DOI:10.19651/j.cnki.emt.2314597

基于生成对抗网络和混合时空神经网络的入侵检测*

倪志伟 行鸿彦 侯天浩 梁欣怡 王心怡
(南京信息工程大学电子与信息工程学院 南京 210044)

摘 要:针对网络入侵检测领域存在检测准确率低的问题,研究异常流量样本少和分类器性能不佳时的入侵检测模型,提出一种基于改进生成对抗网络和混合时空神经网络的入侵检测模型。改进生成对抗网络通过学习异常流量样本的分布特性,生成具有特定标签的人工异常流量样本;融合卷积神经网络和双向长短时记忆神经网络提取攻击流量的时空融合特征,利用注意力机制对时空融合特征进行加权,构建混合时空神经网络对网络流量进行分类预测。在 UNSW-NB15 数据集上对所提模型进行仿真实验,准确率和 F1 分数分别为 92.93%和 94.81%,表明所提模型能够有效改善原始数据集中的类别不平衡性问题,提高对异常流量样本的检测能力和网络入侵的检测准确率。

关键词:网络入侵检测;生成对抗网络;卷积神经网络;双向长短时记忆神经网络;注意力机制
中图分类号:TP393 **文献标识码**:A **国家标准学科分类代码**:510.40

Intrusion detection based on generative adversarial networks and hybrid spatio-temporal neural networks

Ni Zhiwei Xing Hongyan Hou Tianhao Liang Xinyi Wang Xinyi
(School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Aiming at the problem of low detection accuracy in the field of network intrusion detection, we study the intrusion detection model when there are few samples of anomalous traffic and the performance of classifiers is poor, and propose an intrusion detection model based on improved generative adversarial network and hybrid spatio-temporal neural network. The improved generative adversarial network generates artificial anomalous traffic samples with specific labels by learning the distribution characteristics of the anomalous traffic samples; the fusion convolutional neural network and bidirectional long and short-term memory neural network extracts the spatio-temporal fusion features of the attacking traffic, and utilizes the attention mechanism to weight the spatio-temporal fusion features and constructs a hybrid spatio-temporal neural network to classify and predict the network traffic. Simulation experiments of the proposed model are conducted on the UNSW-NB15 dataset, and the accuracy and F1 score are 92.93% and 94.81%, respectively, indicating that the proposed model can effectively improve the problem of category imbalance in the original dataset, and improve the detection capability of abnormal traffic samples and the detection accuracy of network intrusion.

Keywords: network intrusion detection; generative adversarial networks; convolutional neural networks; bi-directional long and short-term memory neural networks; attention mechanisms

0 引 言

随着 4G、5G、物联网、云计算等技术的飞速发展,给人们的生活带来巨大的便利,但随之而来的网络攻击事件日益增多的网络安全问题,给社会安全带来了巨大威胁和挑战。为此,研究网络入侵检测机理,检测异常流量数据,识别网络流量攻击,提供实时的网络保护,具有重要的理论意义和应用价值。

由于真实网络环境中的流量数据的类别分布存在不平衡性,即流量行为中的多数类远高于流量行为中的少数类,而分类器经过数据类别不平衡的数据集训练后,通常会倾向于将少数类误判为多数类,导致入侵检测模型检测性能不佳。采用平衡化处理的数据集进行训练、构建检测能力更强的分类器,已然成为网络入侵检测领域发展中的两大

收稿日期:2023-09-15
* 基金项目:国家自然科学基金(62171228)、国家重点研发计划(2021YFE0105500)项目资助

关键需求。

网络入侵检测领域中传统的分类器模型是利用机器学习构建的，如朴素贝叶斯(NB)[1]、决策树(DT)[2]、支持向量机(SVM)[3]等，但浅层机器学习无法自动学习特征，需要手动提取特征，且在面对高维大规模网络流量数据时，其检测性能不佳，无法满足新型网络环境的需要。深度学习可以自主地学习特征，在入侵检测性能上表现更佳，其中常用的深度学习方法有：循环神经网络(recurrent neural network，RNN)[4]，卷积神经网络(convolutional neural network，CNN)[5]，以及长短时记忆神经网络(long short term memory，LSTM)[6]以及双向长短时记忆神经网络(bi-directional long short term memory，BiLSTM)[7]。尽管深度学习方法相对与机器学习方法有较大提升，但仍存在一些不足之处，如：CNN 能够提取数据的空间特征，但其学习序列数据的相关特征能力不强；LSTM 在 RNN 的基础上增加了遗忘门，可以有效解决长序列训练过程中的存在的梯度爆炸问题，但其只能读取单方向的序列数据，且误报率较高。而混合神经网络可以提高入侵检测性能，文献[8]提出了一种循环长短时记忆神经网络(RNN-LSTM)模型，通过构建 RNN-LSTM 混合模型，提高了模型的入侵检测准确率。

针对网络流量数据类别不平衡问题，传统方法是利用欠采样、过采样、以及混合采样等技术来解决类别不平衡问题，但这些方法存在过拟合和丢失有用信息的风险。文献[9]提出了一种焦点损失网络入侵检测系统(FL-NIDS)，通过在交叉熵损失函数中加入焦点损失，减少了多数类的贡献，并扩大了访问少数类的范围，以克服数据类别不平衡的问题，提高了分类器模型的入侵检测准确率，但仅通过改进交叉熵损失函数，对提高模型分类性能的提升较为有限。文献[10]使用改进的条件变分自动编码器(ICVAE)用于生成新的攻击样本，平衡了训练数据，增加了训练样本的多样性，但 ICVAE 模型存在生成数据模糊的问题。文献[11]使用生成对抗网络(GAN)对攻击样本进行生成，改善了数据集的不平衡问题。文献[12]提出了一种改进的生成对抗网络(CWGAN)，与 GAN 的不同之处在于，CWGAN 增加了外部条件信息，并在生成器中添加距离损失的 Wasserstein 损失函数，强化了模型的样本生成能力，但 CWGAN 模型存在训练困难、生成数据真实性低的缺陷。文献[13]在 CWGAN 的基础上提出了一种基于特征选择-条件 Wasserstein 生成对抗网络(FCWGAN)，该方法先对数据进行特征选择，简化数据结构，再利用 CWGAN 对数据进行学习并生成样本数据，但该方法在通过特征选择进行数据简化的同时也损失了部分关键特征，导致 CWGAN 模型无法更好地学习数据特征，进而降低了生成样本的质量。文献[14]在 CWGAN 的基础上提出了提出一种由由 WGAN 和 ACGAN 组成的 E-WACGAN 模型，该模型集成了 CWGAN、SGAN 和 infoGAN 的优点，相比于 CWGAN，其生成器具有更好的判断数据真实程度和类别的能力，同时可使随机噪声和类别信息的互信息最大化。文献[15]在 CWGAN 的基础上提出了 CVAE-WGAN 模型，该模型将 VAE 和 CWGAN 两个数据生成网络相结合，提高了其数据生成能力，但 E-WACGAN 模型和 CVAE-WGAN 模型均存在网络结构较为复杂、模型训练困难等问题。

本文将构建基于分类器的条件 Wasserstein 生成对抗网络模型(CCWGAN)，生成特定的少数类异常流量数据样本，改善网络流量数据的类别不平衡问题；将利用 CNN 提取网络流量数据的空间特征、BiLSTM 提取网络流量数据的时间特征、引入 Attention 机制对时空融合特征进行加权，构建混合时空神经网络(CNN-BiLSTM-Attention，CBA)分类模型，增强分类器模型的性能，从而提高整体的网络入侵检测的准确率和泛化能力。

1　模型的理论基础

1.1　生成对抗网络

GAN 是一种主要用于生成数据的深度学习网络模型，旨在通过训练网络模型来生成逼真的图像、音频或文本等数据。GAN 的结构如图 1 所示。GAN 模型由一对生成、判别神经网络组成，其中，判别器网络(discriminator network，DN)主要负责判断真假，生成器网络(generator network，GN)主要负责生成假数据。其中，GN 接收如高斯白噪声等随机噪声作为输入，并学习真实样本的数据分布，进而生成近似真实样本的人工样本。

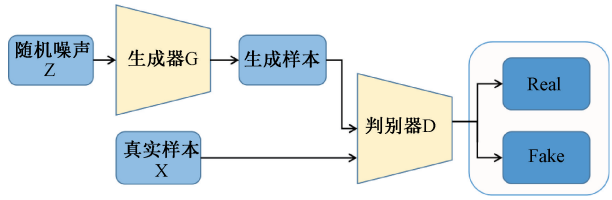


图 1　GAN 基本模型

1.2　卷积神经网络

在网络入侵检测领域，CNN 在提取流量数据的空间特征方面具有较好的表现。CNN 一般为卷积层和池化的交替多层组合形式，即深度 CNN，以提高模型的特征提取能力，因此，本文将利用 CNN 提取流量数据的空间特征。

1.3　双向长短时记忆模型

LSTM 全称为长短时记忆神经网络，相比于 RNN，LSTM 具有更强的记忆能力和控制能力。LSTM 单元包括：遗忘门、输入门、输出门，这些门限将有选择地让信息通过。

而 BiLSTM 由两个逆向正向双 LSTM 网络组成，即双向长短期记忆神经网络，其中，逆向 LSTM 网络用于学习逆向时间序列，正向 LSTM 网络用于学习正向时间序列，从而解决了单向 LSTM 特征提取的单向性问题，所以本文

将应用 BiLSTM 提取流量数据的时间特征。

1.4 注意力机制

注意力(Attention)[16]机制通过模拟人类的注意力工作模式，对不同信息的特征权重进行分配。随着深度神经网络的发展，注意力机制在不同的应用领域得到了广泛的应用，如文本分类，机器翻译，动作识别，语音识别等。Attention 机制的计算过程可以分为两个过程：1)计算权重系数，2)加权求和。其结构如图 2 所示。在过程 1 中，通过计算输入向量 Key 的值与查询向量 $\boldsymbol{F}(Q,K)$ 进行相似度计算得到注意力分数 S，再利用 $softmax$ 函数将 S 进行归一化处理，将其转换为概率分布形式，进而得到权重系数 A；在过程 2 中，将归一化的注意力权重 A 与每个 $Value$ 向量相乘，并将它们相加，以得到最终的注意力向量值。通过上述两个过程，可以获得针对 $Query$ 向量的 $Attention\ Value$，从而实现对输入序列的有针对性的关注，因而，本文将融合 Attention 机制，对信息的特征权重进行分配，提高分类模型的性能。

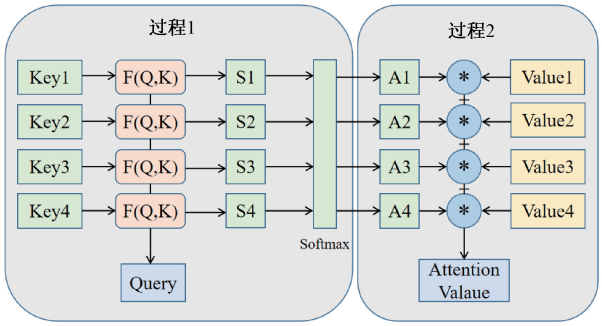


图 2 Attention 机制结构图

2 入侵检测模型的建立

首先，构建 CCWGAN 模型，生成特定的少数类异常流量数据样本，改善网络流量数据的类别不平衡问题；其次，利用 CNN-BiLSTM 提取网络流量数据的时空融合特征、利用 Attention 机制对时空融合特征进行权重分配，构建 CBA 分类模型进行网络流量分类。

2.1 CCWGAN 模型

基于条件 Wasserstein 生成对抗网络 CWGAN (conditional Wasserstein generative adversarial network)是一种改进的 GAN 模型。CWGAN 的结构如图 3 所示，CWGAN 相比于 GAN 的不同之处在于，CWGAN 在 GAN 的基础上增加了条件信息以生成特定的类别的数据，并抛弃了 GAN 使用的 JS 散度，而采用 Wasserstein 距离来评估真实样本与生成样本之间的分布，Wasserstein 距离为：

$$W(P_{data},P_g)=\underset{\gamma\in\prod(p_{data},p_g)}{INF}E_{(x,z)\sim z}[\|x-z\|] \tag{1}$$

其中，P_{data}、P_g 是真实数据分布和生成数据分布。$\prod(p_{data},p_g)$ 是所有边缘分布为 p_{data} 和 p_g 的联合概率。CWGAN 的判别器与判别器的损失函数分别如下式：

$$L_D(x,z)=E_{\tilde{g}\sim p_{g(g)}}[D(\tilde{g}\mid z)]-E_{x\sim p_{data(x)}}[D(x\mid z)]+\lambda E_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{x}\mid y)\|_2-1)^2] \tag{2}$$

$$L_G(z)=-E_{\tilde{g}\sim p_{g(g)}}[D(\tilde{g}\mid z)] \tag{3}$$

其中，D 是判别器，G 是生成器，λ 是人为设定的参数，x 是有标签的真实样本，z 是随机向量，$E(\cdot)$ 是期望值。

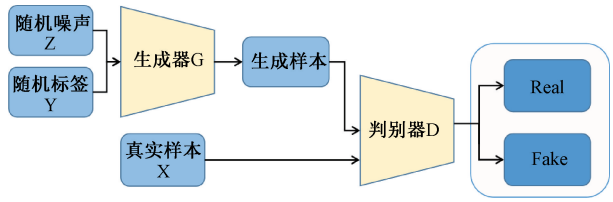


图 3 CWGAN 模型

首先，在 CWGAN 模型的基础上，增加分类器，构建基于分类器的 CWGAN 网络模型(classifier conditional wasserstein generative adversarial network，CCWGAN)，CCWGAN 模型由 3 个部分组成，分别是生成器、鉴别器以及分类器，CCWGAN 模型结构如图 4 所示，CCWGAN 的判别器、判别器的损失函数与 CWGAN 相同，而分类器的损失函数如下：

$$L_C(x,y,z)=CE(C(x),y)+\alpha CE(C(G(z)),argmax(C(G(z)))>K) \tag{4}$$

其中，C 是分类器，y 是类别信息，$E(\cdot)$ 是期望值，α 是无监督损失权重，CE 是交叉熵损失，K 是伪标签阈值。

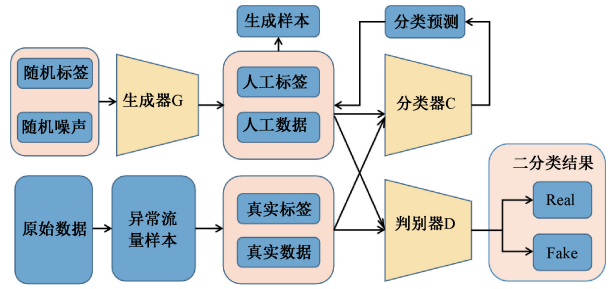


图 4 CCWGAN 基本模型

CCWGAN 模型是基于 Keras 构建的，其神经网络结构如图 5 所示。CCWGAN 模型中的生成器和判别器以及分类器使用全连接网络，主要包括 Dense 层、LeakyReLU 层和 Dropout 层。训练周期分为两个阶段，在每个训练周期的第一阶段，仅训练生成器和判别器，生成器接收随机噪声和随机标签，生成对应的人工异常流量数据和人工标签。使用真实数据和生成器生成的人工样本训练判别器，更新判别器，更好地区分真实样本和人工样本。

1)生成器网络。

生成器使用 Wasserstein 损失函数计算生成器的损失，

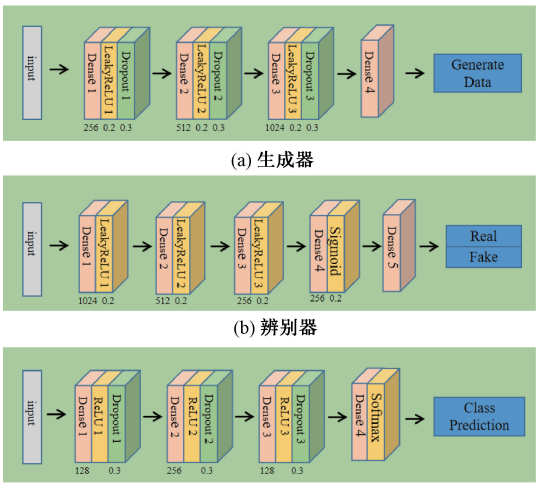


图 5　CCWGAN 神经网络结构

并使用 Adam 优化器进行训练。在训练生成器时，将判别器的权重设置为不可训练，以保持固定。生成器根据随机噪声生成与原始数据大小相同的随机样本，生成器再根据异常流量类别数生成对应的随机样本，3 个全连接层的神经元数分别为 256、512、1 024，LeakyReLU 层的参数 alpha 均设为 0.2，Dropout 层的神经元丢弃率均设为 0.5。

2)判别器网络。

将真实数据、标签和生成数据作为判别器的输入，将标签嵌入为一个 ***One-hot*** 向量，并将其扁平化。在训练判别器时，将生成器的权重设置为不可训练，以保持固定。判别器对真实数据和生成数据进行判断，并返回数据的真假标签。在真实数据和生成数据之间进行加权平均，以生成插值数据。判别器对插值数据进行判断，并返回数据的真假标签。使用 3 个损失函数分别计算真实数据、生成数据和插值数据的损失，其中前两个损失函数为 Wasserstein 损失函数，第 3 个损失函数为梯度惩罚，判别器的前 3 个全连接层的神经元数分别为 1 024、512、256，LeakyReLU 层的参数 alpha 均设为 0.2，第 4 个全链接层的大小设为数据输入的大小，判别器的输出层使用 Sigmoid 函数，并将真实数据、生成数据和插值数据的真假标签作为输出。

3)分类器网络。

在每个训练周期的第 2 阶段，训练生成器和判别器的同时，将分类器增加到训练过程中，把真实标签和生成的标签编码转化为独热编码形式，分别使用真实数据和生成器生成的人工异常流量样本来计算分类器的损失，分类器使用伪标记方法，根据分类器当前状态下最可能的类别来假设标签，仅当模型以高于特定阈值的概率预测样本的类别时，才会保留生成的生成样本和标签，进而增强模型的样本生成能力、提高生成人工样本的质量。

分类器分别构建了对真实数据和生成数据进行分类的两个模型，并使用不同的损失函数进行训练。使用交叉熵损失函数计算分类器的损失，使用 Adamax 优化器进行训练，并计算分类器的准确率作为评价指标。将真实数据和生成数据作为输入，前 3 个全连接层的神经元数分别为 128、256、512，Dropout 层的神经元丢弃率均设为 0.2，第 4 个全连接层的神经元数设为标签种类大小，分类器的输出层使用 Softmax 函数，并将分类器的预测结果作为输出。

2.2　CBA 分类模型

CBA 分类模型如图 6 所示，该模型的神经网络结构主要包括 CNN 单元、BILSTM 单元、Attention 层以及输出层，CBA 分类模型与传统的深度学习网络结构的不同之处在于：CBA 分类模型采用了 CNN 单元与 BILSTM 单元的并行式网络结构而非串行式网络结构，采用并行式网络结构优势为：能够高效地利用 CNN 单元和 BILSTM 单元分别独立提取输入数据的空间特征和时间特征，再将两个单元提取到时间特征和空间特征合并输送到 Attention 层，利用 Attention 层对其进行特权加权，以提高整体模型的分类性能。

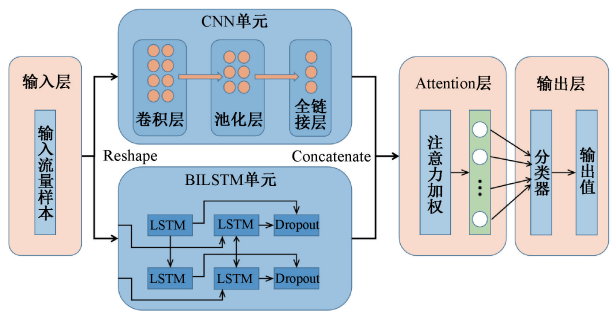


图 6　CBA 分类模型

在 CNN 单元中，卷积层使用 64 个 2×2 大小的卷积核对输入数据进行卷积操作，以提取输入数据的空间特征，激活函数为 ReLU，边缘填充方式为 same；池化层用于降低维度和减小过拟合，池化层使用了 2×2 大小的池化核进行最大池化操作，边缘填充方式为 same；BILSTM 单元中的 BiLSTM 层用于提取数据的时间特征，BILSTM 层的第一层神经元数量设为 64，第二层神经元数量设为 32，并在 BiLSTM 层后添加一个 Dropout 层，以防止模型过拟合，其神经元丢弃比例设置为 0.2；通过 Attention 层对其进行注意力加权，得到加权结果的输出。最后使用 sigmoid 函数作为激活函数进行二分类输出。

2.3　CCWGAN-CBA 入侵检测模型

CCWGAN-CBA 入侵检测模型如图 7 所示，对数据集进行独热编码和最小最大归一化处理。通过 CCWGAN 对数据预处理后的训练集进行少数类异常流量样本生成，将生成的人工异常流量样本加入到原始训练集，改善分类训练过程中的数据类别不平衡性问题，将其合并作为分类模型训练的合并训练集。利用 CNN-BiLSTM 单元提取时空融合特征，引入 Attention 机制给融合后的特征分配不同的权重，构建 CBA 分类模型，最后使用 Sigmiod 函数对数据进行二分类，区分正常流量和异常流量。

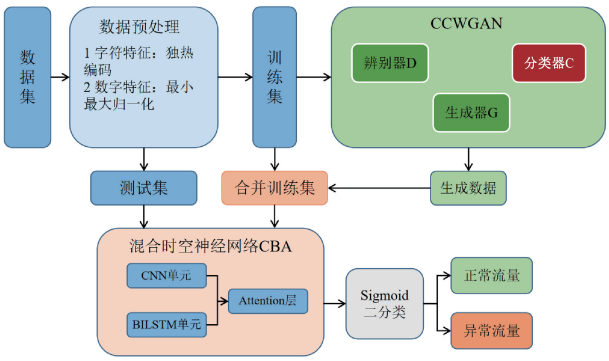


图 7 CCWGAN-CBA 入侵检测模型

3 实验与结果分析

本模型在 Windows 平台上进行了实验，以评估所提出模型的有效性。平台的详细配置如表 1 所示。实验目标数据集为 UNSW-NB15。

表 1 实验配置

实验环境	具体配置
操作系统	Windows 11
CPU	Intel(R) i5-8300H CPU@2.30 GHz
GPU	NVIDIA GeForceGTX 1050Ti
内存	16 GB
框架	TensorFlow

3.1 实验数据与处理

UNSW-NB15[17]是由新南威尔士大学于 2015 年收集并建立的网络入侵检测数据集，能够正确反映当今多样的攻击类型和复杂的网络情况，可以作为网络入侵检测领域新的基准数据集。UNSW-NB15 数据集的总共有 257 673 条数据，其中训练集为 175 341 条，测试集为 82 332 条，包含 43 个特征，一个多分类标签，一个二分类标签，其中多分类标签一共有 10 个类型，分别是正常状态“Normal”和 9 种攻击类型，UNSW-NB15 数据集每个类型的数据分布如表 2 所示。由于 CCWGAN-CBA 模型的目的是从网络流量数据中检测出攻击流量，用“0”和“1”分别表示正常流量和异常流量，进行正常/异常的二分类任务，并删除此数据集的多分类标签，仅保留二分类标签。

为了更好的训练该分类模型，选择使用 *Onehot* 编码方式将数据集中的符号特征 *proto*，*state*，*service* 等非数值型转为数值型的输入，原始特征从 43 维转化为 196 维，再对对数据进行进行 *Minmax* 归一化处理，将数值归一化到 0～1 之间。

$$x^* = \frac{x - x_{\max}}{x_{\max} - x_{\min}} \tag{5}$$

式中：$x_{\max}$ 为数据中的最大值，$x_{\min}$ 为数据中的最小值，x^* 为归一化处理后的数据。

表 2 UNSW-NB15 数据集

序号	类型	训练集数量	测试集数量
1	Normal	56 000	37 000
2	Exploits	33 393	11 132
3	Generic	40 000	18 871
4	Fuzzers	18 184	6 062
5	Reconnaissance	10 491	3 496
6	Shellcode	1 133	378
7	Dos	12 264	4 089
8	Analysis	2 000	677
9	Backdoor	1 746	583
10	Worms	130	44
	共计	175 341	82 332

3.2 评价标准

为了评估所提模型的有效性，使用准确率(Accuracy)、精度(Precision)、召回率(Recall)、F1 分数(F1-score)[18]作为实验过程中的评价指标。Accuracy 是分类模型最常用的评价指标之一，表示所有分类正确的样本占全部样本的比例；Precision 是指预测结果为正样本中实际标签也为正样本的比例，可以用来评估模型对正例的预测能力；Recall 是指所有正样本中被正确预测的比例，可以用来评估模型对正例的覆盖能力；F1-score 是 Precision 和 Recall 的调和平均值，用来衡量分类模型在 Precision 和 Recall 之间的平衡性高低，可以用来综合评估模型的分类能力。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F1-score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{9}$$

式中：TP 表示将正例预测为正例的数量，FP 表示将负例预测为正例的数量，FN 表示将正例预测为负例的数量，TN 表示将负例预测为负例的数量。

3.3 实验结果与分析

由于 UNSW-NB15 数据集存在数据类别不平衡的问题，数据集中的 Analysis、Backdoor、Shellcode、Worms 等 4 类异常类型数据较少，所以本文从训练集选取这 4 类攻击样本作为 CCWGAN 模型的训练数据，并生成对应类型的人工异常流量样本。CCWGAN 的初始参数包括：epoch 设为 300，batch_size 设为 128，分类器分类阈值设为 0.5，使用 Sgd 作为优化器，学习率设为 0.000 5。CCWGAN 模型训练的损失函数如图 8 所示，其中 G_loss、D_loss、C_loss 分别代表了生成器网络、判别器网络、分类器网络的损失函数值，从图中可知，G_loss、D_loss、C_loss 在 epoch＝290 处

左右趋于平稳，CCWGAN 模型训练到达最佳。

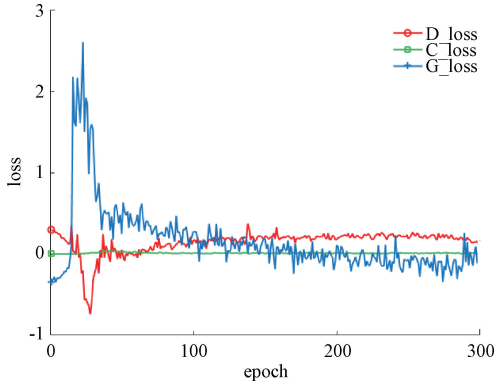


图 8　CCWGAN 模型训练的损失函数

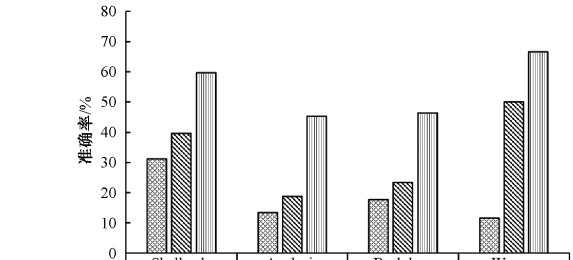
在 CCWGAN 模型训练 290 轮的基础上，对 Analysis、Backdoor、Shellcode、Worms 等 4 种少数类攻击类型进行数据生成，训练集的少数类生成前后数据分布情况如表 3 所示。训练集生成前共有 175 341 条流量样本，数据生成后有 215 341 条流量样本，改善了原始数据集的不平衡度。

表 3　CCWGAN 样本生成前后少数类的数据分布

序号	类型	样本生成前	样本生成后
1	Shellcode	1 133	11 133
2	Analysis	2 000	12 000
3	Backdoor	1 746	11 746
4	Worms	130	10 130

为了验证 CCWGAN 模型的生成样本的有效性，将原始数据集、CWGAN 模型生成后的平衡数据集、CCWGAN 模型生成后的平衡数据集等 3 个数据集中的 4 个少数类异常流量进行分类预测，其准确率结果如图 9 所示。从图中可以看出，CWGAN 模型和 CCWGAN 模型均提高了对于该 4 个少数类的准确率。以 Worms 类为例，原始数据集在分类模型下的准确率仅为 11.59%，经过 CWGAN 模型数据平衡化处理后的分类模型的分类准确率提高到了 50.12%，而 CCWGAN 模型对数据集进行数据平衡化处理后，分类模型的分类准确率进一步提高到了 66.67%，说明 CCWGAN 模型能够有效缓解数据集的类别不平衡性问题，提高分类模型对网络异常流量中的少数类的检测率。

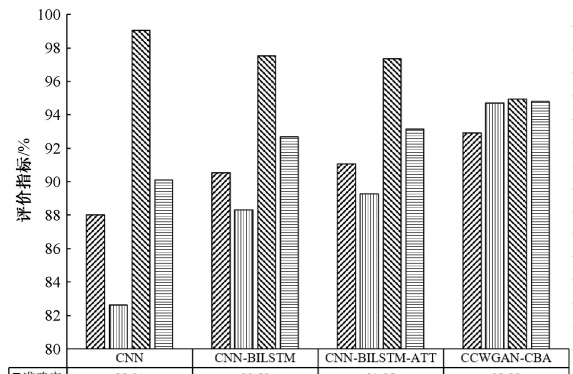
为确保实验结果的可靠性，在 UNSW-NB15 训练集上训练 CBA 分类模型，在 UNSW-NB15 测试集上对该分类模型进行测试，以评估其有效性。该分类模型参数设置如下：epoch 设为 100，batch_size 设为 32，CNN 单元和 BILSTM 单元均使用 relu 函数作为激活函数，模型的损失函数设为二元交叉熵函数，使用准确率作为模型的评价指标，优化器设为 adam，学习率设为 0.001。此外，为了验证 CCWGAN 模型和 CBA 分类模型对网络入侵检测的有效性，将各个模型在 UNSW-NB15 数据集上进行消融实验，



	Shellcode	Analysis	Backdoor	Worms
原始数据集	31.17	13.35	17.74	11.59
CWGAN模型	39.67	18.78	23.33	50.12
ECWGAN模型	59.77	45.31	46.36	66.67

图 9　生成样本的准确率实验

结果如图 10 所示。从图中可以看出 CCWGAN-CBA 模型相比与 CNN 模型、CNN-BiLSTM 模型的分类性能更优。当 CNN-BiLSTM 模型无 Attention 机制时的准确率和 F1 分数分别是 90.52%和 92.69%；当 CNN-BiLSTM 模型加入 Attention 机制后的准确率和 F1 分数提高到了 91.05%和 93.14%。在 CBA 分类模型加入 CCWGAN 模型生成的生成的人工异常流量样本训练后，模型的准确率和 F1 分数进一步提高到了 92.93%和 94.81%。此外，从图中可以看出模型的精度逐渐提高，召回率逐渐降低，即分类器模型对异常流量识别的能力得到增强，实验证明 CCWGAN 模型和构建 CBA 分类模型均能够在一定程度上提升对网络异常流量的检测性能。

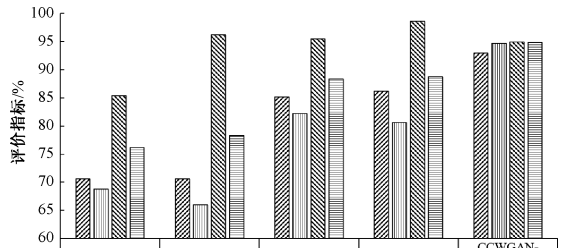


	CNN	CNN-BILSTM	CNN-BILSTM-ATT	CCWGAN-CBA
准确率	88.01	90.52	91.05	92.93
精度	82.62	88.31	89.29	94.69
召回率	99.06	97.53	97.35	94.94
F1分数	90.11	92.69	93.14	94.81

图 10　UNSW-NB15 消融实验

为了说明基于 CCWGAN-CBA 模型的性能，将其与朴素贝叶斯 NB、逻辑回归 LR、决策树 DT、随机森林 RF 等机器学习模型进行对比，结果如图 11 所示。从图中可以看出机器学习模型的精度偏低和召回率偏高，易将异常流量判定为正常流量，导致整体模型分类性能不佳，模型敏感度较低。其中 RF 模型表现相对其他机器学习模型更优，性能更稳定，RF 模型在准确率和 F1 分数上分别达到了 86.16%和 88.69%，而 CCWGAN-CBA 模型在准确率和 F1 分数上分别达到了 92.93%和 94.81%，准确率和 F1 分

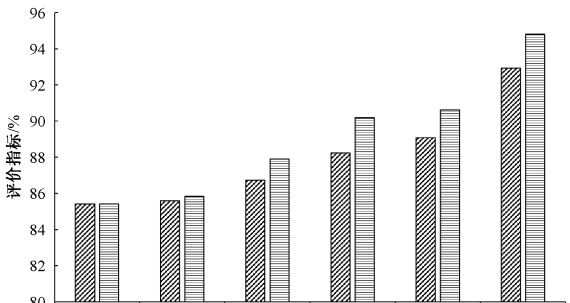
数相比于 RF 模型分别提高了 6.77%和 6.12%，可以看出 CCWGAN-CBA 模型相对与传统机器学习在网络入侵检测上更具有优势。



	NB	LR	DT	RF	CCWGAN-CBA
准确率	70.63	70.59	85.12	86.16	92.93
精度	68.78	65.98	82.20	80.63	94.69
召回率	85.39	96.19	95.46	98.54	94.94
F1分数	76.19	78.27	88.33	88.69	94.81

图 11 CCWGAN-CBA 模型与机器学习模型的对比

为了验证 CCWGAN-CBA 模型的有效性，将现有的先进的入侵检测方法中的模型与 CCWGAN-CBA 模型在 UNSW-NB15 数据集进行实验对比。各个模型的准确率、F1 分数的如图 12 所示。从图中可以看出 CCWGAN-CBA 模型在准确率和 F1 分数上优于其他网络入侵检测模型。首先，CBA 分类模型通过 CNN-BILSTM 融合提取流量数据的时间特征和空间特征，并利用 Attention 机制对时空融合特征进行加权，提高了模型的分类性能。而 RNN-LSTM[8]模型只考虑了流量数据的时间特征、FL-NIDS-CNN[9]模型只考虑了流量数据的空间特征，因此，这两个模型仅取得了 85.42%和 86.73%的入侵检测准确。其次，FCWGAN-BiLSTM[13] 模型、GAN-FS-KNN[11] 模型、ICVAE-DNN[10]模型均对原始数据集进行了数据样本生成处理，改善了数据集的类别不平衡问题，但 FCWGAN-BiLSTM 模型和 GAN-FS-KNN 模型因采用了特征选择方法，在减少了数据集的原始特征数量、提高模型运行的速率的同时，也丢失了部分较为重要的特征，导致模型的分类准确率不是很高。以上模型的检测性能表现最佳的为 ICVAE-DNN 模型，其准确率和 F1 分数达到了 89.08%和 90.61%，但仍低于 CCWGAN-CBA 模型的 92.93% 和 94.81%，说明 CCWGAN-CBA 模型能够更好的生成异常



	RNN-LSTM[8]	CWGAN-BiLSTM[13]	FL-NIDS-CNN[9]	GAN-FS-KNN[11]	ICVAE-DNN[10]	CCWGAN-CBA
准确率	85.42	85.59	86.73	88.24	89.08	92.93
F1分数	85.42	85.84	87.89	90.19	90.61	94.81

图 12 CCWGAN-CBA 模型与深度学习模型的对比

流量样本、减少数据集的类别不平衡度、提高模型对异常流量的检测准确率。

4 结 论

本文利用 CCWGAN 模型对 UNSW-NB15 数据集中 Analysis、Backdoor、Shellcode、Worms 等少数异常样本进行人工异常流量样本生成，改善了数据集的类别不平衡性问题；构建 CBA 分类模型，利用 CNN-BiLSTM 模型提取流量数据的时空特征，并引入 Attention 机制分配特征权重，增强了分类模型的性能。为证明 CCWGAN-CBA 模型的有效性，在 UNSW-NB15 数据集进行了仿真实验，通过生成样本实验可知，CCWGAN 模型可以有效提升 CBA 分类模型对少数类攻击流量的识别准确率；通过与其他模型的对比实验可知，CCWGAN-CBA 模型的入侵检测效果优于 RF 等机器学习模型、ICVAE-DNN 等深度学习模型。

CCWGAN-CBA 模型在 UNSW-NB15 数据集上进行的仿真实验取得了较好的二分类识别效果，但还需要在真实网络环境进行应用测试，以验证模型的实际性能。

参考文献

[1] TALITA A S, NATAZA O S, RUSTAM Z. Naïve bayes classifier and particle swarm optimization feature selection method for classifying intrusion detection system dataset[C]. Journal of Physics: Conference Series. IOP Publishing, 2021, 1752(1): 012021.

[2] DISHA R A, WAHEED S. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique[J]. Cybersecurity, 2022, 5(1): 1.

[3] WISANWANICHTHAN T, THAMMAWICHAI M. A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM[J]. IEEE Access, 2021, 9: 138432-138450.

[4] DEORE B, BHOSALE S. Intrusion detection system based on RNN classifier for feature reduction[J]. SN Computer Science, 2022, 3(2): 114.

[5] MOHAMMADPOUR L, LING T C, LIEW C S, et al. A survey of CNN-based network intrusion detection[J]. Applied Sciences, 2022, 12(16): 8162.

[6] ALQAHTANI A S. FSO-LSTM IDS: Hybrid optimized and ensembled deep-learning network-based intrusion detection system for smart networks[J]. The Journal of Supercomputing, 2022, 78(7): 9438-9455.

[7] ALKADI O, MOUSTAFA N, TURNBULL B, et al. A deep blockchain framework-enabled collaborative intrusion detection for protecting IoT and cloud

networks[J]. IEEE Internet of Things Journal, 2020, 8(12): 9463-9472.

[8] 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.

[9] MULYANTO M, FAISAL M, PRAKOSA SW, et al. Effectiveness of focal loss for minority classification in network intrusion detection systems[J]. Symmetry, 2020, DOI: org/10.3390/sym13010004.

[10] YANG Y, ZHENG K, WU C, et al. Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network[J]. Sensors, 2019, 19: 2528.

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

[12] YU Y, TANG B, LIN R, et al. CWGAN: Conditional wasserstein generative adversarial nets for fault data generation[C]. 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 2019: 2713-2718.

[13] MA Z, LI J, SONG Y, et al. Network intrusion detection method based on FCWGAN and BiLSTM [J]. Computational Intelligence and Neuroscience, 2022, DOI: org/10.1155/2022/6591140.

[14] JIN Q, LIN R, YANG F. E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN [J]. IEEE Systems Journal, 2019, 14(3): 3289-3300.

[15] FASSMEYER P, KORTMANN F, DREWS P, et al. Towards a camera-based road damage assessment and detection for autonomous vehicles: Applying scaled-YOLO and CVAE-WGAN [C]. 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), IEEE, 2021: 1-7.

[16] 童小钟,魏俊宇,苏绍璟,等.融合注意力和多尺度特征的典型水面小目标检测[J].仪器仪表学报,2023,44(1):212-222.

[17] AL-DAWERI M S, ZAINOL ARIFFIN K A, ABDULLAH S, et al. An analysis of the KDD99 and UNSW-NB15 datasets for the intrusion detection system[J]. Symmetry, 2020, 12(10): 1666.

[18] SALIH A A, ABDULAZEEZ A M. Evaluation of classification algorithms for intrusion detection system: A review[J]. Journal of Soft Computing and Data Mining, 2021, 2(1): 31-40.

作者简介

倪志伟,硕士研究生,主要研究方向为网络安全与深度学习。

E-mail: 202212490213@nuist.edu.cn

行鸿彦(通信作者),教授、博士生导师,主要研究方向为微弱信号检测与处理、生物医学信号采集与处理、智能化电子测量技术与仪器。

E-mail: xinghy@nuist.Edu.cn