Closed Predicates in Description Logics: Results on Combined Complexity (Extended Abstract)

Nhung Ngo[1], Magdalena Ortiz[2], and Mantas Šimkus[2]

[1] Free University of Bozen-Bolzano
ngo@inf.unibz.it
[2] Vienna University of Technology
ortiz@kr.tuwien.ac.at|simkus@dbai.tuwien.ac.at

1 Introduction

Description Logics (DLs) are a family of languages for knowledge representation, popular for writing *ontologies* that describe domain-specific knowledge. As fragments of classical first-order logic, DLs make the open-world assumption. That is, the semantics of a knowledge base (KB) is given by *the set of all its models*. This view naturally supports modeling of incomplete knowledge. However, it is not always fully adequate, as in many applications of DLs, incomplete knowledge interacts with *complete information*. For example, in the popular paradigm of *ontology-based data access*, DL ontologies are advocated as a means to leverage domain knowledge when querying data sources. These sources often stem from traditional relational databases, which have a *closed-world* view and are assumed to fully describe the instances included in a relation. E.g., a municipality-provided table of bus routes can be assumed complete. Query answering algorithms should exploit this to exclude irrelevant models and infer more answers.

Combining open and closed world reasoning is not a new topic in DLs [3], but it has received renewed attention in recent years [13, 12, 7, 6, 18]. A prominent proposal for achieving partial closed world reasoning is to enrich KBs with a set of predicates that are to be interpreted as *closed* [12]. In this way, some relations are interpreted under the closed semantics, while others are considered open.

Most work considering closed predicates has focused on the problem of evaluating a query (usually a *conjunctive query*) under the standard *certain answer semantics*, and on *data complexity*, i.e. the complexity measured in the size of the data instance to be queried, while both the ontology and the query are assumed to be fixed. E.g., Franconi et at. ([6]) show that closed predicates lead to the loss of tractability of query answering even in the core fragments of DL-Lite, and provided the first matching upper bounds. An in-depth analysis of the reasons for intractability, and a fine-grained analysis of tractable classes can be found in [12, 13].

The data complexity is an important indicator of whether algorithms can scale to large amounts of data. However, it is also fundamental to understand the *combined complexity*, which takes the size of the ontology and the query along with the data into account, as the size of these two components often has a big impact on the performance of query answering algorithms. Indeed, ontologies can be very large [15], and disregarding their size easily leads to infeasible algorithms. The query size can also have a major impact on the query answering techniques, and avoiding the exponential blow-up in the size of the query was in fact one of the hardest obstacles to overcome for achieving scalable query answering even for DLs of very low data complexity like dialects of *DL-Lite* [1]. It is thus surprising that the combined complexity of query answering in the presence of closed predicates has received so little attention. In this extended abstract, we summarize several results on the combined complexity of reasoning in the presence of closed predicates in a wide range of DLs, which are presented in detail in [14].

2 DLs with closed predicates and query answering

We use $\mathsf{N_C}$, $\mathsf{N_R}$ and $\mathsf{N_I}$ for *concept names*, *role names* and *individual names*, respectively. DL *knowledge bases (KB)* take the form $\mathcal{K} = (\mathcal{T}, \Sigma, \mathcal{A})$. The TBox $\mathcal{T}$ is a set of axioms α, which usually include *concept inclusions* $C \sqsubseteq D$ and *role inclusions* $r \sqsubseteq s$. The specific syntax of concepts C, D, roles r, s and other possible axioms α depends on the DL in question. Σ is a set of concept and role names, called the *closed predicates* of $\mathcal{K}$. Finally, the *ABox* $\mathcal{A}$ is a finite set of assertions of the forms $A(b)$ and $r(a, b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $a, b \in \mathsf{N_I}$.

The notion of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is as usual. We make the *standard name assumption (SNA)*, i.e., $a^{\mathcal{I}} = a$ for all $\mathcal{I}$ and individuals a. Concepts C are interpreted as sets of objects $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and roles r as binary relations $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ in the standard way. The notions of TBox and ABox satisfaction $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$ are also as usual. We say that $\mathcal{I}$ *respects* a given $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ if:
(a) for all $A \in \Sigma \cap \mathsf{N_C}$, if $d \in A^{\mathcal{I}}$, then $A(d) \in \mathcal{A}$, and
(b) for all $r \in \Sigma \cap \mathsf{N_R}$, if $(d, d') \in r^{\mathcal{I}}$, then $r(d, d') \in \mathcal{A}$.
We call $\mathcal{I}$ a *model* of a KB $\mathcal{K} = (\mathcal{T}, \Sigma, \mathcal{A})$, and write $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$, $\mathcal{I} \models \mathcal{A}$, and $\mathcal{I}$ respects the closed predicates in Σ.

The main problem we study here is *query answering*, which is a generalization of standard reasoning tasks such as consistency testing and instance checking. Let $\mathsf{N_V}$ be a countably infinite set of *variables*. *Boolean conjunctive queries (BCQs)*, take the form $q = \alpha_1 \land \ldots \land \alpha_n$, where each α_j is an atom of the form $A(t)$ or $r(t, t')$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $\{t, t'\} \subseteq \mathsf{N_I} \cup \mathsf{N_V}$. Queries are given the standard first order semantics, where all variables are existentially quantified: we write $\mathcal{I} \models q'$ if there is an assignment from the variables in q to objects in $\Delta^{\mathcal{I}}$ that makes all atoms true. A *Boolean union of conjunctive queries (BUCQ)* is of the form $q' = q_1 \lor \ldots \lor q_n$, where each q_i is a BCQ. Given an interpretation $\mathcal{I}$, we write $\mathcal{I} \models q'$ if there is $i \in \{1, \ldots, n\}$ such that $\mathcal{I} \models q_i$, and given a KB $\mathcal{K}$ we write $\mathcal{K} \models q$ if $\mathcal{I} \models q$ holds for every model $\mathcal{I}$ of $\mathcal{K}$. Given $\mathcal{K}$ and a B(U)CQ q as above, the *query entailment problem* is to decide whether $\mathcal{K} \models q$.All results below apply also to query answering (as a decision problem) for queries with answer variables using the standard reduction to query entailment.

3 Results

In the presence of complete data, already consistency checking and instance query entailment in the basic *DL-Lite_{core}* are NP-complete. For the hardness, it is not hard to reduce 3-colorability to deciding consistency of a *DL-Lite_{core}* KB with closed concepts.

	Without closed predicates		With closed predicates	
	Instance queries	B(U)CQs	Instance queries	B(U)CQs
$\textit{DL-Lite}_{core}$	NL [1]	NP [1]	NP **(lb)**	coNEXPTIME-hard **(lb)**
$\textit{DL-Lite}_{\mathcal{R}}$	NL [1]	NP [1]	NP **(ub)**	2EXPTIME **(lb)**
$\mathcal{EL}$	P [2]	NP [11, 16]	EXPTIME **(lb)**	2EXPTIME **(lb)**
$\mathcal{ALCO}$	EXPTIME [17]	2EXPTIME **(lb)**	EXPTIME	2EXPTIME
$\mathcal{SHOQ}$, $\mathcal{SHOI}$	EXPTIME [19, 10, 8]	2EXPTIME [4, 9]	EXPTIME	2EXPTIME

Table 1: Combined complexity of query entailment in description logics with/without closed predicates. All are completeness results, unless stated otherwise. The bold (lb) and (ub) respectively indicate the main lower and upper bounds proved in this work.

For the membership, we provide an algorithm that shows an NP upper bound even for rich *DL-Lite* dialects with Boolean combinations of concepts, role inclusions, and nominals, and for UCQs without existentially quantified variables.

In $\mathcal{EL}$ we can use DBoxes to express disjunction, atomic negation, and nominals, making $\mathcal{EL}$ as expressive as $\mathcal{ALCO}$, and causing consistency testing to require exponential time in the worst case. Tight upper bounds for instance queries and quantifier free UCQs for any DL containing $\mathcal{EL}$ can then be inferred by a reduction to standard reasoning in well-known expressive DLs with nominals.

For more expensive queries, we show that entailment of UCQs is 2EXPTIME-hard for $\textit{DL-Lite}_{\mathcal{R}}$ and $\mathcal{EL}$ based on a reduction from the word problem for *Alternating Turing machines* (ATMs) with exponential work space [5]. Intuitively, using closed concepts can use a simple KB to enforce candidate computations of a given ATM, and then UCQs can test whether the represented computation is correct and accepting. The result is in sharp contrast to the NP upper bound in the setting with no DBox predicate, and coincides with known upper bounds for much richer DLs. A minor variation of the reduction allows shows coNEXPTIME-hardness for $\textit{DL-Lite}_{core}$, but we still don't know if this bound is tight.

A similar reduction allows us to prove 2EXPTIME-hardness of CQ entailment in the standard DL $\mathcal{ALCO}$ (with no DBox). This closes an open problem and singles out nominals as a previously unidentified source of complexity when answering queries over expressive DLs. A summary of our result and its comparison to the case without DBox is provided in Table 1.

In the light of these negative results, we also looked for classes of queries for which the query answering problem has lower complexity. In particular, we considered a restriction on variables called $\mathcal{K}$*-safety* which guarantees that it is sufficient to consider assignments of the variable to individuals in the KB, and there is no need to consider other objects in the interpretation domain . A query is called $\mathcal{K}$-*safe* if all its variables are $\mathcal{K}$-safe. We also generalize $\mathcal{K}$-safe queries to $\mathcal{K}$-*acyclic* by allowing for variables that are not $\mathcal{K}$-safe, but requiring that they induce only *acyclic* subqueries in the original query.

4 Conclusion

Closed predicates seem to have a great potential in applications of DLs, but as our results show, they are computationally costly. It remains a challenge to identify useful restricted settings with better complexity, and to explore other ways for effectively supporting partial completeness at a lower computational cost.

We note that all hardness proofs in the paper are given for KBs $(\mathcal{T}, \Sigma, \mathcal{A})$ where Σ contains only a constant number of concept names (which can be further reduced to a single one in DLs that can express disjointness). Moreover, they also apply to KBs $(\mathcal{T}, \emptyset, \mathcal{A})$ with no closed predicates, but where axioms $A \sqsubseteq \{a, b\}$ with the so-called *one-of* or *enum* are allowed. Hence our results also shed light on the effect of nominal enumerations in lightweight DLs.

Some interesting questions remain open for the *DL-Lite* family, like the precise complexity of (U)CQs in *DL-Lite*$_{core}$. We remark that *inverses* play a very limited role in the hardness results here. They all apply to *1-way DL-Lite KBs*, where only role names r occur in the right-hand side of axioms, and inverse roles r^- on the left-hand side; these KBs have *directed models* in the style of $\mathcal{EL}$ and $\mathcal{ALC}$.

Acknowledgments

This work was partially supported by the Austrian Science Fund (FWF) projects T515, P25207, and the project ORMiE from the Free University of Bozen-Bolzano.

References

1. Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The dl-lite family and relations. *J. Artif. Int. Res.*, 36(1):1–69, September 2009.
2. Franz Baader, Sebastian Brand, and Carsten Lutz. Pushing the EL envelope. In *Proc. of IJCAI 2005*, pages 364–369. Morgan-Kaufmann Publishers, 2005.
3. Marco Cadoli, Francesco M. Donini, and Marco Schaerf. Closed world reasoning in hybrid systems. In *Proc. of ISMIS 1990*, pages 474–481. Elsevier Science Publishing, 1990.
4. Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Regular path queries in expressive description logics with nominals. In *Proc. of IJCAI 2009*, pages 714–720, 2009.
5. Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. Alternation. *Journal of the ACM*, 28(1):114–133, 1981.
6. Enrico Franconi, Yazmin Angélica Ibáñez-García, and Inanç Seylan. Query answering with DBoxes is hard. *Electr. Notes Theor. Comput. Sci.*, 278:71–84, 2011.
7. Enrico Franconi, Volha Kerhet, and Nhung Ngo. Exact query reformulation over databases with first-order and description logics ontologies. *J. Artif. Intell. Res. (JAIR)*, 48:885–922, 2013.

8. Birte Glimm, Ian Horrocks, and Ulrike Sattler. Deciding shoqˆcap knowledge base consistency using alternating automata. In Franz Baader, Carsten Lutz, and Boris Motik, editors, *Proceedings of the 21st International Workshop on Description Logics (DL2008), Dresden, Germany, May 13-16, 2008*, volume 353 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.
9. Birte Glimm, Ian Horrocks, and Ulrike Sattler. Unions of conjunctive queries in SHOQ. In *Proc. of the 11th International Conference on the Principles of Knowledge Representation and Reasoning (KR-08)*, pages 252–262. AAAI Press/The MIT Press, 2008.
10. Jan Hladik. A tableau system for the description logic SHIO. In *IJCAR Doctoral Programme*, 2004.
11. Markus Krötzsch and Sebastian Rudolph. Conjunctive queries for $\mathcal{EL}$ with composition of roles. volume 250 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/Vol-250/`, 2007.
12. Carsten Lutz, Inanç Seylan, and Frank Wolter. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *Proc. of IJCAI 2013*. IJCAI/AAAI, 2013.
13. Carsten Lutz, Inanç Seylan, and Frank Wolter. Ontology-mediated queries with closed predicates. In *Proc. of IJCAI 2015*, 2015.
14. Nhung Ngo, Magdalena Ortiz, and Mantas Simkus. Closed predicates in description logics: Results on combined complexity. In *KR*. AAAI Press, April 2016.
15. Cathy Price and Kent Spackman. Snomed clinical terms. *British Journal of Healthcare Computing and Information Management*, 17(3):27–31, 2000.
16. Riccardo Rosati. On conjunctive query answering in $\mathcal{EL}$. volume 250 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/Vol-250/`, 2007.
17. Klaus Schild. A correspondence theory for terminological logics: Preliminary report. In *Proc. of IJCAI 1991*, pages 466–471. Morgan Kaufmann, 1991.
18. Inanç Seylan, Enrico Franconi, and Jos de Bruijn. Effective query rewriting with ontologies over DBoxes. In *Proc. of IJCAI 2009*, pages 923–925, 2009.
19. Stephan Tobies. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation*. PhD thesis, LuFG Theoretical Computer Science, RWTH-Aachen, Germany, 2001.