
Una Introducción a las Redes Bayesianas

Serafín Moral

Departamento de Ciencias de la Computación
Universidad de Granada

Redes Bayesianas

Sistemas Expertos Probabilísticos

- Representar conocimiento con incertidumbre.
- Después se puede manipular para razonamiento y toma de decisiones.
- Se pueden tratar muchas variables.
- Las reglas (probabilidades) se pueden estimar a partir de datos.
- Los modelos tienen una interpretación clara y bien definida.
- Actualmente están teniendo un gran desarrollo.

Indicios de importancia

- En 1999 J. Pearl uno de los pioneros en Inteligencia Artificial recibió el *IJCAI Award for Research Excellence* (El séptimo de estos premios bianuales). Esta es la distinción más importante en Inteligencia Artificial.
- Evolución de publicaciones en JCR (base de datos de publicaciones) bajo la búsqueda *Bayesian Networks*:
 - 1990-1999: 118 publicaciones
 - 2000-2006: 587 publicacione
- Algunos artículos altamente citados en `scholar.google.com`:
 - Aprendizaje de Hecherman y col. (1995): 1249 citas
 - Clasificación supervisada de Friedman y col. (1997): 880 cirtas
 - Análisis de datos de expresión genética de Friedman y col. (2000): 906 citas
 - Filtrado de clientes de Breese y col. (1998) 1129 citas
 - Libro de Judea Pearl: 8027 citas

Referencias

- E. Castillo, J.M. Gutiérrez, A.S. Hadi (1996) *Sistemas Expertos y Modelos de Redes Probabilísticas*. Monografías de la Academia de Ingeniería. Academia de Ingeniería, Madrid.
- R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter (1999) *Probabilistic Networks and Expert Systems*. Springer-Verlag, Nueva York.
- F.V. Jensen (1996) *An Introduction to Bayesian Networks*. UCL Press, Londres.
- F.V. Jensen (2001) *Bayesian Networks and Decision Graphs*. Springer-Verlag, Nueva York.
- F.V. Jensen, T.D. Nielsen (2007) *Bayesian Networks and Decision Graphs* (2nd Edition). Springer-Verlag, Nueva York.
- U. Kjaerulff, A.L. Madsen (2007) *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer-Verlag.
- J. Pearl (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

Contenido

- Problemas para manejar conocimiento incierto
- Teoría de la Probabilidad
- Independencia
- Redes Bayesianas, D-separación
- Construcción de redes Bayesianas
- Algoritmo de borrado o de eliminación de variables
- El programa Elvira
- Otros temas: configuración de máxima probabilidad, diagramas de influencia, aprendizaje

Sistemas Basados en Reglas

SI es un animal con pelo ENTONCES es un mamífero

Incertidumbre:

SI tiene fiebre y dolor de cabeza, entonces tiene gripe (certeza 0.7)

- MYCIN fue diseñado para determinar tratamientos en infecciones de la sangre con 300 reglas.
- Si una conclusión se obtiene por varias vías, los valores de certeza se combinan.
- Las certezas no eran probabilidades: éstas imponen unas reglas de cálculo muy estrictas.
- Su correcto funcionamiento se basa en un cuidadoso diseño de las reglas en función del uso que se hace de ellas.

Problemas

- La validez de una regla depende del contexto.

Si conozco el nivel de estudios de una persona, obtengo información sobre su nivel de ingresos. Esta información puede ser equivocada y ponerse de manifiesto si conozco el puesto de trabajo concreto que esta persona desarrolla

Si al salir de casa vemos el césped mojado podemos sospechar que ha llovido. Si descubrimos que nos hemos dejado la manguera abierta, dejamos de sospechar que ha llovido.

Problemas

- La validez de una regla depende del contexto.

Si conozco el nivel de estudios de una persona, obtengo información sobre su nivel de ingresos. Esta información puede ser equivocada y ponerse de manifiesto si conozco el puesto de trabajo concreto que esta persona desarrolla

Si al salir de casa vemos el césped mojado podemos sospechar que ha llovido. Si descubrimos que nos hemos dejado la manguera abierta, dejamos de sospechar que ha llovido.

- Las reglas con incertidumbre deberían de poder usarse en ambas direcciones.

Si hay fuego debe de haber humo

Si vemos humo sospechamos la existencia de fuego

Problemas

- La validez de una regla depende del contexto.

Si conozco el nivel de estudios de una persona, obtengo información sobre su nivel de ingresos. Esta información puede ser equivocada y ponerse de manifiesto si conozco el puesto de trabajo concreto que esta persona desarrolla

Si al salir de casa vemos el césped mojado podemos sospechar que ha llovido. Si descubrimos que nos hemos dejado la manguera abierta, dejamos de sospechar que ha llovido.

- Las reglas con incertidumbre deberían de poder usarse en ambas direcciones.

Si hay fuego debe de haber humo

Si vemos humo sospechamos la existencia de fuego

- Correlación entre las informaciones. Si una misma información se repite muchas veces no debe de aumentar nuestra certidumbre.

Probabilidad

La probabilidad como medida de certeza, no presenta ninguno de estos problemas.

Puedo tener $P(\textit{Gripe}|\textit{Fiebre}) = 0.9, P(\textit{Gripe}|\textit{Fiebre}, \textit{Otitis}) = 0.1$.

Presenta otro: **necesito una distribución de probabilidad conjunta.**

- Si tengo 30 variables, X_1, \dots, X_n y cada una de ellas, X_i , toma dos posibles valores $\{a_i, \bar{a}_i\}$, entonces necesitamos partir de las probabilidades de todas las combinaciones

$$(x_1, x_2, \dots, x_n), \quad x_i \in \{a_i, \bar{a}_i\}$$

Si $n = 30$, necesitamos 2^{30} valores, pero inicialmente solemos disponer de unas cuantas probabilidades condicionadas.

Probabilidad

Sólo vamos a considerar la probabilidad sobre conjuntos finitos.

Vamos a suponer un conjunto U finito de sucesos elementales y una familia de conjuntos o sucesos \mathbf{B} (si U es finito esta familia suele ser el conjunto de las partes de U).

Una **medida de probabilidad** sobre (U, \mathbf{B}) es una aplicación $P : \mathbf{B} \rightarrow [0, 1]$, que verifica:

- $P(U) = 1$

- Si A y C son disjuntos $P(A \cup C) = P(A) + P(C)$

Probabilidad Condicional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

Aunque tiene sentido hablar de probabilidad condicionada a sucesos de probabilidad 0, y en ese caso se debe de verificar:

$$P(A \cap B) = P(A|B).P(B)$$

La probabilidad $P(A|B)$ es la probabilidad de A cuando conocemos que B y sólo B es cierto.

El Teorema de la Probabilidad Total

Si un paciente tiene la enfermedad E , entonces un test T resulta positivo con probabilidad 0.95 . Si la enfermedad no está presente el test es positivo con probabilidad 0.03 . Si la probabilidad de sufrir la enfermedad es 0.01 , ¿Cual es la probabilidad de que un paciente cualquiera presente un test positivo?

Queremos la probabilidad de $T+$, pero sólo conocemos la probabilidad de $T+$ condicionado a la enfermedad y a que no se tenga la enfermedad, y además conocemos las probabilidades de tener y no tener la enfermedad.

Si $\{H_i\}_{i \in I}$ es una colección finita de sucesos disjuntos dos a dos y cuya unión es el suceso seguro (U).

$$P(B) = \sum_{i \in I} P(B|H_i)P(H_i)$$

El Teorema de la Probabilidad Total

Si $\{H_i\}_{i \in I}$ es una colección finita de sucesos disjuntos dos a dos y cuya unión es el suceso seguro (U).

$$P(B) = \sum_{i \in I} P(B|H_i)P(H_i)$$

Demostración: $P(B) = P(B \cap U) = P(B \cap (\bigcup_{i \in I} H_i)) =$
 $P(\bigcup_{i \in I} (B \cap H_i)) = \sum_{i \in I} P(B \cap H_i) = \sum_{i \in I} P(B|H_i)P(H_i)$

$$P(T+) = P(T+|E).P(E) + P(T+|\neg E).P(\neg E) =$$
$$0.95 \times 0.01 + 0.03 \times 0.99 = 0.0392$$

El Teorema de Bayes

Si un paciente tiene la enfermedad E , entonces un test T resulta positivo con probabilidad 0.95 . Si la enfermedad no está presente el test es positivo con probabilidad 0.03 . Si la probabilidad de sufrir la enfermedad es 0.01 , ¿Cual es la probabilidad de que un paciente con un test positivo sufra la enfermedad?

Conocemos $P(T+|E)$ y las probabilidades $P(T+|\neg E)$, $P(E)$ y queremos la probabilidad $P(E|T+)$. Es como invertir la probabilidad condicionada.

Si $\{H_i\}_{i \in I}$ es una colección de sucesos disjuntos dos a dos y cuya unión es el suceso seguro (U).

$$P(H_j|B) = \frac{P(H_j \cap B)}{P(B)} = \frac{P(B|H_j) \cdot P(H_j)}{P(B)} = \frac{P(B|H_j) \cdot P(H_j)}{\sum_{i \in I} P(B|H_i) P(H_i)}$$

El Teorema de Bayes

Si un paciente tiene la enfermedad E , entonces un test T resulta positivo con probabilidad 0.95 . Si la enfermedad no está presente el test es positivo con probabilidad 0.03 . Si la probabilidad de sufrir la enfermedad es 0.01 , ¿Cual es la probabilidad de que un paciente con un test positivo sufra la enfermedad?

Si $\{H_i\}_{i \in I}$ es una colección de sucesos disjuntos dos a dos y cuya unión es el suceso seguro (U).

$$P(H_j|B) = \frac{P(B|H_j).P(H_j)}{\sum_{i \in I} P(B|H_i)P(H_i)}$$

En el caso del ejemplo,

$$P(E|T+) = \frac{P(T+|E).P(E)}{P(T+|E).P(E)+P(T+|\neg E).P(\neg E)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.03 \times 0.99} = 0.0095 / 0.0392 = 0.2423$$

Variables Inciertas

Una variable es una magnitud medible en un determinado problema. Es incierta cuando su resultado no puede ser determinado con exactitud.

- Vamos a hablar en términos de variables inciertas. Las variables aleatorias las representaremos por X, Y, Z, \dots
 - Temperatura con valores en $\{\leq 36, 36.5, 37, 37.5, 38, 38.5, 39, 39.5, \geq 40\}$
 - Hepatitis con valores en $\{Presente, Ausente\}$
 - N. de Hijos con valores en $\{0, 1, 2, 3, > 3\}$
- Un valor genérico de la variable X se representará por x
- Un conjunto de variables se representará por \mathbf{X}
- Un valor genérico de \mathbf{X} se representará por \mathbf{x}

Variables Discretas y Continuas

- Una variable es **discreta** si el conjunto de valores posibles es finito (*Presencia de una enfermedad, Número de asignaturas matriculadas, Sexo, Estudios realizados*)
- Una variable es **continua** si toma valores en un intervalo de los números reales (*Altura, Peso, Luminosidad*).
- Nosotros vamos a considerar variables discretas
- Si hay continuas las discretizamos dividiéndolas en un conjunto finito de intervalos

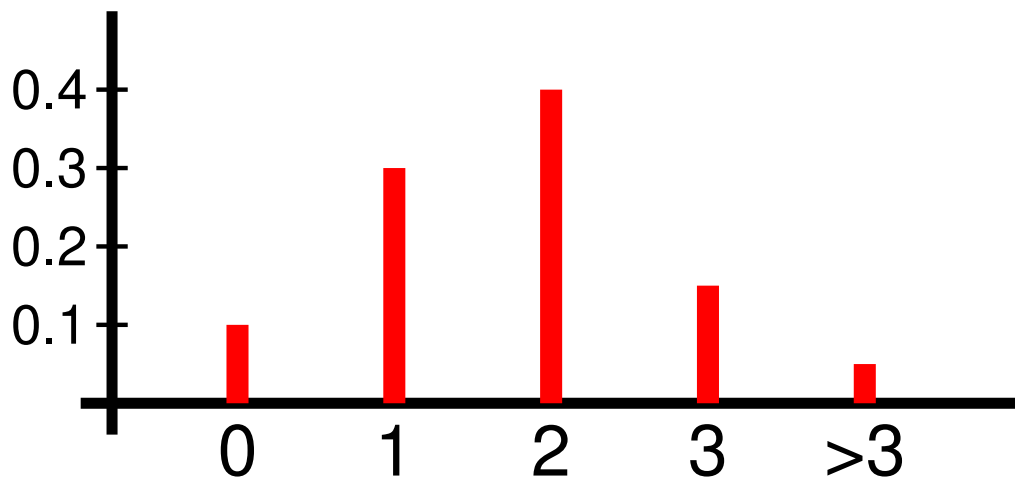
Distribuciones de probabilidad

Una distribución de probabilidad p sobre X es la función que asigna a cada valor x , la probabilidad con que X toma dicho valor. Se notará como $p(x)$.

Ejemplo: Variable **N. de hijos** con valores $\{0, 1, 2, 3, > 3\}$ y la distribución de probabilidad:

p	0	1	2	3	>3
	0.1	0.3	0.4	0.15	0.05

Sus valores deben de sumar 1.



Distribuciones Conjuntas

Si tenemos un conjunto de variables \mathbf{X} una **distribución de probabilidad conjunta** asocia a cada posible valor de estas \mathbf{x} , su probabilidad $p(\mathbf{x})$.

Ejemplo: Tenemos las variables X (*Color de los ojos*) e Y (*Color del pelo*), una distribución conjunta sobre estas variables puede ser

		Y	
		Moreno	Rubio
X	Marrones	0.5	0.15
	Azules	0.05	0.3

También podemos tener distribuciones que dependan de más de dos variables, p.e. $p(x, y, z)$.

Distribuciones Condicionadas

Si tenemos dos variables, X, Y , la distribución de probabilidad de Y dado X , es una función de los conjuntos dónde Y y X toman sus valores en $[0,1]$, dada por

$$p(y|x) = P(Y = y|X = x)$$

Es evidente que $\forall x, \sum_y p(y|x) = 1$

Caso de los test y de las enfermedades $p(t|e)$

	$t+$	$t-$
e	0.95	0.05
$\neg e$	0.03	0.97

Distribuciones Condicionadas

Si condicionamos a varias variables, tenemos que dar el valor de probabilidad de la variable para cada combinación de valores de las variables condicionadas.

Ejemplo:

Sean X *Cáncer de Pulmón*, Y *Fumador* y Z *Sexo*. Supongamos que tenemos que una probabilidad condicionada de X dadas las variables Y y Z , tenemos que dar una tabla de valores como la siguiente:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.5	0.4	0.2	0.1
X=No	0.5	0.6	0.8	0.9

Muchas Variables

¿Qué pasa si el número de variables es elevado?

Supongamos que en el problema de la enfermedad que se detecta con un test, en vez de un sólo test tenemos 10 (T_1, \dots, T_{10}) .

Ahora para especificar el problema y después poder aplicar el teorema de Bayes, deberemos indicar todos los valores

$$p(t_1, \dots, t_{10} | e), \quad \forall t_i \in \{+, -\}, e \in \{\text{pres}, \text{aus}\}$$

Esto constituye un número importante de valores y crece exponencialmente en función del número de tests.

Independencia Condicional

Una hipótesis que permite simplificar el problema: Los tests son condicionalmente independientes dada la enfermedad.

Entonces, podemos expresar

$$p(t_1, \dots, t_{10} | e) = \prod_{i=1}^{10} p(t_i | e)$$

La independencia será definida formalmente más adelante, pero se puede interpretar como que los tests tienen distintos mecanismos de medición, se fijan en distintos factores, no se equivocan siempre en los mismos casos.