

Manejo de Incertidumbre

Eduardo Morales, Enrique Sucar

INAOE

Contenido

- 1 Introducción
- 2 Probabilidad
- 3 Técnicas Empíricas
- 4 Redes Bayesianas
 - Estructura
 - Parámetros
 - Inferencia
 - Aprendizaje
- 5 Redes Bayesianas Causales

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Introducción

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Suar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Los sistemas inteligentes deben ser capaces de representar y razonar con incertidumbre.
- Existen varias causas de incertidumbre que tienen que ver con la información, el conocimiento y la representación:
- **Información:** incompleta, poco confiable, ruido, ...
- **Conocimiento:** impreciso, contradictorio, ...
- **Información:** no adecuada, falta de poder descriptivo

Ejemplos

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Diagnóstico médico e industrial
- Predicción financiera
- Exploración minera / petrolera
- Interpretación de imágenes
- Reconocimiento de voz
- Monitoreo / control de procesos industriales
- Robótica
- Modelado del usuario
- Bioinformática

Efectos de Incertidumbre

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Se pierden varias propiedades de los sistemas que no tienen incertidumbre, basados en lógicas o reglas, lo cual hace el manejo de incertidumbre más complejo:
- Modularidad
- Monotonicidad

Modularidad

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Un sistema de reglas es modular, ya que para saber la verdad de una regla sólo tiene que considerarla a ésta, sin importar el resto del conocimiento
- Pero si hay incertidumbre ya no puedo considerar la regla por si sola, debo tomar en cuenta otras reglas

Monotónico

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Suar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Un sistema es monotónico si al agregar nueva información a su base de datos, no se alteran las conclusiones que seguían de la base de datos original
- Si hay incertidumbre ya no puedo considerar que la certeza en una hipótesis ya no puede cambiar, debo tomar en cuenta otras reglas que involucren a dicha hipótesis

Técnicas No Numéricas

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Lógicas no-monotónicas
- Sistemas de mantenimiento de verdad (TMS, ATMS)
- Teoría de endosos

Técnicas Numéricas

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Empíricas (MYCIN, Prospector)
- Métodos aproximados
- Lógica difusa
- Teoría de Dempster-Shafer
- Probabilísticas - Redes Bayesianas

Axiomas

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- 1 $P(A)$ es una función en $[0, 1]$.
- 2 $P(A, B | C) = P(A | C)P(B | A, C)$ (regla del producto).
- 3 $P(A | B) + P(\neg A | B) = 1$ (regla de la suma).

Probabilidad Condicional

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- $P(H | B)$ condicionada sólo en el contexto B es la probabilidad *a priori*;
- Al incorporar información adicional D se conoce como probabilidad *posterior*, $P(H | D, B)$
- La probabilidad condicional se define como (por simplicidad omitimos el contexto):
$$P(H | D) = P(H, D) / P(D)$$

Regla de Bayes

De la regla del producto:

$$P(D, H | B) = P(D | H, B)P(H | B) = P(H | D, B)P(D | B) \quad (1)$$

De donde:

$$P(H | D, B) = \frac{P(H | B)P(D | H, B)}{P(D | B)} \quad (2)$$

Esta última ecuación se conoce como *La regla o teorema de Bayes*

The term $P(H | B)$ is the *prior* and $P(D | H, B)$ is the *likelihood*

Independencia

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- En ciertos casos la probabilidad de H no es afectada por D , entonces se dice que H y D son *independientes*:
$$P(H, D | B) = P(H | B)$$
- La regla del producto se puede simplificar a:
$$P(A, B | C) = P(A | C)P(B | C)$$

Independencia Condicional

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Dos proposiciones independientes sólo dado el contexto son *marginalmente* independientes; si son independientes dada información adicional, E , entonces son *condicionalmente* independientes:
$$P(H, D \mid B, E) = P(H \mid B, E)$$
- Ejemplo: H representa *regar el jardín*, D la *predicción del clima* y E *lloviendo*

Regla de la Cadena

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Probabilidad de la conjunción de N proposiciones $P(A_1, A_2, \dots, A_N | B)$, llamada la probabilidad *conjunta*
- Generalizando la regla del producto a N proposiciones obtenemos la *regla de la cadena*:
$$P(A_1, A_2, \dots, A_N | B) = P(A_1 | A_2, A_3, \dots, A_N, B)P(A_2 | A_3, A_4, \dots, A_N, B) \cdots P(A_N | B)$$
- Se pueden usar relaciones de independencia condicional para simplificarla

Probabilidad Total

- Dada la partición, $B = \{B_1, B_2, \dots, B_n\}$, del espacio de muestreo Ω , tal que $\Omega = B_1 \cup B_2 \cup \dots \cup B_n$ and $B_i \cap B_j = \emptyset$
- A es la unión de las intersecciones:

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_n \cap A)$$
- Entonces:

$$P(A) = \sum_i P(A | B_i)P(B_i) \quad (3)$$

- Incorporando esto en la regla de Bayes, obtenemos el Teorema de Bayes:

$$P(B | A) = \frac{P(B)P(A | B)}{\sum_i P(A | B_i)P(B_i)} \quad (4)$$

Probabilidad en Sistemas Expertos

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Sean $H = h_1, \dots, h_n$ el conjunto de n posibles hipótesis y $E = e_1, \dots, e_m$, m posibles evidencias. En general lo que queremos encontrar es la h^* más probable dado E .
- Se tiene que calcular $P(h | E)$ para cada subconjunto de hipótesis y seleccionar la de mayor probabilidad utilizando el teorema de Bayes.
- Dos opciones básicas: exhaustivo o independiente

Exhaustivo

$$P(h_i | e_{j1} \dots e_{jk}) = \frac{P(e_{j1} \dots e_{jk} | h_i)P(h_i)}{\sum_{j=1}^n P(e_{j1} \dots e_{jk} | h_j)P(h_j)}$$

- Se requiere calcular las probabilidades condicionales $P(E | h_i)$ para cada combinación de evidencias (en general no se pueden calcular de sus componentes individuales).
- Esto implica que se tienen que conocer un número exponencial de probabilidades!

Independiente

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Suar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

$$P(h_i | e_{j1} \dots e_{jk}) = \frac{P(e_{j1} | h_i) \dots P(e_{jk} | h_i) P(h_i)}{\sum_{l=1}^n P(e_{j1} | h_l) \dots P(e_{jk} | h_l) P(h_l)}$$

- Las evidencias son condicionalmente independientes
- Con esto, sólo se requieren $m \times n$ probabilidades condicionales y $n - 1$ probabilidades a priori.

Análisis

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Suar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- En la opción 1 el sistema se vuelve demasiado complejo, mientras que la opción 2 puede no ser realista para muchas aplicaciones.
- Una alternativa es buscar un compromiso entre ambos extremos, esto se logra mediante las redes Bayesianas (que veremos más adelante)
- Dados estas limitaciones de los enfoques básicos, en los primeros sistemas expertos se propusieron técnicas *ad hoc* para el manejo de incertidumbre.

Técnicas Empíricas

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Son técnicas empíricas o ad-hoc orientadas a resolver aplicaciones específicas y sin un fuerte fundamento teórico.
- Las más conocidas son las que corresponden a dos de los primeros sistemas expertos:
- PROSPECTOR (exploración minera)
- MYCIN (diagnóstico de enfermedades infecciosas en la sangre)

MYCIN

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Las técnicas desarrolladas en MYCIN y Prospector son similares, ambas consideran sistemas basados en reglas a los que se les adicionan Factores de Certeza o Probabilidades Subjetivas, respectivamente:

Si: se observa cierta evidencia E , Entonces: se concluye cierta hipótesis H con probabilidad (certeza, ...) P

- Veremos brevemente el método de MYCIN.

Factores de Certeza

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- MYCIN define un Factor de Certeza que se asocia a cada regla y cada evidencia, y se definen un conjunto de reglas para combinar estos factores.
- Los factores de certeza están en el rango $[-1, +1]$

Reglas de Combinación

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Propagación o reglas en serie:
- $CF(h, e') = CF(h, e) \times \max(0, CF(e, e'))$
- Conjunción (AND) y disjunción (OR) de evidencias:
- $CF(e_1 \text{ and } e_2) = \min[CF(e_1), CF(e_2)]$
- $CF(e_1 \text{ or } e_2) = \max[CF(e_1), CF(e_2)]$

Reglas de Combinación

- Co-conclusión o reglas en paralelo:

$$CF(h, e'_1 \text{ co } e'_2) = \left\{ \begin{array}{l} \frac{CF(h, e'_1) + CF(h, e'_2)(1 - CF(h, e'_1))}{CF(h, e'_1) + CF(h, e'_2)} \\ \text{if } CF(h, e'_i) > 0, i = 1, 2 \\ \\ 1 - \min\{|CF(h, e'_1)|, |CF(h, e'_2)|\} \\ \text{if } -1 < CF(h, e'_1) \times CF(h, e'_2) \leq 0 \\ \frac{CF(h, e'_1) + CF(h, e'_2)(1 + CF(h, e'_1))}{CF(h, e'_1) + CF(h, e'_2)} \\ \text{if } CF(h, e'_i) < 0, i = 1, 2 \end{array} \right\}$$

Ejemplo

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

R1: IF A and (B or C) Then H cf 0.8

R2: If D and F Then B cf 0.6

R3: If F or G Then H cf 0.4

R4: If A Then D cf 0.75

R5: If I Then G cf 0.3

Se conoce:

$CF(A, Ev) = 1,$

$CF(C, Ev) = 0.5,$

$CF(F, Ev) = 0.7,$

$CF(I, Ev) = -0.4$

Análisis

- Aunque pretendía apartarse de probabilidad, se ha demostrado [Heckerman 86] que la técnica de MYCIN corresponde a un subconjunto de probabilidad con una serie de suposiciones implícitas:
 - La evidencia es condicionalmente independiente de la hipótesis y su negación.
 - La red de inferencia debe corresponder a un árbol para que los resultados sean coherentes.
 - Las fórmulas para conjunción y disjunción (min y max) sólo son válidas si uno de los términos es subconjunto del otro.
- Estas suposiciones no son válidas en muchas aplicaciones por lo que el método de MYCIN no se puede generalizar.

Redes Bayesianas

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

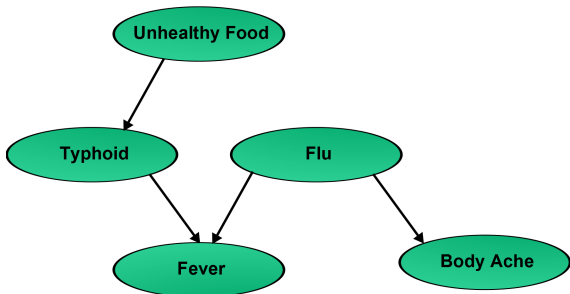
Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Modelos gráficos dirigidos que representan relaciones de dependencia entre variables
- Los nodos representan variables aleatorias y los arcos relaciones de dependencia directa
- La estructura representa una serie de relaciones de independencia condicional

Ejemplo



- *Fever* is independent of *Body ache* given *Flu* (common cause)
- *Fever* is independent of *Unhealthy food* given *Typhoid* (indirect cause)
- *Typhoid* is independent of *Flu* when *Fever* is NOT known (common effect). Knowing *Fever* makes *Typhoid* and *Flu* dependent

Introducción

- In addition to the structure, a Bayesian network considers a set of local parameters, which are the conditional probabilities for each variable given its parents in the graph
- The joint probability of all the variables in the network can be represented based on these local parameters; this usually implies an important saving in the number of required parameters
- Given a Bayesian network we can answer several probabilistic queries. For instance, for the previous example: What is the probability of Fever given Flu? Which is more probable, Typhoid or Flu, given Fever and Unhealthy food?

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Bayesian Networks

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- A Bayesian network (BN) represents the joint distribution of a set of n (discrete) variables, X_1, X_2, \dots, X_n , as a directed acyclic graph (DAG) and a set of conditional probability tables (CPTs)
- Each node, that corresponds to a variable, has an associated CPT that contains the probability of each state of the variable given its parents in the graph
- The structure of the network implies a set of conditional independence assertions, which give power to this representation

Conditional Independence Assertions

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- The conditional independence assertions implied by the structure of a BN should correspond to the conditional independence relations of the joint probability distribution, and vice versa
- If X is conditionally independent of Z given Y :
 - In the probability distribution: $P(X|Y, Z) = P(X|Y)$.
 - In the graph: $I \langle X | Y | Z \rangle$.

D-Separation

- Conditional independence assertions can be verified directly from the structure of a BN using a criteria called *D-separation*
- 3 basic BN structures for 3 variables and 2 arcs:
 - Sequential: $X \rightarrow Y \rightarrow Z$.
 - Divergent: $X \leftarrow Y \rightarrow Z$.
 - Convergent: $X \rightarrow Y \leftarrow Z$.
- In the first two cases, X and Z are conditionally independent given Y , however in the third case this is not true

D-Separation

- Given a graph G , a set of variables A is conditionally independent of a set B given a set C , if there is no trajectory in G between A and B such that:
 - 1 All convergent nodes are or have descendants in C .
 - 2 All other nodes are outside C .

Markov Blanket

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- The *Markov Blanket* of a node X , $MB(X)$, is the set of nodes that make it independent of all the other nodes in G , that is $P(X | G - X) = P(X | MB(X))$
- For a BN, the Markov blanket of X is:
 - the parents of X ,
 - the sons of X ,
 - and other parents of the sons of X .

CPTs

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura

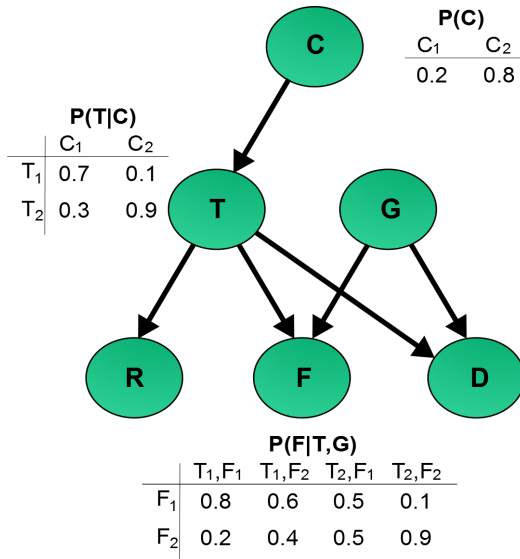
Parámetros
Inferencia

Aprendizaje

Redes
Bayesianas
Causales

- In the case of a BN, the parameters are the conditional probabilities of each node given its parents in the graph
- If we consider discrete variables:
 - Root nodes: vector of marginal probabilities.
 - Other nodes: conditional probability table (CPT) of the variable given its parents in the graph.

Example



Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Probabilistic inference

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Probabilistic inference consists in *propagating* the effects of certain evidence in a Bayesian network to estimate its effect on the unknown variables
- There are basically two variants of the inference problem in BNs:
 - *Single query inference*: obtaining the posterior probability of a single variable, H , given a subset of known (instantiated) variables, \mathbf{E} , that is, $P(H \mid \mathbf{E})$
 - *Conjunctive query inference*: consists in calculating the posterior probability of a set of variables, \mathbf{H} given the evidence, \mathbf{E} , that is, $P(\mathbf{H} \mid \mathbf{E})$

Inference algorithms:

- 1 Probability propagation (Pearl's algorithm).
- 2 Variable elimination.
- 3 Conditioning.
- 4 Junction tree.
- 5 Stochastic simulation.

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Complexity

- In the worst case the inference problem is *NP-hard* for Bayesian networks
- There are efficient (polynomial) algorithms for certain types of structures (singly connected networks)
- For other structures it depends on the connectivity of the graph.
- In many applications, the graphs are *sparse* and in this case there are inference algorithms which are very efficient

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

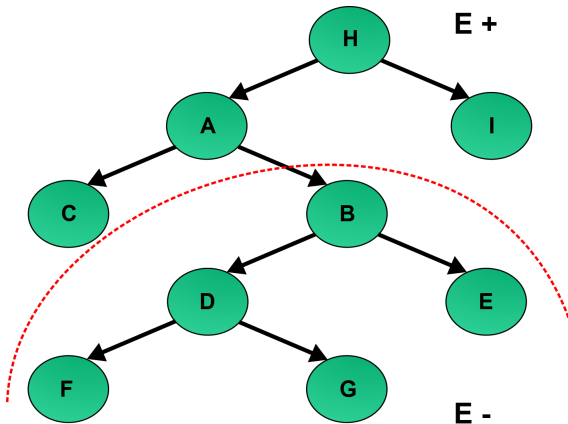
Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Probability propagation in trees

- Given that the BN has a tree structure, any node divides the network into two independent subtrees



Basic equations

- Given certain evidence, \mathbf{E} (subset of instantiated variables), the posterior probability for a value i of any variable B , can be obtained by applying the Bayes rule:

$$P(B_i|\mathbf{E}) = P(B_i)P(\mathbf{E}|B_i)/P(\mathbf{E}) \quad (5)$$

- We can separate the evidence into:
 - \mathbf{E}_- : Evidence in the tree rooted in B .
 - \mathbf{E}_+ : All other evidence.
- Then:

$$P(B_i|\mathbf{E}) = P(B_i)P(\mathbf{E}_-, \mathbf{E}_+ | B_i)/P(\mathbf{E}) \quad (6)$$

Basic equations

- Given that \mathbf{E}_+ and \mathbf{E}_- are independent, by applying the Bayes rule again, we obtain:

$$P(B_i|\mathbf{E}) = \alpha P(B_i|\mathbf{E}_+)P(\mathbf{E}_-|B_i) \quad (7)$$

Where α is a normalization constant.

- We define the following terms:

$$\lambda(B_i) = P(\mathbf{E}_-|B_i) \quad (8)$$

$$\pi(B_i) = P(B_i|\mathbf{E}_+) \quad (9)$$

- Then:

$$P(B_i|\mathbf{E}) = \alpha\pi(B_i)\lambda(B_i) \quad (10)$$

Propagation algorithm

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- The computation of the posterior probability of any node B is decomposed into two parts: (i) the evidence coming from the sons of B in the tree (λ), and the evidence coming from the parent of B , (π)
- We can think of each node B in the tree as a simple processor that stores its vectors $\pi(B)$ and $\lambda(B)$, and its conditional probability table, $P(B | A)$
- The evidence is propagated via a message passing mechanism, in which each node sends the corresponding messages to its parent and sons in the tree

Messages

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- A message sent from node B to its parent A :

$$\lambda_B(Ai) = \sum_j P(B_j | A_i) \lambda(B_j) \quad (11)$$

- A message sent from node B to its son S_k :

$$\pi_k(Bi) = \alpha \pi(B_j) \prod_{l \neq k} \lambda_l(B_j) \quad (12)$$

where l refers to each one of the sons of B

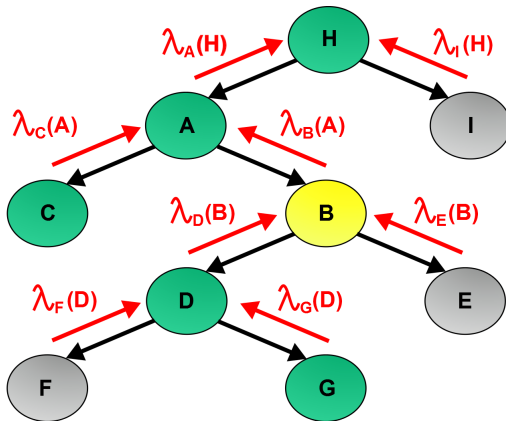
Combination and Propagation

- Each node can receive several λ messages, which are combined via a term by term multiplication for the λ messages received from each son:

$$\lambda(A_i) = \prod_{j=1}^m \lambda_{S_j}(A_i) \quad (13)$$

- The propagation algorithm starts by assigning the evidence to the known variables, and then propagating it through the message passing mechanism until the root of the tree is reached for the λ messages, and the leaves are reached for the π messages

Bottom-up propagation



Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

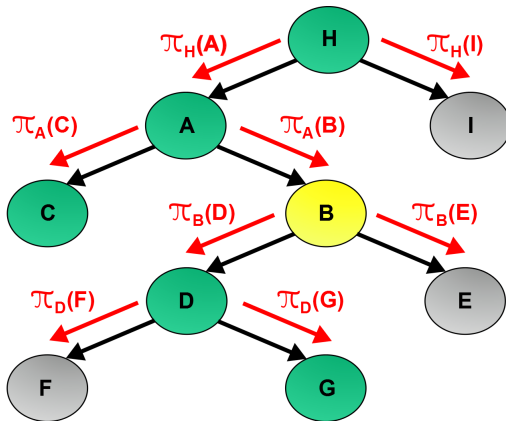
Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Top-down propagation



Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Initial Conditions

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros

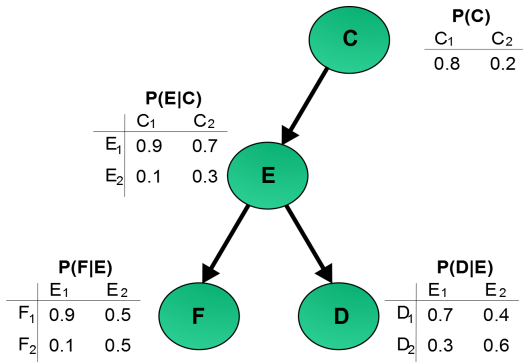
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Leaf nodes: If not known, $\lambda = [1, 1, \dots, 1]$ (a uniform distribution). If known, $\lambda = [0, 0, \dots, 1, \dots, 0]$ (one for the assigned value and zero for all other values).

Root node: If not known, $\pi = P(A)$ (prior marginal probability vector). If known, $\pi = [0, 0, \dots, 1, \dots, 0]$ (one for the assigned value and zero for all other values).

Propagation - example



- Consider that the only evidence is $F = \text{false}$ - initial conditions for the leaf nodes are:
 $\lambda_F = [1, 0]$ and $\lambda_D = [1, 1]$ (no evidence)

Example - λ propagation

- Multiplying the λ vectors by the corresponding CPTs:

$$\lambda_F(E) = [1, 0] \begin{bmatrix} 0,9, 0,5 \\ 0,1, 0,5 \end{bmatrix} = [0,9, 0,5]$$

$$\lambda_D(E) = [1, 1] \begin{bmatrix} 0,7, 0,4 \\ 0,3, 0,6 \end{bmatrix} = [1, 1]$$

- Then, $\lambda(E)$ is obtained by combining the messages from its two sons:

$$\lambda(E) = [0,9, 0,5] \times [1, 1] = [0,9, 0,5]$$

- Propagation to its parent, C :

$$\lambda_E(C) = [0,9, 0,5] \begin{bmatrix} 0,9, 0,7 \\ 0,1, 0,3 \end{bmatrix} = [0,86, 0,78]$$

Example - π propagation

- Given that C is not instantiated, $\pi(C) = [0,8, 0,2]$
- Propagate to its son, E , which also corresponds to multiplying the π vector by the corresponding CPT:

$$\pi(E) = [0,8, 0,2] \begin{bmatrix} 0,9, 0,7 \\ 0,1, 0,3 \end{bmatrix} = [0,86, 0,14]$$

- We now propagate to its son D ; however, given that E has another son, F , we also need to consider the λ message from this other son, thus:

$$\pi(D) = [0,86, 0,14] \times [0,9, 0,5] \begin{bmatrix} 0,7, 0,4 \\ 0,3, 0,6 \end{bmatrix} = [0,57, 0,27]$$

Example - posterior probabilities

- Given the λ and π vectors for each unknown variable, we just multiply them term by term and then normalize to obtain the posterior probabilities:

$$\begin{aligned}P(C) &= [0,86, 0,2] \times [0,86, 0,78] = \alpha[0,69, 0,16] \\ &= [0,815, 0,185]\end{aligned}$$

$$\begin{aligned}P(E) &= [0,86, 0,14] \times [0,9, 0,5] = \alpha[0,77, 0,07] \\ &= [0,917, 0,083]\end{aligned}$$

$$P(D) = [0,57, 0,27] \times [1, 1] = \alpha[0,57, 0,27] = [0,67, 0,33]$$

Introduction

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Learning a Bayesian network includes two aspects:
 - Structure Learning. There are two main types of methods: (i) global methods based on search and score, and (ii) local methods that use conditional independence tests
 - Parameter Learning. When the structure is known, parameter learning consists in estimating the conditional probability tables (CPTs) from data.

Parameter Learning

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- If we have *sufficient* and complete data for all the variables, and we assume the topology of the BN is known, parameter learning is straight forward
- The CPT for each variable can be estimated from the data based on the frequency of each value obtaining a *maximum likelihood* (ML) estimator
- For example, to estimate the CPT of B given it has two parents, A, C :

$$P(B_i | A_j, C_k) \sim NB_i A_j C_k / NA_j C_k \quad (14)$$

Structure learning

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Structure learning consists in obtaining the topology of the BN from the data
- This is a complex problem because: (i) the number of possible structures is *huge* even with a few variables (it is super-exponential on the number of variables); (ii) a very large database is required to obtain good estimates of the statistical measures
- There are several techniques depending on the type of structure – trees, polytrees, general DAG

General case

- For the general case several methods have been proposed, which can be divided into two main classes:
 - 1 Global methods: these perform a heuristic search over the space of network structures, starting from some initial structure, and generating a variation of the structure at each step. The *best* structure is selected based on a score that measures how well the model represents the data. Common scores are BIC and MDL
 - 2 Local methods: these are based on evaluating the (in)dependence relations between subsets of variables given the data, to sequentially obtain the structure of the network. The most well known variant of this approach is the PC algorithm

Causality

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Humans tend to think in causal terms –certain events cause other events
- Causal knowledge tends to be general, it will apply in different situations and contexts
- It helps to predict the effects of interventions (actions)
- It can be used for imagining future scenarios or what if? (counterfactuals)

Associative vs. Causal Models

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Probabilistic models do not necessarily represent causal relations
- Consider the following two directed Bayesian networks:
BN1: $A \rightarrow B \rightarrow C$ and BN2: $A \leftarrow B \leftarrow C$
- They represent the same conditional independence relations; however, both models represent very different causal relations

Causal Bayesian Networks

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

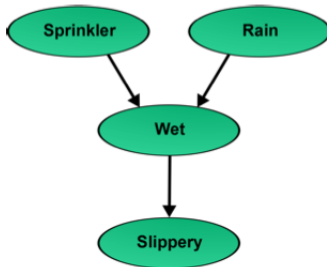
Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- A Causal Bayesian network (CBN) is a directed acyclic graph, G , in which each node represents a variable and the arcs in the graph represent causal relations
- $A \rightarrow B$ represents some physical mechanism such that the value of B is directly affected by the value of A
- This relation can be interpreted in terms of interventions –setting of the value of some variable or variables by an external agent
- Causal networks represent stronger assumptions than Bayesian networks, as all the relations implied by the network structure should correspond to causal relations

Example

- A simple example of a CBN – which basically encodes the following causal relations: (i) Sprinkler causes Wet, (ii) Rain causes Wet, (iii) Wet causes Slippery



Causal Reasoning

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Causal reasoning has to do with answering causal queries from a causal model
- There are several types of causal queries, we will consider:
 - Causal predictions
 - Counterfactuals
- Causal networks represent stronger assumptions than Bayesian networks, as all the relations implied by the network structure should correspond to causal relations

DO Operation

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- If $P(X)$ is the joint probability distribution of the set of variables X , then we define $P_y(X)$ as the distribution resulting from setting the value for a subset of variables, Y , via an intervention
- This can be represented as $do(Y = y)$, where y is a set of constants

Prediction

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Suar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Estimate the effect of an intervention:
 $P_C(X_E | do(X_C = x_C))$
- Procedure:
 - 1 Eliminate all incoming arrows in the graph to all nodes in X_C , thus obtaining a modified CBN.
 - 2 Fix the values of all variables in X_C , $X_C = x_C$.
 - 3 Calculate the resulting distribution in the modified model (via probability propagation as in Bayesian networks).

Conclusiones

- Las redes bayesianas permiten modelar problemas con incertidumbre mediante la representación de las variables de interés y sus relaciones de dependencia / independencia condicional.
- Una vez modelado un problema mediante redes bayesianas, existen algoritmos eficientes para estimar las probabilidades posteriores de alguna variables dadas otras variables (evidencia).
- Es posible aprender la estructura y parámetros de una red bayesiana a partir de datos, conocimiento experto o combinación de ambos.
- Las redes bayesianas forman parte de un conjunto de técnicas conocidas como Modelos Gráficos Probabilistas que incluyen a los clasificadores bayesianos, modelos ocultos de Markov, campos de Markov, diagramas de influencia, etc.

Conclusiones

- Existen, básicamente, otras dos alternativas bien fundamentadas para el manejo de incertidumbre: teoría de Dempster-Shafer y Lógica Difusa.
- La teoría de Dempster-Shafer permite diferenciar entre ignorancia e iguales probabilidades, aunque para problemas grandes se vuelve demasiado compleja
- La lógica difusa tiene la ventaja de que permite representar más fácilmente ciertos conceptos no bien definidos y, en particular las reglas difusas se asemejan a los sistemas basados en reglas. Sin embargo, tiene la dificultad de la falta de una semántica clara.

Referencias

- David Heckerman, Probabilistic Interpretations for MYCIN's Certainty Factors, Proceedings of the First Conference on Uncertainty in Artificial Intelligence (UAI1985)
- Russel and Norvig, Cap. 13 y 14
- Sucar, L. E., *Probabilistic Graphical Models*, Springer 2015
- Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)
- Neapolitan, R.E.: Learning Bayesian Networks. Prentice Hall, New Jersey (2004)

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

Tarea

Manejo de
Incertidumbre

Eduardo
Morales,
Enrique Sucar

Introducción

Probabilidad

Técnicas
Empíricas

Redes
Bayesianas

Estructura
Parámetros
Inferencia
Aprendizaje

Redes
Bayesianas
Causales

- Repite el ejemplo de MYCIN asumiendo que todas las evidencias tienen un CF igual a 1
- Para la red bayesiana de la lámina 50: (i) determina las relaciones de independencia condicional entre las variables; (ii) Si se sabe que el valor de $C = c_1$, calcula la probabilidad posterior de las otras variables.