

黄真川

15996253710
huangzc@lamda.nju.edu.cn
chuan92.com

教育背景

- 2014.9–2017.6 硕士, 南京大学计算机系, 机器学习与数据挖掘研究所 (LAMDA).
2010.9–2014.6 学士, 南京大学计算机系, 保研, GPA Top20% .

实习经历

- 2016.4–2016.7 腾讯, 互动娱乐运营部数据挖掘算法组, 算法工程师.
* 从玩家的画像、游戏行为日志、道具购买等数据库中分析整理千万量级的数据
* 在 Spark 平台上设计实现一种多标记学习算法用于游戏道具推荐, 重点优化头部位置推荐的准确率; 算法训练时间短 (20min 内), 准确率相对于热销推荐有一倍提升
2013.7–2013.10 eBay, StubHub DevOps Team, Software Engineer Intern.
* 基于 libssh2 库使用 C 语言开发内部 scp, curl 工具
* 基于 Spring, RESTful, MongoDB 等开发追踪内部 Maven 仓库依赖变动的工具

科研经历

- 2015.12–2016.4 多模态多标记学习研究.
将在校大学生的奖学金、助学金评定建模成一种多模态多标记学习问题。将多个来源的信息看作不同的模态, 提取过程有效保护隐私, 算法通过正则项引入标记之间的关系, 在半监督的框架下预测生效。相关论文 *College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach* 已被 *IEEE International Conference on Data Mining 2016* 录用
2015.7–2015.12 基于近邻重构的多示例学习研究.
多示例学习是一种半监督学习, 其输入为一组被标注的“包”, 每个“包”都包括许多示例。设计并实现了一种基于近邻重构的多示例学习算法, 可用于文本情感识别、图像感兴趣区域检测等。论文发表于 *2016 International Conference on Intelligence Science and Big Data Engineering*.

竞赛经历

- 2015.12–2016.1 Prudential Life Insurance Assessment, Kaggle Competition, 116th/2619.
* 使用 GBDT, DNN 作为基学习器, 通过 Linear Regression 融合, 解决生命保险评级 (有序回归) 的问题.
* 主要使用 scikit-learn, pandas, xgboost, keras 等工具
2014.12–2015.1 Click-Through Rate Prediction, Kaggle Competition, 108th/1604.
* 移动设备广告点击率预测问题, 四千万条广告曝光数据, 大多为离散特征
* 主要使用 FTRL 和 FM 模型

社区

- Blog chuan92.com.
使用 Jekyll 在 DigitalOcean VPS 上搭建的个人博客, 整理技术方面的一些思考和书评
GitHub github.com/chuan92.
热爱开源, 为 breeze 数值计算库, python-mode 等项目提过 issues 和 pull requests

奖项

- 2014 南京大学优秀毕业生, 南京大学学业奖学金一等
2013 南京大学计算机系优秀学生, 南京大学人民奖学金