Langchain - Task 8

(Auditors - Sameera Sudarshan Kashyap and Pavithra Selvaraj) | GitHub

Initially read and understood about Langchain and setup the development environment by installing necessary packages:

```
Python
pip install langchain-community
pip install langchain-openai
pip install chromadb
```

Migrating to Langchain

We migrated all model queries by using Langchains abstraction over models, we used the `HumanMessage and SystemMessage` helpers to encapsulate the prompts.

```
Python
from langchain_core.messages import HumanMessage, SystemMessage
```

We then declared the OpenAI model using the ChatOpenAI wrapper from Langchain.

```
Python
# example

client = ChatOpenAI(model="gpt-4o", api_key=api_key)
chat_messages = [
        SystemMessage(
            content="You are an expert in solving constraint satisfaction
problems. Your task is to analyze the given problem and provide a structured
solution."
        ),
        HumanMessage(
            content=f"Solve this constraint satisfaction problem by identifying
the correct option and give the output {nl_problem_goal}"
        ),
    ]

    csp_response = client.invoke(chat_messages).content
```

Inference
We used the KB created for task 4 as the base to test the Langchain and RAG implementation which uses OpenAI model and vector-based retrieval to improvise the reasoning ability. We constructed a vector database and stored it in the popular vectordb, ChromaDB. By using text file (prolog_kb_context.txt) containing the Prolog knowledge base as a source, this included:

- **Document Loading**: Used TextLoader from langchain.document_loaders to read and load the knowledge base file.
- **Chunking**: Applied RecursiveCharacterTextSplitter to divide the document into manageable text chunks for embedding.
- **Embedding and Storage**: Used OpenAIEmbeddings() to embed the text chunks and stored them using the Chroma vector store.

Experimentation:
Problem 1:
Monica, Rachel and Phoebe are Female
Ross, Chandler and Joey are male
Ross, Rachel, Monica, Chandler and Joey are all friends
Is the following statement true, false or uncertain?
monica is married to chandler

Answer: True
The answer was correctly inferred because the information was present in the prolog_kb_context.txt file

Problem 2:
Monica, Rachel and Phoebe are Female
Ross, Chandler and Joey are male
Ross, Rachel, Monica, Chandler and Joey are all friends

Is the following statement true, false or uncertain?
Amy is rachel's sister

Initial result: The system was not able to infer the answer even after multiple refinements

Final result: We added the details about Amy into the prolog_kb_context.txt file. After this the system was able to infer the result once the embeddings were regenerated with the additional information. With this we observed that Langchain+RAG setup performs well in inferring the results related to the problem statement where additional context seems helpful in symbolic reasoning

Challenges:
Due to an issue with package versioning, the pyswip library was failing in initialization, this caused confusion and led to hours of debugging, but eventually after reinstalling the latest version of pyswip, we were able to obtain the correct outputs.