Complexity Landscape for Counting Queries
Meghyn Bienvenu, Quentin Manière, Michaël Thomazo

To cite this version:
Meghyn Bienvenu, Quentin Manière, Michaël Thomazo. Complexity Landscape for Counting Queries. 35th International Workshop on Description Logics, Aug 2022, Haifa, Israel. hal-03896410

HAL Id: hal-03896410
https://inria.hal.science/hal-03896410
Submitted on 13 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Complexity Landscape for Counting Queries

Extended Abstract

Meghyn Bienvenu\textsuperscript{1}, Quentin Manière\textsuperscript{1} and Michaël Thomazo\textsuperscript{2}

\textsuperscript{1}CNRS, University of Bordeaux, Bordeaux INP, LaBRI, Talence, France
\textsuperscript{2}Inria, DI ENS, ENS, CNRS, University PSL, Paris, France

Abstract

We summarize our recent work \cite{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 \cite{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 \cite{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 \cite{7, 8, 9}. A recent approach, proposed in \cite{10} as a generalization of \cite{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\textsubscript{core} and DL-Lite$^2$\textsubscript{core} \cite{8}. Recent works have shown that intractability arises even for simple forms of CCQs \cite{11, 12}. However, some interesting tractable cases have also been identified, notably, rooted CCQs \cite{10, 11, 13} and cardinality queries \cite{12}.
coupled with DL-Lite\text{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\text{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 $\mathcal{ELHI}$. 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 $2\text{EXP}$ upper bound in combined complexity for $\mathcal{ELHI}$. A matching lower bound, which applies to both $\mathcal{EL}$ and DL-Lite\text{pos}, is obtained by establishing a novel connection between CCQ answering and OMQA with closed predicates. This yields $2\text{EXP}$-completeness for a wide range of Horn DLs and closes the combined complexity gap for CCQ answering in DL-Lite\text{core}. We further prove a coNEXP lower bound for DL-Lite\text{pos}, which matches an existing coNEXP upper bound, yielding the precise combined complexity for DL-Lite\text{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 $\mathcal{ELHI}$.

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

<table>
<thead>
<tr>
<th>Concept</th>
<th>DL-Lite\text{pos}</th>
<th>DL-Lite\text{core}</th>
<th>DL-Lite$^\mathcal{H}$\text{pos}</th>
<th>DL-Lite$^\mathcal{H}$\text{core}</th>
<th>$\mathcal{EL}(\mathcal{H})$, $\mathcal{EL}(\mathcal{HI})$</th>
<th>$\mathcal{EL}(\mathcal{HI})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>coNP$^\dagger$</td>
<td>coNP$^\dagger$</td>
<td>EXP</td>
<td>coNEXP$^\dagger$</td>
<td>2EXP</td>
<td>2EXP</td>
</tr>
<tr>
<td>Role</td>
<td>coNP$^\dagger$</td>
<td>coNP$^\dagger$</td>
<td>EXP</td>
<td>coNEXP$^\dagger$</td>
<td>2EXP</td>
<td>2EXP</td>
</tr>
<tr>
<td>CCQ</td>
<td>coNEXP$^\dagger$</td>
<td>coNEXP$^\dagger$</td>
<td>2EXP</td>
<td>2EXP</td>
<td>2EXP</td>
<td>2EXP</td>
</tr>
</tbody>
</table>

Table 1
Combined complexity results for CCQs and cardinality queries, all bounds are tight. $^\dagger$: previously known upper / lower bound.
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


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