计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201200077

基于深度学习的网络流量异常检测

杨月麟　毕宗泽
中国科学技术大学软件学院　合肥 230022

摘　要　为了解决网络流量数据的远程依赖性及数据集样本不平衡导致的长尾效应等问题,文中基于视觉 Transformer 提出一种网络流量异常检测模型,将多头自注意力引入残差网络,通过 Feature Embedding 将输入的稀疏高维度特征转化为稠密低维度特征,并加入二维相对位置编码,实现对流量数据位置全局感知,解决网络流量数据的远程依赖性。视觉 Transformer 模块包括编码器与解码器,编码器由 N 个相同的层堆叠组成,每层包括一个多头卷积自注意力层和一个二维卷积前馈网络,解码器在每层中插入一个查询自注意力的附加层,得到合成的流量特征图。同时提出深度自适应特征学习算法,通过半监督学习缓解数据分布不平衡导致的长尾效应问题,根据模型对无标签数据中尾部类别数据识别精确率高的特点,在无标签数据中挑选预测类别为尾部类别的样本加入到已标记集合,通过引入尾部类别样本缓解类别不平衡问题。使用 CIC-IDS-2017 网络入侵检测数据集进行实验评估。通过对比实验证明,模型的尾部样本检测准确率高于其他深度学习模型在提高检测性能的同时减少了检测时间,在网络流量异常检测领域具备实际应用价值。

关键词:深度学习;异常检测;注意力;类别再平衡;残差网络
中图法分类号　TP183

Network Anomaly Detection Based on Deep Learning

YANG Yue-lin and BI Zong-ze
School of Software Engineering,University of Science and Technology of China,Hefei 230022,China

Abstract　This paper proposes a novel and general end-to-end convolutional transformer network for modeling the long-range spatial and temporal dependence on network anomaly detection. The core ingredient of the proposed model is the feature embedding module by just replacing the spatial convolutions with proposed global self-attention in the final three bottleneck blocks of a ResNet,and the multi-head convolutional self-attention layer in encoder and decoder,which learns the sequential dependence of network traffic data. Our model uses an encoder,built upon multi-head convolutional self-attention layers,to map the input sequence to a feature map sequence,and then another deep networks,incorporating multi-head convolutional self-attention layers,decode the target synthesized feature map from the feature maps sequence. We also present a class-rebalancing self-training framework to alleviate the long tail effect caused by the imbalance of data distribution through semi-supervised learning,which is motivated by the observation that existing SSL algorithms produce high precision pseudo-labels on minority classes. The algorithm iteratively retrains a baseline SSL model with a labeled set expanded by adding pseudo-labeled samples from an unlabeled set,where pseudo-labeled samples from minority classes are selected more frequently according to an estimated class distribution. In this paper,CIC-IDS-2017 datasets is used for experimental evaluation. The experiments shows that the accuracy of our model is higher than that of other deep learning models,which improves detection performance while reducing detection time,and has practical application value in the field of network traffic anomaly detection.

Keywords　Deep learning,Anomaly detection,Attention,Class-rebalancing,ResNet

1　引言

网络攻击方式众多,其作用位置各不相同[1],可以通过监测异常流量的方式检测[2]。由于网络拓扑结构复杂,流量数据不仅存在局部的信息关联,也存在远程依赖,数据形式多样且带有高维特征,混有噪声信息,IDS 只能捕获数据的局部信息。研究人员尝试应用深度学习中的计算机视觉技术进行安全防御[3],但数据集样本类别不平衡导致的长尾效应会使训练的模型预测结果更偏向于头部类别。而通常用于解决长尾效应的重采样、重加权等方法,非常依赖标签来重新平衡模型预测偏差。在半监督算法中通常利用在有标签数据上训练的模型对无标签数据生成伪标签,但如果伪标签是由带有预测偏差的模型生成的,而测试集是平衡的,则使用此类伪标签进行模型训练会加剧模型预测偏差并恶化模型质量,使模型更加偏向于头部类别,从而导致所有类别的平均召回率都降低。本文提出了一种结合卷积与 Transformer 的异常检测模型(Convolutional Transformer),以及类别再平衡自训练算法(Class-Rebalancing Self-Training Algorithm,CReST)以解决

通信作者:杨月麟(yangyuelin_k@outlook.com)

以上问题,实验表明该模型能够同时捕捉流量数据的局部与全局信息,且能够有效解决网络流量数据集长尾效应,优于以往的模型。

2 相关工作

近年来深度学习技术在网络安全领域的探索不断推进。Chia 等[4]结合了 Transformer[5] 和 CNN,通过实验证明,相比于 RNN 和 LSTM 等,Transformer 拥有更强大的编码能力,也能更高效地利用 GPU 等高性能设备完成大规模训练。研究人员在计算机视觉领域尝试使用 Transformer[6],将 Self-Attention 机制与 CNN 架构结合或完全替代 CNN。谷歌提出了一种基于 Transformer 网络的计算机视觉模型(Vision Transformer,ViT)[7],将三维图像数据切割成 patch 后转化为序列化数据,使用 Self-Attention 机制完全替代 CNN 后进行图像识别任务,实验表明在大规模数据集上 Transformer 模型的效果超过目前一些 SOTA 结果;麦吉尔大学的研究人员将卷积引入视觉 Transformer[8],提出包含新卷积 token 嵌入的 Transformer 层次结构,以及利用卷积投影的卷积 Transformer 块,将卷积神经网络的移位、缩放和失真不变性引入到 ViT 架构中,同时保持了 Transformers 的动态注意力、全局上下文和更好的泛化能力等优点;通过 Transformer 的自注意机制可以很好地模拟 Tokens Embedding 之间的全局交互,但是缺少在局部区域内进行信息交换的 locality 机制,Luc Van Gool[9] 团队通过在前馈网络中加入 Depth-Wise 卷积,将 locality 机制引入视觉 Transformer;谷歌大脑通过研究输入扰动的鲁棒性以及模型扰动的鲁棒性,发现在接受了足够数据量的预训练后,ViT 模型在各种扰动下都与 ResNet 具有相同的鲁棒性,同时 Transformer 对于移除几乎所有单层都具有鲁棒性[10];AIST 的研究人员通过实验发现,在没有自然图像和人工注释的标签的情况下,Transformer 的预训练仍然能够完成[11]。这些研究成果证明,在网络流量异常检测领域,基于卷积与 Transformer 的模型十分具有前景。

3 类别再平衡自训练算法

在类别不平衡的数据中,已标记和未标记的数据具有大致相同的不平衡分布,基于类别不平衡数据训练的带偏差模型,尾部类别尽管召回率很低,但精确率却出人意料地高。这表明许多尾部类别样本被预测为头部类别,模型在将样本分类为尾部类别时是保守的,但一旦做出这样的预测,就几乎可以确信它是正确的[12]。

本文利用尾部类别的高精确率来减轻低召回率造成的影响,提出类别再平衡自训练算法,其框架如图 1 所示。该算法通过自适应地从未标记的数据集中对伪标签为尾部类别的数据进行重新标记,并加入已标记的数据集中,以补充数据集中的尾部类别数据,然后重新训练该半监督模型。对于 L 类别分类任务,$X=\{(x_n,y_n),n\in(1,\cdots,N)\}$ 表示已标记数据集,其中 $x_n\in\mathbb{R}^d$ 表示训练样本,$y_n\in\{1,\cdots,L\}$ 表示与之相关的类别标签。N_l 表示 X 中类别 l 的训练样本数量,即 $\sum_{l=1}^{L}N_l=N$。假设这些类别按基数降序排列,即 $N_1\geqslant N_2\geqslant\cdots\geqslant N_L$。$X$ 边界类分布是偏斜的,即 $N_1\gg N_L$。通过不平衡比率表示类别不平衡的程度,即 $\gamma=\frac{N_1}{N_L}$。除了已标记数据集 X,数据集 $U=\{u_m\in\mathbb{R}^d:m\in(1,\cdots,M)\}$ 表示与数据集 X 具有相同数据分布的未标记数据集。标签分数 $\beta=\frac{N}{N+M}$ 表示已标记数据的百分比。给定类别不平衡的数据集 X 与数据集 U,类别再平衡自训练算法的目标就是学习一个类别均衡的分类器 $f:\mathbb{R}^d\to\{1,\cdots,L\}$。通过在分类器的预测 $\hat{y}_m=f(u_m)$ 中分配伪标签来使用未标记的数据[13-14],通过相应的伪标签在已标记数据和未标记数据上进行优化。伪标签的质量对模型的最终性能至关重要。因为分类器训练过程中提高了对所有类别的预测,然而,由于训练样本分布偏斜导致分类器本身就具有偏差,未标记数据的伪标签可能会使模型更加偏斜,进一步加剧类别不平衡问题,并导致对尾部类别的识别性能严重下降。自训练模型会经历多次迭代,模型在标记数据集上训练,模型的预测用于生成无标签数据 u_m 的伪标签 $\hat{y}_m$,在下一次迭代中,将生成的伪标签集合 $\hat{U}=\{(u_m,\hat{y}_m)\}_{m=1}^{M}$ 加入已标记数据集,即 $X'=X\cup\hat{U}$。多次迭代下 α 取值对模型准确率的影响如图 2 所示。

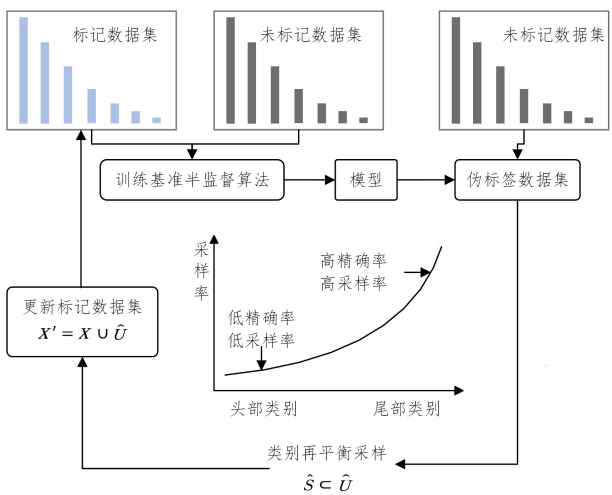

图 1 类别再平衡自训练算法

Fig. 1 Class-rebalancing self-training algorithm

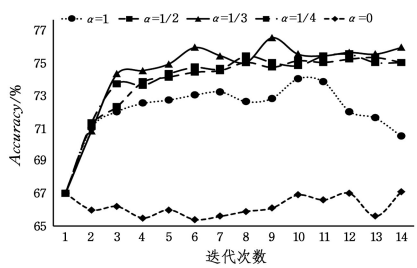

图 2 多次迭代下不同 α 取值的测试准确率

Fig. 2 Test accuracy over generations with different α

本文使用半监督学习算法,同时利用标记数据和未标记数据对模型进行训练,而不是仅仅使用标记数据,不再将伪标签集合 $\hat{U}$ 中的每一个样本都包括在内,而是将标记数据集扩展到选定的子集 $\hat{S}\subset\hat{U}$,即 $X'=X\cup\hat{S}$。本文依据一条类别再平衡规则选择 $\hat{S}$,即类别 l 的频率越低,被预测为类别 l 的无

标签样本被越多地加入伪标签集合$\hat{S}$。通过式(1)计算被预测为类别 l 的无标签样本加入伪标签集合$\hat{S}$的比例：

$$\mu_l=\left(\frac{N_{L+1-l}}{N_1}\right)^{\alpha} \tag{1}$$

其中，$\alpha\geqslant 0$ 用于调整采样率及控制集合 $\hat{S}$ 的大小，当 $\alpha=0$ 时，对于所有类别 l，其$\mu_l=1$，即所有的未标记样本都被保留，算法退化为传统的自训练算法。

4 网络流量异常检测模型

本文结合卷积与视觉 Transformer，将 locality 机制引入网络流量异常检测模型，图 3 给出了本文提出的网络流量异常检测模型网络G_{θ_G}结构，其包括 Feature Embedding 模块F_{θ_F}、位置编码P_{θ_P}、编码器E_{θ_E}、解码器D_{θ_D}以及一个用于特征合成的前馈网络S_{θ_S}。

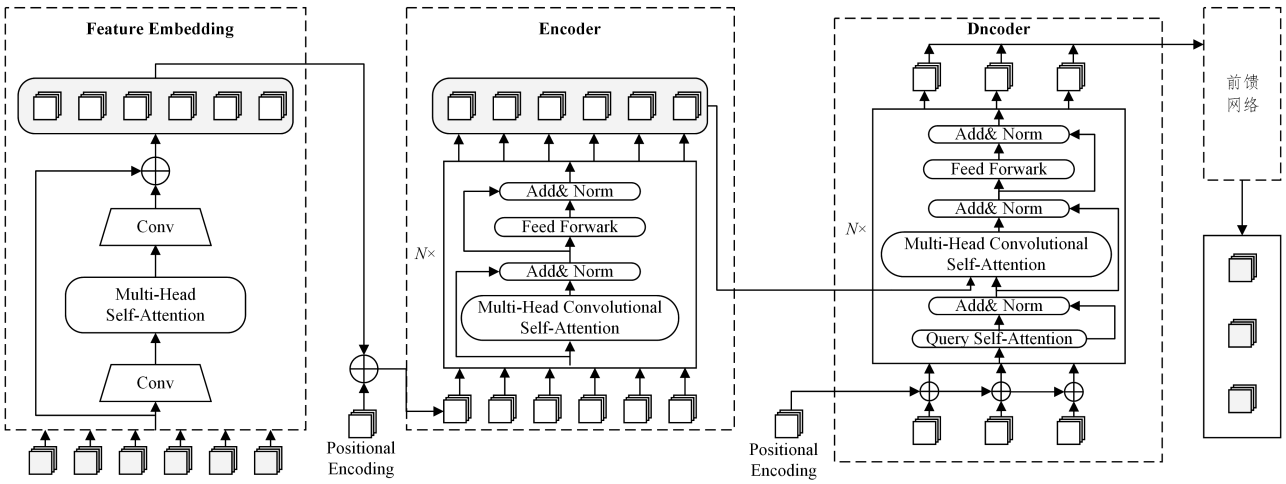

图 3 网络流量异常检测模型结构

Fig. 3 Convolutional Transformer networks G_{θ_G}

Feature Embedding 模块负责将输入的流量序列特征重新编码，将稀疏的高维特征向量转化为稠密的低维度特征向量，使得相似的特征在空间上的位置也是相近的。给定流量序列$X_i\in\mathbb{R}^{H\times W\times 3}$，特征图$H_i\in\mathbb{R}^{H\times W\times d_{model}}$可表示为：

$$J_i=F_{\theta_F}(X_i),i\in[1,n] \tag{2}$$

其中，所有输入流量不仅共享相同的 Embedding 网络架构F_{θ_F}，而且还共享参数θ_F。为了使模型充分利用流量序列的顺序，在编码器和解码器的每一层都添加了位置编码(Positional Encodings)。模型使用的位置编码是一个三维张量，它不同于原始 Transformer 架构中为矢量序列构建的位置编码。位置编码的尺寸与序列特征图相同，使得两者可以直接相加。本文使用不同频率的 sin 函数和 cos 函数：

$$Pos\_Map_{(p,(i,j,2k))}=\sin(n/10000^{2k/d_{\text{model}}}) \tag{3}$$

$$Pos\_Map_{(p,(i,j,2k+1))}=\cos(n/10000^{2k/d_{\text{model}}}) \tag{4}$$

其中，p 是位置标志，(i,j)代表特征空间位置，通道大小为 $2k$，即位置编码的每个维度都与正弦曲线相对应。波长呈现几何增长，范围从 2π 到 $10\,000*2\pi$。对于任意固定偏移 m，$Pos_Map_{(p+m)}$可以被表示为$Pos_Map_{(p)}$的线性函数。对于特定的 Feature Map J_i，其位置嵌入均可被视为：

$$Z_i=\mathcal{J}_i\oplus Pos\_Map_{(i)},i\in[1,n] \tag{5}$$

其中，$\oplus$运算符代表张量元素对位相加。将 ResNet 最后 3 个 bottleneck 模块的 3×3 卷积替换为全局的多头自注意力。ResNet 主干中的 c5 堆栈通常使用 3 个模块，用多头自注意力层替换每个模块的空间 3×3 卷积层。c5 中第一个模块使用 3×3 的卷积核，步幅为 2，对第一个 bottleneck 块使用 2×2 平均池化，步幅为 2。对于 1024×1024 分辨率的输入，c5 堆栈第一块中的多头自注意力层在 64×64 上运行，其余两块在 32×32 上运行。Feature Embedding 模块使用的多头自注意力层如图 4 所示。

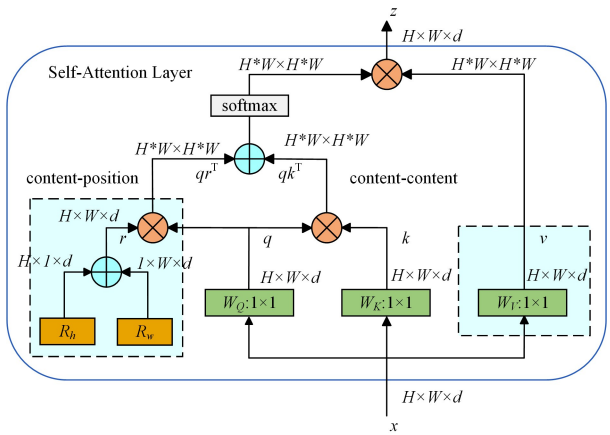

图 4 多头自注意力层

Fig. 4 Multi-head self-attention (MHSA) layer

全局自注意力在二维特征图上执行，该特征图将相对位置编码分别拆分为高度R_h和宽度R_w。logits 是 $qk^{\mathrm{T}}+q\,r^{\mathrm{T}}$，其中 q,k,r 表示 query，key 和 position encodings respectively，$\oplus$和$\otimes$分别表示对应元素求和与矩阵乘法，而 1×1 代表 point-wise convolution。

编码器模型由 N 个相同的层堆叠组成，每层包括多头卷积自注意力层和二维卷积前馈网络，两个子层采用残差连接。为了加快这些残差连接，模型中的所有子层产生相同尺寸的的输出，即$d_{\text{model}}=32$。对于位置编码之后的特征序列 $Z=\{Z_0,\cdots,Z_n\}\in\mathbb{R}^{H\times W\times d}$，其等价特征序列$\hat{Z}=\{\hat{Z}_0,\cdots,\hat{Z}_n\}\in\mathbb{R}^{H\times W\times d}$可被模型学习，编码操作可表示为：

$$\hat{Z}=E_{\theta_E}(Z) \tag{6}$$

卷积自注意力结构如图 5 所示。

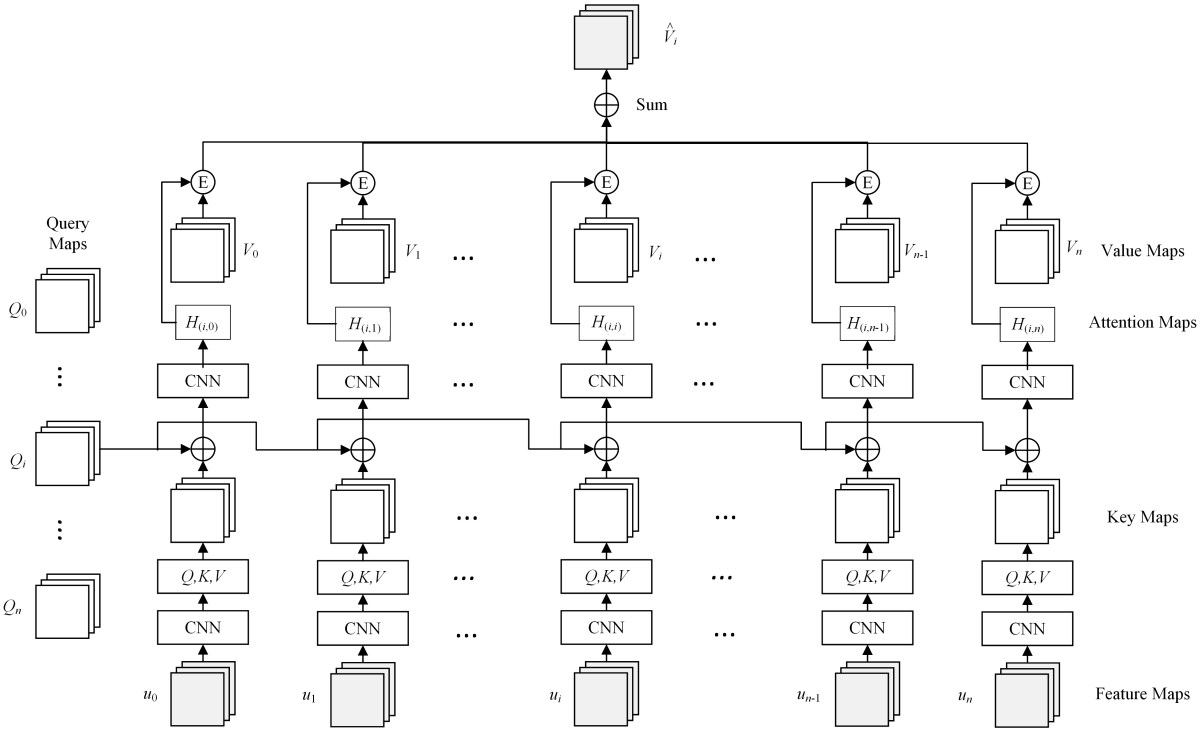

图 5 卷积自注意力结构

Fig. 5 Convolutional self-attention

卷积自注意力可描述为将一组 query map 和一组 key-value map 映射到输出，其中 query，key，value 和输出都是三维张量。给定输入特征图序列 $U=\{U_0,\cdots,U_n\}\in\mathbb{R}^{H\times W\times d_{\text{model}}}$，通过卷积子网络生成 query map 和配对的 key-value map，即 $U'=\{\{Q_0,K_0,V_0\},\cdots,\{Q_n,K_n,V_n\}\}\in\mathbb{R}^{H\times W\times 3}$。给定序列 U_i 的一组 $\{Q_i,K_i,V_i\}$，U_i 和 U_j 之间的 attention map $H_{(i,j)}\in\mathbb{R}^{H\times W\times 1}$ 可以通过使用 key map K_j 将一个兼容子网络 M_{θ_M} 应用到 query map Q_i 生成，如式(7)所示：

$$H_{(i,j)}=M_{\theta_M}(Q_i,K_j) \tag{7}$$

在得到全部相应的 attention map $H_{(i)}=\{H_{(i,1)},H_{(i,2)},\cdots,H_{(i,n)}\}\in\mathbb{R}^{H\times W\times 1}$ 后，对 $H_{(i)}$ 的第三维进行连续的操作，然后将 SoftMax 操作应用在 $H_{(i)}\in\mathbb{R}^{H\times W\times n}$ 的维度 $d=3$ 上：

$$H_{(i)}=\text{SoftMax}(H_{(i)})_d,d=3 \tag{8}$$

最后，由 attention map $H_{(i,j)}$ 与相应的 value map V_j 之积进行元素对应相加得到输出 $\hat{V}_i$，其表达式为：

$$\hat{V}_i=\sum_{j=1}^{n}H_{(i,j)}V_j \tag{9}$$

解码器除了与编码器中相同的两个子层外，还插入了一个 query self-attention 附加层。对于一个 query 序列 $Q=\{Q_0,\cdots,Q_n\}\in\mathbb{R}^{H\times W\times d}$，解码过程如式(10)所示：

$$\hat{Q}=D_{\theta_D}(\hat{Z},Q) \tag{10}$$

编码和解码过程都是并行进行的。前馈网络通过均方误差指导网络的优化过程，如式(11)所示：

$$L_{G_{\theta_G}}=\frac{1}{N}\sum_{i=1}^{N}\|\hat{x}_i-y_i\|_2 \tag{11}$$

其中，N 是训练样本数，$\hat{x}_i$ 和 y_i 是数据预测值与真实值。

5 实验结果及其分析

本文使用 CIC-IDS-2017[15] 数据集，60% 作为训练集，20% 作为交叉验证集，20% 作为测试集，采用准确率 Accuracy、精确率 Precision、召回率 Recall 和 F-score 作为模型性能评估的指标：

$$Accuracy=\frac{TP+FN}{TP+TN+FP+FN} \tag{12}$$

$$Precision=\frac{TP}{TP+FP} \tag{13}$$

$$Recall=\frac{TP}{TP+FN} \tag{14}$$

$$F_1=2*\frac{Precision*Recall}{Precision+Recall} \tag{15}$$

CIC-IDS-2017 数据集中有 9 种异常流量和 1 种正常流量，其中正常流量占据流量总数的绝大部分，其总量远远大于异常流量的总量；在异常流量中，DoS 攻击和 Port Scan 等又占据大部分的比例，其总量远远大于 Heartbleed 等异常流量的总量。这就导致该数据集的类别分布有着非常巨大的不平衡，这种类别间的巨大差异会使训练出的模型的预测值明显偏向头部类别，而远离尾部类别。同时，模型对尾部类别的预测准确率往往十分高，接近于 100%，这看似与模型偏差的结果相矛盾，实则是由于类别间巨大的数量差异致使模型对于尾部类别的预测对准确率的影响微乎其微。因此在评价模型好坏时，使用准确率往往不能反映客观事实。而精确率与召回率则避免了因数量的巨大差异造成的影响，可以较为客观地反映模型的性能。因此本文主要使用精确率与召回率评价模型。

如表 1 所列为本模型是否使用类别再平衡自训练算法的情况下的检测的结果。数据显示，在使用类别再平衡自训练算法后，模型对头部类别的预测精确率均有不同幅度的提升。对于头部类别的预测误检率降低，对于尾部类别的预测精确率和准确率都有非常大的提升。头部类别的 *F-score* 基本持

平，这说明该算法不会影响头部类别较高的预测性能，尾部类别 *F-score* 均有不同幅度的提升，表明算法对于尾部类别的识别具有积极影响，这些数据充分说明算法对于解决类别不平衡问题的有效性。

表 1 使用类别再平衡自训练算法前后的模型性能

Table 1 Model performance before and after using CReST

（单位：%）

	使用 CReST	*Accuracy*	*Precision*	*Recall*	*F-score*
Benign	√	99.89	**99.91**	99.92	**99.92**
	×	**99.90**	99.88	**99.95**	99.91
FTP Parator	√	**99.96**	**92.16**	**85.68**	**88.80**
	×	99.89	88.97	77.79	83.01
SSH Parator	√	**99.95**	**91.76**	**85.51**	**88.54**
	×	98.27	87.78	71.66	78.91
Heart Bleed	√	99.99	**96.77**	**58.82**	**73.17**
	×	99.99	90.91	39.22	54.79
Web Attack	√	99.99	**95.71**	**72.07**	**82.22**
	×	99.99	91.20	50.83	65.28
Infiltration	√	99.99	**92.25**	**45.83**	**61.24**
	×	99.99	89.71	20.83	33.81
Botnet	√	99.99	**95.69**	**66.67**	**78.63**
	×	99.99	89.27	40.22	55.46
Port Scans	√	**99.17**	97.76	**98.41**	**98.08**
	×	98.21	97.76	98.40	98.07
DoS	√	98.93	**97.97**	98.57	**98.33**
	×	99.93	97.91	**98.60**	98.25
DDoS	√	99.89	**97.80**	**99.37**	**98.58**
	×	**98.90**	97.67	99.31	98.53

从图 6 可知，在使用类别再平衡自训练算法前后，头部类别的 *F-score* 基本持平，这说明该算法不会影响头部类别较高的预测性能。尾部类别 *F-score* 均有不同幅度的提升，其中 Infiltration 类别的 *F-score* 提升了 27.43%，提升幅度为81.13%，Botnet 类别的 *F-score* 提升了 23.71%，提升幅度为42.75%。这表明类别再平衡自训练算法对于尾部类别的识别具有非常大的提升，再次印证了本算法的有效性。

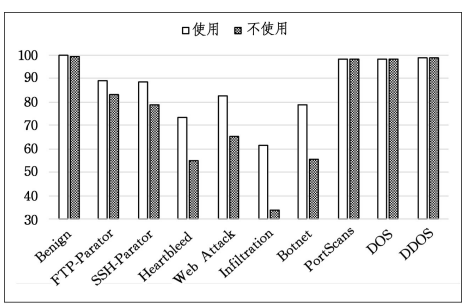

图 6 使用 CReST 前后 *F-score* 的对比

Fig. 6 Comparison of *F-score* beform and after using CReST

其他模型检测的性能如表 2 所列，对于 DoS 和 Probe 攻击，本模型性能表现优于其他模型，在尾部类别的检测中，其检测精确率与召回率几乎与头部类别齐平。

表 2 各模型在 CIC-IDS-2017 数据集上检测的性能

Table 2 Model performance tested on the CIC-IDS-2017 dataset

（单位：%）

	Model	*Accuracy*	*Precision*	*Recall*	*F-score*
Normal	CNN	99.11	99.19	98.76	98.98
	SRDLM	91.13	95.37	90.53	92.88
	Stacked-LSTM	**99.99**	**99.94**	**99.97**	**99.96**
	本模型	99.89	99.91	99.92	99.92
DoS	CNN	83.90	91.60	92.26	91.93
	SRDLM	95.50	95.51	93.87	94.68
	Stacked-LSTM	87.68	92.04	85.78	88.80
	本模型	**98.93**	**97.97**	**98.57**	**98.33**
U2R	CNN	99.81	32.80	14.33	19.95
	SRDLM	94.80	96.92	72.95	83.24
	Stacked-LSTM	91.95	75.87	75.52	75.70
	本模型	**99.98**	**97.67**	**99.31**	**98.53**
R2L	CNN	99.86	21.30	13.47	16.50
	SRDLM	94.80	60.45	46.80	52.76
	Stacked-LSTM	91.77	75.08	75.45	75.26
	本模型	**99.99**	**97.69**	**96.67**	**97.18**
Probe	CNN	80.63	69.10	61.77	65.38
	SRDLM	83.28	96.92	72.95	83.24
	Stacked-LSTM	92.90	80.98	74.85	77.79
	本模型	**99.17**	**97.76**	**98.41**	**98.08**

其他模型在尾部类别上的检测结果则表现不佳，CNN 模型对于 U2R 类别检测的 *F-score* 为 19.95%，在 R2L 类别检测的 *F-score* 为 16.50%，这导致其整体的检测性能变得非常差。其他模型在尾部类别的检测结果同样较差，这体现了本模型的检测效果。各模型的 *F-score* 如图 7 所示。

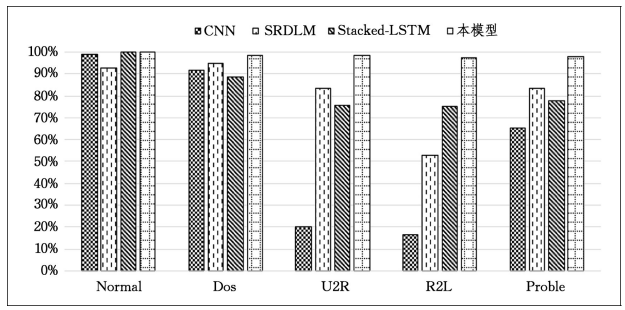

图 7 各模型的 *F-score* 对比

Fig. 7 Comparison of *F-score* of each model

Stacked-LSTM 在头部类别的检测中 *F-score* 最高能达到 99.96%，而在尾部类别的检测中只能达到 75.70% 和 75.26%，这反映了类别不平衡对该模型的巨大负面影响。同样地，CNN 模型在尾部类别的检测结果堪称灾难性的，其对于 U2R 类别检测的 *F-score* 只有 19.95%，在 R2L 类别检测的 *F-score* 只有 16.50%，这导致其整体的多分类检测性能变得非常差。其他模型在尾部类别的检测结果同样不佳，这也体现了本模型检测效果的优越性。同时，为了充分验证类别再平衡自训练算法的有效性，避免其存在模型差异而降低实用性，本文设计了一组对照实验，将类别再平衡算法应用于各种其他模型，其使用前后对尾部类别检测的 *F-score* 对比如图 8 所示。

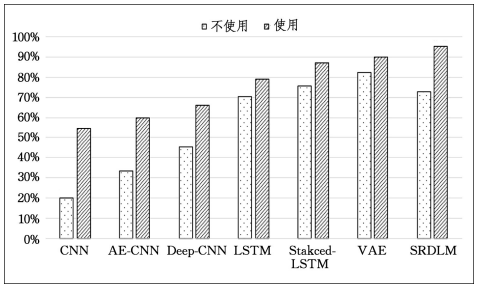

图 8 各模型使用 CReST 前后的 *F-score* 对比

Fig. 8 *F-score* of other models before and after using CReST

在应用了类别再平衡自训练算法后，各种模型对于尾部类别的检测 F-score 均有明显提升，如 CNN 模型使用该算法后对尾部类别的检测 *F-score* 提升了 34.52%，提升幅度为 173%，LSTM 提升了 8.47%，表明该算法不仅对于尾部类别检测 *F-score* 较低的模型有效，而且对于尾部类别检测 *F-score* 较高的模型也同样有效。

在 Feature Embedding 模块处理之前，数据集中各种类别的分布是混乱的，不同类别的数据常常出现在空间的相同或相近位置，而同类别的数据又往往距离非常远，如图 9 所示。这对于模型挖掘数据特征的联系非常不利，便其难以根据数据的空间位置判断其相似性。

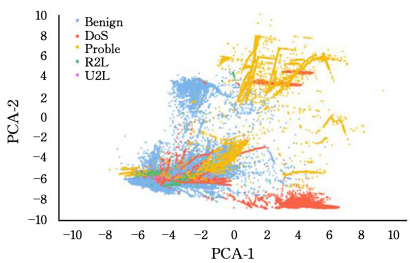

图 9 原始数据特征

Fig. 9 Original data features

经过 Feature Embedding 模块处理后的数据如图 10 所示，不同类别的头部数据相互聚集，在空间上出于更近的位置，形成明显的聚类，不同类别的头部数据则相互远离，在空间上处于更远的位置，形成明显的隔断。尾部数据则散落于空间各处，没有形成聚类，或聚类现象不明显，这是由于尾部数据类别过少，模型难以挖掘其内部特征关联，因此其在空间中的位置也难以实现聚集。

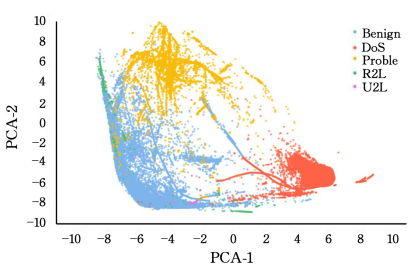

图 10 Feature Embedding 处理后的数据特征

Fig. 10 Data features after Feature Embedding

结束语 随着时代的不断发展，互联网的蓬勃生命力会延续很长时间，有网络存在的地方就会不断产生新的安全问题，这些问题不会随着技术的进步而消失，而是会长期与互联网共存，在不断的对抗中，网络攻击技术会持续更新，而对抗攻击的网络安全检测技术也会随之成长，因此，网络流量异常检测领域仍然有十分巨大的发展空间。同时，随着人工智能技术的兴起，在可预见的未来，这项技术会不断改变我们的生活，逐渐渗透到我们生活的方方面面，在网络安全领域也一定会占据一席之地。人工智能的优势使得我们看待世界的方式发生了改变，因此各行各业都应适应这种变化，积极拥抱变化，不断进步，网络流量异常检测领域也不例外。针对网络环境加速复杂化的现实，人工智能技术拥有比人类更好的适应能力、更快的反应速度、更灵敏的检测技术、更高效的解决方式、更经济的策略、更安全的行为，因此，利用人工智能技术发展网络流量异常检测一定是未来网络安全领域的主要发展方向。

针对这种环境，本文提出了一种新型网络流量异常检测模型，以深度学习中神经网络与注意力为基础，对网络流量特征进行深度提取并分类检测，其效率与性能皆优于以往基于规则的检测方案，并拥有广阔的应用前景。面对海量无标签网络数据，该模型依靠其自身强大的学习能力，能够在保证准确率的情况下高效检测异常流量，通过卷积自注意力提取网络流量的深层远距离关联特征，做到没有漏网之鱼，达到良好的检测效果，特别是在小样本数据上的检测能力堪称一绝，在大样本的数据面前检测能力能够保持较高水准。

总的来说，本文分析了以往技术的不足，从深层次剖析各种不足存在的原因，并对症下药，提出解决这些问题的有效算法和模型，将深度学习技术巧妙融入网络安全领域，提高了网络流量异常检测的能力，这为维护网络安全、保障正常用户合法权益做出了贡献。

参 考 文 献

[1] ZHOU Y, LI J. Research of Network Traffic Anomaly Detection Model Based on Multilevel Autoregression[C]// 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT). 2019: 380-384.

[2] KONG B, LIU Z, ZHOU G, et al. A Method of Detecting the Abnormal Encrypted Traffic Based on Machine Learning and Behavior Characteristics[C]// Proceedings of the 2019 the 9th International Conference on Communication and Network Security(ICCNS 2019). New York, NY, USA: Association for Computing Machinery, 2019: 47-50.

[3] VERMA A K, KAUSHIK P, SHRIVASTAVA G. A Network Intrusion Detection Approach Using Variant of Convolution Neural Network[C]// 2019 International Conference on Communication and Electronics Systems (ICCES). 2019: 409-416.

[4] CHIA Y K, WITTEVEEN S, ANDREWS M. Transformer to CNN: Label-Scarce Distillation for Efficient Text Classification [J]. arXiv: 1909.03508.

[5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]// Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017: 5998-6008.

[6] KHAN S, NASEER M, HAYAT M, et al. Transformers in Vision: A Survey[J]. arXiv: 2101.01169.

[7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image is Worth 16x16 Words: Transformers for Image Recogni-

tion at Scale[C]// International Conference on Learning Representations(ICLR 2021). 2020.
[8] WU H, XIAO B, CODELLA N, et al. CvT: Introducing Convolutions to Vision Transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021.
[9] LI Y, ZHANG K, CAO J, et al. LocalViT: Bringing Locality to Vision Transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2021.
[10] BHOJANAPALLI S, CHAKRABARTI A, GLASNER D, et al. Understanding Robustness of Transformers for Image Classification[C]// Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition (CVPR). 2021.
[11] NAKASHIMA K, KATAOKA H, MATSUMOTO A, et al. Can Vision Transformers Learn without Natural Images? [J]. arXiv: 2103. 13023.
[12] JAMAL M A, BROWN M, YANG M H, et al. Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition From a Domain Adaptation Perspective[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
[13] SOHN K, BERTHELOT D, CARLINI N, et al. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence[C]// Advances in Neural Information Processing Systems (NeurIPS). Curran Associates, Inc. , 2020: 596-608.
[14] XIE Q, LUONG M T, HOVY E, et al. Self-TrainingWith Noisy Student Improves ImageNet Classification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
[15] KURNIABUDI, STIAWAN D, DARMAWIJOY O, et al. CIC-IDS-2017 Dataset Feature Analysis with Information Gain for Anomaly Detection[J]. IEEE Access, 2020, 8: 132911-132921.

YANG Yue-lin, born in 1994, master, engineer. His main research interests include network security and deep learning.

(上接第 532 页)

[11] YOON K, PARK D, YIM Y, et al. Security authentication system using encrypted channel on uav network[C]// 2017 First IEEE International Conference on Robotic Computing (IRC). IEEE, 2017: 393-398.
[12] GORENFLO C, LEE S, GOLAB L, et al. Fastfabric: Scaling hyperledger fabric to 20 000 transactions per second[C]// 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE, 2019: 455-463.
[13] H CHAO, MAHESHWARI A, SUDARSANAN V, et al. UAV traffic information exchange network[OL]. http://dx. doi. org/10. 252018: 14/6. 2018-3347.
[14] DORRI A, KANHERE S S, JURDAK R, et al. Blockchain for IoT security and privacy: The case study of a smart home[C]// 2017 IEEE International Conference on Pervasive Computing and Communications Workshops. PerComWorkshops, 2017: 618-623.
[15] LIANG X, ZHAO J, SHETTY S, et al. Towards data assurance and resilience in IoT using blockchain[C]// MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM). IEEE, 2017: 261-266.
[16] BOZIC N, PUJOLLE G, SECCI S, A tutorial on blockchain and applications to secure network control-planes[C]// 2016 3rd Smart Cloud Networks Systems. SCNS, 2016: 1-8.
[17] VADLAMANI S, EKSIOGLU B, NANDI H M, et al. Jamming attacks on wireless networks: A taxonomic survey[J]. Int. J. Prod. Econ. , 2016(172): 76-94.
[18] ZHOU Y. Summary of UAV Ground Station Development [J]. Aviation Electronics Technology, 2010, 41(1): 1-6.
[19] CAMPION M, RANGANATHAN P, FARUQUE S. UAV swarm communication and control architectures: a review[J]. Journal of Unmanned Vehicle Systems, 2018, 7(2): 93-106.
[20] MCGOVERN S. Blockchain for Unmanned Aircraft Systems [R]. John A. Volpe National Transportation Systems Center (US), 2020.
[21] JENSEN I J, SELVARAJ D F, RANGANATHAN P. Blockchain Technology for Networked Swarms of Unmanned Aerial Vehicles (UAVs)[C]// 2019 IEEE 20th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2019: 1-7.

WANG Yu-chen, born in 1997, postgraduate. Her main research interests include blockchain and recommendation system.

XU Li-zhen, born in 1963, professor. His main research interests include database technology, software engineering and enterprise information.