Razonamiento Probabilístico

Dr. Jesús Antonio González Bernal

Razonamiento Probabilístico

- Explicamos como construir modelos de redes para razonar bajo incertidumbre de acuerdo a las leyes de la teoría de la probabilidad. R&N, pg. 492.

Redes Bayesianas

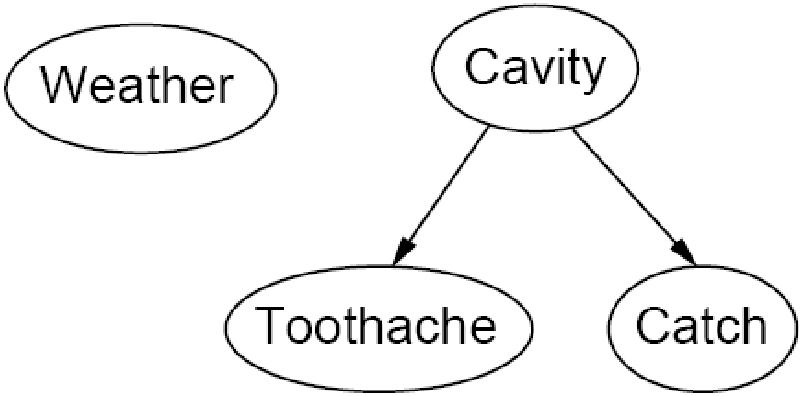
- Son una forma gráfica, simple para representar afirmaciones de independencia y por tanto, una forma compacta de especificar distribuciones conjuntas completas
  - i.e. *P(Tiempo*, *Caries*)
    - Representación con tabla de probabilidades de 4X2
      - *Tiempo*: Soleado, Lluvioso, Nublado, Nevado
      - *Caries*: Cierto, Falso
    - Distribución de probabilidad conjunta de *Tiempo* y *Caries*

Redes Bayesianas

- Sintaxis
  - Un conjunto de nodos, uno por variable
  - Un grafo acíclico dirigido (liga ≈ "influencia directamente")
  - Distribución de probabilidad para cada nodo dados sus padres:
    - $P(X_i \mid Parents(X_i))$
- Las distribuciones condicionales se representan con una tabla de probabilidad condicional (CPT) dada la distribución sobre X_i para cada combinación de valores de los padres

Ejemplo

- La topología de la red codifica las afirmaciones de independencia condicional

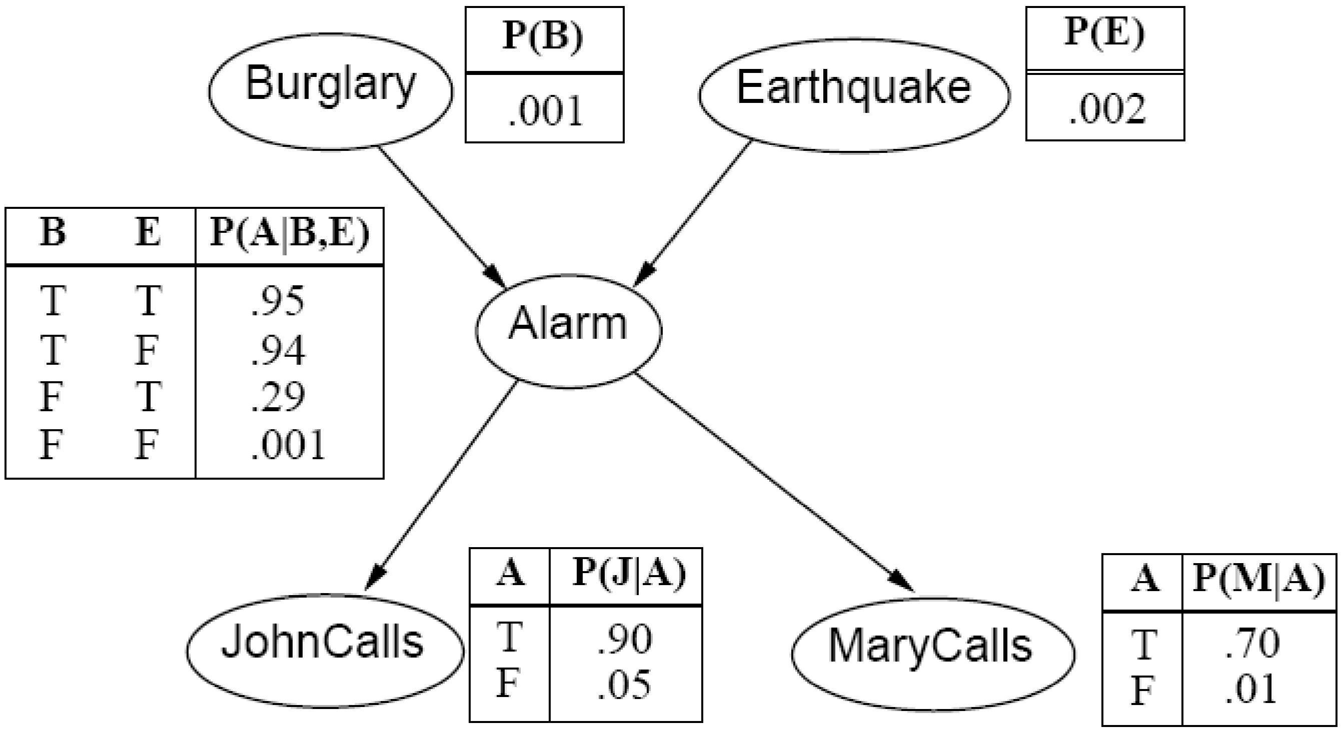


- *Weather* es independiente de las demás variables
- *Toothache* y *Catch* son condicionalmente independientes dada *Cavity*

Ejemplo

- Estoy en el trabajo, el vecino John llama para decirme que mi alarma esta sonando, pero la vecina Mary no me llama. Algunas veces se enciende con temblores ligeros. ¿Habrá un ladrón en casa?
- Variables: *Burglar*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Topología de la red refleja conocimiento "causal":
  - Un ladrón puede encender la alarma
  - Un temblor puede encender la alarma
  - La alarma puede causar que Mary llame
  - La alarma puede causar que John llame

Ejemplo



P(B)
.001

P(E)
.002

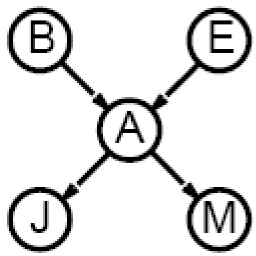
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactes

- Una CPT para X_i booleana con k padres booleanos tiene 2^k renglones para las combinaciones de valores de los padres
- En cada renglón tenemos un número p para X_i = true
  - El número para X_i = false es solo $1 - p$
- Si cada variable tiene no más de k padres, la red completa tendrá $O(n \cdot 2^k)$ números
- Crece linealmente con n, vs $O(2^n)$ para la distribución conjunta completa
- En la red anterior (burglary): $1 + 1 + 4 + 2 + 2 = 10$
  - Versus $2^5 = 32$

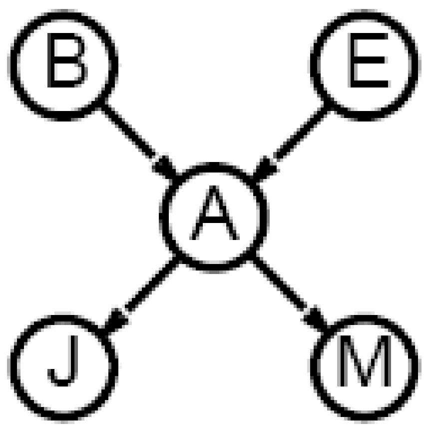


Semántica Global

- La semántica global define la distribución conjunta completa como el producto de las distribuciones condicionales locales:

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

- Por ejemplo:



$$\begin{aligned}
&P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) \\
&= P(j \mid a)P(m \mid a)P(a \mid \neg b, \neg e)P(\neg b)P(\neg e) \\
&= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998) \\
&\approx 0.00063
\end{aligned}$$

Construcción de Redes Bayesianas

- Necesitamos un método para que una serie de afirmaciones probadas localmente de independencia condicional, garanticen la semántica global requerida
  1. Elegir un orden de las variables $X_1, \ldots, X_n$
  2. For $i = 1$ to n

     Añadir X_i a la red

     Seleccionar padres de $X_1, \ldots, X_{i-1}$ tal que

     $\mathrm{P}(X_i \mid Parents(X_i)) = \mathrm{P}(X_i \mid X_1, \ldots, X_{i-1})$
- Esta elección de los padres garantiza la semántica global

$$
\begin{aligned}
P(X_1,\ldots,X_n) &= \prod_{i=1}^{n} P(X_i \mid X_1,\ldots,X_{i-1}) \quad \text{(regla de la cadena)} \\
&= \prod_{i=1}^{n} P(X_i \mid Parents(X_i)) \quad \text{(por construcción)}
\end{aligned}
$$

Construcción de Redes Bayesianas

A partir de la ecuación 14.1:

$$P(X_1 \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

Implica relaciones de independencia condicional que guían en la construcción de la topología de la red. Reescribimos la distribución conjunta en términos de una probabilidad condicional con una conjunción más pequeña. Usando la regla del producto

$$P(x_1 \ldots, x_n) = P(x_n \mid x_{n-1} \ldots, x_1) P(x_{n-1} \ldots, x_1)$$

Se repite el proceso, reduciendo cada probabilidad con conjunciones a una probabilidad con conjunciones a una probabilidad condicional con una conjunción más pequeña.

$$P(x_1 \ldots, x_n) = P(x_n \mid x_{n-1} \ldots, x_1) P(x_{n-1} \mid x_{n-2} \ldots, x_1) \ldots P(x_2 \mid x_1) P(x_1) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$

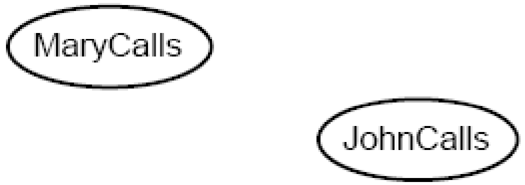
$$\begin{aligned} P(X_1, \ldots, X_n) &= \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \quad \text{(regla de la cadena)} \\ &= \prod_{i=1}^{n} P(X_i \mid Parents(X_i)) \quad \text{(por construcción)} \end{aligned}$$

Orden Correcto de las Variables

- Agregar primero las "*causas raíces*"
- Luego, las "*variables que influyen*"
- Continuar con este proceso hasta llegar a las "*hojas*", que no tienen influencia causal directa sobre las demás variables

Ejemplo --- Red Incorrecta

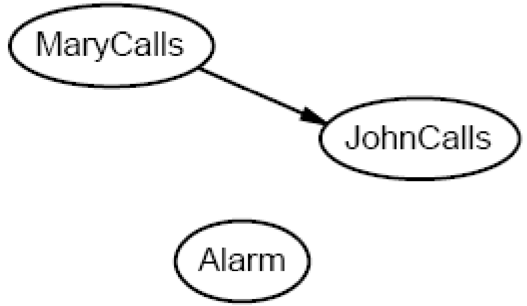
- Suponemos que elegimos el orden M, J, A, B, E



- $P(J \mid M) = P(J)$?

Ejemplo --- Red Incorrecta

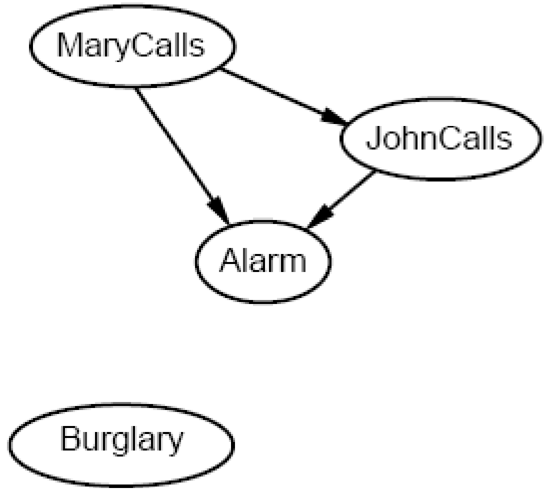
- Suponemos que elegimos el orden *M*, *J*, *A*, *B*, *E*



- P(J | M) = P(J)? No
- P(A | J,M) = P(A | J)? P(A | J,M)=P(A)?

Ejemplo --- Red Incorrecta

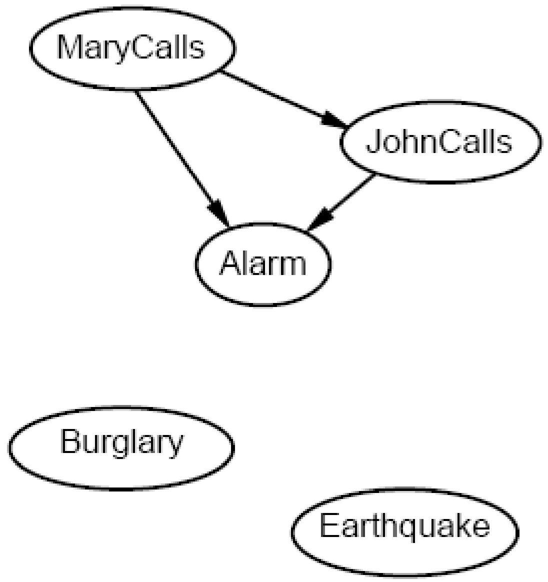
- Suponemos que elegimos el orden *M*, *J*, *A*, *B*, *E*



- P(J | M) = P(J)? No
- P(A | J,M) = P(A | J)? P(A | J,M)=P(A)? No
- P(B | A, J, M) = P(B | A)?
- P(B | A, J, M) = P(B)?

Ejemplo --- Red Incorrecta

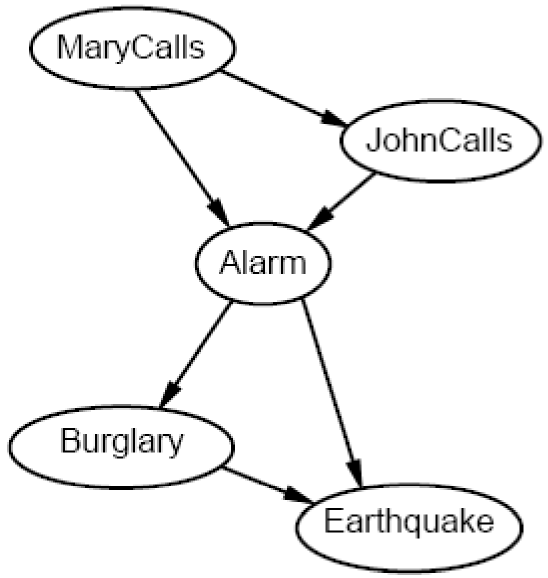
- Suponemos que elegimos el orden *M*, *J*, *A*, *B*, *E*



- P(J | M) = P(J)? No
- P(A | J,M) = P(A | J)? P(A | J,M)=P(A)? No
- P(B | A, J, M) = P(B | A)? Sí
- P(B | A, J, M) = P(B)? No
- P(E | B, A, J, M) = P(E | A)?
- P(E | B, A, J, M) = P(E | A, B)?

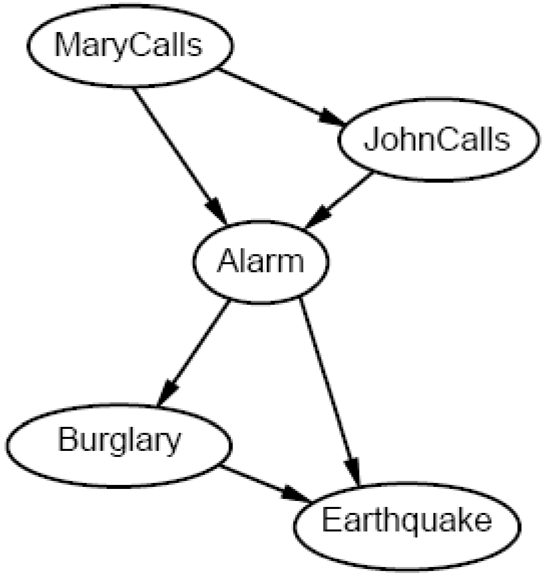
Ejemplo --- Red Incorrecta

- Suponemos que elegimos el orden *M*, *J*, *A*, *B*, *E*



- P(J | M) = P(J)? No
- P(A | J,M) = P(A | J)? P(A | J,M)=P(A)? No
- P(B | A, J, M) = P(B | A)? Sí
- P(B | A, J, M) = P(B)? No
- P(E | B, A, J, M) = P(E | A)? No
- P(E | B, A, J, M) = P(E | A, B)? Sí

Ejemplo --- Red Incorrecta



- Decidir sobre la independencia condicional es difícil en direcciones no causales
  - Difíciles de entender para nosotros
- Estimar probabilidades condicionales es difícil en direcciones no causales
- Red menos compacta: 1 + 2 + 4 + 2 + 4 = 13 números requeridos

Tareas de Inferencia

- Queries simples: calcular la marginal posterior P(Xi|E=e)
  - i.e., P(NoGas|Bomba=vacía, Luces=Encendidas, Enciende=falso)
- Queries conjuntos: P(Xi, Xj|E=e) = P(Xi|E=e)P(Xj|Xi, E=e)
- Decisiones óptimas: Redes de decisión incluyen información de utilidad; se requiere inferencia probabilística para P(salida|acción, evidencia)
- Valor de información: ¿Cuál evidencia buscar en el siguiente paso?
- Análisis de sensibilidad: ¿Cuáles valores de probabilidad son más críticos?
- Explicación: ¿Porqué necesito un nuevo motor de ignición?

Variables Irrelevantes

- Considerando el query *P*(JohnCalls | *Burglary*=*true*)

$$P(J \mid b) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(J \mid a) \sum_m P(m \mid a)$$

- La suma sobre *m* es idéntica a 1; *M* es irrelevante al query

- Teorema 1: *Y* es irrelevante al menos que $Y \in Ancestros(\{X\} \cup E)$
- Aquí, $X = JohnCalls$, $E = \{Burglary\}$, y $Ancestros(\{X\} \cup E) = \{Alarm, Earthquake\}$, entonces *MaryCalls* es irrelevante