DOI:10.11918/202306042

融合稀疏自注意力机制的增量入侵检测模型

金志刚[1], 周峻毅[1,2],武晓栋[1],刘 凯[1]

(1.天津大学 电气自动化与信息工程学院,天津 300072; 2.天津大学 未来技术学院,天津 300072)

摘 要:传统的基于自注意力的入侵检测模型在注意力值计算中存在时间复杂度较高的问题,且多数模型面向静态网络环境。针对上述问题,提出融合稀疏自注意力机制的增量入侵检测模型。首先,引入稀疏度度量公式以降低时间复杂度,在不影响模型检测性能的前提下减轻模型计算压力;其次,构建动态示例存储器,以极小内存开销缓解增量学习中的概念漂移现象;最后,设计类别平衡损失函数,无需动态调整模型即可增强旧类别样本学习能力。推导与实验结果证明:稀疏自注意力机制的时间复杂度更低、分类效果更优;对比其他方案,所提增量学习机制的旧知识记忆能力更强,该入侵检测模型在现代网络环境中有着较好的应用前景。

关键词:入侵检测;自注意力机制;增量学习;深度学习;灾难性遗忘

中图分类号:TP393.08 **文献标志码**:A **文章编号**:0367-6234(2026)03-0020-08

Incremental intrusion detection model incorporating sparse self-attention mechanism

JIN Zhigang[1], ZHOU Junyi[1,2], WU Xiaodong[1], LIU Kai[1]

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;
2. School of Future Technology, Tianjin University, Tianjin 300072, China)

Abstract: Traditional self-attention-based intrusion detection models have high time complexity in the calculation of attention values, and most intrusion detection models are oriented to static network environments. To address the above problems, we proposed an incremental intrusion detection model incorporating a sparse self-attention mechanism. First, we introduced a sparsity metric formula to reduce the time complexity, so as to alleviate the computational pressure of the model without affecting the detection performance of the model; Second, we constructed a dynamic example memory to alleviate the concept drift phenomenon of the model in incremental learning at the cost of a very small amount of memory space; Finally, we designed a category-balanced loss function, which is capable of enhancing the learning ability of the model for old-category samples without dynamically adjusting the model. Derivation and experiments prove that the sparse self-attention mechanism has lower time complexity and better classification effect. Compared with other schemes, the incremental learning mechanism shows a stronger ability to memorize old knowledge. The intrusion detection model has a better application prospect in the modern network environment.

Keywords: intrusion detection; self-attention mechanism; incremental learning; deep learning; catastrophic forgetting

随着信息化社会的快速发展,网络安全问题日益突出。入侵检测技术作为保障网络安全的核心技术之一,始终是网络安全领域的研究热点[1]。

传统基于误用的入侵检测系统因其难以适应现代网络攻击手段、规则库维护成本较高以及实时监测灵敏度较低等缺陷,已经逐渐淡出主流研究视野。近年来,随着硬件算力的大突破和深度学习算法的快速发展,基于深度学习的入侵检测技术凭借异常流量识别率高、分类速度快等优势成为主流研究方向,并获得了一系列研究成果[2-4]。然而,当前入侵检测技术仍然面临两大关键问题。

首先,现代网络流量数据具有显著的相关性特征。例如,电力信息物理系统中存在丰富的相关相似性[5],对于分布式拒绝服务(distributed denial of service,DDoS)攻击,需联合一段时间内的流量数据才能准确识别,否则极有可能将其判断为正常访问

收稿日期:2023-06-09;**录用日期**:2023-08-21;**网络首发日期**:2025-08-01
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250731.1800.004
基金项目:国家自然科学基金(52171337)
作者简介:金志刚(1972—),男,教授,博士生导师;周峻毅(1998—),男,硕士研究生;武晓栋(1996—),男,博士研究生
通信作者:金志刚,zgjin@tju.edu.cn

行为[6]。这表明流量相关性应作为入侵检测模型设计的关键考量因素。其次,目前大部分基于深度学习的入侵检测研究都是面向静态网络流量数据,即假设训练数据静止不变。但在实际应用中,网络流量数据会随着时间动态变化,静态训练数据设定导致模型难以适应动态环境[7-8]。

目前,部分研究已围绕流量数据相关性展开,旨在提高入侵检测模型的分类准确率。Kasongo 等[9]基于入侵检测数据的时序相关性,提出融合深度长短期记忆网络的入侵检测系统,利用信息增益的技术实现特征抽取,减少数据冗余,提升检测准确率; Sethi 等[10]在分布式代理平台背景下,设计基于记忆强化学习的入侵检测系统,利用注意力机制高效且精准地分类现代网络攻击手段; Khan 等[11]融合卷积神经网络和循环神经网络,捕捉流量特征的空间依赖性以及时序相关性; Lan 等[12]提出基于分层注意力的自适应无监督入侵检测模型,通过注意力机制和联合损失函数来强化正常流量的紧凑相关性学习,提升良性样本识别能力,并在此基础上训练单分类支持向量机用于检测未知攻击样本,降低模型在新场景下的假阳性率; Fu 等[13]结合通道注意力机制和长短期记忆网络,通道注意力机制可以动态调节训练过程中的通道权重,长短期记忆网络可以有效学习流量时序相关性,增强对长时间依赖关系的学习能力;石磊等[14]利用自注意力机制对流量数据进行特征提取,拟合数据特征的全局依赖性,再利用双向长短期记忆网络学习数据间的时序相关性,提高了入侵检测时准确率。

此外,国内外学者也开展了增量式入侵检测系统的相关研究。Mahdavi 等[15]提出增量入侵检测系统新框架,基于增量聚类算法实现无先验知识学习,并融合迁移学习方法,使模型在迭代过程中通过迁移学习获取增量知识,从而完成增量学习; Wang 等[16]提出基于极限学习机的自适应增量入侵检测策略,在学习过程中,模型可自适应添加隐藏神经元,在动态的数据流中寻找适配新旧知识的最优模型,实验证明该策略以较低时间成本取得了较为优异的性能;刘强等[17]提出基于堆叠稀疏自编码器(SSAE)和自组织增量神经网络(SOINN)的物联网入侵检测方法,利用 SSAE 进行特征抽取,当新类别来临时,SOINN 将自适应生成局部拓扑结构,对已有的 SOINN 结构施加约束,从而保证对旧类别的记忆。

通过文献调研发现,一方面,现有多数入侵检测研究对数据相关性的关注度不高,且单独使用自注意力机制存在计算复杂度较高的问题;另一方面,多数增量入侵检测机制采用动态调整模型方案,随着任务迭代,模型规模也会不断扩大,这会加大部署端的存储压力,且动态调整过程计算复杂,对部署端的算力要求较高。

综上,本文提出融合稀疏自注意力机制的增量入侵检测模型。稀疏自注意力不仅可以保证充分拟合数据相关性以实现对流量的精准分类,还降低了模型在计算自注意力值时的时间复杂度。同时,通过动态示例存储已学习的旧类别样本,在占用极小存储资源的基础上,配合类别平衡损失函数对模型反向传播进行正则化,以解决灾难性遗忘问题。

1　稀疏自注意力机制

1.1　机制原理

自注意力机制由 Vaswani 等[18]提出,近年来因其出色的流量数据时序相关性捕捉能力,已经被广泛应用于入侵检测领域并取得了较好的成果[19-21]。然而,传统的自注意力机制通过规范点积计算注意力值,导致设备计算压力过大,这与入侵检测设备的资源约束相悖[22]。为此,引入稀疏自注意力机制,在保证相关性学习效率的同时降低时间复杂度。

在数据输入稀疏自注意力层之前,需对无时间特征的入侵检测数据进行位置编码。设原始空间数据为 $X \in \boldsymbol{R}^{m\times n}$,其中 m 为原始空间数据的样本数,n 为原始空间中数据样本的特征维度,$x_i \in \boldsymbol{R}^n$,($i=1,2,\cdots,m$),则位置编码层 P 的编码方式为

$$\begin{cases} \mathrm{PE}_{(\mathrm{pos},2j)} = \sin(\mathrm{pos}/10000^{2j/n}) \\ \mathrm{PE}_{(\mathrm{pos},2j+1)} = \cos(\mathrm{pos}/10000^{2j/n}) \end{cases} \tag{1}$$

式中:pos 表示当前样本 x_i 在所有输入样本中的相对位置;j 为当前样本特征的相对位置。将式(1)的结果与预处理后的入侵检测数据相加,得到融合相对位置信息的流量数据。

传统自注意力机制的注意力计算函数为

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

式中:$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为查询矩阵、键矩阵、值矩阵;d_k 为键向量维度。

网络流量数据作为时序序列,其自注意力分布符合长尾分布,即当前流量数据 x_i 仅与其他少部分流量数据存在强相关性。因此,计算当前流量数据 x_i 与其他数据的自注意力值时,无需与注意力域内所有数据进行规范点积。因此,受文献[23]启发,引入稀疏度度量公式,对查询矩阵中起关键作用的

查询向量进行筛选,具体如式(3)所示:

$$M(\boldsymbol{q}_i,\boldsymbol{K}) = \ln\left(\sum_{j=1}^{m}\exp\left(\frac{\boldsymbol{q}_i\boldsymbol{k}_j^{\mathrm{T}}}{\sqrt{d_k}}\right)\right) - \frac{1}{m}\sum_{j=1}^{m}\frac{\boldsymbol{q}_i\boldsymbol{k}_j^{\mathrm{T}}}{\sqrt{d_k}} \tag{3}$$

式中:$\boldsymbol{q}_i$ 代表查询矩阵 $\boldsymbol{Q}$ 中第 i 个查询向量;$\boldsymbol{k}_j^{\mathrm{T}}$ 代表键矩阵 $\boldsymbol{K}$ 中第 j 键向量的转置。式(3)中前半部分是查询向量 $\boldsymbol{q}_i$ 与所有键向量的对数求和指数(Log-Sum-Exp),后半部分是查询向量 $\boldsymbol{q}_i$ 与所有键向量的算术平均。式(3)的值越大,说明查询向量 $\boldsymbol{q}_i$ 的自注意力值分布与均匀分布差异越大,即查询向量 $\boldsymbol{q}_i$ 在计算自注意力值时更"活跃"。设置采样数,挑选稀疏度值最靠前的 u 个查询向量参与自注意力计算,如式(4)所示:

$$u = c \times \ln m \tag{4}$$

式中 c 为超参数。计算注意力值时只挑选稀疏度靠前的 u 个查询向量参与计算,为了保持查询矩阵的维度不变,将剩余的 $m-u$ 个查询向量直接替换为值向量矩阵 $\boldsymbol{V}$ 的均值 $\mathrm{mean}(v_i)$ 形成新的稀疏查询矩阵 $\overline{\boldsymbol{Q}}$。稀疏自注意力机制的计算公式为

$$\text{Sparse-Attention}(\overline{\boldsymbol{Q}},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\overline{\boldsymbol{Q}}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{5}$$

经过稀疏自注意力机制优化后,自注意力层的时间复杂度由 $O(m^2)$ 降至 $O(m\ln m)$,在充分学习数据相关性的同时降低了时间复杂度,直接减少了模型的训练时间,提高了入侵检测模型的学习效率。

2 增量学习策略

2.1 动态示例存储器

基于回放的增量学习策略可有效增强模型对旧类别数据的记忆,避免灾难性遗忘,且简单高效。传统的基于回放的增量学习策略在当前任务 T^n 中存储部分训练数据,当下一个任务 T^{n+1} 到来时,将存储的旧训练数据与新训练数据共同作为训练集来训练模型,但该策略忽略了入侵检测场景中各类别流量数据的类别不平衡问题[24]。为解决上述问题,构建动态示例存储器 M,使存储器中旧类别样本分布更趋向于真实分布。设入侵检测模型接收的任务流为 $T=\{T^n\}_{n=1}^{N}$,N 代表任务总数,记录并保存每个任务的样本总数量 $\{\mathrm{Size}[T^n]\}_{n=1}^{N}$ 以及任务中各类别样本的数量 $\{\mathrm{Size}[K^{n,j}]\}_{j=1}^{k}$,$k$ 代表每个任务中包含的类别总数。当入侵检测模型处理当前任务 T^n 时,动态示例存储器会按照式(6)更新每个旧类别的存储空间:

$$m^{i,j} = \frac{m\ \mathrm{Size}[K^{i,j}]}{\sum_{i=1}^{n-1}\mathrm{Size}[T^i]} \tag{6}$$

式中:m 代表示例存储器的总容量大小;$m^{i,j}$ 代表第 i 个任务中第 j 类的示例存储器容量。旧类别样本的挑选方法参考文献[25]。随着任务不断更新,动态示例存储器可以保证旧类别样本分布与入侵检测模型已经接收的旧类别训练样本分布相同。动态示例存储器会引导模型学习真实的流量数据分布,以强化对旧类别知识的记忆效果。

2.2 类别平衡损失函数

随着任务的不断迭代,动态示例存储器中的各旧类别样本数量逐渐减少,导致模型对旧类别样本的记忆不断衰退。扩大存储器的总容量虽可增强记忆,但入侵检测设备资源受限,无法承担过大的存储开销。因此构建类别平衡损失函数,在模型的训练过程中对新旧类别样本进行重加权,增强模型对旧样本的记忆,缓解灾难性遗忘。

具体来说,对于入侵检测模型正在处理的任务 T^n 中一个小批次数据 $\{\boldsymbol{X}_b^n,\boldsymbol{Y}_b^n\}=\{x_i^n,y_i^n\}_{i=1}^{b}\subset\{X^n,Y^n\}$(其中 b 是批次大小(batch-size),y_i^n 代表第 i 个样本的真实标签),首先通过公式(7)获取 T^n 的模型 θ^n 的最后一层神经网络的第 y_i^n 个神经元 $N_{y_i^n}^n$ 的梯度信息:

$$g_i^n = \frac{\partial \mathrm{CE}(\varphi^n(x_i^n,\theta^n),y_i^n)}{\partial N_{y_i^n}^n} = \varphi^n(x_i^n,\theta^n)_{y_i^n} - 1 \tag{7}$$

式中:$\mathrm{CE}(\cdot)$ 为交叉熵损失函数;$\varphi^n(x_i^n,\theta^n)$ 为模型对输入 x_i^n 经 θ^n 预测输出的概率分布;$\varphi^n(x_i^n,\theta_i^n)_{y_i^n}$ 代表经 θ^n 预测,样本 x_i^n 被分为 y_i^n 的 softmax 概率。为了放缓入侵检测模型对新类别数据的学习速度,加强对旧类别记忆,通过公式(8)和公式(9)计算新、旧类别的梯度信息均值,并通过公式(10)重新定义新类别与旧类别在损失函数中的权重,实现损失函数重加权。

$$g_{\mathrm{new}} = \frac{\sum_{i=1}^{b}|g_i^n|\cdot\psi(y_i^n\in K^{\mathrm{new}})}{\sum_{i=1}^{b}\psi(y_i^n\in K^{\mathrm{new}})} \tag{8}$$

$$g_{\mathrm{old}} = \frac{\sum_{i=1}^{b}|g_i^n|\cdot\psi(y_i^n\in K^{\mathrm{old}})}{\sum_{i=1}^{b}\psi(y_i^n\in K^{\mathrm{old}})} \tag{9}$$

$$L_{\mathrm{CB}} = \frac{1}{b}\sum_{i=1}^{b}\frac{|g_i^n|}{g_{m,i}}\cdot\mathrm{CE}(\varphi^n(x_i^n,\theta^n),y_i^n) \tag{10}$$

式中,K^{new} 为当前任务的新类别集合;K^{old} 为旧类别集合;$\psi(\cdot)$ 是条件判断函数,当函数中的变量判断

为真时 $\psi(\cdot)=1$,否则 $\psi(\cdot)=0$;$g_{m,i}=\psi(y_i^n\in K^{\text{new}})\cdot g_{\text{new}}+\psi(y_i^n\in K^{\text{old}})\cdot g_{\text{old}}$。利用权重$\frac{|g_i^n|}{g_{m,i}}$调节每个样本在损失函数中的贡献,通常,旧类别样本占比小时,权重值较大;新类别样本占比大时,则权重值较小。在动态示例存储器的基础上,类别平衡损失函数可以主动平衡新类别和旧类别的学习步调,进一步缓解入侵检测模型的灾难性遗忘。

3　模型系统架构

3.1　模型结构及工作流程

本文提出的入侵检测模型架构如图 1 所示。

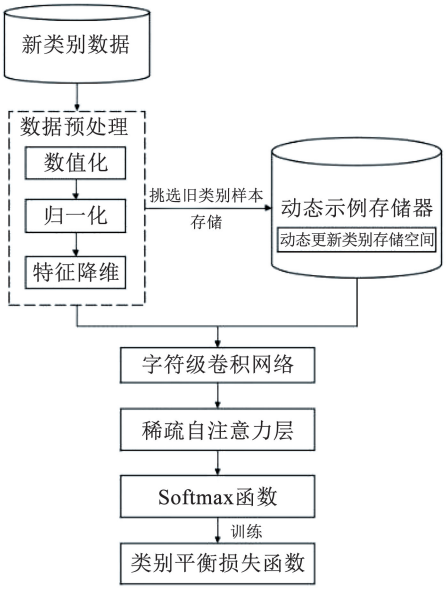

图 1　系统结构

Fig. 1　System structure

工作流程如下:

1)新类别数据预处理:利用独热编码(One-Hot Encoding)将数据中的非数值化特征转换为数值特征;通过式(11)对数据进行归一化处理,避免模型在训练过程中不收敛;利用特征降维的方式(如主成分分析)浓缩数据特征,减少数据的冗余度。

$$x^*=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{11}$$

2)样本筛选与存储器更新:从数据预处理后的新类别数据中挑选代表性样本存储到动态示例存储器当中,按照公式(6)对存储器中各类别样本数量进行调整。

3)特征学习:将新类别训练样本和动态示例存储器中的旧类别样本(不含本次任务新增样本)共同输入字符级卷积网络,进一步学习特征数据,为稀疏自注意力层提供更优输入。

4) 相关性拟合:将字符级卷积网络的输出作为稀疏自注意力层的输入,对入侵检测数据间的潜在相关性进行学习和拟合。

5) 分类预测:利用 Softmax 函数将输出转化为类别概率,对数据进行分类。

6) 模型训练:利用类别平衡损失函数对模型进行训练,优化参数。

3.2　模型的算法流程

输入: 任务 T^n,新类别数据 $\{X_i^n,Y_i^n\}_{i=1}^{Nt}$,旧类别数据 $\{X_i,Y_i\}_{i=1}^{m}$

输出: 模型最优参数 $\boldsymbol{W}$(权重矩阵)、b(偏置项)

1) for i in $\{X_i^n,Y_i^n\}_{i=1}^{Nt}$,利用公式(11)进行数据预处理

2) 挑选代表性样本存入存储器,利用公式(6)动态调整各旧类别样本数

3) 构造待训练数据集 $\{X^n,Y^n\}=\text{processed}(\{X_i^n,Y_i^n\}_{i=1}^{Nt})\cup\{X_i,Y_i\}_{i=1}^{m}$)

4) For i in Epoch:

5) 初始化模型参数

6) 训练模型,根据公式(10)对模型中的权重矩阵 $\boldsymbol{W}$ 和偏置项 b 进行更新

7) Return $\boldsymbol{W}$,b

8) End for

9) End

4　实验

实验的硬件环境:CPU 为 AMD Ryzen 7 5800H,内存为 16 GB(3200 MHz),显卡为 NVIDIA GeForce RTX 3070(显存 8 GB)。实验的软件环境为:Python 3.7.13,Pytorch 1.10.0。

4.1　数据集介绍

UNSW-NB 15 数据集是由 Moustafa 等[26]于澳大利亚网络安全中心收集整理,相较于 KDD'99、NSL-KDD 等数据集,其数据特征与攻击手段更加贴近现代网络攻击。数据集包含训练集和测试集,具体分布如表 1 所示。

表 1　UNSW-NB15 训练集和测试集数据分布

Tab. 1　Data distribution of UNSW-NB15 training set and test set

类别	训练集实例数	测试集实例数
Normal	56 000	37 000
Analysis	2 000	677
Backdoor	1 746	583
DoS	12 264	4 089
Exploits	33 393	11 132
Fuzzers	18 184	6 062
Generic	40 000	18 871
Reconnaissance	10 491	3 496
Shellcode	1 133	378
Worm	130	44

CICIDS 2018 数据集是由加拿大网络安全研究所开发的最新数据集[27]。该数据集涵盖 6 种攻击类别和 1 种正常类别,每条数据含 79 个特征维度,贴合现代流量行为。为避免类别不平衡问题对实验的干扰,删除 Web attack 类别,选取每类 20% 的数据作为测试集。表 2 为 CICIDS 2018 数据集包含的数据类别以及每个类别所含的数据总数。

表 2 CICIDS 2018 数据集数据分布

Tab. 2 Data distribution of CICIDS 2018 dataset

类别	数据总数
Benign	611 215
DDoS	687 742
Infiltration	161 934
Bot	286 191
Brute-Force	380 949
DoS	654 300

4.2 评价指标

采用准确率(Accuracy),精确率(Precision),召回率(Recall),F_1 值作为评价指标,计算过程见式(12)~(15):

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{12}$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{13}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{14}$$

$$F_1 = \frac{2T_P}{2T_P + F_P + F_N} \tag{15}$$

式中:T_P 为正实例被判为正实例的数量;T_N 为负实例被判为负实例的数量;F_P 为负实例被判为正实例的数量;F_N 为正实例被判为负实例的数量。

4.3 超参数设置

首先去除字符级卷积网络对实验的影响,对稀疏自注意力层的超参数进行调参实验,以寻找适配稀疏自注意力层的超参数。用于评估的数据来自训练集中随机选取的 30% 的数据,具体的实验结果如表 3 所示。

表 3 调参实验结果

Tab. 3 Parameter tuning experiment results

c	注意力头数	注意力层数	训练集准确率/%
5	2	1	92.00
10	2	1	91.70
20	2	1	86.88
5	4	1	92.32
5	8	1	92.39
5	4	2	92.67
5	4	3	93.56

如表 3 所示,在固定注意力头数和层数不变的情况下,$c = 5$ 时稀疏自注意力层表现最佳。因此,固定 c 的值为 5,并改变注意力头数和层数。实验结果表明,当注意力层数不变时,自注意力头数为 8 时模型表现最佳,但提升幅度有限。而增加注意力头数会同时增加设备的计算负担,因此选定注意力头数为 4。最后,固定 c 和注意力头数的值,改变注意力层数,由结果得知,随着注意力层数的增加,模型在训练集上的准确率不断提升,但考虑到模型的计算负担,未进一步增加注意力层数。综上,选定超参数 c 为 5,注意力头数为 4,注意力层数为 3。

对于其他超参数,依据实验和经验进一步确定。本文中涉及的超参数设置如表 4 所示。对于特征缩减模块,采用主成分分析对归一化后的数据进行特征缩减。

表 4 超参数设置

Tab. 4 Hyperparameter setting

模块	名称	设定值
稀疏自注意力层	c	5
	注意力头数	4
	稀疏自注意力层数	3
字符级卷积层	卷积核大小	3
	卷积核个数	128
	步长	1
增量机制	动态示例存储器大小	总数据集的 2%
整体参数	学习率	0.0001
	batch-size	128
	Epoch	24
	特征缩减后维度	16
	Dropout	0.005

对于 CICIDS 2018 数据集,为了避免实验冗余,超参数沿用表 4 中的设定值。

4.4 二分类实验

二分类实验主要用于验证稀疏自注意力机制对流量数据的分类效果。首先关闭增量学习机制,将数据设置为静态,对模型进行训练。二分类实验的第一部分为对比实验,将本文提出的模型与传统机器学习模型和类循环神经网络进行对比。需要注意的是,因实验环境、数据集等因素的影响,用于对比的模型结果均来自本文的复现。

从准确率上看,本文模型优于传统基于机器学习的入侵检测模型和基于类循环神经网络的入侵检测模型,说明稀疏自注意力机制可以在时间复杂度较低的情况下很好地拟合流量间的相关性,提高入侵检测的分类效果。进一步分析表 5,在 2 个数据

集上,模型的精确率(Precision)大幅高于其他模型,但召回率(Recall)较低,出现该现象的原因是大部分攻击是连贯出现的,稀疏自注意力机制会将相邻的攻击数据赋予较高的注意力值,进而提高模型识别负实例的能力,提升精确率;而正常流量中会穿插着攻击流量,这导致稀疏自注意力机制会将部分正实例误判为负实例,导致召回率下降。

表 5　本文模型与其他模型性能比较

Tab. 5　Performance comparison between proposed model and other models　%

数据集	模型	准确率	精确率	召回率	F_1 值
UNSW-NB15	SVM	82.91	73.05	98.20	83.78
	RF	81.64	71.41	98.62	82.84
	RNN	89.05	88.71	86.53	87.61
	LSTM	91.67	90.13	91.48	90.80
	GRU	91.08	87.11	94.08	90.46
	Ours	91.97	93.75	88.00	90.78
CICIDS 2018	LSTM	96.22	96.50	98.44	97.45
	GRU	92.04	92.44	97.15	94.74
	Ours	96.83	98.34	97.35	97.84

为进一步验证稀疏自注意力机制的有效性,在 UNSW-NB15 数据集上将传统自注意力机制与本文机制进行对比,结果如图 2 所示。

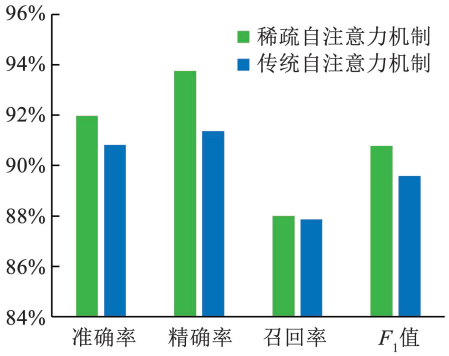

图 2　传统自注意力与稀疏自注意力机制性能对比

Fig. 2　Performance comparison between traditional self-attention and sparse self-attention mechanisms

对比结果表明,稀疏自注意力机制的各项指标均优于传统的自注意力机制。其中精确率(Precision)的提升最为明显,这是因为稀疏自注意力机制更能聚焦于关键信息,更好地学习攻击数据间的强相关性,减少了模型对负实例的误判。而两者召回率(Recall)均相对较低且差距较小,原因为正实例中突发的攻击行为不利于类自注意力模型学习正实例间的相关性,从而导致模型误判。

最后对字符级卷积层进行消融实验,分析字符级卷积层对模型分类准确率的影响。结果如表 6 所示。

表 6　字符级卷积层的消融实验结果

Tab. 6　Ablation experiment sresults of character-level convolutional layers

方法	准确率/%
完整模型(Ours)	91.97
移除字符级卷积层(Without Char-CNN)	90.58

实验表明,字符级卷积层对模型性能有正向贡献。未经过字符级卷积层处理的数据,可能无法使后续的分类模型充分利用数据特征学习流量内的潜在相关性,导致模型分类准确率降低。

4.5　多分类增量学习实验

本部分实验主要验证所提增量学习机制的有效性。首先将 UNSW-NB15 数据集分为 3 个顺序任务:T_1 包含的类别为 Normal、Analysis、Backdoor;T_2 包含 DoS、Exploits、Fuzzers;T_3 包含 Generic、Reconnaissance、Shellcode。将 CICIDS 2018 数据集分为 2 个任务:T_1 包含 Benign、DDoS、Infiltration;T_2 包含 Bot、Brute-Force、DoS。对比结果如表 7 所示。

表 7　增量机制对比实验结果

Tab. 7　Incremental mechanism comparison experiment results

数据集	任务阶段	方法	准确率/%
UNSW-NB15	T_1	文献[28]	90.80
		文献[29]	88.86
		本文方法	89.54
	T_2	文献[28]	64.68
		文献[29]	74.17
		本文方法	69.32
	T_3	文献[28]	55.01
		文献[29]	59.91
		本文方法	61.65
CICIDS 2018	T_1	文献[28]	95.77
		文献[29]	96.91
		本文方法	96.30
	T_2	文献[28]	93.94
		文献[29]	94.72
		本文方法	94.85

分析实验结果,在 UNSW-NB15 数据集上,本文方法在 T_2 阶段优于文献[28],但略逊于文献[29];在 T_3 阶段,优于另外两种方法。这是因为本文方法结合了基于回放和基于正则化的策略,随着增量任务迭代,模型不会大量遗忘已学习过的知识,对早期任务的记忆保持也更为稳定。但当增量学习任务数较少时,本文方法的优势可能不明显;当任务数增大时,本文方法性能优势更为突出。在 CICIDS 2018 数据集上,本文方法在两个任务阶段均取得了有竞

争力的结果。综上,本文增量学习机制在任务较少的情况下与其他方案性能持平,在任务增多时展现出更好的长期记忆能力。

进一步,通过消融实验分析各增量学习模块的作用。由于 UNSW-NB15 可以被分为更多的任务阶段,因此后续的消融实验使用 UNSW-NB15 数据集进行验证。首先,将动态示例存储器替换为传统的存储器,比较二者之间的性能差别。图 3 为完整模型在 T_3 上的混淆矩阵,图 4 为使用传统存储器后的混淆矩阵。

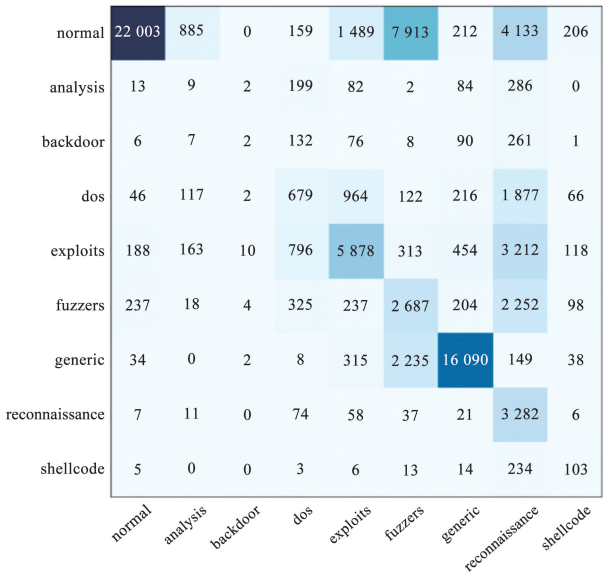

图 3 完整模型的混淆矩阵

Fig. 3 Confusion matrix of complete model

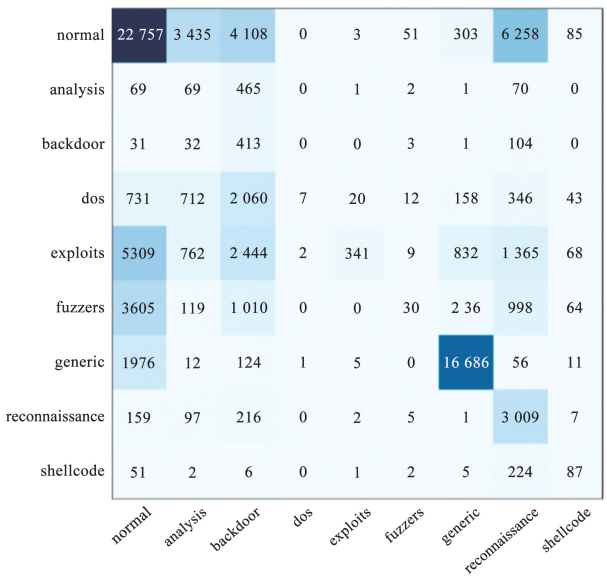

图 4 使用传统存储器后的混淆矩阵

Fig. 4 Confusion matrix with traditional memory

通过图 3 和图 4 可以发现,在使用传统存储器时,模型对 T_1 中的 Analysis、Backdoor 类别的识别率较高,但对 T_2 中的 3 个类别识别能力较差。这是因为传统存储器会为每个旧类别分配相等的存储空间,不符合入侵检测数据固有的类别不平衡分布,导致存储的样本分布发生概念漂移,进而严重影响模型性能。

更进一步,分析类别平衡损失函数对增量学习机制产生的影响。在完整模型的基础上,将损失函数替换为普通交叉熵损失,在 T_3 上得到的混淆矩阵结果如图 5 所示。

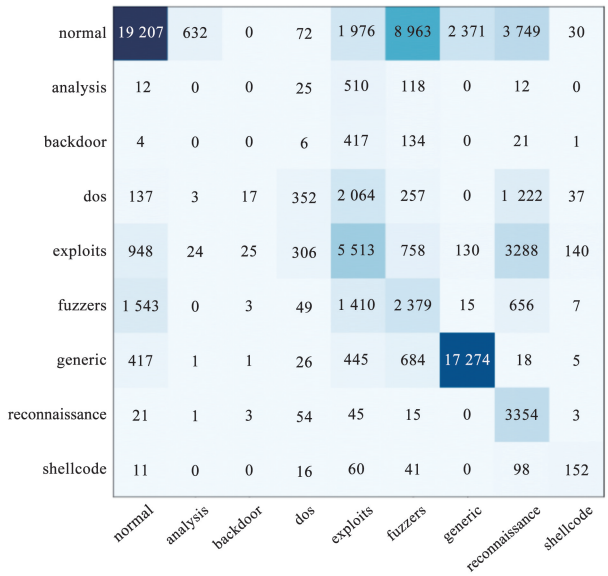

图 5 使用普通交叉熵损失后的混淆矩阵

Fig. 5 Confusion matrix with ordinary cross-entropy loss

对比图 3 和图 5,移除类别平衡损失函数后,模型对 6 种旧类别的遗忘程度均有所加深。结果表明,模型仅依靠动态示例存储器中的旧样本,不足以有效缓解灾难性遗忘。同时,移除类别平衡损失之后,模型对新类别的识别率有所提升,这与类别平衡损失函数的设计初衷相符。

5 结论与展望

本文针对传统入侵检测模型对流量数据的相关性关注度不足、传统自注意力机制计算复杂度高的问题,提出了稀疏自注意力机制。该机制在降低计算复杂度的同时,有效利用数据间相关性对流量数据实现精准分类。此外,本文融合基于回放与基于正则化的方法,构建动态示例存储器和类别平衡损失函数,在不改变模型原有结构的基础上实现了入侵检测模型的增量学习,有效缓解了模型在动态环境中的灾难性遗忘问题,为现代入侵检测模型的设计提供了实用参考。未来工作将更专注于研究鲁棒、高效的增量式入侵检测系统,使其可以更好地适应互联网环境的动态变化,推动科研成果向实际应用转化。

参考文献

[1] LI Xukui, CHEN Wei, ZHANG Qianru, et al. Building auto-encoder intrusion detection system based on random forest feature selection[J]. Computers & Security, 2020, 95: 101851. DOI: 10.1016/j. cose. 2020. 101851

[2] KAN Xiu, FAN Yixuan, FANG Zhijun, et al. A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network[J]. Information Sciences, 2021, 568: 147. DOI: 10.1016/j. ins. 2021. 03. 060

[3] ANDRESINI G, APPICE A, CAFORIO F P, et al. ROULETTE: A neural attention multi-output model for explainable network intrusion detection[J]. Expert Systems with Applications, 2022, 201: 117144. DOI: 10.1016/j. eswa. 2022. 117144

[4] YU Jing, YE Xiaojun, LI Hongbo. A high precision intrusion detection system for network security communication based on multi-scale convolutional neural network[J]. Future Generation Computer Systems, 2022, 129: 399. DOI: 10.1016/j. future. 2021. 10. 018

[5] 陈家琪,王琦,汤奕,等. 考虑双侧特征的电力信息物理系统异常检测方法[J]. 电网技术,2022,46(6):2339.
CHEN Jiaqi, WANG Qi, TANG Yi, et al. Anomaly detection method for cyber physical power system considering bilateral features[J]. Power System Technology, 2022, 46(6): 2339. DOI: 10.13335/j. 1000 - 3673. pst. 2021. 0370

[6] VIEGAS E, SANTIN A O, ABREU V JR. Machine learning intrusion detection in big data era: A multi-objective approach for longer model lifespans[J]. IEEE Transactions on Network Science and Engineering, 2020, 8(1): 366. DOI: 10.1109/TNSE. 2020. 3038618

[7] MOHAMED M R, NASR A A, TARRAD I F, et al. Exploiting incremental classifiers for the training of an adaptive intrusion detection model[J]. International Journal of Network Security, 2019, 21(2): 275

[8] PUTINA A, ROSSI D. Online anomaly detection leveraging stream-based clustering and real-time telemetry[J]. IEEE Transactions on Network and Service Management, 2020, 18(1): 839. DOI: 10.1109/TNSM. 2020. 3037019

[9] KASONGO S M, SUN Yanxia. A deep long short-term memory based classifier for wireless intrusion detection system[J]. ICT Express, 2020, 6(2): 98. DOI: 10.1016/j. icte. 2019. 08. 004

[10] SETHI K, MADHAV Y V, KUMAR R, et al. Attention based multi-agent intrusion detection systems using reinforcement learning[J]. Journal of Information Security and Applications, 2021, 61: 102923. DOI: 10.1016/j. jisa. 2021. 102923

[11] KHAN M A. HCRNNIDS: Hybrid convolutional recurrent neural network-based network intrusion detection system[J]. Processes, 2021, 9(5): 834. DOI: 10.3390/pr9050834

[12] LAN Jinghong, LIU Xudong, LI Bo, et al. A novel hierarchical attention-based triplet network with unsupervised domain adaptation for network intrusion detection[J]. Applied Intelligence, 2023, 53(10): 11705. DOI: 10.1007/s10489 - 022 - 04076 - 0

[13] FU Yanfang, DU Yishuai, CAO Zijian, et al. A deep learning model for network intrusion detection with imbalanced data[J]. Electronics, 2022, 11(6): 898. DOI: 10.3390/electronics11060898

[14] 石磊,张吉涛,高宇飞,等. 基于 Transformer 与 BiLSTM 的网络流量入侵检测[J]. 计算机工程,2023,49(3):29
SHI Lei, ZHANG Jitao, GAO Yufei, et al. Intrusion detection of network traffic based on Transformer and BiLSTM[J]. Computer Engineering, 2023, 49(3): 29. DOI: 10.19678/j. issn. 1000 - 3428. 0065135

[15] MAHDAVI E, FANIAN A, MIRZAEI A, et al. ITL-IDS: incremental transfer learning for intrusion detection systems[J]. Knowledge-Based Systems, 2022, 253: 109542. DOI: 10.1016/j. knosys. 2022. 109542

[16] WANG Chengru, XU Rongfang, Lee S J, et al. Network intrusion detection using equality constrained-optimization-based extreme learning machines[J]. Knowledge-Based Systems, 2018, 147: 68. DOI: 10.1016/j. knosys. 2018. 02. 015

[17] 刘强,张颖,周卫祥,等. 自适应类增量学习的物联网入侵检测系统[J]. 计算机工程,2023,49(2):169
LIU Qiang, ZHANG Ying, ZHOU Weixiang, et al. Adaptive class incremental learning-based IoT intrusion detection system[J]. Computer Engineering, 2023, 49(2): 169. DOI: 10.19678/j. issn. 1000 - 3428. 0063917

[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the Neural Information Processing Systems 30. La Jolla, CA: MIT Press, 2017: 5998. DOI: 10.5555/3295222. 3295349

[19] 郭志民,周劼英,王丹,等. 基于 Transformer 神经网络模型的网络入侵检测方法[J]. 重庆大学学报,2021,44(11):81
GUO Zhimin, ZHOU Jieying, WANG Dan, et al. Network intrusion detection method based on Transformer neural network model[J]. Journal of Chongqing University, 2021, 44(11): 81. DOI: 10.11835/j. issn. 1000 - 582X. 2021. 11. 010

[20] WU Zihan, ZHANG Hong, WANG Penghai, et al. RTIDS: A robust transformer-based approach for intrusion detection system[J]. IEEE Access, 2022, 10: 6437. DOI: 10.1109/ACCESS. 2022. 3182333

[21] ULLAH F, ULLAH S, SRIVASTAVA G, et al. IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic[J]. Digital Communications and Networks, 2023, 9(6): 1541. DOI: 10.1016/j. dcan. 2023. 03. 008

[22] ZHAO Ruijie, GUI Guan, XUE Zhi, et al. A novel intrusion detection method based on lightweight neural network for internet of things[J]. IEEE Internet of Things Journal, 2021, 9(12): 9960. DOI: 10.1109/JIOT. 2021. 3119055

[23] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(12): 11106. DOI: 10.1609/aaai. v35i12. 17325

[24] BEDI P, GUPTA N, JINDAL V. I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems[J]. Applied Intelligence, 2021, 51(2): 1133. DOI: 10.1007/s10489 - 020 - 01886 - y

[25] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. ICaRL: Incremental classifier and representation learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2017: 2001. DOI: 10.1109/cvpr. 2017. 587

[26] MOUSTAFA N, SLAY J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//2015 Military Communications and Information Systems Conference (MilCIS). Piscataway, NJ: IEEE, 2015: 1. DOI: 10.1109/milcis. 2015. 7348942

[27] THAKKAR A, LOHIYA R. A review of the advancement in intrusion detection datasets[J]. Procedia Computer Science, 2020, 167: 636. DOI: 10.1016/j. procs. 2020. 03. 330

[28] DATA M, ARITSUGI M. T-DFNN: An incremental learning algorithm for intrusion detection systems[J]. IEEE Access, 2021, 9: 154156. DOI: 10.1109/ACCESS. 2021. 3127985

[29] CONSTANTINIDES C, SHIAELES S, GHITA B, et al. A novel online incremental learning intrusion prevention system[C]//2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). Piscataway, NJ: IEEE, 2019: 1. DOI: 10.1109/ntms. 2019. 8763842

(编辑　丁晓清)