基于深度信念网络和迁移学习的隐匿FDI攻击入侵检测

郭方洪, 易新伟, 徐博文, 董 辉†, 张文安

(浙江工业大学 信息工程学院，杭州 310014)

摘 要: 成功地检测隐匿虚假数据入侵(false data injection, FDI)攻击是确保电力系统安全运行的关键. 然而,大多数工作通过建立FDI攻击模型模拟真实的入侵行为,得到的模拟数据往往与真实数据存在一定的差异,导致基于机器学习的检测方法出现较差的学习效果. 鉴于此,针对源域中模拟样本数据量大而目标域中真实样本标记少的特点,提出基于深度信念网络 (DBN) 和迁移学习的检测算法. DBN 中的受限玻尔兹曼机 (restrict boltzmann machine, RBM)能够对海量目标域无标签样本进行特征自学习,基于模型的迁移学习方法可以克服数据之间的差异性,同时解决有标签真实样本稀缺的问题. 最后,在IEEE 14-bus电力系统模型上验证了所提出方法的优点和有效性.

关键词: 智能电网；隐匿虚假数据入侵攻击；深度信念网络；迁移学习; 无监督学习

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2020.1469

引用格式: 郭方洪, 易新伟, 徐博文, 等. 基于深度信念网络和迁移学习的隐匿FDI攻击入侵检测 [J]. 控制与决策, 2022, 37(4): 913-921.

Stealthy FDI attack detection based on deep belief network and transfer learning

GUO Fang-hong, YI Xin-wei, XU Bo-wen, DONG Hui†, ZHANG Wen-an

(College of Information Engineering，Zhejiang University of Technology，Hangzhou 310014，China)

Abstract: Successful detection of false data injection (FDI) attacks are essential for ensuring secure power grids operation. However, most work simulates real intrusion behaviors by establishing FDI attack models, and the simulated data obtained is often different from the real data, resulting in poor learning effects based on machine learning detection methods. Motivated by this fact, considering the large amount of simulated sample data in the source domain and a small number of labeled real samples in the target domain, a detection algorithm based on the deep belief network (DBN) and transfer learning is proposed. The restrict boltzmann machine (RBM) in the DBN can automatically extract features from a large number of unlabeled samples in the target domain, and the model-based transfer learning method overcomes the differences between data and solves the problem of the scarcity of labeled real samples. Finally, the IEEE 14-bus power system is employed to show the advantages and effectiveness of the proposed method.

Keywords: smart grid；stealthy false data injection attack；deep belief network；transfer learning；unsupervised learning

0 引 言

传统电网正经历着向智能电网的巨大演变,智能电网以集成、高速和双向通信网络为基础,通过先进的传感和测量技术、先进的控制方法和大数据技术,实现了更安全、更高效的电力管理[1-2]. 广泛的通信基础设施已用于传输和监控配电系统连接点的电网运行参数. 然而,网络的开放性也让智能电网更容易受到各种恶意网络攻击,如拒绝服务攻击[3]、负载重分布攻击[4]和隐匿虚假数据入侵(FDI)攻击等. 与其他攻击不同的是,隐匿FDI攻击可以绕过坏数据检测机制,通过篡改测量数据,使得控制中心获得错误的系统运行状态而做出错误的决策,最终达到获取经济利益或者其他非法目的[5-6].

FDI攻击目前已经得到了广泛的研究,国内外学者提出了针对FDI攻击的保护对策和防御方案. 文献[7]提出了一种基于关键传感器的保护机制,确定了检测攻击所需的相角测量单元(phasor measurement units, PMU)最小数目及其位置,以节省大规模部署

收稿日期: 2020-10-25；**录用日期:** 2021-02-10.

基金项目: 国家自然科学基金青年基金项目(61903333)；浙江省“钱江人才”特殊急需类项目(QJD1902010).

责任编委: 孙秋野.

†通讯作者. E-mail: hdong@zjut.edu.cn.

PMU的费用. 文献[8]表示PMU对GPS的依赖使其更容易遭受攻击. 为此,Wu等[9]提出了一种考虑攻击行为不确定性的广义攻击分离方案. 现有的大多数研究[10-11]均假设攻击者拥有对整个系统网络拓扑和参数的访问权限,然而实际场景中获得整个系统知识是非常困难的. 文献[12]研究了在网络信息不完全情况下的攻击策略,提出可以通过提高网络耦合度降低被攻击的风险.

另一方面,在针对FDI攻击的检测上,文献[13]将等效测量变换与最大加权残差法相结合以检测FDI攻击. 文献[14]提出一种基于机器学习的方法,通过训练分布式的支持向量机(SVM)对隐匿FDI攻击进行检测. 文献[15]从模糊聚类的角度出发,运用数据挖掘中的聚类分析方法区分虚假数据. 文献[16]针对交流系统中的FDI攻击,提出了一种小波变换与递归神经网络(RNN)相结合的方法,获得了较好的检测精度. 文献[17]应用条件深度信念网络(CDBN)识别以窃电为目的的隐匿FDI攻击. 文献[18]提出了一种基于负荷预测的实时FDI检测方法,根据预测值与真实值的偏差是否大于某一阈值来检测FDI攻击. 上述深度学习方法可以通过电力系统中的海量数据进行训练,有效揭示了FDI攻击特征模式. 近年来,深度学习方法在FDI检测上也取得了一定进展,值得注意的是,该方法有效的前提是训练集和测试集具有高度相似性、训练数据充足且有代表性. 由于电力系统中有标记的实际测量数据十分稀缺,目前绝大部分研究往往通过建立FDI注入攻击模型[14,17]来获得海量模拟的正常测量数据(正样本)和被攻击/篡改后的测量数据(负样本). 然而,受电网拓扑、攻击强度和测量噪声等因素的影响,模拟样本与真实样本之间存在较大的差异,通过模拟样本训练好的机器学习模型在真实样本上很可能表现出较差的学习效果.

鉴于此,本文提出了基于深度信念网络(DBN)和迁移学习的检测算法,DBN中的RBM层能对无标签样本进行特征自学习,从海量无标签样本中得到高度抽象的重要特征;而基于模型的迁移方法则能高效利用电力系统大数据,克服数据之间的差异性,挖掘不同数据集的共性,从而解决有标签真实样本稀缺的问题. 具体而言,首先利用海量目标域无标签样本对DBN网络逐层预训练,获得真实样本的分层特征表达;然后通过海量源域有标签样本再训练得到参数共享的DBN网络,将DBN网络参数迁移并对网络结构进行调整;最后利用少量目标域有标签样本对适配层进行训练和全网络微调. 利用MATPOWER和POWERWORLD上的IEEE 14-bus进行实验验证,并与模拟样本训练和少量有标注真实样本微调的DBN模型、ANN和PCA-SVM[14]模型进行对比. 实验结果表明,在源域和目标域具有不同的分布差异下,所提出方法不仅能够有效检测目标域中的隐匿FDI攻击,而且其泛化性能优于上述3种方法.

1 问题描述

1.1 隐匿FDI攻击入侵原理

电力系统控制中心通过持续监测各个仪表上的测量值来估计系统的实时状态. 由于系统状态在一段时间内变化缓慢,将潮流方程在操作点附近做泰勒展开,非线性的交流模型近似为线性的直流模型[19],最终可以表示为

$$z = Hx + e. \tag{1}$$

其中:$z \in \boldsymbol{R}^m$为测量值,包括总线电压、有功功率和无功功率;$x \in \boldsymbol{R}^n$为状态变量,包括总线电压和电压相角;$e \in \boldsymbol{R}^m$为测量噪声,且e服从均值为0、协方差矩阵为W的高斯分布;$H \in \boldsymbol{R}^{m\times n}$为雅可比矩阵. 使用加权最小二乘法获得估计的状态变量$\hat{x}$[20],有

$$\hat{x} = (H^{\mathrm{T}}W^{-1}H)^{-1}H^{\mathrm{T}}W^{-1}z. \tag{2}$$

隐匿FDI攻击是指攻击者通过篡改测量值实现隐藏攻击,以欺骗电力系统控制中心,使系统产生错误的状态估计. 设$z_a = z + a$为被攻击后的测量值,a为攻击向量,如果攻击向量$a = Hc$,其中c为非零向量,且与x具有相同维度,则系统状态估计值为$\hat{x}_a = \hat{x} + c$.

隐匿FDI攻击能够绕过坏数据检测(bad data detection, BDD)机制,因为坏数据检测残差只受测量噪声影响,加入攻击向量后残差γ基本不变,即

$$\begin{aligned}\gamma &= \|z + a - H(\hat{x} + c)\|_2 = \\ &\|z - H\hat{x} + (a - Hc)\|_2 = \|z - H\hat{x}\|_2 = \\ &\|e - H(H^{\mathrm{T}}W^{-1}H)^{-1}H^{\mathrm{T}}W^{-1}e\|_2.\end{aligned} \tag{3}$$

电网一般会在关键节点配制受保护的测量仪表,这种测量仪表针对物理和网络攻击附加了额外的安全措施,确保数据不会被泄露. 因此,攻击者必须确保受保护的测量仪表在攻击向量中所对应的条目必须为零[21],才能避免被检测到. 将攻击向量$a = Hc$缩减到对应的子矩阵$a_r = H_r c$,得到

$$H_r c = 0, \tag{4}$$

其中H_r为雅可比矩阵H的子矩阵. 当式(4)只有零解时,意味着攻击者无法在关键节点进行FDI攻击,从而保护电网安全. 然而,由于受保护仪表高昂的成

本,不可能大规模部署,防御者一般只能在少数关键节点实施保护.

1.2　现有检测方法的局限性

由于电力系统的复杂性与攻击特征的多样性,难以用传统解析的方法构造隐匿FDI的攻击特征与系统故障的关系. 而深度学习方法借助海量数据进行训练,能有效揭示攻击特征模式,进而对传感器测量进行分类检测. 在电力系统中存在海量无标签样本,而有标记的攻击样本和正常样本十分稀缺,通过在实际电网中加入攻击向量采集样本是不现实的. 研究人员主要通过采集历史负荷和估计网络拓扑信息模拟电网正常运行,然后构建攻击模型来生成模拟的攻击样本和正常样本. 然而受以下因素的影响,模拟样本与真实样本之间难以服从相同的分布,使用模拟样本训练的机器学习模型很可能在真实样本上表现出较差的学习效果:

1) 网络拓扑估计误差. 雅可比矩阵 H 大小取决于当前时刻的总线电压、电压相角、支路电导、支路电纳和电路与地面之间的电纳[22]. 由于电力负荷和网络参数扰动的不确定性, H 也随之变化. 研究者往往难以准确地获得支路的电阻和电抗等参数,估计的雅可比矩阵 H 将不可避免地存在一定的误差.

2) 攻击强度误差. 攻击强度是指隐匿FDI攻击对系统状态的影响,即注入的攻击向量能对电网系统中的多个状态造成大小超过设定阈值的篡改. 攻击强度主要取决于攻击向量,由于防御者难以事先判断哪些总线或者状态容易受到攻击,模拟的攻击区域以及虚假数据注入大小都难以确定,其注入的攻击向量将与黑客注入的攻击向量存在差异.

3) 测量噪声误差. 在实现隐匿FDI攻击的前提下,随着模拟样本与真实样本噪声差异变大,两个分布之间的数据相关性将逐渐降低.

网络拓扑估计误差和测量噪声误差因素对模拟的攻击样本和正常样本都会造成影响,而攻击强度误差因素只对模拟的攻击样本造成影响. 由于电网不可避免地偏离最佳运行条件,网络拓扑估计误差和测量噪声误差一般处于具有特定范围的变化区间之内[23]. 设 τ 为模拟的正常样本最大误差, a^+ 为注入的攻击向量,则模拟的正常测量值 z^+ 和被攻击后的测量值 z_a^+ 可以表示为

$$z(1-\tau\%) \leqslant z^+ \leqslant z(1+\tau\%),$$
$$z(1-\tau\%)+a^+ \leqslant z_a^+ \leqslant z(1+\tau\%)+a^+. \quad (5)$$

2　基于深度信念网络和迁移学习的检测机制

为了克服上述缺点,提出如图1所示的检测机制,具体的检测算法流程如图2所示.

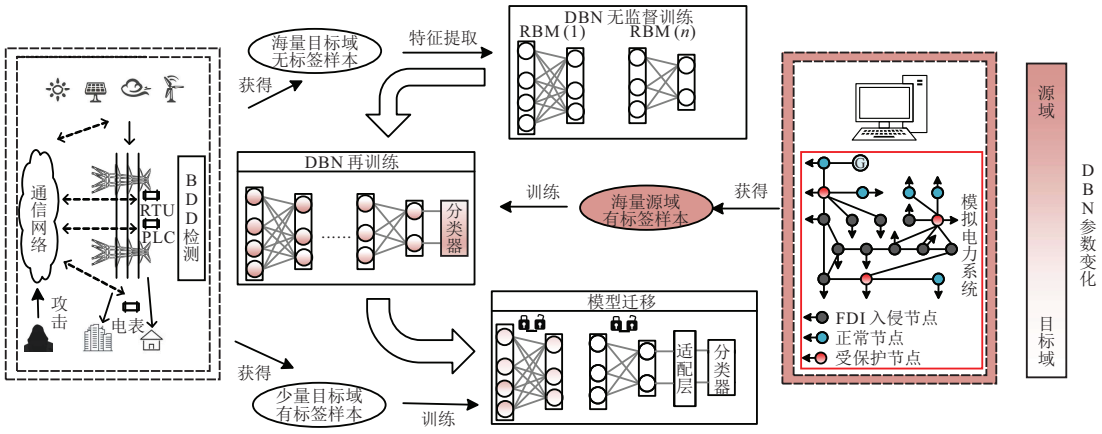

图 1　基于深度信念网络和迁移学习的检测机制

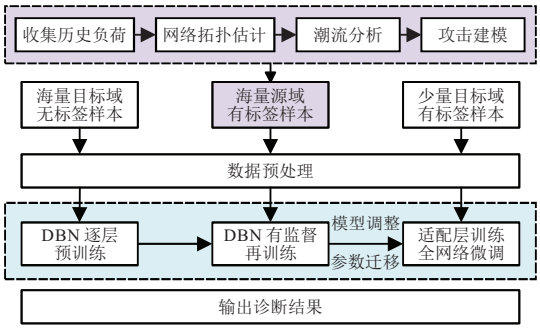

图 2　基于深度信念网络和迁移学习的检测算法流程

首先使用海量目标域无标签样本进行DBN的逐层无监督预训练,得到高度抽象的重要特征;然后利用海量源域有标签样本进行再训练得到参数共享的DBN网络,将参数迁移到目标域网络并冻结,增加或替换隐藏层,获得新的学习空间;接着利用少量目标域有标签样本对增加的适配层进行训练;最后取消冻结,再次利用少量目标域有标签样本进行参数微调得到最终模型.

2.1 DBN无监督预训练

电力系统中存在着海量无标签样本,进行数据挖掘将有助于揭示FDI攻击特征模式,本文通过DBN逐层无监督预训练提取这部分数据的深层特征. DBN网络由多个RBM组成[24],当最后一个RBM预训练结束时,完成对DBN网络参数的初始化. RBM结构如图3所示,若一个RBM包含m个可视层单元,n个隐藏层单元,则可以定义RBM的能量表达式为

$$E(v,h|\theta)=-\sum_{i=1}^{m}a_iv_i-\sum_{j=1}^{n}b_jh_j-\sum_{i=1}^{m}\sum_{j=1}^{n}v_iw_{i,j}h_j. \tag{6}$$

其中:$\theta=a_i$、b_j、$w_{i,j}$为RBM的网络参数,a_i为可见层单元v_i的偏置,b_j为隐藏层单元h_j的偏置,$w_{i,j}$为可见层单元v_i与隐藏层单元h_j的连接权重. 由式(6)定义的能量函数,给出状态为(v,h)、网络参数为θ时的联合概率密度分布为

$$P(v,h|\theta)=\frac{\mathrm{e}^{-E(v,h|\theta)}}{Z(\theta)}, \tag{7}$$

其中$Z(\theta)=\sum_{v,h}\mathrm{e}^{-E(v,h|\theta)}$为归一化因子. 可视层数据$v$的概率分布$P(v|\theta)$对应于$P(v,h|\theta)$的边缘分布为

$$P(v|\theta)=\frac{1}{Z(\theta)}\sum_{h}\mathrm{e}^{-E(v,h|\theta)}. \tag{8}$$

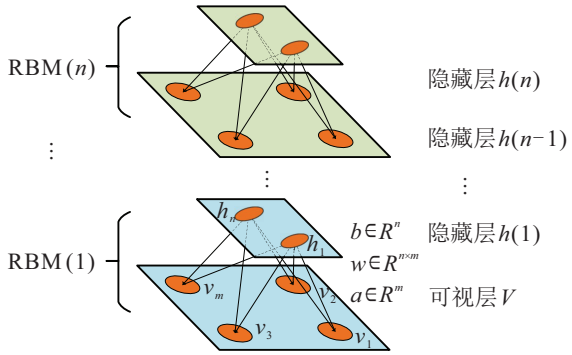

图3 受限玻尔兹曼机(RBM)结构示意图

训练RBM的目标是通过调整参数θ使RBM的概率分布尽可能与输入数据相符合. 设训练样本个数为S,通过随机梯度上升法最大化似然函数$\ln\prod_{i=1}^{S}P(v^i|\theta)$对参数$\theta$求偏导,可以得到

$$\frac{\partial\ln\prod_{i=1}^{S}P(v^i|\theta)}{\partial\theta}=\sum_{i=1}^{S}\Big(-E_{p(h|v^i,\theta)}\Big[\frac{\partial E(v^i,h|\theta)}{\partial\theta}\Big]+E_{p(v,h|\theta)}\Big[\frac{\partial E(v,h|\theta)}{\partial\theta}\Big]\Big), \tag{9}$$

其中:$E_{p(\cdot)}$为关于分布p的数学期望,$p(h|v^i,\theta)$是可视单元为样本v^i时的隐藏层概率分布,$p(v,h|\theta)$为可视层与隐层的联合分布. 利用对比散度算法可以求得RBM最优的参数θ.

由于RBM可见层单元和隐藏层节点单元均为二值变量,传感器测量值为连续分布的实值数据,直接将RBM模型用于攻击测量诊断会难以提取数据的非线性特征. 将DBN模型中的第1个RBM替换为高斯伯努利受限玻尔兹曼机(GBRBM)[25],GBRBM是在RBM的基础上将可见层二值节点替换为带独立高斯噪声的实值变量节点,隐层节点仍保留为二值变量节点.

2.2 模拟测量生成及DBN再训练

电力系统中有标签的测量数量十分稀少,这部分数据不足以训练深度神经网络,因此借助海量源域有标签样本对DBN网络进行再训练.

1) 模拟测量数据集生成. 为了模拟FDI攻击动态数据注入过程,收集电网历史负荷及发电机注入功率数据,通过估计网络拓扑参数再进行分时潮流计算来模拟电网实际运行状态,得到每一时刻总线的电压幅值和相角,进而求出雅可比矩阵H. 当受保护的测量维数r小于状态维数时,由式(4),通过求解H_r的零空间,可以构造满足隐匿FDI攻击的攻击向量a. 需要注意的是,电力系统中各种电气参数的波动会增加FDI攻击被检测到的风险,可以通过构建高度稀疏的攻击向量a最小化篡改测量仪表的数量,最终获得整个模拟测量值数据集$(Z_{\mathrm{FDI},t^{(k)}},Z_{t^{(k)}})_{k=1}^{K}$,其中$k$按照时序关系进行索引.

2) DBN有监督再训练. 根据模拟测量数据集,对预训练好的DBN网络采用反向误差传播算法(BP)对每个隐藏层的权重、偏置进行调整,直至模拟样本测试集准确率基本不变时,完成对网络参数的训练.

2.3 模型迁移

基于模型的迁移方法能够在源域和目标域中共享一些参数信息,有效解决因训练数据少带来的问题[26-27]. 在上述训练好的DBN模型的基础上,结合少量目标域有标签的样本进行迁移. 然而,模拟测量数据与真实测量数据存在一定的分布差距,照搬DBN网络模型结构和参数很可能出现较差的学习效果. 如何迁移已有知识以适配目标域数据以及最大化目标域数据价值是模型迁移的关键. 本文所提出方法在全连接层前增加了batch normalization(BN)层,GBRBM和多个堆叠的RBM能最大程度拟合训

练数据的概率分布, BN 层则能避免前向网络参数变化导致后面全连接层输入数据的分布变化. 另外, 由于有标签真实样本较少, 加入 BN 层有利于解决模型出现过拟合的问题. 如图 4 所示, 当参数迁移后, 在 softmax 分类器前增加或替换全连接层[28], 一方面可以让网络学习新的知识, 另一方面也可以通过调节全连接层的权值保存或者舍弃部分源网络的特征, 达到了既保存源网络信息又能进一步学习目标域信息的目的. 由于目标域有标签的真实样本数量较少, 不足以训练出泛化能力足够强的深度神经网络, 结合深度网络迁移方法中的 finetune 思想, 将网络其他层参数进行冻结, 利用少量目标域有标签样本对增加的 BN 层和全连接层进行训练, 网络训练至测试集准确率处于基本不变的稳定收敛状态时取消冻结. 尽管增加的适配层能够学习目标域信息, 但这是建立在源域训练好的 DBN 模型基础上的, 再次利用少量目标域有标签样本以较小的学习率微调整个网络, 直到测试集准确率处于基本不变的稳定收敛状态, 获得最终模型.

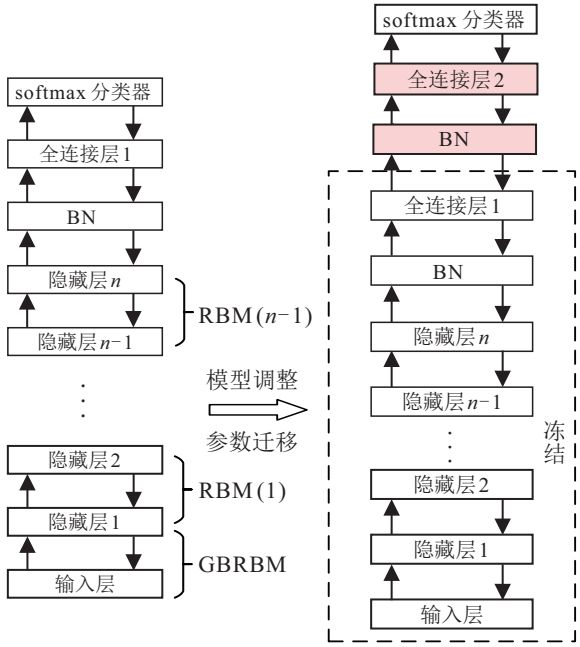

图 4　基于深度信念网络的迁移学习模型

3　仿　真

利用如图 5 所示的 IEEE 14-bus 系统验证基于深度信念网络和迁移学习检测机制的优点和有效性. 收集纽约独立运营商 (NYISO) 2020 年 2 月 ~ 6 月的负荷数据, 其中 11 个区域代表 IEEE 14-bus 的 11 个负荷节点, 随后对负荷数据进行缩放以匹配模拟系统中的电力需求规模. 图 6 为母线 3 上的负荷波动曲线. 通过 POWERWORLD 进行时序潮流计算得到电网系统状态, 将潮流方程线性化求解雅可比矩阵, 最后利用 MATPOWER 工具箱计算得到系统的量测. IEEE 14-bus 系统状态变量维数为 27, 测量变量维数为 54. 将这些测量信息作为模型的输入, 用于检测注入的隐匿 FDI 攻击.

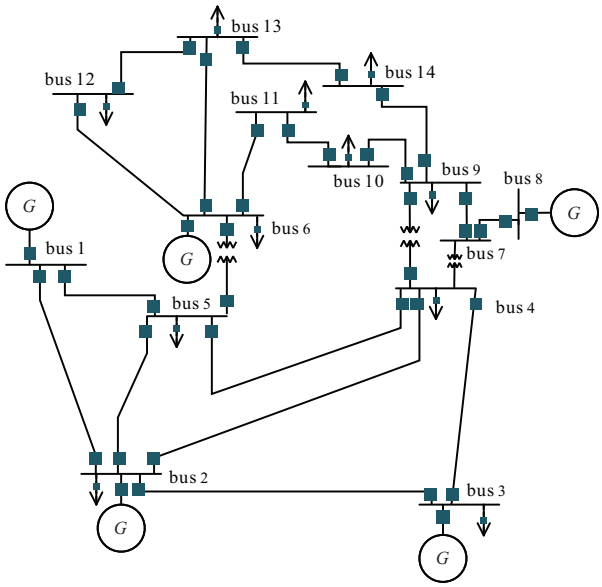

图 5　IEEE 14-bus 系统

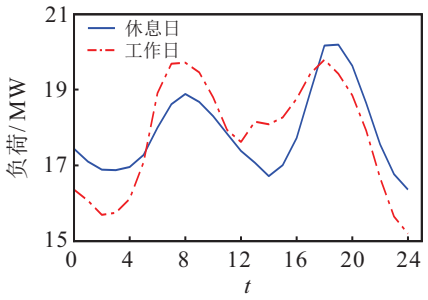

图 6　母线 3 的负荷曲线

3.1　仿真设置

仿真实验中, 以标准 IEEE 14-bus 系统模拟真实电网运行, 所得到的测量值为目标域数据. 为研究抵抗 FDI 攻击的策略, 防御者对真实电网进行攻击建模分析, 不可避免地存在拓扑估计偏差, 因此假设其支路电阻、电抗估计偏差在 10 % 以内, 所得到的测量值为源域数据. 取受保护的测量维数 $r = 2, 4, 6$, 相应地, 攻击者能够选取不同数量的量测节点进行篡改, 从而达到不同的攻击强度. 另外, 设置不同的环境噪声 $e(o, \sigma), \sigma = 0.2, 0.4, 0.6, 0.8, 1$. 从 2 月 ~ 6 月, 在每个整时构建攻击向量 a, 根据不同的 r、σ 和网络拓扑重复 60 次, 分别生成攻击样本和未被攻击的样本, 最后得到目标域无标签样本、有标签样本和源域有标签样本, 并按照 7:3 的比例划分训练集和测试集.

在标准 IEEE 14-bus 系统中, 设置 $r = 2, \sigma = 0.2$, 从目标域数据中挑选一个正常测量, 构建攻击向量, 得到一个隐匿 FDI 攻击测量数据. 攻击对电压相

角状态估计的影响如图7所示. 可见,在节点9、13和14出现了少量偏差. 由于电力系统中电压存在周期性波动,模拟了系统在24小时内的运行状态,并在8:00～10:00和18:00～19:00整点加入隐匿FDI攻击,其对节点电压幅值的影响如图8所示. 节点3、4和5均出现不同程度的测量偏差. 然而坏数据检测残差却变化很小,以18时刻测量为例,加入攻击前残差为0.054 7,加入攻击后残差为0.062 8.

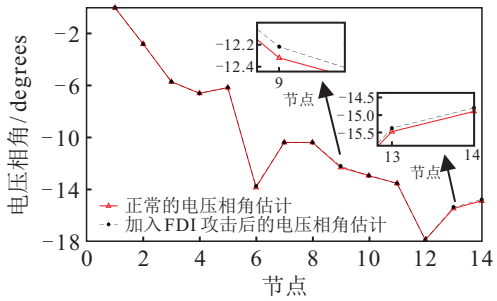

图7 隐匿FDI攻击下的状态估计

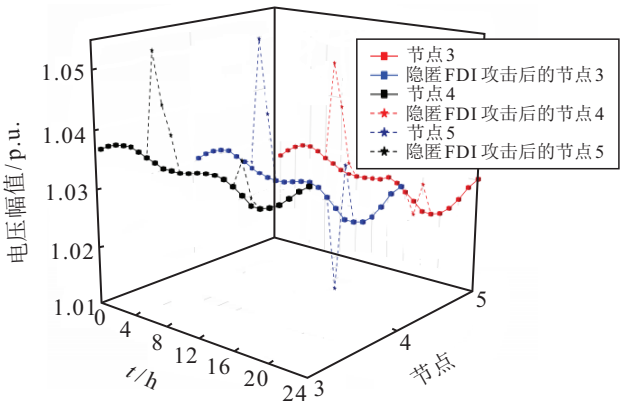

图8 隐匿FDI攻击下电压幅值曲线

源域和目标域测量分别在不同的场景下获得,源域和目标域数据分布差异较大可能会影响模型迁移的效果. 为了研究受保护测量维数t和环境噪声σ对训练样本分布的影响,设置$r=2,\ \sigma=0.4$获得目标域样本,并根据不同的r和σ获得源域样本,其中攻击样本和正常样本比例为1:1. 使用最大均值差异算法(MMD)[29]作为检验统计量衡量源域和目标域数据之间的相关性,如果MMD足够小则认为两个分布相同,反之则为不相同. 结果如图9所示,当源域和目标域中的r与σ相同时,MMD取得最小值0.219 4. 随着源域与目标域中r和σ差值变大,MMD也随之增大,两分布间的数据相关性降低.

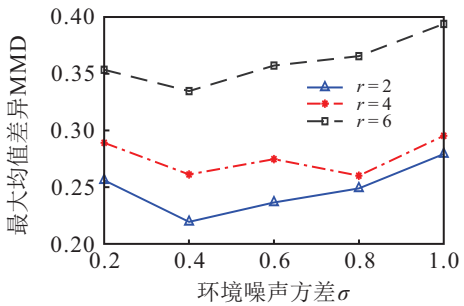

图9 不同环境噪声下的源域和目标域分布差异

3.2 仿真结果分析

设置模型隐藏层个数为6,包括1个GBRBM,3个RBM和2个全连接层. 输入到输出的节点个数为54、40、30、20、12、10、6、2. DBN预训练使用带动量的随机梯度下降优化器,动量设置为0.5,学习率为0.1,DBN再训练、适配层训练和参数微调使用随机梯度下降优化器,学习率分别为0.2、0.01和0.000 1. 本文所提出机制的优越性主要体现在两方面:其一,使用目标域无标签样本进行DBN预训练让模型在未迁移前对目标域样本进行了特征提取,使得模型参数适应于目标域样本的特征表达;其二,在少量目标域有标签的情况下,模型调整后进行适配层训练和微调获得了较好的检测效果. 对以上两点进行实验验证,并采用准确率(ACC)和假阳性率(FPR)作为评价指标,计算方式如下:

$$\begin{aligned}&\mathrm{ACC}=(\mathrm{TN}+\mathrm{TP})/(\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}),\\&\mathrm{FPR}=\mathrm{FP}/(\mathrm{FP}+\mathrm{TN}),\end{aligned}\tag{10}$$

其中TN、TP、FP、FN分别为正确分类的正常样本、正确分类的攻击样本、错误分类的正常样本和错误分类的攻击样本. 假阳性率也被称为误报率,目标是让模型检测准确率高的同时误报率低.

为了证明利用目标域无标签样本进行预训练能使模型获得更好的检测效果,利用相同数量的目标域无标签样本和源域样本进行DBN无监督预训练,然后分别在模型迁移前后对目标域测试集样本进行测试并对比,结果如表1所示. 通过模型迁移,检测精度显著提升,误报率也大幅下降. 另外,相比源域样本,使用目标域无标签样本作为预训练数据集,模型具有更高的检测精度和更低的误报率.

表1 模型迁移前后检测效果 %

预训练数据集	模型迁移前		模型迁移后	
	ACC	FPR	ACC	FPR
目标域无标签样本	63.59	41.14	98.26	0.93
源域样本	57.39	45.13	97.33	1.81

模型迁移过程分为适配层训练和微调两个步骤,目标域有标签训练样本个数对模型迭代收敛效果存在一定的影响. 从理论上讲,足够多的有标签训练样本可以直接用于训练深度模型,而不需要借助迁移学习的方法. 这里取目标域有标签样本分别为250、500、750和1 000组进行模型迁移,适配层训练和微调

过程迭代收敛效果分别如图10和图11所示. 可以发现,从适配层训练到微调,模型可以进一步收敛,且随着有标签样本个数的增加,模型收敛效果更好. 分别在适配层训练后和微调后对目标域测试集样本进行测试并对比,结果如表2所示. 从适配层训练到微调,检测精度从90%以下提升到95%以上,误报率也降低到7%以内. 另外,随着目标域有标签训练样本的增加,检测精度和误报率分别得到了不同程度的提升和下降.

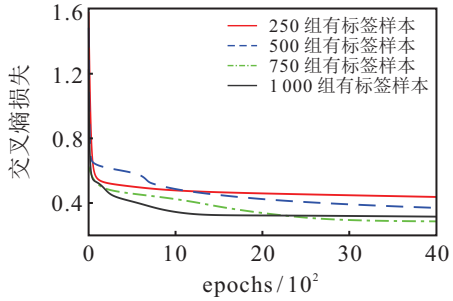

图 10　适配层训练收敛效果

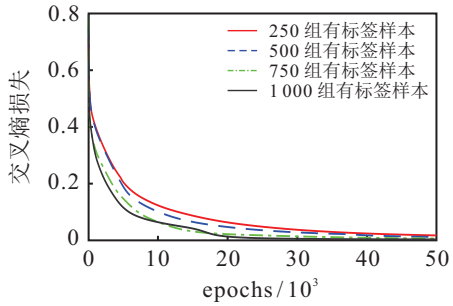

图 11　微调收敛效果

表2　适配层训练和微调检测效果　%

训练样本个数	适配层训练		微调	
	ACC	FPR	ACC	FPR
250	86.58	19.70	95.05	7.18
500	87.60	20.06	96.70	5.26
750	89.67	16.35	97.57	3.46
1000	88.54	14.38	97.69	3.11

由于电力系统中测量数据普遍存在类别不均衡的特点,设置目标域海量无标签样本正负比例为100∶1,源域海量有标签样本正负比例为1∶1,目标域测试集样本正负比例为1∶1,比较不同正负样本比例的目标域有标签训练样本下模型的检测效果如表3所示. 随着样本比例增大,模型准确度显著下降,误报率却保持在1%以内,这是模型对负样本欠拟合造成的.

表3　不同正负样本比例下的检测效果　%

正负样本比例	49∶1	19∶:1	9∶1	4∶1	1∶1
ACC	77.78	78.89	84.72	87.36	97.80
FPR	0.31	0.25	0.56	0.23	1.49

将DBN-finetune、ANN和PCA-SVM[14]模型与所提出方法进行对比,相关设定如下:

1) DBN-Finetune: DBN-Finetune与所提出方法的唯一区别是没有对训练好的DBN模型添加或替换全连接层,而是冻结所有RBM单元隐藏层,使用目标域有标签样本进行微调,微调学习率为0.000 1,当测试集准确率基本不变时完成训练.

2) ANN: 隐藏层数为2,输入到输出的节点个数为54: 30: 10: 2,激活函数为sigmoid,学习率取0.01,使用随机梯度下降优化器,为了避免过拟合,交叉熵损失函数加入L_2正则化项.

3) PCA-SVM: 特征映射后的主成分个数为2,核函数选择高斯核. 其中ANN与PCA-SVM模型分别使用源域样本、相同比例的源域和目标域有标签样本进行训练,两种数据集记为A、B. 通过多次训练和测试,最终得到模型在不同MMD下对目标域测试集的准确率和误报率.

4种检测机制在不同MMD下的检测精度和误报率如表4和表5所示. 由表4和表5可见,使用不同数据集训练的ANN模型检测效果差别很大,当源域数据和目标域数据特征分布存在一定差异时,使用源域数据训练的模型对目标域数据进行测试必

表4　4种检测机制在不同MMD下的检测精度ACC　%

MMD	本文方法	DBN-Finetune	ANN		PCA-SVM	
			A	B	A	B
0.219 4	98.54	92.47	68.45	94.36	61.49	64.32
0.256 1	98.37	92.18	66.34	94.18	58.74	62.61
0.295 8	98.24	91.13	62.62	95.37	54.12	57.35
0.334 7	97.73	89.46	63.15	94.34	54.35	54.45
0.393 9	97.33	86.53	57.24	94.12	50.97	52.86

表5 4种检测机制在不同MMD下的误报率FPR %

MMD	本文方法	DBN-Finetune	ANN		PCA-SVM	
			A	*B*	*A*	*B*
0.219 4	1.43	8.42	28.78	7.34	38.53	34.43
0.256 1	1.94	8.79	33.17	7.65	41.24	36.93
0.295 8	2.26	9.21	37.45	8.43	47.56	43.91
0.334 7	4.06	11.56	37.94	7.20	47.19	47.85
0.393 9	4.59	16.25	41.84	7.96	50.82	49.04

然出现差的检测效果. 而采用PCA降维的训练结果由于丢失了许多主成分信息,检测准确率均在65%以下. 相比之下,基于DBN-Finetune的方法获得了较好的检测效果,这是因为在训练过程中使用有标签的目标域数据进行了微调. 然而,若MMD增大,参数微调则无法解决源域和目标域数据分布差异变大的情况,检测效果显著下降. 本文所提出方法随着MMD的增大,仍均有97%以上的检测精度,误报率也只有少量上升,相比其他3种方法有更好的泛化性.

4 结 论

本文针对电力系统中的隐匿FDI攻击真实测量值与模拟测量值存在一定差异的问题,提出了基于深度信念网络和迁移学习的检测机制. 首先利用DBN中的自学习网络对海量目标域无标签样本进行特征自学习,并用海量源域有标签样本对DBN模型进行再训练;然后将参数迁移到目标域网络并冻结,增加或替换隐藏层,利用少量目标域有标签样本完成对适配层的训练;最后取消冻结,再次利用少量目标域有标签样本进行微调得到合适的检测模型. 设置拓扑估计参数偏差在10%以内,并比较其在不同受保护测量节点和噪声水平下源域和目标域数据的分布差异. 利用NYISO的实际负荷数据验证所提出方法的检测性能,仿真结果表明,在不同分布差异下,这种检测机制均具有较高的检测效果,相比其他机器学习方法具有更好的泛化性和鲁棒性.

参考文献(References)

[1] 王冰玉, 孙秋野, 马大中, 等. 能源互联网多时间尺度的信息物理融合模型[J]. 电力系统自动化, 2016, 40(17): 13-21.
(Wang B Y, Sun Q Y, Ma D Z, et al. A cyber physical model of the energy internet based on multiple time scale[J]. Automation of Electric Power Systems, 2016, 40(17): 13-21.)

[2] 孙秋野, 杨凌霄, 张化光. 人工智能技术在电力系统中的应用与展望[J]. 控制与决策, 2018, 33(5): 938-949.
(Sun Q Y, Yang L X, Zhang H G. Smart energy—Applications and prospects of artificial intelligence technology in power system[J]. Control and Decision, 2018, 33(5): 938-949.)

[3] 王轶楠, 林彦君, 李焕, 等. Dos攻击下电力网络控制系统脆弱性分析及防御[J]. 控制与决策, 2017, 32(3): 411-418.
(Wang Y N, Lin Y J, Li H, et al. Vulnerability analysis and countermeasures of electrical network control systems under DoS attacks[J]. Control and Decision, 2017, 32(3): 411-418.)

[4] 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): 1665-1676.

[5] Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids[J]. ACM Transactions on Information System Security, 2009, 14(1): 1-33.

[6] Yan J Q, Guo F H, Wen C Y. False data injection against state estimation in power systems with multiple attackers[J]. ISA Transactions, 2020, 101: 225-233.

[7] Yang Q Y, Yang J, Yu W, et al. On false data-injection attacks against power system state estimation: Modeling and countermeasures[J]. IEEE Transactions on Parallel Distributed Systems, 2014, 25(3): 717-729.

[8] Sedjelmaci H, Senouci S M, Ansari N. A hierarchical detection and response system to enhance security against lethal cyber-attacks in uav networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 48(9): 1594-1606.

[9] 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.

[10] Kosut O, Jia L Y, Thomas R J, et al. Malicious data attacks on the smart grid[J]. IEEE Transactions on Smart Grid, 2011, 2(4): 645-658.

[11] Moslemi R, Mesbahi A, Velni J M. Design of robust profitable false data injection attacks in multi-settlement electricity markets[J]. IET Generation Transmission & Distribution, 2018, 12(6): 1263-1270.

[12] Li Y C, Wang Y Y. False data injection attacks with incomplete network topology information in smart grid[J]. IEEE Access, 2019, 7: 3656-3664.

[13] Hu Z, Wang Y, Tian X, et al. False data injection attacks identification for smart grids[C]. The 3rd International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE). Piscataway: IEEE, 2015: 139-143.

[14] Esmalifalak M, Liu L C, Nguyen N, et al. Detecting stealthy false data injection using machine learning in smart grid[J]. IEEE Systems Journal, 2017, 11(3): 1644-1652.

[15] Ozay M, Esnaola I, Yarman Vural F T, et al. Machine learning methods for attack detection in the smart grid[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(8): 1773-1786.

[16] Yu J J Q, Hou Y H, Li V O K. Online false data injection attack detection with wavelet transform and deep neural networks[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3271-3280.

[17] He Y B, Mendis G J, Wei J. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism[J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2505-2516.

[18] Deng Y, Zhu K, Wang R, et al. Real-time detection of false data injection attacks based on load forecasting in smart grid[C]. 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids. Piscataway: IEEE, 2019: 1-6.

[19] Sun Y B, Fu M Y, Wang B C, et al. Dynamic state estimation in power systems using a distributed MAP method[C]. The 34th Chinese Control Conference. Hangzhou, 2015: 47-52.

[20] 邬晶, 李依宁, 李少远. 分布式电网CPS系统数据攻击下的状态估计[J]. 控制与决策, 2016, 31(2): 331-336.
(Wu J, Li Y N, Li S Y. State estimation for distributed cyber-physical power systems under data attacks[J]. Control and Decision, 2016, 31(2): 331-336.)

[21] Margossian H, Sayed M A, Fawaz W, et al. Partial grid false data injection attacks against state estimation[J]. International Journal of Electrical Power & Energy Systems, 2019, 110: 623-629.

[22] Liu X, Li Z Y. False data attacks against AC state estimation with incomplete network information[J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2239-2248.

[23] Wang H Z, Ruan J Q, Zhou B, et al. Dynamic data injection attack detection of cyber physical power systems With uncertainties[J]. IEEE Transactions on Industrial Informatics, 2019, 15(10): 5505-5518.

[24] Fischer A, Igel C. Training restricted Boltzmann machines: An introduction[J]. Pattern Recognition, 2014, 47(1): 25-39.

[25] Courville A, Desjardins G, Bergstra J, et al. The spike-and-slab RBM and extensions to discrete and sparse data distributions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(9): 1874-1887.

[26] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge Data Engineering, 2010, 22(10): 1345-1359.

[27] 阎高伟, 贺敏, 汤健, 等. 基于最大均值差异多源域迁移学习的湿式球磨机负荷参数软测量[J]. 控制与决策, 2018, 33(10): 1795-1800.
(Yan G W, He M, Tang J, et al. Soft sensor of wet ball mill load based on maximum mean discrepancy multi-source domain transfer learning[J]. Control and Decision, 2018, 33(10): 1795-1800.)

[28] 王毅星. 基于深度学习和迁移学习的电力数据挖掘技术研究[D]. 杭州: 浙江大学, 2019.
(Wang Y X. Power data mining technology based on deep learning and transfer learning[D]. Hangzhou: Zhejiang University, 2019.)

[29] Smola A, Gretton A, Song L, et al. A hilbert space embedding for distributions[C]. International Conference on Algorithmic Learning Theory. Berlin: Springer, 2007: 13-31.

作者简介

郭方洪 (1988–), 男, 教授, 博士, 从事智能电网可靠性与安全、微电网分布式控制与优化等研究, E-mail: fhguo@zjut.edu.cn;

易新伟 (1997–), 男, 硕士生, 从事信号检测、智能电网安全的研究, E-mail: xwyi700@163.com;

徐博文 (1997–), 男, 硕士生, 从事智能电网安全、微电网分布式优化的研究, E-mail: bwxu@zjut.edu.cn;

董辉 (1979–), 男, 教授, 博士生导师, 从事嵌入式检测技术、智能信息处理等研究, E-mail: hdong@zjut.edu.cn;

张文安 (1982–), 男, 教授, 博士生导师, 从事网络化控制、信息融合、工控系统安全等研究, E-mail: wazhang@zjut.edu.cn.

(责任编辑: 郑晓蕾)