融合 SBIGRU 与注意力机制的虚假数据注入攻击检测

王瑞仁,魏利胜
安徽工程大学 电气工程学院,安徽 芜湖 241000

摘 要:目的 针对当前智能电网虚假数据注入攻击(Smart Grids False Data Injection Attack,SGs FDIA)检测工作仅利用系统状态的空间数据特征来识别攻击,而未考虑连续系统状态中呈现的时间数据相关性问题,研究一种基于堆叠双向门控循环单元(Stacked Bidirectional Gated Recurrent Unit)与混合注意力机制(Hybrid Attention Mechanism,HA)的检测模型 SBIGRU-HA。**方法** 首先,采用 SBIGRU 提取给定时间段内的系统时序特征,捕获数据之间的时序关系;同时,引入残差网络融合原始输入与 SBIGRU 捕获的时序特征;在此基础上,融合坐标注意力(Coordinate Attention,CA)、卷积注意力(Convolutional Block Attention Module,CBAM)、无参注意力 SimAM 三种注意力机制,提取数据的时空特征并为被注入攻击的特征分配更高的权重;最后,将得到的特征表示输入到线性层和 Sigmoid 层,完成攻击检测。**结果** 在 IEEE-14、IEEE-57 节点测试系统上进行仿真实验,实验结果表明:SBIGRU-HA 检测准确率分别达到 98.68%、96.36%,F 得分分别达到 98.32%、95.39%。**结论** SBIGRU-HA 相比较 LSTM、GRU 在各项检测指标上均有所提高,能够完成针对虚假数据的定位检测。

关键词:虚假数据注入攻击;攻击检测;双向门控循环单元;注意力机制
中图分类号:TP391.4 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2026.0003.018

False Data Injection Attack Detection Integrating Stacked Bidirectional GRU and Attention Mechanism

WANG Ruiren, WEI Lisheng
School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, Anhui, China

Abstract: Objective To address the limitation in current smart grid false data injection attack (FDIA) detection methods, which utilize only spatial data features for attack identification while neglecting temporal correlations across sequential system states, this paper proposes SBIGRU-HA, a detection model integrating stacked bidirectional gated recurrent unit (SBIGRU) and hybrid attention (HA). **Methods** First, the SBIGRU extracted the system's temporal features within a given time period and captured the temporal relationships among the data. Meanwhile, a residual network was introduced to fuse the original input with the temporal features captured by the SBIGRU. On this basis, three attention mechanisms, namely coordinate attention (CA), convolutional block attention module (CBAM), and parameter-free attention SimAM, were integrated. These mechanisms were used to extract the spatiotemporal features of the data and assign higher weights to the features with injected attacks. Finally, the obtained feature representation was fed into a linear layer and a Sigmoid layer to complete the attack detection. **Results** Simulation experiments were conducted on the IEEE-14 and IEEE-57 node test systems. The results demonstrated that the SBIGRU-HA model achieved detection accuracies of 98.68% and

收稿日期:2024-03-16 **修回日期:**2024-06-23 **文章编号:**1672-058X(2026)03-0159-08
基金项目:安徽省高校自然科学研究重大项目(KJ2020ZD39);安徽省检测技术与节能装置重点实验室开放基金项目(DTESD2020A02).
作者简介:王瑞仁(1999—),男,安徽合肥人,硕士研究生,从事电网安全研究.
作者简介:魏利胜(1978—),男,安徽巢湖人,博士,教授,从事图像识别、智能化网络控制理论与系统仿真等研究. Email: weilsh@ahpu.edu.cn.
引用格式:王瑞仁,魏利胜.融合 SBIGRU 与注意力机制的虚假数据注入攻击检测[J].重庆工商大学学报(自然科学版),2026,43(3):159-166.
Wang Ruiren, Wei Lisheng. False data injection attack detection integrating stacked bidirectional GRU and attention mechanism [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 159-166.

96.36%, and F scores of 98.32% and 95.39%, respectively, on these two test systems. **Conclusion** Compared with LSTM and GRU, the SBIGRU-HA model demonstrates improvements across all detection metrics and is capable of identifying the specific locations of false data.

Keywords: false data injection attack; attack detection; bidirectional gated recurrent unit; attention mechanism

智能电网集成信息通信技术和其他先进技术并应用于大型电网,使电力的生产、传输、分配和使用更加高效、经济和环保[1]。然而其部分用于传感、通信、存储和数据分析的网络正成为系统运行的主要漏洞[2]。虚假数据注入攻击(False Data Injection Attack,FDIA)是针对智能电网最具破坏性的恶意攻击类型之一,它利用智能电网的拓扑和参数信息设计攻击策略,破坏测量数据从而修改系统估计状态[3],进而可能导致关键线路过载和负载脱落等问题[4]。因此,针对FDIA攻击检测研究对于智能电网安全运行具有重要意义。

近年来,许多专家学者针对虚假数据注入攻击检测进行了深度研究并有许多成果,大致可分为模型驱动和数据驱动两类。模型驱动方面,杨怡等[5]用无迹卡尔曼滤波(Unscented Kalman Filter,UKF)完成FDIA检测。刘鑫蕊等[6]将基于极端梯度提升的日前负荷预测与UKF相结合,同时引入中心极限定理,实现对FDIA的精确检测和修正。Ma等[7]提出一种基于零序电压的三相配电系统FDIA检测方法,通过比较估计出的零序电压向量的L2范数与预定义阈值来检测FDIA的存在。Luo等[8]利用攻击特征的逻辑判断矩阵对FDIA的攻击点进行定位检测。Chakrabarty等[9]首次考虑了针对移相器的FDIA,利用线路电流和节点注入电流与端电压的比值设计检测指标,在量测噪声和负荷波动场景下检测率依然很高。文献[5-9]利用系统模型和参数完成检测,虽然检测效果有一定提升,但无形中增加了对系统模型的依赖。

数据驱动方面,Huang等[10]将攻击矩阵范数小于特定阈值的行置零,以此保证解的收敛性,并与几种矩阵分离算法对比,证明了所提算法的优越性。输电系统依赖监控指令,指令通道属于监视控制与数据采集系统,但针对它的攻击研究较少,Chakrabarty等[11]提出一种能够同时检测虚假数据和虚假命令注入攻击的通用方案,计算成本低,且不需要迭代。Goyel等[12]针对攻击者只知道部分网络拓扑信息的情况,采用一种攻击向量生成策略并且提出一种基于投票的集成学习技术来检测智能电网中的FDIA。Li等[13]首次将联邦学习用于电网攻击检测,提出一种将联邦学习、Transformer、Paillier密码系统相结合的FDIA检测方法,其在保护数据隐私和减少通信开销方面比集中式检测方法更具优势。Cui等[14]提出了一种针对智能电网中FDIA的检测与恢复机制,其采用改进的主成分分析法完成检测,并根据检测结果,设计了一种基于遗传优化的线性二次调节控制器来补偿FDIA对电压的变化。Wu等[15]提出一种由污染状态分离方法、增强的坏数据识别方法和状态恢复算法组成的状态重建方案,完成对虚假数据的定位、剔除与修正。Liu等[16]将深度卷积神经网络和GAN相结合,提出一种黑匣子数据注入攻击方法,可以根据电网的测量数据实时产生最小的干扰。Musleh等[17]将长短期记忆网络和AE相结合,该算法可以学习自动发电控制(Automatic Generation Control,AGC)系统的动态特性。用正常数据训练LSTM-AE,检测阶段将数据输入LSTM-AE后比较残差完成检测,但该方法需要大量的历史正常数据。上述专家学者采用各种数据驱动方法完成检测,但该方法往往涉及庞大的参数量和计算量,不利于实际使用,且检测模型设计时未充分考虑电网测量数据同时具有的时空相关性。

为了同时捕获电网测量的时空相关性,本文探究一种融合堆叠双向门控循环单元(Stacked Bidirectional Gated Recurrent Unit,SBIGRU)与混合注意力机制(Hybrid Attention Mechanism,HA)的检测模型SBIGRU-HA。与以往的FDIA检测工作不同,SBIGRU-HA不仅利用系统状态的空间数据特征来识别攻击,还从连续系统状态中呈现的时间数据相关性中进行学习。SBIGRU用于捕获数据的时序特征,其通过前向和后向门控循环单元分别捕获输入数据的过去和未来信息。采用残差网络来融合原始输入和SBIGRU捕获的时序特征,并结合坐标注意力(Coordinate Attention,CA)、卷积注意力(Convolutional Block Attention Module,CBAM)以及无参注意力SimAM三种注意力机制。最后,将输出送入线性层和Sigmoid层,完成多标签分类检测。

1 融合SBIGRU与注意力机制的检测模型

SBIGRU-HA模型结构如图1所示,其主要分为两部分,前半部分为堆叠的BIGRU网络层,后半部分为融合后的注意力模块,这两部分分别在1.1节和1.2节详细探讨。在1.2.1节、1.2.2节分别研究改进后的CBAM和CA,在1.2.3节阐述SimAM注意力,并探讨这三种注意力是如何融合的。

SBIGRU-HA模型中的SBIGRU是由BIGRU依次串联组成,BIGRU分别通过前向、后向GRU捕获时序输入数据的过去和未来信息。引入残差结构将原始输入与SBIGRU提取后的特征加权融合。SBIGRU-HA中有三种注意力分别为坐标注意力、卷积注意力以及

无参注意力 SimAM。融合后的特征表示分别送入坐标注意力和 CBAM 模块,再将两个模块的输出点积后送入无参注意力 SimAM 模块,依次通过扁平层、线性层以及 Sigmoid 非线性映射得到分类结果。

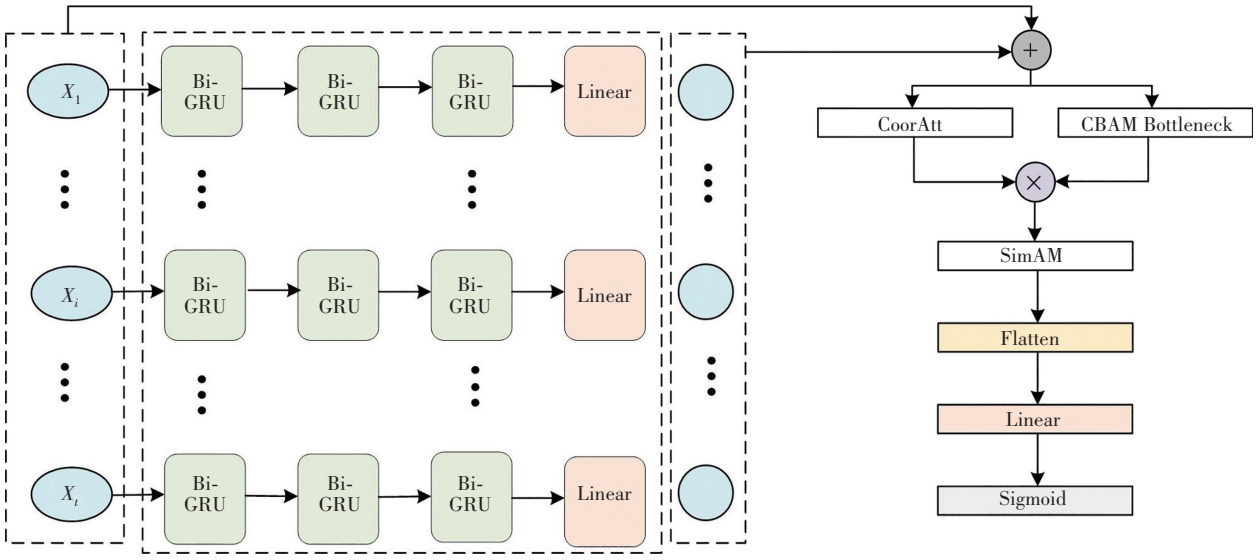


图 1　SBIGRU-HA 模型结构

Fig. 1　Structure of the SBIGRU-HA model

1.1　堆叠双向门控循环单元

GRU 将遗忘门和输入门合并到更新门中,从而直接在当前和历史状态之间创建线性依赖关系。其利用门机制可控的选择添加新信息和删除历史信息。其具体计算公式如下:

$$z_t=\text{Sigmoid}(\boldsymbol{W}_z\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{1}$$

$$\boldsymbol{r}_t=\text{Sigmoid}(\boldsymbol{W}_r\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{2}$$

$$\tilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}\cdot[\boldsymbol{r}_t*\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{3}$$

$$\boldsymbol{h}_t=(1-z_t)*\boldsymbol{h}_{t-1}+z_t*\tilde{\boldsymbol{h}}_t \tag{4}$$

其中,$\boldsymbol{x}_t$ 为输入,$\boldsymbol{z}_t$ 为更新门,$\boldsymbol{r}_t$ 为重置门,Sigmoid 为激活函数,可将数据变为 0~1 范围的数值,tanh 为正弦函数,可将数据变为[-1,1]范围的数值,$\boldsymbol{W}_z$,$\boldsymbol{W}_r$,$\boldsymbol{W}$ 均为权重矩阵,$\boldsymbol{h}_{t-1}$ 为上一时刻的隐藏状态,$\tilde{\boldsymbol{h}}_t$ 为候选隐藏状态,$\boldsymbol{h}_t$ 为传递到下一时刻的隐藏状态。

单向传播不能利用未来的信息,特别是在处理长序列时,可能会遗漏一些重要信息。此外,它容易受梯度消失的影响,可能导致模型的精度下降。在时间序列预测中,使用双向传播可以同时考虑过去和未来信息,从而更好地捕捉序列的过去和未来信息,提高模型的准确性。BIGRU 由正向 GRU 网络和反向 GRU 网络组成,其结合双向循环结构,以增加模型容量和灵活性。结构如图 2 所示,其中 X_i 代表 i 时刻的输入数据,i 时刻前向 GRU 同时接收 i 时刻的时序输入 X_i,以及上一时刻前向 GRU 的隐层输出 $\overrightarrow{\boldsymbol{h}}_{i-1}$;$i$ 时刻后向 GRU 接收 i 时刻的时序输入 X_i,以及 $i+1$ 时刻后向 GRU 的隐层输出 $\overleftarrow{\boldsymbol{h}}_{i+1}$。

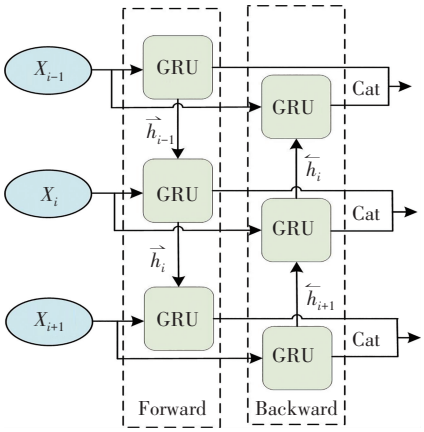


图 2　BIGRU 网络结构

Fig. 2　Architecture of the BIGRU network

SBIGRU 结构如图 3 所示。BIGRU 之间串行连接,最后一个 BIGRU 后接线性层完成与输入的通道对齐。

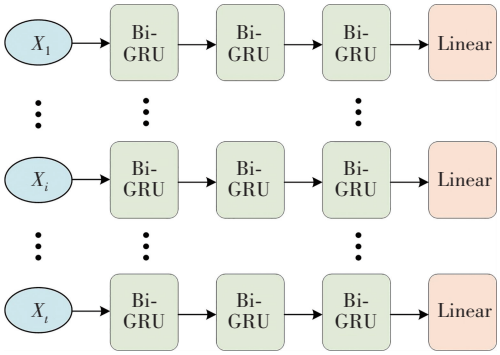


图 3　SBIGRU 模型结构

Fig. 3　Structure of the SBIGRU model

1.2　混合注意力机制

1.2.1　坐标注意力的改进

近年来,注意力机制广泛应用于机器学习任务中。

其中,最经典的注意力是 SE(Squeeze and Excitation)。然而,SE 只考虑通道的重要性,而忽略了位置信息。位置信息对于生成空间特征图是非常重要的,CA 通过池化生成两个独立的方向感知特征图,并对这两个包含方向信息的特征图进行编码,生成两个注意图来捕获每个空间方向的依赖性。为此,采用改进后的坐标注意力,最终的输出不再是原始输入与沿水平和垂直方向注意力特征图的积,而是先将原始输入通过相应的卷积、Batchnorm 以及 Swish 非线性变换来获取增强后的特征表示,然后再与沿水平和垂直方向注意力特征图的积,从而在捕获通道信息和位置关系的同时提升网络的拟合效果。坐标注意力结构如图 4 所示。

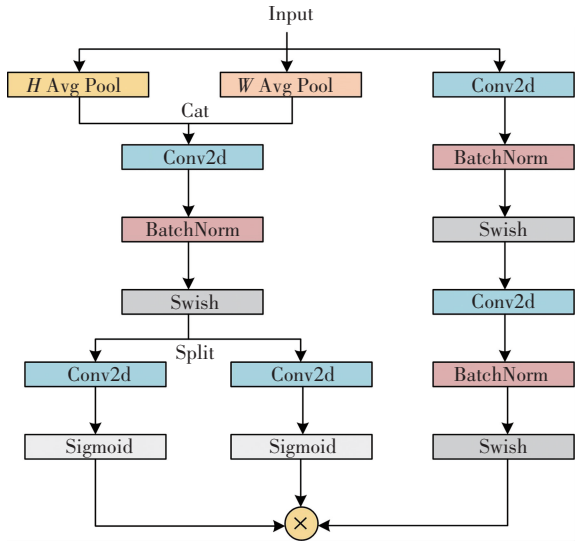


图 4 坐标注意力模型架构

Fig. 4 Architecture of the coordinate attention model

图 4 中,坐标注意力机制的实现步骤如下:

首先,对于给定输入 X,进行沿水平和垂直方向的编码,其中池化核分别为 $(H,1)$,$(1,W)$。高度为 h 的第 c 个通道的输出 $\boldsymbol{Z}_c^h$ 为

$$\boldsymbol{Z}_c^h = \frac{1}{W}\sum_{0\leqslant i\leqslant W}\boldsymbol{X}_c(h,i) \tag{5}$$

相似的宽度为 w 的通道 c 的输出 $\boldsymbol{Z}_c^w$ 为

$$\boldsymbol{Z}_c^w = \frac{1}{H}\sum_{0\leqslant j\leqslant H}\boldsymbol{X}_c(j,w) \tag{6}$$

式(5)、式(6)分别表示图 4 中的 H Avg Pool 和 W Avg Pool。上述两种变换分别对两个空间方向进行分解,得到一对方向感知特征图。这种转换使注意力能够捕捉到沿某一空间方向的长距离依赖和沿另一空间方向的精确信息。通过上述公式的分解,可以得到全局的感受野,编码出准确的位置信息。

其次,将沿两个方向的编码信息拼接后,经过 1×1 的卷积、Batchorm、Swish 函数运算得到中间特征映射 $\boldsymbol{f}$。公式如下:

$$\boldsymbol{f}=\text{Swish}(\text{Conv}([z^h,z^w])) \tag{7}$$

其中,$\boldsymbol{f}\in\mathbf{R}^{C/r\times(H+W)}$,$r$ 表示衰减率,方括号[,]表示沿空间维度的拼接操作,Conv 是 1×1 卷积,Swish 为非线性激活函数。

再次,将 $\boldsymbol{f}$ 沿空间维度方向分解,并进行 1×1 的卷积变换得到与输入 X 相同的维数后,通过 Sigmoid 激活函数得到两个方向上的注意权值:

$$\boldsymbol{g}_c^h=\text{Sigmoid}(\text{Conv}(\boldsymbol{f}^h)) \tag{8}$$

$$\boldsymbol{g}_c^w=\text{Sigmoid}(\text{Conv}(\boldsymbol{f}^w)) \tag{9}$$

其中,Conv 是 1×1 卷积,Sigmoid 为非线性激活函数。

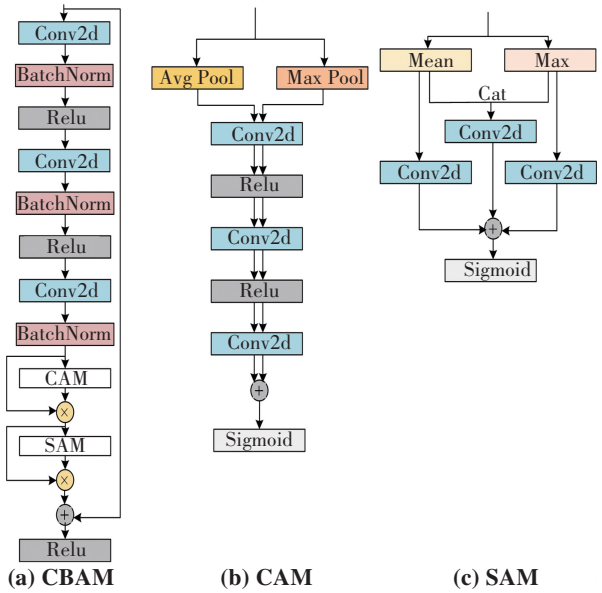
最后,将原始输入经过两层 conv2d、Batchnorm、Swish 后,与上述两个方向的权值点乘得到最终结果。公式如下:

$$\boldsymbol{y}_c(i,j)=K(\boldsymbol{x}_c(i,j))\otimes\boldsymbol{g}_c^h(i)\otimes\boldsymbol{g}_c^w(j) \tag{10}$$

其中,$0\leqslant i<W$,$0\leqslant j<H$,$\boldsymbol{g}_c^h$、$\boldsymbol{g}_c^w$ 分别表示在水平和垂直方向上的坐标注意力加权,K 表示两层 Conv2d、Batchnorm、Swish 运算,⊗表示点积。

1.2.2 CBAM 的改进

CBAM 是一个轻量级的注意力模块,包含两个子模块,通道注意模块(Channel Attention Module,CAM)和空间注意模块(Spatial Attention Module,SAM)。CBAM 模型可以无缝集成到任何 CNN 架构中,并与基本 CNN 进行端到端训练,减少了参数数量和计算能力,并确保作为即插即用模块集成到现有网络架构中。对于 CBAM 的改动主要体现在 CAM、SAM 中,CBAM、CAM、SAM 网络结构如图 5 所示。CAM 中在平均池化和最大池化后多添加一个卷积和 Relu 激活函数。SAM 的输出为原输入分别经过平均池化和最大池化后再分别卷积,并加上融合平均池化和最大池化再卷积的特征,最终再通过 Sigmoid。



(a) CBAM **(b) CAM** **(c) SAM**

图 5 CBAM 架构

Fig. 5 Architecture of CBAM

CBAM 具体计算过程如下。首先,将 F 作为输入,其通过多层感知机后得到中间特征图 $\boldsymbol{F}'$:

$$\boldsymbol{F}'=\mathrm{MLP}(\boldsymbol{F}) \tag{11}$$

其中,$\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$,MLP 是指多层感知机。

其次,将中间特征图 $\boldsymbol{F}'$输入通道注意力模块,$\boldsymbol{F}'$分别经最大池化和平均池化后通过同一个 MLP′后相加,再通过 Sigmoid 完成非线性映射,其输出与中间特征图 $\boldsymbol{F}'$点积后得到 $\boldsymbol{F}''$:

$$\boldsymbol{F}''=\boldsymbol{F}'\otimes\mathrm{Sigmoid}\begin{pmatrix}\mathrm{MLP}'(\mathrm{AvgPool}(\boldsymbol{F}'))+\\ \mathrm{MLP}'(\mathrm{MaxPool}(\boldsymbol{F}'))\end{pmatrix} \tag{12}$$

其中,MLP′表示 SAM 中的多层感知机,AvgPool、MaxPool 分别表示平均池化、最大池化,Sigmoid 为非线性激活函数,⊗表示对应张量点积。

再次,将 $\boldsymbol{F}''$注入空间注意力模块,$\boldsymbol{F}''$分别经平均池化和最大池化后再卷积,其输出与 $\boldsymbol{F}''$经平均池化和最大池化拼接后再卷积的结果相加,再通过非线性激活函数 Sigmoid 完成非线性映射,将其与 $\boldsymbol{F}''$点积后得到 $\boldsymbol{F}'''$:

$$\boldsymbol{F}'''=\boldsymbol{F}''\otimes\mathrm{Sigmoid}\begin{pmatrix}f^{7\times7}(\mathrm{AvgPool}(\boldsymbol{F}''))+\\ f^{7\times7}(\mathrm{MaxPool}(\boldsymbol{F}''))+\\ f^{7\times7}\left(\begin{bmatrix}\mathrm{AvgPool}(\boldsymbol{F}''),\\ \mathrm{MaxPool}(\boldsymbol{F}'')\end{bmatrix}\right)\end{pmatrix} \tag{13}$$

其中,$f^{7\times7}$为卷积,7×7 为卷积核尺寸,方括号[,]表示沿空间维度的拼接操作,AvgPool、MaxPool 分别表示平均池化、最大池化,Sigmoid 为非线性激活函数,⊗表示点积。

最终将原始输入与 $\boldsymbol{F}'''$加权后送入 Relu 得到最终输出 $\boldsymbol{F}''''$:

$$\boldsymbol{F}''''=\mathrm{Relu}(\boldsymbol{F}+\boldsymbol{F}''') \tag{14}$$

1.2.3　融合后的注意力机制

通道注意和空间注意关注一维和二维的关系,而 SimAM 注意力机制通过设计能量函数直接生成真正有效的 3D 权重,无需添加额外的子网络或额外的模型参数。SimAM 注意力使用能量函数 E 来计算目标像素点与周围像素点之间的关系,其计算公式如下:

$$E=\frac{4(\sigma^2+\lambda)}{(t-\mu)^2+2+2\lambda} \tag{15}$$

其中,t 为目标神经元,λ 为常数,μ 和 σ^2 是在该通道内移除的目标神经元的均值和方差。

通过添加 Sigmoid 函数来抑制注意权值的离群值,并与输入特征矩阵的相应元素进行点积运算得到 SimAM 模块的最终输出 $\bar{\boldsymbol{W}}$ 为

$$\bar{\boldsymbol{W}}=\mathrm{Sigmoid}\left(\frac{1}{E}\right)\otimes\boldsymbol{W} \tag{16}$$

其中,$\boldsymbol{W}$ 为 SimAM 模块输入,Sigmoid 为非线性函数,⊗表示点积,E 表示能量函数。

最终将三种注意力相融合,改进后的 CA 和 CBAM 模块并行,两者输出结果点积后输入 SimAM 模块。在 CA 和 CBAM 后加入 SimAM 注意机制,其通过评估每个神经元的重要性来调整特征图的注意分布,使通道和空间注意协同工作。同时,SimAM 可以自适应调整特征映射的权值,更加关注目标的局部区域。这样可以提高目标定位精度,减少定位误差,增强网络的特征提取能力。

2　实验验证及结果分析

2.1　数据集

首先,从理论上而言,如果攻击者能够获得所有系统配置信息(即网格拓扑信息、系统参数、状态估计算法的细节和坏数据检测方法等),并且有能力操纵所有仪表测量,那么很容易成功地启动 FDIA。然而,在实际攻击中,攻击者不太可能已知系统全部信息。

其次,由于电力系统中 PMU 设备越来越多,基于纯 SCADA 测量的状态估计器正逐渐向混合状态估计器发展。对于攻击者而言,FDIA 模型应该进行相应的修改,否则控制中心就会利用 PMU 来检测网络攻击。本实验所采用的数据集来自文献[15]。其设计的非线性攻击模型不仅考虑在有部分网络拓扑信息下的 FDIA,且还可同时处理 SCADA 和 PMU 的测量。

最后,分别在已知 IEEE-14 总线全部拓扑信息和部分 IEEE-57 拓扑结构信息的情况下实施 FDIA。数据的采样频率为 15 min 一次,总共采集 2 480 个数据,最终使用的数据集中包括正常数据 480 个,遭受攻击的数据 2 000 个。

2.2　评价指标及实验设置

检测领域常用的指标分别为准确率($\zeta_{\mathrm{Accuracy}}$)、精度($\zeta_{\mathrm{Precision}}$)、召回率($\zeta_{\mathrm{Recall}}$)和 F 分数(ζ_F),以上指标值越大,代表分类器性能越好,其公式分别如下:

$$\zeta_{\mathrm{Accuracy}}=\frac{N_{\mathrm{TP}}+N_{\mathrm{TN}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}+N_{\mathrm{TN}}} \tag{17}$$

$$\zeta_{\mathrm{Precision}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}} \tag{18}$$

$$\zeta_{\mathrm{Recall}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}} \tag{19}$$

$$\zeta_F=2\times\frac{\zeta_{\mathrm{Precision}}\times\zeta_{\mathrm{Recall}}}{\zeta_{\mathrm{Precision}}+\zeta_{\mathrm{Recall}}} \tag{20}$$

其中,N_{TP} 为真正类,表示模型正确地将正例预测为正例的数量;N_{FP} 为假正类,表示模型错误地将负例预测为正例的数量;N_{TN} 为真负类,表示模型正确地将负例预测为负例的数量;N_{FN} 为假负类,表示模型错误地将正例预测为负例的数量。

通过引入损失函数去寻找最优的学习参数集,并衡量每个小批的实际输出和真实输出之间的差异。选

择交叉熵函数为损失函数以完成多标签分类。采用小批量梯度下降法对网络进行训练以提高收敛速度和避免过拟合。实验中,每个小批量包含 256 个数据实例。在每次迭代中,从训练集中随机选择固定数量的训练样本,即一个 mini-batch,来计算梯度。按照机器学习的惯例,将每批 7/10 的数据分成训练集,3/10 的数据分成验证集。然后,使用 Adam 优化器,初始学习率为 0.001,patience 为 5。所有的测量在训练前首先使用 z-score 方法进行标准化,将每个量测量都标准化为平均值为 0,标准差为 1 的测量数据。实验平台环境为 Windows 11-X64 操作系统,显卡为 NVIDIA-GeForce-GTX-3060,16GB 运行内存,处理器为 12th Gen Intel (R) Core (TM) i5-12600KF 3.70 GHz。在 Pytorch 1.10.1 框架、CUDA11.1 环境下实验。为进一步说明所提模型优越性,采用 LSTM、GRU 模型在相同工况下进行对比实验,实验结果分别记录为 LSTM、GRU。

2.3 实验结果及分析

(1) IEEE-14。首先在已知所有拓扑信息并成功完成 FDIA 的 IEEE-14 节点进行 FDIA 检测。对每个状态量进行分类,0 表示该状态量正常,1 表示受到攻击,通过辨别每个状态量是否受到攻击来获取受攻击的精确位置。其中量测量为 104 维,状态量为 28 维,量测数据为节点有功、无功,支路有功、无功,状态量为系统节点电压的幅值以及相角。比较 SBIGRU-HA 在不同 BIGRU 层数下得到的检测指标,结果如表 1 所示。

表 1 在 IEEE-14 节点系统的性能比较结果

Table 1 Performance comparison results on the IEEE-14 node system

结 构		精确率	召回率	*F* 分数	准确率
SBIGRU-HA	1 层 BIGRU	0.984 6	0.978 2	0.981 4	0.985 6
	2 层 BIGRU	0.985 4	0.979 5	0.982 5	0.986 5
	3 层 BIGRU	0.986 8	0.979 7	0.983 2	0.987 0
	4 层 BIGRU	0.986 3	0.980 5	0.983 4	0.987 1
	5 层 BIGRU	0.986 5	0.977 8	0.982 1	0.986 2
GRU		0.953 6	0.939 7	0.946 6	0.958 0
LSTM		0.955 7	0.937 3	0.946 4	0.959 6

表 1 中可以得出,当 BIGRU 的层数从 1 层增加到 3 层时,4 个指标也都略有增加,而当 BIGRU 的层数从 3 层增加到 5 层时,指标略有下降。这被称为退化问题:随着网络深度的增加,精度趋于饱和,然后迅速退化。考虑增加网络层数会增加网络的复杂度,从而增加训练时间,综合比较后选择 BIGRU 层数为 3。显然 SBIGRU-HA 的检测结果在精确率、召回率、*F* 分数、准确率 4 个指标上均优于传统的 GRU 和 LSTM。

采用 3 层 BIGRU 进行检测,在 IEEE-14 节点系统进行仿真,得出每个状态量的 *F* 分数、准确率,分别如图 6、图 7 所示。

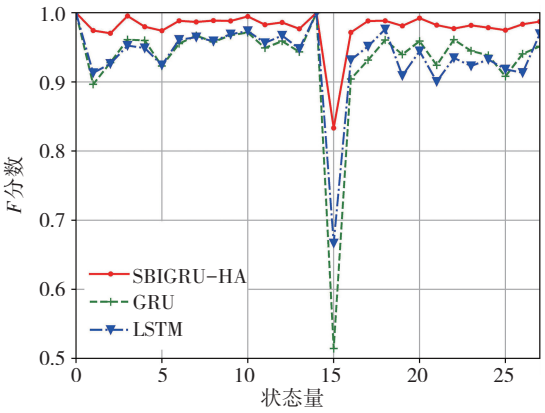


图 6 IEEE-14 节点系统各状态量检测的 *F* 分数

Fig. 6 *F* scores for the detection of each state variable in the IEEE-14 node system

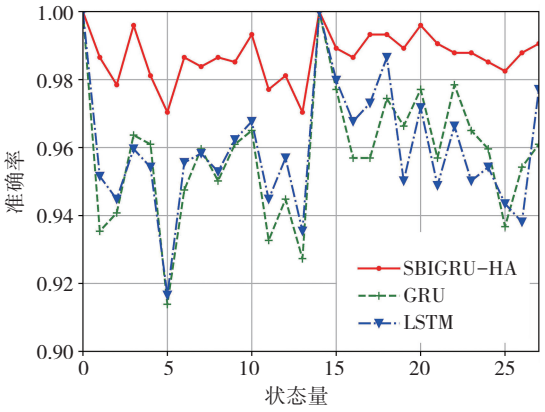


图 7 IEEE-14 节点系统各状态量检测的准确率

Fig. 7 Accuracy of detection for each state variable in the IEEE-14 node system

图 6 中横坐标按照先电压相角后电压幅值的顺序排列,其中 0 号和 14 号状态量分别为平衡节点的电压相角和幅值。从图 6 中可以看出,SBIGRU-HA 在各个状态量的 *F* 分数均优于传统的 LSTM、GRU,相对而言 GRU 的效果最差,其中相差最大的是 1 号和 15 号状态量。1 号状态量的检测 *F* 分数从原先的 89.5%上升到 97.3%,相比于 LSTM 的结果提升 7.8%;而 15 号状态量的检测 *F* 分数由原先的 51.6%提升到 83.3%,相比于 LSTM 的结果足足提升 31.7%。从稳定性而言,LSTM 和 GRU 在各个状态量的 *F* 分数波动较大,其中 GRU 的最低 *F* 分数 仅为 51.5%,而 LSTM 的最低 *F* 分数 为 66.9%,最高准确率都能达到 100%,而 SBIGRU-HA 最低 *F* 分数却能达到 83.3%,最高准确率也能达到 100%,明显稳定于上述两种传统模型。图 6 中 0 号和 14 号状态量检测的 *F* 分数之所以能达到 100%,是因为平衡节点一旦被攻击将很容易被检测,因此攻击者往往是在不攻击平衡节点这一假设前提下发动的。

从图 7 中可以得出,SBIGRU-HA 准确率明显优于

LSTM、GRU。具体而言,LSTM、GRU 的最低检测准确率出现在状态量 5 号,仅为 91.3% 和 91.6%,而在此点 SBIGRU-HA 却能达到 97%。除去参考节点,LSTM 的最高检测准确率出现在状态量 18 号,达到 98.6%,而在此点 SBIGRU-HA 却依旧达到 99.3%;GRU 的最高检测准确率出现在状态量 22 号,达到 97.8%,SBIGRU-HA 结果为 98.8%。SBIGRU-HA 在每个状态量的检测准确率全部优于 LSTM、GRU。从稳定性而言,SBIGRU-HA 仅有 5 号和 13 号状态量的检测率为 97%,其他状态量的检测准确率均在 97% 以上。GRU 和 LSTM 的稳定性相对而言较差,其检测准确率的最低点都出现在状态量 5 号,最低准确率分别为 91.4%、91.7%。换而言之,GRU 和 LSTM 的检测结果没有 SBIGRU-HA 提供的检测结果可信度高。

(2) IEEE-57。在实践中,因为完整的网络信息不仅是保密的,且在控制中心受到高度保护,获取完整的网络信息对于攻击者而言既昂贵又不现实。在已知部分拓扑信息并成功完成 FDIA 的 IEEE-57 节点进行 FDIA 检测。其中量测量为 419 维,状态量为 114 维,量测数据也为节点有功、无功,支路有功、无功,状态量为系统节点电压的幅值以及相角。比较 SBIGRU-HA 在不同 BIGRU 层数下得到的检测指标,结果如表 2 所示。

表 2　在 IEEE-57 节点系统的性能比较结果

Table 2　Performance comparison results on the IEEE-57 node system

结　构		精确率	召回率	*F* 分数	准确率
SBIGRU-HA	1 层 BIGRU	0.960 5	0.943 3	0.951 8	0.966 5
	2 层 BIGRU	0.961 5	0.943 9	0.952 6	0.967 0
	3 层 BIGRU	0.963 6	0.944 3	0.953 9	0.967 9
	4 层 BIGRU	0.963 3	0.942 0	0.952 5	0.967 0
	5 层 BIGRU	0.963 1	0.942 8	0.952 9	0.967 2
GRU		0.935 8	0.921 8	0.928 7	0.951 2
LSTM		0.931 1	0.916 4	0.9237	0.946 6

从表 2 中可以看出,BIGRU 的层数从 1 层增加到 3 层时,精确率、召回率、*F* 分数、准确率都略有增加;而当层数从 3 到 4 时,指标都略有下降;层数增加到 5 层时,精确率相对下降,但召回率、*F* 分数、准确率都略有上升。综合考虑各指标,以及模型复杂度,最终选择 BIGRU 层数为 3 层。当数据维度上升,SBIGRU-HA 模型检测的各项指标相对低维都有所下降,但其与传统 GRU、LSTM 的模型依旧有提升。具体而言,精确率分别相对 GRU、LSTM 分别提高 2.78%、3.25%,召回率分别提高 2.25%、2.79%,*F* 分数分别提高 2.52%、3.02%,准确率分别提高 1.67%、2.13%。

采用 3 层 BIGRU 进行检测,在 IEEE-57 节点系统进行仿真,得出每个状态量的 *F* 分数、准确率的结果,分别如图 8、图 9 所示。

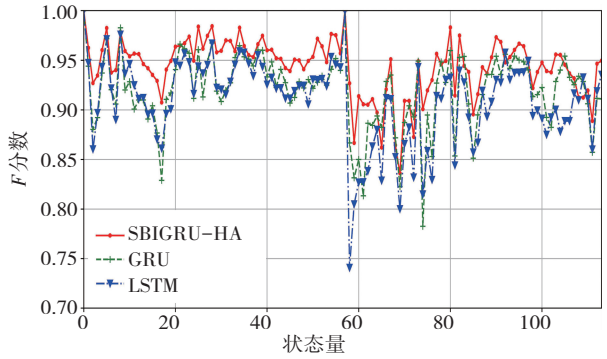


图 8　IEEE-57 节点系统各状态量检测的 *F* 分数

Fig. 8　*F* scores for the detection of each state variable in the IEEE-57 node system

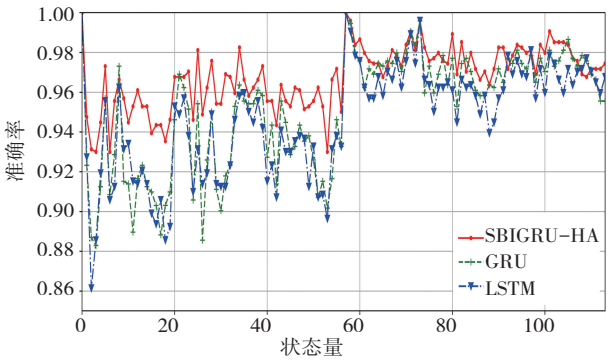


图 9　IEEE-57 节点系统各状态量检测的准确率

Fig. 9　Accuracy of detection for each state variable in the IEEE-57 node system

由图 8 可知,整体而言 SBIGRU-HA 在各个状态量的 *F* 分数优于传统 GRU、LSTM,但有部分点效果差于它们,比如 83、92、97、105、108、109 号状态量,其中有明显的差距的是 108、109。具体而言 SBIGRU-HA、GRU、LSTM 在 108 号状态量的 *F* 分数分别为 0.912 5,0.933 9,0.909 9,与 GRU 的结果相差最大,达到 2.14%;在 109 号状态量的 *F* 分数分别为 0.912 4、0.928 6、0.932 8,与 LSTM 的结果相差最大,达到 2.04%。图 8 中 0 号、57 号状态量分别为平衡节点的幅值和相角,除去平衡节点,SBIGRU-HA 稳定性优于传统 GRU、LSTM,模型检测效果最差的点出现在 69 号,结果为 0.836 3,而 GRU、LSTM 的结果分别为 0.823 1、0.800 7。换而言之,在 SBIGRU-HA 检测效果最差点的结果依旧好于 GRU、LSTM 的结果。而 LSTM 效果最差点出现在 58 号,其结果为 0.741 2,SBIGRU-HA 检测结果为 0.926 9,相对其而言提升 18.57%,GRU 效果最差点出现在 74 号,其结果为 0.782 2,SBIGRU-HA 结果为 0.901 4,相对其而言提升 11.92%。

从图 9 可以看出,LSTM 的检测准确率均在 86% 以上,GRU 的检测准确率均在 88% 以上,而考虑时空特征并融合注意力的 SBIGRU-HA 模型检测准确率均在 92% 以上,除去参考节点,检测效果最优的点都出现在 73 号,检测结果都达到 99% 以上,最高检测率和最低检

测率相差分别达到13%、11%、7%,可在从稳定性而言,SBIGRU-HA模型的效果优于GRU、LSTM。个别检测点的准确率效果差于LSTM、GRU,如8、65、78、83、92、108、109号,LSTM在109号状态量效果优于SBIGRU-HA最多,效果分别为97.71%、96.77%,也就相差0.94%;GRU在8号状态量效果优于SBIGRU-HA最多,效果分别为97.31%、96.64%,也就相差0.67%,可见SBIGRU-HA不仅在准确率上优于GRU、LSTM,在稳定性上也更胜一筹。

3 结 论

针对当前智能电网FDIA检测的时间数据相关性的问题,提出一种融合SBIGRU与三种注意力机制的检测模型,并在IEEE-14以及IEEE-57系统进行仿真实验。所采用数据集中的攻击模型不仅分别考虑已知全部拓扑信息、部分拓扑信息的情况,而且还考虑PMU量测对攻击的影响。SBIGRU-HA不仅从连续系统状态中呈现的时间数据相关性中学习,而且还考虑数据的空间特征,在两个公共电网上的综合实验结果表明,SBIGRU-HA在IEEE-14系统上相比于GRU、LSTM准确率分别提升2.9%、2.74%,在IEEE-57系统则分别提升1.67%、2.13%。在电网攻击检测中相比于误检,漏检对电网的危害更大,因此需更加关注综合指标F分数。SBIGRU-HA在IEEE-14系统上相比于GRU、LSTM的F分数则分别提升3.66%、3.68%,在IEEE-57系统则分别提升2.52%、3.02%。综上,SBIGRU-HA能更好完成FDIA检测。

但在设计SBIGRU-HA模型时,未考虑训练时间、检测时间,未来将进一步权衡检测准确率、检测时间、模型复杂度和训练时间等多种因素,以及改进BIGRU的网络结构并采用更先进的注意力机制,降低模型的假阳性和阴性率。

参考文献(References):

[1] YU X, XUE Y. Smart grids: A cyber-physical systems perspective[J]. Proceedings of the IEEE, 2016, 104(5): 1058-1070.

[2] KHURANA H, HADLEY M, LU N, et al. Smart-grid security issues[J]. IEEE Security and Privacy, 2010, 8(1): 81-85.

[3] LIU Y, NING P, REITER M K. False data injection attacks against state estimation in electric power grids [J]. ACM Transactions on Information and System Security, 2011, 14(1): 13.

[4] YUAN Y, LI Z, REN K. Modeling load redistribution attacks in power systems[J]. IEEE Transactions on Smart Grid, 2011, 2(2): 382-390.

[5] 杨怡, 王勇. 基于AUKF的分布式电源系统虚假数据攻击检测方法[J]. 电工电能新技术, 2021, 40(12): 48-55.
YANG Yi, WANG Yong. Detection method of false data attack in distributed generation system based on AUKF[J]. Advanced Technology of Electrical Engineering and Energy, 2021, 40(12): 48-55.

[6] 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 Xin-rui, CHANG Peng, SUN Qiu-ye. 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.

[7] MA C, LIANG H, JING Y. A novel ZSV-based detection scheme for FDIAs in multiphase power distribution systems[J]. IEEE Transactions on Smart Grid, 2023, 14(2): 1236-1248.

[8] LUO X, LI Y, WANG X, et al. Interval observer-based detection and localization against false data injection attack in smart grids[J]. IEEE Internet of Things Journal, 2021, 8(2): 657-671.

[9] 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.

[10] HUANG K, XIANG Z, DENG W, et al. False data injection attacks detection in smart grid: A structural sparse matrix separation method[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(3): 2545-2558.

[11] CHAKRABARTY S, SIKDAR B. Unified detection of attacks involving injection of false control commands and measurements in transmission systems of smart grids[J]. IEEE Transactions on Smart Grid, 2022, 13(2): 1598-1610.

[12] GOYEL H, SWARUP K S. Data integrity attack detection using ensemble-based learning for cyber-physical power systems [J]. IEEE Transactions on Smart Grid, 2023, 14(2): 1198-1209.

[13] LI Y, WEI X, LI Y, et al. Detection of false data injection attacks in smart grid: A secure federated deep learning approach[J]. IEEE Transactions on Smart Grid, 2022, 13(6): 4862-4872.

[14] CUI J, GAO B, GUO B. A novel detection and defense mechanism against false data injection attack in smart grids[J]. IET Generation, Transmission & Distribution, 2023, 17(20): 4514-4524.

[15] WU T, XUE W, WANG H, 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.

[16] LIU Z, WANG Q, YE Y, et al. AGAN-based data injection attack method on data-driven strategies in power systems[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 3203-3213.

[17] MUSLEH A S, CHEN G, DONG Z Y, et al. Attack detection in automatic generation control systems using LSTM-based stacked autoencoders [J]. IEEE Transactions on Industrial Informatics, 2023, 19(1): 153-165.

责任编辑:代小红