Probabilistic graphical models: current research activities

Jirka Vomlel

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
http://www.utia.cz/vomlel

Aalborg, Denmark, November, 20, 2013
A simple Bayesian network model - Chest Clinic

![Bayesian network diagram]

- **Visit to Asia**
- **Smoker**
- **Tuberculosis**
- **Cancer**
- **Bronchitis**
- **RTG**
- **Dyspnoea**
A simple Bayesian network model - Chest Clinic

Conditional probability tables (CPTs)

\[ P(\text{Visit to Asia}) \]

\[ P(\text{Smoker}) \]

\[ P(\text{Tuberculosis} \mid \text{Visit to Asia}) \]

\[ P(\text{Cancer} \mid \text{Smoker}) \]

\[ P(\text{Bronchitis} \mid \text{Smoker}) \]

\[ P(\text{RTG} \mid \text{Tuberculosis, Cancer}) \]

\[ P(\text{Dyspnoea} \mid \text{Tuberculosis, Cancer, Bronchitis}) \]
Probabilistic inference with the Bayesian network

\[ P(X|\text{Smoker}=\text{true}) \]
Probabilistic inference with the Bayesian network

\[ P(X | \text{Smoker} = \text{true}, \text{Dyspnoea} = \text{true}) \]
Probabilistic inference with the Bayesian network

\[ P(X | \text{Smoker=true, Dyspnoea=true, RTG=true}) \]
Probabilistic inference with the Bayesian network

\[ P(X | \text{Smoker} = \text{true}, \text{Dyspnoea} = \text{true}, \text{RTG} = \text{true}, \text{Visit to Asia} = \text{true}) \]
First, assume a deterministic function. RTG is positive iff the patient has tuberculosis or cancer.

<table>
<thead>
<tr>
<th>RTG</th>
<th>Tuberculosis</th>
<th>Cancer</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
CPT $P(\text{RTG} \mid \text{Tuberculosis, Cancer})$

RTG can have other reasons for being positive and RTG need not be positive even if the patient has tuberculosis or cancer.

<table>
<thead>
<tr>
<th>RTG</th>
<th>Tuberculosis</th>
<th>Cancer</th>
<th>$p$</th>
<th>$p'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$p_0$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$p_0 \times p_1$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$p_0 \times p_2$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$p_0 \times p_1 \times p_2$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$1 - p_0$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>$1 - p_0 \times p_1$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$1 - p_0 \times p_2$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$1 - p_0 \times p_1 \times p_2$</td>
</tr>
</tbody>
</table>

$p_0, p_1, p_2 \in \langle 0, 1 \rangle$
RTG can have other reasons for being positive and RTG need not be positive even if the patient has tuberculosis or cancer.

<table>
<thead>
<tr>
<th>RTG</th>
<th>Tuberculosis</th>
<th>Cancer</th>
<th>( p )</th>
<th>( p' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>( p_0 )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>( p_0 \cdot p_1 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>( p_0 \cdot p_2 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( p_0 \cdot p_1 \cdot p_2 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( 1 - p_0 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>( 1 - p_0 \cdot p_1 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( 1 - p_0 \cdot p_2 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( 1 - p_0 \cdot p_1 \cdot p_2 )</td>
</tr>
</tbody>
</table>

\( p_0, p_1, p_2 \in \langle 0, 1 \rangle \)

This local model is called ”noisy-or”. 
RTG can have other reasons for being positive and RTG need not be positive even if the patient has tuberculosis or cancer.

<table>
<thead>
<tr>
<th>RTG</th>
<th>Tuberculosis</th>
<th>Cancer</th>
<th>p</th>
<th>p'</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.95</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.019</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.019</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.00038</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 − 0.05</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1 − 0.007</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1 − 0.981</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1 − 0.99962</td>
</tr>
</tbody>
</table>

p₀, p₁, p₂ ∈ ⟨0, 1⟩

This local model is called "noisy-or".
Let \( k \) be the number of parents. We need to specify \( k + 1 \) values \( p₀, p₁, \ldots, pₖ \) instead of \( 2^k \) in a general CPT.
Current research activities

- Model elicitation
Current research activities

- Model elicitation
  - learning models from data (using Integer Programming)
  - learning models with local structure of a noisy-or like type.
  - combination of expert knowledge and data (biological pathways and experimental data)
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
  - exploiting determinism
  - exploiting local structure of CPTs
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
  - iterative refinement
  - anytime inference methods
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
- Other types of probabilistic graphical models:
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
- Other types of probabilistic graphical models:
  - models with continuous variables (other than Gaussian)
  - models with causal interpretation of directed edges
  - models with both directed and undirected edges in the model (e.g. chain graphs)
  - modeling temporal and spatial information.
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
- Other types of probabilistic graphical models:
- Finding good strategies with the help of a BN:
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
- Other types of probabilistic graphical models:
  - Finding good strategies with the help of a BN:
    - Decision-Theoretic Troubleshooting
    - Adaptive Testing
Current research activities

- Model elicitation
- Efficient inference with special types of probabilistic models
- Methods of approximate inference
- Other types of probabilistic graphical models:
- Finding good strategies with the help of a BN:
- Classification and regression for medical applications:
Current research activities

• Model elicitation
• Efficient inference with special types of probabilistic models
• Methods of approximate inference
• Other types of probabilistic graphical models:
• Finding good strategies with the help of a BN:
• Classification and regression for medical applications:
  – mortality prediction
  – prediction of medical care costs