山西大学学报(自然科学版)49(4):642—656,2026
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2026007

融合生成对抗网络与混合采样策略的网络攻击检测方法

朱丽波[1],陈湘[2],闵超[3]
(1.海南警察学院 图书馆,海南 海口 570100;
2.海南警察学院 边防管理系,海南 海口 570100;
3.南京大学 信息管理学院,江苏 南京 210023)

摘 要:针对现有网络安全态势评估方法在处理少样本攻击检测精度不足、鲁棒性较差等问题上的局限性,本文提出了一种融合生成对抗网络(Generative Adversarial Networks,GAN)与混合采样(Hybrid Sampling,HS)策略的网络攻击检测方法(GAN-HS)。该方法充分利用GAN在高质量样本生成方面的优势,通过生成器模拟潜在的攻击样本分布,生成伪攻击样本以有效补充和增强少样本数据,从而提升模型对真实且未见样本的检测能力。为应对数据严重不平衡问题,本文还引入了混合采样策略,将GAN生成的伪数据与真实样本有机结合,改善训练数据的分布情况,提高对罕见攻击类型的检测性能。此外,本文在训练过程中采用动态损失函数对网络权重进行自适应调节,使得模型在多类别攻击场景下能够更好地捕捉关键特征,从而提升整体鲁棒性和泛化能力。实验结果表明,与传统方法相比,所提出GAN-HS方法在NSL-KDD(Network Security Laboratory-knowledge Discovery and Data Mining)和CSE-CIC-IDS 2018(Communications Security Establishment & Canadian Institute for Cybersecurity-intrusion Detection System 2018)数据集上*F*1值较最优基线在二分类上分别提高了3.27%和2.42%,在多分类上分别提高了4.37%和4.99%。特别是在罕见攻击检测中展现出明显优势,由此实现了更全面、实时的网络安全态势评估,也为网络安全监测和预警提供了新的技术思路。

关键词:网络安全态势评估;生成对抗网络;深度学习;伪样本生成;少样本攻击检测
中图分类号:TP309 **文献标志码**:A **文章编号**:0253-2395(2026)04-0642-15

A Network Attack Detection Method Integrating Generative Adversarial Networks and Hybrid Sampling Strategies

Zhu Libo[1], Chen Xiang[2], Min Chao[3]
(1. *Libraries, Hainan Police College, Haikou* 570100*, China;*
2. *Department of Border Management, Hainan Police College, Haikou* 570100*, China;*
3. *School of Information Management, Nanjing University, Nanjing* 210023*, China*)

Abstract: Aiming at the limitations of existing network security situation assessment methods in dealing with the problems of insufficient detection accuracy and poor robustness of few sample attacks, this paper proposes a network attack detection method based on generative adversarial networks (GAN) and hybrid sampling (HS) strategy (GAN-HS). This method makes full use of the advantages of GAN in high-quality sample generation, simulates the distribution of potential attack samples through the generator, and generates fake attack samples to effectively supplement and enhance the few-shot data, so as to improve the detection ability of the model for real and unseen samples. In order to deal with the problem of serious data imbalance, this paper also introduces a mixed sampling

收稿日期:2025-10-23;**修回日期**:2026-03-04
基金项目:海南省自然科学基金(725RC797);海南省教育科学规划2025年度一般课题资助项目(QJY202512030)
作者简介:朱丽波(1987—),男,吉林长春人,博士,副教授,研究方向为国家安全与情报支持、数据管理与分析。E-mail:zhulibo@hpl.edu.cn
引文格式:朱丽波,陈湘,闵超.融合生成对抗网络与混合采样策略的网络攻击检测方法[J].山西大学学报(自然科学版),2026,49(4):642-656. DOI:10.13451/j.sxu.ns.2026007.

strategy, which organically combines the fake data generated by GAN with real samples to improve the distribution of training data and improve the detection performance of rare attack types. In addition, a dynamic loss function is used to adaptively adjust the network weights during the training process, so that the model can better capture the key features in multi-class attack scenarios, thereby improving the overall robustness and generalization ability. The proposed GAN-HS method is tested in NSL-KDD (Network security laboratory-knowledge discovery and data mining) and CSE-CIC-IDS 2018 (Communications security establishment & Canadian Institute for Cybersecurity-intrusion detection system 2018) dataset. Experimental results show that compared with traditional methods, the *F*1 score of the proposed GAN-HS method is 3.27% and 2.42% higher than the optimal baseline in binary classification, and 4.37% and 4.99% higher than the optimal baseline in multi-classification. In particular, it shows obvious advantages in rare attack detection, which realizes a more comprehensive and real-time network security situation assessment, and also provides a new technical idea for network security monitoring and early warning.

Key words: network security situation assessment; generative adversarial networks; deep learning; pseudo-sample generation; few-shot attack detection

0 引言

随着数字化与信息化的不断深入发展,网络系统已成为社会各领域的关键基础设施。然而,网络攻击事件频发,手段复杂多样,涉及拒绝服务攻击(Denial of Service,DoS)、远程到本地攻击(Remote to Local,R2L)、用户到根攻击(User to Root,U2R)等多种类型,这些攻击行为会对系统资源和用户数据造成严重威胁[1-2]。网络安全态势评估(Network Security Situation awareness,NSSA)是一种综合性分析方法,旨在通过对网络流量、威胁特征和攻击行为进行检测和分析,实现对当前网络安全状况的实时掌握和趋势预测[3]。NSSA不仅可以识别潜在风险,还能为管理者提供干预措施的决策依据,从而在攻击发生前采取应对策略[4-5]。网络攻击检测通过分析网络流量、系统日志或用户行为等数据,识别异常模式以判断是否存在恶意攻击行为的技术手段。然而,现有的网络攻击检测方法在面对日益庞大的网络流量、快速变化的威胁模式和复杂的多类别攻击时,存在实时响应能力不足、检测率不均衡等问题。此外,随着深度学习技术的引入,尽管分类和检测精度有所提升,但对少见和新型攻击类型的检测能力依旧是研究难点和热点问题[6]。

在NSSA领域,许多学者引入不同的数据分析与机器学习方法来提升检测效果[7]。在网络安全态势评估领域,早期工作主要采用传统的机器学习方法[8]以及深度学习模型[9](如反向传播神经网络和长短期记忆网络(Long short-term Memory,LSTM))来实现异常检测,但这些方法在处理数据不平衡和少样本攻击时存在明显不足。早期方法如灰色关联分析[10]和模糊逻辑推理[11]在多维特征关联建模与威胁量化方面具有一定优势,但在新型攻击识别能力与灵活性上仍显不足[12]。支持向量机结合优化算法[13]虽可增强多类别攻击预测效果,但对少样本攻击类型仍检测不足。近年来,生成对抗网络(Generative Adversarial Network,GAN)因其在生成高质量伪样本方面的优势而逐渐受到关注,Salem等[14]和Hara等[15]的研究分别展示了GAN及其变种在二分类异常检测和半监督学习中的应用效果。然而在多类别和少样本攻击检测任务中,其泛化能力和分类精度仍有提升空间[16]。基于此,最新研究工作尝试通过结合GAN生成器与判别器的对抗博弈机制、混合采样策略以及动态权重交叉熵损失函数,有效地增强了少样本攻击的特征表示和整体分类性能,为实现更加全面和实时的网络安全态势评估提供了新的理论和方法支持[17-20]。自Goodfellow等提出GAN基本框架以来[21],相关工作不断将其引入网络异常检测与态势评估任务中,如利用Wasserstein距离和梯度惩罚提升训练稳定性[22]。同时,GAN还与递归神经网络、变分模型及自适应融合策略相结合,以提升复杂网络环境下的检测精度和系统安全性[23-26]。

尽管已有研究在网络安全态势评估领域取得了很多进展,但其在少样本攻击类型检测和泛化能力方面仍面临挑战。一方面,真实网络环境中的数据分布极不均衡,大量正常流量和

主流攻击类型数据掩盖了少样本攻击的特征,导致传统模型难以学习到少见攻击行为的关键特征。另一方面,新型攻击模式的多样性和突发性使现有模型在应对未知威胁时缺乏灵活性,表现出较低的准确率。此外,现有方法在生成用于数据增强的样本时,生成质量和多样性不足,无法充分模拟实际攻击场景。因此,提升模型在少样本攻击检测中的表现,并提高其对未知攻击模式的适应能力,是网络安全态势评估研究中亟待解决的问题。

为解决现有网络安全态势评估方法在少样本攻击检测和泛化性能方面的不足,本研究提出了一种基于生成对抗网络的网络安全态势评估方法。该方法旨在通过生成器构建高质量的伪样本集,提升训练数据的多样性,并结合判别器提高模型对真实攻击样本的判别能力,以增强模型在不平衡数据集上的表现。同时,本研究引入动态权重分配机制,使模型在多类别网络攻击场景中能够灵活调整关注重点。本文开展的研究工作主要包括以下三个方面:首先,构建了基于GAN的网络安全态势评估框架,其中生成器用于生成高质量的伪攻击样本,判别器用于提升模型对真实和伪样本的区分能力;其次,引入混合采样策略,将GAN生成的样本与原始数据和合成过采样技术(Synthetic Minority Oversampling Technique,SMOTE)算法生成的数据结合,缓解数据不平衡问题;再次,设计了动态损失函数,根据不同攻击类型的权重自适应调整损失,以提升多类别分类任务的检测性能;最后,在NSL-KDD(Network Security Laboratory-knowledge Discovery and Data Mining)和CSE-CIC-IDS2018(Communications Security Establishment & Canadian Institute for Cybersecurity-intrusion Detection System 2018)数据集上进行了实验验证,结果表明,所提方法在少样本攻击检测和整体态势评估精度方面优于传统方法,尤其在罕见攻击类型检测中表现出更高的准确率和鲁棒性。

本文的研究结果具有重要的实际意义和理论价值,它不仅显著提高了少样本攻击检测的准确率和鲁棒性,还为多类别网络攻击情景下的安全态势评估提供了一种新的技术思路。通过引入生成对抗网络生成高质量伪样本、混合采样策略及动态损失函数,该方法有效解决了数据不平衡问题,并提升了对未知攻击的检测能力。从而有助于实时、全面地掌握网络安全状况,为未来构建更加智能和自适应的网络安全防护系统奠定了坚实基础。

1 相关工作

1.1 网络安全态势评估及网络攻击检测研究

网络安全态势评估的相关工作主要包括利用传统机器学习方法对网络攻击进行检测和分类的研究,以及近年来引入深度学习技术以提升检测精度与实时响应能力的方法。传统机器学习方面,Zhao等[10]基于灰色关联分析构建了多维系统安全态势评估模型,强调了多特征维度之间的相关性。为增强态势评估的灵活性,Alali等[11]采用模糊逻辑推理系统,通过推理规则实现威胁程度的量化,适用于动态调整网络安全状态,但该方法在面对新型或未知攻击类型时响应较慢,难以准确识别新特征模式。随着深度学习技术的快速发展,研究者们将神经网络引入网络态势评估任务。Dong等[9]提出了一种改进的反向传播神经网络,通过引入动量因子加速网络训练,提高了模型的收敛速度和检测精度,但在面对高维和非平衡数据时,模型易产生过拟合问题。针对网络安全态势评估任务,Zhang等[12]开发了一种交叉双通道架构,将卷积神经网络与双向LSTM相结合。Hu等[13]将支持向量机与优化的布谷鸟搜索算法相结合,提出了一种针对多类别网络攻击的预测方法,并在NSL-KDD数据集上实现了较高的准确率,但对于少样本攻击类型,如U2R和R2L检测率依然偏低。因此,针对多类别网络安全态势评估任务,尤其是少样本攻击检测问题,现有研究在泛化能力、检测精度和对新型威胁的适应性方面仍存在进一步优化空间。Tan[17]提出了一种专为6G环境设计的深度学习驱动的网络安全态势评估方法,构建深度自编码器模型识别网络内的各种攻击类型。Chen等[18]提出利用多头关注机制改进门控周期单元,利用多头关注机制获取不同位置的网络安全特征信息,从而提高网络态势预测的学习特

性,实现网络安全态势预测。Kalutharage等[19]利用人工智能识别基于流量特征的网络异常,并解释了每个检测到的异常背后最重要的影响因素。该模型集成存储在知识图中的领域知识,以验证检测到的异常是否构成合法攻击。Kostas等[20]提出了一种建模物联网网络攻击的方法,采用了改进的滚动窗特征提取方法。为了减少过拟合,他们应用了一个多步骤的遗传算法特征选择过程。

1.2 生成式对抗网络研究

同时,一些研究开始探索基于GAN的数据增强与异常检测技术,以解决数据不平衡问题和提高对少样本及未知攻击的识别效果,从而为构建更加全面和自适应的网络安全防护系统提供理论与实践支持。GAN由Goodfellow等[21]首次提出,通过生成器与判别器之间的博弈对抗实现高质量数据生成,已广泛应用于图像生成、语音合成和异常检测等领域。在网络安全领域,GAN展现出强大的潜力。Salem等[14]提出了一种基于GAN的异常流量检测方法,通过生成器生成伪攻击样本、判别器区分正常流量和异常流量,有效提高了模型的鲁棒性和对异常行为的检测能力,但其应用场景主要集中于二分类任务。为应对更复杂的分类任务,Hara等[15]设计了对抗自动编码器,结合GAN生成高质量样本,用于半监督学习异常检测,尽管在检测未知数据方面表现良好,但在多类别分类和少样本任务中仍存在局限性。因此,将GAN技术扩展到多类别分类和少样本数据检测,成为提升网络安全态势评估精度和鲁棒性的关键方向。Zhao等[22]提出了一种利用GAN模拟形势发展过程的形势预测模型。为了解决GAN收敛困难的问题,以Wasserstein距离的收敛距离和函数作为计算GAN损失函数的理论依据,同时采用梯度惩罚法。Khekare等[23]通过无缝集成混合生成式对抗网络-递归神经网络,建模无缝集成到流量工程和接入控制方法的基础中。在可访问性控制方法的基础上,提高网络效率和安全性。Wang等[24]提出了一种自适应融合域循环变分生成对抗网络,设计了一种多度量导向的自适应数据融合策略,基于统计指标和时频指标实现了合成数据和真实数据的高质量融合。Da Silva Ruffo等[25]提出了一种基于GAN半监督系统方法,用于开发基于IP流的网络中流量异常的检测系统。每当检测到异常时,就会触发缓解模块,自动阻止可疑IP并恢复正常的网络功能。Bethu[26]提出了一种恶意攻击检测器,旨在保护物联网系统免受此类威胁,系统利用基于GAN的模型有效地从数据集中学习并检测恶意网络活动。

基于当前研究工作中传统机器学习和深度学习方法在网络安全态势评估中的不足,尤其是在少样本攻击检测和多类别分类场景下的精度和鲁棒性问题,提出了一种新型的基于GAN的研究工作。该工作通过利用生成器生成高质量的伪攻击样本,增强数据集的多样性,并结合判别器对真实与生成样本进行区分,从而有效缓解数据不平衡问题;同时,借助混合采样策略和动态损失函数优化模型训练流程,实现更精准的识别和全面的安全态势评估。

2 本文方法

2.1 问题和任务定义

在网络安全态势评估任务中,目标是从大规模网络流量数据中识别不同类型的网络攻击,并量化当前网络的安全状态。设网络流量数据集为$D=\{(x_i,y_i)\}_{i=1}^{N}$,其中$x_i\in\mathbb{R}^d$表示第$i$条网络流量样本的特征向量,$y_i\in\{0,1,\cdots,C-1\}$表示其类别标签,$C$为类别总数。在此任务中,问题的关键在于:(1)数据集中少数类别样本的数量远小于多数类别,即$|D_j|\ll|D_k|$ $(j,k\in\{0,1,\cdots,C-1\},j\neq k)$,导致模型倾向于主流类别,忽视少样本类别;(2)在测试阶段,样本x_{test}属于未知攻击类型时,模型输出预测类别概率$p(y|x_{\text{test}})$往往偏向主流类别而非真实标签。网络攻击检测任务目标是构建一个分类器$f:\mathbb{R}^d\rightarrow\{0,1,\cdots,C-1\}$,在训练数据$D_{\text{train}}$上学习特征映射,准确预测网络攻击。同时,通过生成高质量的伪样本增强少样本类别数据,从而提高整体分类性能,并在面对未知样本时提高模型预测概率$p(y=c|x_{\text{test}})$的准确性,实现多类别攻击类型的全面识别与实时

评估。

2.2 方法框架

本文提出的融合生成对抗网络与混合采样的网络安全态势评估框架,旨在提升少样本攻击检测能力和多类别分类性能。如图1所示,框架的GAN、混合采样策略和多类别分类器三个阶段环环相扣,先由GAN生成高质量伪样本扩展少样本特征空间,再通过混合采样策略实现数据分布平衡,最终结合动态优化的多类别分类器完成精确攻击识别,从而整体提升对少样本和未知攻击的检测能力。在GAN阶段中,生成器$\mathcal{G}$接收随机噪声z作为输入,生成伪样本$\hat{x}=\mathcal{G}(z;\phi)$,其目标是模拟真实少样本类别的分布$P_{\text{data}}(x|y\in\mathcal{Y}_{\text{minority}})$,以提高少样本数据的多样性。判别器$\mathcal{D}(x;\theta)$接受输入$x$(包含真实样本和伪样本),输出判别概率$\mathcal{D}(x)\in[0,1]$。判别器目标是最大化对真实样本和伪样本的区分能力,通过对抗损失函数$\min\limits_{\mathcal{G}}\max\limits_{D}\mathcal{L}(\mathcal{G},\mathcal{D})$进行博弈优化。在混合采样策略阶段中,引入混合采样机制,将生成样本$\hat{x}$、真实少样本x_{minority}、多数类样本x_{majority}合并,构建新的训练集以缓解数据不平衡问题。在多类别分类器阶段,结合生成样本$\hat{x}$和真实数据x,训练多类别分类器$f_\theta(x)$,以最大化分类精度和少样本类别的$F1$值。此框架实现了生成样本和分类任务的联合优化,增强了对少样本攻击类型的检测能力并提升了多类别网络攻击场景下的泛化性能和鲁棒性。

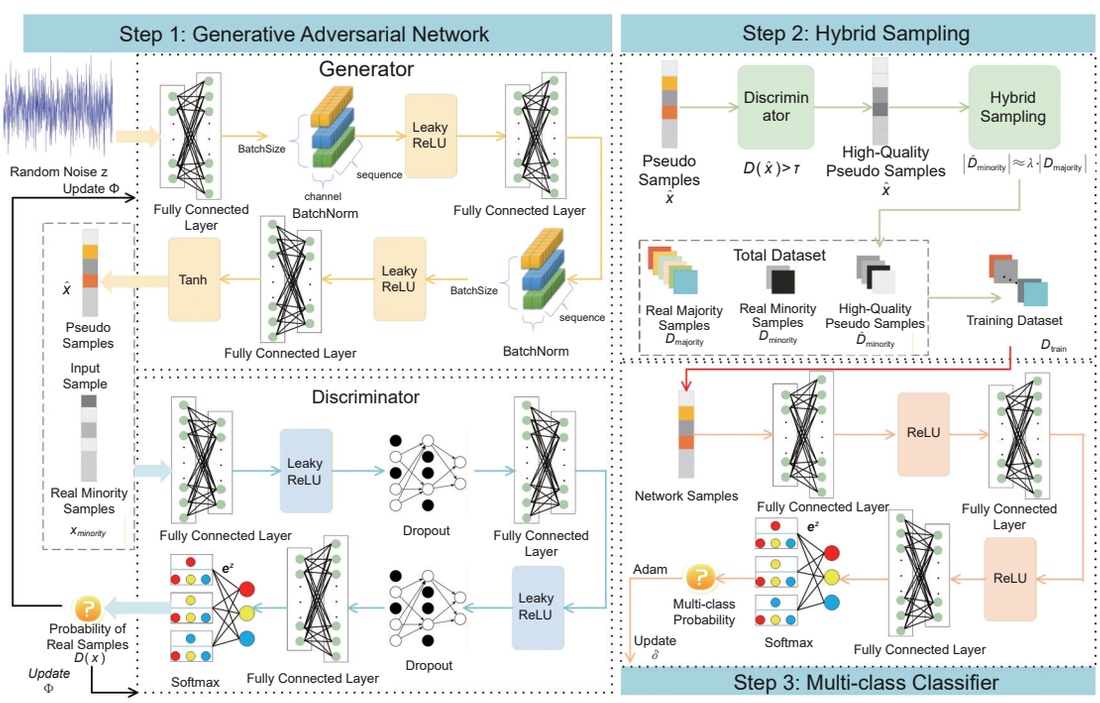


图1 融合生成对抗网络与混合采样的网络安全态势评估框架图

Fig. 1 Framework of network security situation assessment based on GAN-HS

2.2.1 生成对抗网络

本文在网络安全态势评估模型中GAN以生成高质量的伪少样本数据,增强少样本类别的特征表示能力,从而提升模型对不平衡数据集的话应性和多类别攻击的检测性能。GAN由生成器$\mathcal{G}(z;\phi)$和判别器$\mathcal{D}(x;\theta)$组成,通过生成器与判别器之间的对抗博弈,不断优化伪样本的生成质量和模型的区分能力。

其中,生成器$\mathcal{G}(z;\phi)$的输入为随机噪声向量$z\sim\mathcal{N}(0,1)$,通过一系列神经网络层映射生成伪样本$\hat{x}=\mathcal{G}(z;\phi)$,$\phi$为网络参数。其目标是生成与真实少样本攻击数据分布相似的伪样本,使判别器无法区分其真假。生成器的输出$\hat{x}\in\mathbb{R}^d$具有与真实数据相同的特征维度d。判别器$D(x;\theta)$接收输入样本x,θ为网络参数,输出概率值$\mathcal{D}(x)\in[0,1]$,表示输入样本为真实

样本的概率。判别器的目标是尽可能准确地区分真实样本$x\sim P_{\text{data}}(x)$和生成样本$\hat{x}\sim P_{\mathcal{G}}(x)$。生成器和判别器通过对抗损失函数进行优化。对抗损失函数定义为:

$$\mathcal{L}(\mathcal{G},\mathcal{D})=\mathbb{E}_{x\sim P_{\text{data}}(x)}\left[\log\mathcal{D}(x)\right]+\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[\log\left(1-\mathcal{D}\left(\mathcal{G}(z;\phi)\right)\right)\right],\tag{1}$$

其中$P_{\text{data}}(x)$是真实样本分布,$P_{\mathcal{G}}(x)$是生成样本分布,z随机噪声输入。

生成器的目标是生成高质量的伪样本,使判别器无法区分真假,其优化目标为:

$$\mathcal{L}_{\mathcal{G}}=-\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[\log\mathcal{D}\left(\mathcal{G}(z;\phi)\right)\right],\tag{2}$$

即最小化生成样本被判别为伪样本的概率,生成器通过最小化该损失生成尽可能逼真的少样本伪数据。

判别器的目标是最大化区分真实样本和生成样本的能力,其优化目标为:

$$\mathcal{L}_{\mathcal{D}}=-\mathbb{E}_{x\sim P_{\text{data}}(x)}\left[\log\mathcal{D}(x)\right]-\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[\log\left(1-\mathcal{D}\left(\mathcal{G}(z;\phi)\right)\right)\right],\tag{3}$$

即最大化真实样本被正确判别为真实样本的概率,同时最大化生成样本被判别为伪样本的概率。

生成器和判别器通过交替优化进行训练,具体步骤如下:判别器更新过程中,固定生成器的参数ϕ,根据输入真实样本和生成样本更新判别器参数$\theta\leftarrow\theta+\eta\nabla_{\theta}\mathcal{L}_{\mathcal{D}}$,其中,$\eta$为学习率,$\nabla_{\theta}\mathcal{L}_{\mathcal{D}}$表示损失函数对参数$\theta$的梯度。生成器更新过程中,固定判别器的参数$\theta$,根据输入噪声$z$更新生成器参数$\phi\leftarrow\phi-\eta\nabla_{\phi}\mathcal{L}_{\mathcal{G}}$,$\nabla_{\phi}\mathcal{L}_{\mathcal{G}}$表示损失函数$\mathcal{L}_{\mathcal{G}}$对生成器参数$\phi$的梯度,用于指导生成器更新,生成器通过最小化损失$\mathcal{L}_{\mathcal{G}}$来生成尽可能接近真实数据分布的伪样本。在生成器训练完成后,生成的伪样本$\hat{x}_{\text{minority}}=\mathcal{G}(z;\phi)$,$z\sim\mathcal{N}(0,1)$被用于增强少样本攻击类别的数据集。

2.2.2 混合采样策略

在网络攻击检测中,数据类别不平衡是一个普遍存在的问题,尤其是少样本攻击类别的数量远小于正常流量和多数类攻击数据,这会导致分类器更倾向于多数类,降低对少样本攻击的检测能力。为了解决这一问题,本文提出了一种混合采样策略,将GAN生成的伪样本、真实少样本数据和多数类数据相结合,构建平衡的数据集,从而提高模型在少样本类别上的检测精度。混合采样策略通过结合以下三类样本构建新的训练集:

$$D_{\text{train}}=D_{\text{majority}}\cup D_{\text{minority}}\cup\hat{D}_{\text{minority}},\tag{4}$$

其中,多数类样本集合D_{majority}包含真实网络流量中的多数类数据,少样本类集合D_{minority}包含真实少样本攻击类别的数据。生成样本集合$\hat{D}_{\text{minority}}$由生成器$G(z;\phi)$生成的少样本伪数据组成,$\hat{x}=\mathcal{G}(z;\phi)$,其中$z\sim\mathcal{N}(0,1)$。$\left|D_{\text{majority}}\right|\gg\left|D_{\text{minority}}\right|$,即多数类样本数量远大于少样本数量。$\left|\hat{D}_{\text{minority}}\right|$通过生成器动态生成,使得少样本类在训练集中的样本数量接近多数类数量,从而平衡数据分布。

为了提高生成样本的多样性和质量,本文引入以下生成样本选取策略:生成多个批次的伪样本$\{\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_k\}$,并通过判别器$D(x;\theta)$对生成样本进行筛选,仅保留判别概率接近真实样本的伪样本:

$$\hat{D}'_{\text{minority}}=\{\hat{x}\,|\,\mathcal{D}(\hat{x})>\tau\},\tag{5}$$

其中τ为筛选阈值,用于剔除质量较差的伪样本。少样本增强比例根据少样本类别在训练集中所占比例,动态调整生成样本数量,使得:

$$\left|\hat{D}_{\text{minority}}\right|\approx\lambda\cdot\left|D_{\text{majority}}\right|,\tag{6}$$

其中$\lambda\in(0,1]$为平衡系数,用于控制少样本类别的增强力度,避免生成样本数量过多导致过拟合。混合采样的伪代码如代码1所示。

2.2.3 多类别分类器

在本文提出的GAN-HS模型中,多类别分类器$f_{\theta}(x)$负责对输入样本x进行攻击类别预测,其输出为样本属于不同类别的概率分布。该分类器在生成的伪样本$\hat{x}$和真实样本x的基础上进行训练,从而提升模型对少样本攻击类别的检测能力和整体性能。

网络流量样本$x\in\mathbb{R}^d$,其中d是特征向量的维度。预测结果$f_{\theta}(x)\in\mathbb{R}^C$,其中$C$是类别总数。分类器输出$f_{\theta}(x)_c$表示样本$x$属于类别$c$的概率:

$$\sum_{c=1}^{C}f_{\theta}(x)_c=1,\quad f_{\theta}(x)_c\in[0,1]。\tag{7}$$

多类别分类器$f_\theta(x)$由全连接神经网络组成,通过多个全连接层和非线性激活函数,如RelU(Rectified Linear Unit)函数:

$$h^{(l)}=\mathrm{ReLU}\left(W^{(l)}h^{(l-1)}+b^{(l)}\right), \tag{8}$$

其中$W^{(l)}$和$b^{(l)}$分别为第l层的权重矩阵和偏置向量,$h^{(l)}$为第l层的输出特征。

代码1 混合采样策略伪代码

1. 输入:真实数据集D_{majority}(多数类样本),D_{minority}(少样本攻击类样本),生成的伪少样本$\hat{D}_{\text{minority}}$。
2. 输出:平衡后的训练集D_{train}。
3. **步骤1**: 生成伪样本
4. 使用生成器$\mathcal{G}$生成伪少样本$\hat{D}_{\text{minority}}$;
5. $\hat{D}_{\text{minority}}=\mathcal{G}(z)$;#$z$为随机噪声输入
6. **步骤2**: 筛选高质量伪样本 #通过判别器$\mathcal{D}$筛选出质量较高的伪样本
7. $\hat{D}_{\text{minority-prime}}=\{\ \}$;
13. ***for*** $\hat{x}$ ***in*** $\hat{D}_{\text{minority}}$:
14. *if* $\mathcal{D}(\hat{x})>\tau$: #判别器$\mathcal{D}$对伪样本的概率大于阈值$\tau$时,保留该伪样本
15. $\hat{D}_{\text{minority-prime.append}(\hat{x})}$;
16. ***end for***
16. **步骤3**: 计算生成样本的数量
18. #根据训练阶段少样本类别的比例λ动态调整生成的伪样本数量
19. $\lambda=\mathrm{adjust}_\lambda(D_{\text{minority}},D_{\text{majority}})$; # 调整生成样本数量的平衡系数
20. $N_{\text{generated}_{\text{samples}}}=[\lambda\cdot|D_{\text{majority}}|]$;
21. **步骤4**: 构建新的平衡训练集
22. #将多数类样本、少样本攻击类样本和筛选后的伪少样本合并
23. $D_{\text{train}}=D_{\text{majority}}\cup D_{\text{minority}}\cup\hat{D}_{\text{minority-prime}[:\text{num_generated_samples}]}$;
24. #输出平衡后的训练集
25. **return** D_{train}。

输出层使用Softmax激活函数输出多类别概率分布:

$$f_\theta(x)_c=\frac{\exp(o_c)}{\sum_{k=1}^{C}\exp(o_k)}, \tag{9}$$

其中o_c是输出层对应类别c的得分,C为类别总数。

为了进一步增强对少样本类别的检测性能,本文在混合采样策略的基础上设计了动态权重交叉熵损失函数。对于类别c的样本,其损失权重w_c根据类别样本数量$|D_c|$动态调整为:

$$w_c=\frac{1}{|D_c|}。 \tag{10}$$

从而分类器的损失函数变为:

$$\mathcal{L}_{cls}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}w_c\cdot y_{i,c}\log f_\delta(x_i)_c。 \tag{11}$$

这样通过增加少样本类别在损失计算中的权重,进一步提高了对少样本攻击的检测能力。

为了最小化交叉熵损失函数$\mathcal{L}_{cls}$,使用Adam优化算法更新参数δ:

$$\theta\leftarrow\theta-\eta\nabla_\delta\mathcal{L}_{cls}, \tag{12}$$

其中η为学习率,$\nabla_\delta\mathcal{L}_{cls}$为损失函数对参数$\delta$的梯度。

3 数据集和实验设置

3.1 数据集

NSL-KDD(Network Security Lab KDD Cup 99)数据集[27]是一个广泛应用于网络入侵检测领域的数据集,用于评估各种入侵检测方法和算法。该数据集是对KDD Cup 1999竞赛数据集的改进版本,旨在提高原始数据集的质量和多样性。NSL-KDD数据集包含四种类型的网络流量:正常流量(Normal)、DOS攻击、U2R攻击和R2L攻击,为网络安全研究提供了多样化的测试数据。NSL-KDD数据集各类别统计如表1所示。

表1 NSL-KDD数据集各类别统计

Table 1 Statistics of each category of NSL-KDD dataset

数据集	Normal	DoS	Probe	U2R	R2L	合计
KDD Train+	67 343 (53.46%)	45 927 (36.46%)	11 656 (9.25%)	52 (0.04%)	995 (0.79%)	125 973 (100%)
KDD Test+	9 710 (43.07%)	7 456 (33.07%)	2 421 (10.74%)	202 (0.90%)	2 754 (12.22%)	22 543 (100%)

注:表1中括号中的百分数表示正常和不同攻击类型占比。

CSE-CIC-IDS2018 on AWS数据集[28]涵盖多种不同类型的攻击场景,包括Infiltration、DoS、DDoS、Bot和FTP-BruteForce等。数据集包含1 252 835条记录,包含了捕获每台机器的

网络流量和系统日志,并使用CICFlow-V3工具从流量中提取了80个特征,提供了丰富的特征信息用于网络安全检测研究。CSE-CIC-IDS2018 on AWS数据集各类别统计如表2所示。

表2 CSE-CIC-IDS2018 on AWS数据集各类别统计

Table 2 Category statistics of CSE-CIC-IDS2018 on AWS dataset

网络类型	数量	占比
Benign	971 016	77.51%
Infiltration	38 703	3.09%
DoS attacks-Hulk	37 323	2.98%
Bot	137 185	10.95%
DDOS attack-HOIC	57 507	4.59%
DDOS attack-LOIC-UDP	8 377	0.67%
FTP-BruteForce	234	0.18%
DoS attacks-GoldenEye	332	0.03%
DoS attacks-SlowHTTPTest	103	0.01%
SSH-Bruteforce	55	0.004%
合计	1 252 835	100%

3.2 数据预处理

本文在实验前对数据进行了严格的预处理。首先,去除原始数据中的重复项与缺失值,以确保数据完整性与一致性。随后,对连续型特征采用Z-score标准化处理,使各维度特征分布保持均衡;对离散型特征进行独热编码(One-hot Encoding),以便模型能够处理非数值型输入。接着,利用信息增益(Information Gain)方法进行特征选择,筛除冗余或低相关性特征,保留对攻击识别贡献较大的关键特征。为缓解类别不平衡问题,在少样本类别上引入GAN生成伪样本,并结合真实样本与混合采样策略进行数据扩充。最后,将数据集按照70%∶15%∶15%的比例划分为训练集、验证集和测试集,为后续模型训练与性能评估提供高质量的数据基础。

3.3 实验设置

本文实验在Intel Xeon W-2175 CPU(2.50 GHz)、256 GB内存与NVIDIA A4000 GPU(16 GB×2)的环境下进行。训练过程采用Adam优化器,学习率设为1×10^{-4},批次大小为64,训练轮数为50 epoch。生成对抗网络中,随机噪声维度设为100,生成器与判别器均由三层全连接网络组成,激活函数为ReLU,输出层使用Sigmoid。混合采样平衡系数λ取0.5,伪样本筛选阈值τ设为0.7。分类器部分采用三层DNN(Deep Neural Network)结构,损失函数为带动态权重的交叉熵损失,并加入L2正则化以防止过拟合。

3.4 比较方法

(1)SVM。支持向量机(Support Vector Machines,SVM)[29]是一种监督学习模型,广泛应用于分类问题。它通过找到最佳的超平面,将数据集划分为不同的类别,以最大化类间的间隔,从而提高分类精度。SVM在处理小样本和高维数据时具有较好的效果,尤其适用于二分类任务。

(2)DT。决策树(Decision Tree,DT)[30]是一种树状结构的模型,常用于分类和回归任务。通过递归地分裂数据集,根据特征的值进行条件判断,最终形成树的叶子节点,表示类别或预测值。决策树具有直观、易解释的优点,但容易在数据不平衡的情况下产生过拟合。

(3)LSTM。LSTM[31]是一种特殊的循环神经网络,能够有效捕捉序列数据中的长期依赖关系。通过引入遗忘门、输入门和输出门,LSTM在处理时间序列数据和长距离依赖问题时表现优异。

(4)DNN。DNN[32]是由多个全连接层组成的神经网络,通过逐层学习数据的抽象特征,能够对复杂问题进行建模。DNN能够自动提取高维特征,并在多种应用中展现出强大的表现力,但训练过程可能受到过拟合和梯度消失的挑战。

(5)GAN-HS。本文提出的融合GAN和混合采样策略的网络攻击检测方法,沿用3.3节实验设置。

3.5 评价指标

在正常流量和网络攻击二分类任务中,本文使用精确率、召回率、$F1$值作为评价指标[33]。具体地,精确率(Precision,P)强调预测为正例样本中真正例的比例:

$$P=\frac{TP}{TP+FP}, \tag{13}$$

其中TP(True positive)是将正例预测为正例的

数量,FP(False positive)是将负例误预测为正例的数量。

召回率(Recall,R)强调所有正例中被正确检出的比例:

$$R=\frac{TP}{TP+FN}, \tag{14}$$

其中FN(False Negative)是将正例误预测为负例的数量。

$F1$值则为精确率和召回率的调和平均,兼顾两者的权衡:

$$F1=2\times\frac{P\times R}{P+R}, \tag{15}$$

在多类别分类任务中,本文使用准确率(Accuracy,A_{cc})和宏平均$F1$值(Macro-$F1$,$F1_{Macro}$)[34]作为评价指标。A_{cc}通常采用总体正确预测样本数与样本总数之比的形式表示:

$$A_{cc}=\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left(\hat{y}_i=y_i\right), \tag{16}$$

其中N为样本总数,y_i为第i个样本的真实标签,$\hat{y}_i$为模型预测标签,$\mathbb{I}(\cdot)$为指示函数,当预测正确时取1,否则取0。

对于$F1_{Macro}$值,每个类别c,计算该类别的精确率P_c和召回率R_c:

$$\begin{aligned}P_c&=\frac{TP_c}{TP_c+FP_c},\\ R_c&=\frac{TP_c}{TP_c+FN_c},\end{aligned} \tag{17}$$

其中TP_c、FP_c、FN_c分别表示类别c的真正例、假正例和假负例样本数。然后计算每个类别的$F1$值:

$$F1_c=\frac{2\times P_c\times R_c}{P_c+R_c}。 \tag{18}$$

最后,对所有类别的$F1$值取算术平均,得到宏平均$F1$值:

$$F1_{Macro}=\frac{1}{C}\sum_{c=1}^{C}F1_c, \tag{19}$$

其中C为类别总数。

4 实验结果及分析

4.1 二分类实验结果及分析

表3给出了在NSL-KDD和CSE-CIC-IDS2018数据集上的不同方法的二分类(正常/入侵)实验结果。从表3可以看出,本文提出的GAN-HS方法在两个基准数据集上均显著优于其他比较模型。在NSL-KDD数据集上,GAN-HS的P、R和$F1$值分别达到86.27%、87.35%和86.80%,相比传统的SVM(75.85%)和DT(79.15%)方法,$F1$值分别提升了10.95%和7.65%;相比深度学习模型LSTM(83.88%)和DNN(83.21%),也分别提升了2.92%和3.59%。这表明,GAN-HS能够更有效地识别网络攻击样本,尤其在样本不平衡和少样本类别条件下,表现出更强的特征捕获与分类能力。

表3 在2个基准数据集上的不同方法的二分类实验结果(%)

Table 3 Binary classification results of different methods on two benchmark datasets (%)

不同方法	NSL-KDD数据集			CSE-CIC-IDS2018数据集		
	P	R	$F1$	P	R	$F1$
SVM	76.34	75.38	75.85	80.17	81.77	80.96
DT	79.29	79.02	79.15	84.73	82.90	83.80
LSTM	83.57	84.20	83.88	86.50	87.82	87.15
DNN	82.55	83.88	83.21	87.02	88.05	87.53
GAN-HS	**86.27**	**87.35**	**86.80**	**90.29**	**89.61**	**89.95**

注:加粗表示最高的数值。

在CSE-CIC-IDS2018数据集上,GAN-HS的P、R和$F1$值分别为90.29%、89.61%和89.95%,整体性能优于所有对比方法。其中,$F1$值较SVM(80.96%)和DT(83.80%)分别提高8.99%和6.15%,较LSTM(87.15%)和DNN(87.53%)也提升约2.80%。这一结果表明,融合生成对抗网络与混合采样策略的GAN-HS模型不仅在检测准确率上表现突出,还有效提升了召回能力和模型鲁棒性。通过生成高质量的伪样本并动态平衡类别分布,该方法显著增强了对罕见和新型攻击类型的检测性能,实现了更全面、稳定的网络安全态势评估。

4.2 多分类实验结果及分析

表4给出了在2个基准数据集上的不同方法的多分类实验结果。从表4可以看出,本文提出的GAN-HS方法在两个基准数据集上的多分类任务中均取得了最优结果。在NSL-KDD数据集上,GAN-HS的A_{cc}达到84.48%,$F1_{Macro}$为86.98%,分别较SVM提升了16.59%和15.62%,较DT提升了11.31%和11.71%,较LSTM提升了7.41%和6.90%,较DNN提升了4.24%和4.37%。可见,GAN-HS在分类精度与宏平均指标上均有显著提升,表明该方法在

复杂多类别攻击场景下能够有效学习各类别间的特征差异,显著改善了模型对少样本攻击类别的检测能力。

表4 在2个基准数据集上的不同方法的多分类实验结果(%)

Table 4 Multi-class classification results of different methods on four benchmark datasets (%)

不同方法	NSL-KDD数据集		CSE-CIC-IDS2018数据集	
	A_{cc}	$F1_{Macro}$	A_{cc}	$F1_{Macro}$
SVM	67.89	71.36	70.27	73.09
DT	73.17	75.27	75.40	77.56
LSTM	77.07	80.08	78.38	80.28
DNN	80.24	82.61	81.10	82.11
GAN-HS	**84.48**	**86.98**	**85.75**	**87.10**

在CSE-CIC-IDS2018数据集上,GAN-HS同样表现出优异的性能,A_{cc}达到85.75%,$F1_{Macro}$为87.10%。与传统方法SVM(70.27%、73.09%)相比,GAN-HS分别提升了15.48%和14.01%;相较于深度学习模型DNN(81.10%、82.11%),A_{cc}与$F1_{Macro}$也分别提高了4.65%和4.99%。这一结果充分说明,融合生成对抗网络与混合采样策略的GAN-HS模型能够有效平衡数据分布,提升多类别攻击检测的整体性能与鲁棒性,在真实复杂网络环境中具备更强的泛化能力与稳定性。

4.3 消融实验分析

为了验证框架各模块的作用,本文设计了以下比较实验:1) Base-DNN。使用普通深度神经网络(DNN)分类器,无GAN与混合采样。作为基线模型,评估基本分类能力。2) HS-only。仅采用SMOTE-Tomek等传统混合采样方法,以验证传统采样对不平衡数据的改善作用。3) GAN-only。仅引入GAN生成少样本伪数据,不使用SMOTE,验证GAN生成样本的增强效果。消融实验结果如表5所示。

表5 消融实验结果(%)

Table 5 Results of ablation experiments (%)

不同方法	NSL-KDD数据集			CSE-CIC-IDS2018数据集		
	$F1$(二类)	A_{cc}(多类)	$F1_{Macro}$(多类)	$F1$(二类)	A_{cc}(多类)	$F1_{Macro}$(多类)
Base-DNN	80.56	78.32	80.33	82.13	77.10	79.01
HS-only	77.19	79.08	81.90	80.57	80.35	81.38
GAN-only	81.04	81.50	82.09	79.93	81.78	82.12
GAN-HS	**86.80**	**84.48**	**86.98**	**89.95**	**85.75**	**87.10**

从表5可以看出,本文提出的GAN-HS模型在消融实验中取得了最优性能,充分验证了各模块的有效性。在NSL-KDD数据集上,基础模型Base-DNN的二分类$F1$值为80.56%、多分类A_{cc}为78.32%、$F1_{Macro}$为80.33%。当仅加入混合采样策略(HS-only)后,模型在A_{cc}和$F1_{Macro}$上分别提升至79.08%和81.90%,说明混合采样机制能够在一定程度上缓解数据不平衡问题,增强模型的整体识别能力。而仅引入生成对抗网络(GAN-only)时,模型在A_{cc}和$F1_{Macro}$上进一步提升至81.50%和82.09%,表明GAN生成的高质量伪样本有效丰富了少样本类别的特征分布。最终,融合两种机制的GAN-HS模型达到最高性能,$F1$值、A_{cc}和$F1_{Macro}$分别为86.80%、84.48%和86.98%,相较Base-DNN提升幅度分别达到6.24%、6.16%和6.65%。

在CSE-CIC-IDS2018数据集上的结果与之相一致。Base-DNN的$F1$、A_{cc}和$F1_{Macro}$分别为82.13%、77.10%和79.01%,而HS-only与GAN-only分别将$F1$提升至80.57%和79.93%,表明两种机制单独使用均能改善模型性能。融合后的GAN-HS模型在三项指标上分别达到89.95%、85.75%和87.10%,较基线模型提升7.82%、8.65%和8.09%,性能优势尤为明显。实验结果验证了混合采样与GAN模块的互补性:GAN有效扩充了少样本特征空间,而混合采样机制进一步平衡了样本分布,两者结合显著增强了模型的鲁棒性与泛化能力。

4.4 少样本攻击类型检测能力分析

为了验证本文提出方法在少样本攻击类型检测能力,本文列出了DNN和GAN-HS方法在NSL-KDD和CSE-CIC-IDS2018数据集中少样

本攻击类型检测的$F1$值结果分别如图2和图3所示。从图2和图3可以看出,本文提出的GAN-HS网络攻击检测方法在少数类攻击检测任务中表现出显著优势。在NSL-KDD数据集上,模型在Probe、U2R和R2L三类少样本攻击的$F1$值分别达到80.16%、90.25%和86.70%,相较于DNN模型在同类任务上的结果有明显提升。这说明生成对抗网络在增强少样本特征分布、提高模型区分能力方面具有良好效果。同时,混合采样策略有效平衡了少数类与多数类样本比例,使模型在不同攻击类别之间的检测性能更加均衡,避免了传统模型在主流类别上过拟合而忽视罕见攻击类型的现象。在CSE-CIC-IDS2018数据集上,模型同样展现出较强的稳定性与鲁棒性。针对复杂且多样的攻击类型,如DDOS-HOIC、SSH-Bruteforce和DoS-Hulk,模型的$F1$值分别达到91.28%、89.10%和88.33%,整体检测精度保持在高水平。尤其在DDOS和暴力破解类攻击场景中,GAN-HS模型能够生成高质量伪样本,显著缓解真实数据稀缺带来的特征学习不足问题,从而提升检测率与泛化性能。综合来看,GAN-HS模型不仅在单一攻击检测上表现优异,更在多类别复杂攻击识别中展现出出色的自适应学习能力,充分验证了其在少样本网络攻击检测和动态安全态势评估中的实际应用潜力。

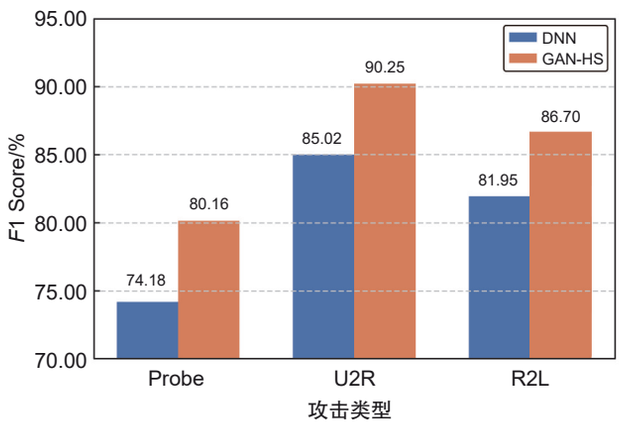


图2 DNN和GAN-HS方法在NSL-KDD数据集中少样本攻击类型检测的$F1$值结果

Fig. 2 $F1$ score results of of DNN and GAN-HS methods for few-shot attack type detection in NSL-KDD dataset

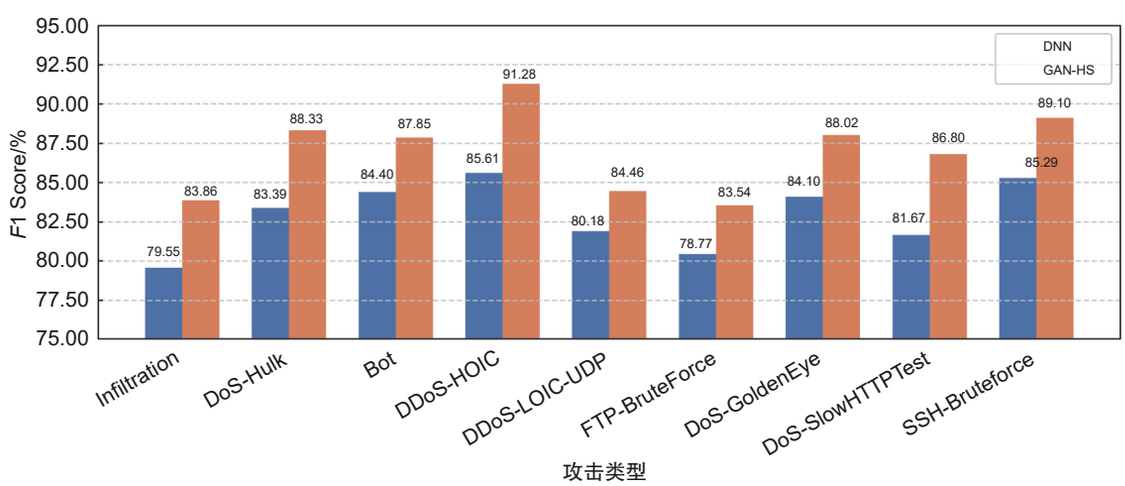


图3 DNN和GAN-HS方法在CSE-CIC-IDS2018数据集中少样本攻击类型检测的$F1$值结果

Fig. 3 $F1$ score results of of DNN and GAN-HS methods for few-shot attack type detection in the CSE-CIC-IDS2018 dataset

4.5 超参数影响分析

1) GAN学习率影响

根据图4和图5所示的实验结果,可以对GAN学习率对模型性能的影响在NSL-KDD和CSE-CIC-IDS2018两个数据集上进行如下分析。对于NSL-KDD数据集,随着GAN学习率的增大,模型性能呈现出“先提升后下降”的变化趋势。当学习率较小(如0.005×10^{-2})时,生成器和判别器的更新幅度有限,模型收敛速度较慢,$F1$(二类)和$F1_{\text{Macro}}$(多类)均处于相对较低水平。随着学习率增至0.01×10^{-2},两项指标均达到最优,其中$F1$值约为86.80%,$F1_{\text{Macro}}$值约为86.98%,表明此时GAN能够在生成样本质量与训练稳定性之间取得良好平衡,有效增强少样本

攻击特征的判别能力。然而,当学习率继续增大时,模型性能明显下降,说明过大的学习率会破坏对抗训练的稳定性,导致生成样本质量下降,从而影响整体分类性能。在CSE-CIC-IDS2018数据集上,GAN学习率对性能的影响趋势与NSL-KDD基本一致,但整体性能水平更高、波动更为明显。当学习率从较小值逐步增大时,$F1$(二类)和$F1_{Macro}$(多类)均显著提升,并在0.01×10^{-2}处达到峰值,分别为89.95%和87.10%,说明在复杂且高度不平衡的攻击场景下,合适的学习率有助于GAN生成更具代表性的少样本攻击数据,从而显著提升多类别检测能力。当学习率继续增大时,两项指标均出现持续下降,表明过高的学习率会导致对抗训练不稳定,影响模型对细粒度攻击特征的学习。综合两组实验结果可以看出,中等学习率是保证GAN训练稳定性与检测性能的关键因素,为后续模型参数设置提供了重要依据。

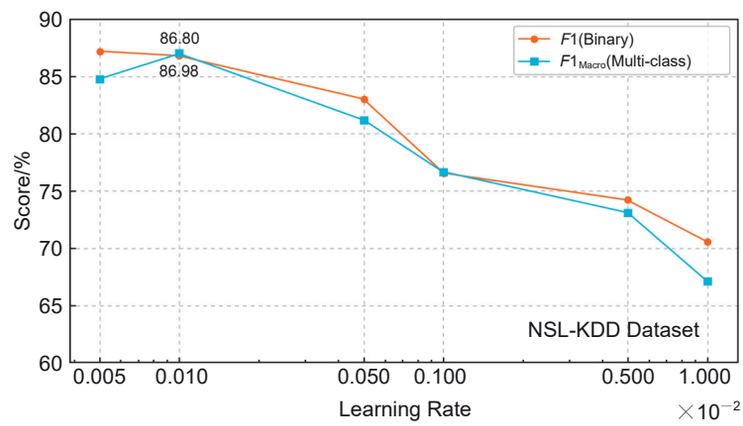


图4 NSL-KDD数据集不同GAN学习率对$F1$(二类)和$F1_{Macro}$(多类)实验结果的影响

Fig. 4 Influence of different GAN learning rates on $F1$ (two-class) and $F1_{Macro}$ (multi-class) experimental results with NSL-KDD dataset

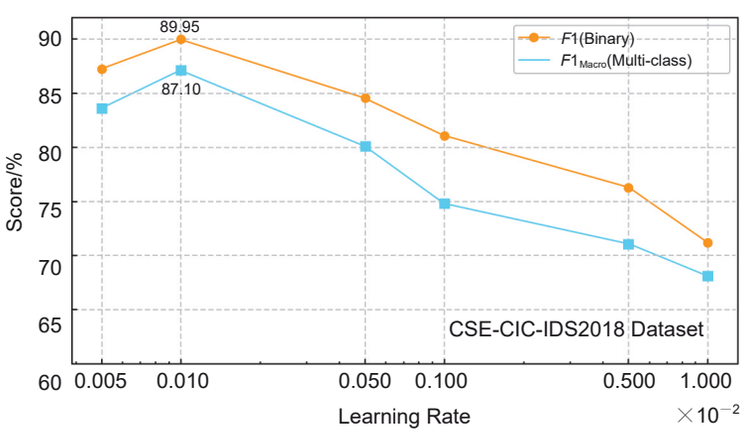


图5 CSE-CIC-IDS2018数据集不同GAN学习率对$F1$(二类)和$F1_{Macro}$(多类)实验结果的影响

Fig. 5 Influence of different GAN learning rates on $F1$ (two-class) and $F1_{Macro}$ (multi-class) experimental results with CSE-CIC-IDS2018 dataset

2)平衡系数λ影响

图6和图7分别展示了NSL-KDD和CSE-CIC-IDS2018数据集中不同平衡系数λ下的$F1$(二类)和$F1_{Macro}$(多类)实验结果影响。对于NSL-KDD数据集,随着平衡系数λ从0.1逐步增大,模型的$F1$与$F1_{Macro}$整体呈现出先显著提升、后逐渐回落的趋势。在较小λ(0.1~0.3)区间内,少样本增强力度不足,模型仍受到类别不平衡的影响,$F1$与$F1_{Macro}$均处于相对较低水平。随着λ增大至0.5,两项指标$F1$值为86.80%,$F1_{Macro}$值为86.98%,说明此时生成样本数量与真实样本比例较为合理,既有效缓解了少样本攻击的欠学习问题,又避免了过多伪样本带来的噪声干扰。当λ进一步增大时,两项指标出现不同程度下

降,表明过强的少样本增强会削弱真实数据分布特征,影响模型泛化性能。在CSE-CIC-IDS2018数据集上,平衡系数λ的影响趋势与NSL-KDD基本一致,但性能变化幅度更加明显。随着λ从0.1增大0.5,$F1$和$F1_{Macro}$均快速提升,并在$\lambda=0.5$处达到最优,分别约为89.95%和87.10%,说明在攻击类型更复杂、数据规模更大的场景下,适度增强少样本攻击数据对模型性能提升尤为关键。然而,当λ继续增大时,两项指标持续下降,反映出过量生成样本会破坏原始数据分布,使模型对真实攻击特征的判别能力下降。综合两组实验结果可以得出结论:平衡系数λ的合理取值对模型性能具有决定性影响,其中$\lambda=0.5$在两个数据集上均取得最优效果,验证了所提出混合采样策略在控制增强强度方面的有效性与稳定性。

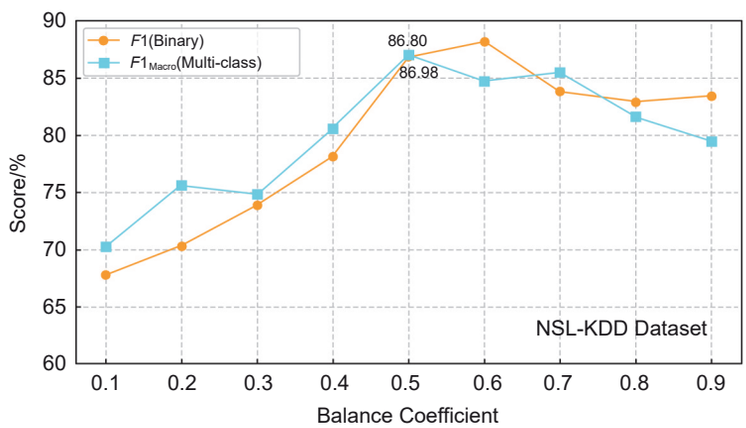


图6 NSL-KDD数据集不同平衡系数λ对$F1$(二类)和$F1_{Macro}$(多类)实验结果的影响

Fig. 6 Impact of different balance coefficient λ on $F1$ (two-class) and $F1_{Macro}$ (multi-class) experimental results with NSL-KDD dataset

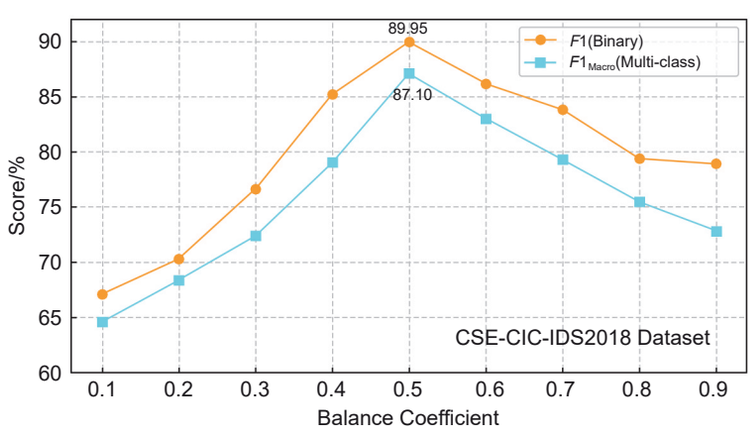


图7 CSE-CIC-IDS2018数据集不同平衡系数λ对$F1$(二类)和$F1_{Macro}$(多类)实验结果的影响

Fig. 7 Impact of different balance coefficient λ on $F1$ (two-class) and $F1_{Macro}$ (multi-class) experimental results with CSE-CIC-IDS2018 dataset

5 结论

综上所述,本文通过引入GAN与HS机制,提出了一种用于网络入侵检测的GAN-HS模型。实验结果表明,该模型在多个基准数据集上均显著优于传统机器学习与深度学习方法,在$F1$、Accuracy与$F1_{Macro}$等指标上均实现了全面提升。其中,GAN模块有效缓解了少样本类别分布不均的问题,而混合采样机制进一步增强了特征多样性与判别性,使模型在复杂网络环境下具有更强的泛化能力。这一结果与研究初衷高度一致,验证了通过生成样本与真实样本协同学习能够提升入侵检测性能的假设。与已有研究相比,本文方法在保持检测精度的同时,提升了对罕见攻击类型的识别能力,体现出更优的鲁棒性与稳定性。然而,本研究仍存在一定局限性,如模型训练时间较长、对生成样本质量较为依赖等。未来工作将考虑引入轻

量化生成结构与自适应特征筛选机制,以进一步提升模型的训练效率和在实际网络安全场景中的应用可行性。

参考文献:

[1] Chen Z H. Research on Internet Security Situation Awareness Prediction Technology Based on Improved RBF Neural Network Algorithm[J]. *J Comput Cogn Eng*, 2022, **1**(3): 103−108. DOI:10.47852/bonviewjcce149145205514.

[2] Liu L X, Xu P H, Fan K F, *et al*. Research on Application of Knowledge Graph in Industrial Control System Security Situation Awareness and Decision-making: A Survey [J]. *Neurocomputing*, 2025, **613**: 128721. DOI:10.1016/j.neucom.2024.128721.

[3] Ren Y M, Feng X J. Application Analysis of Network Security Situational Awareness Model for Asset Information Protection[J]. *Ijacsa*, 2024, **15**(5): 790. DOI: 10.14569/ijacsa.2024.0150579.

[4] 卫鸿婧, 胡治国. 面向僵尸网络DGA攻击的智能检测技术与对抗策略研究[J]. 山西大学学报(自然科学版), 2025, **48**(4): 725−740. DOI:10.13451/j.sxu.ns.2025018.
Wei H J, Hu Z G. The Research on Intelligent Detection Technology and Countermeasures for Botnet DGA Attacks[J]. *J Shanxi Univ Nat Sci Ed*, 2025, **48**(4): 725−740. DOI:10.13451/j.sxu.ns.2025018.

[5] Li W, Li J J, Zhang C T, *et al*. A Priori Algorithm Based Network Security Situational Awareness Multi-source Data Correlation Analysis Method[J]. *J Cyber Secur Mobil*, 2023, **12**(6): 869−892. DOI: 10.13052/jcsm2245-1439.1263.

[6] 席荣荣, 云晓春, 张永铮, 等. 一种改进的网络安全态势量化评估方法[J]. 计算机学报, 2015, **38**(4): 749−758. DOI: 10.3724/SP.J.1016.2015.00749.
Xi R R, Yun X C, Zhang Y Z, *et al*. An Improved Quantitative Evaluation Method for Network Security[J]. *Chin J Comput*, 2015, **38**(4): 749−758. DOI: 10.3724/SP.J.1016.2015.00749.

[7] Wang J K, Zhang D, Gao H J. Structure and Key Technologies of Nuclear Power Plant Network Security Situational Awareness Platform[C]//Proceedings of the 8th International Conference on Cyber Security and Information Engineering. New York: ACM, 2023: 15−22.. DOI: 10.1145/3617184.3617187.

[8] 张红斌, 尹彦, 赵冬梅, 等. 基于威胁情报的网络安全态势感知模型[J]. 通信学报, 2021, **42**(6): 182−194. DOI: 10.11959/j.issn.1000-436x.2021106.
Zhang H B, Yin Y, Zhao D M, *et al*. Network Security Situational Awareness Model Based on Threat Intelligence [J]. *J Commun*, 2021, **42**(6): 182−194. DOI: 10.11959/j.issn.1000-436x.2021106.

[9] Dong G S, Li W C, Wang S W, *et al*. The Assessment Method of Network Security Situation Based on Improved BP Neural Network[C]//The 8th International Conference on Computer Engineering and Networks (CENet2018). Cham: Springer International Publishing, 2019: 67−76. DOI:10.1007/978-3-030-14680-1_9.

[10] Zhao X L, Xu H, Wang T, *et al*. Research on Multidimensional System Security Assessment Based on AHP and Gray Correlation[M]//Trusted Computing and Information Security. Singapore: Springer Singapore, 2020: 177−192. DOI:10.1007/978-981-15-3418-8_13.

[11] Alali M, Almogren A, Hassan M M, *et al*. Improving Risk Assessment Model of Cyber Security Using Fuzzy Logic Inference System[J]. *Comput Secur*, 2018, **74**: 323−339. DOI:10.1016/j.cose.2017.09.011.

[12] Zhang S C, Fu Z Y, An D Z, *et al*. Network Security Situation Assessment Based on BKA and Cross Dual-channel[J]. *J Supercomput*, 2025, **81**(3): 461. DOI:10.1007/s11227-025-06932-5.

[13] Hu J J, Ma D Y, Liu C, *et al*. Network Security Situation Prediction Based on MR-SVM[J]. *IEEE Access*, 2019, **7**: 130937−130945. DOI:10.1109/ACCESS.2019.2939490.

[14] Salem M, Taheri S, Yuan J S. Anomaly Generation Using Generative Adversarial Networks in Host-based Intrusion Detection[C]//2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). New York: IEEE, 2019: 683−687. DOI:10.1109/UEMCON.2018.8796769.

[15] Hara K, Shiomoto K. Intrusion Detection System Using Semi-supervised Learning with Adversarial Auto-encoder [C]//NOMS 2020−2020 IEEE/IFIP Network Operations and Management Symposium. New York: IEEE, 2020: 1−8. DOI:10.1109/NOMS47738.2020.9110343.

[16] Rajkumar M, Karthika J, Abinayaa S S. Multi-view Consistent Generative Adversarial Network for Enhancing Intrusion Detection with Prevention Systems in Mobile Ad Hoc Networks Against Security Attacks[J]. *Comput Secur*, 2025, **150**: 104242. DOI: 10.1016/j.cose.2024.104242.

[17] Tan Q L. Deep Learning-driven Network Security Situation Awareness Method in 6G Environment[J]. *Internet Technol Lett*, 2025, **8**(2): e70006. DOI:10.1002/itl2.70006.

[18] Chen J, Bian H Q, Liang H. A Network Security Situation Prediction Model Enhanced by Multi Head Attention Mechanism[J]. *Informatica*, 2025, **49**(18): 143−156. DOI:10.31449/inf.v49i18.7670.

[19] Kalutharage C S, Liu X D, Chrysoulas C. Neurosymbolic Learning and Domain Knowledge-driven Explainable AI for Enhanced IoT Network Attack Detection and Response[J]. *Comput Secur*, 2025, **151**: 104318. DOI:10.1016/j.cose.2025.104318.

[20] Kostas K, Just M, Lones M A. IoTGeM: Generalizable Models for Behaviour-based IoT Attack Detection[J]. *Comput Netw*, 2025, **270**: 111550. DOI:10.1016/j.comnet.2025.111550.

[21] Goodfellow I, Pouget-abadie J, Mirza M, *et al*. Generative Adversarial Networks[J]. *Commun ACM*, 2020, **63**(11): 139−144. DOI:10.1145/3422622.

[22] Zhao Y, Gan G. Research on Network Security Situation Prediction Based on Improved GAN[C]//2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture. New York: ACM, 2021: 612−617. DOI:10.1145/3495018.3495128.

[23] Khekare G, Kumar K P, Prasanthi K N, *et al*. Optimizing Network Security and Performance through the Integration of Hybrid GAN-RNN Models in SDN-based Access Control and Traffic Engineering[J]. *Int J Adv Comput Sci Appl*, 2023, **14**(12): 596−606. DOI: 10.14569/ijacsa.2023.0141262.

[24] Wang X, Jiang H K, Zeng T, *Et al*. An Adaptive Fused Domain-cycling Variational Generative Adversarial Network for Machine Fault Diagnosis under Data Scarcity [J]. *Inf Fusion*, 2026, **126**: 103616. DOI: 10.1016/j.inffus.2025.103616.

[25] Da Silva Ruffo V G, Lent D M B, Carvalho L F, *et al*. Generative Adversarial Networks to Detect Intrusion and Anomaly in IP Flow-based Networks[J]. *Future Gener Comput Syst*, 2025, **163**: 107531. DOI:10.1016/j.future.2024.107531.

[26] Bethu S. Malicious Attack Detection in IoT by Generative Adversarial Networks[J]. *SN Comput Sci*, 2025, **6**(4): 372. DOI:10.1007/s42979-025-03874-w.

[27] Meena G, Choudhary R R. A Review Paper on IDS Classification Using KDD 99 and NSL KDD Dataset in WEKA[C]//2017 International Conference on Computer, Communications and Electronics (Comptelix). New York: IEEE, 2017: 553−558. DOI: 10.1109/COMPTELIX.2017.8004032.

[28] Farhan R I, Maolood A T, Hassan N F. Performance Analysis of Flow-based Attacks Detection on CSE-CIC-IDS2018 Dataset Using Deep Learning[J]. *Indones J Electr Eng Comput Sci*, 2020, **20**(3): 1413. DOI: 10.11591/ijeecs.v20.i3.pp1413-1418.

[29] Li J Y, Li Y D, Song J G, *et al*. Quantum Support Vector Machine for Classifying Noisy Data[J]. *IEEE Trans Comput*, 2024, **73**(9): 2233−2247. DOI: 10.1109/TC.2024.3416619.

[30] Mienye I D, Jere N. A Survey of Decision Trees: Concepts, Algorithms, and Applications[J]. *IEEE Access*, 2024, **12**: 86716−86727. DOI:10.1109/ACCESS.2024.3416838.

[31] Orvieto A, Zucchet N. Recurrent Neural Networks: Vanishing and Exploding Gradients Are Not the End of the Story[C]//Advances in Neural Information Processing Systems 37. New Orlean: Curran Associates, Inc., 2024: 139402−139443. DOI:10.52202/079017-4425.

[32] Li M, Micheli A, Wang Y G, *et al*. Guest Editorial: Deep Neural Networks for Graphs: Theory, Models, Algorithms, and Applications[J]. *IEEE Trans Neural Netw Learn Syst*, 2024, **35**(4): 4367−4372. DOI: 10.1109/TNNLS.2024.3371592.

[33] Diallo R, Edalo C, Awe O O. Machine Learning Evaluation of Imbalanced Health Data: a Comparative Analysis of Balanced Accuracy, MCC, and *F*1 Score[M]//Practical Statistical Learning and Data Science Methods. Cham: Springer Nature Switzerland, 2024: 283−312. DOI:10.1007/978-3-031-72215-8_12.

[34] Hinojosa Lee M C, Braet J, Springael J. Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted *F*1-scores[J]. *Appl Sci*, 2024, **14**(21): 9863. DOI:10.3390/app14219863.