一种基于自编码器降维的神经卷积网络入侵检测模型

孙敬[1]，丁嘉伟[1]，冯光辉[2]
（1. 郑州工业应用技术学院信息工程学院，河南 郑州 451150；
2. 广州大学计算机科学与网络工程学院，广东 广州 510006）

摘　要：为了提升入侵检测的准确率，鉴于自编码器在学习特征方面的优势以及残差网络在构建深层模型方面的成熟应用，提出一种基于特征降维的改进残差网络入侵检测模型（improved residual network intrusion detection model based on feature dimensionality reduction，IRFD），进而缓解传统机器学习入侵检测模型的低准确率问题。IRFD采用堆叠降噪稀疏自编码器策略对数据进行降维，从而提取有效特征。利用卷积注意力机制对残差网络进行改进，构建能提取关键特征的分类网络，并利用两个典型的入侵检测数据集验证IRFD的检测性能。实验结果表明，IRFD在数据集UNSW-NB15和CICIDS 2017上的准确率均达到99%以上，且F1-score分别为99.5%和99.7%。与基线模型相比，提出的IRFD在准确率、精确率和F1-score性能上均有较大提升。

关键词：网络攻击；入侵检测模型；堆叠降噪稀疏自编码器；卷积注意力机制；残差网络
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025002

A neural convolutional network intrusion detection model based on autoencoder dimension reduction

SUN Jing[1], DING Jiawei[1], FENG Guanghui[2]
1. School of Information Engineering, Zhengzhou University of Industrial Technology, Zhengzhou 451150, China
2. School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

Abstract: In order to improve the accuracy of intrusion detection, considering the advantages of autoencoders in learning features and the mature application of residual networks in constructing deep models, an improved residual network intrusion detection model based on feature dimensionality reduction (IRFD) was proposed. The goal of the proposed IRFD was to solve the issue of low detection accuracy of traditional machine learning based intrusion detection models. In IRFD, the stacking denoising sparse autoencoder was employed to reduce the dimensionality of features and extract effective features. The convolutional attention mechanism was used to improve the residual network and form a classification network that could extract key features. Two typical intrusion detection datasets were em-

收稿日期：2024−08−06；**修回日期：**2024−11−22
基金项目：教育部产学合作协同育人项目（No.220602236285739）；2024年度河南省高等教育教学改革研究与实践项目（本科教育类）（No.2024SJGLX0584）
Foundation Items: The Industry-university Cooperative Education Project of Ministry of Education (No. 220602236285739), 2024 Henan Provincial Higher Education Teaching Reform Research and Practice Project (Undergraduate Education) (No.2024SJGLX0584)

ployed to verify the detection performance of the IRFD. Experimental results demonstrate that the detection accuracy of the proposed IRFD on the both UNSW-NB15 and CICIDS 2017 datasets are over 99%, with F1-score of 99.5% and 99.7%, respectively. Compared with the state-of-the-art models for the intrusion detection, the accuracy, precision, and F1-score performance of the IRFD were significantly improved.

Key words: network attack, intrusion detection model, stacking denoising sparse autoencoder, convolutional attention mechanism, residual network

0 引言

网络入侵检测系统（intrusion detection system，IDS）通过分析监测网络流量来检测网络攻击，进而维护网络安全[1]。目前，机器学习（machine learning，ML）和深度学习（deep learning，DL）已被广泛应用于IDS中[2-3]。基于ML或DL的IDS模型能够从网络流量数据中提取特征并进行分析，进而检测网络流量中是否存在攻击行为。因此，开发高效的IDS模型成为网络安全领域的重要研究内容。

最初，研究人员利用支持向量机、决策树和朴素贝叶斯等典型的机器学习方法来构建入侵检测模型。例如，文献[4]利用随机森林方法构建了一个IDS模型。文献[5]利用朴素贝叶斯方法构建了防御网络攻击的安全路由。文献[6]提出了一种基于增强型支持向量机的入侵检测模型，该模型利用交叉和变异算法智能地选择信息量最大的流量特征，再通过增强支持向量机进行分类，以提升攻击的检测率。文献[7]提出了基于支持向量机的无线传感网络异常入侵检测模型，以实现对网络入侵攻击的检测。文献[8]采用了多层支持向量机来检测网络攻击，从收集到的高维网络数据中提取特征信息，并利用这些特征信息训练模型。

然而，这些传统的机器学习仅能以浅层方式学习数据，难以挖掘数据的深层特征，这也导致基于传统机器学习的入侵检测模型仅适用于小规模、非高维数据，一旦面临大规模的高维数据，这些传统的机器学习方法便难以维持高检测率的性能。

与传统的机器学习方法相比，深度学习方法能更深层地学习并挖掘数据特征，因此，基于深度学习的入侵检测模型在检测攻击的准确率上，优于基于传统机器学习的入侵检测模型。文献[9]提出了一种融合双向门控单元和注意力机制的入侵检测模型，该模型利用卷积神经网络提取数据的非线性特征，并利用注意力机制对不同类型流进行加权处理，进而提高了模型特征提取与分类的性能。文献[10]提出了一种融合卷积神经网络和双向长短时记忆（bidirectional long short-term memory，BiLSTM）网络的入侵检测模型，该模型利用卷积神经网络提取特征，再将这些特征输入BiLSTM网络，进而由BiLSTM检测和识别网络中的恶意行为。

尽管基于深度学习的IDS模型在一定程度上已经取得了较好的检测性能，但仍存在一些亟待解决的问题，且存在较大的研究空间。一方面，随着网络规模的增大，数据量暴增，高维的数据降低了训练模型的效率；另一方面，网络规模的迅速扩张滋生了新的未知攻击，这些未知攻击并不在训练集中，而监督学习模型通常只能检测在训练过程中遇到过的攻击类别，这就要求入侵检测模型具备较高的检测未知攻击的能力。

无监督学习则是一种先利用良性数据进行训练，再观察网络环境中的流量，进而判断流量中是否存在攻击的方法，整个过程不受限于特定的攻击类型。其中，基于自编码器的模型就属于无监督学习方法，它不依赖于用于训练模型的特定样本类型，从而展现出了对未知攻击检测的潜力。为此，本文提出了一种基于特征降维的改进残差网络入侵检测模型（improved residual network in-

trusion detection model based on feature dimensionality reduction，IRFD）。IRFD 先利用堆叠降噪稀疏自编码器提取有效特征，减少冗余特征，进而实现对数据的降维，然后将卷积注意力机制和残差网络进行融合，构建基于卷积注意力机制-残差网络的分类器。本文采用了数据集 UNSW-NB15 和 CICIDS 2017 分析 IRFD 的性能，并进行相应的实验分析和讨论。性能分析结果表明，提出的 IRFD 在这两个数据集上的 F1-score 均达到了 99%。

1　自编码器和稀疏自编码器

1.1　自编码器

考虑包含 n 个数据样本的数据集 $D=\left\{\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots,\boldsymbol{x}^{(n)}\right\}$，且 $|D|=n$。每个样本数据为一个 d 维矢量，用 $\boldsymbol{x}^{(i)}$ 表示第 i 个样本数据：

$$\boldsymbol{x}^{(i)}=\left[x_1^{(i)},x_2^{(i)},\cdots,x_d^{(i)}\right] \tag{1}$$

其中，$i=1,2,\cdots,n$。

自编码器是一种神经网络[11]，旨在以无监督的方式学习一个恒定函数（identity function，IF），以重建原始输入，同时在这个过程中压缩数据，从而形成一个有效的压缩表示。

自编码器由编码器和解码器两部分组成。前者将输入的高维数据转换成低维的输出数据，即压缩表示；后者则从低维表示中重构出原输入数据。与主成分分析或矩阵分解法不同，自编码器更侧重于学习如何从压缩表示中重构原始输入。高质量的中间表示能够捕获潜在变量，从而增强解压缩过程。

自编码器的基本结构如图 1 所示，图 1 中 $g(\cdot)$ 和 $f(\cdot)$ 分别表示编码器和解码器网络，并用 ϕ 和 θ 表示它们各自的网络参数。编码器 g_ϕ 将原始的高维输入 $\boldsymbol{x}$ 压缩成低维表示，$\boldsymbol{z}=g_\phi(\boldsymbol{x})$。解码器则利用该低维表示重构输入，$\boldsymbol{x}'=f_\theta(\boldsymbol{z})=f_\theta(g_\phi(\boldsymbol{x}))$，其中 $\boldsymbol{x}'$ 表示解码器对输入 $\boldsymbol{x}$ 的重构。

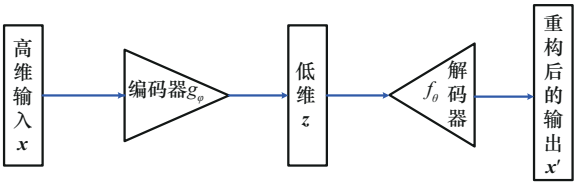


图 1　自编码器的基本结构

在训练阶段，参数 (ϕ,θ) 被联合学习，进而使重构的 $\boldsymbol{x}'$ 逼近于原始输入 $\boldsymbol{x}$，$\boldsymbol{x}\approx\boldsymbol{x}'$，即 $f_\theta(g_\phi(\boldsymbol{x}))$ 逼近于 $\boldsymbol{x}$。有多个指标可评估两个矢量间的相似度，如交叉熵、均方误差损失。与基于交叉熵的损失函数相比，均方误差损失更易计算，自编码器（autoencoder，AE）的均方误差损失函数表示为：

$$L_{\mathrm{AE}}(\boldsymbol{x},\boldsymbol{z})=\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}^{(i)}-f_\theta\left(g_\phi\left(\boldsymbol{x}^{(i)}\right)\right)\right)^2 \tag{2}$$

1.2　稀疏自编码器

与自编码器不同，稀疏自编码器（sparse autoencoder，SAE）对隐藏单元的激活进行“稀疏”约束，增强了鲁棒性。此外，SAE 对具有稀疏特征的数据集特别有效，SAE 的“稀疏”约束确保在任何给定时间内只有小部分隐藏神经元被激活，大多数隐藏神经元在大部分时间内保持稳定。

假定第 ℓ 个隐藏层有 s_ℓ 个神经元。用 $a_j^{(\ell)}(\cdot)$ 表示神经元 s_ℓ 的激活函数，且 $j=1,2,\cdots,s_\ell$。用 $\hat{\rho}_j$ 表示神经元 s_ℓ 被激活的概率。通过实施稀疏性策略，SAE 鼓励隐藏单元仅捕获和表示输入数据的最相关和最有区别的特征，进而形成更有效和更有意义的数据表示，这种稀疏性策略提高了模型的泛化能力，增强了模型的鲁棒性。SAE 通过最小化式（3）所示的目标函数来训练模型：

$$\begin{aligned}L_{\mathrm{SAE}}(\boldsymbol{x},\boldsymbol{z})&=L(\theta)+\beta\sum_{\ell=1}^{L}\sum_{j=1}^{s_\ell}D_{\mathrm{KL}}\left(\rho\|\hat{\rho}_j^{(\ell)}\right)=\\&L(\theta)+\beta\sum_{\ell=1}^{L}\sum_{j=1}^{s_\ell}\rho\log\frac{\rho}{\hat{\rho}_j^{(\ell)}}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j^{(\ell)}}\end{aligned} \tag{3}$$

其中，$L_{\mathrm{SAE}}(\boldsymbol{x},\boldsymbol{z})$ 表示稀疏自编码器的损失函数，

ρ表示稀疏性参数，$\hat{\rho}_j^{(\ell)}$表示在第ℓ个隐藏层中第j个神经元的平均激活值，$D_{\mathrm{KL}}\left(\rho\|\hat{\rho}_j^{(\ell)}\right)$表示$\rho$与$\hat{\rho}_j^{(\ell)}$间的Kullback-Leibler散度，$\beta$为稀疏性限制条件的权重系数。

自编码器的编码和解码运算，将高维数据转换成低维数据，以实现对数据的降维，而SAE的隐藏层保持一定的稀疏性，能进一步对数据进行压缩。尽管这个过程会丢失一定的信息，但在训练过程中会尽可能减少所丢失的信息[12]。

2 基于特征降维的卷积注意力机制-残差网络的入侵检测模型

基于特征降维的卷积注意力机制-残差网络的入侵检测模型框架如图2所示，该模型主要包含3个阶段。首先，对数据进行预处理，即先对数据集UNSW-NB15和CICIDS 2017中的数据进行独热编码和归一化处理，以提高检测攻击的效率。其次，在数据降维阶段，将经特征工程技术处理后的数据作为数据降维阶段的输入数据，再利用堆叠降噪稀疏自编码器对输入数据进行编码和解码。最后，将卷积注意力机制和残差网络相结合，构建检测分类器，以提高对攻击检测的准确率。

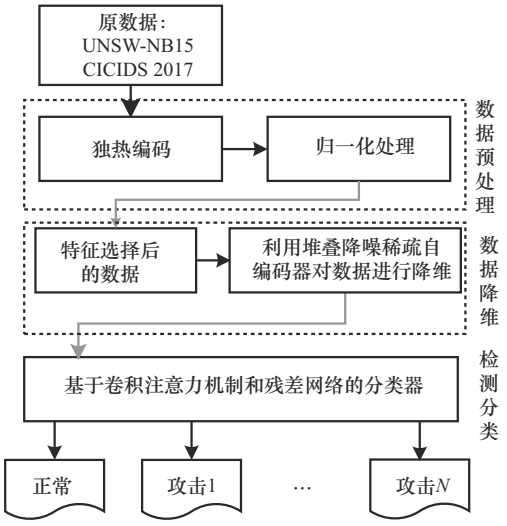


图2 基于特征降维的卷积注意力机制-残差网络的入侵检测模型框架

2.1 数据预处理

利用独特编码将数据集UNSW-NB15和CICIDS 2017中的符号数据转换成数值数据，并将它们的标签数据进行数值编码。由于不同特征数据间的维数差异较大，需要对这些数据进行归一化，使每个特征的取值分布在$[0,1]$。利用最大-最小方法对数据进行归一化处理：

$$\hat{x}=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{4}$$

其中，x表示原始数据，$x_{\min}$和$x_{\max}$分别表示原始数据所在列的最小值和最大值，$\hat{x}$表示对x归一化后的数据。

2.2 基于堆叠降噪稀疏自编码器的特征降维

降噪自编码器通过在原数据中添加额外噪声来增加检测模型的鲁棒性，降低训练过程中的过拟合风险。因此，将降噪自编码器与SAE相结合，充分利用它们各自的优点，形成降噪稀疏自编码器（denoising SAE，DSAE）。

堆叠DSAE是将两个或多个DSAE组合而成，第1个DSAE的隐含层作为第2个DSAE的输入。基于堆叠降噪稀疏自动编码器的框架如图3所示，图3给出了由两个DSAE构成的堆叠DSAE结构。

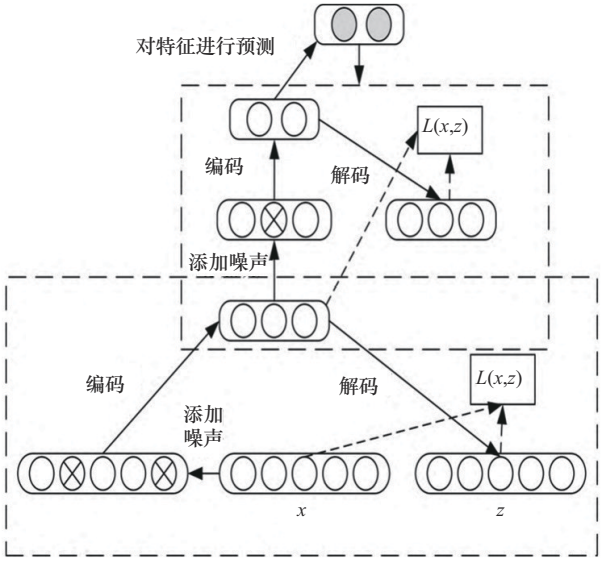


图3 基于堆叠降噪稀疏自动编码器的框架

堆叠DSAE通过逐层编码，提取出深层且有价值的特征，剔除冗余、无价值的特征。在训练模型时，先采用贪婪式分层训练权重策略，逐层进行初始化，然后采用微调的训练方式训练整个网络架构[13]。

2.3　基于卷积注意力机制和残差网络的分类器

作为经典的卷积神经网络（convolutional neural network，CNN），残差网络在堆叠的CNN中引入跳跃结构，实现了跨层连接。残差网络结构如图4所示，其中输入数据$\boldsymbol{x}$一方面通过卷积网络学习得到残差函数$F(\boldsymbol{x})$，另一方面通过跨层连接直接得到$\boldsymbol{x}$，最终形成$F(\boldsymbol{x})+\boldsymbol{x}$。

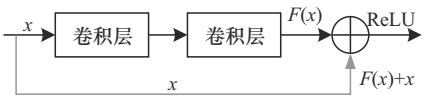


图4　残差网络结构

残差网络更容易学习，并且在加大网络深度的同时，也不会出现梯度消失的现象，提高了训练模型的效率。为了进一步提高检测攻击的准确率，将卷积注意力机制（convolutional block attention mechanism，CBAM）融入残差网络，形成改进的残差网络。CBAM是一个轻量的注意力模块，它从通道和空间两个维度计算注意力权重，以提升模型的性能。它主要由通道注意力和空间注意力两个模块组成，前者关注哪个通道上的特征是有意义的，而后者关注空间中哪些特征是有意义的。此外，由于CBAM是一个低复杂、轻量级模块，它能与各类卷积神经网络结合。为此，将CBAM融入残差网络，对传统的残差网络进行改进，改进的残差网络结构如图5所示，在两个卷积层之间插入一个CBAM，旨在利用CBAM机制在提取深层特征方面的优势，从而提升检测攻击的精度。

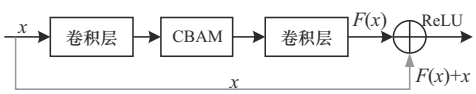


图5　改进的残差网络结构

基于CBAM的残差网络的分类模型结构如图6所示。输入数据先经卷积层（3×3 conv, 48）处理后输入改进的残差网络。在一个残差块3×3与1×1卷积核间插入CBAM。插入CBAM的目的在于选择更优的特征。经过1×1卷积核处理后，可减少训练模型的参数量，提升训练模型的效率。

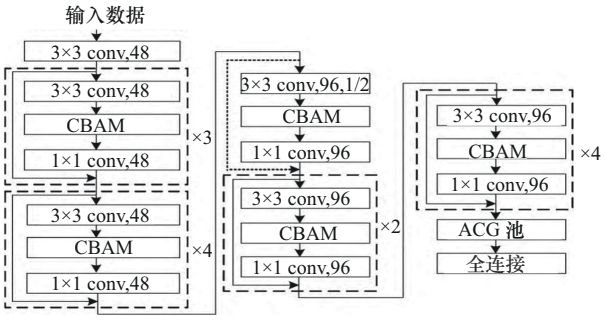


图6　基于CBAM的残差网络的分类模型结构

3　实验及性能分析

利用Jupyter Notebook软件编写检测程序。运行程序的计算机参数：Intel四核处理器（i5-10210U 1.6 GHz），64位操作系统，Windows 11。

3.1　评价指标

为不失一般性，选用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和F1-score这4个指标评估IRFD的性能。这4个指标依赖于TP、TN、FP和FN，它们的释义见表1。

表1　TP、TN、FP和FN的释义

符号	变量名	释义
TP	真阳性	正确分类为攻击类别的攻击样本数量
TN	真阴性	正确分类为正常类别的正常样本数量
FP	假阳性	错误分类为攻击类别的正常样本数量
FN	假阴性	错误分类为正常类别的攻击样本数量

准确率（Accuracy）的定义如下所示。

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{5}$$

精确率（Precision）的定义如下所示。

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{6}$$

召回率（Recall）的定义如下所示。

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{7}$$

F1-score的定义如下所示。

$$\text{F1-score}=2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{8}$$

3.2 基线模型

为了有效地分析IRFD的性能，选择基于机器学习和基于深度学习的检测模型作为基线模型，对比分析它们的性能。采用3种具有代表性的机器学习算法进行对比实验，分别是伯努利朴素贝叶斯（Bernoulli naïve Bayes，BNB）、决策树（decision tree，DT）和支持向量机（support vector machine，SVM）。同时，还选取了两个基于深度学习的检测模型作为基线模型，同IRFD进行对比实验，它们分别是文献[14]提出的基于深度神经网络的两层框架的入侵检测（deep learning network based two-stage IDS，DTSS）模型和文献[15]提出的多层框架的入侵检测模型（multi-stage intrusion detection framework，MSDF），它们均采用深度神经网络构建入侵检测模型。

实验中的BNB、DT和SVM模型参数设置如下。

（1）BNB分类器采用默认参数取值，即拉普拉斯参数$\lambda=1.0$、二值化参数为0。

（2）DT分类器的参数设置：选用基尼系数作为特征划分方法；树的最大深度math_depth为None，即不限制深度；最小样本数min_samples_split为2，其他参数采用默认参数。

（3）SVM：内核kernel为Sigmoid，惩罚系数为1.0，核系数gamma=5，最大迭代次数为50，其他参数取默认参数。此外，DTSS和MSDF模型的参数取值与原文的参数取值相同。

3.3 模型的收敛速度

IRFD训练样本的损失函数值随迭代次数（epoch）的变化情况如图7所示。由图7可知，最初损失值（Loss）随迭代次数的增加迅速下降，当迭代次数达到约30次时，Loss开始趋于平稳。这表明，提出的IRFD能够快速收敛。

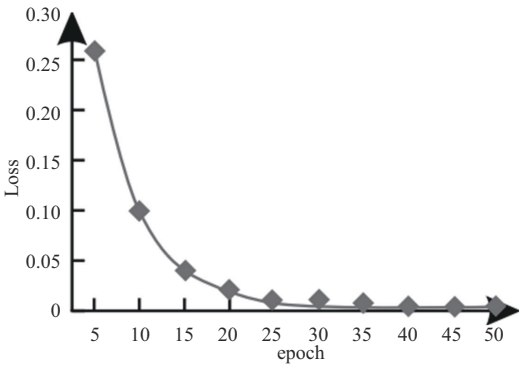


图7 损失函数值随迭代次数epoch的变化情况

3.4 基于数据集UNSW-NB15的性能分析

数据集UNSW-NB15[16]是由澳大利亚网络安全中心收集的实验数据，其包含正常流量和模糊测试、拒绝服务、后门等典型的网络攻击。实验中所采用的数据集UNSW-NB15的各类攻击分布信息见表2，其中包含了175 341条记录，每条记录有47个特征。

表2 数据集UNSW-NB15的各类攻击分布信息

类别	训练集		测试集	
	数量	比例	数量	比例
正常	56 000	31.94%	37 000	44.94%
泛型	40 000	22.81%	18 871	22.92%
漏洞利用	33 393	19.04%	11 132	13.52%
模糊测试	18 184	10.37%	6 062	7.36%
拒绝服务	12 264	6.99%	4 089	4.97%
踩点	10 491	5.98%	3 496	4.25%
渗透分析	2 000	1.14%	677	0.82%
后门	1 746	1.00%	583	0.71%
壳代码	1 133	0.65%	378	0.46%
蠕虫	130	0.07%	44	0.05%
总数	175 341	100%	82 332	100%

IRFD对数据集UNSW-NB15各类样本的分类准确率如图8所示。由图8可知，IRFD对正常、泛型和渗透分析样本的分类准确率达到0.98以

上，而对后门、蠕虫和壳代码样本的分类准确率不及0.8，原因在于这3种样本的训练数量较少，如蠕虫的训练样本数只有130，占比不及1%。极少的训练样本数，降低了检测准确率。

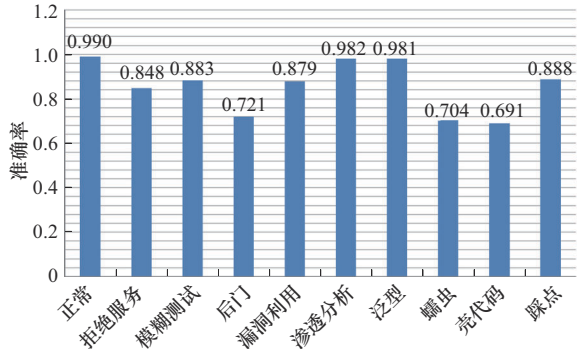
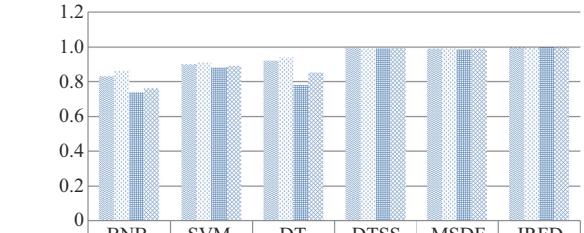


图8　IRFD对数据集UNSW-NB15各类样本的分类准确率

IRFD和基线模型的检测性能（UNSW-NB15数据集）如图9所示。由图9可知，基于传统的机器学习方法的IDS模型（BNB、SVM和DT）的检测性能不及基于深度神经网络的IDS模型（DTSS、MSDF和IRFD）。例如，DT模型的Accuracy、Precision、Recall和F1-score分别为0.920、0.940、0.780和0.852，而DTSS模型的Accuracy、Precision、Recall和F1-score分别为0.994、0.990、0.992和0.992。原因在于传统的机器学习方法为浅层次方法，难以学习到数据的深层特征。相反，深度神经网络通过反复地学习数据，掌握了数据的深层特征，这有利于提高检测率。



	BNB	SVM	DT	DTSS	MSDF	IRFD
Accuracy	0.83	0.901	0.920	0.994	0.991	0.996
Precision	0.86	0.913	0.940	0.99	0.991	0.995
Recall	0.74	0.883	0.780	0.992	0.984	0.995
F1-score	0.76	0.890	0.852	0.992	0.991	0.995

图9　IRFD和基线模型的检测性能（UNSW-NB15数据集）

同属深度神经网络方法，IRFD的检测性能优于DTSS和MSDF模型。IRFD的Accuracy、Precision、Recall和F1-score分别为0.996、0.995、0.995和0.995，均达到99%以上。原因在于IRFD先通过降噪稀疏自编码器对数据进行降维，提取了优质特征的数据，再利用残差网络构建分类器，提高了检测性能。

除了检测精度，检测效率也是评估检测模型的一项重要指标。基线模型和IRFD的执行时间（数据集UNSW-NB15）如图10所示，图10列出了BNB、SVM、DT、DTSS、MSDF和IRFD的执行时间，包括训练模型和测试阶段所消耗的时间。

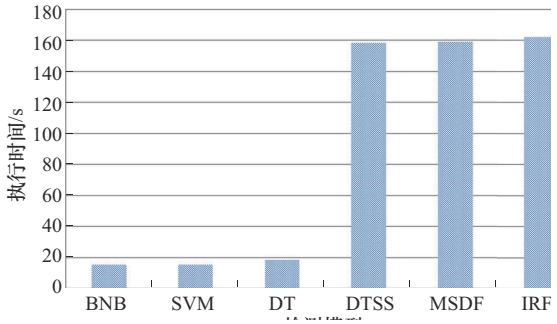


图10　基线模型和IRFD的执行时间（数据集UNSW-NB15）

由图10可知，基于传统机器学习的入侵检测模型的执行时间远低于基于深度学习的入侵检测模型的执行时间。BNB、SVM和DT的执行时间均小于20 s，而DTSS、MSDF和IRFD的执行时间约在160 s。原因在于DTSS、MSDF和IRFD需要反复学习数据、训练模型，这增加了训练时间。此外，IRFD的执行时间最高，达到了162 s。

结合图9和图10的数据可知，高的检测性能是以高的检测时间为代价的。原因在于训练准确的检测模型需要深层次学习数据，需要反复训练模型，这必然增加了执行时间。因此，可根据应用需求，选择相应的检测模型，若应用场景对检测时间有苛刻要求，则可以选择检测时间短的检测模型，反之，则可以选择检测性能高的检测模型。

3.5 基于数据集CICIDS 2017的性能分析

数据集CICIDS 2017是入侵检测的专用数据集[17]，它由加拿大网络安全研究所采集分布，格式类似KDD99。数据集CICIDS 2017包含了多个良性和最新的常见网络攻击，被广泛用于分析车载网络的入侵检测算法的性能。考虑原CICIDS 2017的样本数据较大，笔者采用分层抽样方法，即按规定的比例从不同类中随机抽取样本。分层抽样的优点是样本代表性较好，且抽样误差小。从数据集CICIDS 2017中抽取了56 661条样本数据进行训练和测试。数据集CICIDS 2017的各类攻击分布信息见表3。

表3 数据集CICIDS 2017的各类攻击分布信息

类别	训练集		测试集	
	数量	比例	数量	比例
正常	18 185	40.11%	4 546	40.11%
拒绝服务	15 228	33.59%	3 807	33.59%
端口扫描	6 357	14.02%	1 589	14.02%
暴力	2 214	4.88%	553	4.87%
网络攻击	1 744	3.84%	436	3.84%
网络机器人	1 572	3.46%	394	3.47%
渗入威胁	28	0.06%	8	0.07%

表3列出了样本数据在正常、拒绝服务、端口扫描、暴力、网络攻击、网络机器人和渗入威胁攻击这7类样本的数量及占比。先分析IRFD对上述7类样本的分类性能，IRFD对数据集CICIDS 2017中7类样本的分类准确率如图11所示。

由图11可知，IRFD可较准确地对这7类样本进行分类，对正常、拒绝服务、端口扫描和暴力的分类准确率达到90%以上，只有对渗入威胁分类的准确率未达到0.8。原因在于渗入威胁样本数量非常少，供训练的样本数只有28条。此外，观察图11不难发现，准确率随训练样本数的下降而下降。

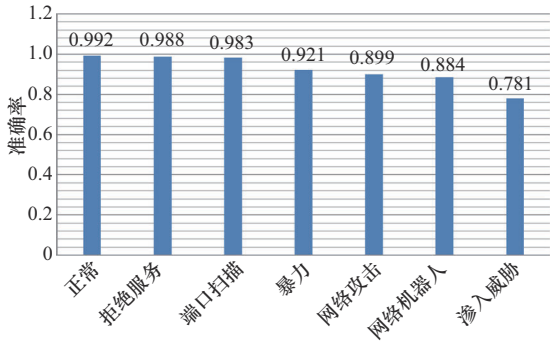
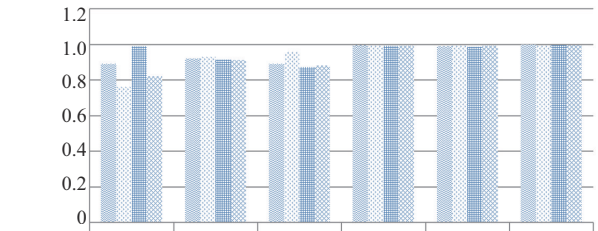


图11 IRFD对数据集CICIDS 2017中7类样本的分类准确率

IRFD和基线模型的检测性能（CICIDS 2017数据集）如图12所示。由图12可知，IRFD的整体的检测性能优于基线模型，它的Accuracy、Precision、Recall和F1-score分别为0.998、0.996、0.995和0.997。这说明IRFD通过提取特征以及基于改进残差网络的分类器可对样本有效地进行分类，从而提升了检测攻击的性能。这些数据表明，IRFD能够较准确地对网络流量进行分类，能鉴别正常和攻击。



	BNB	SVM	DT	DTSS	MSDF	IRFD
Accuracy	0.891	0.922	0.891	0.995	0.991	0.998
Precision	0.762	0.931	0.961	0.992	0.994	0.996
Recall	0.991	0.915	0.871	0.992	0.986	0.995
F1-score	0.821	0.912	0.881	0.992	0.993	0.997

图12 IRFD和基线模型的检测性能（CICIDS 2017数据集）

基线模型和IRFD的执行时间（CICIDS 2017数据集）如图13所示。与图10类似，图13也列出了IRFD和基线模型的执行时间，此执行时间包括训练时间和检测时间。由图13可知，基于深度神经网络的入侵检测模型（DTSS、MSDF和IRFD）的执行时间远高于基于机器学习的入侵检测模型（BNB、SVM和DT）。BNB、SVM和DT模型的执行时间分别为约12.4 s、13.1 s和12.2 s，

均不超过 15 s，而 DTSS、MSDF 和 IRFD 的执行时间分别达到 150.2 s、148.2 s 和 152.1 s。原因在于训练模型需要较长时间，特别是训练神经网络模型，而机器学习算法复杂度较低，执行时间较短。

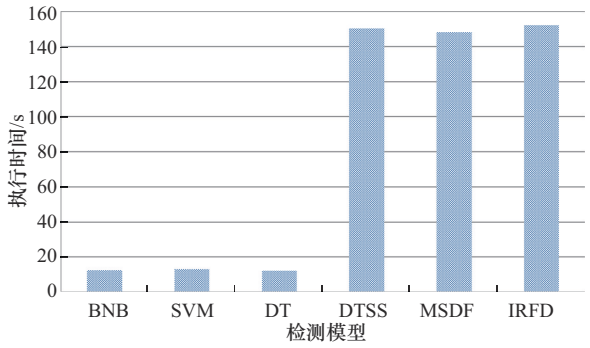


图 13　基线模型和 IRFD 的执行时间（CICIDS 2017 数据集）

4　结束语

本文提出了结合特征降维的改进残差网络入侵检测模型 IRFD。IRFD 利用堆叠降噪稀疏自编码器提取有效特征，以实现对数据的降维处理，并通过残差网络构建分类器，提升对攻击样本检测的准确率。性能分析表明，提出的 IRFD 具有较好的检测性能，在数据集 UNSW-15 和 CICIDS 2017 上均获取了不低于 99.5% 的 F1-score。然而，IRFD 还存在一些不足，其相比于 BNB、SVM 和 DT，IRFD 的执行时间较长，这说明 IRFD 的复杂度较高。后续将进一步对算法进行改进，降低检测时间，拓展 IRFD 的应用场景。

参考文献：

[1] 吴昊, 郝佳佳, 卢云龙. 物联网场景下基于蜜场的分布式网络入侵检测系统研究[J]. 通信学报, 2024, 45(1): 106-118.
WU H, HAO J J, LU Y L. Research on distributed network intrusion detection system for IoT based on honeyfarm[J]. Journal on Communications, 2024, 45(1): 106-118.

[2] 刘奇旭, 肖聚鑫, 谭耀康, 等. 工业互联网流量分析技术综述[J]. 通信学报, 2024, 45(8): 221-237.
LIU Q X, XIAO J X, TAN Y K, et al. Survey of industrial Internet traffic analysis technology[J]. Journal on Communications, 2024, 45(8): 221-237.

[3] 刘涛涛, 付钰, 王坤, 等. 基于 VAE-CWGAN 和特征统计重要性融合的网络入侵检测方法[J]. 通信学报, 2024, 45(2): 54-67.
LIU T T, FU Y, WANG K, et al. Network intrusion detection method based on VAE-CWGAN and fusion of statistical importance of feature[J]. Journal on Communications, 2024, 45(2): 54-67.

[4] FARNAAZ N, JABBAR M A. Random forest modeling for network intrusion detection system[J]. Procedia Computer Science, 2016, 89: 213-217.

[5] NUO Y. A novel selection method of network intrusion optimal route detection based on naive Bayesian[J]. International Journal of Applied Decision Sciences, 2018, 11(1): 1-17.

[6] 赵文灏, 陈曦. 增强支持向量机和遗传算法的 WSN 安全研究[J]. 计算机应用与软件, 2024, 41(2): 300-304, 327.
ZHAO W H, CHEN X. Research on WSN security based on enhanced support vector machine and genetic algorithm[J]. Computer Applications and Software, 2024, 41(2): 300-304, 327.

[7] 肖衡, 龙草芳. 基于机器学习的无线传感网络通信异常入侵检测技术[J]. 传感技术学报, 2022, 35(5): 692-697.
XIAO H, LONG C F. Communication anomaly intrusion detection technology for wireless sensor networks based on machine learning[J]. Chinese Journal of Sensors and Actuators, 2022, 35(5): 692-697.

[8] 程超, 武静凯, 陈梅. 一种基于 RBM-SVM 算法的无线传感网络入侵检测算法[J]. 计算机应用与软件, 2022, 39(5): 325-329.
CHENG C, WU J K, CHEN M. An intrusion detection algorithm for wireless sensor network based on RBM-SVM algorithm[J]. Computer Applications and Software, 2022, 39(5): 325-329.

[9] 杨晓文, 张健, 况立群, 等. 融合 CNN-BiGRU 和注意力机制的网络入侵检测模型[J]. 信息安全研究, 2024, 10(3): 202-208.
YANG X W, ZHANG J, KUANG L Q, et al. A network intrusion detection model integrating CNN-BiGRU and attention mechanism[J]. Journal of Information Security Research, 2024, 10(3): 202-208.

[10] 张明明, 刘凯, 李贤慧, 等. 基于深度学习的恶意行为检测与识别模型研究[J]. 信息安全研究, 2023, 9(12): 1152-1158.
ZHANG M M, LIU K, LI X H, et al. Research on malicious be-

havior detection and identification model based on deep learning[J]. Journal of Information Security Research, 2023, 9(12): 1152-1158.

[11] 谢胜利, 陈泓达, 高军礼, 等. 基于分布对齐变分自编码器的深度多视图聚类[J]. 计算机学报, 2023, 46(5): 945-959.
XIE S L, CHEN H D, GAO J L, et al. Deep multi-view clustering based on distribution aligned variational autoencoder[J]. Chinese Journal of Computers, 2023, 46(5): 945-959.

[12] 贺瑞芳, 赵堂龙, 刘焕宇. 基于去噪图自编码器的无监督社交媒体文本摘要[J]. 软件学报, 2024: 1-21.
HE R F, ZHAO T L, LIU H Y. Denoising graph auto-encoder for unsupervised social media text summarization[J]. Journal of Software, 2024: 1-21.

[13] 张甜甜, 赵庶旭, 王小龙. 基于堆叠降噪自编码器的肝癌亚型分类[J]. 计算机应用与软件, 2024, 41(6): 79-84.
ZHANG T T, ZHAO S X, WANG X L. Classification of liver cancer subtypes based on stacked denoising autoencoder[J]. Computer Applications and Software, 2024, 41(6): 79-84.

[14] ALMUTLAQ S, DERHAB A, HASSAN M M, et al. Two-stage intrusion detection system in intelligent transportation systems using rule extraction methods from deep neural networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(12): 15687-15701.

[15] KHAN I A, MOUSTAFA N, PI D C, et al. An enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 25469-25478.

[16] 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]. 重庆邮电大学学报(自然科学版), 2022, 34(6): 1071-1078.
SHI Y B. Anomaly intrusion detection method based on variational autoencoder and attention mechanism[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, 34(6): 1071-1078.

[17] ENGELEN G, RIMMER V, JOOSEN W. Troubleshooting an intrusion detection dataset: the CICIDS2017 case study[C]// Proceedings of the 2021 IEEE Security and Privacy Workshops (SPW). Piscataway: IEEE Press, 2021: 7-12.

[作者简介]

孙敬（1991– ），女，郑州工业应用技术学院信息工程学院讲师，主要研究方向为计算机应用、人工智能。

丁嘉伟（1986– ），男，郑州工业应用技术学院信息工程学院副教授，主要研究方向为人工智能、管理信息系统。

冯光辉（1982– ），男，广州大学计算机科学与网络工程学院博士生，主要研究方向为差分隐私、联邦学习等。