Knowledge Representation and Learning
Weighted Model Counting

Luciano Serafini
FBK, Trento, Italy

July 12, 2023
<table>
<thead>
<tr>
<th>Task Name</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model checking:</td>
<td>$\phi, I$</td>
<td>$I(\phi)$</td>
</tr>
<tr>
<td>Satisfiability:</td>
<td>$\phi$</td>
<td>$\max_I I(\phi)$</td>
</tr>
<tr>
<td>Maximum Satisfiability:</td>
<td>$\phi, w$</td>
<td>$\max_I I(\phi) \cdot w(I)$</td>
</tr>
<tr>
<td>Model counting:</td>
<td>$\phi$</td>
<td>$\sum_I I(\phi)$</td>
</tr>
<tr>
<td>Weighted model counting:</td>
<td>$\phi, w$</td>
<td>$\sum_I I(\phi) \cdot w(I)$</td>
</tr>
</tbody>
</table>
Definition (Weighted model counting)

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

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

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

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

Suppose that we log what people buy in a supermarket:

<table>
<thead>
<tr>
<th>#</th>
<th>Itemsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$a \ b \ c \ d$</td>
</tr>
<tr>
<td>1</td>
<td>$a \ b \ e \ f$</td>
</tr>
<tr>
<td>7</td>
<td>$a \ b \ c$</td>
</tr>
<tr>
<td>3</td>
<td>$a \ c \ d \ f$</td>
</tr>
<tr>
<td>2</td>
<td>$g$</td>
</tr>
<tr>
<td>1</td>
<td>$d$</td>
</tr>
<tr>
<td>4</td>
<td>$d \ g$</td>
</tr>
</tbody>
</table>

- Every combination of items can be seen as an interpretation on the set of propositions $a, b, \ldots g$. 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 assign to each a weight $w(\mathcal{I})$ which is the number of times an itemset has been observed.
$$\text{Example}$$

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
<th>$f$</th>
<th>$g$</th>
<th>$w(\mathcal{I})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{WMC}(a \land (b \lor c)) &= 4 + 1 + 7 + 3 = 15 \\
\text{WMC}(a \land g) &= 0 \\
\text{WMC}(a \land \neg g) &= 4 + 1 + 7 + 3 = 15 \\
\text{WMC}(a \rightarrow b) &= 4 + 1 + 7 + 2 + 1 + 4 = 19
\end{align*}
\]
Model counting vs. Weighted model counting

- In model counting each interpretation weights 1;
- In \textit{WMC} instead, some models are more important than others, and it makes sense to associate a weight $w(I) \geq 0$ to each interpretation $I$.
- In \textit{weighted model counting} each model of a formula counts for its weight $w(I)$
- 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 $w(I)$ associated to the model $I$ can be interpreted in probabilistically; i.e., the higher the weight of a model the more likely the model;
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 $\phi$ 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 $w$.

$$Z(w) = \sum_{\mathcal{I}} w(\mathcal{I})$$ (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 $W : \{0, 1\}^{|P|} \rightarrow \mathbb{R}^+$

What is a compact way to represent the weight function?

- To explicitly defining the weights for each interpretation we need $2^{|P|}$ 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
  \[ w(I) = \prod_{I \models \phi_i} w_i \] (2)

  or alternatively
  \[ w(I) = \exp \left( \sum_{I \models \phi_i} w'_i \right) \] (3)

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>(w(I))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>6.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>(w(I))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>5.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7.0</td>
</tr>
</tbody>
</table>

- 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 \(p \lor q : 2\),
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 \( w : \text{Lit} \rightarrow \mathbb{R}^+ \) 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

<table>
<thead>
<tr>
<th>p</th>
<th>q</th>
<th>r</th>
<th>$w(x)MCh(\mathcal{I},x)$</th>
<th>$w(\neg x)MCh(\mathcal{I},\neg x)$</th>
<th>$w(\mathcal{I})$</th>
<th>$Pr(\mathcal{I})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.04</td>
<td>0.11</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.36</td>
<td>0.07</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3.2</td>
<td>6.528</td>
<td>0.34</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.2</td>
<td>4.352</td>
<td>0.23</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.2</td>
<td>1</td>
<td>0.72</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.2</td>
<td>1</td>
<td>0.48</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1.2</td>
<td>3.2</td>
<td>2.304</td>
<td>0.12</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.2</td>
<td>3.2</td>
<td>1.536</td>
<td>0.08</td>
</tr>
</tbody>
</table>

$WMC(p \lor \neg q \rightarrow r) = w(001) + w(010) + w(011) + w(101) + w(111) \approx 14.26$

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

$Pr(p \lor \neg q \rightarrow r) = \frac{WMC(p \lor \neg q \rightarrow r)}{WMC(\top)} \approx \frac{14.26}{19.32} \approx 0.74$
Weighted Model counting

Examples (Weights can be associated also to formulas)

<table>
<thead>
<tr>
<th>$w$</th>
<th>$\Delta$ defines fresh variables</th>
<th>$w'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\neg(p \lor q) \rightarrow 0.0$</td>
<td>$f_0 \leftrightarrow \neg(p \lor q)$</td>
<td>$f_0 \rightarrow 0.0$</td>
</tr>
<tr>
<td>$p \rightarrow 0.1$</td>
<td>$f_1 \leftrightarrow p$</td>
<td>$f_1 \rightarrow 0.1$</td>
</tr>
<tr>
<td>$p \lor r \rightarrow 1.2$</td>
<td>$f_2 \leftrightarrow p \lor r$</td>
<td>$f_2 \rightarrow 1.2$</td>
</tr>
<tr>
<td>$q \rightarrow r \rightarrow 2.5$</td>
<td>$f_3 \leftrightarrow q \rightarrow r$</td>
<td>$f_3 \rightarrow 2.5$</td>
</tr>
</tbody>
</table>

\[
WMC(p \lor \neg q \rightarrow r \land \Delta) = \\
w(0011011) + w(0100000) + w(0110011) + w(1010111) + w(1110111) = \\
0 + 1 + 3 + 0.3 + 0.3 = 4.6
\]

\[
WMC(\Delta) = w(0001001) + w(0011011) + w(0100000) + w(0110011) \\
+ w(1000111) + w(1010111) + w(1100110) + w(1110111) \\
= 0 + 0 + 1 + 3 + 0.3 + 0.3 + 0.12 + 0.3 = 5.02
\]

\[
Pr(p \lor \neg q \rightarrow r | \Delta) = \frac{WMC(p \lor \neg q \rightarrow r \land \Delta)}{WMC(\Delta)} = \frac{4.6}{5.02} \approx 0.92
\]
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 function on the set of propositional variables of \( \phi \) and \( \psi \).

1. If \( \phi \) and \( \psi \) do not contain common propositional variables (\( \phi \land \psi \) is decomposable) then:

\[
WMC(\phi \land \psi, w) = WMC(\phi, w|_P(\phi)) \cdot WMC(\psi, w|_P(\psi))
\]

2. If \( \phi \land \psi \) is unsatisfiable (\( \phi \lor \psi \) is deterministic) and \( \phi \) and \( \psi \) contains the same set of propositional variables (\( \phi \lor \psi \) is smooth) then

\[
WMC(\phi \lor \psi) = WMC(\phi) + WMC(\psi)
\]

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 $\phi \lor \psi$ with $p \in \text{props}(\psi) \setminus \text{props}(\phi)$ apply this transformation

  $$\phi \land (p \lor \neg p) \lor \psi$$

- **Smoothing right:** For subformula $\phi \lor \psi$ with $p \in \text{props}(\phi) \setminus \text{props}(\psi)$ apply this transformation

  $$\phi \lor \psi \land (p \lor \neg p)$$

This results in:

$$\left( \phi \land \bigwedge_{p \in \text{props}(\psi) \setminus \text{props}(\phi)} (p \lor \neg p) \right) \lor \left( \psi \land \bigwedge_{q \in \text{props}(\phi) \setminus \text{props}(\psi)} (q \lor \neg q) \right)$$
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 $\land$-node we perform the product of the child nodes;
- at every $\lor$-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 \((a \land b) \lor c\), 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 analyze the two cases:

- First **Smooth** then **determinism**

\[
\begin{align*}
(a \land b) \lor c \\
((a \land b) \land (c \lor \neg c)) \lor (c \land (a \lor \neg a) \land (b \lor \neg b)) \\
(a \land b) \land (\top \lor \bot) \lor (\top \land (a \lor \neg a) \land (b \lor \neg b)) \land c \lor \\
((a \land b) \land (\bot \lor \top)) \lor (\bot \land (a \lor \neg a) \land (b \lor \neg b)) \land \neg c
\end{align*}
\]

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

\[(a \land b) \lor c\]

\[((b \lor c) \land a) \lor (c \land \neg a)\]

\[((b \lor (c \land \neg b)) \land a) \lor (c \land \neg a)\]

\[((b \lor (c \land \neg b)) \land a) \lor (c \land \neg a \land (b \lor \neg b))\]

\[((((b \land (c \lor \neg c)) \lor (c \land \neg b)) \land a) \lor (c \land \neg a \land (b \lor \neg b)))\]

Shannon's exp. on \(a\)

Shannon's exp. on \(b\)

Smoothing

Smoothing

Let us use the resulting formula for weighted model counting of \((a \land b) \lor c\) 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 \((a \land b) \lor c\), 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 analyze the two cases:

- First **Smooth** then **determinism**

\[
(a \land b) \lor c \\
((a \land b) \land (c \lor \neg c)) \lor (c \land (a \lor \neg a) \land (b \lor \neg b)) \\
(a \land b) \land (\top \lor \bot) \lor (\top \land (a \lor \neg a) \land (b \lor \neg b)) \land c \lor \\
((a \land b) \land (\bot \lor \top)) \lor (\bot \land (a \lor \neg a) \land (b \lor \neg b)) \land \neg c
\]

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**

\[
(a \land b) \lor c \\
((b \lor c) \land a) \lor (c \land \neg a) \\
((b \lor (c \land \neg b)) \land a) \lor (c \land \neg a) \\
((b \lor (c \land \neg b)) \land a) \lor (c \land \neg a \land (b \lor \neg b)) \\
(((b \land (c \lor \neg c)) \lor (c \land \neg b)) \land a) \lor (c \land \neg a \land (b \lor \neg b))
\]

Shannon's exp. on \(a\)

Shannon's exp. on \(b\)

Smoothing

Smoothing

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

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

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

- For every formula $\phi$

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

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

$$
Pr(\phi) = \frac{\text{WMC}(\phi, w)}{\text{WMC}(\top, w)} = \frac{1}{Z(w)} \text{WMC}(\phi, w) \quad (6)
$$

- Conditional probability can also be defined:

$$
Pr(\phi \mid \psi) = \frac{\text{WMC}(\phi \land \psi, w)}{\text{WMC}(\top, w)} = \frac{\text{WMC}(\phi \land \psi, w)}{\text{WMC}(\psi, w)} \quad (7)
$$
### Example

<table>
<thead>
<tr>
<th>$w(I)$</th>
<th>$p$</th>
<th>$q$</th>
<th>$r$</th>
<th>$p \land q \rightarrow r$</th>
<th>$\neg p \land q \equiv r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{WMC}(\top) = 11.9 \\
\text{WMC}(p \land q \rightarrow r) = 1.2 + 1.1 + 2.8 + 2.6 + 0.8 + 0.0 + 1.3 = 9.8 \\
\text{WMC}((\neg p \land q) \equiv r) = 1.2 + 2.6 + 0.8 + 2.1 = 5.9
\]

\[
\text{Pr}(p \land q \rightarrow r) = \frac{9.8}{11.9} \approx 0.82
\]
\[
\text{Pr}((\neg p \land q) \equiv r) = \frac{5.9}{11.9} \approx 0.49
\]
\[
\text{Pr}((\neg p \land q) \equiv r) \mid p \land q \rightarrow r = \frac{1.2 + 2.6 + 0.8}{9.8} \approx 0.47
\]
A *Bayesian network* on a set of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ 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 probabilities

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

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

$$Pr(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} Pr(X_i = x_i \mid \mathbf{X}_{\text{par}(i)} = \mathbf{x}_{\text{par}(i)}) \quad (8)$$
Bayesian networks

Example

The following simple Bayesian Network

\[
\begin{align*}
\Pr(A) &= 1 \\
0.3 \\
\Pr(B = 1 | A = a) \\
a &
\begin{array}{c|c}
0 & 0.4 \\
1 & 0.9
\end{array}
\end{align*}
\]

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

\[
\begin{array}{c|c|c}
a & b & P(A = a, B = b) \\
0 & 0 & 0.42 \\
0 & 1 & 0.28 \\
1 & 0 & 0.03 \\
1 & 1 & 0.27
\end{array}
\]
Encoding bayesian networks in \#SAT

| 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      |

nodes are propositional variables

\[ D : \quad \text{John is Doing some work} \]
\[ F : \quad \text{John has Finished his work} \]
\[ G : \quad \text{John is Getting tired} \]
\[ H : \quad \text{John Has a rest} \]

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

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

\[
Pr(F = 0 \mid D = 1) = 1 - Pr(F = 1 \mid D = 1) = 0.5 \\
Pr(F = 0 \mid D = 1') = 1 - Pr(F = 1 \mid D = 0) = 0.9
\]
\[ P(D) = 0.5 \]

### Table 1

| 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 |

### Table 2

<table>
<thead>
<tr>
<th>d</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>( Pr(D, F, G, H = d, f, g, h) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( 0.5 \cdot 0.9 \cdot 0.8 \cdot 1.0 = 0.360 )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>( 0.5 \cdot 0.9 \cdot 0.8 \cdot 0.0 = 0.000 )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>( 0.5 \cdot 0.9 \cdot 0.2 \cdot 0.6 = 0.054 )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>( 0.5 \cdot 0.9 \cdot 0.2 \cdot 0.4 = 0.036 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>( 0.5 \cdot 0.1 \cdot 0.8 \cdot 0.6 = 0.024 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( 0.5 \cdot 0.1 \cdot 0.8 \cdot 0.4 = 0.016 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( 0.5 \cdot 0.1 \cdot 0.2 \cdot 0.0 = 0.000 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( 0.5 \cdot 0.1 \cdot 0.2 \cdot 1.0 = 0.010 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( 0.5 \cdot 0.4 \cdot 0.3 \cdot 1.0 = 0.060 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>( 0.5 \cdot 0.4 \cdot 0.3 \cdot 0.0 = 0.000 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>( 0.5 \cdot 0.4 \cdot 0.7 \cdot 0.6 = 0.084 )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>( 0.5 \cdot 0.4 \cdot 0.7 \cdot 0.4 = 0.056 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>( 0.5 \cdot 0.6 \cdot 0.3 \cdot 0.5 = 0.045 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( 0.5 \cdot 0.6 \cdot 0.3 \cdot 0.5 = 0.045 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( 0.5 \cdot 0.6 \cdot 0.7 \cdot 0.0 = 0.000 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( 0.5 \cdot 0.6 \cdot 0.7 \cdot 1.0 = 0.210 )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( 1.000 )</td>
</tr>
</tbody>
</table>
Encoding BN in WMC

**Example**

1. \( F_0, F_1, G_0, G_1, H_{00}, H_{01}, H_{10}, H_{11}, \)

2. \( w(D) = 0.5 \)
   \( w(F_0) = 0.1 \)
   \( w(F_1) = 0.5 \)
   \( w(G_0) = 0.2 \)
   \( w(G_1) = 0.7 \)
   \( w(H_{00}) = 0.0 \)
   \( w(H_{01}) = 0.4 \)
   \( w(H_{10}) = 0.5 \)
   \( w(H_{11}) = 1.0 \)

3. \( w(\neg D) = 0.5, w(\neg F_0) = 0.9 \ldots \)

4. \( w(F) = W(\neg F) = 1 \ldots \)

5. \( F_0 \leftrightarrow F \land \neg D \quad F_1 \leftrightarrow F \land D \)
   \( G_0 \leftrightarrow G \land \neg D \quad G_1 \leftrightarrow G \land D \)
   \( H_{11} \leftrightarrow H \land F \land G \quad H_{00} \leftrightarrow H \land \neg F \land \neg G \)
   \( H_{01} \leftrightarrow H \land \neg F \land G \quad H_{10} \leftrightarrow H \land F \land \neg G \)
Proposition

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

- for every assignment $x = (x_1, \ldots, x_n)$ to the variables $X_1, \ldots, X_n$. there is a unique interpretation $I_x$ that satisfies $\Phi_B$ and such that $I(X_i) = x_i$.
- For every $I$ that satisfies $\Phi_B$

$$w_B(I) = Pr(X_1 = I(X_1), \ldots, X_n = I(X_n))$$
Answering CPQ’s via WMC

\[
Pr(\phi \mid \psi) = \frac{\text{WMC}(\Phi_B \land \phi \land \psi, w_B)}{\text{WMC}(\Phi_B \land \psi, w_B)}
\]

We can use knowledge compilation. For instance the sd-DNNF reduction of $\Phi_B$ for the previous example is

\[
D \land (F \land F_1 \land (G \land G_1 \land (H \land H_{11} \lor \neg H \land \neg H_{11}))) \lor \\
(\neg G \land \neg G_1 \land (H \land H_{10} \lor \neg H \land \neg H_{10}))) \lor \\
(\neg F \land F_1 \land (G \land G_1 \land (H \land H_{01} \lor \neg H \land \neg H_{01}))) \lor \\
(\neg G \land \neg G_1 \land (H \land H_{00} \lor \neg H \land \neg H_{00}))) \lor \\
\neg D \land (F \land F_0 \land (G \land G_0 \land (H \land H_{11} \lor \neg H \land \neg H_{11}))) \lor \\
(\neg G \land \neg G_0 \land (H \land H_{10} \lor \neg H \land \neg H_{10}))) \lor \\
(\neg F \land F_0 \land (G \land G_0 \land (H \land H_{01} \lor \neg H \land \neg H_{01}))) \lor \\
(\neg G \land \neg G_0 \land (H \land H_{00} \lor \neg H \land \neg H_{00})))
\]
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 $\mathcal{I} = \mathcal{I}^{(1)}, \mathcal{I}^{(2)}, \ldots, \mathcal{I}^{(d)}$ 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.
  
  $w_1 : \phi_1, \ldots, w_k : \phi_k$

- How can we find a tuple of weights $w = (w_1, \ldots, w_k)$ that best fits the observed data?

- One criteria is to find the vector of weights $w$ that maximizes the **Likelihood** of the data, i.e.:

  $$Likelihood(\mathcal{I} \mid w) = Pr(\mathcal{I} \mid w)$$
Maximizing the likelihood of data

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

\[
Pr(\mathbb{I} | \mathbf{w}) = \prod_{i=1}^{d} Pr(I^{(i)} | \mathbf{w})
\]

- We have that \( Pr(I^{(i)} | \mathbf{w}) = \frac{WMC(I^{(i)}|\mathbf{w})}{WMC(\top|\mathbf{w})} \)

\[
Pr(\mathbb{I} | \mathbf{w}) = \prod_{i=1}^{d} \frac{w(I^{(i)} | \mathbf{w})}{w(\top | \mathbf{w})}
\]

- where \( WMC(\top | \mathbf{w}) = \sum_{\mathbb{I} \models \top} w(\mathbb{I} | \mathbf{w}) \)

- and \( w(\mathbb{I} | \mathbf{w}) = \exp \left( \sum_{j=1}^{k} w_{j} \cdot I(\phi_{j}) \right) \)

- we therefore have that:

\[
Likelihood(\mathbb{I} | \mathbf{w}) = \prod_{i=1}^{d} \frac{1}{WMC(\top | \mathbf{w})} \exp \left( \sum_{j=1}^{k} w_{j} \cdot I^{(i)}(\phi_{i}) \right)
\]
Maximizing the log-likelihood of data

Learning weights

\[ \mathbf{w}^* = \arg\max_{\mathbf{w}} \text{Likelihood}(\mathbb{I} \mid \mathbf{w}) \]

which is equivalent to

\[ \mathbf{w}^* = \arg\max_{\mathbf{w}} (\ln (\text{Likelihood}(\mathbb{I} \mid \mathbf{w}))) \]

i.e.,

\[ \mathbf{w}^* = \arg\max_{\mathbf{w}} \left( \sum_{i=1}^{d} \sum_{j=1}^{k} w_j \cdot \mathcal{I}^{(i)}(\phi_j) - d \cdot \ln (\text{WMC}(\top \mid \mathbf{w})) \right) \]

\[ \mathbf{w}^* = \arg\max_{\mathbf{w}} (n_j \cdot w_j - d \cdot \ln (\text{WMC}(\top \mid \mathbf{w}))) \]

where \( n_j \) is the number of observations \( \mathcal{I}^{(i)} \) for which the formula \( \phi_j \) 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.,

\[
\frac{\partial \text{logLik}(\mathbb{I} \mid \mathbf{w})}{\partial w_i} = 0
\]

where

\[
\text{logLik}(\mathbb{I} \mid \mathbf{w}) = n_j \cdot w_j - d \cdot \ln \left( \text{WMC}(\top \mid \mathbf{w}) \right)
\]

**Problem:** calculating \( \frac{\partial \ln(\text{WMC}(\top \mid \mathbf{w}))}{\partial w_i} \), i.e.,

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

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 $\phi_1$, then

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

can be computed analytically

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

- **Observation 1:** the more often $\phi_1$ is satisfied in the observation, the larger it’s weight $w_1$
- the more models of $\phi_1$, i.e., the larger $\#\text{SAT}(\phi_1)$ the smaller $w_1$. 
Special case: we only have \textbf{one formula} $\phi : w$

Derivation of the formula (10).

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

   \[
   \text{Likelihood}(I \mid w) = \prod_{i=1}^{d} \frac{1}{\text{WMC}(\top \mid w)} \exp \left( w \cdot \mathcal{I}^{(i)}(\phi) \right)
   \]
   \[
   = \text{WMC}(\top \mid w)^{-d} \exp \left( \sum_{i=1}^{d} w \cdot \mathcal{I}^{(i)}(\phi) \right)
   \]
   \[
   = \text{WMC}(\top \mid w)^{-d} \exp (n \cdot w)
   \]

2. We then determine the logarithm of the likelihood

   \[
   \text{LogLike}(I \mid w) = n \cdot w - d \cdot \log(\text{WMC}(\top \mid w))
   \]

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

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

   \[
   \frac{\partial \text{LogLike}(I \mid w)}{\partial w} = n - d \cdot \left( \frac{1}{\text{WMC}(\top \mid w)} \right) \cdot \frac{\partial \text{WMC}(\top \mid w)}{\partial w}
   \]
   \[
   = n - d \cdot \left( \frac{e^{w} \cdot \#\text{SAT}(\phi)}{e^{w} \cdot \#\text{SAT}(\phi) + \#\text{SAT}(\neg \phi)} \right)
   \]
We then pose the derivative equal to 0

\[
0 = \frac{\partial \log \text{Like}(\mathbb{I} | w)}{\partial w}
\]

\[
0 = n - d \cdot \left( \frac{e^w \cdot \# \text{SAT}(\phi)}{e^w \cdot \# \text{SAT}(\phi) + \# \text{SAT}(\neg \phi)} \right)
\]

\[
d \cdot \left( \frac{e^w \cdot \# \text{SAT}(\phi)}{e^w \cdot \# \text{SAT}(\phi) + \# \text{SAT}(\neg \phi)} \right) = n
\]

\[
d \cdot e^w \cdot \# \text{SAT}(\phi) = n \cdot e^w \cdot \# \text{SAT}(\phi) + n \cdot \# \text{SAT}(\neg \phi)
\]

\[
e^w = \frac{n \cdot \# \text{SAT}(\neg \phi)}{(d - n) \# \text{SAT}(\phi)}
\]

\[
w = \log \left( \frac{n \cdot \# \text{SAT}(\neg \phi)}{(d - n) \# \text{SAT}(\phi)} \right)
\]
Example of learning weights

Example

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

<table>
<thead>
<tr>
<th>#</th>
<th>Itemsets</th>
<th>$a \ w = \log \left( \frac{15 \cdot 2^6}{7 \cdot 2^6} \right) \approx 0.76$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$a \ b \ c \ d$</td>
<td>$\neg a \ w = \log \left( \frac{7 \cdot 2^6}{15 \cdot 2^6} \right) \approx -0.76$</td>
</tr>
<tr>
<td>1</td>
<td>$a \ b \ e \ f$</td>
<td>$e \ w = \log \left( \frac{1 \cdot 2^6}{21 \cdot 2^6} \right) \approx -3.04$</td>
</tr>
<tr>
<td>7</td>
<td>$a \ b \ c$</td>
<td>$\neg e \ w = \log \left( \frac{21 \cdot 2^6}{1 \cdot 2^6} \right) \approx 3.04$</td>
</tr>
<tr>
<td>3</td>
<td>$a \ c \ d \ f$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$d$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$d$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$d \ g$</td>
<td></td>
</tr>
</tbody>
</table>
Example of learning weights

Example

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

<table>
<thead>
<tr>
<th>#</th>
<th>Itemsets</th>
<th>$a \land b$ w = $\log \frac{12 \cdot (2^7 - 2^5)}{10 \cdot 2^5} \approx 8.21$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$a\ b\ c\ d$</td>
<td>$c \land d$ w = $\log \frac{7 \cdot (2^7 - 2^5)}{15 \cdot 2^5} \approx 7.27$</td>
</tr>
<tr>
<td>1</td>
<td>$a\ b\ e\ f$</td>
<td>$e \land f$ w = $\log \frac{1 \cdot (2^7 - 2^5)}{21 \cdot 2^5} \approx 4.99$</td>
</tr>
<tr>
<td>7</td>
<td>$a\ b\ c$</td>
<td>$a \rightarrow b$ w = $\log \frac{19 \cdot (2^7 - 3 \cdot 2^5)}{3 \cdot 3 \cdot 2^5} \approx 0.75$</td>
</tr>
<tr>
<td>3</td>
<td>$a\ c\ d\ f$</td>
<td>$a \land b \land c \land \neg e \land \neg f \rightarrow g$ w = $\log \left( \frac{11}{11 \cdot (2^7 - 2)} \right) \approx -4.84$</td>
</tr>
<tr>
<td>2</td>
<td>$a\ c\ d\ f$</td>
<td>$a \land b \land \neg c \land \neg d \land e \land f \land \neg g$ w = $\log(21 \cdot (2^7 - 1)) \approx 7.89$</td>
</tr>
</tbody>
</table>


