DOI: 10.13382/j.jemi.B2508113

基于生成对抗网络的流量异常检测方法*

陈万志[1] 尹明悦[1] 王天元[2]
(1.辽宁工程技术大学软件学院 葫芦岛 125105;2.国网辽宁省电力有限公司 营口 115005)

摘 要:针对流量异常检测模型因噪声和离群点干扰导致鲁棒性下降、特征表达能力不足,以及在处理不平衡高维海量数据时少数类检测率偏低等问题,提出一种基于生成对抗网络的流量异常检测方法。首先,采用基于聚类的 SCiForest 算法检测异常点,减少其对后续网络的影响。其次,设计以降噪自编码器为核心组件的生成对抗网络(denoising autoencoder-based generative adversarial network, DGAN),基于重建误差分布之间的 Wasserstein 距离定义其训练目标,生成可信的合成少数类样本,从而有效缓解数据不平衡问题。再次,通过与判别器一致的降噪自编码器(denoising autoencoder, DAE),输入真实样本与合成样本进行重构训练,得到优化后的编码器部分作为特征提取与降维模块,以增强特征的表达能力。最后,将处理后的数据输入融合卷积神经网络和双向门控循环单元的特征模型(feature fusion model of CNN and BiGRU, CNN-BiGRU-FFusion),在充分捕捉空间特征和时序特征的基础上实现分类与检测。在 NSL-KDD 数据集上的准确率和 F1 分数分别达到 92.06%和 92.25%,验证了所提方法在网络流量异常检测任务中的优越性能,并通过 CICIDS2017 数据集的实验进一步验证其可行性。

关键词:流量异常检测;异常点检测;生成对抗网络;降噪自编码器;卷积神经网络;双向门控循环神经网络单元
中图分类号:TP393;TN911.7 **文献标识码**:A **国家标准学科分类代码**:510.40

Traffic anomaly detection method based on generative adversarial networks

Chen Wanzhi[1] Yin Mingyue[1] Wang Tianyuan[2]
(1. College of Software, Liaoning Technical University, Huludao 125105, China; 2. State Grid Yingkou Electric Power Company of Liaoning Electric Power Supply Co. Ltd., Yingkou 115005, China)

Abstract: In response to the problems of decreased robustness and insufficient feature expression ability caused by noise and outlier interference in traffic anomaly detection models, and low minority class detection rates when dealing with imbalanced high-dimensional massive data, a traffic anomaly detection method based on generative adversarial networks was proposed. Firstly, the clustering based on SCiForest algorithm is used to detect outliers and reduce their impact on the subsequent training of the generative adversarial network. Secondly, a denoising autoencoder-based generative adversarial network (DGAN) is designed to generate reliable synthetic minority class samples. The network defines its training target based on the Wasserstein distance between reconstructed error distributions, effectively alleviating the problem of data imbalance. Again, using a denoising autoencoder (DAE) with the same architecture as the generative adversarial network discriminator, real and synthetic samples are input for reconstruction training, and the optimized encoder part is extracted as the feature extraction and dimensionality reduction module to enhance feature expression ability. Finally, the processed data is input into the feature fusion model of CNN and BiGRU (CNN-BiGRU-FFusion) model, which completes classification and detection based on capturing spatial and temporal features. The accuracy and F1 score on the NSL-KDD dataset reached 92.06% and 92.25%, respectively, verifying the superior performance of the proposed method in network traffic anomaly detection tasks. The feasibility of the method was further validated through experiments on the CICIDS2017 dataset.

Keywords: traffic anomaly detection; outlier detection; generative adversarial network; denoising autoencoder; CNN; BiGRU

收稿日期: 2025-01-09 Received Date: 2025-01-09
*基金项目:国家重点研发计划(2018YFB1403303)、辽宁省教育厅高校科研基金(2021LJKZ0327)项目资助

0 引 言

随着互联网的迅速发展,网络安全问题变得愈加严峻。大量敏感数据通过网络传输和存储,网络攻击的数量和复杂性也呈现出快速增长的趋势。不同类型的网络攻击通常伴随着大量异常流量,因此,流量异常检测已成为识别和防范网络攻击的关键环节[1]。

随着网络攻击的不断升级,机器学习和深度学习技术逐渐应用于流量异常检测领域。这些技术能够从海量数据中自动提取特征、发现隐藏的模式,并通过构建分类器有效区分正常与异常行为。相比传统的基于规则的流量异常检测方法,机器学习方法更适用于复杂的网络环境。陈万志等[2]提出一种空间因素背景基的基点分类方法,该方法通过提取训练中不同类别数据的背景基构造基点分类算法,在流量异常检测数据集上的二分类实验中表现出色。沈萍等[3]结合滑动窗口与信息熵设计了特征提取方法并构建孤立森林评分扩展模型,提升了异常类别的识别率。Mohiuddin 等[4]通过包装鲸鱼优化和正弦余弦算法筛选特征,并使用极限梯度提升树(extreme gradient boosting, XGBoost)精准识别异常流量。尽管如此,在处理大规模、高维且不平衡的网络流量数据时,传统机器学习方法仍存在局限性,难以充分应对特征表达能力不足及少数类检测率低等问题。

深度学习技术逐渐成为流量异常检测的研究焦点。梁欣怡等[5]提出一种自监督特征以及卷积神经网络与双向长短期记忆(convolutional neural networks and bidirectional long short-term memory network, CNN-BiLSTM)结合的入侵检测方法,通过自编码器进行数据增强,再利用特征增强的 CNN-BiLSTM 模型进行异常检测,实验结果验证了模型的高效性。李晓佳等[6]提出改进的物联网入侵检测模型,结合 CNN 和循环神经网络(recurrent neural network, RNN)以减少池化层信息丢失与梯度消失问题,并引入自注意力机制以提升多分类任务的表现。Alhassan 等[7]利用基于相关性的方法筛选高效特征,并借助自编码器(autoencoder, AE)区分正常流量与攻击行为。尽管这些深度学习方法在流量异常检测领域展现了强大的潜力,但在现实网络中攻击流量的占比相对相对较低,导致流量异常检测数据集中普遍存在明显的类别不平衡问题,这种不平衡增加了少数类样本检测的困难,使得模型在少数类样本的识别上表现能力不佳。

针对网络流量中正常流量与攻击流量失衡的问题,研究者提出了多种处理不平衡数据的方法。常见的方法有随机过采样、SMOTE[8]和 ADASYN[9]等,通过过采样技术生成新的样本以扩充数据。然而,这些方法可能导致数据集过拟合或引入噪声[10]。与传统方法相比,生成对抗网络(generative adversarial network, GAN)在应对样本失衡问题上表现出色。GAN 通过模拟真实数据分布生成高质量的合成样本以实现数据增强,其中生成器负责生成与真实样本极为相似的合成数据,判别器则通过对抗学习具备强大的样本区分能力[11]。然而,现有 GAN 方法在处理噪声干扰、少数类样本生成质量和特征提取能力等方面仍有提升空间,这为进一步优化 GAN 在流量异常检测中的应用提供了方向。

综上所述,尽管网络流量异常检测领域已有多种创新方法和模型,但仍存在未能生成高质量的少数类样本而导致模型鲁棒性不足的问题。此外,传统过采样策略可能放大数据噪声和离群点,限制特征表达效果。因此,从数据预处理、数据增强、特征提取和检测模型 4 个层面考虑,本文提出一种基于生成对抗网络的流量异常检测方法,以降噪自编码器为核心的生成对抗网络(denoising autoencoder-based generative adversarial network, DGAN)数据增强方法,采用对称降噪自编码器作为 DGAN 的判别器,并使用降噪自编码器的解码器模块作为生成器。通过对抗训练机制,生成高质量的少数类样本,增强数据鲁棒性,有效缓解类别不平衡问题。

设计一种降噪自编码器(denoising autoencoder, DAE)模块,采用与 DGAN 的判别器一致的结构,通过重构输入样本进行训练,并提取优化后的编码器部分作为特征提取与降维模块,进一步增强特征表达能力。融合卷积神经网络和双向门控循环单元的特征模型(feature fusion model of CNN and BiGRU, CNN-BiGRU-FFusion)进行检测,分别捕捉特征之间的空间和时序关系,并以多层感知机(multilayer perceptron, MLP)作为特征融合模块将时空特征进行融合,最终实现更精确的分类和预测,从而显著提升流量异常检测的性能。

1 相关技术

1.1 降噪自编码器

AE 由编码器和解码器两部分组成,训练目标为最小化重构误差。

编码器公式为:

$$\boldsymbol{z} = f(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}) \tag{1}$$

式中:$\boldsymbol{z}$ 为潜在空间中的特征表示;f 为激活函数;$\boldsymbol{W}$ 和 $\boldsymbol{b}$ 是对应的权重与偏置向量。

解码器公式为:

$$\tilde{\boldsymbol{x}} = g(\boldsymbol{W}'\boldsymbol{z} + \boldsymbol{b}') \tag{2}$$

式中:g 为解码器的激活函数;$\boldsymbol{W}'$ 和 $\boldsymbol{b}'$ 是解码器的权重与偏置向量。

AE 被广泛用于降维、特征提取、数据压缩等任务。然而,其在处理含有噪声的数据时鲁棒性较差,难以应对实际应用中的数据噪声问题。

DAE 通过向数据添加随机噪声,迫使模型在训练过程中去除噪声,进而重构原始的干净输入。通过此学习方式既能避免过拟合又能增强隐藏层学习特征的鲁棒性,从而有效提高模型的泛化能力。DAE 的训练目标为最小化重构误差,其公式为:

$$L_{DAE}(\boldsymbol{x},\tilde{\boldsymbol{x}}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}-\tilde{\boldsymbol{x}})^2 \tag{3}$$

与 AE 相比,DAE 更注重提取隐藏在噪声下的原始数据特征,从而在流量异常检测等噪声环境复杂的任务中展现出独特优势,其结构如图 1 所示。

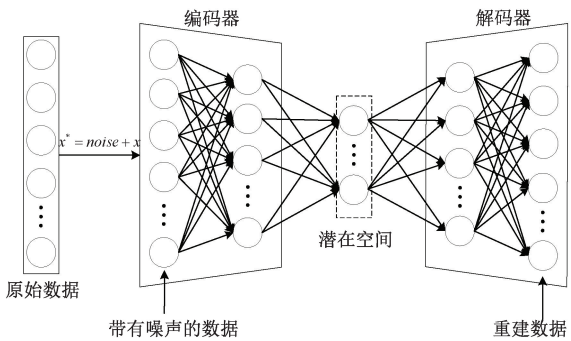


图 1　DAE 网络结构

Fig. 1　Network structure of DAE

1.2　生成对抗网络

GAN 由生成器 G 和判别器 D 组成[12]。生成器 G 尽可能生成接近真实数据的样本,而判别器 D 负责判定其真实性,其整体结构如图 2 所示。

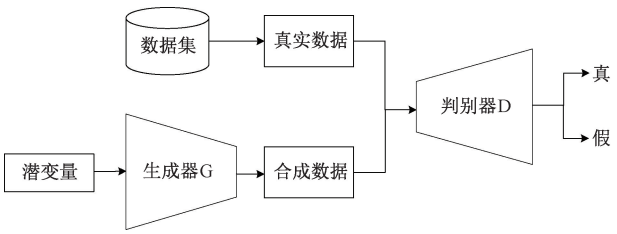


图 2　GAN 整体结构

Fig. 2　The overall structure of GAN

GAN 以零和博弈的方式进行对抗。假设潜在空间的概率分布为 p_z,真实数据的概率分布为 p_{data},GAN 的目标函数 $V(D,G)$ 表达为:

$$V(D,G) = \min_G \max_D E_{\boldsymbol{x}\sim p_{data}}[\log(D_{\theta_D}(\boldsymbol{x}))] + E_{z\sim p_z}[\log(1 - D_{\theta_D}(D_{\theta_G}(z)))] \tag{4}$$

式中:$D_{\theta_D}(\boldsymbol{x})$ 和 $1 - D_{\theta_D}(D_{\theta_G}(z))$ 分别表示判别器判断数据为真实或由生成器 G 生成的概率。

在训练过程中,判别器 D 的目标是最大化目标函数,以区分真假数据;而生成器 G 的目标是最小化目标函数,试图生成更加逼真的样本来“欺骗”判别器,使其无法区分真实样本和生成样本。最终,当 GAN 达到纳什均衡时,目标函数的最小最大问题得到最优解,此时生成器生成的数据与真实数据极为相似。

实际应用中,GAN 的训练往往不稳定,并难以实现理想的纳什均衡。为了解决这些问题,许多研究致力于提升 GAN 的训练稳定性[13-16]。其中,BEGAN 模型[16]提出一种基于重构误差的改进方法,通过计算真实数据与合成数据重构误差分布之间的 Wasserstein 距离,来评估它们的差异性,从而显著增强了 GAN 的训练稳定性。

BEGAN 的目标函数由判别器损失和生成器损失两部分构成。判别器损失的表达式为:

$$L_D = L(\boldsymbol{x};\boldsymbol{\theta}_D) - k_t \cdot L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D) \tag{5}$$

式中:$L(\boldsymbol{x};\boldsymbol{\theta}_D)$ 和 $L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D)$ 分别表示真实样本和生成样本的重构误差,k_t 是动态调整生成器和判别器权重的平衡因子。判别器的优化目标是通过减少真实样本和生成样本的重构误差差异,提高样本的辨别能力。生成器损失的表达式为:

$$L_G = L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D) \tag{6}$$

生成器的优化目标是让生成样本的重构误差分布尽可能接近真实样本的重构误差分布,从而生成更高质量的合成数据。为进一步平衡生成器和判别器的训练过程引入了平衡因子 k_t,其动态更新公式为:

$$k_{t+1} = k_t + \lambda_k \cdot (\gamma \cdot L(\boldsymbol{x};\boldsymbol{\theta}_D) - L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D)) \tag{7}$$

式中:$\gamma \in [0,1]$ 被称为 diversity ratio,用于调整真实样本和生成样本重构误差的权重比例;λ_k 为学习率,用于平滑 k_t 的更新过程。通过动态调整 k_t,模型能够在生成器与判别器之间找到稳定的平衡点,从而改善训练效果。

2　基于生成对抗网络的流量异常检测模型

2.1　总体框架

在网络流量异常检测任务中,各环节面临着不同的挑战。网络流量数据常常受到噪声和离群点的干扰,这些异常数据可能会干扰模型性能,从而导致分类结果不准确。为此,引入 SCiForest 算法检测并隔离离群点,提升数据质量。

此外,在高维且类别分布不平衡的数据集中,模型往往忽略少数类样本特征,导致检测能力下降。传统的解决方案通常通过对少数类样本进行过采样来缓解这一问题。然而,过度采样容易引入噪声。为了解决这一问题,采用 DGAN 方法生成可信的少数类样本,以提升数据的质量和模型的检测能力。

之后,为进一步增强特征表达能力,使用与 DGAN 判别器结构一致的 DAE 模型对扩展数据集进行特征提取与降维,确保特征空间的兼容性与稳定性。

最后,由于网络流量数据同时具备空间特性和时序性,传统的机器学习方法通常难以有效捕捉数据中的局部特征和空间依赖关系。为此,结合 CNN 与 BiGRU 分别提取空间与时间特征,并通过 MLP 模块融合多维信息,实现高精度的流量异常检测。所提出的流量异常检测模型的整体框架如图 3 所示。

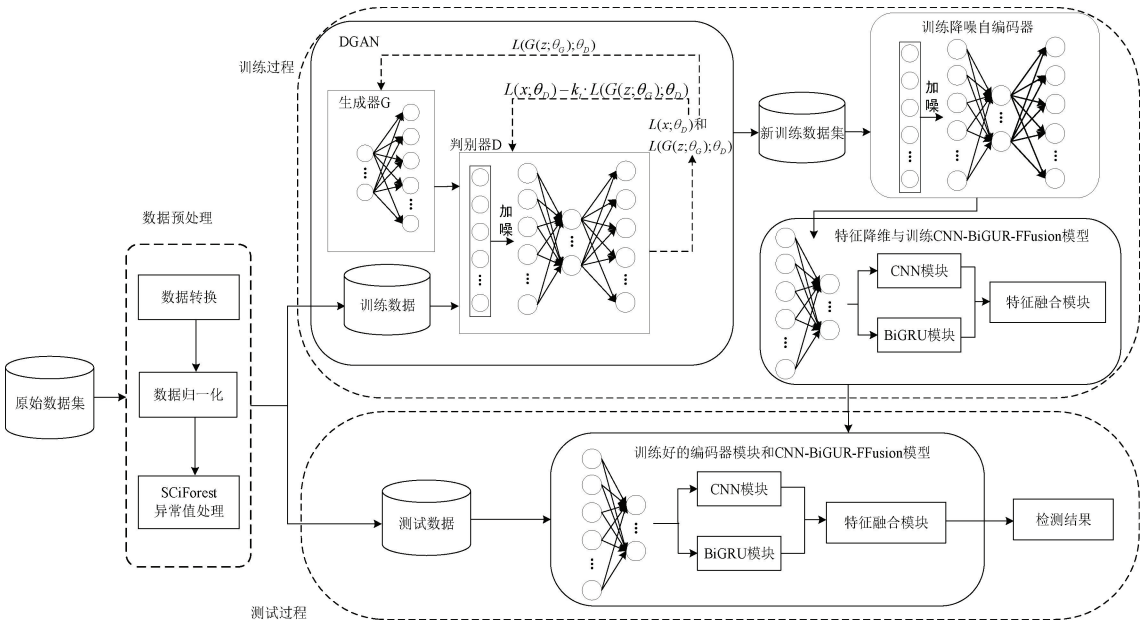


图 3　模型总体框架

Fig. 3　Overall framework of the model

2.2　检测流程

本文提出的基于生成对抗网络的流量异常检测方法由预处理、数据增强、特征降维和流量异常检测 4 个主要流程组成,算法流程如图 4 所示。

1)数据集预处理

(1)数据清洗

原始数据中可能存在特征值异常的“脏数据”,这些数据不仅无法用于模型训练和测试,还可能降低模型的检测性能。因此,首先需对这些数据进行清洗与剔除,以确保模型训练数据的质量。

(2)字符特征的二进制编码

数据集中通常包含字符类型的特征,为使模型能够处理这些数据,首先使用 LabelEncoder 方法将这些字符特征转换为数字编码,随后采用独热编码将每个分类特征映射为多个互斥的二进制特征。在处理过程中,确保测试集与训练集使用相同的类别列,并对测试集中缺失的类别进行填充。

(3)数据规范化处理

由于数据集中各个特征的取值范围差异较为显著,本文采用归一化方法规范数据,归一化处理后,特征值被缩放至相同的量纲范围内,避免因数值差异对模型产生偏差。其公式如下:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

(4)SCiForest 离群点检测

为提升检测准确性,采用 SCiForest 算法剔除异常数据。该算法通过构造一组孤立树对数据进行划分,并计算数据点到达叶子结点的路径长度。路径越短,表示越容易被孤立,异常的可能性越大,最终以平均路径长度作为异常得分依据,得分越高异常性越强。通过去除异常点,可有效缓解噪声干扰,提升模型的性能。

2)DGAN 模型

在流量异常检测中,生成高质量且具有可信度的合成数据,对于增强原始特征的表达能力和提高异常检测的准确性至关重要。传统的过采样方法往往容易导致过拟合或引入噪声,而本文提出的 DGAN 模型能够有效生成高质量的数据,增强数据的鲁棒性。

(1)数据准备

首先,将所使用的流量数据集按流量类型划分为多个子数据集。每个子数据集训练一个单独的 DGAN 模型,从而使每个 DGAN 模型只生成对应流量类型的合成数据。

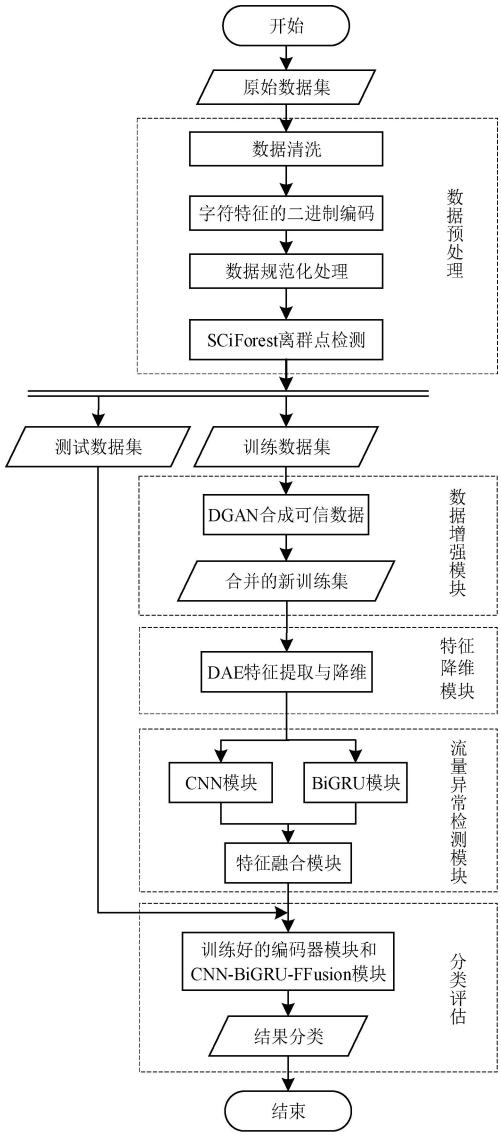


图 4　总体流程

Fig. 4　Overall procedures

(2)判别器的训练

首先,将真实数据和生成器生成的伪造样本进行加噪,为每个样本 $\boldsymbol{x}_i$ 添加高斯噪声,得到加噪样本 $\boldsymbol{x}_i^*$:

$$\boldsymbol{x}_i^* = \boldsymbol{x}_i + \mathcal{N}(0,\sigma^2) \tag{9}$$

其次,将加噪的数据输入到降噪自编码器的编码器部分,使用式(1)将数据映射为潜在空间表示 $\boldsymbol{z}_i$ 。

然后,降噪自编码器的解码器通过式(2)将潜在空间表示 $\boldsymbol{z}_i$ 解码为重构数据 $\tilde{\boldsymbol{x}}_i$ 。

最后,判别器通过式(3)分别计算真实样本和生成样本的重构误差,并根据式(5)对判别器参数进行优化。

(3)生成器的训练

首先,从潜在空间采样得到随机变量 $\boldsymbol{z}_i$,将其输入到由降噪自编码器解码器部分构成的生成器中,生成伪造样本 $\boldsymbol{x}_{fake} = G(\boldsymbol{z};\boldsymbol{\theta}_G)$ 。

之后,生成样本 $\boldsymbol{x}_{fake}$ 被输入到判别器中进行评估,根据公式(3)计算生成样本的重构误差,并将其作为生成器的损失函数。生成器通过最小化该损失函数,使生成样本无限接近真实样本。

(4)平衡因子 k_t 的更新

为了平衡生成器和判别器的训练过程,根据式(7)的动态平衡因子 k_t 平衡生成器和判别器的训练过程,避免生成器和判别器过度优化,从而确保训练保持平衡。

(5)训练终止的标准

在流量异常检测任务中,GAN 的训练终止标准是一个重要的考虑因素,直接影响其生成数据的质量。引入收敛标准作为训练终止标准,其公式为:

$$\mathcal{M} = L(\boldsymbol{x};\boldsymbol{\theta}_D) + |\ \gamma \cdot L(\boldsymbol{x};\boldsymbol{\theta}_D) - L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D)\ | \tag{10}$$

式中: $L(\boldsymbol{x};\boldsymbol{\theta}_D)$ 和 $L(G(\boldsymbol{z};\boldsymbol{\theta}_G);\boldsymbol{\theta}_D)$ 分别表示真实样本和生成样本的重构误差; $\gamma \in [0,1]$ 为 diversity ratio。

当满足下面任一条件时训练终止:①已达自定义训练轮数的上限;②当收敛标准 $\mathcal{M}$ 达到设定阈值时。

(6)训练完成后,利用训练好的 DGAN 根据原始数据中各类攻击样本的数量,生成指定类别的合成攻击样本,并将合成的样本与原始训练数据结合,构建一个新的平衡训练集。

3)基于 DAE 的特征降维

在特征降维阶段,采用基于 DAE 的特征提取与降维方法。考虑到前一阶段的 DGAN 的判别器模型本身已采用对称结构的 DAE 模型用于识别真实和生成样本,其体系结构已具备良好的特征学习能力,因此本模块延续该结构用于降维任务。在体系结构上,二者具有相同的数据输入格式与特征分布,使得结构统一具有较好的兼容性与稳定性。构建 DAE 模型后,使用由原始数据与合成样本构成的扩展数据集对其进行训练。训练完成后提取其中的编码器部分作为特征降维模块。在测试阶段,不再对输入数据添加噪声。此时使用已经训练好的编码器模块直接对输入数据进行特征提取和降维,并将其结果用于后续的异常检测。

4)基于 CNN 和 BiGRU 的特征融合流量异常检测模型

所提出的 CNN-BiGRU-FFusion 流量异常检测模型主要有 CNN 模块、BiGRU 模块和特征融合模块三个部分,采用并行架构将 CNN 模块与 BiGRU 模块结合,通过特征融合模块对捕捉的时空特征进行融合,最终完成分类任务。

(1)CNN 模块

CNN 作为一种深度学习模型,因其在图像和音频等输入类型上的出色表现而广泛应用。卷积层是其核心计

算单元。池化层通过采样减少输入特征的维度,降低计算复杂度。全连接层根据提取的特征执行分类任务。

采用一维卷积神经网络(1 D convolutional neural network, 1 D-CNN)提取数据的空间特征。1 D-CNN 通过卷积操作提取数据中的局部空间信息,池化层进一步提取更为广泛的特征,以提高模型的性能。为减小训练过程中内在协变量的变化,引入批归一化层,以加速模型的训练过程,并采用 ReLU 激活函数加速网络的收敛。

(2)BiGRU 模块

门控循环单元(gate recurrent unit, GRU)对 RNN 进行改进,有效地缓解传统 RNN 的长期依赖和梯度消失问题。然而,传统的 GRU 模型只能处理单向时序数据,这可能导致模型无法捕捉到与当前数据相关的重要信息。为了解决上述问题,采用 BiGRU 从前向和后向同时处理数据,从而有效捕捉流量数据中的时序特征,其公式如下:

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}_r \times [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{11}$$

$$\boldsymbol{z}_t = \sigma(\boldsymbol{W}_z \times [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{12}$$

$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W} \times [\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{13}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \odot \tilde{\boldsymbol{h}}_t \tag{14}$$

式中:$\boldsymbol{r}_t$ 为重置门;$\boldsymbol{z}_t$ 为更新门;$\tilde{\boldsymbol{h}}_t$ 和 $\boldsymbol{h}_t$ 为时刻 t 的候选隐藏状态和最终隐藏状态。

BiGRU 包含两个独立的 GRU 模块,一个处理序列的正向信息,另一个处理反向信息,通过拼接正向和反向的隐藏状态得到综合的隐藏状态。对于时间步 t,BiGRU 的最终隐藏状态 $\boldsymbol{h}_t^{BiGRU}$ 为:

$$\boldsymbol{h}_t^{BiGRU} = [\overrightarrow{\boldsymbol{h}}_t, \overleftarrow{\boldsymbol{h}}_t] \tag{15}$$

(3)特征融合模块

为充分整合 CNN 和 BiGRU 模块提取的空间特征与时序特征,采用 MLP 作为特征融合与分类模块。首先,将 CNN 提取到的空间特征标记为 $\boldsymbol{F}_{CNN}$,BiGRU 输出的双向时序特征表示为 $\boldsymbol{F}_{BiGRU}$,其次,将 $\boldsymbol{F}_{CNN}$ 与 $\boldsymbol{F}_{BiGRU}$ 进行拼接操作构成融合向量:

$$\boldsymbol{F}_{fusion} = \mathrm{Concat}(\boldsymbol{F}_{CNN}, \boldsymbol{F}_{BiGRU}) \tag{16}$$

融合后的向量 $\boldsymbol{F}_{fusion}$ 被输入到 MLP 中,首先经过一层隐藏全连接层进行非线性特征变换,之后连接输出层进行最终分类。整个模块支持端到端训练,损失函数通过反向传播影响 CNN、BiGRU 和 MLP 的权重更新,从而显著提升整体模型的表达能力与分类精度。

3 实验设计与结果分析

所有实验基于 AMD Ryzen 7 8845HS 处理器和 Window 10 操作系统实现,编程语言采用 Python3.9,使用 TensorFlow-GPU 2.4.2 软件环境。

3.1 实验数据集

1)NSL-KDD 数据集是对 KDD Cup 1999 数据集的改进,去除了其中大量冗余数据,并对训练集与测试集比例进行了合理划分[2]。其具体信息如表 1 所示。

表 1 NSL-KDD 数据集信息

Table 1 NSL-KDD dataset information

样本类别	训练样本数	测试样本数
Normal	67 343	9 711
Dos	45 927	7 460
Probe	11 656	2 885
R2L	995	2 421
U2R	52	67
总计	125 973	22 544

2)CICIDS2017 数据集源于加拿大网络安全研究所,涵盖了正常流量和 14 种不同类型的攻击。为避免失衡,本文从正常数据中抽取部分样本与所有攻击数据拼接为实验数据集[5]。CICIDS2017 数据集详细数据分布如表 2 所示。

表 2 CICIDS2017 数据集信息

Table 2 CICIDS2017 dataset information

类别	描述	样本数
正常数据	BENIGN	529 918
攻击数据	DoS Hulk	231 073
	PortScan	158 930
	DDoS	128 027
	Dos GoldenEye	10 293
	FTP-Patator	7 938
	SSH-Patator	5 897
	DoS Slowloris	5 796
	DoS Slowhttptest	5 499
	Bot	1 966
	Web Attack-Brute Force	1 507
	Web Attack-XSS	652
	Infiltration	36
	Web Attack-Sql Injection	21
	Heartbleed	11
总计	-	1 087 564

3.2 实验评估指标

采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数(F1-score)4 个指标评估本文模型性能,计算公式如下:

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

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

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

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{20}$$

式中：*TP* 和 *TN* 分别表示正确分类为攻击和正常的样本数；*FP* 和 *FN* 分别表示错误分类为攻击和正常的样本数。

3.3　超参数设置

超参数的选择对模型性能有重要影响，合适的超参数配置能够显著提高模型的准确率与泛化能力。

1）SCiForest 参数设置

SCiForest 离群点检测用于剔除远离大多数数据的离散点，提升异常检测的精度。核心参数 n_trees 定义了孤立树的数量。

通过实验测试不同 n_trees 对检测性能的影响，最终确定 n_trees = 150 时检测效果最佳。去除离群点前后 NSL-KDD 数据集规模的对比结果如表 3 所示。

表 3　SCiForest 去除离群点前后样本数据大小

Table 3　The size of the sample data before and after outlier removal using SCiForest

样本类别	离群点去除前	离群点取出后
Normal	67 343	63 572
Dos	45 927	44 526
Probe	11 656	11 344
R2L	995	995
U2R	52	52

2）DGAN 参数设置

采用的 DGAN 模型的判别器为一个具有 5 层结构的对称降噪自编码器。判别器的编码器和解码器均由包含 90 个神经元的隐藏层组成，潜在空间的维度设置为 50，生成器的潜在空间和解码器隐藏层与判别器的相应模块保持一致。为提高学习稳定性，模型为每个隐藏层应用批归一化，并采用 ReLU 激活函数。对于训练终止标准，本文设定的收敛度量阈值为 0.06，训练的最大迭代次数（epoch）为 200。当模型的收敛度量低于阈值或达到最大 epoch 时，训练将终止。

利用 DGAN 对 NSL-KDD 数据集中的 Probe、R2L 和 U2R 类别进行样本扩充，分别扩充 5 000、2 000 和 1 000 个样本，扩充后的数据将与原数据集合并，构建新的训练集。同时，由于用于降维和特征提取的 DAE 模型与 DGAN 的判别器在结构上完全一致，因此 DGAN 的配置同样适用于特征降维模块。

为验证 DGAN 模型的训练收敛性与稳定性，本文监测了 NSL-KDD 数据集中 Probe、R2L 和 U2R 三类少数样本在训练过程中生成器与判别器的损失函数变化以及收敛度量 $\mathcal{M}$ 的动态变化，相关训练曲线如图 5 所示。

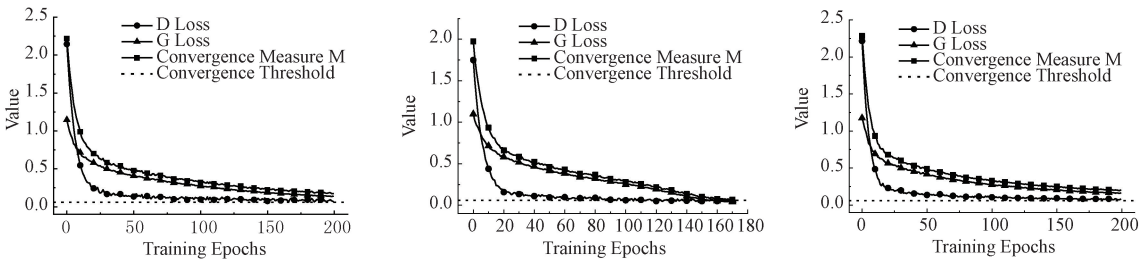


图 5　Probe、R2L 和 U2R 三类训练曲线

Fig. 5　Training curves for Probe, R2L and U2R classes

从图 5 可以观察到，训练初期，3 类样本的生成器和判别器损失均迅速下降，约在第 25 轮后下降速率减缓。对于 Probe 与 U2R 类，生成器与判别器的损失在接近 200 轮次时基本稳定，且收敛度量波动收窄，表明训练过程稳定并达成收敛。R2L 类别则在第 172 轮其收敛度量 $\mathcal{M}$ 率先下降至预设阈值 0.06，触发提前终止机制，显示该类样本训练已满足收敛条件。由此可得出 DGAN 模型具有良好的稳定性和收敛性。

3）CNN-BiGRU-FFusion 模块参数设置

CNN 模块用于提取数据中的空间特征。使用 1D-CNN 模型提取空间特征，其包含两个卷积层，分别使用大小为 3 的 32 和 64 个卷积核。每个卷积层后都接有 2×2 的最大池化层，并应用批归一化来减少内部协变量偏移。最后，全连接层将卷积和池化后的特征展平，并将其传递到后续的特征融合模块。

BiGRU 模块用于提取数据中的时序特征，捕捉前后时序关系。BiGRU 的隐藏层单元数设置为 64，并且采用双向 GRU 结构，以更全面地捕捉数据中的时间依赖关系。此外，为增强模型的泛化能力，BiGRU 模块中引入了 dropout 层，丢弃率设为 0.2，并使用 ReLU 作为激活函数。

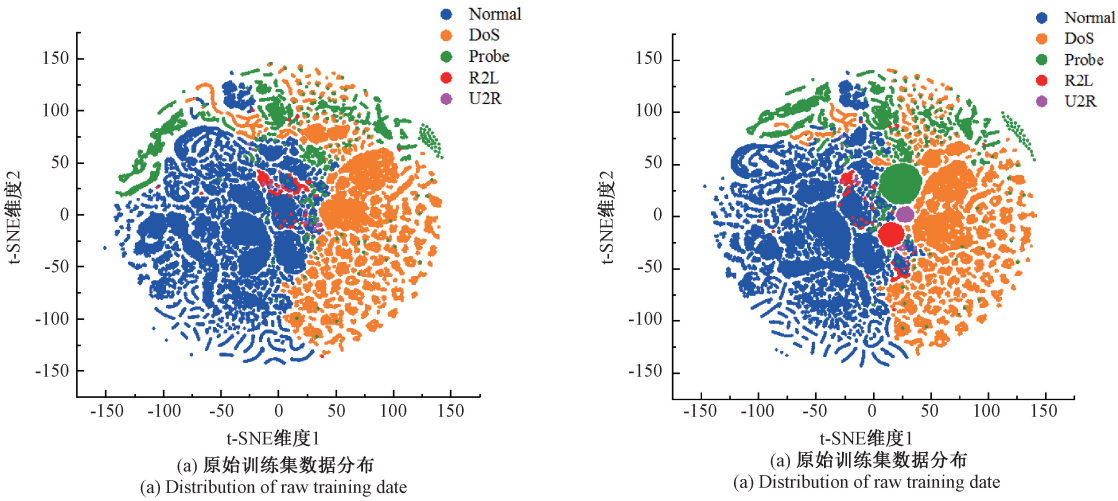
特征融合模块采用 MLP，输入层将 CNN 和 BiGRU 提取的特征进行拼接，形成一个融合向量。隐藏层单元数设置为 64，为适应多类别分类任务，输出层采用 Softmax 激活函数。模型整体采用 Adam 优化器优化，使用交叉熵损失函数。

3.4　实验结果分析

1）DGAN 生成样本质量分析

为评估 DGAN 生成样本的质量与分布特性，采用 t-

分布随机近邻嵌入(t-distributed stochastic neighbor embedding, t-SNE)算法对 NSL-KDD 的原始训练集与 DGAN 增强后的训练集进行可视化对比。该方法可将高维特征映射至二维空间,以直观展示各类样本的分布结构。原始训练集与加入合成样本后的扩展训练集在二维空间中的分布情况如图 6 所示。



(a) 原始训练集数据分布
(a) Distribution of raw training date

图 6　DGAN 数据增强前后训练集数据分布

Fig. 6　Distribution of training set data before and after DGAN data augmentation

从图 6 可以观察到,经过 DGAN 生成样本扩充后的训练集在样本边界上更加清晰,尤其是 Probe、R2L 与 U2R 三个少数类的分布更加明显,显著改善了原始训练集中存在的类别混叠现象。这种分布优化有效缓解了分类器因数据稀疏导致的过拟合风险,从而为后续分类模型的训练提供了更优的数据基础。

2)数据增强比较实验

为了验证 DGAN 算法在数据增强中的优越性,设计不同采样方法的对比实验。在相同的检测模型条件下,采用 ROS[17]、SMOTE[8]、ADASYN[9]、ADASYN-WGAN[18]和本文方法通过 NSL-KDD 数据集进行处理,精确率比较实验结果如图 7 所示。

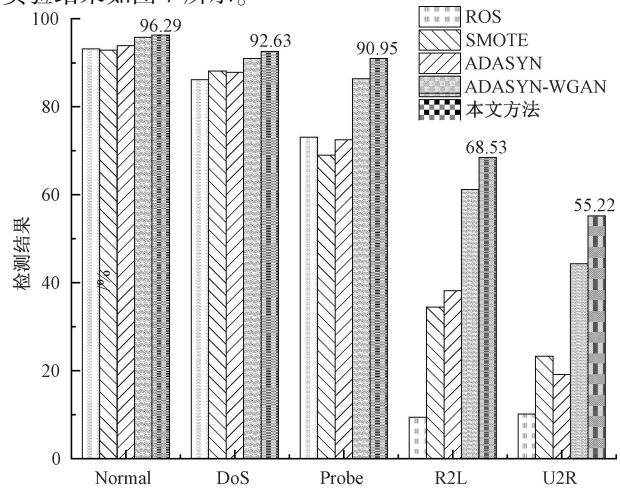


图 7　数据增强方法对比

Fig. 7　Comparison of different data augmentation methods

由图 7 可知,由于 Normal 类属于多数类,且 DoS 类的样本数占比也明显高于其他类别,因此这两个类别的精确率都比较高,而对于少数类 Probe、R2L 和 U2R,本文方法明显优于其他数据增强算法,这是由于 ROS 方法仅对原始数据进行简单的重采样,容易导致数据过拟合现象;SMOTE 方法虽然有效避免了过拟合,但可能加剧类内数据分布的不平衡、并且容易放大噪声,从而影响分类效果;ADASYN 算法根据插值原理合成样本,但其性能受到样本分布不均衡和噪声的影响。本文通过重构误差的改进方法,克服了 WGAN 在训练过程中可能出现的模型崩塌和训练不稳定的问题,使得生成器生成更接近真实数据的可信合成样本,显著提升各类别准确率。

3)特征降维对比实验

为验证用于特征降维的 DAE 模块与 DGAN 的判别器结构一致性是否有助于提升特征降维性能,设计多种特征降维结构进行对比实验。在保持检测模块不变的条件下,分别构建 DAE-Small、DAE-Large、DAE-Deep 以及 VAE 模型 4 种降维结构,并在 NSL-KDD 数据集上进行性能评估。各模型均采用对称结构设计,具体设置如下:DAE-Small 采用隐藏层神经元数为 80,潜在空间维度为 40,表示较小参数规模的结构;DAE-Large 采用 120 个神经元和 60 维潜在空间,表示较大参数规模的结构;DAE-Deep 采用双层编码器结构,分别设置 120 和 90 个神经元,潜在空间维度为 50;VAE 模型的隐藏层采用 90 个神经元,潜在空间维度设置为 50,各模型性能结果如表 4 所示。

表 4　不同降维方法对比结果

Table 4　Comparison results of different dimensionality reduction methods　　(%)

方法	Normal	DoS	Probe	R2L	U2R
DAE-Small	94.61	82.61	79.18	46.35	38.15
DAE-Large	95.86	87.54	83.60	62.17	52.09
DAE-Deep	95.23	88.67	84.95	59.26	49.71
VAE	96.09	90.34	86.28	50.67	47.96
本文	96.29	92.63	90.95	68.53	55.22

从表 4 可以看出,本文所采用的 DAE 模型在各类样本上的精确率均优于对比模型。相比之下,参数规模较小的 DAE-Small 在特征学习方面存在不足,导致精度偏低;DAE-Large 和 DAE-Deep 尽管增加了参数规格,但并未带来更明显的性能提升,且存在一定的过拟合风险;VAE 模型在 Normal 和 DoS 类中表现良好,但由于其结构设计与 DGAN 体系不统一,对于 Rrobe、R2L 和 U2R 三个少数类兼容性较差,导致这 3 个类别的检测性能低于本文方法。上述结果表明,本文设计的 DAE 结构在特征降维方面更具有优势,能有效提升整体分类性能。

4)消融实验

为了验证 SCiForest、DGAN、DAE 和 CNN-BiGRU-特征融合模块结合使用的有效性,基于 NSL-KDD 数据集上进行系列消融实验。以 CNN 和 BiGRU 单独进行异常检测作为基线模型,逐步添加模块,验证各模块对整体性能的贡献,实验结果如表 5 所示。

表 5　消融实验结果

Table 5　Ablation experiment results　　(%)

模型	SCiForest	DGAN	DAE	CNN	BiGRU	特征融合模块	Accuracy	Precision	Recall	F1-score
基线模型:CNN				√			81.23	80.25	81.23	80.74
基线模型:BiGRU					√		80.68	82.08	80.68	81.37
CNN+特征融合模块				√		√	81.75	81.28	81.75	81.51
BiGRU+特征融合模块					√	√	81.15	82.61	81.15	81.87
CNN-BiGRU-FFusion				√	√	√	82.40	83.75	82.40	83.07
DAE+CNN-BiGRU-FFusion			√	√	√	√	84.30	85.15	84.30	84.72
DGAN+DAE+CNN-BiGRU-FFusion		√	√	√	√	√	90.26	90.72	90.26	90.49
本文	√	√	√	√	√	√	92.06	92.45	92.06	92.25

从表 5 可以看出,单独使用 CNN 和 BiGRU 进行流量异常检测时,模型的准确率分别为 81.23%和 80.68%,这是因为 CNN 能够有效提取数据的空间特征,而 BiGRU 擅长捕捉双向的时序特征,因此两者均对异常检测性能有一定提升。接下来,CNN 和 BiGRU 分别与特征融合模块结合,模型的准确率达到 81.75%和 81.15%,说明以 MLP 作为特征融合模块能够充分学习特征之间的关系,并通过反向传播的方式优化网络权重,从而提高分类性能。之后将 CNN 和 BiGRU 提取的特征共同输入到特征融合模块,结合成完整的异常检测模块后,模型的准确率提高至 82.40%,原因是异常检测模型能够充分学习时空特征,进一步增强了分类能力。在此基础上引入 DAE 模块,模型的准确率和 F1 分数分别提升至 84.30%和 84.72%,这是因为 DAE 能够对数据特征进行有效学习和降维,有效增强了特征的表达能力。随后,结合 DGAN 生成高质量的少数类样本,模型的准确率和 F1 分数分别提升至 90.26%和 90.49%,这一提升表明 DGAN 模块能够有效缓解数据失衡问题,生成接近真实分布的少数类样本,从而显著提高少数类的检测性能。最后,通过加入 SCiForest 模块,有效去除了数据中的异常点,进一步优化数据,最终模型的准确率和 F1 分数分别达到了 92.06%和 92.25%。这些实验结果表明,各模块的结合能够有效提升模型性能,充分验证了本文方法的有效性。

5)与其他方法对比

为全面验证所提方法的性能,与其他基于 NSL-KDD 数据集的检测方法进行对比,对比结果如表 6 所示。

表 6　NSL-KDD 数据集上与现有模型的比较结果

Table 6　Comparison results with existing models on the NSL-KDD dataset　　(%)

方法	Accuracy	Precision	Recall	F1-score
文献[19]	86.59	88.55	86.59	86.88
文献[20]	90.99	91.39	90.94	90.89
文献[21]	91.74	91.62	91.86	91.54
文献[22]	90.64	91.78	90.28	91.05
本文	92.06	92.45	92.06	92.25

从表 6 可以看出,本文方法的评价指标均优于对比模型。实验结果表明,通过 DGAN 生成少数类样本并利用 DAE 进行特征降维,之后采用 CNN-BiGRU-FFusion 进行分类的方法为流量异常检测提供了新的解决方案。

6)可行性实验

为了评估本文方法的可行性,在 CICIDS2017 数据集上进行进一步验证,按 3∶7 的比例划分测试集和训练集,实验流程与 NSL-KDD 数据集一致[2]。将本文方法与最新的流量异常检测模型进行比较,实验结果如表 7 所示。

表 7 CICIDS2017 数据集上与现有模型的比较结果

Table 7 Comparison results with existing models on the CICIDS2017 dataset (%)

方法	Accuracy	Precision	Recall	F1-score
文献[23]	97.84	97.73	95.91	96.81
文献[24]	99.12	98.60	98.20	98.80
文献[25]	95.21	88.76	82.59	84.14
文献[26]	99.49	99.26	99.21	99.23
本文	99.63	99.69	99.63	99.66

从表 7 可以看出,本文方法在 CICIDS2017 数据集上的表现突出,准确率为 99.63%,精确率为 99.69%,召回率为 99.63%,F1 分数为 99.66%,验证本文方法在异常检测方面具有明显优势。其主要原因在于本文方法能够生成可信的少数类样本,DAE 的编码器模块用于特征提取和降维,有效增强了数据的表达能力;同时,使用 CNN 和 BiGRU 分别捕捉了数据的空间特征和时序特征,并将二者融合用于异常流量检测,从而显著提升了模型对异常流量的识别能力。

4 结 论

为了解决现有流量异常检测模型中存在的问题,本文提出一种基于生成对抗网络的流量异常检测方法。通过 SCiForest 隔离异常点,并采用以降噪自编码器为核心的 DGAN 方法生成可信的少数类样本,有效缓解失衡问题;采用与 DGAN 判别器相同架构的 DAE 进行特征提取与降维以增强数据特征的表达能力;通过 CNN-BiGRU-FFusion 模型在融合空间特征与时序特征的基础上完成分类检测。在数据集上的实验结果表明,本文方法能够有效识别异常流量。下一步工作将重点在更复杂场景验证所提方法的有效性。

参考文献

[1] 杨宏宇,张豪豪,胡泽,等. 基于深度学习的网络异常流量检测研究综述[J]. 武汉大学学报(理学版),2025,71(2): 159-172.
YANG H Y, ZHANG H H, HU Z, et al. A review of network anomaly traffic detection based on deep learning[J]. Journal of Wuhan University (Science Edition), 2025, 71(2): 159-172.

[2] 陈万志,任鹏江,王天元. 因素空间背景基的流量异常检测基点分类方法[J]. 电子测量与仪器学报,2024,38(6):84-94.
CHEN W ZH, REN P J, WANG T Y. Traffic anomaly detection method based on fundamental point classification by factor space background basis [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(6): 84-94.

[3] 沈萍,陈俊丽. 基于孤立森林评分扩展的流量异常检测方法[J]. 电子测量技术,2024,47(8):157-163.
SHEN P, CHEN J L. Traffic anomaly detection method based on iForest score extension [J]. Electronic Measurement Technology, 2024, 47(8): 157-163.

[4] MOHIUDDIN G, LIN Z, ZHENG J, et al. Intrusion detection using hybridized meta-heuristic techniques with weighted XGBoost classifier [J]. Expert Systems with Applications, 2023, 232: 120596.

[5] 梁欣怡,行鸿彦,侯天浩. 基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法[J]. 电子测量与仪器学报,2022,36(10):65-73.
LIANG X Y, XING H Y, HOU T H. CNN-BiLSTM network intrusion detection method based on self-supervised feature enhancement [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 65-73.

[6] 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 X J, ZHAO G SH, WANG Y, et al. Improved internet of things intrusion detection model for CNN and RNN [J]. Computer Engineering and Applications, 2023, 59(14): 242-250.

[7] ALHASSAN S, ABDUL S G, MICHEAL A, et al. CFS-AE: Correlation-based feature selection and autoencoder for improved intrusion detection system performance[J]. Journal of Internet Services and Information Security, 2024, 14(1): 104-120.

[8] SAYEGH H R, DONG W, ALMADANI A M. Enhanced intrusion detection with LSTM-based model, feature selection, and SMOTE for imbalanced data[J]. Applied Sciences, 2024, 14(2): 479.

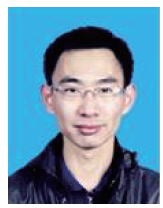
[9] 陈万志,赵林,王天元. 特征增强的改进 LightGBM 流量异常检测方法[J]. 电子测量与仪器学报,2024,38(3):195-207.
CHEN W ZH, ZHAO L, WANG T Y. Improved Light GBM for traffic anomaly detection method with feature enhancement[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(3): 195-207.

[10] OKSUZ K, CAM B C, KALKAN S, et al. Imbalance problems in object detection: A review [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3388-3415.

[11] JABBAR A, LI X, OMAR B. A survey on generative adversarial networks: Variants, applications, and training[J]. ACM Computing Surveys (CSUR), 2021,

54(8): 1-49.
[12] GOODFELLOW I, POUGET A J, MIRZA M, et al. Generative adversarial nets [J]. Advances in Neural Information Processing Systems, 2014, 27.
[13] RADFORD A. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. ArXiv preprint arXiv:1511.06434,2015.
[14] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs [J]. ArXiv preprint arXiv:1606.03498, 2016.
[15] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]. International Conference on Machine Learning, 2017: 214-223.
[16] BERTHELOT D. BEGAN: Boundary equilibrium generative adversarial networks [J]. ArXiv preprint arXiv:1703.10717, 2017.
[17] LEMAÃŽTRE G, NOGUEIRA F, ARIDAS C K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning[J]. Journal of Machine Learning Research, 2017, 18(17): 1-5.
[18] 周万珍,盛媛媛,张永强,等. 基于 ADASYN 和 WGAN 的混合不平衡数据处理方法[J]. 河北工业科技, 2024,41(4):291-298.
ZHOU W ZH, SHENG Y Y, ZHANG Y Q, et al. Hybrid imbalanced data processing based on ADASYN and WGAN [J]. Hebei Journal of Industrial Science and Technology, 2024, 41(4): 291-298.
[19] CUI J, ZONG L, XIE J, et al. A novel multi-module integrated intrusion detection system for high-dimensional imbalanced data [J]. Applied Intelligence, 2023, 53(1): 272-288.
[20] WANG S, XU W, LIU Y. Res-TranBiLSTM: An intelligent approach for intrusion detection in the Internet of Things[J]. Computer Networks, 2023, 235: 109982.
[21] DUAN X, FU Y, WANG K. Network traffic anomaly detection method based on multi-scale residual classifier[J]. Computer Communications, 2023, 198: 206-216.
[22] 陈虹,由雨竹,金海波,等. 改进特征选择和 CNN-BiLSTM 的网络入侵检测方法[J]. 微电子学与计算机,2025, 42(8):132-143.
CHEN H, YOU Y ZH, JIN H B, et al. Network intrusion detection method of improving feature selection and CNN-BiLSTM[J]. Microelectronics and Computer, 2025, 42(8):132-143.
[23] 李润杰,张小庆,刘昌华. 融合 SMOTE-Tomek Link 与集成模型的入侵检测方法[J]. 计算机技术与发展, 2024, 34(7):100-107.
LI R J, ZHANG X Q, LIU CH H. An intrusion detection approach incorporating SMOTE-Tomek link with integrated modeling [J]. Computer Technology and Development, 2024, 34(7): 100-107.
[24] ALSULAMI M H. Residual dense optimization-based Multi-Attention transformer to detect network intrusion against cyber attacks [J]. Applied Sciences, 2024, 14(17): 7763.
[25] BALLA A, HABAEBI M H, ELSHEIKH E A A, et al. Enhanced CNN-LSTM deep learning for SCADA IDS featuring hurst parameter self-similarity [J]. IEEE Access, 2024, 12: 6100-6116.
[26] KHAN S, KHAN M A, ALNAZZAWI N. Artificial neural network-based mechanism to detect security threats in wireless sensor networks[J]. Sensors, 2024, 24(5): 1641.

作者简介



陈万志(通信作者),2015 年于辽宁工程技术大学(中国测绘科学研究院联合培养)获得博士学位,现为辽宁工程技术大学副教授,硕士生导师,主要研究方向为人工智能与智能信息处理、网络与信息安全和工控软件与数据分析。
E-mail: chenwanzhi@lntu.edu.cn

Chen Wanzhi (Corresponding author) received his Ph. D. degree from Liaoning Technical University (China Academy of Surveying and Mapping Science Joint Cultivation) in 2015. Now he is an associate professor and master's supervisor in Liaoning Technical University. His main research interests include artificial intelligence and intelligent information processing, network and information security, and industrial control software and data analytics.

尹明悦,2022 年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为网络安全和入侵检测。
E-mail: 1308897193@qq.com

Yin Mingyue received his B. Sc. degree from Liaoning Technical University in 2022. Now he is a M. Sc. candidate at Liaoning Technical University. His main research interests include network security and intrusion detection.