Two Node Directed Graphical Model Example

Jascha Sohl-Dickstein

March 20, 2012

Let's work with a toy system with one root node T, and two leaf nodes, $E1$ and $E2$. See Figure 1. We know the probabilities $P(T)$, $P(E1|T)$, $P(E2|T)$. There are a few tasks we might want to do.

1. *Infer T if we know $E1$ and $E2$.* We use Bayes' rule to write

$$P(T|E1, E2) = \frac{P(T, E1, E2)}{P(E1, E2)} \tag{1}$$
$$\propto P(T, E1, E2) \tag{2}$$
$$= P(T)P(E1|T)P(E2|T). \tag{3}$$

   This is what the current version of the code does to infer T. We don't have to calculate the $P(E1, E2)$ in the denominator, because it doesn't depend on T, but if we don't calculate it then we have to normalize the resulting expression to make sure the probabilities sum to one. That's what the $\propto$ means – that the answer is proportional to $P(T, E1, E2)$, but doesn't include a normalizing factor.

2. *Infer T if we know $E1$, but are missing data from $E2$.* Because of the directed form our model has, it's easy to exclude leaves that we don't have information

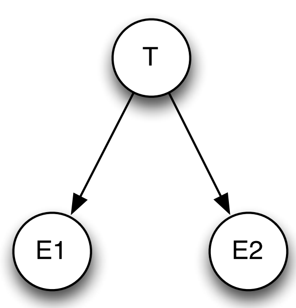

Figure 1: Graphical model for toy system.

about. We use Bayes' rule to write

$$P(T|E1) = \frac{P(T, E1)}{P(E1)} \tag{4}$$

$$\propto P(T, E1) \tag{5}$$

$$= P(T)P(E1|T). \tag{6}$$

We can use this to estimate T in an almost identical way to in item 1, except leaving out the $P(E2|T)$ term.

3. *Predict $E2$ if $E2$ is missing, but we know $E1$.* That is, we want to compute $P(E2|E1)$. This involves a little bit more complicated algebra,

$$P(E2|E1) = \sum_T P(E2, T|E1) \tag{7}$$

$$= \sum_T P(E2|T, E1)P(T|E1) \tag{8}$$

$$= \sum_T P(E2|T)P(T|E1). \tag{9}$$

We used the fact that $E2$ only directly depends on T to say that $P(E2|T, E1) = P(E2|T)$. The sum is over all values of T, in this case $\{0, 1\}$. (The sum is the same as averaging $P(E2|T)$ over the probability distribution $P(T|E1)$. If $P(T|E1)$ was hard to calculate exactly, we could approximate this sum by averaging over samples from $P(T|E1)$.)

So we can estimate the value of missing leaves by calculating the probability distribution over T given the leaves we know, and then calculating the distribution over the missing leaves given the estimated distribution over T.

4. *Update the parameters if we know $E1$, but are missing data from $E2$.* To do this, we calculate $P(T|E1)$ as in item 2. We then update the parameters for $P(T)$ and $P(E1|T)$. We simply leave the parameters for $P(E2|T)$ untouched, since we don't know the value of $E2$. If there are many datapoints, and some of them have $E2$ and others don't, then the parameters for $P(E2|T)$ are just set by averaging over the datapoints for which $E2$ is known.