How to Approximate Ontology-Mediated Queries (Extended Abstract)

Anneke Haga[1], Carsten Lutz[1], Leif Sabellek[1], and Frank Wolter[2]

[1] Department of Computer Science, University of Bremen, Germany
[2] Department of Computer Science, University of Liverpool, UK

The complexity of ontology-mediated querying in popular expressive description logics (DLs) such as $\mathcal{ALC}$ and $\mathcal{ALCI}$ is prohibitively high, namely coNP-complete in data complexity [4] and ExpTime- resp. 2ExpTime-complete in combined complexity [3]. As a consequence, practical implementations resort to approximations of ontology mediated queries (OMQs) [6, 5, 7] that are, however, often of a rather pragmatic nature. The work reported about in this abstract is concerned with a systematic study of OMQ approximations that achieve the following desiderata [2]:

(i) PTime data complexity,
(ii) fixed-parameter tractability (FPT) with the parameter being the size of the OMQ (if possible) and
(iii) improved combined complexity (if possible),

We mainly consider approximation from below, that is, approximations that are sound, but (potentially) incomplete. Recall that an OMQ is a triple $Q = (\mathcal{O}, \Sigma, q)$ where $\mathcal{O}$ is an ontology, q an actual query such as a conjunctive query (CQ), and Σ a signature for the databases $\mathcal{D}$ that Q is evaluated on. Our starting point is the observation that we may attain (the only non-optional) desideratum (i) by relaxing the ontology $\mathcal{O}$ or the database $\mathcal{D}$. Note that relaxing the query q is not promising towards this aim as ontology-mediated querying is coNP-hard already for atomic queries (AQs), that is, CQs of the form $A(x)$.

For ontology relaxing approximation, we choose a DL $\mathcal{L}$ for which ontology-mediated querying is in PTime in data complexity. We then replace $\mathcal{O}$ with every $\mathcal{L}$-ontology $\mathcal{O}'$ such that $\mathcal{O} \models \mathcal{O}'$ (which guarantees soundness) and take the union of all answers. As choices for $\mathcal{L}$, we consider Horn description logics such as $\mathcal{ELI}$ and frontier-one tuple-generating dependencies (TGDs) [1] with the treewidth of the body and head bounded by a constant. For database relaxing approximation, we choose a class $\mathfrak{D}$ of databases for which ontology-mediated querying is in PTime in data complexity. We then replace $\mathcal{D}$ with every database $\mathcal{D}' \in \mathfrak{D}$ such that there is a homomorphism from $\mathcal{D}'$ to $\mathcal{D}$ (which guarantees soundness) and then take the union of all answers. As choices for $\mathfrak{D}$, we consider databases of bounded treewidth and databases that are proper trees.

An OMQ language is a pair $(\mathcal{L}, \mathcal{Q})$ with $\mathcal{L}$ an ontology language and $\mathcal{Q}$ a query language. We study the approximation of OMQ languages $(\mathcal{L}, \mathcal{Q})$ with

$\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCI}\}$ and $\mathcal{Q} \in \{\text{UCQ}, \text{CQ}, \text{AQ}, \text{bELIQ}\}$ where UCQ denotes unions of CQs and bELIQ denotes the class of unary CQs that correspond to $\mathcal{ELI}$-concepts (ELIQs) and of Boolean CQs $\exists x\, q(x)$ with $q(x)$ an ELIQ. The exact problem studied is *approximate OMQ evaluation*, meaning to decide, given an OMQ Q, a database $\mathcal{D}$, and a tuple $\bar{a}$ of constants from $\mathcal{D}$, whether $\bar{a}$ is an approximate answer to Q on $\mathcal{D}$.

In this abstract, we only state explicitly two main results, the first one concerning ontology relaxing approximation.

Theorem 1. *Let $\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCI}\}$ and $\ell, k, k' \geq 1$ with $\ell < k$. Then $\ell, k, 1, k'$-ontology relaxing evaluation is*

1. ExpTime-*complete in combined complexity and* PTime-*complete in data complexity in $(\mathcal{L}, \mathcal{Q})$, $\mathcal{Q} \in \{AQ, CQ, UCQ\}$;*
2. *FPT in $(\mathcal{L}, \mathcal{Q})$, $\mathcal{Q} \in \{CQ_p^{tw}, UCQ_p^{tw} \mid p \geq 1\}$.*

Let us clarify notation. A CQ *has treewidth at most* (ℓ, k) if it admits a tree decomposition in which the size of the bags is bounded by k and the overlap between the bags is bounded by ℓ. Then, $\ell, k, 1, k'$*-ontology relaxing evaluation* means that we replace $\mathcal{O}$ with every set of frontier-one TGDs $\mathcal{O}'$ such that $\mathcal{O} \models \mathcal{O}'$ and the TGDs in $\mathcal{O}'$ are such that the treewidth of their bodies is at most (ℓ, k) while the treewidth of their heads is at most $(1, k')$. With CQ_p^{tw}, we mean CQs of treewidth bounded by the constant p and UCQ_p^{tw} means disjunctions of CQs from CQ_p^{tw}. Note that ontology relaxing approximation indeed achieves desideratum (i) and that in the case of $\mathcal{ALCI}$, it additionally achieves desideratum (iii). Desideratum (ii) is only achieved for (U)CQs of bounded treewidth. In the full paper, we also study ontology relaxing approximation using the DL $\mathcal{ELI}_\bot^u$ in place of TGDs, where we additionally attain linear time data complexity for $(\mathcal{ALCI}, \text{bELIQ})$.

The second main theorem concerns database relaxing approximation.

Theorem 2. *Let $1 \leq \ell < k$. Then ℓ, k-database relaxing evaluation is*

1. 2ExpTime-*complete in combined complexity and FPT (thus in* PTime *in data complexity) in $(\mathcal{ALCI}, \mathcal{Q})$, $\mathcal{Q} \in \{CQ, UCQ, CQ_p^{tw}, UCQ_p^{tw} \mid p \geq 1\}$;*
2. ExpTime-*complete in combined complexity and FPT in $(\mathcal{ALC}, \mathcal{Q})$ and in $(\mathcal{ALCI}, \mathcal{Q})$, $\mathcal{Q} \in \{AQ, bELIQ\}$.*

Here, ℓ, k*-database relaxing evaluation* means that we replace the input database $\mathcal{D}$ with every database $\mathcal{D}'$ of treewidth at most (ℓ, k) that admits a homomorphism to $\mathcal{D}$. Thus also database relaxing approximations achieve desideratum (i). In contrast to ontology relaxing approximations, there are no cases where desideratum (iii) is achieved. However, desideratum (ii) is achieved for a much wider class of queries.

In the full paper, we also study database relaxing approximation using proper trees in place of databases of bounded treewidth for which Point 2 of Theorem 2 can be strengthened to linear time in data complexity (which implies FPT). We also make the surprising observation that tree-database relaxing evaluation is ExpSpace-hard in $(\mathcal{ALC}, \text{CQ})$ and 2ExpTime-hard in $(\mathcal{ALC}, \text{UCQ})$, thus *harder* than non-approximate evaluation which is ExpTime-complete.

We also study approximation from above in the form of ontology strengthening approximation and database strengthening approximation. These are defined dually to ontology/database relaxing approximations and are complete, but (potentially) unsound. For $\mathcal{L}$-ontology strengthening approximation, we replace $\mathcal{O}$ with every $\mathcal{L}$-ontology $\mathcal{O}'$ such that $\mathcal{O}' \models \mathcal{O}$ (which guarantees completeness) and take the intersection of all answers. For $\mathfrak{D}$-database strengthening approximation, we replace $\mathcal{D}$ with every database $\mathcal{D}' \in \mathfrak{D}$ such that there is a homomorphism from $\mathcal{D}$ to $\mathcal{D}'$ (which guarantees completeness) and then take the intersection of all answers.

It turns out that ontology strengthening approximation and database strengthening approximation are less well-behaved than their counterparts that approximate from below. We state the two main theorems that illustrate this. Recall that $\mathcal{ELIU}_\bot$ is the fragment of $\mathcal{ALCI}$ that extends $\mathcal{ELI}_\bot$ with disjunction.

Theorem 3. *Let* $\mathcal{Q} \in \{AQ, CQ, UCQ\}$. $\mathcal{ELI}_\bot$*-ontology strengthening evaluation in* $(\mathcal{ELIU}_\bot, \mathcal{Q})$ *is* 2ExpTime*-complete in combined complexity and FPT.*

So $\mathcal{ELI}_\bot$-ontology strengthening evaluation satisfies desiderata (i) and (ii), but not (iii). In fact, we consider the lower bound for $(\mathcal{ELIU}_\bot, \text{AQ})$ surprising as non-approximate evaluation is only ExpTime-complete [3]. Thus, approximate evaluation from above is significantly harder. The lower bound depends only on disjunction on the *left* hand side of concept inclusions, which are syntactic sugar, but not on the seemingly much more 'dangerous' disjunctions on the right hand side. It is in fact a byproduct of our proofs that, without disjunctions on the left, $\mathcal{ELI}_\bot$-ontology strengthening evaluation in $(\mathcal{ELIU}_\bot, \text{UCQ})$ is ExpTime-complete. $\mathcal{ALCI}$-ontologies can be rewritten in polynomial time into a 'nesting-free' normal form that is often used by reasoners and that has sometimes been presupposed for approximation [7]. The rewriting is not equivalence preserving, but only yields a conservative extension. $\mathcal{ALCI}$-ontologies in this form can in turn be rewritten into an equivalent $\mathcal{ELIU}_\bot$-ontology without disjunction on the left and thus enjoy $\mathcal{ELI}_\bot$-ontology strengthening evaluation in ExpTime. Ontology strengthening evaluation in $(\mathcal{ELIU}_\bot, \mathcal{Q})$ remains a non-trivial open problem.

For the second theorem, we use $\mathfrak{D}_1$ to denote the class of databases that are disjoint unions of trees, multi-edge and self-loops admitted.

Theorem 4. $\mathfrak{D}_1$*-database strengthening approximation is* coNP*-complete in data complexity in* $(\mathcal{ALCI}, UCQ)$*. The lower bound already holds when the ontology is empty. It also holds in* $(\mathcal{EL}, CQ)$*.*

Thus, $\mathfrak{D}_1$-database strengthening approximation does not satisfy our crucial desideratum (i). For $(\mathcal{EL}, \text{CQ})$, the data complexity even *increases* from PTime to coNP-complete when transitioning from non-approximate evaluation to the approximate version.

Acknowledgement. Anneke Haga and Carsten Lutz were supported by DFG CRC 1320 Ease. Frank Wolter was supported by EPSRC grant EP/S032207/1.

References

1. Baget, J., Leclère, M., Mugnier, M., Salvat, E.: Extending decidable cases for rules with existential variables. In: Proc. of IJCAI. pp. 677–682 (2009)
2. Haga, A., Lutz, C., Sabellek, L., Wolter, F.: How to approximate ontology-mediated queries. In: Proc. of KR (2021)
3. Lutz, C.: The complexity of conjunctive query answering in expressive description logics. In: Proc. of IJCAR. LNCS, vol. 5195, pp. 179–193. Springer (2008)
4. Schaerf, A.: On the complexity of the instance checking problem in concept languages with existential quantification. J. of Intel. Inf. Systems **2**, 265–278 (1993)
5. Thomas, E., Pan, J.Z., Ren, Y.: TrOWL: Tractable OWL 2 reasoning infrastructure. In: Proc. of ESWC. LNCS, vol. 6089, pp. 431–435. Springer (2010)
6. Tserendorj, T., Rudolph, S., Krötzsch, M., Hitzler, P.: Approximate OWL-reasoning with Screech. In: Proc. of RR. LNCS, vol. 5341, pp. 165–180. Springer (2008)
7. Zhou, Y., Cuenca Grau, B., Nenov, Y., Kaminski, M., Horrocks, I.: PAGOdA: Pay-as-you-go ontology query answering using a datalog reasoner. J. Artif. Intell. Res. **54**, 309–367 (2015)