

程 刚

13052003867 withchenggang@gmail.com

上海市长宁区仙霞西路 885 弄

教育背景

上海对外经贸大学	应用统计学	本科	2013.09 - 2017.06
----------	-------	----	-------------------

工作经历

携程旅行网	机器学习工程师	2017.11- 2018.11
-------	---------	------------------

- 跟团游个性化推荐项目(基于 xgboost), 负责人:
 - 利用 hql、pyspark、java 对埋点数据进行解析与清洗, 完成前期数据处理与部分特征工程工作。
 - 利用 python 完成 xgboost (rank:pairwise) 模型特征工程, 样本构造, 离线训练, 参数调优等工作。
 - 配合在线服务工程开发, 编写在线打分逻辑, 完成数据保全, 模型校验, 模型自动更新等流程。
- 景点门票个性化推荐项目(基于产品标签), 负责人:
 - 用户行为数据清洗, 处理等基础数据搭建; 利用订单, 浏览, 评论等数据计算冷启动模型。
 - 利用浏览、点击等行为数据计算用户在不同城市对不同标签的长期偏好; 结合埋点加入实时偏好。
- 参与酒店个性化推荐模型, 利用 scala 完成数据处理、spark-xgboost 区分出度假目的的订单, 最后调用 mllib 中的 ALS 算法实现协同过滤模型, 得到用户对酒店星级, 商圈的偏好。

前隆科技	数据挖掘工程师	2017.03 - 2017.11
------	---------	-------------------

- 响应业务需求,为其提供数据支持与分析, 建立市场运营模型, 如 RFM 聚类模型, 为公司精准化营销提供基础。
- 对客服聊天记录进行文本挖掘,分析用户聊天高频词,为客服聊天机器人提供优化策略, 客服转人工率从 60%下降至 40%。
- 利用 python 爬虫定期爬取各贴吧, 论坛, 微博数据, 对数据进情感分析, 并利用 tableau 制成 BI 报表, 为公司网络舆情提供参考。

项目经历

- Kaggle: Home Credit Default Risk, 信贷违约风险预测** 2018.06 - 2018.08
对多表聚合, 完成特征工程, 特征选择(Lasso、LightGBM、FeatureTools), 非平衡数据处理(ENN、Tomek), 模型选择和优化 (GridSearch、BayesOpt、LightGBM、XGBoost), 以及模型融合(Stacking、Weighted Average Rank); 最终排名: **137/7200+** (top2%)
- 大学生挑战杯竞赛, 开发一款简单易用的统计分析软件** 2016.02 - 2016.06
本人工作: 负责基本矩阵运算,统计功能(回归分析等),数据挖掘算法(Apriori 算法等)功能的实现。
项目地址: <https://github.com/chenggang0815/VISAP>

掌握技能

- 熟练使用 python、R 进行机器学习算法开发, 熟悉常见数据结构, 熟悉 scala-saprk (mllib) 机器学习框架。
- 熟练掌握常见的机器学习和推荐算法, 如: lr、rf、svd、ltr、fm、gdbt、xgboost、lightgbm; 熟悉常见的深度算法, 如 rbm、cnn、rnn、lstm、gan 等。
- 数理统计基础扎实,熟悉回归分析,多元统计,时间序列,贝叶斯统计等方面的知识。
- 熟练掌握 hql, mysql 等数据库查询与优化, 熟悉 Linux 工作环境/Shell 命令,了解 Hadoop/Spark 等分布式计算原理。