Querying Clique Guarded Existential Rules

Sen Zheng Renate A. Schmidt

University of Manchester, Manchester, UK

{sen.zheng,renate.schmidt}@manchester.ac.uk

Abstract: We investigate the Boolean conjunctive query answering problem, where from a given a Boolean conjunctive query q, a database $\mathcal{D}$ and a theory Σ, the aim is to check whether $\mathcal{D} \cup \Sigma \models q$. We show that Boolean conjunctive query answering can be decided using ordered resolution with dynamic selection when Σ is expressed as clique guarded existential rules. Clique guarded existential rules subsume rules commonly used in ontology-based query answering systems such as guarded existential rules.

1 Introduction

Query answering against decidable fragments is one of the fundamental problems behind ontology-based data access systems. This work is concerned with checking whether $\mathcal{D} \cup \Sigma \models q$, where $\mathcal{D}$ is a set of ground facts, Σ is a set of clique guarded existential rules and q is a Boolean conjunctive query. The class of clique guarded existential rules is a Horn fragment of the clique guarded fragment [5]. For the latter, the satisfiability is known as 2EXPTIME-complete [5]. This means the class of clique guarded existential rules is decidable. As far as we know, as yet there is no complexity result for (querying) clique guarded existential rules and there is no approach to decide the problem of query answering for clique guarded existential rules.

Our querying approach is based on ordered resolution with selection and can be embedded into the framework of [1]. Hence, the soundness and refutational completeness result from [1] applies immediately. We use the top variable technique to avoid term depth increase in the resolvents. This technique is inspired by the partial hyperresolution technique from [3, 4], where it is used to decide the loosely guarded fragment [6]. We adapt this approach so that the top variable technique can be used for answering queries. The idea of the top variable technique is that we only apply resolution when the positive premises contains the potentially deepest terms.

2 Preliminaries

An *existential rule* R is a first order formula of the form $\forall X \forall Y \phi(X,Y) \rightarrow \exists Z \psi(X,Z)$ where $\phi(X,Y)$ and $\psi(X,Z)$ are conjunctions of atoms, and called the *body* and the *head* of R, respectively. X, Y and Z are variable sets. An existential rule R is *clique guarded* if each pair of free variables of the head co-occurs in at least one atom of the body. The definition of *clique guarded existential rules* (CGERs) is a strict extension of the definition of guarded existential rules [2] since free variables of the head do not require to occur in a single atom of the body. An example of a CGER is $\forall xyzv_1v_2v_3(A_1(x,y,v_1) \wedge A_2(x,z,v_2) \wedge A_3(y,z,v_3) \rightarrow \exists w B(x,y,z,w))$, which is not a guarded existential rule. The class of CGERs can be seen as a Horn fragment of the clique guarded fragment [5].

A *Boolean conjunctive query* (BCQ) q is a first-order formula of the form $q = \exists X \varphi(X)$ where φ is a conjunction of atoms containing only variables and constants. The rule set Σ denotes a set of clique guarded existential rules and the database $\mathcal{D}$ denotes a set of ground atoms. We answer BCQ satisfiability of $\mathcal{D} \cup \Sigma \models q$ by answering $\mathcal{D} \cup \Sigma \cup \neg q \models \perp$.

A term is *flat* if it is a variable or a ground term. A term is *simple* if it is a variable, a constant or a compound term $f(u_1, ..., u_n)$ where $n > 0$, such that $u_1, ..., u_n$ are variables or ground terms. A *flat* (*simple*) literal is a literal so that every term in it is flat (simple). A *flat* (*simple*) clause is a clause so that every literal in it is flat (simple). Assume a clause $\phi \rightarrow \psi$ where ϕ is a conjunction of flat atoms. *Chained variables* in ϕ are variables that occur in multiple atoms of ϕ. A compound term t is *weakly covering* if for every non-ground, compound subterm s of t, it is the case that $var(s) = var(t)$. A literal L is weakly covering if each argument of L is either a ground term, a variable, or a weakly covering term t, such that $var(t) = var(L)$. A clause C is weakly covering if each term t in C is either a ground term, a variable, or a weakly covering term such that $var(t) = var(C)$.

3 Decision Procedures

Now we describe the steps of our approach to deciding query answering for clique guarded existential rules.

Clausal Transformation. We use *CGER-Trans* to denote the clausal transformation for CGERs. There are three major steps to transform a clique guarded existential rule R into a set of clauses:

1. Rewrite implications by using negations and disjunctions, and transform R into negation normal form, obtaining the formula R_{nnf}.

2. Transform R_{nnf} into prenex normal form and apply outer Skolemisation: if $\forall X$ is the subsequence of all universal quantifiers of the ψ-prefix of subformula $\exists y \psi$ of ψ, then $\psi[y/f(X)]$ is the outer Skolemisation of $\exists y \psi$, obtaining the formula R_{sko}.

3. Drop all universal quantifiers and transform R_{sko} into its conjunctive normal form, denoted as a set of clauses, obtaining the Horn cliques guarded clauses.

By *Query-Trans* we denote the clausal transformation for Boolean conjunctive queries. One can obtain a *query clause* by simply negating a Boolean conjunctive query.

Clausal Normal Forms. A clause C is a *Horn clique guarded clause* if a condensed form of Horn clause C satisfies these conditions:

1. C is simple and weakly covering.
2. There is a set of negative flat literals $\mathcal{L}$ containing chained variables in C.
   (a) If there is no more than two chained variables, all variables in C occur in one literal of $\mathcal{L}$.
   (b) If there is more than two chained variables, all chained variables co-occur in at least one literal of $\mathcal{L}$. These literals are called *guards*. All variables in non-guard literals of C occurs as chained variables in guards.

The definition of Horn clique guarded clause is a proper superset of the definition of the clique guarded existential rules because function symbols are allowed in negative literals. A ground fact is a Horn clique guarded clause and a *query clause* is a negative flat clause.

Resolution Refinement. We use an admissible ordering and a selection function making use of the top variable technique to restrict the application of resolution. Let *Query-Refine* denote the refinement using: A lexicographic path ordering $\succ_{lpo}$ based on a precedence $f > a > p$ for f denoting function symbols, a denoting constants and p denoting predicate symbols, and a selection function such that the following conditions all hold:

1. If a clause contains negative non-ground compound literals, then at least one of these literals is selected.
2. If there is no negative non-ground compound literal, but there are positive non-ground compound literals, then the maximality principle with respect to $\succ_{lpo}$ is applied to determine the eligible literals.
3. If a clause contains no non-ground compound literals, select all the negative literals containing top variables.

Condition 3 indicates that the selected literals are not set in advance. Instead, in each inference, these selected literals are determined before resolution by checking whether they contain top variables. Hence, we call this kind of selection the dynamic selection.

4 Termination

We use *Query-Res* to denote the calculus consisting of: the condensation rule, tautology elimination, ordered factoring and ordered resolution with selection defined by *Query-Refine*.

Claim 1 *In an application of* Query-Res*, the resolvents of a set of Horn clique guarded clauses and query clauses are query clauses.*

Claim 2 *In an application of* Query-Res*, the resolvents of a set of Horn clique guarded clauses are Horn clique guarded clauses.*

Claim 1 and Claim 2 show that *Query-Res* can guarantee that given a set of Horn clique guarded clauses and query clauses, all derived clauses are either Horn clique guarded clauses or query clauses.

Claim 3 *In an application of* Query-Res*, given a finite set of fixed-length Horn clique guarded clauses and fixed-length query clauses, each derived clause have a fixed length.*

Claim 3 shows that using *Query-Res*, a derived clause cannot be arbitrarily long. Claim 1, 2 and 3 show that:

Claim 4 Query-Res *decides query clauses and Horn clique guarded clauses.*

The following claim shows the main result of this work:

Claim 5 *The combination of the clausal transformations (*CGER-Trans *and* Query-Trans*) and resolution procedures* Query-Res *decide the Boolean conjunctive query answering problem for the clique guarded existential rules.*

5 Conclusion

We developed a decision procedure for answering Boolean conjunctive queries against clique guarded existential rules, based on ordered resolution and a sophisticated form of selection. Since this is still ongoing work, current work is focused on offering formal proofs to support our claims.

References

[1] Leo Bachmair and Harald Ganzinger. Resolution theorem proving. In Alan Robinson and Andrei Voronkov, editors, *Handbook of Automated Reasoning*, pages 19–99. Elsevier and MIT Press, 2001.

[2] Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, and Andreas Pieris. Datalog+/-: A family of logical knowledge representation and query languages for new applications. In *Proc. LICS'10*, pages 228–242. IEEE, 2010.

[3] Hans de Nivelle and Maarten de Rijke. Deciding the guarded fragments by resolution. *J. Symb. Comput.*, 35(1):21–58, 2003.

[4] Harald Ganzinger and Hans de Nivelle. A superposition decision procedure for the guarded fragment with equality. In *Proc. LICS'99*, pages 295–303. IEEE, 1999.

[5] Erich Grädel. Decision procedures for guarded logics. In *Proc. CADE'16*, pages 31–51. Springer, 1999.

[6] Johan van Benthem. Dynamic bits and pieces. Research Report LP-97-01, Univ. Amsterdam, 1997.