

# 张思凡

13260151230

zhangsifanbj@163.com

北京市海淀区

## 工作经历

微软互联网工程院	SDE 2	2022.11-至今
阿里巴巴 AE 技术部	推荐算法工程师	2020.07-2022.10
阿里巴巴 & 亚马逊 AWS	算法工程师（实习）	2019.05-2019.09

## 项目经历

阿里巴巴 AE 技术部	2020.07-2022.10
-------------	-----------------

分别负责 AE 推荐业务中排序及召回模块，从全链路视角进行迭代升级，不断提升整个推荐链路的业务效率，助力系统精准匹配用户兴趣，取得了加收转化率+7%，UV 价值+12%的效果。

➤ 排序：针对推荐策略实时调控的需要，提出实时深度控制模块，并横向拓展助力全场域提升

深度控制 LTR 作为大模型的后链路，摒弃传统的手工调权，直接设定线上期望业务目标，通过深度控制网络（PID），动态调整各 label 所对应的 loss 融合权重，以寻找当前限制下的最优参数。通过控制离线 auc 及模型参数变化的阈值，来保证在线效果的稳定性。取得 UV 价值+12%的效果。

全场域优化 对实时流程的上下游进行了抽象。样本侧，设置缓存区以归因延迟 label，设置缓冲队列以保持正负样本比例稳定。利用 drop rank 选取大模型 top10%的特征用于 LTR 训练。构建了从数据源、样本构造、到模型训练、线上生效等一整套配置化流程。根据不同场景的业务目标及场域特性，分别选用相应的特征、模型及打分组合公式，并进行超参寻优。于各个场域均取得了 IPV +2%，UV 价值 +2%以上的收益。

➤ 召回：针对多路召回合并时带来的马太效应，提出个性化多路召回融合模块

个性化召回融合 针对线上多路召回( $\geq 10$ )合并时，截断数量由人工拍定造成的马太效应与效率损失问题，提出了个性化召回融合方法。利用黑盒优化的思想，引入投票理论和 borda 计数法，根据 item 的相对顺序及召回对应的权重，确定 item 最终顺序。离线侧，采用生成器动态生成各路召回融合权重；在线侧，利用评估器对不同权重的线上效果进行评估，同时指导生成器的训练。取得了 GMV +2.4%的效果。

➤ 全链路：离线评估召回有效性，研究召回与排序的漏斗模式带来的耦合问题

离线评估 搭建 hitrate 评估流程，通过对超过 20 种召回源进行统一尺度下的评估，以离线验证召回数据源的有效性；通过各路召回组合评估，以发现可互补的召回源；通过粗排打分结果评估，以评定粗排的有效性。

解除耦合 针对推荐系统漏斗模式所带来的上下游链路耦合问题，评估各路召回曝光占比及转化效率，以验证耦合现象的存在。与精排联动，对召回来源进行泛化，削减耦合程度，助力更多召回源上线。

## 发表论文

- Yifan Zhu, **Sifan Zhang**, etc. Social weather: A review of crowdsourcing-assisted meteorological knowledge services through social cyberspace[J]. Geoscience Data Journal, 2020 (SCI-2 区)
- Yubing Nie, Yifan Zhu, **Sifan Zhang**, etc. Academic rising star prediction via scholar's evaluation model and machine learning techniques[J]. Scientometrics, 2019. (SCI-2 区)

## 教育背景

北京理工大学（985）	硕士	计算机学院 计算机科学与技术	2017.09 - 2020.07
暨南大学（211）	本科	电气信息学院 包装工程	2012.09 - 2016.07

## 获奖情况

北京市 2020 年优秀毕业生	2020.07
北京大学生优秀创业团队二等奖，创立了智法互动创业团队，已投入运营	2019.09
工信创新奖学金	2019.04