

袁晓薇

模识实验室，导师：刘康
中国科学院自动化所，北京市海淀区中关村东路 95 号智能化大厦 100190
性别：女 出生年月：1997/10/27
Tel: 86-15010527558 E-mail: yuanxiaowei2022@ia.ac.cn

教育背景

中国科学院自动化所 在读博士	09/2022 – 至今
中国科学院信息工程研究所 硕士（推免） GPA: 3.8/4.0	09/2019 – 06/2022
中央财经大学 本科 加权平均分: 92.07/100	09/2015 – 06/2019

科研工作

RAG 系统的重排优化

基于风险最小化的 zero-shot LLM 重排改进

《Improving Zero-shot LLM Re-Ranker with Risk Minimization》EMNLP 2024 一作

- <https://arxiv.org/abs/2406.13331>
- 当前主流的无监督重排方法 UPR 的核心思想是将 LLM 映射为查询似然模型 QLM 中的文档模型，然而其中存在着有偏估计问题；针对这一偏差，本文基于贝叶斯的最小风险决策提出了一种优化的重排方法UR³，旨在量化并缓解估计的误差；UR³将误差最小化问题转化为最大化的文档生成概率，从而在风险最小化目标下同步查询和文档生成概率的优化。
- 我们的实证结果表明，UR³显著提高了重排结果，特别是提高了于 Top-1 的准确性。它通过使用更少的输入文档实现更高的准确性，更有利于下游的 QA 任务性能提升。

语言模型的训练与推理

面向知识冲突的自适应解码方法

《Discerning and Resolving Knowledge Conflicts through Adaptive Decoding with Contextual Information-Entropy Constraint》ACL Findings 2024 一作

- <https://arxiv.org/abs/2402.11893>
- 针对知识冲突问题：当外部知识与模型参数知识不一致时，由于大模型对内部知识的依赖，可能无法利用外部相冲突的知识；本文提出了一种面向冲突的自适应的解码方法
- 针对当前方法无法有效识别知识冲突是否发生的问题，提出基于上下文的信息熵约束的冲突识别方法；和自适应的解码策略：对于冲突 Token，在解码生成中需要优先考虑外部的知识；对于不冲突的 Token，则需兼顾外部知识和模型内部知识
- 在下游的 QA 任务中：对于冲突数据和其他解码方法相比，有一致提升；对于非冲突数据，保持了稳定的精度

语言与知识图谱的对比预训练方法

《Contrastive Language-knowledge Graph Pre-training》ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP 2024) 一作

- <https://dl.acm.org/doi/10.1145/3644820>
- 针对知识图谱与语言模型的高效协同训练问题，本文介绍了一个简单而有效的知识增强模型，即 COLLEGE (对比的语言-KG 预训练模型)，该模型利用对比学习将事实知识融合进语言模型。这种方法保持了 KG 的原始图结构，以最大程度的提供可用信息；并且避免了异构知识的嵌入融合的问题
- 实验结果表明，在知识密集型的理解任务和知识探测任务上，COLLEGE 方法可以优于先前的 SOTA 结果

获奖情况

2021 年获“2020-2021 年度中国科学院大学三好学生”	05/2021
2018 年获“2017-2018 学年本科生国家奖学金”	10/2018
2018 年获“2017-2018 学年中央财经大学校级三好学生”	09/2018

专业技能

- 熟悉 Python 语言，有 PyTorch 深度学习框架与语言模型训练的实验经验
- 熟悉主流大语言模型如 LLaMA 等模型的推理机制及 RAG 的应用
- 语言技能：英语（六级 568）