基于联合注意力机制和一维卷积神经网络-双向长短期记忆网络模型的流量异常检测方法

尹梓诺* 马海龙 胡 涛

(解放军信息工程大学信息技术研究所 郑州 450001)

摘 要：针对流量数据集中类别不平衡限制了分类模型对少数类攻击流量的检测性能这一问题，该文提出一种基于联合注意力机制和1维卷积神经网络-双向长短期记忆网络(1DCNN-BiLSTM)模型的流量异常检测方法。首先在数据预处理过程中利用BorderlineSMOTE方法对流量数据不平衡训练样本预处理，使得各类流量数据均衡，有助于后续模型对各类数据的充分训练。然后设计联合注意力机制和1DCNN-BiLSTM的模型对流量数据进行训练，提取流量数据的局部和长距离序列特征并进行分类，通过注意力机制将对分类有用的特征按其重要性赋予权值，提高对少数攻击类的检出率。实验结果表明，同几种现有方法相比，该文方法对NSL-KDD和CICIDS2017数据集的检测准确率最高(可达93.17%和98.65%)，对NSL-KDD数据集中的提权攻击(U2R)攻击流量的检出率至少提升13.70%，证明了该文方法提升少数类攻击流量检出率的有效性。

关键词：流量异常检测；类别不平衡；一维卷积神经网络-双向长短期记忆网络；注意力机制

中图分类号：TN915.08; TP393 **文献标识码**：A **文章编号**：1009-5896(2023)10-3719-10

DOI: 10.11999/JEIT220959

A Traffic Anomaly Detection Method Based on the Joint Model of Attention Mechanism and One-Dimensional Convolutional Neural Network-Bidirectional Long Short Term Memory

YIN Zinuo MA Hailong HU Tao

(*Institute of Information Technology*, *PLA Information Engineering University*, *Zhengzhou* 450001, *China*)

Abstract: Considering the problem that the class imbalance of traffic dataset limits the performance of the model to the minority class attack traffic, a traffic anomaly detection method based on the joint model of attention mechanism and One-Dimensional Convolutional Neural Network - Bidirectional Long Short Term Memory (1DCNN-BiLSTM) is proposed. First, in the data preprocessing, the BorderlineSMOTE method is used to preprocess the imbalanced traffic training data, so that the quantities of different categories are balanced, which is helpful for the model to train various types fully. Then, the joint model of attention mechanism and 1DCNN-BiLSTM is designed to extract the local and long-distance sequence features of the traffic data. The features useful for classification are assigned weights according to their importance through the attention mechanism, which makes the model improve the detection rate of attack classes. Experimental results show that the proposed method has the highest accuracy for NSL-KDD and CICIDS2017 datasets (up to 93.17% and 98.65%). The proposed method improves the detection rate of User to Root(U2R) attack traffic in NSL-KDD dataset by at least 13.70%, which proves the effectiveness of the proposed method in improving the detection rate of minority attack traffic.

Key words: Traffic anomaly detection; Class imbalance; One-Dimensional Convolutional Neural Network-Bidirectional Long Short Term Memory (1DCNN-BiLSTM); Attentional mechanism

收稿日期：2022-07-18；改回日期：2022-09-03；网络出版：2022-09-06

*通信作者： 尹梓诺 yinzinuo1997@163.com

基金项目：国家重点研发计划(2018YFB0804002)

Foundation Item: The National Key R&D Program of China (2018YFB0804002)

1 引言

基于网络的计算服务和应用程序在人们的生活中发挥着重要作用，越来越多的网络节点设备连接到互联网中。据统计研究数据库(Statista)估计，到2025年，将有7.5×10^{10}台设备连接到互联网，构成巨大的网络接入设备规模[1]。随着互联网规模指数级增大，网络攻击所使用的协议、操作系统和应用软件的缺陷和漏洞也在不断更新与增多。流量异常检测是保护网络和信息系统安全的有效手段，被广泛用于检测网络流量恶意行为[2]。

随着流量数据不断增加，研究人员引入机器学习方法对大规模流量数据进行分类和预测，实现流量异常检测。早期，研究人员基于传统机器学习方法进行流量异常检测，使用单个分类器[3,4]或融合多个分类器[5–8]进行检测。但研究发现以传统机器学习算法为基础的流量异常检测结果并不理想，其检测性能较依赖特征。大多强调特征工程和特征选择，具有较高的误报率[9]。

近年来，许多深度学习方法[10,11]通过神经网络的搜索空间从原始流量特征中自动提取高级特征，被应用到流量异常检测研究中，取得了一些较好的研究成果。董书琴等人[12]提出一种结合堆叠去噪自编码器和softmax的流量异常检测方法提高对NSL-KDD的检测性能。缪祥华等人[13]将密集连接卷积神经网络应用于流量异常检测，提升对KDD99数据集的检测准确率。Sivamohan等人[14]对长短期记忆网络(Long Short Term Memory, LSTM)、门控循环单元、双向长短期记忆网络3种循环神经网络在CICIDS2017数据集上的检测性能进行对比评估，发现双向长短期记忆网络(Bidirectional Long Short Term Memory, BiLSTM)的检测准确率最优。

大多数基于传统机器学习模型和深度学习模型的流量异常检测方法都需要大量样本数据进行充分学习来获取更好的检测效果。然而流量数据存在严重的类别不平衡，正常样本往往远多于异常样本，在异常样本中各攻击类流量数据所占的比例差别很大[15]。在异常数据较少、流量数据严重不平衡的情况下，将这种不平衡流量数据训练集直接输入传统分类模型进行学习和训练会导致多数类样本淹没少数类样本，少数威胁程度高的攻击流量有可能被错误检测为良性流量或其他攻击类别，这也对网络、设备、用户构成更高的风险。因此，为有效检测网络中的恶意流量，需要解决网络流量异常检测中的类别不平衡问题。

研究人员主要从数据和算法两个角度解决流量异常检测中的类别不平衡问题。在数据方面，主要通过重采样技术均衡各类流量数据，如合成少数类过采样技术 (Synthetic Minority Oversampling TEchnique, SMOTE)[16]、自适应合成抽样技术 [17]、平衡重采样技术[18,19]等。在算法方面，通过改进算法或使用集成方法[20–23]提升检测能力。但现有研究在少数类攻击流量的检出率方面还存在较大提升空间。为解决流量异常检测中的类别不平衡问题，本文提出一种基于联合注意力机制和1维卷积神经网络-双向长短期记忆网络(One-Dimensional Convolutional Neural Network-Bidirectional Long Short Term Memory, 1DCNN-BiLSTM)模型的流量异常检测方法，将数据增强技术与深度学习模型相结合提升少数攻击类的检出率。本文的主要贡献如下：

(1)本文提出一种基于联合注意力机制和1DCNN-BiLSTM模型的流量异常检测方法，融合数据不平衡处理技术和深度学习模型，从均衡数据和改进模型两方面出发，提高对高度不平衡流量数据的检测性能。

(2)本文设计一种联合注意力机制和1DCNN-BiLSTM的深度学习混合模型用于流量异常检测，分别利用1DCNN和BiLSTM提取网络流量数据的局部与长距离序列特征，同时在1DCNN的每个块和BiLSTM末端添加有效的注意力机制，着重关注对分类起重要作用的特征，提高对少数攻击类的检出率。

(3)本文使用NSL-KDD和CICIDS2017流量数据集进行实验，利用多种评估指标将所提方法与一些现有典型机器学习方法和在流量数据不平衡问题上效果较好的方法进行对比，实验结果表明，本文方法能够显著提升少数类攻击流量的检出率，在对不平衡流量数据的检测性能上表现出优越性。

2 基于联合注意力机制和1DCNN-BiLSTM模型的流量异常检测方法

2.1 流量异常检测框架

本文设计流量异常检测方法的目标是在恶意流量样本数量较少的情况下，对流量数据实现优越的检测性能。对此，本文所提流量异常检测方法结合了数据重采样技术和深度学习网络模型，所提方法的整体检测框架如图1所示，主要包含3个模块：数据预处理模块、流量异常检测模块和分类评估模块。

数据预处理模块负责对原始流量特征数据进行量化、归一化以及训练数据重采样等操作，量化和归一化使数据能满足深度学习模型的输入格式要求，更有利于模型的训练和检测，数据重采样能够使流量数据均衡，减轻原始数据类别不平衡对检测结果造成的影响和偏差。

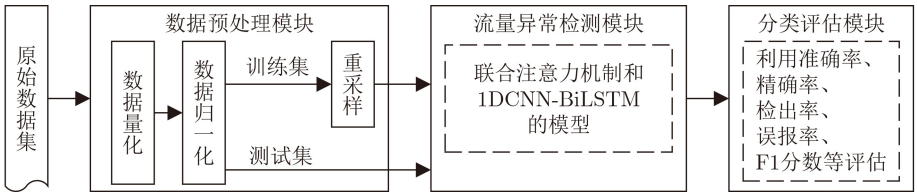

图 1 流量异常检测框架

在流量异常检测模块，本文对预处理后的流量数据设计一种联合注意力机制和1DCNN-BiLSTM的模型进行深度流量特征提取和学习，同时对特征重要性予以考虑，有效检测数量少且威胁程度高的攻击流量。

在分类评估模块中，利用多种检测评价指标对模型的检测结果进行评估和分析。

2.2 数据预处理

(1)量化。流量特征数据包含非数值特征(如NSL-KDD数据集中'protocol type', 'service' 和'flag')，需要将这类特征转换为数值特征，本文采用LabelEncoder()函数进行标签编码。同时，样本类别标签也需转换为数字，对于二分类，将正常和攻击标签分别编码为0和1，对于多分类，将各攻击类型进行独热编码。

(2)归一化。为缩小流量数据集中特征值间的大小差异，避免数值量级差异和单位差异对检测结果的影响，保证检测结果有效，采用Min-Max归一化方法将各特征数据映射到[0,1]区间，其公式如式(1)所示

$$x_n = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

其中，x为特征列X的各特征值，$X_{\min}$和$X_{\max}$分别为特征列X的最小值和最大值。

(3)过采样。流量数据集中不同类别的样本数量差异显著，如在NSL-KDD数据集中，正常样本的数量是提权攻击(User-to-Root, U2R)样本的1 000多倍。深度学习分类器在模型训练过程中，对少数类攻击样本的特征学习不充分，会影响模型的检测性能。因此在数据预处理过程中，需要对少数类攻击流量过采样，使深度学习模型能充分有效地学习流量样本空间中每个类的边界，边界样本对泛化更重要，但它们更容易发生错误分类。对流量异常检测，需要新合成的攻击类样本处于类别边界附近来提供足够的信息用于学习和检测。本文利用borderlineSMOTE算法对异常流量样本进行过采样，首先识别边缘的攻击样本，然后重新生成攻击样本，最后将新生成的样本加入到流量数据训练集。

对于训练集中每个攻击样本x，计算其m个最近邻，若x的最近邻中，正常样本的数量多于攻击样本，那么x作为攻击类的边界样本很可能被错分为正常样本，需对这类边界样本过采样。在采样过程中，计算攻击样本x的k个最近邻攻击样本，并从中随机选择n个($1<n<k$)攻击样本，攻击类流量新样本的生成公式如式(2)所示

$$\boldsymbol{T}_n = \boldsymbol{T}_i + \mathrm{rand}(0,1) \times |\boldsymbol{T}_j - \boldsymbol{T}_i| \tag{2}$$

其中，$\boldsymbol{T}_n$为新生成的样本，$\boldsymbol{T}_i$为边界样本，$\boldsymbol{T}_j$为$\boldsymbol{T}_i$的邻居，rand(0,1)表示生成[0,1]区间的随机数。

2.3 联合注意力机制和1DCNN-BiLSTM的流量检测模型

流量数据本质上可以看作具有前后关联关系的序列数据，因此流量特征数据也具有显著的前后序列依赖关系和同一序列不同特征间的关联关系，如NSL-KDD数据集的Probe攻击可能表现为流量特征在一段时间的持续变化，对于该攻击类型，可以利用序列学习模型学习数据集中“基于主机的流量统计特征”和“基于时间的流量统计特征”来捕获时间前后流量数据的关联关系和深层特征进行检测。

为提升对少数攻击流量的检出率，本文设计一种联合注意力机制和1DCNN-BiLSTM的流量异常检测模型，对流量特征数据进行充分学习，有效提取其深层复杂特征。在该模型中，1DCNN适用于序列处理，可以实现更多非线性转换，对流量序列特征提供较强的局部特征学习能力。网络流量数据遵循时间序列模式，可以根据过去跨长距离的流量连接记录对当前流量连接记录进行分类，但1DCNN在长距离学习建模方面的能力有限。BiLSTM网络主要用于实现长距离序列特征学习。因此，将1DCNN提取的深度流量特征输入BiLSTM，进一步学习深度流量特征向量之间跨长距离的序列关联模式。为进一步提升模型对不平衡流量数据的检测性能，本文在模型中加入有效注意力层，附加在1DCNN网络的各池化层末端和BiLSTM网络末端，在学习特征的过程中，提高与流量类别相关特征的权重，使模型倾向于注意对异常流量检测更重要的特征。根据特征重要性调整权值，更全面地把握流量特征，提高少数类攻击流量的检出率。

联合注意力机制和1DCNN-BiLSTM的模型可

以学习正常和恶意流量数据序列的相关性与局部特征，具有多层结构，包括1维卷积层(Conv1D layer)、池化层(Pooling layer)、注意力层(Attention layer)、双向长短期记忆网络层(BiLSTM layer)、平铺层(Flatten layer)和全连接层(Full Connection layer, FC layer)等，其模型结构如图2所示。将经过预处理的流量数据通过输入层输入模型后，利用隐层计算得到检测结果，通过输出层输出。

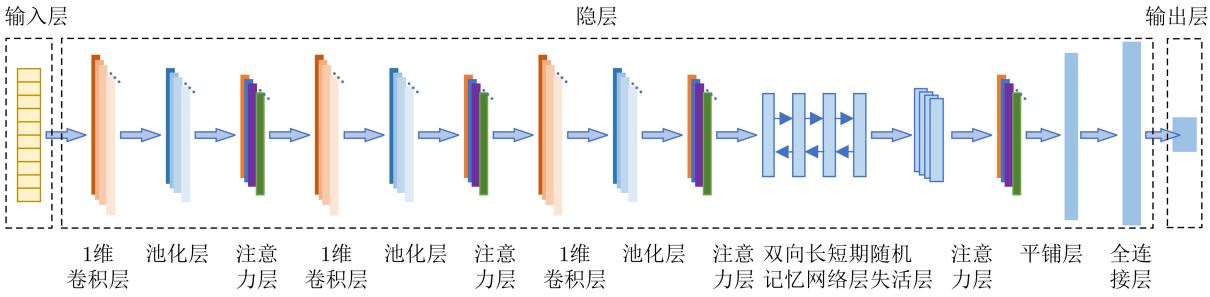

图 2 联合注意力机制和1DCNN-BiLSTM模型结构图

2.3.1 1DCNN

1DCNN是一种以1维网格形式获取序列数据进行特征识别的CNN。虽然1DCNN只有1个维度，但它在特征识别方面同样具有2DCNN的平移不变性优势。基于此，本文将流量特征数据构建为具有良性和恶意标签的序列数据，首先应用1DCNN实现对流量数据的局部特征提取。1DCNN模型通过堆叠1维卷积层和池化层来实现局部空间特征提取功能，解决局部特征丢失问题，其结构如图3所示。

1维卷积层是特征提取的关键，它通过训练流量数据得到一组具有最小损失的最优卷积核，利用卷积核(滤波器)自动提取复杂流量特征。流量数据的第i个样本可表示为m维特征向量$\boldsymbol{x}_i \in R^m$，多个连续向量$\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_j$可表示为$\boldsymbol{x}_{i:j}$，1维卷积仅在流量特征数据序列的垂直方向进行卷积，因此其卷积核的宽度即为流量特征的维度，通过使用滤波器$\boldsymbol{w}$对输入流量数据应用卷积操作来构建一个特征映射，实现局部空间特征提取，其计算公式如式(3)所示

$$\boldsymbol{h}_i = f\left(\boldsymbol{w} \otimes \boldsymbol{x}_{i:j} + \boldsymbol{b}\right) \tag{3}$$

其中，$\boldsymbol{b}$为偏置值，$f(\cdot)$表示卷积计算的非线性激活函数线性整流函数(Rectified Linear Unit, ReLU)。

池化层进一步聚集和保留了卷积层所提取的短期特征，得到最重要的特征。常用的池化方法是最大池化和平均池化。本文利用最大池化层将各卷积层特征向量的最大值合并，作为最终特征值。在1维卷积层和池化层进行操作后，得到了一个$1\times n$维的数据特征，很好地分析并保留流量数据序列的局部特征。

2.3.2 BiLSTM

BiLSTM是一种LSTM变体，它不仅具有LSTM模型的远距离序列学习能力，而且进一步改进LSTM，能够学习序列数据正向和反向的关联关系，使模型在分类问题上更具优势。本文利用输入流量数据训练BiLSTM的正向LSTM和反向LSTM，其结构如图4所示，包含输入层、正向隐层、反向隐层和输出层。正向LSTM提取输入的深度流量特征序列的正向特征，而反向LSTM与之相反，提取深度流量特征序列从后往前的反向特征。输出层对二者的输出数据进行整合。在时间步t利用BiLSTM模型对当前时刻的输入序列元素值$\boldsymbol{x}_t$进行特征提取的正向LSTM和反向LSTM计算如算法1所示。

时间步t上输出向量计算公式如式(4)所示

$$\boldsymbol{H}_t = \boldsymbol{W'}_{hf}\boldsymbol{h'} + \boldsymbol{W''}_{hb}\boldsymbol{h''} + \boldsymbol{b}_h \tag{4}$$

其中，$\boldsymbol{x}_t$为t时刻的输入序列，$\boldsymbol{C}_t$为t时刻的记忆细胞状态，$\boldsymbol{h}_{t-1}$为隐层状态，$\boldsymbol{W}_c, \boldsymbol{W}_f, \boldsymbol{W}_i, \boldsymbol{W}_o$分别为

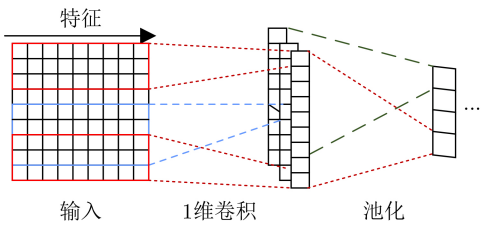

图 3 1DCNN结构图

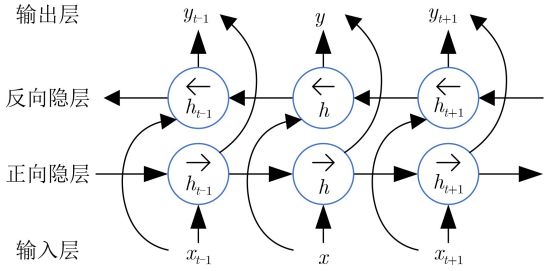

图 4 BiLSTM结构图

算法1　正向LSTM和反向LSTM计算

正向LSTM	反向LSTM
$\boldsymbol{i}_t=\sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_i)$	$\boldsymbol{i}_t=\sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t+1},\boldsymbol{x}_t]+\boldsymbol{b}_i)$
$\boldsymbol{f}_t=\sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_f)$	$\boldsymbol{f}_t=\sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t+1},\boldsymbol{x}_t]+\boldsymbol{b}_f)$
$\boldsymbol{g}_t=\tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_c)$	$\boldsymbol{g}_t=\tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t+1},\boldsymbol{x}_t]+\boldsymbol{b}_c)$
$\boldsymbol{C}_t=\boldsymbol{i}_t\odot\boldsymbol{g}_t+\boldsymbol{f}_t\odot\boldsymbol{C}_{t-1}$	$\boldsymbol{C}_t=\boldsymbol{i}_t\odot\boldsymbol{g}_t+\boldsymbol{f}_t\odot\boldsymbol{C}_{t+1}$
$\boldsymbol{o}_t=\sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_o)$	$\boldsymbol{o}_t=\sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t+1},\boldsymbol{x}_t]+\boldsymbol{b}_o)$
$\boldsymbol{h'}_t=\boldsymbol{o}_t\odot\tanh(\boldsymbol{C}_t)$	$\boldsymbol{h''}_t=\boldsymbol{o}_t\odot\tanh(\boldsymbol{C}_t)$

记忆细胞状态、遗忘门、输入门和输出门的权重矩阵，$\boldsymbol{b}_c,\boldsymbol{b}_f,\boldsymbol{b}_i,\boldsymbol{b}_o$分别表示对应偏置，$\odot$表示两个向量对应元素相乘操作，$\boldsymbol{W'}_{hf},\boldsymbol{W''}_{hb}$分别表示正向和反向输入到隐层权重矩阵。

BiLSTM有效利用网络流量前后数据中存在的时序特征来改进模型训练，使模型全面学习序列特征。

2.3.3 注意力机制

注意力机制的原理是：在大量信息中，将有限的注意力资源聚焦于需要关注的少数关键信息上，忽略无用和不相关信息，对更关键重要的信息进行特征提取。在流量异常检测领域，通过引入注意力机制，对用于检测攻击的不同流量特征赋予对应的权重，更有利于提高少数攻击样本的检出率。本文在1DCNN网络和BiLSTM网络中分别引入注意力机制。对于1DCNN，将注意力层附加在卷积块末端，改善卷积神经网络仅关注局部特征而导致对全局特征学习不准确的情况。对于BiLSTM，注意力机制对其隐层向量输出表达式进行加权求和，检测效果更优。注意力机制通过分配概率代替原始随机分配权重。将卷积块或BiLSTM得到的隐层向量h_t作为注意力层的输入，其处理过程如式(5)—式(7)所示

$$\boldsymbol{e}_t=\boldsymbol{u}\tanh(\boldsymbol{w}_t\boldsymbol{h}_t+\boldsymbol{b}) \tag{5}$$

$$\boldsymbol{a}_t=\frac{\exp(\boldsymbol{e}_t)}{\sum_{j=1}^{t}\exp(\boldsymbol{e}_j)} \tag{6}$$

$$\boldsymbol{s}_t=\sum_{t=1}^{l}\boldsymbol{a}_t\boldsymbol{h}_t \tag{7}$$

其中，$\boldsymbol{a}_t$为权重，$\boldsymbol{s}_t$为对$\boldsymbol{h}_t$加权求和得到的高级流量特征。最后将$\boldsymbol{s}_t$输入到全连接层，得到检测结果。

3　实验分析

3.1 实验环境、评估指标和超参数配置

本文的所有实验均在一台具有32 GB内存、Intel Core i7-8700 3.20 GHz CPU和Nvidia GeForce GT 730 GPU的台式机上进行，使用Python3.5编程，对NSL-KDD和CICIDS2017数据集进行实验，以评估其检测多种经典攻击和现代攻击的有效性。本文使用评估指标准确率(Accuracy)、精确率(Precision)、检出率(Detection Rate, DR)、误报率(False Positive Rate, FPR)和F1-score来评估模型的检测性能。评估指标的计算公式如式(8)所示

$$\left.\begin{aligned}
&\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}\\
&\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}\\
&\text{DR}=\frac{\text{TP}}{\text{TP}+\text{FN}}\\
&\text{FPR}=\frac{\text{FP}}{\text{FP}+\text{TN}}\\
&\text{F1-score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}
\end{aligned}\right\} \tag{8}$$

其中，TP为正确预测为攻击样本的数量，TN为正确预测为正常样本的数量，FP为错误预测为攻击样本的数量，FN为错误预测为正常样本的数量。

由于深度学习模型具有参数化特性，模型的训练和检测结果受参数的影响较大。本文设计的联合注意力机制和1DCNN-BiLSTM的模型结构同2.3节的结构图一致。3个卷积层的卷积核数为64-128-32，3个最大池化层的poolsize设置为1。池化层的输出作为注意力层输入，1DCNN网络中最后一个注意力层的输出作为BiLSTM网络的输入，BiLSTM包含64个单元，其后的dropout层参数为0.5，用于防止过拟合，其后附加注意力层，之后连接Flatten平铺层用于将多维输出1维化，输出层使用全连接层，将模型的输出向量转换为类别标签的维度。对于二分类，输出层包含1个单元，激活函数为sigmoid，用于区分正常和攻击类型；对于多分类，输出层包含n个单元(n为样本类别数)，激活函数为softmax，用于区分多个攻击类型。经过多次调整模型的超参数，得到使模型学习效果最佳的超参数配置。模型的超参数配置为：训练过程损失函数为categorical_crossentropy，采用Adam优化器，学习率为0.001, Epoch=30, Batchsize=32。

3.2 数据集

NSL-KDD数据集KDD CUP99数据集的改进，广泛用于流量异常检测。尽管还有其他更新的数据集，但它仍然被许多最先进的流量异常检测文献用于性能评估。它删除了数据集中的冗余记录，包含正常样本和4种攻击样本，攻击类别包含DoS, Probe, U2R, R2L。该数据集包含41维流量特征和

1维类别标签。在实验中，将KDDTrain+_20 Percent作为训练集，KDDTest+为测试集。其数据分布如表1所示。

CICIDS2017数据集由通信安全机构(Communications Security Establishment, CSE)与加拿大网络安全研究所(Canadian Institute for Cybersecurity, CIC)在2017年收集，是一个具有复杂现代攻击类型的流量数据集。该数据集含有3 119 345个网络传输样本、78维流量特征和1维类别标签，包含正常样本和14种攻击样本。此外，该数据集中包含一些标签和特征缺失的样本，删除这些样本后，共得到2 824 876个样本，其数据分布如表2所示。

3.3 二分类实验

本节基于NSL-KDD和CICIDS2017数据集对流量异常检测方法进行二分类实验。

为验证本文模型的检测性能，实验先比较了3种机器学习的典型分类方法以及在类别不平衡问题上当前较流行且检测效果较好的3种模型在NSL-KDD数据集上的检测性能。3种机器学习方法分别为典型的随机森林(Random Forest, RF)、多层感知机(MultiLayer Perceptron, MLP)和组合模型1DCNN-BiLSTM，3种文献模型分别为文献[14]提出的BiLSTM模型，文献[19]提出的深度CNN模型以及文献[23]提出的朴素贝叶斯决策表和多目标进化特征选择(Naïve Bayes Decision Table and Multi Objective Evolutionary Feature Selection, DTNB+MOEFS)组合模型。同时为验证BorderlineSMOTE的有效性，本实验还比较了模型在不采用重采样方法、采用随机过采样(Random Over-Sampling, ROS)方法以及采用BorderlineSMOTE过采样方法3种情况下对NSL-KDD数据集的检测性能。其中，ROS方法的采样思路是：从少数类攻击流量样本中随机采样复制样本，使少数类流量与多数类流量的样本量相同，从而得到新的均衡数据集。实验结果如表3所示。

从表3中观察到，在异常流量样本少于良性样本的情况下，本文方法对攻击流量取得了最高的检测准确率、精确率、检出率和F1-score，分别为93.17%, 93.52%, 94.55%, 94.03%。本文方法的检测误报率为8.64%，相比其他模型中最低的误报率仅增加了0.69%。对于本文模型，采用BorderlineSMOTE过采样方法的检测准确率相比不采用重采样提升了4.76%，相比采用ROS过采样方法提升了4.05%。MLP和RF模型的检测准确率较低，分别为80.88%和79.92%。这是由于这两种模型在样本量较少的情况下，挖掘能力略差。因此在仅用原始训练集KDDTrain 20%样本构成的训练集KDDTrain+_20Percent进行学习时，准确学习数据特征的能力较差，检测准确率低。相比于CNN模型[19]、BiLSTM模型[14]、DTNB+MOEFS模型[23]和组合模型1DCNN-BiLSTM，本文模型的检测准确率分别提升了10.65%, 7.08%, 6.39%和4.61%，在精确率方面也分别提升了2.00%, 1.06%, 5.68%和5.77%，在检出率方面分别提升了19.74%, 12.61%, 9.69%和6.33%，检出率大幅提升，说明本文方法在区分正常流量和异常流量方面的检测性能优于其他模型。综上分析，本文方法对经典NSL-KDD数据集进行检测时，在正常流量与异常流量不平衡的情况下，检测性能优于其他方法，这验证了本文方法面对经典流量数据集正常和异常数据不平衡问题上的有效性。

为了评估本文方法检测新型现代攻击样本的有效性，实验在CICIDS2017数据集上进行进一步验证。由于CICIDS2017数据集数据量较大，其中良性样本占80.30%，攻击样本占19.70%。为提高实验效率，抽取10%的数据作为实验数据，训练集和测试集按照7:3的比例划分。由于上一实验已证明BorderlineSMOTE过采样方法的有效性，因此，本实验仅对比不同模型的检测性能，实验结果如表4所示。

表 1 NSL-KDD数据集数据分布

类别	正常流量	DoS	Probe	U2R	R2L	总计
KDDTrain+_20Percent	13 449	9 234	2 289	11	209	25 192
KDDTest+	9 711	7 458	2 116	200	3 059	22 544

表 2 CICIDS2017数据集数据分布

类别	正常流量	Bot	DDoS	DoS GoldenEye	DoS Hulk	DoS Slowhttptest	DoS Slowloris	FTP-Patator
数量	2 271 320	1 956	128 025	10 293	230 124	5 499	5 796	7 935

类别	Heartbleed	Infiltration	PortScan	SSH-Patator	Web Attack Brute Force	Web Attack-SQL Injection	Web Attack-XSS	总计
数量	11	36	158 804	2 897	1 507	21	652	2 824 876

表 3　基于NSL-KDD数据集的二分类检测结果(%)

模型	采样方法	准确率	精确率	检出率	误报率	F1-score
MLP		78.73	90.58	69.97	9.69	78.92
	ROS	79.08	92.03	69.34	8.04	79.04
	BorderlineSMOTE	79.92	93.26	71.02	6.72	79.82
RF		80.17	90.84	72.46	9.65	80.62
	ROS	80.52	90.79	73.21	9.81	81.06
	BorderlineSMOTE	80.88	90.18	74.53	10.73	81.61
CNN[19]		81.43	89.04	76.91	12.58	82.38
	ROS	82.17	90.65	76.55	9.95	83.07
	BorderlineSMOTE	84.20	91.69	78.96	8.87	84.91
BiLSTM[14]		85.18	80.16	89.03	17.73	84.36
	ROS	86.73	92.49	83.46	8.96	87.74
	BorderlineSMOTE	87.01	92.54	83.96	7.95	88.04
DTNB+MOEFS[23]		83.04	82.79	82.29	16.73	82.54
	ROS	85.96	86.02	85.14	14.09	85.58
	BorderlineSMOTE	87.57	88.49	86.20	13.30	87.33
1DCNN-BiLSTM		84.37	82.16	85.05	13.56	83.58
	ROS	86.13	84.92	86.23	12.74	85.57
	BorderlineSMOTE	89.06	88.42	88.92	10.34	88.67
本文模型		88.94	90.90	89.53	11.84	90.20
	ROS	89.54	91.66	89.78	10.79	90.71
	BorderlineSMOTE	93.17	93.52	94.55	8.64	94.03

表 4　基于CICIDS2017数据集的二分类检测结果(%)

模型	准确率	精确率	检出率	误报率	F1-score
MLP	94.01	86.94	93.00	5.61	89.87
CNN[19]	95.68	91.36	93.43	3.29	92.38
BiLSTM[14]	98.14	92.86	96.85	3.74	94.81
RF	96.61	93.28	94.20	3.51	93.74
DTNB+MOEFS[23]	96.80	97.40	96.70	3.70	97.05
联合注意力机制和1DCNN-BiLSTM模型	98.65	97.21	99.77	3.07	98.47

由表4可以观察到，本文的流量异常检测的准确率相比RF模型提升了2.11%，相比MLP模型提升了4.94%，相比深度CNN模型[19]提升了3.10%，相比BiLSTM模型[14]提升了0.52%，相比DTNB+MOEFS模型[23]提升了1.91%，检测效果有明显提升。所提模型的检测精确率、检出率和F1-score均高于其他方法，FPR均小于其他方法。进一步验证了本文方法对新型数据集中不平衡流量样本的检测有效性。

3.4　多分类实验

为了验证本文的流量异常检测方法区分不同攻击类型的检测有效性，实验基于NSL-KDD数据集进行模型多分类性能评估。实验对比了机器学习典型算法MLP, RF和现有流行的3种面对流量数据类别不平衡问题效果较好的模型CNN[19], BiLSTM[14], I-SiamIDS[22]，不同模型的检测性能见图5—图8。

图5为6种方法的检出率对比图，其中蓝色、红色、绿色、紫色、黄色分别表示各模型对正常流量、DoS攻击流量 、Probe攻击流量、U2R攻击流量、R2L攻击流量的检出率。由图5可以观察到，本文方法在正常类流量数据的检出率上略小于其他算法，其检出率为92.07%，相对于最高的RF算法，降低了5.46%。但对DoS攻击流量以及样本量较少的U2R和R2L攻击流量，本文方法相比其他方法均取得了最高的检出率，分别为93.66%, 83.00%和84.66%。对U2R攻击流量的检出率，本文方法相比其他方法至少提升了13.70%。对R2L攻击流量的

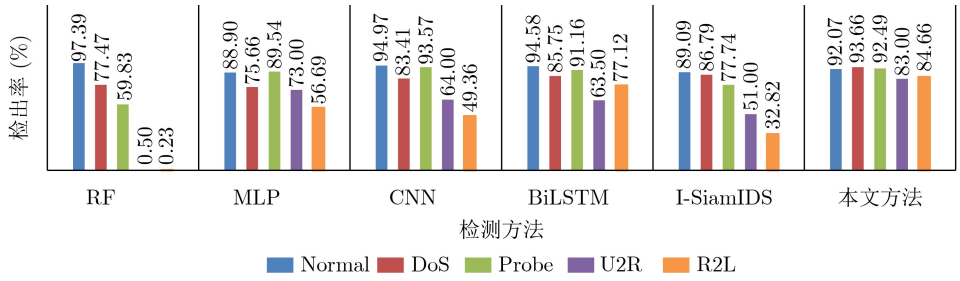

图 5 基于NSL-KDD数据集的多分类检出率

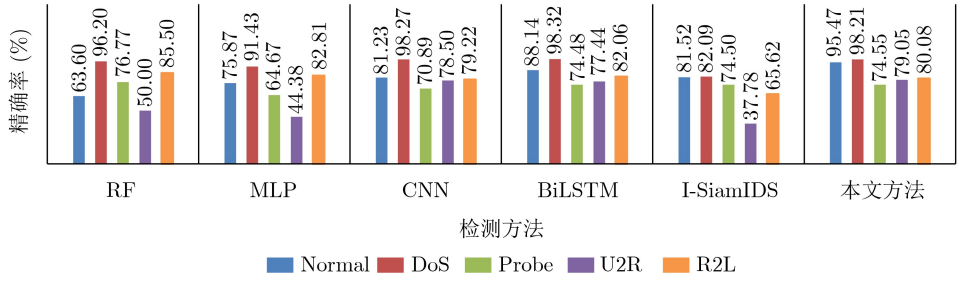

图 6 基于NSL-KDD数据集的多分类检测精确率

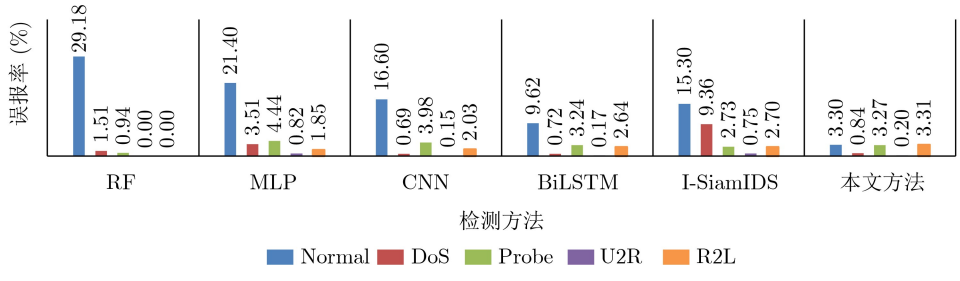

图 7 基于NSL-KDD数据集的多分类检测误报率

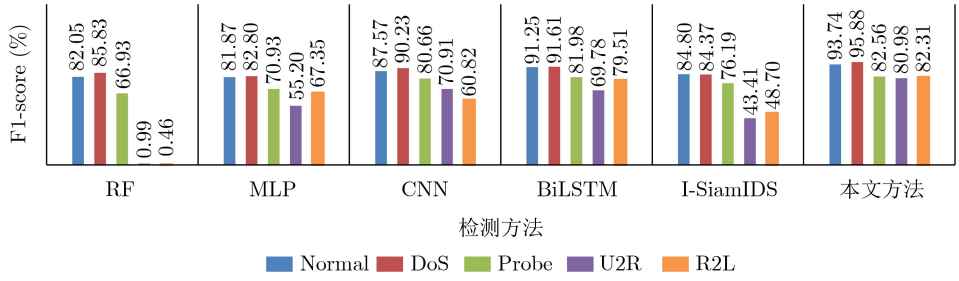

图 8 基于NSL-KDD数据集的多分类检测F1-score

检出率，本文方法相比其他方法至少提升了9.78%。这一结果表明，与基准分类器相比，本文方法能够提高对样本量较少的各攻击样本的检出率，对少数类攻击样本的识别能力较好。

图6为6种方法的检测精确率对比图，其中各颜色分别表示模型对各类别流量的检测精确率。由图6可知，本文方法对DoS, Probe, R2L的检测精确率分别为98.21%, 74.55%, 80.08%，虽略小于个别分类器，但在所有分类模型中，本文方法对正常类和样本量较少的U2R攻击流量均取得了最高的精确率，分别为95.47%和79.05%。对于正常类和4种攻击类，本文方法相对其他方法的检测精确率更稳定。这一结果表明，本文方法对各类样本的检测精确程度良好。

图7为6种方法的多分类检测误报率对比图，其中各颜色分别表示模型对各类别流量的检测误报率。由图可知，本文方法对正常类、DoS类均取得了最低的检测误报率，分别为3.30%和0.84%。对Probe, U2R, R2L 3类的检测误报率略高于个别基准模型，这是由于在检测过程中，本文方法将一定量正常类和DoS攻击流量数据误判为这3类数据，导致检测误报率略高，但本文方法对5类数据的误

报率较小，均不超过4%。其他方法对正常样本的检测误报率较高，这是由于其他对比方法在检测过程中将一些攻击样本识别为正常样本，识别不准确，检测误报率较高，这也说明所对比其他算法无法有效应对数据不平衡问题。

图8是6种方法的多分类检测F1-score对比图，其中各颜色分别表示模型对各类别流量的检测F1-score。F1-score更能反映模型的整体检测性能。由图8可知，本文方法对各类流量数据均取得最高的F1-score，对正常流量，DoS, Probe, U2R, R2L攻击流量的检测F1-score分别为93.74%, 95.88%, 82.56%, 80.98%, 82.31%。这表明本文方法在保证检测精确率的同时，提升了对少数攻击流量数据的检出率。

从图5—图8的各项性能参数对比可以看出，本文模型无论是对样本量较多的正常流量和DoS, Probe攻击流量，还是对样本量较少的U2R, R2L攻击流量，在各个检测性能指标上都较好，虽然在个别评价指标上略差于某个分类器，但在少数攻击流量的检出率等性能上效果最好，相比其他典型模型具有明显提升，能够对少数类攻击流量有效分类。由此验证了本文方法对不平衡流量数据多分类任务的有效性和优越性。

4　结束语

考虑到流量异常检测中类别不平衡问题严重影响了对攻击流量数据的检测准确率和对少数攻击类流量的检出率，本文提出了一种基于联合注意力机制和1DCNN-BiLSTM模型的流量异常检测方法，该方法结合了类不平衡处理技术BorderlineSMOTE和混合深度学习模型。通过BorderlineSMOTE生成新的攻击数据，使各类数据均衡。同时设计了联合注意力机制和1DCNN-BiLSTM的模型对流量数据进行训练，充分提取流量数据的局部特征和长距离序列特征，更好地学习流量特征数据的前后关联关系，并对特征按重要性赋予权重，充分发挥重要特征在流量异常检测中的作用，从而提高检测准确率和检出率。本文使用NSL-KDD和CICIDS2017数据集进行训练与测试。实验结果表明，与一些现有典型的和较为流行的机器学习算法相比，本文方法在二分类和多分类的检测准确率、精确率、检出率、误报率和F1-Score 5种性能评估指标上取得了良好的效果，验证了本文方法检测不平衡攻击流量的有效性。在未来的研究中，将探索其他检测模型和方法，提高对恶意流量的检测性能，以增强模型用于实际网络流量的可能性。

参 考 文 献

[1] Statista Research Department. Number of internet of things (IoT) connected devices worldwide from 2019 to 2021, with forecasts from 2022 to 2030[EB/OL]. https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/, 2022.

[2] SU Yu, QI Kaiyue, DI Chong, *et al*. Learning automata based feature selection for network traffic intrusion detection[C]. 2018 IEEE Third International Conference on Data Science in Cyberspace, Guangzhou, China, 2018: 622–627. doi: 10.1109/DSC.2018.00099.

[3] SYARIF I, PRUGEL-BENNETT A, and WILLS G. Unsupervised clustering approach for network anomaly detection[C]. 4th International Conference on Networked Digital Technologies, Berlin, Germany, 2012: 135–145. doi: 10.1007/978-3-642-30507-8_13.

[4] BO Li and YUAN Chenyuan. The research of intrusion detection based on support vector machine[C]. 2009 International Conference on Computer and Communications Security, Hong Kong, China, 2009: 21–23. doi: 10.1109/ICCCS.2009.43.

[5] TENGL S, ZHANG Zhenhua, TENG Luyao, *et al*. A collaborative intrusion detection model using a novel optimal weight strategy based on genetic algorithm for ensemble classifier[C]. 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design. Nanjing, China, 2018: 761–766. doi: 10.1109/CSCWD.2018.8465148.

[6] SORNSUWIT P and JAIYEN S. Intrusion detection model based on ensemble learning for U2R and R2L attacks[C]. 2015 7th International Conference on Information Technology and Electrical Engineering, Chiang Mai, Thailand, 2015: 354–359. doi: 10.1109/ICITEED.2015.7408971.

[7] NEGANDHI P, TRIVEDI Y, and MANGRULKAR R. Intrusion detection system using random forest on the NSL-KDD dataset[C]. Emerging Research in Computing, Information, Communication and Applications, Singapore, 2019: 519–531. doi: 10.1007/978-981-13-6001-5_43.

[8] KORONIOTIS N, MOUSTAFA N, SITNIKOVA E, *et al*. Towards developing network forensic mechanism for botnet activities in the IoT based on machine learning techniques[C]. International Conference on Mobile Networks and Management, Cham, Switzerland, 2018: 30–44. doi: 10.1007/978-3-319-90775-8_3.

[9] D'HOOGE L, WAUTERS T, VOLCKAERT B, *et al*. Inter-dataset generalization strength of supervised machine learning methods for intrusion detection[J]. *Journal of Information Security and Applications*, 2020, 54: 102564. doi: 10.1016/j.jisa.2020.102564.

[10] TANG T A, MHAMDI L, MCLERNON D, *et al*. Deep learning approach for network intrusion detection in software defined networking[C]. 2016 International Conference on Wireless Networks and Mobile Communications, Fez, Morocco, 2016: 258–263. doi: 10.1109/WINCOM.2016.7777224.

[11] SHONE N, NGOC T N, PHAI V D, *et al*. A deep learning approach to network intrusion detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41–50. doi: 10.1109/TETCI.2017.2772792.

[12] 董书琴, 张斌. 基于深度特征学习的网络流量异常检测方法[J]. 电子与信息学报, 2020, 42(3): 695–703. doi: 10.11999/JEIT190266.
DONG Shuqin and ZHANG Bin. Network traffic anomaly detection method based on deep features learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 695–703. doi: 10.11999/JEIT190266.

[13] 缪祥华, 单小撤. 基于密集连接卷积神经网络的入侵检测技术研究[J]. 电子与信息学报, 2020, 42(11): 2706–2712. doi: 10.11999/JEIT190655.
MIAO Xianghua and SHAN Xiaoche. Research on intrusion detection technology based on densely connected convolutional neural networks[J]. *Journal of Electronics & Information Technology*, 2020, 42(11): 2706–2712. doi: 10.11999/JEIT190655.

[14] SIVAMOHAN S, SRIDHAR S S, and KRISHNAVENI S. An effective recurrent neural network (RNN) based intrusion detection via bi-directional long short-term memory[C]. 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021: 1–5. doi: 10.1109/CONIT51480.2021.9498552.

[15] EBENUWA S H, SHARIF M S, ALAZAB M, *et al*. Variance ranking attributes selection techniques for binary classification problem in imbalance data[J]. *IEEE Access*, 2019, 7: 24649–24666. doi: 10.1109/ACCESS.2019.2899578.

[16] 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: 321–357. doi: 10.1613/jair.953.

[17] HE Haibo, BAI Yang, GARCIA E A, *et al*. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 2008: 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[18] HE Haibo and GARCIA E A. Learning from imbalanced data[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(9): 1263–1284. doi: 10.1109/TKDE.2008.239.

[19] YU Yingwei and BIAN Naizheng. An intrusion detection method using few-shot learning[J]. *IEEE Access*, 2020, 8: 49730–49740. doi: 10.1109/ACCESS.2020.2980136.

[20] CHOWDHURY M M U, HAMMOND F, KONOWICZ G, *et al*. A few-shot deep learning approach for improved intrusion detection[C]. 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, USA, 2017: 456–462. doi: 10.1109/UEMCON.2017.8249084.

[21] CHORAŚ M and PAWLICKI M. Intrusion detection approach based on optimised artificial neural network[J]. *Neurocomputing*, 2021, 452: 705–715. doi: 10.1016/j.neucom.2020.07.138.

[22] BEDI P, GUPTA N, and JINDAL V. I-SiamIDS: An improved Siam-IDS for handling class imbalance in network-based intrusion detection systems[J]. *Applied Intelligence*, 2021, 51(2): 1133–1151. doi: 10.1007/s10489-020-01886-y.

[23] PANIGRAHI R, BORAH S, PRAMANIK M, *et al*. Intrusion detection in cyber-physical environment using hybrid Naïve Bayes-Decision table and multi-objective evolutionary feature selection[J]. *Computer Communications*, 2022, 188: 133–144. doi: 10.1016/j.comcom.2022.03.009.

尹梓诺：女，博士生，研究方向为网络空间安全、网络流量异常检测等.

马海龙：男，副研究员，研究方向为网络空间内生安全技术、网络威胁智能检测以及新型网络体系等.

胡 涛：男，助理研究员，研究方向为新型网络体系结构等.

责任编辑：余 蓉