计算机科学
COMPUTER SCIENCE

基于多尺度记忆残差网络的网络流量异常检测模型

王馨彤，王璇，孙知信

引用本文

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.

WANG Xin-tong, WANG Xuan, SUN Zhi-xin. Network Traffic Anomaly Detection Method Based on Multi-scale Memory Residual Network[J]. Computer Science, 2022, 49(8): 314-322.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于 IPSO-BiLSTM 的网络安全态势预测

Network Security Situation Prediction Based on IPSO-BiLSTM

计算机科学, 2022, 49(7): 357-362. https://doi.org/10.11896/jsjkx.210900103

基于 Transformer 和 LSTM 的药物相互作用预测

Drug-Drug Interaction Prediction Based on Transformer and LSTM

计算机科学, 2022, 49(6A): 17-21. https://doi.org/10.11896/jsjkx.210400150

改进注意力机制的多叉树网络多作物早期病害识别方法

Multi-tree Network Multi-crop Early Disease Recognition Method Based on Improved Attention Mechanism

计算机科学, 2022, 49(6A): 363-369. https://doi.org/10.11896/jsjkx.210500044

基于 Stacking 多模型融合的 IGBT 器件寿命的机器学习预测算法研究

Study on Machine Learning Algorithms for Life Prediction of IGBT Devices Based on Stacking Multi-model Fusion

计算机科学, 2022, 49(6A): 784-789. https://doi.org/10.11896/jsjkx.210400030

基于共同子空间分类学习的跨媒体检索研究

Study on Cross-media Information Retrieval Based on Common Subspace Classification Learning

计算机科学, 2022, 49(5): 33-42. https://doi.org/10.11896/jsjkx.210200157

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220200011

基于多尺度记忆残差网络的网络流量异常检测模型

王馨彤　王　璇　孙知信
南京邮电大学江苏省邮政大数据技术与应用工程研究中心　南京 210023
南京邮电大学国家邮政局邮政行业技术研发中心(物联网技术)　南京 210023
南京邮电大学宽带无线通信与传感网技术教育部重点实验室　南京 210023
(1220045303@njupt.edu.cn)

摘　要　基于深度学习的网络流量异常检测模型通常存在现实环境适应性差、表征能力有限以及泛化能力弱的问题。为此,提出了一种基于多尺度记忆残差网络的网络流量异常检测模型。基于高维特征空间分布分析,证明网络流量数据预处理方法的有效性;将多尺度一维卷积与长短期记忆网络相结合,通过深度学习算法提高模型的表征能力;基于残差网络的思想,实现深度特征提取,同时防止梯度消失、梯度爆炸、过拟合及网络退化现象,加快模型收敛速度,从而实现准确高效的网络流量异常检测。数据预处理可视化结果表明,经独热编码处理后,相较于标准化处理,归一化处理可使正常流量与异常流量数据有效分离;有效性验证实验及性能评估实验结果表明,通过增加恒等映射可加快模型收敛速度,并有效解决网络退化问题;对比实验结果表明,多尺度一维卷积及长短期记忆网络可提升模型的表征能力并使模型具备较强的泛化能力,且本文模型相比当前部分深度学习模型呈现更优的性能指标。

关键词:网络流量异常检测;多尺度记忆残差网络;多尺度一维卷积;长短期记忆网络;残差网络;网络入侵检测
中图法分类号　TP393.0

Network Traffic Anomaly Detection Method Based on Multi-scale Memory Residual Network

WANG Xin-tong, WANG Xuan and SUN Zhi-xin
Post Big Data Technology and Application Engineering Research Center of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
Post Industry Technology Research and Development Center of the State Posts Bureau(Internet of Things Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China
Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract　Network traffic anomaly detection based on deep learning usually has the problems of poor adaptability to real-world environments, limited representation ability and week generalization ability. From the perspective of these problems, a network traffic anomaly detection method based on multi-scale memory residual network is proposed. Based on the analysis of high-dimensional feature space distribution, this paper demon-strates the validity of the approach to network traffic data preprocessing. Combining multi-scale one-dimensional convolution and long short-term memory network, the representation ability is enhanced by deep learning classifiers. To make the network traffic anomaly detection accurate and efficient, by the idea of residual network, the deep feature extraction is implemented, the problems of vanishing/exploding gradients, the over-fitting and network degradation are prevented, and the convergence speed of the model is accelerated. The visualizations of data preprocessing result suggest that, compared with standardization, normalization has better capability to separate the abnormal traffic data from the normal traffic data. The result of validity verification and performance evaluation experiment reveal that, by inserting identity mapping, the convergence speed of the model can be accelerated, and the network degradation problem can be efficiently addressed. The result of contrast experiment indicates the one-dimensional convolution and long short-term memory network can reinforce the representation and generalization ability of our model, and the performance metrics of our model is better than that of the current deep learning model.

到稿日期:2022-02-07　返修日期:2022-03-18
基金项目:国家自然科学基金(61972208)
This work was supported by the National Natural Science Foundation of China(61972208).
通信作者:孙知信(sunzx@njupt.edu.cn)

Keywords Network traffic anomaly detection, Multi-scale memory residual network, Multi-scale one-dimensional convolution, Long short-term memory network, Residual network, Network intrusion detection

1 引言

随着网络技术的发展与网络规模的日益扩大，网络流量呈指数式增长，网络安全威胁与风险问题愈发突出。入侵检测系统(Intrusion Detection System，IDS)[1]是一种对网络安全进行即时监控并对网络攻击做出主动安全响应的网络安全技术。网络流量是主要的网络状态之一，当发生网络入侵行为时，通常会出现网络流量异常现象，因此，网络流量异常检测是当前网络入侵检测系统(Network Intrusion Detection System，NIDS)的研究重点。

然而，网络攻击模式的不断变化增加了网络流量异常检测的难度[2]。基于人工智能的赋能效应，网络空间安全面临新的风险，其中包括网络攻击越来越智能化，大规模攻击越来越频繁，网络攻击的隐蔽性越来越高，网络攻击的对抗博弈性越来越强，重要数据越来越容易被窃取等[3]。维护网络安全是一个攻防博弈的过程，网络流量异常检测作为保障网络安全的先决条件，因可识别未知网络攻击而受到越来越多的关注，故如何构建智能高效的网络异常流量检测模型成为关键。

基于机器学习的网络流量异常检测是当前主流的网络入侵检测手段，依据检测技术可进一步将其划分为基于传统机器学习方法的网络流量异常检测与基于深度学习的网络流量异常检测。传统机器学习方法通常强调特征工程，且在特征选择问题上存在困难，无法有效解决高维海量网络流量的异常检测问题，从而导致模型准确率低，误报率高[4]，表征能力有限。随着网络中海量数据的增加，网络带宽的提升，数据的复杂性和特征的多样性也不断提升，传统机器学习难以达到分析和预测的目的[5]。而深度学习方法能够有效处理大规模网络流量数据，相较于传统机器学习方法，深度学习具备更强的表征性能，可有效提升网络流量异常检测的效率及准确率，故基于深度学习的网络流量异常检测是当前网络攻击的有效防护手段。

本文围绕网络流量异常检测模型展开研究，并提出了一种基于多尺度记忆残差网络的网络流量异常检测模型。本文第2节介绍了相关工作，着重从特征工程、深度学习模型以及模型优化3个方面进行阐述；第3节对本文提出的模型所涉及的相关概念进行介绍；第4节介绍了本文提出的网络流量异常检测模型的总体模型架构，并对多尺度记忆残差模块进行了详细描述；第5节介绍了本文实验过程，并对实验结果进行了评估分析与对比；最后总结全文。

2 相关工作

网络流量异常检测是网络安全领域的经典问题。本节着重从特征工程、深度学习模型以及模型优化3个方面对相关工作进行分析。

特征工程通过转换特征空间来提升数据集建模性能[6]。数据和特征决定了机器学习的上限，而模型和算法只是逼近该上限。在数据集选择方面，相较于目前使用广泛的KDD1999，DARPA1998，DARPA1999，NSL-KDD等数据集，近年来陆续发布的UNSW-NB15，CICIDS2017，CICIDS-001等网络流量数据集涵盖了后门、蠕虫等新型攻击，可反映当前新型的网络入侵行为模式，更具说服力[5]。而在数据预处理方面，网络流量数据中存在非数值型特征，且数据中不同特征属性间存在量纲差异，因此需对其进行数值化处理及无量纲化处理。Lu等[7]提出了一种基于信息增益的网络流量特征选择方法，该方法采用信息增益作为特征重要性度量指标，以直观反映特征是否会对分类效果产生影响；Xiao等[8]采用主成分分析(Principal Component Analysis，PCA)和自编码器(Auto-Encoder，AE)等不同方法去除网络流量数据中的冗余及无关特征。上述工作通过特征选择及数据降维方法对网络流量数据进行处理，旨在降低特征维度，从而提升计算效率，而能否适用于当前网络环境并优化特征空间分布仍有待考证。

深度学习模型因具备深层神经网络结构，能够自主学习重要特征并生成输出，因此比传统机器学习方法更有效[9]。如何构建网络流量异常检测深度学习模型是当前的研究重点。目前，基于多分类器构建的深度学习模型因能够有效提高模型表征能力而被广泛运用，然而却时常存在泛化能力弱的问题。Ma等[10]利用三层堆叠长短期记忆网络(Long Short-Term Memory Network，LSTM)[11]来提取不同深度的网络流量特征，解决了单层LSTM适应性弱的问题，并利用改进残差网络对LSTM模型进行优化；Wu等[12]表明，相比较浅的神经网络，较深的神经网络在数据学习及泛化方面具备更强的性能，并将卷积神经网络(Convolutional Neural Networks，CNN)与门控循环单元(Gated Recurrent Unit，GRU)[13]融入到残差网络中，在有效捕捉网络流量时间特征及空间特征的同时提高学习效率。上述深度学习模型基于残差网络思想[14]来改善模型结构，但在模型优化上还存在问题。

模型结构决定其性能表现，如何实现模型优化是构建智能高效的网络异常流量检测模型的关键。关于模型优化问题，Li等[15]表明Dropout[16]方法在网络状态由训练转向测试时会产生方差偏移，而批归一化(Batch Normalization，BN)[17]则能保持其方差稳定，使用时需通过避免方差偏移风险以克服其组合的局限性；Cooijmans等[18]证明了BN在隐藏层间转化的有效性，从而能够减少时间步之间的内部协变量偏移，并将BN引入LSTM以加快训练收敛速度且能够提升其泛化能力。上述工作在Dropout与BN有效防止梯度消失、梯度爆炸及过拟合等问题的基础上，从数学角度分析其有效性，并提出改进方案。网络退化问题成为当前模型亟需解决的问题。

综上所述，当前网络流量异常检测模型往往存在现实环境适应性差、表征能力有限以及泛化能力弱的问题。在特征工程方面，亟需针对可反映当前网络环境的网络流量数据提出有效的数据预处理方法并证明其有效性；在深度学习模型

方面，如何通过多个分类器在实现深度表征的同时提升模型泛化能力成为关键；在模型优化方面，已存在 Dropout 与 BN 等较为成熟的方案，如何将其进行有机结合并解决网络退化问题成为关键。

为应对上述挑战，本文提出了一种基于多尺度记忆残差网络（Multi-Scale Memory Residual Network，MSMRNet）的网络流量异常检测模型，主要工作概况如下：

（1）基于高维特征空间分布分析，证明本文网络流量数据预处理方法的有效性。

（2）将多尺度一维卷积与长短期记忆网络相结合，通过深度学习算法提高模型表征能力。

（3）基于残差网络的思想，实现深度特征提取，同时防止梯度消失、梯度爆炸、过拟合及网络退化现象，加快模型收敛速度，实现网络流量异常检测。

3 基础理论

本文所提模型主要基于一维卷积、长短期记忆网络及残差网络，下面就相关概念和基础知识予以介绍。

3.1 一维卷积

一维卷积特指一维输入与卷积核间的内积运算，定义如下：

$$x_{out}=\sigma(\boldsymbol{W}\cdot x_{in}+b) \tag{1}$$

其中，x_{in} 表示卷积层的输入，x_{out} 表示卷积层的输出，$\boldsymbol{W}\in P^k$ 表示卷积层的卷积核权值参数，$b\in P^d$ 表示卷积层的偏置，σ 表示激活函数，k 表示卷积核长度，d 表示特征维度。以卷积核长度为 3 且输入与输出维度一致的一维卷积运算为例，运算过程如图 1 所示。

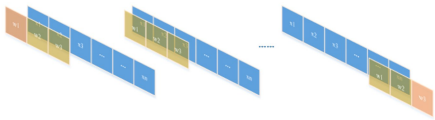


图 1 一维卷积运算过程

Fig. 1 Computational process of one-dimensional convolution

一维卷积层通过共享权值的局部连接，可对网络流量数据进行固定尺度的局部空间特征提取。

3.2 长短期记忆网络

LSTM 网络[11]作为循环神经网络（Recurrent Neural Network，RNN）的变体，在 RNN 基础上通过引入门控机制解决了长序列训练过程中的梯度消失与梯度爆炸问题。LSTM 网络结构如图 2 所示，各 LSTM 单元分别由遗忘门 f_t、输入门 i_t、候选细胞状态 $\tilde{c}_t$ 以及输出门 o_t 构成，计算式如下：

$$f_t=\text{sigmoid}(\boldsymbol{W}_{if}x_t+b_{if}+\boldsymbol{W}_{hf}h_{t-1}+b_{hf}) \tag{2}$$

$$i_t=\text{sigmoid}(\boldsymbol{W}_{ii}x_t+b_{ii}+\boldsymbol{W}_{hi}h_{t-1}+b_{hi}) \tag{3}$$

$$\tilde{c}_t=\tanh(\boldsymbol{W}_{i\tilde{c}}x_t+b_{i\tilde{c}}+\boldsymbol{W}_{h\tilde{c}}h_{t-1}+b_{h\tilde{c}}) \tag{4}$$

$$o_t=\text{sigmoid}(\boldsymbol{W}_{io}x_t+b_{io}+\boldsymbol{W}_{ho}h_{t-1}+b_{ho}) \tag{5}$$

$$c_t=f_t\times c_{t-1}+i_t\times\tilde{c}_t \tag{6}$$

$$h_t=o_t\times\tanh(c_t) \tag{7}$$

其中，x_t 表示 t 时刻的输入；c_t 表示 t 时刻的细胞状态；h_t 表示 t 时刻的隐藏状态，初始时刻的隐藏状态为 0；$\boldsymbol{W}$ 与 b 分别表示各结构间的权重与偏置。以网络流量数据为例，LSTM 单元对其的处理主要分为 3 个阶段。

（1）遗忘阶段。通过遗忘门 f_t 实现上一时刻的细胞状态 c_{t-1} 中网络流量数据的选择性遗忘。

（2）选择记忆阶段。通过输入门 i_t 决定当前时刻网络流量数据输入 x_t 中需存储到当前时刻细胞状态 c_t 的信息。通过候选细胞状态 $\tilde{c}_t$ 对当前输入门 i_t 中的信息进行选择性记忆。

（3）输出阶段。使用 tanh 激活函数对当前时刻细胞状态 c_t 进行缩放，通过输出门 o_t 控制输出到隐藏状态 h_t 的信息。分类结果 y_t 由隐藏状态 h_t 函数变换而来。

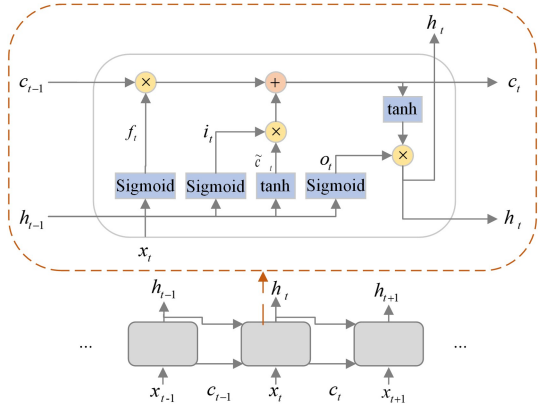


图 2 LSTM 网络结构图

Fig. 2 Structure of LSTM

3.3 残差网络

残差网络（Residual Network，ResNet）[14]由残差块纵向堆叠构成，在普通卷积神经网络的基础上通过将恒等映射[19]作为短路连接（Shortcut Connection）解决了网络退化的问题。与门控机制相比，恒等映射非数据驱动，且无须权重参数控制。残差模块结构如图 3 所示。

残差模块可表示为：

$$y_l=x_l+\mathcal{F}(x_l,\boldsymbol{W}_l) \tag{8}$$

$$x_{l+1}=\text{ReLU}(y_l) \tag{9}$$

其中，x_l 表示第 l 个残差模块的输入；y_l 表示第 l 个残差模块的输出，由恒等映射 x_l 与残差 $\mathcal{F}(x_l,\boldsymbol{W}_l)$ 构成；$\mathcal{F}$ 表示残差函数。

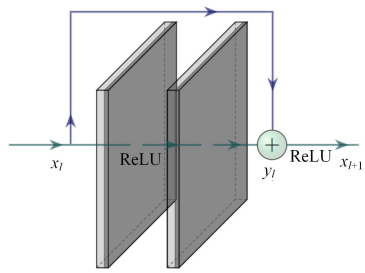


图 3 残差模块结构

Fig. 3 Structure of residual block

深层网络可融合低、中、高层网络的特征，从而使特征层次更加丰富[20]，故网络深度对提高网络流量异常检测模型的准确率具有至关重要的作用。恒等映射的引入使得深层网络的性能较浅层网络明显提升，呈高准确性且易于收敛，同时有效避免了梯度消失与梯度爆炸问题。

4　基于 MSMRNet 的网络流量异常检测模型

4.1　模型架构

结合残差网络及长短期记忆网络思想，本文提出了一种基于 MSMRNet 的网络流量异常检测模型，该模型由数据预处理以及 MSMRNet 模型两部分构成。符号定义如表 1 所列。模型架构如图 4 所示。

表 1　符号定义

Table 1　Definition of notations

符号	描述
X_0	初始网络流量数据
X	经数据预处理后的网络流量数据
X_i	经 $i-1$ 个多尺度记忆残差模块处理后的网络流量数据，$i=\{1,2,\cdots,num\}$，$num-1$ 表示多尺度记忆残差模块个数
Y_0	经多尺度记忆残差网络处理后的网络流量数据
Z_0	经展平层处理后的网络流量数据
Z	经全连接层处理后的网络流量数据
x_{out1}、x_{out2}、x_{out3}	单模块中的多尺度局部空间特征
x_{add}	单模块中的多尺度局部空间融合特征
x_{fus}	单模块中经 BN 与 ReLU 函数处理后的多尺度局部空间特征
x_{fus_lstm}	单模块中对 x_{fus} 进行 LSTM 操作后的全局映射特征

数据预处理：网络流量数据中存在协议类型、服务类型等非数值型数据，而机器学习模型无法处理非数值数据，因此需对网络流量数据进行数值化处理。同时，网络流量数据不同，特征属性间数量级相差较大[5]，因此需对网络流量数据进行无量纲化处理。初始网络流量数据 $X_0=\{x_1,x_2,\cdots,x_f\}$，经数据预处理后获得 $X=\{x_1,x_2,\cdots,x_n\}$，其中，$f=|X_0|$ 表示初始流量数据特征维度，$n=|X|$ 表示经数据预处理后的网络流量数据特征维度。

MSMRNet 模型：将经过数据预处理的网络流量数据 X 输入至 MSMRNet 进行深度特征提取，即 MSMRNet 初始输入 $X_1=X$。MSMRNet 由若干多尺度记忆残差模块堆叠而成，其第 l 个多尺度记忆残差模块以 X_l 作为输入，并生成输出 X_{l+1}，其中，输入 X_l 与输出 X_{l+1} 维度一致。利用 Flatten 层将多维输出 Y_0 一维化，以获得输出 Z_0。经过全连接层对局部特征进行综合处理[21]获得输出 Z。采用 softmax 函数作为分类器实现网络流量分类，获得网络流量分类结果 $\hat{Y}$。$\hat{Y}$ 中各元素分别表示各网络流量类别概率，其最大概率类别即为分类结果，计算式如式(10)所示，$\boldsymbol{W}_d$ 与 b_d 分别表示权重矩阵与偏置项。

$$\hat{Y}=\text{softmax}(Z)=\text{softmax}(\boldsymbol{W}_d^{\mathrm{T}}Z_0+b_d) \tag{10}$$

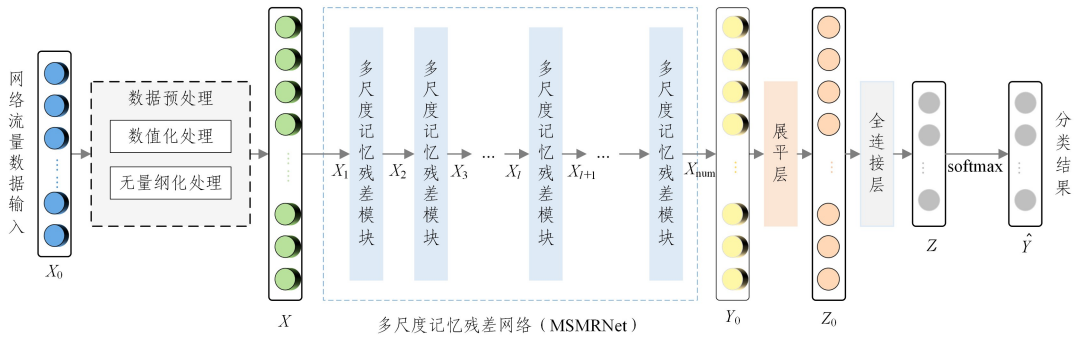


图 4　基于 MSMRNet 的网络流量异常检测模型架构

Fig. 4　Architecture of network traffic anomaly detection method based on MSMRNet

基于 MSMRNet 的网络流量异常检测模型实现方法如算法 1 所示。

算法 1　基于 MSMRNet 的网络流量异常检测模型实现方法

输入：训练集 X_{train}，测试集 X_{test}，标签集 Y

输出：网络流量分类结果 $\hat{Y}_D$

步骤 1　数据预处理

1. 对训练集 X_{train} 及测试集 X_{test} 进行数值化处理，获得 $X_{\_train\_num}$ 及 $X_{\_test\_num}$
2. 对训练集 $X_{\_train\_num}$ 及测试集 $X_{\_test\_num}$ 进行无量纲化处理，获得 $X_{\_train}$ 及 $X_{\_test}$

步骤 2　构建模型

3. 添加若干多尺度记忆残差模块
4. 添加 Flatten 层及全连接层，采用 softmax 函数作为分类器

步骤 3　训练模型

5. 设置实验超参数：优化器 optimizer、单次训练样本数 batch_size、学习率 learning_rate、训练轮数 epoch。设置实验验证集
6. while 未达到预设训练轮数 epochs do
7. 　while 训练集不为空 do
8. 　　以小批次数据集 batch 作为模型输入
9. 　　计算交叉熵损失函数，C 表示网络流量类别数
10. 　　使用 Adam 优化器更新模型参数
11. 　end while
12. 　使用验证集验证模型并进行参数微调
13. end while

步骤 4　保存模型

14. 保存参数微调后的模型

步骤 5　测试模型

15. 载入已保存模型，使用测试集测试该模型
16. return 测试集中网络流量数据分类结果

4.2　多尺度记忆残差模块

多尺度记忆残差模块为 MSMRNet 的核心模块。MSMRNet 通过多尺度记忆残差模块的多尺度一维卷积层来增加网络宽度，并将其与 LSTM 相结合，提高了模型的表征能力及泛化能力；通过堆叠多个多尺度记忆残差模块来增加网络深度，每个多尺度记忆残差模块在多尺度记忆模块的基础上增加恒等映射，以防止深层网络梯度消失、梯度爆炸、过拟合及网络退化现象的出现，加快模型收敛速度，实现模型深度表征。

多尺度记忆模块如图 5(a)所示，多尺度记忆残差模块在其基础上增加恒等映射，如图 5(b)所示。本节从多尺度一维卷积层、LSTM 层及残差学习 3 个方面对多尺度忆残差模块进行介绍。

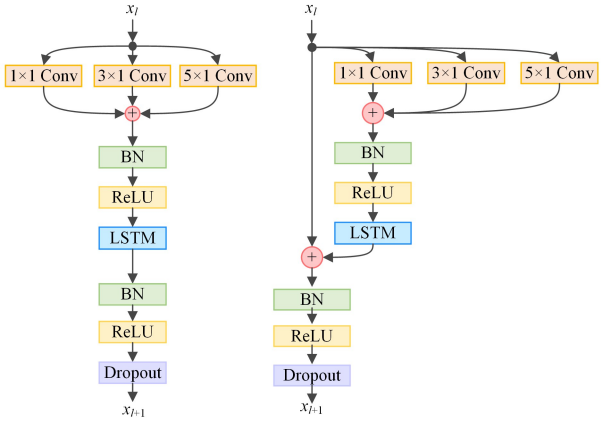


图 5 残差模块结构

Fig. 5 Structure of residual block

4.2.1 多尺度一维卷积层

网络流量异常检测不能仅依赖于离散的局部特征，而应通过不同尺度的卷积核提取不同尺度的网络流量特征，结合多尺度局部特征进行异常检测[22-23]。基于 Inception 结构[24-25]的多尺度特征融合的思想，本文通过多尺度一维卷积层实现多尺度网络流量数据局部空间特征提取，采用的卷积核长度(即特征提取尺度)分别为 1,3,5，计算式分别如下：

$$x_{out1}=\sigma(\boldsymbol{W}_1 \cdot x+b_1) \tag{11}$$

$$x_{out2}=\sigma(\boldsymbol{W}_2 \cdot x+b_2) \tag{12}$$

$$x_{out3}=\sigma(\boldsymbol{W}_3 \cdot x+b_3) \tag{13}$$

其中，$\boldsymbol{W}_1 \in P^{k_1}$，$\boldsymbol{W}_2 \in P^{k_2}$，$\boldsymbol{W}_3 \in P^{k_3}$，$k_1=1$，$k_2=3$，$k_3=5$，激活函数 σ 采用 ReLU 函数。通过逐项相加的方式，即 Add 函数实现特征融合，计算式如下：

$$x_{add}=x_{out1}+x_{out2}+x_{out3} \tag{14}$$

若采用 Concatenate 函数通过合并的方式进行特征融合，则单个元素的信息量不变，特征维度增加；若采用 Add 函数通过逐项相加的方式进行特征融合，则单个元素的信息量增加，特征维度不变。由于 x_{out1}，x_{out2} 与 x_{out3} 维度一致，为减少后续运算参数量，本文采用 Add 函数代替 Concatenate 函数进行特征融合。

利用 BN 层对 x_{add} 进行规范化处理获得 x_{add_BN}，使得输出满足或近似服从正态分布，即 $x_{add\_BN} \sim (0,1)$，从而加快模型收敛速度，防止出现梯度消失的现象，并将 ReLU 作为激活函数以放大特征间的差异，获得最终多尺度一维融合特征 x_{fus}。计算式如下：

$$x_{fus}=\mathrm{ReLU}(x_{add\_BN}) \tag{15}$$

4.2.2 LSTM 层

为防止长序列训练过程中出现梯度消失及梯度爆炸问题，本文通过 LSTM 提取输入数据的全局时间特征。将经多尺度一维卷积处理后的网络流量数据 x_{fus} 输入至 LSTM 层，计算式如下：

$$x_{fus\_lstm}=LSTM(x_{fus}) \tag{16}$$

其中，$x_{fus\_lstm}=\{h_1,h_2,\cdots,h_n\}$ 为 LSTM 层的输出，h_i 表示第 i 个时间步的隐藏状态，$i \in \{1,2,\cdots,n\}$，LSTM 函数公式如式(2)—式(7)所示。LSTM 层输出维度由其隐藏单元个数决定。由于后续残差学习需进行逐项相加操作，因此将隐藏单元个数设为 n，n 表示多尺度记忆残差模块输入数据的特征维度，与经数据预处理后的网络流量数据特征维度一致。

通过 LSTM 层对多尺度局部空间特征进行全局映射，使得网络对多尺度网络流量特征具备更强的全局记忆能力，从而获得更好的特征分析效果，提升网络的表征能力。

4.2.3 残差学习

为解决网络退化问题，本文通过增加恒等映射来实现残差拟合。定义多尺度记忆残差映射如下：

$$y_{res}=x+\mathcal{F}(x)=x+x_{fus\_lstm} \tag{17}$$

其中，x 与 y_{res} 分别表示多尺度记忆残差学习的输入与输出，$\mathcal{F}$ 表示残差映射。传统神经网络通过非线性映射难以拟合恒等映射 x，而本文通过残差映射 $\mathcal{F}(x)$，使得原本输入 x 达到最优时只需拟合 $\mathcal{F}(x)=0$。除此之外，通过 BN 层及 ReLU 函数优化特征分布。为避免出现方差偏移[15]，将 Dropout 层置于 BN 层后使用，通过 Dropout 层[16]防止出现过拟合，得到多尺度记忆残差模块的最终输出 y。

恒等映射的引入未增加模型参数及计算复杂度，且随着多尺度记忆残差模块堆叠个数的增加，MSMRNet 加深，模型变得更复杂，通过逐层残差学习，深层网络训练结果较浅层结果具备更好的效果，网络退化问题得以解决。

5 实验

5.1 数据集介绍

本文采用 UNSW-NB15 数据集[26]中的部分数据作为仿真实验数据，以大约 6∶4 的比例划分训练集与测试集[27]，数据集分布如表 2 所列。UNSW-NB15 数据集共包含 9 种网络攻击类型，且涵盖后门、蠕虫等新型攻击类别，解决了 KDD99 数据集与 NSL-KDD 数据集中现实环境适应性弱以及训练集与测试集分布不同的问题，且不包含冗余数据。

表 2 实验数据集分布

Table 2 Distribution of experimental data set

类别	训练集	测试集
Normal	56000	37000
Analysis	2000	677
Backdoor	1746	583
DoS	12264	4089
Exploits	33393	11132
Fuzzers	18184	6062
Generic	40000	18871
Reconnaissance	10491	3496
Shellcode	1133	378
Worms	130	44
合计	175341	82332

实验数据集包含基本特征、流量特征、内容特征、时间特征及附加特征 5 种特征属性类别，共 42 维特征，其中存在非数值数据，且各特征属性性质不同，缺乏综合性，因此，

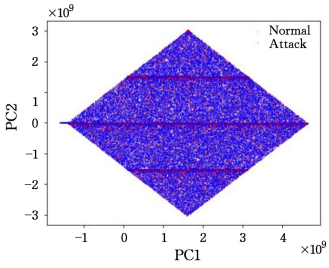
需对其进行数据预处理。

5.2 数据预处理

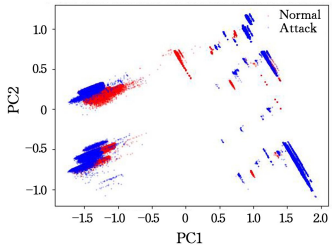
数值化处理：针对实验数据集中类别型特征进行数值化处理。由于特征类别间不存在相对关系，因此本实验采用独热编码对其进行数值化处理。

无量纲化处理：针对经过独热编码处理后的实验数据集中的数值型特征进行无量纲化处理。本实验分别采用标准化与归一化对数据集进行处理。

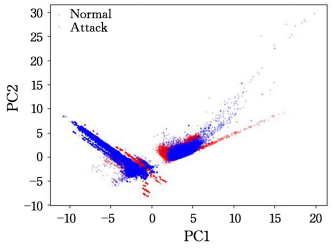
为获得有效的数据预处理方法，本实验通过 PCA 二维可视化方法对数据预处理后的实验数据集特征空间分布进行分析。经过独热编码处理后的实验数据集 PCA 二维可视化结果如图 6(a)所示，可以看出，正常流量数据点与异常流量数据点大面积重叠。经过独热编码与标准化处理后的实验数据集 PCA 二维可视化结果如图 6(b)所示，可以看出，正常流量数据点与异常流量数据点已部分分离，但仍有部分区域重叠。经过独热编码与归一化处理后的实验数据集 PCA 二维可视化结果如图 6(c)所示，可以看出，相比经过独热编码与标准化处理后的结果，正常流量数据点与异常流量数据点已有效分离。因此，本实验采用独热编码与归一化对实验数据集进行数据预处理。



(a)独热编码处理后的实验数据集



(b)独热编码与标准化处理后的实验数据集



(c)独热编码与归一化处理后的实验数据集

图 6 数据预处理后的实验数据集 PCA 二维可视化结果

Fig. 6 Two-dimensional visualizations of data set after data preprocessing using PCA

5.3 实验设置

本文实验基于 Tensorflow 与 Keras 深度学习框架实现。经数据预处理后，网络流量数据特征维度为 194，因此多尺度记忆残差模块参数设置如表 3 所列。

表 3 多尺度记忆残差模块参数设置

Table 3 Parameter of multi-scale memory residual network

层操作名称	输出维度	参数设置	上层操作名称
input_1(InputLayer)	(None,1,194)	—	
conv1d_1(Conv1D)	(None,1,194)	194,1,same	input_1
conv1d_2(Conv1D)	(None,1,194)	194,3,same	input_1
conv1d_3(Conv1D)	(None,1,194)	194,5,same	input_1
add_1(Add)	(None,1,194)	—	conv1d_1 conv1d_2 conv1d_3
batch_normalization_1(Batch Normalization)	(None,1,194)	—	add_1
activation_1(Activation)	(None,1,194)	relu	batch_normalization_1
lstm(LSTM)	(None,1,194)	194,return_sequences=True	activation_1
add_2(Add)	(None,1,194)	—	lstm_1
batch_normalization_2(Batch Normalization)	(None,1,194)	—	add_2
activation_2(Activation)	(None,1,194)	relu	batch_normalization_2
dropout_1(Dropout)	(None,1,194)	0.5	activation_2

本文实验超参数设置如表 4 所列。采用 sparse_categorical_crossentropy 作为损失函数。

表 4 超参数设置

Table 4 Hyper-parameter

参数名称	参数意义	参数值
optimizer	优化器	Adam
learning_rate	学习率	0.001
batch_size	批训练样本数	1 000
epochs	训练轮数	50

5.4 评价指标

本文采用准确率(Accuracy，ACC)、精确率(Precision，P)、召回率(True Positive Rate，TPR)、误报率(False Positive Rate，FPR)以及 F1 分数(F1-Score，F1)作为评估指标，计算式如下：

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

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

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

$$FPR=\frac{FP}{TN+FP} \tag{21}$$

$$F1=\frac{2\times P\times TPR}{P+TPR} \tag{22}$$

其中，ACC 表示预测正确的样本占总样本的比率，其值越高，则网络流量异常检测性能越好。P 表示被预测为异常的样本中预测正确的样本比率，其值越高，则网络流量异常检测效果越好。TPR 表示实际为异常的样本中被预测为异常的样本比率，其值越高，则网络流量异常检测的异常流量漏报率

越低。*FPR* 表示实际为正常的样本中被预测为异常的样本比率，其值越低，则网络流量异常检测对正常流量的判别效果更好。*F*1 表示准确率和召回率的调和平均值，为两者综合考虑结果，其值越高，则网络流量异常检测综合判别性能越好。

5.5 有效性验证实验

为验证 MSMRNet 解决网络退化问题的有效性，本文构建了不同深度的多尺度记忆网络(Multi-Scale Memory Network，MSMNet)与 MSMRNet 进行对比，实验模型如下。

MSMNet-5：由 5 个多尺度记忆模块堆叠而成，共包含 20 个训练参数层、10 个非训练参数层与 1 个全连接层。

MSMRNet-5：由 5 个多尺度记忆残差模块堆叠而成，共包含 20 个训练参数层，10 个非训练参数层与 1 个全连接层。

MSMNet-10：由 10 个多尺度记忆模块堆叠而成，共包含 40 个训练参数层、20 个非训练参数层与 1 个全连接层。

MSMRNet-10：由 10 个多尺度记忆残差模块堆叠而成，共包含 40 个训练参数层、20 个非训练参数层与 1 个全连接层。

图 7 给出了 MSMNet 与 MSMRNet 训练及测试过程中的丢失率对比。

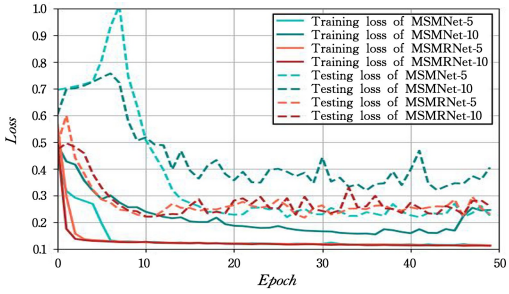


图 7 MSMNet 与 MSMRNet 的丢失率对比

Fig. 7 Comparison of loss rates between MSMNet and MSMRNet

通过纵向比较可知，相比 MSMNet-5，MSMNet-10 在训练及测试过程中的丢失率大幅增加，模型性能明显减弱，随着迭代次数的增加，出现了过拟合现象，并存在网络退化问题；而 MSMRNet-10 相比 MSMRNet-5 在训练过程中的丢失率有所减少，且收敛速度更快，更快达到较优的测试丢失率。通过横向比较可知，MSMRNet-5 相比 MSMNet-5 以及 MSMRNet-10 相比 MSMNet-10 在训练及测试过程中模型收敛速度显著提高，且在更深的模型中，MSMRNet 的性能提升更明显。

由此可知，所提模型通过增加恒等映射，在未增加额外参数的同时，可有效解决网络退化问题，且模型更易于优化收敛，并具备更好的性能。

5.6 性能评估实验

为验证通过增加恒等映射可提升深层网络性能，本文将 MSMNet-5，MSMNet-10，MSMRNet-5 以及 MSMRNet-10 模型的性能评估结果进行对比。

图 8 给出了 MSMNet 与 MSMRNet 模型的 ROC 曲线对比。ROC 曲线可直观反映模型性能，坐标(0，1)表示将所有异常样本预测为异常，且将所有正常样本预测为正常的理想模型。AUC 值表示 ROC 曲线下方面积，可反映模型对样本的分类能力，其值越大，则网络流量异常检测性能更佳。

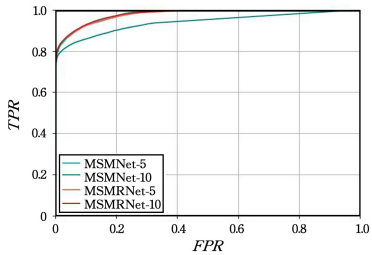


图 8 MSMNet 与 MSMRNet 模型 ROC 曲线对比

Fig. 8 Comparison of ROC curve between MSMNet and MSMRNet

表 5 列出了上述模型性能评估结果的对比。通过纵向对比可知，随着网络深度的增加，MSMNet-10 较 MSMNet-5 模型性能大幅降低，网络退化现象明显；而 MSMRNet-10 较 MSMRNet-5 模型性能有所提升，具备更高的准确率、精确率以及 F1 分数，且误报率更低，ROC 曲线的 AUC 值更优。通过横向对比可知，相同深度的模型中，MSMRNet-5 与 MSMNet-5 性能相近；而 MSMRNet-10 相比 MSMNet-10 性能显著提升。

表 5 MSMNet 与 MSMRNet 性能评估结果的对比

Table 5 Comparison of performance evaluation results between MSMNet and MSMRNet

(单位：%)

方法	*ACC*	*P*	*TPR*	*FPR*	*F*1	*AUC*
MSMNet-1	87.682	82.883	97.832	24.754	89.739	0.979 39
MSMNet-5	89.064	84.964	97.368	21.111	90.744	0.978 87
MSMNet-10	81.881	77.954	93.548	32.414	85.042	0.938 24
MSMRNet-1	87.973	83.032	98.231	26.570	89.994	0.981 60
MSMRNet-5	89.559	86.740	95.661	17.916	90.983	0.978 26
MSMRNet-10	90.725	88.663	95.348	14.938	91.884	0.979 93

由此可知，MSMRNet 在浅层网络中与 MSMNet 性能相近，提升不明显；而在深层网络中，与 MSMNet 相比，其展现出更优的综合性能评估结果，具备更强的网络流量异常检测性能。

5.7 对比实验

本文将 MSMRNet 模型与以下深度学习模型进行对比。

CNN[28]：基于 CNN 进行网络流量异常检测，通过 CNN 有效提取网络流量数据空间特征。

LSTM[11]：基于 LSTM 进行网络流量异常检测，通过 LSTM 有效提取网络流量数据时间特征。

GRU[13]：基于 GRU 进行网络流量异常检测，GRU 在 LSTM 的基础上减少了门控函数个数，通过 GRU 可提升网络流量数据时间特征提取效率。

CNN-LSTM[29]：基于 CNN 与 LSTM 进行网络流量异常

检测，先通过 CNN 提取网络流量数据的空间特征，再通过 LSTM 提取网络流量数据的时间特征。

CNN-GRU[12,28]：基于 CNN 与 GRU 进行网络流量异常检测，先通过 CNN 提取网络流量数据的空间特征，再通过 GRU 提取网络流量数据的时间特征，并在一定程度上可提升时间特征提取效率。

本文的对比实验为基于上述深度学习模块构建网络流量异常检测模型，通过堆叠多个深度学习模块实现深度表征。下文为对比实验结果及分析。表 6 列出了不同深度学习模型网络流量性能评估结果对比。由对比结果可知：CNN，CNN5 及 CNN10 已具备较优的网络流量异常检测性能；LSTM 在本实验数据集上效果略优于 GRU；采用 CNN 进行局部空间特征提取后，LSTM 与 GRU 性能更优；单模块 MSMNet 较 CNN-LSTM 以及 CNN-GRU 具有更高的准确率、精确率、F1 分数及 AUC 值，且误报率更低，总体性能更优；MSMRNet-10 较其余模型准确率高达 90.725%，精确率提升至 88.663%，误报率降至 14.938%，F1 分数增至 91.884%，整体性能更佳。

表 6　不同深度学习模型性能评估结果的对比

Table 6　Comparison of performance evaluation results of different deep learning models

（单位：%）

方法	ACC	P	TPR	FPR	F1	AUC
CNN	86.773	81.319	98.637	27.762	89.145	**0.98204**
CNN-5	87.117	81.936	98.266	26.543	89.361	0.97875
CNN-10	88.533	84.125	97.589	22.562	90.358	0.98018
LSTM	83.750	78.028	98.114	33.849	86.926	0.96934
LSTM-5	83.765	77.652	**99.007**	34.911	87.039	0.97593
LSTM-10	84.423	79.017	97.635	31.465	87.345	0.97129
GRU	83.389	77.578	98.220	34.781	86.687	0.96839
GRU-5	83.922	78.606	97.276	32.438	86.950	0.97011
GRU-10	84.287	79.136	97.048	31.349	87.182	0.97087
CNN-LSTM	86.240	80.712	98.562	28.857	88.749	0.98046
CNN-LSTM-5	87.243	81.985	98.467	26.508	89.474	0.98057
CNN-LSTM-10	84.367	78.676	98.231	32.619	87.373	0.97268
CNN-GRU	86.090	80.778	98.076	28.594	88.590	0.97815
CNN-GRU-5	86.185	80.488	98.129	29.368	88.741	0.98129
CNN-GRU-10	88.401	84.439	96.766	21.849	90.183	0.97742
MSMNet-1	87.682	82.883	97.832	24.754	89.739	0.97939
MSMNet-5	89.064	84.964	97.368	21.111	90.744	0.97887
MSMNet-10	81.881	77.954	93.548	32.414	85.042	0.93824
MSMRNet-1	87.973	83.032	98.231	26.570	89.994	0.98160
MSMRNet-5	89.559	86.740	95.661	17.916	90.983	0.97826
MSMRNet-10	**90.725**	**88.663**	95.348	**14.938**	**91.884**	0.97993

本文针对上述结果做出如下分析：采用 CNN 可实现对网络流量数据局部空间特征的有效提取，且深层 CNN 模型使得所提取的特征层次更丰富；GRU 在结构上对 LSTM 进行简化，提高了计算效率，但在本实验数据集中，采用 LSTM 在一定程度上呈现了更好的性能；经独热编码处理后，网络流量数据较为稀疏，若采用 LSTM 或 GRU 直接对其进行全局时间特征提取，则异常检测性能较弱，而若先采用 CNN 对其进行局部特征提取，可在一定程度上避免稀疏特征造成的影响，提高模型总体表征性能；相比 CNN-LSTM 以及 CNN-GRU 的单尺度一维卷积特征提取，MSMNet 采用多尺度一维卷积对网络流量特征进行处理，具有更丰富的局部空间特征表示能力；MSMRNet-10 通过恒等映射解决了 CNN-LSTM-10 及 MSMRNet-10 的网络退化问题[12]。深层模型具备更强的学习潜力，TPR 虽稍逊于部分模型，但精确率大幅提升，从 F1 分数进行综合评估，本文模型具有更好的网络流量异常检测效果。

结束语　为解决基于深度学习的网络流量异常检测模型存在的环境适应性差、表征能力有限以及泛化能力弱的问题，本文提出一种基于多尺度记忆残差网络的网络流量异常检测模型。本文基于高维特征空间分布分析，证明了网络流量数据预处理方法的有效性；将多尺度一维卷积与长短期记忆网络相结合，通过深度学习算法提高模型表征能力；基于残差网络的思想，实现深度特征提取，同时防止梯度消失、梯度爆炸、过拟合及网络退化问题，加快模型收敛速度，从而实现准确高效的网络流量异常检测。数据预处理可视化结果表明，经独热编码处理后，相较于标准化处理，归一化处理可使正常流量与异常流量数据有效分离；有效性验证实验及性能评估实验结果表明，通过增加恒等映射可加快模型收敛速度，提升网络流量异常检测性能，并有效解决网络退化问题；对比实验结果表明，多尺度一维卷积及长短期记忆网络可提升模型的表征能力并使模型具备较强的泛化能力，且本文模型相较于当前深度学习模型的性能指标更优。

参 考 文 献

[1] ANDERSON J P. Computer security threat monitoring and surveillance[R]. Technical Report, James P. Anderson Company, 1980.

[2] ZHONG Y, CHEN W, WANG Z, et al. HELAD: A novel network anomaly detection model based on heterogeneous ensemble learning[J]. Computer Networks, 2020, 169: 107049.

[3] GUO Y, FANG B X, LI A P, et al. Artificial intelligence enabled cyberspace security defence [J]. Strategic Study of Chinese Academy of Engineering, 2021, 23(3): 98-105.

[4] SU T, SUN H, WANG S. Intrusion detection using convolutional recurrent neural network[C]// Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition. 2019: 413-419.

[5] JIAN S, LU Z, DU D, et al. Overview of network intrusion detection technology[J]. Journal of Cyber Security, 2020, 5(4): 96-122.

[6] NARGESIAN F, SAMULOWITZ H, KHURANA U, et al. Learning feature engineering for classification[C]// International Joint Conference on Artificial Intelligence(IJCAI). 2017: 2529-2535.

[7] LU X, LIU P, LIN J. Network traffic anomaly detection based on information gain and deep learning[C]// Proceedings of the 2019 3rd International Conference on Information System and Data Mining. 2019: 11-15.

[8] XIAO Y, XING C, ZHANG T, et al. An intrusion detection model based on feature reduction and convolutional neural networks[J]. IEEE Access, 2019, 7: 42210-42219.

[9] AHMAD Z, SHAHID K A, WAI SHIANG C, et al. Network intrusion detection system: A systematic study of machine learning and deep learning approaches[J]. Transactions on Emerging Telecommunications Technologies, 2021, 32(1): e4150.

[10] MA W G, ZHANG Y D, GUO J. Abnormal traffic detection method based on LSTM and improved residual neural network optimization[J]. Journal on Communications, 2021, 42(5): 23-40.

[11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[12] WU P, GUO H, MOUSTAFA N. Pelican: A deep residual network for network intrusion detection[C]// 2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops(DSN-W). IEEE, 2020: 55-62.

[13] CHO K, MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Conference on Empirical Methods in Natural Language Processing. 2014: 1724-1734.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[15] LI X, CHEN S, HU X, et al. Understanding the disharmony between dropout and batch normalization by variance shift[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2682-2690.

[16] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[17] IOFFE S, SEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR, 2015: 448-456.

[18] COOIJMANS T, BALLAS N, Laurent C, et al. Recurrent batch normalization[J]. arXiv: 1603.09025, 2016.

[19] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// European Conference on Computer Vision. Cham: Springer, 2016: 630-645.

[20] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// European Conference on Computer Vision. Cham: Springer, 2014: 818-833.

[21] SAINATH T N, KINGSBURY B, Mohamed A, et al. Improvements to deep convolutional neural networks for LVCSR[C]// 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2013: 315-320.

[22] ZHANG J, LING Y, FU X, et al. Model of the intrusion detection system based on the integration of spatial-temporal features [J]. Computers & Security, 2020, 89: 101681.

[23] WANG X, YIN S, LI H, et al. A Network Intrusion Detection Method Based on Deep Multi-scale Convolutional Neural Network[J]. International Journal of Wireless Information Networks, 2020, 27(4): 503-517.

[24] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

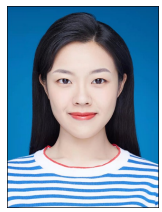
[25] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]// Thirty-first Association for Advancement of Artificial Intelligence(AAAI) Conference on Artificial Intelligence. 2017.

[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). IEEE, 2015: 1-6.

[27] MOUSTAFA N, SLAY J. The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set[J]. Information Security Journal: A Global Perspective, 2016, 25(1/2/3): 18-31.

[28] VINAYAKUMAR R, SOMAN K P, POOMACHANDRAN P. Applying convolutional neural network for network intrusion detection[C]// 2017 International Conference on Advances in Computing, Communications and Informatics(ICACCI). IEEE, 2017: 1222-1228.

[29] WU P, GUO H. LuNET: A deep neural network for network intrusion detection[C]// 2019 IEEE Symposium Series on Computational Intelligence(SSCI). IEEE, 2019: 617-624.



WANG Xin-tong, born in 1998, postgraduate. Her main research interests include cyber security, intrusion detection and machine learning.

SUN Zhi-xin, born in 1964, Ph.D, professor, Ph.D supervisor. His main research interests include network communication and computer network and security.

(责任编辑:何杨)