文章编号:1671-4229(2021)02-0001-12

人工智能赋能网络安全应用

王郁夫[a], 李沛辰[b], 易 波[a], 王兴伟[a*]
(东北大学 a. 计算机科学与工程学院; b. 软件学院, 辽宁 沈阳 110169)

摘 要: 网络安全是一种组织和国家为保护网络空间中使用数据和资产的机密性、完整性及可用性而遵循的与安全风险管理流程相关的方法和行动. 文章针对网络安全中的常见问题,对近年来基于人工智能技术的网络安全解决方案和带来的赋能效果进行系统性的总结,分析了近5年来70余篇论文中的研究工作,介绍了人工智能技术中的决策树、贝叶斯、聚类、支持向量机、卷积神经网络(CNN)、循环神经网络(RNN)、生成对抗网络(GAN)和强化学习(RL)在网络入侵、恶意软件、钓鱼网站以及垃圾邮件4种常见安全问题中的应用,并提出了人工智能在网络安全应用中存在的潜在问题与挑战. 总体而言,人工智能技术在实际应用中提高了网络安全应用程序的准确性、可扩展性、可靠性和性能,在未来的网络安全技术发展中定会成为赋能的重要引擎动力.

关键词: 人工智能; 网络安全; 机器学习; 强化学习
中图分类号: TN 915.08 **文献标志码**: A

Artificial intelligence empowers cybersecurity applications

WANG Yu-fu[a], *LI Pei-chen*[b], *YI Bo*[a], *WANG Xing-wei*[a*]
(a. School of Computer Science and Engineering; b. Software College, Northeastern University, Shenyang 110169, China)

Abstract: Cybersecurity is a method and action related to the security risk management process followed by organizations and countries to protect the confidentiality, integrity and availability of data and assets used in cyberspace. Aiming at the common problems in cybersecurity, this paper systematically summarizes the solutions based on artificial intelligence technology and the empowerment effects brought by them in recent years. We analyze the research work from more than 70 papers in the past five years, and introduce the use of decision trees, Bayes, clustering, support vector machines (SVM), convolutional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN) and reinforcement learning (RL) in four common cybersecurity problems including network intrusion, malware, phishing websites and spam. At the end of the paper, the potential problems and challenges in the application process of artificial intelligence in cybersecurity are given. In general, artificial intelligence technology has improved the accuracy, scalability, reliability and performance of cybersecurity in practical applications, and will surely become an important engine for empowerment in the future development of cybersecurity technology.

Key words: artificial intelligence; cybersecurity; machine learning; reinforcement learning

基金项目: 国家自然科学基金资助项目(61872073, 62032013);辽宁省"兴辽英才计划"资助项目(XLYC1902010)
作者简介: 王郁夫(1995—),男,硕士研究生. E-mail:wyfhj95@163.com
* 通信作者. E-mail:wangxw@mail.neu.edu.cn
引文格式: 王郁夫, 李沛辰, 易波, 等. 人工智能赋能网络安全应用[J]. 广州大学学报(自然科学版),2021,20(2):1-12.

1 研究背景

1969年以来,互联网伴随着人类已经发展了半个多世纪,从最初的军用网络ARPANET到如今的万维网,互联网已经渗透到了我们生活中的方方面面.据《Digital 2021 Global Overview Report》[1]统计,全球互联网使用人数已经达到46.6亿,普及率达到59.5%.在生活中,人们只需要打开自己的网络设备,敲动指尖点击屏幕,就随时可以尽情享受从沟通交流到衣食住行的全方位服务.不仅如此,自从20世纪以来,互联网领导的信息技术革命正在不断推动世界经济的快速发展.以中国的经济水平为例,根据《中国数字经济发展白皮书(2020)》[2]统计,在2005年,我国数字经济占GDP仅为14.2%,但2019年,这一数字上升至36.2%,数字经济已经成为国家经济发展的最大动力.同时,随着2020年以来新冠肺炎疫情在全球范围的爆发,互联网支持的"数字经济"更是成为对冲疫情的影响、重塑经济结构体系、增强治理水平的重要推动力.我国最新发布的《世界互联网发展报告2020》[3]指出,在新冠疫情冲击全球经济社会发展的大环境下,数字经济被视为全球经济复苏新引擎.在未来,世界各国应该大力推进以5G、人工智能、物联网等为代表的信息基础设施建设,数字技术的快速发展,带动了产业深度融合.

随着互联网衍生的数字经济的高速发展,网络安全逐渐成为现代社会万物互联和技术发展过程中愈发明显的治理难题,也逐渐受到世界各国的重视.网络安全不只是通信行业的问题,已经逐渐辐射到社会、经济、军事等更为重要的领域.尽管世界各国都在不断强化网络安全技术,增大网络安全领域的投入力度,但网络安全技术的发展仍旧落后于恶意使用网络技术的步伐,网络威胁的发生频率、恶劣影响和防护的复杂性都在不断升级.令人激动的是,近年来,随着人工智能的发展,人工智能技术在网络安全的应用上给人们提供了一种新的解决网络安全威胁的可行方法.人工智能技术拥有类人的逻辑能力,能够使机器实现对物理世界的认知并实现自主决策,其内在逻辑是通过数据输入理解世界,或通过传感器感知环境,然后运用模式识别实现数据的分类、聚类、回归等分析,并据此做出最优的决策推荐.进一步地,当人工智能运用到网络安全领域时,机器自动化和机器学习等技术能有效且高效地帮助人类预测、感知和识别安全风险,快速检测定位危险来源,分析安全问题产生的原因和危害方式,综合智慧大脑的知识库判断并选择最优策略,采取缓解措施或抵抗威胁,甚至提供进一步缓解和修复的建议.这个过程不仅将人们从繁重、耗时、复杂的任务中解放出来,面对不断变化的风险环境、异常的攻击威胁形态比人更快、更准确,综合分析的灵活性和效率也更高.

因此,为了推进网络安全技术研究的发展,本文主要总结人工智能技术在网络安全问题中的应用所带来的赋能效果,重点介绍了网络安全领域使用的机器学习和深度学习方法及其描述,旨在帮助那些希望开始研究机器学习和深度学习技术应用于网络安全领域的研究者们进行总体的调研.本文中,首先对现今网络环境中的安全问题进行介绍并按照其特点进行基础分类.进而,介绍了人工智能领域的主流技术和模型,并结合上述安全问题的分类举例说明这些技术在网络安全中的应用.在调研中,主要统计了近5年该领域中的研究工作,确保在每一个网络安全分类介绍中都包含机器学习或深度学习中各种基于主流模型衍生来的优秀的研究工作.最后,对现阶段人工智能技术在网络安全中的应用给出了总结,并提出了其在未来的网络安全应用中的风险与挑战.

2 网络安全定义与分类

2.1 网络安全定义

近年来,用于讨论数字设备及信息安全性方面的术语发生了很大的变化.21世纪初,在这种语境下经常使用的术语是计算机安全、IT安全或信息安全.然而,随着时间的推移,网络安全这个新的术语开始变得越来越流行.搜索词"计算机安全和信息安全"的数量在稳步下降,而和"网络安全"有关的各种变体正在超越它们.

虽然网络安全是一个被广泛使用的术语,但是其定义变化很大,主观性较强.到目前为止,网络安全并没有一个通用的、被普遍接受的定义.文献[4]通过研究现有的、由权威提供的网络安全定义,基于各种词汇和语义分析技术,试图更好地理解这些定义的范围、语境以及相关性.最终,基于所进行的分析,提出了一个改进的更具代表性的定义:组织和国家为保护网络空间中使用的数据和资产的机密性、完整性和可用性而遵循的与安全风险管理流程相关的方法和行动.该概念包括指导方针、政策、保障措施、技术、工具以及培训的集合,为网络环境及其用户的状态提供最佳保护.

2.2 网络安全分类

网络安全是一个庞大的研究领域,涉及到方方面面

的技术,通过结合近年来对人工智能和网络安全的相关研究文献[5-8],综合考虑其研究结果,以及本文在 Web of Science、Google Scholar、知网等平台上的检索统计结果,选出了 4 个最受关注的研究方向,分别是网络入侵、恶意软件、网络钓鱼和垃圾邮件. 下面对这些方向进行简要的介绍.

2.2.1　网络入侵

网络入侵是指任何未经授权的访问、操纵、修改或破坏信息的尝试,或远程使用计算机系统发送垃圾邮件、进行黑客攻击或修改其他计算机的行为. 入侵检测系统(IDS)智能地监视计算资源中发生的活动,例如网络流量和计算机使用情况,以分析事件并生成应对措施. IDS 通常监视和分析用户和系统活动,访问系统和数据的完整性,识别恶意活动模式,对入侵产生反应,并报告检测结果.

根据检测原理,文献[9]将网络入侵检测分为以下 3 个方面:误用/签名检测、异常检测和混合检测等.

(1)误用/签名检测

误用检测又称签名检测,是一种已知网络误用发生时产生警告的入侵检测方法. 签名检测技术度量输入事件和已知入侵签名之间的相似性. 它标记与预定义的入侵模式有相似之处的行为. 因此,已知的攻击类型可以立即被检测到,但是签名检测不能检测新的攻击.

(2)异常检测

当被检测对象的行为与预定义的正常模式有显著差异时,异常检测将触发警告. 因此,异常检测技术被设计用于检测与预期的正常模型相偏离的行为. 在网络安全中,异常检测包括检测恶意活动,例如渗透和拒绝服务. 该方法通常包括训练和检测两个步骤. 在训练步骤中,机器学习技术用于在没有攻击的情况下生成正常模式的描述;在检测步骤中,如果事件记录明显偏离正常的模式,则将输入事件标记为攻击[10].

(3)混合检测

大多数 IDS 要么采用误用检测技术,要么采用异常检测技术. 这两种方法都存在缺陷:误用检测技术缺乏检测未知入侵的能力;异常检测技术通常产生很高的虚报率. 为了改进入侵检测技术,研究人员提出了混合检测技术,将异常检测和误用检测技术结合在入侵检测中.

2.2.2　恶意软件

恶意软件是一种通过传播渗透到计算机系统,破坏其安全性、完整性和功能性的软件. 不同类型的恶意软件包括病毒、蠕虫、木马、后门、间谍软件、僵尸网络等. 随着互联网用户的日益普及,恶意软件对计算机系统的安全构成了严重威胁[11-12].

一般来说,恶意软件检测技术被分为 3 类:静态的、动态的和混合的. 静态方法分解和分析源代码而不执行它. 虽然速度很快,但是会产生很高的假阳性率. 此外,无法检测到混淆的恶意软件. 动态分析技术在监视虚拟环境中执行代码相互作用的同时,消耗了大量的时间和内存资源,而混合方法则利用了静态和动态方法的优点.

2.2.3　网络钓鱼

在网络安全领域,钓鱼是一种犯罪性欺诈过程,通过在电子通信中伪装成一个可信赖的实体,试图获取敏感信息,如用户名、密码和信用卡信息. 网络钓鱼一般是通过电子邮件或即时通讯进行的,通常会让用户在一个外观和感觉几乎与合法网站相同的虚假网站上输入详细信息. 网络钓鱼是社会工程技术的一个例子,利用当前网络安全技术的低可用性来欺骗用户.

2.2.4　垃圾邮件

垃圾邮件是指未经请求就通过电子邮件大量发送的信息. 大多数垃圾邮件本质上是商业性的. 但是,无论其是否商业化,垃圾邮件中的许多网站不仅令人厌烦,而且还很危险,因为它们可能包含链接,导向钓鱼网站或包含恶意软件的网站,或包含恶意软件作为文件附件. 垃圾邮件发送者从聊天室、网站、客户列表、新闻组和获取用户地址的病毒中收集电子邮件地址,这些收集到的电子邮件地址有时也会卖给其他垃圾邮件发送者.

3　人工智能技术在网络安全中的应用

“人工智能”这一词语最早起源于 1956 年 8 月,约翰·麦卡锡、马文·闵斯基、克劳德·香农等在美国达特茅斯学院的会议中对人工智能给出了最初的定义:“用机器来模仿人类学习以及其他方面的智能”. 从 90 年代开始,物理计算能力的提升使得人工智能迎来了飞速发展,到了现在,互联网技术的成熟、大数据、云计算等支撑技术的完善让人工智能的发展变得越来越快,人工智能带来的强大学习和计算能力正不断被应用于网络态势分析、计算机视觉、语音识别,自然语言处理等多种领域. 近年来,人工智能技术也被广泛应用于网络安全防护中,能够很好地解决网络入侵、恶意软件、网络钓鱼和垃圾邮件等方面的问题. 在本节中,首先将网络安全中应用的人工智能基础技术进行分类,重点介绍其中的机器学习和深度学习方法,进而,针对网络安全中不同的常见问题,对近年来相关工作中的解决方案和带来的赋能效果进行系统性的总结.

3.1 传统机器学习赋能网络安全

本文将传统的机器学习划分为3类,分别是决策树类机器学习方法、基于贝叶斯类的机器学习方法以及基于聚类的机器学习方法,进而整体上介绍传统的机器学习方法对网络安全的赋能应用.

机器学习中的"决策树"能够通过对历史数据的分析,实现对求得目标的分类或预测. 决策树代表的是对象属性与对象值之间的一种映射关系. 贝叶斯思想被总结为一种条件概率,即在事件B发生的情况下,事件A发生的概率. 在20世纪后,朴素贝叶斯思想被广泛应用于机器学习的决策中,首先在不完全情报下,对部分未知的状态用主观概率估计,然后用贝叶斯公式对发生概率进行修正,最后再利用期望值和修正概率做出最优决策. 聚类是机器学习中一种重要的无监督训练算法,可以将数据点整合为一系列特定的组合. 理论上分为同一类别的数据点具有相同的特征,而不同类别的数据点具有不同的属性. 网络安全问题可以抽象映射为机器学习可以解决的问题,将机器学习应用到网络安全已成为近年来安全领域的研究热点.

(1)网络入侵

Panigrahi 等[13]提出了一种基于c4.5的入侵检测系统,该系统基于目前流行的统一树构造(CTC)算法,能够有效地处理类别不平衡的数据. 已经提出了一种称为监督相对随机采样(SRRS)的随机采样机制的改进版本,用于在检测器预处理阶段从高级不平衡数据集中生成平衡样本. 实验结果表明,该系统在NSL-KDD数据集和CICIDS2017数据集上有很高的检测精度.

移动自组织网络中节点的动态特性给网络带来了安全问题,大多数入侵检测方法都在能量消耗方面取得了较好的检测效果,但信任仍然是一个重要因素. Veeraiah 等[14]提出了一种信任感知模糊聚类和模糊朴素贝叶斯(Trust-aware fuzzy clus-fuzzy NB)的自组网入侵检测方案,模糊朴素贝叶斯通过节点信任表确定节点中的入侵行为. 仿真实验在存在和不存在节点攻击的情况下进行分析,并基于时延、能量、检测率和吞吐量等指标对所提方法进行了验证,仿真结果表明了所提方法的有效性.

数据质量和训练算法被认为是决定入侵检测能力的两个关键因素,现有的研究对数据质量考虑的比较少,而这对于构建一个高性能入侵检测系统非常重要. Gu 等[15]提出了一种基于支持向量机和朴素贝叶斯特征嵌入的入侵检测框架. 具体而言,它是一种数据质量改进技术,即朴素贝叶斯特征嵌入,将原始数据转化为高质量数据,然后利用支持向量机建立入侵检测模型. 实验表明,该方法在UNSW-NB15, CICIDS2017, NSL-KDD, Kyoto 2006 + 等数据集上均取得了良好的效果.

k-means 的简单性和效率使得它在聚类分析中很受欢迎,然而k-means有收敛到局部最优的趋势,并且依赖于聚类中心的初始值. Chen 等[16]提出了一种高效的混合聚类算法,称为QALO-K,该算法结合量子计算和群体智能算法的优点,对k-means算法进行改进,使k-means算法向全局最优方向收敛. 将该方法应用于KDD Cup 99大型数据集进行入侵检测. 仿真结果表明,该算法可以有效地用于数据聚类和入侵检测.

(2)恶意软件

静态恶意软件检测是安全套件中的一个基本层,它试图在执行前将样本分类为恶意或良性. Pham 等[17]提出了一种使用梯度增强决策树算法的静态PE恶意软件检测方法,通过便携式可执行分析和梯度增强决策树算法,适当地降低特征维数来减少训练时间. 在恶意软件研究基准数据集EMBER上,基于超过600 000次训练和200 000次测试,提出的方法有着良好的表现.

为了更有效地检测Android恶意软件,Shang 等[18]提出了一种基于改进朴素贝叶斯分类的Android恶意软件检测模型. 提出了基于改进的朴素贝叶斯的恶意检测算法提高检测率,还提出了一种基于Pearson相关系数的相关方法来处理特权属性,并利用Android应用程序属性之间的相关性对结果进行优化.

Zhang 等[19]提出了一种新的Android恶意软件聚类方法ANDRE,该方法利用异构信息,包括代码相似性、利用反病毒厂商的原始标签和元数据信息,共同学习一种混合表示,将网络中的所有恶意软件嵌入到一个低维、紧凑的混合特征空间中,有效地聚类弱标记恶意软件.

(3)钓鱼网站

在网络钓鱼检测方面,Zhu 等[20]提出了一种基于决策树和最优特征选择的神经网络钓鱼检测模型. 首先,对传统的K-medoids聚类算法进行改进,采用增量选择初始中心的方法去除公共数据集中的重复点. 然后,设计了一种基于新定义的特征评价指标、决策树和局部搜索方法的最优特征选择算法,以剔除负面的、无用的特征. 最后,通过适当调整参数构造神经网络分类器的最优结构,并利用所选最优特征进行训练. 实验结果表明,该模型比现有的许多方法具有更高的性能.

(4)垃圾邮件

针对垃圾邮件数据集存在的严重不平衡问题,Lu 等[21]提出了一种新的web垃圾邮件检测集成分类器,分类器能够自动采样和选择子分类器. 通过构建若干

C4.5 决策树子分类器,利用这些子分类器构造一个集成决策树分类器,用于对测试数据中的实例进行分类.在 WEBSPAM-UK2006 数据集上进行的实验表明,与一些基准系统和最新方法相比,具有显著的分类性能.

负选择算法(NSA)是一种解决垃圾邮件问题的方法,然而,NSA 方法缺乏连续适应性,检测性能较差.Chikh 等[22]提出了一种新的基于改进的 NSA 垃圾邮件检测方法,即聚类 NSA 和果蝇优化组合(CNSA FFO).该系统将实际的 NSA 与 k-means 聚类和 FFO 相结合,提高了经典 NSA 的效率.通过对实际垃圾邮件数据集的性能和准确性测试表明,CNSA FFO 方法能够比传统的 NSA 方法和其他模型更好地检测垃圾邮件.

随着注册用户社交活动的增加,Twitter 社交网络越来越受欢迎,但是也有一些垃圾信息散布者利用 Twitter 传播恶意信息,发布钓鱼链接,用虚假账户在网络上泛滥,并从事其他恶意活动.研究人员提出了许多方法来识别一组垃圾邮件发送者,然而每种方法都针对特定类别的垃圾邮件发送者.Adewole 等[23]提出了一种不同的方法来检测 Twitter 上的垃圾邮件发送者.该方法基于垃圾邮件帐户之间存在的相似性,通过引入 PCA 和优化的 K-means 算法来提高垃圾邮件发送者聚类的初始检测,从 200 多万条 tweets 中随机选择超过 20 万个账号进行聚类,以检测垃圾邮件发送者的聚类,实验结果证明算法取得了良好的成果.

3.2　支持向量机赋能网络安全

支持向量机(SVM)是由统计学习理论(SLT)发展而来的一种新的通用学习方法,主要解决高维空间的小样本学习问题.SVM 的主要思想是求解能够正确划分训练数据集并且几何间隔最大的分离超平面.SVM 作为一种新颖的小样本学习方法,基本不涉及概率测度及大数定律等,能够实现高效的从训练样本到预报样本的推理,大大简化了通常的分类和回归等问题.SVM 可以很好地解决网络安全数据的分类、预测、关联规则学习等问题,从而可以给出预防网络威胁的更优解决方案.

(1)网络入侵

在无线传感器网络(WSN)的入侵检测研究中,Safaldin 等[24]通过使用带有支持向量机的修正二元灰狼优化器(GWOSVM-IDS)提出了一种增强型入侵检测系统.GWOSVM-IDS,旨在通过降低误报率和 WSN 环境中 IDS 产生的特征数量来提高 WSN 环境中的入侵检测精度和检测率,并减少处理时间.

Saleh 等[25]设计了一种基于 SVM 的混合 IDS(HIDS)方法,可以以实时方式成功使用并适合解决多类分类问题.通过应用基于距离的方法来选择信息量最大的训练示例,然后将其用于训练优化支持向量机(OSVM),从而拒绝异常值.之后,使用 OSVM 来拒绝异常值.最后,在拒绝异常值之后,HIDS 可以通过应用优先 K-最近邻分类器成功检测攻击.

Gu 等[26]提出了一种基于具有特征增强的 SVM 集成入侵检测框架,通过对原始特征进行对数边际密度比变换后,使用 SVM 集成构建入侵检测模型.Raman 等[27]提出一种基于支持向量机的 Hyper Clique 改进二元引力搜索算法(HC-IBGSA SVM),能够在检测率和误报率方面提高 SVM 的性能.

(2)恶意软件

在恶意软件检测中,Wadkar 等[28]应用基于线性支持向量机权重的特征排序来识别恶意软件样本在不同时间的不同变化问题.通过长时间分析,基于自动化和可量化的机器学习技术能够高效检测恶意软件样本中的进化变化.Ghouti 等[29]提出了一种仅使用恶意软件二进制文件的图像表示来检测和分类恶意软件的新方案.使用主成分分析在紧凑的子空间中提取恶意软件类别和结构的高度判别特征.然后,设计了一种优化的 SVM 模型将提取的特征进行恶意软件类别分类.

(3)钓鱼网站

Anupam 等[30]提出一种利用网站 URL 的不同属性进行钓鱼网站检测的 SVM 分类方法,在现有数据集上训练的支持向量机二进制分类器能够通过寻找最佳超平面区分两个类别,并预测网站是否为合法网站,将网站分类为网络钓鱼和非网络钓鱼.Rao 等[31]提出一种基于双支持向量机(TWSVM)的新型启发式技术用以检测恶意注册的网络钓鱼站点以及托管在受感染服务器上的站点.通过比较,TWSVM 能够以 98.05% 的显着准确率比较访问网站的登录页面和主页来检测托管在受感染域上的网络钓鱼网站.Ravi 等[32]则讨论了一种基于软件定义网络的新型框架方法,借助网络空间中使用 CANTINA 方法(DMLCA)的深度机器学习来预防网络钓鱼攻击.CANTINA 方法使用 SVM 来处理网络钓鱼攻击的分类问题,同时能够借助 DMLCA 方法提高检测准确率.

(4)垃圾邮件

Olatunji[33]提出了一种基于支持向量机的垃圾邮件检测模型,强调搜索最佳参数以获得更好的性能.Kumaresan 等[34]提出了一种使用 S-Cuckoo 并基于混合内核的支持向量机(HKSVM)的垃圾邮件分类框架.首先,根据文本和图像从电子邮件中提取特征,然后,使用提

出的分类器 HKSVM 模型进行分类,这种基于图像提供的附加特征和 SVM 分类器的修改显著地改进了对垃圾邮件的分类能力.

针对垃圾邮件评论实例不足所导致监督技术面临类别不平衡的问题,Tian 等[35]提出了名为 Ramp One-Class SVM 的鲁棒且非凸的半监督算法,采用 One-Class SVM 来处理欺骗性意见缺乏标记数据的问题,并利用 Ramp 损失函数的非凸特性,消除了异常值和非评论意见的影响.

3.3 卷积神经网络赋能网络安全

1980 年,Fukushima 等[36]提出了一种由卷积层、池化层构成的新的神经网络结构,在此基础上,1998 年,Lecun 等[37]将 BP 算法应用在这种神经结构中,提出了 LeNet-5[37]模型,这也就是卷积神经网络(CNN)的雏形. 相比于人工神经网络(ANN)模型中的单神经元结构,CNN 最大的不同在于其使用卷积层代替了原本的全连接层,使用卷积核进行特征提取,结合局部连接和权值共享的方法,能够在大幅减少训练权值参数的情况下获得全局关系. 随着互联网的不断发展,网络状态分析的数据量正在不断攀升,由于训练参数量大幅减少的优势,CNN 模型及其演变优化后的模型正在被应用于网络流量检测、网络态势分析等场景.

(1)网络入侵

不同类型的 IDS 被设计为仅用于解决单一类型的入侵或多种变体,Shams 等[38]提出了一种新的上下文感知特征提取方法,作为基于 CNN 的多类入侵检测的预处理步骤. 基于此的 IDS 系统可以识别 4 ~ 12 种不同类型的入侵检测. Nguyen 等[39]提出了一种网络入侵检测系统 NIDS 新算法,该算法将 GA、CNN 提取器和 BG 分类器进行合理组合,实现了一种 3 层的特征提取结构. 实验证明,将 CNN 模型作为特征提取器,结合 BG 分类器的混合学习方法,能够提高该算法的最终分类性能. 在许多网络安全的衍生领域中,CNN 模型也提供了入侵检测问题的多种解决方案. Jeong 等[40]首次将 CNN 模型应用在自动驾驶汽车的安全场景下,以解决汽车以太网的入侵检测问题. 提出一种基于特征生成和 CNN 网络的入侵检测系统,建立了一个基于 BroadR-Reach 的物理测试平台,并捕获了真实的 AVTP 包进行性能评价. Gao 等[41]考虑到电网监控下入侵对象规模的多样性和应用场景的复杂性,提出了一种改进的基于上下文感知掩码区域的 Mask R-CNN 模型,即 ID-Net,用于入侵对象检测. 一个调制的可变形卷积操作被集成到主干网络中,可以用于从工程车辆的几何变化中学习鲁棒的特征表示.

(2)恶意软件

Cui 等[42]提出了一种利用 CNN 和智能算法进行恶意代码检测的方法. CNN 用于对恶意代码可执行文件转换成的灰度图像进行识别和分类. 然后采用非支配排序遗传算法 II(NSGA-II)来处理恶意软件族的数据不平衡问题. 为了提高大规模 Android 恶意软件检测的准确性和效率,Lu 等[43]提出了一种基于深度神经网络的高效恶意软件检测框架 DLAMD,结合能够实现快速响应的预测阶段和精准溯源的深度检测阶段,选择自动提取特征内部隐藏模式的 CNN 进行特征选择. 相似地,Wang 等[44]提出了一种基于深度自编码器和串行卷积神经网络结构的混合模型,重构了 Android 应用程序的高维特征,并利用多个 CNN 对 Android 恶意软件进行检测.

(3)钓鱼网站

Adebowale 等[45]重点设计开发了一种基于深度学习的钓鱼检测解决方案,利用通用资源定位器和网站内容,采用 CNN 和长短时记忆算法构建了一种名为智能钓鱼检测系统的混合分类模型,在大数据集情况下提升分类器预测性能. 相似地,Parra 等[46]提出了一种基于云的分布式深度学习框架,用于网络钓鱼和僵尸网络攻击检测及缓解. 该模型使用分布式 CNN 模型作为物联网设备微安全插件嵌入,用于检测网络钓鱼和应用层 DDoS 攻击. 分布式 CNN 模型嵌入到客户端物联网设备的 ML 引擎中,能够在源头检测和保护物联网设备免受网络钓鱼攻击.

(4)垃圾邮件

Liu 等[47]提出了一种新的检测方法,即从用户的角度对恶意网页进行截屏,从而使网络垃圾邮件失效. 采用 CNN 作为分类算法,在真实的 Web 环境中进行了 3 个月的恶意网站检测测试且性能良好. 最近,CNN 开始应用于合成孔径雷达(SAR)图像的自动目标识别(ATR)问题. Oh 等[48]提出了一种基于 CNN 的具有姿态角边缘化学习的 SAR 目标识别网络(SPAM-Net),它边缘化了 SAR 目标在其姿态角上精确估计真实的类别概率.

3.4 循环神经网络赋能网络安全

1933 年西班牙神经生物学家 Rafael Lorente 在研究大脑皮层时发现刺激信号能够在神经回路中循环传递,因此,提出一种反向回路假设. 这种假设之后被神经生物学领域总结为循环反馈系统,并基于此衍生出了各类数学模型. 1990 年,Elman[49]提出了第一个全连接的循环神经网络(RNN),这是一种在时间结构上存在共享特性的神经网络变体,一个序列的当前输出与前面的输出

也是有关的.其单个的神经元相比以往 ANN 中的神经元添加了反馈输入,也就是通过一系列权值共享前馈神经元的依次连接,这样使得循环神经网络在 t 时刻的输入不仅实现与输出的映射,而且能够参考 t 时刻之前所有输入数据对网络的影响.因为 RNN 独有的对高级特性依赖关系如时序特征的提取能力,正在被逐渐应用在网络态势感知等安全问题中.

(1)网络入侵

针对云环境下的入侵检测问题,Balamurugan 等[50]提出了一种归一化 K 均值聚类算法与 RNN 组合而成的新颖算法,包括检查来自用户数据包的审查算法和称为 NK-RNN 的混合分类模型,能够有效地检测到实验证明的入侵者.在车载通信中,针对控制器局域网(CAN)总线缺乏防御的问题,Tariq 等[51]提出了一种基于 RNN 的 CAN 总线消息攻击检测框架(CAN-ADF),采用由动态网络流量特征和 RNN 组成的基于规则的检测方法,实现了准确的入侵检测性能.

(2)恶意软件

Sun 等[52]将恶意代码的静态分析与 RNN 和 CNN 方法相结合.通过使用 RNN,不仅考虑了恶意软件的原始信息,还考虑了将原始代码与时序特征相关联的能力,然后,使用 minhash 从原始代码以及来自 RNN 预测代码的融合中生成特征图像,并使用 CNN 来对特征图像进行分类.

近年来,加密货币交易急剧增加,这一趋势也吸引了网络威胁参与者利用现有漏洞感染目标. Yazdinejad 等[53]提出了一种新颖的 RNN 学习模型,用于寻找加密货币恶意软件威胁,使用 5 种不同的长短期记忆(LSTM)结构进行训练,并通过 10 倍交叉验证技术进行评估.随着物联网设备越来越多地部署在不同行业中,越来越多的应用程序及其不断增强的计算和处理能力使它们成为有价值的攻击目标. Haddad 等[54]探讨了使用 RNN 模型检测 IoT 恶意软件的潜力,使用 RNN 来分析基于 ARM 的 IoT 应用程序执行操作代码.

(3)钓鱼网站

基于之前的工作,Somesha 等[55]提出了一种基于启发式特征的机器学习方法用于钓鱼网站检测,并使用 18 个特征实现了 99.5% 的准确率.在本文中,针对笔者在早期工作分析的 10 种特征,并基于 DNN、LSTM 和 CNN,提出了一种新颖的网络钓鱼 URL 检测模型,实现了 LSTM 最高 99.57% 的准确率.同样为了克服恶意 URL 使用户受害的问题,Shivangi 等[56]提出了一种在 chrome 扩展的检测工具,使用 ANN 和 LSTM 网络来分析 URL,并对其进行分类,旨在帮助用户避免成为恶意 URL、网络钓鱼和社会工程等恶意和欺诈活动的受害者.

(4)垃圾邮件

Makkar 等[57]在检测网络垃圾邮件的研究中,提出一种基于 RNN 深度学习架构来检测隐藏模式,设计了一个名为 FS^2RNN 的 RNN 特征选择方案框架.同时,Makkar 等[58]还提出了一种通过浏览不同网站和网页使用特征提取的恶意图像广告垃圾邮件保护器(SPAMI)框架,使用 CNN、RNN 和 LSTM 模型标记垃圾邮件广告图像. Xu 等[59]提出了一个 Sifter 系统,一种无需劳动密集型特征工程即可以可扩展方式检测在线社交垃圾邮件的系统,能够在 RNN 的支持下处理社交垃圾邮件,摆脱传统的人工特征工程.

3.5　对抗神经网络赋能网络安全

2014 年,Goodfellow[60]提出了一种基于博弈论的新颖的神经网络模型——生成式对抗神经网络(GAN),该网络由两个目标互相冲突的神经网络组成,分别为生成器和鉴别器,通过对抗性过程同时进行训练,当取得纳什均衡时则达到生成器的训练目标.近年来,GAN 已经成为机器学习领域最火热的研究之一,Yann LeCun 更是称之为“过去 10 年间机器学习领域最让人激动的点子”. GAN 的优势在于其提供了神经网络生成困难的解决思路,同时,在最新的研究工作中,GAN 中的鉴别器能够被用作目标神经网络训练过程中的监控器,以防止生成器过拟合.研究者们正在将 GAN 应用在网络安全领域中,着重解决网络攻击样本生成和网络攻击行为检测问题,帮助构建智能有效的网络安全防护机制.

(1)网络入侵

在入侵检测领域,生成对抗网络是一种很有前途的无监督方法,通过对系统进行隐式建模来检测网络攻击. de Araujo-Filho 等[61]提出了 FID-GAN,一种用于使用 GAN 的 CPS 新型基于雾的无监督入侵检测系统,能够通过训练加速重建损失计算的编码器,实现映射到潜在空间数据样本的重建来计算重建损失.入侵检测中流量异常模式样本的不平衡数据是一个具有挑战性的问题,Huang 等[62]提出了一种新颖的不平衡生成对抗网络(IGAN),在典型的 GAN 中引入不平衡数据过滤器和卷积层,为少数类生成新的代表性实例,建立一个基于 IGAN 的入侵检测系统(IGAN-IDS),使用 IGAN 生成的实例来应对类不平衡的入侵检测. Yan 等[63]提出了一种 DoS-WGAN 通用架构,使用 Wasserstein 生成对抗网络(WGAN)和梯度惩罚技术来逃避网络流量分类器.为了将拒绝服务攻击(DoS)流量伪装成正常的网络流量,

DoS-WGAN 会自动合成攻击痕迹,可以击败针对 DoS 案例的现有 NIDS/网络安全防御,使基于 CNN 的 NIDS 检测率从 97.3% 下降到 47.6%. 本架构将在网络攻防博弈中发挥特别重要的作用.

(2)恶意软件

以前基于 GAN 的研究是使用相同的特征量来学习恶意软件检测. 但是现有的学习算法往往使用多个恶意软件,影响了规避的性能,对攻击者来说是不现实的. 为了解决这个问题,Kawai 等[64]应用了具有不同特征量和只有一种恶意软件的差异化学习方法. 为了评估机器学习算法在恶意软件检测中的脆弱性,Taheri 等[65]提出了5 种不同的攻击场景来干扰恶意应用程序. 为了区分对抗样本和良性样本,提出了两种防御机制来对抗攻击. 结果表明,当使用生成对抗网络方法时,提出的攻击模型在用于强化开发的反恶意软件系统时,生成的规避变体将检测率提高了 50%. 为了同时规避恶意软件检测和对抗示例检测,Li 等[66]在双目标 GAN 的基础上开发了一种新的对抗示例攻击方法. 该方法生成的对抗实例中有 95% 以上的实例能够突破防火墙 Android 恶意软件检测系统.

(3)垃圾邮件

垃圾邮件的特征收集通常需要很长时间,因此很难将它们应用到冷启动垃圾邮件审查检测任务中. 因此,Tang 等[67]利用 GAN 来解决这个问题,为新用户从易于访问的特性(EAFs)中生成合成行为特性(SBFs). 首先,为普通用户选择 6 个公认的真实行为特征(RBFs). 然后,训练一个 GAN 框架,包括一个生成器,从包含文本、评级和属性特征的 EAF 中生成 SBFs,以及一个鉴别器来区分 RBFs 和 SBFs.

3.6 强化学习赋能网络安全

1911 年,行为心理学专家 Thorndike 提出了效用法则,指出行为会被记住取决于该行为产生的效用. 1954 年 Minsky 首次提出“强化”和“强化学习”的概念和术语,之后,1989 年,Watkins 等[68]提出了现今强化学习最广泛使用的 Q 学习策略(Q-learning). 2013 年,DeepMind 发表了强化模型模拟人类进行 Atari 游戏的论文,从此,强化学习开始了飞速发展. 强化学习不同于传统的机器学习,其核心思想是通过试错学习如何能最佳地匹配状态(States)和动作(Actions),以期获得最大的回报(Rewards). 在面对网络安全问题时,强化学习能够通过试错的方式不断探索更优的决策,在网络状态变化后,能够智能地动态建模,推动网络安全走向智能化.

(1)网络入侵

Lopez-Martin 等[69]提出了一种新的应用深度强化学习(DRL)算法的入侵检测框架,是对经典 DRL 范式(基于与现场环境的交互)进行概念上的修改,用记录训练入侵的采样函数替换环境. 这种新的伪环境除了对训练数据进行采样外,还会根据训练过程中发现的检测错误产生奖励. 对 AWID 和 NSL-KDD 数据集获得的结果与其他机器学习模型进行了全面的比较,结果表明,与现有的机器学习技术相比,DRL 通过模型和一些参数的调整,可以提高入侵检测的速度和效果. Caminero 等[70]提出了一种新的基于监督 RL 模型和对抗 RL 模型相结合的框架 AE-RL,提供了一个遵循 RL 环境指导方针的模拟环境. 这种方法的原理是为模拟环境提供一种智能行为. 首先,通过从训练数据集中随机提取新样本,根据分类器预测的好坏产生奖励;其次,通过进一步调整初始行为与对抗目标,环境将积极地尝试增加分类器做出准确预测的难度. 实验结果证明这种结构提高了分类器的最终性能. 基于物联网的现代智能家庭是一个具有挑战性的安全环境:设备不断变化,新的漏洞被发现并经常未打补丁,不同用户与设备的交互方式不同,对网络风险的态度也不同. Heartfield 等[71]提出了 MAGPIE,其为第一个智能家庭入侵检测系统,能够自动调整其底层异常分类模型的决策功能,以适应智能家庭不断变化的条件. 该方法将一种新的基于概率聚类的奖励机制应用于非平稳多臂土匪强化学习中,从而达到上述目的. 在真实家庭中的实验评估表明,MAGPIE 显示出了较高的准确性.

(2)恶意软件

最近的研究表明,基于监督学习的恶意软件检测模型很容易受到蓄意攻击,因为其依赖于带有明确标记的静态特征. 为了暴露和展示这些模型中的弱点,Fang 等[72]提出了一种利用强化学习规避反恶意软件引擎的 DQEAF 框架. DQEAF 通过神经网络不断地与恶意软件样本交互来训练 agent. 动作是一组合理的修改,不破坏样本的结构和功能. Agent 可以通过深度强化学习选择最优的功能保持动作序列,目的是规避监督检测器. 实验表明,该方法在 PE 样本中具有较高的成功率,其有效性也得到了其他恶意软件族的验证,显示了良好的鲁棒性. 准确检测移动设备上的恶意软件需要快速处理大量的应用轨迹,基于云的恶意软件检测可以利用安全服务器的数据共享和强大的计算资源来提高检测性能. Xiao 等[73]设计了一种基于云的恶意软件检测游戏,并导出了游戏的纳什均衡,展示了移动设备如何选择卸载速率,在传输成本和检测性能之间进行权衡. 针对时变无

线网络中的动态博弈,提出了一种基于 Q 学习的恶意软件检测策略,并利用 Dyna 体系结构和已知无线信道模型进一步提高了检测性能.

(3)网络钓鱼

Smadi 等[74]首次提出一种将神经网络与强化学习相结合的在线模式下检测钓鱼攻击的新框架. 该模型通过采用强化学习的思想,随时间动态地增强系统,能够自我适应,产生一个新的钓鱼邮件检测系统,反映出新探索行为的变化. 该模型通过在线方式自动向原有数据集添加更多的电子邮件,解决了数据集有限的问题,提出一种新的算法来探索新的数据集中的任何新增添的钓鱼行为. 通过使用已知的数据集进行严格的测试,证明该技术可以有效地处理零日网络钓鱼攻击. 受钓鱼网站进化特性的启发,Chatterjee 等[75]设计了一种基于深度强化学习的网络钓鱼自动检测框架,来建模和检测恶意 URL. 该模型能够适应网络钓鱼网站的动态行为,从而学习与网络钓鱼网站检测相关的特征.

4　问题与挑战

虽然目前人工智能方法已经在网络安全领域证明了其强大的效果,也展现出了相比传统方法在检测精度、灵活程度等方面优势,但是目前仍然面临一些具有挑战性的问题:

如何选择合适的数据集是一个重要问题,大多数现有数据集是比较旧的,使用这些数据集会导致算法在理解新型网络攻击的行为模式方面存在不足. 如 KDD99[76]或 NSL-KDD[77]数据集,这些数据集包含旧的可能已经过时的流量,不能代表最近的攻击场景和流量行为. 因此,需要针对入侵检测等特定问题领域建立最新的数据集,以及研究如何设计可扩展性强的更加灵活的数据集,这可能是网络安全数据科学领域面临的重要挑战.

如果要比较相关工作的运行效果,应该在完全相同环境下进行对比. 但不幸的是,不同的研究工作之间是比较孤立的,在效果和效率方面很难提供一个公平公正的比较,这是由于环境的多样性造成的,包括但不限于:①使用的数据集;②采用的数据集部分;③预处理方法;④超参数配置;⑤硬件平台. 因此,需要使用统一的计算平台和考虑共同的影响因素进行更多的实验研究,以期获得公平的比较结果.

基于人工智能的安全模型通常使用大量静态数据来构建检测系统,训练其学习正常和异常行为的能力. 然而,大型的动态安全系统中的正常行为如果没有被很好的定义,数据库的增长可能会使得正常的行为模式随着时间的推移而改变,这常常导致大量的假警报. 大型动态系统中的安全检测是一个既现实而又复杂的问题,值得研究人员投入更多的关注.

5　总　结

随着互联网日新月异的发展,其面临的威胁与挑战也愈发复杂. 本文总结了人工智能技术应用在网络安全领域的最新进展,涉及到的技术包括传统机器学习、深度学习及强化学习等方法. 本文还调查了近年来网络安全领域中人工智能技术的相关应用,总结出了 4 个最受关注的研究方向,分别是网络入侵、恶意软件、钓鱼网站以及垃圾邮件,之后介绍了人工智能技术在每个研究方向中的具体应用场景. 最后,本文指出了当前面临的问题与挑战.

参考文献:

[1] We are social & Hootsuite, digital 2021 global overview report[R/OL]. (2021-01-23). https:// datareportal. com/reports/digital-2021-global-overview-report, 2021-01-23.

[2] 中国信息通信研究院. 中国数字经济发展白皮书(2020)[R/OL]. (2020-07-03) [2021-03-21]. http://www. cac. gov. cn/files/pdf/baipishu/shuzijingjifazhan. pdf.

[3] 《世界互联网发展报告 2020》《中国互联网发展报告 2020》蓝皮书新闻发布会[R/OL]. (2020-11-23). http://it. people. com. cn/GB/119390/118340/434404/434417/434436/index. html.

[4] Schatz D, Bashroush R, Wall J. Towards a more representative definition of cyber security[J]. Journal of Digital Forensics, Security and Law, 2017, 12(2): 53-74.

[5] Wiafe I, Koranteng F N, Obeng E N, et al. Artificial intelligence for cybersecurity: A systematic mapping of literature[J]. IEEE Access, 2020, 8: 146598-146612.

[6] Abbas N N, Ahmed T, Shah S H U, et al. Investigating the applications of artificial intelligence in cyber security[J]. Scientometrics, 2019, 121(2): 1189-1211.

[7] Mahdavifar S, Ghorbani A A. Application of deep learning to cybersecurity: A survey[J]. Neurocomputing, 2019, 347: 149-176.

[8] Torres J M, Comesaña C I, Garcia-Nieto P J. Machine learning techniques applied to cybersecurity[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(10): 2823-2836.

[9] Dua S, Du X. Data mining and machine learning in cybersecurity[M]. Boca Raton: CRC Press, 2016.

[10] Shafiq M, Tian Z, Bashir A K, et al. CorrAUC: A malicious bot-ioT traffic detection method in ioT network using machine learning techniques[J]. IEEE Internet of Things Journal, 2021, 8(5):3242-3254.

[11] Shafiq M, Tian Z, Bashir A, et al. IoT malicious traffic identification using wrapper-based feature selection mechanisms [J]. Computers & Security, 2020, doi:10.1016/j.cose.2020.101863.

[12] Shafiq M, Tian Z, Bashir A A, et al. Data mining and machine learning methods for sustainable smart cities traffic classification: A survey[J]. Sustainable Cities and Society, 2020,doi:10.1016/j.scs.2020.102177.

[13] Panigrahi R, Borah S, Bhoi A K, et al. A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets[J]. Mathematics, 2021, 9(7): 751.

[14] Veeraiah N, Krishna B T. Trust-aware fuzzy clus-fuzzy NB: Intrusion detection scheme based on fuzzy clustering and Bayesian rule[J]. Wireless Networks, 2019, 25(7): 4021-4035.

[15] Gu J, Lu S. An effective intrusion detection approach using SVM with naïve Bayes feature embedding[J]. Computers & Security, 2021, doi:10.1016/j.cose.2020.102158.

[16] Chen J, Qi X, Chen L, et al. Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection[J]. Knowledge-Based Systems, 2020, doi:10.1016/j.knosys.2020.106167.

[17] Pham H D, Le T D, Vu T N. Static PE malware detection using gradient boosting decision trees algorithm[C]//International Conference on Future Data and Security Engineering. Cham:Springer, 2018: 228-236.

[18] Shang F, Li Y, Deng X, et al. Android malware detection method based on naive Bayes and permission correlation algorithm[J]. Cluster Computing, 2018, 21(1): 955-966.

[19] Zhang Y, Sui Y, Pan S, et al. Familial clustering for weakly-labeled android malware using hybrid representation learning [J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 3401-3414.

[20] Zhu E, Ju Y, Chen Z, et al. DTOF-ANN: An artificial neural network phishing detection model based on decision tree and optimal features[J]. Applied Soft Computing, 2020, doi:10.1016/j.asoc.2020.106505.

[21] Lu X Y, Chen M S, Wu J L, et al. A novel ensemble decision tree based on under-sampling and clonal selection for web spam detection[J]. Pattern Analysis and Applications, 2018, 21(3): 741-754.

[22] Chikh R, Chikhi S. Clustered negative selection algorithm and fruit fly optimization for email spam detection[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(1): 143-152.

[23] Adewole K S, Han T, Wu W, et al. Twitter spam account detection based on clustering and classification methods[J]. The Journal of Supercomputing, 2020, 76(7): 4802-4837.

[24] Safaldin M, Otair M, Abualigah L. Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(2): 1559-1576.

[25] Saleh A I, Talaat F M, Labib L M. A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers[J]. Artificial Intelligence Review, 2019, 51(3): 403-443.

[26] Gu J, Wang L, Wang H, et al. A novel approach to intrusion detection using SVM ensemble with feature augmentation[J]. Computers & Security, 2019, 86: 53-62.

[27] Raman M R G, Somu N, Jagarapu S, et al. An efficient intrusion detection technique based on support vector machine and improved binary gravitational search algorithm[J]. Artificial Intelligence Review, 2020,53:3255-3286.

[28] Wadkar M, Di Troia F, Stamp M. Detecting malware evolution using support vector machines[J]. Expert Systems with Applications, 2020, doi:10.1016/j.eswa.2019.113022.

[29] Ghouti L, Imam M. Malware classification using compact image features and multiclass support vector machines[J]. IET Information Security, 2020, 14(4): 419-429.

[30] Anupam S, Kar A K. Phishing website detection using support vector machines and nature-inspired optimization algorithms [J]. Telecommunication Systems, 2021, 76(1): 17-32.

[31] Rao R S, Pais A R, Anand P. A heuristic technique to detect phishing websites using TWSVM classifier[J]. Neural Com-

puting and Applications, 2021,33(11): 5733-5752.
[32] Ravi R. A performance analysis of software defined network based prevention on phishing attack in cyberspace using a deep machine learning with CANTINA approach (DMLCA)[J]. Computer Communications, 2020, 153: 375-381.
[33] Olatunji S O. Improved email spam detection model based on support vector machines[J]. Neural Computing and Applications, 2019, 31(3): 691-699.
[34] Kumaresan T, Saravanakumar S, Balamurugan R. Visual and textual features based email spam classification using S-cuckoo search and hybrid kernel support vector machine[J]. Cluster Computing, 2019, 22(1): 33-46.
[35] Tian Y, Mirzabagheri M, Tirandazi P, et al. A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM[J]. Information Processing & Management, 2020, doi:10.1016/j.ipm.2020.102381.
[36] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]. Competition and Cooperation in Neural Nets. Berlin, Heidelberg: Springer, 1982.
[37] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[38] Shams E A, Rizaner A, Ulusoy A H. A novel context-aware feature extraction method for convolutional neural network-based intrusion detection systems[J]. Neural Computing and Applications,2021, doi:10.1007/S00521-021-05994-9.
[39] Nguyen M T, Kim K. Genetic convolutional neural network for intrusion detection systems[J]. Future Generation Computer Systems, 2020, 113: 418-427.
[40] Jeong S, Jeon B, Chung B, et al. Convolutional neural network-based intrusion detection system for AVTP streams in automotive Ethernet-based networks[J]. Vehicular Communications, 2021, doi:10.1016/j.vehcom.2021.100338.
[41] Gao F, Ji S, Guo J, et al. ID-Net: an improved mask R-CNN model for intrusion detection under power grid surveillance [J]. Neural Computing and Applications, 2021,33: 9241-9257.
[42] Cui Z, Du L, Wang P, et al. Malicious code detection based on CNNs and multi-objective algorithm[J]. Journal of Parallel and Distributed Computing, 2019, 129: 50-58.
[43] Lu N, Li D, Shi W, et al. An efficient combined deep neural network based malware detection framework in 5G environment[J]. Computer Networks, 2021, doi:10.1016/J.COMNET.2021.107932.
[44] Wang W, Zhao M, Wang J. Effective android malware detection with a hybrid model based on deep autoencoder and convolutional neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(8): 3035-3043.
[45] Adebowale M A, Lwin K T, Hossain M A. Intelligent phishing detection scheme using deep learning algorithms[J/OL]. Journal of Enterprise Information Management, 2020. [2021-06-20]. https://www.emerald.com/insight/content/doi/10.1108/JEIM-01-2020-0036/full/html.
[46] Parra G D L T, Rad P, Choo K K R, et al. Detecting internet of things attacks using distributed deep learning[J]. Journal of Network and Computer Applications, 2020, doi:10.1016/j.jnca.2020.102662.
[47] Liu D, Lee J H. CNN based malicious website detection by invalidating multiple web spams[J]. IEEE Access, 2020, 8: 97258-97266.
[48] Oh J, Youm G Y, Kim M. SPAM-Net: A CNN-based SAR target recognition network with pose angle marginalization learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,31(2):701-714.
[49] Elman J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.
[50] Balamurugan V, Saravanan R. Enhanced intrusion detection and prevention system on cloud environment using hybrid classification and OTS generation[J]. Cluster Computing, 2019, 22(6): 13027-13039.
[51] Tariq S, Lee S, Kim H K, et al. CAN-ADF: The controller area network attack detection framework[J]. Computers & Security, 2020, doi:10.1016/j.cose.2020.101857.
[52] Sun G, Qian Q. Deep learning and visualization for identifying malware families[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(1): 283-295.
[53] Yazdinejad A, HaddadPajouh H, Dehghantanha A, et al. Cryptocurrency malware hunting: A deep recurrent neural network approach[J]. Applied Soft Computing, 2020, doi:10.1016/j.asoc.2020.106630.
[54] Haddad P H, Dehghantanha A, Khayami R, et al. A deep recurrent neural network based approach for internet of things malware threat hunting[J]. Future Generation Computer Systems, 2018, 85: 88-96.
[55] Somesha M, Pais A R, Rao R S, et al. Efficient deep learning techniques for the detection of phishing websites[J].

Sādhanā, 2020, 45(1): 1-18.

[56] Shivangi S, Debnath P, Sajeevan K, et al. Chrome extension for malicious urls detection in social media applications using artificial neural networks and long short term memory networks[C]//2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). Piscataway: IEEE, 2018: 1993-1997.

[57] Makkar A, Obaidat M S, Kumar N. Fs2rnn: Feature selection scheme for web spam detection using recurrent neural networks[C]//2018 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE, 2018: 1-6.

[58] Makkar A, Kumar N, Zomaya A Y, et al. SPAMI: A cognitive spam protector for advertisement malicious images[J]. Information Sciences, 2020, 540: 17-37.

[59] Xu H, Guan B, Liu P, et al. Harnessing the nature of spam in scalable online social spam detection[C]//2018 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE, 2018: 3733-3736.

[60] Goodfellow I J. On distinguishability criteria for estimating generative models[EB/OL]. (2015-05-21)[2021-06-15]. https://arxiv.org/pdf/1412.6515.pdf.

[61] de Araujo-Filho P F, Kaddoum G, Campelo D R, et al. Intrusion detection for cyber-physical systems using generative adversarial networks in fog environment[J]. IEEE Internet of Things Journal, 2020,8(8):6247-6256.

[62] Huang S, Lei K. IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks[J]. Ad Hoc Networks, 2020, doi:10.1016/j.adhoc.2020.102177.

[63] Yan Q, Wang M, Huang W, et al. Automatically synthesizing DoS attack traces using generative adversarial networks[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(12): 3387-3396.

[64] Kawai M, Ota K, Dong M. Improved malgan: Avoiding malware detector by leaning cleanware features[C]//2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). Piscataway: IEEE, 2019: 40-45.

[65] Taheri R, Javidan R, Shojafar M, et al. Can machine learning model with static features be fooled: An adversarial machine learning approach[J]. Cluster Computing, 2020,23:3233-3253.

[66] Li H, Zhou S Y, Yuan W, et al. Adversarial-example attacks toward android malware detection system[J]. IEEE Systems Journal, 2019, 14(1): 653-656.

[67] Tang X, Qian T, You Z. Generating behavior features for cold-start spam review detection with adversarial learning[J]. Information Sciences, 2020, 526: 274-288.

[68] Watkins C J C H, Dayan P. Q-learning[J]. Machine Learning, 1992, 8(3/4): 279-292.

[69] Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Application of deep reinforcement learning to intrusion detection for supervised problems[J]. Expert Systems with Applications, 2020, doi:10.1016/j.eswa.2019.112963.

[70] Caminero G, Lopez M, Carro B. Adversarial environment reinforcement learning algorithm for intrusion detection[J]. Computer Networks, 2019, 159: 96-109.

[71] Heartfield R, Loukas G, Bezemskij A, et al. Self-configurable cyber-physical intrusion detection for smart homes using reinforcement learning[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 1720-1735.

[72] Fang Z, Wang J, Li B, et al. Evading anti-malware engines with deep reinforcement learning[J]. IEEE Access, 2019, 7: 48867-48879.

[73] Xiao L, Li Y, Huang X, et al. Cloud-based malware detection game for mobile devices with offloading[J]. IEEE Transactions on Mobile Computing, 2017, 16(10): 2742-2750.

[74] Smadi S, Aslam N, Zhang L. Detection of online phishing email using dynamic evolving neural network based on reinforcement learning[J]. Decision Support Systems, 2018, 107: 88-102.

[75] Chatterjee M, Namin A S. Detecting phishing websites through deep reinforcement learning[C]//2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). Piscataway: IEEE, 2019, 2: 227-232.

[76] Tavallaee M, Bagheri E, Lu W, et al. A detailed analysis of the KDD Cup 99 data set[C]//2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Piscataway: IEEE, 2009: 1-6.

[77] 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.

【责任编辑：孙向荣】