Reducing the Cost of Probabilistic Knowledge Compilation

Giso H. Dal, Steffen Michels and Peter J.F. Lucas
Radboud University, Nijmegen (The Netherlands)

Outline

Outline

The Problem

Exact inference is nice,
but...

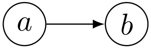
knowledge compilation is
computationally intensive.

Outline

Running Example

a	b	$P(X)$
1	1	0.4
1	2	0.4
2	1	0.05
2	2	0.05
3	1	0
3	2	0.1

Joint distribution



$$P(X) = P(b|a)P(a)$$

Bayesian network

$P(a=1)$	$P(a=2)$	$P(a=3)$
0.8	0.1	0.1

a	$P(b=1\|a)$	$P(b=2\|a)$
1	0.5	0.5
2	0.5	0.5
3	0	1

CPTs

Encoding

Let a BN be defined over variables x. We encode it as Boolean function f by adding for $x \in X$:

$$\underbrace{(x_1 \vee \cdots \vee x_n)}_{\text{at-least-once}} \wedge \underbrace{\bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{n} (\overline{x_i} \vee \overline{x_j})}_{\text{at-most-once}} .$$

Unique symbolic weights ω_j identify distinct probabilities local to x's CPT. We introduce into f clauses that imply each weight ω_j.

Encoding

$P(a=1)$	$P(a=2)$	$P(a=3)$
0.8	0.1	0.1

a	$P(b=1 \mid a)$	$P(b=2 \mid a)$
1	0.5	0.5
2	0.5	0.5
3	0	1

Encoding

$P(a=1)$	$P(a=2)$	$P(a=3)$
0.8	0.1	0.1

a	$P(b=1 \mid a)$	$P(b=2 \mid a)$
1	0.5	0.5
2	0.5	0.5
3	0	1

$P(a=1)$	$P(a=2)$	$P(a=3)$
ω_1	ω_2	ω_2

a	$P(b=1 \mid a)$	$P(b=2 \mid a)$
1	ω_3	ω_3
2	ω_3	ω_3
3	ω_4	ω_5

Encoding

$P(a=1)$	$P(a=2)$	$P(a=3)$
0.8	0.1	0.1

$P(a=1)$	$P(a=2)$	$P(a=3)$
ω_1	ω_2	ω_2

a	$P(b=1\|a)$	$P(b=2\|a)$
1	0.5	0.5
2	0.5	0.5
3	0	1

a	$P(b=1\|a)$	$P(b=2\|a)$
1	ω_3	ω_3
2	ω_3	ω_3
3	ω_4	ω_5

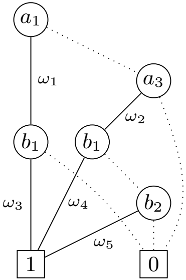
Variables:

$$(a_1 \vee a_2 \vee a_3) \wedge (\overline{a_1} \vee \overline{a_2}) \wedge (\overline{a_1} \vee \overline{a_3}) \wedge (\overline{a_2} \vee \overline{a_3}) \wedge (b_1 \vee b_2) \wedge (\overline{b_1} \vee \overline{b_2})$$

Probabilities:

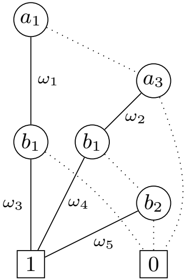
$$(\overline{a_1} \vee \omega_1) \wedge (\overline{a_2} \vee \omega_2) \wedge (\overline{a_3} \vee \omega_2) \wedge (\overline{a_1} \vee \overline{b_1} \vee \omega_3) \wedge (\overline{a_1} \vee \overline{b_2} \vee \omega_3) \wedge (\overline{a_2} \vee \overline{b_1} \vee \omega_3) \wedge (\overline{a_2} \vee \overline{b_2} \vee \omega_3) \wedge (\overline{a_3} \vee \overline{b_1} \vee \omega_4) \wedge (\overline{a_3} \vee \overline{b_2} \vee \omega_5).$$

Inference by Weighted Model Counting

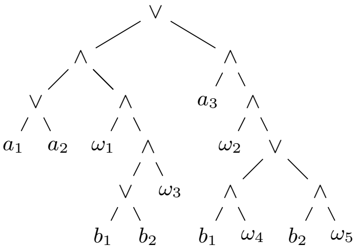


WPBDD

Inference by Weighted Model Counting

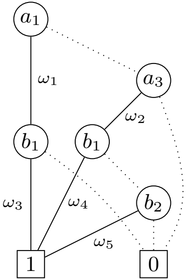


WPBDD

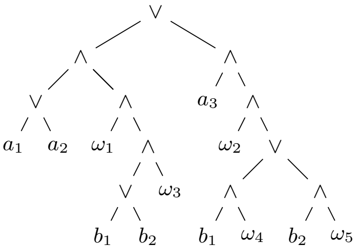


Logical circuit

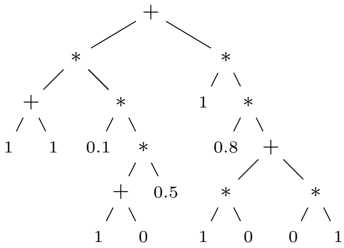
Inference by Weighted Model Counting



WPBDD



Logical circuit



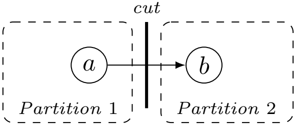
Instantiated arithmetic circuit for $P(b = 1)$

Outline

The Framework

1. Partition
2. Compilation
3. Assembly
4. Inference

Partition and Compile

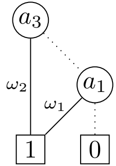


Partitioning

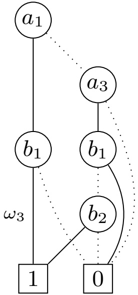
$P(a=1)$	$P(a=2)$	$P(a=3)$
ω_1	ω_2	ω_2

a	$P(b{=}1 \mid a)$	$P(b{=}2 \mid a)$
1	ω_3	ω_3
2	ω_3	ω_3
3	ω_4	ω_5

Capture symbolic structure in CPTs



Partition 1. Compilation with order: $a_3 \leq a_2 \leq a_1$



Partition 2. Compilation with order: $a_1 \leq a_2 \leq a_3 \leq b_1 \leq b_2$

Partition and Compile: What to compile?

Partition 1:

$$(a_1 \vee a_2 \vee a_3) \wedge (\overline{a_1} \vee \overline{a_2}) \wedge (\overline{a_1} \vee \overline{a_3}) \wedge (\overline{a_2} \vee \overline{a_3}) \wedge$$

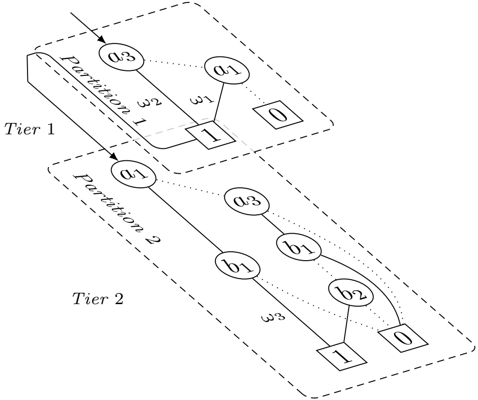
$$(\overline{a_1} \vee \omega_1) \wedge (\overline{a_2} \vee \omega_2) \wedge (\overline{a_3} \vee \omega_2)$$

Partition 2:

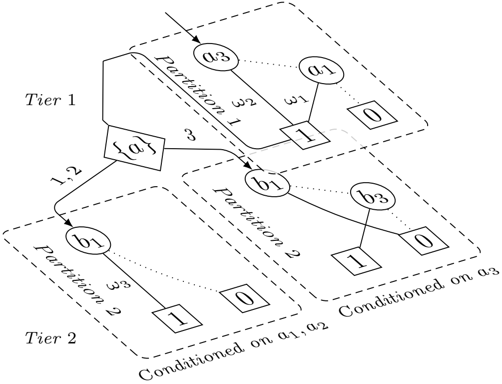
$$(a_1 \vee a_2 \vee a_3) \wedge (\overline{a_1} \vee \overline{a_2}) \wedge (\overline{a_1} \vee \overline{a_3}) \wedge (\overline{a_2} \vee \overline{a_3}) \wedge$$
$$(b_1 \vee b_2) \wedge (\overline{b_1} \vee \overline{b_2}) \wedge$$

$$(\overline{a_1} \vee \overline{b_1} \vee \omega_3) \wedge (\overline{a_1} \vee \overline{b_2} \vee \omega_3) \wedge (\overline{a_2} \vee \overline{b_1} \vee \omega_3) \wedge$$
$$(\overline{a_2} \vee \overline{b_2} \vee \omega_3) \wedge (\overline{a_3} \vee \overline{b_1} \vee \omega_4) \wedge (\overline{a_3} \vee \overline{b_2} \vee \omega_5).$$

Assemby and Inference



Dynamic conditioning



Static conditioning

Outline

An upperbound

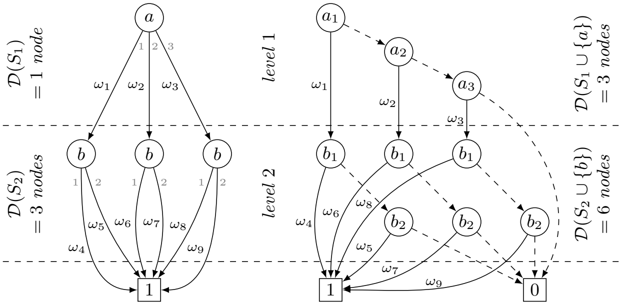
Let a Bayesian network be defined over n variables X.

1. We formulate it as a set of *constraints* $C_x \in C$, where C_x represents the dependencies of variable $x \in X$:
$$C_x = \{x\} \cup \textsc{Parents}(x).$$
2. Compile using ordering $O = \{o_1, \ldots, o_n\}$ imposed on X by π, where $o_i = \pi(x)$, $o_j \leq o_k$ for $j < k$.
3. The upperbound is based on *spanning variables* $S_l \in S$ associated with each evaluation depth, or level l:
$$S_l = \{o_1, \ldots, o_{l-1}\} \ \cap \bigcup_{\{C_x \in C \ : \ \{o_1, \ldots, o_{l-1}\} \subset C_x\}} C_x$$
4. A generalized upperbound is computed by:
$$\sum_{l=1}^{n} \mathcal{D}(S_l), \text{ where } \mathcal{D}(Y) = \prod_{x \in Y} |x|.$$

An Upperbound

$P(a=1)$	$P(a=2)$	$P(a=3)$
ω_1	ω_2	ω_3

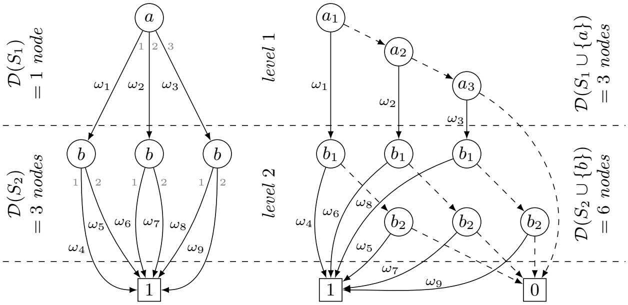
a	$P(b=1\|a)$	$P(b=2\|a)$
1	ω_4	ω_5
2	ω_6	ω_7
3	ω_8	ω_9



An Upperbound

$P(a{=}1)$	$P(a{=}2)$	$P(a{=}3)$
ω_1	ω_2	ω_3

a	$P(b{=}1 \mid a)$	$P(b{=}2 \mid a)$
1	ω_4	ω_5
2	ω_6	ω_7
3	ω_8	ω_9

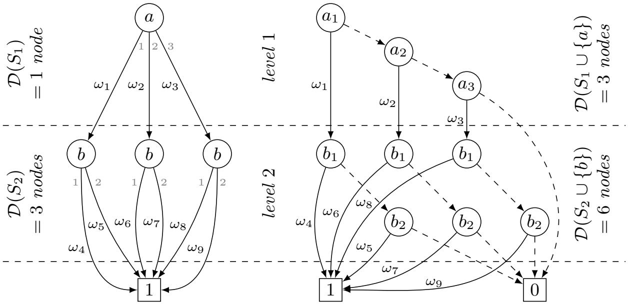


$$C_a = \{a\}, C_b = \{a, b\}.$$

An Upperbound

$P(a{=}1)$	$P(a{=}2)$	$P(a{=}3)$
ω_1	ω_2	ω_3

a	$P(b{=}1\|a)$	$P(b{=}2\|a)$
1	ω_4	ω_5
2	ω_6	ω_7
3	ω_8	ω_9



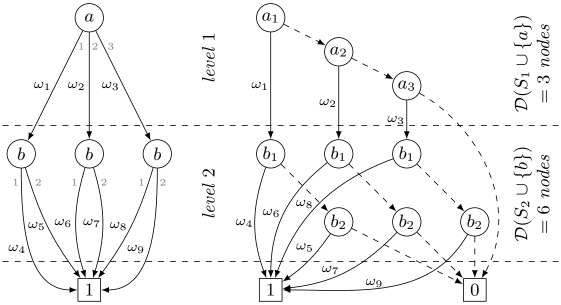
$$C = \; C_a = \{a\}, C_b = \{a, b\}.$$

l	O	S_l	$\mathcal{D}(S_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i)$	$\mathcal{D}(S_l \cup O_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i \cup O_l)$
1	a	$\{\}$				
2	b	$\{a\}$				

An Upperbound

$P(a{=}1)$	$P(a{=}2)$	$P(a{=}3)$
ω_1	ω_2	ω_3

a	$P(b{=}1 \mid a)$	$P(b{=}2 \mid a)$
1	ω_4	ω_5
2	ω_6	ω_7
3	ω_8	ω_9



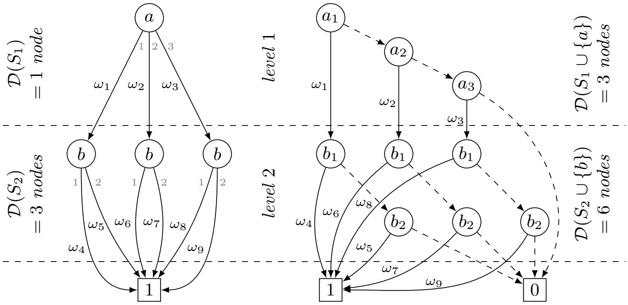
$$C = \; C_a = \{a\}, C_b = \{a, b\}.$$

l	O	S_l	$\mathcal{D}(S_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i)$	$\mathcal{D}(S_l \cup O_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i \cup O_l)$
1	a	$\{\}$	1	1	3	3
2	b	$\{a\}$				

An Upperbound

$P(a{=}1)$	$P(a{=}2)$	$P(a{=}3)$
ω_1	ω_2	ω_3

a	$P(b{=}1 \mid a)$	$P(b{=}2 \mid a)$
1	ω_4	ω_5
2	ω_6	ω_7
3	ω_8	ω_9



$$C = \quad C_a = \{a\}, C_b = \{a, b\}.$$

l	O	S_l	$\mathcal{D}(S_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i)$	$\mathcal{D}(S_l \cup O_l)$	$\sum_{i=1}^{l+1} \mathcal{D}(S_i \cup O_l)$
1	a	$\{\}$	1	1	3	3
2	b	$\{a\}$	3	4	6	9

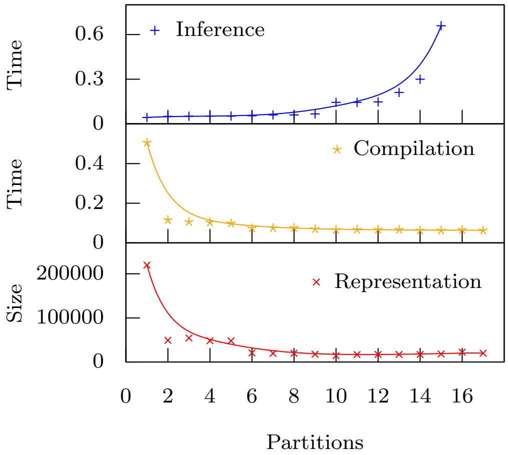
Outline

Compilation Time

Bayesian Network	X	$\mathcal{A}(X)$	P	WPBDD S	WPBDD T	WPBDD S	WPBDD T	SDD S	SDD T	OBDD S	OBDD T
insurance	27	89	2	33183	0.014	348956	0.077	1731415	2.046	1263540	0.289
weeduk	15	90	2	30735	0.406	30733	0.418	-	-	109734	0.176
alarm	37	105	2	2730	0.003	3788	0.003	35004	0.054	10008	0.003
water	32	116	2	49212	0.095	219797	0.506	-	-	-	-
powerplant	40	120	2	2451	0.002	4158	0.003	26662	0.038	11043	0.002
carpo	54	122	2	1937	0.003	2377	0.003	13405	0.028	7179	0.003
win95pts	76	152	2	49405	0.018	810957	0.361	1109210	2.057	4876152	4.997
hepar2	70	162	2	33234	0.025	56574	0.033	188453	2.423	142806	0.161
fungiuk	15	165	2	79682	1.515	234322	7.763	-	-	733551	0.812
hailfinder	56	223	2	225325	0.052	4025502	1.395	10508499	7.51	31493220	12.435
3nt	58	228	2	9844	0.015	858645	0.578	42774722	58.905	15592962	19.493
4sp	58	246	2	83156	0.035	918353	0.352	-	-	20558352	34.598
barley	48	421	2	13721258	11.197	-	-	-	-	-	-
mainuk	48	421	2	9045244	9.002	-	-	-	-	-	-
andes	220	440	2	426513	0.117	-	-	-	-	-	-
pathfinder	135	520	2	143032	0.493	577163	0.717	2287777	23.337	5732988	18.656
mildew	35	616	2	1634250	111.434	5666709	113.264	-	-	-	-
munin1	186	992	4	13196919	6.693	-	-	-	-	-	-
pigs	441	1323	9	3292450	1.534	-	-	-	-	-	-

Compilation cost,
where X and $\mathcal{A}(X)$ are the number of variables in the BN and encoding, S is the number of logical operators that the symbolic representation induces, time T is in seconds, P is number of partitions, and - implies compilation failure due to memory requirements.

Partition vs. Inference Time



Effects of partitioning on the *water* network.

Closing Remarks

- The compilation cost is drastically reduced.
- The representations obtained are much smaller.
- Independent partition orderings increase structure exploitation.
- WMC for huge networks?

Questions?

Email me at
gdal **at** cs.ru.nl
for any further questions.