



杨振坤

(+86) 135-0022-5946 yzknower@163.com

求职意向：推荐算法实习生



教育背景

华南理工大学 (985/双一流)	信息与通信工程	硕士 (推免)	2023.09 ~ 至今
• 研究方向: 图神经网络与推荐系统			
合肥工业大学 (211/双一流)	电子信息工程	本科	2019.09 ~ 2023.07
• GPA : 3.80/4.3	专业排名: 3/143 (前 2%)	安徽省优秀毕业生 (全院 <1%)	

实习经历

腾讯科技有限公司微信事业群 广州 2024.08 ~ 2024.11

问一问推荐算法组

- 针对单机版图召回模型增量训练过程中内存需求激增导致的资源申请失败问题，利用现有的 embedx 图模型框架，成功实现了模型增量训练流程的并行化改造。这一改进显著缩短了训练时间，从原先的 10 小时缩减至半小时，提升了训练效率达 20 倍。
- 通过对新老用户热门 item 流量占比的深入分析，发现新用户存在流量过热现象，需进行流量消热处理。进一步分析发现，系统在高爆低点现象上表现不明显。基于此，我设计并实施了一项实验，通过负样本上采样和正样本下采样的方法对热门样本进行处理。实验结果表明，正样本下采样能够在保持点击率稳定的同时，提升新用户留存率 3%，并将实验进行了推全，增强了系统个性化分发。
- 面对问一问场景下新用户特征稀少，难以精确捕捉用户兴趣点的挑战，创新性地利用相似场景下公众号的数据进行跨域推荐。通过将新用户阅读过的公众号文章映射到问一问中最相似的问题，并据此为新用户推荐相关内容，有效提升了新用户的点击率 2%。

钛动科技股份有限公司 (未来独角兽) 广州 2023.04 ~ 2023.08

DSP 项目部 点击率预测和智能竞价 (被评为年度最佳项目)

- 本项目旨在解决钛动科技 Ads DSP 平台中展示广告的点击率预测问题，通过建立 CTR 预估模型实现广告点击率的精准预测，从而对广告投放策略进行调整，提升广告投放效果，提高广告主的 ROI。
- 对业务实验数据集进行详细的统计特征分析，在特征分析的基础上，进行了数据集的下采样、数据划分、异常值缺失值处理、连续特征归一化、特征构造编码等一系列特征工程构建。在现有模型上进行改进，通过在 DeepFM 模型中引入注意力机制，再利用 GBDT 训练数据集中的连续特征，将 GBDT 的叶子节点信息与离散特征合并成新的输入特征，充分考虑到不同模型的特长，创新性提出了 GBDT-DeepAFM 模型。
- 利用 Pytorch 和 LightGBM 实现模型并进行离线实验，结果表明 GBDT-DeepAFM 模型在 AUC 和 Logloss 都有较好的表现。模型上线后击败第三方 (Cusper) SaaS 服务，实现了 ROI 从 0.7 到 1.4。

科研经历

Efficient Balanced Multi-view Graph Clustering via Good Neighbor Fusion 2023.01 ~ 2024.03

IEEE Transactions on Pattern Analysis and Machine Intelligence 录用 (人工智能顶刊, IF=23.6) 第一作者

- 本论文旨在解决多视角图聚类中的跨视角一致性信息融合、聚类平衡和求解效率的问题。
- 提出利用锚节点判定节点关系的基本框架，考虑跨视角的差异设计跨视角良好邻居投票 (CGNV) 模块，通过投票机制识别可靠的先验矩阵，获得先验标签，增强了从跨视角数据中挖掘一致性信息的能力。
- 引入了一种新的平衡归一化项，通过调整超参数来控制聚类结果的平衡性，并且通过严格的数学推导提供了理论证明，实现了定量控制聚类结果的平衡性。
- 在 COIL20、DIGIT10 等 6 个基准数据集上进行测试，和 AMGL、SwMC、MLAN 等 11 种主流多视角图聚类算法进行对比在 ACC、F1、NMI、ARI 以及运行效率上都取得了最佳性能。

荣誉奖项

- 国家级：2021-2022 学年国家奖学金 (<1%)、2019-2020、2020-2021 学年国家励志奖学金
- 省级：安徽省大学生统计建模大赛一等奖、全国大学生数学建模竞赛安徽赛区二等奖、安徽省机器人大赛三等奖、安徽省计算机设计大赛三等奖
- 校级：2019-2020 学年合肥工业大学校三等奖学金、2020-2021、2021-2022 学年合肥工业大学校一等奖学金、2020-2021、2021-2022 学年合肥工业大学校三好学生

专业技能

- 熟练掌握 Python 语言和 Pytorch 深度学习框架，能快速搭建 CNN、LSTM、Transformer、GCN 等模型。
- 熟悉推荐系统中召回、粗排、精排、重排等链路，掌握 CF、FM、双塔、Wide&Deep、DIN、DPP 等模型。
- 熟悉线性回归、PCA、K-means、SVM、Bagging、随机森林、XGBoost 等机器学习算法。
- 通过 CET-4(552)、CET-6，具有英文论文写作经验，能流利阅读英文文献和工作文档。