

万凡琦

手机: 18907305772

邮箱: fanqiwan@foxmail.com

地址: 广东省广州市, 510006

个人主页: <https://fanqiwan.github.io/>

教育经历

中山大学 计算机学院

计算机科学与技术专业, 硕士, 导师: 权小军教授

2022 年 9 月-至今

➤ 研究兴趣: 多源异构大模型融合、大模型推理强化、大模型高效对齐、任务型对话系统

西安交通大学 电信学部

自动化专业, 本科, 指导教师: 吕红强教授

2018 年 9 月-2022 年 6 月

➤ 学业成绩: 93/100; 排名: 10/180

论文发表

联合腾讯 AI Lab 首次提出了**多源异构大模型融合**研究问题, 旨在整合现有大模型的优势, 减少重复训练开销, 研究成果 FuseLLM 被 ICLR 2024 接收, 该成果在 Twitter 上引发广泛关注 (40w+浏览), 并被机器之心等公众号宣传报道; 基于模型融合的思想推出 FuseChat 系列对话大模型, 是当时 MT-Bench 等权威榜单上最佳 7B 模型, 该成果在 Twitter 上被著名点评人 AK 推荐, 入选 HuggingFace 每日论文精选, 并受邀请在机器之心平台开展直播宣传实测。后续作为指导者提出基于偏好学习的隐式模型融合方法, 两篇论文分别被 ICLR 2025 (WRPO), ICLR 2025 SCI-FM (FuseChat-3.0) 接收, 一篇论文 NeurIPS 2025 (FuseRL) 审稿中。FuseAI 系列开源项目在 GitHub 上获得 500+星标, 相关模型在 HuggingFace 上获得 15 万+下载, 相关算法被 GitHub 5000+星标的模型合并仓库 MergeKit 收录。目前主要关注**推理大模型的强化学习**, 联合阿里通义实验室提出长上下文推理强化学习概念, 以第一作者开展 QwenLong-L1 研究工作, 在长文档问答任务中超过 o3-mini, Qwen3-235B-A22B, 媲美 Claude-3.7-Thinking。

多源异构大模型融合

- **Weighted-Reward Preference Optimization for Implicit Model Fusion, ICLR 2025, 共同第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[HF Daily Papers\]](#) / [\[r/LocalLLaMA\]](#) / [\[魔搭社区\]](#) / [\[Blog\]](#) / [\[Website\]](#)
针对现有模型融合中词表对齐与分布映射困难, 提出加权偏好优化方法 (WRPO) 让目标模型隐式从源模型和自身之间的差异中学习。基于 WRPO 的 FuseChat-3.0-Llama-3.1-8B 在 14 个任务上平均性能提升 16.8%, 成为 AlpacaEval-2 和 Arena-Hard 榜单当时最强 8B 模型。
- **Knowledge Fusion of Large Language Models, ICLR 2024, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[Featured by elvis\]](#) / [\[Featured by AIDB\]](#) / [\[机器之心\]](#)
我们提出 FuseLLM 旨在通过知识融合创建一个统一的基座大模型, 将多个异构大模型的独特优势和能力相结合。FuseLLM-7B 在包括常识、推理、问答和代码等任务的 12 项测试基准中超过 Llama-2-7B。
- **FuseChat: Knowledge Fusion of Chat Models, EMNLP 2025, Under Review, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[Featured by AK\]](#) / [\[HF Daily Papers\]](#) / [\[机器之心\]](#)
FuseChat 是 FuseLLM 在对话大模型融合上的扩展, 旨在将多个结构和规模各异对话大模型的集体知识和各自优势整合到一个更强大的模型中。FuseChat-7B 是当时 MT-Bench 等权威榜单上最佳 7B 模型。
- **FuseRL: Dense Preference Optimization for Heterogeneous Model Fusion, NeurIPS 2025, Under Review, 共同第一作者**
[\[Paper\]](#)
FuseRL 旨在最大化隐式模型融合时不同模型回复的参与训练的利用率。其在 SFT 和 DPO 过程中分别引入来自多个源模型的回复或偏好对, 并结合奖励分数加权优化, 是当时 AlpacaEval-2 and Arena-Hard 上最强 8B 大模型。

大模型推理强化

- **QwenLong-L1: Towards Long-Context Large Reasoning Models with Reinforcement Learning, *Tech Report*, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[r/LocalLLaMA\]](#) / [\[HF Daily Papers\]](#)
首次提出长上下文推理强化学习概念，需要强化模型从长上下文定位收集相关知识再复杂推理的能力。为解决长上下文推理强化学习训练过程不稳定问题，提出渐进式上下文扩展和多阶段难题回放策略。QwenLong-L1-32B 在 7 个长文档问答任务中超过 o3-mini, Qwen3-235B-A22B, 媲美 Claude-3.7-Thinking。
- **FuseO1-Preview: System-II Reasoning Fusion of LLMs, *Tech Report*, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[r/LocalLLaMA\]](#) / [\[Mergekit Support\]](#)
FuseO1-Preview 使用 SCE 合并算法将 DeepSeek-R1-Distill-Qwen-32B, QwQ-32B-Preview, Sky-T1-32B 融合，在多个推理评测上超所有融合模型，在 AIME24 上超越 OpenAI o1-mini，接近 OpenAI o1。

大模型高效对齐

- **Let Large Language Models Find the Data to Train Themselves, *NeurIPS 2025, Under Review*, 第一作者**
[\[Paper\]](#)
首次提出利用大模型自我收集数据训练自身概念，构建 Active Data Search 智能体框架使大模型与外部环境通过 API 交互，爬取/清洗/改写/合成/标注数据用于大模型自我训练，使用强化学习优化数据收集过程，在 1,000 个内部任务以及 3 个公开任务上取得显著提升。
- **Advantage-Guided Distillation for Preference Alignment in Small Language Models, *ICLR 2025, Spotlight*, 第二作者**
[\[GitHub\]](#) / [\[Paper\]](#) / [\[Paper Weekly\]](#)
针对小模型对齐困难的问题，ADPA 引入分布级别优势函数，使用教师大模型对齐前后分布变化，为学生小模型提供细粒度分布级别奖励信号，提升小模型对齐效率。
- **Knowledge Verification to Nip Hallucination in the Bud, *EMNLP 2024*, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#)
我们提出 Knowledge Consistent Alignment，其通过自动生成的问答测试检测指令数据中的外部知识与基座大模型中的内在知识之间的不一致，发现并修正那些诱导大模型产生幻觉的训练数据，从而减轻大模型幻觉。
- **Explore-Instruct: Enhancing Domain-Specific Instruction Coverage through Active Exploration, *EMNLP 2023*, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#) / [\[Paper Weekly\]](#)
我们提出 Explore-Instruct，其利用大模型从广度和深度上自动探索领域空间，构造特定领域的层次化指令微调任务集，从而提高特定领域指令数据多样性。在三个领域中，Explore-Instruct 的表现显著优于 Self-Instruct。

任务型对话系统

- **Multi-Grained Knowledge Retrieval for End-to-End Task-Oriented Dialog, *ACL 2023*, 第一作者**
[\[GitHub\]](#) / [\[HF\]](#) / [\[Paper\]](#)
我们提出多粒度知识检索模型，并使用生成模型对检索内容的交叉注意力作为监督信号蒸馏检索模型，实现 RAG 中检索和生成模型的迭代提升，是当时小规模和大规模知识库场景下的最佳模型。

工作经历

阿里通义实验室 NLP 团队 算法实习生

Qwen-Long & Qwen-Doc 推理强化学习研究。

2025 年 3 月-至今

字节豆包 (Seed) 大语言模型团队 算法实习生

大语言模型自我提升的研究。

2024 年 6 月-2025 年 2 月

腾讯 AI Lab 自然语言处理中心 研究型实习生

大语言模型的指令微调和模型融合的研究。2023 年 3 月-2024 年 5 月

唯品会电商平台 商业项目

电商平台评论的细粒度情感分析，由王睿博士指导。2022 年 4 月-2023 年 1 月

学术竞赛

2023 年兴智杯全国人工智能创新应用大赛

亚军 深度学习模型可解释性任务2022 年 8 月-2023 年 2 月

2022 年科大讯飞 AI 开发者大赛

亚军 基于论文摘要的文本分类和查询性问答任务2022 年 7 月-2022 年 10 月

2022 年“阿里灵杰”问天引擎电商搜索算法赛

季军&技术创新奖 电商搜索（召回&精排）任务2022 年 3 月-2022 年 6 月

所获荣誉

中山大学国家奖学金2023 年 9 月-2024 年 6 月

腾讯 AI Lab 犀牛鸟专项研究项目优秀奖2022 年 9 月-2023 年 9 月

西安交通大学优秀毕业生2018 年 9 月-2022 年 6 月

西安交通大学国家奖学金2018 年 9 月-2019 年 6 月