文章编号:1003－6199(2025)02－0009－06 DOI:10.16339/j.cnki.jsjsyzdh.202502002

基于 ResNet-BiGRU 的入侵检测研究

刘东†, 柳毅

(广东工业大学计算机学院,广东 广州 510006)

摘 要:随着网络流量的爆炸性增长和入侵行为的不断多样化,传统的入侵检测方法在面对海量的网络数据时,存在特征提取不充分、模型存在过拟合、检测分类的准确率不足等问题,提出了一种基于ResNet-BiGRU网络的入侵检测算法。该算法首先使用SMOTE过采样重构训练数据集以缓解正负类样本不平衡问题,提出了一种网络并行的结构充分提取特征,将卷积神经网络和双向门控单元同时提取空间尺度特征和时间序列特征,为了将空间尺度特征和时间尺度特征更好地聚合表达,设计了一个Transformer混合融合块,最后利用Softmax函数进行分类。为验证模型的有效性,在UNSW-NB15数据集进行实验,对比结果表明该模型在二分类和多分类任务上检测准确率分别达到94.1%和82.6%,结果表明提出的算法具有较高的性能和有效性。

关键词:网络安全; 双向门控循环单元; 残差网络; 混合融合块; 入侵检测

中图分类号:TP393.1 **文献标识码**:A

Research on Intrusion Detection Based on ResNet-BiGRU

LIU Dong†, LIU Yi

(Department of Computer Science, Guangdong University of Technology, Guangzhou, Guangdong 510006, China)

Abstract: With the explosive growth of network traffic and the continuous diversification of intrusion behaviors, traditional intrusion detection methods suffer from insufficient feature extraction, overfitting of models, and insufficient accuracy of detection and classification in the face of massive network data. An intrusion detection algorithm based on ResNet-BiGRU network is proposed. The algorithm first uses SMOTE oversampling to reconstruct the training dataset to alleviate the positive and negative class sample imbalance problem, proposes a network parallel structure to extract features adequately, combines the convolutional neural network and the bidirectional gating unit to extract both spatial-scale features and time-series features, and in order to express the spatial-scale features and the temporal-scaled features in a better way by aggregating them, designs a Transformer hybrid fusion block, and finally the Softmax function is utilized for classification. In order to verify the effectiveness of the model, experiments are carried out on the UNSW-NB15 dataset, and the comparison results show that the model achieves detection accuracy of 94.1% and 82.6% on binary and multi-classification tasks, respectively, and the results indicate that the proposed algorithm has high performance and effectiveness.

Key words: networks security; bi-directional gated recurrent units; residual networks; hybrid fusion block; intrusion detection

收稿日期:2023－10－20
基金项目:广东省重点领域研发计划项目(2021B0101200002)
作者简介:刘东(1998—),男,河南驻马店人,硕士生,研究方向:网络与信息安全。
† **通信联系人**,E-mail: 1874209996@qq.com

随着云计算和物联网的快速发展,网络安全问题也逐渐成为人们关注的热点问题。2019 年,大约 6.2 亿账户详细信息被黑客泄漏并在暗网上出售。新冠疫情期间,大多数人必须在家远程通过互联网进行工作和学习,互联网的用户数量急剧增加,网络流量呈现爆炸式增长,信息泄露以及电信诈骗层出不穷。2021 年上半年国家互联网应急中心发布的网络安全检测数据分析报告指出,我国境内感染计算机恶意程序的主机数量约 446 万,同比增长了 46.8%[1],网络攻击数量和种类日益频繁和复杂。因此,准确地检测网络攻击对个人和国家安全都十分重要。

传统的机器学习方法已广泛应用于网络异常流量检测系统,常用的有方法有:KNN(K-Nearest Neighbor)、贝叶斯网络模型、支持向量机、随机森林、决策树算法等,这些方法取得了不错的效果。文献[2]提出了一种主成分分析的方法处理数据样本,采用主成分分析法将高维数据压缩为低维数据,使用支持向量机训练分类进行入侵检测,准确率较高。文献[3]提出了一种基于数据挖掘分类和聚类技术的检测框架,将随机森林算法与加权 k-means 聚类算法相结合,建立一个混合框架进行入侵检测,有效地降低了误报率,但能否适应当前的网络环境有待考证。随着网络攻击日益频繁和攻击种类复杂化,这些传统的浅层机器学习方法已经难以应对复杂的网络攻击场景。

近年来,越来越多的研究人员将深度学习技术引入到入侵检测领域。如基于注意力机制的 Transformer[4]在文本分类、对话任务和其他自然语言处理领域取得了巨大成功,有研究者将 Transformer 引入入侵检测领域,文献[5]提出了一种混合神经网络 DdosTC 结构,结合了 Transformer 和卷积神经网络,用来检测 SDN 上的分布式拒绝服务,该方法有效提高了准确度,但未考虑数据集样本不平衡的问题。文献[6]提出了一种结合了卷积神经网络和双向长短期记忆神经网络的深度学习模型,整合数据的空间和时间特征的学习,获得了较高的检测率和相对较低的误报率,但在模型优化上还存在问题。

综上所述,深度学习在入侵检测领域虽然取得了不错的效果,但仍旧面临着一些挑战:(1)面对不断变化的海量网络数据,现有入侵检测系统的算法准确率还有待进一步提高;(2)目前模型存在过拟合的问题;(3)大多异常检测算法模型单一,考虑信息不够全面,普遍存在特征提取不充分的问题;(4)随着云计算和物联网的快速发展,网络环境日益复杂,对入侵检测模型的适用性也具有更高的要求。

针对上述存在的挑战,本文提出了一种基于 ResNet-BiGRU 网络的入侵检测算法。主要贡献如下:

(1)提出了一种 Transformer 的混合融合块(Transformer Hybrid Fusion Block,THFB),能够更好地融合时间序列特征和局部空间特征,提高分类准确率。

(2)将卷积神经网络和双向门控循环网络结合并行来提取局部空间特征和时间特征,实现了特征的充分提取,提高增加模型的分类能力和泛化能力。

(3)基于残差块的思想,设计多层卷积残差网络层叠深度提取特征,在 TFHB 中设计跳跃连接,有效防止梯度爆炸、梯度消失及网络退化问题。

1 模型设计

为了更快速、更好地进行入侵检测,本方法主要分为三个部分,首先是数据预处理的阶段,其次是特征提取阶段,将空间特征和时间特征使用 CNN 和 BiGRU 分开处理捕捉数据的不同尺度特征,再次是特征融合阶段,利用提出的 Transformer 混合融合块来融合时空特征,最后使用 Softmax 进行分类。该模型的整体流程图如图 1 所示。

1.1 数据预处理

UNSW-NB15 数据集包含不同的数据类型,无法直接用于模型的训练和测试,因此需要对数据集进行预处理。二分类时,将正常流量 Normal 用 0 表示,攻击流量 1 表示。多分类时,将 0、1、2、3、4、5、6、7、8、9 分别分配给 Normal、Generic、Exploits、Dos、Reconnaissance、Analysis、Backdoor、Shellcode、Worms、Fuzzers。

模型的训练和测试只接受数值型数据,首先采用 LabelEncoder 进行编码,将字符型特征转换为数值类型。特征值的范围不一样,特征之间无法直接比较,通过归一化处理,将每个特征值范围都映射到[0,1]内。

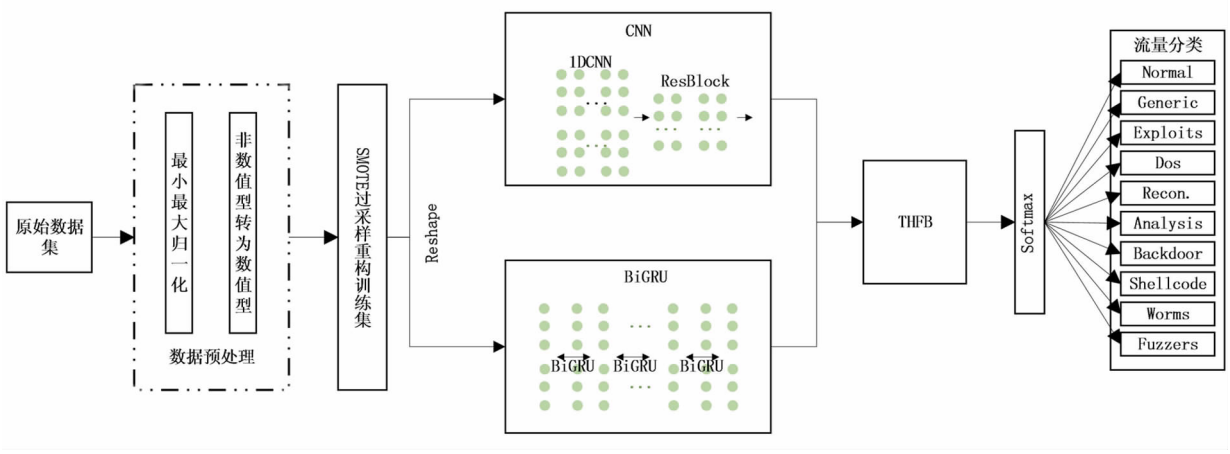

图 1　网络入侵检测模型

UNSW-NB15 数据集中 Worms 和 Shellcode 攻击样本数量较少，在多分类时使用 SMOTE 方法[7]对训练集中的 Worms 和 Shellcode 攻击样本进行过采样操作，为了防止 SMOTE 算法可能出现边缘分布问题，只生成少量的样本。如表 1 所示：

表 1　对数据集部分样本过采样

	Worms	Shellcode
过采样前	127	1054
过采样后	1181	2108

1.2　空间特征提取

为了提高空间特征的表达能力，设计多个卷积神经网络和残差块层叠结构，保证能够学习到复杂的特征变换，由于数据集是序列类数据，在残差块(Residual Block)中用核大小为 3x3 的一维卷积神经网络(Conv1D)进行处理，在每个卷积层输出后应用批量归一化(BatchNormalization)[8]和 ReLU[9]来进行非线性变换，结构如图 2 所示。

假设输入特征为 x，$F(x)$表示从输入特征到输出特征的残差映射，期望最优解为 $y = F(x)+x$，将输入特征 x 和卷积后的输出相加，进行特征融合，缓解梯度消失和梯度爆炸问题。其中激活函数 ReLU 的公式如下：

$$f(x) = \max(0,x) \tag{1}$$

其中，x 表示输入值，ReLU 的作用是保持非负数值不变。

跳跃连接的引入使得深层网络的性能明显提升，在加快收敛速度的同时避免了梯度消失和爆炸的问题。

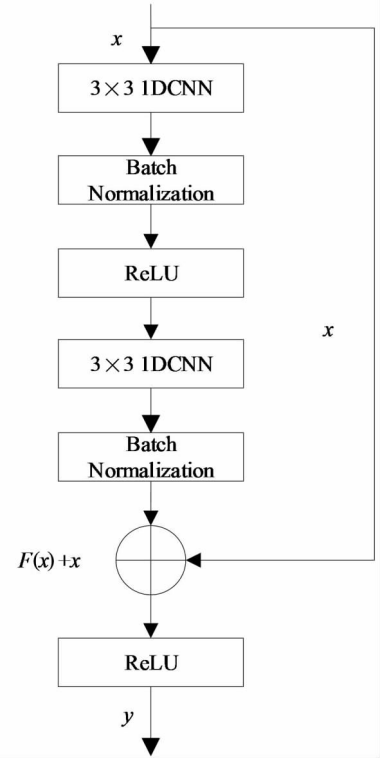

图 2　残差块结构

1.3　时间序列特征提取

门控循环单元 GRU(Gated Recurrent Unit)能够有效地传输时间序列并且解决了循环神经网络中存在的梯度爆炸和梯度消失的问题[10]，BiGRU 在 GRU 的基础上引入双向性，由一个前向 GRU 和一个后向 GRU 组成，能够同时获取前向和后向的所有信息，结构如图 3 所示。

输入层分别将输入特征数据 x_1、x_2、x_3 输入前向网络和后向网络中，$\overrightarrow{f_1}$、$\overrightarrow{f_2}$、$\overrightarrow{f_3}$ 表示 BiGRU 层输出的前向隐藏状态，$\overleftarrow{f_1}$、$\overleftarrow{f_2}$、$\overleftarrow{f_3}$ 表示 BiGRU 层的后向网络输出状态，将两个状态进行组合得到最终的

特征。

$$h_i = [\underset{f_i}{\rightarrow}, \underset{f_i}{\leftarrow}] \tag{2}$$

其中，h_i 表示将两个状态组合的最终输出结果，$i=1,2,3\cdots n$，n 表示输入数据的总数。

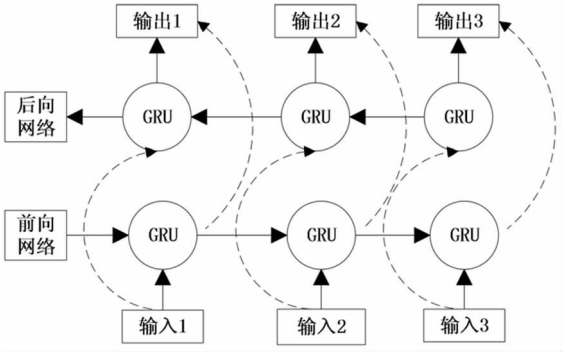

图 3 BiGRU 结构

1.4 特征融合

为了更好地整合 CNN 和 BiGRU 的优势，捕捉数据流中的时空模式，本文设计了 THFB 进行特征融合，整个模块如图 4 所示，THFB 模块包括一个多头注意力机制、一个门控单元和一个改进的前馈神经网络。

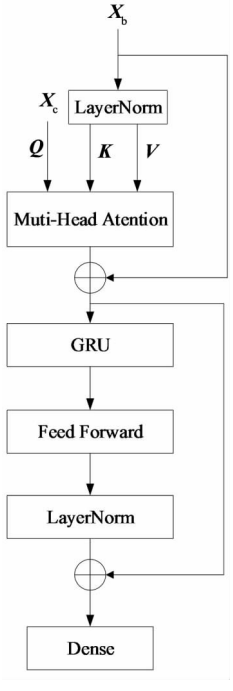

图 4 Transform 混合块结构

在多头注意力中将卷积神经网络的输出表示为高频先验查询(Query)，将 BiGRU 的输出作为键(Key)、值(Value)并计算相互注意力来细化获得的特征，建立空间和时间的关联性，自适应地调整空间和时间特征的权重，获取更具信息量的特征表示，多头注意力计算公式如下：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}})V \tag{3}$$

$$\text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Concat}(\text{head}_1, \text{head}_2,\cdots,\text{head}_n)W^O \tag{4}$$

$$\text{head}_i = \text{Attention}(\boldsymbol{Q}W_i^Q, KW_i^K, VW_i^V) \tag{5}$$

其中，$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 分别代表查询矩阵、关键字矩阵、值矩阵；dk 则表示关键字矩阵 $\boldsymbol{K}$ 的维度，$i=1,2,\cdots n$，n 为注意力头的个数，$\boldsymbol{W}^0$ 表示一个与多注意力头拼接后结果进行线性变换的矩阵。

对前馈神经网络提出改进，进一步聚合时空特征，引入门控机制，提高模型在多个抽象层次上的表示学习能力，充分提取空间信息和时间信息，增强模型的泛化能力。整个过程可以表述为如下公式：

$$X_{\text{fuse}} = \text{Muti}-\text{Attention}(\text{LN}(X_{\text{b}})+X_{\text{c}})+X_{\text{b}} \tag{6}$$

$$X_{\text{thfb}} = X_{\text{fuse}} + \text{FF}(\text{GRU}(X_{\text{fuse}})) \tag{7}$$

在 TFHB 中使用跳跃连接来防止梯度消失和网络退化的问题，帮助网络有效地学习复杂的特征和表示，加快网络收敛。

2 实验分析

2.1 实验数据集

为了验证本文提出模型的分类性能，选取了公开数据集 UNSW-NB15，该数据集基于真实的网络流量，由澳大利亚南威尔士大学于 2015 年开发。该数据集主要由 47 个属性特征和 2 个类别特征组成，包含正常流量和 9 种攻击流量：Generic、Exploits、Dos、Reconnaissance、Analysis、Backdoor、Shellcode、Worms、Fuzzers。UNSW-NB15 数据集分布信息如表 2 所示。

表 2 数据集分布信息

类型	训练集	测试集
Normal	64794	28026
Generic	41282	17589
Exploits	31197	13328
Dos	11395	4958
Recon.	9794	4139
Analysis	1893	784
Backdoor	1633	696
Shellcode	1054	457
Worms	127	47
Fuzzers	17022	7224

2.2　评价标准

为了更全面地验证提出模型的效果，本文使用准确率(Accuracy)、召回率(Recall)、精确度(Precision)、F1 分数(F1-score)、假负率(FNR)和 ROC 曲线来评估模型。Accuracy 是预测的正确样本数量和总样本的数量之比，其值越高代表检测效果越好。Precision 是指正确预测为正例的样本数量占所有预测为正例样本的比例。Recall 是正确预测为正例的样本数量和真实正例的数量之比。F1-score 是精确度和召回率之间的调和均值大小，体现模型综合性能。FNR 是被误分类为负类的样本中，实际上是正类的比例，计算公式如下：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{8}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{9}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{10}$$

$$\text{F1\_score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{11}$$

$$\text{FNR}=\frac{\text{FN}}{\text{FN}+\text{TP}} \tag{12}$$

ROC 曲线通过计算 AUC 来评估模型，AUC 越大，越接近左上角证明模型的性能越好。

2.3　结果分析

本节进行实验结果分析。第一，设置实验将不同的深度神经网络单一模型应用到该数据集上面进行对比来验证提出模型的有效性。

第二，为证明 THFB 模块的有效性，增加实验的可信度。本文构造与不同方式的多尺度记忆网络融合方法进行试验，同时将现有的一些多尺度融合入侵检测方法进行对比。CNN-BiGRU-A 表示本文的模型最后采用了 Add 函数逐项相加的方式融合特征，CNN-BiGRU-C 表示采用了 Concatenate 函数实现特征融合。文献[11]提出了一种多尺度记忆模块堆叠的方式提取特征，将 CNN 和 LSTM 多层次叠加引入残差网络，通过 Add 函数融合特征。文献[12]提出一种 CNN 和 BiLSTM 结合的入侵检测方法，将提取的特征通过 Concatenate 函数进行融合。不同模型的实验结果对比如表 3 所示。

表 3　不同模型的实验结果对比(%)

模型算法	二分类				多分类			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Transformer	88.6	85.9	95.1	91.6	70.1	63.7	70.1	63.8
GRU	90.6	92.3	92.9	92.6	78.1	76.2	78.2	74.8
LSTM	92.3	93.4	94.7	94.0	74.5	72.2	74.5	72.1
CNN-BiGRU-A	91.0	88.1	90.1	93.5	76.6	58.0	68.5	59.4
CNN-BiGRU-C	92.1	93.0	92.3	94.1	78.0	72.0	73.1	75.6
文献[11]	91.0	87.0	90.0	91.3	79.5	76.1	76.7	78.8
文献[12]	92.0	94.0	92.0	92.5	79.8	70.1	79.6	79.4
本文模型	94.1	96.1	94.5	95.3	82.6	82.9	82.6	80.0

从表 3 可以看出：

(1)本文提出的模型在准确率、精确度和 F1 分数这三个指标上的表现均优于对比方法，其中二分类准确率提高 1.8%左右，但在召回率上略低于 Transformer 和 LSTM 算法。对于多分类准确率达到了 82.6%，在召回率上也优于其他对比方法，代表该方法具有更低的漏报率。证明本文提出的模型在综合性能上优于单一模型，在入侵检测时能取得良好的效果。

(2)THFB 有效性分析：相较于 CNN-BiGRU-A、CNN-BiGRU-C、文献[11]与文献[12]的模型，THFB 融合方法在二分类和多分类准确率分别提升了 2.0%和 0.8%，在召回率、F1 分数本文模型的表现优于其他方法。验证了 THFB 的有效性且具备较高的性能。

图 5 是不同模型的 ROC 曲线图，可以看出本文模型得到的曲线相对于其他算法得到的曲线更加接近左上角，并且 AUC 面积大于其他算法得到的 AUC 面积，以上结果表明本文所提出的模型的表现要略优于其他方法。

最后将一些近年来的网络流量异常检测方法和本文提出的模型在同一数据集 UNSW-NB15 上的性能进行对比。

文献[13]是一种利用 PCA 和自编码器来提取特征，使用人工蜂群算法找到 SVM 最佳参数的方法。文献[14]提出一种融合多个卷积神经网络的方法进行入侵检测。文献[15]提出一种基于流量异常分析的入侵检测方法，在多个维度对数据进行

优化，通过随机森林进行验证。文献[16]提出将多层双向门循环单元神经网络和改进的前馈神经网络相结合训练。

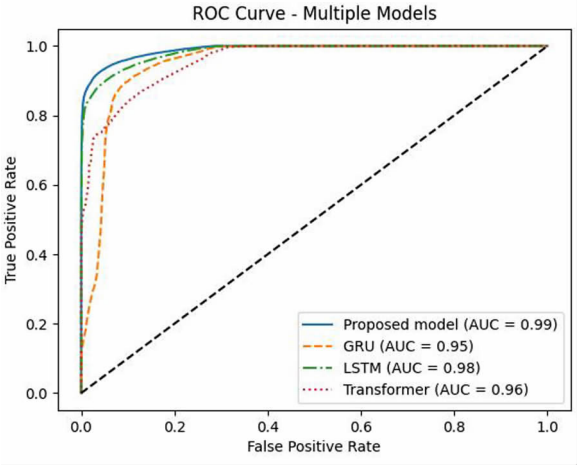

图 5 不同模型的 ROC 曲线图

表 4 在 UNSW-NB15 数据集上不同研究性能比较

模型与算法	评价指标	
	Accuracy	FNR
文献[13]	0.91	0.084
文献[14]	0.88	0.085
文献[15]	0.91	0.079
文献[16]	0.88	0.082
本文模型	0.94	0.054

从表 4 可以看到，本文的模型在准确率和假负率上表现优于其他方法，文献[13]没有考虑到数据冗余的问题，性能表现一般。文献[14]表现最差的原因在于没有考虑到数据样本不平衡和特征冗余问题。文献[15]假负率表现较好是由于该方法通过遗传算法计算出每个分类的最佳均衡分布。这进一步证明本文提出的入侵检测算法具有较高的性能和有效性。

3 结 论

针对网络异常流量检测问题，提出了一种基于 ResNet-BiGRU 网络的入侵检测算法。设计了一种新颖的 Transformer 混合融合块进行时间尺度和空间尺度的特征融合；本文相比于传统的机器学习方法效果有明显的提升，此外通过对比现有的一些混合模型检测方法，本文方法的综合性能有明显提高，在准确率、召回率和 F1 分数上具有优越性。下一步工作将在更多数据集中检验模型的适用性，并对网络模型结构进一步改进，提高模型性能和缩短计算时间。

参考文献

[1] 2021 年上半年我国互联网网络安全检测数据分析报告[EB/OL]. https://www.cert.org.cn/publish/main/46/2021/20210731090556980286517/20210731090556980286517_.html,2021.

[2] NSKH P, VARMA M N, NAIK R R. Principle component analysis based intrusion detection system using support vector machine[C]//2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, 2016: 1344－1350.

[3] ELBASIONY R M, SALLAM E A, ELTOBELY T E, et al. A hybrid network intrusion detection framework based on random forests and weighted k-means[J]. Ain Shams Engineering Journal, 2013, 4(4): 753－762.

[4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.

[5] WANG H, LI W. DDosTC: a transformer-based network attack detection hybrid mechanism in SDN[J]. Sensors, 2021, 21(15): 5047.

[6] SINHA J, MANOLLAS M. Efficient deep CNN-BiLSTM model for network intrusion detection[C]//Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition. 2020: 223－231.

[7] 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.

[8] IOFFE S, SZEGEDY C. Batch normalization: Accele-rating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning, 2015: 448－456.

[9] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning (ICML－10). 2010: 807－814.

[10] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[11] 王馨彤，王璇，孙知信. 基于多尺度记忆残差网络的网络流量异常检测模型[J]. 计算机科学, 2022, 49(8): 314－322.

[12] 马明艳，陈伟，吴礼发. 基于 CNN_BiLSTM 网络的入侵检测方法[J]. 计算机工程与应用，2022，58(10)：116－124.

[13] TIAN Q, LI J, LIU H. A method for guaranteeing wireless communication based on a combination of deep and shallow learning[J]. IEEE Access, 2019, 7: 38688－38695.

[14] LI Y, XU Y, LIU Z, et al. Robust detection for network intrusion of industrial IoT based on multi-CNN fusion[J]. Measurement, 2020, 154: 107450.

[15] 刘新倩，单纯，任家东，等. 基于流量异常分析多维优化的入侵检测方法[J]. 信息安全学报，2019，4(1)：14－26.

[16] 李海涛，王瑞敏，董卫宇，等. 一种基于 GRU 的半监督网络流量异常检测方法[J]. 计算机科学, 2023, 50(3): 380－390.