Towards Hardware-Aware Tractable Learning of Probabilistic Models

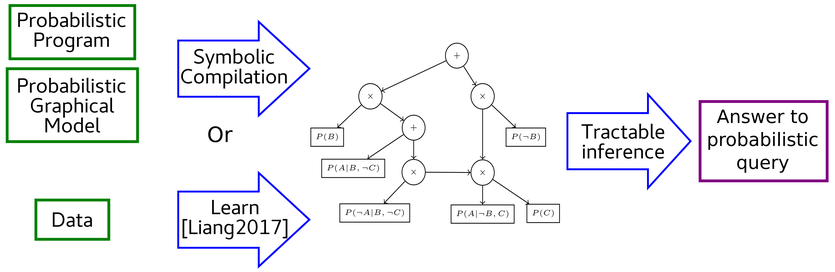
Laura I. Galindez Olascoaga, Wannes Meert, Nimish Shah, Marian Verhelst and Guy Van den Broeck

Background

Tractable probabilistic models possess a set of traits that make them ideal for embedded sensory applications:

-Robustness to missing data allows them to cope with sensor failure.
-Small data needs allow them to adapt to different users quickly.
-Tractability enables reliable inference under constrained resource budgets.

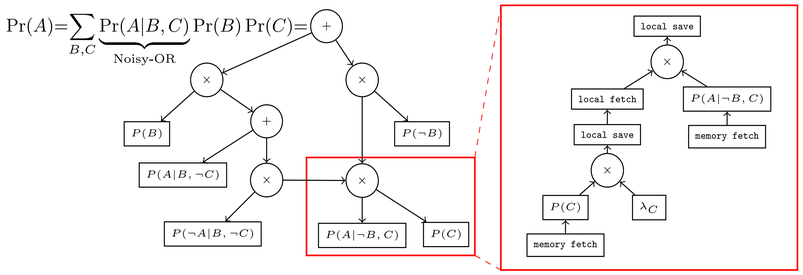
Arithmetic Circuits are one of those tractable representations [Darwiche2009]:



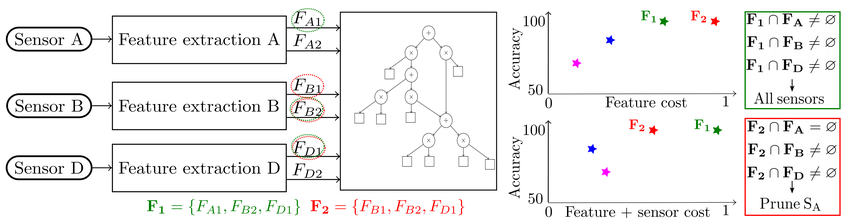
Motivation

Limitations of the state of the art:

1. Tractability notions disregard hardware implementation details and are given in abstract terms (e.g. time and space).



2. Traditional resource aware techniques do not exploit hardware scaling opportunities (e.g. turning sensors off, reducing bits and precision).



Hardware-aware cost

The properties of Arithmetic Circuits allow us to reliably predict resource usage. We propose the Hardware-aware cost, given in terms of a specific hardware resource (e.g. energy consumption).

Scalable system properies:
- (1) Model complexity α
- (2) Feature set $\mathbf{F}$
- (3) Sensor set S
- (4) Precision nb

System configuration $\sigma = \{\alpha, \mathbf{F}, S, nb\}$

Hardware-aware inference cost

$$\mathrm{C_{HA}}(\alpha, \mathbf{S}, \mathbf{F}, nb) = \mathrm{C_{AC}}(\alpha, nb) + \sum_{S \in \mathbf{S}} \mathrm{C_{SI}}(S, \mathbf{F}_S)$$

Sensor interfacing cost

$$\mathrm{C_{SI}}(S, \mathbf{F}) = \mathrm{C_S}(S) + \sum_{F_S \in \mathbf{F}_S} \mathrm{C_F}(F_S)$$

Inference cost

$$\mathrm{C_{AC}}(\alpha, nb) = \mathrm{C_+}(nb) + \mathrm{C_\times}(nb) + \mathrm{C_{mem}}(nb) + \mathrm{C_{cache}}(nb)$$

Map to trade-off space

Goal

Find the optimal configuration for a given accuracy and cost σ^*

Strategy

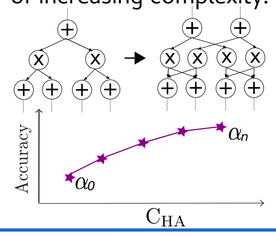
Very hard optimization problem (>NP): we rely on a sequence of greedy searches.

Trade-off space search and optimal configuration selection

Scale each system property individually and get Pareto optimal after each stage.

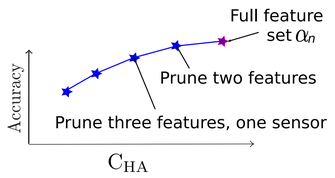
(1) Model complexity α

Learn a set of models α of increasing complexity.



(2) Feature set $\mathbf{F}$ (3) Sensor set S

Greedy feature and sensor pruning algorithm for each model in α that fulfills $\max Acc/\mathrm{C_{HA}}$



(4) Precision nb

Standard IEEE floating point: 64, 32, 16, 8 bits.

References

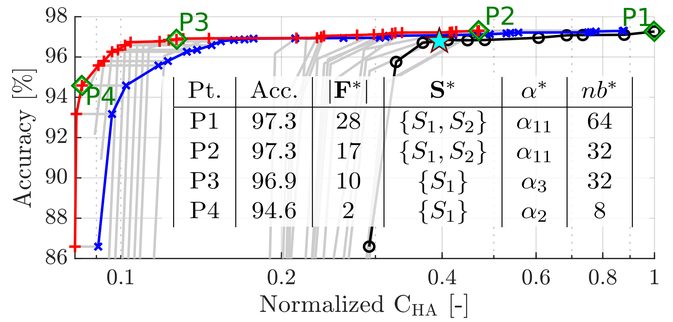
[Darwiche2009]:A. Darwiche. Modeling and Reasoning with Bayesian Networks. Cambridge University Press, 2009.

[Liang2017]: Y. Liang, J. Bekker and G. Van den Broeck. Learning the Structure of Probabilistic Sentential Decision Diagrams. UAI 2017.

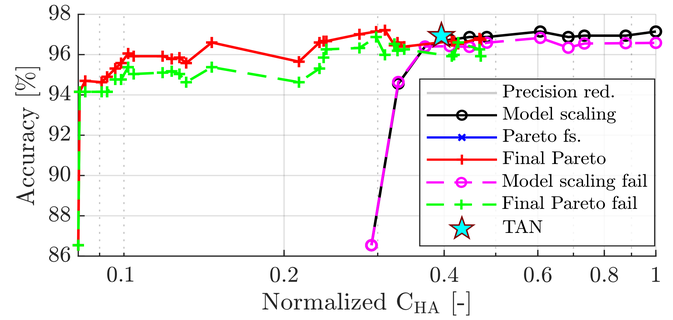
Experiments

Experiments on a smartphone-based Human Activity Recognition benchmark with gyroscope and accelerometer sensor signals, and statistical features. The cost is given in terms of relative energy consumption.

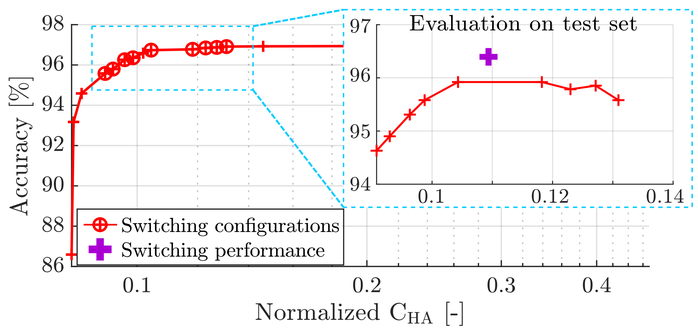
Potential cost savings of up to 90 % with accuracy loss of less than 1 % enabled by a combination of feature and sensor pruning, model simplification and precision reduction.



Pt.	Acc.	$\lvert\mathbf{F}^*\rvert$	$\mathbf{S}^*$	α^*	nb^*
P1	97.3	28	$\{S_1, S_2\}$	α_{11}	64
P2	97.3	17	$\{S_1, S_2\}$	α_{11}	32
P3	96.9	10	$\{S_1\}$	α_3	32
P4	94.6	2	$\{S_1\}$	α_2	8



Potential use for on-line scenarios where resources and performance must be traded-off dynamically



Contact: Laura I. Galindez Olascoaga (laura.galindez@esat.kuleuven.be)

NeurIPS 2019