

CSE458 Bayesian Networks

Lectures 1 & 2

Ann E. Nicholson and Kevin B. Korb

Faculty of Information Technology
Monash University
Clayton, VIC 3168
AUSTRALIA

{annn,korb}@csse.monash.edu.au
HTTP://WWW.CSSE.MONASH.EDU.AU/BAI

Text: *Bayesian Artificial Intelligence*, Kevin B. Korb and Ann E. Nicholson, Chapman & Hall/CRC, 2004.

Bayesian AI Tutorial

Introduction to Bayesian AI

- Reasoning under uncertainty
- Probabilities
- Bayesian philosophy
 - Bayes' Theorem
 - Conditionalization
 - Motivation
- Bayesian decision theory
- How to be an effective Bayesian
- Probabilistic causality
 - Humean causality
 - Prob causality
 - Are Bayesian networks Bayesian?
- Towards a Bayesian AI

Bayesian AI Tutorial

Nicholson & Korb

3

Reasoning under uncertainty

Uncertainty: The quality or state of being not clearly known.

This encompasses most of what we understand about the world — and most of what we would like our AI systems to understand.

Distinguishes *deductive* knowledge (e.g., mathematics) from *inductive* belief (e.g., science).

Sources of uncertainty

- Ignorance
(which side of this coin is up?)
- Complexity
(meteorology)
- Physical randomness
(which side of this coin will land up?)
- Vagueness
(which tribe am I closest to genetically? Picts? Angles? Saxons? Celts?)

Bayesian AI Tutorial

Nicholson & Korb

4

Probability Calculus

Classic approach to reasoning under uncertainty.
(origin: Blaise Pascal and Fermat).

Kolmogorov's Axioms:

1. $P(U) = 1$
2. $\forall X \subseteq U \ P(X) \geq 0$
3. $\forall X, Y \subseteq U$
if $X \cap Y = \emptyset$
then $P(X \vee Y) = P(X) + P(Y)$

Conditional Probability $P(X|Y) = \frac{P(X \wedge Y)}{P(Y)}$

Independence $X \perp\!\!\!\perp Y$ iff $P(X|Y) = P(X)$

Bayesian AI Tutorial

Rev. Thomas Bayes (1702-1761)



Bayesian AI Tutorial

Bayes' Theorem; Conditionalization

— Due to Reverend Thomas Bayes (1764)

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

Conditionalization: $P'(h) = P(h|e)$

Or, read Bayes' theorem as:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Prob of evidence}}$$

Assumptions:

1. Joint priors over $\{h_i\}$ and e exist.
2. Total evidence: e , and only e , is learned.

Bayesian AI Tutorial

Motivation: Breast Cancer

Let $P(h) = 0.01$ (one in 100 women tested have it)
 $P(e|h) = 0.8$ and $P(e|\neg h) = 0.1$
(true and false positive rates). What is $P(h|e)$?

Bayesian AI Tutorial

Motivation: Breast Cancer

Let $P(h) = 0.01$ (one in 100 women tested have it)
 $P(e|h) = 0.8$ and $P(e|\neg h) = 0.1$
(true and false positive rates). What is $P(h|e)$?

Bayes' Theorem yields:

$$\begin{aligned} P(h|e) &= \frac{P(e|h)P(h)}{P(e)} \\ &= \frac{P(e|h)P(h)}{P(e|h)P(h) + P(e|\neg h)P(\neg h)} \\ &= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} \\ &= \frac{0.008}{0.008 + 0.099} \\ &= \frac{0.008}{0.107} \\ &\approx 0.075 \end{aligned}$$

Bayesian AI Tutorial

Motivation

Huge variety of cases where

- Uncertainty dominates considerations
- Getting it right is crucial

Examples and consequences:

- Medicine: death, injury, disease
- Law: false imprisonment, wrongful execution
- Space shuttle: explosion
- Hiring: wasted time and money

Bayesian AI Tutorial

Bayesian Decision Theory

— Frank Ramsey (1926)

Decision making under uncertainty: what action to take (plan to adopt) when future state of the world is not known.

Bayesian answer: Find utility of each possible outcome (action-state pair) and take the action that **maximizes expected utility**.

Example

Action	Rain ($p = .4$)	Shine ($1 - p = .6$)
Take umbrella	30	10
Leave umbrella	-100	50

Expected utilities:

$$E(\text{Take umbrella}) = (30)(.4) + (10)(.6) = 18$$

$$E(\text{Leave umbrella}) = (-100)(.4) + (50)(.6) = -10$$

Bayesian AI Tutorial

Humean Causality

As we shall see, causality and Bayesian networks are intimately related concepts.

David Hume (1737) analyzed A causes B as:

- Whenever A occurs, B occurs
- A and B are contiguous
- A precedes B

This was immediately challenged by Thomas Reid: let A be night and B day; the conditions are satisfied, but neither causes the other.

Leading to: CounterExample \rightarrow new conditions \rightarrow CE ...

Through the next two centuries the “conditions” (sufficiency) account of causality has built up complexity without explanation

Bayesian AI Tutorial

Probabilistic Causality

Salmon (1980): What is this sufficiency nonsense? Either of determinism and indeterminism are *possible* – i.e., it is a contingent fact of the world whether it is deterministic or not.

1. A philosophical analysis of causality should not *presume* determinism.
2. Besides, there is evidence that indeterminism is correct.
3. A probabilistic analysis of causality does predetermine the determinism question, whereas the sufficiency analysis does.

Probabilistic causality

- started by H Reichenbach, IJ Good, P Suppes, W Salmon
- has turned into the theory of Bayesian networks

Bayesian AI Tutorial

Bayesian AI

A Bayesian conception of an AI is:
An autonomous agent which

- Has a utility structure (preferences)
- Can learn about its world and the relation between its actions and future states (probabilities)
- Maximizes its expected utility

The techniques used in learning about the world are (primarily) statistical. . . Hence

Bayesian data mining

Bayesian AI Tutorial

Introduction to Bayesian Networks

- Nodes, structure and probabilities
- Reasoning with BNs
- Understanding BNs

Bayesian AI Tutorial

Bayesian Networks

- Data Structure which represents the dependence between variables.
- Gives concise specification of the joint probability distribution.
- A Bayesian Network is a graph in which the following holds:
 1. A set of random variables makes up the nodes in the network.
 2. A set of directed links or arrows connects pairs of nodes.
 3. Each node has a conditional probability table that *quantifies* the effects the parents have on the node.
 4. Directed, acyclic graph (DAG), i.e. no directed cycles.

Bayesian AI Tutorial

Example: Lung Cancer Diagnosis

A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive XRay would indicate either TB or lung cancer.

Bayesian AI Tutorial

Nodes and Values

Q: What are the nodes to represent and what values can they take?

Nodes can be discrete or continuous; will focus on discrete for now.

- Boolean nodes: represent propositions, taking binary values true (T) and false (F).

Example: *Cancer* node represents proposition “the patient has cancer”.

- Ordered values..

Example: *Pollution* node with values $\{low, medium, high\}$.

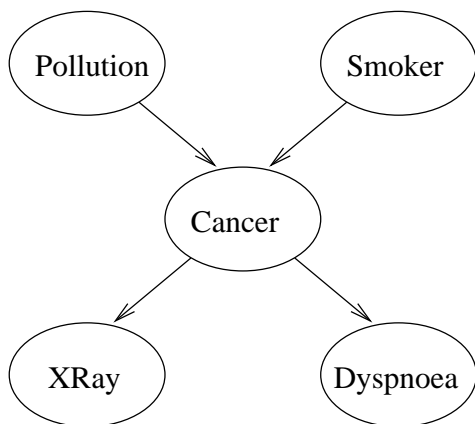
- Integral values.

Example: *Age* node with possible values from 1 to 120.

Lung cancer example: nodes and values

Node name	Type	Values
<i>Pollution</i>	Binary	$\{low, high\}$
<i>Smoker</i>	Boolean	$\{T, F\}$
<i>Cancer</i>	Boolean	$\{T, F\}$
<i>Dyspnoea</i>	Boolean	$\{T, F\}$
<i>XRay</i>	Binary	$\{pos, neg\}$

Lung cancer example: network structure



Note: No explicit representation of other causes of cancer, or other causes of symptoms.

Structure terminology and layout

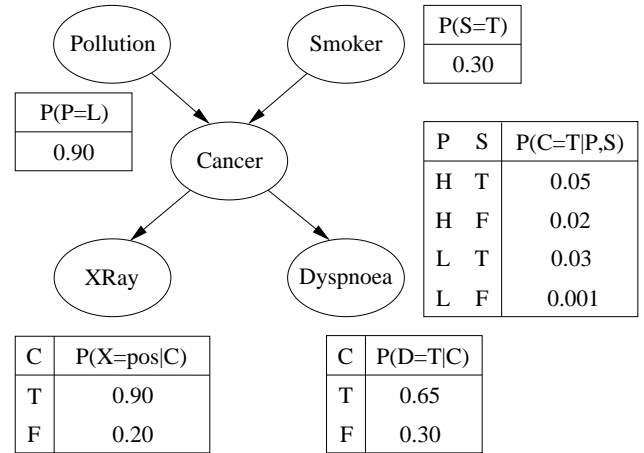
- Family metaphor:
 $Parent \Rightarrow Child$
 $Ancestor \Rightarrow \dots \Rightarrow Descendant$
- Markov Blanket = parents + children + children's parents
- Tree analogy:
 - **root** node: no parents
 - **leaf** node: no children
 - **intermediate** node: non-leaf, non-root
- Layout convention: root notes at top, leaf nodes at bottom, arcs point down the page.

Conditional Probability Tables

Once specified topology, need to specify **conditional probability table (CPT)** for each node.

- Each row contains the conditional probability of each node value for a each possible combination of values of its parent nodes.
- Each row must sum to 1.
- A table for a Boolean var with n Boolean parents contain 2^{n+1} probs.
- A node with no parents has one row (the prior probabilities)

Lung cancer example: CPTs



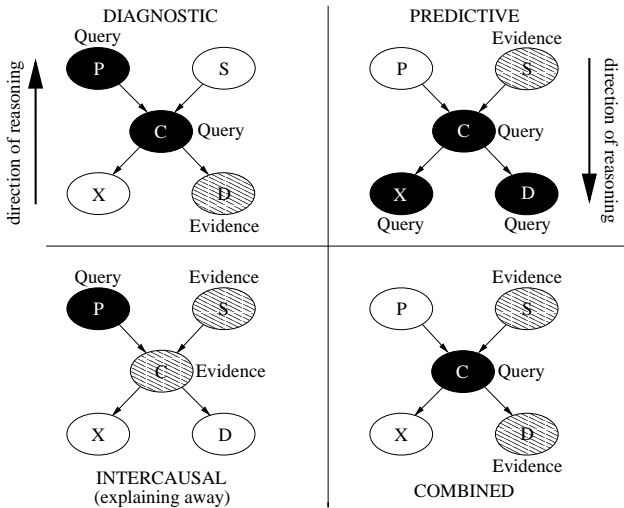
The Markov Property

- Modelling with BNs requires the assumption of the **Markov Property**:
there are no direct dependencies in the system being modeled which are not already explicitly shown via arcs.
- Example: there is no way for smoking to influence dyspnoea except by way of causing cancer.
- BNs which have the Markov property are called Independence-Maps (I-Maps).
- Note: existence of arc does not have to correspond to real dependency in the system being modelled - can be nullified in the CPT.

Reasoning with Bayesian Networks

- Basic task for any probabilistic inference system:
Compute the posterior probability distribution for a set of **query variables**, given new information about some **evidence variables**.
- Also called *conditioning* or *belief updating* or *inference*.

Types of Reasoning



Bayesian AI Tutorial

- Specific evidence: a definite finding that a node X has a particular value, x .

Example: $Smoker=T$

- Negative evidence: a finding that node Y is *not* in state y_1 (but may take any other values).
- “Virtual” or “likelihood” evidence: source of information is not sure about it.

Example:

- $e =$ Radiologist is 80% sure that $Xray=pos$
- Want e.g.:

$$P(Cancer|e) = P(Cancer|Xray, e)P(Xray|e) + P(Cancer|\neg Xray, e)P(\neg Xray|e)$$

- **Jeffrey Conditionalization**

Bayesian AI Tutorial

Reasoning with numbers

- Reasoning with lung cancer example using Netica BN software.

(See Table 2.2 in *Bayesian AI* text.)

Bayesian AI Tutorial

Understanding of Bayesian Networks (Semantics)

- A (more compact) representation of the joint probability distribution.
 - helpful in understanding how to construct network
- Encoding a collection of conditional independence statements.
 - helpful in understanding how to design inference procedures
 - via *Markov property / I-map*:
 - Each conditional independence implied by the graph is present in the probability distribution

Bayesian AI Tutorial

probability distribution

Write $P(X_1 = x_1, \dots, X_n = x_n)$ as $P(x_1, x_2, \dots, x_n)$.

Factorization (chain rule):

$$\begin{aligned}
 P(x_1, x_2, \dots, x_n) &= P(x_1) \times \dots \times P(x_n | x_1, \dots, x_{n-1}) \\
 &= \prod_i P(x_i | x_1, \dots, x_{i-1})
 \end{aligned}$$

Since BN structure implies that the value of a particular node is conditional *only* on the values of its parent nodes, this reduces to

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i | Parents(X_i))$$

provided $Parents(X_i) \subseteq \{x_1, \dots, x_{i-1}\}$.

$$\begin{aligned}
 P(X = pos \wedge D = T \wedge C = T \wedge P = lo \wedge S = F) \\
 &= P(X = pos | D = T, C = T, P = lo, S = F) \\
 &\quad \times P(D = T | C = T, P = lo, S = F) \\
 &\quad \times P(C = T | P = lo, S = F) P(P = lo | S = F) P(S = F) \\
 &= P(X = pos | C = T) P(D = T | C = T) \\
 &\quad \times P(C = T | P = lo, S = F) P(P = lo) P(S = F)
 \end{aligned}$$

Pearl's Network Construction Algorithm

1. Choose the set of relevant variables $\{X_i\}$ that describe the domain.
2. Choose an ordering for the variables, $\langle X_1, \dots, X_n \rangle$.
3. While there are variables left:
 - (a) Add the next variable X_i to the network.
 - (b) Add arcs to the X_i node from some minimal set of nodes already in the net, $Parents(X_i)$, such that the following conditional independence property is satisfied:

$$P(X_i | X'_1, \dots, X'_m) = P(X_i | Parents(X_i))$$

where X'_1, \dots, X'_m are all the variables preceding X_i , including $Parents(X_i)$.

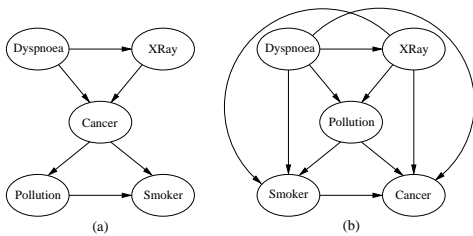
- (c) Define the CPT for X_i .

Compactness and Node Ordering

- Compactness of BN depends upon sparseness of the system.
- The best order to add nodes is to add the “root causes” first, then the variable they influence, so on until “leaves” reached.

→ Causal structure

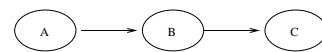
- Alternative structures using different orderings (a) $\langle D, X, C, P, S \rangle$ (b) $\langle D, X, P, S, C \rangle$.



1. These BNs still represent same joint distribution.
2. Structure (b) requires as many probabilities as the full joint distribution! See below for *why*.

Conditional Independence: Causal Chains

Causal chains give rise to conditional independence:

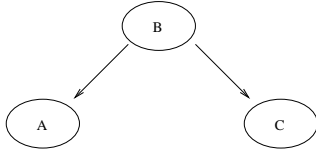


$$P(C | A \wedge B) = P(C | B)$$

Example: “smoking causes cancer which causes dyspnoea”

Conditional Independence: Common Causes

Common causes (or ancestors) also give rise to conditional independence:

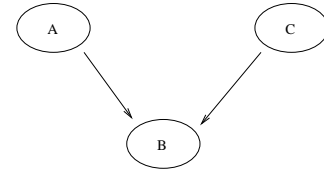


$$P(C|A \wedge B) = P(C|B) \equiv A \perp\!\!\!\perp C | B$$

Example: cancer is a common cause of the two symptoms, a positive XRay result and dyspnoea.

Conditional Dependence: Common Effects

Common effects (or their descendants) give rise to conditional *dependence*:



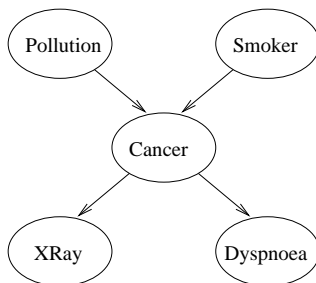
$$P(A|C \wedge B) \neq P(A)P(C) \equiv \neg(A \perp\!\!\!\perp C | B)$$

Example: Cancer is a common effect of pollution and smoking.

Given lung cancer, smoking “explains away” pollution.

D-separation

- Graphical criterion of conditional independence.
 $X \perp\!\!\!\perp Y | Z$
- We can determine whether a set of nodes X is independent of another set Y , given a set of evidence nodes E , via the Markov property:
 $X \perp\!\!\!\perp Y | E \rightarrow X \perp\!\!\!\perp Y | E$.
- Example



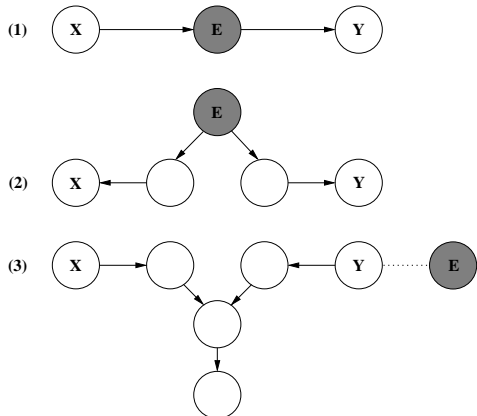
D-separation

How to determine d-separation, $X \perp\!\!\!\perp Y | E$:

- If every undirected path from a node in X to a node in Y is *d-separated* by E , then X and Y are *conditionally independent* given E .
- A set of nodes E *d-separates* two sets of nodes X and Y if every undirected path from a node in X to a node in Y is *blocked* given E .
- A path is *blocked* given a set of nodes E if there is a node Z on the path for which one of three conditions holds:
 - Z is in E and Z has one arrow on the path leading in and one arrow out (chain).
 - Z is in E and Z has both path arrows leading out (common cause).
 - Neither Z nor any descendant of Z is in E , and both path arrows lead in to Z (common effect).

D-separation (cont'd)

- Evidence nodes **E** shown shaded.



Bayesian AI Tutorial

Causal Ordering

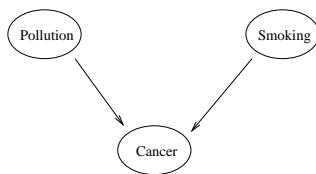
Why does variable order affect network density?

Because

- Using the causal order allows direct representation of conditional independencies
- Violating causal order requires new arcs to re-establish conditional independencies

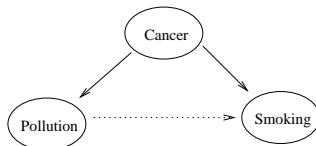
Bayesian AI Tutorial

Causal Ordering (cont'd)



Pollution and *Smoking* are marginally independent.

Given the ordering: Cancer, Pollution, Smoking:



Marginal independence of *Pollution* and *Smoking* must be re-established by adding $Pollution \rightarrow Smoking$ or $Smoking \leftarrow Pollution$

Bayesian AI Tutorial

Bayesian Networks: Summary

- Bayes' rule allows unknown probabilities to be computed from known ones.
- Conditional independence (due to causal relationships) allows efficient updating
- BNs are a natural way to represent conditional independence info.
 - links between nodes: qualitative aspects;
 - conditional probability tables: quantitative aspects.
- Probabilistic inference: compute the probability distribution for query variables, given evidence variables
- BN Inference is very flexible: can enter evidence about any node and update beliefs in any other nodes.

Bayesian AI Tutorial

Inference Algorithms: Overview

- Exact inference
 - Trees and polytrees:
 - * message-passing algorithm
 - Multiply-connected networks:
- Approximate Inference
 - Large, complex networks:
 - * Stochastic Simulation
 - * Other approximation methods
- In the general case, both exact and approximate inference are computationally complex (“NP-hard”).
- Causal inference

Inference in chains

Two node network $X \rightarrow Y$.

- Evidence $X = x$, then $Bel(Y) = P(Y|X = x)$ straight from CPT.
- Evidence $Y = y$

$$\begin{aligned}
 Bel(X = x) &= P(X = x|Y = y) \\
 &= \frac{P(Y = y|X = x)P(X = x)}{P(y)} \\
 &= \alpha P(x)\lambda(x)
 \end{aligned}$$

where

$$\alpha = \frac{1}{P(Y = y)}$$

$P(x)$ is the prior, and $\lambda(x) = P(Y = y|X = x)$ is the **likelihood**.

Since $\sum_i Bel(Y = y_i) = 1$, we can compute α as a **normalizing constant**.

Example: $Flu \rightarrow HighTemp$

Suppose $P(Flu = T) = 0.05$,
 $P(HighTemp = T|Flu = T) = 0.9$,
 $P(HighTemp = T|Flu = F) = 0.2$.

Given evidence $HighTemp = T$, then

$$\begin{aligned}
 Bel(Flu = T) &= \alpha P(Flu = T)\lambda(Flu = T) \\
 &= \alpha \times 0.05 \times 0.9 = \alpha 0.045 \\
 Bel(Flu = F) &= \alpha P(Flu = F)\lambda(Flu = F) \\
 &= \alpha \times 0.95 \times 0.2 = \alpha 0.19
 \end{aligned}$$

We can compute α via

$$Bel(Flu = T) + Bel(Flu = F) = 1 = \alpha 0.045 + \alpha 0.19$$

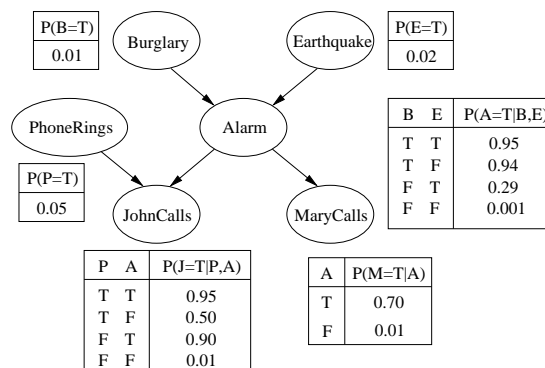
giving

$$\alpha = \frac{1}{0.19 + 0.045}$$

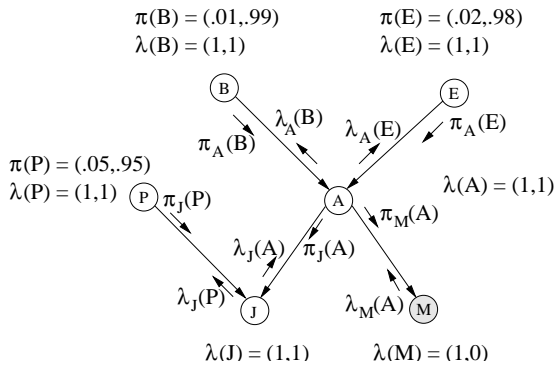
This allows us to finish the belief update:

$$\begin{aligned}
 Bel(Flu = T) &= \frac{0.045}{0.19 + 0.045} = 0.8085 \\
 Bel(Flu = F) &= \frac{0.19}{0.19 + 0.045} = 0.1915
 \end{aligned}$$

Earthquake example

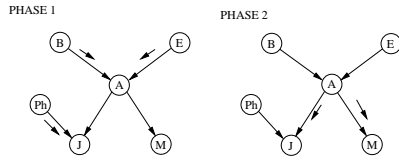


Inference in polytrees: message passing

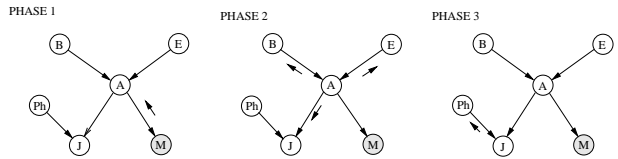


Message propagation

PROPAGATION, NO EVIDENCE



PROPAGATION, EVIDENCE for node M

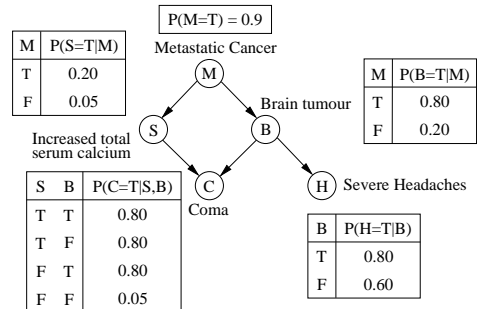


Message-passing algorithm: features

- All computations are local \Rightarrow efficient
- Requires summation over all joint instantiations of parent nodes \Rightarrow exponential in no. of parents.
- No. of propagation steps depends on length of longest path

Inference in multiply connected networks

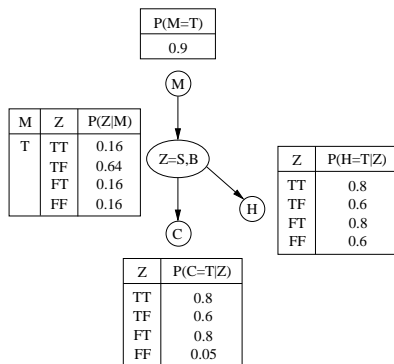
- Networks where two nodes are connected by more than one path
 - Two or more possible causes which share a common ancestor
 - One variable can influence another through more than one causal mechanism
- Example: Cancer network



- Message passing doesn't work - evidence gets "counted twice"

Clustering methods

- Transform network into a probabilistically equivalent polytree by merging (clustering) offending nodes
- Cancer example: new node Z combining B and C



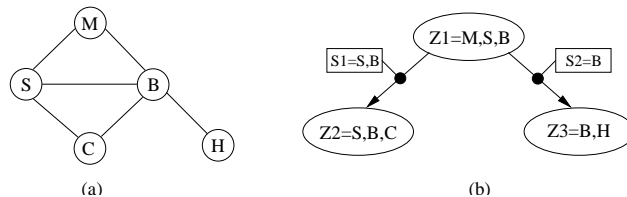
$$P(z|a) = P(b, c|a) = P(b|a)P(c|a)$$

$$P(e|z) = P(e|b, c) = P(e|c)$$

$$P(d|z) = P(d|b, c)$$

Jensen join-tree method

- Jensen Join-tree (Jensen, 1996) version the current most efficient algorithm in this class (e.g. used in Hugin, Netica).



Jensen join-tree method (cont.)

Network evaluation done in two stages

- Compile into join-tree
 - May be slow
 - May require too much memory if original network is highly connected
- Do belief updating in join-tree (usually fast)

Caveat: clustered nodes have increased complexity; updates may be computationally complex

Approximate inference with stochastic simulation

- Use the network to generate a large number of cases that are consistent with the network distribution.
- Evaluation may not converge to exact values (in reasonable time).
- Usually converges to close to exact solution quickly if the evidence is not too unlikely.
- Performs better when evidence is nearer to root nodes, however in real domains, evidence tends to be near leaves (Nicholson&Jitnah, 1998)

Causal modeling

We should like to do causal modeling with our Bayesian networks.

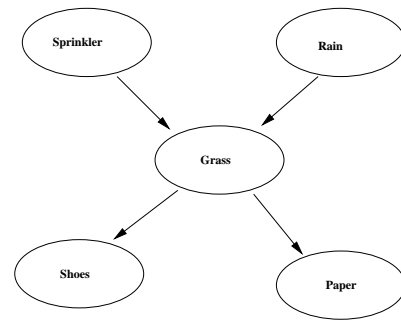
Prerequisite: arcs are truly causal (hence, nodes are properly ordered).

Reasoning about real or hypothetical interventions:

- what if we upgrade quality in manufacturing?
- what if we treat the patient with X, Y, Z?

For planning, control, prediction.

Common practice appears to be: let observation stand for intervention.



If we observe that the lawn is wet:

- We can infer in any direction; everything updates
- We get, e.g., “explaining away” between causes

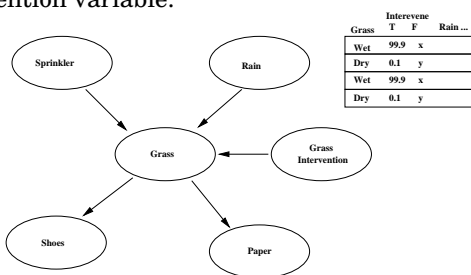
What happens if we *intervene* in a causal process?

Spirtes, et al., (1993), Pearl (2000) answer: cut links to parents and *then* update.

- No explaining away; parents are then unaffected
- Downstream updating is as normal

Causal inference

We prefer (conceptually) to augment the graph with an intervention variable:



- Simplistically, parent connections are severed
- With full generality, X acquires a new parent D_X
 - Allows any degree of control for intervention
 - Allows any kind of interaction with existing parents
 - Bayesian update algorithms unaffected

Inference: Summary

- Probabilistic inference: compute the probability distribution for query variables, given evidence variables
- Causal inference: compute the probability distribution for query variables, given intervention
- BN Inference is very flexible: can enter evidence about any node and update beliefs in any other nodes.
- The speed of inference in practice depends on the structure of the network: how many loops; numbers of parents; location of evidence and query nodes.
- BNs can be used to model causal intervention.

Extensions to Bayesian Networks

- For decision making: decision networks
- For reasoning about changes over time: dynamic Bayesian networks

Bayesian AI Tutorial

Making Decisions

- Bayesian networks can be extended to support decision making.
- **Preferences** between different outcomes of various plans.
 - Utility theory
- **Decision theory** = Utility theory + Probability theory.

Bayesian AI Tutorial

Nicholson & Korb

59

Expected Utility

$$EU(A|E) = \sum_i P(O_i|E, A) U(O_i|A) \quad (1)$$

- E = available evidence,
- A = a nondeterministic action
- O_i = possible outcome state
- U = utility

Bayesian AI Tutorial

Nicholson & Korb

60

Decision Networks

A Decision network represents information about

- the agent's current state
- its possible actions
- the state that will result from the agent's action
- the utility of that state

Also called, *Influence Diagrams* (Howard&Matheson, 1981).

Bayesian AI Tutorial

Type of Nodes

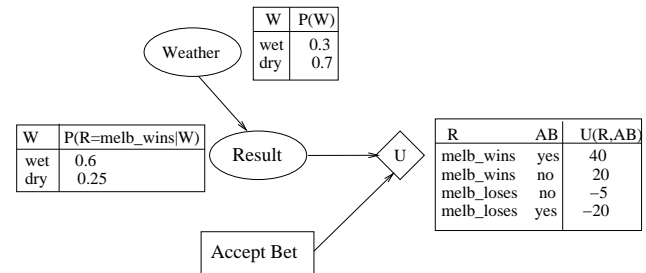
Chance nodes: (ovals) represent random variables (same as Bayesian networks). Has an associated CPT. Parents can be decision nodes and other chance nodes.

Decision nodes: (rectangles) represent points where the decision maker has a choice of actions.

Utility nodes: (diamonds) represent the agent's utility function (also called **value nodes** in the literature). Parents are variables describing the outcome state that directly affect utility. Has an associated table representing multi-attribute utility function.

Example: Football Team

Clare's football team, Melbourne, is going to play her friend John's team, Carlton. John offers Clare a friendly bet: whoever's team loses will buy the wine next time they go out for dinner. They never spend more than \$15 on wine when they eat out. When deciding whether to accept this bet, Clare will have to assess her team's chances of winning (which will vary according to the weather on the day). She also knows that she will be happy if her team wins and miserable if her team loses, regardless of the bet.



Evaluating Decision Networks: Algorithm

1. Add any available evidence.
2. For each action value in the decision node:
 - (a) Set the decision node to that value;
 - (b) Calculate the posterior probabilities for the parent nodes of the utility node, as for Bayesian networks, using a standard inference algorithm;
 - (c) Calculate the resulting expected utility for the action.
3. Return the action with the highest expected utility.

Evaluating Decision Networks: Example

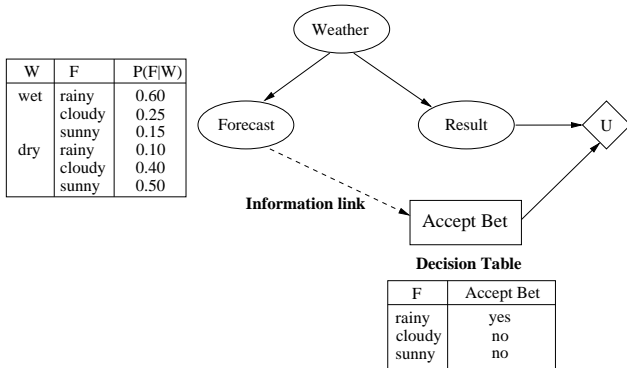
$$\begin{aligned}
 P(R = melb_wins) &= P(W = w)P(R = melb_wins|W = w) \\
 &= +P(W = d)P(R = melb_wins|W = d)
 \end{aligned}$$

$$\begin{aligned}
 EU(AB = yes) &= P(R = wins)U(R = wins|AB = yes) \\
 &+ P(R = loses)U(R = loses|AB = yes) \\
 &= (0.3 \times 0.6 + 0.7 \times 0.25)40 \\
 &+ (0.3 \times 0.4 + 0.7 \times 0.75) - 20 \\
 &= 0.355 \times 40 + 0.645 \times -20 = 14.2 - 12.9 \\
 &= 1.3
 \end{aligned}$$

$$\begin{aligned}
 EU(AB = no) &= P(R = wins)U(R = wins|AB = no) \\
 &+ P(R = loses)U(R = loses|AB = no) \\
 &= (0.3 \times 0.6 + 0.7 \times 0.25)20 \\
 &+ (0.3 \times 0.4 + 0.7 \times 0.75) - 5 \\
 &= 0.355 \times 20 + 0.645 \times -5 = 7.1 - 3.225 \\
 &= 3.875
 \end{aligned}$$

Information Links

- Indicate when a chance node needs to be observed before a decision is made.



Bayesian AI Tutorial

Decision Table Algorithm

- Add any available evidence.
- For each combination of values of the parents of the decision node:
 - For each action value in the decision node:
 - Set the decision node to that value;
 - Calculate the posterior probabilities for the parent nodes of the utility node, as for Bayesian networks, using a standard inference algorithm;
 - Calculate the resulting expected utility for the action.
 - Record the action with the highest expected utility in the decision table.
- Return the decision table.

Bayesian AI Tutorial

Nicholson & Korb

67

Fever problem description

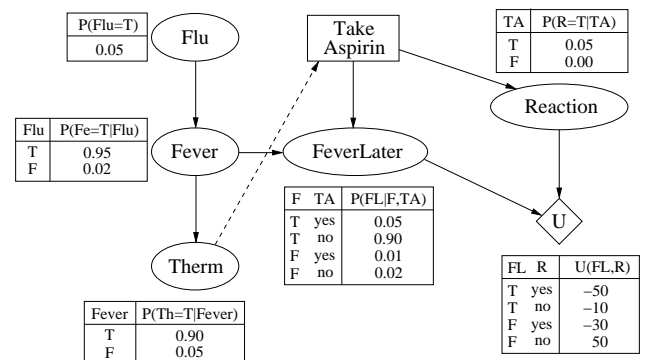
Suppose that you know that a fever can be caused by the flu. You can use a thermometer, which is fairly reliable, to test whether or not you have a fever. Suppose you also know that if you take aspirin it will almost certainly lower a fever to normal. Some people (about 5% of the population) have a negative reaction to aspirin. You'll be happy to get rid of your fever, as long as you don't suffer an adverse reaction if you take aspirin.

Bayesian AI Tutorial

Nicholson & Korb

68

Fever decision network



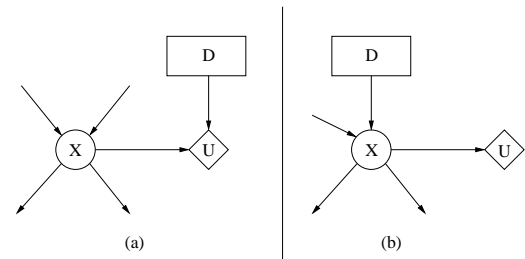
Bayesian AI Tutorial

Fever decision table

Ev.	$Bel(FL=T)$	$EU(TA=yes)$	$EU(TA=no)$	Dec.
None	0.046	45.27	45.29	no
$Th=F$	0.525	45.41	48.41	no
$Th=T$	0.273	44.1	19.13	yes
$Th=T$ & $Re=T$	0.273	-30.32	0	no

Bayesian AI Tutorial

Types of actions

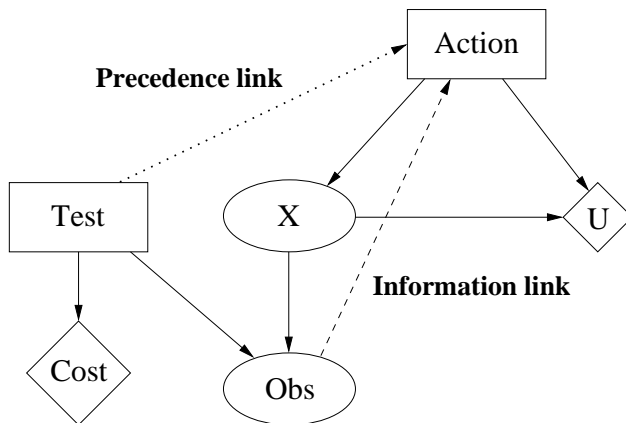


(a) Non-intervening and (b) Intervening

Bayesian AI Tutorial

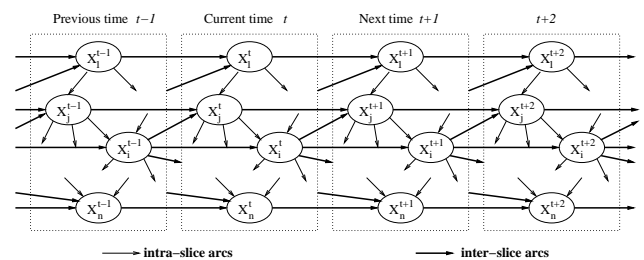
Sequential decision making

- Precedence links used to show temporal ordering.
- Network for a test-action decision sequence



Bayesian AI Tutorial

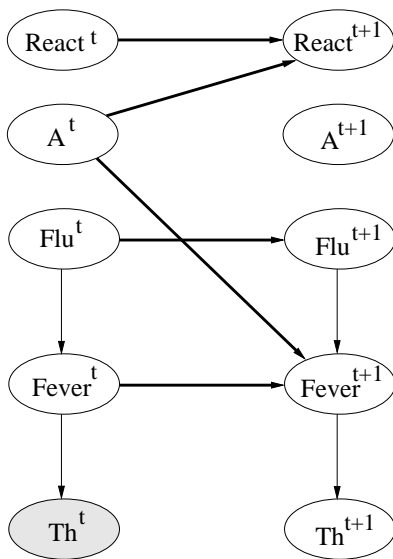
Dynamic Belief Networks



- One node for each variable for each time step.
- **Intra-slice arcs** $X_i^T \rightarrow X_j^T$
- **Inter-slice (temporal) arcs**
 1. $X_i^T \rightarrow X_i^{T+1}$
 2. $X_i^T \rightarrow X_j^{T+1}$

Bayesian AI Tutorial

Fever DBN



Bayesian AI Tutorial

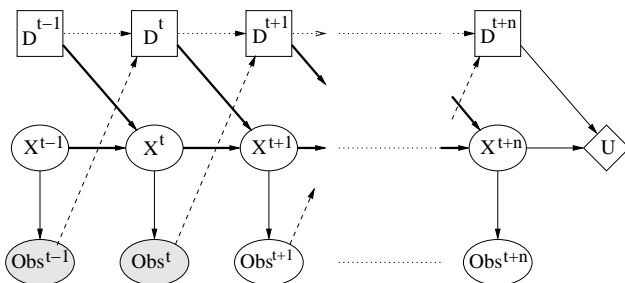
DBN reasoning

- Can calculate distributions for S_{t+1} and further: **probabilistic projection**.
- Reasoning can be done using standard BN updating algorithms
- This type of DBN gets very large, very quickly.
- Usually only keep two time slices of the network.

Bayesian AI Tutorial

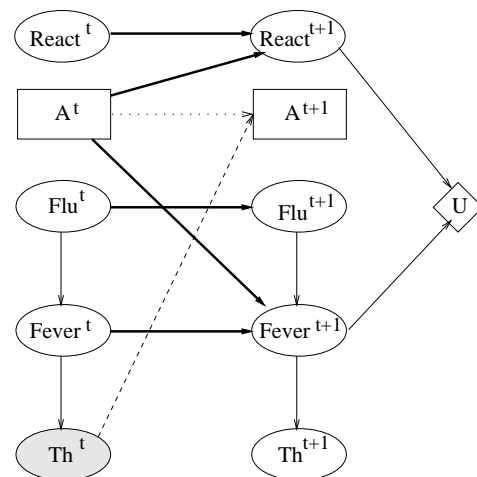
Dynamic Decision Network

- Similarly, Decision Networks can be extended to include temporal aspects.
- Sequence of decisions taken = Plan.



Bayesian AI Tutorial

Fever DDN



Bayesian AI Tutorial

Uses of Bayesian Networks

1. Calculating the belief in query variables given values for evidence variables (above).
2. Predicting values in dependent variables given values for independent variables.
3. Modeling causal interventions.
4. Decision making based on probabilities in the network and on the agent's utilities (Influence Diagrams [Howard and Matheson 1981]).
5. Deciding which additional evidence should be observed in order to gain useful information.
6. *Sensitivity analysis* to test impact of changes in probabilities or utilities on decisions.

Extensions: Summary

- BNs can be extended with decision nodes and utility nodes to support decision making: *Decision Networks* or *Influence Diagrams*.
- BNs and decision networks can be extended to allow explicit reasoning about changes over time.