第 38 卷　第 5 期
2023 年 10 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 38　No. 5
Oct. 2023

DOI:10.13364/j.issn.1672-6510.20220259

面向不平衡数据和特征冗余的网络入侵检测

张翼英，王德龙，渠慧颖，张　傲，张　磊
（天津科技大学人工智能学院，天津 300457）

摘　要：为了解决传统方法因数据不平衡及特征冗余而导致检测准确率不高的问题，提出了一种结合 SMOTE（synthetic minority over-sampling technique）算法采样的 SDAE-LSTM（stacked deep auto-encoder-long short term memory）入侵检测模型。首先，针对数据不平衡问题，采用 SMOTE 算法在少数类样本点之间随机插入样本增加其数量，达到类间平衡的目的。其次，针对特征冗余问题，利用堆叠式深度自编码器（stacked deep auto-encoder，SDAE）进行降维，实现数据的深度特征提取。最后，基于长短期记忆（long short term memory，LSTM）神经网络，精准捕获网络入侵特征，准确地实现入侵检测。通过在 UNSW-NB15 数据集上的大量实验，有效证明了本文模型与其他模型相比有着更好的入侵检测效果。

关键词：不平衡数据；特征冗余；SMOTE；堆叠式深度自编码器；长短期记忆神经网络；网络入侵检测
中图分类号：TP393.08　　**文献标志码**：A　　**文章编号**：1672-6510（2023）05-0057-07

Network Intrusion Detection for Unbalanced Data and Feature Redundancy

ZHANG Yiying，WANG Delong，QU Huiying，ZHANG Ao，ZHANG Lei
（College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China）

Abstract：In order to solve the problem of low detection accuracy caused by data imbalance and feature redundancy in traditional methods，a stacked deep auto-encoder-long short term memory（SDAE-LSTM）intrusion detection model combined with synthetic minority over-sampling technique（SMOTE）sampling is proposed in our current study. Firstly，aiming at the problem of data imbalance，a SMOTE method was used to randomly insert samples between a few sample points to increase their number，so as to achieve the goal of category balance. Secondly，aiming at the problem of feature redundancy，the stacked deep auto-encoder（SDAE）was used to reduce the dimension and realize the depth feature extraction of data. Finally，based on the long short term memory（LSTM）neural network，the network intrusion characteristics was accurately captured and the intrusion detection was accurately implemented. Through a large number of experiments on UNSW-NB15 datasets，our proposed model is effectively proved to have better intrusion detection effect than other models.

Key words：unbalanced data；feature redundancy；SMOTE；SDAE；LSTM neural network；network intrusion detection

入侵检测系统[1]能够及时发现已知的网络攻击，是一种积极主动的网络安全防御技术，在网络安全领域备受关注，是该领域研究的热点之一[2]。入侵检测的关键是对网络流量进行正常与异常的分类[3]，并根据分类的结果采取相应的应对措施以保障网络的安全运行，减少受害者的财产损失。

迄今为止，国内外学者和研究人员已经提出了大量的入侵检测方法对网络攻击进行检测，其中有两个问题成为研究热点，一是如何解决数据不平衡的问题以提高入侵检测的准确率，二是如何进行特征提取才能更好地获取数据的特征，从而提高对网络攻击识别的准确率。针对数据不平衡问题，罗文华等[4]首先利

收稿日期：2022–11–16；**修回日期**：2023–02–18
基金项目：国家自然科学基金项目（61807024）
作者简介：张翼英（1973—），男，辽宁人，教授；**通信作者**：王德龙，硕士研究生，delongwang@mail.tust.edu.cn

用 KNN(K-nearest neighbor)算法选出了与多数类样本距离最近的少数类样本，之后使用 DBSCAN (density-based spatial clustering of applications with noise)算法对选出的样本进行聚类，形成新的数据集，提高了模型的召回率。汪祖民等[5]提出了一种基于 DBSCAN_GAN_XGBoost 的入侵检测模型，该方法对数量较少的攻击样本进行了扩充，并利用 DBSCAN 算法对扩充后的样本数据进行聚类，极大提高了稀有攻击类别的检测准确率，但对多数量的攻击类型的检测准确率较低。章缙等[6]在传统随机森林的基础上加入了上采样和加权投票等优化手段，大大提高了模型的检测能力，但优化后的随机森林(optimized random forest，ORF)依然存在模型训练时间较长及检测准确率不高的问题。Hamid 等[7]提出了一种基于人工神经网络(artificial neural network，ANN)和小波变换(wavelet transform，WTF)的方法解决数据不完备的问题，提高了少量攻击类别的检测准确率。

针对特征提取，葛继科等[8]利用多层卷积神经网络对数据集进行特征选择，去除了数据的冗余特征，之后使用 sigmoid 激活函数对攻击行为进行二分类预测，减少了检测时间，提高了准确率。夏栋梁等[9]利用蚁群算法(ant colony optimization，ACO)对数据进行降维，提高了分类精度，减小了误报率，但降维所用时间较长。刘辉[10]提出了一种基于主成分分析(principal component analysis，PCA)和多层感知机(multi-layer perceptron，MLP)的入侵检测模型，该模型首先采用 PCA 对数据进行降维，然后将降维后的数据送入 MLP 神经网络分类器，对网络入侵行为进行分类，提高了入侵检测系统的准确率，但 PCA 只能进行线性降维，对非线性特征描述不够全面，以至于误报率较高。赵荷等[11]凭借深度信念网络(deep belief network，DBN)对入侵特征进行提取，并利用递归特征(recursive feature addition，RFA)对影响模型检测性能的特征进行选择，有效降低了模型训练时间。

上述方法虽然都解决了数据不平衡或者数据特征冗余的问题，但普遍存在检测准确率不高及对网络攻击分类不够精确的问题。基于此，本文提出一种结合 SMOTE(synthetic minority over-sampling technique)算法采样的 SDAE-LSTM(stacked deep auto-encoder-long short term memory)入侵检测模型，旨在解决攻击类别不平衡的问题，降低数据的维度，提高准确率，实现对入侵行为的精准捕获，为网络入侵检测提供一种行之有效的方法。

1　相关理论知识

1.1　SMOTE 算法

SMOTE 算法的核心是通过线性变换函数在一些距离较近的少数类数据中获得新数据，使原数据集类别间的数量相对平衡[12]。对少数类中的每个样本 x，计算欧氏距离，得到与其距离较近的 K 条数据。之后，根据不平衡率对样本集的采样比例进行设置，然后确定采样倍率 N。接下来随机选取样本 x 的 K 近邻中的多个样本，记所选样本为 y。对于所选样本 y，按照式(1)利用原始样本 x 构造新样本 z。

$$z = x + \text{rand}(0,1) \times |x - y| \tag{1}$$

其中 rand(0，1)表示生成 0 ~ 1 之间的随机数。

1.2　堆叠式深度自编码器原理

网络入侵检测数据具有维度高、数据量大的特点[13]，因此在构建入侵检测模型时所要面临的一个重要任务便是对所用数据集进行特征降维。自编码器(auto-encoder，AE)拥有极强的非线性拟合能力，这使它能够逼近任何拥有非线性特征的函数，除此之外，AE 还能够自动学习数据特征。基于此，本文选择堆叠式深度自编码器(stacked deep auto-encoder，SDAE)进行特征降维，这样既能自动进行深层特征提取，又同时保持了数据的一致[14]，得到最有用的特征。

堆叠式深度自编码器由多个自编码器相互叠加构成[15]，其凭借上一层隐藏层的表示作为下一层的输入，获得更加抽象的表示。图 1 为两个稀疏自编码器相互级联而成的两层堆叠式深度自编码器。

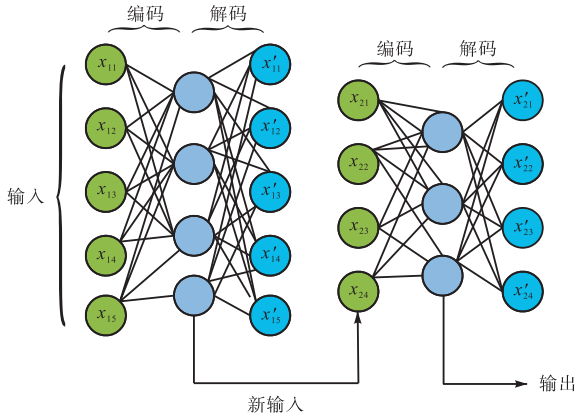


图 1　堆叠式深度自编码器

Fig. 1　Stacked deep auto-encoder

SDAE 是一种无监督的神经网络模型，能够通过

不断学习得到输入数据的深层表示。将输入层数据 x 进行转换，便可以得到其隐藏层的表示 $h(x)$，之后由隐藏层对其进行重构，还原出新的输入数据 $\bar{x}$。使重构后的数据 $\bar{x}$ 能够尽量还原 x 便是 SDAE 的训练目标。通常情况下会定义均方误差描述 SDAE 的损失函数，如式(2)所示。

$$L_s = L(w,b,c) = \frac{1}{2}\sum_{k=1}^{n}(x_k - \bar{x}_k)^2 \tag{2}$$

式中：w 表示网络权重系数，b 表示隐藏层偏置，c 表示输出层偏置。

为了防止过拟合，通常会在损失函数上加权重衰减，如式(3)所示。

$$L_s = L(w,b,c) = \frac{1}{2}\sum_{k=1}^{n}(x_k - \bar{x}_k)^2 + L_{wd} \tag{3}$$

权重衰减的公式为

$$L_{wd} = \frac{1}{2}\lambda(\|\boldsymbol{W}\|_F^2 + \|\boldsymbol{W}^*\|_F^2) \tag{4}$$

式中：$\|\boldsymbol{W}\|_F$ 为权重矩阵 $\boldsymbol{W}$ 的 F–范数，$\boldsymbol{W}^*$为 $\boldsymbol{W}$ 的转置矩阵，λ 为权重衰减系数。模型的最优参数 w、b、c 通过梯度下降算法求解得出。

1.3　长短期记忆神经网络

长短期记忆(long short term memory，LSTM)神经网络是循环神经网络(recurrent neural network，RNN)的一种优化模型，其单元结构如图 2 所示。LSTM 的关键是细胞(cell)的状态 C_t，为了删除或增加 cell 中的相关信息，LSTM 利用输入门、输出门和遗忘门控制信息的通过方式，并保护和控制 cell 的状态。

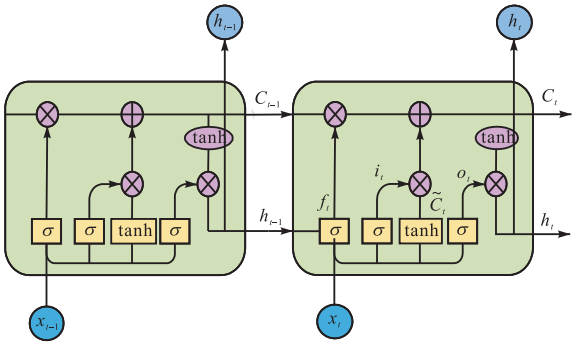


图 2　LSTM 单元结构

Fig. 2　LSTM unit structure

遗忘门 cell 状态中的信息是否需要删除由一个 sigmoid 层决定，计算方法如式(5)所示。对于输入 x_t 和 h_{t-1}，遗忘门会输出一个值域为[0，1]的数字，并将其放入 cell 状态 C_{t-1} 中。当值为 1 时，尽数保留；当值为 0 时，则全部清除。

$$f_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, x_t] + \boldsymbol{b}_f) \tag{5}$$

式中：x_t 为 LSTM 神经网络在 t 时刻的输入，h_t 为 t 时刻隐藏层的状态，f_t 为 t 时刻遗忘门的输出状态，σ 为 sigmoid 激活函数，$\boldsymbol{W}_f$为权重矩阵，$\boldsymbol{b}_f$为偏置向量。

cell 通过两个阶段增加新信息。第一阶段是决定信息的弃留，该阶段由一个包含 sigmoid 层的输入门完成，计算过程见式(6)。第二阶段 LSTM 会为上一步保留的信息生成一个向量 $\tilde{\boldsymbol{C}}_t$，对细胞状态进行更新，此过程由 tanh 函数实现，计算过程见式(7)。

$$i_t = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, x_t] + \boldsymbol{b}_i) \tag{6}$$

$$\tilde{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_c \cdot [h_{t-1}, x_t] + \boldsymbol{b}_c) \tag{7}$$

在输入门和遗忘门的基础之上，按照式(8)将细胞状态 C_{t-1} 进行更新。

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{\boldsymbol{C}}_t \tag{8}$$

式中：$f_t \cdot C_{t-1}$ 表示想要删除的信息，$i_t \cdot \tilde{\boldsymbol{C}}_t$ 表示新增加的信息。

LSTM 的输出内容是由输出门决定的，计算过程如式(9)和式(10)所示。首先利用 sigmoid 函数决定要输出的内容，然后用 tanh 函数把 cell 的状态值转换到 – 1 ~ 1 之间，并凭借 sigmoid 函数的非线性作用得到最终输出。

$$o_t = \sigma(\boldsymbol{W}_o \cdot [h_{t-1}, x_t] + \boldsymbol{b}_o) \tag{9}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{10}$$

2　结合 SMOTE 采样的 SDAE-LSTM 入侵检测模型

针对传统的网络入侵检测技术因数据不平衡及特征冗余而导致检测准确率不高的问题，提出了一种结合 SMOTE 采样的 SDAE-LSTM 入侵检测模型，如图 3 所示。检测流程共分为 3 个阶段：第一阶段，针对数据不平衡问题，使用 SMOTE 采样方法，对数据中的低频攻击样本进行增量处理，以达到各类别数据量平衡的目的，从而使入侵检测数据能够得到更加充分的训练，凭借此解决数据在传输过程中出现的丢失问题；第二阶段，利用堆叠式深度自编码器对采样后的数据进行特征降维，既能自动进行深层特征提取，又能够保持数据的一致，得到最有用的特征；第三阶段，将提取的深度特征送入 LSTM 进行网络异常检测、特征识别、分类并将分类结果输出。模型训练流程如图 4 所示，具体如下：

(1)初始化 SDAE 的权重参数 w 和隐藏层偏置 b。

(2)将经过预处理和 SMOTE 平衡化后的数据 x

送入 SDAE 模型。

(3) 计算各层的输出值，并通过误差反向传播方法对 SDAE 进行微调。

(4) 更新各层参数，直至达到最优。

(5) LSTM 训练。

(6) 对损失函数进行计算。

(7) 观察训练结果并进行参数调整，直到模型效果达到最优。

训练完成后，将测试集送入模型中进行网络异常检测，并得出分类结果。

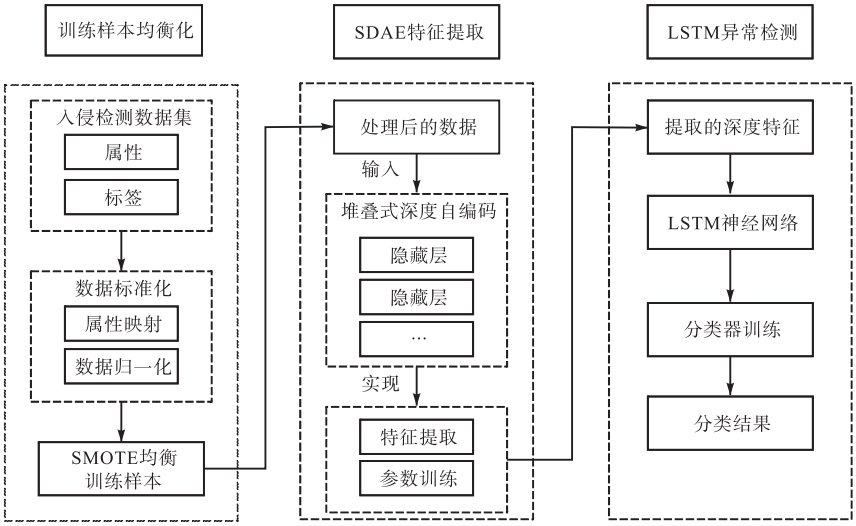


图 3　结合 SMOTE 采样的 SDAE-LSTM 入侵检测模型

Fig. 3　SDAE-LSTM intrusion detection model combined with SMOTE sampling

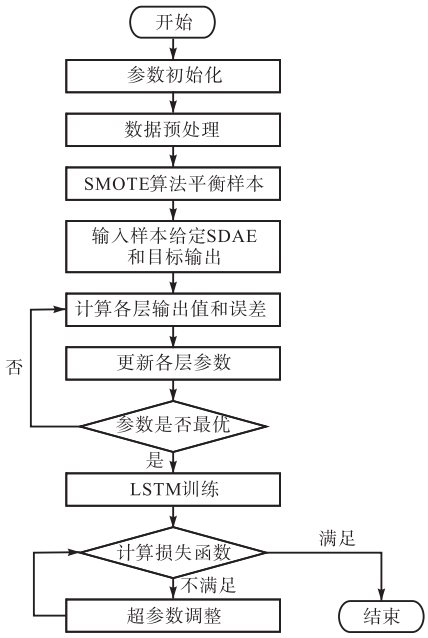


图 4　模型训练流程

Fig. 4　Model training process

3　实　验

实验所用数据集为 UNSW-NB15，其拥有真实的网络攻击行为和正常行为，可以全方位地反映当前网络攻击的多样性[16]。该数据集包含 9 种攻击类型和正常数据 Normal，分为训练集和测试集，各类数据分布见表 1。

表 1　UNSW-NB15 数据集各类数据分布

Tab. 1　UNSW-NB15 datasets distribution of various data

序号	攻击类型	样本数	
		训练集	测试集
1	Analysis	2 000	677
2	Shellcode	1 133	378
3	Backdoor	1 746	583
4	Worms	130	44
5	Fuzzers	18 184	6 062
6	Generic	40 000	18 871
7	Exploit	33 393	11 132
8	DoS	12 264	4 089
9	Reconnaissance	10 491	3 496
10	Normal	56 000	37 000

3.1　模型评价指标

为合理评价本文入侵检测模型的性能，本实验采用准确率(A)、精确率(P)、召回率(R)、误报率(F)和 F_1 值作为本文模型的评价标准。精确率和召回率可以在一定程度上表征模型的查准率与查全率。准确率为预测正确的样本个数占全部样本个数的比例，其比值越大意味着模型的检测能力越好。误报率为所有误报样本占正常样本的比例。

$$A=\frac{\sum_{i=1}^{n}N_{\mathrm{TP}_i}+N_{\mathrm{TN}_i}}{\sum_{i=1}^{n}(N_{\mathrm{TP}_i}+N_{\mathrm{TN}_i}+N_{\mathrm{FN}_i}+N_{\mathrm{FP}_i})} \tag{11}$$

$$P=\frac{\sum_{i=1}^{n}N_{\mathrm{TP}_i}}{\sum_{i=1}^{n}(N_{\mathrm{TP}_i}+N_{\mathrm{FP}_i})} \tag{12}$$

$$R=\frac{\sum_{i=1}^{n}N_{\mathrm{TP}_i}}{\sum_{i=1}^{n}(N_{\mathrm{TP}_i}+N_{\mathrm{FN}_i})} \tag{13}$$

$$F=\frac{\sum_{i=1}^{n}N_{\mathrm{FP}_i}}{\sum_{i=1}^{n}(N_{\mathrm{TN}_i}+N_{\mathrm{FP}_i})} \tag{14}$$

$$F_1=\frac{2P\cdot R}{P+R} \tag{15}$$

其中：N_{TP} 表示实例是正例并被预测为正例的数量，N_{FP} 表示实例是负例但被预测为正例的数量，N_{FN} 表示实例是正例但被预测为负例的数量，N_{TN} 表示实例是负例并被预测为负例的数量。

3.2　数据预处理

(1)属性映射

网络入侵检测数据多为字符型[17]，不能直接输入，需要对其进行处理后才能使用。属性映射便是将数据集中的符号特征转换为数字型特征，使所有数据均为数值型，便于入侵检测模型进行处理。UNSW-NB15 数据集共有 45 个特征属性，其中 state、proto、attack_cat 和 service 是字符型特征，它们无法直接输入，需要对其进行数值化处理。本实验选用 one-hot 独热编码对字符型特征进行数值化转换。

(2)数据归一化

入侵检测数据集进行属性映射后需要进行归一化处理[18]，以消除不同特征间的量纲给检测结果带来的负面影响，便于对模型进行综合评价。例如 dur 的取值有 0.655 618，而 sload 的取值有 124 104.4，如此巨大的量纲差异会对模型的训练及收敛速度造成很大影响。此外，数据集经归一化处理后能够有效避免奇异样本造成的不良影响。本文选取 min ~ max 标准化方法进行数据归一化处理，将数据映射到[0，1]区间内，如式(16)所示。

$$x'=(x-x_{\min})/(x_{\max}-x_{\min}) \tag{16}$$

式中：x 为经过转换后的数据，x' 为经过归一化处理后的数据，$x_{\min}$ 和 $x_{\max}$ 分别为特征最小值和特征最大值。

3.3　SMOTE 技术均衡训练样本

鉴于 UNSW-NB15 数据集存在数据不平衡的问题，本研究选用 SMOTE 采样算法对数据集中的稀有攻击类型样本进行扩充，以使各类别样本比例变得均衡[19]。采样前后各类型训练集样本数对比见表 2。

表 2　采样前后各类型训练集样本数对比

Tab. 2　Comparison of various types of data before and after sampling

序号	攻击类型	训练集样本数	
		采样前	采样后
1	Analysis	2 000	10 852
2	Shellcode	1 133	10 750
3	Backdoor	1 746	11 022
4	Worms	130	11 036
5	Fuzzers	18 184	18 184
6	Generic	40 000	40 000
7	Exploit	33 393	33 393
8	DoS	12 264	12 264
9	Reconnaissance	10 491	10 491
10	Normal	56 000	56 000

3.4　实验设置与结果分析

由于神经网络模型中参数较多，故在进行实验之前需要对参数进行设置，本实验通过对比分析，将 SDAE 的结构及 LSTM 的参数进行设置。

SDAE 结构设置：本文所用 SDAE 模型深度为 4，每层神经元个数分别设置为 128、64、64、32，激活函数为 ReLU 函数。

LSTM 模型参数设置：本文所用 LSTM 模型层数为两层，每层节点个数为 128，学习率为 0.002，批大小为 128，训练次数为 500，优化器选用 Adam。

3.4.1　SMOTE 算法性能分析

为了验证本文所用采样方法的性能，实验对稀有攻击样本采样前后各类型的召回率和精确率进行了对比，结果如图 5 和图 6 所示。

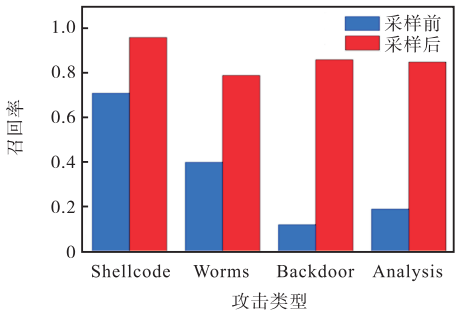


图 5　稀有攻击类型样本采样前后的召回率对比

Fig. 5　Comparison of recall before and after sampling rare attack samples

实验结果表明，对稀有攻击类型样本进行 SMOTE 采样后，Shellcode、Worms、Backdoor 和 Analysis 的召回率及精确率都有明显提高。攻击类型 Analysis 和 Backdoor 在进行 SMOTE 采样之前，其召

回率都普遍较低，这是因为二者样本数量过少，数据不能充分学习，使其行为难以检测，从而导致召回率较低。分别对稀有攻击类型的数据样本进行采样，其样本数量得到增加，提高了稀有攻击类别数据的占比，使其能够充分被分类器学习，从而使召回率得以提高。使用 SMOTE 技术对所用样本进行平衡化处理，能够有效提高稀有攻击类型的召回率和精确率，

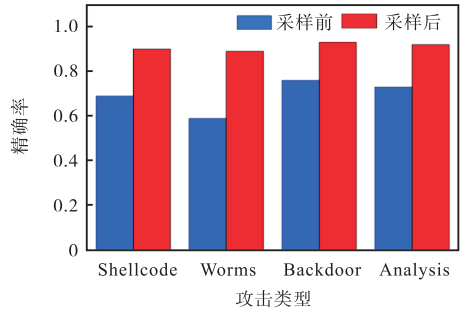


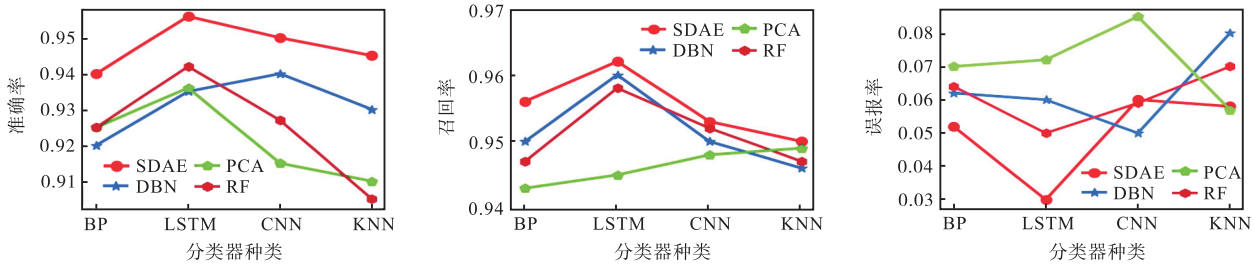
图 6　稀有攻击类型样本采样前后的精确率对比

Fig. 6　Comparison of precision before and after sampling rare attack samples

从而使模型获得更优异的入侵检测性能。

3.4.2　SDAE 特征提取模型性能分析

为了验证堆叠式深度自编码器 SDAE 优异的深层特征提取能力，本文以 UNSW-NB15 数据集为基础，以准确率、召回率和误报率为评价指标，在 LSTM、BP、KNN 和 CNN 这 4 种分类器上，将 SDAE、主成分分析（PCA）、深度信念网络（DBN）和随机森林（RF）这 4 种特征提取模型的性能进行了对比，如图 7 所示。由实验结果可知，当对某一种分类器进行固定时，以上 4 种特征提取模型都表现出了较好的性能，分类准确率和召回率均较高，误报率较小，但相比较而言 SDAE 特征提取模型的准确率最高，其次是 DBN 和 RF，最差为 PCA。SDAE 准确率之所以高是因为其具有强大的特征提取能力，使冗余特征被消除，留下的是数据最有用的特征，能够让分类器进行精准学习，使模型更好地识别出各种攻击类型数据的类别，提高检测准确率。



（a）准确率　　（b）召回率　　（c）误报率

图 7　不同特征提取模型在不同分类器上的准确率、召回率和误报率

Fig. 7　Accuracy, recall and false alarm of different feature extraction models on different classifiers

3.4.3　模型整体性能分析

为了进一步验证本文所提出的结合 SMOTE 采样的 SDAE-LSTM 模型在网络入侵检测方面的整体性能，本文选取文献综述中所提到的 ORF[6]、ANN-WTF[7]、PCA-MLP[10]、DBN-RFA[11]这 4 种入侵检测模型作为对比模型，以 UNSW-NB15 数据集为基础，以准确率、精确率、召回率和 F_1 值为评价指标，对比结果见表 3。

表 3　不同入侵检测模型的性能

Tab. 3　Performance of different intrusion detection models

模型	准确率	精确率	召回率	F_1
本文模型	0.958	0.953	0.961	0.955
DBN-RFA 模型	0.928	0.936	0.972	0.932
ANN-WTF 模型	0.917	0.931	0.969	0.924
PCA-MLP 模型	0.908	0.902	0.913	0.905
ORF 模型	0.780	0.800	0.780	0.790

由表 3 可知：本文模型对网络攻击分类的准确率、精确率与 F_1 值均高于其他模型。本文模型的召回率高于 PCA-MLP 模型和 ORF 模型，稍低于 DBN-RFA 模型与 ANN-WTF 模型。DBN-RFA 和 ANN-WTF 两种入侵检测模型的总体性能优于 PCA-MLP 模型，但逊色于本文模型，那是因为虽然 DBN 和 ANN 都能够对数据进行降维，但却无法实现数据的深度特征提取，而 SDAE 能够对数据进行深度特征提取，使分类器可以精准学习，故而性能优于二者。在对比结果中，PCA-MLP 模型分类性能较差，一方面是因为 PCA 只能进行线性降维，无法对非线性特征进行全面描述，另一方面是因为 MLP 是单一的神经网络模型，分类效果不及其他几种模型，从而导致其性能较差。ORF 模型在 5 种对比模型中性能最差，虽然作者对其进行了优化，但随机森林仍存在一些分类能力较差的决策树，使检测模型的分类准确率变低。从以上 5 种模型的对比结果可以看出，本文模型的精确率及对网络攻击识别的准确率较高，召回率稍有不足，总体上优于其他对比模型。

4 结　语

本文提出一种结合 SMOTE 采样的 SDAE-LSTM 入侵检测算法，对网络异常行为进行了检测。针对数据不平衡问题，采用 SMOTE 采样方法，在少量攻击类样本点之间随机插入样本，增加其数量，以达到类间平衡的目的。针对特征冗余问题，利用堆叠式深度自编码器进行数据降维，得到数据的主要特征，让分类器进行精准学习，使模型更好地识别出各种攻击类型数据的类别。最后，基于 LSTM 神经网络，精准捕获网络入侵特征，准确地实现入侵检测。实验结果表明，通过 SMOTE 采样与 SDAE 深层特征提取，大大降低了数据的冗余度，模型准确率达到 0.958，为网络入侵检测提供了一种行之有效的方法。

参考文献：

[1] 张仁杰，陈伟，杭梦鑫，等. 基于变分自编码器的不平衡样本异常流量检测[J]. 计算机科学，2021，48(7)：62–69.

[2] 饶海兵，朱苏磊，杨春夏. 基于空时特征融合和注意力机制的网络入侵检测模型[J]. 计算机与现代化，2022(6)：116–121.

[3] 王伟. 基于深度学习的网络流量分类及异常检测方法研究[D]. 合肥：中国科学技术大学，2018.

[4] 罗文华，许彩滇. 利用改进 DBSCAN 聚类实现多步式网络入侵类别检测[J]. 小型微型计算机系统，2020，41(8)：1725–1731.

[5] 汪祖民，王冬昊，梁霞，等. 基于 DBSCAN_GAN_XGBoost 的网络入侵检测方法[J]. 郑州大学学报(工学版)，2022，43(3)：44–51.

[6] 章缙，李洪赭，李赛飞. 针对基于随机森林的网络入侵检测模型的优化研究[J]. 计算机与数字工程，2022，50(1)：106–110.

[7] HAMID Y，SHAH F A，SUGUMARAN M. Wavelet neural network model for network intrusion detection system[J]. International journal of information technology，2019，1(2)：251–263.

[8] 葛继科，刘浩因，李青霞，等. 基于改进 CNN-LSTM 的网络入侵检测模型研究[J]. 软件工程，2022，25(1)：56–58.

[9] 夏栋梁，刘玉坤，鲁书喜. 基于蚁群算法和改进 SSO 的混合网络入侵检测方法[J]. 重庆邮电大学学报(自然科学版)，2016，28(3)：406–413.

[10] 刘辉. 基于主成分分析和多层感知机神经网络的入侵检测方法研究[J]. 软件工程，2020，23(7)：10–12.

[11] 赵荷，盖玲. 深度置信网络结合递归特征添加的网络入侵检测方法[J]. 计算机应用与软件，2020，37(11)：304–310.

[12] CHAWALA N V，BOWYER K W，HALL L O，et al. SMOTE：synthetic minority over-sampling technique[J]. Journal of artificial intelligence research，2002，16(1)：321–357.

[13] HAN X，CUI R B，LAN Y F，et al. A Gaussian mixture model based combined resampling algorithm for classification of imbalanced credit data sets[J]. International journal of machine learning and cybernetics，2019，10(12)：3687–3699.

[14] VINCENT P，LATOCHELLE H，LAJOIE I，et al. Stacked denoising auto-encoders：learning useful representations in a deep network with a local denoising criterion[J]. The journal of machine learning research，2010，11(12)：3371–3408.

[15] PARK S，SEO S，KIM J. Network intrusion detection using stacked denoising auto-encoder[J]. Advanced science letters，2017，23(10)：9907–9911.

[16] NG W W Y，HU J J，YEUNG D S，et al. Diversified sensitivity-based under sampling for imbalance classification problems[J]. IEEE Transactions on cybernetics，2017，45(11)：2402–2412.

[17] 魏明军，张鑫楠，刘亚志，等. 一种基于 SSA-BRF 的网络入侵检测方法[J]. 河北大学学报(自然科学版)，2022，42(5)：552–560.

[18] 声秋，李友国，高渊，等. 基于对抗深度学习的物联网安全检测方法[J]. 电子设计工程，2022，30(11)：50–54.

[19] 王伟，代红，赵斯祺. 基于特征优化和 BP 神经网络的入侵检测方法[J]. 计算机工程与设计，2021，42(10)：2755–2761.

责任编辑：郎婧