Towards Distributed Search for Graphical Models

Lars Otten
Dept. of Computer Science
University of California, Irvine
Scenario

- **Given is a graphical model and a query:**
  - Bayesian network with variables $X$ and conditional probability tables $P(x_i|par_i)$.
  - Query: find most probable explanation.
  
  
  \[
  MPE = \max_x P(x) = \max_x \prod_{i} P(x_i|par_i)
  \]
  
  - Problem is NP-hard.

- \begin{array}{|c|c|c|}
  \hline
  B & A & P(B|A) \\
  \hline
  0 & 0 & 0.2 \\
  1 & 0 & 0.8 \\
  0 & 1 & 0.0 \\
  1 & 1 & 1.0 \\
  \hline
\end{array}
Our Main Application

- **Computational Biology:**
  - Finding most likely *haplotypes* in general pedigrees.
  - Encoded as Bayesian network.
  - Corresponds to solving the MPE problem.
Reminder: Depth-First Search

- **Standard depth-first search procedure:**
  - Instantiate variables one at a time.
    - Backtrack in case of inconsistencies.
  - **Time complexity: \( exp(n) \).**
    - Linear space.
AND/OR Search Spaces [Mateescu&Dechter]

- Use decomposition and caching:
  - Time and space: \( \exp(w^*) \)
Distributed AND/OR Search
Parallel Search

- How can we search in parallel?
  - Research topic already in 1980's and 90's.
- Search tree: 'Cut off' at some depth (frontier).
  - Solve subtrees in parallel.
- Speedup at most linear.
Cutset Parallelization for AND/OR

• **Set parallelization frontier:**
  - Implies cutset in AND/OR search graph:
Redundancies Introduced

- Limited communication assumed.
  - Caching lost across subproblems.
  - Can introduce redundancies.
Cutset Parallelization

- \{A,B\} yields 8 subproblems.
  - Conditioning and decomposition!
Parallelization on a Computational Grid

- **Master node:**
  - Explores master search space through Branch and Bound and generates subproblems.

- **Worker nodes:**
  - Solve assigned subproblems, e.g., via vanilla AND/OR BaB.

- **Distributed computing, master-worker paradigm.**
  - distributed.net (1997), SETI@home (1999, now BOINC), Folding@home (2000), etc.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Distributed AND/OR Branch and Bound

- Central search space in gray.
  - Explored gradually through BaB.
Preliminary Experiments

- Bayesian networks of varying complexity:
  - Easy and complex pedigree instances.
  - Mastermind game instances.
- Sequential: 3.0 GHz system.
- Parallel: Condor workload management system.
  - 5, 10, or 15 dedicated worker nodes.
  - Each node with 2.33 - 3.0 GHz.
- http://graphmod.ics.uci.edu/~lars/cgi-bin/daoopt/webstats.py