Incorporate Symbolic Reasoning on Multi-hop Retrieval Augmented Generation

Nguyễn Như Giáp
Hanoi University of Science and Technology
baluanhugiap@example.com

***Abstract*—Multi-hop Retrieval-Augmented Generation (RAG) tasks challenge language models to reason over multiple pieces of evidence. However, conventional RAG systems often suffer from unreliable reasoning due to their purely text-based retrieval and generation pipeline. In this work, we propose a method to incorporate reliable reasoning into multi-hop RAG by integrating Knowledge Graphs (KGs) into the retrieval process. Specifically, we leverage structured relationships from KGs to guide the retrieval of more semantically relevant documents and support interpretable, step-by-step reasoning paths. Our approach improves factual consistency and response accuracy in complex question answering. Experimental results on multi-hop QA benchmarks demonstrate the effectiveness of our KG-enhanced RAG framework compared to traditional RAG methods.**

***Index Terms*—Multi-hop Question Answering, Retrieval-Augmented Generation, Knowledge Graph, Symbolic Reasoning, Natural Language Processing, Reliable Inference, Question Decomposition**

I. Introduction

Large Language Models (LLMs) have demonstrated strong capabilities in generating natural language responses across various tasks. However, they often suffer from major limitations, such as hallucinating facts and relying on outdated knowledge. Retrieval-Augmented Generation (RAG) was introduced to overcome these issues by retrieving external documents during inference time to ground the language model's outputs with up-to-date and relevant information.

While effective, most existing RAG systems adopt a single-step retrieval process, assuming that all necessary information can be gathered in one call to a retrieval system. This assumption breaks down in multi-hop question answering tasks, where reasoning must span multiple, scattered facts from different sources.

Multi-hop retrieval refers to a process in which the system performs sequential retrieval steps—using intermediate results to inform the next search—to incrementally gather evidence for complex queries. This multi-step mechanism is crucial when the required information is not co-located in a single document but rather distributed across several sources.

Such complex reasoning tasks demand more than shallow retrieval: they require chaining multiple pieces of evidence and integrating them to form an accurate and consistent answer. Unfortunately, traditional RAG systems lack the structure to support such compositional reasoning, often leading to incomplete or incorrect answers.

To address this limitation, our work proposes a novel method that incorporates structured symbolic reasoning into the RAG pipeline using Knowledge Graphs. By guiding the retrieval and reasoning process with structured entity relationships, our approach enables more reliable inference and improves performance on multi-hop question answering tasks.

II. Related Work

A. Chain-of-Retrieval

Chain-of-Retrieval (CoRAG) enhances standard RAG by supporting multiple, sequential retrieval steps before generating a final answer. Each retrieval builds upon intermediate sub-queries and sub-answers, which the model dynamically formulates. Training involves rejection sampling to craft effective retrieval chains, and at inference, strategies such as greedy decoding, best-of-N, or Monte Carlo Tree Search control chain length and quality. CoRAG demonstrates significant performance gains—even exceeding 10 EM points on multi-hop QA datasets and showing strong generalization across benchmarks like KILT. However, CoRAG remains dependent solely on natural-language inference over text: errors in early retrieval or reasoning steps can cascade, and models may struggle to resolve reasoning conflicts arising from noisy or incomplete text sources.

B. Interleaving Retrieval and Chain-of-Thought (IRCoT)

IRCoT interleaves retrieval calls with chain-of-thought (CoT) reasoning steps. At each CoT sentence, the model retrieves additional evidence to inform the next reasoning step, enabling dynamic, context-aware information gathering. Trivedi et al. apply IRCoT with GPT-3 and Flan-T5, showing up to +21% retrieval recall and +15% QA gains on datasets like HotpotQA, MuSiQue, and IIRC. The approach also reduces factual errors in reasoning by 40–50%, compared to baselines. Yet, similar to CoRAG, IRCoT relies exclusively on raw-text reasoning, making it fragile in the presence of noisy or ambiguous documents. The lack of structured symbolic guidance may lead to misinterpretation or hallucination in multi-hop inference.

C. Limitations and Motivation for Symbolic Integration

Although CoRAG and IRCoT have advanced multi-hop RAG, they both depend entirely on natural-language inference. This reliance exposes them to two major issues:

- **Error propagation**: Mistakes or uncertainties in early retrieval or CoT steps can compound, derailing the reasoning chain.

- **Noisy text data**: Raw text is unstructured and inconsistent, making it difficult for LMs to perform reliable, compositional reasoning over multiple documents.

These limitations prompt the question: *Could symbolic reasoning tools, such as Knowledge Graphs, enhance reliability and interpretability in multi-hop RAG?* Our work explores this direction by integrating Knowledge Graphs to guide both retrieval and step-wise inference.

III. Proposed Method

We propose a method that enhances the reliability of reasoning in Multi-hop Retrieval-Augmented Generation (RAG) by integrating Knowledge Graphs (KGs) into the retrieval and reasoning pipeline. Our approach is structured into three main stages: (1) query decomposition, (2) sub-query reasoning with knowledge graphs, and (3) final answer aggregation.

A. Query Decomposition

To handle complex multi-hop questions, we first decompose the original query into several simpler sub-queries. Each sub-query targets a specific aspect of the original question, allowing for more precise retrieval and reasoning.

For instance, given a query such as:

"Which online betting platform provides a welcome bonus of up to $1000 in bonus bets for new customers' first losses, runs NBA betting promotions, and is anticipated to extend the same sign-up offer to new users in Vermont, as reported by both CBSSports.com and Sporting News?"

We decompose it into sub-queries like:

- Q1: What online betting platforms offer a welcome bonus of up to $1000 for new customers?
- Q2: Which platforms run NBA betting promotions?
- Q3: Are there platforms expected to extend the offer to new users in Vermont as reported by CBSSports.com and Sporting News?

B. Sub-query Reasoning with Knowledge Graphs

Each sub-query is used to retrieve relevant documents via vector-based similarity search. From these retrieved documents, we construct a local Knowledge Graph by extracting entities and their relationships. This graph offers a structured representation of the textual content and enables symbolic reasoning.

By reasoning over the KG, we can more accurately infer intermediate answers for each sub-query, reducing reliance on noisy, unstructured text. This step also mitigates the risk of error propagation by enabling interpretable reasoning paths.

C. Final Answer Aggregation

Once the answers to all sub-queries are obtained, we aggregate the (sub-question, answer) pairs and use them as additional context alongside the original retrieved documents. The combined evidence—both in textual and graph form—is used to answer the original complex query.

This final reasoning step incorporates both symbolic structure (from the KG) and semantic content (from raw text), leading to more accurate and consistent responses. The proposed pipeline is designed to be modular, allowing future integration with other multi-hop RAG techniques like CoRAG or IRCoT.

IV. Experimental Evaluation

A. Dataset

We evaluate our proposed method on the MultiHop-RAG dataset [**?**], a benchmark designed to assess retrieval-augmented generation systems on complex multi-hop queries. The dataset contains 2,556 questions, each requiring evidence aggregation from two to four documents. It also includes annotated gold answers and supporting documents for objective evaluation of both retrieval and reasoning components.

B. Case Study: Reasoning Quality

To assess the reasoning capability of our method, we conduct a case study comparing our Knowledge Graph-enhanced approach with a strong Chain-of-Retrieval baseline. For example, consider the following complex query:

> *"Which online betting platform provides a welcome bonus of up to $1000 in bonus bets for new customers' first losses, runs NBA betting promotions, and is anticipated to extend the same sign-up offer to new users in Vermont, as reported by both CBSSports.com and Sporting News?"*

Chain-of-Retrieval Result: *DraftKings Sportsbook* — Incorrect

Our Method: *Caesars Sportsbook* — Correct

The supporting evidence reveals that although DraftKings is a well-known betting provider, it does not match all the criteria outlined in the query. Our method, leveraging KG-based reasoning, accurately links multiple conditions across retrieved documents and identifies Caesars Sportsbook as the correct answer.

C. Reasoning Reliability and Noise Robustness

We observe that traditional embedding-based retrieval methods struggle when entity overlap introduces partial similarities. For example, the term "bet" may appear across multiple documents, misleading the model toward superficially similar but contextually irrelevant content. Our approach mitigates this by explicitly modeling relationships between entities using Knowledge Graphs, allowing the system to distinguish relevant evidence and suppress noise.

D. Observations

- Knowledge Graphs help disambiguate similar terms by introducing structured links (e.g., "welcome bonus" $\rightarrow$ "$1000" $\rightarrow$ "first loss" $\rightarrow$ "Vermont").
- KG-based reasoning allows for better handling of long-range dependencies between facts spread across multiple sources.
- The incorporation of structured intermediate answers reduces the likelihood of hallucinations, improving factual accuracy.

V. Conclusion and Future Work

This report presented a method to improve the reliability of reasoning in multi-hop Retrieval-Augmented Generation (RAG) by incorporating Knowledge Graphs into the retrieval and reasoning process. The proposed approach involves decomposing complex queries into sub-questions, retrieving relevant documents, constructing Knowledge Graphs from these documents, and using symbolic reasoning to derive intermediate answers. These answers are then aggregated to form a complete and accurate response to the original question.

Through case studies and qualitative evaluation on the MultiHop-RAG dataset, we observed that our method offers improvements in response accuracy and robustness, especially in scenarios where relevant information is distributed across multiple documents. By using structured knowledge representations, the model is better equipped to handle ambiguity and reduce the propagation of errors during reasoning.

Future Work. In future work, we plan to expand this approach by integrating Knowledge Graphs into other advanced multi-hop RAG techniques such as IRCoT. We also aim to explore graph-based filtering techniques that can help identify and use only the most relevant information from large retrieval results, improving the efficiency and interpretability of the system.

References