Informatics 2D: Reasoning and Agents

Alex Lascarides

School of Informatics
The University of Edinburgh

Lecture 23a: Introduction to Bayesian Networks

Where are we?

Last time ...

- Using JPD tables for probabilistic inference
- Concepts of absolute and conditional independence
- Bayes' rule

Today ...

- **Introduction to Bayesian Networks**

Representing knowledge in an uncertain domain

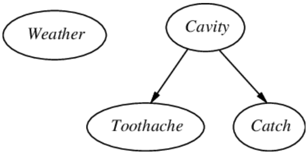
- Full joint probability distributions can become intractably large very quickly
- Conditional independence helps to reduce the number of probabilities required to specify the JPD
- Now we will introduce **Bayesian networks** (BNs) to systematically describe dependencies between random variables
- Roughly speaking, BNs are graphs that connect nodes representing variables with each other whenever they depend on each other

Bayesian networks

- A BN is a directed acyclic graph (DAG) with nodes annotated with probability information
- The nodes represent random variables (discrete/continuous)
- Links connect nodes. If there is an arrow from X to Y, we call X a **parent** of Y
- Each node X_i has a conditional probability distribution (CPD) attached to it
- The CPD describes how X_i depends on its parents, i.e. its entries describe $\mathbf{P}(X_i|Parents(X_i))$

Bayesian networks

- Topology of graphs describes conditional independence relationships
- Intuitively, links describe **direct effects** of variables on each other in the domain
- Assumption: anything that is not directly connected does not directly depend on each other
- In previous dentist/weather example:



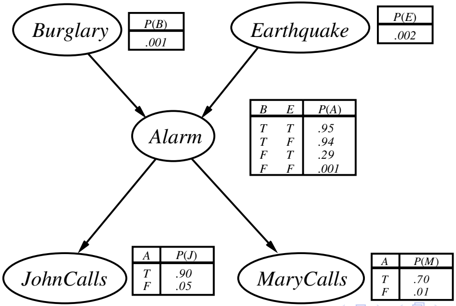
Arcs and Independence

Each variable is conditionally independent of its non-descendants, given its parents.

$$\text{If } X \notin Parents^*(Y), \text{ then}$$
$$\mathbf{P}(X|Parents(X), Y) = \mathbf{P}(X|Parents(X))$$

Example

- New burglar alarm has been fitted, fairly reliable but sometimes reacts to earthquakes
- Neighbours John and Mary promise to call when they hear alarm
- John sometimes mistakes phone for alarm, and Mary listens to loud music and sometimes doesn't hear alarm



Example – things to note

- No perception of earthquake by John or Mary
- No explicit modelling of phone ring confusing John, or of Mary's loud music
  (summarised in uncertainty regarding their reaction)
- Actually this uncertainty summarises any kind of failure
  - almost impossible to enumerate all possible causes,
  - and we don't have estimates for their probabilities anyway
- Each row in CPTs contains a **conditioning case**,
  one row for each possible combination of values of the parents.
- We often omit $P(\neg x_i|Parents(X_i))$ from CPT for node X_i
  (computes as $1 - P(x_i|Parents(X_i))$)
- $\mathbf{P}(J|M, A, B, E) = \mathbf{P}(J|A)$ and $\mathbf{P}(M|J, A, B, E) = \mathbf{P}(M|A)$

Summary

- BNs consist of two components:
  1. A graphical that captures conditional independence among RVs (more later)
  2. A CPT for each RV: $P(X|Parents(X))$
- Probabilities/uncertainty in a BN can be due to:
  1. Your choice not to model certain factors
  2. Genuine ignorance about what factors are relevant
- Next time: **BNs are a compact representation of JPDs**