doi:10.1631/FITEE.1601325

题目：基于机器学习的自动化恶意代码分类与新恶意代码检测技术

概要：恶意软件的爆炸式增长对信息安全构成重大威胁。基于签名机制的传统反病毒系统无法将未知的恶意软件分类到相应的恶意家族和检测新的恶意软件。因此，我们提出一种基于机器学习的恶意软件分析系统，由数据处理系统，决策系统和新的恶意软件检测系统三个子系统组成。数据处理系统包含灰度图像的纹理特征，Opcode 特征和 API 特征等三种特征提取方法。决策系统被用来分类恶意软件和证实可疑的恶意软件。最后，检测系统使用共享近邻聚类算法（shared nearest neighbor, SNN）来发现新的恶意软件。我们在 Kingsoft,，ESET NOD32 和 Anubis 收集的二万多恶意样本集上对所提出的方法进行了评估。结果表明，我们的系统可以有效地分类未知恶意软件，准确率可达 98.9%。同时新恶意软件的成功检测率为 86.7%。

关键词：恶意代码分类；机器学习；*n*-gram；灰度图；特征提取；恶意代码检测