Tema 5: Razonamiento con incertidumbre

- Probabilidad incondicional

+ Conceptos básicos

* $0 \leq P(a) \leq 1$; $P(a)=1$: true ; $P(a)=0$: false
* $P(a \vee b) = P(a) + P(b) + P(a \wedge b)$; si a y b son disjuntos: $P(a \vee b) = P(a) + P(b)$
* $P(\neg a) = 1 - P(a)$
* $\sum_{i=1}^{n} P(D=d_i) = 1$; La suma de las probabilidades de cada variable que forma D, es 1

+ Eventos atómicos

Proposiciones elementales que dan valor a todas las variables del conjunto.
Toda proposición es equivalente a la disyunción de un conjunto de eventos atómicos

Ej: patata = (patata ∧ queso) ∨ (patata ∧ ¬queso)

De forma general: $P(a) = \sum_{e_i \in e(a)} P(e_i)$ # P(a) es el sumatorio de las probabilidades de los eventos atómicos de a

Estos eventos se organizan mediante una tabla de distribución conjunta, en la que se coloca como fila la consecuencia y como columnas las variables que las preceden.

Ej: Estudiamos los casos para p, sabiendo que puede ser consecuencia de q, r o ambas

	q	q	¬q	¬q
	r	¬r	r	¬r
p				
¬p				

Estas podrían ser las probabilidades posibles para cada caso 2 filas x 4 colum (ya que "q y r" se estudian a la vez)

Ej uso de forma general: $P(q \vee p)$ o $P(p \vee q)$

$P(p \vee q) = P(p,q,r) + P(p,q,\neg r) + P(p,\neg q,\neg r) + P(p,\neg q,r) + P(\neg p,q,r) + P(\neg p,q,\neg r)$

Suma de todos los eventos que satisfacen $P(p \vee q)$

- Probabilidad condicional

+ Conceptos básicos

$$\boxed{P(a|b) = \frac{P(a \wedge b)}{P(b)}}$$

Probabilidad de que se de a sabiendo que se da b.

$P(a \wedge b)$ se puede escribir como $P(a,b)$

Hablamos siempre de creencia, no de evidencia conocida

+ Normalización

A la hora de calcular una probabilidad condicional, podemos calcularla con la formula anterior o calcular la probabilidad de que se cumpla y no; y normalizarla (hacer que ambas sumen 1)

Ej: $p(p|q)$

1) $p(p|q) = \frac{P(p \wedge q)}{P(q)}$ 2) $P(p|q) = \alpha \left[P(p,q,r), P(p,q,\neg r) \right]$

2) Habría que estudiar tanto el caso de p, como el de ¬p; y sumar los casos positivos y negativos por separado para después normalizarlo

+ Independencia probabilística

Dos variables son independientes si conocer el valor de ella no afecta el grado de creencia de la otra.

Ej: $P(X|Y) = P(X)$; $P(Y|X) = P(Y)$; $P(X,Y) = P(X) \cdot P(Y)$

Ej₁: $P(q|p,r) = P(q|p)$ # q es independiente de r (independencia condicional)

Ej₂: Sabiendo que q y r dependen de p, podemos decir que:

$$P(p,q,r) = P(p) \cdot P(q|p) \cdot P(r|p)$$

<u>De forma general</u>: $P(\text{Causa}, E_i \ldots, E_n) = P(\text{Causa}) \prod_i P(E_i|\text{Causa})$

La probabilidad de una causa y una serie de efectos (q ocurre debido a p) es el producto de la probabilidad de la causa y la probabilidad de cada evento dada la causa. Es decir, producto de la probabilidad con sus precesores, si no tiene, sería la probabilidad de dicha variable

IA

Tema 5: Razonamiento con incertidumbre

- Redes bayesianas

Una red bayesiana es un grafo acíclico dirigido

+ Características básicas

1º Cada nodo representa una variable aleatoria

2º Los arcos entre X e Y definen los padres $A \rightarrow B$ (B depende de A)

3º Cada nodo contiene una distribución probabilística con sus padres $P(X_i | padres(X_i))$

Con las redes bayesianas podemos ver claramente las independencias entre ciertas variables del "mundo"

La mejora de las redes bayesianas frente a la DCC es de: $n \cdot 2^k$ frente a 2^n (siendo n las variables y K la influencia max ~~media~~)

+ Construcción de redes bayesianas

Para facilitar la construcción de la red, comenzar por las proposiciones que no dependan de otra y desarrollar a partir de ahí

+ D-separa

Una vez dibujada la red bayesiana se puede deducir la independencia mediante d-separación. Para que dos variables X e Y estén d-separadas y por lo tanto sean condicionalmente independientes:

1- Pasa por un elemento que es parte del conjunto dado ($Z \in \mathbb{Z}$) con arcos en el mismo sentido o divergentes.

Ej1: Dada {Z} el camino entre X e Y es condicionalmente independiente (d-separa)

1) $X \leftarrow A \rightarrow Z \rightarrow B \leftarrow Y$ # Z d-separa independientemente del resto del camino

2) $X \rightarrow A \leftarrow Z \leftarrow B \leftarrow Y$

3) $X \rightarrow A \leftarrow Z \rightarrow B \leftarrow Y$

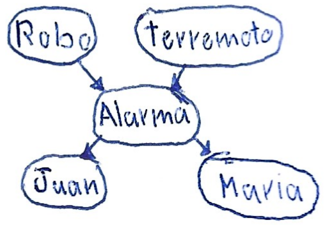
2- Pasa por un elemento que no es parte del conjunto dado ($Z \notin \mathbb{Z}$) con arcos convergentes y ningún descendiente de Z está en $\mathbb{Z}$

Ej2: Dada {C} el camino entre X e Y es condicionalmente independiente (d-separa)

$X \rightarrow C \leftarrow B \rightarrow Z \leftarrow Y$ # Aunque nos den C, lo que hace que d-separe es Z

+ Inferencia probabilística en una red bayesiana

Calcular la probabilidad dada una consulta (efecto) y evidencias.
Para entender mejor como desarrollarlo use el ejemplo de las diapos.



$P(Robo|juan, maria) = \langle P(r|j,m); P(\neg r|j,m)\rangle = 1)$

$1) = \alpha \langle \sum_{t}\sum_{A} P(r,t,A,j,m); \sum_{t}\sum_{A} P(\neg r, t, A, j, m)\rangle = 2)$

$2) = \alpha \langle \sum_{t}\sum_{A} P(r)\, P(t)\, P(A|r,t)\, P(j|A)\, P(m|A);$
$\sum_{t}\sum_{A} P(\neg r)\, P(t)\, P(A|\neg r,t)\, P(j|A)\, P(m|A)\rangle$

1) Estamos buscando la probabilidad de un robo si juan y maría han llamado, por ello, estudiaremos los casos separados para r y $\neg r$, después lo normali-
2) zaremos. Para estudiar $P(r|j,m)$, sabiendo que j y m son evidencias y por lo tanto siempre positivas; hay que hacer una ~~producto~~ suma del producto de $P(r)$, $P(A|r,t)$, $P(j|A)$ y $P(m|A)$; comprobando tanto para A, como T el caso positivo y negativo. Luego habría que hacer lo mismo con el caso de $P(\neg r|j,m)$

+ Algoritmo de eliminación de variables

Partimos del ejercicio anterior como ejemplo. Por la probabilidad a estudiar conocemos las probabilidades de las evidencias $P(r|j,m) \Rightarrow P(j|A)$ y $P(m|A)$

$P(m|A) = F_M(A)$ aquí se pone la variable que no conozcamos y forme parte de P()

Una vez tenemos claras las probabilidades de las evidencias, pasamos a las antecesoras, en este caso $P(A|r,t)$ o $f_A(A,R,T)$ # No conocemos ninguna variable
En este caso nos la da la red bayesiana.
Pasaríamos a las siguientes antecesoras hasta acabar

* Eliminación de una variable

Si queremos eliminar una variable de un modelo:

1º Buscar las probabilidades en las que interviene la variable a eliminar.
Ej: Eliminaremos A del ejemplo.
Las funciones en las que aparecen son $\Rightarrow P(m|a)$, $P(j|a)$ y $P(a|r,t)$ o $f_M(A)$, $f_J(A)$ y $f_A(A,R,T)$

2º Generamos la tabla sin la variable, para ello, multiplicamos los valores de la variable a eliminar en cada caso, teniendo en cuenta que, los valores que tendrán mas cambios serán los de la función de referencia, calculamos los casos positivos y negativos y los sumamos.

3º Repetimos el paso 2º hasta quedarnos con una función que solo tenga como variable desconocida la de consulta y la multiplicamos por la tabla de la función de la variable de consulta. Al normalizarla, nos dará la probabilidad que buscabamos al inicio

IA

Tema 5: Razonamiento con incertidumbre

+ Observaciones algoritmo de eliminación

Comenzamos con 1 factor por variable y vamos reduciendo:

1º Por cada variable que no sea de evidencia o consulta sustituimos los factores por el resultante de eliminar una variable (Partir de las evidencias hasta la consulta).

* Eliminar variable: buscar todos los factores dónde aparece, multiplicarlos y sumar los casos positivos y negativos (obteniendo el factor sin variable)

Si la variable es la de consulta solo multiplicamos, ya que queremos el porcentaje positivo y negativo

2º Cuando lleguemos al factor único que solo depende de la variable de consulta, normalizamos el factor y obtenemos el resultado de la inferencia probabilística.

* Operaciones
  - Multiplicar tablas: partiendo de $f_1(x,y)$ y $f_2(y,z)$, se define el producto $f_{12}(x,y,z)$, cuya entradas es la multiplicación de ambas
  - Agrupar tablas: si $f_{XY}(Y,Z)$ es un factor (multiplicar XY), su agrupamiento sería eliminar la Y, es decir, dejar el factor como $f_{\bar{Y}}(z)$, resultante de hacer la suma de las columnas para "y" y "¬y"

Toda variable que no sea antecesora en la red se elimina

+ Inferencia exacta en redes bayesianas

La inferencia exacta es inviable al menos que se haga mediante muestreo

* Muestreo o Sampling

Consiste en generar eventos en base a una distribución de probabilidad.

Ej muestreo simple: $P(A) = \langle \alpha, 1-\alpha \rangle$, generamos una muestra de números aleatorios entre $[0,1]$, si sale un valor $< \alpha$ a es positiva, sino es ¬a

De esta forma podemos hacernos una idea del modelo aproximado

* Muestreo a priori

Simplemente con cada variable de independiente a antecesores vamos generando muestras de forma aleatoria y rellenando el modelo. Importante, si una probabilidad padre $P(A) = a$, sale negativa (¬a), la probabilidad de los dependientes se harán con ¬a, por ejemplo, $P(B \mid \neg a, c)$

* Muestreo con rechazo

Con el modelo anterior, generar varios resultados y devolver la probabilidad de un evento dado los resultados.

Ej: Si generamos 100 muestras y de estas 15 son a y de esas 15 solo 2 son b, la $P(b \mid a) = \langle 2, 13 \rangle$, normalizamos después

+ Ponderación por verosimilitud

Consiste en aplicar un peso en base a la probabilidad del evento del que depende, si este es una variable de evidencia.

* Para realizar el algoritmo, hacemos el muestreo normal, pero cuando llegamos a una variable de evidencia calculamos W x P (e|predecesores), con esto obtenemos un resultado, además del peso de la creencia de este. Las evidencias tendrán el signo de la probabilidad que se busca.

IA

Tema 5: Redes Bayesianas

Tipos de problemas (conv. julio 2023)

+ Calcula la probabilidad dada unas premisas

$P(a, \neg b, c, e, f)$

1º Miramos en la red bayesiana si hay alguna variable con dependencia que no aparezca en la red bayesiana; en este caso sería d, que es padre, tanto de e, como de f, por lo que hay que calcular la probabilidad sumando, tanto el caso positivo, como el negativo.

$P(a, \neg b, c, e, f) = \overbrace{P(a, \neg b, c, d, e, f)}^{1)} + P(a, \neg b, c, \neg d, e, f)$

2º Desarrollamos las probabilidades, empezando por el producto de las variables sin padres y multiplicándolas por el producto de las probabilidades de las variables dependientes

1) $P(a, \neg b, c, d, e, f) = P(a) \cdot P(\neg b) \cdot P(c|a, \neg b) \cdot P(d|a, \neg b) \cdot P(e|d) \cdot P(f|d)$

+ Calcular la probabilidad por eliminación de variables

$P(B|\neg a, e)$

Conocemos las evidencias $\neg a$ y e; por lo tanto, sabemos $P(\neg a)$ y $P(e|d)$.

Como queremos llegar a B nos fijamos en sus hijos; en este caso son C y D

1º Organizamos las tablas iniciales con los valores sin dependencia e iniciales (evidencias)

#F_A no habría que ponerla porque es siempre $\neg a$ y no tiene padre; Fe siempre es e

B	$f_B(B)$
b	0'1
¬b	0'9

#tomamos los valores ¬a

C	$f_c(C, B)$
b c	0'7
b ¬c	0'3
¬b c	0'2
¬b ¬c	0'8

D	$f_d(D, B)$
b d	0'8
b ¬d	0'2
¬b d	0'7
¬b ¬d	0'3

E	$f_e(D)$
d	0'1
¬d	0'8

F	$f_f(D, F)$
d f	0'4
¬d f	0'9
d ¬f	0'6
¬d ¬f	0'1

2º Eliminamos variables

1) Eliminamos D

~~[illegible]~~ $f_{\times D}(B, F) = f_e(D) \times f_d(D, B) \times f_f(D, F)$

B	F	d	¬d	$f_{-D}(B, F)$
b	f	0'1 x 0'8 x 0'4	0'8 x 0'2 x 0'9	Suma de la fila
b	¬f	0'1 x 0'8 x 0'6	0'8 x 0'2 x 0'1	
¬b	f	0'1 x 0'7 x 0'4	0'8 x 0'3 x 0'9	
¬b	¬f	0'1 x 0'7 x 0'6	0'8 x 0'3 x 0'1	

3º Multiplicamos los factores restantes y normalizamos

+ <u>D-separa</u>

D-separa si:

*Dada un valor {c}, este aparece en ~~[illegible]~~ la red entre los valores a analizar como:

$[X \rightarrow \ldots \rightarrow c \rightarrow \ldots \rightarrow Y]$, $[" \leftarrow c \rightarrow "]$ o $[" \leftarrow c \rightarrow "]$

*Dada un valor {c}, no aparece entre los valores, pero hay otro tal que:

$[X \rightarrow \ldots \rightarrow Z \leftarrow \ldots \rightarrow Y]$