Bayesian Networks

Lyle Ungar

Heavily adapted from slides by Mitch Marcus

Semantics of belief nets: conditional independence
Active trails
D-separation / Markov separation
Next: HMM’s, a kind of Bayes Net
Bayesian Networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

**Syntax:**
- A set of nodes, one per random variable
- A set of directed edges (link ≈ "directly influences"), yielding a directed, acyclic graph
- A conditional distribution for each node given its parents: $P(X_i | \text{Parents}(X_i))$

**In the simplest case, the conditional distribution is represented as a Conditional Probability Table (CPT) giving the distribution over $X_i$ for each combination of parent values**

Why not just specify the full probability distribution?
The Required-By-Law Burglar Example

I'm at work, and my neighbor John calls to say my burglar alarm is ringing, but my neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology ideally reflects causal knowledge:
1. A burglar can set the alarm off
2. An earthquake can set the alarm off
3. The alarm can cause Mary to call
4. The alarm can cause John to call
Local Markov Assumption (LMA)

- **LMA**: A variable $X$ is independent of its non-descendants given its parents:
  - $(X \perp \text{NonDescendants}_X | \text{Parents}_X)$

- The LMA lets us decompose the joint in terms of the CPTs:
  $$P(E,B,A,J,M) =$$
  $$P(E)P(B)P(A|B,E)P(J|A)P(M|A)$$
Conditional Independence Properties

- **Symmetry**: \((X \perp Y|Z) \rightarrow (Y \perp X|Z)\)
- **Decomposition**: \((X \perp Y,W|Z) \rightarrow (X \perp Y|Z)\)
- **Weak union**: \((X \perp Y,W|Z) \rightarrow (X \perp Y|Z,W)\)
- **Contraction**: \((X \perp W|Y,Z),(X \perp Y|Z) \rightarrow (X \perp Y,W|Z)\)
B ⊥ J ?
B ⊥ J|A ?
J ⊥ M|A ?
B ⊥ E|A ?
Explaining away
A trail \( \{X_1, X_2, \ldots, X_k\} \) in the graph (no cycles) is an **active trail** if for each consecutive triplet in the trail:

- \( X_{i-1} \rightarrow X_i \rightarrow X_{i+1} \), and \( X_i \) is not observed
- \( X_{i-1} \leftarrow X_i \leftarrow X_{i+1} \), and \( X_i \) is not observed
- \( X_{i-1} \leftarrow X_i \rightarrow X_{i+1} \), and \( X_i \) is not observed
- \( X_{i-1} \rightarrow X_i \leftarrow X_{i+1} \), and \( X_i \) is observed or one of its descendants is observed

Variables connected by active trails are not conditionally independent
D-separation

◆ Variables $X_i$ and $X_j$ are independent if there is no active trail between $X_i$ and $X_j$.
  ● given a set of observed variables $O \subseteq \{X_1, \ldots, X_m\}$
Examples
Semantics

- Local semantics give rise to global semantics
- Local semantics: given its parents, each node is conditionally independent of everything except its descendants
- Global semantics (why?):
  \[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i)) \]
  \[ P(J \land M \land A \land \neg B \land \neg E) = \]
  \[ P(\neg B)P(\neg E)P(A | \neg B \land \neg E)P(J | A)P(M | A) \]
How to build a belief net?

- **Ask people**
  - Try to model causality in structure
  - Estimate probabilities from data

- **Automatic construction**
  - How to cast belief network building as search?
How to learn a belief net?

◆ Pick the loss function that you want to minimize
  ● \(- \log(\text{likelihood}) + \text{const} \times \text{(\# of parameters)}\)
    ■ Parameters, \(\theta\), are the numbers in the conditional probability tables
  ● Likelihood = \(p(X|\ \theta) = \prod_i p(x_i|\theta)\)
    ■ Probability of the data given the model

◆ Do stochastic gradient descent (or annealing)
  ● Randomly change the network structure by one link
  ● Re-estimate the parameters
  ● Accept the change if the loss function is lower
    ■ Or sometimes even if it is higher
Hidden-latent variables

- Sometimes not every variable is observable
- In this case, to do estimation, use EM
  - If you know the values of the latent variable, then it is trivial to estimate the model parameters (here, the values in the conditional probability tables)
  - If you know the model (the probabilities), you can find the expected values of the non-observed variables
- Thus, given a known structure, one can find a (local) maximum likelihood model.
Belief Network with Hidden Variables: Example
Belief Net Extensions

- Hidden variables
- Decision (action) variables
- Continuous distributions
- “Plate” models
  - E.g. Latent Dirichlet Allocation (LDA)
- Dynamical Belief Nets (DBNs)
  - E.g. Hidden Markov Models (HMMs)

What is the simplest Belief net with hidden variables?
What you should know

- **Belief nets provide a compact representation of joint probability distributions**
  - Guaranteed to be a consistent specification
  - Graphically show conditional independence
  - Can be extended to include “action” nodes

- **Networks that capture causality tend to be sparser**

- **Estimating belief nets is easy if everything is observable**
  - But requires serious search if the network structure is not known
Gene transcriptional response to pathogen infection
Bayesian network questions

• What does the belief network diagram mean?
  - Why are the diagrams good?
  - Why are the diagrams bad?

• How many parameters in the model below?
  - A) <12  B) 12-16  C)17-20  D) more than 20
  - How does this change if a variable can take on 3 values?

• What is conditionally independent of what?
  - P(D|C) = P(D|C,E)?
  - P(D|F) = P(D|F,E)?
  - P(F|D) = P(F|D,A)?
  - P(F|D) = P(F|D,C)?
Bayesian network questions

• **What does the belief network diagram mean?**
  - Why are the diagrams good? - clearly show conditional independence
  - Why are the diagrams bad? - don’t show equation form (binary, Gaussian…)

• **How many parameters in the model below?**
  - A) <12     B) 12-16     C)17-20     D) more than 20
  - How does this change if a variable can take on 3 values?

• **What is conditionally independent of what?**
  - $P(D|C) = P(D|C,E)$? - yes
  - $P(D|F) = P(D|F,E)$? - no
  - $P(F|D) = P(F|D,A)$? - yes
  - $P(F|D) = P(F|D,C)$? - no

$P(A),P(B),P(C): 3$ parameters

$P(D|A,B,C):8$    $P(E|C):2$    $P(F|D,E):4$    $3+8+2+4=17$
Bayesian network questions

• How are most belief nets built?
  • A) interview people
  • B) machine learning

• How to build a belief net?
  • Sequentially add variables, checking for conditional independence
  • Is it sufficient to check $P(C|A,B) = P(C|A)$, $P(C|B)$ or $P(C)$?
    • A) yes
    • B) no

• What makes a belief net better or worse?

• How can you build a better belief net?