Modeling and Reasoning with Bayesian Networks
Software Packages for BNs

Software Packages for Graphical Models / Bayesian Networks
http://www.cs.ubc.ca/ murphyk/Bayes/bnsoft.html

SamIam from UCLA

GeNIe/SMILE from the University of UPitt
http://www2.sis.pitt.edu/ genie/

Hugin lite from Hugin:
http://www.hugin.com

MSBN from Microsoft Research:
http://www.research.microsoft.com/dtas/msbn/
Reasoning with BNs

What types of queries can be posed to a Bayesian network?

Probability of Evidence: the probability of some variable instantiation \( e, Pr(e) \)

\[
Pr(X = yes, D = no) ?
\]

The variables \( E = \{X, D\} \) are called evidence variables

Other types of evidence?

\[
Pr(X = yes \lor D = no) ?
\]
A Bayesian Network

- Visit to Asia? (A)
- Smoker? (S)
- Tuberculosis? (T)
- Lung Cancer? (C)
- Bronchitis? (B)
- Tuberculosis or Cancer? (P)
- Positive X-Ray? (X)
- Dyspnoea? (D)
Bayesian network tools do not usually provide direct support for computing the probability of arbitrary pieces of evidence.

But such probabilities can be computed indirectly.

The *Case-Analysis Method*:

\[
Pr(X = yes \vee D = yes) = Pr(X = yes, D = yes) + Pr(X = yes, D = no) + Pr(X = no, D = yes)
\]

This can always be done, but is only practical when the number of evidence variables \( E \) is relatively small.
The Auxiliary-Node Method: We can add an auxiliary node $E$ to the network, declare nodes $X$ and $D$ as the parents of $E$, and then adopt the following CPT for $E$:

| $x$ | $d$ | $e$   | $\Pr(e|x, d)$ |
|-----|-----|-------|---------------|
| yes | yes | yes   | 1             |
| yes | no  | yes   | 1             |
| no  | yes | yes   | 1             |
| no  | no  | yes   | 0             |

There are some techniques for representing deterministic CPTs which do not suffer from exponential blowup.
Reasoning with BNs

- Prior and posterior marginals: $P(s)$, $P(s|e)$
- $S \subset V$ is small
- Most available BN tools support only marginals over single variables
- Though the algorithms underlying these tools are capable of computing some other types of marginals
- Most tools do not provide direct support for accommodating soft evidence, with the expectation that users will utilize the method of virtual evidence for this purpose
**Reasoning with BNs**

- **Most probable explanation (MPE):** identify an instantiation $x_1, \ldots, x_n$ for which $P(x_1, \ldots, x_n|e)$ is maximal.

- Identify the most probable instantiation of network variables given some evidence

- Choosing each value $x_i$ so as to maximizes the probability $Pr(x_i|e)$ does not necessarily lead to a most probable explanation
Reasoning with BNs

- Maximum a posteriori hypothesis (MAP): find an instantiation \( m \) of variables \( M \subset V \) for which \( P(m|e) \) is maximal
- Finding the most probable instantiation for a subset of network variables
- The variables in \( M \) are known as MAP variables
- MPE is a special case of MAP, and is much easier to compute algorithmically
- Very little support for this type of queries in BN tools
- A common method for approximating MAP is to compute an MPE and then project the result on the MAP variables.
Modeling with Bayesian networks

1. Define the network variables and their values.
   - Query variables
   - Evidence variables
   - Intermediary variables

2. Define the network structure.
   - Guided by causal interpretation of network structure.
   - What is the set of variables that we regard as the direct causes of $X$?

3. Define the CPTs.
Diagnosis I: medical diagnosis
The flu is an acute disease characterized by fever, body aches and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils which leads to a sore throat and can be associated with fever.
BNs for medical diagnosis

- Cold?
- Flu?
- Tonsillitis?

- Chilling?
- Body Ache?
- Sore Throat?
- Fever?

- Condition
- Chilling?
- Body Ache?
- Sore Throat?
- Fever?
Specification of CPTs

- The CPT for a condition, such as tonsillitis, must provide the belief in developing tonsillitis by a person about whom we have no knowledge of any symptoms.

- The CPT for a symptom, such as chilling, must provide the belief in this symptom under the possible conditions.

- The probabilities are usually obtained from a medical expert, based on known medical statistics or subjective beliefs gained through practical experience.

- Another key method for specifying the CPTs is by estimating them directly from medical records of previous patients.
Diagnosis II: medicine diagnosis

A few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is a scanning test which has a false positive of 1% and a false negative of 10%. The second is a blood test, which detects progesterone with a false positive of 10% and a false negative of 30%. The third test is a urine test, which also detects progesterone with a false positive of 10% and a false negative of 20%. The probability of a detectable progesterone level is 90% given pregnancy, and 1% given no pregnancy. The probability that insemination will impregnate a cow is 87%.
A Bayesian Network

Pregnant? (P)

Progestrone Level (L)

Scanning Test (S)

Urine Test (U)

Blood Test (B)

<table>
<thead>
<tr>
<th>P</th>
<th>θ_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>.87</td>
</tr>
<tr>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

| P | S  | θ_{S|P} |
|---|----|--------|
| yes | -ve | .10    |
| no  | +ve | .01    |

| P | L     | θ_{L|P} |
|---|-------|--------|
| yes | undetectable | .10 |
| no  | detectable    | .01   |

| L | B     | θ_{B|L} |
|---|-------|--------|
| detectable | -ve | .30    |
| undetectable | +ve | .10    |

| L | U     | θ_{U|L} |
|---|-------|--------|
| detectable | -ve | .20    |
| undetectable | +ve | .10    |
Suppose now that a farmer is not too happy with this and would like three negative tests to drop the probability of pregnancy to no more than 5%.

The farmer is willing to buy more accurate test kits for this purpose, but needs to know the false positive and negative rates of the new tests, which would ensure the above constraint

**Sensitivity analysis**: understand the relationship between the parameters of a Bayesian network, and the conclusions drawn based on the network
Sensitivity Analysis

Which network parameters do we have to change, and by how much, so as to ensure that the probability of pregnancy given three negative tests would be no more than 5%?

Samlam tool

1. If the false negative rate for the scanning test were about 4.63% instead of 10%.

2. If the probability of pregnancy given insemination were about 75.59% instead of 83%.

3. If the probability of a detectable progesterone level given pregnancy were about 99.67% instead of 90%.
What is interesting about the above results of sensitivity analysis is that they imply that improving the blood and urine tests cannot help.

Sensitivity analysis is an important mode of analysis when developing Bayesian networks.

It can be performed quite efficiently since its computational complexity is similar to that of computing posterior marginals.
Network Granularity

- One of the issues that arises when building Bayesian networks: How fine grained should the network be? Do we need to include an intermediary variable in the network?
- Modeling convenience
- Bypass: the process of removing a variable, redirecting its parents to its children, and then updating the CPTs of these children.
A Bayesian Network

<table>
<thead>
<tr>
<th>$P$</th>
<th>$\theta_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>.87</td>
</tr>
</tbody>
</table>

| $P$ | $S$ | $\theta_s|_p$ |
|-----|-----|-------------|
| yes | $-$ve | .10         |
| no  | $+$ve | .01         |

| $P$ | $B$ | $\theta_b|_p$ |
|-----|-----|-------------|
| yes | $-$ve | .36         |
| no  | $+$ve | .106        |

| $P$ | $U$ | $\theta_u|_p$ |
|-----|-----|-------------|
| yes | $-$ve | .27         |
| no  | $+$ve | .107        |
Network Granularity

- Intermediary variables cannot be bypassed in general.

- A general case in which an intermediary variable can be bypassed without affecting model accuracy $(Pr(q, e) = Pr'(q, e))$:
  - A node $X$ which has a single child $Y$

$$
\theta'_{y|uv} = \sum_{x} \theta_{y|xv} \theta_{x|u}
$$

$U$ are the parents of $X$ and $V$ are the other parents of $Y$

- Even though a variable may be bypassed without affecting the model accuracy, one may wish not to bypass it simply because the bypass procedure will lead to a large CPT.
Diagnosis III: digital circuit

Consider a digital circuit. Given some values for the circuit primary inputs and output (test vector), our goal is to decide whether the circuit is behaving normally. If not, our goal is then to decide the most likely health states of its components.
The BN structures can be generated automatically by software
The values of variables representing circuit wires (primary inputs, outputs, or internal wires):
\{low, high\}

The values for health variables: \{ok, faulty\}, or \{ok, stuckat0, stuckat1\}

Specifying CPTs: if our goal is to compute the probability of some health state \(x, y, z\) given some test vector \(a, b, e\), then this probability \(Pr(h|i, o)\) is independent of the probabilities \(Pr(a)\) and \(Pr(b)\)
## Diagnosis III: digital circuit

| $X$   | $\theta_x$ | $A$ | $X$ | $C$ | $\theta_{c|a,x}$ |
|-------|------------|-----|-----|-----|------------------|
| ok    | .99        | high| ok  | high| 0                |
| faulty| .01        | low | ok  | high| 1                |
|       |            | high| faulty| high| .5              |
|       |            | low | faulty| high| .5              |
Diagnosis III: digital circuit

| $X$    | $\theta_x$ | $A$      | $X$ | $C$ | $\theta_c|a,x$ |
|--------|------------|----------|-----|-----|---------------|
| ok     | .99        | high     | ok  | high| 0             |
| stuckat0 | .005   | low      | ok  | high| 1             |
| stuckat1 | .005   | high     | stuckat0 | high| 0             |
|        |            | low      | stuckat0 | high| 0             |
|        |            | high     | stuckat1 | high| 1             |
|        |            | low      | stuckat1 | high| 1             |
Diagnosis III: digital circuit

The network with fault modes satisfies the following crucial property: given the values of health variables \((X, Y, Z)\), and given the values of input/output variables \((A, B, E)\), there is at most one instantiation of the remaining variables \((C, D)\) which is consistent with these values.

MAP queries on this network, where MAP variables are \(X, Y, Z\), and evidence variables are \(A, B, E\), can be reduced to MPE queries by simply projecting the result of an MPE query on the MAP variables \(X, Y,\) and \(Z\).

This has a major computational implication.
Diagnosis III: digital circuit

- The extension of the diagnosis problem: we assume that we have two test vectors instead of only one.
- Our goal now is to find the most probable health state of the circuit given these two test vectors.
- Does the health of a component stay the same during each of the two tests?
- Do we want to allow for the possibility of intermittent faults?
BNs for diagnosis
Channel Coding

We need to send four bits $U_1$, $U_2$, $U_3$, and $U_4$ from a source $S$ to a destination $D$ over a noisy channel, where there is a 1% chance that a bit will be inverted before it gets to the destination. To improve the reliability of this process, we will add three redundant bits $X_1$, $X_2$, and $X_3$ to the message, where $X_1$ is the XOR of $U_1$ and $U_3$, $X_2$ is the XOR of $U_2$ and $U_4$, and $X_3$ is the XOR of $U_1$ and $U_4$. Given that we received a message containing seven bits at destination $D$, our goal is to restore the message generated at the source $S$. 

BNs for Channel Coding

\[ \begin{align*}
U_1 & \rightarrow Y_1 \\
U_2 & \rightarrow Y_2 \\
U_3 & \rightarrow Y_3 \\
U_4 & \rightarrow Y_4 \\
X_1 & \rightarrow Y_5 \\
X_2 & \rightarrow Y_6 \\
X_3 & \rightarrow Y_7
\end{align*} \]
Channel Coding

Decoder quality measures

- Word Error Rate (WER)
- Bit Error Rate (BER)

Queries to pose

- MPE
- Posterior Marginal $Pr(u_i|y_1, \ldots, y_7)$
When SamBot goes home at night, he wants to know if his family is home before he tries the doors. (Perhaps the most convenient door to enter is double locked when nobody is home). Often when SamBot’s wife leaves the house she turns on an outdoor light. However, she sometimes turns on this light if she is expecting a guest. Also, SamBot’s family has a dog. When nobody is home, the dog is in the back yard. The same is true if the dog has bowel trouble. Finally, if the dog is in the back yard, SamBot will probably hear her barking, but sometimes he can be confused by other dogs barking. SamBot is equipped with two sensors: a light-sensor for detecting outdoor lights and a sound-sensor for detecting the barking of dogs. Both of these sensors are not completely reliable and can break. Moreover, they both require SamBot’s battery to be in good condition.
BN for SamBot problem

ExpectingCompany → FamilyHome → OutdoorLight → LightSensor → LightSensorBroken → Battery → SoundSensor → SoundSensorBroken → OtherBarking → DogOutside → DogBarking → DogBowel
Dealing with Large CPTs

One of the major issues that arise when building Bayesian network models is the potentially large size of CPTs.

One approach for dealing with large CPTs is to try to develop a micro model which details the relationship between the parents and their common child.

The goal here is to reveal the local structure of this relationship in order to specify it using a smaller number of parameters than $2^n$. 

Noisy-or model

\[ C_1 \quad C_2 \quad \ldots \quad C_n \]

\[ E \]

\[ C_1 \quad Q_1 \quad C_2 \quad Q_2 \quad \ldots \quad C_n \quad Q_n \]

\[ E \]

\[ L \]
Interpret parents $C_1, \ldots, C_n$ as causes, and variable $E$ as their common effect.

The intuition is that each cause $C_i$ is capable of establishing the effect $E$ on its own, regardless of other causes, except under some unusual circumstances which are summarized by the suppressor variable $Q_i$.

When the suppressor $Q_i$ of cause $C_i$ is active, cause $C_i$ is no longer able to establish $E$.

The leak variable $L$ is meant to represent all other causes of $E$ which were not modelled explicitly. Hence, even when none of the causes $C_i$ is active, the effect $E$ may still be established by the leak variable $L$. 
Noisy-or model

Cold?  Flu?  Tonsillitis?

Chilling?  Body Ache?  Sore Throat?  Fever?

Condition

Chilling?  Body Ache?  Sore Throat?  Fever?
Noisy-or model

The noisy-or model can be specified using $n + 1$ parameters

$$
\theta_{q_i} = Pr(Q_i = \text{active})
$$

$$
\theta_l = Pr(L = \text{active})
$$

The full CPT for variable $E$, with its $2^n$ parameters, can be induced from the $n + 1$ parameters

$$
P(E = \text{passive}|\alpha) = (1 - \theta_l) \prod_{i \in I\alpha} \theta_{q_i}
$$

where $\alpha$ is an instantiation of the parents $C_1, \ldots, C_n$, and $I\alpha$ is the set of the indices of causes in $\alpha$ that are active.
Other Representations of CPTs

- The noisy-or model is only one of several other models for local structure. Each one of these models is based on some assumption about the way parents interact with their common child.
- Most often, we have some local structure in the relationship between a node and its parents, but that structure does not fit nicely into any of the existing canonical models such as noisy-or.
- For these irregular structures, there are several non-tabular representations that are not necessarily exponential in the number of parents.
## Other Representations of CPTs

### Decision Trees

A decision tree for a set of CPTs is shown below, along with a table of probabilities for each variable.

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Pr(E=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Other Representations of CPTs

If-Then Rules

If $C_1 = 1$ then $Pr(E = 1) = 0.0$
If $C_1 = 0 \land C_2 = 1$ then $Pr(E = 1) = 0.9$
If $C_1 = 0 \land C_2 = 0 \land C_3 = 1$ then $Pr(E = 1) = 0.3$
If $C_1 = 0 \land C_2 = 0 \land C_3 = 0 \land C_4 = 1$ then $Pr(E = 1) = 0.6$
If $C_1 = 0 \land C_2 = 0 \land C_3 = 0 \land C_4 = 0$ then $Pr(E = 1) = 0.8$
Other Representations of CPTs

Deterministic CPTs can be represented compactly by a set of propositional sentences

| A       | X       | C       | $\theta_{c|a,x}$ |
|---------|---------|---------|-----------------|
| high    | ok      | high    | 0               |
| low     | ok      | high    | 1               |
| high    | stuckat0| high    | 0               |
| low     | stuckat0| high    | 0               |
| high    | stuckat1| high    | 1               |
| low     | stuckat1| high    | 1               |

$(X = ok \land A = high) \lor X = stuckat0 \implies C = low$

$(X = ok \land A = low) \lor X = stuckat1 \implies C = high$
Other Representations of CPTs

A word of caution on how these representations of CPTs are sometimes used by Bayesian network tools

- Many of these tools will expand these representations into their corresponding CPTs before they perform inference.

- In such a case, these representations are only being utilized in addressing the modeling problem since the size of expanded CPT is still exponential in the number of parents.

- The reason why these tools perform this expansion before inference is that many algorithms for inference in Bayesian networks require a tabular representation of CPTs as they cannot operate on the above representations directly.