Lecture 28 of 42

Reasoning under Uncertainty: Introduction to Graphical Models, Part 2 of 2

William H. Hsu
Department of Computing and Information Sciences, KSU

KSOL course page: http://snipurl.com/v9v3
Course web site: http://www.kddresearch.org/Courses/CIS730
Instructor home page: http://www.cis.ksu.edu/~bhsu

Reading for Next Class:

Hugin Bayesian Network tutorials: http://www.hugin.com/developer/tutorials/
Building, learning BNs: http://bit.ly/2IeNgz, http://bit.ly/2yWocz
Kevin Murphy's survey on BNs, representation: http://bit.ly/4ihafj

Lecture Outline

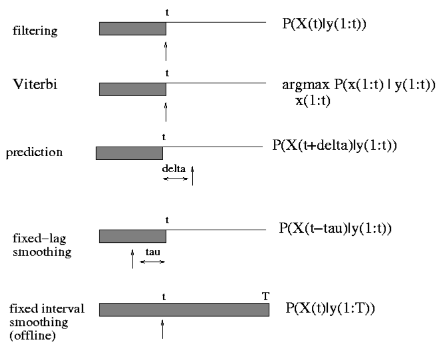
- **Reading for Next Class:** Murphy tutorial, Part 1 of 3; Hugin tutorial
- Last Class: 14.1 – 14.2 (p. 492 – 499), R&N 2e
- Today: Graphical Models, Sections 14.3 – 14.5 (p. 500 – 518), R&N 2e
- Coming Week: Graphical Models Concluded, Intro to Learning

Temporal Probabilistic Reasoning: Review

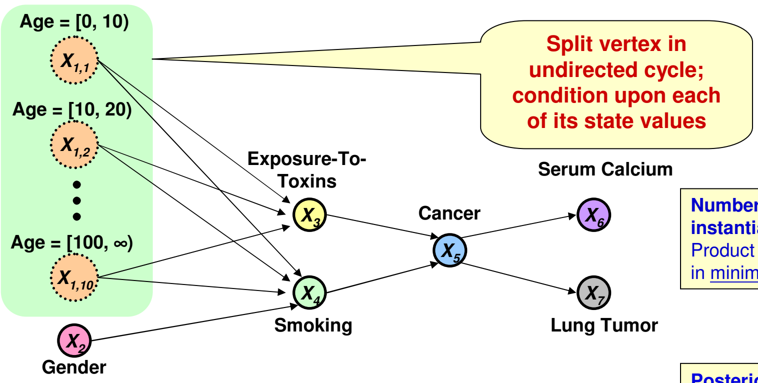
- **Goal: Estimate** $P(X_t^i \mid y_{1 \ldots r})$

Adapted from Murphy (2001), Guo (2002)

- **Filtering:** $r = t$
  - **Intuition: infer current state from observations**
  - **Applications: signal identification**
  - **Variation: Viterbi algorithm**
- **Prediction:** $r < t$
  - **Intuition: infer future state**
  - **Applications: prognostics**
- **Smoothing:** $r > t$
  - **Intuition: infer past hidden state**
  - **Applications: signal enhancement**
- **CF Tasks**
  - **Plan recognition by smoothing**
  - **Prediction cf. *WebCANVAS* – Cadez *et al.* (2000)**



Inference by Loop Cutset Conditioning



Number of network instantiations: Product of arity of nodes in minimal loop cutset

Posterior: marginal conditioned upon cutset variable values

- **Deciding Optimal Cutset:** **NP-*hard***
- **Current Open Problems**
  - **Bounded cutset conditioning: ordering heuristics**
  - **Finding randomized algorithms for loop cutset optimization**

Inference by Variable Elimination [1]: Factoring Operations

Enumeration is inefficient: repeated computation
e.g., computes $P(J = true|a)P(M = true|a)$ for each value of e

Variable elimination: carry out summations right-to-left,
storing intermediate results (factors) to avoid recomputation

$$
\begin{aligned}
\mathbf{P}(B|J = true, M = true) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{B} \Sigma_e \underbrace{P(e)}_{E} \Sigma_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(J = true|a)}_{J} \underbrace{P(M = true|a)}_{M} \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) P(J = true|a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) f_{\bar{A}JM}(b,e) \text{ (sum out } A) \\
&= \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (sum out } E) \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}
$$

Inference by Variable Elimination [2]: Pointwise Product

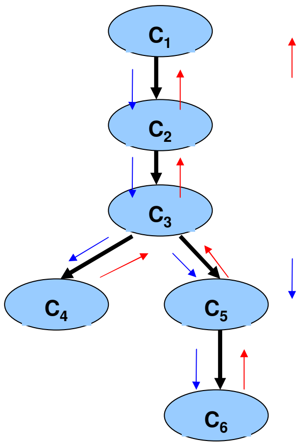
Pointwise product of factors f_1 and f_2:
$f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \times f_2(y_1, \ldots, y_k, z_1, \ldots, z_l)$
$= f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)$
E.g., $f_1(a,b) \times f_2(b,c) = f(a,b,c)$

Summing out a variable from a product of factors: move any constant factors outside the summation:

$$\Sigma_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \, \Sigma_x \, f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assuming $f_1, \ldots, f_i$ do not depend on X

Pearl's Propagation Algorithm: Review



Upward (child-to-parent) λ messages

$\Psi'(C_i')$ modified during λ message-passing phase

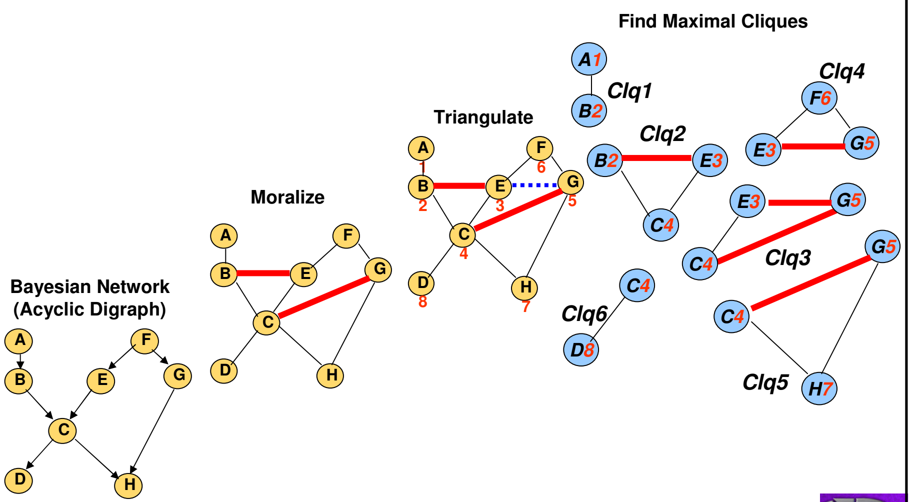
Downward π messages

$P'(C_i')$ is computed during π message-passing phase

Multiply-connected case: exact, approximate inference are #P-complete (counting problem is #P-complete iff decision problem is NP-complete)

Adapted from Neapolitan (1990), Guo (2000)

Inference by Clustering [1]: Moralization, Triangulation, Cliques



Adapted from Neapolitan (1990), Guo (2000)

Inference by Clustering [2]: Junction Tree Algorithm

Input: list of cliques of **triangulated, moralized graph** G_u

Output:

- **Tree of cliques**
- **Separators nodes S_i**,
- Residual nodes R_i and **potential probability $\Psi(Clq_i)$** for all cliques

Algorithm:

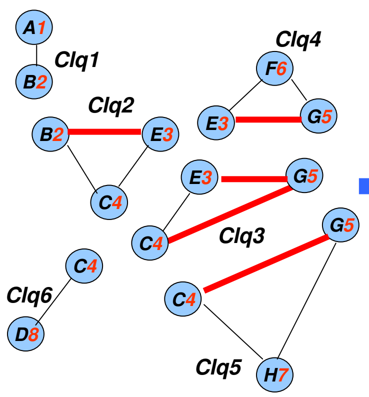
1. $S_i = Clq_i \cap (Clq_1 \cup Clq_2 \cup \ldots \cup Clq_{i-1})$
2. $R_i = Clq_i - S_i$
3. If $i > 1$ then identify a $j < i$ such that Clq_j is a parent of Clq_i
4. Assign each node v to a unique clique Clq_i that $v \cup c(v) \subseteq Clq_i$
5. Compute $\Psi(Clq_i) = \Pi_{f(v)\ Clqi} = P(v \mid c(v))$ {1 if no v is assigned to Clq_i}
6. Store Clq_i , R_i , S_i, and $\Psi(Clq_i)$ at each vertex in the tree of cliques

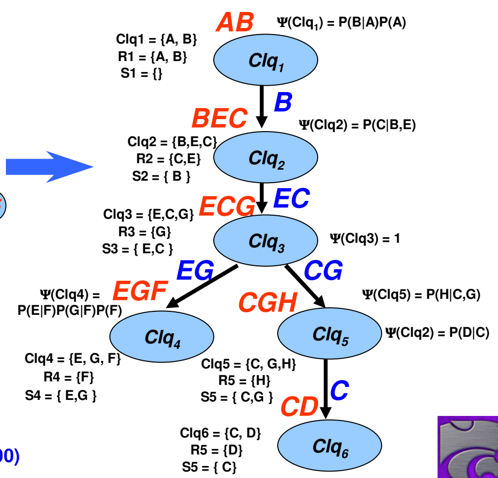
Inference by Clustering [3]: Clique Tree Operations

R_i: residual nodes

S_i: separator nodes

$\Psi(Clq_i)$: potential probability of Clique i





Adapted from Neapolitan (1990), Guo (2000)

Learning Structure from Data

- **General-Case BN Structure Learning: *Use Inference to Compute Scores***
- **Optimal Strategy: Bayesian Model Averaging**
  - Assumption: models $h \in H$ are mutually exclusive and exhaustive
  - *Combine predictions of models in proportion to marginal likelihood*
    - Compute conditional probability of hypothesis h given observed data D
    - *i.e.,* compute expectation over unknown h for unseen cases
    - Let $h \equiv$ structure, parameters $\Theta \equiv$ CPTs

$$P\left(\vec{x}^{(m+1)} / D\right) = P\left(x_1, x_2, \ldots, x_n / \vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(m)}\right)$$

$$= \sum_{h \in H} P\left(\vec{x}^{(m+1)} / D, h\right) \cdot P(h / D)$$

Posterior Score

Marginal Likelihood

Prior over Parameters

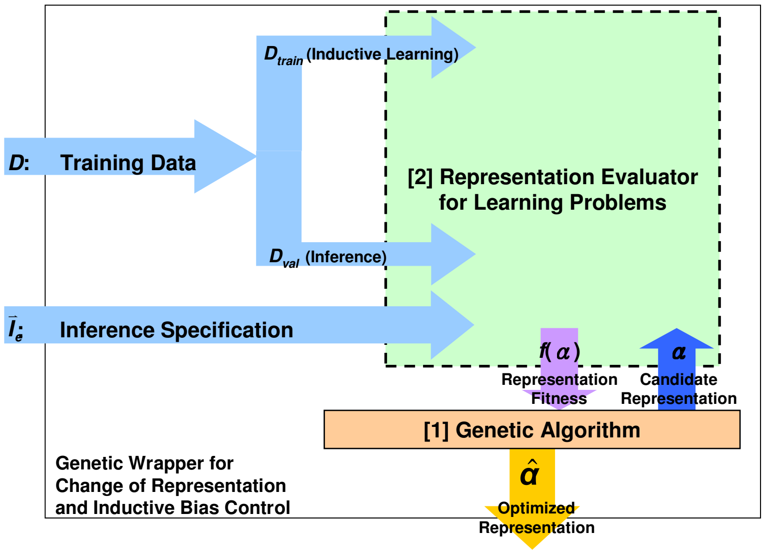
$$P(h / D) \propto P(D / h) \cdot P(h)$$

$$= P(h) \cdot \int P(D / h, \Theta) \cdot P(\Theta / h)\, d\Theta$$

Prior over Structures

Likelihood

Genetic Algorithms for Parameter Tuning in Learning



Terminology

- **Uncertain Reasoning**
  - Ability to perform inference in presence of uncertainty about
    - premises
    - rules
  - Nondeterminism
- **Representations for Uncertain Reasoning**
  - Probability: measure of belief in sentences
    - Founded on Kolmogorov axioms
    - prior, joint *vs.* conditional
    - Bayes's theorem: P(A | B) = (P(B | A) * P(A)) / P(B)
  - Graphical models: graph theory + probability
  - Dempster-Shafer theory: upper and lower probabilities, reserved belief
  - Fuzzy representation (sets), fuzzy logic: degree of membership
  - Others
    - Truth maintenance system: logic-based network representation
    - Endorsements: evidential reasoning mechanism

Summary Points

- **Last Class: Reasoning under Uncertainty and Probability**
  - Uncertainty is pervasive
    - Planning
    - Reasoning
    - Learning (later)
  - What are we uncertain about?
    - Sensor error
    - Incomplete or faulty domain theory
    - "Nondeterministic" environment
- **Today: Graphical Models**
- **Coming Week: More Applied Probability**
  - Graphical models as KR for uncertainty: Bayesian networks, *etc.*
  - Some inference algorithms for Bayes nets