Formula SampleSearch

Vibhav Gogate

Pedro Domingos

University of Washington

Rina Dechter

University of California, Irvine

**With help from Andrew Gelfand
University of California, Irvine**

Formula-based Probabilistic Inference

- Probabilistic Inference is variable-based
  - Variable/Bucket Elimination (Eliminate variables)
  - AND/OR search (condition on variables)
  - Importance/Gibbs sampling (sample variables)
- We change this viewpoint in our UAI 2010 paper
  - Eliminate/Condition/Sample assignments to formulas instead of variables!
  - Generalizes Variable-based inference
    - A variable is a unit clause.
  - Provably more efficient.

Formula-based search space

- Express a Markov network as a set of soft (weighted) (R) and hard (H) formulas.

$$Z = \sum_{x \in Sol(H)} \mathbf{exp}\left(\sum_i w_i \phi_i(x) \right) = \sum_{r \in R} \# Sol(H \wedge r) \times \mathbf{exp}\left(\sum_i w_i \phi_i(x_r) \right)$$

where $\phi_i(x)$ is $\mathbf{1}$ if x satisfies R_i and 0 otherwise.

Formula SampleSearch

- Given a distribution Q(R), we can rewrite Z as:

$$Z = \sum_{r \in R} \#Sol(H \wedge r) \times \mathbf{exp}\left(\sum_i w_i \phi_i(x_r) \right) \frac{Q(r)}{Q(r)}$$

$$Z = E_Q\left[\frac{\#Sol(H \wedge r) \times \mathbf{exp}\left(\sum_i w_i \phi_i(x_r) \right)}{Q(r)} \right]$$

Formula SampleSearch: Issues

$$Z = E_Q \left[\frac{\#Sol(H \wedge r) \times \mathbf{exp}\left(\sum_i w_i \phi_i(x_r)\right)}{Q(r)} \right]$$

- Rejection problem: #Sol(H ∧ r) might be zero
  - Use SampleSearch (Gogate and Dechter, AISTATS 2007).

- How to count #Sol(H ∧ r)?
  - sometimes exact counting is feasible
  - Use SampleSearch for model counting (Gogate and Dechter, AAAI 2007).

Implementation highlights

- Importance distribution from the output of IJGP (with i-bound =3)
- Rao-Blackwellisation or w-cutset sampling (Bidyuk and Dechter, 2007)
- Search implemented using the minisat solver (Sorenson and Een, 2003).
- Rejection control (Liu, 2001) this helped a lot!
  - Reject a sample if its weight is too low with some probability (Changhe and Druzdzel, 2007).
- **Tails matter (ε-cutoff)**: Replaced all small probabilities which are less than a threshold in Q(R) with the threshold (Changhe and Druzdzel, 2006).

References

- **Vibhav Gogate and Pedro Domingos (2010): Formula-based Probabilistic Inference, UAI-2010.**

- **Vibhav Gogate and Rina Dechter, "SampleSearch: A scheme that searches for consistent samples", AISTATS-2007.**

- **Vibhav Gogate and Rina Dechter, "Approximate Inference Algorithms for Hybrid Bayesian networks with Discrete constraints", UAI 2005.**