Representing Probabilistic Relations in RDF

Yoshio Fukushige
Network Development Center
Matsushita Electric Industrial Co., Ltd.
4-5-15 Higashi-shinagawa, Shinagawa-ku, Tokyo 140-8632, Japan
fukushige.yoshio@jp.panasonic.com

Abstract

Probabilistic inference will be of special importance when one needs to know how much we can say with what all we know given new observations. Bayesian Network is a graphical probabilistic model with which one can represent probabilistic relations intuitively and several effective algorithms for inference are developed. This paper reports a now ongoing work in its design stage which provides a vocabulary for representing probabilistic knowledge in a RDF graph which is to be mapped to a Bayesian Network to do inference on it.

1 Introduction

In the real world, especially in the scientific fields like Life Science, or in applications like contents classification and recommendation, it is often the case that relationship between resources holds probabilistically, or we can make statements only with uncertainty. Such relationships can be well described with probabilistic model. And probabilistic inference will be of special importance when one needs to know how much we can say with what we know incompletely.

In this position paper, I report my ongoing work which provides a vocabulary for representing probabilistic knowledge in a RDF graph. I introduce Bayesian Networks and list the requirements for the representing language and candidate vocabulary.

2 Bayesian Network

A Bayesian Network(BN) (Pearl 88) [1] is a graphical model to represent probabilistic relations. It is a directed acyclic graph (DAG), representing probabilistic dependencies among a set of propositions. A node represents a set of exhaustive and exclusive set of propositions (called 'variable' or 'partition'). A link represents a direct dependency between the variables. Each node is accompanied with a conditional probability table (CPT) that represents the probabilistic relationship between the variables. The posterior probability distributions ("beliefs")for each variable could be calculated by propagating beliefs.

Figure 1 shows an example illustration of a BN (CPTs are not shown). It has 5 nodes and 5 links connecting them.

Advantages of employing BN are among others:

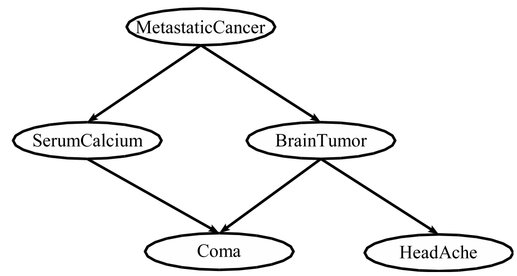

Figure 1: example Bayesian Network

- the relations are expressed by a graph, which is a familiar notion in the Semantic Web, and thus intuitive and easy to understand
- effective calculation algorithms (including simulation methods) have been developed

3 Requirements for the representation language

The aim of this work is not to just represent Bayesian networks in the Semantic Web, but to get a language (or extension vocabulary) which can describe probabilistic relations in a way that is Semantic Web compatible and easy to map to a BN. It is to put together the distributed information in the Semantic Web, and do probabilistic inference in the BN.

The components of a BN are nodes and links and CPT's attached to the nodes. A node represents a set of exhaustive and mutually exclusive propositions (called partition).

The representation language should be able to express:

- a partition, i.e. a set of exhaustive and exclusive propositions
- propositions in such a way that they are distinguished from ground facts
- a probability with which a proposition holds with/without certain conditions

4 Vocabulary for RDF representation

RDF is a W3C standard as one of the fundamental building blocks of the Semantic Web. By representing probabilistic relations in RDF, one gets advantage of reusing existing vocabularies and tools for RDF processing, and one can treat the probabilistic relations themselves as resources and incorporate them into RDF graphs.

To provide a vocabulary that satisfy the requirements above, I introduced the following classes and predicates:

classes `prob:Partition`, `prob:ProbabilsticStatement`, `prob:Clause`, `prob:Probability`,

predicates `prob:predicate`, `prob:hasProbability`, `prob:condition`, `prob:case`, `prob:about`

Details are omitted because of the limit of the space. Details are to be available at<http://www.w3.org/2005/08/08-prob/>.

Points to note are:

- Conditions are linked with `prob:Partition`'s, not with each cases.
- Introduction of `prob:Clause`'s
  `prob:Clause` is a generalization of the RDF reification. `prob:Clause`represents has one `prob:predicate` and zero or more 'terms.' (cf. the pattern 2 of N-ary relationship representations in [9])

The following is an example graph which represents a probabilistic relation: "if cond0, then case1 has probability prob1 and case2 has probability prob2" (in a Turtle [2] serialization)

```
[a prob:Partition;
 prob:condition :cond0;
 prob:case
  [a prob:ProbabilisticStatement;
   prob:about :case1;
   prob:hasProbability :prob1],
  [a prob:ProbabilisticStatement;
   prob:about :case2;
   prob:hasProability :prob2].
].
```

5 Related works

(Ding & Peng 2004) [4] and (Gu, Pung & Zhang 2004) [5] are close works to this. They proposes to augment OWL to allow probabilistic markups, and provides a set of transformation rules to convert the probabilistically annotated OWL ontology into a BN.

(Holi & Hyvönen 2004) [6] is an attempt to express and calculate overlapping of concepts in RDF(S) and OWL using BN.

Works in combinatorial use of BNs and Description Logics includes, among others, (Koller, Levy & Pfeffer 1997) [7] which presents P-CLASSIC; a probabilistic version of the Description Logic CLASSIC, and (Yelland 2000) [8] which incorporates Description Logics into BNs.

While others address T-Box knowledge, (Fukushige 2004)[3] proposes a method to encode probabilistic relations in A-Box, which is in the same direction with this work.

6 Conclusion and future works

This position paper reported a ongoing work which provides a vocabulary for representing probabilistic knowledge in a RDF graph.

Open issues (other than implementing issues) include:

- Relationship with rule languages
- How to standardize Query Languages against BN store
- How to learn BNs from data or/and partial description in RDF.
- How to deal with / avoid cyclic probabilistic description in RDF
- How to deal with continuous probabilistic distributions
- Examination of computational complexity

References

[1] Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, 1988

[2] Becket, D.: Turtle - Terse RDF Triple Language, http://www.ilrt.bris.ac.uk/discovery/2004/01/turtle/

[3] Fukushige, Y.:Representing Probabilistic Knowledge in the Semantic Web, position paper for the W3C Workshop on Semantic Web for Life Sciences, http://www.w3.org/2004/09/13-Yoshio/PositionPaper.html, 2004

[4] Ding, Zh., Peng, Y.: A Probabilistic Extension to Ontology Language OWL, Proc. of the 37th Hawaii International Conference in System Sciences, 2004

[5] Gu, T., Pung, H.K., Zhang, D.Q.: A Bayesian Approach for Dealing with Uncertain Contexts, Proc. of the Second International Conference on Pervasive Computing (Pervasive 2004) ,"Advances in Pervasive Computing", Austrian Computer Society, vol. 176, 2004

[6] oli, M., Hyvönen, E.: A Method for Modeling Uncertainty in Semantic Web Taxonomies, Proceedings of the 13th international World Wide Web conference, pp.296-297, 2004

[7] Koller, D., Levy, A., Pfeffer, A.: P-CLASSIC: A tractable probabilistic description logic, Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97), pp.390-397, 1997

[8] Yelland, P.M.: An Alternative Combination of Bayesian Networks and Description Logics, Proceedings of the KR2000 Conference, pp.225-234, 2000

[9] Noy, N., Rector, A.: Defining N-ary Relations on the Semantic Web: Use With Individuals, W3C Working Draft, http://www.w3.org/TR/2004/WD-swbp-n-aryRelations-20040721/, 2004