Bounded Implication for Existential Rules (Extended Abstract)

Cristina Civili[1] and Riccardo Rosati[2]

[1] School of Informatics, University of Edinburgh, UK

[2] DIAG, Sapienza Università di Roma, Italy

The problem This paper deals with the property of *boundedness* in rule languages. Boundedness is an important notion that formalizes the fact that a rule set Σ can be unfolded into a finite set Σ' of acyclic (i.e., non-recursive) rules such that Σ and Σ' are equivalent on every database: it is therefore a crucial property for optimizing the processing of rules. Such a property has been extensively studied, especially for the Datalog rule language [9,4], and, recently, for Answer Set Programming [16].

In Datalog, the (uniform) boundedness of a program P can be defined as the existence of an integer k such that, for every database D, the number of iterated applications (in a forward chaining manner) of P to D that are necessary to compute the minimal model of P and D is bounded by k. This definition of boundedness is equivalent to the existence of a finite, non-recursive program that is equivalent to P. Also, it is well-known that a Datalog query is bounded if and only if it is equivalent to a first-order sentence [1,12].

More recently, rule-based languages have been used in the context of *ontology-based data access* [15]. In this framework, the main focus is on the problem of answering conjunctive queries over an ontology expressed by a set of rules, and one of the most studied properties is the *first-order rewritability of conjunctive queries (CQFO-rewritability)* over an ontology, which corresponds to the above mentioned first-order expressibility in Datalog: an ontology $\mathcal{O}$ is CQFO-rewritable if every conjunctive query q over the ontology can be equivalently rewritten into a first-order query q', i.e., q' is such that, for every database instance $\mathcal{D}$, the evaluation of q over $\mathcal{O}$ and $\mathcal{D}$ coincides with the evaluation of q' over $\mathcal{D}$. Notably, in the case when the ontology is expressed as a set P of Datalog rules, the CQFO-rewritability of P and the boundedness of P are equivalent properties.

Existential rules, which extend Datalog rules to the presence of existentially quantified variables and multiple atoms in rule heads, have been proposed and studied in the last years as a specification language for ontology-based data access [5,2,14]. An existential rule (or simply rule) σ over a relational schema $\mathbf{S}$ is an expression of the form $\forall\mathbf{x}\forall\mathbf{y}(\Phi(\mathbf{x},\mathbf{y},\bar{a}) \rightarrow \exists\mathbf{z}\Psi(\mathbf{x},\mathbf{z},\bar{b}))$, where $\Phi(\mathbf{x},\mathbf{y},\bar{a})$ (the body of σ) and $\Psi(\mathbf{x},\mathbf{z},\bar{b})$ (the head of σ) are conjunctions of atoms over $\mathbf{S}$. We call $\mathbf{x}$ the *frontier variables* ($\mathcal{F}(\sigma)$), and $\mathbf{z}$ the *existential head variables* of σ ($\mathcal{E}_H(\sigma)$), while $\bar{a},\bar{b}$ are the constants occurring in σ (*Const*(σ)). Several recent studies have focused on the first-order rewritability property for existential rules (e.g., [7,2,8]). On the other hand, the notion of boundedness for existential rules has not been deeply investigated. To our knowledge, one of the most relevant recent approaches to this problem is presented in [2], where the notion of *acyclic graph of rule dependencies (aGRD)* is defined, which corresponds to a form of boundedness for existential rules. Instead, we start our study from a notion of boundedness for existential rules that generalizes the definition of boundedness provided for Datalog to existential rules in a simpler way: we call such a notion *strict boundedness*. More precisely, we say that a set Σ of existential rules is strictly bounded if it is logically equivalent to a finite and acyclic set of rules.

However, it can be immediately verified that, for arbitrary sets of existential rules, the above notion of boundedness is much stronger than the first-order rewritability of conjunctive queries. That is, while strict boundedness of a rule set implies its CQFO-rewritability, there exist rule sets that are CQFO-rewritable but are not strictly bounded. Notice that the same property holds for the above mentioned notion of aGRD.

The main goal of this paper is to answer the following question: *is it possible to generalize the notion of boundedness for Datalog to existential rules, in such a way that the correspondence with the notion of first-order rewritability of conjunctive queries is preserved?* Actually, from the forward chaining perspective, such a generalization has been provided by the *bounded derivation depth property* of the *chase* of existential rules [5,10]. However, we would like to characterize this property in terms of equivalent representations of the set of rules, and see how the alternative notion of boundedness as existence of a finite and non-recursive equivalent rule set has to be extended to capture first-order rewritable rule sets.

Our contribution Our approach to the study of boundedness for existential rules is inspired by the work in query rewriting for existential rules [2,13,6,11]. In particular, we extend the techniques presented in [2,13] to address the problem of computing an unfolding of a set of existential rules, and the problem of defining an appropriate notion of redundancy between rules.

First, we define a notion of boundedness for existential rules that weakens strict boundedness by giving up the acyclicity condition: we call such a notion *weak implication-boundedness*. It is based on the idea of looking for a finite representation of all the *single-head* rules (i.e., rules with one atom in the head) that are logically implied by the initial rule set, and on a notion of equivalence between rule sets, called *R-equivalence*, that is different from (and stronger than) the standard logical equivalence. R-equivalence is based on a notion of *redundancy* between two rules.

Given two rules $\sigma : \Phi(\mathbf{x}, \mathbf{y}, \bar{a}) \rightarrow \exists \mathbf{z}\ \Psi(\mathbf{x}, \mathbf{z}, \bar{b})$, and $\sigma' : \Phi'(\mathbf{x}', \mathbf{y}', \bar{c}) \rightarrow \exists \mathbf{z}'\ \Psi(\mathbf{x}', \mathbf{z}', \bar{d})$, we say that σ is redundant with respect to σ' if there exists a specialization of σ' $\eta(\sigma') = \sigma'_s$ (where $\eta : \mathcal{F}(\sigma') \rightarrow \mathcal{F}(\sigma') \cup \mathit{Const}(\sigma)$) and a bijective function $\epsilon : \mathcal{F}(\sigma'_s) \rightarrow \mathcal{F}(\sigma)$ such that the following first-order sentences are valid:

$$\forall \mathbf{x} \forall \mathbf{y}\ body(\sigma)(\mathbf{x}, \mathbf{y}, \bar{a}) \rightarrow \exists \mathbf{y}'\ \epsilon(body(\sigma'_s))(\mathbf{x}, \mathbf{y}', \bar{e})$$

$$\forall \mathbf{x}' \forall \mathbf{z}'\ head(\sigma'_s)(\mathbf{x}', \mathbf{z}', \bar{d}) \rightarrow \exists \mathbf{z}\ \epsilon^{-}(head(\sigma))(\mathbf{x}', \mathbf{z}, \bar{f})$$

where $\bar{e} = \bar{c} \cup \bar{a} \cup \bar{b}$ and $\bar{f} = \bar{d} \cup \bar{a} \cup \bar{b}$.

Given two sets of rules Σ, Σ', we say that Σ' *R-entails* Σ if, for each non-tautological rule $\sigma \in \Sigma$ there exists a rule $\sigma' \in \Sigma'$ such that σ is redundant w.r.t. σ'. Moreover, we say that Σ and Σ' are *R-equivalent* if both Σ' R-entails Σ and Σ R-entails Σ'.

Let Σ be a set of rules over a signature $\mathbf{S}$. We define the *SH-closure* of Σ as the set $\Sigma^{\star s} = \{\sigma \mid \sigma$ is a single-head rule over $\mathbf{S}$ and $\mathit{Const}(\Sigma)$, and $\Sigma \models \sigma\}$.

We say that a set Σ of rules is *weakly implication-bounded* if $\Sigma^{\star s}$ is R-equivalent to a finite set of rules.

It is immediate to verify that strict boundedness always implies weak implication-boundedness, while the converse does not always hold. Furthermore, it turns out that weak implication-boundedness is not equivalent to CQFO-rewritability. More precisely, it is possible to show that weak implication-boundedness does not imply CQFO-rewritability for single-head ternary set of rules (i.e., rules over relations of arity ≤ 3). Similarly, it can be shown that weak implication-boundedness does not imply CQFO-rewritability for binary set of rules (i.e., over relations of arity ≤ 2). On the other hand, we are able to prove the correspondence between weak implication-boundedness and the notion of *AFO-rewritability*, that is, first-order rewritability of all atomic queries (i.e., conjunctive queries consisting of a single atom).

To arrive at a notion of boundedness that corresponds to CQFO-rewritability, we define a second notion of *strong implication-boundedness* for existential rules. Roughly speaking, such a notion is obtained from weak implication-boundedness by discarding the restriction to single-head rules in the deductive closure of the rule set, and by considering projections of such a deductive closure.

Let Σ be a set of rules over a signature $\mathbf{S}$. We call *closure of* Σ the set $\Sigma^\star = \{\sigma \mid \sigma \text{ is a rule over } \mathbf{S} \text{ and } \mathit{Const}(\Sigma), \text{ and } \Sigma \models \sigma\}$. Moreover, let σ, σ' be two rules. We say that σ' is *head-unifiable* w.r.t. σ if there exists a homomorphism $\mu : \mathcal{F}(\sigma) \to \mathcal{F}(\sigma') \cup \mathit{Const}(\sigma')$ and an isomorphism $\epsilon : \mathcal{E}_H(\sigma) \to \mathcal{E}_H(\sigma')$ such that $head(\mu(\epsilon(\sigma))) = head(\sigma')$. Then, we call *projection* of a rule set Σ with respect to a rule σ the set $\Pi_\sigma(\Sigma) = \{\sigma' \in \Sigma \mid \sigma' \text{ is head-unifiable with } \sigma\}$.

We say that a set Σ of rules over a schema $\mathbf{S}$ is *strongly implication-bounded* if, for each rule σ over $\mathbf{S}$, $\Pi_\sigma(\Sigma^\star)$ is R-equivalent to a finite set of rules.

The notion of weak implication-boundedness has the desired correspondence with the CQFO-rewritability. Namely, every set of rules Σ is strongly implication-bounded if and only if Σ is CQFO-rewritable. This in turn implies the correspondence between strong boundedness and the bounded derivation depth property [10].

Moreover, the equivalence between weak and strong implication-boundedness actually extends to two broad classes of existential rules: *single-head binary* rules, that is, single-head rules over relations of arity not greater than 2, and *frontier-guarded* [2] rules.

We believe that the equivalence between weak and strong implication-boundedness is a very important property for a set of existential rules. In particular, the above correspondence could be exploited in the optimization of query answering over ontologies expressed by rule sets belonging to the above classes.

Finally, it is possible to show that checking strong (or, equivalently, weak) implication-boundedness is undecidable for single-head binary rules, and (using results from [3]) decidable for frontier-guarded rules. These results complement the well-known undecidability of (strict) boundedness for Datalog [9].

Acknowledgments This research is partially supported by the UK EPSRC Project VADA (grant n. M025268) and the EU Project Optique (grant n. FP7-318338).

References

1. Miklós Ajtai and Yuri Gurevich. Datalog vs first-order logic. *J. Comput. Syst. Sci.*, 49(3):562–588, 1994.
2. Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artificial Intelligence*, 175(9–10):1620–1654, 2011.
3. Vince Bárány, Michael Benedikt, and Balder ten Cate. Rewriting guarded negation queries. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013*, pages 98–110, 2013.
4. Michael Benedikt, Balder ten Cate, Thomas Colcombet, and Michael Vanden Boom. The complexity of boundedness for guarded logics. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, Kyoto, Japan, July 6-10, 2015*, pages 293–304, 2015.
5. Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Semantic Web J.*, 14:57–83, 2012.
6. Andrea Calì, Georg Gottlob, and Andreas Pieris. Advanced processing for ontological queries. *Proc. of the VLDB Endowment*, 3(1):554–565, 2010.
7. Andrea Calì, Georg Gottlob, and Andreas Pieris. Towards more expressive ontology languages: The query answering problem. *Artificial Intelligence*, 193:87–128, 2012.
8. Cristina Civili and Riccardo Rosati. A broad class of first-order rewritable tuple-generating dependencies. In *Proc. of the 2nd Datalog 2.0 Workshop*, 2012.
9. Haim Gaifman, Harry G. Mairson, Yehoshua Sagiv, and Moshe Y. Vardi. Undecidable optimization problems for database logic programs. *J. of the ACM*, 40(3):683–713, 1993.
10. Georg Gottlob, Stanislav Kikot, Roman Kontchakov, Vladimir V. Podolskii, Thomas Schwentick, and Michael Zakharyaschev. The price of query rewriting in ontology-based data access. *Artif. Intell.*, 213:42–59, 2014.
11. Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proc. of the 27th IEEE Int. Conf. on Data Engineering (ICDE 2011)*, pages 2–13, 2011.
12. Phokion G. Kolaitis and Christos H. Papadimitriou. Some computational aspects of circumscription. *J. ACM*, 37(1):1–14, 1990.
13. Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. Sound, complete and minimal ucq-rewriting for existential rules. *Semantic Web*, 6(5):451–475, 2015.
14. Markus Krötzsch and Sebastian Rudolph. Extending decidable existential rules by joining acyclicity and guardedness. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI 2011)*, pages 963–968, 2011.
15. Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.
16. Heng Zhang and Yan Zhang. First-order expressibility and boundedness of disjunctive logic programs. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.