DOI: 10.19783/j.cnki.pspc.241690

基于关联差异的电力信息物理系统虚假数据注入攻击检测

郭晓利[1,2]，王 月[1]，李 斌[1]

(1. 东北电力大学计算机学院，吉林 吉林 132012；2. 吉林省电力大数据智能处理工程技术研究中心，吉林 吉林 132012)

摘要：为了保障智能电网的安全稳定运行，快速、准确的虚假数据注入攻击(false data injection attacks, FDIA)检测至关重要。现有基于数据驱动的 FDIA 检测模型主要依赖固定的判别阈值进行异常识别，但这一方法存在显著不足：攻击者可通过持续试探与分析模型响应行为，逐步调整注入攻击幅度，从而绕过检测机制，导致检测精度下降。针对这一问题，提出基于关联差异的 FDIA 检测模型。首先，从数据间关联的角度出发，设计基于关联差异的 FDIA 检测模型结构。其次，通过嵌入位置信息并引入修正因子以约束注意力作用范围，提出基于位置修正因子的先验关联提取方法。然后，结合量测数据序列的细粒度和多尺度特性，提出基于双流粒度对齐的序列关联提取方法。最后，引入拓扑关联，定义关联差异，并设计基于关联差异的对抗性判别准则推理方法，通过对抗训练放大正常与攻击量测数据的可区分度，得到判别准则。实验结果表明，所提模型相比于现有检测模型具有更高的检测准确率和鲁棒性，在面对不同幅值的注入攻击时表现优异。

关键词：虚假数据注入攻击；关联差异；量测数据；对抗性判别准则

False data injection attack detection in cyber-physical power systems based on correlation discrepancy

GUO Xiaoli[1,2], WANG Yue[1], LI Bin[1]

(1. School of Computer Science, Northeast Electric Power University, Jilin 132012, China; 2. Jilin Engineering Technology Research Center of Intelligent Electric Power Big Data Processing, Jilin 132012, China)

Abstract: To ensure the secure and stable operation of smart grids, fast and accurate detection of false data injection attacks (FDIA) is critical. Existing data-driven FDIA detection models primarily rely on fixed discrimination thresholds for anomaly identification. However, this approach has notable limitations: attackers can iteratively probe and analyze model responses, gradually adjusting the magnitude of injected attacks to bypass detection, thereby reducing detection accuracy. To address this issue, this paper proposes a FDIA detection model based on correlation discrepancy. First, a detection framework centered on data correlation discrepancies is designed. Second, a position-aware correction factor is embedded to constrain attention scopes, enabling prior correlation extraction with enhanced positional awareness. Then, leveraging the fine-grained and multi-scale characteristics of measurement data sequences, a dual-stream granularity alignment method is developed to capture sequential correlations. Finally, topological correlations are incorporated to define correlation discrepancies, and an adversarial discrimination criterion is formulated through adversarial training to amplify the distinguishability between normal and attacked measurements, resulting in an effective discrimination criterion. Experimental results demonstrate that the proposed model achieves superior detection accuracy and robustness compared with existing methods and performs well under injection attacks of varying magnitudes.

This work is supported by the National Natural Science Foundation of China (No. 52377081).

Key words: false data injection attack; correlation discrepancy; measurement data; adversarial discrimination criteria

0 引言

随着新型电力系统的发展，先进信息通信技术在电力系统中被广泛应用，电力系统已经演变为信息物理紧密耦合的电力信息物理系统(cyber-physical power system, CPPS)[1-3]，提高了电力系统的广泛感知、高效通信和高性能计算的能力。然而，信息流和电力潮流频繁互动，也增加了系统的脆弱性[4-5]。其中虚假数据注入攻击(false data injection attacks, FDIA)代表了一类新型的网络攻击[6-9]，攻击者根据已知部分电网参数设计攻击向量并篡改终端采集的

基金项目：国家自然科学基金项目资助(52377081)；吉林省自然科学基金项目资助(20220101234JC)

量测数据，使得这些数据可以绕过不良数据检测(bad data detection, BDD)误导控制中心产生错误的状态估计[10]，严重威胁电网安全运行。

目前，国内外对电网 FDIA 的检测进行了大量研究，主要研究思路有：基于系统模型的检测方法和基于数据驱动的检测方法。系统模型检测方法利用电网系统模型和不同测量集来求解更精准的状态估计。例如，文献[11]在保留原有加权最小二乘估计器的同时引入附加卡尔曼滤波器，设置检测阈值，通过两个估计测量的偏差检测 FDIA。为了适应电网运行状态的动态变化，文献[12]结合动态状态估计器和卡尔曼滤波器进行状态估计。由于 FDIA 具有显著稀疏性[13]，即集中攻击少量关键量测设备或时间点，以最小代价影响电力系统状态估计。因此，也有研究者通过观察攻击前后量测数据的物理规律变化来进行 FDIA 检测，例如，文献[14]利用测量矩阵的低秩特性和攻击矩阵的稀疏性，将 FDIA 检测转化为矩阵分离问题。然而，系统模型方法的准确性受到模型精度、计算复杂度与检测延迟的制约[15]。

与基于电网系统模型检测的方法不同，基于数据驱动的检测方法在检测过程中不涉及电网系统模型和参数，而是学习历史数据中的隐含特征，以挖掘 FDIA 的异常特性。例如，文献[16]提出了一种新型 FDIA 检测方法，该方法利用基于集成学习的递归神经网络(recurrent neural network, RNN)来拟合测量数据的非线性特征，但随着时间序列长度的增加，训练过程中会出现梯度爆炸或消失的问题。进而，长短期记忆网络(long short-term memory, LSTM)[17]和门控循环单元(gated recurrent unit, GRU)[18]逐渐被用来捕捉量测数据的长期依赖关系。

同时，电网是由发电机和负载相互连接组成的网络[19]，其运行特性高度依赖于其拓扑结构。因此，也有研究者通过量测数据的空间特征来解决 FDIA 检测问题，文献[20]提出一种基于图神经网络(graph neural network, GNN)的 FDIA 检测方法。考虑到传统基于标量的图神经网络难以适应电力系统中负载、发电量及拓扑结构的高度动态变化，文献[21]将电网建模为加权双向图，并利用路由机制构造蕴含位置、方向与连接关系等系统属性信息的向量表征，进而提出一种基于向量的胶囊神经网络算法来检测 FDIA。作为 GNN[22]的一种变体，图卷积网络(graph convolutional network, GCN)[23]将卷积推广到图域，结合电网固有的拓扑结构有效捕获空间信息，提高了 FDIA 检测性能。也有研究者提出基于量测数据时空相关性的 FDIA 检测方法，例如，文献[24]并行挖掘电网量测数据的时序和空间特征信息，相比以往仅提取时间或空间特征的检测模型检测准确率和效率更高。

以上基于数据驱动的方法假设攻击行为是静态的且攻击幅值是固定的，通过设定固定的可区分阈值来实现攻击检测。然而在实际场景中，攻击者会逐步试探和分析模型的响应，进而动态调整攻击幅值以规避检测，这种逐步调整攻击幅值的行为导致传统方法在应对变化的攻击模式时表现不佳。

已有研究表明，序列关联是正常量测数据的一个关键特征，即每个时间点的数据都与整个序列高度相关。与正常数据不同，攻击数据具有稀有性，通常难以与整个序列建立广泛的关联。攻击数据具有独特的时序依赖性，其关联并非均匀分布，而是集中出现在具有相似攻击模式的前置相邻时间点，这种关联模式即为先验关联。基于这两种关联来分析攻击行为，将某一时间点的序列关联与先验关联之差定义为关联差异，攻击数据的关联差异通常小于正常数据的关联差异[25]。但该定义方法仅考虑了 FDIA 对量测数据时间模式的破坏，未考虑 FDIA 也会影响量测数据间的空间关联。

针对上述问题并结合已有研究基础，本文进一步引入拓扑关联，即基于电网拓扑结构分析量测数据间的空间关联，更全面刻画 FDIA 特征。针对 FDIA 攻击者调整攻击幅值的攻击行为，提出了一种基于关联差异(association discrepancy, AD)的电网 FDIA 检测模型。该模型将关联差异重新定义为，在特定时间点上由序列、先验与拓扑关联整合而成的综合差异度量，并基于关联差异对抗训练得到最大化的判别准则，实现检测阈值的动态寻优，以应对变化的攻击模式。

1　虚假数据注入攻击相关问题描述

在电网中，来自终端仪表的测量数据流是完成电力系统状态估计不可或缺的部分。测量方程表示为

$$\boldsymbol{z} = f(\boldsymbol{x}) + \boldsymbol{e} \tag{1}$$

式中：$\boldsymbol{z}$ 为聚合测量数据组成的测量向量；$f(\cdot)$ 为测量值与状态向量之间的非线性函数关系，由特定电网的拓扑结构和参数决定；$\boldsymbol{x}$ 为当前时刻电网的状态向量；$\boldsymbol{e}$ 为测量噪声。

新型电力系统中电网 FDIA 的潜在场景如图 1 所示。可以看出：FDIA 攻击者通过精心设计虚假量测数据，绕过状态估计中的不良数据检测，进而导致错误的状态估计结果。

当存在注入虚假数据时，状态估计的测量方程式(1)变为式(2)。

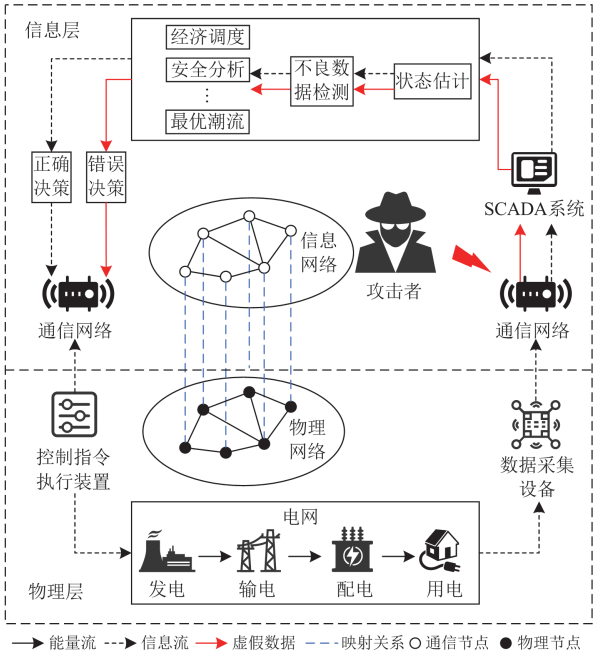


图 1 电网 FDIA 潜在场景

Fig. 1 Possible scenarios of FDIA in power grids

$$z_{\mathrm{A}}=f(\boldsymbol{x})+\boldsymbol{e}+\boldsymbol{a} \tag{2}$$

式中：z_{A} 为当前时刻观察到的测量向量；$\boldsymbol{a}$ 为注入的攻击向量。

定义 $\hat{\boldsymbol{x}}_{\mathrm{A}}$ 为在虚假数据注入后从测量向量 z_{A} 估计的状态，计算如式(3)所示。

$$\hat{\boldsymbol{x}}_{\mathrm{A}}=\hat{\boldsymbol{x}}+\boldsymbol{c} \tag{3}$$

式中：$\hat{\boldsymbol{x}}$ 为未被攻击的原始测量向量 $\boldsymbol{z}$ 估计的状态向量；$\boldsymbol{c}$ 为攻击向量引起的状态估计误差。

基于攻击者注入的具有虚假数据的测量数据，通过加权最小二乘法来估计状态，如式(4)所示。

$$J(\hat{\boldsymbol{x}}_{\mathrm{A}})=[z_{\mathrm{A}}-f(\boldsymbol{x}_{\mathrm{A}})]^{\mathrm{T}}\boldsymbol{\Delta}^{-1}[z_{\mathrm{A}}-f(\boldsymbol{x}_{\mathrm{A}})] \tag{4}$$

式中：$\boldsymbol{\Delta}$ 为测量误差的协方差矩阵。

考虑到式(2)中的测量噪声 $\boldsymbol{e}$ 是高斯分布并且具有零均值，则估计的状态变量如式(5)所示。

$$\hat{\boldsymbol{x}}_{\mathrm{A}}=\hat{\boldsymbol{x}}+(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\boldsymbol{a} \tag{5}$$

式中：$\boldsymbol{F}$ 为测量函数 $f(\cdot)$ 的雅可比矩阵。

检测残差 r 被定义为观测到的 $\boldsymbol{z}$ 与估计的状态向量 $\hat{\boldsymbol{x}}$ 获得的测量值之间的差，表示为

$$r=\left\|\boldsymbol{z}-f(\hat{\boldsymbol{x}})\right\|_2 \tag{6}$$

式中：$\|\cdot\|_p$ 表示计算 p 范数。

使用检测残差 r 与预定义的不良数据检测阈值 τ 进行比较，如果 $r<\tau$，则量测数据正常，否则，则认为存在不良量测数据。

攻击者通常在波动的测量残差范围内设计攻击向量，使虚假数据能够有效地绕过不良数据的检测。从式(2)、式(3)和式(6)，可以得到攻击后检测残差 r_{A} 为

$$r_{\mathrm{A}}=r_{\mathrm{N}}+\tau_{\mathrm{A}} \tag{7}$$

式中：r_{N} 表示无攻击发生时的正常残差；τ_{A} 为攻击引起的检测偏差。τ_{A} 定义如式(8)所示。

$$\tau_{\mathrm{A}}=\left\|\boldsymbol{a}+f(\hat{\boldsymbol{x}})-f(\hat{\boldsymbol{x}}+\boldsymbol{c})\right\|_2 \tag{8}$$

令 $\tau_{\mathrm{A}}=0$，由式(8)可得到攻击向量 $\boldsymbol{a}$。

$$\boldsymbol{a}=f(\hat{\boldsymbol{x}}+\boldsymbol{c})-f(\hat{\boldsymbol{x}}) \tag{9}$$

在这种情况下，$r_{\mathrm{A}}=r_{\mathrm{N}}$，注入的虚假数据理想地绕过残差检测。

2　基于关联差异的 FDIA 检测模型

2.1　检测模型架构

所提出的基于关联差异的检测模型如图 2 所示，主要包括以下 5 个部分。

1) 时序特征提取模块，为关联提取模块提供高质量的特征输入。

2) 序列关联提取模块，设计两种卷积网络提取每个量测数据时间点的序列关联。

3) 先验关联提取模块，使用基于位置修正因子的注意力机制提取每个时间点的先验关联。

4) 拓扑关联提取模块，引入多头自注意力机制提取每个时间点下量测数据间的空间关联。

5) 关联差异对抗训练模块，基于 3 种关联定义关联差异，设计对抗训练机制动态调整检测阈值。

检测模型的输入为时间上连续的待检测量测数据样本，其中输入特征包含节点的电压幅值、有功功率、无功功率和支路有功、无功功率等量测量。输出为检测结果标签，标签“0”表示样本未遭受 FDIA，标签“1”表示样本遭受 FDIA。

2.2　基于位置修正因子的先验关联提取方法

由于自动编码器(autoencoder, AE)能够高效学习时序数据复杂的非线性特征并捕捉时序数据中的动态变化，设计基于双向门控循环单元的自编码器 BiGRU-AE 来提取量测数据中的时间依赖关系。使用 BiGRU 层替代 AE 中的全连接层，在减少参数量的同时，保持较强的学习能力。

GRU 简化了 LSTM 的结构，将输入门和遗忘门合并为更新门，减少了训练参数并加快了收敛速度，其结构如图 3 所示。

GRU 的机制可表示为

$$\boldsymbol{z}_t=\sigma(\boldsymbol{W}_{\mathrm{z}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{z}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{z}}) \tag{10}$$

$$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_{\mathrm{r}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{r}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{r}}) \tag{11}$$

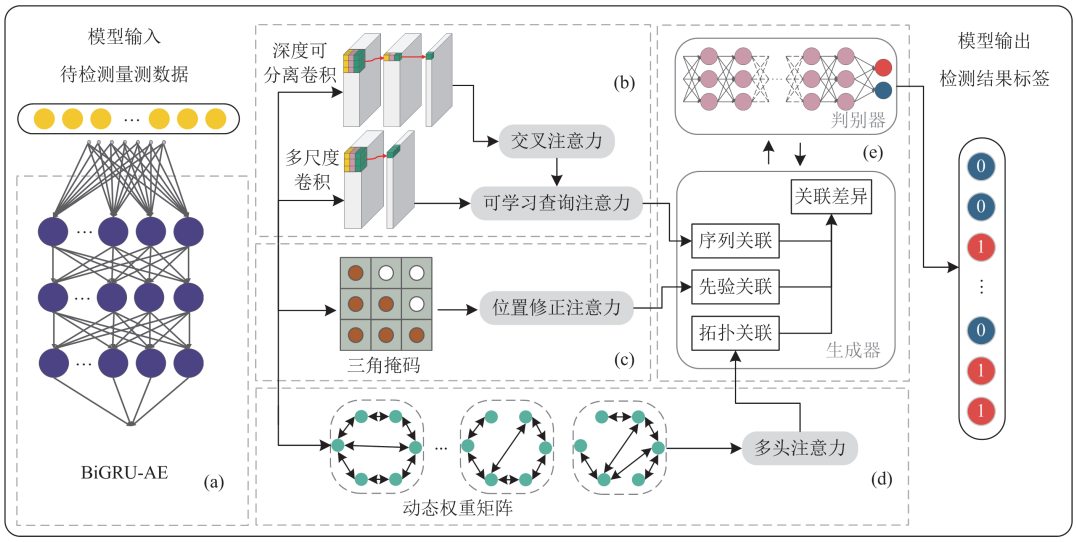


图 2　基于关联差异的 FDIA 检测模型结构

Fig. 2 Structure of the FDIA detection model based on association discrepancy

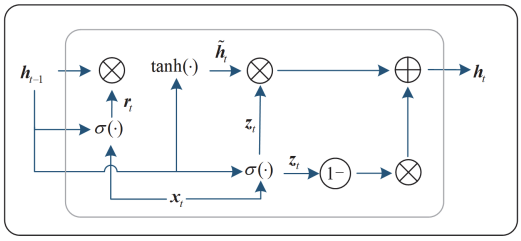


图 3　GRU 结构

Fig. 3 Structure of GRU

$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{U}_{\mathrm{h}}(\boldsymbol{r}_t \otimes \boldsymbol{h}_{t-1}) + \boldsymbol{W}_{\mathrm{h}}\boldsymbol{x}_t + \boldsymbol{b}_{\mathrm{h}}) \tag{12}$$

$$\boldsymbol{h}_t = (1-\boldsymbol{z}_t) \otimes \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \otimes \tilde{\boldsymbol{h}}_t \tag{13}$$

式中：$\tilde{\boldsymbol{h}}_t$ 为候选隐藏状态；$\boldsymbol{z}_t$、$\boldsymbol{r}_t$ 分别为更新门和重置门在 t 时刻的输出；$\sigma(\cdot)$ 为 Sigmoid 函数；$\boldsymbol{W}_{\mathrm{z}}$、$\boldsymbol{W}_{\mathrm{r}}$、$\boldsymbol{W}_{\mathrm{h}}$ 和 $\boldsymbol{U}_{\mathrm{z}}$、$\boldsymbol{U}_{\mathrm{r}}$、$\boldsymbol{U}_{\mathrm{h}}$ 为可学习的权重矩阵，控制信息流的传递；$\boldsymbol{x}_t$ 为 t 时刻的输入；$\boldsymbol{b}_{\mathrm{z}}$、$\boldsymbol{b}_{\mathrm{r}}$ 和 $\boldsymbol{b}_{\mathrm{h}}$ 为学习的偏置向量，调整每个门的输出；$\boldsymbol{h}_t$、$\boldsymbol{h}_{t-1}$ 分别为 t、$t-1$ 时刻的隐藏状态；tanh(·) 为双曲正切函数；“$\otimes$”为逐元素相乘运算。

BiGRU-AE 结构如图 4 所示，数值特征 F_t 首先通过嵌入层映射为低维向量 $\boldsymbol{f}_t$。然后，向量 $\boldsymbol{f}_t$ 被送入 BiGRU 层，生成前向隐藏状态 $\overrightarrow{\boldsymbol{h}}_t$ 和后向隐藏状态 $\overleftarrow{\boldsymbol{h}}_t$。最终，将向量 $\overrightarrow{\boldsymbol{h}}_t$、$\overleftarrow{\boldsymbol{h}}_t$ 进行拼接，并输入线性层，经过 LeakyReLU 激活函数处理，得到经 BiGRU-AE 提取的特征表示 $\boldsymbol{F}_{\mathrm{BA}}$，如式(14)所示。

$$\boldsymbol{F}_{\mathrm{BA}} = \mathrm{LeakyReLU}(\boldsymbol{W}_{\mathrm{s}}[\overrightarrow{\boldsymbol{h}}_t;\overleftarrow{\boldsymbol{h}}_t]) \tag{14}$$

式中：$\boldsymbol{W}_{\mathrm{s}}$ 为可学习的权重矩阵；$[\cdot;\cdot]$ 为向量拼接操作。

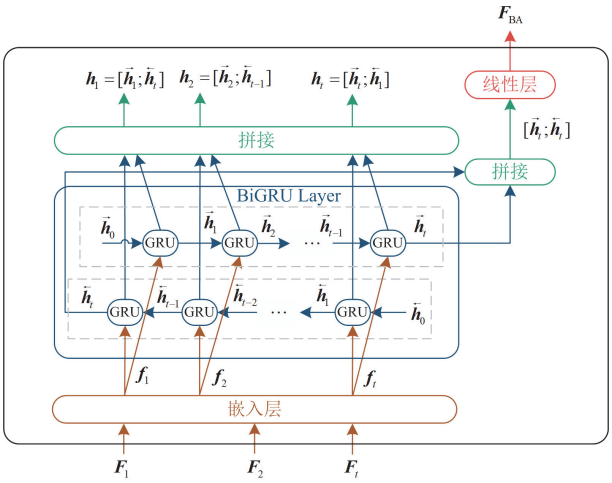


图 4　BiGRU-AE 结构

Fig. 4 Structure of BiGRU-AE

FDIA 具有局部集中性，仅影响局部节点或时段。自注意力机制虽能提取数据关联性，但全局计算注意力权重会弱化局部攻击特征。为此通过嵌入位置信息并引入修正因子约束注意力范围，增强攻击局部特征的捕捉能力。

首先，将输入序列的位置信息嵌入到注意力权重计算中。对输入特征 $\boldsymbol{F}_{\mathrm{BA}}$ 进行线性变换后得到

$$\boldsymbol{Q} = \boldsymbol{F}_{\mathrm{BA}} \cdot \boldsymbol{P}_{\mathrm{Q}} \tag{15}$$

$$\boldsymbol{K} = \boldsymbol{F}_{\mathrm{BA}} \cdot \boldsymbol{P}_{\mathrm{K}} \tag{16}$$

$$\boldsymbol{V} = \boldsymbol{F}_{\mathrm{BA}} \cdot \boldsymbol{P}_{\mathrm{V}} \tag{17}$$

式中：$\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 分别为查询向量、键向量和值向量；$\boldsymbol{P}_{\mathrm{Q}}$、$\boldsymbol{P}_{\mathrm{K}}$ 和 $\boldsymbol{P}_{\mathrm{V}}$ 为对应的权重矩阵，分别作用于输入特征以学习互补的特征表示。

修正因子 $\boldsymbol{\lambda}$ 是与输入序列等长的可学习权重向量，通过动态学习位置信息调节各位置注意力权重，实现对时间点的精准关注。

$$\boldsymbol{\lambda}=\mathrm{Softmax}(F_{\mathrm{C}}(\boldsymbol{F}_{\mathrm{BA}})) \tag{18}$$

式中：$F_{\mathrm{C}}(\cdot)$ 为全连接层；Softmax(·) 为激活函数。

接着，引入一个下三角掩码矩阵 $\boldsymbol{T}_{\mathrm{CM}}$，作用于注意力权重计算过程。该掩码矩阵确保序列中任一时间点仅能关注到之前的时间点，从而有效防止未来信息泄露，严格保证了先验关联的因果性。

$$\boldsymbol{T}_{\mathrm{CM}}=\begin{cases}1, & a>b\\ 0, & a\leqslant b\end{cases} \tag{19}$$

式中：a、b 为对应时间点在序列中的位置索引。

其次，利用自注意力机制计算时间点的关联程度，得到自注意力分数 $\boldsymbol{A}_{\mathrm{P}}$，如式(20)所示。

$$\boldsymbol{A}_{\mathrm{P}}=\boldsymbol{Q}\times\frac{\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}}\times\boldsymbol{T}_{\mathrm{CM}} \tag{20}$$

式中：d_K 为 $\boldsymbol{V}$ 的维度。

最后，应用修正因子 $\boldsymbol{\lambda}$，使注意力集中于先前临近时间点，得到加权求和后的输出 $\boldsymbol{P}$，即先验关联，如式(21)所示。

$$\boldsymbol{P}=\mathrm{Softmax}(\boldsymbol{\lambda}\times\boldsymbol{A}_{\mathrm{P}})\times\boldsymbol{V} \tag{21}$$

2.3 基于双流粒度对齐的序列关联提取方法

正常量测数据在不同时间点之间存在长期稳定的全局信息关联，但 FDIA 攻击会通过注入虚假数据打破这一稳定性。为了提取量测数据点与整个量测数据序列之间的关联性，设计了一种基于双流粒度对齐的序列关联提取方法。该方法采用了一种组合式架构以提取序列关联：首先利用深度可分离卷积与多尺度卷积并行捕获局部及多粒度特征，然后通过交叉注意力与可学习查询注意力机制融合全局上下文信息，有效反映了 FDIA 对量测数据全局时间模式的破坏，结构如图 5 所示。

所提方法设计双分支结构，分别采用深度可分离卷积与多尺度卷积提取特征，前者通过轻量化计算捕捉高维量测数据的局部关联性，后者通过多时间跨度提取序列变化模式，共同增强模型对量测数据的多层次表征能力。

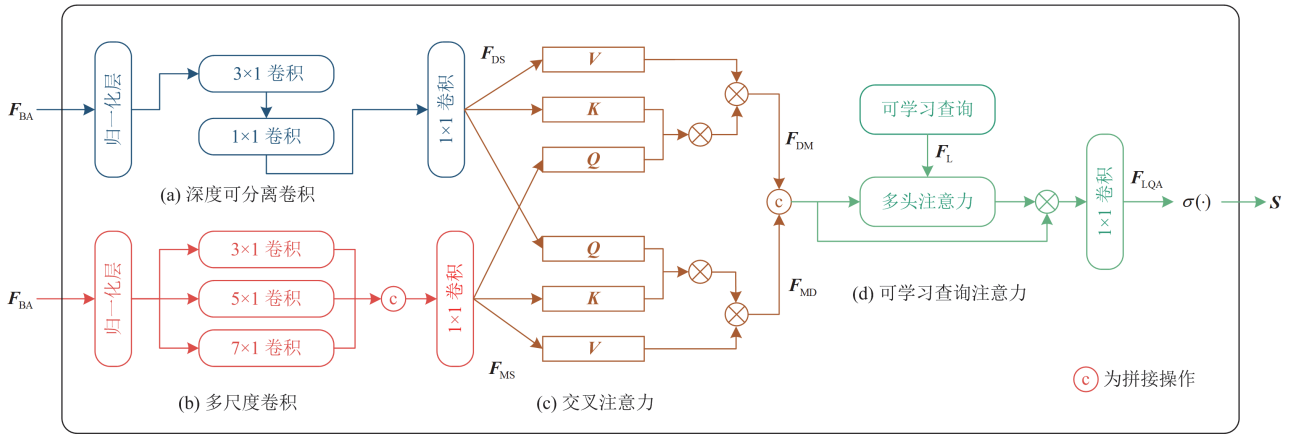


图 5 序列关联提取方法结构

Fig. 5 Structure of the sequential association extraction method

深度可分离卷积由逐通道的深度卷积和逐点卷积组成，具体过程如下所示。

首先，给定输入特征 $\boldsymbol{F}_{\mathrm{BA}}$，使用逐通道卷积和逐点卷积来编码 $\boldsymbol{F}_{\mathrm{BA}}$，其输出通过求和进行融合并发送到1×1卷积中。

$$\boldsymbol{F}_{\mathrm{DS}}=C_{1\times1}(C_{\mathrm{DP}}(C_{\mathrm{DC}}(N_{\mathrm{orm}}(\boldsymbol{F}_{\mathrm{BA}})))) \tag{22}$$

式中：$\boldsymbol{F}_{\mathrm{DS}}$ 为深度可分离卷积的输出；$C_{1\times1}(\cdot)$ 为一维卷积；$C_{\mathrm{DP}}(\cdot)$ 为逐点卷积，卷积核大小设置为1×1；$C_{\mathrm{DC}}(\cdot)$ 为逐通道卷积，卷积核大小设置为 3×1；$N_{\mathrm{orm}}(\cdot)$ 为归一化层，负责标准化输入数据。

多尺度卷积沿着同一空间维度并行编码多尺度上下文信息，过程如下。

给定输入特征 $\boldsymbol{F}_{\mathrm{BA}}$，使用 3 个并行的一维卷积来编码 $\boldsymbol{F}_{\mathrm{BA}}$，其输出通过求和进行融合并发送到1×1卷积中。

$$\boldsymbol{F}_{\mathrm{MS}}=C_{1\times1}(C_{\mathrm{AT}}(\sum_{i=1}^{3}C_{\mathrm{D}i}(N_{\mathrm{orm}}(\boldsymbol{F}_{\mathrm{BA}})))) \tag{23}$$

式中：$\boldsymbol{F}_{\mathrm{MS}}$ 为多尺度卷积的输出；$C_{\mathrm{AT}}(\cdot)$ 为特征拼接操作；$C_{\mathrm{D}i}(\cdot)$ 为一维卷积，卷积核大小分别设置为3×1、3×1和7×1。

其次，计算 $\boldsymbol{F}_{\mathrm{DS}}$ 和 $\boldsymbol{F}_{\mathrm{MS}}$ 之间的交叉注意力，在细粒度特征和多尺度特征间进行信息交换和整合。

具体来说，对于深度可分离卷积分支，将 $\boldsymbol{F}_{\mathrm{DS}}$ 作为键和值矩阵，将 $\boldsymbol{F}_{\mathrm{MS}}$ 视为查询矩阵。计算过程为

$$\boldsymbol{F}_{\mathrm{DM}} = C_{\mathrm{MS}}(\boldsymbol{F}_{\mathrm{MS}}, \boldsymbol{F}_{\mathrm{DS}}, \boldsymbol{F}_{\mathrm{DS}}) \tag{24}$$

式中：C_{MS} 为计算深度可分离卷积的交叉关注。

对于多尺度分离卷积分支，使用类似的方式计算交叉注意力，如式(25)所示。

$$\boldsymbol{F}_{\mathrm{MD}} = C_{\mathrm{DS}}(\boldsymbol{F}_{\mathrm{DS}}, \boldsymbol{F}_{\mathrm{MS}}, \boldsymbol{F}_{\mathrm{MS}}) \tag{25}$$

式中：C_{DS} 为计算多尺度卷积的交叉关注。

随后，引入可学习的查询向量，主动从输入序列中检索关键信息，得到融合后特征 $\boldsymbol{F}_{\mathrm{LQA}}$，该过程如式(26)所示。

$$\boldsymbol{F}_{\mathrm{LQA}} = C_{1\times 1}(L_{\mathrm{QA}}(\boldsymbol{F}_{\mathrm{L}}, C_{\mathrm{AT}}(\boldsymbol{F}_{\mathrm{DM}}, \boldsymbol{F}_{\mathrm{MD}}), C_{\mathrm{AT}}(\boldsymbol{F}_{\mathrm{DM}}, \boldsymbol{F}_{\mathrm{MD}}))) \tag{26}$$

式中：L_{QA} 为可学习查询注意力；$\boldsymbol{F}_{\mathrm{L}}$ 为可学习查询向量，旨在精准捕获任务相关特征，其参数会在训练过程中动态调整。

最后，经过 Sigmoid 激活函数对融合后的特征进行处理，得到序列关联矩阵 $\boldsymbol{S}$。

$$\boldsymbol{S} = \sigma(\left\|\boldsymbol{F}_{\mathrm{LQA}}\right\|) \tag{27}$$

2.4 基于关联差异的对抗性判别准则推理方法

为了计算关联差异，前文分别对先验关联和序列关联进行建模。然而，采样设备依赖电网的物理拓扑关系紧密连接，使正常量测数据在拓扑上呈现出高度一致性。FDIA 会破坏这种一致性，导致某些节点或量测点间的空间关系异常，因此引入多头自注意力机制对量测数据的空间关联进行提取。相比图注意力机制，多头自注意力并行学习节点关联，多视角建模空间关系，并基于输入动态生成拓扑关联权重矩阵，自适应表征量测节点间实时关联强度，即拓扑关联。

给定输入特征矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times d}$，$n$ 为输入图的节点个数，d 为输入的特征维度。类似于式(15)—式(17)，对节点 i、j 特征进行变换后得到对应的 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$，并计算 $\boldsymbol{Q}_i$ 和 $\boldsymbol{K}_j$ 的相似度 $\boldsymbol{A}_{\mathrm{T}}$。

$$\boldsymbol{A}_{\mathrm{T}} = \frac{\boldsymbol{Q}_i \times \boldsymbol{K}_j^{\mathrm{T}}}{\sqrt{d_K}} \tag{28}$$

然后，利用门控机制对多头注意力输出进行加权，控制不同注意力头的贡献，如式(29)所示。

$$\boldsymbol{A}_{\mathrm{gated}} = \frac{1}{n}\sum_{k=1}^{n} \sigma(\boldsymbol{W}_{\mathrm{gate},k}\boldsymbol{A}_{\mathrm{T}} + \boldsymbol{b}_{\mathrm{gate},k}) \tag{29}$$

式中：$\boldsymbol{A}_{\mathrm{gated}}$ 为加权处理后的相似度；$\boldsymbol{W}_{\mathrm{gate},k}$ 为第 k 个注意力头的权重，控制该头在最终输出中的贡献大小；$\boldsymbol{b}_{\mathrm{gate},k}$ 为第 k 个注意力头的偏置，对注意力头的输出进行调整。

相似度经过归一化，得到动态注意力权重 α_{ij} 为

$$\alpha_{ij} = \mathrm{Softmax}(\boldsymbol{A}_{\mathrm{gated}}) \tag{30}$$

通过上述过程，每个节点都可以得到与其相邻节点之间的动态注意力权重。将所有节点的动态注意力权重平均池化，得到该时刻的拓扑关联矩阵 $\boldsymbol{T}$。

$$\boldsymbol{T} = \frac{1}{n}\sum_{i=1}^{n} \sigma(\boldsymbol{W}_i \cdot \mathrm{Reshape}(\boldsymbol{\theta}) + \boldsymbol{b}_i) \tag{31}$$

式中：$\boldsymbol{W}_i$ 为第 i 个节点输出的权重，控制其在最终输出中的贡献大小；$\boldsymbol{\theta}$ 为动态注意力权重 α_{ij} 的组合矩阵；Reshape(·) 表示矩阵重塑为一维向量；$\boldsymbol{b}_i$ 为可学习的偏置，对第 i 个节点的输出进行调整。

根据提取到的 3 种关联，定义关联差异为先验关联和序列关联间的 JS 散度与拓扑关联的乘积，其衡量了不同状态下先验关联和序列关联间的相似度以及节点的空间关联强度。关联差异如式(32)所示。

$$A_{\mathrm{ssDis}}(\boldsymbol{P},\boldsymbol{S},\boldsymbol{T};\boldsymbol{X}) = \left[\frac{1}{2}\begin{pmatrix} K_{\mathrm{L}}(\boldsymbol{P}\,\|\,\dfrac{\boldsymbol{P}+\boldsymbol{S}}{2}) + \\ K_{\mathrm{L}}(\boldsymbol{S}\,\|\,\dfrac{\boldsymbol{P}+\boldsymbol{S}}{2}) \end{pmatrix} \otimes \boldsymbol{T}\right] \tag{32}$$

式中：$K_{\mathrm{L}}(\cdot\|\cdot)$ 为计算 $\boldsymbol{P}$ 和 $\boldsymbol{S}$ 的 KL 散度；$A_{\mathrm{ssDis}}(\boldsymbol{P},\boldsymbol{S},\boldsymbol{T};\boldsymbol{X}) \in \mathbb{R}^{n\times 1}$ 为输入序列 $\boldsymbol{X}$ 的逐点关联差异。攻击数据的 $A_{\mathrm{ssDis}}(\boldsymbol{P},\boldsymbol{S},\boldsymbol{T};\boldsymbol{X})$ 比正常数据小，因此是可区分的。

引入 BiGRU-AE 的重构损失优化模型，重构损失会迫使模型挖掘非相邻区域间的长期依赖关系，从而增大攻击数据的重构难度。输入序列 $\boldsymbol{X}$ 的损失函数 $\boldsymbol{L}$ 的表达式为

$$\begin{aligned} &\boldsymbol{L}(\hat{\boldsymbol{X}},\boldsymbol{P},\boldsymbol{S},\boldsymbol{T},\gamma;\boldsymbol{X}) = \\ &\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_{\mathrm{F}}^{2} - \gamma \times \left\|A_{\mathrm{ssDis}}(\boldsymbol{P},\boldsymbol{S},\boldsymbol{T};\boldsymbol{X})\right\|_{1} \end{aligned} \tag{33}$$

式中：$\hat{\boldsymbol{X}} \in \mathbb{R}^{n\times d}$ 为 $\boldsymbol{X}$ 的重构；γ 为用于权衡损失的变量；$\|\cdot\|_{\mathrm{F}}$ 表示计算 Frobenius 范数。

定义 $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ 的判别准则为

$$\boldsymbol{D}(\boldsymbol{X}) = \mathrm{Softmax}(-A_{\mathrm{ssDis}}(\boldsymbol{P},\boldsymbol{S},\boldsymbol{T};\boldsymbol{X})) \otimes \left[\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_{2}^{2}\right] \tag{34}$$

式中：$\boldsymbol{D}(\boldsymbol{X}) \in \mathbb{R}^{n\times 1}$ 表示 $\boldsymbol{X}$ 的逐点异常判据。使用判别准则作为检测模型分类决策的边界，以确定待检测样本的状态。

接着，利用对抗训练的思想[26]，将检测模型设计为生成器，并通过多次训练判别准则来最大化可区分性，从而实现检测阈值的动态优化。在训练过程中，生成器和判别器共同优化，确保判别准则能

够持续适应攻击幅值的变化。为了避免过拟合，训练过程中采用了早停策略和正则化手段，采用二元交叉熵函数 $\boldsymbol{L}_{\mathrm{D}}$ 作为判别器的损失函数。

$$\boldsymbol{L}_{\mathrm{D}}(y,\hat{y})=-\frac{1}{N}\sum_{l=1}^{N}[y_l\ln(\hat{y}_l)+(1-y_l)\ln(1-\hat{y}_l)] \quad (35)$$

式中：y_l 为第 l 条样本真实标签；$\hat{y}_l$ 为生成器对第 l 条样本的预测标签；N 为测试样本数量。

3 算例分析

3.1 数据说明

使用 IEEE14、IEEE30 和 IEEE118 总线系统生成仿真数据集，其中每个系统的线路拓扑及参数信息从 Matpower 获得。为了确保生成数据能够准确反映实际电力系统的运行场景，本研究采用纽约州 5 个月的真实负荷数据，将其按比例分配至各测试系统，并以 5 min 为固定间隔对负荷数据进行降采样，以模拟真实负荷的动态波动特性[27]。最终，生成一个包含 4 万条量测数据的样本集，所有数据均对应系统正常运行状态。

攻击样本通过对量测数据进行虚假数据注入获得。为了生成具有代表性的攻击样本，假设攻击者掌握部分电网拓扑信息，并根据系统的负荷分布和潮流特性随机选择关键节点或线路进行攻击，根据式(1)使用 PyPower 求解攻击向量。定义攻击幅值为攻击引起的状态量偏移程度，设置攻击幅值的范围为正常状态的 0.1%~2.5%，旨在反映不同强度的 FDIA。考虑到 FDIA 通常具有稀疏性，设定攻击持续时间为 15 个时间步长。

生成的量测数据中，标签“0”表示正常状态，“1”表示遭受攻击。为了模拟真实情况，数据集中正常样本与攻击样本的比例设置为 39:1，虽然数据存在一定的不平衡，但这种设置更符合实际场景中 FDIA 的稀有性。训练集和测试集的划分比例为 3:1，以保证模型训练与评估的合理性。

3.2 实验环境和对比方法

本文所提模型 AD 基于 PyCharm 202303、Python 3.8.18、和 Pytorch 1.13.0 深度学习环境开发。训练模型的 PC 配置如下：Intel(R) Xeon(R) Gold 6240 CPU @ 2.60 GHz、32 GB 内存。设置初始学习率为 0.0001，训练过程的批处理大小为 128，迭代次数为 10。

将 AD 与现有基于数据驱动的时序特性检测模型(CNN[28]和极限梯度提升决策树[29](extreme gradient boosting, XGBoost))、空间特征检测模型(GNN、GCN)以及基于时空相关性的 FDIA 检测模型(LSTM-GCN[30]、LSTM-AE[31])进行比较。

3.3 评价指标

通过准确率(Accuracy)、精度(Precision)、召回率(Recall)、F1 分数(F1-Score)和假阳性率(false positive rate, FPR)评估模型性能。其中，准确率衡量模型的整体检测表现；精度表示检测为攻击的样本中实际为攻击的占比，体现模型对攻击样本检测的准确性；召回率表示实际攻击样本中被成功检测出的比例，反映模型的识别能力；F1 分数兼顾精度与召回率的平衡；假阳性率反映模型的误报情况。准确率、精度、召回率和 F1 分数越大，表明模型性能越优；假阳性率越小，表明模型误报越少。以上指标表达式分别如式(36)—式(39)所示。

$$A_{\mathrm{ccuracy}}=\frac{T_{\mathrm{n}}+T_{\mathrm{p}}}{T_{\mathrm{n}}+F_{\mathrm{n}}+T_{\mathrm{p}}+F_{\mathrm{p}}} \quad (36)$$

$$P_{\mathrm{recision}}=\frac{T_{\mathrm{p}}}{T_{\mathrm{p}}+F_{\mathrm{p}}} \quad (37)$$

$$R_{\mathrm{ecall}}=\frac{T_{\mathrm{p}}}{F_{\mathrm{n}}+T_{\mathrm{p}}} \quad (38)$$

$$F_{\mathrm{PR}}=\frac{F_{\mathrm{p}}}{F_{\mathrm{p}}+F_{\mathrm{n}}} \quad (39)$$

式中：T_{n} 表示模型检测正确的无攻击样本数；T_{p} 表示模型检测正确的攻击样本数；F_{p} 表示模型检测为攻击而实际为无攻击的样本数；F_{n} 表示模型检测为无攻击而实际为攻击的样本数。

F1 分数为精度和召回率的调和平均数，其定义为

$$F_{\text{1-Score}}=2\times\frac{P_{\mathrm{recision}}\times R_{\mathrm{ecall}}}{P_{\mathrm{recision}}+R_{\mathrm{ecall}}} \quad (40)$$

此外，通过接受者操作特征(receiver operating characteristic, ROC)曲线展示样本真阳性率与假阳性率之间的关系，ROC 曲线下的面积越大，表示模型的检测性能越好。

3.4 实验结果分析

除鲁棒性实验外，设置每个实验环境测量噪声标准差为 2%，设置攻击幅值为正常状态的 2.5%。

3.4.1 检测模型性能对比

7 种 FDIA 检测模型在 IEEE14 总线系统下的性能如表 1 所示。由表 1 可以看出：与单独提取量测数据时间或空间特征的检测模型相比，LSTM-AE 综合利用了时空关系，取得了较好的效果，精度和召回率均超过 95%。AD 的精度和召回率超过 99%，这是由于 AD 通过针对每个时间点建模，根据 FDIA 的稀有性和局部集中性进行检测。

表 1 各模型在 IEEE14 系统中的性能比较

Table 1 Performance comparison of various models in the IEEE14 system

%

模型	$P_{recision}$	R_{ecall}	$F_{1\text{-}Score}$	$A_{ccuracy}$	F_{PR}
CNN	51.48	64.24	57.16	90.37	6.73
GNN	63.89	85.19	73.02	93.70	5.35
XGBoost	71.46	86.92	78.43	95.22	3.86
GCN	84.41	89.00	86.65	97.26	1.83
LSTM-GCN	88.65	86.81	87.72	97.57	1.23
LSTM-AE	96.38	98.73	97.54	99.50	0.41
AD	99.53	99.07	99.30	99.86	0.05

上述模型在 IEEE30 总线系统中的性能如表 2 所示。由表 2 可以看出：随着系统复杂度增加，各检测模型的表现都有所下降。然而，AD 的综合性能仍优于表现较好的 LSTM-AE。这是因为 LSTM-AE 计算复杂度较高，容易产生更多的误报，导致精度下降。与 IEEE14 总线系统相比，AD 的精度有所下降，但召回率有所提高。随着系统复杂度增加，更多正常样本被误判为攻击样本，导致精度降低。然而，AD 通过关注时间点的关联差异，提升了复杂系统中的攻击样本检测率和召回率。这归因于其序列关联提取方法有效提取细粒度与多尺度特征，能精准捕捉 FDIA 对全局时间模式的破坏。

表 2 各模型在 IEEE30 系统中的性能比较

Table 2 Performance comparison of various models in the IEEE30 system

%

模型	$P_{recision}$	R_{ecall}	$F_{1\text{-}Score}$	$A_{ccuracy}$	F_{PR}
CNN	42.65	57.06	48.81	88.03	8.53
GNN	42.15	53.13	47.00	88.02	8.10
XGBoost	61.26	77.43	68.40	92.85	5.44
GCN	71.06	73.03	72.03	94.33	3.31
LSTM-GCN	78.30	79.75	79.01	95.76	2.46
LSTM-AE	95.78	97.11	96.44	99.28	0.48
AD	98.17	99.54	98.85	99.77	0.21

在 IEEE118 总线系统中，7 种检测模型的性能表现出较大的波动，如表 3 所示。由于 IEEE118 系统规模大、拓扑复杂，增加了攻击面，检测变得困难。在大规模系统中，拓扑关联的提取至关重要，可有效反映 FDIA 对电网空间分布的破坏。通过对抗训练，AD 进一步放大数据的全局差异，使模型在高维数据场景下依然表现优异。

在 IEEE14 总线系统中，上述检测模型的 ROC 曲线如图 6 所示。可以看出：与其他模型相比，AD 的 ROC 曲线拐点更靠近点(0,1)，这得益于 AD 对 FDIA 局部特征的建模，可以有效识别稀有且局部化的攻击数据。

表 3 各模型在 IEEE118 系统中的性能比较

Table 3 Performance comparison of various models in the IEEE118 system

%

模型	$P_{recision}$	R_{ecall}	$F_{1\text{-}Score}$	$A_{ccuracy}$	F_{PR}
CNN	34.47	53.70	41.99	85.16	11.34
GNN	33.69	47.57	39.44	85.39	10.40
XGBoost	46.78	68.98	55.75	89.05	8.72
GCN	60.97	71.41	65.78	92.57	5.08
LSTM-GCN	74.53	78.59	76.51	95.17	2.98
LSTM-AE	91.29	96.99	94.05	98.77	1.03
AD	97.79	97.45	97.62	99.53	0.22

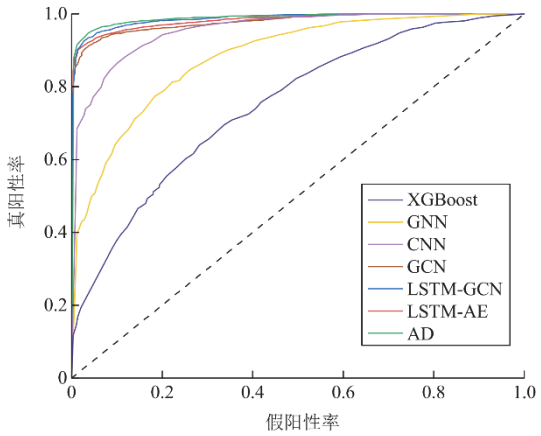


图 6 各模型在 IEEE14 系统中的 ROC 曲线

Fig. 6 ROC curves of various models in the IEEE14 system

3.4.2 模型运行时间对比

上述检测模型在不同总线系统下的运行时间如表 4—表 6 所示。可以看出：无论是训练收敛时间、单个 Epoch 的训练时间还是检测执行时间，AD 均优于其他检测模型。这是由于 AD 对每个时间点单独建模，仅关注当前时间点及其相关的上下文信息，大幅减少了数据量。

3.4.3 模型鲁棒性分析

选择数据采集噪声和攻击幅值来评估 AD 在不同环境下的鲁棒性。在 3 种总线系统中引入不同级别的噪声，噪声遵循均值为 0 且标准差在 0.02~0.10 内的正态分布，各模型的 F1 分数如图 7 所示。

表 4 各模型在 IEEE14 系统的运行时间

Table 4 Runtime of various models in the IEEE14 system

模型	训练收敛时间/s	单个 Epoch 训练时间(最小值)/s	检测执行时间/s
CNN	102	10	0.895
GNN	129	11.6	1.584
XGBoost	68	5.9	0.249
GCN	89	8.6	1.459
LSTM-GCN	137	12.9	1.856
LSTM-AE	176	15.6	1.699
AD	52	4.8	0.236

表 5 各模型在 IEEE30 系统的运行时间

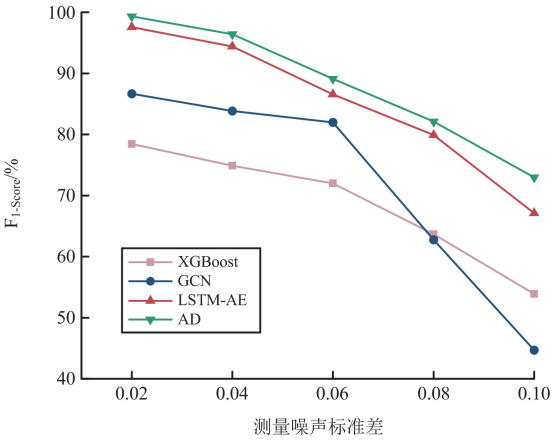
Table 5 Runtime of various models in the IEEE30 system

模型	训练收敛时间/s	单个 Epoch 训练时间(最小值)/s	检测执行时间/s
CNN	177	11.2	0.93
GNN	194	19	1.743
XGBoost	157	15.6	0.496
GCN	159	13.2	1.576
LSTM-GCN	202	18	2.606
LSTM-AE	242	22.8	2.096
AD	131	12.1	0.65

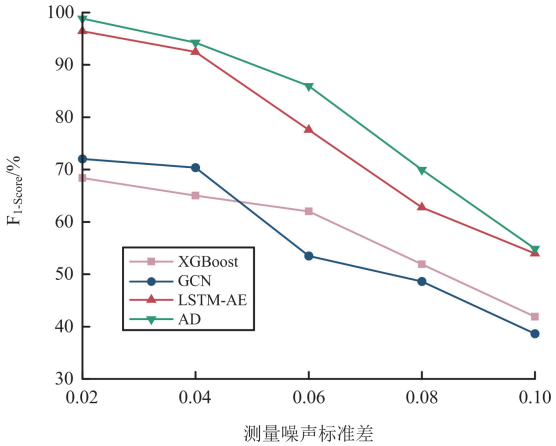
表 6 各模型在 IEEE118 系统的运行时间

Table 6 Runtime of various models in the IEEE118 system

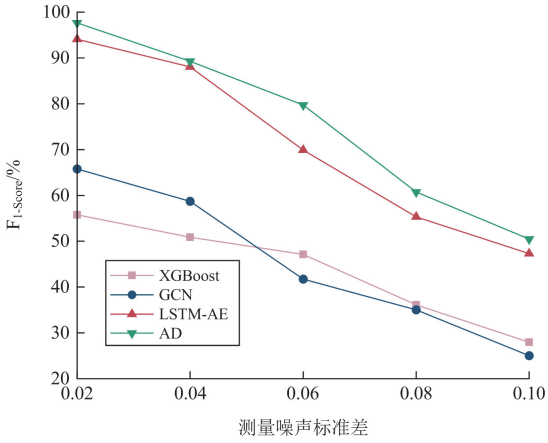
模型	训练收敛时间/s	单个 Epoch 训练时间(最小值)/s	检测执行时间/s
CNN	271	26.9	1.087
GNN	337	30.5	2.393
XGBoost	254	23.3	0.815
GCN	264	25.9	2.95
LSTM-GCN	386	37.7	2.484
LSTM-AE	381	33.1	2.392
AD	235	22.2	0.941



(a) IEEE14系统



(b) IEEE30系统



(c) IEEE118系统

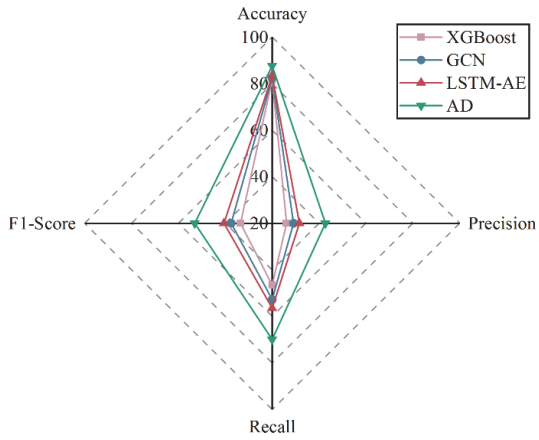
图 7 各模型在不同系统的 F1 分数

Fig. 7 F1-Score of various models in different systems

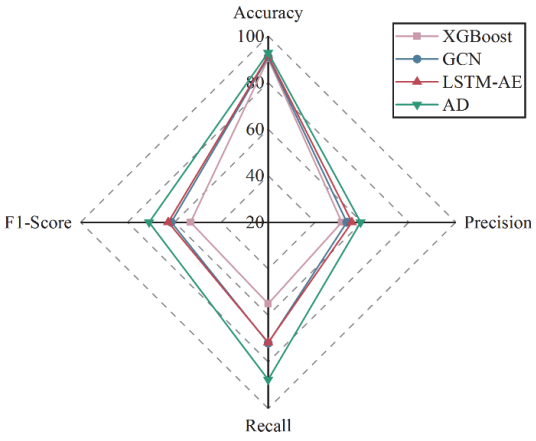
由图 7(a)可以看出：在 IEEE14 总线系统中，随着测量噪声标准差增加，XGBoost 和 GCN 的 F1 分数比 AD 下降更快，这表明 AD 的抗噪能力更强。由于该算法聚焦于当前时间点及其邻近局部时间点，将噪声的瞬时影响约束于局部时间范围内。XGBoost 对噪声较敏感，原因在于其基模型决策树的划分机制容易在训练过程中过度拟合噪声细节，导致模型泛化性能下降。

由图 7 可以看出：随着总线系统复杂度的增加，GCN 的性能优势逐渐减弱。这是因为 GCN 对图结构的依赖使其在系统拓扑结构复杂度提升时更易受到噪声干扰。相比之下，LSTM-AE 的 F1 分数下降较为平缓，得益于自动编码器通过数据降维和重构从噪声中提取有效特征，从而减小模型复杂度增加带来的负面影响。

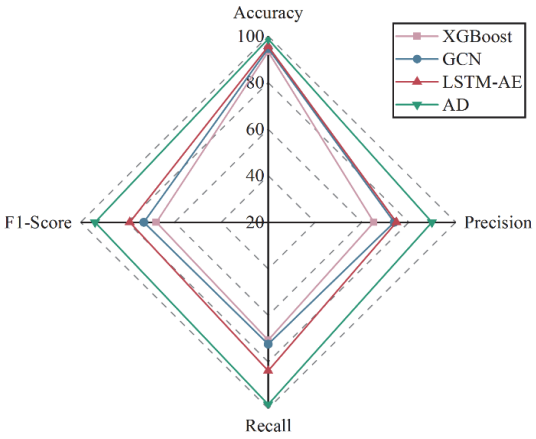
为检验 AD 模型在面对不同幅值攻击时的表现，在 IEEE14 总线系统中进行了攻击测试，各模型性能如图 8 所示。



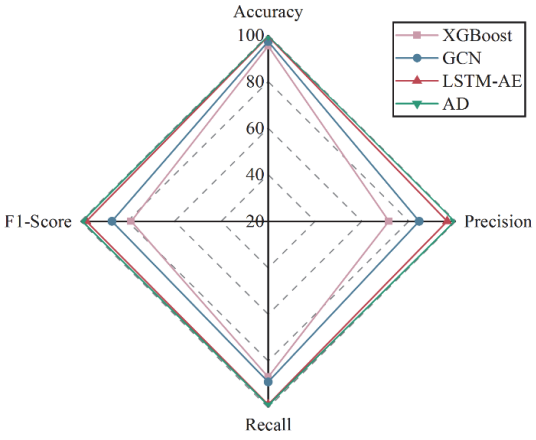
(a) 攻击幅值为 0.1%



(b) 攻击幅值为 0.5%



(c) 攻击幅值为 1%



(d) 攻击幅值为 2.5%

图 8 各检测模型在不同攻击幅值条件下的检测性能

Fig. 8 Detection performance of various models under different attack magnitudes

由图 8 可以看出：在低注入攻击幅值 0.1%时，各模型的检测效果均不理想。当注入攻击幅值增加到 0.5%时，3 种模型的性能都有明显改善，但 AD 的各项指标始终表现更优。

4 结论

针对电网 FDIA 攻击者调整注入攻击幅值导致检测模型精度下降的问题，本文从量测数据关联差异的角度提出了一种新型检测模型，将序列间固有的关联特性引入检测中。在 IEEE14、IEEE30 和 IEEE118 节点系统上进行仿真验证，得出以下结论。

1) AD 对每个时间点单独建模，规避了全局建模数据量大、计算复杂度高的问题，加速了模型收敛。

2) 与现有检测模型相比，AD 在准确率、精度和召回率等指标上表现更佳，并在不同噪声和攻击幅值情况下展现了较强的鲁棒性。

3) 即使在较低注入攻击幅值下，AD 仍能保持较高的检测精度，表明其在攻击者动态调整攻击幅值情况下具有较强的适用性。

该模型在性能上表现出色，在未来工作中，本文将探讨如何在检测到 FDIA 后定位具体受攻击位置，以抑制攻击影响的传播，提高电网整体安全性。

参考文献

[1] 席磊, 王艺晓, 何苗, 等. 基于反向鲸鱼-多隐层极限学习机的电网 FDIA 检测[J]. 中国电力, 2024, 57(9): 20-31.
XI Lei, WANG Yixiao, HE Miao, et al. FDIA detection in power grid based on opposition-based whale optimization algorithm and multi-layer extreme learning machine[J]. Electric Power, 2024, 57(9): 20-31.

[2] 庞清乐, 韩松易, 周泰, 等. 基于 ASRUKF 和 IMC 算法的电力信息物理系统虚假数据注入攻击检测[J]. 智慧电力, 2024, 52(7): 111-118.
PANG Qingle, HAN Songyi, ZHOU Tai, et al. False data injection attack detection of cyber-physical power system based on ASRUKF and IMC algorithms[J]. Smart Power, 2024, 52(7): 111-118.

[3] 蔡晔, 汤丽, 唐夏菲, 等. 基于 NSGA-II 的电力信息物理系统骨干网络辨识[J]. 电力系统自动化, 2023, 47(12): 38-46.
CAI Ye, TANG Li, TANG Xiafei, et al. Backbone network identification of cyber-physical power system based on non-dominated sorting genetic algorithm-II[J]. Automation of Electric Power Systems, 2023, 47(12): 38-46.

[4] 席磊, 白芳岩, 王文卓, 等. 基于海马优化深层极限学习机的电力信息物理系统 FDIA 检测[J]. 电力系统保护与控制, 2025, 53(4): 14-26.
XI Lei, BAI Fangyan, WANG Wenzhuo, et al. Cyber-physical power system FDIA detection based on seahorse optimized deep extreme learning machine[J]. Power System

Protection and Control, 2025, 53(4): 14-26.

[5] 席磊, 董璐, 程琛, 等. 基于混合黑猩猩优化极限学习机的电力信息物理系统虚假数据注入攻击定位检测[J]. 电力系统保护与控制, 2024, 52(14): 46-58.
XI Lei, DONG Lu, CHENG Chen, et al. Location detection of a false data injection attack in a cyber-physical power system based on a hybrid chimp optimized extreme learning machine[J]. Power System Protection and Control, 2024, 52(14): 46-58.

[6] 刘鑫蕊, 吴泽群. 面向智能电网的空间隐蔽型恶性数据注入攻击在线防御研究[J]. 中国电机工程学报, 2020, 40(8): 2546-2559.
LIU Xinrui, WU Zequn. Online defense research of spatial-hidden malicious data injection attacks in smart grid[J]. Proceedings of the CSEE, 2020, 40(8): 2546-2559.

[7] 黄冬梅, 王一帆, 胡安铎, 等. 融合无监督和有监督学习的虚假数据注入攻击检测[J]. 电力工程技术, 2024, 43(2): 134-141.
HUANG Dongmei, WANG Yifan, HU Anduo, et al. Detection method of false data injection attack based on unsupervised and supervised learning[J]. Electric Power Engineering Technology, 2024, 43(2): 134-141.

[8] 陈将宏, 饶佳黎, 李伟亮, 等. 基于向量自回归模型的电网虚假数据注入攻击检测[J]. 电力科学与技术学报, 2024, 39(3): 1-9.
CHEN Jianghong, RAO Jiali, LI Weiliang, et al. Detection method of false data injection attacks on power grids based on vector auto-regression model[J]. Journal of Electric Power Science and Technology, 2024, 39(3): 1-9.

[9] 谢云云, 严欣腾, 燕子敖, 等. 面向交直流混联电网的虚假数据注入攻击策略优化[J]. 电力工程技术, 2023, 42(4): 94-101.
XIE Yunyun, YAN Xinteng, YAN Zi'ao, et al. Strategy optimization of false data injection attack on AC-DC hybrid systems[J]. Electric Power Engineering Technology, 2023, 42(4): 94-101.

[10] 李卓, 谢耀滨, 吴茜琼, 等. 基于深度学习的电力系统虚假数据注入攻击检测综述[J]. 电力系统保护与控制, 2024, 52(19): 175-187.
LI Zhuo, XIE Yaobin, WU Qianqiong, et al. Review of deep learning-based false data injection attack detection in power systems[J]. Power System Protection and Control, 2024, 52(19): 175-187.

[11] PEI Chao, XIAO Yang, LIANG Wei, et al. A deviation-based detection method against false data injection attacks in smart grid[J]. IEEE Access, 2021, 9: 15499-15509.

[12] TIAN Jiwei, WANG Buhong, LI Jing, et al. Exploring targeted and stealthy false data injection attacks via adversarial machine learning[J]. IEEE Internet of Things Journal, 2022, 9(15): 14116-14125.

[13] ZHOU Tailin, XIAHOU Kaishun, ZHANG L L, et al. Real-time detection of cyber-physical false data injection attacks on power systems[J]. IEEE Transactions on Industrial Informatics, 2020, 17(10): 6810-6819.

[14] LI Boda, DING Tao, HUANG Can, et al. Detecting false data injection attacks against power system state estimation with fast go-decomposition approach[J]. IEEE Transactions on Industrial Informatics, 2018, 15(5): 2892-2904.

[15] 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): 8602-8622.
YANG Yuze, LIU Wenxia, LI Chengze, et al. Overview of FDIA detection methods for power SCADA systems[J]. Proceedings of the CSEE, 2023, 43(22): 8602-8622.

[16] WANG Yufeng, ZHANG Zhihao, MA Jianhua, et al. KFRNN: An effective false data injection attack detection in smart grid based on Kalman filter and recurrent neural network[J]. IEEE Internet of Things Journal, 2021, 9(9): 6893-6904.

[17] 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 detection method for interactive demand response[J]. Automation of Electric Power Systems, 2021, 45(3): 15-23.

[18] 陈冰, 唐永旺. 基于Bi-GRU和自注意力的智能电网虚假数据注入攻击检测[J]. 计算机应用与软件, 2021, 38(7): 339-344, 349.
CHEN Bing, TANG Yongwang. Detection of false data injection attacks in smart grid based on Bi-GRU and self attention[J]. Computer Applications and Software, 2021, 38(7): 339-344, 349.

[19] YAN Bingjing, JIANG Zhenze, YAO Pengchao, et al. Game theory based optimal defensive resources allocation with incomplete information in cyber-physical power systems against false data injection attacks[J]. Protection and Control of Modern Power Systems, 2024, 9(2): 115-127.

[20] 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, 2021, 13(1): 807-819.

[21] LI Yucheng, WANG Yuanyuan. Developing graphical detection techniques for maintaining state estimation integrity against false data injection attack in integrated electric cyber-physical system[J]. Journal of systems

architecture, 2020, 105.

[22] LI Xueping, WANG Yaokun, LU Zhiguang. Graph-based detection for false data injection attacks in power grid[J]. Energy, 2023, 263.

[23] 符杨, 张语涵, 田书欣, 等. 抵御多点虚假数据攻击的主动配电网状态估计方法[J]. 智慧电力, 2023, 51(4): 69-76, 83.
FU Yang, ZHANG Yuhan, TIAN Shuxin, et al. Active distribution network state estimation method to resist multi-point false data attacks[J]. Smart Power, 2023, 51(4): 69-76, 83.

[24] 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. Explanatable power grid false data injection attack detection method based on MGAT-TCN model[J]. Automation of Electric Power Systems, 2024, 48(2): 118-127.

[25] XU Jiehui, WU Haixu, WANG Jianmin, et al. Anomaly transformer: time series anomaly detection with association discrepancy[J]. arXiv preprint arXiv: 2110.02642, 2021.

[26] ZHANG Ying, WANG Jianhui, CHEN Bo. Detecting false data injection attacks in smart grids: a semi-supervised deep learning approach[J]. IEEE Transactions on Smart Grid, 2020, 12(1): 623-634.

[27] XU Wangkun, HIGGINS M, WANG Jianhong, et al. Blending data and physics against false data injection attack: an event-triggered moving target defence approach[J]. IEEE Transactions on Smart Grid, 2022, 14(4): 3176-3188.

[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] 刘鑫蕊, 常鹏, 孙秋野. 基于 XGBoost 和无迹卡尔曼滤波自适应混合预测的电网虚假数据注入攻击检测[J]. 中国电机工程学报, 2021, 41(16): 5462-5476.
LIU Xinrui, CHANG Peng, SUN Qiuye. Grid false data injection attacks detection based on XGBoost and unscented Kalman filter adaptive hybrid prediction[J]. Proceedings of the CSEE, 2021, 41(16): 5462-5476.

[30] LI Houjun, DOU Chunxia, YUE Dong, et al. End-edge-cloud collaboration-based false data injection attack detection in distribution networks[J]. IEEE Transactions on Industrial Informatics, 2023, 20(2): 1786-1797.

[31] MUSLEH A S, CHEN Guo, DONG Zhaoyang, et al. Attack detection in automatic generation control systems using LSTM-based stacked autoencoders[J]. IEEE Transactions on Industrial Informatics, 2022, 19(1): 153-165.

收稿日期：2024-12-18；　修回日期：2025-03-26

作者简介：

郭晓利(1968—)，女，硕士，教授，硕士生导师，研究方向为电力人工智能与大数据、电力系统安全防御；E-mail: 243589657@qq.com

王　月(2001—)，女，硕士研究生，研究方向为电力系统网络安全；E-mail: wangyue012024@163.com

李　斌(2000—)，男，通信作者，硕士研究生，研究方向为电力系统网络安全。E-mail: louis412613@163.com

（编辑　张　颖）