



1st SouthStat Meeting

Encontro Sul Brasileiro de Estatística e Ciência de Dados

14 E 15 DE DEZEMBRO DE 2023. CURITIBA-PR.

Can open-source Large Language Models replace GPT-4 in the text-to-SQL task?

Ruanitto Roberto Docini¹

UFPR (Federal University of Parana)

Luiz Oliveira²

UFPR (Federal University of Parana)

ABSTRACT

This work has as its goal to explore how open-source large language models, made available to the community, behave in the text-to-SQL task, where a natural language question is translated automatically into a SQL code, and executed in a database to retrieve the values that answer the question proposed. The reason why this is pertinent to execute such analysis is based on the fact that nowadays, the state-of-the-art is predominantly dominated by proprietary models of big tech companies, in the majority of the cases, gpt-4, owned by OpenAI, with closed code and training techniques. Such a fact is a concern, since, is a limiter to the democratization of these solutions and also for the enhancement of them, once that the access to the resource is limited to those who have the monetary means to pay for the model Api. Moreover, for it's closed implementation, the utilization of a private models, limit the comprehension of certain aspects of the final inference result, what makes the exploration of it's limitations harder, e even it's enhancement more complicated because of the lack of informations of how the model is built and trained. Under this aspect, we conducted experimentations to comprehend if open-source models can achieve the same performance that closed source models in this task. For our tests, a benchmark data source, called Spider, was used, containing 10181 questions and 5693 SQL queries, about 200 different databases covering 128 unique domains. We adapted the best performing technique in the area, called DAIL-SQ, replacing the language model component, that originally uses gpt-4 for three open-source families of LLMs, utilizing different sizes of architecture in each one of them, for us to explore also, how model size can interfere in getting results closer to the leading solutions in the field. Results show that these models, in general, don't present a satisfactory result when we replicate the techniques that are today considered state-of-the-art, being 25% the best performance of execution accuracy obtained when we adapt the most successful solutions with all the experimented open-source large language models.

Key-words: large language models; text-to-SQL; open-source; prompt-engineering; SQL queries.