Bases Formales de Computación
Redes de Bayes

Camilo Rueda [1]

[1] Universidad Javeriana-Cali

PUJ 2008

Razonamiento bajo incertidumbre

Incertidumbre: Cualidad o estado de ser no conocido con claridad

Razonar (probabilísticamente) bajo incertidumbre:

- razonar con conocimiento incompleto de probabilidades:

  lógica probabilística

- idear representación de distribución probabilística completa de historias de estados:

  redes de Bayes

Redes de Bayes

Razonamiento Bayesiano:

> Inferencia probabilística basada en combinación de información a priori con datos observados

Redes de Bayes:

> Lenguaje gráfico para especificar distribuciones probabilíticas complejas en razonamiento Bayesiano

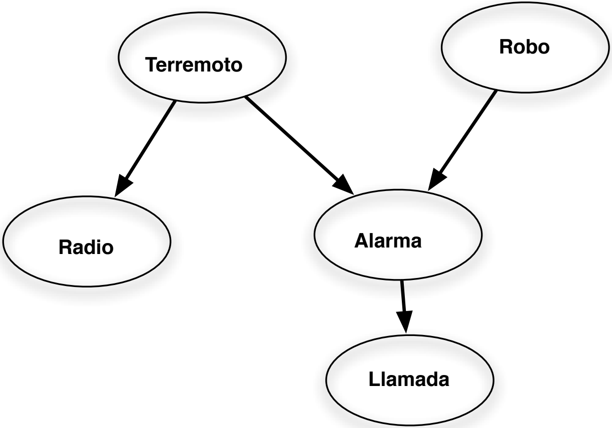
Redes de Bayes: un ejemplo

- Hay o no hay robo en la casa
  (esto no depende de otras cosas en el modelo)
- Hay o no hay terremoto
  (esto no depende de otras cosas en el modelo)
- La alarma puede dispararse por robo o por terremoto
- El vecino hace una llamada basado en si oye la alarma
- el reporte del radio depende de si hubo un terremoto

Cuáles son las rtelaciones causales?

Redes de Bayes: causalidad

Expresar causalidad



Notación

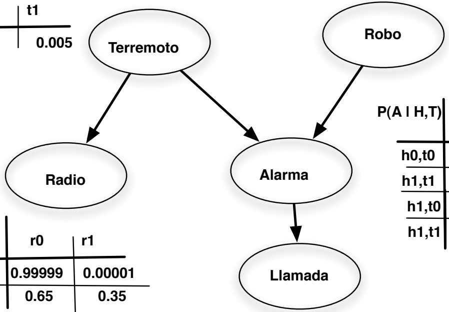
- " Probabilidad de A dado que se conoce (con exactitud) B": $P(A|B)$
- $P(X|Y)$: tabla 2-D con todos los valores: $P(X = x_i|Y = y_j)$
- Probabilidadess condicionales pueden definirse en términos de probabildades incondicionales:

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

- Por consiguiente la regla de la cadena: $P(X,Y) = P(X|Y)P(Y)$
- y teorema de Bayes:
  $P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$
  $P(X|Y) = P(Y|X)P(X)/P(Y)$

Especificar distribución de probabilidad en RB

- Probabilidades a priori de todo nodo raiz
- Probs condicionales de nodos internos, dada toda combinación de predecesores directos



P(T)

t0	t1
0.995	0.005

P(H)

h0	h1
0.99	0.01

P(A I H,T)	a0	a1
h0,t0	0.999	0.001
h1,t1	0.7	0.3
h1,t0	0.2	0.8
h1,t1	0.05	0.95

P(R I T)	r0	r1
t0	0.99999	0.00001
t1	0.65	0.35

P(L I A)	l0	l1
a0	0.95	0.05
a1	0.3	0.7

Distribución conjunta

distribución conjunta:
distribución probabilística sobre todos los datos posibles

Si hay n variables, la DC requiere 2^n valores
Sean las variables: $gripa(g)$, $alergia(a)$, $sinusitis(s)$,

g	a	s	0,027
g	a	$\tilde{s}$	0,003
g	$\tilde{a}$	s	0,162
g	$\tilde{a}$	$\tilde{s}$	0,108
$\tilde{g}$	a	s	0,014
$\tilde{g}$	a	$\tilde{s}$	0,056
$\tilde{g}$	$\tilde{a}$	s	0,0063
$\tilde{g}$	$\tilde{a}$	$\tilde{s}$	0,6237

Definición formal de Red de Bayes

una Red de Bayes $\mathcal{B} = (G, \theta)$ sobre $X_1, \ldots X_n$ es un grafo G en el que:

- cada nodo X_i está asociado con una tabla de probabilidad condicional (TPC): $P(X_i, Padres(X_i))$
- TPC: distribución sobre X_i para cada combinación de valores de los padres de X_i
- $\mathcal{B}$ representa la distribución conjunta sobre $X_1 \ldots X_n$ mediante la regla de la cadena para RB:

$$\boxed{P(x_1, \ldots, x_n) = \prod_i P(x_i \mid Padres(X_i))}$$

Cálculo de probabilidades a partir de DC

Sea la red de Bayes

$$G \rightarrow S \leftarrow A$$

y las TPC:

$$\begin{array}{c|c} g & 0{,}3 \\ \tilde{g} & 0{,}7 \end{array} \qquad \begin{array}{c|c} a & 0{,}1 \\ \tilde{a} & 0{,}9 \end{array} \qquad \begin{array}{c|cccc} & ga & g\tilde{a} & \tilde{g}a & \tilde{g}\tilde{a} \\ s & 0{,}9 & 0{,}6 & 0{,}2 & 0{,}01 \\ \tilde{s} & 0{,}1 & 0{,}4 & 0{,}8 & 0{,}99 \end{array}$$

entonces:

- $P(g, \tilde{a}, s) = P(g)P(\tilde{a})P(s \mid g, \tilde{a}) = 0{,}3 \times 0{,}9 \times 0{,}6 = 0{,}162$
- $P(\tilde{a} \mid s, g) = P(\tilde{a}, s, g)/P(s, g)$
  $P(s, g) = P(g, \tilde{a}, s) + P(g, a, s)$

Tipos de razonamiento

- causal (“hacia abajo”)
- de diagnóstico o evidencial (“hacia arriba”)
  cuáles son las razones más probables de una observación
- intercausal (“explicar descarte de hipótesis”)

Tipos de razonamiento

- causal (“hacia abajo”)
  - probabilidad a priori de robo $P(h) = 0{,}01$
  - probabilidad a priori de terremoto $P(t) = 0{,}005$
  - probabilidad de que el vecino llame $P(l)$ es $0{,}0568$
  - qué tan confiable es el vecino en detectar robos (i.e. $P(l \mid h)$) ?
- de diagnóstico o evidencial (“hacia arriba”)
  cuáles son las razones más probables de una observación
- intercausal (“explicar descarte de hipótesis”)

Tipos de razonamiento

- causal (“hacia abajo”)
- de diagnóstico o evidencial (“hacia arriba”)
  cuáles son las razones más probables de una observación
  - suponer que el vecino llama
  - la probabilidad de robo sube $P(h \mid l) = 0{,}325$
  - la probabilidad de terremoto también sube $P(t \mid l) = 0{,}1034$
- intercausal (“explicar descarte de hipótesis”)

Tipos de razonamiento

- causal (“hacia abajo”)
- de diagnóstico o evidencial (“hacia arriba”)
  cuáles son las razones más probables de una observación
- intercausal (“explicar descarte de hipótesis”)
  supongamos que el radio reporta terremoto.
  Cómo afecta esto las creencias?
  - la probabilidad de terremoto sube sustancialmente
    $P(t \mid l, r) = 0{,}9993$
    (mientras que $P(t \mid l) = 0{,}021$)
  - la probabilidad de robo baja $P(h \mid l, r) = 0{,}0268$

robo y terremoto son independientes pero el reporte del radio explica la llamada

Reducción de parámetros en RB

Número de parámetros independientes para 5 variables es $2^5 - 1 = 31$

En la RB hay 10 parámetros independientes.

qué permite hacer esto?

Reducción de parámetros en RB

Número de parámetros independientes para 5 variables es
$2^5 - 1 = 31$
En la RB hay 10 parámetros independientes.
qué permite hacer esto?

No toda distribución es representable en RB
Solo las que cumplen los supuestos de independencia condicional

La estructura de la RB permite identificar variables condicionalmente independientes

Independencia condicional

Sean $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ conjuntos de variables y P una distribución.

X es independiente de **Y** dado **Z** en P (escrito, $\mathcal{I}(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z})$) si

para toda asignación de valores $\mathbf{x}, \mathbf{y}, \mathbf{z}$ a $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$

$P(\mathbf{x} \mid \mathbf{z}, \mathbf{y}) = P(\mathbf{x} \mid \mathbf{z})$

o sea,

$$\boxed{\begin{array}{l} P(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z}) \\ P(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z}) P(\mathbf{Y} \mid \mathbf{Z}) \end{array}}$$

Independencia condicional (2)

Cómo “fluye ” la influencia de variables en una RB?
Cuándo el observar el valor de X cambia nuestra creencia sobre el valor de Y?

Independencia condicional (2)

Cómo "fluye " la influencia de variables en una RB?
Cuándo el observar el valor de X cambia nuestra creencia sobre el valor de Y?

- Caso obvio: cuando X es padre de Y y también cuando X es hijo de Y:

dos nodos directamente conectados están correlacionados

- cuando X es ancestro o descendiente de Y ?

Independencia condicional (2)

Cómo “fluye ” la influencia de variables en una RB?
Cuándo el observar el valor de X cambia nuestra creencia sobre el valor de Y?

- Caso obvio: cuando X es padre de Y y también cuando X es hijo de Y:

  dos nodos directamente conectados están correlacionados

- cuando X es ancestro o descendiente de Y ? si.
  Cuando observamos robo, la probabilidad de alarma sube por consiguiente, también la de llamada

Independencia condicional: otros casos

si se dispara la alarma, la probabilidad de terremoto aumenta:
$P(t \mid a) > P(t)$
luego también sube la de reporte en el radio.

Independencia condicional: otros casos

Nodos en camino de un arco hacia arriba y luego hacia abajo están correlacionados

Independencia condicional: otros casos

Nodos en camino de un arco hacia arriba y luego hacia abajo están correlacionados

y en camino de uno hacia abajo y luego hacia arriba?
p.ej. observar robo, cambia la probabilidad de terremoto?

Independencia condicional: otros casos

Nodos en camino de un arco hacia arriba y luego hacia abajo están correlacionados

Nodos en camino en forma de V, NO están correlacionados.
$\mathcal{I}(H, T \mid \emptyset)$

Independencia condicional: otros casos

si la alarma se dispara, la probabilidad de llamada aumenta: $(P(l \mid a) > P(l))$
si además hubo un robo, cambia esto la probabilidad de llamada?

Independencia condicional: otros casos

si la alarma se dispara, la probabilidad de llamada aumenta: $(P(l \mid a) > P(l))$
si además hubo un robo, cambia esto la probabilidad de llamada? no!
$P(l \mid a, h) = P(l \mid a)$. O sea, $\mathcal{I}(H, L \mid A)$

Independencia condicional: otros casos

Nodos en camino de dependencia, pero con evidencia de nodos intermedios NO están correlacionados
p.ej. H y L dado A, o también A y R dado T

Independencia condicional: otros casos

Los nodos T y H son independientes. Pero, si se sabe A?

Independencia condicional: otros casos

Los nodos T y H son independientes. Pero, si se sabe A? se vuelven dependientes!

Independencia condicional: otros casos

Nodos en camino de un arco hacia arriba y luego hacia abajo están correlacionados

Nodos en camino en forma de V, NO están correlacionados.
$\mathcal{I}(H, T \mid \emptyset)$

Nodos en camino de dependencia, pero con evidencia de nodos intermedios NO están correlacionados
p.ej. H y L dado A, o también A y R dado T

Nodos en camino en V, pero con evidencia de nodos intermedios (o de sucesores) se vuelven correlacionados

Caminos activos

caminos activos implican correlación (de X con Y)

Caso 1: activo ssi Z no es 1

Caso 2: activo ssi Z no es 1

Caso 3: activo ssi Z no es 1

Caso 4: activo ssi Z o uno de sus descendientes es 1

Camino activo: definición general

Un camino $X_1, \ldots, X_n$ es activo dado $\mathbf{Z}$ si

- cuando existe el camino $X_{i-1} \rightarrow X_i \leftarrow Xi + 1$, entonces X_i o uno de sus descendientes está en $\mathbf{Z}$
- ningún otro nodo en el camino está en $\mathbf{Z}$

Dos nodos X, Y están $d-$separados dado $\mathbf{Z}$ si no hay caminos activos entre ellos.

Cuando X, Y están $d-$separados dado $\mathbf{Z}$ entonces $\mathcal{I}(X, Y \mid Z)$

Inferencia en RB

- $A \rightarrow B$:
  $P(B) = P(a)P(B \mid a) + P(\tilde{a})P(B \mid \tilde{a})$
  en general: $P(B) = \sum_A P(A)P(B \mid A)$
  (componentes en TPC)
- $A \rightarrow B \rightarrow C$:
  $P(c) = \sum_b P(b)P(c \mid b)$ (válido ssi $\mathcal{I}(A, C \mid B)$)
  (componentes en TPC o ya calculados)
- generalizable a cualquier cadena

Complejidad para n variables y tablas de k valores: $O(nk^2)$

Algoritmo de Inferencia en RB: la idea

Por qué no es necesario calcular toda probabilidad conjunta?

$$p(d) = \sum_{a,b,c} P(a,b,c,d)$$

$$= \sum_{a,b,c} P(a)P(b \mid a)P(c \mid b)P(d \mid c)$$

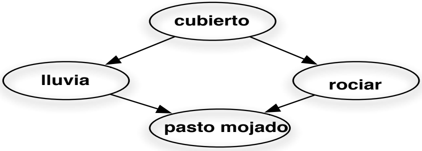
$$= \sum_{a} \sum_{b} \sum_{c} P(a)P(b \mid a)P(c \mid b)P(d \mid c)$$

$$= \sum_{c} \sum_{b} \sum_{a} P(a)P(b \mid a)P(c \mid b)P(d \mid c)$$

$$= \sum_{c} P(d \mid c) \sum_{b} P(c \mid b) \sum_{a} P(a)P(b \mid a)$$

Algoritmo de Inferencia en RB: la idea (2)

Calcular “factores” sucesivamente
$f_A(b) = \sum_a P(a)P(b \mid a)$ que corresponde a $P(B)$
luego $f_B(c) = \sum_b f_A(b)P(c \mid b)$, etc. (programación dinámica)



$$P(m) = \sum_{l,r,c} P(m \mid l,r)P(l \mid c)P(r \mid c)P(c)$$

$$= \sum_{l,r} P(m \mid l,r) \sum_c P(l \mid c)P(c)$$

definimos $f_C(l,r) = \sum_c P(l \mid c)P(r \mid c)P(c)$ y luego,
$\sum_{l,r} P(m \mid l,r) f_1(l,r)$

Algoritmo de Inferencia en RB: eliminación de variables

- el algoritmo opera sobre factores o tablas
- factor f sobre $Y_1, \ldots, Y_n$: función $f(Y_1, \ldots, Y_n) \in \mathbb{R}$
- factores pueden sumarse y multiplicarse

Algoritmo de Inferencia en RB: eliminación de variables

- el algoritmo opera sobre factores o tablas
- factor f sobre $Y_1, \ldots, Y_n$: función $f(Y_1, \ldots, Y_n) \in \mathbb{R}$
  ejemplo:
  - una DC $P(H, T, A, R, L)$ es factor sobre H, T, A, R, L
  - una TPC $P(A \mid H, T)$ es factor sobre H, T
- factores pueden sumarse y multiplicarse

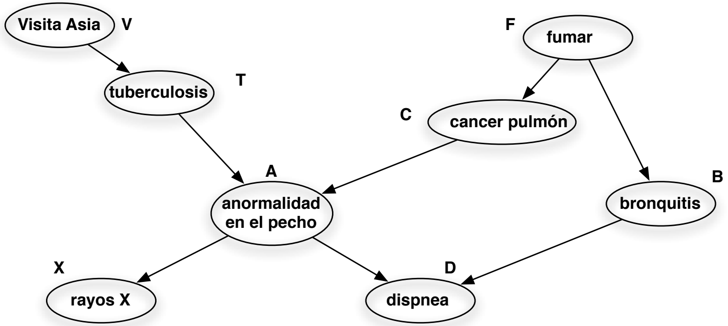
Algoritmo de Inferencia en RB: eliminación de variables

- el algoritmo opera sobre factores o tablas
- factor f sobre $Y_1, \ldots, Y_n$: función $f(Y_1, \ldots, Y_n) \in \mathbb{R}$
  ejemplo:
  - una DC $P(H, T, A, R, L)$ es factor sobre H, T, A, R, L
  - una TPC $P(A \mid H, T)$ es factor sobre H, T
- factores pueden sumarse y multiplicarse
  - $f'(C, L, R) = f_1(C) \times f_2(L, C) \times f_3(R, C)$
    equivalente a $P(C, L, R) = P(C)P(L \mid C)P(R \mid C)$
  - $f''(L, R) = \sum_C f'(C, L, R))$
    equivalente a $P(l, r) = \sum_c P(c, l, r)$

Algoritmo de eliminación de variables

procedure eliminar-variables(
  Grafo sobre $X_1, \ldots, X_n$, // estructura de la RB
  $P(X_i \mid padres(X_i))$, // TPC de nodos RB
  $Y_1, \ldots, Y_k$ // variables de query
)
  Sea $\mathcal{F}$ el conjunto de factores $\{P(X_i \mid padres(X_i)) : i = 1 \ldots n\}$
  Sea $\{Z_1, \ldots, Z_m\} = \mathbf{X} - \mathbf{Y}$
  para $i = 1, \ldots, m$
    Extraiga de $\mathcal{F}$ todo factor $f_1, \ldots, f_r$ que mencione a Z_i
    Sea $f' = \prod_j f_j$
    sea $f'' = \sum_{Z_i} f'$
    Inserte f'' en $\mathcal{F}$
  $retorne(\prod_{f \in \mathcal{F}} f)$

Algoritmo: un ejemplo



Algoritmo: un ejemplo (2)

calcular

$$\boxed{\begin{aligned} P(D) = \textstyle\sum_{A,B,C,T,F,X,V} & P(V)P(T \mid V)P(F)P(C \mid F) \\ & P(B \mid F)P(A \mid P,T)P(X \mid A) \\ & P(D \mid A,B) \end{aligned}}$$

- eliminar V: $f_V(T) = \sum_V P(T \mid V)P(V)$
- eliminar F: $f_F(B,C) = \sum_F P(B \mid F)P(C \mid F)P(F)$
- eliminar X: $\sum_X P(X \mid A)$
- calcular probabilidad de C: $f_B(C) = \sum_B f_F(B,C)$
- eliminar T: $f_T(A,C) = \sum_T P(A \mid T,C)f_V(T)f_B(C)$
- eliminar C: $f_P(A,B) = \sum_C f_F(B,C) \times f_T(A,C)$
- eliminar A,B: $\sum_{A,B} P(D \mid A,B)f_P(A,B)$