Complexity Landscape for Counting Queries

Extended Abstract

Meghyn Bienvenu[1], Quentin Manière[1] and Michaël Thomazo[2]

[1]CNRS, University of Bordeaux, Bordeaux INP, LaBRI, Talence, France
[2]Inria, DI ENS, ENS, CNRS, University PSL, Paris, France

Abstract

We summarize our recent work [1] on extending the study of counting queries to Horn description logics outside the DL-Lite family.

Keywords

Ontology-mediated query answering, counting queries

1. Introduction

Ontology-mediated query answering (OMQA) facilitates access to data through the use of ontologies, which provide a convenient vocabulary for query formulation and capture domain knowledge that can be exploited to obtain more complete query results. The OMQA approach has been extensively studied over the past fifteen years [2, 3, 4], leading to the identification of ontology languages that are well suited to OMQA due to their attractive computational properties. Particular attention has been paid to Horn description logics of the DL-Lite and $\mathcal{EL}$ families [5, 6].

While most work on OMQA considers that the user query is a conjunctive query (CQ), there has been significant interest in exploring the possibility of adopting more expressive query languages for OMQA. In particular, several works have investigated ways of equipping CQs with some form of counting [7, 8, 9]. A recent approach, proposed in [10] as a generalization of [8], considers *counting conjunctive queries (CCQs)* that are syntactically defined like standard CQs except that some variables may be designated as *counting variables*. In each model of the knowledge base, we can count the number of possible assignments to the counting variables that make the query answer hold. As the count value may differ between models, the goal is to identify intervals that provide upper and lower bounds on the count values across all models.

The problem of answering CCQs is intractable, in both data and combined complexity, for common DL-Lite dialects such as DL-Lite$_{\mathsf{core}}$ and DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$[8]. Recent works have shown that intractability arises even for simple forms of CCQs [11, 12]. However, some interesting tractable cases have also been identified, notably, rooted CCQs [10, 11, 13] and cardinality queries [12]

DL 2022: 35th International Workshop on Description Logics, August 7–10, 2022, Haifa, Israel
meghyn.bienvenu@u-bordeaux.fr (M. Bienvenu); quentin.maniere@u-bordeaux.fr (Q. Manière); michael.thomazo@inria.fr (M. Thomazo)
0000-0001-6229-8103 (M. Bienvenu); 0000-0001-9618-8359 (Q. Manière); 0000-0002-1437-6389 (M. Thomazo)

	DL-Lite$_{\mathsf{pos}}$	DL-Lite$_{\mathsf{core}}$	DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$	DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$	$\mathcal{EL}(\mathcal{H}_\bot), \mathcal{EL}(\mathcal{HI})$	$\mathcal{EL}(\mathcal{H})\mathcal{I}_\bot$
Concept	NL	coNP	NL	coNP⸸	EXP	coNEXP
Role	NL	coNP	coNP⸸	coNP⸸	EXP	coNEXP
CCQ	coNEXP†	coNEXP†	2EXP	2EXP	2EXP	2EXP

Table 1
Combined complexity results for CCQs and cardinality queries, all bounds are tight. †/ ⸸: previously known upper / lower bound.

coupled with DL-Lite$_{\mathsf{core}}$ ontologies. Query rewriting techniques have also begun to be explored [14]. However, despite these advances, we still have only a partial understanding of CCQ answering in common DL-Lite dialects, and the precise combined complexity has remained elusive: the current bounds for DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ are between coNEXP and coN2EXP [8]. Moreover, to the best of our knowledge, CCQ answering has not yet been studied for DLs outside the DL-Lite family.

2. Contributions

In [1], we extend the study of CCQ answering to other well-known Horn description logics, such as $\mathcal{EL}$ and the more expressive $\mathcal{ELHI}_\bot$. The techniques used in the DL-Lite context do not readily transfer to $\mathcal{EL}$ due to the presence of conjunction, and in any case, our results show that they do not achieve the optimal combined complexity even for DL-Lite. We therefore develop a new approach based upon the observation that there exists a model minimizing the count value that consists of an arbitrary structure $\mathcal{I}^*$ containing all assignments for the counting variables, augmented with structures that are tree-shaped, provided we ignore edges to and from $\mathcal{I}^*$. Importantly, we can bound the size of the central component $\mathcal{I}^*$, which enables us to explore all possible options for $\mathcal{I}^*$. Checking whether a given $\mathcal{I}^*$ can be extended to a model preserving the minimum count value can be done by specifying a set of *patterns* (intuitively representing a pair of adjacent elements), and testing via local consistency conditions whether they can be coherently assembled. This latter step takes inspiration from a CQ answering technique for existential rules [15], and is also similar in spirit to type-elimination style procedures, which have been employed for reasoning with expressive DLs, see e.g. [16, 17].

Using this new approach, we are able to establish a 2EXP upper bound in combined complexity for $\mathcal{ELHI}_\bot$. A matching lower bound, which applies to both $\mathcal{EL}$ and DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$, is obtained by establishing a novel connection between CCQ answering and OMQA with closed predicates. This yields 2EXP-completeness for a wide range of Horn DLs and closes the combined complexity gap for CCQ answering in DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$. We further prove a coNEXP lower bound for DL-Lite$_{\mathsf{pos}}$, which matches an existing coNEXP upper bound, yielding the precise combined complexity for DL-Lite$_{\mathsf{core}}$ as well. We also explore how to shrink the size of the models implicitly generated by our procedure, producing models with bounded size which we use to show that CCQ answering is coNP-complete in data complexity for all logics between $\mathcal{EL}$ and $\mathcal{ELHI}_\bot$.

In addition to CCQs, we also investigate the special case of cardinality queries, which correspond to Boolean atomic CCQs and allow us to ask for (bounds on) the number of members of a given concept or role. We obtain a complete picture of data and combined complexity of answering cardinality queries in $\mathcal{ELHI}_\bot$ and its various sublogics. While the data complexity is coNP-complete for all considered logics, the combined complexity ranges from NL or coNP in DL-Lite logics to EXP or coNEXP for $\mathcal{EL}$ and its extensions. We achieve these results using a variety of techniques: refinements of our approach for general CCQs, adaptations of existing constructions, and further reductions involving closed predicates. Table 1 summarizes the combined complexity results for both CCQs and cardinality queries.

3. Perspectives

We have extended the study of CCQ answering to Horn DLs outside the DL-Lite family, establishing a complete picture of the combined and data complexity of the problems of answering CCQs and cardinality queries in $\mathcal{ELHI}_\bot$ and its various sublogics. Interestingly, the new techniques we devised not only allowed us to close some open questions concerning the combined complexity of CCQ answering in DL-Lite, but also extend to non-Horn DLs: the 2EXP procedure can be adapted to handle $\mathcal{ALCHI}$ KBs as our investigations, not presented in [1], have shown.

Going forward, the main challenge is to develop practical algorithms. A first direction is to look for restrictions on the query or ontology that ensure polynomial data complexity for logics of the $\mathcal{EL}$ family. Unfortunately, our results on cardinality queries show that restrictions as have been considered for DL-Lite [11, 10, 12] are not sufficient to obtain tractability in $\mathcal{EL}$, so novel restrictions need to be identified. Second, it would be desirable, for $\mathcal{EL}$ but also for DL-Lite, to develop more refined coNP procedures that are amenable to implementation using SAT solvers. We believe that our improved understanding of the structure of optimal models will prove helpful for both of these research directions.

Acknowledgments

Partially supported by ANR project CQFD (ANR-18-CE23-0003).

References

[1] M. Bienvenu, Q. Manière, M. Thomazo, Counting queries over $\mathcal{ELHI}_\bot$ ontologies, in: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2022, p. (To appear).

[2] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, Journal of Data Semantics (JoDS) 10 (2008) 133–173.

[3] M. Bienvenu, M. Ortiz, Ontology-mediated query answering with data-tractable description logics, in: Tutorial Lectures of the 11th Reasoning Web International Summer School (RW), 2015, pp. 218–307.

[4] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: a survey, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI), 2018, pp. 5511–5519.

[5] D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Tractable reasoning and efficient query answering in description logics: The DL-Lite family, Journal of Automated Reasoning (JAR) 39 (2007) 385–429.

[6] F. Baader, S. Brandt, C. Lutz, Pushing the $\mathcal{EL}$ envelope, in: Proceedings of the 19th international joint conference on Artificial intelligence (IJCAI), 2005, pp. 364–369.

[7] D. Calvanese, E. Kharlamov, W. Nutt, C. Thorne, Aggregate queries over ontologies, in: Proceedings of the 2nd International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW), 2008, pp. 97–104.

[8] E. V. Kostylev, J. L. Reutter, Complexity of answering counting aggregate queries over DL-Lite, Journal of Web Semantics (JWS) (2015) 94–111.

[9] C. Feier, C. Lutz, M. Przybylko, Answer counting under guarded TGDs, in: Proceedings of the 24th International Conference on Database Theory (ICDT), 2021.

[10] M. Bienvenu, Q. Manière, M. Thomazo, Answering counting queries over DL-Lite ontologies, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, pp. 1608–1614.

[11] D. Calvanese, J. Corman, D. Lanti, S. Razniewski, Counting query answers over a DL-Lite knowledge base, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, pp. 1658–1666.

[12] M. Bienvenu, Q. Manière, M. Thomazo, Cardinality queries over DL-Lite ontologies, in: Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI), 2021, pp. 1801–1807.

[13] C. Nikolaou, E. V. Kostylev, G. Konstantinidis, M. Kaminski, B. C. Grau, I. Horrocks, Foundations of ontology-based data access under bag semantics, Journal of Artificial Intelligence (AIJ) (2019) 91 – 132.

[14] D. Calvanese, J. Corman, D. Lanti, S. Razniewski, Rewriting count queries over DL-Lite TBoxes with number restrictions, in: Proceedings of the 33rd International Workshop on Description Logics (DL), 2020.

[15] M. Thomazo, J. Baget, M. Mugnier, S. Rudolph, A generic querying algorithm for greedy sets of existential rules, in: Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2012.

[16] S. Rudolph, M. Krötzsch, P. Hitzler, Type-elimination-based reasoning for the description logic $\mathcal{SHIQ}b_s$ using decision diagrams and disjunctive datalog, Logical Methods in Computer Science (LMCS) 8 (2012).

[17] T. Eiter, C. Lutz, M. Ortiz, M. Simkus, Query answering in description logics: the knots approach, in: Proceedings of the 16th International Workshop on Logic, Language, Information and Computation (WoLLIC), 2009, pp. 26–36.