

陈家慰

<https://chen700564.github.io>

教育背景

中国科学院软件研究所, 中文信息处理实验室, 在读博士研究生

2019.08-至今

导师：韩先培研究员、孙乐研究员

电子科技大学, 信息与计算科学, 理学学士

2015.08-2019.07

GPA 3.91/4, 排名第 1

研究经历概述

- 研究方向：大模型的知识增强、代码大模型。(NeurIPS 2024, AAAI 2024)
- 此前的研究方向：弱监督学习、信息抽取。(COLING2024, ACL2023, ACL2022, EMNLP2021)
- 实习经历：百度 NLP、阿里巴巴通义实验室、字节跳动 Seed
- 发表论文：一作 7 篇, CCF-A 4 篇, CCF-B 2 篇;
- 谷歌学术引用次数: 460+

工作经历

字节跳动 (ByteDance), SEED, 实习生

2024.04-至今

- 探究代码大模型的在预训练过程中多语言能力的动态构建建立过程, 并设计一个多语言预训练数据配比的方法, 用于更好地预训练多语大模型。相关论文发表在 ICLR 2025。该研究为字节跳动豆包等大模型提供技术支持。

阿里巴巴 (Alibaba), 通义实验室, 实习生

2023.08-2024.03

- 探究和提升大模型的外部知识记忆和内部知识组织能力, 使大模型能够按需生成已被内化的知识, 从而在超越检索增强生成范式。最终构建了一个大语言模型的自检索框架, 相关论文发表在 NeurIPS 2024。该研究为检索增强生成产品通义晓蜜提供技术支持。

百度 (Baidu), 自然语言处理部门, 实习生

2022.07-2023.03

- 进行语言模型上下文学习 (in-context learning) 能力探究, 赋予语言模型更强的上下文学习能力。最终构建了一个上下文学习能力的元函数预训练算法, 相关论文发表在 ACL 2023。该研究为百度信息抽取产品提供技术支持。
- 百度文心一言数据标注。为文心一言标注任务数据、开放问答数据和人类偏好数据。

主要研究经历

代码大模型多语言能力演化过程探究

该工作探究预训练阶段代码大模型的多语言能力演化过程, 采用三种方式追溯大模型学习新语言时的内部状态: 在新语言习得时, 大模型逐渐从使用主导语言的知识系统逐渐演变为使用新语言自己的知识系统。基于上述假设和实验结果, 发现了对于很多语言, 使用主导语言的知识系统优于使用自身知识系统, 为此, 本工作提出一种不同语言预训练语料的配比方法, 使大模型达到最优的多语言系统状态。相关论文发表在 ICLR 2025。

大语言模型的自检索框架

该工作构建了一个大语言模型驱动信息检索系统——自检索 (Self-Retrieval), 包含了构建索引、检索和自评估三个过程: (1) 外部语料库首先被内化到大语言模型的参数中; (2) 利用大语言模型强大语义理解和生成能力根据查询生成相关文档; (3) 利用大模型自我评估能力, 评估生成文档的相关性并进行排序。相关论文发表在 NeurIPS 2024。

大语言模型的检索增强生成评估

该工作构建了一个大模型检索增强生成评估基准 (RGB), 它用于评估大模型检索增强生成的四种重要能力: (1) 噪声鲁棒性; (2) 拒识能力; (3) 信息聚合能力; (4) 反事实鲁棒性。实验表明, 大语言模型在这四种能力上仍然表现不足, 有时无法生成可靠的内容。相关论文发表在 **AAAI 2024**。

- **语言模型上下文学习能力的元函数预训练**

该工作提出一个命名实体识别上下文学习能力的元函数预训练算法, 它能够将命名实体识别上下文学习能力注入给语言模型。为了注入上下文学习能力, 该工作将上下文学习后获得命名实体抽取器与经过微调得到的命名实体抽取器进行对齐, 使语言模型具备元函数的能力, 能够针对不同的上下文范例转变为特定的抽取器。相关论文发表在 **ACL 2023**。

- **小样本命名实体识别的自描述预训练**

该工作针对小样本命名实体识别的有限信息挑战和知识不匹配挑战, 提出自描述机制, 自描述机制通过将实体提及映射到公共的概念空间来对齐不匹配的预训练数据与下游数据。该工作基于维基百科数据预训练了概念自描述网络, 其能够有效地进行知识迁移, 提升下游任务小样本命名实体识别的性能。相关论文发表在 **ACL 2022**。

评测竞赛

- 2022 年全国大数据与计算智能挑战赛, 高鲁棒性要求下的事件检测任务, 第二名
- 2022 年语言与智能技术竞赛, 情感可解释任务, 三等奖
- CASE@ACL2021: 政治领域细粒度事件分类评测, Zero-Shot 场景第一名, 有监督场景第二名
- 2017 年高教社杯全国大学生数学建模竞赛, 全国一等奖

论文

1. *The Rise and Down of Babel Tower: Investigating the Evolution Process of Multilingual Code Large Language Model*. **Jiawei Chen**, Wentao Chen, Jing Su, Jingjing Xu, Hongyu Lin, Mengjie Ren, Yaojie Lu, Xianpei Han, Le Sun., ICLR 2025 (第一作者)
2. *Self-Retrieval: End-to-End Information Retrieval with One Large Language Model*. Qiaoyu Tang*, **Jiawei Chen***, Zhuoqun Li, Bowen Yu, Yaojie Lu, Cheng Fu, Haiyang Yu, Hongyu Lin, Fei Huang, Ben He, Xianpei Han, Le Sun, Yongbin Li., NeurIPS 2024 (第一作者, CCF-A)
3. *Benchmarking Large Language Models in Retrieval-Augmented Generation*. **Jiawei Chen**, Hongyu Lin, Xianpei Han, Le Sun., AAAI 2024 (第一作者, CCF-A)
4. *Learning In-context Learning for Named Entity Recognition*. **Jiawei Chen**, Yaojie Lu, Hongyu Lin, Jie Lou, Wei Jia, Dai Dai, Hua Wu, Boxi Cao, Xianpei Han, Le Sun., ACL 2023 (第一作者, CCF-A)
5. *Few-shot Named Entity Recognition with Self-describing Networks*. **Jiawei Chen***, Qing Liu*, Hongyu Lin, Xianpei Han, Le Sun., ACL 2022 (第一作者, CCF-A)
6. *Few-shot Named Entity Recognition via Superposition Concept Discrimination*. **Jiawei Chen**, Hongyu Lin, Xianpei Han, Yaojie Lu, Shanshan Jiang, Bin Dong, Le Sun., LREC-COLING 2024 (第一作者, CCF-B)
7. *Honey or Poison? Solving the Trigger Curse in Few-shot Event Detection via Causal Intervention*. **Jiawei Chen**, Hongyu Lin, Xianpei Han, Le Sun., EMNLP 2021 (第一作者, CCF-B)
8. *Retentive or Forgetful? Diving into the Knowledge Memorizing Mechanism of Language Models*. Boxi Cao, Qiaoyu Tang, Hongyu Lin, Xianpei Han, **Jiawei Chen**, Tianshu Wang, Le Sun., LREC-COLING 2024 (CCF-B)
9. *Towards Scalable Automated Alignment of LLMs: A Survey*. Boxi Cao, Keming Lu, Xinyu Lu, **Jiawei Chen**, Mengjie Ren, Hao Xiang, Peilin Liu, Yaojie Lu, Ben He, Xianpei Han, Le Sun, Hongyu Lin, Bowen Yu., Under review, arXiv 2406.01252

* Equally Contribution