Lower and Upper Approximations for Depleting Modules of Description Logic Ontologies

William Gatens, Boris Konev, and Frank Wolter

University of Liverpool, UK

It is known that no algorithm can extract the minimal depleting Σ-module from ontologies in expressive description logics (DLs). Thus research has focused on algorithms that approximate minimal depleting modules 'from above' by computing a depleting module that is not necessarily minimal. The first contribution of this paper is an implementation (AMEX) of such a depleting module extraction algorithm for expressive acyclic DL ontologies that uses a QBF solver for checking conservative extensions relativised to singleton interpretations. To evaluate AMEX and other module extraction algorithms we propose an algorithm approximating minimal depleting modules 'from below' (which also uses a QBF solver). We present experiments based on NCI (the National Cancer Institute Thesaurus) that indicate that our lower approximation often coincides with (or is very close to) the upper approximation computed by AMEX, thus proving for the first time that an approximation algorithm for minimal depleting modules can be almost optimal on a large ontology in a non-tractable DL.

We use standard notation from logic and description logic (DL), details can be found in [1]. In a DL, concepts are constructed from countably infinite sets $\mathsf{N_C}$ of *concept names* and $\mathsf{N_R}$ of *role names* using the concept constructors defined by the DL. A *signature* Σ is a finite subset of $\mathsf{N_C} \cup \mathsf{N_R}$. The Σ-*reduct* $\mathcal{I}|_\Sigma$ of an interpretation $\mathcal{I}$ is obtained from $\mathcal{I}$ by setting $\Delta^{\mathcal{I}|_\Sigma} = \Delta^{\mathcal{I}}$, and $X^{\mathcal{I}|_\Sigma} = X^{\mathcal{I}}$ for all $X \in \Sigma$, and $X^{\mathcal{I}|_\Sigma} = \emptyset$ for all $X \notin \Sigma$. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-*inseparable*, in symbols $\mathcal{T}_1 \equiv_\Sigma \mathcal{T}_2$, if $\{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{T}_1\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{T}_2\}$.

Definition 1. *Let $\mathcal{M} \subseteq \mathcal{T}$ be TBoxes and Σ a signature. Then $\mathcal{M}$ is a* depleting Σ-module *of $\mathcal{T}$ if $\mathcal{T} \setminus \mathcal{M} \equiv_{\Sigma \cup \mathsf{sig}(\mathcal{M})} \emptyset$.*

Every depleting module $\mathcal{M}$ of $\mathcal{T}$ is inseparable from $\mathcal{T}$ for its signature and, in particular, $\mathcal{T} \equiv_\Sigma \mathcal{M}$. It follows from results in [10] that $\mathcal{T} \models \varphi$ iff $\mathcal{M} \models \varphi$ holds for any second-order sentence φ using symbols from Σ only. Thus, a TBox and its depleting Σ-module can be equivalently replaced by each not only in applications using entailed CIs between Σ-concepts but also in data access applications with data given in Σ. For further discussion of the properties of depleting modules see [3, 10].

Approximating Depleting Modules While the inseparability-based notion of a module has its theoretic appeal, unfortunately, checking if a subset $\mathcal{M}$ of $\mathcal{T}$ is a depleting Σ-module of $\mathcal{T}$ for some given signature Σ is undecidable already for general TBoxes formulated in $\mathcal{EL}$ and for acyclic $\mathcal{ALC}$-TBoxes [10, 13]. Therefore, we introduce lower and upper approximations of depleting Σ-modules.

Assume that $\mathcal{T}_1$ and $\mathcal{T}_2$ are TBoxes and Σ a signature. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are *1-Σ-inseparable*, in symbols $\mathcal{T}_1 \equiv^1_\Sigma \mathcal{T}_2$, if $\{\mathcal{I}|_\Sigma \mid \sharp\Delta^{\mathcal{I}} = 1 \text{ and } \mathcal{I} \models \mathcal{T}_1\} = \{\mathcal{I}|_\Sigma \mid \sharp\Delta^{\mathcal{I}} = 1 \text{ and } \mathcal{I} \models \mathcal{T}_2\}$. If $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-inseparable, then they are 1-Σ-inseparable.

Role%	0%					50%					100%				
$\lvert\Sigma\rvert$	Star	AMEX	Hybrid	1-dep	Diff/200	Star	AMEX	Hybrid	1-dep	Diff/200	Star	AMEX	Hybrid	1-dep	Diff/200
NCI*															
100	3834.21	722.21	710.65	671.68	10	3887.17	972.68	960.44	960.39	3	3915.18	1013.23	1000.79	1000.70	4
250	5310.96	1721.28	1705.71	1705.61	4	5452.52	1882.65	1870.87	1870.83	4	5539.39	1924.77	1912.95	1912.89	5
500	6977.33	2725.74	2700.00	2699.96	2	7186.09	2933.90	2919.23	2919.15	3	7237.22	2987.75	2977.62	2977.58	2
750	8235.36	3573.97	3542.57	3542.49	2	8437.07	3801.24	3786.05	3786.01	2	8579.98	3902.12	3892.36	3892.26	4
1000	9273.62	4341.25	4305.41	4305.38	1	9525.81	4570.55	4553.91	4553.81	4	9542.00	4621.42	4612.19	4606.46	3
NCI* ($\sqsubseteq$)															
100	58.74	69.53	58.74	58.74	0	291.91	326.68	291.91	291.89	2	345.01	357.58	345.01	344.89	5
250	330.79	386.45	330.79	330.78	1	652.09	716.64	652.09	652.09	0	775.00	808.03	775.00	775.00	0
500	852.14	1007.20	852.14	852.14	0	1173.34	1274.27	1173.34	1173.34	0	1387.67	1444.68	1387.67	1387.67	0
750	1352.47	1571.46	1352.47	1352.47	0	1681.12	1816.79	1681.12	1681.12	0	1935.47	2009.62	1935.47	1935.47	0
1000	1788.02	2046.62	1788.02	1788.02	0	2152.83	2315.19	2152.83	2152.83	0	2434.06	2519.63	2434.06	2434.06	0
NCI* ($\equiv$)															
100	2760.96	310.25	310.25	309.21	122	2759.11	319.08	319.11	318.23	114	2782.54	318.79	318.79	317.73	130
250	3989.74	622.65	622.63	621.89	110	4000.93	623.38	623.25	622.50	104	3973.78	624.51	624.23	623.47	102
500	4994.77	1003.76	1003.75	1002.95	108	4983.10	1002.14	1002.04	1001.32	101	4986.77	999.87	999.87	999.08	101
750	5539.78	1310.33	1310.31	1309.38	124	5531.60	1313.51	1311.54	1310.67	90	5525.28	1307.71	1307.71	1306.85	106
1000	5886.91	1573.06	1573.14	1572.11	122	5901.34	1577.34	1572.14	1571.10	102	5903.37	1576.95	1571.18	1570.08	103

Table 1. Modules of NCI* and its fragments

Definition 2. *Let $\mathcal{M} \subseteq \mathcal{T}$ be TBoxes and Σ a signature. Then $\mathcal{M}$ is a* 1-depleting Σ-module *of $\mathcal{T}$ if $\mathcal{T} \setminus \mathcal{M} \equiv^1_{\Sigma \cup \mathsf{sig}(\mathcal{M})} \emptyset$.*

In contrast to Σ-inseparability which is undecidable, 1-Σ-inseparability can be decided by reduction to the validity of quantified Boolean formulas (QBF). By definition, every depleting Σ-module of $\mathcal{T}$ is a 1-depleting Σ-module of $\mathcal{T}$, so 1-depleting Σ-modules are a good candidate for approximating depleting modules from below.

Theorem 1. *Given an $\mathcal{ALCQI}$-TBox $\mathcal{T}$ and signature Σ, the unique minimal 1-depleting Σ-module of $\mathcal{T}$ can be computed in polynomial time with each call to a QBF solver treated as a constant time oracle call.*

Our upper approximation of depleting Σ-modules is also based on 1-Σ-inseparability but uses an additional syntactic dependency check to ensure that a depleting module is extracted. Let $\mathcal{T}$ be an *acyclic* TBox and Σ a signature. We say that $\mathcal{T}$ *has a direct Σ-dependency* if there exists $\{A, X\} \subseteq \Sigma$ with $A \prec^+_{\mathcal{T}} X$, where $\prec^+_{\mathcal{T}}$ is the transitive closure of the relation $\prec_{\mathcal{T}} \subseteq \mathsf{N_C} \times (\mathsf{N_C} \cup \mathsf{N_R})$ defined by setting $A \prec_{\mathcal{T}} X$ iff there exists an axiom of the form $A \sqsubseteq C$ or $A \equiv C$ in $\mathcal{T}$ such that X occurs in C. Although one can construct TBoxes $\mathcal{T}$ and depleting Σ-modules $\mathcal{M}$ of $\mathcal{T}$ such that $\mathcal{T} \setminus \mathcal{M}$ contains direct $\Sigma \cup \mathsf{sig}(\mathcal{M})$-dependencies (see [10]), for typical depleting Σ-modules $\mathcal{M}$, the set $\mathcal{T} \setminus \mathcal{M}$ should not contain direct $\Sigma \cup \mathsf{sig}(\mathcal{M})$-dependencies because such dependencies indicate a semantic link between two distinct symbols in $\Sigma \cup \mathsf{sig}(\mathcal{M})$.

Theorem 2. *Given an acyclic $\mathcal{ALCQI}$ TBox $\mathcal{T}$ and signature Σ, the unique minimal depleting Σ-module s.t. $\mathcal{T} \setminus \mathcal{M}$ contains no direct $\Sigma \cup \mathsf{sig}(\mathcal{M})$-dependencies can be computed in polynomial time with each call to the QBF solver being treated as a constant time oracle call.*

Experiments and Evaluation To evaluate how close depleting module extraction algorithms can approximate minimal depleting modules we compared

- our new system AMEX, in which the inseparability check is implemented by reduction to the validity of QBF and uses the QBF solver sKizzo [2];
- $\top\bot^*$ locality-based module extraction [3, 15] as implemented in the OWL-API library version 3.2.4.1806 (called STAR-modules for ease of pronunciation);
- a hybrid approach in which one iterates AMEX and STAR-module extraction. This results in a depleting module contained in both the AMEX and the STAR-module;
- the algorithm computing the minimal 1-depleting module. The inseparability check was again implemented using the reduction to the validity of QBF and uses sKizzo.

We used fragments of the NCI Thesaurus version 08.09d taken from the Bioportal [20]. The results given in Table 1 show the average sizes of the modules extracted by the four algorithms from the set of CIs $\mathsf{NCI}^\star(\sqsubseteq)$, the set of CEs $\mathsf{NCI}^\star(\equiv)$ and the union of both $\mathsf{NCI}^\star$ over 200 random signatures for each signature size combination of 100 to 1000 concept names and 0%, 50%, and 100% of role names. In addition, in each case we give the number of signatures (out of 200) in which there is a difference between the hybrid module and the minimal 1-depleting module. It can be seen that

- in $\mathsf{NCI}^\star$ and $\mathsf{NCI}^\star(\sqsubseteq)$ the hybrid module almost always coincides with the minimal 1-depleting module (and therefore with the minimal depleting module).
- in $\mathsf{NCI}^\star(\equiv)$, in 50% of all cases the hybrid module coincides with the minimal 1-depleting module. On average the minimal 1-depleting module is less than 0.3% smaller than the hybrid module.
- in all three TBoxes, hybrid modules are only slightly smaller than AMEX-modules.
- in $\mathsf{NCI}^\star(\equiv)$, AMEX-modules are significantly smaller than STAR-modules.
- in $\mathsf{NCI}^\star(\sqsubseteq)$, STAR-modules are slightly smaller than AMEX modules.
- in $\mathsf{NCI}^\star$, AMEX-modules are still significantly smaller than STAR-modules, but less so than in $\mathsf{NCI}^\star(\equiv)$.

In the full version of the paper [5] we also apply the hybrid approach and the lower approximation to the full NCI Thesaurus version 08.09d, which additionally contains role inclusions, domain and range restrictions, and disjointness axioms. The results are very similar to the results for $\mathsf{NCI}^\star$: hybrid modules are on average significantly smaller than STAR modules and are often identical to the minimal 1-depleting module.

Conclusion We have shown that for the NCI Thesaurus one can compute efficiently depleting modules that are consistently very close to the minimal depleting modules and often coincide with the latter. The experiments also show that for TBoxes with many axioms of the form $A \equiv C$, AMEX-modules can be significantly smaller than STAR-modules and that a hybrid approach can lead to significantly smaller modules than 'pure' STAR-modules.

This is only the first step towards a novel systematic evaluation of the quality of upper approximations of modules using lower approximations. It would be of great interest to compute lower approximations for a more comprehensive set of cyclic ontologies and compare them with the upper approximations given by STAR-modules and by the hybrid approach. It is also interesting to investigate n-depleting modules (based on inseparability for interpretations of size at most n) with $n > 1$. These modules can still be extracted by using QBF solvers and the same algorithm; the cost is much higher, though, since the length of the encoding into a QBF is exponential in n.

References

1. F. Baader, D. Calvanes, D. McGuiness, D. Nardi, and P. Patel-Schneider, *The Description Logic Handbook: Theory, implementation and applications*, Cambridge University Press, Cambridge, UK, 2003.
2. M. Benedetti, 'sKizzo: a QBF decision procedure based on propositional skolemization and symbolic reasoning', Technical Report 04-11-03, ITC-irst, (2004).
3. B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler, 'Modular reuse of ontologies: theory and practice', *Journal of Artificial Intelligence Research (JAIR)*, **31**, 273–318, (2008).
4. B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler, 'Extracting modules from ontologies: A logic-based approach', in *Modular Ontologies*, 159–186, Springer, (2009).
5. W. Gatens, B. Konev, and F. Wolter, 'Lower and upper approximations for depleting modules of description logic ontologies', in *ECAI'14*, to appear. (2014).
6. W. Gatens, B. Konev, and F. Wolter, 'Module extraction for acyclic ontologies', in *WoMO*, (2013).
7. M. Horridge and S. Bechhofer, 'The OWL API: A Java API for OWL ontologies', *Semantic Web*, **2**(1), 11–21, (2011).
8. B. Konev, R. Kontchakov, M. Ludwig, T. Schneider, F. Wolter, and M. Zakharyaschev, 'Conjunctive query inseparability of OWL 2 QL TBoxes', in *AAAI*, pp. 221–226. AAAI Press, (2011).
9. B. Konev, C. Lutz, D. Walther, and F. Wolter, 'Formal properties of modularisation', in *Modular Ontologies*, 25–66, Springer, (2009).
10. B. Konev, C. Lutz, D. Walther, and F. Wolter, 'Model-theoretic inseparability and modularity of description logic ontologies', *Artificial Intelligence*, **203**, 66–103, (2013).
11. R. Kontchakov, L. Pulina, U. Sattler, T. Schneider, P. Selmer, F. Wolter, and M. Zakharyaschev, 'Minimal module extraction from DL-Lite ontologies using QBF solvers', in *IJCAI*, pp. 836–841, (2009).
12. R. Kontchakov, F. Wolter, and M. Zakharyaschev, 'Logic-based ontology comparison and module extraction, with an application to DL-Lite', *Artificial Intelligence*, **174**(15), 1093–1141, (2010).
13. C. Lutz and F. Wolter, 'Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$', *Journal of Symbolic Computing*, **45**(2), 194–228, (2010).
14. R. Nortje, K. Britz, and T. Meyer, 'Reachability modules for the description logic SRIQ', in *LPAR*, pp. 636–652, (2013).
15. U. Sattler, T. Schneider, and M. Zakharyaschev, 'Which kind of module should I extract?', in *DL*. CEUR-WS.org, (2009).
16. J. Seidenberg, 'Web ontology segmentation: Extraction, transformation, evaluation', in *Modular Ontologies*, 211–243, (2009).
17. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, eds., H. Stuckenschmidt, C. Parent, and S. Spaccapietra, volume 5445 of *LNCS*, Springer, 2009.
18. B. Suntisrivaraporn, 'Module extraction and incremental classification: A pragmatic approach for ontologies', in *ESWC*, pp. 230–244, (2008).
19. C. Vescovo, P. Klinov, B. Parsia, U. Sattler, T. Schneider, and D. Tsarkov, 'Empirical study of logic-based modules: Cheap is cheerful', in *ISWC*, pp. 84–100, (2013).
20. P. Whetzel, N. Fridam Noy, N. Shah, P. Alexander, C. Nyulas, T. Tudorache, and M. Musen, 'BioPortal', *Nucleic Acids Research*, (Web-Server-Issue), 541–545, (2011).