Solving Marginal MAP Exactly by Probabilistic Circuit Transformations

YooJung Choi
Joint work with Tal Friedman and Guy Van den Broeck
Marginal MAP (MMAP)

Given a set of query variables $Q \subset X$ and evidence $e$,

$$\arg\max_{q \in \text{val}(Q)} p(q, e) = \arg\max_{q \in \text{val}(Q)} \sum_{h \in \text{val}(H)} p(q, h, e)$$

$\Rightarrow$ i.e. MAP of a marginal distribution on $Q$

$\Rightarrow$ in general, $\mathsf{NP}^{\mathsf{PP}}$-hard
MMAP on PCs

\[
\arg \max_{q \in \text{val}(Q)} p(q, e) = \arg \max_{q \in \text{val}(Q)} \sum_{h \in \text{val}(H)} p(q, h, e)
\]

- Smooth + decomposable ⇒ tractable marginal
  - e.g. \( p(X_1 = 1, X_2 = 0) \)

- + deterministic ⇒ tractable MAP
  - e.g. \( \max_{X_1X_2X_3} p(X_1, X_2, X_3) \)

- MMAP: **NP-hard** even for PCs that are tractable for marginals & MAP
  - Intuition: need the PC \( p(Q) \) to be deterministic
MMAP on PCs

\[
\text{arg max } p(q, e) = \text{arg max } \sum_{q \in \text{val}(Q)} p(q, h, e)
\]

- Smooth + decomposable $\Rightarrow$ tractable marginal
  - e.g. $p(X_1 = 1, X_2 = 0)$

- + deterministic $\Rightarrow$ tractable MAP
  - e.g. $\max_{X_1 X_2 X_3} p(X_1, X_2, X_3)$

- MMAP: \textbf{NP-hard} even for PCs that are tractable for marginals & MAP
  - Intuition: need the PC $p(Q)$ to be deterministic

$Q = \{X_1, X_2\}$
MMAP on PCs

• Enforce circuit constraints to get linear-time MMAP
  • E.g. constrained pseudo-tree (AND/OR search) [Marinescu, Dechter & Ihler ‘14], constrained vtree ((P)SDD) [Oztok, Choi & Darwiche ‘16]
  • Marginal determinism (aka Q-determinism)
    • Circuit size may blow up
    • Need to change the circuit for a different query variable set \( Q \)
• Branch-and-bound search [Huang, Chavira & Darwiche ‘06; Mei, Jiang & Tu ‘18]

Our approach: iterative circuit transformations
Bounds on MMAP

Upper bound through a single feedforward pass [Huang et al. '06]

Lower bound: $p(q)$ for any $q \in val(Q)$ works

Q: can we tighten these bounds further?

⇒ transform the PC to get better bounds
Circuit pruning for MMAP

Some parts of the circuit may be irrelevant for the MMAP solution

- Example: computing $p(X_1 = 1, X_2 = 0)$
  - Only the highlighted edges are used
  - Remaining edges propagate zero

- $X_1 = 1, X_2 = 0$ is the MMAP solution for $Q = \{X_1, X_2\}$
  - Pruning any black edge does not affect the MMAP solution

Q: can we efficiently identify which edges can be safely pruned?
Edge bounds for MMAP

For every edge, what is the maximum marginal probability $p(q)$ that uses/activates that edge?

\[ \forall (n, c): \text{define } \text{EB}(n, c) \geq \max_{q \in C_{n,c}} p(q) \]

\[ C_{n,c} = \{ q \in \text{val}(Q): p(q) \text{ "activates" edge } (n, c) \} \]

\[ (X_1 = 1, X_2 = 0) \in C_{n,c}, \quad (X_1 = 1, X_2 = 0) \notin C_{n,c} \]

\[ C_{r,1} = \{(X_1 = 1, X_2 = 0), (X_1 = 1, X_2 = 1)\} \]
Edge bounds for MMAP

For every edge, what is the maximum marginal probability \( p(q) \) that uses/activates that edge?

\[ \forall (n, c): \text{define } EB(n, c) \geq \max_{q \in C_{n,c}} p(q) \]

\( C_{n,c} = \{ q \in \text{val}(Q): p(q) \text{ "activates" edge (n, c)} \} \)

Given a lower bound \( l \) on MMAP, we can safely prune any edge \((n, c)\) if \( EB(n, c) < l \).

For a smooth and decomposable PC, all edge bounds can be computed with a single feedforward & backward pass through the circuit.
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Q-deterministic sum => max
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Backward pass: tighten $\text{EB}(n, c)$ at every Q-deterministic sum $n$
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Backward pass: tighten $\text{EB}(n, c)$ at every $Q$-deterministic sum $n$
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Backward pass: tighten $\text{EB}(n, c)$ at every $\text{Q}$-deterministic sum $n$
Feedforward pass: upper-bound on MMAP [Huang et al. '06]

Backward pass: tighten $EB(n,c)$ at every $Q$-deterministic sum $n$
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Backward pass: tighten $\text{EB}(n,c)$ at every $\mathbf{Q}$-deterministic sum $n$

e.g. using $p(X_1 = 0, X_2 = 1) = 0.256$ as lower bound
Feedforward pass: upper-bound on MMAP [Huang et al. ’06]

Backward pass: tighten $\mathbb{EB}(n, c)$ at every Q-deterministic sum $n$

e.g. using $p(X_1 = 0, X_2 = 1) = 0.256$ as lower bound
Iterative MMAP solver

- Prune edges
- Split on a query variable

Each split tightens the bound

After splitting on $Q_1, \ldots, Q_n$

$\Rightarrow$ linear-time MMAP for $Q = \{Q_1, \ldots, Q_n\}$
Empirical evaluation

Example run

Average run time in seconds (# instances solved)

| Dataset      | (30%, 30%, 40%) | (50%, 20%, 30%) |
|--------------|------------------|------------------|
|              | MaxSPN (ours)    | MaxSPN (ours)    |
| NLTCS        | 0.004 (10)       | 0.01 (10)        |
| MSNBC       | 0.01 (10)        | 0.03 (10)        |
| KDD          | 0.02 (10)        | 0.04 (10)        |
| Plants       | 0.27 (10)        | 2.95 (10)        |
| Audio        | 188.59 (10)      | 2041.33 (6)      |
| Jester       | 265.50 (10)      | 2913.04 (2)      |
| Netflix      | 344.71 (10)      | - (0)            |
| Accidents    | 0.54 (10)        | 109.56 (10)      |
| Retail       | 0.03 (10)        | 0.06 (10)        |
| Pumsb-star   | 273.70 (10)      | 2208.27 (7)      |
| DNA          | 2809.44 (4)      | 505.75 (9)       |
| Kosarek      | 1.60 (10)        | 48.74 (10)       |
| MSWeb        | 25.70 (10)       | 1543.49 (10)     |
| Book         | - (0)            | 7.25 (10)        |
| EachMovie    | - (0)            | 93.66 (10)       |
| WebKB        | - (0)            | 102.37 (10)      |
| Reuters-52   | - (0)            | 22.91 (10)       |
| 20 NewsGrp.  | - (0)            | 88.13 (10)       |
| BBC          | - (0)            | 766.93 (9)       |
| Ad           | - (0)            | 344.81 (10)      |

Total Solved | 124 | 199 | 105 | 187
Conclusion

• Iterative pruning and splitting to tighten MMAP bounds
  • Each split may (worst-case) double the circuit size, but pruning can be effective *in practice*

• Also an iterative MPE solver for non-deterministic PCs

• Can we generalize the bounds to other queries that require determinism for tractability?