Bayes Networks: Representation and Inference

Manuela M. Veloso
Carnegie Mellon University
Computer Science Department

Thanks to past instructors

15-780 Graduate AI – Spring 2013

Readings:
• Russell & Norvig: chapter 14

Bayes Rule

• \( P(A \mid B) = \frac{P(A, B)}{P(B)} \)

• \( P(A, B) = P(A \mid B) \cdot P(B) \)

• \( P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \)
Independence

Variables $R$ and $S$ are independent if:

$$P(S \mid R) = P(S)$$

And we can then derive:

- $P(\neg S \mid R) = P(\neg S)$
- $P(S,R) = P(S)P(R)$
- $P(R \mid S) = P(R)$
Conditional Independence

Two dependent random variables may become independent when conditioned on a third variable:

\[ P(A,B \mid C) = P(A \mid C) P(B \mid C) \]

Example:

\begin{align*}
P(\text{liked movie}) &= 0.5 \\
P(\text{slept}) &= 0.4 \\
P(\text{liked movie, slept}) &= 0.1 \\
P(\text{liked movie} \mid \text{long}) &= 0.4 \\
P(\text{slept} \mid \text{long}) &= 0.6 \\
P(\text{like movie, slept} \mid \text{long}) &= 0.24
\end{align*}

Bayesian Networks

Bayesian networks are directed acyclic graphs with nodes representing random variables and edges representing dependency assumptions.

\begin{align*}
P(\text{Lo}) &= 0.5 \\
P(\text{Li} \mid \text{Lo}) &= 0.4 \\
P(\text{Li} \mid \neg \text{Lo}) &= 0.7 \\
P(\text{S} \mid \text{Lo}) &= 0.6 \\
P(\text{S} \mid \neg \text{Lo}) &= 0.2
\end{align*}
Constructing a Bayesian Network

- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies
- Step 3: Populate the CPTs

Example Problem

An alarm system:
- B – Did a burglary occur?
- E – Did an earthquake occur?
- A – Did the alarm sound off?
- M – Mary calls
- J – John calls

How do we construct the network for this problem with these 5 random variables?
Factoring Joint Distributions

We can try to get conditional dependencies of one variable on others (to represent graphically as earlier).

Using the chain rule we can always factor a joint distribution:

\[
\]

Example: Variables ordered alphabetically. Other orderings fine too.

A Bayesian Network

\[
\]

Number of params in CPT:

- A: 2^4
- B: 2^3
- E: 4
- J: 2
- M: 1

Total: 31 parameters
Using Domain Knowledge

An alarm system:

- B – Did a burglary occur?
- E – Did an earthquake occur?
- A – Did the alarm sound off?
- M – Mary calls
- J – John calls

Domain knowledge: story… (Judea Pearl)

Number of parameters:
- A: 4, B: 1, E: 1, J: 2, M: 2
- Total: 10 parameters

By using domain knowledge we saved 21 parameters.

Constructing a Bayesian Network: Revisited

- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies
  - Select an ordering of the variables
  - Add them one at a time
  - For each new variable X added, select the minimal subset of nodes as parents, such that
    X is independent from all other nodes in the current network, given its parents
- Step 3: Populate the CPTs
Bayesian Network Inference

- Given a network, we want to INFER values for all sorts of unobserved variables or joint variables, or conditional variables – anything!

- For example, in our previous network the only observed variables are the phone calls from John and Mary. However, what we are really interested in is the value of the variable burglary, given our observations.

- INERENCE: for example: P (B | J, ¬M)
Inference

A simpler question: How can we compute a joint distribution from the network?
For example, \( P(B, \neg E, A, J, \neg M) \).

\[
P(B, \neg E, A, J, \neg M) = P(B)P(\neg E)P(A | B, \neg E)P(J | A)P(\neg M | A)
\]

\[
= 0.05 * 0.9 * 0.85 * 0.7 * 0.2
\]

\[
= 0.005355
\]

Inference for Partial Joints

We are now interested in queries of the form:
\( P(B | J, \neg M) \)

Conditionals can be written as joints:

\[
P(B | J, \neg M) = \frac{P(B, J, \neg M)}{P(B, J, \neg M) + P(\neg B, J, \neg M)}
\]

How do we compute a joint that does not include all the variables of the network (partial joint)?

We can “add” the missing variables in all their possible combinations.
Computing: $P(B,J, \neg M)$

$$P(B) = .05 \quad P(E) = .1$$

$$P(B,J, \neg M,A,E) + P(B,J, \neg M, \neg A,E) + P(B,J, \neg M,A,\neg E) + P(B,J, \neg M, \neg A, \neg E) =$$

$$0.0007 + 0.00001 + 0.005 + 0.0003 = 0.00601$$

Computing: $P(B,J, \neg M)$

$$P(B) = .05 \quad P(E) = .1$$

$$P(B,J, \neg M,A,E) + P(B,J, \neg M, \neg A,E) + P(B,J, \neg M,A,\neg E) + P(B,J, \neg M, \neg A, \neg E) =$$

$$0.0007 + 0.00001 + 0.005 + 0.0003 = 0.00601$$

How can we reuse computations?
Computing: $P(B, J, \neg M)$

\[
P(B, J, \neg M, A, E) + P(B, J, \neg M, \neg A, E) + P(B, J, \neg M, A, \neg E) + P(B, J, \neg M, \neg A, \neg E) = 
\]

\[\sum_{a \in \{A, \neg A\}} \sum_{e \in \{E, \neg E\}} P(B)P(e)P(a \mid B,e)P(M \mid a)P(J \mid a)\]

Store as a function of $a$ and use whenever necessary (no need to recompute each time)

Computational Complexity

- We can reuse computations to reduce the running time
- However, there are still cases in which this algorithm will lead to exponential running time.
- Exact Bayesian Inference is NP-Hard.
Stochastic Inference

We can easily sample the joint distribution to obtain possible instances:
1. Sample the free variables
2. For every other variable: If all parents have been sampled, sample based on conditional distribution

We end up with a new set of assignments for B, E, A, J and M which are a random sample from the joint

Using Sampling For Inference

- Revisiting our problem to compute \( P(B \mid J, \neg M) \)
- Looking at the samples we can count:
  - \( N \): total number of samples
  - \( N_c \): total number of samples in which the condition holds \((J, \neg M)\)
  - \( N_B \): total number of samples where the joint is true \((B, J, \neg M)\)
- For a large enough \( N \):
  - \( N_c / N = P(J, \neg M) \)
  - \( N_B / N = P(B, J, \neg M) \)
- And so, we can set:
  \[
P(B \mid J, \neg M) = P(B, J, \neg M) / P(J, \neg M) \approx N_B / N_c
\]
Using Sampling For Inference

• Lets revisit our problem to compute $P(B \mid J, \neg M)$
• Looking at the samples we can count:
  - $N$: total number of samples
  - $N_c$: total number of samples in which the condition holds (J, $\neg M$)
  - $N_B$: total number of samples where the joint is true (B, J, $\neg M$)
• For a large enough $N$:
  - $N_c / N \approx P(J, \neg M)$
  - $N_B / N \approx P(B, J, \neg M)$
• And so, we can set:
  \[ P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)} \approx \frac{N_B}{N_c} \]

Problem: What if the condition rarely happens?
We would need lots and lots of samples, and most would be wasted

Weighted Sampling

• Again: $P(B \mid J, \neg M)$
• Given an assignment to parents, we assign a value of 1 to $J$ and 0 to $M$
• This way, all samples will contain the correct values for the conditional variables
• We record the probability of this assignment ($w = p_1^* p_2$) and we weight the new joint sample by $w$
Summary

• Bayes Rule
• Joint distribution, independence, conditional independence
• Attributes of Bayesian networks
• Constructing a Bayesian network
• Inference in Bayesian networks
  – exact, stochastic, weighted stochastic