



RAG & CausalRAG

论文调研

2025年8月13日

马斌

目录

1

CausalRAG论文调研

2

RAG技术调研

3

总结

4

引用目录



1 CausalRAG论文调研 --- 概况

■ 调研目录:

- [1] CausalRAG(会议: Findings-2025.7) --- 用LLM识别和估计因果路径, 构建精炼因果图并生成因果摘要
- [2] ADviaRAG --- RAG助力构建AD生物标志物因果网络, 为每个节点生成潜在原因, 结合Chain-of-Thought增强推理
- [3] LACR --- LLM辅助因果恢复





1 CausalRAG论文调研 --- [1] CausalRAG

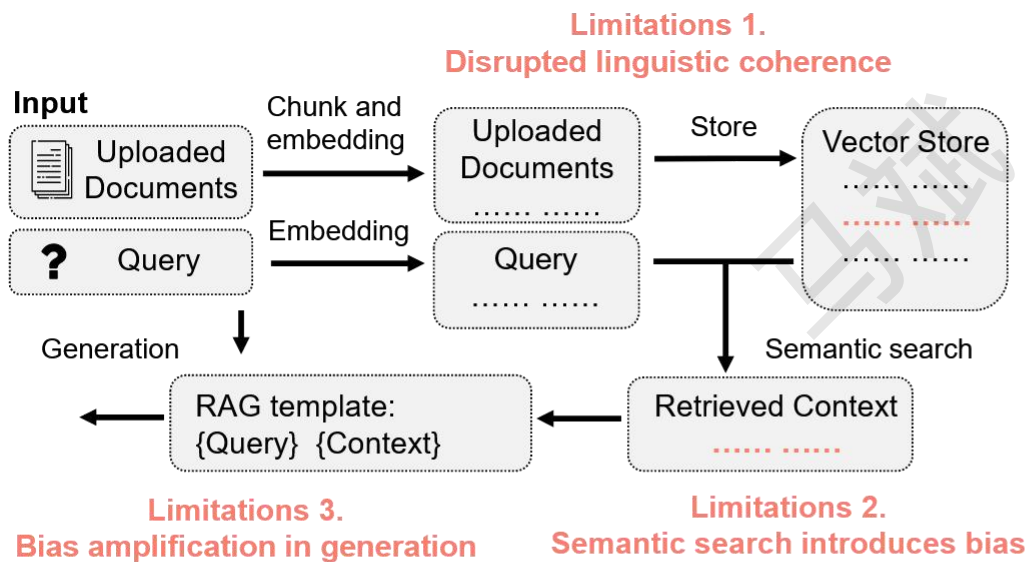
■ 发现:

□ 常规RAG的三个问题:

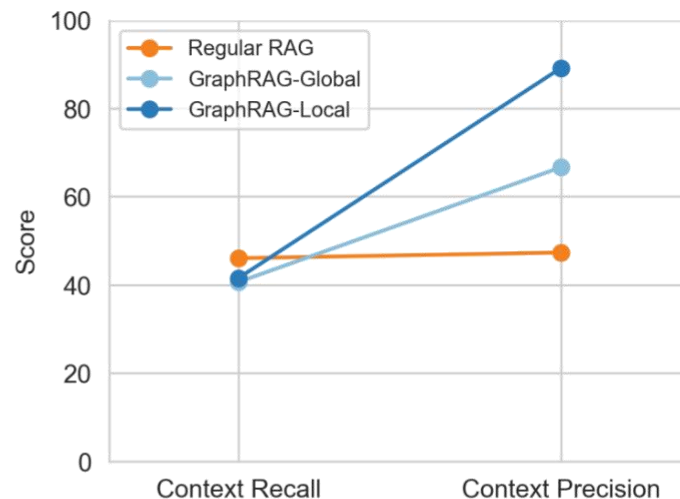
1. 语言连贯性被破坏
2. 生成中的偏差放大
3. 语义搜索引入偏差

□ 常规RAG --- 召回率高, 准确度低

GraphRAG --- 召回率低, 准确度高



(a) Analysis of Limitations in regular RAG System



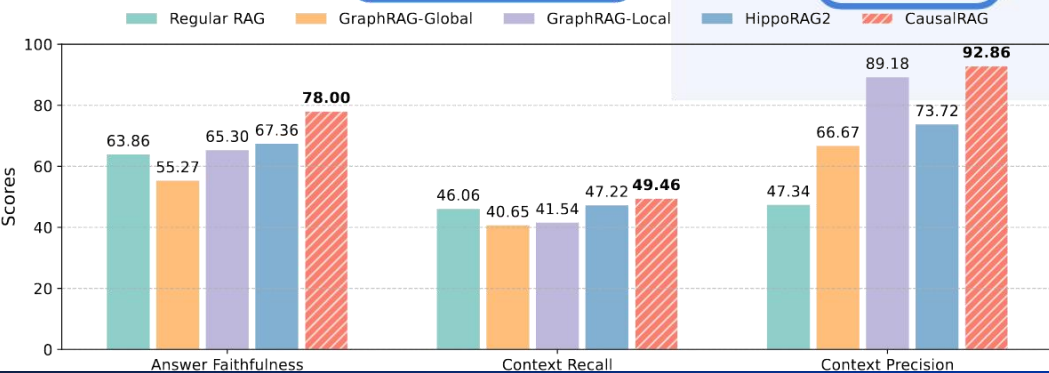
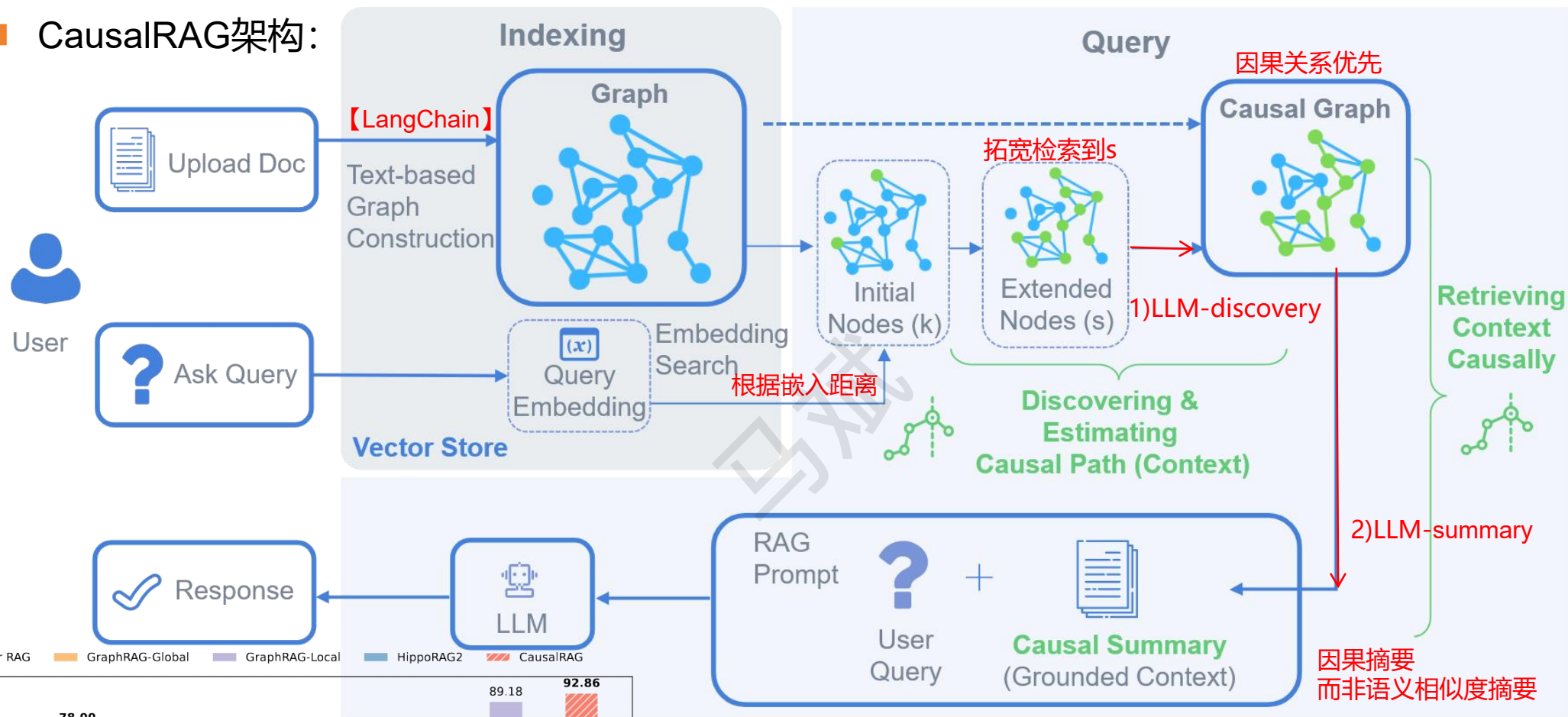
(b) Experimental study on context recall and precision

Figure 1: Analytical and experimental studies reveal limitations in regular RAG and GraphRAG. (a) identifies three key retrieval and generation issues in regular RAG; (b) evaluates RAG via context precision and recall, showing regular RAG excels in recall but lacks precision. GraphRAG improves precision but trades off some recall.



1 CausalRAG论文调研 --- [1] CausalRAG

CausalRAG架构:





2 RAG技术调研



【根据[1] CausalRAG 的调研】

- 最近的研究集中在沿两个主要方向改进 RAG：
 - 通过设计更具适应性和动态性的检索框架来提高效率和集成机制
(Gan et al., 2024; Ravuru et al., 2024; Zhang et al., 2024a);
 - 改进外部知识的表示以促进检索和推理，GraphRAG是一种主要方法
(Edge et al., 2024; Guo et al., 2024; Potts, 2024)。
- 尽管取得了这些进步，但现有的 RAG 架构仍然面临影响检索质量和响应准确性的关键限制，这主要是由于三个问题：
 - 文本分块设计导致的上下文完整性中断;
 - 依赖语义相似性而不是因果相关性进行检索;
 - 在选择真正相关的文件方面缺乏准确性。



2 RAG技术调研



【根据[4] GraphRAGsurvey 的调研】

■ 现存主要挑战：

- 自适应LLM提示词：考虑引入动态优化提示词的循环
- 增强LLM对图问题的理解：仅依赖序列标记，一个序列有过多节点时限制性能。要专门的非线性算法组件
- RAG动态图：更新的计算开销大
- 跨领域 RAG 的多模态图：处理多模态的输入数据
- RAG可缩放图形检索：优化索引或分区，如超图或使用GNN
- 改进图构建技术：分层图、GNN
- GraphRAG的可解释性：节点与边的可视化
- 整合用户交互：纳入用户反馈循环
- 鲁棒性和偏差缓解：考虑开发图偏差感知模块



3 总结



- 本次调研中发现:
- CausalRAG已经有相关研究, 已大大解决:
 - 图可解释性问题 --- 因果摘要
 - 文件选择的逻辑性 --- 检索中因果关系优先
- 现存问题:
 - 图构建仍为静态 --- 可考虑整合现有动态图构建方法,
如[4]中提到的 DynaGRAG [176], RAG4DyG [197], DRAGIN [161], Multi-Armed Bandit Enhanced RAG [173]
 - 图构建过程开销大 --- 引入优化算法组件
 - 检索过程仍兼用了传统的语义相关性检索 --- 可考虑替换为或分层兼用GNN策略



4 引用目录



- [1] N. Wang, X. Han, J. Singh, J. Ma, and V. Chaudhary, “CausalRAG: Integrating Causal Graphs into Retrieval-Augmented Generation,” in Findings of the Association for Computational Linguistics: ACL 2025, W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar, Eds., Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 22680–22693. doi: 10.18653/v1/2025.findings-acl.1165.
- [2] X. Zhou et al., “Accelerating Causal Network Discovery of Alzheimer Disease Biomarkers via Scientific Literature-based Retrieval Augmented Generation,” June 12, 2025, arXiv: arXiv:2504.08768. doi: 10.48550/arXiv.2504.08768.
- [3] Y. Zhang, Y. Zhang, Y. Gan, L. Yao, and C. Wang, “Causal Graph Discovery with Retrieval-Augmented Generation based Large Language Models,” June 18, 2024, arXiv: arXiv:2402.15301. doi: 10.48550/arXiv.2402.15301.
- [4] Z. Zhu, T. Huang, K. Wang, J. Ye, X. Chen, and S. Luo, “Graph-based Approaches and Functionalities in Retrieval-Augmented Generation: A Comprehensive Survey,” Apr. 08, 2025, arXiv: arXiv:2504.10499. doi: 10.48550/arXiv.2504.10499.



谢谢！

2025年8月13日

马斌

