中国科学技术大学

图数据库

(202X 年 7 月 – 202X 年 7 月)

LLM FOR DB

张昱　课题组

系统软件与软件安全研究组 (S4Plus)

中国科学技术大学 · 计算机科学与技术学院

目 录

1 工作日志

1.1 2021-2022 学年

2 概述

Part I

论文篇

3 程序分析相关论文

3.1 [arXiv2024 Effective Bug Detection in Graph Database Engines: An LLM-based Approach]

[?] 现在许多领域需要使用图数据库存储数据，但是目前的数据库 bug 检测工具限制了查询的语言，以及查询的环境。并且需要使用者有预备的知识来生成查询语句来检测其中的 bug.

3.1.1 novel paradigm tool: DGDB

使用 LLMs 例如 Chatgpt 来生成高质量的查询语句，并部署到不同的图数据库中，来进行全面的自动化检测。

具体来说，在其随机生成一个属性图后，将图数据，查询生成指令，查询生成约束交给 LLM，让其生成满足这些条件的查询语句，并在多种不同的数据库中运行，比较器结果，发现问题。

3.1.2 contribution

- 提出了以 prompt 来帮助 gpt 生成非空的查询语句，并且丰富其生成的查询语句类型，提高其检测效率。
- 提出了一个图数据库错误检测的简单范式，使使用人员不需要有先验知识，并且可以应用到不同查询语言的数据库当中。
- 做了相应的实验并找到了传统检测方法无法检测到的错误，证明了提出方法的有效性。

3.1.3 框架

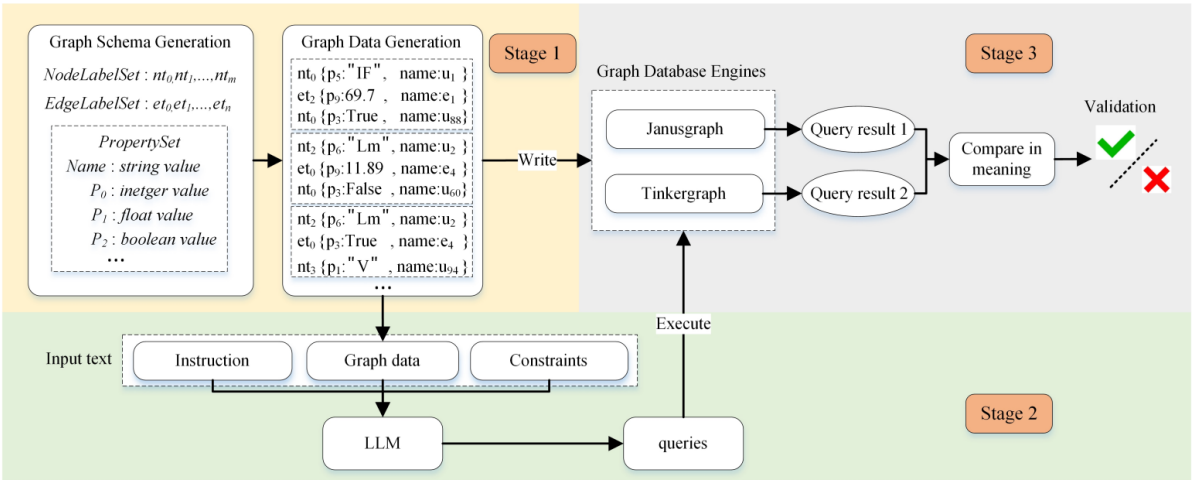

图 1: DGDB 框架图

3.1.4 结果

实验结果表明，本文的方法可以有效地发现两种查询语言的最新版本的图形数据库引擎中的错误。

3.1.5 局限性

- 实验完整度不够
- 虽然本文一直强调本文的方法可以使使用者不需要关于图数据库的相关知识，但是在利用 LLM 的过程中以及比对不同数据库的返回结果，若是两个数据库返回结果不同，还是需要有相关的知识来判别。

3.2 VLDB2024:D-Bot: Database Diagnosis System using Large Language Models

3.2提出目前数据库的诊断主要依靠数据库管理者，但是往往数据库管理者的处理效率十分低，这在很多情况下是不可接受的。并且现在的方法只支持有限的诊断情况，例如劳动密集型的数据库更新。

本文提出了一种基于 LLM 的数据库诊断系统 D-Bot，其可以自动的从诊断文件中获取信息，并且生成诊断报告。

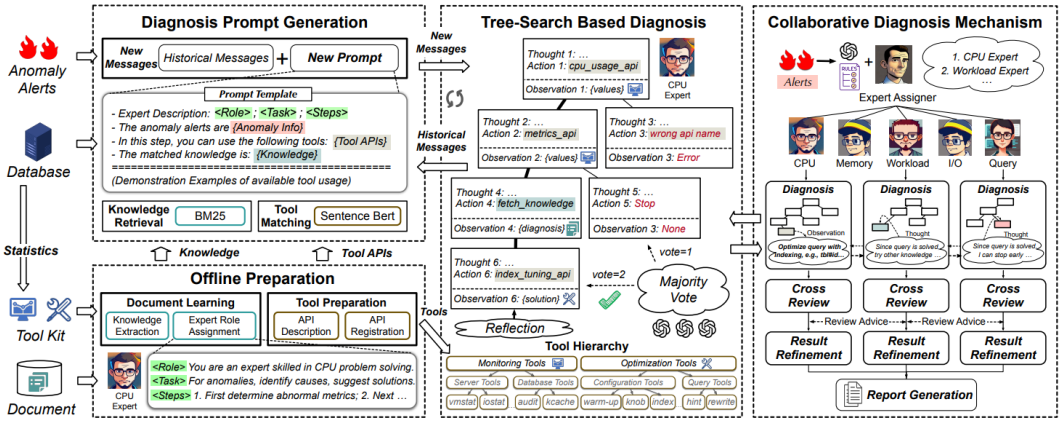

Figure 3: Database Diagnosis in *D-Bot*.

图 2: dbot$_d$*iagnosis*

3.2.1 D-Bot

[H] 相比于传统方法和 GPT4，本文提出的 D-bots 的效果更加优秀。下列是其相应的技术：

- 从文件离线提取信息
- 自动 prompt 生成 e.g. 知识匹配和工具检索
- 使用树检索方法对根本原因进行分析
- 多根因复杂异常协同机制

3.2.2 传统方法与 D-Bot

[redacted]

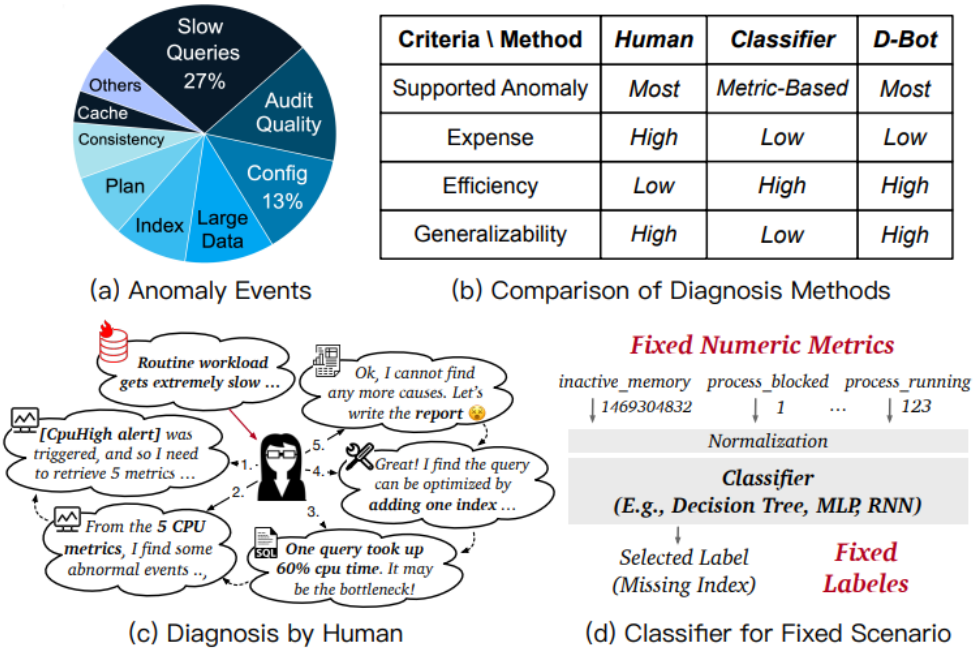

图 3: ways

- 用人工的方法, 延迟太大, 并且需要很大的时间消耗来培养数据库管理人员
- 用半自动化的方式, 虽然一定程度减少了时间消耗以及人工成本, 但是因为其基于小模型, 没有人工方式的思维能力, 无法准确的找到根本原因
- D-Bot
  1. 精准诊断
  2. 节约时间
  3. 高泛化性

3.2.3 实验

实验设置了许多对比实验以及相对应的消融实验

对比试验

- HumanDBA 从业两年的系统诊断人员
- D-Bot(GPT-4) 使用 GPT-4 的 D-BOT

- D-Bot(GPT-3.5) 使用 GPT-3.5 的 D-BOT
- DNN 具有 ReLU 激活的两层神经网络将输入的异常度量向量分类为一个或多个根本原因
- DecisionTree 使用决策树算法来标记输入度量值的根本原因
- GPT-4 传统 GPR-4
- GPT-3.5 传统 GPT-3.5

相融实验

- 无提取知识的 D-bot
- 采用思维链的 D-bot

Table 4: Performance on different anomalies.

Diagnosis Method	Single Cause Anomaly		Multi-Cause Anomaly	
	Acc	HEval	Acc	HEval
HumanDBA	0.955	0.720	0.487	0.806
D-Bot (GPT-4)	0.754	0.500	0.655	0.669
D-Bot (GPT-3.5)	0.542	0.370	0.533	0.493
DNN	0.352	N/A	0.036	N/A
DecisionTree	0.331	N/A	0.086	N/A
GPT-4	0.351	0.39	0.105	0.151
GPT-3.5	0.266	0.2	0.144	0.130

图 4: RESULT

3.2.4 结论

在本文中，提出了一种利用大型语言模型（LLM）的数据库诊断系统。从文档中进行了离线知识提取，并从现有工具中准备了函数 API。我们将合适的知识和 API 匹配到 LLM 提示中进行在线诊断，并提出了一种基于树搜索的算法，以准确有效地利用工具并利用知识进行分析。设计了一种协作诊断机制，通过多个 LLM 的协作提高了效率。实验结果表明，D-Bot 与基线和人类 DBA 相比取得了显着的改进。

3.3 VLDB2024 Combining Small Language Models and Large Language Models for Zero-Shot NL2SQL

论文提出了一种名为 Z-NL2SQL 的新框架，通过结合小型语言模型（SLMs）和大型语言模型（LLMs），解决零样本 NL2SQL 问题。框架分为两个阶段：SQL 草图生成和 SQL 查询补全及修正。通过综合利用 SLMs 的架构识别和 LLMs 的复杂语言推理能力，Z-NL2SQL 在多种基准测试中实现了最优性能，比现有方法的执行准确率提高了 10% 至 20%。

3.3.1 background

自然语言到 SQL（NL2SQL）是一种将自然语言问题转化为 SQL 查询的技术，旨在降低非技术用户访问和分析数据的门槛。当前模型面临的问题是新环境（数据库和语言现象）中的泛化能力不足，尤其在缺乏标注数据的零样本情况下。现有方法主要依赖小型模型（如 BART、T5）或大型模型（如 GPT-4），但它们在复杂推理和精确架构匹配方面各有不足。

3.3.2 contribution

- 双模型协同框架：

  通过分解任务，SLMs 负责生成 SQL 草图，LLMs 负责补全和修正。这种组合充分发挥了两类模型的优势。

- 数据库感知序列化：

  提出了一种新型序列化方法，使得 SLM 能够在新数据库中更好地泛化。

- 问题感知对齐器：

  使用多级匹配策略，引导 LLM 精确补全 SQL 查询，确保与数据库内容的一致性。

- 执行驱动选择策略：

  通过 SQL 执行结果来选择最优查询，从而提升模型的可靠性和准确性。

3.3.3 框架

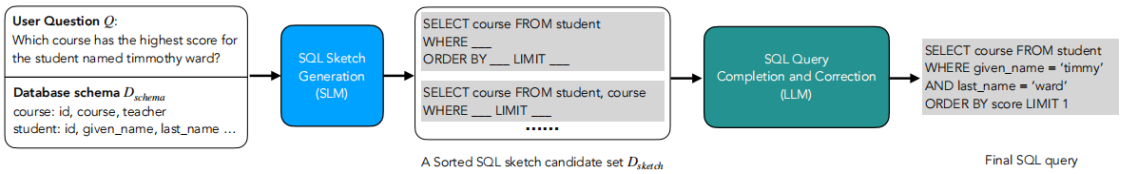

Figure 4: ZeroNL2SQL Overview. Given a user question Q and a database schema D_{schema}, (1) an *SQL Sketch Generation* Module generates a list of SQL sketch candidates, and (2) an *SQL Query Completion* Module completes the SQL query.

图 5: overview

3.3.4 实验

数据集：使用 Spider 数据集进行训练，并在 Dr.Spider、KaggleDBQA 和 GeoQuery 数据集上测试。这些数据集模拟了不同的零样本环境，包括数据库、语言和 SQL 的变化。

基线模型：对比了多种 SOTA 方法，包括 SMBOP、RESDSQL 和微调的 LLaMA2。

评估指标：采用执行准确率（Execution Accuracy, EX）来衡量模型生成的 SQL 查询在数据库中的执行结果。

其也做了一些消融实验，来验证系统中的部分做了哪些优化例如:

- SQL sketch 生成模块中每个组件的作用
- 评估 SQL 查询完成和更正模块中每个组件的作用

3.3.5 结论

- 论文提出了一种创新性框架，通过结合 SLMs 和 LLMs，显著提升了零样本 NL2SQL 任务的性能。
- 实验证明，该框架在执行准确率和泛化能力方面均超越了现有的最先进方法。
- 未来的工作可能包括进一步优化多级匹配策略及处理更复杂的 SQL 查询。

References