Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 17

Oct, 19, 2015

Slide Sources
D. Koller, Stanford CS - Probabilistic Graphical Models
D. Page, Whitehead Institute, MIT

Several Figures from
“Probabilistic Graphical Models: Principles and Techniques” D. Koller, N. Friedman 2009
Simple but Powerful Approach: Particle Filtering

Idea from **Exact Filtering**: should be able to compute \( P(X_{t+1} \mid e_{1:t+1}) \) from \( P(X_t \mid e_{1:t}) \)

“.. One slice from the previous slice…”

Idea from **Likelihood Weighting**

- Samples should be weighted by the probability of evidence given parents

**New Idea**: run multiple samples simultaneously through the network
Particle Filtering

- Run all \textbf{N samples together} through the network, one slice at a time

**STEP 0:** Generate a population on \textbf{N} initial-state samples by sampling from initial state distribution \( P(X_0) \)

\[ \begin{array}{c}
\text{Rain}_0 \\
\text{true} \\
\text{false}
\end{array} \]

\[ N = 10 \]
STEP 1: Propagate each sample for $x_t$ forward by sampling the next state value $x_{t+1}$ based on $P(X_{t+1} | X_t)$.
**STEP 2**: Weight each sample by the likelihood it assigns to the evidence

- E.g. assume we observe *not umbrella* at t+1
Particle Filtering

STEP 3: Create a new population from the population at $\mathbf{X}_{t+1}$, i.e. resample the population so that the probability that each sample is selected is proportional to its weight.

- Start the Particle Filtering cycle again from the new sample.
Is PF Efficient?

In practice, approximation error of particle filtering remains bounded overtime.

It is also possible to prove that the approximation maintains bounded error with high probability (with specific assumption: probs in transition and sensor models $>0$ and $<1$).
422 big picture: Where are we?

Deterministic

- Logics
  - First Order Logics
- Ontologies
  - Temporal rep.
  - Full Resolution
  - SAT

Stochastic

- Belief Nets
- Approx. : Gibbs
- Approx. : Particle Filtering
- Markov Chains and HMMs
  - Forward, Viterbi….
- Undirected Graphical Models
  - Markov Networks
  - Conditional Random Fields
- Markov Decision Processes and Partially Observable MDP
  - Value Iteration
  - Approx. Inference
- Reinforcement Learning

Applications of AI

Hybrid: Det + Sto

Prob CFG
Prob Relational Models
Markov Logics

Representation Reasoning Technique
Lecture Overview

Probabilistic Graphical models

- Intro
- Example
- Markov Networks Representation (vs. Belief Networks)
- Inference in Markov Networks (Exact and Approx.)
- Applications of Markov Networks
Probabilistic Graphical Models

From “Probabilistic Graphical Models: Principles and Techniques” D. Koller, N. Friedman 2009
Misconception Example

• Four students (Alice, Bill, Debbie, Charles) get together in pairs, to work on a homework
• But only in the following pairs: AB, AD, DC, BC
• Professor misspoke and might have generated misconception
• A student might have figured it out later and told study partner

Four random vars

eg A random var two values

a1 Alice has the misc.

a0 Alice doesn’t have the misc.
Example: In/Dependencies

Are A and C independent because they never spoke?
No, because A might have figure it out and told B who then told C
But if we know the values of B and D,….

And if we know the values of A and C
Which of these two Bnets captures the two independencies of our example?

- (A ⊥ C | B D)
- (B ⊥ D | A C)

In a. B \not\perp D | C
In b. same

- a. B
- b. D
- c. Both
- d. None
Factors define the local interactions (like CPTs in Bnets)
What about the global model? What do you do with Bnets?
How do we combine local models?

As in BNets by multiplying them!

\[
\tilde{P}(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(A, D)
\]

\[
P(A, B, C, D) = \frac{1}{Z} \tilde{P}(A, B, C, D)
\]

\[
P(A, B) = 
\]

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Unnormalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a^0) (b^0) (c^0) (d^0)</td>
<td>300000</td>
</tr>
<tr>
<td>(a^0) (b^0) (c^0) (d^1)</td>
<td>300000</td>
</tr>
<tr>
<td>(a^0) (b^0) (c^1) (d^0)</td>
<td>300000</td>
</tr>
<tr>
<td>(a^0) (b^0) (c^1) (d^1)</td>
<td>30</td>
</tr>
<tr>
<td>(a^0) (b^0) (c^1) (d^1)</td>
<td>500</td>
</tr>
<tr>
<td>(a^0) (b^1) (c^0) (d^0)</td>
<td>500</td>
</tr>
<tr>
<td>(a^0) (b^1) (c^0) (d^1)</td>
<td>5000000</td>
</tr>
<tr>
<td>(a^0) (b^1) (c^1) (d^1)</td>
<td>500</td>
</tr>
<tr>
<td>(a^1) (b^0) (c^0) (d^0)</td>
<td>100</td>
</tr>
<tr>
<td>(a^1) (b^0) (c^0) (d^1)</td>
<td>100</td>
</tr>
<tr>
<td>(a^1) (b^0) (c^1) (d^1)</td>
<td>10</td>
</tr>
<tr>
<td>(a^1) (b^1) (c^0) (d^0)</td>
<td>100000</td>
</tr>
<tr>
<td>(a^1) (b^1) (c^0) (d^1)</td>
<td>100000</td>
</tr>
<tr>
<td>(a^1) (b^1) (c^1) (d^1)</td>
<td>100000</td>
</tr>
</tbody>
</table>

Normalized

<table>
<thead>
<tr>
<th>(\phi_1[A, B])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d^0) (a^0) 100</td>
</tr>
<tr>
<td>(d^0) (a^1) 1</td>
</tr>
<tr>
<td>(d^1) (a^0) 1</td>
</tr>
<tr>
<td>(d^1) (a^1) 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\phi_2[B, C])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c^0) (d^0) 1</td>
</tr>
<tr>
<td>(c^0) (d^1) 100</td>
</tr>
<tr>
<td>(c^1) (d^0) 100</td>
</tr>
<tr>
<td>(c^1) (d^1) 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\phi_3[C, D])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d^0) (c^0) 1</td>
</tr>
<tr>
<td>(d^0) (c^1) 1</td>
</tr>
<tr>
<td>(d^1) (c^0) 1</td>
</tr>
<tr>
<td>(d^1) (c^1) 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\phi_4[D, A])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a^0) (b^0) 0.04</td>
</tr>
<tr>
<td>(a^0) (b^1) 0.04</td>
</tr>
<tr>
<td>(a^1) (b^0) 4.1 \times 10^{-6}</td>
</tr>
<tr>
<td>(a^1) (b^1) 0.69</td>
</tr>
</tbody>
</table>

CPSC 422, Lecture 17
Multiplying Factors (same seen in 322 for VarElim)
Factors do not represent marginal probs. !

Marginal $P(A,B)$
Computed from the joint

<table>
<thead>
<tr>
<th></th>
<th>$a^0 b^0$</th>
<th>$a^0 b^1$</th>
<th>$a^1 b^0$</th>
<th>$a^1 b^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^0 b^0$</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a^0 b^1$</td>
<td></td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a^1 b^0$</td>
<td></td>
<td></td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>$a^1 b^1$</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
</tbody>
</table>
Step Back.... From structure to factors/potentials

In a Bnet the joint is factorized....

In a Markov Network you have one factor for each maximal clique
Directed vs. Undirected

Indepedendencies

\[
\begin{align*}
(F \perp H \mid S) \\
(C \perp S \mid F, H) \\
(M \perp C, H, S \mid F)
\end{align*}
\]

Factorization

\[
\begin{align*}
P(s, f, h, m, c) &= P(s) \times P(f \mid s) \times P(h \mid s) \times P(m \mid f) \times P(c \mid f, h) \\
P(A, B, C, D) &= \frac{1}{Z} \Phi_1(A, B) \times \Phi_2(B, C) \times \Phi_3(C, D) \times \Phi_4(A, D)
\end{align*}
\]
General definitions

Two nodes in a Markov network are independent if and only if every path between them is cut off by evidence.

So the markov blanket of a node is…?

eg for A C

eg for C
Markov Networks Applications (1): Computer Vision

Called Markov Random Fields

- Stereo Reconstruction
- Image Segmentation
- Object recognition

Typically pairwise MRF

- Each \textit{vars} correspond to a \textit{pixel} (or \textit{superpixel})
- Edges (factors) correspond to interactions between adjacent pixels in the image
  - E.g., in segmentation: from generically penalize discontinuities, to road under car
Image segmentation

classifying each superpixel independently with a Markov Random Field!
Markov Networks Applications (2): Sequence Labeling in NLP and BioInformatics

Conditional random fields (next class Wed)

**KEY**

- **B-PER**: Begin person name
- **I-PER**: Within person name
- **B-LOC**: Begin location name
- **I-LOC**: Within location name
- **OTH**: Not an entity

**Diagram**

Mrs. Green spoke today in New York Green chairs the finance committee

Recognize names of persons, locations etc. Named Entities

CPSC 422, Lecture 17
Learning Goals for today’s class

You can:

- Justify the need for undirected graphical model (Markov Networks)
- Interpret local models (factors/potentials) and combine them to express the joint
- Define independencies and Markov blanket for Markov Networks
- Perform Exact and Approx. Inference in Markov Networks
- Describe a few applications of Markov Networks
How to prepare….

- Keep Working on assignment-2!
- Go to Office Hours
- Learning Goals (look at the end of the slides for each lecture – will post complete list)
- Revise all the clicker questions and practice exercises
- Will post more practice material today
How to acquire factors?