

CycleSQL

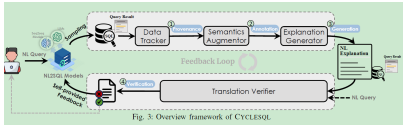
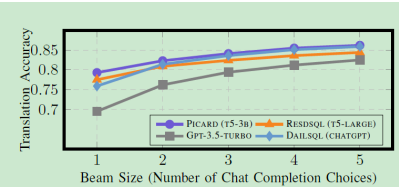
论文: Grounding Natural Language to SQL Translation with Data-Based Self-Explanations

背景

痛点: 现有模型无法一次性生成正确的SQL

主要贡献

- 1、循环反馈机制
- 2、对于NL2SQL融入更加丰富的解释
 - data-level
 - operation-level
- 3、大量、高质量评测



Data Tracking

- 具体方式: 通过重写SQL召回数据层面的信息
- 思想: 查询结果如何存储在底层数据库中派生出来的

- 1、将查询结果转换为where语句
如果为*, 则跳过此规则
- 2、提取原始SQL中的所有列
- 3、消除聚合函数和Group by语句

聚合函数可能会折叠某些数据行, 本质上隐藏了追踪数据来源。

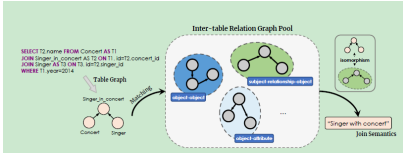
Semantics Enrichment

捕获query与SQL之间的语义关系

Explanation Generation

将上述源信息转换为NL表达

方式: 通过有向图来生成



Explanation: The query returns a result with one column of aggregation (type count) and one row. That is, for flights with aircraft Airbus A340-300, there are 2 flights in total. (intermediate reasoning step ❶)

示例

Translation Verification.

- 根据查询结果生成的NL解释作为自供反馈来验证底层NL2SQL的准确性。
- 使用LLM判断生成的解释是否包含原始NL查询 (NLI任务)

具体方式: 自建具有蕴含和矛盾对的NLI模型

