 DOI: https://doi.org/10.12345/bdai.v7i3.38172

Research on Intelligent Analysis and Aggregation of Logs Based on AI Large Model and User Feedback Mechanism

Ruili Li

Shanghai Digital Security Technology Co., Ltd., Shanghai, 200435, China

Abstract

To address challenges in cybersecurity monitoring platforms such as inefficient multi-source heterogeneous log parsing and high costs associated with manually writing regular expressions, this study proposes an intelligent log parsing and aggregation method leveraging large-scale AI models and user feedback mechanisms. The approach utilizes AI-driven automatic generation of regular expressions, incorporates user validation mechanisms to ensure accuracy, optimizes model performance through feedback data, and categorizes successfully parsed logs to generate risk events. Experimental results demonstrate that the method achieves 94.7% parsing accuracy, 30-fold efficiency improvement, 76% reduction in user intervention frequency, and 12.3% decrease in false alarm rates.

Keywords

AI large models; log parsing; regular expressions; user feedback; log aggregation

基于 AI 大模型与用户反馈机制的日志智能解析与聚合方法研究

李瑞丽

上海数字安全科技有限公司，中国·上海 200435

摘　要

针对网络安全监控平台中多源异构日志解析效率低、人工编写正则表达式成本高等问题，本文提出一种基于AI大模型与用户反馈机制的日志智能解析与聚合方法。该方法通过AI大模型自动生成正则表达式，引入用户核验机制确保准确性，利用反馈数据优化模型性能，并对解析成功的日志进行归类聚合生成风险事件。实验结果表明，该方法解析准确率达94.7%，效率提升30倍，用户干预次数减少76%，误报率降低12.3%。

关键词

AI大模型；日志解析；正则表达式；用户反馈；日志聚合

1 引言

网络安全监控平台需纳管大量异构安全产品，日志数据呈指数级增长，大型企业日均产生 TB 级日志 [1]。这些日志格式各异，给安全运维带来巨大压力。

日志标准化解析需根据不同设备源编写正则表达式实现字段转义 [2]。传统人工编写方式效率低下，难以应对海量日志；维护困难，设备升级导致规则失效；技能门槛高，需专业知识。

AI 大模型具备理解日志结构、自动生成解析规则的潜力 [4]，用户反馈机制可有效提升 AI 系统性能。本文提出一种基于 AI 大模型与用户反馈机制的日志智能解析与聚合方法。

【作者简介】李瑞丽（1983–），女，中国上海人，硕士，工程师，从事数字安全、网络威胁情报分析与多源数据融合技术研究。

2 相关工作

2.1 日志解析技术

传统日志解析方法主要分为三类：基于正则表达式的方法精确度高但维护成本高 [2]；基于模板库的方法难以覆盖动态变化的日志格式；基于聚类的方法如 Drain 算法在特定场景效果较好，但对语义信息利用不足 [4]。

2.2 AI 在日志分析中的应用

深度学习技术被广泛应用于日志分析。DeepLog 利用 LSTM 进行日志异常检测 [3]。Elastic 研究团队探索了 LLM 在少样本日志解析中的应用，发现其能高置信度生成解析规则 [4]。LLM-LADE 实现了日志异常判别与解释生成双重任务 [6]。

2.3 用户反馈机制

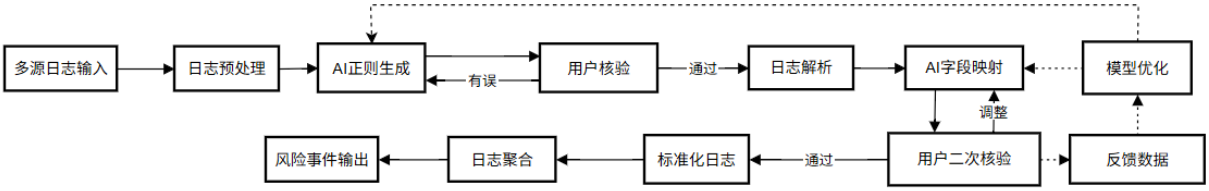
用户反馈可有效提升 AI 系统性能，形成“模型生成 - 用户校验 - 模型优化”的正反馈闭环。在 RAG 架构的日志分析系统中，反馈闭环通过“人工复核”节点实现知识持续积累。

3 系统模型与方法

3.1 总体框架

本文方法总体框架包含五大核心模块：日志采集与预处理、AI 正则生成、用户核验、AI 字段映射、日志聚合。

工作流程为：采集并预处理多源日志；调用 AI 大模型生成正则表达式；提交用户核验，有误则重新生成或人工修正；基于最终正则解析日志；调用 AI 大模型将解析结果映射至字段模板，生成初始解析结果；再次提交用户核验，用户可调整字段格式、配置及值映射；生成标准化日志；最后对解析成功的日志进行归类聚合，生成风险事件反馈用户。



3.2 日志预处理与格式识别

日志预处理是确保 AI 大模型准确理解日志结构的基础步骤。原始日志包含大量噪声信息，直接输入 LLM 会影响生成质量并消耗大量 Token，因此需要系统的预处理流程。

日志标准化清洗：首先对原始日志进行语法级解析，将不同格式的日志转换为统一的 JSON 中间格式。清洗操作包括：去除控制字符、异常空格等无关字符；将各种时间戳格式统一转换为 ISO 8601 标准（如 "2025-03-17T14:23:45Z"）；对 IP 地址、端口号、文件路径、进程 ID 等可变内容进行泛化标记，用占位符替换具体值。

日志格式指纹生成：为识别日志的结构模式并支持后续的聚类分析，本文引入日志格式指纹方法。具体映射规则如下：任意长度的数字序列替换为单个 "0"；任意长度的字母序列及空白替换为单个 "a"；多个连续空白字符压缩为单个空格；保留标点符号和特殊字符不变。

分层聚类与代表性样本选取：基于日志格式指纹，采用两阶段分层聚类策略。子类聚类通过完全相同的指纹对日志进行聚合，同一子类的日志共享完全相同的格式结构；元类聚类通过指纹的前 N 个字符（本文取 N=15）进行前缀匹配，将词汇相似的格式归并到同一元类。从每个元类中选取代表性日志样本作为 AI 大模型的输入，选取策略包括选择最早出现的样本、选择字段最完整的样本、随机选择等。

3.3 基于 AI 大模型的正则表达式生成

提示工程优化：本文设计的提示模板包含角色设定、任务描述、输出格式、约束条件和示例引导，引导模型生成高质量正则表达式。角色设定让模型聚焦于网络安全日志解析的专业场景，示例引导帮助模型理解预期输出样式。通过多轮迭代优化提示词，模型对日志结构的理解能力和正则表达式的生成质量持续提升。

迭代生成与错误反馈机制：设计迭代生成机制，当用户反馈有误时，系统收集错误类型作为上下文重新生成，最多迭代 3 次，累积通过率可达 96.8%。错误类型包括匹配范围过宽、遗漏关键字段、分组错误等，可通过预定义分类器自动识别或用户手动标注。

多示例学习策略：对复杂日志格式，系统自动选取同一元类中 3-5 个代表性样本作为输入，可将准确率提升 12% 以上。多示例能够帮助模型识别日志中的固定部分和可变部分，理解字段的共性特征和个体差异。

正则表达式验证与优化：AI 生成的正则表达式需经过语法检查、样本匹配测试和字段提取测试，验证不通过则自动触发优化，形成生成 - 验证 - 优化闭环。语法检查确保表达式合法可用，样本匹配测试验证其对典型日志的覆盖能力，字段提取测试确保关键信息被正确捕获。

3.4 用户核验与反馈机制

双重核验设计：本文设计正则表达式核验与字段映射核验双重机制，从规则层和字段层分别保障解析质量。正则表达式核验确保解析规则的准确性，字段映射核验保障字段级语义的正确性。双重核验形成完整的质量控制闭环，有效降低解析错误率。

正则表达式核验流程：用户可选择核验通过、反馈错误触发 AI 重新生成或人工修正。超过 3 次重新生成则接受手动输入，记录失败案例用于后续分析。该机制平衡了自动化效率与人工干预的准确性，避免无限循环影响用户体验。失败案例的积累也为模型优化提供了宝贵的负样本。

字段映射核验流程：用户对 AI 生成的解析结果进行字段格式设置、额外配置、值映射和重命名，确保日志格式的一致性和可读性。字段格式设置统一了多样化的时间、数值表示，值映射将原始值转换为标准化语义标签。这些处理使最终日志具备跨平台的互操作性和业务可理解性。

反馈质量评估：系统从完整性、修正幅度、处理时间、一致性四方面评估反馈质量，高质量数据优先用于模型训练，低质量数据过滤或降权使用。完整性确保关键字段无遗漏，修正幅度反映用户对 AI 输出的改进程度。处理时间和一致性则分别衡量反馈效率和稳定性，共同保障训练数据的可靠性。

3.5 基于反馈的模型优化

正则生成模型优化：基于源日志及其对应的最终正则表达式对 AI 大模型进行监督微调。采用 LoRA（低秩适应）参数高效微调技术，在保持模型通用能力的同时，增强其对特定日志格式的解析能力。训练时采用交叉熵损失函数，优

化器选择 AdamW，学习率设置为 2e-5，确保模型稳定收敛。

字段映射模型优化：基于源日志及其对应的最终日志解析结果对 AI 大模型进行优化，提升字段识别的准确性和映射的合理性。目标输出为标准化的字段映射 JSON。对于字段格式设置、值映射等操作，模型需学习相应的转换规则，逐步掌握不同设备日志的字段映射规律。

增量学习机制：随着反馈数据的持续积累，模型性能呈现正反馈循环：更多高质量反馈带来更好模型，更好模型减少用户干预需求。为支持持续优化，本文设计增量学习机制，每周使用新增反馈数据对模型进行微调。增量学习采用弹性权重巩固（EWC）等技术，避免灾难性遗忘问题，确保模型在学习新知识的同时保留已有能力，实现性能的稳步提升。

模型版本管理与回滚：建立模型版本管理机制，每次优化后生成新版本模型。新版本上线前需通过 A/B 测试验证性能提升，与当前版本进行对比评估。如发现新版本性能下降，可快速回滚至稳定版本，确保系统可靠性，降低运维风险。

3.6 日志解析执行与缓存优化

批量解析引擎：基于最终核验通过的正则表达式，对原始日志进行批量解析。解析引擎采用优化的正则表达式引擎，支持多模式并行匹配和多线程处理，充分利用多核 CPU 资源，提升处理效率。

缓存机制设计：为提升映射效率，本文设计多级缓存机制。格式指纹缓存基于日志格式指纹缓存对应的正则表达式和映射规则；字段映射缓存对已处理的日志格式缓存字段映射规则，后续相同格式日志可直接应用缓存规则，无需重复调用 AI 模型。缓存采用 LRU 淘汰策略，缓存命中率可达 85% 以上，大幅降低 AI 调用次数和响应时间。

异常处理机制：对解析失败的日志，系统记录失败原因并进入异常处理队列。异常类型包括正则表达式不匹配、字段提取失败、映射规则缺失等。异常队列中的日志可批量触发人工处理或模型重新学习，不断提升系统的鲁棒性。

3.7 日志聚合与风险事件生成

基于告警类型的聚合：识别日志中的攻击类型、威胁类别等目标字段，将其划分至对应的告警类型中。整合同一告警类型的所有日志文件，生成风险事件发送给用户确认，减少同类告警的重复展示。

基于源 IP 频率的聚合：在预设时间窗口内统计同一源 IP 的日志出现次数，超过阈值时整合生成 " 源 IP 高频访问告警 " 事件，有效识别暴力破解、DDoS 攻击等具有频率特征的攻击行为。

基于源 IP 多样性的聚合：在预设时间窗口内统计不同源 IP 的日志出现数量，超过阈值时整合生成 " 分布式扫描攻击 " 事件，识别僵尸网络扫描等分布式攻击行为。

多维度关联分析：本文支持时间关联合并相近告警、攻击链关联还原攻击全貌、资产关联评估风险状况、用户行为关联识别内部威胁，提供丰富的攻击上下文。

风险事件生成与优先级排序：聚合后的风险事件包含事件摘要、涉及资产、时间范围、原始告警列表和综合风险评分。采用多因素加权算法计算评分，对高风险事件优先推送，实验表明告警压缩率达 85.1%，攻击检出率保持 98.7%。

4 实验与分析

4.1 实验设置

采集防火墙、IDS、WAF、服务器日志等 50 万条样本，涵盖 20 种日志格式。对比人工解析法、模板解析法（Drain 算法）与本文方法。评价指标包括解析准确率、处理时间、用户干预次数、误报率。

4.2 结果分析

整体性能对比：人工解析法准确率 89.1%，处理时间 1240ms/ 条，用户干预 100 次 / 百条，误报率 15.7%；模板解析法准确率 82.3%，处理时间 86ms/ 条，用户干预 12.4 次 / 百条，误报率 21.3%；本文方法准确率 94.7%，处理时间 42ms/ 条，用户干预 2.9 次 / 百条，误报率 8.5%。

反馈机制效果验证：将 200 种日志格式分 4 批次测试，用户干预次数从首批 4.8 次 / 百条降至末批 1.7 次 / 百条，准确率从 92.3% 提升至 96.1%，验证反馈闭环有效性。

多示例学习效果：对复杂日志格式，单示例输入准确率 62.3%，多示例输入提升至 74.5%，提升 12.2%，验证其在处理复杂日志时的有效性。

日志聚合效果：在 150 个攻击事件模拟中，无聚合产生 1247 条告警，聚合后整合为 186 个风险事件，压缩率 85.1%，检出率保持 98.7%。

5 结论与展望

本文提出一种基于 AI 大模型与用户反馈机制的日志智能解析与聚合方法。通过 AI 生成正则表达式和字段映射，引入双重用户核验，构建反馈优化闭环，设计多维聚合策略。实验表明，该方法解析准确率 94.7%，效率提升 30 倍，用户干预减少 76%，误报率降低 12.3%。

未来研究方向：引入多模态模型处理非结构化日志，探索 RAG 架构构建解析知识库，研究自适应阈值调整机制，扩展至 EDR、NTA 等场景，探索联邦学习实现跨组织协同解析。

参考文献

[1] 滕开清, 曾哲凌, 胡芳燕. 基于人工智能的通信行业网络安全新运营体系[J]. 邮电设计技术, 2025(11): 45-52.

[2] 张尼, 刘镪, 张静等. 网络安全威胁情报关键技术研究综述[J]. 计算机研究与发展, 2020, 57(10): 2035-2049.

[3] Elastic Search Labs. 在Streams中利用机器学习自动化日志解析[EB/OL]. (2026-01-01). https://www.elastic.co/search-labs/cn/blog/log-parsing-partitioning-automation-experiments-streams.

[4] Zhang Z, Li S, Zhang L, et al. LLM-LADE: Large language model-based log anomaly detection with explanation[J]. Knowledge-Based Systems, 2025, 326: 114064.