

Evaluating Probabilistic Reasoning Systems

Adnan Darwiche and Rina Dechter

The *probabilistic reasoning evaluation* took place during the month preceding the UAI conference. Its results will be presented at the applications workshop, and a full report on the methodology, benchmarks and results will be posted at the evaluation webpage following the UAI conference.

Motivation

Over the past two decades a variety of exact and approximate algorithms were developed across several communities (e.g. UAI, NIPS, SAT/CSPs) for answering optimization and likelihood queries over probabilistic graphical models. Since all these tasks are NP-hard, theoretical guarantees are rare and empirical evaluation becomes a central evaluation tool. Yet, the empirical comparison between algorithms requires agreement on representations, benchmarks and evaluation criteria which is challenging, especially in the context of approximation algorithms.

Some communities have already addressed similar challenges through yearly empirical evaluations and competitions (e.g. SAT, CSP and planning) which proved effective, leading to algorithmic advances and to software development and dissemination. We believe that such an effort could benefit probabilistic inference algorithms as well. Probabilistic reasoning presents additional challenges, however, as it tends to be harder, requires heterogenous knowledge representation frameworks, and must deal with the issue of evaluating approximate inference algorithms.

Goals

Our goal is to use the evaluation as a process that will help establish some standards for evaluating probabilistic reasoning systems based on both exact and approximate algorithms. Another long term goal is to reinforce a tradition of building and sharing probabilistic reasoning systems that allow easy access to state-of-the-art inference algorithms by members of the broader scientific and engineering communities. We hope to achieve a number of objectives:

- Increase the utilization of probabilistic inference algorithms in real-world applications by reducing the investment needed for building applications based on probabilistic reasoning.

- Allow newer members of the inference community to quickly capitalize on the expertise of more senior members of the community by providing broader access to existing code.
- Foster an environment where reported empirical results are accompanied by the very systems used to obtain them.

The actual UAI'08 *probabilistic reasoning evaluation* took place during the month preceding the conference and its results are presented and discussed during the applications workshop. The evaluation includes both Bayesian and Markov networks and consider three inference tasks: probability of evidence (partition function), most probable explanations (also called MPE or energy minimization), and node marginals. The evaluation will consider both exact and approximate algorithms, especially anytime algorithms that improve their approximations with time. Details of the evaluation can be found at:

<http://graphmod.ics.uci.edu/uai08/Evaluation>

A full report on the methodology, benchmarks and the results will be posted at the evaluation webpage following the conference.

Organizing Committee

- Fahiem Bacchus: <http://www.cs.toronto.edu/~fbacchus/>
- Jeff Bilmes: <http://ssli.ee.washington.edu/people/bilmes/>
- (co-chair) Adnan Darwiche: <http://www.cs.ucla.edu/~darwiche/>
- (co-chair) Rina Dechter: <http://www.ics.uci.edu/~dechter/>
- Hector Geffner: <http://www.tecn.upf.es/~hgeffner/>
- Alexander Ihler: <http://www.ics.uci.edu/~ihler/>
- Joris Mooij: <http://www.jorismooij.nl/>
- Kevin Murphy: <http://www.cs.ubc.ca/~murphyk/>