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

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

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

Are databases accessible to non-experts?

- Traditional data retrieval is challenging
- Query by Example (QBE) tries but fails

Traditional data retrieval is challenging

- Non-expert users struggle to formulate complex SQL query

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

QBE

SQuID: Semantic similarity-aware Query Intent Discovery

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

Query by Example (QBE) tries but fails

- SQuID produces efficient queries

SQuID outperforms positive and unlabeled learning

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

SQuID outperforms query reverse engineering

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

SQuID is scalable

- SQuID achieves high accuracy with very few examples

Example tuples

IMDb benchmark queries

- SQuID outperforms query reverse engineering
- SQuID outperforms positive and unlabeled learning

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 applies abduction to reject irrelevant similarities

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

SQuID is scalable

- SQuID is aware of semantic context, captures association entities, and discovers semantic context
- SQuID is aware of semantic context, captures association entities, and discovers semantic context

Basic semantic properties

- Directly affiliated with entities

Derived semantic properties

- Aggregate over basic property of associated entities

SQuID produces efficient queries

- SQuID achieves high accuracy with very few 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

- SQuID achieves high accuracy with very few examples

SQuID applies abduction to reject irrelevant similarities

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