



宋子恒

出生年月: 1993.07.08

18810712120

毕业时间: 2019.06

ziheng_song@126.com

意向职位: 数据挖掘

北京 海淀区北京邮电大学

教育背景

2016.09-至今	北京邮电大学	计算机科学与技术	(硕士)
2012.09-2016.07	西安电子科技大学	通信工程	(本科)

实习经历

2018.04-至今 今日头条 Data 数据挖掘实习生

- (1) 使用 xgboost 构建中银用户通过拒绝模型,计算特征重要性,以及评价模型性能
 - 使用银行提供的授信结果作为样本
 - 使用用户基础信息 app 安装信息 广告展现点击信息作为主要特征
 - Auc 0.7
- (1) 构建广告点击以及抖音财经特征仓库
 - 使用 hive spark 从抖音和广告日志中抽取信息聚合为训练模型所需要的特征表
- (2) 使用 spark 计算题目热度优化今日头条健康频道答题页点击转化率
 - 结合头条主端用户画像库, 使用 spark 统计题目覆盖的人数
 - 按照随机策略和热度排序策略积累用户数
 - 提升转化率 5%

2017.11-2018.03 百度 商务搜索部(凤巢) 策略研发实习生

- (1) 调研 FM, FFM 算法
- (2) 配合 shell 脚本使用公司内部平台搭建数据处理 pipeline
 - 使用 hive 与 udf 结合处理前端埋点日志
 - 例行产出数据报告, 自动监控异常指标
 - 按照不同粒度生成报表, 使用 shell 自动发送邮件给运营
- (1) 使用 hive, hadoop, python 负责日常广告相关指标的统计和分析
 - 使用 hadoop streaming 结合 pandas 库分析不同广告样式上点击率的差异
 - 配合 shell 脚本以及 hive python 构建常用的数据分析工具

项目经历

- 2018.07 CCIR-2018 知乎用户推荐比赛 排名 10 / 315
 - 使用 hive sql 实现基于物品的协同过滤算法 计算相关分数, 对用户推荐相关性 top10 文章
 - 处理数据总量 1000w 行 80GB
- 2018.05 第二届腾讯社交广告算法大赛 初赛成绩 230 / 3000
 - 使用 sklearn, pandas 完成数据的预处理
 - 使用 xgboost 完成模型的训练和预测
 - 多类模型做融合 auc 0.7448

专业技能

1. 熟悉常用的机器学习算法理论及框架(Decision Tree, SVM, LR, GBDT, Xgboost, Lightgbm)
2. 熟练使用 hive, Hadoop, Spark 等大数据处理框架
3. 熟悉常用的数据结构与算法
4. 熟悉 Python, Java, C