A reasoner for generalized Bayesian dl-programs

Livia Predoiu

Institute of Computer Science, University of Mannheim, Germany
livia@informatik.uni-mannheim.de
1 Motivation

2 generalized dl-programs

3 generalized Bayesian dl-programs
   - Using generalized Bayesian dl-programs for reasoning with Ontologies and uncertain mappings

4 Reasoner Architecture

5 Conclusions & Outlook
Motivation

- Knowledge in the Semantic Web is provided on independent peers
- Domains overlap, but no (global) reference ontology
- Mappings need to be created *dynamically* and *automatically*.
- Automatically created mappings are *uncertain hypotheses* (oversimplifying, erroneous)
- Uncertainty can be modelled with probabilities
Motivation cont’d

Integrated reasoning with ontologies and uncertain mappings provides

- insight into the (un)certainty of the reasoning results
- a natural ranking method over the reasoning results
Syntax:

- A generalized dl-program $KB = (L, P)$ consists of a description logic knowledge base $L$ in the DLP fragment and a Datalog program $P$

Semantics:

- $L$ is translated into its Datalog equivalent $L'$
- $L' \cup P$ are interpreted as Datalog programs
Syntax

- A generalized Bayesian dl-program is a 4-tuple \( KB = (L, P, \mu, Comb) \) where
  - \((L, P)\) is a generalized dl-program
  - \(\mu(r, v)\) is a probability function over all truth valuations \(w\) of the head atom associated with each rule \(r\) in \(\text{ground}(P)\) and every truth valuation \(v\) of the body atoms of \(r\)
  - \(Comb\) is a combining rule, which defines how rules of \(r \in \text{ground}(P)\) with same head atom can be combined.
Semantics

- each generalized Bayesian dl-program $KB = (L, P, \mu, Comb)$ encodes the structure of a Bayesian Network $BN$

- Translation from $KB$ to $BN$
  - $(L, P)$ is translated into its Datalog equivalent $D = L' \cup P$
  - a ground atom $a$ is active iff it belongs to the canonical model of $D$; $r \in ground(D)$ is active iff all its atoms are active
  - every active atom corresponds to a node in $BN$
  - every active rule corresponds to direct influence relations between the atoms involved and is translated to arcs in $BN$
  - $\mu$ is the conditional probability density for each active rule
  - for at least 2 active rules with same head, the combining rule $Comb$ generates a joint conditional distribution from the individual ones of the involved rules.
Using generalized Bayesian dl-programs for reasoning with Ontologies and uncertain mappings

Intuitively

- \( L = O_1 \cup O_2 \) encodes the ontologies
- \( P, \mu \) encodes the mappings

Mappings:

- \( Q(O_i) \) denotes the matchable elements of the ontology \( O_i \)
- **Matching**: Given two ontologies \( O \) and \( O' \), determine correspondences between \( Q(O) \) and \( Q(O') \).
- **Correspondences** are 5-tuples \((id, e, e', r, n)\) such that
  - \( id \) is a unique identifier;
  - \( e \in Q(O) \) and \( e' \in Q(O') \);
  - \( r \in R \) is a semantic relation (here: implication);
  - \( n \) is a degree of confidence in the correctness. (here: probabilities)
Ontology 1
(1) \text{Technical\_Report} \sqsubseteq \forall \text{keyword}.\text{Topic} \sqcap \forall \text{author}.\text{Person};
(2) \text{Book} \sqsubseteq \text{Publication};
(3) \text{Article} \sqsubseteq \text{Publication};
(4) \text{Collection} \sqsubseteq \text{Publication};
(5) \text{Publication} \sqsubseteq \forall \text{keyword}.\text{Topic} \sqcap \forall \text{author}.\text{Person}.

Ontology 2
(1) \text{Paper} \sqsubseteq \text{Publication};
(2) \text{Proceedings} \sqsubseteq \text{Publication};
(3a) \top \sqsubseteq \forall \text{includes}.\text{Paper};
(3b) \top \sqsubseteq \forall \text{includes}^{-1}.\text{Proceedings};
(4a) \top \sqsubseteq \forall \text{published\_by}.\text{Publisher};
(4b) \top \sqsubseteq \forall \text{published\_by}^{-1}.\text{Publication};
(5a) \top \sqsubseteq \forall \text{about}.\text{Subject};
(5b) \top \sqsubseteq \forall \text{about}^{-1}.\text{Publication};
(6a) \top \sqsubseteq \forall \text{author}.\text{Person};
(6b) \top \sqsubseteq \forall \text{author}^{-1}.\text{Publication}.

Translation of Ontology 1
(1a) \text{Topic}(y) \leftarrow \text{Technical\_Report}(x) \land \text{keyword}(x, y);
(1b) \text{Person}(y) \leftarrow \text{Technical\_Report}(x) \land \text{author}(x, y);
(2) \text{Publication}(x) \leftarrow \text{Book}(x);
(3) \text{Publication}(x) \leftarrow \text{Article}(x);
(4) \text{Publication}(x) \leftarrow \text{Collection}(x);
(5a) \text{Topic}(y) \leftarrow \text{Publication}(x) \land \text{keyword}(x, y);
(5b) \text{Person}(y) \leftarrow \text{Publication}(x) \land (x, y).

Translation of Ontology 2
(1) \text{Publication}(x) \leftarrow \text{Paper}(x);
(2) \text{Publication}(x) \leftarrow \text{Proceedings}(x);
(3a) \text{Paper}(y) \leftarrow \text{includes}(x, y);
(3b) \text{Proceedings}(x) \leftarrow \text{includes}(x, y);
(4a) \text{Publisher}(y) \leftarrow \text{published\_by}(x, y);
(4b) \text{Publication}(x) \leftarrow \text{published\_by}(x, y);
(5a) \text{Subject}(y) \leftarrow \text{about}(x, y);
(5b) \text{Publication}(x) \leftarrow \text{about}(x, y);
(6a) \text{Person}(y) \leftarrow \text{author}(x, y);
(6b) \text{Publication}(x) \leftarrow \text{author}(x, y).

Mappings
(1) \mathcal{O}_1' : \text{Publication}(x) \overset{(0.9, 0.2)}{\leftrightarrow} \mathcal{O}_2' : \text{Publication}(x);
(2) \mathcal{O}_1' : \text{Article}(x) \overset{(0.7, 0.2)}{\leftarrow} \mathcal{O}_2' : \text{Paper}(x);
(3) \mathcal{O}_1' : \text{Person}(x) \overset{(0.9, 0.2)}{\leftarrow} \mathcal{O}_2' : \text{Person}(x);
(4) \mathcal{O}_1' : \text{Collection}(x) \overset{(0.7, 0.2)}{\leftarrow} \mathcal{O}_2' : \text{Proceedings}(x);
(5) \mathcal{O}_1' : \text{keyword}(x, y) \overset{(0.7, 0.2)}{\leftarrow} \mathcal{O}_2' : \text{about}(x, y);
(6) \mathcal{O}_1' : \text{author}(y, x) \overset{(0.7, 0.2)}{\leftarrow} \mathcal{O}_2' : \text{author}(x, y).
Example Network:

Two types of queries:
- ground queries
- non-ground queries (information retrieval)
Reasoner Architecture

A reasoner for generalized Bayesian dl-programs

URSW 2008
Livia Predoiu
Conclusions and Outlook:

- ongoing implementation of a reasoner for integrated reasoning with deterministic ontologies and probabilistic mappings

Next steps

- finish implementation
- experiments
- add support for and experiment with more than two ontologies (mapping chains)
- enhance the efficiency of the reasoning procedure (e.g. by approximation)
  - logical
  - probabilistic
  - Combination of logical & probabilistic
- add further matchers & trust
- add mapping conflict resolution