



傅苗

<https://miao-fu.com>

基本信息

26岁

4年经验

18569499324

miaofu1997@qq.com

荣誉奖项

2017年11月

全国大学生数学建模竞赛
国家级一等奖

2018年4月

美国大学生数学建模竞赛
Honorable Mention

2019年6月

湖南省优秀毕业生

专业技能

Programming

Python, Go, Scala, SQL, Bash,
JavaScript, Ruby, LaTeX

Frameworks

Hadoop, Spark, Flink, TensorFlow,
PyTorch, Vue, Jekyll

Tools & Platforms

Git, Hive, Clickhouse, MySQL,
Kafka, HBase, MATLAB, ChatGPT

教育背景

2015.9-2019.6

湖南工商大学

信息与计算科学 GPA:3.43/4 (专业前7%)

数据结构、数据库原理、算法导论、信息论与编码、操作系统

概率论、统计学、运筹学、数学分析、高等代数、常微分方程、离散数学

2017.6-2017.8

Brown University 布朗大学

Summer Session

Statistical Inference (统计推断)、Linear Algebra (线性代数)

工作与项目经验

2021.3-2023.8

深圳萨摩耶数字科技有限公司

数据科学工程师

智能投放决策系统 (Intelligent-Decision)

- 负责系统的整体研发进度管理，包括排期规划、任务分配以及开发设计文档的撰写，确保项目按期上线；
- 开发了预测模型数据预处理模块(输入项)、结果存储以及后续运算适配模块(输出项)；
- 针对不同账期(MOB)的分析场景，设计并开发了投放效益计算流程，将损益数据按月和11个渠道类型进行聚合分析，成功处理每月3000多个投放渠道的数据，大大提升了数据处理效率和分析准确性；
- 训练线性回归(LR)和增长率折算模型，准确预测了成本项指标的变化趋势，评估了各投放渠道在未来账期的损益表现，预测准确率达90%以上，有效支持了未来投放决策的优化；
- 设计并开发了输入项相关大数据集群表监控系统，并实现了数据当天展示的效果；
- 对外输出了与计算流程相关的2项专利，提升了公司在技术创新领域的竞争力。

设备指纹系统 (Device-Fingerprint)

- 从Kafka读取设备指纹数据流，通过Spark进行分布式的批量预测任务。训练了三胞胎孪生神经网络生成高维向量指纹，并在cubetree向量数据库中进行匹配。上线后，系统每天处理千万级的数据量，实现了50毫秒内的匹配响应时间，匹配准确率达到单场景96.8%，跨场景71.24%，成功替换了原采购的设备指纹系统，为公司节省了40万人民币/年；
- 将每日新增的百万级指纹数据通过FlinkSQL导入Clickhouse，并创建和维护物化视图，以提高数据查询和存储效率，实现秒级响应。通过优化查询效率，开发并监控设备指纹系统的性能指标，包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)及F1值。

省呗App用户客群分析

- 用户画像与转化率分析：**基于省呗App在2019年至2023年4月的数据，分析了注册用户的登录、完件、授信、交易的转化率表现；计算了以性别、年龄、职业等10多个细分用户群体的转化占比。通过这些分析，识别出了更优质的用户群体，并据此调整了营销与放款策略，成功提高了客群质量，降低了市场获客成本。
- 淘金计划：**分析了累计超300万名注册未交易沉默用户的末次登录时间，并同时结合用户属性(未完件状态、授信拒绝、性别、职业等)进行分段打标。通过将细分的用户群体推送给渠道方进行精准营销，实现了存量用户激活率提升约13%，有效促进了再经营目标的达成。
- 用户留存与LTV分析：**提取近30天内各类用户群体(未完件、授信通过、拒绝户、C类户)的登录数据和交易及中收信息，计算出用户的30天留存率和每用户平均收入，并据此准确测算出用户生命周期价值(LTV)，为优化用户运营策略提供数据支持。

语言能力

英语六级:

565/710

TOEFL:

99/120

GRE:

321/340

省呗App中收业务分析与优化

- **会员卡数据日报**: 设计并开发了会员卡数据日报，对不同类型会员卡的曝光点击率、开通收益、支付成功率、退款率等关键指标进行了全面跟踪和分析。实现了每日自动化的数据跑批和报表更新，并将结果接入观远BI报表系统，大幅提升了运营效率。
- **次贷导流项目**: 针对本平台授信被拒用户，构建了用户分群模型并为每个群体打标，推送定制的第三方次贷平台落地页列表。通过精准分群、个性化推送及A/B Testing优化策略，每月次贷导流的注册率平均提升了约8%，交易发起率提升了约5%，显著提高了项目的收入和利润增长。

大数据集群维护与SQL效率优化

- 负责大数据集群的日常运维和巡检，确保各系统在高并发和大数据量环境下的稳定运行；针对业务慢SQL查询进行了深入分析和优化，大幅提升了SQL执行效率；解决了Hive小文件带来的存储问题和处理瓶颈，合并小文件以减少I/O开销，提高了数据读取和写入Hive的性能。以上优化措施显著增强了集群的整体性能和资源利用率。
- 编写了风控输入项数据由Hive导入HBase的Bulkload任务脚本，并负责任务的后续运维。在并发量增加的情况下，显著提高了风控数据的写入和读出效率。

2019.4-2021.3

中国电子 中电工业互联网有限公司

大数据工程师

混凝土泵车缺失数据治理

结合马尔科夫链蒙特卡洛方法 (MCMC) 和随机变量，对期望最大化 (EM) 算法进行改进，成功实现了对混凝土泵车行驶数据的高效填补与校正。通过从 MCMC 转移矩阵中采样生成缺失数据点，通过迭代优化使样本估计逐步收敛至真值，有效解决了行驶数据缺失问题。

2018.10-2019.6

移动商务智能湖南省重点实验室

助理研究员

基于大规模群体决策的网络化在线服务社交推荐

在科研人员与学术文献的复杂异构网络中，运用可重启随机游走(RWR)算法，深入挖掘数据间的潜在关联，为未来的科研合作关系提供精准且可靠的推荐。本人主要负责RWR推荐模型的实现，采用MATLAB和Python编程，系统地分析异构网络中节点的相似度，成功实施随机游走算法并输出高质量的合作推荐结果。

相关论文发表在SCI期刊《IEEE》 DOI: 10.1109/TCSS.2019.2932288