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

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+浏览与 1800+喜欢），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。FuseLLM 和 FuseChat 系列开源项目在 GitHub 上获得 400+星标，相关模型在 HuggingFace 上获得 40000+下载。研究内容还涉及**大模型高效对齐**（幻觉缓解，垂直领域）以及**任务型对话系统**（检索增强生成）。

多源异构大模型融合

- **FuseChat: Knowledge Fusion of Chat Models**, ***NeurIPS 2024***, ***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。
- **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。

大模型高效对齐

- Self-Evolution Fine-Tuning for Policy Optimization, ***EMNLP 2024, Under Review***, **第四作者**
  [GitHub] / [Paper]
  我们提出 Self-Evolution Fine-Tuning，其通过自适应校正模型，提升模型生成的回复质量，并使用改写后数据微调模型，从而在保持对齐算法稳定性和有效性的同时，消除样本标注需求。
- **Knowledge Verification to Nip Hallucination in the Bud**, ***EMNLP 2024, Under Review***, **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

- Retrieval-Generation Alignment for End-to-End Task-Oriented Dialogue System, ***EMNLP 2023***, **第四作者**
  [GitHub] / [Paper]
- **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。

其他

- *BlockPruner: Fine-grained Pruning for Large Language Models*, ***EMNLP 2024, Under Review*, 第二作者**
  [GitHub] / [Paper]
- *PsyCoT: Psychological Questionnaire as Powerful Chain-of-Thought for Personality Detection*, ***EMNLP 2023 Findings***, **第三作者**
  [GitHub] / [Paper]
- Clustering-Aware Negative Sampling for Unsupervised Sentence Representation, ***ACL 2023 Findings***, **第二作者**
  [GitHub] / [Paper]

工作经历

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。 2023 年 6 月-至今

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，由黄昕庭博士和闭玮博士共同指导。 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

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