

# 刘柱

18636204417 · liuzhu22@mails.tsinghua.edu.cn · <https://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 - 2026.6

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**: 语言学启发的模型评估与可解释性, 计算词义理解, 语义地图模型

罗马第一大学, 自然语言处理方向, 访问学生 2025.3 - 2025.9

导师: Roberto Navigli; **研究方向**: 词义理解, 多语数据增强

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

**研究方向**: 视频字幕标注, 概率生成模型, 多模态深度学习

大连理工大学, 数字媒体技术, 学士 2015.9 - 2019.6

## 实习经历

- 在清华大学 THUNLP 组实习, 导师为孙茂松教授。 2023.09-现在
- 在罗马第一大学 NLP 组实习, 导师为 Roberto Navigli 教授。 2025.03-2025.08

## 论文发表

### 模型评估

- Zhu Liu** and Ying Liu. Ambiguity Meets Uncertainty: Investigating Uncertainty Estimation for Word Sense Disambiguation. In Findings of *ACL 2023*, pages 3963–3977, Toronto, Canada.
- Zhu Liu**, Zhen Hu, Ying Liu CoMeDi Shared Task: Models as Annotators in Lexical Semantics Disagreements. CoMeDi workshop in *COLING 2025*, Abu Dhabi, UAE.
- Zhu Liu**, Cunliang Kong, Ying Liu, and Maosong Sun. A Top-down Graph-based Tool for Modeling Classical Semantic Maps: A Crosslinguistic Case Study of Supplementary Adverbs. *NAACL 2025*, pages 4567–4576, Albuquerque, New Mexico. Association for Computational Linguistics.
- Zhu Liu**, Zhen Hu, Lei Dai, Ying Liu XISM: an eXploratory and Interactive Graph Tool to Visualize and Evaluate Semantic Map Models. arxiv:2507.04070, 2025. **Under review**.
- Zhu Liu** Evaluating Distributed Representations for Multi-Level Lexical Semantics: A Research Proposal[J]. arXiv preprint arXiv:2406.00751, 2024.

### 模型可解释性

- Zhu Liu**, Cunliang Kong, Ying Liu, and Maosong Sun. Fantastic Semantics and Where to Find Them: Investigating Which Layers of Generative LLMs Reflect Lexical Semantics. In Findings of *ACL 2024*, pages 14551–14558, Bangkok, Thailand.
- Zhu Liu**, Ying Liu, Kangyang Luo, Cunliang Kong, and Maosong Sun. From the New World of Word Embeddings: A Comparative Study of Small-World Lexico-Semantic Networks in LLMs[J]. arxiv:2502.11380, 2025. **Under review**.
- Kangyang Luo, Yuzhuo Bai, Cheng Gao, Shuzheng Si, Yingli Shen, **Zhu Liu**, Zhitong Wang, Cunliang Kong, Wenhao Li, Yufei Huang, Ye Tian, Xuantang Xiong, Lei Han, Maosong Sun. GLTW: Joint Improved Graph Transformer and LLM via Three-Word Language for Knowledge Graph Completion. *ACL 2025 Findings*.

### 多模态

- 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

## 1. 模型评估

### 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。此外，不确定性也会带来标注者之间的分歧，如何预测多样标注中的平均值和方差，是 CoMeDi@Coling 2025 的一项核心任务，本研究对该问题的探索亦被该工作坊收录。

### 基于自上而下图算法的语义图模型构建与应用

2023.10-

语义地图模型是语言类型学中研究多功能语言形式的重要工具，现有的模型构建都是在小规模的数据量下由专家手工标注的，因而难以拓展。本文开发了一种自上而下的算法，将语义地图经典的连接性假设进行拓展，从而可以在大规模数据下高效地自动构建语义地图。同时，本文还设计针对图的评价指标对生成的语义地图进行评估。该工作获得顶会 NAACL 2025 接收，并获得“清华大学第六届语言文字学青年学术论坛”优秀论文一等奖，同时该工具也在不断在线维护开发。

### 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语义不变的，例如“行人走便道”和“便道走行人”。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。

## 2. 模型可解释性

### 大语言模型的语义理解性能评估

2023.10-

生成式大语言模型具有强大的通用性能，但模型内部对于语义的理解仍需评估。本研究着眼于该类模型内部表征的可解释，重点从表征方面探究模型学习到哪些词汇语义知识，并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，合作导师为孙茂松老师，目前已被 Findings of ACL 2024 录用。作为对该工作的拓展，另一个工作也研究了模型的词库中的输入向量构成的语义网络空间，并且对比了不同规模的模型，发现它们都呈现小世界特性。该工作投稿至 EMNLP25 上。

## 3. 多模态

### 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。

## 荣誉获奖

### 清华大学第四/五/六届语言文字学青年学术论坛

2022.11

一年一度的校内博士生论文，分别获得优秀论文三等奖、一等奖、一等奖。

### 第八届上海图书馆开放数据竞赛

2023.10

本次参赛作品为“案例分析：文化遗产图像自动描述和交互聊天——以广西高山村壁画为例”，获得优秀创意论文奖。

### 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。

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 **ACL 2025**、**NAACL SRW 2025**、**CoMeDi 2024**、**EMNLP 2023** 和 **Rep4NLP workshop** 的审稿人。
- 担任 **ACL 2025** 学生研究论坛（SRW）的联合主席。
- 担任 **Coling 2025**、**NAACL 2025** 的会议组织志愿者。
- 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。