Homework 4: LLMs

1 Finetuned Errors Analysis

Here we discuss and demonstrate errors exists in the Text2SQL task on finetuned model.

1.1 Error Types

- **Syntax Error: Incomplete SQL statement.** In generated SQL statements, there are a large number of statements that are wrong due to incomplete. These statements are truncated because of length limitations, this problem could be caused by the model actually failing to generate a proper statement, while the answers are not violating the length limitations.

  Here is one example:

```
SELECT DERIVED_TABLEalias0.LENGTH FROM
DERIVED_TABLE AS DERIVED_TABLEalias0
WHERE DERIVED_TABLEalias0.DERIVED_FIELDalias0 =
( SELECT MAX( DERIVED_FIELDalias1.DERIVED_FIELDalias1 ) FROM
( SELECT RIVERalias0.TRAVERSE FROM RIVER AS RIVERalias0 WHERE
```

- **Syntax Error: Missing/Wrong Delimiter.** Another syntax error is due to missing/wrong delimiter of the SQL statement, here is one example:

```
SELECT MOUNTAINalias0.TRAVERSE FROM MOUNTAIN AS MOUNTAINalias0
WHERE MOUNTAINalias0.MOUNTAIN_NAME = "mckinley"
```

  From my experiments and observation, there are no executable but wrong-in-return SQL statements are generated.

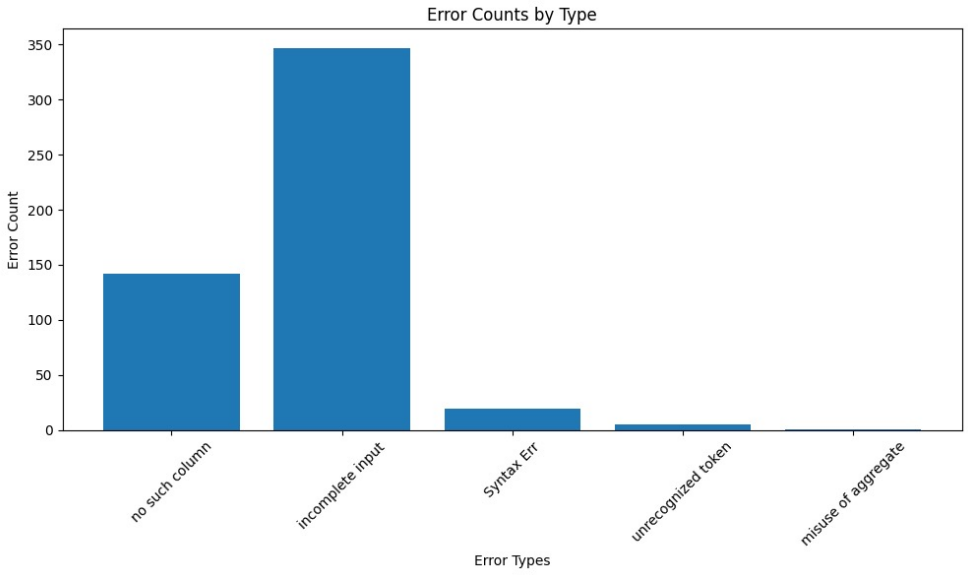

Figure 1: Error Types distribution

2 Execution Accuracy vs. Exact Match Accuracy

Execution accuracy calculates by comparing the return value of database while the exact match accuracy calculates by comparing the referenced SQL statement.

Execution accuracy allows more flexible answers while exact match accuracy does not. However, execution accuracy also provides a possible noisy accuracy measurement. For example, when we ask *What is the biggest financial city in California?* and *What is the biggest financial city in the West of the USA?*, these two questions have the same answers but should execute different SQL statements. In the case of running execution accuracy, we will allow the model to generate the same answers, while exact match accuracy does not.

3 Open-ended Exploration

Refer to Figure 1, we can observe that the second and third most common errors are caused by no such column and delimiter missing syntax error, so I tried to add this information in the prompt to help improve the accuracy. This actually reduces the errors caused by delimiter missing, but accuracy does not change too much.

Here, the prompt becomes: *Write a SQL query based on the following question. Remember to add a delimiter; .*