Reasoning Under Uncertainty: Belief Networks

Computer Science cpsc322, Lecture 27

(Textbook Chpt 6.3)

Nov, 13, 2013
Big Picture: R&R systems

Environment

Deterministic

Arc Consistency

Search

Vars + Constraints

SLS

Stochastic

Belief Nets

Var. Elimination

Decision Nets

Var. Elimination

Markov Processes

Value Iteration

Problem

Static

Constraint Satisfaction

Query

Sequential

Planning

Representation

Reasoning Technique

CPSC 322, Lecture 2
Key points Recap

• We model the environment as a set of random variables:
  \[ X_1, \ldots, X_n \]
  Joint probability distribution (JPD): \[ P(X_1, \ldots, X_n) \]

• Why the joint is not an adequate representation?

“Representation, reasoning and learning” are “exponential” in ….

Solution: Exploit marginal and conditional independence

\[
\begin{align*}
  P(X|Y) &= P(X) \\
  P(X|Y, Z) &= P(X|Z)
\end{align*}
\]

But how does independence allow us to simplify the joint?

CHAIN RULE.
Lecture Overview

• Belief Networks
  • Build sample BN
  • Intro Inference, Compactness, Semantics
  • More Examples
Belief Nets: Burglary Example

There might be a **burglar** in my house

The **anti-burglar alarm** in my house may go off

I have an agreement with two of my neighbors, **John** and **Mary**, that they **call** me if they hear the alarm go off when I am at work

**Minor earthquakes** may occur and sometimes the set off the alarm.

Variables: **B A M J E** \( n = 5 \)

Joint has \( 2^{5} - 1 \) entries/probs
Belief Nets: Simplify the joint

- Typically order vars to reflect causal knowledge (i.e., causes *before effects*)
  - A burglar (B) can set the alarm (A) off
  - An earthquake (E) can set the alarm (A) off
  - The alarm can cause Mary to call (M)
  - The alarm can cause John to call (J)

\[
P(B, E, A, M, J)
\]

- Apply Chain Rule

\[
P(B) \cdot P(E|B) \cdot P(A|BE) \cdot P(M|AEB) \cdot P(J|MABE)
\]

- Simplify according to marginal&conditional independence
Belief Nets: Structure + Probs

- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities

- Directed Acyclic Graph (DAG)

\[\Rightarrow P(B) \times P(E) \times P(A|B,E) \times P(M|A) \times P(J|A)\]
### Burglary: complete BN

**Burglary**

<table>
<thead>
<tr>
<th>$P(B=T)$</th>
<th>$P(B=F)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>.999</td>
</tr>
</tbody>
</table>

**Earthquake**

<table>
<thead>
<tr>
<th>$P(E=T)$</th>
<th>$P(E=F)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.002</td>
<td>.998</td>
</tr>
</tbody>
</table>

**Alarm**

<table>
<thead>
<tr>
<th>$P(\neg A \mid B,E)$</th>
<th>$P(A \mid B,E)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P(A=T \mid B,E)$</td>
</tr>
<tr>
<td>$T$</td>
<td>.95</td>
</tr>
<tr>
<td>$T$</td>
<td>.94</td>
</tr>
<tr>
<td>$F$</td>
<td>.29</td>
</tr>
<tr>
<td>$F$</td>
<td>.001</td>
</tr>
</tbody>
</table>

**John Calls**

<table>
<thead>
<tr>
<th>$A$</th>
<th>$P(J=T \mid A)$</th>
<th>$P(J=F \mid A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>.90</td>
<td>.10</td>
</tr>
<tr>
<td>$F$</td>
<td>.05</td>
<td>.95</td>
</tr>
</tbody>
</table>

**Mary Calls**

<table>
<thead>
<tr>
<th>$A$</th>
<th>$P(M=T \mid A)$</th>
<th>$P(M=F \mid A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>.70</td>
<td>.30</td>
</tr>
<tr>
<td>$F$</td>
<td>.01</td>
<td>.99</td>
</tr>
</tbody>
</table>

*Call for any other reasons*
Lecture Overview

• Belief Networks
  • Build sample BN
  • Intro Inference, Compactness, Semantics
  • More Examples
Burglary Example: Bnets inference

Our BN can answer any probabilistic query that can be answered by processing the joint!

(Ex1) I'm at work,
- neighbor John calls to say my alarm is ringing,
- neighbor Mary doesn't call.
- No news of any earthquakes.
- Is there a burglar?

(Ex2) I'm at work,
- Receive message that neighbor John called,
- News of minor earthquakes.
- Is there a burglar?

Set digital places to monitor to 5
Burglary Example: Bnets inference

Our BN can answer any probabilistic query that can be answered by processing the joint!

(Ex1) I'm at work,

- neighbor John calls to say my alarm is ringing,
- neighbor Mary doesn't call.
- **No news of any earthquakes.**
- Is there a burglar?

The probability of Burglar will:
- A. Go down
- B. Remain the same
- C. Go up
Bayesian Networks – Inference Types

Diagnostic

- Burglary
  - P(B) = 0.001
  - P(B) = 0.016
- Alarm
- JohnCalls
  - P(J) = 1.0

Predictive

- Burglary
  - P(B) = 1.0
- Alarm
- JohnCalls
  - P(J) = 0.66

Intercausal

- Earthquake
  - P(E) = 1.0
- Alarm
  - P(A) = 1.0
- JohnCalls
  - P(J) = 0.011
  - P(J) = 0.003

Mixed

- Earthquake
  - P(¬E) = 1.0
- Alarm
  - P(A) = 0.003
- JohnCalls
  - P(M) = 1.0
  - P(M) = 0.033
BNnets: Compactness

<table>
<thead>
<tr>
<th>Burglary</th>
<th>Earthquake</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(B=T))</td>
<td>(P(B=F))</td>
</tr>
<tr>
<td>.001</td>
<td>.999</td>
</tr>
<tr>
<td>(P(E=T))</td>
<td>(P(E=F))</td>
</tr>
<tr>
<td>.002</td>
<td>.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B)</th>
<th>(E)</th>
<th>(P(A=T \mid B,E))</th>
<th>(P(A=F \mid B,E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>.95</td>
<td>.05</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>.94</td>
<td>.06</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>.29</td>
<td>.71</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>.001</td>
<td>.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alarm</th>
<th>John Calls</th>
<th>Mary Calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(J=T \mid A))</td>
<td>(P(J=F \mid A))</td>
<td>(P(M=T \mid A))</td>
</tr>
<tr>
<td>T</td>
<td>.90</td>
<td>.10</td>
</tr>
<tr>
<td>F</td>
<td>.05</td>
<td>.95</td>
</tr>
</tbody>
</table>

\[ \text{JPD} = 2^5 - 1 \]

2 + 2 + 4 + 1 + 1 + 1 = 10
BNets: Compactness

In General:
A CPT for boolean $X_i$ with $k$ boolean parents has $2^k$ rows for the combinations of parent values.

Each row requires one number $p_i$ for $X_i = true$.
(the number for $X_i = false$ is just $1-p_i$)

If each variable has no more than $k$ parents, the complete network requires $O(n \cdot 2^k)$ numbers.

For $k<< n$, this is a substantial improvement,

- the numbers required grow linearly with $n$ vs. $O(2^n)$ for the full joint distribution
BNets: Construction General Semantics

The full joint distribution can be defined as the product of conditional distributions:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \] (chain rule)

Simplify according to marginal & conditional independence

- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i)) \]
BNets: Construction General Semantics (cont')

\[ P (X_1, \ldots, X_n) = \prod_{i=1}^{n} P (X_i | \text{Parents}(X_i)) \]

• Every node is independent from its non-descendants given its parents
Lecture Overview

• Belief Networks
  • Build sample BN
  • Intro Inference, Compactness, Semantics
  • More Examples
Other Examples: Fire Diagnosis (textbook Ex. 6.10)

Suppose you want to diagnose whether there is a fire in a building:

- you receive a noisy report about whether everyone is leaving the building.
- if everyone is leaving, this may have been caused by a fire alarm.
- if there is a fire alarm, it may have been caused by a fire or by tampering.
- if there is a fire, there may be smoke raising from the bldg.
Other Examples (cont’)

• Make sure you explore and understand the Fire Diagnosis example (we’ll expand on it to study Decision Networks)

• Electrical Circuit example (textbook ex 6.11)

• Patient’s wheezing and coughing example (ex. 6.14)

• Several other examples on
Realistic BNet: Liver Diagnosis
Source: Onisko et al., 1999

$n \approx 60 \sim 2^{10} \sim (2^{10})^6$

$10^{18}$ entries

$n \leq 60 \times 2^4 = 10^3$
Realistic BNet: Liver Diagnosis
Source: Onisko et al., 1999
Realistic BNet: Liver Diagnosis
Source: Onisko et al., 1999

Assuming there are ~60 nodes in this Bnet and assuming they are all binary, how many numbers are required for the JPD vs BNet

<table>
<thead>
<tr>
<th></th>
<th>JPD</th>
<th>BNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\sim 10^{18}$</td>
<td>$\sim 10^3$</td>
</tr>
<tr>
<td>B</td>
<td>$\sim 10^{30}$</td>
<td>$\sim 10^{18}$</td>
</tr>
<tr>
<td>C</td>
<td>$\sim 10^{13}$</td>
<td>$\sim 10^{14}$</td>
</tr>
<tr>
<td>D</td>
<td>$\sim 10$</td>
<td>$\sim 10^3$</td>
</tr>
</tbody>
</table>
Answering Query under Uncertainty

Probability Theory

Static Belief Network & Variable Elimination

Hidden Markov Models

Dynamic Bayesian Network

Natural Language Processing

Email spam filters

Diagnostic Systems (e.g., medicine)

Monitoring (e.g., credit cards)

Student Tracing in tutoring Systems

Some Applications

you will know

you will know a little

you will know
Learning Goals for today’s class

You can:
Build a Belief Network for a simple domain

Classify the types of inference
Diagnostic, Predictive, Intercausal, Mixed

Compute the representational saving in terms on number of probabilities required
Next Class (Wednesday!)

Bayesian Networks Representation

- **Additional Dependencies** encoded by BNets
- More **compact representations** for CPT
- Very simple but extremely useful Bnet (**Bayes Classifier**)
Belief network summary

• A belief network is a directed acyclic graph (DAG) that effectively expresses independence assertions among random variables.

• The parents of a node $X$ are those variables on which $X$ directly depends.

• Consideration of causal dependencies among variables typically help in constructing a Bnet.