Applications

- Medical diagnosis
- Genetic linkage analysis
- Probabilistic planning
Constructing Bayesian Networks

- Define network variables & their domains
  - Query & evidence variables, usually from problem statement
  - Intermediary variables, harder to determine

- Define network structure (edges)
  - What variables are direct causes of $X$?

- Define network CPTs
  - Determined objectively from problem statement
  - Reflection of subjective beliefs
  - Estimated from data
Medical Diagnosis

- Flu characterized by fever and body aches, can be associated with chilling and sore throat
- Cold associated with chilling, can cause sore throat
- Tonsillitis (inflammation of tonsils) causes sore throat, can be associated with fever
Medical Diagnosis

- Flu characterized by fever and body aches, can be associated with chilling and sore throat
- Cold associated with chilling, can cause sore throat
- Tonsillitis (inflammation of tonsils) causes sore throat, can be associated with fever
- Capture this knowledge with Bayesian network, diagnose condition given symptoms
Medical Diagnosis

- Flu characterized by fever and body aches, can be associated with chilling and sore throat
- Cold associated with chilling, can cause sore throat
- Tonsillitis (inflammation of tonsils) causes sore throat, can be associated with fever

- Capture this knowledge with Bayesian network, diagnose condition given symptoms
- Query, evidence, intermediary variables?
Medical Diagnosis

- Flu characterized by fever and body aches, can be associated with chilling and sore throat
- Cold associated with chilling, can cause sore throat
- Tonsillitis (inflammation of tonsils) causes sore throat, can be associated with fever

- Capture this knowledge with Bayesian network, diagnose condition given symptoms
- Query, evidence, intermediary variables?
- Flu, cold, tonsillitis;
Medical Diagnosis

- Flu characterized by fever and body aches, can be associated with chilling and sore throat
- Cold associated with chilling, can cause sore throat
- Tonsillitis (inflammation of tonsils) causes sore throat, can be associated with fever

- Capture this knowledge with Bayesian network, diagnose condition given symptoms
- Query, evidence, intermediary variables?
- Flu, cold, tonsillitis; chilling, body ache, sore throat, fever
Medical Diagnosis: Network Structure

- Cold
- Flu
- Tonsillitis
- Chilling
- Body Ache
- Sore Throat
- Fever
Medical Diagnosis: Network Structure

- Cold
- Flu
- Tonsillitis

- Chilling
- Body Ache
- Sore Throat
- Fever
Medical Diagnosis: Network Structure

- **Cold**
  - Chilling
  - Body Ache

- **Flu**
  - Chilling
  - Body Ache
  - Sore Throat

- **Tonsillitis**
  - Sore Throat
  - Fever
Medical Diagnosis: Naive Bayes

Condition

Chilling  Body Ache  Sore Throat  Fever

Condition: normal, cold, flu, tonsillitis

Cold  Flu  Tonsillitis

Chilling  Body Ache  Sore Throat  Fever
Medical Diagnosis: Naive Bayes

Condition: normal, cold, flu, tonsillitis

Model inaccuracy: *single fault* assumption
Medical Diagnosis: Naive Bayes

Condition: normal, cold, flu, tonsillitis

Model inaccuracy: *single fault* assumption

- Cold $\Rightarrow$ Fever & Sore Throat independent
Medical Diagnosis: Naive Bayes

Condition: normal, cold, flu, tonsillitis

Model inaccuracy: *single fault* assumption

- Cold ⇒ Fever & Sore Throat independent
- Body Ache ⇒ Flu ↑ ⇒ Cold & Tonsillitis ↓
Medical Diagnosis: Naive Bayes

Condition: normal, cold, flu, tonsillitis

Model inaccuracy: single fault assumption

- Cold ⇒ Fever & Sore Throat independent
- Body Ache ⇒ Flu ↑ ⇒ Cold & Tonsillitis ↓
- No Fever ⇒ Cold ↑
Medical Diagnosis: CPTs

From medical experts, based on known stats or subjective beliefs

- CPTs for conditions (roots of graph): $\text{Pr}(\text{condition})$ without knowledge of symptoms
- CPTs for symptoms: $\text{Pr}(\text{Chilling}|\text{Cold, Flu})$ etc
Medical Diagnosis: CPTs

From medical experts, based on known stats or subjective beliefs

- CPTs for conditions (roots of graph): Pr(condition) without knowledge of symptoms
- CPTs for symptoms: Pr(Chilling|Cold, Flu) etc

Estimated directly from medical records

<table>
<thead>
<tr>
<th>Case</th>
<th>Cold</th>
<th>Flu</th>
<th>Tonsillitis</th>
<th>Chilling</th>
<th>Body Ache</th>
<th>Sore Throat</th>
<th>Fever</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T</td>
<td>F</td>
<td>?</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>?</td>
<td>T</td>
<td>F</td>
<td>?</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Maximize \[ \prod_{i=1}^{N} \Pr(d_i) \]
Given symptoms (chilling, body ache, sore throat, fever), pose MAP query on Cold, Flu, & Tonsillitis

Reduces to MPE if evidence covers all four symptoms
Genetic Linkage Analysis

- Goal: map genes onto chromosome (i.e., determine location of genes on chromosome)
- Useful for detecting and predicting diseases
- Input: pedigree, observed genotype/phenotype
Pedigree

- **Genes:** $G_1$, $G_2$
- **Gene has alleles (states):** $G_1(A, a)$, $G_2(B, b)$
- **Individual carries two alleles per gene, from father & mother**
- **Genotype:** collection of alleles of individual, composed of two haplotypes
- **Phenotype:** observable traits of individual
Genotype and Phenotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/A</td>
<td>blood type A</td>
</tr>
<tr>
<td>A/B</td>
<td>blood type AB</td>
</tr>
<tr>
<td>A/O</td>
<td>blood type A</td>
</tr>
<tr>
<td>B/B</td>
<td>blood type B</td>
</tr>
<tr>
<td>B/O</td>
<td>blood type B</td>
</tr>
<tr>
<td>O/O</td>
<td>blood type O</td>
</tr>
</tbody>
</table>

- Deterministic
- However, not in reverse direction: Blood type A $\Rightarrow$ A/A or A/O
### Genotype and Phenotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/A</td>
<td>blood type A</td>
</tr>
<tr>
<td>A/B</td>
<td>blood type AB</td>
</tr>
<tr>
<td>A/O</td>
<td>blood type A</td>
</tr>
<tr>
<td>B/B</td>
<td>blood type B</td>
</tr>
<tr>
<td>B/O</td>
<td>blood type B</td>
</tr>
<tr>
<td>O/O</td>
<td>blood type O</td>
</tr>
</tbody>
</table>

- **Deterministic**

- However, not in reverse direction: Blood type A ⇒ A/A or A/O

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>H/H</td>
<td>healthy</td>
</tr>
<tr>
<td>H/D</td>
<td>healthy</td>
</tr>
<tr>
<td>D/D</td>
<td>ill with probability 0.9</td>
</tr>
</tbody>
</table>

- **Not deterministic**

- **Penetrance**: Pr(phenotype | genotype)
Recombination

Independent genes: recombination frequency $\theta = \frac{1}{2}$

Linked genes: $\theta < \frac{1}{2}$

Closer on chromosome $\approx$ more linked

Genetic linkage analysis: Estimate distances between, and hence locations of, genes on chromosome based on data containing recombination information.
Recombination

Independent genes: \( \theta = \frac{1}{2} \)

Linked genes: \( \theta < \frac{1}{2} \)

Closer on chromosome \( \approx \) more linked

Genetic linkage analysis: Estimate distances between, and hence locations of, genes on chromosome based on data containing recombination information.
Recombination

- Independent genes: recombination frequency $\theta = 1/2$
- Linked genes: $\theta < 1/2$
- Closer on chromosome $\approx$ more linked
- Genetic linkage analysis: Estimate distances between, and hence locations of, genes on chromosome based on data containing recombination information
Recombination Frequencies

Gene 1
Gene 2
Gene 3

$\theta_1$
$\theta_2$
Pedigree as Bayesian Network

- **Child: 3**
- **Father: 1**
  - $GP_{i1}/GM_{i1}$
  - $GP_{i2}/GM_{i2}$
  - $GP_{i3}/GM_{i3}$
- **Mother: 2**
  - Phenotype:
    - $P_{i1}$, $P_{i2}$, $P_{i3}$
  - Selectors:
    - $SP_{31}/SM_{31}$
    - $SP_{32}/SM_{32}$
    - $SP_{33}/SM_{33}$

$\triangleright$ 3 genes: $GP_{i1}/GM_{i1}$, $GP_{i2}/GM_{i2}$, $GP_{i3}/GM_{i3}$

$\triangleright$ Phenotype: $P_{i1}, P_{i2}, P_{i3}$

$\triangleright$ Selectors: $SP_{31}/SM_{31}$, $SP_{32}/SM_{32}$, $SP_{33}/SM_{33}$
Pedigree as Bayesian Network

Encoding recombination frequencies

|       |       | \[\theta_{sp_{32}|sp_{31}}\] |
|-------|-------|-----------------------------|
| \(SP_{31}\) | \(SP_{32}\) | \(1 - \theta_{12}\) |
| \(p\)     | \(p\)     | \(\theta_{12}\)            |
| \(p\)     | \(m\)     | \(\theta_{12}\)            |
| \(m\)     | \(p\)     | \(\theta_{12}\)            |
| \(m\)     | \(m\)     | \(1 - \theta_{12}\)        |

Priors on \(GP/GM\) of founders obtained from population stats
Given \( e \) on subset of genotype \((GP, GM)\) and phenotype \((P)\), \( Pr(e) \) indicates likelihood of \( \hat{\theta} \) (recombination frequencies)

Can compute preferred \( \hat{\theta} \), or search for most likely \( \hat{\theta} \)

Given \( \hat{\theta} \) and relative order of genes, determine their locations on chromosome
Planning: Find plan (sequence of actions) to go from initial to goal state

Probabilistic planning: In uncertain domains, find plan to reach goal with high probability
Slippery Griper

Goal: Block painted and held, gripper clean

Probabilistic initial state
- Block not painted, not held
- Gripper clean, but dry with probability 0.7

Probabilistic action effects
- Paint: Paints block w/p 1; makes gripper dirty w/p 1 if it holds block, w/p 0.1 if not
- Pick-up: Succeeds w/p 0.5 if gripper wet, 0.95 if gripper dry
- Dry: Dries wet gripper w/p 0.8; doesn’t affect dry gripper
Conformant Probabilistic Planning

- Probabilistic initial state and action effects

- Conformant: Action effects not observable (can’t decide next action by observing environment)

- Need straight-line plan with max probability of success, for given horizon (number of steps)

- Example 2-step plan: [paint, pickup] (succeeds with probability 0.7335)
Brute-force Approach

- Compute success probability for all plans of length one
- Given success probabilities of plans of length $i$, compute probability of success for plans of length $i + 1$
- Iterate to planning horizon $n$
- Pick $n$-step plan with max success probability
- Exponential in planning horizon $n$
Suppose action variable $A_i$ represents action taken at step $i$

Planning $\iff$ $\max \Pr(\text{goal}|A_1, \ldots, A_n)$
$\iff$ $\max \Pr(\text{goal}, A_1, \ldots, A_n)$
Planning Domain as Bayesian Network

- Suppose action variable $A_i$ represents action taken at step $i$

- Planning $\iff \max \Pr(\text{goal}|A_1, \ldots, A_n)$
  $\iff \max \Pr(\text{goal}, A_1, \ldots, A_n)$

- How to define Bayesian network?
Planning Domain as Bayesian Network

Paint: Paints block with probability 1

<table>
<thead>
<tr>
<th>Painted₀</th>
<th>Action₁</th>
<th>Painted₁</th>
<th>Pr(.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>paint</td>
<td>T</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>*</td>
<td>T</td>
<td>1</td>
</tr>
<tr>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
</tr>
</tbody>
</table>
Painting Domain as Bayesian Network

**Diagram:**

- **Painted**
- **Held**
- **Dry**
- **Clean**

**Action**

- **Painted**
- **Held**
- **Dry**
- **Clean**

**Table:**

<table>
<thead>
<tr>
<th>Clean₀</th>
<th>Held₀</th>
<th>Action₁</th>
<th>Clean₁</th>
<th>Pr(·)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>paint</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>paint</td>
<td>F</td>
<td>.1</td>
</tr>
<tr>
<td>F</td>
<td>*</td>
<td>*</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Text:**

Paint: Makes gripper dirty w/p 1 if it holds block, w/p 0.1 if not
Planning Domain as Bayesian Network

Add one layer per time step

Initial state → priors on $Painted_0, Held_0, Dry_0, Clean_0$
Goal state → evidence $e$ on $Painted_n, Held_n, Dry_n, Clean_n$
Optimal plan → $\text{argmax} \Pr(e, A_1, \ldots, A_n)$

Priors on $Action_i$ irrelevant
Plan Assessment

\[ \Pr(e, A_1, \ldots, A_n) \]
Plan Search

\[
\Pr\left(\text{t}^\prime, C, B\right) \leq 0.101
\]
Plan Search

\[
\max_{C, B} \Pr(e, t', C, B) \leq
\]

Diagram:

```
    T
   / \  
  C   C
 / \ / \  
B   B B   B
 / | / | / |
.tcb .tcb' .tc'b .tc'b' .t'cb .t'cb' .t'c'b .t'c'b' .009 .079 .117 .053 .023 .051 .031 .055
```

Inequality:

\[
\max_{C, B} \Pr(e, t', C, B) \leq
\]
Plan Search

\[
\max_{C,B} \Pr(e, t', C, B) \leq 0.101
\]
Upper Bounds

$$\max_{A_{k+1}, \ldots, A_n} \Pr(e, a_1, \ldots, a_k, A_{k+1}, \ldots, A_n)$$
Upper Bounds

\[
\max_{A_{k+1}, \ldots, A_n} \Pr(e, a_1, \ldots, a_k, A_{k+1}, \ldots, A_n)
\]

Some OR-nodes are decision nodes on \( A_{k+1}, \ldots, A_n \)

Turn these into max (instead of +)

Root of arithmetic circuit gives upper bound
Summary of the Course

- Bayesian networks as compact representations of probability distributions
- Inference by logical encoding and compilation to arithmetic circuits
- Inference by variable elimination
- Applications