CWGAN-DNN:一种条件 Wasserstein 生成对抗网络入侵检测方法

贺佳星，王晓丹，宋亚飞，来　杰
(空军工程大学防空反导学院，西安，710051)

摘要　针对现有的基于机器学习的入侵检测系统对类不平衡数据检测准确率低的问题，提出一种基于条件 Wasserstein 生成对抗网络(CWGAN)和深度神经网络(DNN)的入侵检测(CWGAN-DNN)。CWGAN-DNN 通过生成样本来改善数据集的类不平衡问题，提升对少数类和未知类的检测效率。首先，通过变分高斯混合模型(VGM)对原始数据中的连续特征进行处理，将连续特征的高斯混合分布进行分解；然后利用 CWGAN 学习预处理后数据的分布并生成新的少数类数据样本、平衡训练数据集；最后，利用平衡训练集对 DNN 进行训练，将训练得到的 DNN 用于入侵检测。在 NSL-KDD 数据集上进行的实验结果表明：利用 CWGAN 生成的数据进行训练，DNN 的分类准确率和 F1 分数提升了 5%，AUC 下降了 2%；与其他类均衡方法相比，CWGAN-DNN 的准确率至少提升了 3%、F1 分数和 AUC 提升了 1%。

关键词　入侵检测；类均衡技术；生成对抗网络；深度神经网络；高斯混合模型
DOI　10.3969/j.issn.1009-3516.2021.05.011
中图分类号　TP393.08　　**文献标志码**　A　　**文章编号**　1009-3516(2021)05-0067-08

CWGAN-DNN:An Intrusion Detection Method Based on Conditional Wasserstein Generative Adversarial Network

HE Jiaxing, WANG Xiaodan, SONG Yafei, LAI Jie
(Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China)

Abstract　In order to solve the problem of low detection accuracy of class imbalance data by existing intrusion detection systems based on machine learning, an intrusion detection method based on conditional Wasserstein generative adversarial network (CWGAN) and deep neural network (DNN) is proposed. CWGAN-DNN improve the class imbalance problem of data sets by generating samples, and the detection efficiency of intrusion detection system (IDS) on minority and unknown classes is increased. Firstly, the data are preprocessed by the variation Gaussian mixture model (VGM) to decompose the mixed distribution of continuous features. And then the CWGAN is used to learn the distribution of original dataset and generate minority-class data to balance the training dataset, and train the DNN with balanced dataset. Finally, the trained DNN is used for intrusion detection. The experimental results on NSL-KDD dataset show that the data generated by CWGAN can improve DNN's classification accuracy and F1 score by 5%, but AUC

收稿日期：2021-04-20
基金项目：国家自然科学基金(61876189;61273275;61806219;61703426)
作者简介：贺佳星(1997—)，男，陕西西安人，硕士生，研究方向：网络入侵检测、生成对抗网络。E-mail:hejiaixng@stu.xidian.edu.cn
通信作者：王晓丹(1967—)，女，陕西汉中人，教授，博士生导师，研究方向：机器学习、模式识别和计算机视觉。E-mail:afeu_wxd@163.com

引用格式：*贺佳星，王晓丹，宋亚飞，等. CWGAN-DNN:一种条件 Wasserstein 生成对抗网络入侵检测方法[J]. 空军工程大学学报(自然科学版)，2021，22(5)：67-74. HE Jiaxing, WANG Xiaodan, SONG Yafei, et al. CWGAN-DNN: An Intrusion Detection Method Based on Conditional Wasserstein Generative Adversarial Network[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(5): 67-74.*

decreases by 2%. Compared with other equalization methods, the accuracy, F1 score and AUC of CWGAN-DNN are improved by at least 3%, 1% and 1%.

Key words intrusion detection; class balancing method; generate adversarial network; deep neural network; Gaussian mixture model

近年来,多种基于机器学习的入侵检测系统(intrusion detection system, IDS)得到了广泛的研究,这些入侵检测方法具备良好的检测性能。但是实际的网络活动中,正常的流量和行为中占绝对主导地位,异常行为的数量较少。正常行为和攻击行为、不同攻击行为之间的类不平衡问题在入侵检测的数据集中普遍存在,极大影响了 IDS 的检测性能[1]。

针对入侵检测中的数据集类不平衡问题,文献[2]和[3]分别使用随机欠采样(random under-sampling,RUS)和随机过采样(random over-sampling,ROS)来解决 IDS 中的类不平衡问题。文献[4]更进一步将 RUS 和 ROS 技术结合起来应用。合成少数过度采样技术(synthetic minority over-sampling technique, SMOTE)[5-6]在数据生成领域中表现良好,文献[7~9]将 SMOTE 技术应用在 IDS 模型中,平衡数据集,提升训练效果。但 SMOTE 本身依赖于插值进行过采样,生成样本的拟合度比较低。

生成对抗网络(generative adversarial networks,GAN)[10]技术,能够从给定数据中学习其分布,并根据学习到的分布生成新的样本数据。利用生成对抗网络技术,文献[11]提出了一种将 GAN、粒子群算法和极限学习机结合的入侵检测方法,用 GAN 生成少数类样本。文献[12]提出一种基于 CGAN 的入侵检测技术,利用 CGAN 生成少数类样本,均衡训练数据集。

除类不平衡问题外,入侵检测数据集还存在着连续特征的混合模型问题。普通的 GAN 无法模拟一个 2D 数据集上的高斯混合分布的所有混合分量,学习到的数据分布存在失真[13]。例如,在入侵检测领域的常见数据集 NSL-KDD 数据集[14]包含 3 个离散特征和 38 个连续特征,共 41 维特征向量。本文利用核密度估计法对所有的连续特征进行估计后发现,38 个连续特征中的 22 个都存在高斯混合分布的情况。文献[11]提出的基于 GAN 的入侵检测方法均未考虑连续特征的混合分布问题。

为解决上面提到的入侵检测数据集中的类不平衡问题和连续特征的混合分布问题,本文提出了一种基于条件 Wasserstein 生成对抗网络(conditional Wasserstein generative adversarial network, CWGAN)和 DNN 的方法。首先,CWGAN-DNN 对原始数据集进行过滤,选择少数类样本,在训练的过程中使用条件向量控制生成样本的类别,保证只生成少数类样本;其次,对数据集中的连续特征利用变分高斯混合模型(variational Gaussian mixture model,VGM)[15]进行分解;最后,利用 CWGAN 学习处理后的数据集并生成新的少数类数据。达到平衡训练数据集,改善入侵检测系统性能的目的。

1 相关工作

1.1 数据生成技术

入侵检测数据集,如 NSL-KDD[10]通常组织为一个二维的表格 $\boldsymbol{T}$,表格中的每一行代表一条流量,每一列代表一维特征。$\boldsymbol{T}$ 中包含 N_c 个连续特征列 $\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_{N_c}\}$ 和 N_d 个离散特征列 $\{\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_{N_d}\}$。$\boldsymbol{T}$ 中的每一列都可以看作是一个随机变量,所有随机变量遵循一个未知的联合分布:

$$P\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_{N_c},\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_{N_d}\} \tag{1}$$

其中的任意一行可以表示为:

$$r_j=\{c_{1,j},c_{2,j},\cdots,c_{N_c,j},d_{1,j},d_{2,j},\cdots,d_{N_dj}\},\ j\in\{1,2,\cdots,n\} \tag{2}$$

式中:r_j 是来自联合分布的一个观测样本。

在数据的生成领域,合成数据的通常做法是通过将表中的每一列视为随机变量,建模一个联合多元概率分布,然后从中取样进行生成[16]。定义输入原始数据集为 $s=(r,y)$,生成器输出合成样本为 $s_G=[G(z,y'),y']$,其中:y、z 和 y' 分别表示原始类标签、高斯噪声和少数类标签。

1.2 生成对抗网络

条件生成对抗网络(conditional generative adversarial networks,CGAN)在原始 GAN[10]的基础上引入类别信息,以生成指定类的样本。生成器同时将噪声和类别信息作为输入,鉴别器在接收样本的时候也会收到对应类别信息。CGAN 的目标函数可以表示为:

$$\min_G\max_D V(G,D)=\min_G\max_D E_{x\sim p_r}[\log D(x\mid y)]+E_{z\sim p_z}[\log(1-D(G(z\mid y)))] \tag{3}$$

式中:z 为随机噪声;x 为原始样本;y 为类别信息;p_z 为 z 的分布;G 为生成器;D 为鉴别器;p_r 为真实数据 x 的分布;$G(z)$ 为 G 生成的伪数据;$D(x)$ 为 D

给样本打出的分数；$E(\cdot)$为期望值。

GAN 和 CGAN 中采用的 Jensen-Shannon 散度会导致模式崩溃和梯度消失问题[17]。针对该问题 WGAN-GP[18]（Wasserstein GAN-gradient penalty）将计算损失函数的 JS 数度改为泥土移动（earth mover，EM）距离，并增加了梯度惩罚项，来满足 Lipschitz 约束，避免训练过程中判别器无法收敛的情况出现，解决了 GAN 和 CGAN 中的梯度消失和模式崩溃问题。WGAN-GP 的目标函数为：

$$V_{(G,D)}=\max_{D\in 1-\text{lipschitz}}\{E_{x\sim p_r}[D(x)]-E_{x\sim p_g}[D(x)]-\lambda E_{x\sim p_{\text{penaty}}}[\|\nabla_x D(x)\|-1]^2\} \tag{4}$$

式中：λ 是人为指定的参数；$\|\nabla_x D(x)\|$ 表示对 $D(x)$ 中 x 的计算范式；$x\sim p_{\text{penaty}}$ 表示从 p_r 上的一个点和 p_g 的一个点的连线上取中间位置的 x。

2　基于 CWGAN 的入侵检测方法

CWGAN-DNN 的框架由数据预处理模块、CWGAN 模块和深度神经网络（DNN）3 个部分组成，如图 1 所示。第 1 步，将原始数据集进行处理；第 2 步，利用 CWGAN 模块生成均衡训练集；第 3 步，DNN 模块使用均衡数据集进行训练，并在测试数据集上执行入侵检测。

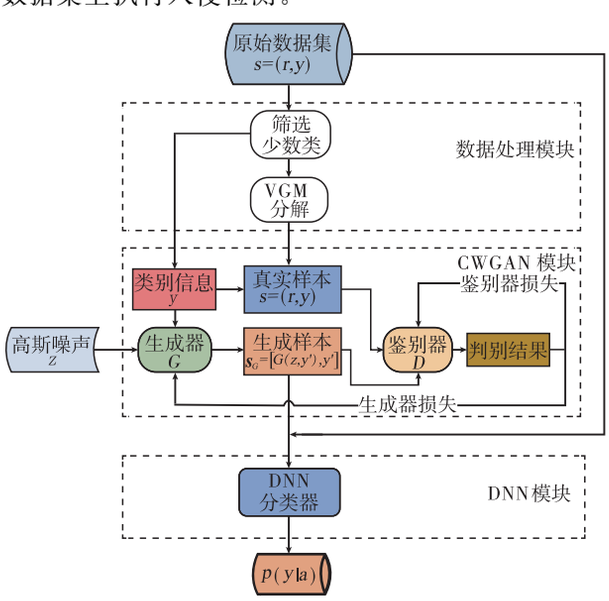


图 1　CWGAN-DNN 入侵检测方法的结构

2.1　数据预处理

数据预处理部分提取少数类信息，并对混合分布的连续特征进行分解。以原始样本 $s=(r,y)$ 为输入，首先筛选出需要进行增强的少数类数据 (r,y')，而后进行 VGM 分解得到少数类样本数据 $s'=(r',y')$。

对于高斯混合分布的连续特征列，利用 VGM 来进行处理。入侵检测数据集中，离散特征可以表示为 one-hot 向量[19]，而高斯混合分布的连续特征难以全面地表示，本文选择将其利用 VGM 分解。

连续特征列中的每个值，通过 VGM 都将其转换为一个指示所在分量的 one-hot 向量 $\boldsymbol{v}_*$ 和一个指示在分量下值大小的标量 a_* 的并联。

利用 VGM 分解连续特征的混合分布，首先对于每一个连续列，使用 VGM 估计其分量的个数，并训练高斯混合模型；其次，对于 $\boldsymbol{C}_i$ 中的每个值 $c_{i,j}$，计算其在每个分量下的概率；最后在所求得的概率最大的子分布上采样并进行归一化。

通过 VGM 分解将 $c_{i,j}$ 分解为 $\boldsymbol{v}_{i,j}\oplus a_{i,j}$，便于 GAN 更好地学习混合分布。可以将数据表格中的一行 $\boldsymbol{r}_j$ 改写为连序特征和离散特征的串联 $\boldsymbol{r}'_j$：

$$\boldsymbol{r}'_j=\boldsymbol{v}_{1,j}\oplus a_{1,j}\oplus\cdots\oplus\boldsymbol{v}_{N_c,j}\oplus a_{N_c,j}\oplus\boldsymbol{d}_{1,j}\oplus\cdots\oplus\boldsymbol{d}_{N_d} \tag{5}$$

2.2　CWGAN 模块

CWGAN 的整个训练过程见表 1，其中 θ_G、η_{θ_G}、θ_D、η_{θ_D} 分别为生成器的网络参数、梯度和鉴别器的网络参数、梯度。

表 1　CWGAN 训练过程

算法 1：基于 CWGAN 和 VGM 的少数类数据生成

输入：$s=(\boldsymbol{r},\boldsymbol{y})$，其中 $\boldsymbol{r}$ 为特征向量，$\boldsymbol{y}$ 为类别标签

输出：$\boldsymbol{s}_G=[G(\boldsymbol{z},\boldsymbol{y}'),\boldsymbol{y}']$

$c_{i,j}=\boldsymbol{v}_{i,j}\oplus a_{i,j}$ / * 连续变量分解 * /

$\boldsymbol{r}'_j=\boldsymbol{v}_{1,j}\oplus a_{1,j}\oplus\cdots\oplus\boldsymbol{v}_{N_c,j}\oplus a_{N_c,j}\oplus\boldsymbol{d}_{1,j}\oplus\cdots\oplus\boldsymbol{d}_{N_d}$ / * 特征向量分解 * /

while D has not converged to 0.5 **do** / * CWGAN 训练 * /

for t steps **do** / * 优化鉴别器 * /

从 $p_{\text{data}}(\boldsymbol{r}',\boldsymbol{y}')$ 中采样 $\{(\boldsymbol{r}'_i,\boldsymbol{y}'_i)\}_{i=1}^{n_z}$

从 $p_z(\boldsymbol{z})$ 中采样 $\{(\boldsymbol{z}_i)\}_{i=1}^{n_z}$

$$\eta_{\theta_D}\leftarrow\nabla\theta_D\Big[\frac{1}{n_z}\sum_{i=1}^{n_z}\{D((\boldsymbol{r}'_i,\boldsymbol{y}'_i))-D(G(\boldsymbol{z}_i,\boldsymbol{y}'_i),\boldsymbol{y}'_i)-\lambda E_{(\boldsymbol{r},\boldsymbol{y})\sim p_{\text{penaty}}}[\|\nabla_{(\boldsymbol{r},\boldsymbol{y})}D(\boldsymbol{r},\boldsymbol{y})\|-1]^2\}\Big]$$

$\theta_D\leftarrow\theta_D+\alpha_D\cdot\text{Adam}(\theta_D,\eta_{\theta_D})$

end

从 $p_z(\boldsymbol{z})$ 中采样 $\{(\boldsymbol{z}_i)\}_{i=1}^{n_z}$. / * 优化生成器 * /

$$\eta_{\theta_G}\leftarrow\nabla\theta_G\Big[\frac{1}{n_z}\sum_{i=1}^{n_z}(D(G(\boldsymbol{z}_i,\boldsymbol{y}'_i),\boldsymbol{y}'_i))\Big]$$

$\theta_G\leftarrow\theta_G-\alpha_G\cdot\text{Adam}(\theta_G,\eta_{\theta_G})$

end

return u / * 生成样本 * /

训练 CWGAN 过程中,依次交替训练生成器和鉴别器,训练主要的步骤如下:

1)首先将随机噪声向量 $\boldsymbol{z}$ 与类别标签 $\boldsymbol{y}'$ 输入 G,训练并得到生成样本 $\boldsymbol{s}_G$;

2)固定生成器 G,对 D 进行训练,更新 θ_D;

3)固定鉴别器 D,对 G 进行训练,更新 θ_G;

4)在鉴别器的损失值未达到 0.5 之前,循环执行步骤 1)~3),G 和 D 交替训练,使得生成的样本不断接近真实样本。

2.2.1 鉴别器 D

鉴别器 D 由一个 3 层的全连接网络组成。在训练的过程中,将 $\boldsymbol{s}'$ 和 $\boldsymbol{s}_G$ 混合作为判别器 D 的输入,输出为样本属于真实样本 $\boldsymbol{s}'$ 和伪样本 $\boldsymbol{s}_G$ 的分类概率值,然后通过激活函数将概率值转换为预测标签。对于一个特定的输入 $(\boldsymbol{r},\boldsymbol{y})$,需要判断它来自 $\boldsymbol{s}'$ 而不是 $\boldsymbol{s}_G$ 的概率 $D(\boldsymbol{r},\boldsymbol{y})$,见表 1 中第 7、8 行。

在 WGAN 模型当中,不再对鉴别器的值取对数。得到损失值后,可以计算梯度 η_{θ_D},通过 Adam 算法更新鉴别器的网络参数 θ_D。

2.2.2 生成器 G

生成器 G 同样采用一个全连接网络,将少数类标签 $\boldsymbol{y}'$ 和高斯噪声 $\boldsymbol{z}$ 作为输入,训练过程中将类别信息 $\boldsymbol{y}'$ 与噪声 $\boldsymbol{z}$ 连接为 $(\boldsymbol{z},\boldsymbol{y}')$,$G$ 生成样本为 $\boldsymbol{s}_G=[G(\boldsymbol{z},\boldsymbol{y}'),\boldsymbol{y}']$。表 1 中第 11、12 行给出了公式。

2.3 DNN 部分

深度神经网络模块(DNN)被用来执行入侵检测。在 CWGAN-DNN 当中,DNN 被设计为一个包含 2 个隐藏层的全连接网络。

在训练阶段,DNN 模型以均衡训练集作为输入。最终给出它们在不同入侵类别上的概率分布 $p(\boldsymbol{y}|\boldsymbol{r})$。在测试阶段,DNN 模块以测试数据集上未曾出现过的样本作为输入,检测 CWGAN 对于 DNN 网络的提升效果。

3 实验及分析

本文过实验来评估 CWGAN-DNN 的性能。将 CWGAN-DNN 与传统的入侵检测方法、基于类均衡的方法和一些先进的入侵检测方法进行了比较。

3.1 基准数据集

采用 NSL-KDD 数据集对 CWGAN-DNN 进行评估。NSL-KDD 是评估入侵检测领域的经典基准数据集,剔除了 KDDCUP99 中的冗余数据,并对训练集和测试集的构成进行了调整,更适用于网络入侵检测的研究。NSL-KDD 中的每条数据包括 41 个特征。将 NSL-KDD 预先划分为训练集和测试集,分别为 KDDTrain+_20 和 KDDTest+。NSL-KDD 数据集的详细信息如表 2 所示。

表 2 NSL-KDD 数据集样本类型的分布

Label	KDDTrain+_20	KDDTest+	KDDTest-21
Normal	13 449	9 711	2 152
Dos	9 234	7 458	4 342
Probe	2 289	2 421	2 402
U2R	11	200	200
R2L	209	2 754	2 754
合计	25 192	22 544	11 850

从表 1 中可以看出,训练集 KDDTrain+_20 存在严重的类别不平衡现象,U2R 和 R2L 攻击样本的数量严重偏少。

3.2 评价指标

为了定量评价 CWGAN-DNN 的性能,本文采用准确率(accuracy)、精确率(precision)、召回率(recall)和 F1 分数(F1 score)作为主要指标来衡量本模型的多分类性能:

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

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

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

$$F1=\frac{2\times TP}{2\times TP+FP+FN} \tag{9}$$

式中:TP、TN、FP 和 FN 分别表示真阳性、真阴性、假阳性和假阴性。

此外,文中采用 AUC(area under the ROC curve,AUC)来反映综合能力。ROC(receiver operating characteristic curve,ROC)是研究学习器泛化性能的有效工具[20]。AUC 是 ROC 曲线下的面积,可以更进一步的比较分类器的性能。

3.3 实验设置

3.3.1 实验设置

为了测试基于 CWGAN-DNN 的网络入侵检测方法的性能,本文设计了以下实验:

实验 1:CWGAN-DNN 模型的训练实验。

实验 2:CWGAN-DNN 与传统入侵检测方法的性能对比实验。

选择一些常见的机器学习方法进行比较。朴素贝叶斯和决策树是经典的机器学习方法,它们能以较低的开销提供较好的性能。随机森林[21]是一种由多个决策树构成的集成学习方法,但比决策树具有更强的泛化能力。支持向量机是一种经典而高效的分类方法,但不适用于大数据[22-23]。多层感知器(MLP)[24]是一种最简单的深度学习模型,具有稳定的分类能力。

实验 3:CWGAN-DNN 与不同数据类均衡方法

的性能对比实验。

在类平衡方法方面，选用了随机过采样（ROS）、SMOTE 和自适应综合过采样（adaptive synthetic，ADASYN）[25]技术进行对比，并将其与第 2.3 节中叙述的 DNN 结合。利用类平衡方法生成样本，然后将平衡后的样本输入 DNN 进行入侵检测，将这些方法分别记做 ROS＋DNN、SMOTE＋DNN 和 ADASYN＋DNN。这些方法中的 DNN 参数都保持一致，参考表 2 进行设置。

表 2　DNN 的网络结构

序号	layer	size	activation
1	Fully-connected	100	ReLU
2	Fully-connected	100	ReLU
3	Fully-connected	40	ReLU
4	Fully-connected	5	Softmax

实验 4：CWGAN-DNN 与现有的入侵检测模型的性能对比实验。

将 CWGAN-DNN 与几种较为先进的入侵检测方法进行比较。基于模糊的神经网络（fuzziness-based neural network，NN）是一种半监督学习方法，它可以提高通过模糊分类进行入侵检测的泛化能力；文献[26]在最小二乘支持向量机（LSSVM）之前引入了一种基于互信息的特征选择（MIFS），它对特征进行贪婪选择，提升入侵检测的效果；基于 LSSVM＋MIFS，文献[27]进一步提出了一种灵活的 MIFS（FMIFS）方法，是一种无需经验参数的自适应特征选择方法；文献[28]将一维的入侵检测数据集转化为二维灰度图，并利用一个 2 层的卷积神经网络（CNN）进行入侵检测。这 4 种方法，均是在标准的 KDDTrain＋20％数据集上进行训练，并在 KDDTest＋数据集上进行了测试，这一点上与本文方法相同。本文采用参考文献中给出的性能数据进行对比，与 CWGAN-DNN 进行比较。

实验 5：不同数据生成率对 CWGAN-DNN 性能影响对比实验。

为了验证生成样本的效果，本文还对数据的生成率对于 CWGAN-DNN 入侵检测性能的影响，通过一系列不同生成率的实验来评估 CWGAN-DNN。

3.3.2　实验环境及参数设置

本文实验基于 Windows10 操作系统下 Pytorch 和 sklearn 框架进行实现。CWGAN 模块中，鉴别器 D 和生成器 G 的体系结构相对灵活，可以根据具体情况进行设置。本文根据文献[16]确定了 CWGAN 的结构，并在验证集上通过网格搜索调整 CWGAN 的参数。生成器和鉴别器的各层的网络大小分别设置为 100-256-256-63 和 63-256-256-5。

CWGAN 的学习率（在算法 1 中表示为 α_D 和 α_G）都设置为 0.001。批大小（用 m 表示）为 500，在每次迭代时优化 D 和 G 一次。生成样本个数的数量设置为原始样本的数量的 1.5 倍，即数据的生成比（生成样本数量：原始样本数量）为 1.5∶1。对 CWGAN 模块进行训练，直到所有类的 D 收敛到 0.5，此时鉴别器已经无法分辨生成样本和真实样本，即生成器具备了生成高质量样本的能力，这意味着 CWGAN 已经完成优化。

DNN 使用具有 100 个神经元的全连接层作为输入层。之后采用 2 个大小分别为 100 和 40 的全连接层作为隐藏层。输出层的大小等于类数 5。学习率设置为 0.000 1。最后使用 Softmax 函数，将输出向量转换为概率分布 $p(\boldsymbol{y}|\boldsymbol{r})$。表 2 描述了 DNN 模块的详细结构。

作为对比试验，传统入侵检测方法和类均衡方法均使用 sklearn 框架进行实现，参数选择上使用网格搜索法进行优化；在与其它的先进入侵检测方法进行对比时，参考对应文献中的数据。

3.4　实验结果及分析

3.4.1　实验 1：CWGAN 和 DNN 的训练

利用 CWGAN 为 Probe、R2L 和 U2R 等 3 个少数类生成样本，训练过程中鉴别器 D 的损失曲线见图 2。

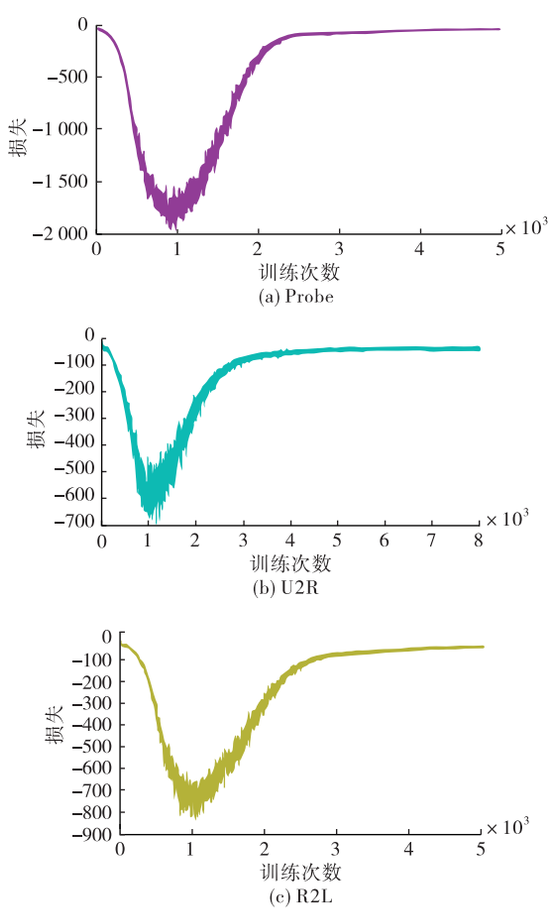


图 2　NSL-KDD 上少数类样本训练鉴别器的损失曲线

图 2（a）、（b）和（c）分别对应 Probe、U2R 和

R2L 类,可以看到 CWGAN 在 Probe、U2R 和 R2L 类上分别在 3 000、5 000 和 3 500 次训练后鉴别器损失收敛。

3 个类的损失函数曲线在收敛之后均存在不同程度的震荡,其中 U2R 类的震荡最为明显,R2L 次之,Probe 的震荡幅度最小。这是因为 KDD Train +数据集当中 U2R 类的仅有 11 个样本,较难收敛。

以图 2(b)中 Probe 类的损失曲线为例进行分析。在 0～1 000 次迭代时,G 无法生成有效的数据,D 可以很轻易的区分出生成样本,损失值较大;在 1 000～3 000 次迭代过程中,G 的生成能力不断提高,D 给出生成样本的得分越来越接近原始样本的得分,损失值不断降低,并在 4 000 次后达到动态稳定。

在 DNN 模块训练过程中,将平衡后的数据集作为输入。DNN 模块在训练后,在测试数据集上进行入侵检测。以交叉熵函数为损失,采用 Adam 算法进行优化,将学习率设置为 0.000 1。DNN 的损失曲线如图 3 所示,分类器已经收敛。

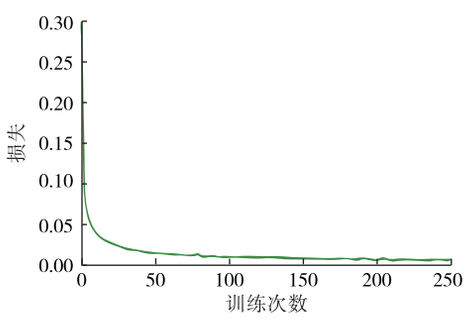


图 3 DNN 分类器训练过程中的损失曲线

3.4.2 实验 2:CWGAN-DNN 与传统入侵检测方法的性能对比实验

如表 3 所示,与传统的机器学习方法相比,CWGAN-DNN 的准确度、和 F1 分数分别提升了最少 3%,而 AUC 比 MLP 有 2%的下降。在传统的方法中,MLP 的准确率较好,在准确度和 F1 分数相比其他机器学习方法有 1%～2%下降的情况下,AUC 提升至少 9%。MLP 作为一种深度学习模型,具有良好的分类能力,但在处理类不平衡数据方面仍然存在一定的困难。朴素贝叶斯在 NSL-KDD 上取得了所有机器学习算法中最高的 AUC(83%),因为它是一个生成模型,受数据类不均衡的影响较小,但它表达能力的不足导致它的分类性能较差。随机森林在 2 个数据集上取得了不错的成绩,显示了其作为集成模型的优秀泛化能力。然而,上述结果均反映出传统方法对数据的表达能力不足和对类不平衡数据的处理能力较弱的缺陷。

表 3 CWGAN 与传统机器学习方法比较 单位:%

模型	accuracy	F1 Score	AUC
Bayes	60	53	83
DT	75	71	72
RF	76	72	82
SVM	70	67	80
Ours	79	75	90

3.4.3 实验 3:CWGAN-DNN 与不同数据类均衡方法的性能对比实验

如表 4 所示,对类均衡方法,CWGAN-DNN 与常见的类均衡方法相比,CWGAN-DNN 的准确度、F1 分数和 AUC 分别提升了最少 3%、1%和 1%。采用均衡样本训练的 DNN 的表现都比仅使用原始训练集更好,因为类均衡方法有助于解决类不平衡问题。SMOTE 和 ADASYN 方法可以较好地从生成的样本中学习少数类,其准确度与 RF 等机器学习算法相同,F1 分数和 AUC 均有一定的提升。然而相比于仅使用原始数据集进行训练的 DNN,类均衡方法在提升准确度和 F1 分数的同时,AUC 有至少 2%的下降,体现了分类能力和泛化能力的取舍。

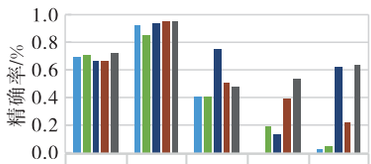
表 4 CWGAN 与其他类均衡方法的比较 单位:%

模型	accuracy	F1 Score	AUC
DNN	74	70	92
ROS+DNN	76	74	89
SMOTE+DNN	76	74	88
ADASYN+DNN	75	72	86
Ours	79	75	90

如图 4 所示,所有的类均衡方法对少数类的检测能力均有一定的提升。

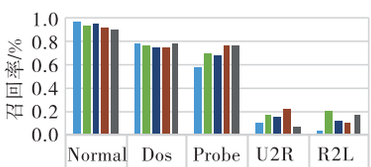
在图 4(a)中可以看到,ADASYN 方法对 Probe 和 R2L 两类的精确度提升明显,在 U2R 类上表现一般。而 CWGAN 可以大幅度提升 U2R 和 R2L 两个少数类的检测精确度,和本文选择的方法相比均有不同程度的提升。在 U2R 和 R2L 类上,CWGAN-DNN 对比 DNN 分别提升了 55%和 61%,对比其他类均衡方法,分别了 15%～40%和 1%～58%的提升。

图 4(b)中,CWGAN 在 Probe 和 R2L 两个少数类上的召回率提升幅度与其他性能较好的类均衡方法相近。但在 U2R 类上,CWGAN-DNN 的召回率仅为 6%。根据式(16)的描述,recall 较低,说明 FP 的值较大,有大量的其他类的样本被检测为 U2R 类。这是因为训练集当中 U2R 类仅有 11 个样本,导致 CWGAN 在 U2R 样本上的进行训练时无法学习到 U2R 类样本的特征,生成的样本质量较低。



	Normal	Dos	Probe	U2R	R2L
DNN	0.70	0.93	0.41	0.00	0.03
ROS+DNN	0.71	0.86	0.41	0.20	0.06
ADASYN+DNN	0.67	0.95	0.76	0.15	0.63
SMOTE+DNN	0.67	0.96	0.51	0.40	0.23
Ours	0.73	0.96	0.49	0.55	0.64

（a）类均衡方法在各类上的精确度



	Normal	Dos	Probe	U2R	R2L
DNN	0.97	0.78	0.58	0.10	0.02
ROS+DNN	0.94	0.77	0.70	0.17	0.19
ADASYN+DNN	0.95	0.75	0.67	0.15	0.11
SMOTE+DNN	0.92	0.74	0.76	0.21	0.09
Ours	0.90	0.78	0.77	0.06	0.17

（b）类均衡方法在各类上的召回率

图 4　类均衡方法在各类别上的性能比较

3.4.4　实验 4：CWGAN-DNN 与现有的入侵检测模型的性能对比实验

如表 5 所示，在与其他的先进的入侵检测算法进行比较时，CWGAN-DNN 在准确率、召回率和 F1 分数 3 项指标上都追平或超过其他的方法，而在精确度和 AUC 指标上，CWGAN-DNN 仅落后 CNN 4%和 2%。CWGAN-DNN 性能的提高得益于 CWGAN 对少数类的高代表性合成样本，缓解类不平衡问题，模拟了未知异常。

表 5　CWGAN 与其他先进方法的对比

模型	准确度	召回率	精确度	F1 分数	AUC
NN	0.75	0.75	0.77	0.70	0.93
LSSVM+MIFS	0.76	0.75	0.73	0.73	0.85
LSSVM+FMILS	0.78	0.78	0.68	0.72	0.84
CNN	0.78	0.78	0.83	0.74	0.93
Ours	0.79	0.78	0.79	0.75	0.91

3.4.5　实验 5：不同数据生成率对 CWGAN-DNN 性能影响对比实验。

具有不同生成比率的 IGAN-IDS 的精度如表 4 所示。与没有合成样本的结果（$r=0$）相比，可以观察到 CWGAN-DNN 在训练过程中生成样本时表现出更高的性能（精确度提高 4%），说明了 CWGAN 的积极作用。进一步比较了不同生成比率的 IGAN-IDS 的细节性能，如图 5 所示。从图 5 可以看出，CWGAN 在生成率为 1.5 时获得最高的准确度、召回率和 F1 分数。而在生成率为 0.5 和 1 时，AUC 的值最低。在实验中，建议生成率应控制在 1.5。

综上所述，CWGAN-DNN 确实可以通过生成样本来提高入侵检测的性能，而且生成的比率应该受到严格的控制。

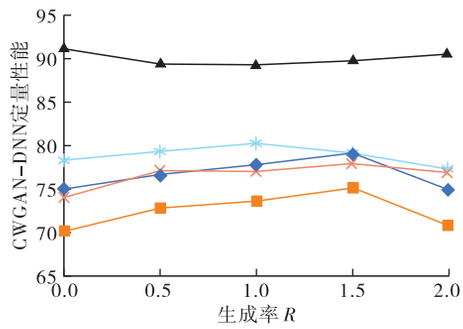


图 7　CWGAN-DNN 在不同生成率下的表现

4　结语

为解决入侵检测系统中由训练样本类不均衡导致的检测识别率和泛化能力较差的问题，本文通过 VGM 和 CWGAN 技术生成少数类新的训练样本，改善训练集中的类不平衡问题，并在均衡训练集上训练一个全连接 DNN 网络进行入侵检测。将本文提出的 CWGAN-DNN 入侵检测技术进行性能评估，并与传统入侵检测方法、类均衡方法和主流入侵检测技术进行了比较。同时对于 CWGAN 在少数类的检测性能和 CWGAN 数据生成率对检测性能的影响。实验结果表明，CWGAN-DNN 相比于传统的基于机器学习和类均衡技术的入侵检测方法，能够有效提升少数类的检测性能，从而提高总体分类能力和泛化能力。

参考文献

[1] YANG Y, ZHENG K, WU C, et al. Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational Auto Encoder and Deep Neural Network[J]. Sensors, 2019, 19(11):2528.

[2] KUANG L, ZULKERNINE M. An Anomaly Intrusion Detection Method Using the CSI- KNN Algorithm[C]// Proceedings of the 2008 ACM Symposium on Applied Computing. New York, NY, USA: ACM, 2008:921-926.

[3] ABDULHAMMED R, FAEZIPOUR M, ABUZNEID A, et al. Deep and Machine Learning Approaches for Anomaly-Based Intrusion Detection of Imbalanced Network Traffic[J]. IEEE Sensors Letters, 2019, 3(1):1-4.

[4] DAVID A C, NITESH V C, AARON S. Combating

Imbalance in Network Intrusion Datasets[C]// 2006 IEEE International Conference on Granular Computing. Atlanta, GA, USA:IEEE,2006:732-737.

[5] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic Minority Over-Sampling Technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.

[6] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning[C]//International Conference on Intelligent Computing. Hefei, China:Springer,2005:878-887.

[7] JULIAN L, ALBERTO F, FRANCISCO H, et al. Addressing Data-Complexity for Imbalanced Datasets: A Preliminary Study on the Use of Preprocessing for C4.5[C]// 2009 the Ninth International Conference on Intelligent Systems Design and Applications. Pisa, Italy:IEEE,2009:523-528.

[8] QAZI N, RAZA K. Effect of Feature Selection, SMOTE and under Sampling on Class Imbalance Classification[C]// 2012 UKSim 14th International Conference on Computer Modelling and Simulation. Cambridge, UK:IEEE,2012:145-150.

[9] TESFAHUN A, BHASKARI D L. Detection Using Random Forests Classifier with SMOTE and Feature Reduction[C]// 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies. Pune, India:IEEE,2013:127-132.

[10] 王坤峰,苟超,段艳杰,等.生成式对抗网络 GAN 的研究进展与展望[J].自动化学报,2017,43(3):321-332.

[11] 杨彦荣,宋荣杰,周兆永. 基于 GAN-PSO-ELM 的网络入侵检测方法[J]. 计算机工程与应用,2020,56(12):66-72.

[12] 彭中联,万巍,荆涛,等.基于改进 CGANs 的入侵检测方法研究[J].信息网络安全,2020, 20(5): 47-56.

[13] SRIVASTAVA A, VALKOV L, RUSSELL C. VEEGAN: Reducing Mode Collapse in GANs Using Implicit Variational Learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. [S.l.]:ACM, 2017: 3310-3320.

[14] TAVALLAEE M, BAGHERI E, LU W, et al. A Detailed Analysis of the KDD CUP 99 Dataset[C]// 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Ottawa, ON, Canada:IEEE, 2009:1-6.

[15] FU K S. Pattern Recognition and Machine Learning [M].[S.l.]:Springer,1970.

[16] XU L, SKOULARIDOU M, CUESTA-INFANTE A. Modeling Tabular Data Using Conditional GAN [C]//The 33rd Conference on Neural Information Processing Systems. Vancouver, Canada:[s. n.], 2019:1-15.

[17] MIRZA, M, OSINDERO S. Conditional Generative Adversarial Nets[Z]. ArXiv:1411.1784 2014.

[18] HUANG S, LEI K. IGAN-IDS: An Imbalanced Generative Adversarial Network towards Intrusion Detection System in Ad-Hoc Networks[J]. Ad Hoc Networks, 2020, 105: 102177.

[19] 周志华.机器学习[M]. 北京: 清华大学出版社, 2016: 33-35.

[20] PANDA M, PATRA M. Network Intrusion Detection Using Naive Bayes[J]. International Journal of Computer Science and Network Security, 2007:258-263.

[21] WANG Y, WONG J, MINER A. Anomaly Intrusion Detection Using One Class SVM[C]//Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop. West Point, NY, USA:IEEE, 2004: 358-364.

[22] HASAN M A M, NASSER M, PAL B, et al. Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS)[J]. Journal of Intelligent Learning Systems & Applications, 2014, 6(1):45-52.

[23] MORADI M, ZULKERNINE M. A Neural Network Based System for Intrusion Detection and Classification of Attacks[J]. Proceedings of 2004 IEEE International Conference on Advances in Intelligent Systems.[S.l.]:IEEE, 2014:140-01-04.

[24] HE H, BAI Y, GARCIA E A, et al. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning[C]// 2008 IEEE International Joint Conference on Neural Networks. Hong Kong, China: IEEE, 2008: 1322-1328.

[25] ASHFAQ R A R, WANG X Z, HUANG J Z, et al. Fuzziness Based Semi-Supervised Learning Approach for Intrusion Detection System[J]. Information Sciences, 2017,378(1): 484-497.

[26] AMIRI F, YOUSEFI M M R, LUCAS C, et al. Mutual Information-Based Feature Selection for Intrusion Detection Systems[J]. Journal of Network & Computer Applications, 2011, 34(4):1184-1199.

[27] AMBUSAIDI M, HE X, NANDA P, et al. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm[J]. IEEE Transactions on Computers, 2016,65(10) 2986-2998.

[28] LI Z, QIN Z, HUANG K et al. Intrusion Detection Using Convolutional Neural Networks for Representation Learning[C]//International Conference on Neural Information Processing. Guangzhou, China: Springer, 2017:858-866.

(编辑:徐楠楠)