Latent Tree Models and Approximate Inference in Bayesian Networks
Yi Wang, Nevin L. Zhang, Tao Chen — The Hong Kong University of Science and Technology

The Problem
- Probabilistic inference
  - Given Bayesian network $\mathcal{N}$ and evidence $E = e$
  - What is $P_X(Q|E = e)$?
- Intractable for general BNs

Our Contribution
- A novel approximate inference method

Latent Tree Models
- Tree-structured Bayesian networks
  - Manifest variables at leaf nodes are observed
  - Latent variables at internal nodes are hidden
- Two merits
  - Computationally simple
  - Model complex relationships among manifest variables

Structure Learning
- Basic ideas
  - In LTM, siblings are more closely correlated than nodes located far apart
  - If $\mathcal{M}$ approximates $\mathcal{N}$ well, then
    Nodes closely correlated in $\mathcal{M}$ are closely correlated in $\mathcal{N}$
- Introduce latent variables for closely correlated nodes in $\mathcal{N}$
- A hierarchical clustering procedure

Cardinalities of Latent Variables
- Set cardinalities at $C$
- Extreme case 1: Large $C (\geq \prod_{X \in X} |X|)$
  - Represents BN $\mathcal{N}$ exactly, best approximation
  - High inferential complexity
- Extreme case 2: $C = 1$
  - Poorest approximation of $\mathcal{N}$
  - Lowest online cost
- Extreme case 1 $\rightarrow$ Extreme case 2
  - Approximation accuracy decreases
  - Inferential efficiency improves

Offline Phase
- Inputs: (1) BN $\mathcal{N}$ over $X$, (2) Parameter $C$
- Output: An LTM $\mathcal{M}$
  - Uses $X$ as manifest variables
  - Cardinalities of latent variables upper bounded by $C$
  - Small KL divergence
  $$D[\mathcal{P}_N(X)||\mathcal{P}_M(X)] = \sum_X \mathcal{P}_N(X) \log \frac{\mathcal{P}_N(X)}{\mathcal{P}_M(X)}$$

Parameter Learning
- Find optimal parameter $\theta^* = \arg \min_{\theta} D[\mathcal{P}_N(X)||\mathcal{P}_M(X|\theta)]$
- Solve an equivalent problem
  - Generate data set $\mathcal{D}$ with $N$ samples from $\mathcal{N}$
  - Run EM to learn MLE $\hat{\theta} = \arg \max_{\theta} \log \mathcal{P}_M(\mathcal{D}|\theta)$
- As $N \rightarrow \infty$, $\hat{\theta}$ almost surely converges to $\theta^*$

Experimental Results
- Evaluated on 8 networks
  - Increase $C$ for
    - Higher approximation accuracy
    - Longer online phase
  - Versus clique tree propagation
    - Good approximation accuracy, average KL < 0.01
    - Low online cost, faster by one to two orders of magnitude
  - Versus loopy belief propagation
    - Comparable or higher approximation accuracy
    - Faster by five times to one order of magnitude

Conclusion
- A novel approximate method for probabilistic inference
  - Tradeoff between efficiency and accuracy by changing $C$
  - Good approximation accuracy at low online cost
  - Offline phase takes a long time due to EM algorithm
- Suitable for applications
  - Allow long offline phase
  - Demand good online performance