Conjunctive Queries: Minimization

(Chapter 16 of DBT)

[DBT] Database Theory, https://github.com/pdm-book/community

Complexity of Query Evaluation

Theorem: CQ-Evaluation is NP-complete, and in PTIME in data complexity

Proof:

(NP-membership) Guess-and-check:

- Consider a database D, a CQ $Q(x_1,\ldots,x_k)$:- body, and a tuple $(a_1,\ldots,a_k)$ of values
- Guess a substitution h : terms(body) → terms(D)
- Verify that h is a match of Q in D, i.e., h(body) ⊆ D and $(h(x_1),\ldots,h(x_k)) = (a_1,\ldots,a_k)$

(NP-hardness) Reduction from 3-colorability

(in PTIME) For every substitution h : terms(body) → terms(D), check if h(body) ⊆ D

and $(h(x_1),\ldots,h(x_k)) = (a_1,\ldots,a_k)$

Complexity of Static Analysis

Theorem: Let Q_1 and Q_2 be conjunctive queries. The problem of deciding whether there exists a query homomorphism from Q_2 to Q_1 is NP-complete

Proof:

(NP-membership) Guess a substitution, and verify that is a query homomorphism

(NP-hardness) Easy reduction from **CQ**-Evaluation

By applying the homomorphism theorem we get that:

Corollary: **CQ**-Equivalence and **CQ**-Containment are NP-complete

Minimizing Conjunctive Queries

- **Goal:** minimize the number of joins in a query

- A conjunctive query Q_1 is minimal if there is no conjunctive query Q_2 such that:
  1. $Q_1 \equiv Q_2$
  2. Q_2 has fewer atoms than Q_1

- The task of CQ minimization is, given a conjunctive query Q, to compute a minimal one that is equivalent to Q

Homomorphism Theorem

A query homomorphism from $Q_1(x_1,\dots,x_k)$:- $body_1$ to $Q_2(y_1,\dots,y_k)$:- $body_2$ is a substitution $h : terms(body_1) \rightarrow terms(body_2)$ such that:

1. h is a homomorphism from $body_1$ to $body_2$
2. $(h(x_1),\dots,h(x_k)) = (y_1,\dots,y_k)$

Homomorphism Theorem: Let Q_1 and Q_2 be conjunctive queries. It holds that:

$Q_1 \subseteq Q_2$ iff there exists a query homomorphism from Q_2 to Q_1

Minimization by Deletion

By exploiting the homomorphism theorem we can show the following:

Theorem: Consider a conjunctive query $Q_1(x_1,\ldots,x_k)$:- $body_1$.

If Q_1 is equivalent to a conjunctive query $Q_2(y_1,\ldots,y_k)$:- $body_2$ where $|body_2| < |body_1|$,

then Q_1 is equivalent to a query $Q_3(x_1,\ldots,x_k)$:- $body_3$ such that $body_3 \subseteq body_1$

$$\Downarrow$$

The above theorem says that to minimize a conjunctive query $Q_1(x_1,\ldots,x_k)$:- body we simply need to remove some atoms from body

Minimization Procedure

Minimization($Q(x_1,\ldots,x_k)$:- body)

Repeat until no change

choose an atom $\alpha \in$ body such that the variables $x_1,\ldots,x_k$ appear in body \ $\{\alpha\}$

if there is a query homomorphism from $Q(x_1,\ldots,x_k)$:- body to $Q(x_1,\ldots,x_k)$:- body \ $\{\alpha\}$

then body := body \ $\{\alpha\}$

Return $Q(x_1,\ldots,x_k)$:- body

Note: if there is a query homomorphism from $Q(x_1,\ldots,x_k)$:- body to $Q(x_1,\ldots,x_k)$:- body \ $\{\alpha\}$, then the two queries are equivalent since there is trivially a query homomorphism from the latter to the former query

Minimization Procedure: Example

(a,b,c,d are constants)

Q(x) :- R(x,y), R(x,b), R(a,b), R(u,c), R(u,v), S(a,c,d)

$\{y \mapsto b\}$

Q(x) :- R(x,b), R(a,b), R(u,c), R(u,v), S(a,c,d)

$\{v \mapsto c\}$

Q(x) :- R(x,b), R(a,b), R(u,c), S(a,c,d)

minimal query

Note: the mapping $x \mapsto a$ is not valid since x is a distinguished variable

Uniqueness of Minimal Queries

Natural question: does the order in which we remove atoms from the body of the input conjunctive query matter?

Theorem: Consider a conjunctive query Q. Let Q_1 and Q_2 be minimal conjunctive queries such that $Q_1 \equiv Q$ and $Q_2 \equiv Q$. Then, Q_1 and Q_2 are isomorphic (i.e., they are the same up to variable renaming)

Therefore, given a conjunctive query Q, the result of Minimization(Q) is unique (up to variable renaming) and is called the core of Q