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

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

Objetivos de la sesión

1. Entender que es y que representa una red Bayesiana
2. Conocer como se puede utilizar una red Bayesiana para extraer conocimiento
3. Entender las bases de la inferencia exacta en redes Bayesianas
4. Entender las bases de la inferencia aproximada en redes Bayesianas
5. Conocer las bases de los métodos de muestreo aleatorio

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

Índice

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

Probabilidad como media de incertidumbre (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

Probabilidad como media de incertidumbre (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 Alcoyano la liga del 2022’
  - ¿P(a)?
  - P(a|Fidel) = 0,7
  - P(a|Conocimiento previo) : Medida de conocimiento, si este conocimiento previo permanece constante podemos escribir P(a)
  - ¿Consistencia interna?

5

Probabilidad como media de incertidumbre (III)

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)

6

Probabilidad como media de incertidumbre (IV)

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

7

Probabilidad como media de incertidumbre (V)

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

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

8

Probabilidad como media de incertidumbre (VI)

Probabilidad condicionada

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

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

9

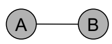
Probabilidad como media de incertidumbre (VII)

Probabilidad condicionada

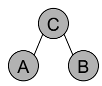
Pepe y Juan lanzan la misma moneda, primero lanza pepe

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)



10

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

Teorema 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

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

Constante de normalización P(*B*)

$$P(B) = \sum_i P(B \mid A_i)P(A_i)$$

Regla de la cadena

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

11

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 (I)

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

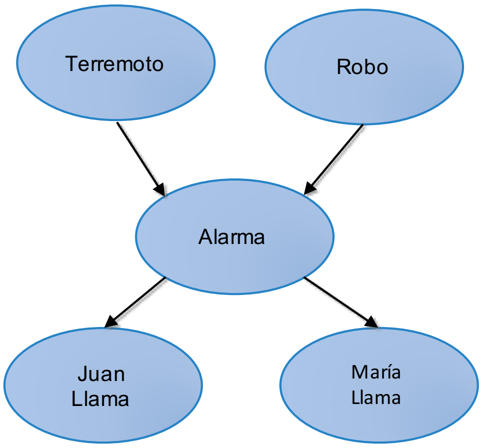
Su finalidad principal es calcular la distribución conjunta de las variables nodo

12

Redes Bayesianas (II) Semántica

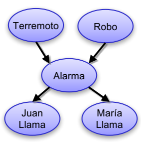
Dada la siguiente red bayesiana,

¿Qué distribución representa



Redes Bayesianas (III)
Semántica

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



$$P(T) \cdot P(R) \cdot P(A|T,R) \cdot P(J|A) \cdot P(M|A)$$

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

14

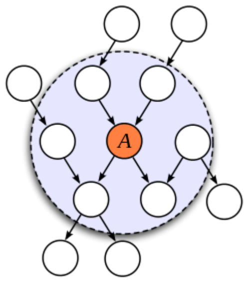
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

Redes Bayesianas (IV)

Semántica

- Cobertura de Markov
  - Un nodo A es condicionalmente independiente de todos los nodos de la red dados:
    - Sus padres
    - Sus hijos
    - Los padres de sus hijos



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

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

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

17

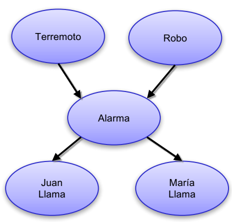
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial

Universitat d'Alacant
Universidad de Alicante

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)

18

Inferencia (III)

Ejemplo 1

De esta manera tenemos que:

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

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

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

19

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

Inferencia (IV)

Ejemplo 2

¿P(R|J+,M+)? Si sabemos que:

P(T) = 0,001

P(R) = 0,002

P(J|A) =

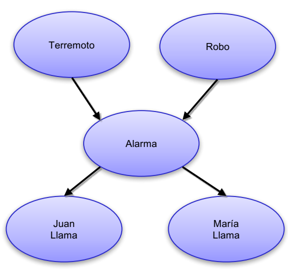
A	J
0	0,05
1	0,9

P(A|T,R)=

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

P(M|A)=

A	M
0	0,01
1	0,7



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

Inferencia (V)

Ejemplo 2

De esta manera tenemos que:

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

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

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

21

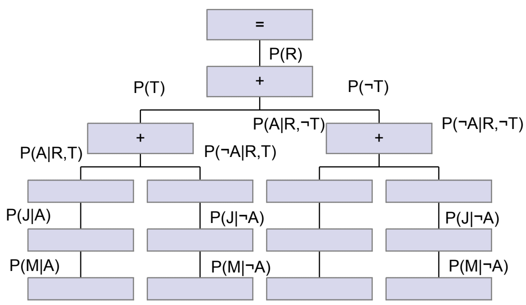
Dpnt. de Ciència de la Computació i Intel·ligència Artificial
Dpto. de Ciencia de la Computación e Inteligencia Artificial
Universitat d'Alacant
Universidad de Alicante

Inferencia (VI)

Ejemplo 2

Para calcular descomponemos utilizando un árbol:

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

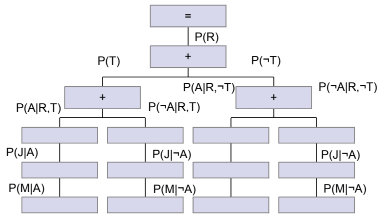


22

Inferencia (VII)

Ejemplo 2

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**

23

Inferencia (VIII)

Ejercicio 1

¿P(J|R)?

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

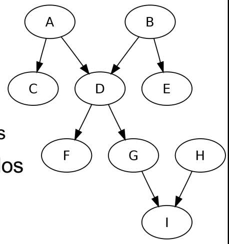
$$= P(R) \cdot \sum_T (P(T) \cdot \sum_A (P(A) \mid R,T) \cdot P(J \mid A) \cdot \sum_M P(M \mid A)))$$

24

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
  - Se puede calcular en tiempo lineal



25

Inferencia aproximada (I)

Sobre la inferencia exacta

- Redes con conexión múltiple son intratables utilizando inferencia exacta
- Complejidad NP-hard en el caso general

Inferencia utilizando algoritmos de muestreo aleatorio (Monte Carlo)

- Existen varios algoritmos
  - Muestreo directo
  - Muestreo por rechazo
  - Gibbs Sampling

26

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

Inferencia aproximada (II)

Muestreo directo

- rb: red Bayesiana
- ALGORITMO **Muestreo_Directo**(*rb*) retorna un evento extraido de *rb*
  - X = <vector de sucesos con n elementos>
  - Para cada variable X_i en $X_1,\ldots,X_n$ hacer
    - X_i = Obtener una muestra aleatoria de $P(X_i|Padres(X_i))$
  - Devolver X

Para responder cualquier pregunta de la red
- Obtener un vector de eventos X[]
- Contar apariciones en X[] de las evidencias
- Dividir por suficientesMuestras

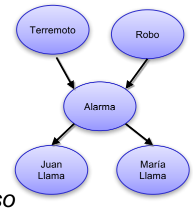
27

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

Inferencia aproximada (III)

Ejemplo de muestreo de una red mediante muestreo directo



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 = <falso,falso,cierto,cierto,falso>*

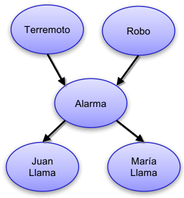
28

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

Inferencia aproximada (IV)

Ejemplo de uso

¿P(R|J,M)?



P(T) = 0,001;

P(R) = 0,00

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[ ] que cumpla este patrón X=<?,cierto,?,cierto,cierto>*
3. *Devolver C/numeroDeMuestras*

29

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

Inferencia aproximada (V)

¿Problema del muestreo directo?

Otros tipos de muestreo aleatorio

- Muestreo por rechazo
- Gibbs Sampling

30

Inferencia aproximada (VI)
Muestreo por rechazo

- ALGORITMO **Muestreo_por_Rechazo**(*B,c,rb*) retorna estimación P(B|c)
- Para j = 1 hasta num_muestras hacer
  - **x** = Muestro_Directo(rb)
  - Si **x** es consistente con la evidencia c:
  - N[y] = N[y] +1, donde y es el valor de B en **x**
- Devolver Normalizar(N)

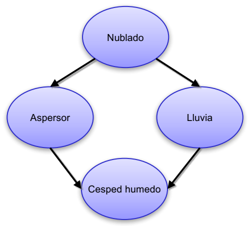
Entradas:
- B: variable buscada (pregunta)
- c: valores observados de las variables conocidas C
- rb: red bayesiana

Variables locales:
- N: vector de recuento para cada valor de B, inicialmente 0

31

Inferencia aproximada (VII)
Muestreo por rechazo

Ejemplo:



Queremos estimar P(Lluvia|Aspersor=cierto)
Extraemos 100 muestras, de las cuales 73 tienen el aspersor apagado
Nos quedamos con las 27 que coinciden con la evidencia
De las 27:
- En 8 Lluvia = cierto
- En 19 Lluvia es falso

P(Lluvia|Aspersor=cierto) = Normalizar(<8,19>) = <0.296,0.704

32

Inferencia aproximada (VIII)

Muestreo por Gibbs (MCMC)

- ALGORITMO **Muestreo_por_Gibbs**(*B,c,rb,N*) retorna estimación P(B|c)
- Inicializar **x** con valor aleatorios para las variables en Z
- Para j = 1 hasta num_muestras hacer
  - Para cada Z_i en Z hacer
    - **x**[Z_i] = muestrear P(Z_i|mb(Z_i))
    - N[x] = N[x] + 1 donde x es el valor de B en **x**
- Devolver Normalizar(N)

Entradas:
- B: variable buscada (pregunta)
- c: valores observados de las variables conocidas C
- rb: red bayesiana

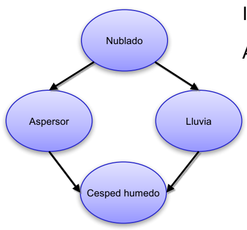
Variables locales:
- N: vector de recuento para cada variable B (inicialmente vale 0)
- Z, las variables sin evidencia en rb
- x: el estado de la red, copiado inicialmente de c

Funciones
- mb() retorna la cobertura de Markov un nodo

33

Inferencia aproximada (IX)

Ejemplo



Queremos estimar P(Lluvia|Aspersor=cierto,Cesped=cierto)

Las variables conocidas: Aspersor y Cesped, se fijan a su valor.

Las variables desconocidas: Nublado y Lluvia se establecen aleatoriamente.

Imaginemos que el estado inicial es: [cierto,cierto,falso,cierto]

Ahora las variables sin evidencia se muestrean repetidamente en orden arbitrario. Por ejemplo

- Se muestrea Nublado, dado su recubrimiento de Markov. Por tanto extraemos de P(Nublado|Aspersor=cierto,Lluvia=Falso) Asumamos que el resultado es falso y por tanto el nuevo estado es [falso,cierto,falso,cierto]
- Se muestrea Lluvia dado su recubrimiento: P(Lluvia|Nublado=falso, aspersor=cierto, cesped=cierto). Asumamos que el resultado es cierto. El nuevo estado es [falso,cierto,cierto,cierto]

Todo estado visitado mediante este proceso es una muestra que contribuye a estimar la pregunta

Por ejemplo, si durante es te proceso se visitan 20 estados donde lluvia es cierto y 60 donde es falso P(L|A=cierto,C=cierto)=Normalizar(<20,60>)=<0.25,0.75>

34

Para saber más

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

35