Probabilistic Reasoning

Chapter 14[Russel]

"…probability theory is more fundamentally concerned with the structure of reasoning and causation than with numbers."

Glenn Shafer and Judea Pearl
Introduction to Readings in Uncertain Reasoning,
Morgan Kaufmann, 1990

Outline

- Syntax
- Semantics

Bayesian networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

Syntax:

a set of nodes, one per variable

a directed, acyclic graph (link ≈ "directly influences")

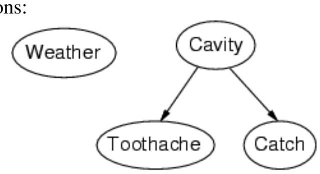
a conditional distribution for each node given its parents:

$$\mathbf{P}(X_i \mid Parents(X_i))$$

In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example

Topology of network encodes conditional independence assertions:



Weather is independent of the other variables

Toothache and *Catch* are conditionally independent given *Cavity*

Example

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes.

John and Mary have promised to call you at work when they hear the alarm John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too. Mary, on the other hand, likes rather loud music and sometimes misses the alarm altogether

Is there a burglar?

Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

Network topology reflects "causal" knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Bayesian Networks

A Bayesian network specifies a joint distribution in a structured form

Represent dependence/independence via a directed graph

Nodes = random variables and Edges = direct dependence

Structure of the graph ⇔ Conditional independence relations

In general,

$$p(X_1, X_2, ...X_N) = \Pi\ p(X_i \mid \text{parents}(X_i))$$

The full joint distribution

The graph-structured approximation

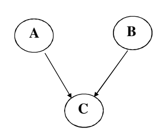
Requires that graph is acyclic (no directed cycles)

2 components to a Bayesian network

- The graph structure (conditional independence assumptions)
- The numerical probabilities (for each variable given its parents)

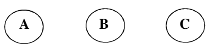
Example of a simple Bayesian network

$p(A,B,C) = p(C|A,B)p(A)p(B)$ ⟷



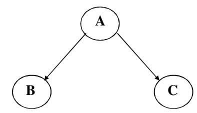
- Probability model has simple factored form
- Directed edges => direct dependence
- Absence of an edge => conditional independence
- Also known as belief networks, graphical models, causal networks
- Other formulations, e.g., undirected graphical models

Examples of 3-way Bayesian Networks



Marginal Independence:
p(A,B,C) = p(A) p(B) p(C)

Examples of 3-way Bayesian Networks

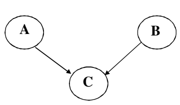


Conditionally independent effects:
p(A,B,C) = p(B|A)p(C|A)p(A)

B and C are conditionally independent
Given A

e.g., A is a disease, and we model
B and C as conditionally independent
symptoms given A

Examples of 3-way Bayesian Networks

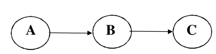


Independent Causes:
p(A,B,C) = p(C|A,B)p(A)p(B)

“Explaining away” effect:
Given C, observing A makes B less likely
e.g., earthquake/burglary/alarm example

A and B are (marginally) independent
but become dependent once C is known

Examples of 3-way Bayesian Networks



Markov dependence:
p(A,B,C) = p(C|B) p(B|A)p(A)

Example

Consider the following 5 binary variables:

- B = a burglary occurs at your house
- E = an earthquake occurs at your house
- A = the alarm goes off
- J = John calls to report the alarm
- M = Mary calls to report the alarm

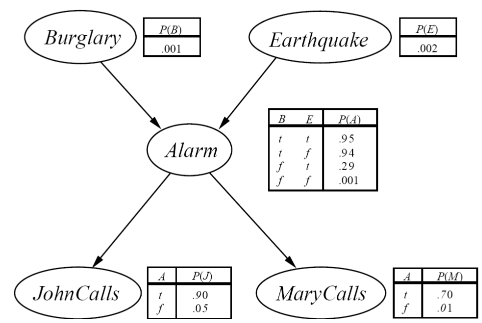
What is P(B | M, J) ? (for example)

We can use the full joint distribution to answer this question

Requires $2^5 = 32$ probabilities

Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

The Desired Bayesian Network



the probability that thr: alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call

$P(j \wedge m \wedge a \wedge \sim b \wedge \sim e)$

$= P(j|a)P(m|a)P(a| \sim b \wedge \sim e)P(\sim b)P(\sim e)$

$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 =$ 0.0006'2.