Towards Soundness Preserving Approximation for TBox Reasoning in OWL 2

Yuan Ren
Dpt. of Computing Science
University of Aberdeen
y.ren@abdn.ac.uk

Jeff Z. Pan
Dpt. of Computing Science
University of Aberdeen
jeff.z.pan@abdn.ac.uk

Yuting Zhao
Dpt. of Computing Science
University of Aberdeen
yuting.zhao@abdn.ac.uk

ABSTRACT

Large scale semantic web applications require efficient and robust description logic (DL) reasoning services. In this paper, we present a soundness preserving tractable approximative reasoning approach for TBox reasoning in $\mathcal{R}$, a fragment of OWL2-DL supporting $\mathcal{ALC}$ GCIs and role chains with 2ExpTime-hard complexity. We first rewrite the ontologies into $\mathcal{EL}^+$ with an additional complement table maintaining the complementary relations between named concepts, and then classify the approximation. Preliminary evaluation shows that our approach can classify existing benchmarks in large scale efficiently with a high recall.

1. MOTIVATION

The Web Ontology Language (OWL) has been so phenomenally successful that OWL 2 (the second version of OWL) is currently undergoing standardisation by the W3C OWL Working Group [1]. It provides not only OWL2-DL, as a decidable extension of OWL-DL, but also three tractable profiles, including OWL2-EL, OWL2-QL and OWL2-RL. These profiles are all based on description logics (DLs). So far, successful applications of ontologies have benefited from the reasoning services provided by DL reasoners.

However, reasoning such as TBox classification in DLs is usually hard; e.g., the DL $\mathcal{SROIQ}$ the adjacent logic of OWL2-DL, is N2ExpTime-complete. Traditionally, such expressive DLs are handled by tableau algorithm, upon which many DL reasoners, such as FaCT++, Pellet, Racer and HermiT, have been built. These reasoners implemented various optimisation techniques [4, 3, 5, 12, 11, 6] to improve the efficiency. However, their performance and robustness are still restricted by the worst case complexity bounds. Our early evaluation has already shown that reasoning with even very small ontologies can be very difficult [2]. Thus, reasoning with expressive DLs makes up a major bottleneck for real-time knowledge processing in the upcoming semantic web. On the other hand, TBox reasoning in tractable profiles such as the OWL2-EL is only PTime-complete. This brings a new challenge: can users use OWL2-DL to build their ontologies and still enjoy the tractable services provided by the profiles?

[1]http://www.w3.org/2007/OWL/wiki/OWL_Working_Group
[2]http://www.abdn.ac.uk/~csc303/benchmark/RBenchmarkTest.zip

Approximation has been identified as a potential way to combine the modelling flexibility of expressive OWL2-DL and tractability of less expressive profiles. Semantic approximations [8] use reasoners of the more expressive language to precompute all the entailed axioms (in the less expressive language), but this precomputation could be really expensive. Syntactic approximations [1, 2, 9, 10, 13] find the syntactic upper and lower bounds of the ontology in the less expressive language; therefore, they could be very useful when reasoners of the more expressive language do not exist or cannot load the original theory. However, many of these approximation approaches have not been evaluated or proven lowly effective on today's large scale semantic web.

In this paper, we present a soundness preserving approximative reasoning approach for OWL 2. As a first step, we focus on TBox reasoning of $\mathcal{R}$, a fragment of OWL 2 covering all the $\mathcal{ALC}$ expressiveness and role chain. We first approximate the TBox into a $\mathcal{EL}^+$ TBox with an additional complement table maintaining the complementary relations between named concepts, and then classify the approximation. Preliminary evaluation shows that our approach can classify existing benchmarks in large scale efficiently with a high recall.

2. APPROACH

$\mathcal{EL}^+$ has the same types of role inclusions (RIs) as $\mathcal{R}$, so we preserve all the RIs. Regarding concept axioms, we represent non-$\mathcal{EL}^+$ concept expressions with new concept names and define their complements. The complementary relations will be maintained in a separate complement table CT. The reasoning will be enriched by the complementary relations to recover the semantics of approximated concepts.

Given an $\mathcal{R}$ TBox, the approximation generates a pair (T, CT) by the following steps:

1. Initialising T and CT as empty sets.
2. For each concept expression C in the TBox, introducing an unique fresh concept name N_C to represent C, an unique fresh concept name nN_C to represent $\neg C$. Names for $\top$, $\bot$ and atomic concepts are kept.
3. Adding each pair of (N_C, nN_C) and (nN_C, N_C) into CT.
4. For each concept expression C, if C constructed by an $\mathcal{EL}^+$ constructor, e.g C is $D \sqcap E$, adding definition of N_C, e.g. $N_C \equiv N_D \sqcap N_E$, into T.

5. For each concept expression C, if C is not constructed by an $\mathcal{EL}^+$ constructor, e.g. C is $\forall r.D$, then $\neg C$ is constructed by an $\mathcal{EL}^+$ constructor ($\exists r.\neg D$). Thus, adding definition of nN_C, e.g. $nN_C \equiv \exists r.nN_D$, into T.
6. For each concept expression $C \sqsubseteq D$ ($C \equiv D$) in TBox, adding $N_C \sqsubseteq N_D$ ($N_C \equiv N_D$) into T.
7. Adding all the RIs into T.

The resulting T will be an $\mathcal{EL}^+$ TBox, and CT containing all the complementary relations for concept names in T.

In the reasoning phase, we enrich the classification of T with information maintained in CT by the following heuristics:

1. If $A \sqsubset B \sqcap C$ and $(B, C) \in CT$, then $A \sqsubseteq \bot$.
2. If $A \sqsubseteq B$ and $(A, nA), (B, nB) \in CT$, then $nB \sqsubseteq nA$.
3. If $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq \bot$ then for each $1 \leq i \leq n$, $A_1 \sqcap \cdots \sqcap A_{i-1} \sqcap A_{i+1} \sqcap \cdots \sqcap A_n \sqsubseteq nA_i$ where $(A_i, nA_i) \in CT$.
4. If $A \sqsubseteq \exists r.\bot$, then $A \sqsubseteq \bot$.

The complement-enriched reasoning algorithm can still be performed in PTIME and it is soundness guaranteed. With this algorithm, we can efficiently compute the subsumption between named concepts.

3. EVALUATION

We implemented our approximation and reasoning algorithm in our REL reasoner, which is part of the TrOWL ontology reasoning infrastructure. To evaluate its performance in practice, we compared with the latest versions of mainstream reasoners on 5 real world ontologies taken from the HermiT benchmark[3]. According to the HermiT evaluation [7], these ontologies are most difficult ones for existing OWL-DL reasoners. To focus on TBox reasoning, we removed all the ABox axioms from these ontologies. The results are shown in Tab.1, in which time unit is second. t/o means time out after 5 minutes. e/o means error occurred. As we can see, REL performs more efficiently than all the other reasoners (HermiT and FaCT++ takes longer time in organising the taxonomy on Cyc after classification and the overall time is longer than REL). Furthermore, REL is the only one that can classify the FMA ontology.

Table 1: Efficiency Evaluation

Ontology	HermiT	FaCT++	RacerPro	Pellet	REL
Wine	1.297	0.829	19.469	1.719	0.579
Cyc	3.609*	1.329*	56.047	99.672	5.016
Tambis	1.265	e/o	2.126	e/o	0.625
FMA	e/o	e/o	t/o	e/o	21.718
DLP	151.517	0.922	3.922	4.735	0.594

In order to evaluate the completeness, we counted the number of retrieved subsumptions between named concepts, and compared with the results of FaCT++ (according to the HermiT report, FaCT++ yields correct results on all these ontologies). The recall is shown in Tab.2. As we can see, the recall is all over 95%.

[3] http://hermit-reasoner.com/2009/JAIR_benchmarks/

Table 2: Completeness Evaluation

Ontology	Recall
Wine	96.8%
Cyc	100%
Tambis	98.7%
FMA	N/A
DLP	100%

4. REFERENCES

[1] P. Groot, H. Stuckenschmidt, and H. Wache. Approximating description logic classification for semantic web reasoning. In A. Gómez-Pérez and J. Euzenat, editors, *ESWC*, volume 3532 of *Lecture Notes in Computer Science*, pages 318–332. Springer, 2005.

[2] P. Hitzler and D. Vrandecic. Resolution-based approximate reasoning for owl dl. In Y. G. et al., editor, *Proceedings of the 4th International Semantic Web Conference, Galway, Ireland, November 2005*, volume 3729 of *Lecture Notes in Computer Science*, pages 383–397. Springer, Berlin, NOV 2005.

[3] I. Horrocks. Reasoning with expressive description logics: Theory and practice. In *In: Andrei Voronkov, (ed) Proc. 18th Int. Conf. on Automated Deduction (CADE-18)*, pages 1–15. Springer, 2002.

[4] I. Horrocks, U. Sattler, and S. Tobies. Practical reasoning for expressive description logics. In *Proceedings of the 6th International Conference on Logic Programming and Automated Reasoning*, pages 161–180, London, UK, 1999. Springer-Verlag.

[5] I. Horrocks, U. Sattler, and S. Tobies. Practical reasoning for very expressive description logics. *Logic Journal of the IGPL*, 8:2000, 2000.

[6] E. K. Hudek and G. Weddell. Binary absorption in tableaux-based reasoning for description logics. In *In Proc. DL 2006*, 2006.

[7] B. Motik, R. Shearer, and I. Horrocks. Hypertableau Reasoning for Description Logics, 2008. Submitted to a journal.

[8] J. Z. Pan and E. Thomas. Approximating OWL-DL Ontologies. In *the Proc. of the 22nd National Conference on Artificial Intelligence*, pages 1434–1439, 2007.

[9] M. Schaerf and M. Cadoli. Tractable reasoning via approximation. *Artificial Intelligence*, 74:249–310, 1995.

[10] H. Stuckenschmidt and F. van Harmelen. Approximating terminological queries. In *FQAS '02: Proceedings of the 5th International Conference on Flexible Query Answering Systems*, pages 329–343, London, UK, 2002. Springer-Verlag.

[11] D. Tsarkov and I. Horrocks. Efficient reasoning with range and domain constraints. In *In Proc. of the 2004 Description Logic Workshop (DL 2004*.

[12] D. Tsarkov, I. Horrocks, and P. F. Patel-Schneider. Optimizing terminological reasoning for expressive description logics. *J. Autom. Reason.*, 39(3):277–316, 2007.

[13] H. Wache, P. Groot, and H. Stuckenschmidt. Scalable instance retrieval for the semantic web by approximation. In *WISE Workshops*, volume 3807 of *Lecture Notes in Computer Science*, pages 245–254. Springer, 2005.