On the first-order rewritability of conjunctive queries over binary guarded existential rules (extended abstract)

Cristina Civili, Riccardo Rosati

Dipartimento di Ingegneria informatica, automatica e gestionale
Sapienza Università di Roma

Abstract. We study conjunctive query answering and first-order rewritability of conjunctive queries for binary guarded existential rules. In particular, we prove that the problem of establishing whether a given set of binary guarded existential rules is such that all conjunctive queries admit a first-order rewriting is decidable, and present a technique for solving this problem. These results have a important practical impact, since they make it possible to identify those sets of binary guarded existential rules for which it is possible to answer every conjunctive query through query rewriting and standard evaluation of a first-order query (actually, a union of conjunctive queries) over a relational database system.

1 Introduction

Ontology-based query answering [1, 13, 6], now a consolidated topic in knowledge representation and database theory, studies the problem of answering expressive queries posed to a knowledge base that represents information both at the intensional and at the extensional level. The knowledge base is interpreted under the open-world assumption, which constitutes the main difference between this form of query answering and the standard query answering over relational databases. *Description Logics* or, more recently, *existential rules* are mostly used as the formalism to express knowledge bases, and *conjunctive queries (CQs)* are the query language usually considered in this setting. In this paper we focus on the framework existential rules, a.k.a. Datalog+/-, which extends Datalog with existential variables in the head of rules [6, 2].

Query rewriting is a well-known approach to ontology-based query answering (see, e.g., [8, 15, 11, 12]). In this approach, the knowledge base is divided into an *intensional* component, called the *ontology* (which in this paper is a set of existential rules), and an *extensional* component, which is usually a relational database instance. A query (CQ) posed to the knowledge base is first rewritten into a new query using the ontology only; the reformulated query constitutes a so-called *perfect* rewriting of the initial query, in the sense that its evaluation over the database produces exactly the *certain answers* to the query, i.e., the answers that hold in all the models of the knowledge base. This modularized strategy has several benefits, especially when the ontology $\mathcal{O}$ and the given query q enjoy the property called *first-order rewritability* (or *FO-rewritability*): this property holds if and only if there exists a *FO-rewriting* of q for $\mathcal{O}$, i.e., a first-order query that constitutes a perfect rewriting of the given query q with respect to the ontology $\mathcal{O}$. FO-rewritability has an important practial implication, since it allows for solving the ontology-based query answering problem through standard evaluation of an SQL query over a relational database [8].

Most of the existing research related to FO-rewritability has tried to identify ontology languages that enjoy such a property: i.e., languages such that the FO-rewritability holds for every ontology $\mathcal{O}$ in this language and for every CQ q [8, 1, 7, 2, 9, 10]. For non-FO-rewritable ontology languages, the work has mostly focused on the identification of the FO-rewritability property for *single* pairs constituted by an ontology $\mathcal{O}$ and a query q. In particular, [3] has studied this problem when $\mathcal{O}$ is expressed in a fragment of existential rules and when q is a conjunctive query; while [5] have considered the case when $\mathcal{O}$ is expressed in Horn Description Logics and q is a query retrieving all the instances of a predicate.

In this paper we study a problem that is in the middle between the FO-rewritability of an ontology language and the FO-rewritability of a a specific query for a specific set of existential rules. We are interested in the problem of deciding whether a given set of existential rules $\mathcal{R}$ is such that *all* CQs over such an ontology admit a FO-rewriting for $\mathcal{R}$. In particular, we consider the class of *binary* and *guarded* existential rules (denoted by *BGER*s) and prove that such a task is decidable. To this aim, we prove the following crucial property: if a set of existential rules $\mathcal{R}$ is such that all *atomic queries* (that is, the conjunctive queries composed of a single atom) admit a FO-rewriting for $\mathcal{R}$, then *all conjunctive queries* admit a FO-rewriting for $\mathcal{R}$. So, to decide the FO-rewritability of all CQs for a set *BGER*s, it suffices to decide the FO-rewritability of atomic queries for such a set of existential rules.

We then use the results of [3], which shows that deciding the FO-rewritability of an atomic query for a set of *BGER*s is decidable: since the number of relevant atomic queries for a finite set of existential rules is finite, this result immediately implies the decidability of the problem of deciding CQ-FO-rewritability of a set of *BGER*s. This result has an important practical impact, since it proves the possibility of identifying those sets of *BGER*s for which it is always possible to answer a conjunctive query through query rewriting and standard evaluation of a first-order query (actually, a union of conjunctive queries) over a relational database system.

2 Deciding CQ-FO-rewritability of BGERs

An *existential rule* ρ over a signature Σ is a first-order logic expression of the form $\forall \bar{\mathbf{x}} \forall \bar{\mathbf{y}}\ \Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \rightarrow \exists \bar{\mathbf{z}}\ \alpha(\bar{\mathbf{x}}, \bar{\mathbf{z}})$ where $\alpha(\bar{\mathbf{x}}, \bar{\mathbf{z}})$ is an atom, called the head of ρ ($head(\rho)$), $\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a conjunction of atoms, called the body of ρ ($body(\rho)$), and $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}$ are sequence of variables. An atom is an expression of the form $R(t_1, \ldots, t_n)$ where R is a predicate (or relation name) in Σ and $t_1, \ldots, t_n$ are called terms. We refer to the number of terms in R as the *arity* of R. We only consider variables as terms.

We will use a simplified notation for existential rules in which we omit the universal quantifiers of the body and we replace the conjunction symbol with a comma (e.g. $\rho : P(x, y), S(y, z) \rightarrow \exists w\ T(x, w)$).

A *binary* existential rule is an existential rule where all the atoms have at most an arity of 2. A *guarded* existential rule is an existential rule such that there exists an atom in its body that contains all the universally quantified variables of the rule; such an atom is called a guard. In this work we focus on *binary guarded existential rules (BGER)*, i.e., sets of existential rules that are both binary and guarded.

Notice that the TGD mentioned above is binary, but not a guarded, while the TGD $\rho : P(x, y), S(y, x) \rightarrow \exists w\ T(x, w))$ is both binary and guarded.

A database over a signature Σ is a set of ground atoms. Given a signature Σ, a set of existential rules $\mathcal{R}$ over Σ, and a database D over Σ, a model for $\langle \mathcal{R}, D \rangle$ is a first-order interpretation that satisfies all formulas in $\mathcal{R} \cup D$.

Concerning queries, we consider *conjunctive queries* (CQs) over a signature Σ, that are represented through expressions of the form $q(\bar{\mathbf{x}}) \leftarrow \exists \bar{\mathbf{y}} \Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, where $\bar{\mathbf{x}}$ are the distinguished variables of the query, $\bar{\mathbf{y}}$ are the existentially quantified variables of the query, and $\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a conjunction of atoms of the form $R(t_1, \ldots, t_n)$, where $R \in \Sigma$ and $t_1, \ldots, t_n$ are variables in $\bar{\mathbf{x}}$ or $\bar{\mathbf{y}}$. A boolean conjunctive query (BCQ) is a CQ with no distinguished variables. We also consider *atomic queries*, i.e., CQs of the above form where $\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is constituted of a single atom.

The certain answers to q over $\langle \mathcal{R}, D \rangle$ (notation $cert(q, \langle \mathcal{R}, D \rangle)$ are all the tuples $\bar{a}$ of constants such that $\mathcal{I} \models q(\bar{a})$ for every interpretation $\mathcal{I}$ of $\langle \mathcal{R}, D \rangle$, where $q(\bar{a})$ is the BCQ obtained by replacing $\bar{\mathbf{x}}$ with $\bar{a}$ in q. If q is a BCQ, the certain answer to q over $\langle \mathcal{R}, D \rangle$ is true if $\mathcal{I} \models q$ for every interpretation $\mathcal{I}$ of $\langle \mathcal{R}, D \rangle$ (notation $\langle \mathcal{R}, D \rangle \models q$), and false otherwise (notation $\langle \mathcal{R}, D \rangle \not\models q$). Then, let q be a CQ, and let q' be a first-order query, we say that q' is a *perfect rewriting* of q w.r.t. a set of existential rules $\mathcal{R}$ if, for each database D, $cert(q, \langle \mathcal{R}, D \rangle) = cert(q', \langle \emptyset, D \rangle)$. Moreover, we say that q is first-order rewritable (or *FO-rewritable*) for $\mathcal{R}$ if there exists a FO query q', such that q' is a perfect rewriting of q w.r.t. $\mathcal{R}$.

Definition 1. *Let $\mathcal{R}$ be a set of BGERs. We say that $\mathcal{R}$ is* CQ-FO-rewritable *if every CQ over $\mathcal{R}$ is FO-rewritable for $\mathcal{R}$. Moreover, we say that $\mathcal{R}$ is* atom-FO-rewritable *if every atomic query over $\mathcal{R}$ is FO-rewritable for $\mathcal{R}$.*

We now present the main result of this paper.

Theorem 1. *Let $\mathcal{R}$ be a set of BGERs. If $\mathcal{R}$ is atom-FO-rewritable then $\mathcal{R}$ is CQ-FO-rewritable.*

To prove the theorem, we make use of a conjunctive query rewriting technique for *BGER*s. Specifically, we make use of the general technique presented in [12] for rewriting conjunctive queries over existential rules, which can also be applied to *BGER*s. Such a technique generates a perfect rewriting in the form of a union of CQs, in particular a set of non-redundant CQs (i.e., no CQ is contained into another CQ): such a set may be finite (which implies that q is FO-rewritable for $\mathcal{R}$) or infinite. We classify the rewriting steps of this technique (that perform a form of resolution between a CQ and an inclusion axiom of the set of existential rules) into two categories: those that involve only "descendants" of a single atom of the initial query, and call such resolution steps *single-ancestor rewriting steps*, and those that involve descendants of at least two atoms of the initial query, and call such resolution steps *multiple-ancestor rewriting steps*.

We are then able to prove the following lemma, which states that, if all atomic queries are FO-rewritable for a set of existential rules $\mathcal{R}$, then the application of single-ancestor rewriting steps only cannot lead to generating an unbounded number of rewritings of a CQ.

Lemma 1. *Let $\mathcal{R}$ be a set of BGERs such that $\mathcal{R}$ is atom-FO-rewritable and let q be a conjunctive query. If the rewriting of query q for $\mathcal{R}$ only applies single-ancestor rewriting steps, then it generates a finite set of CQs.*

The above lemma allows us to prove Theorem 1. Indeed, from such a lemma, it follows that, under the hypothesis that $\mathcal{R}$ is atom-FO-rewritable, if a CQ is not FO-rewritable for $\mathcal{R}$, then it must be possible to perform an infinite sequence of rewriting steps containing infinite multiple-ancestor rewriting steps (and generating non-redundant CQs). However, we show that, due to the restricted form of *BGER*s, this is impossible: roughly, the reason is that every multiple-ancestor step either eliminates a variable occurring in the initial query or generates an isolated subquery (i.e., a subquery not connected by existential variables to the rest of the query).

Example 1. Let $\mathcal{R}$ be the following set of *BGER*s:

$$P(x,y), R(y,x) \rightarrow \exists z\ S(y,z)$$
$$R(y,x) \rightarrow P(x,y)$$
$$S(x,y), S(y,x) \rightarrow \exists z\ R(x,z)$$

$\mathcal{R}$ is atom-FO-rewritable, since it can be verified that all the atomic queries over the signature of $\mathcal{R}$ are FO-rewritable. Thus, by Theorem 1, $\mathcal{R}$ is also CQ-FO-rewritable.

As an example, we provide a perfect rewriting of the query $q() \leftarrow S(x,y)$:

$$q() \leftarrow S(x,y) \qquad q() \leftarrow R(z_0,x)$$
$$q() \leftarrow P(x,z_0), R(z_0,x) \qquad q() \leftarrow S(z_0,z_1), S(z_1,z_0)$$

In [3], it has been proved that the FO-rewritability of an answer-guarded conjunctive query over a set of *BGER*s is decidable. In particular, the following property directly follows from Theorem 11 in [3]:

Proposition 1. *For every set $\mathcal{R}$ of BGERs, and for every atomic conjunctive query q, one can effectively find a GN-Datalog program that computes the certain answers to q.*

GN-Datalog programs are a subclass of Datalog programs that enjoys the following property:

Proposition 2 ([4], Corollary 8.9). *For GN-Datalog programs, boundedness over finite instances is decidable and coincides with boundedness over unrestricted instances.*

From the above propositions, and from Theorem 1, it is possible to derive the following technique for deciding the FO-rewritability of a set of *BGER*s $\mathcal{R}$ over a finite signature Σ:

1. compute the set $\mathcal{Q}$ of all possible atomic queries over Σ;
2. for each atomic query $q \in \mathcal{Q}$, find the GN-Datalog program P that computes the certain answers to q over $\mathcal{R}$ and add it to $\mathcal{P}_{\mathcal{R}}$;

3. if there exists an unbounded program $P \in \mathcal{P}_{\mathcal{R}}$, then return false, otherwise return true.

Based on the above technique, we are able to show the following property.

Theorem 2. *Let $\mathcal{R}$ be a set of BGERs. Establishing the CQ-FO-rewritability of $\mathcal{R}$ is decidable.*

3 Conclusions

The work that is the closest to the present one is [14], which shows a property analogous to Theorem 1 for the description logic $\mathcal{ELI}$, which corresponds to a subclass of binary guarded existential rules.

We are currently working at extending the results presented in this paper to more expressive existential rules. Also, we would like to study the possibility of optimizing the technique for deciding the FO-rewritability of atomic queries for the case of *BGER*s.

Acknowledgments. This research has been partially supported by the EU under FP7 project Optique (grant n. FP7-318338).

References

1. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research*, 36:1–69, 2009.
2. J.-F. Baget, M. Leclère, M.-L. Mugnier, and E. Salvat. On rules with existential variables: Walking the decidability line. *Artificial Intelligence*, 175(9–10):1620–1654, 2011.
3. V. Bárány, M. Benedikt, and B. ten Cate. Rewriting guarded negation queries. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013*, pages 98–110, 2013.
4. V. Bárány, B. ten Cate, and M. Otto. Queries with guarded negation. *Proc. of the 38th Int. Conf. on Very Large Data Bases (VLDB 2012)*, 5(11):1328–1339, July 2012.
5. M. Bienvenu, C. Lutz, and F. Wolter. First-order rewritability of atomic queries in horn description logics. In *Proc. of the 23st Int. Joint Conf. on Artificial Intelligence (IJCAI 2013)*, pages 754–760. AAAI Press, 2013.
6. A. Calì, G. Gottlob, and T. Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Semantic Web J.*, 14:57–83, 2012.
7. A. Calì, G. Gottlob, and A. Pieris. Towards more expressive ontology languages: The query answering problem. *Artificial Intelligence*, 193:87–128, 2012.
8. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
9. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Data complexity of query answering in description logics. *Artificial Intelligence*, 195:335–360, 2013.
10. C. Civili and R. Rosati. A broad class of first-order rewritable tuple-generating dependencies. In *Proc. of the 2nd Datalog 2.0 Workshop*, 2012.
11. G. Gottlob, G. Orsi, and A. Pieris. Ontological queries: Rewriting and optimization. In *Proc. of the 27th IEEE Int. Conf. on Data Engineering (ICDE 2011)*, pages 2–13, 2011.

12. M. König, M. Leclere, M.-L. Mugnier, and M. Thomazo. Sound, complete and minimal ucq-rewriting for existential rules. *Semantic Web J.*, 2013.
13. R. Kontchakov, C. Lutz, D. Toman, F. Wolter, and M. Zakharyaschev. The combined approach to query answering in *DL-Lite*. In *Proc. of the 12th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 247–257, 2010.
14. C. Lutz and F. Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of the 24th Int. Workshop on Description Logic (DL 2011)*, 2011.
15. H. Pérez-Urbina, B. Motik, and I. Horrocks. Tractable query answering and rewriting under description logic constraints. *J. Applied Logic*, 8(2):186–209, 2010.