

蒋浩的简历

个人简介

- 手机: 15850730108
- 邮箱: nju.jianghao@foxmail.com
- 四年数据工作经验, 三年算法研发经验; 数据科学方向的全栈工程师: 懂算法、会技术、能数据分析、又能独立开发数据产品;
- 希望加入一支使用深度学习技术的团队, 并在团队中做长时间的积淀;

学历

- 硕士 2011 - 2014 南京大学 计算机科学与技术系
- 本科 2007 - 2011 南京大学 计算机科学与技术系

工作经历

2018.04 - 2018.06 技术负责人 创业 上海

- **加入和离开原因:** 在前几份工作中, 一直期望能从事一段时间前端和后端的研究, 以更好地理解技术和产品; 在 CTO 好友的邀请下加入一家电商公司, 进军共享自动售卖机领域, 对我来说是个不错的学习契机; 但公司在后来遇到融资问题, 管理层也出现分歧, 思考再三后选择离开;
- **负责技术研发:** 完善电商 app 的搜索和推荐功能, 登录短信验证功能, 上线电商的拼团功能; 完成售货机的前端界面开发, 后台管理系统开发, 使用 MQTT 实现客户端到后端的长连接、实现机器货品和广告数据的热更新;
- **负责团队管理:** 负责前端团队、app 端团队的招聘、开发进度管理、代码审核; 组织技术分享、聚餐等团建活动;

2017.04 - 2018.04 高级算法工程师 网易 新闻推荐部 杭州

- **工作内容:** 参与网易新闻的头条、短视频、PUSH 的用户模型和推荐策略的研发工作;
- **丰富用户模型的特征工程:** 编写、部署和监控基础日志解析任务, 添加新的行为、并将其转为用户的兴趣点; 利用数据挖掘方法, 不断构造新的特征 (例如频繁项集、黄金点击、时长百分位、性别预测、时间段偏好等), 并通过 AB 测试进行特征选择;
- **高维度向量近邻检索的研究:** 完成 LSH (Locality Sensitive Hashing) 在多种距离标准下的 Java 实现, 解决大规模高维度用户向量在近邻搜索中的时间问题, 方便快速生成不同用户向量的近邻数据供推荐使用;
- **深度学习在用户模型的应用:** 使用 Tensorflow, 结合用户的基础特征和时序行为, 将多种深度学习模型 (DNN、LSTM、RBM 等) 作为编码器, 应用于用户向量的生成;
- **推荐策略优化:** 负责 KNN 推荐策略的开发、与不同用户模型组合的尝试、参数调试、AB 测试和上线; 目前 KNN 是网易新闻线上主力推荐策略之一;
- **新用户内容推荐:** 使用 bandit 算法, 结合平台用户的先验数据, 对新用户的冷启动推荐策略进行改进, 提升了新用户的留存率;
- **其它:** 使用 Levenshtein 对地理位置文本信息标准化; 使用 Greedy Set Cover 算法计算兴趣点的用户覆盖度, 优化用户的召回率; 团队技术分享; 等等

2015.10 - 2017.04 算法工程师 丁香园 杭州

- **工作内容:** 参与丁香园大数据部门的早期创建, 推动多项项目的实施, 在离开时, 贡献了当时部门项目的 90% 的代码;
- **推动公司日志改造:** 利用半年时间, 制定和统一公司所有产品线 (很多) 的日志打点规范, 并推动各产品线打点日志开发上线; 使用 Spark Streaming + SparkSQL + Hive 完成公司基础日志在集群上的解析和入库;

参与常用指标的 HiveQL 分析；改变了公司之前不重视各产品日志的统一收集、清洗、入库、分析，不能有效驱动业务，无法进行数据挖掘的状况；

- **流式计算**：推动流式计算（Kafka + Spark Streaming）在数据部门的广泛应用；
- **数据挖掘产品1，医生指数**：利用机器学习方法，设计了医生活跃度、影响力、专业度三个指标，通过医生在公司产品的实时行为、静态属性快速近实时地计算得到，指标很好地反映了医生用户在活跃、影响力、专业程度上的表现，为公司在数字广告、运营、数据变现方面提供商业支持；
- **数据挖掘产品2，医生画像**：利用机器学习和 NLP，结合公司的医学知识库，根据医生对内容的不同行为，实时计算用户的短期兴趣模型和长期兴趣模型，为公司在内容推荐、数字广告、运营、数据变现提供支持；
- **其他**：在公司的医生问答产品中根据患者的提问，使用推荐算法和 NLP 推荐科室与医生；使用 word2vec 优化搜索；参与医学知识图谱的研发；做过一段时间的 HiveQL 分析师；带团队新人；2016 年在公司产品部门的个人绩效为 A+，(top 5%, 7/189)

2014.9 - 2015.9 咨询顾问 Ernst & Young 金融咨询部门 上海

- **项目经历**：我所在的团队帮助安盛天平车险理赔业务完成新老系统的数据迁移咨询项目，3 个人的团队 4 个月为 E&Y 创造 350w 的收入；参与太平洋保险客户分级服务咨询项目，基于统计学方法，从客户属性与保单属性中设计一套保险客户的等级计算规则，识别高质量客户，并设计报表指标来帮助太保更好地管理服务质量；
- **保险业数据挖掘**：收集和抓取中外保险行业的数据，从数据趋势中分析国内的保险公司在科技、研发、模式上的项目机遇，为团队 BD 提供指导与支持；基于对特征工程、通用线性模型、随机森林、GBDT、神经网络、集成学习（bagging, boosting, stacking）、因子分析等数据挖掘方法的理解与熟练使用，带领数据团队参加了多个 kaggle 上保险相关的数据挖掘比赛：固定资产风险预测、火灾损失率预估、车险持有者的驾驶行为分析；

其他

- 博客：<https://mult1vac.me>
- 关注 KDD Cup、Kaggle、Data Science Bowl 等数据挖掘比赛，利用业余时间参加了一些 Kaggle 的数据挖掘比赛，比较好的成绩为：
 - Walmart 购物者购物行为预测(top 2%)
 - Otto Group 产品分类竞赛(top 5%)
 - Home Depot 搜索相关性预测(top 10%)
- 本科项目：和同学一起制作了一个基于 Arm Linux 的智能家居机器人，可以自动调节或通过网络控制家中电子设备，并在当年物联网比赛中获奖；

技能树

- 良好的英文阅读和写作能力；
- 作中熟练使用 Python/Scala/Java，熟悉 R；
- 熟悉 Mac/Linux 下的工程开发，熟练使用 Git，熟练使用 Maven/SBT，熟悉单元测试方法；
- 熟练使用常用机器学习算法：通用线性模型、随机树和森林、GBDT、深度学习（RNN，LSTM，seq2seq 等）、非监督模型（k-means，LDA，Factor Analysis）；
- 熟练使用 Python 下的 NumPy、sklearn、XGBoost、Keras、TensorFlow、gensim 进行机器学习应用的开发，熟悉 Python 下的爬虫开发和后端开发；
- 熟练使用 Java/Scala 基于 Hadoop、Spark、Spark Streaming、Hive、HBase、Redis、Kafka 的开发，熟悉 Elastic Search 并能快速搭建和优化搜索引擎；熟练使用 Hive-SQL 进行大规模数据分析和挖掘；

点击 [此处](#) 或扫描最右侧二维码，获取我最新简历的网页版。

