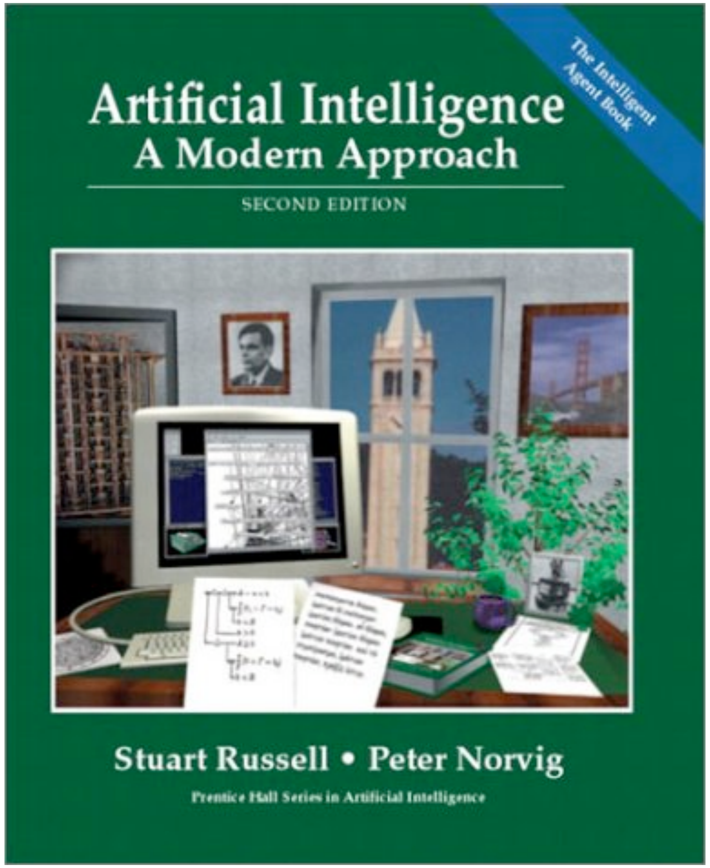


Bayesian Networks

Chapter 14

Section 1 – 2

Updating the Belief State

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Let D now observe toothache with probability 0.8 (e.g., “the patient says so”)
- How should D update its belief state?

Updating the Belief State

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Let E be the evidence such that $P(toothache|E) = 0.8$
- We want to compute $P(c \wedge t \wedge pc|E) = P(c \wedge pc|t,E)\ P(t|E)$
- Since E is not directly related to the cavity or the probe catch, we consider that c and pc are independent of E given t, hence: $P(c \wedge pc|t,E) = P(c \wedge pc|t)$

Updating the Belief State

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	~~0.108~~ 0.432	~~0.012~~ 0.048	~~0.072~~ 0.018	~~0.008~~ 0.002
¬Cavity	~~0.016~~ 0.064	~~0.064~~ 0.256	~~0.144~~ 0.036	~~0.576~~ 0.144

To get these 4 probabilities we normalize their sum to 0.8

To get these 4 probabilities we normalize their sum to 0.2

- Let E be the evidence such that P(Toothache|E) = 0.8
- t∧pc|E) = P(c∧pc|t,E) P(t|E)
- Since E is not directly related to the cavity or the probe catch, we ... e independent of E given t, he ... t)

Issues

- If a state is described by n propositions, then a belief space contains 2^n states (possibly, some have probability 0)
- → Modeling difficulty: many numbers must be entered in the first place
- → Computational issue: memory size and time

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Toothache and pcatch are independent given cavity (or ¬cavity), but this relation is hidden in the numbers ! [Verify this]
- Bayesian networks explicitly represent independence among propositions to reduce the number of probabilities defining a belief state

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:

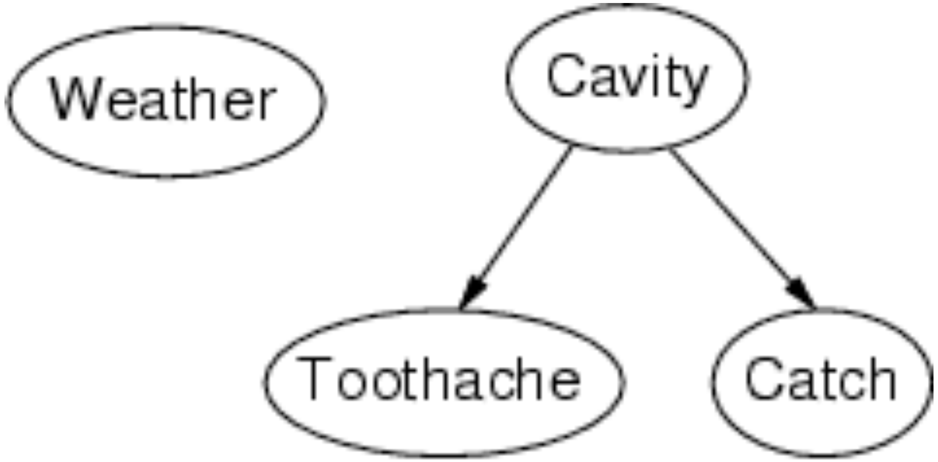
$$\mathbf{P}(X_i \mid Parents(X_i))$$

- 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

Following slides by Hwee Tou Ng (Singapore)

Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

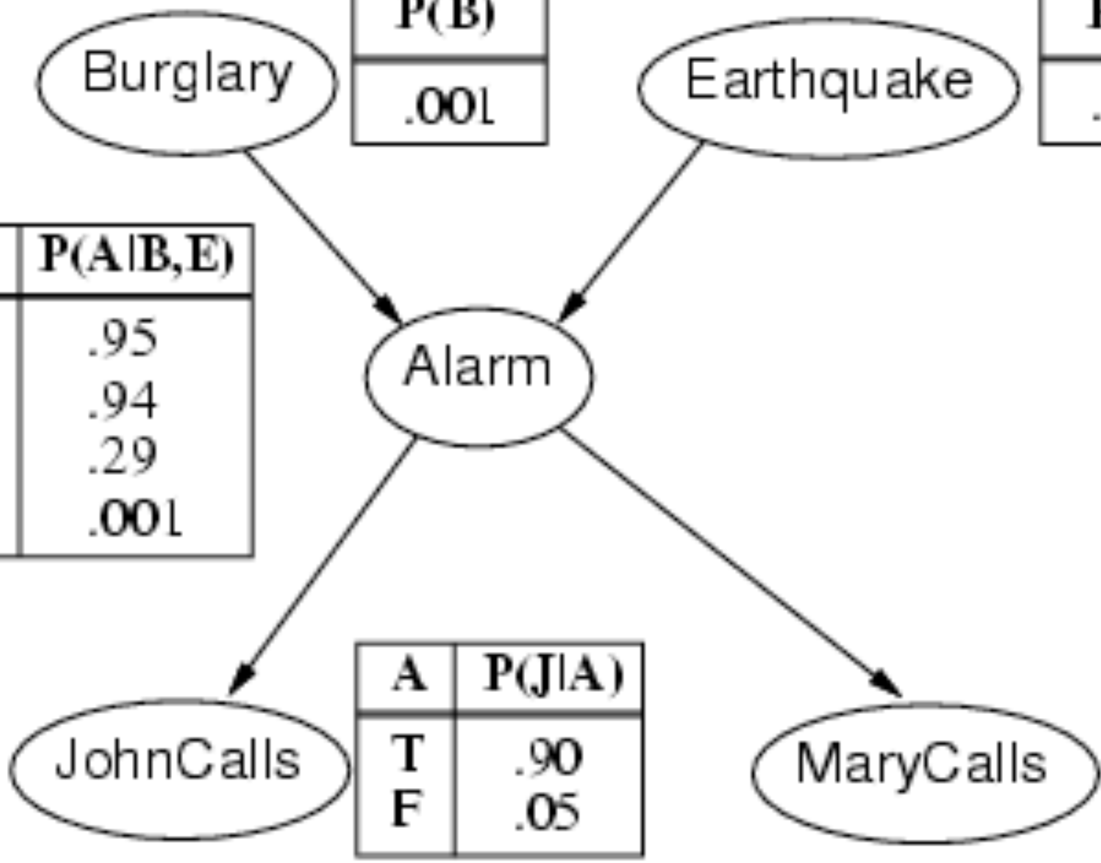
Example

- 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. 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

Example contd.



P(B)
.001

P(E)
.002

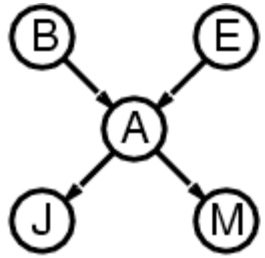
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

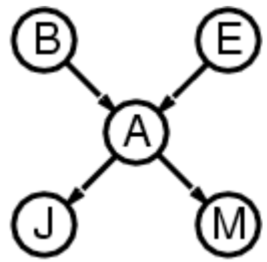


- Each row requires one number p for X_i = *true* (the number for X_i = *false* is just $1-p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct