Knowledge Representation and Learning

Weighted Model Counting

Luciano Serafini

FBK, Trento, Italy

July 12, 2023

Reasoning tasks on Propositional Logic

Task Name	Input	Output
Model checking:	ϕ,$\mathcal{I}$	$\mathcal{I}(\phi)$
Satisfiability:	ϕ	$\max_{\mathcal{I}} \mathcal{I}(\phi)$
Maximum Satisfiability:	ϕ, w :	$\max_{\mathcal{I}} \mathcal{I}(\phi) \cdot w(\mathcal{I})$
Model counting:	ϕ	$\sum_{\mathcal{I}} \mathcal{I}(\phi)$
Weighted model counting:	ϕ, w	$\sum_{\mathcal{I}} \mathcal{I}(\phi) \cdot w(\mathcal{I})$

Definiton of Weighted Model Counting

Definition (Weighted model counting)

Let $\mathcal{P}$ be a set of propositional variables. Given a *weight function* $w : \{0,1\}^{|\mathcal{P}|} \rightarrow \mathbb{R}^+$, the problem of weighted model counting is the problem of computing the summation of the weights of the models that satisfies a formula ϕ.

$$\mathrm{WMC}(\phi, w) = \sum_{\mathcal{I}\in\{0,1\}^{|\mathcal{P}|}} w(\mathcal{I}) \cdot \mathcal{I}(\phi)$$

An alternative and equivalent formulation of weighted model counting is the following:

$$\mathrm{WMC}(\phi, w) = \sum_{\substack{\mathcal{I}\in\{0,1\}^{|\mathcal{P}|} \\ \mathcal{I}\models\phi}} w(\mathcal{I})$$

Example

Example

Suppose that we log what people buy in a supermarket:

#	*Itemsets*						
4	a	b	c	d			
1	a	b			e	f	
7	a	b	c				
3	a		c	d		f	
2							g
1				d			
4				d			g

- Every combination of items can be seen as an interpretation on the set of propositions $a, b, \dots g$. and the number of times we observe such a combination could be considered the weight of the model.

- We have 2^7 possible itemsets (interpretations $\mathcal{I}$), and we can assigns to each a weight $w(\mathcal{I})$ which is s the number of times an itemset has been observed.

Example

Example

$\mathcal{I}$							
a	b	c	d	e	f	g	$w(\mathcal{I})$
1	1	1	1	0	0	0	4
1	1	0	0	1	1	0	1
1	1	1	0	0	0	0	7
1	0	1	1	0	1	0	3
0	0	0	0	0	0	1	2
0	0	0	0	1	0	0	1
0	0	0	0	1	0	1	4

$$
\begin{aligned}
\textsc{wmc}(a \wedge (b \vee c)) &= 4 + 1 + 7 + 3 = 15 \\
\textsc{wmc}(a \wedge g) &= 0 \\
\textsc{wmc}(a \wedge \neg g) &= 4 + 1 + 7 + 3 = 15 \\
\textsc{wmc}(a \rightarrow b) &= 4 + 1 + 7 + 2 + 1 + 4 = 19
\end{aligned}
$$

Model counting vs. Weighted model counting

- in model counting each interpretation weights 1;
- In WMC instead, some models are more important than others, and it makes sense to associate a weight $w(\mathcal{I}) \geq 0$ to each interpretation $\mathcal{I}$.
- in weighted model counting each model of a formula counts for its weight $w(\mathcal{I})$
- this interpretation of weighted models can be used to represent some form of uncertainty about the world. E.g., by associating probability of a formula to be true.
- the weight $w(\mathcal{I})$ associated to the model $\mathcal{I}$ can be interpreted in probabilistically; i.e., the higher the weight of a model the more likely the model;

Weighted model counting vs. MaxSAT

- Weight functions have been defined also in MaxSAT but there are some differences:
- In MaxSAT we allow negative weights, in WMC we don't
- in MaxSAT Weights are used for defining an order on the interpretations;
- the nominal value of the weight function is not important
- two weight function are equivalent for MaxSAT if they define the same order on interpretations.
- in weighted model counting instead we are really interested in the nominal value of the weight of an interpretation.

The partition function $Z(w)$

Proposition

If ϕ is valid, then $\text{WMC}(\phi, w)$ is equal to $\sum_{\mathcal{I}:\mathcal{P}\rightarrow\{0,1\}} w(\mathcal{I})$

- The quantity $\sum_{\mathcal{I}:\mathcal{P}\rightarrow\{0,1\}} w(\mathcal{I})$ is called partition function of w.

$$Z(w) = \sum_{\mathcal{I}} w(\mathcal{I}) \tag{1}$$

- Computing $Z(w)$ is a source of complexity. In general we have to compute $w(\mathcal{I})$ for all the 2^n interpretations

Specifying $W : \{0,1\}^{|\mathcal{P}|} \rightarrow \mathbb{R}^+$

What is a compact way to represent the weight funciton?

- To explicitly defining the weights for each interpretation we need $2^{|\mathcal{P}|}$ parameters;
- Alternatively one can select n formulas $\phi_1, \ldots, \phi_n$ and associate a weight to each one $w_1, \ldots, w_n$, and define

$$w(\mathcal{I}) = \prod_{\mathcal{I} \models \phi_i} w_i \tag{2}$$

  or alternatively

$$w(\mathcal{I}) = \exp\left(\sum_{\mathcal{I} \models \phi_i} w_i'\right) \tag{3}$$

- There is no free lunch. There are weight function that cannot be defined with less then $2^{|\mathcal{P}|}$ formulas.
- But in many cases it is possible. In this cases we say that w factorizes w.r.t., $\phi_1, \ldots, \phi_n$.

Specifying $W : \{0,1\}^{|\mathcal{P}|} \to \mathbb{R}^+$

Example

Consider the following two weight functions

p	q	$w(\mathcal{I})$
0	0	1.0
0	1	2.0
1	0	3.0
1	1	6.0

p	q	$w(\mathcal{I})$
0	0	2.0
0	1	3.0
1	0	5.0
1	1	7.0

- The left weight function can be expressed using two weighted formulas; i.e. $3 : p$ and $2 : q$ using definition (2), indeed the weight of the model that satisfies both p and q is the product of the weight of p and q, so we say that it factorizes)
- The second can be expressed with the weighted formulas $p \vee q : 2$,

Specifying $W : \{0,1\}^{|\mathcal{P}|} \rightarrow \mathbb{R}^+$ by literals

Specifying weights on literals

$$w(\mathcal{I}) = \prod_{p \in P} w(p)^{\mathcal{I}(p)} \cdot w(\neg p)^{1-\mathcal{I}(p)}$$

$$WMC(\phi, w) = \sum_{\mathcal{I} \models \phi} \prod_{p \in P} w(p)^{\mathcal{I}(p)} \cdot w(\neg p)^{1-\mathcal{I}(p)}$$

$$= \sum_{\mathcal{I} \models \phi} \exp\left(\sum_{p \in P} v(p) \cdot \mathcal{I}(p) + v(\neg p) \cdot (1 - \mathcal{I}(p)) \right)$$

where $w : Lit \rightarrow \mathbb{R}^+$ is a mapping from the set of literals (i.e., p and $\neg p$ for p propositional variable) to positive real numbers. ($v(\cdot) = \log(W(\cdot))$)

Weighted Model counting

Example

w

$$
\begin{aligned}
p &\to 1.2 \\
\neg p &\to 3.4 \\
q &\to 3.2 \\
\neg q &\to 1.0 \\
r &\to 0.4 \\
\neg r &\to 0.6
\end{aligned}
$$

p	q	r	$w(x)^{MCh(\mathcal{I},x)} w(\neg x)^{MCh(\mathcal{I},\neg x)}$						$w(\mathcal{I})$	$Pr(\mathcal{I})$
0	0	0	1	3.4	1	1.0	1	0.6	2.04	0.11
0	0	1	1	3.4	1	1.0	0.4	1	1.36	0.07
0	1	0	1	3.4	3.2	1	1	0.6	6.528	0.34
0	1	1	1	3.4	3.2	1	0.4	1	4.352	0.23
1	0	0	1.2	1	1	1.0	1	0.6	0.72	0.04
1	0	1	1.2	1	1	1.0	0.4	1	0.48	0.02
1	1	0	1.2	1	3.2	1	1	0.6	2.304	0.12
1	1	1	1.2	1	3.2	1	0.4	1	1.536	0.08

$$WMC(p \vee \neg q \to r) = w(\,001\,) + w(\,010\,) + w(\,011\,) + w(\,101\,) + w(\,111\,) \approx 14.26$$

$$WMC(\top) = w(\,000\,) + w(\,001\,) + \cdots + w(\,111\,) \approx 19.32$$

$$Pr(p \vee \neg q \to r) = \frac{WMC(p \vee \neg q \to r)}{WMC(\top)} \approx \frac{14.26}{19.32} \approx 0.74$$

Weighted Model counting

Examples (Weights can be associated also to formulas)

w	Δ defines fresh variables	w'
$\neg(p \vee q) \rightarrow 0.0$	$f_0 \leftrightarrow \neg(p \vee q)$	$f_0 \rightarrow 0.0$
$p \rightarrow 0.1$	$f_1 \leftrightarrow p$	$f_1 \rightarrow 0.1$
$p \vee r \rightarrow 1.2$	$f_2 \leftrightarrow p \vee r$	$f_2 \rightarrow 1.2$
$q \rightarrow r \rightarrow 2.5$	$f_3 \leftrightarrow q \rightarrow r$	$f_3 \rightarrow 2.5$

$$
\begin{aligned}
&WMC(p \vee \neg q \rightarrow r \wedge \Delta) = \\
&w(0011011) + w(0100000) + w(0110011) + w(1010111) + w(1110111) = \\
&0 + 1 + 3 + 0.3 + 0.3 = 4.6
\end{aligned}
$$

$$
\begin{aligned}
WMC(\Delta) &= w(0001001) + w(0011011) + w(0100000) + w(0110011) \\
&+ w(1000111) + w(1010111) + w(1100110) + w(1110111) \\
&= 0 + 0 + 1 + 3 + 0.3 + 0.3 + 0.12 + 0.3 = 5.02
\end{aligned}
$$

$$
Pr(p \vee \neg q \rightarrow r | \Delta) = \frac{WMC(p \vee \neg q \rightarrow r \wedge \Delta)}{WMC(\Delta)} = \frac{4.6}{5.02} \approx 0.92
$$

Algorithm for Weighted Model Counting

- Exact method based on knowledge compilation. Generalization of model counting algorithm
- Approximated methods (not covered in the course): based on rectangular approximation[1] or by reducing it to (unweighted) model counting[2]. See[3] for a survey.

---

[1] Ermon et al. 2013.

[2] Colnet and Meel 2019.

[3] Chakraborty, Meel, and Vardi 2021.

Properties of WMC

Let w be a weight funciton on the set of propositinal variables of ϕ and ψ.

1. If ϕ and ψ do not contain common propositional variables ($\phi \wedge \psi$ is decomposable) then:

$$\text{WMC}(\phi \wedge \psi, w) = \text{WMC}(\phi, w|_{\mathcal{P}(\phi)}) \cdot \text{WMC}(\psi, w|_{\mathcal{P}(\psi)})$$

2. If $\phi \wedge \psi$ is unsatisfiable ($\phi \vee \psi$ is deterministic) and ϕ and ψ contains the same set of propositional variables ($\phi \vee \psi$ is smooth) then

$$\text{WMC}(\phi \vee \psi) = \text{WMC}(\phi) + \text{WMC}(\psi)$$

3. A formula is in smooth deterministic decomposable negated normal form (sd-DNNF) if
    - negation appears only in front of atoms (NNF);
    - every conjunction is decomposable;
    - every disjunction is smooth and deterministic.

Conversion to sd-DNNF

We use the same rules used for transforming in d-DNNF (Shannon's expansion) with the following additional rule

- Smoothing left: For subformula $\phi \vee \psi$ with $p \in props(\psi) \setminus props(\phi)$ apply this transformation

$$\phi \wedge (p \vee \neg p) \vee \psi$$

- Smoothing right: For subformula $\phi \vee \psi$ with $p \in props(\phi) \setminus props(\psi)$ apply this transformation

$$\phi \vee \psi \wedge (p \vee \neg p)$$

This results in:

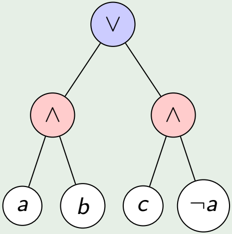
$$\left(\phi \wedge \bigwedge_{p \in props(\psi) \setminus props(\phi)} (p \vee \neg p)\right) \vee \left(\psi \wedge \bigwedge_{q \in props(\phi) \setminus props(\psi)} (q \vee \neg q)\right)$$

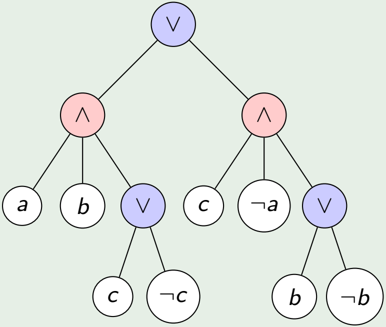
Reduction to sd-DNNF

Example

Smoothing $(a \land b) \lor (c \land \neg a)$ results in

$$(a \land b \land (c \lor \neg c)) \lor ((c \land \neg a) \land (b \lor \neg b))$$





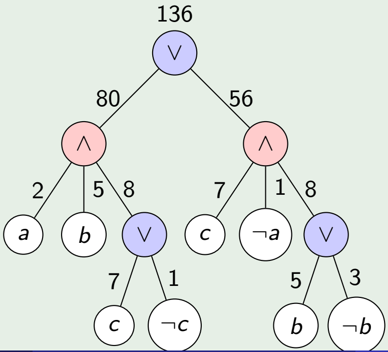
Weighted model counting of sd-DNNF formulas

Every leaf (literal) is associated with its weight, and as in d-DNNF,

- at every $\wedge$-node we perform the product of the child nodes;
- at every $\vee$-node we perform the sum of the child nodes.

Example

Consider the following weighted literals: $a : 2$, $\neg a : 1$, $b : 5$, $\neg b : 3$, $c : 7$, and $\neg c : 1$.



Interference between smoothing and determinism

Example

consider the formula $(a \wedge b) \vee c$, This formula is neither smooth nor deterministic. Should we try to first smooth it and then make it deterministic by applying Shannon's expansion? or should we proceed in the opposite direction? Let's analize the two cases:

- First Smooth then determinism

$$
\begin{aligned}
&(a \wedge b) \vee c \\
&((a \wedge b) \wedge (c \vee \neg c)) \vee (c \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \\
&(a \wedge b) \wedge (\top \vee \bot)) \vee (\top \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \wedge c \vee \\
&((a \wedge b) \wedge (\bot \vee \top)) \vee (\bot \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \wedge \neg c
\end{aligned}
$$

However notice that the formula in blue is not deterministic and we should repeat the application of Shannon's expansion. This method of proceeding, though it is correct will result in exploding the formula.

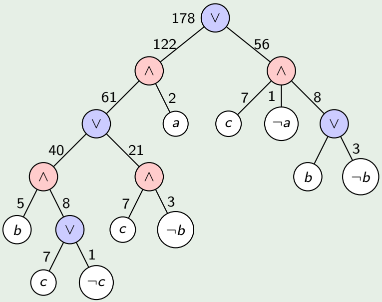
Interference between smoothing and determinism

Example

- First determinism then Smooth

$(a \wedge b) \vee c$	Shannon's exp. on a
$((b \vee c) \wedge a) \vee (c \wedge \neg a)$	Shannon's exp. on b
$((b \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a)$	Smoothing
$((b \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a \wedge (b \vee \neg b)$	Smoothing
$(((b \wedge (c \vee \neg c)) \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a \wedge (b \vee \neg b))$	

Let us use the resulting formula for weighted model counting of $(a \wedge b) \vee c$ with the weighted literals: $a : 2$, $\neg a : 1$, $b : 5$, $\neg b : 3$, $c : 7$, and $\neg c : 1$.



Interference between smoothing and determinism

Example

consider the formula $(a \wedge b) \vee c$, This formula is neither smooth nor deterministic. Should we try to first smooth it and then make it deterministic by applying Shannon's expansion? or should we proceed in the opposite direction? Let's analize the two cases:

- First Smooth then determinism

$$
\begin{aligned}
&(a \wedge b) \vee c \\
&((a \wedge b) \wedge (c \vee \neg c)) \vee (c \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \\
&(a \wedge b) \wedge (\top \vee \bot)) \vee (\top \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \wedge c \vee \\
&((a \wedge b) \wedge (\bot \vee \top)) \vee (\bot \wedge (a \vee \neg a) \wedge (b \vee \neg b)) \wedge \neg c
\end{aligned}
$$

However notice that the formula in blue is not deterministic and we should repeat the application of Shannon's expansion. This method of proceeding, though it is correct will result in exploding the formula.

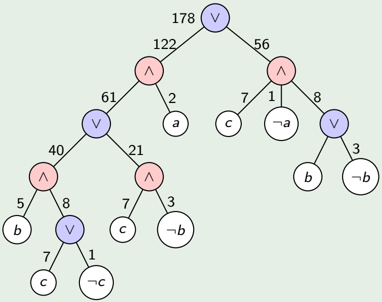
Interference between smoothing and determinism

Example

- First determinism then Smooth

$$
\begin{aligned}
&(a \wedge b) \vee c && \text{Shannon's exp. on } a \\
&((b \vee c) \wedge a) \vee (c \wedge \neg a) && \text{Shannon's exp. on } b \\
&((b \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a) && \text{Smoothing} \\
&((b \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a \wedge (b \vee \neg b) && \text{Smoothing} \\
&(((b \wedge (c \vee \neg c)) \vee (c \wedge \neg b)) \wedge a) \vee (c \wedge \neg a \wedge (b \vee \neg b))
\end{aligned}
$$

Let us use the resulting formula for weighted model counting of $(a \wedge b) \vee c$ with the weighted literals: $a : 2$, $\neg a : 1$, $b : 5$, $\neg b : 3$, $c : 7$, and $\neg c : 1$.



$\textsc{wmc}$ and probabulity

- The weight function w define the probability measure on the space of all the propositional interpretations of a finite set of propositional variable $\mathcal{P}$.

$$\Pr(\mathcal{I}) = \frac{w(\mathcal{I})}{\sum_{\mathcal{I}\in\mathbb{I}} w(\mathcal{I})} \tag{4}$$

- Foe every formula ϕ

$$\Pr(\phi) = \sum_{\mathcal{I}} \mathcal{I}(\phi) \cdot \Pr(\mathcal{I}) \tag{5}$$

- By replacing (4) in (5) we obtain:

$$\Pr(\phi) = \frac{\textsc{wmc}(\phi, w)}{\textsc{wmc}(\top, w)} = \frac{1}{Z(w)}\textsc{wmc}(\phi, w) \tag{6}$$

- Conditional probability can also be defined:

$$\Pr(\phi \mid \psi) = \frac{\frac{\textsc{wmc}(\phi\wedge\psi, w)}{\textsc{wmc}(\top, w)}}{\frac{\textsc{wmc}(\psi, w)}{\textsc{wmc}(\top, w)}} = \frac{\textsc{wmc}(\phi \wedge \psi, w)}{\textsc{wmc}(\psi, w)} \tag{7}$$

WMC and probability

Example

$w(\mathcal{I})$	p	q	r	$p \wedge q \rightarrow r$	$(\neg p \wedge q) \equiv r$
1.2	0	0	0	1	1
1.1	0	0	1	1	0
2.8	0	1	0	1	0
2.6	0	1	1	1	1
0.8	1	0	0	1	1
0.0	1	0	1	1	0
2.1	1	1	0	0	1
1.3	1	1	1	1	0
11.9					

$$\textsc{wmc}(\top) = 11.9$$

$$\textsc{wmc}(p \wedge q \rightarrow r) = 1.2 + 1.1 + 2.8 + 2.6 + 0.8 + 0.0 + 1.3 = 9.8$$

$$\textsc{wmc}((\neg p \wedge q) \equiv r) = 1.2 + 2.6 + 0.8 + 2.1 = 5.9$$

$$\Pr(p \wedge q \rightarrow r) = \frac{9.8}{11.9} \approx 0.82$$

$$\Pr((\neg p \wedge q) \equiv r) = \frac{5.9}{11.9} \approx 0.49$$

$$\Pr((\neg p \wedge q) \equiv r) \mid p \wedge q \rightarrow r) = \frac{1.2 + 2.6 + 0.8}{9.8} \approx 0.47$$

Definition (Bayesian Network)

A *Bayesian network* on a set of random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ is a pair $\mathcal{B} = (G, Pr)$ is a pair composed of a directed acyclic graph $G = ([n], E)$ (where $[n] = \{1, \ldots, n\}$) and Pr specifies the conditional probababilities

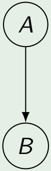
$$Pr(X_i = x_i \mid \boldsymbol{X}_{\text{par}(i)} = \boldsymbol{x}_{\text{par}(i)})$$

for every $X_i \in \boldsymbol{X}$. $\mathcal{B}$ uniquely define the join distribution on $\boldsymbol{X}$

$$Pr(\boldsymbol{X} = \boldsymbol{x}) = \prod_{i=1}^{n} Pr(X_i = x_i \mid \boldsymbol{X}_{\text{par}(i)} = \boldsymbol{x}_{\text{par}(i)}) \tag{8}$$

Example

The following simple Bayesian Netsork



$Pr(A) = 1$
0.3

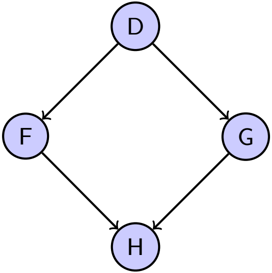
a	$Pr(B = 1 \mid A = a)$
0	0.4
1	0.9

specifies the joint probability distribution $P(A, B) = P(A) \cdot P(B \mid A)$

a	b	$P(A = a, B = b)$
0	0	0.42
0	1	0.28
1	0	0.03
1	1	0.27

Encoding bayesian networks in #SAT

P(D)
0.5



D	$P(F\|D)$
true	0.5
false	0.1

D	$P(G\|D)$
true	0.7
false	0.2

F	G	$P(H\|F,G)$
true	true	1.0
true	false	0.5
false	true	0.4
false	false	0.0

[4]

[4] Sang, Beame, and Kautz 2005.

Semantics of BN

- nodes are propositional variables

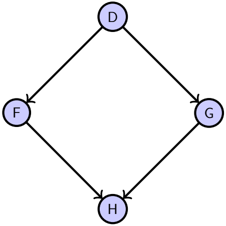
D :	John is Doing some work
F :	John has Finished his work
G :	John is Getting tired
H :	John Has a rest

- tables associated to noses (conditional probability table (CPT)) specifies conditional probabilities of the node. w.r.t, its parents

$$Pr(F = 1 \mid D = 1) = 0.5$$
$$P(F = 1 \mid D = 0) = 0.1$$

$$Pr(F = 0 \mid D = 1) = 1 - Pr(F = 1 \mid D = 1) = 0.5$$
$$Pr(F = 0 \mid D = 1`) = 1 - Pr(F = 1 \mid D = 0) = 0.9$$

P(D)
0.5



D	$P(F\|D)$
true	0.6
false	0.1

D	$P(G\|D)$
true	0.7
false	0.2

F	G	$P(H\|F,G)$
true	true	1.0
true	false	0.5
false	true	0.4
false	false	0.0

d	f	g	h	$Pr(D,F,G,H = d,f,g,h)$
0	0	0	0	$0.5 \cdot 0.9 \cdot 0.8 \cdot 1.0 = 0.360$
0	0	0	1	$0.5 \cdot 0.9 \cdot 0.8 \cdot 0.0 = 0.000$
0	0	1	0	$0.5 \cdot 0.9 \cdot 0.2 \cdot 0.6 = 0.054$
0	0	1	1	$0.5 \cdot 0.9 \cdot 0.2 \cdot 0.4 = 0.036$
0	1	0	0	$0.5 \cdot 0.1 \cdot 0.8 \cdot 0.6 = 0.024$
0	1	0	1	$0.5 \cdot 0.1 \cdot 0.8 \cdot 0.4 = 0.016$
0	1	1	0	$0.5 \cdot 0.1 \cdot 0.2 \cdot 0.0 = 0.000$
0	1	1	1	$0.5 \cdot 0.1 \cdot 0.2 \cdot 1.0 = 0.010$
1	0	0	0	$0.5 \cdot 0.4 \cdot 0.3 \cdot 1.0 = 0.060$
1	0	0	1	$0.5 \cdot 0.4 \cdot 0.3 \cdot 0.0 = 0.000$
1	0	1	0	$0.5 \cdot 0.4 \cdot 0.7 \cdot 0.6 = 0.084$
1	0	1	1	$0.5 \cdot 0.4 \cdot 0.7 \cdot 0.4 = 0.056$
1	1	0	0	$0.5 \cdot 0.6 \cdot 0.3 \cdot 0.5 = 0.045$
1	1	0	1	$0.5 \cdot 0.6 \cdot 0.3 \cdot 0.5 = 0.045$
1	1	1	0	$0.5 \cdot 0.6 \cdot 0.7 \cdot 0.0 = 0.000$
1	1	1	1	$0.5 \cdot 0.6 \cdot 0.7 \cdot 1.0 = 0.210$
				1.000

Encoding BN in WMC

$\Phi_{\mathcal{B}}$ and $w_{\mathcal{B}}$

1. For every node p with $k > 0$ parents introduce 2^k new propositional variables $p_{\boldsymbol{b}}$ for $\boldsymbol{b} \in \{0,1\}^k$.
2. $w_{\mathcal{B}}(p_{\boldsymbol{b}}) \triangleq \Pr(p = 1 \mid \text{par}(p) = \boldsymbol{b})$.
3. $w(\neg p_{\boldsymbol{b}}) \triangleq 1 - w(p_{\boldsymbol{b}})$.
4. set the weight of all the other literals to 1
5. For every $p_{\boldsymbol{b}}$ add

$$p_{\boldsymbol{b}} \leftrightarrow p \wedge \left(\bigwedge_{\substack{i=1\\ b_i=1}}^{k} p_i \wedge \bigwedge_{\substack{i=1\\ b_i=0}}^{k} \neg p_i \right)$$

Example

1. $F_0,\ F_1,\ G_0,\ G_1,\ H_{00},\ H_{01},\ H_{10},\ H_{11},$
2. $$\begin{array}{rr} w(D) = 0.5 & \\ w(F_0) = 0.1 & w(F_1) = 0.5 \\ w(G_0) = 0.2 & w(G_1) = 0.7 \\ w(H_{00}) = 0.0 & w(H_{01}) = 0.4 \\ w(H_{10}) = 0.5 & w(H_{11}) = 1.0 \end{array}$$
3. $w(\neg D) = 0.5,\ w(\neg F_0) = 0.9\ \ldots$
4. $w(F) = W(\neg F) = 1 \ldots$
5. $$\begin{array}{ll} F_0 \leftrightarrow F \wedge \neg D & F_1 \leftrightarrow F \wedge D \\ G_0 \leftrightarrow G \wedge \neg D & G_1 \leftrightarrow G \wedge D \\ H_{11} \leftrightarrow H \wedge F \wedge G & H_{00} \leftrightarrow H \wedge \neg F \wedge \neg G \\ H_{01} \leftrightarrow H \wedge \neg F \wedge G & H_{10} \leftrightarrow H \wedge F \wedge \neg G \end{array}$$

Property of the encoding

Proposition

Let $\mathcal{B}$ be a Bayesian networks on the boolean random variables $X_1, \dots, X_n$ that defines the joint probability distribution $Pr(X_1, \dots, X_n)$.

- *for every assignment $\boldsymbol{x} = (x_1, \dots, x_n)$ to the variables $X_1, \dots, X_n$. there is a unique interpretaiton $\mathcal{I}_{\boldsymbol{x}}$ that satisfies $\Phi_{\mathcal{B}}$ and such that $\mathcal{I}(X_i) = x_i$*
- *For every $\mathcal{I}$ that satisfies $\Phi_{\mathcal{B}}$*

$$w_{\mathcal{B}}(\mathcal{I}) = Pr(X_1 = \mathcal{I}(X_1), \dots, X_n = \mathcal{I}(X_n))$$

$$Pr(\phi \mid \psi) = \frac{\text{WMC}(\Phi_{\mathcal{B}} \wedge \phi \wedge \psi, w_{\mathcal{B}})}{\text{WMC}(\Phi_{\mathcal{B}} \wedge \psi, w_{\mathcal{B}})} \tag{9}$$

We can use knowledge complilation. For instance the sd-DNNF reduction of $\Phi_{\mathcal{B}}$ for the previous example is

$$\begin{aligned}
D\wedge \quad & (F \wedge F_1 \wedge (G \wedge G_1 \wedge (H \wedge H_{11} \vee \neg H \wedge \neg H_{11})) \vee \\
& \qquad\qquad (\neg G \wedge \neg G_1 \wedge (H \wedge H_{10} \vee \neg H \wedge \neg H_{10}))) \vee \\
& (\neg F \wedge F_1 \wedge (G \wedge G_1 \wedge (H \wedge H_{01} \vee \neg H \wedge \neg H_{01})) \vee \\
& \qquad\qquad (\neg G \wedge \neg G_1 \wedge (H \wedge H_{00} \vee \neg H \wedge \neg H_{00}))) \vee \\
\neg D\wedge \quad & (F \wedge F_0 \wedge (G \wedge G_0 \wedge (H \wedge H_{11} \vee \neg H \wedge \neg H_{11})) \vee \\
& \qquad\qquad (\neg G \wedge \neg G_0 \wedge (H \wedge H_{10} \vee \neg H \wedge \neg H_{10}))) \vee \\
& (\neg F \wedge F_0 \wedge (G \wedge G_0 \wedge (H \wedge H_{01} \vee \neg H \wedge \neg H_{01})) \vee \\
& \qquad\qquad (\neg G \wedge \neg G_0 \wedge (H \wedge H_{00} \vee \neg H \wedge \neg H_{00})))
\end{aligned}$$

Learning weights

- Suppose we have a set of observations of itemsets, as for instance the one we have seen at the beginning of the class. i.e., our observations are a sequence of possible repeated interpretations $\mathbb{I} = \mathcal{I}^{(1)}, \mathcal{I}^{(2)}, \dots, \mathcal{I}^{(d)}$ where d the the size of the observations.
- and we want to model the probability distribution obtained via weighted model counting with a set of weighted formulas. $w_1 : \phi_1, \dots, w_k : \phi_k$
- How can we find a tuple of weights $\boldsymbol{w} = (w_1, \dots, w_k)$ that best fits the observed data?
- One criteria is to find the vector of weights $\boldsymbol{w}$ that maximizes the Likelihood of the data, i.e.:

$$Likelihood(\mathbb{I} \mid \boldsymbol{w}) = \Pr(\mathbb{I} \mid \boldsymbol{w})$$

Maximizing the likelihood of data

- we assume that each observation in $\mathbb{I} = (\mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(d)})$ is independent from all the others.

$$\Pr(\mathbb{I} \mid \boldsymbol{w}) = \prod_{i=1}^{d} Pr(\mathcal{I}^{(i)} \mid \boldsymbol{w})$$

- We have that $Pr(\mathcal{I}^{(i)} \mid \boldsymbol{w}) = \frac{\text{WMC}(\mathcal{I}^{(i)} \mid \boldsymbol{w})}{\text{WMC}(\top \mid \boldsymbol{w})}$

$$\Pr(\mathbb{I} \mid \boldsymbol{w}) = \prod_{i=1}^{d} \frac{w(\mathcal{I}^{(i)} \mid \boldsymbol{w})}{w(\top \mid \boldsymbol{w})}$$

- where $\text{WMC}(\top \mid \boldsymbol{w}) = \sum_{\mathcal{I} \models \top} w(\mathcal{I} \mid \boldsymbol{w})$
- and $w(\mathcal{I} \mid \boldsymbol{w}) = \exp\left(\sum_{j=1}^{k} w_j \cdot \mathcal{I}(\phi_j)\right)$
- we therefore have that:

$$Likelihood(\mathbb{I} \mid \boldsymbol{w}) = \prod_{i=1}^{d} \frac{1}{\text{WMC}(\top \mid \boldsymbol{w})} \exp\left(\sum_{j=1}^{k} w_j \cdot \mathcal{I}^{(i)}(\phi_i)\right)$$

Maximizing the log-likelihood of data

Learning weights

$$\boldsymbol{w}^* = \operatorname*{argmax}_{\boldsymbol{w}} \mathit{Likelihood}(\mathbb{I} \mid \boldsymbol{w})$$

which is equivalent to

$$\boldsymbol{w}^* = \operatorname*{argmax}_{\boldsymbol{w}} \left(\ln\left(\mathit{Likelihood}(\mathbb{I} \mid \boldsymbol{w})\right)\right)$$

i.e.,

$$\boldsymbol{w}^* = \operatorname*{argmax}_{\boldsymbol{w}} \left(\sum_{i=1}^{d}\sum_{j=1}^{k} w_j \cdot \mathcal{I}^{(i)}(\phi_j) - d \cdot \ln\left(\text{WMC}(\top \mid \boldsymbol{w})\right)\right)$$

$$\boldsymbol{w}^* = \operatorname*{argmax}_{\boldsymbol{w}} \left(n_j \cdot w_j - d \cdot \ln\left(\text{WMC}(\top \mid \boldsymbol{w})\right)\right)$$

where n_j is the number of observations $\mathcal{I}^{(i)}$ for which the fornula ϕ_j is true.

Maximizing the log-likelihood of data

- Try maximization with gradient ascent approach, by putting to zeros the partial derivatives of the log likelihood, i.e.,

$$\frac{\partial logLik(\mathbb{I} \mid \boldsymbol{w})}{\partial w_i} = 0$$

where

$$logLik(\mathbb{I} \mid \boldsymbol{w}) = n_j \cdot w_j - d \cdot \ln\left(\textsc{wmc}(\top \mid \boldsymbol{w})\right)$$

- **Problem:** calculating $\frac{\partial \ln(\textsc{wmc}(\top|\boldsymbol{w}))}{\partial w_i}$, i.e.,

$$\frac{\partial\left(\ln\left(\sum_{\mathcal{I}} \exp\left(\sum_{j=1}^{k} w_j \cdot \mathcal{I}(\phi_j)\right)\right)\right)}{\partial w_j}$$

requires exponential amount of time. Use approximative techniques[5].

[5] Richardson and Domingos 2006.

Special case: we only have one formula

- If we consider only one formula ϕ_1, then

$$\frac{\partial \ln \left(\sum_{\mathcal{I}} \exp(w_1 \cdot \mathcal{I}(\phi_1))\right)}{\partial w_1}$$

can be computed analytically

$$w_1 = \ln \left(\frac{n_1 \cdot \#SAT(\neg\phi_1)}{(d - n_1)\#SAT(\phi_1)} \right) \tag{10}$$

- **Observation 1:** the more often ϕ_1 is satisfied in the obsesrvation, the larger it's weight w_1
- the more models of ϕ_1, i.e., the larger $\#SAT(\phi_1)$ the smaller w_1.

Special case: we only have one formula $\phi : w$

Derivation of the formula (10).

1. THe likelihood w.r.t., a single formula $w : \phi$ of the data $\mathbb{I} = \mathcal{I}^{(1)}, \dots, \mathcal{I}^{(d)}$

$$
\begin{aligned}
Likelihood(\mathbb{I} \mid w) &= \prod_{i=1}^{d} \frac{1}{\textsc{wmc}(\top \mid w)} \exp\left(w \cdot \mathcal{I}^{(i)}(\phi)\right) \\
&= \textsc{wmc}(\top \mid w)^{-d} \exp\left(\sum_{i=1}^{d} w \cdot \mathcal{I}^{(i)}(\phi)\right) \\
&= \textsc{wmc}(\top \mid w)^{-d} \exp\left(n \cdot w\right)
\end{aligned}
$$

2. We then determine the logarithm of the likelihood

$$
LogLike(\mathbb{I} \mid w) = n \cdot w - d \cdot \log(\textsc{wmc}(\top \mid w))
$$

where n is the number of $\mathcal{I}^{(i)}$'s that satisfy ϕ.

3. We then compute the derivative w.r.t, w

$$
\begin{aligned}
\frac{\partial LogLike(\mathbb{I} \mid w)}{\partial w} &= n - d \cdot \left(\frac{1}{\textsc{wmc}(\top \mid w)}\right) \cdot \frac{\partial \textsc{wmc}(\top \mid w)}{\partial w} \\
&= n - d \cdot \left(\frac{e^{w} \cdot \#\textsc{sat}(\phi)}{e^{w} \cdot \#\textsc{sat}(\phi) + \#\textsc{sat}(\neg\phi)}\right)
\end{aligned}
$$

Special case: we only have one formula $\phi : w$

4. We then pose the derivative equal to 0

$$0 = \frac{\partial LogLike(\mathbb{I} \mid w)}{\partial w}$$

$$0 = n - d \cdot \left(\frac{e^w \cdot \#\textsc{sat}(\phi)}{e^w \cdot \#\textsc{sat}(\phi) + \#\textsc{sat}(\neg\phi)} \right)$$

$$d \cdot \left(\frac{e^w \cdot \#\textsc{sat}(\phi)}{e^w \cdot \#\textsc{sat}(\phi) + \#\textsc{sat}(\neg\phi)} \right) = n$$

$$d \cdot e^w \cdot \#\textsc{sat}(\phi) = n \cdot e^w \cdot \#\textsc{sat}(\phi) + n \cdot \#\textsc{sat}(\neg\phi)$$

$$e^w = \frac{n \cdot \#\textsc{sat}(\neg\phi)}{(d-n)\#\textsc{sat}(\phi)}$$

$$w = \log \left(\frac{n \cdot \#\textsc{sat}(\neg\phi)}{(d-n)\#\textsc{sat}(\phi)} \right)$$

Example of learning weights

Example

Suppose that we have $\mathbb{I} = \mathcal{I}^{(1)}, \dots, \mathcal{I}^{(22)}$ are summarized in the following table:

#	Itemsets						
4	*a*	*b*	*c*	*d*			
1	*a*	*b*			*e*	*f*	
7	*a*	*b*	*c*				
3	*a*		*c*	*d*		*f*	
2							*g*
1				*d*			
4				*d*			*g*

$$a \quad w = \log\left(\frac{15 \cdot 2^6}{7 \cdot 2^6}\right) \approx 0.76$$

$$\neg a \quad w = \log\left(\frac{7 \cdot 2^6}{15 \cdot 2^6}\right) \approx -0.76$$

$$e \quad w = \log\left(\frac{1 \cdot 2^6}{21 \cdot 2^6}\right) \approx -3.04$$

$$\neg e \quad w = \log\left(\frac{21 \cdot 2^6}{1 \cdot 2^6}\right) \approx 3.04$$

Example of learning weights

Example

Suppose that we have $\mathbb{I} = \mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(22)}$ are summarized in the following table:

#	Itemsets						
4	a	b	c	d			
1	a	b			e	f	
7	a	b	c				
3	a		c	d		f	
2							g
1				d			
4				d			g

$$a \wedge b \quad w = \log \frac{12 \cdot (2^7 - 2^5)}{10 \cdot 2^5} \approx 8.21$$

$$c \wedge d \quad w = \log \frac{7 \cdot (2^7 - 2^5)}{15 \cdot 2^5} \approx 7.27$$

$$e \wedge f \quad w = \log \frac{1 \cdot (2^7 - 2^5)}{21 \cdot 2^5} \approx 4.99$$

$$a \rightarrow b \quad w = \log \frac{19 \cdot (2^7 - 3 \cdot 2^5)}{3 \cdot 3 \cdot 2^5} \approx 0.75$$

$$a \wedge b \wedge c \wedge \neg e \wedge \neg f \rightarrow g \qquad w = \log \left(\frac{11}{11 \cdot (2^7 - 2)} \right) \approx -4.84$$

$$a \wedge b \wedge \neg c \wedge \neg d \wedge e \wedge f \wedge \neg g \qquad w = \log(21 \cdot (2^7 - 1)) \approx 7.89$$

bibliography

Chakraborty, Supratik, Kuldeep S Meel, and Moshe Y Vardi (2021). "Approximate model counting". In: *Handbook of Satisfiability*. IOS Press, pp. 1015–1045.

Colnet, Alexis de and Kuldeep S Meel (2019). "Dual hashing-based algorithms for discrete integration". In: *International Conference on Principles and Practice of Constraint Programming*. Springer, pp. 161–176.

Ermon, Stefano et al. (2013). "Taming the curse of dimensionality: Discrete integration by hashing and optimization". In: *International Conference on Machine Learning*. PMLR, pp. 334–342.

Richardson, Matthew and Pedro Domingos (Feb. 2006). "Markov Logic Networks". In: *Mach. Learn.* 62.1-2, pp. 107–136. ISSN: 0885-6125. DOI: 10.1007/s10994-006-5833-1. URL: http://dx.doi.org/10.1007/s10994-006-5833-1.

Sang, Tian, Paul Beame, and Henry Kautz (2005). "Solving Bayesian networks by weighted model counting". In: *Proc.AAAI-05*. Vol. 1, pp. 475–482.