Lecture 2: Introduction to Bayesian Networks

Overview:

- Syntax
- Earthquake Example
- Updating Beliefs
- Representation of the Joint Distribution
- Network construction
- Conditional Independence, D-separation and Causal Ordering
- Dynamic (Temporal) Networks
- Uses of Bayesian Networks
- BN Software and Web Resources

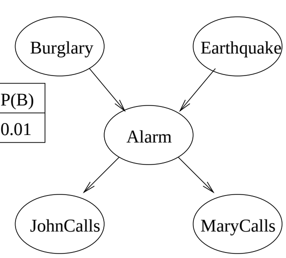
Bayesian Networks

- Data Structure which represents the dependence between variables;
- Gives concise specification of the joint probability distribution.
- A Bayesian Network is a graph in which:
  1. A set of random variables makes up the nodes in the network.
  2. A set of directed links or arrows connects pairs of nodes.
  3. Each node has a conditional probability table that *quantifies* the effects the parents have on the node.
  4. Directed, acyclic graph (DAG), i.e. no directed cycles.

Example: Earthquake (Pearl,R&N)

- You have a new burglar alarm installed.
- It is reliable about detecting burglary, but responds to minor earthquakes.
- Two neighbours (John, Mary) promise to call you at work when they hear the alarm.
  - John always calls when hears alarm, but confuses alarm with phone ringing (and calls then also)
  - Mary likes loud music and sometimes misses alarm!
- Given evidence about who has and hasn't called, estimate the probability of a burglary.

Earthquake Example: Network Structure



P(E)
0.02

P(B)
0.01

B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

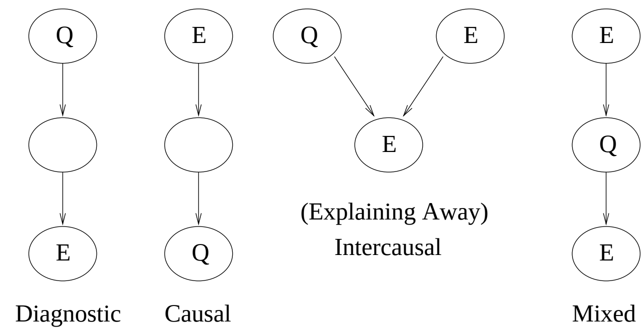
Earthquake Example: Notes

- Assumptions: John and Mary don't perceive burglary directly; they do not feel minor earthquakes.
- Note: no info about loud music or telephone ringing and confusing John. Summarised in uncertainty in links from `Alarm` to `JohnCalls` and `MaryCalls`.
- Once topology is specified, need to specify **conditional probability table** for each node.
  - Each row contains the cond prob of each node value for a **conditioning case**.
  - Each row must sum to 1.
  - A table for a Boolean var with n Boolean parents contain 2^{n+1} probs.
  - A node with no parents has one row (the prior probabilities)

Inference in Bayesian Networks

- Basic task for any probabilistic inference system:

  Compute the posterior probability distribution for a set of **query variables**, given values for some **evidence variables**.
- Also called *Belief Updating*.
- Types of Inference:



Kinds of Inference

- Diagnostic inferences: from effect to causes.

  `P(Burglary|JohnCalls)`
- Causal Inferences: from causes to effects.

  `P(JohnCalls|Burglary)`

  `P(MaryCalls|Burglary)`
- Intercausal Inferences: between causes of a common effect.

  `P(Burglary|Alarm)`

  P(Burglary|Alarm $\wedge$ Earthquake)
- Mixed Inference: combining two or more of above.

  P(Alarm|JohnCalls $\wedge$ $\neg$EarthQuake)

  P(Burglary|JohnCalls $\wedge$ $\neg$EarthQuake)

Semantics of Bayesian Networks

- A (more compact) representation of the joint probability distribution.
  - helpful in understanding how to construct network
- Encoding a collection of conditional independence statements.
  - helpful in understanding how to design inference procedures

The two views are *equivalent*.

Representing the joint probability distribution

$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$

$= P(x_1, x_2, ..., x_n)$

$= P(x_1) \times P(x_2|x_1) ... \times P(x_n|x_1 \wedge ... x_{n-1})$

$= \prod_i P(x_i|x_1 \wedge ... x_{i-1})$

$= \prod_i P(x_i|Parents(X_i))$

Example: $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$

$= P(J|A)P(M|A)P(A|\neg B \wedge \neg E)P(\neg B)P(\neg E)$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.0067.$

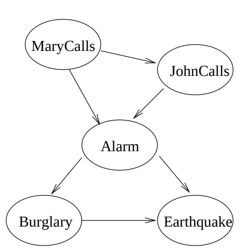
Network Construction Algorithm

1. Choose the set of relevant variables X_i that describe the domain.
2. Choose an ordering for the variables.
3. While there are variables left:
   (a) Pick a variable X_i and add a node to the network for it.
   (b) Set $Parents(X_i)$ to some minimal set of nodes already in the net such that the conditional independence property is satisfied.
   $P(X_i|X_{i-1}, ..., X_1) = P(X_i|Parents(Xi))$
   (c) Define the CPT for X_i

Compactness and Node Ordering

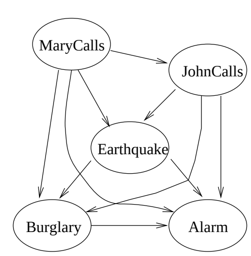
- Compact BN has order $k < N$ (i.e. it is sparsely connected).
- The optimal order to add nodes is to add the "root causes" first, then the variable they influence, so on until "leaves" reached.
- Examples of poor ordering (which still represent same joint distribution):

1. MaryCalls, JohnCalls, Alarm, Burglary, Earthquake.



Compactness and Node Ordering (cont.)

2. MaryCalls, JohnCalls, Earthquake, Burglary, Alarm.

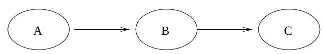


Same number of probabilities as full joint!

See below for *why*

Conditional Independence: Causal Chains

Causal chains give rise to conditional independence:



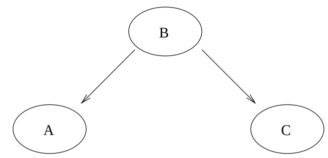
$$P(C|A \wedge B) = P(C|B)$$

Example

- A = Jack's flu
- B = severe cough
- C = Jill's flu

Conditional Independence: Common Causes

Common causes (or ancestors) also give rise to conditional independence:



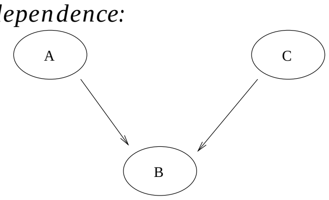
$$P(C|A \wedge B) = P(C|B)$$

Example

- A = Jack's flu
- B = Joe's flu
- C = Jill's flu

Conditional Dependence: Common Effects

Common effects (or their descendants) give rise to conditional *dependence:*



$$P(A|C \wedge B) \neq P(A)P(C)$$

Example

- A = flu
- B = severe cough
- C = tuberculosis

  Given a severe cough, flu "explains away" tuberculosis.

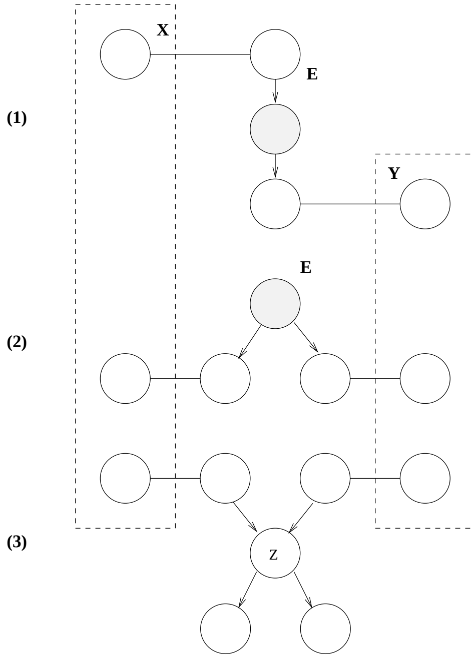
D-separation

Can we determine whether a set of nodes X is independent of another set Y, given a set of evidence nodes E?

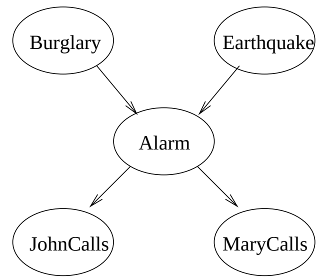
- If every undirected path from a node in X to a node in Y is *d-separated* by E, then X and Y are *conditionally independent* given E.
- A set of nodes E *d-separates* two sets of nodes X and Y if every undirected path from a node in X to a node in Y is *blocked* given E.
- A path is *blocked* given a set of nodes E if there is a node Z on the path for which one of three conditions holds:
  1. Z is in E and Z has one arrow on the path leading in and one arrow out (chain).
  2. Z is in E and Z has both path arrows leading out (common cause).
  3. Neither Z nor any descendant of Z is in E, and both path arrows lead in to Z (common effect).

D-separation (cont'd)

- Evidence nodes **E** shown shaded.



D-separation: earthquake example



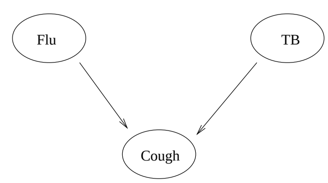
Causal Ordering

Why does variable order affect network density?

Because

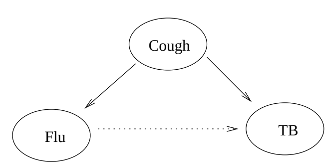
- Using the causal order allows direct representation of conditional independencies
- Violating causal order requires new arcs to re-establish conditional independencies

Causal Ordering (cont'd)



Flu and TB are marginally independent.

Given the ordering: Cough, Flu, TB:



Marginal independence of Flu and TB must be re-established by adding $Flu \rightarrow TB$ or $Flu \leftarrow TB$

Example: Cancer

Metastatic cancer is a possible cause of a brain tumor and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also possibly associated with a brain tumor. (Example from: Pearl, 1988.)

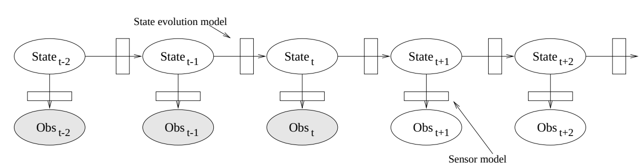
Example: Asia

A patient presents to a doctor with shortness of breath. The doctor considers that possibles causes are tuberculosis, lung cancer and bronchitis. Other additional information that is relevant is whether the patient has recently visited Asia (where tuberculosis is more prevalent), whether or not the patient is a smoker (which increases the chances of cancer and bronchitis). A positive xray would indicate either TB or lung cancer. (Example from (Lauritzen, 1988).)

Example: A Lecturer's Life

Dr. Ann Nicholson spends 60% of her work time in her office. The rest of her work time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get some real work done). When she is not in her office, she leaves her light on only 5% of the time. 80% of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10% of the time she is not in her office, she is still logged onto the computer. Suppose a student checks Dr. Nicholson's login status and sees that she is logged on. What effect does this have on the student's belief that Dr. Nicholson's light is on?

Dynamic Belief Networks



- The values of state variables at time t depend only on the values at $t-1$.
- Can calculate distributions for S_{t+1} and further: **probabilistic projection**.
- Can be done using standard BN updating algorithms
- This type of DBN gets very large, very quickly.
- Usually only keep two time slices of the network.
- (See examples in Lecture 6 on BN Case Studies)

Uses of Bayesian Networks

1. Calculating the belief in query variables given values for evidence variables (above).
2. Predicting values in dependent variables given values for independent variables.
3. Decision making based on probabilities in the network and on the agent's utilities (Influence Diagrams [Howard and Matheson 1981]; see Lecture 4).
4. Deciding which additional evidence should be observed in order to gain useful information.
5. *Sensitivity analysis* to test impact of changes in probabilities or utilities on decisions.

BN Software and Web Resources

- Netica: `www.norsys.com`
- Hugin: `www.hugin.com`
- Analytica: `www.lumina.com`
- JavaBayes: `http://www.cs.cmu.edu/ javabayes/Home/`
- Bayesware: `http://www.bayesware.com`
- Genie: `http://www2.sis.pitt.edu/ genie`
- BN info site (follow links from subject homepage)
  - Bayesian Belief Network site (Russell Greiner)
  - Summary of BN software and links to software sites (Kevin Murphy)

Summary

- Bayes' rule allows unknown probabilities to be computed from known ones.
- Conditional independence (due to causal relationships) allows efficient updating
- Bayesian networks are a natural way to represent conditional independence info.
  - links between nodes: qualitative aspects;
  - conditional probability tables: quantitative aspects.
- Robust and easy to use Bayesian network software is now readily available.

References

E. Charniak (1991), "Bayesian Networks Without Tears", *Artificial Intelligence Magazine*, pp. 50-63, Vol 12.

P. Haddaway, "An Overview of Some Recent Developments in Bayesian Problem-Solving Techniques". In *Artificial Intelligence Magazine*, pages 50-63, Vol 20, No. 2, 1999.

F. V. Jensen (1996), *An introduction to Bayesian Networks*, Springer.

S.L Lauritzen and D.J. Spiegelhalter (1988) "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems", *Journal of the Royal Statistical Society*, 50(2), pp. 157-224.

J. Pearl (1988), *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann.

Poole, D., Mackworth, A., and Goebel, R. (1998) *Computational Intelligence: a logical approach*. Oxford University Press.

Russell & Norvig (1995), *Artificial Intelligence: A Modern Approach*, Prentice Hall.