Dynamic StarAI
Dynamic Models and Statistical Relational AI

Tutorial at KI 2019

Tanya Braun, Marcel Gehrke, Ralf Möller
Universität zu Lübeck
Wrap Up

• Probabilistic relational models (PRMs) (Ralf)
• Answering static queries (Tanya)
• Answering continuous queries (Marcel)
• Take home messages (Ralf)
  • Putting PRMs in context
  • LJT and LDJT research relevant for all variants of PRMs
## Encoding a Joint Distribution

### Table:

<table>
<thead>
<tr>
<th>Travel(X)</th>
<th>Epid</th>
<th>Sick(X)</th>
<th>g_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>5</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>0</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>6</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>true</td>
<td>6</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>false</td>
<td>2</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>9</td>
</tr>
</tbody>
</table>

### Diagram:

- **Nat(D)** = natural disaster (D)
- **Man(W)** = man-made disaster (W)

The diagram includes nodes for **Nat(D)**, **Man(W)**, **Epid**, **Travel(X)**, and **Sick(X)**, connected by edges labeled **g_1**, **g_2**, and **g_3**.
Semantics of a PRM

- Joint probability distribution $P_G$ by grounding

\[ P_G = \frac{1}{Z} \prod_{f \in gr(G)} f \]

\[ Z = \sum_{\nu \in r(v, gr(G))} \prod_{f \in gr(G)} f_i(\pi_{rv(f_i)}(\nu)) \]

$\pi_{variables}(\nu) =$ projection of $\nu$ onto variables
Probabilistic Relational Models

- 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 98]
  - Primula [Jaeger 95]
  - ProbLog [De Raedt et al. 07]
• Define junction tree and relational snapshot interface such that grounding is avoided (e.g., with “node merging” strategies)

• Answer queries w.r.t. single nodes (fewer atoms to consider)

• Do we need to compute $Z$?

• For dynamic (sequential) reasoning, copy junction tree pattern, not the model
  • Tradeoff: memory requirements vs. performance
  • Generalization required for reasoning ("taming evidence")
PRMs and variants

• Probabilistic Relational Models (raw PRMs) [Poole 03, Tagipour et al. 13]

• Markov Logic Networks (MLNs) [Richardson and Domingos 06]
  • Use logical formulas to specify potential functions

• Probabilistic Soft Logic (PSL) [Bach, Broecheler, Huang Getoor et al. 17]
  • Use density functions to specify potential functions
Markov Logic Networks (MLNs)

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

- An MLN is a set of constraints \((w, \Gamma(x))\)
  - \(w = \text{weight}\)
  - \(\Gamma(x) = \text{FO formula}\)

- **Weight** of a world = product of \(\exp(w)\) terms
  - for all MLN rules \((w, \Gamma(x))\) and groundings \(\Gamma(a)\) that hold in that world

- 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. $D$

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

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

- Sum allows for component-wise optimization during weight learning

- $Z = \sum_{\omega \in \{0,1\}^m} \ln(weight(\omega))$
- $P(\omega) = \frac{\ln(weight(\omega))}{Z}$
MLN Reasoning

• Answering queries for $m$ discrete atoms

$$P(\text{Attends}(alice)) = \sum_{\omega \in \{\text{FALSE, TRUE}\}^m \land \omega \models \text{Attends}(alice)} P(\omega)$$

• Approximated lifted reasoning for MLNs
  Lifted Weighted Model Counting [Gogate and Domingos 11, 16]
  Lifted MAP [Sarkhel, Venugopal et al. 14]
Transforming MLNs into PRMs ...

Hard constraint

\[ \infty \text{ Presents}(X, P, C) \Rightarrow \text{Attends}(X, C) \]

Soft constraint, weight = \( \exp(3.75) \)

\[ 3.75 \text{ Publishes}(X, C) \land \text{FarAway}(C) \Rightarrow \text{Attends}(X, C) \]

\begin{tabular}{|c|c|c|}
\hline
\text{Presents}(X, P, C) & \text{Attends}(X, C) & g1 \\
\hline
\text{TRUE} & \text{TRUE} & 1000 \\
\text{TRUE} & \text{FALSE} & 1 \\
\text{FALSE} & \text{TRUE} & 1000 \\
\text{FALSE} & \text{FALSE} & 1000 \\
\hline
\end{tabular}

\[(X, P, C) \in \text{Dom}(X) \times \text{Dom}(P) \times \text{Dom}(C)\]

Possibly shattering required
Transforming MLNs into PRMs ...

\[ \Pr(P = \text{Presents}(X,P,C) \Rightarrow \text{Attends}(X,C)) \]

\[ \Pr(P = \text{Publishes}(X,C) \land \text{FarAway}(C) \Rightarrow \text{Attends}(X,C)) \]

\[ \Pr(P = \text{FarAway}(C)) \]

\[ \Pr(P = \text{Attends}(X,C)) \]

<table>
<thead>
<tr>
<th>Publishes(X,C)</th>
<th>FarAway(C)</th>
<th>Attends(X,C)</th>
<th>g2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>1</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>exp(3.75)</td>
</tr>
<tr>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>exp(3.75)</td>
</tr>
</tbody>
</table>
... and back?

- Represent/approximate potential functions using formulas with parfactor input atoms, e.g., for (approximate) explanation purposes

- Alternative: Represent potential functions using algebraic decision diagrams (ADDs) [Chavira and Darwiche 07]
  - Requires “multiplications” used for elimination to be defined over ADDs
Probabilistic Soft Logic (PSL)

- What if ranges of atoms are real-valued?
- Need to define potential function as density function

\[
\text{range}(\text{distance}(C)) = [0, 1] \\
0 = “close” \\
1 = “far away”
\]

[Bach, Broecheler, Huang Getoor et al. 17]
Probabilistic Soft Logic (PSL): Example

- First Order Logic weighted rules
  \[
  0.3 : \text{friend}(B, A) \land \text{votesFor}(A, P) \rightarrow \text{votesFor}(B, P) \\
  0.8 : \text{spouse}(B, A) \land \text{votesFor}(A, P) \rightarrow \text{votesFor}(B, P) \\
  \]

- Evidence
  \[
  I(\text{friend}(\text{John, Alex})) = 0.5 \\
  I(\text{spouse}(\text{John, Mary})) = 0.8 \\
  I(\text{votesFor}(\text{Alex, Romney})) = 0.6 \\
  I(\text{votesFor}(\text{Mary, Obama})) = 0.1 \\
  \]

- Inference
  \[
  - I(\text{votesFor}(\text{John, Obama})) = ? \\
  - I(\text{votesFor}(\text{John, Romney})) = ?
  \]
PSL’s Interpretation of Logical Connectives

- Continuous truth assignment
  Łukasiewicz relaxation of AND, OR, NOT
  - \( I(\ell_1 \land \ell_2) = \max \{0, I(\ell_1) + I(\ell_2) - 1\} \)
  - \( I(\ell_1 \lor \ell_2) = \min \{1, I(\ell_1) + I(\ell_2)\} \)
  - \( I(\neg \ell_1) = 1 - I(\ell_1) \)

- Distance to satisfaction \( d \)
  - Implication: \( \ell_1 \rightarrow \ell_2 \) is satisfied iff \( I(\ell_1) \leq I(\ell_2) \)
  - \( d = \max \{0, I(\ell_1) - I(\ell_2)\} \)

- Example
  - \( I(\ell_1) = 0.3, I(\ell_2) = 0.9 \Rightarrow d = 0 \)
  - \( I(\ell_1) = 0.9, I(\ell_2) = 0.3 \Rightarrow d = 0.6 \)
Digression: Maximum Entropy Principle

• Given:
  • States $s = s_1, s_2, \ldots, s_n$
  • Density $p_s = (p_1, p_2, \ldots, p_n)$

• Maximum Entropy Principle:
  • W/o further information, select $p_s$, s.t. entropy $H$ is maximized

$$H = - \sum_s p_s(s) \log p_s(s) = -p_s \log p_s$$

• w.r.t. $k$ constraints (expected values of features $f_i$)

$$\sum_s p_s(s) f_i(s) = D_i \quad i \in \{1..k\}$$

Minimize component-wise weighted difference:

$$\sum_{i} \lambda_i \left( \sum_s p_s(s) f_i(s) - D_i \right)$$
Maximum Entropy Principle

• Maximize Lagrange functional for determining \( p_s \)

\[
L = -p_s \log p_s - \sum_i \lambda_i \left( \sum_s p_s(s) f_i(s) - D_i \right) - \mu \left( \sum_s p_s(s) - 1 \right)
\]

• Partial derivatives of \( L \) w.r.t. \( p_s \) \( \Rightarrow \) roots:

\[
p_s(s) = \frac{\exp \left( - \sum_i \lambda_i f_i(s) \right)}{Z}
\]

where \( Z \) is for normalization
(Boltzmann-Gibbs distribution)

• "Global" modeling: additions/changes to constraints/rules
  influence the whole joint probability distribution

Probability values should sum up to 1

Features

Weights
PSL Probability Distribution

Density function $f(I)$ for assignment $I$ in the style of Boltzmann-Gibbs distributions:

$$f(I) = \frac{1}{Z} \exp[-\sum_{r \in R} \lambda_r (d_r(I))^-]$$

- a possible continuous truth assignment
- Normalization constant
- For all rules
- Weight of formula $r$
- Distance to satisfaction of rule $r$

Suggestion:

- Compile model into PRM with junction tree representation
- Variable elimination on subsets of model
  - Algebraic convolution of density functions $f$
  - Compilation into convex optimization problem
PRMs: Take Home Messages

• Knowledge representation with PRMs
  • Real-valued range values (PSL)
  • Discrete range values, adopt propositional logic (MLNs)
  • Discrete range values, drop propositional logic (raw PRMs)

• Our PRM research might be relevant for all formalisms
  • (Dynamic) junction tree provides for smaller submodels that need to be handled (also holds for knowledge compilation)
  • Organization of junction tree w.r.t. grounding avoidance
  • Taming evidence in the dynamic case

• PRMs are an active area of research
  • Generalized sequential structures
  • Compilation approaches
  • Varying domains (towards first-order structures)
  • Transfer, Explanation
Bibliography

• [Chavira and Darwiche 07]

• [De Raedt et al. 07]

• [Domingos et al. 06]

• [Gogate and Domingos 11]

• [Jaeger 95]

• [Bach, Broecheler, Huang Getoor et al. 17]

• [Poole 03]
  David Poole. First-order probabilistic inference. IJCAI 2003: 985-991

• [Tagipour et al. 13]
Bibliography

- [Schmidt et al. 90]

- [Dantsin 91]

- [Ng & Subramanian 93]

- [Poole et al. 93]

- [Fuhr 95]

- [Rölleke & Fuhr 98]

- [Sarkhel, Venugopal et al. 14]
  Somdeb Sarkhel, Deepak Venugopal, Parag Singla, Vibhav Gogate.: Lifted MAP Inference for Markov Logic Networks. AISTATS 2014: 859-867

- [Sato 95]

- [van den Broeck 13]
The End

*PRMs are a true backbone of AI, and this tutorial emphasized only some central topics. We definitely have not cited all publications relevant to the whole field of PRMs here. We would like to thank all our colleagues for making their slides available (see some of the references to papers for respective credits). Slides or parts of it are almost always modified.