动态融合。

2.2.1 PCNN

为增强检测方法的学习能力及泛化能力, 引入 PCNN 来提取量测数据的多尺度特征。

经过图注意力提取空间相关性后, 输入数据由 3 个不同大小的卷积核的并行卷积进一步提取特征相关性。给定输入特征向量 $\mathbf{x}_{\text{in}} \in \mathbf{R}^{N_{\text{in}} \times 1 \times B_{\text{in}}}$, N_{in} 为一维输入向量的长度, 1 为通道数, B_{in} 为特征向量的数目。经过并行卷积后, 输出为 C 通道的特征, 则并行卷积的大、中、小卷积核分别为 $(\mathbf{k}_S^{l_S \times C, B_{\text{in}}}, \mathbf{b}_S)$ 、 $(\mathbf{k}_M^{l_M \times C, B_{\text{in}}}, \mathbf{b}_M)$ 、 $(\mathbf{k}_L^{l_L \times C, B_{\text{in}}}, \mathbf{b}_L)$, 其中 \mathbf{k} 表示卷积核; l_S 、 l_M 、 l_L 分别为一维卷积的卷积核长度; \mathbf{b}_L 、 \mathbf{b}_M 、 \mathbf{b}_S 分别为大、中、小卷积对应的偏置。则卷积层的输出如下所示:

$$[\mathbf{x}_S, \mathbf{x}_M, \mathbf{x}_L] = [\mathbf{k}_S^{l_S \times C, B_{\text{in}}}, \mathbf{k}_M^{l_M \times C, B_{\text{in}}}, \mathbf{k}_L^{l_L \times C, B_{\text{in}}}] \cdot \mathbf{x}_{\text{in}} + [\mathbf{b}_S, \mathbf{b}_M, \mathbf{b}_L] \quad (8)$$

批量归一化层(batch normalization, BN)^[27]通过归一化, 将每层的输入向量转化为正态分布, 从而加速训练过程并提高模型的泛化能力。输入 BN 层的批量数据 $\mathbf{x}_B = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_f\}$, 其中 f 为批量数据的数量, \mathbf{x}_i 是批量数据中的第 i 个特征向量, 则转化过程如下所示:

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (9)$$

式中: ε 为一个非常小的常数; 均值 μ_B 、方差 σ_B^2 的计算式如下:

$$\mu_B = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i \quad (10)$$

$$\sigma_B^2 = \frac{1}{k} \sum_{i=1}^k (\mathbf{x}_i - \mu_B)^2 \quad (11)$$

则 BN 层输出为:

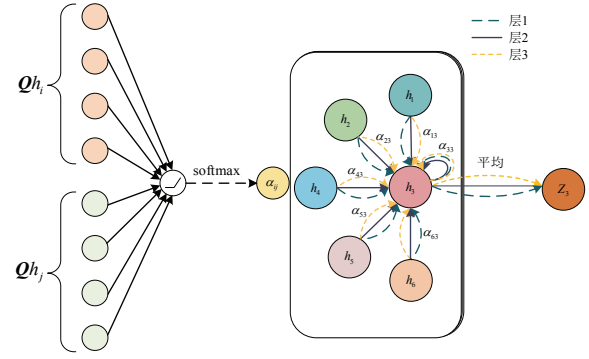


图2 GAT结构示意图

Fig.2 GAT structure diagram

$$\mathbf{y}_B \leftarrow \mathbf{y}_i = \gamma_i \hat{\mathbf{x}}_i + \beta_i \equiv \text{BN}_{\gamma_i, \beta_i}(\mathbf{x}_B) \quad (12)$$

式中: \mathbf{y}_B 为批量的归一化输出特征向量 \mathbf{y}_i ; γ_i 、 β_i 为可学习的参数, γ_i 为缩放因子, 用于调整规范化后数据的尺度; β_i 为平移因子, 用于调整规范化后的数据, 使其均值可以被适当地调整。

经过 BN 层归一化后的特征经过 ReLU 激活。最后, 为了进一步提高 PCNN 的泛化能力, 本文在并行卷积旁引入 1 条残差连接线来连接原始特征 \mathbf{x}_h 和卷积提取的特征 \mathbf{x}_L 、 \mathbf{x}_M 、 \mathbf{x}_S 。

2.2.2 AFF

在处理多尺度特征时, 为了提高检测方法对特征间互补信息的利用, 本文采用 AFF 增强 PCNN, 从而实现对不同尺度特征动态融合。

经过并行卷积提取多尺度特征后, 输入特征为 \mathbf{x}_h 、 \mathbf{x}_L 、 \mathbf{x}_M 、 $\mathbf{x}_S \in \mathbf{R}^{C \times 1 \times K}$, 其中, C 为通道数, K 为特征的长度。特征融合过程如图 3 所示。则基于注意力机制的特征融合计算式如下所示:

$$\mathbf{x}_f = (\mathbf{x}_h \oplus \mathbf{x}_L) \otimes w + (1 - w) \otimes (\mathbf{x}_S \oplus \mathbf{x}_M) \quad (13)$$

式中: \oplus 表示加法运算; \otimes 表示点乘运算; w 为随中间特征变化的权重。 w 的计算方法如下式所示:

$$w(A) = s(G(A) \oplus P(A)) \quad (14)$$

式中: A 为中间特征, 由 \mathbf{x}_h 、 \mathbf{x}_L 、 \mathbf{x}_M 、 \mathbf{x}_S 相加得到; $G(A)$ 为全局特征; $P(A)$ 为局部特征。 $G(A)$ 和 $P(A)$ 的计算式如下所示:

$$G(A) = b(H_2(r(b(H_1(g(A)))))) \quad (15)$$

$$P(A) = b(H_2(r(b(H_1(A)))))) \quad (16)$$

式中: H_1 和 H_2 表示逐点卷积, 卷积核大小分别为 $\frac{C}{q} \times C \times 1 \times 1$ 、 $C \times \frac{C}{q} \times 1 \times 1$, q 为通道缩减比; b 表示 BN 层; r 表示 ReLU 激活函数; g 表示全局平均池化。 g 的计算式如下所示:

$$g(A) = \frac{1}{K} \sum_{i=1}^K A_i \quad (17)$$

式中: A_i 表示中间特征 A 中的元素。

2.3 GAT-MPFCNN 整体框架

本文所提检测方法主要分为 3 部分: 1) 采用图注意力网络进行特征拓扑关系的提取; 2) 利用多尺度并行融合卷积网络进行多尺度特征提取和融合, 然后进行平均池化; 3) 利用两层全连接层进行分类。模型的输入为数据采集与监控(supervisory control and data acquisition, SCADA)和相量测量单元(phasor measurement unit, PMU)混合量测数据, 而输出与各节点的电压幅值和相角状态量相关联。模型的每个输出标签对应一个状态量, 且每个标签用二元数值 0 或 1 表示, 其中 0 意味着该状态量为正常状态, 1 则表示状态量遭受攻击。图 4 给出检测方法的整体框架。

检测采样时刻 t 电力系统是否受到 FDIA 的流程如下: t 时刻采集的量测数据送入控制中心, 而后传统的 BDD 会剔除部分坏数据。通过 BDD 后的量测数据经过数据预处理并输入本文所提出的检测方法进行攻击定位检测。其中, 定位检测模型输出结果为长度为 $2N$ 的二进制向量。若输出结果全为 0, 则表示为正常量测数据输入云端数据库, 若输出结果存在 1, 则进行 FDIA 定位并预警。

3 仿真研究

3.1 数据集

本文采用 IEEE-14 和 57 节点测试系统来评估所提出检测方法的性能。在电力系统中, 通常使用 SCADA 数据进行 FDIA 攻击构建。但在足够的同步 PMU 部署下状态估计问题变成线性问题, 这能提高状态估计的速度和精度^[28]。然而, 由于成本问题, 电力系统多采用 SCADA 和 PMU 混合状态估计。鉴于电力系统结构庞大复杂, 大多数攻击者难以知晓完整的拓扑结构。因此, 本文根据文献^[29]构建了基于 SCADA 和 PMU 混合状态估计的考虑网络拓扑结构部分可知的 FDIA 攻击模型。其中, SCADA 量测数据主要由节点电压幅值、节点注入有功功率和无功功率、支路有功功率和无功功率组成。而 PMU 量测数据为 PMU 安装处的节点电压幅值和相角组成。

本文设置了两种不同的攻击场景: 1) IEEE-14 节点上攻击者掌握全部拓扑信息; 2) IEEE-57 节点

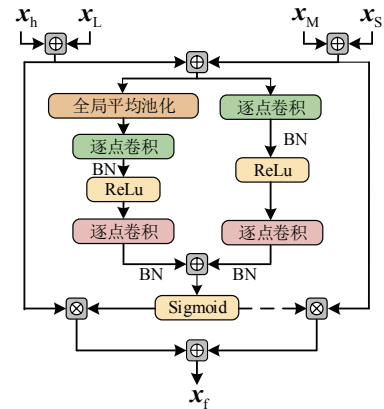


图 3 特征融合过程示意图

Fig.3 Schematic of feature fusion process

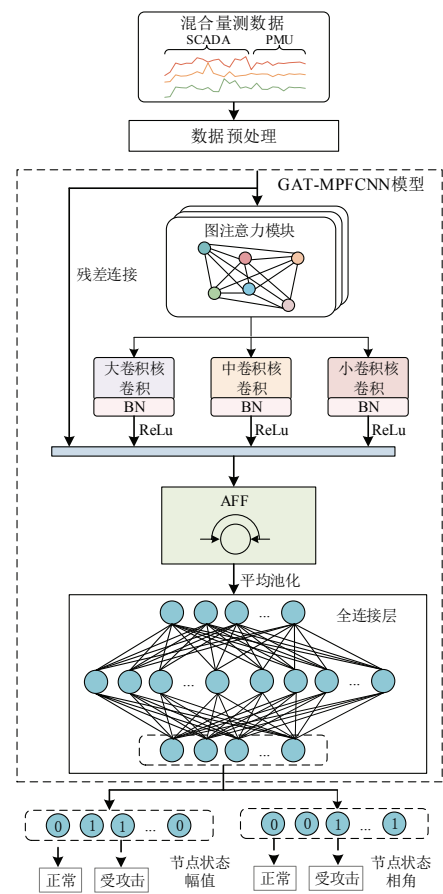


图 4 GAT-MPFCNN 整体框架

Fig.4 Overview of GAT-MPFCNN

上攻击者掌握部分拓扑信息。考虑到线路过载会对电力系统造成极大破坏, 通过对电力系统进行 FDIA 攻击可使指定线路发生过载。在这两种场景下, 本文采用广东某地采集的负荷数据, 采样频率为每 15 min 采样一次。由于实际测量中存在随机噪声, 本文在 SCADA 量测数据中加入均值为 0, 标准差为 0.02 的高斯噪声, 在 PMU 量测数据中加入均值

为 0, 标准差为 0.001 的高斯噪声^[30]。

对于 IEEE-14 和 IEEE-57 节点测试系统, 生成 2 480 个实例, 用于训练和测试模型。且由于 FDIA 定位为多标签分类问题, 针对不同的负荷工况, 随机选择攻击区域和过载线路, 以生成更平衡的攻击数据集。

3.2 实验设置

3.2.1 评价指标

本文采用准确率(I_{Acc})、精度(I_{Pre})、召回率(I_{Rec})、F1 值(I_{F1})作为评估指标来评估本文所提模型和其他定位检测模型的性能。其定义分别如下所示:

$$I_{Acc} = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}} \quad (18)$$

$$I_{Pre} = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (19)$$

$$I_{Rec} = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (20)$$

$$I_{F1} = \frac{2n_{TP}}{2n_{TN} + n_{FP} + n_{FN}} \quad (21)$$

式中: n_{TP} 、 n_{TN} 、 n_{FP} 、 n_{FN} 分别为真正例、真负例、假负例、假正例的数目。由于在实际应用场景中, 高精度与高召回率分别代表少误检和少漏检。而 F1 值作为这两者的调和平均数, 综合反映了模型性能。因此, 本文将 F1 值作为重要的评估指标。

3.2.2 数据预处理

由于 SCADA 和 PMU 组成的混合量测数据存在部分重合, 为了剔除冗余数据并防止量测数据维度过多造成模型检测效率低和耗时长, 本文使用主成分分析法(principal component analysis, PCA)^[31]对数据进行降维。经过 PCA 降维后, 能去除数据中的冗余和噪声^[32], 以较低维度保留关键信息。本文选择主成分贡献率大于 90% 的特征, 然后将降维后的数据输入检测模型进行检测与定位。

3.2.3 训练设置

FDIA 攻击的仿真模拟在 MATLAB R2022a 用 MATPOWER 工具箱构建。而定位检测方法是在 Python 3.9.16 中使用 Pytorch 2.0.1 和 Scikit-learn 1.2.2 库实现。本文将数据集划分为两部分, 分别用于训练和测试, 比例为 7:3。在图注意力模块构建中, 设置图注意力层为 3。而并行卷积模块的卷积核大小分别为 9、5、3。在注意力特征融合模块中, 设置通道缩减比为 2。两层全连接层的隐藏层维度大小设置为 130。整个模型使用 Adam 优化器进行

训练, 初始学习率为 0.005, 批处理大小设置为 32, 模型采用二元交叉熵损失函数作为损失函数。

3.3 检测模型性能分析

为了评估本文所提出的定位检测方法的性能, 将本文所提方法和现有的 FDIA 定位检测方法进行性能对比, 包括融合卷积注意力模块的卷积神经网络(convolutional neural network-convolutional block attention module, CNN-CBAM)^[33]、MGAT^[34]、基于深度学习的位置检测架构(deep learning based locational detection architecture, DLLD)^[17]、多隐层极限学习机(multi layer extreme learning machine, ML-ELM)^[35]、K 近邻算法(k-nearest neighbor, KNN)^[36]、AdaBoost^[37], 并对结果进行分析。各定位检测方法在 IEEE-14 节点测试系统的定位性能指标如表 1 所示, 在 IEEE-57 节点测试系统的定位性能指标如表 2 所示。

由表 1 与表 2 可知, GAT-MPFCNN 在 IEEE-14 和 IEEE-57 节点测试系统上的准确率、精度、召回率和 F1 值都优于其他检测方法。且 GAT-MPFCNN 在 IEEE-14 和 IEEE-57 节点测试系统都具有较高的 F1 值, 分别为 98.40% 和 95.29%。这表明本文所提定位检测方法在 FDIA 定位检测时能够保持较低错误率并精确定位到攻击注入位置。且其他模型在 IEEE-57 节点测试系统的 F1 值相较于 IEEE-14 节点

表 1 IEEE-14 节点定位性能指标对比

模型	准确率	精度	召回率	F1 值
GAT-MPFCNN	0.987 4	0.985 4	0.982 5	0.984 0
DLLD	0.977 8	0.974 6	0.970 0	0.972 3
CNN-CBAM	0.975 3	0.974 2	0.962 1	0.968 1
MGAT	0.971 1	0.964 8	0.960 9	0.962 9
ML-ELM	0.971 9	0.974 5	0.952 7	0.963 4
Adaboost	0.946 9	0.934 8	0.931 4	0.933 1
KNN	0.953 3	0.971 7	0.909 0	0.939 3

表 2 IEEE-57 节点定位性能指标对比

模型	准确率	精度	召回率	F1 值
GAT-MPFCNN	0.967 1	0.952 4	0.953 3	0.952 9
DLLD	0.956 1	0.930 4	0.944 0	0.937 1
CNN-CBAM	0.950 2	0.943 2	0.912 2	0.927 4
MGAT	0.940 0	0.906 9	0.916 7	0.911 8
ML-ELM	0.927 4	0.931 9	0.911 6	0.921 6
Adaboost	0.935 5	0.909 6	0.905 1	0.907 3
KNN	0.928 0	0.934 1	0.853 6	0.892 1

测试系统都下降到 95%以下, 这表明现有定位检测方法相较于本文的检测方法的可扩展性差, 面对大规模的电力网络系统的鲁棒性不足。

图 5 和图 6 展现了各方法在两系统上各节点状态的检测准确率。以 IEEE-14 节点测试系统为例, 每个节点有两个状态量, 由于攻击前后参考节点状态量不变, 删掉参考节点对应的 2 个标签, 共 26 个标签。从图 5 和图 6 仿真结果可见, GAT-MPFCNN 在 IEEE-14 和 IEEE-57 节点测试系统上的各节点检测准确率几乎都高于其他方法, 且分别稳定在 96% 和 92% 以上。在 IEEE-14 节点测试系统中, 其他定位检测模型的准确率在标签 13 处下降到 97% 以下, 其中 Adaboost 的准确率只有 90.19%。而在 IEEE-57 节点测试系统中, 其他模型的准确率在标签 45 处下几乎都在 90% 以下, 但本文所提出检测模型的准确率仍在 95% 左右。这表明 GAT-MPFCNN 能更好地识别 CPPS 中的 FDIA。

从图 7 和图 8 可以看出, GAT-MPFCNN 在 IEEE-14 和 IEEE-57 节点测试系统中的 F1 值第一四分位数分别为 98.09% 和 92.90%。这表明在 2 个测试系统中, 本文提出的检测方法在至少 75% 的节点的 FDIA 检测上都表现出较高的准确性和召回率的平衡。在进行 FDIA 定位检测时, 检测模型能在避免漏检的情况下不发出假警报, 从而减少电力系统的人力和财力的浪费。虽然在 IEEE-57 节点测试系统中 GAT-MPFCNN 对一些节点的定位检测 F1 值在 80% 到 90% 之间, 但电力系统工作人员能采用其他方式对其进行辅助保护。虽然其他定位检测模型能在 14 节点测试系统上具有良好的检测性能, 但在 57 节点测试系统上多数都表现一般。

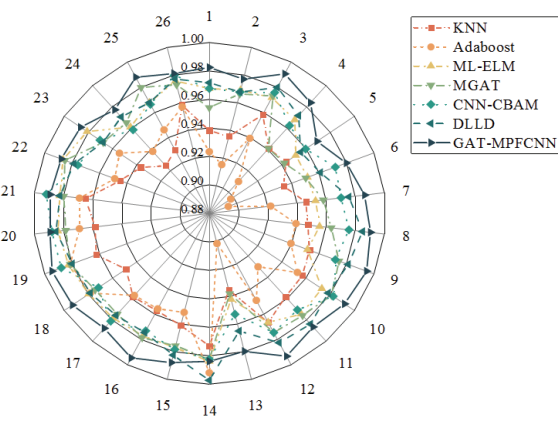


图 5 IEEE-14 节点定位检测精度

Fig.5 Location detection accuracy for IEEE-14 bus

3.4 消融实验

为了验证所提模型中每个模块对 FDIA 定位的必要性, 本文对 GAT-MPFCNN 进行消融实验。在

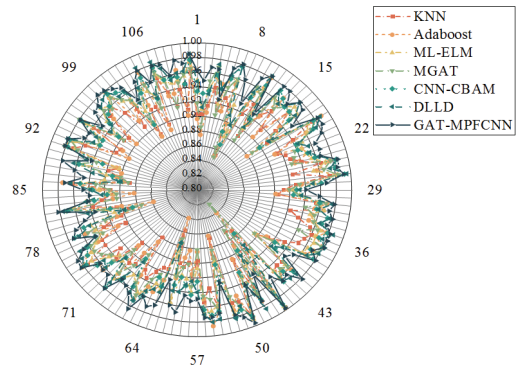


图 6 IEEE-57 节点定位检测精度

Fig.6 Location detection accuracy for IEEE-57 bus

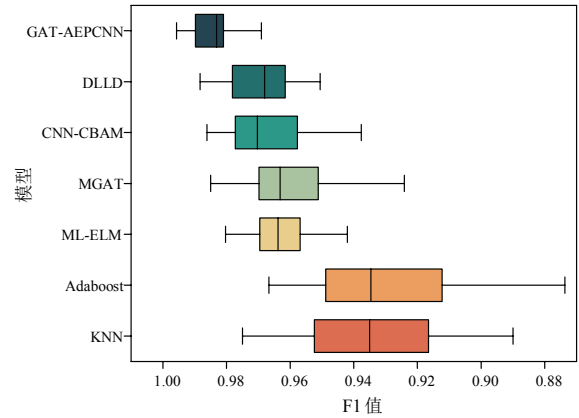


图 7 IEEE-14 节点定位检测 F1 值箱线图

Fig.7 Diagram of IEEE-14 bus localization detection F1 score box-line

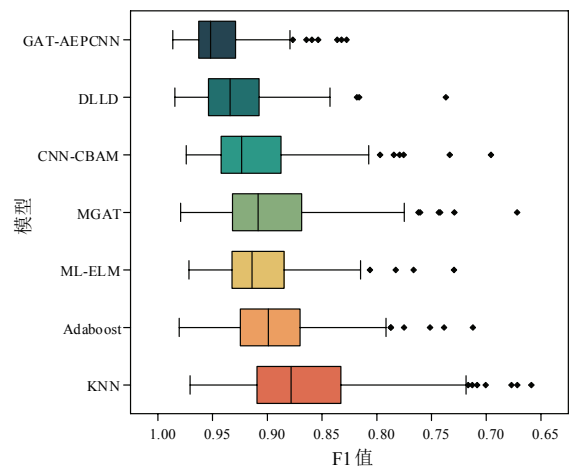


图 8 IEEE-57 节点定位检测 F1 值箱线图

Fig.8 Diagram of IEEE-57 bus localization detection F1 score box-line

实验过程中保持参数一致, 实验结果如表 3 所示, 而“-”表示禁用此模块。

由表 3 可见, 去除各个模块后模型的召回率和 F1 值都有所下降, 这表明检测模型的每个组成部分都是必要的。在 2 个测试系统中, 禁用 GAT 后, 检测方法的召回率分别下降了 0.74%、1.11%, F1 值则下降了 0.38%、0.4%, 模型的召回率下降较快, 这说明 GAT 对召回率的影响比 F1 值大, GAT 在提高检测方法定位 FDIA 的能力方面起着重要作用。在实验中较于去除 GAT 和去除 AFF, 去除 PCNN 导致模型的指标下降最明显。去除 PCNN 后模型召回率和 F1 值都下降了 2%~3%, 这表明 PCNN 对模型性能的影响最大。并且在 IEEE-57 节点测试系统中, 去除 PCNN 后模型性能下降更为明显, 这表明 PCNN 的泛化能力强, 在大规模电力系统中也能表现良好。

3.5 模型可解释性分析

作为一种非线性降维算法, t-分布邻域嵌入算法 (t-distributed stochastic neighbor embedding, t-SNE)^[38]可以将高维元素输出转换到低维空间以便于可视化和分析数据中的模式。因此, 为了增强模型的可解释性, 本文使用 t-SNE 可视化 GAT-MPFCNN 的定位过程。

以 IEEE-14 节点测试系统为例, 其结构示意图如图 9 所示。采用 t-SNE 进行可视化的对比图如图 10 所示。图中从左至右表示测试集样本经过 GAT-MPFCNN 的定位过程。由于 FDIA 可以一次攻击多个节点从而导致多个状态量同一时刻发生改变, 而每个标签对应一个状态量, 故一个样本可能有多个标签。为便于分析, 本文只对样本的部分标签进行上色。由于样本标签中, 一半标签对应节点的相角, 另一半对应节点的幅值, 因此本文对表示相角的部分标签进行上色。

图 10 中, 1 号表示正常样本, 2 号表示节点 2 相角状态量被攻击的样本, 而 3 号、4 号、5 号、6 号、7 号、8 号分别表示节点 3、6、8、9、10、13 角状态量被攻击的样本, 而 9 号表示带有剩余标签的样本。

图 10(a)展示了将 GAT-MPFCNN 的输入数据映射到二维空间的结果。从该子图可以看出, 正常样本被受攻击样本环绕, 后者分布相对分散, 因此 FDIA 的检测和定位较为困难。从图 10(b)可以看出, 经过 GAT 提取空间特征后, 部分正常样本与异常样本得到有效分离, 从而增强了模型对 FDIA 的检测

表 3 消融实验的定位结果

Table 3 Localization results of ablation study

模型	IEEE-14		IEEE-57	
	召回率	F1 值	召回率	F1 值
GAT-MPFCNN	0.982 5	0.983 9	0.953 3	0.952 9
-GAT	0.975 1	0.980 1	0.942 2	0.948 9
-PCNN	0.963 0	0.968 2	0.927 5	0.933 5
-AFF	0.978 1	0.978 9	0.944 3	0.947 6

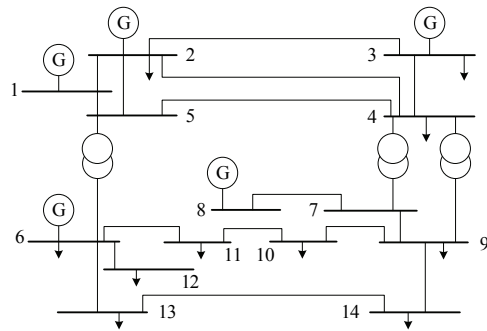


图 9 IEEE-14 节点测试系统结构示意图

Fig.9 Schematic diagram of IEEE14 bus test system structure

性能。然而, 准确定位 FDIA 仍存在挑战。在图 10(c)中, 由于 MPFCNN 进一步捕获并融合了多尺度特征, 使得不同标签的样本更有效地聚集, 同时样本间的距离也得以增加, 提升了检测方法的定位性能。从图 10(d)可得出, 双层全连接层对学习的特征进行整合, 不同标签的样本得以更好地聚集, 进一步提升了模型的定位准确性。

4 结论

1) 本文提出一种基于 GAT-MPFCNN 的 FDIA 定位检测方法。其通过引入 GAT 动态捕捉量测数据间的拓扑关系, 能够有效利用节点之间的相关性, 以适应易变、复杂的电力系统拓扑的动态变化; 采用 MPFCNN 提取并动态融合量测数据的多尺度特征, 以提高检测方法的泛化能力和学习能力。

2) 本文所提检测方法在 IEEE-14 节点测试系统上仿真验证得到的 F1 值、召回率分别为 98.40%、98.25%, 在 IEEE-57 节点测试系统上则为 95.29%、95.33%。结果表明该检测方法的综合性能均优于目前较先进的 FDIA 定位检测方法。此外, 本文利用消融实验验证了 GAT-MPFCNN 中每个模块的有效性。并且本文通过可视化检测方法的定位过程, 提高了检测方法的解释性。

3) 在定位 FDIA 后, 为降低其对 CPPS 造成的

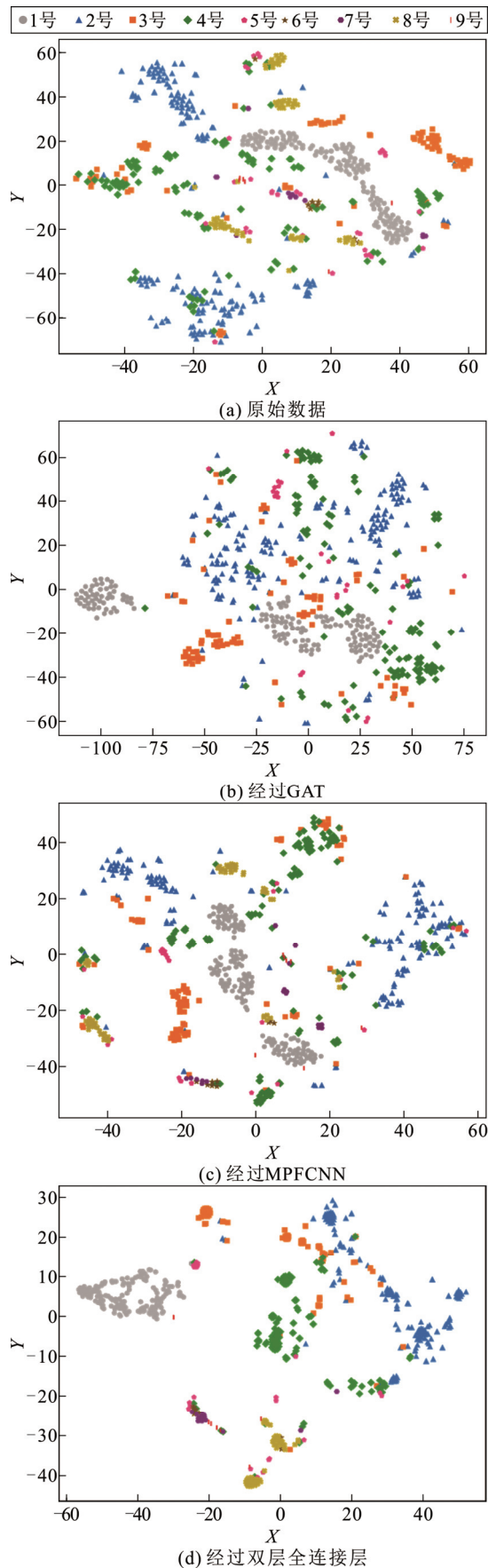


图 10 IEEE-14 节点 t-SNE 可视化对比图

Fig.10 Comparison map of IEEE-14 bus t-SNE visualization

影响, 采取有效的防御措施是至关重要的。因此, 未来研究将重点探索攻击后的 FDIA 防御策略, 目标是构建一个集定位与防御于一体的 FDIA 防御框架, 以增强 CPPS 的安全性和稳定性。

参考文献 References

- [1] 龚立, 王先培, 田猛, 等. 电力信息物理系统韧性的概念与提升策略研究进展[J]. 电力系统保护与控制, 2023, 51(14): 169-187. GONG Li, WANG Xianpei, TIAN Meng, et al. Concepts and research progress on enhancement strategies for cyber physical power system resilience[J]. Power System Protection and Control, 2023, 51(14): 169-187.
- [2] 王琦, 李梦雅, 汤奕, 等. 电力信息物理系统网络攻击与防御研究综述(一)建模与评估[J]. 电力系统自动化, 2019, 43(9): 9-21. WANG Qi, LI Mengya, TANG Yi, et al. A review on research of cyber-attacks and defense in cyber physical power systems part one modelling and evaluation[J]. Automation of Electric Power Systems, 2019, 43(9): 9-21.
- [3] 苏盛, 吴长江, 马钧, 等. 基于攻击方视角的电力 CPS 网络攻击模式分析[J]. 电网技术, 2014, 38(11): 3115-3120. SU Sheng, WU Changjiang, MA Jun, et al. Attacker's perspective based analysis on cyber attack mode to cyber-physical system[J]. Power System Technology, 2014, 38(11): 3115-3120.
- [4] 田猛, 王先培, 董政呈, 等. 基于拉格朗日乘子法的虚假数据攻击策略[J]. 电力系统自动化, 2017, 41(11): 26-32. TIAN Meng, WANG Xianpei, DONG Zhengcheng, et al. Injected attack strategy for false data based on Lagrange multipliers method[J]. Automation of Electric Power Systems, 2017, 41(11): 26-32.
- [5] 张博, 刘绚, 于宗超, 等. 基于人工智能的电力系统网络攻击检测研究综述[J]. 高电压技术, 2022, 48(11): 4413-4426. ZHANG Bo, LIU Xuan, YU Zongchao, et al. Review on artificial intelligence-based network attack detection in power systems[J]. High Voltage Engineering, 2022, 48(11): 4413-4426.
- [6] 王琦, 邵伟, 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.
- [7] LIU X, BAO Z, LU D, et al. Modeling of local false data injection attacks with reduced network information[J]. IEEE Transactions on Smart Grid, 2015, 6(4): 1686-1696.
- [8] 巨云涛, 于燕玲, 张紫枫, 等. 计及坏数据辨识的微网群三相分布式状态估计方法[J]. 高电压技术, 2022, 48(4): 1251-1263. JU Yuntao, YU Yanling, ZHANG Zifeng, et al. Three-phase distributed state estimation method of microgrid group considering bad data identification[J]. High Voltage Engineering, 2022, 48(4): 1251-1263.
- [9] 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]. 中国电机工程学报, 2023, 43(22): 8602-8621. YANG Yuze, LIU Wenxia, LI Chengze, et al. Review of FDIA detection methods for electric power SCADA system[J]. Proceedings of the CSEE, 2023, 43(22): 8602-8621.
- [10] LIU C S, DENG R L, HE W L, et al. Optimal coding schemes for detecting false data injection attacks in power system state estimation[J]. IEEE Transactions on Smart Grid, 2022, 13(1): 738-749.
- [11] JORJANI M, SEIFI H, VARJANI A Y. A graph theory-based approach to detect false data injection attacks in power system AC state estimation[J]. IEEE Transactions on Industrial Informatics, 2021, 17(4): 2465-2475.
- [12] CHAKRABARTY S, SIKDAR B. Detection of malicious command injection attacks on phase shifter control in power systems[J]. IEEE Transactions on Power Systems, 2021, 36(1): 271-280.
- [13] 董运昌, 王启明, 曹杰, 等. 基于过采样和级联机器学习的电网虚假数据注入攻击识别[J]. 电力系统自动化, 2023, 47(8): 179-188. DONG Yunchang, WANG Qiming, CAO Jie, et al. Identification of false data injection attacks in power grid based on oversampling and cascade machine learning[J]. Automation of Electric Power Systems, 2023, 47(8): 179-188.

- [14] CHEN B R, WU Q H, LI M S, et al. Detection of false data injection attacks on power systems using graph edge-conditioned convolutional networks[J]. *Protection and Control of Modern Power Systems*, 2023, 8(1): 16.
- [15] PARIZAD A, HATZIADONIU C J. A real-time multistage false data detection method based on deep learning and semisupervised scoring algorithms[J]. *IEEE Systems Journal*, 2023, 17(2): 1753-1764.
- [16] DOU C X, WU D, YUE D, et al. A hybrid method for false data injection Attack detection in smart grid based on variational mode decomposition and OS-ELM[J]. *CSEE Journal of Power and Energy Systems*, 2022, 8(6): 1697-1707.
- [17] WANG S Y, BI S Z, ZHANG Y J A. Locational detection of the false data injection attack in a smart grid: a multilabel classification approach[J]. *IEEE Internet of Things Journal*, 2020, 7(9): 8218-8227.
- [18] BOYACI O, NARIMANI M R, DAVIS K R, et al. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks[J]. *IEEE Transactions on Smart Grid*, 2022, 13(1): 807-819.
- [19] YU J J Q. Graph construction for traffic prediction: a data-driven approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(9): 15015-15027.
- [20] 李叶, 毛伊敏, 陈志刚. 基于 Winograd 卷积的并行深度卷积神经网络优化算法[J]. *信息与控制*, 2023, 52(4): 466-482.
LI Ye, MAO Yimin, CHEN Zhigang. Winograd-based parallel deep convolutional neural network optimization algorithm[J]. *Information and Control*, 2023, 52(4): 466-482.
- [21] GUO N N, LIN C, YAN H, et al. Real-time pantograph anomaly detection using unsupervised deep learning and K-Nearest Neighbor classification[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 3512013.
- [22] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//*Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE, 2021: 3560-3569.
- [23] 王韵琳, 冯天波, 孙宁, 等. 融合注意力与多尺度特征的电力绝缘子缺陷检测方法[J]. *高电压技术*, 2024, 50(5): 1933-1942.
WANG Yunlin, FENG Tianbo, SUN Ning, et al. Defect detection method for power insulators based on attention and multi-scale context information[J]. *High Voltage Engineering*, 2024, 50(5): 1933-1942.
- [24] 巫春玲, 郑克军, 徐先峰, 等. 基于自适应插值强跟踪扩展卡尔曼滤波的电力系统动态状态估计研究[J]. *电网技术*, 2023, 47(5): 2078-2088.
WU Chunling, ZHENG Kejun, XU Xianfeng, et al. Dynamic state estimation of power system based on adaptive interpolation strong tracking extended Kalman filter[J]. *Power System Technology*, 2023, 47(5): 2078-2088.
- [25] 高正男, 杨帆, 胡姝博, 等. 面向新能源电力系统状态估计的伪波动数据清洗[J]. *高电压技术*, 2022, 48(6): 2366-2377.
GAO Zhengnan, YANG Fan, HU Shubo, et al. Pseudo-fluctuation data cleaning for state estimation of new energy power system[J]. *High Voltage Engineering*, 2022, 48(6): 2366-2377.
- [26] 王渝红, 吴恒帅, 于光远, 等. 基于图注意力网络的风电场汇集并网系统次同步振荡预警方法[J]. *高电压技术*, 2023, 49(7): 2995-3005.
WANG Yuhong, WU Hengshuai, YU Guangyuan, et al. Subsynchronous oscillation early warning method for wind farm integration grid system based on graph attention network[J]. *High Voltage Engineering*, 2023, 49(7): 2995-3005.
- [27] ALSOBHI W, ALAFIF T, ZONG W W, et al. Adaptive batch normalization for training data with heterogeneous features[C]// *Proceedings of 2023 International Conference on Smart Computing and Application*. Hail, Saudi Arabia: IEEE, 2023: 1-6.
- [28] VALVERDE G, CHAKRABARTI S, KYRIAKIDES E, et al. A constrained formulation for hybrid state estimation[J]. *IEEE Transactions on Power Systems*, 2011, 26(3): 1102-1109.
- [29] WU T, XUE W L, WANG H Z, et al. Extreme learning machine-based state reconstruction for automatic attack filtering in cyber physical power system[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(3): 1892-1904.
- [30] 席磊, 田习龙, 余涛, 等. 基于相关特征-多标签级联提升森林的电网虚假数据注入攻击定位检测[J]. *南方电网技术*, 2024, 18(5): 39-50, 61.
XI Lei, TIAN Xilong, YU Tao, et al. Locational detection of false data injection attack in power grid based on relevant features multi-label cascade boosting forest[J]. *Southern Power System Technology*, 2024, 18(5): 39-50, 61.
- [31] CHEN J H, LIAO C M. Dynamic process fault monitoring based on neural network and PCA[J]. *Journal of Process Control*, 2002, 12(2): 277-289.
- [32] 葛泉波, 程惠茹, 张明川, 等. 基于 PCA 和 ICA 模式融合的非高斯特征检测识别[J]. *自动化学报*, 2024, 50(1): 169-180.
GE Quanbo, CHENG Huiru, ZHANG Mingchuan, et al. Non-Gaussian feature detection and recognition based on PCA and ICA pattern fusion[J]. *Acta Automatica Sinica*, 2024, 50(1): 169-180.
- [33] 周先军, 王茹, 刘航, 等. 基于 CNN-CBAM 的虚假数据注入攻击辨识研究[J]. *光通信研究*, 2024(3): 230125.
ZHOU Xianjun, WANG Ru, LIU Hang, et al. Research on false data injection attack identification based on CNN-CBAM[J]. *Study on Optical Communications*, 2024(3): 230125.
- [34] 苏向敬, 邓超, 栗风永, 等. 基于 MGAT-TCN 模型的可解释电网虚假数据注入攻击检测方法[J]. *电力系统自动化*, 2024, 48(2): 118-127.
SU Xiangjing, DENG Chao, LI Fengyong, et al. Interpretable detection method for false data injection attack on power grid based on multi-head graph attention network and time convolution network model[J]. *Automation of Electric Power Systems*, 2024, 48(2): 118-127.
- [35] 席磊, 何苗, 周博奇, 等. 基于改进多隐层极限学习机的电网虚假数据注入攻击检测[J]. *自动化学报*, 2023, 49(4): 881-890.
XI Lei, HE Miao, ZHOU Boqi, et al. Research on false data injection attack detection in power system based on improved multi layer Extreme Learning Machine[J]. *Acta Automatica Sinica*, 2023, 49(4): 881-890.
- [36] OZAY M, ESNAOLA I, YARMAN VURAL F T, et al. Machine learning methods for attack detection in the smart grid[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(8): 1773-1786.
- [37] WANG X Y, LUO X Y, ZHANG Y Y, et al. Detection and isolation of false data injection attacks in smart grids via nonlinear interval observer[J]. *IEEE Internet of Things Journal*, 2019, 6(4): 6498-6512.
- [38] 吴旻昊, 王建功, 朱英刚, 等. 基于 t-SNE 降维与聚类的主动配电网运行方式在线识别[J]. *电力建设*, 2023, 44(8): 52-60.
WU Minhao, WANG Jiangong, ZHU Yinggang, et al. Online identification of active distribution network operation mode based on t-SNE dimensionality reduction and clustering[J]. *Electric Power Construction*, 2023, 44(8): 52-60.



XI Lei
Ph.D., Professor



LI Zongze
Corresponding author

席磊

1982—, 男, 博士, 教授, 博导
主要从事自动发电控制、人工智能、电力系统网络攻击与防御方面的研究工作
E-mail: xilei2014@163.com

李宗泽(通信作者)

2001—, 男, 硕士生
从事信息物理系统网络攻击与防御方面的研究
E-mail: Lizongze0608@163.com