基于人工智能的网络信息安全威胁检测技术研究

胡梁嘉
杭州安恒信息技术股份有限公司　浙江杭州　310000

【摘　要】为突破传统威胁检测技术的瓶颈，解决无法识别未知攻击和检测效率低等方面的现实难题，本次研究中提出基于人工智能的网络信息安全威胁检测技术。文中首先明确了人工智能在网络信息安全威胁检测中的适配性，并结合网络安全威胁检测的技术特点梳理智能检测流程，分别探究机器学习与深度学习在威胁检测中的应用方法，通过实验方式，验证传统IDS检测模型、单一机器学习模型（孤立森林）、单一机器学习模型（随机森林）与CNN-BiLSTM模型，在精确度、准确度、召回率与误报率等关键性能指标方面的差异，进一步验证了智能威胁检测技术的应用成果。

【关键词】人工智能；网络信息安全；威胁检测技术；孤立森林；CNN-BiLSTM模型

Research on Network Information Security Threat Detection Technology Based on Artificial Intelligence

Hu Liangjia
Hangzhou Anheng Information Technology Co., Ltd.　Hangzhou, Zhejiang Province　310000

【Abstract】To break through the bottlenecks of traditional threat detection technologies and solve the practical problems such as inability to identify unknown attacks and low detection efficiency, this research proposes an artificial intelligence-based network information security threat detection technology. Firstly, this paper clarifies the adaptability of artificial intelligence in network information security threat detection and summarizes the intelligent detection process based on the technical characteristics of network security threat detection. It then explores the application methods of machine learning and deep learning in threat detection. Through experimental methods, it verifies the differences in key performance indicators such as accuracy, precision, recall rate, and false alarm rate among traditional IDS detection models, single machine learning models (Isolation Forest), single machine learning models (Random Forest), and CNN-BiLSTM models. This further validates the application results of intelligent threat detection technology.

【Key words】Artificial Intelligence; Network Information Security; Threat Detection Technology; Isolation Forest; CNN-BiLSTM Model

数字经济背景下，网络安全问题日趋复杂，网络攻击呈现出隐蔽化和多样化的特征，且会随着网络技术的迭代而更新，经常让人们防不胜防。以往采取的传统威胁检测技术对于已知攻击特征库的依赖度较大，一旦面临未知威胁便无法立即响应，甚至在已经发生数据泄漏和系统异常后才被发现，网络信息安全威胁防御的主动性不强。基于人工智能的网络信息安全威胁检测技术则可借助强大的学习功能和模型算法不断适应新的网络数据环境，有助于实现威胁检测模型自适应的目标。因此，研究智能威胁检测技术对于提高网络安全防护水平具有积极意义。

1.人工智能在网络信息安全威胁检测中的适配性分析

一是在传统的威胁检测技术中，通常是采取特征匹配与规则推理的方式识别网络安全威胁，缺陷表现如下：仅支持对规则库中已有攻击行为的识别，针对那些经过变形处理的代码无效；特征匹配的过程耗时较长，特别是在数据信息量暴涨的今天，采取特征匹配方式难以实时检测信息安全威胁；经常出现误判现象，如在正常的运行环境下出现流量峰期易被判定为异常攻击，一定程度上增加了安全运维的工作量。而人工智能技术的融入，可通过机器学习的手段持续优化威胁检测算法，提高检测效率的同时，还能增强系统响应能力。

二是网络环境日趋复杂，对网络信息安全威胁检测技术提出了智能检测的基本需求。第一，应该准确辨别正常行为与异常攻击行为，支持对信息安全威胁的高效检出；第二，应具备对未知威胁和攻击行为的识别能力，不依赖特征对照物便可准确识别新的攻击行为；第三，在数秒内便可完成对海量数据信息的扫描识别与检测，实现对信息安全威胁的实时检测；第四，能够持续学习和了解攻击行为与模式，自动更新检测模型和检测策略，适应不同的网络数据环境。

三是人工智能技术中的机器学习功能、深度学习功能与强化学习功能等可满足当前的网络信息安全威胁检测需求。机器学习是基于历史数据来挖掘各种攻击行为的潜在特征，可高效识别攻击行为并完成分类；深度学习是在机器学习的基础上，挖掘网络数据的抽象特征，可以有效解决因特征不充分无法检测信息安全威胁的问题；强化学习支持信息安全威胁模型的自行迭代，即结合网络环境以及攻击模式的变化动态调整检测策略，具备较强的自适应能力[1]。

2.基于人工智能的网络信息安全威胁检测技术

2.1 智能检测的关键流程

基于人工智能的网络信息安全威胁检测（流程见图 1）

是通过数据采集、预处理、特征提取、推理和响应机制等一系列操作实现威胁检测与预警的目标。其中的数据采集环节主要是对终端服务器、系统日志和设备等原始数据的采集，包括流量数据、操作行为和系统运行数据等，可以作为模型训练与检测的基础数据；数据预处理则是指对原始数据的去重、降噪，并对数据进行结构化转换的过程，可保障数据质量；特征提取指的是提炼预处理数据中的关键网络威胁特征，例如流量特征、行为特征和系统状态等；模型推理是指将提取的特征数据输入 AI 模型后，利用模型判断当前的行为是否存在异常；响应机制是指当模型发现异常后，触发警报，通过邮件或者弹窗等方式做出提醒，同时驱动网络安全设备做出拦截动作。

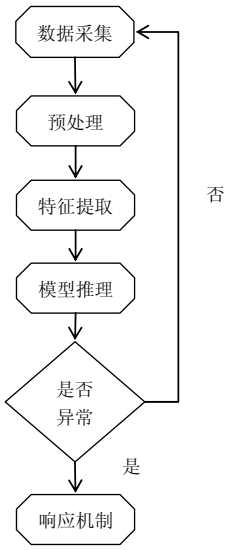

图 1 智能检测流程图

2.2 机器学习在网络信息安全威胁检测中的应用

2.2.1 监督学习在已知威胁中的应用

监督学习的实质是，将已经标注好的攻击数据作为样本数据进行模型训练，确保模型掌握已知攻击的特征与模式，以实现对同类型攻击行为的准确识别。其应用原理是先给定一个训练集：

$$D=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\} \quad (1)$$

式（1）中，x_1 表示网络数据的特征量；$y_1\in\{0,1\}$ 表示标签，（0 为正常，1 为攻击），该模型是基于映射关系 $f:X\to Y$ 来预测新的特征向量 x_{new}。

现阶段较为常用的监督学习算法包括卷积神经网络法、随机森林法和 XGBoost 等，在密码暴力攻击以及 DDoS 攻击中的作用较为显著，可以通过模型学习的方式确定网络异常连接数与异常流量峰值，实现对外部攻击的准确识别[2]。

2.2.2 无监督学习在未知威胁中的应用

无监督学习与前者的区别在于，无需提前标注数据，仅需学习正常网络数据便可总结出常规行为的模型特征，在威胁检测中只需识别出现偏离常规行为的特征即可。其应用优势表现为可对未知威胁进行有效检测，能够弥补传统威胁检测中的短板。本次研究中选用孤立森林算法识别异常网络行为，是基于异常点容易被孤立的原理构建随机决策树来检测异常数据。为能实现量化异常的目标，可利用如下计算公式：

$$s(x,n)=2-\frac{E(h(x))}{c(n)} \quad (2)$$

式（2）中，$h(x)$ 表示样本 x 从根节部位至叶子结点的路径数量，直接反映被孤立样本的难易程度，其中异常样本的 $h(x)$ 小于正常样本；$E(h(x))$ 表示 x 在森林中所有决策树平均路径长度；$c(n)$ 表示样本量 n 的正常样本路径长度。

异常程度的判断标准为，当 $s(x,n)\approx 1$ 时，样本异常程度偏高；当其 ≈ 0.5 时，代表样本正常；当其 <0.5 时，表示样本无异常特征。

2.3 深度学习在网络信息安全威胁检测中的应用

深度学习的应用实际上就是构建多层神经网络，自动提取网络数据中深层的抽象特征，有效打破传统威胁检测的限制，摆脱对特征工程的依赖。当前的网络威胁更为复杂，此种威胁检测技术的适配度更高。本次研究中，决定联合应用卷积神经网络（CNN）与双向长短期记忆网络（BiLSTM），形成 CNN-BiLSTM 模型，旨在充分发挥二者的空间特征提取作用与时间序列特征提取作用，满足现阶段以时空特征为表现的威胁检测场景[3]。

2.3.1 基于 CNN 的空间特征提取

该模式是利用卷积层与池化层进行空间特征数据提取，公式为：

$$y_{i,j}^{l}=\sigma\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}w_{m,n}^{l}\cdot x_{i+m,j+n}^{l-1}+b^{l}\right) \quad (3)$$

式（3）中，$y_{i,j}^{l}$ 表示 l 层卷积层输出特征图；$w_{m,n}^{l}$ 表示卷积核权重；$x_{i+m,j+n}^{l-1}$ 表示 l–1 层输入特征；b^{l} 表示偏置项；σ 表示激活函数。

进行网络信息安全威胁检测中，该模型是通过对网络流量表的卷积操作，提取特征矩阵中的长度、协议和传输方向等关键空间特征，据此判断是否存在异常攻击。

2.3.2 基于 BiLSTM 的时间特征提取

BiLSTM 分别为前向、后向 LSTM 两个部分，可实现对前后向时间序列的全面捕捉，在一些典型时间序列的网络数据安全检测中较为适用。通过对前向 LSTM 正向时序网络数据特征和后向 LSTM 反向时序网络数据特征的融合，可以形成一个完整的时序特征，通过观察时序特征便可识别出有无多次反复登录等异常操作，可准确捕捉暴力破解密码的攻击行为[4]。

2.3.3CNN–BiLSTM 模型

该模型的工作流程为，数据输入–CNN 特征提取 – 池化层–BiLSTM 时序特征捕捉 – 全连接 – 输出。先是利用 CNN 层提取输入数据的空间特征，经由池化层处理后，进入 BiLSTM 时序特征捕捉环节，最后在全连接层融合特征数据，输出最终的判断结果。

3.智能威胁检测模型的性能验证

实验数据由真实网络数据与攻击流量数据共同组成，保障 DDoS、SQL 和远程代码等多种网络攻击行为的检测，并将样本数据划分为训练集、测试集与验证集，所有数据均经过标准化处理。利用 Python 搭建实验平台。为验证人工智能威胁检测模型的应用效果，设计了 3 组检测模型，分别为传统 IDS 检测模型、单一机器学习模型和 CNN-BiLSTM 模型，实验中所有模型均处于相同的网络环境下，通过对比研究可进一步验证智能检测模型的应用成果。实验结果见表 1：

表 1　各组模型性能评价指标（%）

模型	准确率	精确率	召回率	F1 值	误报率
传统 IDS 检测模型	82.5	78.3	75.6	76.9	8.9
单一机器学习模型（孤立森林）	92.1	90.5	88.7	89.6	3.2
单一机器学习模型（随机森林）	94.8	93.7	92.5	93.1	2.1
CNN-BiLSTM 模型	98.2	97.8	97.5	97.6	1.5

从上表数据中不难看出，CNN-BiLSTM 模型无论是在准确率还是在精确率方面都独具优势，相较于传统 IDS 模型分别提升了 15.7 和 19.5 个百分点，这标志着深度学习的混合模型架构可以清晰地区分攻击行为与正常网络行为。此外，CNN-BiLSTM 模型的误报率仅有 1.5%，可很大程度上控制无效告警率，起到减轻网络安全运维压力的重要作用。而在单一机器学习模型中的随机森林与孤立森林相比，随机森林对攻击行为的识别效率更高，究其原因是随机森林是基于已知攻击的威胁检测模型，与本次研究的威胁场景更为匹配，而孤立森林则适用于未知威胁检测的场景。

4.人工智能网络信息安全威胁检测技术的应用案例

4.1 案例背景

某金融公司的内网遭受 APT 攻击，攻击表现为利用钓鱼邮件植入代码，并利用网络漏洞进行横向移动，目的是窃取核心金融数据。公司已有的传统 IDS 检测模型未及时发现异常攻击行为，致使公司内网被连续攻击多天，最终是核心服务器发现异常访问。

4.2AI 检测技术的应用

这一攻击行为引起了公司高层的重视，立即组织部署 BiLSTM 与孤立森林模型搭建了智能威胁检测系统，对公司内网日志以及数据流量进行全面追溯与检测。通过对用户登录行为的分析发现有多个员工账号存在凌晨登录的情况，正常来讲，登录时间应为白天 9:00～18:00 之间，经计算异常分数≈0.98，结合异常判断标准将其判定为高度异常；利用 BiLSTM 模型分析内网流量时序发现有多终端同时向核心服务器发送数据包的行为，由于数据包较小且模拟了常规的传输频率，流量较为规律未被传统 IDS 检测模型识别。经过模型验证，发现异常攻击行为，智能威胁检测系统立即触发警报，同时驱动内网防火墙拦截异常 IP，通过阻断流量移动的路径和锁定钓鱼邮件的植入节点一举解决了 APT 攻击问题[5]。

4.3 案例分析

在该案例中，之所以传统 IDS 检测模型无法发挥作用是由于攻击者模拟了正常网络数据行为，加之规则库中没有此类攻击特征，无法通过特征对比准确识别攻击行为。然而，孤立森林通过对用户行为的分析准确捕捉了异常登录行为，同时通过对网络流量模式的检测，提取了异常行为特征，并准确定位异常节点，整个攻击处理过程仅用时 15min，有效杜绝了该金融公司核心金融数据的泄露，为公司减少了大量的经济损失。

结语：

基于人工智能的网络信息安全威胁检测技术可以打破传统威胁检测中的技术瓶颈，实现对未知威胁的准确识别与定位，相对来说检测效率和检出率均优于传统威胁检测技术，可以将智能威胁检测技术作为打造网络信息安全防护体系的关键技术手段。此外，应该认识到网络环境趋于复杂化、数据海量化和数据行为多元化的层面发展，为网络信息安全威胁检测带来了很多新挑战，今后仍需不断探究智能技术与网络信息安全技术的融合方法，开发出自适应性较强和支持智能化检测的技术体系，为网络安全提供可靠的技术支撑。

参考文献

[1]曹铮,李志,宋益博.基于人工智能的电子信息安全检测与防护技术研究[J].移动信息,2025,47(6):223-225.
[2]张作东.基于人工智能的航空信息安全防护技术研究[J].新潮电子,2025(4):22-24.
[3]王瑞涵.基于人工智能的威胁检测与防护系统[J].移动信息,2024,46(1):129-131.
[4]周志强.基于知识图谱模型的网络威胁识别与预测系统的研究与实现[D].北京邮电大学,2024.
[5]张济鸿,谭龙广,傅东波.基于人工智能的计算机网络信息安全防护模式研究[J].移动信息,2025,47(8):192-194.