

教育背景

中国科学技术大学, 计算机科学与技术学院, 计算机应用技术	2021.9 – 至今
在读硕士研究生, GPA: 4.06 / 4.3, 研究方向: 推荐系统, 自然语言处理	
中国科学技术大学, 少年班学院, 计算机科学与技术	2017.9 – 2021.6
工学学士, GPA: 3.97 / 4.3, 排名: 3 / 253	

实习与项目经历

OPPO 研究院, 行知实验室, 研究实习生	2023.2 – 至今
------------------------	-------------

- 用户每天在 OPPO 的众多硬件端会产生大量的时序行为数据, 其中蕴含着丰富的用户信息。该项目主要关注如何基于用户的 **App 相关时序行为数据** 得到预训练用户表征, 以此赋能各种下游业务场景。
- 针对用户 **多样化且富含噪声** 的时序行为数据, 我设计了一种 **通用预训练方法**, 在游戏拉新、用户召回、App 推荐等多个下游业务场景取得了平均 1.3% 的 AUC 相对提升, 相关工作正在接受审稿中。

智慧教育知识图谱 (LUNA), 实验室项目 [项目链接]	2021.5 – 2023.2
-------------------------------	-----------------

- 随着教育信息化浪潮席卷全球, 线上教育资源呈爆炸式增长。该项目旨在对互联网上的海量 **多模态** 教育资源进行收集、标准化和分析, 并向学生和教师提供各种 **智能化教育服务**。
- 我主要负责 **教育资源搜索系统**, 基于 BM25 算法和预训练语言模型构建了完整的 **召回 – 排序** 流程, 并进一步采用 **量化和知识蒸馏** 技术对所使用的预训练语言模型进行压缩, 大幅降低服务时延和对服务器资源的需求。

微软亚洲研究院, 社会计算组, 研究实习生	2020.9 – 2021.5
-----------------------	-----------------

- 我主要对新闻理解与新闻推荐进行研究, 为 **MSN 新闻服务** 开发了基于预训练语言模型的 **多语言新闻推荐模型**, 在 8 个地区取得了 13%~32% 的 AUC 相对提升以及 10%~43% 的 NDCG@10 相对提升。

华为技术有限公司, 联合研究项目	2020.3 – 2020.12
------------------	------------------

- 该项目旨在将图像超分、图像去噪等大规模 **生成式模型** 进行压缩与加速, 从而能够部署到移动端进行实时处理。
- 我首先复现了若干图像超分方法, 然后利用 **剪枝和量化** 技术, 在保持生成质量的同时将模型规模减小约 70%。

发表论文

- Untargeted Attack against Federated Recommendation Systems via Poisonous Item Embeddings and the Defense (AAAI'23, CCF-A, 一作)

该工作对于针对 **联邦推荐系统** 的非定向投毒攻击及其防御机制进行研究。我们首先提出了一种基于 **聚类** 的高效 **非定向模型投毒攻击** 方法, 仅需 1% 的恶意客户端就可使推荐模型的性能下降约 40%。我们进一步设计了一种基于 **物品表征均匀度** 的防御策略, 能够有效提升联邦推荐系统对于各种非定向投毒攻击的抵抗能力。[Paper] [Github]

- Tiny-NewsRec: Effective and Efficient PLM-based News Recommendation (EMNLP'22, CCF-B, 一作)

为使预训练语言模型更好地适应新闻领域, 我们提出了一种基于 **对比学习** 的自监督 **后训练** 方法。此外, 为降低预训练语言模型在提供线上实时服务时对于计算资源的需求, 我们还设计了一个 **两阶段多教师知识蒸馏** 框架, 在保持模型性能的同时将模型大小缩减了 50%~70%, 并将推理速度提升了 2~8 倍。[Paper] [Github]

- HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation (ACL'21, CCF-A, 共作)

针对如何更准确地从历史新闻浏览记录中提取用户 **多样化的兴趣**, 我们设计了一种 **层级兴趣树** 结构, 分别捕获用户的综合兴趣、粗粒度主题兴趣和细粒度主题兴趣, 在改善推荐性能的同时大幅提升了推荐结果的多样性。[Paper]

- NewsBERT: Distilling Pre-trained Language Model for Intelligent News Application (Findings of EMNLP'22, CCF-B, 共作)

大规模预训练语言模型在赋能诸多智能新闻应用的同时也给线上服务延迟带来了挑战, 为此我们设计了一个 **教师 – 学生联合蒸馏** 框架, 在学生模型模仿教师模型的训练过程中, 通过 **动量蒸馏** 的方法向其注入教师模型的梯度信息, 在保持模型性能的同时大幅降低了模型的大小与推理时延。[Paper]

获奖情况

华为奖学金 (校前 2%), 中国科学技术大学	2019, 2022
硕士研究生一等奖学金, 中国科学技术大学	2021, 2022
“明日之星”优秀实习生奖, 微软亚洲研究院	2021
国家奖学金 (全国前 0.2%), 中华人民共和国教育部	2020

专业技能

- 编程语言: 熟练使用 Python, 熟悉 SQL, Shell, 了解 C/C++, HTML, JavaScript
- 框架与工具: 熟练使用 PyTorch, \LaTeX , Git, 了解 Flask, Django, Vue.js, Docker
- 语言能力: 英语四级: 654, 英语六级: 597, TOEFL: 105