Containment of Conjunctive Queries in Datalog Programs

Jef Wijsen

April 17, 2010

1 Problem Statement

Let $q : H \leftarrow B$ be a rule-based conjunctive query. Let Answer be the relation name that occurs in H. Let P be a datalog program, without negation, such that Answer occurs in the head of some rule of P. Is there an algorithm to decide whether $q \sqsubseteq P$? Here, $\sqsubseteq$ has its standard meaning, that is, $q \sqsubseteq P$ if and only if for every database I, $q(I) \subseteq P(I)$.

Example 1 Let P be the datalog program with two rules:

$$\begin{aligned} S(x,y) &\leftarrow G(x,z), G(z,y) \\ S(x,y) &\leftarrow S(x,z), S(z,y) \end{aligned}$$

Let q be the following rule:

$$S(x,y) \leftarrow G(x,u), G(u,w), G(w,y)$$

Let $I = \{G(1,2), G(2,3), G(3,4)\}$. Then $q(I) = \{S(1,4)\}$ and $P(I) = \{S(1,3), S(2,4)\}$. Consequently, $q \not\sqsubseteq P$.

2 Decidability Result

Theorem 1 *Let $q : H \leftarrow B$ be a rule-based conjunctive query. Let P be a datalog program such that the relation name of H occurs in the head of some rule of P. Let ν be a valuation that maps each variable of q to a new constant. Then, $q \sqsubseteq P$ if and only if $\nu(H) \in P(\nu(B))$.*

Proof. $\boxed{\Rightarrow}$ Assume $q \sqsubseteq P$. Since $\nu(H) \in q(\nu(B))$ is obvious, we have $\nu(H) \in P(\nu(B))$.

$\boxed{\Leftarrow}$ Assume $\nu(H) \in P(\nu(B))$. Let I be an arbitrary ground database. Let L be a ground atom such that $L \in q(I)$. We need to show $L \in P(I)$.

We can assume a valuation θ such that $\theta(B) \subseteq I$ and $\theta(H) = L$. Consider the following two sequences:

$$\begin{array}{llll} \nu(B) = T_P^0(\nu(B)), & T_P^1(\nu(B)), & T_P^2(\nu(B)), & \ldots \\ \theta(B) = T_P^0(\theta(B)), & T_P^1(\theta(B)), & T_P^2(\theta(B)), & \ldots \end{array}$$

We show hereafter that for $k \in \{0, 1, 2, \ldots\}$, $\theta(\nu^{-1}(T_P^k(\nu(B)))) \subseteq T_P^k(\theta(B))$. Consequently, $\theta(\nu^{-1}(P(\nu(B))) \subseteq P(\theta(B))$. From $\nu(H) \in P(\nu(B))$, it follows $\theta(\nu^{-1}(\nu(H))) \in \theta(\nu^{-1}(P(\nu(B)))$, hence $\theta(\nu^{-1}(\nu(H))) \in P(\theta(B))$. Since $\theta(\nu^{-1}(\nu(H))) = \theta(H) = L$, we have $L \in P(\theta(B))$. Since $\theta(B) \subseteq I$, we have $L \in P(I)$ by *Monotonicity* of datalog programs.

The proof of $\theta(\nu^{-1}(T_P^k(\nu(B)))) \subseteq T_P^k(\theta(B))$ for $k \geq 0$ runs by induction on increasing k, as follows:

Induction basis $k = 0$. We have $\theta(\nu^{-1}(T_P^0(\nu(B)))) = \theta(\nu^{-1}(\nu(B))) = \theta(B)$ and $T_P^0(\theta(B)) = \theta(B)$. Consequently, $\theta(\nu^{-1}(T_P^0(\nu(B)))) \subseteq T_P^0(\theta(B))$.

Induction step $k \rightarrow k + 1$. The induction hypothesis is $\theta(\nu^{-1}(T_P^k(\nu(B)))) \subseteq T_P^k(\theta(B))$. We need to show $\theta(\nu^{-1}(T_P^{k+1}(\nu(B)))) \subseteq T_P^{k+1}(\theta(B))$. Let L be an atom such that $L \in \theta(\nu^{-1}(T_P(T_P^k(\nu(B)))))$. We need to show that $L \in T_P^{k+1}(\theta(B))$. Two cases can occur:

- $L = \theta(\nu^{-1}(R(\vec{a})))$ where R is an extensional relation name and $R(\vec{a}) \in T_P^k(\nu(B))$. Then, $L \in \theta(\nu^{-1}(T_P^k(\nu(B))))$. By the induction hypothesis, $L \in T_P^k(\theta(B))$. Since the relation name of R is extensional, $L \in T_P(T_P^k(\theta(B))) = T_P^{k+1}(\theta(B))$.
- P contains a rule $H_0 \leftarrow B_0$ and there is a valuation μ such that $\mu(B_0) \subseteq T_P^k(\nu(B))$ and $L = \theta(\nu^{-1}(\mu(H_0)))$. From $\mu(B_0) \subseteq T_P^k(\nu(B))$, it follows $\theta(\nu^{-1}(\mu(B_0))) \subseteq \theta(\nu^{-1}(T_P^k(\nu(B))))$. By the induction hypothesis, $\theta(\nu^{-1}(\mu(B_0))) \subseteq T_P^k(\theta(B))$. By the definition of T_P, we have $\theta(\nu^{-1}(\mu(H_0))) \in T_P(T_P^k(\theta(B)))$. Consequently, $L \in T_P^{k+1}(\theta(B))$.

$\square$

3 Exercises

1. Let Q be a union-of-rules and P a datalog program. Can it be decided whether $Q \sqsubseteq P$?