How to Compute Probability in MLN: An Tutorial in Examples *

Guangchun Cheng

March 8, 2012

1 Preliminary

This article illustrates how to compute the probability in Markov logic network (MLN) using examples. Also we show how the Markov logic "softens" the rigorous constraints in First-order logic but still keep its powerful express ability. From last blog post, we showed that each first-order logic rule is corresponding to one (actually a series of grounded) clique in Markov networks, therefore, the following formula is to compute the satisfiability of the FOL formula from facts.

$$P(X = x) = \frac{1}{Z} exp\left(\sum_i w_i n_i(x)\right) \tag{1}$$

where $n_i(x)$ is "the number of true groundings of FOL formula F_i in x". One may ask a question: since the value of X is already assigned, i.e. x, then the truth value of formula F_i is also determined with either TRUE or FALSE, why are there several different groundings that make F_i true? To understand this, we have to understand one important property of MLN, this is:

"*An MLN can be viewed as a* ***template*** *for constructing Markov networks.*"

This is an important sentence understanding MLN and the computation of the probability. It is probably the bridge between first-order logic and Markov networks. This **template** MLN (weighted first-order formulas according to my understanding) has different instantiations, each of which is part of X or X's component. For example, the FOL formula

$$w \qquad (\forall x) Smokes(x) \Rightarrow Cancer(x) \tag{2}$$

may have a lot of grounded instances, such as

$$Smokes(A) \Rightarrow Cancer(A)$$

$$Smokes(B) \Rightarrow Cancer(B)$$

*More can be found at http://guangchun.wordpress.com/

$$Smokes(C) \Rightarrow Cancer(C)$$

......

Then in equation (1), the variable X should be like X=(Smokes(A)=1 Cancer(A)=1 Smokes(B)=0 Cancer(B)=1 Smokes(C)=1 Cancer(C)=0,...). We can see that the instantiation (grounding) (Smokes(A), Cancer(A)) becomes part of variable X (first two components in this case). The same for B and C. Suppose we only have $\{A, B, C\}$, then $n_i = 2$ because the formula in (2) is false under (Smokes(C)=1, Cancer(C)=0), and true under both (Smokes(A)=1, Cancer(A)=1) and (Smokes(B)=0, Cancer(B)=1).

Equipped with these preliminary knowledge, let's look at how the probability in equation (1) is computed through three examples. The first one is the simplest with only one first-order formula and only one constant. The second one deals with more constants, and the third extends to multiple first-order formulas.

2 Single Formula Single Constant

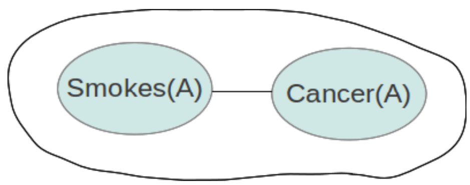

Figure 1: Single Formula Single Constants

In this simple example, we suppose first-order formula (2) is the only rule we have, and its weight is w. There is only one constant $\{A\}$. See figure 1 for the corresponding Markov network. The X in equation (1) is a 2D variable $(Smokes(A), Cancer(A))$. Thus there are totally 4 different probability in this world, or say 4 different possible worlds, or 4 different states of the world: $x \in \{(0,0),(0,1),(1,0),(1,1)\}$.

Since we only have one first-order formula, i can only be $\{1\}$ in equation (1), and the summation is only for one item. Following this and the discussion in previous section, we have

$$P(0,0) = \frac{1}{Z}e^{w*1} = \frac{e^w}{Z}, \qquad P(0,1) = \frac{1}{Z}e^{w*1} = \frac{e^w}{Z}$$

$$P(1,0) = \frac{1}{Z}e^{w*0} = \frac{1}{Z}, \qquad P(1,1) = \frac{1}{Z}e^{w*1} = \frac{e^w}{Z}$$

where $Z = e^w + e^w + e^0 + e^w = 3e^w + 1$ is the partition factor. $P(1,0) = \frac{1}{3e^w+1}$ is because when Smokes(A)=1, Cancer(A)=0, the truth value of formula (2) is false and thus $n_1 = 0$. But it is true for each of the other three groundings, which gives $n_1 = 1$.

This finishes MLN modeling: get joint distribution P(Smokes(A), Cancer(A))

Now let's verify if MLN generalizes (or softens) first-order logic. First, it is obvious that the probability for (Smokes(A)=1, Cancer(A)=0) is not zero but $P(1,0) = \frac{1}{3e^w+1}$. So it does "softens" the canonical first-order logic requirement of either TRUE or FALSE. Now it gets some in-betweens. Second, when given fact that A smokes (Smokes(A)=1), the probability A has cancer, using Bayes theory, is

$$P(Cancer(A) = 1|Smokes(A) = 1) = \frac{P(Cancer(A) = 1, Smokes(A) = 1)}{P(Smokes(A) = 1)}$$

where the marginal probability

$$P(Smokes(A) = 1) = \frac{1}{Z} + \frac{e^w}{Z} = \frac{e^w + 1}{Z}$$

and the joint probability

$$P(Cancer(A) = 1, Smokes(A) = 1) = \frac{1}{Z}e^{w*1} = \frac{e^w}{Z}$$

Therefore,

$$P(Cancer(A) = 1|Smokes(A) = 1) = \frac{e^w}{e^w + 1} \tag{3}$$

From (3) we can see

$$\lim_{w\to+\infty} P(Cancer(A) = 1|Smokes(A) = 1) = \lim_{w\to+\infty} \frac{e^w}{e^w + 1} = 1$$

which is consistent with first-order logic. The same conclusion can be achieved for $P(Cancer(A) = 1|Smokes(A) = 0)$ and $P(Cancer(A) = 0|Smokes(A) = 0)$.

3 Single Formula Multiple Constants

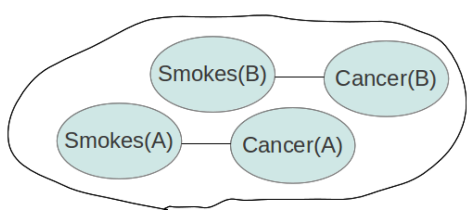

Figure 2: Single Formula Multiple Constants

Suppose we have one more person in our system Bob (B), but still we only have one rule, the first-order formula(2). Now we have two grounded formulas, as shown in figure 2. Now it may be the best time to understand "An MLN can be viewed as a **template** for constructing Markov networks," because both grounded formulas are derived from the same first-order formula (2).

The world we are modelling this time is in 4D: X=(Smokes(A), Cancer(A)Smokes(B), Cancer(B)). We have to obtain the probability P(Smokes(A), Cancer(A)Smokes(B), Cancer(B)) through (3) when given different truth value assignments to each of the four components.

What's the range of i in equation (1)? Is it $\{1, 2\}$ because we now have two grounded formulas ans thus two cliques? **NO!** Still the i can only be in $\{1\}$, not $\{1, 2\}$. Careful readers may have notice that true groundings have already been characterized in n_i, or in order words, the equation (1) can be written as

$$P(X = x) = \frac{1}{Z} exp \left(\sum_{i}^{f} \sum_{j}^{n_i(x)} w_i \right) \tag{4}$$

where f is the number of first-order formulas (not grounded ones). This formula can be read as "for each first-order formula, add the weights whenever its grounding is true". Pay attention to the order in this sentence.

Since we only look at 4D binary world here, each component of X can only be 0 or 1, we totally have $2^4 = 16$ different possible states of the world. When computing $P(X)$, we assume that we don't have only facts or evidences (such as $Smokes(A) = 0$). I think this is equivalently saying: each of the 16 different states have equal happening probability. If we think in this way, the computation method we discuss here can be easily extended for inference in MLNs, where the only difference is not all states happen and the happening probabilities are different–such facts show up several times while others do not.

First, let's look at how the probability P(X=(0,0,0,0)) is computed as an example, i.e. Smokes(A)=0, Cancer(A)=0, Smokes(B)=0, Cancer(B)=0. According to equation (1) or (4), how many true groundings of first-order formula (2) in (0,0,0,0)?

- Smokes(A)=0 and Cancer(A)=0, so $Smokes(A) \Rightarrow Cancer(A)$ is true;
- Smokes(B)=0 and Cancer(B)=0, so $Smokes(A) \Rightarrow Cancer(A)$ is true;

Therefore, there are $n_i(x) = 2$ true grounding of only formula F_i ($i = 1$), i.e. formula (2). We only have one formula, so

$$P(0, 0, 0, 0) = \frac{1}{Z} exp \left(\sum_{i=1}^{1} \sum_{j}^{n_i(x)} w_i \right) = \frac{1}{Z} exp \left(\sum_{j=1}^{2} w_1 \right) = \frac{1}{Z} e^{w*2}$$

Similarly, we can get the probabilities of different "possible worlds", shown in Table 1 (Z omitted). From Table 1, we can get $Z = 9e^{2w} + 6e^{w} + 1$.

Once joint distribution $P(X)$ is obtained, different probabilities such as conditional probability and marginal probability can be easily computed. For example (C short for Cancer, S for Smokes)

$$\begin{aligned} & P(C(A) = 1, C(B) = 1 | S(A) = 1, S(B) = 1) \\ = & \frac{P(C(A) = 1, C(B) = 1, S(A) = 1, S(B) = 1)}{P(S(A) = 1, S(B) = 1)} \end{aligned} \tag{5}$$

Table 1: Probabilities of different worlds

X				$n_i(x)$	$e^{n_i(x)*w_i}$
Smokes(A)	Cancer(A)	Smokes(B)	Cancer(B)		
0	0	0	0	2	e^{2w}
0	0	0	1	2	e^{2w}
0	0	1	0	1	e^{w}
0	0	1	1	2	e^{2w}
0	1	0	0	2	e^{2w}
0	1	0	1	2	e^{2w}
0	1	1	0	1	e^{w}
0	1	1	1	2	e^{2w}
1	0	0	0	1	e^{w}
1	0	0	1	1	e^{w}
1	0	1	0	0	1
1	0	1	1	1	e^{w}
1	1	0	0	2	e^{2w}
1	1	0	1	2	e^{2w}
1	1	1	0	1	e^{w}
1	1	1	1	2	e^{2w}

where the joint probability in the numerator is in Table 1, and the marginal probability in denominator can be got by adding items in Table 1 where (Smokes(A)=1 and Smokes(B)=1):

$$P(S(A) = 1, S(B) = 1) = \frac{e^{2w} + 2ew + 1}{Z}$$

Therefore, the probability in (5) is

$$P(C(A) = 1, C(B) = 1|S(A) = 1, S(B) = 1) = \frac{e^{2w}}{e^{2w} + 2e^{w} + 1}$$

It is 1 when $w \to +\infty$.

Similarly, we can get $P(C(A) = 0, C(B) = 0|S(A) = 1, S(B) = 1) = \frac{1}{e^{2w}+2e^{w}+1}$, and it is zero when $w \to +\infty$.

4 Multiple Formulas Multiple Constants

To keep this article short, we leave this section. The method is the same as the previous, and the only difference is when computing the probability we have to enumerate i over 1,2,3,...,N (N is the number of first-order formulas). But for each i, the procedure is totally the same as it is shown "Single Formula Multiple Constants". For example, if we add the following first-order formula, we will get the grounded formulas shown in figure 3 where different shadows represents different first-order formulas. Note that in the figure, we do not consider the relationship of Friends(A,A) or Friends(B,B).

Do try it by yourself!

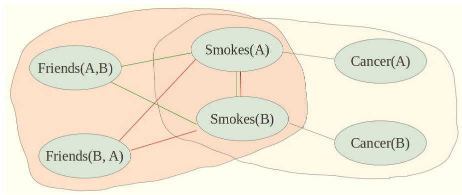

Figure 3: Multiple Formulas Multiple Constants

5 Summary

Some notes for MLN and computation of the joint probability:

1. MLN is a generative model which models the joint distribution of the predicates given constants in a system. The representation of MLN is weighted first-order formulas, and its reasoning is through converting the first-order logic to Markov network.

2. The i in equation (1) runs over different first-order formulas, not the grounded formulas. The Markov network of Markov logic is constructed from the grounded formulas/predicates, but these formulas are grouped based on which formula (*template*) they are derived from. This is reflected as $n_i(x)$ in formula (1), because those "grouped" grounded formulas from the same first-order formula have the same weight w_i.

3. In MLN, the predicates are organised in a hierarchical manner: MLN$\Longrightarrow$FOL groups$\Longrightarrow$grounded formulas. The middle layer (i.e. FOL group) is directly related to the knowledge/rules in first-order logic form.

6 Reference

Matthew Richardson and Pedro Domingos, "Markov Logic Networks." Machine Learning, 62 (2006), pp 107-136.