

# 刘柱

(+86) 186-3620-4417 · liuzhu22@mails.tsinghua.edu.cn · juniperliuzhu.netlify.app

## 教育背景

清华大学, 计算语言学方向, 在读博士研究生 2022.9 - 现在

课程: 自然语言处理, 语料库语言学 (A), 汉语词汇专题研讨 (A), 汉语词法句法专题研讨 (A), 语言类型学, 词法学

导师: 刘颖和孙茂松; **研究方向**: 计算语言学, 词义消歧, 不确定性, 可解释大语言模型

南方科技大学, 计算机科学与技术, 硕士研究生 2019.9 - 2022.6

GPA 3.67/4.0 高级算法 (A), 贝叶斯数据分析 (A), 概率视角下的机器学习 (A); 导师: 郑锋

大连理工大学, 数字媒体技术, 学士 2015.9 - 2019.6

## 论文发表

- Liu, Z., Kong, C., Liu, Y., Sun, M. (2024). Fantastic Semantics and Where to Find Them: Investigating Which Layers of Generative LLMs Reflect Lexical Semantics. arXiv preprint arXiv:2403.01509.
- Zhu Liu and Ying Liu. 2023. Ambiguity Meets Uncertainty: Investigating Uncertainty Estimation for Word Sense Disambiguation. In Findings of the Association for Computational Linguistics: *ACL 2023*, pages 3963–3977, Toronto, Canada.
- Liu, Z., Wang, T., Zhang, J., Zheng, F., Jiang, W., Lu, K. (2022). Show, Tell and Rephrase: Diverse Video Captioning via Two-Stage Progressive Training. *IEEE Transactions on Multimedia*.
- Wang, Teng, Zhu Liu, Feng Zheng, Zhichao Lu, Ran Cheng, and Ping Luo. "Semantic-Aware Pretraining for Dense Video Captioning." arXiv preprint arXiv:2204.07449 (2022).
- Ma Jian, Liu Zhu, Shi Yurong, et al. Multimodal Solar Flare Prediction Model Based on Sunspot Group Data [J]. *Spacecraft Environmental Engineering*, 2021,38(3):256-262.

## 科研经历

大语言模型的词汇语义表示 2023.10-

生成式大语言模型具有显著的通用性能。本研究着眼于该类模型的可解释方向, 重点从表征方面探究模型学习到哪些词汇语义知识, 并不断与传统的较小规模模型、双向模型进行对比分析。该研究在清华大学自然语言处理与社会人文计算实验室科研访问期间展开, 目前已经投稿至 *ACL 2024* 短文。

大语言模型对主宾互逆句成分的及物性分析 2023.11-

汉语是典型的 SVO 语言, 通常情况保持一个意义的同时, 主宾不可互逆, 但仍存在不少的语言事实是主宾可逆的。本次研究以大语言模型为参照, 考察主宾可逆情况下各个成分 (SVO) 表征的及物性, 从而为两类现象提供一个计算角度的解释。该研究成果获得“清华大学第五届语言文字学青年学术论坛”优秀论文一等奖。

词义消歧的不确定估计和分析 2022.9-2023.1

词义消歧旨在确定某个词汇在上下文中的含义。传统的词义消歧模型将这个过程视作一个分类任务: 所有可能的词义选项视作相互语义正交的候选类别, 然后从中选择一个最佳选项。但真实世界中的词汇语义是一个连续概念, 无法确定唯一的最佳释义。本次研究试图考虑词汇语义的不确定性和模糊性, 并试图寻找背后的词汇语言学因素。本次研究成果获得“清华大学第四届语言文字学青年学术论坛”优秀论文三等奖, 并被顶会 *ACL 2023* 接收。

基于变分自编码模型的视频字幕多样性生成 2021.4-2022.9

视频字幕生成 (Video Captioning) 用语言描述一段视频。多样性字幕生成旨在发掘从视频到多个描述语句的这种一对多的映射关系, 从而更加逼近自然状态下的语言分布。本次研究把这一问题建模为条件隐变量模型, 利用条件变分自编码器来刻画这种多样性。本次科研最终发表在国际期刊: *IEEE Transactions on Multimedia* 上。

空间太阳风暴事件预测 2019.9-2020.4

该项目通过太阳表面的磁图预测太阳风暴事件。我们利用了一些先进的图片分类模型和缓解样本不平衡的方法。其间团队参加了太阳耀斑的预测比赛, 并获得第一名。后续该项目继续得到军工委资助, 作为二作在中文核心期刊《航天器环境工程》中发表一篇文章。

## 竞赛获奖

---

### ActivityNet：大规模视频行为密集视频标注

2021.5-6

该比赛是计算机视觉顶级会议 CVPR 工作坊组织的比赛，我们队伍最终获得**第二名**，并受邀进行发言，之后发表一篇技术报告。

### 太阳风暴识别和预警人工智能挑战赛

2020.10

该比赛依托阿里天池，是中科院空间所发起的空间预报比赛，我们队伍最终获得**第一名 (1/70)**，成功获得相关方的 30 万科研资助，并在之后顺利结题。该项目同时申请了学校的竞赛奖学金资助。

## 学术服务

---

- 担任 **EMNLP 2023** 的审稿人。
- 2023 年春季学期担任校内《Python 与数字人文》助教。
- 2020 年担任校内《C 和 C++ 编程设计》的助教。