计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(9):136−144 [doi: 10.15888/j.cnki.csa.008657]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于无监督学习的智能电网入侵检测①

李　洋[1], 余亚聪[2], 张立武[1], 邱兰馨[3], 曹　委[1], 秦中元[2]

[1](南瑞集团有限公司 (国网电力科学研究院有限公司), 南京 211106)
[2](东南大学 网络空间安全学院, 南京 211189)
[3](国网浙江省电力有限公司信息通信分公司, 杭州 310016)
通信作者: 秦中元, E-mail: zyqin@seu.edu.cn

摘　要: 智能电网通过引入信息和通信技术服务, 带来了传统电网的技术演变, 与此同时在安全方面也带来了严重的挑战. 本文提出了一种智能电网入侵检测系统安全架构和一种基于无监督学习的新型入侵检测系统 (intrusion detection system, IDS). 我们设计了区域式训练 (block-training) 架构, 不仅可以减轻数据中心的计算压力, 还可以对本地流量进行特征训练. 我们还提出了一种基于交叉验证的递归特征消除的差分自编码器算法 (RFECV-VAE). RFECV-VAE 综合了 RFECV 和 VAE 模型, 在特征选择过程使用递归特征消除交叉验证法 (recursive feature elimination cross-validation, RFECV), 异常检测采用差分自编码器 (variational autoencoders, VAE), 它可以对大规模高维数据进行高精度异常检测. 最后, 本文选择深度自编码器、深度自编码器高斯混合模型、单类支持向量机、隔离森林、差分自编码器作为对比算法, 采用准确率、ROC_AUC、*F*1_*score* 和训练时间等指标来进行性能评估. 实验结果表明, RFECV-VAE 算法结果优于其他比较算法.

关键词: 智能电网; 入侵检测; 差分自编码器; 无监督学习; 机器学习

引用格式: 李洋,余亚聪,张立武,邱兰馨,曹委,秦中元.基于无监督学习的智能电网入侵检测.计算机系统应用,2022,31(9):136–144. http://www.c-s-a.org.cn/1003-3254/8657.html

Intrusion Detection in Smart Grid Based on Unsupervised Learning

LI Yang[1], YU Ya-Cong[2], ZHANG Li-Wu[1], QIU Lan-Xin[3], CAO Wei[1], QIN Zhong-Yuan[2]

[1](NARI Group Corporation (State Grid Electric Power Research Institute Co. Ltd.), Nanjing 211106, China)
[2](School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China)
[3](Information and Communication Branch of State Grid Zhejiang Electric Power Co. Ltd., Hangzhou 310016, China)

Abstract: The smart grid (SG) constitutes a technological evolution of the traditional grid by introducing information and communication technology (ICT) services. Although using of ICT has advantages, it poses some serious challenges to security. In this study, we propose a security architecture of the smart grid intrusion detection system and a novel intrusion detection system (IDS) based on unsupervised learning. We design block-training architecture which can not only reduce the computing burden in the data center but also train the characteristics of local traffic. We also propose a variational autoencoder based on recursive feature elimination with cross-validation (RFECV-VAE). The RFECV-VAE is a combination of RFECV (for feature selection) and VAE model (for anomaly detection) and can detect large-scale and high-dimensional data with high accuracy. Finally, we choose deep autoencoder (DAE), deep autoencoding Gaussian mixture model (DAGMM), one-class support vector machine (OCSVM), isolation forest (IF), and VAE as comparison algorithms and accuracy, ROC_AUC, *F*1_*score*, and training duration for performance evaluation. The experimental results show that RFECV-VAE outperforms the comparison algorithms.

Key words: smart grid; intrusion detection; variational autoencoder (VAE); unsupervised learning; machine learning

① 基金项目: 国家电网有限公司总部管理科技项目 (SGZJXT00JSJS2000455)
收稿时间: 2021-11-25; 修改时间: 2021-12-22; 采用时间: 2022-01-05; csa 在线出版时间: 2022-05-31

众所周知, 电力是主要能源之一, 在工业和生活中发挥着不可替代的作用. 随着信息和通信技术 (information and communication technology, ICT) 的最新进展, 智能电网 (smart grid, SG) 提供了一个经济、高效、可持续的电力系统, 已经被广泛引入. 智能电网生态系统通常由多个智能设备组成, 包括智能计量、收集和监测系统, 它们能够产生大量通过互联网进行传输的数据. 然而, 在许多物理网络系统中 (如智能电网), 标准通信协议缺乏基本的安全措施, 如加密和认证, 这使得工业网络特别容易受到攻击[1]. 2020 年 4 月, 葡萄牙的一家跨国天然气和电力能源公司受到 Ragnar Locker 勒索软件的攻击, 被索求巨额赎金. 同年 6 月, 巴西电力公司 Light S.A.同样被黑客勒索了巨额赎金.

在当今的智能电网系统中, 特别是在高级计量基础设施 (advanced metering infrastructure, AMI) 中, 通常会收集和传输用户的地理位置、身份表示和电力消耗. 一旦攻击者破坏了智能电力终端, 如智能电表, 用户的私人信息将会面临被泄露的风险. 攻击者可以根据用户分时段的用电情况推断出用户的用电模式, 从而推断出用户的出行习惯, 并在用户不在家时实施入室盗窃. 所以先进的入侵检测安全架构是智能电网的一个重要组成部分. 它不仅可以通过智能终端以规定的方式收集和分析用户的电力数据, 还能提供双向的通信.

目前, 智能电力系统主要由一系列承担不同角色的智能嵌入式电力终端组成, 如配电终端单元、变压器终端单元、馈线终端单元等. 这些智能嵌入式终端通过与智能电网的互动传输电力信息, 使整个网络更加智能化. 然而目前智能电力系统主要存在以下两个问题: (1) 电力信息复杂且数据量庞大, 常规的入侵检测系统难以应对如此繁重的计算压力. (2) 在引入大量异构电力智能终端设备的同时, 这些设备本身也存在大量漏洞, 容易被攻击者利用, 成为进一步攻击电网主站的跳板. 一旦电力智能终端被入侵, 整个电网将面临被破坏的可能.

为了解决智能终端被入侵的问题, 近年来国内外学者提出了许多研究方案. Hinton 等人[2] 提出了使用自编码器来识别异常数据, Zong 等人[3] 提出深度自编码高斯混合模型用于无监督的异常检测, 解决了单一高斯分布不能适应复杂分布的问题. 但他们只是根据重建误差来检测异常, 缺乏客观性和可变性.

本文对于智能电网的入侵检测进行了深入的研究, 提出了区域式训练 (block-training, BT) 架构, 使智能终端的 IDS 适应本地流量特征, 并通过合理分配计算资源, 进一步分散计算中心的计算压力. 此外, 还提出了基于交叉验证的递归特征消除的差分自编码器算法 (RFECV-VAE), 在特征选择过程使用递归特征消除交叉验证法, 异常检测采用差分自编码器, 经过实验验证, 该算法更适合于高维数据和大规模数据集, 具有较高的准确性和较少的检测时间.

本文的其余部分组织如下: 第 2 节概述了当前领域内的相关研究工作, 第 3 节介绍了本文的研究方法, 包括 BT 架构和 RFECV-VAE 算法, 第 4 节对实验结果进行分析并且对模型属性进行相关评估, 第 5 节给出了本文的结论和对日后发展方向的思考.

1 智能电网入侵检测相关工作

1.1 智能电网领域中的网络安全威胁

由于网络威胁层出不穷, 关键基础设施的网络安全, 特别是智能电网安全越来越引起人们的重视.智能电网安全问题可能来自许多方面, 如黑客攻击、网络犯罪和网络战争.

Hahn[4] 提出, 为了使攻击对系统产生负面影响, 攻击者不仅需要知道如何破坏电网的网络元素, 还需要知道如何控制网络元素来操纵物理系统. 为了解决智能电网通信中存在的认证问题, Aghapour 等人[5] 提出通过使用基于知识、占有和生物识别的 3 个认证因素来加强对用户的身份认证, 以达到防御伪装攻击的目的.

1.2 入侵检测算法研究

近年来, 在异常检测领域提出了许多新方法.在本节中, 我们回顾并总结了近年来提出的入侵检测算法.

支持向量机 (support vector machines, SVM)[6] 是一种可以将 n 维空间数据进行分类的方法. Winter 等人[7] 提出了一种感应式网络 IDS, 它使用 OCSVM (one-class SVM) 作用于网络流量的识别并进行分析. Wagner 等人[8] 提出了一种基于 SVM 的处理大量网络流记录的检测方法.

聚类方法在检测数据集的独特性方面表现良 Casas 等人[9] 提出了一种基于异常检测的 IDS, 它通过使用各种无监督聚类方法将网络中的数据包随机地收集成流, 以达到检测网络流异常的目的. Hosseinpour

等人[10]提出了一种基于无监督聚类和人工免疫系统的分布式 IDS, 取得了不错的检测结果.

决策树 (decision tree, DT) 根据树的每个节点的值建立规则, 并生成一个树模型. Thaseen 等人[11]讨论了基于决策树的各种算法在入侵检测分类中的效果和影响. Stevanovic 等人[12]提出了一种有效的方法来检测僵尸网络. 结果显示, 在有监督的机器学习方法中, 随机森林 (random forest, RF) 方法表现最好. Zhou[13]提出, 异常现象有两个明显的特征: 少和不同. 这两个特点使得异常现象更容易受到一种叫做隔离的机制的影响. 所以他设计了一个可以有效构建隔离实例的二叉树结构, 称为隔离树 (iTree). 由于对隔离的敏感性, 离群点和正常点都分布在 iTree 的两端, 较浅的点更可能是离群点, 较深的点更可能是正常点. Jiang 等人[14]提出, 他们提出的 PSO-XGBoost 模型显示出比其他替代模型更高的分类精度.

人工神经网络 (artificial neural network, ANN) 的目标是模拟人体的神经网络. Song 等人[15]提出了一个使用反向传播神经网络分类器和统计特征向量的异常检测系统. Siniosoglou 等人[16]根据自编码器和对抗生成网络, 提出了一种适用于智能电网的入侵检测系统, 并且通过实验证明了该系统的有效性. Abuadlla 等人[17]提出了一种 IDS, 以检测基于流量的数据中的一些特定的入侵行为. Vinayakumar 等人[18]创建了一个高效的 IDS, 基于深度神经网络, 可以通过监督学习方法识别突发的入侵行为. Mendonça 等人[19]提出了一种基于树-卷积神经网络分层算法和软根-符号激活函数的算法.该模型减少了生成模型的训练时间, 并被用于检测 DDoS、网络攻击. Andresini 等人[20]提出了一种新颖的深度学习方法, 该方法使用卷积神经网络, 为计算机网络提供了一种有效的方法来分析网络流量, 以区分恶意活动. Rajadurai 等人[21]提出了结合多种机器学习算法的堆叠式集合学习.

1.3 差分自编码器

自编码器 (autoencoder, AE) 是一种神经网络方法, 基本上由一个编码器和一个解码器构成, 它能够以无监督的方式将输入向量重建为输出向量[22]. Hinton 等人[2]提出, 使用自编码器可以降低高维数据的维度. Zong 等人[3]提出了深度自动编码高斯混合模型, 用于无监督的异常检测. Vincent 等人[23]提出了一种具有更强的特征学习能力的基于去噪自编码器的结构. Kingma 等人[24]介绍了一种基于随机变异推理和学习的算法, 在具有连续潜变量和后验分布的大数据集的情况下, 能有效地推断和学习有向概率模型. An 等人[25]提出了利用差分自编码器中重建概率来检测异常. Li 等人[26]提出了基于随机森林算法的自编码器入侵检测系统, 该算法可以预测自编码器的结果, 在检测时间和检测精度上表现更好.

差分自编码器是一种有向概率图形模型, 其后验概率由神经网络逼近, 形成一种类似于自动编码器的结构. VAE 的结构如图 1 所示[25].

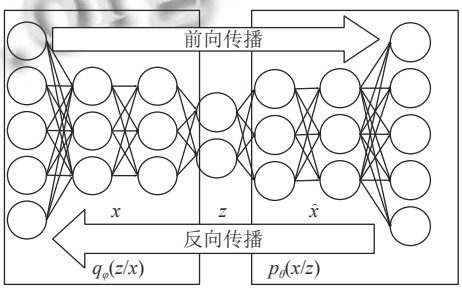

图 1 差分自编码器架构

图 1 的左半部分是编码过程, 右半部分是解码过程. x 是模型的原始输入, z 是模型的潜变量. $q_\varphi(z|x)$是近似后验. z 是由采样和输入数据 x 的参数生成的, 它不仅包含 x 的信息, 而且满足高斯分布, 便于后续梯度下降或其他优化技术的应用. $p_\theta(x|z)$代表在给定潜伏变量 z 时, 数据 x 的可能性. $\hat{x}$是基于潜伏变量 z 产生的新样本. 从训练好的 VAE 概率编码器中提取大量样本进行测试, 在每次计算中, 输入是编码器的每个样本, 输出是概率解码器根据算法输出的均值和方差参数.根据输出的均值和方差, 可以计算出从分布中产生原始数据的概率. 异常判断的标准是, 重建概率低于所划定的阈值的数据点将被归类为异常点. 重建概率和重建误差在许多方面是不同的[25]. 重建概率阈值的决定是比较客观、合法和容易获得的.

2 基于区域式训练架构模型的智能电网入侵检测模型

智能电网信息系统由 3 个重要组成部分组成: 智能终端、数据中心和中央系统.目前的智能电网系统存在两个问题: (1) 智能终端的计算资源分配不均, 导致有的区域没有计算资源对细微的异常进行识别, 而有

的区域计算资源处于空闲状态. (2) 不同类型和地区的终端受到的攻击具有局域性. 为了解决这些问题, 本文提出了基于区域式训练架构模型的智能电网入侵检测模型, 如图 2 所示.

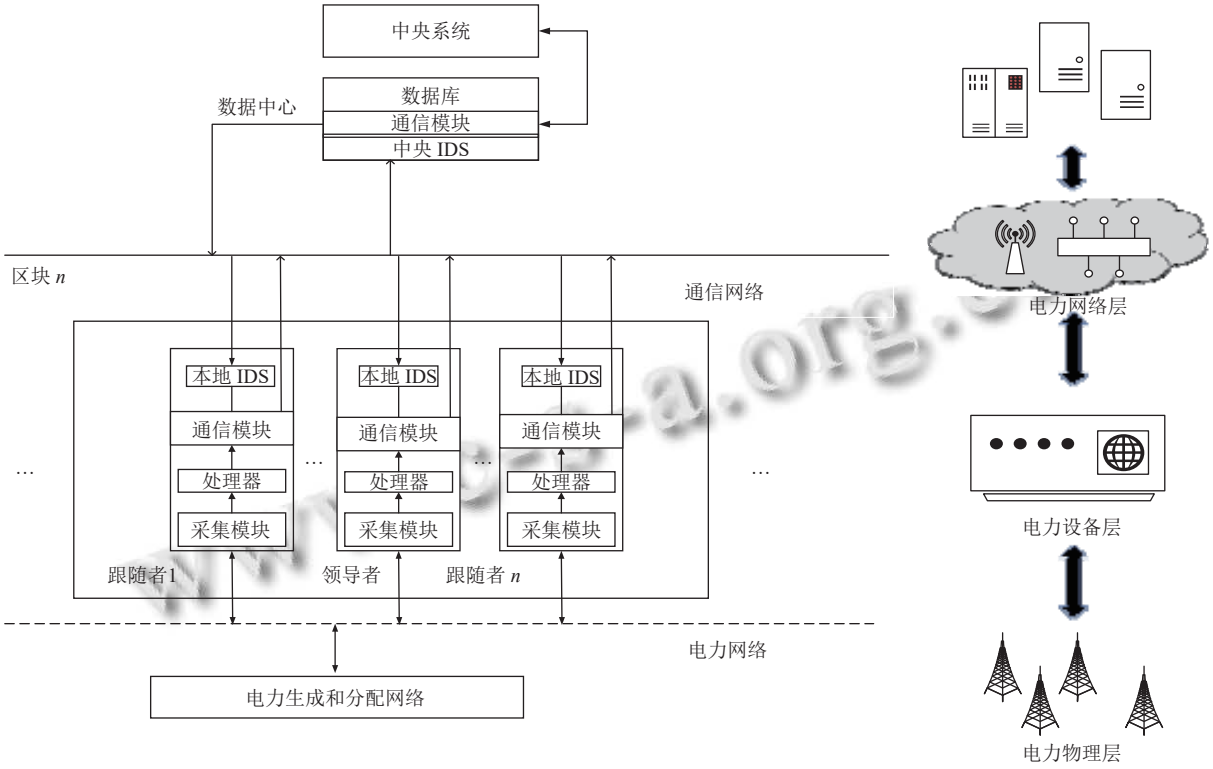

图 2 区域式训练系统架构模型

如图 2 所示, 我们在逻辑上将网络分为电力网络和通信网络. 电力网络主要负责传输能量流, 而通信网络则负责传输信息流. 该架构中主要分为 4 层, 分别是电力物理层、电力设备层、电力网络层和电力应用层. 下面对各层进行介绍.

2.1 电力物理层

电力物理层涉及与发电、输电和配电有关的物理层设施, 负责底层电力的生成、运输和分配等功能.

2.2 电力设备层

电力设备层包括各种电力终端, 如智能电表、配电终端单元、变压器终端单元等. 电力采集模块从电力网络中采集电力信息, 经处理器处理后, 电力信息通过通信模块与通信网络进行交互. 我们将逻辑上相邻的一些智能终端组合成一个区块, 在一个区块中, 拥有最高计算能力的终端将被选为区块的领导者 (leader), 其他成员为追随者 (follower). Leader 可以根据一段时间内的实际流量来对区块的入侵识别模型进行训练, 在训练完模型后, 它将参数推送给同一区块的 follower 终端, follower 终端根据 leader 发送的参数更新自己识别模型的参数.如果 leader 超过一定的时间没有发送更新的参数, 可以认为 leader 已经失去了作为领导者的能力, 该区块中将重新选举出一个新的 leader. 通过这种方式, 普通终端只参与异常的识别, 而不参与模型的训练, 这样可以分摊数据中心的计算压力, 并且可以减少绝大多数的终端重复计算造成的计算资源浪费.

2.3 电力网络层

电力网络层由无线网络、互联网、电力专用网等组成. 电力网络层负责电力设备和网络层之间的信息流传输. 该层是网络攻击的主要目标, 攻击者通过截取、窃听、篡改该层的信息报文进行攻击.

2.4 电力应用层

电力应用层主要包括电力数据中心和中央系统控制平台. 除了存储所有的电力信息外, 该层还负责每个区块的 leader 的选举和管理. 为了保证数据中心的安全, 数据中心的 IDS 必须具有最高的安全级别. 因此, 数据中心的 IDS 使用网络中出现的所有信息流进行训

练以保证数据中心的安全.

总而言之, 区块训练和 leader-follower 模式不仅可以分散数据中心的计算压力, 减少部分终端重复计算的资源浪费, 还可以使特定范围内的终端准确地识别自己网段的异常情况.

3 基于交叉验证的递归特征消除的差分自编码器算法

本文提出的 RFECV-VAE 算法分为训练和测试两部分. 图 3 是 RFECV-VAE 训练过程的流程图, 图 4 是测试过程流程图. 在图 3 中, 左边部分展示了特征选择过程, 右边部分展示了训练过程. 训练的目的是通过对正常数据的训练, 获得决定编码和解码结果的参数 θ 和 Φ. 图 4 中, 本算法将通过对计算出的测试数据的重建概率和设定的阈值进行比较以检测其是否为异常数据. 该算法的实现步骤如下.

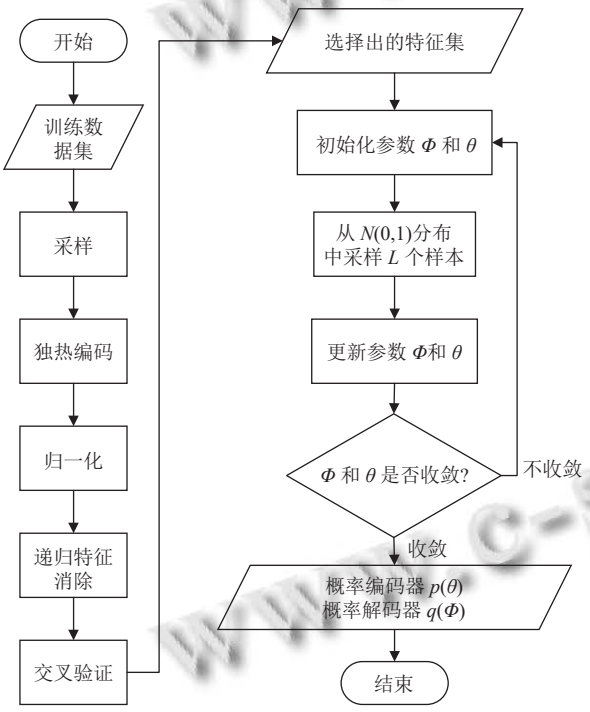

图 3 RFECV-VAE 算法训练流程图

3.1 预处理过程

预处理包括以下步骤: 采样、独热编码、归一化和特征选择.

(1) 采样

基准数据集将被分为两部分: 训练集和测试集. 由于基准数据集中存在 DDoS 攻击, 异常数据量远大于正常数据量, 所以有必要对数据集进行抽样.

本文提出的算法是基于无监督学习的. 因此, 训练模型只需要对正常数据集进行训练, 以学习正常数据的特征. 为了验证模型识别异常的能力, 测试集中的异常数据量应该与正常数据量相似.

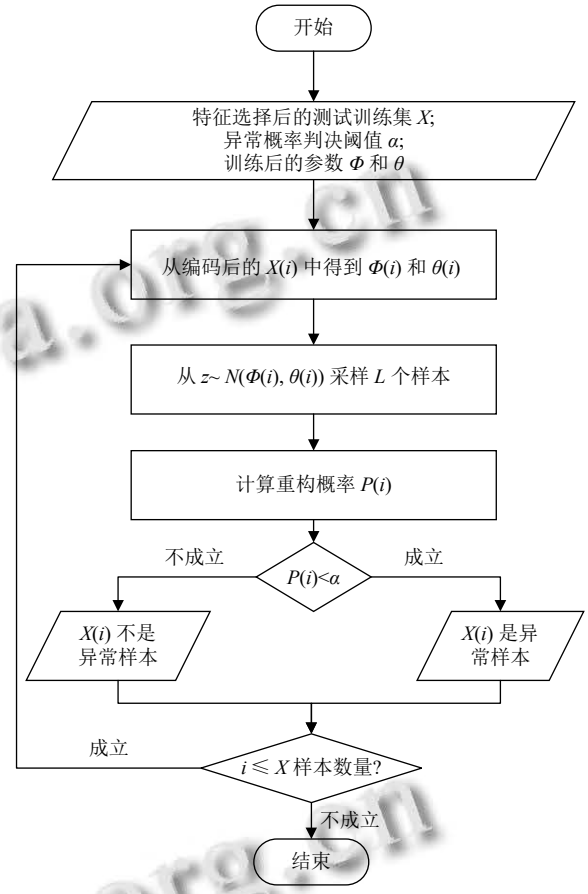

图 4 RFECV-VAE 算法训练流程图

(2) 独热编码

每条数据的特征类型分为数字特征和字符特征. 为了使模型能够学习字符特征, 我们对字符特征进行一次独热编码, 将其转为数字特征, 使得字符特征之间的距离计算更加合理, 但会造成特征数量的增加.

(3) 归一化

为了使数据更具可比性, 采用最大和最小归一化方法, 使所有特征指标在处理后处于同一数量级, 以减少极端特征取值对准确率的影响. 计算公式见式 (1), 其中, x_{max} 表示该特征在所有样本中的最大值, x_{min} 表示该特征在所有样本中的最小值, x_{norm} 表示特征归一化后取值, 取值范围在 0 到 1 之间.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

3.2 特征选择

经过独热编码, 特征维度明显增加. 为了降低模型学习的难度, 我们对数据进行特征筛选, 选择最能代表数据的特征. 这里我们使用递归特征消除交叉验证法. 递归特征消除 (recursive feature elimination, RFE) 的主要思想是反复建立模型并选择最好 (或最差) 的特征, 然后把选择的特征放在一边, 最后对剩下的特征重复这个过程, 直到所有的特征都被遍历. 这个过程中消除的顺序就是特征的排序. 交叉验证 (cross-validation, CV) 的目的是为了对 RFE 的结果进行验证, 具体的步骤如下: 首先, 根据在 RFE 阶段确定的特征重要性, 依次选择不同数量的特征. 之后对选定的特征集进行交叉验证, 最后确定平均得分最高的特征集.

3.3 VAE 算法

异常检测任务是以无监督的方式执行的, 这意味着只有正常的数据样本可以被用来训练 VAE. 训练过程中, 概率解码器g_θ和编码器f_φ分别对原始输入变量空间和潜在变量空间的多向正态分布进行参数化. 测试过程是通过从训练好的 VAE 模型, 对每个测试用例产生的平均值和方差参数来计算从分布中产生的原始数据的概率, 也称为重构概率 (reconstruction probability, RP), 重构概率和预先设定的阈值进行比较, 比较结果作为异常的判定准则. 其中, 重构概率计算方式是通过对$E_{q_\varphi(z|x)}[\log p_\theta(x|z)]$的 Monte Carlo 估计.

RP 通过使用原始输入变量分布参数的随机潜在变量来计算的. 这基本上等同于从近似后验分布中提取的一些潜在变量产生数据的概率. 当重建概率大于阈值时, 代表该数据为异常数据.

4 实验分析

4.1 实验环境

实验环境包括硬件设备和软件环境. 硬件设备方面, 我们采用了智能配电终端 PDZ 932. 该终端集成了供电信息采集、存储和传输、负荷控制、设备通信联网和状态监测、决策和本地分析、协同计算和主站通信等功能. 模型的训练是在 Core i7 处理器和 GTX1050Ti 显卡的计算机中进行的. 软件环境方面, 我们使用 Python 3 环境和 TensorFlow 库.

4.2 数据准备

我们在 2021 年 4 月和 2021 年 6 月期间从电力终端收集了大约 300 万条传输信息, 其中每一条都代表了电力终端和主站之间的一次通信. 由于电力终端的传输报文是基于 TCP/IP 协议, 我们对报文进行分析和挖掘, 可以实现对电力网络层的安全防护. 在实际收集到的流量包中, 主要有以下 4 类异常情况: 拒绝服务攻击 (DoS)、远程机器未授权认证 (R2L)、本地用户未授权访问 (U2R) 和端口检测 (Probing).

4.3 数据预处理

首先, 需要对从电力终端采集的数据进行重建, 以平衡正常样本和异常样本. 经过数据采样, 我们构建了训练集和测试集, 数据结构如表 1 所示.

表 1 训练数据集和测试数据集的分布

数据集类型	正常数据	异常数据
训练数据集	680 946	0
测试数据集	291 834	238 047

表 1 中, 正常数据按照 7:3 划分, 70% 的正常数据作为训练集数据, 另外 30% 作为测试机的正常数据. 攻击数据共包含 238 047 条各种攻击类型的数据. 由于该算法属于无监督学习, 训练集不包括异常数据, 模型只学习正常数据的特征. 测试集为了检测不同类型的异常, 设置正常数据和异常数据的比例在 1:1 左右.

然后, 在独热编码阶段, 采样数据的维度会增加, 因为字符类型的特征被替换成数字特征. 对于所收集的数据集, 字符类型的特征是协议类型 (protocol type)、标志 (flag) 和服务 (service). 在独热编码之后, 特征从 41 维变成 118 维.

最后, 对所有的数字特征进行归一化, 使其取值全部处于同一范围内.

4.4 特征选择

本实验特征选择阶段选择的模型是随机森林, 评价标准是准确率. 图 5 显示了模型准确率与所选数据特征数量的变化.

从图 5 中可以看出, 当选择的特征数量达到 16 个左右时, 准确率达到了峰值. 继续增加特征选择的数量并不会明显提高准确率, 反而会增加维度, 不利于模型的训练. 因此, 我们选择最适合本实验的特征, 对原始数据进行处理. 最后所选特征为 dst_host_count, dst_bytes, logged_in, count, srv_count, same_srv_rate, service_ecr_i, diff_srv_rate, dst_host_srv_count, dst_host_same_srv_rate, protocol_type_icmp, dst_host_diff_srv_rate, service_http, src_bytes, dst_host_rerror_rate, dst_host_same_src_port_rae.

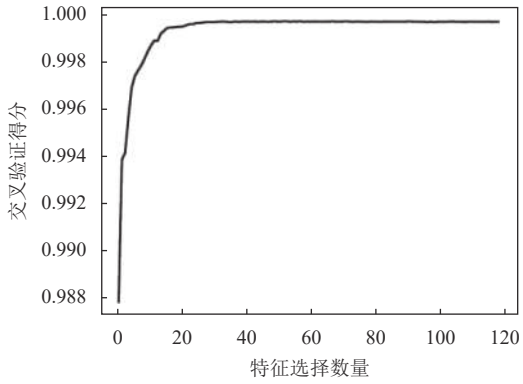

图 5 交叉验证得分和特征选择数量的关系曲线图

4.5 参数设置

为了避免单一抽样的随机性, 我们进行了 20 轮实验, 将每个实验的结果叠加, 取平均值作为最终结果. 模型的具体参数设置见表 2.

表 2 训练模型中参数的设置

参数	取值	参数含义
Input_dim	16	输入X的维度
Latent_dim	10	隐含层Z的维度
Learning_rate	0.000 5	模型的学习率
Batch_size	32	每次训练的数据数量
Train_iter	900	训练轮次
Hidden_units	128	隐含层单元数量
Judge_loss	0.93	异常概率判决阈值

4.6 实验评估

除了选择准确率和训练时间作为评价标准, 我们另外选择了 ROC_AUC 和 *F*1_*score* 作为额外的评价标准. ROC 是一条以真阳性率 (TPR) 为纵坐标, 以假阳性率 (FPR) 为横坐标的曲线. AUC 则是 ROC 曲线下面积, AUC 值越大, 则代表分类效果越好. *F*1_*score* 的作用是协调准确性 (*Precision*) 和召回率 (*Recall*). *F*1_*score* 的计算公式如式 (2):

$$F1\_score = 2 \times \frac{Precison \times Recall}{Precision + Recall} \tag{2}$$

4.7 实验结果

这里将实验结果分析分为两部分. 一是参数设置对模型的影响, 二是相同条件下不同算法的结果比较.

在实验 1 中, 我们研究了重构概率阈值选择和迭代次数对模型识别效果的影响. 我们以表 2 的参数设置为基准, 每次只改变其中一个参数变量. 重建概率从 0.90 开始到 0.99 结束, 而迭代次数从 1 开始到 2 000 次结束, 模型的准确率、*F*1_*score* 和 ROC_AUC 的变化见图 6 和图 7.

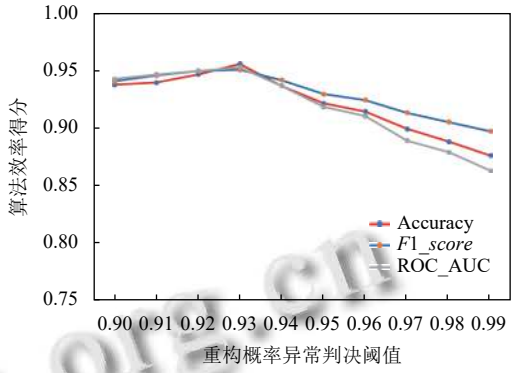

图 6 算法效率和概率判决阈值之间的关系

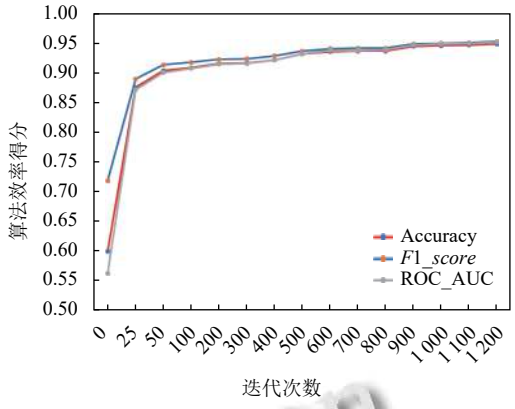

图 7 算法效率和迭代次数之间的关系

如图 6 所示, 当重建概率阈值增加时, 模型识别异常的能力先增加后减少, 出现峰值时, 判决阈值为 0.93. 这说明, 当重构概率判决阈值为 93% 时, 分类效果最好, 当阈值继续增加时, 并不会提升识别的效果, 反而会有所减弱, 原因是阈值提高时, 会有更多的正常数据被判别为异常. 在图 7 中, 随着迭代次数的增加, 模型识别异常的能力总体呈现上升趋势. 但是, 当迭代次数达到一定数量时, 模型参数已经收敛, 识别效果并没有明显增加, 因此我们选择迭代次数 900 轮作为最佳迭代次数.

在实验 2 中, 我们选择了几种经典算法作为比较算法, 有深度自编码器 (deep autoencoder, DAE)[24], 深度自编码器高斯混合模型 (DAGMM)[2], 单类支持向量机 (OC-SVM)[7], 隔离森林 (IF)[13], 差分自编码器 (VAE)[25]. 算法对比结果见表 3.

如表 3 中所示, RFECV-VAE 模型不仅在准确

率、*F*1_*score*、ROC_AUC 指标下表现出最好的性能, 同时, 训练时间也优于大多数算法. 综上可得出, RFECV-VAE 算法不仅在异常识别准确度方面表现优异, 而且还降低了收敛时间, 以使得整个系统变得轻量, 非常适合智能电力网络中高识别率和海量数据的要求.

表 3 不同算法的识别结果对比

算法	Accuracy	*F*1_*score*	ROC_AUC	训练时间 (s)
DAE[24]	0.870	0.755	0.766	3 703.00
DAGMM[2]	0.832	0.703	0.836	3 984.27
OC-SVM[7]	0.805	0.711	0.848	3 753.62
IF[13]	0.887	0.928	0.908	3 948.33
VAE[25]	0.933	0.938	0.944	4 032.52
RFECV-VAE	**0.946**	**0.950**	**0.949**	**3 440.21**

4.8 算法复杂度分析

RFECV-VAE 算法由两部分组成, 分别是特征选择 RFECV 部分和差分自编码器 VAE 部分. 本节分析算法的复杂度.

RFECV 包括递归特征消除和交叉验证. 在递归特征消除部分, 算法会在所有特征中遍历, 依次选择删除其中某一个特征, 并且重新构造模型, 计算每个特征的重要性, 时间复杂度为 O(n). 在交叉验证阶段, 算法根据 RFE 阶段得出的特征重要性排名, 依次选择不同数目的特征进行打分, 选出最适合的特征数目, 时间复杂度为 O(n). 故 RFECV 算法的时间复杂度为 O(n). 算法在整个过程中保存每个特征的重要性分数, 故空间复杂度为 O(n).

VAE 部分的训练阶段, 使用的样本集为正常样本数据, 每次可以训练多个固定数目的样本, 并反向更新参数, 每个样本只参与一次训练过程, 故 VAE 算法的时间复杂度为 O(n), 空间复杂度为 O(1).

5 总结

为了处理智能电网中的异常检测任务, 我们提出了一种智能电网 IDS 安全架构 (区域式训练模式) 和一种新的无监督算法 (RFECV-VAE). 从宏观的角度来看, 区域式训练不仅分担了数据中心的计算压力, 而且使智能终端更适应某一地区的流量特点. 从个体角度来看, 新的无监督算法显示出比其他算法更好的性能, 并且优化了训练时间, 使整个系统变得效率且轻量. 实验表明, 本文提出的安全架构和检测算法非常适用于智能电网的入侵检测场景.

参考文献

1 Liu Q, Hagenmeyer V, Keller HB. A review of rule learning-based intrusion detection systems and their prospects in smart grids. IEEE Access, 2021, 9: 57542–57564. [doi: 10.1109/ACCESS.2021.3071263]

2 Hinton G, Salakhutdinov R. Discovering binary codes for documents by learning deep generative models. Topics in Cognitive Science, 2011, 3(1): 74–91. [doi: 10.1111/j.1756-8765.2010.01109.x]

3 Zong B, Song Q, Min MR, *et al*. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. Proceedings of the 6th International Conference on Learning Representations. Vancouver, 2018. 97–106.

4 Hahn A. Cyber security of the smart grid: Attack exposure analysis, detection algorithms, and testbed evaluation [Ph.D. Thesis]. Ames: Iowa State University, 2013.

5 Aghapour S, Kaveh M, Martín D, *et al*. An ultra-lightweight and provably secure broadcast authentication protocol for smart grid communications. IEEE Access, 2020, 8: 125477–125487. [doi: 10.1109/ACCESS.2020.3007623]

6 Liao HJ, Lin CHR, Lin YC, *et al*. Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 2013, 36(1): 16–24. [doi: 10.1016/j.jnca.2012.09.004]

7 Winter P, Hermann E, Zeilinger M. Inductive intrusion detection in flow-based network data using one-class support vector machines. Proceedings of the 4th IFIP International Conference on New Technologies, Mobility and Security. Paris: IEEE, 2011. 1–5.

8 Wagner C, François J, State R, *et al*. Machine learning approach for IP-flow record anomaly detection. Proceedings of the 10th International Conference on Research in Networking. Valencia: Springer, 2011. 28–39.

9 Casas P, Mazel J, Owezarski P. UNADA: Unsupervised network anomaly detection using sub-space outliers ranking. Proceedings of the 10th International Conference on Research in Networking. Valencia: Springer, 2011. 40–51.

10 Farahnakian F, Amoli PV, Hosseinpour F, *et al*. Artificial immune system based intrusion detection: Innate immunity using an unsupervised learning approach. International Journal of Digital Content Technology and Its Applications, 2014, 8(5): 1–12.

11 Thaseen S, Kumar CA. An analysis of supervised tree based classifiers for intrusion detection system. Proceedings of 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering. Salem: IEEE, 2013.

294–299.
12 Stevanovic M, Pedersen JM. An efficient flow-based botnet detection using supervised machine learning. Proceedings of 2014 International Conference on Computing, Networking and Communications (ICNC). Honolulu: IEEE, 2014. 797–801.
13 Liu FT, Ting KM, Zhou ZH. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 3.
14 Jiang H, He Z, Ye G, *et al*. Network intrusion detection based on PSO-XGBoost model. IEEE Access, 2020, 8: 58392–58401. [doi: 10.1109/ACCESS.2020.2982418]
15 Song S, Ling L, Manikopoulo CN. Flow-based statistical aggregation schemes for network anomaly detection. Proceedings of 2006 IEEE International Conference on Networking, Sensing and Control. Ft. Lauderdale: IEEE, 2006. 786–791.
16 Siniosoglou I, Radoglou-Grammatikis P, Efstathopoulos G, *et al*. A unified deep learning anomaly detection and classification approach for smart grid environments. IEEE Transactions on Network and Service Management, 2021, 18(2): 1137–1151. [doi: 10.1109/TNSM.2021.3078381]
17 Abuadlla Y, Kvascev G, Gajin S, *et al*. Flow-based anomaly intrusion detection system using two neural network stages. Computer Science and Information Systems, 2014, 11(2): 601–622. [doi: 10.2298/CSIS130415035A]
18 Vinayakumar R, Alazab M, Soman KP, *et al*. Deep learning approach for intelligent intrusion detection system. IEEE Access, 2019, 7: 41525–41550. [doi: 10.1109/ACCESS.2019.2895334]
19 Mendonça RV, Teodoro AAM, Rosa RL, *et al*. Intrusion detection system based on fast hierarchical deep convolutional neural network. IEEE Access, 2021, 9: 61024–61034. [doi: 10.1109/ACCESS.2021.3074664]
20 Andresini G, Appice A, Malerba D. Nearest cluster-based intrusion detection through convolutional neural networks. Knowledge-based Systems, 2021, 216: 106798. [doi: 10.1016/j.knosys.2021.106798]
21 Rajadurai H, Gandhi UD. A stacked ensemble learning model for intrusion detection in wireless network. Neural Computing and Applications, 2020: 1–9.
22 Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. Gold Coast: ACM, 2014. 4–11.
23 Vincent P, Larochelle H, Lajoie I, *et al*. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 2010, 11: 3371–3408.
24 Kingma DP, Welling M. Auto-encoding variational Bayes. Proceedings of the 2nd International Conference on Learning Representations. Banff, 2014.
25 An J, Cho S. Variational autoencoder based anomaly detection using reconstruction probability. Proceedings of 2015 Special Lecture on IE. 2015. 1–18.
26 Li XK, Chen W, Zhang QR, *et al*. Building auto-encoder intrusion detection system based on random forest feature selection. Computers & Security, 2020, 95: 101851.

(校对责编: 牛欣悦)