DOI: 10.13382/j.jemi.B2205445

基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法*

梁欣怡 行鸿彦 侯天浩

(南京信息工程大学江苏省气象灾害预报预警与评估协同创新中心 南京 210044)

摘 要:针对网络入侵检测中攻击样本和流量特征不足的问题,提出一种基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法,实现在流量数据中检测异常网络流量的目标。通过分析流量特征数据分布差异,采用 IQR 异常值处理方法进行数据预处理,使用自编码器对攻击样本进行数据增强,构建 CNN-BiLSTM 神经网络和自编码器组成半自监督模型,分别提取高维流量特征和自监督特征,将组合特征作为最终特征输入到分类模型中进行预测分类,实现网络入侵检测。实验结果表明,与其他入侵检测方法相比,所提方法在准确率和 F1 分数上分别达到了 85.7%和 85.1%,能够有效提高网络入侵的检测精度以及对未知攻击的检测能力。

关键词: 深度学习;自监督学习;数据增强;网络入侵检测

中图分类号: TP393;TN911.7 **文献标识码**: A **国家标准学科分类代码**: 510.40

CNN-BiLSTM network intrusion detection method based on self-supervised feature enhancement

Liang Xinyi Xing Hongyan Hou Tianhao

(Jiangsu Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Aiming at the problem of insufficient attack samples and traffic characteristics in network intrusion detection, a CNN-BiLSTM network intrusion detection method based on self-supervised feature enhancement was proposed to detect abnormal network traffic in traffic data. By analyzing the difference in the distribution of traffic characteristic, IQR outlier processing method was used for data preprocessing, and autoencoder was used to enhance the number of attack samples. A semi-self-supervised model composed of CNN-BilSTM neural network and autoencoder was constructed to extract high-dimensional traffic characteristics and self-supervised features respectively. The combined features are input into the classification model as the final features for prediction and classification, so as to realize the function of network intrusion detection. The experimental results show that compared with other intrusion detection methods, the accuracy and F1 score of the proposed method are 85.7% and 85.1% respectively, which can effectively improve the detection accuracy of network intrusion and the detection ability of unknown attacks.

Keywords: deep learning; self-supervised learning; data enhancement; network intrusion detection

0 引 言

随着网络信息领域关键技术的不断突破与发展,互联网已经成为推动国家各个领域高质量发展的关键基础设施,对实现国家经济快速发展具有重大意义。与此同时,网络攻击事件频繁发生,严重危害我国国家安全,影响我国信息化建设的健康发展。如何有效监测网络攻击事件[1]已经成为影响国家安全、社会公共利益的突出问题。而网络入侵检测系统通过对网络进行实时监视,能够有效感知网络攻击,对于维护网络空间安全起着重要的作用,具有重要的研究意义。

尽管网络入侵检测技术已经发展了数十年,但是现有的网络入侵检测技术仍然面临着日益复杂的互联网攻击和海量数据入侵检测的挑战,准确检测异常流量对于网络安全性和可靠性尤为重要,基于检测技术可以将入

收稿日期: 2022-05-01 Received Date: 2022-05-01

*基金项目:国家重点研发计划(2021YFE0105500)、国家自然科学基金(62171228)项目资助

侵检测系统划分为基于误用的入侵检测和基于异常的入侵检测[2]。由于基于误用的入侵检测系统难以检测零日攻击,所以基于异常的入侵检测是目前网络入侵检测领域研究的重点。

近年来,机器学习已经广泛应用于网络入侵领域。Gao 等[3]使用不同的分类器,如决策树、随机森林、KNN 等作为基本分类器,提出了自适应集成学习投票算法,提高了入侵检测的准确性。Verma 等[4]提出了在检测网络入侵时使用机器学习分类算法 XGBoost 和 AdaBoost 来训练 NIDS 模型,表明基于异常的机器学习入侵检测具有很大的改进潜力。但是现有的基于传统的机器学习方法是简单的浅层特征学习,面对大规模高维网络流量数据,往往需要进行复杂的特征提取工程,且准确率较低。而深度学习网络以其强大的算力和学习能力,不需要复杂的特征工程就可以自动进行高维数据的特征选择,更适合用于网络入侵检测。Su 等[5]提出了一种结合了双向长短时记忆(Bidirectional long short term memory, BiLSTM)和注意力机制的流量异常检测模型 BAT,利用注意力机制对由 BiLSTM 模型生成的网络流向量进行筛选,获得网络流分类的关键特征,能够有效提高异常检测能力;Abolhasanzadeh 等[6]提出使用自动编码器对流量特征进行降维,减少了入侵检测系统时间和空间的复杂性,且优于 PCA 等传统降维方法;Ieracitano 等[7]提出一种基于统计分析和自动编码器(auto encoder, AE)的入侵检测模型以提取最优特征,实现了更好的分类性能。

虽然上述方法都在一定程度上改善了检测效果,但仍然存在如下问题:在实际流量中,入侵状态在一般情况下颇为少见,正常流量和入侵流量之间的不平衡往往会导致分类器偏向多数类结果;面对日益复杂的网络环境和海量的入侵数据[8],现有的许多方法泛化能力不够,不能有效检测未知攻击。

考虑到上述问题,提出一种基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法。在对攻击类流量进行数据增强的同时,还利用自监督模型提取自监督特征对流量特征进行增强,辅助 CNN-BiLSTM 网络完成后续的分类任务,有效提高入侵检测的准确率。

1 理论基础

自编码器作为一种数据增强技术也常用于对攻击样本数据进行不平衡处理[9],而且自编码器作为自监督模型的一种,最主要的目的就是学习到更丰富的信息表征[10],可以用来解决攻击样本不足和流量特征不足的问题。而卷积神经网络和双向长短时记忆网络可以提取网络流量的高维时空流量特征,结合自编码器提取自监督特征完成后续的分类任务。

1.1 自编码器

自监督学习主要是利用辅助任务从大规模的无监督数据中挖掘自身的监督信息,通过构造的监督信息对网络进行训练,从而学习到对下游任务有价值的表征。自编码器(AE)作为自监督模型的一种,最主要的目的就是将输入信息作为学习目标,学习到更丰富的信息表征,可应用于降维、降噪、异常值检测、数据生成、深度神经网络的预训练。

基本的自编码器[11]可以看作一个 3 层的神经网络结构:输入层、隐层、输出层,其网络结构如图 1 所示。图中 x_i 表示输入节点,y_i 表示输出节点,"+1"表示偏置项。设编码函数以及解码函数为 $f(x)$ 和 $g(x)$,其中 f 和 g 为神经网络。

1)编码器(encoder)部分:

输入变量集合 X 通过一个神经网络(编码器)降维压缩得到隐变量集合 H。

$$H = f(X) = \sigma(W_1 X + b_1) \tag{1}$$

2)解码器(decoder)部分:

隐变量集合 H 通过另一个神经网络(解码器)将降维后的特征进行重构,得到输出变量集合 Y。

$$Y = g(H) = \sigma(W_2 H + b_2) \tag{2}$$

自编码器[12]的训练目标是使得还原后得到的重构特征尽可能地与原始特征保持一致,即目标函数 $L(X,Y)$ 达到最小值。

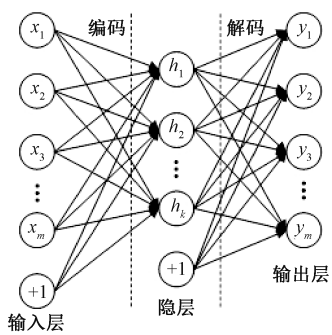

图 1 自编码器结构图

Fig. 1 Structure diagram of autoencoder

1.2 卷积神经网络

卷积神经网络(convolutional neural network, CNN)是一种带有卷积结构的深度神经网络,在空间特征提取方面效果极佳,常用于信号处理及图像分类当中。基本的 CNN 结构如图 2 所示,其中,卷积层的主要作用是提取输入图片的局部特征;为防止过拟合,往往在卷积层后面接池化层以降低数据维度;卷积层和池化层交替叠加的深层网络能够迭代提取更复杂的特征,并由全连接层整合输入到分类器进行分类。

1.3 双向长短时记忆神经网络

长短时记忆神经网络(long short-term memory,

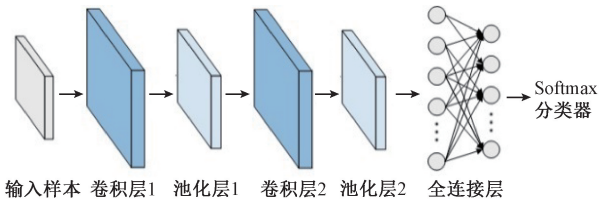

图 2　卷积神经网络结构

Fig. 2　Structure diagram of convolutional neural network

LSTM)是循环神经网络中的一种,其基本结构如图 3 所示,与普通的循环神经网络相比,LSTM 的特殊之处在于它包括 3 个门:输入门 i_t ,遗忘门 f_t 和输出门 o_t ,可以更好地捕捉到较长距离的依赖关系,具体工作原理如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{8}$$

式中:输入门 i_t ,遗忘门 f_t ,输出门 o_t ,前一时刻隐层状态 h_{t-1},当前时刻隐层状态 h_t ,当前时刻输入 x_t ,临时单元状态 $\tilde{C}_t$,当前时刻单元状态 C_t ,上一时刻单元状态 C_{t-1}。

遗忘门用于决定遗忘和丢弃的信息,记忆门通过 sigmoid 函数和 tanh 函数决定需要更新的单元和更新的单元信息,输出门根据单元状态,通过 sigmoid 函数和 tanh 函数确定输出值。

BiLSTM 是一种特殊的 LSTM 网络,由前向 LSTM 与后向 LSTM 组合而成,广泛应用于自然语言处理任务中,与 LSTM 相比,可以更好地捕捉双向的依赖关系。考虑到入侵检测数据集中包含具有时间顺序的网络流量,因此应用 BiLSTM 来提取时间特征具有高度适配性。

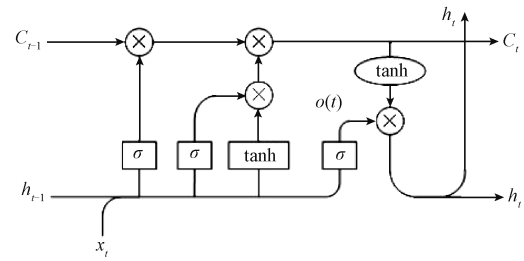

图 3　LSTM 结构图

Fig. 3　Structure diagram of LSTM

2　基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法

由于流量特征具有时空特性,CNN-BiLSTM 模型凭借出色的时空特征提取能力被应用于网络入侵检测,但许多研究并没有考虑到数据不平衡和流量特征不足的问题,因此本文提出一种基于自监督特征增强的 CNN-BiLSTM 方法,在 CNN-BiLSTM 模型的基础上增加了两个自编码器进行改进,两个自编码器分别负责数据增强和特征增强,有效提高了整体模型的检测性能。

2.1　基于数据增强和自监督特征增强的 CNN-BiLSTM 模型

基于数据增强和自监督特征增强的 CNN-BiLSTM 模型(CNN-BiLSTM network intrusion detection model based on data enhancement and self-supervised feature enhancement, AE-CNN-BiLSTM-AE)如图 4 所示,主要由数据增强模型,特征增强模型以及 CNN-BiLSTM 模型组成,其中数据增强自编码器 1 生成攻击类流量样本,扩充攻击类样本数量,完成数据增强的任务;CNN-BiLSTM 模型负责提取流量的高维时空特征;特征增强自编码器 2 负责学习流量数据集更丰富的信息表征,生成自监督特征,进行特征增强,辅助 CNN-BiLSTM 网络完成后续的分类任务。

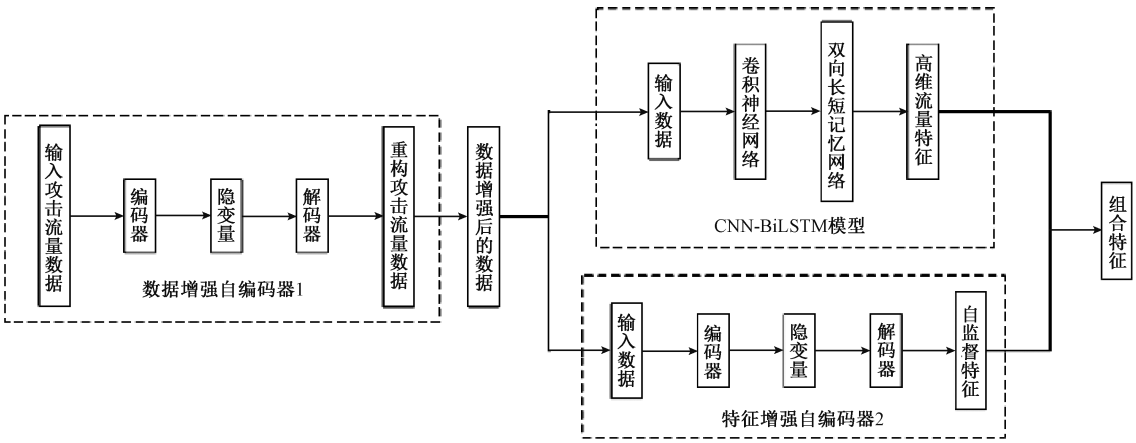

图 4　AE-CNN-BiLSTM-AE 整体模型

Fig. 4　Overall model diagram of AE-CNN-BiLSTM-AE

1)增强模型

用于数据增强的自编码器 1 和用于特征增强的自编码器 2 采用相同的模型结构,由输入层,全连接层,批量正则化层,输出层组成,具体模型结构如图 5 所示。

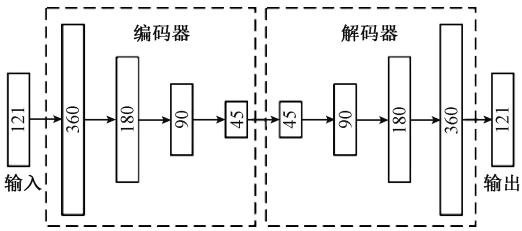

图 5 自编码器模型结构图

Fig. 5 Model structure diagram of autoencoder

每个全连接层后面都接一个批量正则化层,为简化表示,批量正则化层未在图中给出。

2)CNN-BiLSTM 模型

为提取高维流量特征,提出一种 CNN-BiLSTM 模型,包括输入层,二维卷积层,池化层,全连接层,BiLSTM 层以及输出层,其具体结构如图 6 所示。

CNN 网络采用两组卷积-池化层完成对空间特征的提取;接着通过全连接层输入到 BiLSTM 网络中提取时间特征,两层 BiLSTM 分别使用 128 个和 64 个神经元,最终输出 CNN-BiLSTM 神经网络提取的高维时空特征。

优选地,上述模型都使用 AdamW 优化器,自编码器 1 和自编码器 2 采用 tanh 激活函数,CNN-BiLSTM 中的

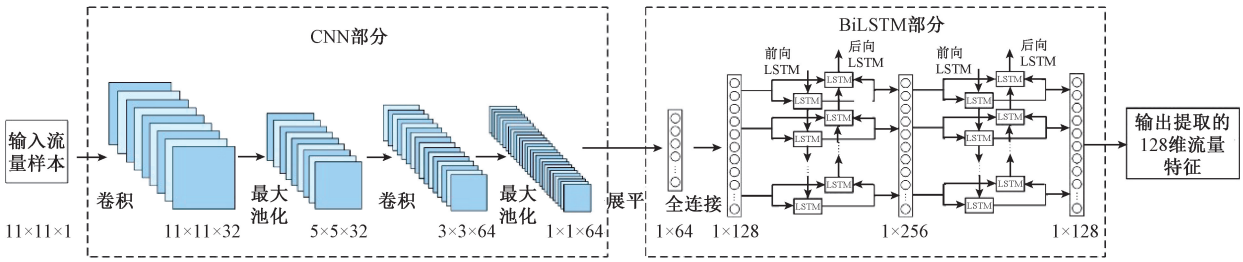

图 6 CNN-BiLSTM 模型结构图

Fig. 6 Model structure diagram of CNN-BiLSTM

CNN 层采用 ReLU 函数作为激活函数,BiLSTM 层采用 sigmoid 函数作为激活函数,并且在初始化参数时采用 kaming 初始化方法和 xavier 初始化方法。

2.2 检测流程

对于提出的基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法,本文采用 NSL-KDD 数据集[13]进行训练与测试,具体检测流程图如图 7 所示,下面将根据图 7 对网络入侵检测流程进行详细阐述。

1)对入侵检测数据集进行数据预处理,包括符号特征数值化,异常值处理和归一化处理

(1)符号特征数字化

为方便模型训练,需要将 NSL-KDD 数据集中的 3 个符号特征 protocol_type,service,flag 转换成数字特征表示。经过独热编码,上述特征分别转换成 3、70 和 11 个数值特征,再加上原有的 38 个数字特征,则原来的 41 维特征经数值化处理之后变为 122 维。由于特征 num_outbound_cmds 全 0,对训练帮助不大,将其删除得到 121 维特征。

(2)四分位距(IQR)异常值处理

离群值是指在数据中与其他数值相比差异较大的值,其存在往往是无法避免的。数据集中,过大或过小的极端数据都是离群值,可能会影响到分析结果,尤其是在分类预测时,需要对那些离群值进行谨慎处理。

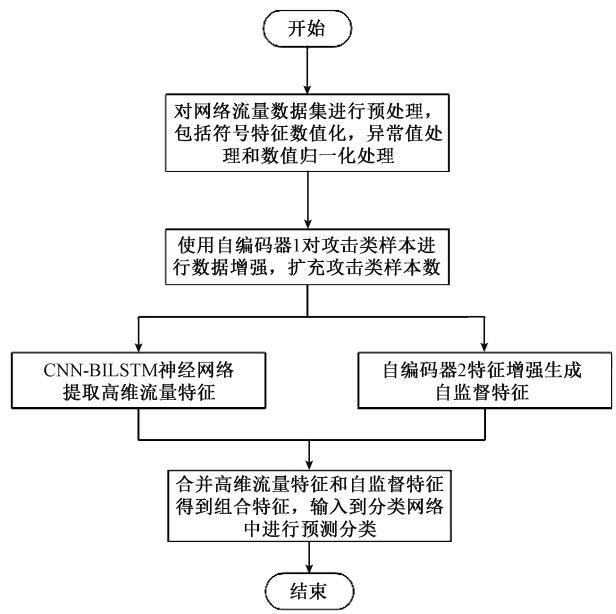

图 7 检测流程

Fig. 7 Detection flow chart

大部分研究者对入侵检测数据集的预处理仅采用数值化和归一化,忽略了对离群值的处理。因此在数据归一化之前,对 38 个数值特征做数据分析,发现 duration,src_bytes,dst_bytes 这 3 个特征数据分布差异过大,为避

免离群值对检测结果的影响,需对其进行异常值处理。

$$IQR = Q_3 - Q_1 \tag{9}$$

$$OF = Q_3 + 1.5 \cdot IQR \tag{10}$$

对流量采用四分位距异常值处理方法,其具体算法流程如下:首先计算该特征所有数据的第一四分位数 Q_1 和第三四分位数 Q_3,根据式(9)计算出四分位距 IQR;再由式(10)计算出异常值边界 OF;最后按算法 1 所示算法对特征进行处理。

算法 1 四分位距异常值处理方法

输入:X 指数据集中流量样本的集合$\{x_1, x_2, x_3, \cdots, x_n\}$

F 指每个流量样本需要 IQR 处理的特征集合$\{f_1, f_2, f_3\}$

OF 指每个特征的异常值边界集合$\{of_1, of_2, of_3\}$

输出:经过四分位距异常值处理的流量样本 X

```
for each x ∈ X do
  for each f ∈ F do
      if of = 0
        then if x > 0
        then x ← 1
        else x = 0
      else
        if x > of
          then x ← of
  end
end
```

(3)归一化处理

对经过标准化和四分位距异常值处理后的数据集根据式(11)进行 Min-Max Scaling 处理,将数值归一化到 0~1 之间:

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

式中:$x_{\max}$ 为样本数据的最大值,$x_{\min}$ 为样本数据的最小值,x^* 为归一化之后的数据。

2)使用自编码器 1 对攻击类样本进行数据增强

将数据预处理后训练集中的攻击类样本的 121 维特征 x_i 输入到深度自编码器 1 中,输出重构样本 $\hat{x}_i$,则 x_i 和 $\hat{x}_i$ 经过 *log_softmax* 分类器和 *softmax* 分类器的数据分布分别为 $p_1(x_i)$ 和 $q_1(x_i)$。

$$D_{KL1} = \sum_i^n p_1(x_i) \log \frac{p_1(x_i)}{q_1(x_i)} \tag{12}$$

$$MSE = \frac{1}{n} \sum_i^n (x_i - \hat{x}_i)^2 \tag{13}$$

$$L_1 = 0.5MSE + 0.5D_{KL1} \tag{14}$$

使用上述结合 KL 散度 D_{KL1} 和 MSE 损失的自定义损失函数 L_1 作为评估标准进行迭代预训练 500 轮,采用训练中损失最小时对应的最佳模型进行数据增强生成 58 630 条攻击流量样本,并将生成的 58 630 个攻击样本与训练集样本合并得到最终训练集。

3)使用 CNN-BiLSTM 模型和自编码器 2 分别提取高维流量特征和自监督特征

首先将最终训练集每个样本的 121 维特征转换成 11×11 维特征输入到 CNN 网络中提取流量空间特征,再通过全连接层输入到 BiLSTM 网络提取流量的时间特征,最终输出 128 维高维流量特征;接着将最终训练集特征输入到自编码器 2 中生成 121 维自监督特征,辅助 CNN-BiLSTM 模块完成后续的分类任务;最后合并高维流量特征和自监督特征,得到特征增强后的最终特征,输入到分类网络中进行预测分类。

将最终训练集的 121 维特征 x_i' 通过深度自编码器 2 生成的自监督特征记做 $\hat{x}_i'$,则 x_i' 和 $\hat{x}_i'$ 经过 *log_softmax* 分类器和 *softmax* 分类器的数据分布分别为 $p_2(x_i')$ 和 $q_2(x_i')$。

$$D_{KL2} = \sum_i^N p_2(x_i') \log \frac{p_2(x_i')}{q_2(x_i')} \tag{15}$$

$$L_c = -\frac{1}{N} \sum_i^N [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \tag{16}$$

$$L_2 = 0.8L_c + 0.2D_{KL2} \tag{17}$$

式中:L_c 指预测分类值和真实类别之间的交叉熵损失;y_i 表示样本 i 的标签,攻击为 1,正常为 0;$p_i(x)$ 指样本 i 预测为攻击类的概率;

使用上述结合 KL 散度 D_{KL2} 和交叉熵损失 L_c 的自定义损失函数 L_2 作为评估标准对半自监督模型进行迭代训练更新模型参数。

4)最后将测试集输入训练完成的模型测试模型性能。

3　实验与分析

为了评估入侵检测方法的有效性,实验使用深度学习开源框架 PyTorch 进行测试。在 Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz NVIDIA GeForceGTX1050, RAM8GB 的环境中进行。

3.1　数据集与评估标准

1)数据集

实验的网络流量数据集采用 NSL-KDD 数据集中的 KDDTrain+和 KDDTest+文件。NSL-KDD 数据集是 KDDcup99 数据集[14]的改进版本,由于其不含冗余和重复记录,作为基准数据集被广泛应用于许多入侵检测系统中。其中训练集 KDDTrain+包括 22 种攻击类型的标签样本,测试集 KDDTest+包含 39 种攻击类型的标签样本,因此采用 NSL-KDD 数据集能够评估模型的泛化能力,使检测更为准确。数据集的样本分布情况如表 1 所

示,其中每个流量样本包含 41 个特征,包括 38 个数值(例如“int64”或“float64”)和 3 个符号值(例如“object”)。此外,虽然 KDDTrain+和 KDDTest+都包含多个类标签,但本文的目的是检测出攻击,只进行二分类任务,据此对数据集标签进行替换,正常流量标签为 0,异常流量标签为 1。

表 1 NSL-KDD 数据集的样本分布情况

Table 1 Sample distribution of NSL-KDD dataset

NSL-KDD 数据集	正常流量样本数	异常流量样本数	总计样本数
训练集 KDDTrain+	67 348	58 630	125 973
测试集 KDDTest+	9 711	12 833	22 544

2)评估标准

为了评估所提方法的有效性,使用准确率(Accuracy),精度(Precision),召回率(Recall)和 F1 分数作为实验过程中的衡量指标。整体精度性能是通过分析入侵检测模型的 F1 分数来衡量的。F1 分数是精度和召回的调和平均值,其中精确度能够衡量入侵检测系统识别攻击的能力,而召回率可以被认为是系统查找所有攻击的能力。F1 分数越高,算法在"精度"和"召回率"之间取得的平衡就越好。相反,当一个指标以牺牲另一个指标为代价来改进时,F1 得分会受到影响。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 = \frac{2 \times R \times P}{R + P} \tag{21}$$

式中:*TP* 和 *FP* 分别表示正确预测和错误预测流量为正常类型的样本数,*TN* 和 *FN* 分别表示正确预测和错误预测流量为攻击类型的样本数。*TP*、*TN*、*FP* 和 *FN* 四者之和为总样本数。

3.2 实验结果与分析

自编码器结构的不确定性在于网络的深度和每个隐藏层的神经元数目。目前,还没有成熟的理论方法来选择自编码器的最优网络结构。如果模型过于简单,可能无法有效提取输入向量的压缩表示,相反,更深层的自编码器模型意味着更好的非线性表示能力,更有可能学到更高层次的特征表示,以处理更复杂的数据。因此需要根据网络入侵检测的需要,设置合适的自编码器结构以提取最优的自监督特征辅助 CNN-BiLSTM 模型训练。在实验中,分别测试不同层数和不同神经元数目的自编码器结合 CNN-BiLSTM 模型组成的半自监督模型(CNN-BiLSTM-AE)在 KDDTest+上的表现。

如表 2 所示,实验结果表明,随着隐藏层数的增加,半自监督模型分类效果逐渐增强,采用 9 层 $AE_{121\text{-}360\text{-}180\text{-}90\text{-}45\text{-}90\text{-}180\text{-}360\text{-}121}$ 的效果是最好的,因此将此结构作为最终的自编码器模型。

表 2 不同自编码器结构的半自监督模型(CNN-BiLSTM-AE)在 KDDTest+上的实验结果

Table 2 Experimental results of semi-self-supervised models with different autoencoder structures (CNN-BILSTM-AE) on KDDTest+ (%)

AE 模型结构	准确率 Accuracy	精度 Precision	召回率 Recall	F1 分数
$AE_{121\text{-}32\text{-}121}$	80.54	81.24	81.31	80.52
$AE_{121\text{-}45\text{-}121}$	81.95	82.74	82.95	81.92
$AE_{121\text{-}64\text{-}32\text{-}64\text{-}121}$	82.74	83.47	83.29	82.45
$AE_{121\text{-}90\text{-}45\text{-}90\text{-}121}$	82.99	83.59	83.52	82.97
$AE_{121\text{-}128\text{-}64\text{-}32\text{-}64\text{-}128\text{-}121}$	83.07	83.33	83.07	83.01
$AE_{121\text{-}180\text{-}90\text{-}45\text{-}90\text{-}180\text{-}121}$	83.37	84.11	84.20	83.35
$AE_{121\text{-}256\text{-}128\text{-}64\text{-}32\text{-}64\text{-}128\text{-}256\text{-}121}$	83.76	84.36	84.25	83.74
$AE_{121\text{-}360\text{-}180\text{-}90\text{-}45\text{-}90\text{-}180\text{-}360\text{-}121}$	**84.17**	**84.77**	**84.67**	**84.15**

为了验证数据增强技术和特征增强技术对于入侵检测的有效性,将半自监督模型分别与未经数据增强和自监督特征增强的模型进行对比。各个模型在 KDDTest+上的测试结果如表 3 所示,传统的 CNN-BiLSTM 的准确率和 F1 分数是 82.1%和 82.1%;经过 AE 数据增强后的 CNN-BiLSTM 准确率和 F1 分数是 84.19%和 84.17%;经过自监督特征增强之后的 CNN-BiLSTM 模型的准确率和 F1 分数是 84.17%和 84.15%。实验结果表明数据增强和特征增强能够提升 CNN-BiLSTM 的检测能力,而将两者结合后的 AE-CNN-BiLSTM-AE 模型的准确率和 F1 分数达到了 85.7%和 85.1%,效果提升最为显著,可见通过数据增强和自监督特征增强能够显著提高入侵检测模型的准确率。

由于网络入侵检测的最终目的是在流量数据中检测出异常网络流量,所以检测准确率是本文最关心的指标。虽然采用数据增强和自监督增强技术会使模型复杂度略高于其他模型,但就准确率和 F1 分数而言,AE-CNN-BiLSTM-AE 模型带来的提升是其他模型所不能及的。

表 3　不同模型在 KDDTest+上的实验结果

Table 3　Experimental results of different models on KDDTest+　　(%)

模型	准确率 Accuracy	精度 Precision	召回率 Recall	F1 分数
CNN	82.53	83.29	83.45	82.5
CNN-BiLSTM	83.69	84.38	84.41	83.66
AE-CNN-BiLSTM	84.19	84.77	84.65	84.17
CNN-BiLSTM-AE	84.17	84.77	84.67	84.15
AE-CNN-BiLSTM-AE	**85.7**	**84.7**	**87.5**	**85.1**

损失函数作为衡量模型预测能力的关键,也会对模型的训练效果造成影响,因此根据不同的场合选取合适的损失函数至关重要。考虑到 MSE 损失和交叉熵损失作为分类损失能够衡量模型的预测与真实值之间的偏离程度,KL 散度作为分布损失能够比较两个概率分布之间的接近程度,因此本文根据网络入侵检测的应用场景,引入权重系数作为损失分量的重要性衡量指标,结合 MSE 损失和 KL 散度定义了损失函数 L_1 作为数据增强的评估标准,结合交叉熵损失和 KL 散度定义了损失函数 L_2 作为所提模型的评估标准。为了使模型的训练能够达到最优效果,在定义损失函数时需要选择合适的权重系数,在实验中,对 L_1 和 L_2 分别选取 11 组不同的权重系数进行测试。

采用不同 L_1 损失函数的 AE-CNN-BiLSTM-AE 模型在 KDDTest+上的实验结果如表 4 所示,可见 MSE 损失和 KL 散度的权重系数为 0.5 和 0.5 时,检测效果最好。

表 4　采用不同 L_1 损失函数的 AE-CNN-BiLSTM-AE 模型在 KDDTest+上的实验结果

Table 4　Experimental results of AE-CNN-BilSTM-AE model with different L_1 loss functions on KDDTest+　　(%)

MSE 损失权重	KL 散度权重	准确率 Accuracy	精度 Precision	召回率 Recall	F1 分数
0	1.0	83.66	84.38	84.44	83.64
0.1	0.9	79.92	80.87	81.42	79.87
0.2	0.8	79.72	80.58	80.94	79.68
0.3	0.7	80.58	81.44	81.79	80.54
0.4	0.6	84.20	84.59	84.34	84.16
0.5	**0.5**	**85.7**	**84.7**	**87.5**	**85.1**
0.6	0.4	81.16	82.04	82.43	81.12
0.7	0.3	84.03	84.74	84.79	84.01
0.8	0.2	82.26	83.08	83.31	82.24
0.9	0.1	81.50	82.36	82.68	81.47
1.0	0	82.08	82.92	83.19	82.05

采用不同 L_2 损失函数的 AE-CNN-BiLSTM-AE 模型在 KDDTest+上的实验结果如表 5 所示,结果表明在交叉熵损失和 KL 散度的权重系数为 0.8 和 0.2 时,AE-CNN-BiLSTM-AE 模型的检测准确率和 F1 分数最高,因此本文采用式(14)与(17)所定义的损失函数作为模型的评估标准。

表 5　采用不同 L_2 损失函数的 AE-CNN-BiLSTM-AE 模型在 KDDTest+上的实验结果

Table 5　Experimental results of AE-CNN-BilSTM-AE model with different L_2 loss functions on KDDTest+　　(%)

交叉熵损失权重	KL 散度权重	准确率 Accuracy	精度 Precision	召回率 Recall	F1 分数
0	1.0	54.50	52.66	52.87	51.16
0.1	0.9	82.93	83.54	83.46	82.90
0.2	0.8	82.53	83.07	82.94	82.50
0.3	0.7	81.71	82.55	82.86	81.67
0.4	0.6	82.78	83.53	83.65	82.76
0.5	0.5	82.13	82.95	83.21	82.10
0.6	0.4	83.99	84.69	84.73	83.97
0.7	0.3	82.81	83.57	83.70	82.79
0.8	**0.2**	**85.7**	**84.7**	**87.5**	**85.1**
0.9	0.1	83.31	83.40	83.19	83.22
1.0	0	84.12	84.39	84.11	84.07

将所提方法与传统的机器学习方法对比,结果如图 8 所示,随机森林、朴素贝叶斯[15]、可支持向量机[3]都是经典机器学习方法,在准确率和 F1 分数上不尽理想,表明采用的深度学习模型具有一定的优势。而 XGBoost[4]是大规模并行 boosting tree 的工具,是一种集成多个弱分类器的机器学习算法,其准确率和 F1 分数达到了 84.25%和 83.87%,即便 XGBoost 优于大部分经典机器学习算法,但其测试结果还是略低于本文所采用的深度学习方法,充分验证了面对海量的高维网络流量数据,深度学习的强大算力优势。

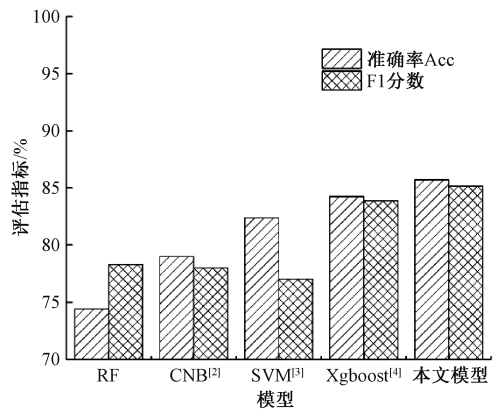

图 8　本文模型与其他机器学习模型的对比结果

Fig. 8　Comparison of the proposed model with other machine learning models

为了证明半自监督模型在入侵检测上的有效性,

图 9 给出了所提方法与其他深度学习方法在准确率和 F1 分数上的对比结果。可以看到,半自监督模型的准确率和 F1 分数明显优于 GB-RBM[16]、TSODE[17] 和 LSTM[18]模型,而 LCVAE[19] 模型的准确率虽然达到了 85.51%,略低于本文 85.7%的准确率,但是其 F1 分数仅有 80.78%,远不及本文 85.1%的 F1 分数。由于所提模型与其余模型相比在准确率和 F1 分数上具有一定的优势,说明利用 AE 进行数据增强和自监督特征增强的方法具有一定研究意义,而且该方法的优点在于提供了一种自监督特征增强的新思路,并通过实验仿真表明是切实可行的,能有效提高检测准确率。

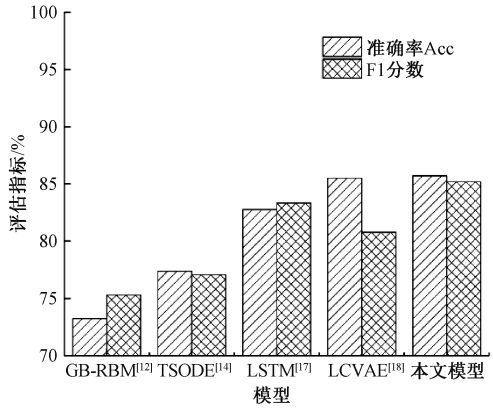

图 9 本文模型与其他深度学习模型的对比结果

Fig. 9 Comparison between the proposed model and other deep learning models

4 结 论

本文提出一种基于自监督特征增强的 CNN-BiLSTM 网络入侵检测模型,为避免离群值对检测结果的影响,在数据预处理阶段采用 IQR 异常值处理方法,利用自编码器对攻击类样本进行扩充,将 CNN-BiLSTM 模型提取的高维流量特征和自编码器生成的自监督特征组合输入到分类网络进行识别分类,解决了攻击样本和流量特征不足的问题,能够有效提高网络入侵的检测精度,改善对未知攻击的检测能力。由于近年来生成对抗网络(GAN)在生成领域展示出了非常大的优势,在接下来的工作中,将继续对半自监督网络入侵检测方法展开研究,把生成对抗网络应用到数据生成和自监督特征提取中,并在更新的入侵检测数据集和实测流量上进行下一步研究,从而提升模型检测性能和泛化性。

参考文献

[1] 苏醒. 基于网络行为的计算机网络安全预警与响应系统研究[J]. 电子测量技术, 2019, 42(21): 123-126.
SU X. Researching on computer network security early warning and response system based on network behavior[J]. Electronic Measurement Technology, 2019, 42(21): 123-126.

[2] 蹇诗婕, 卢志刚, 杜丹, 等. 网络入侵检测技术综述[J]. 信息安全学报, 2020, 5(4): 96-122.
JIAN SH J, LU ZH G, DU D, et al. Overview of network intrusion detection technology [J]. Journal of Cyber Security, 2020, 5(4): 96-122.

[3] GAO X, SHAN C, HU C, et al. An adaptive ensemble machine learning model for intrusion detection[J]. IEEE Access, 2019, 7: 82512-82521.

[4] VERMA P, ANWAR S, KHAN S, et al. Network intrusion detection using clustering and gradient boosting[C]. 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2018: 1-7.

[5] 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.

[6] ABOLHASANZADEH B. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features [C]. 2015 7th Conference on Information and Knowledge Technology (IKT). IEEE, 2015: 1-5.

[7] IERACITANO C, ADEEL A, MORABITO F C, et al. A novel statistical analysis and autoencoder driven intelligent intrusion detection approach [J]. Neurocomputing, 2020, 387: 51-62.

[8] 施媛波. 变分自编码器和注意力机制的异常入侵检测方法[J/OL]. 重庆邮电大学学报(自然科学版): 1-8[2022-03-11].
SHI Y B. Anomaly intrusion detection method based on variational autoencoder and attention mechanism[J/OL]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition): 1-8 [2022-03-11].

[9] 宁亚飞, 赵英亮, 吴美荣, 等. 时空卷积自编码网络异常行为检测[J]. 国外电子测量技术, 2020, 39(10):104-108.
NING Y F, ZHAO Y L, WU M R, et al. Study of spatiotemporal convolutional auto-encoder for anomaly detection [J]. Foreign Electronic Measurement Technology, 2020, 39(10):104-108.

[10] 于晓升, 许茗, 王莹, 等. 基于卷积变分自编码器的异常事件检测方法[J]. 仪器仪表学报, 2021, 42(5):151-158.
YU X SH, XU M, WANG Y, et al. Anomaly detection method based on convolutional variational auto-encoder[J].

Chinese Journal of Scientific Instrument, 2021, 42(5): 151-158.

[11] 来杰, 王晓丹, 向前, 等. 自编码器及其应用综述[J]. 通信学报, 2021, 42(9): 218-230.
LAI J, WANG X D, XIANG Q, et al. Review on autoencoder and its application [J]. Journal on Communications, 2021, 42(9): 218-230.

[12] 苏鹏, 王常顺, 卢萌萌. 基于变分自编码器的视频异常事件检测方法[J]. 电子测量与仪器学报, 2020, 32(10):179-185.
SU P, WANG CH SH, LU M M. Video anomaly detection and localization via variational autoencoder[J]. Journal of Electronic Measurement and Instrumentation, 2020, 32(10):179-185.

[13] DHANABAL L, SHANTHARAJAH S P. A study on NSL-KDD dataset for intrusion detection system based on classification algorithms [J]. International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4(6): 446-452.

[14] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE, 2009: 1-6.

[15] BELAVAGI M C, MUNIYAL B. Performance evaluation of supervised machine learning algorithms for intrusion detection[J]. Procedia Computer Science, 2016, 89: 117-123.

[16] IMAMVERDIYEV Y, ABDULLAYEVA F. Deep learning method for denial of service attack detection based on restricted boltzmann machine[J]. Big Data, 2018, 6(2): 159-169.

[17] FATANI A, ABD ELAZIZ M, DAHOU A, et al. IoT intrusion detection system using deep learning and enhanced transient search optimization [J]. IEEE Access, 2021, 9: 123448-123464.

[18] LI Z, RIOS A L G, XU G, et al. Machine learning techniques for classifying network anomalies and intrusions[C]. 2019 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2019: 1-5.

[19] XU X, LI J, YANG Y, et al. Toward effective intrusion detection using log-cosh conditional variational autoencoder[J]. IEEE Internet of Things Journal, 2020, 8(8): 6187-6196.

作者简介

梁欣怡,2021 年于南京信息工程大学获得学士学位,现为南京信息工程大学硕士研究生,主要研究方向为信号处理。
E-mail: 943430161@qq.com

Liang Xinyi received her B. Sc. degree from Nanjing University of Information Science & Technology. Now she is a M. Sc. candidate at Nanjing University of Information Science & Technology. Her main research interest includes signal processing.

行鸿彦(通信作者),1983 年于太原理工大学获得学士学位,1990 年于吉林大学获得硕士学位,2003 年于西安交通大学获得博士学位,现为南京信息工程大学教授、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

Xing Hongyan (Corresponding author), received his B. Sc. degree from Taiyuan University of Technology in 1983, M. Sc. degree from Jilin University in 1990, and Ph. D. degree from Xi-an Jiaotong University in 2003. Now he is a professor and supervisor for Ph. D. student in Nanjing University of Information Science &Technology. His main research interests include weak signal detection, bio-medical signal collection and processing, and design of intelligent electronic measurement technology and instrument.