基于半监督学习的物联网流量异常检测研究

王忠贤 易伟
郑州科技学院，中国·河南 郑州 450064

摘要：物联网设备在局域网环境中的广泛部署使其面临严峻的安全挑战。异常流量检测是保障其安全的关键，然而在实际场景中，异常样本极其稀缺，导致传统监督学习方法因样本不均衡而性能不佳。为解决标记异常样本稀缺的问题，本文提出了一种基于半监督学习的深度自编码器异常检测模型。该模型通过编码器 - 解码器结构学习正常流量的潜在特征表示，并利用少量标记样本微调一个集成在编码器后的分类器，形成结合重构误差与分类损失的综合目标函数。另外还设计了一套针对网络流量数据的增强策略以缓解数据不平衡。

关键词：物联网安全；异常检测；半监督学习；深度自编码器；数据增强

Anomaly Detection Method For Iot Traffic Based on Semi–supervised Deep Autoencoder

Wang Zhongxian, Yi Wei
Zhengzhou University of Science and Technology, China Henan Zhengzhou 450064

Abstract: The extensive deployment of IoT devices in the local area network environment makes them face severe security challenges. Anomaly detection is crucial to ensure their security, in practical scenarios, anomaly samples are extremely rare, resulting in poor performance of traditional supervised learning methods due to the unbalanced samples. To address the scarcity of labeled anomaly samples this paper proposes a semi-supervised learning-based deep autoencoder anomaly detection model. The model learns the latent feature representation of normal traffic through an encoder-decoder structure and fine-unes a classifier integrated after the encoder using a small amount of labeled samples to form a comprehensive objective function that combines the reconstruction error with the classification loss. Additionally, a set of augmentation is designed to alleviate data imbalance for network traffic data.

Keywords: Internet of things security; Anomaly detection; Semi-supervised learning; Deep autoencoder; Data augmentation

0 引言

随着物联网技术在智能家居、工业控制等领域的深度融合，局域网环境下的物联网设备数量激增，其安全问题日益凸显。以网络摄像头和智能门锁为代表的设备，其正常流量模式通常稳定且可预测 (如持续的视频流或间歇的控制指令)。任何与既定模式的显著偏离，如摄像头突发大量 TCP 连接请求或智能门锁与未知 IP 频繁通信，都可能是设备遭受攻击或发生故障的信号。然而，物联网环境中的异常流量样本在实际数据集中占比极低 (通常低于 5%)，这种严重的样本不平衡使得依赖大量标记数据的传统监督学习模型难以有效学习异常特征，导致检测准确率和召回率低下。无监督学习方法 (如孤立森林) 虽不依赖标签，但对于高级持续威胁等产生的、与正常流量模式高度相似的复杂异常，其检测能力有限，泛化性不足。

为了在标记样本稀缺的条件下实现高效的异常检测，半监督学习提供了一个有前景的方向。它能够同时利用大量易获取的正常流量数据 (未标记) 和少量宝贵的异常样本 (已标记)。本文提出一种基于深度自编码器的半监督异常检测框架。该模型通过编码器 - 解码器结构学习正常流量的潜在特征表示，并利用少量标记样本微调一个集成在编码器后的分类器，形成结合重构误差与分类损失的综合目标函数。

1 异常检测场景与数据困境

识别出设备后，需要持续监控其流量以检测异常。然而，在实际的物联网应用中，异常样本稀缺是一个普遍存在的问题。传统监督学习因缺乏正样本（异常）而表现不佳。无监督方法虽无需标签，但对复杂、隐蔽的异常模式泛化能力不足，难以检测高级持续威胁（APT）攻击。

2 深度自编码器与半监督训练策略

2.1 模型架构设计

为了对物联网设备异常检测中出现的标记异常样本稀缺的问题，提出一种基于半监督学习与深度自编码器的物

联网异常检测方法，其中深度自编码器的模型架构设计是关键。

编码器：编码器部分采用 3 层全连接层，其结构为 256 → 128 → 64。在物联网流量异常检测中，输入的流量特征包含丰富的信息，数据包长度序列、端口熵等。这些特征经过第一层全连接层，由 256 个神经元对其进行初步的特征变换和组合。这一层提取出一些局部的特征模式。这些局部特征模式在经过第二层全连接层时，由 128 个神经元进一步对其进行整合和提炼。最后，经过第三层全连接层，将特征压缩至 64 维的低维隐空间，将原始的复杂流量特征有效地压缩成低维的潜在表示。

解码器：解码器与编码器呈对称结构，其作用是将低维隐空间的特征表示还原成原始特征。从 64 维的潜在表示开始，经过第一层 128 个神经元的全连接层，将特征进行初步的扩展和重构，尝试恢复一些丢失的细节信息。再经过第二层 256 个神经元的全连接层，进一步对特征进行丰富和细化，使其更接近原始的流量特征。最终输出与输入特征维度相同的重构特征。

分类器：在编码层后接入 Softmax 分类器，其目的是利用少量的标记异常样本对模型进行有监督的训练。Softmax 分类器将编码层输出的低维特征映射到不同的类别概率空间，通过优化来调整分类器的权重，其中 y 表示样本的真实类别，z 表示编码层输出的特征，表示在特征 z 的条件下样本属于类别 y 的概率。通过这种方式，将无监督学习和有监督学习相结合，形成半监督目标函数，其中是一个超参数，用于平衡重构误差和分类误差的权重。

2.2 训练流程优化

2.2.1 无监督预训练

在训练的初始阶段，使用 10 万条正常流量数据进行无监督预训练。最小化重构误差（MSE），使其学会正常流量的特征表示。

2.2.2 有监督微调

加入 4% 的标记异常样本（约 4000 条）进行有监督微调，同时优化重构损失和分类损失，让模型学会区分正常和异常。

2.2.3 异常判别

在线检测时，计算新流量的重构误差。若误差超过阈值，则判定为异常。

2.3 数据集与对比实验

为应对样本量不平衡问题，本文提出一种流量数据扩展策略，具体可采用以下单一方法或多种方法组合的形式实现：

2.3.1 时间序列反转

针对流量数据时序依赖性强、模型易过度拟合单一时间流向的问题，采用时间序列逆向处理策略：将 24 小时历史流量序列沿时间轴翻转，使原始序列的起始片段转化为终止片段、终止片段转化为起始片段。

2.3.2 特征变换

考虑到流量数据特征（如数据包大小、传输速率）常存在分布偏态问题，对核心特征实施针对性数学变换：包括对数转换（用于压缩数值跨度较大的特征）、平方根运算（缓解右偏分布特征的极端值影响）、指数变换（增强低数值特征的区分度）等。

2.3.3 时空变换

若流量数据携带有地理空间标识（如设备接入位置的经纬度坐标），则针对空间维度实施时空变换：对地理坐标施加微小随机扰动（如基于高斯分布的 ±0.001 经纬度偏移），模拟不同地理位置（如同一区域内不同家庭、同一楼栋不同楼层）的设备接入场景。

2.3.4 采样变换

针对时间序列类流量数据的“多分辨率分析需求”，采用多时间尺度重采样策略：一方面通过下采样将分钟级流量数据聚合为小时级（如按 60 分钟滑动窗口求和），降低时间分辨率以凸显宏观流量趋势；另一方面通过上采样（如线性插值）将小时级数据拆解为分钟级，提升时间粒度以捕捉微观流量波动。

2.3.5 噪声注入

为模拟现实场景中数据不确定性（如物联网设备传感器的采集误差、无线传输链路的电磁干扰），向流量数据中注入符合实际扰动特性的随机噪声。

2.3.6 数据切片

采用时间片段化重组方法：将完整的流量时序数据（如 24 小时序列）切割为若干等长独立子片段（如每 2 小时 1 个片段），再通过随机排列子片段的顺序构建新的样本序列（如将“片段 1– 片段 2– 片段 3”重组为“片段 3– 片段 1– 片段 2”）。

2.3.7 缺失模拟

针对实际应用中“数据传输中断、设备离线导致的流量数据缺失”问题，采用缺失场景模拟策略：基于掩码机制随机掩盖部分时间步（如随机选取 10% 的时间点）或特征维度（如随机掩盖“数据包重传率”特征）的数值，构建含缺失值的流量样本。表 1 为训练集和测试集中不同流

表 1 训练集和测试集中不同流量类型的样本数量

报文类型	协议类型	非恶意报文	恶意报文
ClientHello报文	密码套件	提供密码套件0x002f(TLS_RSA_WITH_AES_128_CBC_SHA)	提供3个过时的密码套件，包括0x0004(TLS_RSA_WITH_RC4_128_MD5)
	扩展块	大多数TLS流中支持0x000d，以及0x0005（状态请求）、0x3374（下一个协议协商）、0xff01	客户端支持的TL扩展块值相同
ClientKeyExchange 报文	密钥长度	256位椭圆曲线密码作为公钥	2048 位RSA 公钥
ServerHello 报文		--	过时的密码套件
Certificate报文		0.1%为自签名	70%是自签名

表 2 超参数的取值

超参数	值
批量大小	128
预训练迭代次数	100
迭代次数	50
预训练学习率	0.001
学习率	0.001
权重衰减系数 λ	1e-6
平衡因子n	1

表 3 不同标记异常样本比例取值对应的 AUC 值（百分比）

γ_l	0	0.01	0.02	0.04	0.08
设备 1	99.84	99.92	99.89	99.92	99.94
设备 2	95.40	98.37	99.04	99.39	99.60
设备 3	98.95	99.58	99.82	99.82	99.85
设备4	96.70	99.04	99.35	99.67	99.76
设备 5	98.38	99.39	99.73	99.75	99.75
设备 6	98.06	99.52	99.85	99.81	99.90
设备 7	98.80	99.26	99.90	99.90	99.85
设备 8	97.79	99.73	99.64	99.74	99.79
设备 9	98.34	99.40	99.40	99.39	99.50

表 4 不同污染率取值对应的 AUC 值（百分比）

γ_p	0	0.01	0.02	0.05	0.10
设备 1	99.97	99.76	99.89	99.54	99.77
设备 2	99.94	99.92	99.86	98.84	97.75
设备 3	99.98	99.97	99.98	99.99	99.96
设备4	99.98	81.17	99.40	99.56	93.05
设备 5	99.79	99.76	98.18	93.11	93.38
设备 6	99.82	97.70	98.69	99.15	93.06
设备 7	99.89	99.90	99.89	99.90	99.91
设备 8	99.74	99.95	99.88	97.58	96.96
设备 9	99.94	99.93	99.55	98.86	98.99

量类型的样本数量列表。

N-BaIoT 数据包含 9 类物联网设备的真实流量数据，其中 7 台设备同时受到 BASHLITE 与 Mirai 两种僵尸网络的感染，另外 2 台则仅被 BASHLITE 单一恶意程序入侵。该数据集提供正常网络流量与异常恶意流量，涵盖 100 毫秒、500 毫秒、1.5 秒、10 秒和 1 分钟五种时间窗口的采样数据。其 23 维特征分为四类：第一类包含 8 个特征，可捕捉单方向数据包的传输规模特性；第二类涵盖 4 个特征，聚焦于单位时间内数据包总量的统计需求，形成用于量化数据传输频次的计数特征；第三类涉及 3 个特征，侧重描述数据包传输过程中时延波动的特性，构建反映数据传输稳定性的抖动特征；第四类包含 8 个特征，可同时关联入

表 5 不同已知异常类个数对应的 AUC 值（百分比）

K_l	1	2	3	4	5	6	7	8	9	10
设备 1	81.32	97.61	98.36	99.22	99.26	99.91	99.92	99.92	99.79	99.98
设备 2	79.78	86.14	93.91	97.03	98.33	98.81	99.88	99.90	99.97	99.96
设备 3	92.27	97.24	97.75	99.99	99.99					
设备4	95.47	95.59	95.82	97.44	99.09	99.32	99.67	99.53	100.00	100.00
设备 5	74.49	96.88	97.36	97.92	98.83	99.79	99.89	99.85	99.93	99.99
设备 6	93.99	92.67	96.83	97.49	99.63	99.27	99.99	99.96	99.86	99.99
设备 7	97.39	97.36	99.49	99.99	99.99					
设备 8	72.17	93.49	93.38	96.15	97.84	99.64	99.91	99.82	99.97	100.00
设备 9	94.82	94.13	97.48	98.95	99.33	99.52	99.46	99.96	99.98	99.99

站与出站双向数据包的传输规模，实现对双方向数据交互体量的综合表征。上述所有超参数的具体取值细节已系统汇总于表 2，为后续实验复现、结果验证及方法改进提供清晰参考。

在半监督学习的研究范式中，存在三个关键调控参数。第一个参数为训练集中标记样本的占比，其核心作用是调控标记样本在训练数据中的分布权重：当该占比取值为 0 时，半监督学习范式将退化为无监督学习；取值为 1 时，则完全转化为监督学习，二者分别对应不同的样本依赖场景。第二个参数是无标记样本的污染率，具体指无标记训练样本中异常样本的占比，的值越趋近于 1，表明无标记样本中异常样本的掺杂程度越高，模型对“正常 - 异常”特征边界的学习易受干扰，进而导致异常检测任务的难度显著提升。第三个参数为标记样本所含的异常类数量。关于上述三个参数的具体实验结果分析如下（表 3-4）。

3 结语

本文针对局域网内物联网设备因异常流量样本稀缺，导致传统监督学习异常检测方法性能受限的问题，本文提出基于半监督学习的深度自编码器模型与配套流量数据增强策略，通过编码器 - 解码器学习正常流量特征，结合少量标记样本微调分类器、构建综合目标函数，有效突破样本不均衡瓶颈，为物联网设备局域网安全防护提供了切实可行的技术路径，也为后续相关领域的异常检测研究提供了有益参考。

参考文献：

[1] 张明，李强，王丽. 基于半监督学习的未知异常检测方法 [J]. 计算机研究与发展，2025,62(4):892-905.

[2] 刘畅，陈明 基于深度学习的半监督对抗鲁棒网络入侵检测系统 [J]. 计算机工程与应用，2025,61(8): 12-120.

[3] 赵鑫，李娜，张伟. 基于半监督联邦学习的物联网入侵检测方法 [J]. 电子与信息学报，2024,46(3):987-995.

[4] Morichetta A, Bek Bulatova V, Dustdar S. FL-SERENADE: Federated Learning for SEmi-supeRvisEd Network Anomaly DEtection[C]//Proceedings of the 16th International Conference on Pervasive and Embedded Computing and Communication Systems. Vienna: SCITEPRESS, 2023: 45-56.

[5] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM Computing Surveys, 2009, 41(3): 1-58.

基金项目：河南省科技攻关项目：基于深度学习的物联网设备识别异常检测关键技术研究 (项目编号：242102210094)。

作者简介：王忠贤（1985.12-），男，汉族，河南郸城，本科，实验师，研究方向：计算机科学与技术、大数据分析与处理等，单位：郑州科技学院。

易伟（1984.06-），男，汉族，河南信阳，本科，副教授，研究方向为计算机科学与技术、物联网工程等，单位：郑州科技学院。