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，由权小军教授指导。 2022 年 9 月-至今

- 研究兴趣：任务型对话系统；指令微调；模型融合

西安交通大学 电信学部

自动化专业，本科，由吕红强教授指导。 2018 年 9 月-2022 年 6 月

- 学业成绩：93/100；排名：10/180

论文发表

我的研究兴趣集中在文本生成领域。之前，我主要研究**对话系统**。在大语言模型出现后，我的研究方向转向了**指令微调**（例如，构建特定领域大语言模型，减轻大语言模型幻觉）以及**模型融合**（例如，结合多个不同结构大语言模型的能力）。

模型融合

- **FuseChat: Knowledge Fusion of Chat Models**, ***Tech Report***, **第一作者**.
  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个人优势整合到一个更强大的模型中。FuseChat-7B 在 MT-Bench 上取得了 8.22 均分，是目前最强 7B 对话大模型。该论文在 Twitter 上被著名点评人 AK 推荐，入选 HuggingFace 每日论文精选，并受邀在机器之心公众号直播。
- **Knowledge Fusion of Large Language Models**, ***ICLR 2024***, **第一作者**.
  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各自能力相结合。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。该论文在 Twitter 上被广泛关注，获得 40w+浏览与 1000+喜欢，并被机器之心，Paper Weekly 等公众号宣传报道。
  FuseLLM 和 FuseChat 开源项目在 GitHub 获得 300+星标，相关模型在 HuggingFace 获得 20000+下载。

指令微调

- **Knowledge Verification to Nip Hallucination in the Bud**, ***ACL 2024, Under Review***, **第一作者**.
  我们提出知识一致性对齐（KCA），其通过自动生成的问答测试检测训练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**.
  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的多样性。在三个特定领域中，Explore-Instruct 的表现显著优于 Self-Instruct。

对话系统

- Retrieval-Generation Alignment for End-to-End Task-Oriented Dialogue System, ***EMNLP 2023***, **第四作者**.
  我们在检索器训练过程中引入最大边际似然，以解决端到端任务型对话系统中的检索-生成不一致问题。
- **Multi-Grained Knowledge Retrieval for End-to-End Task-Oriented Dialog**, ***ACL 2023***, **第一作者**.
  我们提出了一种多粒度知识检索器（MAKER），并为检索器训练引入了一种新颖的蒸馏目标。在 MultiWOZ 2.1 和 CamRest 测试基准上，MAKER 在小规模知识库和大规模知识库配置下都取得了 SOTA 性能。

其他

- *PsyCoT: Psychological Questionnaire as Powerful Chain-of-Thought for Personality Detection*, ***EMNLP 2023 Findings***, **第三作者**.
- Clustering-Aware Negative Sampling for Unsupervised Sentence Representation, ***ACL 2023 Findings***, **第二作者**.

工作经历

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-至今

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