Team 9 Project Proposal

Explainable Natural Language Query Interface for Relational Databases Using a Multi-Agent System

AIM

Build intelligent database exploration tools that are both powerful and explainable across diverse database schemas, allow users input natural language queires and receive relevant tables with highlighted records.

BACKGROUND & VALUE PROPOSITION (2 marks)

Understanding the data & the current data interpretation practice

Consist of 200 database with multiple tables, 10,181 questions, and 5,693 corresponding complex SQL queries. On average, each database in Spider has 28 columns and 9 foreign keys. The average question lenghth and SQL length are about 13 and 21 respectively.

Spider contains both databases with multiple tables in different domains and complex SQL queries. It tests the ability of a system to generalise to not only new SQL queries and database schemas, but also new domains.

The task DO NOT evaluate model performance on generating values, predicting correct SQL structures and columns is more realistic and critical. Evaluation include Componenet Matching (Using SET to compare, order do not matter e.g. in SELECT statement), Exact Matching, and Execution Accuracy (False Positive like produce NULLs).

SQL query has 4 level of hardness. Select From Where —> JOIN + Group By –> + Having –> + Sub Query

Model : Seq2Seq ==> + Attention == > Copy (attention based copying operation)

Example Split – Question for the same database can appear in both train and test Database Split – All question for the same database are in the same split.

Some other leader baord work done by PROMPT ENGINEERING. E.g. First open AI Text to SQL, all info commented by a Pound/Hash sign '#'

What is bottleneck(s)/challenge(s) in client's data interpretation

What is the value proposition for applying data science techniques

In order to help the user see the data, concrete the reasoning part, increase ability to trust.

DELIVERABLES & TIMELINE & COSTS (3 marks)

Expected outcomes of the project

Utlising multi agent system to E.g. - a Schema Intelligence Agent that automatically maps database tables and relationships into structured formats (such as knowledge graphs), - a Query Understanding Agent that processes natural language queries, - a Table Relevance Agent that identifies relevant tables, - a Record Discovery Agent that finds relevant records within those tables.

Data analytics & visualisation outputs

You'll build a web dashboard or other visualization tools where users input natural language queries and receive relevant tables with highlighted records. If possible, accompanied by clear explanations of why each component was selected.

Presentation, reports and codes

Timeline of their completion

METHODS (5 marks)

Preliminary interrogation of the data

- Task 1 2 3 4 5 ?

Choice of different data exploration/analytical techniques to test

How to make the process understandable and the output usable

Visualisation to communicate the processes/results

How to use collaborative tools to ensure data and code version controls