Bayesian belief networks

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Administration

- **Problem set 6 is due today**
- **Problem set 7 is out:**
  - Due on November 5
  - No programming part

- **Midterms:**
  - See the instructor
- **PS 1-5:**
  - See the TA
Modeling uncertainty with probabilities

- We need to define the full joint probability distribution over random variables defining the domain of interest.
- With the known full joint we are able to handle an arbitrary probabilistic inference problem.

Problems:

- **Space complexity.** To store a full joint distribution we need to remember $O(d^n)$ numbers.
  - $n$ – number of random variables, $d$ – number of values
- **Inference (time) complexity.** To compute some queries requires $O(d^n)$ steps.
- **Acquisition problem.** Who is going to define all of the probability entries?

Medical diagnosis example.

- **Space complexity.**
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: 2*2*2*3*2=48
  - We need to define at least 47 probabilities.
- **Time complexity.**
  - Assume we need to compute the marginal of Pneumonia=T from the full joint distribution
    \[
    P(Pneumonia = T) = \sum_{i=T,F} \sum_{j=T,F} \sum_{k=h,n,l} \sum_{u} P(Pneumonia = i, Fever = j, Cough = k, WBCcount = l, Pale = u)
    \]
  - Sum over: 2*2*3*2=24 combinations
Bayesian belief networks (BBNs)

Bayesian belief networks.
- Represent the full joint distribution over the variables more compactly with a smaller number of parameters.
- Take advantage of conditional and marginal independences among random variables.

- **A and B are independent**
  \[ P(A, B) = P(A)P(B) \]

- **A and B are conditionally independent given C**
  \[ P(A, B \mid C) = P(A \mid C)P(B \mid C) \]
  \[ P(A \mid C, B) = P(A \mid C) \]

Bayesian belief networks (general)

Two components: \( B = (S, \Theta_S) \)
- **Directed acyclic graph**
  - Nodes correspond to random variables
  - (Missing) links encode independences

- **Parameters**
  - Local conditional probability distributions for every variable-parent configuration

\[
P(X_i \mid pa(X_i))
\]

Where:
\( pa(X_i) \) - stand for parents of \( X_i \)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>0.94</td>
<td>0.06</td>
</tr>
<tr>
<td>F</td>
<td>0.29</td>
<td>0.71</td>
</tr>
<tr>
<td>F</td>
<td>0.001</td>
<td>0.999</td>
</tr>
</tbody>
</table>
**Bayesian belief network.**

```
<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>E</th>
<th>Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>Earthquake</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>P(A</td>
<td>B,E)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>E</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>0.06</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0.71</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>P(J</td>
<td>A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>0.90</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>P(M</td>
<td>A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>0.7</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.01</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>
```

**Full joint distribution in BBNs**

**Full joint distribution** is defined in terms of local conditional distributions (obtained via the chain rule):

\[
P(X_1, X_2, \ldots, X_n) = \prod_{i=1,\ldots,n} P(X_i \mid pa(X_i))
\]

**Example:**

Assume the following assignment of values to random variables:

\[B = T, E = T, A = T, J = T, M = F\]

Then its probability is:

\[
P(B = T, E = T, A = T, J = T, M = F) =
P(B = T)P(E = T)P(A = T \mid B = T, E = T)P(J = T \mid A = T)P(M = F \mid A = T)
\]

CS 1571 Intro to AI
Bayesian belief networks (BBNs)

Bayesian belief networks
- Represents the full joint distribution over the variables more compactly using the product of local conditionals.
- But how did we get to local parameterizations?

Answer:
- **Graphical structure** encodes **conditional and marginal independences** among random variables.
- **A and B are independent** \( P(A, B) = P(A)P(B) \)
- **A and B are conditionally independent given C**
  \[
  P(A | C, B) = P(A | C) \\
  P(A, B | C) = P(A | C)P(B | C)
  \]
- **The graph structure implies the decomposition !!!**

Independences in BBNs

3 basic independence structures:

1. **Burglary**
   - **Alarm**
   - **JohnCalls**

2. **Burglary**
   - **Earthquake**
   - **Alarm**
   - **JohnCalls**
   - **MaryCalls**
Independences in BBNs

1. JohnCalls is independent of Burglary given Alarm
   \[ P(J \mid A, B) = P(J \mid A) \]
   \[ P(J, B \mid A) = P(J \mid A)P(B \mid A) \]

2. Burglary is independent of Earthquake (not knowing about the Alarm)
   \[ P(B, E) = P(B)P(E) \]
   But Burglary and Earthquake become dependent once I know the Alarm!!
Independences in BBNs

1. MaryCalls is independent of JohnCalls given Alarm
   \[
   P(J \mid A, M) = P(J \mid A)
   \]
   \[
   P(J, M \mid A) = P(J \mid A)P(M \mid A)
   \]

Independences in BBN

- BBN distribution models many conditional independence relations among distant variables and sets of variables
- These are defined in terms of the graphical criterion called d-separation
- **D-separation and independence**
  - Let X, Y and Z be three sets of nodes
  - If X and Y are d-separated by Z, then X and Y are conditionally independent given Z
- **D-separation**
  - A is d-separated from B given C if every undirected path between them is blocked
- **Path blocking**
  - 3 cases that expand on the three basic independence structures
Undirected path blocking

A is d-separated from B given C if every undirected path between them is blocked

• 1. Path blocking with a linear substructure

X  \[\longrightarrow\]  Z  \[\longrightarrow\]  Y

X in A  \[\quad\]  Z in C  \[\quad\]  Y in B

• 2. Path blocking with the wedge substructure

X  \[\longrightarrow\]  Z  \[\longrightarrow\]  Y

X in A  \[\quad\]  Z in C  \[\quad\]  Y in B
Undirected path blocking

A is d-separated from B given C if every undirected path between them is **blocked**

- 3. Path blocking with the vee substructure

X in A

Y in B

X \( \rightarrow \) Z \( \leftarrow \) Y

Z or any of its descendants **not** in C

Independences in BBNs

- Earthquake and Burglary are independent given MaryCalls

\[ \text{?} \]
Independences in BBNs

- Earthquake and Burglary are independent given MaryCalls \( F \)
- Burglary and MaryCalls are independent (not knowing Alarm) \(?\)

---

Independences in BBNs

- Earthquake and Burglary are independent given MaryCalls \( F \)
- Burglary and MaryCalls are independent (not knowing Alarm) \( F \)
- Burglary and RadioReport are independent given Earthquake \(?\)
Independences in BBNs

- Earthquake and Burglary are independent given MaryCalls \( F \)
- Burglary and MaryCalls are independent (not knowing Alarm) \( F \)
- Burglary and RadioReport are independent given Earthquake \( T \)
- Burglary and RadioReport are independent given MaryCalls \( ? \)
Bayesian belief networks (BBNs)

Bayesian belief networks
- Represents the full joint distribution over the variables more compactly using the product of local conditionals.
- So how did we get to local parameterizations?

\[ P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid pa(X_i)) \]

- The decomposition is implied by the set of independences encoded in the belief network.

Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:

\[ P(B=T, E=T, A=T, J=T, M=F) = \]
Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:

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

\[ = P(J = T \mid B = T, E = T, A = T, M = F) P(B = T, E = T, A = T, M = F) \]

\[ = P(J = T \mid A = T) P(B = T, E = T, A = T, M = F) \]

CS 1571 Intro to AI
Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:

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

\[ = P(J=T \mid B=T, E=T, A=T, M=F)P(B=T, E=T, A=T, M=F) \]
\[ = P(J=T \mid A=T)P(B=T, E=T, A=T, M=F) \]
\[ = P(M=F \mid B=T, E=T, A=T)P(B=T, E=T, A=T) \]
\[ = P(M=F \mid A=T)P(B=T, E=T, A=T) \]
\[ = P(A=T \mid B=T, E=T)P(B=T, E=T) \]

CS 1571 Intro to AI
Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:

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

\[ = P(J=T \mid B=T, E=T, A=T, M=F)P(B=T, E=T, A=T, M=F) \]

\[ = P(J=T \mid A=T)P(B=T, E=T, A=T, M=F) \]

\[ = P(J=T \mid A=T)P(M=F \mid B=T, E=T, A=T)P(B=T, E=T, A=T) \]

\[ = P(A=T \mid B=T, E=T)P(B=T, E=T) \]

\[ = P(B=T)P(E=T) \]

\[ = P(J=T \mid A=T)P(M=F \mid A=T)P(A=T \mid B=T, E=T)P(B=T)P(E=T) \]

Bayesian belief network.

- In the BBN the **full joint distribution** is expressed using a set of local conditional distributions.
Parameter complexity problem

- In the BBN the **full joint distribution** is

\[ P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i)) \]

- **What did we save?**

**Parameters:**
- full joint: \( 2^5 = 32 \)
- BBN: \( 2^3 + 2(2^2) + 2(2) = 20 \)

**Parameters to be defined:**
- full joint: \( 2^5 - 1 = 31 \)
- BBN: \( 2^2 + 2(2) + 2(1) = 10 \)

---

Model acquisition problem

**The structure of the BBN**
- typically reflects causal relations
  (BBNs are also sometime referred to as **causal networks**)
- Causal structure is intuitive in many applications domain and it is relatively easy to define to the domain expert

**Probability parameters of BBN**
- are conditional distributions relating random variables and their parents
- Complexity is much smaller than the full joint
- It is much easier to obtain such probabilities from the expert or learn them automatically from data