

邵维奇

188-1091-9359 | shaoweiqi@alu.ruc.edu.cn

个人网站:https://victorswq.github.io/x/

北京/上海/广州/深圳 | 求职意向:短视频推荐算法



教育经历

中国人民大学

2020年09月 - 2023年06月

人工智能 硕士 高瓴人工智能学院

北京

导师: [陈旭](#) 准聘副教授 GPA: 3.47 / 4.00 荣誉奖项: 2020/2021/2022中国人民大学二等学业奖学金, SIGIR-AP2023最佳论文提名 研究方向: 基于用户多兴趣推荐系统/基于多智能体的推荐系统 (2023中稿会议:SIGIR-AP,WWW,IJCNN) (2023中稿期刊:TOIS) 项目经历: [RecBole](#) 开源推荐系统代码库开发

中央财经大学

2016年09月 - 2020年07月

信息管理与信息系统 本科 信息学院

北京

导师: [李海峰](#) 副教授 GPA: 2.80 / 4.00 荣誉奖项: 2020中央财经大学校级社会实践优秀奖学金 课外学习: ACCA特许公认会计师(8门) 资格证书: JAVA计算机二级, CET-4/6 毕业设计: 基于强化学习的五子棋智能体, 基于深度强化学习和蒙特卡洛树搜索, 采用强化学习离线对抗学习方法, 让计算机通过自我对局提升下棋能力

工作经历

阿里巴巴

2023年08月 - 2023年11月

算法工程师 阿里妈妈外投广告算法团队 主管: 亮博(王亮) Mentor: 翊冬(贾易东) 乙骨(刘嘉熙)

北京

- **字节user embedding模型**: 随着短视频媒体(快手, 抖音)快速崛起并占据大量用户时长, 阿里广告主有强烈的需求通过站外媒体引流到淘宝站内, 外投广告算法团队的任务是将广告投放到其他的媒体平台, 对接字节侧流量进行广告外部投放时, 阿里侧的每个参竞广告, 其预估曝光价值由媒体侧精排 CTR 模型、阿里侧精排 CVR 模型和机制模型协同给出, **媒体字节侧的user embedding向量表征信息含有丰富的用户信息**, 有利于阿里侧精排 CVR中的用户建模, 可以对用户画像进行准确描述, 阿里精排CVR模型利用字节user embedding信息提高CVR转换率预估, **将字节侧328维用户特征向量加入阿里CVR精排模型丰富了模型特征**, 针对用户信息缺失, 制定全零填充用户向量表征冷启动, **结果在离线AUC指标上提升了13个百分点**。后续打算针对用户向量表征制定合理高效的特征提取模块神经网络, 采用更加高效的冷启动模型预估精排 CVR。
- **退款率模型开发**: 为了促进用户消费, 淘宝和天猫会设置购物狂欢大促节, 相比于非大促期间, **大促期间用户购买量增加**, 但是用户的退款率也会增加, 同时用户退款存在时间延迟, 无法进行及时跟踪并修改数据标签, 而在阿里侧CVR精排模型不断做增量更新线上模型, 无法及时更新的退款数据会导致CVR精排模型虚高CVR真实转换率, **如何提前修正该数据进而增强CVR精排模型预估变得非常重要**。只生成了数据, 还未来得及进行下一步实验便进入了其他项目。后续可以采用的方法包括对用户建立合理的退款率预估模型提前更新数据标签, 利用修正后的数据训练模型从而淡化模型偏差。

百度

2021年06月 - 2023年04月

算法实习生 搜索策略部 主管: 夏龙/吕晓勇 Mentor: 夏龙/陈旭/赵佳枢/殷大伟

北京

- **研究型实习生**: 在[殷大伟](#)博士, [夏龙](#)博士, [赵佳枢](#)助理教授和[陈旭](#)准聘副教授的指导下, 做有关**用户多兴趣推荐算法**方面的研究工作, 针对用户多兴趣进行建模提取用户多兴趣, 不同用户兴趣数量不一致, 同一个用户的兴趣数量也在不断发生变化, **如何针对用户多兴趣进行高效建模**会遇到用户兴趣数量不一致, **用户兴趣演化**, **时间距离**等因素影响用户兴趣划分, 我们提出了SPLIT模型建模用户多兴趣, SPLIT设计了物品分配器模块和子序列建模模块, 物品分配器模块动态切分用历史点击序列形成兴趣子序列, 将切分好的兴趣子序列放入子序列建模模块形成用户多兴趣, 将用户多兴趣按权重相结合并计算其与候选物品之间的相关性, 将相关性较强的物品推荐给用户, 在离线实验中, 模型在八个三种不同公开类型的离线数据上, **Precision, MRR, Recall, NDCG**等指标均取得了超过7%的提升, 文章中稿SIGIR-AP2023, 并荣获**最佳论文提名**。
- **业务型实习生**: 百度好看视频是百度的网页版短视频, 视频中存在大量的OCR信息, 如何利用好这些信息用于模型搜索将提升模型搜索效果, 百度视频搜索采用预训练好的ERNIE模型, 输入框长度固定, 视频OCR信息长度可能超过输入框, 也有可能比输入框长度短, 在**固定输入框的条件下**采用特征长度自适应筛选, 将包括标题, 文字, 内容等OCR信息送入训练增加搜索相关性从而增加搜索准确性, 离线实验有所提升。后续打算在固定输入框前面加入神经网络模块增大输入框长度, 从而完全利用视频OCR信息, 提高搜索准确率。(Mentor: 苏立新/陈丽娟)

项目经历

RecBole 开源推荐系统代码库2020年09月 - 2020年12月

序列化推荐系统论文代码复现 AIBOX北京

- 随着学届和工业界推荐系统的不断发展，统一的推荐系统开源代码库有助于科研人员以及工业界人员进行**更加公平，更加快速**的综合实验，RecBole是基于深度学习的开源推荐系统代码库，包括序列化推荐系统，基于内容的推荐系统，基于知识图谱的推荐系统等，该开源代码库具有很多优点，包括统一的数据格式，评价标准，模型格式，日志记录等，快速上手同时易于理解，该开源模型代码在 [GitHub](#) 已获得 **3300+ stars**。我根据相应论文复现了**六个序列化推荐系统**，包括 **FOSSIL，SHAN，HGN，HRM，NPE，RepeatNet**。

研究经历

Sequential Recommendation with User Evolving Preference Decomposition.2021年06月 - 2023年11月

(SIGIR-AP 2023, Best Paper Honorable Mention)

WeiQi Shao, Xu Chen, Long Xia, Jiashu Zhao, Jingsen Zhang and Dawei Yin.

- 首篇从用户**多/动态/进化**兴趣角度出发, 动态切分用户点击序列形成用户多兴趣子序列, 从而提升推荐效果，提出的且分方法也可用作**动态序列切分框架**。
- 通过创新强化学习动作空间 (**增加一个动作**)来实现动态调节强化学习输出空间，从而动态切分用户点击序列；对于生成好的子序列，采用不同类型的**序列用户建模**提取用户的多兴趣。

Aggregation Recommendation from Heterogeneous Videos via Multi-Agent Reinforcement Learning. (IJCNN2023,Computer Class A Conference)2022年03月 - 2022年05月

Yabin Zhang, WeiQi Shao(Co-First Author), Xu Chen, Yali Du, Xiaoxiao Xu, Dong Zheng, Changhua Pei, Shuai Zhang, Peng Jiang, Kun Gai.

- 快手视频推荐**-针对三种不同的业务(**直播/广告/短视频**)；给每一种业务设置一个**源智能体**进行推荐相应的业务，三个业务源智能体给最终聚合智能体采用**多智能体强化学习**推荐不同类型视频提升推荐效果。
- 面向不同类型视频进行两阶段选择：源视频选择，视频聚合推荐。

Debiased Recommendation with User Feature Balancing. (TOIS 2023, CCF-A)2022年02月 - 2022年02月

Mengyue Yang, Guohao Cai, Furui Liu, Jiarui Jin, Zhenhua Dong, Xiuqiang He, Jianye Hao, WeiQi Shao, Jun Wang, Xu Chen.

Recommendation with Causality enhanced Natural Language Explanations.2022年03月 - 2022年03月

(WWW 2023, CCF-A)

Jingsen Zhang, Xu Chen, Jiakai Tang, WeiQi Shao, Zhenhua Dong and Rui Zhang.

其他

- 学习上刻苦认真(考研笔试第一名)，英语一85，品学兼优;热爱运动
- [GitHub主页](#)/[谷歌学术](#)/[博客主页](#)/[个人网站](#)