

刘亚慧

☎ +86 178 6380 7812 ✉ yahuiliu.nlp@foxmail.com

年龄: 29 岁 性别: 女 政治面貌: 中共党员

研究方向: 自然语言处理 英语: CET-6: 487 分

个人主页: <https://yahui19960717.github.io/>



教育背景

苏州大学 计算机科学与技术学院 2018.09 - 至今 硕博连读
计算机科学与技术 (NLP 实验室 HLT-LA 组, 导师: 张民教授、李正华教授)
山东农业大学 信息科学与工程学院 2014.09 - 2018.06 本科
计算机科学与技术

论文经历

- Annotation Error Detection in Painstakingly Annotated Data: Part-of-Speech Tagging as a Case Study. (一作 2025.05 ESWA SCI 一区)
Yahui Liu, Zhenghua Li, Chen Gong, Shilin Zhou, Min Zhang
论文内容: 重新标注了 CTB7E 的测试集, 为标注错误检测任务提供一个真实的错误检测数据集, 并提出在 MC Dropout 设定下基于 KL 散度的标注错误检测方法。
- MuCPAD: A Multi-Domain Chinese Predicate-Argument Dataset (一作 2022.07 NAACL CCF-B)
Yahui Liu, Haoping Yang, Chen Gong, Qingrong Xia, Zhenghua Li, Min Zhang
论文内容: 构建了跨领域语义角色标注数据集 MuCPAD, 该数据集覆盖 6 个领域、3.1 万句子、9.2 万个谓词实例。基于这个数据提供了一个 baseline, 并提出两个策略来提升 baseline 的性能。
- 一种轻量级的汉语语义角色标注规范 (一作 2020.12 中文信息学报 CCF-T1 中文期刊)
刘亚慧, 杨浩苹, 李正华, 张民
论文内容: 提出了一个面向多源多领域的中文谓词论元标注规范 (53 页)。
- Exploring the PR Tradeoff in Sparse Span Extraction Tasks: A Case Study of NER and SRL
Yahui Liu, Zhenghua Li, Min Zhang (一作 SCI 一区在投)
论文内容: 为 NER/SRL 任务构建了一个预测后处理分析框架。基于稀疏的 span 结构, 系统性地探究了四种调整策略 (增加、删除及其组合) 对性能的影响, 发现了“删易加难”的不对称现象, 为评估和理解模型在预测阶段的不确定性提供了新的视角。

项目经历

- 阿里达摩院语义分析项目 2019.03 - 2020.09
项目内容: 1) 负责撰写并优化标注规范, 定期与阿里团队沟通确认需求; 2) 承担标注人员的招聘与管理工作, 监督数据标注流程, 确保数据质量与交付进度; 3) 撰写会议纪要、双周报及月报, 记录与跟踪数据、模型方面的工作内容, 确保项目进度透明和信息同步。

竞赛经历

- CCL 2025 CFSP(汉语框架语义分析) (2/156) 苏州大学 第二名 2025.04 - 2025.08
- CCL 2024 CFSP(汉语框架语义分析) (1/79) 苏州大学 第一名 2024.03 - 2024.07
技术报告: Yahui Liu, Chen Gong, and Min Zhang. 2024. Leveraging LLMs for Chinese Frame Semantic Parsing. In Proceedings of the 23rd CCL (Volume 3: Evaluations), pages 21–31.
- CCL 2023 CFSP(汉语框架语义分析) (3/21) 苏州大学 第三名 2023.03 - 2023.08
技术报告: 刘亚慧, 李正华, 张民. 2023. CCL23-Eval 任务 3 系统报告: 苏州大学 CFSP 系统. In Proceedings of the 22nd CCL (Volume 3: Evaluations), pages 84–93.