V CONFERENCIA NACIONAL DE CIENCIAS DE LA COMPUTACIÓN
CCBOL'98
Del 16 al 20 de Noviembre 1998 en Potosí

Conferencia C1

MODELOS DE REDES PROBABILÍSTICAS EN SISTEMAS EXPERTOS

Prof. José Manuel Gutiérrez
Dpto. de Matemática Aplicada y Ciencias de la Computación
Universidad de Cantabria (España)
http://ccaix3.unican.es/~gutierjm

Introducción

En las dos últimas décadas se ha producido un notable desarrollo en el área de la inteligencia artificial y, en particular, en la de los sistemas expertos. Los primeros sistemas expertos utilizaron reglas para representar explícitamente el conocimiento disponible sobre un problema concreto (reglas de transacciones bursátiles, reglas de control de tráfico, etc.) y mecanismos de inferencia lógica para obtener conclusiones en base a un conocimiento concreto. La simplicidad de estos sistemas y la facilidad con que podían definirse motivaron un gran desarrollo en este área y la aparición de numerosos sistemas expertos basados en esta metodología (sistemas expertos de control de tráfico, de automatización de procesos complicados, de diagnóstico médico, etc.).

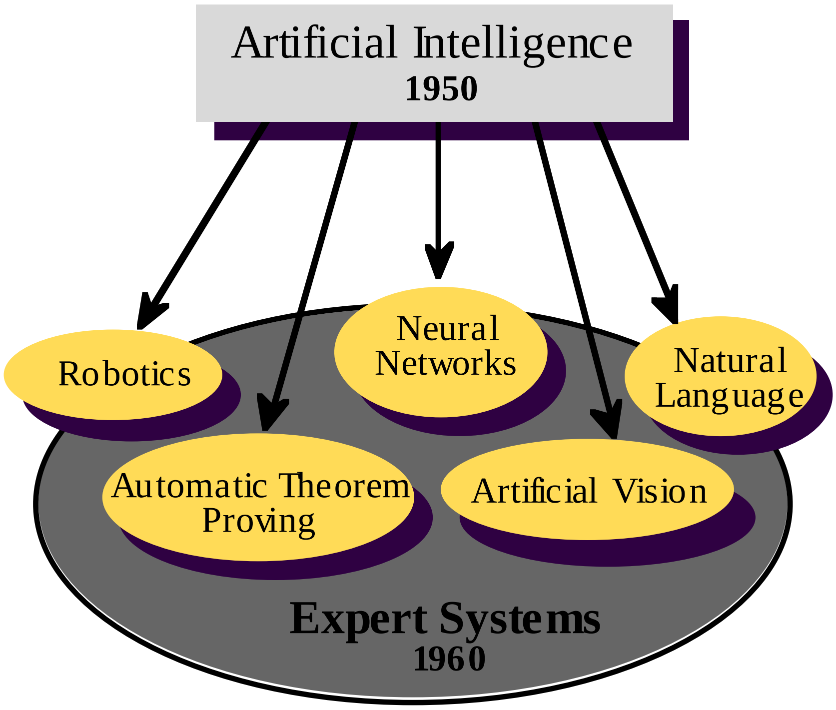
El gran inconveniente de los sistemas expertos basados en reglas es que sólo pueden aplicarse a situaciones deterministas donde las premisas de una regla se cumplen, o no. Sin embargo, hay muchas situaciones prácticas que implican incertidumbre. Los sistemas expertos basados en probabilidad son los modelos más populares para tratar estas situaciones. En este caso, el conocimiento sobre un problema dado se representa explícitamente por medio de una función de probabilidad conjunta de las variables que intervienen en el problema. El inconveniente de estos sistemas es que la definición de una función de probabilidad conjunta no es una tarea intuitiva y, por tanto, su definición no es tan simple como la de un sistema experto basado en reglas. Las redes probabilísticas (redes Bayesianas y redes de Markov) permiten definir de forma gráfica (por medio de un grafo) las relaciones de dependencia entre las variables y definir la función de probabilidad conjunta a través una factorización de funciones locales de probabilidad. Esto permite simplificar el proceso de definición del sistema experto de forma considerable. Esta característica ha hecho a estos modelos populares y ha motivado su gran difusión en los últimos años.

En esta conferencia se introducen los conceptos básicos sobre los sistemas expertos y se analizan en especial detalle los modelos de redes probabilísticas desde un punto de vista teórico y práctico.

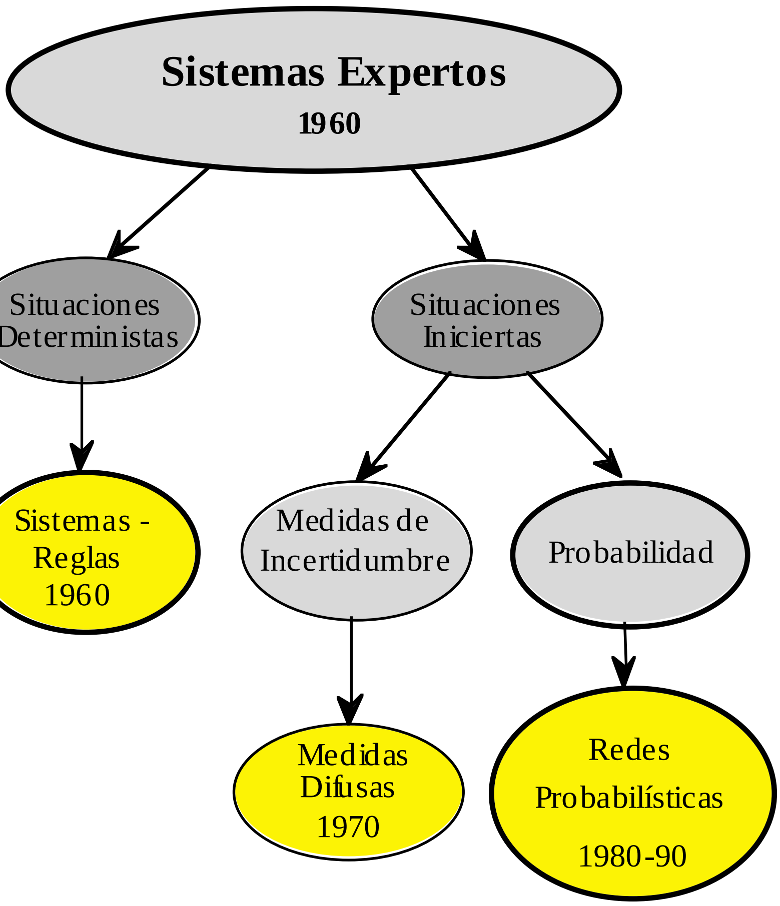
Inteligencia Artificial y Sistemas Expertos

No hace mucho tiempo, se creía que algunos problemas como la demostración de teoremas, el reconocimiento de la voz y el de patrones, ciertos juegos (como el ajedrez o las damas), y sistemas altamente complejos de tipo determinista o estocástico, debían ser resueltos por personas, dado que su formulación y resolución requieren ciertas habilidades que sólo se encuentran en los seres humanos (por ejemplo, la habilidad de pensar, observar, memorizar, aprender, ver, oler, etc.). Sin embargo, el trabajo realizado en las tres últimas décadas por investigadores procedentes de varios campos, muestra que muchos de estos problemas pueden ser formulados y resueltos por máquinas.

El amplio campo que se conoce como *inteligencia artificial* (IA) trata de estos problemas, que en un principio parecían imposibles, intratables y difíciles de formular utilizando ordenadores. Hoy en día, el campo de la IA engloba varias subáreas tales como los sistemas expertos, la demostración automática de teoremas, el juego automático, el reconocimiento de la voz y de patrones, el procesamiento del lenguaje natural, la visión artificial, la robótica, las redes neuronales, etc (una revisión de los campos que componen la IA se puede encontrar en Castillo, Gutiérrez y Hadi, 1997). Aunque los sistemas expertos constituyen una de las áreas de investigación en el campo de la IA, la mayor parte de las restantes áreas, si no todas, disponen de una componente de sistemas expertos formando parte de ellas.



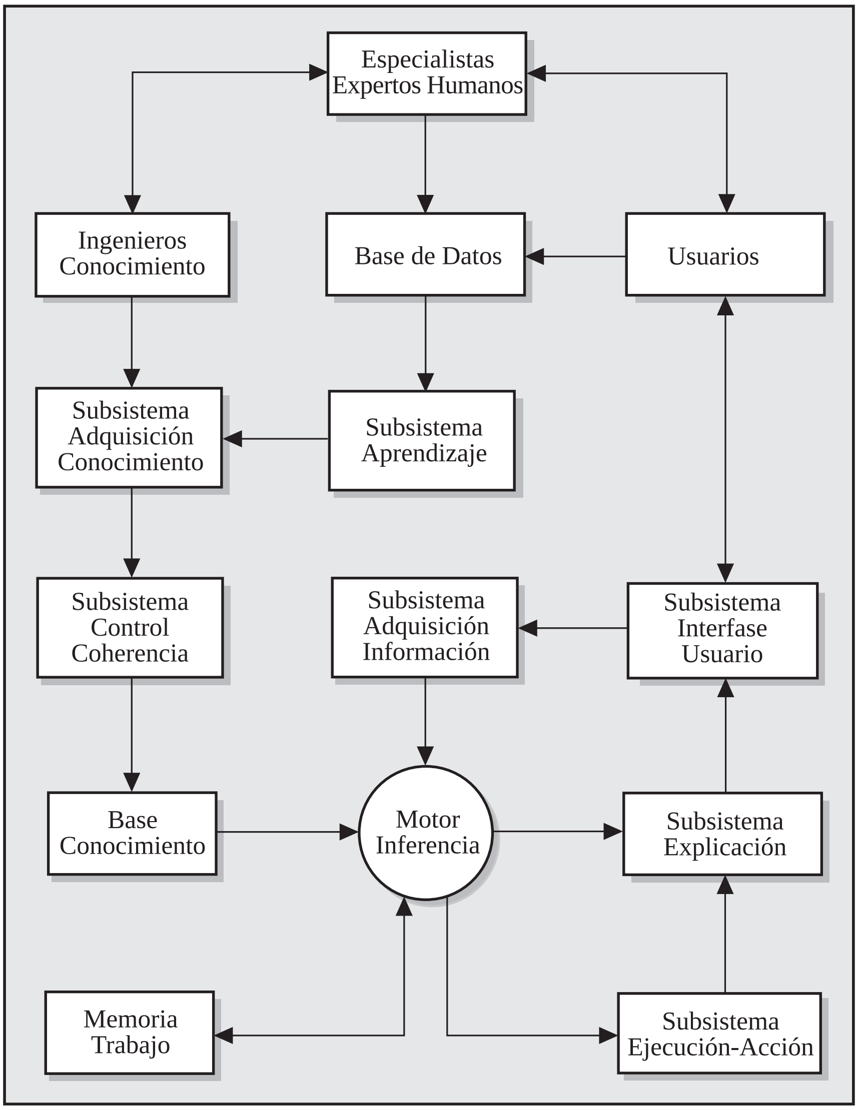
En esta conferencia se trata de introducir un nuevo tipo de sistemas expertos que utilizan grafos para representar el conocimiento y probabilidad para representar las incertidumbre. Estos sistemas se conocen como redes probabilísticas.



Componentes de un Sistema Experto

Los sistemas expertos son máquinas que piensan y razonan como un experto lo haría en una cierta especialidad o campo. Por ejemplo, un sistema experto en diagnóstico médico requeriría como datos los síntomas del paciente, los resultados de análisis clínicos y otros hechos relevantes, y, utilizando éstos, buscaría en una base de datos la información necesaria para poder identificar la correspondiente enfermedad. Un Sistema Experto de verdad, no sólo realiza las funciones tradicionales de manejar grandes cantidades de datos, sino que también manipula esos datos de forma tal que el resultado sea inteligible y tenga significado para responder a preguntas incluso no completamente especificadas.

La siguiente figura ilustra los distintos componentes de un sistema experto



Tipos de Sistemas Expertos

Los problemas con los que pueden tratar los sistemas expertos pueden clasificarse en dos tipos: problemas esencialmente deterministas y problemas esencialmente estocásticos. Por ejemplo, en el campo médico las relaciones entre síntomas y enfermedades se conocen sólo con un cierto grado de certeza (la presencia de un conjunto de síntomas no siempre implica la presencia de una enfermedad). Estos tipos de problemas pueden también incluir algunos elementos deterministas, pero se trata fundamentalmente de problemas estocásticos.

Consecuentemente, los sistemas expertos pueden clasificarse en dos tipos principales según la naturaleza de problemas para los que están diseñados: deterministas y estocásticos. Los problemas de tipo determinista pueden ser formulados usando un conjunto de reglas que relacionen varios objetos bien definidos. Los sistemas expertos que tratan problemas deterministas son conocidos como *sistemas basados en reglas*.

En situaciones inciertas, es necesario introducir algunos medios para tratar la incertidumbre. Por ejemplo, algunos sistemas expertos usan la misma estructura de los sistemas basados en reglas, pero introducen una medida asociada a la incertidumbre de las reglas y a la de sus premisas. En este caso se pueden utilizar algunas fórmulas de propagación para calcular la incertidumbre asociada a las conclusiones. Durante las últimas décadas han sido propuestas algunas medidas de incertidumbre. Algunos ejemplos de estas medidas son los *factores de certeza*, usados en las conchas para generar sistemas expertos tales como el sistema experto MYCIN; la *lógica difusa*, etc.

Otra medida intuitiva de incertidumbre es la *probabilidad*, en la que la distribución conjunta de un conjunto de variables se usa para describir las relaciones de dependencia entre ellas, y se sacan conclusiones usando fórmulas muy conocidas de la teoría de la probabilidad. Este es el caso del sistema experto PROSPECTOR, que utiliza el teorema de Bayes para la exploración de mineral.

Los sistemas expertos que utilizan la probabilidad como medida de incertidumbre se conocen como *sistemas expertos probabilístico* y la estrategia de razonamiento que usan se conoce como *razonamiento probabilístico*, o *inferencia probabilística*.

Sistemas Expertos Probabilísticos

En los primeros sistemas expertos, se eligió la probabilidad como medida para tratar la incertidumbre pero, desgraciadamente, muy pronto se encontraron algunos problemas, debidos al uso incorrecto de algunas hipótesis de independencia, utilizadas para reducir la complejidad de los cálculos. Como resultado, en las primeras etapas de los sistemas expertos, la probabilidad fue considerada como una medida de incertidumbre poco práctica. La mayoría de las críticas a los métodos probabilísticos se basaban en el altísimo número de parámetros necesarios, la imposibilidad de una asignación o estimación precisa de los mismos, o las hipótesis poco realistas de independencia.

Consecuentemente, en la literatura de la época, surgieron medidas alternativas a la probabilidad, como los factores de certeza, las credibilidades, las plausibilidades, las necesidades o las posibilidades, para tratar la incertidumbre. Sin embargo, con la aparición de las redes probabilísticas (principalmente las redes Bayesianas y Markovianas, la probabilidad ha resurgido de forma espectacular, y es, hoy en día, la más intuitiva y la más aceptada de las medidas de incertidumbre.

Consideremos el siguiente ejemplo de diagnóstico médico, donde se supone que se tiene un conjunto de enfermedades dado (E1,...,En) y un conjunto de síntomas asociados (S1,...,Sm) a éstas. La pregunta a la que trata de responder un sistema experto probabilístico de diagnóstico médico es: *Supuesto que el paciente presenta un subconjunto de síntomas S, ¿qué enfermedad es más probable que tenga?*. Para responder a esta pregunta han de calcularse las probabilidades $P(Ei \mid S)$, $i = 1, 2, \ldots, n$. Por ejemplo, en un caso práctico podíamos haber obtenido:

Enfermedad	P(Di | S)	
1	0.2	
2	0.1	
3	0.8	→ más probable
4	0.4	
5	0.0	→ menos probable
6	0.7	

El problema de estos sistemas es que suponen que se conoce la función de probabilidad conjunta de todas las enfermedades y síntomas. Sin embargo, en la práctica no se conocen todos los datos necesarios para definir la función de probabilidad conjunta pues, en casos reales, estos pueden constituir una cantidad ingente de información. Por ejemplo, para un caso de diagnóstico médico con 100 enfermedades binarias y 100 síntomas binarios se requieren más de 10^{60} parámetros para especificar la función de probabilidad conjunta.

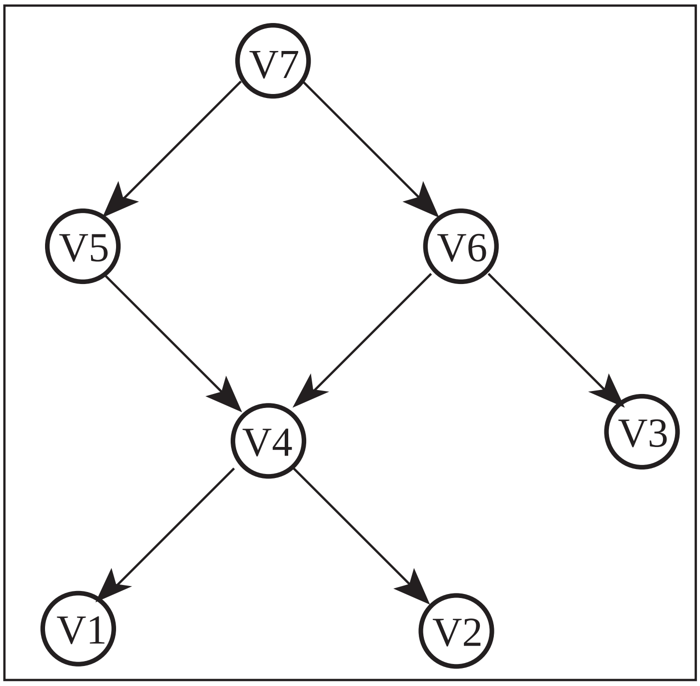
Los modelos de redes probabilísticas utilizan grafos para definir relaciones de dependencia entre las variables del modelo y simplificar la estructura de la función de probabilidad conjunta. Así, serán necesario un número inferior de parámetros para especificar el modelo. Esta simplificación se realiza en base a una factorización de la probabilidad.

Dado un conjunto de variables (V1,...,Vn), una función de probabilidad siempre puede escribirse como un producto de funciones de probabilidad condicionada:

P(V1, …, Vn) = P(V1) P(V2 | V1) P(Vi | V1, …, Vi–1) P(Vn | V1, …, Vn–1)

Redes Probabilísticas. Redes Bayesianas

Una red Bayesiana utiliza un grafo dirigido para representar las relaciones entre las variables del modelo. Por ejemplo, el siguiente grafo representa las relaciones existentes en un conjunto de siete variables



En base a estas relaciones, la función de probabilidad del modelo puede escribirse como el producto de las probabilidades de cada uno de los nodos condicionado a sus padres:

P(V1, V2, V3, V4, V5, V6, V7)=

P(V1 | V4) P(V2 | V4) P(V3 | V6) P(V4 | V5,V6) P(V5 | V7) P(V6 | V7) P(V7)

Estas funciones de probabilidad condicionada vienen determinadas directamente por las relaciones de independencia contenidas en el grafo (la notación I(A, B | C) significa que los conjuntos de variables A y B son condicionalmente independientes una vez que se conoce el conjunto C).

P(1 | 4)	I(1; 23567 | 4)
P(2 | 4)	I(2; 13567 | 4)
P(3 | 6)	I(3; 12457 | 6)
P(4 | 5,6)	I(4; 7 | 5,6)
P(5 | 7)	I(5; 7 | ∅)
P(6 | 7)	I(6; 7 | ∅)
P(7)	I(7; ∅ | ∅)

El criterio gráfico que se utiliza para saber que relaciones de independencia condicional están contenidas en un grafo se llama criterio de *d-separación.*

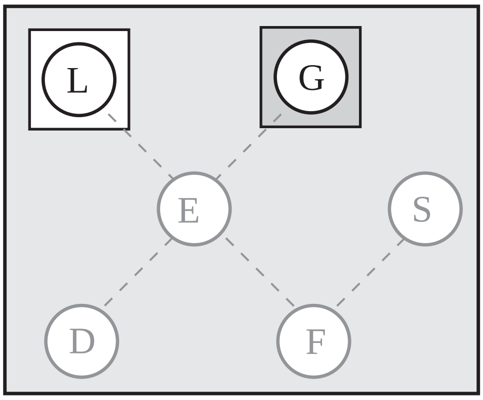
Criterio de d-separación

Sean X, Y y Z tres subconjuntos disjuntos de nodos en un grafo dirigido acíclico; entonces se dice que Z d-separa X e Y si y sólo si a lo largo de todo camino no dirigido entre cualquier nodo de X y cualquier nodo de Y existe un nodo intermedio A tal que, o bien:

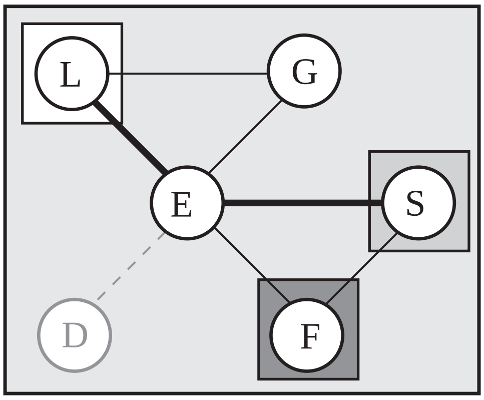
- A es un nodo de aristas convergentes en el camino y ni A ni sus descendientes están en Z, o bien
- A no es un nodo de aristas convergentes en el camino y A está en Z.

Una definición equivalente se muestra a continuación:
Sean X, Y y Z tres subconjuntos disjuntos en un grafo dirigido acíclico, entonces se dice que Z D-separa a X e Y si y sólo si Z separa X e Y en el grafo moral (grafo no dirigido obtenido añadiendo aristas entre nodos con hijos comunes) del menor subconjunto ancestral que contenga a los nodos de X, Y y Z.

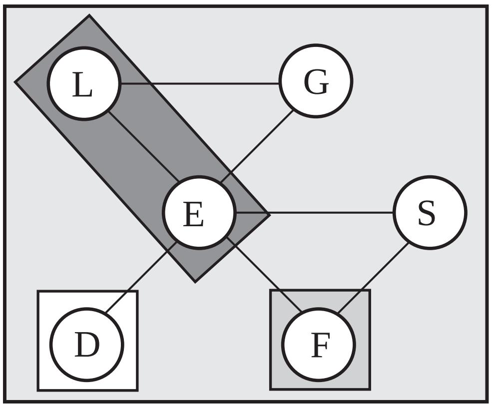
Esta definición alternativa fué propuesta por Lauritzen y otros.



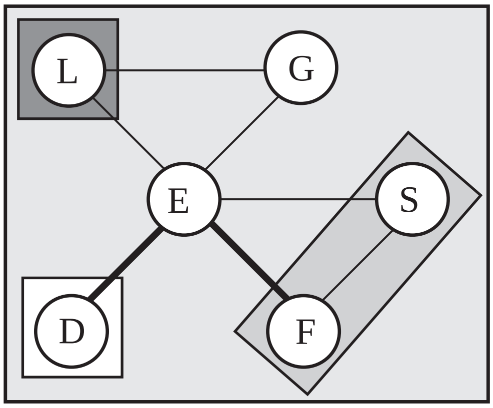
(a) $I(L, G \mid Ø)$



(b) $D(L, S \mid F)$



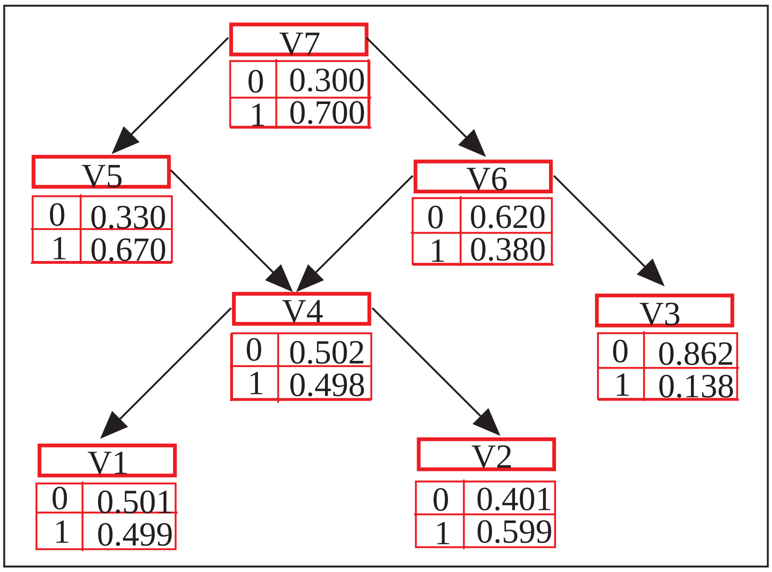
(c) $I(D, F \mid \{L, E\})$



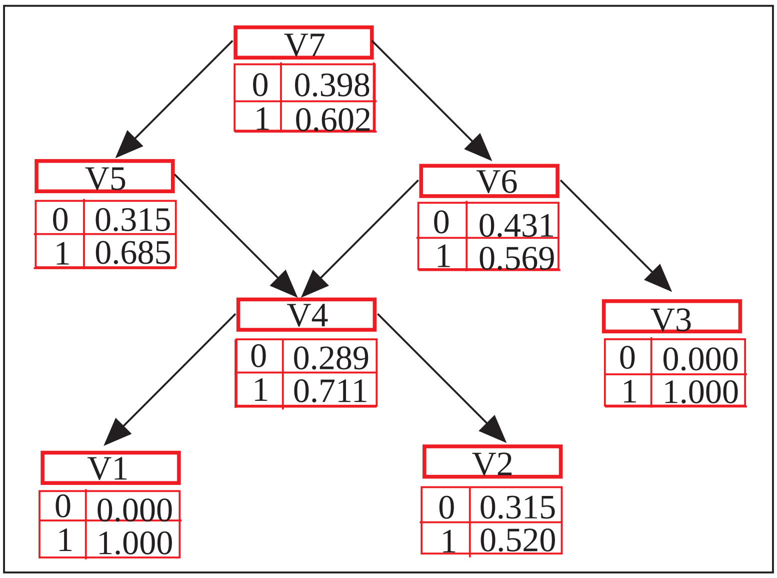
(d) $D(D, \{S, F\} \mid L)$

Propagación de Evidencia

Una vez definida la red Bayesiana, el proceso de proagación de evidencia consiste en el cálculo de las funciones de probabilidad de los nodos condicionados a la evidencia obervada. Inicialmente, cuando no se conoce evidencia, estas probabilidades son las probabilidades marginales del modelo:



Sin embargo, cuando se oberva cierta evidencia (supongamos que V1=1 y V3=1), las nuevas probabilidades condicionadas muetran el efecto de la evidencia en el resto de los nodos. Este proceso se denomina propagación de evidencia pues ésta se propaga por el grafo acutalizando las probabilidades de los nodos.



Bibliografía

Castillo, E., Gutiérrez, J.M. and Hadi, A.S. (1997*) Expert Systems and Probabilistic Network Models*. Springer Verlag, New York. Versión castellana publicada por la Academia de Ingeniería (1998)

Jensen, F.V. (1996) *An Introduction to Bayesian Networks.* Springer-Verlag, New York.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.