ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-1079

Querying in $\mathcal{EL}^+$ with Nonmonotonic Rules

Matthias Knorr[1] and **José Júlio Alferes**[1]

Abstract. A general top-down algorithmization for the Well-Founded MKNF Semantics - a semantics for combining rules and ontologies - was recently defined based on an extension of SLG resolution for Logic Programming with an abstract oracle to the parametric ontology language. Here we provide a concrete oracle with practical usage, namely for $\mathcal{EL}^+$ which is tractable for reasoning tasks like subsumption. We show that the defined oracle remains tractable (wrt. data complexity) so that the combined (query-driven) approach of non-monotonic rules with that oracle is tractable as well.

1 Introduction

It is frequently being claimed that integrating open world with closed world reasoning is a key issue for practical large-scale ontology applications. In fact, e.g. [8] describes a large case study about matching patient records for clinical trials criteria containing up to millions of assertions. There, open world reasoning is needed in radiology and laboratory data, because, for example, unless a lab test asserts a negative finding, no arbitrary assumptions about the results can be made. However, in pharmacy data, the closed world assumption can be used to infer that a patient is not on a medication if it is not asserted. Combining ontologies and their underlying Description Logics (DL) with nonmonotonic rules is thus an important problem in Knowledge Representation.

Among the various proposals for combining rules and ontologies, the 3-valued MKNF semantics[2] [4] represents one with several favorable properties: the integration is tight; the data complexity of reasoning on the ontology alone is maintained, in particular, if the ontology is tractable then the combined approach is tractable as well. In [1] a general top-down query procedure which extends SLG resolution with tabling [3] with an abstract oracle to a parametric ontology language was defined for that framework. The oracle receives a query as input and the knowledge already derived and replies with a (possibly empty) set of atoms, defined in the rules, whose truth suffices to prove the initial query.

However, no concrete oracle is provided. It is simply stated that, under some conditions, if such an oracle exists then the integration is possible, and tractable if the ontology language is tractable as well. Here, we provide a concrete oracle, with practical usage, namely for $\mathcal{EL}^+$ a fragment of the light-weight description logics $\mathcal{EL}^{++}$ which for reasoning tasks like subsumption is tractable and part of the W3C recommendations[3] for the Semantic Web.

Our approach includes a preprocessing step that applies the subsumption algorithm[4] for $\mathcal{EL}^+$ to compute all the subsumption relationships contained in the ontology and then removes redundant information w.r.t. answering queries. The resulting reduced ontology is translated into rules and can be directly combined with the set of rules contained in the hybrid knowledge base (KB), and SLG resolution with tabling can be applied for querying. We show that our approach is correct w.r.t. [1] and maintains its tractable data complexity.

2 Preliminaries for $\mathcal{EL}^+$ and Hybrid MKNF

In the following, we use $\mathcal{EL}^+$ also admitting an ABox containing standard concept and role assertions. We refer for syntax and semantics of $\mathcal{EL}^{++}$ to [2], noting that $\mathcal{EL}^+$ is further restricted by not allowing nominals and concrete domains. We also recall the inference problems from [2], only adding one non-standard reasoning task similar to instance checking for concepts. A pair of individuals (a, b) is an *instance of a role* r in ABox $\mathcal{A}$ w.r.t. a CBox $\mathcal{C}$ if $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$ for every common model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{C}$. We also refer to [2] for the subsumption algorithm for $\mathcal{EL}^{++}$, we recall however the intermediate normal form: For CBox $\mathcal{C}$, the notion $\mathsf{BC}_\mathcal{C}$ represents the smallest set of concept descriptions that contains the top concept $\top$, and all concept names used in $\mathcal{C}$ while $\mathsf{R}_\mathcal{C}$ denotes the set of all role names used in $\mathcal{C}$. Then, $\mathcal{C}$ is in *normal form* if

1. all GCIs have one of the following forms, where $C_1, C_2 \in \mathsf{BC}_\mathcal{C}$ and $D \in \mathsf{BC}_\mathcal{C} \cup \{\bot\}$:

(1)	$C_1 \sqsubseteq D$	(3)	$\exists r.C_1 \sqsubseteq D$
(2)	$C_1 \sqcap C_2 \sqsubseteq D$	(4)	$C_1 \sqsubseteq \exists r.C_2$

2. all role inclusions are of the form $r \sqsubseteq s$ or $r_1 \circ r_2 \sqsubseteq s$

The subsumption algorithm itself computes the full class hierarchy corresponding to axioms of the form (1), and also axioms of the form (4).

For notions on hybrid MKNF we refer to [4] and for the top-down procedure to [1], in particular for the definitions of a partial oracle and the transformation doubling the KB.

3 An Oracle for $\mathcal{EL}^+$

In short, our oracle is built as follows: We use the algorithm for subsumption from [2] to compute the complete class hierarchy of the CBox of the ontology to preprocess the ontology. The obtained result together with the CBox is then simplified by removing all statements which are redundant for querying instances. The outcome of that plus the ABox are then transformed into a set of rules which can be used in a top-down manner, by using SLG alone, yielding the desired oracle. We now detail the simplification and the transformation processes needed for this oracle.

[1] CENTRIA, FCT/UNL, Portugal, email: {mknorr, jja}@di.fct.unl.pt
[2] An extension of [7] and based on the logics of minimal knowledge and negation as failure (MKNF) of [6].
[3] http://www.w3.org/TR/owl2-overview/
[4] Like the one included in Pellet (http://clarkparsia.com/pellet/) or CEL (http://lat.inf.tu-dresden.de/systems/cel/)

3.1 Simplifying the Ontology

The subsumption algorithm in [2] computes two mappings, namely S and R, which correspond to axioms of the form (1) for S and (4) for R, in particular all explicit and implicit statements of the form (1). But not all the statements are required for querying for instances.

Example 1 *Consider the following hybrid MKNF KB consisting of a CBox in $\mathcal{EL}^+$, one rule, and some facts.*

$$C \sqsubseteq \exists r.D \qquad G(X) \leftarrow \mathbf{not}\, D(X)$$
$$\exists r.C \sqsubseteq D \qquad C(a). \quad C(b).$$
$$C_1 \sqcap C_2 \sqsubseteq D \qquad r(a,b).$$

One can verify that $G(a)$ is not derivable due to $\mathbf{not}\, D(a)$ and the axiom of the form (3). On the other hand, it is possible to derive $G(b)$ since the axiom of the form (4) does not permit to derive $D(b)$ even if $C(a)$ and $r(a,b)$ are known - $D(b)$ does not hold in all models.

This idea can be formalized in the following definition and it can be shown that answers to instance queries are identical in C and C'.

Definition 1 *Let $\mathcal{C}$ be a CBox in $\mathcal{EL}^+$ and S and R be the mappings obtained from the subsumption algorithm. We obtain the reduced CBox $\mathcal{C}'$ from $\mathcal{C}$ by removing all GCIs of the form (4) from $\mathcal{C}$ and by adding for each $D \in S(C)$ a GCI $C \sqsubseteq D$.*

3.2 Transformation into Rules

We now detail the transformation of the reduced CBox and the ABox into rules limiting to the case where the ontology alone is consistent, but considering inconsistencies introduced by the hybrid KB. Moreover, care must be taken so that whenever something is proven false in the ontology then its default negation also has to hold in the rules.

For this purpose, a doubling of the KB was defined in [1] that introduced a new predicate A^d for each predicate A in the KB and combined the original and the new predicates appropriately such that an inconsistency can be discovered when querying successfully $C(a)$ and its doubled negation $\mathbf{not} C^d(a)$, for some $C(a)$. To maintain this behavior, in Def. 2 we provide in general two rules for each statement we transform. Moreover, in [4] the definitions require to add an axiom $\neg C \sqsubseteq NC$ (resp. $\neg R \sqsubseteq NR$) for each concept name C (resp. role R[5]) which also appears in the head of some rule and a new predicate NC (resp. NR). These axioms do not affect the simplifications in the previous subsection but are not expressible in $\mathcal{EL}^+$. So, we do not transform them into rules but use them implicitly in the cases (i1) to (i3) in the following definition.

Definition 2 *Let $\mathcal{K} = (\mathcal{O}, \mathcal{P})$ be a hybrid MKNF KB and $\mathcal{O}$ a reduced CBox $\mathcal{C}$ in $\mathcal{EL}^+$ plus an ABox $\mathcal{A}$. We define $\mathcal{P}^d_{\mathcal{O}}$ from $\mathcal{O}^+$, where C,D, C_1 and C_2 are concept names, R, S, T are role names, and a, b are individual names, as the smallest set containing:*

(a1) *for each $C(a) \in \mathcal{A}$:*
$C(a) \leftarrow$ *and* $C^d(a) \leftarrow \mathbf{not}\, NC(a)$.

(a2) *for each $R(a,b) \in \mathcal{A}$:*
$R(a,b) \leftarrow$ *and* $R^d(a,b) \leftarrow \mathbf{not}\, NR(a,b)$.

(c1) *for each GCI $C \sqsubseteq D \in \mathcal{C}$:*
$D(X) \leftarrow C(X)$ *and* $D^d(X) \leftarrow C^d(X), \mathbf{not}\, ND(X)$.

(c2) *for each $C_1 \sqcap C_2 \sqsubseteq D \in \mathcal{C}$:*
$D(X) \leftarrow C_1(X), C_2(X)$ *and*
$D^d(X) \leftarrow C_1^d(X), C_2^d(X), \mathbf{not}\, ND(X)$.

(c3) *for each $\exists R.C \sqsubseteq D \in \mathcal{C}$:*
$D(X) \leftarrow R(X,Y), C(Y)$ *and*
$D^d(X) \leftarrow R^d(X,Y), C^d(Y), \mathbf{not}\, ND(X)$.

(r1) *for each RI $R \sqsubseteq S \in \mathcal{C}$: $S(X,Y) \leftarrow R(X,Y)$ and*
$S^d(X,Y) \leftarrow R^d(X,Y), \mathbf{not}\, NS(X,Y)$.

(r2) *for each $R \circ S \sqsubseteq T \in \mathcal{C}$: $T(X,Z) \leftarrow R(X,Y), S(Y,Z)$ and*
$T^d(X,Z) \leftarrow R^d(X,Y), S^d(Y,Z), \mathbf{not}\, NT(X,Z)$.

(i1) *for each $C \sqsubseteq \bot \in \mathcal{C}$:* $NC(X) \leftarrow$.

(i2) *for each $C_1 \sqcap C_2 \sqsubseteq \bot \in \mathcal{C}$:*
$NC_2(X) \leftarrow C_1(X)$ *and* $NC_1(X) \leftarrow C_2(X)$.

(i3) *for each $\exists R.C \sqsubseteq \bot \in \mathcal{C}$:*
$NC(Y) \leftarrow R(X,Y)$ *and* $NR(X,Y) \leftarrow C(Y)$.

[5] Here we use capital letters for all predicates appearing in the DL-part.

The obtained program $\mathcal{P}^d_{\mathcal{O}}$ can then be used in combination with the doubled rules to obtain the correct partial oracle for $\mathcal{EL}^+$, to be integrated in the general top-down procedure of **SLG($\mathcal{O}$)** and we can simplify the interaction by integrating the two programs into one top-down procedure on rules only. It can be shown that this transformation is correct for consistent KBs and that discovery of inconsistencies is maintained. Moreover, the polynomial data complexity is kept.

Theorem 1 *Let $\mathcal{K} = (\mathcal{O}, \mathcal{P})$ be a hybrid MKNF KB with $\mathcal{O}$ in $\mathcal{EL}^+$. Then* **SLG($\mathcal{O}$)** *evaluation of a query in $\mathcal{K}_{\mathcal{EL}^+} = (\emptyset, (\mathcal{P}^d \cup \mathcal{P}^d_{\mathcal{O}}))$ is decidable with data complexity* P.

4 Conclusions

We have presented an approach in the spirit of [1] which allows us to combine non-monotonic rules with a DL in $\mathcal{EL}^+$. Our approach remains tractable w.r.t. data complexity and allows us to discover possible inconsistencies between the rules and the ontology. Future work includes: the (non-trivial) extension to $\mathcal{EL}^{++}$; comparing our work to a procedure which simply transforms the entire ontology into rules without any simplifications; and, based on that comparison, consider the extension to ELP ([5]), a set of rules of DL expressions which extend $\mathcal{EL}^{++}$, whose algorithmization also transforms its expressive rules into datalog rules, a process in which a simplification similar to the one here presented might be useful, too.

REFERENCES

[1] J. J. Alferes, M. Knorr, and T. Swift, 'Queries to hybrid MKNF knowledge bases through oracular tabling', in *The Semantic Web - ISWC 2009*, pp. 1–16. Springer, (2009).

[2] F. Baader, S. Brandt, and C. Lutz, 'Pushing the $\mathcal{EL}$ envelope', in *IJCAI'05: 19th Int. Joint Conf. on Artificial Intelligence*, pp. 364–369. Morgan Kaufmann, (2005).

[3] W. Chen and D. S. Warren, 'Tabled Evaluation with Delaying for General Logic Programs', *Journal of the ACM*, **43**(1), 20–74, (1996).

[4] M. Knorr, J. J. Alferes, and P. Hitzler, 'A coherent well-founded model for hybrid MKNF knowledge bases', in *Proceedings of the 18th European Conference on Artificial Intelligence, ECAI2008*, pp. 99–103. IOS Press, (2008).

[5] M. Krötzsch, S. Rudolph, and P. Hitzler, 'ELP: Tractable rules for OWL 2', in *Proceedings of the 7th International Semantic Web Conference (ISWC-08)*, volume 5318 of *Lecture Notes in Computer Science*, pp. 649–664. Springer, (2008).

[6] V. Lifschitz, 'Nonmonotonic databases and epistemic queries', in *International Joint Conference on Artifical Intelligence, IJCAI'91*, pp. 381–386, (1991).

[7] B. Motik and R. Rosati, 'A faithful integration of description logics with logic programming', in *20th Int. Joint Conf on Artificial Intelligence (IJCAI)*, pp. 477–482, Hyderabad, India, (6–12 2007). AAAI Press.

[8] C. Patel, J. J. Cimino, J. Dolby, A. Fokoue, A. Kalyanpur, A. Kershenbaum, L. Ma, E. Schonberg, and K. Srinivas, 'Matching patient records to clinical trials using ontologies.', in *ISWC/ASWC*, volume 4825 of *Lecture Notes in Computer Science*, pp. 816–829. Springer, (2007).