Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 17

Feb, 24, 2021

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
422 big picture: Where are we?

StarAI (statistical relational AI)
Hybrid: Det + Sto
Prob CFG
Prob Relational Models
Markov Logics

Deterministic

Logics
First Order Logics

Ontologies
Temporal rep.

Query

Stochastic

Belief Nets
Approx. : Gibbs
Markov Chains and HMMs
Forward, Viterbi….
Approx. : Particle Filtering

Undirected Graphical Models
Markov Networks
Conditional Random Fields

Planning

Markov Decision Processes and
Partially Observable MDP

Applications of AI

• Value Iteration
• Approx. Inference
Reinforcement Learning

Representations
Reasoning
Techniques
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

- **Temporal Models**
  - 6.2, 15.1-2, 15.3.1, 15.3.3

- **Relational Models**
  - 6.3-4, 17.5, (18.6.2)

- **Continuous Models**
  - 5.5, 7, 14.1-2, 14.3.1-2, 14.5.1-3

- **Undirected Models**
  - 4.3-7

- **Bayesian Networks**
  - 3.3-4, 5.1-4

- **Learning Undirected Models**
  - 20.1-2, 20.3.1-2

- **Exact Inference**
  - 9.1-4, 10.1-2

- **Approx. Inference**
  - 11.3.1-5, 12.1, 12.3.1-3

- **Decision Making**
  - 22.1-2, 23.1-2, 23.4-5

- **MAP Inference**
  - 13.1-4

- **Advanced Approx. Inference**
  - 8, 10.3, 11, 12.3-4

- **Structure Learning**
  - 17.3-4, 18.1, 18.3-4, 18.6

- **Advanced Learning**
  - 18.5, 19, 20

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

CPSC 422, Lecture 17
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
a' Alice has the misc.
\(\bar{a} \) Alice doesn't have the misc.

CPSC 422, Lecture 17
Example: In/Dependencies

Are A and C independent because they never spoke?

a. Yes  b. No  c. Cannot Tell

No, because A might have figured 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 ⊥ D | C
In b. same

a. b.

- Both
- None

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!

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

\[
P(A, B, C, D) = \frac{1}{2} \tilde{P}(A, B, C, D)
\]
Multiplying Factors (same seen in 322 for VarElim)

*unrelated to our running example*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^1$</td>
<td>$b^1$</td>
</tr>
<tr>
<td>$a^1$</td>
<td>$b^2$</td>
</tr>
<tr>
<td>$a^2$</td>
<td>$b^1$</td>
</tr>
<tr>
<td>$a^2$</td>
<td>$b^2$</td>
</tr>
<tr>
<td>$a^3$</td>
<td>$b^1$</td>
</tr>
<tr>
<td>$a^3$</td>
<td>$b^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b^1$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$b^1$</td>
<td>$c^2$</td>
</tr>
<tr>
<td>$b^2$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$b^2$</td>
<td>$c^2$</td>
</tr>
</tbody>
</table>

| $a^1$ | $b^1$ | $c^1$ | $0.5 \cdot 0.5 = 0.25$ |
| $a^1$ | $b^1$ | $c^2$ | $0.5 \cdot 0.7 = 0.35$ |
| $a^1$ | $b^2$ | $c^1$ | $0.8 \cdot 0.1 = 0.08$ |
| $a^1$ | $b^2$ | $c^2$ | $0.8 \cdot 0.2 = 0.16$ |
| $a^2$ | $b^1$ | $c^1$ | $0.1 \cdot 0.5 = 0.05$ |
| $a^2$ | $b^1$ | $c^2$ | $0.1 \cdot 0.7 = 0.07$ |
| $a^2$ | $b^2$ | $c^1$ | $0.1 \cdot 0.1 = 0$ |
| $a^2$ | $b^2$ | $c^2$ | $0.0 \cdot 0.2 = 0$ |
| $a^3$ | $b^1$ | $c^1$ | $0.3 \cdot 0.5 = 0.15$ |
| $a^3$ | $b^1$ | $c^2$ | $0.3 \cdot 0.7 = 0.21$ |
| $a^3$ | $b^2$ | $c^1$ | $0.9 \cdot 0.1 = 0.09$ |
| $a^3$ | $b^2$ | $c^2$ | $0.9 \cdot 0.2 = 0.18$ |

In this example, A has three values: $a^1$, $a^2$, $a^3$. 

CPSC 422, Lecture 17
Factors do not represent marginal probs.!

\begin{center}
\begin{tabular}{|c|c|}
\hline
\(a^0\ b^0\) & 0.13 \\
\hline
\(a^0\ b^1\) & 0.69 \\
\hline
\(a^1\ b^0\) & 0.14 \\
\hline
\(a^1\ b^1\) & 0.04 \\
\hline
\end{tabular}
\end{center}

Marginal \(P(A,B)\)

Computed from the joint
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

\[
\Phi_1(A, B, D) \\
\Phi_2(B, D, C) \\
\Phi_3(C, E, F) \\
\Phi_4(E, G)
\]
Directed vs. Undirected

Independencies

\[(F \perp H \mid S)\]
\[(C \perp S \mid F, H)\]
\[(M \perp C, H, S \mid F)\]

Factorization

\[P(S, F, H, M, C) = P(S)P(F \mid S)P(H \mid S)P(M \mid F)\]
\[P(C \mid F, H)\]

\[P(A, B, C, D) = \frac{1}{\sum} \Phi_1(AB)\]
\[\Phi_2(BC)\]
\[\Phi_3(CD)\]
\[\Phi_4(AD)\]
General definitions

Two nodes in a Markov network are independent if and only if all paths between them are blocked by evidence.

So the markov blanket of a node is...?

eg for A C

eg for C

a. All the parents of its children
b. The whole network

C. All its neighbors
Markov Networks Applications (1): Computer Vision

Called Markov Random Fields

- Stereo Reconstruction
- Image Segmentation
- Object recognition

Typically pairwise MRF

- Each vars correspond to a pixel (or 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 (1): Computer Vision

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

**Simple Example**

- $A_{ij}$
- $A_{i-1, j}$
- $A_{i, j-1}$
- $A_{i, j}$

Factors:

- $A_{ij}$
- $A_{i, j-1}$
- $A_{i-1, j}$

Values:

- road: 100
- car: 50
- 1
- 100
Markov Networks Applications (1): Computer Vision

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

Simple example:

Factors only for the labels road and car

Factors between any two nodes (one above the other)
Conditional random fields (next class Fri)

5 possible states (similar to HMM)

recognize names of persons
locations etc

named entities
Combining CRFs and Neural Models

SEMANTIC IMAGE SEGMENTATION WITH DEEP CONVOLUTIONAL NETS AND FULLY CONNECTED CRFS

International Conference on Learning Representations (ICLR), San Diego, California, USA, May 2015.

Liang-Chieh Chen Univ. of California, Los Angeles; George Papandreou Google Inc.; Iasonas Kokkinos INRIA; Kevin Murphy Google Inc.; Alan L. Yuille Univ. of California, Los Angeles

1. Use CNN to generate a rough prediction of segmentation (smooth, blurry heat map)
2. Refine this prediction with a conditional random field (CRF)
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
Less than Two weeks to Midterm, Mon, March 8

How to prepare....

• Keep Working on assignment-2!
• Go to Office Hours
• Learning Goals (look at the end of the slides for each lecture – complete list will be posted)
• Revise all the clicker questions and practice exercises
• More practice material will be posted next week
• Check questions and answers on Piazza
How to acquire factors?