Journal of Computer Applications 计算机应用 | ISSN 1001-9081 CODEN JYIIDU | 2019-05-31 http://www.joca.cn

文章编号:1001-9081(2019)05-0000-00 doi:10.11772/j.issn.1001-9081.2019020327

基于改进卷积神经网络的网络入侵检测方法

杨宏宇*，王峰岩

(中国民航大学 计算机科学与技术学院，天津 300300)

(*通信作者电子邮箱 yhyxlx@hotmail.com)

摘 要: 针对基于深度学习的网络入侵检测技术存在检测效率低、模型训练易出现过拟合和泛化能力较弱的问题，提出一种基于改进卷积神经网络（Improved Convolutional Neural Network，ICNN）的入侵检测模型（ICNN Based Intrusion Detection Model，IBIDM）。与传统“卷积-池化-全连接”层叠式网络设计方式不同，该模型采用跨层聚合网络的设计方式，首先将预处理后的训练集数据作为输入数据前向传播并提取网络特征，对跨层聚合网络的输出数据执行合并操作；然后，根据分类结果计算训练误差并通过反向传播过程进行迭代优化至模型收敛；最后，利用训练好的分类器对测试数据集进行分类测试。实验结果表明，IBIDM 具有较高的入侵检测准确率和真正率，且误报率较低。

关键词: 网络入侵检测；卷积神经网络；前向传播；跨层聚合；迭代优化

中图分类号:TP18;TP393.08 **文献标志码**：**A**

Network intrusion detection model based on improved convolutional neural network

YANG Hongyu*, WANG Fengyan

(School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aiming at the problems of network intrusion detection technology based on deep learning, such as low detection efficiency, easy over-fitting and weak generalization ability of model training, this paper proposes an improved convolutional neural network (ICNN) based Intrusion Detection Model (IBIDM). Different from the traditional "convolution-pool-full connection" cascading network design method, the model adopted the design method of cross-layer aggregation network. Firstly, the pre-processed training set data was forwardly propagated as input data and network features were extracted, and the merge operation was performed on the output data of the cross-layer aggregation network. Then, the training error was calculated according to the classification result and iteratively optimized to the model convergence by the back propagation process. Finally, a classification test experiment was performed on the test data set using a trained classifier. The experimental results show that the IBIDM has high intrusion detection accuracy and true positive rate, and the false positive rate is low.

Keywords: network intrusion detection; convolutional neural network; forward propagation; cross-layer aggregation; iterative optimization

0 引言

随着新型网络攻击特征的不断出现，适应性强且稳定有效的入侵检测方法成为一项迫切需要。目前，通用的网络身份验证机制和防火墙技术虽然能够满足用户基本的安全防护需求，但是防护能力相对较弱，一旦遭遇专业黑客的恶意攻击，这些防护措施就形同虚设。目前以误用检测[1,2]和异常检测[3-6]为代表的入侵检测方法普遍存在检测精度低和特征提取效率低、误报率高等不足。随着人工智能方法在入侵检测系统（Intrusion Detection System，IDS）中的应用研究，基于人工智能的检测方法已经成为 IDS 研究的热点之一。

目前在入侵检测方法中应用的人工智能方法主要包括神经网络、遗传算法和免疫算法等，例如基于深度 BP 神经网络的入侵检测[7]和基于模糊系统的协同演化网络入侵检测[8]。这些方法在对数据集的处理过程中容易出现关键信息丢失的情况，造成样本数据集“失真”，数据集样本特征提取效率不高，导致检测结果波动性较大。为解决此问题，陈虹等[9]提出基于优化数据处理的深度置信网络入侵检测方法，Qu 等[10]提出基于深度置信网络的入侵检测模型，Yin 等[11]提出基于递归神经网络的入侵检测的深度学习方法，Shone 等[12]提出基于无监督特征学习的非对称深度自动编码器（Nonsymmetric deep autoencoder，NDAE），Yuan 等[13]

收稿日期: 2019-02-27； **修回日期**:2019-04-05； **录用日期**:2019-05-06。

基金项目：国家自然科学基金民航联合研究基金项目(U1833107)；国家科技重大专项 (2012ZX03002002).

作者简介：杨宏宇(1969-)，男，吉林长春人，教授，博士，CCF 会员，主要研究方向：网络信息安全；王峰岩(1993-)，男，河南南阳人，硕士研究生，主要研究方向：网络与信息安全。

利用遗传算法和支持向量机（Support Vector Machine，SVM）建立了一种网络入侵检测分类器，魏明军等[14]提出一种多种群克隆选择算法，通过改变该算法的匹配规则，进行入侵检测仿真实验。上述 5 种方法在一定程度上提高了入侵检测的准确率，但在实验过程中参数调优比较困难，且特征提取效率较低，计算任务量大。为了克服上述方法的不足，贾凡等[15]和王明等[16]利用卷积神经网络，分别实现基于卷积神经网络的入侵检测算法（Intrusion Detection Algorithm Based on Convolutional Neural Network，IDABCNN）和网络入侵检测系统（Network Intrusion Detection Model Based on Convolutional Neural Network，NIDMBCNN）。与其它机器学习方法相比，采用卷积神经网络的入侵检测方法显著提高了分类的准确性，但该方法在模型训练过程中的收敛速度不理想，泛化能力差，导致真正率较低并且误报率较高。卷积神经网络（Convolutional Neural Network，CNN）作为一种半监督式神经网络，由于具备将低级入侵流量数据特征抽象表示为高级特征的能力且特征学习能力突出，所以近年来逐渐被应用于入侵检测领域。Donghwoon 等[17]建立了三个不同深度的 CNN 模型，验证了网络深度对入侵检测性能的影响，Vinayakumar[18]等通过时间序列的方法对网络流量进行建模，利用 CNN 及 CNN-LSTM 等变体架构分别进行了入侵检测实验。通过对文献[17]和[18]的分析表明，与检测效果良好的变分自动编码器（Variational Auto-Encoder，VAE），多层感知机（Multi-Layer Perception，MLP）等检测模型相比，CNN 的特征提取能力优势明显且分类效果更好。

上述研究虽然在样本识别能力和性能上均有提升，但是在网络训练上存在过拟合和泛化能力差等不足，检测精度和检测效率还有待提高。为了避免网络训练过拟合并提升泛化能力，本文利用卷积神经网络的结构特性并结合跨层聚合设计理念，提出一种基于深度学习的入侵检测模型。该模型以本文提出的改进卷积神经网络为基础，结合跨层设计方式，利用预处理后的原始样本数据集进行模型训练，经过循环特征提取和迭代优化，使模型达到良好的收敛效果。通过对已训练好的分类器进行分类测试，实验结果验证了本文方法具有较好的检测效果。

1 改进卷积神经网络

1.1 设计思路

针对传统的分类检测算法在训练过程中存在的大量的关键参数需要人为设置，且训练过程中容易导致关键特征信息丢失和参数调优困难等问题。本文考虑利用卷积神经网络(Convolutional Neural Network，CNN)的端到端半监督式的网络训练分类器，利用 CNN 多层特征检测网络，在数据训练过程中自主学习样本特征并发现其中的规律，无需人为设置大量关键参数，达到简化实现过程的目的。为此，本文提出一种改进卷积神经网络（ICNN），其设计思路如下：

（1）采用跨层聚合的网络设计方式，使入侵检测模型从第二次卷积操作开始，将卷积后的结果保存，之后再单独进行卷积、池化、全连接操作。

（2）对第 3 次卷积操作的输出结果执行同样操作后，使用 Tensorflow 中的 concat()函数对跨层聚合网络的输出数据执行合并操作。

（3）根据 SoftMax 的分类结果计算 Loss 值，进行反向传播，通过迭代优化网络权值和偏置，直至达到良好的收敛效果。

1.2 改进卷积神经网络结构

改进卷积神经网络（ICNN）结构如图 1 所示。ICNN 包含 5 个卷积层、2 个池化层、4 个全连接层和 1 个 SoftMax 层。其中，Conv1，Conv2，Conv33 个卷积层使用 Relu 激活函数以增大网络稀疏性。为防止训练过程中出现过拟合现象，在跨层聚合模块的两个全连接层 FC1_layer 和 FC3_layer 使用正则化方法 Dropout。

该卷积网络结构中的 concat()是执行合并操作的函数，Softmax 层（Softmax Layer）用于对入侵检测模型的训练输出结果进行分类。

1.3 模型训练

ICNN 的模型训练由前向传播过程和反向传播过程组成。

（1） 前向传播

在 ICNN 中，训练的前向传播过程设计如下：

首先，在网络中分批次进行训练，每次训练都从预处理后的训练数据集中随机选取一个固定大小的块（batch）作为输入，训练时输入的数据参数维度按照(batch_size,H,W,channel)四维参数设置。每次训练时，从数据集选取大小为 M 的块（batch），输入数据的高度和宽度分别设为 1 和 122，通道（channel）为单通道。

在 ICNN 中（如图 1 所示），Input Data 首先经过两个卷积层（Conv1_Relu 和 Conv2_Relu），利用卷积层的多卷积核对初始输入的所有特征图进行卷积运算，每一个卷积核都对应一个特定特征图进行网络权重学习（即网络特征提取），Conv2_Relu 输出结果在激活函数 Relu 的作用下分别作为 Conv3_Relu 和 Conv4 的输入数据，卷积运算和激活函数 Relu 公式定义为：

$$y_j^l = s\left(\sum_{i\in M_j} y_i^{l-1} w_{ij}^l + b_i^l\right) \qquad (1)$$

$$Relu(y)=\begin{cases} y & (y>0) \\ 0 & (y\le 0) \end{cases} \qquad (2)$$

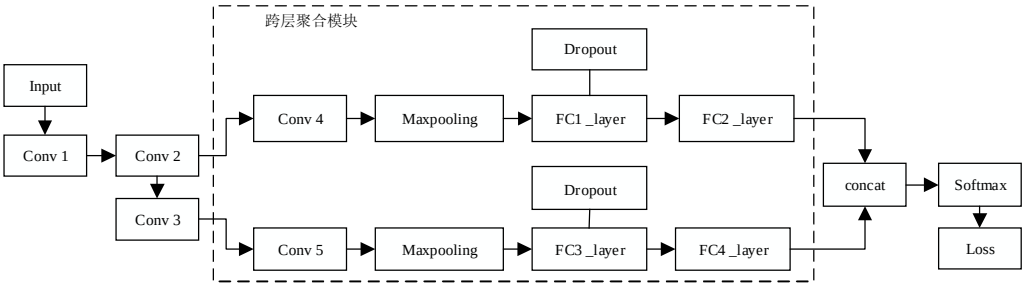

图 1 ICNN 结构

Fig.1 ICNN structure

其中，y_j^l 表示每一个卷积层经过一个卷积核的处理得到的结果，l 表示卷积层数，j 表示卷积核的序列号，σ 是激活函数（两个卷积层均使用激活函数），w 表示权重，b 为偏置。然后，由跨层聚合模块进行处理。对于卷积层（Conv4）每一个卷积核输出的特征图都会经过池化层（Max_pooling1）进行下采样操作，目的是对数据进行降维和区域最大化特征提取，输出的结果在两个全连接层 FC1_layer，FC2_layer 的作用下，得到 $M/2$ 维数据集合。池化层计算公式为：

$$z_j^l = b\left(w_j^l down\left(z_j^{l-1}\right)+b_j^l\right) \qquad (3)$$

其中，$down(z)$表示对矩阵元素 z 的下采样操作。池化操作和普通的神经元计算类似，也有其权重和偏置。

同样地，第 3 个卷积层（Conv3_Relu）的输出结果在卷积层（Conv5），池化层（Max_pooling2）和两个全连接层 FC3_layer，FC4_layer 的作用下，输出 $M/2$ 维数据集合。

然后利用 Tensorflow 中的 concat()函数对 FC2_layer 和 FC4_layer 层的输出结果执行合并操作，产生大小为 M 维的数据集合。

最后，用 SoftMax 层进行数据分类。

（2） 反向传播

在 ICNN 中，训练的反向传播过程设计如下：

首先，利用 Softmax 层对训练集的样本分类结果，计算出整体的误差参数值 Loss。

最后，根据 Loss 值进行反向传播（Back Propagation，BP）。反向传播的目的是根据实际输出值与理想输出值之间的误差迭代调整各层网络的权重和偏置，直至模型达到良好收敛效果。在反向传播过程中，为了快速找到最优权重 w 和偏置 b，使网络的输出 $f(x)$能够拟合所有的训练输入 x，设定一个损失函数 $C(w,b)$，用以找出最优的参数组合，以此量化模型拟合程度。而随机梯度下降（Stochastic Gradient Descent，SGD）算法提供了最小化此损失函数的方式。损失函数定义为：

$$C(w,b)\equiv\frac{1}{2n}\sum_x y^{(x)}-a^2 \qquad (4)$$

其中，w 表示网络权重的集合，b 为所有偏置的集合，n 是训练输入数据的个数，a 表示当输入为 x 时输出的向量。参数 w 和 b 的定义为：

$$w\to w_k^{'}\equiv w_k-h\frac{qC}{qw_k} \qquad (5)$$

$$b\to b_l^{'}\equiv b_l-h\frac{qC}{qb_l} \qquad (6)$$

其中，η 表示学习率，偏导数 $\theta C/\theta w_k$ 和 $\theta C/\theta b_l$ 表示任意权重和任意偏置的变化率。

损失函数值计算过程如图 2 所示。计算步骤设计如下：

步骤 1　设置初始激活值 a^1 并输入；

步骤 2　计算加权和 $z^l=w^l a^{l-1}+b^l$ 和各层节点激活值 $a^l=\sigma(z^l)$，其中 $l=(1,2,3,\ldots,L)$，进行前向传播；

步骤 3　计算各网络的输出层误差 $\delta^l=$ s $_a c$ ¤ $\sigma' z^L$ 并输出；

步骤 4　根据获取的每一项输出层误差 $\delta^l=((w^{l+1})^T\delta^{l+1}$ ¤ $\sigma' z^l)$，其中，$l=(L\text{-}1,L\text{-}2,\ldots)$，进行反向传播；

步骤 5　计算并输出损失函数的任意权重的变化率 $\theta C/\theta w_{jk}^l = a^{l-1}{}_k\,\delta_j^l$ 和任意偏置的变化率 $\theta C/\theta b_j^l=\delta_j^l$；

步骤 6　将步骤 5 的结果，分别代入公式（5）和（6）获取权重 w 和偏置 b，然后根据公式（4）获取损失函数值。

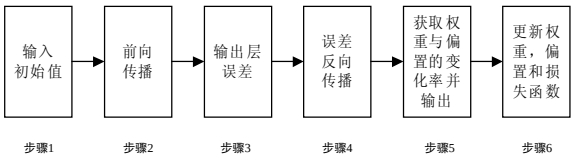

图 2 损失函数值计算流程图

Fig.2 Flow chart of loss function calculation

ICNN 的反向传播过程，是从最后一层开始向后计算误差向量，这种反向移动的进行是基于损失函数在网络中的输出结果。为了把握损失函数随前面层的权重和偏置变化的规律，通过随机梯度下降的方式和重复使用链式法则，反向获取所需的表达式结果，从而可以得到最优的参数组合（即最小化损失函数）。

在 ICNN 训练过程中，随着网络层数的不断增加和卷积核数量的不断增多，网络抽取到检测样本的特征信息也不断增多，网络自身通过不断的特征筛选和参数优化，最终使模型收敛，这体现了 ICNN 的深度学习过程对入侵检测样本建模的有效性。

2 网络入侵检测模型

2.1 网络入侵检测模型框架

以 ICNN 为基础，本文提出一个基于 ICNN 的网络入侵检测模型（ICNN Based Intrusion Detection Model，IBIDM），该模型主要由数据预处理模块、ICNN 网络训练模块和分类检测模块构成，IBIDM 的架构如图 3 所示。

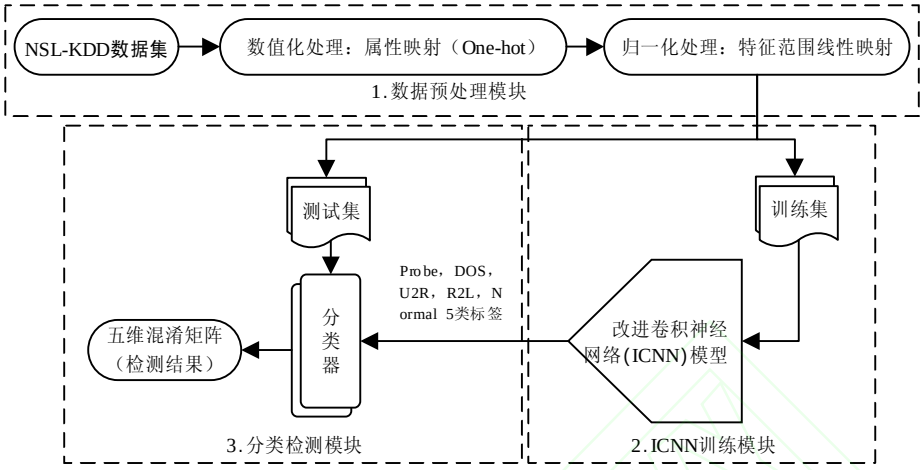

图 3 IBIDM 的架构

Fig.3 IBIDM architecture

2.2 模块设计

（1） 数据预处理模块

该模块包括数值化处理和归一化处理两项操作，其目的是为网络训练提供规范的输入数据集，该模块的处理过程设计如下：

首先，采用属性映射(One-hot)方法对数据集中连续型和离散型的符号型数据进行数值化处理。

然后，将处理后的数值型数据进行特征范围线性映射，得到规范的网络输入数据集；

（2）ICNN 训练模块

该模块包括前向传播特征提取和反向传播迭代优化两个过程。

前向传播特征提取：将预处理后的训练数据集作为 Input Data 进行前向传播，利用 ICNN 的自主学习能力进行特征提取。

反向传播迭代优化：根据 ICNN 训练得到的 Loss 值进行误差反向传播，经过反复的参数优化，直至达到良好收敛效果。

（3）分类检测模块

根据 Probe、DOS、U2R、R2L 和 Normal 五个样本类别为分类标签训练分类器，将预处理后的测试数据集作为 Test Data 输入训练好的分类器，分类器对检测样本进行分类检测，输出五维混淆矩阵，即为检测结果。

3 数据特征分析及预处理

3.1 数据集选取及特征描述

在本文研究中，选取 NSL-KDD CUP 数据集[19-20]作为基准数据集，用于检测模型的训练和性能测试。与 KDD CUP 99 数据集相比，NSL-KDD CUP 数据集删除了大量冗余数据，使其样本数据的比例分配更加均衡合理、可利用性更高，因此能够更好满足本研究的检测模型的验证实验需求。上述两个数据集的样本数据特征相同，数据集中每一项入侵记录均有 42 维特征，详细分为 38 维数字特征，3 维符号特征和 1 个攻击类型标签（label）。NSL-KDD CUP 数据集数据类型主要包含：正常类数据（Normal）和 4 大攻击类型数据（Probe，DOS，U2R，R2L），4 大攻击类型数据又可具体细分为 39 个小类。

NSL-KDD CUP 数据集主要包含训练集（KDDTrain）、测试集（KDDTest+）和测试集（KDDTest^{-21}）三个子数据集，其样本类别分布与特征分类情况分别如表 1 和表 2 所示。

表 1 样本类别分布表

Tab.1 Sample category distribution table

样本类别	训练集 (KDDTrain)	测试集 (KDDTest+)	测试集 (KDDTest^{-21})
Probe	45927	2421	4342
DOS	11656	7458	2402
R2L	995	2754	2754
U2R	52	200	200
Normal	67343	9711	2152
总量	125973	22544	11850

表 2 特征分类表

Tab.2 Feature classification table

序号	样本特征	类型	序号	样本特征	类型
1	duration	连续	22	is_guest_login	离散
2	protocol_type	符号	23	count	连续
3	service	符号	24	srv_count	连续
4	flag	符号	25	serror_rate	连续
5	src_bytes	连续	26	srv_serror_rate	连续
6	dst_bytes	连续	27	rerror_rate	连续

7	land	离散	28	srv_rerror_rate	连续
8	wrong_fragment	连续	29	same_srv_rate	连续
9	urgent	连续	30	diff_srv_rate	连续
10	hot	连续	31	srv_diff_host_rate	连续
11	num_failed_logins	连续	32	dst_host_count	连续
12	root_shell	离散	33	dst_host_srv_count	连续
13	num_compromised	连续	34	dst_host_same_srv_rate	连续
14	root_shell	离散	35	dst_host_diff_srv_rate	连续
15	su_attempted	离散	36	dst_host_same_src_port_rate	连续
16	num_root	连续	37	dst_host_srv_diff_port_rate	连续
17	num_file_creations	连续	38	dst_host_serror_rate	连续
18	num_shells	连续	39	dst_host_srv_serror_rate	连续
19	num_access_files	连续	40	dst_host_rerror_rate	连续
20	num_outbound_cmds	连续	41	dst_host_srv_rerror_rate	连续
21	is_host_login	离散	42	label	标签

注：连续型和离散型特征都属于数值特征

3.2 数据预处理

数据预处理包括数值化处理和归一化处理两个过程。

（1）数值化处理

由于 ICNN 的输入项为数字矩阵，因此采用独热（One-hot）编码方法，将测试数据集中具有符号型特征的数据映射为数字特征向量。该处理主要针对以下 3 个特征：

特征 1 对于 protocol_type 型特征的 3 种类型的属性：TCP，UDP 和 ICMP，将其分别编码为二进制向量（1,0,0），（0,1,0）和（0,0,1）；

特征 2 将 service 型特征所包含的 70 种符号属性通过编码变为 70 维二进制特征向量；

特征 3 将 flag 型特征所包含的 11 种类符号属性通过编码变为 11 维二进制特征向量。

经过数值化处理，上述 3 类符号型特征变成 84 维二进制特征向量，加上数据集本身的 38 维数字特征，数据集中每一项记录的 42 维特征最终变为 122 维二进制特征向量。

（2）归一化处理

在数据集中，连续型特征数据之间取值范围差异明显，例如，num_root 型特征的取值范围是[0,7468]，而 num_shells 型特征的取值范围是[0,5]，可见两者的最小值和最大值的范围差距很大。为了便于运算处理和消除量纲，采用归一化的处理方法，将每个特征的取值范围统一线性映射在[0,1]区间内。归一化计算公式为：

$$X_{norm}=\frac{X-X_{min}}{X_{max}-X_{min}} \tag{7}$$

其中，X_{max} 表示特征的最大值，X_{min} 表示特征的最小值。

4 实验与结果

4.1 实验环境配置与超参数设置

本文研究中，对检测模型的检测验证和对比测试实验均在 linux 操作系统上进行。使用 Python 的深度学习库 Tensorflow 编程实现本文的 ICNN 和 IBIDM。为提高运算效率，减少训练时间，采用 Tensorflow-GPU[21]进行并行计算加速。实验的软硬件配置环境如表 3 所示。

表 3 实验环境配置

Tab.3 Experimental environment configuration

硬件配置	软件环境
Intel Core i7-7700 CPU @ 2.80GHz	Ubuntu16.04
Nvidia GeForce GTX 1050(6G)	Python3.5.2
16.0GB RAM	Pycharm2017

经过多次和小范围的参数组合搭配训练，根据训练和测试结果，确定 ICNN 训练的超参数设置如下：

网络学习率为 0.1，此时网络的训练速度最佳，且特征学习过程较为详细；

权值衰减系数为 0.001，此时 ICNN 的复杂度对损失函数的影响最小；

正则化方式 Dropout 的权重失活率大小设为 0.5；

实验总的迭代训练次数为 300 次；

实验过程的每次迭代训练从训练集数据中选取的数据量大小是 1000 项，每个 batchsize 迭代训练次数为 50 次。

4.2 评价指标

本文采用准确率（Accuracy，AC）、真正率（True Positive Rate，TPR）和误报率（False Positive Rate，FPR）作为评估 IBIDM 检测性能的 3 个关键指标。

指标 1 *AC*：分类器准确预测的样本数与测试集总样本数的比值。

$$AC=\frac{TP}{TP+FP} \tag{8}$$

其中，*TP*（True Positive）为真正类，表示本属于正类的样本被准确预测为正类的样本数；*FP*（False Positive）为假正类，表示本属于负类的样本被错误预测为正类的样本数。

指标 2 *TPR*：分类器将正样本准确预测为正样本的数量与所有正样本数量的比值。

$$TPR=\frac{TP}{TP+FN} \tag{9}$$

其中，*FN*（False Negative）为假负类，表示本属于正类的样本被错误预测为负类的样本数。

指标 3 *FPR*：分类器将负样本错误预测为正样本的数量与所有负样本数量的比值。

$$FPR=\frac{FP}{FP+TN} \tag{10}$$

其中，*TN*（True Negative）为真负类，表示本属于负类的样

本被准确预测为负类的样本数。

4.3 实验设计

采用 KDDTest+和 KDDTest^{-21} 测试集分别开展 3 个检测实验。每个实验具体设计如下：

实验 1：以测试集 KDDTest+和 KDDTest^{-21} 中的五种数据类型 Normal，Probe，DOS，U2R，R2L 为 5 类标签，分别进行分类检测，根据分类检测模块所输出的五维混淆矩阵（检测结果），计算 IBIDM 对 4 个攻击类型的准确率、真正率和误报率指标，并根据评估指标值绘制两测试集关于误报率与真正率的 ROC 曲线图。

实验 2：应用 NSL-KDD CUP 数据集，对其它 3 种在入侵检测方法中应用的典型神经网络 LeNet-5[22]，DBN（Deep Belief Nets）[23]和 RNN（Recurrent Neural Network）[24]进行训练和检测，将检测结果与 IBIDM 的检测效果进行对比分析。

实验 3：应用 NSL-KDD CUP 数据集，对基于 CNN 入侵检测技术的 2 个典型入侵检测模型 IDABCNN 和 NIDMBCNN[15-16]分别进行训练和检测，将检测效果与 IBIDM 的检测效果进行对比分析。

4.4 实验结果

4.4.1 分类检测与结果

该实验包括训练和测试两个过程。在训练过程中，ICNN 模型的整体训练误差与迭代次数的关系如图 4 所示。由图 4 可见，随着迭代次数的增加训练误差逐渐减小；当迭代次数为 50 时，误差值最小，这说明 ICNN 模型结构设计和模型训练的超参数设置较为合理，能够满足检测要求。

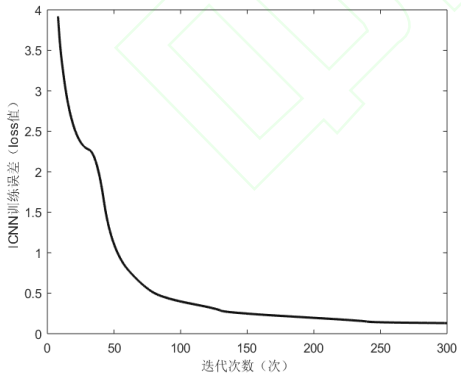

图 4 迭代次数与 ICNN 训练误差的关系图

Fig.4 Relationship between iteration number and ICNN training error

在测试过程中，KDDTest +和 KDDTest^{-21} 测试集在分类测试后输出的五维混淆矩阵分别如表 4 和表 5 所示，这两个矩阵即为具有五类标签样本的测试分类结果，其中粗体数字表示各样本类别被准确预测的数量。

表 4 KDDTest+测试集分类混淆矩阵

Tab.4 KDDTest+ test set classification confusion matrix

预测数量 / 实际数量	Probe	DOS	R2L	U2R	Normal
Probe	**2322**	52	6	0	41
DOS	75	**6946**	101	0	736
R2L	6	0	**2637**	7	104
U2R	9	0	6	**174**	11
Normal	27	75	176	14	**9419**

表 5 KDDTest^{-21} 测试集分类混淆矩阵

Tab.5 KDDTest^{-21} test set classification confusion matrix

预测数量 / 实际数量	Probe	DOS	R2L	U2R	Normal
Probe	**3986**	156	92	0	108
DOS	96	**1997**	128	0	171
R2L	77	0	**2219**	94	364
U2R	17	0	12	**148**	23
Normal	52	104	183	33	**1780**

测试集 KDDTest +和 KDDTest^{-21} 四类攻击类型的分类评估指标结果分别如图 5 和图 6 所示。

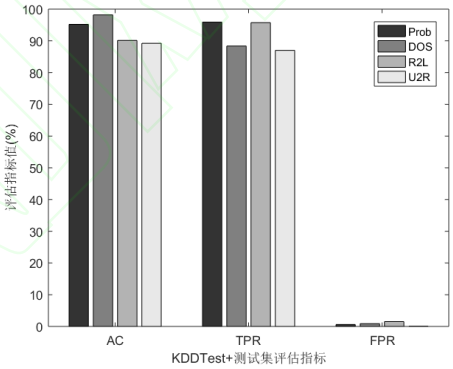

图 5 NSL KDDTest+测试集评估指标

Fig.5 NSL KDDTest+ test set evaluation index

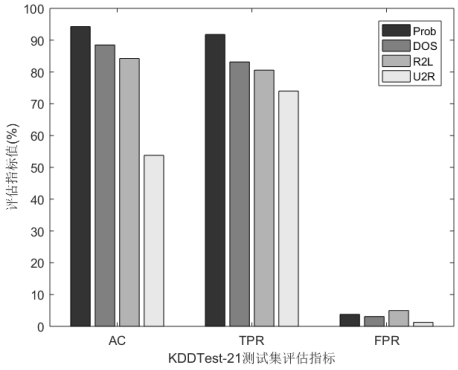

图 6 NSL KDDTest^{-21} 测试集评估指标

Fig.6 NSL KDDTest^{-21} test set evaluation index

由图 7 与图 8 可见，IBIDM 对 U2R 型数据的检测准确率低于其它 3 种类型，这是由于 U2R 的样本数据量太小，致使模型训练过程中，特征提取量较少，因而测试时分类器对其数据检测能力较弱。

为了更直观地评估 IBIDM 测试实验，利用测试集 KDDTest +与 KDDTest^{-21} 分类测试后的样本数据绘制了关于误报率（FPR）与召回率（TPR）的 ROC 曲线图，ROC 曲线如图 7 和图 8 所示。

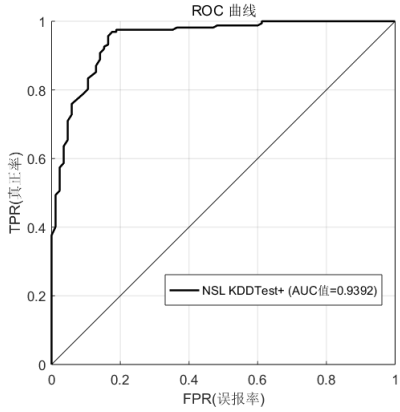

图 7 NSL KDDTest+测试集的 ROC 曲线

Fig.7 ROC curve of NSL KDDTest+ test set

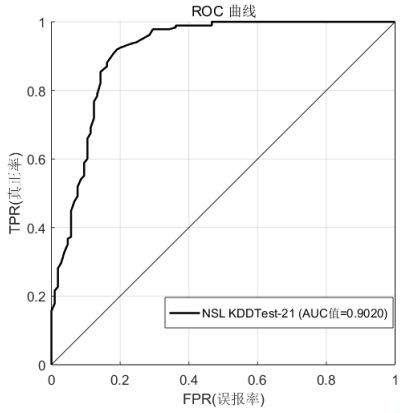

图 8 NSL KDDTest^{-21}测试集的 ROC 曲线

Fig.8 ROC curve of NSL KDDTest^{-21} test set

由图 7 和图 8 可见，IBIDM 五分类测试的 ROC 曲线拟合程度较好，通过计算得到 KDDTest +与 KDDTest^{-21} 的 AUC 的值分别为 0.9392 和 0.9020，说明通过 ICNN 训练的分类模型对正类样本的排序能力很强。

4.4.2 对比实验与结果

为了进一步验证 IBIDM 在网络入侵检测中的有效性，对在 IDS 中应用的 3 种典型神经网络 LeNet-5[22]，RNN[23]和 DBN[24]检测模型进行检测实验。分别对 IBIDM 和上述 3 个模型进行了五分类检测训练和检测测试。

表 6 检测结果对比

Tab.6 Comparison of test results

数据集	入侵检测模型	AC(%)	TPR(%)	FPR(%)
KDDTest+	IBIDM	92.94	95.55	0.76
	LeNet-5	86.54	91.31	2.03
	DBN	92.45	95.68	0.85
	RNN	93.08	94.39	0.92
KDDTest^{-21}	IBIDM	90.91	92.42	3.96
	LeNet-5	82.36	87.93	3.24
	DBN	88.69	89.48	4.46
	RNN	79.47	80.58	1.75

LeNet-5，RNN、DBN和IBIDM的检测实验结果对比如表6所示。由表6可见，在测试集KDDTest+上，IBIDM与其它3个模型相比，检测准确率高于LeNet-5和DBN，略低于RNN，真正率略低于DBN并且略高于LeNet-5和RNN，而误报率均低于其它3个模型；在测试集KDDTest^{-21}上，IBIDM的检测准确率和真正率均高于其它三类模型，误报率低于DBN，略高于LeNet-5和RNN，这说明ICNN对入侵检测数据样本的识别能力较强。

为了更好地验证IBIDM的有效性，对IDABCNN[15]和NIDMBCNN[16]分别进行模型训练和测试，为保证实验结果具有可比性，检测样本均使用NSL-KDD数据集，检测结果如表7所示。

表 7 3 种模型检测结果

Tab.7 3 model test results

数据集	入侵检测模型	AC(%)	TPR(%)	FPR(%)
KDDTest+	IBIDM	92.94	95.55	0.76
	IDABCNN	92.78	90.61	0.95
	NIDMBCNN	97.34	91.33	0.82
KDDTest^{-21}	IBIDM	90.64	91.74	3.87
	IDABCNN	89.33	87.49	3.63
	NIDMBCNN	88.16	83.71	1.02

由表7可见，在测试集KDDTest+上，IBIDM的检测的准确率低于NIDMBCNN，略高于IDABCNN，而真正率略高于IDABCNN和NIDMBCNN，误报率也更低，侧面表示出NSL-KDD数据集对提升实验检测效果更有帮助。在测试集KDDTest^{-21}上，IBIDM的检测准确率和真正率均高于其它两类模型，但误报率略高于后两者，所以IBIDM有进一步的提升空间。综上研究，说明IBIDM的跨层聚合的网络结构设计方式能够保证模型在训练过程中能够更好地提取特征信息，同时也反映出IBIDM的拟合程度较为理想，训练出的网络模型对样本数据分类效果较好。

5 实验与结果

针对目前基于深度学习技术的网络入侵检测方法普遍存在检测效率较低、模型训练过程中易出现过拟合和泛化能力较差的情况，本文提出一种基于改进卷积神经网络的网络入侵检测模型，利用预处理后的训练集和测试集数据，在 IBIDM 中进行了分类训练和测试实验。实验结果表明，IBIDM 的入侵检测准确率和真正率较高，误报率较低。总体上看，IBIDM 发挥了深度学习模型对样本数据的特征提取优势。

虽然本文方法在解决基于深度学习的入侵检测方法存在的诸多问题上取得了良好效果，但是应用深度学习技术进行入侵检测仍面临一大难题，即在模型优化过程中参数值无法收敛到全局最优，下一步对 IBIDM 的改进重点集中在以下两个方面：（1）针对随机梯度下降算法（SGD）在模型训练过程中易出现梯度弥散，损失函数易陷入局部最优解的情况，考虑使用模拟退火、粒子群算法或者蚁群算法等群智能优化算法代替其进行参数调优。（2）尝试使用其他数据集进行训练和测试验证实验，根据实验结果对算法和方法进行继续改进，进一步提升入侵检测模型的泛化能力和有效性。

参考文献

[1] Zheng K, Cai Z, Zhang X, et al. Algorithms to speedup pattern matching for network intrusion detection systems[J]. Computer Communications, 2015, 62(C): 47-58.

[2] Sung J S, Kang S M, Kwon T G. Pattern Matching Acceleration for Network Intrusion Detection Systems[C]// International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation. Berlin Heidelberg: Springer-Verlag, 2005: 289–298.

[3] 李鹏, 周文欢. 基于 k-means 和决策树的混合入侵检测算法[J]. 计算机与现代化, 2017(12): 12-16.

[4] 戚名钰, 刘铭, 傅彦铭等. 基于 PCA 的 SVM 网络入侵检测研究[J]. 信息网络安全, 2015(2): 15-18.

[5] Xu Y, Zhao H. Intrusion Detection Alarm Filtering Technology Based on Ant Colony Clustering Algorithm[C]// Sixth International Conference on Intelligent Systems Design and Engineering Applications. Washington, DC, USA: IEEE Computer Society., 2016: 470-473.

[6] Wang X. Design of temporal sequence association rule based intrusion detection behavior detection system for distributed network[J]. Modern Electronics Technique, 2018, 41(3): 108-114.

[7] Roy S S, Mallik A, Gulati R, et al. A Deep Learning Based Artificial Neural Network Approach for Intrusion Detection[C]// International Conference on Mathematics and Computing. Berlin Heidelberg Germany: Springer-Verlag, 2017: 44-53.

[8] Yong M A. A network intrusion detection schemer based on fuzzy inference and Michigan genetic algorithm[J]. Electronic Design Engineering, 2016, 24(11): 107-110.

[9] 陈虹, 万广雪, 肖振久. 基于优化数据处理的深度信念网络模型的入侵检测方法[J]. 计算机应用, 2017, 37(6): 1636-1643. (Chen H, Wang X, Xiao Z J. Intrusion detection method of deep belief network model based on optimization of data processing [J]. Journal of Computer Applications, 2017, 37 (6) : 1636-1643.)

[10] Qu F, Zhang J, Shao Z, et al. An Intrusion Detection Model Based on Deep Belief Network[C]// In Proceedings of the 2017 VI International Conference on Network, Communication and Computing, Kunming, China, 8-10 December, 2017: 97-101.

[11] Yin C, Zhu Y, Fei J, et al. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks[J]. IEEE Access, 2017, 5(99): 21954-21961.

[12] 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.

[13] Yuan Q, Lintao L V. Network intrusion detection method based on combination of improved ant colony optimization and genetic algorithm[J]. Journal of Chongqing University of Posts & Telecommunications, 2017, 29(1): 85-89.

[14] 魏明军, 王月月, 金建国. 一种改进免疫算法的入侵检测设计[J]. 西安电子科技大学学报, 2016, 43(2): 126-131.

[15] 贾凡, 孔令智. 基于卷积神经网络的入侵检测算法[J]. 北京理工大学学报, 2017, 37(12): 1271-1275.

[16] 王明, 李剑. 基于卷积神经网络的网络入侵检测系统[J]. 信息安全研究, 2017, 3(11): 990-994.

[17] Donghwoon K, Kathiravan N, Sang C S, et al. An empirical study on network anomaly detection using convolutional neural networks[C] //In 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Piscataway, New Jersey, USA: IEEE, 2018: 1595-1598.

[18] Vinayakumar R, Soman K P, Poornachandran P. Applying convolutional neural network for network intrusion detection[C] //In 2017 International Conference on Advanced Computing, Communications and Informatics, Piscataway, New Jersey, USA: IEEE, 2017: 1222-1228.

[19] 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.

[20] NSL-KDD dataset [EB/OL]. [2018-07-20] http://nsl.cs.unb.ca/NSL-KDD/.

[21] TensorFlow-GPU [EB/OL]. [2018-07-20]. https://www.tensorflow.org/

[22] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[23] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[24] Chung J, Gulcehre C, Cho K, et al. Gated Feedback Recurrent Neural Networks[C]// International Conference on Machine Learning. New York, USA, 2015: 2067-2075.

This work is partially supported by the Civil Aviation Joint Research Fund Project of National Natural Science Foundation of China (U1833107); the National Science and Technology Major Project (2012ZX03002002).

YANG Hongyu, born in 1969, Ph. D., professor. His research interests include network information security.

NING Fengyan, born in 1993, M. S. candidate. His research interests include network information security