Bayesian Networks and Decision-Theoretic Reasoning for Artificial Intelligence

Jack Breese  
Microsoft Research

Daphne Koller  
Stanford University
Overview

- Decision-theoretic techniques
  - Explicit management of uncertainty and tradeoffs
  - Probability theory
  - Maximization of expected utility

- Applications to AI problems
  - Diagnosis
  - Expert systems
  - Planning
  - Learning
Science- AAAI-97

- Model Minimization in Markov Decision Processes
- Effective Bayesian Inference for Stochastic Programs
- Learning Bayesian Networks from Incomplete Data
- Summarizing CSP Hardness With Continuous Probability Distributions
- Speeding Safely: Multi-criteria Optimization in Probabilistic Planning
- Structured Solution Methods for Non-Markovian Decision Processes
Applications

Microsoft’s cost-cutting helps users

04/21/97

A Microsoft Corp. strategy to cut its support costs by letting users solve their own problems using electronic means is paying off for users. In March, the company began rolling out a series of Troubleshooting Wizards on its World Wide Web site.

Troubleshooting Wizards save time and money for users who don’t have Windows NT specialists on hand at all times, said Paul Soares, vice president and general manager of Alden Buick Pontiac, a General Motors Corp. car dealership in Fairhaven, Mass.
Teenage Bayes

Microsoft Researchers Exchange Brainpower with Eighth-grader

Teenager Designs Award-Winning Science Project

.. For her science project, which she called "Dr. Sigmund Microchip," Tovar wanted to create a computer program to diagnose the probability of certain personality types. With only answers from a few questions, the program was able to accurately diagnose the correct personality type 90 percent of the time.

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Course Contents

» Concepts in Probability
  ◆ Probability
  ◆ Random variables
  ◆ Basic properties (Bayes rule)
■ Bayesian Networks
■ Inference
■ Decision making
■ Learning networks from data
■ Reasoning over time
■ Applications
Probabilities

- Probability distribution $P(X|\xi)$
  - $X$ is a random variable
    - Discrete
    - Continuous
  - $\xi$ is background state of information
Discrete Random Variables

- Finite set of possible outcomes

\[ X \ni \{x_1, x_2, x_3, \ldots, x_n\} \]

\[ P(x_i) \geq 0 \]

\[ \sum_{i=1}^{n} P(x_i) = 1 \]

X binary: \[ P(x) + P(\overline{x}) = 1 \]
Continuous Random Variable

- Probability distribution (density function) over continuous values

\[ X \in [0,10] \quad P(x) \geq 0 \]

\[
\int_{0}^{10} P(x) \, dx = 1
\]

\[
P(5 \leq x \leq 7) = \int_{5}^{7} P(x) \, dx
\]
More Probabilities

- **Conditional**
  \[ P(x \mid y) \equiv P(X = x \mid Y = y) \]
  - Probability that \( X=x \) given we know that \( Y=y \)

- **Joint**
  \[ P(x, y) \equiv P(X = x \land Y = y) \]
  - Probability that both \( X=x \) and \( Y=y \)
Rules of Probability

- **Product Rule**

\[ P(X, Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X) \]

- **Marginalization**

\[ P(Y) = \sum_{i=1}^{n} P(Y, x_i) \]

For binary $X$:

\[ P(Y) = P(Y, x) + P(Y, \bar{x}) \]
Bayes Rule

\[ P(H, E) = P(H \mid E)P(E) = P(E \mid H)P(H) \]

\[ P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} \]

\[ P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e, h) + P(e, \bar{h})} \]

\[ = \frac{P(e \mid h)P(h)}{P(e \mid h)P(h) + P(e \mid \bar{h})P(\bar{h})} \]
Course Contents

■ Concepts in Probability
  » Bayesian Networks
    ◆ Basics
    ◆ Additional structure
    ◆ Knowledge acquisition

■ Inference

■ Decision making

■ Learning networks from data

■ Reasoning over time

■ Applications
Bayesian networks

- Basics
  - Structured representation
  - Conditional independence
  - Naïve Bayes model
  - Independence facts
Bayesian Networks

\[ S \in \{no, light, heavy\} \rightarrow Smoking \rightarrow Cancer \]

\[
\begin{align*}
P(S=\text{no}) & : 0.80 \\
P(S=\text{light}) & : 0.15 \\
P(S=\text{heavy}) & : 0.05 \\
\end{align*}
\]

\[
\begin{align*}
P(C=\text{none}) & : 0.96 \\
P(C=\text{benign}) & : 0.03 \\
P(C=\text{malignant}) & : 0.01 \\
\end{align*}
\]

\[
\begin{align*}
P(C=\text{light}) & : 0.88 \\
P(C=\text{benign}) & : 0.08 \\
P(C=\text{malignant}) & : 0.04 \\
\end{align*}
\]

\[
\begin{align*}
P(C=\text{heavy}) & : 0.60 \\
P(C=\text{benign}) & : 0.25 \\
P(C=\text{malignant}) & : 0.15 \\
\end{align*}
\]
Product Rule

\[ P(C, S) = P(C|S) \cdot P(S) \]

<table>
<thead>
<tr>
<th>( S \downarrow )</th>
<th>( C \Rightarrow )</th>
<th>none</th>
<th>benign</th>
<th>malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.768</td>
<td>0.024</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>light</td>
<td>0.132</td>
<td>0.012</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>heavy</td>
<td>0.035</td>
<td>0.010</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>
## Marginalization

<table>
<thead>
<tr>
<th>$S$</th>
<th>$C$</th>
<th>none</th>
<th>benign</th>
<th>malig</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>none</td>
<td>0.768</td>
<td>0.024</td>
<td>0.008</td>
<td>.80</td>
</tr>
<tr>
<td>light</td>
<td>light</td>
<td>0.132</td>
<td>0.012</td>
<td>0.006</td>
<td>.15</td>
</tr>
<tr>
<td>heavy</td>
<td>heavy</td>
<td>0.035</td>
<td>0.010</td>
<td>0.005</td>
<td>.05</td>
</tr>
<tr>
<td>total</td>
<td>total</td>
<td>0.935</td>
<td>0.046</td>
<td>0.019</td>
<td></td>
</tr>
</tbody>
</table>

$P(Smoke)$

$P(Cancer)$
Bayes Rule Revisited

\[
P(S \mid C) = \frac{P(C \mid S)P(S)}{P(C)} = \frac{P(C, S)}{P(C)}
\]

<table>
<thead>
<tr>
<th>$S \downarrow$</th>
<th>$C \mapsto$</th>
<th>none</th>
<th>benign</th>
<th>malign</th>
</tr>
</thead>
<tbody>
<tr>
<td>$no$</td>
<td>0.768/.935</td>
<td>0.024/.046</td>
<td>0.008/.019</td>
<td></td>
</tr>
<tr>
<td>$light$</td>
<td>0.132/.935</td>
<td>0.012/.046</td>
<td>0.006/.019</td>
<td></td>
</tr>
<tr>
<td>$heavy$</td>
<td>0.030/.935</td>
<td>0.015/.046</td>
<td>0.005/.019</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Cancer=$</th>
<th>none</th>
<th>benign</th>
<th>malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(S=\text{no})$</td>
<td>0.821</td>
<td>0.522</td>
<td>0.421</td>
</tr>
<tr>
<td>$P(S=\text{light})$</td>
<td>0.141</td>
<td>0.261</td>
<td>0.316</td>
</tr>
<tr>
<td>$P(S=\text{heavy})$</td>
<td>0.037</td>
<td>0.217</td>
<td>0.263</td>
</tr>
</tbody>
</table>
A Bayesian Network

- Age
- Gender
- Exposure to Toxics
- Smoking
- Cancer
- Serum Calcium
- Lung Tumor
Independence

Age and Gender are independent.

\[ P(A,G) = P(G)P(A) \]

\[ P(A|G) = P(A) \quad A \perp G \]

\[ P(G|A) = P(G) \quad G \perp A \]

\[ P(A,G) = P(G|A) P(A) = P(G)P(A) \]

\[ P(A,G) = P(A|G) P(G) = P(A)P(G) \]
Conditional Independence

Cancer is independent of Age and Gender given Smoking.

\[ P(C|A,G,S) = P(C|S) \quad C \perp A,G \mid S \]
More Conditional Independence: Naïve Bayes

Serum Calcium and Lung Tumor are dependent

Serum Calcium is independent of Lung Tumor, given Cancer

\[ P(L|SC,C) = P(L|C) \]
Naïve Bayes in general

2n + 1 parameters:

\[ P(h), P(e_i \mid h), P(e_i \mid \overline{h}), \quad i = 1, \ldots, n \]
More Conditional Independence: Explaining Away

Exposure to Toxics and Smoking are independent

\[ E \perp S \]

Exposure to Toxics is dependent on Smoking, given Cancer

\[ P(E = \text{heavy} \mid C = \text{malignant}) > P(E = \text{heavy} \mid C = \text{malignant}, S=\text{heavy}) \]
Put it all together

\[ P(A, G, E, S, C, L, SC) = \]

\[ P(A) \cdot P(G) \cdot \]

\[ P(E \mid A) \cdot P(S \mid A, G) \cdot \]

\[ P(C \mid E, S) \cdot \]

\[ P(SC \mid C) \cdot P(L \mid C) \]
General Product (Chain) Rule for Bayesian Networks

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

\( \text{Pa}_i = \text{parents}(X_i) \)
Conditional Independence

A variable (node) is conditionally independent of its non-descendants given its parents.

- **Age**
- **Gender**
- **Exposure to Toxics**
- **Smoking**
- **Cancer**
- **Serum Calcium**
- **Lung Tumor**

Cancer is independent of Age and Gender given Exposure to Toxics and Smoking.
Cancer is independent of Genetic Damage given Exposure to Toxics and Smoking.

Another non-descendant

- Age
- Gender
- Exposure to Toxics
- Smoking
- Cancer
- Genetic Damage
- Serum Calcium
- Lung Tumor

Parents: Smoking, Gender
Non-Descendants: Cancer, Genetic Damage
Independence and Graph Separation

- Given a set of observations, is one set of variables dependent on another set?
- Observing effects can induce dependencies.
- d-separation (Pearl 1988) allows us to check conditional independence graphically.
Bayesian networks

- Additional structure
  - Nodes as functions
  - Causal independence
  - Context specific dependencies
  - Continuous variables
  - Hierarchy and model construction
Nodes as functions

- A BN node is a conditional distribution function
  - its parent values are the inputs
  - its output is a distribution over its values
Any type of function from $Val(A,B)$ to distributions over $Val(X)$

$lo : 0.7$
$med : 0.1$
$hi : 0.2$
Causal Independence

- **Burglary** causes **Alarm** iff motion sensor clear
- **Earthquake** causes **Alarm** iff wire loose
- Enabling factors are independent of each other
Fine-grained model

- **Burglary**
  - $m$
  - $\overline{m}$
  - $b$
  - $\overline{b}$
  - $1-r_B$
  - $r_B$
  - 0
  - 1

- **Earthquake**
  - $w$
  - $\overline{w}$
  - $e$
  - $\overline{e}$
  - $1-r_E$
  - $r_E$
  - 0
  - 1

- **Motion sensed**
- **Wire Move**
- **Deterministic or**
- **Alarm**
Noisy-Or model

Alarm false only if all mechanisms independently inhibited

\[ P(a) = 1 - \prod_{\text{parent } X \text{ active}} r_X \]

# of parameters is linear in the # of parents
CPCS Network
Context-specific Dependencies

- *Alarm* can go off only if it is *Set*
- A burglar and the cat can both set off the alarm
- If a burglar comes in, the cat hides and does not set off the alarm
Asymmetric dependencies

Node function represented as a tree

- **Alarm** independent of
  - **Burglary, Cat** given $\overline{s}$
  - **Cat** given $s$ and $b$

### Diagram

- **Alarm-Set**
- **Burglary**
- **Cat**

```
A            S    C
\( (a: 0, \overline{a}: 1) \) \( \overline{s} \) \( s \) \( \overline{c} \) \( c \) \( (a: 0.9, \overline{a}: 0.1) \)
\( (a: 0.01, \overline{a}: 0.99) \)

\( (a: 0.6, \overline{a}: 0.4) \)
```

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Asymmetric Assessment

Asymmetric Assessment of Printer Output

Assessment Hierarchy:

1. Print Data
   1. Yes
      1. Printer Location
         1. Local
            1. LOCAL Transport
               1. Yes
                  1. Local Printer OK
                     1. Normal
                     1. Abnormal
               1. No
                  1. Network
                     1. NET Transport
                        1. Yes
                           1. Net Printer OK
                              1. Normal
                              1. Abnormal
                        1. No
                           1. Abnormal
      1. No
         1. Abnormal

- Net OK
- Local OK
- Net Transport
- Local Transport
- Location
- Printer Output
Continuous variables

Outdoor Temperature

97°

A/C Setting

hi

Indoor Temperature

Function from \( Val(A,B) \)
to density functions
over \( Val(X) \)

\[ P(x) \]

\[ X \]
Gaussian (normal) distributions

\[ P(x) = \frac{1}{\sqrt{2\pi \sigma}} \exp\left(\frac{-(x-\mu)^2}{2\sigma}\right) \]

\[ N(\mu, \sigma) \]

different mean

different variance
Gaussian networks

\[ X \sim N(\mu, \sigma^2_X) \]
\[ Y \sim N(ax + b, \sigma^2_Y) \]

Each variable is a linear function of its parents, with Gaussian noise

Joint probability density functions:
Composing functions

- Recall: a BN node is a function
- We can compose functions to get more complex functions.
- The result: A hierarchically structured BN.
- Since functions can be called more than once, we can reuse a BN model fragment in multiple contexts.
Bayesian Networks

- Knowledge acquisition
  - Variables
  - Structure
  - Numbers
What is a variable?

- Collectively exhaustive, mutually exclusive values

\[ x_1 \lor x_2 \lor x_3 \lor x_4 \]

\[ \neg(x_i \land x_j) \quad i \neq j \]

- Values versus Probabilities

Risk of Smoking

Smoking

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Clarity Test: Knowable in Principle

- Weather  \{Sunny, Cloudy, Rain, Snow\}
- Gasoline: Cents per gallon
- Temperature  \{ \geq 100F, < 100F\}
- User needs help on Excel Charting  \{Yes, No\}
- User’s personality  \{dominant, submissive\}
Structuring

Network structure corresponding to “causality” is usually good.

Extending the conversation.
Do the numbers really matter?

- Second decimal usually does not matter
- Relative Probabilities

- Zeros and Ones
- Order of Magnitude: $10^{-9}$ vs $10^{-6}$
- Sensitivity Analysis
Bayesian Networks and Structure

- Causal independence: from $2^n$ to $n+1$ parameters
- Asymmetric assessment: similar savings in practice.
- Typical savings (#params):
  - 145 to 55 for a small hardware network;
  - 133,931,430 to 8254 for CPCS !!
Course Contents

- Concepts in Probability
- Bayesian Networks
  » Inference
- Decision making
- Learning networks from data
- Reasoning over time
- Applications
Inference

- Patterns of reasoning
- Basic inference
- Exact inference
- Exploiting structure
- Approximate inference
Predictive Inference

How likely are elderly males to get malignant cancers?

\[ P(C=\text{malignant} \mid \text{Age}>60, \text{Gender}=\text{male}) \]
How likely is an elderly male patient with high Serum Calcium to have malignant cancer?

\[ P(C=\text{malignant} \mid \text{Age}>60, \text{Gender}=\text{male}, \text{Serum Calcium} = \text{high}) \]
Explaining away

- If we see a lung tumor, the probability of heavy smoking and of exposure to toxics both go up.
- If we then observe heavy smoking, the probability of exposure to toxics goes back down.
Inference in Belief Networks

- Find $P(Q=q|E=e)$
  - $Q$  the query variable
  - $E$ set of evidence variables

$$P(q | e) = \frac{P(q, e)}{P(e)}$$

$X_1,..., X_n$ are network variables except $Q, E$

$$P(q, e) = \sum_{x_1,..., x_n} P(q, e, x_1,..., x_n)$$
Basic Inference

\[ P(b) = ? \]
Product Rule

\[ P(C, S) = P(C|S) \cdot P(S) \]

<table>
<thead>
<tr>
<th>$S \downarrow$</th>
<th>$C \Rightarrow$</th>
<th>none</th>
<th>benign</th>
<th>malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.768</td>
<td>0.024</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>light</td>
<td>0.132</td>
<td>0.012</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>heavy</td>
<td>0.035</td>
<td>0.010</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>
Marginalization

<table>
<thead>
<tr>
<th>S</th>
<th>none</th>
<th>benign</th>
<th>malig</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.768</td>
<td>0.024</td>
<td>0.008</td>
<td>0.80</td>
</tr>
<tr>
<td>light</td>
<td>0.132</td>
<td>0.012</td>
<td>0.006</td>
<td>0.15</td>
</tr>
<tr>
<td>heavy</td>
<td>0.035</td>
<td>0.010</td>
<td>0.005</td>
<td>0.05</td>
</tr>
<tr>
<td>total</td>
<td>0.935</td>
<td>0.046</td>
<td>0.019</td>
<td></td>
</tr>
</tbody>
</table>

\[ P(Cancer) \]

\[ P(Smoke) \]
Basic Inference

\[ P(b) = \sum_a P(a, b) = \sum_a P(b \mid a) P(a) \]

\[ P(c) = \sum_b P(c \mid b) P(b) \]

\[ P(c) = \sum_{b,a} P(a, b, c) = \sum_{b,a} P(c \mid b) P(b \mid a) P(a) \]

\[ = \sum_b P(c \mid b) \sum_a P(b \mid a) P(a) P(b) \]
Inference in trees

\[ P(x) = \sum_{y_1, y_2} P(x \mid y_1, y_2) P(y_1, y_2) \]

because of independence of \( Y_1, Y_2 \):

\[ = \sum_{y_1, y_2} P(x \mid y_1, y_2) P(y_1) P(y_2) \]
Polynomials

- A network is *singly connected* (a polytree) if it contains no undirected loops.

**Theorem:** Inference in a singly connected network can be done in linear time*.

Main idea: in variable elimination, need only maintain distributions over single nodes.

* in network size including table sizes.
The problem with loops

The grass is dry only if no rain and no sprinklers.

\[ P(\overline{g}) = P(\overline{r}, \overline{s}) \sim 0 \]
The problem with loops contd.

\[
P(\bar{g}) = \underbrace{P(\bar{g} \mid r, s)}_{0} P(r, s) + \underbrace{P(\bar{g} \mid r, \bar{s})}_{0} P(r, \bar{s})
+ \underbrace{P(\bar{g} \mid \bar{r}, s)}_{0} P(\bar{r}, s) + \underbrace{P(\bar{g} \mid \bar{r}, \bar{s})}_{1} P(\bar{r}, \bar{s})
\]

\[
= P(\bar{r}, \bar{s}) \sim 0
\]

\[
\neq P(\bar{r}) P(\bar{s}) \sim 0.5 \cdot 0.5 = 0.25
\]
Variable elimination

\[ P(c) = \sum_b P(c \mid b) \sum_a P(b \mid a) P(a) \]
Inference as variable elimination

- A factor over $X$ is a function from $\text{val}(X)$ to numbers in $[0,1]$:  
  - A CPT is a factor  
  - A joint distribution is also a factor
- BN inference:  
  - factors are multiplied to give new ones  
  - variables in factors summed out
- A variable can be summed out as soon as all factors mentioning it have been multiplied.
Variable Elimination with loops

Complexity is exponential in the size of the factors
Join trees*

A join tree is a partially precompiled factorization

* aka junction trees, Lauritzen-Spiegelhalter, Hugin alg., …
Exploiting Structure

Idea: explicitly decompose nodes

Noisy or:

- Burglary
- Earthquake

Deterministic or

- Motion sensed
- Wire Move
- Alarm
Noisy-or decomposition

Smaller families
⇒ Smaller factors
⇒ Faster inference
Inference with continuous variables

- Gaussian networks: polynomial time inference regardless of network structure
- Conditional Gaussians:
  - discrete variables cannot depend on continuous

\[ S \sim N(a_F w + b_F, \sigma^2_F) \]

- These techniques do not work for general hybrid networks.
Computational complexity

- **Theorem:** Inference in a multi-connected Bayesian network is NP-hard.

Boolean 3CNF formula $\phi = (u \lor \overline{v} \lor w) \land (\overline{u} \lor \overline{w} \lor y)$

$$\text{Probability } (\cdot) = \frac{1}{2^n} \cdot \# \text{ satisfying assignments of } \phi$$

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Stochastic simulation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>e</td>
<td>a</td>
<td>c</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>e</td>
<td>a</td>
<td>c</td>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>

\[
P(b) = 0.03
\]

**Burglary**

\[
P(a) = \begin{array}{cccc}
0.98 & 0.7 & 0.4 & 0.01 \\
\end{array}
\]

**Earthquake**

\[
P(e) = 0.001
\]

**Alarm**

\[
P(c) = \begin{array}{cc}
a & \bar{a} \\
0.8 & 0.05
\end{array}
\]

**Call**

\[
P(n) = \begin{array}{cc}
e & \bar{e} \\
0.3 & 0.001
\end{array}
\]

\[
P(b|c) \sim \frac{\text{# of live samples with } B=b}{\text{total # of live samples}}
\]

Samples:

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Likelihood weighting

\[
P(b|c) = \frac{\text{weight of samples with } B=b}{\text{total weight of samples}}
\]

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\neg b)</td>
<td>e</td>
<td>a</td>
<td>c</td>
<td>(\neg n)</td>
<td>0.8</td>
</tr>
<tr>
<td>b</td>
<td>(\neg e)</td>
<td>a</td>
<td>(\neg c)</td>
<td>n</td>
<td>0.95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

P(c) 0.8 0.05
Other approaches

- Search based techniques
  - search for high-probability instantiations
  - use instantiations to approximate probabilities
- Structural approximation
  - simplify network
    - eliminate edges, nodes
    - abstract node values
    - simplify CPTs
  - do inference in simplified network
CPCS
Network
Course Contents

- Concepts in Probability
- Bayesian Networks
- Inference
  » Decision making
- Learning networks from data
- Reasoning over time
- Applications
Decision making

- Decisions, Preferences, and Utility functions
- Influence diagrams
- Value of information
Decision making

- Decision - an irrevocable allocation of domain resources
- Decision should be made so as to maximize expected utility.
- View decision making in terms of
  - Beliefs/Uncertainties
  - Alternatives/Decisions
  - Objectives/Utilities
A Decision Problem

Should I have my party inside or outside?

- **in**
  - **dry** → **Regret**
  - **wet** → **Relieved**

- **out**
  - **dry** → **Perfect!**
  - **wet** → **Disaster**
Value Function

- A numerical score over all possible states of the world.

<table>
<thead>
<tr>
<th>Location?</th>
<th>Weather?</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>in</td>
<td>dry</td>
<td>$50</td>
</tr>
<tr>
<td>in</td>
<td>wet</td>
<td>$60</td>
</tr>
<tr>
<td>out</td>
<td>dry</td>
<td>$100</td>
</tr>
<tr>
<td>out</td>
<td>wet</td>
<td>$0</td>
</tr>
</tbody>
</table>
Preference for Lotteries

\[ \approx \]

\begin{align*}
0.2 &\quad $40,000 \\
0.8 &\quad $0
\end{align*}

\[ \approx \]

\begin{align*}
0.25 &\quad $30,000 \\
0.75 &\quad $0
\end{align*}
Desired Properties for Preferences over Lotteries

If you prefer $100 to $0 and \( p < q \) then

\[
\begin{array}{c}
p \\ \text{\$100}
\end{array}
\quad \sim \quad
\begin{array}{c}
q \\ \text{\$100}
\end{array}
\begin{array}{c}
1-p \\ \text{\$0}
\end{array}
\quad \begin{array}{c}
1-q \\ \text{\$0}
\end{array}
\]

(always)
Expected Utility

Properties of preference ⇒ existence of function $U$, that satisfies:

\[ \sum_i p_i U(x_i) < \sum_i q_i U(y_i) \]
Some properties of $U$

$0.8 \rightarrow \frac{0.2}{G102} \approx \frac{1}{G112} \uparrow \Rightarrow U \neq \text{monetary payoff}$
Attitudes towards risk

- **$U(l)$**: Utility function for different $l$ values.
- **$U($500)$**: Certain equivalent or insurance/risk premium.
- **$U$** vs. **$\$ reward$** graph:
  - **Risk Averse** (Convex $U$): Prefers certainty over uncertainty.
  - **Risk Neutral** (Linear $U$): No preference for certainty vs. uncertainty.

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Are people rational?

\[ 0.2 \cdot U(\$40k) > 0.25 \cdot U(\$30k) \]

\[ 0.8 \cdot U(\$40k) > U(\$30k) \]

\[ 0.8 \cdot U(\$40k) < U(\$30k) \]
Maximizing Expected Utility

Choose the action that maximizes expected utility

\[ EU(\text{in}) = 0.7 \cdot U(\$50) + 0.3 \cdot U(\$60) = 0.7 \cdot 0.632 + 0.3 \cdot 0.699 = 0.652 \]

\[ EU(\text{out}) = 0.7 \cdot U(\$100) + 0.3 \cdot U(\$0) = 0.7 \cdot 0.865 + 0.3 \cdot 0 = 0.605 \]

Choose in
Multi-attribute utilities
(or: Money isn’t everything)

- Many aspects of an outcome combine to determine our preferences.
  - vacation planning: cost, flying time, beach quality, food quality, …
  - medical decision making: risk of death (micromort), quality of life (QALY), cost of treatment, …

- For rational decision making, must combine all relevant factors into single utility function.
Influence Diagrams

Earthquake
  Newcast

Burglary
  Alarm

Call
  Goods
    Recovered
  Go Home?
    Miss Meeting
    Big Sale

Utility

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Decision Making with Influence Diagrams

- Earthquake
  - Newcast
- Burglary
- Alarm
- Call

- Go Home?
  - Neighbor Phoned Yes
  - No Phone Call No

- Goods Recovered
  - Miss Meeting
  - Big Sale
  - Utility

Expected Utility of this policy is 100
Value-of-Information

- What is it worth to get another piece of information?
- What is the increase in (maximized) expected utility if I make a decision with an additional piece of information?
- Additional information (if free) cannot make you worse off.
- There is no value-of-information if you will not change your decision.
Value-of-Information in an Influence Diagram

How much better can we do when this arc is here?

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Value-of-Information is the increase in Expected Utility

<table>
<thead>
<tr>
<th>Phonecall?</th>
<th>Newscast?</th>
<th>Go Home?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Quake</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No Quake</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>Quake</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>No Quake</td>
<td>No</td>
</tr>
</tbody>
</table>

Expected Utility of this policy is 112.5
Course Contents

- Concepts in Probability
- Bayesian Networks
- Inference
- Decision making
  - Learning networks from data
- Reasoning over time
- Applications
Learning networks from data

- The learning task
- Parameter learning
  - Fully observable
  - Partially observable
- Structure learning
- Hidden variables
The learning task

Input: training data
Output: BN modeling data

- Input: fully or partially observable data cases?
- Output: parameters or also structure?
Parameter learning: one variable

- Unfamiliar coin:
  - Let $\theta =$ bias of coin (long-run fraction of heads)

- If $\theta$ known (given), then
  - $P(X = heads \mid \theta) = \theta$

- Different coin tosses independent given $\theta$
  - $P(X_1, \ldots, X_n \mid \theta) = \theta^h (1-\theta)^t$
    - $h$ heads, $t$ tails
Maximum likelihood

- Input: a set of previous coin tosses
  - $X_1, \ldots, X_n = \{H, T, H, H, H, T, T, H, \ldots, H\}$
    - $h$ heads, $t$ tails
- Goal: estimate $\theta$
- The likelihood $P(X_1, \ldots, X_n \mid \theta) = \theta^h (1-\theta)^t$
- The maximum likelihood solution is:
  $$\theta^* = \frac{h}{h+t}$$
Bayesian approach

Uncertainty about $\theta \Rightarrow$ distribution over its values

$$P(X = \text{heads}) = \int_{0}^{\infty} P(X = \text{heads} \mid \theta) P(\theta) \, d\theta = \int_{0}^{\infty} \theta \, P(\theta) \, d\theta$$
Conditioning on data

\[ P(\theta) \overset{D}{\longrightarrow} P(\theta | D) \propto P(\theta) P(D | \theta) = P(\theta) \theta^h (1-\theta)^t \]

\( h \) heads, \( t \) tails
Good parameter distribution:

\[ \text{Beta}(\alpha_h, \alpha_t) \propto \theta^{\alpha_h - 1} (1 - \theta)^{\alpha_t - 1} \]

*Dirichlet distribution generalizes Beta to non-binary variables.*

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
General parameter learning

- A multi-variable BN is composed of several independent parameters ("coins").

![Diagram](A → B)

Three parameters:

\[ \theta_A, \theta_B|a, \theta_B|\overline{a} \]

- Can use same techniques as one-variable case to learn each one separately

Max likelihood estimate of \( \theta_B|\overline{a} \) would be:

\[ \theta^*_B|\overline{a} = \frac{\text{#data cases with } b, \overline{a}}{\text{#data cases with } \overline{a}} \]
### Partially observable data

**Diagram:**
- **Variables:** `B`, `E`, `A`, `C`, `N`, `b`, `a`, `c`, `n`.
- **Events:** Burglary, Earthquake, Alarm, Call, Newscast.

**Table:**

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
</table>

**Fill in missing data with “expected” value**

- **Expected** = distribution over possible values
- **Use** “best guess” BN to estimate distribution

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Intuition

- In fully observable case:

\[ \theta^*_{n|e} = \frac{\text{#data cases with } n, e}{\text{#data cases with } e} = \frac{\sum_j I(n,e \mid d_j)}{\sum_j I(e \mid d_j)} \]

\[ I(e \mid d_j) = \begin{cases} 1 & \text{if } E=e \text{ in data case } d_j \\ 0 & \text{otherwise} \end{cases} \]

- In partially observable case \( I \) is unknown.

Best estimate for \( I \) is:

\[ \hat{I}(n, e \mid d_j) = P_{\theta^*}(n, e \mid d_j) \]

Problem: \( \theta^* \) unknown.
Expectation Maximization (EM)

Repeat :

- **Expectation (E) step**
  - Use current parameters $\theta$ to estimate filled in data.

$$\hat{I}(n, e \mid d_j) = P_\theta (n, e \mid d_j)$$

- **Maximization (M) step**
  - Use filled in data to do max likelihood estimation

$$\tilde{\theta}_{n \mid e} = \frac{\sum_j \hat{I}(n, e \mid d_j)}{\sum_j \hat{I}(e \mid d_j)}$$

- Set: $\theta := \tilde{\theta}$

until convergence.
Structure learning

Goal:
find “good” BN structure (relative to data)

Solution:
do heuristic search over space of network structures.
Search space

Space = network structures
Operators = add/reverse/delete edges
Heuristic search

Use scoring function to do heuristic search (any algorithm). Greedy hill-climbing with randomness works pretty well.
Scoring

- Fill in parameters using previous techniques & score completed networks.
- One possibility for score:
  likelihood function: $\text{Score}(B) = P(\text{data} \mid B)$

  Example: $X, Y$ independent coin tosses
typical $\text{data} = (27 \ h-h, \ 22 \ h-t, \ 25 \ t-h, \ 26 \ t-t)$

  Maximum likelihood network structure:

  ![Diagram of network with directed edge from X to Y]

  Max. likelihood network typically fully connected

  *This is not surprising: maximum likelihood always overfits…*
Better scoring functions

- MDL formulation: balance fit to data and model complexity (# of parameters)
  \[ \text{Score}(B) = P(\text{data} \mid B) - \text{model complexity} \]

- Full Bayesian formulation
  - prior on network structures & parameters
  - more parameters ⇒ higher dimensional space
  - get balance effect as a byproduct*

* with Dirichlet parameter prior, MDL is an approximation to full Bayesian score.
Hidden variables

■ There may be interesting variables that we never get to observe:
  ◆ topic of a document in information retrieval;
  ◆ user’s current task in online help system.

■ Our learning algorithm should
  ◆ hypothesize the existence of such variables;
  ◆ learn an appropriate state space for them.
Randomly scattered data
Actual data
Bayesian clustering (Autoclass)

- (hypothetical) class variable never observed
- if we know that there are $k$ classes, just run EM
- learned classes = clusters
- Bayesian analysis allows us to choose $k$, trade off fit to data with model complexity

naïve Bayes model:

```
Class
    /\      /\     /\...
   /  \    /  \    /  \...
E_1  E_2  ...  E_n
```
Clustered distributions
Detecting hidden variables

- Unexpected correlations $\Rightarrow$ hidden variables.

Hypothesized model

```
Cholesterolemia
  └── Test1 ─── Test2 ─── Test3
```

Data model

```
Cholesterolemia
  └── Test1 ─── Test2 ─── Test3
```

“Correct” model

```
Cholesterolemia
  └── Test1 ─── Test2 ─── Test3

Hypothyroid
```

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Course Contents

- Concepts in Probability
- Bayesian Networks
- Inference
- Decision making
- Learning networks from data
  » Reasoning over time
- Applications
Reasoning over time

- Dynamic Bayesian networks
- Hidden Markov models
- Decision-theoretic planning
  - Markov decision problems
  - Structured representation of actions
  - The qualification problem & the frame problem
  - Causality (and the frame problem revisited)
Dynamic environments

- Markov property:
  - past independent of future given current state;
  - a conditional independence assumption;
  - implied by fact that there are no arcs $t \rightarrow t+2$. 

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Dynamic Bayesian networks

- State described via random variables.
- Each variable depends only on few others.
An HMM is a simple model for a partially observable stochastic domain.
Hidden Markov models (HMMs)

Partially observable stochastic environment:

- Mobile robots:
  - states = location
  - observations = sensor input

- Speech recognition:
  - states = phonemes
  - observations = acoustic signal

- Biological sequencing:
  - states = protein structure
  - observations = amino acids
HMMs and DBNs

- HMMs are just very simple DBNs.
- Standard inference & learning algorithms for HMMs are instances of DBN algorithms
  - Forward-backward = polytree
  - Baum-Welch = EM
  - Viterbi = most probable explanation.
Acting under uncertainty

Markov Decision Problem (MDP)

- Overall utility = sum of momentary rewards.
- Allows rich preference model, e.g.:

  rewards corresponding to “get to goal asap” = \[
  \begin{cases}
  +100 & \text{goal states} \\
  -1 & \text{other states}
  \end{cases}
  \]
The optimal action at time $t$ depends on the entire history of previous observations.

Instead, a distribution over $State(t)$ suffices.
Structured representation

Preconditions

Move:

Turn:

Probabilistic action model

- allows for exceptions & qualifications;
- persistence arcs: a solution to the frame problem.
Causality

- Modeling the effects of interventions
- Observing vs. “setting” a variable
- A form of persistence modeling
Cold temperatures can cause the distributor cap to become cracked.

If the distributor cap is cracked, then the car is less likely to start.
Setting vs. Observing

The car does not start. Will it start if we replace the distributor?
Predicting the effects of interventions

The car does not start. Will it start if we replace the distributor cap?

What is the probability that the car will start if I replace the distributor cap?
Mechanism Nodes

<table>
<thead>
<tr>
<th>$M_{\text{start}}$</th>
<th>Distributor</th>
<th>Starts?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Starts</td>
<td>Cracked</td>
<td>Yes</td>
</tr>
<tr>
<td>Always Starts</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>Never Starts</td>
<td>Cracked</td>
<td>No</td>
</tr>
<tr>
<td>Never Starts</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>Normal</td>
<td>Cracked</td>
<td>No</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>Inverse</td>
<td>Cracked</td>
<td>Yes</td>
</tr>
<tr>
<td>Inverse</td>
<td>Normal</td>
<td>No</td>
</tr>
</tbody>
</table>
Assumption: The mechanism relating Dist to Start is unchanged by replacing the Distributor.
Course Contents

- Concepts in Probability
- Bayesian Networks
- Inference
- Decision making
- Learning networks from data
- Reasoning over time
  » Applications
Applications

- Medical expert systems
  - Pathfinder
  - Parenting MSN
- Fault diagnosis
  - Ricoh FIXIT
  - Decision-theoretic troubleshooting
- Vista
- Collaborative filtering
Why use Bayesian Networks?

- Explicit management of uncertainty/tradeoffs
- Modularity implies maintainability
- Better, flexible, and robust recommendation strategies
Pathfinder

- Pathfinder is one of the first BN systems.
- It performs diagnosis of lymph-node diseases.
- It deals with over 60 diseases and 100 findings.
- Commercialized by Intellipath and Chapman Hall publishing and applied to about 20 tissue types.
Studies of Pathfinder Diagnostic Performance

- Naïve Bayes performed considerably better than certainty factors and Dempster-Shafer Belief Functions.
- Incorrect zero probabilities caused 10% of cases to be misdiagnosed.
- Full Bayesian network model with feature dependencies did best.
Commercial system: Integration

- Expert System with advanced diagnostic capabilities
  - uses key features to form the differential diagnosis
  - recommends additional features to narrow the differential diagnosis
  - recommends features needed to confirm the diagnosis
  - explains correct and incorrect decisions
- Video atlases and text organized by organ system
- “Carousel Mode” to build customized lectures
- Anatomic Pathology Information System
On Parenting: Selecting problem

- Diagnostic indexing for Home Health site on Microsoft Network
- Enter symptoms for pediatric complaints
- Recommends multimedia content
On Parenting : MSN

Original Multiple Fault Model
Single Fault approximation
On Parenting: Selecting problem

Describe the child

This feature is designed to help you find information relevant to questions you answer about childhood symptoms. Keep in mind that any information you find is in no way comprehensive.

Also, this feature is NOT intended to be used to diagnose medical conditions or replace the advice of a healthcare professional. Always contact your healthcare provider for medical advice.

Age: Toddler
Sex: Female

Complaint:
- Abdominal pain
- Abnormal control of body movements
- Biting or hitting
- Blood in stool
- Blood in urine
- Blood in vomit
- Bluish or purplish skin
- Breath-holding
- Breathlessness or difficulty breathing
- Colic or gas pain
- Constipation
- Cough
- Delayed development
- Delayed speech
- Diarrhea
- Difficulty swallowing
### Performing diagnosis/indexing

**Describe the child**

in the drop-down boxes at the right. Relevant information will appear below.

<table>
<thead>
<tr>
<th>Age: Toddler</th>
<th>Sex: Female</th>
</tr>
</thead>
</table>

**Complaint:** Abdominal pain

---

Localized pain: Can the child localize, or point to, the site of the pain?

- [ ] No, unable to localize
- [ ] Below the navel to the child's left
- [ ] Above the child's navel
- [ ] Either of the child's sides
- [ ] Below the navel to the child's right
- [ ] Above the navel to the child's right
- [ ] Above the navel to the child's left
- [ ] Don't Know

---

**Results so far**

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral gastroenteritis</td>
<td>🔴</td>
</tr>
<tr>
<td>Psychosomatic pain</td>
<td>🔴</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>
RICOH Fixit

- Diagnostics and information retrieval
FIXIT: Ricoh copy machine
Online Troubleshooters

Microsoft now offers advanced inference engine technology to help you easily troubleshoot problems with Microsoft products. These Troubleshooting Wizards are the electronic version of our best engineers. Try them and see!

Choose a Troubleshooting Wizard from the list below, then click Next.

- **Access**: Conversion Troubleshooter
- **Access**: Help exporting to Active Server Pages and viewing them in your web browser.
- **DirectX**: Help with common issues when you run applications with DirectX
- **Excel**: PivotTable Troubleshooter
- **Excel**: Workbook Troubleshooter
- **Excel**: Video Troubleshooter
- **Exchange**: Directory Synchronization Troubleshooter
- **Exchange**: Internet Mail Connector Troubleshooter
- **Games**: Help with problems when you install or start your Games
- **Games**: Help with games display problems
- **Games**: Help with sound problems when you run your games
- **Games**: Help with common issues when you run your games
- **Internet Explorer 3.0**: Help browsing the web
Define Problem

Troubleshooting Wizards

Print Troubleshooter

The Print Troubleshooter lists recommended troubleshooting steps in the order of greatest benefit and least cost to you (the user).

What type of problem are you having?

- My document didn't print at all.
- Graphics look incomplete or incorrect.
- Fonts are missing or do not look as they did on the screen.
- The printout is garbled or contains garbage.
- I only got part of the page I expected.
- Printing is unusually slow.

Next
Gather Information

Troubleshooting Wizards

Print Troubleshooter

This table tracks your status in the troubleshooting process. If you need to change your response to a question, you can do so below:

<table>
<thead>
<tr>
<th>Problem:</th>
<th>Print Output</th>
</tr>
</thead>
</table>

Are you printing from an MS-DOS-based or a Windows-based application?

- I am printing from MS-DOS or from an MS-DOS application.
- I am printing from a Windows application.
- I don't want to do this now.

Next
Get Recommendations

Print Troubleshooter

This table tracks your status in the troubleshooting process. If you need to change your answer to a question, you can do so below:

<table>
<thead>
<tr>
<th>Problem:</th>
<th>Print Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Environment:</td>
<td>○ MS-DOS ○ Windows ○ Unknown</td>
</tr>
<tr>
<td>Printing over Network:</td>
<td>○ No (Local printer) ○ Yes (Network printer) ○ Unknown</td>
</tr>
<tr>
<td>Printer Driver Set Offline:</td>
<td>○ Online ○ Unknown</td>
</tr>
</tbody>
</table>

Is your printer turned on and on-line?

1. Make sure the printer is properly plugged into a power outlet.
2. Turn on the printer's power switch.
3. Make sure the printer is **on line**. Most printers have an On Line button with a light.
4. Make sure the light is on.

If you need more information on any of these steps, consult your printer's manual.

○ It worked! I turned it on and now I can print.
○ Yes, my printer is on, but it still won't print.
○ I don't want to do this now.

Next
Vista Project: NASA Mission Control

Decision-theoretic methods for display for high-stakes aerospace decisions
Costs & Benefits of Viewing Information

Decision quality

Quantity of relevant information
Status Quo at Mission Control
Time-Critical Decision Making

- Consideration of time delay in temporal process

Diagram:
- Action A, t
- Duration of Process
- State of System H, t₀
- State of System H, t’
- Utility

Nodes:
- E₁, t₀
- E₂, t₀
- Eₙ, t₀
- E₁, t’
- E₂, t’
- Eₙ, t’
Simplification: Highlighting Decisions

- Variable threshold to control amount of highlighted information

<table>
<thead>
<tr>
<th></th>
<th>15.6</th>
<th>14.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>10.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>5.4</td>
<td>4.8</td>
</tr>
<tr>
<td>He Pres</td>
<td>17.7</td>
<td>14.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>33.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>10.2</th>
<th>10.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>12.8</td>
<td>12.5</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>He Pres</td>
<td>15.8</td>
<td>15.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>32.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>
Simplification: Highlighting Decisions

- Variable threshold to control amount of highlighted information

<table>
<thead>
<tr>
<th></th>
<th>15.6</th>
<th>14.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>10.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>5.4</td>
<td>4.8</td>
</tr>
<tr>
<td>He Pres</td>
<td>17.7</td>
<td>14.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>33.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>10.2</th>
<th>10.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>12.8</td>
<td>12.5</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>He Pres</td>
<td>15.8</td>
<td>15.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>32.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>
Simplification: Highlighting Decisions

- Variable threshold to control amount of highlighted information

<table>
<thead>
<tr>
<th></th>
<th>1st Set</th>
<th>2nd Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>15.6</td>
<td>14.2</td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>10.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>5.4</td>
<td>4.8</td>
</tr>
<tr>
<td>He Pres</td>
<td>17.7</td>
<td>14.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>33.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1st Set</th>
<th>2nd Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>10.2</td>
<td>10.6</td>
</tr>
<tr>
<td>Fuel Pres</td>
<td>12.8</td>
<td>12.5</td>
</tr>
<tr>
<td>Chamb Pres</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>He Pres</td>
<td>15.8</td>
<td>15.7</td>
</tr>
<tr>
<td>Delta v</td>
<td>32.3</td>
<td>63.3</td>
</tr>
</tbody>
</table>
What is Collaborative Filtering?

- A way to find cool websites, news stories, music artists etc
- Uses data on the preferences of many users, not descriptions of the content.
- Firefly, Net Perceptions (GroupLens), and others offer this technology.
Bayesian Clustering for Collaborative Filtering

- Probabilistic summary of the data
- Reduces the number of parameters to represent a set of preferences
- Provides insight into usage patterns.
- Inference:

\[ P(\text{Like title } i \mid \text{Like title } j, \text{Like title } k) \]
Applying Bayesian clustering

```
title 1

user classes

title 1       title 2       ...       title n
```

<table>
<thead>
<tr>
<th>titles</th>
<th>class1</th>
<th>class2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>title1</td>
<td>p(like)=0.2</td>
<td>p(like)=0.8</td>
<td></td>
</tr>
<tr>
<td>title2</td>
<td>p(like)=0.7</td>
<td>p(like)=0.1</td>
<td></td>
</tr>
<tr>
<td>title3</td>
<td>p(like)=0.99</td>
<td>p(like)=0.01</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
MSNBC Story clusters

Readers of commerce and technology stories (36%):
- E-mail delivery isn’t exactly guaranteed
- Should you buy a DVD player?
- Price low, demand high for Nintendo

Sports Readers (19%):
- Umps refusing to work is the right thing
- Cowboys are reborn in win over eagles
- Did Orioles spend money wisely?

Readers of top promoted stories (29%):
- 757 Crashes At Sea
- Israel, Palestinians Agree To Direct Talks
- Fuhrman Pleads Innocent To Perjury

Readers of “Softer” News (12%):
- The truth about what things cost
- Fuhrman Pleads Innocent To Perjury
- Real Astrology

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Top 5 shows by user class

Class 1
- Power rangers
- Animaniacs
- X-men
- Tazmania
- Spider man

Class 2
- Young and restless
- Bold and the beautiful
- As the world turns
- Price is right
- CBS eve news

Class 3
- Tonight show
- Conan O’Brien
- NBC nightly news
- Later with Kinnear
- Seinfeld

Class 4
- 60 minutes
- NBC nightly news
- CBS eve news
- Murder she wrote
- Matlock

Class 5
- Seinfeld
- Friends
- Mad about you
- ER
- Frasier
Richer model
What’s old?

Decision theory & probability theory provide:

- principled models of belief and preference;
- techniques for:
  - integrating evidence (conditioning);
  - optimal decision making (max. expected utility);
  - targeted information gathering (value of info.);
  - parameter estimation from data.
What’s new?

Bayesian networks exploit domain structure to allow compact representations of complex models.
Some Important AI Contributions

- Key technology for diagnosis.
- Better more coherent expert systems.
- New approach to planning & action modeling:
  - planning using Markov decision problems;
  - new framework for reinforcement learning;
  - probabilistic solution to frame & qualification problems.
- New techniques for learning models from data.
What’s in our future?

- Better models for:
  - preferences & utilities;
  - not-so-precise numerical probabilities.
- Inferring causality from data.
- More expressive representation languages:
  - structured domains with multiple objects;
  - levels of abstraction;
  - reasoning about time;
  - hybrid (continuous/discrete) models.