Probabilistic Relational Modeling

Statistical Relational AI

Tutorial at ICCS 2019

Marcel Gehrke, University of Lübeck

Thanks to Ralf Möller for making his slides publicly available.
Agenda: Probabilistic Relational Modeling

• Application
  • Information retrieval (IR)
  • Probabilistic Datalog
• Probabilistic relational logics
  • Overview
  • Semantics
  • Inference problems
• Scalability issues
  • Proposed solutions

Goal: Overview of central ideas

*We would like to thank all our colleagues for making their slides available (see some of the references to papers for respective credits). Slides are almost always modified.
Application

- Probabilistic Datalog for information retrieval [Fuhr 95]:

  0.7 \text{term}(d1,ir).
  0.8 \text{term}(d1,db).
  0.5 \text{link}(d2,d1).
  \text{about}(D,T) :\neg \text{term}(D,T).
  \text{about}(D,T) :\neg \text{link}(D,D1), \text{about}(D1,T).

- Query/Answer

  :\neg \text{term}(X,ir) \land \text{term}(X,db).

  \begin{align*}
  X &= 0.56 \ d1
  \end{align*}

Application: Probabilistic IR

• Probabilistic Datalog

0.7 \text{term}(d1,\text{ir}).
0.8 \text{term}(d1,\text{db}).
0.5 \text{link}(d2,d1).

\text{about}(D,T) :- \text{term}(D,T).
\text{about}(D,T) :- \text{link}(D,D1), \text{about}(D1,T).

• Query/Answer

q(X) :- \text{term}(X,\text{ir}).
q(X) :- \text{term}(X,\text{db}).

:-q(X)
X = 0.94\ d1
Application: Probabilistic IR

• Probabilistic Datalog

  0.7 term(d1,ir).
  0.8 term(d1,db).
  0.5 link(d2,d1).

  about(D,T) :- term(D,T).

  about(D,T) :- link(D,D1), about(D1,T).

• Query/Answer

  :- about(X,db).

  X = 0.8 d1;
  X = 0.4 d2
Application: Probabilistic IR

• Probabilistic Datalog

0.7 \text{term}(d1,ir).
0.8 \text{term}(d1,db).
0.5 \text{link}(d2,d1).

\text{about}(D,T):=-\text{term}(D,T).
\text{about}(D,T):=-\text{link}(D,D1),\text{ about}(D1,T).

• Query/Answer

:- \text{about}(X,db) & \text{about}(X,ir).

X = 0.56\ d1;
X = 0.28\ d2 \ # NOT naively 0.14 = 0.8*0.5*0.7*0.5
Solving Inference Problems

• QA requires proper probabilistic reasoning

• Scalability issues
  • Grounding and propositional reasoning?
  • In this tutorial the focus is on lifted reasoning in the sense of [Poole 2003]
    • Lifted exact reasoning
    • Lifted approximations

• Need an overview of the field: Consider related approaches first

Application: Probabilistic IR

• Uncertain Datalog rules: Semantics?

0.7 \text{term}(d1,\text{ir}).
0.8 \text{term}(d1,\text{db}).
0.5 \text{link}(d2,d1).
0.9 \text{about}(D,T):- \text{term}(D,T).
0.7 \text{about}(D,T):- \text{link}(D,D1), \text{about}(D1,T).
Application: Probabilistic IR

• Uncertain Datalog rules: Semantics?

0.7 \text{term}(d1, \text{ir}).
0.8 \text{term}(d1, \text{db}).
0.5 \text{link}(d2,d1).
0.9 \text{temp1}.
0.7 \text{temp2}.

about(D,T) :- \text{term}(D,T), \text{temp1}.
about(D,T) :- \text{link}(D,D1), about(D1,T), \text{temp2}.
Probabilistic Datalog: QA

• Derivation of lineage formula with Boolean variables corresponding to used facts


• Probabilistic relational algebra


• Ranking / top-k QA

Probabilistic Relational Logics: Semantics

- **Distribution semantics** (aka grounding or Herbrand semantics) [Sato 95]
  Completely define discrete joint distribution by "factorization"
  Logical atoms treated as random variables
  - Probabilistic extensions to Datalog [Schmidt et al. 90, Dantsin 91, Ng & Subramanian 93, Poole et al. 93, Fuhr 95, Rölleke & Fuhr 97 and later]
  - Primula [Jaeger 95 and later]
  - BLP, ProbLog [De Raedt, Kersting et al. 07 and later]
  - Probabilistic Relational Models (PRMs) [Poole 03 and later]
  - Markov Logic Networks (MLNs) [Domingos et al. 06]
- **Probabilistic Soft Logic (PSL)** [Kimmig, Bach, Getoor et al. 12]
  Define density function using log-linear model
- **Maximum entropy semantics** [Kern-Isberner, Beierle, Finthammer, Thimm 10, 12]
  Partial specification of discrete joint with “uniform completion”

---

Inference Problems w/ and w/o Evidence

• **Static case**
  • Projection (margins),
  • Most-probable explanation (MPE)
  • Maximum a posteriori (MAP)
  • Query answering (QA): compute bindings

• **Dynamic case**
  • Filtering (current state)
  • Prediction (future states)
  • Hindsight (previous states)
  • MPE, MAP (temporal sequence)
% Intensional probabilistic facts:
0.6::heads(C) :- coin(C).

% Background information:
coin(c1).
coin(c2).
coin(c3).
coin(c4).

% Rules:
someHeads :- heads(_).

% Queries:
query(someHeads).
0.9744

https://dtai.cs.kuleuven.be/problog/
ProbLog

- **Compute marginal probabilities** of any number of ground atoms in the presence of evidence
- **Learn the parameters** of a ProbLog program from partial interpretations
- **Sample** from a ProbLog program
  - Generate random structures (use case: [Goodman & Tenenbaum 16])
- **Solve decision theoretic problems:**
  - Decision facts and utility statements


Markov Logic Networks (MLNs)

- Weighted formulas for modelling constraints [Richardson & Domingos 06]

- An MLN is a set of constraints \((w, \Gamma(x))\)
  - \(w = \text{weight}\)
  - \(\Gamma(x) = \text{FO formula}\)
- \textit{weight} of a world = product of \(\exp(w)\)
  - for all MLN rules \((w, \Gamma(x))\) and groundings \(\Gamma(a)\) that hold in that world
- \textbf{Probability} of a world = \(\frac{\text{weight}}{Z}\)
  - \(Z = \text{sum of weights of all worlds}\) (no longer a simple expression!)
Why exp?

- Log-linear models
- Let $D$ be a set of constants and $\omega \in \{0,1\}^m$ a world with $m$ atoms w.r.t. $D$

$$\text{weight}(\omega) = \prod_{\{(w,\Gamma(x))\in MLN \mid \exists a \in D^n : \omega \models \Gamma(a)\}} \exp(w)$$

$$\ln(\text{weight}(\omega)) = \sum_{\{(w,\Gamma(x))\in MLN \mid \exists a \in D^n : \omega \models \Gamma(a)\}} w$$

- Sum allows for component-wise optimization during weight learning

- $Z = \sum_{\omega \in \{0,1\}^m} \ln(\text{weight}(\omega))$

- $P(\omega) = \frac{\ln(\text{weight}(\omega))}{Z}$
Factor graphs

• Unifying representation for specifying **discrete distributions** with a factored representation
  • Potentials (weights) rather than probabilities

• Also used in engineering community for defining **densities w.r.t. continuous domains**
  [Loeliger et al. 07]

Scalability Issues

Scalability: Proposed solutions

• Limited expressivity
  • Probabilistic databases

• Knowledge Compilation
  • Linear programming
  • Weighted first-order model counting

• Approximation
  • Grounding + belief propagation (TensorLog)
Wrap-up Statistical Relational AI

• Probabilistic relational logics
  • Overview
  • Semantics
  • Inference problems

• Dealing with scalability issues (avoiding grounding)
  • Reduce expressivity (liftable queries)
  • Knowledge compilation (WFOMC)
  • Approximation (BP)

Next: Exact Lifted Inference
Mission and Schedule of the Tutorial*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>20 min</td>
</tr>
<tr>
<td>• StaR AI</td>
<td>✓</td>
</tr>
<tr>
<td>Overview: Probabilistic relational modeling</td>
<td>30 min</td>
</tr>
<tr>
<td>• Semantics (grounded-distributional, maximum entropy)</td>
<td>✓</td>
</tr>
<tr>
<td>• Inference problems and their applications</td>
<td></td>
</tr>
<tr>
<td>• Algorithms and systems</td>
<td></td>
</tr>
<tr>
<td>Scalable static inference</td>
<td>40 + 30 min</td>
</tr>
<tr>
<td>• Exact propositional inference</td>
<td></td>
</tr>
<tr>
<td>• Exact lifted inference</td>
<td></td>
</tr>
<tr>
<td>Scalable dynamic inference</td>
<td>50 min</td>
</tr>
<tr>
<td>• Exact propositional inference</td>
<td></td>
</tr>
<tr>
<td>• Exact lifted inference</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>10 min</td>
</tr>
</tbody>
</table>

*Thank you to the SRL/StaRAI crowd for all their exciting contributions! The tutorial is necessarily incomplete. Apologies to anyone whose work is not cited.