廖桓萱

■ huanxuanliao@gmail.com · • 15575986157 · • Xnhyacinth · • https://xnhyacinth.github.io/

☎ 教育背景

中国科学院大学自动化研究所, 北京

2023 - 至今

在读硕士研究生模式识别,自动化研究所模式识别国家重点实验室,导师何世柱 and 刘康

华北电力大学, 北京

2019 - 2023

学士智能科学与技术(创新), 1/13

👺 实习/项目经历

蚂蚁集团北京

2024年10月-2025年3月

实习 研究型实习生 (超长上下文)

基于全局和局部的提示词压缩

- 实现了基于 Qformer 和 MLP 的动态路由 soft token 全局压缩和基于分类的 hard token 局部压缩 长上下文建模综述
 - 基于自己的长上下文建模仓库完成综述撰写

KV Cache Channel 压缩

• 对 token 进行动态 channel 裁剪 (topk 和 topp) 而不是只针对长度压缩,同时考虑恢复裁剪的 channel

♀ 研究兴趣

- 大模型/多模态大模型
- 长上下文建模
- 推理, 问答, 压缩, KV Cache

■ 主要文章

(* stands for equal contribution; Listed in reverse chronological order.)

- [12] A Comprehensive Survey on Long Context Language Modeling. (Preprint, 2025)
- [11] <u>Huanxuan Liao</u>, Shizhu He, Yupu Hao, Jun Zhao, Kang Liu. DATA: Decomposed Attention-based Task Adaptation for Rehearsal-Free Continual Learning. (Preprint, 2025)
- [10] Wangtao Sun, Haotian Xu, <u>Huanxuan Liao</u>, Xuanqing Yu, Zhongtao Jiang, Shizhu He, Jun Zhao, Kang Liu. VaiBot: Shuttle Between the Instructions and Parameters of Large Language Models. (Preprint, 2025)
- [9] Xiangnnan Wu, Shizhu He, <u>Huanxuan Liao</u>, Jun Zhao, Kang Liu. Differentiated Instruction for Student Language Models: Contrastive Error Self-Correction via Refinement Feedback of Large Language Models. (Preprint, 2025)
- [8] Yupu Hao, Pengfei Cao, Zhuoran Jin, <u>Huanxuan Liao</u>, Yubo Chen, Kang Liu, Jun Zhao. Evaluating Personalized Tool-Augmented LLMs from the Perspectives of Personalization and Proactivity. (Preprint, 2025)
- [7] <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Kang Liu, Jun Zhao. Neural-Symbolic Collaborative Distillation: Advancing Small Language Models for Complex Reasoning Tasks. (AAAI 2025, CCF-A)
- [6] Yupu Hao, Pengfei Cao, Zhuoran Jin, <u>Huanxuan Liao</u>, Yubo Chen, Kang Liu, Jun Zhao. CITI: Enhancing Tool Utilizing Ability in Large Language Models without Sacrificing General Performance. (AAAI 2025, CCF-A)
- [5] <u>Huanxuan Liao</u>, Shizhu He, Yupu Hao, Xiang Li, Yuanzhe Zhang, Kang Liu, Jun Zhao. SKIntern: Internalizing Symbolic Knowledge for Distilling Better CoT Capabilities into Small Language Models. (COLING 2025, CCF-B)

- [4] <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Kang Liu, Shengping Liu, Jun Zhao. Awakening Augmented Generation: Learning to Awaken Internal Knowledge of Large Language Models for Question Answering. (COLING 2025, CCF-B)
- [3] <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Yanchao Hao, Shengping Liu, Kang Liu, Jun Zhao. From Instance Training to Instruction Learning: Task Adapters Generation from Instructions. (NeurIPS 2024, CCF-A)
- [2] Yixuan Weng, Zhiqi Wang, <u>Huanxuan Liao</u>, Shizhu He, Shengping Liu, Kang Liu, Jun Zhao. LMTuner: An user-friendly and highly-integrable Training Framework for fine-tuning Large Language Models. (Preprint, 2023)
- [1] <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Kang Liu, Jun Zhao. Dynamic Weighted Neural Bellman-Ford Network for Knowledge Graph Reasoning. (CCKS 2023, Springer)

♡ 获奖情况

奖学金 北京优秀毕业生 2023年6月 国家奖学金 2022年12月 国家奖学金 2021年12月 四方社会奖学金 2020年12月 竞赛 国家三等奖,信息安全竞赛 2022年8月 国家级优秀,全国大学生创新创业项目 2021年12月 北京市三等奖, 互联网+ 2021年8月

◇ 专业服务

会议审稿: ACL ARR 2024, 2025.

开源社区: Awesome-LLM-Long-Context-Modeling 1.3k+ Stars.