文章编号：1674-0629(2025)06-0026-13 中图分类号：TM73；TP393. 08 文献标志码：A
DOI：10. 13648/j. cnki. issn1674-0629. 2025. 06. 003

基于 Focal Loss^{IM}-Transformer 的电网虚假数据注入攻击检测

席磊[1,2]，和昀[1]，李子豪[1]，曹利锋[1]，李宗泽[1]，石雨凡[3]
（1. 三峡大学电气与新能源学院，湖北 宜昌 443002；2. 三峡大学梯级水电站运行与控制湖北省重点实验室，湖北 宜昌 443002；3. 国网山西省电力公司运城市经济技术开发区供电公司，山西 运城 044000）

摘要：虚假数据注入攻击对电力信息物理系统造成严重安全威胁。由于受到攻击样本与正常样本之间存在类别不平衡特性，导致机器学习检测方法偏向于多数类的预测，影响其对攻击的检测精度。为此，提出了基于 Focal Loss^{IM}-Transformer 的虚假数据注入攻击检测。Transformer 利用其自注意力机制能够捕捉数据中的长期依赖性，进而识别不平衡的虚假数据注入攻击数据。Focal Loss^{IM} 通过引入调制因子来更好地匹配虚假数据注入攻击样本的分布和特性，来增强检测方法对不平衡数据的识别能力，以提高检测方法对攻击的检测精度。通过在 IEEE 14 节点系统、IEEE 30 节点系统和 IEEE 57 节点系统进行仿真，验证了所提方法的有效性。相较于传统损失函数和其他检测方法，所提方法显示出更好的泛化能力和对少数类的识别能力，且辨识精度高、误报率低。

关键词：电力信息物理系统；虚假数据注入攻击；不平衡数据；Transformer；Focal Loss

False Data Injection Attack Detection Based on Focal Loss^{IM}-Transformer

XI Lei[1,2], HE Yun[1], LI Zihao[1], CAO Lifeng[1], LI Zongze[1], SHI Yufan[3]
（1. College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei 443002, China；2. Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang, Hubei 443002, China；3. Yuncheng Economic and Technological Development Zone Power Supply Company, State Grid Shanxi Electric Power Company,Yuncheng, Shanxi 044000, China）

Abstract: False data injection attacks pose serious security threats to cyber-physical power system. Due to the class imbalance property between the attacked samples and the normal samples, machine learning detection methods tend to predict the majority of classes, which affects their detection accuracy of the attacks. Therefore, a false data injection attack detection based on Focal Loss^{IM} Transformer is proposed. Transformer utilizes itself attention mechanism to capture long-term dependencies in data, thereby identifies imbalanced false data injection attack data. Focal Loss^{IM} enhances the detection method´s ability to identify imbalanced data by introducing modulation factors to better match the distribution and characteristics of false data injection attack samples, thereby improving the detection accuracy of the detection method for attacks. The effectiveness of the proposed method is verified through simulations on IEEE 14-node system, IEEE 30-node system, and IEEE 57-node system. Compared with traditional loss functions and other detection methods, the proposed method exhibits better generalization ability and recognition ability for minority classes, with high recognition accuracy and low false alarm rate.

Key words: power information physical system; false data injection attack; unbalanced data; Transformer; Focal Loss

基金项目：国家自然科学基金资助项目(52477104)；宜昌市自然科学研究项目(A23-2-001)。
Foundation item: Supported by the National Natural Science Foundation of China (52477104); the Natural Sciences Research Project of Yichang City (A23-2-001).

0　引言

随着信息化的发展，电网已发展成为信息技术与传统物理架构深度耦合的电力信息物理系统（cyber-physical power system，CPPS）[1−3]。该系统通过传感器、控制单元和执行器的协作，实时采集、分析数据并根据数据做出相应决策，从而提升电网的效率与灵活性。其中，通信网络作为信息技术中的重要组成部分，是CPPS中传感器、控制单元和执行器之间数据交换的关键渠道。然而，CPPS对通信网络的高度依赖，也使其暴露在网络攻击的目标范围之中。

虚假数据注入攻击[4−6]（false data injection attack，FDIA）作为一种隐蔽性强、破坏性大的网络攻击，对电网的安全稳定运行构成严重威胁[7]，对数据保护和实时监测提出了严峻挑战。FDIA通过精心篡改传感器数据，使被篡改的数据在形式上与正常数据高度相似，从而绕过传统的不良数据检测（bad data detection，BDD[8]）机制，进而误导电网工作人员做出错误决策。近年来，全球多个地区的电网遭遇过类似的网络攻击，其中2015年乌克兰电网事件尤为典型，攻击者成功篡改了电网的关键数据，导致电力供应中断并造成大规模的社会经济损失。尽管我国电网在系统架构和网络安全防护方面已采取了多层次的防护措施，但随着信息化的推进以及攻击手段的不断发展，传统的防护措施并不完全可靠。故研究CPPS中的FDIA的检测和防御方法具有重要意义。

目前，针对CPPS中的FDIA的检测方法主要分为两大类：基于模型驱动[9−12]的检测方法和基于数据驱动[13−15]的检测方法。基于模型驱动的方法需要建立精确的电力系统模型，通过对量测数据进行状态估计来辨识量测数据中的异常。文献[16]提出了一种基于向量自回归（vector autoregression，VAR）模型和残差分析方法，有效地检测了CPPS中的FDIA。文献[17]利用图论增强电力系统交流状态估计以实现FDIA的检测。然而，由于CPPS的复杂性和非线性特性，构建精确的CPPS模型变得愈发困难，从而导致基于模型驱动的方法难以实施。

基于数据驱动的方法则摒弃了复杂的建模过程，仅需从海量的电网数据中挖掘CPPS各尺度上的关联性，便可实现对FDIA的检测。文献[18]提出了一种结合反向学习的鲸鱼优化多隐层极限学习机的FDIA检测方法，解决了无法精确获取受攻击位置的问题。文献[19]提出利用长短期记忆（long short-term memory，LSTM）网络所具有的时间序列预测方面的强大记忆能力来识别潜在的受损测量值，解决了电力系统中识别受损测量值的难题。文献[20]提出了利用卷积神经网络（convolutional neural network，CNN）对高维历史量测数据进行高效特征提取及分类，解决了高维数据处理和分类的难题。然而，由于受到攻击的样本与正常样本之间存在类别不平衡特性，导致基于数据驱动的机器学习检测方法偏向于多数类的预测，影响了其对攻击的检测精度。此外，FDIA攻击模式可能随技术演进或防御策略升级动态变化[21]，这类别不断积累的特性要求检测模型在无需完全重新训练的情况下增量学习新类别特征，同时避免对历史知识的遗忘，这对电网系统的实时监控尤为重要。

Transformer算法[22−23]借助其自注意力机制提高了对输入序列的处理能力，使算法能全面关注序列中各位置的信息并分析位置间的依赖关系，能够识别不平衡FDIA数据。然而Transformer常用的二元交叉熵损失函数（binary cross-entropy，BCE）对少数类样本的识别能力仍需提高。Focal Loss损失函数[24]通过调整损失函数中的焦点，使算法更加关注那些难以正确分类的少数类样本。然而，Focal Loss在其标准形式中少量的可调节参数难以完全适应FDIA样本的特异性和复杂性。为此本文在原有的Focal Loss函数的基础上进行了改进，通过改变调制因子以获得额外的可调整参数，从而形成了一种Focal Loss变体函数，即Focal Loss^{IM}，来调整损失函数对于不平衡样本的损失贡献。Focal Loss^{IM}被集成到Transformer算法中以增强算法对难以分类的少数类样本的关注能力，提升了对不平衡FDIA量测数据的识别性能。

综上，本文提出基于Focal Loss^{IM}-Transformer的虚假数据注入攻击检测。Transformer利用其自注意力机制能够捕捉数据中的长期依赖性，进而识别不平衡的虚假数据注入攻击数据；Focal Loss^{IM}通过引入调制因子来更好地匹配虚假数据注入攻击样本的分布和特性，增强了检测方法对不平衡数据的识别能力，提高了检测方法对攻击的检测精度。所提方法从算法架构和损失函数调整两个层面解决了数

据不平衡问题，在IEEE 14节点系统、30节点系统以及57节点系统上进行的仿真实验验证了该方法的有效性。

1 FDIA

1.1 状态估计及不良数检测

状态估计[25-26]是电力系统监控与控制的核心，通过数据采集与监视控制系统(supervisory control and data acquisition，SCADA)[27]和相量测量单元(power management unit，PMU)[28]收集的量测数据来估计当前电网的运行状态，其公式表述为：$\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{x})+\boldsymbol{e}$，其中$\boldsymbol{z}=[z_1, z_2, \cdots, z_m]^{\mathrm{T}}$为测量向量，包括母线电压、有功功率和无功功率注入，以及支路有功潮流和无功潮流等；$\boldsymbol{x}=[x_1, x_2, \cdots, x_n]^{\mathrm{T}}$为系统状态变量；$m$和$n$分别为测量的数量和系统状态数量，且两者之间有一定冗余度，即$m>n$；$\boldsymbol{h}(\cdot)$为量测值与系统状态之间的非线性映射，通常取决于电网节点的电压幅度和相位角；$\boldsymbol{e}=[e_1, e_2, \cdots, e_m]^{\mathrm{T}}$为测量误差向量。

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

在此基础上，采用线性加权最小二乘法(weighted least squares，WLS)得到状态估计的目标函数$J(\boldsymbol{x})$，$\boldsymbol{R}$为测量误差的协方差矩阵，其逆矩阵$\boldsymbol{R}^{-1}$用于对不同的测量误差赋予相应的权重。通过求解以上式子可得到系统状态向量估计值$\hat{\boldsymbol{x}}$。

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

为确保状态估计结果的准确性和可靠性，不良数据检测技术[29-30]被广泛采用，其中基于残差向量的方法应用最为广泛。该方法定义残差r为实际测量值$\boldsymbol{z}$与通过状态估计模型得到的预测值$\boldsymbol{h}(\boldsymbol{x})$之间的差异，即$r=\left\|\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right\|$。在此基础上，通过$r$与预设阈值$\tau$进行比较，有效鉴别异常数据：如果$r>\tau$，检测机制将判断SCADA系统的量测数据中含有错误数据；$r<\tau$，则判断数据正常。

1.2 FDIA原理

FDIA是一种针对状态估计过程的安全威胁，攻击者通过向信息系统中的量测单元发起攻击，篡改或伪造传感器数据，向控制系统传递虚假的电网状态信息。这些虚假信息被控制系统错误地用于状态估计和决策支持，导致调度中心做出错误的调度决策或负荷分配，最终危害电网的安全运行。针对CPPS的FDIA示意图如图1所示。

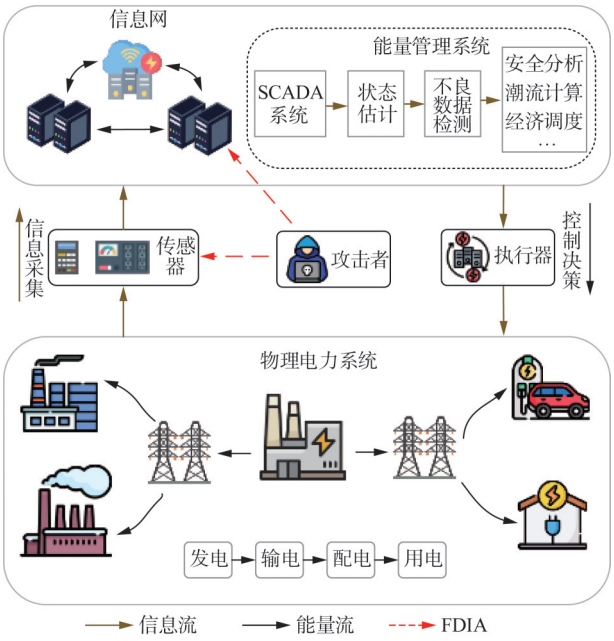

图1 针对CPPS的FDIA示意图

Fig. 1 Sketch diagram of FDIA for CPPS

攻击者利用系统拓扑和状态估计算法设计攻击向量$\boldsymbol{a}=\boldsymbol{Hc}$其中$\boldsymbol{H}$为系统的拓扑信息矩阵，$\boldsymbol{c}$为状态变量的攻击注入向量。攻击者通过选择适当的攻击向量使得篡改后的测量向量$\boldsymbol{z}_{\mathrm{m}}$为：$\boldsymbol{z}_{\mathrm{m}}=\boldsymbol{z}+\boldsymbol{a}$。在此攻击模式下，系统受到攻击后的残差$r_{\mathrm{m}}$变为：

$$\begin{aligned} r_{\mathrm{m}}&=\boldsymbol{z}_{\mathrm{m}}-\boldsymbol{h}(\hat{\boldsymbol{x}})=(\boldsymbol{z}+\boldsymbol{a})-\boldsymbol{h}(\boldsymbol{x}+\boldsymbol{c})\\ &=\boldsymbol{z}+\boldsymbol{Hc}-\boldsymbol{h}(\boldsymbol{x}+\boldsymbol{c})\\ &=r+\boldsymbol{Hc}-(\boldsymbol{h}(\boldsymbol{x}+\boldsymbol{c})-\boldsymbol{h}(\boldsymbol{x})) \end{aligned} \tag{3}$$

函数$\boldsymbol{h}$是关于$\boldsymbol{x}$的近似线性，即$\boldsymbol{h}(\boldsymbol{x}+\boldsymbol{c})\approx\boldsymbol{h}(\boldsymbol{x})+\boldsymbol{Hc}$，则攻击引起的额外项$\boldsymbol{Hc}-\boldsymbol{h}(\boldsymbol{x}+\boldsymbol{c})+\boldsymbol{h}(\boldsymbol{x})$在数值上很小或接近于0，导致$r_{\mathrm{m}}$近似等于$r$。攻击者通过精心设计的攻击向量$\boldsymbol{a}$在不显著改变残差值的情况下可成功绕过基于残差的BDD检测。

1.3 FDIA模型构建

在实际情况下FDIA受多种限制，如攻击者的资源限制和电力系统的防护措施。在资源受限时FDIA通常表现为攻击向量的稀疏性，即只篡改部分测量点以逃避监控。这种策略减少了攻击者需操控的数据点数量，降低了攻击成本，同时保持了攻击的有效性。具体公式如下。

$$\begin{cases}\alpha_i=\min_{\mathrm{c}}\left\|\boldsymbol{Hc}\right\|\\ \text{s.t.}\quad a_i=\boldsymbol{Hc}=1\end{cases} \tag{4}$$

式中：α_i 为被篡改的传感器数量，其中目标函数 $\|\boldsymbol{Hc}\|$ 的最小化表示了寻求最少的非零元素，体现了向量的稀疏性。约束条件 $a_i = \boldsymbol{Hc} = 1$ 确保了至少有一个测量点 i 受到攻击，满足攻击基本要求。

基于上述现实限制提出的 FDIA 模型旨在最小化攻击成本，同时通过简化攻击目标数量提高攻击的隐蔽性。本文针对电力系统中的单个节点发起攻击，旨在通过改变该节点的测量数据来误导系统的状态估计，而不引起系统其他部分的异常。确保至少有一个量测点被攻击，减少被篡改的量测点数量，其主要过程分为两个阶段进行。

攻击准备阶段：确定攻击区域中各个节点的电压幅值 U_m 和相位角 U_a、支路上的有功功率 P 和无功功率 Q，以及发电机注入的有功功率 P_{Gi} 和无功功率 Q_{Gi}。

$$P_i = V_i \sum_{j=1}^{n} V_j \left(G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}\right) \tag{5}$$

$$Q_i = V_i \sum_{j=1}^{n} V_j \left(G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}\right) \tag{6}$$

$$P_{Gi} = P_G - P_i \tag{7}$$

$$Q_{Gi} = Q_G - Q_i \tag{8}$$

式中：P_i、Q_i 分别为节点 i 的有功功率和无功功率；V_i 和 V_j 分别为节点 i 和 j 的电压幅值；θ_{ij} 为节点 i 和 j 的电压相角差；G_{ij} 和 B_{ij} 分别为支路电导和电纳；P_G 和 Q_G 分别为发电机的有功功率和无功功率。

攻击阶段：攻击者通过系统拓扑矩阵 $\boldsymbol{H}$ 构建攻击向量 $\boldsymbol{a} = \boldsymbol{Hc}$，该攻击向量 $\boldsymbol{z}_{attack}$ 会被映射到受影响的量测数据上，即 $\boldsymbol{z}_{attack} = \boldsymbol{z} + \boldsymbol{a}$，从而在不触发任何警报的情况下替代原始数据。攻击的有效性和隐蔽性将在第 3 节中验证。

2　Focal Loss^{IM}-Transformer 检测模型框架

本文提出的 FDIA 检测模型框架基于 Transformer 架构[22]，主要包括 3 个组成部分：嵌入层、编码层和输出层。其中，编码层是模型的核心，由以下 4 个关键组件构成：多头注意力机制、前馈全连接网络、残差连接及归一化层。其中，Transformer 架构的自注意力机制天然具备捕捉动态时序关联的能力，其全局依赖建模特性可为动态新增攻击类别的检测提供潜在支持。如图 2 所示。

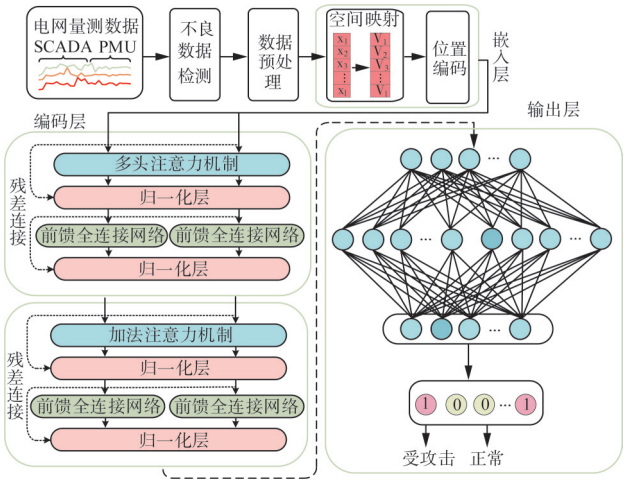

图 2　基于 Focal Loss^{IM}-Transformer 的 FDIA 检测模型框架

Fig. 2　Framework of Focal Loss^{IM}-Transformer based FDIA detection model

2.1　嵌入层

电力数据通常具有显著的序列特征，这些特征不仅反映了当前时刻的测量值，还与前后时刻的数据密切相关。为了有效处理这些序列特征，本文通过嵌入层将输入的电力数据映射到高维空间。在嵌入层中数据的每个时刻会被转换为一个向量，表示该时刻的数据特征。这一转换不仅保留了数据的数值信息，还通过位置编码注入了每个数据点的时间位置信息，从而使模型能够理解数据的序列关系。通过这种方式，嵌入层能够捕捉到输入数据的时序依赖性，并为后续的模型处理提供充分的时序信息。

假设输入的 FDIA 数据样本是特征向量，分别为 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、$\boldsymbol{x}_3$、$\boldsymbol{x}_4$、…、$\boldsymbol{x}_i$，其中 $\boldsymbol{x}_i$ 为第 i 个特征向量。嵌入层采用全连接网络直接处理这些特征向量，将原始特征嵌入到空间映射，该过程通过公式 $\boldsymbol{V} = \boldsymbol{wx} + \boldsymbol{b}$ 完成，其中 $\boldsymbol{w}$ 为权重矩阵，定义为原始特征空间到嵌入空间的映射；$\boldsymbol{b}$ 为偏置向量；$\boldsymbol{V}$ 为转换后的嵌入向量，它将作为 Transformer 模型输入的一部分。由于 Transformer 模型依赖于位置编码来理解序列中元素的位置信息，对于 FDIA 数据还需要对嵌入向量加入位置编码，位置编码（positional encoding，PE）通过正弦函数为序列中每个位置生成独特的编码向量，公式如下。

$$P_{PE}(p_{pos}, 2i) = \sin\left(\frac{p_{pos}}{10\,000^{2i/d_{model}}}\right) \tag{9}$$

$$P_{\mathrm{PE}}(p_{\mathrm{pos}}, 2i+1)=\cos\left(\frac{p_{\mathrm{pos}}}{10\,000^{2i/d_{\mathrm{model}}}}\right) \tag{10}$$

式中：$P_{\mathrm{PE}}(\cdot)$为位置编码函数；p_{pos}为序列中元素的位置索引；i为向量中的维度索引；d_{model}为模型中向量的维度，最终的输入向量是嵌入向量和位置编码的和：$\boldsymbol{V}'=\boldsymbol{V}+p_{\mathrm{pos}}$。确保每个特征向量包含原始特征的信息，并且融入相应位置信息，使得Transformer模型能够处理具有序列特性的FDIA数据。

2.2 编码层

2.2.1 多头注意力机制

传统的注意力机制通常只能关注到序列中的局部特征，而电力数据中各类测量值，如电压、电流、有功功率、无功功率之间的关系复杂多样，单一的注意力头无法同时关注到这些特征之间的多维依赖性。每个注意力头聚焦于不同特征子空间，如一些注意力头关注时间序列中的趋势和周期性变化，而其他注意力头则关注节点间的协同变化和特征之间的细微依赖。通过这种方式模型能够捕捉到细微的变化，增强了对攻击信号的敏感度。具体计算流程如图3所示。

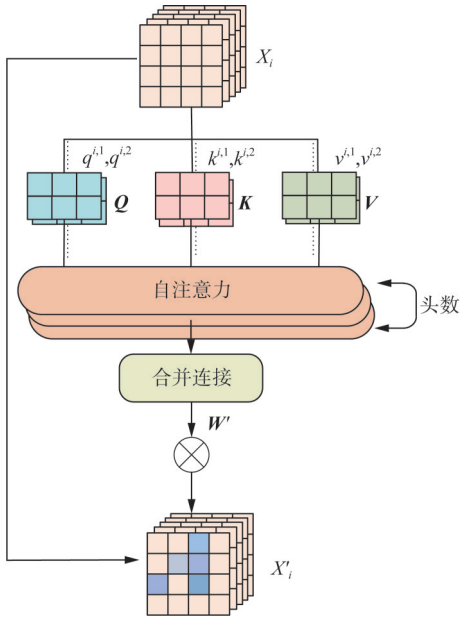

图3 多头注意力机制流程图

Fig. 3 Flowchart of the multi-head attention mechanism

对于序列中的每个数据点，模型通过3组权重矩阵转换为查询(query)、键(key)和值(value)。模型通过计算矩阵$\boldsymbol{Q}$与$\boldsymbol{K}$的点积，得到表示数据点间关系的注意力分数。通过softmax函数对注意力分数进行归一化后，这些分数用于对$\boldsymbol{V}$进行加权求和，得到每个数据点在全序列信息下新的表示，此表示为整个序列的信息，极大地增强了模型对受攻击和未受攻击序列之间差异的捕捉能力，公式表示如下。

$$\begin{aligned}H_i&=\mathrm{Attention}(Q_i,K_i,V_i)\\&=\mathrm{softmax}\left(\frac{Q_iK_i^T}{\sqrt{d_k}}\right)V_i\end{aligned} \tag{11}$$

$$\mathrm{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(H_1,H_2,\cdots,H_n)\boldsymbol{W}^{\mathrm{o}} \tag{12}$$

式中：$\mathrm{Attention}(Q_i, K_i, V_i)$为注意力函数，$Q_i$、$K_i$、$V_i$分别为第$i$个注意力头的查询、键、值矩阵线性映射；$\mathrm{MultiHead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$为多头函数，$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$分别为输入序列的全局查询矩阵、全局键矩阵和全局值矩阵；$\mathrm{Concat}(H_1, H_2, \cdots, H_n)$为沿特征维度拼接所有头的输出，$H_n$为第$n$个注意力头的输出；$\boldsymbol{W}^0$为输出层的权重矩阵；$d_k$为键向量的维度，其在注意力得分计算中作为缩放因子，防止了点积运算的结果过大，从而减轻了梯度消失问题。多头注意力机制通过全局依赖关系建模能够捕捉输入序列中隐含的通用攻击模式特征，如电压突变、相位偏移等，而非依赖特定攻击类别的先验知识。这种特性使模型在面对未知攻击类别时仍可通过特征相似性进行检测，从而提高了对FDIA攻击的感知能力。

2.2.2 前馈全连接网络

在自注意力层提取序列内部关系后，前馈全连接网络(feedforward neural network，FFN)对这些输出进行进一步处理。在FDIA检测的场景中，FFN层增加了模型对非线性特征的捕获能力，FDIA所注入的假数据可能会造成数据的非线性和复杂模式，仅通过线性处理难以区分正常与异常数据。前馈全连接网络通过ReLU激活函数引入非线性，使模型能够学习更加复杂的函数映射，从而区分细微的异常模式，计算公式表示为：

$$\mathrm{FFN}(\boldsymbol{x})=\mathrm{Relu}(0,\boldsymbol{x}\boldsymbol{W}_1+5)\boldsymbol{W}_2+6 \tag{13}$$

式中：$\mathrm{FFN}(\boldsymbol{x})$为前馈全连接网络函数，$\boldsymbol{x}$为从自注意力层传入的输出；$\boldsymbol{W}_1$和$\boldsymbol{W}_2$为权重矩阵；$b_1$和$b_2$为偏置项；$\mathrm{Relu}(\cdot)$为ReLU激活函数，它对每个元素逐个应用，为正的输入返回输入值，为负的输入

返回 0。FFN 层的输出是序列每个位置的新特征表示，包含了自注意力层捕获的序列内部依赖性和其学习到的复杂数据模式。FFN 层的引入不仅能处理攻击者可能引入的各种非线性模式，还能通过非线性变换捕捉到更深层次的数据特征。

2.2.3　归一化层及残差连接

每个自注意力和前馈网络单元的输出后面均设计有残差连接和归一化层，用于缓解深度网络训练中的梯度消失问题，并且加速训练过程提高模型的泛化能力。前馈全连接网络的输出与输入数据 x 进行元素级相加，从而形成残差连接 $x + \mathrm{Sublayer}(x)$，有利于信息在模型中的有效流动，并使得梯度可以直接反向传播至较深的层次，减少训练中信息丢失的风险，紧接着残差连接之后的是层归一化(layer normalization)，其公式为：

$$\mathrm{LayerNorm}(x + \mathrm{Sublayer}(x)) \tag{14}$$

式中：LayerNorm(·)为层归一化函数；Sublayer(x)为子层本身实现的函数。

归一化层通过计算输入的均值和方差对输入进行缩放和平移变换，使得网络中的每一层都有相似的数据分布，提高网络的训练效率和稳定性。通过残差连接和归一化层的策略构建更深的网络模型来捕获不平衡 FDIA 攻击样本数据，同时避免训练过程中可能遇到的数值不稳定问题。

2.3　输出层

输出层接收来自最后一个 Transformer 编码层的特征表示后通过 softmax 函数计算每个类别的概率，形成最终的分类决策。设 $\boldsymbol{z}$ 表示编码层输出的特征向量，输出层通过权重矩阵 $\boldsymbol{W}$ 和偏置矩阵 $\boldsymbol{b}$ 将 $\boldsymbol{z}$ 转换为对应类别的原始得分 $\boldsymbol{z}'$，公式如下。

$$\boldsymbol{z}' = \boldsymbol{W}\boldsymbol{z} + \boldsymbol{b} \tag{15}$$

随后 softmax 函数应用于 $\boldsymbol{z}'$，转换原始得分为概率分布，其中每个元素 p_i 对应于数据样本属于类别 i 的概率，公式如下。

$$\boldsymbol{p}_i = \mathrm{softmax}(\boldsymbol{z}') = \frac{\exp(z_i')}{\sum_j \exp(z_j')} \tag{16}$$

式中 z_i' 和 z_j' 分别为类别 i 和 j 的原始得分。

输出层会产生两个概率值，一个对应于“受攻击”的类别，另一个对应于“未受攻击”的类别。得益于编码层的深层特征学习，输出层提供一个明确的分类机制，通过概率输出给出模型的置信度。

2.4　Focal Loss^{IM} 损失函数

Focal Loss 损失函数为解决在检测任务中的类别不平衡问题。在目标检测方法中，通常会出现大量的简单负样本占据主导地位，从而压制少数的正样本的学习效果，导致模型难以学习到更有区分性的特征来改善检测性能。Focal Loss 是对传统的二元交叉熵损失函数 BCE 的改进，传统的 BCE 表达式为：

$$B_{\mathrm{CE}}(p,\ y) = \begin{cases} -\log(p), & y = 1 \\ -\log(1-p), & \text{其他} \end{cases} \tag{17}$$

式中：$B_{\mathrm{CE}}(p,\ y)$ 为二元交叉熵损失函数；y 为真实标签；p 为模型预测为正类的概率。当 $y = 1$ 时，损失为 $-\log(p)$；当 $y = 0$ 时，损失为 $-\log(1-p)$。BCE 损失函数在处理正负样本时是平等对待的，导致检测精准度不佳。

Focal Loss 损失函数则通过降低易分类样本的权重，使得模型的训练更加关注于难以区分和少数的样本，从而解决了类别不平衡问题。其核心在于引入一个调制因子，这个因子随着样本被正确分类的概率的增加而增大损失的衰减。式(18)为该调制因子的数学表达式[23]。

$$\mathrm{Focal\ Loss}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{18}$$

式中：p_t 为模型对于实际类别 y 的预测概率，当 $y = 1$ 时 $p_t = p$，$y \neq 1$ 时 $p_t = 1 - p$；α_t 为平衡正负样本的权重系数，用于减少类别不平衡对模型的影响，γ 是调节模型关注程度的超参数，用于减少易分类样本对损失函数的贡献，使模型更加关注难分类样本。当 $\gamma = 0$ 时，Focal Loss 将退化为普通的交叉熵损失。

然而对于 Focal Loss 函数虽然调整 α_t 和 γ 是优化 Focal Loss 以提升模型性能的有效手段，但它们存在一定的局限性和性能上限，仅调整这两个参数难以达到最优性能。本文通过修改调制因子产生 Focal Loss 的变体函数 $f^{\mathrm{IM}}_{\mathrm{FocalLoss}}$，引入更多可调整的参数，来调整损失函数对于不同样本的损失贡献。以优化 Transformer 模型对 FDIA 检测的性能进行更加精细的调整。提高整体模型对 FDIA 数据识别的性能。产生变体函数如下。

$$f^{\mathrm{IM}}_{\mathrm{FocalLoss}} = \alpha_t/(1+\beta(1-p_t)^{\gamma})\log(p_t) \tag{19}$$

式中 β 为概率调整的权重系数。Focal Loss^{IM} 采用分数形式的调制因子 $\alpha_t/(1+\beta(1-p_t)^{\gamma})$ 调整损失函

数对原始的BCE Loss进行放缩。Focal Loss^{IM}通过一个随p_t的增加而递减的调制因子对α_t进行调整，减少损失函数对于容易分类样本的惩罚。

当模型预测的概率$p_t \approx 1$时对于那些易于正确分类的样本，损失有所下降，但在$1+\beta(1-p_t)^{\gamma}$的作用下，α_t减少的速度将趋于平稳，有助于防止损失值降至极低，从而减轻模型过拟合的风险。当模型预测的概率$p_t \approx 0$时，$(1-p_t)^{\gamma} \approx 1$，对于难以分类的样本损失减少的速度更为缓慢，此时的损失值几乎不受影响，模型在训练过程中对这些难分类样本给予足够的重视。Focal Loss^{IM}损失函数使得模型在提高其泛化能力和减少对训练数据过度拟合方面取得平衡，确保难分类样本在模型训练过程中获得足够的关注。

为了进一步展示Focal Loss^{IM}在类别不平衡问题中的有效性，本文绘制了$1+\beta(1-p_t)^{\gamma}$随p_t的变化曲线。图4展示了不同γ和β值下调制因子如何随着分类概率的变化而调整样本的损失权重。

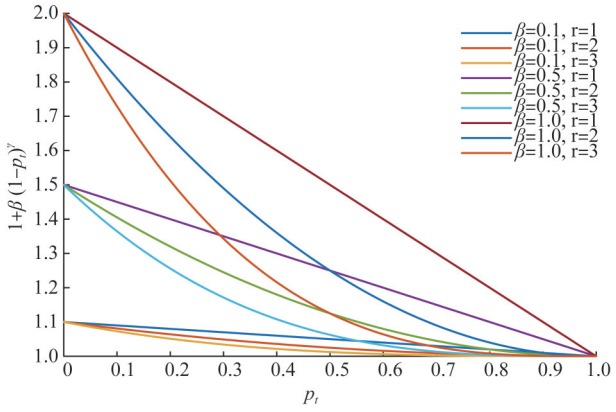

图4 $1+\beta(1-p_t)^{\gamma}$随p_t的变化曲线

Fig. 4 Curves of $1+\beta(1-p_t)^{\gamma}$with respect to the variations of p_t

通过调整β和γ能够控制调制因子的衰减速度，进而调整模型在训练过程中对类别不平衡问题的应对策略。当γ和β增加时调制因子对少数类样本的关注增强，模型能够有效地改善类别不平衡问题下的性能，尤其是在样本类别分布严重不平衡时，Focal Loss^{IM}显著提升了对少数类样本的检测能力。

3 仿真研究

3.1 实验数据生成

在MATLAB环境下，使用Matpower对IEEE 14节点系统、IEEE 30节点系统和IEEE 57节点系统进行仿真实验，以生成数据样本。这些节点系统的网络拓扑、节点数据和支路参数均从来Matpower获取。仿真中使用的负荷数据来自某地市调度中心，用于模拟虚假数据注入攻击。对于每个测试系统，各自生成了15 100个样本数据，其中包括1 500个受攻击样本，从而在总样本中形成1∶10的不平衡比例，即受攻击样本与全部样本的比例，构造出3个不平衡的数据集。

本文通过模拟仿真为电力系统构建FDIA攻击的前提条件，收集的电力负荷数据展示了特定电网在各个区域的真实电力需求。执行了标幺化过程，将此数据集成入模拟系统中，确保数据与系统参数的一致性。对于模拟中未明确指定的负荷节点，赋予零负荷状态，以更精确地模拟特定区域的负荷情形。基于设定的功率因数0.92，计算相应的无功功率。在Matpower中设定潮流计算所需的参数，对电力系统模型进行15 100次迭代潮流计算，逐次更新不同节点处的有功功率与无功功率负荷并保存，为后续FDIA攻击策略的制定与执行提供了基准数据，并为攻击的模拟和评估提供框架。

为进行单节点FDIA攻击的模拟，本文提取潮流计算后关键的系统运行数据，包括节点和支路数据以及发电机注入的有功功率和无功功率数据。在这一基础上，创建一个带有模拟测量噪声的量测数据集以模仿电力系统中的测量不确定性。每个量测数据项被赋予了基于其噪声方差的权重以增强状态估计的可靠性。采用直流潮流模型构建并标幺化雅可比矩阵，解决状态估计问题并计算了实际与估计量测之间的残差，从而为FDIA攻击的检测与效果评估提供了必要的前提。最后对电力系统的某些特定节点实施1 500次模拟攻击，即可通过调整这些节点的状态值分析该节点攻击前后电压幅值状态值的变化。得到攻击点处攻击前后状态图，该处1 500个样本攻击前后状态值微小变化表明FDIA攻击被精心设计以避免在攻击节点触发BDD系统的警报。如图5—7所示。

通过图5—7可知，攻击点处1 500个样本的状态值微小变化表明FDIA被精心设计以避免在攻击节点触发BDD系统的警报。随后在3个节点测试系统随机各进行4次状态估计，以验证攻击有效果，如图8—10所示。

通过图8—10可知，真实状态值和估计状态值

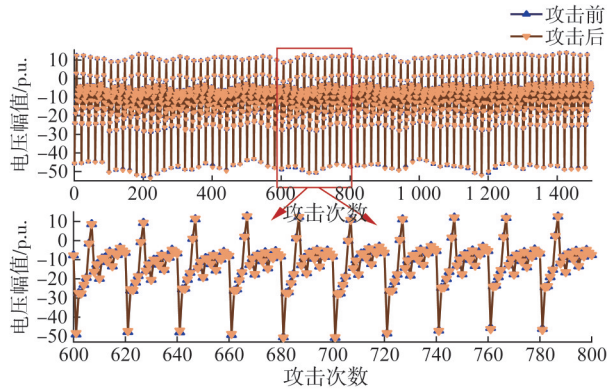

图 5　IEEE 14 节点系统攻击点处攻击前后对比

Fig. 5　Comparison of IEEE 14-node system attack points before and after attack

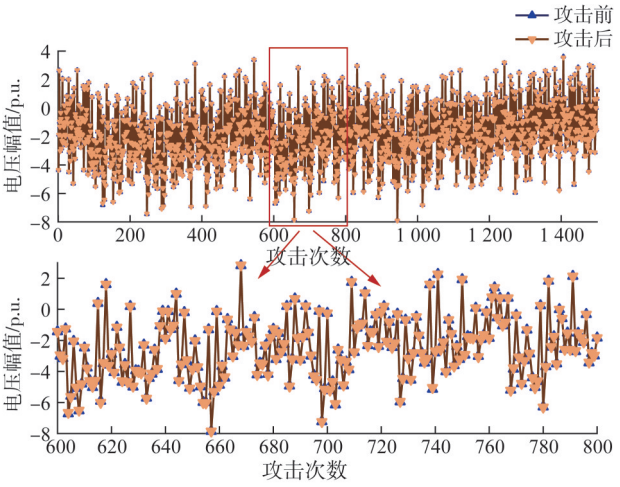

图 6　IEEE 30 节点系统攻击点处攻击前后对比

Fig. 6　Comparison of IEEE 30-node system attack points before and after attack

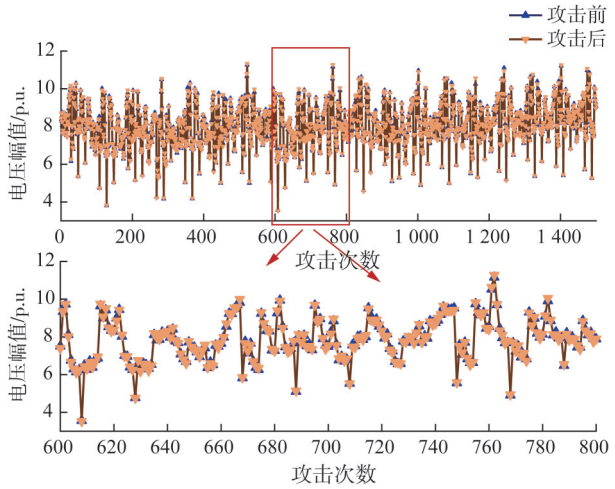

图 7　IEEE 57 节点系统攻击点处攻击前后对比

Fig. 7　Comparison of IEEE 57-node system attack points before and after attack

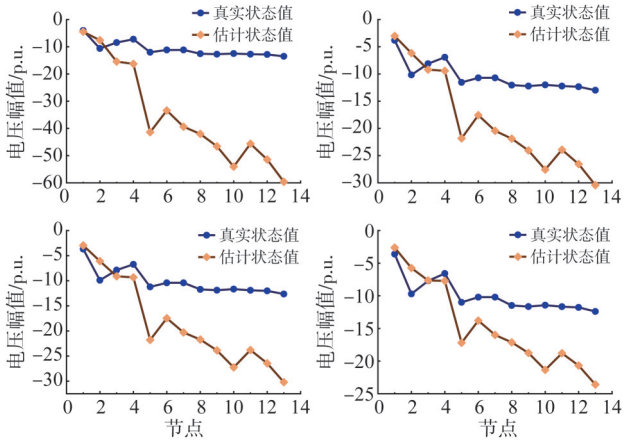

图 8　IEEE 14 节点系统估计性能验证

Fig. 8　Performance validation of IEEE 14-node system estimation

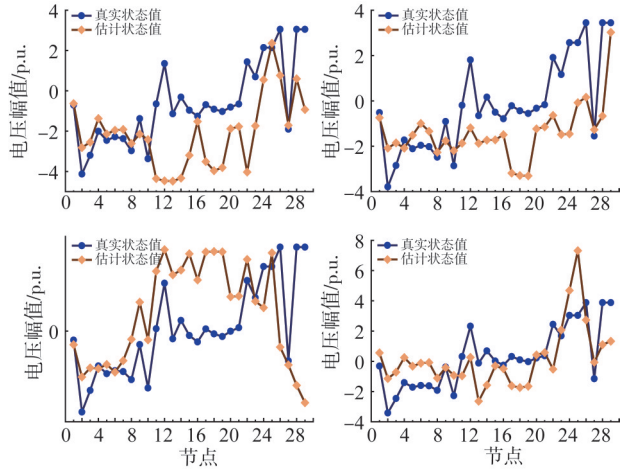

图 9　IEEE 30 节点系统估计性能验证

Fig. 9　Performance validation of IEEE 30-node system estimation

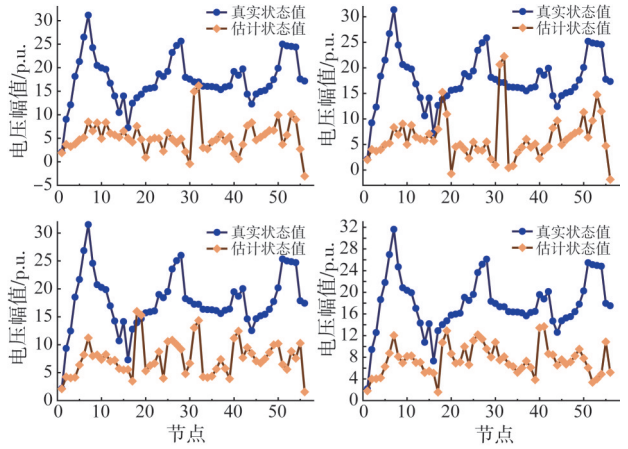

图 10　IEEE 57 节点系统估计性能验证

Fig. 10　Performance validation of IEEE 57-node system estimation

之间出现了明显的差异。这表明了在未受攻击和受攻击后的状态值之间的差异。可见FDIA成功地干扰了状态估计器的准确性，证明了攻击是有效的。

通过对攻击点处状态图和估计性能验证图可知，攻击虽然在攻击节点处变化是微妙的，躲过了BDD检测，但在整个电网中产生了扩散效应，从而影响了整个系统的状态估计。最终证明攻击是有效且隐蔽的，满足虚假注入攻击的条件。最终生成了反映攻击影响的量测数据集，为后续的识别提供了数据准备。

3.2 评价指标及训练设置

使用精确度(precision，记作$R_{\text{Precision}}$)、召回率(recall，记作R_{Recall})、特异性(specificity，记作$R_{\text{Specificity}}$)、F1分数(F1-Score，记作$S_{\text{F1-score}}$)、准确率(accuracy，记作R_{Accuracy})作为模型评估主要参数，公式如式(20)所示。

$$R_{\text{Precision}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}} \tag{20}$$

$$R_{\text{Recall}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}} \tag{21}$$

$$R_{\text{Specificity}}=\frac{N_{\text{TN}}}{N_{\text{TN}}+N_{\text{FP}}} \tag{22}$$

$$S_{\text{F1-score}}=\frac{2\times R_{\text{Recall}}\times R_{\text{Precision}}}{R_{\text{Recall}}+R_{\text{Precision}}} \tag{23}$$

$$R_{\text{Accuracy}}=\frac{N_{\text{TP}}+N_{\text{TN}}}{N_{\text{TP}}+N_{\text{TN}}+N_{\text{FP}}+N_{\text{FN}}} \tag{24}$$

式中N_{TP}、N_{TN}、N_{FP}、N_{FN}分别为真正例、真负例、假负例、假正例的数量。考虑到数据集中存在正负样本不平衡的情况。在这种情况下，仅使用准确率来评估定位检测性能会导致误导性结果，缺乏足够的可信度。为了更准确地衡量检测方法的性能，引入了精确率、召回率和F1-score等综合评估指标。

在数据处理方面，本文将数据集按照7∶2∶1的比例划分为训练集、验证集和测试集，以进行模型的训练和评估。Transformer模型的超参数配置为：编码层数量为2层，多头注意力头数设为8，前馈神经网络的隐藏层维度为128。所有实验均采用Adam优化器，批次大小为32，初始学习率设定为0.001。

3.3 参数敏感性分析

3.3.1 类别不平衡敏感性分析

针对FDIA数据中攻击样本占比低的特性，通过调节Focal Loss$^{\text{IM}}$中的类别权重参数α_t，分析模型对不平衡数据的适应性。在IEEE 14节点系统中，设置α_t=0.25、0.5、0.75，实验结果如表1所示。当α_t=0.5时，召回率和F1分数分别达到79.9%和84.3%，表明合理的类别权重设置能有效平衡少数类样本的损失贡献。

表1 IEEE 14节点系统不同α_t值的性能对比

Tab.1 Performance comparison of different values of α_t in the IEEE 14-node system

α_t	R_{Recall}	$S_{\text{F1-score}}$	$R_{\text{Specificity}}$
0.25	0.763	0.812	0.986
0.5	0.799	0.843	0.998
0.75	0.751	0.826	0.992

3.3.2 多头注意力机制的特征提取

多头注意力机制的计算中头数的选择是Transformer编码层关键的一步，为探究多头注意力机制中头数对特征提取能力的影响，在IEEE 30节点系统中对比了不同头数的检测性能与计算效率。如表2所示，当注意力头数从4增加至8时，F1分数提升3.2%，而头数进一步增加至16时，模型性能趋于稳定。实验表明，8头设计在特征提取效率与计算成本间达到了平衡。

表2 IEEE 30节点系统不同头数的性能对比

Tab.2 Performance comparison of different headcounts in the IEEE 30-node system

注意力头数	R_{Recall}	$S_{\text{F1-score}}$	训练时间/(s·epoch^{-1})
4	0.891	0.914	5.8
8	0.931	0.946	7.2
16	0.928	0.943	9.5

3.4 同算法下对比分析

本部分采用基于Transformer算法对IEEE 14节点系统、IEEE 30节点系统和IEEE 57节点系统进行FDIA检测。这些系统面临的共同挑战是数据集的不平衡性。本文引入Focal Loss$^{\text{IM}}$，并将其性能与传统的BCE Loss及标准Focal Loss进行比较，模型训练对比如图11—13所示。

通过图11可知，在IEEE 14节点系统中，Focal Loss$^{\text{IM}}$在训练集上表现出较快的收敛速度和更高的准确率，与BCE Loss和Focal Loss相比，损失下降更为平滑。在验证集上，该损失函数展示了出色的泛化能力，准确率稳步提升，且损失曲线呈现较小的波动，表明算法对未见样本的预测性能有所

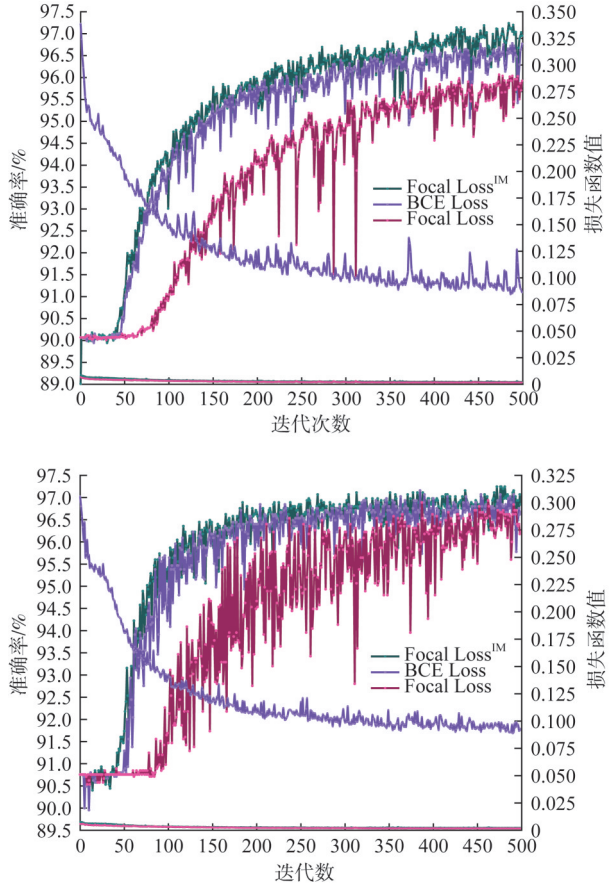

图 11　IEEE 14 节点系统模型训练对比

Fig. 11　Comparison of model trainings for the IEEE 14-node system

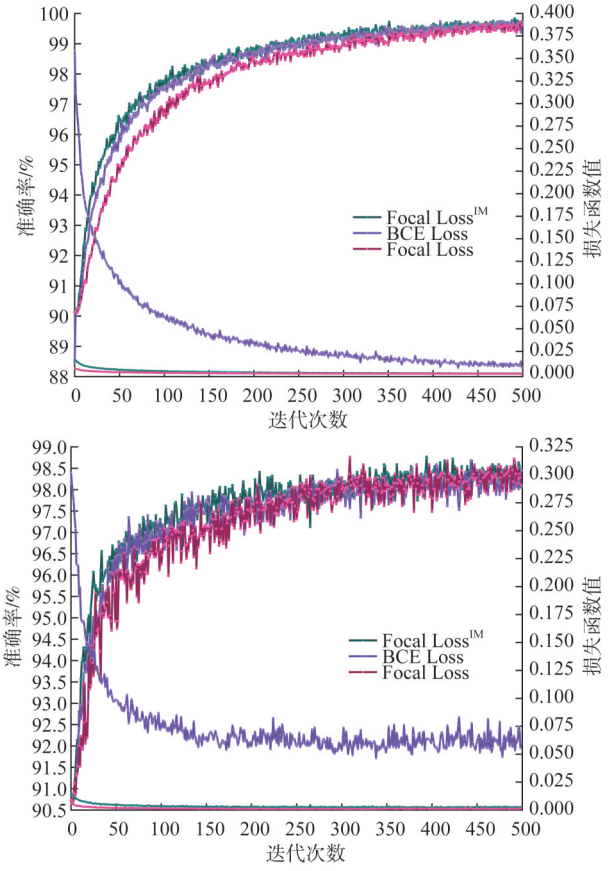

图 12　IEEE 30 节点系统模型训练对比

Fig. 12　Comparison of model trainings for the IEEE 30-node system

增强。由图 12 可知，对于 IEEE 30 节点系统中，Focal Loss^{IM}同样在训练初期迅速减少了损失，而且准确率增长快于其他两种损失函数。在验证集上，算法在使用 Focal Loss 变体函数时达到了更高的准确率，同时保持了较低的损失，展现其在不平衡数据集上的优势。由图 13 可知，在 IEEE 57 节点系统中，Focal Loss^{IM}依旧在早期训练阶段显示了其对不平衡类别的适应能力，损失曲线平稳而下降趋势明显。验证集上的表现也证实了其稳定性和泛化能力，算法在整个训练过程中维持了较高的准确率。模型测试集对比如表 3—5 所示。

在 IEEE 14，30 和 57 节点系统的不平衡数据集分析中，Focal Loss^{IM}在关键性能指标上相较于 BCE Loss 和原始 Focal Loss 显示出显著提升。由表 1 可知，在 IEEE 14 节点系统中，Focal Loss^{IM}将准确率提升至 96. 71 %，比 BCE Loss 高出 0. 49 %，与 Focal Loss 的提升为 0. 45 %。其召回率达到 79. 9 %，较 BCE Loss 和 Focal Loss 分别提高了 8. 0 %和 8. 8 %，而 F1 得分提升到 84. 3 %，分别比 BCE Loss 和 Focal Loss 高出 3. 5 %和 3. 4 %。由表 2 可知，在 IEEE 30 节点系统中，F1 得分达到 94. 6 %，相比 BCE Loss 和 Focal Loss 的提升分别为 3. 3 %和 3. 6 %。由表 3 可知，而在 IEEE 57 节点系统中，Focal Loss^{IM}在召回率上实现了 11. 4 %的提升至 89. 9 %，与 Focal Loss 相比，F1 得分也显著提高，从 87. 2 %提升至 93. 7 %，比 BCE Loss 的提升更是达到 3. 4 %和 6. 5 %。突显 Focal Loss^{IM}在提高召回率和 F1 得分上的显著优势，减漏检攻击的风险，同时保持对非攻击样本的高特异性。

3. 5　不同算法对比分析

为验证所提方法在 FDIA 检测中的优越性，本文选取了基于数据驱动的检测模型与损失函数组合的方法进行对比分析，包括 Focal Loss^{IM}-Transformer、Focal Loss^{IM}-CNN、BCE Loss-CNN、Focal Loss^{IM}-LSTM、BCEloss-LSTM、Focal Loss^{IM}-

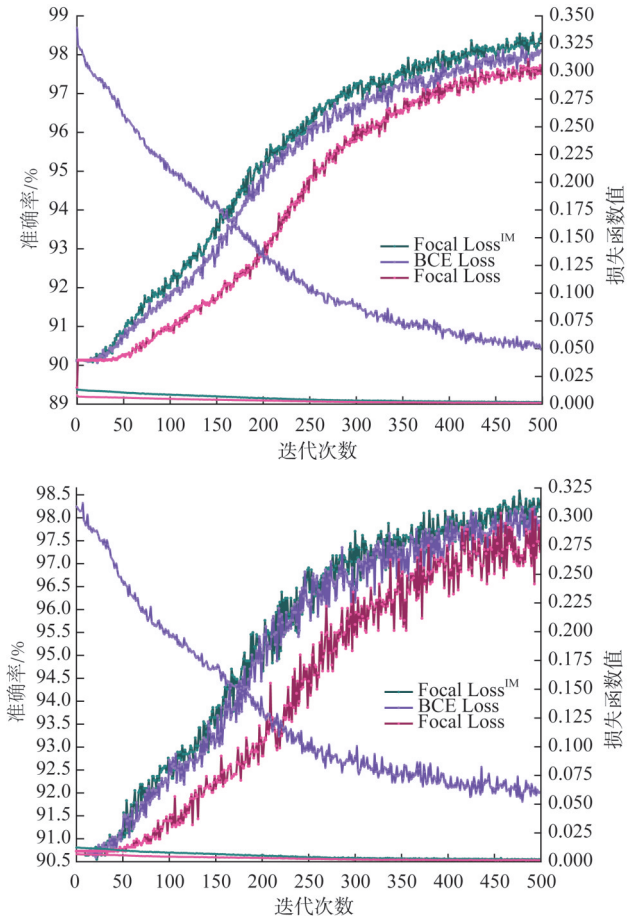

图 13 IEEE 57 节点系统模型训练对比

Fig. 13 Comparison of model training for the IEEE 57-node system

表 3 IEEE 14 节点系统测试集指标对比

Tab. 3 Comparison of test set indexes for the IEEE 14-node system

损失函数	R_{Accuracy}	$R_{\text{Precision}}$	R_{Recall}	$R_{\text{Specificity}}$	$S_{\text{F1-score}}$
Focal LossIM	0. 967	0. 892	0. 799	0. 998	0. 843
BCE Loss	0. 962	0. 923	0. 719	0. 993	0. 808
Focal Loss	0. 962	0. 937	0. 711	0. 994	0. 809

表 4 IEEE 30 节点系统测试集指标对比

Tab. 4 Comparison of test set indexes for the IEEE 30-node system

损失函数	R_{Accuracy}	$R_{\text{Precision}}$	R_{Recall}	$R_{\text{Specificity}}$	$S_{\text{F1-score}}$
Focal Loss$^{\text{IM}}$	0. 989	0. 962	0. 931	0. 997	0. 946
BCE Loss	0. 981	0. 960	0. 870	0. 996	0. 913
Focal Loss	0. 981	0. 943	0. 879	0. 994	0. 910

Mlp、BCE Loss-MLP、Focal Loss$^{\text{IM}}$-RNN 和 BCE Loss-RNN。这些模型涵盖了从时序建模、优化分类到集成学习的多种技术框架，代表了当前主流的

表 5 IEEE 57 节点系统测试集指标对比

Tab. 5 Comparison of test set indexes for the IEEE 57 bus system

损失函数	R_{Accuracy}	$R_{\text{Precision}}$	R_{Recall}	$R_{\text{Specificity}}$	$S_{\text{F1-score}}$
Focal Loss$^{\text{IM}}$	0. 986	0. 978	0. 899	0. 998	0. 937
BCE Loss	0. 980	0. 968	0. 846	0. 997	0. 903
Focal Loss	0. 974	0. 980	0. 785	0. 998	0. 872

FDIA 检测方法。在 IEEE 14 节点系统、IEEE 30 节点系统和 IEEE 57 节点系统上，对比实验分别从准确率、精确度、召回率、特异性和 F1 分数等多个指标进行了系统性评估，如表 4—6 所示。

表 6 IEEE 14 节点系统测试集指标对比

Tab. 6 Comparison of test set indexes for the IEEE 14-node system

损失函数-模型	R_{Accuracy}	$R_{\text{Precision}}$	R_{Recall}	$R_{\text{Specificity}}$	$S_{\text{F1-score}}$
FocalLossIM-Transformer	0. 967	0. 892	0. 799	0. 998	0. 843
BCE Loss-CNN	0. 954	0. 880	0. 679	0. 989	0. 767
Focal LossIM-CNN	0. 962	0. 936	0. 707	0. 994	0. 806
BCE Loss-LSTM	0. 951	0. 949	0. 594	0. 996	0. 731
Focal LossIM-LSTM	0. 947	0. 940	0. 562	0. 996	0. 703
BCE Loss-MLP	0. 956	0. 931	0. 655	0. 994	0. 769
Focal LossIM-MLP	0. 954	0. 925	0. 643	0. 994	0. 759
BCE Loss-RNN	0. 962	0. 923	0. 723	0. 993	0. 811
Focal LossIM-RNN	0. 963	0. 937	0. 715	0. 994	0. 811

表 7 IEEE 30 节点系统测试集指标对比

Tab. 7 Comparison of test set indexes for the IEEE 30-node system

损失函数-模型	R_{Accuracy}	$R_{\text{Precision}}$	R_{Recall}	$R_{\text{Specificity}}$	$S_{\text{F1-score}}$
FocalLossIM-Transformer	0. 989	0. 962	0. 931	0. 997	0. 946
BCE Loss-CNN	0. 964	0. 897	0. 771	0. 989	0. 829
Focal LossIM-CNN	0. 965	0. 898	0. 775	0. 989	0. 832
BCE Loss-LSTM	0. 970	0. 933	0. 789	0. 993	0. 855
Focal LossIM-LSTM	0. 971	0. 929	0. 798	0. 993	0. 859
BCE Loss-MLP	0. 947	0. 916	0. 571	0. 994	0. 703
Focal LossIM-MLP	0. 949	0. 918	0. 591	0. 994	0. 719
BCE Loss-RNN	0. 962	0. 905	0. 733	0. 991	0. 810
Focal LossIM-RNN	0. 962	0. 882	0. 753	0. 988	0. 812

由表 6—8 可知，在 IEEE14、30 和 57 节点系统的不平衡数据测试集上进行的实验结果表明，相比于其他机器学习方法，Focal Loss$^{\text{IM}}$-Transformer 在所有考察的性能指标上均最优，尤其是在识别少数类(即攻击样本)的能力上，即受攻击样本的召回率和 F1 值均展现了显著的优势。高召回率意味着攻

表8　IEEE 57节点系统测试集指标对比

Tab. 8　Comparison of test set indexes for the IEEE 57-node system

损失函数-模型	$R_{Accuracy}$	$R_{Precision}$	R_{Recall}	$R_{Specificity}$	$S_{F1\text{-}score}$
FocalLoss^{IM}-Transformer	0. 986	0. 978	0. 899	0. 998	0. 937
BCE Loss-CNN	0. 954	0. 935	0. 635	0. 995	0. 756
Focal Loss^{IM}-CNN	0. 943	0. 891	0. 558	0. 992	0. 686
BCE Loss-LSTM	0. 945	0. 963	0. 526	0. 998	0. 680
Focal Loss^{IM}-LSTM	0. 951	0. 927	0. 610	0. 994	0. 736
BCE Loss-MLP	0. 956	0. 942	0. 647	0. 995	0. 767
Focal Loss^{IM}-MLP	0. 957	0. 923	0. 671	0. 993	0. 777
BCE Loss-RNN	0. 957	0. 943	0. 659	0. 995	0. 776
Focal Loss^{IM}-RNN	0. 959	0. 924	0. 687	0. 993	0. 788

击事件被有效识别，直接关联到系统的安全性。高F1值表明模型在减少误报（保持高精确率）和减少漏报（保持高召回率）之间达到了较好的平衡。因此，本文提出的Focal Loss^{IM}-Transformer在不平衡FDIA数据识别上具有显著优势。

4　结语

针对FDIA识别数据不平衡问题，本文提出了基于Focal Loss^{IM}-Transformer算法的FDIA检测方法。

通过在IEEE 14节点系统、30节点系统和57节点系统上的仿真验证了攻击的隐蔽性和有效性，根据攻击样本的不平衡比例所设计的Focal Loss^{IM}，能更好地关注少数类（即攻击样本），且保持了对多数类的高精度识别，以此更好地适应FDIA数据不平衡的问题。本文提出的Focal Loss^{IM}-Transformer在准确率上超越了机器学习检测方法，且在不平衡数据集中最为关键的性能指标上也展现了优势，即在召回率和F1得分上Transformer显示出其对少数类的高识别能力，保持了高特异性，减少了误报率。

本文验证了Focal Loss^{IM}-Transformer在FDIA检测问题中的有效性，适合于处理对安全性要求极高的不平衡数据集。未来将进一步优化模型结构和损失函数以增强模型对动态新增攻击类别的适应性，解决类别不断积累的长期挑战，同时提升对新型攻击模式的泛化能力

参考文献

[1]　林峰，梅勇，朱益华，等. 网络攻击对电力系统典型场景全过程影响综述[J].南方电网技术，2023,17(11):61－75.
LIN Feng，MEI Yong，ZHU Yihua，et al. Overview of the overall influence of cyber attack on typical scenarios of power systems [J]. Southern Power System Technology，2023，17(11)：61－75.

[2]　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.

[3]　陈家琪，王琦，汤奕，等. 考虑双侧特征的电力信息物理系统异常检测方法 [J]. 电网技术，2022，46(6)：2339－2348.
CHEN Jiaqi，WANG Qi，TANG Yi，et al. Anomaly detection method for cyber physical power system considering bilateral features[J]. Power System Technology，2022，46(6)：2339－2348.

[4]　LIU C S，LIANG H，CHEN T W. Network parameter coordinated false data injection attacks against power system AC state estimation [J]. IEEE Transactions on Smart Grid，2021，12(2)：1626－1639.

[5]　王琦，邰伟，汤奕，等. 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.

[6]　ZHOU T L，XIA H K，ZHANG L L. Real-time detection of cyber-physical false data injection attacks on power systems [J]. IEEE Transactions on Industrial Informatics，2021，17(10):6810－6819.

[7]　单瑞卿，盛阳，苏盛，等. 考虑攻击方身份的电力监控系统网络安全风险分析[J]. 电力科学与技术学报，2022，37(5)：3－16.
SHAN Ruiqing，SHENG Yang，SU Sheng，et al. Network security risk analysis of power monitoring system considering the identity of attacker[J]. Journal of Electric Power Science and Technology，2022，37(5):3－16.

[8]　TIAN M，DONG Z C，ZHAO L，et al. Computation of worst-case operation scenarios against false data injection attacks considering load demand and generation uncertainties [C]// IECON 2020 the 46th Annual Conference of the IEEE Industrial Electronics Society，October 18－21，2020，Singapore，Singapore. New York：IEEE，2020：3359－3364.

[9]　CHEN B Y，LI H B，ZHOU B.Real-time identification of false data injection attacks：a novel dynamic-static parallel state estimation based mechanism [J]. IEEE Access，2019 (7)：95812－95824.

[10]　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.

[11]　黄冬梅，丁仲辉，胡安铎，等. 低成本对抗性隐蔽虚假数据注入攻击及其检测方法[J]. 电网技术，2023，47(4)：1531－1540.
HUANG Dongmei，DING Zhonghui，HU Anduo，et al. Low-cost adversarial stealthy false data injection attack and detection method[J]. Power System Technology，2023，47(4)：1531－1540.

[12]　LIU Z，WANG Q，YE Y，et al. A GAN-based data injection

attack method on data-driven strategies in power systems[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 3203 - 3213.

[13] GUO F, YAO S H, ZHANG N, et al. XGBoost based fake data injection attack detection method for power grid[C]//2022 2nd International Conference on Electrical Engineering and Control Science, December 16 - 18, 2022, Nanjing, China. Nanjing: ICECS, 2022: 404 - 407.

[14] JIN S, SHENG S, XUE Y, et al. Review on data-driven based electricity theft detection method and research prospect for low false positive rate[J]. Automation of Electric Power Systems, 2022, 46(1):3 - 14.

[15] CHEN Z, ZHU J, LI S, et al. Detection of false data injection attack in automatic generation control system with wind energy based on fuzzy support vector machine[C]// Proceedings of the IECON 2020 46th Annual Conference of the IEEE Industrial Electronics Society, October 18 - 21, 2020, Singapore, Singapore.New York: IEEE, 2020:122 - 132.

[16] CHEN Y, HAYAWI K, ZHAO Q, et al. Vector auto-regression-based false data injection attack detection method in edge computing environment[J]. Sensors, 2022, 22(18):6789 - 6797.

[17] 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.

[18] 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.

[19] HUANG D M, HE L A, SUN J Z, et al. Distributed detection method for a false data attack in a power grid based on edge computing[J]. Power System Protection and Control, 2021, 49(13):1 - 9.

[20] 席磊, 程琛, 田习龙. 基于改进卷积神经网络的电网虚假数据注入攻击定位方法[J]. 南方电网技术, 2025, 19(1): 75 - 85.
XI Lei, CHENG Chen, TIAN Xilong. Improved convolutional neural network-based localization method for false data injection attacks on power grids [J]. Southern Power System Technology, 2025, 19(1): 75 - 85.

[21] LIU X, LI Z Y. Local load redistribution attacks in power systems with incomplete network information [J]. IEEE Transactions on Smart Grid, 2014, 5(4): 1665 - 76.

[22] 陈冰, 唐永旺. 基于Transformer编码器的智能电网虚假数据注入攻击检测 [J]. 计算机应用与软件, 2022, 39(7): 336 - 342.
CHEN bing, TANG Yongwang. False data injection attacks detecting based on transformer encoder in smart grid [J]. Computer Applications and Software, 2022, 39(7):336 - 342.

[23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, August 11 - 13, 2017, Long Beach, California, USA. California: ICNIPS, 2017: 113 - 122.

[24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. 2017 IEEE International Conference on Computer Vision (ICCV), October 22 - 29, 2017, Venice, Italy. New York: IEEE, 2017: 2999 - 3009.

[25] 邵振国, 林洪洲, 陈飞雄, 等. 采用区间动态状态估计的局部不可观系统谐波源定位 [J]. 电工技术学报, 2023, 38(9): 2391 - 2402.
SHAO Zhenguo, LIN Hongzhou, CHEN Feixiong , et al. Harmonic source location in the partial unobservable system based on interval dynamic state estimation [J]. Transactions of China Electrotechnical Society, 2023, 38(9): 2391 - 2402.

[26] 徐艳春, 王格, 孙思涵, 等. 基于改进广义极大似然估计的配 电网状态估计方法[J]. 南方电网技术, 2022, 16(6): 23 - 32.
XU Yanchun, WANG Ge, SUN Sihan, et al. Distribution network state estimation method based on improved generalized maximum likelihood estimation [J]. Southern Power System Technology, 2023, 16(6): 23 - 32.

[27] WANG Jie, ZHAO Chunhui. Robust control performance monitoring for varying-dimensional time-series data based on SCADA systems[J]. IEEE Transactions on Instrumentation and Measurement, 2022(71):1 - 14.

[28] 李浩, 张禄亮, 栾云飞, 等. 基于子系统划分和注入电流比的配电网故障定位方法[J]. 电力系统保护与控制, 2023, 51(8): 63 - 72.
LI Hao, ZHANG Luliang, LUAN Yunfei, et al. Fault location method for a distribution network based on subsystem division and injection current ratio [J]. Power System Protection and Control, 2023, 51(8): 63 - 72.

[29] 汪际峰, 李鹏, 梁锦照, 等. 电力系统数字化历程与发展趋势[J]. 南方电网技术, 2021, 15(11): 1 - 8.
WANG Jifeng, LI Peng, LIANG Jinzhao, et al. Development history and trends of power system digitalization[J]. Southern Power System Technology, 2021, 15(11):1 - 8.

[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.
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.

收稿日期: 2025-02-08

作者简介:

席磊(1982), 男, 高级工程师(教授级), 教授, 博士, 从事电力系统运行与控制、自动发电控制、信息物理系统网络攻击与防御、智能控制方法的研究, xilei2014@163.com;

和昀(1999), 男, 硕士研究生, 研究方向为信息物理系统网络攻击与防御, 18869029937@163.com;

李宗泽(2001), 男, 通信作者, 硕士研究生, 研究方向为信息物理系统网络攻击与防御, lizongze0608@163.com。