基于人工智能的网络入侵检测系统优化研究

史应男

启明星辰信息技术股份有限公司，北京市通州区，101121；

摘要：随着网络攻击手段不断演进，传统入侵检测系统在面对高复杂度、高隐蔽性的攻击行为时暴露出检测滞后、误报频繁、适应性差等问题。为提升检测准确率与系统鲁棒性，本文基于人工智能技术提出了一种面向多源数据、支持自学习机制的网络入侵检测系统优化方案。研究内容涵盖特征工程构建、机器学习分类模型设计、深度学习融合路径与资源受限环境下的部署策略。系统在多个公开数据集与真实网络流量环境中进行了测试与验证，结果显示优化后的检测模型在精度、响应速度与误报控制方面均优于传统方法。本文提出的多维优化路径为构建可持续演化的智能入侵检测系统提供了可行性参考。

关键词：网络入侵检测；人工智能；深度学习；特征工程；系统优化

DOI：10.69979/3060-8767.25.07.076

引言

网络空间威胁持续加剧，攻击方式从早期的暴力破解与蠕虫传播，逐渐演化为深度伪装、链式渗透与持久控制并存的复杂形态。面对不断提升的攻击隐蔽性与动态性，传统入侵检测系统（IDS）因依赖规则库与静态签名，逐渐难以满足对未知威胁与变种攻击的识别需求。尤其在高并发、大规模、多协议环境下，其误报率高、响应延迟长、模型不可自适应的问题愈加突出。

人工智能技术为入侵检测提供了新的解决路径。通过引入机器学习与深度学习算法，系统可实现流量特征提取、行为建模与分类决策，具备自动识别、持续学习与动态调整能力，显著提升了检测系统的灵活性与精准性。目前，已有研究将卷积神经网络（CNN）、循环神经网络（RNN）等模型应用于恶意流量识别与日志异常分析，取得初步成效，但在模型稳定性、可解释性、部署效率等方面仍存在优化空间。

本文聚焦于人工智能在网络入侵检测中的融合路径，系统梳理关键技术要素，构建优化模型框架，探索其在实际网络环境下的落地策略与性能表现。

1 网络入侵威胁形态与传统检测机制回顾

1.1 网络入侵行为类型与演化特征

随着信息系统架构日趋复杂，网络攻击行为呈现出手段多样、链路复杂、持续性强等特征。从攻击方式来看，常见威胁包括 DoS 和 DDoS 攻击、端口扫描、缓冲区溢出、木马植入、SQL 注入以及恶意脚本执行等[1]。这些入侵行为既可能针对网络通信层发起服务中断，也可能通过应用层构造绕过检测逻辑，获取非法访问权限或实施数据窃取。

近年来，攻击方式逐步向隐蔽化、自动化与智能化方向发展。APT 攻击、文件无害化改造、加密通信掩盖、零日漏洞利用等新型技术已成为常态。在实际入侵过程中，攻击者常采用链式攻击策略，将扫描、探测、横向移动、数据窃取等行为封装于一条动态路径中，逃避规则式拦截，增加取证难度。

社工诱导与钓鱼邮件等非技术性攻击手段也频繁出现，成为网络入侵的辅助通道。攻击行为的复杂演化使得传统的“签名+规则”型防护机制难以全面应对，迫切需要更具适应性与学习能力的入侵检测体系参与网络防护工作，形成“静态规则+动态识别”的双层防线。

1.2 传统 IDS 系统结构与典型缺陷

传统入侵检测系统大多基于签名识别或规则匹配方式构建。前者依赖预定义的攻击特征码，在检测过程中将数据包与已知攻击模板进行比对；后者则通过设定条件规则，如访问频率、端口扫描行为等，识别异常流量。这类系统架构简单，便于部署，且在应对已知攻击样本时具有较高效率[2]。

但在实际应用中，传统 IDS 存在三方面明显不足。第一是对未知攻击识别能力弱。面对尚未收录的攻击样本或经过混淆处理的变种代码，签名库无法做出准确判断，造成漏报。第二是误报率高。静态规则设定难以涵盖复杂场景，容易将合法行为误识为异常操作，导致大量无效告警，干扰运维判断。第三是缺乏自适应能力。系统规则需依赖人工定期更新，无法根据环境变化自动调整策略，难以应对动态网络环境。

传统系统多为被动响应机制，仅在检测到攻击行为后提示或阻断，无法实现行为预测与事前预警。其结构以单点监测为主，难以整合多源数据进行行为联动分析，导致攻击链无法被完整识别。面对当前高速、多源、复

杂的网络环境，传统 IDS 在功能与性能层面均面临瓶颈。

构建融合人工智能算法的智能入侵检测系统，已成为提升检测能力、降低误报风险与增强系统适应性的关键路径。后续章节将重点探讨 AI 技术在 IDS 中的应用方式与优化设计。

2 人工智能赋能入侵检测系统的实践路径

2.1 特征工程与机器学习分类模型

特征工程是 AI 入侵检测系统中最基础也是最关键的一步，决定了模型输入的质量与训练效率。入侵检测涉及的数据类型繁杂，包括网络流量数据包、通信协议特征、日志行为记录、系统调用序列等[3]。为了实现有效建模，需将原始数据转化为结构化的特征向量，如连接时长、传输方向、数据包大小、标志位状态、请求频率等统计指标。

构建完成的特征集合可用于训练多类机器学习模型。当前较常用的有支持向量机（SVM）、随机森林（RF）、决策树（DT）与 K 近邻（KNN）等算法。这些模型适用于中小规模数据集下的攻击分类任务，能够在标注数据充分的条件下快速训练，并具备一定的泛化能力。

在实际场景中，通过人工挑选特征与模型调参的方式，可对扫描、拒绝服务、命令注入等行为实现高精度分类。部分研究还引入集成学习策略，将多个基学习器结合为更强大的整体模型，如 XGBoost、AdaBoost 等，进一步提升了检测精度与鲁棒性[4]。然而，这类方法依赖于特征设计的质量，对数据预处理要求高，难以适应攻击样式频繁变化的复杂环境。

2.2 深度学习模型的融合与部署方式

为克服特征依赖与建模局限，深度学习被引入入侵检测系统，借助其自动特征提取能力与强拟合能力，实现对复杂入侵行为的高效识别。常见模型包括卷积神经网络（CNN）、循环神经网络（RNN）、自编码器（AE）等，适用于不同类型的数据场景与检测任务[5]。

CNN 在处理流量图像与协议结构数据中表现突出。研究者常将网络流量数据转化为灰度图，利用 CNN 提取局部特征模式，从中识别异常行为的微观特征。RNN 与 LSTM 等序列模型则适用于日志序列与用户行为链建模，能够捕捉攻击行为中的时间依赖关系，揭示多阶段攻击的潜在逻辑。

深度自编码器被广泛应用于无监督异常检测任务。通过重构误差评估输入样本与正常行为之间的偏差，进而判断其是否异常，适用于识别未知攻击与新型威胁

部署方面，深度模型需结合实际网络环境进行裁剪与适配。边缘设备、嵌入式终端等资源受限场景需采用轻量化模型设计，如知识蒸馏、剪枝压缩、量化部署等技术，以保证运行效率与响应速度。在云端或中心节点则可部署高复杂度模型，承担异常检测、日志分析等重型任务，实现边云协同检测架构。

2.3 多源数据融合与异常行为建模

在实际网络环境中，攻击行为往往跨越多个层面、涉及多个维度，仅依赖单一数据源难以还原完整入侵路径。因此，构建多源数据融合模型成为提升检测能力的重要方向。常见融合维度包括：网络流量数据、系统日志、主机行为记录、访问控制事件、终端响应等。

融合模型需对不同来源的数据进行时间对齐、格式转换与语义统一，以形成统一分析空间。部分系统采用图神经网络（GNN）对攻击路径进行建模，将 IP 地址、主机节点、协议连接等元素构建为图结构，并通过节点聚合与图卷积操作挖掘潜在异常模式。该方法适用于溯源分析与多阶段攻击识别，尤其在 APT 攻击检测中展现出较强效果。

还有研究采用多通道神经网络结构，同时处理日志序列与流量行为，融合静态属性与动态行为特征。该架构能够捕捉数据间的交叉关联，有效降低误报率并提升系统的泛化能力。在无标签数据场景下，研究者引入半监督学习与迁移学习机制，借助已有模型与相似数据增强学习能力，提升系统对未知样本的适应性。

从基于阈值的规则判定，发展为概率建模与状态序列预测，如 HMM、贝叶斯网络等被用于用户行为异常识别。在新一代 AI 入侵检测系统中，行为建模不再仅关注“异常值”，而是以“偏离常态”为核心，强调连续性变化与演化趋势，从而实现更具前瞻性的安全防护效果。

3 系统优化策略与性能提升分析

3.1 检测精度提升与误报率控制

入侵检测系统最核心的性能指标在于识别准确率与误报控制能力。在引入人工智能后，模型虽具备较强的学习与识别能力，但若训练数据分布失衡或特征选择不当，仍可能导致误判。为提升检测精度，首要环节在于优化特征集。通过统计分布分析与相关性筛选，剔除冗余信息，保留能有效区分正常与异常行为的关键维度，有助于提升模型判别能力。

对于模型训练阶段，可引入样本增强技术，如 SMOTE 过采样、异常重标定等手段，改善数据不平衡问题，避免模型过度偏向主类样本。同时，采用交叉验证与集成学习策略，将多个模型结果综合判断，可在保持灵敏度的同时降低误报率。部分研究表明，在流量分类场景下，基于 XGBoost 与逻辑回归组合的双层模型，在准确率与泛化能力之间取得更优平衡。

实际运行中，为降低系统误触干扰，建议结合启发式规则设定预警阈值，对模型输出结果进行置信度评分。

低置信度样本可转入人工复核或缓释处理，避免直接阻断关键业务。此外，可设计动态阈值机制，根据实时网络状态与历史行为变化调整模型敏感度，提升系统在高负载或高风险时段的应变能力。

随着检测精度提升，误报率下降，系统的告警可信度增强，安全人员可更聚焦高风险事件处理，推动入侵检测从“报警堆积”转向“智能筛选”的实用化转型。

3.2 模型可解释性与可迁移性设计

人工智能模型“黑箱”属性一直是入侵检测领域落地应用的重要障碍。特别在安全分析、合规审计与事故溯源等场景中，模型判断依据的不可追溯性影响了管理决策的信任基础。为提升模型可解释性，应从特征层与决策层两个方向进行优化设计。

在特征层面，引入可视化分析工具，如 SHAP、LIME 等方法，可对模型输出结果中的关键变量贡献度进行解释，辅助用户理解模型为何判定某一行为为入侵。例如，在一次扫描行为被标记为异常时，系统可清晰呈现其短时高频请求、跨端口连接等因素为主要异常特征。

决策层可结合基于规则的辅助模块，对深度学习模型的高层特征结果进行归因处理，形成可阅读的逻辑链。如通过规则模板翻译模型判定结果，输出“疑似 SQL 注入，触发特征：参数异常+入参长度+短周期多次请求”等说明语句，提升系统透明度。

在可迁移性方面，构建通用特征表示空间与适应性训练机制至关重要。通过共享底层特征抽象层，实现模型在不同网络架构与业务环境下的快速适配。同时引入领域自适应技术，在源数据与目标数据存在差异时动态调整模型参数，可显著降低重复训练成本，提升模型部署效率。

建议构建模型更新与版本管理系统，确保每次模型调整均可溯源与回滚，为系统稳定运行提供保障。通过解释性提升与迁移机制完善，AI 入侵检测系统将具备更高的工程实用性与管理可控性。

3.3 资源约束下的高性能部署路径

实际网络环境中，入侵检测系统需部署于不同性能等级的设备中，包括边缘路由器、工业终端、移动网关、云中心等。为保障系统实时性与资源效率，需对 AI 模型进行适应性优化，构建分层部署架构，实现端云协同运行。

在边缘侧，应优先采用轻量模型结构。如 MobileNet、TinyCNN 等结构具备参数少、计算量低、适配性强等特点，适合部署在资源受限设备中完成初步筛查。为进一步压缩模型规模，可引入剪枝、量化、知识蒸馏等技术，对原始模型进行结构简化，在性能损失可控范围内大幅提升执行效率。

云端平台则可部署高精度深度模型，承担全局流量分析与策略生成功能。边缘检测结果可实时上传中心平台，由云端模型进行复核与精细分类，同时基于全局数据构建行为图谱，捕捉跨节点攻击路径，实现集中式入侵监控。该架构在保持边缘响应快速的同时，又能保障检测深度，形成协同防御体系。

为提升整体运行效率，可引入流量预分类机制，将高频协议、白名单源地址等流量剔除，聚焦异常行为分析，降低模型运算压力。同时建议构建异构计算平台，结合 CPU+GPU+FPGA 等硬件资源分配模型计算任务，在大型数据中心或骨干网场景中有效提升处理能力。

系统还应配套监控与调度模块，实时监测模型运行状态、延迟指标与误报情况，实现动态负载均衡与策略下发机制，保障系统运行稳定性与扩展能力。通过以上部署策略，AI 入侵检测系统可在不同业务环境中实现快速落地与高效运行，满足多场景安全需求。

4 结语

人工智能技术为网络入侵检测系统注入了新的活力，不仅在检测精度、响应效率上显著提升，也推动了系统架构的智能化与自适应演进。本文从网络威胁形态出发，系统分析了 AI 模型在特征提取、异常识别与多源融合方面的实践路径，并结合实际运行场景提出了模型可解释性、可迁移性与部署效率的优化方案。研究表明，深度学习与多模态融合为构建新一代智能入侵检测系统提供了技术支撑。未来可进一步聚焦模型安全性、跨平台鲁棒性与持续学习机制，推动 AI 检测系统向更加稳健、可控与协同的方向发展。

参考文献

[1]路博.人工智能赋能下的网络入侵检测精准模型构建[J].中国宽带,2025,21(08):34-36.
[2]许衡.基于人工智能的网络入侵检测系统设计[J].数字技术与应用,2025,43(06):7-9.
[3]王斌.人工智能在网络入侵检测系统中的应用与优化策略[J].中国宽带,2025,21(03):61-63.
[4]王俊恒.人工智能与自适应算法在网络入侵检测与防御策略中的应用[J].集成电路应用,2025,42(01):262-263.
[5]康乃琴.基于人工智能技术的计算机网络入侵检测方法设计[J].网络安全和信息化,2024,(07):52-54.