

Factor Graphs

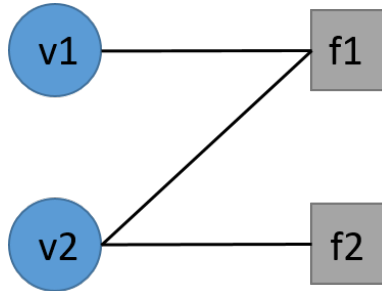
Feiran Wang

1 Variables and Factors

DeepDive uses factor graphs to perform learning and inference. A factor graph is a type of probabilistic graphical model. There are two types of nodes in a factor graph, (random) variables and factors. A random variable can be used to quantitatively describe an event. For example, we can use a random variable to denote if John smokes. If John smokes, the random variable takes a value of 1, and 0 if John does not smoke. For now, DeepDive only supports boolean variables, so we will constrain our discussion to boolean variables.

A factor is a function of variables, and is used to evaluate the relations among variable(s). For example, a function $\text{imply}(A, B)$ means if A , then B . Now suppose we have relation that “if John smokes then he has cancer”. Here we have two variables, one indicating if John smokes, the other indicating if John has cancer. Thus, $\text{imply}(\text{smoke}, \text{cancer})$ expresses the rule above.

The figure shows an example of a factor graph, where v_1 and v_2 are two variables, and f_1, f_2 are two factors. Factor f_1 is connected with v_1 and v_2 , while f_2 is connected with v_2 . We will use this example to illustrate some basic concepts about factor graphs.



2 Possible Worlds and Probabilities

A possible world is a particular possible assignment to every variable, denoted by I . We can also think of it as each variable taking a particular value.

- How many possible worlds in the factor graph above? Each variable can take value 0 or 1, and there are two variables. So we have four possible worlds. The possible worlds are shown in the table below, with each column representing a possible world.

v_1	0	0	1	1
v_2	0	1	0	1

How do we define the probability of a possible world? We define it through factor functions. We give different weight to factor functions, to express the relative influence of each factor on the probability. Factors with larger weight have greater impacts on the probability. The probability of a possible world graph is then defined to be proportional to some measure of weighted combination of factor functions (for how to define such a measure, please refer to [Factor Graphs and

the Sum-Product Algorithm] <http://www.comm.utoronto.ca/~frank/papers/KFL01.pdf>), i.e., for the above graph,

$$\Pr(I) \propto \text{measure}\{w_1 f_1(v_1, v_2) + w_2 f_2(v_2)\}.$$

Here, w_1, w_2 are weights associated with factor functions.

- If f_1 is imply function with weight $w_1 = 1$, and f_2 is isTrue with weight $w_2 = 0.5$ (for explanation on types of factor functions, see http://deeplive.stanford.edu/doc/inference_rule_functions.html). What is the probability of possible world $v_1 = 1, v_2 = 0$ proportional to (in terms of measure)?

Here, $f_1(v_1, v_2) = \text{imply}(1, 0) = 0$, and $f_2(v_2) = \text{isTrue}(0) = 0$. Thus, the answer is easily computed by $\text{measure}(w_1 f_1(v_1, v_2) + w_2 f_2(v_2)) = \text{measure}(1 \cdot 0 + 0.5 \cdot 0) = \text{measure}(0)$.

It is not convenient to express the probability to be proportional to something rather than to have an absolute value. To define absolute probabilities of possible worlds, we can simply normalize the probabilities above against all possible worlds. That is, we define the probability of possible world I as

$$\Pr(I) = \frac{\text{measure}\{w^T f(I)\}}{\sum_J \text{measure}\{w^T f(J)\}},$$

where the sum is over all possible worlds.

- What's the probability of the possible world $v_1 = 1, v_2 = 0$?

3 Marginal Inference and Weight Learning

Now, we can perform marginal inference on factor graphs. A marginal inference is to infer the probability of one variable taking a particular value. For example, if we would like to infer whether John has cancer, and it is expressed using a variable v_1 , this means we would like to infer the probability of $v_1 = 1$. It is straightforward to define the probability of it as just the sum of probabilities of possible worlds that contains the specific value for that variable. This is similar to marginal probability and joint probability. The marginal inference for the event $\{v_1 = 1\}$ is expressed as

$$\Pr\{v_1 = 1\} = \sum_{I: v_1=1} \Pr(I).$$

- What is the result of the marginal inference $\Pr\{v_1 = 1\}$?

In DeepDive, you can assign factor weights manually, or you can let DeepDive learn weights automatically. In order to learn weights automatically, you must have enough training data available. DeepDive chooses weights that agree most with the training data. Formally, the training data is just set of possible worlds, and we choose weights by maximizing the probabilities of these possible worlds.