

杨子逸

手机：19924680829

邮箱：yangzy39@mail2.sysu.edu.cn

个人主页：<https://yangzy39.github.io>

实习意向：大模型算法 NLP 算法



教育背景

中山大学 计算机学院 2023.09 - 2026.06

计算机技术专业，硕士研究生，导师：权小军教授

- 研究方向：大模型融合 偏好优化 长思维链推理

中山大学 计算机学院 2019.09 - 2023.06

计算机科学与技术专业，本科

- 相关课程：自然语言处理 人工智能 机器学习与数据挖掘

论文

[1] Ziyi Yang, Fanqi Wan, Longguang Zhong, Tianyuan Shi, and Xiaojun Quan. Weighted-reward preference optimization for implicit model fusion. **ICLR 2025**, poster

[2] Ziyi Yang, Fanqi Wan, Longguang Zhong, Canbin Huang, Guosheng Liang and Xiaojun Quan. FuseChat-3.0: Preference Optimization Meets Heterogeneous Model Fusion. **SCI-FM @ ICLR 2025**, poster

[3] Longguang Zhong, Fanqi Wan, Ziyi Yang, Guosheng Liang, and Xiaojun Quan. FuseRL: Dense Preference Optimization for Heterogeneous Model Fusion. **NeurIPS 2025**, under review

[4] Fanqi Wan, Ziyi Yang, Longguang Zhong, Ruijun Chen, Xiaojun Quan. FuseChat: Knowledge Fusion of Chat Models. **NeurIPS 2025**, under review

[5] Tianyuan Shi, Canbin Huang, Fanqi Wan, Longguang Zhong, Ziyi Yang, Weizhou Shen, Xiaojun Quan, Ming Yan. Mutual-Taught for Co-adapting Policy and Reward Models. **ACL 2025**, under review

科研经历

首次提出了基于偏好优化的隐式模型融合研究问题，旨在从偏好优化的角度出发，将多个功能强大源模型的能力隐式地整合并迁移到一个规模更小更加高效的目标模型中。相关研究成果发表于机器学习领域顶级会议 ICLR，并获得了港理工杨红霞教授和哈工深杨晶教授等团队的关注

基于加权奖励偏好优化的隐式模型融合 2024.06 - 2024.10

[Paper] / [Github] / [HF] / [AI Time]

• 研究内容

- 针对现有模型融合中的词表对齐与分布矩阵对齐困难、效率低下等问题，提出加权奖励偏好优化方法 (WRPO)，其核心思想是让目标模型隐式地从源模型和自身回复之间的差异中进行学习
- 为解决目标模型与源模型间的分布偏差，提升融合稳定性与效率，提出内部奖励加权方案，并结合渐进式调整策略，使目标模型逐步学习来自源模型的偏好样本

• 研究成果

- WRPO 显著优于相同参数规模的模型融合方法，媲美 106 倍参数规模的模型集成方法；在 AlpacaEval-2 评测集上超越所有参与融合的模型，相较于传统偏好学习 DPO 方法提升了 7.7 分
- 以第一作者身份在国际机器学习顶会 **ICLR2025** 发表论文一篇

FuseChat-3.0：偏好优化邂逅异构模型融合 2024.10 - 2024.12

[Paper] / [Github] / [HF] / [魔搭社区]

• 研究内容

- FuseChat-3.0 是对 WRPO 核心思想的实践延伸，进一步改进并拓展了融合数据的领域和规模
- 针对数学和代码领域，引入基于规则验证的数据合成方案，在 DPO 损失中增加长度约束项

- 研究成果

- 融合模型 FuseChat-3.0-8B 在 14 个基准测试集上的平均性能相比目标模型 Llama-3.1-8B-Instruct 提升 **16.8%**，成为 AlpacaEval-2 和 Arena-Hard 榜单中**最强 8B 模型**
- 成果被阿里云**魔搭社区**报导，并获国家超级计算广州中心 2024 “天河之星”优秀应用入围奖
- 以第一作者身份在国际机器学习顶会 **SCI-FM @ ICLR2025** 发表论文一篇

FuseRL：面向异构模型融合的密集奖励偏好优化

2024.10 - 2025.01

[Paper]

- 研究内容

- FuseRL 提出一个新的异构模型融合框架，核心思想是最大化源模型回复的利用率
- 对每条输入引入来自源模型的多个回复或偏好对，结合奖励分数进行加权 SFT 或加权偏好优化

- 研究成果

- FuseRL 结合偏好优化算法 DPO/SimPO 在 AlpacaEval-2 测试集上分别提升 3.0/5.9 分
- 以第三作者身份在国际机器学习顶会 **NeurIPS2025** 投稿论文一篇

项目经历

FuseChat: 对话大模型融合

2023.10 - 2024.05

[Paper] / [Github] / [HF] / [mergekit]

- 负责部分：主要负责模型合并算法的设计与优化、模型训练以及性能评测

- 项目内容：

- 针对以往多教师蒸馏融合可扩展性差的问题，提出成对教师蒸馏加模型合并的两阶段融合框架
- 为解决现有词表对齐方法适用场景受限的问题，提出基于统计的全局映射矩阵对齐方法
- 为减轻模型合并中不同模型的知识干扰问题，提出自动化合并系数计算的细粒度模型合并算法

- 项目成果：

- 融合得到的 FuseChat-7B 是当时多轮对话评测集 MT-Bench 上**最佳 7B 模型**
- 该成果入选 HuggingFace 每日论文精选，Github 项目仓库收获超过 **600 星标**
- 提出的模型合并算法 SCE 被 Github 超过 **5600 星标** 的知名模型合并仓库 mergekit 收录

FuseO1: 推理大模型融合

2025.01 - 至今

[Blog] / [Github] / [HF]

- 负责部分：主要负责评测框架搭建、训练数据集构建以及推理模型行为分析

- 项目内容：

- 将 FuseChat 中提出的 SCE 模型合并方法应用于推理大模型融合，复现 DeepSeek-R1 完整流程
- 探究长文本/文档场景下的强化学习，研究推理模型长思维链出现背后各种行为的可解释性

- 项目成果：

- 通过 SCE 合并算法将 DeepSeek-R1-32B, QwQ-32B, Sky-T1-32B 进行融合，在 AIME24 数学评测集上相较 DeepSeek-R1-32B 提升 4.8 分，得到当时**最强的 32B 推理模型**
- 融合模型发布第三天在 HuggingFace 首页亮相，目前总下载量超过 **10 万次**