Bayes Nets

Many slides courtesy of
Dan Klein, Stuart Russell,
or Andrew Moore

CS 5300 / CS 6300
Artificial Intelligence
Spring 2009

Hal Daumé III
hal@cs.utah.edu

www.cs.utah.edu/~hal/courses/2009S_AI

Announcements

- Midterm
  - Oops :)
  - After we grade, we'll "adjust" to make it reasonable

Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

[redacted]

- Given a joint distribution, we can reason about unobserved variables given observations (evidence)
- General form of a query:

[redacted]

Stuff you care about *Stuff you already know*

- This kind of posterior distribution is also called the belief function of an agent which uses this model

Independence

- Two variables are *independent* if:

[redacted]

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

[redacted]

  - We write: [redacted]
- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence

- N fair, independent coin flips:

[redacted]

H	0.5
T	0.5

[redacted]

H	0.5
T	0.5

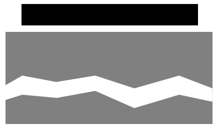
[redacted]

[redacted]

H	0.5
T	0.5

[redacted]

[redacted]



Example: Independence?

[redacted]

T	P
warm	0.5
cold	0.5

[redacted]

T	W	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

[redacted]

W	P
sun	0.6
rain	0.4

[redacted]

T	W	P
warm	sun	0.3
warm	rain	0.2
cold	sun	0.3
cold	rain	0.2

Conditional Independence

- P(Toothache,Cavity,Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(catch | toothache, cavity) = P(catch | cavity)
- The same independence holds if I don't have a cavity:
  - P(catch | toothache, ¬cavity) = P(catch| ¬cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)

Conditional Independence

- Unconditional (absolute) independence is very rare (why?)
- Conditional independence is our most basic and robust form of knowledge about uncertain environments:

[redacted]

- What about this domain:
  - Traffic
  - Umbrella
  - Raining
- What about fire, smoke, alarm?

The Chain Rule I

- Can always factor any joint distribution as an incremental product of conditional distributions

[redacted]

- Why?
- This actually claims nothing…
- What are the sizes of the tables we supply?

The Chain Rule II

- Trivial decomposition:

[redacted]

- With assumption of conditional independence:

[redacted]

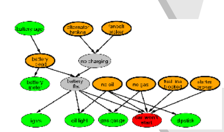
- Conditional independence is our most basic and robust form of knowledge about uncertain environments
- Bayes' nets / graphical models help manage independence

Birthday Paradox

- What's the probability that no two people in this room have the same birthday?

Probabilistic Models

- Models are descriptions of how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables

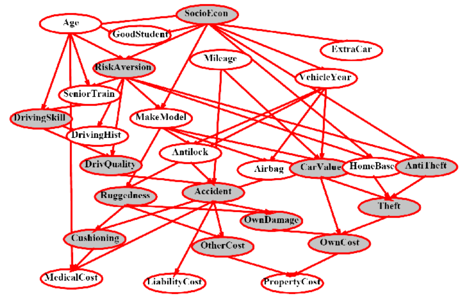


- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

Bayes' Nets: Big Picture

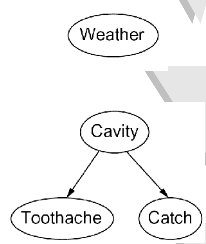
- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for 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



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
- For now: imagine that arrows mean direct causation



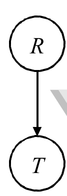
Example: Coin Flips

- 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

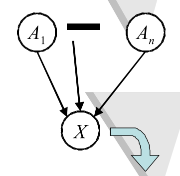
- Let's build a causal graphical model
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

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

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- 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

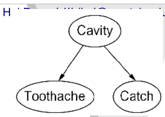


[redacted]

  - CPT: [redacted]bility 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:

[redacted]

  - Example:

[redacted]

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

[redacted]

h	0.5
t	0.5

[redacted]

h	0.5
t	0.5

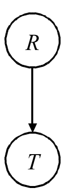
[redacted]

h	0.5
t	0.5

[redacted]

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

Example: Traffic



[redacted]

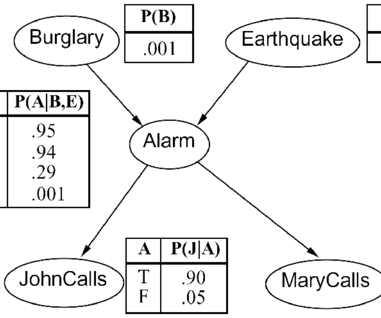
r	1/4
¬r	3/4

[redacted]

[redacted]

r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

Example: Alarm Network



P(B)
.001

P(E)
.002

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

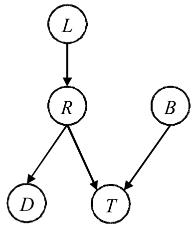
A	P(J|A)
T	.90
F	.05

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

[redacted]

Example: Traffic II

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame



Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?
- How big is an N-node net if nodes have k parents?
- Both give you the power to calculate
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Building the (Entire) Joint

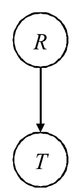
- We can take a Bayes' net and build the full joint distribution it encodes

[redacted]

  - Typically, there's no reason to build ALL of it
  - But it's important to know you could!
- To emphasize: every BN over a domain implicitly represents some joint distribution over that domain

Example: Traffic

- Basic traffic net
- Let's multiply out the joint



[redacted]

r	1/4
¬r	3/4

[redacted]

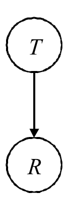
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

[redacted]

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Example: Reverse Traffic

- Reverse causality?



[redacted]

t	9/16
¬t	7/16

[redacted]

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

[redacted]

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

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

Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)