Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity

Anna Fariha
afariha@cs.umass.edu
University of Massachusetts Amherst

Alexandra Meliou
ameli@cs.umass.edu
University of Massachusetts Amherst

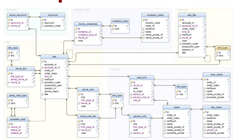
Are databases accessible to non-experts?

65% structured data resides in relational databases

Traditional data retrieval is challenging

- Non-expert users struggle to formulate complex SQL query

complex schema



no SQL expertise

```
SELECT person.name
FROM person, castinfo, movietogenre, genre
WHERE person.id = castinfo.person_id
  AND castinfo.movie_id = movietogenre.movie_id
  AND movietogenre.genre_id = genre.id
  AND genre.name = 'Comedy'
GROUP BY person.id
HAVING count(*) >= 40
```

Query by Example (QBE) tries but fails

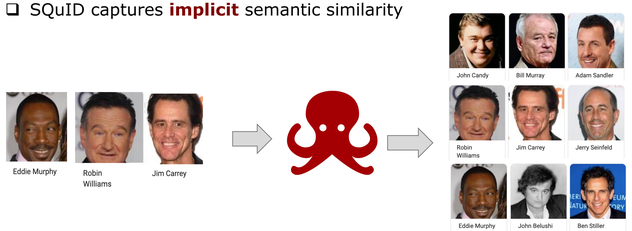
- Traditional QBE systems only see the **type** of the examples and **overlooks semantic similarity**

QBE

list of all actors

SQuID: Semantic similarity-aware Query Intent Discovery

- SQuID is aware of **semantic similarity**
- SQuID captures **implicit** semantic similarity

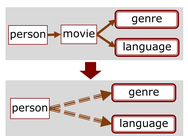


Basic semantic properties

- Directly affiliated with entities

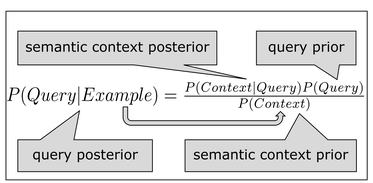
Derived semantic properties

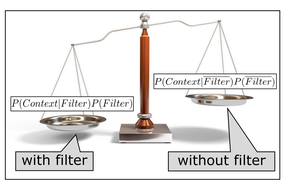
- Aggregate over basic property of associated entities



SQuID applies abduction to reject irrelevant similarities

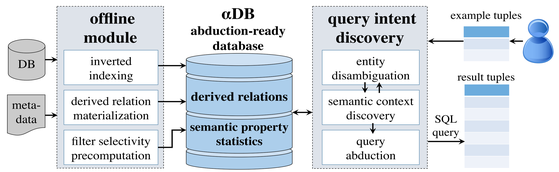
- **Abduction**: most likely **explanation** of an **observation**
- **Probabilistic abduction model**: most likely query given examples



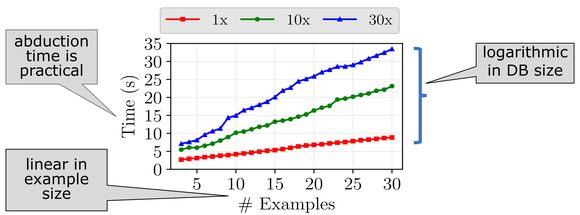


SQuID architecture

- **Real-time** performance through **abduction-ready database** – precomputed offline
- **Online module** discovers semantic context, captures the **most likely query intent**, and constructs SQL query

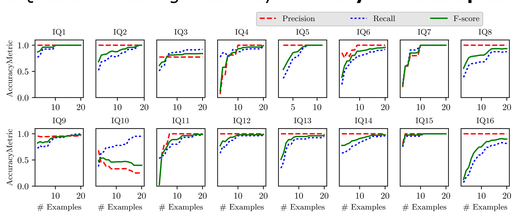


SQuID is scalable



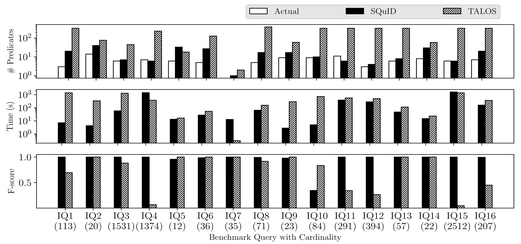
IMDb benchmark queries

- SQuID achieves high accuracy with **very few examples**



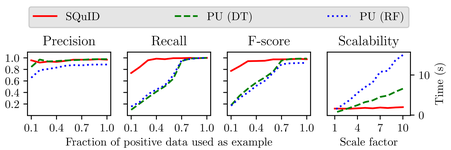
SQuID outperforms query reverse engineering

- QRE over-fits the examples, and fails to generalize intent

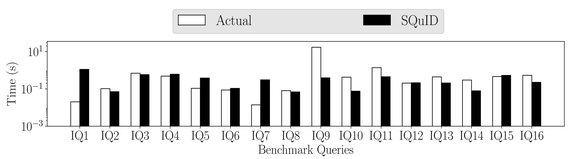


SQuID outperforms positive and unlabeled learning

- PU learning requires 70% data as example
- PU learning is not scalable



SQuID produces efficient queries



Datasets

IMDb · dblp · UCI Machine Learning Repository

squid.cs.umass.edu

UMASSCS