Computational Intelligence 2008–2009

Tutorial III

Peter Lucas and Marina Velikova and Nivea Ferreira
Institute for Computing and Information Sciences
University of Nijmegen

Probabilistic reasoning using Pearl's algorithm

Pearl's algorithm is an object-oriented algorithm for probabilistic reasoning, assuming that the question of how to compute the marginal probability distribution $P(V_i)$ for any vertex V_i in a Bayesian network can be answered in terms of the information which should be sent by other vertices in the graph to V_i in order to make it possible for V_i to compute this probability distribution locally. This information is described in terms of two types of *messages*: messages sent to the vertex in the same direction as the arcs in the graph, called *causal parameters*, and messages in the reverse direction, called *diagnostic parameters*. As there may be more than one incoming/outgoing arc for any vertex V_i, those messages must actually be combined by V_i, to obtain what are called a *compound causal* $\pi(V_i)$ and a *compound diagnostic parameter* $\lambda(V_i)$. Those compound parameters are again combined in order to compute the marginal probability $P^*(V_i)$ (known as the *data fusion lemma*):

$$P^*(V_i) = \alpha \cdot \pi(V_i) \cdot \lambda(V_i)$$

where α is a normalisation constant to ensure that $P^*(v_i) + P^*(\neg v_i) = 1$.

A node V_i can compute its compound causal parameter from its own probability function (expressed in terms of conditional probability values) and the causal parameter it receives from its parents:

$$\pi(V_i) = \sum_{\rho(V_i)} P(V_i \mid \rho(V_i)) \cdot \prod_{j=1}^{m} \pi_{V_i}^{V_j}(V_j)$$

with parents $\rho(V_i) = V_1 \wedge \cdots \wedge V_j \wedge \cdots \wedge V_m$ The compound causal parameter for a vertex describes the combined influence on this vertex probabilities of all evidence entered for all its non-descendants.

Further, the compound diagnostic parameter of V_i is computed from the separate diagnostic parameter that it receives from its children:

$$\lambda(V_i) = \prod_{j=1}^{m} \lambda_{V_j}^{V_i}(V_i)$$

The compound diagnostic parameter for a vertex describes the combined influence of all evidence that has been entered for its descendants.

A vertex V_i can calculate the causal parameter that it needs to send to a successor V_{i_j} from its compound causal parameter and the diagnostic parameters it receives from its *other* successors:

$$\pi^{V_i}_{V_{i_j}} = \alpha \cdot \pi_{V_i}(V_i) \cdot \prod_{k=1,\ldots,m,k\neq j} \lambda^{V_i}_{V_{i_k}}(V_i)$$

Finally, a vertex V_i can calculate the diagnostic parameter that it needs to send to a predecessor V_{j_k} from its own probability function and the diagnostic parameter it receives from its successors:

$$\lambda^{V_{j_k}}_{V_i}(V_{j_k}) = \alpha \cdot \sum_{V_i} \lambda_{V_i}(V_i) \cdot [\sum_{\rho(V_i)\setminus\{V_{i_k}\}} P(V_i \mid (\rho(V_i) \setminus \{V_{i_k}\}) \wedge V_{i_k}) \cdot \prod_{l=,\ldots,n,l\neq k} \pi^{V_{j_l}}_{V_i}(V_{j_i})]$$

The causal parameter is a parameter that a node sends to a descendant to provide this descendant with information regarding its non-descendants, while a diagnostic parameter that a node sends to its predecessor in order to provide the information concerning the vertices in the subtree whose root is the vertex at hand.

Computation rules for the separate causal and diagnostic parameters allow a vertex to correctly propagate a piece of given evidence to its neighbours. Note that he topology of directed trees and singly-connected networks determines that at most one chain exists between any two nodes. Consequently, parameters updating originated from a node cannot reach this node from another chain.

Note that those compound parameters have to take entered evidence in the Bayesian network into account, which means that the values of those parameters may (but need not!) change after evidence has been entered into the network.

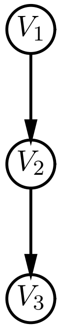
Exercise 1

Let $\mathcal{B} = (G, P)$ be a Bayesian network with directed acyclic graph $G = (V(G), A(G))$ and joint probability distribution P, as shown in Figure 1.

a. Which probabilistic information is needed by vertex V_2 in order to locally compute $P(V_2)$. Give your answer in terms of causal parameters and diagnostic parameters. Use the data fusion lemma to compute $P(V_2)$.

b. Assume that $V_1 = true$ has been entered into the network. Now, compute the causal and diagnostic parameters for $\mathcal{B}$. Also compute the compound causal and diagnostic parameters for V_3. Finally, apply the data fusion lemma to determine $P^*(V_3)$.

c. Now assume that in addition to $V_1 = true$ the value *false* has been entered for V_3. Compute $P^*(V_i)$, for $i = 1, 2, 3$.



$P(v_1) = 0.4, P(\neg v_1) = 0.6$

$P(v_2|v_1) = 0.3, P(\neg v_2|v_1) = 0.7$
$P(v_2|\neg v_1) = 0.9, P(\neg v_2|\neg v_1) = 0.1$

$P(v_3|v_2) = 0.7, P(\neg v_3|v_2) = 0.3$
$P(v_3|\neg v_2) = 0.1, P(\neg v_3|\neg v_2) = 0.9$

Figure 1: Bayesian network.

Exercise 2

Consider Figure 2 and the following probability distribution:

$P(v_4 \mid v_3) = 0.3$ $P(\neg v_4 \mid v_3) = 0.7$
$P(v_4 \mid \neg v_3) = 0.5$ $P(\neg v_4 \mid \neg v_3) = 0.5$
$P(v_3 \mid v_1, v_2) = 0.4$ $P(\neg v_3 \mid v_1, v_2) = 0.6$
$P(v_3 \mid \neg v_1, v_2) = 0.2$ $P(\neg v_3 \mid \neg v_1, v_2) = 0.8$
$P(v_3 \mid v_1, \neg v_2) = 0.7$ $P(\neg v_3 \mid v_1, \neg v_2) = 0.3$
$P(v_3 \mid \neg v_1, \neg v_2) = 0.3$ $P(\neg v_3 \mid \neg v_1, \neg v_2) = 0.7$

$P(v_1) = 0.1$ $P(\neg v_1) = 0.9$
$P(v_2) = 0.8$ $P(\neg v_2) = 0.2$

a. Compute the marginal probability distribution $P(V_4)$ using data fusion.

b. Next, assume that we know that v_2 ($V_2 = \mathit{true}$) holds. What is the value of $P^*(v_2)$ and $P^*(\neg v_2)$ according to data fusion? Compute $P^*(V_4)$, i.e. $P^*(v_4)$ and $P^*(\neg v_4)$, and also $P^*(V_1)$ using data fusion.

Compare your results with those obtained when using standard probability theory.

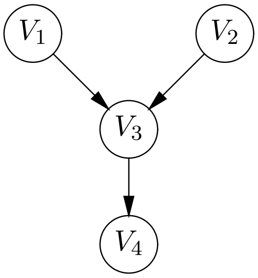


Figure 2: Bayesian network.