Exact and Approximate Weighted Model Integration Using Knowledge Compilation

Pedro Zuidberg Dos Martires, Anton Dries, Luc De Raedt

{firstname.lastname}@cs.kuleuven.be

Probabilistic Inference

Probabilistic inference algorithms are targeted towards:

- either continuous distributions: symbolic inference, Hamilton Monte Carlo, variational Bayesian Inference, ...
- or discrete distributions: SAT, weighted model counting, ...

We want to combine state-of-the-art from both
→ best of both worlds!

Weighted Model Integration

$$\texttt{working} \leftrightarrow (\texttt{cooling} \land (\texttt{t}^2 < 30)) \lor (\texttt{t} < 5)$$

$$p(\texttt{cooling}) = 0.99$$

$$\texttt{t} \sim \mathcal{N}_{\texttt{t}}(20, 5)$$

Question:

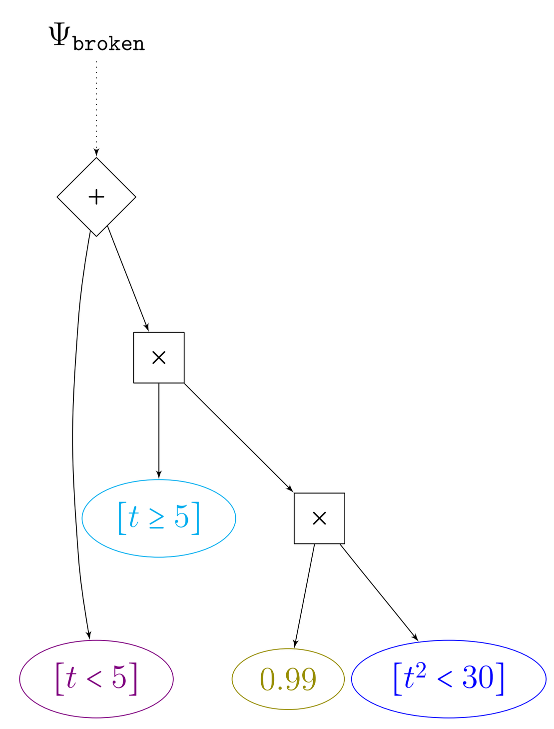
$$p(\texttt{working}) = ?$$

In general:

$$p(x|e) = \frac{p(e|x)p(x)}{\int_x p(x,e)}$$

Two Algorithms

Exact: Symbo



$$[\texttt{t}<5] + 0.99[\texttt{t}\geq 5][\texttt{t}^2<30]$$

$$p(\texttt{working}) = \int \left([\texttt{t}<5] + 0.99[\texttt{t}^2<30][\texttt{t}\geq 5]\right)\mathcal{N}_{\texttt{t}}(20,5)dt$$

Integrals become easily intractable.

Approximate: Sampo

$$t \approx [2.8, 35.1, 5.4, 22.2, 21.4]$$

$$\begin{bmatrix}[2.8<5]\\ [35.1<5]\\ [5.4<5]\\ [22.2<5]\\ [21.4<5]\end{bmatrix} + \begin{bmatrix}[2.8\geq 5]\\ [35.1\geq 5]\\ [5.4\geq 5]\\ [22.2\geq 5]\\ [21.4\geq 5]\end{bmatrix} \times \begin{bmatrix}0.99\\ 0.99\\ 0.99\\ 0.99\\ 0.99\end{bmatrix} \times \begin{bmatrix}[7.84<30]\\ [1232.01<30]\\ [29.16<30]\\ [492.84<30]\\ [457.96<30]\end{bmatrix}$$

$$= \begin{bmatrix}1\\ 0\\ 0\\ 0\\ 0\end{bmatrix} + \begin{bmatrix}0\\ 1\\ 1\\ 1\\ 1\end{bmatrix} \times \begin{bmatrix}0.99\\ 0.99\\ 0.99\\ 0.99\\ 0.99\end{bmatrix} \times \begin{bmatrix}1\\ 0\\ 1\\ 0\\ 0\end{bmatrix} = \begin{bmatrix}1\\ 0\\ 0.99\\ 0\\ 0\end{bmatrix}$$

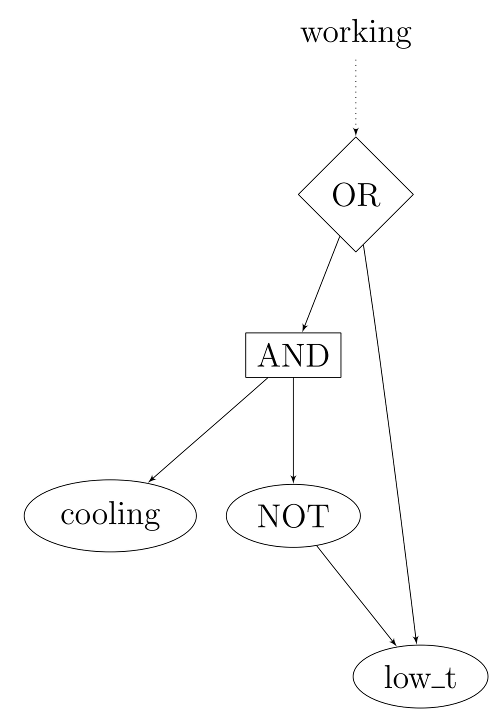
$$p(\texttt{working}) = \frac{1}{5}\sum_{i=1}^{5} \Psi^{MC}_{\texttt{broken},i} = 1.99/5 = 0.398$$

Pure vector calculus and can be executed on the GPU!
→ cheap probabilistic inference
→ embarrassingly parallelizable

Knowledge Compilation

offline: compile theory (expensive)
online: fast inference (cheap)

- evaluation in linear time
- conditioning in poly-time
- repeated querying



$$\texttt{working} \leftrightarrow \texttt{cooling} \lor \texttt{low\_t}$$

Algebraic Model Counting

Generalized framework for probabilistic inference:

- define specific semiring $(\mathcal{A}, \oplus, \otimes, e^{\oplus}, e^{\otimes})$ for specific task

Link to belief propagation:

- sum-product: $\oplus$ is normal addition
- max-product: $\oplus$ is maximization

We defined a custom probability density semiring with custom elements:

$$\mathcal{A} = \{(a, \mathcal{V}(a))\}$$

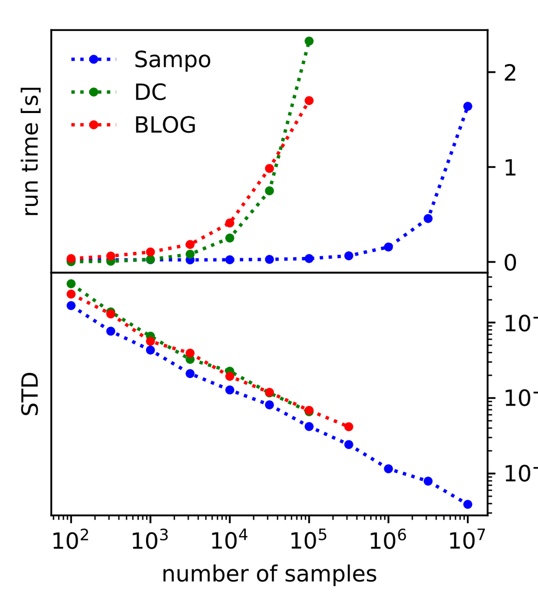
$$a = [\texttt{t}<5] + 0.99[\texttt{t}^2<30][\texttt{t}\geq 5]$$

$$V(a) = \{t\}$$

Results

- Symbo is faster on 9/10 benchmark problems than PSI, excluding knowledge compilation
- Symbo is faster on 7/10 benchmark problems than PSI, including knowledge compilation

Logical reasoning generally improves symbolic inference!



- **Sampling on the GPU → constant time complexity**
- **Avoid sampling categorical variables → reduction in variance**

Adnan Darwiche. "Modeling and reasoning with Bayesian networks"
Angelika Kimmig, Guy Van den Broeck, and Luc De Raedt. "Algebraic Model Counting"
Gehr, Misailovic, and Vechev, "PSI: Exact Symbolic Inference for Probabilistic Programs
Davide Nitti, Tinne De Laet, and Luc De Raedt. "Probabilistic logic programming for hybrid relational domains"
Brian Milch, Bhaskara Marthi, and Stuart Russell. "BLOG: Relational modeling with unknown objects"