杨子逸

手机：19924680829
邮箱：yangzy39@mail2.sysu.edu.cn
个人主页：https://yangzy39.github.io
研究方向：模型融合、推理模型、强化学习

教育背景

中山大学　计算机学院 2023.09 - 2026.06
硕士研究生　计算机技术　导师：权小军教授

中山大学　计算机学院 2019.09 - 2023.06
本科　计算机科学与技术

实习经历

阿里通义实验室　NLP 文档智能团队 2025.05 - 至今

- **业务模型优化**：主导 Qwen-Doc 模型的 RL 流程优化，基于 DAPO 算法，在 14 个 B 端业务评测集上实现**平均性能提升 5.17 分**，成功推动模型上线，当前客户日调用量达**千万级**
- **基础模型研发**：作为核心成员参与 QwenLong 基座模型研发，实习初期参与 L1 模型技术报告撰写，该模型性能媲美 o3-mini；现阶段作为核心成员参与 L1.5 系列模型全链路研发，贡献**自进化数据合成方案**与**多任务 RL 动态均衡采样器**，并搭建了多领域模型性能评测框架
- **前沿技术探索**：开展长上下文场景的**模型自进化强化学习**研究，在无需任何人工标注的情况下，模型根据输入的长文档进行自我出题、解题、校验，并结合 RL 算法同步优化三种角色能力，基于 Qwen2.5-7B 的实验在八个长文本评测集上取得**超过 50% 的平均性能提升**，成果计划投稿 ICLR 2026

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。其余研究领域包括大模型自我进化，自适应推理，长上下文推理等

基于加权奖励偏好优化的隐式模型融合 [ICLR 2025] / [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 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 [ICLR SCI-FM] / [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 [Preprint]

FuseRL 的核心思想是最大化隐式模型融合过程中不同源模型回复的利用率，通过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，显著提升模型指令遵循能力

策略与奖励模型协同适应的互教学习 [ACL 2025 Main]

互教学习（Mutual-Taught）是一种面向大模型的自提升方法，通过 EM 算法实现策略模型和奖励模型的协同进化，在无需额外人工标注的情况下即可显著提升模型性能

ThinkSwitcher：何时深入思考，何时快速决策 [Preprint]

ThinkSwitcher 设计了一种动态思维链切换框架，使推理模型能根据任务复杂度自适应调整推理深度，在保持复杂任务高准确率的同时**降低了 20% 至 30% 的计算成本**

项目经历

FuseChat: 基于成对蒸馏与参数合并的对话大模型融合 [Paper] / [Github] / [HF] / [mergekit]

- **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**
  - 项目 Github 仓库收获超过 **600stars**，SCE 算法被知名模型合并库 **mergekit (6.2k+ stars)** 收录

FuseO1: 推理大模型融合 [Blog] / [Github] / [HF]

- **项目内容**:
  - 将 FuseChat 项目中提出的 SCE 模型合并方法应用于推理大模型融合，通过合并 DeepSeek-R1-32B, QwQ-32B 和 Sky-T1-32B 得到 FuseO1-32B 模型，验证了该方法在推理领域的有效性
- **项目成果**:
  - 融合模型 FuseO1-32B 在 AIME、LiveCodeBench、GPQA 等多个主流推理评测集上超所有源模型，成为当时**最强的 32B 推理模型**，AIME24 性能超越 OpenAI o1-mini，接近 OpenAI o1
  - 模型发布 3 天内登上 HuggingFace 首页，总下载量**超 10 万次**

QwenLong-L1：推理模型长上下文强化学习 [Paper] / [Github] / [HF] / [Daily Papers]

- **项目内容**:
  - 通过强化学习算法增强模型从长上下文定位相关知识并进行多步复杂推理的能力
  - 针对长上下文强化学习中训练不稳定的问题，提出渐进式上下文扩充和难题回顾采样策略
- **项目成果**:
  - QwenLong-L1-32B 在 7 个长文档问答任务中超过 o3-mini、Qwen3-plus, 媲美 Claude-3.7-Thinking
  - 项目 Github 仓库收获约 **300stars**，被机器之心、量子位等公众号报导

发表论文

[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] 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**, main

[4] Longguang Zhong, Fanqi Wan, **Ziyi Yang**, Guosheng Liang, and Xiaojun Quan. FuseRL: Dense Preference Optimization for Heterogeneous Model Fusion. **AAAI 2026**, under review

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

[6] Guosheng Liang, Longguang Zhong, **Ziyi Yang**, Xiaojun Quan. ThinkSwitcher: When to Think Hard, When to Think Fast **EMNLP 2025**, under review

[7] Fanqi Wan, Weizhou Shen, Shengyi Liao, Yingcheng Shi, Chenliang Li, **Ziyi Yang**, Ji Zhang, Fei Huang, Jingren Zhou, Ming Yan. QwenLong-L1: Towards Long-Context Large Reasoning Models with Reinforcement Learning. Tech report