A Single Approach to Decide Chase Termination on Linear Existential Rules

Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo, Federico Ulliana

A Single Approach to Decide Chase Termination on Linear Existential Rules

Michel Leclère[1], Marie-Laure Mugnier[1], Michaël Thomazo[2], Federico Ulliana[1]

[1] Univ. Montpellier, LIRMM, Inria, CNRS
leclere@lirmm.fr, mugnier@lirmm.fr, ulliana@lirmm.fr
[2] Inria, DI ENS, ENS, CNRS, PSL University
michael.thomazo@inria.fr

The chase procedure is a fundamental tool for solving many issues involving tuple-generating dependencies, such as data integration, data-exchange, query answering using views or query answering on probabilistic databases. In the last decade, tuple-generating dependencies raised a renewed interest under the name of *existential rules* for the ontology-mediated query answering (OMQA) problem. In this context, the aim is to query a knowledge base $(I, \mathcal{R})$, where I is an instance (or factbase) and $\mathcal{R}$ is a set of existential rules (see e.g. the survey chapters [CGL09,MT14]). A fundamental property of the chase is that it allows one to compute a (possibly infinite) *universal model* of $(I, \mathcal{R})$, i.e., a model that can be homomorphically mapped to any other model of $(I, \mathcal{R})$. Hence, the answers to a conjunctive query (and more generally to any kind of query closed by homomorphism) over $(I, \mathcal{R})$ can be defined by considering solely this universal model.

The chase starts from an instance and exhaustively performs a sequence of rule applications with respect to a redundancy criterion, which differs according to the considered chase variant. We focus in this paper on the main variants, namely: *semi-oblivious* [Mar09] (aka skolem [Mar09]), *restricted* [BV81,FKMP05] (aka standard [One12]) and *core* [DNR08]. All these produce homomorphically equivalent results but terminate for increasingly larger subclasses of existential rules. The question of whether a chase variant terminates on *all instances* for a given set of existential rules is known to be undecidable when there is no restriction on the kind of rules [BLMS11,GM14]. A number of *sufficient* syntactic conditions for termination have been proposed in the literature for the semi-oblivious chase (see e.g. [One12,GHK$^+$13,Roc16] for syntheses), as well as for the restricted chase [CDK17] (note that the latter paper also defines a sufficient condition for non-termination). However, only few positive results exist regarding the termination of the chase on specific classes of rules. Decidability was shown for the semi-oblivious chase on guarded-based rules (linear rules, and their extension to (weakly-)guarded rules) [CGP15]. Decidability of the core chase termination on guarded rules for a fixed instance was shown in [Her12].

In this work, we provide new insights on the chase termination problem for *linear* existential rules, which are precisely of the form $\forall \mathbf{x} \forall \mathbf{y}.[\alpha_1(\mathbf{x}, \mathbf{y}) \rightarrow \exists \mathbf{z}.\alpha_2(\mathbf{x}, \mathbf{z})]$, where α_i is an atom and $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are pairwise disjoint tuples of variables. Linear rules form a simple yet important subclass of guarded existential rules, which generalizes inclusion dependencies [Fag81] and positive inclusions in DL-Lite$_\mathcal{R}$ [CDL$^+$07] (which can be seen as inclusion dependencies restricted to unary and binary predicates). Concerning the ontology-mediated query answering problem, we note that linear rules are first-order rewritable, hence OMQA on conjunctive queries can be solved by query rewriting. However, it is well known that the size and the unusual form of the rewritten query may give rise to practical efficiency issues. The materialization of ontological inferences in the data is often a good alternative to query rewriting, provided that some chase algorithm terminates. Finally, having the choice of how to process a set of linear rules may extend the applicability of query answering techniques that combine query rewriting and materialization [BLMS11].

The question of whether a chase variant terminates on all instances for a set of linear existential rules can be asked under two forms: Does *every* (fair) chase sequence terminate? Does *some* (fair) chase sequence terminate? It is well-known that these two questions have the same answer for the semi-oblivious and the core chase variants, but not for the restricted chase. Indeed, this last one may admit both terminating and non-terminating sequences over the same knowledge base. We show that the termination problem is decidable for linear existential rules, whether we consider *any version* of the problem and *any chase variant*.

We study chase termination by exploiting in a novel way a graph structure, namely the *derivation tree*, which was originally introduced to solve the ontology-mediated (conjunctive) query answering problem for the family of greedy-bounded treewidth sets of existential rules [BMRT11,Tho13], a class of rules that generalizes guarded-based rules and in particular linear rules. We first use derivation trees to show the decidability of the termination problem for the semi-oblivious and restricted chase variants, and then generalize them to *entailment trees* to show the decidability of termination for the core chase. For any chase variant we consider, we adopt the same high-level procedure: starting from a finite set of canonical instances (representative of all possible instances), we build a (set of) tree structures for each canonical instance, while forbidding the occurrence of a specific pattern, we call *unbounded-path witness*. The built structures are finite thanks to this forbidden pattern, and this allows us to decide if the chase terminates on the associated canonical instance. By doing so, we obtain a uniform approach to study the termination of several chase variants, which we believe to be of theoretical interest per se. The derivation tree is moreover a simple structure and the algorithms built on this notion are likely to lead to an effective implementation.

Besides providing new theoretical tools to study chase termination, we obtain the following results for linear existential rules:

- a new proof of the decidability of the semi-oblivious chase termination, building on different objects than the previous proof provided in [CGP15]; we show that our algorithm provides the same complexity upper-bound;
- the decidability of the restricted chase termination, for both versions of the problem, i.e., termination of all (fair) chase sequences and termination of some (fair) chase sequence; to the best of our knowledge, these are the first positive results on the decidability of the restricted chase termination;
- a new proof of the decidability of the core chase termination, with different objects than previous work on the core chase termination reported in [Her12].

The full paper is available as a technical report [LMTU18].

References

[BLMS11] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.

[BMRT11] Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, and Michaël Thomazo. Walking the complexity lines for generalized guarded existential rules. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 712–717. IJCAI/AAAI, 2011.

[BV81] Catriel Beeri and Moshe Y. Vardi. The implication problem for data dependencies. In Shimon Even and Oded Kariv, editors, *Automata, Languages and Programming, 8th Colloquium, Acre (Akko), Israel, July 13-17, 1981, Proceedings*, volume 115 of *Lecture Notes in Computer Science*, pages 73–85. Springer, 1981.

[CDK17] David Carral, Irina Dragoste, and Markus Krötzsch. Detecting chase (non)termination for existential rules with disjunctions. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 922–928. ijcai.org, 2017.

[CDL+07] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. Autom. Reasoning*, 39(3):385–429, 2007.

[CGL09] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. Datalog extensions for tractable query answering over ontologies. In Roberto De Virgilio, Fausto Giunchiglia, and Letizia Tanca, editors, *Semantic Web Information Management - A Model-Based Perspective*, pages 249–279. Springer, 2009.

[CGP15] Marco Calautti, Georg Gottlob, and Andreas Pieris. Chase termination for guarded existential rules. In Tova Milo and Diego Calvanese, editors, *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 91–103. ACM, 2015.

[DNR08] Alin Deutsch, Alan Nash, and Jeffrey B. Remmel. The chase revisited. In Maurizio Lenzerini and Domenico Lembo, editors, *Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2008, June 9-11, 2008, Vancouver, BC, Canada*, pages 149–158. ACM, 2008.

[Fag81] Ronald Fagin. A normal form for relational databases that is based on domians and keys. *ACM Trans. Database Syst.*, 6(3):387–415, 1981.

[FKMP05] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[GHK+13] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. Acyclicity notions for existential rules and their application to query answering in ontologies. *J. Artif. Intell. Res.*, 47:741–808, 2013.

[GM14] Tomasz Gogacz and Jerzy Marcinkowski. All-instances termination of chase is undecidable. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, volume 8573 of *Lecture Notes in Computer Science*, pages 293–304. Springer, 2014.

[Her12] André Hernich. Computing universal models under guarded tgds. In Alin Deutsch, editor, *15th International Conference on Database Theory, ICDT '12, Berlin, Germany, March 26-29, 2012*, pages 222–235. ACM, 2012.

[LMTU18] Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo, and Federico Ulliana. A single approach to decide chase termination on linear existential rules. *CoRR*, abs/1602.05828, 2018.

[Mar09] Bruno Marnette. Generalized schema-mappings: from termination to tractability. In Jan Paredaens and Jianwen Su, editors, *Proceedings of the Twenty-Eigth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2009, June 19 - July 1, 2009, Providence, Rhode Island, USA*, pages 13–22. ACM, 2009.

[MT14] Marie-Laure Mugnier and Michaël Thomazo. An introduction to ontology-based query answering with existential rules. In *Reasoning Web. Reasoning on the Web in the Big Data Era - 10th International Summer School 2014, Athens, Greece, September 8-13, 2014. Proceedings*, pages 245–278, 2014.

[One12] Adrian Onet. *The chase procedure and its applications*. PhD thesis, Concordia University, Canada, 2012.

[Roc16] Swan Rocher. *Querying Existential Rule Knowledge Bases: Decidability and Complexity. (Interrogation de Bases de Connaissances avec Règles Existentielles : Décidabilité et Complexité)*. PhD thesis, University of Montpellier, France, 2016.

[Tho13] Michaël Thomazo. *Conjunctive Query Answering Under Existential Rules - Decidability, Complexity, and Algorithms*. PhD thesis, Montpellier 2 University, France, 2013.