Contribution ID: **8** Type: **not specified**

基于 LLM 的钓鱼邮件识别智能体

Tuesday, 26 August 2025 17:40 (20 minutes)

随着数字化办公的普及，钓鱼邮件已成为网络攻击的主要载体。因存储着大量核心知识产权、未公开研究数据、涉密项目信息及人员敏感数据等，科研机构成为重点网络攻击目标，因此，构建具备多模态信息融合与深度推理能力的智能化钓鱼邮件识别机制，对提升科研机构网络安全防护水平、保障科研活动的连续性与安全性具有重要意义。
传统钓鱼邮件识别方法多依赖于单一模态分析，存在识别维度有限、复杂场景适应性弱等问题。针对这一局限性，本研究将思维链（Chain-of-Thought）技术与大语言模型 (Large Language Model,LLM) 结合构建了具有工具调用能力和强大推理能力的钓鱼邮件识别智能体，能够将复杂的识别任务拆解为结构化的推理步骤，并在关键节点通过工具调用补充所需的多模态信息，提升判断的置信度。
该智能体采用四模块架构，其中信息增强与推理决策为核心模块。LLM 对输入模块预处理后的邮件数据进行初步分析，生成包含置信度的判断结果，并决策是否调用信息增强模块；信息增强模块则提供用于获取开源威胁情报、网站截图与网站代码三类跨模态特征信息的工具，用于增强分析依据；随后，推理决策模块融合初步结论与获取的增强信息进行综合研判，输出最终的识别结果。
本研究不仅能够显著提升识别准确率与结果可解释性，还有效缓解了 LLM 在特定领域知识覆盖不足及信息时效性滞后的问题，为钓鱼邮件识别提供了一种新的技术路径，为提升网络安全防护能力提供了理论支撑与实践参考。

Summary

Primary author: 苑, 新阳 (中国科学院高能物理研究所)

Presenter: 苑, 新阳 (中国科学院高能物理研究所)

Session Classification: 科研信息化管理与系统

Track Classification: 人工智能与应用