

FATIDS: 面向类不平衡样本的 物联网入侵检测方法

王 鹏, 宋亚飞, 王晓丹, 路艳丽, 向 前

(空军工程大学 防空反导学院, 西安 710051)

摘 要: 针对传统物联网入侵检测方法大多依赖复杂的特征预处理技术且对全局特征的建模能力不强, 难以有效表示高维序列的抽象特征, 从而对类不平衡数据的泛化性能较差的问题, 提出了一种基于 FATIDS 的物联网入侵检测方法, 通过自注意力机制实现了端到端的特征选择和特征提取, 动态调整对序列特征的注意力权值, 提高了针对高维序列特征的全局建模能力。为解决物联网入侵检测面临的样本类不平衡问题, 利用 Focal Loss 动态缩放模型梯度, 自适应降低简单样本的权重, 并聚焦于分类困难的类别。最后, 在公开的物联网入侵检测数据集 ToN_IoT 和 DS2OS 上验证了本文方法的有效性, 实验结果表明: FATIDS 在 ToN_IoT 的准确率、精确率、召回率和 F_1 分数 4 项指标分别为 99.60%、97.51%、96.59% 和 97.02%, 在 DS2OS 四项指标分别为 99.47%、99.93%、95.77% 和 97.42%, 本文方法相较其他先进方法实现了更强的检测性能。此外, 还进一步验证了重要超参数对本文方法性能的影响。

关键词: 物联网入侵检测; Transformer; Focal Loss; 网络安全

中图分类号: TP183; TP391.4 **文献标志码:** A **文章编号:** 1671-5497(2025)12-3986-14

DOI: 10.13229/j.cnki.jdxbgxb.20240403

FATIDS: an IoT intrusion detection method for class-imbalanced samples

WANG Peng, SONG Ya-fei, WANG Xiao-dan, LU Yan-li, XIANG Qian
(Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China)

Abstract: Network security issues are becoming increasingly prominent, and IoT network security urgently needs further investigations. Traditional IoT intrusion detection methods have weak feature representation capability for sequence data, and most of the methods based on machine learning and deep learning rely on complex feature preprocessing techniques and have weak global modeling capability for high-dimensional sequence data. To address the above problems, we propose a FATIDS-based IoT intrusion detection method, which achieves end-to-end feature selection and feature extraction through the self-attention mechanism, dynamically adjusts the attention to sequence features, and improves the global modeling

收稿日期: 2024-04-16.

基金项目: 国家自然科学基金项目(61876189, 61703426, 61273275); 陕西省高校科协青年人才托举计划项目(20190108); 陕西省创新人才推进计划项目(2020KJXX-065).

作者简介: 王鹏(1999-), 男, 博士研究生. 研究方向: 智能信息处理, 模式识别. E-mail: peng_wang2022@163.com

通信作者: 王晓丹(1966-), 女, 教授, 博士. 研究方向: 智能信息处理, 模式识别. E-mail: afeu_wang@163.com

capability for high-dimensional sequence features. To solve the imbalance problem faced by IoT intrusion detection, the Focal Loss is utilized to dynamically scale the model gradient, adaptively reduce the weight of simple samples, and focus on classes that are difficult to classify. Finally, the performance of the proposed method is validated on the ToN_IoT and DS2OS standard datasets, and the experimental results show that the proposed method achieves superior detection performance compared to other remarkable methods, and the impact of important hyperparameters on the performance of the proposed method is also validated.

Key words: intrusion detection of internet of things; transformer; Focal lLoss; cyber security

0 引 言

随着物联网(Internet of things, IoT)技术的迅猛发展,其在社会生活的各个方面已经得到了广泛应用,如智能物流、车联网、工业物联网、智能医疗和环境监测等领域^[1-3]。物联网通过将传感器和设备与互联网集成,极大提高了生产生活效率,但大量物联网设备暴露在网络中,各方面的问题日益凸显,尤其是物联网设备会产生大量的信息流量,需要可靠的智能处理技术保障整个体系的安全。入侵检测系统(Intrusion detection system, IDS)^[4]是一种积极主动的安全防护技术,无须转发任何流量,IDS可以从IoT中收集网络信息,如系统日志、审计记录、网络行为等,并对其进行分析,判断网络中是否存在潜在的攻击行为,并及时做出响应,将损失降到最低。随着物联网技术的普及,这种设备集成的体系在惠及人民的同时也开启了新的网络犯罪缺口。基于人工智能的入侵检测系统因其能够有效学习大量数据并识别之前未知的恶意活动已经成为工业界和学术界研究的热点问题。

近期的网络威胁暴露了传统防御框架的弱点,传统网络防御系统中使用的技术通常是基于静态和启发式的攻击特征,无法获取抽象的时序特征表示,导致识别性能和泛化性能十分有限。随着神经网络在其他领域的迅猛发展,机器学习也逐渐应用于物联网的入侵检测任务中,Pajouh等^[5]提出了一种基于双层降维和双层分类模块的新型入侵检测模型,旨在检测用户到根(U2R)和远程到本地(R2L)攻击等恶意活动,该模型采用了主成分分析和线性判别分析的降维模块,将高维数据集转化为特征较少的低维数据集,然后利用朴素贝叶斯和K-NearestNeighbor的双层分类模块识别可疑行为。Saba等^[6]提出了一种两阶段

混合方法,首先利用遗传算法选择适当的特征,然后采用机器学习(ML)算法,包括支持向量机(SVM)、集成分类器和决策树对选择的特征进行分类。Albulayhi等^[7]提一种新颖的特征选择和提取方法,该方法首先使用两种基于熵的方法(即信息增益(IG)和增益比(GR))选择和提取不同比例的相关特征。然后,使用数学集合论(联合和交集)提取最佳特征,利用4种机器学习算法,通过实验验证了模型的有效性。Islam等^[8]对以浅层机器学习和深度学习为基础的物联网环境中的入侵检测系统进行了全面分析和讨论,机器学习方法仅能提取序列数据的浅层特征,在物联网入侵检测任务中的泛化性能仍有待提高,基于深度学习的物联网入侵检测系统性能更加突出。

为了提取物联网流量的深层序列特征,增强对序列特征的表示能力,深度学习被引入物联网入侵检测领域。Saba等^[9]针对基于异常的入侵检测系统提出了一种基于卷积神经网络(CNN)的入侵检测方法,利用CNN有效提取相邻序列之间的关联信息,在NID数据集和BoT-IoT数据集实现了较好的检测效果。Abd Elaziz等^[10]结合群智能优化算法和深度学习算法,提出了一种基于CNN-CapSA的物联网入侵检测方法,首先利用深度神经网络和CapSA优化算法提出一种高效的特征选择技术,从物联网入侵检测序列数据中获取最佳特征,然后利用分类器进行分类,在4个数据集上验证了模型的优越性和竞争力。但是卷积神经网络受卷积核感受野的限制,仅能提取相邻序列的局部信息,对高维序列特征的提取能力存在局限性。为进一步提高序列数据的建模能力,Almiani等^[11]提出了一种全自动入侵检测系统,使用多层递归神经网络提取序列数据的连续相关信息,在NSL-KDD数据集上使用各种典型指标评估了模型性能。Jeyanthi等^[12]提出了一种

基于物联网的医疗入侵检测系统,该系统采用循环神经网络(RNN)和双向长短期记忆(BiLSTM)算法提取序列数据的复杂特征模式,实现了物联网入侵检测和分类,并在IoTID20数据集上验证了方法的有效性。为进一步降低循环神经网络在处理序列数据时存在梯度消失和爆炸的风险,通过CNN与RNN结合提取更加稳健的融合特征,李晓佳等^[13]提出一种面向CNN和RNN的物联网入侵检测模型,通过引入多头注意力机制,解决了CNN的原始池化层存在信息丢失的问题和RNN在处理长序列数据时存在梯度消失的问题。Altunay等^[14]通过CNN、长短期记忆(LSTM)和CNN+LSTM混合生成的深度学习架构,提出了3种不同的模型检测物联网中的入侵活动,在使用UNSW-NB15和X-IIoTID数据集进行的研究中,提出的循环神经网络和卷积神经网络能够提取局部和全局序列特征,实现了更好的检测性能。然而循环神经网络存在记忆消失问题,在串行处理长序列数据时会丢失重要特征,提取高维序列的全局特征时仍存在局限性。以上方法主要通过两阶段实现特征选择和特征提取,模型性能受限于复杂的特征预处理技术,随着Transformer的发展,其利用自注意力机制动态计算序列数据的全局相关性,自适应调整注意力权重,能够有效提取序列数据的深层全局特征,故本文进一步研究了Transformer在物联网入侵检测任务中的应用。同时,为了解决物联网入侵检测数据面临的类不平衡的问题,Dina等^[15]提出一种基于Focal Loss的深度学习的物联网入侵检测方法,缓解了数据集的不平衡问题,取得了较好的检测效果。

综上,物联网流量具有维度高、时序性和类别不平衡的特点,现有物联网入侵检测方法主要存在以下不足:

(1)模型训练前大多需要复杂的特征选择和特征降维,需要进一步加强端到端的物联网入侵检测研究。

(2)针对长程序列全局特征的建模能力较弱,无法有效提取物联网流量的高可分辨特征。

(3)模型自身对类别不平衡样本不敏感,处理类不平衡问题时泛化性能较弱。

针对以上问题,本文提出一种名为FATIDS的物联网入侵检测模型,通过自注意力机制自适

应地调整对物联网序列特征的注意力,自适应实现端到端的特征选择和特征提取,并采用Focal Loss动态缩放模型梯度^[16],自适应降低多数类样本的权重,使得模型聚焦于分类困难的少数样本类别,从而解决物联网攻击样本的类不平衡问题,提高模型的泛化性能。本文以公开物联网入侵检测数据集ToN_IoT^[17]和DS2OS^[18]作为基准数据集,验证了所提方法的有效性,并进一步探索了重要超参数对FATIDS检测性能的影响。

本文的主要贡献为以下3个方面。

(1)为实现自适应特征选择和提取全局深层特征,构建端到端的物联网入侵检测方法,本文提出了基于FATIDS的物联网入侵检测方法,利用自注意力机制动态调整对物联网网络流量特征的权重。相较其他先进的方法,本文方法具有了更优异的识别性能。

(2)为解决物联网攻击样本的不平衡问题,本文利用Focal Loss对交叉熵损失函数进行改进,自适应降低对多数样本类别的损失权重,聚焦分类困难的少数样本类别,提高了本文方法针对不平衡样本的泛化性能。

(3)利用公开真实环境下的ToN_IoT和DS2OS物联网入侵检测数据集进行评估,验证了本文方法的有效性,并探索了重要超参数对性能的影响。

1 模型结构

本节首先介绍了本文方法的整体框架,然后对重要模块的原理和细节进行介绍和分析。FATIDS的整体结构如图1所示。

在图1中,FATIDS网络架构主要分为数据集、模型结构和损失构建3部分。数据集部分展示了部分物联网入侵检测数据的样例。为增强序列数据的深层全局特征提取能力,进一步提高物联网入侵检测能力,本文提出一种面向类不平衡样本的物联网入侵检测方法,称为FATIDS,模型主要由输入层、线性映射层、位置编码、FATIDS特征编码器、Softmax层和输出层构成,最后,在损失构建部分使用Focal Loss改进原始的交叉熵损失函数。针对高维且类不平衡的物联网流量数据,从模型层面,FATIDS在特征提取阶段利用自注意力机制能够自适应特征选择与特征增强,而后利用前馈神经网络有效挖掘全局特征的关联关

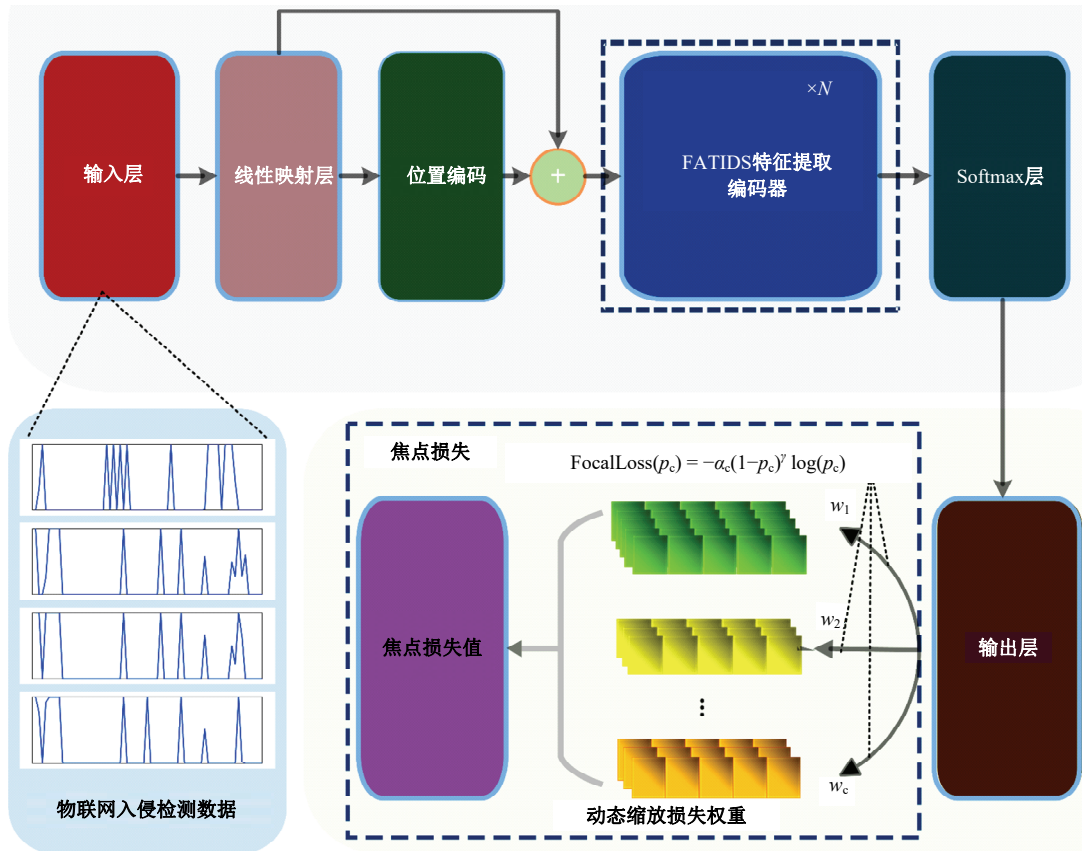


图 1 FATIDS 网络架构示意图

Fig. 1 FATIDS architecture diagram

系,提取类内和类间高分辨特征,实现了端到端的特征提取与检测;损失构建层面,使用 Focal Loss 根据预测的概率动态调整类别的损失权重,自适应聚焦检测困难的少数样本类别,从而克服模型对多数类的倾向性,进一步消除类不平衡问题的不良影响,提高物联网入侵检测能力的泛化性能。

1.1 位置编码

为了获得更加丰富的特征表示,首先对输入的序列数据进行线性映射,假设序列为 $U = [u_1, u_2, \dots, u_L]$, 其中 L 表示序列长度,现将序列特征向量映射到 d 维,作为模型的特征维度。映射计算方式为:

$$x_i = W_e u_i + b_e \quad (1)$$

式中: x_i 为映射后的高维序列; W_e 、 b_e 分别为参数矩阵和偏置项。

卷积神经网络和循环神经网络通过连续处理时间序列得到位置信息。而 Transformer 是一种基于自注意力机制的网络,自注意力机制对位置信息是不敏感的^[19],需要添加位置编码为时间序列提供位置信息。

位置编码在嵌入层之后增加位置信息,位置编码嵌入位置信息的计算方式为:

$$f(x_i, i) = W(x_i + p_i) \quad (2)$$

式中: i 为 x_i 的位置; $p_i \in \mathbb{R}^d$ 为一个可训练的 d 维的位置向量, d 取决于 x_i 。位置编码为每个位置的序列数据增加相互独立的位置信息,位置向量跟随数据训练自动调整,能够有效表示网络流量特征之间的相关性。通过位置编码,得到具有位置信息的高维序列 Y 。

1.2 FATIDS 特征编码器

特征编码器是 FATIDS 的重要组成部分,特征编码器主要由多头注意力模块、前馈神经网络、批归一化层和残差连接层组成。针对高维流量序列数据,多头注意力模块能够提取序列数据的多种特征模式,与前馈神经网络结合以实现自适应全局特征选择和特征提取,有效挖掘类内和类间的特征关系,提取不同类别的高可分辨特征,增强少数类别的检测性能;由于流量数据量较大,利用批归一化层^[20]对隐藏层数据进行归一化处理,抑制内部协变量偏移,增强模型泛化性能;为避免浅层信息丢失,使用残差连接层^[21]抑制梯度问题,促进信息流

动,提高特征提取的稳定性;此外,为增强前馈神经网络的非线性特征表达能力,使用 GELU 激活函数学习更加复杂的非线性映射关系,并使用随机失

活层随机置零神经元,抑制模型因类别不平衡导致的迅速过拟合,提高特征提取的泛化性能。FATIDS 特征编码器的结构如图 2 所示。

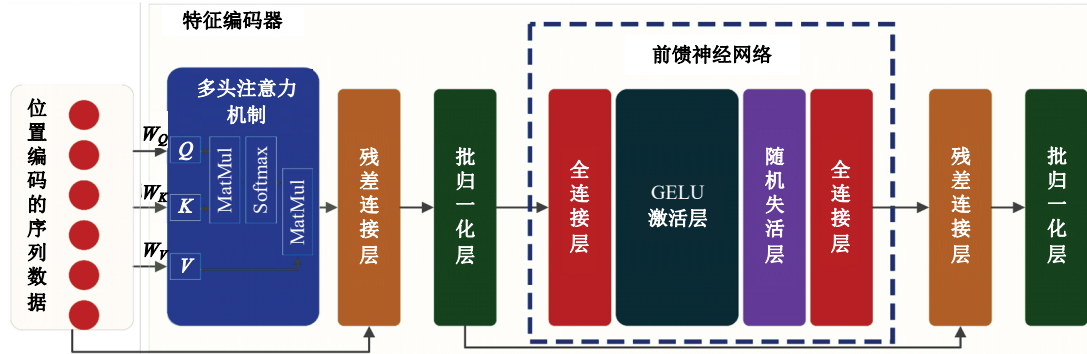


图 2 FATIDS 特征编码器结构图

Fig. 2 Schematic diagram of FATIDS feature encoder

参照图 2,将嵌入位置信息的高维序列 $Y = [y_1, y_2, \dots, y_n]$ 中的每个元素通过映射得到查询向量 $Q = YW_Q$,键向量 $K = YW_K$ 和值向量 $V = YW_V$,其中 $W_Q, W_K, W_V \in \mathbb{R}^{d_{\text{model}} \times d_h}$, d_h 表示第 h 个注意力头的维度。通过 Q, K, V 计算序列元素之间的注意力权值矩阵 $A \in \mathbb{R}^{n \times n}$:

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_h}}\right) \quad (3)$$

通过注意力权值矩阵 A 对值 V 进行加权,同时注意力权值矩阵通过数据驱动,故通过计算序列数据相关性,自注意力机制能够动态调整序列数据的注意力权值,以实现自适应的特征选择。

$$\text{SH}(X) = AV \quad (4)$$

多头注意力模块为 H 个自注意力模块对输入序列在不同层次上计算注意力权重,捕捉时间序列数据的长期依赖关系和局部特征,并将各尺度的输出进行拼接:

$$\text{MH}(X) =$$

$$[\text{SH}_1(X), \text{SH}_2(X), \dots, \text{SH}_H(X)]W_{\text{MH}} \quad (5)$$

式中:参数矩阵 $W_{\text{MH}} \in \mathbb{R}^{H \cdot d_H \times d_{\text{model}}}$; $\text{MH}(X)$ 为多头注意力模块的输出结果。多头注意力模块将嵌入位置信息的由高维向量组成的序列,通过可学习的参数矩阵分裂为查询向量、键向量和值向量,而后通过查询向量和键向量计算序列元素之间的相关性,得到注意力权值矩阵,将值向量与注意力权值矩阵进行运算得到单头注意力模块的输出,通过合并单头注意力模块的输出得到多头注意力模块的输出结果。

将多头注意力模块的输出与输入完成残差连接和批归一化处理后,作为前馈神经网络的输入,

前馈神经网络包含 4 层,第 1 层先将输入映射到高维空间,第 2 层使用非线性激活层 GELU 增强对特征的非线性表达能力,为防止过拟合问题,第 3 层使用随机失活层随机将神经元的输出置零,降低模型的复杂度,第 4 层再进行降维处理,过程为:

$$E_l = (\text{Dropout}(\text{GELU}(\text{BN}(\text{MH} + Y)W_1)))W_2 \quad (6)$$

$$O_l = \text{BN}(E_l + (\text{BN}(\text{MH} + Y))) \quad (7)$$

式中: O_l 为第 l 层特征编码器的输出; $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{FFN}}}$, $W_2 \in \mathbb{R}^{d_{\text{FFN}} \times d_{\text{model}}}$ 分别为两层维度变换的参数矩阵; $\text{BN}(\cdot)$ 为批归一化函数; $\text{Dropout}(\cdot)$ 为随机失活层。堆叠的特征编码器通过串行连接,将上一个特征编码器的输出作为输入,最终输出深层全局特征。

经过堆叠的 l 层特征编码器处理后,最后使用 Softmax 作为分类器,输出的结果为:

$$Y = \text{Softmax}(O_l) = \frac{\exp(O'_i)}{\sum_{k=1}^K O'_k} \quad (8)$$

式中: O_l 为第 l 层编码器的输出; $\exp(\cdot)$ 为 e 的指数函数; K 为总的类别数。

1.3 损失构建

数据集样本的类别不平衡问题是普遍存在的,包括目标识别、生物医学、航空航天和物联网入侵检测等^[22]。由于正常行为的样本数量远大于恶意样本的数量,给模型的准确性和可靠性带来了巨大挑战。常用于分类的损失函数为交叉熵损失函数,可表示为:

$$\text{CrossEntropy}(p, y) = \begin{cases} -\log(p), & y=1 \\ -\log(1-p), & \text{其他} \end{cases} \quad (9)$$

式中: $y \in \{-1, 1\}$ 为数据集中的正类和负类样本; $p \in [0, 1]$ 为在标签为1时模型的估计概率。交叉熵损失函数能够有效降低平均损失,但是在类不平衡数据集上,交叉熵损失倾向于将目标预测为多数类样本,导致少数类样本即恶意样本的识别精度降低,损害了模型的识别精度和可靠性。

为了解决物联网入侵检测面临的类不平衡问题,进一步提高模型识别性能,本文提出使用Focal Loss改进传统的交叉熵损失函数,模型的预测概率 p_c 定义为:

$$p_c = \begin{cases} p, & y=1 \\ 1-p, & \text{其他} \end{cases} \quad (10)$$

在类不平衡数据集中,交叉熵损失函数表现较差的原因是多数类样本主导了模型的损失,对少数类样本的关注程度不足,模型倾向于将样本分为多数类,以提高准确率。Focal Loss通过引入一个平衡因子调整样本的关注度。平衡因子通过对交叉熵损失的计算结果进行缩放,将损失的关注程度聚焦于难分样本,使得易分类样本的损失减小,难分类样本的损失增大。Focal Loss计算损失的过程为:

$$\text{Focal Loss}(p_c) = -\alpha_c(1-p_c)^\gamma \log(p_c) \quad (11)$$

$$\alpha_c = \begin{cases} \alpha, & y=1 \\ 1-\alpha, & \text{其他} \end{cases} \quad (12)$$

式中: p_c 为模型预测的概率; α 和 γ 为可调整的超参数。

在模型训练阶段,因类别不平衡少数类样本的检测难度远高于多数类样本,模型更倾向于将样本分为多数类样本,采用Focal Loss能够缓解模型的倾向性和过拟合问题,增强少数类样本的检测能力。通过分类器可输出类别预测的概率 $p_c \in [0, 1]$,由 p_c 可知样本分类的难易程度, p_c 趋近0说明样本分类困难, p_c 趋近1说明样本分类简单。在交叉熵损失函数中增加乘项 $1-p_c$,通过样本分类的难易程度动态调整损失的权重。此外,通过调整参数 α_c 和 γ 的值,提高控制损失对多数类和少数类样本的关注程度,Focal Loss能够使模型更加关注难以分类的少数类样本,抑制易分类多数类样本的影响。因此,在类不平衡的情况下,模型能够更好地学习到少数类样本的特征,

提高入侵检测任务的性能。

通过结合模型和训练损失两个层面,提取类内和类间的高可分辨特征,根据样本检测的难易程度动态放缩梯度,从特征提取和训练优化两个阶段克服类不平衡问题,提高模型泛化性能。

2 实验结果与分析

2.1 数据集

物联网入侵检测数据集除了数据来源的异质性外,还面临着严重的类不平衡问题。为实现对真实世界物联网和工业物联网的模拟,本文使用物联网公开的真实入侵检测数据集 ToN_IoT 和 DS2OS。ToN_IoT 是在操作系统、物联网或工业物联网服务和网络系统下生成的包含合法和攻击事件的集合。训练测试数据集包括物联网流量共 461 043 条(包括正常流量 30 万条、攻击流量 161 043 条),包含了 44 个特征。ToN_IoT 数据集中包含 9 种异常类型,分别是 Scanning、DoS、DDoS、Ransomware、Backdoor、Injection、XSS、Password Cracking 和 MITM 攻击类。按照 8:2 的比例划分训练和测试数据集,具体分布如表 1 所示。

表 1 ToN_IoT 数据集的样本类别分布

Table 1 Sample category distribution of the ToN_IoT dataset

类别	训练集数量	测试集数量	总量
normal	240 000	60 000	300 000
DDoS	16 000	4 000	20 000
DoS	16 000	4 000	20 000
injection	16 000	4 000	20 000
mitm	834	209	1 043
password	16 000	4 000	20 000
ransomware	16 000	4 000	20 000
scanning	16 000	4 000	20 000
xss	16 000	4 000	20 000
backdoor	16 000	4 000	20 000

DS2OS 数据集是由从不同物联网设备捕获的数据组成,以体现样本的异质性特点。数据集包含 8 种类型的物联网设备通信流量,即光控制器、温度计、运动传感器、洗衣机、电池、恒温器、智能门和智能手机。训练测试数据集包括物联网流量共 357 953 条(包括正常流量 347 936 条、攻击流量 10 018 条),包含了 12 个特征。DS2OS 包含 7

种类型的攻击,即 DoS、Probing、Malicious Control、Malicious Operation、Network Scan、Spying 和 Wrong Setup 攻击类。按照 8:2 的比例划分训练和测试数据集,具体分布如表 2 所示。

表 2 DS2OS 数据集的样本类别分布

Table 2 Sample category distribution of the DS2OS dataset

类别	训练集数量	测试集数量	总量
normal	278 348	69 587	347 935
scan	1 238	309	1 547
malitiousOperation	644	161	805
DoSattack	4 624	1 156	5 780
spying	426	106	532
dataProbing	274	68	342
wrongSetUp	98	24	122
malitiousControl	711	178	889

由于数据集包含大量非数字特征,需要将数据集的非数字特征转换为数字特征,以便模型训练顺利进行。在数据的预处理阶段,通常使用独热编码对特征进行数字化处理,但是考虑到采取独热编码会使特征维度非常庞大,不利于模型的特征提取。因此,仅对数据集中部分特征采用独热编码,剩余的非数字特征均采用标签编码,即不同类别采用不同的数字标签表示,相同类别采用相同数字标签表示。

将数据集中的非数字特征转换为整型数据,主要包含独热编码、标签编码。ToN_IoT 数据集进行独热编码的非数字特征包含 proto、conn_state 和 service 3 种,根据特征中标签的种类实现了特征的扩展,在所属标签下置 1,其余为 0。进行标签编码的非数字特征包含 dns_query、dns_AA 和 weird_notice 等 23 种特征。DS2OS 数据集均使用标签编码,并丢弃时间戳特征,使用 11 种特征。

经过数据转换后,得到数字化特征。因为数据集中不是所有特征的数值范围都在同一范围内,在模型训练中不利于模型的收敛。采用 min-max 标准化,使得数据集中数据在 [0, 1] 之间。min-max 标准化公式的计算过程为:

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (13)$$

式中: $X_{\min}$ 为特征列中的最小值; $X_{\max}$ 为特征列中的最大值; X_i 为特征第 i 行的值。

2.2 评价指标

为了验证本文方法的识别性能,采用以下评价指标评估本文模型的性能表现:准确率、精确率、召回率、 F_1 分数。评价指标的计算需要使用下述 4 个参数: TP——被模型预测为正常类的正常样本; TN——被模型预测为异常类的异常样本; FP——被模型预测为正常类的异常样本; FN——被模型预测为异常类的正常样本。

精确率 (Acc) 是正常样本和异常样本被正确分类所占比例,计算方式为:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

精确率 (Precision): 又名查准率,即预测为正确的正常数据占实际为正常数据的比例,计算方式为:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (15)$$

召回率 (Recall): 又名查全率,预测为正确的正常数据占实际为正常数据的比例,计算方式为:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (16)$$

F_1 分数值 (F_1 score): 因精确率和召回率为一对矛盾的度量指标,为平衡精确率和召回率而提出的精确率和召回率的调和平均值,计算方式为:

$$F_1 \text{ score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

2.3 多分类性能实验

为了验证本文方法的在物联网入侵检测任务中的检测性能,在 ToN_IoT 和 DS2OS 物联网入侵检测数据集进行验证实验。首先在 ToN_IoT 物联网入侵检测数据集上验证多分类的识别性能,本文方法对 10 类样本的检测性能结果如表 3 所示。

FATIDS 在 ToN_IoT 物联网入侵检测数据集上的 10 分类平均准确率、精确率、召回率和 F_1 分数分别为 99.60%、97.51%、96.59% 和 97.02%。如表 3 所示,本文方法在 normal、Ransomware、XSS 和 Backdoor 4 类样本中实现了较高的识别性能,精确率、召回率和 F_1 分数均达到了 99.76% 以上。在 DDoS、Password 和 Scanning 3 类样本中精确率、召回率和 F_1 分数均达到了 98.35% 以上。在 DoS 攻击检测中精确率、召回率和 F_1 分数分别达到了 98.92%、97.91% 和 98.41%,召回率指标相对较低;在 Injection 攻击

表 3 FATIDS在 ToN_IoT数据集上的多分类性能

Table 3 Multi-classification performance of FATIDS on ToN_IoT dataset

类别及平均值	准确率	精确率	召回率	F_1 分数
normal	—	99.98	100.00	99.99
DDoS	—	98.80	98.51	98.65
DoS	—	98.92	97.91	98.41
Injection	—	96.22	98.85	97.51
MITM	—	83.14	73.71	78.14
Password	—	98.86	98.88	98.87
Ransomware	—	100.00	100.00	100.00
Scanning	—	99.34	98.35	98.84
XSS	—	99.86	99.76	99.81
Backdoor	—	100.00	99.95	99.97
平均值	99.60	97.51	96.59	97.02

检测中精确率、召回率和 F_1 分数分别达到了 96.22%、98.85% 和 97.51%，精确率指标相对较低。结合表 1 和图 3 分析可知，DoS 攻击在样本数量与其他样本相同时，检测难度相对较大，DoS 和 DDoS 攻击类型都易被错分为 Injection 攻击类型。在 MITM 攻击检测中精确率、召回率和 F_1 分数分别达到了 83.14%、73.71% 和 78.14%，MITM 攻击样本仅有 1 043 个，占 ToN_IoT 数据集的 0.23%，检测效果相对其他样本类别下降较明显，其最易被错分为 Password 攻击类型，样本数量为 20 000 个，所占样本比例并不高，可见分类错误并非完全因为类不平衡问题。由于网络攻击样式本身检测难度不相同，当前类别权重的放缩能力有限，下一步将针对个别难分样本设置错分代价，以确保高效检测各类别的攻击样式。

为了直观对比 ToN_IoT 数据集中各类别的分类效果，FATIDS 对 ToN_IoT 数据集分类的混淆矩阵和评价指标如图 3 所示。

进一步在 DS2OS 物联网入侵检测数据集上验证多分类的识别性能，评价本文方法对 8 类样本的检测性能。检测结果如表 4 所示。

FATIDS 在 DS2OS 物联网入侵检测数据集上 8 分类的平均准确率、精确率、召回率和 F_1 分数分别为 99.47%、99.93%、95.77% 和 97.42%。如表 4 所示，在 scan、maliciousOperation、spying、dataProbing、wrongSetUp 和 maliciousControl 6 类样本中精确率、召回率和 F_1 分数均达到了 100%。在 normal 类型中精确率、召回率和 F_1 分数分别达

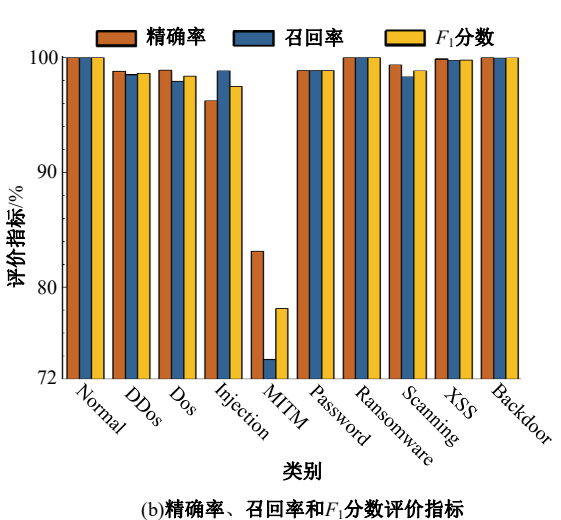
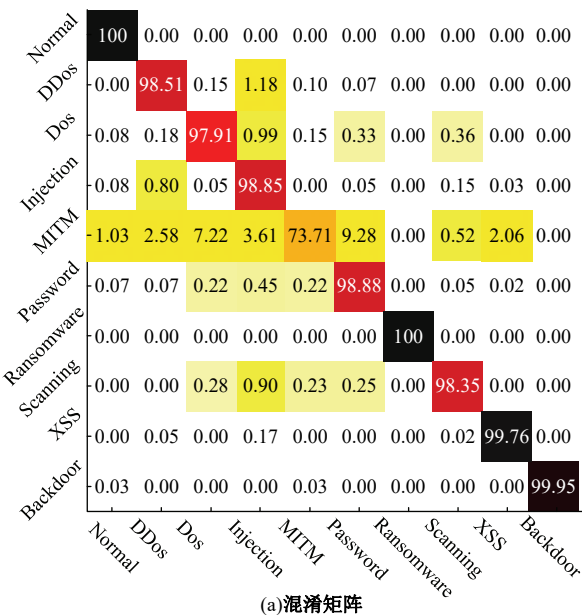


图 3 FATIDS在 ToN_IoT数据集上的多分类实验结果

Fig. 3 Experimental results of FATIDS for multi-classification on ToN_IoT dataset

表 4 FATIDS在 DS2OS数据集上的多分类性能

Table 4 Multi-classification performance of FATIDS on DS2OS dataset

类别及平均值	准确率	精确率	召回率	F_1 分数
normal	—	99.45	100.00	99.72
scan	—	100.00	100.00	100.00
maliciousOperation	—	100.00	100.00	100.00
DoSattack	—	100.00	66.20	79.66
spying	—	100.00	100.00	100.00
dataProbing	—	100.00	100.00	100.00
wrongSetUp	—	100.00	100.00	100.00
maliciousControl	—	100.00	100.00	100.00
平均值	99.47	99.93	95.77	97.42

到了 99.45%、100.00% 和 99.72%,而在 DoSattack 类型中精确率、召回率和 F_1 分数分别达到了 100.00%、66.20% 和 79.66%。结合表 2 和图 4 分析可知,DoSattack 类型召回率指标较低,33.80% 的 DoSattack 被错分为 normal 类型,然而 DoSattack 类型样本在 DS2OS 数据集中数量相较于其他攻击类型更多,可见 DoSattack 类型检测效果较差并非完全因为类不平衡问题,而是样本本身分类难度较大,下一步工作需着重提高个别难分攻击类型的检测能力,根据攻击类型的难易程度展开代价敏感研究。

为了直观对比 DS2OS 数据集中各类别的分类效果,FATIDS 对 DS2OS 数据集分类的混淆矩阵和评价指标如图 4 所示。

2.4 入侵检测方法对比实验

为了进一步验证本文方法的优越性,与先进的基线模型与入侵检测方法进行对比实验。本文选择 ExtraTrees-IDS^[23]、E-GraphSAGE^[24]、LSTM、GRU-FCN^[25]、ResNet、XCM^[26]、TST^[27] 和 GMS-IDS^[13] 共 8 种方法进行对比实验。为了实现最优的模型性能,对比实验的模型架构和参数根据参考文献进行设计。为了全面评估本文方法的先进性和有效性,选择准确率、精确率、召回率和 F_1 分数在数据集 ToN_IoT 验证各方法的检测性能,各方法对比实验在 10 分类任务中的检测结果如表 5 所示。

由表 5 可知,本文提出的 FATIDS 在准确率、精确率、召回率和 F_1 分数 4 项指标中分别达到了 99.60%、97.51%、96.59% 和 97.02%,在对比实验中实现了最优的检测性能,相较位居第二的 GMS-IDS 在准确率、精确率、召回率和 F_1 分数 4 项指标中分别提高了 1.44、0.65、1.76 和 1.32 个百分点。相较传统的基于 Transformer 的 TST 模型,准确率、精确率、召回率和 F_1 分数分别提高了 1.70、5.12、3.41 和 4.26 个百分点。相较 XCM、ResNet、GRU-FCN、LSTM 和 E-GraphSAGE 五种先进方法,准确率、精确率、召回率和 F_1 分数分别提高了 1.93、5.91、5.55 和 5.73 个百分点。相较传统的 ExtraTrees-IDS 物联网入侵检测方法,精确率、召回率和 F_1 分数分别提高了 41.09、11.35 和 29.12 个百分点,体现了 FATIDS 的优越性和有效性。

在 DS2OS 物联网入侵检测数据集上,选择

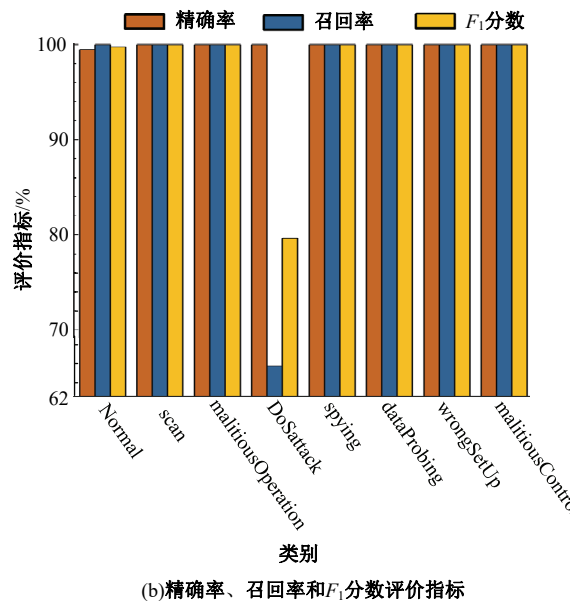
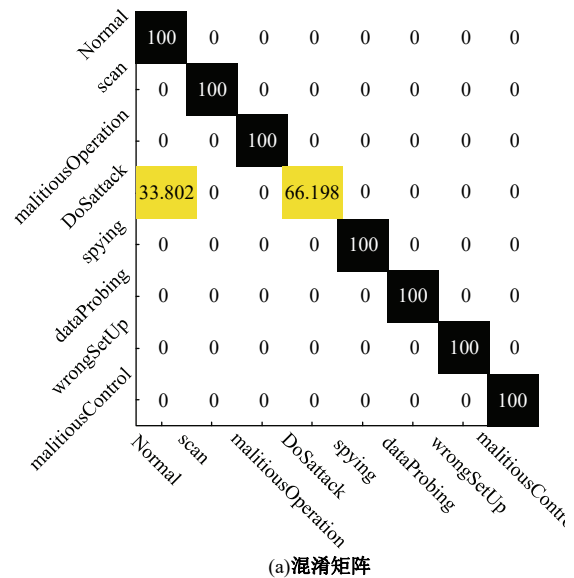


图 4 FATIDS 在 DS2OS 数据集上的多分类实验结果

Fig. 4 Experimental results of FATIDS for multi-classification on DS2OS dataset

LR、SVM、ANN^[28]、DRL、DRL with GAN^[29]、LSTM、HDRaNN^[30]、TCN^[31] 和 TST 共 9 种方法进行对比实验。各方法在 8 分类任务中的检测结果如表 6 所示。

如表 6 所示,本文提出的 FATIDS 在准确率、精确率、召回率和 F_1 分数 4 项指标中分别达到了 99.47%、99.93%、95.77% 和 97.42%,在对比实验中实现了最优的检测性能,相较位居第二的 TST 在准确率、精确率、召回率和 F_1 分数 4 项指标中分别提高了 0.04、0.02、3.59 和 2.11 个百分点。相较 ANN,在 DS2OS 数据集上准确率、精确率、召回率和 F_1 分数 4 项指标分别提高了 0.07、

表 5 物联网入侵检测模型在 ToN_IoT 数据集的对比实验结果

Table 5 Comparative experimental results of IoT intrusion detection models on ToN_IoT dataset %

IDS	准确率	精确率	召回率	F_1 分数
ExtraTrees-IDS	—	56.42	85.24	67.90
E-GraphSAGE	—	82.79	81.82	82.30
LSTM	92.70	76.49	76.00	75.97
GRU-FCN	96.17	88.30	86.57	87.30
ResNet	97.67	91.60	91.04	91.29
XCM	96.05	87.68	83.96	85.02
TST	97.90	92.39	93.18	92.76
GMS-IDS	98.16	96.86	94.83	95.70
FATIDS	99.60	97.51	96.59	97.02

表 6 物联网入侵检测模型在 DS2OS 数据集的对比实验结果

Table 6 Comparative experimental results of IoT intrusion detection models on DS2OS dataset %

IDS	准确率	精确率	召回率	F_1 分数
LR	98.30	45.82	27.75	31.56
SVM	98.20	44.50	24.12	27.71
ANN	99.40	99.22	95.63	97.03
DRL	98.96	63.00	76.00	67.00
DRL with GAN	99.02	67.00	86.00	72.00
LSTM	99.22	88.35	92.64	89.47
HDRaNN	98.56	89.01	89.74	89.27
TCN	99.22	88.35	92.64	89.47
TST	99.43	99.91	92.18	95.31
FATIDS	99.47	99.93	95.77	97.42

0.71、0.14 和 0.39 个百分点。相较 TCN、HDRaNN、LSTM、DRL with GAN 和 DRL 五种先进方法在准确率、精确率、召回率和 F_1 分数 4 项指标中分别提高了 0.25、10.92、3.13 和 7.95 个百分点以上。相较传统的基于 LR 和 SVM 的物联网入侵检测方法在准确率、精确率、召回率和 F_1 分数 4 项指标中分别提高了 1.17、54.11、68.02 和 65.86 个百分点以上,体现了 FATIDS 的优越性和有效性。

综上分析可得,FATIDS 能够提取物联网入侵检测序列数据的全局深层信息,并通过 Focal Loss 解决物联网入侵检测中的类不平衡问题,使用准确率、精确率、召回率和 F_1 分数 4 项指标进行评估,在 ToN_IoT 工业物联网数据集和异质设备

物联网数据集 DS2OS 上均实现了更加优异的检测性能。

2.5 超参数对比实验

超参数在深度学习模型中具有关键作用,在 FATIDS 中,模型训练的学习率 Lr、FATIDS 的编码器堆叠层数 Layer、多头自注意力机制的头数 Head 和 Focal Loss 的 Gamma 参数 γ 共 4 个超参数对模型性能具有重要影响,其中 Lr 影响模型的收敛效果,FATIDS 的编码器堆叠层数 Layer 影响模型深度,多头自注意力机制的头数 Head 影响网络流量特征模式的提取,Focal Loss 的 Gamma 参数 γ 调节对难分样本的关注程度。设置学习率 $Lr = \{0.0001, 0.001, 0.01, 0.1\}$,FATIDS 的编码器堆叠层数为 $Layer = \{1, 3, 5, 7\}$,多头自注意力机制的头数 $Head = \{4, 8, 16, 32\}$,Focal Loss 的 Gamma 参数 $\gamma = \{0, 1.0, 2.0, 3.0\}$,验证超参数对模型的性能影响。

FATIDS 在 ToN_IoT 物联网入侵数据集上的超参数实验结果见图 5。

如图 5(a)所示,随着堆叠层数 Layer 的增加,FATIDS 的准确率、精确率、召回率和 F_1 分数呈现先升高后降低的趋势,在 $Layer = 3$ 时实现了更优的性能,4 项指标分别为 99.60%、97.51%、96.59% 和 97.02%,适度增加编码器层数能够加深模型深度,提取抽象和高可分辨特征,模型过深会导致过拟合问题反而会损害检测性能。如图 5(b)所示,对于多头自注意力机制的头数 Head 超参数,随着注意力头数 Head 的增加,FATIDS 的准确率、召回率和 F_1 分数呈现先升高后缓慢降低的趋势,在 $Head = 16$ 时除精确率相较 $Head = 8$ 降低了 0.73%,其他 3 项指标均实现了更优的性能,注意力头数增加能够提取更多的序列数据特征模式,但随着特征模式过度增加而前馈神经网络提取能力有限,会引起过拟合问题导致检测性能缓慢下降。如图 5(c)所示,随着学习率 Lr 的增加,FATIDS 的准确率、精确率、召回率和 F_1 分数呈现先迅速提高后缓慢降低的趋势,在 $Lr = 10^{-3}$ 时实现了更优的性能,由波动趋势可知学习率对模型影响较大。如图 5(d)所示,令 $\alpha_c = 1$,Focal Loss 的 Gamma 参数 $\gamma = 0$ 时,Focal Loss 即交叉熵损失函数,随着 Gamma 参数 γ 的增加,FATIDS 的准确率、精确率、召回率和 F_1 分数呈现先降低后提高,然后又降低的趋势, $\gamma = 1.0$ 时 FA-

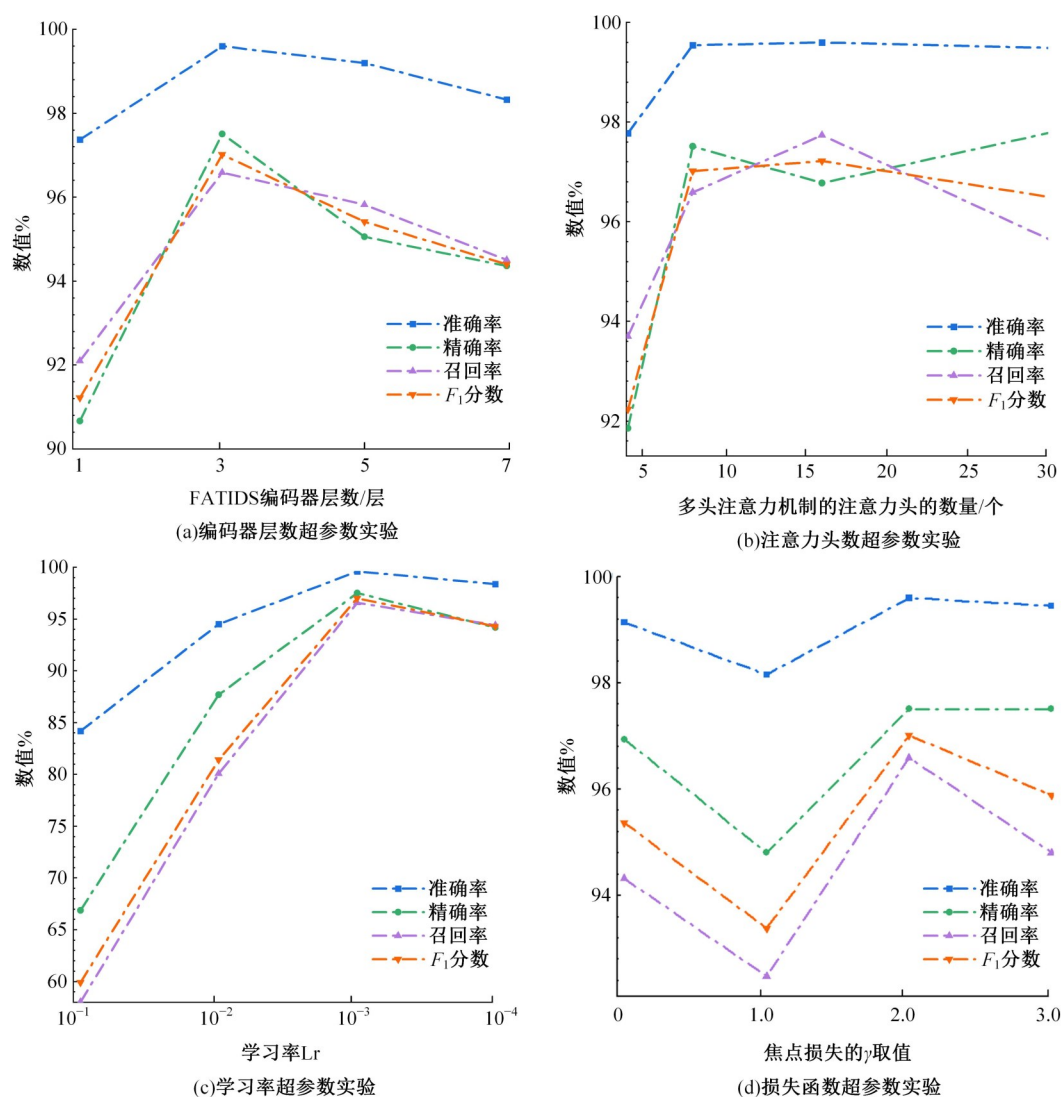


图5 FATIDS在ToN_IoT数据集上的超参数实验结果

Fig. 5 Hyperparameter experimental results of FATIDS on the ToN_IoT dataset

TIDS的性能低于使用交叉熵损失函数,在 $\gamma = 2.0$ 时实现了更优的性能,可见合理调整Gamma系数能够提高对难分样本的检测能力,Focal Loss对Gamma系数较为灵敏。

FATIDS在DS2OS物联网入侵数据集上的超参数实验结果见图6。

如图6(a)所示,随着堆叠层数Layer的增加,FATIDS的准确率、精确率、召回率和 F_1 分数呈现先提高后降低的趋势,其中准确率、精确率降低幅度较小,在Layer=3、Layer=5时实现了更优的性能,4项指标分别为99.47%、99.93%、95.77%和97.42%,模型深度能够影响特征提取的有效性,模型过浅会导致欠拟合问题,过深会导致过拟合问题。如图6(b)所示,随着注意力头数

Head的增加,FATIDS的准确率、精确率、召回率和 F_1 分数呈现先升高后降低的趋势,在Head=16时4项评价指标均实现了更优的性能,注意力头数增加能够提取更多的序列数据特征模式,但是注意力头数过多会引起过拟合问题,导致检测性能下降。如图6(c)所示,随着学习率Lr的增加,FATIDS的准确率、精确率、召回率和 F_1 分数呈现先迅速升高后降低的趋势,在 $Lr = 10^{-3}$ 时实现了更优的性能,由波动趋势可知学习率对模型检测性能影响较大。如图6(d)所示,令 $\alpha_c = 1$,Focal Loss的Gamma参数 $\gamma = 0$ 时,Focal Loss即交叉熵损失函数,随着Gamma参数 γ 的增加,FATIDS的准确率、精确率、召回率和 F_1 分数呈现先提高后降低的趋势,其中准确率和精确率波动幅

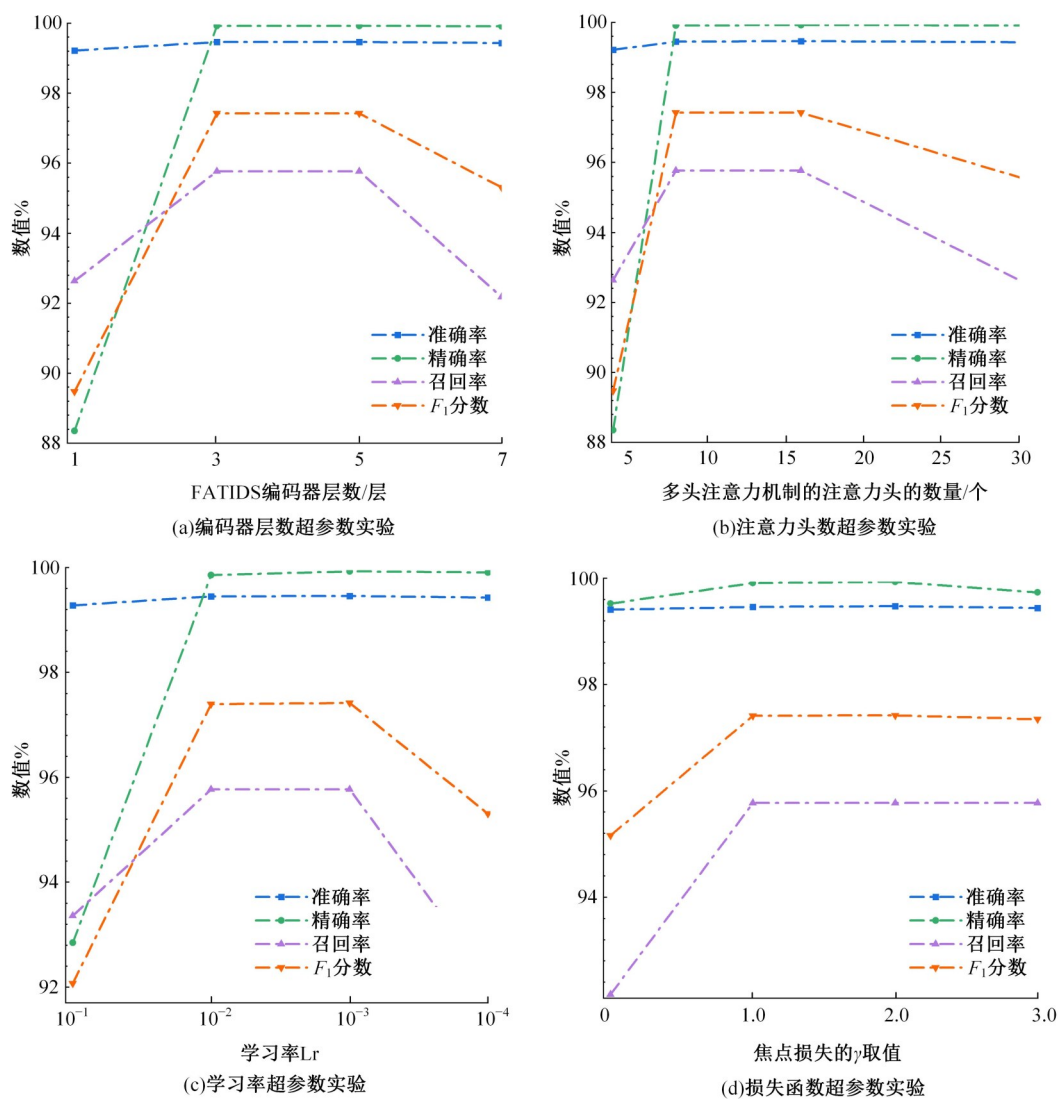


图 6 FATIDS在DS2OS数据集上的超参数实验结果

Fig. 6 Hyperparameter experimental results of FATIDS on the DS2OS dataset

度较小,小于0.06%, $\gamma=1.0$ 、 $\gamma=2.0$ 、 $\gamma=3.0$ 时召回率均为95.77%,综合4项评价指标在 $\gamma=2.0$ 时,实现了更优的性能,说明合理调整Gamma系数能够提高对难分样本的检测能力。

3 结束语

针对物联网入侵检测问题,本文提出了一种名为FATIDS的面向类不平衡样本的物联网入侵检测方法,采用Transformer动态调整模型对物联网网络流量特征的注意力,更加高效地提取网络流量的全局深层特征,并采用Focal Loss动态缩放模型梯度解决物联网入侵检测面临的类不平衡问题,自适应降低多数类样本的损失权重,增强模型对少数类样本的注意力。为验证本文方法的有效性,在物联网入侵检测公开数据集ToN_IoT

和DS2OS上验证了本文方法的有效性。对比其他先进的方法,在准确率、精确率、召回率和 F_1 分数指标上,本文方法均实现了更加优异的性能,最后验证了重要超参数对模型性能的影响。

参考文献:

- [1] Heidari A, Jabraeil J M A. Internet of Things intrusion detection systems: a comprehensive review and future directions[J]. Cluster Computing, 2022, 1: 1-28.
- [2] Kaur B, Dadkhah S, Shoeleh F, et al. Internet of Things (IoT) security dataset evolution: challenges and future directions[J]. Internet of Things, 2023, 22: No. 100780.
- [3] Khraisat A, Alazab A. A critical review of intrusion detection systems in the Internet of Things: tech-

- niques, deployment strategy, validation strategy, attacks, public datasets and challenges[J]. *Cybersecurity*, 2021, 4(1): No. 18.
- [4] Alani M M, Awad A I. An intelligent two-layer intrusion detection system for the Internet of Things[J]. *IEEE Transactions on Industrial Informatics*, 2023, 19(1): 683-692.
- [5] Pajouh H H, Javidan R, Khayami R, et al. A Two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks[J]. *IEEE Transactions on Emerging Topics in Computing*, 2019, 7(2): 314-323.
- [6] Saba T, Sadad T, Rehman A, et al. Intrusion detection system through advance machine learning for the Internet of Things networks[J]. *IT Professional*, 2021, 23(2): 58-64.
- [7] Albulayhi K, Abu Al-haija Q, Alsuhibany S A, et al. IoT intrusion detection using machine learning with a novel high performing feature selection method [J]. *Applied Sciences*, 2022, 12(10): No. 5015.
- [8] Islam N, Farhin F, Sultana I, et al. Towards machine learning based intrusion detection in IoT networks[J]. *Computers, Materials & Continua*, 2021, 69(2): 1801-1821.
- [9] Saba T, Rehman A, Sadad T, et al. Anomaly-based intrusion detection system for IoT networks through deep learning model[J]. *Computers and Electrical Engineering*, 2022, 99: No. 107810.
- [10] Abd Elaziz M, Al-qaness M A A, Dahou A, et al. Intrusion detection approach for cloud and IoT environments using deep learning and capuchin search algorithm[J]. *Advances in Engineering Software*, 2023, 176: No. 103402.
- [11] Almiani M, Abughazleh A, Al-rahayfeh A, et al. Deep recurrent neural network for IoT intrusion detection system[J]. *Simulation Modelling Practice and Theory*, 2020, 101: No. 102031.
- [12] Jeyanthi D V, Indrani B. IoT-based intrusion detection system for healthcare using RNNBiLSTM deep learning strategy with custom features[J]. *Soft Computing*, 2023, 27(16): 11915-11930.
- [13] 李晓佳, 赵国生, 汪洋, 等. 面向CNN和RNN改进的物联网入侵检测模型[J]. *计算机工程与应用*, 2023, 59(14): 242-250.
Li Xiao-jia, Zhao Guo-sheng, Wang Yang, et al. Improved intrusion detection model of Internet of Things for CNN and RNN[J]. *Computer Engineering and Applications*, 2023, 59(14): 242-250.
- [14] Altunay H C, Albayrak Z. A hybrid CNN+LSTM-based intrusion detection system for industrial IoT networks[J]. *Engineering Science and Technology, an International Journal*, 2023, 38: No. 101322.
- [15] Dina A S, Siddique A B, Manivannan D. A deep learning approach for intrusion detection in Internet of Things using focal loss function[J]. *Internet of Things*, 2023, 22: No. 100699.
- [16] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J/OL]. [2024-04-02]. <https://arxiv.org/pdf/1708.02002v2>
- [17] Alsaedi A, Moustafa N, Tari Z, et al. TON_IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven intrusion detection systems[J]. *IEEE Access*, 2020, 8: 165130-165150.
- [18] Khare S, Totaro M. Ensemble learning for detecting attacks and anomalies in IoT smart home[C]//3rd International Conference on Data Intelligence and Security (ICDIS), South Padre Island, USA, 2020: 56-63.
- [19] 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 (NIPS'17), Red Hook, USA, 2017: 6000-6010.
- [20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on International Conference, Lille, France, 2015: 448-456.
- [21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[J/OL]. [2024-04-03]. arXiv preprint arXiv:1512.03385.
- [22] Aguiar G, Krawczyk B, Cano A. A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework[J]. *Machine Learning*, 2024, 113: 4165-4243.
- [23] Wang Z G, Oates T. Imaging time-series to improve classification and imputation[J/OL]. [2024-04-04]. <https://www.ijcai.org/Proceedings/15/Papers/553.pdf>
- [24] Lo W W, Layeghy S, Sarhan M, et al. E-Graph-SAGE: a graph neural network based intrusion detection system for IoT[C]//NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 2022: 1-9.
- [25] Čeponis D, Goranin N. Investigation of dual-flow

- deep learning models LSTM-FCN and GRU-FCN efficiency against single-flow CNN models for the host-based intrusion and malware detection task on univariate times series data[J]. *Applied Sciences*, 2020, 10(7): No. 2373.
- [26] Fauvel K, Lin T, Masson V, et al. XCM: an explainable convolutional neural network for multivariate time series classification[J]. *Mathematics*, 2021, 9(23): No. 3137.
- [27] Zerveas G, Jayaraman S, Patel D, et al. A transformer-based framework for multivariate time series representation learning[C]//*Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event, Singapore, 2021: 2114-2124.
- [28] Hasan M, Islam M, Islam I, et al. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches[J]. *Internet of Thing*, 2019, 7: No. 100059.
- [29] Benaddi H, Jouhari M, Ibrahimi K, et al. Anomaly detection in industrial IoT using distributional reinforcement learning and generative adversarial networks[J]. *Sensors*, 2022, 22(21): No. 8085.
- [30] Huma Z E, Latif S, Ahmad J, et al. A hybrid deep random neural network for cyberattack detection in the Industrial Internet of Things[J]. *IEEE Access*, 2021, 9: 55595-55605.
- [31] Bai S J, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J/OL]. [2024-04-06]. <https://doi.org/10.48550/arXiv.1803.01271>