

# 曾行健

✉ zxbibobibobi@163.com

## 教育经历

东南大学  
硕士研究生 网络空间安全学院  
指导老师: 袁亚丽副研究员

南京, 江苏  
2023.09至今

东南大学  
本科毕业 网络空间安全学院

南京, 江苏  
2019.09 - 2023.07

## 研究经历

### Learning from Massive Highly Imbalanced Data via Hybrid-sampling with Self-paced Curriculum. [Code](#)

- 在投论文。
- 本工作面向大规模不平衡数据分类场景, 考虑少数类内分布不平衡、类间重叠、数据类间不平衡等数据层面挑战, 基于对现有不平衡方法缺陷的分析, 提出一种基于混合采样的集成不平衡学习框架**SCHE** (Self-paced Curriculum Hybrid Ensemble)。**SCHE**借鉴了经典的深度学习范式: 自步-课程学习 (Self-paced Curriculum Learning), 基于先验和学习反馈来指导对不平衡数据集的采样策略。
- 基于**Python**实现了**SCHE**, 并在具有不同不平衡程度、不同大小的多个数据集上进行了一系列实验。实验结果表明, 与经典的不平衡学习方法以及近年提出的baseline相比, **SCHE**在多个指标 (AUPRC、F1、G-mean、MCC指数) 上以可接受的计算开销取得了有竞争力甚至更好的结果。

### Research on Imbalanced Learning Methods for Credit Card Fraud Detection

- 本科毕设, 主要工作是相关工作总结和算法设计。
- 面向数据不平衡的信用卡欺诈检测场景, 设计了一种新颖的过采样算法以增强基于欠采样 (under-sampling) 的集成不平衡学习方法。
- 该算法基于聚类过程, 对少数类样本进行适当的数据增强, 在一定程度上扩大训练集大小, 避免了传统过采样/欠采样方法由于采取满采样率, 在极度不平衡的欺诈检测场景下面临的噪声引入/信息丢失等问题。

## 研究兴趣

- 数据挖掘。包括不平衡学习 (Imbalanced Learning) 和异常检测 (Anomaly Detection)。
- 加密流量分类 (Encrypted Traffic Classification)。对在流量工程背景下应用机器学习/深度学习方法解决有关问题感兴趣。
- 对匿名网络的攻击 (与防御)。包括网站指纹攻击 (Website Fingerprinting), 流关联攻击 (Flow Correlation)。

## 个人总结

- 习惯 (有经验于) 使用python进行项目开发。
- 擅长文献阅读与信息总结。
- 习惯快速学习新领域。
- 对研究抱有热情。