# 基于人工智能的网络安全威胁检测技术研究

孙少山[1]　范慧博[1]　姜禹含[2]　周雪琪[3]
1.国家电投集团内蒙古能源有限公司　内蒙古通辽　028011
2.蒙东协合新能源有限公司　辽宁沈阳　110179
3.北京中企时代科技有限公司　北京　100086

**摘　要：**在网络攻击朝着智能化、隐蔽化方向发展的时候，依靠静态规则和特征识别的办法，已经难以满足日益严重的安全需求，本文围绕着人工智能技术在互联网空间安全领域内运作机制的问题，全方位展开细致探究，文章对人工智能的进步历史及其理论根基做了详细梳理，而且全面归纳了当下现存的威胁检测办法所存在的缺点以及遇到的实际困难，接着针对机器学习与深度学习这类算法在异常行为识别、恶意代码种类划分这些重要任务上怎样具体实施，特别注重分析了联邦学习和生成对抗网络等前沿技术和经典AI模型的融合更新潜能，研究发现整合人工智能技术，可以提升自动化处理的能力水平，在提高对不熟悉威胁的识别精准度并加强数据保护措施等方面，有助于显著改善网络安全防御系统适应外部环境变化的水平与精确程度，从而为创建出更加有效的动态协同安全防范体系赋予了有力的技术保障基础。

**关键词：**人工智能；网络安全；威胁检测；机器学习；深度学习；入侵检测

## 一、人工智能技术概述

### （一）人工智能的基本概念与发展历程

是一种融合多学科知识的人工智能技术体系，主要目的是给机器赋予像人一样的感知、推理、学习与决策功能，自从1956年达特茅斯会议上首次提出这个概念之后，这一领域就不断经历过几次技术跃迁和调整研究方向的情况，在早期的时候是以符号主义作为理论基础去构造专家系统以模拟逻辑推理过程，当时开发出一个叫DENDRAL系统的产物就是典型示例之一，这推动了人工智能针对特定场景展开应用探索的发展趋势出现，在进入20世纪90年代之后伴随着大数据数量增长且计算能力明显提高的大环境下，连接主义渐渐成为主流研究框架，在此期间依靠机器学习为背景的各种算法能够凭借数据来促使自身持续改善，而且确实也做到了脱离实验室范畴迈向实际应用领域的成功转变情况发生。

进入21世纪之后，深度学习领域取得的重大突破，促使人工智能技术迈入了一个全新的发展阶段，2012年AlexNet在图像识别任务上表现出色，这显示了深层神经网络处理复杂数据的巨大潜力，于是它开始逐渐向自然语言处理以及计算机视觉等关键领域延伸，当下，人工智能正在沿着从感知智能走向认知智能，并最终达到决策智能的技术发展路线前进，它的应用范围持续扩大，给解决各种复杂问题赋予了更新颖的方案。

### （二）人工智能的关键技术

人工智能技术体系可大致分为理论基础和技术应用两个重要部分，与网络安全紧密相关的技术被总结成三种重要方向，第一，机器学习中监督学习类型的应用例如支持向量机，逻辑回归，无监督的学习方法有K-Means聚类、DBSCAN等等手段，借助对已有数据的

个人简介：
1.孙少山（1971.08——），男，蒙古族，河北望都人，本科学历，高级工程师，国家电投集团内蒙古能源有限公司科技与数智化部主任。主要研究方向为火电厂生产全过程管理，科技与数智化管理。
2.范慧博（1992.04——），男，汉族，吉林辽源人，男，本科学历，工程师，国家电投集团内蒙古能源有限公司科技与数智化部科技管理。主要研究方向计算机科学与技术，新能源工程与科学。
3.姜禹含（1998.04——），女，汉族，吉林松原人，硕士学历，工程师，蒙东协合新能源有限公司。主要研究方向计算机科学与技术，新能源工程与科学。
4.周雪琪（1996.09——），女，汉族，内蒙古通辽人，本科学历，助理工程师，北京中企时代科技有限公司综合部人力资源管理。主要研究方向为企业战略管理、数字化转型。

数据特征进行提取来完成分类预测的功能，在深度学习层面，以深神经网为主体的各类算法包括卷积神经网络、循环神经网络及自编码器等多个分支，有着自适应的特征抽取特性，且能够应对复杂的高维度非线性问题，在技术层面，跨域技术的融合包括联邦学习所采取的分布式模式保证着数据的安全传输，GAN则是利用生成器和判定机博弈过程产生更优数据成果，在强化学习当中则是依靠动态变化策略选择提高系统匹配水平。

多模态技术协同整合之后，人工智能系统在网络环境下应对海量异构数据的水平能够得到明显改善，而且可以准确找出正常行为模式里的异常之处，在新型攻击手段层出不穷的情形下，保证其检测性能始终处于稳定可靠的状况当中。

**（三）人工智能在网络安全领域的应用现状**

人工智能技术在网络安全防护方面的应用逐渐深入，在恶意代码检测层面，把机器学习算法同样本的静态属性（文件哈希值，字符串特性）以及动态行为特性（系统调用序列）融合起来，能有效地改进病毒和木马的识别精准度，同传统的依照特征识别的办法相比，它的识别准确率提升了大约30%。就网络入侵检测而言，借助深度学习的智能化手段可以立即处理大量网络流量的数据，并且精确地辨别出SQL注入或者分布式拒绝服务攻击之类的危险情况，某个商用系统应用了卷积神经网络加上循环神经网络相配合的体系结构规划，而且将响应的时间缩减到以毫秒为单位的程度。

## 二、网络安全威胁检测技术现状

**（一）传统网络安全威胁检测技术**

传统的威胁检测技术大多靠规则匹配，特征提取之类的手段来构建模型，基本上可归类为三种常见模式，第一种是依循签名的检测体制，凭借预先设置的攻击特性库执行网络流量或者文件的精确比对，譬如防火墙端口过滤规则，杀毒软件引擎里面的病毒特征码，它的优势是对已知威胁有较高的识别精准度，但应对不知晓的攻击形式以及零日漏洞却显得比较吃力。第二种则是依赖异常表现的监测办法，凭借动态阈值评判是否存在脱离了平常的行为模式，比如系统CPU资源存在起伏，异地登入的情况等等，然而过高的误报比例常因各种参数设定不确定引发；第三种即是以状态感知型框架为主，在对网络互动的上下文环境进行全面考量之后作出操作是否合法性的判定，并且增强了应对复杂事项的灵活性，不过当碰上诸如加密交流还有种种麻烦的坏事时依然显得捉襟见肘。

这种技术依靠人工制定的规则和特征参数，但是现在的网络攻击变得非常自动化，而且种类繁多，这种技术规则更新慢，检测范围小等缺点，很难达到动态防御体系所需要的实时反应能力和全面覆盖的要求。

**（二）网络安全威胁检测面临的挑战**

当下网络安全威胁有着动态发展的显著特点，这就给监测工作带来严重挑战，攻击手段的智能化倾向特别明显，一些黑客利用人工智能技术制造对抗样本，通过小范围的特征改变来避开现有的防御机制，有研究显示，在对抗性攻击环境中，传统的机器学习模型识别正确率大概会下滑40%以上，威胁形式的隐蔽性主要表现在高级持续性威胁（APT）当中，这种攻击一般采取“低烈度，长周期”的渗透策略，在很长一段时间里慢慢执行攻击行为，于是就冲破了依靠即时观察的传统防护体系对于完整攻击流程的跟踪能力。

**（三）人工智能技术在网络安全威胁检测中的优势**

人工智能技术依靠数据推动的自适应学习机制，在破解传统网络安全检测难题时表现出了明显的优势，从检测效能上看，机器学习模型可以很好地应对大规模网络数据流，某大型数据中心统计的数据表明，其AI系统日均处理量达到10TB，相比传统办法提升了大约5倍；对于未知威胁识别而言，深度学习算法冲破了人工规则的限制，凭借生成对抗网络（GAN）生成的合成样本来改良训练集，把零日攻击检测精确度提升到88.2%；在动态适应性上，强化学习利用与攻击场景的即时互动来改善决策逻辑，联邦学习则在保证数据隐私的同时推动跨域协同防御，某多机构联合防护平台经过这种技术改进之后，威胁情报共享效率提高了40%。

## 三、基于人工智能的网络安全威胁检测技术

**（一）基于机器学习的威胁检测方法**

机器学习依靠数据驱动的预测模型来开展威胁检测，其中涉及的关键步骤包含安全事件建模、数据预处理、特征提取以及模型训练，就网络安全而言，针对已知攻击类型的分类问题，监督学习方法中的随机森林和支持向量机（SVM）被广泛采用，对网络流量数据执行源IP地址、端口信息以及包长度等核心特征的提取之后，建立起分类模型用以辨别正常通信与潜在威胁行为，通过研究显示，在CICIDS2017数据集当中，SVM算法的检测准确率可以达到92.3%。

无监督学习方法在未知威胁检测上具有明显应用价值，K-Means聚类算法凭借对流量特征间相似度加以分类，把偏离中心区的数据视为潜在威胁，该方法在僵尸

网络识别里得到过实践检验，半监督学习结合了有标签和无标签数据，依靠少量已知样本引导模型找出新出现的攻击模式，从而解决恶意样本稀少的问题，特征工程是核心环节，采用归一化、降维等预处理手段改善数据质量，有研究显示，网络流量PCA之后，模型训练效率提高35%，这些技术手段一起构成了完整的机器学习威胁检测框架，在提升检测精度和运行效率上形成协同作用。

**（二）基于深度学习的威胁检测技术**

深度学习依靠多层神经网络的非线性映射，做到了特征提取的自动化进程，冲破了传统机器学习对人工特征规划的依赖，显示了在复杂威胁检测方面的明显长处，卷积神经网络（CNN）凭借卷积层和池化层来提取局部空间特征，很适合做网络流量的空间模式剖析，把输入数据变成二维矩阵以后，就能精确识别分布式拒绝服务攻击（DDoS）之类的异常流量特点，有研究显示它在UNSW-NB15数据集上的检测精确率可以做到95.8%，循环神经网络（RNN）及其变种长短期记忆网络（LSTM），由于擅长处理序列数据，所以备受重视，可以很好地捕捉攻击行为的时间依赖性，在高级持续性威胁（APT）的追踪以及恶意软件动态行为分析方面表现不错，实验数据表明LSTM对多阶段攻击的识别准确率比传统方法高大概27%。

自编码器凭借重构误差评价机制做到异常检测，在加密流量分类上表现出明显的优势，可以精确地辨别合法和非法的通讯，无需对原始数据实施解析，生成对抗网络（GAN）通过生成器来生成高仿真的恶意样本，很好地解决了训练集分布偏斜的现象，NASA所开发的CyberGAN模型生成的数据同真实流量特征的契合度高达98%以上，这些深度学习技术的应用极大地改善了安全防御体系在复杂环境中的适应性以及潜在威胁的感知能力。

**（三）基于人工智能的威胁检测技术的融合与创新**

单一的人工智能技术存在固有的局限性，于是促使多模态融合创新得以发展，进而形成了多层次协同监测体系，在技术整合方面，机器学习和深度学习进行了深度耦合，从而实现功能互补，通过随机森林完成了特征提取之后，再把数据交给卷积神经网络（CNN）开展分类处理，某种混合架构在恶意软件识别方面的F1分数明显提高到0.93，联邦学习同差分隐私技术相互结合，很好地解决了跨域数据协作时的隐私安全问题，在医疗领域既保障了患者信息的安全，又使得威胁检测准确率维持在92%以上。

自适应学习机制是创新研究的主要议题，依靠在线优化动态调节模型参数以应对新型威胁，某个动态防御框架采用强化学习算法即时更新监测策略，即使在攻击模式改变的时候，也能维持85%以上的检测精确率，可解释性技术的加入极大地改善了算法的可信度，LIME和SHAP这些工具凭借对特征重要性的可视化分析，使得安全专家可以深刻认识人工智能的决策过程，某家银行执行该方案以后，安全事件回应速度加快了大约40%，这种跨学科融合既改良了威胁察觉水平，又促使网络安全保护朝着现代主动预测方向转变。

**总结**

人工智能技术给网络安全威胁检测带来了革新性的解决办法，通过融合机器学习准确的分类能力，深度学习自行提取特征的机制，多模态数据协同优化的策略，从而冲破了传统检测手段的固有限制，从技术应用的角度来讲，依靠人工智能的监测平台在效率改进，环境适应性，功能拓展等方面有着明显的优势，而且凭借生成对抗网络（GAN）提升对未知威胁的识别精准度，利用联邦学习加强数据隐私保护这些实际例子，为创建动态防御体系给予了可信的技术支撑。当下人工智能技术在网络安全方面的应用碰上不少难题，对抗性攻击存在模型失效的风险，黑箱特性使算法可靠性降低，数据安全和隐私保护之间的矛盾也没得到解决，以后的研究要着重关注这几个方面：打造更可靠的防御体系来提高系统的稳定性，研发具备高透明度的安全检测办法以加强决策的可解释性，形成跨平台协同的标准化防护机制，经过不断的技术更新和实际操作改善，人工智能大概会在网络安全方面取得突破性成果，给数字化社会的安全运行给予强有力的保证。

**参考文献**

[1]王炜森.面向物联网安全威胁的人工智能检测技术研究[J].电脑编程技巧与维护，2024（12）：129-131.

[2]沈朗捷.基于人工智能的港口网络安全威胁检测与防御技术[J].中国信息界，2024（9）：30-32.