INTELIGENCIA ARTIFICIAL (INF371)

RAZONAMIENTO PROBABILÍSTICO CON REDES BAYESIANAS (PARTE 1)

Dr. Edwin Villanueva Talavera

Contenido

- Incerteza
- Probabilidades
- Inferencia con Distribución de Probabilidad Conjunta
- Independencia
- Regla de Bayes

Bibliografía:

Capitulo 13.1, 13.2, 13.3, 13.4 y 13.5 del libro:

Stuart Russell & Peter Norvig "Artificial Intelligence: A modern Approach", Prentice Hall, Third Edition, 2010

Incerteza

- Incerteza es inherente al mundo real. esta aparece por la observabilidad parcial del entorno, no determinismo del entorno o las acciones y falta de conocimiento
- Ejemplo:
  - Consideremos un taxi automático que tiene que llevar pasajeros al aeropuerto. El agente tiene que decidir la acción: A_t = salir al aeropuerto con t min de anticipación al vuelo
  - La incerteza puede aparecer por:
    - Observabilidad parcial (estado de carreteras, trafico, accidentes, etc.)
    - Sensores ruidosos (de trafico, del motor, gps, etc.)
    - Incerteza de las acciones (neumático agujereado, fallas mecánicas, etc.)
  - Agentes puramente lógicos no consiguen lidiar con incerteza:
    - Podrían acabar deduciendo acciones sin utilidad practica :

      "A_{1440} llegará a tiempo al aeropuerto"

Incerteza

- Existen diversos enfoques para lidiar con incerteza:
  - Lógica No Monotonica: se asume hipótesis a menos que haya evidencia de lo contrario. Problema para determinar cuales y cuantas hipótesis
  - Reglas con Factores de Incerteza: Problema al combinar reglas contradictorias
  - Lógica Difusa: aborda incerteza desde el punto de vista de grado de impreción o ambigüedad
  - Probabilidades: herramienta efectiva para abordar incertezas

- Probabilidades son un medio por el cual los agentes pueden modelar sus grados de creencia dadas las evidencias disponibles:

Ej. $P(\text{llegar a tiempo al aeropuerto} \mid A_{25}) = 0.1$

Probabilidades

- Las probabilidades son una forma de resumir la incerteza que viene de:
  - Ignorancia: falta de conocimiento de todos los factores relevantes y condiciones iniciales
  - laxitud: cuando la lista de reglas (o excepciones) del fenómeno en cuestión es demasiado grande para enumerarlos
- Existen varias perspectivas para entender probabilidades:
  - Fecuentista: probabilidades son frecuencias de experimentos repetidos. Ej. “probabilidad de una moneda salir cara = 0.5” se interpreta como: en el limite de infinitas repeticiones saldrá cara 50% de veces
  - Objetivista: la probabilidad es una propiedad real de los objetos o eventos. En el ejemplo anterior, 0.5 seria una propiedad de la moneda
  - Subjetivista: las probabilidades son relativas al agente, el cual las usa para caracterizar sus creencias

    Esta visión es la mas aceptada actualmente en IA y es la que usaremos en adelante

Probabilidades

- Probabilidad Subjetiva o Bayesiana:
  - Establece el grado de creencia del agente en una proposición, dadas las evidencias disponibles
  - Probabilidad puede cambiar cuando llegan nuevas evidencias:
    - $P(\text{llegar a tiempo al aeropuerto} \mid A_{60}) = 0.85$
    - $P(\text{llegar a tiempo al aeropuerto} \mid A_{60}, \text{Lluvia}) = 0.5$

- Al igual que en la lógica clásica, las proposiciones son verdaderas o falsas, lo que cambia es que, el agente probabilístico atribuye grados de creencia a las proposiciones:
  - Probabilidad 0 de una proposición significa creer que ella es falsa con total convicción
  - Probabilidad 1 de una proposición significa creer que ella es verdad con total certeza

Probabilidades

Conceptos:

- Espacio muestral: conjunto de todos los posibles resultados (*outcomes*) de un experimento aleatorio. Ej. lanzar 2 dados:
  - $\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$
- Proposiciones: son afirmaciones de situaciones/eventos del mundo:
  - "Obtener una suma par al lanzar dos dados"
- Una proposición describe un conjunto de resultados del espacio muestral. Si ocurre alguno de dichos resultados entonces la proposición se verifica
  - Ejemplo, la proposición α = "obtener una suma par al lanzar dos dados" esta integrada por los resultados elementales: $\{(1,1), (1,3), \ldots\}$
  - Ejercicio: ¿Cuáles son resultados elementales que cumplen la proposición α = "obtener un numero primo al lanzar dos dados" ?

Probabilidades

Conceptos:

- Variable aleatoria: función que mapea resultados del espacio muestral a valores numéricos. Notación: primera letra en **mayúscula**!

  Ej. Lanzar 2 dados:

  $\Omega = \{(1,1), (1,2), (2,1), \ldots\}$ Espacio muestral

  $Suma = \{2, 3, \ldots\}$ Variable Aleatoria (suma total)

- Dominio de una variable aleatoria: conjunto de valores posibles

  $\text{Dominio}(Suma) = \{2, 3, \ldots, 12\}$

  Notación: valores de una variable aleatoria siempre en **minúscula**!

- Tipos de Variables aleatorias:
  - Booleanas: dominio = {*true*, *false*}. Ej: *SumaPar*
  - Discretas: dominio es un conjunto contable. Ej: *Estaciones*
  - Continuas: dominio continuo. Ej: *Temperatura*

Probabilidades

Conceptos:

- Proposiciones elementales pueden ser expresadas como atribuciones particulares de variables aleatorias

  Ej. La proposición α = "obtener suma par al lanzar dos dados" puede ser expresada como: $SumaPar = true$

Notación: las proposiciones con variables booleanas del tipo $X = true$ pueden ser abreviadas con el nombre de la variable en minúscula (x) y su negación con $\neg x$:

$$\text{sumapar} \leftrightarrow SumaPar = true$$

$$\neg\text{sumapar} \leftrightarrow SumaPar = false$$

- Proposiciones complejas pueden ser construidas a partir de proposiciones elementales y conectivos lógicos . Ejemplos:
  - $Clima = soleado \vee Estacion = verano$
  - $AlgunDadoPar = true \wedge \neg sumapar$

Probabilidades

Conceptos:

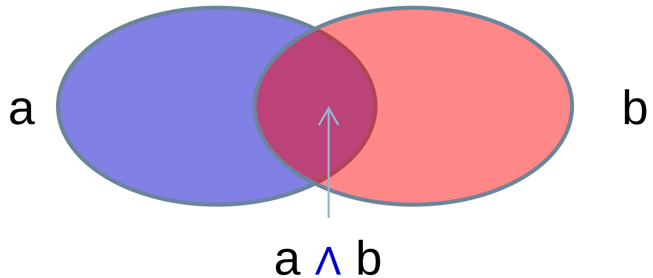
- Evento Atómico: Atribución particular de cada variable aleatória del estado del entorno
  - Son mutuamente exclusivos: solo un evento puede ocurrir en un instante
  - Son exhaustivos: algún evento tiene que ocurrir
- Ejemplo: Imaginemos un entorno con dos variables booleanas: *Caries* y *DolorDiente*, se tiene 4 eventos atómicos distintos:
  - $Caries = true \wedge DolorDiente = true$
  - $Caries = true \wedge DolorDiente = false$
  - $Caries = false \wedge DolorDiente = true$
  - $Caries = false \wedge DolorDiente = false$

Probabilidades

Axiomas de probabilidades (Kolmogorov):

- Para cualquier proposición ø:
  - $0 \leq P(ø) \leq 1$
  - $P(\Omega) = 1$, probabilidad de todo el espacio de eventos atómicos
  - $P(ø) = \sum_{\omega \in ø} P(\omega)$, ω son los eventos atómicos donde ø se cumple

  La probabilidad de cualquier proposición puede ser calculada desde que se tenga especificado las probabilidades de cada evento atómico (probabilidad conjunta completa)
- Para cualquier par de proposiciones *a*, *b*:
  - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$ (principio inclusion-exclusion)



Ejercicio: siendo:
a: obtener numero par al tirar un dado
b: obtener numero>4 al tirar un dado
Calcular $P(a \vee b)$

Probabilidades

- **Probabilidad incondicional o "*a priori*"**: Es el grado de creencia en una proposición en la ausencia de otras informaciones:

  Ejemplo: $P(Caries = true) = 0.1$; $P(Clima = soleado) = 0.72$

- **Distribución de Probabilidades de uma VA**: probabilidades de todos los posibles valores de la variable

  $$\text{Ej. } \mathbf{P}(Clima) = \begin{aligned} P(Clima &= soleado) = 0.72 \\ P(Clima &= lluvioso) = 0.1 \\ P(Clima &= nublado) = 0.08 \\ P(Clima &= nevoso) = 0.01 \end{aligned}$$

  $\mathbf{P}(Clima) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalizado, i.e., suma da 1)

Notación: la notación $\mathbf{P}(X)$ **(en negrita)** indica que se trata de una distribución de probabilidades sobre los diferentes valores de la variable aleatoria X. El resultado es un vector de valores, diferente de la notación $P(X=x)$ que indica la probabilidad (escalar) de un valor específico $X=x$

Probabilidades

- Distribución de Probabilidad Conjunta: probabilidades de todas las combinaciones de valores (eventos atómicos) de un conjunto de variables aleatorias

Ej. **P**$(Clima, Caries)$:

	Clima			
Caries	*soleado*	*lluvioso*	*nublado*	*nevoso*
true	0.144	0.02	0.016	0.02
false	0.576	0.08	0.064	0.08

La suma de todos los valores de la distribución de Prob. Conjunta debe ser **1**

- Una distribución de probabilidad conjunta especifica la prob. de todo evento atómico. Es por lo tanto una especificación completa de la incerteza sobre el entorno definido por esas variables
  - Cualquier cuestión sobre el dominio puede ser respondida a partir de la distribución de probabilidad conjunta!

Probabilidades

- Probabilidad Condicional o "a posteriori". Una vez que se observa los valores de algunas variables aleatorias (evidencias) las probabilidades *a priori* no son mas aplicables. En ese caso se debe usar probabilidades condicionales ("*a posteriori*"), que miden el grado de creencia en una proposición, dada la evidencia . Ejemplos:
  - $P(Caries = true \mid DolorDiente = true) = 0.8$
  - $P(Caries = true \mid dolordiente, Caries = true) = 1$
  - $P(Caries = true \mid dolordiente, Clima = soleado) = 0.8$
- Distribución de Probabilidad Condicional: La distribución condicional $\mathbf{P}(X|Y)$ especifica el conjunto de valores $P(X = x_i \mid Y = y_j)$ para cada i, j posible.
  - Ej. $\mathbf{P}(Caries \mid Clima)$:

Caries \ *Clima*	*soleado*	*lluvioso*	*nublado*	*nevoso*
true	0.2	0.2	0.2	0.2
false	0.8	0.8	0.8	0.8

La suma de las probabilidades para un valor fijo de la variable condicionante debe ser **1**

Probabilidades

Propiedades de Probabilidades Condicionales:

- Pueden ser definidas en términos de probabilidades *a priori*. Así, para cualquier par de proposiciones *a*, *b* se cumple:
  - $P(a \mid b) = P(a \wedge b) \,/\, P(b)$, si $P(b) > 0$
- Regla del producto:
  - $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$
  - Se puede generalizar para VAs. Ej:

$$\mathbf{P}(Clima, Caries) = \mathbf{P}(Caries \mid Clima)\ \mathbf{P}(Clima)$$

Clima

Caries	*soleado*	*lluvioso*	*nublado*	*nevoso*
true	0.144	0.02	0.016	0.02
false	0.576	0.08	0.064	0.08

Clima

Caries	*soleado*	*lluvioso*	*nublado*	*nevoso*
true	0.2	0.2	0.2	0.2
false	0.8	0.8	0.8	0.8

soleado	*lluvioso*	*nublado*	*nevoso*
0.72	0.1	0.08	0.1

No es multiplicación matricial !

Probabilidades

Propiedades de Probabilidades Condicionales:

- Regla de la Cadena: es obtenida a partir de aplicaciones sucesivas de la regla del producto:

$$\begin{aligned}\mathbf{P}(X_1,\ldots,X_n) &= \mathbf{P}(X_1, \ldots ,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots ,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1, \ldots ,X_{n-2})\, \mathbf{P}(X_n \mid X_1, \ldots ,X_{n-1}) \\ &= \ldots \\ &= \prod_{i= 1\ldots n} \mathbf{P}(X_i \mid X_1, \ldots , X_{i-1})\end{aligned}$$

Inferencia con Distribución de Probabilidad Conjunta

- Inferencia probabilística: proceso de calculo de probabilidades de proposiciones de consulta (*queries*) a partir de evidencias observadas

- Cuando se dispone la distribución de probabilidad conjunta de un conjunto de variables aleatorias, esta constituye la base de conocimiento de la cual se puede derivar cualquier inferencia probabilística relativa a las variables modeladas

probabilidad conjunta

Inferencia con Distribución de Probabilidad Conjunta

- Ejemplo: dada la dist. de prob conjunta **P**(*Caries, DolorDiente, Sonda*)

DolorDiente

	dolordiente		*¬dolordiente*	
Caries	*sonda*	*¬sonda*	*sonda*	*¬sonda*
caries	0.108	0.012	0.072	0.008
¬caries	0.016	0.064	0.144	0.576

Sonda: variable aleatoria que indica si la sonda dental encontró anomalía en el diente

La probabilidad de cualquier proposición ø es: $P(\o) = \sum_{\omega \in \o} P(\omega)$, la suma de probabilidades de los eventos atómicos ω donde la proposición se cumple

Inferencia con Distribución de Probabilidad Conjunta

- **Ejemplo**: dada la dist. de prob conjunta **P**(*Caries, DolorDiente, Sonda*)

DolorDiente

	dolordiente		*¬dolordiente*	
Caries	*sonda*	*¬sonda*	*sonda*	*¬sonda*
caries	0.108	0.012	0.072	0.008
¬caries	0.016	0.064	0.144	0.576

Sonda: variable aleatoria que indica si la sonda dental encontró anomalía en el diente

La probabilidad de cualquier proposición ø es: $P(ø) = \sum_{\omega \in ø} P(\omega)$, la suma de probabilidades de los eventos atómicos ω donde la proposición se cumple

- **Ejercicio**: Cuánto es $P(dolordiente)$?

$$P(dolordiente) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inferencia con Distribución de Probabilidad Conjunta

- Ejemplo: dada la dist. de prob conjunta **P**(*Caries, DolorDiente, Sonda*)

	DolorDiente			
	dolordiente		*¬dolordiente*	
Caries	*sonda*	*¬sonda*	*sonda*	*¬sonda*
caries	0.108	0.012	0.072	0.008
¬caries	0.016	0.064	0.144	0.576

Sonda: variable aleatoria que indica si la sonda dental encontró anomalía en el diente

La probabilidad de cualquier proposición ø es: $P(ø) = \sum_{\omega \in ø} P(\omega)$, la suma de probabilidades de los eventos atómicos ω donde la proposición se cumple

- Ejercicio: Cuánto es $P(dolordiente \lor caries)$?

$$P(dolordiente \lor caries) = 0.108 + 0.012 + 0.016 + 0.064 + 0.072 + 0.008 = 0.28$$

Inferencia con Distribución de Probabilidad Conjunta

Calculo de Probabilidades Condicionales:

- Se usa la forma: $P(a \mid b) = P(a \wedge b) \,/\, P(b)$
- Ejemplo: Calcular $P(\neg caries \mid dolordiente)$

DolorDiente

Caries	*dolordiente*		*¬dolordiente*	
	sonda	*¬sonda*	*sonda*	*¬sonda*
caries	0.108	0.012	0.072	0.008
¬caries	0.016	0.064	0.144	0.576

$$
\begin{aligned}
P(\neg caries \mid dolordiente) &= P(\neg caries \wedge dolordiente) \,/\, P(dolordiente) \\
&= (.016 + .064\,) \,/\, (.108 + .012 + .016 + .064) \\
&= 0.4
\end{aligned}
$$

- Ejercicio: Calcular $P(caries \mid dolordiente)$

Inferencia con Distribución de Probabilidad Conjunta

Calculo de Distribuciones de Probabilidades Condicionales

- Idea: computar la distribución sobre la variable de consulta, fijando las variables de evidencia y sumando sobre las demás variables
- Ejemplo: Calcular $\mathbf{P}(Caries \mid dolordiente)$

DolorDiente

Caries	*dolordiente*		*¬dolordiente*	
	sonda	*¬sonda*	*sonda*	*¬sonda*
caries	0.108	0.012	0.072	0.008
¬caries	0.016	0.064	0.144	0.576

Constante de normalización

$\boxed{\alpha = 1/\, P(dolordiente)}$

$$\mathbf{P}(Caries \mid dolordiente) = \mathbf{P}(Caries \wedge dolordiente) \,/\, P(dolordiente) = \alpha\, \mathbf{P}(Caries \wedge dolordiente)$$

$$= \alpha[\mathbf{P}(Caries, dolordiente, sonda) + \mathbf{P}(Caries, dolordiente, \neg sonda)]$$

$$= \alpha[\ \langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle\ ]$$

$$= \alpha\ \langle 0.12, 0.08\rangle \quad = \langle 0.6, 0.4\rangle$$

No es necesario calcular α, solo normalizar el vector para que sume 1

Inferencia con Distribución de Probabilidad Conjunta

Inferencia Probabilística por enumeración

- Objetivo: Calcular la distribución de probabilidades de variables de consulta X, dados valores específicos **e** de variables de evidencia **E**
- Sea **Y** el conjunto de variables no observadas, la consulta $\mathbf{P}(X|\mathbf{e})$ puede ser calculada como:

$$\mathbf{P}(X|\mathrm{e}) = \alpha \sum_{\mathrm{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

la sumatoria es sobre las diferentes combinaciones de valores de las variables en **Y**

- Note que la unión de las variables X, **Y, E** constituye el conjunto completo de las variables del dominio
  - Así, $\mathbf{P}(X, \mathbf{e}, \mathbf{y})$ es simplemente un subconjunto de probabilidades de la distribución de probabilidad conjunta total

Inferencia con Distribución de Probabilidad Conjunta

Limitaciones de Inferencia Probabilística por enumeración

- Complejidad de tiempo (peor caso): $O(d^n)$
  - donde d es la cardinalidad (numero de valores de una variable) mas alta entre todas las variables; n es el número de variables.
- Complejidad de espacio $O(d^n)$ para almacenar la distribución conjunta
- Como encontrar las probabilidades para $O(d^n)$ elementos?

Independencia

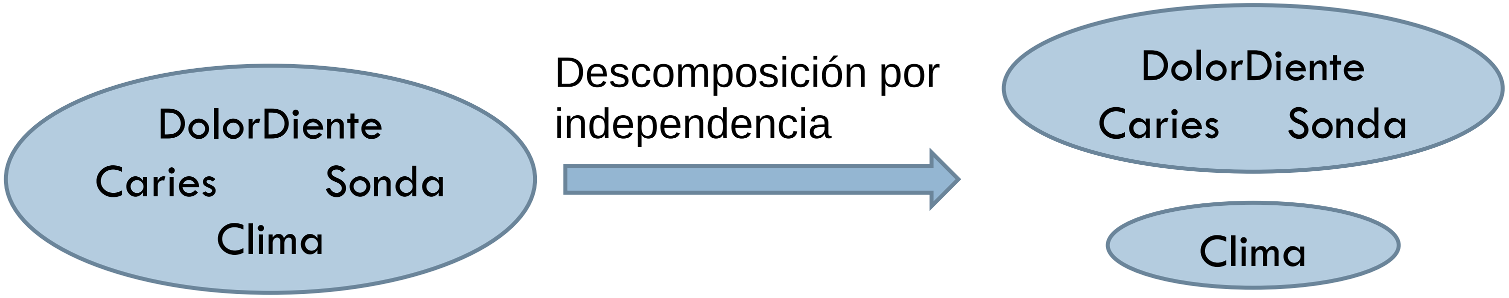
Independencia Total o Marginal

- Dos variables aleatorias A, B son independientes si y solamente si:

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

- Ejemplo: Si al dominio {*DolorDiente*, *Caries*, *Sonda*} aumentamos la variable aleatoria *Clima*, se cumplirá:

  P(*DolorDiente*, *Caries*, *Sonda*, *Clima*) = **P**(*Caries*, *DolorDiente*, *Sonda*) **P**(*Clima*)



  - La probabilidad conjunta puede ser especificada con 12 valores (8 para **P**(*Caries*, *DolorDiente*, *Sonda*) y 4 para **P**(*Clima*)), en vez de 32 valores
  - Independencia total es rara

Independencia

Independencia Condicional

- Si se sabe que hay caries, la probabilidad de la sonda encontrar una anomalía en el diente no depende de tener dolor de diente
  - $\mathbf{P}(Sonda \mid dolordiente, caries) = \mathbf{P}(Sonda \mid caries)$
- La misma independencia ocurre si se sabe que no hay caries
  - $\mathbf{P}(Sonda \mid dolordiente, \neg caries) = \mathbf{P}(Sonda \mid \neg caries)$
- Esto indica que *Sonda* es condicionalmente independiente de *DolorDiente* cuando se conoce el valor de *Caries*
  - $\mathbf{P}(Sonda \mid DolorDiente, Caries) = \mathbf{P}(Sonda \mid Caries)$

Independencia

Independencia Condicional

- Escribiendo la distribución conjunta usando la regla de la cadena:

$$\mathbf{P}(DolorDiente, Sonda, Caries) = \mathbf{P}(DolorDiente \mid Sonda, \mathrm{Caries})\mathbf{P}(Sonda \mid Caries)\mathbf{P}(Caries)$$

$$= \mathbf{P}(DolorDiente \mid \mathrm{Caries})\ \mathbf{P}(Sonda \mid \mathrm{Caries})\ \mathbf{P}(\mathrm{Caries})$$

- En dicho ejemplo, el numero de valores para especificar la distribución conjunta pasa de 7 a 5 valores

- En la mayoría de casos, el uso de independencia condicional reduce el tamaño de la distribución conjunta de exponencial a lineal en el numero de variables

Regla de Bayes

Regla de Bayes

- De la regla del producto : $P(a \wedge b) = P(a|b)\, P(b) = P(b|a)\, P(a)$

  Se deduce la regla de Bayes: $\boxed{P(a|b) = P(b|a)\, P(a) \;/\; P(b)}$

- En la forma de distribuciones de probabilidades:

$$\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\,\mathbf{P}(Y) \;/\; \mathbf{P}(X) = \alpha\, \mathbf{P}(X|Y)\,\mathbf{P}(Y)$$

- Es útil para derivar probabilidades de diagnóstico ($P(Cause|Effect)$) a través de probabilidades causales ($P(Effect|Cause)$):
  - $P(Cause|Effect) = P(Effect|Cause)\, P(Cause) \;/\; P(Effect)$

- Ejemplo: Sea: C = Infectado con Covid19 , T = Tos

$$P(c) = 0.001; \quad P(t) = 0.1; \quad P(t|c) = 0.8; \quad P(c|t) = ?$$

$$P(c|t) = P(t|c)\, P(c) \;/\; P(t) = 0.8 \times 0.001/0.1 = 0.008$$

Preguntas?