

教育背景

◇ 本科，自动化（2018年9月-2022年6月）

清华大学，自动化系

◇ 博士生，自然语言处理（2022年9月-至今）

中国科学院，软件研究所，中文信息处理实验室

指导教师：孙乐研究员，韩先培研究员，林鸿宇副研究员，陆焱杰副研究员



一作论文 4篇 CCF-A（包含1篇 ICML spotlight），1篇 CCF-B

面向大模型推理能力开展系统性研究：认知视角能力诊断[4,5]，结构范式推理增强[1,3]，真实场景复杂应用[2,6]

1. StructRAG: Boosting Knowledge Intensive Reasoning of LLMs via Inference-time Hybrid Information Structurization.

Zhuoqun Li, Xuanang Chen, Haiyang Yu, Hongyu Lin, Yaojie Lu, Qiaoyu Tang, Fei Huang, Xianpei Han, Le Sun, Yongbin Li. (ICLR 2025)

出发点：对于知识密集型推理任务，此前基于段落块的 RAG 会引入大量文本噪声，不利于大模型进行推理。

贡献：提出结构化推理增强框架 StructRAG，首先判断出解决当前任务所需的最优结构类型，然后将原始纯文本知识自组织为该类型的结构化知识，从而仅将有效信息保留并显示地关联起来，实现系统推理性能的显著提升。

2. DeepSolution: Boosting Complex Engineering Solution Design via Tree-based Exploration and Bi-point Thinking.

Zhuoqun Li, Haiyang Yu, Xuanang Chen, Hongyu Lin, Yaojie Lu, Fei Huang, Xianpei Han, Yongbin Li, Le Sun. (ACL 2025)

出发点：针对复杂需求进行方案设计广泛存在于真实工程场景中，能否构建一个系统自动完成此任务？

贡献：提出针对真实复杂工程需求的自动方案设计系统 DeepSolution，通过双模式树进行方案探索和迭代审核，从而找到最优方案优化路径并生成可靠解决方案；此外，构建首个针对真实复杂工程场景方案设计的评测基准。

3. PaperRegister: Boosting Flexible-grained Paper Search via Hierarchical Register Indexing.

Zhuoqun Li, Xuanang Chen, Hongyu Lin, Yaojie Lu, Xianpei Han, Le Sun. (ACL 2026)

出发点：此前论文检索系统通常使用摘要构建索引，难以应对灵活粒度的检索需求，阻碍了大模型有效推理。

贡献：提出基于层次索引树的结构化论文检索系统 PaperRegister，离线时根据层次化 schema 进行内容抽取，进而构建层次化索引树；在线时根据输入查询选出对应粒度索引，进而完成准确论文检索，辅助大模型深度推理。

4. MetaphorVU: Towards Metaphorical Video Understanding.

Zhuoqun Li, Boxi Cao, Guiping Jiang, Fangrui Lv, Ruotong Pan, Jianan Wang, Xiangyu Wu, Hongyu Lin, Yaojie Lu, Yong Du, Ruyin Jia, Liyan, Tingting Gao, Han Li, Xianpei Han, Le Sun. (ICML 2026 spotlight)

出发点：此前的视频理解能力评估主要集中于物体识别和事件追踪，而对高阶认知能力的评估仍相对稀缺。

贡献：提出首个系统性的隐喻视频理解评测基准 MetaphorVU-Bench，实验显示当前多模态大模型隐喻视频理解性能较弱，核心原因是缺乏隐喻跨域映射能力；为此，构建了隐喻知识图谱，可以有效支撑大模型的跨域映射。

5. Meta-Cognitive Analysis: Evaluating Declarative and Procedural Knowledge in Datasets and Large Language Models.

Zhuoqun Li, Hongyu Lin, Yaojie Lu, Hao Xiang, Xianpei Han, Le Sun. (COLING 2024)

出发点: 对大模型推理时的知识利用进行细粒度认知分析, 从而制定更高效知识注入策略, 进而提升推理能力。
贡献: 提出基于元认知理论的知识解耦方法和基于提示注入的效用分析方法, 实验显示在大部分任务上事实性知识更重要, 只有对于简单推理问题过程性知识更重要; 此外, 增加预训练步数可显著提升大模型的知识利用能力。

6. PraMem: Practice-derived Experiential Memory for Long-horizon Behavior Prediction.

Zhuoqun Li, Boxi Cao, Jiawei Chen, Hanshu Zhou, Ruoxi Xu, Guiping Jiang, Ruotong Pan, Tingting Gao, Han Li, Xiangyu Wu, Hongyu Lin, Yaojie Lu, Xianpei Han, Le Sun. (在投)

出发点: 长序列用户行为预测是各电商平台上普遍存在的真实复杂需求, 能否基于大模型实现准确预测?
贡献: 提出练习驱动的经验记忆管理方法 PraMem, 将长历史序列从上下文负担转变为帮助大模型挖掘用户行为规律并警示自身偏差的宝贵资源, 从而在此资源上通过迭代练习动态维护经验记忆, 最终可辅助完成准确预测。

参与论文

- Self-retrieval: End-to-end Information Retrieval with One Large Language Model.

Qiaoyu Tang*, Jiawei Chen*, **Zhuoqun Li**, Bowen Yu, Yaojie Lu, Haiyang Yu, Hongyu Lin, Fei Huang, Ben He, Xianpei Han, Le Sun, Yongbin Li. (NeurIPS 2025)

- Harvesting Event Schemas from Large Language Models.

Jialong Tang, Hongyu Lin, **Zhuoqun Li**, Yaojie Lu, Xianpei Han, Le Sun. (CCKS 2023)

实习&技能

英语能力: CET-6

编程能力: Python, Bash, LaTeX 等

实习经历:

通义实验室-自然语言智能, 2024年4月-2025年2月

实习成果: 一作论文[1,2], 专利[1,2]

快手-基础大模型与应用部, 2025年9月-至今

实习成果: 一作论文[4,5], 专利[3,4]

获奖情况

中国科学院大学 三好学生 (2024)

中国科学院大学 三好学生标兵 (2025)

工程开发

参与开发国产大模型套件 PaddleNLP

核心贡献: 向量检索模型组件主要完成人

主要工作: 提交 10 次 PR, 支持向量检索模型的国产化训练、评估、以及量化加速

代码地址: https://github.com/PaddlePaddle/PaddleNLP/tree/develop/slm/pipelines/examples/contrastive_training

专利成果

1. 余海洋, **李卓群**, 黄非, 李永彬. 一种基于多类型结构化知识的检索增强智能体方案. (专利局已受理)
2. 唐乔裕, **李卓群**, 余海洋, 黄非, 李永彬. 一种新的仅由大模型驱动的检索增强生成系统. (专利局已受理)
3. 吴翔宇, **李卓群**, 曹博希, 林鸿宇, 陆焱杰, 韩先培, 孙乐. 一种基于隐喻知识图谱的隐喻视频理解增强方法. (专利局已受理)
4. 吴翔宇, **李卓群**, 陈嘉伟, 曹博希, 林鸿宇, 陆焱杰, 韩先培, 孙乐. 一种基于经验记忆的长序列行为预测增强方法. (专利局已受理)
5. 孙乐, **李卓群**, 林鸿宇, 王天舒, 曹博希, 陆焱杰, 周炜翔, 韩先培. 基于任务指示的参考知识链接任务通用解决方法及装置. CN118395972A
6. 林鸿宇, 陆焱杰, 唐家龙, **李卓群**, 韩先培, 孙乐. 一种基于大规模语言模型的事件模板归纳方法和系统. CN116450827A
7. **李卓群**, 陶焜. 基于文本匹配和意图识别融合处理的智能对话方法及装置. CN115238050B

个人主页

主页: <https://li-z-q.github.io>

GitHub: <https://github.com/Li-Z-Q>

Google Scholar: <https://scholar.google.com/citations?user=IrSf23QAAAAJ>