SQL Test Case Generation Using Multi-Objective Optimization

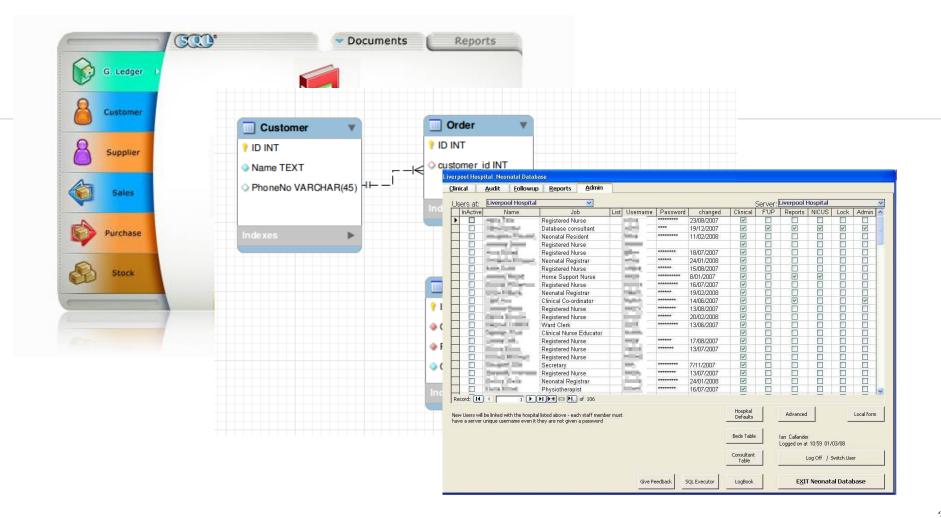


20184228 Seah Kim 20184400 Jeonggwan Lee 20186473 Liu Lingjun 20186505 Nick Heppert



Table of Contents

- ✓ Introduction of motivated paper
- ✓ Our challenge
- ✓ Our goal and expectation



SELECT items .* FROM invoice JOIN items ON invoice.id = items.invoiceid Howstroudnue cest this query?



Search-based Test Data Generation for SQL Queries

Jeroen Castelein et al. ICSE 18'



Search-based Test Data Generation for SQL Queries (Jeroen Castelein et al. ICSE 18')

- ✓ Modeled the problem of test data generation for SQL queries as a search-based problem.
- ✓ Implemented three different search approaches in tool, named EvoSQL
 - (random search, biased random search, genetic algorithms)
- Executed on 2,135 queries extracted from 4 software systems



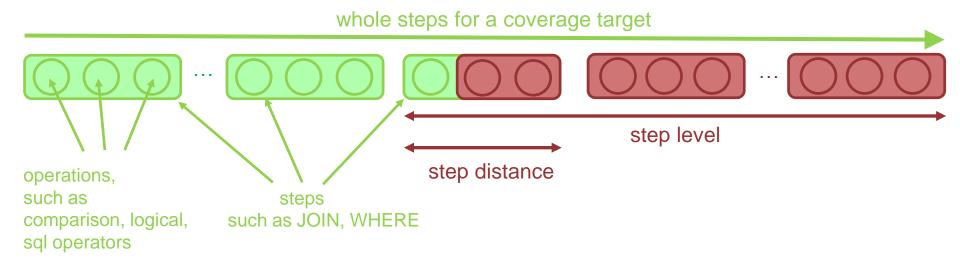
Fitness Function

```
FROM Cars
JOIN Tires
ON Cars.tire_id = Tires.id
WHERE model = 'Ferrari'
Step 2
```

If Step 1 doesn't produce an output, database stops its execution before proceeding to Step 2



Fitness Function



Fitness function = step level + step distance

Problem: Inefficiency of single-target strategy

1. The order of each coverage target is not optimized

2. Inefficient allocation of the budget might happen, such as infeasible coverage target.

Solution

Multi-Objective Optimization!





SELECT * FROM Product WHERE (Category = 'Toy')



SELECT * FROM Product WHERE (Category = 'Toy')

| id | category |
|----|----------|
| 1 | Toy |
| 2 | Food |
| 3 | Car |

Product



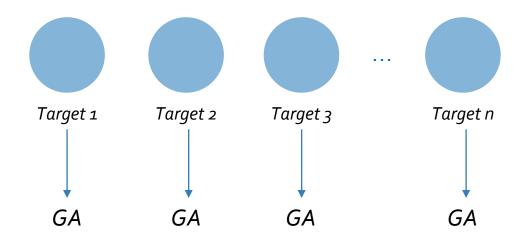
SELECT * FROM Product WHERE (Category != 'Toy')

| id | category |
|----|----------|
| 1 | Toy |
| 2 | Food |
| 3 | Car |

Product

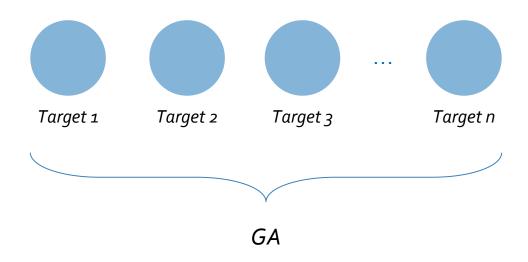


Challenge 1. The order of each coverage target being executed is not optimized



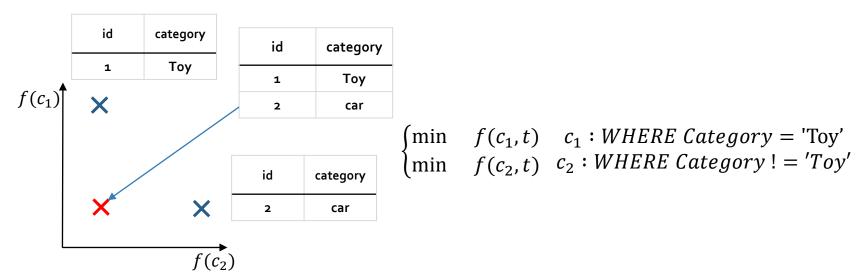


Challenge 1. The order of each coverage target being executed is not optimized





SELECT * FROM Product WHERE (Category = 'Toy')





Challenge 2. Inefficient allocation of the budget might happen, such as infeasible coverage target.

"A > 10 and A < 10"? \rightarrow infeasible, no solution from GA

$$\begin{cases} \min & f(c_1,t) & f(c_1,t) => feasible \\ \min & f(c_2,t) & f(c_2,t) => Infeasible \\ \dots & \dots \\ \min & f(c_m,t) & f(c_m t) => feasible \end{cases}$$

Solution from EvoSQL is to manually remove infeasible coverage targets.



Our Expectation

- ✓ Our MOO version can solve more complex queries simultaneously compared to single-objective strategy.
- Each solution is guided by hinting the similar solution from other coverage target.
- ✓ It doesn't consume improper budget for handling the infeasible coverage target .

Questions?

Thanks for your attention!

