Default Logics for Plausible Reasoning with Controversial Axioms

Thomas Scharrenbach[1], Claudia d'Amato[2], Nicola Fanizzi[2], Rolf Grütter[1], Bettina Waldvogel[1], and Abraham Bernstein[3]

[1] Swiss Federal Institute for Forest, Snow and Landscape Research WSL Birmensdorf, Switzerland
{thomas.scharrenbach, rolf.gruetter, bettina.waldvogel}@wsl.ch
[2] Università degli Studi di Bari Bari, Italy {claudia.damato, fanizzi}@di.uniba.it
[3] University of Zurich, Department of Informatics Zurich, Switzerland
{bernstein}@ifi.uzh.ch

Abstract. Using a variant of Lehmann's Default Logics and Probabilistic Description Logics we recently presented a framework that invalidates those unwanted inferences that cause concept unsatisfiability without the need to remove explicitly stated axioms. The solutions of this methods were shown to outperform classical ontology repair w.r.t. the number of inferences invalidated. However, conflicts may still exist in the knowledge base and can make reasoning ambiguous. Furthermore, solutions with a minimal number of inferences invalidated do not necessarily minimize the number of conflicts. In this paper we provide an overview over finding solutions that have a minimal number of conflicts while invalidating as few inferences as possible. Specifically, we propose to evaluate solutions w.r.t. the quantity of information they convey by recurring to the notion of entropy and discuss a possible approach towards computing the entropy w.r.t. an ABox.

1 Introduction

In the Semantic Web, knowledge is represented by ontologies expressed in the Web Ontology Language OWL. The current standard, OWL2 [1], defines different profiles all of which have some Description Logics as a rough syntactic variant. These Description Logics (DL) are decidable fragments of first-order logics where knowledge is explicitly expressed in axioms and assertions. DL knowledge bases have well-defined model-theoretic semantics. They allow to express knowledge on different levels of expressivity and enable to infer new conclusions from existing knowledge.

When ontologies evolve or one ontology is mapped to another, contradictions may be introduced that cause the knowledge base as a whole to be inconsistent. Yet, for an inconsistent knowledge base any conclusion—even meaningless ones—becomes trivially true. One cause of inconsistency is given by assertions of concepts that are inferred to be unsatisfiable. Hence, it is desirable to prevent concepts from being inferred unsatisfiable. A knowledge base can become inconsistent for other reasons, but *we propose to start off with conflict-free conceptualizations and apply a method that never infers any concept to be unsatisfiable.*

In the Semantic Web, agents interacting with an ontology assume that both the query and the answer are expressible in OWL2. Furthermore, the answer should have meaningful semantics but not infer conflicts. We therefore demand any formalism allowing for plausible reasoning on controversial information to fulfill the following properties:

1. Permanence: The formalism for knowledge representation is not changed.
2. Coherency: No concept is inferred to be unsatisfiable
3. Autonomy: The procedure shall work automatically.
4. Originality: The original information should be kept.
5. Conservation: As little inferred information as possible shall be lost.

We presented a method for solving unsatisfiable concepts [2] using a combination of Lehmann's Default Logics [3] and Lukasiewicz' Probabilistic Description Logics [4]. Instead of removing (explicit) axioms, we propose to *invalidate those inferences that cause concepts to be inferred unsatisfiable* [5]. While it is possible to reason with *all* information provided, we may still produce contradicting inferences. In this paper we show that minimizing the number of inferences invalidated does not necessarily minimize the number of those conflicts. For finding optimal solutions we propose to evaluate these w.r.t. their information content which requires the definition of the entropy of a solution. We discuss a possible approach towards computing the entropy w.r.t. an ABox and give an outlook on future work.

2 Procedure

For each unsatisfiable concept U of an ontology, its *justifications* $J^k_{U \sqsubseteq \bot}$ [6], i.e. the minimal sets of axioms explaining the conflict, are determined in a first step. *Each of these justifications is split up into two sets*: one that contains all axioms which contain the unsatisfiable concept, $\Gamma^k_{U \sqsubseteq \bot}$ and one that contains all other axioms of that very justification, $\Theta^k_{U \sqsubseteq \bot}$ [2]. Afterwards, the *root unsat justifications* are determined, which are those justifications that do not depend on any other justification [7].

According to the partition scheme of Lehmann's Default Logics, the axioms of the root justifications are put into partitions $\mathcal{U}_0, \ldots, \mathcal{U}_N$ and a separate TBox $\mathcal{T}_\Delta$ such that all concepts in $\mathcal{T}_\Delta \cup \mathcal{U}_n$ are satisfiable for $n = 0, \ldots N$. Thanks to the splitting, we do not have to perform additional satisfiability checks for computing the partition. The resulting Default TBox is a family of (classical) TBoxes: $\mathcal{DT} = (\mathcal{T}_\Delta \cup \mathcal{U}_0, \ldots, \mathcal{T}_\Delta \cup \mathcal{U}_N)$. For such a Default TBox we may either use the inference methods provided by Probabilistic Description Logics [4] or stick to classical reasoning on the single partitions, separately. Either approach defines a deductive closure of the Default TBox as a set of OWL2 axioms, but we prefer the latter approach to change the formalism for reasoning only as little as possible.

Instead of putting all axioms of the root unsat justifications into the partitions, we showed in [5] that we indeed have to put *only two axioms of each root unsat justification into the partitions*—one of each $\Theta^k_{U\sqsubseteq\bot}$ and one of each $\Gamma^k_{U\sqsubseteq\bot}$—while we may put the remaining axioms into $\mathcal{T}_\Delta$. While potentially invalidating less inferences, however, finding partitions may become non-deterministic.

We propose to approximate an optimal solution by a (stochastic) search process: On the one hand, the number of possible solutions is exponential in the number of axioms in the justifications. On the other hand, once the justifications are known, *finding a single valid solution can be performed efficiently*, because the complexity of the approach is dominated by the complexity of finding justifications—a task which has to be performed anyhow.

3 Minimizing Conflicts by Minimizing the Entropy

By invalidating the inferences of the kind $\mathcal{DT} \models U \sqsubseteq \bot$ we ignore the conflicts during reasoning. Yet, inferences such as the co-occurrence of $\mathcal{DT} \models A$ and $\mathcal{DT} \models \neg A$ are still possible but not desired. Hence, a performance measure that assesses the quality of a solution must not only take into account the number of inferences invalidated but, even more important, the number of conflicts still remaining.

Assume the simple TBox $\mathcal{T} = \{B \sqsubseteq A, C \sqsubseteq B, C \sqsubseteq \neg A,\}$ which has two Default TBoxes as potential solutions:

$$\mathcal{DT}^0 \text{ with } \mathcal{T}^0_\Delta = \{C \sqsubseteq B\}, \quad \mathcal{U}^0_0 = \{B \sqsubseteq A\}, \quad \mathcal{U}^0_1 = \{C \sqsubseteq \neg A\}$$
$$\mathcal{DT}^1 \text{ with } \mathcal{T}^1_\Delta = \{C \sqsubseteq \neg A\}, \quad \mathcal{U}^1_0 = \{B \sqsubseteq A\}, \quad \mathcal{U}^1_1 = \{C \sqsubseteq B\}$$

In contrast to the latter, the first Default TBox $\mathcal{DT}^0$ preserves the inference $C \sqsubseteq A$. Yet, in the presence of an ABox that infers the assertion $C(i)$, the assertion $A(i)$ as well as its complement $\neg A(i)$ can be inferred. The second Default TBox $\mathcal{DT}^1$, in contrast, infers only $\neg A(i)$. It is preferred over $\mathcal{DT}^0$, because it contains fewer conflicts than $\mathcal{DT}^1$.

Conflicts potentially reduce the information content of a knowledge base. For minimizing the number of conflicts as well as the number of inferences invalidated we are currently investigating *qualitative measures based on the entropy* of a possible solution. As opposed to methods based on the structure of an ontology [8], we propose that an entropy-measure should take into account the ambiguity of different ABoxes.

In information theory, the entropy measures the average information content of a random variable we are missing when the value of the random variable is not known [9]. If we know the probability mass function p of the random variable X, we may explicitly denote the entropy by $\mathcal{H}(X) = -\sum_{n=0}^{N} p(x_n) \log p(x_n)$. In case $p(x_n) = 0$, then $p(x_n) \log p(x_n) = 0$. We propose to approximate the probability mass function $p_\mathcal{A}$ for the axioms $B \sqsubseteq A \in \mathcal{DT}$ by counting assertions for the concept $(\neg B \sqcup A)$ found by the instance retrieval service of the reasoning process:

$$p_\mathcal{A}(B \sqsubseteq A) = \frac{|\{x \in \mathcal{A}^\mathcal{I} \quad | \quad \mathcal{T}, \mathcal{A} \models (\neg B \sqcup A)(x)\}|}{\sum_{D \sqsubseteq C \in \mathcal{DT}} |\{y \in \mathcal{A}^\mathcal{I} \quad | \quad \mathcal{T}, \mathcal{A} \models (\neg D \sqcup C)(y)\}|}$$

The entropy of a Default TBox $\mathcal{DT}$ measures the information content of its axioms w.r.t. an ABox $\mathcal{A}$: $\mathcal{H}(\mathcal{DT}, \mathcal{A}) = -\sum_{B \sqsubseteq A \in (\mathcal{DT})} p_{\mathcal{A}}(B \sqsubseteq A) \log p_{\mathcal{A}}(B \sqsubseteq A)$. For the Default TBoxes in the example above, we obtain an entropy of $\mathcal{H}(\mathcal{DT}^0) = -\log(1/3)$ and $\mathcal{H}(\mathcal{DT}^1) = -\log(1/2)$ which would make us choose $\mathcal{DT}^1$ rather than $\mathcal{DT}^0$. Our current hypothesis is that *a Default TBox with minimal entropy also minimizes the number of explicit conflicts w.r.t. an ABox.* A prototype implementation is available [4].

4 Conclusion

We recently introduced a framework that never infers any concept to be unsatisfiable while keeping all originally provided information. This allows plausible reasoning on ontologies that possibly contain controversial information—as it is the case for mapped or dynamic ontologies. Finding solutions is non-deterministic and requires optimization techniques that, in turn, require a performance measure for evaluating the quality of possible solutions.
While reasoning ignores conflicts, they are still present in the knowledge base and may lead to sub-optimal results. It was shown that solutions invalidating a minimal number of inferences do not necessarily minimize the number of conflicts still present. For minimizing these we proposed to use an entropy-based performance measure. We provided a definition for the entropy of a solution w.r.t an ABox which is currently being further investigated.

References

1. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S.: OWL 2 Web Ontology Language Primer. W3C Recommendation, W3C (2009)
2. Scharrenbach, T., Grütter, R., Waldvogel, B., Bernstein, A.: Structure Preserving TBox Repair using Defaults. In: 23rd Intl. Workshop on Description Logics. (2010)
3. Lehmann, D.: Another perspective on default reasoning. Ann. Math. Artif. Intell **15** (1995) 61–82
4. Lukasiewicz, T.: Expressive probabilistic description logics. Art. Intell. **172**(6-7) (2008) 852–883
5. Scharrenbach, T., d'Amato, C., Fanizzi, N., Grütter, R., Waldvogel, B., Bernstein, A.: Unsupervised conflict-free ontology evolution without removing axioms. In: 4th International Workshop on Ontology Dynamics (IWOD-2010). (to appear).
6. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Proc. of IJCAI 2003. (2003) 355–362
7. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.: Debugging unsatisfiable classes in owl ontologies. Journal of Web Semantics **3**(4) (2005) 268–293
8. Doran, P.S., Tamma, V., Payne, T.R., Palmisano, I.: An entropy inspired measure for evaluating ontology modularization. In: Proc. of KCAP2009. (2009) 73–80
9. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal **27** 379–423, 623–656

[4] http://www.wsl.ch/info/mitarbeitende/scharren/owl-defaults/