CS4725/CS6705

Chapter 14: Probabilistic Reasoning

Full joint distributions

- Can answer any question about probabilities of events in a domain
- However:
  - They become unreasonably large as the number of variables grows
  - Specifying probabilities for atomic events (combinations of values for all variables) is unnatural and usually very difficult

Bayesian networks

- **Bayesian network:** A graphical structure used to represent the dependencies among variables and to give a concise specification of any full joint probability distribution

Bayesian networks

- A **Bayesian network** is a directed graph in which each *node* is labelled with probability information.
  1. Nodes represent random variables (discrete or continuous).
  2. If there is an arrow from X to Y, X is a *parent* of Y.
  3. Each node X_i has a conditional probability distribution $P(X_i \mid Parents(X_i))$ that quantifies the effect of the parents on the node.
  4. The graph has no directed cycles (DAG = directed, acyclic graph).

Constructing Bayesian networks

- The structure of a Bayesian network captures the independence relationships among the variables.
- Idea: An arrow from X to Y indicates that X has a direct influence on Y.
- Once a domain expert has specified the *structure* of a network, the next task is to specify a conditional probability distribution for each variable, given its parents.

Bayesian network: example

- Consider the following example:
  - Burglar alarm: quite reliable at detecting burglaries, but also responds sometimes to earthquakes
  - Two neighbours, John and Mary, have promised to phone you at work if they hear the alarm.
  - John always calls when he hears the alarm, but sometimes calls mistakenly when he hears your phone ring.
  - Mary often misses the alarm altogether.
  - Given who has called or not called, what is the probability of a burglary?
  - [Example on the board]

Conditional probability tables

- Each node in a Bayesian network has an associated **conditional probability table**, showing the probabilities of different values of the node, given each possible combination of values for its parent nodes.
- [Example on the board]

BNs as representations of full joint distributions

- Every entry in a full joint distribution can be calculated from the information in a BN.
- [Details and example on the board]

Constructing BNs

- Try to add nodes in an order such that the "root causes" are added first, then the variables that they influence, and so on.
- The parents of node X_i should be all those nodes previously added that *directly influence* X_i
- Adding nodes in an incorrect order can lead to much more complicated networks. [See burglar alarm example in the book.]

Exact inference in BNs

- Probabilistic inference: Given some observed **event** (i.e., some assignment of values to a set of **evidence variables**), what is the probability distribution for a set of **query variables**?
  - For example: Given that both John and Mary phoned, what is the probability that there was a burglary?

Exact inference: notation

- Let X be the query variable.
- Let **E** be the set of evidence variables
- Let **e** be a particular observed event.
- Let **Y** be the nonevidence variables.

- Typically, we are looking for the posterior probability distribution **P**(X | **e**).

Inference by enumeration

- One approach to inference in Bayesian networks:
  - Just compute sums of products of conditional probabilities from the network
  - [Example on the board]

Complexity of exact inference

- If a network is **singly connected** (at most one undirected path between any two nodes), then time and space complexity are **linear** in the size of the network (number of CPT entries).
- For **multiply connected** networks, complexity can be exponential in the worst case.

Approximate inference

- For large, multiply connected networks, we cannot use exact inference.
- **Monte Carlo** algorithms (or randomized sampling):
  - Basic idea: generate many random samples based on the probability distributions associated with the variables
  - Answer a query based on the fraction of sample cases that satisfy the query.
  - Higher number of samples: more accurate estimates

Other approaches to uncertain reasoning

- Default reasoning
- Rule-based approaches
- Dempster-Shafer theory: interval-valued degrees of belief
- Fuzzy logic