AI赋能的安全风险评估与预警系统研究

金梦依 李 源 朱志敏*
浙江大华技术股份有限公司 浙江 杭州 310053

摘 要：AI技术为安全风险评估与预警提供了新范式，通过机器学习整合多源异构数据，突破传统方法局限。研究构建了包含数据预处理、算法融合、模型训练优化及验证迭代的评估模型，设计了分层架构的预警系统，涵盖感知层、数据层、计算与应用层，实现风险实时识别、等级判定及响应触发。系统采用多重安全机制保障数据与运行安全，推动风险管理从被动应对转向主动预防，提升复杂系统安全防护韧性与效能。

关键词：AI赋能；安全风险评估；预警系统

引 言

安全风险评估与预警是复杂系统稳定运行的关键环节，传统方法受限于数据处理能力与动态监测不足，难以应对非线性风险变化。AI技术的发展为解决这一难题提供了可能，其在多源数据整合、动态建模及模式识别上的优势，可重塑风险管理模式。本文聚焦AI赋能的安全风险评估与预警系统，探讨模型构建方法，设计系统架构与功能模块，旨在通过数据驱动与算法优化，提升风险评估精度与预警时效性，为安全管理提供科学支撑。

1 AI 赋能安全风险评估与预警的重要性

AI技术的深度融入正在重塑安全风险评估与预警的范式，其核心价值在于突破传统方法对复杂数据处理能力的局限，通过机器学习算法对多源异构信息进行实时整合与分析，从而挖掘出潜藏在海量数据中的风险关联规律。这种数据处理模式不仅能够覆盖设备运行参数、环境变量、历史事故记录等多维指标，还能通过深度学习模型不断优化风险识别的敏感度，使原本依赖人工经验的定性判断逐渐转向基于数据驱动的定量评估，大幅降低了主观偏差对评估结果的干扰。在动态风险监测层面，AI系统依托边缘计算与物联网设备的协同联动，可实现对风险因子的持续追踪与动态建模，当关键指标偏离正常阈值时，能够通过预设的算法模型快速生成风险等级预判，并同步触发相应的预警机制，这种实时响应能力有效缩短了从风险出现到人工干预的时间差，为风险处置争取了关键窗口。相较于传统定期评估模式，AI驱动的动态预警机制更能适应复杂系统中风险因子的非线性变化，尤其在工业生产、网络安全等领域，可精准捕捉设备老化、异常访问等细微风险信号，避免因小概率事件累积引发系统性危机。AI赋能的安全风险评估体系还具备自我进化能力，通过持续吸纳新的风险案例与处置数据，模型能够不断迭代优化评估逻辑，逐步提升对新型风险的预判能力，这种适应性学习特性使其能够应对技术迭代与环境变化带来的未知风险挑战。在复杂系统中，不同风险因子的耦合效应往往是引发重大事故的关键，而AI的多变量分析能力可有效解析这种耦合关系，构建更为立体的风险传导模型，为制定针对性防控策略提供科学依据，进而推动安全风险管理从被动应对向主动预防转型，显著提升整体安全防护体系的韧性与效能。

2 AI 赋能的安全风险评估模型构建

2.1 多源异构数据预处理

安全风险评估所涉及的数据来源极为广泛，涵盖设备传感器采集的运行参数、环境监测站反馈的温湿度及有害气体浓度等环境数据，还有过往事故详细记录形成的历史数据等，这些数据不仅来源多样，且在格式、结构与语义上存在显著差异，呈现出典型的多源异构特性。在将这些数据纳入风险评估模型前，需进行精细的预处理操作。针对数据缺失问题，运用数据挖掘领域的多重填补法，依据数据的整体分布特征与属性间的关联关系，对缺失值进行合理推测与填充，避免因数据缺失导致关键信息遗漏。对于异常值，借助基于密度的空间聚类算法（DBSCAN），识别出偏离正常数据簇的孤立点，并通过与领域专家知识结合，判断其是否为真实异常或数据采集错误，若是错误则进行修正或剔除。为统一数据格式，针对结构化数据，采用标准化的数据编码方案，规范时间戳、数值单位等关键字段；对于非结构化的文本数据，如事故报告、设备维护日志等，利用自然语言处理中的词嵌入技术（如Word2Vec），将文本转

通讯作者：朱志敏(1990-08-16)，男，工程师，从事安防行业产品与解决方案工作。E-mail:442098132@qq.com

化为计算机易于处理的低维向量表示，从而打破数据格式壁垒，为后续分析奠定坚实基础。

2.2 风险评估算法选择与融合

风险评估算法的恰当选取与有效融合是构建精准评估模型的核心。在众多算法中，逻辑回归算法能够基于历史数据，通过对风险因素与风险发生概率之间的线性关系建模，快速识别出具有显著影响的风险因子，但其对复杂非线性关系的刻画能力有限。决策树算法则以树形结构对数据进行划分，依据不同属性的取值来构建决策规则，可直观呈现风险评估的逻辑流程，适用于处理离散型数据，但在面对连续型变量较多的数据集时，容易出现过拟合现象。神经网络中的长短期记忆网络（LSTM），因其独特的门控机制，在处理时间序列数据方面表现卓越，能够捕捉风险因素随时间的动态变化趋势，有效应对如设备运行状态随时间波动这类具有时序特征的风险场景。为充分发挥各算法优势，采用Stacking集成学习策略，将逻辑回归、决策树作为初级学习器，对数据进行初步特征提取与风险模式识别，再将其输出作为LSTM网络的输入，利用LSTM强大的非线性处理能力对初级学习器的结果进行深度融合与再学习，从而构建出一个能够兼顾线性与非线性关系、适应多种数据类型的综合性风险评估算法体系[1]。

2.3 评估模型的训练与优化

评估模型的训练过程是一个不断调整模型参数以最小化预测误差的迭代过程。以深度学习框架TensorFlow为基础搭建风险评估模型，采用随机梯度下降（SGD）算法作为优化器，在每一次迭代中，随机选取一批训练数据（mini-batch），计算模型在这批数据上的损失函数值（如均方误差损失函数，用于衡量预测风险值与真实风险值之间的偏差），并根据梯度信息更新模型参数，使模型朝着损失函数值减小的方向优化。为防止模型过拟合，引入L2正则化技术，在损失函数中添加正则化项，对模型参数进行约束，避免参数值过大导致模型过于复杂。通过调整学习率这一关键超参数，控制模型在训练过程中的参数更新步长，初期设置较大学习率以加快收敛速度，随着训练推进，采用指数衰减策略逐渐减小学习率，使模型在后期能够更精细地调整参数，提高模型的泛化能力。为提升训练效率，利用GPU并行计算技术，将大规模的训练数据分配到多个计算核心上同时进行计算，大幅缩短模型训练时间，加速模型收敛至最优解。

2.4 评估模型的验证与迭代

完成模型训练后，需对其性能进行严格验证，以确保模型在实际应用中的可靠性。从训练数据集中划分出一部分数据作为验证集，将模型在验证集上的预测结果与真实值进行对比，运用准确率、召回率、F1值等多种评估指标综合衡量模型性能。准确率反映模型正确预测为正样本（即存在风险情况）的比例，召回率体现模型对实际存在风险样本的覆盖程度，F1值则是综合考虑准确率与召回率的平衡指标。若模型在验证集上表现不佳，如准确率较低或召回率不达标，需对模型进行迭代优化。第一，返回模型训练环节，调整超参数设置，如增加神经网络的隐藏层节点数、改变激活函数类型等，以探索更优的模型结构；第二，重新审视数据预处理过程，检查是否存在数据质量问题未被妥善处理，或者是否遗漏了关键风险特征，对数据进行再次清洗、特征工程优化后，重新训练模型并进行验证，如此循环往复，通过不断迭代，逐步提升模型的预测精度与稳定性，使其能够精准识别各类安全风险，为实际应用提供可靠的风险评估支持。

3 AI 赋能的安全风险预警系统设计

3.1 系统总体架构设计

AI赋能的安全风险预警系统采用分层架构设计，以实现数据处理、模型计算与应用服务的解耦与协同。底层为感知层，部署物联网传感器、智能监控设备及边缘计算节点，实时采集设备运行的振动频率、温度变化、能耗波动等物理量，以及网络流量、访问行为等虚拟空间数据，构建全域感知数据源网络。数据经边缘节点预处理后，通过5G/工业以太网传至中层数据层。该层由分布式文件系统（HDFS）与时序数据库（InfluxDB）组成，分别负责非结构化日志持久化存储与高频采样数据时序化管理，还集成数据清洗引擎与特征工程模块，为上层提供标准化特征向量。上层为计算与应用层，有基于GPU集群的模型训练平台与轻量化推理引擎，训练平台支持多算法并行调优与模型版本管理，推理引擎通过容器化部署实时计算风险评分，将结果与预警信号推送至可视化应用层，该层用WebGL技术构建三维动态风险热力图，支持多维度筛选查看风险分布。整个架构通过微服务网关实现各层间的通信调度，借助服务注册与发现机制确保系统在节点扩容或故障时的自适应调整[2]。

3.2 预警功能模块设计

预警功能模块基于风险等级传导机制实现多层级预警逻辑，核心包含风险识别、等级判定与响应触发三个子模块。风险识别模块利用集成特征提取器，从预处理数据中解析出设备异常振动频谱、环境参数突变曲线、网络数据包异常特征等关键风险指标，结合预训练风险

特征库进行模式匹配，实时标记潜在风险点。等级判定模块采用多维度加权算法，将识别到的风险指标映射至预设五级风险等级体系。算法权重通过强化学习动态调整，如在工业场景提升设备温度异常权重，网络场景增加异常访问频次影响因子，最终输出风险等级量化值与置信度区间。响应触发模块依等级判定结果执行差异化预警：低等级（1-2级）经系统内消息队列推送至设备维护终端，附风险趋势预测；中等级（3级）触发声光报警，生成含风险位置、影响范围的处置指引；高等级（4-5级）启动联动控制，向关联设备发紧急停机或隔离信号，并通过API接口同步预警信息至应急指挥系统。各子模块通过事件总线实现异步通信，降低耦合，提升响应，确保预警逻辑在高并发数据输入时的处理效率。

3.3 数据交互与接口设计

数据交互架构采用发布-订阅模式实现跨模块、跨系统的数据流转，内部接口基于RESTful规范设计，外部接口则通过API网关提供标准化接入协议。系统内部，感知层与数据层之间采用MQTT协议进行数据传输，边缘节点作为发布者将采集数据按主题分类推送至消息代理服务器，数据层订阅相关主题实现数据的异步接收，这种设计减少了节点间的直接耦合，提升了数据传输的容错性。数据层与计算层通过基于gRPC的接口进行高效通信，利用ProtocolBuffers序列化格式压缩数据传输量，使特征向量在GPU集群间的传输延迟控制在毫秒级，满足实时推理的时间敏感需求。与外部系统的交互采用OpenAPI 3.0规范，接口包含设备状态查询、风险历史数据导出、预警规则配置等功能，通过API网关实现请求限流、身份认证与数据加密，支持第三方系统通过OAuth 2.0协议获取访问令牌后进行数据交互。针对海量历史数据的批量导出需求，设计基于HTTPRange的断点续传接口，允许客户端分块下载大型数据集，同时提供数据格式转换服务，支持将原始时序数据导出为CSV、Parquet等多种格式，适配不同数据分析工具的导入需求。

3.4 系统安全与可靠性设计

系统安全机制从数据传输、存储与计算三个维度构建防护体系，确保AI模型与风险数据的完整性与机密性。数据传输环节，采用TLS 1.3协议加密所有通信链路。针对边缘节点与中心服务器的无线传输，额外部署基于混沌加密算法的轻量级加密模块，在不明显增加能耗的前提下增强抗截获能力。存储安全上，敏感数据如设备密钥、模型参数等，用AES-256算法加密后存储。分布式存储系统通过多副本机制实现数据冗余，每个数据块在不同物理节点保存3个以上副本，且副本分布遵循机架感知策略，防止单点故障致数据丢失。计算安全通过可信执行环境（TEE）隔离保护模型推理过程，防止恶意进程篡改结果。部署异常行为检测模块，实时监控GPU/CPU使用率突变、内存访问越界等情况，发现攻击迹象立即隔离计算节点。可靠性设计采用故障自愈机制，心跳检测程序实时监控节点运行状态，节点离线时自动迁移任务至备用节点，迁移靠预加载模型快照无缝衔接，确保预警服务持续可用，系统平均无故障运行时间（MTBF）设计目标不低于10000小时[3]。

结语

综上所述，AI赋能的安全风险评估与预警系统通过多源异构数据预处理、多算法融合建模及分层架构设计，实现了风险评估从定性到定量、从定期到实时的转变。模型的自我进化能力与系统的安全可靠性设计，使其能适应复杂环境与新型风险挑战。该研究推动了安全风险管理向主动预防转型，未来可进一步优化算法融合策略与跨系统协同机制，提升在更广泛领域的适配性与应用效能。

参考文献

[1]朱艳,刘铭宸.AI算法赋能网络文化安全风险预警机制的策略[J].传播力研究,2024,8(1):145-147.

[2]刘阳,方雨,樊佳靓.AI技术赋能财务应用的风险及对策研究[J].商业经济,2025(1):165-168.

[3]唐建华.AI赋能智慧监狱安防体系的深度应用[J].数码设计,2024(21):91-93.