文章编号:1003－6199(2025)02－0001－08 DOI:10.16339/j.cnki.jsjsyzdh.202502001

基于无监督学习的入侵流量检测分类

方君[1],王茜[2],孙雪丽[2†]

(1. 海军航空大学信息融合研究所,山东 烟台 264001;2. 海军航空大学航空基础学院,山东 烟台 264001)

摘 要:针对现有入侵流量检测模型对小样本分类准确率低的问题,提出了一种基于 Wasserstein Divergence Objective for GANs (WGAN-div)和 Information Maximizing Generative Adversarial Nets(Info GAN)的无监督学习入侵流量分类模型。首先,通过对不平衡的数据训练集进行过采样改善数据分布,然后对非数据部分进行独热编码处理并与数据部分整合,降低预处理复杂度,最后利用 Info GAN 模型进行数据训练,并在 NSL-KDD、CICIDS2017、UNSW-NB15 数据集进行性能评估和算法效能对比。实验结果表明,算法在多分类任务准确率分别达到 91.0%、97.1%、79.9%,二分类任务准确率可达 90.9%、96.9%、86.1%。相比于经典深度学习算法,Info GAN 模型的准确率更高,误报率更低,具备较高的可靠性和工程应用价值。

关键词:入侵流量检测;生成对抗网络;过采样;不平衡数据集

中图分类号: TP393.0 **文献标识码**:A

Intrusion Traffic Detection and Classification Based on Unsupervised Learning

FANG Jun[1], WANG Qian[2], SUN Xueli[2†]

(1. Institute of Information Fusion, Naval Aviation University, Yantai, Shandong 264001, China;
2. School of Aviation Basis, Naval Aviation University, Yantai, Shandong 264001, China)

Abstract: To solve the problem that the classification accuracy of model small samples is low, an unsupervised learning intrusion traffic classification model based on Wasserstein divergence objective for GANs (WGAN-div) and Information Maximizing Generative Adversarial Nets(Info GAN)is presented. Firstly, the unbalanced data training set is oversamped to improve the data distribution. Then, the non-data part is processed by independent thermal coding and integrated with the data part to reduce the complexity of pretreatment. Finally, the Info GAN model is used for data training. Performance evaluation and algorithm efficiency comparison were carried out in NSL-KDD, CICIDS2017 and UNSW-NB15 data sets. The experimental results show that the accuracy of multi-classification task is 91.0%, 97.1%, 79.9% respectively, and the accuracy of binary classification task is 90.9%, 96.9%, 86.1% respectively. Compared with the classical deep learning algorithm, the Info GAN model has higher accuracy and lower false positive rate, and has higher reliability and engineering application value.

Key words: intrusion traffic detection; generative adversarial nets; oversampling; unbalanced datasets

入侵检测是指在计算机及数据网络正常开放运行的同时,进行的一种安全监测和保障[1],其目标则是针对入侵方式进行实时检测与识别。目前,与防火墙等传统网络防御技术相比,网络入侵检测系统(NIDS)能够更好地对网络异常流量进行检测识别,从而防止网络受到可能的入侵,以确保其机

收稿日期:2024－11－10
基金项目:国家自然科学基金面上项目(62371465);山东省泰山学者专项人才工程专项(TS201511020)
作者简介:方君(1979—),男,安徽安庆人,博士,副教授,研究方向:通信信号处理。
† **通信联系人**,E-mail: snow0530@126.com

密性、完整性和可用性[2]。

早期的入侵流量检测依赖于基于规则的检测方法[3]，这些方法能够实现对规则约束内的入侵流量类型进行识别，但对于既定规则外的未知入侵流量无法做到有效识别。同时专家知识库的建立需要大量的先验知识和时间，在互联网信息流量巨大的现代条件下显然无法适用。基于机器学习的入侵检测算法能够从图像、文本等原始数据中直接进行特征表示学习，其特征层参数通过程序从数据中进行学习调整[4]，免除了人工特征处理的步骤，在处理大数据方面相比于浅模型具备更大优势。文献[5]采用一种基于信息增益和主成分分析的数据处理方法，使用支持向量机(SVM)进行入侵流量分类，但是该方法为了使训练数据平衡，没有使用全部数据集，而是进行抽取，在不平衡数据处理和数据集完整性方面存在不足。文献[6]提出一种自学习入侵检测系统(STL-IDS)对数据集进行特征学习和降维，提升了 SVM 对攻击的预测精度，但是存在识别率不高、模型泛化性不强等问题。文献[7]提出一种 BAT-MC 模型，结合了双向长短时记忆(BLSTM)和注意力机制，采用卷积层对数据集进行处理，能够自动完成网络流量层次结构的学习。文献[8]针对传统入侵检测方法受限于数据集类不平衡以及所选特征代表性不强等问题，提出一种基于 VAE-CWGAN 和特征统计重要性融合的检测方法。

从无标签的数据中提取相应的特征信息，使得训练数据的获取更加方便[9]，主要缺陷为对相同数据集的检测性能通常低于监督学习算法。生成式对抗网络(Generative Adversarial Networks，GAN)是一种无监督的深度学习模型，能够从估计的概率分布中生成与现有数据相似的新数据[10]。在入侵检测中，可通过该特性设计数据扩充技术以解决入侵数据集的不平衡问题；也可使用该特性实现模型对数据特征关系的学习，以实现对入侵流量的聚类。文献[11]提出一种 ACGAN-SVM 算法，实现了合成攻击流量数据、扩充数据集的目的，算法对四种入侵流量数据集进行了分类测试，结果表明使用扩充后的数据集算法在分类精确率、召回率指标上性能均得到了提升。文献[12]提出了一种 GAN-RF 算法，使用生成对抗网络对 CICIDS2017 数据集进行数据扩充，然后使用随机森林(RF)模型进行入侵流量检测分类，相比单独使用 RF 模型，性能均得到提升。文献[13]使用 BiGAN(Bidirectional Generative Adversarial Networks)实现了对 NSL-KDD 数据集的入侵检测，并与原始 GAN 的检测性能进行了对比，结果表明，使用 BiGAN 模型的入侵流量二分类检测性能优于原始 GAN。文献[14]对 BiGAN 的损失函数进行了改进，算法实现了对 10% 的 KDD-99 入侵流量数据集的分类识别。

以上文献分别从数据扩充和分类检测两方面使用生成对抗网络进行了方法改进，并在入侵检测性能上获得了提升。但尚未有方法对这两方面同时进行 GAN 性能改善，且部分算法的检测性能还有较大提升空间。基于此，本文提出了一种基于 WGAN-div 和 Info GAN(Information Maximizing Generative Adversarial Nets)的无监督学习算法，使用 WGAN-div 算法对 NSL-KDD 数据集训练集进行数据扩充，解决了入侵流量占比不平衡的问题，提升了深度学习算法对入侵流量特征的提取能力；使用 InfoGAN 模型对 NSL-KDD、CIC-IDS2017、UNSW-NB15 数据集分别进行入侵流量分类性能测试，实验结果表明，多分类任务准确率分别达到 91.0%、97.1%、79.9%，二分类任务准确率可达 90.9%、96.9%、86.1%，同时算法在精确率、召回率等指标上误报率低，具备较强的泛化性，有较高的工程应用价值。

1 基于 WGAN-div 与 Info GAN 的无监督检测模型

1.1 基于 WGAN-div 的数据扩充方法

原始 GAN 在训练中存在梯度消失、训练不稳定及模式崩溃等问题。用于量化两个概率分布之间的相似度的度量通常采用 KL 散度(Kullback-Leibler Divergence)和 JS 散度(Jensen-Shannon Divergence)，KL 散度的计算如式(1)：

$$\mathrm{KL}(p\parallel q)=\int_x p(x)\lg\frac{p(x)}{q(x)}\mathrm{d}x=E_{x\sim p}\left[\lg\frac{p(x)}{q(x)}\right] \tag{1}$$

JS 散度是基于 KL 散度的变体，解决了后者的非对称问题[15-17]，其计算如式(2)：

$$\mathrm{JS}(p\parallel q)=\frac{1}{2}D_{\mathrm{KL}}\left(p\parallel\frac{p+q}{2}\right)+\frac{1}{2}D_{\mathrm{KL}}\left(q\parallel\frac{p+q}{2}\right) \tag{2}$$

原始 GAN 正是使用 JS 散度来训练生成器减小真实分布与生成分布之间的距离，将判别器最优解代入，其目标函数可表示为式(3)：

$$\min_G \max_D V(D,G) = \mathrm{KL}\left(p_{\mathrm{data}} \parallel \frac{p_{\mathrm{data}}+p_G}{2}\right)+ \mathrm{KL}\left(p_G \parallel \frac{p_{\mathrm{data}}+p_G}{2}\right)-2\lg 2 = 2\mathrm{JS}(p_{\mathrm{data}} \parallel p_G)-2\lg 2 \tag{3}$$

由式(2)和 $D(x)$ 表达式推导可得式(4):

$$\mathrm{KL}(p_G \parallel p_{\mathrm{data}}) = E_{x\sim p_G}\left[\lg \frac{p_G(x)/(p_G(x)+p_{\mathrm{data}}(x))}{p_{\mathrm{data}}(x)/(p_G(x)+p_{\mathrm{data}}(x))}\right]= E_{x\sim p_G}[\lg(1-D(x))]- E_{x\sim p_G}[\lg(D(x))] \tag{4}$$

结合式(3)进行代换运算,可用 JS 散度形式表示 $E_{x\sim p_G}[\lg(1-D(x))]$,移项得式(5):

$$E_{x\sim p_G}[\lg(D(x))]= 2\mathrm{JS}(p_{\mathrm{data}} \parallel p_G)- \mathrm{KL}(p_{\mathrm{data}} \parallel p_G)- E_{x\sim p_{\mathrm{data}}}[\lg(D(x))]- 2\log 2 \tag{5}$$

式中仅前两项与生成器相关,即生成器目标函数表示为式(6):

$$L_G = 2\mathrm{JS}(p_{\mathrm{data}} \parallel p_G)- KL(p_{\mathrm{data}} \parallel p_G) \tag{6}$$

WGAN 为了解决模式崩溃问题,采用 Wasserstein 距离来衡量分布间的距离[18]。其计算公式如式(7):

$$W(p_{\mathrm{data}},p_G) = \inf_{\gamma\in\prod(p_{\mathrm{data}},p_G)} E_{(x-y)\sim\gamma}[\parallel x-y \parallel] \tag{7}$$

由于 $\parallel x-y \parallel$ 难以直接进行计算,因此引入利普希茨(Lipschitz)连续条件,即:设 $\exists k \geqslant 0$,定义域内任意两个 x_1,x_2 满足如式(8)的关系:

$$|f(x_1)-f(x_2)| \leqslant k|x_1-x_2| \tag{8}$$

则称 k 为 $f(x)$ 的 Lipschitz 常数。则目标函数转化为:

$$V(G,D) = \max_{D\in 1-\mathrm{Lipschitz}} \{E_{x\sim p_{\mathrm{data}}}[D(x)]- E_{x\sim p_G}[D(x)]\} \tag{9}$$

WGAN-GP 采用梯度惩罚的方式,以在样本的过渡区满足约束,其表达式如式(10):

$$V_{\mathrm{GP}}(G,D) = \max_{D\in 1-\mathrm{Lipschitz}} \{E_{x\sim p_{\mathrm{data}}}[D(x)]- E_{x\sim p_G}[D(x)]\}= -\lambda E_{x\sim p_y}[(\parallel \nabla \mathrm{D(x)} \parallel_2 -1)^2] \tag{10}$$

其中∇表示梯度算子,p_y 是从真实数据 p_{data} 和伪数据 p_G 分布点之间沿直线均匀采样得到的分布。WGAN-div 通过引入 Wasserstein 散度,证明了 L 约束的可去除性。W 散度的表达式如式(11)所示。

$$W_{k,p}[p_{\mathrm{data}},p_G]= \inf_{f\in C_c^1(\Omega)} E_{x\sim p_{\mathrm{data}}}[f(x)]- E_{x\sim p_G}[f(x)]+ kE_{x\sim p_u}[\parallel \nabla f(x) \parallel^p] \tag{11}$$

利用判别器参数化 $f \in C_c^1$,可将最大最小问题表示为式(12):

$$V_{\mathrm{div}(G,D)} = \min_G \max_D \{E_{x\sim p_G}[D(x)]- E_{x\sim p_{\mathrm{data}}}[D(x)]- kE_{x\sim p_u}[\parallel \nabla_x D(x) \parallel^p]\} \tag{12}$$

1.2　基于 Info GAN 的入侵检测方法

Info GAN 通过改进输入噪声矢量以解决该问题:将输入噪声分解为两部分,一部分为不可压缩的噪声 z,另一部分为潜向量 $\boldsymbol{c}$,用于针对数据分布的结构化语义特征。用 $\boldsymbol{c}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_L\}$ 表示潜向量的集合,并且满足式(13)的关系:

$$P\{c_1,c_2,\cdots,c_L\}= \prod_{i=1}^{L} P(c_i) \tag{13}$$

Info GAN 的目的在于训练网络在无监督的条件下发现并恢复每个样本数据中的潜向量。因此使用信息正则化方法来解决此问题,互信息熵的计算公式如式(14):

$$I(X;Y) = H(X)-H(X \mid Y) = H(Y)-H(Y \mid X) \tag{14}$$

在给定 $p_G(x)$ 的情况下,应当保持 $p_G(x \mid \boldsymbol{c})$ 具备一个较小的熵,因此采用的目标函数如式(15):

$$\min_G \max_D V_1(D,G) = V(D,G)-\lambda I(\boldsymbol{c};G(z,\boldsymbol{c})) \tag{15}$$

实践中,$I(\boldsymbol{c};G(z,\boldsymbol{c}))$ 难以直接进行最大化,因此引入辅助分布 $q(\boldsymbol{c} \mid x)$,其计算过程如式(16):

$$I(\boldsymbol{c};G(z,\boldsymbol{c})) = H(\boldsymbol{c})-H(\boldsymbol{c} \mid G(z,\boldsymbol{c})) = E_{x\sim G(z,\boldsymbol{c}),\boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)} \lg p(\boldsymbol{c}' \mid x)+H(\boldsymbol{c}) \tag{16}$$

由式(1)可构造 $\lg p(\boldsymbol{c}' \mid x)$ 的 KL 散度,取值恒大于等于 0,利用 $q(\boldsymbol{c} \mid x)$ 将其代换为式(17):

$$E_{\substack{x\sim G(z,\boldsymbol{c})\\ \boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)}} \lg p(\boldsymbol{c}' \mid x) = E_{x\sim G(z,\boldsymbol{c})}[D_{\mathrm{KL}}(p(\boldsymbol{c} \mid x) \parallel q(\boldsymbol{c} \mid x))+ E_{\boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)} \lg q(\boldsymbol{c}' \mid x)] \geqslant E_{\substack{x\sim G(z,\boldsymbol{c})\\ \boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)}} q(\boldsymbol{c}' \mid x) \tag{17}$$

因此可得结论如下式:

$$I(\boldsymbol{c};G(z,\boldsymbol{c})) \geqslant E_{\substack{x\sim G(z,\boldsymbol{c})\\ \boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)}} \lg q(\boldsymbol{c}' \mid x)+H(\boldsymbol{c}) \tag{18}$$

由于 $E_{\substack{x\sim G(z,\boldsymbol{c})\\ \boldsymbol{c}'\sim p(\boldsymbol{c}\mid x)}} \lg q(\boldsymbol{c}' \mid x)$ 取值与生成器和辅助分布 $q(\boldsymbol{c} \mid x)$ 有关,将其简写为 $L_1(G,Q)$,则目标函数转换为式(19):

$$\min_{G,Q}\max_{D} V_{\text{Info}}(D,G,Q) = V(D,G) - \lambda L_1(G,Q) \tag{19}$$

综合以上两节，提出了一种基于 WGAN-div 的数据扩充方法以解决入侵流量数据集的不平衡问题，提升深度学习模型对少数样本特征的提取能力；提出一种基于 Info GAN 的入侵流量检测模型。将两种方法结合起来，能够得到一种入侵检测性能良好的算法。

2 实验分析

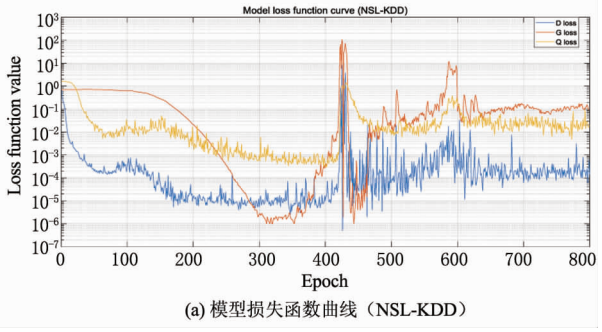
2.1 入侵流量数据集预处理

本文选用数据集为：NSL-KDD 数据集[5]、CICIDS2017 数据集[3]、UNSW-NB15 数据集[18]。

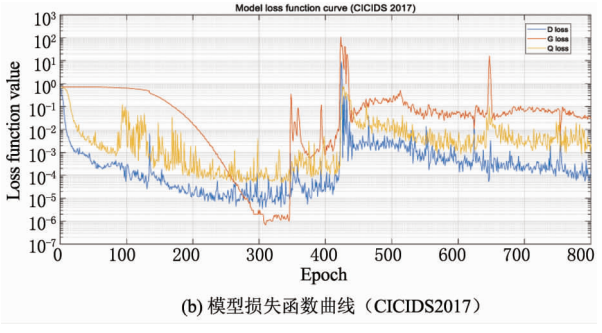
由文献可知，NSL-KDD 数据集共包括 5 种类型流量数据，其中一种是正常流量，标签为 Normal，另外四种是攻击流量，标签分别为 DoS、Probing、R2L、U2R。训练集和测试集都存在明显的数据不平衡现象，数据量最少的样本仅占比 0.04%，样本比例差距高达 1336.5∶1，需要对数据集进行平衡化的处理。CICIDS2017 数据集攻击流量包含若干小类别，统一将数据按照大类别以 6 种攻击类型流量进行标识，分别是 Botnet、Brute Force、DoS、Infiltration、PortScan、Web Attack。该数据集中各类数据也存在明显的分布不平衡问题，样本比例差距高达 83340∶1。本文设计在小样本、数据分布不均衡的条件下进行入侵检测，需要对数据集进行平衡化的处理。UNSW-NB15 数据集包含 9 种入侵流量，分别是 Fuzzers、Analysis、Backdoors、Reconnaissance、DoS、Exploit、Generic、Shellcode、Worms。观察数据分布可发现，正常流量(Normal 流量)在训练集和测试集中占比最大，Worms 流量占比最小，两者样本数量比例差异最高可达 840.91∶1，这表明该数据集也存在样本数量不平衡的问题，Analysis、Backdoors、Shellcode、Worms 等都属于小样本类别数据，需要对其进行数据扩充，以提升对小样本类别的入侵检测性能。另外，UNSW-NB15 预先划分的训练集和测试集中均存在大量的重复冗余数据，在进行预处理前需要对其进行移除。

2.2 实验结果

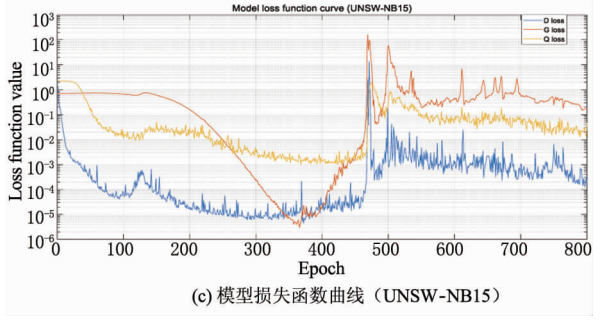
将预处理完毕的数据集划分为训练集和测试集。其中 CICIDS2017 采用 1∶9 的比例划分训练集和测试集，另外两种数据集采用公开数据集中给出的划分方法确定训练集和测试集。其中，每种数据集均包括原数据集和经 SMOTE、ADASYN、WGAN-div 过采样扩充的扩充数据集。采用的深度学习模型包括 Info GAN 以及经典机器学习算法和深度学习算法。图 1 表明 Info GAN 模型中生成器损失函数 G_loss、判别器损失函数 D_loss、分类器损失函数 Q_loss 在训练集训练过程中随训练轮数 epoch 的变化情况。随着训练轮数增加，模型的损失函数值首先逐步降低，在 epoch＝305～360 之间取值达到最小，而后随着训练轮数增长，取值在一定区间内上下波动，最终趋于稳定，这表明模型已经接近收敛。



(a) 模型损失函数曲线（NSL-KDD）



(b) 模型损失函数曲线（CICIDS2017）



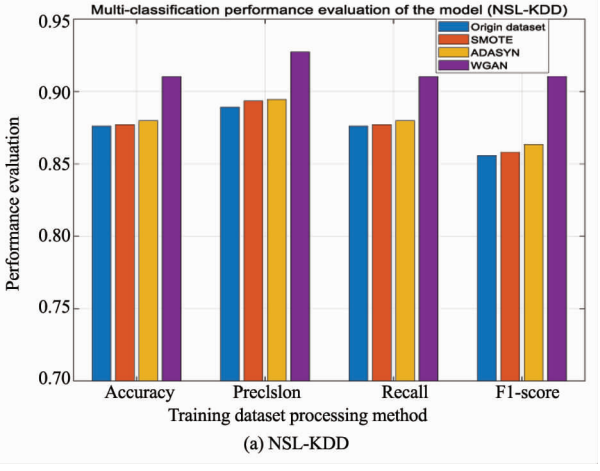
(c) 模型损失函数曲线（UNSW-NB15）

图 1 Info GAN 模型损失函数曲线

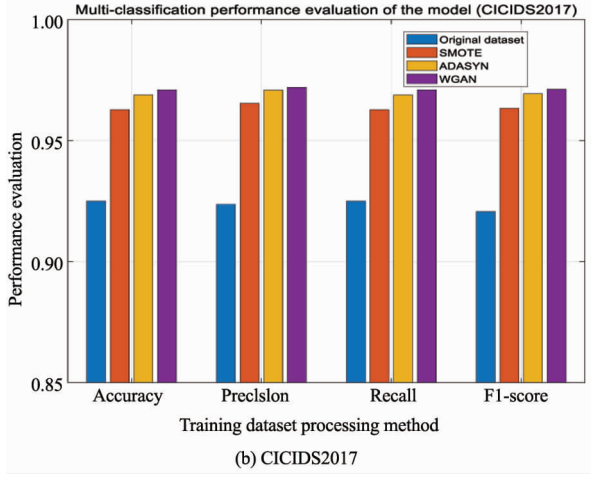
2.2.1 多分类模型性能对比

图 2 是 Info GAN 模型在各数据集的不同训练集下进行训练得到的模型多分类性能评估图。整体上看 Info GAN 模型对三种数据集都能够实现有效的多分类，分类准确率取决于数据集的数据分布。通过 t-SNE 图可得知，UNSW-NB15 数据集中训练集数据类间混叠较严重，会对分类准确率造成干扰。

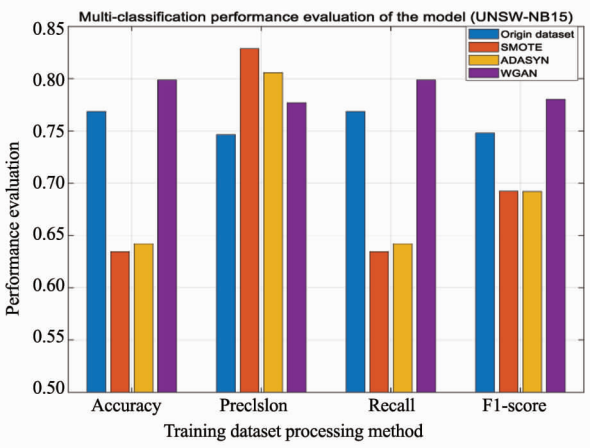
在 NSL-KDD、CICIDS2017 数据集的攻击流量分类任务中，过采样训练集均可提升模型的分类识别性能。在多分类性能评估中，Info GAN 模型在三种数据集上均展现出一定的分类能力。通过对模型损失函数曲线的分析可知，随着训练轮数增加，模型损失函数值先下降后趋于稳定，表明模型逐渐收敛。在 NSL-KDD 和 CICIDS2017 数据集的攻击流量分类任务中，过采样训练集对模型分类识别性能有提升作用，且 WGAN-div 扩充后的训练集效果最佳。这是因为 WGAN-div 生成的数据特征与原样本更契合，类内聚集性强，能有效提升对少数样本的分类识别率。例如，在 NSL-KDD 数据集中，WGAN-div 处理后的训练集使模型准确率、精确率、召回率和 F1 值均达到最高。而 UNSW-NB15 数据集训练集数据类间混叠较严重，对分类准确率产生一定干扰，但 WGAN-div 仍能在一定程度上改善分类效果。进一步分析数据发现，模型在处理类间差异较大的数据时，能够更好地学习到不同类别的特征，但对于类间混叠严重的数据，仍面临挑战，未来可考虑结合特征工程或改进模型结构来提高分类性能。



(a) NSL-KDD



(b) CICIDS2017



(c) UNSW-NB15

图 2　Info GAN 模型多分类性能评估

2.2.2　二分类模型性能对比

通过使用二分类的手段对流量进行划分，可以方便快捷地筛选入侵流量并进行处理。为了检验在二分类情况下 Info GAN 模型对各类型数据集的分类性能，将多种攻击流量标签进行合并，统一改写为“Attack”类型，便得到了用于二分类检测识别的训练集和测试集。图 3 展示的是 Info GAN 模型在各数据集的不同训练集下进行训练得到的模型的二分类性能评估图。图 3(a)是模型对 NSL-KDD 数据集的二分类测试集评估结果，整体上，二分类性能指标优于多分类结果，这主要得益于二分类简化了分类任务，避免了攻击流量间相互区分的难题。在 NSL-KDD 数据集上，WGAN-div 扩充的训练集使模型准确率提升明显，相比其他过采样算法对 Info GAN 模型分类效果提升最大。在 CICIDS2017 和 UNSW-NB15 数据集上，SMOTE 和 ADASYN 算法过采样的训练集导致 Info GAN 模型分类效果下降，而 WGAN-div 扩充的模型性能指标均有所提升，再次证明 WGAN-div 数据扩充技术的有效性和稳定性。深入对比不同数据集的二分类结果，发现模型在不同数据分布下的表现有所差异，对于数据分布相对简单的数据集(如 NSL-KDD 在二分类时)，模型能够更高效地学习到区分正常与攻击流量的特征，而对于复杂数据集(如 CICIDS2017 和 UNSW-NB15)，虽然 WGAN-div 有提升作用，但仍有进一步优化空间，可能需要更多的数据增强或模型改进策略。

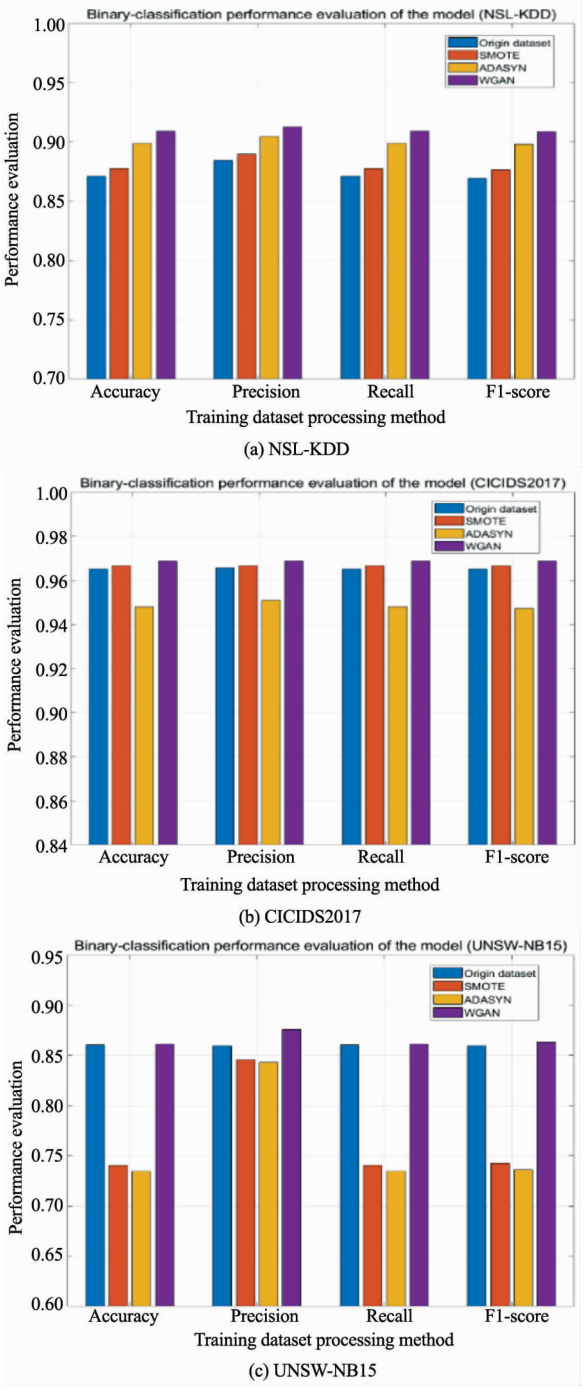


图 3 Info GAN 模型二分类性能评估

2.2.3 与现有算法性能对比

为验证 Info GAN 算法模型的实际性能和应用价值，采用文献[14]和文献[15]提出的算法、经典机器学习算法和深度学习算法与本文算法进行性能对比，分别在多分类和二分类的情况下针对准确率、精确率、召回率和 F1 值进行评估。表 1 中最后四行表示使用 Info GAN 模型，但使用了不同类型的训练集最终得到的结果，其余算法均使用数据集的原训练集进行模型训练。

表 1 的数据说明，在多分类任务中，Info GAN 分类模型性能显著优于其他算法，在三种数据集的测试集分类上均达到最高准确率和最低误报率，充分体现了算法设计的合理性和模型分类的准确性、泛化性及可靠性。例如，在 NSL-KDD 数据集上，Info GAN 模型的准确率达到 91.0%，相比其他算法有明显提升。在二分类任务中，虽然 Info GAN 算法性能略低于部分有监督算法（如 RF、CNN-LSTM 等），但作为无监督学习算法，其无需标注训练数据的优势不可忽视，在实际应用中可大幅节省数据标注时间，提高训练效率。深入分析不同算法的性能差异，发现有监督算法在有大量标注数据时能充分学习数据特征，但标注数据的获取往往成本高昂。而 Info GAN 模型能够直接从无标签数据中学习有效特征，在处理小样本和不平衡数据集时表现出色，尤其在数据标注困难的场景下具有更大的应用潜力。同时，对比使用不同训练集（原训练集、SMOTE、ADASYN、WGAN-div 扩充训练集）的 Info GAN 模型结果，进一步证实了 WGAN-div 数据扩充技术对提升模型分类性能的重要作用。

2.2.4 复杂度分析

（1）WGAN-div 计算复杂度

生成器与判别器训练：WGAN-div 在训练过程中需要同时训练生成器和判别器。在每一轮训练中，生成器生成伪数据并输入判别器，判别器计算损失函数并更新参数，然后固定判别器参数，生成器根据判别器的反馈更新自身参数。这个过程涉及复杂的神经网络计算，其计算复杂度较高。假设生成器和判别器的神经网络结构分别具有 p 和 q 个参数，在训练过程中，计算生成器损失函数和判别器损失函数的时间复杂度分别约为 $O(p)$ 和 $O(q)$。由于每一轮训练都需要多次计算损失函数并更新参数，因此整个 WGAN-div 数据扩充过程的计算复杂度与训练轮数以及生成器和判别器的参数数量密切相关，大致为 $O(t(p+q))$。

样本插值与梯度惩罚（WGAN-GP 部分）：WGAN-GP 为了满足约束条件，采用在真伪样本中随机插值进行惩罚的方式。这一操作涉及对样本的额外处理，其计算复杂度与数据集中样本数量 n 以及插值计算的复杂度有关。假设插值计算的复杂度为 $O(k)$（k 取决于具体的插值算法），则这部分操作的计算复杂度约为 $O(nk)$。在整体

WGAN-div 数据扩充过程中，虽然这部分计算复杂度相对生成器和判别器训练较低，但也不可忽视。

表 1　算法性能比较

(a)二分类算法性能比较

二分类算法	NSL-KDD				CICIDS2017				UNSW-NB15			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
RF	0.929	0.946	0.919	0.933	0.940	0.849	0.969	0.905	0.903	0.988	0.867	0.924
SVM	0.837	0.769	0.993	0.867	0.799	0.992	0.328	0.493	0.653	0.998	0.462	0.659
CNN-LSTM	0.867	0.885	0.867	0.865	0.977	0.977	0.977	0.977	0.826	0.874	0.826	0.830
SMOTE	0.878	0.890	0.878	0.877	0.967	0.967	0.967	0.967	0.741	0.846	0.741	0.742
ADASYN	0.899	0.905	0.899	0.898	0.948	0.951	0.948	0.947	0.735	0.843	0.735	0.736
Proposed	0.909	0.913	0.909	0.909	0.969	0.969	0.969	0.969	0.861	0.876	0.861	0.864

(b)多分类算法性能比较

多分类算法	NSL-KDD				CICIDS2017				UNSW-NB15			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
RF	0.753	0.814	0.753	0.715	0.944	0.970	0.944	0.953	0.755	0.755	0.695	0.724
SVM	0.702	0.689	0.702	0.656	0.799	0.757	0.799	0.723	0.702	0.689	0.702	0.656
CNN-LSTM	0.827	0.850	0.838	0.844	0.970	0.970	0.970	0.970	0.772	0.826	0.799	0.813
SMOTE	0.877	0.894	0.877	0.858	0.963	0.965	0.963	0.963	0.635	0.829	0.635	0.692
ADASYN	0.880	0.895	0.880	0.864	0.969	0.971	0.969	0.969	0.642	0.806	0.642	0.692
Proposed	0.910	0.927	0.910	0.911	0.971	0.972	0.971	0.971	0.799	0.777	0.799	0.780

(2)InfoGAN 计算复杂度

生成器与判别器对抗训练：Info GAN 模型基于原始 GAN 的对抗训练框架，其生成器和判别器的训练过程与 WGAN-div 类似，也涉及复杂的神经网络计算。同样假设生成器和判别器的参数数量分别为 p' 和 q'，训练轮数为 t'，则这部分计算复杂度约为 $O(t'(p'+q'))$。

潜向量处理与互信息计算：Info GAN 通过改进输入噪声矢量，引入了潜向量，并在目标函数中考虑了潜向量与生成数据之间的互信息熵。计算互信息熵需要对数据分布进行估计和计算，这增加了一定的计算复杂度。假设处理潜向量和计算互信息熵的复杂度为 $O(l)$（l 取决于具体的计算方法和数据分布情况），则这部分操作在每一轮训练中都需要执行，因此其计算复杂度约为 $O(t'l)$。在整个 Info GAN 模型训练过程中，这部分计算复杂度虽然相对生成器和判别器对抗训练较低，但对于模型整体性能和计算资源需求也有一定影响。

综合上述各部分的计算复杂度分析，基于 WGAN-div 和 Info GAN 的无监督学习入侵流量分类模型的总计算复杂度可以表示为 $O(t(p+q)+nk)+O(t'(p'+q')+t'l)$。

3　结　论

提出了一种基于 WGAN-div 和 Info GAN 的无监督学习入侵流量分类模型，算法通过改进输入量为噪声和潜向量的叠加量，解决了原始 GAN 可解释性差的问题。相比于有监督的深度学习模型，算法直接使用无标签数据进行训练和分类，节省了标注数据标签的工作量。通过 WGAN-div 算法对数据训练集进行特征学习，并合成少数样本数据以实现数据扩充的目的，改善小样本在不平衡数据集中的分布，解决了深度学习网络在不平衡数据集的学习中易忽略小样本特征信息的问题，通过过采样前后的对比及与 SMOTE、ADASYN 过采样效果的比较说明了 WGAN-div 算法对于分类模型性能提高的重要作用。

本文提出的算法在 NSL-KDD、CICIDS2017、UNSW-NB15 的测试集上进行了入侵流量分类性能的测试和算法效能对比，使用的性能指标包括准确率、精确率、召回率、调和平均值。结果显示，算

法在多分类任务上性能均达到最佳，准确率分别达到 91.0%，97.1%，79.9%，同时保持较高的二分类准确率和精确率，性能优于大部分深度学习算法。整体来看，本文提出的算法误报率低、泛化性好，具备较高的可靠性和工程应用价值。

参考文献

[1] MUKHERJEE B, HEBERLEIN L T, LEVITT K N. Network intrusion detection[J]. IEEE Network, 1994, 8(3): 26－41.

[2] AHMAD Z, KHAN S A, SHIANG W C, et al. Network intrusion detection system: a systematic study of machine learning and deep learning approaches[J]. Transactions on Emerging Telecommunications Technologies, 2021, 32(1): e41507.

[3] PANIGRAHI R, BORAH S. A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems [J]. International Journal of Engineering & Technology, 2018, 7(3.24): 479－482.

[4] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436－444.

[5] SALO F, NASSIF A B, ESSEX A. Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection[J]. Computer Networks, 2019, 148: 164－175.

[6] AL-QATF M, LASHENG Y, AL-HABIB M, et al. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection[J]. IEEE Access, 2018, 6: 52843－52856.

[7] SU T, SUN H, ZHU J, et al. BAT: deep learning methods on network intrusion detection using NSL-KDD dataset[J]. IEEE Access, 2020, 8:29575－29585.

[8] 刘涛涛，付钰，王坤，等. 基于 VAE-CWGAN 和特征统计重要性融合的网络流量异常检测方法[J/OL]. 通信学报，2024.

[9] LIU H, LANG B. Machine learning and deep learning methods for intrusion detection systems: a survey[J]. Applied Sciences, 2019, 9(20): 4396.

[10] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139－144.

[11] VU L, NGUYEN Q U. Handling imbalanced data in intrusion detection systems using generative adversarial networks [J]. Journal of Research and Development on Information and Communication Technology, 2020, 2020(1): 1－13.

[12] LEE J H, PARK K H. GAN-based imbalanced data intrusion detection system[J]. Personal and Ubiquitous Computing, 2021, 25(1): 121－128.

[13] ZHANG X. Network intrusion detection using generative adversarial networks[D]. New Zealand: University of Canterbury, 2020.

[14] CHEN H, JIANG L. Efficient GAN-based method for cyber-intrusion detection[J]. arXiv preprint arXiv:1904.02426, 2019.

[15] 顾伟，行鸿彦，侯天浩. 基于网络流量时空特征和自适应加权系数的异常流量检测方法[J]. 电子与信息学报，2024，46(6)：2647－2654.

[16] BAGUI S, LI K. Resampling imbalanced data for network intrusion detection datasets[J]. Journal of Big Data, 2021, 8(1): 1－41.

[17] LIU X, LI T, ZHANG R, et al. A gan and feature selection-based oversampling technique for intrusion detection[J]. Security and Communication Networks, 2021, 2021(1): 9947059.

[18] ALEESA A M, YOUNIS M, MOHAMMED A A, et al. Deep-intrusion detection system with enhanced UNSW-NB15 dataset based on deep learning techniques[J]. Journal of Engineering Science and Technology, 2021, 16(1): 711－727.