DOI: 10.3969/j.issn.2096 - 8299.2021.06.014

面向入侵检测的机器学习方法综述

王亮亮，谷朝阳

(上海电力大学 计算机科学与技术学院，上海 200090)

摘 要: 入侵检测方法是基于网络的入侵检测系统的核心，可以是基于特征的，也可以是基于异常的。基于特征的检测方法具有较高的检测率，但不能检测到未知新型攻击；基于异常的检测方法可以检测到新型攻击，但误报率较高。为了降低入侵检测的误报率并提高其检测率，许多机器学习技术被应用到入侵检测系统中。通过对大量带有入侵数据训练样本的学习，构建了一个用于区分正常状态和入侵状态的入侵检测模型。针对目前入侵检测系统存在的高误报率、低检测速度和低检测率等问题，对机器学习技术在入侵检测系统中的的优势、系统检测的通用数据集以及系统评估指标进行了详细阐述，并对未来研究趋势进行了展望。

关键词: 入侵检测系统；机器学习；数据集；评估指标

中图分类号: TP309.1 **文献标志码**: A **文章编号**: 2096 - 8299(2021)06 - 0591 - 06

Overview of Machine Learning Methods for Intrusion Detection

WANG Liangliang, GU Zhaoyang

(*School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai* 200090, *China*)

Abstract: The intrusion detection method is the core of the network-based intrusion detection system. It can be feature-based or anomaly-based. Feature-based methods have higher detection rates, but cannot detect unknown new types of attacks. Anomaly-based detection methods can detect new types of attacks, but the false positive rate is high. In order to reduce the false alarm rate and improve the detection rate, many machine learning techniques are applied to intrusion detection systems. It learns a large number of training samples with intrusion data to build an intrusion detection model that distinguishes between normal and intrusion states. This article describes the advantages of machine learning technology in intrusion detection systems, the general data set for system detection, and the system evaluation indicators and future prospects for high intrusion detection rates, low detection speeds, and low detection rates in current intrusion detection systems.

Key words: intrusion detection systems; machine learning; data set; evaluation indicators

入侵检测系统(Intrusion Detection System, IDS)是一种对网络传输进行即时监视，在发现可疑传输时发出警报或者采取主动反应措施的网络安全设备。与其他网络安全设备不同，IDS 是一种积极主动的安全防护技术，重在对网络、系统的运行状况进行监视，尽可能发现各种攻击企图、攻

收稿日期: 2020-02-28

通信作者简介: 王亮亮(1984—)，男，博士，讲师。主要研究方向为密码学与信息安全。E-mail: llwang@shiep.edu.cn。

基金项目: 国家自然科学基金(61802249, U1936213)；上海高校青年教师培养资助计划(ZZsdl18006)。

击行为或攻击结果,以保证网络系统资源的机密性、完整性和可用性。理想的 IDS 应该具有高准确率、低误报率、低计算成本等特点[1]。IDS 中有误用(签名)和异常(行为)两种检测方法[2]。在误用检测中,系统存储已知的攻击特征,并在网络流量中寻找这些特征,如果有匹配的,则认为是攻击。基于误用的 IDS 使用包含攻击特征的数据库来检测数据中的入侵,具有良好的检测率[3],且可以检测到误报率较低的攻击,但无法检测到没有定义特征的新攻击[4]。基于异常的 IDS 通常通过在局域网中记录正常活动流量来建立模型,一旦系统监测到与该模型的任何偏差,就会将其视为异常或攻击。但由于新攻击不存在特征码,因此该检测方案无法检测新的攻击[5]。基于异常的检测技术寻找异常行为,这也是它能够检测新攻击的原因[6]。总之,IDS 为系统或网络的检测正常行为设定阈值,通过吸收系统或网络在学习阶段的正常行为来完成检测,任何违反某个阈值的进程都被视为可能的入侵。然而,基于网络的检测技术由于很难定义系统的正常行为而面临误报问题。为了解决这一问题,许多机器学习技术在 IDS 上得到了应用。

1 机器学习

本文主要介绍监督学习、无监督学习和深度学习在入侵检测系统中的应用。监督学习包括支持向量机(Support Vector Machine,SVM)、朴素贝叶斯分类器和决策树等。SVM 通过寻找一个超平面对样本进行分类,要求间隔最大化[7],$w \cdot x_i + b$ 即为分离超平面。假设给定一个特征空间上的训练数据集 $T = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_N, y_N)\}$,$x_i$ 为第 i 个特征向量,y_i 为类标签,$i = 1, 2, 3, \cdots, N$。实际情况下几乎不存在完全线性可分的数据。为了解决这个问题,引入“软间隔”概念,即允许某些点不满足约束。采用 hinge 损失,优化问题为

$$\min_{w,b,\xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \tag{1}$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant 1 - \xi_i \qquad \xi_i \geqslant 0$$

式中:ξ_i——松弛变量,$\xi_i = \max(0, 1 - y_i(w \cdot x_i + b))$,即一个 hinge 损失函数;

C——惩罚参数,$C > 0$,C 值越大,对分类的惩罚越大。

然后用拉格朗日乘子法得到拉格朗日函数,再求其对偶问题。

朴素贝叶斯通过估计每个类别出现的概率以及每个类别条件下每个属性出现的概率,使用最大似然估计法获得某样本属于某类别的概率值,最后选择最大概率作为推测结果。

决策树通过信息增益方法对样本特征进行选择生成决策树,再通过修剪决策树防止过拟合。决策树具有高效的性能,常用 ID3,C4.5,CART 等不同的算法生成决策树。决策树算法描述如表 1 所示。

表 1 决策树算法描述

算法	算法描述
ID3 算法	在决策树的各级节点上,使用信息增益作为属性的选择标准。确定每个节点的保留属性
C4.5 算法	在决策树的各级节点上,使用信息增益作为属性的选择标准。确定每个节点的保留属性。技能处理离散属性和连接属性
CART 算法	是一种非常有效的十参数分类和回归算法。通过构建树、修建树、评估树来构建二义树

无监督学习包括 K-means 聚类和主成分分析法(Principal Component Analysis,PCA)[8]。K-means 算法是一种聚类算法,目的是将相似的样本划分到同一簇中。具体来说,首先,在样本中随机选取 k 个样本点作为各个簇的中心点,计算所有样本点与各个簇中心之间的距离;然后,根据计算距离把各个样本点划入最近的簇中,再根据簇中已有的样本点,重新计算簇中心;最后,重复以上操作直至达到停止条件。PCA 利用线性代数进行数据降维,将多个变量转换为少数几个不相关的综合变量来比较全面地反映整个数据集。这些综合变量称为主成分,各主成分之间不相关。PCA 主要通过奇异值分解或特征值分解方法计算协方差矩阵的特征值与特征向量,然后选择其中最大的 k 个特征值,将其对应的 k 个特征向量分别作为列向量组成特征向量矩阵,将数据转换到 k 个特征向量构建的新空间中完成数据降维。

深度学习是机器学习研究中的一个新领域。通过建立、模拟人脑分析学习的神经网络,模仿人脑机制来解释数据,主要方法有循环神经网络(Recurrent Neural Network,RNN)、卷积神经网络(Convolutional Neural Networks,CNN)、长短期记忆网络(Long Short-Term Memory,LSTM)等。

RNN 是一个序列到序列的模型,不仅考虑当前时刻的输入,还考虑对前面内容的记忆。其模型如图 1 所示。其中,x_t表示 t 时刻的输入,o_t表示 t 时刻的输出,s_t表示 t 时刻的记忆。权值共享,图 1 中的 $\boldsymbol{u}$ 是输入层到隐藏层的权重矩阵,$\boldsymbol{W}$ 是隐藏层上一次的值作为这次输入的权重矩阵,$\boldsymbol{V}$ 是隐藏层到输出层的权重矩阵。每一个输入值都只与它本身的那条路线建立权连接,不会与别的神经元连接。

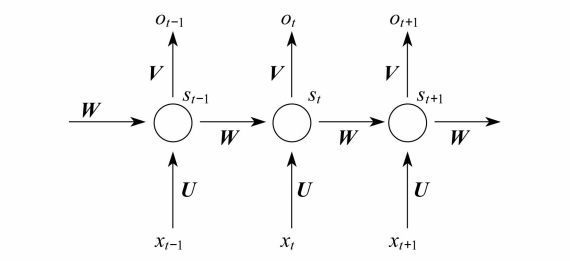

图 1 RNN 模型

CNN 主要包含 3 层,分别为卷积层、池化层和全连接层。卷积层是通过卷积核提取特征,再通过池化层降维,最后通过全连接层分类。因有用信息与需要处理信息距离较远而导致的 RNN "长依赖"问题,需要通过 LSTM 来解决。LSTM 由 3 个门来控制细胞状态,分别为忘记门、输入门和输出门。忘记门决定细胞状态丢弃部分信息,输入门决定给细胞状态添加新信息,输出门决定输出的细胞信息。

监督学习在未知攻击中表现不好,但在已知攻击中表现良好,误报率较低。相反,无监督学习能够检测未知攻击,但误报率较高。深度学习自动提取高级特征,为后期分类做好预备工作,但随着网络深度的增加,会大大增加训练和测试时间。机器学习用于入侵检测的简单流程如图 2 所示。

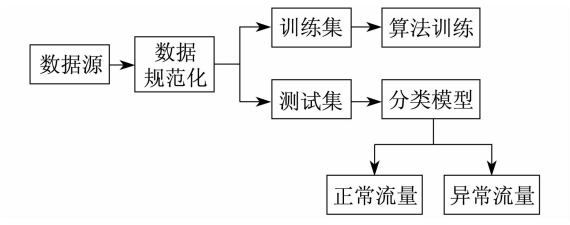

图 2 机器学习用于入侵检测的简单流程

1.1 监督学习

监督学习重在建立不仅能够基于数值特征区分至少两个类别,而且对于新的未观察到的样本误差最小的模型[9]。为了建立该模型,分类器需要一个包含正常样本和攻击样本的标记训练数据集。由于监督学习能为分类器提供更多的信息,理论上有助于提高检测率。然而,监督学习也存在一些问题,如不能保证标签准确,且如果训练数据集包含噪声,则会造成较高的误报率[10]。

文献[5]提出了一种高速决策的实时入侵检测系统。该系统使用基于决策树的 C4.5 分类模型,使用正向选择排序(Forword Selection Ranking,FSR) 和后向淘汰排序(Backwared Elimination Ranking,BER) 技术从 KDD 99 入侵数据集中的 41 个特征中选择了 9 个最佳特征。文献[11]提出了一种新的基于决策树的雾计算入侵检测系统。该系统由于雾节点产生的数据量大,计算能力有限,因此安全性存在问题。研究者根据分类和回归树产生最大信息增益的特征构造二叉树,利用 KDD CUP 99 数据集对系统进行了评估。结果表明,该决策树具有最佳的检测精度、攻击覆盖率和可接受的检测时间。文献[12]提出了一种基于 SVM 的异常流量检测算法和一种计算数据特征归一化熵算法。具体来讲,在受到攻击的情况下,由于数值偏离正常值,因此可以设置阈值以发现网络中的任何异常行为; 利用 SVM 建立分类器检测网络异常流量; 为了提高 SVM 参数的质量,采用粒子群优化算法对参数进行更精确的估计。采用 DARPA 和 KDD CUP 99 数据集对该方法进行了性能评价。结果表明,该方法能准确地检测出不同类型的攻击。文献[13]通过 LSTM 学习网络数据的特征和模式,将其分为良性或攻击性两种类别。作为一种深度学习算法,一方面 LSTM 通过对原始数据进行处理,降低了传统机器学习的特征工程负担; 另一方面,通过将网络数据包的序列输入到能够长期记忆序列的深度算法中,LSTM 可以用来识别与窗口大小无关的长数据包序列中攻击模式是否重复。文献[14]使用了最真实的入侵检测数据集 NSL-KDD,通过从训练集和测试集中消除冗余,创建了两个具有挑战性的测试集。同时,结合多个基础学习器构建集成学习器,包括决策树、贝叶斯和 KNN 等,提高了检测的准确性。文献[15]使用 SVM 作为分类器,提高了入侵检测的效率,但仍存在受大数据集影响和训练时间过长而导致系统崩溃的问题。为了最大限度地提高单个特征提取算法的效率,研究者开发了一个高效的入侵检测系统,实现了

线性判别分析(Linear Discviminant Analysis, LDA) 和 PCA 特征提取算法的集成。

1.2 无监督学习

无监督学习不使用标记数据,而使用统计模型,在没有任何先验知识的情况下,基于两个假设函数将数据划分为正常和异常[16]。

文献[17]基于会话的数据预处理模块从数据包的报头部分提取头特征,并在会话中选择网络应用层的有效负载一起形成记录。随后,这些记录被输入到叠加去噪自编码器中,以获得对正常和恶意网络流量进行分类的基本特征。文献[18]将训练过程分为无监督预训练和监督微调两种。在无监督预训练过程中,扩张卷积自编码(DCAEs) ,从大量未标记样本中学习特征。然后,利用反向传播算法和少量标记样本进行监督微调,以增强从未标记数据中学习到的表示。使用不同的原始网络流量和无监督的预训练使模型更具适应性和灵活性,但训练过程耗时较长。文献[19]提出了一种新的多线程 K-均值方法,将聚类技术应用于异常检测中。

1.3 深度学习

文献[20]利用不同的降维方法去除网络流量数据中的冗余和不相关特征,利用 CNN 自动提取降维数据的特征。为了降低计算成本,研究者将原始的流量矢量格式转换为图像格式,并使用 KDD CUP 99 数据集对所提出的 CNN 模型的性能进行了评估。结果表明,CNN-IDS 模型的时效性高于传统算法。文献[21]提出了一种使用递归神经网络(RNN-IDS) 进行入侵检测的深度学习方法,研究了该模型在二元分类和多分类中的性能,以及神经元的数量和不同的学习率对所提出模型的性能影响,并将其与人工神经网络、随机森林、支持向量机以及其他机器学习方法进行了比较。结果表明,RNN-IDS 在二元分类和多分类中的性能均优于传统的机器学习。文献[22]提出了一种 LSTM 用于雾到物通信中的分布式网络攻击检测,证明了深度模型的有效性和高效率。文献[23]针对传统 BP 神经网络在检测速度、精度、复杂度等方面的缺陷,提出了一种基于 DBN 的网络入侵检测算法,将数据通过双层 RBM 结构降维,再用 BP 神经网络反向微调结构参数,从而简化了数据的复杂度,减少了 BP 神经网络的计算量。

2 数据集

具有代表性的数据集是评价和比较不同网络入侵检测系统质量的良好基础。本文主要介绍 3 种具有代表性的数据集的由来、检测攻击类型和特征数量。

2.1 KDD CUP 99

KDD CUP 99[24] 数据集是为“第三届国际知识发现和数据挖掘工具竞赛”而开发的。该竞赛与“第五届国际知识发现和数据挖掘大会”同时举行。完整的数据集有 400 万个数据点、42 个特征和 4 种攻击类型。

2.2 UNSW-NB15

澳大利亚网络安全中心(ACCS) 的网络范围实验室设计了基于网络的 UNSW NB 15 数据集。该数据集有超过 250 万条记录,49 个特征。存在模糊攻击、分析攻击、后门攻击、拒绝服务攻击、漏洞攻击、通用攻击、侦察攻击、外壳代码攻击和蠕虫攻击等 9 类攻击。

2.3 NSL-KDD

为了使机器学习算法在 KDD CUP 99 上更好地工作,研究者通过删除重复记录来减小数据量,创建了 NSL-KDD 数据集。该数据集在包含 KDD CUP 99 数据集基本记录和数据特性的基础上,具有以下特点: 由于训练集中没有冗余数据,分类器不会给出偏差结果,使得检测率更为准确; 从每个困难级别组中选择的记录数与 KDD 数据集中记录的百分比成正比[25]。

3 评估指标

目前,评价入侵检测系统性能的指标主要有以下 4 个。

(1) 召回率(查全率) ,针对原有样本而言,表示样本中的正例有多少被预测正确了。召回率存在两种可能: 一是把原来的正类预测为正类(TP) ; 二是把原来的正类预测为负类(FN) 。其公式为

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

(2) 精确率(查准率) ,针对预测结果而言,表示预测为正的样本中有多少是真正的正样本。

预测为正存在两种可能: 一是把正类预测为正类(TP); 二是把负类预测为正类(FP)。其公式为

$$P = \frac{TP}{TP + FP} \quad (3)$$

(3) AUC(Area Under Curve)针对样本分布不均匀问题,同时考虑了分类器对于正类和负类的分类能力,在样本不平衡的情况下,依然能够对分类器作出合理的评价。

(4) F_1 值,综合召回率和精确度的调和值。召回率和精确度都集中在阳性样本上,但它们都不能反映出模型处理阴性样本的能力。

4 研究展望

现有的入侵检测系统解决方案面临着检测率不均衡、检测精度低、实时高速网络异常检测困难等问题。针对检测率低这一问题,混合机器学习算法通过结合多种机器学习技术予以改善,如级联不同的分类器以提高系统的性能和精度[26]。基于三角形区域的最近邻(TANN)[27]是一种使用无监督和有监督学习技术的混合学习。该学习首先利用 K-means 聚类得到攻击类簇的中心; 然后通过计算两个中心之间的三角形区域来创建数据的新特征签名; 继而 KNN 分类器利用新特征对分类攻击进行改进。该学习继承了有监督和无监督学习的优点,具有良好的性能和无标记能力。但存在计算复杂和时间消耗等问题。针对训练集中很多类型样本个数少而造成的检测率不均衡问题,文献[28]使用单侧选择减少多数类别的噪声样本,再通过合成少数样本过采样技术(Synthetic Minority Over-sampling Technique, SMOTE)增加样本数量,建立了一个平衡的数据集,使模型充分学习少数样本的特征,减少了模型训练的时间; 并通过使用 CNN 提取空间特征,使用双向长期短期记忆(BiLSTM)提取时间特征,从而形成了一个深层次的网络模型。文献[29]针对某些恶意样本提出了一种基于元学习框架和 Few-shot 的检测方法。该方法可用于区分和比较包括正常未受影响样本和恶意样本的网络流量样本,能够在未经训练的数据集中,基于学习到的先验知识,使用有限数量的标签检测出新类型样本。

总之,使用机器学习技术进行分类的缺点是对手有可能试图绕过分类器进行攻击,研究这类攻击的领域称为“对抗性机器学习”,该学习在图像分类和垃圾邮件检测等领域得到了广泛的探索,但在其他领域,如入侵检测等方面的探索很少,具有广阔的发展前景与扩展空间。

5 结 语

本文梳理了近年来机器学习在入侵检测系统中的相关研究: 首先简要阐述了不同机器学习方法在入侵检测系统上的应用; 然后介绍了常用数据集和分类器评估指标的具体内容; 最后,对目前存在的问题和未来的研究趋势进行了分析和展望。

参考文献:

[1] ILLY P, KADDOUM G, MOREIRA C M, et al. Securing Fog-to-Things environment using intrusion detection system based on ensemble learning [C]//IEEE Wireless Communications and Networking Conference (WC-NC). Valencia: IEEE, 2019: 32-41.

[2] DAVIS J J, CLARK A J. Data preprocessing for anomaly based network intrusion detection: a review [J]. Computers & Security, 2011, 30(6/7): 353-375.

[3] DANESHPAZHOUH A, SAMI A. Entropy-based outlier detection using semi-supervised approach with few positiveexamples [J]. Pattern Recognition Letters, 2014(49): 77-84.

[4] NISIOTI A, MYLONASA, YOO P D, et al. From intrusion detection to attacker attribution: a comprehensive survey of unsupervised methods [J]. Communications Surveys & Tutorials, 2018, 20(4): 3369-3388.

[5] RATHORE M M, SAEED F, REHMAN A, et al. Intrusion detection using decision tree model in high-speed environment [C]//International Conference on Soft-computing & Network Security. Tami Nadu, 2018: 2361-2382.

[6] DIRO A A, CHILAMKURTI N, KUMAR N. Lightweight cybersecurity schemes using elliptic curve cryptography in publish-subscribe fog computing [J]. Mobile networks & applications, 2017, 22(5): 848-858.

[7] SHIJOE J, MALATHI D, BHARATH R, et al. A survey on anomaly based host intrusion detection system [J]. Journal of Physics Conference Series, 2018(5): 12000-12021.

[8] BHARTI K, JAIN S, SHUKLA S. Fuzzy K-mean clustering via J48 for intrusion detection system [J]. International Journal of Computer Science and Technologies, 2010(4): 318-351.

[9] ANDRÉ OBERTHÜR, WARNATP. Supervised classification [J]. Encyclopedia of Cancer, 2014(10): 62-68.

[10] JIANG S Y, SONG X, WANG H, et al. A clustering-based method for unsupervised intrusion detections [J]. Pattern recognition letters, 2006, 27(7): 802-810.

[11] KAI P, LEUNG V C M, ZHENG L X, et al. Intrusion detection system based on decision tree over big data in fog envi-

ronment [J]. Wireless Communications & Mobile Computing,2018(5):1-10.

[12] NAILA B A,MOHAMED G. A genetic clustering technique for anomaly-based intrusion detection systems [C]//2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Busan,2015:65-70.

[13] DIRO A,CHILAMKURTI N. Leveraging LSTM networks for attack detection in Fog-to-Things communications [J]. IEEE Communications Magazine,2018,56(9):124-130.

[14] ILLY P, KADDOUM G, MOREIRA C M, et al. Securing Fog-to-Things environment using intrusion detection system based on ensemble learning [C]//IEEE Wireless Communications and Networking Conference (WCNC). Marrakesh: IEEE,2019:362-372.

[15] 张瑞霞,王勇. 融合 PCA 和 LDA 的入侵检测算法 [J]. 计算机技术与发展,2009,19(11):132-134.

[16] ABUROMMAN A A, REAZ M B I. Ensemble of binary SVM classifiers based on PCA and LDA feature extraction for intrusion detection [C]//2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). Xi'an, China: IEEE,2016:636-640.

[17] YU Y,LONG J,CAI Z. Session-based network intrusion detection using a deep learning architecture [C]//14th International Conference,Kiakyushu,Japen,2017:216-222.

[18] YU Y, LONG J, CAI Z P. Network intrusion detection through stacking dilated convolutional autoencoders [J]. Security and Communication Networks,2017(11):1-10.

[19] PATHAK V, ANANTHANARAYANAV S. A novel multi-threaded K-means clustering approach for intrusion detection [C]//IEEE International Conference on Software Engineering & Service Science. Suzhou, China: IEEE,2012:757-760.

[20] XIAO Y,XING C,ZHANG T, et al. An intrusion detection model based on feature reduction and convolutional neural networks [J]. IEEE Access,2019(17):42210-42219.

[21] 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(9):1-3.

[22] MEENA G, CHOUDHARY R R. A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA [C]//2017 International Conference on Computer, Communications and Electronics (Comptelix). Jaipur, India: IEEE, 2017:553-558.

[23] 徐东辉,王勇,樊汝森. 一种基于 DBN 的网络入侵检测算法 [J]. 上海电力学院学报,2013,29(6):589-592.

[24] ZHAO J,ZHU Y C. Research on intrusion detection method based on SOM neural network in cloud environment [J]. Computer Science and Application,2016,6(8):135-146.

[25] KARATAS G,DEMIR O,SAHINGOZ O K. Deep learning in intrusion detection systems [C]//International Congress on Big Data. San Francisco,USA,2018:321-331.

[26] ELTANBOULY S, BASHENDY M. Machine learning techniques for network anomaly detection: a survey [C]//2020 IEEE International Conference on Informatics,IoT,and Enabling Technologies (ICIoT). Doha,Qatar,2020:156-162.

[27] TSAI C F,LIN C Y. A triangle area based nearest neighbors approach to intrusion detection [J]. Pattern Recognition, 2010,43(1):222-229.

[28] JIANG K,WANG W,WANG A,et al. Network intrusion detection combined hybrid sampling with deep hierarchical network [J]. IEEE Access,2020(9):32464-32476.

[29] SANKARANARAYANAN S, JAIN A, CHELLAPPA R, et al. Regularizing deep networks using efficient layerwise adversarial training [C]//National Conference on Artificial Intelligence. Sri Lanka,2018:1-9.

(责任编辑 谢 冉)