

卫文康

电话: (+86) 18211209213 | 邮箱: awk669@qq.com

个人网站: wenkangwei.github.io/about/, <https://wenkangwei.gitbook.io/paper-read> | 学术网站: meritpages.com/wenkan

个人总结

我是一个学习能力强并热爱技术创新及落地的 6 年+算法经验的工程师。我毕业后 3 年内在 SHEIN 从算法工程师 (S4-2) 晋升至资深算法工程师 (S6-1)，作为项目经理/主开发推动过 **3+个部门级项目** 在搜推场景落地推广，也曾申请相关发明专利。我目前对搜推算法/多模态/多模态模型在搜推算法应用/模型目标建模/loss 设计等很感兴趣。

教育背景

Clemson University(克莱姆森大学)	美国南卡罗莱纳州
● 硕士(全奖本硕连读): 计算机工程, 副修计算机科学	2019.08 - 2021.05
● 本科: 电子工程	2016.08 - 2019.05
● 院内优等生名单 Dean's List 年级排名前 1%(2016.06 - 2019.05); 校内优等生名单 President's List (2019.06). 校内 "CUhackit" 极客比赛一等奖 (2018.01), Kaggle 木薯叶疾病图像识别比赛铜牌 (2020.02) 等	

专业技能

编程工具: Python(Tensorflow, PyTorch, PySpark 等), SQL, C/C++, Linux, Spark, Hadoop, AWS/Azure/GCP 等
专业能力: 搜索/推荐算法 (DSSM/DCN/DIN/蒸馏学习/全链路一致性建模/长序列建模/多模态入模/排序模型建模等), 特征交叉 (DCN/DeepFM 等), 机器学习 (如无监督聚类等), NLP (SFT/word2vec/transformer 等), 数据分析等

工作经历

SHEIN(希音)	深圳, 南山区
● AI 实验室, 高级算法工程师 (S5) 资深算法工程师 (S6)	2022.06 - 至今
负责调研实现和落地推广通用算法能力到搜推场景 (级联学习, 多模态 LLM 大模型搜推应用等项目)	
● 推荐算法, 算法工程师 (S4)	2021.10 - 2022.06
负责列表页场景精排模型建模迭代	
● 发明专利(美国+中国): 对于电商商品的粗排打分模型进行训练的方法及相关产品	受理日期: 2024.04.09
描述: 一种基于 Min-Max 思想和蒸馏学习的 listwise 多目标级联粗排训练方法	

部分项目经历

多模态大模型在搜推应用探索 - 项目负责人 项目经理	2024.06 - 至今
● 目的: 探索多模态和 LLM 大模型在搜推领域应用价值, 设计和落地多模态能力在搜推应用使用方法	
● 行动 1(多模态生成和评估): 联合 NLP 同学设计商品描述文本数据, 对多模态 CLIP 模型, Matryoshka for Bert 嵌入模型 fine-tune (SFT) 导出图文向量(维度可选). 利用类目相似度和可视化工具评估召回 top-N 商品相似度	
● 行动 2(多模态搜推模型建模探索): Step1: Late-Fusion 方式建模并消融实验验证多模态向量在推荐模型直接入模和行为序列向量的线上差异和模型表现; Step2: 参考 MMBee/MAKE/SimTier/DCN 等方法思路设计出多模态向量融合模块, 提高信息流粗排离线 CTR AUC +0.5%; Step3: step2 基础上利用 cross-attention, 多模态行为序列特征交叉, 多模态 loss 建模等方法让模型学习多模态信息增益, 提升粗排模型效果 (技术细节可详谈)	
● 结果: 在信息流场景离线评估点击 AUC 和加车 AUC 各提升 +0.5% ~ 2%, 多个国家大区线上点击率显著提升+1%以上, 目前流量覆盖 100% 到 2+个大区	
从 0 到 1 搭建和落地搜推粗排精排链路一致性建模(级联学习) - 项目负责人 项目经理	2023.05 - 2023.12
● 目的: 为解决公司搜推排序链路(粗排精排)打分为一致确保召回优质商品推送给用户, 从数据到模型系统化搭建公司搜推场景首个粗排级联学习链路	
● 行动 1: 独立调研蒸馏思路和开发支持精排(大模型)粗排(小模型)联合训练框架功能(方案 1), 搭建从 0 到 1 搭建同时支持 pairwise 和 list-wise 的蒸馏学习粗排数据流(方案 2)	
● 行动 2: 基于 list-wise loss 蒸馏学习多目标优化思路发明一种计算复杂度低 list-wise 序蒸馏 loss, 在信息流, 列表, 搜索等场景离线粗排精排打分序列 NDCG 相对提升最高+5%	

卫文康

电话: (+86) 18211209213 | 邮箱: awk669@qq.com

个人网站: wenkangwei.github.io/about/, <https://wenkangwei.gitbook.io/paper-read> | 学术网站: meritpages.com/wenkan

- 结果 1: 3 个月内联合搜推 10+ 个同学落地级联粗排方案到搜索推荐所有场景, 流量 100% 覆盖美国欧洲等大区, 所有搜推实验的场景内人均销售指标显著提升在 +0.55% ~ 4.7%, 其中商详场景曾单个实验大盘人均销售 GMV 显著提升最高达 +1.41%
- 结果 2: 发表发明专利(中国+美国, 已受理): 基于 Min-Max 思想和蒸馏学习 list-wise 多目标级联学习训练算法

列表页精排大区模型开发 - 模型开发

2023.02 - 2023.03

- 目的: 为了减少模型维护成本, 探索多个国家维度模型合并成一个大区模型的方法并提升精排模型线上效果
- 行动 1: 探索和利用不同动态权重模型结构(如站点/国家特征向量, 动态权重, LHUC)对不同国家数据同时学习, 并与不同国家独立模型进行离线评估对比
- 行动 2: 利用 LHUC 结构带来不同国家线上效果持平或微正提升后, 利用 slot-shuffle 特征选择工具选取一批重要性较高的用户统计特征, 进一步提升精排对用户刻画能力
- 结果: 在列表页场景离线评估中不同国家点击 AUC 提升 0.6%~1.8%, 加车 AUC 提升 0.89%~2%。线上指标 uv 销售提升 +0.78%, pv 点击率提升 +0.59%

分布式增量训练框架开发 - 项目主开发

2023.01 - 2023.05

- 目的: 实现 TensorFlow 分布式增量训练能力提高模型训练速度和对 CPU 资源的利用率达到提效功能
- 行动 1: 调研并向业务算法团队统一实现 TensorFlow Estimator 和 Keras 分布式训练框架(同步/异步/混步训练策略), 以及外围功能如商品向量增量增删更新, 不同云存储适配(google cloud/azure/aws), 模型训练容错机制等
- 行动 2: 利用 TensorFlow 分布式增量训练框架(异步策略)打破单机训练的数据并行瓶颈, 将基线精排模型训练时间从 21H 降低至 6H, 减少训练时间 2/3
- 结果: 消融验证功能并推广 TensorFlow 分布式增量训练框架到搜推所有场景, 功能覆盖率达到 100%