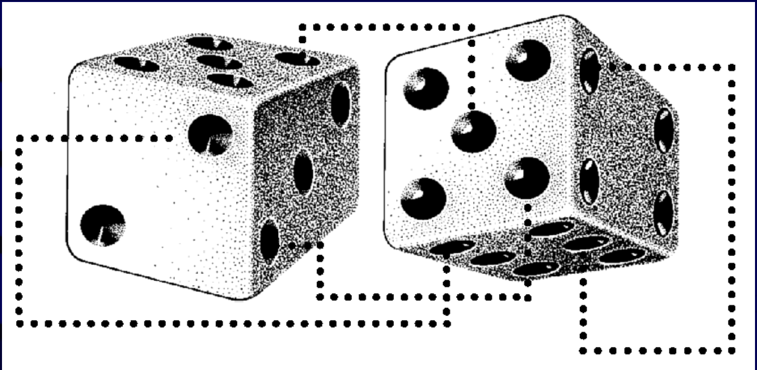
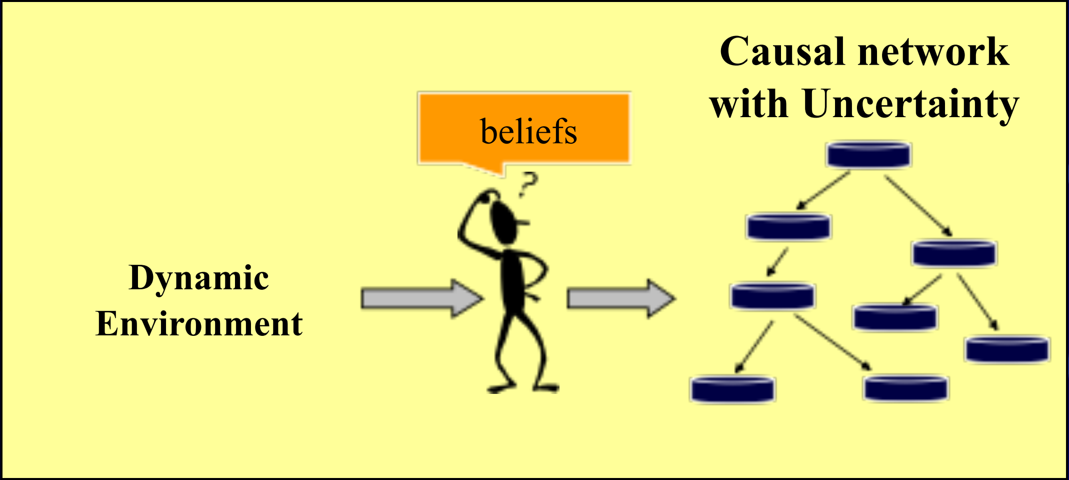
Introduction to Bayesian Networks



Based on the Tutorials and Presentations:

(1) Dennis M. Buede Joseph A. Tatman, Terry A. Bresnick;
(2) Jack Breese and Daphne Koller;
(3) Scott Davies and Andrew Moore;
(4) Thomas Richardson
(5) Roldano Cattoni
(6) Irina Rich

Discovering Causal Relationship from the Dynamic Environmental Data and Managing Uncertainty - are among the basic abilities of an intelligent agent



Overview

- **Probabilities basic rules**
- **Bayesian Nets**
- **Conditional Independence**
- **Motivating Examples**
- **Inference in Bayesian Nets**
- **Join Trees**
- **Decision Making with Bayesian Networks**
- **Learning Bayesian Networks from Data**
- **Profiling with Bayesian Network**
- **References and links**

Probability of an event

- Define the *probability* of an event A, $P(A)$ to be a number between 0 and 1 such that:
- $P(A) = 1$ if the event A is certain.
- $P(A \cup B) = P(A) + P(B)$ if the events A and B are mutually exclusive
  - » i.e. it is not possible for both A and B to occur simultaneously.
  - » $A \cup B$ is the event "either A happens or B happens"

Conditional probability

- A *conditional probability statement* has the form:
- "Given that B happened, the probability that A happened is x."
  - » This is written: $P(A \mid B) = x$.
- This does *not* imply that whenever B is true the probability of A is x
  - » P(Snow today | Snow yesterday)
  - » $\neq$ P(Snow today | Snow yesterday, the month is July)

Probabilities

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

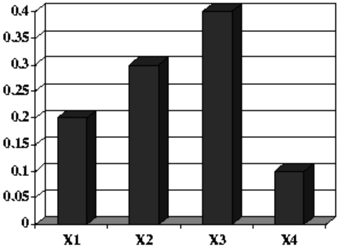
Discrete Random Variables

- Finite set of possible outcomes

$$X \in \{x_1, x_2, x_3, \ldots, x_n\}$$

$$P(x_i) \geq 0$$

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



X binary: $P(x) + P(\overline{x}) = 1$

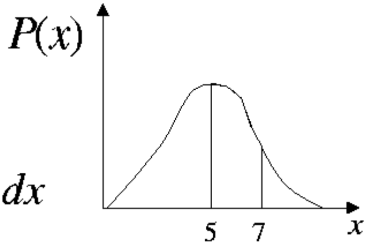
Continuous Random Variable

- Probability distribution (density function) over continuous values

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

$$\int_0^{10} P(x)\,dx = 1$$

$$P(5 \leq x \leq 7) = \int_5^7 P(x)\,dx$$



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

$$P(x, y) \equiv P(X = x \wedge Y = y)$$

  - Probability that both X=x and Y=y

Conditional independence

- Two events A and B are said to be (probabilistically) independent if:

$$P(A,B) = P(A)P(B)$$

equivalently $P(A \mid B) = P(A)$ $\quad (P(B)>0)$

- Two events A and B are said to be (probabilistically) independent given C if:

$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$

equivalently $P(A \mid B,C) = P(A \mid C)$ $\quad (P(B)>0)$

The fundamental rule

Conditional and joint probabilities are combined with the following rule:

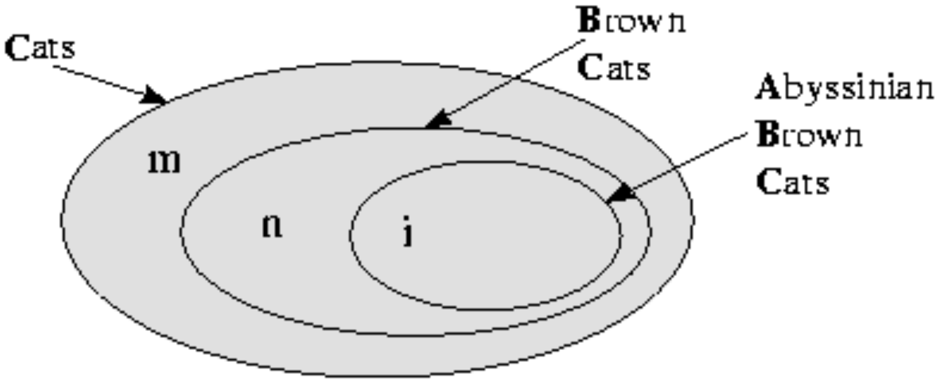
The 'Fundamental Rule'

$$P(A \mid B)\, P(B) = P(A, B)$$

Fundamental Rule given a context

$$P(A \mid B, C)\, P(B \mid C) = P(A, B \mid C)$$

Instance of "Fundamental rule"



$$P(A \mid B,C) = i/n \quad P(B \mid C) = n/m \quad P(A,B \mid C) = i/m$$

$$P(A \mid B,C)\, P(B \mid C) = P(A,B \mid C)$$

Rules of Probability

■ Product Rule

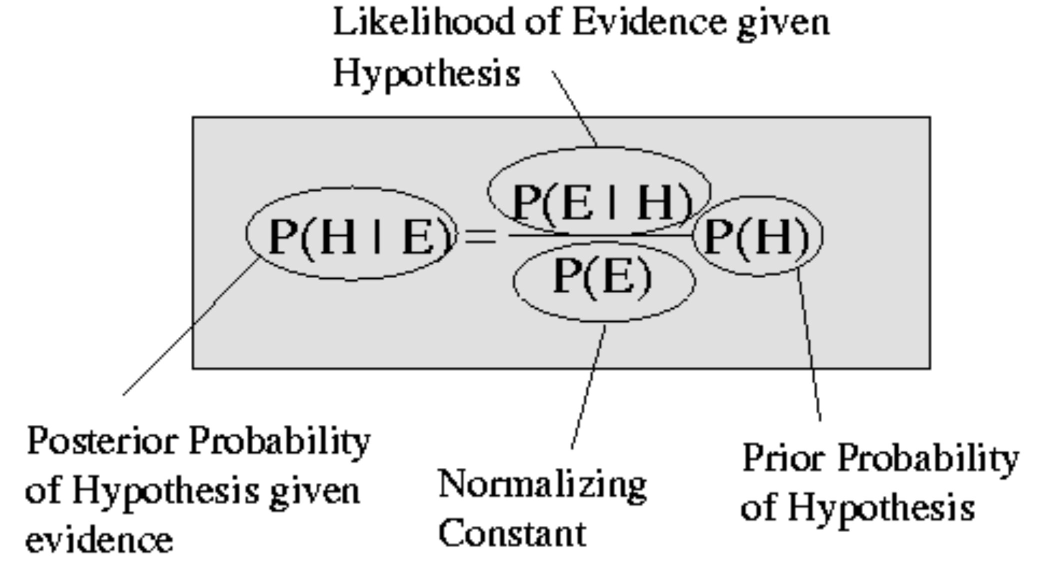
$$P(X,Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X)$$

■ Marginalization

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

X binary: $P(Y) = P(Y, x) + P(Y, \overline{x})$

Bayes rule


Likelihood of Evidence given Hypothesis
P(E | H)
P(H | E) =
P(H)
P(E)
Posterior Probability of Hypothesis given evidence
Normalizing Constant
Prior Probability of Hypothesis

Bayes rule example (1)

$$P(\text{Cancer} \mid \text{Test+}) = \frac{P(\text{Test+} \mid \text{Cancer})P(\text{Cancer})}{P(\text{Test +})}$$

$P(\text{Test+} \mid \text{Cancer}) = 0.9 \quad P(\text{Test–} \mid \text{Cancer})=0.1$

$P(\text{Test+} \mid \text{No Cancer}) = 0.01 \quad P(\text{Test–} \mid \text{No Cancer}) =0.99$

$P(\text{Cancer})=0.0001$

$$\begin{aligned} P(\text{Test+}) &= P(\text{Test+} \mid \text{Cancer})P(\text{Cancer}) + \\ &\quad P(\text{Test+} \mid \text{No cancer})P(\text{No Cancer}) \\ &= 0.9\text{x}0.0001 + 0.01\text{x}(1–0.0001) \\ &= 0.0010899 \end{aligned}$$

Bayes rule example (2)

$$P(\text{Cancer} \mid \text{Test+}) = \frac{P(\text{Test+} \mid \text{Cancer})P(\text{Cancer})}{P(\text{Test+})}$$

$$P(\text{Test+} \mid \text{Cancer}) = 0.9$$

$$P(\text{Cancer}) = 0.0001$$

$$P(\text{Test+}) = 0.0010899$$

$$P(\text{Cancer} \mid \text{Test+}) = 0.9 \text{x} 0.0001 / 0.0010899$$

$$= 0.08$$

Overview

– Probabilities basic rules
– **Bayesian Nets**
– Conditional Independence
– Motivating Examples
– Inference in Bayesian Nets
– Join Trees
– Decision Making with Bayesian Networks
– Learning Bayesian Networks from Data
– Profiling with Bayesian Network
– References and links

What are Bayesian nets?

อธิบายความไม่แน่นอน (ค่าสถิติ)

– Bayesian nets (BN) are a network-based framework for representing and analyzing models involving uncertainty;

– BN are different from other knowledge-based systems tools because uncertainty is handled in mathematically rigorous yet efficient and simple way

– BN are different from other probabilistic analysis tools because of network representation of problems, use of Bayesian statistics, and the synergy between these

Definition of a Bayesian Network

Knowledge structure:

variables are nodes

arcs represent probabilistic dependence between variables

conditional probabilities encode the strength of the dependencies

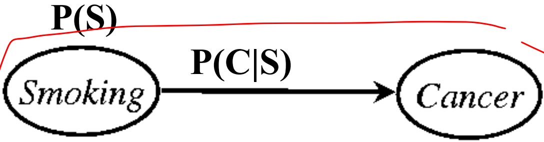
Computational architecture:

- computes posterior probabilities given evidence about some nodes
- exploits probabilistic independence for efficient computation

Bayesian Networks

direct graph

$S \in \{no, light, heavy\}$



$C \in \{none, benign, malignant\}$

P(S)

$P(S{=}no)$	0.80
$P(S{=}light)$	0.15
$P(S{=}heavy)$	0.05

ไม่สูบเลย

P(C|S)

$Smoking{=}$	no	$light$	$heavy$
$P(C{=}none)$	0.96	0.88	0.60
$P(C{=}benign)$	0.03	0.08	0.25
$P(C{=}malig)$	0.01	0.04	0.15

Product Rule

?

- $P(C,S) = P(C|S)\,P(S)$

$$P(C|S) = \frac{P(S \cap C)}{P(S)}$$

$S\Downarrow$ $C\Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

Marginalization

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

Bayes Rule Revisited

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

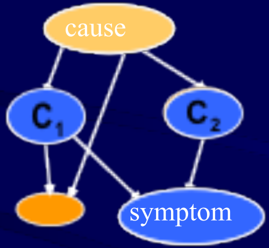
$S\Downarrow$ $C\Rightarrow$	*none*	*benign*	*malig*
no	0.768/.935	0.024/.046	0.008/.019
light	0.132/.935	0.012/.046	0.006/.019
heavy	0.030/.935	0.015/.046	0.005/.019

Cancer=	*none*	*benign*	*malignant*
P(S=no)	0.821	0.522	0.421
P(S=light)	0.141	0.261	0.316
P(S=heavy)	0.037	0.217	0.263

What Bayesian Networks are good for?

- Diagnosis: P(cause|symptom)=?
- Prediction: P(symptom|cause)=?
- Classification: P(class|data)
- Decision-making (given a cost function)

Com [illegible]

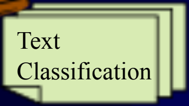


Medicine

Speech recognition

Bio- informatics

Stock market



Computer troubleshooting

Why learn Bayesian networks?

- Combining domain expert knowledge with data

- Efficient representation and inference
- Incremental learning
- Handling missing data: **<1.3 2.8 ?? 0 1 >**
- Learning causal relationships:

Overview

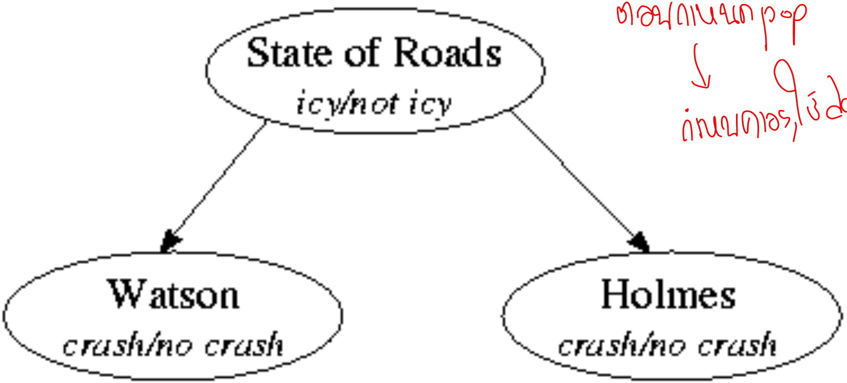
- Probabilities basic rules
- Bayesian Nets
- **Conditional Independence**
- Motivating Examples
- Inference in Bayesian Nets
- Join Trees
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- References and links

"Icy roads" example

(เอา case มาสร้าง network

- Inspector Smith is waiting for Holmes and Watson who are both late for an appointment.
- Smith is worried that if the roads are icy one or both of them may have crashed his car.
- Suddenly Smith learns that Watson has crashed.
- Smith thinks: *If Watson has crashed, probably the roads are icy, then Holmes has probably crashed too!*
- Smith then learns it is warm outside and roads are salted
- Smith thinks: *Watson was unlucky; Holmes should still make it.*

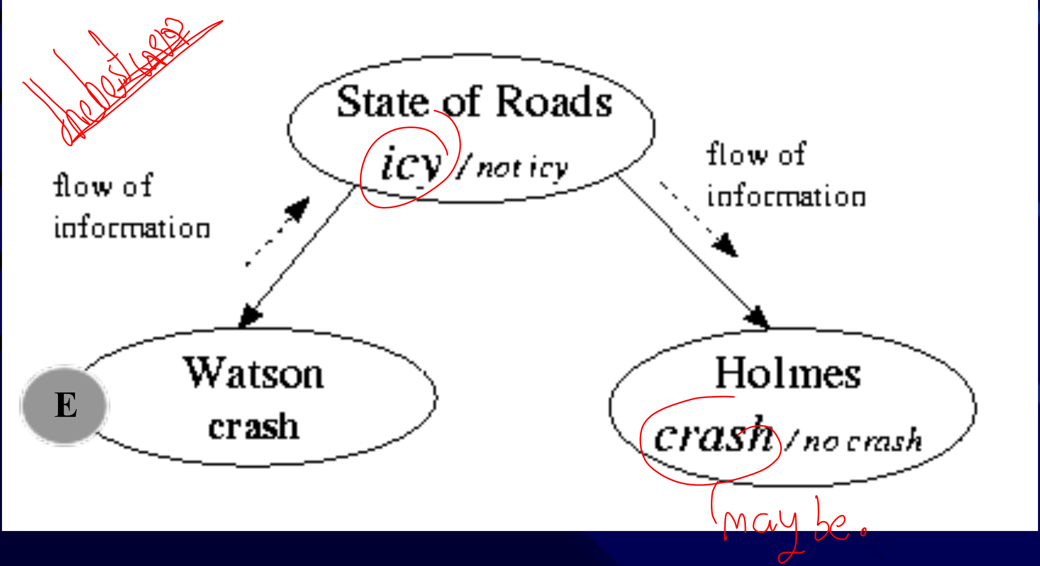
Causal relationships



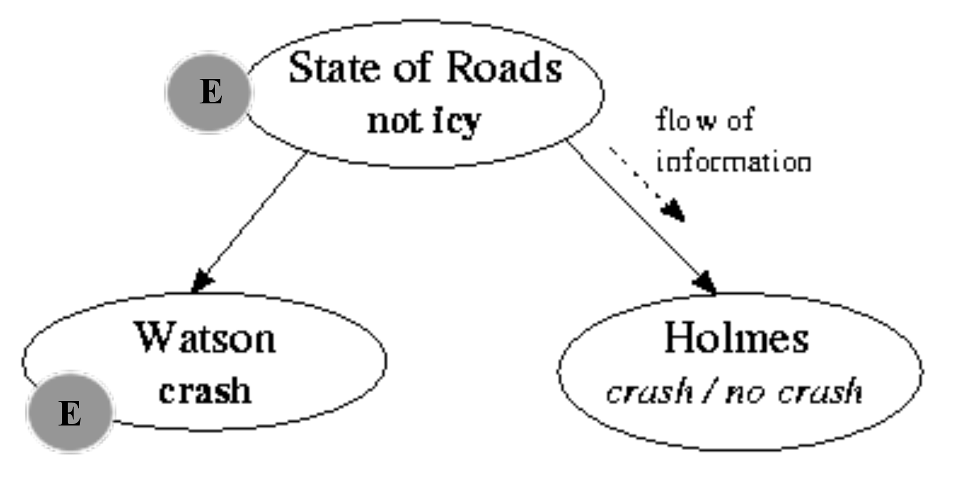
ตอนที่เป็นแค่ pop

กำหนดเองไม่ได้

Watson has crashed !



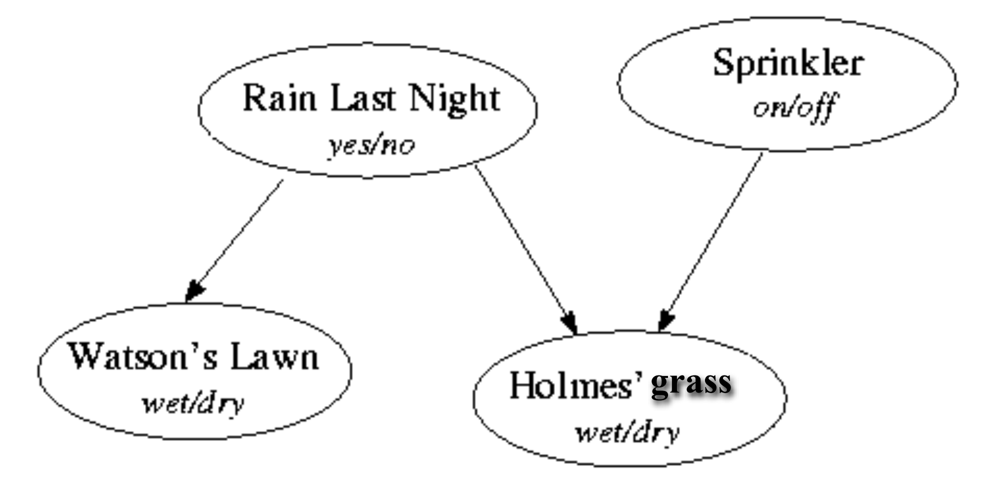
… But the roads are salted !



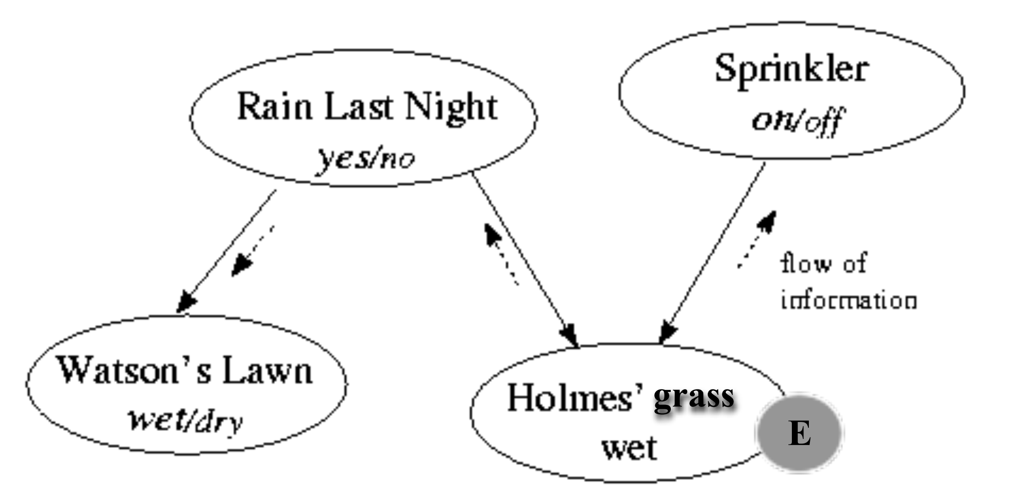
"Wet grass" example

- One morning as Holmes leaves for work, he notices that his grass is wet. He wonders whether he has left his sprinkler on, or it has rained.
- Glancing over to Watson's lawn he notices that it is also wet.
- Holmes thinks: *Since Watson's lawn is wet, it probably rained last night.*
- He then thinks: *If it rained then that explains why my* ***grass*** *is wet, so probably the sprinkler is off.*

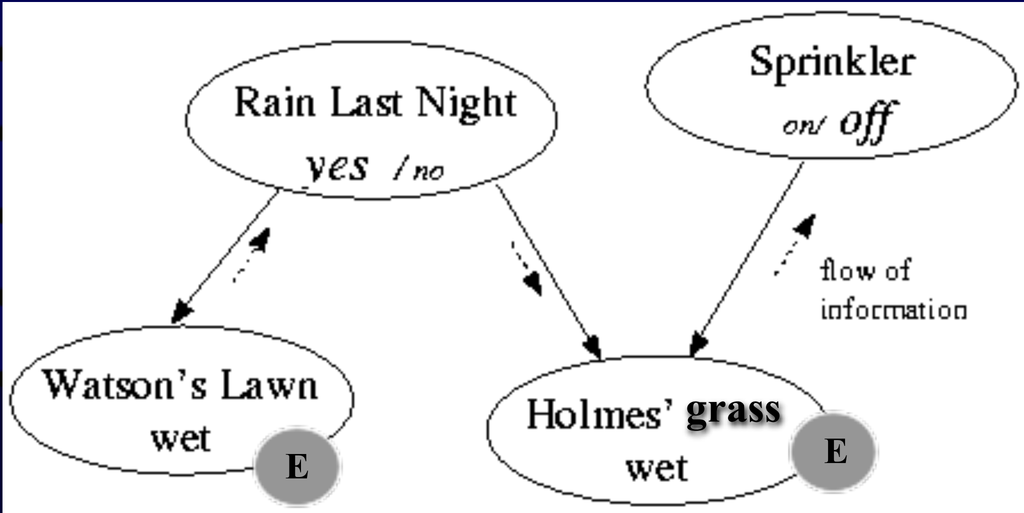
Causal relationships



Holmes' grass is wet !



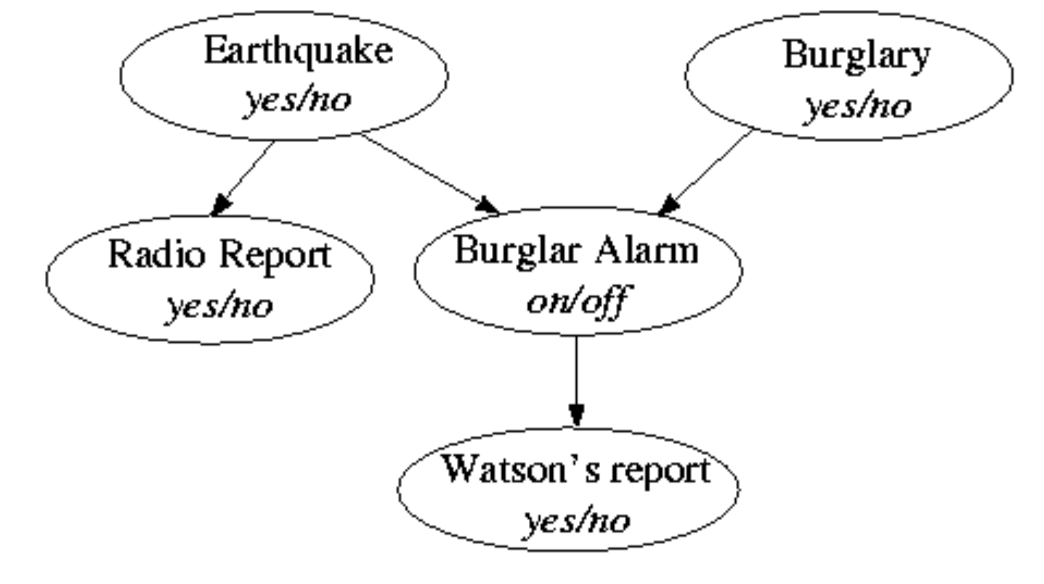
Watson's lawn is also wet !



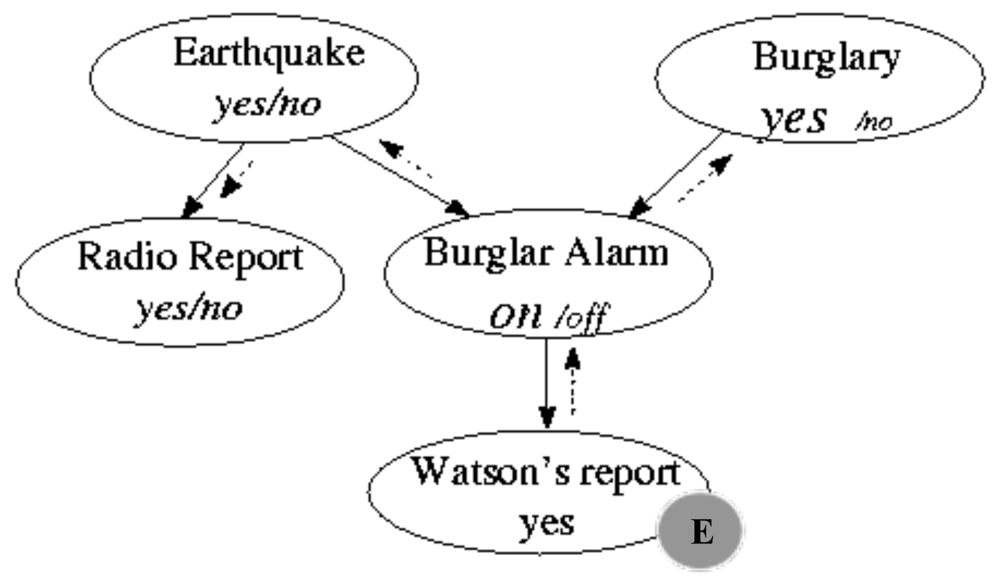
"Burglar alarm" example

- Holmes is at work when he receives a call from Watson, informing him that his alarm has gone off.
- Holmes thinks it is likely that the alarm really went off, although Watson sometimes play practical jokes.
- Holmes is on his way home when he hears a report on the radio, that there was an earthquake in the vicinity.
- Since the burglar alarm has been known to go off when there is an earthquake, Holmes reckons that a burglary is unlikely.
- Holmes goes back to work. (Leaving the noise for Watson)

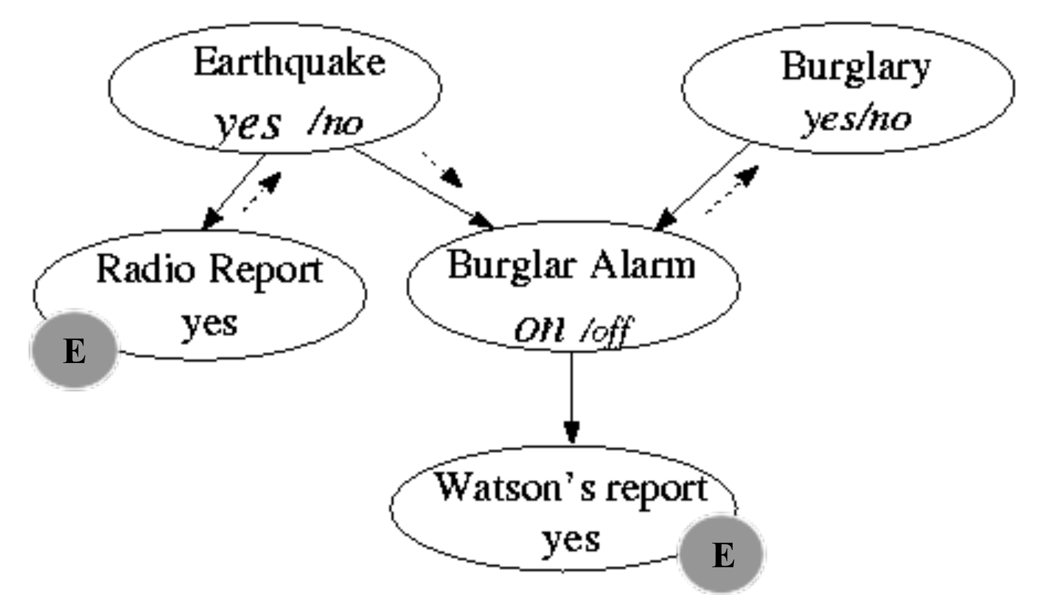
Causal relationships



Watson reports about alarm

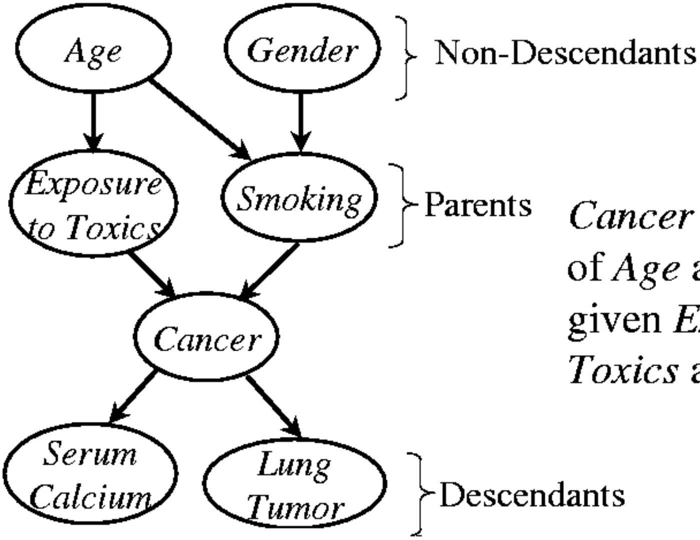


Radio reports about earthquake



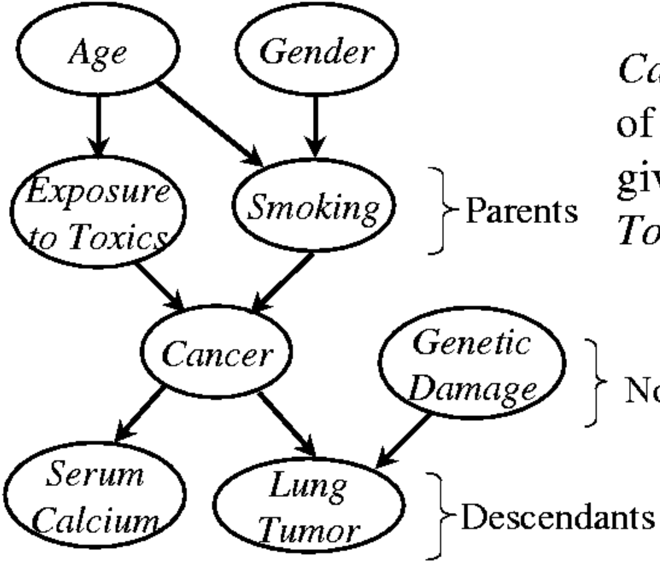
Conditional Independence

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



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

Another non-descendant



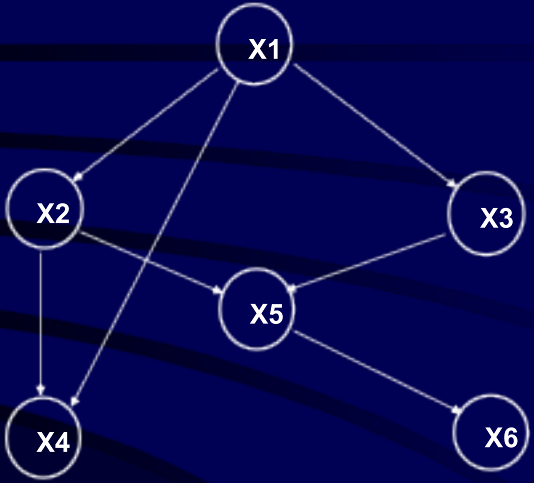
Cancer is independent of *Genetic Damage* given *Exposure to Toxics* and *Smoking*.

General Product (Chain) Rule for Bayesian Networks

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \boldsymbol{Pa_i})$$

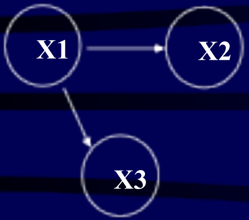
$$\boldsymbol{Pa_i} = parents(X_i)$$

Sample of General Product Rule



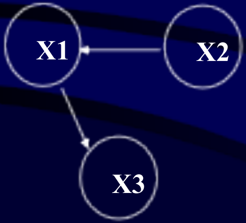
p(x1, x2, x3, x4, x5, x6) = p(x6 | x5) p(x5 | x3, x2) p(x4 | x2, x1) p(x3 | x1) p(x2 | x1) p(x1)

Arc Reversal - Bayes Rule



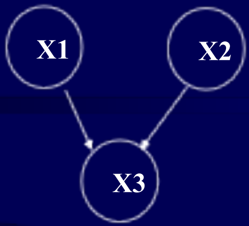
$p(x1, x2, x3) = p(x3 \mid x1)\, p(x2 \mid x1)\, p(x1)$

is equivalent to



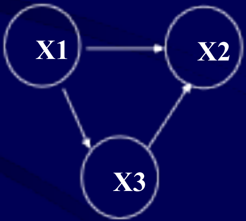
$p(x1, x2, x3) = p(x3 \mid x1)\, p(x2 , x1)$

$= p(x3 \mid x1)\, p(x1 \mid x2)\, p(x2)$



$p(x1, x2, x3) = p(x3 \mid x2, x1)\, p(x2)\, p(x1)$

is equivalent to



$p(x1, x2, x3) = p(x3, x2 \mid x1)\, p(x1)$

$= p(x2 \mid x3, x1)\, p(x3 \mid x1)\, p(x1)$

D-Separation of variables

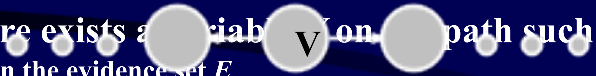
- Fortunately, there is a relatively simple algorithm for determining whether two variables in a Bayesian network are conditionally independent: *d-separation*.

- Definition: *X* and *Z* are *d-separated* by a set of evidence variables *E* iff every undirected path from *X* to *Z* is "blocked".
- A path is "blocked" iff one or more of the following conditions is true: ...

A path is blocked when:

- There exists a variable *V* on the path such that
  - it is in the evidence set *E*
  - the arcs putting *V* in the path are "tail-to-tail"

- Or, there exists a variable *V* on the path such that

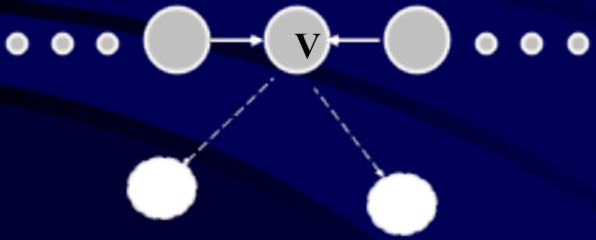


  - it is in the evidence set *E*
  - the arcs putting *V* in the path are "tail-to-head"

- Or, ...

… a path is blocked when:

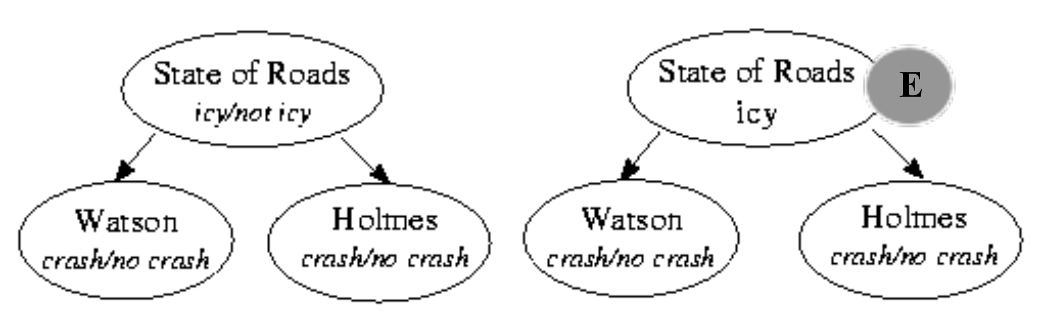
- … **Or, there exists a variable *V* on the path such that**
  - **it** is NOT **in the evidence set *E***
  - neither are any of its descendants
  - **the arcs putting *V* on the path are "head-to-head"**



D-Separation and independence

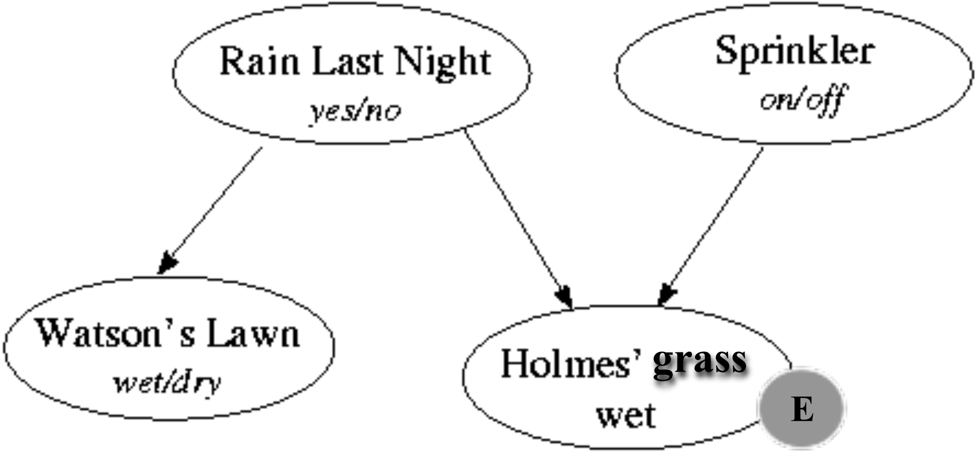
- Theorem [Verma & Pearl, 1998]:
  - If a set of evidence variables E d-separates X and Z in a Bayesian network's graph, then X and Z will be independent.
- *d*-separation can be computed in linear time.
- Thus we now have a fast algorithm for automatically inferring whether learning the value of one variable might give us any additional hints about some other variable, given what we already know.

Holmes and Watson: "Icy roads" example



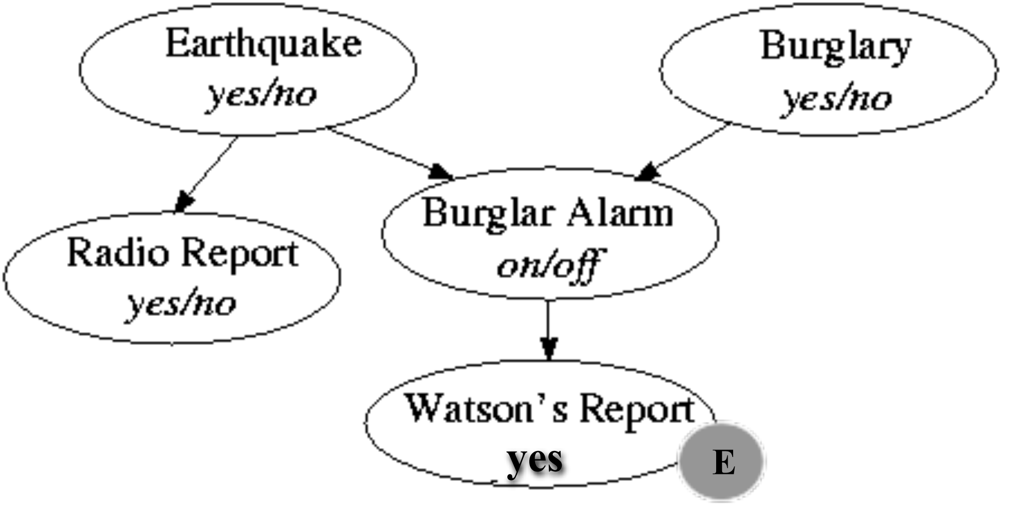
Watson and Holmes are d-connected given no evidence, but
Watson and Holmes are d-separated given State of Roads

Holmes and Watson: "Wet grass" example



Watson's Lawn and Sprinkler are d-connected given Holmes' grass

Holmes and Watson: "Burglar alarm" example

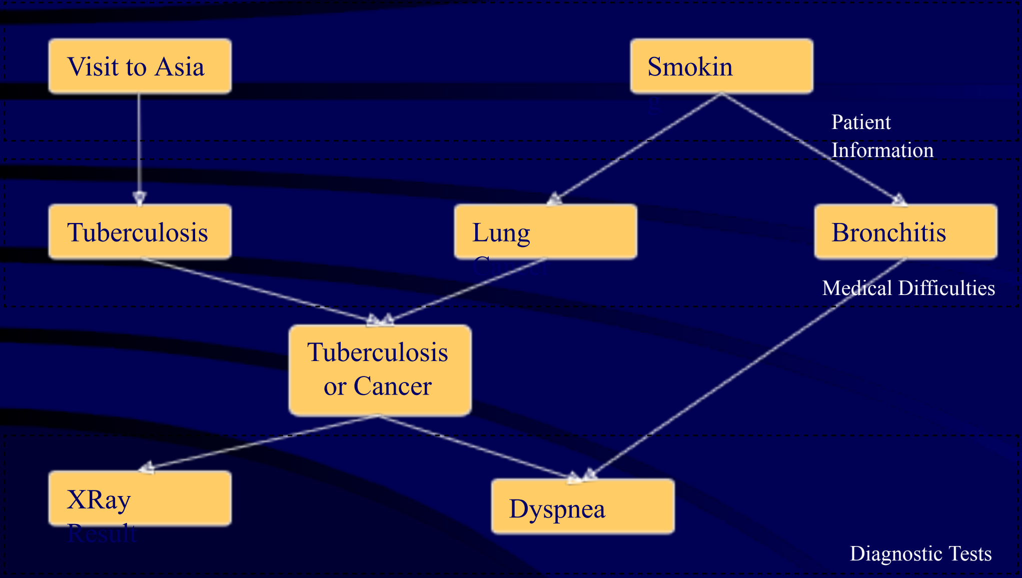


Radio Report and Burglary are d-connected given Watson's Report

Overview

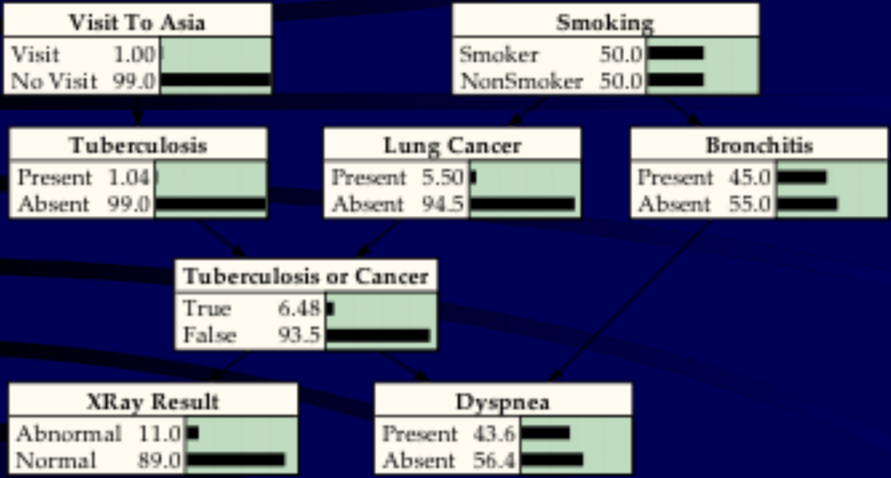
- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- **Motivating Examples**
- Inference in Bayesian Nets
- Join Trees
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- References and links

Example from Medical Diagnostics



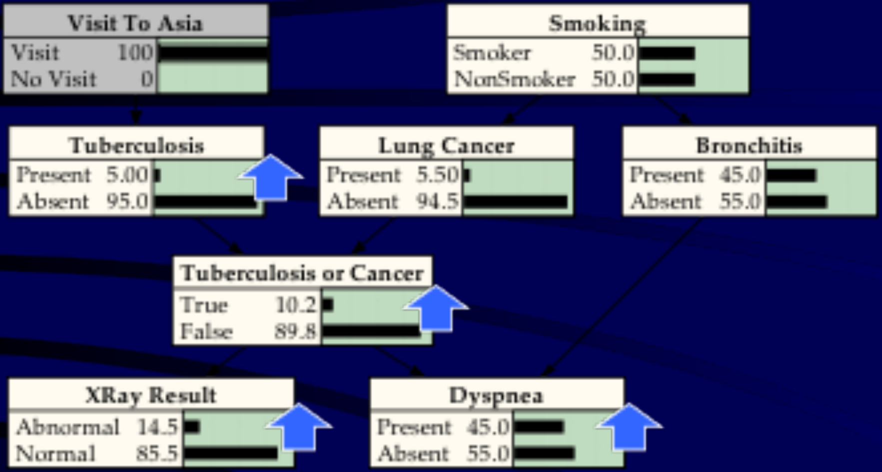
Network represents a knowledge structure that models the relationship between medical difficulties, their causes and effects, patient information and diagnostic tests

Example from Medical Diagnostics



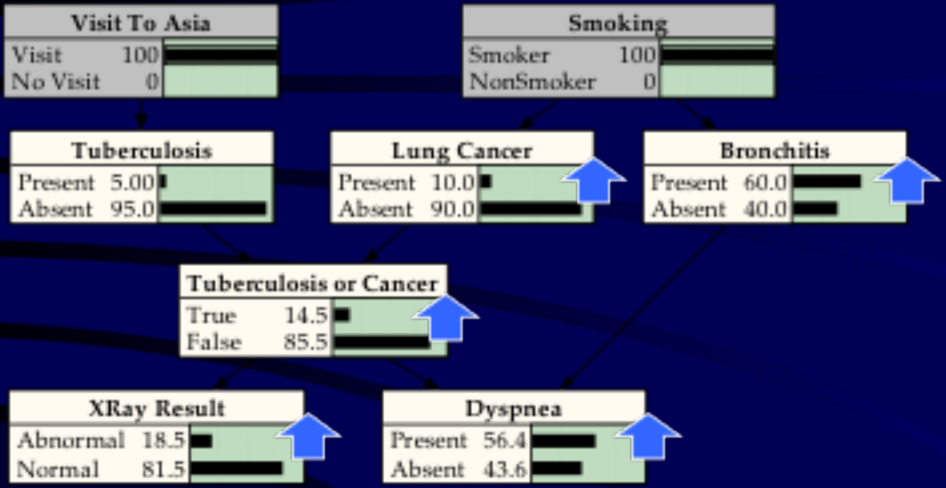
- Propagation algorithm processes relationship information to provide an unconditional or marginal probability distribution for each node
- The unconditional or marginal probability distribution is frequently called the belief function of that node

Example from Medical Diagnostics



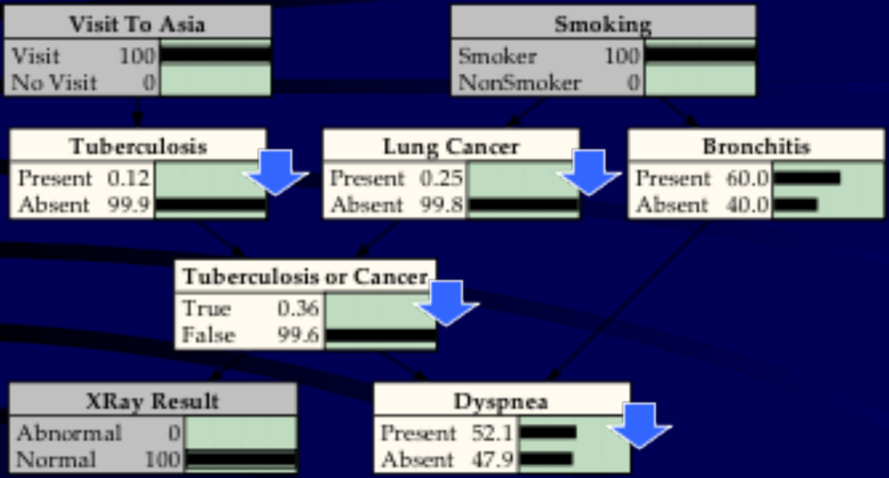
- **As a finding is entered, the propagation algorithm updates the beliefs attached to each relevant node in the network**
- **Interviewing the patient produces the information that "Visit to Asia" is "Visit"**
- **This finding propagates through the network and the belief functions of several nodes are updated**

Example from Medical Diagnostics



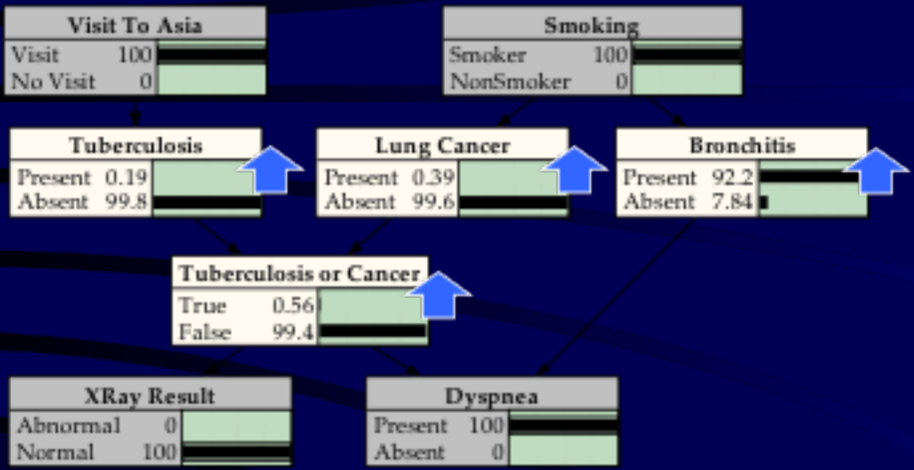
- **Further interviewing of the patient produces the finding "Smoking" is "Smoker"**
- **This information propagates through the network**

Example from Medical Diagnostics



- Finished with interviewing the patient, the physician begins the examination
- The physician now moves to specific diagnostic tests such as an X-Ray, which results in a "Normal" finding which propagates through the network
- Note that the information from this finding propagates backward and forward through the arcs

Example from Medical Diagnostics



- The physician also determines that the patient is having difficulty breathing, the finding "Present" is entered for "Dyspnea" and is propagated through the network
- The doctor might now conclude that the patient has bronchitis and does not have tuberculosis or lung cancer

Overview

- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- **Inference in Bayesian Nets**
- Join Trees
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- References and links

Inference Using Bayes Theorem

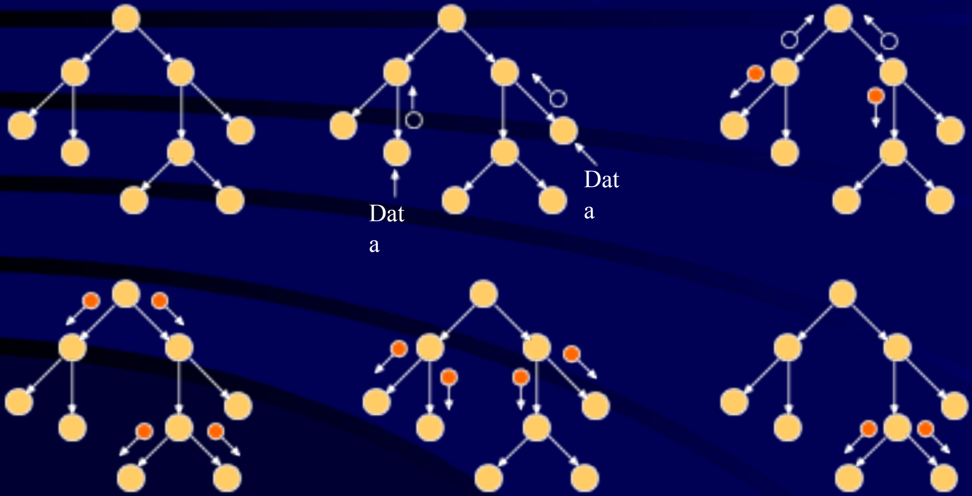
- The general probabilistic inference problem is to find the probability of an event given a set of evidence;

- This can be done in Bayesian nets with sequential applications of Bayes Theorem;

- In 1986 Judea Pearl published an innovative algorithm for performing inference in Bayesian nets.

Propagation Example

"The impact of each new piece of evidence is viewed as a perturbation that propagates through
the network via message-passing between
neighboring variables . . . ." (Pearl, 1988, p 143



- The example above requires five time periods to reach equilibrium after the introduction of data (Pearl, 1988, p 174)

Basic Inference

$$P(b) = ?$$

Product Rule

- $P(C,S) = P(C|S)\,P(S)$

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

Marginalization

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

Basic Inference

$$\textcircled{A} \longrightarrow \textcircled{B} \longrightarrow \textcircled{C}$$

$$\underbrace{P(b)} = \sum_a P(a, b) = \sum_a P(b \mid a)\, P(a)$$

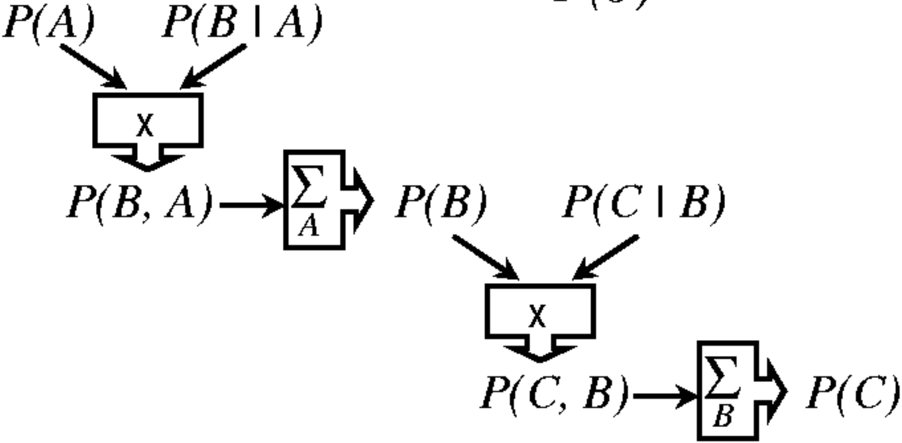
$$P(c) = \sum_b P(c \mid b)\, \overbrace{P(b)}$$

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

$$= \sum_b P(c \mid b) \underbrace{\sum_a P(b \mid a)\, P(a)}_{P(b)}$$

Variable elimination

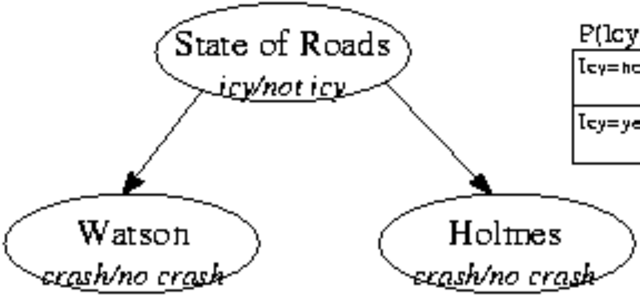
$$P(c) = \sum_b P(c \mid b) \underbrace{\sum_a P(b \mid a)\, P(a)}_{P(b)}$$



"Icy roads" example

- Inspector Smith is waiting for Holmes and Watson who are both late for an appointment.
- Smith is worried that if the roads are icy one or both of them may have crashed his car.
- Suddenly Smith learns that Watson has crashed.
- Smith thinks: *If Watson has crashed, probably the roads are icy, then Holmes has probably crashed too!*
- Smith then learns it is warm outside and roads are salted
- Smith thinks: *Watson was unlucky; Holmes should still make it.*

Bayes net for "Icy roads" example



P(Icy)

Icy=no	0.3
Icy=yes	0.7

P(Watson I Icy)	Icy = yes	Icy = no
Watson Crash = yes	0.8	0.1
Watson Crash= no	0.2	0.9

P(Holmes I Icy)	Icy = yes	Icy = no
Holmes Crash = yes	0.8	0.1
Holmes Crash = no	0.2	0.9

Extracting marginals

To find P(Holmes Crash) we first compute P(Holmes Crash, Icy) using the fundamental rule:

e.g. P(H Crash = yes, Icy = yes)

= P(H Crash=yes | Icy =yes)P(Icy=yes)

P(Holmes,Icy)	Icy = yes	Icy = no	P(H Crash)
Holmes Crash = yes	0.8 x0.7=0.56	0.1 x0.3=0.03	0.56+0.03=0.59
Holmes Crash = no	0.2 x0.7=0.14	0.9 x0.3=0.27	0.14+0.27=0.41

Then summing each row gives us the requred probabilities.
By symmetry P (W Crash) is the same.

Updating with Bayes rule (given evidence "Watson has crashed")

After we discover that Watson has crashed we can compute P(Icy | W Crash =y) using Bayes rule:

$$P(\text{Icy} \mid \text{W Crash}=y) = \frac{P(\text{W Crash} = y \mid \text{Icy})P(\text{Icy})}{P(\text{W Crash} = y)}$$

$$= (0.8\text{x}0.7, 0.1\text{x}0.3)/0.59$$

$$= (0.95, 0.05)$$

Extracting the marginal

- To calculate P(H Crash | W Crash = y) we first calculate P(H Crash, Icy | W Crash)

P(H | W=y, Icy)	Icy = yes	Icy = no	
Holmes Crash = yes	0.8 x0.95=0.76	0.1 x0.05=0.005	0.765
Holmes Crash = no	0.2 x0.95=0.19	0.9 x0.05=0.045	0.235

Again, summing gives us P(H Crash | W Crash = yes)

P(H Crash | W Crash, Icy=no) = P(H Crash | Icy=no)
= (0.1,0.9)

Alternative perspective

Icy = no	0.3
Icy = yes	0.7

	Icy = yes	Icy = no
Watson Crash = yes	0.56	0.03
Watson Crash = no	0.14	0.27

	Icy = yes	Icy = no
Holmes Crash = yes	0.56	0.03
Holmes Crash = no	0.14	0.27

We represent the model as two joint tables, P(Watson, Icy) and P(Holmes, Icy) with a table for the overlap P(Icy).

Alternative perspective

Icy=no	0.3
Icy=yes	0.7

	Icy = yes	Icy = no
Watson Crash = yes	0.56/0.59 = 0.95	0.03/0.59= 0.05
Watson Crash = no	0	0

	Icy = yes	Icy = no
Holmes Crash = yes	0.56	0.03
Holmes Crash = no	0.14	0.27

If evidence on Watson arrives of the form $P^{New}(\text{W Crash}) = (1,0)$

then $$P^{New}(\text{W Crash, Icy})$$

$$= P(\text{Icy} \mid \text{W Crash})\, P^{New}(\text{W Crash}) = \frac{P(\text{W Crash, Icy})}{P(\text{W Crash})}\, P^{New}(\text{W Crash})$$

Alternative perspective

Icy=no	0.95
Icy=yes	0.05

	Icy = yes	Icy = no
Watson Crash = yes	0.56/0.59 = 0.95	0.03/0.59= 0.05
Watson Crash = no	0	0

	Icy = yes	Icy = no
Holmes Crash = yes	0.56*0.95/0.7= 0.76	0.03*0.05/0.3= 0.005
Holmes Crash = no	0.14*0.95/0.7= 0.19	0.27*0.05/0.3= 0.045

The table for Icy can then be updated by marginalizing the table for Watson. The table for Holmes can then be updated using the same rule:

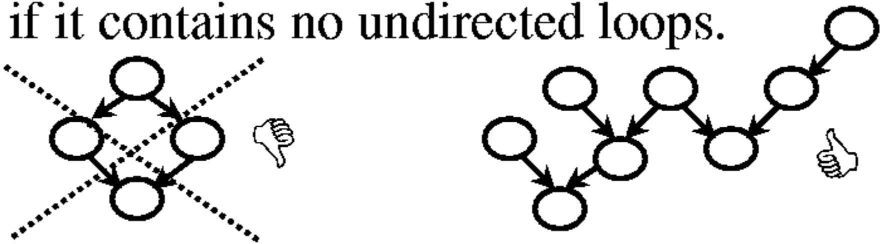
$$P^{New}(H\ Crash, Icy) = \frac{P(H\ Crash, Icy)}{P(Icy)} P^{New}(Icy)$$

Overview

- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- Inference in Bayesian Nets
- **Join Trees**
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- References and links

Polytrees

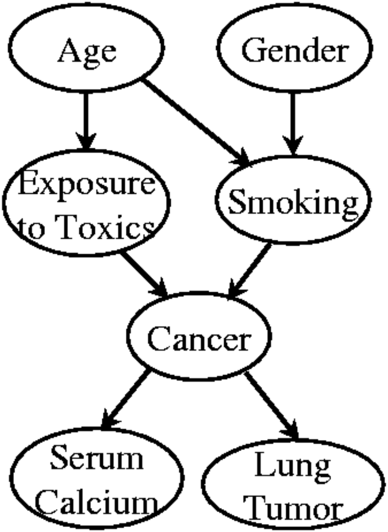
- A network is *singly connected* (a *polytree*) if it contains no undirected loops.

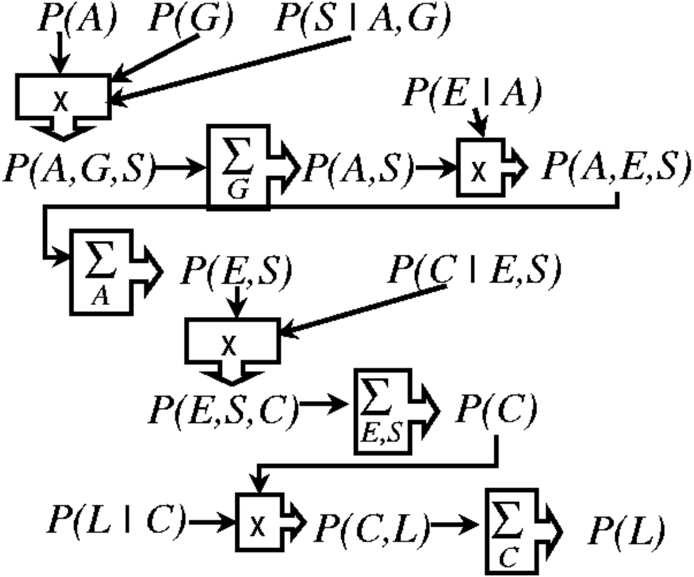


Theorem: Inference in a singly connected network can be done in linear time*.

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

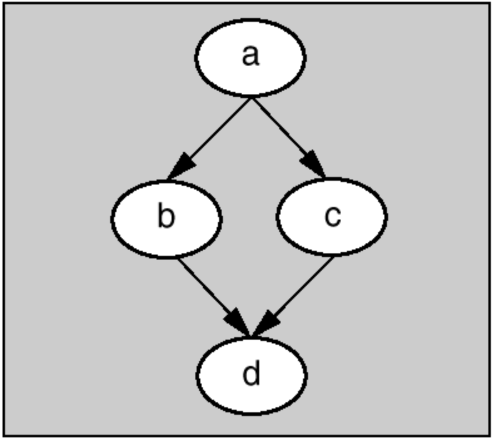
Variable Elimination with loops





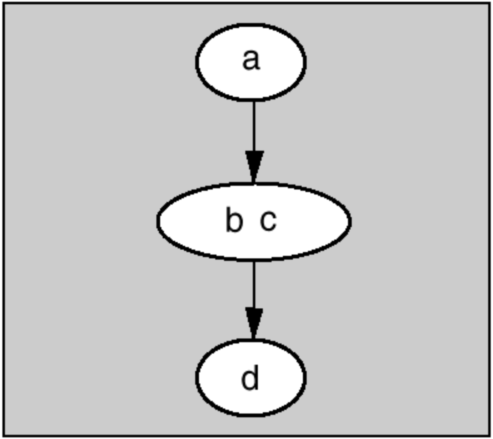
Complexity is exponential in the size of the factors

Join Trees



A Multiply Connected Network.

There are two paths between node a *and node* d.



A Clustered, Multiply Connected Network.

By clustering nodes b *and* c, *we turned the graph into a singly connected network.*

Graphical Method of Building the Junction Tree

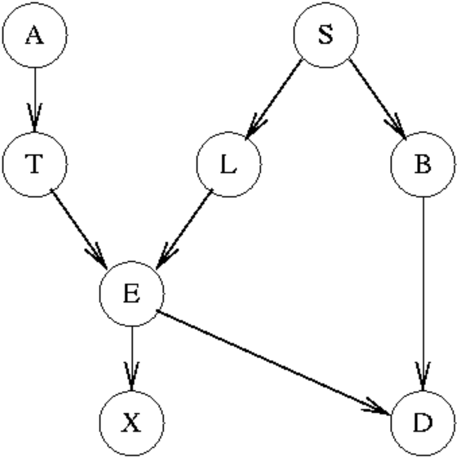
The Junction Tree can be constructed through a series of graph operations

- **Marry the Parents** ("moralize the graph"): Add an undirected edge between every pair of parents of a node (unless they are already connected.
- **Make All Arrows Undirected**
- **Triangulate the Graph:** Add edges so that every cycle of length 4 or more contains a chord.
- **Identify the maximal Cliques:** A clique is a complete graph. A maximal clique is a maximal complete subgraph.
- **Form Junction Graph:** Create a cluster node for each clique and label it with the variables in the clique.

  Create an edge between any pair of cluster nodes that share variables.

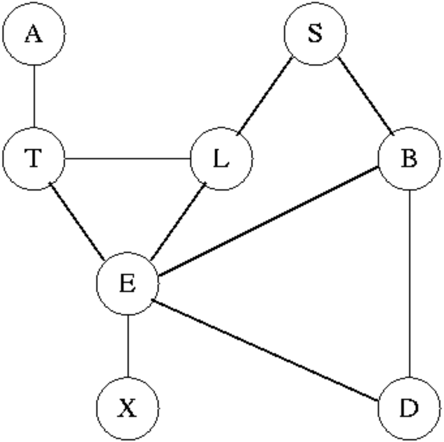
  Place a separator node on the edge labeled with the set of variables shared by the cluster nodes it joins.
- **Form the junction tree:** Compute a maximum weighted spanning tree of the junction graph where the weight on each edge is the number of variables in the separator of the edge.

Example



$$P(U) = P(A)P(S)P(T|A)P(L|S)P(B|S)P(E|L,T)P(D|B,E)P(X|E)$$

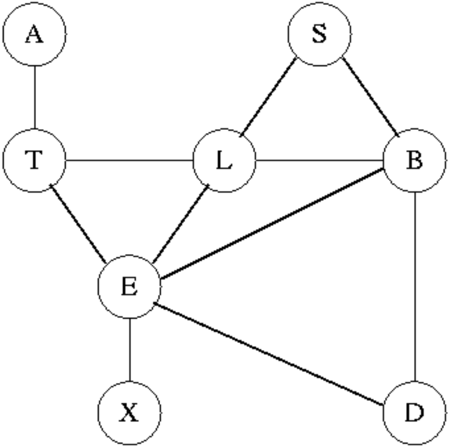
Step 1: Moralize the Graph



We join **T** and **L** because they are parents of **E**.

We join **E** and **B** because they are parents of **D**.

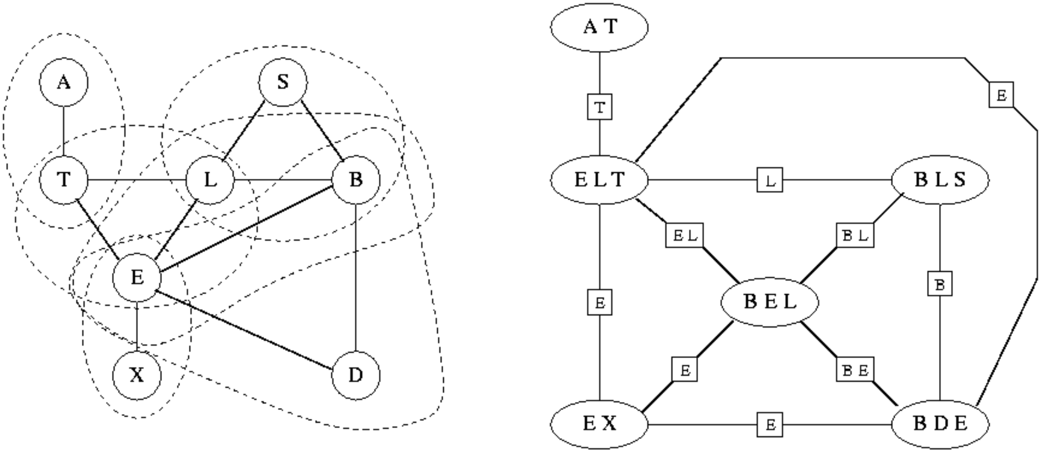
Step 2: Triangulate the Moral Graph



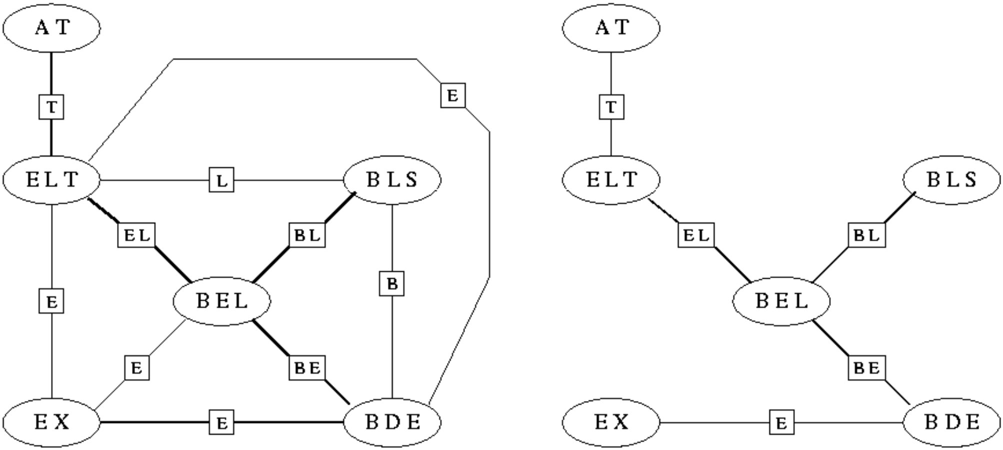
There is a cycle of length four with no shortcuts: **E**, **L**, **S**, **B**.

We have a choice of where to add the shortcut. Either **LB** or **SE** would work.

Step 3: Cliques and Junction Graph



Step 4: Junction Tree



Notice that the running intersection property holds (this is guaranteed by the maximum weight spanning tree and the moralizing and triangulating edges).

Overview

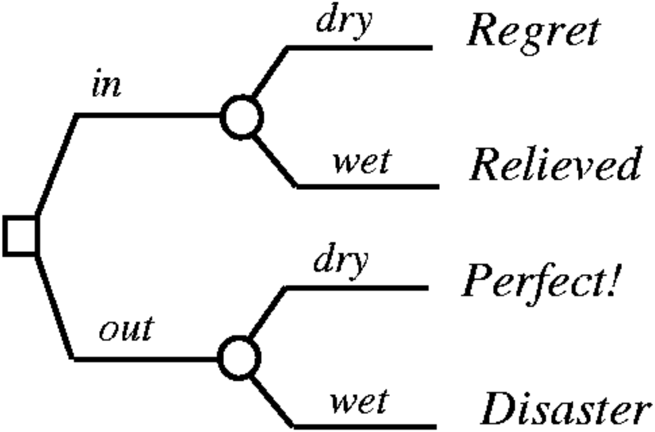
- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- Inference in Bayesian Nets
- Join Trees
- **Decision Making with Bayesian Networks**
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- References and links

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?

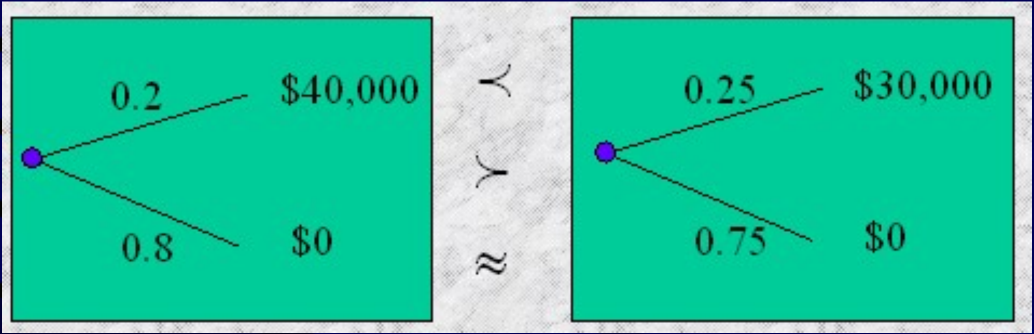


Value Function

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

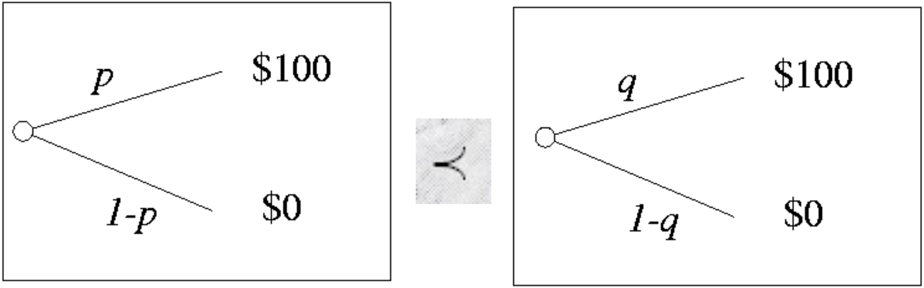
Location?	Weather?	Value
in	dry	$50
in	wet	$60
out	dry	$100
out	wet	$0

Preference for Lotteries



Desired Properties for Preferences over Lotteries

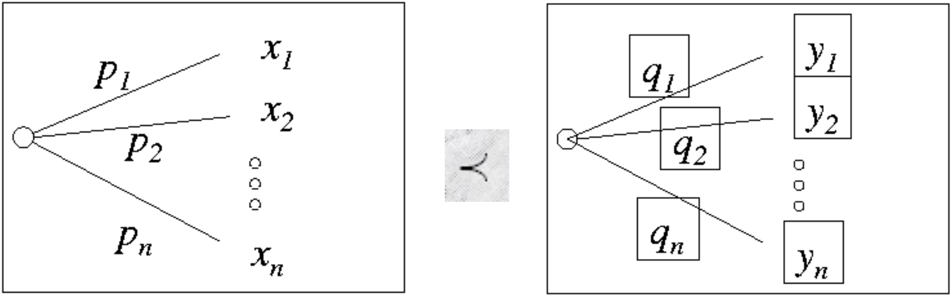
If you prefer \$100 to \$0 and $p < q$ then



(always)

Expected Utility

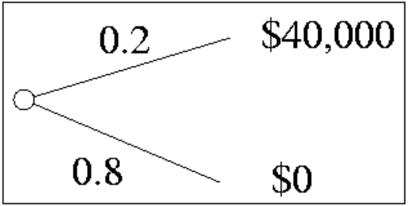
Properties of preference $\Rightarrow$
existence of function U, that satisfies:



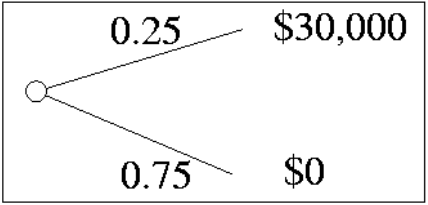
iff

$$\Sigma_i\, p_i\, U(x_i) \quad < \quad \Sigma_i\, q_i\, U(y_i)$$

Are people rational?

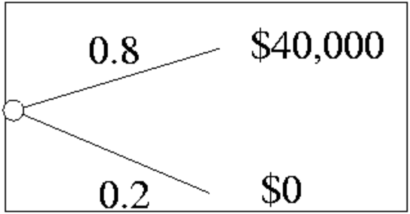


$\succ$

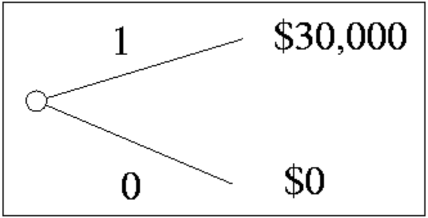


$$0.2 \bullet U(\$40k) \quad > \quad 0.25 \bullet U(\$30k)$$

$$0.8 \bullet U(\$40k) \quad > \quad U(\$30k)$$

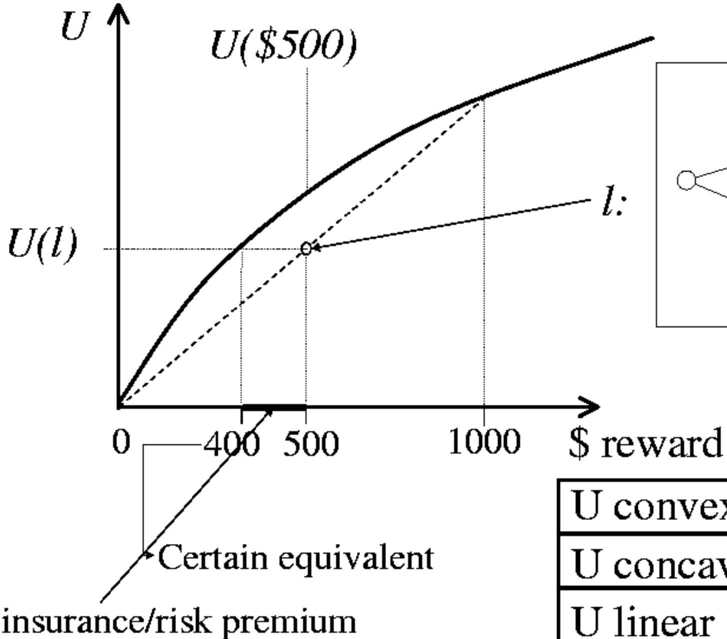


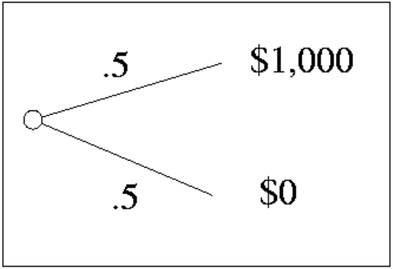
$\prec$



$$0.8 \bullet U(\$40k) \quad < \quad U(\$30k)$$

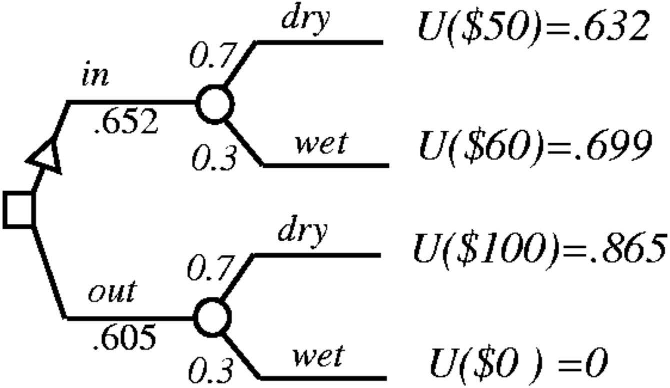
Attitudes towards risk





U convex	risk averse
U concave	risk seeking
U linear	risk neutral

Maximizing Expected Utility



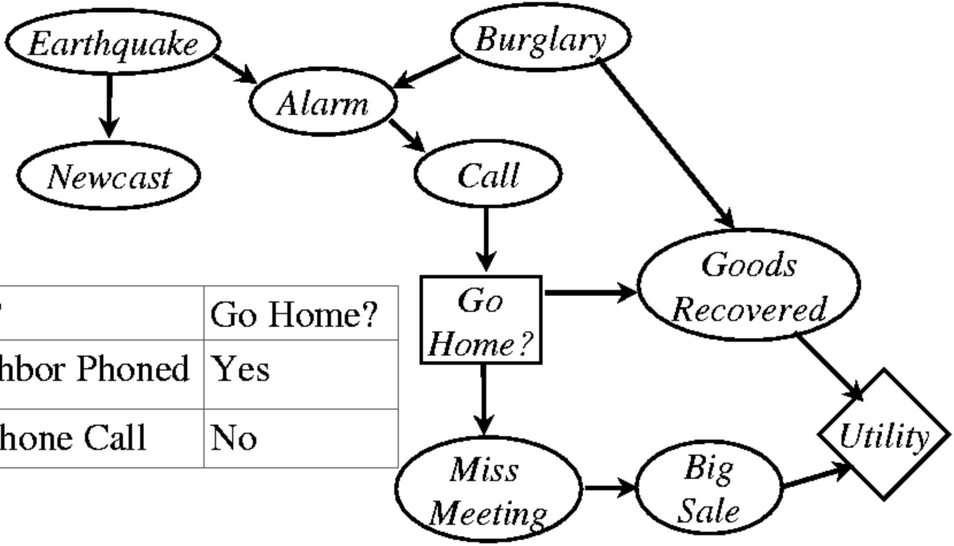
choose the action that maximizes expected utility

$$EU(in) = 0.7 \cdot .632 + 0.3 \cdot .699 = .652$$

$\Longrightarrow$ Choose *in*

$$EU(out) = 0.7 \cdot .865 + 0.3 \cdot 0 = .605$$

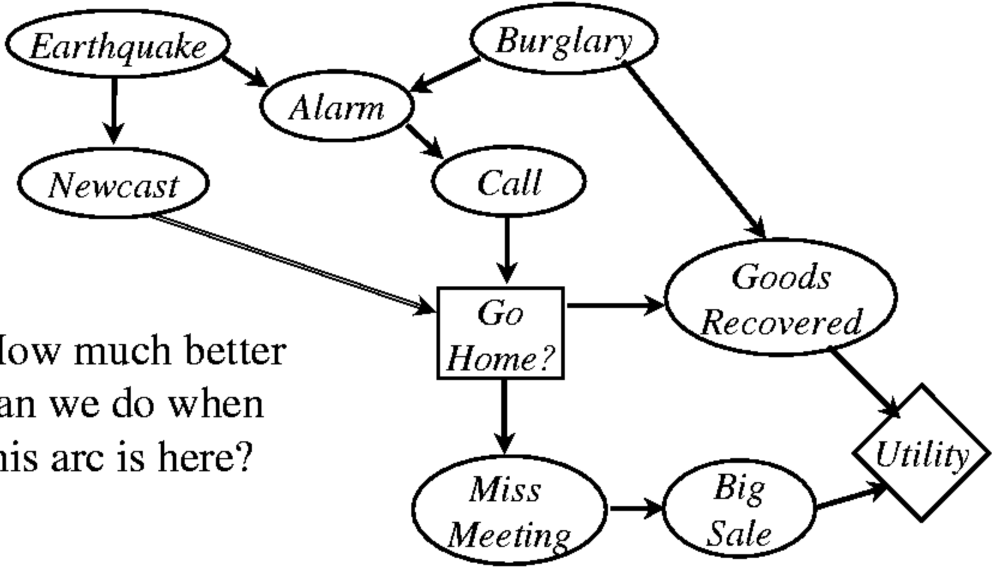
Decision Making with Influence Diagrams



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

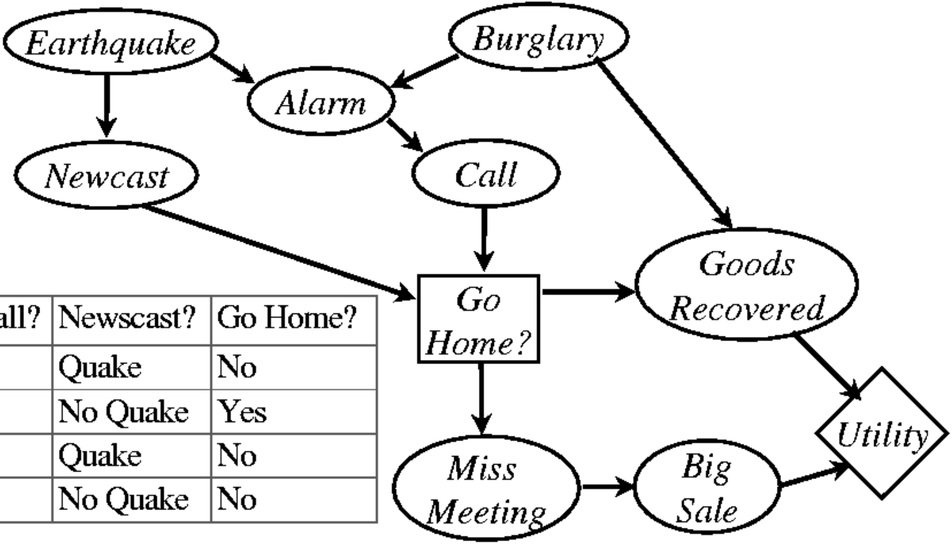
Expected Utility of this policy is 100

Value-of-Information in an Influence Diagram



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

Value-of-Information is the increase in Expected Utility



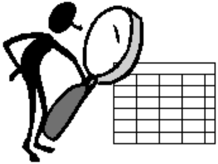
Phonecall?	Newscast?	Go Home?
Yes	Quake	No
Yes	No Quake	Yes
No	Quake	No
No	No Quake	No

Expected Utility of this policy is 112.5

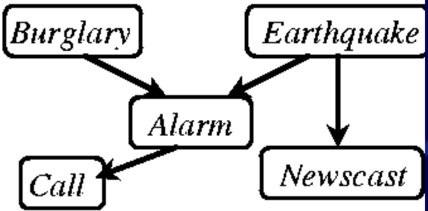
Overview

- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- Inference in Bayesian Nets
- Join Trees
- Decision Making with Bayesian Networks
- **Learning Bayesian Networks from Data**
- Profiling with Bayesian Network
- References and links

The learning task



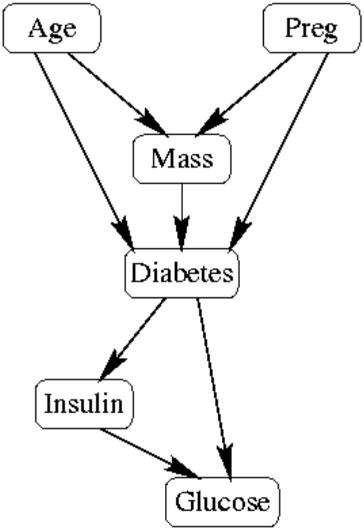
B	E	A	C	N
$\bar{b}$	e	a	c	$\bar{n}$
b	$\bar{e}$	$\bar{a}$	$\bar{c}$	n
		⋮		



Input: training data *Output: BN modeling data*

- Input: fully or partially observable data cases?
- Output: parameters or also structure?

Known structure, Fully Observable



Preg	Glucose	Insulin	Mass	Age	Diabetes
5	121	112	26.2	30	0
10	101	180	32.9	63	0
7	137	0	32.0	39	0
12	100	105	30.0	46	0
9	140	0	32.7	45	1
1	102	0	39.5	42	1
2	99	160	36.6	21	0
2	174	120	44.5	24	1
1	111	0	32.8	45	0
5	117	105	39.1	42	0

Learning Process

- **Discretize the Data**

$$\text{Glucose} < 100 \Rightarrow 0$$

$$100 \leq \text{Glucose} < 120 \Rightarrow 1$$

$$120 \leq \text{Glucose} < 140 \Rightarrow 2$$

$$140 \leq \text{Glucose} \Rightarrow 3$$

- **Count Cases**

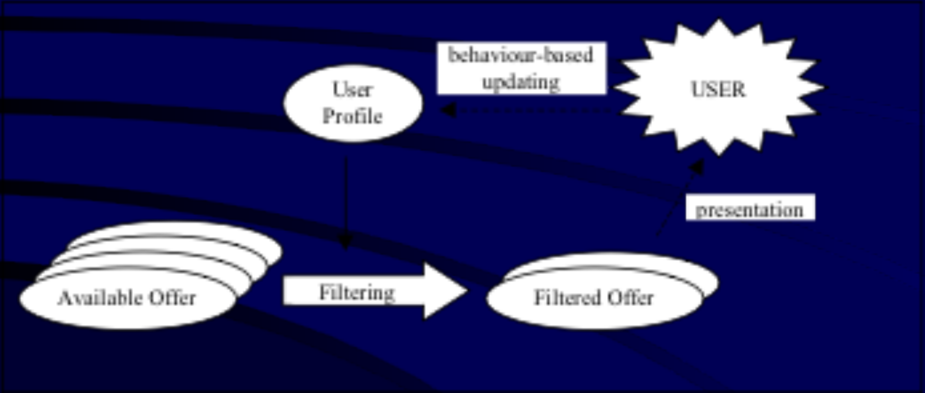
$$P(Mass = 0 | Preg = 1, Age = 2) = \frac{N(Mass = 0, Preg = 1, Age = 2)}{N(Preg = 1, Age = 2)}$$

Read more about Learning BN in:
http://http.cs.berkeley.edu/~murphyk/Bayes/learn.html

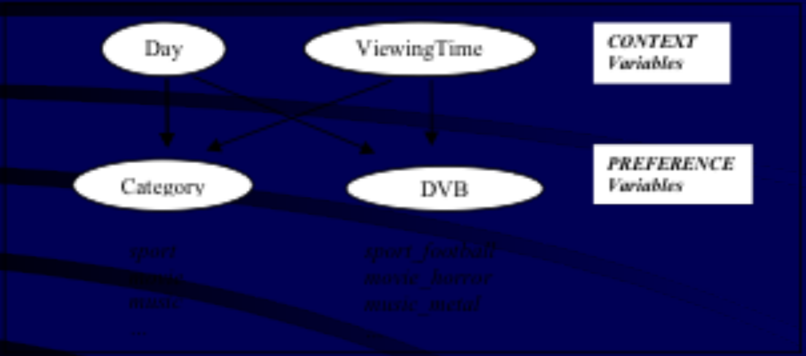
Overview

- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- Inference in Bayesian Nets
- Join Trees
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- **Profiling with Bayesian Network**
- References and links

User Profiling: the problem



The BBN encoding the user preference



- Preference Variables:
  what kind of TV programmes does the user prefer and how much?
- Context Variables:
  in which (temporal) conditions does the user prefer …? 0

BBN based filtering

1) From each item of the input offer extract:

- the classification
- the possible (empty) context

2) For each item compute

Prob (*<classification>* | *<context>*)

3) Items with highest probabilities are the output of the filtering

Example of filtering

The input offer is a set of 3 items:

1. a concert of classical music on Thursday afternoon
2. a football match on Wednesday night
3. a subscription for 10 movies on evening

The probabilities to be computed are:

1. P (MUS = CLASSIC_MUS | Day = Thursday, ViewingTime = afternoon)
2. P (SPO = FOOTBAL_SPO | Day = Wednesday, ViewingTime = night)
3. P (CATEGORY = MOV | ViewingTime = evening)

BBN based updating

- The BBN of a new user is initialised with uniform distributions
- The distributions are updated using a Bayesian learning technique on the basis of user's actual behaviour
- Different user's behaviours -> different *learning weights*:
  1) the user declares their preference
  2) the user watches a specific TV programme
  3) the user searches for specific kind of programmes

Overview

- Probabilities basic rules
- Bayesian Nets
- Conditional Independence
- Motivating Examples
- Inference in Bayesian Nets
- Join Trees
- Decision Making with Bayesian Networks
- Learning Bayesian Networks from Data
- Profiling with Bayesian Network
- **References and links**

Basic References

- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kauffman.
- Oliver, R.M. and Smith, J.Q. (eds.) (1990). Influence Diagrams, Belief Nets, and Decision Analysis, Chichester, Wiley.
- Neapolitan, R.E. (1990). Probabilistic Reasoning in Expert Systems, New York: Wiley.
- Schum, D.A. (1994). The Evidential Foundations of Probabilistic Reasoning, New York: Wiley.
- Jensen, F.V. (1996). An Introduction to Bayesian Networks, New York: Springer.

Algorithm References

- Chang, K.C. and Fung, R. (1995). Symbolic Probabilistic Inference with Both Discrete and Continuous Variables, IEEE SMC, 25(6), 910-916.
- Cooper, G.F. (1990) The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence, 42, 393-405,
- Jensen, F.V, Lauritzen, S.L., and Olesen, K.G. (1990). Bayesian Updating in Causal Probabilistic Networks by Local Computations. Computational Statistics Quarterly, 269-282.
- Lauritzen, S.L. and Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. J. Royal Statistical Society B, 50(2), 157-224.
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kauffman.
- Shachter, R. (1988). Probabilistic Inference and Influence Diagrams. Operations Research, 36(July-August), 589-605.
- Suermondt, H.J. and Cooper, G.F. (1990). Probabilistic inference in multiply connected belief networks using loop cutsets. International Journal of Approximate Reasoning, 4, 283-306.

Key Events in Development of Bayesian Nets

- 1763 Bayes Theorem presented by Rev Thomas Bayes (posthumously) in the Philosophical Transactions of the Royal Society of London
- 19xx Decision trees used to represent decision theory problems
- 19xx Decision analysis originates and uses decision trees to model real world decision problems for computer solution
- 1976 Influence diagrams presented in SRI technical report for DARPA as technique for improving efficiency of analyzing large decision trees
- 1980s Several software packages are developed in the academic environment for the direct solution of influence diagrams
- 1986? Holding of first Uncertainty in Artificial Intelligence Conference motivated by problems in handling uncertainty effectively in rule-based expert systems
- 1986 "Fusion, Propagation, and Structuring in Belief Networks" by Judea Pearl appears in the journal *Artificial Intelligence*
- 1986,1988 Seminal papers on solving decision problems and performing probabilistic inference with influence diagrams by Ross Shachter
- 1988 Seminal text on belief networks by Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference
- 199x Efficient algorithm
- 199x Bayesian nets used in several industrial applications
- 199x First commercially available Bayesian net analysis software available

Software

- Many software packages available
  - See Russell Almond's Home Page
- Netica
  - www.norsys.com
  - Very easy to use
  - Implements learning of probabilities
  - Will soon implement learning of network structure
- Hugin
  - www.hugin.dk
  - Good user interface
  - Implements continuous variables