Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Tema 7: Redes Bayesianas

- Repaso de probabilidad
  - Distribuciones de probabilidad conjunta
  - Teorema de Bayes y regla de la cadena
- Redes Bayesianas
  - Definición
  - Inferencia
    - Exacta
    - Aproximada

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Breve introducción a la probabilidad

Introducción (I)

- Teoría de la probabilidad
  - Dos aproximaciones, frecuencial y bayesiana
- Aproximación frecuencial
  - La probabilidad P de un evento a
    P(a) se define por la frecuencia de a basada en las observaciones pasadas
  - 60% de los nacimientos en España son niñas
    - a = 'Elegir al azar a un bebe y que sea niña'
    - P(a) = 0.6
  - Utilizamos el pasado para predecir el presente

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Breve introducción a la probabilidad

Introducción (II)

- Aproximación Bayesiana
  - Razonar sobre creencias en condiciones de incertidumbre
  - Deseamos conocer la probabilidad de que una nueva arquitectura de ordenador funcione correctamente
    - No hay instancias previas
  - a ='gana el CD Guadalajara la liga del 2022'
    - ¿P(a)?
    - P(a|Diego) = 0,7
    - P(a|Conocimiento previo) : Medida de conocimiento, si este conocimiento previo permanece constante podemos escribir P(a)
    - ¿Consistencia interna?

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Breve introducción a la probabilidad

Axiomas de probabilidad

Axiomas de la probabilidad

I. P(a) debe ser un nº entre [0,1]

II. Si a es un evento cierto, entonces P(a)=1

III. Si a y b son mutuamente exclusivos entonces P(a v b) = P(a) + P(b)

De esta manera…

- P(a + ¬a) = 1 (por el 2º axioma)

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Breve introducción a la probabilidad

Variables y distribuciones de probabilidad

A = 'Ganador de la liga en el 2022'
A = {a1,a2,a3,….}

- ¿P(a1 + a2) = 1?
  - No es exhaustivo
- P(a1 + a2 + a3 + …) = 1
- Probabilidad total

$$\sum_{i=1}^{n} P(a_i) = 1$$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Breve introducción a la probabilidad

Distribuciones conjuntas y marginalización

- A: 'funciona el monitor' = {m1,m2,m3}
  B: 'funciona la tarjeta de video' = {v1,v2}
  ¿P(A,B)?
- Distribución conjunta
  P(A,B) = {P(m1,v1), P(m1,v2), P(m2,v1), P(m2,v2), P(m3,v1), P(m3,v2)}
- Marginalización

$$P(a) = \sum_{i} P(a, b_i)$$

Breve introducción a la probabilidad

Probabilidad condicionada (I)

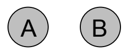
- El contexto K: P(A) = P(A|K)
- P(A|B) = P(A|B,K)
- Sucesos independientes
  P(A|B) = P(A)
- Sucesos Condicionalmente Independientes
  P(A|B,C) = P(A|C)   A y B son C.I. dado C
- Sucesos dependientes

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

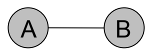
Breve introducción a la probabilidad

Probabilidad condicionada (II)

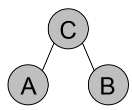
- **Pepe y Juan lanzan la misma moneda, primero lanza pepe**
  a : 'Pepe obtiene cara'
  b : 'Juan obtiene cara'
  P(A|B) = P(A)



- **Igual que antes pero la moneda tiene cierta tendencia a sacar cara (no sabemos cual)**
  P(B|A) > P(A)



- 'A' y 'B' son dependientes con una variable C: 'la moneda tiene tendencia a sacar cara'. Aunque 'A' y 'B' no son independientes si lo son respecto a 'C'
  P(A|C) = P(A|B,C)



- ¿Faltan dependencias en el grafo anterior?

Breve introducción a la probabilidad
Regla de Bayes

- Sabemos que:
  $P(A|B)\,P(B) = P(A,B)$
  $P(B|A)\,P(A) = P(B,A) = P(A,B)$
- Regla de Bayes

$$P(A\,|\,B) = \frac{P(B\,|\,A)P(A)}{P(B)} = \alpha \cdot P(B\,|\,A)P(A)$$

- Constante de normalización P(*B*)

$$P(B) = \sum_i P(B\,|\,A_i)P(A_i)$$

- Regla de la cadena

$$P(A,B) = P(A)P(B\,|\,A)$$
$$P(A,B,C) = P(A)P(B\,|\,A)P(C\,|\,B,A)$$

Redes Bayesianas
Definición

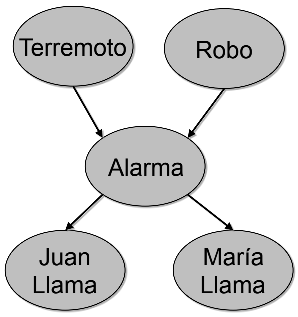
- Una red bayesiana es:
  - Un grafo acíclico dirigido para representar dependencias entre variables y mostrar una descripción escueta de cualquier distribución de probabilidad conjunta completa
- Esta formada por
  - Un conjunto de variables aleatorias que forman los nodos de la red. Cada nodo X tendrá adjunta una distribución P(X|Padres(X))
  - Un conjunto de enlaces que determinan la influencia (dependencia) entre nodos. Si X se conecta con Y se dice que X influencia a Y
- Su finalidad principal es calcular la distribución conjunta de las variables nodo

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas

Semántica (I)

- Dada la siguiente red bayesiana, ¿qué distribución representa?



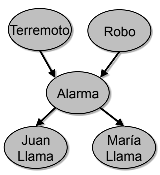
Sistemas Inteligentes

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas

Semántica (II)

- P(T,R,A,J,M) =



P(T) · P(R) ·
P(A|T,R) ·
P(J|A) · P(M|A)

- ¿ |P(T,R,A,J,M)| sin independencia condicional?
  - $2^5 = 32$
- ¿Y con independencia condicional?
  - $2+2+2^3+2^2+2^2 = 20$

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas
Inferencia

- ¿Para que queremos la distribución conjunta?
  - A partir de la distribución conjunta podemos contestar cualquier pregunta relativa a la red…
- Varios tipos de inferencia en redes Bayesianas
  - Exacta (caso general)
  - Casos especiales (Kim&Pearl…)
  - Aproximada

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas
Inferencia Exacta (I)

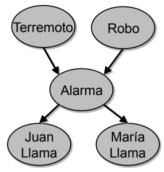
- Inferencia exacta general (funciona para todas la RR.BB.)
  - Regla de inferencia general
    (Donde B son las variables buscadas, C las conocidas y D las desconocidas)

$$P(B \mid C) = \alpha \cdot \sum_{D} P(B,D,C)$$

- Problema: Mucha complejidad

Redes Bayesianas
Inferencia Exacta (II)

- Ejemplo 1)
  - ¿Cuál es la probabilidad de que suene la alarma si llama María?



$$P(B \mid C) = \alpha \cdot \sum_{D} P(B,D,C)$$

  - P(R,T,A,J,M) =
    = P(R)·P(T)·P(A|R,T)·P(J|A)·P(M|A)

Redes Bayesianas
Inferencia Exacta (III)

- De esta manera obtenemos que

$$P(A \mid M) = \alpha \cdot \sum_{R} \sum_{T} \sum_{J} P(R,T,A,J,M) =$$

$$= \alpha \cdot \sum_{R} \sum_{T} \sum_{J} P(R) \cdot P(T) \cdot P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A) =$$

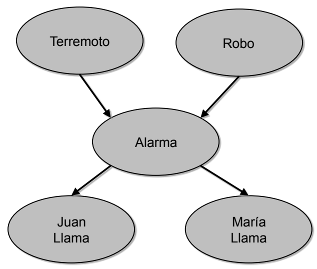
$$= \alpha \cdot P(M \mid A) \cdot \sum_{R} \left(P(R) \sum_{T} \left(P(T) \cdot P(A \mid R,T) \cdot \underbrace{\sum_{J} P(J \mid A)}_{1} \right) \right)$$

Redes Bayesianas
Inferencia Exacta (IV)

- Ejemplo 2) ¿P(R|J+,M+)? Si sabemos que:

P(T) = 0,001

P(R) = 0,002



P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,94
1	0	0,29
1	1	0,95

P(J|A) =

A	J
0	0,05
1	0,9

P(M|A)=

A	M
0	0,01
1	0,7

Redes Bayesianas
Inferencia Exacta (V)

- Entonces:

$$P(R \mid J,M) = \alpha \sum_T \sum_A P(R,T,A,J,M) =$$

$$\alpha \sum_T \sum_A P(R) \cdot P(T) \cdot P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A) =$$

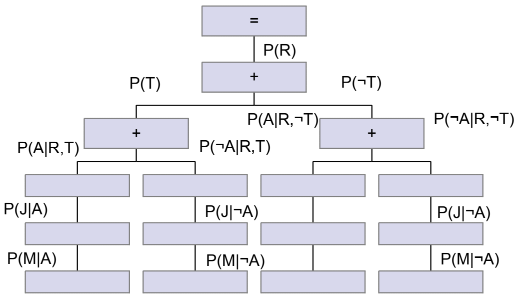
$$\alpha \cdot P(R) \cdot \sum_T \left(P(T) \cdot \sum_A \left(P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A) \right) \right)$$

Redes Bayesianas

Inferencia Exacta (VI)

- Para calcular descomponemos utilizando un árbol

$$\alpha \cdot P(R) \cdot \sum_T P(T) \cdot \sum_A P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A)$$



Redes Bayesianas

Inferencia Exacta (VII)

P(R|J,M)			**P(¬R|J,M)**	
1	P(A|R,T)*P(J|A)*P(M|A)	0,5985	*1* P(A|¬R,T)*P(J|A)*P(M|A)	0,1827
	P(¬A|R,T)*P(J,¬A)*P(M|¬A)	0,000025	P(¬A|¬R,T)*P(J,¬A)*P(M|¬A)	0,000355
	P(T)* SUM 1	0,00059853	P(T)* SUM 1	0,00018306
2	P(A|R,¬T)*P(J|A)*P(M|A)	0,5922	*2* P(A|¬R,¬T)*P(J|A)*P(M|A)	0,00063
	P(¬A|R,¬T)*P(J,¬A)*P(M|¬A)	0,00003	P(¬A|¬R,¬T)*P(J,¬A)*P(M|¬A)	0,0004995
	P(¬T) * SUM 2	0,59163777	P(¬T) * SUM 2	0,00112856
	TOTAL R+	**0,00118447**	TOTAL ¬R	**0,00130899**
	Cómo alpha*(R,¬R) = 1			
P(R|J,M)=	**R**	**0,47503178**		
	¬R	**0,52496822**		

Redes Bayesianas
Inferencia Exacta (VIII)

- Ejemplo 3) ¿P(J|R)?

$$P(J \mid R) = \sum_T \sum_A \sum_M P(R) \cdot P(T) \cdot P(A \mid R,T) \cdot P(J \mid A) \cdot P(M \mid A) =$$

$$= P(R) \cdot \sum_T (P(T) \cdot \sum_A (P(A) \mid R,T) \cdot P(J \mid A) \cdot \sum_M P(M \mid A)))$$

Redes Bayesianas
Inferencia exacta en poliárboles

- Existen algoritmos más eficientes para tipos específicos de redes
  - Modelo de Kim y Pearl
    - Método de inferencia para redes bayesianas.
    - Solo aplicable a un poliárbol.
      - No existe más de un camino entre cada pareja de nodos
    - Se basa en el paso de dos tipos de mensajes entre nodos
      - Para actualizar la credibilidad
      - Para introducir nueva evidencia

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas
Inferencia aproximada (I)

- Sobre la inferencia exacta
  - Redes con conexión multiple son intratables utilizando inferencia exacta
- Inferencia utilizando algoritmos de muestreo aleatorio (Monte Carlo)
  - Existen varios algoritmos
    - Muestreo directo
    - Muestreo por rechazo
    - MCCM (monte Carlo para cadenas de Markov)…
  - Veremos el de muestreo directo

Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas
Inferencia aproximada (II)

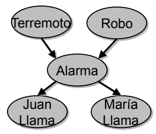
- Muestreo directo
  - ALGORITMO Muestreo Directo
    - Desde k=1 hasta suficientesMuestras
      - s[k][ ] = <vector de sucesos con n elementos>
      - Desde i=1 hasta n hacer
        - $S[k]_i$ = Obtener una muestra aleatoria de P(X_i| Padres(X_i))
      - fDesde
    - fDesde
    - Devolver x
  - Para responder cualquier pregunta de la red
    - Contar apariciones en s[ ] de las evidencias
    - Dividir por suficientesMuestras

Redes Bayesianas

Inferencia aproximada (III)

- Ejemplo) Muestrear la red

P(T) = 0,001;P(R) = 0,002



P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,29
1	0	0,94
1	1	0,95

P(J|A) =

A	J
0	0,05
1	0,9

P(M|A)=

A	J
0	0,01
1	0,7

1. Muestreo a partir de P(Terremoto) = <0,001 0,999>. **supongamos (s.)** que devuelve *falso*
2. Muestreo(P(Robo)) s. devuelve *falso*
3. Muestreo(P(Alarma|<Robo=falso, Terremoto=falso>)) s. devuelve *cierto*
4. Muestreo(P(Juan|<A=cierto>)) s. Devuelve *cierto*
5. Muestreo(P(Maria|<A=cierto>)) s. Devuelve *falso*
6. *X[k] = <falso,falso,cierto,cierto,falso>*

Redes Bayesianas

Inferencia aproximada (IV)

- Ejemplo) ¿P(R|J,M)?

P(T) = 0,001;P(R) = 0,002

P(A|T,R)=

T	R	A
0	0	0,001
0	1	0,29
1	0	0,94
1	1	0,95

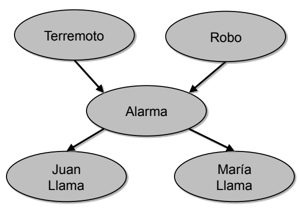
P(J|A) =

A	J
0	0,05
1	0,9

P(M|A)=

A	J
0	0,01
1	0,7

1. *Para obtener P(R|J,M)*
2. *C = Contar X[k] que cumpla este patrón X=<?,cierto,?,cierto,cierto>*
3. *Devolver C/numeroDeMuestras*



Bibliografía

Base

- Inteligencia Artificial, Un enfoque moderno. Stuart Russell y Peter Norvig, ed Pearson [pág 561]

Repaso probabilidad e introducción a las RR.BB.

- Probability Theory and Bayesian Belief Bayesian Networks. Norman Fenton. http://www.dcs.qmul.ac.uk/~norman/BBNs/BBNs.htm