深度置信网络融合局部保持投影的入侵检测模型

武玉坤[1,2]　李　伟[2]　陈沅涛[3]

[1](杭州职业技术学院　浙江 杭州 310018)

[2](浙江工业大学计算机科学与技术学院　浙江 杭州 310023)

[3](长沙理工大学计算机与通信工程学院　湖南 长沙 410114)

摘　要　网络入侵检测系统(NIDS)提供了比其他传统网络防御技术(如防火墙系统)更好的网络安全解决方案。提出一种深度置信网络(DBN)与局部保持投影技术相融合的入侵检测模型。深度置信网络用于原始数据的特征学习;采用局部保持投影(LPP)融合深层特征,进一步去除冗余和无关特征。最后使用 Softmax 分类器进行分类。研究该方法在 NSL-KDD 数据集和 UNSW-NB15 数据集上的准确率、检测率、误报率等分类指标,并与常规的机器学习分类方法及其他文献中最新的方法进行比较。实验结果表明 DBN-LPP 模型提高了入侵检测的综合性能,其性能优于传统的机器学习分类方法及其他方法,为入侵检测提供了一种新的研究方法。

关键词　入侵检测　深度学习　深度置信网络　局部保持投影

中图分类号　TP391　　**文献标志码**　A　　DOI:10.3969/j.issn.1000-386x.2024.06.010

INTRUSION DETECTION MODEL BASED ON DEEP BELIEF NETWORK FUSING LOCALITY PRESERVING PROJECTION

Wu Yukun[1,2]　Li Wei[2]　Chen Yuantao[3]

[1](*Hangzhou Vocational and Technical College*, *Hangzhou 310018*, *Zhejiang*, *China*)

[2](*College of Computer Science and Technology*, *Zhejiang University of Technology*, *Hangzhou 310023*, *Zhejiang*, *China*)

[3](*School of Computer and Communication Engineering*, *Changsha University of Science and Technology*, *Changsha 410114*, *Hunan*, *China*)

Abstract　Network intrusion detection systems (NIDS) provide a better solution to network security than other traditional network defense technologies, such as firewall systems. This paper proposes an intrusion detection model that combines deep belief network (DBN) and local preserving projection (LPP). The DBN was used for feature learning of the original data, and the LPP was used to fuse the deep features to further remove redundant and irrelevant features. Softmax classifier was used for classification. In addition, the accuracy, detection rate, false alarm rate and other classification indicators of this method on the NSL-KDD data set and UNSW-NB15 data set were studied and compared with the conventional machine learning classification method and the latest model method in other literature. The experimental results show that the DBN-LPP model improves the comprehensive performance of intrusion detection system, and its performance is better than traditional machine learning classification methods and other methods. This paper provides a new research method for intrusion detection.

Keywords　Intrusion detection　Deep learning　Deep belief network　Locality preserving projection

---

收稿日期:2020 - 12 - 23。国家自然科学基金项目(61502422,61972056);浙江省自然科学基金项目(LY18F020028);浙江省科技厅公益项目(2017C33108);浙江省教育厅一般科研项目(Y202044619);杭州职业技术学院高层次人才科研启动项目(RCXY202243)。
武玉坤,教授,主研领域:机器学习,数据挖掘。**李伟**,教授。**陈沅涛**,副教授。

0 引　言

互联网的高速发展正在改变着人们的生活、学习和工作,同时我们也面临着各种各样的网络安全问题,如何识别网络攻击特别是未曾出现的攻击是一个关键的技术问题。网络入侵检测系统(NIDS)提供了更好的解决方案相对于传统的防火墙系统,能帮助网络管理员检测网络中的攻击、漏洞和蓄意破坏。NIDS 分为基于签名的网络入侵检测系统(SNIDS)和基于异常的网络入侵检测系统(ADNIDS)两种形式。在 SNIDS 中,系统根据 NIDS 中预先安装的规则检测攻击。将网络流量与更新后的攻击签名数据库进行比较,以检测网络流量数据集中的入侵。

基于异常的检测系统通过研究网络流量中不正常行为对网络流量中的未知或异常行为进行分类,偏离正常流量模式的网络行为被归为入侵。其优点是可以预测未知和新的攻击。因此,我们重点研究这类入侵检测系统。异常检测方法可以应用于网络安全、信用卡欺诈检测、军事应用和许多医疗应用[1]。实际上入侵检测本质上是一个分类问题,用来区分网络流量行为是正常的还是异常的。各种机器学习技术已被广泛应用于入侵检测系统中,如支持向量机(SVM)[2]、K 最近邻(KNN)[3]、朴素贝叶斯(NB)[4-5]、人工神经网络(ANN)[6]、随机森林(RF)[7-8]和自组织映射(SOM)[9],以及其他相关机器学习方法[10-11]。

然而,传统的机器学习方法大都属于浅层学习,往往强调特征工程;它们不能有效地解决实际网络应用环境中出现的海量入侵数据分类问题。随着数据集的动态增长,多个分类任务会导致准确率下降。此外,浅层结构限制了在入侵检测问题中学习复杂非线性关系的能力,浅层学习不适合大规模数据的高维学习的智能分析和预测要求。相反,深度学习有潜力从数据中提取更好的表示,从而创建更好的模型。因此,入侵检测技术在经历了一个相对较慢的时期后又得到了快速的发展。

深度学习理论[12]提出后,深度学习理论与技术在机器学习领域迅速崛起。在此背景下,相关的理论论文和实践研究成果层出不穷,并取得了令人瞩目的成果,在大数据时代,深度学习算法在大多数典型的机器学习应用中都取得了优异的效果,尤其是在语音识别、图像识别[13]和动作识别[14-16]等领域。近年来,深度学习理论和技术得到了飞速的发展,这意味着人工智能的新时代已经开启,为发展智能入侵检测技术提供了全新的途径。

深度学习是一种新的机器学习方法,具有克服传统机器学习方法不足的潜力。深度学习模型包含多个隐含层,这些隐含层的特征并不是由人工设计的,而是自动从输入数据中学习的[13]。深度信念网络(Deep Belief Network,DBN)是一种流行的深度学习模型,近年来已成功应用于入侵检测。Gao 等[17]用训练 DBN 作为入侵检测的分类器。同样,Alom 等[18]也通过一系列实验利用 DBN 来进行检测入侵,Zhang 等[19]使用遗传算法寻找最优的 DBN 网络结构来进行物联网的入侵检测,Yang 等[20]使用密度峰值聚类联合 DBN 进行入侵检测系统的构建。DBN 的成功主要归功于两个方面:首先,DBN 可以从输入的数据中自动学习有用的特征,从而消除了对特征选择的必要性。其次,DBN 是由一些训练好的受限玻尔兹曼机(RBM)构造的,这使得它比人工神经网络或支持向量机更有可能有效地学习复杂的非线性关系。

尽管 DBN 等深度学习模型能够在很大程度上自动从原始数据中学习有用的特征,但所学习的深度特征往往是高维的,包含冗余信息[21],这可能会降低分类精度,导致训练时间增加。为了进一步提高学习到的深度特征的质量,可以考虑结合一种特征融合方法——局部保持投影(LPP),将 DBN 学习到的深度特征融合在一起,提取出最具代表性的信息,减小维数。已有多项研究证实,与主成分分析(PCA)[22]相比,LPP 在特征融合方面具有显著优势。

本文提出了一种深度置信网络融合局部保持投影(LPP)进行入侵检测的新方法。首先,利用一系列预先训练的受限玻尔兹曼机(RBM)构造深度置信网络(DBN),对原始网络流量数据进行特征学习。然后采用局部保持投影(LPP)融合深度特征,进一步提高学习特征的质量。最后,将融合的深度特征输入 Softmax 进行分类检测。将该方法应用于 NSL-KDD 数据集和 UNSW-NB15 数据集进行入侵检测。结果表明,该方法克服了传统的特征提取方法依赖于手工特征提取的缺点,具有较高的效率和可靠性。

1 相关技术与新方法

1.1 RBM

RBM 作为基于能量生成模型的一个特例,能够为未知的数据分布[23]提供一个学习模型。每个 RBM 包含一个可见层和一个隐藏层,如图 1 所示。同一层的单元没有连接。相邻两层的单元具有对称的定向

连接[24-25]。

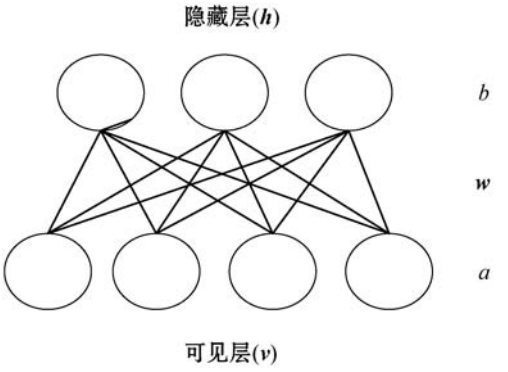

图 1 受限玻尔兹曼机结构

对于图 1 所示 RBM,其中可见向量 $\boldsymbol{v}=\{v_1,v_2,\cdots,v_I\}\in\{0,1\}$,隐藏向量 $\boldsymbol{h}=\{h_1,h_2,\cdots,h_J\}\in\{0,1\}$,权重矩阵 $\boldsymbol{w}$,可见偏置 a,隐藏偏置 b,则一组状态$(\boldsymbol{v},\boldsymbol{h})$的能量函数为:

$$E(\boldsymbol{v},\boldsymbol{h})=-\sum_{i=1}^{I}a_iv_i-\sum_{j=1}^{J}b_jh_j-\sum_{i=1}^{I}\sum_{j=1}^{J}v_iw_{ij}h_j \tag{1}$$

式中:I 和 J 分别为可见层与隐藏层的单元数量。可见层与隐藏层的联合概率分布为:

$$p(\boldsymbol{v},\boldsymbol{h})=\frac{\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})}}{Z} \tag{2}$$

$$Z=\sum_{v}\sum_{h}\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})} \tag{3}$$

式中:Z 为归一化因子,也称作配分函数。

可见层与隐藏层神经元向量的边缘概率分布分别为:

$$p(\boldsymbol{v})=\sum_{h}p(\boldsymbol{v},\boldsymbol{h})=\frac{\sum_{h}\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})}}{\sum_{v}\sum_{h}\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})}} \tag{4}$$

$$p(\boldsymbol{h})=\sum_{v}p(\boldsymbol{v},\boldsymbol{h})=\frac{\sum_{v}\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})}}{\sum_{v}\sum_{h}\mathrm{e}^{-E(\boldsymbol{v},\boldsymbol{h})}} \tag{5}$$

可见层与隐藏层的条件概率分别为:

$$p(\boldsymbol{v}\mid\boldsymbol{h})=\prod_{i}p(v_i\mid\boldsymbol{h}) \tag{6}$$

$$p(\boldsymbol{h}\mid\boldsymbol{v})=\prod_{j}p(h_j\mid\boldsymbol{v}) \tag{7}$$

通过式(8)计算隐藏单元的状态,然后通过式(9)计算重构可见单元的状态,通过最小化可见单元与其重构单元之间的差值,使得隐藏单元成为可见单元降维后的新的表示形式。换句话说,隐藏单元表示从可见单元中提取的特征。

$$p(h_j=1/\boldsymbol{v})=\frac{1}{1+\exp\left(-b_j-\sum_{i=1}^{I}v_iw_{ij}\right)} \tag{8}$$

$$p(v_i=1/\boldsymbol{h})=\frac{1}{1+\exp\left(-a_i-\sum_{j=1}^{J}h_jw_{ij}\right)} \tag{9}$$

RBM 的训练分为正向阶段和负向阶段。正向阶段将数据从可见层传输到隐藏层,负向阶段将数据从隐层传输到可见层进行重构。RBM 学习的正向阶段和负向阶段分别由式(8)和式(9)表示。训练前的目标是确保训练后的 RBM 模型尽可能接近地描述输入数据的分布。即寻找 RBM 的最优参数 $\theta=\{a,b,\boldsymbol{w}\}$ 来使得 $p(\boldsymbol{v})$ 最大,公式如下:

$$\frac{\partial\log p(\boldsymbol{v})}{\partial\theta}=-<\frac{\partial E(\boldsymbol{v},h)}{\partial\theta}>_{\text{data}}+<\frac{\partial E(\boldsymbol{v},\boldsymbol{h})}{\partial\theta}>_{\text{model}} \tag{10}$$

式中:$<\cdot>_{\text{data}}$和$<\cdot>_{\text{model}}$分别表示式(7)和式(2)两种概率分布的数学期望,由于$<\cdot>_{\text{model}}$的计算复杂度巨大,其值很难获取,对比散度算法(Contrastive Divergence (CD))[26]被采用来更新模型的参数,如式(11)-式(13)所示。

$$\Delta w_{ij}^{n}=m\Delta w_{ij}^{n-1}+\eta(<v_ih_j>_{\text{data}}-<v_ih_j>_k) \tag{11}$$

$$\Delta b_j^{n}=m\Delta b_j^{n-1}+\eta(<h_j>_{\text{data}}-<h_j>_k) \tag{12}$$

$$\Delta a_i^{n}=m\Delta a_i^{n-1}+\eta(<v_i>_{\text{data}}-<v_i>_k) \tag{13}$$

式中:$\eta\in[0,1]$是学习率,用来调整模型的学习速度;$m\in[0,1]$是动量参数;n 是训练的迭代次数;k 表示 k 步对比散度,实践证明 $k=1$ 已工作得较好[27]。

1.2 DBN 的构建与训练

DBN 的构建过程如图 2 所示。每个 RBM 层都使用下层 RBM 的激活概率作为输入数据进行训练,输出作为上一层 RBM 的输入。可见层和第一层隐层形成第一个 RBM,第一隐层和第二隐层形成第二个 RBM,第二隐层和第三隐层形成第三个 RBM,最后,在顶层添加一个 Softmax 分类器。

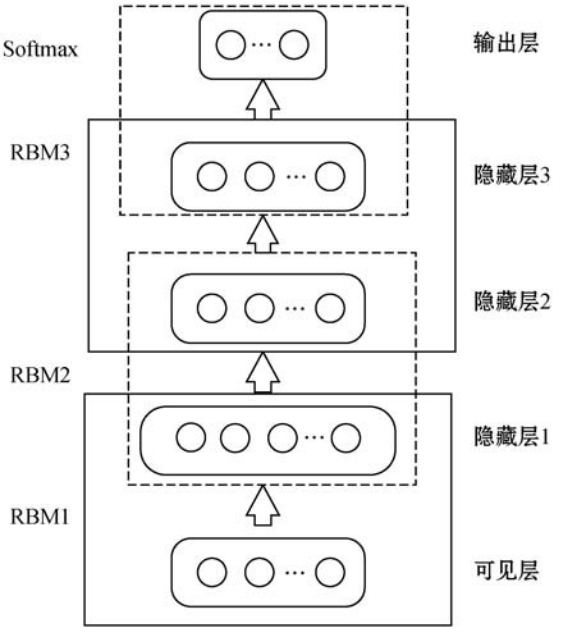

图 2 深度置信网络模型

DBN 的训练过程包括预训练与微调,预训练过程

即分别训练每个 RBM,微调阶段将进一步减少训练误差,提高分类精度。采用反向传播算法对标签数据进行参数微调。与无监督预训练不同的是,有监督的微调可以同时更新所有的参数,直到达到最大迭代次数。当预训练和微调阶段都完成后,DBN 可以用于实际的入侵检测分类。

1.3　融合局部保持投影

标准 DBN 学习到的深层特征通常是高维的,包含了一定的噪声。LPP 是一种数据融合方法,能够有效地获取流形数据内在结构的有用信息[27]。本文将 LPP 用于去除高维深度特征的冗余信息,获取有价值的低维表示。

LPP 的目标是在线性近似下保留原始数据集的局部结构。假设 DBN 学习到的深度特征数据集为 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}\in\mathbf{R}^D$,其中 n 是样本数量,D 是样本维度;LPP 的目的是寻找一个映射矩阵 $\boldsymbol{W}$ 来获取一个低维的数据集 $\boldsymbol{Y}=\{y_1,y_2,\cdots,y_n\}\in\mathbf{R}^d$,其中 $d\ll D$,$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的对应关系如式(14)所示。

$$\boldsymbol{Y}=\boldsymbol{W}^{\mathrm{T}}x_i\quad i=1,2,\cdots,n \tag{14}$$

映射矩阵 $\boldsymbol{W}\in\mathbf{R}^{D\times d}$ 是通过优化式(15)所示的目标函数。

$$\min\sum_{i,j}^{n}(y_i-y_j)^2\boldsymbol{S}_{ij} \tag{15}$$

式中:$\boldsymbol{S}_{ij}$ 是权值矩阵,代表了两样本的关系。通过近邻图计算,矩阵内部元素的定义为:

$$S_{ij}=\begin{cases}\dfrac{\exp(\|x_i-x_j\|^2)}{t} & x_j\in N_k(x_i)\text{或}x_i\in N_k(x_j)\\ 0 & \text{其他}\end{cases} \tag{16}$$

t 是总体样本方差。从权值矩阵 $\boldsymbol{S}_{ij}$ 的设置中可以看出,在对应近邻样本的位置上赋非零权值,而相距较远的样本则赋零。则在投影中达到保留样本的局部邻域的目的。

将式(15)中的最小优化问题转化为最小化以下代数方程:

$$\frac{1}{2}\sum_{ij}\|y_i-y_j\|^2\boldsymbol{S}_{ij}=\frac{1}{2}\sum_{ij}(\boldsymbol{W}^{\mathrm{T}}x_i-\boldsymbol{W}^{\mathrm{T}}x_j)\boldsymbol{S}_{ij}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{D}-\boldsymbol{S})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{17}$$

式中:$\boldsymbol{D}$ 是 N 阶对角矩阵,且 $\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{S}_{ij}$,$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{S}$ 是拉普拉斯矩阵。然后,通过求解广义特征值问题得到变换矩阵:

$$\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}=\lambda\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{18}$$

式中:λ 是特征值,$\boldsymbol{W}$ 对应着特征向量,前 d 个特征向量 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$ 与前 d 个最小的非零特征值相对应。最终降维后的数据表示如下:

$$\boldsymbol{Y}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\quad\boldsymbol{W}^{\mathrm{T}}=(\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d) \tag{19}$$

降维后的向量 $\boldsymbol{Y}$ 保留了原始深度特征的局部重要信息,作为分类器的最终输入特征。通常,Softmax 是基于深度特征进行分类的较优选择。该方法采用 DBN 作为特征提取器,LPP 作为特征优化器,Softmax 作为分类器。

1.4　方法流程

该方法的流程如图 3 所示,一般步骤如下。

对数据集的训练集与测试集分别进行预处理:利用一系列预训练的 RBM 构建 DBN 模型,学习训练样本的深层特征;利用 LPP 对深度特征进行提炼,进一步去除冗余信息,提高特征的质量;将提炼的深度特征输入 Softmax 分类器进行入侵检测分类;用测试样本验证该方法的有效性。

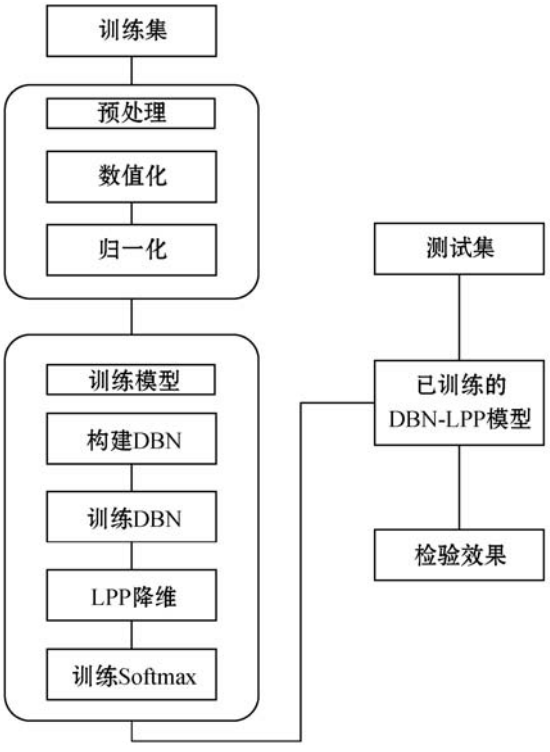

图 3　DBN-LPP 流程

2　实验与分析

2.1　实验设置

2.1.1　实验数据

为了验证本文提出算法的有效性,本文采用 2 个比较流行的网络入侵数据集来进行实验分析。

1) NSL-KDD 数据集。2009 年生成的 NSL-KDD 数据集[28-29]被广泛应用于入侵检测实验中。最近的文献[30-32]中研究者都是使用 NSL-KDD 作为基准数据

集,NSL-KDD 数据集不仅有效地解决了 KDDCUP99 数据集固有的冗余记录的问题,也使记录的数量更加合理。数据集包括训练集 KDDTrain + ,测试集 KDDTest + 和 KDDTest-21。数据集包括普通记录和四种不同类型的攻击记录,如表 1 所示。KDDTest-21 数据集是 KDDTest + 的子集,并且更难于分类。每条记录有 41 个特征和 1 个类标签,特征包括基本特征、内容特征和流量特征[30]。根据其特点,攻击有四种类型:DoS(拒绝服务攻击)、R2L(远程主机的未授权访问)、U2R(非法的本地超级用户特权访问)和 Probe(端口监视或扫描)。测试集具有特定的攻击类型,这些攻击类型在训练集中没有出现,为入侵检测提供了更加现实的理论依据。

表 1 NSL-KDD 数据集

数据集	**Total**	Normal	Dos	Probe	R2L	U2R
KDDTrain +	**125 973**	67 343	45 927	11 656	995	52
KDDTest +	**22 544**	9 711	7 458	2 421	2 754	200
KDDTest-21	**11 850**	2 152	4 342	2 402	2 754	200

2) UNSW-NB15。UNSW-NB15[33-34]是一个新数据集。该数据集反映了一个更现代、更复杂的威胁环境。该数据集是由网络数据采集分析工具 TcpDump 抓取原始的网络数据包,然后由黑客流量监控工具 Argus,开源入侵检测工具 Zeek-IDS 和 12 种算法生成了带有类标签的 49 个特征[35]。完整的数据集总共包含 25 400 443 条记录。根据分层抽样方法从完整的数据集中抽取了一个子集,该子集的训练集包含 175 341 条记录,测试集包含 82 332 条记录。该子集仅仅包含 43 个特征。子集中含有 10 种类别,1 个正常类别,9 种攻击类别。具体如表 2 所示。

表 2 UNSW-NB15 数据集

类别	UNSW_NB15_Training-set	UNSW_NB15_Testing-set
Normal	56 000	37 000
Generic	40 000	18 871
Exploits	33 393	11 132
Fuzzers	18 184	6 062
DoS	12 264	4 089
Reconnaissance	10 491	3 496
Analysis	2 000	677
Backdoor	1 746	583
Shellcode	1 133	374
Worms	130	44
Total	**175 341**	**82 332**

图 4 和图 5 分别显示了 NSL-KDD 和 UNSW-NB15 两种数据集通过 UMAP(Uniform Manifold Approximation and Projection)[35]投影到二维空间的原始训练数据的样本空间分布。可以看出 2 种数据集在低维空间都是非线性可分的,这符合流形学习技术的应用,而 LPP 能够有效地恢复内在流形结构的重要信息[36],学习到数据的本质特征。

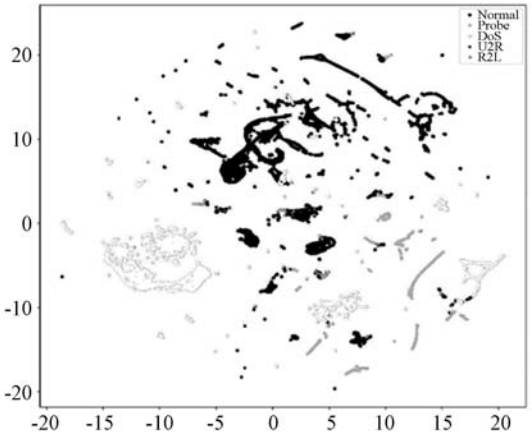

图 4 NSL-KDD 数据集可视化

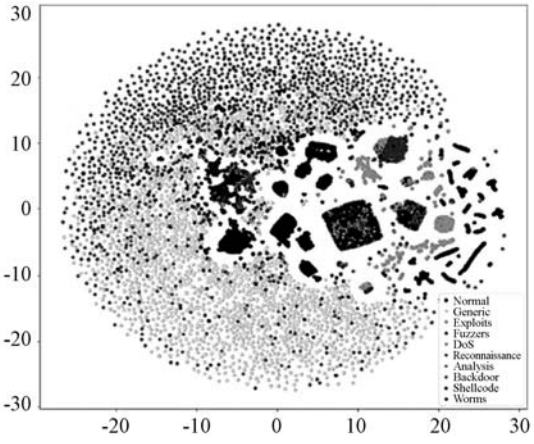

图 5 UNSW-NB15 数据集可视化

2.1.2 数据预处理

1) 字符字段数值化。NSL-KDD 数据集中有 38 个数值特征字段和 3 个非数值特征字段。由于本文所提模型的输入值应该是一个数字矩阵,因此必须将一些非数值特征(如协议类型、服务和标志字段)转换为数值形式。例如字段 protocol_type 有三种类型的属性,tcp、udp 和 icmp,采用 one-hot 编码为二进制向量(1,0,0)(0,1,0)和(0,0,1),类似地,services 有 70 种属性特征,字段 flag 有 11 种属性特征。最终 41 维特征经过变换后映射成 122 维特征。

UNSW-NB15 数据集中包含 39 个数值型特征和 3 个类别字段,针对 UNSW-NB15 中的 proto、service 和 state 非数值型属性进行同样的编码,最终 42 维特征映射为 196 维。

2) 归一化。NSL-KDD 数据集和 UNSW-NB15 数

据集的一些特征的取值有非常大的范围,比如字段 duration 的取值其他特征值也存在较大的差异,从而使特征值无法比较,不适合处理。因此,根据式(20),利用 max-min 归一化将所有特征值映射到[0,1]范围,进行归一化处理。

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{20}$$

式中:x_i为数据的属性值,$x_{\min}$为该数据属性的最小值,$x_{\max}$为最大值。

2.2　评估标准

在入侵检测性能评估的对比实验中,采用准确率(Acc)、检测率(R)、精度(P)、F1 分数(F1)误报率(FAR)等评价指标来衡量模型的性能,其中准确率、检测率、精度、F1、误报率的定义分别为:

$$A_{\text{cc}} = \frac{T_{\text{P}} + T_{\text{N}}}{T_{\text{P}} + T_{\text{N}} + F_{\text{P}} + F_{\text{N}}} \tag{21}$$

$$D_{\text{R}} = R_{\text{ecall}} = \frac{T_{\text{P}}}{T_{\text{P}} + F_{\text{N}}} \tag{22}$$

$$P_{\text{recision}} = \frac{T_{\text{P}}}{T_{\text{P}} + F_{\text{P}}} \tag{23}$$

$$F_{\text{1-score}} = \frac{2(P_{\text{recision}} \times R_{\text{ecall}})}{P_{\text{recision}} + R_{\text{ecall}}} \tag{24}$$

$$F_{\text{AR}} = \frac{F_{\text{P}}}{F_{\text{P}} + T_{\text{N}}} \tag{25}$$

式中:T_{P}(True Positive)表示正确地预测出攻击记录的数量;T_{N}(True Negative)表示正确地预测出正常记录的数量;F_{P}(False Positive)表示错误地预测出攻击记录的数量;F_{N}(False Negative)表示错误地预测出正常记录的数量。

2.3　实验结果与讨论

本文使用当前最广泛的深度学习框架 TensorFlow 来进行编程。该实验是在深度学习机上进行的,该机器配置英特尔 C621@ 主频 2.10GHz,睿频 4.00 GHz,64 GB 内存,8 GB 显存。设计了两组实验,研究了 DBN-LPP 模型对正常、异常二分类和两种数据集中的多分类的性能。

为了验证提出方法的先进性,本文设计了几组实验,一是比较特征融合前与融合后的结果对比,二是与其他的特征优化方法进行对比,如 PCA 等,三是与其他机器学习方法进行比较,如逻辑回归(LR)、决策树(J48)、朴素贝叶斯(NB)、随机森林(RF)、K 近邻分类(KNN)、支持向量机(SVM)、多层感知机(MLP)等机器学习方法的性能。在这些机器学习算法中我们也进行了特征融合前后的对比。

2.3.1　实验参数的设置

在本研究中,提出的方法的主要参数列于表 3。深度学习中神经网络结构的选择仍然是一个巨大的挑战。目前,还没有一种成熟的理论方法来选择深度学习模型的最优结构[37]。本文在确定 DBN 架构时参考其他文献中的模型结构。NSL-KDD 数据集设置 DBN 结构为 122-110-70-30,UNSW-NB15 数据集的 DBN 结构为 196-100-80-40。输入层的单元数由样本的维度决定,输出层的单元数由入侵检测分类数量决定。

表 3　模型参数

超参数	NSL-KDD 数据集	UNSW-NB15 数据集
隐层数量(RBM 个数)	3	3
第一隐层神经元数量	110	100
第二隐层神经元数量	70	80
第三隐层神经元数量	30	40
迭代次数	100	100
学习率	0.001	0.001
动量	0.5	0.5
LPP 中的 d	13	15
LPP 中的 k	12	12

在 LPP 算法中,融合特征维数 d 和最近邻 k 的个数应预先确定。在大多数情况下,k 值根据经验设置为 12。实验结果显示随着融合特征维数 d(从 1 增加到 20)的增加,准确率不断变化。本文采用 10 折交叉验证方法,NSL-KDD 数据集对 d 的最佳选择为 13,UNSW-NB15 数据集中对 d 的最佳选择为 15。

2.3.2　二分类实验

在二分类实验中,本文尝试了各种常规机器学习算法,如 LR、J48、朴素贝叶斯(NB)、随机森林(RF)、多层感知器(MLP)、支持向量机(SVM)和其他分类算法。从表 4、表 5 中可以看出,本文提出的方法的准确率在 NSL-KDD 数据集和 UNSW-NB15 数据集上的二分类准确率分别为 85.73% 和 87.10%,而使用原始特征与 PCA 降维后的方法都低于本文所提出的方法。在 UNSW-NB15 数据集上使用 PCA 降维后的相关指标略低于使用原始特征分类,使用 LPP 方法融合特征,在准确率、精度、召回率、F1 各个方面的性能指标都大部分都高于原始特征分类性能与使用 PCA 降维后的分类性能,这充分证明了 LPP 更能够学习到高维非线性数据的低维表示,在非线性流形数据的分类方面的效果优于 PCA,PCA 更适合线性可分的情况。

表 4 NSL-KDD 二分类性能比较(%)

模型算法	原始特征				PCA 特征融合				LPP 特征融合			
	Acc	P	R	F1	Acc	P	R	F1	Acc	P	R	F1
LR	79.02	86.20	75.12	80.32	81.20	87.42	78.23	82.53	81.35	88.61	80.35	84.28
J48	49.43	53.65	82.43	64.95	61.40	72.73	51.62	60.43	73.70	83.23	67.23	74.22
NB	51.56	54.30	89.73	67.62	77.80	85.56	73.40	79.03	79.45	87.52	75.42	81.02
RF	71.40	96.32	51.83	67.32	73.80	94.93	57.03	71.34	75.82	95.87	61.34	74.81
KNN	75.20	95.82	59.16	73.12	70.40	89.20	54.73	67.82	77.63	92.46	56.92	70.46
SVM	77.20	95.50	63.05	75.92	72.43	88.83	59.03	70.92	80.52	90.41	68.23	77.76
MLP	81.21	91.02	63.27	74.52	81.89	93.10	67.10	78.02	82.37	95.36	72.81	82.57
DBN	82.28	92.56	83.45	87.76	83.29	93.84	85.42	89.43	**85.73**	95.51	87.23	91.18

表 5 UNSW-NB15 二分类性能比较(%)

模型算法	原始特征				PCA 特征融合				LPP 特征融合			
	Acc	P	R	F1	Acc	P	R	F1	Acc	P	R	F1
LR	80.62	74.81	97.53	84.72	71.62	69.82	85.92	76.80	81.42	74.92	91.05	82.20
J48	84.23	79.92	95.24	86.93	67.63	67.53	81.24	73.42	84.71	98.20	86.43	91.94
NB	55.62	**99.82**	19.42	32.53	64.03	73.35	33.05	47.82	67.32	85.42	50.07	52.66
RF	83.82	77.62	**99.42**	87.13	75.12	71.42	98.93	81.43	90.30	98.82	86.73	92.38
KNN	84.43	79.32	96.92	87.24	65.73	63.75	87.83	73.85	81.32	93.23	77.87	84.86
SVM	81.62	75.13	99.63	85.64	63.40	67.83	63.82	65.72	82.45	99.80	86.75	92.82
MLP	82.65	75.83	98.92	85.83	52.03	55.73	63.20	59.24	83.14	94.43	72.54	82.05
DBN	85.53	87.28	84.29	84.85	81.32	87.59	84.67	85.23	**87.10**	88.05	87.61	87.78

2.3.3 多分类实验

为了进一步验证本文提出的模型的优越性,使用测试数据测量了的模型在多分类中的性能,并与 LR、J48、NB、RF、KNN、SVM、MLP 等传统的机器学习算法进行了比较。从表 6 中可以看出,数据集 NSLKDD 多分类中 DBN-LPP 模型在准确率、召回率、精度和 F1-score 方面比传统的机器学习算法具有优势,U2R 和 R2L 两种数量较少的类别的检测率较低,这主要是由于在训练过程中,分类器更偏重了多类别的数据,但 DBN-LPP 模型在少数类别 U2R 和 R2L 的检测中达到了最高的检测率,分别为 19.53% 和 43.23%,而 DBN-PCA 模型在这两种类别的检测方面效果较差。这充分证明 DBN-LPP 模型在检测少数类别和未知类别攻击方面更有效率。在误报率方面,本文的模型也达到了最低的 1.98%。

表 6 NSL-KDD 多分类性能比较(%)

模型	Normal	Probe	DoS	U2R	R2L	Acc	P	R	F1	FAR
LR	92.85	71.66	81.09	0.00	2.11	75.62	92.04	62.57	74.50	7.15
J48	92.90	58.69	82.72	8.50	7.12	76.69	92.31	64.43	75.89	7.10
NB	52.52	11.90	35.20	15.50	40.38	48.75	56.09	45.90	50.48	47.48
RF	97.37	58.53	80.24	0.50	7.55	76.49	96.84	60.69	74.62	2.63
KNN	92.78	59.40	82.25	3.50	3.56	76.51	92.16	64.19	75.68	7.22
SVM	92.82	61.71	74.85	0.00	0.00	72.28	91.26	56.73	69.97	7.18
MLP	96.10	65.30	85.40	2.50	14.56	80.22	95.85	68.21	79.70	3.90
DBN	97.04	69.85	83.11	5.50	12.56	80.82	96.84	68.53	80.26	2.96
DBN-PCA	97.73	72.97	87.96	0.00	0.00	80.58	94.45	80.58	84.08	2.27
DBN-LPP	98.02	74.56	89.62	19.53	43.23	85.36	95.98	84.86	87.08	1.98

表 7 展示了本文模型在 UNSW-NB15 数据集多分类上的性能表现,从整体准确率、检测率、精度以及 F1-score 和误报率方面,DBN-LPP 模型都优于传统的机器学习算法以及 DBN-PCA 模型。各个分类器在 Analysis 和 Backdoor 的类别中检测率较低,这主要是由于这两种攻击类别的特征与 Exploits 攻击类别相似,很容易把它们误分为 Exploits。在 DoS 检测率方面,NB 模型的效果最好,达到了 70.11%,神经网络模型 MLP、DBN、DBN-PCA、DBN-LPP 等模型在 DoS 检测方面效果低于相关传统的机器学习模型。DBN-LPP 在少类别 Analysis、Backdoor、Shellcode、Worms 中也取得了较好的检测率,从表 7 中也可以看出 DBN-LPP 在各个方面都优于 DBN-PCA 模型,这也再次证明了 LPP 算法在优化特征方面好于 PCA。

表 7　UNSW-NB15 多分类性能比较(%)

类别	LR	J48	NB	RF	KNN	SVM	MLP	DBN	DBN-PCA	DBN-LPP
Normal	72.90	75.49	57.78	76.42	74.56	57.64	74.31	69.68	65.92	78.31
Generic	80.72	96.58	96.29	96.73	96.63	96.24	96.41	96.34	92.75	97.43
Exploits	89.71	75.82	42.05	76.24	74.48	74.51	86.20	87.42	80.39	74.84
Fuzzers	48.36	51.29	42.48	53.33	42.33	75.01	45.53	55.10	58.41	60.53
DoS	9.58	9.67	**70.11**	10.37	19.44	0.00	7.65	8.24	8.65	7.84
Recon	75.68	74.53	36.76	78.52	58.94	0.57	77.46	79.81	78.66	81.34
Analysis	2.45	4.89	0.00	5.17	1.48	0.00	0.59	0.00	0.00	15.76
Backdoor	4.24	10.66	0.00	11.49	2.56	0.00	8.06	0.34	0.00	21.73
Shellcode	58.42	59.78	0.00	60.85	14.47	0.00	60.32	59.26	61.32	67.84
Worms	0.00	12.81	0.00	4.55	11.11	0.00	36.36	0.00	0.00	43.52
Acc	68.24	68.73	48.08	72.45	70.38	68.91	73.95	75.77	78.41	79.82
P	76.18	82.57	52.97	83.36	82.05	73.58	82.26	79.99	80.67	85.89
R	97.23	96.21	50.92	96.46	94.01	96.27	97.28	98.90	94.62	97.65
F1	85.43	88.87	51.92	89.44	87.63	83.41	89.14	88.45	89.32	92.33
FAR	37.24	24.89	55.40	23.58	25.19	42.36	25.69	30.32	25.86	15.66

2.3.4　与其他最新入侵检测模型的对比

此外,为了展示本文所提模型的优势,将 DBN-LPP 模型与其他最新的入侵检测技术进行了比较,比较的方法与结果如表 8 所示。可以看出,本文所提出的方法在 NSL-KDD 数据集上的总体性能在准确率和误报率方面相比其他最新的方法都达到了最优,但在检测率方面略低于 S-NDAE 模型。在 UNSW-NB15 数据集上,本文模型在准确率方面达到了 79.82%,高出 EM Clustering 模型 1.35 百分点,整个比较结果充分验证了本文算法的有效性。

表 8　本文方法与其他文献方法的比较(%)

模型	数据集	Acc	DR	FAR
S-NDAE[38]	NSL-KDD(KDDTest+)	85.42	85.42	14.58
RNN[39]	NSL-KDD(KDDTest+)	81.29	69.73	26.89
CNN[40]	NSL-KDD(KDDTest+)	79.48	68.66	27.90
CVAE[41]	NSL-KDD(KDDTest+)	80.10	80.10	8.18
LSTM[42]	NSL-KDD(KDDTest+)	82.78	N/A	N/A
SHIA[43]	NSL-KDD(KDDTest+)	78.50	78.50	N/A
FL-NIDS[44]	NSL-KDD(KDDTest+)	75.13	48.71	N/A
DBN-LPP	NSL-KDD(KDDTest+)	86.36	84.86	1.98
EM Clustering[34]	UNSW-NB15	78.47	N/A	N/A
SHIA[43]	UNSW-NB15	65.10	65.10	N/A
FFDNN[45]	UNSW-NB15	77.16	N/A	N/A
FL-NIDS[44]	UNSW-NB15	73.39	40.73	N/A
DBN-LPP	UNSW-NB15	79.82	97.65	15.66

3　结　语

该方法是利用基于 DBN 框架进行特征学习结合 LPP 降维,并利用 Softmax 进行分类。实验结果表明,该模型提高了分类精度。在二分类和多分类中也表现出良好的性能。与以往的逻辑回归、J48、朴素贝叶斯、

RF、SVM 等模型以及其他文献分类方法相比,本文的方法在 NSL-KDD 和 UNSW-NB15 数据集上获得了更高的综合性能。本文方法下一步改进,使用 DBN 的多个阶段的混合特征学习模型来实现良好的特征表示,同时尝试结合更多的降维机制。此外,通过在大数据 Spark 平台中实现该系统,可以容纳更大的数据集以及减少模型的训练和测试时间。

参 考 文 献

[1] Ahmed M, Mahmood A N, Hu J K. A survey of network anomaly detection techniques[J]. Journal of Network and Computer Applications,2016,60(1):19 - 31.

[2] Reddy R, Ramadevi Y, Sunitha K, Effective discriminant function for intrusion detection using SVM[C]//International Conference on Advances in Computing, Communications and Informatics,2016(9):1148 - 1153.

[3] Li W C, Yi P, Wu Y, et al. A new intrusion detection system based on KNN classification algorithm in wireless sensor network[J]. Journal of Electrical and Computer Engineering,2014,2014:1 - 8.

[4] Amor N B, Benferhat S, Elouedi Z. Naive Bayesian networks in intrusion detection systems[C]//14th European Conference on Machine Learning,2003:11.

[5] Valdes A, Skinner K. Adaptive, model-based monitoring for cyber attack detection[C]//3rd International Workshop on Recent Advances in Intrusion Detection,2000:80 - 93.

[6] Ingre B, Yadav A. Performance analysis of NSL-KDD dataset using ANN[C]//International Conference on Signal Processing and Communication Engineering Systems,2015:92 - 96.

[7] Farnaaz N, Jabbar M A. Random forest modeling for network intrusion detection system[J]. Procedia Computer Science, 2016,89(1):213 - 217.

[8] Zhang J, Zulkernine M, Haque A. Random-forests-based network intrusion detection systems[J]. IEEE Transactions on Systems Man and Cybernetics Part C: Applications and Reviews,2008,38(5):649 - 659.

[9] Hoz E D, Hoz E M, Ortiz A, et al. Feature selection by multi-objective optimization: Application to network anomaly detection by hierarchical self-organizing maps[J]. Knowledge-Based Systems,2014,71:322 - 338.

[10] Khan J A, Jain N. A survey on intrusion detection systems and classification techniques[J]. International Journal of Engineering Research & Technology,2016,2(5):202 - 208.

[11] Buczak A L, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection[J]. IEEE Communications Surveys and Tutorials,2016,18(2): 1153 - 1176.

[12] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015,521(7553):436 - 444.

[13] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural Networks,2015,61(1):85 - 117.

[14] Liu L, Shao L, Li X L, et al. Learning spatio-temporal representations for action recognition: A genetic programming approach[J]. IEEE Transactions on Cybernetics,2016,46 (1):158 - 170.

[15] Liu A, Su Y T, Nie W Z, et al. Hierarchical clustering multi-task learning for joint human action grouping and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(1):102 - 114.

[16] Wu J X, Zhang Y, Lin W Y. Good practices for learning to recognize actions using FV and VLAD[J]. IEEE Transactions on Cybernetics,2016,46(12):2978 - 2990.

[17] Gao N, Gao L, Gao Q L, et al. An intrusion detection model based on deep belief networks[C]//2nd International Conference on Advanced Cloud & Big Data,2014:247 - 252.

[18] Alom M Z, Bontupalli V, Taha T M. Intrusion detection using deep belief networks[C]//National Aerospace and Electronics Conference,2015:339 - 344.

[19] Zhang Y, Li P, Wang X H. Intrusion detection for IoT based on improved genetic algorithm and deep belief network [J]. IEEE Access,2019,7:31711 - 31722.

[20] Yang Y Q, Zheng K F, Wu C H, et al. Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks[J]. Applied Sciences,2019,9(2):238.

[21] Zabalza J, Ren J C, Zheng J B, et al. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging[J]. Neurocomputing,2016,185:1 - 10.

[22] Ding X, He Q B, Luo N W. A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification[J]. Journal of Sound and Vibration,2015,335:367 - 383.

[23] Bengio Y. Learning deep architectures for AI[J]. Foundations & Trends in Machine Learning,2009,2(1):1 - 127.

[24] Shao H D, Jiang H K, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network[J]. Measurement Science & Technology,2015,26(11):1 - 17.

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

[26] Hinton G H. Training products of experts by minimizing contrastive divergence[J]. Neural Computation,2002:14(8): 1771 - 1800.

[27] He X F. Locality preserving projections[J]. Advances in Neural Information Processing Systems,2003,16(1):186 -

197.
[28] Tavallaee M, Bagheri E, Lu W, et al. A detailed analysis of the KDD CUP 99 data set[C]//IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009:1 - 6.
[29] Revathi S, Malathi A. A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection[J]. International Journal of Engineering Research & Technology,2013,2(11):1848 - 1853.
[30] Paulauskas N, Auskalnis J. Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset [C]//Open Conference of Electrical, Electronic and Information Sciences,2017:1 - 5.
[31] Bhattacharjee P, Fujail A, Begum S A. Intrusion detection system for NSL-KDD data set using vectorised fitness function in genetic algorithm[J]. Advanced Computer Science and Technology,2017,10(2):235 - 246.
[32] Ashfaq R A, Wang X Z, Huang J Z, et al. Fuzziness based semi-supervised learning approach for intrusion detection system[J]. Information Sciences,2017,378(2):484 - 497.
[33] Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//Military Communications and Information Systems Conference,2015:1 - 6.
[34] 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): 18 - 31.
[35] McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction[EB]. arXiv:1802.03426,2020.
[36] Shao H D, Jiang H K, Wang F, et al. An enhancement deep feature fusion method for rotating machinery fault diagnosis[J]. Knowledge-Based Systems,2017,119:200 - 220.
[37] Kim S, Choi Y, Lee M. Deep learning with support vector data description[J]. Neurocomputing,2015,165:111 - 117.
[38] Shone N, Ngoc T N, Phai V, et al. A deep learning approach to network intrusion detection[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2 (1):41 - 50.
[39] Yin C L, Zhu Y F, Fei J L, et al. A deep learning approach for intrusion detection using recurrent neural networks[J]. IEEE Access,2017,5:21954 - 21961.
[40] Wu K H, Chen Z G, Li W. A novel intrusion detection model for a massive network using convolutional neural networks[J]. IEEE Access,2018,6:50850 - 50859.
[41] Lopez-Martin M, Carro B, Sanchez-Esguevillas A, et al. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT[J]. Sensors, 2017,17(9):1967.
[42] Li Z D, Rios A L, Xu G Y, et al. Machine learning techniques for classifying network anomalies and intrusions [C]//IEEE International Symposium on Circuits and Systems,2019:1 - 5.
[43] Vinayakumar R, Alazab M, Soman K P, et al. Deep learning approach for intelligent intrusion detection system[J]. IEEE Access,2019,7:41525 - 41550.
[44] Mulyanto M, Faisal M, Prakosa S W, et al. Effectiveness of focal loss for minority classification in network intrusion detection systems[J]. Symmetry,2021,13(1):4.
[45] Kasongo S M, Sun Y X. A deep learning method with wrapper based feature extraction for wireless intrusion detection system[J]. Computers & Security,2020,92:101752.

(上接第 28 页)

[4] Snow P, Deery B, Lu J, et al. Factom business processes secured by immutable audit trails on the blockchain[R]. Factom,2018.
[5] Kleinaki A S, Mytis-Gkometh P, Drosatos G, et al. A blockchain-based notarization service for biomedical knowledge retrieval[J]. Computational and Structural Biotechnology Journal,2018,16:288 - 297.
[6] 北京市方圆公证处课题组."区块链"在公证实践中的应用[J].中国公证,2019(4):54 - 55.
[7] 林晔晗,潘玲娜,刘文添.依托区块链技术打造智慧信用生态系统-广州互联网法院"网通法链"上线[EB/OL]. [2021 - 01 - 19]. https://www.chinacourt.org/article/detail/2019/03/id/3808242.shtml.
[8] 邵奇峰,金澈清,张召,等.区块链技术:架构及进展[J].计算机学报,2018,41(5):969 - 988.
[9] 于戈,聂铁铮,李晓华,等.区块链系统中的分布式数据管理技术-挑战与展望[J].计算机学报,2021,44(1): 28 - 54.
[10] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [EB/OL]. [2021 - 01 - 19]. https://bitcoin.org/bitcoin.pdf.
[11] Androulaki E, Barger A, Bortnikov V, et al. Hyperledger fabric: A distributed operating system for permissioned blockchains[C]//13th EuroSys Conference,2018:1 - 15.
[12] 曾诗钦,霍如,黄韬,等.区块链技术研究综述:原理、进展与应用[J].通信学报,2020,41(1):134 - 151.
[13] 钱卫宁,邵奇峰,朱燕超,等.区块链与可信数据管理:问题与方法[J].软件学报,2018,29(1):150 - 159.
[14] Rowstron A, Druschel P. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems[C]//IFIP/ACM International Conference on Distributed Systems Platforms and Open Distributed Processing, 2001:329 - 350.