

研究方向

自然语言处理与机器学习, 视觉语言跨模态研究, 文本生成

教育背景

- 2018-2023 **硕博连读**, 统计学 (机器学习).
大数据学院, 复旦大学
导师: 范剑青, 魏忠钰
- 2014-2018 **理学学士**, 数学与应用数学.
数学科学学院和大数据学院, 复旦大学

工作与实习经历

- 2023.09- **阿里通义实验室**.
视觉文本大模型 Qwen-VL
医疗领域大模型的 Agent 应用与评测
- 2019.09- **微软亚洲研究院**.
2020.03 基于 Diffusion 的文本生成, 改进 Transformer, 知识增强的生成模型
导师: 段楠, 宫叶云
- 2018.05- **香港中文大学**.
2018.07 文本生成
导师: 黄锦辉

获奖情况

- 2020.07 字节跳动奖学金 (全国 10 人)
- 2019.10 复旦大学国家奖学金
- 2014.02 中国数学奥林匹克竞赛 (CMO) 银牌

学术服务

ACL2023, ACL2024 多模态领域主席 (Area Chair)
ACL, EMNLP, NeurIPS, ICLR, ICML, IJCAI, AACL, NLPCC 的审稿人

项目与研究经历

(视觉文本) 大语言模型

- 围绕增强视觉文本大模型的 Agent 能力的主题, 全方位地收集评测数据与构建评测环境, 分析当前 VL 模型的不足, 并通过 SFT 和预训练来进行补足。
- 针对当前的 LLMs 缺乏交互式评测的问题, 我们在医学领域创建 AI Hospital, 通过使用多代理系统, 创建实时互动诊断环境。AI Hospital 中包括患者、医学检察员和主任医生, 用大语言模型来模拟不同角色。该过程包括收集病历, 进行医学检查, 以及让 LLMs 作为实习医生进行诊断和合作制定治疗计划。(Arxiv2024)

跨模态表示学习.

- 跨模态生成模型的 *exposure bias* 会导致与图片内容无关的高频 N-Gram 的出现, 我们提出先构造描述图片所需的词汇集合, 然后利用这个词汇集合作为限制进行生成。(ACL2019)
- 针对现有图片表征仅捕捉底层信息而无法建模对应的文本描述当中的高层信息的问题, 我们在场景图的基础上引入主题节点来隐式地学习高层主题概念, 并使用文本重建作为辅助任务来从文本数据当中直接学习概念表征。(IJCAI2021)
- 在图文检索任务的跨模态对比学习中, 仅在句子层级有匹配的监督信号, 这导致模型在面对仅有局部不匹配的负样例时难以做出正确决策, 因此我们通过构建短语层级的监督信号来进行模型训练。(ICMR2022)
- 在图文检索任务的表征学习当中, 使用难负样本进行对比学习可以帮助匹配模型学到更好的表示, 但是通过在数据集内部采样来获取负样本的方式限制了样本难度。我们通过句子改写来生成难负样本, 再使用它们来训练匹配模型。(NAACL2022)
- 为了将跨模态知识挖掘和跨模态预训练两个任务联合起来, 共同的提升, 我们提出了一个持续学习的框架, 并以多模态图谱作为基础来将二者联系起来并实现相互地持续地提升。(Arxiv)

文本生成.

- 针对问题生成单一的现象, 我们使用问题类型和变分自编码器来增强多样性。(IJCAI2018)
- 针对生成的问题过于机械的问题, 我们将 VQA 的数据作为负样例, VQG 作为正样例来训练分类器, 来指导生成器生成更自然流畅的问题。(COLING2018)
- 针对生成模型缺乏常识的问题, 我们将检索到的外部知识加入到预训练生成模型当中, 并通过在知识上进行筛选来减小噪声以实现更好地融合。(COLING2020)
- 我们将 SAN 和 FFN 统一到 Mask Attention Network 的框架下。在这个过程中, 我们发现它们在局部建模上存在不足, 因此提出 Dynamic MAN 来补强。我们改进的模型相较于 Transformer 在翻译任务上提高了 5%-10%。(NAACL2021)
- 我们提出了 AR-Diffusion 来更好地利用上下文的次序关系进行文本生成, 既提高了生成质量, 超过了自回归的模型, 又降低了资源消耗, 它可以比现有的扩散语言模型快 100 倍到 600 倍。(Arxiv)

论文发表 Google Scholar

- **Zhihao Fan**, Jialong Tang, Wei Chen, Siyuan Wang, Zhongyu Wei, Jun Xi, Fei Huang, Jingren Zhou. AI Hospital: Interactive Evaluation and Collaboration of LLMs as Intern Doctors for Clinical Diagnosis. (Arxiv2024)
- Tong Wu*, **Zhihao Fan***, Xiao Liu, Hai-Tao Zheng, Yeyun Gong, Yelong Shen, Jian Jiao, Juntao Li, Zhongyu Wei, Jian Guo, Nan Duan, Weizhu Chen. AR-Diffusion: Auto-Regressive Diffusion Model for Text Generation (NeurIPS2023, Equal Contribution)
- **Zhihao Fan**, Zhongyu Wei, Siyuan Wang and Xuanjing Huang. Bridging by Word: Image Grounded Vocabulary Construction for Visual Captioning. (ACL2019)
- **Zhihao Fan**, Zhongyu Wei, Siyuan Wang, Ruize Wang, Zejun Li, Haijun Shan, Xuanjing Huang. TCIC: Theme Concepts Learning Cross Language and Vision for Image Captioning. (IJCAI2021)
- **Zhihao Fan**, Zhongyu Wei, Piji Li, Yanyan Lan and Xuanjing Huang. A Question Type Driven Framework to Diversify Visual Question Generation. (IJCAI2018)
- **Zhihao Fan**, Yeyun Gong, Dayiheng Liu, Zhongyu Wei, Siyuan Wang, Jian Jiao, Nan Duan, Ruofei Zhang and Xuanjing Huang. Mask Attention Networks: Rethinking and Strengthen Transformer. (NAACL2021)
- **Zhihao Fan**, Yeyun Gong, Zhongyu Wei, Siyuan Wang, Yameng Huang, Jian Jiao, Xuanjing Huang, Nan Duan and Ruofei Zhang. An Enhanced Knowledge Injection Model for Commonsense Generation. (COLING2020)

- **Zhihao Fan**, Zhongyu Wei, Siyuan Wang, Yang Liu and Xuanjing Huang. A Reinforcement Learning Framework for Natural Question Generation using Bi-discriminators. (COLING2018)
- **Zhihao Fan**, Zhongyu Wei, Zejun Li, Siyuan Wang, Jianqing Fan. Negative Sample is Negative in Its Own Way: Tailoring Negative Sentences for Image-Text Retrieval. (NAACL2022 Findings)
- **Zhihao Fan**, Zhongyu Wei, Zejun Li, Siyuan Wang, Haijun Shan, Xuanjing Huang and Jianqing Fan. Constructing Phrase-level Semantic Labels to Form Multi-Grained Supervision for Image-Text Retrieval. (ICMR2022)
- **Zhihao Fan**, Zhongyu Wei, Jingjing Chen, Siyuan Wang, Zejun Li, Jiarong Xu, Xuanjing Huang. A Unified Continuous Learning Framework for Multi-modal Knowledge Discovery and Pre-training. (Arxiv2023)
- **Zhihao Fan**, Zhongyu Wei, Jingjing Chen, Zejun Li, Siyuan Wang, Jiarong Xu, Xuanjing Huang. Graph Interpretation for Image-Text Matching: Link Prediction on Concept-Enhanced Cross-Modal Graph via Meta-path based Transformer (Arxiv2023)
- Zejun Li, **Zhihao Fan**, Huaixiao Tou, Zhongyu Wei. MVP: Multi-Stage Vision-Language Pre-Training via Multi-Level Semantic Alignment. (ACMMM2022)
- Zejun Li, **Zhihao Fan**, Jingjing Chen, Qi Zhang, Xuanjing Huang, Zhongyu Wei. Unifying Cross-Lingual and Cross-Modal Modeling Towards Weakly Supervised Multilingual Vision-Language Pre-training. (ACL2023)
- Siyuan Wang, Zhongyu Wei, **Zhihao Fan**, Zengfeng Huang, Weijian Sun, Qi Zhang and Xuanjing Huang. PathQG: Neural Question Generation from Facts. (EMNLP 2020)
- Siyuan Wang, Zhongyu Wei, **Zhihao Fan**, Yang Liu and Xuanjing Huang. A Multi-Agent Communication Framework for Question-Worthy Phrase Extraction and Question Generation. (AAAI 2019)
- Siyuan Wang, Zhuohan Long, **Zhihao Fan**, Zhongyu Wei, Xuanjing Huang. Benchmark Self-Evolving: A Multi-Agent Framework for Dynamic LLM Evaluation. (Arxiv2024)
- Zejun Li, Ye Wang, Mengfei Du, Qingwen Liu, Binhao Wu, Jiwen Zhang, Chengxing Zhou, **Zhihao Fan**, Jie Fu, Jingjing Chen, Xuanjing Huang, Zhongyu Wei. Reform-eval: Evaluating large vision language models via unified re-formulation of task-oriented benchmarks. (Arxiv2023)