StarAI

Stable Inference over Time in Dynamic PRMs

Tutorial ECAI 2020

Tanya Braun, Marcel Gehrke, Ralf Möller
Universität zu Lübeck
Agenda

• Probabilistic relational models (PRMs) [Ralf]
• Exact symmetries and changing domains in static PRMs [Tanya]
• Stable inference over time in dynamic PRMs [Marcel]
  • Reasoning over time
  • Keeping reasoning polynomial
• Summary [Tanya]
Reasoning over Time

Keep the past independent from the future
Lifted: Dynamic Model

- Marginal distribution query: $P(A^i_\pi | E_{0:t})$ w.r.t. the model:
  - Hindsight: $\pi < t$ (was there an epidemic $t - \pi$ days ago?)
  - Filtering: $\pi = t$ (is there an currently an epidemic?)
  - Prediction: $\pi > t$ (is there an epidemic in $\pi - t$ days?),

Gehrke et al. (2018)
Reasoning over Time: Naïve

- Given temporal pattern
  - Instantiate and unroll pattern for $T$ timesteps
  - Infer on unrolled model
    - Works for all types of queries
- Problems:
  - Huge model (unrolled for $T$ timesteps)
  - Redundant temporal information and calculations
Reasoning over Time: Interfaces

• Main idea:
  Use temporal conditional independences to perform inference on smaller model
  • Normally only a subset of random variables influence next time step → interface variables
  • State description of interface variables from time slice $t - 1$ suffice to perform inference on time slice $t$

→ Makes past independent from the present (and the future)
Reasoning over Time: Interfaces

• Build a helper structure of clusters (junction tree)
  • Cluster = set of randvars occurring together during calculations
    • Each cluster collects all information currently present in a model encoded in the randvars contained in the cluster
  • Ensure interface variables part of one cluster
    • Cluster acts basically as a gateway to the future
    • Query over interface variables collects state description of interface variables

• Proceed forward one time step at a time, using the same structure

• Algorithms:
  • Propositional: Interface Algorithm (Murphy, 2002)
  • Lifted: Lifted Dynamic Junction Tree Algorithm (G et al, 2018)
Lifted Dynamic Junction Tree Algorithm: LDJT

• Input
  • Temporal model $G$
  • Evidence $E$
  • Queries $Q$

• Algorithm
  1. Identify interface variables
  2. Build FO jtree structures $J$ for $G$
  3. Instantiate $J_t$
  4. Restore state description of interface variables from $m_{t-1}$
  5. Enter evidence $E_t$ into $J_t$
  6. Pass messages in $J_t$
  7. Answer queries $Q_t$
  8. Store state description of interface variables in $m_t$
  9. Proceed to next time step (step 3)

G et al. (2018)
LDJT: Identify Interface Variables

- Use temporal conditional independences to perform inference on smaller model (Murphy (2002))

\[ \mathbf{I}_{t-1} = \{ A^i_{t-1} \mid \exists \phi(\mathcal{A})_c \in G : A^i_{t-1} \in \mathcal{A} \land A^j_{t-1} \in \mathcal{A} \} \]

- Set of interface variable $\mathbf{I}_{t-1}$ consists of all PRVs from time slice $t - 1$ that occur in a parfactor with PRVs from time slice $t$
LDJT: Construct FO jtree Structure

- Turn model in 1.5 time slice model
- Suffices to perform inference over time slice $t$
- From 1.5 time slice model construct FO jtree structure
- Ensure $I_{t-1}$ is contained in a parcluster and $I_t$ is contained in a parcluster
- Label parcluster with $I_{t-1}$ as in-cluster and parcluster with $I_t$ as out-cluster
LDJT: Query answering

- Instantiate FO jtree structure
- Restore state description of interface variables
- Enter evidence
- Pass messages
- Query answering:
  - Find parcluster containing query term
  - Extract submodel
  - Answer query with LVE
LDJT: Proceed in time

- Calculate forward message $m_3$ using out-cluster ($C_3^2$)
- Eliminate $Travel(X)_3$ from $C_3^2$’s local model
- Instantiate next FO jtree and enter $m_3$
- Enter evidence and pass messages

G et al. (2018)
Reasoning over Time: Interfaces

• Forward pass for filtering and prediction queries
  • Keep current instantiation of FO jtree in memory

• Backward pass for hindsight queries (G et al., 2019)
  • Different instantiation approaches
    • Trade-off between memory and runtime

• Other query types possible
  • e.g., MPE (G et al., 2019a)

• All have one problem:
  they see evidence over time
Keeping Reasoning Polynomial

Why evidence screws everything up and how approximating symmetries might save us
Taming Reasoning

- Evidence can ground a model over time
- Non-symmetric evidence
  - Observe evidence for some instances in one time step
  - Observe evidence for a subset of these instances in another time step
  - Split the logical variable slowly over time
- Vanilla junction trees for each time step
- Forward message carries over splits, leading to slowly grounding a model over time

G et al. (2020)
Evidence over Time

- $D_3(x_1) = \text{true}$
- Split $g_3^2$ into
  - $g_3^2'$ for $x_1$ and
  - $g_3^2''$ for $X \neq x_1$

- $m_3$ consists of
  - $m^{12}$
  - $m^{32}$
  - $g_3^2'$ and $g_3^2''$ with $D_3(X)$ eliminated

\[ R_2(X) A_2(X) \quad R_3(X) \quad R_4(X) \]
\[ R_3(X) A_3(X) \quad g_3^2 \quad m^{12} \quad m^{32} \quad g_3^2 \quad m^{23} \]
\[ g^E \quad m_2 \quad m^{21} \quad g^E \quad m_3 \quad m^{21} \]
\[ g_3^2 \quad m_2 \quad m^{21} \quad g_4^2 \quad m_{12} \quad m^{32} \]
\[ g_3^2 \quad m^{23} \quad g_3^2 \quad m^{23} \]
\[ C_3 \quad C_3 \quad C_4 \quad C_4 \]

G et al. (2020)
Evidence over Time

- $D_4(x_2) = \text{true}$
- Split $g_4^2$ into
  - $g_4^2$' for $x_2$ and
  - $g_4^2$'' for $X \neq x_2$
- $m_4$ consists of
  - $m^{12}$ (containing $m_3$)
  - $m^{32}$
  - $g_4^2$' and $g_4^2$'' with $D_4(X)$ eliminated

\[ \begin{array}{c}
\text{In-cluster } C_3^1 \\
R_2(X) A_2(X) \\
R_3(X) \\
g^E m_2 m^{21}
\end{array} \quad \begin{array}{c}
\text{Out-cluster } C_3^2 \\
R_3(X) A_3(X) \\
D_3(X) \\
g^E m_2 m^{21}
\end{array} \quad \begin{array}{c}
\text{In-cluster } C_4^1 \\
R_3(X) A_3(X) \\
D_3(X) \\
g^E m_3 m^{21}
\end{array} \quad \begin{array}{c}
\text{Out-cluster } C_4^2 \\
R_4(X) A_4(X) \\
D_4(X) \\
g^E m_3 m^{21}
\end{array} \]

\[ \begin{array}{c}
\text{In-cluster } C_3^1 \\
R_2(X) A_2(X) \\
R_3(X) \\
g^E m_2 m^{21}
\end{array} \quad \begin{array}{c}
\text{Out-cluster } C_3^2 \\
R_3(X) A_3(X) \\
D_3(X) \\
g^E m_2 m^{21}
\end{array} \quad \begin{array}{c}
\text{In-cluster } C_4^1 \\
R_3(X) A_3(X) \\
D_3(X) \\
g^E m_3 m^{21}
\end{array} \quad \begin{array}{c}
\text{Out-cluster } C_4^2 \\
R_4(X) A_4(X) \\
D_4(X) \\
g^E m_3 m^{21}
\end{array} \]

G et al. (2020)
Undoing Splits

• Need to undo splits to keep reasoning polynomial w.r.t. domain sizes

• Where can splits be undone efficiently?

• How to undo splits?

• Is it reasonable to undo splits?
  • Effect of slight differences in evidence?
  • Impact of evidence vs. temporal behaviour of model?
Approximating Symmetries in Static Models

• Approximate symmetries while entering evidence (Singla et al. 2014, Venugopal and Gogate 2014)
  • Model does not blow up
  • Approximate inference results

• Other results for approximating symmetries exists (Van den Broeck and Darwiche 2013, Van den Broeck and Niepert 2015, Mladenov et al. 2017)

• We want to be as exact as possible
  • Use benefits of temporal model for symmetries
Where Can Splits Be Undone Efficiently?

- Evidence causes splits in a logical variable in the same way in all factors in a model.
- LDJT always instantiates a vanilla junction tree.
- Forward message carries over splits.

G et al. (2020)
How to Undo Splits?

• The colour passing algorithm can efficiently identify exact symmetries
  • Presented in previous section (Ahmadi et al. 2013)

• Evidence causes differences in distributions

• Need to find approximate symmetries to undo splits caused by evidence

• Need a way to merge factors
Comparing Parfactors

• Comparing all marginals is expensive
• Comparing the joint distribution over the complete interface is expensive
Comparing Parfactors

- Comparing marginals of a subset of PRVs can determine non-similar factors similar

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>7</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>1</td>
</tr>
</tbody>
</table>

- $P(A(x_1 = true)):$ \( \frac{2}{3} \)
- $P(R(x_1 = true)):$ \( \frac{5}{12} \)

G et al. (2020)
Comparing Parfactors

- Potentials determine distributions
- Similar ratios in potentials lead to similar marginals and similar factors

\[
\begin{array}{|c|c|c|}
\hline
R(X) & A(X) & g \\
\hline
false & false & 4 \\
false & true & 3 \\
true & false & 2 \\
true & true & 1 \\
\hline
\end{array}
\]

- \( P(A(x_1 = true)) \): \( \frac{4}{10} \)
- \( P(R(x_1 = true)) \): \( \frac{3}{10} \)
- \( P(A(x_1 = true) \land R(x_1 = true)) \): \( \frac{1}{10} \)
Identifying Similar Groups

- Groups are equal if they have the same full joint distribution

- Full joint distribution computationally hard to get
  - Use parfactors as vector
  - If vectors of two groups point in same direction, they have the same full joint distribution

G et al. (2020)
Find Approximate Symmetries

• Cosine similarity for similarity of vectors

\[ \cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \]

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>7</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>2</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>4</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>2</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>4</td>
</tr>
</tbody>
</table>

\[ \cos(\theta) = \frac{0.2 + 7.4 + 4.2 + 1.4}{\sqrt{0 + 49 + 16 + 1 \cdot \sqrt{4 + 16 + 4 + 16}}} \approx 0.7785 \]
Find Approximate Symmetries

- Cosine similarity for similarity of vectors

\[
\cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}
\]

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>3</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>2</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>1</td>
</tr>
</tbody>
</table>

- \(\cos(\theta) = \frac{4 \cdot 3.9 + 3 \cdot 3.1 + 2 \cdot 2.1 + 1 \cdot 0.9}{\sqrt{16 + 9 + 4 + 1} \cdot \sqrt{15.21 + 9.61 + 4.41 + 0.81}} \approx 0.9993\)

G et al. (2020)
Find Approximate Symmetries

- Cosine similarity for similarity of vectors

\[ \cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \]

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>3</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>2</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R(X)</th>
<th>A(X)</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>8</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>6</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>4</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>2</td>
</tr>
</tbody>
</table>

- \[ \cos(\theta) = \frac{4 \cdot 8 + 3 \cdot 6 + 2 \cdot 4 + 1 \cdot 3}{\sqrt{16+9+4+1} \cdot \sqrt{64+36+16+4}} = 1 \]

- Cluster splits with \( 1 - \cos(\theta) \) as distance function
Cluster Groups

- Density-based clustering as unknown number of clusters
- Cosine similarity as distance function

G et al. (2020)
Cluster Groups

- Density-based clustering as unknown number of clusters
- Cosine similarity as distance function

G et al. (2020)
Merge Clusters

- Merge groups of cluster by calculating mean of cluster while accounting for groundings
- Replace old groups with merged group in temporal message
Merging Parfactors

- Merge similar parfactors based on distance function while accounting for groundings

| \( |\mathcal{D}(X)| = 4 \) | \( |\mathcal{D}(X')| = 4 \) | \( |\mathcal{D}(X'')| = 2 \) |
|---|---|---|
| \( \begin{array}{ccc} R(X) & A(X) & g \\ false & false & 4 \\ false & true & 3 \\ true & false & 2 \\ true & true & 1 \end{array} \) | \( \begin{array}{ccc} R(X') & A(X') & g \\ false & false & 7.9 \\ false & true & 6 \\ true & false & 3.9 \\ true & true & 2.1 \end{array} \) | \( \begin{array}{ccc} R(X'') & A(X'') & g \\ false & false & 15.7 \\ false & true & 12.2 \\ true & false & 8.1 \\ true & true & 3.8 \end{array} \) |

| \( |\mathcal{D}(X)| = 10 \) |
|---|
| \( \begin{array}{ccc} R(X) & A(X) & g \\ false & false & \frac{(4\cdot4+7.9\cdot4+15.7\cdot2)}{10} = 7.9 \\ false & true & \frac{(3\cdot4+6.4+12.2\cdot2)}{10} = 6.04 \\ true & false & \frac{(2\cdot4+3.9\cdot4+8.1\cdot2)}{10} = 3.98 \\ true & true & \frac{(1\cdot4+2.1\cdot4+3.8\cdot2)}{10} = 2 \end{array} \) |
Is It Reasonable to Undo Splits?

- Approximate forward message
- For each time step the temporal behaviour is multiplied on the forward message
- Indefinitely bounded error due to temporal behaviour
Taming Reasoning

• Need to undo splits to keep reasoning polynomial w.r.t. domain sizes

• Where can splits be undone efficiently?
  • Undo splits in a forward message

• How to undo splits?
  • Find approximate symmetries
  • Merge based on groundings

• Is it reasonable to undo splits
  • Yes, due to the temporal model behaviour (indefinitely bounded error)

G et al. (2020)
Results

- DBSCAN for Clustering
- ANOVA for checking fitness of clusters

\[
\begin{array}{|c|c|c|c|}
\hline
\pi & \text{Max} & \text{Min} & \text{Average} \\
\hline
0 & 0.0001537746121 & 0.00000000001720 & 0.0000191206488 \\
2 & 0.00000000851654 & 0.0000000000001 & 0.0000000111949 \\
4 & 0.0000000000478 & 0 & 0.0000000000068 \\
\hline
\end{array}
\]
Wrap-up Stable Inference over Time

• Reasoning over time
  • Unrolling of model infeasible
  • Using interface variables to separate past from future

• Keeping reasoning polynomial
  • Evidence yielding a splintered model
  • Taming effects of evidence
    • Using approximate symmetries to identify groups of parfactors
    • Merging a group into a single parfactor
    • Error indefinitely bounded
Bibliography
Alphabetically sorted
Bibliography

• Ahmadi et al. (2013)

• G et al. (2018)

• G et al. (2019)

• G et al. (2019a)
Bibliography

• G et al. (2020)

• Mladenov et al. (2017)

• Murphy (2002)

• Venugopal and Gogate (2014)
• Van den Broeck and Darwiche (2013)

• Van den Broeck and Niepert (2015)