UROP Report1

Jiaye Liu

Student ID: 76613365

2024-12-09

**Project Overview**

Large Language Models (LLMs) are powerful tools for generating and understanding text, but they can sometimes provide inaccurate or made-up answers, known as "hallucinations." This happens because LLMs rely on patterns in their training data rather than verified facts, it does not understand language but return based on probability. Knowledge Graphs (KGs) store factual information in a structured way, using entities and relationships. By integrating KGs with LLMs, we aim to **improve the accuracy and reliability of AI-generated responses by grounding the generated responses in the factual and strutured data** in the KG.

This project investigates whether using KGs to support LLMs through Retrieval-Augmented Generation (RAG) can enhance the LLM’s accuracy. The ultimate goal is to develop a system where users can upload textbooks and receive cheat sheets that summarize key information using both the LLM and the KG.

**What Has Been Done So Far**

1. I **experimented with converting unstructured data into a Knowledge Graph** using an Neo4j LLM Knowledge Graph Builder. The **results were satisfactory for liberal arts subjects, but less effective for math**, which is challenging due to the complex formulas and logical structures. It is also a big challenge is to create and maintain high-quality KGs, as they require expert input and substantial manual effort, which is a crucial reason of limiting their widespread use. KG will be stored once the UROP funding is secured, which will be used to purchase the neo4j plan.
2. I **find some probability questions for later testing**.
3. I have **developed a starter** for this project. I used **Langchain to integrated a KG with an LLM to help guide its responses through RAG**. The idea is that the LLM will first retrieve relevant facts from the KG, and then use this information to generate a more accurate response. In it, prompts are set, and with some few-shot examples. The code will be executed once the UROP funding is secured, which will be used to purchase the OpenAI API.

**Next Steps**

1. For math related subjects, attempt to **process data in a more structured format**, such as LaTeX, or explore using an existing KG. If both does not work, consider developing a extractor from the fundamental, like using spacy.
2. **Testing** the LLM with KG (using RAG) and compare it to a standard LLM chat model without any KG support. Evaluating the accuracy and relevance of the responses generated by both systems, focusing on identifying measurable improvements brought by the KG.
3. **Generalizing** Beyond Specific Subjects. Currently, the tests focus on probability-related questions. The next step is to generalize the system to handle a broader range of subjects, where the user input, KG, and LLM extraction can adapt to various domains. This will make the system more flexible and capable of addressing different areas of knowledge.
4. **Refining the system to reduce computational costs.** Specifically, I consider removing the use of KG memory and session IDs, which are not essential for the current code, but added for further development. These features are necessary for conversational AI or dealing with follow-up questions, which is common is probability test.
5. **Exploring the most recent Retrieval-Augmented Generation (RAG) approaches** to see if any new techniques or optimizations can improve our current model. This research could offer insights.

**Conclusion**  
 This project demonstrates the potential benefits of integrating Knowledge Graphs with Large Language Models to improve the accuracy of AI-generated text. Although there are challenges in building high-quality KGs, they are still ways to handle it. The next steps will focus on refining the KG extraction process, testing the hybrid system, future refining, etc. Ultimately, the goal is to improve accuracy of generated text and create a robust system that help students and professors with their education.

**Appendix: Starter of Integrating LLM With KG Using Langchain**

<https://github.com/Jiaye2027/KG/blob/main/kg-as-rag-in-llm.ipynb>

