Epistemic and Statistical Probabilistic Ontologies

Fabrizio Riguzzi, Elena Bellodi, Evelina Lamma, Riccardo Zese

ENDIF – University of Ferrara, Italy

fabrizio.riguzzi@unife.it, elena.bellodi@unife.it, evelina.lamma@unife.it, riccardo.zese@student.unife.it
Outline

1. Representing Uncertainty
2. Probabilistic Logic Languages
3. Distribution Semantics
4. Probabilistic Ontologies under DISPONTE semantics
5. Epistemic vs Statistical Probability
6. Inference in Probabilistic Ontologies
7. Examples
8. Query answering for Probabilistic OWL DL Ontologies
9. Related works
10. Conclusions
11. References
Representing Uncertainty

Uncertainty Representation

- Semantic Web
  - Incompleteness or uncertainty are intrinsic of much information on the World Wide Web
  - Most common approaches: probability theory, Fuzzy Logic

- Logic Programming
  - Uncertain relationships among entities characterize many complex domains
  - Most common approaches: probability theory → Distribution Semantics (Sato, 1995)[6]
    - It underlies **Probabilistic Logic Languages** (ICL, PRISM, ProbLog, LPADs), ...
    - They define a probability distribution over normal logic programs
    - The distribution is extended to a joint distribution over worlds and queries
    - The probability of a query is obtained from this distribution by marginalization
Example: Program $T$, development of an epidemic or pandemic, if somebody has the flu and the climate is cold.

$C_1 = \text{epidemic} : -\text{flu}(X), \text{epid}(X), \text{cold}$.

$C_2 = \text{pandemic} : -\text{flu}(X), \neg \text{epid}(X), \text{pand}(X), \text{cold}$.

$C_3 = \text{flu}(\text{david})$.

$C_4 = \text{flu}(\text{robert})$.

$F_1 = 0.7 :: \text{cold}$.

$F_2 = 0.6 :: \text{epid}(X)$.

$F_3 = 0.3 :: \text{pand}(X)$.

- Distributions over facts
- Worlds obtained by selecting or not every grounding of each probabilistic fact
Distribution Semantics

Case of no function symbols: finite set of groundings of each probabilistic fact $F$

- a ProbLog fact $p :: F$ is interpreted as $F : p \lor \text{null} : 1 - p$.

Atomic choice: selection of a value for a grounding of a probabilistic fact $F$: $(F_i, \theta_j, k)$, where $\theta_j$ is a substitution grounding $F_i$ and $k \in \{0, 1\}$.

Composite choice $\kappa$: consistent set of atomic choices

- $\kappa = \{(F_2, \{X/david\}, 1), (F_2, \{X/david\}, 0)\}$ not consistent

- Boolean random variable $X_{ij}$, for each $(F_i, \theta_j, k)$
Distribution Semantics

- **Selection** $\sigma$: a total composite choice (one atomic choice for every grounding of each probabilistic fact)

$$
\sigma = \{(F_1, \{\}, 1), (F_2, \{X/david\}, 1), (F_3, \{X/david\}, 1),
(F_2, \{X/robert\}, 0), (F_3, \{X/robert\}, 0)\}
$$

A selection $\sigma$ identifies a logic program $w_\sigma$ called **world**:

$$
w_\sigma = T_C \cup \{F_i\theta_j | (F_i, \theta_j, 1) \in \sigma\}, \text{ where } T_C \text{ is the set of certain rules of } T \text{ (a normal logic program)}
$$

The probability of $w_\sigma$ is

$$
P(w_\sigma) = P(\sigma) = \prod_{(F_i, \theta_j, 1) \in \kappa} p_i \prod_{(F_i, \theta_j, 0) \in \kappa} (1 - p_i)
$$

For the example above:

$$
P(w_\sigma) = 0.7 \times 0.6 \times 0.3 \times (1 - 0.6) \times (1 - 0.3)
$$

Finite set of worlds: $W_T = \{w_1, \ldots, w_m\}$

$P_W$ distribution over worlds: $\sum_{w \in W_T} P(w) = 1$
Distribution Semantics

Conditional probability of a query Q: \( P(Q|w) = 1 \) if \( w \models Q \) and 0 otherwise

Joint distribution of the worlds and queries \( P(Q,w) \):

\[
P(Q, w) = P(Q|w)P(w)
\]

\[
P(Q) = \sum_{w \in W_T} P(Q, w) = \sum_{w \in W_T} P(Q|w)P(w) = \sum_{w \in W_T: w \models Q} P(w)
\]

In the example T has 5 Boolean random variables

- \( F_1 \rightarrow X_{11} \) (1 grounding)
- \( F_2 \rightarrow X_{21} \) and \( X_{22} \) (2 groundings)
- \( F_3 \rightarrow X_{31} \) and \( X_{32} \) (2 groundings)

and thus 32 worlds. The query epidemic is true in 5 of them. By the sum of their probability, we obtain \( P(\text{epidemic}) = 0.588 \).
**Idea:** annotate each axiom of an ontology with a probability and assume that each axiom is independent of the others (see URSW2011)

**DISPONTE** semantics exploits the translation of a probabilistic ontology into a first order logic theory

A probabilistic ontology defines thus a distribution over normal theories (worlds) obtained by including an axiom in a world with a probability given by the annotation

The probability of a query is again computed from this distribution with marginalization:

\[
P(Q) = \sum_w P(Q, w) = \sum_w P(Q|w)P(w) = \sum_{w: w|= Q} P(w)
\]

What’s new w.r.t. URSW2011?
We can specify two kinds of probability for OWL DL axioms, under the DISPONTE semantics:

1. **Epistemic probability**
   - $p ::_e E$ where $p \in [0, 1]$ and $E$ is any (TBox, RBox or ABox) axiom
   - $p \rightarrow$ represents our degree of belief in axiom $E$
   - e.g., $p ::_e C \sqsubseteq D$ represents the fact that we believe in the truth of $C \sqsubseteq D$ with probability $p$.

2. **Statistical probability**
   - $p ::_s E$ where $p \in [0, 1]$ and $E$ is a TBox or RBox axiom
   - $p \rightarrow$ represents information regarding random individuals from certain populations
   - e.g., $p ::_s C \sqsubseteq D$ means instead that a random individual of class $C$ has probability $p$ of belonging to $D$.

3. Any unannotated axiom $E$ is certain.
Observations

1. **Epistemic probability**
   - \( p ::_e C \sqsubseteq D \) represents the fact that we believe in the truth of \( C \sqsubseteq D \) with probability \( p \)
   - If two individuals \( i \) and \( j \) belong to class \( C \), the probability that they both belong to \( D \) under the epistemic probability is \( p \)

2. **Statistical probability**
   - \( p ::_s C \sqsubseteq D \) means that a random individual of class \( C \) has probability \( p \) of belonging to \( D \)
   - If two individuals \( i \) and \( j \) belong to class \( C \), thus the probability that they both belong to \( D \) under statistical probability interpretation is \( p \times p \).
Explanations for a query

- Each *atomic choice* is a triple \((F_i, \theta_j, k)\)
  - \(F_i\) is the formula obtained by translating the \(i\)-th axiom \(E_i\)
  - \(\theta_j\) is a substitution
  - \(k \in \{0, 1\}\). \(k\) indicates whether \((F_i, \theta_j, k)\) is chosen to be included in a world \((k = 1)\) or not \((k = 0)\)
- If \(F_i\) is obtained from an unannotated axiom, then \(\theta_j = \emptyset\) and \(k = 1\)
- If \(F_i\) is obtained from an axiom of the form \(p ::_e E_i\), then \(\theta_j = \emptyset\)
- If \(F_i\) is obtained from an axiom of the form \(p ::_s E_i\), then \(\theta_j\) instantiates the variables occurring in the logical translation of axiom \(E_i\).
- Boolean random variables \((X_{ij})\) are, again, associated to (instantiations of) logical formulas \((F_i)\) by substitution \(\theta_j\)
Inference and Query answering

- Similarly to the case of probabilistic logic programming, the probability of a query $Q$ given a probabilistic ontology $O$ can be computed by first finding the explanations for $Q$ in $O$
- **Explanation**: subset of axioms of $O$ that is sufficient for entailing $Q$
- All the explanations for $Q$ must be found, corresponding to all ways of proving $Q$
- Probability of $Q \rightarrow$ probability of the DNF formula

$$F(Q) = \bigvee_{e \in E_Q} \left( \bigwedge_{(F_i, \theta_j, 1) \in e} X_{ij} \bigwedge_{(F_i, \theta_j, 0) \in e} \overline{X_{ij}} \right)$$

where $E_Q$ is the set of explanations and $X_{ij}$ is a random variable with $k = 1$ and probability $p_i$ (and $\overline{X_{ij}}$ is a random variable with $k = 0$ and probability $(1 - p_i)$)
- We exploit an underlying DL reasoner for computing explanations, and Binary Decision Diagrams for making these explanations mutually incompatible.
Example 1.1 - *people+pets* ontology

- *fluffy* is a *Cat* with (epistemic) probability 0.4 and *tom* is a *Cat* with probability 0.3; *Cats* are *Pets* with (epistemic) probability 0.6

\[
\begin{align*}
0.4 & \therefore_e fluffy : Cat \\
0.3 & \therefore_e tom : Cat \\
0.6 & \therefore_e Cat \sqsubseteq Pet
\end{align*}
\]

- Everyone who has a pet animal (*hasAnimal Pet*) is a *PetOwner*, *kevin* has two animals, *fluffy* and *tom*

\[
\exists hasAnimal Pet \sqsubseteq PetOwner
\]

\[
(kevin, fluffy) : hasAnimal
\]

\[
(kevin, tom) : hasAnimal
\]

- \( Q = kevin : PetOwner \) has two (mutually exclusive) explanations: 
  \{ (1), (3), not (2) \} and \{ (2), (3) \}

\[
P(Q) = 0.4 \times 0.6 \times (1 - 0.3) + 0.3 \times 0.6 = 0.348
\]
Example 1.2 - people+pets ontology

- If we replace epistemic with statistical probability in axiom:

\[
0.6 \triangledown_s \text{Cat} \sqsubseteq \text{Pet}
\]  

(7)

- then for \(Q = \text{kevin: PetOwner}\) we have instances of axiom (7) in (mutually exclusive) explanations: \{(1), (7)/fluffy, not (2)\}, \{(1), (7)/fluffy, (2), not ((7)/tom)\} and \{(2),(7)/tom\}

\[
P(Q) = 0.4 \times 0.6 \times (1 - 0.3) + 0.4 \times 0.6 \times 0.3 \times (1 - 0.6) + 0.3 \times 0.6 = 0.3768
\]
BUNDLE system

Binary decision diagrams for Uncertain reasoNing on Description Logic thEories

- BUNDLE performs inference over probabilistic OWL DL ontologies that follow the DISPONTE semantics.
- It exploits an underlying ontology reasoner able to return all explanations for a query, such as Pellet [7].
- Explanations for a query in the form of a set of sets of axioms.
- Pellet has been extended to record not only used axioms, but their instantiations too, in order to correctly handle statistical probability.
- BUNDLE performs a double loop over the set of explanations and over the set of (instantiated) axioms in each explanation, in which it builds a BDD representing the set of explanations.
- JavaBDD library for the manipulation of BDDs.
- BUNDLE has been implemented in Java and will be available for download from http://sites.unife.it/bundle.
(Laskey, and da Costa, 2005) [4] proposed PR-OWL, an upper ontology that provides a framework for building probabilistic ontologies and allows to use the first-order probabilistic logic MEBN; instead we tried to provide a minimal extension to DL.

(Koller et al., 1997) [3] present a probabilistic description logic based on Bayesian networks that deals with statistical terminological knowledge, but, differently from us, does not allow probabilistic assertional knowledge about concept and role instances.

(Jaeger, 1994) [2] allows assertional knowledge about concept and role instances together with statistical terminological knowledge. We can also represent epistemic information with terminological knowledge.
(Ding, and Peng, 2004)[1] propose a probabilistic extension of OWL that admits a translation into Bayesian networks. The semantics assigns a probability distribution \( P(i) \) over individuals and a probability to a class \( C \) as \( P(C) = \sum_{i \in C} P(i) \), while we assign a probability distribution over theories.

In (Nilsson, 1986)’s probabilistic logic [5]: a probabilistic interpretation \( Pr \) defines a probability distribution over the set of interpretations \( \mathcal{I} \). The probability of a logic formula \( \phi \) according to \( Pr \), denoted \( Pr(\phi) \), is the sum of all \( Pr(I) \) such that \( I \in \mathcal{I} \) and \( I \models \phi \)

while a probabilistic knowledge base may have multiple models that are probabilistic interpretations, a probabilistic ontology under the distribution semantics defines a single distribution over interpretations.

Worth to mention also alternative approaches to modeling imperfect knowledge in ontologies, based on fuzzy logic.
Conclusions and future works

- DISPONTE semantics for probabilistic ontologies inspired by the distribution semantics of probabilistic logic programming
  - two ways (epistemic and statistical) to specify the probability of the axioms of an ontology
- The problem of inference in DISPONTE remains decidable if it was so in the underlying description logic
  - BUNDLE system able to compute the probability of queries from an uncertain OWL DL ontology
  - Computing explanations of a query is exponential in time
  - Computing the probability of a DNF formula of independent Boolean random variables is a \#P-complete problem (\#P over the number of computed explanations)

Future works

- Extension to treat different degrees of statistical probability, by choosing which variables in the logical translation are subject to instantiation and which not (as proposed by (Halpern, 1990))
Thanks.

Questions?
Zhongli Ding and Yun Peng.
A probabilistic extension to ontology language OWL.

Manfred Jaeger.
Probabilistic reasoning in terminological logics.

Daphne Koller, Alon Y. Levy, and Avi Pfeffer.
P-classic: A tractable probabilistic description logic.

Of starships and klingons: Bayesian logic for the 23rd century.

Probabilistic logic.

Taisuke Sato.
A statistical learning method for logic programs with distribution semantics.

Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz.
Pellet: A practical OWL-DL reasoner.