研究与开发

基于深度学习和SVM-RFE的网络入侵检测模型

叶青[1]，张延年[2]，吴昊[2]
（1. 江苏省高港中等专业学校，江苏 泰州 225324；
2. 南京交通职业技术学院电子信息工程学院，江苏 南京 211188）

摘　要：网络入侵检测系统是对抗各种网络威胁的有效手段。然而，网络入侵数据中存在大量冗余信息和分布不平衡问题，为此，提出基于深度学习和支持向量机的递归特征消除算法的网络入侵检测（DLRF）模型。DLRF模型利用基于支持向量机的递归特征消除算法进行特征权重排序，选择重要特征。同时，结合过采样和欠采样技术解决数据样本分布不平衡的问题。利用3个深度学习算法构建集成框架的基学习器，并利用深度神经网络构建元学习器，进而提升DLRF模型检测网络攻击的性能。通过两个典型的网络入侵数据集UNSW-NB15和数据集CICIDS 2017验证DLRF模型的性能。性能分析表明，DLRF模型在这两个数据集上的准确率分别为0.906 8、0.996 8，F1值（F1-score）分别为0.906 8、0.996 0。

关键词：入侵检测模型；深度学习；递归特征消除；集成学习
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025100

Deep learning and support vector machine-recursive feature elimination-based network intrusion detection model

YE Qing[1], ZHANG Yannian[2], WU Hao[2]
1. Jiangsu Province Gaogang Secondary Vocational School, Taizhou 225324, China
2. Nanjing Vocational Institute of Transport Technology, School of Electronic Information Engineering, Nanjing 211188, China

Abstract: Network intrusion detection system has gained attention as an effective means of combating various cyber threats. However, there are a lot of redundant information and unbalanced distribution problems in network intrusion data, therefore, deep learning and support vector machine-recursive feature elimination-based network intrusion detection model (DLRF) was proposed. The features were sorted by the support vector machine-recursive feature elimination algorithm and the important features were selected. Moreover, both oversampling and under-sampling techniques were utilized to tackle the unbalance problem of data sample distribution. Three deep learning algorithms were used

收稿日期：2024−10−11；修回日期：2024−12−18
通信作者：吴昊，fuisf_89ui@163.com
基金项目：南京交通职业技术学院重大课题项目（No.JZ2304）；南京交通职业技术学院重点课题项目（No.JZ2306）
Foundation Items: The Nanjing Vocational Institute of Transport Technology Major Program (No.JZ2304), The Nanjing Vocational Institute of Transport Technology Key Program (No.JZ2306)

to build the base learner of the ensemble framework, and deep neural network was used to build the meta-learner, so as to improve the performance of DLRF model to detect network attacks. The proposed framework was experimented with two publicly available and popular network traffic datasets, namely UNSW-NB15 and CICIDS-2017. The accuracy rates of the DLRF model on these two datasets are 0.906 8 and 0.996 8 respectively, and the F1-score are 0.906 8 and 0.996 0.

Key words: intrusion detection model, deep learning, recursive feature elimination, ensemble learning

0　引言

近年来，网络基础设施在容量、系统负载和复杂性方面快速扩展。人们的生活越来越依赖智能家电、智能手机和云服务的应用。特别是在2020年新冠病毒疫情时，互联网几乎成为每个行业的生命线[1-2]。然而，网络基础设施的大面积部署以及网络用户数量的爆发式增长，对互联网的安全提出了严峻的挑战。

未经管理员同意，访问不同系统、窃取或修改重要数据、安装恶意软件或中断日常服务的任何攻击性操作均被视为网络攻击。网络攻击可分为目标攻击和非目标攻击两种。前者是指有针对性的攻击，并试图渗透到特定的业务中，而后者是指非针对性的攻击，并试图破坏尽可能多的系统或设备。目标攻击主要包括鱼叉式钓鱼和僵尸攻击。此外，恶意软件、拒绝服务（denial of service，DoS）、分布式拒绝服务（distributed denial of service，DDoS）、侦察、后门等是一些常见的目标攻击。

网络入侵检测系统（network intrusion detection system，NIDS）是增强网络安全[3]、防御网络攻击的有效手段。NIDS通过对网络流量进行分析，获取正常流量和网络攻击的特征，进而达到检测攻击的目的。依据Kwon等[4]的研究结果，执行NIDS的过程主要分3个步骤：（1）收集和跟踪网络流量数据；（2）清理原始数据，并对数据格式进行相应转换，使其满足下一个步骤的需求；（3）利用处理后的数据训练模型对样本数据进行鉴别。

设计一个有效的入侵检测模型是构建NIDS的基础。而基于异常的入侵检测是NIDS的重要分支。基于异常的入侵检测模型常采用机器学习（machine learning，ML）和深度学习（deep learning，DL）构建分类模型，应对网络攻击的多变性，增强模型的稳健性。在高维的大数据集场景下，通常基于DL的检测模型优于基于ML的检测模型[5]。

现多数研究采用数据集UNSW-NB15和数据集CICIDS 2017验证检测模型的性能。然而，多数检测模型属二分类。由于每类攻击遵循不同的攻击程序，研究多分类方案更有意义，并且研究多分类的检测模型更有助于设计预防技术[6]。此外，基于集成的检测模型的检测性能优于单个分类器的检测性能。

为此，提出基于深度学习和支持向量机的递归特征消除算法的网络入侵检测（deep learning and support vector machine-recursive feature elimination-based network intrusion detection，DLRF）模型。与现有的入侵检测模型不同：（1）DLRF模型同时采用过采样和欠采样技术处理数据分布不平衡问题，并利用基于支持向量机的递归特征消除算法对数据特征进行选择，用表征能力更强的特征训练模型，进而提升模型效率；（2）DLRF模型利用3个深度学习方法构成集成学习模型，并对传统的stacking模型进行改进，依据基学习器预测的均方误差设置基学习器的学习权重，提升集成学习模型的学习效率。

利用数据集UNSW-NB15和数据集CICIDS 2017验证DLRF模型的性能。性能分析表明，

DLRF 模型在数据集 UNSW-NB15 和数据集 CICIDS 2017 上的准确率和 F1 值（F1-score）分别为0.906、0.996和0.905、0.996。

1 基于stacking的网络入侵检测模型

1.1 模型框架

DLRF 模型的总体框架如图 1 所示，主要由数据预处理、特征选择、数据增强和模型训练 4 个阶段构成。在数据预处理阶段，对数据进行清洗，清除冗余数据，再对数据进行编码和归一化。在特征选择阶段，利用基于支持向量机的递归特征消除（support vector machine-recursive feature elimination，SVM-RFE）算法对特征权重排序，并选择重要的特征。

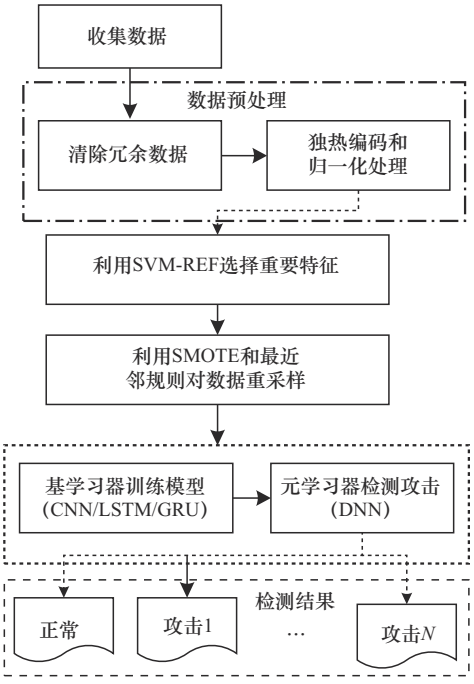


图 1 DLRF 模型的总体框架

考虑数据分布不平衡的问题，利用 SMOTE 进行过采样，增强少数类样本数量，并利用最近邻规则（edited nearest neighbor，ENN）进行欠采样，减少多数类样本数量，进而平衡各类样本的数量。

在模型训练阶段，利用数据训练基学习器和元学习器。将卷积神经网络（convolutional neural network，CNN）、门控递归单元（gated recurrent unit，GRU）和长短期记忆（long short-term memory，LSTM）网络作为基学习器，利用深度神经网络（deep neural network，DNN）作为元学习器。最后，通过元学习器对样本进行最终分类预测。

1.2 数据预处理

采用数据集 UNSW-NB15 和数据集 CICIDS 2017 分析 DLRF 模型的性能。数据集 UNSW-NB15 有 82 332 条测试样本和 175 341 条训练样本，每个样本有 49 个特征。数据集 UNSW-NB15 包含正常和 6 类攻击，共 7 类样本。这 6 类攻击分别是 Worms、Generic、Fuzzers、Reconnaissance（Reconna）、Shellcode 和 DoS 类攻击。

先对数据集 UNSW-NB15 进行预处理，删除一些冗余特征以及对特征进行独热编码。处理过程如下：（1）由于“id”列无用，将其从数据集 UNSW-NB15 中删除；（2）将“stime”和“ltime”特征从数据集 UNSW-NB15 中删除；（3）将与交换机相关的特征从数据集 UNSW-NB15 中删除，这些特征是“sport”“scrip”“dstip”和“dsport”；（4）分类特征，如“proto”和“service”具有广泛的取值范围，为此，利用标签编码器对这些特征进行编码，为每类样本生成唯一的新数值；（5）删除一些缺失值的特征，如“ct ftp cmd”“is ftp login”“ct flw http method”。

数据集 CICIDS 2017 是入侵检测的专用数据集，包含多个良性和最新的常见网络攻击，也广泛使用分析车载网络的入侵检测算法的性能。考虑原数据集 CICIDS 2017 中的样本数据很大，从中随机抽取了 56 661 个样本数据进行训练和测试，包含良性（Benign）、拒绝服务（DoS）、端口扫描（Portscan）、暴力（Bruteforce）、Web（Webattack）、Bot 和 DDoS 攻击 7 类样本。

与处理数据集UNSW-NB15一样，也对数据集CICIDS 2017进行类似的处理，删除冗余特征值，并进行编码，为每类样本生成唯一的编码，将符号型数据转换成二进制数值型数据。

接下来，对编码后数据进行归一化处理，将数据映射至[0,1]区间。为此，DLRF模型采用最大-最小值方法对样本数据进行归一化处理，如式（1）所示：

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

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，x_n表示对x归一化后的数据。

1.3　基于SVM-RFE的特征选择

特征选择是指从海量的特征集中选择最相关特征，并剔除不相关特征，以提高模型性能和降低预测模型计算成本的过程。此外，通过剔除噪声特征，保留重要特征，进而提高训练模型效率。

DLRF模型采用基于支持向量机的递归消除算法选择特征。基于支持向量机的递归特征消除（support vector machine-recursive feature elimination，SVM-RFE）算法是结合支持向量机（SVM）和递归特征消除（RFE）算法对特征进行排序，消除冗余特征，保留重要特征的方法[7]。

SVM通过寻找一个最优分类超平面，使样本离超平面的距离尽可能大，基于SVM分类的示意图如图2所示，两类样本的特征分别是特征1和特征2。图2中的实线表示最优分类超平面。

令$\{(x_i, y_i), i = 1, 2, \cdots, n\}$表示样本集，其具有$n$个样本，其中$x_i$、$y_i$分别表示第$i$个样本的输入值和输出值（标签）。SVM在分类时先建立以x为自变量的回归估计函数$f(x)$：

$$|y - f(x)| = \begin{cases} 0, y - f(x) \leqslant \varepsilon \\ y - f(x) - \varepsilon, y - f(x) > \varepsilon \end{cases} \tag{2}$$

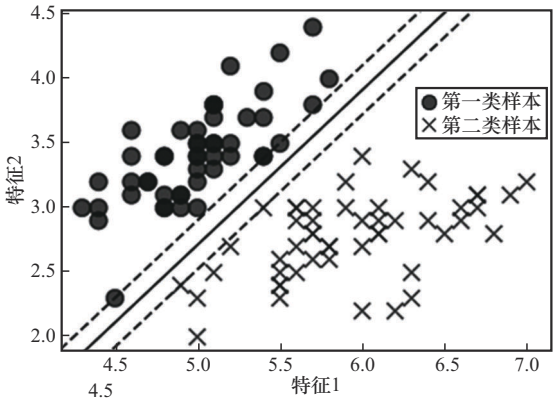


图2　基于SVM分类的示意图

其中，y泛指输出值的估计值，ε表示不敏感损失函数。

依式（2）可知，当样本落入以$f(x)$为中心且距离为2ε的区域时，回归损失为0。反之，若样本落入其他区域，则进行惩罚。SVM在分类样本时就是通过构建合适的函数$f(x)$，在ε尽可能小时，降低对样本惩罚的次数。据此，可建立优化问题：

$$\begin{cases} \min\left\{ L_P = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i \right\} \\ \text{s.t. } y_i\left(\boldsymbol{w}^{\mathrm{T}} x_i + b\right) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0, i = 1, 2, \cdots, n \end{cases} \tag{3}$$

其中，$\boldsymbol{w} = \left[w_1, w_2, \cdots, w_p\right]$为特征向量，$C$表示惩罚因子，$\xi_i$为松驰变量，$b$为函数$f(x)$的偏置。

SVM-RFE算法先利用SVM评估特征对分类效果的贡献，即计算各特征的权重。然后，剔除小权重的特征，保留重要特征。利用SVM-RFE选择特征的基本思路[8]如下，依据目标函数L_p，计算剔除第i个特征对L_p的影响：

$$\Delta L_p(i) = \frac{\partial L_p}{\partial w_i}\Delta w_i + \frac{\partial^2 L_p}{\partial w_i^2}(\Delta w_i)^2 \tag{4}$$

再得到每个特征的权重w_i值，最后依据权重w_i对特征进行排序：

$$\text{Rank}(i) = (w_i)^2 \tag{5}$$

1.4 数据重采样

不平衡数据是分类领域的共性问题。所谓不平衡数据是指数据中各类样本数并不是均匀分布的，这意味着数据集偏向一个或多个类，其他类的样本数较少。因此，若训练模型的数据分布不平衡，则所训练模型可能会偏向多类样本，不利于对少类样本的检测。据此，需对数据进行重采样，以解决数据分布不平衡的问题。

为此，本文采用合成少数类过采样技术（synthetic minority oversampling technique，SMOTE）增强少数类的样本数据。SMOTE依据少数类样本的邻居样本信息，为其增加一定的样本数，进而增强数据[9]。以样本x_i为例，计算离邻近样本的欧氏距离；再利用k近邻法，搜索出离样本x_i距离最近的k个少数类样本；然后，再从这k近邻样本中随机选择一个$\hat{x}_i$，并依式（6）为样本x_i生成一个新的样本x_{new}：

$$x_{\text{new}}=x_i+\left(\hat{x}_i-x_i\right)\times\delta \tag{6}$$

其中，$\delta\in[0,1]$中一个随机数。

对于多数类样本，利用最近邻规则（ENN）削弱样本数量。它的核心思想是剔除与其k近邻的大多数类别不同的样本，进而减少多数类的样本数量。以样本x_i为例，如果它的k近邻点有超过一半的样本都不属于多数类，就删除此样本x_i。

1.5 基于stacking集成的检测模型

基于stacking集成的检测模型如图3所示，由两层组成。第一层由CNN、GRU和LSTM 3个深度学习的分类器构成，它们称为基学习器。第二层的元学习器由DNN构成。将3个基学习器的预测结果按纵向拼接作为元学习器的输入。

选择CNN、RGU和LSTM作为基学习器的原因如下。

（1）尽管CNN广泛应用于图像处理领域，但它也被广泛应用于网络入侵检测系统[10]以及物联网数据分析中。CNN可从网络流量中捕获空间相关性，能从高维输入数据中自动提取局部相关性，进而选择更优的特征。而池化层在不丢失重要信息的基础上，通过局部相关性压缩特征图的维度，降低了过拟合的概率。

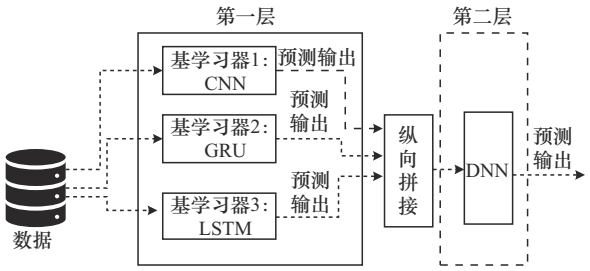


图3 基于stacking集成的检测模型

（2）LSTM和GRU都是RNN的高级变体，能够对长期时间依赖关系进行建模。它们在时间性数据领域得到广泛应用。LSTM的体系结构比GRU更复杂，并且在计算上比GRU更昂贵。然而，GRU占用的计算资源更少，训练时间更快[11-12]。由于不同的攻击类型网络入侵呈现不同模式，并且这些模式可以随着时间的推移分散在多个数据包中，为了有效地识别这些模式，DLRF模型同时使用了LSTM和GRU，增强了从时间特征中检索信息的能力，提升了集成模型的稳健性和整体性能。

1.6 改进的stacking集成模型

从第1.5节可知，传统的stacking集成模型是将第一层的3个基学习器的预测结果直接按纵向拼接，拼接结果作为元学习器的输入，这忽略了3个基学习器的学习性能不同的客观事实。由于基学习器的学习性能并不相同，它们的输出结果可能存在非常大的差异。为此，对传统的stacking集成方法进行改进，即依据基学习器的预测性能设置权重[13]。

为此，先计算基学习器预测性能的均方根误差（root mean square error，RMSE）。令E_ℓ表示

第 ℓ 个基学习器的 RMSE，其定义如式（7）所示：

$$E_\ell = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^\ell - \hat{y}_i\right)^2}, \ell = 1,2,3 \quad (7)$$

其中，$\hat{y}_i$表示样本x_i的真实标签值，y_i^ℓ表示由第ℓ个基学习器的预测值，n为样本数。

然后，再依据E_ℓ计算基学习器的权重λ_ℓ：

$$\lambda_\ell = \frac{E^*}{E_\ell} = \frac{1}{E_\ell}\min_{\ell=1,2,3}\{E_\ell\} \quad (8)$$

其中，$E^* = \min_{\ell=1,2,3}\{E_\ell\}$表示 3 个基学习器最小 RMSE 值。

最后，将基学习器的输出与它的权重相乘，相乘结果作为元学习器的输入。此外，为了防止过拟合，采用 5 折交叉验证。

基学习器的基于 5 折交叉验证如图 4 所示，给出了训练基学习器和元学习器的过程，主要步骤如下。

步骤 1　将经过数据预处理后的数据进行拆分，拆分成训练集 Data_t 和测试集 Data_s；并将训练集 Data_t 拆分成 5 等份，分别标记为 D_1、D_2、D_3、D_4、D_5。

步骤 2　基学习器训练模型，并进行预测。基学习器先从 D_1、D_2、D_3、D_4、D_5 选择一份数据作为测试，其他作为训练集。训练完后，就对测试数据进行预测，并记录预测结果。由于采用了 5 折交叉验证，此过程重复 5 次，可得 5 个预测值。同时，将训练后的模型对测试数据 Data_s 进行预测，可得到 5 个预测结果，并计算它们的均值。最终，将 5 个预测值和 1 个均值进行纵向拼接，形成基学习器的输出，如图 4 所示。令 BP_ℓ 表示基学习器 ℓ 的输出值，将 BP_ℓ 与基学习器的预测效果（RMSE）进行加权，得到加权后的输出值 $B\widehat{P}_\ell$。

步骤 3　将 $B\widehat{P}_\ell$ 值作为新特征加入数据集中，构成新的训练集。由于 DLRF 模型采用了 3 个基学习器，可得到 $B\widehat{P}_1$、$B\widehat{P}_2$、$B\widehat{P}_3$。将 $B\widehat{P}_1$、$B\widehat{P}_2$、$B\widehat{P}_3$ 加入数据集，进而得到新的训练集 $\text{Data}_\text{t}^\text{new}$。

步骤 4　用 $\text{Data}_\text{t}^\text{new}$ 训练元学习器。训练结束后，就对 Data_s 进行预测，并评估预测性能。

2　性能分析

利用数据集 UNSW-NB15 和数据集 CICIDS 2017 验证 DLRF 模型的性能。采用 Scikit-learn 库训练模型。同时，将数据集划分为训练集、测试

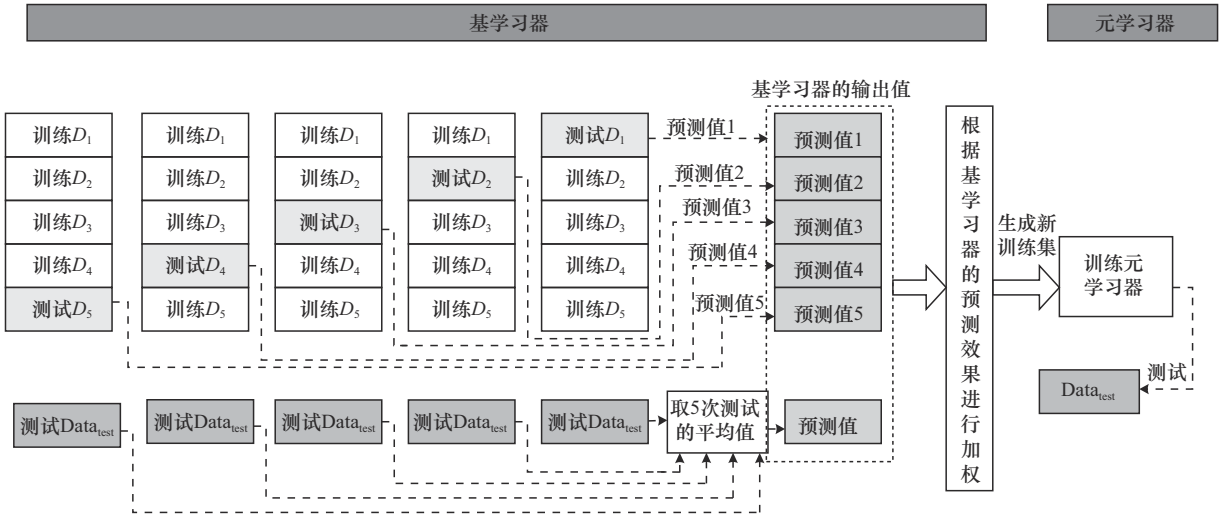


图 4　基学习器的基于 5 折交叉验证

集和验证集。数据集的70%用于训练模型，10%用于验证模型，而余下的20%用于测试DLRF模型的性能。

2.1 实验环境及相关参数

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，Window 11。

CNN模型由输入层、两个卷积层、一个最大池化层和全连接层组成。选用ReLU激活函数。LSTM模型由一个致密（Dense）层和一个丢弃（Dropout）层组成。每层也采用ReLU激活函数，损失函数为分类交叉熵，并使用学习率为0.001的Adam优化器。GRU是另一种短期记忆解决方案，其基本原理与LSTM基本相同。GRU使用隐藏状态，由复位门和更新门组成，采用Tanh激活函数。

2.2 性能指标

选用准确率（Accuracy）、精确率（Precision）和F1值（F1-score）作为评价指标。将正常流量（Normal）和良性（Benign）作为正例，其他攻击全部视为反例。TP表示真实值为正例，正确预测为正例的样本数目；TN表示真实值为反例，正确预测为反例的样本数目；FP表示真实值为反例，错误预测为正例的样本数目；FN表示真实值为正例，错误预测为反例的样本数目[14]。

准确率（Accuracy）表示分类正确的样本数量与总样本数量的比值：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{9}$$

与准确率不同，精确率（Precision）表示预测为正例的样本中，真实值为正例的比例：

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

召回率（Recall）专用于评估检测模型对正例的分类能力，其等于正确预测为正例的样本数占总正例样本数的比值：

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

考虑指标Precision和Recall存在一定的片面性，同时也选用F1值指标。F1值对Precision和Recall进行了融合，其定义如式（12）所示。F1值越大，模型分类的综合性能越好。

$$\text{F1-score}=2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

2.3 基于数据集UNSW-NB15的DLRF模型的检测性能

本节利用数据集UNSW-NB15验证DLRF性能，并将3个基学习器作为3个基准模型。同时，选择文献[15]提出的基于集成特征选择的入侵检测模型（ensemble feature selection-based IDS，ENFS）、文献[16]提出的基于深度学习的入侵检测模型（deep learning-based IDS，DIDS）和文献[17]提出的基于非线性增广显式特征的入侵检测模型（detection model based on nonlinear augmented explicit features，NAEF）作为基准模型。

这3个基准模型均利用数据集UNSW-NB15验证它们的性能。DLRF模型在数据集UNSW-NB15上的预测性能见表1。

表1 DLRF模型在数据集UNSW-NB15上的预测性能

模型		准确率	精确率	召回率	F1值
基准模型	ENFS	0.797 0	0.967 0	0.773 0	0.859 0
	DIDS	0.853 8	0.905 3	0.896 3	0.898 0
	NAEF	0.857 0	0.902 0	0.873 0	0.883 0
基学习器	CNN	0.891 0	0.897 0	0.892 0	0.891 0
	LSTM	0.890 0	0.895 0	0.890 0	0.890 0
	GRU	0.884 0	0.894 0	0.884 0	0.882 0
所提模型	DLRF	0.906 8	0.909 3	0.906 7	0.906 8

从表 1 可知，DLRF 模型的预测性能最优，准确率达到 0.906 8，F1 值达到 0.906 8。DLRF 模型性能优于单个深度模型，包括 CNN、LSTM 和 GRU。这 3 个基学习器的准确率在 0.884 0~0.891 0 区间，F1 值在 0.882 0~0.891 0 区间。

此外，ENFS 模型、DIDS 模型和 NAEF 模型的准确率和 F1 值均未达到 0.9，它们的预测性能劣于 DLRF 模型。原因在于：DLRF 模型将多个深度模型组合到一个集成框架中，这些深度模型从多方面捕获网络流量的特征，进而使 DLRF 模型能够做出更准确的分类结果。此外，DLRF 模型在第二层使用了元学习器来学习如何组合基本模型的权重来做出最终预测，这提高了模型精度。

与此同时，DLRF 模型通过 SMOTE 和 ENN 解决数据集分布不平衡问题，这有助于提高分类的准确率。相反，ENFS 模型的精确率达到 0.967 0，远高于 DLRF 模型。然而，ENFS 模型没有考虑数据分布不平衡问题，它只限于检测某一类攻击。总体上，DLRF 模型通过将多个深度模型组合到一个集成框架以及采用数据增强技术解决数据分布不平衡问题，提升模型的分类性能。

DLRF 模型在数据集 UNSW-NB15 上的混淆矩阵如图 5 所示，显示了 DLRF 模型在分类 7 类样本的检测率。从图 5 可知，DLRF 模型可准确地检测这些攻击。检测 Worms、Reconnaissance（Reconna）和 Generic 3 类攻击的准确率超过 0.95。而检测 Shellcode 和 Normal 的准确率也低于 0.90。此外，DLRF 模型能准确检测 89% 的 DoS 攻击，但只能准确检测 53% 的 Fuzzers 攻击，将 32% 的 Fuzzers 攻击误判为 Normal。原因在于：Fuzzers 攻击与 Normal 流量非常相似，很难区分它们。

总体上，DLRF 模型在数据集 UNSW-NB15 上具有良好的性能，只是在识别 Fuzzers 攻击方面存在困难。

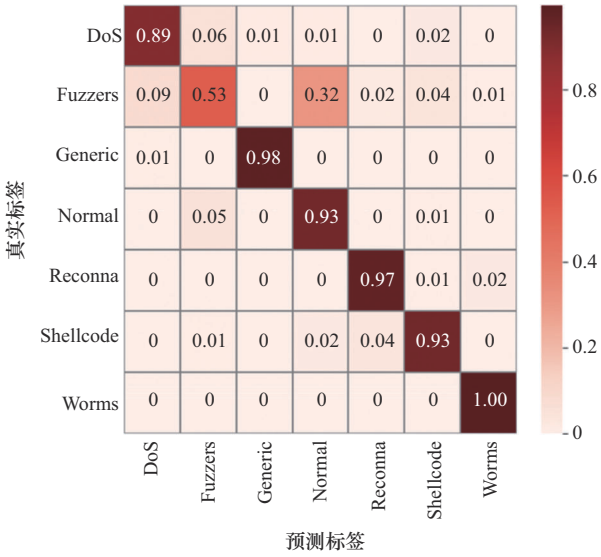


图 5　DLRF 模型在数据集 UNSW-NB15 上的混淆矩阵

2.4　基于数据集 CICIDS 2017 的 DLRF 性能分析

DLRF 模型在数据集 CICIDS2017 上的预测性能见表 2。

表 2　DLRF 模型在数据集 CICIDS 2017 上的预测性能

模型		准确率	精确率	召回率	F1 值
基准模型	ENFS	0.989 0	0.999 0	0.988 1	0.993 0
	RNGU	0.989 9	0.984 2	0.986 3	0.986 0
	SABiLSTM	0.999 2	0.991 5	0.990 9	0.991 2
基学习器	CNN	0.975 0	0.977 0	0.976 0	0.973 0
	LSTM	0.972 0	0.974 0	0.973 0	0.971 0
	GRU	0.960 1	0.964 0	0.961 0	0.960 0
所提模型	DLRF	0.996 8	0.995 1	0.997 0	0.996 0

表 2 中的 RNGU 模型和 SABiLSTM 模型分别是文献[18]提出的基于递归神经网络和 GRU 的入侵检测模型（RNGU）和文献[19]提出的基于自注意力机制的双向长短期记忆网络的检测模型（SABiLSTM）。RNGU 模型和 SABiLSTM 模型均利用神经网络构建入侵检测模型，并利用数据集 CICIDS 2017 验证模型性能。为此，选择它们作为基准模型。

从表 2 可知，DLRF 模型的检测性能优于其

他模型，它的准确率、精确率、召回率和F1值的性能分别达到0.996 8、0.995 1、0.997 0和0.996 0。而ENFS模型只能获取0.989 0的准确率，并且ENFS模型未考虑数据的不平衡分布问题。SABi-LSTM模型的检测性能优于ENFS和RNGU模型，原因在于：SABiLSTM模型采用自注意力机制。自注意力机制具有关注重要特征的能力，这有利于提升模型的检测性能。

此外，DLRF模型的预测性能优于RNGU模型。RNGU模型的准确率为0.989 9，而DLRF模型的准确率达到0.996 8。DLRF模型具有高的召回率，达到0.997，这意味着该模型具有较低的漏检率和较高的检测率。

DLRF模型在数据集CICIDS 2017上的混淆矩阵如图6所示。从图6可知，除了对Benign类型的检测率未达到100%，只有99%，对其他6类攻击的检测率均达到100%，这直观地说明：DLRF模型可有效地检测这些攻击。

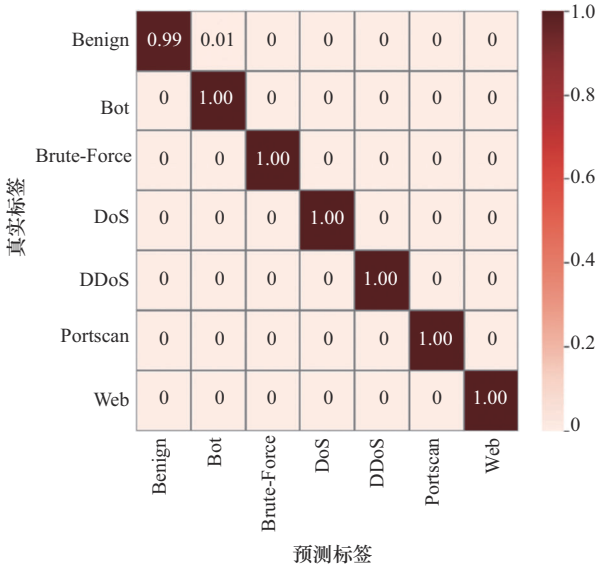


图6　DLRF模型在数据集CICIDS 2017上的混淆矩阵

2.5　DLRF模型的复杂度

模型复杂度是衡量算法性能和效率的重要指标。设计一个低复杂度的入侵检测模型有助于提高性能，并减少所需计算资源。为此，本节分析DLRF模型的复杂度。

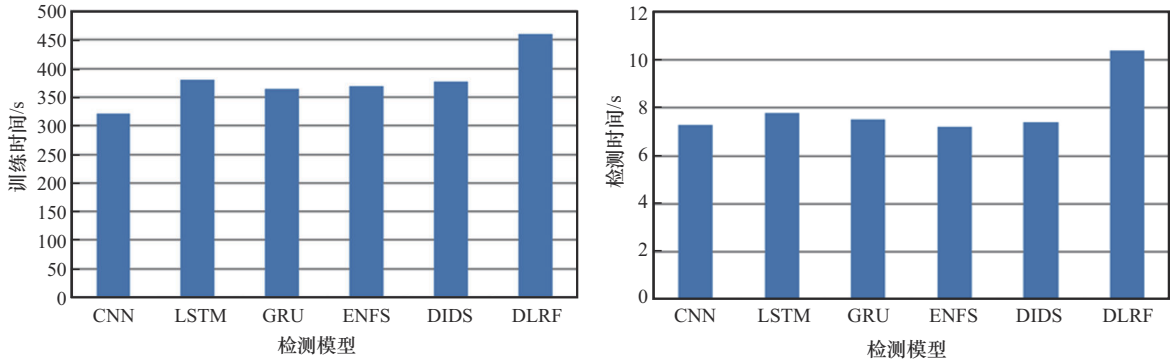
先从理论角度分析DLRF模型的复杂度。DLRF模型主要涉及重采样、基学习器学习样本和预测测试样本3个运算。而重采样和特征提取属于数据预处理阶段。假定训练数据集中有N个样本。若T_{base}表示一个基学习器学习一个样本的时间，则一个基学习器学习N个样本的时间复杂度为$O(N\times T_{\text{base}})$。若$T_{\text{meta}}$表示一个元学习器学习一个样本的时间，则它学习$N$个样本的时间复杂度为$O(N\times T_{\text{meta}})$。假定预测数据集中有$M$个样本，则预测阶段的时间复杂度为$O(M\times T_{\text{test}})$。因此，DLRF模型的总时间复杂度为$O(N\times T_{\text{base}}+N\times T_{\text{meta}}+M\times T_{\text{test}})$。

接下来，通过记录训练时间和检测时间分析DLRF模型的复杂度。

各模型在数据集UNSW-NB15上的检测时间和训练时间如图7所示，单位为s。从图7可知，DLRF模型的训练时间和检测时间最长。其中DLRF模型的训练时间为460 s，检测时间约10 s。原因在于：DLRF模型采用了集成框架，需要训练多个深度学习模型，这增加了训练时间。

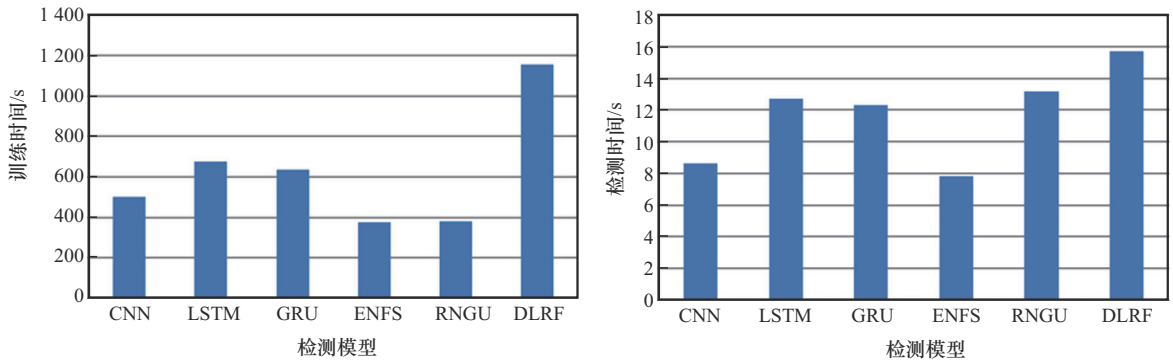
各模型在数据集CICIDS 2017上的检测时间和训练时间如图8所示。相比于基学习器，DLRF模型增加了额外的训练时间。原因在于：DLRF模型不仅需要训练第一层的基学习器，还需训练第二层的元学习器。从表2可知，这种额外的训练时间提升了模型的预测性能和泛化性能。

由于测试数据必须通过整个集成模型，所提DLRF模型比基学习器需要更多时间。预先训练的集成模型先通过基学习器处理测试数据，然后将其传递给元学习器，并由后者做出最终预测。这些过程需要更多计算来聚合不同模型的预测，从而提高DLRF模型的准确性和性能，这必然增加了测试时间。然而，与单个模型相



(a) 训练时间　　(b) 检测时间

图 7　各模型在数据集 UNSW-NB15 上的训练时间和检测时间



(a) 训练时间　　(b) 检测时间

图 8　各模型在数据集 CICIDS 2017 上的训练时间和检测时间

比，集成模型可以更有效地处理不同类型的网络流量。

总体而言，尽管集成方法需要比基学习器更多的检测时间和训练时间，但它最终会提高性能，使其成为一种很有前途的网络入侵检测方法。

此外，DLRF 模型采用 LSTM 和 GRU 来检测广泛的时间特征。上述数据表明：LSTM 在这两个数据集上的检测精度都优于 GRU。原因在于：LSTM 可以更好地处理更大的数据集，而本文使用的数据集也非常大。LSTM 采用了 3 个门（包括一个输入门、一个输出门、一个遗忘门）和一个额外的存储单元，使其能够捕获输入数据中更长期的依赖关系。另一方面，GRU 的体系结构更简单，只有两个门，分别为一个更新门和一个重置门，这使得它的效率较低。尽管 LSTM 具有更高的检测精度，但 GRU 比 LSTM 具有更快的训练时间和更高的计算效率。原因在于：GRU 采用了更简单的架构。简单的架构参数少，易计算，降低了训练时间。

总之，DLRF 模型将多个深度学习模型组合到一个集成框架中，以捕获网络流量的不同方面并做出明智的决策。元学习器将基学习模型的预测结合在一起以提高精度。DLRF 模型采用了数据增强技术或类别加权解决了不平衡分类问题，在数据集 UNSW-NB15 和数据集 CICIDS 2017 上的表现优于单个模型和最先进的模型。然而，它在识别 Fuzzers 攻击方面存在困难，需要进一步研究。尽管训练时间较长，但对于数据集 UNSW-NB15 和数据集 CICIDS 2017，DLRF 模型分别可在 10.4 s 和 15.7 s 内检测出攻击。

3 结束语

在5G等超高速网络技术进步的推动下，接入互联网的设备数量快速增长。黑客/入侵者继续使用新技术发动大规模网络攻击，使网络流量更容易受到攻击。本文提出了一种基于人工智能的集成模型来识别各种类型的网络攻击。与以往研究不同，本文提出的DLRF模型不仅能判断网络流量是良性的还是不正常的，还能对流量中的攻击类型进行分类。

由于数据集UNSW-NB15和CICIDS 2017非常接近真实的网络流量，利用它们验证DLRF模型的性能，这有利于提高DLRF模型的效率、稳健性和实用性。采用SVM-RFE选择可能的最佳流特征，并采用SMOTE和ENN技术对不平衡类进行重采样和欠采样。性能分析表明，DLRF模型在数据集UNSW-NB15和CICIDS 2017中分别获得0.906 8和0.996 8的准确率。

DLRF模型的主要缺陷是受训练阶段使用的标签数据的限制，不具有零日攻击能力。后期，将改进模型，增强它的防御零日攻击能力。

参考文献：

[1] 谢添丞, 吴凌淳, 林羽丰, 等. 面向大数据的网络流量监控与分析算法综述[J]. 无线电通信技术, 2022, 48(5): 782-793.
XIE T C, WU L C, LIN Y F, et al. Survey of network traffic monitoring and analysis for big data[J]. Radio Communications Technology, 2022, 48(5): 782-793.

[2] 张涛, 费稼轩, 王琦, 等. 电力信息物理系统跨域攻击协同防御架构及机制研究[J]. 电子学报, 2024, 52(4): 1205-1218.
ZHANG T, FEI J X, WANG Q, et al. Research of architecture and mechanism of coordinative defense methods for cross-domain attacks of cyber physical system[J]. Acta Electronica Sinica, 2024, 52(4): 1205-1218.

[3] 董卫魏, 王曦, 钟昕辉, 等. 基于人工智能技术的轻量级网络入侵检测系统设计[J]. 现代电子技术, 2024, 47(5): 108-111.
DONG W W, WANG X, ZHONG X H, et al. Design of light-weight network intrusion detection system based on artificial intelligence technology[J]. Modern Electronics Technique, 2024, 47(5): 108-111.

[4] KWON D, KIM H, KIM J, et al. A survey of deep learning-based network anomaly detection[J]. Cluster Computing, 2019, 22(1): 949-961.

[5] AHMAD J, FARMAN H, JAN Z. Deep learning methods and applications[J]. Springer Briefs in Computer Science, 2018: 31-42.

[6] HU R, WU Z Y, XU Y, et al. Multi-attack and multi-classification intrusion detection for vehicle-mounted networks based on mosaic-coded convolutional neural network[J]. Scientific Reports, 2022, 12(1): 6295.

[7] 李安琪, 杨琳, 蔡言颜, 等. 基于递归特征消除-随机森林模型的江浙沪农田土壤肥力属性制图[J]. 地理科学, 2024, 44(1): 168-178.
LI A Q, YANG L, CAI Y Y, et al. Digital mapping of soil fertility attributes in croplands in Jiangsu, Zhejiang and Shanghai based on recursive feature elimination-random forest model[J]. Scientia Geographica Sinica, 2024, 44(1): 168-178.

[8] 罗红郊, 马晓琴, 孙妍, 等. 基于RFE特征选择的PSO-SVM用电量预测算法[J]. 电子设计工程, 2023, 31(20): 172-176.
LUO H J, MA X Q, SUN Y, et al. PSO-SVM electricity consumption prediction algorithm based on RFE feature selection[J]. Electronic Design Engineering, 2023, 31(20): 172-176.

[9] 李淑琪, 光彪, 赵玉凤, 等. SMOTE__ENN结合AdaBoost在临床预测模型中的应用探析[J]. 中国卫生统计, 2023, 40(6): 817-821.
LI S Q, GUANG B, ZHAO Y F, et al. Application of SMOTE__ENN combined with AdaBoost in clinical prediction model[J]. Chinese Journal of Health Statistics, 2023, 40(6): 817-821.

[10] 高鑫泽, 吕国, 杨宵, 等. 基于CNN-LSTM模型的入侵检测算法研究[J]. 河北建筑工程学院学报, 2024, 42(3): 216-221.
GAO X Z, LYU G, YANG X, et al. Research on intrusion detection algorithm based on CNN-LSTM model[J]. Journal of Hebei Institute of Architecture and Civil Engineering, 2024, 42(3): 216-221.

[11] YANG S D, YU X Y, ZHOU Y. LSTM and GRU neural network performance comparison study: taking Yelp review dataset as an example[C]//Proceedings of the 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). Piscataway: IEEE Press, 2020: 98-101.

[12] BELLA H K , VASUNDRA S. A novel framework based on extra tree regression classifier and grid search LSTM for intrusion detection in IoT and cloud environment[J]. International Journal of Intelligent Engineering and Systems, 2024, 17(4): 504-517.

[13] 孙玉芹, 王敏, 田方, 等. 基于熵权法Stacking集成学习的多

分类窃电检测[J]. 科学技术与工程, 2024, 24(30): 12996-13004.
SUN Y Q, WANG M, TIAN F, et al. Stacking integrated learning of multi-classification electricity theft detection based on entropy weight method[J]. Science Technology and Engineering, 2024, 24(30): 12996-13004.

[14] 梅元清, 郭肇强, 周慧聪, 等. 面向对象软件度量阈值的确定方法: 问题、进展与挑战[J]. 软件学报, 2023, 34(1): 50-102.
MEI Y Q, GUO Z Q, ZHOU H C, et al. Deriving object-oriented metric thresholds: research problems, progress, and challenges[J]. Journal of Software, 2023, 34(1): 50-102.

[15] DAS S, SAHA S, PRIYOTI A T, et al. Network intrusion detection and comparative analysis using ensemble machine learning and feature selection[J]. IEEE Transactions on Network and Service Management, 2022, 19(4): 4821-4833.

[16] ALEESA A M, YOUNIS M, MOHAMMED A A, et al. Deep-intrusion detection system with enhanced UNSW-NB15 dataset based on deep learning techniques[J]. Journal of Engineering Science and Technology, 2021, 16(1): 711-727.

[17] YU X, LU Y, JIANG F, et al. A cross-domain intrusion detection method based on nonlinear augmented explicit features[J]. IEEE Transactions on Network and Service Management, 2024(99): 1.

[18] HENRY A, GAUTAM S. Intrusion detection using big data and deep learning techniques[C]//Proceedings of the Computing, Communication and Learning: First International Conference. Waranga: Springer, 2023:220-230.

[19] ATTIQUE D, HAO W, PING W, et al. Explainable and data-efficient deep learning for enhanced attack detection in IIoT ecosystem[J]. IEEE Internet of Things Journal, 2024, 11(24): 38976-38986.

[作者简介]

叶青（1973– ），男，江苏省高港中等专业学校讲师，主要研究方向为人工智能、物联网应用。

张延年（1977– ），男，南京交通职业技术学院电子信息工程学院副教授，主要研究方向为大数据、人工智能。

吴昊（1973– ），男，南京交通职业技术学院电子信息工程学院教授，主要研究方向为智能交通、计算机网络。