Conjunctive Queries: Minimization

(Chapter 16 of DBT)

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(x₁,...,xₖ) : body, and a tuple (a₁,...,aₖ) 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(x₁),...h(xₖ)) = (a₁,...,aₖ)

(NP-hardness) Reduction from 3-colorability

(in PTIME) For every substitution h : terms(body) → terms(D), check if h(body) ⊆ D and (h(x₁),...h(xₖ)) = (a₁,...,aₖ)

Complexity of Static Analysis

Theorem: Let Q₁ and Q₂ be conjunctive queries. The problem of deciding whether there exists a query homomorphism from Q₂ to Q₁ 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 is minimal if there is no conjunctive query Q₂ such that:
  1. Q₁ ≡ Q₂
  2. Q₂ has fewer atoms than Q₁

- The task of CQ minimization is, given a conjunctive query Q₁, to compute a minimal one that is equivalent to Q₁
Homomorphism Theorem

A query homomorphism from $Q_1(x_1,\ldots,x_k) : \text{body}_1$ to $Q_2(y_1,\ldots,y_k) : \text{body}_2$ is a substitution $h : \text{terms(body}_1) \rightarrow \text{terms(body}_2)$ such that:

1. $h$ is a homomorphism from $\text{body}_1$ to $\text{body}_2$
2. $(h(x_1),\ldots,h(x_k)) = (y_1,\ldots,y_k)$

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

$Q_1 \subseteq Q_2$ if and only if 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 $Q_1(x_1,\ldots,x_k) : \text{body}_1$.

If $Q_1$ is equivalent to a conjunctive query $Q_2(y_1,\ldots,y_k) : \text{body}_2$ where $|\text{body}_2| < |\text{body}_1|$, then $Q_1$ is equivalent to a query $Q_3(x_1,\ldots,x_k) : \text{body}_3$ such that $\text{body}_3 \subseteq \text{body}_1$.

Minimization Procedure

Minimization($Q(x_1,\ldots,x_k) : \text{body}$)

Repeat until no change

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

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

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

Return $Q(x_1,\ldots,x_k) : \text{body}$

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

$(y \mapsto b)$

$(v \mapsto c)$

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 $\text{Minimization}(Q)$ is unique (up to variable renaming) and is called the **core** of $Q$. 