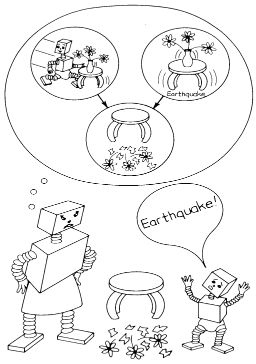
Sesión 6:
Redes Bayesianas - Inferencia



Probabilistic Reasoning in a Causal Network

[Neapolitan 90]

Inferencia en Redes Bayesianas

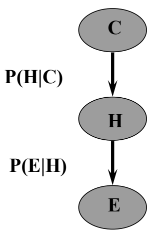
- Introducción
- Propagación en árboles (y poliárboles)
- Algoritmo de eliminación
- Propagación en redes multi-conectadas
  - Condicionamiento
  - Simulación
  - Agrupamiento
- Abducción

Propagación de Probabilidades

El razonamiento probabilístico o propagación de probabilidades consiste en propagar de los efectos de la evidencia a través de la red para conocer la probabilidad *a posteriori* de las variables.

La propagación consiste en darle valores a ciertas variables (evidencia), y obtener la probabilidad posterior de las demás variables dadas las variables conocidas (instanciadas).

Inferencia bayesiana



***Causal*:**
C -> H

***Evidencial*:**
E -> H

***Mixta*:**
C,E -> H

Tipos de Técnicas

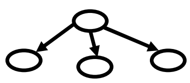
- Calcular probabilidades posteriores:
  - Una variable, cualquier estructura: algoritmo eliminación (*variable elimination*)
  - Cualquier variable, estructuras sencillamente conectadas (árboles, poliárboles): propagación
  - Cualquier variable, cualquier estructura:
    - Agrupamiento (*junction tree*)
    - Simulación estocástica
    - Condicionamiento

Tipos de Técnicas

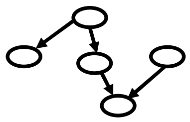
- Obtener variable(s) de mayor probabilidad dada cierta evidencia – abducción:
  - Abducción total
  - Abducción parcial

Tipos de Estructuras

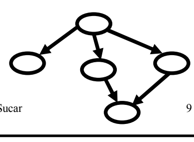
Redes conectadas en forma sencilla:



- **Árboles**
- **Poliárboles**



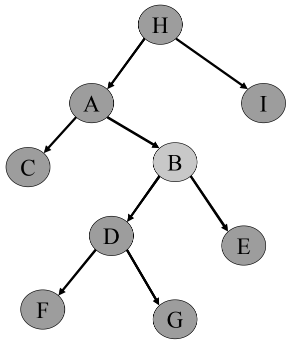
Redes multiconectadas:



Propagación en Árboles

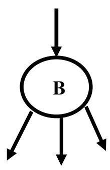
Cada nodo corresponde a una variable discreta, $B\{B_1, B_2,\ldots, B_n)$ con su respectiva matriz de probabilidad condicional, $P(B|A)=P(B_j| A_i)$

Propagación en Árboles

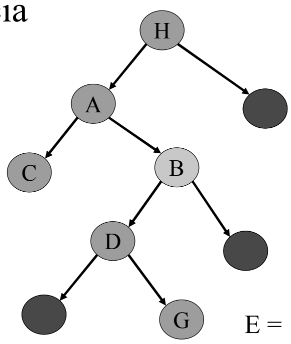


Dada cierta evidencia *E* --representada por la instanciación de ciertas variables-- la probabilidad posterior de cualquier variable *B*, por el teorema de Bayes:

$$P(B_i \mid E)=P(Bi) P(E \mid B_i) / P(E)$$



Evidencia



E = {I,F,E}

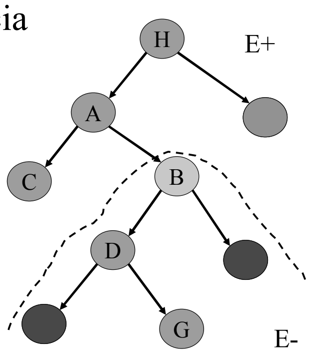
Evidencia

Ya que la estructura de la red es un árbol, el Nodo *B* la separa en dos subárboles, por lo que podemos dividir la evidencia en dos grupos:

E-: Datos en el árbol que cuya raíz es *B*

E+: Datos en el resto del árbol

Evidencia



Entonces:

$P(B_i \mid E) = P(B_i)\, P(E^-, E^+ \mid B_i) / P(E)$

Pero dado que ambos son independientes y aplicando nuevamente Bayes:

$P(B_i \mid E) = \alpha\, P(B_i \mid E^+)\, P(E^- \mid B_i)$

Donde α es una constante de normalización

Definiciones:

Si definimos los siguientes términos:

$\lambda(B_i) = P(E^- \mid B_i)$

$\pi(B_i) = P(B_i \mid E^+)$

Entonces:

$P(B_i \mid E) = \alpha\, \pi(B_i)\, \lambda(B_i)$

Desarrollo

- En base a la ecuación anterior, se puede integrar un algoritmo distribuido para obtener la probabilidad de un nodo dada cierta evidencia
- Para ello se descompone el cálculo de cada parte:
  - Evidencia de los hijos (λ)
  - Evidenica de los demás nodos (π)

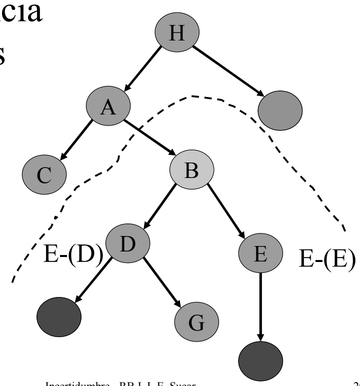
Evidencia de los hijos (λ)

- Dado que los hijos son condicionalmente independientes dado el padre:

$$\lambda (B_i) = P (E^- \mid B_i) = \Pi_k P (E_k^- \mid B_i)$$

- Donde E_k^- corresponde a la evidencia del subárbol del hijo k

Evidencia hijos



Evidencia de los hijos (λ)

- Condicionando respecto a los posibles valores de los hijos de B:

$$\lambda (B_i)= \Pi_k \left[\Sigma_j P (E_k^- \mid B_i, S_j^k) P(S_j^k \mid B_i) \right]$$

- Donde S^k es el hijo k de B, y la sumatoria es sobre los valores de dicho nodo (teorema de probabilidad total)

Evidencia de los hijos (λ)

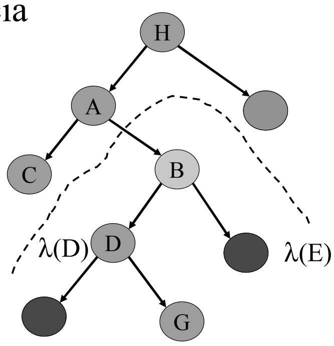
- Dado que B es condicionalmente independiente de la evidencia dados sus hijos:

$$\lambda(B_i) = \Pi_k \left[\Sigma_j P(E_k^- \mid S_j^k)\, P(S_j^k \mid B_i) \right]$$

- Substituyendo la defínción de λ:

$$\lambda(B_i) = \Pi_k \left[\Sigma_j P(S_j^k \mid B_i)\, \lambda(S_j^k) \right]$$

Evidencia hijos



Evidencia de los hijos (λ)

- Recordando que λ es un vector (un valor por cada posible valor de B), lo podemos ver en forma matricial:

$$\boxed{\lambda} = \boxed{\lambda}\ \boxed{P(S \mid B)}$$

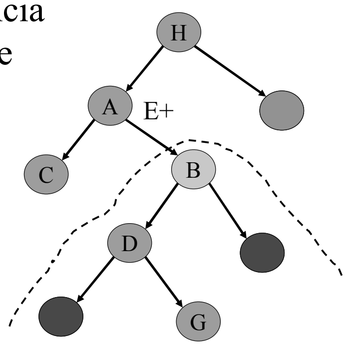
Evidencia de los demás nodos (π)

- Condicionando sobre los diferentes valores del nodo padre (A):

$$\pi\ (B_i) = P\ (B_i \mid E^+) = \Sigma_j\ P\ (B_i \mid E^+, A_j)\ P(A_j \mid E^+)$$

- Donde A_j corresponde a los diferentes valores del nodo padre de B

Evidencia padre



Evidencia de los demás nodos (π)

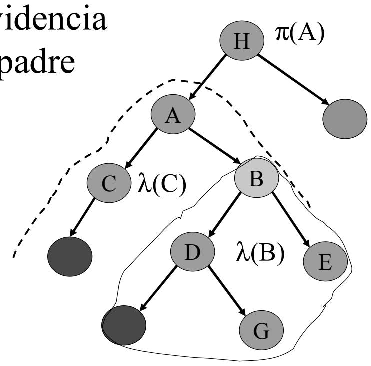
- Dado que B es independiente de la evidencia "arriba" de A dado A:

$$\pi\ (B_i) = \Sigma_j\ P\ (B_i \mid A_j)\ P(A_j \mid E^+)$$

- La $P(A_j \mid E^+)$ corresponde a la P posterior de A dada toda la evidencia excepto B y sus hijos, por lo que se puede escribir como:

$$P(A_j \mid E^+) = \alpha\ \pi\ (A_i)\ \Pi_{k \neq B}\ \lambda_k\ (A_i)$$

Evidencia padre



Evidencia de los demás nodos (π)

- Substituyendo $P(A_j \mid E^+)$ en la ecuación de π :

$$\pi\,(B_i) = \Sigma_j\, P\,(B_i \mid A_j)\,[\,\alpha\, \pi\,(A_i)\, \Pi_{k\neq B}\, \lambda_k\,(A_i)\,]$$

- De forma que se obtiene combinando la π de del nodo padre con la λ de los demás hijos

Evidencia de los demás nodos (π)

- Dado que también π es un vector, lo podemos ver en forma matricial (donde P_A es el producto de la evidencia de padre y otros hijos):

$$\boxed{\pi} = \boxed{P\,(B \mid A)}\;\boxed{P_A}$$

Algoritmo

Mediante estas ecuaciones se integra un algoritmo de propagación de probabilidades en árboles.

Cada nodo guarda los valores de los vectores π y λ, así como las matrices de probabilidad *P*.

La propagación se hace por un mecanismo de paso de mensajes, en donde cada nodo envía los mensajes correspondientes a su padre e hijos:

Mensaje al padre (hacia arriba) -- nodo B a su padre A:

$$\lambda_B(A_i) = \sum_j P(B_j|A_i)\lambda(B_j)$$

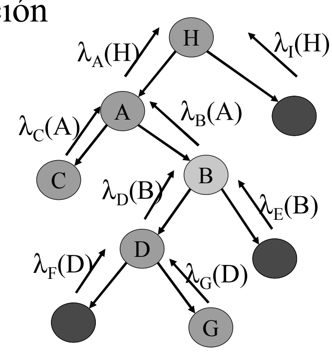
Mensaje a los hijos (hacia abajo) -- nodo B a su hijo S_k :

$$\pi_k(B_i) = \alpha\pi(B_j)\prod_{l\neq k}\lambda_l(B_j)$$

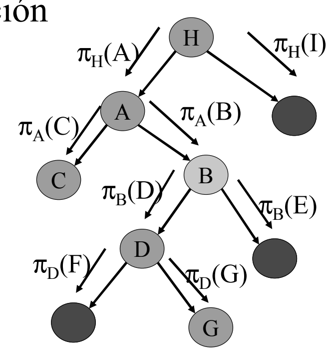
Al instanciarse ciertos nodos, éstos envían mensajes a sus padres e hijos, y se propagan hasta a llegar a la raíz u hojas, o hasta encontrar un nodo instanciado.

Así que la propagación se hace en un solo paso en un tiempo proporcional al diámetro de la red.

Propagación λ



Propagación π



Condiciones Iniciales

- Nodos hoja no conocidos:

  $\lambda(B_i) = [1,1, \ldots]$

- Nodos asignados (conocidos):

  $\lambda(B_i) = [0,0, ..1, 0, \ldots, 0]$ (1 para valor asignado)

  $\pi(B_i) = [0,0, ..1, 0, \ldots, 0]$ (1 para valor asignado)

- Nodo raíz:

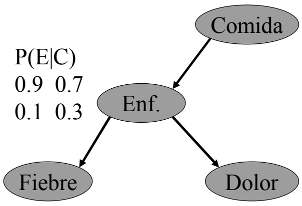
  $\pi(A) = P(A)$, (probabilidad marginal inicial)

Ejemplo

P(C)
0.8 0.2

P(E|C)
0.9 0.7
0.1 0.3

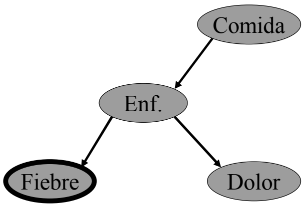
P(F|E)
0.9 0.5
0.1 0.5



P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

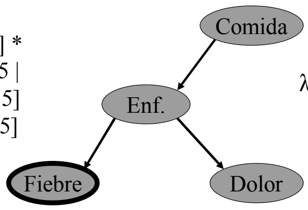
F=si
λ=[1,0]



λ=[1,1]

Ejemplo

λ_F= [1,0] *
[.9 .5 |
.1 .5]
= [.9 .5]



λ_D= [1,1] *
[.7 .4 |
.3 .6]
= [1 1]

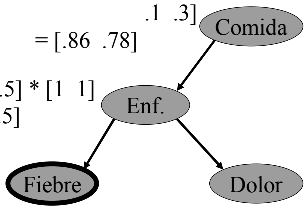
P(F|E)
0.9 0.5
0.1 0.5

P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

$\lambda(C) = [.9\ .5] * [.9\ \ .7|$
$\quad .1\ \ .3]$
$= [.86\ \ .78]$

$\lambda(E) = [.9\ .5] * [1\ \ 1]$
$= [.9\ .5]$



P(E|C)
0.9 0.7
0.1 0.3

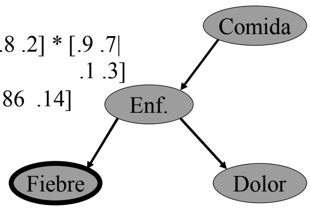
P(F|E)
0.9 0.5
0.1 0.5

P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

$\pi(C) = [.8\ \ .2]$

$\pi(E) = [.8\ .2] * [.9\ .7|$
$\quad .1\ .3]$
$= [.86\ \ .14]$



P(E|C)
0.9 0.7
0.1 0.3

P(F|E)
0.9 0.5
0.1 0.5

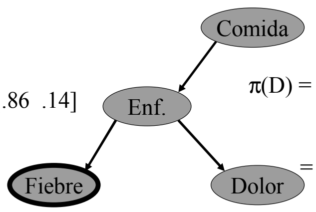
P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

$\pi(C) = [.8\ \ .2]$

$\pi(E) = [.86\ \ .14]$

$\pi(D) = [.86\ .14] * [.9\ \ .5]$
$[.7\ .4|$
$.3\ .6]$
$= [.5698\ \ .2742]$

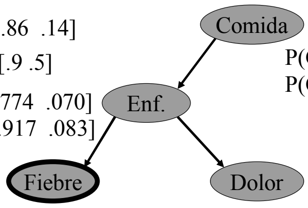


P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

$\pi(C) = [.8\ \ .2]$
$\lambda(C) = [.86\ \ .78]$
$P(C)=\alpha[.688\ \ .156]$
$P(C)= [.815\ \ .185]$

$\pi(E) = [.86\ \ .14]$
$\lambda(E) = [.9\ .5]$
$P(E)=\alpha[.774\ \ .070]$
$P(E)= [.917\ \ .083]$



$\pi(D) = [.57\ \ .27]$
$\lambda(D)=[1,1]$
$P(D)=\alpha[.57\ \ .27]$
$P(D)= [.67\ \ .33]$

Demo 1

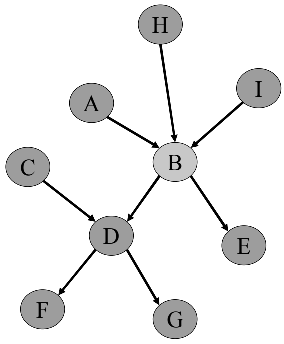
- Ejemplo en MatLab

Propagación en Poliárboles .

Un poliárbol es una red conectada en forma sencilla, pero en la que un nodo puede tener varios padres:

$$P(B \mid A_1, A_2, \ldots, A_n)$$

Propagación en Poliárboles



Algoritmo

- El método es muy similar al de árboles, con algunas consideraciones adicionales:
  - Considerar la probabilidad condicional del nodo dados todos sus padres para el cálculo de π y λ
  - Enviar los mensajes λ a cada uno de los padres de un nodo

Propagación en Redes Multiconectadas

Una red multiconectada es un grafo no conectado en forma sencilla, es decir, en el que hay múltiples trayectorias entre nodos (MCG).

En este tipo de RP los métodos anteriores ya no aplican, pero existen otras técnicas alternativas:

- **Condicionamiento**
- **Simulación estocástica**
- **Agrupamiento**

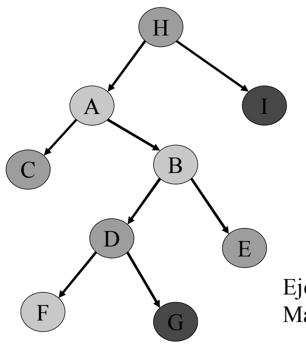
Demo 2

- Propagación en Elvira

Abducción

- La "abducción" se define como encontrar la mejor "explicación" (valores de un cierto conjunto de variables) dada cierta evidencia
- Normalmente se buscan los valores del conjunto "explicación" que tiene mayor probabilidad
- En general, el conjunto de mayor probabilidad NO es igual a los valores individuales de mayor probabilidad

Abducción



Ejemplo:
Max P(A,B,F|G,I)

Referencias

- Pearl 88 – Cap. 4,5
- Neapolitan 90 – Cap. 6,7,8
- Notas Jordan – Cap. 4