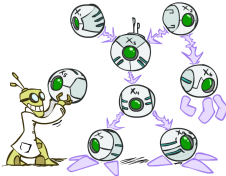


CSE 473: Artificial Intelligence

Bayes' Nets

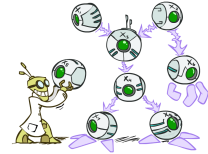


Dieter Fox

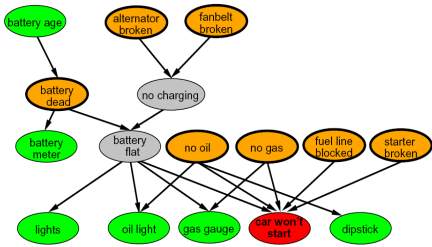
[Most slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at <http://ai.berkeley.edu>.]

Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
 - Unless there are only a few variables, the joint is WAY too big to represent explicitly
 - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
 - More properly called **graphical models**
 - We describe how variables locally interact
 - Local interactions chain together to give global, indirect interactions
 - For about 10 min, we'll be vague about how these interactions are specified

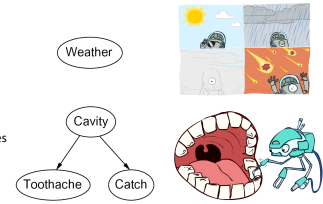


Example Bayes' Net: Car



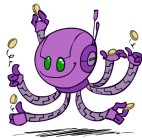
Graphical Model Notation

- Nodes: variables (with domains)**
 - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions**
 - Similar to CSP constraints
 - Indicate "direct influence" between variables
 - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)



Example: Coin Flips

- N independent coin flips



- No interactions between variables: **absolute independence**

Example: Traffic

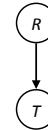
- Variables:**
 - R: It rains
 - T: There is traffic



- Model 1: independence



- Model 2: rain causes traffic



- Why is an agent using model 2 better?

Example: Traffic II

- Let's build a causal graphical model!
- Variables
 - T: Traffic
 - R: It rains
 - L: Low pressure
 - D: Roof drips
 - B: Ballgame
 - C: Cavity

Example: Alarm Network

- Variables
 - B: Burglary
 - A: Alarm goes off
 - M: Mary calls
 - J: John calls
 - E: Earthquake!

Bayes' Net Semantics

Bayes' Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
 - A collection of distributions over X , one for each combination of parents' values

$$P(X|a_1 \dots a_n)$$
 - CPT: conditional probability table
 - Description of a noisy "causal" process

A Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs

- Bayes' nets **implicitly** encode joint distributions
 - As a product of local conditional distributions
 - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together.

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

- Example:

$$P(+cavity, +catch, -toothache)$$

Probabilities in BNs

- Why are we guaranteed that setting

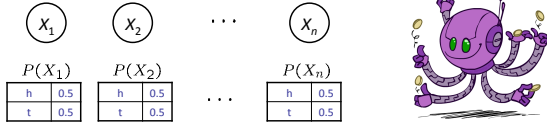
$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$
 results in a proper joint distribution?
- Chain rule (valid for all distributions):

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1 \dots x_{i-1})$$
- Assume conditional independences:

$$P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$$
 → Consequence:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$
- Not every BN can represent every joint distribution
 - The topology enforces certain conditional independencies

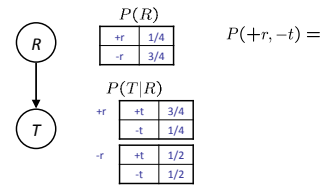
Example: Coin Flips



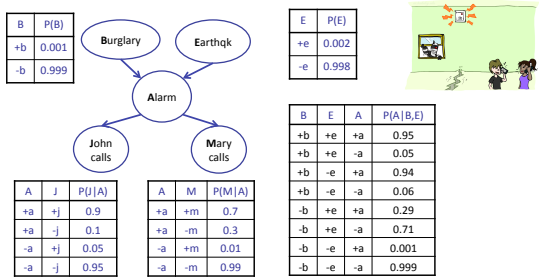
$P(h, h, t, h) =$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic

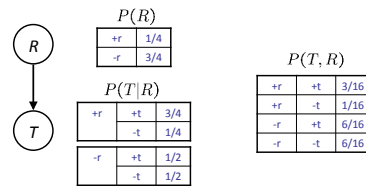


Example: Alarm Network



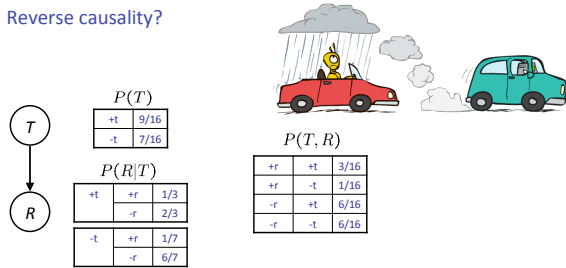
Example: Traffic

■ Causal direction



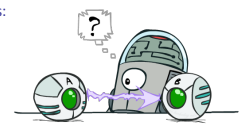
Example: Reverse Traffic

■ Reverse causality?



Causality?

- When Bayes' nets reflect the true causal patterns:
 - Often simpler (nodes have fewer parents)
 - Often easier to think about
 - Often easier to elicit from experts
- BNs need not actually be causal
 - Sometimes no causal net exists over the domain (especially if variables are missing)
 - E.g. consider the variables *Traffic* and *Drips*
 - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
 - Topology may happen to encode causal structure
 - Topology really encodes conditional independence



$P(x_i|x_1, \dots, x_{i-1}) = P(x_i|parents(X_i))$

Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution
- Next: how to answer queries about that distribution
 - Today:
 - First assembled BNs using an intuitive notion of conditional independence as causality
 - Then saw that key property is conditional independence
 - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)

