Bayesian Belief Network

Phone Message

V 1.1

Introduction

Foundation

D-separation

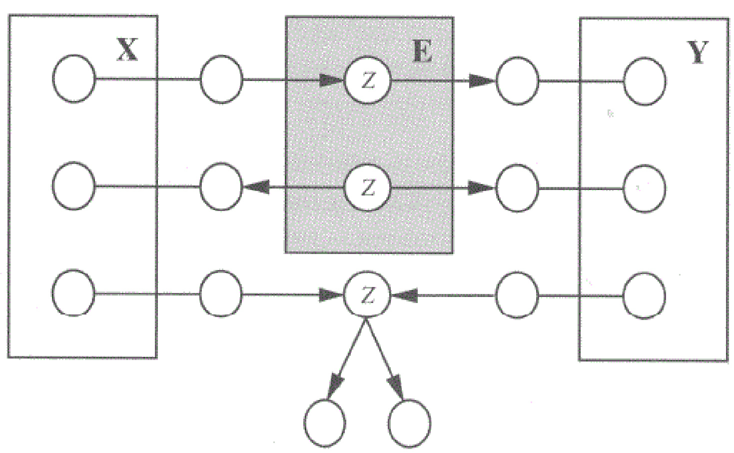

Figure 1. Direction Dependent Separation

A set of nodes E d-separates two sets of nodes of X and Y if every undirected path from a node in X to a node Y is blocked given E. A path is blocked given a set of nodes E if there is a node Z on the path for which one of three conditions hold:

(1) Z is in E and Z has one arrow on the path leading in and one arrow out

(2) Z is in E and Z has both path arrows leading out.

(3) Neither Z nor any descendant of Z is in E, and both path arrows lead in to Z. [1]

To understand the three conditions mentioned above, I will use Figure 4.

(a) Whether Linda is happy or Darnell leaves a message is independent given evidence that there is a message on the answering machine.

(b) Linda being happy and Darnell calling are independent if it is known that Linda leaves a message

(c) Linda being happy and Darnell leaving a message are independent given no evidence at all.

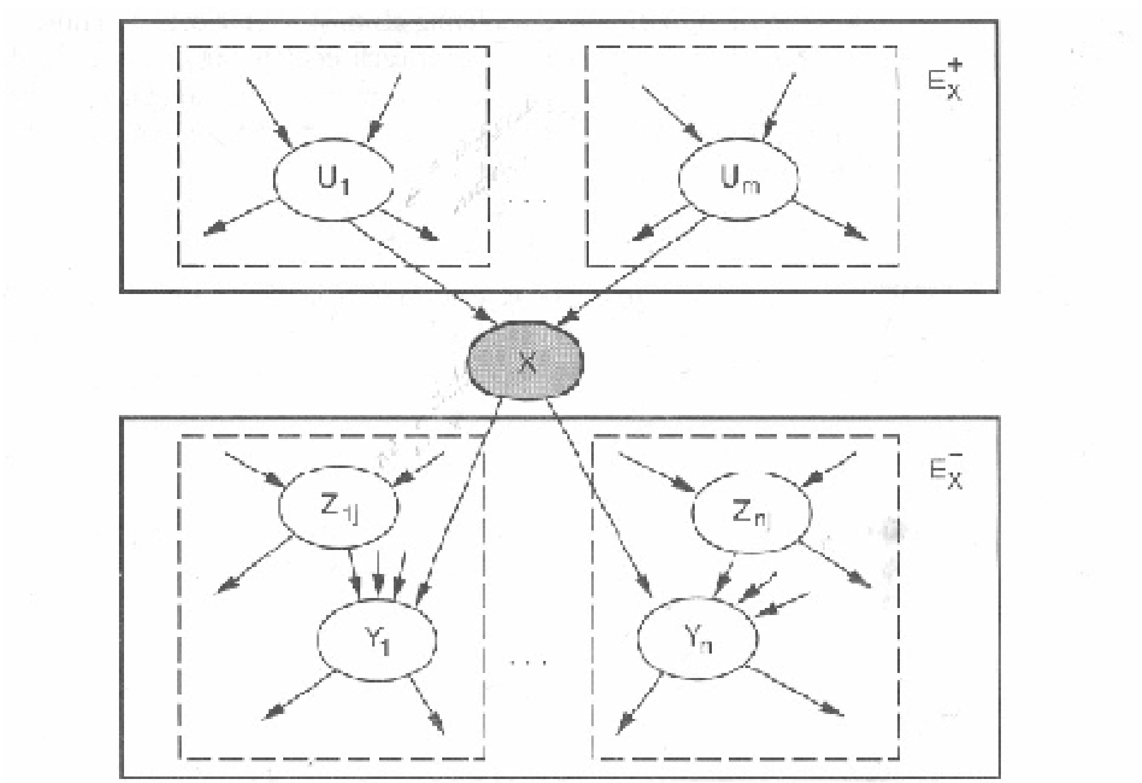

Figure 2. Generic Singly Connected Network. The network is partitioned according to the parents and children of X

$\mathcal{E}^+$ is the causal support for X – the evidence variables “above “ X that are connected to X through its parents.
$\mathcal{E}^-$ is the evidential support for X – the evidence variables “below “ X that are connected to X through its children.[2]

function BELIEF-NET-ASK(X) **returns** a probability distribution over the values of X
inputs: X, a random variable

SUPPORT-EXCEPT(X, *null*)

---

function SUPPORT-EXCEPT(X, V) **returns** $\mathbf{P}(X|E_{X\setminus V})$

if EVIDENCE?(X) **then return** observed point distribution for X
else
calculate $\mathbf{P}(E^-_{X\setminus V}|X)$ = EVIDENCE-EXCEPT(X, V)
$U \leftarrow$ PARENTS[X]
if U is empty
then return $\alpha\, \mathbf{P}(E^-_{X\setminus V}|X)\, \mathbf{P}(X)$
else
for each U_i **in** U
calculate and store $\mathbf{P}(U_i|E_{U_i\setminus X})$ = SUPPORT-EXCEPT(U_i, X)
return $\alpha\, \mathbf{P}(E^-_{X\setminus V}|X) \sum_{\mathbf{u}} \mathbf{P}(X|\mathbf{u}) \prod_i \mathbf{P}(U_i|E_{u_i\setminus X})$

---

function EVIDENCE-EXCEPT(X, V) **returns** $\mathbf{P}(E^-_{X\setminus V}|X)$

$\mathbf{Y} \leftarrow$ CHILDREN[X] $-$ V
if Y is empty
then return a uniform distribution
else
for each Y_i **in Y do**
calculate $\mathbf{P}(E^-_{Y_i}|y_i)$ = EVIDENCE-EXCEPT(Y_i, null)
$\mathbf{Z}_i \leftarrow$ PARENTS[Y_i] $-$ X
for each Z_{ij} **in** $\mathbf{Z}_i$
calculate $\mathbf{P}(Z_{ij}|E_{Z_{ij}\setminus Y_i})$ = SUPPORT-EXCEPT(Z_{ij}, Y_i)
return $\beta \prod_i \sum_{y_i} P(E^-_{Y_i}|y_i) \sum_{\mathbf{z}_i} \mathbf{P}(y_i|X, \mathbf{z}_i) \prod_j P(z_{ij}|E_{Z_{ij}\setminus Y_i})$

Figure 3.

$$\mathcal{P}(X \mid \mathcal{E}) = \mathcal{P}(X \mid \mathcal{E}^+, \mathcal{E}^-) \quad (1)$$

$$\mathcal{P}(X \mid \mathcal{E}^+, \mathcal{E}^-) = \frac{\mathcal{P}(\mathcal{E}_X^+) \mid X, \mathcal{E}^-)\,\mathcal{P}(X \mid \mathcal{E}^+, \mathcal{E}^-)}{\mathcal{P}(X \mid \mathcal{E}^+, \mathcal{E}^-)} \quad (1a)$$

$$\mathcal{P}(X \mid \mathcal{E}_X^+) = \sum_u \mathcal{P}(X \mid u) \prod_i \mathcal{P}(u_i \mid \mathcal{E}\mathcal{U}_i \setminus X) \quad (1b)$$

$$\mathcal{P}(\mathcal{E}_X^- \mid X) = \prod_i \sum_{\mathcal{Y}_i} \sum_{\mathcal{Z}_i} \mathcal{P}\left(\mathcal{E}^-_{\mathcal{Y}_i} \,\middle|\, X, \mathcal{Y}_i, \mathcal{Z}_i\right) \mathcal{P}\left(\mathcal{E}^+_{\mathcal{Y}_i \setminus X} \,\middle|\, X, \mathcal{Y}_i, \mathcal{Z}_i\right) \mathcal{P}(\mathcal{Y}_i, \mathcal{Z}_i \mid X) \quad (1c)$$

Figure 3 is a pseudo-code representation of the approach used to query a Bayesian network. It is a backward chaining algorithm for solving probabilistic queries,

Message

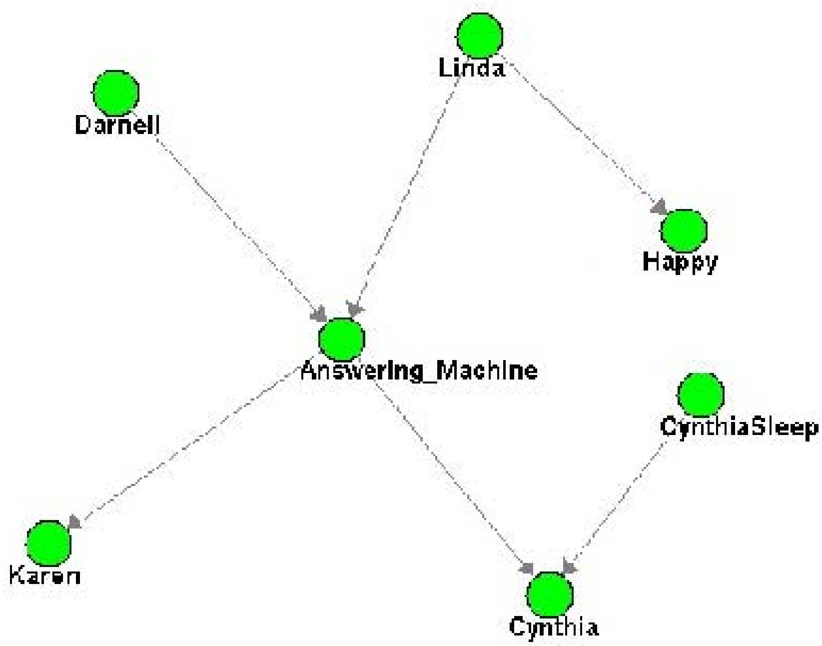

Figure 4.

$$\begin{aligned} L &= \text{Linda} \\ D &= \text{Darnell} \\ M &= \text{Answering Machine} \\ K &= \text{Karen} \\ S &= \text{CynthiaSleep} \\ C &= \text{Cynthia} \\ H &= \text{Happy} \end{aligned}$$

Scenario

$$\mathcal{E}^+ = \{\} \tag{2}$$

$\mathcal{P}(D \mid \mathcal{E}^+)$

The goal here is to find the probability of Darnell leaving a message given the evidence above.

$$\mathcal{P}(D \mid \mathcal{E}^+) = \sum_u \mathcal{P}(D \mid u) \prod_i \mathcal{P}(u_{i|} \mid \mathcal{E}\mathcal{U}_i \setminus D) \tag{3}$$

$$\sum_u \mathcal{P}(D \mid u) \prod_i \mathcal{P}(u_{i|} \mid \mathcal{E}\mathcal{U}_i \setminus D) = \mathcal{P}(D = t \mid \mathcal{E}^+) + \mathcal{P}(D = f \mid \mathcal{E}^+)$$

$\mathcal{P}(L \mid \mathcal{E}^+)$

The goal here is to find the probability of Linda leaving a message given the evidence above. In the case of Darnell, he will either leave a message or won't It is different for Linda because we have three possible outcomes. This will require that we know the probability of each outcome given the evidence above.

$$\mathcal{P}(L \mid \mathcal{E}^+) = \sum_u \mathcal{P}(L \mid u) \prod_i \mathcal{P}(u_{i|} \mid \mathcal{E}\mathcal{U}_i \setminus L) \tag{4}$$

$$\sum_u \mathcal{P}(L \mid u) \prod_i \mathcal{P}(u_{i|} \mid \mathcal{E}\mathcal{U}_i \setminus L) = \mathcal{P}(L = t \mid \mathcal{E}^+) + \mathcal{P}(L = m \mid \mathcal{E}^+) + \mathcal{P}(L = f \mid \mathcal{E}^+)$$

$\mathcal{P}(\mathcal{E}_L^- \mid L = t)$

It is necessary to determine the evidence below Linda given we know that Linda will leave a message. This is done by looking at the evidence variables below Linda that are connected to Linda through its children. Equation (6a) introduces two new variables $\mathcal{V}_M$ and $\mathcal{V}_H$. They help clarify how to calculate $\mathcal{P}(\mathcal{E}_L^- \mid L = t)$. The variable $\mathcal{B}$ is a normalizing constant can be assigned the value of 1.[3]

$$\mathcal{P}(\mathcal{E}_L^- \mid L = t) = \mathcal{B} \times \mathcal{V}_M \times \mathcal{V}_H \tag{5}$$

$\mathcal{V}_M$

Equation (7a) shows that it necessary to use the various probability distributions for some of the random variables above and below $\mathcal{M}$ (answering machine). $\mathcal{P}(\mathcal{E}_M^- \mid m)$ (the evidence beneath the answer machine for each discrete value of M) requires the introduction of two new variables: $\mathcal{V}_{\text{K1}}$ and $\mathcal{V}_{\text{C1}}$.

$$\begin{aligned} \mathcal{V}_M = \sum_m \mathcal{P}(\mathcal{E}_M^- \mid M) \\ \sum_{zm} \mathcal{P}(M \mid L, D) \times \prod \mathcal{P}(\mathcal{Z}\, m,\ j \mid \mathcal{E}_{\text{zm},.j} \setminus M) \end{aligned} \tag{6a}$$

$$\begin{aligned} \mathcal{V}_M = \sum_m \mathcal{P}(\mathcal{E}_M^- \mid m)\, ((\mathcal{P}(m \mid L = t, D = t) \times \mathcal{P}(D = t) + \\ \mathcal{P}(m \mid L = t, D = f) \times \mathcal{P}(D = f)) \end{aligned} \tag{6b}$$

$$\begin{aligned} \mathcal{V}_M = \mathcal{P}(\mathcal{E}_M^- \mid M = t) * (\mathcal{P}(M = t \mid L = t, D = t)\, \mathcal{P}(D = t) + \\ (\mathcal{P}(M = t \mid M = t, D = f)\, \mathcal{P}(D = f)) + \\ (\mathcal{P}(\mathcal{E}_M^- \mid M = f) * \\ (\mathcal{P}(M = f \mid D = t, L = t)\, \mathcal{P}(D = t) + \\ \mathcal{P}(M = f \mid D = f, L = t)\, \mathcal{P}(D = f)) \end{aligned} \tag{6c}$$

$\mathcal{P}(\mathcal{E}_M^- \mid M = t)$

Since the random variable $\mathcal{M}$ has two distinct possibilities (True/False), the first case will look at the evidence below the answering machine given the machine has a message on it.

$$\mathcal{P}(\mathcal{E}_M^- \mid M = t) = \mathcal{B} \times \mathcal{V}_{C1} \times \mathcal{V}_{K1} \tag{7a}$$

$$\mathcal{V}_{K1} = \sum_k \mathcal{P}(\mathcal{E}_k^- \mid K = f) \sum_{zk} \mathcal{P}(K = f \mid M) \times \prod \mathcal{P}(\mathcal{Z}_{k'\,j} \setminus M) \tag{8a}$$

$$\mathcal{V}_{K1} = \mathcal{P}(K = f \mid M = t) \tag{8b}$$

$$\mathcal{V}_{C1} = \sum_c \mathcal{P}(\mathcal{E}_C^- \mid C = t) \sum_{zc} \mathcal{P}(C = t \mid M, S) \times \prod \mathcal{P}(\mathcal{Z}_{C,,j} \setminus C) \tag{9a}$$

$$\mathcal{V}_{C1} = \mathcal{P}(C = t \mid M = t, S = t) \times \mathcal{P}(S = t \mid) + \mathcal{P}(C = t, M = t, S = f) \times \mathcal{P}(S = f \mid) \tag{9b}$$

$\mathcal{P}(\mathcal{E}_M^- \mid M = f)$

$$\mathcal{P}(\mathcal{E}_M^- \mid M = f) = \mathcal{B} \times \mathcal{V}_{K2} \times \mathcal{V}_{C2} \tag{10a}$$

$$\mathcal{V}_{K2} = \sum_k \mathcal{P}(\mathcal{E}_k^- \mid K = f) \sum_{zk} \mathcal{P}(K = f \mid M) \times \prod \mathcal{P}(\mathcal{Z}_{k'\,j} \setminus M) \tag{11a}$$

$$\mathcal{V}_{K2} = \mathcal{P}(K = f \mid M = f) \tag{11b}$$

$$\mathcal{V}_{C2} = \sum_c \mathcal{P}(\mathcal{E}_C^- \mid C = t) \sum_{zc} \mathcal{P}(C = t \mid M, S) \times \prod \mathcal{P}(\mathcal{Z}_{C,,j} \setminus C) \tag{12a}$$

$$\mathcal{V}_{C2} = (C = t \mid M = f, S = t) \times \mathcal{P}(S = t \mid) + \mathcal{P}(C = t, M = f, S = f) \times \mathcal{P}(S = f \mid) \tag{12b}$$

$\mathcal{V}_H$

$$\mathcal{V}_H = \sum_H \mathcal{P}(\mathcal{E}_H^- \mid \mathcal{H}) \sum_{\mathcal{Z}\mathrm{H}} \mathcal{P}(H = t \mid L) \prod_j \mathcal{P}(\mathcal{Z}_{H,,j} \setminus H) \tag{13}$$

$$\mathcal{V}_H = \mathcal{P}(H = t \mid L = f)$$

$\mathcal{P}(\mathcal{E}_L^- \mid L = m)$

$$\mathcal{P}(\mathcal{E}_L^- \mid L = m) = \mathcal{B} \times \mathcal{V}_{M1} \times \mathcal{V}_{H1} \tag{14}$$

$$\mathcal{V}_{M1} = \sum_m \mathcal{P}(\mathcal{E}_M^- \mid m) \sum_{zm} \mathcal{P}(m \mid D, L) \times \prod \mathcal{P}(\mathcal{Z}\mathrm{m},\ j \mid \mathcal{E}_{\mathrm{zm},.j} \setminus M) \tag{15}$$

$$\mathcal{V}_{M1} = \sum_m \mathcal{P}(\mathcal{E}_M^- \mid m)\, ((\mathcal{P}(m \mid D = t, L = m) \times \mathcal{P}(D = t) + \mathcal{P}(M \mid L = m, D = f) \times \mathcal{P}(D = f))$$

$$\mathcal{V}_{\text{M1}} = (\mathcal{P}(\mathcal{E}_M^- \mid M = t) * (\mathcal{P}(M = t \mid L = m, D = t)\,\mathcal{P}(D = t) + (\mathcal{P}(M = t \mid L = m, D = f)\,\mathcal{P}(D = f)) + (\mathcal{P}(\mathcal{E}_M^- \mid M = f) * (\mathcal{P}(M = f \mid D = t, L = m)\,\mathcal{P}(D = t) + \mathcal{P}(M = f \mid D = f, L = m)\,\mathcal{P}(D = f))$$

$\mathcal{V}_{\mathbf{H1}}$

$$\mathcal{V}_{\text{H1}} = \sum_{H} \mathcal{P}(\mathcal{E}_H^- \mid \mathcal{H}) \sum_{\mathcal{Z}\text{H}} \mathcal{P}(H = t \mid L) \prod_{j} \mathcal{P}(\mathcal{Z}_{H,,j} \setminus H) \tag{16}$$

$$\mathcal{V}_{\text{H1}} = \mathcal{P}(H = t \mid L = m)$$

$\mathcal{P}(\mathcal{E}_L^- \mid L = f)$

The approach is similar to the one used to calculate $\mathcal{P}(\mathcal{E}_L^- \mid L = t)$.

$$\mathcal{P}\,(\mathcal{E}_L^- \mid l = F) = \mathcal{B} \times \mathcal{V}_{\text{M2}} \times \mathcal{V}_{\text{H2}} \tag{17}$$

$\mathcal{V}_{\mathbf{M2}}$

$$\mathcal{V}_{\text{M2}} = \sum_{m} \mathcal{P}(\mathcal{E}_M^- \mid m) \sum_{zm} \mathcal{P}(m \mid D, L) \times \prod \mathcal{P}(\mathcal{Z}\text{m}, j \mid \mathcal{E}_{\text{zm},.j} \setminus M) \tag{18}$$

$$\mathcal{V}_{\text{M2}} = \sum_{m} \mathcal{P}(\mathcal{E}_M^- \mid M)\,(\mathcal{P}(M \mid D = t, L = f) \times \mathcal{P}(D = t) + \mathcal{P}(M \mid L = f, D = f) \times \mathcal{P}(D = f)$$

$$\mathcal{V}_{\text{M2}} = (\mathcal{P}(\mathcal{E}_M^- \mid M = t) * (\mathcal{P}(M = t \mid L = f, D = t)\,\mathcal{P}(D = t) + (\mathcal{P}(M = t \mid L = f, D = f)\,\mathcal{P}(D = f)) + (\mathcal{P}(\mathcal{E}_M^- \mid M = f) * (\mathcal{P}(M = f \mid D = t, L = f)\,\mathcal{P}(D = t) + \mathcal{P}(M = f \mid D = f, L = f)\,\mathcal{P}(D = f))$$

$\mathcal{V}_{\mathbf{H2}}$

$$\mathcal{V}_{\text{H2}} = \sum_{H} \mathcal{P}(\mathcal{E}_H^- \mid \mathcal{H}) \sum_{\mathcal{Z}\text{H}} \mathcal{P}(h = T \mid L) \prod_{j} \mathcal{P}(\mathcal{Z}_{H,,j} \setminus H) \tag{19}$$

$$\mathcal{V}_{\text{H2}} = \mathcal{P}(H = t \mid L = f)$$

Verification

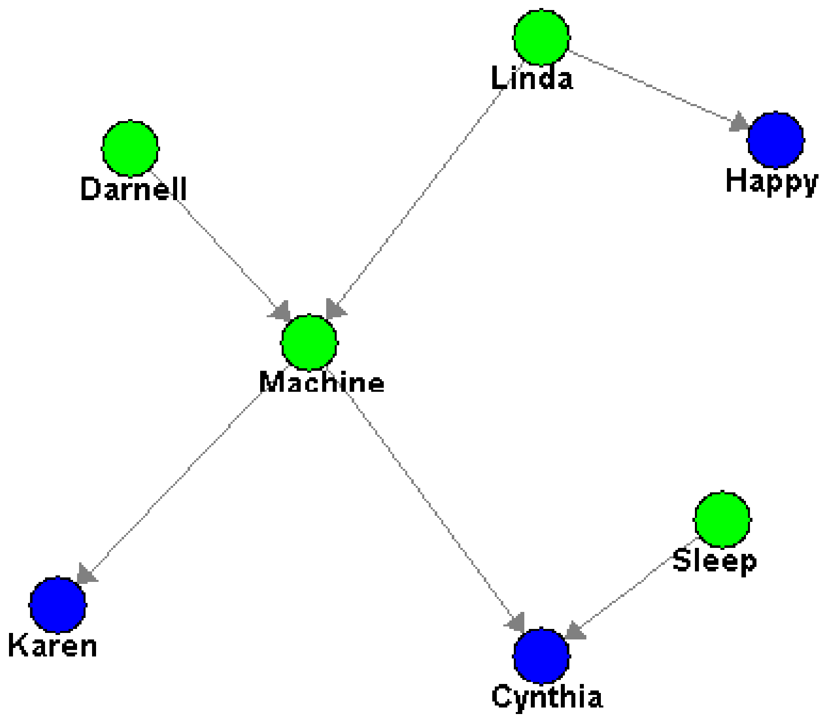

Figure 5.

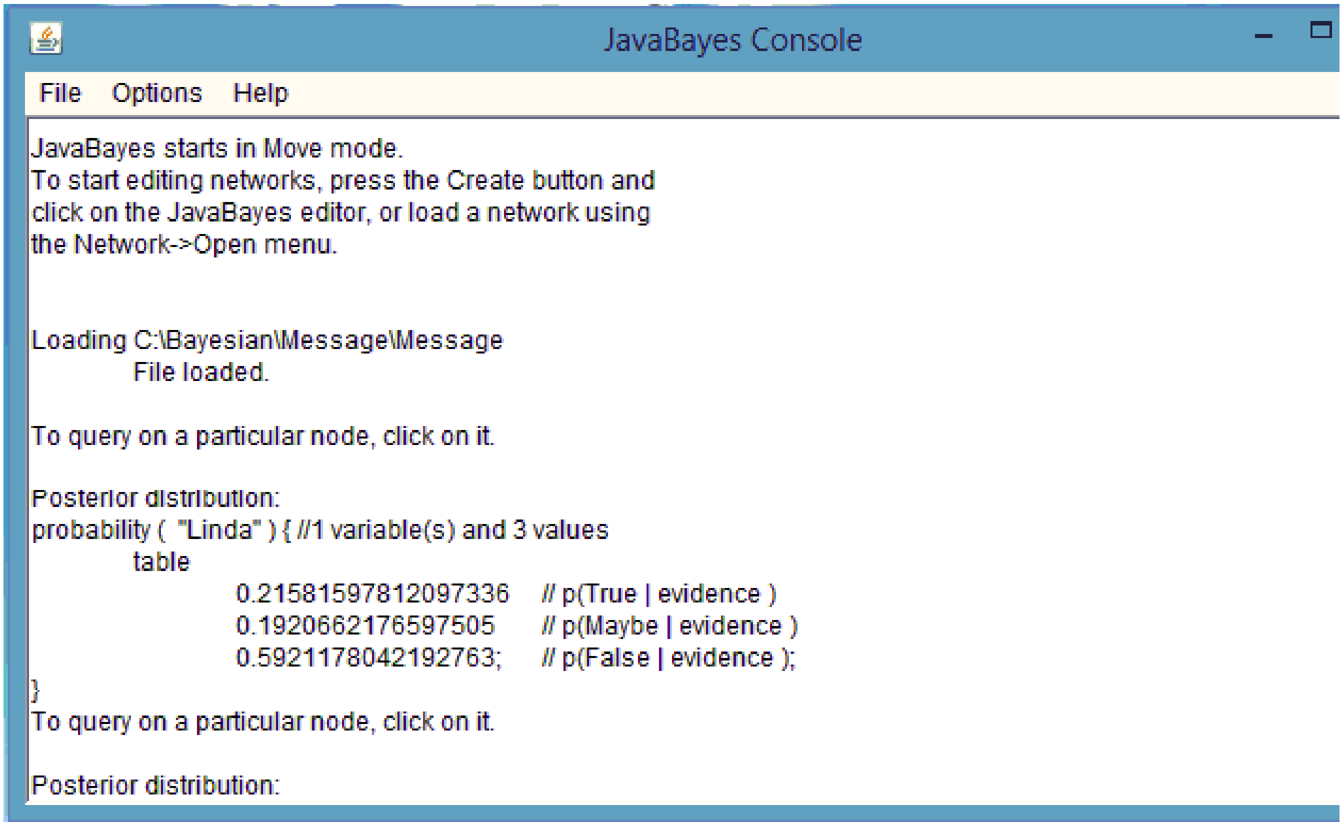

Figure 6.

$$\mathcal{V}_M = 0,03657$$
$$\mathcal{V}_H = 0.0005$$
$$\mathcal{P}(\mathcal{E}_L^- \mid L = t) = 1.8325^{E-5}$$
$$\mathcal{V}_{\text{K1}} = .10$$
$$\mathcal{V}_{\text{C1}} = .77$$
$$\mathcal{P}(\mathcal{E}_M^- \mid M = t) = .077$$
$$\mathcal{V}_{\text{K2}} = .95$$
$$\mathcal{V}_{\text{C2}} = .0385$$
$$\mathcal{P}(\mathcal{E}_M^- \mid M = f) = .036575$$
$$\mathcal{V}_{\text{M1}} = .040650$$
$$\mathcal{V}_{\text{H1}} = .2$$
$$\mathcal{P}(\mathcal{E}_L^- \mid L = m) = .00813$$
$$\mathcal{V}_{\text{M2}} = .0527668$$
$$\mathcal{V}_{\text{H2}} = .95$$
$$\mathcal{P}(\mathcal{E}_L^- \mid L = f) = .05012$$
$$\mathcal{P}(L \Rightarrow t \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = \alpha * .997 * 0.000018 \tag{20}$$
$$\mathcal{P}(L \Rightarrow t \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = .21581$$

$$\mathcal{P}(L \Rightarrow m \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = \alpha * .002 * 0.00813012 \tag{21}$$
$$\mathcal{P}(L \Rightarrow m \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = .1920662177$$
$$\mathcal{P}(L \Rightarrow f \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = \alpha * .001 * 0.50128 \tag{22}$$
$$\mathcal{P}(L \Rightarrow f \mid C \Rightarrow t, K \Rightarrow f, H \Rightarrow t) = .592117$$
$$\alpha = 11812.00207 \tag{23}$$

Notes

(1) Russell, Stuart J., and Peter Norvig. Artificial Intelligence: A Modern Approach. Englewood Cliffs, NJ: Prentice Hall, 1995. Print.

(2) Jensen, Finn V., and Thomas Dyhre. Nielsen. Bayesian Networks and Decision Graphs. New York: Springer, 2007. Print.

(3) Nevatia, Ram. Class. University of Southern California, Los Angeles. July-Aug. 2000. Lecture.