

葛晨笛

电话: 19801280260 | 邮箱: gcd23@mails.tsinghua.edu.cn
个人网站: <https://gcd19.github.io/>

教育经历

清华大学 2023年09月 - 2026年06月
计算机科学与技术 硕士 计算机系 北京
研究方向: 多模态大语言模型、自动机器学习、图机器学习 **导师**: 朱文武教授

清华大学 2019年08月 - 2023年06月
计算机科学与技术 本科 计算机系 (GPA: 3.8/4.0) 北京
经济与金融 辅修 经济管理学院

实习经历

阿里巴巴集团 2024年05月 - 2025年02月
大模型算法实习生, 第一作者成果被 CCF-A 类会议 ICML 2025 [1] 接收。 北京 / 杭州

- 研究多模态大语言模型在持续指令微调中的灾难性遗忘问题, 提出 D-MoLE 方法, 结合任务敏感的层级 LoRA-MoE 分配与模态感知课程机制, 在参数预算有限的条件下兼顾新任务适应与旧任务能力保留。方法在多模态持续学习基准上抗遗忘能力较 SOTA 平均提升约 15%。成果为阿里多平台 (淘宝、闲鱼、1688 等) 联合电商审核大模型在任务持续演化场景下的训练策略探索提供了方法支持。

腾讯 TEG 2023年04月 - 2023年12月
推荐算法实习生, 第一作者成果发表于 CCF-A 类会议 AAAI 2025 [2] (Oral)。 北京

- 研究跨域推荐中的架构定制与信息迁移问题, 针对固定模型适配能力不足及源域行为噪声引发的性能退化, 提出 BiGNAS 框架, 结合任务定制的 GNN 超网络与行为重要性建模, 自动搜索各域最优架构, 并动态评估源域行为价值以引导跨域信息迁移。方法在微信广告跨域推荐数据 (公众号、朋友圈、视频号) 上验证, AUC 平均相对提升 3.4%。成果帮助微信多场景广告推荐业务探索跨域结构定制与信息迁移的建模路径。

主要研究工作

[1] **Chendi Ge**, Xin Wang, Zeyang Zhang, Hong Chen, Jiawei Fan, Longtao Huang, Hui Xue, Wenwu Zhu. Dynamic Mixture of Curriculum LoRA Experts for Continual Multimodal Instruction Tuning. **ICML 2025 (CCF-A, 第一作者)**.

- 提出基于动态混合专家的持续多模态指令微调方法 D-MoLE, 结合层级 LoRA-MoE 的任务敏感分配与梯度驱动的模式课程机制, 缓解持续多模态指令微调中的架构冲突与模式不平衡问题。在多模态持续学习基准上平均抗遗忘性能较 SOTA 提升约 15%, 并在 MME、MMM、POPE 等基准中展现出良好的通用能力保留效果。

[2] **Chendi Ge**, Xin Wang, Ziwei Zhang, Yijian Qin, Hong Chen, Haiyang Wu, Yang Zhang, Yuekui Yang, Wenwu Zhu. Behavior Importance-Aware Graph Neural Architecture Search for Cross-Domain Recommendation. **AAAI 2025 Oral (CCF-A, 第一作者)**.

- 提出行为重要性感知的图神经架构搜索框架 BiGNAS, 通过联合优化任务定制的 GNN 架构与源域行为重要性建模, 提升跨域推荐中的结构适应性与信息迁移效果。在 Amazon 和微信广告跨域推荐数据上, AUC 较 SOTA 平均相对提升 5.6% 和 3.4%, 有效缓解因固定架构性适配不足与源域行为噪声导致的性能退化。

专业技能

- 编程语言: Python、C/C++、Java、Rust 等
- 深度学习与大模型框架: PyTorch、Transformers、PEFT、DeepSpeed、vLLM、SGLang、verl 等