

赵卫国

应届硕士 • 166-2080-8852 • aszhaoweiguo@hotmail.com

教育背景

2016.09 – 2019.01	哈尔滨工业大学	计算机科学与技术	硕士	排名: 22 / 115
2012.09 – 2016.06	青岛大学	网络工程	学士	排名: 5 / 61

获奖经历

2017.10	2017 开放学术精准画像大赛	第三名 (3/406)
2017.09	哈尔滨工业大学	一等奖学金
2014.04	第五届“蓝桥杯”全国软件大赛山东赛区本科 C/C++ 组	二等奖

项目经历

2017.09 – 至今 基于神经网络的安卓恶意代码检测 独立完成

任务描述: 对提供的安卓应用程序, 分析检测识别其中的恶意应用程序。

方法:

a) 静态分析安卓应用程序, 利用反编译工具获取函数调用关系图和操作码序列, 使用污点分析技术获取 API 序列; b) 将函数调用关系图和 API 序列利用 LSTM 学习特征向量, 将操作码构造成灰度图使用 CNN 学习特征向量; c) 使用 SVM + Stacking 技术进行分类。对安卓应用程序的最终分类准确率为 96%。

2017.07 – 2017.10 2017 开放学术精准画像大赛 队长 第 3 名 (TOP 1%, 3 / 406)

任务描述: Task 1: 基于学者名+组织名的搜索结果, 确定学者的个人主页和个人信息。Task 2: 基于学者的论文信息, 挖掘学者的研究兴趣。Task 3: 基于学者的论文信息, 预测学者截止 2017 年 6 月的论文被引用量。

方法:

Task 1: 使用搜索结果标题、URL 等构造特征, 样本正负比 1:9, 尝试过采样和负采样方法, 使用 XGBoost 确定个人主页, CNN 识别照片, 朴素贝叶斯确定性别, 正则表达式确定邮件等信息; Task 2: 基于论文题目, 合著作者、论文引用关系等构造特征, 使用 PCA 降维后训练三个分类器并将其输出结果后处理, 选出学者研究兴趣; Task 3: 基于业务需要建立特征(h-index、论文数量等), 使用 XGBoost 二分类后, 对正例数据使用 XGBoost 回归预测学者论文被引用量。

2016.10 – 2016.12 微博用户画像 独立完成

任务描述: 给定用户微博文本数据以及社交网络图, 预测用户的性别(二分类)、年龄(三分类)、地域(八分类), 最终采用加权准确率作为评价指标。

方法:

a) 使用卡方分析提取关键词, 构造 BOW 特征; b) 利用微博文本, 使用 Doc2Vec 技术训练出用户的 Document Vector; c) 利用社交关系, 使用 Graph Embedding 训练用户的 Node Vector 作为特征; d) 综合上述特征, 使用 PCA 降维、SVM 基分类器, 并采用 Stacking 技术进行模型融合。e) 对地域分类额外构建了地域常识词典; 对性别分类额外建立性别倾向性词典。最终准确率为性别 88.3%, 年龄 64.8%, 地域 72.7%。

个人技能

- 熟悉 Python、C++ 语言, Linux 平台;
- 熟悉 LR、SVM、XGBoost 等常见机器学习算法, 了解 CNN、LSTM 等深度学习算法;
- 熟练使用 Scikit-learn、Numpy 等机器学习库, 了解 Tensorflow 深度学习库。
- 个人博客: <https://weiguo Zhao.github.io/>