基于人工智能的威胁检测与防护系统

王瑞涵
（中国民用航空西南地区空中交通管理局 成都 610202）

摘　要　文中针对传统的基于签名匹配的威胁检测系统存在的局限，探讨了人工智能技术在网络安全防护中的应用。通过分析异常检测、恶意软件检测和自动化安全响应 3 个方面，阐明了机器学习和深度学习模型可以实现对未知威胁的检测和主动防御。研究认为，人工智能驱动的网络安全防护系统代表了技术发展的方向，但还需进一步的数据积累和模型优化，以实现更智能的商业安全产品的开发。

关键词：人工智能；威胁检测；防护系统；系统设计
中图分类号　TP391

Threat Detection and Protection System Based on Artificial Intelligence

WANG Ruihan
（Southwest Air Traffic Management Bureau of Civil Aviation of China，Chengdu 610202，China）

Abstract　In response to the limitations of traditional signature matching based threat detection systems，this paper explores the application of artificial intelligence technology in network security protection. This paper clarifies that machine learning and deep learning models can achieve detection and active defense against unknown threats by analyzing three aspects：anomaly detection，malware detection，and automated security response. Research suggests that artificial intelligence driven network security protection systems represent the direction of technological development，but further data accumulation and model optimization are needed to achieve more intelligent development of commercial security products.

Keywords　Artificial intelligence，Threat detection，Protection system，System design

0　引言

随着信息技术的迅速发展，网络安全威胁日益严峻。各类网络病毒、木马程序不断涌现，给用户的信息安全带来了巨大的风险。传统的基于签名的威胁检测与防护手段，难以应对日新月异的未知威胁，检测率和准确率较低[1]。为有效应对各类网络安全威胁，亟需研发更智能化的威胁检测与防护系统。近年来，人工智能技术飞速发展，在图像识别、自然语言处理等领域展现出强大的学习和分析能力。将人工智能技术应用于网络安全领域，可实现对网络流量、文件等的高效智能分析，发现潜在的威胁，并进行主动防御。

1　传统威胁检测与防护系统的局限性

传统的威胁检测与防护系统主要依赖已知威胁的签名匹配来实现检测。这需要事先收集各种威胁的特征码，并将其收录到签名库中，当系统检测到文件或网络流量中的特征码与签名库中的某一签名匹配时，即判断存在威胁[2]。这种签名匹配机制可以有效地对已记录在案的老旧威胁进行屏蔽，但很难应对种类繁多、日新月异的未知威胁。例如，通用型木马 Hydra 利用代码变异技术，每天可自动生成上万个变种，其变异速度远超过传统签名更新的速度。2017 年 WannaCry 勒索软件爆发时，文件创建时间戳变异技术产生了数百万种变型，导致基于签名的防护系统完全失效。另外，针对不同的威胁类型，传统系统需设计不同的检测模块，存在模块隔离、无法共享信息、检测联动能力差的问题。例如，针对病毒的静态代码检测系统，难以关联网络流量中的异状连接行为，无法整体识别多级联动攻击；针对 Web 攻击的 WAF 设备，对文件型恶意代码检测能力十分有限，模块割裂导致威胁信息孤立，稽查盲区大，攻击征兆易被忽略。当前，商用威胁检测系统的误报率也普遍过高。基于静态签名匹配的恶意软件检测系统，经常将正常软件误判为病毒；基于行为检测和可疑连接的入侵检测系统也存在大量误报，高误报率不仅会耗费安全分析人员的大量时间来进行人工校验，也容易忽略真正的威胁。总体来说，传统的基于签名和规则的威胁检测与防护系统，效率低、更新不灵活、无法应对未知威胁、误报率高的问题日益凸显，亟需引入更智能的检测与防御技术。

作者简介：王瑞涵（1994—），本科，助理工程师，研究方向为网络信息安全。

2 人工智能在威胁检测与防护系统中的应用

2.1 异常检测与入侵检测系统

相较于基于签名匹配的入侵检测方法，基于人工智能的异常检测系统可以更好地应对未知的攻击[3]。这类系统通过机器学习算法，对网络流量、系统调用、用户操作等大数据进行训练，建立正常行为基线，实时监控 Sysmon 日志、网络流量包特征等，探测可能的攻击行为尝试，不依赖已知攻击模式，对 APT 等前沿高级威胁具有较强的检测能力。例如，Darktrace 的自主神经系统通过无监督学习分析网络常规流量，判断异常连接行为。在拦截勒索软件 WannaCry 初期攻击时，其检测到某台主机尝试向大量随机 IP 发送 SMB 流量，并将其评判为“未知的森林蠕变攻击”，采取隔离措施，有效阻止了后续的勒索软件传播。Deep Instinct 公司基于深度学习的 NPU 智能处理芯片，可以进行近零日未知恶意代码检测。其训练样本达到上亿量级，检出率高达 99.9%，实现了毫秒级检测响应速度。

另外，通过算法优化和模型融合，一些商用入侵检测系统也逐步引入了机器学习技术。例如，Darktrace 的自适应协同引擎整合了多种深度学习算法，实现了对内部威胁的实时监控；Cisco 的网络入侵检测系统结合了威胁情报、行为分析和机器学习，检测比例提高了 60%以上；Palo Alto 的 ML-Powered NGFW，训练样本超过 10 亿，检测准确率达 99.6%。总之，相比规则方法，经人工智能算法训练后的检测模型更具主动预判能力，可以更快、更精准地识别威胁，降低误报率，并对未知攻击模式实现早期预警。

2.2 恶意软件检测与分析

相较于基于静态签名的方法，使用机器学习算法训练出的恶意软件检测模型，可以实现对未知威胁的精准检测。这类模型通过分析样本文件的文本特征、程序行为特征等构建检测模型，不依赖固定签名，对异种变种恶意代码具备较强的适应性[4]。例如，Invincea Labs 使用深度神经网络对超过 100 万份恶意软件样本进行模型训练，检测精确率达 99.8%；Endgame 公司将机器学习与沙箱动态运行技术相结合，对文件运行时的行为进行建模，实现了对未知勒索病毒的准确检出；卡巴斯基实验室的集成型恶意软件分类系统，使用了 12 种机器学习算法的组合，检出率达 99.7%；微软基于大规模云端威胁情报，运用卷积神经网络、LSTM 等深度学习模型，实现了对未知威胁的实时检测。

另外，部分企业还尝试使用强化学习的方法进行恶意软件检测与对抗。DeepMind 基于深度 Q 网络的恶意软件检测器，可在恶意软件不断变种的对抗环境中，通过试错自我增强，提高检测准确性。例如，以色列理工学院使用标记树状态空间表示恶意软件行为，构建强化学习模型，检测准确率高达 98%。总之，机器学习训练的恶意软件检测模型，可以突破基于签名的局限，对日益复杂多变的恶意代码实现精准、高效的检测，这是进化和增强传统防病毒技术的重要路径。未来，还将构建端到端的人工智能驱动型的智能恶意软件识别与防控系统。

2.3 自动化安全决策与响应

当威胁事件发生时，人工智能可以通过大数据分析快速做出最优的自动化安全决策，实现主动防御。例如，Darktrace 的自主响应技术针对检测到的异常，可以选择最小化干预的方式，包括流量限制、临时隔离等，避免了过激反应；Deep Instinct 构建了基于深度强化学习的决策系统，面对恶意事件时，可以选择隔离、拦截、解除等不同策略，并通过持续学习优化响应方案；微软使用老师-学生框架进行安全决策的深度强化学习，老师模型指导学生模型学习最佳反应决策。此外，一些安全厂商也在构建端到端的自动化安全响应系统[5]。Darktrace 的自主反应系统可以在检测到威胁时，立即对恶意主机实施隔离，实现毫秒级的自动化对抗，无需人工干预；Cynet 的自动响应机制能对可疑活动进行实时干预，包括解除进程、阻止文件操作等；Palo Alto Networks 的自动化安全操作系统，直接利用机器学习模型输出的检测结果，自动触发防火墙访问控制、网络隔离等防御措施。

总体来说，人工智能驱动的自动化安全决策与响应系统，可以代替人工分析判断，对疑似威胁实时做出并执行最优防御策略。这不仅缩短了事件响应时间，也避免了人为错误，提升了网络安全防线的主动防御能力。

3 基于人工智能的威胁检测与防护系统设计

3.1 数据采集和预处理

构建高效的基于人工智能的威胁检测系统，需要大量高质量的样本数据进行模型训练。数据采集可以从多个维度进行，包括通过网络流量捕获、沙箱执行、蜜罐集聚等方式搜集各类恶意样本；也可以从安全厂商、病毒总结平台获取已标注的恶意文件样本；还可以收集正常业务系统的运行日志、网络流量等建立正常行为基线。例如，Darktrace 公司声称其自主神经系统训练的样本规模超过 1000 亿；卡巴斯基实验室收集了上亿量级的恶意代码样本用于训练；INVINCEA 实验室的深度学习恶意软件检测系统使用了超过 1500 万个训练样本。这需要长时间的样本搜集积累和标注工作。

数据预处理则需要对采集到的训练样本进行清洗、平衡和特征工程等。移除错误标注的样本、处理样本不平衡问题、数据增强技术合成等方法可用于增强数据量。然后，需要解析样本文件，提取文本特征、程序行为特征、功能调用序列等有效信息，并进行编码、归一化处理，转换为模型可直接输入的训练数据。例如，美国 Jang 等学者提取了巨量恶意软件的字符串签名特征和函数调用图特征，构建了一个规模达 150GB 的恶意软件检测训练数据集；英国 Ni 等学者针对 DNS 流量数据，经过统计流量特征提取和 PCA 降维，获得了用于异常检测模型训练的稠密流特征表达。充分的样本数据规模和高质量的数据预处理，是构建高效威胁检测模型的基础，也是目前系统构建过程中非常关键的环节。

3.2 特征提取和选择

在完成样本数据预处理后，需要对数据进行特征工程，

提取对模型有意义的特征子集。针对不同的数据类型，可以提取不同的特征。对于文件样本，可以提取字符串签名特征、Imports/Exports API特征、函数调用特征、控制流程图特征等对恶意行为具有代表性的静态程序特征。对于网络流量，可以提取源目的IP、端口特征，也可以通过统计方法提取流量的封包数量、大小、时间分布等统计流量特征。对于系统日志，可以提取系统调用、注册表访问等运行时的行为特征。然后，需要进行特征选择，移除冗余和无关特征。基于过滤法的Relief、FCBF算法可以评估特征重要性并排除冗余特征；基于包装法的RFE算法可以递归排除对模型贡献小的特征。此外，还可以采用基于压缩的特征选择方法，如基于Lasso的压缩感知能移除对模型压缩性有害的特征。一个具有区分能力的优化特征子集，可以提高模型的检测效率和泛化能力。通过特征工程的预处理样本数据，可提高后续模型训练的速度和效果。同时，特征选择也具有降维压缩的作用，减少存储和计算成本，适用于实际大规模商业系统的部署。

3.3 模型训练和评估

在完成样本数据预处理和特征提取后，需要选择合适的机器学习或深度学习模型，进行模型训练。针对不同的威胁检测任务，可以选择适用的模型。如基于静态分析的恶意文件检测，可以采用CNN深度神经网络分析程序特征，也可以用GBDT、随机森林等集成树模型对文件特征进行分类。基于网络流量分析的异常检测，可以采用自编码器、LSTM等进行时间序列预测建模。基于运行日志的攻击行为检测可以用HMM模型对行为序列进行建模。

在模型训练过程中，需要关注模型的性能指标并采取优化措施。如对CNN、RNN等深度网络采用不同网络结构、激活函数、正则化方法来提高准确率。同时，需要采用交叉验证、早停等技术防止过拟合。在模型训练完成后，在单独的测试集上评估性能，关注指标包括检测率、误报率、AUC值等。在商业部署时，可以建立增量学习机制，使用新出现的样本不断优化模型。通过不断试验和优化机器学习模型，可以持续提升威胁检测与分析的性能，逐步逼近甚至超越人工分析的效果。

4 结语

随着网络攻击手段的日益复杂，传统的基于签名和规则的安全防护手段正面临新的挑战。同时，人工智能和深度学习技术的进步为网络安全领域带来了新的机遇。若要实现以人工智能驱动的网络安全防护系统，则还存在数据采集、模型解释性、实时性等挑战，这需要进一步加大创新力度和技术积累。人工智能安全代表了网络安全技术发展的主流与方向。未来，将出现基于人工智能技术的更智能化的商业安全产品，使网络世界变得更安全。

参考文献

[1] 张雨.某机构网络安全升级典型案例[J].网络安全和信息化，2020(2)：120-123.

[2] 朱平哲.网络入侵检测与防护算法系统的实现[J].安徽电子信息职业技术学院学报，2019，18(3)：7-12.

[3] 安星硕，曹桂兴，苗莉，等.智慧边缘计算安全综述[J].电信科学，2018，34(7)：135-147.

[4] 倪建武.计算机网络信息安全威胁及其防护策略[J].信息与电脑(理论版)，2012(16)：19-20.

[5] 邱通平.基于行为的企业内网威胁检测技术分析[J].计算机光盘软件与应用，2012(5)：113，105.

(上接第119页)

好的师生关系，使学生能有效借助相应的平台进行及时反馈，帮助教师调整教学进度、教学内容以及教学方法等，真正实现面向全体学生的公平教育。

3.5 信息技术在教学管理中的应用

教学管理是教学工作的重要环节，学习不同的学科知识，需要不同教师来完成相应的教学任务，但不同的教师对课堂的管理存在差异。如果不能对学生进行有效的约束，则会在很大程度上因课堂教学秩序而影响教学质量，借助信息技术来进行教学管理，可以规避因教学管理的疏忽而造成的教学质量低下。班主任作为教学管理的主要参与者，对学生的震慑力更强，其可以借助信息技术来进行有效监管，督促学生认真参与课堂教学，从而为教师带来良好的课堂秩序，有效推进教学工作[6]。

4 结语

在信息化时代，教育信息化对教学改革工作非常重要。因此，在学科教学改革过程中，学科教学与信息技术的融合成为一种必然趋势。但在实际的融合过程中，应注重策略和方法，使信息技术更好地辅助学科教学工作，提升学科教学质量。本文深入分析了学科教学与信息技术的融合，提出了学科教学与信息技术融合的有效策略。通过构建信息化教育平台，并在学科教学中应用信息技术手段，可以推动学科教学改革工作的不断深入，提升学科教学质量。

参考文献

[1] 杜鹃，王琳琳.人工智能技术在教育中的应用研究[J].物联网技术，2023，13(6)：157-159.

[2] 郭锦松.现代教育技术在小学数学课堂教学中的应用[J].教师博览，2023(15)：67-68.

[3] 伍佩仪.运用现代教育技术优化小学美术教学[J].家长，2023(4)：121-123.

[4] 王谦.依托现代教育技术优化小学科学实验教学[J].小学生(下旬刊)，2023(1)：46-48.

[5] 王丽.现代教育技术与小学数学课堂的有效整合[J].西部素质教育，2022，8(20)：146-148.

[6] 吕能新，吕绍瑜.论现代教育技术与小学语文教学的深度融合[J].安徽教育科研，2022(26)：90-92.