

Demonstrating Pronto: a Non-Monotonic Probabilistic OWL Reasoner

Pavel Klinov¹, Bijan Parsia¹

The University of Manchester, Manchester, M13 9PL, UK
{pklinov,bparsia}@cs.man.ac.uk

Abstract. The demonstration presents Pronto - a prototype of a non-monotonic probabilistic reasoner for very expressive Description Logics. Pronto is built on top of the OWL DL reasoner Pellet, and is capable of performing default probabilistic reasoning in the Semantic Web. It can handle uncertainty in terminological and assertional DL axioms. The demonstration covers Pronto's features and capabilities as well as current challenges and limitations. It describes how an involved realistic problem of breast cancer risk assessment can be formalized in terms of probabilistic reasoning in Pronto¹.

1 Introduction

One of the limitations of current Description Logic (DL) reasoners is the inability to handle uncertain knowledge. It is a serious obstacle to the expansion of the Semantic Web because many domains of human interest contain knowledge that cannot be represented with absolute certainty. One example of an uncertain domain is medicine, in particular, disease diagnosing. Symptoms, causes and consequences of many diseases are uncertain which complicates conceptualization of such domains in formal ontologies and thus restricts machine understanding.

This demonstration presents Pronto - a probabilistic DL reasoner prototype [1]. Pronto is an attempt to provide reasoning services for P-*SHIQ*(D) - a very expressive formalism that is a probabilistic generalization of OWL with the exception of nominals [2]. In addition, the demo will show how Pronto can aid in the probabilistic ontology-based modeling of a realistic medical problem by providing representation and reasoning services for breast cancer risk assessment.

2 Research Background

Pronto provides means for the representation of uncertain ontological statements and offers a collection of reasoning services. It uses the syntax of Lukasiewicz's conditional constraints to express uncertain OWL axioms [2] and Lehmann's lexicographic entailment to perform default probabilistic reasoning [3] [2].

¹ **Acknowledgment:** Pronto was developed when the first author was an intern at Clark & Parsia, LLC (www.clarkparsia.com) during the summer 2007.

2.1 Representation of Uncertain Knowledge

Pronto represents uncertain ontological knowledge using P-*SHIQ*(D) formalism [2] that is a probabilistic generalization of the very expressive DL *SHIQ*(D). It is heavily based on the earlier developed approach to default probabilistic reasoning with *conditional constraints* [4]. In the context of P-*SHIQ*(D), conditional constraints are expressions of the form $(D|C)[l, u]$ where C and D are arbitrary *SHIQ*(D) concepts and $[l, u]$ is a closed interval within $[0, 1]$.

Using conditional constraints Pronto is capable of representing uncertainty in both, TBox and ABox axioms. Informally, constraint $(D|C)[l, u]$ can represent uncertain TBox axioms with the semantics: " C is a sub-concept of D with probability between l and u ". An uncertain ABox axiom $a : C$ can be represented by a constraint $(C|\top)[l, u]$ for probabilistic individual a .

The important feature of P-*SHIN*(D) supported by Pronto is the ability to capture *default* terminological knowledge. That is, TBox constraints are default in the sense that they represent statements that are generally true but might be overridden by more specific statements. The demo will exemplify this possibility.

2.2 Default Probabilistic Reasoning

The main reasoning task in P-*SHIQ*(D) is entailing new conditional constraints, both, terminological and assertional, from probabilistic knowledge bases. Given that standard notion of logical entailment is too weak [4], it has been proposed to use Lehmann's lexicographic entailment that obeys desirable non-monotonic properties [3]. In particular, it allows to resolve conflicts between conditional constraints by preferring some constraints to others [4].

Computing probabilistic entailments involves other reasoning procedures, such as probabilistic satisfiability and probabilistic consistency [2].

3 Technological Basis

One of the principal requirements for Pronto was that the uncertainty could be gradually introduced into existing OWL ontologies and that the existing OWL reasoning services should be retained. To meet that requirement, Pronto was designed on top of the OWL reasoner Pellet [5] that performs reasoning with the classical parts of ontologies

Pronto also employs methods for solving linear optimization and related sub-problems. Following Lukasiewicz's approaches to probabilistic default reasoning and probabilistic logic programming, Pronto reduces the probabilistic satisfiability to the problem of solvability of a corresponding linear system. In addition, lexicographic entailment is reduced to a number of logical entailments each of which is computed by performing linear optimization. Pronto makes use of an LP solver to perform these tasks ².

Finally, Pronto uses OWL 1.1 axiom annotations to associate probability intervals with uncertain OWL axioms.

² Operations Research library, available at: <http://opsresearch.com/OR-Objects/>

4 Demonstration

It has been chosen to use Life Sciences domain, in particular, medical informatics to demonstrate Pronto's capabilities. One relevant medical problem is assessing women risk of developing breast cancer. Given that neither all the relevant risk factors are known nor their impact is sufficiently investigated, the assessment cannot be done with absolute certainty. There have been proposed few statistical approaches to probabilistically estimate the breast cancer risk, for example, Gail model used by the National Cancer Institute (NCI) risk calculator³ [6].

The demo aims to show how Pronto might aid in approaching the same problem by incorporating the statistical knowledge into a cancer ontology. It will present how the problem can be modeled using a probabilistic ontology, and how risk assessment can be performed on the basis of probabilistic default reasoning. Extended version of OWLSight⁴ will serve as the interface to Pronto.

4.1 Probabilistic Model Demonstration

The model of breast cancer risk assessment consists of two major parts - classical OWL ontology and the probabilistic part that represents domain uncertainty. The classical part of the ontology models two types of risk of developing breast cancer. First is *absolute* risk, i.e., the risk that can be measured without the reference to other categories of women. Second, the ontology models relative breast cancer risk, i.e., the risk comparably to an average woman.

The ontology defines risk factors that are relevant to breast cancer. It makes the distinction between the factors that should be known to a woman, e.g., age, family cancer history, breastfeeding, and those that can only be inferred on the basis of other factors or by the examination, e.g., BRCA gene mutation.

Following the assumption that the subjective probabilities representing risk factors for a certain individual can be combined with objective probabilities representing the statistical knowledge, the probabilistic part combines a set of uncertain ABox and TBox axioms. ABox axioms define risk factors that are relevant to a particular individual. TBox axioms model generic probabilistic relationships between the risk factors and classes of women.

It will be shown how to express various dependencies between risk factors. One possibility is to represent how the presence of one risk factor allows to guess on the presence of others. This is the principal way to use *inferred* risk factors.

It will be demonstrated how to capture the impact of combining risk factors, i.e., if they are known to strengthen or weaken each other. Classical part of the ontology provides classes that are combinations of multiple risk factors. The model can define the risk for women having multiple risk factors to be higher (or lower) than if they had just one of the factors. This is possible using the *overriding* feature of the default probabilistic reasoning.

³ <http://www.cancer.gov/bcrisktool>

⁴ Web based ontology browser: <http://pellet.owldl.com/ontology-browser/>

Finally, the ontology contains a number of ABox axioms that represent risk factors for specific individuals. The motivation is that individual women can supply the knowledge about the risk factors that are known to them, e.g., age.

4.2 Probabilistic Reasoning Demonstration

The modeling described above is necessary to reduce the problem of assessing breast cancer risks to the standard lexicographic entailment implemented in Pronto. Risk assessment for a particular woman corresponds to the entailment of an ABox constraint. For example, $(WomanWithBRCInLongTerm|\top)[0.6, 0.8]$ implies that some woman's risk of developing cancer in life time is 60%-80%. The reasoning will be demonstrated on a number of test probabilistic individuals.

It will also be presented how Pronto justifies the results of the risk assessment by generating the *explanations* for the entailments. In particular, it can retrieve exactly those risk factors and generic statistical axioms that caused the inference for a particular woman and filter out all the irrelevant risk factors. In addition to being useful for end users, this capability can aid the model developers in testing the accuracy and adequacy of their model.

5 Conclusion

Although the demo does not pretend to cover all the aspects of default probabilistic reasoning in the Semantic Web, it is expected to help the attendees learn the following important things:

- Main features, e.g., representational and reasoning capabilities of Pronto.
- How Pronto can help in modeling real life problems that involve uncertainty.
- How current OWL ontologies can be reused in probabilistic models.
- Current challenges and limitations of probabilistic DL reasoning.

References

1. Klinov, P., Parsia, B., Mazlack, M.J.: Pronto - a non-monotonic probabilistic description logic reasoner. Submitted to ESWC (2008)
2. Lukasiewicz, T.: Probabilistic description logics for the semantic web. Technical Report Nr. 1843-06-05, Institut für Informationssysteme, Technische Universität Wien (2007)
3. Lehmann, D.: Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence* **15(1)** (1995) 61–82
4. Lukasiewicz, T.: Probabilistic default reasoning with conditional constraints. *Annals of Mathematics and Artificial Intelligence* **34(1-3)** (2002) 35–88
5. Sirin, S., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y.: Pellet: A practical OWL-DL reasoner. Technical Report CS 4766, University of Maryland, College Park (2005)
6. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Shairer, C., Mulvihill, J.J.: Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* **81(25)** (1989) 1879–1886