TESIS

Sistema experto bayesiano para ecocardiografía

Francisco Javier Díez Vegas

Licenciado en Ciencias Físicas
por la Universidad Autónoma de Madrid

Presentada en la

FACULTAD DE CIENCIAS

de la

UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA

para la obtención

del grado de

Doctor en Ciencias Físicas

Madrid, 1994

TESIS DOCTORAL

SISTEMA EXPERTO BAYESIANO PARA ECOCARDIOGRAFÍA

Por: Francisco Javier Díez Vegas

Director de la Tesis: D. José Mira Mira

Tribunal calificador

Presidente: D. Sebastián Dormido Bencomo

Vocales: D. Álvaro Barreiro García

D. Enrique Castillo Ron

D. Serafín Moral Callejón

Secretaria: Dª Ana E. Delgado García

Vocal suplente: D. Manuel Yuste Llandres

Madrid, 12 de Julio de 1994

A mis padres

Resumen y objetivos

Las redes bayesianas (RR.BB.) constituyen una metodología para la construcción de sistemas expertos, que surgió en la década de los 80 y se ha expandido notablemente en los últimos seis años. Su principal cualidad es que realizan el tratamiento de la incertidumbre a partir de un modelo causal probabilístico.

Esta tesis tiene dos objetivos principales. El primero, de naturaleza teórica, consiste en un estudio de las redes bayesianas. El segundo es la aplicación de las ideas desarrolladas a la construcción de un sistema experto para diagnóstico mediante ecocardiografía, denominado DIAVAL.

La introducción de esta memoria presenta brevemente la inteligencia artificial desde un punto de vista personal, insistiendo en la representación del conocimiento y en el tratamiento de la incertidumbre. También mencionamos los errores que suele cometer el ser humano en la estimación subjetiva de la incertidumbre. Con ello, intentamos justificar por qué hemos utilizado una red bayesiana en vez de construir un sistema experto estándar, es decir, basado en reglas.

La parte II realiza un estudio completo de las RR.BB., de forma sistemática, analizando para cada uno de sus aspectos la **teoría**, los **algoritmos** y la **implementación**, con especial interés por las implementaciones distribuidas.

En cuanto a la propagación de la evidencia, la presente memoria introduce la *puerta OR graduada*, una extensión de la puerta OR binaria que simplifica la adquisición del conocimiento y ahorra tiempo de computación en la inferencia, y el *condicionamiento local*, un nuevo algoritmo para el cálculo de la probabilidad en redes con bucles, que supera ampliamente a los métodos de condicionamiento anteriores y es capaz de competir con los algoritmos más utilizados hoy en día.

Proponemos también un *modelo de aprendizaje secuencial* (ajuste de parámetros) basado en distribuciones gaussianas para las probabilidades, y un método de *explicación del razonamiento*, que distingue dos tipos de interacción (caso general y puerta OR) y seis clases de enlaces. Concluimos la parte II con una valoración de las RR.BB.: historia, contexto, cualidades y limitaciones.

La parte III describe la construcción del sistema experto DIAVAL, según las cinco etapas clásicas: identificación, conceptualización, formalización, implementación y evaluación. Hemos intentado estudiar con detalle la obtención del conocimiento en sistemas bayesianos, desde el campo de la medicina. En cuanto a la implementación de DIAVAL, hemos puesto el mayor esmero en dotar al sistema experto de un interfaz flexible y fácil de manejar, conociendo la importancia de este factor en su aceptación por parte de los médicos. La evaluación del sistema ha sido satisfactoria, aunque todavía es necesario mejorar el programa antes de su implantación definitiva.

La memoria concluye resumiendo las lecciones aprendidas y las principales aportaciones, valorando los logros y limitaciones de DIAVAL, y señalando las líneas de investigación que quedan abiertas.

Índice

Agradecimientos xv
Nota sobre el uso del idioma xvii
Lista de figuras xix
Lista de tablas xxi

Parte I INTRODUCCIÓN **1**

1 Presentación **3**
1.1 Proyecto para ecocardiografía 3
1.2 Desarrollo, metodología y objetivos 5
1.3 Organización de la tesis 6
1.4 Guía de lectura 9

2 Fundamentos de inteligencia artificial **11**
2.1 Inteligencia artificial y sistemas expertos 11
2.1.1 Orígenes históricos 11
2.1.2 Ramas y tendencias en I.A. 13
2.1.3 Sistemas expertos. Definición y estructura 15
2.1.4 Ejemplos clásicos de sistemas expertos 17
2.1.5 Ventajas y limitaciones de los sistemas expertos 21
2.1.6 Juicio crítico 23
2.2 Representación del conocimiento y razonamiento 23
2.2.1 Lógica 24
2.2.2 Reglas 25
2.2.3 Redes asociativas 26
2.2.4 Marcos 27
2.2.5 Sistemas híbridos 28
2.3 Tratamiento de la incertidumbre 28
2.3.1 Método probabilístico clásico 29
2.3.2 Factores de certeza (MYCIN) 32
2.3.3 Lógica difusa 33
2.3.4 Otros métodos 35
2.4 Crítica de los sistemas basados en reglas 36
2.4.1 Representación del conocimiento 36
2.4.2 Tratamiento de la incertidumbre 37
2.4.3 Conclusión 40

3 Estimación subjetiva de la incertidumbre **43**
3.1 Errores en la estimación de la probabilidad 44
3.1.1 Estudios psicológicos 44
3.1.2 Errores en medicina 46
3.1.3 Discusión 48
3.2 Expresiones cualitativas de la probabilidad 48
3.3 Conclusión 49
Apéndice: Prevalencia, sensibilidad y especificidad 50

Parte II REDES BAYESIANAS **53**

4 Fundamentos de redes bayesianas **55**
4.1 Presentación intuitiva 55
4.2 Definición formal de red bayesiana 68
4.2.1 Estructura de la red. Teoría de grafos 68
4.2.2 Nociones de probabilidad 70
4.2.3 Definición de red bayesiana 74
4.3 Semántica de las redes bayesianas 76

5 Propagación de evidencia en poliárboles **79**
5.1 Caso general 79
5.1.1 Definiciones básicas 80
5.1.2 Computación de los mensajes 81
5.1.3 Comentarios 83
5.1.4 Implementación distribuida 85
5.2 La puerta OR 88
5.2.1 La puerta OR binaria 88
5.2.2 La puerta OR graduada 89
5.2.3 Algoritmo de propagación 93
5.2.4 Implementación distribuida 96
5.2.5 Semántica 98
5.2.6 Comentarios 100

6 Tratamiento de los bucles **101**
6.1 Métodos previos 102
6.1.1 Condicionamiento 102
6.1.2 Agrupamiento 104
6.1.3 Simulación 107
6.2 Condicionamiento local 108
6.2.1 Poliárbol asociado y asignación de bucles 108
6.2.2 Propagación de evidencia 112
6.2.3 Eficiencia del algoritmo 116
6.2.4 Ventajas del condicionamiento local 119
Apéndice: Complejidad de algoritmos 123

7 Aprendizaje en redes bayesianas **127**
7.1 Nociones básicas 127
7.1.1 Tipos de aprendizaje 127
7.1.2 Crítica del aprendizaje por lotes 128
7.2 Modelo de aprendizaje secuencial 129
7.2.1 Introducción 129
7.2.2 Bases del modelo 129
7.2.3 Propiedades estadísticas 132
7.2.4 El caso general 133
7.2.5 La puerta OR 136
7.2.6 Implementación distribuida 138
7.2.7 Discusión 139
Apéndice: Demostraciones 141

8 Explicación del razonamiento **147**
8.1 Planteamiento del problema 147
8.1.1 Importancia 147
8.1.2 Cuestiones básicas 148
8.1.3 Tipos de explicación 149
8.2 Métodos previos de explicación 150
8.2.1 Sistemas heurísticos 150
8.2.2 Sistemas probabilísticos 152
8.3 Modelo de explicación 154
8.3.1 Clases de enlaces 154
8.3.2 Opciones para nodos, parámetros y enlaces 156
8.3.3 Implementación distribuida 160
8.3.4 Discusión 161

9 Implementación del razonamiento bayesiano **163**
9.1 Características del Lisp 163
9.1.1 Programación en Lisp 163
9.1.2 Entornos de programación 166
9.1.3 Ventajas e inconvenientes 167
9.2 Introducción al GoldWorks 167
9.2.1 Marcos e instancias 168
9.2.2 Programación orientada a objetos 169
9.2.3 Motor de inferencia: reglas 170
9.2.4 Interfaces 172
9.2.5 Valoración 173
9.3 Razonamiento bayesiano 175
9.3.1 Representación de la red 175
9.3.2 Propagación de evidencia 175
9.3.3 Aprendizaje 176
9.3.4 Explicación 176

10 Valoración de las redes bayesianas **179**
10.1 Orígenes y evolución 179
10.1.1 Desarrollo histórico 179
10.1.2 Investigación actual en RR.BB. 180
10.1.3 Modelos afines 181
10.1.4 Aplicaciones 183
10.2 Las redes bayesianas en el contexto de la I.A. 184
10.2.1 Ramas y tendencias de la I.A. 184
10.2.2 Diagnóstico 185
10.2.3 Computación neuronal 185
10.3 Cualidades 187
10.3.1 Fundamento normativo 187
10.3.2 Razonamiento causal 187
10.3.3 Tratamiento de independencias y correlaciones 188
10.3.4 Representación estructurada 189
10.3.5 Procesamiento distribuido 189
10.4 Limitaciones 189
10.4.1 Complejidad de los algoritmos 189
10.4.2 Obtención del conocimiento 190
10.4.3 Razonamiento temporal 191
10.4.4 Limitación en el rango de aplicaciones 192
10.4.5 Control del razonamiento 192
10.4.6 Explicación 193
10.5 Conclusión 193

Parte III APLICACIÓN A LA CARDIOLOGIA **195**

11 Desarrollo del sistema experto DIAVAL **197**
11.1 Fases de desarrollo 197
11.2 Identificación del problema 200
11.2.1 Clases de enlaces 200
11.2.2 Objetivos: funcionalidades del sistema 201
11.2.3 Viabilidad y oportunidad del proyecto 201
11.3 Obtención del conocimiento 203
11.3.1 Metodología 204
11.3.2 Dificultades inherentes al campo de la medicina 206
11.3.3 Factores humanos 207

12 Conocimiento médico cardiológico **209**
12.1 Proceso de razonamiento en medicina 209
12.1.1 Diagnóstico, pronóstico y tratamiento 209
12.1.2 Conceptos básicos sobre diagnóstico 210
12.2 Fisiopatología cardíaca 212
12.2.1 Anatomía y fisiología 212
12.2.2 Valvulopatías 212
12.2.3 Cavidades, pericardio, arterias y venas 214
12.2.4 Isquemia 214

12.2.5 Enfermedades del miocardio 215
12.2.6 Malformaciones congénitas 216
12.3 Técnicas de diagnóstico 216
12.3.1 Datos clínicos 217
12.3.2 Pruebas complementarias 218
12.3.3 Ecocardiograma 220
12.4 Técnicas terapéuticas 222

13 Implementación de DIAVAL **225**
13.1 Red bayesiana cardiológica 225
13.1.1 Contenido de la red 225
13.1.2 Construcción de la red 226
13.2 Propiedades del interfaz de usuario 227
13.2.1 Características deseables 227
13.2.2 Descripción general del interfaz 229
13.2.3 Entrada de datos 231
13.3 Presentación de los resultados 235
13.3.1 Diagnóstico 235
13.3.2 Explicación del razonamiento 238
13.3.3 Informe final 239
13.4 Entorno de apoyo 241

14 Evaluación de DIAVAL **243**
14.1 Introducción 243
14.2 Evaluación realizada 245
14.2.1 Método de evaluación 245
14.2.2 Resultados obtenidos 248
14.2.3 Discusión 250
14.3 Propuesta de evaluación objetiva 251

Parte IV CONCLUSIONES **253**

15 Resultados obtenidos **255**
15.1 Lecciones aprendidas 255
15.2 Principales aportaciones 256
15.2.1 Aportaciones a las redes bayesianas 256
15.2.2 Construcción de un sistema experto bayesiano 257
15.3 Valoración del sistema experto DIAVAL 258
15.3.1 Logros 258
15.3.2 Limitaciones 259
15.4 Futuro desarrollo 260
15.4.1 Razonamiento bayesiano 260
15.4.2 Sistemas expertos 262
15.5 Resumen de conclusiones 264

BIBLIOGRAFÍA **265**

Agradecimientos

Son muchas las personas a las que debo estar agradecido, por múltiples motivos. Entre quienes me han ayudado en la realización del doctorado, doy las gracias en primer lugar al Prof. José Mira, director de la tesis, porque además de ofrecerme sus orientaciones y su apoyo en todo momento, me ha dejado la libertad suficiente para investigar por caminos diferentes de los que inicialmente habíamos proyectado.

Doy las gracias también al Prof. Sebastián Dormido, director del Dpto. de Informática y Automática de la UNED, por la forma en que se interesa por todas y cada una de las personas que están cerca de él; cuantas veces he necesitado de él por los motivos más variados, siempre he podido contar con su ayuda. El agradecimiento es extensivo a todos los demás miembros del Departamento, por haber creado un ambiente de trabajo tan agradable. Me siento especialmente afortunado por tener excelentes compañeros de doctorado; la amistad que ha surgido entre nosotros es uno de los frutos más valiosos de esta tesis.

También he tenido la suerte de contar con un estupendo grupo de médicos que tanto me ha ayudado en la realización del proyecto: Río Aguilar, Antonio Baño, Enrique González, Elena Iturralde, Javier Jiménez, Mónzer Khanji, Íñigo Lozano, Javier Lozano, Gregorio Rodríguez, Sonia Rodríguez y Susana Zubillaga, casi todos ellos del Hospital de La Princesa, de Madrid. Les agradezco, además de la paciencia que han tenido conmigo, todo lo que me han enseñado sobre medicina y sobre calidad humana en el trato a los pacientes. También he conocido enfermeras encantadoras. Dudo mucho que algún día un sistema experto pueda sustituir a un médico —espero que nunca llegue ese día—; de lo que estoy seguro es de que nunca un ordenador podrá sustituir a una enfermera como Marisa en las mil tareas que realiza cada mañana, incluyendo la de atender a cierto "tísico" embarcado en un proyecto interminable.

Estoy agradecido al Ministerio de Educación y Ciencia por la beca predoctoral que me concedió, y a la Comisión Interministerial de Ciencia y Tecnología por la financiación de los proyectos TIC-88-315, TIC-92-136 y TIC-94-0095, en los que he participado bajo la dirección del Prof. José Mira, los cuales nos han aportado el apoyo económico necesario para desarrollar la investigación reflejada en esta memoria.

También doy las gracias al Ministerio de Educación y Ciencia por la beca que me permitió pasar tres meses en la Universidad de California Los Angeles. Dicho viaje surgió a partir de una invitación del Prof. Judea Pearl, a quien agradezco los medios que me facilitó para que mi estancia allí resultara sumamente provechosa. Por el mismo motivo debo agradecer profundamente a Moisés Goldzsmidt todo lo que me ayudó, tanto dentro como fuera de la UCLA, y también a David Heckerman por la cordialidad con que apoyó mi trabajo.

Otra estancia que me resultó muy provechosa fue la visita a la Universidad de Santiago de Compostela, en Enero de 1990, donde Diego Cabello, Roque Marín y los demás miembros de su Departamento me atendieron estupendamente.

Hay numerosos investigadores que con sus críticas, comentarios y sugerencias han enriquecido el contenido de esta tesis. Además de varias personas ya citadas, doy las gracias a Enrique Castillo, Luis Miguel Castejón, Marek Druzdzel, Serafín Moral,

Richard Neapolitan, Mark Peot, Santiago Sandoval, Ross Shachter, Sampath Srinivas y José Ramón Velasco. A muchos de ellos debo agradecerles los trabajos que me han enviado, y por el mismo motivo doy las gracias a Steen Andreassen, Riccardo Bellazzi, Bruce D'Ambrosio, Peter Haddawy, Charles Kahn, Uffe Kjærulff, Augustine Kong, Peter Sember, Jonathan Stillman, Jaap Suermondt, Chus Taboada, Ingrid Zuckerman y a los investigadores de la Universidad de Aalborg (Dinamarca), que han hecho públicos sus trabajos mediante una red de ordenadores. También obtuve numerosas referencias en una visita al *Knowledge Systems Laboratory*, de la Universidad de Stanford, y en otra visita a la Universidad Politécnica de Madrid, por la que doy las gracias a los profesores José Cuena y Juan Pazos.

El rumbo de investigación que tomó esta tesis se debe a un accidente de motocicleta que sufrí el 20 de Octubre de 1990. Durante mi convalecencia en Burgos, pude "perder el tiempo" dando vueltas a unos modelos de probabilidad, sin saber que ya estaban inventados. Pero el esfuerzo mereció la pena y dio más tarde sus frutos. Por eso doy las gracias a la chica que me atropelló y al casco que me salvó la vida.

He recibido otros muchos apoyos menos relacionados con el contenido de esta tesis, pero igualmente valiosos para su realización. En primer lugar, doy las gracias a D. Estanislao Pérez Ubierna, hombre singular, que con la fundación de una beca para "estudiantes burgaleses de física o química", me dio la posibilidad de continuar mis estudios en Madrid, lo cual me hubiera resultado sumamente difícil sin su ayuda. Él, que era además un gran aficionado a la medicina, habría disfrutado al ver la culminación de esta tesis, en la que había puesto tanta ilusión. Le pido disculpas por no haber terminado unos meses antes, aunque estoy seguro de que él se alegra ahora más aún que si hubiera estado entre nosotros.

Llega el momento de mencionar a mis padres, a quienes está dedicada esta tesis. Es tanto lo que les debo que resulta difícil encontrar las palabras adecuadas. Con el paso del tiempo, voy aprendiendo a valorar sus innumerables sacrificios para que mis hermanos y yo pudiéramos estudiar con toda libertad, en el ambiente más favorable posible, incluso en los momentos de mayor escasez. Por todo el cariño que nos han dado, por su ejemplo de trabajo, de entrega y de espíritu de superación, me siento muy feliz de poder ofrecerles como pequeña recompensa esta tesis, que es fruto de su esfuerzo más que del mío.

Doy las gracias también a todos los profesores que he tenido, en particular a Tino Barriuso, burgalés multidimensional, físico y poeta, maestro y amigo. Él es el "culpable" de que yo hiciera esta carrera e incluso de que iniciara este doctorado. Su ejemplo y su confianza en mí han significado un estímulo inigualable desde el día en que le conocí.

Casi terminando, quiero recordar a todos mis amigos, y pedir disculpas a los que han soportado los largos meses de encierro al final de este doctorado: por los cumpleaños que no he felicitado, por las cartas que no he escrito, por los cafés que no hemos compartido, por las noches de salsa en que he estado ausente... Menos mal que todo llega a su fin.

Sería injusto concluir sin mencionar a quien merece el mayor agradecimiento, pues todo, incluso lo que recibimos por medio de otros, procede de Él. La lista de beneficios desborda toda imaginación; y entre los primerísimos lugares de esa lista interminable, debo darle gracias muy especialmente por la cantidad de personas extraordinarias que me ha permitido encontrar en todas las etapas de mi vida.

Nota sobre el uso del idioma

Como burgalés que soy, me siento especialmente obligado a cuidar con esmero la lengua que nació en mi tierra. Ante todo, voy a intentar que la tesis esté escrita en *español*, una de las lenguas más hermosas que existen, y no en "*espanglish*", que es un monstruo híbrido y malsonante. No se trata de defender el inmovilismo lingüístico; al contrario, esta memoria utiliza numerosos neologismos. Lo importante es que éstos se introduzcan sólo cuando sea necesario y se haga teniendo en cuenta criterios estéticos. Aunque seamos "de ciencias" no tenemos derecho a maltratar la lengua de San Juan de la Cruz y de Cervantes.

Muchos de los atentados que se cometen contra nuestro idioma se deben al uso continuo de bibliografía anglosajona. Basta observar la lista de referencias incluida en esta tesis para comprobar que la mayor parte de los términos que utilizamos proceden del inglés. Pero debemos hacer un esfuerzo por buscar la mejor traducción y no la primera que nos venga a la mente. En este sentido, uno de los mayores peligros que nos amenazan son los "falsos amigos". En informática, por ejemplo, el uso ha convertido ya en norma hablar de *librerías de programas* a pesar de que la traducción correcta de *library* es *biblioteca*, que —naturalmente— es más adecuada para denominar un archivo donde las rutinas no se compran, sino que sólo se consultan. Luchando contra esta tendencia, en la presente memoria no *asumiremos* hipótesis, ni *clamaremos* nuestras opiniones, ni utilizaremos discos *removibles*[1]...

Otro de los peligros procede de los términos ingleses polisémicos, tales como *language*, que puede significar *lenguaje* o *idioma*. Sin duda sería más correcto decir *idiomas de programación* en vez de *lenguajes de programación*; también aquí el uso ha impuesto una traducción incorrecta. Hay otros errores tan graves que, por muy extendidos que estén, nos negamos a admitir. Uno de ellos se debe al término *random*, que puede significar *aleatorio* o *arbitrario* (en los dos sentidos del término castellano). El ordenador que hemos utilizado en nuestro proyecto tiene *memoria de acceso arbitrario* (*random access memory*, RAM), o dicho con más propiedad, *memoria de acceso directo*; si le hubiéramos instalado *memoria de acceso aleatorio*, el funcionamiento del sistema experto sería impredecible.

También hay que procurar no introducir neologismos irreflexivamente. Por ejemplo, la palabra "*lincar*" resulta innecesaria, pues existe en castellano *enlazar*, que es la fiel traducción del verbo inglés *link*; sin embargo, su uso está tan extendido que probablemente acabará imponiéndose. Esperemos que no ocurra lo mismo con aberraciones tales como "*apdatear*", "*rebutear*" o "*debujear*". ¿Por qué no decir *actualizar*, *reiniciar* o *depurar*, que suenan mucho mejor?

Más discutida fue la utilización de *implementar* y sus derivados. Es cierto que existen expresiones como *realizar* o *llevar a cabo*, pero su significado es más amplio, por lo que finalmente la Real Academia Española ha incluido en su diccionario este

[1] Suponemos que el objetivo de un disco *removible* sería poder agitar los datos; en cambio, para almacenar y transportar información es más recomendable utilizar discos *extraíbles* (*removable disks*).

Hemos reunido en una lista varias docenas de falsos amigos entre el inglés y el español. Sería incluso divertido comentar anécdotas que han provocado algunos de ellos, pero no es éste el lugar apropiado.

verbo, con el significado de "poner en funcionamiento". A nosotros nos parece conveniente utilizar también el término *implementación*, con la confianza de que algún día la Academia dará su conformidad.

Por otra parte, hay algunas expresiones que aún no tienen una traducción unánimemente aceptada. Así, se habla frecuentemente de los *conjuntos borrosos*, según una de las posibles traducciones literales de *fuzzy sets*. Nosotros, en cambio, preferimos denominarles *conjuntos difusos*, pues el hecho de que éstos no tengan una frontera bien definida es una característica intrínseca, mientra que el término *borroso* sugiere que se trata de un fenómeno de observación. En cuanto a la traducción de *interface*, nos parece acertada la decisión de la Academia al optar por *interfaz*, en vez de "*interfase*", que correspondería a *interphase*.

Además de los términos que se emplean aquí con sentido metafórico (*marcos, demonios, facetas*...), utilizamos el término *instancia* como neologismo, para traducir "*instance*", que significa *ejemplo, caso particular* o *elemento de una clase*.

Por una cuestión de puro gusto personal, preferimos decir *indexado* en vez de *indizado* (ambos están admitidos por la Academia), porque el primero se utiliza con mucha más frecuencia en informática y porque nos suena mejor, aunque sea más anglosajón. Igualmente, nos suena mejor *diseñador del sistema experto* que *ingeniero del conocimiento*.

Hay algunas expresiones realmente difíciles de traducir. Una de ellas es *pattern matching*. En nuestra opinión, hay que empezar reconociendo que, en este contexto, el verbo *match* se emplea con dos significados. El primero de ellos se refiere al hecho de *comparar* dos patrones para ver si coinciden; por eso creemos que la expresión anterior debe traducirse por *comparación de patrones*. El segundo significado indica que los patrones efectivamente coinciden: tal es el sentido en frases como "*The assertion matches the pattern*", que traduciremos como "La afirmación *iguala* el patrón".

A pesar del interés por cuidar la lengua, tememos que esta memoria pueda contener numerosos errores semánticos, gramaticales y estilísticos, por lo que agradeceremos todas las correcciones y sugerencias que recibamos.

Lista de figuras

1.1 Sistema completo proyectado originalmente. 4
1.2 Sistema experto realizado en esta tesis. 5
1.3 Pequeño ejemplo que será desarrollado en la sección 4.1. 7

2.1 Estructura convencional de un sistema experto. 16
2.2 Ejemplo de marco utilizado en el sistema experto PIP. 20
2.3 Método clásico de diagnóstico probabilístico. 30
2.4 Nodo C con dos causas y un efecto. 38

4.1 Nodo X con un hijo Y_1. 56
4.2 Nodo X con dos hijos. 59
4.3 Nodo X con dos padres. 63
4.4 Nodo X con dos padres y dos hijos. 66
4.5 Un ciclo y dos bucles. 70
4.6 Un pequeño poliárbol. 75

5.1 Propagación de evidencia mediante intercambio de mensajes. 81
5.2 Padres de Y_j. 83
5.3 Computaciones realizadas en el nodo X. 87
5.4 Ejemplo de puerta OR. 91
5.5 Computaciones realizadas en la puerta OR. 97

6.1 Conocer el valor de A nos permite romper el bucle. 102
6.2 Árbol de cúmulos. 105
6.3 Computaciones realizadas en el cúmulo C_1. 106
6.4 Una parte de la red y de su poliárbol asociado. 109
6.5 Bucles adyacentes. 110
6.6 Cadena de rombos. 117
6.7 La escalera contiene N cuadrados y $2N + 2$ nodos. 120
6.8 Inclusión de un nodo-acción. 123

7.1 Mensajes para la familia X (en función de los parámetros). 131
7.2 Aprendizaje en el nodo X. 139

8.1 Pantalla de explicación para los efectos de un nodo. 157
8.2 Intervalos de un parámetro. 158

11.1 Fases de desarrollo de un sistema experto (tomado de [24, pág. 139]). . 198

12.1 Proceso básico de razonamiento en medicina. 210
12.2 Porción de la red para la insuficiencia mitral. 213

13.1 Pantalla para introducir un nuevo eco. 230
13.2 Parámetros del eco modo M. 233

13.3 Interpretación del electrocardiograma. 235
13.4 Pantalla para indicar las intervenciones valvulares. 236
13.5 Diagnósticos obtenidos 237
13.6 Explicación de las causas de un diagnóstico. 238

Lista de tablas

4.1 Probabilidad del paludismo, $P(+x|u_1, u_2)$. 63

5.1 Parámetros $c_x^{u_1}$. 91
5.2 Parámetros $c_x^{u_2}$. 91
5.3 Caso general y puerta OR 96

6.1 Operaciones necesarias para la cadena de rombos. 118

Parte I

INTRODUCCIÓN

Capítulo 1

Presentación

1.1 Proyecto para ecocardiografía

La inteligencia artificial es la rama de la informática que trata de reproducir las capacidades de la mente humana mediante computadoras. Dentro de ella, los sistemas expertos son programas de ordenador que poseen gran cantidad de conocimiento sobre un campo concreto, que en nuestro caso será la ecocardiografía.

Ésta es una nueva tecnología de diagnóstico que ha experimentado un extraordinario desarrollo en los últimos veinte años. Su fundamento consiste en obtener imágenes del corazón mediante el análisis de las ondas ultrasónicas reflejadas. Su cualidad principal es que se trata de una técnica no invasiva (al menos el ecocardiograma habitual, que es transtorácico), por lo que aporta gran cantidad de información sin riesgo para el paciente. El mayor inconveniente es que las imágenes obtenidas son de calidad muy variable, de modo que su interpretación resulta difícil e incluso entre cardiólogos con larga experiencia surgen diferencias en el diagnóstico. Por esta razón sería deseable aplicar al problema las técnicas modernas de visión artificial, con el fin de obtener imágenes más nítidas, de las que puedan extraerse medidas y observaciones precisas.

Nuestro objetivo inicial fue construir un sistema experto para ecocardiografía que incluyera el tratamiento digital de las imágenes, tal como muestra la figura 1.1. Este esquema indica que el médico controla el ecocardiógrafo y observa en él las imágenes, que son también almacenadas como grabaciones de vídeo. La tarjeta de visión obtiene las imágenes a partir del ecocardiógrafo o de la cinta de vídeo, las digitaliza y extrae la información que contienen, la cual queda disponible para el médico y para el sistema experto. Éste realiza el diagnóstico utilizando, además del ecocardiograma, la historia clínica —contenida en la base de datos— y los síntomas, signos y pruebas complemetarias (electrocardiograma, analítica, radiografías, etc.), introducidos en el sistema por el personal médico o paramédico. (Sería deseable en un futuro que toda la información procedente de las diversas fuentes se introdujera automáticamente, liberando así al personal sanitario de labores administrativas, aunque este deseo es claramente utópico a corto plazo.)

Sin embargo, pronto se vio que el proyecto así definido resultaba demasiado ambicioso para una sola tesis doctoral, por lo que se decidió dividirlo en dos partes. Una se ocuparía del tratamiento digital de los ecocardiogramas: filtrado de imágenes, segmentación, reconocimiento y medición. La otra consistiría en la interpretación de

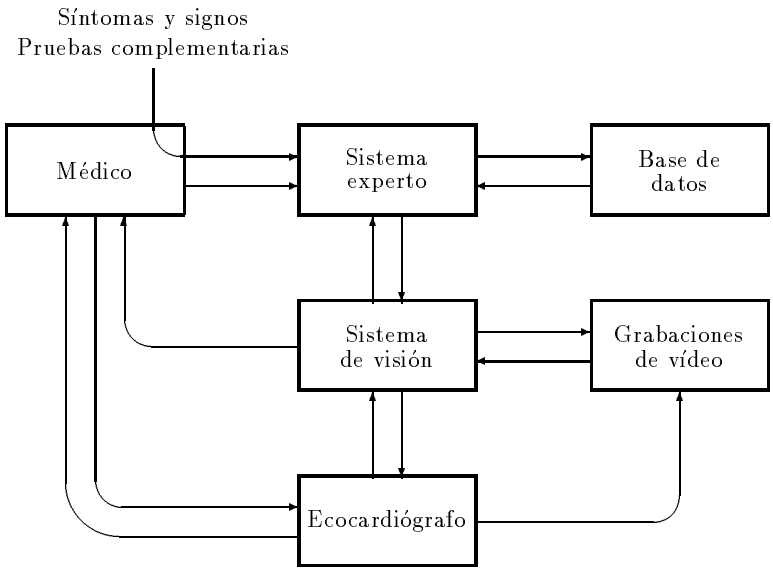


Figura 1.1: Sistema completo proyectado originalmente.

la información obtenida, con el fin de llegar a un diagnóstico. La primera de ellas es ante todo un trabajo sobre visión artificial, que está siendo realizado por el Prof. José Luis Fernández Vindel, también en el Departamento de Informática y Automática de la UNED. La segunda parte se sitúa dentro de la inteligencia artificial para medicina, y constituye el objeto de la presente tesis doctoral.

En el momento actual el sistema de visión se encuentra en fase de desarrollo, por lo que el sistema experto no tiene aún acceso directo al análisis del ecocardiograma, sino que los resultados han de ser introducidos directamente por el médico, tal como muestra la figura 1.2. Esto plantea varios inconvenientes. El primero de ellos es que el médico no puede contar con la información que proporcionaría el sistema de visión: imágenes filtradas de mayor nitidez, y medidas objetivas cuya obtención resulta lenta y tediosa incluso para un cardiólogo experimentado; tampoco puede contar con el asesoramiento que le proporcionaría el sistema experto a la hora de realizar la prueba. Por último, está el inconveniente de que el médico pierde tiempo introduciendo en el ordenador los resultados obtenidos. En consecuencia, es deseable que el sistema de visión se encuentre operativo lo antes posible, con el fin de que pueda integrarse con nuestro sistema experto.

En resumen, esta tesis se centra en el desarrollo de un sistema experto para el diagnóstico de cardiopatías, denominado DIAVAL[1] que utiliza el ecocardiograma como fuente principal de información y los síntomas, signos y otras pruebas como datos complementarios. Todos estos resultados serán introducidos por el médico —o

[1] El nombre se debe a que, inicialmente, el sistema experto se iba a limitar al DIAgnóstico de VALvulopatías; posteriormente su capacidad se extendió hasta abarcar toda la ecocardiografía, pero mantuvo su nombre original.

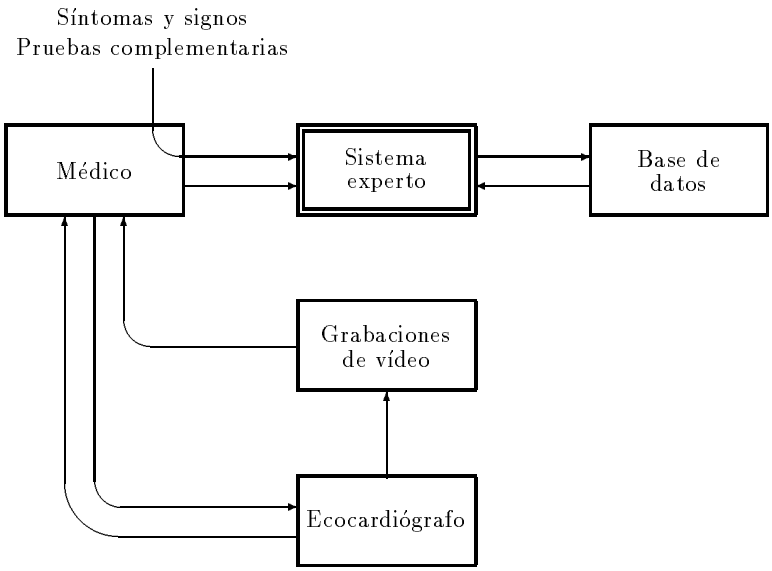


Figura 1.2: Sistema experto realizado en esta tesis.

por el personal paramédico— en el sistema, el cual realiza un diagnóstico y elabora un informe que puede pasar a la impresora.

1.2 Desarrollo, metodología y objetivos

Inicialmente, estaba previsto construir el sistema experto utilizando la metodología estándar, es decir, mediante reglas de producción. Sin embargo, pronto la tesis tomó un rumbo diferente. Desde las primeras entrevistas con los médicos que han colaborado en el proyecto, ellos mismos sugirieron utilizar un modelo patofisiológico de la evolución de las enfermedades cardíacas. Nos encontramos así con una red causal construida por nuestros expertos, que era imposible de traducir a reglas, por las razones que se explicarán en la sección 2.4.

Ante esta limitación, desconociendo la literatura publicada sobre redes bayesianas en los últimos años, empezamos a investigar la aplicación del teorema de Bayes con el fin de realizar diagnósticos a partir de nuestra red causal. Así redescubrimos los axiomas de independencia y sus propiedades de asimetría, desarrollamos la puerta OR como modelo de interacción entre las causas e implementamos un primer algoritmo para el cálculo de la probabilidad. (Esto ocurrió en 1990, el mismo año en que aparecieron varios de los trabajos más importantes sobre este campo.) Posteriormente, pudimos saber que todos estos hallazgos habían sido realizados y superados unos años antes por varios investigadores, especialmente por el Prof. Judea Pearl, de la Universidad de California Los Angeles, que suele ser considerado como el creador de las redes bayesianas.

Sin embargo, el esfuerzo no fue inútil, pues el conocimiento adquirido nos permitió

desarrollar posteriormente un algoritmo propio, el condicionamiento local, capaz de competir y en ciertos aspectos superar al método más utilizado en la actualidad para calcular la probabilidad en redes bayesianas. También descubrimos que la generalización de la puerta OR que estábamos utilizando había sido propuesta anteriormente por Max Henrion, aunque él no había dado una definición formal del modelo ni había aprovechado las ventaja que ofrece en la computación de la probabilidad. Además, esta tesis desarrolla un modelo estadístico de aprendizaje (ajuste de parámetros) y propone un método de explicación del razonamiento para redes bayesianas.

Por tanto, la **metodología** escogida para realizar el sistema experto cardiológico fue impuesta por el propio modelo fisiopatológico, y consiste en la aplicación estructurada del teorema de Bayes en una red causal, conocida como red bayesiana.

Éste es un campo muy reciente, que surgió en la década de los 80 y se ha expandido sobre todo en los últimos seis años. Actualmente, numerosas universidades de todo el mundo, incluyendo Stanford, Carnegie-Mellon y el MIT, y varias empresas privadas están investigando sobre redes bayesianas. Sin embargo, prácticamente no existe ninguna publicación en nuestro idioma sobre este campo. Es más, hasta donde conocemos, tampoco en otros idiomas existe un libro o una tesis que trate todos los aspectos de las redes bayesianas: fundamentos, modelización, algoritmos, implementaciones (recursiva, iterativa y distribuida), obtención del conocimiento, aprendizaje y explicación.

En consecuencia, esta tesis se ha propuesto dos **objetivos** principales. El primero consiste en realizar un estudio a fondo sobre todos los temas relacionados con las redes bayesianas, intentando ofrecer algunas aportaciones originales varios de ellos. El segundo es eminentemente práctico: el desarrollo de un sistema experto para ecocardiografía que funcione en el Servicio de Cardiología del Hospital de la Princesa, de Madrid, y desde allí pueda implantarse posteriormente en otros centros sanitarios.

1.3 Organización de la tesis

La tesis está dividida en cuatro partes. La primera es una introducción, que comienza con el presente capítulo, dedicado a explicar las motivaciones y objetivos del trabajo realizado. Puesto que vamos a desarrollar un sistema experto, dedicamos el próximo capítulo a explicar brevemente el contexto de nuestra investigación. Nuestro propósito no es simplemente dar una introducción a la inteligencia artificial y los sistemas expertos, sino mostrar una visión crítica y personal sobre el tema. En este sentido es especialmente importante la sección 2.4, que explica por qué no hemos utilizado una representación basada en reglas, a pesar de ser ésta la metodología más habitual hoy en día.

El capítulo 3 es de naturaleza diferente a los demás, pues no trata principalmente sobre matemáticas, ni sobre informática, ni sobre medicina. En él se recogen varios estudios psicológicos sobre los errores que suele cometer el ser humano en la estimación de la probabilidad. Al mostrar que el sentido común innato tiende a equivocarse sistemáticamente en este tipo de problemas buscamos, por un lado, resaltar la utilidad que un sistema basado en la teoría de la probabilidad puede tener como ayuda a los expertos humanos, especialmente en medicina, y por otro, intentamos refutar la opinión generalizada de que un sistema inteligente debe imitar el funcionamiento de la mente humana; nosotros, en cambio, sostenemos que los sistemas de ayuda al

diagnóstico deben fundamentarse sobre todo en los principios normativos de la teoría de la decisión.

La parte II de la tesis se dedica al estudio teórico de las redes bayesianas, independientemente del campo de aplicación. Antes de abordar las complejidades matemáticas, la sección 4.1 desarrolla progresivamente el ejemplo de la figura 1.3, en que los nodos representan las siguientes variables:

$$\begin{aligned} U_1 &= \text{País-de-origen} \\ U_2 &= \text{Tipo-sanguíneo} \\ X &= \text{Paludismo} \\ Y_1 &= \text{Gota-gruesa} \\ Y_2 &= \text{Fiebre} \end{aligned}$$

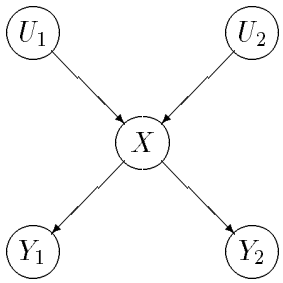


Figura 1.3: Pequeño ejemplo que será desarrollado en la sección 4.1.

La finalidad de esta sección es que el lector no familiarizado con el tema pueda tener una visión intuitiva de qué son las redes bayesianas, por qué se introducen los axiomas de independencia y cómo se propaga la evidencia hasta llegar a un diagnóstico; más concretamente, nos interesa aclarar el significado de los mensajes π y λ, que desempeñarán un papel tan importante en capítulos posteriores. Insistiremos sobre todo en la íntima conexión entre el sentido común y los resultados matemáticos, para mostrar que los axiomas y las propiedades que se van a introducir a continuación no son arbitrarios.

La sección 4.2 es puramente matemática: consiste en una definición formal de las redes bayesianas. Por fin, el capítulo concluye con unas reflexiones sobre la semántica de las redes bayesianas. De nuevo, insistiremos en la relación entre las propiedades matemáticas de las RR.BB. y el razonamiento causal intuitivo.

Según David Marr [140, sec. 2.1], en el tratamiento computacional de un problema pueden distinguirse tres niveles: teoría, algoritmo e implementación. Así, en el ejemplo de la aritmética, tal como él lo propone, la **teoría** trata sobre las propiedades asociativa, conmutativa, de elemento neutro, etc., que son independientes de si la representación es decimal, binaria o romana. Los **algoritmos** son los métodos que se utilizan para realizar las operaciones aritméticas; en el caso de una representación

decimal, el algoritmo consiste en las reglas que todos conocemos para la suma, resta, multiplicación y división. La **implementación** consiste en traducir dicho algoritmo a un lenguaje de programación (ensamblador, C, Pascal, Lisp, etc.) con el fin de que pueda ser comprensible para el procesador de una computadora.

Siguiendo este esquema, después de haber dedicado el capítulo 4 a los fundamentos teóricos, en los dos siguientes estudiaremos los algoritmos destinados al cálculo de la probabilidad, tanto para poliárboles (sec. 5.1) como para redes con bucles (cap. 6), comparando los métodos previos con el que nosotros proponemos. Para cada uno de ellos discutiremos cuáles son las implementaciones posibles; entre ellas, nos interesa especialmente la distribuida (paralela), aunque mencionaremos también la recursiva y la iterativa.

La sección 5.2 utiliza el mismo esquema de tres niveles para el estudio de la *puerta OR*: teoría (5.2.2), algoritmo (5.2.3) e implementación (5.2.4). De nuevo, el esquema se repite en el capítulo dedicado al *aprendizaje*: teoría (7.2.2), algoritmo (7.2.3-7.2.5) e implementación (7.2.6).

Al tratar la *explicación del razonamiento* nos gustaría poder aplicar el mismo esquema de tres niveles. Lamentablemente, aún no existe una teoría sobre este tema, ni en los sistemas expertos en general, ni en las redes bayesianas en particular, por lo que nos limitaremos a definir los tipos de explicación y a hacer una revisión de los métodos que se han propuesto como soluciones *ad hoc* para resolver aspectos parciales del problema. Después, mostraremos el método de explicación utilizado en el sistema experto que hemos desarrollado y la forma de realizar una implementación distribuida.

El capítulo 9 muestra cómo se ha implementado en nuestro sistema el razonamiento bayesiano. Su mayor interés consiste en que describe la estructura y el funcionamiento de una herramienta concreta, *GoldWorks*, destinada a la construcción de sistemas expertos, explicando cómo se utilizan en la práctica la programación orientada a objetos, los marcos, las reglas, las agendas, los demonios, etc.

La parte II concluye con un capítulo destinado a situar las redes bayesianas en el contexto de la inteligencia artificial y a discutir sus ventajas y limitaciones.

Después de estudiar teoría de las redes bayesianas, la parte III está dedicada al desarrollo de una aplicación concreta, el sistema experto DIAVAL, para el diagnóstico de enfermedades cardíacas mediante ecocardiograma, con un énfasis especial en las valvulopatías. El capítulo 11 explica cuáles han sido las fases que se han seguido a la hora de diseñar el sistema y describe la primera de ellas, conocida como *identificación del problema*. La sección 11.3 está dedicada a la obtención del conocimiento necesario para construir una red bayesiana, con especial interés por el campo de la medicina.

El capítulo siguiente sintetiza los conocimientos médicos que hemos necesitado. Empieza con una sección dedicada a los conceptos básicos, comunes a todas las especialidades de la medicina: diagnóstico, pronóstico, hallazgos patognomónicos, etc., y a continuación explica brevemente cuáles son los datos que considera nuestro modelo y las enfermedades que puede diagnosticar.

Una vez que tenemos el conocimiento médico necesario, vamos a mostrar cómo lo hemos estructurado en forma de red bayesiana y cómo hemos diseñado el interfaz, intentando ajustarnos a las características ideales comentadas en la sec. 13.2.1. Describiremos sobre todo la entrada de datos y la presentación de los resultados.

El capítulo 14 se ocupa de la evaluación del sistema. Tras revisar en la bibli-

ografía los trabajos anteriores, describe el método de evaluación empleado, discute los resultados obtenidos y, finalmente, propone un método de evaluación más rigurosa.

Finalmente, la parte IV contiene las conclusiones: las lecciones aprendidas, las aportaciones realizadas, la valoración de los resultados y las cuestiones que han quedado abiertas como caminos para futuros trabajos de investigación.

1.4 Guía de lectura

La primera parte consiste en una introducción a la tesis. El lector familiarizado con los sistemas expertos y el tratamiento de la incertidumbre comprobará que la mayor parte del capítulo 2 le resulta conocida. Puede ser interesante, sin embargo, detenerse en la sección 2.4, en que se hace una crítica de los sitemas basados en reglas, pues recopila varios artículos relativamente recientes que no suelen aparecer en la literatura sobre sistemas expertos. Es conveniente conocer estos temas para poder situar las redes bayesianas en el contexto de la inteligencia artificial y valorar sus cualidades y limitaciones (cap. 10).

El capítulo 3, dedicado a los errores humanos en la estimación de la probabilidad, no es necesario para comprender el resto de la memoria.

Dentro de la parte II, los tres primeros capítulos presentan la teoría y los algoritmos para el cálculo de probabilidades en redes bayesianas. La presentación intuitiva de la sección 4.1 puede ayudar a entender el sentido de las definiciones y axiomas que aparecerán en capítulos posteriores.

La sección 5.1 se limita a presentar el algoritmo de Kim y Pearl [126, 172]; el único detalle digno de señalar es que se ha modificado la definición de $\pi(x)$ y de $\pi_X(u_i)$ de acuerdo con la sugerencia de Peot y Shachter [181], con el fin de permitir un tratamiento coherente del algoritmo de condicionamiento local, que se presentará en el capítulo siguiente. Esta sección es necesaria para comprender el resto de la parte II. También conviene conocer la puerta OR (sec. 5.2), porque hablaremos de ella con frecuencia en capítulos posteriores.

El capítulo 6 presenta el condicionamiento local (sec. 6.2), como extensión del algoritmo de Kim y Pearl mostrado en el capítulo anterior. La sección 6.1, "Métodos previos", además del interés que presenta por sí misma, resulta útil para comparar nuestro algoritmo con los métodos desarrollados por otros investigadores.

A continuación, se encuentra el capítulo más denso en expresiones matemáticas. Para las cuestiones que se van a discutir posteriormente, basta saber que el algoritmo de aprendizaje propuesto se reduce a las expresiones que aparecen en la figura 7.2 (pág. 139).

El tema de la generación de explicaciones lingüísticas (cap. 8) puede resultar mucho más ameno y no requiere una lectura tan detenida como los anteriores. Lo mismo puede decirse del capítulo siguiente, decicado a la implementación del razonamiento bayesiano; las dos primeras secciones aportarán pocos conocimientos a quien esté ya familiarizado con el Lisp, la programación orientada a objetos o las herramientas para construir sistemas expertos. La sección 9.3 sirve de base al capítulo 13, en que se describe la implementación del sistema experto para cardiología.

La parte II concluye con la valoración de las redes bayesianas. Para situarlas en el contexto de la inteligencia artificial y ponderar sus ventajas y limitaciones, es

conveniente conocer los temas tratados en el capítulo 2.

La tercera parte describe el sistema experto DIAVAL. A pesar de que se trata de una aplicación particular, la mayor parte de su contenido es válida para cualquier otro campo de la medicina, especialmente si se trata de construir un sistema experto basado en redes bayesianas. Entre los temas que aparecen en ella se encuentran los siguientes: las fases de desarrollo de un sistema experto (sec. 11.1), la obtención del conocimiento (sec. 11.3), la construcción de la red (sec. 13.1), el interfaz de usuario (sec. 13.2) y la evaluación del sistema (cap. 14).

Quien desee conocer más a fondo el caso de la medicina, podría encontrar especialmente interesante el capítulo 12, que contiene de forma muy resumida el conocimiento médico introducido en el sistema. En él se presentan los conceptos médicos básicos (sec. 12.1.2) y las técnicas de diagnóstico (12.3) y tratamiento (12.4) utilizadas en cardiología, que son esencialmente las mismas utilizadas en muchas otras especialidades.

En la parte IV, dedicada a las conclusiones, puede encontrarse un resumen de los resultados obtenidos; en particular se indican allí cuáles son las aportaciones más relevantes de esta tesis (sec. 15.2) y se valora el sistema experto que hemos construido (sec. 15.3). Por último, la sección 15.4 puede ser interesante como fuente de inspiración para futuras líneas de investigación, tanto de naturaleza teórica (redes bayesianas) como aplicada (sistemas expertos, especialmente para medicina).

Capítulo 2

Fundamentos de inteligencia artificial

2.1 Inteligencia artificial y sistemas expertos

2.1.1 Orígenes históricos

En 1943, Warren McCulloch y Walter Pitts [143] sugirieron la semejanza entre el funcionamiento del cerebro y las máquinas de Turing y propusieron las neuronas formales como modelo simplificado orientado a simular el comportamiento de las neuronas reales.

Otro de los pioneros de la inteligencia artificial (I.A.) fue Claude Shannon; además de realizar aportaciones de gran importancia sobre teoría de la información, estudió en 1950 la posibilidad de utilizar ordenadores para jugar al ajedrez [205]. También los trabajos de Norbert Wiener, creador de la cibernética [190, 259], y los de von Neumann, autor de la teoría de autómatas [243], ejercieron gran influencia en el surgimiento de la inteligencia artificial.

En 1950, Alain Turing discutió en profundidad la posibilidad de construir una máquina inteligente. Para formalizar la cuestión, propuso el famoso *test de Turing* [238]: una máquina puede considerarse inteligente si es capaz de mantener una conversación a través de un teletipo, como si fuera un ser humano. La literatura posterior suele presentar el test de Turing poniendo el énfasis en la capacidad de engañar a una persona, a pesar de que la formulación original la cuestión de simular inteligencia como un juego de imitación, no como un juego de engaño [145].

A pesar de estos antecedentes, suele considerarse 1956 como el año en que nació la inteligencia artificial, pues el término fue acuñado por John McCarthy al convocar la Conferencia de Darmouth, basada en "la conjetura de que todo aspecto del aprendizaje o cualquier otra característica de la inteligencia puede en principio describirse con tanta precisión que pueda ser simulado mediante una máquina" [37, pág. 11].

En esta conferencia, Allen Newell y Herber Simon presentaron el programa *Logic Theorist* [156], capaz de demostrar teoremas matemáticos. Más tarde desarrollaron el *General Problem Solver (GPS)* [157], entre sus numerosas aportaciones a la inteligencia artificial.

Otros de los participantes en la conferencia fueron Marvin Minsky, que ha realizado contribuciones destacadas sobre la representación del conocimiento —a él debe-

mos el concepto de *marco* (sec. 2.2.4)—, Arthur Samuel, que construyó un jugador de damas capaz de aprender hasta llegar a derrotar a su creador, y Alex Bernstein, que construyó un jugador de ajedrez, el primer programa de I.A. capaz de realizar búsqueda heurística, entendiendo como *heurística* el conjunto de reglas que procede más bien de la experiencia que de planteamientos teóricos, y sirve para guiar el razonamiento.

Algunos de los campos que *aún* se estudian dentro de la I.A. son los siguientes:[1]

- *Teoría de juegos.* Se estudiaron desde el sencillo tres-en-raya hasta el difícil ajedrez. Hoy en día ya no suelen considerarse como problemas de I.A., al menos los juegos más simples, pues se resuelven con algoritmos de búsqueda que no requieren el uso de conceptos.

- *Demostración automática de teoremas.* Fue, junto con el anterior, el primer campo donde se desarrollaron programas "inteligentes". Sigue siendo aún hoy un área importante que aporta ideas valiosas para otros relacionados con él, especialmente para los sistemas expertos.

- *Visión artificial.* La visión de bajo nivel (filtrado) no pertenece a la I.A., pero la de medio nivel (segmentación) y, sobre todo, la de alto nivel (reconocimiento) están dentro de ella.

- *Robótica.* Trata de dar cada vez mayor flexibilidad y autonomía a los robots para que puedan adpatarse a nuevas situaciones. Está muy relacionado con la *visión artificial* y con los problemas de *planificación*.

- *Comprensión del lenguaje.* Éste es un campo muy complejo, incluso cuando el texto está ya introducido en forma de caracteres en el ordenador. El problema se complica más aún cuando se trata de escritura manual o de una conversación, hasta el punto de que el *reconocimiento del habla* se considera en sí mismo como una rama de la I.A., y el *reconocimiento de caracteres escritos* constituye uno de los problemas básicos de la visión artificial.

- *Traducción automática.* Muy relacionado con el anterior, éste es el campo donde se produjo el fracaso más estrepitoso de la I.A., debido a que se intentó hacer traducciones casi palabra por palabra, sin caer en la cuenta de la complejidad del problema. Hoy en día vuelve a ser objeto de intensa investigación, aunque abordado con las técnicas de comprensión y generación de lenguaje natural.[2]

- *Programación automática.* Consiste en el diseño de sistemas que generen código eficiente a partir de especificaciones parciales o imprecisas. La capacidad de un sistema para programarse a sí mismo es uno de los rasgos más característicos de la inteligencia.

- *Interfaces inteligentes.* Su objetivo es que el usuario pueda manejar una base de datos o un sistema operativo, por ejemplo, en lenguaje natural y el sistema

[1] La palabra "*aún*" indica que las fronteras de la inteligencia artificial se desplazan con el tiempo, como discutiremos más adelante.

[2] Se habla de *lenguaje natural* al referirse a los idiomas humanos (español, inglés, etc.), en contraposición a los lenguajes informáticos (FORTRAN, LISP, etc.).

"adivine" sus intenciones, corrija sus errores y le dé los consejos oportunos. La disponibilidad de este tipo de interfaces es fundamental para la implantación masiva de los ordenadores en la vida cotidiana.

- *Sistemas expertos.* Es el campo donde se han conseguido los mayores éxitos prácticos de la I.A. En la sección 2.1.3 trataremos su definición e historia, y en capítulos posteriores abordaremos diferentes aspectos, tales como inferencia, aprendizaje, explicación, obtención del conocimiento, etc.

Hay otros cuatro campos que, por plantear cuestiones fundamentales, se encuentran relacionadas con todos los mencionados anteriormente:

- *Lenguajes de programación.* Aunque su importancia teórica es de segundo orden, en la práctica es muy útil contar con las herramientas adecuadas que faciliten el trabajo. Los lenguajes más característicos de la inteligencia artificial son LISP (sec. 9.1), PROLOG (sec. 2.2.1) y los orientados a objetos (sec. 9.2.2).

- *Representación del conocimiento.* Hoy en día se reconoce como una de las cuestiones claves de la I.A., pues de ella dependen el almacenamiento de la información, la expresividad, la eficiencia del razonamiento, el aprendizaje... Trataremos este punto en la sección 2.2.

- *Razonamiento aproximado:* La incertidumbre surge en prácticamente todas las aplicaciones reales de la I.A. En la sección 2.3 hablaremos sobre las fuentes de incertidumbre y los métodos para tratarla.

- *Aprendizaje.* Este es un campo tan extenso que resulta imposible intentar citar brevemente todos los tipos que existen, dependiendo del área de aplicación, fuentes de información, técnicas, objetivos, etc. El capítulo 7 está dedicado al aprendizaje en redes bayesianas.

2.1.2 Ramas y tendencias en I.A.

Dentro de la I.A. podemos distinguir tres grandes ramas según el método que utilizan:

Simbólica: Su característica esencial es que representa los conceptos mediante símbolos. El razonamiento consiste en manipular los símbolos de acuerdo con ciertas reglas obtenidas a partir de principios lógicos.

Probabilística: Se basa en el cálculo de probabilidades, concretamente en la aplicación del teorema de Bayes, para realizar un razonamiento. Aunque la inferencia se desarrolla a nivel numérico (cuantitativo), hay luego un paso al nivel cualitativo, pues de otro modo estaríamos en el campo de las matemáticas, no en el de la inteligencia artificial.

Conexionista: Consiste en utilizar una red de neuronas artificiales (pueden ser procesadores reales o simulados mediante ordenador). Suelen utilizarse sobre todo en tareas de clasificación. La red se programa a través de ejemplos y la información se almacena en forma de *pesos* (números) asociados a los enlaces entre neuronas.

En general, cuando se habla de inteligencia artificial se trata de la rama simbólica, pues las redes neuronales tienen un origen independiente (por cierto, anterior a la I.A. simbólica) y se han desarrollado tanto en los últimos años que constituyen un campo amplio y bien diferenciado, dotado de gran vitalidad. Los esfuerzos por integrar ambas corrientes aún no han dado los frutos deseados. Por otra parte, la rama probabilística permanece unida al tronco de la I.A., pues como ya hemos mencionado, necesita utilizar los métodos simbólicos para pasar del nivel cuantitativo al cualitativo.

Como veremos en capítulos posteriores, las redes bayesianas se sitúan en la corriente probabilística en cuanto a la inferencia y al aprendizaje; sin embargo, la capacidad de explicación utiliza casi siempre procesamiento simbólico.

Si en vez de considerar los métodos atendemos a los objetivos, podemos reconocer en la I.A. tres tendencias:

Psicológica: Se fundamenta en el estudio de la mente humana con el fin de reproducir sus capacidades mediante sistemas informáticos. La tendencia equivalente a ésta dentro de la computación neuronal sería la **biológica**, basada en el estudio de los sistemas nerviosos animales.

Teórica: Estudia los principios teóricos de la inteligencia que deben regir tanto al razonamiento humano como el automático. Se apoya sobre todo en el estudio de la lógica, la epistemología, la teoría de la probabilidad, la lingüística, etc.

Pragmática: Supone que la forma más eficaz de desarrollar la I.A. consiste en tratar de resolver problemas concretos y, a partir de la experiencia, generalizar los resultados para poder aplicarlos a áreas diferentes.

Personalmente, nos situamos principalmente en la tendencia teórica, aunque sintiendo gran aprecio por las otras dos. La tendencia psico-biológica puede ser muy útil como fuente de inspiración, del mismo modo que la observación del vuelo de aves e insectos puede inspirar el desarrollo de la aeronáutica. (Hay que reconocer que el ser humano sigue siendo la única "máquina" verdaderamente inteligente y no es previsible, al menos a medio plazo, que pueda ser superada por la I.A.) Sin embargo, a la hora de construir sistemas artificiales nos parece más provechoso —por ejemplo— el estudio de la aerodinámica o de los fundamentos de la inteligencia, pues no se trata de reproducir lo que ya existe en la naturaleza sino de construir máquinas que nos faciliten lo que ésta no proporciona.

Igualmente, consideramos muy valiosa la aproximación pragmática a la I.A. siempre que busque métodos generales y no se limite a resolver problemas particulares con soluciones *ad hoc*, pues entonces no sería siquiera una ingeniería, mucho menos una ciencia. Sin embargo, es muy útil afrontar problemas concretos y reales porque con ello se plantean nuevas cuestiones y se decantan las que son verdaderamente importantes, guiando así la investigación teórica. La realimentación teoría-aplicaciones ha sido especialmente provechosa en el desarrollo de esta tesis, que no es el fruto de elucubraciones sobre teoría de la probabilidad, sino que el estudio de las redes bayesianas, la generalización de la puerta OR, el modelo de aprendizaje y el método de explicación han ido surgiendo al intentar construir un sistema experto para cardiología. Además, sería absurdo sentarse a esperar una teoría completa sobre la inteligencia, que quizá no consigamos nunca; en cambio, la tendencia pragmática ha

producido sistemas expertos que, a pesar de todas sus limitaciones, están resultando muy eficaces a la hora de resolver problemas concretos del mundo real.

2.1.3 Sistemas expertos. Definición y estructura

Los primeros años de la I.A. estuvieron dedicados a estudiar métodos de razonamiento generales, basados sobre todo en la lógica y en técnicas de búsqueda. Después de dos décadas de investigación, se llegó a la conclusión de que la potencia de un sistema depende más del conocimiento específico que posee sobre un campo concreto que de las estrategias de razonamiento que utiliza [87]. Surgió así en la década de los 60 el área de los sistemas expertos, que se desarrolló con fuerza en la década siguiente.

Las características fundamentales de un sistema experto son:

1. *Dominio reducido:* Los SS.EE. se diferencian de otros anteriores en que, en vez de tratar de resolver problemas generales, se especializan en un campo limitado, como puede ser la interpretación de espectrogramas químicos, la prospección geológica, las enfermedades infecciosas, etc.

2. *Competencia en su campo:* De ahí viene su nombre: al igual que un experto humano es una persona que sabe mucho sobre su especialidad, de modo que puede dar respuestas rápidas y fiables, un sistema experto debe resolver los mismos problemas con una eficiencia comparable a la de un especialista humano. La complejidad de los problemas abordados diferencian los SS.EE. de muchos programas anteriores de I.A., que no trataban cuestiones del mundo real, sino modelos muy simplificados (en inglés, *toy problems*, problemas "de juguete").

   Sin embargo, los sistemas expertos se están aplicando en varios tipos de problemas del mundo real; las *tareas* que realizan pueden clasificarse como interpretación, predicción, diagnóstico, diseño, planificación, monitorización, depuración, reparación, instrucción y control [222, 246]. El libro de Waterman [246] contiene un estudio de los SS.EE. más importantes que se habían desarrollado hasta 1986 en agricultura, computación, control de procesos, derecho, electrónica, fabricación, física, geología, ingeniería, matemáticas, medicina, meteorología, química, tecnología espacial, tratamiento de la información y, lamentablemente, aplicaciones militares.

3. *Separación conocimiento / inferencia:* En nuestra opinión, ésta es una característica presente en todos los programas de I.A., y distingue los SS.EE. de programas anteriores (por ejemplo, las ayudas al diagnóstico médico basadas en árboles de decisión), en que el conocimiento sobre el campo estaba implícito en el código del programa.

   Esta separación tiene como primera ventaja la facilidad de ampliar o depurar el conocimiento que posee el sistema, sin tener que modificar el algoritmo de inferencia.

4. *Capacidad de explicación:* Es otra ventaja de la propiedad anterior. Al tener representado explícitamente el conocimiento, el sistema puede "reflexionar" sobre la información que posee y la forma en que la ha utilizado. En capítulos posteriores insistiremos en su importancia para que el usuario final acepte el consejo que ofrece la máquina.

5. *Flexibilidad en el diálogo:* Consiste en la capacidad de generar preguntas de acuerdo con el razonamiento, tal como puede hacer —por ejemplo— un médico en su consulta, a diferencia de los programas que presentan un cuestionario rígido con gran cantidad de preguntas irrelevantes.

   (Los SS.EE. destinados a la interpretación de datos constituyen una excepción, pues en este caso toda la información está disponible desde el primer momento y no es necesario generar preguntas.)

6. *Tratamiento de la incertidumbre:* Esta característica es una consecuencia de la complejidad de los problemas que abordan los SS.EE., pues en el mundo real hay muchas fuentes de incertidumbre, asociadas a los datos (inexactitudes e imprecisiones), al dominio en sí (relaciones probabilísticas) o al conocimiento disponible (información incompleta, causas desconocidas, falta de acuerdo entre los expertos, etc.). En la sección 2.3 estudiaremos detenidamente los tipos de incertidumbre, especialmente en el campo de la medicina.

Estructura de los sistemas expertos

La estructura clásica de un sistema experto se encuentra representada en la figura 2.1. En ella se observa que el *motor de inferencia* es el elemento central encargado de coordinar todos los demás. Entre ellos, tiene especial importancia la *base de conocimiento* pues, como su nombre indica, contiene el conocimiento relativo al dominio, ya sea un campo de la medicina, de la ingeniería, etc. Generalmente consta de un conjunto de reglas, un conjunto de marcos, una red o una combinación de varios de ellos (ver la sec. 2.2). En la figura 2.1 se puede observar que hay una flecha desde la base de conocimiento hasta el motor de inferencias pero no a la inversa, lo cual indica que éste lee la base de conocimiento pero no la modifica, ya que en los sistemas expertos convencionales no hay aprendizaje.

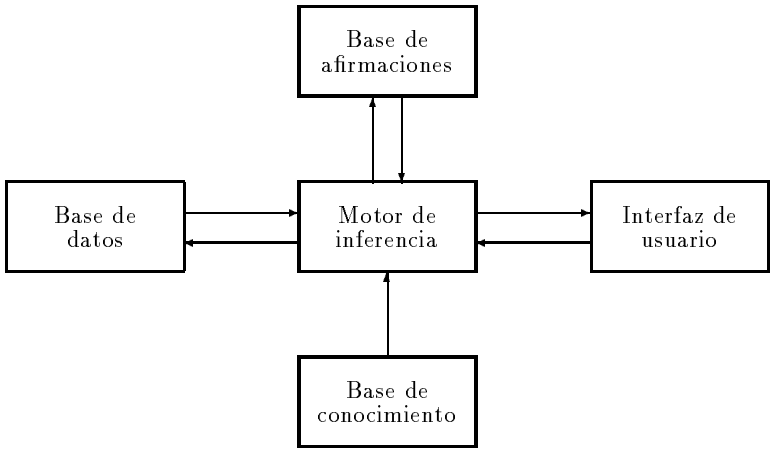


Figura 2.1: Estructura convencional de un sistema experto.

A diferencia de la base de conocimiento, que es estática, existe una *base de afirmaciones* dinámica, donde el motor de inferencia almacena las conclusiones que va

obteniendo, y a su vez busca en ella las premisas que le permitan obtener otras nuevas. El contenido de la base de afirmaciones es diferente para cada consulta que se realiza.

El tercer elemento clave de un sistema experto es el *interfaz de usuario*. Aparentemente, su importancia es secundaria. Sin embargo, la aceptación de un sistema experto por parte de sus destinatarios depende en gran medida de que el interfaz sea potente, fácil de manejar y agradable. En la sección 13.2 describiremos con detalle las características ideales de un buen interfaz.

Por último, hay un componente de poca importancia conceptual, pero que resulta imprescindible en la práctica: una *base de datos* gestionada por el programa, pues en casi todas las aplicaciones de la vida real se hace necesario almacenar y recuperar de forma eficiente grandes cantidades de información.

Conviene insistir en la diferencia entre base de datos y base de conocimiento. La primera almacena información puntual; por ejemplo, "el paciente tiene 36 años, pesa 75 Kg., mide 1'73, padeció fiebre reumática, etc.". La base de conocimiento, en cambio, contiene información sobre como relacionar los datos y los conceptos entre sí: "la fiebre reumática produce estenosis mitral en el 30% de las personas menores de 40 años".

Ésta es la estructura clásica, aplicable sobre todo a los sistemas basados en reglas. En los sistemas basados en marcos, por el contrario, tanto la información relativa al dominio como las conclusiones que obtiene el sistema se almacenan dentro de los marcos, de modo que no es posible establecer la distinción entre base de conocimiento y base de afirmaciones. Algunos sistemas basados en redes almacenan la nueva información añadiendo nodos y arcos, con lo que tampoco es aplicable para ellos la distinción anterior.

2.1.4 Ejemplos clásicos de sistemas expertos

Describimos en esta sección cinco sistemas clásicos (DENDRAL, MYCIN, PIP, CASNET y PROSPECTOR) especialmente relacionados con las redes bayesianas. Existen otros sistemas expertos famosos, tales como INTERNIST/CADACEUS en medicina, que no vamos a estudiar porque no presentan ninguna característica relevante en relación con el tema de esta tesis.

DENDRAL

Es considerado por muchos como el primer sistema experto. Fue desarrollado en la Universidad de Stanford a partir de 1965 [25, 27]. En él observamos muchas de las características básicas de los sistemas expertos: dominio reducido (su objetivo era determinar la estructura de moléculas químicas mediante espectrometría de masas), separación conocimiento / inferencia (introdujo el uso de reglas para representar el conocimiento), razonamiento simbólico (utiliza conceptos como cetona y aldehído), formación de hipótesis, búsqueda heurística (una búsqueda exhaustiva sería imposible) y alcanzó una eficacia superior a cualquier experto humano.

MYCIN

Es sin duda el sistema experto más famoso, el que ha contado con más recursos, el que ha dado lugar a más proyectos derivados, el que ha marcado el paradigma de

los sistemas expertos en la actualidad. Se desarrolló también en la Universidad de Stanford en la década de los 70 [26], como sistema de consulta para el diagnóstico y tratamiento de enfermedades infecciosas.

Su base de conocimiento contiene unas 400 reglas y su motor de inferencia utiliza encadenamiento hacia atrás (sec. 2.2.2); el tratamiento de la incertidumbre se basa la combinación de factores de certeza (sec. 2.3.2). Una de las novedades más significativas que introdujo MYCIN fue la capacidad de explicación, que comentaremos en el capítulo 8.

Una extensión de MYCIN fue TEIRESIAS [60, 61], que introdujo el uso de metareglas (es decir, reglas sobre cómo utilizar otras reglas) con tres fines: construir la base de conocimiento mediante un diálogo con el experto humano, guiar el razonamiento mediante objetivos explícitos y mejorar la capacidad de explicación a partir de dichos objetivos.

La separación conocimiento / inferencia de que hemos hablado permitió utilizar ambas componentes de forma independiente. Por una parte, la base de conocimiento fue utilizada en GUIDON para construir un sistema tutor que instruyera a los estudiantes sobre enfermedades infecciosas [43, 45]. Por otra parte, al "vaciar" el sistema de las reglas sobre enfermedades infecciosas, quedó EMYCIN (el nombre significa *Essential MYCIN*) [241, 242], que incluye las estrategias de razonamiento y las facilidades de explicación, y fue utilizado para construir nuevos sistemas expertos [15, 86].

Dos de los sistemas más importantes surgidos a partir de MYCIN fueron CENTAUR, en que Aikins combinó el uso de marcos y reglas en lo que llamó *prototipos* [2, 3], y ONCOCIN, destinado a la gestión de protocolos de oncología, en que el razonamiento temporal desempeña un papel esencial [26, págs. 604–612],[211].

CASNET

Fue desarrollado en la Universidad de Rutgers por Weiss y Kulikowski [133, 248] entre principios y mediados de los años 70 para el diagnóstico del glaucoma. A pesar de ser uno de los primeros sistemas expertos, incluye características ausentes en la mayor parte de los sistemas posteriores.

El conocimiento se representa mediante una red causal, *Causal ASociational NETwork*. Los nodos se agrupan en tres niveles: observaciones, estados patofisiológicos y enfermedades, más un cuarto grupo correspondiente a planes terapéuticos. Los arcos representan relaciones causales, temporales y jerárquicas (de clasificación). Cada relación causal tiene asociado un factor de confianza que indica —informalmente— la probabilidad de que se produzca el efecto. El diagnóstico consiste en hallar caminos que unan estados fisiológicos entre sí y con las observaciones disponibles, tratando de explicar los hallazgos y la evolución de la enfermedad.

La evaluación del sistema produjo resultados muy satisfactorios (ver las referencias anteriores) y, del mismo modo que MYCIN dió lugar a EMYCIN, CASNET dió lugar a EXPERT, una herramienta con la que se construyeron sistemas expertos para reumatología y endocrinología.

La representación del conocimiento de CASNET tiene mucha semejanza con la de las redes bayesianas en cuanto que se trata de una red causal, aunque existen en realidad grandes diferencias, que se manifiestan sobre todo en el método de razonamiento empleado: en las RR.BB. la propagación es puramente numérica, mientras que CASNET utiliza técnicas predominantemente heurísticas.

PROSPECTOR

La experiencia de MYCIN inspiró la construcción de PROSPECTOR [82, 81] en el *Stanford Research Institute*, SRI. Se trata de un sistema destinado a la exploración geológica, en el cual el conocimiento viene representado por una red de reglas y la inferencia se realiza mediante técnicas bayesianas. Sin embargo, se diferencia de las redes bayesianas principalmente en tres puntos:

1. PROSPECTOR utiliza una red de inferencia que relaciona hallazgos con hipótesis, aunque sin asignar un significado causal a los enlaces.

2. El modelo matemático de PROSPECTOR es informal (más bien deberíamos decir que es inconsistente, como veremos en la sección 2.3.1). En realidad, no se trata un sistema verdaderamente bayesiano.

3. La inferencia de PROSPECTOR examina sólo ciertas reglas, de acuerdo con la evidencia recibida, mientras que los algoritmos de redes bayesianas deben propagar la evidencia sobre toda la red.

A pesar de las deficiencias que presenta desde el punto de vista matemático, PROSPECTOR constituyó el primer éxito comercial de la I.A., pues este sistema experto ayudó a detectar un depósito de molibdeno valorado en 100 millones de dólares. (El otro gran éxito comercial fue el sistema experto R1, también llamado XCON, que aún se utiliza en *Digital Equipment Corporation* para configurar las estaciones de trabajo VAX [144].)

PIP

El sistema experto PIP (*Present Illness Program*) se desarrolló durante los años 70 en el M.I.T. [168, 234]. Su interés en relación con la presente tesis se debe a que es el primer sistema experto basado en marcos (sec. 2.2.4). La figura 2.2 muestra un ejemplo en que podemos observar los diferentes campos que caracterizan una determinada entidad clínica, el síndrome nefrótico. Observamos que hay diferentes tipos de hallazgos relativos a una hipótesis; algunos de ellos bastan por sí mismos para confirmarla ("es-suficiente") o para descartarla ("no-puede-haber"), mientras que otros sólo aportan una cierta evidencia que debe ser ponderada en el contexto de los demás hallazgos. Esta distinción refleja los dos tipos de razonamiento médico que Szolovits y Pauker identifican en la práctica médica: categórico y probabilístico [234].

En cuanto al control del razonamiento, hay tres estados en que puede encontrarse un marco. Inicialmente, todos ellos están *inactivos*. Cuando se introduce en el sistema un determinado hallazgo, los marcos correspondientes a las enfermedades que podrían explicarlo son *activadas* y los marcos relacionados con ellos (por ejemplo, mediante los campos "puede-complicarse-por" o "diagnóstico-diferencial"), pasan a estar *semiactivos*. Con ello, se pretender simular el comportamiento del médico que investiga una las razones a favor o en contra de una determinada enfermedad (un marco activo), aunque teniendo en cuenta otras enfermedades relacionadas con ella (marcos semiactivos), hasta llegar finalmente a una conclusión que englobe toda la información disponible.

Nombre: Síndrome nefrótico

Es-un-tipo-de: Estado clínico
Hallazgo: Concentración de albúmina baja
Hallazgo: Proteinuria intensa
Hallazgo: Proteinuria > 25 gramos / 24 horas
Hallazgo: Edema simétrico masivo
Hallazgo: Edema simétrico y facial o peri-orbitario
Hallazgo: Concentración de colesterol elevada
Hallazgo: Lípidos en la orina presentes
No-debe-haber: Proteinuria ausente
Es-suficiente:
Edema masivo y proteinuria > 25 gramos / 24 horas
Criterios mayores:
Concentración de sero-albúmina
Baja: 1.0
Alta: -1.0
Proteinuria
> 25 gramos / 24 horas: 1.0
Intensa: 0.5
Ausente o leve: -1.0
Edema
[...]
Criterios menores:
Concentración de colesterol
Elevada: 1.0
No elevada: -1.0
Lípidos en la orina
Presentes: 1.0
Ausentes: -0.5
Puede-estar-producida-por:
Glomerulonefritis aguda
Glomerulonefritis crónica
Drogas nefrotóxicas
Síndrome nefrótico idiopático
[...]
Puede-complicarse-por:
Hipovolemia
Celulitis
Puede-producir:
Retención de sodio
Diagnóstico-diferencial:
Si presión venosa yugular elevada,
considerar: pericarditis constrictiva.
Si ascitis presente,
considerar: cirrosis.
[...]

Figura 2.2: Ejemplo de marco utilizado en el sistema experto PIP.

Otros sistemas expertos

En la Universidad de Santiago de Compostela se han desarrollado varios sistemas expertos, que nos han servido de inspiración en este proyecto. El programa CAEMF, desarrollado por Roque Marín en 1987, es —según nuestros conocimientos— el primer sistema experto español [138]. Un año más tarde, Senén Barro presentó un sistema de monitorización inteligente para unidades de cuidados coronarios, que nos resultó útil como referencia por su relación con la cardiología [11]. Dentro del mismo grupo de trabajo, Ramón Pérez Otero diseñó la herramienta MEDTOOL [164, 165], que ha sido utilizada posteriormente para la implementación de TAO, un sistema consejero para oncología [9, 10]. En el capítulo 13 mencionaremos el interfaz gráfico de dicho programa, que aparece descrito en [139] y [235].

Dentro de los sistemas expertos basados en redes bayesianas, los más conocidos son PATHFINDER y MUNIN, cuyo avanzado desarrollo ha dado lugar a productos comercializados. Hablaremos de ambos sistemas con más detenimiento en la sección 10.1.4.

2.1.5 Ventajas y limitaciones de los sistemas expertos

Entre las principales ventajas, podemos señalar las siguientes (adaptando la descripción de [246, págs. 12–15]):

- *Permanencia:* Los expertos humanos pueden morir, cambiar de empresa o perder facultades por la edad o la falta de ejercicio, malográndose así el fruto de largos años de formación, problema que no existe en absoluto para un sistema experto.

- *Duplicación:* El experto humano se encuentra en un único lugar físico y es, obviamente, irreproducible, mientras que cuando hemos construido un sistema experto podemos fabricar un número ilimitado de copias destinadas a todos los lugares donde sean necesarias.

- *Fiabilidad:* Un sistema experto responderá siempre de la misma manera ante un cierto problema, mientras que un experto humano puede estar condicionado por factores emocionales, prejuicios personales, tensión, fatiga, etc.

- *Rapidez:* Un experto humano razona con rapidez cuando se trata de las cuestiones que resuelve cotidianamente, pero necesita un tiempo considerable cuando se le presenta un caso poco frecuente y debe consultar información tal como informes anteriores, datos de catálogo o referencias bibliográficas;[3] sin embargo, un sistema experto puede realizar casi instantáneamente no sólo el cálculo numérico y la gestión de bases de datos (historia del paciente, datos de catálogo), sino también la consulta de bases de conocimiento sobre temas especializados.

[3] Por ejemplo, observé en cierta ocasión que, ante un enfermo procedente de un país tropical, una doctora de medicina general declaró que necesitaba consultar un libro especializado —que en ese momento no se encontraba en el consultorio— para poder diagnosticar o rechazar cierta enfermedad.

- *Bajo coste:* Es cierto que puede resultar caro inicialmente construir un sistema experto, pero una vez construido produce grandes beneficios. Además de la posibilidad de duplicación antes citada, los sistemas expertos funcionan generalmente en computadores de precio reducido y pueden estar activos 24 horas al día, incluso en ambientes perjudiciales y lugares peligrosos.

A pesar de ello, presentan también grandes carencias:

- *Sentido común:* Para un ordenador no hay nada obvio. Un sistema experto médico puede admitir fácilmente que hay un paciente varón de 80 g. de peso y 175 m. de estatura que está embarazado desde hace tres años, salvo que el diseñador del programa haya indicado explícitamente que ésos son datos absurdos. Por más que nos esforcemos en depurar la base de conocimiento, nunca podremos estar completamente seguros de que el sistema experto no va a dar un resultado contrario al sentido común más elemental.

- *Flexibilidad:* La flexibilidad de un ser humano es ilimitada; inmediatamente reconoce cuándo se trata de una excepción o cuándo el caso excede sus conocimientos, y puede adaptarse a situaciones inesperadas buscando nuevas soluciones, mientras que, a pesar de los continuos progresos que se están realizando, los sistemas expertos de la actualidad y del futuro próximo seguirán siendo terriblemente rígidos.

- *Lenguaje natural:* Todavía estamos muy lejos de tener un sistema que pueda formular preguntas flexibles y mantener una conversación informal con un usuario o con un paciente. Lo mismo podemos decir en cuanto a la capacidad de generar explicaciones e informes.

- *Experiencia sensorial:* Casi todos los sistemas expertos de la actualidad se limitan a recibir información mediante un teclado y un ratón. Sólo algunos disponen de tratamiento de imágenes, y aun éste resulta ridículo comparado con la capacidad visual humana, por no hablar de la capacidad auditiva, táctil y olfativa.

- *Perspectiva global:* Un experto humano es capaz de detectar inmediatamente con gran fiabilidad cuáles son las cuestiones centrales y cúales son secundarias, y saber que hay detalles que encierran la clave del problema mientras que una infinidad de ellos es realmente insignificante.

En el caso de la ecocardiografía se observan claramente todas estas limitaciones y algunas más. Un médico experto es capaz rechazar a primera vista datos absurdos, de adaptarse a casos extraños, y mientras realiza el eco puede hablar con el paciente, captar el latido del corazón, observar su peso, estatura, color de la piel y mil detalles más, considerar su situación familiar y laboral, comprender su estado anímico, tranquilizarle cuando sea necesario, y tener en cuenta una cantidad ilimitada de información que nunca estará disponible para un sistema experto. Por tanto, la idea de sustituir un médico por un ordenador, a parte de todos los problemas jurídicos, éticos y psicológicos que pueda plantear, es irrealizable hoy en día y seguirá siéndolo, al menos durante muchos años.

2.1.6 Juicio crítico

En 1967, Marvin Minsky afirmó que "en una generación, el problema de crear inteligencia artificial estará resuelto" [148, pág. 2]. Hoy en día las expectativas son mucho más limitadas. A medida que conocemos más sobre la inteligencia, caemos en la cuenta de lo compleja que es la mente humana y lo difícil que resulta duplicar incluso algunas de las capacidades aparentemente más sencillas.

Si miramos retrospectivamente la historia de la I.A., podemos concluir que los mayores fracasos se han dado al intentar abarcar problemas generales; el proyecto de traducción automática que se desarrolló en EE.UU. en la década de los 60 constituye el más famoso de ellos, por las expectativas puestas en él, por el dinero invertido y por los disparates que se generaban en las traducciones. Igualmente, el *Proyecto de la Quinta Generación*, desarrollado en Japón entre 1982 y 1992 con un presupuesto de mil millones de dólares, ha quedado muy lejos de alcanzar sus objetivos.

Por el contrario, la I.A. ha cosechado éxitos notables al reducir su campo de aplicación. El más destacado, tanto por los resultados obtenidos como por la repercusión comercial y por la cantidad de investigación generada, lo constituyen los sistemas expertos. En el campo de la medicina son varios los sistemas que han alcanzado una competencia comparable o incluso superior a la de expertos humanos.

Por otro lado, hoy en día ya no se considera importante que las máquinas dupliquen o superen las capacidades del ser humano, sino que colaboren con él, aprovechando las cualidades que cada uno posee. En este sentido, es significativa la afirmación que aparece como "información para los autores" en la revista *Knowledge Acquisition*:[4]

> Lo importante no es la inteligencia *artificial*, sino la extensión de la inteligencia *natural* mediante sistemas basados en conocimiento.

Por tanto, al desarrollar en esta tesis un proyecto de inteligencia artificial, no tratamos de construir un sistema experto que sustituya a los médicos; tan sólo aspiramos a poner en sus manos un instrumento que les ayude a realizar el diagnóstico con mayor rapidez y fiabilidad.

2.2 Representación del conocimiento y razonamiento

En los primeros años de la I.A., se insistió sobre todo en buscar mecanismos de razonamiento potentes y eficientes aplicables a todo tipo de problemas. Posteriormente, fue descubriéndose de forma progresiva el valor del conocimiento que posee el sistema y, por tanto, la importancia de la forma de representarlo, hasta el punto de que hay quien considera que ésta es la piedra angular de la inteligencia artificial [20]. El aprendizaje automático es otro de los aspectos en que la representación del conocimiento desempeña un papel esencial.

Brachman y Levesque señalan la necesidad de buscar un compromiso entre *expresividad* y *eficiencia* al escoger el método de representación [20]. La expresividad puede definirse *grosso modo* como la capacidad de realizar afirmaciones no explícitas. Así, el "lenguaje" de una base de datos es muy poco expresivo: no podemos afirmar "hay al menos una persona con ingresos superiores a 150.000 ptas. mensuales" a menos

[4] Tomado de W. J. Clancey [46]. Cursiva en el original.

que indiquemos quién es ésa persona y cuáles son sus ingresos exactos. La lógica de predicados, en el otro extremo, es mucho más expresiva, pero resulta completamente ineficiente incluso en problemas muy reducidos.

El debate sobre cuál es el método de representación más adecuado está lejos de llegar a una conclusión. Por ello, gran parte de la investigación actual ha adoptado una tendencia pragmática, que consiste en estudiar qué combinación de los métodos existentes puede resultar más conveniente en cada caso concreto. Es lo que se conoce como representaciones híbridas, de las que hablaremos al final de esta sección. (Muchos de los artículos más importantes sobre este tema se encuentran recogidos en [20]. Para una introducción sencilla, ver [189].)

2.2.1 Lógica

El empleo de la lógica en I.A. se remonta a los orígenes de este campo, como hemos mencionado anteriormente. Sin embargo, el decidir hasta qué punto la lógica resulta válida y útil para resolver problemas del mundo real sigue siendo tema de intenso debate en el que no vamos a entrar. (Dos referencias recientes que representan las dos tendencias opuestas son [160] y [19].)

En I.A. se ha utilizado sobre todo la lógica de predicados, desarrollada por Frege en el siglo pasado, pues la lógica proposiciones, de Boole, es tan sencilla que resulta poco expresiva. La lógica de predicados, en cambio, admite los cuantificadores universal ($\forall$) y existencial ($\exists$).

La principal ventaja de la lógica es su expresividad, frente a otros métodos de representación, en que algunos problemas ni siquiera pueden enunciarse. Su mayor desventaja desde el punto de vista práctico es la ineficiencia, debida a la falta de estructuración. Otra gran desventaja práctica es que resulta difícil y engorroso, casi imposible, tratar la incertidumbre en la lógica de predicados.

Existen en la actualidad numerosas extensiones de la lógica; las más importantes dentro de la I.A. son las pertenecientes a las ramas modales, no-monótonas y multivaluadas. (Para no incluir aquí una lista interminable de referencias, remitimos al lector a la *Encyclopedia of Artificial Intelligence* [206].)

PROLOG

Por último, dentro de este apartado suele citarse el lenguaje PROLOG, diseñado por A. Colmeraruer y P. Roussel en la Universidad de Marsella (Francia) hacia 1970 a partir de las ideas de R. A. Kowalski sobre *programación lógica*, la cual pretende ser una implementación de la lógica de predicados, aunque con importantes diferencias:

- Este lenguaje no incluye toda la expresividad de la lógica de predicados, sino que sólo admite cláusulas de Horn (cláusulas del tipo $P_1 \wedge \ldots \wedge P_n \Rightarrow C$).

- La deducción de PROLOG se basa en la *hipótesis del mundo cerrado*: toda afirmación que no se encuentra entre las afirmaciones explícitas ni se puede deducir de ellas, se considera falsa.

- Para que el razonamiento sea mínimamente eficiente en problemas reales hace falta incluir instrucciones de control, con lo que se pierde la ventaja más pregonada de la programación lógica: su carácter puramente declarativo.

- Otra cuestión, también relacionada con el problema del control de la ejecución, es que el orden de la inferencia depende de la ordenación de las cláusulas, por lo que la eficiencia exige tener muy en cuenta este factor a la hora de escribir un programa; éste es otro motivo por el que se pierde el carácter puramente declarativo de la programación lógica.

2.2.2 Reglas

Como acabamos de explicar, el principal problema de la lógica es que resulta muy ineficiente incluso en problemas relativamente reducidos. Por eso surgieron los sistemas basados en reglas, que sacrifican la expresividad a cambio de la eficacia. Aunque las reglas se habían utilizado ya en sistemas anteriores, fue MYCIN el primer sistema experto que las utilizó en la forma en que hoy es habitual. Según Gregorio Fernández [88],

> Desde el punto de vista de la lógica formal, [MYCIN] utiliza lógica de proposiciones con algunas extensiones que la aproximan a la lógica de predicados (existencia de predicados predefinidos y ciertas posibilidades cuantificar) y a las lógicas multivaluadas (factores de certidumbre).

El procedimiento de inferencia en MYCIN consiste en buscar una regla que nos permita confirmar la hipótesis buscada. Por ejemplo, para saber si un organismo es bacteroide escogemos la regla siguiente [240, pág. 71]:

```
($AND (SAME CNTXT GRAM GRAMNEG)
      (SAME CNTXT MORPH ROD)
      (SAME CNTXT AIR ANAEROBIC))
(CONCLUDE CNTXT IDENTITY BACTEROIDES TALLY .6)
```

que significa

SI el organismo es gram-negativo
Y tiene forma de bastón
Y es anaeróbico
ENTOCES el organismo es bacteroide (con una certeza de 0'6).

(Observar que el término `CNTXT` corresponde a una variable que se asociará al organismo correspondiente en cada caso. Como hemos mencionado anteriormente, ésta es una de las características que sitúa las reglas más cerca de la lógica de predicados que de la lógica proposicional.)

Para poder ejecutar la regla y concluir que se trata de un organismo bacteroide, hace falta confirmar cada una de sus premisas. El sistema trata primero de ver si es gram-negativo buscando en la base de afirmaciones otra regla de la que podamos deducirlo (y así sucesivamente, en lo que se conoce como *encadenamiento hacia atrás* de las reglas) o preguntando al usuario que está introduciendo la información. Si se confirma la primera premisa, el sistema tratará de confirmar la segunda, y si se confirman todas, concluirá que es bacteroide; si no, examinará otras reglas.

El encadenamiento hacia atrás corresponde a una inferencia *basada en hipótesis*, pues son éstas el objetivo que guía la búsqueda. Existe también la inferencia *basada*

en datos: un dato confirmado puede disparar una o más reglas, lo cual puede provocar que se disparen otras, y así sucesivamente; este proceso se conoce como *encadenamiento hacia delante* (es el que utiliza, por ejemplo, el sistema PROSPECTOR).

Las reglas constituyen la representación estándar para los sistemas expertos actuales. A pesar de sus innegables ventajas, a la hora de aplicarlas en la práctica resultan no ser tan declarativas ni tan modulares como proclaman sus defensores. En la sección 2.4 trataremos la crítica de las reglas, tanto desde el punto de vista de la representación del conocimiento como en lo relativo al tratamiento de la incertidumbre.

2.2.3 Redes asociativas

A finales de los 60, Ross Quillian [184] introdujo las redes en I.A. como modelo asociativo de la memoria humana, concretamente en el campo del lenguaje natural (comprensión y traducción); para ello, desarrolló redes que representaran el significado de las palabras (vocabulario inglés). Un método similar fue utilizado por Patrick Winston [261] en su sistema de aprendizaje aplicado a visión artificial (el ejemplo más típico es la representación de un arco); por Jaime Carbonell, en el sistema SCHOLAR [31]; por Roger Shank [204], cuyo objetivo era representar conceptos, en lugar de los términos de un idioma particular; por John Sowa [213], en sus *grafos conceptuales* y por muchos otros [214].

Las redes de Quillian fueron llamadas *redes semánticas* porque representaban el significado de las palabras de la lengua inglesa. Otros formalismos inspirados en su trabajo suelen ser llamados también redes semánticas, aunque —en nuestra opinión— resulta incorrecto utilizar dicho término cuando se trata, por ejemplo, de *redes de clasificación* o de *redes causales*. La denominación adecuada para los sistemas de representación que relacionan conceptos mediante nodos y arcos es *redes asociativas*.

Las dos características que, a nuestro juicio, distinguen los distintos tipos de redes son:

Tipo de conocimiento representado: Las redes semánticas, como su propio nombre indica, representan el *significado* de los conceptos. En consecuencia, sus arcos corresponden a relaciones lingüísticas, espaciales, temporales, causales y de muchos otros tipos, por lo que llevan etiquetas como "produce", "es-parte-de", "pertenece-a", "desemboca-en", etc.[5] Las redes de clasificación, en cambio, sólo admiten dos tipos de arcos: "instancia-de" y "subclase-de". En las redes causales hay un solo tipo de arco, que va siempre de la causa al efecto. (La red de CASNET mencionada anteriormente, las redes bayesianas y las redes cualitativas (sec. 10.1.3) son casos particulares de redes causales.)

Método de inferencia: Como consecuencia de lo anterior, las redes semánticas tienen un mecanismo de inferencia propio, basado en el significado de cada arco. Por ejemplo, un arco del tipo "pertenece-a" y otro del tipo "desemboca-en" darán lugar a conclusiones muy diferentes. Las redes de clasificación, en cambio, suelen utilizarse principalmente como forma estructurada de almacenar información, pues el único mecanismo de razonamiento que admiten por

[5] Existen redes semánticas con un solo tipo de arco, que actúa simplemente como conector, sin otro significado; pero no vamos a hablar de ellas aquí para no complicar la exposición.

sí mismas es la herencia de propiedades y valores; en estos modelos, la mayor parte de la inferencia suele realizarse por otro método, como puede ser un conjunto de reglas (sistemas híbridos). Las redes causales, por su parte, utilizan mecanismos de inferencia propios, que nada tienen que ver con los anteriores.

El concepto de red es tan importante y tan útil que aparece con frecuencia de una u otra forma en todos los tipos de representación avanzados. Veamos a continuación un ejemplo, los marcos.

2.2.4 Marcos

El concepto de *marco* (*frame*) fue introducido por Marvin Minsky en 1975 [149] como forma estructurada y flexible de representar el conocimiento, con el fin de reproducir la eficiencia del sentido común en los problemas del mundo real. A pesar del tiempo transcurrido —o quizá debido a ello— el significado y la utilización de los marcos también es objeto de un intenso debate (ver referencias en [20] y [189, cap. 4]).

En la página 19 ya hablamos del programa PIP, el primer sistema experto basado en marcos. En la figura 2.2 mostrábamos un ejemplo de cómo se pueden utilizar los marcos para representar el conocimiento médico de cara al diagnóstico.

En nuestra opinión, existen básicamente dos modos en que los marcos son utilizados:

1. *Patrones para la comparación (pattern matching).* En este caso, los marcos son fijos, en el sentido de que la información que contienen no se modifica. Podemos definir, por ejemplo, cada enfermedad como una instancia. El método de diagnóstico consiste en activar las instancias cuyas propiedades coincidan con la información disponible y desactivar las que son inconsistentes con los datos hasta seleccionar aquélla que mejor se ajusta a los datos.

2. *Almacenes dinámicos de información.* Dentro de esta línea hay un rango muy amplio de posibilidades, dependiendo de la capacidad de procesamiento asociada a cada instancia. En un extremo, éstos constituyen una especie de armarios en cuyos cajones (*slots*) se almacenan los datos, separados del mecanismo de inferencia, que suele venir dado por un conjunto de reglas; la representación correspondiente a los marcos sería entonces equivalente a una lógica de predicados sin cuantificadores (casi una lógica proposicional), pero con una estructura fija para la información, semejante a una base de datos relacional. En el otro extremo, cada marco (o cada instancia) se convierte en un procesador conceptual que realiza toda la inferencia, al estilo de la programación orientada a objetos. Existen además sistemas híbridos, como los *prototipos* de Janice Aikins [2, 3], en que las reglas se almacenan en las propiedades (*slots*) de las instancias, de donde pasan al motor de inferencias.

Una de las características más típicas de los marcos, tal como se conciben en la actualidad, es su ordenación jerárquica (una relación de orden parcial), en que un nodo superior indica una clase más amplia. Los que están por debajo no sólo *heredan* las propiedades de sus superiores, sino que pueden heredar tambien valores y métodos (procedimientos) asociados a ellas. La herencia de valores constituye así un mecanismo de *razonamiento por defecto*, muy típico de los sistemas basados en marcos y de la programación orientada a objetos.

2.2.5 Sistemas híbridos

Como hemos dicho al principio de esta sección, no parece que el debate para decidir cuál es el mejor método de representación vaya a llegar una conclusión. Probablemente incluso resulta erróneo plantear así la cuestión, y sería más productivo estudiar problema por problema qué método —o qué combinación de métodos— se adecúa mejor a cada caso. Hoy en día hay pocas aplicaciones basadas en un sólo tipo de representación. Así, la mayor parte de los sistemas expertos están construidos a base de reglas, aunque utilizan también objetos que, en mayor o menor medida, utilizan muchas de las ideas propuestas para los marcos, y a su vez tanto las reglas como los objetos suelen agruparse en una red o en varias de ellas simultáneamente atendiendo a diferentes criterios, como es el caso del sistema experto desarrollado en esta tesis, que combina una red de clasificación con una red causal (ver la sec. 13.1).

2.3 Tratamiento de la incertidumbre

Al hablar de la inteligencia artificial, dijimos que el tratamiento de la incertidumbre constituye uno de los campos fundamentales, que afecta en mayor o menor medida a todos los demás. En particular, al definir las características esenciales de los sistemas expertos, vimos que una de las más típicas y complejas es el tratamiento de la incertidumbre. Las fuentes más habituales de incertidumbre pueden agruparse en tres tipos: deficiencias de la información, características del mundo real y deficiencias del modelo. Podemos ver que en medicina se dan todas ellas:

- *Información incompleta.* En muchos casos la historia clínica completa no está disponible, y el paciente es incapaz de recordar todos los síntomas que ha experimentado y cómo se ha desarrollado la enfermedad. Además, en otras ocasiones, las limitaciones prácticas impiden contar con todos los medios que deberían estar disponibles, por lo que el médico debe realizar su diagnóstico en base a la información que posee, aunque ésta sea muy limitada.

- *Información errónea.* En cuanto a la información suministrada por el paciente, puede que éste describa incorrectamente sus síntomas e incluso que trate de mentir deliberadamente al médico. También es posible que el diagnóstico anterior, contenido en la historia clínica, haya sido erróneo. Y tampoco es extraño que las pruebas de laboratorio den falsos positivos y falsos negativos. Por estas razones, el médico siempre debe mantener una duda razonable frente toda la información disponible.

- *Información imprecisa.* Hay muchos datos en medicina que son difícilmente cuantificables. Tal es el caso, por ejemplo, de los síntomas como el dolor o la fatiga. Incluso en un método tan técnico como la ecocardiografía hay muchas observaciones que en la práctica deben ser cuantificadas subjetivamente, como son el prolapso valvular o la aquinesia ventricular.

- *Mundo real no determinista.* A diferencia de las máquinas mecánicas o eléctricas, cuyo funcionamiento se rige por leyes deterministas, los profesionales de la medicina comprueban a diario que cada ser humano es un mundo, en que las leyes generales no siempre resultan aplicables. Muchas veces las mismas

causas producen efectos diferentes en distintas personas, sin que haya ninguna explicación aparente. Por ello, el diagnóstico médico debe estar siempre abierto a tratar con la probabilidad y con las excepciones.

- *Modelo incompleto.* Por un lado, hay muchos fenómenos médicos cuya causa aún se desconoce. Por otro, es frecuente la falta de acuerdo entre los expertos de un mismo campo. Finalmente, aunque toda esta información estuviera disponible, sería imposible, por motivos prácticos, incluirla en un sistema experto.

- *Modelo inexacto.* Por último, todo modelo que trate de cuantificar la incertidumbre, por cualquiera de los métodos que existen, necesita incluir un elevado número de parámetros; por ejemplo, en el caso de las redes bayesianas, necesitamos especificar todas las probabilidades a priori y condicionales. Sin embargo, esta información nunca está disponible, por lo que debe ser estimada de forma subjetiva. Es deseable, por tanto, que nuestro modelo de razonamiento pueda tener en cuenta sus propias inexactitudes.

En resumen: el tratamiento de la incertidumbre es esencial en el diseño de sistemas expertos, muy especialmente en el caso de la medicina. Veamos a continuación cuáles son los métodos más utilizados. Puede encontrarse una revisión más extensa en [158].

2.3.1 Método probabilístico clásico

El desarrollo de programas de diagnóstico basados en técnicas bayesianas comenzó en los años 60. Entre los sistemas de esa década destacan el de Warren, Toronto y Veasy para el diagnóstico de cardiopatías congénitas [245], los de Gorry y Barnett [95, 96] y el programa creado por de Dombal y sus colaboradores para el diagnóstico del dolor abdominal agudo [65].

El fundamento de estos sistemas se encuentra en la aplicación del teorema de Bayes. Supongamos que tenemos un conjunto de diagnósticos, $d_1, \ldots, d_n$ y un conjunto de hallazgos o evidencias $e_1, \ldots, e_m$. La probabilidad del conjunto de hipótesis dado el conjunto de evidencia es, según dicho teorema,

$$P(d_1, \ldots, d_n | e_1, \ldots, e_m) = \frac{P(e_1, \ldots, e_m | d_1, \ldots, d_n) \cdot P(d_1, \ldots, d_n)}{\sum_{d'_1, \ldots, d'_n} P(e_1, \ldots, e_m | d'_1, \ldots, d'_n) \cdot P(d'_1, \ldots, d'_n)} \tag{2.1}$$

Sin embargo, esta expresión es imposible de aplicar por la enorme cantidad de información que requiere: necesitaríamos conocer todas las probabilidades a priori $P(d_1, \ldots, d_n)$ y todas las probabilidades condicionales $P(e_1, \ldots, e_m | d_1, \ldots, d_n)$.

Por ello se introduce la *hipótesis* de que los diagnósticos son exclusivos (no puede haber dos de ellos a la vez) y exhaustivos (no hay otros diagnósticos posibles). Hay, por tanto, una variable de n valores que corresponde al diagnóstico, y hay m variables —binarias en general— correspondientes a los posibles hallazgos, tal como muestra la figura 2.3. En consecuencia, la probabilidad de un diagnóstico cualquiera, d_i, se expresa como

$$P(d_i | e_1, \ldots, e_m) = \frac{P(e_1, \ldots, e_m | d_i) \cdot P(d_i)}{\sum_j P(e_1, \ldots, e_m | d_j) \cdot P(d_j)} \tag{2.2}$$

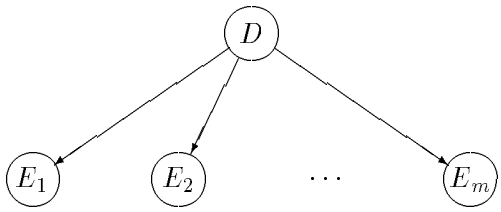


Figura 2.3: Método clásico de diagnóstico probabilístico.

Aun así, la fórmula resulta inaplicable incluso en problemas reducidos, debido a la necesidad de conocer las probabilidades condicionales $P(e_1, \ldots, e_m|d_j)$ para todas las combinaciones posibles de hallazgos. Por tanto, se hace necesario introducir una nueva *hipótesis*, la de *independencia condicional*: los hallazgos (los síntomas, podríamos decir) son independientes entre sí para cada diagnóstico.[6]

En forma matemática, se expresa así:

$$P(e_1, \ldots, e_m|d_i) = P(e_1|d_i) \cdot \ldots \cdot P(e_m|d_i), \qquad \forall d_i \tag{2.3}$$

Con esto tenemos por fin una fórmula con un número mucho más reducido de parámetros,

$$P(d_i|e_1, \ldots, e_m) = \frac{P(e_1|d_i) \cdot \ldots \cdot P(e_m|d_i) \cdot P(d_i)}{\sum_j P(e_1|d_j) \cdot \ldots \cdot P(e_m|d_j) \cdot P(d_j)} \tag{2.4}$$

A partir de ella, resulta muy sencillo comparar la probabilidad de dos diagnósticos, con la posibilidad de incorporar secuencialmente nuevos hallazgos,

$$\frac{P(d_i|e_1, \ldots, e_m)}{P(d_j|e_1, \ldots, e_m)} = \frac{P(e_1|d_i)}{P(e_1|d_j)} \cdot \ldots \cdot \frac{P(e_n|d_i)}{P(e_n|d_j)} \cdot \frac{P(d_i)}{P(d_j)} \tag{2.5}$$

Esto es lo que se conoce como método clásico de diagnóstico probabilístico.

Discusión

Aunque este método sirvió de base a los sistemas de diagnóstico ya citados, con resultados satisfactorios para pequeños problemas, presenta serias deficiencias:

1. La hipótesis de diagnósticos exclusivos y exhaustivos es pocas veces aplicable en casos reales [210]. En la mayor parte de los problemas de diagnóstico en medicina, por poner un ejemplo, pueden darse dos enfermedades simultáneamente, con lo que el método clásico resulta totalmente inadecuado. Por otra parte, suele ser muy difícil o imposible especificar todas las causas que pueden producir un conjunto de hallazgos.

2. Igualmente, la hipótesis de independencia condicional, tal como se introduce en el método clásico, es muy cuestionable [234]. Normalmente, los hallazgos correspondientes a cada diagnóstico están fuertemente correlacionados, por lo

[6] Para una exposición más detallada de la independencia condicional, con discusión crítica, ver el capítulo 4.

que dicha hipótesis resulta inadmisible, pues lleva a sobreestimar la importancia de los hallazgos asociados entre sí. (Más adelante veremos que las redes bayesianas resuelven este problema introduciendo causas intermedias, con lo que la hipótesis de independencia condicional resulta ser mucho más razonable.)

3. Además, sigue existiendo el problema de la gran cantidad de parámetros necesarios en el modelo, incluso después de introducir las dos hipótesis anteriores. Por ejemplo, si queremos considerar n_d diagnósticos y n_e hallazgos posibles, necesitamos $n_d \cdot n_e$ probabilidades condicionales, $P(e_j|d_i)$, más n_d probabilidades a priori, $P(d_i)$, con lo que necesitamos un total de $n_h \cdot (n_e + 1)$ parámetros. Si pensamos en el número de diagnósticos existentes y, sobre todo, en la cantidad de hallazgos habituales en cualquier campo de la medicina, vemos que el número de parámetros resulta imposible de tratar incluso limitando considerablemente el alcance de nuestro modelo.

4. Por último, desde el punto de vista de la construcción de sistemas expertos, el método probabilístico clásico presenta el inconveniente de que la información no está estructurada, lo cual conlleva dos consecuencias negativas: por un lado, resulta complicado el mantenimiento de la base de conocimiento, ya que ésta consiste exclusivamente en un montón de parámetros; por otro lado, la falta de estructura impide explicar por qué se ha escogido un diagnóstico y se han descartado los demás.

A lo largo de esta tesis iremos viendo cómo las redes bayesianas resuelven cada uno de estos problemas, tratando de demostrar que las objeciones tradicionales contra el uso de la probabilidad en inteligencia artificial han perdido ya su validez.

Método bayesiano subjetivo

Antes de concluir esta sección dedicada a los métodos probabilísticos, debemos mencionar otro modelo bayesiano, importante por su aplicación en el sistema experto PROSPECTOR, del que ya hemos hablado en la sección 2.1.4. El modelo matemático se encuentra descrito en el artículo original de Duda, Hart y Nilsson [82].

Como dijimos anteriormente, PROSPECTOR es un sistema experto basado en reglas de producción. Para el tratamiento de la incertidumbre, a cada una de ellas se le añaden dos parámetros probabilísticos, que indican suficiencia y necesidad. Combinando los factores de cada regla con los de sus premisas, obtenemos dos nuevos factores para la conclusión correspondiente (uno mide la evidencia a favor de la conclusión y otro la evidencia en contra), que serán propagados sucesivamente al disparar nuevas reglas. Como estas probabilidades y factores proceden de estimaciones subjetivas realizadas por el experto, la red será en general inconsistente (es decir, contendrá infomación contradictoria), por lo que es necesario recurrir a métodos de interpolación para obtener aproximaciones razonables.

Al igual que en el método clásico, este modelo recurre a la independencia condicional para poder actualizar la probabilidad, aunque aquí la hipótesis no se aplica de forma global a toda la evidencia disponible, sino a los distintos hallazgos relacionados con una misma hipótesis.

Podemos concluir, por tanto, que el método bayesiano subjetivo, aunque basado en la teoría de la probabilidad, se aparta de ella debido a las dificultades prácticas

que ésta presenta. En la página 19 ya mencionamos que, a pesar de las apariencias, la filosofía y el método de las redes bayesianas son completamente diferentes de los utilizados en PROSPECTOR.

2.3.2 Factores de certeza (MYCIN)

Como consecuencia de las dificultades que presentaba el método probabilístico clásico, los creadores del programa MYCIN diseñaron un modelo *ad hoc* que, en vez de buscar un fundamento teórico, trataba de reproducir la forma en que el ser humano combina intuitivamente distintas fuentes de información.

La idea básica es la misma que en PROSPECTOR: utilizar dos factores, $MB(H,E)$ y $MD(H,E)$, que miden la fuerza con que la evidencia E actúa a favor o en contra de la hipótesis H, respectivamente. De estas dos medidas se deduce un único factor de certeza, CF, que se utiliza en la propagación de las reglas:

$$CF(H,E) \equiv MB(H,E) - MD(H,E) \tag{2.6}$$

Este factor puede calcularse directamente a partir de las probabilidades $P(H)$ y $P(H|E)$ mediante la siguiente expresión [99, 209]:

$$CF(H,E) = \begin{cases} \dfrac{P(H|E) - P(H)}{1 - P(H)} & \text{si } P(H|E) \geq P(H) \\ \dfrac{P(H|E) - P(H)}{P(H)} & \text{si } P(H|E) < P(H) \end{cases} \tag{2.7}$$

Cuando tenemos una conjunción de premisas como antecedente de una regla, la medida de evidencia a favor de la conclusión viene dada por el mínimo de las medidas correspondientes a las premisas; para medir la evidencia en contra, se toma el máximo:

$$MB(H_1 \wedge H_2, E) = \min(MB(H_1,E)\,,\, MB(H_2,E)) \tag{2.8}$$

$$MD(H_1 \vee H_2, E) = \max(MD(H_1,E)\,,\, MD(H_2,E)) \tag{2.9}$$

Para una disyunción de premisas se intercambian el papel del máximo y del mínimo.

Si hay varias reglas que aportan evidencia a favor o en contra de una hipótesis, hace falta una fórmula que indique cómo combinar los factores de certeza correspondientes. Buchanan y Shortliffe [26, cap. 10] explican por qué la fórmula original utilizada en MYCIN resultaba insatisfactoria y cómo fue sustituida por la fórmula de van Melle [242]:

$$CF_{comb}(x,y) = \begin{cases} x + y(1-x) & \text{si } x > 0 \text{ e } y > 0 \\ \dfrac{x+y}{1 - \min|x|,|y|} & \text{si } x \cdot y < 0 \\ -CF_{comb}(-x,-y) & \text{si } x < 0 \text{ e } y < 0 \end{cases} \tag{2.10}$$

Discusión

A pesar del éxito que obtuvo MYCIN en cuanto a la calidad de sus resultados [266], el modelo de los factores de certeza fue duramente criticado. Los creadores del programa, insatisfechos con el modelo, consultaron con el matemático J. Barclay Adams, quien señaló varias deficiencias e inconsistencias [1]:

1. El modelo de factores de certeza incluye también hipótesis de independencia condicional (volveremos sobre este punto más adelante).

2. La ecuación (2.7) puede dar un factor de certeza mayor a la hipótesis que debería tenerlo menor según la definición de CF a partir de probabilidades.

3. La utilización de máximos y mínimos para ponderar las premisas de cada regla (ecs. (2.8) y (2.9)) significa introducir nuevas hipótesis sobre las correlaciones entre ellas, que vienen a coincidir con las de la lógica difusa, de la que hablaremos en la sección siguiente, y que en general no se cumplen.

.

La crítica más fuerte que se ha hecho contra los factores de certeza procede de Heckerman, quien, basándose en una interpretación probabilística de los factores de certeza, demostró las graves inconsistencias a que estos dan lugar, incluso con reglas tan sencillas como "$E \Longrightarrow H$" (reglas de una sóla premisa), tanto en la propagación paralela (varias fuentes de evidencia apuntando a una misma hipótesis) como en la secuencial (encadenamiento de reglas) [99].[7] Para que este modelo fuera consistente deberían cumplirse varias hipótesis muy exigentes que no aparecen explícitamente en su planteamiento y que son falsas en general.

Horvitz y Heckerman [117] demostraron, además, que la utilización de factores de certeza es inconsistente porque, dicho *grosso modo*, el factor de certeza de la regla "$E \Longrightarrow H$" incluye tanto la probabilidad a priori de H como el cambio en la probabilidad de H debido a E, por lo que al combinar varias reglas de forma paralela estamos incluyendo repetidamente la probabilidad a priori en lugar de considerarla una sola vez.

Además de estas inconsistencias específicas del modelo de MYCIN, existen otras que, por ser comunes a todos los sistemas basados en reglas, serán tratadas en la sección 2.4.2.

2.3.3 Lógica difusa

La lógica difusa[8] fue desarrollada por Lofti Zadeh como una extensión de la teoría de conjuntos difusos que él mismo había creado [267]. Esta teoría surgió ante la necesidad de tratar matemáticamente conjuntos cuyos límites no están bien definidos. Por ejemplo, el conjunto de personas casadas tiene límites claros, mientras que el conjunto de personas altas no los tiene: si tres hombres miden respectivamente 1'65, 1'75 y 1'85, su *grado de pertenencia* a dicho conjunto es diferente, pero no se puede establecer un límite para el conjunto.

Los conjuntos clásicos pueden considerarse como un caso particular de conjuntos difusos, e igualmente las operaciones clásicas de unión, intersección y complementariedad resultan ser un caso particular de las correspondientes en la nueva teoría.

[7] Cuando David Heckerman inicio su tesis doctoral, su director de tesis, Edward Shortliffe le indicó que estudiase la teoría de Dempster-Shafer con el fin de corregir el modelo de factores de certeza de MYCIN. Sin embargo, insatisfecho también con esta teoría, prefirió utilizar redes bayesianas en su tesis doctoral [103] (comunicación personal). Heckerman es hoy en diá uno de los investigadores más brillantes en este nuevo campo.

[8] La explicación de por qué utilizamos el término "difusa" en vez de "borrosa" se encuentra en el apartado *Nota sobre el uso del idioma*, al comienzo de esta tesis (pág. xvii).

La pertenencia a un conjunto difuso induce una *distribución de posibilidad* [268], semejante a una distribución de probabilidad, aunque generalmente no está normalizada. Sobre estas distribuciones podemos definir *modificadores difusos*, que son operadores matemáticos correspondientes a los modificadores del lenguaje natural: muy, un poco, bastante, extremadamente... [270]. Del mismo modo podemos hablar de *cuantificadores difusos*: casi todos, muchos, pocos, algunos, etc. [269].

Se comprueba fácilmente que casi todos los conceptos que utilizamos en la vida real corresponden a conjuntos difusos, incluso en medicina, que es el campo que más nos interesa: presión alta, dolor agudo, fatiga leve, tumor grande, síntoma evidente, técnica sensible, diagnóstico complejo, pronóstico grave, terapia arriesgada, alta mortalidad, etc., etc.

A partir de la teoría de conjuntos difusos, podemos definir la *lógica difusa*, que utiliza reglas cuyas premisas incluyen afirmaciones difusas. Por ejemplo,

SI la curva es cerrada
Y la velocidad es alta
ENTONCES la curva resulta casi siempre muy peligrosa.

Vemos que esta regla incluye además dos modificadores difusos: "casi siempre" y "muy".

Hemos presentado aquí la lógica difusa de forma cualitativa, sin entrar en el formalismo matemático sobre grados de pertenencia, distribuciones de posibilidad, modificadores difusos y propagación de evidencia mediante encadenamiento de reglas. Una exposición sencilla del funcionamiento de las reglas difusas realizada por el propio Zadeh puede encontrarse en [272].

Discusión

La lógica difusa constituye una forma sencilla de representar explícitamente las imprecisiones del lenguaje natural que aparecen en todos los campos de la actividad humana, como ya hemos comentado. Por ello, no es extraño que la teoría de Zadeh se haya desarrollado de forma tan prodigiosa en los últimos años, dando lugar a gran cantidad de libros, revistas y congresos, y a numerosas aplicaciones prácticas, no sólo en el campo de los sistemas expertos, sino también en la construcción de máquinas y aparatos.

Sin embargo, desde el punto de visto teórico han surgido fuertes críticas contra ella. En cuanto a la teoría de conjuntos difusos, existe la cuestión de la semántica. Éste problema se manifiesta a la hora de dar una definición operativa del grado de pertenencia. Por ejemplo, si una persona mide 1'75, ¿cuál es su grado de pertenencia para el conjunto de personas altas? Si le asignamos 0'6, ¿por qué no podríamos haberle asignado 0'7 ó 0'4? La única forma conocida de dar no sólo un sentido objetivo sino también una definición operativa para los grados de pertenencia y las distribuciones de posibilidad es traducirlos a afirmaciones probabilísticas, tal como proponen independientemente Giles [94] y Cheeseman [39]. Así, grado de pertenencia de 0'7 para el conjunto de personas altas midiendo 1'75 significaría que la probabilidad de que una persona que mide 1'75 sea considerara alta es del 70%.[9]

[9]Existe además el problema del contexto. Por ejemplo, una mujer que mida 1'70 puede conside-

Aquí surge un dilema para los defensores de la lógica difusa. Si rechazan esta interpretación, tendrán que buscar una semántica alternativa y de ella deducir *definiciones operativas* para el grado de pertenencia y para la posibilidad; a nuestro juicio, que quizá se deba a la ignorancia sobre el tema, no existe tal interpretación como fundamento universalmente aceptado de la lógica difusa. Por el contrario, si se acepta la interpretación probabilística de los grados de pertenencia, la forma de combinarlos vendrá dictada por las leyes de la probabilidad. Tratarlos de otro modo significa introducir inconsistencias o, en el mejor de los casos, aplicar aproximaciones cuya validez habría que demostrar.

No es extraño, por tanto, que en la lógica difusa se encuentren hipótesis implícitas que, por ser falsas, pueden llevar a conclusiones erróneas. Por ejemplo, según la regla del mínimo para combinación de conjuntos difusos, la posibilidad de la conjunción de dos proposiciones A y B es el mínimo de sus correspondientes posibilidades: $\mathrm{Pos}(A \wedge B) = \min(\mathrm{Pos}(A), \mathrm{Pos}(B))$; sin embargo, en el caso de que A y B sean sucesos excluyentes entre sí, la posibilidad de su conjunción es exactamente cero [39, 262]. Podríamos utilizar medidas diferentes del mínimo, pero todas ellas están introduciendo hipótesis de correlación o independencia, sin que en muchos casos sean conscientes de ello quienes las están aplicando. Otras inconsistencias de la lógica difusa aparecen en [137, sec. 11].

Por último, un argumento más para criticar el uso de la regla del mínimo y preferir el uso de técnicas bayesianas se debe a la pérdida de sensibilidad que introduce dicha regla [221]. Por ejemplo, si tenemos dos premisas con posibilidades de 0'9 y 0'2 respectivamente, el valor de la conclusión que resulta de aplicar el mínimo sería el mismo que si tuviéramos 0'2 y 0'2.

Además de estas críticas específicas de la lógica difusa, hay otras deficiencias comunes a todos los sistemas basados en reglas, que serán tratadas en la sección 2.4.

Nuestra opinión personal es que los conjuntos difusos constituyen una aportación muy valiosa a la inteligencia artificial, aunque para que los grados de pertenencia tengan un significado objetivo hay que apoyarse en la teoría de la probabilidad.

Sin embargo, pasar de ahí a la lógica difusa es un salto en que se introducen de forma injustificada axiomas, hipótesis y aproximaciones que en unos casos son innecesarios y en otros inconsistentes. A nuestro entender, resultaría más correcto adherirse a la interpretación probabilística de los grados de pertenencia y desarrollar con ayuda de la teoría de la probabilidad un método semejante a la lógica difusa, pero en que todas las hipótesis estuvieran explícitas y todas las aproximaciones justificadas. No pretendemos afirmar que sea ésta una tarea fácil, pero creemos que debería intentarse el esfuerzo.

2.3.4 Otros métodos

Entre las técnicas numéricas para el tratamiento de la incertidumbre hay que citar la teoría de Dempster-Shafer [68, 201], una extensión de la teoría de la probabilidad que mide para cada proposición tanto la evidencia a favor como la evidencia en contra. Sin

rarse alta, pero no un hombre de la misma estatura. El concepto de "joven", por poner otro ejemplo, es especialmente sensible al contexto; un estudiante de 30 años se considera mayor, mientras que un catedrático de la misma edad se considera joven. El método bayesiano acepta de forma natural el contexto, al considerarlo como una de las variables que condicionan la probabilidad.

embargo, su formalismo matemático es complejo, por lo que no la vamos a desarrollar aquí. (Una presentación clara, con un ejemplo bien desarrollado se encuentra en [225].) Además, esta teoría debe considerar cada combinación posible de hipótesis que desea evaluar; la explosión combinatoria resultante hace que este método sea inaplicable en problemas del mundo real, a menos que se utilicen aproximaciones. Por esta razón, aunque el método de Dempster-Shafer sigue estudiándose en la teoría, prácticamente no se emplea hoy en día para la construcción de sistemas expertos.

Existen además otros **métodos no numéricos** para el tratamiento de la incertidumbre, basados fundamentalmente lógicas no monotónicas, tales como los modelos de razonamiento por defecto (el más conocido es el de Reiter [187]), los sistemas de suposiciones razonadas (originalmente llamados *truth maintenance systems*, aunque sería más correcto denominarlos *reason maintenace systems*) de Doyle [76] y la teoría de justificaciones (*theory of endorsements*) de Cohen y Grinberg [48, 49]. Estos métodos consisten en que, cuando no hay información suficiente, se hacen suposiciones que posteriormente podrán ser corregidas al recibir nueva información. El problema principal que presentan se debe a su naturaleza cualitativa, por lo que no pueden considerar los distintos grados de certeza o incertidumbre de las hipótesis. Suelen presentar además problemas de explosión combinatoria. En consecuencia, se estudian más por su importancia teórica (fundamentación de la inteligencia artificial) que por las aplicaciones prácticas a que puedan dar lugar.

Dentro de los métodos no numéricos debemos incluir también las redes cualitativas de Wellman [249, 250], aunque no vamos a tratar de ellas aquí sino en la sec. 10.1.3, al hablar de los modelos simplificados de redes bayesianas.

2.4 Crítica de los sistemas basados en reglas

Hoy en día el método más utilizado para la construcción de sistemas expertos consiste en el encadenamiento de reglas. En cambio, esta tesis toma un camino completamente diferente, tanto para la representación del conocimiento y el razonamiento como para el tratamiento de la incertidumbre. El objetivo de esta sección es explicar por qué hemos decidido apartarnos del camino habitual y explorar una técnica novedosa.

2.4.1 Representación del conocimiento

La ventaja más pregonada de los sistemas basados en reglas es su modularidad, que se traduce en la posibilidad de seguir añadiendo nuevas reglas a medida que se obtiene nueva información; el motor de inferencia se encargará después de encadenar y combinar todas las reglas que se encuentran en la base de conocimiento. Otra de las ventajas de las reglas es la flexibilidad, que se deriva de utilizar una representación modular.

Sin embargo, estas ventajas tienen su contrapartida. En efecto, basta pensar que modularidad implica representación no estructurada. En consecuencia, a medida que la base de datos va creciendo, resulta más difícil de mantener, por varias razones.

La primera es que la modularidad nunca es tan grande en la práctica como algunos afirman en la teoría. En efecto, suele decirse que las reglas permiten representar el conocimiento de forma declarativa en vez de tener que preocuparse de detalles de

programación. Pero esto no es cierto. A la hora de construir la base de conocimiento, el diseñador debe tener muy en cuenta cómo ordena las reglas y cómo ordena las premisas dentro de cada regla, imaginando cuál va a ser el flujo de información en la ejecución del programa. De otro modo, las preguntas que genere el sistema probablemente parecerán caóticas al usuario que está realizando una consulta. En consecuencia, una de las cuestiones fundamentales es ordenar cuidadosamente las cláusulas de cada regla para determinar en qué orden van a ser examinadas las premisas, y ordenar también las reglas dentro de la base de conocimiento para ver cuál de las que se pueden aplicar en un caso determinado va a ser investigada en primer lugar.

Un método muy habitual para controlar el razonamiento consisten en asignar prioridades a las reglas. Otro método también muy usado se basa en incluir ciertas cláusulas con el fin de tener en cuenta el contexto de la consulta. El resultado de todas estas técnicas *ad hoc* es que las reglas acaban perdiendo sus características más esencial: el carácter declarativo. A la larga, puede resultar mucho más complicado predecir cómo va a comportarse un sistema de este tipo que escribir un algoritmo. Hay técnicas más sofisticadas para controlar el razonamiento, tales como las metareglas (pág. 18), que aún no se han desarrollado lo suficiente como para ser práctica habitual en la construcción de sistemas expertos.

Otra consecuencia negativa de la falta de estructura de una base de conocimiento es la dificultad para verificar su contenido. Cuando un sistema basado en reglas crece por encima de cierto límite, resulta casi imposible predecir qué efecto va a tener la adición de una nueva regla y comprobar que no va a generar inconsistencias. El método habitual para la evaluación de sistemas expertos consiste no en su verificación sino en su validación: ver cómo se comporta ante un conjunto limitado de casos y comparar los resultados con los ofrecidos por expertos humanos (ver el cap. 14). Sin embargo, aunque el sistema haya funcionado correctamente en los ejemplos de prueba, no tenemos garantía de que vaya a actuar correctamente en casos futuros.

Podemos resumir estas críticas diciendo que la flexibilidad y la modularidad de las reglas, aunque en principio parecen facilitar la labor de construir sistemas expertos, plantean quizá más problemas de los que resuelven. Si esto es así en cuanto a la representación del conocimiento, resulta aún más claro en el tratamiento de la incertidumbre, como vamos a mostrar a continuación. Luego, en capítulos posteriores, veremos cómo las redes bayesianas, al basarse en una representación estructurada del conocimiento resuelven los problemas que aquí hemos planteado.

2.4.2 Tratamiento de la incertidumbre

Hemos presentado tres métodos de tratamiento de la incertidumbre basados en reglas: el método bayesiano subjetivo (PROSPECTOR), los factores de certeza (MYCIN) y la lógica difusa. La teoría de Dempster-Shafer puede utilizarse sobre una red, aunque la mayor parte de las aplicaciones se han construido en forma de reglas. Cada uno de dichos métodos presenta deficiencias específicas que fueron comentadas en las secciones correspondientes. Vamos a dedicar ahora este apartado a discutir los problemas inherentes a toda representación basada en reglas, cualquiera que sea el mecanismo que utilice para combinar los grados de credibilidad.

Hemos comentado anteriormente que la ventaja más pregonada de las reglas es su modularidad, y hemos visto también los problemas que ésta origina en cuanto a

la representación del conocimiento. Mostraremos a continuación que los problemas relativos al tratamiento de la incertidumbre proceden también de dicha modularidad. (Un tratamiento más detallado puede encontrarse en [174, sec. 1.2] y en [155, cap. 4].)

Empezaremos considerando un ejemplo sencillo. Supongamos que una alteración C puede tener dos causas A y B, y un efecto D, como indica la figura 2.4. Al intentar construir un sistema basado en reglas que nos permita realizar inferencias sobre este modelo tan simple, encontramos varias dificultades.

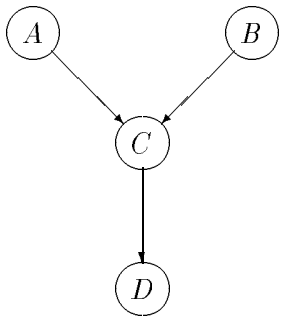


Figura 2.4: Nodo C con dos causas y un efecto.

La primera consiste en que las reglas no permiten establecer inferencias bidireccionales. En efecto, según nuestro modelo, la presencia de C nos hace sospechar la presencia de A como mecanismo causante. Recíprocamente, si descubrimos A, pensaremos que probablemente ha producido C. Ahora bien, si incluimos en la base de conocimiento dos reglas, "$A \Longrightarrow C$" y "$C \Longrightarrow A$", el aumento de la credibilidad de una de ellas aumenta la credibilidad de la otra, y viceversa, dando lugar a un ciclo sin fin.

Existen dos formas de intentar solucionarlo. Una de ellas consiste en incluir sólo una de las dos reglas, con lo cual estamos limitando la capacidad de inferencia de nuestro sistema. La otra consiste en no actualizar la credibilidad de cada proposición o variable más que una sola vez (ésta es la solución adoptada por GoldWorks, sec. 9.2.3). El inconveniente de esta segunda opción es que entonces las actualizaciones de la credibilidad no se transmiten a las reglas encadenadas. Así, en nuestro ejemplo anterior, una vez que la regla "$A \Longrightarrow C$" dispara la regla "$C \Longrightarrow D$", la credibilidad de D no se modificará aunque B aporte nueva evidencia a favor o en contra de C. El resultado es que el orden de llegada de la información puede alterar las conclusiones derivadas.

La segunda dificultad reside en que las reglas no distinguen entre causas y efectos. En el ejemplo anterior, vemos que tanto A como B como D pueden aportar evidencia a favor de C, pero de forma diferente. En efecto, al observar D aumenta la credibilidad de C y, por tanto, aumenta la credibilidad de B (y también de A) como causa posible. Supongamos que luego observamos también A. De nuevo aumenta la credibilidad de C pero en este caso la credibilidad de B no aumentará sino que disminuirá, pues A puede explicar la presencia de C, con lo cual disminuye nuestra sospecha sobre B. En resumen: un aumento en la credibilidad de C puede conducir a un aumento

o a una reducción de la credibilidad de B, dependiendo de cuál sea el origen de la evidencia. Sin embargo, en un sistema basado en reglas, "$C \Longrightarrow B$" hará aumentar la credibilidad de B, sea cual sea el origen de la evidencia.

Por supuesto, podemos escribir reglas más sofisticadas, tales como "$C \wedge \neg A \Longrightarrow B$", pero entonces el sistema se vuelve mucho más complejo, pues los casos imaginables son casi ilimitados, por lo que en la práctica esta solución se hace inviable.

La tercera dificultad está relacionada con la anterior. Imaginemos que hemos deducido B a partir de D, pero luego observamos A, con lo cual debe disminuir la credibilidad de B. Sin embargo, no nos está permitido incluir la regla "$A \Longrightarrow \neg B$ ni otras similares, pues la presencia de A por sí misma no excluye B: ambas causas pueden estar presentes.

Existe, finalmente, un cuarto problema. Supongamos que tenemos dos informes médicos (o dos indicios de cualquier clase) que apuntan a un mismo diagnóstico. Ahora bien, si descubrimos que el segundo de ellos se basó en el diagnóstico del primero, la fiabilidad conjunta de ambos informes es menor que si los dos médicos hubieran llegado a la misma conclusión independientemente. La incapacidad de tratar fuentes de evidencia correlacionadas es otra de las limitaciones de los sistemas basados en reglas.

Heckerman y Henrion [105, 106] llegaron a la conclusión de que el origen de los problemas mencionados consiste en aplicar al razonamiento con incertidumbre un método, las reglas, que es válido solamente en el campo de la lógica, donde todas las proposiciones son ciertas o falsas, pero nunca dudosas. En efecto, en el campo de la lógica sí existe la *modularidad semántica*, la cual significa que podemos deducir una conclusión a partir de unas premisas

—independientemente de cómo fueron deducidas dichas premisas, e

—independientemente de que existan otras proposiciones o reglas.

Estas dos propiedades se denominan desacoplo (*detachment*) y localidad (*locality*), respectivamente [174, pág. 5]. Sin embargo, en el razonamiento incierto no se cumplen dichas propiedades, tal como hemos discutido en los ejemplos anteriores. De aquí se deduce que **solamente es correcto utilizar reglas en dominios deterministas**, pues éstas son incapaces de tratar las correlaciones que surgen a causa de la incertidumbre.

Hemos mencionado anteriormente que los métodos de tratamiento de la incertidumbre basados en reglas, especialmente el de PROSPECTOR y el de MYCIN, surgieron ante las dificultades que presentaba el tratamiento probabilístico clásico, principalmente por las hipótesis de independencia que requería. Sin embargo, un estudio en profundidad demuestra que los métodos basados en reglas no aportan una solución alternativa, sino que inconscientemente introducen las mismas hipótesis e incluso otras mucho más estrictas. Como tales hipótesis son generalmente falsas en las aplicaciones reales, los sistemas basados en reglas, incluida la lógica difusa, realizan un tratamiento inconsistente —y por tanto incorrecto— de la incertidumbre.

En capítulos posteriores veremos cómo las redes bayesianas resuelven de forma natural estos problemas mediante la construcción de modelos causales en que todas las hipótesis de dependencia e independencia se encuentran explícitamente representadas.

2.4.3 Conclusión

El debate sobre cuál es el método más adecuado para representar la incertidumbre sigue abierto hoy en día. Aunque aún estamos lejos de llegar a un acuerdo, podemos afirmar que el método probabilístico vuelve a ser un serio candidato como solución al problema. Esta vía cuenta cada vez con más adeptos que la defienden, muchos de ellos de forma apasionada.

Como señala Glenn Shafer [202], se han superado ya aquellos tiempos en que el tratamiento puramente probabilístico era menospreciado por la mayoría de los estudiosos de la inteligencia artificial. Este hecho tiene una explicación histórica: cuando surgió este nuevo campo, los ordenadores resultaban muy útiles en el cálculo numérico intensivo, pero no eran inteligentes. Precisamente la inteligencia artificial se esforzaba en distinguirse de los métodos algorítmicos dedicados al tratamiento de números. Además, la inteligencia humana razona a nivel simbólico y no funciona como una máquina calculadora. Por eso, continúa Shafer, "los prejuicios contra los números en general y contra la probabilidad en particular no han desaparecido completamente de la inteligencia artificial". Así se entiende por qué han surgido tantos métodos no numéricos para el tratamiento de la incertidumbre (sec. 2.3.4).

Afortunadamente, estos prejuicios van retrocediendo. Una de las razones es que hoy en día se comprende mejor que en la resolución de problemas hay varios niveles. La distinción de David Marr entre teoría, algoritmo e implementación es realmente clarificadora [140, sec. 2.1]. Así, lo que hace que un sistema experto sea tal no es que utilice reglas en lugar de algoritmos probabilísticos, sino que sea capaz de razonar a nivel conceptual, aunque tenga por debajo un nivel numérico-algorítmico.

Otra de las razones a favor de la utilización de números es de orden práctico: los factores cuantitativos han resultado sumamente útiles en todos los sistemas expertos en que han sido empleados, empezando por MYCIN y PROSPECTOR, mientras que los métodos no numéricos, a pesar de su interés teórico, resultan muy poco eficientes en las aplicaciones reales.

Por último, en la psicología cognitiva se han realizado numerosos estudios que demuestran que la mente humana no siempre alcanza la solución óptima, sino que tiende a cometer cierto tipo de errores. (A este tema dedicamos el próximo capítulo.) Por tanto, la I.A. ya no insiste tanto en imitar el modo de razonamiento humano, sino en complementarlo en aquellos aspectos en que es más deficiente, y es natural pensar que para esta meta debamos emplear técnicas diferentes de las que utiliza la mente humana.

Y dentro de los métodos numéricos, la teoría de la probabilidad vuelve a gozar de un puesto relevante, y continúa ganando posiciones frente a técnicas alternativas. La razón principal es que con el surgimiento de las redes bayesianas se han superado las objeciones más fuertes que existían contra el método probabilístico clásico.

Por un lado, se han realizado varios estudios empíricos que la señalan como superior a los demás métodos empleados habitualmente [100, 111, 124, 262]. Por otro, hay argumentos teóricos (algunos ya han sido señalados anteriormente) para justificar por qué la probabilidad es no sólo el mejor método sino *el único* adecuado.

El estudio más extenso, redactado en el lenguaje agresivo que caracteriza a su autor, es el realizado por Peter Cheeseman [40]. Otro artículo suyo sobre el mismo tema se encuentra en [38]. Su conclusión es que "los problemas tratados típicamente

con métodos difusos pueden ser tratados mejor probabilísticamente" [39].

Otro de los defensores más apasionados de la probabilidad es Dennis Lindley, quien afirma: "La única descripción satisfactoria de la incertidumbre es la probabilidad" [137] y explica por qué los demás métodos resultan insatisfactorios.

Igualmente, Heckerman y Horvitz plantean desde un punto de vista axiomático cuáles son las propiedades deseables en un sistema de razonamiento incierto, y llegan a la conclusión de que el método probabilístico es el único que cumple estas condiciones [106, 117].

Como era de esperar, no todos los investigadores están de acuerdo con estas afirmaciones. Shafer, naturalmente, defiende su propia teoría (ver [202] y las réplicas que aparecen a continuación de ese artículo), mientras que Zadeh [271] y Kosko [132] sostienen que la lógica difusa resuelve problemas que escapan al tratamiento probabilístico. Es de prever que este intenso debate se prolongará durante mucho tiempo. De todos modos, resulta interesante la siguiente reflexión, teniendo en cuenta que procede Davis, Buchanan y Shortliffe [63], los creadores del MYCIN, quienes al comentar el artículo que en 1977 publicaron en la revista *Artificial Intelligence* [62] afirman:

> [...] En la última década, la investigación sobre tratamiento de la incertidumbre en I.A. ha avanzado notablemente. Muchas de las restricciones que limitaban las opciones para tratar la incertidumbre en MYCIN (en particular, los argumentos en contra de adoptar un método bayesiano estadístico clásico) ya no son válidos. Por ejemplo, las redes bayesianas proporcionan ahora un método viable para construir grandes sistemas de diagnóstico sin utilizar las hipótesis burdas e inherentemente defectuosas de MYCIN sobre la independencia condicional y la modularidad del conocimiento (ver, por ejemplo, [53, 106]).

Capítulo 3

Estimación subjetiva de la incertidumbre

Antes de introducirnos en el estudio de las redes bayesianas, vamos a dedicar un breve capítulo a revisar algunos de los estudios psicológicos sobre la estimación subjetiva de la probabilidad y de la utilidad, especialmente en el campo de la medicina, con el fin de mostrar en qué medida los sistemas expertos bayesianos pueden ser útiles para el diagnóstico.

La referencia clásica es el libro de Kahneman, Slovic and Tversky, "*Judgement under uncertainty: Heuristics and biases*" [123], en que se recogen varios artículos sobre el tema. En el primer párrafo de esta obra, los autores afirman:[1]

> El enfoque del estudio del juicio que este libro representa procede de tres líneas de investigación que se desarrollaron en los años 50 y 60: la comparación de la predición clínica con la predicción estadística, iniciada por Paul Meehl; el estudio de la probabilidad subjetiva en el paradigma bayesiano, introducido en la psicología por Ward Edwards; y la investigación de las heurísticas y las estrategias de razonamiento, para las cuales Herbert Simon ofreció un programa y Jerome Bruner un ejemplo. [...]
>
> Con la introducción de las ideas bayesianas en la investigación psicológica, por Edwards y sus colaboradores, los psicólogos dispusieron por primera vez de un modelo de actuación óptima ante la incertidumbre, con el cual poder comparar los juicios humanos. La adecuación de éstos a los modelos normativos había de convertirse en uno de los principales paradigmas de investigación sobre el juicio en presencia de incertidumbre.

En la última sección del capítulo discutiremos la importancia de estas investigaciones en relación con el tema de esta tesis.

[1] Hemos traducido *judgement* por *juicio*, aunque también tiene aquí el sentido de *opinión* o *estimación*.

3.1 Errores en la estimación de la probabilidad

3.1.1 Estudios psicológicos

El primer trabajo importante sobre este tema es el artículo de Tversky y Kahneman [239], publicado en 1974, en que, a partir de experimentos psicológicos, muestran algunos errores típicos en la estimación de la probabilidad. Estos autores identifican tres heurísticas que emplea el ser humano: *representatividad*, *disponibilidad* y *ajuste y anclaje*; cada una de ellas lleva a una serie de errores. Mostramos ahora los que se relacionan con estas tres y a continuación los que afectan al razonamiento causal.

Representatividad

A la hora de juzgar la probabilidad de que A pertenezca a la clase B o de que A haya sido producido por B, la gente suele considerar únicamente en qué medida A es representativo de B, sin tener en cuenta otros factores, como la probabilidad a priori o el tamaño de la muestra. Algunos de los errores originados por esta heurística son los siguientes:

- *Insensibilidad a la probabilidad a priori.* Cuando a los sujetos que participaban en el estudio se les indicaba solamente la composición de la muestra (por ejemplo, "hay un 30% de X y un 70% de Y"), ellos estimaban adecuadamente la probabilidad (0'3 y 0'7, respectivamente), pero al introducir información irrelevante, los sujetos ignoraban la probabilidad a priori, afirmando que ambas probabilidades eran iguales: 0'5. Más adelante veremos los equívocos que sufren muchos médicos acerca de la probabilidad a priori.

- *Conservadurismo.* En varios estudios se ha demostrado que el ser humano tiende a mantener su juicio inicial sobre la probabilidad, subestimando la fuerza de la evidencia recogida posteriormente. Más aún, se ha comprobado también que el ser humano tiende a valorar excesivamente la evidencia a favor de su hipótesis inicial, minusvalorando la que se opone a ella. La resistencia al cambio y la ponderación selectiva de la evidencia puede tener consecuencias importantes, especialmente en el campo del diagnóstico médico.

- *Equívocos sobre el azar.* Hay varios ejemplos curiosos sobre este punto. Así, muchos sujetos piensan que al lanzar una moneda al aire varias veces es más probable obtener una secuencia como "C+C++C", en que las caras (C) y las cruces (+) están distribuidas al azar, que una secuencia como "CCC+++", en que las caras y las cruces están ordenadas; del mismo modo, consideran que, para una moneda sin trucar, la secuencia "CCCCCC" es menos probable que las anteriores. Otro ejemplo es la *falacia del jugador*, por la que se supone que un número tiene más probabilidades de aparecer cuando no ha aparecido en las últimas veces que se ha lanzado el dado.[2] Detmer et al. [69] mostraron en un

[2] Hace unos meses, cierto periódico madrileño daba algunos consejos sobre cómo jugar a la lotería primitiva. Decía más o menos así: "En principio, todas las combinaciones son igualmente probables. Sin embargo, basándonos en la experiencia, podemos excluir algunas que no han aparecido nunca. Por ejemplo, no ha habido ningún caso en que salieran seis números consecutivos. Tampoco han salido nunca seis números de la misma decena, ni seis números terminados en la misma cifra..."

experimento que una parte de los cirujanos encuestados caía en este error al presentarles un problema similar sobre un hospital imaginario.

- *Equívocos sobre la regresión.* Existe un fenómeno estadístico, conocido como *regresión a la media*, por el que después de haber obtenido un resultado alto es más probable que el siguiente sea inferior al primero, y viceversa. Este fenómeno aparece en muchos casos en la vida cotidiana, por ejemplo en cuanto al rendimiento de un individuo en pruebas consecutivas o en cuanto a la estatura o la inteligencia de los hijos respecto de los padres. Sin embargo, esto lleva muchas veces a conclusiones erróneas, tal como narran Tverky y Kahneman:

  > En una discusión sobre entrenamiento de pilotos, instructores experimentados señalaron que a la felicitación por un aterrizaje excepcionalmente suave suele seguir un aterrizaje peor que el anterior, mientras que a una dura reprimenda por un aterrizaje brusco suele seguir una mejora en el siguiente. Los instructores concluyeron que las recompensas verbales perjudican el aprendizaje, mientras que los castigos verbales lo benefician, en contra de la doctrina psicológica establecida.

Disponibilidad

Al estimar la probabilidad de un suceso, la gente se basa en el número de casos que recuerda —no hay otro modo de hacer una estimación subjetiva—. Sin embargo, esta heurística puede llevar a varios tipos de errores.

- Hay casos que se graban con más fuerza en la memoria, lo cual puede dar lugar a una estimación de su frecuencia (y por tanto de su probabilidad) mayor que la real.Tversky [239] afirma: "El impacto que el ver una casa ardiendo produce sobre la probabilidad subjetiva de tales accidentes es probablemente mayor que el impacto de leer en el periódico la noticia de un incendio." Existe comprobación empírica de este fenómeno que, por otra parte, todos hemos experimentado en la vida cotidiana.

- Puede darse una explicación para el hecho de que muchas veces se observen correlaciones espurias e, inversamente, que se pasen por alto correlaciones existentes: la razón es que cuando la asociación de dos fenómenos se recuerda mejor que otras asociaciones, esto puede dar lugar a la "observación" de que ambos fenómenos están correlacionados, y viceversa.

Ajuste y anclaje

Muchas veces, la gente estima una cantidad a partir de un valor inicial, realizando un *ajuste* para el problema planteado. Tversky y Kahneman denominan *anclaje* a la influencia que el valor inicial ejerce sobre la respuesta de los sujetos estudiados.

- *Ajuste insuficiente.* En un experimento, a los sujetos de un grupo se les pidió que juzgaran si la proporción de países africanos en las Naciones Unidas era superior al 10%, y qué porcentaje estimaban como correcto. A un segundo grupo se le preguntó si la proporción era mayor del 65%, y cúal era la que ellos

estimaban. La respuesta promedio del primer grupo fue el 25% y la del segundo el 45%.

- *Errores sobre conjunción y disyunción de sucesos.* Para la conjunción de sucesos, la probabilidad es menor que la de cada uno de ellos; para la disyunción, la probabilidad es mayor. Por un fenómeno similar al del ajuste insuficiente, el ser humano tiende a atribuir a la conjunción una probabilidad mayor que la que le corresponde en realidad, y para la disyunción ocurre lo mismo en sentido contrario.

Atribución causal

Hay varios experimentos que confirman la tendencia del ser humano a ordenar su percepción del mundo de modo causal. Sin embargo, esto puede dar lugar a inconsistencias y paradojas en los modelos resultantes. Algunas de ellas son las siguientes:

- *Causalidad ficticia.* La tendencia a percibir la realidad en forma causal puede llevar a percibir causalidad donde se sabe que no existe, cayendo así en una contradicción encubierta. De este modo, sujetos que reconocen que entre peso y estatura hay correlación pero no de causalidad, afirman que es más razonable decir "Juan pesa mucho porque es alto" que "Juan es alto porque pesa mucho".

- *Asimetría en la inferencia.* Existen problemas en que la relación entre dos variables es completamente simétrica, y sin embargo la gente tiende a realizar inferencias en un sentido con mayor confianza que en el contrario. Por ejemplo, a muchas personas les parece más fiable precedir la edad del hijo a partir de la del padre que realizar la predicción inversa, cuando en realidad ambas son equivalentes. Existen otros experimentos que confirman que el ser humano tiende a poner más confianza en el razonamiento deductivo-causal (inferir el efecto a partir de la causa) que en el razonamiento de diagnóstico (inferir la causa a partir del efecto).

- *Efecto retrospectivo.* Algunos sucesos que a priori se juzgaban como improbables pueden parecer "razonables" o "muy posibles" una vez que el sujeto conoce que han sucedido, e incluso éste encuentra con relativa facilidad una explicación de por qué *tenía* que suceder así.

3.1.2 Errores en medicina

La capacidad de un médico para diagnosticar y tratar las enfermedades se debe sobre todo a los conocimientos que ha acumulado durante largos años de estudio y experiencia. Sin embargo, es de suponer que a la hora de aplicar dichos conocimientos utiliza las mismas heurísticas que hemos comentado anteriormente, por lo que en medicina pueden presentarse los mismos riesgos de error que acabamos de comentar.

En efecto, Detmer et al. [69] comprobaron que médicos experimentados también puede cometer el mismo tipo de errores cuando se les plantean cuestiones similares sobre casos clínicos; más concretamente, estos autores observaron la insensibilidad al tamaño de la muestra, la falacia del jugador, el efecto retrospectivo y otros relacionados con la disponibilidad.

En un trabajo más extenso, Berwick, Fineberg y Weinstein [17] elaboraron un questionario con información sobre enfermedades ficticias —para evitar que la experiencia pudiera interferir con los resultados de la investigación— y lo distribuyeron a 281 sujetos de varios grupos: estudiantes de medicina, residentes, médicos investigadores y médicos dedicados en exclusiva a la práctica clínica. La conclusión es que este último grupo obtuvo unos resultados mucho peores que los tres anteriores. Las deficiencias más importantes observadas fueron las siguientes:

1. *Falta de acuerdo sobre el significado de conceptos básicos.* El 45% de los médicos dedicados en exclusiva a la práctica sanitaria respondieron incorrectamente cuando se les preguntó el significado de conceptos básicos, como la prevalencia, la incidencia y el valor p.

2. *Desconocimiento de los principios estadísticos fundamentales*, como el regreso a la media y el teorema del límite central. Solamente el 32% de los encuestados respondieron correctamente a una pregunta que consistía en simplemente en aplicar el teorema de Bayes; entre los médicos dedicados en exclusiva a la práctica clínica, tan sólo el 21% respondió correctamente.

3. *Aceptación de conclusiones injustificadas,* debido principalmente a la "detección" de correlaciones donde un sencillo análisis estadístico hubiera mostrado que las coincidencias eran puramente casuales. En otros casos se dio precisamente el fenómeno contrario: el rechazo de conclusiones que sí podían deducirse de los datos disponibles.

4. *Estimación incorrecta de la utilidad.* Incluso definiendo la utilidad simplemente en términos de mortalidad, sin tener que considerar tampoco el coste y el riesgo de las pruebas, muchos médicos evaluaron incorrectamente la conveniencia de intervenir quirúrgicamente en la enfermedad hipotética; en general, eran reacios a operar incluso cuando los riesgos de no hacerlo superaban el peligro de la intervención.

Los errores médicos debidos al desconocimiento de la probabilidad no se dan solamente en experimentos de laboratorio sobre casos ficticios. El trabajo de Eddy [83] es sumamente interesante para conocer la gravedad de los errores que pueden cometerse incluso en publicaciones de revistas prestigiosas. Al repasar la literatura sobre la utilidad de la radiografía para detectar el cáncer de mama, muestra que en numerosas referencias se confunde la sensibilidad con el valor predictivo positivo (ver el apéndice de este capítulo), lo cual conduce a una estimación incorrecta de la probabilidad —en una ocasión el error alcanza un factor de 150—; incluso hay cierto autor que desaconseja la realización de la prueba, en contra de las más elementales nociones de teoría de la decisión.

Otra de las conclusiones del estudio de Eddy es la confusión que existe en muchos médicos sobre el papel que la probabilidad a priori desempeña en el diagnóstico, hasta el punto de que algunos de ellos afirman que para un paciente particular ésta no es aplicable y hay que considerar únicamente los datos disponibles sobre cada persona. El despreciar la prevalencia es probablemente el error que con más frecuencia se repite en la bibliografía médica (ver también [169]).

3.1.3 Discusión

Hemos revisado varios estudios que muestran la tendencia del ser humano a cometer cierto tipos de errores ante la estimación de la incerdumbre. Los profesionales de la medicina tampoco están exentos de tropezar en las sutilezas y paradojas de la teoría de la probabilidad. Sin embargo, no conviene exagerar la importancia de este peligro. Los estudios de Christensen-Szalanski et al. demuestran que en la práctica médica el efecto de estas tendencias es sólo cuantitativo —es decir, no conlleva errores cualitativos— [42] y que la gravedad de los errores disminuye considerablemente cuando existe un aprendizaje ligado a la experiencia [41].

Ciertamente, estos trabajos cuestionan hasta qué punto los resultados de los experimentos psicológicos son válidos en la vida real, pero tampoco los invalidan por completo. De hecho, los ejemplos que Dawes [64] ofrece sugieren que también en la vida real el ser humano comete el mismo tipo de errores en la estimación de la probabilidad. Y el trabajo de Eddy [83] mencionado anteriormente demuestra que incluso en trabajos escritos y reflexionados es posible que aparezcan errores a la hora de calcular la probabilidad.

La conclusión es que el ser humano tiende a equivocarse sistemáticamente cuando se enfrenta a problemas de incertidumbre; el peligro existe también en el campo de la medicina y, a pesar de que a veces se ha exagerado su importancia, hay evidencia suficiente de que este tipo de errores se dan y de que pueden llevar a conclusiones totalmente equivocadas.

3.2 Expresiones cualitativas de la probabilidad

Hemos mencionado algunos de los errores que aparecen en medicina como consecuencia por desconocimiento de la teoría de la probabilidad. Vamos a abordar ahora otra fuente de imprecisiones e incluso de graves errores: la utilización de expresiones cualitativas de la probabilidad.

Willian Schwartz [195] fue uno de los primeros en señalar los peligros de esta imprecisión:

> Un conflicto aparentemente irreconciliable relativo a la conveniencia de una operación puede explicarse con frecuencia como resultado de los diferentes significados asociados por dos médicos al término "baja probabilidad". Para uno puede significar bastante menos del 1%, mientras que para otro significa el 10 o incluso el 15%.

Bryant y Norman [23] realizaron un estudio en que pidieron a 16 médicos que asignaran probabilidades numéricas a 30 expresiones cualitativas extraídas de informes clínicos. Sólo 9 de estas asignaciones tuvieron un rango de variación inferior al 50%. En algunos casos el rango fue incluso superior al 90%, lo cual indica la casi total incertidumbre sobre el significado de tales expresiones.

En un experimento similar, Kong et al. [131] obtuvieron valores numéricos diferentes, aunque la ordenación de las expresiones (según la probabilidad asignada) fue la misma. Lo más interesante de este trabajo es el estudio de la influencia que el *contexto* ejerce sobre la estimación de la probabilidad.

Nakao y Axelrod [154] publicaron un trabajo con un título muy expresivo: "*Numbers are better than words: Verbal specifications of frequency have no place in*

medicine". En él, estudiaron la asignación de probabilidad para 22 términos, dividiendo a los sujetos del estudio en dos grupos: médicos y no médicos. Para 10 de los términos hubo una concordancia mayor entre los médicos que entre los demás participantes; no obstante, incluso las diferencias que aparecieron dentro del primer grupo fueron tan significativas que los autores concluyen: "Las expresiones cualitativas de frecuencia deben ser eliminadas de las comunicaciones médicas; de no ser posible, el autor debería especificar numéricamente la frecuencia que quiere indicar al utilizar tal expresión".

Algunos experimentos psicológicos revelan que las personas suelen preferir expresar la incertidumbre cualitativamente pero prefieren recibir la información cuantitativamente [28]. (Esta referencia contiene además bastantes resultados curiosos o sorprendentes sobre la actitud psicológica ante ambos tipos de expresiones.)

3.3 Conclusión

Como síntesis de lo expuesto hasta aquí, podemos afirmar que el tratamiento humano de la incertidumbre está sujeto a numerosas fuentes de error e imprecisión. En el campo de la medicina se han llevado a cabo múltiples estudios que confirman e incluso amplían el abanico de peligros existentes. Entre ellos, Komaroff [130] ofrece varios ejemplos de la variabilidad e inexactitud de los datos en medicina, a parte de los ya mencionados.

Ante esta situación, varios investigadores de la inteligencia artificial especializados en aplicaciones médicas han señalado las ventajas que la implantación de ordenadores podrá aportar en la práctica clínica. Entre ellos, Willian Schwartz [195] propone el *análisis de la decisión*, una rama de la teoría de la probabilidad, como solución a estos problemas, tratando de responder a las objeciones planteadas. También Pauker y Kassirer [169] insisten en las ventajas de esta técnica, mostrando cómo aplicar a la medicina el análisis de la sensibilidad y los procesos de Markov. Igualmente, Shortliffe, Buchanan y Feigenbaum [210] hicieron una amplia revisión de los métodos disponibles en ese momento (1979), tanto dentro como fuera de la inteligencia artificial, destacando su utilidad de cara a la medicina.

Las posibilidades de aplicación de los ordenadores al campo sanitario son ilimitadas. Ni siquiera se ha explotado todavía una pequeña parte de su utilidad en labores de gestión. Pero no es ése el tema que nos interesa ahora, sino las aportaciones que la ciencia de la computación puede realizar a un nivel superior: el del razonamiento médico.

La discusión de los temas tratados en este capítulo nos lleva a concluir que hay al menos dos puntos en que la medicina puede beneficiarse de la implantación de sistemas inteligentes. Por un lado, estos programas avanzados pueden conducir a los médicos a un razonamiento más riguroso y a una expresión más precisa de sus conclusiones. Por otro lado, los sistemas expertos pueden colaborar de forma sinergística con el médico con el fin de lograr el mejor rendimiento posible. En este sentido, conviene recordar lo que, basándose en el trabajo de Clarke [47], afirma Schwartz [194]:

> Si extraemos el conocimiento probabilístico de un médico y lo combinamos usando un cálculo normativo, puede dar como resultado una precisión en el diagnóstico superior a la del médico.

Aunque nosotros preferiríamos expresar la misma idea de forma diferente:

> Si ponemos en manos de un médico un sistema experto normativo que le ayude a sacar el máximo provecho del conocimiento que posee, el médico podrá realizar diagnósticos mucho más precisos.

Así enlazamos ya los temas planteados en este capítulo con el objetivo de la presente tesis doctoral. La relación entre ellos es triple. Por un lado, la demostración experimental de que el ser humano sistemáticamente se equivoca en la estimación subjetiva de la probabilidad es un fuerte argumento a favor de la introducción de sistemas expertos médicos que puedan apoyar a los especialistas en aquéllos aspectos en que la mente humana se muestra más propensa a cometer errores.

Por otro lado, esta inclinación psicológica al error es un argumento en contra de los sistemas basados en reglas, que fueron diseñados para imitar la forma de razonamiento que utiliza el ser humano. A la luz de estas investigaciones, podemos concluir que el hecho de que una máquina imite a las personas no siempre es una virtud. Nuestro objetivo, en cambio, consiste en construir *sistemas normativos*, es decir, sistemas que tengan un sólido fundamento teórico, aunque para ello deban emplear un modo de razonamiento distinto al de la mente humana.

En tercer lugar, a la hora de construir una red bayesiana pocas veces se conocen las probabilidades objetivas que vamos a necesitar, por lo que deberemos basarnos en estimaciones subjetivas de los expertos humanos que colaboran en nuestro proyecto (en la sec. 11.3 hablaremos detenidamente sobre el problema de la obtención del conocimiento). El conocer las desviaciones típicas que suele cometer el ser humano en este tipo de problemas puede ayudarnos a ser extremadamente cautos y a plantear las preguntas de modo que reduzcamos las posiblidades de error. En esta línea, los capítulos 28 a 32 del libro de Kahneman, Slovic y Tversky [123] estudian la forma de corregir estas desviaciones; entre ellos, el trabajo de Fishhoff [90] ofrece sugerencias concretas y prácticas que pueden resultarnos útiles en la interacción con los médicos a la hora de extraer el conocimiento.

También la sección 3.2 es importante en cuanto a la obtención del conocimiento para un sistema experto probabilístico, pues las referencias citadas aportan una valiosa información al establecer una correspondencia, basada en experimentos psicológicos, entre las expresiones cualitativas que encontramos en la bibliografía o que obtenemos de nuestros expertos y las probabilidades cualitativas que necesitamos para construir la red bayesiana.

Apéndice: Prevalencia, sensibilidad y especificidad

Repasamos en este apéndice brevemente algunas de las definiciones básicas sobre probabilidad, que tienen un interés especial en el campo de la medicina. Hay un concepto, la *prevalencia*, que afecta a una sola variable, mientras que los otros cuatro se refieren siempre a la relación entre dos variables.[3]

[3] En una *interpretación objetivista* —nuestra postura personal— cada **probabilidad** se corresponde con una **frecuencia**; por tanto, existe una definición operacional que nos permite realizar experimentos en que la probabilidad sea calculada a partir de estudios estadísticos. Puede ocurrir que, en ausencia de datos experimentales, debamos recurrir a una estimación subjetiva de la proba-

Prevalencia: Es la frecuencia con que aparece una determinada enfermedad dentro de una población. Suele expresarse en tanto por ciento o en tanto por mil. Esta frecuencia coincide con la probabilidad de que una determinada persona —de la que no poseemos más información— padezca la enfermedad E. Es decir, la prevalencia viene dada por $P(+e)$.[4]

Sensibilidad: Es la probabilidad de que aparezda cierto hallazgo H en presencia de cierta enfermedad E: $P(+h|+e)$. El hallazgo puede ser un síntoma, un signo, el resultado de una prueba de laboratorio, etc. Una sensibilidad elevada (cercana al 100%) significa que la enfermedad casi siempre produce dicho hallazgo.

Especificidad: Es la probabilidad de que cierto hallazgo H esté ausente cuando la enfermedad E está ausente: $P(\neg h|\neg e)$. Una especificidad elevada significa que $P(+h|\neg e)$ es muy pequeña:

$$P(+h|\neg e) = 1 - P(\neg h|\neg e) \tag{3.1}$$

y, por tanto, que es muy improbable que la presencia de H se deba a otras causas distintas de E.

Valor predictivo positivo: Es la probabilidad de que exista la enfermedad E en presencia del hallazgo H: $P(+e|+h)$. Se relaciona con los tres conceptos anteriores mediante el teorema de Bayes:

$$\begin{aligned} P(+e|+h) &= \frac{P(+e)\,P(+h|+e)}{P(+h)} \\ &= \frac{P(+e)\,P(+h|+e)}{P(+h|+e)\,P(+e) + P(+h|\neg e)\,P(\neg e)} \\ &= \frac{P(+e)\,P(+h|+e)}{P(+h|+e)\,P(+e) + [1 - P(\neg h|\neg e)] \cdot [1 - P(+e)]} \end{aligned} \tag{3.2}$$

Por tanto, los factores que más influyen en el valor predictivo positivo son la prevalencia y la sensibilidad.

Valor predictivo negativo: Es la probabilidad de que la enfermedad E se encuentre ausente en ausencia del hallazgo H: $P(\neg e|\neg h)$. Se relaciona con los tres primeros conceptos también mediante el teorema de Bayes:

$$\begin{aligned} P(\neg e|\neg h) &= \frac{P(\neg e)\,P(\neg h|\neg e)}{P(\neg h)} \\ &= \frac{P(\neg e)\,P(\neg h|\neg e)}{P(\neg h|+e)\,P(+e) + P(\neg h|\neg e)\,P(\neg e)} \\ &= \frac{[1 - P(+e)]\,P(+h|+e)}{[1 - P(+h|+e)]\,P(+e) + P(\neg h|\neg e)[1 - P(+e)]} \end{aligned} \tag{3.3}$$

bilidad (en la sec. 11.3 hablaremos sobre la obtención del conocimiento en la construcción de un sistema experto bayesiano), pero eso no significa que nos adhiramos a la *interpretación subjetivista*, del mismo modo que estimar "a ojo" el peso de una persona no significa aceptar que éste sea un dato subjetivo.

[4] Para una variable binaria X, el valor $+x$ corresponde al valor "presencia de X" o "resultado positivo en el test X". El valor $\neg x$ significa "ausencia de X" o "resultado negativo en el test X". Mantendremos la misma notación a lo largo de toda esta memoria.

El valor predictivo negativo depende, sobre todo, de la prevalencia y de la especificidad.

Existen además otros conceptos, como la *razón de probabilidad*, que se introducen solamente para simplificar algunas ecuaciones, y carecen de una interpretación inmediata. Tampoco nos detenemos a comentar la *incidencia* (el número de casos nuevos de una enfermedad que aparecen en una cierta población en un determinado periodo de tiempo), porque en esta tesis no vamos a abordar los aspectos del razonamiento temporal.

Parte II

REDES BAYESIANAS

Capítulo 4

Fundamentos de redes bayesianas

4.1 Presentación intuitiva

Antes de presentar formalmente la teoría matemática de las redes bayesianas, intentaremos explicar mediante ejemplos sencillos el significado intuitivo de las definiciones y axiomas que luego introduciremos, para que no dé la impresión de que estamos haciendo mera teoría sin conexión con el razonamiento de sentido común. Puesto que esta tesis se orienta esencialmente al desarrollo de una aplicación cardiológica, los ejemplos están tomados del campo de la medicina. Aunque hemos tratado de que sean lo más realistas posible, hemos buscado sobre todo una aproximación cualitatitiva y no pretendemos que los factores numéricos sean exactos.

En una red bayesiana, cada nodo corresponde a una variable, que representa a su vez una entidad del mundo real, como puede ser la edad o el sexo de un paciente, la presencia de una enfermedad, un síntoma o el resultado de una prueba de laboratorio. De aquí en adelante hablaremos indistintamente de nodos y variables, y los denotaremos con letras mayúsculas, tales como X. Utilizaremos la misma letra en minúscula, x, para indicar un valor cualquiera de la variable X.

Entendemos por *probabilidad a priori* la probabilidad de una variable estimada a partir de la experiencia previa, es decir, antes de tener en cuenta la información relevante para el caso particular que estamos diagnosticando.

Ejemplo 1. La red bayesiana no trivial más simple que podemos imaginar consta de dos variables, que llamaremos X e Y_1, y un arco desde la primera a la segunda, como indica la figura 4.1. Por el momento, baste decir que el arco indica *influencia causal*; más adelante precisaremos el sentido de esta expresión. Utilizaremos frecuentemente el término *enlace* como sinónimo de *arco*.

Por concretar el ejemplo, podemos suponer que X representa Paludismo e Y_1 representa Gota-gruesa, la prueba más habitual para determinar la presencia de dicha enfermedad.

Cuando X es una variable binaria correspondiente a una anomalía, $+x$ indica la presencia de dicha anomalía (en nuestro ejemplo significaría "el paciente tiene paludismo") y $\neg x$ indica su ausencia ("el paciente no tiene paludismo"). Si X representa

Figura 4.1: Nodo X con un hijo Y_1.

un test (por ejemplo, Gota-gruesa), $+x$ indica que el test ha dado un resultado positivo y $\neg x$ un resultado negativo.

En la práctica, la información cuantititativa de una red bayesiana viene dada por la probabilidad a priori de los nodos que no tienen padres —$P(x)$— y por la probabilidad condicional de los nodos con padres —$P(y_1|x)$—. Así, en nuestro ejemplo, se supone que conocemos

$$\begin{cases} P(+x) = 0'003 \\ P(-x) = 0'997 \end{cases}$$

lo cual significa que el 3 por mil de la población padece paludismo y, por tanto, la probabilidad a priori de que una persona tenga la enfermedad (es decir, la probabilidad cuando no conocemos nada más sobre esa persona) es del 0'3%. En medicina, esta probabilidad a priori se conoce como *prevalencia* de la enfermedad.

También debemos conocer $P(y|x)$, que es la probabilidad del efecto condicionada en el valor de la causa:

$$\begin{Bmatrix} P(+y_1|+x) = 0'992 & P(+y_1|\neg x) = 0'0006 \\ P(\neg y_1|+x) = 0'008 & P(\neg y_1|\neg x) = 0'9994 \end{Bmatrix}$$

El significado de esta probabilidad condicional es el siguiente: cuando hay Paludismo, el test de la Gota-gruesa da positivo en el 99'2% de los casos. Este valor se conoce como *sensibilidad* del test. Cuando no hay paludismo, el test da positivo (se dice entonces que ha habido un "falso positivo") en el 0'06% de los casos. La probabilidad de que el test dé negativo cuando la enfermedad buscada está ausente —en nuestro caso es el 99'94%— se llama *especificidad*. En todos los problemas de dignóstico, no sólo en el campo de la medicina, tratamos de encontrar las pruebas que ofrezcan el grado más alto de sensibilidad y especifidad con el menor coste posible (en términos de dinero, tiempo, riesgo, etc.).[1]

Naturalmente,

$$\begin{cases} P(+y_1|+x) + P(\neg y_1|+x) = 1 \\ P(+y_1|\neg x) + P(\neg y_1|\neg x) = 1 \end{cases}$$

o, en forma abreviada,

$$\sum_{y_1} P(y_1|x) = 1, \quad \forall x \tag{4.1}$$

[1] Recordemos que esta definición de *sensibilidad* y *especificidad* es aplicable solamente a un enlace entre variables binarias de tipo presente/ausente o positivo/negativo.

Conociendo la probabilidad a priori de X y la probabilidad condicional $P(Y_1|X)$, podemos hallar la probabilidad a priori de Y_1:

$$\begin{cases} P(+y_1) = P(+y_1|+x) \cdot P(+x) + P(+y_1|\neg x) \cdot P(\neg x) \\ P(\neg y_1) = P(\neg y_1|+x) \cdot P(+x) + P(\neg y_1|\neg x) \cdot P(\neg x) \end{cases}$$

que puede escribirse en forma abreviada como

$$P(y_1) = \sum_x P(y_1|x) \cdot P(x) \tag{4.2}$$

En nuestro ejemplo,

$$\begin{cases} P(+y_1) = 0'00357 \\ P(\neg y_1) = 0'99643 \end{cases}$$

Esto significa que si hacemos el test de la gota gruesa a una persona de la que no tenemos ninguna información, hay un 0'357% de probabilidad de que dé positivo y un 99'643% de que dé negativo.

Vamos a ver ahora cómo podemos calcular la probabilidad a posteriori, es decir, la probabilidad de una variable dada la evidencia observada **e**:

$$P^*(x) \equiv P(x|\mathbf{e}) \tag{4.3}$$

a) Supongamos que la gota gruesa ha dado positivo: $\mathbf{e} = \{+y_1\}$. ¿Cuál es ahora la probabilidad de que nuestro paciente tenga paludismo? Si la prueba tuviera una fiabilidad absoluta, responderíamos que el 100%. Pero como es posible que haya habido un falso positivo, buscamos $P^*(+x)$, es decir, $P(+x|+y)$. Para ello, aplicamos el **teorema de Bayes**:

$$P^*(+x) = P(+x|+y_1) = \frac{P(+x) \cdot P(+y_1|+x)}{P(+y_1)} = \frac{0'003 \cdot 0'992}{0'00357} = 0'83263 \tag{4.4}$$

Es decir, de acuerdo con el resultado de la prueba, hay un 83% de probabilidad de que el paciente tenga paludismo.

También podemos calcular $P^*(\neg x)$:

$$P^*(\neg x) = P(\neg x|+y_1) = \frac{P(\neg x) \cdot P(+y_1|\neg x)}{P(+y_1)} = \frac{0'997 \cdot 0'0006}{0'00357} = 0'16737 \tag{4.5}$$

Esto significa que hay un 16'7% de probabilidad de que haya habido un falso positivo. Naturalmente, se cumple que

$$P^*(+x) + P^*(\neg x) = 1 \tag{4.6}$$

La expresión general del teorema de Bayes que hemos aplicado es

$$P^*(x) = P(x|y) = \frac{P(x) \cdot P(y|x)}{P(y)} \tag{4.7}$$

Por razones que quedarán claras más adelante, vamos a reescribirla como

$$P^*(x) = \alpha \cdot P(x) \cdot \lambda_{Y_1}(x) \tag{4.8}$$

donde hemos introducido dos nuevas magnitudes

$$\lambda_{Y_1}(x) \equiv P(\mathbf{e}|x) = P(y_1|x) \tag{4.9}$$

$$\alpha \equiv [P(\mathbf{e})]^{-1} = [P(y_1)]^{-1} \tag{4.10}$$

La propiedad fundamental de α es que no depende de x. Poco a poco iremos viendo su importancia.

Vamos a repetir ahora el cálculo anterior aplicando esta reformulación del teorema de Bayes. En primer lugar tenemos que, cuando el test da positivo,

$$\mathbf{e} = \{+y_1\} \Longrightarrow \begin{cases} \lambda_{Y_1}(+x) = P(+y_1|+x) = 0'992 \\ \lambda_{Y_1}(\neg x) = P(+y_1|\neg x) = 0'0006 \end{cases} \tag{4.11}$$

Esto significa que un resultado positivo en la prueba se explica mucho mejor con la enfermedad presente que con la enfermedad ausente (en la proporción de 0'992/0'0006=1.650), lo cual concuerda, naturalmente, con el sentido común.

Por tanto,

$$\begin{cases} P^*(+x) = \alpha \cdot 0'003 \cdot 0'992 = \alpha \cdot 0'00298 \\ P^*(\neg x) = \alpha \cdot 0'997 \cdot 0'0006 = \alpha \cdot 0'000598 \end{cases}$$

Podríamos calcular α a partir de su definición (4.10), pero resulta mucho más sencillo aplicar la condición de normalización (ec. (4.6)), de donde deducimos que

$$\alpha = [0'00298 + 0'000598]^{-1}$$

y finalmente

$$\begin{cases} P^*(+x) = 0'83263 \\ P^*(\neg x) = 0'16737 \end{cases}$$

que es el mismo resultado que habíamos obtenido anteriormente por la aplicación del teorema de Bayes en su forma original.

Examinando la ecuación (4.8) vemos que en la probabilidad a posteriori, $P^*(x)$, influyen dos factores: por un lado, la probabilidad a priori $P(x)$ y por otro el resultado del test $\lambda_{Y_1}(x)$. En esta ecuación también aparece α que, aunque tiene un significado bien definido, actúa aquí como si fuera simplemente una constante de normalización.

b) Y si la gota gruesa diera un resultado negativo, $\mathbf{e} = \{\neg y_1\}$, ¿cuál sería la probabilidad de que el paciente tuviera paludismo? En ese caso,

$$\mathbf{e} = \{\neg y_1\} \Longrightarrow \begin{cases} \lambda_{Y_1}(+x) = P(\neg y_1|+x) = 0'008 \\ \lambda_{Y_1}(\neg x) = P(\neg y_1|\neg x) = 0'9994 \end{cases} \tag{4.12}$$

Es decir, un resultado negativo en la prueba de la gota gruesa se explica mucho mejor (en la proporción de 0'9994/0'008 = 125) cuando no hay paludismo que cuando lo hay.

Aplicando la ecuación (4.8) como en el caso anterior, obtenemos

$$\begin{cases} P^*(+x) = \alpha \cdot 0'003 \cdot 0'008 = 0'000024 \\ P^*(\neg x) = \alpha \cdot 0'997 \cdot 0'9994 = 0'999976 \end{cases}$$

donde hemos calculado α a partir de la condición de normalización.

El resultado tan bajo para $P^*(+x)$ se explica por dos razones: por un lado, la probabilidad a priori era de sólo un 0'3%; por otro, la alta sensibilidad de la prueba (99'94%) es un argumento convincente para descartar la enfermedad.

En resumen, la ecuación (4.8) nos dice que la probabilidad de una hipótesis ($+x$ ó $\neg x$) dada cierta evidencia ($+y_1$ ó $\neg y_1$, en nuestro ejemplo) es mayor cuanto más probable fuera dicha hipótesis a priori y cuanto mejor explique esa hipótesis la evidencia observada.

Ejemplo 2. Vamos a ampliar el modelo anterior añadiendo un nuevo efecto producido por el paludismo, la Fiebre, que representaremos mediante la variable Y_2, tal como muestra la figura 4.2.

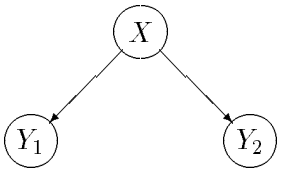


Figura 4.2: Nodo X con dos hijos.

La probabilidad condicional para este segundo enlace XY_2 viene dada por

$$\left\{\begin{array}{ll} P(+y_2|+x) = 0'98 & P(+y_2|\neg x) = 0'017 \\ P(\neg y_2|+x) = 0'02 & P(\neg y_2|\neg x) = 0'983 \end{array}\right\}$$

que indica la probabilidad de que un paciente (una persona, en general) tenga fiebre depediendo de si tiene o no paludismo. Vemos aquí que, para el paludismo, la fiebre tiene mucha menor especificidad (98'3%) que la gota gruesa (99'94%). Así, este sencillo modelo tiene en cuenta que hay muchas otras causas que pueden producir fiebre, aunque no las incluya explícitamente.

Podemos calcular la probabilidad a priori de que un enfermo tenga fiebre,

$$P(+y_2) = \sum_x P(+y_2|x) \cdot P(x) = 0'01989$$

pero éste es un resultado que carece de importancia para el diagnóstico.

a) Supongamos que encontramos un paciente con fiebre, $\mathbf{e} = \{+y_2\}$, y queremos hallar la probabilidad de que tenga paludismo. En primer lugar, expresamos el teorema de Bayes en la forma

$$P^*(x) = \alpha \cdot P(x) \cdot \lambda_{Y_2}(x) \tag{4.13}$$

Ahora α vale $[P(+y_2)]^{-1}$, pero podemos prescindir de su significado y considerarla simplemente como una constante de normalización.

Para un paciente con fiebre,

$$\mathbf{e} = \{+y_2\} \Longrightarrow \left\{\begin{array}{l} \lambda_{Y_2}(+x) = P(+y_2|+x) = 0'98 \\ \lambda_{Y_2}(\neg x) = P(+y_2|\neg x) = 0'017 \end{array}\right. \tag{4.14}$$

de modo que

$$\begin{cases} P^*(+x) = \alpha \cdot 0'003 \cdot 0'98 = 0'148 \\ P^*(\neg x) = \alpha \cdot 0'997 \cdot 0'017 = 0'852 \end{cases}$$

lo cual significa que hay un 14'8% de probabilidad de que el paciente tenga paludismo. Compárese con el 83'3% correspondiente a un resultado positivo de la gota gruesa (ec. (4.4)). La diferencia se debe de que esta prueba es un signo muy específico de la enfermedad, mientras que la fiebre puede estar producida por muchas otras causas.

b) Vamos a estudiar ahora el caso en que tenemos las dos observaciones y ambas indican la presencia de la enfermedad: $\mathbf{e} = \{+y_1, +y_2\}$. Al intentar calcular la probabilidad de que esa persona tenga paludismo, $P(+x|+y_1, +y_2)$ nos damos cuenta de que nos falta información, pues para aplicar el teorema de Bayes,

$$P(+x|+y_1, +y_2) = \frac{P(+y_1, +y_2|x) \cdot P(x)}{P(+y_1, +y_2)} \tag{4.15}$$

necesitamos conocer $P(+y_1, +y_2|x)$ y $P(+y_1, +y_2)$.

Con la información disponible es imposible calcular estas expresiones. Por ello vamos a introducir la *hipótesis de independencia condicional*. Examinemos primero el caso en sabemos con certeza que hay paludismo ($X = +x$). Entonces es razonable pensar que la probabilidad de que el paciente tenga o no tenga fiebre no depende de si hemos realizado el test de la gota gruesa ni del resultado que éste haya dado: la fiebre depende sólo de si hay paludismo (dando por supuesto, como parece razonable, que las demás causas de fiebre no influyen en el resultado del test). La afirmación "conociendo $X = x$, el valor de y_2 no depende del de y_1" se expresa matemáticamente como

$$P(y_2|+x, y_1) = P(y_2|+x) \tag{4.16}$$

o dicho de otro modo[2]

$$P(y_1, y_2|+x) = P(y_1|+x) \cdot P(y_2|+x) \tag{4.17}$$

Observar que estas expresiones son simétricas para Y_1 e Y_2.

Supongamos ahora que *no* hay paludismo ($X = \neg x$). La probabilidad de que el paciente presente fiebre no depende de si la gota gruesa ha dado negativo (como era de esperar) o ha dado un falso positivo por alguna extraña razón. Así tenemos

$$P(y_2|\neg x, y_1) = P(y_2|\neg x) \tag{4.18}$$

o lo que es lo mismo

$$P(y_1, y_2|\neg x) = P(y_1|\neg x) \cdot P(y_2|\neg x) \tag{4.19}$$

Uniendo (4.17) y (4.19), tenemos

$$P(y_1, y_2|x) = P(y_1|x) \cdot P(y_2|x) \tag{4.20}$$

[2] Ambas expresiones son equivalentes cuando $P(+x) \neq 0$, pues

$$P(y_1, y_2|+x) = \frac{P(y_1, y_2, +x)}{P(+x)} = \frac{P(y_2|y_1, +x) \cdot P(y_1, +x)}{P(+x)} = P(y_2|+x, y_1) \cdot P(y_1|+x)$$

que es lo que se conoce como *independencia condicional*. Definiendo

$$\lambda(x) \equiv P(y_1, y_2|x) \tag{4.21}$$

podemos expresar dicha propiedad como

$$\lambda(x) = \lambda_{Y_1}(x) \cdot \lambda_{Y_2}(x) \tag{4.22}$$

Con esta hipótesis ya podemos calcular la probabilidad buscada. La ecuación (4.15) es equivalente a

$$P^*(x) = \alpha \cdot P(x) \cdot \lambda(x) \tag{4.23}$$

En nuestro ejemplo, a partir de (4.11) y (4.14) tenemos

$$\mathbf{e} = \{+y_1, +y_2\} \Longrightarrow \begin{cases} \lambda(+x) = 0'97216 \\ \lambda(-x) = 0'000102 \end{cases} \tag{4.24}$$

El valor de α se calcula al normalizar, de modo que tenemos

$$\begin{cases} P^*(+x) = 0'9663 \\ P^*(\neg x) = 0'0337 \end{cases}$$

Naturalmente, cuando tenemos dos hallazgos a favor del paludismo, la probabilidad resultante (96'6%) es mayor que la correspondiente a cada uno de ellos por separado (83'3% y 14'8%).

c) En el caso de que tengamos un hallazgo a favor y otro en contra, podemos ponderar su influencia mediante estas mismas expresiones. Por ejemplo, si hay fiebre ($+y_2$) pero hay un resultado negativo en la prueba de la gota gruesa ($\neg y_1$), las ecuaciones (4.12), (4.14) y (4.22) nos dicen que

$$\mathbf{e} = \{\neg y_1, +y_2\} \Longrightarrow \begin{cases} \lambda(+x) = 0'008 \cdot 0'98 = 0'00784 \\ \lambda(\neg x) = 0'9994 \cdot 0'017 = 0'01699 \end{cases} \tag{4.25}$$

Vemos que hay más evidencia a favor de $\neg x$ que de $+x$ (en la proporción aproximada de 2 a 1), debido sobre todo al 0'008 correspondiente a la gota gruesa, lo cual es un reflejo de la alta sensibilidad de esta prueba (99'2%). Es decir, si hubiera paludismo, es casi seguro que lo habríamos detectado; al no haberlo detectado, tenemos una buena razón para descartarlo.

Al tener en cuenta además la probabilidad a priori de la enfermedad, nos queda finalmente

$$\begin{cases} P^*(+x) = \alpha \cdot 0'003 \cdot 0'00784 = 0'0014 \\ P^*(\neg x) = \alpha \cdot 0'997 \cdot 0'01699 = 0'9986 \end{cases}$$

Por tanto, la ponderación de la evidencia ha modificado la probabilidad desde 0'3% (valor a priori) hasta 0'14% (valor a posteriori).

d) Aún podemos obtener más información de este ejemplo. Imaginemos que tenemos un paciente con fiebre ($Y_2 = +y_2$) y todavía no hemos realizado la prueba

de la gota gruesa. ¿Qué probabilidad hay de que ésta dé un resultado positivo o negativo? Es decir, ¿cuánto vale $P(y_1|+y_2)$?

Por la teoría elemental de la probabilidad sabemos que

$$\begin{aligned} P^*(y_1) &= P(y_1|+y_2) = \sum_x P(y_1|\,x,+y_2)\cdot P(x|+y_2) \\ &= \sum_x P(y_1|\,x,+y_2)\cdot \frac{P(x,+y_2)}{P(+y_2)} \end{aligned}$$

Aplicando la independencia condicional dada en (4.17) y definiendo[3]

$$\pi_{Y_1}(x) \equiv P(x,+y_2) = P(x)\cdot P(+y_2|\,x) \tag{4.26}$$

$$\alpha \equiv [P(+y_2)]^{-1} \tag{4.27}$$

podemos reescribir la expresión anterior como

$$P^*(y_1) = \alpha\cdot\sum_x P(y_1|x)\cdot\pi_{Y_1}(x) \tag{4.28}$$

Sustituyendo los valores numéricos, tenemos que

$$\mathbf{e} = \{+y_2\} \Longrightarrow \begin{cases} \pi_{Y_1}(+x) = 0'003\cdot 0'98 \;\; = 0'00294 \\ \pi_{Y_1}(\neg x) = 0'997\cdot 0'017 = 0'01695 \end{cases} \tag{4.29}$$

y finalmente

$$\begin{cases} P^*(+y_1) = 0'14715 \\ P^*(\neg y_1) = 0'85285 \end{cases} \tag{4.30}$$

Resulta muy interesante comparar la ecuación (4.28) con (4.2). Al buscar la probabilidad a priori $P(y_1)$ utilizábamos $P(x)$; ahora, al calcular $P^*(y_1)$, utilizamos $\pi_{Y_1}(x)$, que indica la probabilidad de X tras considerar la evidencia relativa a X *diferente* de Y_1.

Vemos así cómo la información que aporta el nodo Y_2 modifica la probabilidad de X, y en consecuencia también la de Y_1. El carácter simultáneamente ascendente y descendente del mecanismo de propagación es lo que nos permite utilizar la red tanto para inferencias *abductivas* (cuál es el diagnóstico que mejor explica los hallazgos) como *predictivas* (cuál es la probabilidad de obtener cierto resultado en el futuro). Un mismo nodo Y_1 puede ser fuente de información u objeto de predicción, dependiendo de cuáles sean los hallazgos disponibles y el objetivo del diagnóstico.

Ejemplo 3. Consideremos una red bayesiana en que un nodo X, que representa la variable Paludismo, tiene dos padres, U_1 = País-de-origen y U_2 = Tipo-sanguíneo, dos de los factores que influyen en la probabilidad de contraer la enfermedad, tal como muestra la figura 4.3.

La variable U_1 podría tener muchos valores, tantos como países de origen quisiéramos considerar. Sin embargo, vamos a suponer que los agrupamos en tres zonas, de alto, medio y bajo riesgo, que denotaremos por u_1^+, u_1^0 y u_1^-, respectivamente. La variable U_2 (Tipo-sanguíneo) puede tomar dos valores: $u_2^\dagger$ ó $u_2^\ddagger$.

[3] $\pi_{Y_1}(x)$ lleva este subíndice porque recoge toda la evidencia relativa a X *diferente* de Y_1.

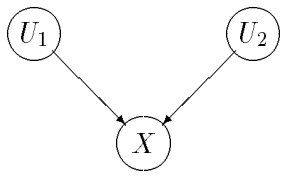


Figura 4.3: Nodo X con dos padres.

Las probabilidades a priori para U_1 y U_2 son:

$$\begin{cases} P(u_1^+) = 0'10 \\ P(u_1^0) = 0'10 \\ P(u_1^-) = 0'80 \end{cases} \qquad \begin{cases} P(u_2^\dagger) = 0'60 \\ P(u_2^\ddagger) = 0'40 \end{cases} \tag{4.31}$$

Esto significa que la mayor parte de las personas que vamos a examinar proceden de una zona de bajo riesgo u_1^- y que el primer tipo sanguíneo, $u_2^\dagger$, es más frecuente que el segundo.

Las probabilidades condicionadas aparecen en la tabla 4.1. En ella vemos que, efectivamente, la zona u_1^+ es la de mayor riesgo y u_1^- la de menor. También observamos que el tipo sanguíneo $u_2^\dagger$ posee mayor inmunidad que el $u_2^\ddagger$.

$U_2 \setminus U_1$	u_1^+	u_0^+	u_1^-
$u_2^\dagger$	0'015	0'003	0'0003
$u_2^\ddagger$	0'022	0'012	0'0008

Tabla 4.1: Probabilidad del paludismo, $P(+x|u_1, u_2)$.

La probabilidad de que una persona (de la que no tenemos ninguna información) padezca paludismo es

$$P(x) = \sum_{u_1, u_2} P(x|u_1, u_2) \cdot P(u_1, u_2) \tag{4.32}$$

De nuevo tenemos el problema de que no conocemos $P(u_1, u_2)$. Podemos entonces hacer la hipótesis de *independencia a priori* entre ambas variables; es decir, suponemos que los grupos sanguíneos se distribuyen por igual en las tres zonas de riesgo. Ésta es una hipótesis que habría que comprobar empíricamente. Si llegáramos a la conclusión de que existe una correlación entre ambas variables, deberíamos trazar un arco desde la una hasta la otra e introducir las correspondientes tablas de probabilidades condicionadas.

Estamos observando aquí una propiedad esencial de las RR.BB.: no sólo los arcos aportan información sobre dependencias causales, sino que también *la ausencia de un arco es una forma (implícita) de aportar información.* En nuestro caso implica que U_1 y U_2 son independientes. Matemáticamente se expresa así:

$$P(u_2|u_1) = P(u_2) \tag{4.33}$$

o bien

$$P(u_1, u_2) = P(u_1) \cdot P(u_2) \tag{4.34}$$

Con esta hipótesis podemos por fin calcular la probabilidad de X:

$$P(x) = \sum_{u_1} \sum_{u_2} P(x|u_1, u_2) \cdot P(u_1) \cdot P(u_2) \tag{4.35}$$

En nuestro caso, el valor obtenido es $P(+x) = 0'3$, que concuerda con el de los ejemplos anteriores.

a) Supongamos que nos enteramos de que la persona en cuestión procede de una zona de alto riesgo. ¿Cual es la probabilidad de que padezca la enfermedad? Una de las formas posibles de realizar el cálculo es ésta:

$$P^*(x) = P(x|u_1^+) = \frac{P(x, u_1^+)}{P(u_1^+)}$$

Si definimos

$$\pi(x) \equiv P(x, \mathbf{e}) = P(x, u_1^+) \tag{4.36}$$

$$\alpha \equiv [P(\mathbf{e})]^{-1} = [P(u_1^+)]^{-1} \tag{4.37}$$

la ecuación anterior se convierte en

$$P^*(x) = \alpha \cdot \pi(x) \tag{4.38}$$

Podemos obtener $\pi(x)$ del siguiente modo:

$$\pi(x) = \sum_{u_2} P(x, u_1^+, u_2) = \sum_{u_2} P(x|u_1^+, u_2) \cdot P(u_1^+, u_2)$$

y aplicando la independencia a priori de las causas podemos expresar la ecuación anterior como

$$\pi(x) = \sum_{u_1} \sum_{u_2} P(x|u_1, u_2) \cdot \pi_X(u_1) \cdot \pi_X(u_2) \tag{4.39}$$

que es un resultado completamente general.

En el ejemplo que estamos tratando,

$$\mathbf{e} = \{u_1^+\} \Longrightarrow \left\{ \begin{array}{ll} \pi_X(u_1^+) = P(u_1^+) & \\ \pi_X(u_1^0) = 0 & \pi_X(u_2^\dagger) = P(u_2^\dagger) \\ \pi_X(u_1^-) = 0 & \pi_X(u_2^\ddagger) = P(u_2^\ddagger) \end{array} \right\} \tag{4.40}$$

y en consecuencia

$$\mathbf{e} = \{u_1^+\} \Longrightarrow \left\{ \begin{array}{l} \pi(+x) = 0'00202 \\ \pi(\neg x) = 0'09798 \end{array} \right. \tag{4.41}$$

Sustituyendo este resultado en la ecuación (4.38) y normalizando, hallamos que $P^*(+x) = 0'0202$; es decir, una persona originaria de una zona de alto riesgo tiene una probabilidad del 2% de padecer paludismo (frente al 0'3% general).

Las expresiones $\pi_X(u_i)$ que hemos utilizado en la deducción no son nuevas: aparecieron ya en la ecuación (4.26). Recordemos que el significado de $\pi_X(u_i)$ es

que transmite a X el impacto de toda la evidencia relativa a U_i. Como no hay evidencia relativa a U_2, $\pi_X(u_2)$ coincide con la probabilidad a priori.

b) Imaginemos ahora que por alguna razón tenemos certeza absoluta de que el enfermo padece paludismo. Antes de hacer un análisis de sangre, podemos predecir qué resultado es más probable, considerando cuál de los dos tipos sanguíneos explica mejor la presencia de la enfermedad:

$$P^*(u_2) = P(u_2|+x) = \frac{P(u_2)\cdot P(+x|u_2)}{P(+x)}$$

o bien

$$P^*(u_2) = \alpha \cdot P(u_2)\cdot \lambda_X(u_2) \tag{4.42}$$

donde

$$\lambda_X(u_2) \equiv P(+x|u_2) = \sum_{u_1} P(+x|u_1, u_2)\cdot P(u_1) \tag{4.43}$$

que en nuestro ejemplo vale

$$\mathbf{e} = \{+x\} \Longrightarrow \begin{cases} \lambda_X(u_2^\dagger) = 0'00204 \\ \lambda_X(u_2^\ddagger) = 0'00344 \end{cases} \tag{4.44}$$

Efectivamente, los valores de la tabla 4.1 nos han llevado a la conclusión de que el paludismo se explica mejor con el tipo sanguíneo $u_2^\ddagger$. Aplicando (4.42), obtenemos

$$\begin{cases} P^*(u_2^\dagger) = 0'471 \\ P^*(u_2^\ddagger) = 0'529 \end{cases} \tag{4.45}$$

Observamos que inicialmente el grupo $u_1^\dagger$ era el más probable (60%), pero ahora es el menos probable (47%) porque explica peor el paludismo.

El cálculo que hemos realizado para X y U_2 es idéntido al que hicimos en el ejemplo **1.a** para Y_1 y X. Vemos de nuevo que un mismo nodo puede ser fuente de información u objeto de predicción, dependiendo de la evidencia disponible.

c) Mostraremos ahora otra de las propiedades más características de las RR.BB.: la aparición de correlaciones entre los padres de un nodo. Continuando con el caso anterior, supongamos que además de tener la certeza de que el enfermo padece paludismo sabemos que procede de un país de alto riesgo; es decir, $\mathbf{e} = \{+x, u_1^+\}$. Aplicaremos de nuevo la ecuación (4.42), aunque ahora

$$\lambda_X(u_2) \equiv P(+x, u_1^+|u_2) = P(+x|u_1^+, u_2)\cdot P(u_1^+|u_2)$$

La independencia condicional nos dice que $P(u_1^+|u_2) = P(u_1^+)$, y tenemos, por tanto,

$$\begin{aligned} \lambda_X(u_2) &= P(x|u_1^+, u_2)\cdot P(u_1^+) \\ &= \sum_{u_1} P(x|u_1, u_2)\cdot \pi_X(u_1) \end{aligned} \tag{4.46}$$

donde $\pi_X(u_1)$ es el vector que apareció en la ecuación (4.40). Al realizar los cálculos obtenemos

$$\mathbf{e} = \{+x, u_1^+\} \Longrightarrow \begin{cases} \lambda_X(u_2^\dagger) = 0'0015 \\ \lambda_X(u_2^\ddagger) = 0'0028 \end{cases} \tag{4.47}$$

y de ahí

$$\begin{cases} P^*(u_2^\dagger) = 0'445 \\ P^*(u_2^\ddagger) = 0'555 \end{cases} \tag{4.48}$$

Si comparamos este resultado con el de la ecuación (4.45), observamos que la probabilidad de $u_2^\dagger$ ha descendido ligeramente como resultado de conocer la zona de origen: $U_1 = u_1^+$. Éste es el fenómeno que queríamos mostrar. Antes de conocer el valor de x, U_1 y U_2 eran *a priori independientes*, por lo que la probabilidad de u_2 no variaba al conocer el valor u_1 (ver la ec. (4.33)). Sin embargo, la independencia se pierde tanto al conocer "$X = +x$" como "$X = \neg x$". Dicho de otro modo,

$$P(u_2|u_1) = P(u_2) \tag{4.49}$$

$$P(u_2|u_1, x) \neq P(u_2|x) \tag{4.50}$$

Recordando el ejemplo 2 (pág. 59), vemos que allí ocurría precisamente lo contrario: las variables Y_1 e Y_2 estaban correlacionadas al principio, pero se volvían *condicionalmente independientes* al conocer el valor de X. Esta asimetría, que es un reflejo del sentido de la causalidad, se conoce como **separación direccional** y es la característica esencial de las redes bayesianas.

Ejemplo 4. Por último, consideremos el caso en que tenemos un nodo con dos causas y dos efectos (fig. 4.4). Las probabilidades condicionadas son las mismas que en los ejemplos anteriores. Por no extender demasiado esta sección, vamos a considerar solamente el caso en que tenemos un paciente que procede de una zona de alto riesgo y presenta fiebre, pero la prueba de la gota gruesa ha dado un resultado negativo. Es decir, $\mathbf{e} = \{u_1^+, \neg y_1, +y_2\}$.

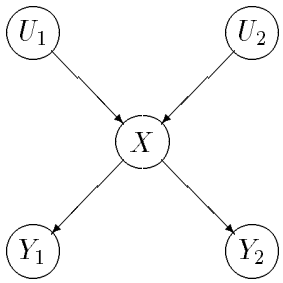


Figura 4.4: Nodo X con dos padres y dos hijos.

El teorema de Bayes nos dice que

$$P^*(x) = P(x|u_1^+, \neg y_1, +y_2) = \frac{P(x) \cdot P(u_1^+, \neg y_1, +y_2|x)}{P(u_1^+, \neg y_1, +y_2)} \tag{4.51}$$

Nuevamente necesitamos utilizar unos valores, $P(u_1^+, \neg y_1, +y_2|x)$, que no conocemos. (Si tuviéramos estos valores podríamos calcular también el denominador de la fracción.) Hemos introducido ya dos hipótesis:

1. Independencia a priori de los nodos que no tienen ningún antepasado común.
2. Independencia condicional de los dos efectos de X cuando conocemos con certeza el valor de X.

Vamos a enunciar ahora la tercera y última hipótesis, la independencia condicional (para cada valor x) entre los padres y los hijos de X

$$P(y_1, y_2|x, u_1, u_2) = P(y_1, y_2|x) \tag{4.52}$$

o lo que es lo mismo

$$P(u_1, u_2, y_1, y_2|x) = P(u_1, u_2|x) \cdot P(y_1, y_2|x) \tag{4.53}$$

La interpretación de estas dos ecuaciones es clara: la probabilidad de los efectos de X depende solamente del valor que toma X, no de la combinación de factores que nos ha llevado a dicho valor. En nuestro ejemplo significa que, si hay certeza de que una persona padece paludismo, la probabilidad de que tenga fiebre y de que detectemos la enfermedad en la prueba de laboratorio no depende del país de origen ni del tipo sanguíneo. Lo mismo podemos decir de la ausencia de paludismo.[4]

De la ecuación (4.53) se deduce fácilmente, sumando sobre u_2, que

$$P(u_1, y_1, y_2|x) = P(u_1|x) \cdot P(y_1, y_2|x) \tag{4.54}$$

con lo que la ecuación (4.51) se convierte en

$$P^*(x) = \alpha \cdot \pi(x) \cdot \lambda(x) \tag{4.55}$$

Recordemos que ya habíamos definido anteriormente $\pi(x)$ y $\lambda(x)$:

$$\pi(x) \equiv P(x) \cdot P(u_1^+|x) = P(x, u_1^+) \tag{4.56}$$

$$\lambda(x) \equiv P(\neg y_1, +y_2|x) \tag{4.57}$$

y que sus valores estaban dados por (4.41) y (4.25), respectivamente. Tras unos cálculos sencillos obtenemos que $P^*(+x) = 0'00942$; es decir, con estos hallazgos, la probabilidad de que el paciente tenga paludismo es menor del 1%.

Podríamos calcular ahora la probabilidad del tipo sanguíneo en función de la evidencia, $P^*(u_2)$, pero lo vamos a omitir para no alargar más la exposición.

La ecuación (4.55) es la fórmula fundamental para el cálculo de la probabilidad en redes bayesianas. En ella aparecen dos términos importantes, $\pi(x)$ y $\lambda(x)$. El primero de ellos transmite el impacto de la evidencia correspondiente a **las causas** de X. En nuestro caso, el único hallazgo "por encima" de X era $U_1 = u_1^+$. Si no tuviéramos ninguna evidencia, $\pi(x)$ sería simplemente la probabilidad a priori $P(x)$.

El segundo, $\lambda(x)$, transmite el impacto de la evidencia correspondiente a **los efectos** de X. En el ejemplo anterior, recogía la influencia de $\neg y_1$ y $+y_2$. Si no tuviéramos ninguna evidencia, $\lambda(x)$ sería un vector constante y podríamos prescindir de él al aplicar la ecuación (4.55), sin alterar el resultado.

[4] Si por alguna razón pensáramos que esta hipótesis no es cierta, deberíamos añadir a nuestro modelo nuevos arcos con el fin de representar las influencias existentes (por ejemplo, entre el país de origen y otras causas de la fiebre) y asignarles las tablas de probabilidad oportunas.

Recapitulación

En esta sección hemos visto las propiedades más importantes de las RR.BB. En primer lugar, que la red contiene información cualitativa (la estrucura del grafo) y cuantitativa (las probabilidades a priori y condicionales). Esta red constituye nuestro modelo causal y —salvo que introduzcamos algún mecanismo de aprendizaje— es invariable.

El proceso de diagnóstico consiste en introducir la evidencia disponible (asignar valores a las variables conocidas) y calcular la probabilidad a posteriori de las variables desconocidas. Se trata en realidad de un proceso de inferencia, aunque no es simbólica sino numérica.

Hemos visto además que este modelo permite tanto un razonamiento diagnóstico (cuál es la causa más probable) como predictivo (qué valor de cierta variable aparecerá con mayor probabilidad). Por otra parte, hemos comentado ya que una ventaja de las RR.BB. es que un mismo nodo puede ser fuente de información u objeto de predicción dependiendo de cuál sea la evidencia disponible, como ocurría con X o con Y_1 en los ejemplos anteriores.

Y hemos comprobado también en el ejemplo 4 que es posible realizar un cálculo incremental, modificando la probabilidad de las variables a medida que va llegando nueva evidencia, sin tener que recalcular todos los mensajes $\pi()$ y $\lambda()$.

4.2 Definición formal de red bayesiana

En la sección anterior hemos presentado de forma intuitiva qué son las redes bayesianas y cómo se propaga la evidencia. Allí vimos la importancia de las hipótesis de independencia. Ahora vamos a dar una definición matemática formal.

4.2.1 Estructura de la red. Teoría de grafos

Nuestro punto de partida consiste en un conjunto finito de *nodos* $\mathcal{N}$. Cada uno de ellos representa una *variable*, que puede ser discreta o continua. Esta relación biunívoca entre nodos y variables nos permite utilizar indistintamente ambos términos. Según indicamos en la sección anterior, utilizaremos una letra mayúscula (X, por ejemplo) para indicar un nodo o variable, y la correspondiente letra minúscula, x, para indicar un valor cualquiera de dicha variable. Es importante señalar que los valores de una variable deben constituir un conjunto exclusivo y exhaustivo. "Exclusivo" significa que dos valores diferentes no pueden ser ciertos al mismo tiempo; "exhaustivo" significa que la variable no puede tomar ningún valor fuera de este conjunto.

Así, en contraposición a lo que ocurría en el método probabilístico clásico (sec. 2.3.1), al construir una R.B. sería un error definir una variable llamada Enfermedad, cuyos valores representasen los posibles diagnósticos correspondientes a la fiebre: neumonía, amigdalitis, paludismo, etc.[5] La razón es que este conjunto de diagnósticos no es exclusivo (un enfermo puede padecer dos enfermedades a la vez) ni exhaustivo (difícilmente incluiremos en nuestro modelo todas las enfermedades diagnosticables). Lo correcto es asignar a cada una de estas patologías una variable y, en consecuencia, un nodo en la red; de este modo, el diagnóstico de múltiples

[5] Las redes de semejanza de Heckerman (sec. 10.1.3) consituyen una notable excepción a esta regla.

enfermedades resulta algo completamente natural en RR.BB.

Introducimos a continuación algunas definiciones básicas sobre grafos:

▷ **Arco.** Es un par ordenado de nodos (X, Y).

Esta definición de arco corresponde a lo que en otros lugares se denomina *arco dirigido*. En la representación gráfica, un arco (X, Y) viene dado por una flecha desde X hasta Y, tal como muestran las figuras de los ejemplos anteriores.

▷ **Grafo dirigido.** Es un par $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ donde $\mathcal{N}$ es un conjunto de nodos y $\mathcal{A}$ un conjunto de arcos definidos sobre los nodos.

Si hubiéramos definido los arcos como pares no ordenados, tendríamos un grafo no dirigido.[6]

▷ **Padre.** X es un *padre* de Y si y sólo si existe un arco (X, Y).

Los padres de X se representan como $pa(X)$. Por semejanza con el convenio utilizado para variables y sus valores, $pa(x)$ representará el vector formado al asignar un valor a cada nodo del conjunto $pa(X)$.

▷ **Hijo.** Y es un *hijo* de X si y sólo si existe un arco (X, Y).

▷ **Antepasado.** X es un *antepasado* de Z si y sólo si existe (al menos) un nodo Y tal que X es padre de Y e Y es antepasado de Z.

▷ **Descendiente.** Z es un *descendiente* de X si y sólo si X es un antepasado de Z.

▷ **Familia** X**.** Es el conjunto formado por X y los padres de X, $pa(X)$.

▷ **Nodo terminal.** Es el nodo que no tiene hijos.

▷ **Camino.** Un *camino* entre X_1 y X_N en una sucesión de nodos $\{X_1, \ldots, X_N\}$ pertenecientes a un grafo $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, tal que $X_i \neq X_j$ para $1 \leq i < j \leq N$ y

$$(X_i, X_{i+1}) \in \mathcal{A} \quad \text{ó} \quad (X_{i+1}, X_i) \in \mathcal{A}, \qquad \forall i, 1 \leq i < N$$

Es decir, dos nodos consecutivos X_i y X_{i+1} están unidos por un arco del primero al segundo o viceversa. Observar que esta definición corresponde a lo que en otros lugares se conoce como *camino abierto*.

▷ **Ciclo.** Es una sucesión de nodos $\{X_1, \ldots, X_N\}$ pertenecientes a un grafo $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, tal que *(1)* $X_i \neq X_j$ para $1 \leq i < j \leq N$, *(2)* para todo $i < N$ existe en $\mathcal{A}$ un arco (X_i, X_{i+1}), y *(3)* existe además un arco (X_N, X_1).

▷ **Bucle.** Sucesión de nodos $\{X_1, \ldots, X_N\}$ pertenecientes a un grafo $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, tal que *(1)* $X_i \neq X_j$ para $1 \leq i < j \leq N$, *(2)* para todo $i < N$ existe en $\mathcal{A}$ un arco (X_i, X_{i+1}) o (X_{i+1}, X_i), *(3)* existe además un arco (X_N, X_1) ó (X_1, X_N) y *(4)* los arcos no forman un ciclo.

[6] Las redes de Markov se basan en grafos no dirigidos, mientras que las redes bayesianas corresponden a grafos dirigidos.

Tanto el ciclo como el bucle corresponden a lo que a veces se denominan *caminos cerrados simples*. La diferencia es que en un ciclo los arcos van de cada nodo al siguiente (nunca a la inversa), mientras que la definición de bucle permite que los arcos tengan cualquiera de los dos sentidos, con la única condición de que no formen un ciclo. La distinción entre ambos es muy importante para el tema que nos ocupa porque las redes bayesianas suelen contener bucles pero, por definición, nunca pueden contener ciclos.

Ejemplo. En la figura 4.5.*a*, vemos que entre B y C hay dos *caminos*: $\{B, A, C\}$ y $\{B, D, C\}$, y lo mismo ocurre en 4.5.*b* y 4.5.*c*. El primero de estos tres grafos es un *ciclo*, mientras que que los dos últimos son *bucles*. □

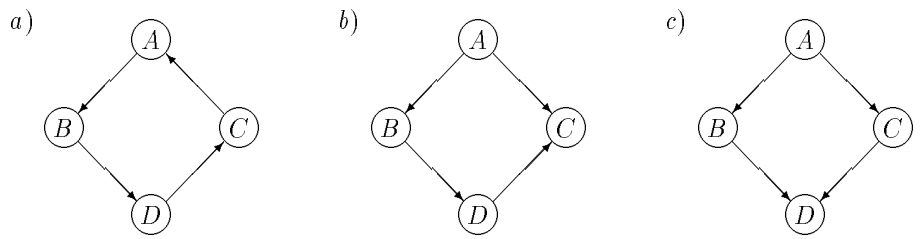


Figura 4.5: Un ciclo y dos bucles.

▹ **Grafo conexo.** Un grafo es *conexo* si entre dos cualesquiera de sus nodos hay al menos un camino.

▹ **Grafo simplemente conexo.** Un grafo es *simplemente conexo* si entre dos cualesquiera de sus nodos hay exactamente un camino.

Por tanto, un grafo simplemente conexo es un grafo conexo que no contiene ciclos ni bucles.

▹ **Grafo múltiplemente conexo.** Es el que contiene ciclos o bucles.

▹ **Poliárbol.** Es un grafo dirigido simplemente conexo.

La diferencia entre árbol y poliárbol es que, en el primero, cada nodo tiene como máximo un padre.

▹ **Grafo acíclico.** Es el grafo en que no hay ciclos.

Un grafo acíclico puede ser dirigido o no-dirigido.

4.2.2 Nociones de probabilidad

Antes de definir las redes bayesianas, vamos a repasar brevemente algunos conceptos básicos de la teoría de la probabilidad. Empezamos por establecer la siguiente notación: $P(x)$ indica la probabilidad de que la variable X tome el valor x; es decir,

$$P(x) \equiv P(X = x) \tag{4.58}$$

Para un conjunto de variables $\bar{X} = \{X_1, \ldots, X_n\}$, la probabilidad de la n-tupla $(x_1, \ldots, x_n)$ es la probabilidad de que $X_i = x_i$ para todo i:

$$P(x_1, \ldots, x_n) \equiv P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) \tag{4.59}$$

▹ **Probabilidad conjunta.** Dado un conjunto de variables $\mathcal{N} = \{X_1, \ldots, X_n\}$, definimos la *probabilidad conjunta* como una aplicación que a cada n-tupla $(x_1, \ldots, x_n)$ le asigna un número real no negativo de modo que[7]

$$\sum_{x_i} P(x_1, \ldots, x_n) = 1 \tag{4.60}$$

▹ **Probabilidad marginal.** Dada una distribución de probabilidad conjunta $P(x_1, \ldots, x_n)$, la *probabilidad marginal* para un subconjunto de variables $\mathcal{N}' = \{Y_1, \ldots, Y_{n'}\} \subset \mathcal{N}$ viene dada por

$$P(y_1, \ldots, y_{N'}) = \sum_{x_i \notin \mathcal{N}'} P(x_1, \ldots, x_n) \tag{4.61}$$

La distribución marginal para una variable es, por tanto,

$$P(x_i) = \sum_{x_j \neq x_i} P(x_1, \ldots, x_n) \tag{4.62}$$

Con el fin de simplificar las expresiones siguientes, vamos a utilizar la notación $\bar{x} \equiv (x_1, \ldots, x_n)$ para representar la n-tupla correspondiente a los valores de las variables del conjunto $\bar{X} = \{X_1, \ldots, X_n\}$.

▹ **Probabilidad condicional.** Dados dos subconjuntos disjuntos de variables, $\bar{X} = \{X_1, \ldots, X_n\}$, con $P(\bar{x}) = P(x_1, \ldots, x_n) > 0$, e $\bar{Y} = \{Y_1, \ldots, Y_m\}$,, la *probabilidad condicional* del segundo dado el primero se define como

$$P(\bar{y}|\bar{x}) = \frac{P(\bar{y}, \bar{x})}{P(\bar{x})} \tag{4.63}$$

▹ **Independencia.** Dos subconjuntos $\bar{X}$ y $\bar{Y}$ son *independientes* si se cumple una de estas tres condiciones:

1. $P(\bar{x}) = 0$
2. $P(\bar{y}) = 0$
3. $P(\bar{y}|\bar{x}) = P(\bar{y})$

▹ **Independencia condicional.** Dos subconjuntos $\bar{X}$ y $\bar{Y}$ son *condicionalmente independientes* dado un tercer conjunto $\bar{Z} = \{Z_1, \ldots, Z_p\}$ si se cumple una de estas tres condiciones:

1. $P(\bar{x}|\bar{z}) = 0$
2. $P(\bar{y}|\bar{z}) = 0$

[7] Para una definición más formal, ver [155, cap. 2].

3. $P(\bar{y}|\bar{x},\bar{z}) = P(\bar{y}|\bar{z})$

▷ **Separación.** El conjunto $\bar{Z}$ *separa* los conjuntos $\bar{X}$ e $\bar{Y}$ si estos dos últimos son condicionalmente independientes dado $\bar{Z}$.

▷ **Hallazgo.** Es la determinación del valor de una variable, $E = e$, a partir de un dato (una observación, una medida, etc.).

▷ **Evidencia.** Es un conjunto de hallazgos: $\mathbf{e} = \{E_1 = e_1, \ldots, E_r = e_r\}$.

En nuestra presentación, supondremos que la evidencia corresponde a nodos terminales, aunque los resultados obtenidos bajo esta hipótesis son fácilmente generalizables.

▷ **Probabilidad a priori.** Es la probabilidad de una variable o subconjunto de variables cuando no hay evidencia.

La probabilidad a priori de $\bar{X}$ coincide, por tanto, con la probabilidad marginal $P(\bar{x})$.

▷ **Probabilidad a posteriori.** Es la probabilidad de una variable o subconjunto de variables condicionada a la evidencia $\mathbf{e}$. La representaremos mediante P^*:

$$P^*(\bar{x}) \equiv P(\bar{x}|\mathbf{e}) \tag{4.64}$$

Mostramos a continuación algunas propiedades de la probabilidad.

Proposición 4.1 *Dados dos subconjuntos disjuntos, $\bar{X}$ e $\bar{Y}$, se cumple*

$$P(\bar{x}) = \sum_{y_j} P(\bar{x},\bar{y})$$

La demostración es inmediata a partir de la definición de probabilidad marginal dada en (4.61).

Proposición 4.2 *Para un subconjunto cualquiera de variables, $\bar{X}$, si $P(x_{j+1},\ldots,x_n) > 0$, se cumple*

$$P(\bar{x}) = P(x_1,\ldots,x_n) = P(x_1,\ldots,x_j|x_{j+1},\ldots,x_n)\cdot P(x_{j+1},\ldots,x_n)$$

Es evidente a partir de la definición de probabilidad condicional (ec. (4.63)). Equivale a decir que, para dos subconjuntos disjuntos $\bar{X}$ y $\bar{Y}$, tenemos que $P(\bar{x},\bar{y}) = P(\bar{x}|\bar{y})\cdot P(\bar{y})$ cuando $P(\bar{y}) > 0$.

Proposición 4.3 *Dados tres subconjuntos disjuntos $\bar{X}$, $\bar{Y}$ y $\bar{Z}$, si $P(\bar{y},\bar{z}) > 0$, se cumple*

$$P(\bar{x},\bar{y}|\bar{z}) = P(\bar{x}|\bar{y},\bar{z})\cdot P(\bar{y}|\bar{z})$$

Demostración. Vemos primero que

$$P(\bar{z}) = \sum_{\bar{y}} P(\bar{y}, \bar{z}) > 0$$

Aplicando la proposición anterior, tenemos

$$P(\bar{x}, \bar{y}|\bar{z}) = \frac{P(\bar{x}, \bar{y}, \bar{z})}{P(\bar{z})} = \frac{P(\bar{x}|\bar{y}, \bar{z}) \cdot P(\bar{y}, \bar{z})}{P(\bar{z})} = P(\bar{x}|\bar{y}, \bar{z}) \cdot P(\bar{y}|\bar{z})$$

como queríamos demostrar. □

Proposición 4.4 *Para dos subconjuntos independientes, $\bar{X}$ e $\bar{Y}$, se cumple*

$$P(\bar{x}, \bar{y}) = P(\bar{x}) \cdot P(\bar{y})$$

Proposición 4.5 *Para dos subconjuntos $\bar{X}$ e $\bar{Y}$ condicionalmente independientes dado el subconjunto $\bar{Z}$, se cumple*

$$P(\bar{x}, \bar{y}|\bar{z}) = P(\bar{x}|\bar{z}) \cdot P(\bar{y}|\bar{z})$$

Las dos últimas proposiciones se deducen fácilmente de las definiciones de independencia e independencia condicional, respectivamente.

Teorema 4.6 (Teorema de Bayes) *Para dos subconjuntos $\bar{X}$ e $\bar{Y}$, si $P(\bar{x}) > 0$ y $P(\bar{y}) > 0$, se cumple*

$$P(\bar{y}|\bar{x}) = \frac{P(\bar{y}) \cdot P(\bar{x}|\bar{y})}{P(\bar{x})}$$

Demostración. Por la definición de probabilidad condicional,

$$P(\bar{y}|\bar{x}) \cdot P(\bar{x}) = P(\bar{y}, \bar{x})$$

$$P(\bar{x}|\bar{y}) \cdot P(\bar{y}) = P(\bar{x}, \bar{y})$$

A partir de aquí, la demostración es inmediata. □

Proposición 4.7 (Fórmula de Bayes con condicionamiento) *Dados tres subconjuntos $\bar{X}$, $\bar{Y}$ y $\bar{Z}$, si $P(\bar{x}, \bar{z}) > 0$ y $P(\bar{y}, \bar{z}) > 0$, se cumple*

$$P(\bar{y}|\bar{x}, \bar{z}) = \frac{P(\bar{y}|\bar{z}) \cdot P(\bar{x}|\bar{y}, \bar{z})}{P(\bar{x}|\bar{z})}$$

Demostración. Basta comprobar que

$$\begin{aligned} P(\bar{y}|\bar{x}, \bar{z}) &= \frac{P(\bar{y}, \bar{x}, \bar{z})}{P(\bar{x}, \bar{z})} = \frac{P(\bar{x}|\bar{y}, \bar{z}) \cdot P(\bar{y}|\bar{z})}{P(\bar{x}, \bar{z})} \\ &= \frac{P(\bar{x}|\bar{y}, \bar{z}) \cdot P(\bar{y}|\bar{z})}{P(\bar{x}|\bar{z})} \end{aligned}$$

pues $P(\bar{x}, \bar{z}) > 0$ implica que $P(\bar{x}) > 0$. □

4.2.3 Definición de red bayesiana

La propiedad fundamental de una red bayesiana es la separación direccional (llamada *d-separation* por Pearl [172, 174]), que se define así:

▹ **Separación direccional.** Dado un grafo dirigido acíclico conexo y una distribución de probabilidad sobre sus variables, se dice que hay *separación direccional* si, dado un nodo X, el conjunto de sus padres, $pa(X)$, separa condicionalmente este nodo de todo otro subconjunto $\bar{Y}$ en que no haya descendientes de X. Es decir,

$$P(x|pa(x), \bar{y}) = P(x|pa(x)) \tag{4.65}$$

Es habitual definir las redes bayesianas a partir de grafos dirigidos acíclicos (en inglés, *directed acyclic graph, DAG*). Sin embargo, nos parece importante incluir la especificación "*conexo*" por tres razones. La primera, porque muchos de los algoritmos y propiedades de las redes bayesianas sólo son correctos para grafos conexos, por lo que es mejor incluir esta característica en la definición que tener que añadirla como nota a pie de página en casos particulares. La segunda razón es que, aun en el caso de que tuviéramos un modelo con dos partes inconexas, podríamos tratarlo como dos redes bayesianas independientes. Y la tercera, porque los modelos del mundo real con que vamos a trabajar son siempre conexos; si hubiera dos partes inconexas no tendríamos uno sino dos modelos independientes.

La definición de separación direccional, aunque pueda parecer extraña a primera vista, es sencilla, y ya fue introducida en los ejemplos de la sección 4.1. En efecto, volviendo a la figura 4.4 de dicha sección (pág. 66), recordamos que, una vez conocido el valor de x, podíamos calcular la probabilidad de y_1 sin que influyeran los valores de las demás variables. Es decir, el conjunto $pa(Y_1) = \{X\}$, separa condicionalmente Y_1 de todas las demás variables.

A partir de las definiciones anteriores, podemos caracterizar las redes bayesianas así:

▹ **Red bayesiana.** Es un grafo dirigido acíclico conexo más una distribución de probabilidad sobre sus variables, que cumple la propiedad de separación direccional.

El término *direccional* hace referencia a la asimetría de dicha propiedad, que se manifiesta en las siguientes propiedades de las redes bayesianas, ilustradas con el ejemplo de la figura 4.6:

1. Si A no tiene padres, entonces $P(x|pa(x)) = P(x|\emptyset) = P(x)$, y la ecuación (4.65) se traduce en $P(e|a) = P(e)$ para cada nodo E que no sea uno de los descendientes de A; en otras palabras, E es a priori independiente de A. En consecuencia, dos nodos cualesquiera D y E que no tengan ningún antepasado común son independientes a priori.

2. Si D es descendiente de A y antepasado de H, y no existe ningún otro camino desde A hasta H, entonces estos dos nodos quedan condicionalmente separados por D:

$$P(h|d, a) = P(h|d). \tag{4.66}$$

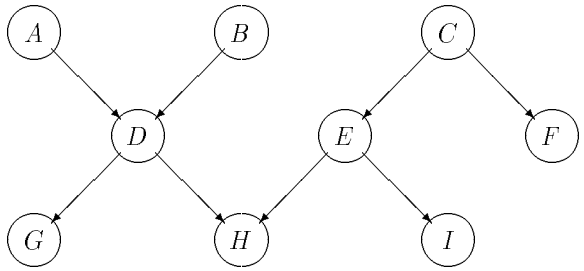


Figura 4.6: Un pequeño poliárbol.

3. Si tanto G como H son hijos de D y no tienen ningún otro antepasado común, este último separa G y H, haciendo que sean condicionalmente independientes:

$$P(g|d,h) = P(g|d). \tag{4.67}$$

4. Dos nodos que son independientes (a priori o condicionalmente) siguen siéndolo mientras no se conozca el valor de ninguno de sus descendientes. Por ejemplo, A es independiente de F aunque conozcamos el valor de I:

$$P(a|f,i) = P(a|i)\,, \tag{4.68}$$

y del mismo modo

$$P(e|f,c,h) = P(e|c,h)\,. \tag{4.69}$$

En general, la independencia (a priori o condicional) de dos nodos, A y E, se pierde al conocer el valor de H, pues en este caso la propiedad de separación direccional ya no es aplicable.

En la definición de red bayesiana, hemos partido de una distribución de probabilidad conjunta para las variables: $P(\mathcal{N}) = P(\bar{x})$. Aparentemente, en el caso de variables binarias, harían falta $2^N - 1$ parámetros. (Serían 2^N si no existiera la ligadura (4.60).) Sin embargo, las condiciones de independencia dadas por la separación direccional imponen nuevas restricciones, que reducen los grados de libertad del sistema. Ciertamente, una de las propiedades más importantes de una red bayesiana es que su distribución de probabilidad puede expresarse mediante el producto de las distribuciones condicionadas de cada nodo dados sus padres, tal como nos dice el siguiente teorema. (Recordemos que, para un nodo X sin padres, $pa(X) = \emptyset$ y, por tanto, $P(x|pa(x)) = P(x)$.)

Teorema 4.8 (Factorización de la probabilidad) *Dada una red bayesiana, su distribución de probabilidad puede expresarse como*

$$P(x_1, \ldots, x_n) = \prod_i P(x_i|pa(x_i))$$

Demostración. Es fácil construir una ordenación de las variables en que los padres de cada nodo aparezcan siempres después de él. Supongamos, sin périda de generalidad, que la ordenación $\{X_1, \ldots, X_n\}$ cumple dicha propiedad. Aplicando sucesivamente la proposición 4.2, podemos escribir

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i+1}, \ldots, x_n)$$

Ahora bien, por la forma en que hemos escogido la ordenación, el conjunto $\{X_{i+1}, \ldots, X_n\}$ incluye todos los padres de X y, en consecuencia, la separación direccional (ec. (4.65)) nos dice que

$$P(x_i | x_{i-1}, \ldots, x_1) = P(x_i | pa(x_i))$$

con lo que concluimos la demostración. □

La importancia de este teorema es que nos permite describir una red bayesiana a partir de la probabilidad condicionada de cada nodo (o la probabilidad a priori en el caso de nodos sin padres), en vez de dar la distribución de probabilidad conjunta, que requeriría un número de parámetros exponencial en el número de nodos y plantearía el grave problema de verificar la propiedad de separación direccional; sin embargo, el número de parámetros requerido para dar las probabilidades condicionadas es proporcional al número de nodos (suponiendo que el número de padres y el número de valores posibles están acotados para cada variable).

Podríamos haber definido las propiedades de independencia en términos de caminos activados o bloqueados, al estilo de Pearl, Geiger y Verma [174, págs. 317-318], [177], seguido también por Charniak [36]. La presentación que hemos escogido se parece más a la propuesta por Neapolitan [155, cap. 5].

4.3 Semántica de las redes bayesianas

Hemos definido ya las redes bayesianas desde un punto de vista matemático formal. La cuestión que nos planteamos ahora es su semántica, es decir, ¿qué interpretación se le puede dar a una red bayesiana? ¿Cómo se corresponde nuestro modelo con el mundo real? ¿Por qué podemos hablar de causas y efectos en una R.B.?

Esta cuestión está ya parcialmente respondida en la sección 4.1, que fue situada antes de la definición formal de R.B. precisamente para mostrar que los conceptos y axiomas introducidos no pareciesen arbitrarios, sino que responden a las propiedades de la causalidad, según nuestra concepción intuitiva del mundo real.

Es importante señalar que la estructura de la red, por sí misma, aporta gran cantidad de información cualitativa. En efecto, un arco XY indica, ya antes de conocer el valor concreto de la probabilidad condicional, que hay una correlación entre ambas variables: el valor que toma X influye sobre la probabilidad de Y, y viceversa. Es lo que llamamos *influencia causal directa*. Tal es la relación que existe, por ejemplo, entre el país de origen y el paludismo, o entre el paludismo y la fiebre. Profundizando un poco más, observamos que la existencia de un camino entre dos variables X e Y, con variables intermedias $\bar{Z}$, indica que hay una *influencia causal indirecta* entre ambas.

Tal como hemos discutido en la presentación intuitiva de las RR.BB., cuando nuestro sentido común, basado en la experiencia, nos dice que la influencia de una variable X sobre uno de sus efectos Y_1 (por ejemplo, del paludismo sobre la prueba de la gota gruesa) no depende de cuáles han sido las causas o mecanismos que han producido X, ni depende tampoco de si X a dado lugar a otros efectos, entonces la red contendrá un arco desde X hasta Y_1, y no habrá ningún arco que conecte Y_1 con las demás variables. Por tanto, *la ausencia de arcos* es también una forma de expresar información. El hecho de que Y_1 depende solamente de su causa, X, se traduce matemáticamente diciendo que, conocido el valor de X, la probabilidad de Y_1 es independiente de los valores que toman esas otras variables, o dicho de otro modo, X separa Y_1 de dichas variables. Empezamos a ver aquí la relación entre el concepto de padre y el de causa, entre el de hijo y el de efecto, entre el arco y el de influencia causal directa, entre el de independencia en los mecanismos causales y el de independencia probabilística.

En este punto es donde se manifiesta la importancia del sentido de los arcos y su relación con la idea de causalidad. Volviendo al ejemplo del paludismo, el hecho de que las variables País-de-origen y Tipo-sanguíneo no tengan ningún padre común significa que son a priori independientes, es decir, que el país no influye en el grupo sanguíneo y viceversa, de modo que, si no hay más evidencia, no podemos obtener ninguna información sobre el país de origen a partir del grupo sanguíneo ni al contrario. Sin embargo, el hecho de que ambas variables tengan un hijo común significa que, una vez conocido el valor de ese nodo, surgen correlaciones entre los padres.[8] Podemos decir, usando la terminogía de Pearl [172], que el camino entre U_1 y U_2 permanece *bloquedo* hasta que sea *activado* por la llegada de información sobre X o sobre alguno de sus descendientes.

Para el caso de los efectos de una variable ocurre precisamente lo contrario: todo médico sabe que hay correlación entre la fiebre y el test de la gota gruesa. Sin embargo, tal como discutimos en la sección 4.1, la correlación desaparece cuando averiguamos si el paciente tiene o no tiene paludismo. Es decir, el camino entre Y_1 e Y_2 está *activado* en principio, y *se bloquea* sólo al conocer el valor de X. De esta asimetría entre padres e hijos, reflejo de la asimetría que existe en el mundo real entre causas y efectos, procede el nombre de separación *direccional*.

Por tanto, hay dos formas de justificar los enlaces que introducimos u omitimos al construir nuestra red. La primera es de naturaleza teórica: formamos un modelo causal a partir de la experiencia de un especialista y trazamos los arcos correspondientes al modelo; la relación que hemos discutido entre los mecanismos causales y las propiedades matemáticas de independencia nos permite fundamentar nuestro modelo. El otro camino para justificar la red consiste en realizar una comprobación empírica a partir de un conjunto suficientemente amplio de casos, utilizando las herramientas estadísticas que se emplean habitualmente para detectar correlaciones.

Hay otro punto relativo a la semántica de las redes bayesianas, que vamos a mencionar sólo brevemente, pues aún está muy discutido. Nos referimos al debate entre los que defienden que las redes expresan causalidad y los que sostienen que éstas sólo expresan correlaciones entre variables. En realidad, no se trata de un debate limitado al campo de las redes bayesianas, sino que la existencia de la causalidad es

[8] La correlación que aparece entre las causas se aprecia mucho más claramente en el caso de la puerta OR (sec. 5.2).

una cuestión que se han planteado matemáticos y filósofos por lo menos desde el siglo XVII, a partir de las teorías de Hume. Para no entrar en esta cuestión citaremos solamente tres trabajos, de Pearl y Verma [179, 176] y de Druzdzel y Simon [80], que muestran cómo recientemente han surgido argumentos matemáticos para defender la interpretación causal frente a la meramente correlacional.

En resumen, lo que hemos intentado mostrar en esta sección es que la información cualitativa que expresa la estructura de una R.B. es más importante aún que la información cuantitativa, como lo demuestra el hecho de que se han construido *redes cualitativas* (sec. 10.1.3), capaces de razonar en base a las propiedades de independencia de las redes bayesianas, incluso en ausencia de valores numéricos. Por este motivo, Neapolitan [155] ha sugerido en nombre de *redes de independencia* (*independence networks*) como el más adecuado para las RR.BB.

Podríamos sintetizar todo lo dicho anteriormente repitiendo lo que Laplace afirmó en la introducción de su famoso libro *Théorie Analytique des Probabilités*:[9]

> La teoría de la probabilidad no es, en el fondo, más que el sentido común reducido al cálculo.

[9]Cita tomada de Druzdzel [77].

Capítulo 5

Propagación de evidencia en poliárboles

La sección 4.2 planteó la teoría de las redes bayesianas (RR.BB.) mediante una definición matemática formal. De acuerdo con los tres niveles propuestos por David Marr [140, sec. 2.1]: **teoría**, **algoritmos** e **implementación**, vamos a dedicar ahora dos capítulos a la computación de la probabilidad. En éste primero estudiaremos las redes simplemente conexas (poliárboles) y en el siguiente veremos cómo tratar los bucles. Primero desarrollaremos con detalle los algoritmos y discutiremos después las diferentes implementaciones posibles: recursiva, iterativa y —la que más nos interesa— distribuida.

5.1 Caso general

En realidad, dada una R.B., a partir de las probabilidades condicionales podríamos calcular la probabilidad conjunta según el teorema 4.8, y luego aplicar las ecuaciones (4.61) y (4.63) para calcular las probabilidades marginal y a posteriori, respectivamente. Sin embargo este método tendría complejidad exponencial incluso en el caso de poliárboles. Además, al añadir nueva evidencia tendríamos que repetir casi todos los cálculos. Por esta razón conviene encontrar algoritmos mucho más eficientes.

El algoritmo para poliárboles que presentamos en esta sección, basado en el paso de mensajes π y λ, fue desarrollado por Kim [125] a partir del que Pearl había propuesto para árboles [170]. Al hablar de "caso general" nos referimos a que es válido para cualquier distribución condicional $P(x|\, pa(x))$, en contraposición a modelos particulares como las puertas AND y OR, que estudiaremos más adelante.

La principal limitación de este algoritmo es que no permite tratar los bucles, que aparecen inevitablemente al desarrollar modelos del mundo real, por lo que en sí mismo resulta de muy poca utilidad y los constructores de RR.BB. recurren a otros que, aun perdiendo las ventajas de éste, son aplicables a todo tipo de RR.BB. Sin embargo, aquí lo vamos a estudiar con detalle por dos razones. Primera, por su sencillez y elegancia, que nos permitirán comprender mejor las propiedades de las RR.BB. Y segunda, porque el algoritmo de condicionamiento local que presentaremos en el próximo capítulo, aplicable también a redes múltiplemente conexas, es una extensión de éste, que utiliza los mismos conceptos y definiciones.

Para comprender mejor el desarrollo matemático que vamos a realizar, puede ser

útil al lector repasar la sección 4.1, en que aparecen sencillos ejemplos numéricos que explican por qué se intruducen las definiciones de π y λ y cómo se propaga la evidencia.

5.1.1 Definiciones básicas

Una de las propiedades fundamentales de un poliárbol es que hay un único camino entre cada par de nodos. En consecuencia, la influencia de cada hallazgo se propaga hasta un nodo X bien a través de los padres o a través de los hijos de éste, por lo que para cada nodo X podemos hacer una partición de la evidencia (recordamos que la evidencia es el conjunto de hallazgos) en dos subconjuntos, tales que

$$\mathbf{e} = \mathbf{e}_X^+ \cup \mathbf{e}_X^- \tag{5.1}$$

$$\mathbf{e}_X^+ \cap \mathbf{e}_X^- = \emptyset \tag{5.2}$$

donde $\mathbf{e}_X^+$ representa la evidencia "por encima de X" y $\mathbf{e}_X^-$ "por debajo de X" en el sentido antes mencionado.

De forma similar, la eliminación de un enlace XY divide a la red —y por tanto también la evidencia— en dos partes, una que queda "por encima" del enlace y otra que queda "por debajo". Las llamaremos $\mathbf{e}_{XY}^+$ y $\mathbf{e}_{XY}^-$, respectivamente. Al igual que en el caso anterior, se cumple

$$\mathbf{e} = \mathbf{e}_{XY}^+ \cup \mathbf{e}_{XY}^- \tag{5.3}$$

$$\mathbf{e}_{XY}^+ \cap \mathbf{e}_{XY}^- = \emptyset \tag{5.4}$$

Ejemplo. En la figura 4.6 (pág. 75), si tuviéramos $\mathbf{e} = \{+f, +g, \neg i\}$, entonces $\mathbf{e}_E^+ = \{+f\}$ y $\mathbf{e}_E^- = \{+g, \neg i\}$. Del mismo modo, $\mathbf{e}_H^+ = \{+f, +g, \neg i\}$ y $\mathbf{e}_H^- = \emptyset$. La eliminación del enlace EH dividiría la red en dos partes, y tendríamos $\mathbf{e}_{EH}^+ = \{+f, \neg i\}$ y $\mathbf{e}_{EH}^- = \{+g\}$. □

Basándonos en la partición de la evidencia, podemos establecer las siguientes definiciones (ver fig. 5.1):

$$\pi(x) \equiv P(x, \mathbf{e}_X^+) \tag{5.5}$$

$$\lambda(x) \equiv P(\mathbf{e}_X^- | x) \tag{5.6}$$

$$\pi_X(u_i) \equiv P(u_i, \mathbf{e}_{U_i X}^+) \tag{5.7}$$

$$\lambda_{Y_j}(x) \equiv P(\mathbf{e}_{XY_j}^- | x) \tag{5.8}$$

El sentido de estas definiciones es el siguiente:

- $\pi(x)$ indica qué valor de X es más probable según la evidencia relacionada con las causas de X (es decir, según la evidencia "por encima" de X).
- $\lambda(x)$ indica qué valor de X explica mejor los hallazgos correspondientes a los efectos de X (la evidencia "por debajo" de X).
- $\pi_X(u)$ indica qué valor de U es más probable según la evidencia "por encima" del enlace UX.

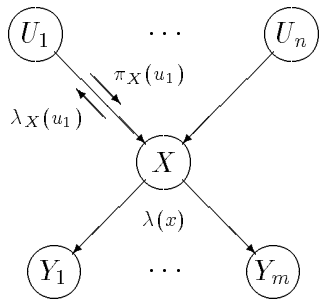


Figura 5.1: Propagación de evidencia mediante intercambio de mensajes.

- $\lambda_Y(x)$ indica qué valor X explica mejor la evidencia "por debajo" del enlace XY.

Para entender mejor estas explicaciones, conviene volver a los ejemplos de la sección 4.1.

Conviene señalar que las definiciones anteriores, aunque tomadas del libro de Pearl [174], han sido modificadas de acuerdo con la propuesta de Peot y Shachter [181], con el fin de permitir un tratamiento coherente de los bucles según el algoritmo de condicionamento local que presentaremos en el próximo capítulo.

5.1.2 Computación de los mensajes

Recordemos una vez más que nuestro objetivo es calcular la probabilidad a posteriori de cada nodo, definida en la ecuación (4.64). A partir, de ahí,

$$\begin{aligned} P^*(x) &= P(x|\,\mathbf{e}) = \alpha\, P(x, \mathbf{e}_X^+, \mathbf{e}_X^-) \\ &= \alpha\, P(x, \mathbf{e}_X^+)\, P(\mathbf{e}_X^-|\, x, \mathbf{e}_X^+) \end{aligned}$$

donde hemos definido

$$\alpha \equiv [P(\mathbf{e})]^{-1} \tag{5.9}$$

Ahora bien, por la separación direccional sabemos que $P(\mathbf{e}_X^-|\, x, \mathbf{e}_X^+) = P(\mathbf{e}_X^-|\, x)$, de modo que, aplicando las definiciones anteriores llegamos a

$$P^*(x) = \alpha\, \pi(x)\, \lambda(x) \tag{5.10}$$

Necesitamos, por tanto, calcular los tres factores que aparecen en esta expresión. Empecemos con $\pi(x)$. Según su definición,

$$\pi(x) = P(x, \mathbf{e}_X^+) = \sum_{\bar{u}} P(x|\,\bar{u}) P(\bar{u}, \mathbf{e}_X^+)$$

Como las causas de X no tienen ningún antepasado común por estar en un poliárbol (red simplemente conexa), todas ellas y las ramas correspondientes son

independientes mientras no consideremos la evidencia relativa a X o a sus descendientes:

$$\begin{aligned} P(\bar{u}, \mathbf{e}_X^+) &= P(u_1, \mathbf{e}_{U_1X}^+, \ldots, u_n, \mathbf{e}_{U_nX}^+) \\ &= \prod_{i=1}^{n} P(u_i, \mathbf{e}_{U_iX}^+) = \prod_{i=1}^{n} \pi_X(u_i) \end{aligned} \tag{5.11}$$

Por tanto,

$$\pi(x) = \sum_{\bar{u}} P(x|\,\bar{u}) \prod_{i=1}^{n} \pi_X(u_i)\,. \tag{5.12}$$

El paso siguiente consiste en calcular $\pi_X(u_i)$ o, lo que es lo mismo, $\pi_{Y_j}(x)$, puesto que en una R.B. todos los nodos son equivalentes; es sólo una cuestión de notación. La evidencia que está por encima del enlace XY_j, $\mathbf{e}_{XY_j}^+$, podemos descomponerla en varios subconjuntos: la que está por encima de X y la que está por debajo de cada enlace XY_k para los demás efectos Y_k de X (fig. 5.1). Sabemos además que X separa $\mathbf{e}_X^+$ de $\mathbf{e}_{XY_k}^-$, y separa también los subconjuntos $\mathbf{e}_{XY_k}^-$ entre sí. Con estas consideraciones, obtenemos

$$\begin{aligned} \pi_{Y_j}(x) &= P(x, \mathbf{e}_{XY_j}^+) = P(x, \mathbf{e}_X^+, \mathbf{e}_{XY_k}^- \; k \neq j) \\ &= P(x, \mathbf{e}_X^+) \prod_{k \neq j} P(\mathbf{e}_{XY_k}^-|\,x) \\ &= \pi(x) \prod_{k \neq j} \lambda_{Y_k}(x) \end{aligned} \tag{5.13}$$

Para calcular esta expresión, es necesario hallar $\lambda_{Y_k}(x)$ —o $\lambda_{Y_j}(x)$, pues el resultado obtenido será válido para todos los efectos de X—. Indicaremos por $\bar{V}$ el conjunto de causas de Y_j (o del efecto considerado) distintas de X, tal como muestra la figura 5.2. Por simplificar la notación, escribiremos $\mathbf{e}_{\bar{V}Y_j}^+ = \mathbf{e}_{V_1Y}^+ \cup \ldots \cup \mathbf{e}_{V_pY}^+$, con lo que nos queda $\mathbf{e}_{XY_j}^- = \mathbf{e}_Y^- \cup \mathbf{e}_{\bar{V}Y}^+$.

Recordemos que Y_j separa $\mathbf{e}_{Y_j}^-$ del resto de la red que está por encima de Y_j, e igualmente los padres de Y_j separan Y_j de $\mathbf{e}_{\bar{V}Y_j}^+$. Aplicando repetidamente la proposición 4.3, resulta

$$\begin{aligned} \lambda_{Y_j}(x) &= P(\mathbf{e}_{XY_j}^-|\,x) \\ &= \sum_{y_j} \sum_{\bar{v}} P(\mathbf{e}_{Y_j}^-, y_j, \mathbf{e}_{\bar{V}Y_j}^+, \bar{v}|\,x) \\ &= \sum_{y_j} \sum_{\bar{v}} P(\mathbf{e}_{Y_j}^-|\,y_j)\, P(y_j|\,\bar{v}, x)\, P(\mathbf{e}_{\bar{V}Y_j}^+, \bar{v}|\,x) \end{aligned}$$

Puesto que las causas de Y_j son independientes a priori, podemos razonar como en la ecuación (5.11) para llegar a

$$P(\bar{v}, \mathbf{e}_{\bar{V}Y_j}^+|\,x) = P(\bar{v}, \mathbf{e}_{\bar{V}Y_j}^+) = \prod_{l=1}^{p} P(v_l, \mathbf{e}_{V_lY_j}^+) = \prod_{l=1}^{p} \pi_{Y_j}(v_l)$$

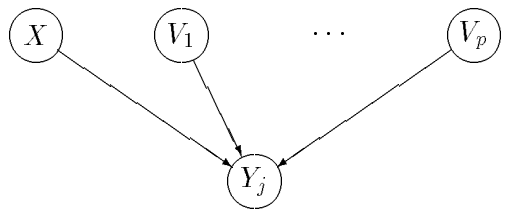


Figura 5.2: Padres de Y_j.

y, en consecuecia,

$$\lambda_{Y_j}(x) = \sum_{y_j} \left[\lambda(y_j) \sum_{\bar{v}} P(y_j | \, x, \bar{v}) \prod_{l=1}^{p} \pi_{Y_j}(v_l) \right] \tag{5.14}$$

Finalmente, hay que calcular $\lambda(x)$, lo cual resulta bastante sencillo:

$$\begin{aligned} \lambda(x) &= P(\mathbf{e}^-_{XY_1}, \ldots, \mathbf{e}^-_{XY_m} | \, x) \\ &= \prod_{j=1}^{m} P(\mathbf{e}^-_{XY_j} | \, x) = \prod_{j=1}^{m} \lambda_{Y_j}(x) \end{aligned} \tag{5.15}$$

Para completar el algoritmo, hay que hallar la constante α que aparece en (5.10). Realizar el cálculo a partir de la definición 5.9 resultaría muy complicado en general. Sin embargo, sabemos que

$$\sum_x P^*(x) = \alpha \sum_x \pi(x) \, \lambda(x) = 1 \tag{5.16}$$

con lo que podemos obtener α como

$$\alpha = \left[\sum_x \pi(x) \, \lambda(x) \right]^{-1} \tag{5.17}$$

En la práctica, calcularemos $\pi(x)$ y $\lambda(x)$ para cada nodo y normalizaremos su producto de acuerdo con la ecuación (5.16).

5.1.3 Comentarios

Las fórmulas que acabamos de deducir son recursivas: $\pi(x)$ se calcula a partir de $\pi_X(u_i)$; $\pi_{Y_j}(x)$ a partir de $\pi(x)$ y de $\lambda_{Y_k}(x)$, etc. Necesitamos por tanto una condición de terminación para que el algoritmo esté completo. Por otro lado, necesitamos explicar cómo introducir en este esquema la evidencia observada. Resolveremos ambos problemas del siguiente modo:

Para un nodo U sin padres, $\mathbf{e}^+_U = \emptyset$, por lo que $\pi u = P(u)$, que es uno de los parámetros que definen la red. En este caso el problema de terminación ya lo teníamos resuelto.

Para un nodo terminal Y (nodo sin hijos), hace falta conocer $\lambda(y)$. Si no hay ninguna información sobre este nodo, asignamos el mismo número para cada valor y; por ejemplo, $\lambda(y) = 1$ para todo y. Vemos en la ecuación (5.10) que un vector

$\lambda(x)$ constante no modifica el valor de $P^*(x)$. También vemos, a partir de la ecuación (5.14), que para un vector constante $\lambda(y) = c$ podemos alterar el orden de los sumatorios y llegar a

$$\begin{aligned}\lambda_{Y_j}(x) &= c \sum_{\bar{v}} \left[\prod_{l=1}^{p} \pi_{Y_j}(v_l) \sum P(y_j | x, \bar{v}) \right] \\ &= c \sum_{\bar{v}} \left[\prod_{l=1}^{p} \pi_{Y_j}(v_l) \right] = c \sum_{\bar{v}} P(\bar{v}, \mathbf{e}^+_{\bar{V} Y_j}) = c\, P(\mathbf{e}^+_{\bar{V} Y_j}), \quad \forall x \qquad (5.18)\end{aligned}$$

que es de nuevo un vector constante y no transmite ninguna información, pues según las ecuaciones (5.10) y (5.14), un vector λ constante no influye en el resultado final.

Si hay un nodo terminal Y de valor conocido y_0 (es decir, la afirmación "$Y = y_0$" es parte de la evidencia), asignamos a $\lambda(y_0)$ un número positivo cualquiera y 0 a los demás valores. Por ejemplo,

$$\begin{cases} \lambda(y_0) = 1 \\ \lambda(y) \ = 0 \quad \text{para } y \neq y_0 \end{cases}$$

lo cual implica, según (5.10),

$$\begin{cases} P(y_0) = 1 \\ P(y) \ = 0 \quad \text{para } y \neq y_0 \end{cases}$$

Vemos que, efectivamente, la probabilidad se ajusta a la afirmación de partida, "$Y = y_0$"; además sólo el valor y_0 cuenta en el sumatorio de la ecuación (5.14), por lo que podemos concluir que esta asignación de $\lambda(y)$ para nodos terminales es completamente coherente, y así queda completo el algoritmo de propagación de evidencia en poliárboles.

Ejemplo. Volvamos de nuevo a la red de la figura 4.6. Recordemos que, además de tener la estructura de la red, conocemos las probabilidades a priori de los nodos sin padres: $P(a)$, $P(b)$ y $P(c)$, y las probabilidades condicionales: $P(d|\, a, b)$, $P(e|\, c)$, etc.

Supongamos que $\mathbf{e} = \{+f, +g, \neg i\}$. La asignación de landas para los nodos terminales será:

$$\begin{cases} \lambda(+f) = 1 \\ \lambda(\neg f) = 0 \end{cases} \quad \begin{cases} \lambda(+g) = 1 \\ \lambda(\neg g) = 0 \end{cases} \quad \begin{cases} \lambda(+h) = 1 \\ \lambda(\neg h) = 1 \end{cases} \quad \begin{cases} \lambda(+i) = 0 \\ \lambda(\neg i) = 1 \end{cases}$$

Queremos calcular $P^*(e)$. Empezamos buscando $\pi(e)$:

$$\begin{aligned}\pi(e) &= \sum_c P(e|\, c)\, \pi_E(c) = \sum_c P(e|\, c)\, P(c) \\ \pi_E(c) &= \pi(c)\, \lambda_F(c) \\ \lambda_F(c) &= \sum_f \lambda(f)\, P(f|\, c) = P(+f|\, c)\end{aligned}$$

Con esto concluimos el cálculo en esta rama del árbol. Continuamos con otras ramas:

$$\begin{aligned}
\lambda(e) &= \lambda_I(e)\,\lambda_H(e) \\
\lambda_I(e) &= \sum_i \lambda(i)\,P(i|\,e) = P(\neg i|\,e) \\
\lambda_H(e) &= \sum_h \lambda(h) \sum_d P(h|\,d,e)\,\pi_H(d)
\end{aligned}$$

Deberíamos calcular ahora $\pi_H(d)$. Sin embargo, podemos saber ya que $\lambda_H(e)$ va a ser un vector constante porque $\lambda(h)$ también lo es. Podemos demostrarlo mediante el argumento numérico de la ecuación (5.18). Otra forma de razonarlo es a partir de las propiedades de independencia condicional: cuando el valor de H no se conoce, D y E son independientes; recordando además que D separa G de E, tenemos

$$\begin{aligned}
\lambda_H(e) &= P(\mathbf{e}^-_{EH}|\,e) = P(+g|\,e) \\
&= \sum_d P(+g|\,d,e)\,P(d|\,e) = \sum_d P(+g|\,d)\,P(d) = P(+g)
\end{aligned}$$

que es un vector constante (no depende de e).

Por fin nos queda $\lambda(e) = \lambda_I(e)$ y basta normalizar el producto $\pi(e)\ \lambda(e)$ para conocer $P^*(x)$. Del mismo modo podemos calcular la probabilidad a posteriori de cualquier otra variable, aprovechando —naturalmente— los resultados ya obtenidos. □

5.1.4 Implementación distribuida

El algoritmo que hemos presentado se presta inmediatamente a una implementación recursiva, según hemos comentado anteriormente. Vamos a ver ahora cómo podemos diseñar un algoritmo distribuido a partir de las expresiones que hemos deducido anteriormente. (Veremos también que este método puede llevarnos a una implementación iterativa, que presenta la ventaja de que requiere mucha menos memoria de cálculo que la implementación recursiva.)

En la implementación distribuida, cada procesador corresponderá a un nodo y, por tanto, a una variable. La información que debe almacenar puede ser estática o dinámica. Aquí, "estática" significa "independiente de la evidencia observada", tal como la estructura de la red y las probabilidades condicionales. En caso de que los nodos no sean procesadores físicos reales sino que estén simulados mediante un programa, lo primero que cada nodo necesita conocer son sus causas y sus efectos; esto puede realizarse fácilmente definiendo dos listas con punteros hacia los nodos correspondientes, las cuales codifican la topología de la red. A continuación hay que introducir la información numérica estática, a saber, las probabilidades a priori y condicionales.

La información dinámica consiste, en primer lugar, en los valores de λ para los nodos terminales, tal como hemos explicado anteriormente, y en los mensajes π y λ correspondientes a la propagación de evidencia. La propiedad más importante que se deriva de los axiomas de independencia descritos en el capítulo anterior es que, en poliárboles, cada enlace descompone la red en dos partes cuya única interacción se

transmite a través de dicho enlace, y los mensajes intercambiados están desacoplados en el sentido de que $\pi_Y(x)$ puede calcularse independientemente de $\lambda_Y(x)$ y viceversa. En la figura 5.3, que muestra las computación realizada en el nodo X, esta propiedad aparece como la ausencia de lazos cerrados en el flujo de información. Se puede comprobar también, observando las fórmulas de la sección 5.1.2, que toda la información requerida por un nodo para computar sus mensajes se encuentra almacenada localmente.

Un nodo X está en disposición de enviar un mensaje a su vecino W cuando y sólo cuando ha recibido ya los mensajes procedentes de todos sus demás vecinos. Un nodo X con n causas y m efectos que ha recibido q mensajes se encuentra en uno de tres estados posibles:

1. $q \leq n + m - 2$. Esto significa que X está esperando al menos dos mensajes, por lo que todavía no puede calcular ninguno de los que debe enviar.

2. $q = n + m - 1$. En este caso, X ha recibido un mensaje de cada vecino excepto de uno, que llamaremos W. Por eso X puede entonces calcular el mensaje que debe enviar a W, pero no los demás.

3. $q = n + m$. Cuando X ha recibido todos los mensajes que estaba esperando, puede calcular por fin los que le faltaban por enviar.

Al principio, $q=0$ para todos los nodos, pues aún no ha circulado ningún mensaje; por tanto, todos los nodos con un solo vecino ($n+m=1$) se encuentran en el estado 2; los demás se encuentran todavía en estado 1. Es posible demostrar que siempre hay algún nodo dispuesto a enviar un mensaje, por lo que el proceso no se interrumpe nunca hasta que el algoritmo se haya completado. En vez de realizar la demostración, que es sencilla pero engorrosa, volvamos una vez más a la figura 4.6.

Antes de que empiece la propagación, todos los nodos que tienen un solo vecino (A, B, F, G and I) se hallan en estado 2, y los demás en estado 1. Cuando aquéllos envían sus mensajes respectivos, C y D pasan al estado 2, y lo mismo ocurre en el paso siguiente con E y H. Cuando los mensajes $\pi_H(e)$ y $\lambda_H(e)$ llegan a su destino, estos dos últimos nodos pasan al estado 3, de modo que pueden enviar ya a sus vecinos los mensajes que faltaban, y en dos pasos más queda concluido el proceso.

La discusión anterior es interesante para demostrar que no es necesario tener un mecanismo global de control, por lo que el modelo puede implementarse como una *red asíncrona* en que el número de mensajes recibido determina qué mensajes puede calcular y enviar cada nodo.

Si el algoritmo se implementa secuencialmente y la computación necesaria en cada nodo está acotada (limitando el número de padres y valores), el tiempo de computación es proporcional al número de nodos. En este caso resulta más eficiente realizar la propagación de evidencia en dos fases: recolección de mensajes hacia el nodo pivote y distribución desde él, como sugieren Lauritzen y Spiegelhalter para árboles de cliques.

En cambio, si hay un procesador por cada nodo, el tiempo de computación es proporcional a la longitud máxima que exista dentro de la red. La versión que hemos presentado aquí, basada en tres estados diferentes para cada nodo, se diferencia ligeramente de la de Pearl [174] en que evita computar y enviar mensajes prematuros

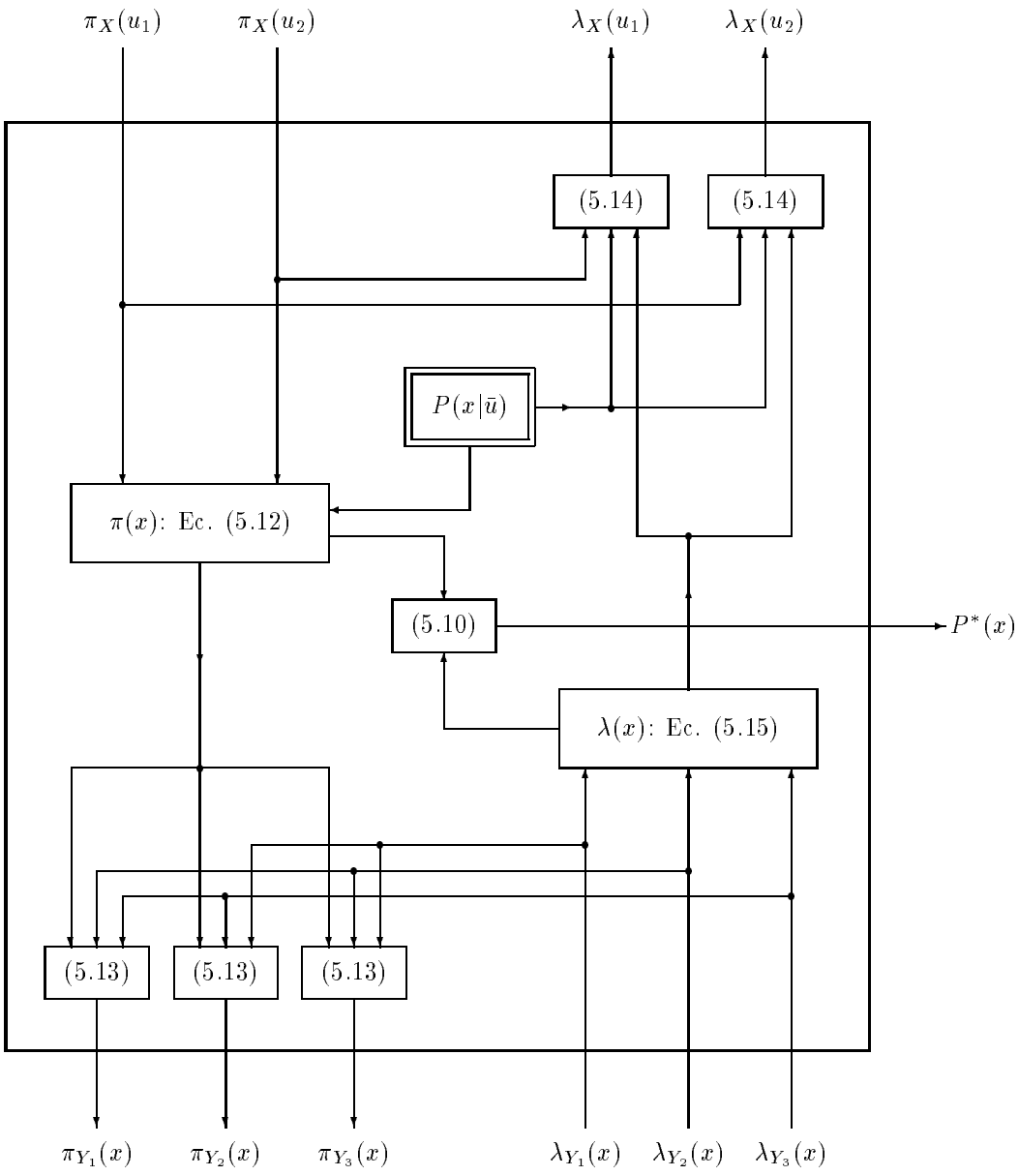


Figura 5.3: Computaciones realizadas en el nodo X.

carentes de sentido. La distinción no tiene importancia si disponemos de un procesador físico (*hardware*) por cada nodo. Pero si los procesadores conceptuales (los nodos) están simulados por un número menor de procesadores reales, el despilfarro computacional de enviar mensajes inútiles puede resultar demasiado caro en términos de eficiencia. En este último caso, en que los nodos hacen cola para acceder a un número limitado de procesadores físicos, encontramos el problema típico de la programación distribuida, a saber, cuál de los mensajes debe computarse primero con el fin de lograr la máxima eficiencia.

5.2 La puerta OR

5.2.1 La puerta OR binaria

Hemos visto que, en el caso general, la probabilidad condicional viene dada por una tabla, tal como la que aparece en la página 63. Por tanto, el número de parámetros requerido para una familia crece exponencialmente con el número de padres. Esto conlleva varios inconvenientes. El más grave es la obtención de dichos parámetros: si obtenemos los resultados a partir de una base de datos, necesitamos gran cantidad de casos para que los parámetros obtenidos sean fiables; si la ausencia de una base de datos nos obliga a recurrir a la estimación subjetiva de un experto humano, resultará muy complicado para él responder a tantísmas preguntas correspondientes a una casuística compleja: "¿Cuál es la probabilidad de que el paciente presente fiebre dado que tenga paludismo, neumonía y apendicitis y no tenga amigdalitis, ni meningitis, ni etc., etc.?"

El segundo problema que plantea el modelo general, una vez obtenidos los parámetros, es la cantidad de espacio de almacenamiento que requiere. Y por último, otro grave inconveniente es que el tiempo de computación para la propagación de evidencia crece también exponencialmente con el número de padres de la familia considerada.

Por estas razones, es conveniente buscar modelos simplificados de interacción causal que simplifiquen la construcción de RR.BB. y la computación de la probabilidad. Pearl [174] los llama modelos *canónicos* porque son aplicables a numerosos campos, no son soluciones *ad hoc* para resolver un problema concreto de un dominio particular. Los más famosos entre ellos son las puertas AND y OR probabilísticas, que suponen una generalización de los correspondientes modelos deterministas. En esta sección vamos a hablar sobre todo de la puerta OR, pues es la que con más frecuencia aparece en la práctica; además, las definiciones y resultados presentados aquí son fácilmente generalizables a la puerta AND.

En la puerta OR binaria (*noisy OR-gate* [174, sec. 4.3.2]) se supone que cada causa U_i actúa para producir el efecto X, pero existe un *inhibidor* I_i que bloquea la influencia; es como si U_i estuviera inactiva. Por tanto, el parámetro fundamental es la probabilidad de que actúe el inhibidor (q_i) o bien su complementario, $c_i = 1 - q_i$, la probabilidad de que la causa U_i actuando en ausencia de otras causas llegue a producir X:

$$P(+x|+u_i, \neg u_j \; j \neq i) = c_i = 1 - q_i$$

Tenemos así la probabilidad de X en el caso de que haya una única causa presente y las demás están ausentes. Para hallar la probabilidad de X en el caso de que

haya más de una causa presente, se introduce la hipótesis de que X sólo está ausente cuando todas las causas están ausentes o cuando para cada causa U_i que está presente ha actuado el correspondiente inhibidor I_i. Se supone que no sólo las causas sino también los inhibidores actuán independientemente, lo cual implica la independencia en sentido probabilístico. En consecuencia,

$$P(-x|\bar{u}) = \prod_{i \in T_U} q_i$$

donde T_U indica el subconjunto de las causas de X que están presentes ($T_U \subset \bar{U}$).

A partir de aquí podemos construir la tabla $P(x|\bar{u})$ y aplicar el algoritmo de propagación general desarrollado en la sección anterior. Pero así habríamos resuelto uno sólo de los inconvenientes anteriores (el de la obtención de los parámetros), pues ya vimos que la complejidad de este algoritmo crecía exponencialmente con el número de padres. Por fortuna, existen expresiones para la puerta OR que llevan a un tiempo de propagación proporcional al tamaño de la familia. Dichas expresiones se encuentran en [174]. Nosotros, en vez de deducirlas aquí, las presentaremos como un caso particular de las correspondientes a la puerta OR graduada que vamos a estudiar a continuación.

5.2.2 La puerta OR graduada

Existe una generalización de la puerta OR binaria, que fue propuesta por Max Henrion [110] como modelo para la obtención del conocimiento. Sin embargo, el nombre de "puerta OR graduada", su formulación matemática y los algoritmos de propagación son originales del presente trabajo; han sido publicados previamente de forma resumida en [72].

Para llegar a una formulación matemática del modelo es necesario introducir previamente el siguiente concepto:

Definición 5.1 (Variable graduada) *Es la variable X que puede estar ausente o presente con g_X grados de intensidad. Tiene por tanto g_X+1 valores posibles, a los que asignaremos enteros tales que $X = 0$ significa "ausencia de X" y los números sucesivos indican grados de mayor intensidad.*

Ejemplo. Supongamos que la variable X =Neumonía puede estar ausente o presente con tres grados de intensidad ($g_X = 3$): leve, moderada o severa. Entonces $X = 0$ significa "el paciente no tiene neumonía", $X = 1$ significa "el paciente tiene neumonía leve", etc. □

Observar que el concepto de *graduada* no es sinónimo de *multivaluada*. De hecho, son dos conceptos independientes: por un lado, no todas las variables multivaluadas representan distintos grados de intensidad, y por otro lado, hay variables binarias graduadas, como son las que hemos visto hasta ahora de tipo presente/ausente o positivo/negativo, cuyos valores representábamos por $+x$ y $\neg x$. (La definición de variable graduada nos dice que a $\neg x$ le corresponde el valor 0 y a $+x$ el valor 1.) Más aún, *las variables que intervienen en la puerta OR binaria son siempre variables graduadas*, pues no tiene sentido plantear dicho modelo para variables no graduadas, tales como el sexo.

El modelo de interacción de la puerta OR es bastante general; aquí lo vamos a definir en el contexto de las RR.BB., aunque sería aplicable a otros métodos de tratamiento de la incertidumbre (ver la sec. 2.3). Por simplificar la escritura, llamaremos $\tilde{U}_i$ al conjunto de todas las causas de X excluida U_i:

$$\tilde{U}_i \equiv \bar{U} - U_i$$

Definición 5.2 (Puerta OR graduada) *En una red bayesiana, dada una variable graduada X con n padres $U_1, \ldots, U_n$ (también variables graduadas), decimos que interaccionan mediante una* puerta OR graduada *cuando se cumplen las dos condiciones siguientes:*

$$1.\ P(X = 0|\bar{U} = 0) = 1 \tag{5.19}$$

$$2.\ P(X \leq x|\bar{u}) = \prod_i P(X \leq x|U_i = u_i, \tilde{U}_i = 0) \tag{5.20}$$

Podemos utilizar la notación $x^0 \equiv$"$X = 0$" para expresar ambas condiciones en forma abreviada como

$$1.\ P(x^0|\bar{u}^0) = 1 \tag{5.21}$$

$$2.\ P(X \leq x|\bar{u}) = \prod_i P(X \leq x|u_i, \tilde{u}_i^0) \tag{5.22}$$

Intentaremos ahora explicar el significado de esta definición. La primera condición es fácil de interpretar: significa que, si todas las causas que pueden producir X están ausentes, entonces tenemos la seguridad de que también X estará ausente. Más adelante relajaremos esta restricción.

La segunda condición (ec. (5.20)) nos dice que $X \leq x$ sólo cuando ninguna de las causas U_i (actuando como si las demás causas estuvieran ausentes) ha elevado X a un grado superior a x. Dicho con otras palabras, el grado que alcanza X es el *máximo* de los grados producidos por las causas actuando independientemente. Al igual que en la puerta OR binaria, el resultado es el máximo de los valores de las entradas; esta coincidencia era de esperar, pues el modelo graduado es tan sólo una generalización del caso binario.

La importancia de esta definición es que permite calcular todos los valores de $P(x|\bar{u})$ a partir de un reducido número de parámetros. Para la familia X, serán las probabilidades X condicionadas a que una sola de las causas esté presente:

$$c_{X=x}^{U_i=u_i} \equiv P(X = x|\, U_i = u_i, \tilde{U}_i = 0) \tag{5.23}$$

que podemos escribir en forma abreviada como

$$c_x^{u_i} \equiv P(x|u_i, \tilde{u}_i^0) \tag{5.24}$$

En principio, el número de parámetros para el enlace U_iX es $(g_{U_i} + 1) \cdot (g_X + 1)$. Sin embargo, la suma de las probabilidades debe ser la unidad y, en consecuencia,

$$c_{x^0}^{u_i} = 1 - \sum_{x=1}^{g_X} c_x^{u_i}. \tag{5.25}$$

Por otra parte, la primera condición de la definición de la puerta OR (ec. (5.20)) es equivalente a decir que

$$c_x^{u_i^0} = \begin{cases} 1 & \text{para } x = 0 \\ 0 & \text{para } x \neq 0\,. \end{cases} \tag{5.26}$$

Por tanto, sólo se necesitan $g_{u_i} \cdot g_X$ parámetros para este enlace.

Ejemplo. Supongamos que tenemos una porción de red representada por la figura 5.4, y que cada una de las tres variables puede tomar los siguientes valores:

$$\begin{aligned} U_1 &= \text{Neumonía} \rightarrow \{\text{ausente, leve, moderada, severa}\} \\ U_2 &= \text{Paludismo} \rightarrow \{\text{ausente, presente}\} \\ X &= \text{Fiebre} \rightarrow \{\text{ausente, leve, elevada}\} \end{aligned}$$

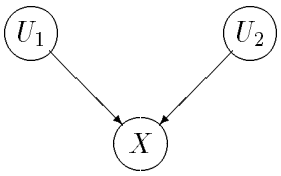


Figura 5.4: Ejemplo de puerta OR.

Vemos que $g_{U_1} = 3$, $g_{U_2} = 1$ y $g_X = 2$. En el modelo general, para esta familia necesitaríamos una tabla con 16 parámetros (hay 4×2=8 combianaciones posibles de u_1 y u_2; para cada una de ellas deberíamos dar tres valores $P(x|u_1, u_2)$, pero como la suma de los tres debe ser la unidad, sólo hace falta dar dos, de modo que necesitamos en total 8×2=16 parámetros). Sin embargo, para la puerta OR graduada, indicaremos los valores de $c_x^{u_1}$ y $c_x^{u_2}$, tal como muestran las tablas 5.1y 5.2.

$X \setminus U_1$	leve	moderada	severa
leve	0'50	0'40	0'20
elevada	0'20	0'50	0'80

Tabla 5.1: Parámetros $c_x^{u_1}$.

$X \setminus U_2$	presente
leve	0'20
elevada	0'75

Tabla 5.2: Parámetros $c_x^{u_2}$.

En ellas se observa que el número de parámetros se ha reducido a la mitad. Si hubiéramos tenido más causas en vez de sólo dos, el ahorro habría sido mayor. A partir

de estas dos pequeñas tablas, aplicando los axiomas anteriores podemos construir la tabla $P(x|u_1, u_2)$ completa necesaria para aplicar el algoritmo general. Sin embargo, existe una solución mucho más eficiente que evita tener que calcular dicha tabla, como mostraremos en la próxima sección. □

Causas no explícitas

Si en el modelo general tomamos $P(+x|\neg u_1, \neg u_2) > 0$, esto significa que X puede estar presente incluso cuando U_1 y U_2 están ausentes. Sin embargo, en la puerta OR, la propiedad (5.20) nos dice que cuando todas las causas de X están ausentes, sabemos con certeza que X estará ausente.

La propiedad en sí es razonable, pero existe el problema de que en la práctica es imposible considerar explícitamente todas las causas, pues éstas pueden ser muy numerosas e incluso muchas de ellas serán desconocidas; esto se ve especialmente claro en el caso de la medicina. La cuestión es importante aunque, afortunadamente, tiene una solución muy sencilla: para cada nodo X incluiremos un nodo X^* que agrupe todas las causas que no aparezcan explícitamente en el modelo.

Podemos suponer que el valor de este nodo siempre es **presente** ($\pi(+x^*) = 1$) y que su eficiencia para producir X viene dada por los parámetros $c_x^{x^*}$ (de forma abreviada, c_x^*); en caso de que X sea una variable binaria, basta conocer un solo número, c_{+x}^*, pues $c_{\neg x}^*$ será siempre 0.

Al desarrollar el algoritmo de propagación para la puerta OR, veremos que el impacto de cada causa U_i se traduce en una $Q_{U_i}(x)$, y todas éstas se combinan de acuerdo con la ecuación (5.33). Por tanto, podemos aplicar, la propiedad asociativa del producto y agrupar varias causas en una sola sin violar los principios axiomáticos de las redes bayesianas. Lo que intento decir es que está matemáticamente justificado incluir las causas no explícitas en un solo nodo y asignar al enlace correspondiente unos valores c_x^* que combinan los parámetros de todas ellas como si se tratara de una sola causa.

La puerta AND

Del mismo modo que hemos definido la puerta OR, podemos introducir la puerta AND graduada como generalización de la puerta AND binaria probabilística, la cual es a su vez era una generalización de la puerta AND determinista.

Definición 5.3 (Puerta AND graduada) *En una red bayesiana, dada una variable graduada X con n padres $U_1, \ldots, U_n$ (también variables graduadas), decimos que interaccionan mediante una* puerta AND graduada *cuando se cumplen las dos condiciones siguientes:*

1. $P(X = 0|\bar{U} = 0) = 1$ (5.27)
2. $P(X \geq x|\bar{u}) = \prod_i P(X \geq x|U_i = u_i, \tilde{U}_i = 0)$ (5.28)

La interpretación es similar a la puerta OR: la primera condición es la misma que en aquélla, y significa, igualmente, que cuando todas las causas de X están ausentes, X también está ausente. La segunda implica que $X \geq x$ sólo cuando cada causa (actuando como si las demás estuvieran ausentes) ha elevado X al menos a un grado

x. Por tanto, la puerta AND corresponde al mínimo, del mismo modo que la puerta OR correspondía al máximo.

5.2.3 Algoritmo de propagación

Hemos resuelto ya los dos primeros problemas que presentaba el modelo general, pues ya no necesitamos obtener ni almacenar un número exponencial de parámetros por familia. Veamos a continuación cómo podemos resolver el tercero, es decir, cómo podemos realizar eficientemente la propagación de evidencia. Empezamos introduciendo la siguiente definición:

$$Q(x) \equiv P(X \leq x, \mathbf{e}_X^+) \tag{5.29}$$

Es fácil obtener $Q(x)$ a partir de $\pi(x)$

$$Q(x) = \sum_{x'=0}^{x} \pi(x') \tag{5.30}$$

y viceversa

$$\pi(x) = \begin{cases} Q(x) - Q(x-1) & \text{para } x \neq 0 \\ Q(0) & \text{para } x = 0 \end{cases} \tag{5.31}$$

Queremos encontrar ahora un algoritmo eficiente para calcular $Q(x)$. Aplicando las ecuaciones (5.22) y (5.11), podemos escribir

$$\begin{aligned} Q(x) &= \sum_{\bar{u}} P(X \leq x|\bar{u})\, P(\bar{u}, \mathbf{e}_X^+) \\ &= \sum_{\bar{u}} \left[\prod_i P(X \leq x|u_i, \tilde{u}_i^0)\, P(u_i|\mathbf{e}_{U_iX}^+) \right] \end{aligned}$$

En esta expresión podemos invertir el orden del productorio y de los sumatorios. En efecto,

$$\begin{aligned} &\sum_{u_1} \left[\prod_{i=1}^{n} P(X \leq x|u_i, \tilde{u}_i^0)\, P(u_i, \mathbf{e}_{U_iX}^+) \right] \\ &= \left[\sum_{u_1} P(X \leq x|u_1, \tilde{u}_1^0)\, P(u_1, \mathbf{e}_{U_1X}^+) \right] \cdot \prod_{i=2}^{n} P(X \leq x|u_i, \tilde{u}_i^0)\, P(u_i, \mathbf{e}_{U_iX}^+) \\ &= P(X \leq x,\, \mathbf{e}_{U_1X}^+,\, \tilde{u}_1^0) \prod_{i=2}^{n} P(X \leq x|u_i, \tilde{u}_i^0)\, P(u_i|\mathbf{e}_{U_iX}^+) \\ &= Q_{U_1} \cdot \prod_{i=2}^{n} P(X \leq x|u_i, \tilde{u}_i^0)\, P(u_i|\mathbf{e}_{U_iX}^+) \end{aligned}$$

donde hemos introducido la definición

$$Q_{U_i}(x) \equiv P(X \leq x, \mathbf{e}_{U_iX}^+, \tilde{u}_i^0) \tag{5.32}$$

que es la probabilidad de X considerando toda la evidencia por encima del enlace U_iX, suponiendo que todas las demás causas de X estuvieran ausentes.

Sustituyendo este resultado en la expresión de $Q(x)$ tenemos

$$Q(x) = Q_{U_1}(x) \cdot \sum_{u_2,\ldots,u_n} \left[\prod_{i=2}^{n} P(X \leq x | u_i, \tilde{u}_i^0)\, P(u_i | \mathbf{e}^+_{U_i X}) \right]$$

y repitiendo la misma operación n veces llegamos a

$$Q(x) = \prod_i Q_{U_i}(x) \tag{5.33}$$

Lo que necesitamos ahora es una fórmula sencilla para calcular $Q_{U_i}(x)$. Para ello, definimos un nuevo conjunto de parámetros $C_x^{u_i}$:

$$C_x^{u_i} \equiv P(X \leq x | u_i, \tilde{u}_i^0) \tag{5.34}$$

que podemos calcular a partir de las $c_x^{u_i}$, según las ecuaciones (5.24) y (5.25):

$$\begin{aligned} C_x^{u_i} &= \sum_{x'=0}^{x} c_x^{u_i} = c_{x^0}^{u_i} + \sum_{x'=1}^{x} c_x^{u_i} = 1 - \sum_{x'=1}^{g_X} c_x^{u_i} + \sum_{x'=1}^{x} c_x^{u_i} \\ &= 1 - \sum_{x'=x+1}^{g_x} c_x^{u_i} \end{aligned} \tag{5.35}$$

Estos nuevos parámetros pueden ser almacenados junto con la descripción de la red (para ahorrar tiempo de computación) o calculados cuando se los necesita, aunque también es posible definir la red a partir de las $C_x^{u_i}$ en lugar de las $c_x^{u_i}$.

Desde aquí, el cálculo de $Q_{U_i}(x)$ es inmediato:

$$\begin{aligned} Q_{U_i}(x) &= \sum_{u_i} P(X \leq x | u_i, \tilde{u}_1^0)\, P(u_i, \mathbf{e}^+_{U_i X}) \\ &= \sum_{u_i} C_x^{u_i}\, \pi_X(u_i) \qquad (5.36) \\ &= \sum_{u_i} \pi_X(u_i) \left[1 - \sum_{x'=x+1}^{g_X} c_x^{u_i} \right] \qquad (5.37) \end{aligned}$$

En el caso de que tengamos además unas c_x^* correspondientes a las causas no explícitas en el modelo, podemos manejarlas como si se tratara de una causa similar a las demás y calcular la respectiva $Q^*(X)$, que deberá incluirse en el productorio de la ecuación (5.33). El tratamiento de las causas no explícitas es, por tanto, muy sencillo.

Hemos resuelto ya la primera parte del problema: cómo calcular $\pi(x)$ para la familia X en tiempo proporcional al número de padres. También el cálculo de $\lambda(x)$ y el de $\pi_X(u_i)$ o $\pi_{Y_j}(x)$ están resueltos, pues podemos aplicar las ecuaciones (5.15) y (5.13) ya que en ellas no aparece $P(x|\bar{u})$ y por tanto no varían al pasar del caso general a la puerta OR. Lo que nos falta por resolver es cómo calcular $\lambda_Y(u_i)$, es decir, el mensaje que X envía a cada uno de sus padres.

La ecuación (5.14) puede escribirse para la familia X como

$$\lambda_X(u_i) = \sum_x \left[\lambda(x) \sum_{\tilde{u}_i} P(x|\bar{u}) \prod_{j \neq i} \pi_X(u_j) \right] \tag{5.38}$$

Observar que, dentro de esta expresión, el valor de u_i en $\bar{u} = (u_1, \ldots, u_n)$ está fijo (depende de qué $\lambda_X(u_i)$ estamos calculando), mientras que el valor de las demás variables u_j va cambiando según indica el sumatorio.

Ahora bien, una forma de fijar el valor u_i' para la variable U_i es asignarle un vector $\pi(u_i)$ definido así:

$$[\pi(u_i')]_{U_i=u_i} = \begin{cases} 1 & \text{para } u_i = u_i' \\ 0 & \text{para } u_i \neq u_i' \end{cases} \tag{5.39}$$

puesto que entonces, según (5.10) y (5.13),

$$[P(u_i)]_{U_i=u_i'} = [\pi_X(u_i)]_{U_i=u_i'} = \begin{cases} 1 & \text{para } u_i = u_i' \\ 0 & \text{para } u_i \neq u_i' \end{cases}$$

y también

$$[P(x|\bar{u})]_{U_i=u_i'} = \sum_{u_i} P(x|\bar{u})\,[\pi(u_i)]_{U_i=u_i'}$$

Sustituyendo este resultado en la ecuación (5.38), tenemos

$$\lambda_X(u_i') = \sum_x \left[\lambda(x) \sum_{\bar{u}} P(x|\bar{u}) \prod_j \pi_X(u_j) \right]_{U_i=u_i'}$$

o bien

$$\lambda_X(u_i) = \sum_x \lambda(x)\,[\pi(x)]_{U_i=u_i} \tag{5.40}$$

Aquí, $[\pi(x)]_{U_i=u_i}$ debe calcularse como hicimos anteriormente, es decir, con las ecuaciones (5.31) y (5.33), aunque ahora la ecuación (5.36) se simplifica para convertirse en

$$[Q_{U_i}(x)]_{U_i=u_i} = C_x^{u_i} \tag{5.41}$$

de acuerdo con el valor de $[\pi_X(u_i)]_{U_i=u_i'}$ indicado anteriormente.

Dicho con otras palabras, el algoritmo de la puerta OR puede expresarse así: para calcular $\pi(x)$, transformamos cada mensaje $\pi_X(u_i)$ en $Q_{U_i}(x)$, y los multiplicamos todos para obtener $Q(x)$, a partir del cual es muy sencillo obtener $\pi(x)$.

Cuando queremos calcular $\lambda_X(u_i)$ seguimos un procedimiento similar: para las causas U_j distintas de U_i tomamos las mismas $Q_{U_j}(x)$ que antes; para U_i, en cambio, tomamos la $Q_{U_i}(x)$ correspondinte al valor u_i según la ecuación (5.41), y repetimos —para cada valor de U_i— el mismo proceso que en el cálculo de $\pi(x)$.

Algoritmo para la puerta AND

Como los pasos necesarios para desarrollar este algoritmo son esencialmente los mismos que para la puerta OR, presentamos solamente los resultados. Las definiciones son

$$Q^{\dagger}(x) \equiv P(X \geq x, \mathbf{e}_X^+) \tag{5.42}$$

$$Q_{U_i}^{\dagger}(x) \equiv P(X \geq x, \mathbf{e}_{U_iX}^+, \tilde{u}_i^0) \tag{5.43}$$

y el algoritmo consiste en aplicar las siguientes expresiones:

$$Q^{\dagger}(x) = \prod_i Q^{\dagger}_{U_i}(x) \tag{5.44}$$

$$Q^{\dagger}_{U_i}(x) = \sum_{u_i} \pi_X(u_i) \left[1 - \sum_{0}^{x-1} c_x^{u_i} \right] \tag{5.45}$$

$$\pi(x) = \begin{cases} Q^{\dagger}(x) - Q^{\dagger}(x+1) & \text{para } x \neq g_X \\ Q^{\dagger}(x) & \text{para } x = g_X \end{cases} \tag{5.46}$$

Las demás ecuaciones son las mismas que en los casos anteriores.

5.2.4 Implementación distribuida

La implementación distribuida de las puerta OR y AND es muy similar a la del caso general. Sin embargo, ahora los parámetros c_x^u (las tablas 5.1 y 5.2, por ejemplo) no son una característica del nodo X ni de esta familia en conjunto, sino que están asociados a cada enlace UX. Observar en la figura 5.5 y en la tabla 5.3 dónde se almacenan las c_x^u. El nodo X debe saber solamente qué tipo de interacción debe aplicar (caso general, puerta OR o puerta AND); en caso de que se trate de una de las dos puertas, los parámetros se encontrarán almacenados en los enlaces correspondientes (salvo los parámetros c_x^*, correspondientes a las causas no explícitas, que se almacenarán en el propio nodo para no tener que añadir un nodo que representara *OTRAS-CAUSAS-DE-X*.) Comparando la figura 5.5 con la relativa al caso general (fig. 5.3), observamos que ahora los enlaces no son canales de información pasivos sino procesadores activos que transforman cada mensaje $\pi_X(u)$ en $Q_U(x)$ y generan además $\lambda_X(u)$, liberando así al nodo X de algunas computaciones.

		Modelo general	Puerta OR
Nodo X	Almacena	$P(x\|u)$	c_x^*
	Recibe	$\pi_X(u_i)$, $\lambda_{Y_j}(x)$	$Q_{U_i}(X)$, $\lambda_{Y_j}(x)$
	Envía	$\lambda_X(u_i)$, $\pi_{Y_j}(x)$, $P^*(x)$	$\lambda(x)$, $\pi_{Y_j}(x)$, $P^*(x)$
Enlace U_iX	Almacena		$c_x^{u_i}$
	Recibe		$\pi_X(u_i)$, $\lambda(x)$, $Q_{U_{i'}}(x)$
	Envía		$Q_{U_i}(x)$, $\lambda_X(u_i)$

Tabla 5.3: Caso general y puerta OR

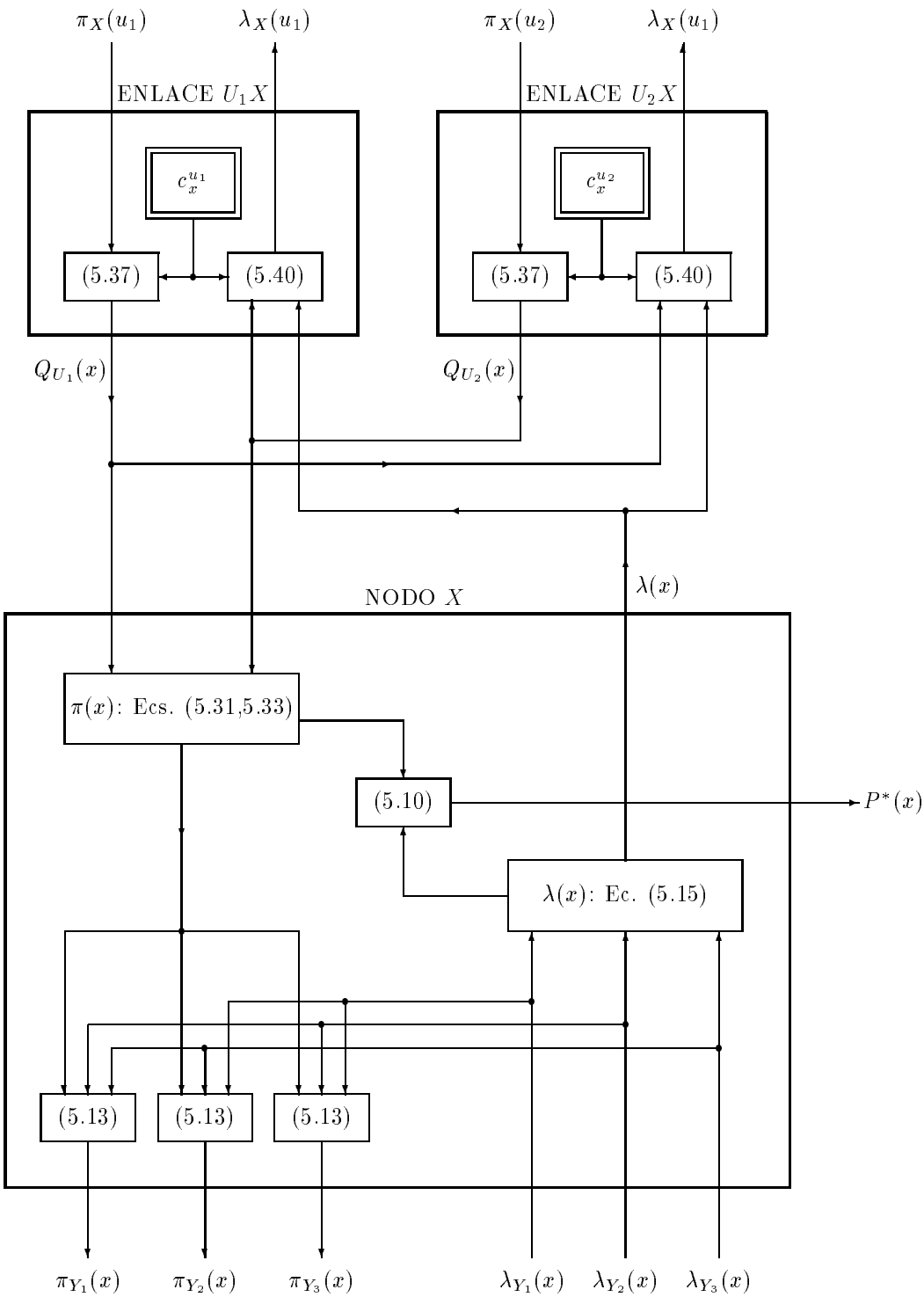


Figura 5.5: Computaciones realizadas en la puerta OR.

5.2.5 Semántica

Sería interesante desarrollar el ejemplo de la página 91 para ver cómo funciona la propagación de evidencia en la puerta OR y comprobar que los resultados obtenidos matemáticamente coinciden con lo previsible mediante nuestro sentido común. Lamentablemente, desarrollar con detalle un solo ejemplo con los múltiples casos posibles nos ocuparía mucho más espacio de lo que podemos permitirnos. Por otra parte, puede encontrarse un excelente ejemplo en [174, sec. 4.3.2]; aunque allí se describe solamente la puerta binaria, el tratamiento resulta muy similar al que pudiéramos realizar para el caso graduado.

Lo que vamos a discutir en esta sección es la relación entre el modelo general y la puerta OR. En la sección 4.3 hablamos de la semántica de las redes bayesianas, refiriéndonos al modelo general. Allí vimos la relación entre los axiomas de independencia y los mecanismos causales percibidos intuitivamente. De igual modo, estudiar la semántica de la puerta OR consiste en establecer una relación entre los axiomas de la definición 5.2 y nuestros conceptos de causalidad. Esta cuestión fue abordada parcialmente al introducir dicha definición. En efecto, allí se mostró que la ecuación (5.20) significa que el efecto X está ausente cuando todas las causas que lo producen están ausentes, lo cual concuerda naturalmente con el sentido común, y la ecuación (5.20) significa que el grado que alcanza el efecto X es el máximo de los que producirían sus causas actuando independientemente.

Al igual que discutimos al hablar de la semántica de las redes bayesianas en general, podemos afirmar aquí que hay dos formas posibles de justificar la utilización de la puerta OR como modelo simplificado al construir nuestra red. La primera —la teórica— consiste en crear en nuestra mente un modelo de cómo actúan las causas a la hora de producir el efecto considerado. Por ejemplo, si suponemos que las diferentes causas de una enfermedad actúan independientemente, en el sentido de que, tal como dice la definición de puerta OR, el grado más probable de la enfermedad es el máximo de los que producirían las causas, entonces estamos en condiciones de aplicar nuestro modelo simplificado; si no es así, debemos recurrir al modelo general.

La segunda forma de justificar nuestro modelo consiste en realizar estudios empíricos sobre un amplio número de casos y ver hasta qué punto la puerta OR puede considerarse como aproximación satisfactoria, y en esto pueden utilizarse dos criterios. Uno de ellos, el más estricto, consitiría en exigir que los resultados estadísticos para la familia X se ajustaran a los predichos por la expresión 5.20; el otro criterio, más flexible, se conformaría con que la red en su conjunto ofreciera diagnósticos acertados, dentro de ciertos límites, aunque las predicciones para la familia X no fueran completamente correctas.[1]

Por último, al hablar de la semántica debemos insistir en la diferencia que existe entre el modelo general y la puerta OR. Si volvemos al ejemplo 3 de la sección 4.1 (pág. 62), recordamos que había una interacción entre el país de origen y el tipo de sangre como *factores condicionantes* de la probabilidad de contraer paludismo. Sin embargo, en el ejemplo de la sección 5.2.2, teníamos dos causas, neumonía y

[1] Estos dos criterios pueden aplicarse también a sistemas expertos no bayesianos. Por ejemplo, en la evaluación de MYCIN [266] se tomó el segundo de ellos y se consideró que la realización del programa había sido un éxito. Sin embargo, con el criterio más estricto, se habría cuestionado la validez del programa, pues éste contiene numerosas inconsistencias, como explicamos en la sección 2.4.2.

paludismo, cada una de las cuales por sí misma era capaz de producir fiebre, interactuando mediante una puerta OR graduada. Por tanto, podemos afirmar que la puerta OR refleja mejor el concepto intuitivo de causalidad que utilizamos en nuestra vida cotidiana. Así, cuando decimos que "A es una **causa** de C" entendemos que "A produce o puede producir C". Nótese el contraste con el primer ejemplo, referido al caso general: los nodos País-de-origen y Tipo-sanguíneo son los padres del nodo Paludismo, pero nadie diría "el país de origen produce paludismo" y menos aún "el tipo de sangre produce paludismo", sino "el país de origen y el tipo de sangre son dos **factores condicionantes** que influyen en la probabilidad de contraer paludismo".

De esta diferencia se deduce una ventaja más de la puerta OR frente al modelo general, además de las que habíamos mencionado anteriormente: a la hora de generar *explicaciones lingüísticas*, si los nodos de la familia X interactúan mediante una puerta OR podemos decir "la causa que [con mayor probabilidad] ha producido X es U_i", "la presencia de U_i explica por qué se ha producido X, y por tanto ya no es necesario sospechar la presencia de U_j" o "al descartar U_i por dichas razones, aumenta nuestra sospecha de que la causa más probable de X es U_j". En el modelo general, no es posible —al menos no es fácil— generar este tipo de explicaciones a partir de una tabla de probabilidades.

Semántica de la puerta AND

Al contrario que la puerta OR, que es aplicable con frecuencia en numerosos dominios por la estrecha relación que guarda con el sentido intuitivo de causalidad, las ocasiones en que se puede aplicar la puerta AND son muy reducidas, al menos en el campo de la medicina. Dentro de las familias que interactúan mediante este modelo, el nodo hijo corresponde siempre a una variable binaria, mientras que los padres pueden ser variables multivaluadas.

En el caso de un nodo hijo X binario, podemos interpretar la condición (5.28) imaginando que cada nodo padre (cada causa) debe realizar una especie de aportación a X. Si $Q^{\dagger}_{U_i}(+x) = 1$, eso significa que tenemos la certeza de que la causa U_i ha satisfecho su "aportación"; un valor entre 0 y 1 indica la probabilidad de que haya sido así. Cuando todas las causas han cumplido su "aportación", X estará presente ($X = +x$); si falla una sola de ellas, X estará ausente.

De nuevo tenemos dos posibilidades de justificar la aplicación de la puerta AND. Si el experto humano a quien estamos consultando nos dice que para que haya un cierto efecto debe darse una serie de condiciones y que si falla una sola de ellas el efecto no se da, aplicaremos el modelo. Ésta es la justificación teórica. La justificación empírica consiste, naturalmente, en aplicar las pruebas estadísticas habituales, como hemos comentado para otros modelos.

La interpretación de la puerta AND con un nodo hijo multivaluado no está clara; sólo en la medida en que corresponde a la función mínimo puede preverse alguna utilidad para ella. Los motivos principales para estudiarla aquí han sido su semejanza con la puerta OR y el hecho de que la puerta AND binaria puede tratarse como caso particular del modelo que hemos presentado.

5.2.6 Comentarios

Concluimos la sección y el capítulo con un par de comentarios breves. El primero de ellos es que las puertas OR y AND no son aplicables solamente a redes bayesianas que incluyan todas las probabilidades condicionales en forma numérica, sino que estos modelos suponen, por encima de su formalismo matemático, un mecanismo conceptual de interacción. Las redes cualitativas, de las que hablaremos en la sección 10.1.3, utilizan las propiedades de estos modelos para razonar a nivel simbólico, prescindiendo de los valores numéricos concretos.

El segundo comentario es que la puerta OR graduada no es la única generalización posible del caso binario. David Heckerman [101] y Sampath Srinivas [219] han desarrollado muy recientemente dos modelos orientados a aplicaciones diferentes. El primero de ellos es más semejante que el segundo a la puerta OR graduada (también Heckerman lo diseñó pensando en aplicaciones médicas), mientras que el de Srinivas está orientado sobre todo a la detección de fallos en circuitos eléctricos. Podría ser interesante discutir aquí las conversaciones que hemos mantenido con ambos investigadores, sobre todo con el segundo de ellos, acerca de las ventajas de cada uno de estos modelos, aunque para ello deberíamos adentrarnos en detalles ajenos a los objetivos de esta tesis.

Capítulo 6

Tratamiento de los bucles

En la sección anterior hemos presentado un algoritmo que permite propagar la evidencia y calcular la probabilidad a posteriori en poliárboles, con dos importantes cualidades: su complejidad es proporcional al número de nodos (suponiendo que el número de valores por cada variable y el número de padres por cada nodo están acotados) y se presta inmediatamente a una implementación distribuida. Por desgracia, la situación se complica considerablemente al pasar de poliárboles a redes múltiplemente conexas.

En primer lugar, si aplicáramos ingenuamente dicho algoritmo a una red con bucles, no sólo sería teóricamente incorrecto (ni siquiera tendría sentido la partición de la evidencia en $\mathbf{e}_X^+$ y $\mathbf{e}_X^-$, o en $\mathbf{e}_{XY}^+$ y $\mathbf{e}_{XY}^-$ [174, pág. 195]), sino que además el algoritmo, al encontrar un bucle, no terminaría nunca, por estar basado en fórmulas recursivas.

Pero lo grave no consiste en no poder aplicar el algoritmo anterior: el origen del problema es mucho más profundo. En efecto, como Cooper [54] demostró, el cómputo de la probabilidad en redes generales es un problema NP (ver el apéndice de este capítulo). Aunque todavía no se ha encontrado una demostración matemática completa, existe el convencimiento unánime de que los problemas NP no tienen solución de complejidad polinómica; es decir, el tiempo necesario para su resolución por ordenador crece exponencialmente con el tamaño del problema. Por tanto, si tratamos de construir RR.BB. más detalladas y que cubran un campo de diagnóstico más amplio, la complejidad de los algoritmos puede crecer tan rápidamente que en seguida nos encontramos con que nuestro programa de ordenador se vuelve terriblemente lento. Sin embargo, no podemos renunciar a tratar los bucles, pues en las aplicaciones reales éstos aparecen con suma frecuencia y trabajar con poliárboles sería una restricción excesiva a la hora de construir nuestros modelos.

Otro inconveniente que acarrean los bucles es que ya no es tan sencillo realizar una computación distribuida con un procesador por cada nodo, con lo que se pierde la belleza del algoritmo anterior. Pero el problema no es sólo de "estética", sino también de eficiencia. Hoy en día se está investigando intensamente sobre la computación en paralelo, con el fin de obtener el máximo rendimiento de las máquinas de múltiples procesadores, que previsiblemente se extenderán de forma considerable en un futuro próximo.

Por estas dos razones, es importante desarrollar algoritmos que traten los bucles de la forma más eficiente y que, a ser posible, admitan una implementación distribuida.

Éste es el objetivo del presente capítulo. En él, ofrecemos primero una revisión de los algoritmos empleados habitualmente para resolver este problema y presentaremos después el método de condicionamiento local, desarrollado en esta tesis.

6.1 Métodos previos

En 1986, Pearl [172] propuso tres métodos para el tratamiento de los bucles, dos exactos (condicionamiento y agrupamiento) y uno aproximado (simulación estocástica). Estos métodos han seguido desarrollándose desde entonces y a su vez han surgido otros nuevos. En esta sección haremos un breve repaso de ellos, primero por la importancia que tienen en sí, ya que se están aplicando a problemas reales, y por su relación con el método de condicionamiento local que propondremos después.

Sería interesante tomar una pequeña red como ejemplo e ilustrar cada uno de los métodos sobre ella. Sin embargo, esto nos ocuparía demasiado espacio. Por otra parte, ese desarrollo se encuentra ya en [174, sec. 4.4].

6.1.1 Condicionamiento

Fue propuesto por Pearl en [172] y desarrollado en [171] y [174, sección 4.4.2]. Suermondt y Cooper [231] completaron el método con un algoritmo de inicialización para la red, que faltaba en la formulación original.

El fundamento teórico del método consiste en la independencia condicional. Por ejemplo, dado el bucle de la figura 6.1, una vez que conocemos el valor de la variable A los nodos B y C son independientes, como si no tuvieran un antepasado común. Es decir, la independencia condicional hace las veces de independencia a priori, rompiendo así los bucles.[1]

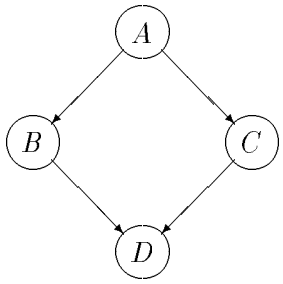


Figura 6.1: Conocer el valor de A nos permite romper el bucle.

Por tanto, cuando hay un solo bucle, el método consiste básicamente en condicionar sobre los valores de una variable, realizando la propagación de la evidencia para cada valor de dicha variable, como si estuviéramos en un poliárbol, y ponderar

[1] Hay que señalar, sin embargo, que no sólo el nodo que está en la parte más alta del bucle es capaz de romperlo por condicionamiento: cualquier nodo, salvo los nodos que están en la parte más baja del bucle (es decir, salvo los nodos que no tienen hijos en el bucle), es un candidato para aplicar condicionamiento.

luego los valores obtenidos para cada valor. En caso de que haya varios bucles, será necesario condicionar sobre varias variables, que constituirán el *conjunto de corte* (en inglés, *cutset*).

Con ello conseguimos resolver los bucles, pero el método en su versión original resulta muy ineficiente por varias razones. En primer lugar, hay que buscar el conjunto de corte. Esto ya es una dificultad no trivial, pues encontrar el conjunto óptimo es un problema NP [229]. En la refencia citada, se ofrece un algoritmo heurístico para hallar un conjunto de corte de tamaño reducido.[2] Otros algoritmos para resolver el mismo problema pueden encontrarse en [223].

El siguiente paso consiste en inicializar la red [231]. Por fortuna, ésta es una operación que puede realizarse una sola vez, almacenando el resultado, pues no depende de la evidencia observada.

Finalmente, debemos calcular la probabilidad a posteriori. El método consiste en tomar uno de los nodos evidencia y realizar la propagación para cada combinación posible de valores de las variables que forman el conjunto de corte. Para cada una de ellas debemos calcular un peso con el fin de ponderar al final la probabilidades a posteriori correspondientes a las diferentes combinaciones. Luego tomamos otro de los nodos evidencia y repetimos todo el proceso, y así sucesivamente con cada uno de ellos. Si el conjunto de corte tiene N variables $A_1, \ldots, A_N$, cada A_i con n_i valores posibles, y tenemos N_E nodos evidencia, hay que repetir el algoritmo de propagación $n_1 \cdot \ldots \cdot n_N \cdot N_E$ veces sobre toda la red. Es decir, la complejidad del algoritmo crece proporcionalmente con el número de nodos, exponencialmente con el tamaño del conjunto de corte y proporcionalmente con el número de hallazgos, por lo que resulta ser totalmente ineficiente.

Puesto que este algoritmo condiciona sobre toda la red, podemos llamarlo *condicionamiento global* para distinguirlo de los métodos que vamos a estudiar a continuación.

El algoritmo original fue modificado por Peot y Shachter [181] dando como resultado lo que podemos llamar *condicionamiento por bloques*. Un *bloque* se define como una porción de la red tal que entre dos nodos cualesquiera hay al menos dos caminos que no comparten ningún arco común. Tanto las redes de la figura 6.5 (pág. 110) como de la figura 6.7 (pág. 120) están formadas por un único bloque.

La ventaja de esta modificación es que la propagación de mensajes *entre bloques* es la misma que la propagación entre nodos en un poliárbol. La desventaja es que *dentro de cada bloque* la propagación es la misma que en el condicionamiento global: hay que repetir el condicionamiento para cada combinación de valores del conjunto de corte. Por eso, si los bloques son grandes, el algoritmo sigue resultando ineficiente.

Otra limitación de este método es que no explica cómo reconocer los bucles que componen una red, y por tanto —a nuestro juicio— el algoritmo está incompleto.

Por último, existe otro método llamado *condicionamiento local* (también podríamos llamarlo *condicionamiento por bucles*) que será presentado con detalle en la sección 6.2.

En cuanto a los sistemas expertos basados en estos algoritmos, puede afirmarse que ni el condicionamiento global ni el condicionamiento por bloques han sido utilizados

[2] Es curioso que dicho algoritmo es incorrecto, puesto que puede incluir en el conjunto de corte un nodo que ni siquiera pertenezca a un bucle. Al comentárselo al autor del algoritmo, nos respondió que el error había sido descubierto anteriormente por Jonathan Stillman.

en aplicaciones reales (al menos sus autores nunca lo han hecho). Sin embargo, el condicionamiento local es el algoritmo utilizado por el sistema experto DIAVAL, descrito en esta tesis.

También hay que señalar que ni el condicionamiento global ni el condicionamiento por bloques permiten una implementación distribuida con un procesador por cada nodo, mientras que el condicionamiento local sí la admite, como veremos en la sección 6.2.4.

6.1.2 Agrupamiento

Otro de los métodos más importantes para la propagación de evidencia en RR.BB. consiste en agrupar las variables para formar nodos de mayor tamaño, de modo que desaparezcan los bucles. Éste es en realidad el método más antiguo, pues fue utilizado ya en los sistemas precursores [29, 50, 180]. Sin embargo, el método no alcanzó gran difusión hasta el trabajo de Lauritzen y Spiegelhalter [135], que fue luego mejorado por Jensen y sus colaboradores [122].

El primer paso del método consiste en añadir enlaces para moralizar el grafo. (Se dice que el grafo es *moral* cuando todos los padres de cada nodo están "casados", es decir, unidos por arcos.) Luego se suprime la dirección de cada arco, con lo cual nos queda un grafo no dirigido.[3]

El paso siguiente consiste en triangular el grafo. Se dice que un grafo está *triangulado* cuando para cada ciclo (hablamos ahora de ciclos no dirigidos) de longitud mayor que tres hay al menos un arco que conecta dos nodos no consecutivos.[4] Por tanto, la triangulación consiste en añadir los arcos necesarios para cortar los ciclos de cuatro o más nodos. Como algoritmo para este proceso, Lauritzen y Spiegelhalter proponen la *búsqueda de máxima cardinalidad*, una técnica heurística desarrollada por Tarjan y Yannakakis [236]. Más adelante explicaremos la importancia de la triangulación por su relación con la complejidad del algoritmo.

Por último, hay que formar los cliques. Se define un *clique* como un "conjunto de nodos completamente conexo maximal", es decir, un conjunto tal que (a) cualquiera de sus nodos está conectado con todos los demás y (b) no hay ningún otro clique diferente que incluya al primero [155]. Cada variable suele estar incluida en varios cliques. El árbol de cliques, construido en la fase estática del método, es independiente de la evidencia observada.

Después hay que inicializar la red asignando a cada clique C_i una función de sus variables, llamada *función potencial*, que consiste en una distribución de probabilidad marginal obtenida a partir de las tablas de probabilidad condicional y de la evidencia disponible. Estas funciones se utilizarán para actualizar las probabilidades a posteriori cuando llegue nueva evidencia. La probabilidad correspondiente a una variable se calcula marginalizando y normalizando la tabla de probabilidad de uno de los cliques que contienen dicha variable. La propagación de evidencia es en esencia muy similar al caso de los poliárboles.

La variante de este método propuesta por Shachter, Andersen y Szolovits [200]

[3] En este método juega un papel importante la teoría de grafos, por lo que puede ser conveniente repasar las definiciones dadas en la sección 4.2.1.

[4] La falta de espacio nos impide explicar con más detalle y con ejemplos estas definiciones. Una excelente presentación se encuentra en [155].

define los *árboles de cúmulos* (*cluster trees*) como una generalización de los árboles de cliques. En su algoritmo, el mensaje que el cúmulo C_i transmite a C_j (fig. 6.2) se calcula mediante la expresión

$$M_{ij} = \sum_{C_i \setminus C_j} \Psi_i \prod_{k \in K_{i-j}} M_{ki} \tag{6.1}$$

donde $C_i \setminus C_j$ representa las variables de C_i no incluidas en C_j, y K_{i-j} incluye los cúmulos vecinos de C_i con la excepción de C_j. Si K_i indica todos los cúmulos adyacentes a C_i, la probabilidad de una configuración de éste último viene dada por

$$P_i = \alpha \, \Psi_i \prod_{k \in K_i} M_{ki}. \tag{6.2}$$

Al igual que en el capítulo anterior, α actúa aquí como constante de normalización. Podemos ver que Ψ_i y P_i dependen solamente de las variables contenidas en C_i, y el mensaje M_{ki} es una función de las variables que pertenecen a la vez a C_i y a C_j. (La intersección de dos cúmulos adyacentes se conoce como *separador* de ambos.)

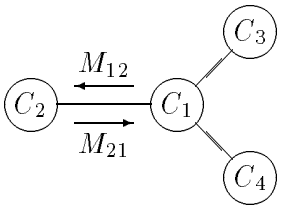


Figura 6.2: Árbol de cúmulos.

Todo lo que dijimos sobre propagación en poliárboles es válido también para los árboles de cúmulos, con las siguientes diferencias:

1. Un poliárbol es un grafo dirigido —cada arco va de una causa a uno de sus efectos— mientras que los árboles de cúmulos son grafos no dirigidos, por lo que no hay distinción entre causas y efectos.

2. En un poliárbol cada nodo representa una variable; en un árbol de cúmulos, cada nodo agrupa varias variables. Por tanto, en este último caso, la probabilidad de cada variable debe obtenerse por marginalización a partir de alguno de los cúmulos a que pertenece.

3. Los mensajes transmitido en un poliárbol dependen de una sola variable; en cambio, en un árbos de cúmulos los mensajes transmitidos entre dos nodos son funciones de las variables que pertenecen a ambos.

6.1.2.1 Implementación distribuida

También para este algoritmo es posible tener una implementación distribuida. Estudiaremos aquí la versión de [200] por su simplicidad y su generalidad.

El esquema de la figura 6.3 muestra las computaciones realizadas en el cúmulo C_1 que aparecía en la figura 6.2. Cada nodo debe almacenar la siguiente información:

- una *lista de las variables* agrupadas en el cúmulo;
- las *tablas de probabilidad condicional* (la tabla correspondiente a cada familia se asigna a uno de los cúmulos que la contienen); la tabla de algunos cúmulos puede ser igual a la unidad en todos sus valores;
- una *función potencial* Ψ_i, obtenida a partir de las tablas de probabilidad condicional y de la evidencia disponible asociada a dicho cúmulo, $\mathbf{e}_i$. Así es como se introducen en la red los hallazgos observados. Esta función se asigna al inicializar el árbol de cúmulos y no varía durante la propagación de evidencia.

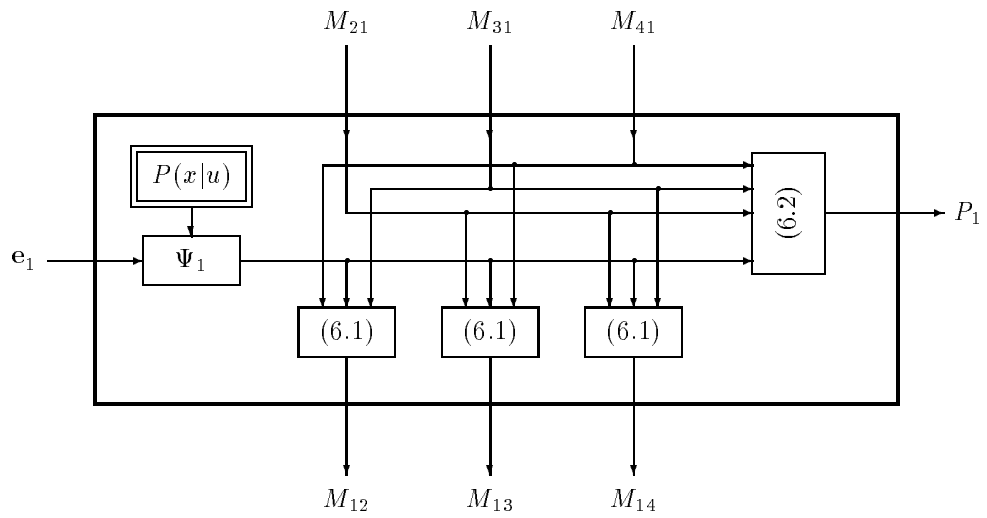


Figura 6.3: Computaciones realizadas en el cúmulo C_1.

Además de tener la evidencia $\mathbf{e}_i$ como entrada, el nodo C_i recibe unos mensajes M_{ji} de todos sus vecinos C_j y los utiliza para calcular P_i y los mensajes M_{ij} que debe enviar. En la figura 6.3 observamos que, al igual que en las figs. 5.3 y 5.5 correspondientes a poliárboles, aquí tampoco hay lazos cerrados en el flujo de información. En la figura 6.3 debería haber una entrada más indicando el conjunto $C_1 \backslash C_j$ como dato, pero éste es más una propiedad del enlace que del nodo y la hemos omitido para ganar en claridad.

En realidad, la figura 6.3 es esencialmente la misma que la figura 5.3. La diferencia aparente se debe a que el algoritmo de los poliárbles realiza la computación de $\pi(x)$ y $\lambda(x)$ como paso intermedio.

6.1.2.2 Aplicaciones

El método de agrupamiento en la versión de Jensen et al. [122] se ha convertido en el algoritmo estándar para el cálculo de la probabilidad en RR.BB. El programa comercial HUGIN, desarrollado por investigadores vinculados a la Universidad de Aalborg y considerado como la herramienta más eficaz para el desarrollo y la computación de RR.BB., está basado en esta técnica [4]. En sus primeras versiones utilizaba la

búsqueda de máxima cardinalidad [236] como algoritmo de triangulación, mientras que versiones más modernas se basan en los métodos heurísticos investigados por Uffe Kjærulff [127]. Sobre esta herramienta el grupo de Aalborg ha desarrollado varios sistemas expertos [7, 5, 6, 163].

Otra herramienta muy difundida, desarrollada en la Universidad de Stanford y en Rockwell Science Center, es el programa IDEAL (*Influence Diagram Evaluation and Analysis in Lisp*) [220], que sigue siendo utilizado por numerosos investigadores como banco de ensayo para poner a prueba sus modelos. Además de ofrecer ayudas para la construcción de RR.BB., el usario puede escoger entre varios algoritmos, no sólo de agrupamiento, y comparar su eficacia. También el sistema comercial PATHFINDER [107, 108], mencionado en la sección 10.1.4, utiliza la propagación en árbol de cliques para evaluar su red de semejanza (sec. 10.1.3).

6.1.3 Simulación

El hecho de que la inferencia exacta en RR.BB. generales es un problema NP [54], se utilizaba con frecuencia como argumento en favor de los métodos aproximados, suponiendo implícitamente que la complejidad de los algoritmos estocásticos propuestos era polinómica. Sin embargo, se ha demostrado muy recientemente que "aproximar la inferencia probabilística en redes bayesianas es [también] un problema NP (*NP-hard*)" [59]. De todos modos, la simulación estocástica sigue siendo una sólución útil para aplicaciones prácticas.

El método consiste en estimar la probabilidad calculando la frecuencia de cada configuración en una serie de ensayos. La R.B. constituye el modelo probabilístico que genera las pruebas. La ventaja de la simulación frente a los métodos exactos es que su complejidad depende sólo del número de nodos y arcos, y no le afecta la topología de la red. Como contrapartida, la simulación es muy sensible a los valores numéricos, que pueden hacer que la aproximación converja con excesiva lentitud.

El primer método de simulación fue el *muestro lógico* (*logic sampling*), propuesto por Henrion [109]. La red se utiliza como generador de muestras, y se toman en cuenta sólo aquéllas que concuerdan con la evidencia observada. El inconveniente de este método tan sencillo es que genera gran cantidad de muestras inútiles, sobre todo cuando aumente el número de hallazgos, con lo que la eficiencia del algoritmo resulta muy pobre.

En el algoritmo de simulación propuesto por Pearl [173], las variables correspondientes a hallazgos toman sólo los valores observados, con el fin de generar siempre muestras útiles. Aunque la idea ha sido mejorada por otros investigadores (hay numerosas referencias en [59]), vamos a presentar aquí el método de simulación directa por su sencillez y porque se presta de forma natural a una implementación distribuida.

La idea básica del algoritmo consiste en asignar aleatoriamente un valor a cada nodo (excepto a los que corresponden a hallazgos) según la distribución de probabilidad de sus vecinos. En una implementación secuencial sólo se activa un nodo cada vez. Sin embargo, en una implementación distribuida, la simulación podría llevarnos a un resultado erróneo si dos nodos adyacentes se activaran a la vez. Esta dificultad se resolvió introduciendo la *inversión de arcos* como estrategia de control distruido para garantizar que cada procesador se activara con frecuencia suficiente pero sin tener dos nodos vecinos activados simultáneamente.

El primer paso consiste en inicializar la red asignando una orientación a cada enlace, con la única condición de que el grafo resultante sea acíclico. La ordenación causal de la R.B. satisface esta restricción, pero podría valer cualquier otra ordenación acíclica. Entonces empieza la simulación. Cada procesador está en espera hasta que todos sus enlaces apunten hacia dentro. En ese momento se activa y realiza las siguientes operaciones:

1. Calcula la distribución condicional de la variable dados los valores asignados a sus vecinos en la simulación.

2. Obtiene un valor para su variable mediante un proceso aleatorio de acuerdo con la distribución calculada.

3. Invierte la dirección de los arcos, de modo que apunten todos hacia fuera. Entonces este nodo queda de nuevo en espera, mientras se activan otros que antes estaban ociosos.

Se puede demostrar que siguiendo estos pasos nunca se van a activar a la vez dos nodos vecinos, que cada procesador se disparará con suficiente frecuencia y que el proceso converge (ver las referencias que aparecen en [173]).

Observar que tanto los cómputos como la estrategia de control son completamente locales, por lo que la *implementación distribuida* en forma de red de procesadores asíncronos no plantea el más mínimo problema.

6.2 Condicionamiento local

Vamos a presentar ahora el método de *condicionamiento local* para la propagación de evidencia en redes con bucles. Este método es una extensión del algoritmo presentado en el capítulo anterior para poliárboles, por lo que ambos tienen mucho en común. De hecho, el condicionamiento local se aplica sobre un *poliárbol asociado* en el que siguen vigentes muchas de las propiedades y definiciones de la sección 5.1.1. Por tanto, un paso previo a la propagación de evidencia consiste en hallar el poliárbol a partir de la R.B. original.

6.2.1 Poliárbol asociado y asignación de bucles

Como vimos en la sección 6.1.1, todo método de condicionamiento consta de dos fases:

1. Seleccionar un conjunto de nodos, llamado *conjunto de corte*, que rompa los bucles de la red.

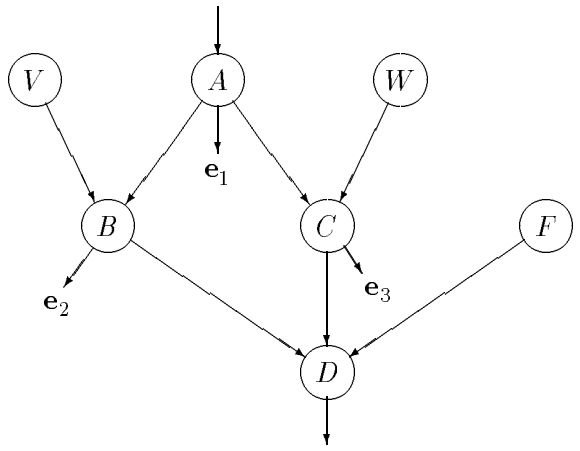
2. Propagar la evidencia en el poliárbol resultante.

En dicha sección mencionamos cuáles son los algoritmos utilizados para hallar el conjunto de corte. Aquí vamos a presentar uno algoritmo diferente que no sólo obtiene un conjunto de corte sino también un *poliárbol asociado*, resultante de eliminar algunos enlaces y añadir *nodos ficticios* (*phantom nodes*). Este nuevo algoritmo detecta además a qué bucles pertenece cada nodo. El proceso consiste en realizar una búsqueda en profundidad sobre el grafo no dirigido. Podemos describir el algoritmo mediante el siguiente ejemplo.

6.2.1.1 Poliárbol asociado

Supongamos que la figura 6.4.*a* muestra una porción de una red que no contiene ningún otro bucle. Partiendo de un nodo escogido arbitrariamente (A en nuestro ejemplo), que llamaremos *nodo pivote*, recorremos la red ignorando la dirección de los arcos y marcando los nodos visitados: B, V, etc. Al llegar a un nodo que no tiene ningún otro enlace por donde poder continuar, retrocedemos (en inglés este proceso se denomina *backtracking*) y seguimos recorriendo las demás ramas del nodo anterior.

a)



b)

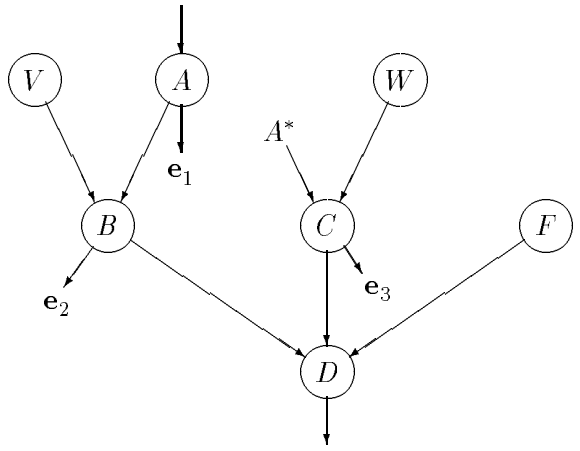


Figura 6.4: Una parte de la red y de su poliárbol asociado.

Si *al avanzar* desde el nodo C (no al retroceder) llegamos a un nodo A que ya ha sido marcado, significa que hemos detectado un bucle; entonces hemos de incluir en el conjunto de corte el nodo que está *en el extremo superior del enlace* y eliminar el enlace AC. En su lugar, añadiremos un enlace desde un *nodo ficticio* A^* hasta el nodo que estaba en la parte inferior del enlace (fig. 6.4.*b*). Un nodo de la red original puede dar lugar a varios nodos ficticios si rompe varios bucles.

Tras eliminar tantos arcos como sea necesario, al final quedará entre cada par de

nodos *como máximo* un camino; y habrá *como mínimo* un camino porque si hemos eliminado un enlace AC esto significa que hemos llegado a A desde C (o viceversa) por otro camino. Tenemos por tanto una red simplemente conexa y podemos utilizar las mismas definiciones de $\mathbf{e}_X^+$, $\mathbf{e}_X^-$, $\mathbf{e}_{XY}^+$ y $\mathbf{e}_{XY}^-$ que en poliárboles (sec. 5.1.1). Por otra parte, la razón más importante de haber eliminado algunos enlaces es que ahora los mensajes de propagación de evidencia circularán por los arcos del poliárbol asociado y no por los de la red original, como veremos en la sección 6.2.2.

El conjunto de corte y el poliárbol asociado obtenidos no son únicos: dependen del nodo pivote elegido y del orden en que visitemos los vecinos de cada nodo. Más adelante página 111 sugeriremos cómo escoger un nodo pivote que lleve a una computación más eficiente.

6.2.1.2 Asignación de bucles

Acabamos de ver cómo podemos detectar los bucles mediante una búsqueda en profundidad. Nuestro propósito es extender el algoritmo anterior para que, además, nos indique a qué bucles pertenece cada nodo, es decir, sobre qué variables debe ser condicionado cada nodo. Este apartado está lleno de detalles técnicos de poca importancia conceptual, por lo que no es necesario para comprender el resto de la exposición.

Consideremos el ejemplo de la figura 6.5. Podemos recorrer los nodo en el orden A, B, G, F, D, C. Cuando intentamos seguir adelante, C detecta que A estaba marcado. Entonces eliminamos el enlace AC y añadimos `A` a la lista de bucles de C. Al retroceder (*backtracking*), cada nodo recoge los nodos de sus ramas, con dos excepciones que luego describiremos.

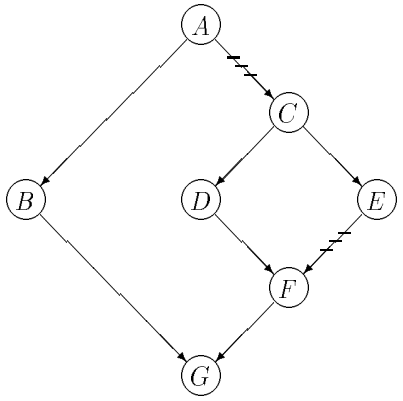


Figura 6.5: Bucles adyacentes.

La búsqueda continúa, hasta que al pasar de E a F, vemos que éste también estaba marcado: hemos detectado otro bucle. Sin embargo, hay una sutil diferencia. Antes, habíamos llegado a A desde abajo y, por tanto, condicionábamos sobre el nodo de llegada.[5] Ahora, en cambio, hemos cerrado el bucle desde arriba, por lo que no condionaremos sobre el nodo de llegada, F, sino sobre el nodo que acaba de detectar el bucle, E. Este nodo pasa a formar parte de su propia lista de bucles, para que C

[5] Tal como hemos indicado anteriormente, incluimos en el conjunto de corte el nodo que está *en el extremo superior* del enlace.

recoja E como bucle. Además, debemos informar a F de que hemos encontrado un bucle desde su padre E.[6]

Seguimos retrocediendo. El nodo D recoge los bucles de C: A y E. También F recoge A y E pero "recuerda" que habíamos cerrado un bucle desde su padre E; por tanto, sólo entrega A a G. Esta es la primera excepción en la transmisión de bucles.

La otra excepción es que un nodo se excluye a sí mismo de los bucles recogidos en una rama. En nuestro ejemplo, A no está incluido en su propia lista de bucles, de modo que si tiene antepasados en el árbol de búsqueda, éstos tampoco incluirán el nodo A en su lista de bucles.

En resumen, cada nodo guarda una lista de los bucles a que pertenece, así como una lista auxiliar en que anota los bucles que fueron cerrados desde alguno de sus padres. Al retroceder de Y a X, el segundo recoge todos los nodos del primero, excluyéndose tanto a sí mismo como a los nodos de la lista auxiliar de Y.

Tras aplicar este algoritmo una vez, vemos que el nodo C ha fundido dos condicionamientos distintos, en A y en E, relativos a dos bucles diferentes. Esto nos sugiere, como regla heurística, escoger como nodo pivote aquél que rompe más bucles, para que la dimensión de los mensajes transmitidos sea lo más reducida posible (ver las secciones siguientes).

6.2.1.3 Algoritmos previos para hallar el conjunto de corte

Como hemos mencionado anteriormente, el problema de encontrar el conjunto de corte más pequeño es NP. El algoritmo iterativo descrito en [229], al buscar los bucles visita cada nodo $3n$ veces, donde n es el número de nodos en la red. Por tanto, su complejidad —en el peor de los casos— es $O(n^2)$. En cambio, el algoritmo recursivo que nosotros hemos descrito visita cada nodo como mucho n veces y no necesita actualizar el grado (el número de vecinos) de cada nodo. Su complejidad es $O(e)$, con e igual al número de enlaces, porque recorre cada enlace sólo una vez. Esto supone una ventaja para redes poco densas (*sparse networks*), donde $e \ll n^2$.

Más aún, el algoritmo de [229] puede incluir incorrectamente en el conjunto de corte un nodo que no pertenezca a ningún bucle. Para corregir el error, Stillman [223] propuso una versión modificada, que suele encontrar un conjunto de corte más pequeño para la mayoría de las redes. La referencia citada muestra además una red para la que tanto la versión original como la modificada dan un conjunto de corte de tamaño $\Omega(\lfloor \frac{n}{4} \rfloor)$ a pesar de que el conjunto de corte óptimo contiene sólo dos nodos. En cambio, en la misma red, nuestro algoritmo encuentra un conjunto de tan sólo dos o tres nodos, dependiendo del nodo pivote escogido. Sin embargo, no es posible establecer en general qué algoritmo encontrará el conjunto de corte más pequeño; el resultado depende de cada red.

En cualquier caso, los algoritmos previos para hallar conjuntos de corte no sirven para el método de condicionamiento local porque no hallan el poliárbol asociado ni indican en qué variable hay que condicionar cada nodo.

[6] En la versión dinámica del algoritmo (sec. 6.2.2.4), veremos que, en el primer caso encontramos un bucle al intentar calcular $\pi_C(a)$; en el segundo, al buscar $\lambda_F(e)$. En consecuencia, $\lambda_Y(x)$ y $\pi_Y(x)$ romperán los bucles de modo diferente.

6.2.2 Propagación de evidencia

Una vez que hemos obtenido un conjunto de corte y un poliárbol asociado, la segunda fase del método de condicionamiento consiste en propagar la evidencia con el fin de calcular la probabilidad a posteriori $P^*(x)$ de cada nodo.

Vamos a mostrar el algoritmo de propagación a partir de la figura 6.4, suponiendo que la red no contiene ningún otro bucle. Tomaremos A como nodo condicionante con el fin de romper el bucle. Para hacer más claro nuestro estudio, lo dividiremos en dos casos: cuando los mensajes condicionados recorren los enlaces correspondientes al bucle desde el nodo ficticio A^* hasta el nodo real A, y cuando los recorren en sentido contrario.

6.2.2.1 Propagación desde un nodo ficticio

Examinemos la red de la figura 6.4. El cálculo de $\pi(a)$ no plantea ningún problema, pues hemos supuesto que no hay ningún bucle por encima de A. Veamos entonces cómo hallar $\lambda(a)$. Según su definición (ec. (5.15)),

$$\lambda(a) = P(\mathbf{e}_A^-|a) = P(\mathbf{e}_1|a)\, P(\mathbf{e}_{AB}^-|a) = \lambda_{\mathbf{e}_1}(a)\, \lambda_B(a) \tag{6.3}$$

Tenemos que calcular $\lambda(a)$ para cada valor concreto a, por lo que podemos aplicar la *independencia condicional* con el fin de romper el bucle. En consecuencia, tenemos

$$\begin{aligned}\lambda_B(a) &= P(\mathbf{e}_{AB}^-|a) = \sum_{b,v} P(\mathbf{e}_B^-, b, \mathbf{e}_{BV}^+, v|a)\\ &= \sum_{b,v} P(\mathbf{e}_B^-|b,a)\, P(b|v,a)\, P(\mathbf{e}_{VB}^+, v)\\ &= \sum_b \left[\lambda(b|a) \sum_v P(b|v,a)\, \pi_B(v)\right]\end{aligned} \tag{6.4}$$

donde

$$\lambda(b|a) \equiv P(\mathbf{e}_B^-|b,a) \tag{6.5}$$

Podemos definir también

$$\lambda_D(b|a) \equiv P(\mathbf{e}_{BD}^-|b,a) \tag{6.6}$$

con el fin de expresar $\lambda(b|a)$ como

$$\lambda(b|a) = \lambda_{\mathbf{e}_2}(b)\, \lambda_D(b|a) \tag{6.7}$$

El valor de $\lambda_D(b|a)$ se obtiene así:

$$\begin{aligned}\lambda_D(b|a) &= \sum_{d,c,f} P(\mathbf{e}_D^-, d, \mathbf{e}_{CD}^+, c, \mathbf{e}_{FD}^+, f|b,a)\\ &= \sum_{d,c,f} P(\mathbf{e}_D^-|d)\, P(d|c,f,b)\, P(\mathbf{e}_{CD}^+|c,a)\, P(c|a)\, P(\mathbf{e}_{FD}^+, f)\\ &= \sum_d \left[\lambda(d) \sum_{c,f} P(d|b,c,f)\, \pi_D(c|a)\, \pi_D(f)\right]\end{aligned} \tag{6.8}$$

donde hemos realizado la siguiente sustitución:

$$\pi_D(c|a) \equiv P(c, \mathbf{e}^+_{CD}|a) \tag{6.9}$$

Igualmente, introducimos

$$\pi(c|a) \equiv P(c, \mathbf{e}^+_C|a) \tag{6.10}$$

para calcular

$$\pi_D(c|a) = P(\mathbf{e}_3|c)\, P(\mathbf{e}^+_C, c|a) = \lambda_{\mathbf{e}_3}(c)\, \pi(c|a) \tag{6.11}$$

y

$$\pi(c|a) = \sum_w P(\mathbf{e}^+_{WC}, c, w|a) = \sum_w P(c|w, a)\, \pi_C(w). \tag{6.12}$$

Observar que no es necesario transmitir un mensaje $\lambda_C(a)$, ya que la evidencia $\mathbf{e}_3$ se ha propagado hasta A a lo largo del camino C–D–B–A. Como ya hemos indicado, los mensajes no circulan por los arcos de la red original sino por los del poliárbol asociado, donde hay un sólo camino desde cada hallazgo hasta cada nodo.

6.2.2.2 Propagación hacia un nodo ficticio

Ahora, debemos propagar la evidencia a lo largo del bucle en sentido contrario, es decir, recogiendo toda la evidencia que está por encima de A y transmitiendo su efecto hacia A^*, de modo que podamos calcular la probabilidad a posteriori de cada nodo.

Para el B tenemos ya $\lambda(b|a)$ dada por ecuación (6.7). Si tuviéramos también

$$\pi(b, a) \equiv P(b, \mathbf{e}^+_B, a) \tag{6.13}$$

podríamos calcular

$$\begin{aligned} P^*(b) &= P(b|\mathbf{e}) = \sum_a P(b, a|\mathbf{e}) = \alpha \sum_a P(b, a, \mathbf{e}) \\ &= \alpha \sum_a \lambda(b|a)\, \pi(b, a) \end{aligned} \tag{6.14}$$

La expresión para $\pi(b, a)$ nos queda

$$\begin{aligned} \pi(b, a) &= \sum_v P(b, \mathbf{e}^+_{VB}, v, \mathbf{e}^+_{AB}, a) \\ &= \sum_v P(b|v, a)\, P(\mathbf{e}^+_{VB}, v)\, P(\mathbf{e}^+_{AB}, a) \\ &= \sum_b P(b|v, a)\, \pi_B(v)\, \pi_B(a) \end{aligned} \tag{6.15}$$

que es en esencia la ecuación (5.12), con la única diferencia de que queremos conocer el valor de $\pi(b)$ para cada valor a, y por eso no sumamos sobre esta variable. El valor de $\pi_B(a)$ es exactamente el mismo que indica la ecuación (5.13), pues es un mensaje que llega desde fuera del bucle:

$$\pi_B(a) = \lambda_{\mathbf{e}_1}(a)\, \pi(a) \tag{6.16}$$

Del mismo modo

$$\begin{aligned}\pi(d) &= \sum_{a,b,c,f} P(d, \mathbf{e}^+_{CD}, c, \mathbf{e}^+_{BD}, b, a, \mathbf{e}^+_{CF}.f) \\ &= \sum_{a,b,c,f} P(d|b,c,f)\,\pi_D(c|a)\,\pi_D(b,a)\,\pi_D(f) \qquad (6.17)\end{aligned}$$

con $\pi_D(b,a)$ definido por

$$\pi_D(b,a) \equiv P(b,a,\mathbf{e}^+_{BD}) \qquad (6.18)$$

y calculado así:

$$\pi_D(b,a) = \lambda_{\mathbf{e}_2}(b)\,\pi(b,a). \qquad (6.19)$$

Nuestro próximo objetivo es definir $\lambda_D(c;a)$ tal que podamos luego escribir

$$P^*(c) = \sum_a \lambda_{\mathbf{e}_3}(c)\,\lambda_D(c;a)\,\pi(c|a) \qquad (6.20)$$

en concordancia con las ecuaciones (5.10) y (5.15).

Curiosamente, la expresión $\lambda_D(c;a)$ resulta más fácil de calcular que de definir. En efecto, la definición intuitiva $\lambda_D(c;a) \equiv P(\mathbf{e}^-_{CD}, a|c)$ no valdría, porque nuestro objetivo es calcular la influencia de $\mathbf{e}^-_{CD}$ sobre C a lo largo del camino A–B–D–C para cada valor de a, pero no queremos que el asignar un valor a C modifique la probabilidad de A a través del enlace AC. En consecuencia, debemos descomponer $\mathbf{e}^-_{CD}$ en dos subconjuntos:

— la evidencia conectada con C sin pasar por A, que es $\mathbf{e}_{C-A} \equiv \mathbf{e}^-_{CD} \cap \mathbf{e}^-_{AB}$,[7] y

— la evidencia que está por encima del enlace AB, $\mathbf{e}^+_{AB}$.

Así, podemos factorizar $\mathbf{e}^-_{CD}$ y obtenemos

$$P(c|\mathbf{e}) = \sum_a P(\mathbf{e}_3|c)\,P(\mathbf{e}_{C-A}|c,a)\,P(\mathbf{e}^+_C, c|a)\,P(\mathbf{e}^+_{AB}, a)$$

Al comparar este resultado con la ecuación (6.20), concluimos que la definición de $\lambda_D(c;a)$ debe ser

$$\lambda_D(c;a) \equiv P(\mathbf{e}_{C-A}|c,a)\,P(a,\mathbf{e}^+_{AB}). \qquad (6.21)$$

Puesto que $\pi_D(b,a) = P(\mathbf{e}_{A-D}|b,a)P(a,\mathbf{e}^+_{AB})$, tenemos finalmente

$$\lambda_D(c;a) = \sum_d \lambda(d)\left[\sum_{b,f} P(d|b,c,f)\,\pi_D(b,a)\,\pi_D(f)\right]. \qquad (6.22)$$

[7] Observar que la definición es simétrica, pues $\mathbf{e}_{A-C}$, la evidencia conectada con A sin pasar por C, es la misma que $\mathbf{e}_{C-A}$.

6.2.2.3 Recapitulación

Ahora que tenemos todas las fórmulas, veamos que no son tan complicadas como parecen. En primer lugar, hemos roto el bucle sustituyendo el arco AC por un enlace desde el nodo ficticio A^* hasta C. Cuando los mensajes viajan desde A^* hasta A recorriendo el poliárbol (a través de los arcos que pertenecían al bucle), están condicionados en a. Observar, por ejemplo, las ecuaciones (6.8) o (6.11). Cuando viajan en sentido opuesto, son proporcionales a $P(a, \mathbf{e}^+_{AB})$; las ecuaciones (6.17) y (6.20) reúnen ambos casos.

En la implementación del algoritmo no es necesario preocuparse por esta distinción. Cuando solicitamos un mensaje a un nodo ficticio A^*, éste nos da una respuesta diferente para cada valor a, de modo que la propagación se realiza en paralelo en vez de mezclar el efecto de todos estos valores. Esto se logra simplemente al no realizar la suma sobre a (comparar las ecs. (6.12) y (6.15) con las correspondientes ecuaciones para poliárboles).

Tampoco es necesario tener implementaciones diferentes para $\lambda_Y(x)$, $\lambda_Y(x;z)$ y $\lambda_Y(x|z)$. Las ecuaciones (6.4), (6.8) y (6.22) son **formalmente equivalentes** a las correspondientes definiciones para poliárboles. La única diferencia es que ahora las expresiones que aparecen pueden ser matrices (dependen de varias variables) en lugar de vectores (en un poliárbol, cada mensaje depende de una sola variable). Por tanto, sólo necesitamos cuatro funciones —o cuatro procedimientos si estamos en Programación Orientada a Objetos—: $\pi(X)$, $\lambda(X)$, $\pi_Y(X)$ y $\lambda_Y(X)$.[8]

El último punto importante que debemos mencionar es que podemos reducir la dimensión de los mensajes cuando el condicionamiento ya no es necesario. Este fenómeno ha aparecido ya al calcular $\lambda_B(a)$, dada por ecuación (6.4): al cerrar el bucle, el condicionamiento desaparece. También se reduce la dimensión de los mensajes al salir del bucle: los nodos de la red adyacentes al bucle pero no incluidos en él pueden sumar sobre a como si pertenecieran a un poliárbol. Por ejemplo, en la red de la figura 6.4, $\lambda_D(f)$ se puede obtener como

$$\lambda_D(f) = \sum_d \left[\lambda(d) \sum_{b,c,a} P(d|b,c,f)\, \pi_D(b,a)\, \pi_D(c|a) \right] \tag{6.23}$$

Como F no pertenece a este bucle (F "lo sabe" gracias a su lista de bucles), el mensaje $\lambda_D(F)$ no está condicionado en A.

6.2.2.4 Generación dinámica del poliárbol asociado

Hemos comenzado exponiendo un método para hallar el poliárbol asociado y asignar bucles a cada nodo. Luego, hemos derivado las fórmulas para propagar la evidencia. Mostraremos ahora que las dos fase pueden integrarse en único algoritmo si $\pi(X)$, $\lambda(X)$, $\pi_Y(X)$ y $\lambda_Y(X)$, además de calcular la probabilidad, se ocupan de actualizar las listas de bucles. (Observar el paralelismo entre la esta sección y la 6.2.1.)

Volvamos al ejemplo de la figura 6.5. Escogemos A como nodo pivote y, en la red original, intentamos calcular sucesivamente $\lambda_B(A)$, $\lambda_G(B)$, $\pi_G(F)$, $\pi_F(D)$, $\pi_D(C)$ y

[8] *Notación:* $\pi(X)$ puede representar tanto un vector $\pi(x)$ como una matriz; por ejemplo, $\pi(x, a|b)$.

$\pi_C(A)$. El nodo A está entonces marcado como "visitado". Por tanto, eliminamos el arco AC y tenemos que $\pi(C) = \pi(c|a)$.[9]

El paso siguiente consiste en calcular $\lambda_E(C)$ y $\lambda_F(E)$. Pero F es también un nodo visitado. Por tnato, $\lambda_F(E)$ actualiza las listas de nodos tal como se explicó en la sección 6.2.1.2 y no devuelve ningún valor porque el enlace EF ya ha sido eliminado. Seguimos adelante, calculando $\lambda_E(c;e)$, $\pi_D(c,e|a)$, $\pi_F(d,e|a)$ y $\pi(f,e|a)$. Observar que e y a desempeñan un papel diferente en estas ecuaciones porque estamos propagando evidencia desde un nodo real E y desde un nodo ficticio A^*. El nodo G no pertenece al bucle cortado por E, así que podemos sumar sobre e y la propagación continúa con $\pi_G(f|a)$, $\lambda_G(b|a)$ y $\lambda_B(a)$.

Cuando hemos completado este primer paso de mensajes y tenemos ya el poliárbol asociado, podemos realizar un segundo paso en que los mensajes cruzan los enlaces en sentido opuesto.

Hay una propiedad que puede llevar a ahorrar bastantes cálculos en este método cuando el poliárbol asociado se genera dinámicamente. Por definición, un bucle está *desbloqueado* cuando un nodo que está en su parte inferior no ha recibido evidencia desde sus efectos (su λ no contiene información).[10] En este caso, las ramas del bucle son condicionalmente independientes. En la práctica, esto implica que, si no necesitamos conocer la probabilidad de este nodo, podemos eliminarlo, el bucle queda abierto y evitamos el condicionamiento en esta parte de la red [8]. Por ejemplo, en el grafo de la figura 6.4.a, si si no hay ningún hallazgo por debajo de D, podemos eliminar este nodo y los enlaces BD, CD y FD. Desde el punto de vista de la implementación esto significa que hay que visitar primero los efectos de un nodo antes que sus causas, para ver si podemos eliminar algún nodo *antes* de que se haya cerrado el bucle.

6.2.3 Eficiencia del algoritmo

Puesto que la inferencia probabilística es NP para RR.BB. [54], se supone que es imposible encontrar un algoritmo de complejidad polinómica para redes múltiplemente conexas. Sin embargo, la complejidad para una red particular depende de su estructura tanto como del número de nodos. Nuestro próximo objetivo es comparar nuestro algoritmo con otros algoritmos exactos descritos en la sección 6.1, mediante el análisis de varios tipos de redes.

Suermondt y Cooper [230] estudiaron dos métodos exactos, condicionamiento global (C.G.) [172, 174] y propagación en árbol de cliques (P.A.C.) [135, 122], e identificaron dos estructuras para las cuales uno de los algoritmos tiene complejidad lineal y el otro complejidad exponencial, y viceversa (ver también [155, pp. 297–305]). Aquí mostraremos que, para ambas estructuras, nuestro algoritmo de condicionamiento local (C.L.) tiene complejidad lineal y requiere mucha menos computación que cualquiera de los dos anteriores.

En el caso de una estructura con largos bucles paralelos que comparten un nodo superior común, la complejidad de la P.A.C. es exponencial en el peor caso, porque la

[9] En la implementación no es necesario crear nodos ficticios: tan sólo son un concepto útil para deducir las fórmulas. Tampoco es necesario formar el conjunto de corte; éste era necesario en el condicionamiento global, pero en el condicionamiento local carece de importancia práctica.

[10] En un bucle "en pajarita" ($\bowtie$), donde dos causas están conectadas con dos efectos, cualquiera de los dos nodos inferiores puede hacer que el bucle esté desbloqueado.

triangulación mediante búsqueda de máxima cardinalidad genera grandes cúmulos.[11] Pero es un problema NP encontrar la triangulación óptima [252]. El C.G., en cambio, siempre tiene complejidad lineal para este problema, y el C.L. es aún más eficiente por las razones expuestas en sección 6.2.4.2.

Estudiemos ahora la cadena de rombos (*diamond ladder*, fig. 6.6), la estructura en que la P.A.C. tiene complejidad lineal mientras que la complejidad del C.G. es exponencial. Podemos escoger el nodo superior D_0 como pivote. Recorremos la red en el orden D_0, B_1, D_1, C_1, B_2, etc. Para cada rombo i, $\pi_{D_i}(C_i)$ estará condicionado sobre D_{i-1}, pero $\lambda(D_i)$ es un vector (depende de una sola variable), de modo que el condicionamiento no se acumula.

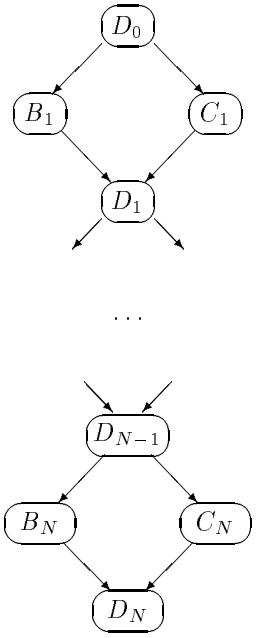


Figura 6.6: Cadena de rombos.

La tabla 6.1 muestra el número de operaciones matemáticas (sumas, multiplicaciones y unas pocas divisiones) necesarias para calcular cada expresión, en el caso de variables binarias, según las fórmulas derivadas anteriormente. En consecuencia, el número de operaciones necesarias es 123 para cada rombo. El cálculo de $P(D_0|\mathbf{e})$ supone 5 operaciones más. Si la cadena tiene N rombos, el número de nodos es $n = 3N + 1$ y el número de operaciones necesarias resulta ser $123N + 5 = 41n - 36$.

Suermondt y Cooper [230] estiman que la P.A.C., *además de la triangulación*, necesita $330N + 111 = 110n - 119$ operaciones para la propagación de evidencia en esta red. Para un solo bucle pequeño ($N = 1$), las operaciones necesarias son 128 para el C.L. frente a 441 para la P.A.C. Estas estimaciones pueden variar dependiendo de la implementación considerada, pero para el propósito de esta sección, nos basta mostrar que el C.L. es más eficiente que la P.A.C. incluso en el caso que resultaba

[11] Otros métodos de triangulación podrían dar lugar a complejidad lineal para la P.A.C. en esta estructura.

Desde D^*_{i-1} hacia D_{i-1}	$\pi(c_i\|d_{i-1})$	0
	$\pi_{D_i}(c_i\|d_{i-1})$	6
	$\lambda_{D_i}(b_i\|d_{i-1})$	10
	$\lambda_{B_i}(d_{i-1})$	6
Desde D_{i-1} hacia D^*_{i-1}	$\pi_{B_i}(d_{i-1})$	4
	$\pi(d_i)$	46
	$\lambda_{D_i}(c_i; d_{i-1})$	28
Probabilidad	$P^*(b_i)$	9
	$P^*(d_i)$	5
	$P^*(c_i)$	9
Total		123

Tabla 6.1: Operaciones necesarias para la cadena de rombos.

más favorable para ésta frente a los métodos de condicionamiento "tradicionales".

De todos modos, hay reconocer que la técnica disponible más eficiente para este problema particular es el agrupamiento. Tanto el método propuesto en [174, pp. 197–204] como el algoritmo de agrupamiento (*clustering algorithm*) [200] necesitan tan sólo $84N + 5 = 28n - 33$ operaciones, como puede deducirse examinando los mensajes π y λ o los mensajes M_{ij}, respectivamente, aunque el C.L. podría aún tener otras ventajas sobre estos algoritmos.

6.2.3.1 Comentario marginal

Para ver hasta qué punto la complejidad de los algoritmos *exactos* depende de la topología de la red, consideremos la adición de un arco desde D_0 hasta D_N en la figura 6.6. El número de nodos sigue siendo el mismo y el número de enlace varía solamente de $4N$ a $4N + 1$, pero el tiempo de computación se duplica, porque hay que aplicar un condicionamiento adicional (D_0, por ejemplo) sobre toda la red. En la P.A.C., la búsqueda de máxima cardinalidad en sentido descendente sobre el nuevo grafo lleva a la presencia de D_0 en cada clique, que es precisamente lo mismo que condicionar toda la red sobre este nodo [200]. (Aquí se ve más claramente la semejanza entre los diferentes algoritmos que estamos estudiando.) Este aumento de complejidad parece ser inherente a la red, y no es de esperar que otro algoritmo pudiera mejorar esencialmente estos resultados. Sin embargo, la presencia de un pequeño bucle adicional —por ejemplo, el resultante de añadir un enlace de C_1 a C_2— apenas aumenta el número de computaciones necesarias en cualquiera de los métodos que conocemos.

La complejidad de las RR.BB. es, en nuestra opinión, una cuestión abierta, que resulta de sumo interés por su relación con los problemas NP, un tema que aún está por resolver.

6.2.4 Ventajas del condicionamiento local

6.2.4.1 Dos versiones del algoritmo

Hay dos variantes principales en la implementación del algoritmo, dependiendo de si propagamos la evidencia en un poliárbol asociado *estático* generado previamente o en uno *dinámico* hallado mientras se propaga la evidencia.

La importancia de la primera versión es que permite una **implementación distribuida**, con un procesador por cada nodo incluso en redes con bucles. No es necesario tener un nodo pivote ni realizar los dos pasos de mensajes. Basta tener un poliárbol asociado con una asignación de bucles para cada nodo, obtenido mediante el algoritmo de búsqueda en profundidad de la sección 6.2.1 o mediante cualquier otro de que dispusiéramos. A partir de ahí, la propagación puede empezar simultáneamente en ambos sentidos sobre cada enlace del poliárbol asociado. La diferencia con el caso del auténtico poliárbol es que ahora la complejidad de los cálculos que realiza cada nodo depende no sólo del número de vecinos y de los valores de las variables, sino también del número de bucles a que pertenece.

La otra versión, la generación dinámica del poliárbol, resulta ventajosa cuando hay bucles desbloqueados (pág. 116). En general, no es rentable este ahorro de computación cuando van a aparecer nuevos hallazgos que podrían cerrar algún bucle, con lo que habría que aplicar otra vez el algoritmo desde el principio. Pero el ahorro es provechoso cuando toda la evidencia está ya disponible antes de calcular la probabilidad o cuando la evidencia llega por bloques, como es el caso de la medicina, en que una nueva prueba añade varios nodos —y quizá también varios bucles— a la red. La conclusión es que resulta mucho más efectivo calcular la probabilidad en una red que contenga sólo los nodos terminales correspondientes a las pruebas ya realizadas.

6.2.4.2 Comparación con otros métodos

a) Condicionamiento

El condicionamiento local (C.L.) es más efectivo que el condicionamiento global (C.G.) siempre que haya bucles, por dos razones:

• El método original aplicaba el condicionamiento sobre toda la red, por lo que su complejidad crecía exponencialmente con el tamaño del conjunto de corte. Sin embargo, el C.L. condiciona sólo dentro de cada bucle, con lo que puede reducir drásticamente la complejidad del algoritmo: ya vimos, por ejemplo, que para la cadena de rombos (sec. 6.2.3) ésta pasaba de ser exponencial a lineal.

• En el C.G., era necesario realizar primero una inicialización de la red [231] y luego calcular recursivamente los pesos transmitidos desde cada hallazgo a todos los demás nodos. Eso hacía que la complejidad del algoritmo, además de crecer exponencialmente con el tamaño del conjunto de corte, fuera proporcional al número de hallazgos (sec. 6.1.1). En el C.L., el tiempo de propagación es independiente del número de hallazgos, una ventaja que aparecía ya en el condicionamiento por bloques (C.B.) de Peot y Shachter [181] y está relacionada con la nueva definición de $\pi(x)$ y $\pi_X(Y_j)$ (sec. 5.1.1).

Además de esta nueva definición, el C.L. y el C.B. comparten tambíen la intuición de que no es necesario condicionar sobre toda la red. Sin embargo, hay varias diferencias entre ellos:

1. Desde un punto de vista conceptual, es importante insistir en que la evidencia se propaga sobre un poliárbol asociado, para poder definir con propiedad los mensajes π y λ. En el trabajo de Peot y Shachter, estos mensajes son computados correctamente a pesar de estar incorrectamente definidos. Por ejemplo, hay que distinguir explícitamente entre $\pi_Y(x, z)$ y $\pi_Y(x|z)$, pues de otro modo, estaríamos contando dos veces la misma evidencia en exprexiones como (6.17).

2. A la hora de hallar el conjunto de corte, el C.B. se basa en el algoritmo que aparece en [229] (que resultó ser defectuoso), pero no indica cómo detectar los bloques.[12] En el C.L., por otra parte, se incluye un algoritmo para hallar el conjunto de corte, construir el poliárbol asociado y asignar los bucles.

3. El algoritmo de Peot y Shachter aplica el condicionamiento globalmente dentro de cada bloque, examinando iterativamente cada asignación de valores para las variables del conjunto de corte correspondiente. Por tanto, todos los nodos que pertenecen al mismo bloque están condicionados en las mismas variables. En el C.L., sin embargo, los mensajes consisten en matrices que se propagan por el poliárbol asociado; la lista de bucles asociada a cada nodo permite decidir localmente, es decir, de forma individualizada para cada nodo, qué condicionamiento es necesario.

4. En consecuencia, la propagación es más eficiente en el C.L. porque se basa en bucles en vez de basarse en bloques. Por ejemplo, la escalera que muestra la figura 6.7 consta de N bucles adyacentes pero está formada por un solo bloque. En el C.B., toda la red estaría condicionada sobre el correspondiente conjunto de corte, cuyo tamaño es $O(n)$, donde $n = 2N + 2$ representa el número de nodos. Esto da lugar a una complejidad exponencial en el número de nodos. Sin embargo, en el C.L. cada nodo pertenece sólo a uno o dos bucles, independientemente del número de "escalones", con lo que el C.L. logra mantener la complejidad lineal también para esta estructura. El ejemplo no es un caso forzado ni extraño, pues la misma diferencia entre el C.B. y el C.L. existe en muchas otras redes compuestas de bucles adyacentes.

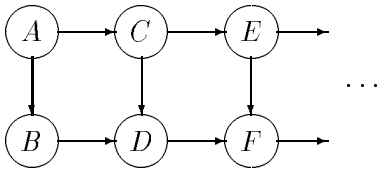


Figura 6.7: La escalera contiene N cuadrados y $2N + 2$ nodos.

5. El C.L. consigue mejor rendimiento en implementaciones paralelas. Para entender el porqué, supongamos que la escalera considerada anteriormente forma parte de una red más amplia y que el nodo C ha recibido ya un mensaje desde la parte derecha de la red. Tan pronto como reciba un mensaje desde A o desde D

[12] Por eso dijimos anteriormente que, en nuestra opinión, se trata de un algoritmo incompleto.

(dependiendo de qué enlace se haya eliminado al construir el poliárbol), podrá enviar ya un mensaje a sus vecinos de fuera de la escalera; no es necesario esperar a que todos los cálculos de la red estén completos.

b) Agrupamiento

Hoy en día, el algoritmo más utilizado para RR.BB. generales es la propagación en árbol de cliques (P.A.C.) [135, 122]. Sin embargo, el C.L. es en general más eficiente que la P.A.C., incluso sin considerar la triangulación. De hecho, para bucles aislados ambos métodos tienen el mismo orden de complejidad (sec. 6.2.3), pero el C.L. requiere menos computación. Para bucles anidados o adyacentes, la complejidad de la P.A.C. puede crecer exponencial (por culpa de la triangulación basada en búsqueda de máxima cardinalidad), mientras que con el C.L. aumenta linealmente cuando un nodo puede romper varios bucles.

Hay otros métodos de agrupamiento, tales como los propuestos en [174, pp. 197–204] y [200], que, al menos en algunos casos, pueden ser más eficientes que el C.L. Desgracidamente, estas técnicas no explican cómo escoger los cúmulos que puedan ahorrar computación en un caso general. La dificultad reside en encontrar la mejor triangulación, lo cual es un problema NP [253].[13] Ya hemos visto que la búsqueda de máxima cardinalidad puede llevar a una propagación de evidencia muy ineficiente. Hay métodos heurísticos [127] o basados en simulación [129]. En cualquier caso, hace falta un tiempo adicional para compilar el árbol de cúmulos a partir de la R.B., operación que no es necesaria en el C.L.

Es muy interesante el trabajo de Shachter et al. [200] en que muestran que su algoritmo de agrupamiento puede resultar equivalente a cualquier otro método exacto si se eligen los cúmulos apropiados. En el caso del C.L., la elección consiste en trazar un arco desde cada nodo del conjunto de corte hasta cada nodo del bucle (o de los bucles) que rompe y formar luego los cliques. La conclusión que obtenemos es que el algortimo de búsqueda en profundidad que describimos en la sección 6.2.1 está realizando implícitamente una triangulación de la red. En general este método añade más arcos de los necesarios. Este es otro tema abierto de investigación: el tratar de decidir cuáles de estos arcos son necesarios y cuáles no.

En cualquier caso, aunque algún algoritmo de agrupamiento pudiera ser el método más eficiente para calcular la probabilidad en el caso general, el C.L. seguiría teniendo importantes ventajas sobre él:

- El C.L. puede ahorrar muchos cálculos cuando hay bucles desbloqueados. En consecuencia, sigue siendo más eficiente en muchas aplicaciones prácticas (ver la sec. 6.2.4.1).
- En RR.BB. es posible responder a preguntas tales como "¿Cuál es la probabilidad $P((x_2 \vee x_3) \wedge x_6)$?", añadiendo un nodo con dos arcos procedentes de X_2 y X_3 a través de una puerta OR, y un segundo nodo con dos arcos, desde el primero y desde X_6 mediante una puerta AND [174, págs. 223ss]. En general esto ocasiona nuevos buclues que pueden aumentar notablemente la complejidad de la computación, pero al menos el C.L. no necesita una compilación de la red para hallar los cúmulos.

[13] Yannakakis [265] demostró que es un problema NP hallar la triangulación minimal de un grafo. Sin embargo, aquí el problema es diferente, pues debemos tener en cuenta tambíen el número de valores de cada variable.

• Por la misma razón, el C.L. no necesita recompilar la red cada vez que introducimos una pequeña modificación en el modelo. Esto es una ventaja significativa a la hora de depurar una base de conocimiento dada por una R.B. de grandes dimensiones.

• Otra ventaja importante del C.L. es que conserva la belleza de los mensajes π y λ por separado en vez de mezclarlos en funciones de potencial. Esos mensajes constituyen un valioso instrumento a partir del cual generar explicaciones lingüísticas, puesto que ofrecen una visión intuitiva y causal de por qué la probabilidad de $P^*(x)$ ha variado, en función de los cambios en las probabilidades de sus vecinos, mientras que los mensajes que circulan entre los cúmulos no tienen una interpretación sencilla. Según lo que conozco, el condicionamiento local es el único método que se presta a una capacidad de explicación distribuida, basada en el análisis de las π y las λ de cada nodo.

• Por último, en el C.L. es inmediato el tratamiento de las puertas AND y OR. En efecto, el algoritmo que desarrollamos en la sección 5.2.3 es también aplicable a redes con bucles; la única diferencia es que ahora las $Q(x)$ podrán estar condicionadas, dando lugar a expresiones como $Q(x, a)$ y $Q(x|a)$, cuyo tratamiento es idéntico al de los casos anteriores. Ésta es una de las grandes ventajas del C.L. frente a los métodos de agrupamiento, porque el primero permite aprovechar el gran ahorro de cálculo que ofrecen las puertas AND y OR.

6.2.4.3 Otras capacidades y posibles ampliaciones

La esencia del C.L. es que la propagación se desarrolla en paralelo para los valores de ciertas variables en vez de mezclar su efecto mediante un sumatorio. Esta propagación condicionada puede aplicarse a cualquier variable que escojamos, con lo cual obtenemos un *razonamiento por casos* simplemente no sumando sobre las variables cuya cauística nos interesa.

Así, al buscar *la mejor explicación*, el objetivo no es calcular la probabilidad a posteriori de cada variable sino considerar un subconjunto de variables y buscar la asignación de variables que, en conjunto, explique mejor la evidencia observada [174, cap. 5]. En la sección 6.2.2 vimos cómo obtener la distribución marginal $P(x, y, \ldots)$. Por tanto, encontrar la mejor explicación es tan sencillo como seleccionar previamente las variables que han de componer la explicación con el fin de no sumar sobre ellas al propagar la evidencia. El único problema de razonar por casos o buscar la mejor explicación es que la complejidad crece exponencialmente con el número de variables seleccionadas.

La propagación en condicionada puede aplicarse también para *predecir el efecto de una acción* sobre la evolución del sistema (fig. 6.8). Por ejemplo, en una R.B. médica podemos incluir un nodo-acción CIRUGIA que indique la posibilidad de operar al paciente. Podemos aplicar una función de evaluación al ESTADO FUTURO del paciente para tratar de predecir si el estado del paciente mejorará con la operación.

No tiene sentido sumar los valores sobre tal variable-acción, puesto que no se trata de una variable aleatoria con una probabilidad a priori, sino que la decisión está bajo control humano. No nos interesa conocer la probabilidad de que suceda, sólo queremos examinar su impacto sobre los nodos que están por debajo (sus efectos). Observar que una diferencia importante entre el nodo ESTADO FUTURO y los nodos terminales correspondientes a posibles hallazgos es que el valor del primero nunca va a ser conocido al realizar la propagación; por tanto el efecto de un nodo acción se

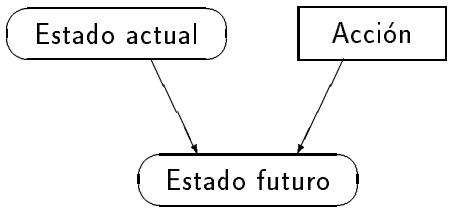


Figura 6.8: Inclusión de un nodo-acción.

propagara únicamente hacia abajo y el condicionamiento sobre esta variable afectará únicamente a sus descendientes, lo cual concuerda con nuestro sentido común.

De este modo, es inmediato ampliar el condicionamiento local de RR.BB. a diagramas de influencia (ver la sección 10.1.3), sin tener que añadir la modificación que Cooper propuso para otros algoritmos [52].

La técnica de propagación condicional es útil también para determinar la sensibilidad y especificidad de un enlace causal, definidas en la sección 4.1. En un poliárbol no es necesario aplicar condicionamiento porque la interacción entre una causa y un efecto inmediato tiene lugar solamente a través de su enlace, pero en un bucle como el de la figura 6.4.*a* la interacción entre B y D a través de B–A–C–D podría ser tan importante como la anterior, debido a la correlación que establece entre las dos causas de D.

Apéndice: Complejidad de algoritmos

La complejidad de los algoritmos es uno de los temas más interesantes de la teoría de la computación, no sólo por su interés conceptual, sino también por sus repercusiones prácticas.[14]

Los dos aspectos fundamentales de la complejidad son el *espacio* que ocupa en memoria la ejecución de un algoritmo y el *tiempo* que tarda en completarse, expresados en función del tamaño del problema. Más concretamente, supongamos que en cierto lenguaje de computación hacen falta n bits para codificar el enunciado del problema. Se dice que la complejidad temporal [o espacial] del algoritmo es polinómica si existen dos constantes c y k tales que el algoritmo necesita, cualquiera que sea el problema concreto planteado, un tiempo [o un espacio de memoria] menor o igual que cn^k. Si el número de operaciones necesarias en el peor de los casos es ck^n, la complejidad del algoritmo es exponencial.

La *complejidad espacial* de un problema siempre es mayor para una implementación recursiva que para una versión iterativa. Los algoritmos exactos que hemos presentado en este capítulo vienen dados por expresiones recursivas; sin embargo, también es posible elaborar una implementación iterativa basada en el recuento de los mensajes recibidos, tal como explicábamos al tratar la implementación distribuida (sec. 5.1.4).

[14] Un libro sencillo y ameno pero riguroso, que nos ha resultado muy útil, es el de Wilf [260]. Para un tratamiento matemático más formal, ver [115].

El tema de la *complejidad temporal* no es tan sencillo. Vimos en la sección 6.2.3.1 que la adición de un solo enlace en una red arbitrariamente grande puede duplicar el tiempo de computación, independientemente del algoritmo exacto empleado. Esto esta relacionado con el hecho de que la computación de la probabilidad en redes bayesianas es un problema NP-completo, según demostró Cooper [54]. Recientemente se ha demostrado que también la computación por simulación estocástica es NP-completa. Veamos qué significa esto.

Problemas NP-completos

Además de los cinco problemas NP-completos citados en este capítulo, existen numerosos ejemplos que se han ido descubriendo, tanto dentro como fuera del campo de las redes bayesianas. Algunos casos célebres son los siguientes:

- *Satisfabilidad:* Dada una expresión booleana, tal como "$(A_1 \wedge A_2 \wedge A_3) \vee (A_4 \wedge A_5)$", ¿es posible encontrar una asignación de variables (por ejemplo, $\{A_4 = +a_4, A_5 = \neg a_5\}$, tal que se satisfaga la expresión? Aunque limitemos a tres el número máximo de variables, el problema —conocido entonces como *3SAT*— sigue siendo NP-completo.

- *Programación lineal entera:* Dada una mátriz A de dimensión $n \times m$ y un vector columna $\mathbf{b}$ de n enteros, ¿existe un vector columna $\mathbf{x}$ de n enteros tal que $A\mathbf{x} \geq \mathbf{b}$?

- *Número cromático:* Dado un grafo y un número entero k, ¿es posible pintar el grafo con k colores de modo que no haya dos nodos adyacentes del mismo color¿

- *Circuito de Hamilton:* Dado un grafo, ¿existe un circuito que visite una y sólo una vez cada nodo?

- *Problema del viajante:* Dados un conjunto de puntos en el plano (las ciudades) y una longitud l, ¿existe un recorrido de longitud menor que l que visite todos los puntos?

En estos ejemplos se intuye una de las características esenciales de los problemas NP:[15] cuando la respuesta es afirmativa, hay una forma sencilla (en el sentido de que requiere poca computación) de demostrar que ésa es la respuesta correcta; en nuestros ejemplos, basta mostrar un caso (una asignación, un vector, etc.) que cumpla la condición exigida.

Ahora bien, aunque sea sencillo comprobar que la respuesta es afirmativa cuando nos ofrecen una demostración, es muy difícil encontrar la respuesta correcta cuando sólo conocemos el enunciado. "Difícil" significa aquí que el tiempo de computación necesario es muy elevado. En realidad, todos los algoritmos que se conocen son de complejidad exponencial. (Algunos de estos algoritmos tienen complejidad menor en casos particulares, e incluso en el promedio, pero en el caso más desfavorable su complejidad es exponencial.)

[15] Hemos presentado aquí ejemplos de *problemas de decisión*, en que la respuesta es "sí" o "no". Cada uno de ellos lleva asociado un problema cualitativo equivalente: hallar una asignación de variables, un vector $\mathbf{x}$, el número cromático, un circuito de Hamilton o el recorrido más corto. En cuanto a la complejidad de los algoritmos, es prácticamente equivalente discutir el problema de decisión o su versión cualitativa.

Más aún, los problemas NP-completos están relacionados entre sí de tal manera que, si existiera un algoritmo de complejidad polinómica para *uno* de ellos, dicho algoritmo podría adaptarse para cualquiera de los otros problemas de la familia, de modo que *todos* ellos tendrían también complejidad polinómica. Reciprocamente, si no existe un algoritmo eficiente para uno de los problemas, tampoco existirá para ninguno de los demás.

Aún no se ha demostrado si la complejidad de la familia NP es polinómica —ésta es la cuestión pendiente más importante de la teoría de la computación—, pero la insistencia con que se han buscado algoritmos para los ejemplos citados nos hace ser pesimistas frente a la posibilidad de que sea polinómica, y más bien parece ser exponencial.

Como consecuecia, existe el convencimiento unánime de que va a ser imposible encontrar un algoritmo eficiente para redes bayesianas de gran tamaño. De forma más gráfica podríamos decirlo así: dado un sistema experto bayesiano, al añadir un pequeño número de nodos o enlaces a la red, el tiempo de computación podría pasar, por ejemplo, de 3 segundos a 40 minutos, con lo cual el programa se volvería prácticamente inútil. Por tanto, para la futura aplicación de las RR.BB. a problemas complejos, se hace necesario introducir aproximaciones que simplifiquen los cálculos. Existe ya un cierto número de propuestas, y en la sección 15.4.1 ofreceremos nuestras propias sugerencias.

Capítulo 7

Aprendizaje en redes bayesianas

El aprendizaje es una de las cuestiones fundamentales de la inteligencia artificial. Podemos definirlo como "el proceso por el que un sistema mejora su rendimiento". Esta definición tan amplia incluye muchas formas posibles de aprendizaje, que pueden ser clasificadas con criterios muy diferentes [193].

Por otra parte, vamos a distinguir el *aprendizaje automático* (*machine learning*) de la *adquisición del conocimiento* (*knowledge acquisition* o *knowledge elicitation*). El aprendizaje automático consiste en obtener el conocimiento a partir de la experiencia o de cierta información previamente almacenada (base de datos, base de casos o conjunto de entrenamiento). La adquisición del conocimiento, en cambio, se centra en la interacción con un experto humano en orden a obtener una descripción computable del conocimiento que éste posee. En este capítulo nos vamos a centrar en el aprendizaje en RR.BB. y trataremos la cuestión de la adquisición de conocimiento en la sección 11.3.

7.1 Nociones básicas

7.1.1 Tipos de aprendizaje

A diferencia de otras especialidades de la inteligencia artificial, en nuestro campo el problema del aprendizaje está bien definido, y se puede plantear así: el objetivo consiste en construir la R.B. que, basándose en los casos observados anteriormente, dé el mejor diagnóstico y la mejor predicción en los casos futuros. Además, la teoría bayesiana concreta aún más el problema: buscamos la R.B. más probable dados los casos observados.

El primer paso en la construcción de una red bayesiana consiste en elegir las variables correspodientes a las magnitudes y conceptos que queremos representar. Después determinaremos las relaciones cualitativas entre ellas trazando los enlaces causales y, finalmente, incluiremos la información cualitativa en forma de probabilidades condicionales.

La situación óptima se daría si tuviéramos una base de datos con un gran número de casos, en que cada variable tuviera un valor determinado. Para esa situación ideal, existen algoritmos que hallan tanto la estructura como los parámetros de la red. El proceso se conoce como *aprendizaje por lotes* (*batch learning*).

Sin embargo, en muchos casos no tenemos disponible una base de datos de tales

características, por lo que debemos recurrir a especialistas que, haciendo uso de su experiencia y de la literatura publicada en su campo, nos dirán cómo construir la red. En este caso, suele ser relativamente sencillo construir la estructura de la red; la dificultad consiste en obtener las probabilidades condicionales. Lo más que podemos conseguir son números aproximados que deben corregirse a medida que el sistema está funcionando. Este proceso, conocido como *aprendizaje secuencial* (*sequential learning*), es en realidad un ajuste de parámetros.

7.1.2 Crítica del aprendizaje por lotes

El primer trabajo sobre aprendizaje de poliárboles a partir de bases de datos fue realizado por Rebane and Pearl [185]. Después aparecieron métodos similares para la construcción de RR.BB. generales; entre ellos destaca el trabajo de Cooper y Herskovits [55, 56, 113] por la influencia que ha ejercido y sigue ejerciendo sobre otros investigadores. En la actualidad podemos detectar un gran interés por el aprendizaje de RR.BB. a partir de bases de datos, con la publicación de numerosos artículos sobre este tema (se pueden encontrar varios ejemplos en [102] y en un libro de reciente publicación [218]). Quizá sea la razón el auge que han experimentado en los últimos las redes neurales, cuya virtud principal consiste en el aprendizaje a partir de un conjunto de ejemplos. Sin embargo, frente a este entusiasmo, quisiéramos expresar nuestras reticencias acerca del aprendizaje por lotes.

En primer lugar, comprobamos que la mayoría de estos algoritmos suponen que tenemos una base de datos en que está determinado el valor de cada variable para cada caso, pues de otro modo el marco matemático se complica tanto que todavía no existe una teoría para abordar el problema. Sin embargo, en la práctica existen muy pocas bases de datos que cumplan ese requisito.

Por otra parte, hay que tener en cuenta que no todas las variables que intervienen en nuestro problema son observables (si todas lo fueran, el diagnóstico sería trivial y no necesitaríamos una R.B.). Por eso, aunque tengamos una base de datos completa, debemos tomarla con sentido crítico, recordando que el valor almacenado para una variable no observable (o no observada) no se obtuvo directamente sino que fue inferido a partir de los valores de otras variables, por lo que estamos en una situación similar al caso de una base de datos incompleta.

Es cierto que existen algoritmos, como el de Pearl y Verma [178], capaces de encontrar variables ocultas examinando las correlaciones entre los datos. Aun así, sigue habiendo dos inconvenientes. El primero es que hace falta una cantidad muy grande de datos para que los resultados sean fiables; el problema se agrava cuando aumenta la proporción de variables ocultas frente a variables registradas. Y el segundo inconveniente es que puede ocurrir que las variables "descubiertas" no correspondan a ningún concepto que podamos identificar; por ejemplo, que no correspondan a ninguna enfermedad o anomalía, si se trata de una base de datos médica.

Estas dificultades están más acusadas en el campo de la medicina por el hecho de que a los especialistas sólo les interesa almacenar en sus bases de datos unos pocos datos observados, junto con el diagnóstico final, mientras que la construcción de una R.B. requiere especificar numerosas variables intermedias necesarias para que tengan validez las hipótesis de independencia. Por eso, la base de datos que resulta satisfactoria para un médico está muy lejos de ser tan detallada y completa como requiere la metodología de las RR.BB.

Estas razones, fundamentalmente la ausencia de una base de datos completa sobre valvulopatías cardíacas, nos ha llevado a estudiar modelos de aprendizaje secuencial, en que más que tratar de obtener la red de forma completamente automatizada, la construiremos con la ayuda de los cardiólogos y dotaremos al sistema de un mecanismo automático que ajuste los parámetros del modelo. No pretendemos afirmar que el aprendizaje secuencial resuelva todas las dificultades mencionadas anteriormente, ni mucho menos. Cuando tenemos información incompleta, el problema del aprendizaje se vuelve realmente difícil de tratar matemáticamente, tanto para un método como para el otro.

7.2 Modelo de aprendizaje secuencial

7.2.1 Introducción

El trabajo más importante sobre aprendizaje secuencial fue realizado por Spiegelhalter y Lauritzen [217], quienes, ante las dificultades que presenta el aprendizaje por lotes, sostienen que la estructura de la red debe mantenerse fija y que el aprendizaje automático debe limitarse al ajuste de parámetros. Con el fin de simplificar el problema hasta hacerlo tratable, su modelo introduce algunas hipótesis, fundamentalmente la independencia local y global de los parámetros (en seguida explicaremos en qué consisten).

Cada vez que la red examina un nuevo caso, los parámetros Θ se actualizan. Por desgracia, si el nuevo caso es incompleto, es decir, si hay variables que no tienen asignado un valor, generalmente aparecen dependencias entre los parámetros. Sin embargo, la única solución tratable conocida consiste en ignorar tales dependencias y trabajar con una distribución de probabilidad actualizada para cada familia, permitiendo así un tratamiento secuencial de los datos.

Por último, queda el problema de la representación de las tablas de probabilidad condicional $P(x|pa(x), \bar{\theta}_X)$; Spiegelhalter y Lauritzen proponen tres modelos de distribución de probabidad para sus parámetros: discretización de las variables, distribuciones de Dirichlet y distribuciones gaussianas para las disparidades (*log-odds*) relativas a cierto estado de refencia. El segundo modelo se aplica en [163] y [215].

Nosotros vamos a presentar aquí un cuarto modelo, que supone una distribución a priori gaussiana $\mathcal{N}(\mu_i, \sigma_i)$ para cada parámetro Θ_i.

7.2.2 Bases del modelo

Introducimos en primer lugar las hipótesis que constituyen la base de nuestro modelo. Suponemos que hemos diagnosticado ya N casos; cada caso i está dado por un conjunto de hallazgos, la evidencia $\mathbf{e}_i$.

Notación. Siguiendo con la misma convención que utlizábamos para las variables ordinarias, utilizaremos la Θ mayúscula para representar un parámetro, que corresponderá en general a una probabilidad condicionada, y la θ minúscula para representar un valor concreto de dicho parámetro; igualmente, $\bar{\Theta}$ representará un conjunto de parámetros y $\bar{\theta}$ un vector en que cada uno de dichos parámetros toma un valor determinado.

Hipótesis 1 (Independencia de los casos) *Los casos* $\mathbf{e}_i$ *son independientes dados los parámetros:*

$$P(\mathbf{e}_1, \ldots, \mathbf{e}_N|\bar{\theta}) = \prod_{i=1}^{N} P(\mathbf{e}_i|\bar{\theta}) \tag{7.1}$$

Esta independencia condicional parece razonable: la probabilidad de un nuevo caso depende sólo de los parámetros del modelo (las probabilidades condicionales y a priori), no de los casos que hemos encontrado hasta antes ni de los que vamos a encontrar en el futuro. Esta hipótesis es la clave de la actualización secuencial de la probabilidad:

$$P(\bar{\theta}|\mathbf{e}_1, \ldots, \mathbf{e}_N) = \alpha\, P(\mathbf{e}_N|\bar{\theta}) \cdot P(\bar{\theta}|\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) \tag{7.2}$$

donde $\alpha = P(\mathbf{e}_1, \ldots, \mathbf{e}_{N-1})/P(\mathbf{e}_1, \ldots, \mathbf{e}_N)$ es un factor de normalización que no depende de $\bar{\theta}$.

Hipótesis 2 (Independencia de los parámetros) *Todos los parámetros son independientes a priori dos a dos. Es decir,*

$$P(\theta_i|\theta_j) = P(\theta_i), \qquad \forall i, \forall j, i \neq j \tag{7.3}$$

La hipótesis incluye tanto la *independencia local*, es decir, la independencia dentro de cada familia, como la *independencia global* (entre diferentes familias). A diferencia de la hipótesis anterior, que parecía bastante razonable, no está claro que la segunda se cumpla para una elección arbitraria de los parámetros. Spiegelhalter y Lauritzen [217] citan el caso de análisis de pedigríes como ejemplo en que se viola claramente la independencia global.

Ambas hipótesis aparecieron ya en [217]. La característica específica de nuestro modelo es la siguiente:

Hipótesis 3 (Distribuciones normales) *La probabilidad inicial de cada parámetro* Θ_i *viene dada por una districión gaussiana*

$$P(\theta_i|\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) = \mathcal{N}(\mu_i, \sigma_i^2) \tag{7.4}$$

con

$$0 < \mu_i < 1 \tag{7.5}$$

$$\sigma_i \ll \; min(\mu_i, 1 - \mu_i). \tag{7.6}$$

Según esta hipótesis, la probabilidad de Θ_i alcanza cierto valor positivo también fuera del intervalo [0,1]. Sin embargo, en nuestro modelo, tanto para el caso general como para la puerta OR, Θ_i va a representar una probabilidad condicional, por lo que su valor debe estar comprendido siempre entre 0 y 1. Esta contradicción no plantea problemas en la práctica, pues la condición (7.6) nos garantiza que la probabilidad de Θ_i fuera de ese intervalo es tan pequeña que puede considerarse nula. Nosotros trabajaremos con valores medios y varianzas, de modo que, dentro de cierta aproximación, las colas de la distribución no influyen en los resultados. Podríamos dar una definición correcta de $f(\bar{\theta})$, al estilo de la ecuación (7.12), pero eso complicaría la exposición sin aportar nada importante.

Las dos últimas hipótesis pueden fundirse diciendo que

$$P(\bar{\theta}|\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) = f(\bar{\theta}) \tag{7.7}$$

donde f representa el producto de distribuciones normales de una variable, y sustituyendo este resultado en la ecuación (7.2),

$$P(\bar{\theta}|\mathbf{e}_1, \ldots, \mathbf{e}_N) \propto P(\mathbf{e}_N|\bar{\theta})\, f(\bar{\theta}) \tag{7.8}$$

Esta expresión sintetiza las tres hipótesis, y será por tanto el punto de partida de nuestro estudio. Lo que necesitamos ahora es calcular $P(\mathbf{e}_N|\bar{\theta})$.

Para cada nodo X (fig. 7.1), haremos una partición del conjunto de parámetros en tres subconjuntos: $\bar{\Theta}_X$ contiene los parámetros relativos a la familia X, es decir, los que determinan $P(x|u)$; los subconjuntos $\bar{\Theta}_X^-$ y $\bar{\Theta}_X^+$ contienen los que determinan la probabilidad en las familias por debajo de X y *por encima de los padres* de X, respectivamente. Igualmente, $\bar{\Theta}_{U_iX}^+$ representa los parámetros que están por encima del enlace U_iX. Es interesante comparar la figura 7.1 con la 5.1 (pág. 81) para comprobar cómo los mensajes dependen ahora explícitamente de los parámetros.

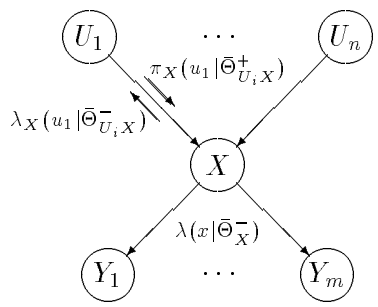


Figura 7.1: Mensajes para la familia X (en función de los parámetros).

Aplicaremos los axiomas de independencia condicional de las redes bayesianas para separar probabilísticamente no sólo las variables —como en el capítulo anterior— sino también los parámetros (que, según [217], podrían tratarse en este sentido como si fueran variables ordinarias):

$$P(\mathbf{e}_N|\bar{\theta}) = \sum_x P(\mathbf{e}_X^-|x, \bar{\theta})\, P(x, \mathbf{e}_X^+|\bar{\theta}) \tag{7.9}$$

Podemos seguir aplicando la separación condicional, con lo que obtenemos

$$\begin{aligned}
P(\mathbf{e}_X^-|x, \bar{\theta}) &= P(\mathbf{e}_X^-|x, \bar{\theta}_X^-) \\
P(x, \mathbf{e}_X^+|\bar{\theta}) &= P(x, \mathbf{e}_X^+|\bar{\theta}_X, \bar{\theta}_X^+) \\
P(x, \bar{u}|\bar{\theta}_X, \bar{\theta}_X^+) &= P(x, \bar{u}|\bar{\theta}_X) \\
P(u_i, \mathbf{e}_X^+|\bar{\theta}_X^+) &= P(u_i, \mathbf{e}_{U_iX}^+|\bar{\theta}_{U_iX}^+)
\end{aligned}$$

y llegamos así a

$$P(\mathbf{e}_N|\bar{\theta}) = \sum_x \lambda(x|\bar{\theta}_X^-)\, \pi(x|\bar{\theta}_X, \bar{\theta}_X^+) \tag{7.10}$$

Las expresiones anteriores son válidas tanto para el caso general (tablas de probabilidad) como para las puertas AND y OR. Estos modelos sólo difieren entre sí en la forma de $P(\mathbf{e}_N|\bar{\theta})$, aunque en todos los casos será una función lineal de los parámetros. Por eso vamos a estudiar en la próxima sección algunas propiedades estadísticas comunes que aplicaremos después a cada modelo particular.

7.2.3 Propiedades estadísticas

Supongamos que tenemos una distribución normal multivariada de variables no correlacionadas, o dicho de otro modo, una distribución dada como producto de distribuciones normales de una variable:

$$f(\bar{\theta}) = \prod_i \mathcal{N}(\mu_i, \sigma_i) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\left(\frac{\theta_i - \mu_i}{\sigma_i}\right)^2\right] \tag{7.11}$$

Supongamos que se cumplen también las condiciones (7.5) y (7.6).

Estudiaremos una nueva distribución definida así:

$$f'(\bar{\theta}) = \begin{cases} c\,(a + \sum_i b_i\theta_i) f(\bar{\theta}) & \text{si } 0 \le \theta_i \le 1, \forall i \\ 0 & \text{en otro caso} \end{cases} \tag{7.12}$$

donde c es una constante de normalización. Podemos suponer, sin pérdida de generalidad, que es positiva. Entonces,

$$a + \sum_i b_i\theta_i \ge 0 \qquad \text{si } \; 0 \le \theta_i \le 1, \forall i \tag{7.13}$$

para que la distribución sea siempre no-negativa.

La condición de normalización de $f'(\bar{\theta})$ implica que, muy aproximadamente,

$$c = a + \sum_i b_i\mu_i. \tag{7.14}$$

(La demostración de ésta ecuación y las siguientes se encuentra en el apéndice.)

Con la misma aproximación, los momentos de la nueva distribución son

$$\mu_i' \equiv E'[\Theta_i] = \mu_i + \Delta_i \tag{7.15}$$

$$\sigma_i'^2 \equiv E'[\Theta_i^2] - \mu_i'^2 = \sigma_i^2 - \Delta_i^2 \tag{7.16}$$

donde

$$\Delta_i \equiv \frac{b_i\,\sigma_i^2}{a + \sum_j b_j\mu_j} \tag{7.17}$$

(Observar que la constante de normalización c es irrelevante. Esto supone una ventaja a la hora de ajustar los parámetros, porque permite que los mensajes vengan multiplicados por constantes.)

La covarianza está dada por

$$\mathrm{cov}(\Theta_i, \Theta_j) = E'[\Theta_i\Theta_j] - \mu_i'\mu_j' = -\Delta_i\Delta_j \tag{7.18}$$

Aparentemente, la ecuación (7.16) podría dar un valor negativo para $\sigma_i'^2$, sin embargo, en el apartado E del apéndice se demuestra que $\sigma_i'^2 > 0$, como era de esperar. También se puede ver, a partir de las ecuaciones (7.16) y (7.17), que la desviación estándar se reduce cuando $b_i \neq 0$. Esta propiedad es importante para el ajuste de parámetros.

7.2.4 El caso general

Intentaremos hallar el valor más probable de los parámetros de la familia X. Para ello vamos a aplicar las hipótesis anteriores

$$\begin{aligned} P(\bar{\theta}_X|\mathbf{e}_1,\ldots,\mathbf{e}_N) & \\ &= \int_{\bar{\Theta}_X^+}\int_{\bar{\Theta}_X^-} P(\bar{\theta}|\mathbf{e}_1,\ldots,\mathbf{e}_N)\, d\bar{\theta}_X^+\, d\bar{\theta}_X^- \\ &\propto \int_{\bar{\Theta}_X^+}\int_{\bar{\Theta}_X^-} P(\mathbf{e}_N|\bar{\theta})\, f(\bar{\theta}_X^+)\, f(\bar{\theta}_X)\, f(\bar{\theta}_X^-)\, d\bar{\theta}_X^+\, d\bar{\theta}_X^- \end{aligned} \tag{7.19}$$

Podríamos dar un parámetro para cada asignación de valores de la familia X: $P(x|\bar{u},\bar{\theta}_X)$. Pero las probabilidades para los valores de X dada una configuración $\bar{u}$ deben sumar la unidad, de modo que habría una ligadura entre ellos. En su lugar, utilizaremos los siguientes parámetros independientes:

$$P(x|\bar{u},\bar{\theta}_X) = \begin{cases} \theta_x^{\bar{u}} & \text{para } x \neq x_0 \\ 1 - \sum\limits_{x'\neq x_0} \theta_{x'}^{\bar{u}} & \text{para } x = x_0 \end{cases} \tag{7.20}$$

Con estos parámetros, la ecuación (7.10) se convierte en

$$\begin{aligned} P(\mathbf{e}_N|\bar{\theta}) &= \sum_x \lambda(x|\bar{\theta}_X^-) \sum_{\bar{u}} P(x|\bar{u},\bar{\theta}_X) \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) \\ &= \lambda(x_0|\bar{\theta}_X^-) \sum_{\bar{u}} \left[1 - \sum_{x\neq x_0} \theta_x^{\bar{u}}\right] \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) + \\ &\qquad + \sum_{x\neq x_0} \lambda(x|\bar{\theta}_X^-) \sum_{\bar{u}} \theta_x^{\bar{u}} \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) \\ &= \lambda(x_0|\bar{\theta}_X^-) \prod_i \left(\sum_{u_i} \pi_X(u_i|\bar{\theta}_{U_i X}^+)\right) + \\ &\qquad + \sum_{x\neq x_0} [\lambda(x|\bar{\theta}_X^-) - \lambda(x_0|\bar{\theta}_X^-)] \sum_{\bar{u}} \theta_x^{\bar{u}} \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) \end{aligned} \tag{7.21}$$

Sustituyendo este resultado en la ecuación (7.19) y realizando la integración llegamos a

$$\begin{aligned} P(\bar{\theta}_X|\mathbf{e}_1,\ldots,\mathbf{e}_N) \propto \Bigg[& \hat{\lambda}(x_0) \prod_i \left(\sum_{u_i} \hat{\pi}_X(u_i)\right) + \\ & + \sum_{x\neq x_0} [\hat{\lambda}(x) - \hat{\lambda}(x_0)] \sum_{\bar{u}} \theta_x^{\bar{u}} \prod_i \hat{\pi}_X(u_i) \Bigg] f(\bar{\theta}_X) \end{aligned} \tag{7.22}$$

Aquí hemos introducido el valor esperado de varias funciones; por ejemplo, recordando que las distribuciones gaussianas $f(\bar{\theta})$ están normalizadas,

$$\begin{aligned} \hat{\lambda}(x) \equiv E[\lambda(x|\bar{\Theta}_X^-)] &= \int \lambda(x|\bar{\theta}_X^-)\, f(\bar{\theta})\, d\bar{\theta} \\ &= \iiint \lambda(x|\bar{\theta}_X^-)\, f(\bar{\theta}_X^+)\, f(\bar{\theta}_X)\, f(\bar{\theta}_X^-)\, d\bar{\theta}_X^+\, d\bar{\theta}_X\, d\bar{\theta}_X^- \\ &= \int \lambda(x|\bar{\theta}_X^-)\, f(\bar{\theta}_X^-)\, d\bar{\theta}_X^- \end{aligned}$$

e igualmente

$$\hat{\pi}_X(u_i) \equiv E[\pi_X(u_i|\bar{\Theta}^+_{U_iX})] = \int \pi_X(u_i|\bar{\theta}^+_{U_iX})\, f(\bar{\theta}^+_{U_iX})\, d\bar{\theta}^+_{U_iX}$$

La ecuación (7.22) puede expresarse como

$$P(\bar{\theta}_X|\mathbf{e}_1,\ldots,\mathbf{e}_N) = \left[a + \sum_{\bar{u}} \sum_{x\neq x_0} b_x^{\bar{u}}\theta_x^{\bar{u}}\right] f(\bar{\theta}_X) \tag{7.23}$$

con

$$a \equiv \hat{\lambda}(x_0) \prod_i \left(\sum_{u_i} \hat{\pi}_X(u_i)\right) \tag{7.24}$$

$$b_x^{\bar{u}} \equiv [\hat{\lambda}(x) - \hat{\lambda}(x_0)] \prod_i \hat{\pi}_X(u_i) \tag{7.25}$$

Por comparación con los resultados de la sección 7.2.3, y teniendo en cuenta que

$$a + \sum_{\bar{u}} \sum_{x} b_x^{\bar{u}}\mu_x^{\bar{u}} = \sum_x \hat{\lambda}(x)\,\hat{\pi}(x) \tag{7.26}$$

(ver la demostración en el apartado F del apéndice), obtenemos finalmente la expresión que buscábamos:

$$\Delta_x^{\bar{u}} = (\sigma_x^{\bar{u}})^2 \frac{[\hat{\lambda}(x) - \hat{\lambda}(x_0)] \prod_i \hat{\pi}_X(u_i)}{\sum_{x'} \hat{\lambda}(x')\,\hat{\pi}(x')} \tag{7.27}$$

En conclusión: la nueva distribución para el parámetro $\Theta_x^{\bar{u}}$ ha cambiado su valor medio de $\mu_x^{\bar{u}}$ a $\mu_x^{\bar{u}} + \Delta_x^{\bar{u}}$ y su varianza de $(\sigma_x^{\bar{u}})^2$ a $(\sigma_x^{\bar{u}})^2 - (\Delta_x^{\bar{u}})^2$. Sin embargo, ahora la distribución marginal para cada parámetro ya no es gaussiana y, además, aparecen generalmente correlaciones cuando el caso considerado no es completo, es decir, cuando la evidencia observada no es suficiente para determinar exactamente los valores de todas las variables.

A pesar de todo, para tener un modelo tratable, *supondremos* que la nueva distribución puede aproximarse por un producto de distribuciones gaussianas univaluadas,

$$P'(x|u,\bar{\theta}_X) = \begin{cases} \mathcal{N}(\mu_x^{\bar{u}} + \Delta_x^{\bar{u}}, \sigma_x^{\bar{u}2} - \Delta_x^{\bar{u}2}) & \text{para } x \neq x_0 \\ 1 - \sum_{x^*\neq x_0} P'(x^*|u,\bar{\theta}_x) & \text{para } x = x_0 \end{cases} \tag{7.28}$$

con el fin de recuperar las hipótesis de partida y poder tratar el caso $\mathbf{e}_{N+1}$ del mismo modo que $\mathbf{e}_N$, teniendo así un método secuencial de aprendizaje.

Comentarios

• La aproximación de la ecuación (7.28) es válida cuando $\Delta_x^{\bar{u}}$ puede considerarse pequeña comparada con $\min(\mu_x^{\bar{u}}, 1 - \mu_x^{\bar{u}})$. De otro modo, la distribución resultante se alejará bastante de una gaussiana y, además, las correlaciones dadas por la ecuación (7.18) no podrían ser despreciadas si las varianzas son amplias y los valores

observados eran poco probables a priori. Por tanto, las aproximaciones que utiliza este método sólo son justificables cuando las $\sigma_x^{\bar{u}}$ son pequeñas, es decir, cuando el modelo original es relativamente preciso.

• Los mensajes $\hat{\lambda}(x)$, $\hat{\pi}_X(u_i)$ y $\hat{\pi}(x)$ pueden obtenerse de forma local, es decir, considerando los mensajes que ha recibido el nodo X y los parámetros de su familia. Esta propiedad es consecuencia de la hipótesis de *independencia global* y es la base de la capacidad de aprendizaje distribuido (ver la sección 7.2.6).

• La separación condicional de los parámetros y el hecho de que sean independientes a priori, hace que el cálculo de los nuevos mensajes sea realmente sencillo. Por ejemplo, a partir de

$$\pi(x|\bar{\theta}_X, \bar{\theta}_X^+) = \sum_{\bar{u}} P(x|\bar{u}, \bar{\theta}_X) \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) \tag{7.29}$$

teniendo en cuenta que los conjuntos $\bar{\Theta}_X$ y los $\bar{\Theta}_{U_i X}^+$ son disjuntos dos a dos, podemos integrar por separado para llegar a

$$\begin{aligned}
\hat{\pi}(x) &= E[\pi(x|\bar{\Theta}_X, \bar{\Theta}_X^+)] = E\left[\sum_{\bar{u}} P(x|\bar{u}, \bar{\Theta}_X) \prod_i \pi_X(u_i|\bar{\Theta}_{U_i X}^+)\right] \\
&= \sum_{\bar{u}} E[P(x|\bar{u}, \bar{\Theta}_X)] \prod_i E[\pi_X(u_i|\bar{\Theta}_{U_i X}^+)] \\
&= \sum_{\bar{u}} \hat{P}(x|\bar{u}) \prod_i \hat{\pi}_X(u_i)
\end{aligned} \tag{7.30}$$

con

$$\hat{P}(x|\bar{u}) \equiv \begin{cases} \mu_x^{\bar{u}} & \text{para } x \neq x_0 \\ 1 - \sum\limits_{x' \neq x_0} \mu_{x'}^{\bar{u}} & \text{para } x = x_0 \end{cases} \tag{7.31}$$

La ecuación (7.31) para $\hat{P}(x|\bar{u})$ es equivalente a la ec. (7.20) para $P(x|u, \bar{\theta}_x)$. La única diferencia es que ahora hay que tomar valores medios en vez de la distribución original. Lo mismo ocurre para $\hat{\pi}(x)$ al comparar (7.30) con (7.29). Por tanto, la propagación de evidencia en este modelo es formalmente equivalente al caso "tradicional" si empleamos los valores medios como si se trataran de probabilidades condicionales exactamente determinadas. En otras palabras, para la propagación de no hace falta tener en cuenta las distribuciones correspondientes a los parámetros: tomamos el valor medio de cada uno de ellos y despreciamos —de momento— su desviación estándar.

• Podemos ver que los resultados matemáticos coinciden con lo que predice el sentido común. Por ejemplo, la ecuación (7.27) afirma que $\Delta_x^{\bar{u}} = 0$ cuando $\sigma_x^{\bar{u}} = 0$, por lo que un parámetro $\Theta_x^{\bar{u}}$ determinado con exactitud no sufrirá ajustes, como era de esperar.

• También observamos que $\Delta_x^{\bar{u}} = 0$ cuando $\hat{\lambda}(x) = \hat{\lambda}(x_0)$, lo cual significa que los parámetros de una familia no se modifican hasta que llegue evidencia (a favor o en contra) desde alguno de sus efectos, como es natural. En el caso de que $\hat{\lambda}(x) \neq \hat{\lambda}(x_0)$ y $\sigma_x^{\bar{u}} \neq 0$, entonces $\Delta_x^{\bar{u}} \neq 0$, al menos para ciertos valores de x y $\bar{u}$. De acuerdo con la ecuación (7.16), la desviación estándar de un parámetro se reduce cada vez que se observa evidencia para el estado de configuración (asignación de valores) correspondiente.

• Podemos tratar la probabilidad a priori de un nodo sin padres como si fuera una probabilidad condicional ordinaria, añadiendo un nodo ficticio que represente una

variable binaria cuyo valor es siempre **presente**. Con este artificio, todos los nodos de la red reciben un tratamiento homogéneo.

• Las ecuaciones que hemos deducido siguen siendo válidas si aplicamos condicionamiento local, pues el formalismo que hemos desarrollado no varía aunque los mensajes estén condicionados. Por ejemplo, para la red de la figura 6.4 (pág. 109) el ajuste de los parámetros de la familia B viene dado por

$$\Delta_b^{v,a} = (\sigma_b^{v,a})^2 \, \frac{[\hat{\lambda}(b|a) - \hat{\lambda}(b_0|a)] \, \hat{\pi}_B(a) \, \hat{\pi}_B(v)}{\sum_{b'} \sum_{a'} \hat{\lambda}(b'|a') \, \hat{\pi}(b', a')}$$

y el de la familia D por

$$\Delta_d^{b,c,f} = (\sigma_d^{b,c,f})^2 \, \frac{[\hat{\lambda}(d) - \hat{\lambda}(d_0)] \sum_a \hat{\pi}_D(b,a) \, \hat{\pi}_D(c|a) \, \hat{\pi}_D(f)}{\sum_{d'} \hat{\lambda}(d') \, \hat{\pi}(d')}$$

Los mensajes que aparecen en estas expresiones son los que obtuvimos en la sección 6.2.2, con la única diferencia de que allí tomábamos el valor exacto y ahora tomamos el valor medio. Vemos así que el condicionamiento local nos permite extender el modelo de aprendizaje para aplicarlo a las redes con bucles con la misma facilidad que a los poliárboles.

• Por último, una ventaja que presenta este modelo de aprendizaje es que permite explicar por qué se ha modificado el valor de cierto parámetro. Así, para el ejemplo 3 del capítulo 4 (pág. 62) podría generar un mensaje como éste: "La probabilidad del paludismo para la zona de origen u_1^+ y el grupo sanguíneo $u_2^\ddagger$ ha descendido de 0'022 a 0'02198 porque hemos encontrado un paciente de esas características en que la prueba de la gota gruesa ha dado un resultado negativo".

7.2.5 La puerta OR

Vimos en el capítulo anterior las numerosas ventajas que presenta la puerta OR en la construcción de redes bayesianas. Sin embargo, el modelo de aprendizaje que hemos estudiado para el caso general no es directamente aplicable porque ésta no utiliza como parámetros las probabilidades condicionales $P(x|\bar{u})$, sino un conjunto mucho más reducido, que introdujimos en la sección 5.2.2. El conjunto de parámetros correspondiente a la familia X se descompone en varios subconjuntos, uno por cada enlace:

$$\bar{\Theta}_X = \bigcup_i \bar{\Theta}_X^{U_i} \tag{7.32}$$

y cada subconjunto $\bar{\Theta}_X^{U_i}$ está formado por varios parámetros,

$$\bar{\Theta}_X^{U_i} \equiv \{\Theta_x^{u_i}\} \tag{7.33}$$

de modo que $\Theta_x^{u_i} \equiv c_x^{u_i}$.

Por tanto, dedicaremos esta sección a adaptar el algoritmo de aprendizaje a la puerta OR, siguiendo un desarrollo similar al del caso general. Tomamos como punto

de partida la ecuación (7.10). Para simplificar las expresiones, empezamos definiendo $R_{U_i}(x)$:

$$R_{U_i}(x) \equiv \begin{cases} [\lambda(x) - \lambda(x+1)] \prod_{j \neq i} Q_{U_j}(x) & \text{para } x < g_X \\ \lambda(x) \prod_{j \neq i} Q_{U_j}(x) & \text{para } x = g_X \end{cases} \tag{7.34}$$

En el caso de que las probabilidades condicionales vengan dadas por distribuciones parametrizadas debemos realizar las siguientes sustituciones:

$$\left.\begin{array}{lcl} \lambda(x) & \rightarrow & \lambda(x|\bar{\theta}_X^-) \\ Q_{U_i}(x) & \rightarrow & Q_{U_i}(x|\bar{\theta}_X^{U_i}, \bar{\theta}_{U_iX}^+) \\ R_{U_i}(x) & \rightarrow & R_{U_i}(x|\bar{\theta}_X^-, \bar{\theta}_X^{U_j}, \bar{\theta}_{U_jX}^+ \; j \neq i) \end{array}\right\} \tag{7.35}$$

En el apartado G del apéndice se demuestra que

$$P(\mathbf{e}_N|\bar{\theta}) = \sum_x R_{U_i}(x|\bar{\theta}_X^-, \bar{\theta}_X^{U_j}, \bar{\theta}_{U_jX}^+ \; j \neq i) \cdot \\ \cdot \left[\sum_{u_i} \pi_X(u_i|\bar{\theta}_{U_iX}^+) - \sum_{u_i} \sum_{x'=x+1}^{g_X} c_{x'}^{u_i} \pi_X(u_i|\bar{\theta}_{U_iX}^+) \right] \tag{7.36}$$

Podemos ver que R_{U_i} no comparte ningún parámetro con los términos que se encuentran dentro del corchete. Igualmente, el conjunto $\bar{\Theta}_X^{U_i} = \{c_x^{u_i}\}$ tampoco comparte ningún parámetro con $\bar{\Theta}_{U_iX}^+$. Por tanto, realizamos una integración similar a la de (7.19) y así obtenemos

$$P(\bar{\theta}_X^{U_i}|\mathbf{e}_1, \ldots, \mathbf{e}_N) = \left[\sum_x \hat{R}_{U_i}(x) \cdot \sum_{u_i} \hat{\pi}_X(u_i) - \right. \\ \left. - \sum_x \sum_{u_i} \sum_{x'=x+1}^{g_X} \hat{R}_{U_i}(x)\, \hat{\pi}_X(u_i)\, c_{x'}^{u_i} \right] f(\bar{\theta}_X^{U_i}) \tag{7.37}$$

Ahora bien, en esta expresión estamos sumando sobre x y sobre x' con la condición de que $x' > x$, lo cual es obviamente igual a sumar sobre x' y sobre x con $x < x'$. Por tanto, podemos reescribir la expresión anterior como

$$\begin{aligned} P(\bar{\theta}_X^{U_i}|\mathbf{e}_1, \ldots, \mathbf{e}_N) &= \left[\sum_x \hat{R}_{U_i}(x) \cdot \sum_{u_i} \hat{\pi}_X(u_i) - \right. \\ &\quad \left. - \sum_{u_i} \sum_{x'} \sum_{x=0}^{x'-1} \hat{R}_{U_i}(x) \hat{\pi}_X(u_i) c_{x'}^{u_i} \right] f(\bar{\theta}_X^{U_i}) \\ &= \left(a + \sum_{u_i} \sum_x b_x^{u_i}\, c_x^{u_i} \right) f(\bar{\theta}_X^{U_i}) \end{aligned} \tag{7.38}$$

con

$$a \equiv \sum_x \hat{R}_{U_i}(x) \tag{7.39}$$

$$b_x^{u_i} \equiv \hat{\pi}_X(u_i) \sum_{x'=0}^{x-1} \hat{R}_{U_i}(x') \tag{7.40}$$

De nuevo podemos aplicar el estudio realizado en la sección 7.2.3, teniendo en cuenta que también ahora

$$a + \sum_{u_i} \sum_{x} b_x^{u_i} \mu_x^{u_i} = \sum_{x} \hat{\lambda}(x)\, \hat{\pi}(x) \tag{7.41}$$

(la demostración está en el apartado H del apéndice). Si recordamos además que $\mu_x^{u_i} = \hat{c}_x^{u_i}$, obtenemos como resultado final

$$\Delta_x^{u_i} = \frac{-\,(\sigma_x^{u_i})^2\, \hat{\pi}_X(u_i) \sum_{x'=0}^{x-1} \hat{R}_{U_i}(x')}{\sum_{x'} \hat{\lambda}(x')\, \hat{\pi}(x')} \tag{7.42}$$

En el caso de variables binarias, $g_{U_i} = g_X = 1$, y sólo hay un parámetro $\Theta_X^{U_i}$ para cada enlace $U_i X$. La ecuación (7.34) nos dice que

$$\hat{R}_{U_i}(X=0) = [\hat{\lambda}(X=0) - \hat{\lambda}(X=1)] \prod_{j \neq i} \hat{Q}_{U_j}(X=0) \tag{7.43}$$

Utilizando la notación $\lambda_x = \hat{\lambda}(X=x)$, $\pi_x = \hat{\pi}(X=x)$ y $\pi_X^{U_i} = \hat{\pi}_X(U_i=1)$, el resultado se simplifica y queda

$$\Delta_X^{U_i} = \frac{(\sigma_X^{U_i})^2\, \pi_X^{U_i}\, [\lambda_1 - \lambda_0] \prod_{j \neq i} \hat{Q}_{U_j}(X=0)}{\lambda_0\, \pi_0 + \lambda_1\, \pi_1} \tag{7.44}$$

Comentarios

• Podemos repetir aquí los mismos comentarios que para el caso general. También en la puerta OR el condicionamiento local nos permite extender el modelo de aprendizaje para aplicarlo a las redes con bucles.

• Además observamos que, según esta última ecuación, cuando $\pi_X^{U_i} = 1$ (es decir, cuando sabemos con certeza que U_i está presente), la evidencia a favor de X ($\lambda_1 > \lambda_0$) hace que $\Delta_X^{U_i}$ sea positivo, mientras que $\Delta_X^{U_i}$ es negativo para $\lambda_1 < \lambda_0$. Éste era el resultado esperado, puesto que el parámetro $c_X^{U_i}$ representa la probabilidad de que U_i en solitario produzca X.

7.2.6 Implementación distribuida

El atractivo del algoritmo de aprendizaje que hemos presentado es que la distribución $P(\bar{\theta})$ de los parámetros que determinan la probabilidad condicional de la familia de X, $P(x|\bar{u}, \bar{\theta}_X)$, puede almacenarse en el nodo X —mediante las $\mu_x^{\bar{u}}$ y las $\sigma_x^{\bar{u}}$— y la actualización de la distribución se basa solamente en la probabilidad a posteriori de esta familia tras la propagación de la evidencia. En otras palabras, el almacenamiento y la actualización secuencial de los parámetros pueden realizarse de forma completamente distribuida.

La figura 7.2 muestra el proceso de aprendizaje distribuido correspondiente al nodo X. El esquema debería incluirse dentro de la figura 5.3 para indicar explícitamente

la íntima unión que existe entre la propagación de evidencia y el aprendizaje, aunque esto complicaría el esquema y se perdería en claridad. Observar que todas las entradas que aparecen en la ecuación (7.27) proceden de nodos vecinos de X, por lo que el proceso es totalmente local.

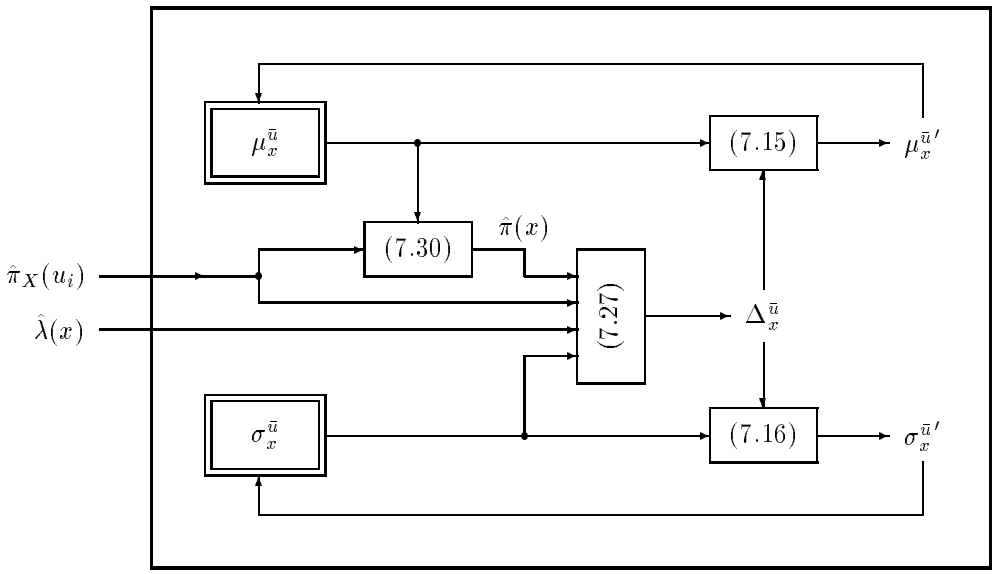


Figura 7.2: Aprendizaje en el nodo X.

A diferencia de las figuras anteriores correspondientes al cómputo de la probabilidad, en la figura 7.2 encontramos el bucle de realimentación típico de los procesos de aprendizaje.

Otra propiedad importante del algoritmo de ajuste de parámetros es que la red está aprendiendo a medida que realiza el diagnóstico. Ésta es una característica esencial de la verdadera computación neuronal, aunque muchos de los modelos de neuronas artificiales, especialmente los basados en retropropagación (*backpropagation*), separan la fase de entrenamiento de la fase de ejecución. Sin embargo, nuestro modelo de aprendizaje integra de forma natural ambos procesos.

Antes de concluir esta sección hay que señalar que las propiedades de computación local y de integración diagnóstico-aprendizaje no son exclusivas de nuestro modelo, sino que se dan también en otros modelos de aprendizaje secuencial en redes bayesianas, en particular a los de Spiegelhalter y Lauritzen que mencionamos al principio de este capítulo.

7.2.7 Discusión

Nuestro modelo de aprendizaje parte de un postulado sencillo: suponemos que cada parámetro tiene una distribución normal dada por su media μ y su varianza σ. Examinamos un nuevo caso $\mathbf{e}_N$ y queremos ver cómo se modifica la distribución. Los cálculos

matemáticos son un poco complicados, debido más a la notación que a los cálculos en sí, pero concluyen con una fórmula concisa para calcular la Δ que nos dice cómo se modifican la media y la varianza.

El algoritmo resultante es por tanto muy simple. Tomamos el valor medio de cada parámetro como si fuera un valor exacto y realizamos la propagación con la evidencia disponible. Cuando el caso está cerrado, es decir, cuando no vamos a recibir más información, utilizamos los mensajes λ y π que ya hemos obtenido y con ellos calculamos la Δ correspondiente a cada parámetro, para ajustar su media y su distribución. La complejidad del algoritmo para un poliárbol es, por tanto, proporcional al número de parámetros; para una red con bucles, el algoritmo necesita realizar más operaciones, pues los mensajes dependen de varias variables y hay que realizar las sumas oportunas sobre las variables que no nos interesan, del mismo modo que sumábamos para obtener las probabilidades a posteriori.

Ventajas

El aprendizaje secuencial, al igual que el aprendizaje por lotes, sigue siendo tan sólo una solución aproximada. Sin embargo, su principal ventaja frente a aquél es que puede utilizarse también en problemas en que no existe una base de datos disponible. En la práctica, este método se aplica construyendo una R.B. con la información facilitada por el experto humano y dejando que el sistema ajuste automáticamente los parámetros según el algoritmo que hemos desarrollado.

Este capítulo presenta un cuarto modelo de distribución para los parámetros, distinto de los tres que propusieron Lauritzen y Spiegelhalter [217]. Una de las ventajas de nuestro modelo es que almacena un valor medio para cada parámetro, que podemos utilizar para propagar la evidencia, mientras que en otros modelos hay que realizar alguna operación previa para obtener este valor a partir de la descripción de las distribuciones. Por otra parte, el utilizar distribuciones gaussianas permite integrar fácilmente tanto estimaciones subjetivas ("la probabilidad es del 60 o del 70%") como resultados experimentales (U produce X en el $67\pm4\%$ de los pacientes) o análisis estadísticos sobre una base de datos.

Otra ventaja importante de nuestro modelo frente a otros anteriores es que puede adaptarse también a la puerta OR, como ya hemos visto, para aprovechar todas las ventajas que ésta presenta (ver la sec. 5.2).

Limitaciones

Una de las objeciones contra el aprendizaje secuencial (cualquiera que sea el modelo escogido para representar los parámetros) procede de las fuertes hipótesis de independencia global y local (ver sec. 7.2.2) cuya validez es difícil de garantizar. Incluso en el caso de que las hipótesis sean ciertas para el modelo inicial, éste ha de ser relativamente exacto, la hipótesis ya no serán válidas, ni siquiera de forma aproximada en cuanto empecemos a ajustar los parámetros en función de los nuevos casos. La crítica es que si tenemos ya un modelo relativamente exacto no es muy importante ajustarlo, y si no lo es, el método de aprendizaje resulta poco fiable. La respuesta a esta objeción es que puede haber modelos con bastante imprecisión (por ejemplo, $\sigma \approx 0'10$), en los que las aproximaciones del modelo siguen siendo razonables.

Otro de los inconvenientes del aprendizaje secuencial es que cuando no disponemos de experimentos o bases de datos que nos den la desviación típica a la vez que la media, resulta muy difícil determinar la varianza a partir de una estimación subjetiva. Si a los expertos humanos ya les cuesta asignar un número a una probabilidad, les resulta mucho más difícil aún decir con qué margen de error la asignan. Por otra parte, la cuestión es importante, pues el valor de σ es crucial en el ajuste de parámetros (ecs. (7.27) y (7.42)). En resumen, éste es un problema abierto que requiere más investigación.

Podemos decir en conclusión que el problema del aprendizaje en RR.BB. está definido con precisión, pero no está resuelto cuando los casos son incompletos, por lo que los métodos de aprendizaje, tanto por lotes como secuencial, se limitan a buscar aproximaciones razonables.

Apéndice: Demostraciones

A Demostración de la ecuación (7.14)

En primer lugar, vemos que para una distribución gaussiana univariada

$$f(\theta_i) = \frac{1}{\sqrt{2\,\pi}\,\sigma_i}\,\exp\left[-\left(\frac{\theta_i - \mu_i}{\sigma_i}\right)^2\right]$$

se cumplen las siguientes propiedades

$$\begin{aligned}
\int f(\theta_i)\,d\theta_i &= 1 \\
\int \theta_i\,f(\theta_i)\,d\theta_i &= \mu_i \\
\int \theta_i^2\,f(\theta_i)\,d\theta_i &= \mu_i^2 + \sigma_i^2 \\
\int \theta_i^3\,f(\theta_i)\,d\theta_i &= \mu_i^3 + 3\,\mu_i\,\sigma_i^2
\end{aligned}$$

La primera de ellas es simplemente la condición de normalización. Las demás se deducen de la definición de $f(\theta_i)$ haciendo el cambio de variable necesario en cada caso.

La ecuación (7.14) se demuestra imponiendo la condición de normalización a $f'(\bar{\theta})$:

$$\int f'(\bar{\theta})\,d\bar{\theta} = 1$$

donde $d\bar{\theta} = d\theta_1 \cdot \ldots \cdot d\theta_{N_\Theta}$ y la integral se extiende sobre todo el espacio $\Re^{N_\Theta}$. Sin embargo, la propiedad (7.6) nos asegura que $f(\bar{\theta}) \approx 0$ fuera del intervalo $[0,1]^{N_\Theta}$, de modo que

$$\begin{aligned}
\int f'(\bar{\theta})d\bar{\theta} &= \int_{[0,1]^{N_\Theta}} c\,(a + \sum_i b_i\,\theta_i)\,f(\bar{\theta})\,d\bar{\theta} \\
&\approx \int_{\Re^{N_\Theta}} c\,(a + \sum_i b_i\,\theta_i)\,f(\bar{\theta})\,d\bar{\theta} \\
&= c\left(a + \sum_i b_i\,\mu_i\right)
\end{aligned}$$

de donde se deduce que, dentro de la mencionada aproximación,

$$c = \frac{1}{a + \sum_i b_i \, \mu_i}$$

B Demostración de la ecuación (7.15)

En las integrales que aparecen a continuación, cada Θ_i debería variar entre 0 y 1. Sin embargo, vamos a integrar de $-\infty$ a $+\infty$, aplicando la misma aproximación que en el apartado anterior.

$$\begin{aligned}
\mu_i' &= E'[\Theta_i] = \int \theta_i \, c \, (a + \sum_j b_j \, \theta_j) \, f(\bar{\theta}) \, d\bar{\theta} \\
&= \frac{a \, \mu_i + b_i \, (\mu_i^2 + \sigma_i^2) + \mu_i \sum_{j \neq i} b_j \, \mu_j}{a + \sum_j b_j \, \mu_j} \\
&= \frac{a \, \mu_i + \mu_i \sum_j b_j \, \mu_j + b_i \, \sigma_i^2}{a + \sum_j b_j \, \mu_j} \\
&= \mu_i + \frac{b_i \, \sigma_i^2}{a + \sum_j b_j \mu_j} = \mu_i + \Delta_i
\end{aligned}$$

C Demostración de la ecuación (7.16)

$$\begin{aligned}
E'[\Theta_i^2] &= \int \theta_i^2 \, c \, (a + \sum_j b_j \, \theta_j) \, f(\bar{\theta}) \, d\bar{\theta} \\
&= \frac{a \, (\mu_i^2 + \sigma_i^2) + b_i \, (\mu_i^3 + 3 \, \mu_i \, \sigma_i^2) + (\mu_i^2 + \sigma_i^2) \sum_{j \neq i} b_j \, \mu_j}{a + \sum_j b_j \, \mu_j} \\
&= \mu_i^2 + \sigma_i^2 + 2 \, \frac{b_i \, \sigma_i^2}{a + \sum_j b_j \mu_j} \\
&= \mu_i^2 + \sigma_i^2 + 2 \, \mu_i \, \Delta_i \\
\mu_i'^2 &= \mu_i^2 + \Delta_i^2 + 2 \, \mu_i \, \Delta_i
\end{aligned}$$

Uniendo ambos resultados, es inmediato que

$$\sigma_i'^2 \equiv E'[\Theta_i^2] - \mu_i'^2 = \sigma_i^2 - \Delta_i^2$$

D Demostración de la ecuación (7.18)

$$
\begin{aligned}
E'[\Theta_i\Theta_j] &= \int \theta_i\theta_j\, c\,(a + \sum_k b_k\,\theta_k)\, f(\bar{\theta})\, d\bar{\theta} \\
&= \frac{\mu_i\mu_j a + \mu_j b_i\,(\mu_i^2 + \sigma_i^2) + \mu_i b_j\,(\mu_j^2 + \sigma_j^2) + \mu_i\mu_j \sum_{k\neq i, k\neq j} b_k\,\mu_k}{a + \sum_k b_k\,\mu_k} \\
&= \mu_i\mu_j + \mu_j\Delta_i + \Delta_j\mu_i \\
\mu_i'\mu_j' &= \mu_i\mu_j + \mu_i\Delta_j + \Delta_i\mu_j + \Delta_i\Delta_j
\end{aligned}
$$

Por tanto

$$\mathrm{cov}(\Theta_i, \Theta_j) = E'[\Theta_i\Theta_j] - \mu_i'\mu_j' = -\Delta_i\Delta_j$$

E Demostración de $\sigma_i'^2 > 0$

Primero vamos a demostrar, como resultado previo, que

$$|b_i\,\sigma_i| \leq a + \sum_j b_j\,\mu_j$$

Estudiemos dos casos:

a) Caso $b_i \geq 0$. De la ecuación (7.6) deducimos que $\sigma_i < \mu_i$, y por tanto,

$$b_i\,\sigma_i \leq b_i\,\mu_i$$

De la ecuación (7.13), tomando $\Theta_i = 0$ y $\Theta_j = \mu_j$ para $j \neq i$, obtenemos

$$a + \sum_{j\neq i} b_j\,\mu_j \geq 0$$

Sumando estas dos expresiones tenemos el resultado buscado.

b) Caso $b_i < 0$. De la ecuación (7.6) deducimos que $\sigma_i < 1 - \mu_i$; como $-b_i > 0$, tenemos

$$|b_i\,\sigma_i| = -b_i\,\sigma_i < -b_i + b_i\,\mu_i$$

De la ecuación (7.13) deducimos

$$a + b_i + \sum_{j\neq i} b_i\,\mu_i \geq 0$$

o lo que es lo mismo,

$$-b_i \leq a + \sum_{j\neq i} b_i\,\mu_i$$

que sustituido en la expresión anterior completa la demostración del resultado previo.

De él podemos deducir que

$$\left|\frac{b_i\,\sigma_i}{a + \sum_j b_j\,\mu_j}\right| \leq 1$$

y por tanto

$$1 - \left(\frac{b_i\,\sigma_i}{a + \sum_j b_j\,\mu_j}\right)^2 \geq 0$$

con lo que llegamos a

$$\begin{aligned}
\sigma_i'^2 &= \sigma_i^2 - \Delta_i^2 = \sigma_i^2 - \left(\frac{b_i\,\sigma_i^2}{a + \sum_j b_j\,\mu_j}\right)^2 \\
&= \sigma_i^2 \left[1 - \left(\frac{b_i\,\sigma_i}{a + \sum_j b_j\,\mu_j}\right)^2\right] \geq 0
\end{aligned}$$

F Demostración de la ecuación (7.26)

Según las ecuaciones (7.24) y (7.25),

$$\begin{aligned}
a + \sum_{\bar{u}} \sum_x b_x^{\bar{u}} \mu_x^{\bar{u}} &= \sum_x \lambda(x_0|\bar{\theta}_X^-) \sum_{\bar{u}} \prod_i \pi_X(u_i|\bar{\theta}_{U_i X}^+) + \\
&\quad + \sum_{\bar{u}} \sum_{x \neq x_0} [\hat{\lambda}(x) - \hat{\lambda}(x_0)]\, \mu_x^{\bar{u}} \prod_i \hat{\pi}_X(u_i) \\
&= \hat{\lambda}(x_0) \sum_{\bar{u}} \left(1 - \sum_{x \neq x_0} \mu_x^{\bar{u}}\right) \prod_i \hat{\pi}_X(u_i) + \\
&\quad + \sum_{x \neq x_0} \hat{\lambda}(x) \sum_{\bar{u}} \mu_x^{\bar{u}} \prod_i \hat{\pi}_X(u_i)
\end{aligned}$$

Si aplicamos aquí las ecuaciones (7.30) y (7.31) llegamos a

$$\begin{aligned}
a + \sum_{\bar{u}} \sum_x b_x^{\bar{u}} \mu_x^{\bar{u}} &= \sum_x \hat{\lambda}(x) \sum_{\bar{u}} \hat{P}(x|\bar{u}) \prod_i \hat{\pi}_X(u_i) \\
&= \sum_x \hat{\lambda}(x)\, \hat{\pi}(x)
\end{aligned}$$

como queríamos demostrar.

G Demostración de la ecuación (7.36)

Para simplificar la notación, supongamos primero que estamos en el caso "tradicional", en que los parámetros están completamente determinados. Las ecuaciones (5.10) y (5.31) nos dicen que

$$\begin{aligned}
P(\mathbf{e}_N) &= \sum_x \lambda(x)\, \pi(x) \\
&= \lambda(X=0)\, Q(X=0) + \sum_{x \neq 0} \lambda(x)\, [Q(x) - Q(x-1)] \\
&= \sum_x \lambda(x)\, Q(x) - \sum_{x \neq 0} \lambda(x)\, Q(x-1) \\
&= \sum_x \lambda(x)\, Q(x) - \sum_{x \neq g_X} \lambda(x+1)\, Q(x) \\
&= \sum_x \left[\lambda(x) \prod_i Q_{U_i}(x)\right] - \sum_{x \neq g_X} \left[\lambda(x+1) \prod_i Q_{U_i}(x)\right]
\end{aligned}$$

Aplicando la definición de R_{U_i} (ec. (7.34)) podemos simplificar este resultado:

$$P(\mathbf{e}_N) = \sum_x R_{U_i}(x)\, Q_{U_i}(x)$$

y finalmente, recordando la ecuación (5.37)

$$P(\mathbf{e}_N) = \sum_x R_{U_i}(x) \left[\sum_{u_i} \pi_X(u_i) - \sum_{u_i} \left(\pi_X(u_i) \sum_{x'=x+1}^{g_x} c_x^{u_i} \right) \right]$$

Basta ahora condicionar cada mensaje sobre los parámetros correspondientes de acuerdo con (7.35) y se obtiene ya la ecuación (7.36).

H Demostración de la ecuación (7.41)

Según las ecuaciones (7.39) y (7.40) y recordando que $\mu_x^{u_i} = \hat{c}_x^{u_i}$,

$$\begin{aligned} a + \sum_{u_i} \sum_x b_x^{u_i} \mu_x^{u_i} &= \sum_x \bar{R}_{U_i}(x) \sum_{u_i} \hat{\pi}_X(u_i) - \\ &\quad - \sum_{u_i} \sum_x \sum_{x'=0}^{x-1} \hat{R}_{U_i}(x')\, \hat{\pi}_X(u_i)\, \hat{c}_x^{u_i} \end{aligned}$$

Tal como razonábamos en la sección 7.2.5, sumar sobre x y sobre x' con $x' < x$ es lo mismo que sumar sobre x' y x con $x > x'$. Por otra parte, x y x' son variables mudas en los sumatorios, por lo que pueden ser sustituidas o intercambiadas. En consecuencia,

$$\begin{aligned} \sum_{x=0}^{g_X} \sum_{x'=0}^{x-1} \hat{R}_{U_i}(x')\, \hat{c}_x^{u_i} &= \sum_{x'=0}^{g_X} \sum_{x=x'+1}^{g_X} \hat{R}_{U_i}(x')\, \hat{c}_x^{u_i} \\ &= \sum_{x=0}^{g_X} \sum_{x'=x+1}^{g_X} \hat{R}_{U_i}(x)\, \hat{c}_{x'}^{u_i} \end{aligned}$$

En consecuencia,

$$a + \sum_{u_i} \sum_x b_x^{u_i} \mu_x^{u_i} = \sum_x \hat{R}_{U_i}(x) \sum_{u_i} \hat{\pi}_X(u_i) \left[1 - \sum_{x'=x+1}^{g_X} \hat{c}_{x'}^{u_i} \right]$$

Por otro lado, sabíamos que

$$Q_{U_i}(x|\bar{\theta}_X^{U_i}, \bar{\theta}_{U_i X}^{+}) = \sum_{u_i} \pi_X(u_i|\bar{\theta}_{U_i X}^{+}) \left[1 - \sum_{x'=x+1}^{g_X} c_x^{u_i} \right]$$

Como los parámetros $c_x^{u_i}$ no pertenecen al conjunto $\bar{\Theta}_{U_i X}^{+}$, los valores esperados $\hat{c}_x^{u_i}$ y $\hat{\pi}_X(u_i)$ pueden calcularse independientemente, de modo que

$$\hat{Q}_{U_i}(x) = \sum_{u_i} \hat{\pi}_X(u_i) \left[1 - \sum_{x'=x+1}^{g_X} \hat{c}_{x'}^{u_i} \right]$$

Uniendo estos resultados tenemos

$$a + \sum_{u_i} \sum_{x} b_x^{u_i} \mu_x^{u_i} = \sum_{x} \hat{R}_{U_i}(x) \hat{Q}_{U_i}(x)$$

e invirtiendo las operaciones que hicimos en el apartado G, concluimos que

$$a + \sum_{u_i} \sum_{x} b_x^{u_i} \mu_x^{u_i} = \sum_{x} \hat{\lambda}(x)\, \hat{\pi}(x)$$

como queríamos demostrar.

Capítulo 8

Explicación del razonamiento

8.1 Planteamiento del problema

8.1.1 Importancia

Los capítulos anteriores han presentado los aspectos numéricos de las redes bayesianas, básicamente la inferencia probabilística y el aprendizaje. Sin embargo, para que una R.B. pueda considerarse como verdadero sistema experto, hace falta que pueda explicar su proceso de razonamiento.[1] La importancia de esta propiedad se debe principalmente a tres motivos.

Por un lado, durante el desarrollo del programa resulta muy útil que éste explique cómo y por qué ha llegado a una conclusión, pues de otro modo hay muchos errores que pueden pasar inadvertidos, e incluso una vez advertidos pueden resultar muy difíciles de localizar y corregir. En consecuencia, esta capacidad es muy importante para la obtención del conocimiento, la depuración y el mantenimiento del sistema experto.

Por otro lado, la explicación ayuda al usuario a entender cómo funciona el sistema. La transparencia de funcionamiento no sólo hace que el usuario se sienta más satisfecho, sino que de ella puede depender que éste acepte o rechace los resultados obtenidos. Cuando el resultado que aporta la máquina no coincide con lo que el usuario esperaba, conviene que presentar una justificación detallada para que éste pueda reconocer si era un error propio o del sistema. Ya hemos dicho que el objetivo del sistema no es reemplazar al ser humano, sino darle consejos útiles y razonados que faciliten su labor, ya que es éste quien debe asumir finalmente reponsabilidad de las decisiones.

La capacidad de explicación es especialmente necesaria en el campo de la medicina. Desde los primeros sistemas de ayuda al diagnóstico se vio claramente su importancia [95, 44]. Es famoso el estudio de Teach y Shortiliffe [237] en que se demuestra que los médicos son muy reticentes a aceptar el consejo de un ordenador si no pueden confirmar su fundamento.

[1] En los sistemas de diágnostico se habla a veces de *explicación* como un "conjunto de hipótesis capaz de justificar las anomalías observadas" (con este significado utilizamos el término en la pág. 122). En cambio, en este capítulo nos referimos a un concepto diferente: la *explicación del razonamiento* trata de mostrar cómo y por qué el sistema ha llegado a sus conclusiones.

Una tercera ventaja de incluir explicaciones en un sistema es que éste puede ser aprovechado por un usuario poco experto como instrumento de aprendizaje: un estudiante de medicina, por ejemplo, introduce los datos de un paciente real o hipotético y compara su diagnóstico con el que ofrece la máquina. Para que la lección sea útil es necesario que el sistema tutor explique cómo ha llegado al diagnóstico y por qué se diferencia del que ha dado el alumno.

En conclusión, la capacidad de explicación se considera una propiedad tan esencial de los sistemas expertos que, en nuestra opinión, resulta pretencioso considerar como tal a una red bayesiana que carezca de ella, a pesar de que frecuentemente se comete este abuso del término. Por eso me parece oportuno distinguir entre *red bayesiana* y *sistema experto bayesiano*.

8.1.2 Cuestiones básicas

El tema de la explicación es muy complejo y extenso, por lo que puede descomponerse en varias tareas o subcampos. Seguiremos aquí la división que se hizo en el *Seminario de la AAAI sobre Explicación*, en 1988 [255, 256], aunque adaptando las denominaciones:

Adquisición del conocimiento: Los diseñadores de sistemas expertos están cada vez más de acuerdo en que la capacidad de explicación no es algo que se añade al programa en la última fase de su construcción, sino que debe tenerse en cuenta desde el inicio de su desarrollo, y es por tanto inseparable del proceso de obtención del conocimiento. La relación entre estos dos aspectos se da en ambas direcciones: la adquisición del conocimiento debe obtener también la información necesaria para la explicación del razonamiento, y a su vez la capacidad de explicación orienta la obtención del conocimiento una vez realizado el primer prototipo.

Tipos de preguntas: Determinar qué facilidades de explicación va a proporcionar el sistema es lo mismo que definir cómo va a ser la interacción con el usuario. Hay sistemas que presentan toda la explicación como un bloque cuando lo solicita el usuario, y éste se limita, como mucho, a escoger el tipo de explicación. Por ejemplo, el sistema ABEL [167], ofrece toda la explicación a la vez según el nivel elegido por el usuario: clínico, patofisiológico o intermedio.

En la mayor parte de los sistemas expertos, en cambio, el programa responde a las preguntas que se le plantean. Por tanto, una de las primeras cuestiones que debe determinar el diseñador es qué tipo de explicaciones va a ofrecer el sistema y, más concretamente, qué tipos de preguntas va responder. En la sección 8.3, vamos a proponer un modelo de explicación capaz de responder a una serie de preguntas diferente dependiendo de si el foco de atención se sitúa sobre un nodo, un parámetro o un enlace.

Contenido de la explicación: La cuestión principal que se plantea es cómo generar el contenido bruto de la explicación, es decir, cómo relacionar la evidencia observada con las conclusiones, en función de la base de conocimiento. Para ello se pueden seguir dos métodos:

- *Introspección:* En este método, el sistema reconstruye su proceso de inferencia (suele llamarse *traza* al registro de las operaciones realizadas) mostrando, por ejemplo, cómo se han encadenado las reglas o por qué una de ellas no ha llegado a ejecutarse. Éste es el procedimiento habitual en la mayor parte de los sistemas que incluyen capacidad de explicación.
- *Reconstrucción:* Consiste en generar la explicación independientemente del proceso de razonamiento, utilizando conocimiento específico para esta tarea. Más adelante expondremos los argumentos de los partidarios de este método.

Modelo de usuario: Otra de las cuestiones fundamentales es determinar cuáles son los conocimientos que posee el usuario, pues la explicación que necesita un principiante contendrá demasiada información trivial para un experto, mientras que la explicación que satisface a un experto resultará probablemente insuficiente para un principiante. Paris [166] defiende que no sólo el grado de detalle sino también el contenido de la explicación debe adaptarse al usuario.

Presentación: Por último, queda la cuestión de cómo mostrar al usuario la información seleccionada; lo primero que hay que decidir es si la presentación va a ser gráfica, lingüística o mixta. En algunos sistemas el usuario se encuentra limitado a escoger entre unos cuantos patrones de preguntas que el sistema le propone; en otros, tiene la libertad de escribir sus preguntas en lenguaje natural, y existen también modelos de explicación basados en técnicas de hipertexto [153].

8.1.3 Tipos de explicación

Ya hemos dicho que, a diferencia de otros aspectos de las redes bayesianas, aquí no disponemos aún de una teoría establecida sobre la explicación. Mas aún, una cuestión previa para llegar a una teoría sería definir adecuadamente el problema: ¿Qué clase de explicación quiere el usuario? Y después: ¿Qué clase de explicación puede ofrecer una red bayesiana? Suermondt [228, sec. 2.1] presenta algunas consideraciones sobre la explicación desde un punto de vista epistemológico, aunque no son de gran ayuda práctica para el problema que nos ocupa.

Puesto que no disponemos aún de una teoría completa, nos limitaremos a clasificar los tipos de explicación. En primer lugar, para un sistema basado en el conocimiento, y en particular para una red bayesiana, podemos distinguir dos clases:

Explicación de la base de conocimiento: Consiste en describir el modelo causal representado en la red, el cual pretende ser una descripción aproximada de una pequeña parte del mundo real. A veces se la denomina también *explicación estática* [112].

Explicación de la inferencia: Trata de mostrar la relación entre la evidencia introducida y las conclusiones del razonamiento. También se conoce como *explicación dinámica* [112], y puede ser de dos tipos:

- *Explicación de las variaciones.* Toma como punto de partida el orden en que se han introducido en el sistema los diferentes hallazgos, para estudiar

cómo han ido variando las conclusiones; se trata de mostrar, por ejemplo, la evolución en la distribución de probabilidad de una variable o por qué la hipótesis anteriormente más verosímil ha pasado a un segundo o tercer lugar.

- *Explicación de los resultados.* A diferencia de la anterior, este segundo tipo considera la evidencia como un todo, independientemente del momento en que haya introducida, y se interesa sobre todo por la relación entre los hallazgos, las hipótesis intermedias y las conclusiones definitivas.

En redes bayesianas podemos establecer además otra clasificación, ortogonal a la primera, que distingue dos niveles [196]:

Nivel micro: Estudia la propagación local de la evidencia, es decir, en qué medida la distribución de probabilidad de un nodo se ve afectada por las probabilidades de sus vecinos.

Nivel macro: Rastrea las principales líneas de razonamiento sobre la red como un todo o sobre ciertos caminos.

8.2 Métodos previos de explicación

Ya antes de la aparición de los sistemas expertos, algunos programas de ayuda al diagnóstico en medicina incluían una facilidad rudimentaria de explicación basada en lo que Swartout llama *texto enlatado* (*canned text*) [232], que consistía en seleccionar, según la información introducida, algunas frases almacenadas previamente. El problema es que el diseñador del programa tenía que anticipar las posibles respuestas necesarias, y que éstas debían ser modificadas cada vez que se introducían cambios en el programa.

La capacidad de explicación que podemos considerar mínimamente "inteligente" surge con la creación de de los sistemas expertos. Trataremos primero los sistemas basados en búsqueda heurística y en la sección siguiente los sistemas basados en inferencia numérica.

8.2.1 Sistemas heurísticos

El primer sistema experto con capacidad de explicación fue MYCIN. El programa podía responder a dos preguntas: *por qué* (*why*) y *cómo* (*how*). La primera significaba "¿*Por qué* me solicita este dato?". En este caso, el sistema respondía mostrando la regla que estaba tratando de ejecutar, traducida del lenguaje Lisp al inglés, e introduciendo el comando "`WHY`" sucesivamente el usuario podía recorrer la cadena de reglas que el sistema estaba encadenando. La pregunta "`HOW 3.1`" significaba "¿*Cómo* se ha deducido el resultado [3.1]?". El sistema respondía mostrando la regla correspondiente. La ventaja de MYCIN frente a proyectos previos es que generaba la explicación automáticamente a partir de la base de conocimiento. Más adelante, el sistema fue ampliado para que admitiera preguntas en inglés con texto libre.

Las explicaciones generadas por MYCIN resultaban útiles para los diseñadores del programa, pero aún estaban lejos de ser satisfactorias para un usuario normal. El problema es que el sistema explicaba los pasos de su ejecución, pero no sus líneas

de razonamiento. Por eso se desarrolló NEOMYCIN [44, 97], un proyecto que incluía conocimiento causal explícito en forma de meta-reglas (es decir, reglas sobre cómo tratar otras reglas) y una organización jerárquica de las hipótesis. Con ello conseguía explicar no sólo los pasos que estaba siguiendo sino también su *estrategia de razonamiento*. (Éste y otros subproyectos del programa MYCIN relacionados con la capacidad de explicación se encuentran descritos en [26, caps. 17–20].)

Entre las investigaciones que surgieron a partir de MYCIN, tiene especial interés de cara a las redes bayesianas el trabajo de Wallis y Shortliffe [244], quienes diseñaron un modelo de explicación causal basado en una red semántica que habían construido a partir de la base de conocimiento del sistema experto MYCIN. Inspirándose en el trabajo de Jaime Carbonell [31], que había asociado *etiquetas de relevancia* y *distancias* a una red semántica con el fin de crear un sistema tutor, Wallis y Shortliffe asignaron a una medida de *complejidad* a cada concepto y a cada regla (indicando si cuán difícil de comprender puede resultar al usuario) y una medida de *importancia* a cada concepto. Definieron además dos parámetros, la *experiencia* (*expertise*) del usuario, para representar su nivel de conocimientos, y el *detalle*, para indicar si el usuario desea una explicación general o minuciosa. Estos dos parámetros permitían seleccionar de forma dinámica las reglas y los conceptos en función de los conocimientos y los deseos del usuario.

El sistema XPLAIN, de Swartout [232], siguió un método diferente, basado en un *modelo del dominio*, que contenía información descriptiva sobre el campo de aplicación, y en *principios del dominio*, que indicaban cómo razonar refinando las hipótesis de trabajo. La idea más original consistía en usar un programador automático para generar el programa de consulta a partir de esta información, y los pasos seguidos en este proceso constituían la base de la explicación.

Weiner [247] diseñó BLAH como sistema capaz de sintetizar toda una línea de razonamiento en una sola afirmación, de localizar y desplazar de forma adecuada el foco de atención y de tener en cuenta las expectativas del usuario. La forma de estructurar la explicación se basa en estudios psicológicos sobre el comportamiento humano; por ejemplo, se ha observado que las personas, a la hora de justificar una elección entre dos alternativas, explican antes las diferencias que las semejanzas, y así es como estructura su explicación el sistema BLAH.

En la sección anterior, al discutir cómo generar el contenido de la explicación, dijimos que hay dos métodos fundamentales. Los sistemas mencionados hasta ahora se basan en el primero de ellos, la *introspección*, que consiste en examinar las operaciones realizadas. Sin embargo, otros autores, como Wick y Thomson [258], sostienen que la explicación es un problema diferente de la inferencia y requiere, por tanto, un conocimiento específico. Por eso proponen la *explicación reconstructiva* como proceso prácticamente independiente de la inferencia. Sus argumentos principales son que los pasos que sigue un sistema experto pueden resultar incomprensibles para un usuario normal y que los expertos humanos, a la hora de explicar sus conclusiones, siguen un proceso diferente del que utilizaron para llegar a ellas. En cambio, los detractores de este método señalan el peligro de que la explicación reconstructiva no refleje fielmente el razonamiento del sistema. (Para conocer más detalles sobre esta discusión, ver [255].) Todos los argumentos son válidos, y no es necesario tomar una postura extrema: existen muchas soluciones intermedias posibles dependiendo del grado de

acoplamiento entre la inferencia y el método de explicación.

Hemos citado aquí los sistemas más conocidos que incluyen capacidad de explicación, destacando aquellos aspectos más relacionados con el modelo que vamos a presentar en este capítulo. Para una revisión más completa, ver [35], [228, sec. 2.2] y [257].

8.2.2 Sistemas probabilísticos

Los primeros sistemas bayesianos de diagnóstico se basaban en el método probabilístico clásico, tal como fue presentado en la sección 2.3.1. Para un cierto conjunto de hallazgos $\bar{e} = \{e_1, \ldots, e_n\}$, la probabilidad de una hipótesis d_i frente a otra d_j viene dada por la ecuación (2.5), que en forma logarítmica puede expresarse como

$$\log \frac{P(d_i|\bar{e})}{P(d_j|\bar{e})} = \log \frac{P(d_i)}{P(d_j)} + \sum_{k=1}^{n} \frac{P(s_k|d_i)}{P(s_k|d_j)}$$

El primer término del segundo miembro es independiente de la evidencia observada. Son, por tanto, los términos del sumatorio los que aumentan o disminuyen la probabilidad de d_i frente a d_j. Resulta así muy sencillo averiguar cuáles son los datos que más han contribuido a favor o en contra de una determinada hipótesis. Éste es básicamente el método utilizado en el sistema MEDAS de Ben-Bassat et al. [14], en el sistema GLADYS de Spiegelhalter and Knill-Jones [216] y en el sistema para la clasificación de apoplejías de Reggia and Perricone [186].

Barth y Norton [12] propusieron un modelo de razonamiento bayesiano no estructurado y simétrico (a diferencia de las redes bayesianas, que poseen una estructura asimétrica). Es posible definir una medida del efecto que un determinado hallazgo E ejerce sobre cierta hipótesis H, a partir de la variación en las probabilidades de E y H y de la correlación entre ellas [161]. Se consigue así una capacidad de explicación local. Si el diseñador del programa introduce además una estructura entre las variables, el sistema es capaz de buscar cadenas de influencia para llegar a una explicación global.

Por último, antes de centrarnos en las redes bayesianas, citaremos el trabajo de Strat y Lowrance [224, 225], que genera explicaciones en un sistema de razonamiento basado en la teoría de Dempster-Shafer (ver [68, 201] y la sec. 2.3.4). Aunque se discute hasta qué punto dicha teoría puede considerarse una extensión de la inferencia probabilística, coincide con ella al menos en utilizar una técnica puramente numérica y, en consecuencia, la explicación presenta esencialmente las mismas dificultades en ambos modelos. El método de Strat y Lowrance consiste en realizar un *análisis de sensibilidad* para ver en qué medida contribuye cada hallazgo al apoyo (*support*), plausibilidad, especificidad y consonancia de cada hipótesis. Los resultados pueden presentarse al usuario en lenguaje natural o en forma gráfica.

Explicación en redes bayesianas

Dentro de los sistemas probabilísticos, vamos a centrarnos en las redes bayesianas, que constituyen el objeto de esta tesis. En el nivel micro, Elsaesser y Henrion [84, 85] describen cómo pasar de la información cuantitativa a la descripción cualitativa; por

ejemplo, cómo pasar de valores numéricos de probabilidad a etiquetas lingüísticas como "probable" o "muy probable". También dentro del nivel micro, Sember y Zuckerman [197] abordan un problema diferente: cómo explicar los cambios en la probabilidad a posteriori de un nodo mediante el análisis de los mensajes π y λ. (Naturalmente, su estudio se limita a poliárboles, pues antes del condicionamiento local no era posible aplicar dicho método a redes con bucles.)

Otro trabajo muy interesante relacionado con el tema que estamos tratando es el de Jensen et al. [121] sobre la detección de conflictos en redes bayesianas. Para un conjunto de hallazgos $\bar{S}$, definen una *medida del conflicto* como

$$\mathrm{conf}(s_1, \ldots, s_n) = \log_2 \frac{P(s_1) \cdot \ldots \cdot P(s_n)}{P(s_1, \ldots, s_n)}$$

La expresión anterior trata de reflejar la idea de que la medida del conflicto debe aumentar cuando la probabilidad de que los hallazgos se presenten de forma conjunta es mucho menor que la probabilidad de aparezcan independientemente. Este método tiene la ventaja de el conflicto local se calcula a partir de medidas locales, con lo cual es posible tratar de descubrir a qué se debe. Un inconveniente es que, al basarse en el algoritmo de propagación en árbol de cliques, la medida local no refleja la estructura de la red original sino la de este árbol. Aun así, el método parece prometedor, al menos para abordar uno de los aspectos de la explicación.

Dentro del *nivel macro* hay que citar el sistema NESTOR, presentado por Greg Cooper [50, 51] en 1984, que puede considerarse como una de las primeras redes bayesianas, aunque es algo más general en el sentido de que las probabilidades que definen el modelo pueden venir dadas por intervalos en vez de estar determinadas con exactitud. NESTOR fue pionero sobre todo en cuanto a la capacidad de explicación, en que ofrecía dos posibilidades. Por un lado, para cada hipótesis podemos pedir al sistema que muestre, en forma de texto, cuáles son las cadenas que relacionan dicha hipótesis con los hallazgos que constituyen la evidencia. Por otra parte, también es posible seleccionar dos hipótesis y ver (numérica y gráficamente) cómo los hallazgos van modificando sucesivamente las probabilidades de cada una de ellas.

Tenemos que esperar ocho años, hasta 1992, para ver un nuevo sistema con capacidad de generar macroexplicaciones. En la tesis de Jaap Suermondt [228], dirigida por Cooper, se presenta la metodología INSITE, que consiste en medir la influencia de la evidencia observada, con la posibilidad de examinar cadenas de nodos particulares. Como medida del impacto existen varias funciones posibles; por defecto, el sistema toma la entropía cruzada (*cross-entropy*).

Hay otros dos métodos, desarrollados por Druzdzel y Henrion. El primero de ellos [78, 77] se basa en el concepto de *escenario*, definido como la asignación de valores para las variables (discretas) de un subconjunto. El método de explicación consiste en seleccionar las variables relevantes y hallar los escenarios más probables (por ejemplo, mediante el algoritmo de Sy [233]). Después hay que ordenar las variables y unirlas con frases para traducir cada escenario en una historia causal de cómo se han producido los hechos [77, sec. 8.3.3].

El otro método que proponen Druzdzel y Henrion [112, 77], se basa en las *redes cualitativas* desarrolladas por Wellman [249, 250], que mantienen los mismos axiomas de independencia que las redes bayesianas pero, como su nombre indica, no utilizan

información numérica sino solamente relaciones cualitativas entre variables graduadas. La influencia de un nodo sobre otro puede ser positiva (si hace aumentar la probabilidad de los valores más altos), negativa, nula o ambigua (desconocida). La puerta OR desempeña un importante papel en este modelo [251]. La inferencia puede realizarse de forma local para cada nodo, de modo la complejidad del algoritmo es proporcional al número de nodos y de hallazgos. El método de explicación se basa en examinar el impacto de la propagación de evidencia entre nodos vecinos. El resultado es una cadena de enlaces entre las dos variables de interés, en que cada eslabón se explica dependiendo de si la influencia es positiva o negativa y del tipo de interacción: predictiva (de causa a efecto), diagnóstica (de efecto a causa) o intercausal (entre dos causas de un efecto) [77, sec. 8.3.2]. El método está aún en fase de desarrollo [79].

Algunos de los ejemplos citados aquí y otros más se encuentran descritos en las tesis doctorales de Jaap Suermondt [228, págs. 42–72] y Marek Druzdzel [77, sec. 3.3].

8.3 Modelo de explicación

Después de definir los conceptos básicos sobre explicación y de haber revisado brevemente algunos de los sistemas conocidos, tanto en el campo heurístico como en el probabilístico, vamos a dedicar la esta sección a describir un nuevo modelo de explicación, desarrollado en la presente tesis doctoral.

Antes de pasar adelante, debemos señalar que, al contrario del modelo de aprendizaje descrito en el capítulo anterior, que puede considerarse como un tema cerrado, nuestro modelo de explicación deja abiertas numerosas líneas posibles para futuras extensiones, según comentaremos más adelante.

8.3.1 Clases de enlaces

En los capítulos anteriores hemos tratado las redes bayesianas desde una perspectiva puramente matemática, en que todos los nodos y todos los enlaces eran equivalentes. Sin embargo, a la hora de generar explicaciones es necesario tener en cuenta que unos nodos corresponden a datos cualitativos, otros a datos cuantitativos, a diagnósticos, a hipótesis o estadios intermedios, etc. Por tanto, el punto de partida consiste en distinguir varios tipos de nodos y varias clases de enlaces. A cada elemento de la red (nodo o enlace) le vamos a asignar una etiqueta simbólica, con lo nuestra red bayesiana pasa a ser, además, una red semántica (sec. 2.2.3).

Empecemos recordando que, para una familia (formada por un nodo y sus padres), hay dos tipos de interacción posibles, que fueron definidos en el capítulo 5. En el *caso general*, la probabilidad condicional viene dada por una tabla, $P(x \mid pa(x))$, que representa una matriz de tantas dimensiones como nodos forman la familia. Para la *puerta OR*, la probabilidad condicional se obtiene a partir de los parámetros $c_x^{u_i}$. La diferencia entre ambos modos de interacción no es sólo una cuestión matemática, sino que corresponde a dos interpretaciones semánticas muy diferentes: en el caso general, los padres de X actúan como factores que *influyen* en la probabilidad de X (sec. 4.3); en la puerta OR, los padres de X son las distintas causas que pueden *producir* presencia de X (sec. 5.2.5).[2]

[2]En realidad, existe un tercer modelo, la puerta AND, definida en el mismo capítulo. Sin embargo,

Recordamos también que existe una relación entre el tipo de interacción y las variables que forman la familia. Así, todos los nodos que intervienen en la puerta OR corresponden necesariamente a variables graduadas, según las definiciones introducidas en las sección 5.2. Como consecuencia de estas diferencias, podemos distinguir seis clases de enlaces, dos para el caso general y cuatro para la puerta OR. Las dos primeras son:

Influencia: Es la relación que se da entre dos nodos que interactúan según el modelo general. El nodo padre es en este caso uno de los factores que influyen en la probabilidad del nodo hijo. Por ejemplo, la probabilidad de que el paciente haya sufrido fiebre reumática depende de la edad, del sexo (es más frecuente en las mujeres) y del país de origen (se da con más frecuencia en los países subdesarrollados). Cada uno de los tres enlaces entre el nodo "enfermedad reumática" y sus padres será del tipo *influencia*.

Parámetro: Es el tipo de enlace que existe entre una anomalía y una medida objetiva de su severidad, como ocurre en el caso del área de la válvula mitral (medido en el eco bidimensional), que define el grado de estenosis. Aunque nuestro sistema experto no trabaja directamente con variables continuas, sino que realiza una conversión a variables discretas mediante la definición de intervalos, conviene tener en cuenta que se trata de un tipo particular de variables.

Las cuatro clases de enlaces correspondientes a la puerta OR son las siguientes:

Causalidad: Aunque todos los enlaces de una red bayesiana reflejan el concepto de causalidad, en el sentido amplio del término, nos referimos aquí a la relación "X produce Y". Por ejemplo, la fiebre reumática produce (con una cierta probabilidad) estenosis mitral, la hipertensión pulmonar produce la dilatación de la arteria pulmonar, etc. Ésta es la clase de enlace habitual entre los nodos que corresponden a diagnósticos y a estados fisiológicos.

Tipo: Este enlace indica que el nodo padre es un subtipo del nodo hijo. Así, el enlace que une la estenosis mitral reumática con la estenosis mitral pertenece a esta clase. Una propiedad característica de estos enlaces es que sus parámetros vienen dados por la matriz identidad, es decir,

$$c_x^{u_i} = \begin{cases} 1 & \text{si } u_i = x \\ 0 & \text{si } u_i \neq x \end{cases}$$

lo cual refleja el hecho de que la estenosis mitral reumática leve siempre produce estenosis mitral leve, la moderada produce estenosis mitral moderada, etc.

Manifestación: Esta clase de enlace se parece al de causalidad en que el nodo hijo es siempre un dato cualitativo, aunque ambos nodos corresponden a conceptos distintos. Se diferencia del anterior en que aquí el nodo hijo es más una característica asociada al nodo padre que un efecto producido por él. Por ejemplo,

en la red bayesiana de nuestro sistema experto (sec. 13.1) no hemos encontrado ningún caso en que se presente este tipo de interacción. Además, las ventajas que la puerta OR presenta de cara a la explicación no aparecen en la puerta AND, por lo que en la práctica este modelo puede englobarse dentro del caso general.

la dilatación del anillo mitral es uno de los rasgos que aparecen en el síndrome de Barlow, pero no se considera que el primero sea una *causa* del segundo en el sentido estricto del término.

Observación: En este caso, el nodo padre y el nodo hijo corresponden a un mismo concepto, aunque el segundo se refiere a la observación en una prueba concreta. Así, hay un enlace de este tipo entre la "derrame pericárdico" y la "derrame pericárdico observado en el ecocardiograma". Es importante representar en la red dos nodos diferentes para tener en cuenta que, en general, los métodos de diagnóstico tienen una sensibilidad y una especificidad limitadas, y también para diferenciar entre las posibles observaciones de una misma anomalía; así, siguiendo con el ejemplo anterior, puede haber también otro nodo correspondiente al "derrame pericárdico observado en la radiografía", unido por un enlace de la misma clase.

8.3.2 Opciones para nodos, parámetros y enlaces

Nodos

Vamos a describir primero las opciones de explicación disponibles para los nodos que no corresponden a datos numéricos, pues estos últimos reciben un tratamiento especial en nuestro modelo. Antes de continuar adelante, conviene describir dos propiedades que caracterizan los nodos e influyen en la explicación:

- *Valor normal.* La mayor parte de los nodos tienen un valor que corresponde a la ausencia de anomalías: por ejemplo, "No alcoholismo", "No estenosis", "Tensión arterial normal", "Ritmo sinusal", "Válvula aórtica bicúspide", etc. Hay otros nodos, como el sexo o la edad en que no cabe hablar de valores normales y valores patológicos.

- *Varios tipos de anomalías.* Dentro de las variables que pueden tener valores anormales, puede ocurrir que éstos correspondan a distintos grados de una misma anomalía (la estenosis puede ser leve, moderada o severa) o a varios tipos de anomalías (el ritmo cardíaco puede corresponder a fibrilación auricular, *flutter*, marcapaso, etc.).

Las distintas opciones de explicación que pueden existir para un nodo son las siguientes:

1. *Prevalencia.* Para los nodos que tienen un valor normal, podemos mostrar la prevalencia (la probabilidad a priori) de los valores anómalos; si estos valores corresponden a distintos grados, mostramos además la suma de las probabilidades de todos ellos. Para los nodos que no tienen un valor normal carece de sentido hablar de prevalencia, y sólo podemos mostrar la probabilidad a priori de los distintos valores.

2. *Probabilidad (a posteriori).* Cuando se han introducido ciertos hallazgos en el sistema y se ha realizado la propagación de la evidencia, es posible mostrar la probabilidad a posteriori, según las mismas consideraciones que hicimos para la prevalencia.

3. *Causas o factores.* Para un nodo con padres, podemos mostrar una lista de sus causas (si interactúan mediante la puerta OR) o de los factores que influyen sobre el nodo (en el caso general). Las causas pueden ordenarse según la eficiencia a priori o a posteriori. La eficiencia se define como la probabilidad de que la causa produzca o haya producido la anomalía representada en el nodo.

4. *Probabilidad condicional.* Aparece solamente para los nodos que tienen padres e interactúan mediante el caso general. La explicación consiste en este caso en mostrar la tabla de probabilidad condicional en función de los valores de los nodos.

5. *Efectos.* Para un nodo con hijos, podemos mostrar los efectos que produce, clasificados según el tipo de enlace que los une, y dentro de cada tipo, por el grado de sensibilidad (cuando esté definido). La figura 8.1 muestra una pantalla de explicación tomada del sistema experto DIAVAL, en que aparecen los efectos correspondientes al nodo "Estenosis mitral reumática"; el usuario puede escoger uno de ellos con el fin de obtener más información o volver a la pantalla correspondiente a dicho nodo.

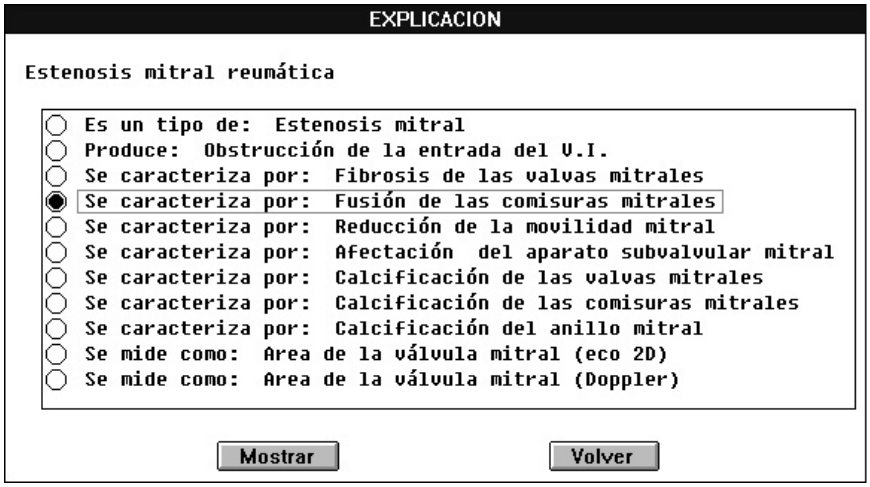


Figura 8.1: Pantalla de explicación para los efectos de un nodo.

6. *Definición.* Consiste en una cadena que podemos asignar al nodo con el fin de ofrecer información adicional al usuario. Puede ser interesante, por ejemplo, en el caso de enfermedades congénitas que aparecen en muy raras ocasiones.

7. *Imagen.* Para los nodos que corresponden a observaciones cualitativas (estamos pensando ya en la ecocardiografía), puede resultar útil al usuario observar una o varias imágenes que le ayuden a distinguir entre la apariencia normal y la patológica.

8. *Pantalla.* En el caso de que el sistema experto cuente con un interfaz gráfico, esta opción permitirá al usuario desplazarse a la pantalla en que aparece reflejado el nodo considerado.

Parámetros

De cara a la explicación, denominamos *parámetros* a los nodos que llevan asociado un dato numérico, que puede proceder de una medida o de un cálcula a partir de algunas medidas. La variable asociada a un nodo-parámetro tomará un valor cuantitativo (en nuestro modelo no hemos considerado variables continuas), que vendrá determinado por el valor numérico.

Las opciones de explicación para un parámetro son las siguientes:

1. *Valor.* Esta opción se presentará como "*Valor medido*" o "*Valor calculado*", dependiendo del tipo de parámetro; muestra al usuario el valor numérico, el valor cualitativo correspondiente y la desviación respecto del valor medio —en caso de que éste se conozca—.

2. *Intervalos.* Se calculan a partir de unos límites que separan cada intervalo del siguiente, permitiendo así discretizar la variable continua con el fin de obtener un valor cualitativo.[3]

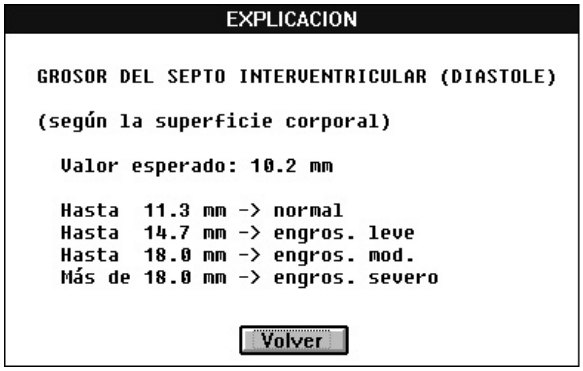


Figura 8.2: Intervalos de un parámetro.

3. *Significado.* En general, un parámetro representa una medida de cierta anomalía. Por ejemplo, el área de la válvula mitral nos da el grado de estenosis. La opción *significado* nos permite pasar de un nodo-parámetro al nodo de la red asociado a la anomalía correspondiente.

4. *Fórmula.* Esta opción aparece solamente en el caso de parámetros calculados a partir de otros.

5. *Imagen.* Al igual que en el caso de los nodos, para algunos parámetros puede ser útil mostrar al usuario una imagen que explique cómo realizar la medida.

6. *Pantalla.* Similar a la opción existente para los nodos.

[3]En la ecocardiografía, hay muchos parámetros que se correlacionan con la superficie corporal. En este caso, los intervalos pueden calcularse con respecto al valor medio de la población general o con respecto a la superficie corporal del paciente considerado. Por tanto, a la hora de generar explicaciones debemos preguntar al usuario qué tipo intervalos desea consultar.

Enlaces

En el capítulo 5 comentábamos que una de las diferencias entre el caso general y la puerta OR es que, en el primero todo el procesamiento se realiza en los nodos, mientras que en la segunda los enlaces cobran una especial importancia, actuando como procesadores capaces de almacenar información y de computar mensajes. Por tanto, para los enlaces correspondientes al caso general (enlaces de las clases influencia y parámetro), no tiene sentido generar explicaciones. Tampoco para los enlaces de la clase tipo, pues su eficiencia (el conjunto de probabilidades condicionales $c_x^{u_i}$) es la matriz unidad. Para estas tres clases de enlaces toda la información relevante se encuentra en los nodos situados en sus extremos.

Para las otras tres clases de enlaces, causal, manifestación y observación, podemos tener las siguientes opciones:

1. *Causa.* Corresponde al nodo que actúa como padre en el enlace.

2. *Efecto.* Es el nodo hijo, es decir, el nodo al que apunta el enlace.

3. *Sensibilidad* y *Especificidad.* Son dos valores numéricos, cuya definición aparece en el apéndice del capítulo 3.

4. *Mecanismo causal.* Es una cadena que el diseñador puede asignar a los enlaces cuando desea explicar al usuario la relación causa-efecto. Por ejemplo, la frase

   > "Cuando el *derrame pericárdico* es severo, el líquido acumulado dificulta la dilatación del corazón; es lo que se conoce como *taponamiento cardíaco*."

   explica el mecanismo causal entre los dos nodos que aparecen en cursiva.

5. *Eficiencia.* Viene dada por las probabilidades condicionales $c_x^{u_i}$.

Navegación por la red

La facilidad de explicación que posee el sistema experto se pondrá en marcha cuando el usuario solicite más información sobre un nodo o un enlace. Al hablar del interfaz gráfico de DIAVAL, veremos que esta operación puede realizarse simplemente pulsando cierto botón cuando el ratón se encuentra sobre la imagen correspondiente al nodo, o abriendo un menú.

Aparecerá entonces una lista de opciones generada dinámicamente dependiendo de cuáles se encuentran disponibles para el nodo o el enlace en cuestión, según hemos explicado anteriormente. Dentro de las opciones existentes, hay algunas que se limitan a presentar cierta información en pantalla, después de lo cual vuelven al punto de partida. Así sucede para la prevalencia de un nodo, para los intervalos de un parámetros o para la sensibilidad de un enlace.

Hay otras opciones, sin embargo, que pueden desplazar el foco de interés de la explicación. Por ejemplo, al mostrar la lista de causas o de factores de un nodo, el usuario podrá seleccionar uno de ellos. Si se trata de un factor, el foco de interés pasa al nodo correspondiente, pues el enlace (de tipo *influencia* o *parámetro*) no contiene información relevante. Si se trata de una causa, el foco de atención se detendrá en el

enlace, de modo que el usuario pueda consultar la sensibilidad, la especificidad, etc. o bien pueda desplazarse a los nodos unidos por el enlace.

De este modo, es posible navegar por la red, ascendiendo y descendiendo por las ramas que la componen y solicitando la información disponible para cada nodo o cada enlace.

Antes de concluir esta sección, conviene aclarar que los ejemplos que hemos escogido para explicar las clases de enlaces están tomados del sistema experto bayesiano para ecocardiografía que describiremos en la parte III. Sin embargo, el modelo de explicación es aplicable a cualquier otro campo, incluso fuera de la medicina, siempre que incluya los dos tipos de interacción y las clases de variables que hemos considerado aquí.

Recordemos también que, al igual que ocurre en otros sistemas expertos, las explicaciones generadas por este método pueden ser útiles tanto para el diseñador del sistema como para el usuario final. Al diseñador le sirve para depurar la base de conocimiento, es decir, para observar con mayor facilidad la información introducida en la red y para investigar las fuentes de errores en caso de que el sistema llegue a conclusiones inaceptables.

8.3.3 Implementación distribuida

En los capítulos anteriores hemos visto cómo los algoritmos presentados admitían implementaciones distribuidas, de modo que los nodos y los enlaces podían procesar información simultáneamente. También para el modelo de explicación que acabamos de presentar es posible tener una implementación en que cada nodo o enlace actúe como procesador, aunque en este caso la discusión tiene menos interés práctico, pues la atención del usuario que solicita explicaciones se centrará en cada instante sobre un solo punto, de modo que no es necesaria la computación paralela.

Estudiemos primero la información estática que debe almacenar cada procesador. Parte de ella coincide con la que requería el sistema con el fin de calcular el impacto de la evidencia: tipo de interacción para cada nodo (general o puerta OR), probabilidades condicionales y a priori, eficiencia de los enlaces, etc. Hay otra información destinada específicamente a la explicación: valor normal para un nodo (si existe y cuál es), tipo de parámetro (medido o calculado), tipo de enlace (influencia, causal, observación, etc.).

Cuando el interés del usuario se centra sobre un nodo o un enlace, éste se activa y responde a los mensajes que recibe, mostrando la información solicitada. Según comentábamos en la sección anterior, hay opciones de explicación en que no varía el foco de atención; mientras sea así, el nodo o enlace sigue activo, hasta que el usuario decida poner fin a la consulta. Para las opciones de explicación que desplazan el foco de interés, el procesador que estaba respondiendo a los mensajes se desactiva, ordenando a su vecino que se active y cediéndole el control de la sesión.

Toda la información que un procesador necesita para responder a las solicitudes del usuario se encuentra contenida en el propio nodo o enlace, o quizá en sus vecinos inmediatos. Por tanto, toda la computación —incluido el control del proceso— se realiza de forma local. La conclusión es que, al igual que en la propagación de evidencia y en el aprendizaje, también para la explicación del razonamiento es posible tener una implementacion totalmente distribuida, en que los nodos y los enlaces

(mejor dicho, ciertas clases de enlaces) actúan como procesadores asíncronos, con lo que es posible tener sistemas expertos en que todas sus operaciones se realizan de modo distribuido con computación local.

8.3.4 Discusión

Vamos a concluir este capítulo situando nuestro modelo dentro de los diferentes tipos de explicación mencionados en la sección 8.1.3, comparándolo con otros sistemas realizados previamente y sugiriendo posibles líneas de desarrollo para el futuro.

En primer lugar, observamos que el método propuesto es capaz de *explicar la base de conocimiento*: causas, efectos, prevalencia (probabilidad a priori), valores para cada nodo, probabilidades condicionales, eficiencia de los enlaces, etc. En este sentido, se diferencia de la mayor parte de los modelos de explicación, tanto de sistemas expertos basados en reglas como de redes bayesianas, que se limitan a explicar sus conclusiones.

En cuanto a la *explicación de la inferencia*, decíamos que puede ser de dos tipos: explicación de las variaciones y explicación de los resultados. La implementación que hemos realizado de este modelo en el sistema experto DIAVAL se ocupa exclusivamente de la explicación de los resultados. Sin embargo, resultaría sencillo extender el modelo de modo que puediera describir las variaciones debidas a la llegada de nueva evidencia: bastaría ir almacenando los resultados obtenidos, con el fin de mostrar posteriormente cómo ha evolucionado la distribución de probabilidad para cada nodo y para cada enlace.

Dentro de los dos niveles de explicación posibles en redes bayesianas, nuestro modelo se sitúa en el *nivel micro*: las explicaciones ofrecidas se refieren a un solo nodo o enlace, sin tener en cuenta la interacción más que con sus vecinos inmediatos. El usuario puede desplazarse por la red, pero en cada instante existe un único foco de atención. En el futuro, sería deseable extender nuestro modelo de modo que pudiera ofrecer también explicaciones en el nivel macro, es decir, que fuera capaz de considerar la red en su conjunto, describiendo las cadenas de propagación de evidencia que justifican las conclusiones obtenidas.

Para ello, resultaría útil aprovechar la experiencia de Wallis y Shortliffe [244], cuyo trabajo hemos mencionado en la sección 8.2.1. Al igual que ellos definían la complejidad e importancia para conceptos y reglas, nosotros podríamos asignar los mismos parámetros para los nodos y enlaces de una red bayesiana, con el fin de generar diferentes explicaciones en función de los conocimientos y los deseos del usuario. Sería interesante también aprovechar las ideas de Weiner [247], mencionadas en la misma sección, con el fin de estructurar la información de modo que se presente al usuario de la forma más clara y compacta posible. Ésta es una línea abierta de investigación que podría llevar al desarrollo de una capacidad de explicación a nivel macro para redes bayesianas, en que se sintetizaran los trabajos sobre explicación en sistemas heurísticos ya citados, los métodos propuestos por Suermondt [228] y Druzdzel [77] para generar macro-explicaciones en redes bayesianas y el modelo de explicación que hemos introducido en este capítulo.

Capítulo 9

Implementación del razonamiento bayesiano

En la primera versión del sistema experto DIAVAL, hemos utilizado una herramienta comercial, *GoldWorks*, que está implementada sobre un dialecto de *Lisp* llamado *Common Lisp*. En este capítulo vamos a describir los aspectos de la implementación del sistema relativos al razonamiento bayesiano. El capítulo 13 es, en cierto sentido, una continuación de éste, pues describe la implementación del sistema experto DIAVAL. La división en dos capítulos se debe a que aquí, dentro de la parte II tratamos los aspectos generales de las redes bayesianas, independientemente del dominio de aplicación, mientras que en la parte III estudiaremos los aspectos concretos de un sistema experto para medicina.

9.1 Características del Lisp

La historia del *Lisp* está íntimamente ligada al surgimiento de la inteligencia artificial. Fue diseñado por John McCarthy en 1958, con el fin de disponer de un lenguaje de programación capaz de tratar símbolos y listas, que admitiera invocación recursiva de funciones [141, 142].

Aunque *Lisp* nació tan sólo dos años después que FORTRAN, su filosofía es completamente diferente. Ambos lenguajes se han mantenido hasta hoy, aunque por razones muy diferentes. FORTRAN sigue utilizándose por la gran cantidad de librerías matemáticas que se han ido creando a lo largo de los años; sin embargo, se trata de un lenguaje anticuado, que se mantiene por inercia. *Lisp*, en cambio, es un lenguaje que goza de excelente salud, especialmente desde la creación de *Common Lisp*, un estándar que permite a un programador ejecutar un mismo programa sobre diferentes implementaciones en máquinas distintas. *Common Lisp* sigue siendo en esencia el mismo lenguaje creado por McCarthy hace 36 años; tan sólo se le han añadido nuevas funciones para facilitar la tarea del programador.

9.1.1 Programación en Lisp

En todos los lenguajes de programación existen varios tipos de datos: caracteres, cadenas, números, etc. En *Lisp*, además de los anteriores, existen dos que son característicos de este lenguaje: las listas y los símbolos.

La esencia del *Lisp* consiste en el tratamiento de listas; de hecho, su nombre procede de *List Processing*. Un programa (la invocación de una función o procedimiento) se expresa en forma de lista. Por ejemplo:[1]

```
(+ (* 2 3) 5)
→ 11
```

El programa anterior viene dado por una lista de tres elementos, el segundo de los cuales es a su vez una lista, que corresponde a la invocación de la función producto. Como se ve en el ejemplo, al evaluar una lista, el primer elemento indica cuál es la función y los demás actúan como argumentos.

Los símbolos son el otro tipo de dato característico de *Lisp*. Un símbolo puede tener asignados un valor (mediante `setq`), una función (mediante `defun`) y una lista de propiedades, de la que no vamos a hablar aquí. Por ejemplo:

```
(setq abc 3)

(defun abc (Numero) (* Numero 2))
```

"`(Numero)`" indica los argumentos de la función y "`(* Numero 2)`" es el *cuerpo* de la función.[2]

Podemos evaluar un símbolo para obtener su valor o podemos utilizarlo para invocar la función que representa:

```
abc
→ 3

(abc 5)
→ 10
```

Cuando un símbolo o una lista van precedidos por una coma, *Lisp* no los evalúa:

```
'abc
→ ABC

'(abc 5)
→ (ABC 5)
```

Existen numerosas funciones en *Common Lisp* para el manejo de listas: `reverse`, `append`, `substitute`, `remove`, `remove-if`, `sort`, etc. Si queremos modificar una lista aplicando una función a cada uno de sus elementos, podemos utilizar `mapcar`:[3]

```
(mapcar #'abc '(2 3 5 8))
→ (4 6 10 16)
```

Imaginemos ahora que, en vez de multiplicar por dos los valores de una lista de números, deseamos definir una función que multiplique los elementos por un factor arbitrario. Una solución posible sería la siguiente:

[1] Representamos con una flecha (→) el resultado de evaluar una expresión.

[2] Utilizamos la convención de representar con minúsculas las funciones y con el primer carácter en mayúscula las variables.

[3] La expresión `#'abc` indica la función correspondiente al símbolo `abc`.

```
(defun multiplicar-lista (Factor Lista-de-numeros)
   (defun multiplicar-elemento (Numero) (* Factor Numero))
   (mapcar #'multiplicar-elemento Lista-de-numeros))
```

Vemos aquí una muestra de la flexibilidad de *Lisp*, que nos permite definir una función dentro de otra función. El resultado es el siguiente:

```
(multiplicar-lista 3 '(1 2 4))
→ (3 6 12)
```

Ésta es una solución válida pero insatisfactoria, porque genera una función que no vamos a utilizar en otro lugar (de hecho, el valor de `multiplicar-elemento` varía cada vez que invocamos `multiplicar-lista` con un factor distinto). Una solución más elegante consistiría en utilizar el *cálculo landa*, una de las características más típicas de *Lisp*:

```
(defun multiplicar-lista (Factor Lista-de-numeros)
   (mapcar #'(lambda (Numero) (* Factor Numero))
           Lista-de-numeros))
```

Aunque es posible programar en *Lisp* utilizando un estilo similar al de otros lenguajes, *Lisp* posee un estilo propio de programación. Por ejemplo, para sumar las raíces cuadradas de varios números podríamos escribir el siguiente programa:

```
(let ((Resultado 0))
   (dolist (Numero '(4 25 36))
      (setq Resultado (+ Resultado (sqrt Numero)))))
→ 13
```

`let` declara una variable; `dolist` asigna sucesivamente a la variable `Numero` cada uno de los elementos de la lista y ejecuta "`(setq ...)`" para cada uno de ellos. Sin embargo, el estilo propio de *Lisp* nos permite resolver este problema en una sola línea, sin utilizar variables intermedias, mediante la función `apply`, que aplica una función a una lista de números:

```
(apply #'+ (mapcar #'sqrt '(4 25 36)))
≡ (apply #'+ '(2 5 6)) ≡ (+ 2 5 6)
→ 13
```

Veamos con otro ejemplo la extraordinaria flexibilidad de *Lisp*. Supongamos que queremos definir una función que tome una lista de funciones y nos muestre el resultado de aplicar cada una de ellas a una lista de números:

```
(defun aplicar-funciones
       (Lista-de-funciones Lista-de-numeros)
   (mapcar #'(lambda (Funcion)
                (apply Funcion Lista-de-numeros))
           Lista-de-funciones))
```

El resultado es el siguiente:

```
(aplicar-funciones '(+ * max rest) '(1 2 3 4 5))
→ (15 120 5 (2 3 4 5))
```

Vemos que `mapcar` ejecuta la expresión `lambda` varias veces, asignando sucesivamente a la variable `Funcion` los distintos valores de la lista `(+ * max rest)`, y formando una nueva lista con los resultados obtenidos.

En otro lenguaje de programación sería muy difícil, si no imposible, realizar operaciones como éstas. La flexibilidad de *Lisp* se debe principalmente a que *tanto los programas como los nombres de las funciones son datos*: los primeros son listas y los segundos son símbolos.

Por simplificar la exposición, hemos utilizado ejemplos numéricos, que a primera vista podrían parecer un tanto rebuscados. Sin embargo, estas posibilidades que ofrece *Lisp* resultan sumamente útiles a la hora de resolver problemas de inteligencia artificial. Por ejemplo, en la sección 9.2.1 hablaremos de los *demonios*, funciones que el diseñador de un sistema experto asigna a un campo de un marco o de una instancia, y que se disparan ante ciertos acontecimientos. Se trata, por tanto, tenemos una lista de funciones (los demonios), desconocida a priori, que *GoldWorks* invocará cuando sea necesario. El mecanismo utilizado será similar al que hemos descrito anteriormente, en que invocábamos una lista de funciones con una misma lista de argumentos.

9.1.2 Entornos de programación

Hemos visto cuáles son algunas de las peculiaridades del *Lisp* en cuanto lenguaje de programación. En *Common Lisp* existen además unas facilidades para la depuración de programas, incluidas en la propia definición del estándar. Los detalles pueden variar de una implementación a otra, pero las características son esencialmente las mismas.

1. `step` es un programa que permite ejecutar una función comprobando paso a paso los resultados que se van obteniendo, así como examinar el estado del sistema en cada instante (asignación de variables, etc.).

2. `trace` selecciona un conjunto de funciones, de modo que *Lisp* emite un mensaje cada vez que el programa entra o sale de alguna de ellas, y muestra el resultado que devuelve cada función. `untrace` deselecciona alguna o todas las funciones anteriores.

3. `debugger` es un programa que se invoca automáticamente cada vez que *Lisp* encuentra un error de ejecución. El programador puede entonces comprobar el estado del sistema para ver qué ha fallado. También es posible invocar esta facilidad de ayuda voluntariamente incluyendo en el código un *punto de ruptura*, mediante la función `break`.

Éstas son las principales ayudas de depuración presentes en toda implementación de *Common Lisp*. Además, cada versión puede incluir otras propias. En particular, *GoldHill Common Lisp*, ofrece dos funciones, `time` y `meter`, para medir el tiempo empleado en ejecución de un programa o en la invocación de las funciones que nos interesen, respectivamente.

Aun con estas facilidades, la mayor ventaja que ofrece *Lisp* para depurar programas es la posibilidad de trabajar indistintamente en modo intérprete o compilado: una función interpretada puede invocar funciones compiladas, y viceversa. Cuando el programador ha comprobado que la función interpretada se comporta correctamente, puede decidir compilarla para mejorar su eficiencia, y continuar así el proceso de forma incremental, con lo cual se ahorra gran cantidad de tiempo.

Por otra parte, los entornos de *Lisp* disponibles hoy en día ofrecen una gran integración entre el intérprete, el editor de programas y el compilador, pudiendo pasar automática o fácilmente de uno a otro, con lo que el desarrollo de aplicaciones resulta mucho más cómodo y agradable que en otros lenguajes de programación.

9.1.3 Ventajas e inconvenientes

Ya hemos comentado que la mayores ventajas del *Lisp* son la facilidad para manejar listas, la posibilidad de trabajar indistintamente en modo intérprete y en modo compilado, y las ayudas de depuración, todo ello integrado en entornos amigables.

Los mayores inconvenientes del *Lisp* son, en cambio, el tamaño y la ineficiencia. Una de las razones del tamaño que ocupan las versiones de *Common Lisp*, tanto en disco duro como en memoria dinámica (RAM), se debe a la gran cantidad de funciones que contiene el lenguaje —más de 800— y a las capacidades de ayuda que lo acompañan. Otra de las razones es que *Lisp* utiliza unas celdillas elementales con las que se van encadenando los elementos de las listas, por lo que ocupa más memoria que si cada dato estuviera simplemente almacenado en un registro. El estilo propio del *Lisp*, basado en la recursividad, también favorece un elevado consumo de memoria.

Cuando ésta se agota, el sistema debe buscar las celdillas que ya no están en uso, para reutilizarlas en nuevas listas; es lo que se conoce como *recogida de basura* (*garbage collection*). El inconveniente es que, en la mayor parte de las implementaciones, el programa debe interrumpir el proceso que está realizando, hasta que la recogida de basura termina.

En cuanto a la ineficiencia del *Lisp*, la razón principal estriba en que el procesamiento de una lista requiere ir buscando uno a uno los elementos que la componen, mediante un sistema de punteros. Por otra parte, la invocación de una función se realiza mediante el símbolo que la representa; esto es lo que permite trabajar a la vez en modo compilado e interpretado, pero tiene el inconveniente de la compilación no resulta tan eficaz como en otros lenguajes de programación.

Por estos motivos, *Lisp* es un lenguaje sumamente útil y cómodo para desarrollar programas de inteligencia artificial, donde la flexibilidad es un requisito primordial. Sin embargo, en problemas en que el tipo de datos y los pasos necesarios para resolverlo están definidos a priori, es preferible utilizar un lenguaje más eficiente.

9.2 Introducción al GoldWorks

Describimos brevemente en esta sección las características más importantes de la herramienta que hemos utilizado para la creación de la primera versión del sistema experto DIAVAL. Nosotros hemos trabajado con la versión para ordenadores tipo PC 386, que se ejecuta sobre *Microsoft Windows*, aunque existen también versiones para ordenadores *Macintosh* y estaciones de trabajo *Sun*.

9.2.1 Marcos e instancias

Los objetos constituyen una forma estructurada de representar el conocimiento, y se clasifican en dos tipos: **marcos** (*frames*) e **instancias** (*instances*). Los primeros corresponden a lo que en otros sistemas expertos se conoce como *clases*. Cada marco puede poseer varios hijos o submarcos (*frame children*), que representan las subclases de que se compone, y a la vez puede tener varios padres. Tenemos por tanto una red de clasificación (sec. 2.2.3) formada a partir de marcos padres y marcos hijos.

Como ejemplo, podemos considerar el interfaz gráfico del propio programa *GoldWorks*, que consiste en un conjunto de marcos. El marco BASIC-IMAGE tiene un padre, GRAPHIC-OBJECT, y tres hijos: ACTUATOR, SCREEN-TEMPLATE y GRAPHIC-IMAGE. Éste último, a su vez, posee varios submarcos: FIXED-MENU, TEXT-IMAGE, NUMERIC-IMAGE, etc. El usuario construye su interfaz creando las imágenes y los menús como instancias de los marcos predefinidos o de nuevos marcos adaptados a sus necesidades.

Por poner otro ejemplo, en el capítulo 13 veremos que en nuestra implementación de la red bayesiana cardiológica el marco DATO posee varios submarcos: SÍNTOMA, AUSCULTACION (es decir, signo observado en la auscultación), RAYOS-X (hallazgo radiológico), DATO-ECO (hallazgo ecocardiográfico), etc. A su vez, DATO-ECO se compone de MODO-M, BIDIMENSIONAL, DOPPLER y MAPA-COLOR, correspondientes a los distintos tipos de datos que describiremos en la sección 12.3.

En *GoldWorks*, existe un marco general, denmominado *TOP-FRAME*, que engloba todos los demás, tanto los pertenecientes al interfaz gráfico como los definidos por el diseñador del sistema experto.

Cada marco posee varios **campos** (*slots*), que corresponden a las propiedades características de sus instancias. La importancia de esta representación consiste en el mecanismo de *herencia*, por el que cada marco, además de sus campos específicos, posee todos los campos de sus padres.

Algunos de los campos del marco BASIC-IMAGE son: x, y, height, width, value y label. Los cuatro primeros, que indican la localización y el tamaño de la imagen, son una herencia del marco GRAPHIC-OBJECT. Los dos últimos, que indican respectivamente el valor representado por la imagen y la etiqueta que aparece debajo de ella, son propios de BASIC-IMAGE. Tanto unos como otros serán heredados por los descendientes de este marco.

Las **instancias** son casos particulares o elementos de una clase determinada. Así, dentro del marco SÍNTOMA tenemos las instancias Disnea, Hemoptisis, Dolor-torácico, Cianosis-periférica, etc. Naturalmente, cada instancia hereda los campos de su marco padre.[4]

La misión principal del campo de una instancia es contener un valor (o una lista de valores), aunque no se limita a un simple almacenamiento estático, sino que cada campo lleva asociadas varias propiedades, que se conocen como **facetas** (*facets*). Entre ellas, `:doc-string` contiene documentación sobre el programa, `:constraints` comprueba que el valor asignado es de cierto tipo (una lista, una cadena, un número dentro de cierto rango, etc.), `:print-name`, `:explanation-string` y `:query-form`

[4]En *GoldWorks*, cada instancia sólo puede tener un marco padre, lo cual supone una limitación a la hora de representar el conocimiento.

están destinadas al desarrollo del interfaz, y así hasta un total de 14 facetas predefinidas más otras que puede crear el diseñador del sistema. En el tema que estamos tratando, nos interesan especialmente las facetas asociadas a mecanismos de inferencia.

Entre éstas, la faceta `:multivalued` indica si el campo puede tener varios valores; de no ser así, la asignación de un nuevo valor en un campo elimina el valor anterior. Por ejemplo, el campo causas del marco NODO es multivaluado (un nodo puede tener varias causas), mientras que el campo edad del marco PERSONA es univaluado.

Otra de las facetas más importantes es `:default-values`, que la instancia hereda de su marco padre. Así, los campos transtorácico y tranesofágico del marco DATO-ECO tienen asignados los valores por defecto `T` y `NIL`, respectivamente, lo cual significa que —mientras no se indique lo contrario— suponemos que el ecocardiograma es sólo transtorácico.

Por último, dentro de los mecanismos de inferencia asociados a las facetas tienen especial importancia los **demonios** (*daemons*), que reciben este nombre porque son funciones que están "al acecho" y se disparan ante ciertos acontecimientos. Los *demonios de acceso*, indicados por la faceta `:when-accessed`, se disparan cuando buscamos el valor de un campo. Por ejemplo, puede haber un demonio que se ocupe de abrir una base de datos y extraer el valor solicitado, o un demonio que formule una pregunta al usuario del sistema.

Existen también en *GoldWorks* los *demonios de modificación*, indicados por la faceta `:when-modified`, que se ejecutan cuando cambia el valor del campo, y tienen como misión actualizar la información que dependen de dicho valor. Por ejemplo, si se modifica el peso del paciente, se dispara un demonio que recalcula la superficie corporal, y esto a su vez dispara otro demonio que actualiza en la pantalla este valor.

9.2.2 Programación orientada a objetos

La programación orientada a objetos (POO) surgió hacia 1970 en los laboratorios de Xerox (Palo Alto, California), con Alan Kay como principal responsable del proyecto. Sus investigaciones se concretaron en la creación del lenguaje Smalltalk. Desde entonces, otros lenguajes de programación y varias herramientas para la construcción de sistemas expertos han incorporado las ideas esenciales de la POO.

En realidad, las últimas versiones de *GoldWorks* incluyen dos implementaciones diferentes de POO. Por un lado, *GoldWorks* se basa en un dialecto de *Common Lisp*, llamado *GoldHill Common Lisp*; la versión 4.4 de este lenguaje se ajusta a la especificación CLOS, que es el *Lisp* orientado a objetos (del mismo modo que C++ es el C orientado a objetos). Por otro lado, *GoldWorks* incluye una especie de POO —es la que vamos a describir aquí— en que los marcos desempeñan el papel de las clases y las instancias el papel de los objetos. Los *métodos* de la POO reciben en *GoldWorks* el nombre de *gestores* (*handlers*).

Cada clase o subclase lleva asociado un gestor específico para cada uno de los mensajes que va a recibir. Por ejemplo, el siguiente programa

```
(define-handler (CONMUTADOR open) (Ventana)
   ( ... ))
```

define un gestor para la clase CONMUTADOR, de modo que cuando enviamos a una

instancia el mensaje `:open` se ejecuta la función o procedimiento que hemos indicado. (`Ventana` actúa aquí como argumento.) La instrucción

```
(send-msg 'CONMUTADOR-12 :open 'VENTANA-45)
```

envía a la instancia CONMUTADOR-12 (que pertenece a la clase CONMUTADOR) la orden de abrirse en una cierta ventana. Si no hubiese un gestor definido para este mensaje en la clase CONMUTADOR, tomaríamos el gestor correspondiente a la clase superior, según los mecanismos de herencia habituales. Por tanto, la ordenación jerárquica de clases y subclases resulta útil no sólo para la herencia de los valores y propiedades de los campos, sino también en la programación orientada a objetos.

El interfaz gráfico de *GoldWorks* está implementado según este modelo de POO, de modo que el usuario tiene una gran flexibilidad a la hora de personalizar el interfaz, creando sus propias subclases dentro de las clases predefinidas en el programa; en unos casos, al diseñador le interesará tomar los gestores y demonios que proporciona el *GoldWorks*, mientras que en otros casos preferirá definir los suyos propios para adaptar el interfaz a sus necesidades.

9.2.3 Motor de inferencia: reglas

Aunque en la implementación de nuestro sistema no hemos utilizado esta facilidad que ofrece el programa *GoldWorks*, vamos a describir aquí brevemente en qué consiste el motor de inferencia, con el único propósito de ofrecer una visión completa del programa.

Para comprender el funcionamiento de las reglas es preciso conocer antes qué son las **afirmaciones** (*assertions*) y los **patrones** (*patterns*). A cada valor asignado en el campo de una instancia corresponde una afirmación, tales como "`(edad pedro-martínez 68)`" o "`(causas estenosis-mitral fiebre-reumatica)`". Hay también afirmaciones que no están asociadas a campos; algunas son afirmaciones no estructuradas, mientras que otras se ajustan a una estructura (*relation*) definida por el diseñador del programa. El conjunto de todas ellas se conoce como *base de afirmaciones* (ver sec. 2.1.3).

Cada afirmación lleva asociados un *factor de certeza* y una *estructura de dependencia*, que indica cómo se ha obtenido (si era un valor por defecto, si lo ha indicado el usuario, si se ha deducido aplicando una regla) y sirve para generar explicaciones y para actualizar la información cuando sea necesario; para cada afirmación, es importante conocer de cuáles depende y cuáles dependen de ella.

Los **patrones** pueden entenderse como afirmaciones generalizadas, con posibilidad de incluir variables. Por ejemplo, en "`(edad ?paciente 68)`", la variable es `?paciente`; se reconoce porque en *GoldWorks* las variables van precedidas de un signo de interrogación. Una afirmación *iguala* un patrón si existe una asignación de variables tal que éste coincide con aquélla. Por ejemplo la asignación de variables `?paciente = pedro-martínez` hace que la afirmación "`(edad pedro-martínez 68)`" iguale el patrón "`(edad ?paciente 68)`".

Una regla consta de dos componentes: el *antecedente* y el *consecuente* (sec. 2.2.2). Básicamente, el antecedente consiste en una serie de patrones, llamados cláusulas; algunos de ellos pueden ser simples afirmaciones, es decir patrones sin variables. Cuando en la base de afirmaciones existe una que iguala el primer patrón, se examina

el segundo para ver si existe otra que lo iguale, teniendo en cuenta que algunas variables ya tienen asignado un valor. Cuando se han igualado sucesivamente todos los patrones de una regla, ésta pasa a la agenda en espera de ser ejecutada.

GoldWorks ofrece además otros tipos de patrones más complejos. Uno de ellos sirve para comprobar si existe una instancia de un marco determinado que tenga un cierto valor. Supongamos que tenemos el siguiente patrón:

```
(instance ?p is paciente
   with sexo var\'on
   with-unknown diab\'etico ?
   with edad ?e)
```

Si la variable `?e` no tiene aún un valor asignado, habrá una igualación del patrón para cada paciente varón del que no sabemos todavía si es diabético, asignando a `?e` la edad de dicho paciente. En cambio, si las cláusulas anteriores de la regla habían asignado ya un valor a la variable `?e` (edad), este patrón se igualará solamente si existe un paciente varón con una dicha edad del que no sabemos todavía si es diabético o no.

Hay otros tipos de patrones que sirven para formar conjunciones y disyunciones (`and`, `or`), eliminar afirmaciones, asignar valores a las variables que aparecen en la regla, evaluar expresiones *Lisp*, incluir comentarios, mostrar resultados, etc.

La función `forward-chain` se encarga de disparar las reglas que se encuentran listas en la agenda. Para cada una de ellas se ejecutarán las acciones indicadas en su consecuente, lo cual puede dar lugar a que se confirme el antecedente de otras nuevas reglas (o de las mismas reglas para nuevos valores), con lo que se crean más elementos en la agenda, que serán disparados sucesivamente.

Otro modo de encadenamiento de reglas que ofrece nuestra herramienta consiste en el encadenamiento hacia atrás. El procedimiento consiste en formular un patrón como objetivo (*an attempt*) e investigar si a partir de las reglas y afirmaciones que posee el sistema puede deducirse alguna afirmación que lo iguale. Por ejemplo, la orden

```
(query (make-attempt
          (instance ?p is paciente
             with diabético T)))
```

trata de comprobar si existe algún paciente diabético. Para ello se examinan las reglas que podrían conducir a este resultado y se trata de confirmar el antecedente de alguna de ellas, y así sucesivamente.

Una propiedad importante de las reglas es el tipo de dependencia. En la mayor parte de los casos, cuando el antecedente de una regla deja de ser cierto, hay que eliminar la conclusión que habíamos obtenido de ella; es lo que se conoce como *dependencia lógica*. Sin embargo, hay otros casos en que el consecuente de una regla sigue siendo válido simplemente porque el antecedente lo fue en el pasado, aunque ya no lo sea. Por ejemplo, para la regla "Si el carrete recibe luz, el carrete está velado", el consecuente sigue siendo cierto si el carrete recibió la luz en el pasado, aunque ya no la reciba. La opción `:dependency` indica cómo queremos que se comporte la regla.

Ésta es una descripción muy breve de las capacidades del *GoldWorks*. En realidad, el programa ofrece mecanismos mucho más sofisticados de controlar la inferencia. Uno

de ellos consiste en definir las reglas como `:forward`, `:backward` o `:bidirectional` para determinar en que tipo de encadenamiento es aplicable. También existe la posibilidad de asignar a cada regla una prioridad, de modo que las que tienen mayor prioridad se disparan antes. Otro mecanismo para controlar la inferencia consiste en agrupar las reglas en conjuntos, que pueden activarse o desactivarse cuando convenga. Finalmente, es posible asignar cada regla y cada conjunto de reglas a un *patrocinador* (*sponsor*), el cual controla su propia agenda, donde almacena las reglas que van a ser disparadas. Los patrocinadores se organizan en un árbol jerárquico, que se encarga de distribuir entre sus ramas los recursos disponibles, es decir, la proporción de reglas que va a disparar cada patrocinador.

El punto más débil de este potente motor de inferencia se encuentra en el tratamiento de la incertidumbre. *GoldWorks* proporciona una función que calcula la certeza de la conclusión de una regla a partir de las certezas de sus premisas y de la certeza de la propia regla (de modo semejante al MYCIN) y la combina con la anterior certeza de dicha afimación.[5] En este punto, *GoldWorks* presenta el grave inconveniente de que no propaga automáticamente las modificaciones en la certeza, de modo que el resultado depende del orden en que se introducen la información al sistema, tal como reconoce el mismo manual del programa. El usuario tiene la posibilidad de definir una función con este objetivo, pero la tarea puede resultar sumamente complicada por los bucles y ciclos que aparecen en el proceso de inferencia.

De todos modos, hay que decir en defensa del *GoldWorks* que el cálculo de factores de certeza en sistemas basados en reglas plantea serias dificultades por la falta de una teoría consistente, tal como discutíamos en la sección 2.4.2; es imposible que una herramienta ofrezca una implementación satisfactoria para un problema que aún no está resuelto a nivel teórico. Para paliar esta situación, el programa da la máxima flexibilidad al diseñador del programa, permitiendo que los factores de certeza tomen cualquier tipo de valores, incluso no numéricos.

9.2.4 Interfaces

Una de las mayores ventajas del *GoldWorks* es que posee numerosos interfaces más o menos integrados en el conjunto del programa, entre los que se encuentran los siguientes:

- *Interfaz gráfico.* Se basa en el entorno de ventanas *MS-Windows* y permite crear diferentes tipos de ventanas e imágenes, según las necesidades del diseñador del sistema experto. Nuestro programa DIAVAL se beneficia de este potente interfaz, según describe el capítulo 13, cuyas figuras muestran varios ejemplos de los tipos de ventanas que pueden crearse.
- *Interfaz para* dBASE III. Ha sido utilizado en el desarrollo del sistema experto DIAVAL para almacenar los resultados y los informes de los ecocardiogramas en un formato estándar. Maneja tanto ficheros de datos (`.DBF`), como índices (`.NDX`) y campos "memo" (`.MEM`).
- *Interfaz para* Lotus 1-2-3. Lee y escribe valores y fórmulas en archivos generados por la hoja de cálculo.

[5] *GoldWorks* utiliza la fórmula original de MYCIN en vez de la fórmula mejorada introducida por van Melle (sec. 2.3.2).

• *Intérprete* ASCII. Lee ficheros ASCII y los convierte en datos tipo *Lisp* o en instancias de *GoldWorks*.

• *Interfaz* DDE. Establece canales de comunicación con otros programas, tales como *Microsoft Excel* o *Microsoft Word*, mediante intercambio dinámico de datos (*Dynamic Data Exchange*).

• *Interfaz con otros lenguajes.* Conocido como FFI, *Foreign Function Interface*. Permite utilizar en *Lisp* datos y funciones generados por otros lenguajes, y viceversa. En particular, puede utilizar y crear librerías del tipo DLL (*Dynamic Link Libraries*), que constituyen la característica esencial del entorno *Microsoft Windows*.

9.2.5 Valoración

En general, podemos afirmar que las principales ventajas e inconvenientes del *GoldWorks* se deben a que está implementado en *Lisp*. Entre las primeras podemos citar las siguientes:

1. *Potencia.* El *GoldWorks* admite varios tipos de representación del conocimiento, que básicamente se ajustan a dos paradigmas: *objetos*, con marcos, instancias, campos, facetas, demonios, todo ello organizado en una red de clasificación con mecanismos de herencia y programación orientada a objetos; y *reglas*, cada una de ellas con prioridad, factor de certeza, direccionalidad, dependencia, justificación, explicación (aunque rudimentaria), y además conjuntos de reglas, patrocinadores, agendas, encadenamiento hacia delante, hacia atrás y mixto, etc.

2. *Flexibilidad.* Hay que señalar ante todo la gran flexibilidad que *GoldWorks* ofrece en cuanto al tipo de patrones que puede utilizar en las reglas, a diferencia de otras herramientas para desarrollo de sistemas expertos, que imponen un patrón rígido tanto para el antecedente como para el consecuente. También hay numerosas posibidades de ampliar y personalizar la herramienta, definiendo nuevas facetas para los campos, nuevas relaciones para las cláusulas de las reglas, nuevos tipos de imágenes para el interfaz, etc., etc. Podemos recordar, como ejemplo, que el programa ofrece una función para combinar los factores de certeza, suponiendo que éstos son números entre 0 y 1, pero es posible sustituir esta función por otra e incluso utilizar valores no numéricos para representar el grado de certeza.

3. *Integración con Lisp.* La flexibilidad mencionada se debe sobre todo a que *GoldWorks* está implementado sobre *Lisp*. Ya dijimos en la sección 9.1 que las principales ventajas de este lenguaje son la posibilidad de trabajar a la vez en modos intérprete y compilado, la facilidad para manejar listas y símbolos, y las ayudas para depurar programas.

4. *Entorno de desarrollo cómodo.* Está basado en *Microsoft Windows*. Posee editores para marcos, instancias, reglas, etc.; es capaz de mostrar la red de objetos y la red de reglas, y permite además supervisar el motor de inferencia inspeccionando las agendas. Al trabajar con reglas, resulta utilísimo también el *Partial Matcher*, en que podemos observar gráficamente qué instancias y afirmaciones satisfacen cada cláusula de una regla, y realizar operaciones similares orientadas

a depurar la base de conocimiento. Por otra parte, resulta inmediato pasar del *GoldWorks* al intérprete de *Lisp* o al editor de programas con sólo pulsar una tecla.

5. *Interfaces.* En la primera versión del sistema experto DIAVAL nos han resultado especialmente útiles el interfaz gráfico y el interfaz para dBASE III. En particular, el módulo *Graphics Layout Tool* supone una gran ayuda a la hora de diseñar las pantallas para el interfaz de usuario.

A pesar de sus importantes ventajas, *GoldWorks* presenta también serios inconvenientes:

1. *Tamaño.* Existen dos versiones del programa: la de desarrollo y la ejecutable. La primera contiene todos los elementos necesarios para crear sistemas expertos; la segunda sólo permite ejecutar un sistema ya desarrollado, sin posibilidad de modificarlo. El mayor inconveniente de la primera es el gran tamaño que ocupa. Para que funcione aceptablemente necesita un mínimo de 8 Mb. de memoria RAM y 13 Mb. de disco duro. La versión ejecutable, es decir, la que se entrega al usuario final para que pueda utilizar el sistema experto, necesita aproximadamente la mitad de memoria y de capacidad de disco, lo cual sigue siendo un tamaño considerable, teniendo en cuenta que se trata de la versión ejecutable. Afortunadamente, la fabricación de ordenadores progresa tan rápidamente que la capacidad que hace muy pocos años parecía exagerada hoy resulta habitual.

2. *Complejidad.* No resulta nada sencillo manejar el *GoldWorks*. Además de conocer el entorno *Microsoft Windows*, es imprescindible saber programar en *Lisp*, concretamente en *GoldHill Common Lisp*. Hace falta también conocer el GMACS, un editor de texto semejante al EMACS, orientado a la creación de programas en *Lisp*. Después, el diseñador debe familiarizarse con todos los conceptos de marcos, instancias, facetas, demonios, herencia, encadenamiento hacia delante y hacia atrás, conjuntos de reglas, agendas, patrocinadores, etc., tal como los utiliza el *GoldWorks*. Por último, debe aprender a manejar otras herramientas más específicas para la aplicación concreta que va a desarrollar: el interfaz gráfico (junto con el *Graphics Layout Tool*), el interfaz de bases de datos, etc. La complejidad del programa se refleja en la extensa documentación que lo acompaña: 14 manuales que superan en total las 2.800 páginas. Además, en la práctica se hace necesario tener a mano el libro de referencia del *Common Lisp* [91], lo cual implica 900 páginas más. La conclusión es que, a diferencia de otras herramientas más sencillas, el diseño de un sistema experto con *GoldWorks* debe correr a cargo de un especialista en este tipo de programación.

3. *Lentitud.* En general, la flexibilidad y la eficiencia son dos dimensiones contrapuestas. En *GoldWorks* se ha optado por la primera en detrimento de la segunda. En el desarrollo de nuestro sistema experto hemos trabajado con un ordenador 386 de 33 Mhz. con procesador matemático y 12 Mb. de RAM. Cargar la versión de desarrollo completa lleva más de diez minutos.

   La recogida de basura (ver sec. 9.1.3) tarda más de 10 segundos, y durante ese tiempo el usuario debe esperar hasta que se reanude el proceso interrumpido.

Naturalmente, al aumentar la memoria RAM este proceso se realiza con menor frecuencia, pero aumenta el tiempo requerido. La recogida de basura es una característica de todos los sistemas que funcionan sobre *Lisp*, aunque en otras implementaciones del lenguaje no hay una interrupción tan prolongada como en el *GoldWorks*.

4. *Errores de implementación.* Resulta desesperante cuando se trabaja con el programa observar que no se comporta como indican los manuales, y peor aún comprobar que muchos de estos fallos hacen que el sistema se quede "colgado", con lo que es necesario reiniciar el programa (recordemos que esta operación tarda más de diez minutos), perdiendo a veces valiosa información. Cada nueva versión del *GoldWorks* corrige los errores de la anterior pero introduce otros diferentes. La frecuencia con que se repiten estos fallos hacen que el diseñador pierda muchísimo tiempo luchando contra las deficiencias del programa.

En resumen: el *GoldWorks* es una herramienta potente y flexible pero ineficiente y difícil de manejar. Por tanto, puede resultar muy útil para diseñar la primera versión de un sistema experto porque permite realizar un prototipo en un tiempo relativamente breve, experimentando nuevas técnicas hasta encontrar la solución más adecuada. Sin embargo, para la realización de una versión definitiva conviene emplear una herramienta más eficiente y robusta, aunque no sea tan flexible, o incluso programar directamente en un lenguaje de propósito general, como C o Pascal.

9.3 Razonamiento bayesiano

9.3.1 Representación de la red

Para comprender este sección, conviene recordar lo expuesto en los capítulos 4, 5 y 6. En ellos vimos que una red se compone fundamentalmente de dos tipos de elementos: nodos y enlaces, y para cada uno de ellos vamos a definir un marco.

El marco NODO tiene varios campos. Dos de ellos, que son multivaluados, corresponden a las causas y los efectos del nodo. Hay unos campos que resultan importantes de cara a la inferencia, tales como el número de valores (n-valores) que puede tomar la variable, los valores posibles y el tipo de interacción: caso general o puerta OR; en el caso general, el nodo debe almacenar también la probabilidad condicional (prob-condic), o la probabilidad a priori (pi) si se trata de un nodo sin padres. Otros campos están orientados hacia el interfaz y hacia la explicación del razonamiento, tales como el nombre completo (por ejemplo, "calcificación de las valvas aórticas").

El marco ENLACE tiene dos campos, causa y efecto, que indican los dos nodos unidos por un enlace. Esta información es redundante, pero resulta muy útil, sobre todo cuando se está desarrollando el sistema. En el caso de la puerta OR, los enlaces cobran una especial importancia, pues en ellos se almacena la eficiencia causa-efecto (sec. 5.2) y la explicación del mecanismo causal.

9.3.2 Propagación de evidencia

Puesto que el algoritmo de propagación de evidencia (caps. 5 y 6) se comprende mejor como paso de mensajes entre los nodos, la primera implementación que realizamos

estaba basada en programación orientada a objetos, descrita anteriormente. Sin embargo, por motivos de eficiencia y de facilidad de depuración, se vió que resultaba más conveniente utilizar funciones *Lisp* en vez de los gestores de mensajes propios de la POO.

Para realizar la inferencia, inicializamos la red asignando a los nodos terminales (que serán instancias del marco DATO) el vector $\lambda(x)$ correspondiente a la información disponible. En general, esta asignación será realizada por el usuario desde el interfaz gráfico. A continuación, invocamos la función `prob-post` con el nombre de un nodo como argumento, con el fin de calcular la probabilidad a posteriori de dicho nodo, $P^*(x)$. Esto da lugar a la invocación recursiva de las funciones `pi`, `landa`, `pi-e-c` y `landa-e-c`, que calculan respectivamente $\pi(x)$, $\lambda(x)$, $\pi_X(u_i)$ y $\lambda_{Y_j}(x)$. El algoritmo de propagación de evidencia utilizado es el *condicionamiento local*, tal como fue descrito en el capítulo 6. Necesitamos, por tanto, disponer un campo en cada instancia-nodo que nos indique si el nodo está en espera y cuáles son las variables en que debe ser condicionado (sec. 6.2.1). El cálculo de $\pi(x)$ y de $\pi_X(u_i)$ debe consultar el valor del campo interacción para el nodo X, pues la computación para la puerta OR es muy diferente del caso general. Los valores obtenidos para los nodos y para los enlaces son almacenados en los campos de las instancias correspondientes, con el fin de ahorrar computación a la hora de calcular la probabilidad a posteriori de un nuevo nodo.

9.3.3 Aprendizaje

La implementación del modelo de aprendizaje descrita en el capítulo 7 es realmente sencilla y hay muy poco que añadir a lo dicho anteriormente. Como explicábamos allí, primero es necesario realizar la inferencia tomando los valores medios de las probabilidades como si fueran valores exactamente determinados. Cuando el usuario ha introducido toda la información disponible y se ha llegado a un diagnóstico, el caso puede darse por cerrado y, en el momento de archivarlo en la base de datos, se invoca la función que actualiza los valores de las probabilidades y sus correspondientes desviaciones estándares.

Para todo nodo que interactúa mediante el modelo general, tomamos la probabilidad condicional, que será una matriz almacenada en el campo prob-condic de la instancia asociada, y el valor de la desviación estándar correspondiente, que será otra matriz almacenada en el campo s-prob-condic. Basta entonces invocar la función *Lisp* que se encarga de calcular la $\Delta_x^{\bar{u}}$ dada por la ecuación (7.27) y, a partir de ella, deducir los nuevos valores de la probabilidad condicional y de su desviación estándar —ecs. (7.15) y (7.16)— que serán almacenadas en los campos correspondientes.

Para los nodos que interactúan mediante la puerta OR, las probabilidades condicionales se toman del campo efic de cada enlace, se actualizan de forma similar mediante funciones acordes con las ecuaciones (7.42), (7.15) y (7.16) y se almacenan en los mismos campos de donde se habían tomado.

9.3.4 Explicación

Algunos de los campos que hemos definido para la inferencia son también útiles a la hora de explicar el razonamiento. Naturalmente, los más importantes son los que recogen la información numérica estática (la que define la red) y dinámica (la

resultante de la inferencia). También el campo interacción juega un papel importante al determinar si vamos a presentar los padres de un nodo como *causas* (para la puerta OR) o como *factores* que influyen sobre la probabilidad del nodo (en el caso general). Igualmente, la clase de enlace (ver la sec. 8.3.1) influye a la hora de presentar los efectos, tal como puede comprobarse en la figura 8.1 (pág. 157).

Además de éstos, existen otros campos que son específicos de la explicación, es decir, que no intervienen en la inferencia, tales como el nombre completo de un nodo, la fórmula de un parámetro o el mecanismo-causal de un enlace. La lista de opciones disponible para cada elemento, sea nodo, parámetro o enlace, se genera dinámicamente; así, si el campo fórmula de un parámetro está vacío, ésta opción no aparece en el menú que se muestra al usuario.

En el capítulo 13 describiremos cómo hemos implementado la capacidad de explicación dentro de un entorno gráfico, integrándola con la entrada de datos y con la exposición de resultados.

Capítulo 10

Valoración de las redes bayesianas

10.1 Orígenes y evolución

10.1.1 Desarrollo histórico

Aunque la expansión de las redes bayesianas (RR.BB.) es un fenómeno muy reciente, la idea fundamental se remonta a los años 20, en que el geneticista Sewal Wright [263] utilizó por primera vez una representación gráfica asociada a una distribución de probabilidad. Los estadísticos, que inicialmente rechazaron estas técnicas, fueron reconociendo más tarde la importancia y la utilidad de los métodos gráficos para representar las propiedades de descomposición de las tablas estadísticas. Por tanto, la primera línea de investigación que condujo a lo que hoy conocemos como redes bayesianas fue desarrollada por matemáticos especializados en probabilidad y estadísitica.

Sin embargo, la gran difusión que las RR.BB. han alcanzado en el campo de la inteligencia artificial se debe sobre todo a los trabajos de Judea Pearl y su grupo de investigadores de la Universidad de California Los Angeles (UCLA). Pearl, que tenía ya reconocido prestigio en el campo de la I.A. por sus publicaciones sobre búsqueda heurística, explica en un reciente artículo cómo se interesó por este tema [175]:

> La idea de estudiar la computación probabilística distribuida en modelos gráficos empezó a fraguarse en mi mente a finales de los 70, cuando leí el artículo de Rumelhart sobre comprensión escrita [191]. En dicho artículo, Rumelhart presentaba evidencia convincente de que la comprensión de textos debe ser un proceso distribuido que combine tanto inferencias ascendentes como descendentes. [...] Así empecé a investigar la posibilidad de lograr computación distribuida en un marco "puramente" bayesiano.

En 1982, Pearl [170] propuso un algoritmo para calcular probabilidad en árboles, basado en el paso de mensajes entre nodos vecinos. Al año siguiente, la tesis doctoral de uno de sus alumnos, Jin Hyung Kim [125], extendió el modelo a poliárboles (en los árboles cada nodo sólo podía tener un padre). En este trabajo, que apareció de forma resumida en [126], podemos encontrar ya las redes bayesianas tal como hoy las

conocemos, con las propiedades de independencia y el algoritmo π-λ de propagación (ver la sec. 5.1).

En el célebre artículo de Pearl [172] de 1986, se usan por primera vez los términos *Bayesian networks, belief networks* y *causal networks*. Dicho artículo describe la puerta OR (ver la sec. 5.2), sugiere tres métodos para tratar los bucles (sec. 6.1) y propone un algoritmo para construir la red a partir de un conjunto de datos (sec. 7.1).

La tercera línea de investigación que condujo a las RR.BB. (recordemos que las dos primeras provenían de la teoría estadístico-probabilística y de la inteligencia artificial) surgió en el campo de la teoría de la decisión [119]. Los *diagramas de influencia*, de los que hablaremos en la sección 10.1.3, son una extensión de las RR.BB. que, además de los nodos correspondientes a variables, incluyen otros para representar decisiones (*decision nodes*) y resultados —ganancias o pérdidas— (*value nodes*) correspondientes a las decisiones. En un principio, la forma de resolver estos diagramas consistía en convertirlos en árboles de decisión, hasta que Shachter [198] desarrolló el algoritmo de *inversión de arcos*, mucho más eficiente.

La publicación más importante en el campo de las RR.BB. es el libro de Pearl, de 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* [174]. En él, además de presentar de forma clara y ordenada los resultados obtenidos hasta ese momento, se discuten la relación de las RR.BB. con las redes de Markov, los sistemas expertos, el tratamiento de la incertidumbre, la lógica no monótona y otros aspectos de la inteligencia artificial.

También en el año 1988 hubo otro avance de gran importancia en este campo: la publicación del algoritmo de Lauritzen y Spiegelhalter [135]. Ambos autores, especialistas en probabilidad y estadística, desarrollaron un método conocido como *propagación en árbol de cliques*, el primer algoritmo que permitía calcular la probabilidad con eficacia razonable incluso en redes con bucles.[1] Este método, con una modificación introducida por Jensen *et al.* [122] es hoy en día el método estándar para el cálculo en RR.BB.

Los artículos que hemos citado aparecen recogidos en el libro de Shafer y Pearl de 1990 [203]. Prácticamente todas las demás publicaciones sobre RR.BB. son posteriores a este año, según se puede comprobar en la bibliografía de esta memoria, lo cual es una prueba de la "juventud" del campo que estamos tratando.

10.1.2 Investigación actual en RR.BB.

En el grupo formado por Judea Pearl en la Universidad de California Los Angeles realizaron su doctorado, entre otros, Jin Hyung Kim, Ingrid Zuckerman, Dan Geiger, Hector Geffner, Thomas Verma y Moisés Goldszmidt, aunque actualmente ninguno de ellos pertenece ya a dicha universidad. En capítulos posteriores de esta tesis mencionaremos los trabajos que ha realizado cada uno de ellos.

Los ya mencionados Lauritzen y Spiegelhalter, aunque desarrollaron sus trabajos más famosos en la Universidad de Cambridge (Reino Unido), son los "padres" del grupo que dirigen Stig Andersen y Steen Andreassen en la Universidad de Aalborg (Dinamarca), uno de los más importantes que existen en el campo de las redes bayesianas (o *causal probabilistic networks*, como ellos las denominan). A este grupo

[1] Un método muy similar había sido propuesto anteriormente por Cannings y sus colaboradores [30, 29] en el campo de la genética, para el cálculo de pedigríes.

pertenecen también Finn Jensen, Frank Jensen, Uffe Kjærulff, Kristian Olesen y otros que citaremos más adelante.

La Universidad de Stanford (California), cuna de los sistemas expertos, es otro de los centros de donde han surgido destacados investigadores sobre RR.BB. Al hablar de los diagramas de influencia hemos citado ya a Howard, Matheson y Shachter. También Gregory Cooper, David Heckerman, Eric Horvitz, Edward Herskovits, Jaap Suermondt y Sampath Srinivas se formaron en ella. En la actualidad existe allí un grupo de jóvenes investigadores, dirigido por Ross Shachter, que está realizando una brillante labor.

Peter Szolovits, cuyos trabajos sobre inteligencia artificial en medicina se consideran clásicos en la especialidad y director del *Clinical Decision Making Group* del *Massachusetts Institute of Technology*, ha comenzado a investigar recientemente en el campo de las redes bayesianas. También en el M.I.T. se encuentra Carlos Rojas-Guzmán, quien está utilizando algoritmos genéticos para computar redes bayesianas.

En la Universidad Carnegie-Mellon, Marek Druzdzel ha concluido su doctorado, dirigido por Max Henrion, y ha pasado a la Universidad de Pittsburgh, donde se encuentran también Cooper, Spirtes, Glymour y Scheines.

Entre las empresas privadas que realizan investigación sobre redes bayesianas destaca *Rockwell*, situada en Palo Alto, junto a la Universidad de Stanford; en ella trabajan varios de los investigadores más importantes de este campo. También *Microsoft* está formando un potente grupo, dirigido por David Heckerman, el investigador más destacado de la joven generación. Igualmente, IBM ha contado con Thomas Verma para su centro de investigación de Nueva York.

En España, hay dos únicos grupos especializados en este campo. Uno de ellos se encuentra en la Universidad de Granada, dirigido por Serafín Moral y Luis de Campos. El otro está compuesto por Enrique Castillo y José María Sarabia, de la Universidad de Cantabria. Estos cuatro autores son matemáticos, especializados en probabilidad y estadística. Hay además investigadores en Madrid y Barcelona que han estudiado ocasionalmente las redes bayesianas y su relación con otros modelos. En varios países hispanoamericanos se realiza también investigación sobre este tema; entre ellos se encuentran, al menos, Colombia, Cuba y Venezuela.

El propósito de citar aquí esta lista de grupos ha sido mostrar la gran expansión que han tenido las redes bayesianas en todo el mundo. Si consideramos 1988 como el año de su asentamiento, no deja de sorprender que en tan sólo seis años las universidades más famosas y varias empresas importantes hayan formado grupos dedicados a esta nueva rama de la inteligencia artificial. Es de esperar que también en nuestro país y en todo el mundo de habla hispana alcancen el grado de desarrollo que deseamos. A este objetivo intenta contribuir la presente tesis doctoral.

10.1.3 Modelos afines

Vamos a presentar en esta sección tres modelos muy relacionados con las redes bayesianas. Los dos primeros suponen sendas simplificaciones, motivadas por dos factores: por la relativa dificultad de construir RR.BB. generales y por la complejidad de los algoritmos de propagación. El tercer modelo, por el contrario, consiste en una ampliación de las redes bayesianas, con el fin de aplicar la teoría probabilística al campo del análisis de la decisión.

Redes de semejanza

Las redes de semejanza (RR.SS.) fueron desarrolladas por David Heckerman [104], con el objetivo principal de facilitar la adquisición del conocimiento en la construcción de modelos probabilísticos.

Básicamente, una R.S. es un tipo particular de red bayesiana en que hay un nodo principal, cuyos valores corresponden a los diagnósticos posibles. Por tanto, aparte de las propiedades de independencia características de las RR.BB., el axioma principal de las redes de semejanza es que el conjunto de diagnósticos es exclusivo, lo cual equivale a suponer que no puede haber dos enfermedades presentes a la vez en un mismo enfermo.

Dos de los conceptos fundamentales de las RR.SS. son el *mapa global del conocimiento* y las *particiones*. El primero de ellos viene dado por la estructura topológica de la red total, en que aparecen las propiedades de independencia condicional típicas de las redes bayesianas. Una partición es una herramienta para valorar las probabilidades asociadas a la red, mediante la determinación de las semejanzas y diferencias entre cada par de enfermedades.

Las dos ventajas principales de las RR.SS. es que simplifican la obtención del conocimiento y la construcción del modelo. Una ventaja adicional es que, para cada pareja de diagnósticos, es posible explicar al usuario qué hallazgos contribuyen a favor o en contra de cada una, o resultan indiferentes en el diagnóstico diferencial.

Sin embargo, el mayor inconveniente que presentan es que en muchas aplicaciones reales el axioma fundamental —la exclusividad de los diagnósticos— no se cumple. En ecocardiografía, por ejemplo, no se puede suponer que haya una sola enfermedad; más bien sucede lo contrario: las valvulopatías están correlacionadas de tal modo que la presencia de una de ellas nos hace sospechar la presencia de otras, y de forma recíproca para su ausencia.

Redes cualitativas

Las redes cualitativas fueron desarrolladas por Wellman [249, 250] como simplificación de las redes bayesianas. Su objetivo principal consiste en tomar la estructura causal de éstas, incluidos los axiomas de independencia, pero sin utilizar información numérica.

Cada nodo tiene una serie de valores entre los que existe una relación de orden parcial, de modo semejante a las variables graduadas que definimos en la sección 5.2.2. Cada enlace lleva una etiqueta simbólica, que puede ser "+", "-" ó "?". La primera de ellas denota una correlación positiva entre causa y efecto, es decir, que un valor más alto para la causa indica un valor más alto para el efecto.

La propagación de la evidencia se realiza igualmente a nivel cualitativo, asignando a las variables las etiquetas anteriores, combinando según ciertas reglas la evidencia transmitida por cada enlace.

Las redes de semejanza poseen las mismas ventajas las redes cualitativas: la facilidad para la construcción de la red y la eficiencia de los algoritmos de propagación. También para ellas existe un modelo de explicación, desarrollado por Druzdzel y Henrion [79, 77], del que ya hemos hablado en la sección 8.2.2.

Su mayor inconveniente es la incapacidad para discriminar los distintos grados de evidencia. Por ejemplo, dada una hipótesis la combinación de un hallazgo que aporta

evidencia decisiva a favor con un hallazgo que apenas aporta evidencia en contra da como resultado una ambigüedad (se representa como ?), de modo que el sistema no llega a ninguna conclusión. Por tanto, nos parece que la aplicabilidad de las redes cualitativas en su estado actual es muy limitada, y se reduce a algunos casos concretos en que, por alguna razón particular las ambigüedades no existan o puedan resolverse de algún modo.

Diagramas de influencia

Los diagramas de influencia fueron creados por Howard y Matheson [118] como modelos gráficos para el análisis de la decisión. Aunque surgieron independientemente de las redes bayesianas, son en realidad una extensión de éstas, o si se quiere, las RR.BB. son un caso particular de aquéllos. La diferencia es que los diagramas de influencia, además de los nodos correspondientes a variables probabilísticas (que se representan mediante un círculo o un óvalo, como es habitual en RR.BB.), existen *nodos de decisión* (representados por cuadrados o rectángulos), que corresponden a las posibles decisiones que están a nuestro alcance, y *nodos valor*, que indican la utilidad asociada a los valores de las variables probabilísticas y de las decisiones tomadas.

En la sección 6.2.4.3, al hablar de la posibilidad de utilizar el condicionamiento local para calcular diagramas de influencia ya vimos un ejemplo en que aparecía un nodo de decisión, dado por la figura 6.8.

El tema de los algoritmos es importante en relación con la aplicabilidad de los diagramas de influencia. El método inicial propuesto por sus diseñadores era bastante ineficiente. La posibilidad de utilizar diagramas de influencia en problemas reales no surgió hasta que Ross Shachter [198, 199] desarrolló la *inversión de arcos* como técnica mucho más eficiente a la hora de evaluar los diagramas. Más tarde, Cooper [52] demostró que es posible adaptar cualquier algoritmo de redes bayesianas de modo que pueda aplicarse a los diagramas de influencia.

La ventaja principal de los diagramas de influencia frente a las redes bayesianas puras es que éstos son capaces de determinar cuál es la decisión más eficaz según la evidencia disponible. Es decir, no se limitan a dar un diagnóstico, sino que además recomiendan qué acción conviene realizar en cada caso.

Desde el punto de vista médico, los diagramas de influencia permiten extender los modelos probabilísticos de las redes bayesianas con el fin emplearlos en el tratamiento de los pacientes. Para ello, es preciso añadir a la red los nodos correspondientes a las posibles acciones terapéuticas y un nodo-valor que represente la utilidad (en términos de mortalidad, de calidad de vida y e incluso de coste económico) asociada a cada posible resultado de la intervención. De este modo, el sistema experto podrá ofrecer una propuesta de tratamiento y una predicción de la evolución del paciente en función de la decisión escogida.

10.1.4 Aplicaciones

A pesar de que las redes bayesianas constituyen una metodología muy joven, son muchas las aplicaciones que se están desarrollando recientemente, la mayor parte de ellas en el campo de la medicina. A parte de los proyectos de investigación que existen en numerosas universidades y empresas privadas de todo el mundo, existen ya al menos dos aplicaciones comerciales basadas en RR.BB.

La primera que surgió es HUGIN [4], una herramienta destinada a la construcción de redes bayesianas. Fue desarrollada por varios investigadores de la Universidad de Aalborg (Dinamarca), a partir de su experiencia en el proyecto MUNIN [7], que consistió en un sistema experto para el diagnóstico de enfermedades musculares.

El otro producto comercial que existe en la actualidad es el sistema experto PATHFINDER, desarrollado por David Heckerman en colaboración con médicos e investigadores de inteligencia artificial de la Universidad de Stanford [107, 108]. Este sistema se basa en las redes de semejanza [104], de las que hemos hablado anteriormente. Una de las principales ventajas de PATHFINDER es la posibilidad de consultar una amplia base de imágenes médicas.

Otras aplicaciones, la mayor parte de ellas en el campo de la medicina, se encuentran referenciadas en [5] y [13].

10.2 Las redes bayesianas en el contexto de la I.A.

10.2.1 Ramas y tendencias de la I.A.

Dentro de las tres ramas de la inteligencia artificial (simbólica, probabilística y conexionista) descritas en la sección 2.1.2, las RR.BB. se sitúan claramente dentro de la segunda, aunque hay ciertos aspectos en los que participan de las otras dos.

Por un lado, cada nodo de la red bayesiana puede considerarse como un procesador, según hemos mostrado en capítulos anteriores para cada una de las computaciones que las redes bayesianas realizan (la recapitulación se encuentra en la sección 10.2.3 de este mismo capítulo). En consecuencia, las RR.BB. pueden considerarse también dentro de la computación conexionista si se entiende ésta en su sentido más amplio, sin restringirse a lo que suelen ser las redes neuronales artificiales de la actualidad.

Por otro lado, la explicación en RR.BB. (cap. 8) consiste en pasar del nivel cuantitativo al cualitativo, y para ello utiliza razonamiento simbólico, por lo que este aspecto de las redes bayesianas puede considerarse dentro de la primera rama mencionada. Además, existen las redes cualitativas como modelo simplificado de las RR.BB. que en vez de utilizar probabilidades numéricas para la propagación de evidencia se basan en la transmisión de mensajes simbólicos entre los nodos, lo cual constituye un argumento más para considerarlas dentro del razonamiento simbólico.

Dijimos en la sección 2.1.2 que dentro de la inteligencia artificial podemos considerar tres tendencias: psicológica, teórica y pragmática. En este caso, no cabe duda de que las RR.BB. deben clasificarse dentro de la teórica, pues se basan en un estudio matemático de los principios que deben guiar el razonamiento y la decisión en presencia de incertidumbre. En cuanto a la tendencia psicológica, a pesar de que las RR.BB. recogen la noción intuitiva de causalidad, está claro que no pueden ser incluidas en ella porque los métodos que utilizan son completamente diferentes de los empleados por la mente humana, como quedó patente en el capítulo 3. Tampoco pueden ser incluidas en la tendencia pragmática porque buscan ante todo un fundamento normativo, aunque eso lleve a una notable pérdida de eficiencia en la computación.

10.2.2 Diagnóstico

Puesto que el objetivo principal de las redes bayesianas es el diagnóstico, vamos a comentar en esta sección sus semejanzas y diferencias con otros sistemas que se han diseñado en I.A. con el mismo objetivo.

En primer lugar, las RR.BB. son las sucesoras de los primeros programas de inteligencia artificial basados en técnicas probabilísticas, cuyo origen es anterior al de los sistemas expertos. En la sección 2.3.1 ya citamos varios sistemas de diagnóstico bayesiano que se desarrollaron en la década de los 60 [65, 95, 96, 245].

También están muy relacionadas con los *sistemas basados en conocimiento profundo*, aunque éstos surgieron de forma independiente. Lo que tienen en común es que se apoyan en un modelo del mundo real, en base al cual realizan el razonamiento. En este sentido, las RR.BB. pueden considerarse como una forma de *diagnóstico basado en fundamentos* (*diagnosis from first principles*), aunque se diferencian de los modelos que utilizan la lógica como método de razonamiento [61, 66, 93, 188]. Estos modelos suponen que existe una descripción determinista del sistema que van a diagnosticar. Las RR.BB., en cambio, pueden tratar de forma natural relaciones probabilísticas y, además son capaces de tener en cuenta las probabilidades a priori, que constituyen una valiosa información utilizada siempre por los expertos humanos, aunque sea muchas veces de forma inconsciente, según comentábamos en la sección 3.1.2).

Si un nodo de la red representa una variable correspondiente a la presencia de una anomalía, podemos diagnosticar que ésta se encuentra presente cuando su probabilidad excede un cierto umbral. Debido a la llegada de nueva evidencia, la probabilidad puede cruzar el umbral varias veces a lo largo de una consulta, con lo cual las RR.BB. constituyen una especie de *razonamiento no monótono* basado en la probabilidad, no en la lógica, que es el método habitual para este tipo de razonamiento. Además, puede ocurrir que varias variables correspondientes a anomalías superen el umbral de probabilidad, de modo que el *diagnóstico de múltiples fallos* es algo trivial en RR.BB., mientras que otros sistemas son incapaces de obtener más de un diagnóstico.

10.2.3 Computación neuronal

Las redes neuronales surgieron a partir de los trabajos de McCulloch y Pitts [143]. Sin embargo, las dificultades que los primeros sistemas presentaban incluso para resolver problemas muy simples [150], hizo que durante algunos descendiera notablemente el interés por este campo. Sin embargo, una serie de avances producida en los años 80 nos ha conducido a la gran expansión que las redes neuronales (RR.NN.) están experimentando en la actualidad [192].

A diferencia de las redes bayesianas, en que existe un único modelo definido con precisión y universalmente aceptado, el paradigma de las RR.NN. incluye numerosos modelos que difieren significativamente no sólo en los métodos sino incluso en los planteamientos básicos, tales como el propio concepto de red neuronal. Por este motivo resulta difícil comparar las RR.BB. con las "redes neuronales" en general. Por otra parte, una comparación exhaustiva queda fuera del propósito de esta sección y será objeto de un estudio posterior. Nos limitaremos aquí a ofrecer algunas indicaciones.

Desde el punto de vista de la propagación de evidencia, un nodo de una red baye-

siana puede considerarse como una neurona compleja que intercambia información con sus "neuronas" vecinas. Además, hemos visto en los ejemplos de capítulos anteriores que el mensaje que un nodo X envía a Y depende de los mensajes que el primero ha recibido de sus demás vecinos, con lo cual podemos encontrar fenómenos de amplificación e inhibición semejantes a los que existen en los sistemas nerviosos de los seres vivos.

Sin embargo, el hecho de que el mensaje transmitido de X a Y depende de los mensajes procedentes de *todos* los demás vecinos de X hace que en una R.B. no sea posible distinguir entre campo axónico y campo dendrítico, puesto que cada nodo intercambia mensajes con todos sus vecinos. En consecuencia, las redes bayesianas se parecen más a las redes de Hopfield [116][2] y difieren de otros modelos, tales como los que se basan en la retropropagación [192].

Otra diferencia entre ambos paradigmas es que, aunque la propagación es bidireccional en RR.BB., los axiomas que las definen excluyen la posibilidad de realimentación (recordar que los algoritmos exactos de propagación tenían que cortar de algún modo los bucles para poder realizar un cálculo de probabilidad consistente). Sin embargo, la realimentación desempeña un papel esencial en los modelos avanzados de computación neuronal.

De todos modos, hay una diferencia mucho más profunda. Se dice que una representación mediante una red es *localista* si existe una relación biunívoca entre nodos y conceptos. En cambio, en una *representación distribuida* un patrón (un concepto) se encuentra almacenado en el conjunto de pesos de la red como un todo y no puede asociarse a un nodo o enlace determinado; la propiedad recíproca, conocida como *superposición* (*superimposition*), significa que cada nodo interviene en la representación de varios conceptos. Por ejemplo, un estado estable de una red de Hopfield corresponde a uno de los patrones que ella puede reconocer.

La principal ventaja de una representación distribuida es la tolerancia a fallos: aunque se deterioren algunos nodos o enlaces, cada concepto sigue estando presente, con menor precisión, en el resto de la red. Sin embargo, en el caso de una representación localista, cada concepto está asociado a un nodo concreto, por lo que si falla el procesador que lo implementa, la red nunca más podrá procesar la información correspondiente.

Todos los sistemas nerviosos de los seres vivos y la mayor parte de las redes neuronales artificiales utilizan representaciones distribuidas, mientras que las redes bayesianas emplean una representación localista en que cada variable corresponde a un nodo, por definición. Por este motivo, las RR.BB. no son tolerantes a fallos: si eliminamos un nodo, la red será incapaz de considerar el hallazgo asociado o de diagnosticar la anomalía correspondiente. Como contrapartida, las RR.BB. presentan la ventaja de que pueden explicar cómo la evidencia transmitida ha influido en el diagnóstico resultante.

Por último, desde el punto de vista del aprendizaje, el algoritmo de ajuste de parámetros presentado en el capítulo 7 actúa secuencialmente a medida que la red examina y diagnostica nuevos casos. Ésta es una diferencia con respecto a los modelos de retropropagación (los más famosos dentro de las RR.NN.), que separan entrenamiento

[2] La semejanza es mayor cuando la evidencia se propaga en RR.BB. mediante simulación estocástica (sec. 6.1.3), y bajo ciertas condiciones llega a haber equivalencia [120].

y ejecución. Sin embargo, hay que advertir que existen otros modelos neuronales con capacidad local de aprendizaje [151] e incluso propuestas de utilizar máquinas de Boltzmann para construir redes bayesianas [134, 253].

10.3 Cualidades

10.3.1 Fundamento normativo

La principal ventaja de las redes bayesianas frente a método alternativos para el tratamiento de la incertidumbre es que éstas se basan en una teoría matemática que indica qué probabilidades son necesarias, cómo deben obtenerse y cómo han de combinarse.[3]

Por tanto, las RR.BB. gozan de este apoyo que no tienen los sistemas desarrollados *ad hoc*, tales como el método subjetivo de PROSPECTOR, los factores de certeza del MYCIN o los innumerables modelos de la lógica difusa. La única justificación posible para estos métodos es construir un sistema y ver que funciona de forma más o menos satisfactoria. Sin embargo, puede repetirse el caso de MYCIN, cuya evaluación resultó completamente satisfactoria [266], a pesar de que tenía graves inconsistencias, que habrían quedado en evidencia si se hubieran escogido los casos de prueba oportunos. Todas las ventajas que vamos a enumerar a continuación no son, en el fondo, más que consecuencia de utilizar un fundamento normativo, que nos permite obtener *todas* y *únicamente* las inferencias que están justificadas.

10.3.2 Razonamiento causal

Una de las características esenciales de las RR.BB. es que emplean razonamiento causal, lo cual implica varias ventajas importantes frente a otros sistemas. La primera de ellas es que el diagnóstico que realizan se basa en un modelo explícito, que tiene en cuenta la influencia entre causas y efectos. En el sistema experto que hemos desarrollado en esta tesis, la red consituye un modelo fisiopatológico de las anomalías cardíacas. Por tanto, tenemos un razonamiento en base a *conocimiento profundo*, frente a otros sistemas en que sólo recogen relaciones entre evidencia y diagnósticos, sin saber cómo ni por qué se producen.

Una de las capacidades fundamentales que permite el conocimiento profundo es la capacidad de explicación del modelo. Como hemos visto en el capítulo 8, una R.B. no sólo es capaz de explicar cómo se ha realizado la inferencia, sino también el modelo en sí, independientemente de la evidencia observada.

Otra ventaja de utilizar razonamiento causal es que la red admite *varios tipos de inferencias*:

[3] Aunque muchos de los principales partidarios de las RR.BB. defienden la interpretación subjetivista de la probabilidad (ver la nota al pie de la pág. 50), nuestra postura personal es que en el campo de la medicina es posible y por tanto deseable utilizar probabilidades objetivas [226]. Sólo en el caso de que no se hayan realizado los estudios estadísticos necesarios tendremos que recurrir a la estimación de los expertos humanos, pero siendo conscientes de que estamos obteniendo estimaciones subjetivas de magnitudes objetivas. El debate entre los partidarios de una y otra interpretación es un tema muy complejo en el que ahora no vamos a entrar. Por fortuna, todos los partidarios de las RR.BB. estamos de acuerdo en los principios axiomáticos y, a partir de ahí, podemos trabajar juntos aunque nuestras interpretaciones filosóficas sean diferentes.

Razonamiento abductivo: Dado un hallazgo (positivo o negativo), buscamos cuál es la causa causa o la combinación de causas que mejor lo explica, es decir, la que con más probabilidad nos llevaría a dicho hallazgo.

Razonamiento predictivo: Dada una anomalía, podemos realizar una estimación de la probabilidad con el objetivo de dar un pronóstico de la enfermedad o de encontrar el resultado más probable de una futura prueba.

Razonamiento intercausal: Si hay dos causas de un mismo efecto que interactúan mediante la puerta OR, la ausencia de una de ellas nos hace sospechar la presencia de la otra, y viceversa: si una de ellas es explicación suficiente del hallazgo, entonces éste ya no es motivo para sospechar la presencia de la otra, aunque tampoco la descarta (en inglés, esta propiedad se conoce como *explaining away*). En el modelo general también hay inferencia intercausal, aunque su interpretación no es tan inmediata como en el caso de la puerta OR.

Una de las limitaciones de los sistemas basados en reglas es que sólo admiten un tipo de inferencia, de los hallazgos hacia las hipótesis, sin tener en cuenta la distinción entre los tres distintos tipos de evidencia, por lo que surgen las inconsistencias que comentamos en la sección 2.4.2. Sin embargo, las RR.BB. no tienen ninguna dificultad en distinguir entre probabilidad a priori, evidencia procedente de las causas, evidencia procedente de los efectos e influencia intercausal, para tratar cada una de ellas de forma adecuada.

Quien conoce las propiedades de las redes bayesianas se sorprende con frecuencia al comprobar que muchos sistemas de razonamiento, incluso los basados en modelos causales, son incapaces o tienen dificultades para diagnosticar múltiples enfermedades (o múltiples fallos), para hallar el diagnóstico más probable, para no excluir otros posibles diagnósticos, para considerar hallazgos negativos, para ponderar evidencia contradictoria y, sobre todo, para realizar el razonamiento intercausal, capacidades que emergen de forma completamente natural en el paradigma bayesiano.

10.3.3 Tratamiento de independencias y correlaciones

Ésta es una ventaja de las RR.BB. tanto frente al método probabilístico clásico como frente a los sistemas basados en reglas.

En efecto, vimos en la sección 2.3.1 que la crítica más fuerte contra la utilización de la probabilidad en inteligencia artificial era que se basaba en hipótesis de independencia condicional que en general resultaban inadmisibles. Las RR.BB. resuelven este problema introduciendo todas las causas intermedias necesarias para lograr una descripción detallada en que los axiomas de independencia tengan sentido, según explicábamos al hablar de su semántica (sec. 4.3).

También los sistemas expertos basados en reglas utilizan hipótesis, muchas veces sin que los diseñadores sean conscientes de ellas. Así, la de independencia entre hallazgos suele aparecer entre las hipótesis explícitas; la utilización del mínimo para calcular la certidumbre de una reglas supone una correlación implícita particular entre sus premisas, la cual suele ser falsa (ver la sec. 2.4.2). Las RR.BB. tienen la posibilidad de considerar las correlaciones entre las causas y las correlaciones entre los efectos, para ponderar adecuadamente las diferentes fuentes de evidencia.

10.3.4 Representación estructurada

Es frecuente mencionar la modularidad como una de las principales ventajas de los sistemas basados en reglas. Sin embargo, en la sección 2.4 vimos que, en cuanto al tratamiento de la incertidumbre conduce a inconsistencias y en cuanto a la construcción de la base de conocimiento plantea el problema de la validación de su contenido.

En cambio, las RR.BB. utilizan una representación estructurada en forma de red, con lo que la información está ordenada y fácilmente accesible, a la vez que mantienen también una forma particular de modularidad: podemos examinar por separado cada nodo, viendo cuáles son sus causas, sus efectos y la tabla de probabilidad condicional correspondiente. Es posible además crear herramientas gráficas para el desarrollo de la R.B., al estilo de IDEAL [220] o de HUGIN [4], que facilitan notablemente la labor. Realizar una modificación en una red bayesiana significa añadir o modificar los enlaces, los nodos o las tablas de probabilidad, sin que exista el peligro de caer en inconsistencias: basta ejecutar un sencillo algoritmo para comprobar que no se han introducido bucles.

Es cierto que en una R.B. todos los nodos (salvo los terminales) son equivalentes, pero es así solamente en cuanto al tratamiento matemático de la incertidumbre. En la sección 13.1 veremos cómo es posible que los nodos estén clasificados con respecto a diferentes criterios y cómo los enlaces pueden llevar varios tipos de etiquetas, de modo que la red puede poseer otras estructuras, además de la puramente causal.

10.3.5 Procesamiento distribuido

En el futuro, es de prever un gran desarrollo de la computación distribuida, con la implantación de máquinas dotadas de varios procesadores. A lo largo de las capítulos anteriores hemos tratado de mostrar cómo es posible conseguir implementaciones distribuidas para cada uno de los algoritmos que hemos ido describiendo: para la propagación de evidencia en poliárboles (sec. 5.1.4), para la puerta OR (sec. 5.2.4), para los algoritmos de agrupamiento (sec. 6.1.2.1), para la simulación estocástica (sec. 6.1.3), para el condicionamiento local (sec. 6.2.4.1), para el aprendizaje secuencial (sec. 7.2.6) y para nuestro modelo de explicación del razonamiento (sec. 8.3.3).

Podemos concluir diciendo que una de las ventajas de las RR.BB. es que permiten construir sistemas expertos completamente distruibuidos, en que tanto la inferencia como la inferencia como el aprendizaje y la explicación del razonamiento pueden implementarse en paralelo.

10.4 Limitaciones

Las ventajas de las RR.BB. son numerosas. La presente tesis las defiende con argumentos teóricos y las aplica en la construcción de un sistema experto. Sin embargo, sería falso afirmar que esta metodología carece de inconvenientes. Veamos cuáles son.

10.4.1 Complejidad de los algoritmos

En una implementación secuencial, el algoritmo para poliárboles requiere un tiempo proporcional al número de nodos (suponiendo que el número de valores y de padres

para cada variable está acotado); en una implementación paralela, el tiempo necesario es proporcional a la longitud del camino más largo dentro de la red. Sin embargo, se ha demostrado que el cálculo de probabilidad en redes con bucles es un problema NP, tanto si se trata de buscar una solución exacta [54] como aproximada dentro de cierto límite [59]. Por tanto, se supone —todavía no existe una demostración completa— que es imposible hallar un algoritmo cuya complejidad para este problema sea una función polinómica del tamaño de la red.

En términos prácticos esto significa que el tiempo necesario para computar la probabilidad en una red bayesiana crece de forma extraordinaria con el número de nodos y enlaces. Sabemos, más concretamente, que lo que hace aumentar la complejidad con tanta rapidez es el número y la longitud de los bucles.

Sin embargo, los bucles son algo de lo que no podemos prescindir, pues vienen dados por las relaciones de causalidad del mundo real. Cuando reconocemos que la adición de un enlace va a introducir un bucle de gran longitud sin modificar apenas la probabilidad del nodo de llegada, podemos decidir prescindir de dicho enlace, siendo conscientes de que estamos introduciendo una simplicación en el modelo. (Todo modelo es una simplificación de la realidad. Lo que se trata es de buscar un compromiso entre la simplicidad, para que sea tratable, y la fidelidad, para que sus predicciones sean acertadas.)

El problema es de gran importancia para que las redes puedan ser aplicadas a problemas de mayor tamaño, y por eso, se está investigando mucho actualmente sobre algoritmos que simplifiquen la computación aprovechando las particularidades de cada problema concreto.

10.4.2 Obtención del conocimiento

Una de las ventajas de las redes bayesianas frente a otros sistemas de razonamiento aproximado es que utilizan probabilidades numéricas, con lo cual permiten obtener no solamente resultados cualitativos sino también cuantitativos. El inconveniente es que debemos obtener un gran número de probabilidades a priori y condicionales que, aun siendo mucho menor que el requirido por el método probabilístico clásico, sigue siendo considerable.

Si tuviéramos bases de datos con gran cantidad de casos en que todas las variables estuvieran determinadas, resultaría sencillo obtener las probabilidades buscadas, mediante *apredizaje por lotes*, tal como explicábamos en el capítulo 7. Pero en la práctica esta situación ideal ocurre muy pocas veces, por lo que es necesario recurrir a la estimación subjetiva de expertos humanos. Podríamos aplicar entonces un algoritmo de *aprendizaje secuencial*, como el que hemos desarrollado en esta tesis, aunque tampoco ese método está libre de problemas; el principal de ellos es que necesitamos conocer el error (la desviación estándar) de la estimación subjetiva, lo cual resulta sumamente difícil.

Aun así, hay que decir que ésta no es una deficiencia de las RR.BB. sino una dificultad intrínseca de los problemas que estamos abordando. Hay otros métodos que no requieren tantos parámetros, incluso hay técnicas puramente cualitativas, pero ésta no es una verdadera solución al problema, sino que simplemente está menospreciando la información que resulta difícil de obtener. En cambio, los partidarios de las redes bayesianas preferimos utilizar números imprecisos antes que prescindir de ellos.

10.4.3 Razonamiento temporal

Si intentamos representar el conocimiento correspondiente al mundo real en forma de red bayesiana, encontramos con frecuencia una dificultad importante: la definición y los algoritmos nos prohíben trazar ciclos. Recientemente, Castillo, Gutiérrez y Hadi [32] han demostrado que cualquier conjunto consistente de probabilidades condicionales y/o marginales definidas pueden representarse mediante una red bayesiana, y por tanto, la imposibilidad de representar ciclos no constituye una limitación de éstas, a pesar de lo que se creía hasta entonces. Sin embargo, aún no está claro cómo podríamos convertir una red con ciclos en una red bayesiana. En nuestra opinión, éste no es un problema aislado, sino que debe ser tratado dentro del contexto del razonamiento temporal.

En efecto, cuando un sistema presenta realimentación, es muy importante tener en cuenta las constantes de tiempo del proceso. Si la interacción fuera instantánea, podrían darse varias situaciones:

1. El sistema llega a un punto de equilibrio instantáneamente, de modo que podríamos considerar el fenómeno simplemente como una transición entre dos estados, regida por una cierta probabilidad.

2. El sistema se dispara, llegando a valores extremos que implican su propia destrucción.

3. El sistema oscila con una frecuencia infinita, lo cual vendría a ser una forma de equilibrio dinámico.

Por tanto, el tratamiento de la realimentación nos parece inseparable del problema del razonamiento temporal.

Ésta es una cuestión que aún no está resuelta de forma completamente satisfactoria en el campo de las redes bayesianas, aunque en estos momentos se están realizando avances esperanzadores.

El primer método que se utilizó consistía en discretizar la variable *tiempo*; para cada instante t se construía una copia de la red original y se trazaban nuevos arcos hacia los nodos correspondientes al instante siguiente, $t+\Delta t$ [18, 58, 67, 128, 159].

Recientemente, Provan [183] ha propuesto tres modos de simplificar el problema. Uno de ellos consiste en eliminar de la red correspondiente al intervalo t los nodos terminales cuyo valor es desconocido en ese instante. Otro modo consiste en seguir solamente la evolución de algunas variables, que pueden seleccionarse según dos criterios alternativos: las que cambian en el tiempo o las variables observables.

Más interesante aún parece la propuesta de Dagum y Galper [57], que en vez de hacer una copia de la red para cada intervalo de tiempo, multiplica el número de nodos sólo para las variables que evolucionan con el tiempo. La ventaja principal es que cada variable puede tener un número de nodos diferente, dependiendo de cuál sea su ritmo de variación en el tiempo; es decir, cada variable posee su propio Δt.

El tema es muy interesante y es de esperar que se produzcan avances importantes en un futuro próximo, pues el razonamiento temporal es de gran importancia en muchos campos del diagnóstico, especialemente en muchas ramas de la medicina.

10.4.4 Limitación en el rango de aplicaciones

Hemos mencionado anteriormente que el fundamento causal es una de las grandes cualidades de las redes bayesianas. Sin embargo, esta característica representa un inconveniente cuando deseamos aplicarlas a un dominio en que las relaciones causales no son bien conocidas, aunque en este caso resulta difícil constuir un sistema experto cualquiera que sea la técnica empleada.

Dentro de las tareas que realizan los sistemas expertos (pág. 15) las RR.BB. están especialmente indicadas en problemas de diagnóstico, y por tanto también en las tareas relacionadas, tales como interpretación, predicción y monitorización. Sin embargo, en problemas de diseño, planificación, instrucción, reparación y control, las redes bayesianas encuentran serias dificultades, pues en ellos desempeñan un papel muy importante el razonamiento temporal y la toma de decisiones.

Del primer aspecto ya hemos hablado más arriba. En cuanto al segundo, los *diagramas de influencia*, de los que hablaremos en la sección 10.1.3, tienen mucho que aportar, pues son extensiones de las redes bayesianas que incluyen nodos-acción con el fin de poder aplicar los métodos de la teoría de la decisión, que es una ciencia apoyada en un sólido fundamento matemático.

10.4.5 Control del razonamiento

Hoy en día, una de las cuestiones menos estudiadas en el campo de las RR.BB. es *cómo generar preguntas*. En los sistemas basados en reglas, el tener una premisa de valor desconocido puede dar lugar a que se examine una regla o a que se solicite del usuario la información necesaria. Sin embargo, los algoritmos para el cálculo de probabilidad en RR.BB. propagan la evidencia disponible por toda la red, y son incapaces de averiguar qué hallazgos adicionales podrían resultar más interesantes de cara al diagnóstico.

En el sistema experto desarrollado en esta tesis este problema no era excesivamente importante, pues en el momento en que el médico está realizando el informe del ecocardiograma introduce toda la información de que dispone, por lo que no es fundamental la capacidad de generar preguntas. Sin embargo, ésta es una cuestión que deberá resuelta para mejorar la capacidad de nuestro sistema y para poder realizar aplicaciones a otros campos de la medicina.

A nuestro modo de ver, existen dos soluciones posibles al problema. Una de ellas consiste en aplicar la teoría de la decisión para medir el valor de la información en función del coste de obtenerla y de la utilidad que se espera obtener. Sin embargo, este método puramente matemático plantea el problema de la explosión combinatoria, por lo que incluso después de añadir hipótesis simplificadoras sigue siendo prácticamente inaplicable.

La otra solución consiste en aplicar métodos heurísticos que indiquen cómo generar la pregunta más oportuna dependiendo de la información recibida hasta ese momento. Esta heurística podría ser de carácter general, es decir, independiente del dominio aplicado, o dependiente de la aplicación concreta, de acuerdo con los protocolos que los expertos humanos tienen establecidos.

10.4.6 Explicación

Ya al comienzo del capítulo dedicado a este tema afirmábamos que aún no existe siquiera una teoría embrionaria que nos permita buscar una solución completa al problema. Tanto para sistemas heurísticos como para redes bayesianas, tenemos que conformarnos con modelos aislados que tratan de resolver, cada uno a su manera, aspectos particulares de casos concretos. El modelo que hemos propuesto supone un paso más en la misma dirección y deja abiertas varias vías para su posible desarrollo, aunque el resultado nos parece aún bastante pobre. Si queremos que las redes bayesianas —y los sistemas expertos en general— tengan una amplia aceptación en la vida cotidiana, debemos dotarles de una capacidad de explicación potente y amigable.

10.5 Conclusión

De las cualidades y deficiencias que hemos comentado se deduce que las redes bayesianas están especialmente indicadas para abordar los problemas de *diagnóstico* en que es posible construir un modelo *causal* y que requieren un tratamiento avanzado de la incertidumbre. La medicina es uno de los campos en que se cumplen estas condiciones.

En cuanto al tamaño del problema, los algoritmos exactos disponibles en la actualidad permiten calcular la probabilidad en redes de unos cientos de nodos, siempre que no haya demasiados bucles; para modelos más complejos, las RR.BB. pueden resultar ineficientes, salvo que se utilicen métodos aproximados adaptados a la naturaleza del problema.

En problemas en los que desempeñan un papel importante el *razonamiento temporal* y el *control de la inferencia* (plantear los objetivos y solicitar la información oportuna), es más aconsejable utilizar reglas en vez de RR.BB. Ejemplos de este tipo de problemas son el diseño, la planificación y el control automático. Quien construye un sistema experto debe decidir en cada caso qué metodología se adecúa mejor al problema que quiere resolver.

Parte III

APLICACIÓN A LA CARDIOLOGIA

Capítulo 11

Desarrollo del sistema experto DIAVAL

En la parte II hemos estudiado las redes bayesianas en sus aspectos generales. Vamos a describir ahora en esta tercera parte la aplicación de esos conceptos en el desarrollo de un sistema experto para ecocardiografía, denominado DIAVAL. En la sección 1.1 se planteó el problema que hemos tratado de resolver. En los próximos capítulos mostraremos las distintas etapas seguidas durante la construcción del sistema experto, desde los inicios hasta su implantación en un hospital.

11.1 Fases de desarrollo

Dado que la aparición de los sistemas expertos es un fenómeno relativamente reciente, aún no existe una metodología completamente establecida. Sin embargo, es habitual distinguir cinco fases o tareas en la construcción de un sistema experto [24, sec. 5.1],[246, cap. 12], aunque entendiendo que no se trata de un proceso lineal, sino de un crecimiento evolutivo en que la realimentación desempeña un papel esencial. Las cinco fases son las siguientes (ver fig. 11.1):

1. **Identificación.** Consiste en definir el problema, determinar los objetivos y estimar los recursos necesarios para concluir el proyecto: participantes y colaboradores, medios materiales, tiempo de realización, etc.

2. **Conceptualización.** En esta fase, el diseñador del sistema recoge información sobre el tipo de problemas que va a resolver el sistema experto, el método que siguen los especialistas humanos y los conceptos que aparecen en el razonamiento. Esta tarea se conoce como *adquisición del conocimiento*; algunos de sus aspectos son independientes del tipo de sistema experto que vamos a construir, mientras que otros dependen de si vamos a utilizar reglas o redes bayesianas.[1]

3. **Formalización.** Una vez que tenemos el conocimiento necesario, hay que estructurarlo de modo que pueda ser introducido en un computador. Para ello, debemos escoger un *método de representación del conocimiento*, como pueden ser

[1] Según la formulación original que aparece en [24], las cinco tareas que estamos describiendo aquí se refieren a la adquisición del conocimiento. Sin embargo, nos parece que es en esta segunda fase donde se obtiene la información necesaria, generalmente mediante la interacción con uno o varios expertos; las etapas posteriores se ocupan de hacer computable el conocimiento ya recopilado.

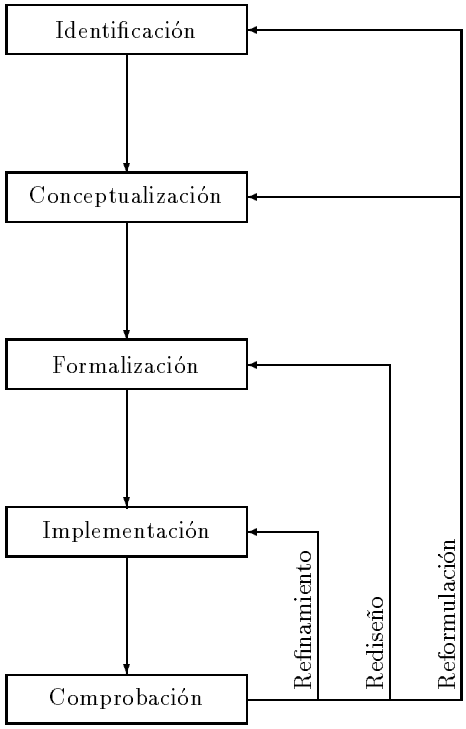


Figura 11.1: Fases de desarrollo de un sistema experto (tomado de [24, pág. 139]).

las reglas, los marcos, las redes o una combinación de ellos (ver la sec. 2.2). En una situación ideal, la representación debe venir dictada únicamente por las características del dominio de aplicación; sin embargo, cuando se va a utilizar un programa comercial, es necesario tener en cuenta ya en esta fase las posibilidades que ofrece la herramienta y las consiguientes limitaciones.

4. **Implementación.** Consiste en programar sobre el ordenador las funciones y procedimientos que van a realizar la *inferencia* (la utilización de una herramienta puede hacer esta tarea mucho más fácil o incluso innecesaria) y de codificar la *base de conocimiento* diseñada anteriormente. Puede incluirse también en esta fase el desarrollo del *interfaz* de usuario y del *entorno* de apoyo: bases de datos, gestión de ficheros, control de la impresora, etc.

5. **Comprobación.** En la última fase el sistema es puesto a prueba con el fin de corregir los posibles errores e incluso rediseñar la base de conocimiento o reformular el problema (fig. 11.1). Además de estas comprobaciones periódicas para examinar incrementalmente el desarrollo del sistema, es habitual realizar una *evaluación* global del comportamiento del programa, tal como se describe en el capítulo 14.

Insistimos de nuevo en que estas fases deben ser entendidas como tareas o procesos dentro de un ciclo evolutivo más que como pasos independientes. Por eso es esencial la realización temprana de un pequeño prototipo que permita evaluar el sistema lo antes posible, para evitar tener que hacer modificaciones esenciales cuando el sistema haya crecido demasiado.

En el sistema experto DIAVAL se han seguido aproximadamente estas cinco fases, según se explica a continuación:

1. **Identificación.** Hubo inicialmente dos o tres reuniones con un grupo de médicos para definir el problema que se quería abordar, según indica la sección 1.1. Posteriormente se comprobó que el proyecto inicial resultaba demasiado ambicioso, por lo que fue necesario redefinir los objetivos. El resultado se concretó en las especificaciones que mencionaremos más adelante en este capítulo como "Funcionalidades del sistema" (sec. 11.2.2).

2. **Conceptualización.** El punto de partida de nuestro trabajo consistió en el estudio de la estenosis mitral, un tema suficientemente reducido para poder empezar y suficientemente complejo como para que no resultase trivial. La elección resultó acertada, pues este tema tan concreto sirvió para que surgieran ya todos los conceptos básicos sobre el diagnóstico médico (sec. 12.1.2) e incluso para indicar la conveniencia de representar el conocimiento médico en forma de red causal. Posteriormente, abordamos el estudio de otras valvulopatías y de las demás alteraciones cardíacas descritas en el capítulo 12.

3. **Formalización.** Nuestra idea inicial suponía utilizar el método estándar para representar el conocimiento: las reglas de producción. Sin embargo, éstas resultan inadecuadas para razonar a partir de modelos causales, especialmente cuando hay fuentes importantes de incertidumbre. Por ello se hizo necesario buscar un formalismo más adecuado, y fue así como llegamos a las redes bayesianas. En consecuencia, la base de conocimiento de DIAVAL viene dada por

un modelo cardiológico causal y probabilístico; el capítulo 12 resume el conocimiento contenido en el sistema experto, mostrando algunos ejemplos de cómo se ha representado en forma de red bayesiana.

4. **Implementación.** Para la primera versión de nuestro sistema, hemos escogido una herramienta comercial *GoldWorks*, por las facilidades que ofrece para trabajar con marcos y para desarrollar interfaces gráficos y con bases de datos. Al hablar de las redes bayesianas, el capítulo 9 estuvo dedicado a la implementación de los algoritmos generales, independientes del dominio (inferencia, aprendizaje y explicación); en esta parte III, el capítulo 13 trata los aspectos de la implementación específicos del sistema experto DIAVAL.

5. **Comprobación.** El capítulo 14 estará dedicado a la evaluación de nuestro sistema experto, principalmente mediante la elaboración de un cuestionario, la observación y el diálogo con los expertos.

11.2 Identificación del problema

11.2.1 Clases de enlaces

En la sección 1.1 vimos que el proyecto inicial consistía en un sistema capaz de analizar los ecocardiogramas y realizar un diagnóstico (fig. 1.1). Posteriormente se dividió en dos partes: un sistema de visión para el análisis de las imágenes y un sistema experto de diagnóstico; esta segunda parte constituye el objeto del proyecto que estamos describiendo en la presente tesis doctoral (fig. 1.2).

La primera cuestión que debemos plantearnos es la utilidad que puede tener un sistema experto en este campo. En muchos centros sanitarios, el servicio de ecocardiografía carece de ordenadores, y los informes se escriben a mano o a una máquina. En el Hospital de la Princesa, en cambio, existe un programa de ordenador, realizado hace diez años por un médico, que ofrece importantes ventajas: tratamiento matemático de los parámetros (cálculo de nuevos parámetros, porcentajes de desviación sobre los valores esperados, etc.), gestión de una base de datos, e impresión de un informe con los resultados.

El programa ha resultado sumamente útil durante este tiempo y sigue utilizándose hasta el día de hoy. Sin embargo, en la actualidad se ha quedado obsoleto. Por un lado, el antiguo programa se ejecuta sobre un ordenador HP-86B, cuyo sistema operativo, incluyendo el formato de los disquetes, es incompatible con los ordenadores actuales. Por otra, el programa está escrito en BASIC compilado, con lo que su actualización resulta dificultosa, la ejecución es lenta y el interfaz es rudimentario, pues se basa en la introducción de códigos numéricos, que el médico debe consultar en un archivador tradicional.

Por tanto, resultaba necesario, o al menos conveniente, sustituir el antiguo programa por otro que funcionara sobre los actuales ordenadores personales, rápidos y económicos, que utilizara una base de datos compatible con otras aplicaciones, que aprovechara los entornos gráficos basados en ventanas y en el uso del ratón, y que no requiriese la memorización o consulta de unos códigos.

11.2.2 Objetivos: funcionalidades del sistema

Sin embargo, nuestro objetivo no se limita a construir una base de datos mejorada, sino que pretendemos aplicar las posibilidades que ofrece la inteligencia artificial en el campo del diagnóstico médico. Nuestros objetivos principales son:

Tratamiento avanzado de datos: En este punto, la primera tarea consiste en la *comprobación* de la información que introduce el usuario: ver que el nombre y apellidos son cadenas de caracteres, que la edad y el peso son números, etc. Conviene también tener en cuenta cuál es el rango posible y el rango habitual, para rechazar los datos absurdos y pedir confirmación ante los que resulten extraños. Deseamos además ofrecer una *interpretación* de los valores medidos, señalando si corresponden a intervalos de normalidad, alteración leve, moderada o severa, teniendo en cuenta la edad y las características del paciente. En este sentido, es muy importante estudiar la *relación* de unos datos con otros para detectar inconsistencias y deducir nueva información.

Diagnóstico: Es nuestro objetivo principal. En todas las ramas de la medicina, y especialmente en la cardiología, es importante combinar distintas fuentes de información, cada una de ellas con sus propias causas de incertidumbre. Intentamos construir un sistema basado en *razonamiento causal probabilístico*, que tenga en cuenta conceptos como prevalencia, sensibilidad y especificidad. La capacidad de *explicación* es esencial para la aceptación por los médicos, y es deseable también que el sistema experto pueda *aprender* de los casos diagnosticados.

Informe: El sistema debe ser capaz de elaborar por sí mismo un informe en que se recojan los datos más relevantes y las conclusiones del diagnóstico. Aunque el médico tendrá siempre la posibilidad de modificarlo cuando lo considere oportuno, conviene que el informe sea acertado en el mayor número de casos posible, para que el médico no tenga que perder tiempo corrigiéndolo.

Gestión de la base de datos: El sistema experto ha de realizar, por lo menos, las mismas funciones que el antiguo programa: almacenamiento, búsquedas y estadística. Sería deseable incluir además técnicas avanzadas de recuperación y tratamiento inteligente de la información.

Respecto a los objetivos del proyecto, debemos señalar que, en cierto modo, los intereses de los médicos eran contrapuestos a los nuestros. Para ellos, lo más importante era tener una buena base de datos con un interfaz amigable y robusto, mientras que para la presente tesis doctoral, ésos son aspectos secundarios. En cambio, la parte central de nuestro proyecto, el diagnóstico ecocardiográfico, a algunos médicos les parecía irrealizable, poco útil e incluso peligroso, tal como comentaremos más adelante.

11.2.3 Viabilidad y oportunidad del proyecto

Según Donald Waterman [246, cap. 11], antes de realizar un sistema experto hay que asegurarse de que el proyecto es *posible*, está *justificado* y es *apropiado*. Veamos cada uno de estos puntos por separado.

Es posible

En nuestro caso se cumplen las características que Waterman señala para garantizar la viabilidad del sistema experto, y son las siguientes:

- Hay auténticos expertos en el área, son capaces de explicar su modo de razonamiento y suelen estar de acuerdo entre ellos en sus conclusiones.

- La tarea de diagnóstico a partir del ecocardiograma se basa solamente en capacidades cognitivas, no requiere capacidad sensorial ni habilidad manual. No es necesario aplicar *sentido común* tal como se entiende en inteligencia artificial; es decir, no es imprescindible razonar sobre relaciones espaciales y temporales, ni sobre los conocimientos e intenciones de las personas, etc.

- El diagnóstico en cardiología no es excesivamente complicado; el conocimiento está relativamente bien estructurado y no tiene una extensión desmesurada. De hecho, veremos en capítulos posteriores cómo podemos construir una red que incluya la mayor parte del conocimiento relevante para nuestra tarea.

Está justificado

Para que el desarrollo de un sistema experto esté justificado, hace falta que los beneficios que se esperan obtener compensen el esfuerzo realizado. En nuestro caso, dado que se trata de una tesis de investigación, no es necesario tener en cuenta la rentabilidad económica; basta considerar la *utilidad* del sistema desde el punto de vista de su implantación en un servicio hospitalario.[2] Como hemos comentado anteriormente, para los médicos resultará útil cualquier base de datos que tenga un interfaz amigable. Sin embargo, nos interesa juzgar la utilidad del sistema de diagnóstico, que constituye el componente central de nuestro proyecto.

Para un ecocardiografista veterano, con muchos años de experiencia, la utilidad de un sistema de diagnóstico será bastante reducida. En cambio, para los residentes que realizan su rotación por el Servicio de Ecocardiografía, el sistema experto podrá resultar útil en los siguientes aspectos:

- Interpretación de los datos numéricos (medidas sobre el ecocardiograma), extrayendo información cualitativa de ellos y ofreciendo ayuda dinámica sobre rangos e intervalos.

- El hecho de solicitar la información de modo ordenado y teniendo en cuenta los antecedentes y los síntomas del paciente impide que pueda pasarse por alto información relevante. La sistematización en la entrada de datos y en la elaboración del informe cumple así una labor pedagógica.

- El sistema puede actuar como *consejero* o *tutor* en cuanto al razonamiento causal, evitando algunos de los errores que comentábamos en el capítulo 3. En este sentido es especialmente útil la capacidad de explicación, tanto del modelo

[2] Waterman ofrece unos criterios de justificación basados en la sustitución de un experto humano por un sistema experto. Sin embargo, en los últimos años, la inteligencia artificial en general y los sistemas expertos en particular están orientados a colaborar con el ser humano más que a competir con él. Por esta razón, hemos decidido considerar aquí unos factores de rendimiento distintos de los que propone este autor.

como del razonamiento. Incluso cuando el sistema ofrezca un diagnóstico erróneo puede resultar provechoso para el usario investigar dónde y por qué ha fallado.

- La capacidad de diagnosticar ciertas enfermedades poco frecuentes (por ejemplo, ciertas enfermedades congénitas), considerando su prevalencia, sus causas y sus efectos, ordenados según su importancia relativa, puede ayudar en algunos casos difíciles y, sobre todo, para evitar que se pasen por alto ciertas enfermedades. En este sentido, es importante la capacidad del modelo probabilístico para sintetizar información procedente de fuentes diversas: historia clínica, electrocardiograma, diferentes tipos de ecocardiograma, etc.

Es apropiado

Por último, cabe preguntarse si no se podía haber construido un sistema tradicional, basado en programación algorítmica, que hubiera resultado mucho más reducido de tamaño y más eficiente que un sistema experto. La alternativa natural sería considerar un árbol de decisión, mejorado con técnicas bayesianas, al estilo —por ejemplo— de Reggia y Perricone [186].

Sin embargo, hemos considerado necesario utilizar una red bayesiana por las siguientes razones:

- Una red bayesiana puede diagnosticar múltiples enfermedades que se encuentren presentes simultáneamente, mientras que los métodos algorítmicos suelen conducir a un único diagnóstico.

- La separación entre base de conocimiento (en nuestro caso, la red) y el motor de inferencia (los algoritmos de propagación), común a todos los sistemas expertos, facilita la depuración y ampliación del modelo, a la vez que permite generar explicaciones —tanto del modelo como de la inferencia— y posibilita cierta capacidad de aprendizaje.

- Una red bayesiana permite realizar varios tipos de inferencia: deductiva, abductiva, intercausal y predictiva, mientras que los modelos algorítmicos basados en árboles de decisión sólo permiten un tipo de inferencia.

- Por estas razones, que han sido explicadas con más detalle en la parte II, decidimos que las ventajas de realizar un sistema experto compensan ampliamente el esfuerzo invertido.

11.3 Obtención del conocimiento

Es frecuente encontrar la afirmación de que la obtención del conocimiento supone un "cuello de botella" en la realización de un sistema experto, en el sentido de que ésta es la fase que mayor tiempo consume. Esto puede ser cierto en el caso de un sistema basado en reglas, especialmente en el caso de utilizar una herramienta comercial que lleve un motor de inferencia incorporado y ofrezca facilidades para la creación del interfaz.

En nuestro proyecto, hay otras fases que han consumido bastante más tiempo. Ha sido necesario, por un lado, un estudio teórico de las redes bayesianas: representación del conocimiento, propagación de evidencia, modelo de explicación, modelo de aprendizaje, etc., y la implementación de los algoritmos correspondientes. Por otro lado, la creación del interfaz ha supuesto un esfuerzo comparable o quizá superior al invertido en la obtención del conocimiento. Sin embargo, hay que señalar que esto ha sido así por tratarse de un proyecto de investigación en que hemos creado un sistema experto bayesiano casi "desde la nada". Si tuviéramos que realizar ahora una nueva aplicación para otro campo de la medicina, podríamos aprovechar gran parte o la totalidad de los algoritmos de inferencia y de las funciones definidas para el interfaz; el esfuerzo se concentraría sobre todo en la obtención del conocimiento, haciendo cierta la afirmación de que, en general, ésta es la fase de desarrollo que más tiempo consume.

11.3.1 Metodología

En primer lugar, conviene recordar que las redes bayesianas implican un *razonamiento basado en modelo*, a diferencia de la mayor parte de los sistemas expertos, que utilizan *razonamiento heurístico*. Esto significa que, en la adquisición del conocimiento, no es tan importante observar cómo diagnostican los expertos humanos para deducir las reglas y las estrategias (meta-reglas) que utilizan, sino construir un modelo —en nuestro caso un modelo patofisiológico— suficientemente sencillo para que sea manejable y suficientemente amplio para que sea competente. Por eso, muchos de los consejos que aparecen en los libros sobre adquisición del conocimiento (por ejemplo, observar cómo resuelve el experto humano unos casos preseleccionados) resultan poco útiles a la hora de construir una red bayesiana.[3] Las dos únicas refencias que hemos encontrado sobre cómo realizar la construcción de una red bayesiana a partir de la interacción con expertos humanos son el trabajo de Spiegelhalter et al. [215] y el de Luis Enrique Sucar et al. [226].

En el caso de un sistema experto bayesiano, podemos dividir la adquisición del conocimiento en dos fases. La primera de ellas consiste en recopilar la *información cualitativa*, es decir, es identificar las anomalías y los datos relevantes y formar una red causal con las relaciones entre ellos. La segunda fase se ocupa de recoger la *información cuantitativa*: las probabilidades a priori y las probabilidades condicionales.

Para nuestro proyecto hemos utilizado dos **fuentes de información** de muy distinta naturaleza: los cardiólogos y los libros de cardiología. Cada una de ellas da lugar, obviamente, a un tipo de interacción muy diferente.

Los libros

Empezamos hablando sobre los libros, no porque sean más importantes en cuanto fuente de información, sino porque la interacción con ellos resulta mucho más sencilla.

[3] Una revisión de los métodos habituales puede encontrarse en la referencia [114], que apareció en un número especial de la revista *SIGART Newsletter* dedicado a la obtención del conocimiento. También la revista *IEEE Transactions on Systems, Man and Cybernetics* dedicó un número especial (vol. 19, no. 3, 1989) al mismo tema. En ambos casos, los artículos incluidos suponen explícita o implícitamente que se va a construir un sistema basado en reglas.

La principal ventaja de trabajar con libros es la disponibilidad: el diseñador de la base de conocimiento puede consultarlos cualquier día, a cualquier hora y en cualquier lugar, sin límite de tiempo. Otra ventaja es que ofrecen la información sistematizada y aportan algunos datos estadísticos precisos, lo cual resulta fundamental a la hora de construir una red bayesiana.

El principal inconveniente de los libros es que no son interactivos: el diseñador no puede pedirles aclaraciones, y fácilmente acaba encontrándose perdido entre tal cantidad de información, sin saber diferenciar lo esencial de lo secundario. Por este motivo, es importante empezar consultando un libro básico (en nuestro caso hemos utilizado el libro de Ferreira [89]), despúes un libro más avanzado (por ejemplo, el de Sokolow y McIlroy [212]) y tener además un libro de consulta para buscar información específica (tal como el de E. Braunwald [21]). La obra de Monaghan [152], dedicada específicamente a la ecocardiografía, trata de forma breve y concisa los aspectos de esta técnica, incluyendo el estudio Doppler. Además, resulta muy útil tener siempre a mano un diccionario de términos médicos [70].[4]

Los médicos

Por mucho interés que ponga el diseñador del sistema experto en consultar los libros sobre el tema de su trabajo, está claro que nunca podrá llegar a ser un especialista, pues ello requeriría largos años de estudio y de experiencia. Además, no es éste el objetivo que se persigue: conviene que el diseñador posea un amplio conocimiento de la materia en cuestión, sobre todo para facilitar el diálogo con los expertos, pero no es necesario que él mismo se convierta en especialista. Incluso cuando el conocimiento ha sido extraído de los libros, siempre es necesario que sea supervisado por un verdadero experto, para corregir los errores y para discernir lo esencial de lo irrelevante.

La mayor ventaja de trabajar con expertos humanos es que el diseñador puede plantearles las preguntas que le interesan y seguir dialogando hasta que se aclaren todas las dudas. En el campo de las redes bayesianas, es muy conveniente contar con especialistas que tengan varios años de experiencia, pues muchos de los factores probabilísticos que se requieren no aparecen en los libros, ni siquiera en los artículos de revistas especiliazadas, por lo que es necesario recurrir a estimaciones subjetivas.

En cambio, el peor inconveniente es la disponibilidad. No siempre es fácil contar con expertos motivados para trabajar en un proyecto, e incluso en el caso de que lo estén, suelen tener tal cantidad de trabajo que resulta sumamente difícil concertar las reuniones con ellos. Por este motivo es provechoso el consejo clásico de grabar en cinta magnetofónica las entrevistas, con el fin de extraer luego detenidamente toda la información relevante. Existen además unos factores humanos que hay que tener muy en cuenta en la relación con los médicos para evitar prejuicios y recelos; hablaremos de ellos en seguida.

[4] Citamos aquí estos libros simplemente porque son los que hemos utilizado en nuestro proyecto; obviamente, una lista de libros recomendados sobre cardiología sería muy diferente.

11.3.2 Dificultades inherentes al campo de la medicina

Obtención de la información cualitativa

Una de las primeras dificultades que encuentra quien desea construir un sistema experto para medicina es aprender la terminología. Éste no es un problema grave y se resuelve en gran medida con la ayuda de un diccionario especializado, aunque hay conceptos básicos que no aparecen en los libros. La sección 12.1.2 recoge algunos de ellos.

Otro problema que sorprende al diseñador, especialmente si ha recibido su formación en el campo de las ciencias exactas, es la falta de estructura en el conocimiento médico. Quizá haya en ello una parte de culpa de los propios especialistas, pero hay que reconocer que el cuerpo humano es incomparablemente más complicado que cualquiera de los sistemas que estudian la física, las matemáticas o la ingeniería. Las clasificaciones resultan difíciles porque los límites suelen ser difusos y por las numerosas excepciones e interrelaciones que aparecen. Para complicar más la situación, hay muchas enfermedades cuyas causas y evolución están aún por estudiar.

En consecuencia, es frecuente la falta de acuerdo entre expertos, no sólo en cuanto al diagnóstico y al tratamiento, sino también en cuanto a la descripción y clasificación de las enfermedades. Por ejemplo, para E. Braunwald [21], la *insuficiencia mitral* y el *prolapso mitral* son dos entidades diferentes. Para Sokolow y McIlroy [212, págs. 396–397], hay dos tipos de insuficiencia mitral: una que es hemodinámicamente insignificante —es decir, que no afecta a la circulación normal de la sangre—, indicando entre paréntesis que se trata del prolapso mitral, y una insuficiencia significativa, debida a otras causas. Para Ferreira [89, pág. 148], sin embargo, el prolapso es uno de los rasgos que puede presentar la insuficiencia mitral.

Ante las contradicciones que se dan entre diferentes especialistas, e incluso de uno de ellos consigo mismo, el diseñador del sistema experto fácilmente se siente confundido y, en ausencia de un criterio de referencia definitivo, debe buscar un consenso razonable dentro de toda la información que recibe.

Obtención de la información cualitativa

Si ya la adquisición del conocimiento cualitativo puede resultar complicada, mucho más lo es la obtención de los datos numéricos. Por más que revisemos la bibliografía médica, difícimente vamos a encontrar la información que necesitamos, pues las descripciones que aparecen en la literatura son casi siempre cualitativas. Veamos como ejemplo la siguiente afirmación extraída de un libro especializado [152, pág. 93]:

> El tumor primario *más común* en el corazón *adulto* es el mixoma y *el 75%* de ellos se localiza en la aurícula izquierda, *habitualmente* en mujeres. [Cursiva añadida.]

En esta breve cita, aparecen dos términos difusos, *adulto* y *habitualmente*. Esto nos plantea varios interrogantes: ¿Desde qué edad se considera a una persona como adulta? ¿Distingue entre *adultos* y *ancianos* o los engloba a todos en el mismo grupo? ¿Qué frecuencia debemos entender por *habitualmente*? Los estudios psicológicos citados en la sección 3.2 pueden ofrecer una cierta ayuda a la hora de convertir las expresiones cualitativas en probabilidades numéricas, aunque las variaciones en las asignaciones son excesivamente amplias, según explicábamos allí. Por otro lado, además

de los términos difusos, aparece una proposición comparativa, "*más común*", que no indica cuáles son las prevalencias de los tumores, ni siquiera la proporción relativa entre ellas.

La existencia de un número concreto, "*el 75%*", tampoco es de gran ayuda. ¿Se trata de un número exacto medido experimentalmente o es una estimación subjetiva? Aun suponiendo que sea un resultado experimental, tampoco podemos introducir este dato directamente en nuestra red bayesiana, porque no expresa una probabilidad condicional entre causas y efectos. Es más, aunque se tratara de la probabilidad o frecuencia de una causa condicionada sobre el efecto, como suele aparecer en la bibliografía médica ("el 70% de las estenosis mitrales son de origen reumático"), el problema sería el mismo: necesitamos conocer la probabilidad del efecto condicionada en sus causas o —en el caso de la puerta OR— la probabilidad de que la causa produzca el efecto: "La fiebre reumática produce estenosis mitral en el 18% de los casos" éste es un dato que aparece alguna vez en la literatura, pero mucho menos frecuentemente de lo que desearíamos.

Sin embargo, el "75%" mencionado no es una información que debamos despreciar. La razón es que nuestra red bayesiana puede deducir la probabilidad de que el tumor sea un mixoma, y el resultado debe coincidir con el que hemos encotrado en el libro. Si no es así, tendremos que revisar las probabilidades obtenidas previamente —recordemos que generalmente éstas proceden de estimaciones subjetivas— hasta llegar a una concordancia.

Por último, incluso contando con resultados numéricos procedentes de estudios estadísticos, la disparidad entre distintos investigadores puede ser sorprendente. Así, hay quien afirma que el prolapso mitral se da en el 20% de las mujeres jóvenes, mientras que otros sitúan esta proporción en el 1 ó el 2%. La diferencia se debe a la técnica y al criterio empleados. A la hora de construir una red bayesiana debemos preguntar a nuestros expertos cuál es la estimación más razonable en función del criterio que decidamos establecer.

11.3.3 Factores humanos

A parte de las dificultades que plantea el conocimiento médico, el diseñador de un sistema experto debe luchar también contra los prejuicios de los especialistas que van a colaborar en el proyecto. Sus objeciones más frecuentes son las siguientes:

- "Es imposible construir un sistema experto porque para un ordenador 2 y 2 siempre son 4, mientras que en medicina no hay reglas fijas."

- "La cardiología (o cualquier otra rama de la medicina) es un campo muy complejo, por lo que resultaría imposible introducir tal cantidad de información en un ordenador."

- "En el diagnóstico desempeña un papel muy importante la intuición del médico, mientras que un ordenador nunca podrá tener intuición."

- "Aunque fuera posible tener un sistema experto, ¿para que quiero yo una máquina que haga mi trabajo? Me basta con tener una base de datos que me facilite la labor burocrática."

Estas objeciones de los médicos, conscientes o inconscientes, podrían resumirse así: "Es imposible construir un sistema experto para mi especialidad. Y si llegara a construirse sería mi enemigo."

Conviene que el diseñador del sistema experto explique detalladamente a sus médicos colaboradores que algunas de estas objeciones son falsas y que otras, aun teniendo una parte de verdad, no constituyen un obstáculo insalvable. En este sentido, hay dos comparaciones que nos han ayudado en el diálogo con los especialista de nuestro proyecto:

Metáfora del teléfono: Podemos imaginar la interacción del médico con el sistema experto como si se tratara de la consulta por teléfono con otro médico más veterano. De hecho, en la práctica diaria los médicos residentes utilizan con frecuencia el interfono para consultar a los adjuntos, e incluso médicos experimentados consultan a sus colegas de otras especialidades.

La metáfora del teléfono demuestra que, al menos en principio, es posible realizar una consulta a un sistema experto de modo que éste, basándose en la información recibida y en la que solicita, puede ofrecer consejos provechosos aun sin tener un contacto directo con el paciente.[5]

Metáfora del libro: Para un médico puede ser muy útil tener a mano un libro especializado, sobre todo a la hora de consultar datos numéricos. La ventaja de un sistema experto frente a un libro es que muestra la información de forma dinámica y adaptada al caso que estamos considerando.

Estas dos metáforas pueden ayudarnos a convencer a nuestros médicos colaboradores de que la realización del sistema experto es posible, de que el sistema va a ser útil y de que no supone una amenaza para su puesto de trabajo, del mismo modo que la consulta mediante interfono o la consulta de libros son ayudas para su trabajo en vez de ser sus rivales.

Es importante que el diseñador del sistema experto no descuide los factores humanos en su interacción con los médicos, para eliminar los prejuicios y conseguir que éstos colaboren de buena gana durante toda la realización del proyecto.

Afortunadamente, tenemos la impresión subjetiva de que estos prejuicios desaparecieron por completo en el momento en que los expertos se sentaron frente al sistema para realizar su evaluación. La inteligencia artificial dejó de ser algo desconocido y amenazador, al observar un programa de ordenador capaz de ofrecerles información valiosa y de realizar diagnósticos e informes a pesar de sus limitaciones; en resumen, ahora han comprobado que el sistema experto puede ser más o menos útil como ayuda en su trabajo, pero nunca podrá ser un competidor.

[5] La *prueba del teléfono* fue propuesta como método sencillo para examinar la viabilidad de un sistema experto: en general, dado un problema, la realización de un sistema experto es posible si y sólo si la consulta telefónica con un especialista pudiera llevarnos a la solución del problema. Así se pueden excluir, por ejemplo, los problemas que requieren una interacción gráfica compleja o destreza manual.

Capítulo 12

Conocimiento médico cardiológico

Los dos elementos esenciales de un sistema experto son el conocimiento que contiene y las técnicas de inferencia que utiliza. En los capítulos anteriores, al estudiar las redes bayesianas, hemos tratado ya extensamente el primero; ahora nos vamos a ocupar del segundo, es decir, de la *representación computacional del conocimiento médico* que permite a nuestro sistema DIAVAL realizar diagnósticos ecocardiográficos.

Para ello, estudiamos en primer lugar los conceptos básicos comunes a todas las ramas de la medicina, insistiendo sobre todo en el diagnóstico, y luego pasaremos a describir muy brevemente el caso concreto de la cardiología.

12.1 Proceso de razonamiento en medicina

12.1.1 Diagnóstico, pronóstico y tratamiento

El tratamiento médico de una enfermedad puede dividirse en tres fases, tal como muestra la figura 12.1. La primera de ellas, el **diagnóstico**, consiste en averiguar cuál es la enfermedad que padece el paciente, describiendo los mecanismos implicados (lo que se conoce como *patofisiología*). La segunda fase consiste en emitir un **pronóstico** sobre la evolución previsible de la enfermedad: esperanza y calidad de vida para el paciente, posibles complicaciones, etc. Por último, hay que aplicar un **tratamiento terapéutico**, que puede ser *médico* —es decir, basado en fármacos— o *quirúrgico*; en esta tercera fase es esencial realizar un *seguimiento* del paciente para ver si el enfermo responde adecuadamente al tratamiento.

Sin embargo, este esquema tan simple se complica por la presencia de bucles de realimentación, más o menos importantes dependiendo de cada especialidad médica. Así, si un paciente no responde al tratamiento, empeorará el pronóstico y tal vez sea necesario incluso corregir el diagnóstico inicial.

A nosotros nos interesa aquí sobre todo el problema del *diagnóstico*, donde las redes bayesianas demuestran sus mejores cualidades (ver cap. 10).[1] Según Ledley

[1] Como desarrollo posterior, podríamos extender nuestra base de conocimiento de modo que sea capaz de realizar también pronósticos y recomendaciones de terapia, aunque este último punto resultaría más difícil, pues un tratamiento eficaz debe tener en cuenta la evolución del paciente, para lo que se hace necesario utilizar razonamiento temporal.

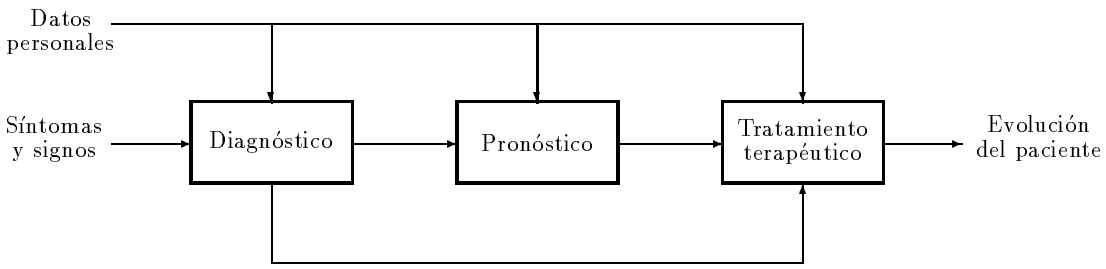


Figura 12.1: Proceso básico de razonamiento en medicina.

y Lusted [136], si pidiéramos a un médico que explicara cómo realiza el proceso de diagnóstico, podría respondernos algo así:

> Primero, obtengo los hechos relativos al caso a partir de la historia clínica, la exploración física y las pruebas de laboratorio. En segundo lugar, evalúo la importancia relativa de los distintos signos y síntomas. Algunos de los datos pueden ser primordiales, otros tienen menor importancia. En tercer lugar, para realizar un diagnóstico diferencial, hago una lista de las enfermedades que guardan una semejanza razonable con el caso considerado. Después voy excluyendo una a una las enfermedades de la lista, hasta que queda claro que el caso encaja en una categoría de enfermedad definida, o que se trata de una enfermedad entre varias posibles, o bien que no se puede determinar su naturaleza exacta.

Aunque esta descripción es demasiado simple, tal como explican los propios autores, puede resultarnos útil para entender que el diagnóstico consiste en un proceso de *refinamiento*: a medida que vamos obteniendo nueva información, acotamos cada vez más el conjunto de los posibles diagnósticos.

Veremos en la próxima sección cuáles son los conceptos claves que influyen en el proceso de diagnóstico. Es importante que el diseñador de un sistema experto médico esté familiarizado con ellos, tanto si va a obtener el conocimiento mediante un diálogo con los especialistas como si lo va a extraer de libros de medicina.

12.1.2 Conceptos básicos sobre diagnóstico

El diagnóstico puede considerarse como un proceso cuya entrada viene dada por los datos disponibles y su salida por las posibles enfermedades o anomalías que queremos diagnosticar.

La información de **entrada** para el sistema de diagnóstico médico se compone de los *datos personales* del enfermo (sexo, edad, peso, estatura, profesión), su *historia clínica* (enfermedades sufridas en pasado, factores hereditarios y demás antecedentes) y los *hallazgos*, que pueden ser de dos tipos:

Síntomas: manifestaciones que el enfermo experimenta como consecuencia de una alteración; son, por tanto, subjetivos, como el dolor o la disnea (sensación de ahogo al respirar).

Signos: manifestaciones objetivas que el médico reconoce o provoca. La cianosis (coloración azul-violeta), por ejemplo, es un signo que se observa espontáneamente; la percusión, en cambio, estudia un ruido provocado por el médico.

Los signos proceden de la exploración física o de las pruebas complementarias, según se explicará más adelante.

Se conoce como *síndrome* un conjunto de síntomas y signos que suelen aparecer conjuntamente, en general sin una causa conocida. (Cuando hay una causa conocida es más frecuente hablar de enfermedad que de síndrome.)

El espacio de **salida** de un sistema de diagnóstico viene dado por las anomalías que queremos diagnosticar. Entre ellas, consideramos *entidades clínicas* las que producen síntomas y/o signos; si no, diremos que se trata solamente de *entidades anatómicas*.[2] También podemos distinguir, de forma muy simplificada, las anomalías *orgánicas*, es decir, aquéllas en que hay una alteración anatómica, de las meramente *funcionales*; así, una "insuficiencia valvular funcional" significa que la válvula no sufre alteraciones anatómicas pero no cierra completamente debido a un exceso de presión.

Según su *origen*, las enfermedades se clasifican en *congénitas* y *adquiridas*. Una enfermedad adquirida puede ser *primaria*, si no se debe a otra alteración previa, o *secundaria* a otra alteración. Así, una pericarditis provocada por una infección viral es primaria, aunque existe también, por ejemplo, una pericarditis secundaria a infarto agudo de miocardio. Una alteración primaria es *idiopática* cuando no se conoce la causa que la ha provocado. Dentro de las enfermedades adquiridas, se consideran *degenerativas* las que aparecen como consecuencia del paso del tiempo. Son alteraciones *yatrogénicas* las que provocan el médico o los medicamentos en su intención de curar.

En cuanto a su *desarrollo*, una enfermedad puede ser *aguda*, si surge súbitamente, o *crónica*, si se mantiene a lo largo del tiempo. Una manifestación es *recidivante* si reaparece después de un periodo de aparente normalidad; si aparece súbitamente se denomina *paroxística*.

En la **relación** entre un hallazgo H, que puede ser un signo o un síntoma, y una enfermedad o una anomalía E, podemos definir varios conceptos: la *sensibilidad* $P(+h|+e)$, la *especificidad* $P(\neg h|\neg e)$, el *valor predictivo positivo* $P(+e|+h)$, el *valor predictivo negativo* $P(\neg h|\neg e)$ y otros de menor relevancia. Una especificidad igual a la unidad (100%) significa que la única causa posible del signo es esa enfermedad. Se dice que entonces que el hallazgo es *patognómico* y, en este caso, basta su presencia para diagnosticar la enfermedad. Recíprocamente, un resultado negativo en una prueba muy sensible es suficiente para descartar la enfermedad. Una descripción más detallada de estos conceptos y de la relación entre ellos puede encontrarse en el apéndice del capítulo 3.

El *diagnóstico diferencial* se da cuando existe un hallazgo o un conjunto de hallazgos que pueden deberse a varias causas. Consiste en buscar evidencia adicional que permita confirmar la presencia de una de las causas, con lo cual las demás quedan prácticamente excluidas. Otra forma de resolver el diagnóstico diferencial se basa

[2] La *anatomía* estudia la estructura de los órganos del cuerpo, mientras que la *fisiología* estudia su función. La *patofisiología* trata el comportamiento anormal de los órganos y la *etiopatología* el origen y desarrollo de las enfermedades.

en ir descartando las causas posibles hasta que quede solamente una de ellas como explicación de los hallazgos.

12.2 Fisiopatología cardíaca

Esta sección contiene un resumen muy breve de la cardiología. Nuestro objetivo no es sintetizar en pocas páginas una especialidad tan compleja, sino dar una idea cualitativa del conocimiento médico incluido en nuestra red bayesiana.

12.2.1 Anatomía y fisiología

El corazón es un órgano situado en el tórax. Consta de cuatro cavidades: dos aurículas y dos ventrículos. La aurícula izquierda recibe la sangre que las venas pulmonares aportan desde los pulmones; su contracción hace que la sangre pase por la apertura de la válvula mitral al ventrículo izquierdo. Esta válvula consta de dos valvas, anterior y posterior, sujetas por las cuerdas tendíneas a los músculos papilares, los cuales se tensan durante la sístole, impidiendo que la sangre retroceda hacia la aurícula. Las paredes del ventrículo izquierdo son mucho más musculosas que las del derecho, pues deben bombear la sangre al circuito sistémico, de mayor recorrido que el circuito pulmonar. A la salida del ventrículo izquierdo se encuentra la válvula aórtica, formada generalmente por tres valvas, o más exactamente, por tres senos de Valsalva. El cierre de la válvula aórtica impide que la sangre retroceda desde la arteria aorta al ventrículo.

En la parte derecha del corazón, la válvula tricúspide permite el paso de la sangre desde la aurícula al ventrículo pero no a la inversa; la válvula pulmonar desempeña la misma función entre el ventrículo derecho y la arteria pulmonar.

12.2.2 Valvulopatías

Puesto que las válvulas tienen dos objetivos, permitir el paso de la sangre en un sentido e impedirlo en el contrario, hay dos afectaciones principales que pueden sufrir: la *estenosis* es un estrechamiento que dificulta el paso de la sangre hacia adelante; la *insuficiencia* es un defecto en el cierre de la válvula, que permite el retroceso de la sangre (regurgitación). Ambas anomalías pueden estar presentes en una misma válvula.

La estenosis más común es la de la válvula mitral, y su origen es casi siempre reumático, es decir, por enfermedad inflamatoria del endocardio valvular, que afecta más a las mujeres que a los varones; excepcionalmente, puede haber estenosis congénita o degenerativa. También la estenosis tricúspide suele deberse a enfermedad reumática. En las válvulas aórtica y pulmonar, sin embargo, más de la mitad de los casos de estenosis son congénitos; el resto es de origen reumático o degenerativo.

La insuficiencia de la válvula mitral se debe en la mayor parte de los casos a una mala coaptación (mal ajuste en el cierre), provocada por una afectación reumática de las valvas o del aparato subvalvular, con engrosamiento y acortamiento. Otras veces se debe a la dilatación del anillo mitral, que puede ser idiopática o secundaria a la dilatación del ventrículo izquierdo; también se observa frecuentemente en el síndrome de Barlow. En los infartos agudos de miocardio puede producirse insuficiencia mitral por disfunción e incluso por rotura de un músculo papilar. La insuficiencia aórtica también suele ser de origen reumático, mientras que en las válvulas del lado derecho

la causa más frecuente es la dilatación del ventrículo derecho (para la tricúspide) y de la arteria pulmonar (para la válvula pulmonar), provocadas generalmente por la hipertensión pulmonar.

El sistema experto DIAVAL realiza el diagnóstico de las estenosis a partir de las características de la válvula observadas en el eco bidimensional: movilidad de las valvas, grosor, calcificación, dilatación del anillo, etc. y los gradientes de presión calculados mediante el eco Doppler, con el fin de diagnosticar las *estenosis*. El diagnóstico de las *insuficiencias* se basa en los datos del eco Doppler (monodimensional y mapa de color), donde se observa la regurgitación del flujo sanguíneo, más que en la información anatómica que proporciona el eco bidimensional.

También tiene en consideración nuestro sistema la endocarditis infecciosa, que se manifiesta en forma de vegetaciones observables en el eco bidimensional, y puede producir insuficiencia aguda por destrucción del tejido valvular, la cual requiere una intervención inmediata.

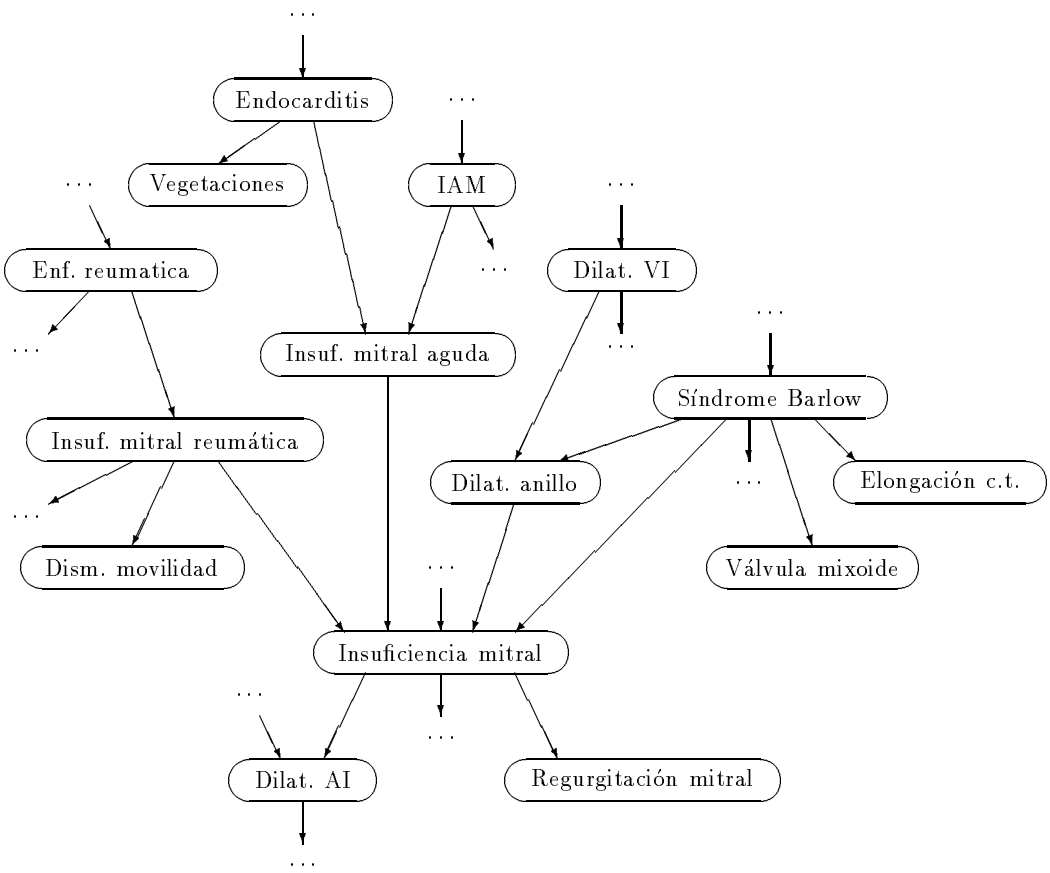


Figura 12.2: Porción de la red para la insuficiencia mitral.

La figura 12.2 muestra una pequeña porción de la red bayesiana de DIAVAL, correspondiente a la insuficiencia de la válvula mitral. Los nodos que no tienen descendientes corresponden a los datos que introduce el usuario; los cinco que aparecen en este ejemplo son todos observaciones del ecocardiograma.

12.2.3 Cavidades, pericardio, arterias y venas

Aurículas y ventrículos

DIAVAL combina los parámetros medidos y las observaciones cualitativas introducidas por el médico, con el fin de determinar si alguna de las cavidades está *dilatada* y en qué grado, y si alguna de las paredes está *engrosada*, lo cual indica la hipertrofia del músculo cardíaco en esa zona. También influye en el diagnóstico la *ecogenicidad* de las paredes cardíacas,[3] pues cuando ésta se encuentra aumentada hay que sospechar una afectación del músculo correspondiente.

Dentro de las cavidades pueden aparecer diferentes tipos de *masas*. Se conoce como *trombo* un coágulo sanguíneo que permanece en el lugar de su formación. Si se desprende y pasa al torrente sanguíneo puede provocar la obstrucción de una arteria, y recibe el nombre de *embolia*. Si ésta se localiza en el cerebro, da lugar a un *accidente cerebro-vascular* (ACV); cuando se desea determinar si éste es de origen cardíaco, resulta aconsejable realizar un ecocardiograma, a ser posible transesofágico, por su mayor sensibilidad.

Pericardio

El pericardio es una membrana que rodea el corazón, y consta de dos capas, interna y externa. Su inflamación se conoce como pericarditis. En ecocardiografía, nos interesa especialmente el diagnóstico de la *pericarditis constrictiva*, llamada así porque dificulta el llenado ventricular durante la diástole. Este fenómeno se debe al *fibrosamiento* del pericardio, que a veces se encuentra además *calcificado*.

En muchos casos, la pericarditis viene acompañada de *derrame pericárdico*, que cuando alcanza cierta gravedad, produce *taponamiento cardíaco*, el cual puede contribuir también a dificultar el llenado ventricular.

Arterias y venas

En la construcción de DIAVAL hemos considerado la *dilatación de la aorta*, valorando su gravedad (leve, moderada o severa) y su localización (en la rama ascendente, en el cayado o en la rama descendente), el *aneurisma de aorta*, la *disección de aorta*, que puede ser de tipo A (si empieza en la raíz) o B (si empieza en la rama ascendente), y la *dilatación* de la arteria pulmonar, de las venas pulmonares y de las cavas superior e inferior.

12.2.4 Isquemia

Se entiende por isquemia la falta de riego sanguíneo el cierta parte del organismo, con la consiguiente afectación de los tejidos. El electrocardiograma es una de las pruebas más útiles para detectar la enfermedad isquémica. El eco se utiliza sobre todo para evaluar las consecuencias producidas. Uno de los aspectos fundamentales que debe valorar es la *función ventricular*, medida como *fracción de eyección*, que es la proporción entre la sangre admitida y expulsado por el ventrículo izquierdo en cada latido.

[3] La ecogenicidad es la capacidad de los tejidos para reflejar los ultrasonidos. En el ecocardiograma se manifiesta como un color blanco de mayor intensidad.

Otro aspecto en que el eco aporta información muy útil es la valoración de la *contractilidad segmentaria*, para determinar qué partes del ventrículo se contraen mejor que las demás. En nuestro sistema experto dividimos la pared ventricular en ocho segmentos; para cada uno de ellos puede haber contractilidad normal, hipoquinesia, aquinesia y disquinesia.[4]

Dentro de la enfermedad isquémica, una de las complicaciones posibles del infarto agudo de miocardio (IAM) es el *aneurisma* (dilatación anormal) del ventrículo izquierdo. También puede producirse la *comunicación interventricular*, por rotura del tabique que separa ambas cavidades. Otra de las consecuencias ocasionales es la rotura de las cuerdas tendíneas o de los músculos papilares que sujetan la válvula mitral, con lo que se produce la insuficiencia mitral —en este caso aguda— mencionada anteriormente.

12.2.5 Enfermedades del miocardio

Éste es sin duda el capítulo peor conocido de toda la cardiología. Según Sokolow y McIlroy [212, pág. 527], "todas las clasificaciones son arbitrarias y ninguna resulta totalmente satisfactoria, debido a que existen muchas causas desconocidas y al solapamiento entre las miocardiopatías". Incluso el término *cardiomiopatía* resulta controvertido, pues según su etimología debería incluir cualquier enfermedad del miocardio; sin embargo, existe el acuerdo de reservar este nombre para las afecciones idiopáticas y para algunos casos particulares de miocardiopatías secundarias.

Miocardiopatía dilatada

A veces también se la conoce como *miocardiopatía congestiva*. Consiste en una dilatación de los ventrículos, especialmente del izquierdo. Aunque generalmente es idiopática, se incluye como excepción dentro de las miocardiopatías la que está provocada por el alcoholismo. La disminución de la contractilidad ventricular es uno de los rasgos que permiten el diagnóstico diferencial de esta enfermedad frente a la dilatación producida por patología valvular o por otras causas.

Miocardiopatía hipertrófica

Tiene una prevalencia mucho menor que la miocardiopatía dilatada, y un fuerte componente hereditario. Se caracteriza por un engrosamiento de las paredes del ventrículo izquierdo, muy especialmente del septo interventricular, hasta el punto de que un cociente entre el grosor del septo y el de la pared posterior superior a 1'3 (ver la fig. 13.2) orienta el diagnóstico diferencial de esta miocardiopatía frente a la hipertrofia debida a otras causas. El movimiento sistólico anterior (SAM) de la válvula mitral y la escasa movilidad del septo son otros de los hallazgos ecocardiográficos típicos de esta enfermedad.

En algunos casos, el engrosamiento del septo puede llegar a obstruir el tracto de salida del ventrículo izquierdo; es lo que se conoce como *miocardiopatía hipertrófica obstructiva*, cuyos efectos son similares a los de la estenosis aórtica ya citada.

[4] La disquinesia se produce cuando un segmento de músculo cardíaco está tan dañado que, al contraerse los demás, él se mueve en dirección contraria a causa de la presión intraventricular.

Miocardiopatía restrictiva

Es una forma muy infrecuente de afectación cardíaca. Se caracteriza fundamentalmente por la incapacidad de las paredes ventriculares para distenderse durante la diástole. En muchos casos está provocada por las denominadas *enfermedades de depósito*, las cuales pueden hacer que la materia acumulada en las paredes o en endocardio ocupe una parte considerable del ventrículo; en este caso recibe el nombre de *miocardiopatía restrictiva obliterativa*.

Es importante el diagnóstico diferencial entre la miocardiopatía restrictiva y la pericarditis constrictiva mencionada anteriormente. En la ecocardiografía, ambas se manifiestan como una dificultad en el llenado ventricular, y no es fácil distinguir una de otra, por lo que a veces se hace necesario acudir a la resonancia magnética nuclear para poder establecer el diagnóstico definitivo.

12.2.6 Malformaciones congénitas

El sistema experto DIAVAL está orientado especialmente a la ecocardiografía del adulto, que se diferencia de la cardiología infantil sobre todo en el capítulo de las enfermedades congénitas. La razón es que hay anomalías relativamente frecuentes entre los recién nacidos, pero cuya gravedad dificulta que el paciente pueda sobrevivir hasta la edad adulta.

La variedad de malformaciones congénitas es prácticamente ilimitada. En nuestro sistema experto hemos considerado las siguientes: válvula aórtica bicúspide (lo normal es que sea tricúspide), comunicación interventricular (CIV), comunicación interauricular (CIA), ductus arterioso persistente (consiste en que la comunicación entre el tronco pulmonar y la aorta no se ha cerrado después del nacimiento), estenosis pulmonar infundibular, estenosis aórtica, coartación (estrechez) de la aorta, atresia tricúspide (significa que esta válvula no ha llegado a desarrollarse), enfermedad de Fallot (estenosis pulmonar, CIV y posición anómala de la aorta) y enfermedad de Ebstein (mala implantación de la válvula pulmonar).

12.3 Técnicas de diagnóstico

La labor de diagnóstico consiste en recoger toda la información posible sobre el paciente con el fin de llegar a un diagnóstico. Sin embargo, ya hemos comentado anteriormente que el proceso no es tan sencillo. En realidad, el médico va obteniendo conclusiones intermedias y realizando diagnósticos provisionales que orientan la recogida de información. Habitualmente, el médico empieza por solicitar los datos que se pueden obtener con mayor facilidad, menor riesgo, mayor rapidez y menor coste. Entre ellos, naturalmente, se encuentran los que puede obtener en su propia consulta: son los datos clínicos, de los que vamos a hablar a continuación. Después, encargará las pruebas complementarias oportunas. Se conoce como *historia clínica* el informe en que se recoge toda la información disponible sobre un paciente. Veamos a continuación los elementos que incluye.

12.3.1 Datos clínicos

Son los más fáciles, rápidos y baratos de obtener. Se recogen siempre en la consulta del médico antes de solicitar las pruebas complementarias.

Filiación

Desde el punto de vista administrativo es necesario conocer el nombre del paciente y conviene registrar también el número de la Seguridad Social o el DNI. De cara al diagnóstico, los datos más importantes son la edad del paciente (o mejor, la fecha de nacimiento), el sexo y, para ciertas enfermedades, la zona geográfica de origen. A veces se anotan también su domicilio, teléfono, estado civil, lugar de nacimiento, etc. La primera que aparece en el sistema DIAVAL durante la introducción de un nuevo eco recoge además los datos relativos al ecocardiograma: número, cinta y hora de grabación en vídeo, tipo de eco, etc. (fig. 13.1).

Motivo principal de la consulta

En general, el enfermo acude a la consulta aquejado de algún síntoma. Otras veces se trata de una revisión, que puede deberse a un examen rutinario o a alguna enfermedad o intervención que sufrió anteriormente. En los servicios especializados, tales como la cardiología, los pacientes generalmente han sido examinados previamente por el médico de atención primaria o por otro especialista, quien ha solicitado la consulta, porque se trata de un enfermo que va a ser dado de alta, que tiene un diagnóstico dudoso o que va a ser intervenido quirúrgicamente. Cuando el enfermo ya ha sido examinado, el motivo de la consulta suele estar bien definido y es de gran ayuda para guiar al médico que realiza el ecocardiograma.

Por tanto, uno de los elementos fundamentales de la historia clínica es la *enfermedad actual*: cuál es, cuándo empezó, dónde se localiza, qué intensidad tiene, a qué se debe, cómo ha evolucionado, qué manisfestaciones la acompañan, etc.

El sistema DIAVAL utiliza una serie de pantallas que aparecen opcionalmente con el fin de recoger esta información. Para algunas enfermedades y signos el programa tiene en cuenta el grado de intensidad, pero en su versión actual no es capaz de considerar otros matices, como el momento en que apareció y la forma en que ha evolucionado.

Anamnesis

Se conoce como *anamnesis* (etimológicamente significa "recuerdo") la parte de la historia clínica que recoge todos los datos del paciente anteriores a la enfermedad. Aunque en ella también se incluye a veces la filiación, vamos a referirnos ahora a los antecedentes y a los factores de riesgo y hereditarios.

Los *antecedentes* del enfermo, tales como las enfermedades que ha sufrido el paciente, la medicación que ha recibido y las intervenciones que se le han practicado, son relevantes de cara al diagnóstico, pues es posible, por ejemplo, que la enfermedad reumática haya producido una afectación cardíaca valvular o que la tuberculosis haya ocasionado una pericarditis constrictiva. En la sección 12.4, dedicada a las técnicas terapéuticas, hablaremos de las posibles complicaciones derivadas de los tratamientos e intervenciones, tal como los considera DIAVAL.

Hay que tener en cuenta además los *factores hereditarios*. La miocardiopatía hipertrófica, por ejemplo, es una enfermedad con una alta probabilidad de transmisión. Por tanto, los hijos de quienes padecen la enfermedad deben someterse a exámenes ecocardiográficos periódicos con el fin de determinar si existe la enfermedad y cómo evoluciona.

Entre los *factores de riesgo* en cardiología podemos citar la hipertensión arterial, las hiperlipidemias (entre ellas, el exceso de colesterol) y el tabaquismo, por la posibilidad de que se produzcan enfermedades coronarias. El alcoholismo es un factor de riesgo para la miocardiopatía dilatada, y la drogadicción lo es para la endocarditis.

En nuestro sistema experto, estos antecedentes y factores vienen representados por nodos que se sitúan como causas en la parte más alta de la red, desde donde propagan la evidencia recibida aumentando o disminuyendo la probabilidad de que aparezcan ciertas anomalías.

Exploración física

Una vez concluido el interrogatorio incial del paciente, el médico comienza la exploración física. Deberá examinar el aparato circulatorio (corazón, arterias y venas), el aparato respiratorio, el abdomen, el aparato locomotor, el sistema neurológico, etc.

En cardiología, naturalmente, tiene especial importancia el estudio del aparato circulatorio y de las alteraciones pulmonares por la gran relación que éstas guardan con el funcionamiento del corazón. Hay síntomas y signos que se asocian con un mal funcionamiento del lado derecho del corazón; en este caso la sangre se acumula en el circuito venoso sistémico y se producen edemas (hinchazón por acumulación de líquido) en los miembros inferiores y en el abdomen; también puede manifestarse como un aumento de la presión venosa yugular, hepatomegalia, edemas en miembros inferiores, ascitis, etc. Cuando falla el lado izquierdo, la sangre acumulada en el sistema venoso pulmonar puede provocar hemoptisis, disnea (sensación de ahogo al respirar), ortopnea (disnea que obliga al paciente a estar de pie o sentado) y otras manifestaciones similares.[5]

Los cuatro apartados clásicos de la exploración física son *inspección*, *auscultación*, *palpación* y *percusión*. Basándose en la *auscultación*, que consiste en escuchar con ayuda del fonendoscopio el funcionamiento del corazón o de los pulmones, un cardiólogo experimentado es capaz detectar o descartar con gran fiabilidad la presencia de valvulopatía y de otras enfermedades, mediante el reconocimiento de ruidos normales, desdoblamientos, soplos, chasquidos, roces y otros sonidos. La *palpación* y la *percusión* son dos métodos sencillos que, a pesar de las innovaciones tecnológicas introducidas en medicina, aún pueden aportar información orientativa valiosa de cara al diagnóstico.

12.3.2 Pruebas complementarias

A veces se conocen también como *pruebas de laboratorio*. Pueden clasificarse en dos grupos: *invasivas* y *no invasivas*. Las primeras utilizan generalmente un catéter o una sonda, que se introduce en el organismo del paciente con el fin de obtener información no accesible desde el exterior. Tal es el caso de la *hemodinámica*, que

[5] Hay veces en que no existe una frontera bien definida entre síntoma (lo que el enfermo experimenta) y signo (lo que observa el médico).

utiliza el catéter para medir presiones y flujos dentro del corazón. Entre las pruebas no invasivas se encuentran la electrocardiografía, la ecocardiografía y la resonancia magnética. Describiremos a continuación algunas de ellas y trataremos a parte con más detalle la ecocardiografía, por constituir el objeto principal de esta tesis.

Electrocardiografía

El electrocardiograma (ECG) consiste en el registro de la actividad eléctrica del corazón. En nuestro sistema experto hemos supuesto que el cardiólogo ha realizado ya la interpretación del ECG, de modo que en la pantalla correspondiente (fig. 13.3) no figuran observaciones primarias, tales como "onda P bífida", sino resultados, como "fibrilación auricular", "crecimiento de la aurícula izquierda", "signos de necrosis", etc.

Existen otras pruebas electrocardiográficas especiales, como el vectograma, la prueba de esfuerzo o ergometría y el registro Holter, que no hemos considerado en la primera versión de DIAVAL.

Radiografía

Esta técnica se basa en el distinto grado de absorción de los rayos Röntgen (rayos X) por los tejidos del cuerpo humano, con lo que es posible obtener información del interior del organismo mediante una técnica de bajo riesgo y coste reducido.

En cardiología, resulta útil para detectar la dilatación de las cavidades, de las arterias aorta y pulmonar, el derrame pericárdico y la calcificación del pericardio y de las válvulas. Además, el estudio de la vascularización pulmonar permite observar anomalías como la hipertensión precapilar o postcapilar, plétora pulmonar, edema intersticial o alveolar, etc., que en muchos casos son consecuencia de afectación del lado izquierdo del corazón o son causa de alteraciones en el lado derecho.

Otras pruebas

Además de los rayos X y la ecocardiografía, existen otras técnicas que permiten obtener imágenes del corazón. La *angiografía* consiste en introducir en el corazón mediante un catéter una solución de contraste (radiopaca), con el fin de detectar la obstrucción de las arterias coronarias. Mediante este método es posible determinar también el tamaño de las cavidades. Por tanto, el cateterismo, además de medir directamente los gradientes de presión, es capaz de proporcionar imágenes cardíacas, aunque su coste económico es elevado y, al ser una técnica invasiva, conlleva cierto riesgo para el paciente.

La *cardiología nuclear*, una rama de la *gammagrafía*, se basa en la inyección intravenosa de radioisótopos que, al ser absorbidos de forma selectiva en distintos lugares de corazón, permiten estudiar el movimiento de las paredes cardíacas, las zonas mal perfundidas y las zonas inflamadas.

La *resonancia magnética nuclear* es una técnica muy reciente que permite observar con extraordinaria precisión los tejidos del cuerpo humano. Sin embargo, su implantación es aún muy escasa debido a su elevado coste económico y a que todavía hay muchos puntos que están por estudiar.

En la versión actualmente disponible del sistema experto DIAVAL no hemos incluido ninguna de estas tres técnicas por dos razones. La primera, por su complejidad, comparable o superior a la de la ecocardiografía, lo cual requeriría un estudio en profundidad que desborda los límites de nuestro proyecto. Y la segunda, porque a la hora de realizar un ecocardiograma no suele estar disponible la información procedente de las pruebas más costosas y arriesgadas, que se realizan sólo cuando el eco no ha dado un diagnóstico definitivo.

12.3.3 Ecocardiograma

La ecocardiografía es una de las técnicas más importantes para el estudio del corazón, y constituye el objeto principal del proyecto descrito en esta memoria. Su nombre se debe a que está basada en el registro de los ultrasonidos reflejados por las estructuras que forman el corazón.

Más detalladamente, su fundamento consiste hacer incidir sobre el corazón un haz de ultrasonidos (es decir, vibraciones de frecuencia igual o superior a 20.000 Hz.), emitido por un transductor, que consta básicamente de un cristal piezoeléctrico. Cuando el haz de ultrasonidos encuentra una frontera entre dos estructuras de diferente densidad, da lugar a dos ondas, una que se transmite hacia adelante y otra que se refleja y es recogida por el transductor, el cual la convierte en una señal eléctrica que será analizada por el ecocardiógrafo [33]. Originalmente, el registro de la imagen se realizaba en papel; hoy en día se utiliza un monitor de vídeo, con la posibilidad de grabar las imágenes para un estudio posterior. Dependiendo del tipo de ultrasonido emitido y del tratamiento de la señal reflejada, podemos distinguir varios tipos de registro; los que se emplean en la actualidad son el modo M, el bidimensional, el Doppler modo M (pulsado o continuo; en blanco y negro o codificado en color) y el mapa Doppler color. También es posible distinguir entre eco transtorácico y eco transesofágico, como explicaremos más adelante.

Las principales ventajas de la ecocardiografía son que se trata de una técnica inocua para el paciente y de bajo costo (aunque mayor que el de la electrocardiografía), y que proporciona gran cantidad de información. A ello han contribuido sobre todo las nuevas técnicas de registro y la mejora en la resolución de los aparatos. Vamos a describir a continuación brevemente los diversos modos de ecocardiografía que se utilizan en la actualidad.[6]

Eco bidimensional

Es la técnica que más se utiliza hoy en día, y sirve de guía para los demás tipos de estudio (modo M, Doppler, etc.). Su principal cualidad es que muestra una imagen del corazón como si se hubiera dado un corte en un determinado plano, y que se trata de una imagen dinámica, en que es posible estudiar el movimiento de las paredes y de las válvulas cardíacas.

[6] Aunque todos los libros de medicina hablan primero del eco en modo M y después del bidimensional, según el orden histórico en que aparecieron, nos parece más didáctico invertir este orden, por varias razones: en primer lugar, el eco bidimensional es más sencillo de comprender, porque utiliza dos dimensiones homogéneas (espaciales), en vez de una espacial y otra temporal; por tanto, esta más próximo a la anatomía, y las imágenes obtenidas resultan mucho más familiares. Otra razón es que, en la práctica diaria, el ecocardiografista utiliza primero el bidimensional para obtener una imagen del corazón, y sólo entonces aplica el modo M, guiado por el bidimensional.

Una vez situado el transductor en un determinado punto, es posible rotarlo sobre su eje y variar el ángulos de incidencia con el fin de obtener cortes longitudinales, transversales, oblicuos, etc. En la práctica, los cortes habituales son: parasternal de eje largo y de eje corto, apical de cuatro o cinco cámaras (las dos aurículas, los dos ventrículos y la aorta), subcostal y suprasternal.

La mayor parte de las pantallas que muestra el sistema DIAVAL se refieren al eco bidimensional: una de ellas para introducir los parámetros medidos con esta técnica y otras para la válvula mitral, la válvula aórtica, las válvulas tricúspide y pulmonar, las cavidades y paredes, las arterias y venas, el pericardio, etc.

Modo M

El eco modo M sirve para observar el movimiento de las paredes y las válvulas: el eje de abcisas representa el tiempo y el de ordenadas la profundidad. Con el surgimiento del eco bidimensional, el modo M ha pasado a un segundo plano de importancia, pero sigue siendo útil como complemento del anterior por su mayor resolución y porque permite registrar en una sola imagen el movimiento de una determinada estructura (una válvula, por ejemplo), por lo cual resulta más apropiado para realizar medidas, a veces sincronizadas con el ECG. Por tanto, los principales datos procedentes del eco modo M que consideramos en nuestro sistema experto son los parámetros correspondientes a los diámetros de las cavidades y el grosor de las paredes, tal como muestra la figura 13.2.

Eco Doppler

El efecto Doppler aparece en numerosos fenómenos físicos, y se debe a la variación entre la frecuencia emitida (ν_e) y la recibida (ν_r) cuando existe una velocidad relativa entre emisor y receptor. En la ultrasonografía, existe un doble efecto Doppler cuando la estructura que refleja el sonido se encuentra en movimiento. La diferencia entre ambas frecuencias viene dada por la siguiente expresión:

$$\Delta\nu = \nu_r - \nu_e = \frac{2\,\nu_e\,v\,\cos\theta}{c}$$

donde v es la velocidad relativa entre emisor y receptor, c es la velocidad de las ondas ultrasónicas en el medio transmisor y θ es el ángulo entre el haz de ultrasonidos y la velocidad relativa entre el emisor y el punto de reflexión. La frecuencia aumenta ($\Delta\nu > 0$) cuando el emisor se acerca al receptor, y disminuye cuando se aleja.

Existen dos variedades principales en la ecocardiografía Doppler. La primera de ellas, conocida como *Doppler pulsado*, consiste en enviar pulsos de sonido y recogerlos con un mismo transductor. Resulta muy útil para medir con precisión la velocidad en un punto concreto. Sin embargo, presenta el problema de la *suplantación* (en inglés, *aliasing*), por lo que no permite estudiar flujos de alta velocidad. El *Doppler continuo*, en cambio, utiliza dos transductores, de modo que puede emitir y recibir sin interrupción; tiene la ventaja de que permite detectar movimientos de alta velocidad y el inconveniente de que no permite determinar su profundidad. Por tanto, ambas modalidades son complementarias en el estudio ecocardiográfico.

La principal utilidad de esta técnica reside en el estudio de los flujos sanguíneos

y, mediante la aplicación de la ecuación simplificada de Bernouilli,

$$\Delta P = 4\,v^2$$

del cálculo de gradientes de presión. La insuficiencia de una válvula se detecta como un flujo de velocidad en el momento en que la válvula debería estar cerrada; la estenosis se detecta como un aumento del gradiente de presión a través de la válvula durante su apertura.

La presentación del registro del eco Doppler puede realizarse en blanco y negro (el eje de abcisas corresponde al tiempo y el de ordenadas a la velocidad) o codificada en color (superpuesta al eco normal): los flujos que se acercan al transductor aparecen en rojo y los que se alejan en azul; los colores amarillos y verdosos indican turbulencias. El eco Doppler codificado en color superpuesto al eco bidimensional se conoce como *mapas de color* y permite localizar directamente los flujos y las turbulencias.

En DIAVAL existen dos pantallas dedicadas al eco Doppler en modo M, una para los parámetros y otra para las observaciones cualitativas, y otras dos pantallas similares para el mapa de color.

Eco transesofágico

Hemos distinguido varios modos de ecocardiograma en función de la señal emitida y de la presentación del registro. Otro criterio posible de clasificación es la localización del transductor. En el *eco transtorácico*, el transductor se sitúa externamente en la región precordial, mediante la aplicación de un gel para facilitar la transmisión del sonido. Sin embargo, el aire de los pulmones y el tejido adiposo de las personas obesas dificulta en muchas ocasiones la observación del corazón. Por eso se desarrolló posteriormente el *eco transesofágico*, que consiste en introducir una sonda por el tubo digestivo hasta situarse a la altura del corazón, con lo que se consigue una resolución muy superior. Este segundo método presenta la desventaja de que es molesto para el paciente, por lo que la mayor parte de las veces basta realizar un eco transtorácico, reservando la otra modalidad para los casos dudosos o para aquéllos en que es importante realizar un diagnóstico muy preciso.

La distinción entre ambos tipos de eco es importante para nuestro sistema experto porque implica que, al ajustar las probabilidades condicionales de la red bayesiana, debemos tener en cuenta que el transesofágico es más específico y mucho más sensible que el otro, lo cual puede influir en el resultado del diagnóstico. Por ejemplo, en la detección de vegetaciones, un resultado negativo mediante un eco transesofágico permite excluir la endocartidis con mucha mayor certeza que si fuera transtorácico.

12.4 Técnicas terapéuticas

Dentro del tratamiento clínico, independientemente de la especialidad considerada, pueden distinguirse dos tipos principales según el método utilizado: el que se basa en medicamentos y el que utiliza la cirugía. En cuanto a su objetivo, un tratamiento puede ser *corrector* si trata de subsanar completamente la anomalía o *paliativo* si se limita a atenuar sus consecuencias.

En la primera versión de nuestro sistema experto nos hemos limitado al diagnóstico, por lo que el tema del tratamiento terapéutico tiene un interés menor en esta

tesis. En la red bayesiana, sólo consideramos el tratamiento *a posteriori*, es decir, teniendo en cuenta las intervenciones que ha sufrido el paciente por su relación con la enfermedad actual. Por ejemplo, si el paciente sufrió hace diez años una valvulotomía mitral percutánea, es posible que la intervención le produjera una insuficiencia mitral y es muy probable que la válvula se haya reestenosado con el paso del tiempo. Más importante aún es conocer si al paciente se le ha implantado una prótesis valvular, y si ésta es de tipo biológico o mecánico, pues las complicaciones que presenta cada una de ellas son muy diferentes.

El sistema experto DIAVAL distingue tres tipos de intervenciones valvulares: comisurotomía, prótesis (considerando los diferentes modelos y tamaños que existen en la actualidad) y anuloplastia. Dentro de las intervenciones coronarias, hay que distinguir entre angioplastia, *bypass* y *stent*. Además, tanto para las intervenciones coronarias como para las valvulares conviene tener en cuenta el mes y el año en que se realizó la intervención, porque el periodo de tiempo transcurrido influye notablemente en la probabilidad de que hayan surgido complicaciones.

En cuanto a los tratamientos no quirúrgicos, DIAVAL tiene en cuenta si al paciente se le han administrado fármacos cardiotóxicos, que afectan fundamentalmente a la función ventricular sistólica, y la radioterapia, que puede producir pericarditis constrictiva y derrame pericárdico.

Capítulo 13

Implementación de DIAVAL

En la presente memoria, la implementación del sistema experto aparece descrita en dos capítulos diferentes. El primero de ellos se situó en la parte II, porque trataba los temas generales del razonamiento bayesiano, independientemente del dominio de aplicación. En este capítulo, en cambio, discutiremos las características particulares de DIAVAL, el sistema experto para ecocardiografía que hemos desarrollado.

13.1 Red bayesiana cardiológica

13.1.1 Contenido de la red

En la sección 9.2.1 hablábamos de la posibilidad que ofrece el *GoldWorks* de representar el conocimiento mediante marcos e instancias, y dentro de los ejemplos anticipábamos parte del contenido de esta sección al hablar de la red de clasificación que aparece en DIAVAL. Antes de continuar, conviene aclarar que la red de clasificación es, en principio, independiente de la red causal. Ambas coinciden en su carácter jerárquico, pero se diferencian por su significado: en la primera, los enlaces indican *inclusión* (una subclase dentro de una clase más amplia) o *pertenencia* (un elemento de una clase), mientras que en la red bayesiana todos los enlaces son *causales*.

Recordemos que en la red de clasificación de *GoldWorks* existe un marco, denominado TOP-FRAME, del que dependen todos los demás, tanto los que ofrece la propia herramienta para la creación del interfaz gráfico como los que nosotros hemos utilizado para definir la red, clasificados en dos submarcos: NODO y ENLACE. El primero de ellos posee a su vez dos submarcos: ALTERACION y DATO. Las alteraciones corresponden a los nodos que pueden tener hijos, mientras que los datos son los nodos terminales de la red. La distinción es útil tanto para la inferencia como para la generación de explicaciones. Hasta aquí, todo es completamente general para redes bayesianas, independientemente del dominio de aplicación.

Para construir un sistema experto en ecocardiografía, hemos visto la necesidad de incluir un nuevo tipo de nodo: DATO-PERSONAL, cuyas instancias son Sexo, Edad y País-de-origen. Conviene además distinguir qué clases de datos vamos a tener. Por eso hemos definido varios submarcos: DATO-ECO, RAYOS-X, ECG, y dentro del primero, MODO-M, BIDIMENSIONAL, DOPPLER y MAPA-COLOR. Un tipo particular de dato es el PARÁMETRO, que se caracteriza por llevar una medida numérica asociada. A su vez, los parámetros pueden proceder del modo M, del eco bidimensional, etc.

La red de DIAVAL consta de 324 nodos. Además de los **tres datos personales** ya mencionados, existe otro grupo que contiene **86 alteraciones**. Algunas de ellas describen *afectaciones anatómicas*, tales como la obstrucción de las entradas y salidas de los ventrículos o la dilatación de las cavidades. Otras representan *estados patofisiológicos*: hipertensión arterial, hipertensión pulmonar, alteración de la relajación, etc. Por último, hay un nodo-alteración para cada *diagnóstico* posible: distintos tipos de valvulopatías, endocarditis, miocardiopatías, enfermedades congénitas, etc. Podríamos haber definido varios submarcos para cada uno de estos tres tipos de alteración, pero la frontera entre ellos es difusa y, además, vimos en la práctica que no era necesario.

Dentro de los nodos de la red, el grupo más numeroso está formado por **325 datos**. Entre ellos, hay **52 parámetros**: 12 del modo M, 13 del bidimensional, 21 del Doppler, 6 del mapa de color.

Estos nodos están unidos entre sí por **335 enlaces**: 126 de ellos son causales, 36 son de influencia, 15 pertenecen a la clase "tipo", 71 correponden a observaciones, 58 a manifestaciones y 37 a parámetros. (Las seis clases de enlaces están descritas en la sec. 8.3.1.) Hay parámetros que sólo influyen indirectamente en la inferencia, al intervenir en el cálculo de otros.

Conviene mencionar también que nuestro sistema experto incluye otra forma de clasificación, de la que no habíamos hablado hasta ahora. Su objetivo es agrupar los nodos en *capítulos*, con el fin de que la presentación de los diagnósticos y de los datos del informe se realice de forma ordenada. En la actualidad, DIAVAL cuenta con **21 capítulos**, 8 para los antecedentes (intervenciones valvulares, intervenciones no valvulares, motivo de petición, enfermedades padecidas, síntomas, factores de riesgo, factores hereditarios y E.C.G.) y 13 para los hallazgos del eco: válvula mitral, v. aórtica, v. tricúspide, v. pulmonar, cavidades y paredes, función ventricular sistólica, llenado ventricular, contractilidad segmentaria, pericardio, arterias y venas, miocardiopatías, alteraciones poco frecuentes y comentarios.

Cada uno de ellos está implementado como instancia del marco CAPÍTULO, que cuenta con cuatro campos: título, todo-normal, datos y diagnósticos; en los dos últimos es donde se agrupan los nodos correspondientes.

De este modo, a pesar de que la red bayesiana sólo distingue entre alteraciones y datos, las dos clasificaciones que acabamos de mencionar permiten estructurar el modelo cardiológico del sistema experto, con el fin de que la interacción con el usuario se asemeje lo más posible al modo en que los médicos organizan su conocimiento y expresan sus conclusiones.

13.1.2 Construcción de la red

Una vez que hemos definido todos los elementos que componen la red, es necesario introducir la información en el ordenador. Para ello conviene crear una serie de funciones y procedimientos que nos faciliten la labor.[1]

[1] En la herramienta comercial HUGIN mencionada en la sec. 10.1.4, existe un interfaz gráfico que simplifica este trabajo. Nosotros nos hemos limitado a crear algunos pequeños programas para uso personal, en que la utilización de ventanas es bastante reducida. La única ventaja de nuestro método frente a HUGIN es que admite varios tipos de nodos, de enlaces y de interacciones, mientras que para dicha herramienta todos los nodos son iguales e interaccionan mediante el modelo general (no considera la puerta OR), y los enlaces no necesitan etiquetas semánticas.

En cuanto a la *introducción de la información*, los programas de apoyo que hemos creado se encargan de realizar las actualizaciones oportunas. Por ejemplo, hemos definido unos demonios con el fin de que, si afirmamos que X es padre de Y, automáticamente se establece que Y es hijo de X —y viceversa— y se crea el enlace correspondiente. Una actualización similar se realiza cuando retractamos la afirmación anterior o cuando eliminamos un enlace. Existe otro pequeño programa que se encarga de determinar el tipo de interacción de cada nodo y la clase de cada enlace, deduciéndolo automáticamente o preguntando al diseñador, que es quien determina en la mayor parte de los casos el tipo de interacción y de enlace.

Otra misión de estos programas de ayuda es garantizar la *consistencia* de la red. En este sentido, el punto más importante es comprobar que la red no contenga ciclos. También hay que comprobar que la clase asignada a cada enlace no está en contradicción ni con el tipo de interacción del nodo hijo ni con el marco a que pertenece el nodo. En el caso más favorable, una inconsistencia en la red puede llevar a resultados absurdos o al menos extraños a la hora de generar explicaciones. En el caso peor, las inconsistencias pueden llevar a que el algoritmo de inferencia entre en un bucle sin fin o produzca un mensaje de error.

13.2 Propiedades del interfaz de usuario

Uno de los componentes fundamentales de nuestro sistema, que se ha desarrollado con especial esmero, es el *interfaz de usuario*. El objetivo es que el personal médico y paramédico pueda interactuar con el programa de forma sencilla y eficiente.

13.2.1 Características deseables

Existe una extensa bibliografía sobre el diseño de interfaces. El libro más interesante que hemos encontrado es el de Brown y Cunningham [22], que aporta muchos consejos claros y concretos. Wexelblat [254] estudia específicamente los interfaces para sistemas expertos y ofrece numerosas referencias. También es interesante el breve artículo de Miller y Masarie [147], en que critican los primeros sistemas expertos porque dirigían la consulta mediante una serie de preguntas, sin que el usuario pudiera tomar la iniciativa sobre la información que deseaba introducir; el sistema MAGIT, desarrollado en la Universidad de Santiago de Compostela [139, 235], recoge la sugerencia de Miller y Masarie de permitir al usuario la entrada libre de información mediante un interfaz gráfico.

Los primeros estudios sobre el diseño de interfaces, clasificaban a los *usuarios* en una escala unidimensional, de principiantes a expertos. Posteriormente, Shneiderman [208] introdujo la distinción entre usuarios principiantes, ocasionales y habituales. Esta clasificación puede resultarnos útil a la hora de diseñar un interfaz: el programa debe ser sencillo para el principiante, ofreciéndole todas las indicaciones necesarias, pero sin entorpecer la interacción del usuario veterano que apenas requiere ayuda, y a la vez hay que tener en cuenta que el usuario ocasional conoce ya el funcionamiento del programa, aunque fácilmente olvida detalles puntuales.

Estas consideraciones son válidas para cualquier aplicación informática. En el caso particular de un sistema experto, hay que tener presente además el distinto grado de conocimiento que cada usuario posee sobre ese campo. En nuestro caso, tendríamos

en un extremo a los médicos residentes, que han adquirido una sólida formación sobre medicina, en particular sobre cardiología, pero aún no están familiarizados con la ecocardiografía al iniciar su rotación en este servicio. En el otro extremo, nos encontramos con especialistas veteranos que tienen en su haber largos años de experiencia, con innumerables ecos realizados. Por tanto, hay que distinguir el tipo de información que cada uno de ellos espera recibir del sistema experto.

Finalmente, a la hora de diseñar un interfaz, conviene imaginarse siempre el peor de los casos, es decir, pensar en un usuario que no posee conocimientos de informática, que incluso tiene fobia a los ordenadores, que puede encontrar dificultades en lo que a nosotros nos parece más trivial, y que es capaz de hacer barbaridades en cuanto le dejemos la más mínima oportunidad. Esta mentalidad, que no es pesimista sino precavida, nos ayudará a diseñar un interfaz tan sencillo y tan robusto como esté a nuestro alcance.

Teniendo en cuenta las referencias citadas y las reflexiones anteriores, mostramos a continuación las características más importantes que, a nuestro juicio, debe poseer el interfaz de un sistema experto:

Facilidad: Debe estar diseñado de modo que el personal médico y paramédico pueda manejar el programa incluso en el caso de que sus conocimientos informáticos sean nulos. El objetivo es que basten unas breves explicaciones práctica para que el usuario aprenda a utilizar el sistema experto.

Familiaridad: La presentación debe ser semejante a otras aplicaciones, para que los usuarios que ya tienen experiencia con otros programas puedan aprovechar sus conocimientos. El programa debe comportarse como el usuario espera que se comporte.

Consistencia: Si la *familiaridad* se refiere a la semejanza con otros programas, la *consistencia* significa que debe comportarse de forma homogénea a lo largo de todas las partes que lo componen: uniformidad en la entrada de datos, en la presentación de la información, en la forma de acceder a la ayuda, etc.

Comodidad: Conviene que el control del programa pueda realizarse mediante el ratón o mediante la pulsación de muy pocas teclas, con el fin de que la interacción resulte rápida y eficiente.

Flexibilidad: Como ayuda al usario, conviene que el programa solicite la información en un orden establecido, pero ofreciendo a la vez la posibilidad de alterar dicho orden y moverse libremente por las distintas pantallas.

Ayuda: Éste es uno de los puntos clave en el diseño del interfaz. Es preciso que el usuario tenga en todo momento un acceso inmediato a la ayuda, y que ésta sea *sensible al contexto*, es decir, adaptada a la tarea que se está realizando. La ayuda debe ser siempre *opcional*, de modo que resulte útil para el principiante pero sin entorpecer al usuario experto.

Claridad: Otro de los puntos fundamentales es la *presentación de la información*. El usuario debe tener siempre presente o fácilmente accesible toda la información que necesita y, por otro lado, hay que evitar abrumarle con más información de la necesaria. El equilibrio entre los dos extremos es primordial.

Realimentación: Para mantener informado al usuario y evitar errores, conviene que el resultado de toda operación que éste realice aparezca reflejado en la pantalla del ordenador. Así se podrán comprobar los datos introducidos y realizar las correcciones necesarias.

Comprobación: En un sistema experto, hay que evitar que la introducción de datos incorrectos pueda llevar a conclusiones absurdas. Para ello, es necesario que el sistema posea conocimiento —en forma explícita o compilada— sobre cuáles son los valores posibles y cuáles son improbables aunque admisibles. También es importante que el sistema solicite la confirmación del usuario antes de realizar una operación que pueda llevar a la pérdida accidental de información.

Robustez: Ya hemos mencionado anteriormente que, a la hora de diseñar el interfaz, conviene ponerse siempre en el peor caso, para evitar que el programa pueda quedarse "colgado". Cuando la base de datos no puede abrirse, cuando no está conectada la impresora, cuando no queda espacio en el disco duro, el sistema deberá ofrecer un mensaje al usuario en que explique claramente cuál es el error y cómo subsanarlo: no basta decir "`Syntax error in line 1425`" o "`Fatal error: #287`" y detener la ejecución. Igualmente, el sistema deberá ser lo más fiable posible frente a apagones eléctricos y fallos del ordenador, de modo que nunca llegue a perderse una cantidad significativa de información.

Seguridad: Todo sistema experto, muy especialmente en el caso de la medicina, deberá estar dotado contraseñas y otras medidas de seguridad que protejan la confidencialidad de los datos frente a intrusos. Las implicaciones legales de la informática médica constituyen un campo que, aunque aún se encuentra en sus comienzos, se está desarrollando con extraordinaria rapidez ante las exigencias de la sociedad.

Compatibilidad: Es conveniente también que el sistema experto sea compatible con otros programas. Este punto afecta especialmente al diseño de la base de datos, que debe poseer un formato estándar con el fin de que otras aplicaciones puedan acceder a ella para realizar estudios estadísticos y de gestión siempre que sea necesario, o en el caso de que el sistema experto actual sea sustituido por otro programa en el futuro.

Reaprovechamiento: Finalmente, es deseable que los elementos utilizados en la construcción del interfaz se encuentren separados del código del programa o de la base de conocimiento, con el fin de que pueda ser utilizado posteriormente en la realización de otras aplicaciones, sin tener que empezar el diseño desde cero. (A diferencia de los criterios anteriores, éste no se refiere a la interacción con el usuario, sino a la utilidad para el diseñador del programa.)

13.2.2 Descripción general del interfaz

En la realización del interfaz de usuario del sistema experto DIAVAL, hemos utilizado la herramienta comercial *GoldWorks*. Por un lado, porque es la misma con que hemos implementado el razonamiento bayesiano de nuestro sistema, con lo cual los dos aspectos, inferencia e interfaz, quedan totalmente integrados. Por otro, porque

GoldWorks resulta sumamente potente, cómodo y flexible a la hora de construir un interfaz gráfico para ordenadores personales tipo PC, aunque presenta el grave inconveniente de la lentitud, según comentaremos más adelante. Vamos a describir en lo que resta de este capítulo el interfaz de nuestro sistema experto, mostrando en qué medida se ajusta a las características deseables que hemos enumerado anteriormente.

Como rasgos generales, destacamos que el programa se ejecuta sobre *MS-Windows*, por lo que la información se presenta en forma de pantallas, que el usuario puede controlar mediante el ratón o mediante el teclado, en muchos casos indistintamente. De este modo garantizamos ya algunas de las características deseables señaladas, tales como la familiaridad, comodidad, realimentación (el usuario ve lo que está introduciendo) y posibilidad de navegar por el interfaz, abriendo y cerrando diferentes ventanas.

El *desarrollo de una sesión* comienza con la aparición de un menú de cinco opciones: introducir un nuevo eco, consultar los ecos anteriores, analizar estadísticamente la base de datos, realizar copias de seguridad o salir del programa. En caso de que el usuario decida introducir un nuevo eco, aparece un nuevo menú en que se le pregunta si el paciente tiene ecos anteriores. Si la respuesta es afirmativa, tratará de localizar al paciente en la base de datos (la búsqueda puede realizarse por nombre y apellidos, por D.N.I. o por número de eco), con el fin de extraer del último eco la información que no ha variado desde entonces: datos personales, antecedentes, etc. En cualquier caso, hay que consultar la base de datos para asignar un número al nuevo eco.

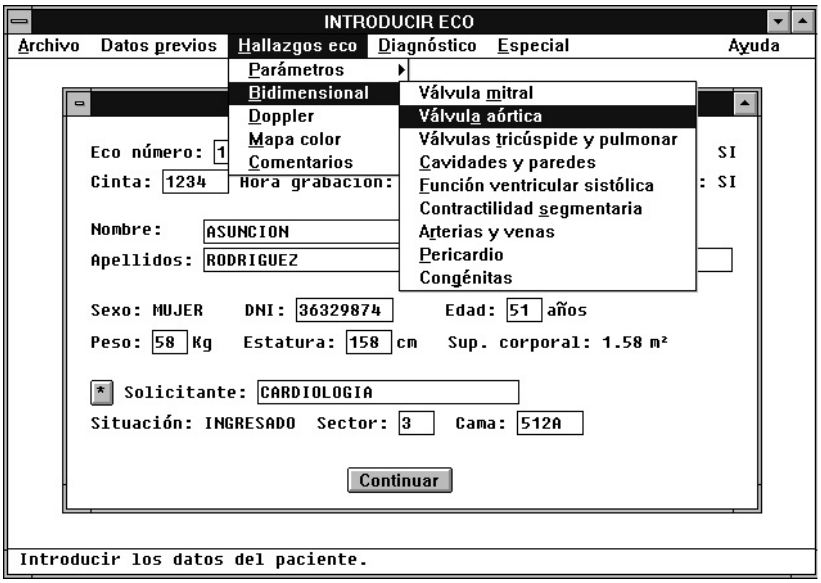


Figura 13.1: Pantalla para introducir un nuevo eco.

A continuación se entra ya en la pantalla principal, que contiene un título y una línea de comandos con varias opciones, los cuales permanecen fijos durante toda la sesión. Dentro de ella hay otra ventana que dará paso a otras a medida que avanza la sesión; la pantalla que aparece en la figura 13.1 está destinada a recoger los datos administrativos del paciente. Otras pantallas están dedicadas al motivo de petición, síntomas, intervenciones valvulares, ECG, parámetros del modo M, válvula mitral

(en el eco 2D), observaciones cualitativas del eco Doppler, comentarios, diagnósticos, médico que ha realizado el eco, etc., hasta un total de 28 pantallas disponibles en la versión actual. Muchas de estas pantallas se corresponden con los capítulos mencionados en la sección 13.1.1.

En la línea de comandos, las opciones Archivo, Especial y Ayuda, se incluyen por semejanza con la mayoría de los programas de *MS-Windows* (para satisfacer el criterio de familiaridad); al marcar con el ratón sobre la primera de ellas (o al pulsar las teclas Alt y A), se despliega un menú de persiana con nuevas opciones: Grabar, Fichas y Salir. Algunas de las opciones de un menú de persiana conllevan la ejecución de un comando (tal como abrir una nueva ventana) o pueden desplegar un nuevo menú del mismo tipo. De acuerdo con los criterios de comodidad y claridad, se recomienda que no haya más de dos niveles de persianas desplegables anidados.

Al pulsar el botón Continuar de la pantalla interior, ésta será reemplazada sucesivamente por otras relativas a la historia clínica del paciente: síntomas, antecedentes, etc. Sin embargo, para que el médico no tenga que pasar necesariamente por todas ellas (en muchos casos no habrá ninguna información que introducir), se ofrece primero un menú en que el usuario señala cuáles desea visitar. Las pantallas escogidas aparecerán en el orden preestablecido; cada una de ellas dispone de dos botones en su parte inferior, para que el usuario pueda avanzar o retroceder cuando lo desee. También es posible acceder a la línea de comandos y desde allí seleccionar la pantalla buscada.

Una vez recorridas todas las pantallas relativas a los datos previos, aparece un menú que permite indicar los capítulos en que todo está normal (ver sec. 13.1.1). Por ejemplo, si el usuario señala "Pericardio normal", eso significa que todos los nodos correspondientes este capítulo toman su valor normal; es decir, que no hay derrame pericárdico, ni engrosamiento, ni calcificación, etc. De este modo se evita tener que visitar la pantalla correspondiente.

Con el resto de las pantallas se forma un nuevo menú, para que el usuario señale las que desea visitar, pues en general hay varios capítulos sobre los cuales el médico no desea introducir información. Cuando se ha llegado a la última de estas pantallas y se pulsa el botón Continuar, el sistema pregunta al usuario si desea que el programa realice un diagnóstico. En caso afirmativo, comienza la propagación de evidencia en la red bayesiana con el fin de extraer las conclusiones, como luego explicaremos.

Por ahora, baste señalar que la opción de Ayuda está siempre accesible desde la línea de comandos. Al seleccionarla, se abre una ventana con la información relativa a la operación que estamos realizando. Por ejemplo, si la ventana interior es "DATOS ADMINISTRATIVOS", la ayuda explica cómo utilizar el ratón o el tabulador para moverse por los diferentes campos. Y así, existen ventanas de ayuda diferentes para las pantallas de parámetros, las que contienen información cualitativa, la de diagnóstico, la que indica el médico que ha realizado el eco, etc. Ésta es una de las formas de *ayuda sensible al contexto* disponibles en nuestro programa; existen otras, específicas de cada tipo de datos, que vamos a mencionar a continuación.

13.2.3 Entrada de datos

En esta sección describiremos los diferentes tipos de datos que admite nuestro sistema y las características de ayuda y comprobación disponibles para cada uno de ellos. Los hemos agrupado en cuatro clases: datos administrativos, parámetros, observaciones cualitativas e intervenciones.

Datos administrativos

Conviene volver de nuevo sobre la figura 13.1, en que aparece la ventana correspondiente a los datos personales. El programa asigna automáticamente el número de eco y la fecha, aunque este último valor puede ser modificado si el usuario lo desea. Hay también otros campos en que se introducen el nombre, los apellidos, la edad, el D.N.I., etc. Cada uno de ellos tiene asociada una función de comprobación. Así, el nombre y los apellidos deben ser cadenas de caracteres. El D.N.I. debe ser un número positivo de un máximo de 8 cifras. Para los demás campos numéricos existe un máximo, un mínimo, un "límite alto" y un "límite bajo". Si el valor introducido queda fuera del rango máximo-mínimo, el sistema muestra un mensaje de error; por ejemplo, si la edad es menor de 0 ó mayor de 200. Si el valor queda fuera del rango alto-bajo (tal como una de edad menor de 14 o mayor de 100), el sistema pide confirmación al usuario, pues tales valores son posibles pero muy poco frecuentes, por lo que es posible que el usuario haya cometido un error al teclearlos.[2]

La superficie corporal se calcula automáticamente en función del peso y de la talla del paciente, según la fórmula de Devereux.

El campo Solicitante lleva a su izquierda un botón que abre un menú, en el cual se muestran los diferentes servicios del hospital y los ambulatorios que han podido solicitar el eco. Su finalidad es que el usuario sólo tenga que teclear el nombre si se trata de un solicitante no habitual; en la mayor parte de los casos, la selección puede realizarse mediante el ratón.

Además de los valores que el usuario introduce mediante el teclado. hay cuatro campos, Sexo, Transtorácico, Transesofágico y Situación, que conmutan entre dos valores —VARÓN/MUJER, SÍ/NO o INGRESADO/AMBULANTE— al ser señalados con el ratón. También es posible seleccionarlos desde el teclado, mediante desplazamientos con el tabulador; de este modo se pretende dar la mayor libertad a los usuarios, pues cada uno de ellos se encuentra más a gusto con un tipo particular de interacción.

Los campos Sector y Cama sólo admiten un valor cuando el paciente está ingresado; en otro caso, aparece un mensaje de error. Los valores de estos dos campos se borran al indicar que el paciente está ambulante. Aunque estos valores no intervienen en el diagnóstico, tratamos de comprobar la consistencia de los datos introducidos siempre que sea posible.

Observamos también que en esta pantalla hay campos que tienen asignado un *valor por defecto*. Por ejemplo, el valor de la fecha viene dado por el día de hoy; igualmente, se supone que el eco es transtorácico y no transesofágico, que el solicitante es el Servicio de Cardiología, etc. El asignar valores por defecto es otra de las formas de ahorrar tiempo a la hora de introducir la información.

Parámetros

Hay otro tipo de datos, los parámetros, que corresponden en general a las medidas realizadas sobre el ecocardiograma. En el interfaz aparecen agrupados en diversas

[2]En el hospital en que hemos desarrollado nuestro sistema experto no suelen realizarse ecocardiogramas a menores de 14 años. Para otro hospital, habría que modificar el "límite bajo" de la edad.

pantallas, tal como muestra la figura 13.2. A cada uno de ellos le corresponde una línea, que consta de los siguientes elementos:

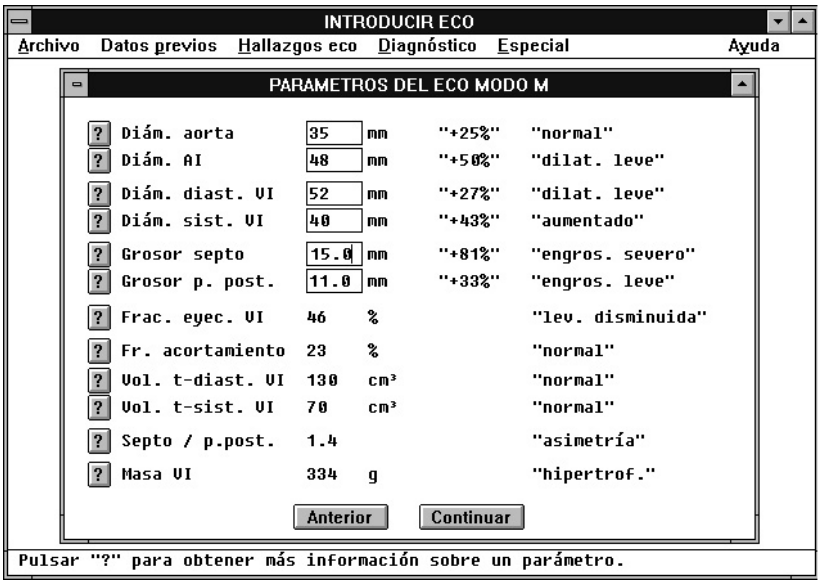


Figura 13.2: Parámetros del eco modo M.

- Un *botón*, marcado con el signo de interrogación, que sirve para solicitar ayuda sobre el parámetro; al señalarlo con el ratón —o mediante el teclado— se invoca la facilidad de ayuda, en que aparecen las opciones correspondientes a los parámetros, que serán algunas de las siguientes: valor, intervalos, significado, fórmula, imagen y pantalla, según se quedó explicado en la sección 8.3.2.

- El *nombre* del parámetro, escrito en forma abreviada.

- El *valor* medido, acompañado de la unidad correspondiente: mm, cm^2, cm/s, %, g, mmHg., etc.

- El *incremento* del parámetro con respecto al valor esperado. Para los parámetros que son —de forma estadística y aproximada— proporcionales a la superficie corporal, el valor esperado se calcula en función de ella (si está disponible, naturalmente) y así se indica en las pantallas de ayuda.

- El *intervalo* correspondiente al valor introducido. Para los valores que se correlacionan con la superficie corporal, los límites de los intervalos se toman proporcionales a ésta.

Al igual que en la pantalla de datos administrativos, el sistema comprueba cuidadosamente todos los datos y realiza las actualizaciones oportunas. Por ejemplo, cuando el usuario introduce el diámetro diastólico del ventrículo izquierdo (VI), el sistema comprueba que el valor se encuentra dentro de los límites admisibles: entre 1 y 200 mm.; en caso negativo, muestra un mensaje de error y rechaza el valor. Si el diámetro se encuentra dentro de los límites posibles pero tiene un valor muy poco

frecuente (menor de 15 mm. o mayor de 100 mm.), el programa solicita confimación al usuario, para asegurarse de que no se trata de un error tipográfico. Para este parámetro, se realiza otra comprobación adicional: el diámetro diastólico debe ser mayor que el sistólico, en caso de que éste haya sido introducido.[3] A continuación, se muestra el incremento con respecto al valor esperado (en la figura 13.2 es el 9%) y se le asigna una interpretación dependiendo del intervalo en que se encuentre (en nuestro ejemplo, es "normal"). Para este parámetro, el valor esperado es proporcional a la superficie corporal, en caso de que haya sido calculada en la pantalla de datos administrativos. Por último, se calculan todos los parámetros que dependen del diámetro diastólico: la fracción de eyección, el volumen telediastólico y la masa del VI, en el caso de que los demás parámetros que intervienen en los cálculos se encuentren ya disponibles.

Observaciones cualitativas

Hay también pantallas que recogen observaciones cualitativas, las cuales pueden estar asociadas a variables binarias o multivaluadas. Para cada uno de estos dos tipos de variables hemos definido un marco, como hijo de los marcos que ofrecía el interfaz gráfico de *GoldWorks*. Cada observación cualitativa viene representada en nuestro interfaz por una imagen, que será una instancia del marco apropiado.

1. *Variables binarias.* Tienen dos valores, anomalía (por ejemplo, "Crecimiento VI") y sin-anomalía ("No crecimiento VI"). Inicialmente, la imagen aparece en gris, indicando que aún no se ha introducido información. Cuando el usuario la señala con el ratón, aparece la anomalía en rojo, como valor anormal. Cuando vuelve a señalarla con el ratón, aparece el valor de sin-anomalía en color verde. Así, el botón izquierdo del ratón sirve para conmutar entre ambos valores; el botón central se utiliza para volver al color gris inicial, es decir, para retractar la información introducida; el botón de la derecha invoca la capacidad de explicación a partir del nodo de la red representado en dicha imagen.

2. *Variables multivaluadas.* Tienen varios valores posibles; por ejemplo, para la variable RITMO son ritmo sinusal, fibrilación auricular, "flutter", marcapaso y otros. En general, el primero de ellos corresponde al valor normal (el corazón sano suele estar en ritmo sinusal); en este caso, se mostrará en color verde al ser seleccionado. Los demás valores aparecen en gris. Al igual que en el caso de las variables binarias, se utiliza el botón central del ratón para cancelar la información introducida (la imagen vuelve entonces al color gris) y el botón derecho para invocar la facilidad de ayuda.

Otra forma de ayuda sensible al contexto consiste en la *línea de mensaje* que aparece en la parte inferior de la pantalla cuando el usuario sitúa el ratón sobre ciertas imágenes. En la figura 13.3, la línea de mensaje que aparece se debe a que, en el momento de capturar la pantalla, el ratón se encontraba sobre la imagen "Bloqueo AV".

[3] El motivo de realizar tantas comprobaciones es doble: por un lado, asegurar la consistencia de los datos para evitar conclusiones absurdas, como pudiera ser una fracción de eyección negativa; por otro, evitar que el sistema pueda abortarse o quedarse bloquedo, como podría ocurrir al dividir por 0.

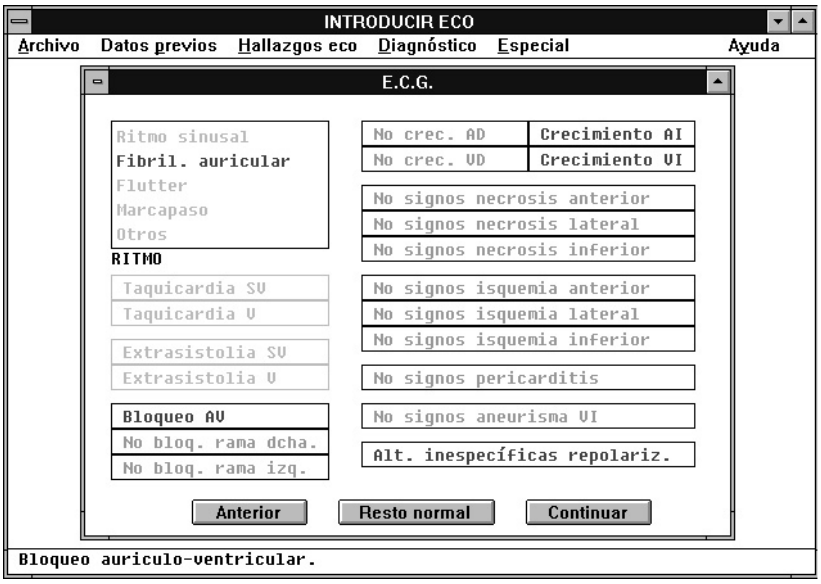


Figura 13.3: Interpretación del electrocardiograma.

Intervenciones

Los dos tipos de imágenes que acabamos de describir son insuficientes para abordar la complejidad de las distintas intervenciones posibles. Por eso, se vio la necesidad de diseñar un nuevo tipo de pantalla más flexible. La figura 13.4 muestra la ventana en que se indican las intervenciones valvulares que el paciente ha sufrido. Al pulsar uno de los tres botones, Valvuloplastia, Prótesis o Anuloplastia, aparece un menú diferente en que el usuario selecciona las características de la intervención: por ejemplo, válvula donde se realizó la anuloplastia, modelo de anillo implantado (se selecciona con el ratón), número, mes y año. Con esta información, el sistema elabora una *línea de texto* para cada intervención. Cuando el usuario selecciona una de ellas, aparece de nuevo en la pantalla el menú correspondiente, con el fin de que se pueda modificar o eliminar la información introducida.

13.3 Presentación de los resultados

13.3.1 Diagnóstico

Una vez que hemos realizado la propagación de evidencia sobre la red bayesiana, tenemos una distribución de probabilidad para los valores de cada variable. El método para llegar al diagnóstico se basa en el establecimiento de dos umbrales:

- *umbral de certeza* y
- *umbral de relevancia.*

El primero indica cuándo consideramos que hay suficiente certeza para poder establecer el diagnóstico. Su valor por defecto es 0'65, lo cual significa que, una vez realizada la inferencia y determinado el valor más probable de una variable, para

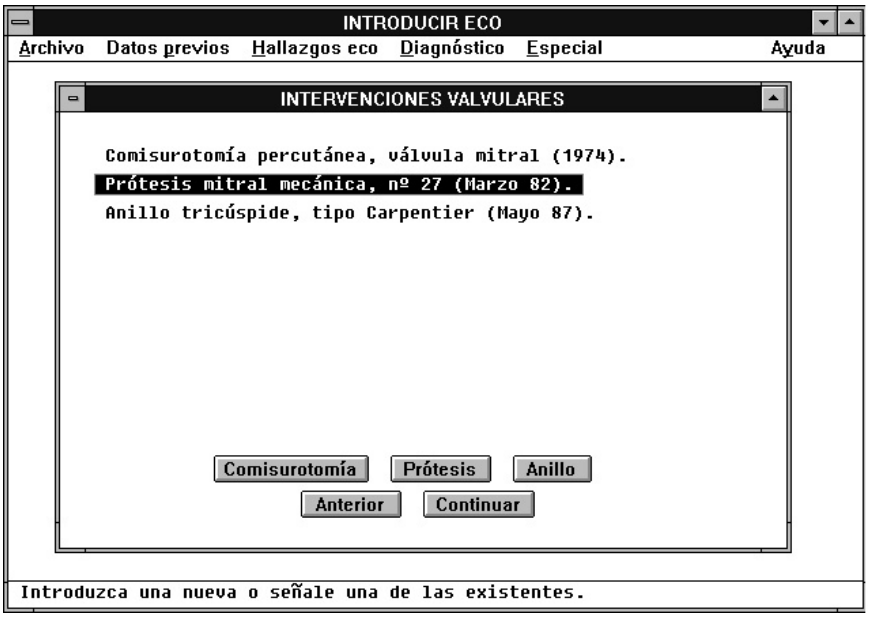


Figura 13.4: Pantalla para indicar las intervenciones valvulares.

aceptarlo como diagnóstico es condición necesaria que su probabilidad alcance como mínimo el 65%.

Sin embargo, no todas las variables cuyo valor se ha determinado van a formar parte del diagnóstico, por dos motivos. En primer lugar, porque la red contiene muchos nodos correspondientes a estados fisiológicos que, según los criterios médicos, no tienen relevancia suficiente como para figurar en las conclusiones de un informe. Por otro lado, un paciente suele tener como máximo tres o cuatro enfermedades, y todas las demás podrán excluirse con una certeza superior al umbral; en general, nos interesa mostrar sólo los diagnósticos positivos (las enfermedades encontradas), omitiendo los diagnósticos negativos.[4]

Por tanto, conviene asignar a cada nodo dos factores: la *relevancia para diagnóstico positivo* (RDP) y la *relevancia para diagnóstico negativo* (RDN). Los nodos correspondientes a las enfermedades más importantes, tales como la estenosis mitral o la miocardiopatía hipertrófica, tienen una RDP elevada; podemos asignarle un valor de 10. Su RDN es mucho menor: le hemos dado un valor de 5. A los estados patofisiológicos, como la hipertensión pulmonar o la alteración del llenado del ventrículo izquierdo, les asignamos generalmente una RDP cercana a 7 y una RDN cercana 3. Las alteraciones anatómicas (el engrosamiento del septo interventricular, por ejemplo), suelen tener una RDP próxima a 6 y un RDN de 2.

Con estas premisas, es fácil formar la lista de diagnósticos a partir de las probabilidades obtenidas. El proceso es el siguiente:

1. Examinar cada nodo para ver si alguno de sus valores supera el *umbral de probabilidad*. En caso afirmativo, asignar al nodo una relevancia, que será igual

[4]Diagnósticos positivos son, por ejemplo, la estenosis mitral leve, moderada y severa. La ausencia de estenosis es un diagnóstico negativo.

a la RDP o a la RDN, dependiendo de si el valor más probable es positivo o negativo, respectivamente. En el caso de variables graduadas (sec. 5.2.2), se calcula la suma de las probabilidades correspondientes a los diagnósticos positivos.

2. Seleccionar aquellos nodos que superan el *umbral de relevancia* y formar con ellos una lista de diagnósticos.

3. Eliminar de la lista un diagnóstico cuando existe otro más específico. El motivo es que no hace falta diagnosticar una enfermedad como la *estenosis mitral* si la información disponible nos ha permitido determinar la presencia de *estenosis mitral de origen reumático*.

Los valores de RDP y RDN no han de ser necesarimente fijos. En la primera versión del sistema experto, la única variación posible se debe al motivo de petición. Por ejemplo, si en la solicitud de ecocardiograma se dice "Descartar endocarditis" habrá que elevar la RDN del nodo correspondiente, pues en este caso un diagnóstico negativo es relevante. Lo mismo puede afirmarse en caso del seguimiento posterior a una intervención, para comprobar —por ejemplo— que no ha habido reestenosis. En el futuro, sería deseable que RDP y RDN variaran dinámicamente en función del proceso de inferencia, ya que si los hallazgos introducidos hacían pensar en la presencia de X, la ausencia de X también es significativa. Para ello, habría que definir alguna *medida de la sorpresa* para que en el diagnóstico (y posteriormente en el informe) quedasen recogidos los resultados inesperados.

En resumen, el usuario tiene dos grados de libertad en cuanto a la selección de los diagnósticos. Si desea que aparezcan sólo los más seguros, descartando los dudosos, deberá elevar el umbral de certeza. Si desea que aparezcan los más importantes, descartando los menos significativos, deberá elevar el umbral de relevancia.

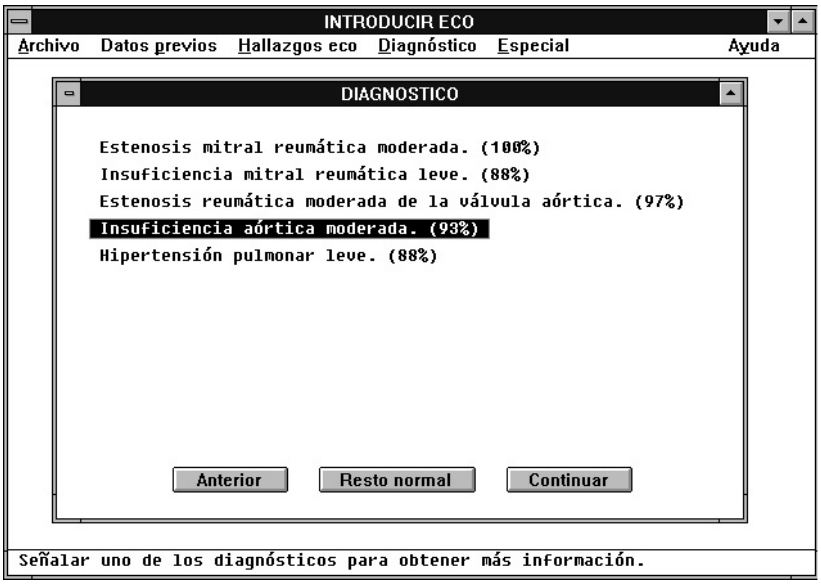


Figura 13.5: Diagnósticos obtenidos

La figura 13.5 recoge los diagnósticos obtenidos para el caso real de una paciente de 58 años, cuyo informe completo mostraremos más adelante. El umbral de certeza seleccionado era 0'65 y el de relevancia 7'5, los valores que hemos asignado por defecto al sistema a partir de la experiencia.

Como **extensiones** de este método, además de la variacíon dinámica de la relevancia diagnóstica, mencionada anteriormente, sería posible refinar el proceso de selección asignando una RDP diferente a cada valor de un nodo, dependiendo de su gravedad: así, la estenosis mitral severa tendría una RDP mayor que la moderada, y ésta mayor que la estenosis leve. Otra posible extensión consiste en establecer dos umbrales de relevancia distintos: uno para diagnósticos positivos (relacionado con la RDP) y otro para los negativos (que afectaría a la RDN), de modo que el usuario tenga un grado más de libertad a la hora de seleccionar los diagnósticos.

13.3.2 Explicación del razonamiento

La capacidad de explicación que describimos en el capítulo 8 se encuentra totalmente integrada con el interfaz. Ya hemos comentado anteriormente que es posible pulsar el botón que aparece a la izquierda de cada parámetro, con el fin de obtener informacion sobre los intervalos, el modo de obtención, la fórmula, etc.; también es posible solicitar información sobre el *significado* de un parámetro, es decir, sobre su nodo padre, y a partir de ahí seguir recorriendo los demás nodos de la red, mediante una serie de ventanas. Cuando se trata de una observación cualitativa, basta pulsar el botón derecho del ratón para invocar la facilidad de explicación, tal como dijimos en la sección 13.2.3.

Igualmente, desde la pantalla de diagnósticos es posible solicitar la explicación de los resultados. La pantalla de la figura 13.5 contiene cinco diagnósticos. Al seleccionar con el ratón uno de ellos, aparece una ventana con varias opciones; algunas de ellas son las mismas que hemos comentado en el capítulo 8: causas, efectos, prevalencia, etc. En el caso de que ya se haya realizado un diagnóstico, el usuario puede acceder además a una pantalla, tal como la que muestra la figura 13.6, en que aparecen de forma ordenada las causas del diagnóstico; desplazándose a través la red por medio pantallas como ésta es posible investigar cómo se ha propagado la evidencia.

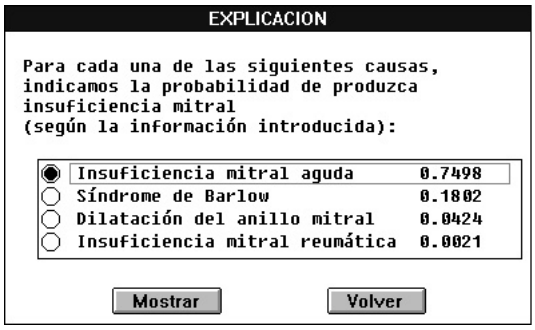


Figura 13.6: Explicación de las causas de un diagnóstico.

Otra de las opciones que aparecen al señalar con el ratón un diagnóstico es la posibilidad de modificarlo, asignando un nuevo valor al nodo correspondiente, y este

nuevo valor sustituirá al antiguo en la pantalla de diagnósticos. Por tanto, el usuario siempre tiene la posibilidad de corregir los resultados de la máquina cuando no está de acuerdo con ellos. El diagnóstico corregido figurará en el informe y se almacenará en la base de datos.[5]

En caso de que el usuario desee acceder a un diagnóstico que no haya sido seleccionado por el programa, se puede utilizar la línea de comandos de la pantalla principal, y mediante dos menús se llega al nodo buscado, con lo cual es posible solicitar explicación sobre ese nodo o incluirlo entre los resultados.

De este modo se satisface una exigencia totalmente razonable de los médicos: que sean ellos y no el programa quien decida cuál es el diagnóstico correcto. El sistema experto ofrece sus conclusiones pero es siempre el ser humano quien tiene la última palabra.

13.3.3 Informe final

Una vez que se ha establecido el diagnóstico y se ha indicado quién es el médico que ha realizado el eco (mediante una última pantalla), el sistema experto elabora el informe y lo muestra dentro de un editor de texto, con el fin de que el usuario pueda hacer las correcciones que desee.

La redacción del informe final es sencilla. Tiene dos partes, una en que se recogen los datos introducidos y otra en que se muestran las conclusiones. La primera consiste en mostrar de forma ordenada los **hallazgos** disponibles, clasificados por capítulos y filtrados según un tercer umbral, semejante al umbral de relevancia para el diagnóstico. A diferencia del caso del diagnóstico, para los datos no interviene el umbral de probabilidad, pues se supone que no existe incertidumbre sobre la información introducida por el usuario. La parte final consiste simplemente en transcribir los **diagnósticos** generados por el sistema y corregidos por el usuario y en añadir el nombre del médico que va a firmar el informe.

Transcribimos a continuación el informe elaborado por el sistema experto DIAVAL para el caso citado anteriormente; tan sólo se han cambiado algunos datos administrativos, que no influyen en el resultado del diagnóstico, y el nombre del médico que ha realizado el eco.

Datos administrativos:

N^o eco: 105337. Fecha: 18/6/94. Cinta: 649. Hora: 12805.
ASUNCIÓN RODRÍGUEZ GARCÍA. DNI: 36329874.
Edad: 51 años. Peso: 58 Kg. Estatura: 158 cm. Sup. corporal: 1'58 m^2.
Solicitante: Cardiología.
Ingresada, sector 3, cama 512A.

Síntomas:

Disnea de grado II.

E.C.G.:

Ritmo sinusal.

---

[5] Es necesario realizar la corrección en esta pantalla, y no en el informe mediante el editor de texto, con el fin de que en la base de datos se registre el valor correcto determinado por el médico.

Válvula mitral:
Area (eco 2D): 1'2 cm^2.
Velocidad E: 1'64 m/s.
Gradiente máximo: 10'4 mmHg.
Gradiente medio: 7'0 mmHg.
THP: 255 ms.
Area (THP): 0'88 cm^2.
Engrosamiento moderado de las valvas.
No calcificación de las valvas.
Movilidad levemente reducida.
Calcificación leve de las valvas.
Fusión moderada de las comisuras, simétrica.
Afectación moderada del aparato subvalvular.
Regurgitación leve.
Score mitral: 9.

Válvula aórtica:
Velocidad máxima: 3'65 m/s.
Gradiente máximo: 53'3 mmHg.
Válvula tricúspide.
Engrosamiento severo de las valvas.

Función ventricular normal.

Contractilidad segmentaria normal.

Pericardio normal.

Cavidades y paredes:
P.A.P.S.: 33 mmHg.

DIAGNÓSTICO:
Estenosis mitral reumática moderada.
Insuficiencia mitral reumática leve.
Estenosis reumática moderada de la válvula aórtica.
Insuficiencia aórtica moderada.
Hipertensión pulmonar leve.

Fdo.: Dr. Joaquín del Pozo

Observar que en el informe no aparecen la interpretación de los parámetros (ver la pantalla de la fig. 13.2) ni las probabilidades asociadas a cada diagnóstico (fig. 13.5), pues a pesar de nuestra sugerencia en este sentido, los médicos han insistido en que el informe escrito no debe contener esta información.

Cuando el usuario se encuentra satisfecho con el informe, bien porque era correcto desde el principio o porque lo ha modificado mediante el editor de texto, puede dar la orden de almacenarlo en la base de datos y enviarlo a la impresora.

13.4 Entorno de apoyo

Base de datos

La base de datos está diseñada en formato `.DBF`, compatible con el programa comercial *DBase III*, de modo que, aunque está generada por el sistema experto, resulta accesible a otros programas, con el fin de que pueda ser examinada para realizar estudios estadísticos diferentes de los que ofrecezca el propio sistema.

Existe un fichero principal, `ECOS.DBF`, en que se registran los datos administrativos, indexado por apellidos, por DNI y por número de eco, con el fin de realizar que la búsqueda sea más eficiente. En otro fichero, `PARAM.DBF` se almacenan los parámetros en los casos en se hayan medido. Las intervenciones poseen un formato especial, y por ello se recogen en un fichero independiente, `INTERV.DBF`. Por último, existe un otro fichero, `INFORMES.MEM` en que se archivan los informes finales.

Utilidades

Completaremos las capacidades del sistema con un módulo dedicado a la gestión de las copias de seguridad y otro a la realización de estadísticas sobre la base de datos. Son dos módulos que aún están pendientes de desarrollar, pues todavía no conocemos bien cuáles son las necesidades de los médicos, por lo que habrá que esperar a que el sistema experto lleve un tiempo en funcionamiento para obtener las especificaciones de diseño.

Capítulo 14

Evaluación de DIAVAL

14.1 Introducción

La construcción de un sistema experto quedaría incompleta si no se realizara una evaluación del programa resultante. En este capítulo vamos a revisar brevemente los métodos que se han propuesto en la literatura sobre el tema para pasar luego a describir la evaluación que hemos realizado y los resultados que hemos obtenido. Finalmente, propondremos un método de evaluación objetiva para nuestro sistema experto, a partir de las sugerencias encontradas en la bibliografía.

El primer sistema experto al que se aplicó una evaluación exhaustiva fue MYCIN, del que hemos hablado ya en varias ocasiones (sec. 2.1.4). Buchanan y Shortliffe [26, cap. 30] describen las tres evaluaciones que realizaron del programa y los resultados obtenidos. Inicialmente, los diseñadores estaban sorprendidos de que, a pesar de las mejoras introducidas en el programa, los expertos que lo juzgaban nunca le daban un índice de aceptación de los resultados superior al 75%; un estudio posterior reveló que el acuerdo entre los propios especialistas se sitúa entre el 70 y el 80%, por lo que los resultados de MYCIN podían considerarse más que satisfactorios. Otra de las lecciones aprendidas es que conviene pedir a expertos humanos que den su solución al mismo problema planteado al sistema, con el fin de que quienes juzgan los resultados no sepan si están evaluando a una máquina o a un ser humano. Esta especie de "test de Turing" es útil para evitar los prejuicios que los evaluadores puedan tener contra los ordenadores. Por último, los investigadores del proyecto MYCIN descubrieron la conveniencia de desarrollar cuestionarios breves con el fin de que los médicos participantes en la evalución devolvieran sus respuestas lo más pronto posible, pues en la segunda evaluación de MYCIN habían tardado un año en recoger todos los cuestionarios enviados. En [16] y [34] pueden encontrarse referencias sobre otros sistemas expertos famosos que también han sido evaluados de forma rigurosa.

Sintetizando las ideas de Suen et al. [227] y de Alun Preece [182], podemos distinguir tres tareas dentro del proceso de evaluación:

Verificación: Consiste en comprobar que el programa satisface las *especificaciones* planteadas inicialmente. La verificación del contenido del programa debe realizarse a nivel sintáctico (inconsistencias, redundancias, ciclos, etc.) y semántico (rangos de error, valores ilegales, reglas incorrectas, etc.).[1]

[1] En el desarrollo de programas convencionales, la verificación se ocupa sobre todo de la *consis-*

Validación: Su objetivo es determinar en qué medida el programa se comporta de modo satisfactorio a la hora de *resolver problemas* del mundo real. Suele realizarse mediante el examen de una serie de casos seleccionados por un grupo de expertos —como en el caso de MYCIN [266]— o extraídos de la literatura especializada —como en el caso de INTERNIST/CADACEUS [146]—.

Aceptación: Indica en qué medida el sistema resulta útil para el usuario al que está destinado. Es un aspecto que en la evaluación de los primeros sistemas expertos apenas se tenía en cuenta, pues éstos estaban concebidos como proyectos de investigación más que como sistemas destinados a su implantación masiva. Sin embargo, en la actualidad se reconoce que este punto es tan importante como los dos anteriores.

Entre los trabajos publicados sobre la evaluación, uno de los primeros que abarca prácticamente todos los aspectos del tema es el de Gaschnig et al. [92], quienes señalan la conveniencia de evaluar no sólo los resultados finales, sino también las conclusiones intermedias que genera el sistema, pues puede ocurrir que en el reducido conjunto de casos examinados se hayan obtenido resultados correctos de formal accidental, a pesar de utilizar un razonamiento erróneo.

Hay otro trabajo interesante, claro y muy práctico, en que Dianne Berry y Anna Hart [16] exponen los puntos que deben ser examinados en un sistema experto, y describen siete métodos para realizar la evaluación:

1. *Primer ensayo (walkthrough)*, en que el diseñador del sistema o los expertos que colaboran en el proyecto juzgan el sistema con mentalidad de usuario final. Se trata de un método muy informal, que sólo es válido para una evaluación orientativa del prototipo inicial.

2. *Cuestionarios.* Las preguntas pueden ser cerradas (es decir, con una lista limitada de respuestas posibles) o abiertas. La ventaja de las primeras es que se prestan más fácilmente a análisis estadísticos; la ventaja de las segundas es su flexibilidad. Tanto para unas como para otras, al formular las preguntas hay que tener cuidado de no predisponer al usuario hacia una respuesta concreta.

3. *Entrevistas.* Las autoras aconsejan que se realicen de forma individual, que haya una lista de preguntas preparada y que se graben en cintas magnetofónicas con el fin de que puedan ser transcritas posteriormente.

4. *Observación formal.* Consiste en tomar notas detalladas de la interacción del usuario con el programa, bien mediante observación directa o mediante grabaciones en vídeo. Puede ser útil pedir al usuario que piense en voz alta, expresando sus dificultades, dudas, comentarios, impresiones, etc.

5. *Registro de entrada (system logging).* Para poder utilizar este método, es necesario instalar en el ordenador una aplicación que almacene automáticamente las operaciones que el usuario realiza: pulsaciones de teclas, actuación sobre las ventanas, etc. La mayor dificultad que esta técnica plantea, a parte de contar

tencia y *completitud*. En los sistemas expertos, en cambio, la consistencia es más difícil de probar, y la completitud es casi imposible. Por eso en los sistemas expertos la verificación tiene menor importancia que la validación.

con la aplicación necesaria, es el análisis de la gran cantidad de información recogida.

6. *Experimentos simples.* Sirven para comparar dos versiones del sistema o dos posibles implementaciones de algún aspecto particular. En el diseño de la prueba hay que seguir los pasos tradicionales del método experimental: seleccionar la hipótesis de trabajo y las variables dependientes e independientes, asignar los sujetos a los grupos, tratar de controlar las posibles fuentes de error y, finalmente, realizar un análisis estadístico de los datos obtenidos.

7. *Diarios.* El pedir al usuario que redacte un diario en que recoja sus comentarios y sugerencias a medida que utiliza el sistema puede ser muy útil para la evaluación del sistema a largo plazo. Sin embargo, en la práctica es difícil conseguir que los usuarios registren en el diario de forma sistemática toda la información que el diseñador del programa desearía.

Por último, estas autoras recomiendan llevar a cabo *pruebas de campo*, para ver cómo se comporta el sistema en el entorno en que va a funcionar finalmente, pues hay muchos aspectos de la evaluación del sistema que resultan difíciles de medir mientras se está desarrollando el programa en un laboratorio.

Existen otras muchas referencia,s bibliográficas que aportan sugerencias y consejos sobre este tema. Wyatt y Spiegelhalter [264], por ejemplo, reflexionan sobre los aspectos legales de los sistemas expertos médicos y ofrecen además sugerencias sobre cómo prevenir los errores (*biases*) que pueden surgir en la evaluación de los sistemas expertos. Suen et al. [227] insisten en la conveniencia de realizar tanto la verificación como la validación desde las primeras etapas de desarrollo del sistema. O'Keefe, Balci y Smith [162], entre otros muchos consejos prácticos, proponen un método de análisis de la sensibilidad que consiste en partir de un caso en que el sistema experto se comporta adecuadamente y realizar variaciones en la entrada de datos para determinar los límites de validez del sistema.

En cuanto a la *evaluación del interfaz*, Sherman [207] realiza un amplio estudio bibliográfico y clasifica en cuatro grupos los métodos existentes: técnicas empíricas, tales como entrevistas y cuestionarios, técnicas analíticas (las que implican un tratamiento matemático de los datos), modelos teóricos, que no necesitan la participación de un usuario en los experimentos, y técnicas especiales diseñadas para casos particulares.

14.2 Evaluación realizada

14.2.1 Método de evaluación

Como vemos, son muchas las posibilidades a la hora de realizar la evaluación de un sistema experto. En la práctica, la elección de un método u otro dependerá de los recursos económicos, de la disponibilidad de expertos dispuestos a colaborar y de las limitaciones de tiempo existentes.

Para la evaluación de DIAVAL, hemos utilizado en mayor o menor medida los cuatro primeros métodos de los siete enunciados anteriormente: a parte de los ensayos previos, hemos pedido a varios médicos que pensaran en un paciente real o

imaginario e introdujeran en el sistema experto los datos correspondientes. A la vez que les explicábamos el funcionamiento del programa, hemos ido observando las operaciones que realizaban y recordando los comentarios que nos hacían, con el fin de transcribirlos posteriormente. Una vez completada la sesión, el médico debía responder el cuestionario que vamos a mostrar en seguida.

Como se puede observar, en él aparecen preguntas cerradas, con cinco respuestas posibles, y preguntas abiertas, que en el impreso que se ha entregado a los médicos contenían líneas en blanco para que éstos pudieran expresar sus opiniones. Para las primeras, expresamos entre corchetes el promedio de las respuestas y la desviación estimada, s_{n-1}. Con estas 25 preguntas hemos tratado de evaluar los aspectos más significativos del programa.

Cuestionario sobre DIAVAL

1. **Modelo cardiológico:**

   - ¿Es correcta la información cualitativa (la relación de causas y efectos en la red)?

     Correcta 1. 2. 3. 4. 5. Incorrecta *[R: 1'4 ± 1'3]*

   - En general, ¿es correcta la información cuantitativa (las probabilidades que aparecen en la red)?

     Correcta 1. 2. 3. 4. 5. Incorrecta *[R: 2'1 ± 1'1]*

   - ¿Considera que los datos y diagnósticos que incluye el modelo son suficientes para abordar el campo de la ecocardiografía?

     Suficientes 1. 2. 3. 4. 5. Insuficientes *[R: 1'6 ± 0'9]*

2. **Interfaz:**
   (Nos referimos al sistema de pantallas, menús, etc.)

   - ¿Es fácil de aprender?

     Fácil 1. 2. 3. 4. 5. Difícil *[R: 1'3 ± 0'5]*

   - ¿Es cómodo y agradable de manejar?

     Cómodo 1. 2. 3. 4. 5. Engorroso *[R: 2'1 ± 1'3]*

   - ¿La interacción con el programa es flexible o se siente forzado/a por él?

     Flexible 1. 2. 3. 4. 5. Forzado/a *[R: 1'7 ± 0'8]*

   - ¿Cómo le parece la presentación de datos y resultados?

     Clara 1. 2. 3. 4. 5. Confusa *[R: 1'3 ± 0'5]*

   - La ayuda que ofrece el sistema para su manejo, ¿es suficiente?

     Suficiente 1. 2. 3. 4. 5. Insuficiente *[R: 1'3 ± 0'6]*

3. Explicación

- ¿Considera útiles las explicaciones incluidas en el programa: intervalos de los parámetros, causas, efectos, prevalencia, etc.?

  Útiles 1. 2. 3. 4. 5. Inútiles *[R: 1'2± 0'8]*

- ¿Le parecen suficientes?

  Suficientes 1. 2. 3. 4. 5. Insuficientes *[R: 1'4± 1'1]*

- ¿Son fáciles de comprender?

  Comprensibles 1. 2. 3. 4. 5. Complicadas *[R: 1'7± 1'5]*

4. Diagnóstico

- En general, ¿es correcto el diagnóstico que ofrece el programa?

  Correcto 1. 2. 3. 4. 5. Incorrecto *[R: 2'0± 1'0]*

- ¿Le parece útil que el programa realice un diagnóstico?

  Útil 1. 2. 3. 4. 5. Inútil *[R: 2'4± 1'5]*

- Comentarios sobre el diagnóstico.

5. Informe

- ¿Es correcto en su contenido?

  Correcto 1. 2. 3. 4. 5. Incorrecto *[R: 1'4± 0'6]*

- ¿Es adecuada la forma de presentar las conclusiones?

  Adecuada 1. 2. 3. 4. 5. Inadecuada *[R: 1'2± 0'4]*

- Comentarios sobre el informe.

6. Comentarios generales

- Valoración general del programa:

  Positiva 1. 2. 3. 4. 5. Negativa *[R: 1'6± 0'7]*

- Si realiza o tuviera que realizar informes ecocardiográficos, ¿le gustaría trabajar con un programa como éste?

  Sí 1. 2. 3. 4. 5. No *[R: 1'4± 0'6]*

- ¿Por qué?
- ¿Qué es lo que más le ha gustado del sistema?
- ¿Y lo que menos le ha gustado o le ha decepcionado?
- ¿Qué añadiría al sistema?
- ¿Qué suprimiría?
- Otros comentarios, sugerencias, opiniones...

14.2.2 Resultados obtenidos

Para la evaluación del sistema experto hemos contado con seis cardiólogos: tres residentes y tres especialistas veteranos, uno de ellos dedicado a la ecocardiografía pediátrica. Una primera valoración cualitativa nos permite observar que los juicios de los observadores han sido bastante positivos, aunque llama la atención la dispersión de las respuestas a algunas cuestiones. Examinemos ahora por separado cada uno de los apartados del cuestionario.

Modelo cardiológico

Como era de esperar, ha sido mucho más positiva la valoración de la información cualitativa que la de la cuantitativa. Esto se debe a que la primera es mucho más objetiva y más fácil de obtener que las probabilidades que aparecen en la red. En general, hemos comprobado que muchas de las probabilidades condicionales que contiene la red son demasiado elevadas, como veremos al hablar del diagnóstico.

Sin embargo, debemos señalar que, al diseñar el cuestionario, pensábamos que los usuarios iban a mostrar gran interés por conocer el modelo bayesiano, utilizando la capacidad de explicación que ofrece el sistema experto, pero no ha sido así. La presión de trabajo que existe sobre ellos nos ha impedido demostrarles con detalle todas las capacidades del programa y, por otro lado, les ha impedido dedicar el tiempo suficiente a explorar el modelo, por lo que, a nuestro juicio, las respuestas dadas a estas dos preguntas se basan más en una primera impresión que en un conocimiento real de la cuestión.

En cuanto a la amplitud del modelo, los expertos han estado de acuerdo en que es más que suficiente para abordar el campo de la ecocardiografía, según nos lo han manifestado también oralmente.

Interfaz

El interfaz ha recibido una calificación muy positiva, sobre todo la facilidad de manejo, la claridad en la presentación de los resultados y las ayudas que ofrece. El consenso ha sido ligeramente inferior a la hora de evaluar la comodidad y la flexibilidad. Según las opiniones expresadas verbalmente, este resultado puede deberse a las distintas experiencias que tenían los usuarios con respecto a otros programas, las cuales han influido en sus expectativas frente al nuestro.

Estar junto a los usuarios mientras interactuaban con el sistema experto nos ha permitido observar además la facilidad y rapidez con que algunos de ellos, familiarizados con el entorno *MS-Windows*, aprendieron a manejar el programa. Otros, en cambio, tenían dificultades en el uso del ratón y en el control del cursor.

Explicación

Los médicos han apreciado notablemente la utilidad de las explicaciones, en especial la presentación de los intervalos correspondientes a cada parámetro, tal como muestra la figura 8.2 (pág. 158). Sin embargo, sería deseable que el programa ofreciera explicaciones más completas y más claras.

A nuestro juicio, la explicación del modelo es suficiente, pero la explicación del razonamiento debe desarrollarse mucho más, para que el usuario pueda comprender el porqué de las conclusiones, especialmente cuando éstas son erróneas.

Diagnóstico

Es en este punto donde se da la mayor disparidad en las respuestas. En primer lugar, señalemos que en varias ocasiones el sistema experto ha sorprendido a los médicos que lo juzgaban —e incluso a su propio diseñador— por la precisión de los diagnósticos, aunque otras veces ha generado resultados incorrectos. Según las leyes de la estadística, el hecho de que cada médico haya introducido un solo caso de prueba en el sistema puede explicar que haya tales diferencias en la valoración del diagnóstico.

Otra cuestión diferente es juzgar la utilidad de que el sistema ofrezca un diagnóstico. Algunos médicos lo aprecian positivamente, mientras que a otros les parece inútil, dado que el médico ya ha extraído sus conclusiones mientras realizaba el eco, por lo que creen que el sistema experto no puede aportarles nada nuevo.

A nuestro juicio, las diferencias tan acusadas en las opiniones de los médicos se deben sobre todo a la distinta mentalidad de cada uno de ellos. Hay quienes intentan que la ecocardiografía sea una técnica lo más objetiva posible, y por eso valoran positivamente un programa que les ayuda a evaluar matemáticamente la información recogida. En cambio, otros especialistas se fían más de sus impresiones subjetivas en la observación del eco y desconfían de la información que les ofrezca un computador.

Hablando de la evaluación del diagnóstico, conviene comentar los tres tipos de errores que, a nuestro juicio, comete con más frecuencia el sistema DIAVAL. El primero de ellos es que a veces genera resultados de más, tales como la hipertensión pulmonar, sin que haya evidencia que lo justifique. También tiende el sistema a dar diagnósticos demasiado específicos; así, en ocasiones puede deducir que la insuficiencia es aguda cuando de los datos introducidos sólo permiten concluir que hay insuficiencia. Otra equivocación consiste en asignar a una enfermedad un grado mayor del que le corresponde, diagnosticando, por ejemplo, estenosis severa en vez de estenosis leve.

Estos tres errores tienen un origen común: la asignación de probabilidades excesivamente altas a los enlaces causales, por lo que el sistema tiende a afirmar con demasiada certeza la presencia del efecto cada vez que recibe evidencia a favor de la causa, o a deducir que una de las causas es mucho más probable que las otras.

Informe

Dentro del funcionamiento del sistema DIAVAL, el aspecto que ha producido la impresión más favorable entre los médicos que lo juzgaron es la elaboración de un informe en que los datos y los diagnósticos aparecen de forma ordenada, concisa y completa. Este comportamiento del sistema es consecuencia de la agrupación de los datos y diagnósticos en capítulos, según hemos explicado anteriormente en la sección 13.1.1.

Impresión general

La valoración general del sistema ha sido bastante positiva. Más aún, nos ha sorprendido que, incluso en los cuestionarios que contenían las respuestas más críticas

a preguntas anteriores, se manifestaba el interés o el deseo trabajar con un sistema como éste.

Entre las cualidades que los evaluadores han destacado al responder a las cuestiones abiertas, se repiten sobre todo la facilidad de manejo, la claridad y la comodidad que ofrece el programa. Uno de los encuestados señala como virtud del sistema la objetividad y otro las futuras posibilidades de desarrollo.

Entre las deficiencias, los usuarios han sido unánimes al quejarse de la lentitud del programa, la cual es consecuencia de la complejidad de las operaciones que realiza, tanto en la presentación de las pantallas del interfaz como en el razonamiento bayesiano. Esta situación se ha visto agravada por la utilización de un ordenador relativamente lento, un 386 de 33 Mhz., cuyos 12 Mb. de memoria estaban casi saturados, de modo que el Lisp, aun habiendo compilado todas las funciones, debía pararse frecuentemente a "recoger basura" (ver la sec. 9.1.3); para el usuario resulta desesperante tener que esperar durante 10 segundos a que se reanude la ejecución del programa, y estas paradas se repiten con demasiada frecuencia. Ciertamente, la velocidad es el primer aspecto que debe ser mejorado si queremos que el sistema experto llegue a ser operativo.

14.2.3 Discusión

Los resultados obtenidos, aunque sean por sí mismos sugerentes, deben ser tomados con precaución, pues la evaluación realizada está lejos de cumplir los requisitos de un experimento científico. El primer motivo es el tamaño de la muestra. Hemos comentado ya que cada médico ha introducido en el sistema un sólo caso de prueba, por lo que en los resultados obtenidos pueden tener excesiva influencia las variaciones estadísticas. El número limitado de participantes —hemos contado tan sólo con seis médicos— también contribuye a introducir variaciones estadísticas significativas.

Hay que tener en cuenta además los factores humanos que cuestionan la validez de los resultados. La amistad que nos une a los evaluadores del sistema y el hecho de que los cuestionarios no fueran completamente anónimos han podido conducir a un juicio más benévolo que el que hubieran dado observadores independientes. En sentido contrario, las deficiencias que contenía el sistema experto en el momento de realizar la evaluación, tanto en el código del programa como en la base de conocimiento, el inconveniente de trabajar con un ordenador relativamente lento y escaso de memoria, y las presiones de tiempo que pesaban sobre los evaluadores han contribuido probablemente hacia un juicio peyorativo de ciertos aspectos del programa.

A pesar de estas deficiencias, la evaluación que hemos realizado nos permite extraer algunas conclusiones. Por un lado, tanto las impresiones subjetivas de los usuarios como la calidad de los diagnósticos nos alienta a completar la implatación del sistema en el servicio de ecocardiografía para el que fue concebido. Por otro, se ve claramente la necesidad de optimizar el código del programa con el fin de ahorrar tiempo y memoria. La utilización de un ordenardor más potente, tal como un 486 de 66 Mhz. con 16 Mb. de memoria, puede multiplicar notablemente la velocidad de ejecución.

En cuanto a los diagnósticos que ofrece el programa, es preciso revisar las probabilidades contenidas en la red, con el fin de corregir los resultados erróneos que aparecen en algunas ocasiones.

14.3 Propuesta de evaluación objetiva

La evaluación que hemos realizado se basa sobre todo en estimaciones subjetivas y, por las razones señaladas anteriormente, su validez es muy limitada. Por tanto, sería deseable contar con una evaluación rigurosa una vez corregidas las deficiencias que se han detectado.

Una de las dificultades que se plantean en el campo de la medicina al intentar medir objetivamente la calidad de un sistema experto es que no siempre se conoce *la* respuesta correcta: hay veces en que ni siquiera la autopsia de una persona aclara cuál era la enfermedad que padecía. En consecuencia, no es posible contar con un criterio absolutamente fiable para juzgar los resultados del sistema, por lo que se hace necesario establecer comparaciones entre los diagnósticos que emiten diferentes especialistas sobre un mismo conjunto de casos.

Nuestra propuesta de evaluación objetiva consiste en examinar una determinada serie de casos clínicos reales. Nos parece conveniente que dentro de ellos haya algunos *casos típicos*, escogidos al azar entre los pacientes ambulantes e ingresados, y también casos especiales, seleccionados para cubrir las *enfermedades poco frecuentes*. Cada uno de ellos contendrá una breve historia clínica del paciente, que incluya los datos personales, antecedentes y el ECG, más los parámetros y las observaciones cualitativas procedentes del ecocardiograma. Necesitamos, por tanto, la colaboración de un médico que nos ayude en esta selección, a ser posible distinto de los que van a juzgar los informes, con el fin de que no haya interfencias en el experimento.

El tamaño de la muestra debe ser suficientemente amplio para que resulte estadísticamente significativo y a la vez no puede ser demasiado grande, teniendo en cuenta que cada médico tendrá que examinar uno por uno todos los casos seleccionados. Nuestra estimación es que un conjunto de 15 casos, 10 típicos y 5 infrecuentes, puede ser un compromiso razonable.

Posteriormente se solicitará a varios médicos, especialistas veteranos y residentes, que señalen qué diagnóstico les parece más probable y cuáles otros podrían considerarse como razonables. A los diagnósticos realizados por los médicos uniremos el deducido por el sistema DIAVAL.

A partir de ahí, existen dos posibilidades de valorar los datos obtenidos. Una de ellas consiste en realizar un análisis estadístico para medir el acuerdo entre los diferentes diagnósticos. La otra posibilidad consiste en realizar una evaluación ciega, en que los médicos participantes juzguen las conclusiones de los demás, incluido el sistema experto.

Creemos que esta propuesta puede significar una evaluación rigurosa, sin que su complejidad exceda los recursos que tenemos a nuestro alcance. El único requisito necesario sería contar con la participación de especialistas procedentes de otros hospitales, para que la evaluación del sistema sea completamente objetiva.

Parte IV

CONCLUSIONES

Capítulo 15

Resultados obtenidos

15.1 Lecciones aprendidas

Existe una regla que dice: "La terminación de una tesis doctoral siempre lleva un año más de lo previsto"; y hay quien añade "... incluso después de aplicar esta regla". En nuestro caso, el plazo de los dos años se ha cumplido al pie de la letra. Ciertamente, ésta es una de las principales lecciones aprendidas, y lo citamos aquí porque, al tratarse de un sistema experto para medicina, este fenómeno puede tener consecuencias importantes de cara a la culminación del proyecto.

En efecto, a diferencia de otros temas de investigación, en que el doctorando depende solamente de los medios técnicos y de su propio esfuerzo, en la realización de un verdadero sistema experto se hace imprescindible la colaboración de uno o varios especialistas, por lo que hay que tener muy en cuenta los factores humanos de los que hablábamos en la sección 11.3.3, y el principal de ellos es la *motivación*. Si no se cuida adecuadamente este punto, es posible que los expertos abandonen antes de concluir el proyecto. En nuestro caso, el contar con un grupo amplio de médicos nos ha salvado de este peligro; pero conviene aprender la lección.

Como consecuencia se deduce que es necesario establecer claramente desde el principio los compromisos que asume cada persona: cuánto tiempo va a dedicar el experto y qué resultados va a aportar el diseñador del sistema. Éste es uno de los puntos que deben quedar resueltos en la fase de identificación (sec. 11.1). La construccción temprana de un pequeño prototipo puede ayudar a eliminar algunos prejuicios y a mantener el interés de los especialistas colaboradores, pero es necesario además que el proyecto se concluya lo antes posible.

Por tanto, conviene que, antes de empezar la interacción con los médicos o con los expertos de que se trate, el diseñador del programa tenga resueltos los problemas teóricos y que incluso disponga de una implementación del mecanismo de inferencia, con el fin de que transcurra el menor tiempo posible desde que se inicia la obtención del conocimiento hasta que el sistema empieza a funcionar.

Sin embargo, este deseo entra en conflicto con los objetivos de una tesis doctoral, que no puede limitarse a aplicar a un caso particular las técnicas conocidas. Generalmente, al principio sólo existe una idea bastante difusa de la metodología que se va a utilizar, y las ideas van surgiendo e implementándose a medida que se conoce mejor el problema en cuestión. En nuestro proyecto, el estudio teórico, la creación de nuevos algoritmos y su implementación han supuesto una serie de aportaciones

interesantes de cara al doctorado, pero a costa retrasar considerablemente la culminación del proyecto, poniendo en peligro su culminación, debido al desánimo de los médicos colaboradores.

En resumen, hay que buscar un compromiso entre el objetivo de aportar resultados novedosos para la tesis doctoral y la necesidad de obtener resultados concretos y concluir el proyecto en el menor tiempo posible. No existe una solución sencilla, pero conocer esta dificultad puede ayudar considerablemente en el planteamiento inicial.

15.2 Principales aportaciones

Resumimos en esta sección las principales aportaciones incluidas en esta tesis doctoral, separando las que se refieren a las redes bayesianas en general de las relativas al sistema experto DIAVAL.

15.2.1 Aportaciones a las redes bayesianas

En los capítulos anteriores, especialmente en los de la parte II, hemos intentado ofrecer una visión completa sobre este campo. Salvo desconocimiento por nuestra parte, la presente tesis doctoral es la primera obra que trata todos los aspectos de las redes bayesianas: axiomas, semántica, algoritmos, aprendizaje, explicación, obtención y representación del conocimiento, etc. En cuanto a las aportaciones novedosas, destacaríamos las siguientes, especialmente la primera:

1. **Condicionamiento local:** Es un nuevo algoritmo para la computación exacta de la probabilidad en redes bayesianas. Consiste en una extensión del algoritmo desarrollado por Kim y Pearl (sec. 5.1). La diferencia principal de nuestro método con respecto al anterior es que no solamente puede aplicarse a poliárboles, sino también a redes con bucles. El condicionamiento local es mucho más eficiente que las técnicas de condicionamiento anteriores, y supera en varios aspectos a los algoritmos de agrupamiento, que hoy en día constituyen el método estándar para la propagación de evidencia.

2. **Puerta OR graduada:** Este modelo había sido propuesto informalmente por Max Henrion [110] como simplificación de la obtención del conocimiento. En la sección 5.2 hemos definido matemáticamente el modelo y hemos desarrollado un algoritmo que permite la computación de la probabilidad en tiempo proporcional al número de padres del nodo (los algoritmos anteriores requerían un tiempo exponencial). Este método es aplicable también a redes con bucles, con la condición de que la inferencia se realice mediante condicionamiento local, pues otros métodos de propagación son incapaces de aprovechar las ventajas de la puerta OR. La diferencia entre el modelo general y este modelo particular no es solamente una cuestión matemática, sino que responde a una semántica diferente del concepto de causalidad (sec. 5.2.5). La puerta OR graduada, que aparece con muchísima frecuencia en sistemas expertos reales (en medicina, es la forma habitual de interacción entre las causas), simplifica la obtención del conocimiento, ahorra espacio en la representación de la red, acelera considerablemente la computación de la probabilidad y ofrece posibilidades de explicación que no se dan en el caso general.

3. **Modelo de aprendizaje secuencial:** Ante los problemas que plantea el aprendizaje por lotes (sec. 7.1.2), hemos decidido desarrollar un modelo de aprendizaje secuencial (sección 7.2); aunque es muy similar a los propuestos por Spiegelhalter y Lauritzen [217], posee dos ventajas: *(1)* por basarse en distribuciones gaussianas de probabilidad, acepta de forma natural los resultados de investigación que aparecen en la literatura médica, y *(2)* se puede aplicar también a la puerta OR graduada, directamente, manteniendo un número reducido de parámetros. Además, se integra completamente con el algoritmo de condicionamiento local ya mencionado.

4. **Modelo de explicación:** La sección 8.3 desarrolla un modelo de explicación local en redes bayesianas, basado en la distinción entre dos tipos de interacción y seis clases de enlaces, y deja abiertas varias posibilidades de extensión para el futuro. La principal innovación de este modelo es que describe también el conocimiento contenido en la red, no solamente las conclusiones de la inferencia, y puede explicar tanto los resultados finales como los cambios producidos al introducir nueva evidencia. Hemos visto además cómo integrar el método de explicación dentro de un interfaz gráfico para sistemas expertos bayesianos.

En cuanto a la presentación de los temas, hemos pretendido darle una organización coherente, adhiriéndonos a la descripción en tres niveles que propone David Marr [140, sec. 2.1]: **teoría**, **algoritmo** e **implementación**, mostrando la posibilidad de realizar una *implementación distribuida* para cada algoritmo de las redes bayesianas.

15.2.2 Construcción de un sistema experto bayesiano

Las redes bayesianas constituyen un tema de investigación eminentemente matemático. Sin embargo, a la hora de aplicar a un problema del mundo real las ideas desarrolladas, se plantean una serie de cuestiones de muy distinta índole, que hemos debido abordar en la construcción del sistema experto DIAVAL. Algunas de ellas son las siguientes:

1. **Obtención del conocimiento:** Se han publicado varios libros y numerosos artículos sobre cómo superar este "cuello de botella", pero casi todos se refieren a sistemas basados en reglas; los trabajos de Spiegelhalter et al. [215] y de Luis Enrique Sucar et al. [226] son las únicas excepciones que conocemos. La sección 11.3 describe cómo se ha obtenido el conocimiento necesario para construir un sistema experto médico, señalando las dificultades propias de este campo y los factores humanos que intervienen, con el objetivo de dar una visión realista del problema. También hemos intentado resumir los conceptos médicos que nos parecen más útiles de cara a la construcción de un sistema experto.

2. **Representación del conocimiento:** Una vez obtenido el conocimiento (en la fase de *conceptualiación*), hay que pasar a su representación computacional (es la fase que se conoce como *formalización*), La distinción entre dos tipos de interacción y seis clases de enlaces (influencia, parámetro, causalidad, tipo, manifestación y observación) es un esquema útil a la hora de construir un modelo patofisiológico en forma de red bayesiana, y sirve de base al modelo de explicación propuesto en esta tesis. De este modo, el sistema resultante es capaz de integrar en un mismo formalismo la inferencia, la explicación y el aprendizaje.

3. **Implementación:** El elemento fundamental de un sistema de razonamiento bayesiano es el mecanismo de propagación de evidencia. Sin embargo, para que pueda comportarse como sistema experto necesita además un interfaz que permita al usuario introducir la información y examinar los resultados. En nuestro proyecto hemos tratado de poner el máximo cuidado en la elaboración de un interfaz gráfico sencillo y amigable, dotado de facilidades de comprobación de datos, ayuda y explicación, conforme a las características ideales sugeridas en la literatura. El seleccionar los diagnósticos y redactar el informe en base a tres umbrales (sec. 13.3) es una de las novedades introducidas en el presente trabajo.

La evaluación del sistema experto ha resultado satisfactoria en términos generales y ha servido para determinar los aspectos que aún deben mejorarse. Una vez que se haya optimizado el código del programa para ganar en velocidad y se hayan ajustado las probabilidades condicionales de la red, dos tareas relativamente sencillas, el sistema experto DIAVAL se encontrará listo para entrar en funcionamiento, en espera solamente de que el servicio de ecocardiografía para el que está destinado disponga del ordenador adecuado.[1] Existe por nuestra parte el compromiso de instalar el programa y realizar su mantenimiento como parte del proyecto, con el fin que la presente tesis doctoral no sea solamente un trabajo de investigación, sino que pueda resultar útil en el campo de la medicina, donde las necesidades asistenciales son cada día mayores.

15.3 Valoración del sistema experto DIAVAL

Donald Waterman [246, págs. 139–141] distingue cinco etapas en la evolución de un sistema experto: prototipo de demostración, prototipo de investigación, prototipo de campo, modelo de producción y sistema comercial. Teniendo en cuenta el tiempo invertido, la extensión de la base de conocimiento, las características del interfaz y el hecho de que ya se encuentra funcionando experimentalmente en un entorno real, el sistema experto DIAVAL puede ser considerado como *prototipo de campo*; de los requisitos señalados por Waterman, el único que aún le falta por cumplir es una evaluación exhaustiva sobre mayor número de casos. Con una implementación más eficiente, podrá convertirse en un modelo de producción e incluso en un sistema comercial.

En cuanto a características de DIAVAL, sus cualidades y deficiencias están relacionadas con las propiedades de la redes bayesianas que expusimos en las secciones 10.3 y 10.4. A continuación, comentamos brevemente los resultados obtenidos.

15.3.1 Logros

Uno de los frutos del trabajo que hemos llevado a cabo es el desarrollo de una serie de programas que podrán ser utilizados en el futuro como herramienta para la construcción de nuevos sistemas expertos. Las principales cualidades de DIAVAL —y de los demás sistemas que podamos crear en el futuro a partir del *software*

[1] Según nuestras estimaciones, sería necesario contar al menos con un procesador 486 de 50 ó 66 Mhz. y con 16 Mb. de memoria.

desarrollado— son *(1)* que permite un *tratamiento normativo de la incertidumbre*, es decir, un tratamiento ajustado a los axiomas de la teoría de la probabilidad, *(2)* se basa en un *modelo causal* y *(3)* es capaz de realizar tres tipos de *razonamiento*: *abductivo*, *deductivo* e *intercausal* (ver la sec. 10.3 para una descripción más detallada).

Otra de sus cualidades es la *integración de inferencia, explicación y aprendizaje*. Aunque a veces se denomina sistema experto a lo que es simplemente una red bayesiana, a nuestro juicio esto es un abuso del término, pues una de las características esenciales de los sistemas expertos es la capacidad de explicación. En la literatura sobre redes bayesianas no se conoce ningún sistema experto que posea a la vez las capacidades de explicación y aprendizaje; tampoco en la literatura sobre sistemas basados en reglas hemos encontrado ejemplos de programas que incluyan estos dos aspectos.

Por último, el *interfaz* ha sido cuidado con gran esmero, de modo que resulte lo más cómodo, sencillo y agradable para el usuario. La evaluación preliminar nos ha confirmado que esta característica contribuye decisivamente a la aceptación del sistema experto por parte de los médicos.

15.3.2 Limitaciones

Una de las desventajas más importante de las redes bayesianas frente a los sistemas basados en reglas es que tampoco existe un control del razonamiento, sino que la inferencia actúa sobre la red como un todo. Como consecuencia, las RR.BB. son incapaces de realizar un *diagnóstico diferencial* explícito, aunque implícitamente la puerta OR resuelve este problema mediante el cálculo de probabilidad.

Otra desventaja, consecuencia de la anterior, es que las RR.BB. son *incapaces de generar preguntas* que orienten al usuario en la búsqueda de información relevante. En el caso de la ecocardiografía esto no supone un grave problema, pues cuando el médico se sienta a redactar el informe no suele recabar información adicional, y el orden en que se introducen los datos puede sistematizarse mediante una serie de pantallas, tal como hemos explicado en el capítulo anterior.

Quizá el mayor inconveniente que presentan hoy en día las RR.BB. es la *complejidad de los algoritmos* disponibles. En nuestro sistema experto, el diagnóstico se demora más de un minuto, y si ampliamos el modelo de modo que incluya más nodos y bucles, el tiempo de computación puede aumentar considerablemente. Por tanto, se hace necesario desarrollar algoritmos aproximados, según hemos comentado en varias ocasiones.

La capacidad de *explicación* que ofrece nuestro sistema es otro de los aspectos que deben ser mejorados. Las explicaciones que ofrece actualmente el sistema son útiles para explorar el modelo, pero resultan insuficientes para comprender los resultados de la inferencia. Aunque la capacidad de explicación es una de las cualidades de DIAVAL, todavía está lejos de ser completamente satisfactoria.

Por último, existe el problema de que las redes bayesianas son incapaces de considerar la evolución de las enfermedades; al menos, hoy en día el razonamiento temporal en RR.BB. resulta engorroso y poco eficiente. Ésta es una de las cuestiones que debemos resolver antes de poder aplicar nuestro modelo al *seguimiento* de los pacientes, una de los aspectos fundamentales de la ecocardiografía.

15.4 Futuro desarrollo

El trabajo desarrollado en la presente tesis doctoral deja abiertas varias líneas de investigación, que hemos dividido en dos grupos: las que se refieren a la parte teórica (redes bayesianas) y las relativas a las aplicaciones (sistemas expertos).

15.4.1 Razonamiento bayesiano

La sección 10.4 hablaba de las deficiencias de las RR.BB. Cada una de ellas supone una vía abierta para futuras investigaciones: desarrollo de algoritmos más eficientes, obtención automática o semiautomática del conocimiento mediante métodos de aprendizaje, razonamiento temporal, aplicación a problemas diferentes del diagnóstico, control del razonamiento (incluyendo la generación automática de preguntas), explicación, etc. Sin embargo, en este capítulo nos vamos a referir únicamente a las extensiones que surgen como consecuencia de las aportaciones realizadas en la presente tesis doctoral.

Teoría

Las redes bayesianas vienen dadas por un conjunto de axiomas, que fueron introducidos en la sección 4.2. Se trata de una definición matemática completa y cerrada. El progreso, en nuestra opinión, puede darse en dos líneas opuestas: por la extensión o por la simplificación. En cuanto al segundo aspecto, existen las *redes de semejanza* y las *redes cualitativas*, mencionadas en la sección 10.1.3.

En cuanto a la extensión de las redes bayesianas, nos interesa especialmente su relación con las teorías de Lofti Zadeh. Nuestra postura personal es de aprecio hacia los conjuntos difusos, para los cuales es posible dar una interpretación probabilística (sec. 2.3.3), de modo que parece posible una integración de éstos en el marco de las redes bayesianas.

Sin embargo, nuestra actitud frente a la lógica difusa es bastante crítica, por las razones expuestas en la citada sección 2.3.3. El problema fundamental que plantea este formalismo es la ausecia de una interpretación objetiva de los factores numéricos que intervienen. La lógica difusa puede considerarse como una *aproximación útil* en muchos casos, pero no como *modelo normativo*. Por tanto, sería deseable partir de la teoría probabilística y tratar de llegar los mismos resultados que en lógica difusa, pero de forma crítica, es decir, intentando descubrir cuáles de ellos son válidos y cuáles no, en qué casos, cuáles son las hipótesis y aproximaciones necesarias, etc. La labor no es sencilla, y tampoco es fácil predecir las posibilidades de éxito, pero puede merecer la pena intentarlo.

Algoritmos

El condicionamiento local admite muchas **variantes** (sec. 6.2), y por eso sería interesante desarrollar una **investigación empírica** con el fin comprobar cuál de ellas resulta más eficiente o, planteando mejor la cuestión, en qué casos conviene aplicar cada una de las diferentes versiones del algoritmo.

Por otro lado, conviene recordar que el cálculo de la probabilidad en redes bayesianas es un problema NP (ver el apéndice del cap. 6), y por tanto es prácticamente

seguro que ningún algoritmo exacto será eficiente cuando aumente el tamaño de la red. Sin embargo, en el futuro sería deseable contar con redes bayesianas que incluyeran gran cantidad de nodos y bucles. Ni siquiera parece descabellado imaginar un sistema experto bayesiano que abarcara la mayor parte de las especialidades de la medicina. Por ello se hace necesario contar cuanto antes con **métodos aproximados** que permitan el cálculo de probabilidad en grandes redes.

El algoritmo de condicionamiento local sugiere varias aproximaciones que simplificarían considerablemente la computación. Una de ellas consiste en observar que ciertos enlaces transmiten más información en un sentido que en el contrario. Por ejemplo, para una válvula aórtica bicúspide (es una malformación congénita) la probabilidad de contraer estenosis es mucho mayor que para una válvula normal. Sin embargo, el hecho de tener estenosis apenas aumenta la probabilidad de que la válvula sea bicúspide, pues se trata de un enlace con una especificidad muy baja. El propagar la evidencia en un sólo sentido de estos **enlaces débiles** es una aproximación que apenas modifica el resultado final, y sin embargo puede simplificar considerablemente el tiempo de computación al reducir el número de bucles. Nuestros sondeos preliminares así lo sugieren, pero éste es un tema que requiere un estudio mucho más detallado. Hay otras aproximaciones similares que pueden significar una aportación importante a la teoría de algoritmos, y que aún están por investigar.

Otra de las extensiones posibles, que afecta tanto a la teoría como a los algoritmos, consiste en desarrollar **modelos simplificados** de interacción, al estilo de la puerta OR. Por ejemplo, entre los factores que influyen en la probabilidad de un infarto de miocardio (hipertensión, tabaquismo, colesterol, etc.) podría buscarse un modelo de interacción que requiese menos parámetros que el modelo general, con el fin de facilitar la adquisición del conocimiento, ahorrar espacio de almacenamiento y se aumentar la velocidad de computación.

Explicación

El modelo de explicación presentado en la sección 8.3 constituye el tema más abierto de la parte II, dedicada a las redes bayesianas. Las posibilidades de extensión son diversas, desde las más sencillas a las más complejas.

Una de ellas consiste en la **explicación del cambio**, es decir, en fijar varios instantes dentro de la consulta con el sistema experto, correspondientes a distintos subconjuntos de evidencia introducida. Para cada uno de esos instantes habría que realizar la propagación de evidencia y almacenar los resultados, con el fin de mostrar la evolución de la probabilidad para los nodos o los enlaces que nos interesen.

Otra posible extensión, muy relacionada con la anterior, es el **razonamiento hipotético**. Consiste en responder a preguntas del tipo "¿Cuál sería el diagnóstico si hubiéramos observado la presencia de X o si no hubiéramos observado Y?" Su realización sería muy sencilla: bastaría guardar los resultados obtenidos para compararlos con los del razonamiento hipotético.

La **presentación gráfica de los resultados** es una de las posibles ampliaciones del interfaz, que resultaría especialmente interesante para los dos temas propuestos: la explicación del cambio y el razonamiento hipotético. En el eje de coordenadas representaríamos el tiempo, o más bien los distintos subconjuntos de evidencia disponible, y en el eje de abcisas la probabilidad para una o varias variables. Otra gráfica posible

consistiría en mostrar en un diagrama de barras la probabilidad de cada uno de los diagnósticos, para ver cuál de ellos domina sobre los demás.

Una extensión más ambiciosa, aunque realizable, consiste en intergrar el razonamiento bayesiano en un entorno de **hipertexto y multimedia**, con lo cual se podrían desarrollar sistemas expertos mucho más flexibles y potentes, sobre todo desde el punto de vista de la explicación. El entorno multimedia permitiría ofrecer no sólo imágenes estáticas sino también vídeo y sonido. Por otro lado, las posibilidades de navegar por la red bayesiana mediante un menú de opciones, que describíamos en la sección 8.3, resultan rudimentarias comparadas con las facilidades que ofrecen ya hoy los entornos de hipertexto.[2] Nuestro objetivo consiste en diseñar una aplicación de este tipo en que la información que se muestra al usuario no esté almacenada previamente en su totalidad, como suele ocurrir en los sistemas de hipertexto actuales, sino que pueda generarse dinámicamente a lo largo de la consulta. La integración de los sistemas expertos en entornos de hipertexto es una vía de investigación que apenas ha empezado a surgir, y nos gustaría poder aportar en el futuro alguna idea a este campo.

Implementación

Para la primera implementación de las ideas desarrolladas en esta tesis se ha utilizado una herramienta comercial, *GoldWorks*, que funciona sobre *Common Lisp*. Esto ha significado un gran ahorro de tiempo, especialmente a la hora de desarrollar el interfaz gráfico y el interfaz con la base de datos. Sin embargo, esto plantea dos inconvenientes. Por un lado, que el sistema resultante es excesivamente lento y ocupa mucho espacio en el ordenador, tanto en disco duro como en memoria dinámica (RAM). El segundo inconveniente es que el sistema experto requiere el uso de dicha herramienta comercial, por lo que no puede funcionar en un servicio de ecocardiografía sin que éste compre la correspondiente licencia, lo cual dificulta considerablemente la posible implantación del programa en distintos hospitales.

Por tanto, sería deseable realizar una **implementación directa en Common Lisp**, lo cual no requeriría gran esfuerzo, pues podríamos aprovechar gran parte del código fuente desarrollado, o una **implementación en lenguaje C**, que sería mucho más eficiente y más fácil de integrar con los programas de hipertexto mencionados; esta segunda opción supondría un esfuerzo mucho mayor, porque habría traducir todo el programa desde el principio, y además es mucho más duro trabajar en C que en Lisp. Los argumentos a favor y en contra de ambas opciones son muy fuertes, y aún no hemos tomado una decisión sobre este punto.

15.4.2 Sistemas expertos

Cardiología

Naturalmente, la primera línea de trabajo ha quedado abierta en esta tesis consiste en completar la implantación del programa DIAVAL. De la evaluación preliminar se deduce la necesidad de realizar un *ajuste de las probabilidades* contenidas en la red,

[2] Es interesante señalar que el modelo de explicación desarrollado en esta tesis se ha inspirado en el trabajo de Johanna Moore y William Swartout [153] sobre las ventajas que un interfaz basado en hipertexto puede ofrecer de cara a la explicación.

a ser posible mediante una búsqueda dentro de la literatura médica, con el fin de sustituir en lo posible las *estimaciones subjetivas* por *medidas objetivas*.

Una vez resuelto este punto, deberíamos llevar a cabo una *evaluación objetiva*, más rigurosa y exhaustiva que la realizada anteriormente, tal como proponíamos en la sección 14.3.

En la versión actual del sistema experto aún no se encuentra operativa la capacidad de *aprendizaje*, debido a que todavía no tenemos ni la precisión inicial requerida para las probabilidades ni las desviaciones asociadas a las distribuciones (cap. 7). Por este motivo, una de las ventajas de contar con medidas objetivas es que el sistema experto podría aprovechar la técnica de aprendizaje que hemos desarrollado.

Otra mejora posible consiste en *refinar* el sistema experto, matizando mucho más toda la información que contiene. Por ejemplo, al dolor torácico le hemos asignado cuatro valores posibles: "Ausente", "Sugerente de angina", "Sugerente de pericarditis" e "Inespecífico"; sería conveniente considerar además el grado de intensidad, la localización, la duración, la forma de aparición, etc., y así para los demás síntomas y observaciones cualitativas. En la misma línea, sería deseable ampliar la red para que incluyera enfermedades y síntomas poco frecuentes, que en la primera versión hemos preferido no considerar.

Tampoco sería difícil *extender* el modelo causal, para que pueda abarcar otras subespecialidades de la cardiología. De hecho, sólo una pequeña parte del conocimiento que posee nuestra red es específico de la ecocardiografía; la mayor se refiere a anomalías y estados patofisiológicos, de modo que bastaría añadir los nodos correspondientes a la hemodinámica o a la resonancia magnética para tener un sistema experto aplicable también a estas técnicas. Igualmente, podríamos ampliar el programa con el fin de implantarlo en un ambulatorio como consejero de médicos no especializados en cardiología.

No olvidemos que nuestro que nuestro objetivo inicial consistía en construir un sistema capaz de realizar el análisis de imágenes ecocardiográficas (sec. 1.1). El proyecto que está desarrollando el Prof. José Luis Fernández Vindel en el Dpto. de Informática y Automática de la UNED se encuentra en fase muy avanzada, por lo que en un futuro próximo podremos plantear la integración del *sistema de visión* con el sistema experto, tal como estaba previsto inicialmente.

Aplicación a otros dominios

Las líneas de trabajo que quedan abiertas no se limitan a la cardiología. Hay numerosos campos de la medicina en que las redes bayesianas son fácilmente aplicables. La ventaja principal es que en la realización de un segundo sistema experto bayesiano podríamos aprovechar casi todo el trabajo invertido en el primero. Por un lado, ya estamos familiarizados con la herramienta (*GoldWorks*) y hemos implementado todos los elementos necesarios: razonamiento bayesiano, módulo de explicación, marcos y funciones para la creación de un interfaz gráfico y de un interfaz con la base de datos, etc. Por otro, hemos adquirido ya algunos conocimiento básicos sobre medicina y tenemos la experiencia de haber realizado un sistema experto en este campo (ver las "lecciones aprendidas" que hemos comentado en este capítulo).

El interés principal de realizar un segundo proyecto consistiría, a nuestro juicio, en descubrir cuánto tiempo se tarda en construir un sistema experto bayesiano una

vez superados todos los obstáculos que hemos encontrado en la primera ocasión, y en conocer hasta qué punto es generalizable la metodología utilizada. Queda así una vía abierta para la realización de varias aplicaciones más en otros campos, pensando incluso en la posibilidad de introducir las redes bayesianas en problemas ajenos a la medicina.

15.5 Resumen de conclusiones

Como síntesis de los temas discutidos en esta memoria, ofrecemos a continuación un resumen muy breve de las conclusiones obtenidas.

- Las redes bayesianas, a diferencia de los sistemas basados en reglas, son capaces de tratar adecuadamente la incertidumbre y de obtener todas las inferencias posibles a partir de un modelo causal (caps. 2 y 10).
- De los errores sistemáticos que comete el ser humano en la estimación de la incertidumbre se deduce que la inteligencia artificial puede suponer una ayuda en campos como la medicina, y que el razonamiento artificial no tiene por qué ser una duplicación del razonamiento humano (cap. 3).
- Los mayores inconvenientes de las redes bayesianas son la complejidad temporal de los algoritmos que utilizan y la incapacidad de razonar en función de un objetivo (cap. 10).
- La puerta OR —binaria o graduada— simplifica la obtención del conocimiento, acelera la computación de la probabilidad y ofrece posibilidades de explicación que no se dan en el caso general (sec. 5.2).
- El condicionamiento local es mucho más eficiente que los algoritmos de condicionamiento anteriores y supera en algunos aspectos a los algoritmos más utilizados hoy en día (sec. 6.2).
- En redes bayesianas es posible tener implementaciones totalmente distribuidas, con un procesador por cada nodo, tanto para la inferencia como para la explicación y el aprendizaje (caps. 4 a 7).
- El sistema experto DIAVAL utiliza las redes bayesianas para realizar el diagnóstico ecocardiográfico. Sus principales cualidades son el razonamiento normativo y el interfaz gráfico avanzado. Su limitación más importante es la lentitud de la versión actual (caps. 13 y 14).

Bibliografía

[1] J. B. Adams. Probabilistic reasoning and certainty factors. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 12, págs. 263–271. Addison-Wesley, Reading, MA, 1984.

[2] J. S. Aikins. Prototypical knowledge for expert systems. *Artificial Intelligence*, **20**:163–210, 1983.

[3] J. S. Aikins. A repr¡esentation scheme using both frames and rules. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 23, págs. 424–452. Addison-Wesley, Reading, MA, 1984.

[4] S. K. Andersen, K. G. Olesen, F. V. Jensen y F. Jensen. HUGIN: A shell for building Bayesian belief universes for expert systems. En: *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, págs. 1080–1085, Detroit, MI, 1989.

[5] S. Andreasen, F. V. Jensen y K. G. Olesen. Medical expert sustems based on causal probabilistic networks. *International Journal of Biomedical Computing*, **28**:1–30, 1991.

[6] S. Andreassen, R. Hovorka, J. Benn, K. G. Olesen y E. R. Carson. A model-based approach to insulin adjustment. En: *Proceedings of the Third Conference on Artificial Intelligence in Medicine*, págs. 239–248, Maastrich, The Netherlands, 1991. Springer-Verlag.

[7] S. Andreassen, M. Woldby, B. Falck y S. K. Andersen. MUNIN — A causal probabilistic network for interpretation of electromyographic findings. En: *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI–87)*, págs. 366–372, Milan, 1987.

[8] M. Baker y T. E. Boult. Pruning Bayesian networks for efficient computation. En: P. P. Bonissone, M. Henrion, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 6*, págs. 225–232. Elsevier Science Publishers, Amsterdam, 1991.

[9] A. Barreiro. *TAO/Medtool: Un sistema consejero en oncología.* Tesis doctoral, Universidad de Santiago, Santiago de Compostela, 1991.

[10] A. Barreiro, R. P. Otero, R. Marín y J. Mira. A modular knowledge base for the follow-up of clinical protocols. *Methods of Information in Medicine*, **32**:373–381, 1993.

[11] S. Barro. *SUTIL: Sistema cuasi-integral de monitorización inteligente en U.C.C.* Tesis doctoral, Universidad de Santiago, Santiago de Compostela, 1988.

[12] S. W. Barth y S. W. Norton. Knowledge engineering within a generalized bayesian framework. En: J. F. Lemmer y L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence 2*, págs. 103–114. Elsevier Science Publishers, Amsterdam, 1988.

[13] R. Bellazzi, S. Quaglini, C. Berzuini y M. Stefanelli. GAMEES: A probabilistic environment for expert systems. *Computer Methods and Programs in Biomedicine*, **35**:177–191, 1991.

[14] M. Ben-Bassat, R. W. Carlson, V. K. Puri, M. D. Davenport, J. A. Schriver, M. Latif, R. Smith, L. D. Portigal, E. H. Lipnick y M. H. Weil. Pattern-based interactive diagnosis of multiple disorders: The MEDAS system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2**:148–160, 1980.

[15] J. S. Bennet y R. S. Englemore. Experience using EMYCIN. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 16, págs. 314–328. Addison-Wesley, Reading, MA, 1984.

[16] D. C. Berry y A. E. Hart. Evaluating expert systems. *Expert Systems*, **7**:199–208, 1990.

[17] D. M. Berwick, H. V. Fineberg y M. C. Weinstein. When doctors meet numbers. *American Journal of Medicine*, **71**:991–998, 1981.

[18] C. Berzuini. Representing time in causal probabilistic networks. En: M. Henrion, R. D. Shachter, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 5*, págs. 15–28. Elsevier Science Publishers, Amsterdam, 1990.

[19] L. Birnbaum. Rigor mortis: A response to Nilson's "Logic and artificial inteligence". *Artificial Inteligence*, **47**:57–177, 1991.

[20] R. J. Brachman y H. J. Levesque (eds.). *Readings in Knowledge Representation.* Morgan Kaufmann, San Mateo (CA), 1985.

[21] E. Braunwald. *Heart Desease.* Saunders & Co., Philadelphia, 1980.

[22] J. R. Brown y S. Cunningham. *Programming the User Interface: Principles and Examples.* John Wiley & Sons, New York, 1989.

[23] G. D. Bryant y G. R. Norman. Expressions of probability: Words and numbers. *New England Journal of Medicine*, **302**:411, 1980.

[24] B. G. Buchanan, D. Barstow, R. Bechtal, J. Bennett, W. Clancey, C. Kulikowski, T. Mitchell y D. A. Waterman. Constructing an expert system. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (eds.), *Building Expert Systems*, cap. 5, págs. 127–167. Addison-Wesley, Reading, MA, 1983.

[25] B. G. Buchanan y E. A. Feigenbaum. DENDRAL and Meta-DENDRAL: Their applications dimension. *Artificial Intelligence*, **11**:5–24, 1978.

[26] B. G. Buchanan y E. H. Shortliffe (eds.). *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

[27] B. G. Buchanan, G. Sutherland y E. A. Feigenbaum. Heuristic DENDRAL: A program for generating explanatory hypotheses in organic chemistry. En: B. Meltzer y D. Michie (eds.), *Machine Intelligence 4*. Edinburgh University Press, Edinburgh, 1969.

[28] D. V. Budescu, S. Weinberg y T. S. Wallsten. Decisions based on numerically and verbally expressed uncertainties. *Organizational Behavior and Human Decision Processes*, **36**:391–405, 1988.

[29] C. Cannings. Probability functions on complex pedigrees. *Advances in Applied Probability*, **10**:26–61, 1978.

[30] C. Cannings, E. A. Thompson y M. H. Skolnick. Recursive derivation of likelihoods on pedigrees of arbitrary complexity. *Advances in Applied Probability*, **8**:622–625, 1976.

[31] J. R. Carbonell. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, **11**:190–202, 1970.

[32] E. Castillo, J. M. Gutiérrez y A. S. Hadi. Modelling probabilistic networks based on conditional/marginal probabilities. Enviado a *Artificial Intelligence*, Enero 1994.

[33] J. L. Chamorro Moreno. Principios físicos de la ultrasonografía. *Bioingeniería y Clínica*, **1**:17–36, 1985.

[34] B. Chandrasekaran. On evaluating AI systems for medical diagnosis. *AI Magazine*, **4**:34–37, 1988.

[35] B. Chandrasekaran, M. C. Tanner y J. R. Josephson. Explaining control strategies in problem solving. *IEEE Expert*, **4**:9–24, 1989.

[36] E. Charniak. Bayesian Networks without tears. *AI Magazine*, **12**:50–63, 1991.

[37] E. Charniak y D. McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, MA, 1985.

[38] P. Cheeseman. In defense of probability. En: *Proceedings of the 9th International Joint Conference on Artificial Intelligence (IJCAI–85)*, págs. 1002–1009, Los Angeles, CA, 1985.

[39] P. Cheeseman. Probabilistic versus fuzzy reasoning. En: L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, págs. 85–102. Elsevier Science Publishers, Amsterdam, 1986.

[40] P. Cheeseman. An inquiry into computer understanding. *Computational Intelligence*, **4**:58–66, 1988.

[41] J. J. J. Christensen-Szalanski y L. R. Beach. Experience and the base-rate fallacy. *Organizational Behavior and Human Performance*, **29**:270–278, 1982.

[42] J. J. J. Christensen-Szalanski y J. B. Bushyhead. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance*, **7**:928–935, 1981.

[43] W. J. Clancey. Dialogue management for rule-based tutorials. En: *Proceedings of the 6th International Joint Conference on Artificial Intelligence (IJCAI–79)*, págs. 155–161, Tokyo, 1979.

[44] W. J. Clancey. The epistemology of rule-based expert systems — A framework for explanation. *Artificial Intelligence*, **20**:215–251, 1983.

[45] W. J. Clancey. Use of MYCIN's rules for tutoring. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 26, págs. 464–489. Addison-Wesley, Reading, MA, 1984.

[46] W. J. Clancey. Notes on "Heuristic classification". *Artificial Intelligence*, **59**:191–196, 1993.

[47] J. R. Clarke. *Surgical Judgement Using Decision Analysis*. Prager, New York, 1984.

[48] P. Cohen y M. Grinberg. A framework for heuristic reasoning about uncertainty. En: *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83)*, págs. 355–357, Karlsruhe, Germany, 1983.

[49] P. Cohen y M. Grinberg. A theory of heuristic reasoning about uncertainty. *AI Magazine*, págs. 17–23, 1983.

[50] G. F. Cooper. *NESTOR: A Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. Tesis doctoral, Dept. of Computer Science, Stanford University, STAN–CS–84–1031, 1984.

[51] G. F. Cooper. A dignostic method that uses causal knowledge and linear programming in the application of Bayes' formula. *Computer Methods and Programms in Biomedicine*, **22**:223–237, 1986.

[52] G. F. Cooper. A method for using belief networks as influence diagrams. En: *Proceedings of the 4th Workshop on Uncertainty in AI*, págs. 55–63, University of Minnesota, Minneapolis, 1988.

[53] G. F. Cooper. Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis*, **5**:39–52, 1989.

[54] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, **42**:393–405, 1990.

[55] G. F. Cooper y E. Herskovits. A Bayesian method for constructing Bayesian belief networks from databases. En: *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, págs. 86–94, Los Angeles, CA, 1991. Morgan Kaufmann, San Mateo, CA.

[56] G. F. Cooper y E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, **9**:309–348, 1992.

[57] P. Dagum y A. Galper. Forecasting sleep apnea with dynamic network models. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 64–71, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[58] P. Dagum, A. Galper y E. Horvitz. Dynamic network models for forecasting. En: *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, págs. 41–48, Stanford University, 1992. Morgan Kaufmann, San Mateo, CA.

[59] P. Dagum y M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, **60**:141–153, 1993.

[60] R. Davis. Meta-rules: Reasoning about control. *Artificial Intelligence*, **15**:179–222, 1980.

[61] R. Davis y B. Buchanan. Meta-level knowledge. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 28, págs. 507–530. Addison-Wesley, Reading, MA, 1984.

[62] R. Davis, B. G. Buchanan y E. H. Shortliffe. Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence*, **8**:15–45, 1977.

[63] R. Davis, B. G. Buchanan y E. H. Shortliffe. Retrospective on "Production rules as a representation for a knowledge-based consultation program". *Artificial Intelligence*, **59**:181–189, 1993.

[64] R. M. Dawes. *Rational Choice in an Uncertain World.* Harcourt Brace Jovanovich, Orlando, FL, 1988.

[65] F. T. de Dombal, J. R. Leaper, J. R. Staniland, A. McCann y J. Horrocks. Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, **2**:9–13, 1972.

[66] J. de Kleer y B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, **32**:97–130, 1987.

[67] T. Dean y K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, **5**:142–150, 1989.

[68] A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, **30**:205–247, 1968.

[69] D. E. Detmer, D. G. Fryback y K. Gassner. Heuristics and biases in medical decision-making. *Journal of Medical Education*, **53**:682–683, 1978.

[70] *Diccionario Terminológico de Ciencias Médicas*. Salvat Editores, S. A., Barcelona, undécima edición, 1974.

[71] F. J. Díez. Local conditioning in Bayesian networks. Technical Report (R–181), Cognitive Systems Laboratory, University of California, Los Angeles, 1992. Enviado a *Artificial Intelligence*.

[72] F. J. Díez. Parameter adjustement in Bayes networks. The generalized noisy OR–gate. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 99–105, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[73] F. J. Díez y J. Mira. Distributed reasoning and learning in Bayesian expert systems. En: C. A. Ntuen (ed.), *Advances in Fault-Diagnosis Problem Solving*. CRC Press, Boca Raton, FL, 1993. En prensa.

[74] F. J. Díez y J. Mira. Distributed inference in Bayesian networks. *Cybernetics and Systems*, **25**:39–61, 1994.

[75] F. J. Díez Vegas y J. Mira Mira. Causal Bayesian reasoning in medicine. *Cybernetics and Systems*, **23**:417–429, 1992.

[76] J. Doyle. A truth maintenance system. *Artificial Intelligence*, **12**:231–272, 1979.

[77] M. J. Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense*. Tesis doctoral, Dept. of Engineering and Public Policy, Carnigie Mellon University, 1993.

[78] M. J. Druzdzel y M. Henrion. Using scenarios to explain probabilistic inference. En: *Proceedings of the AAAI-90 Workshop on Explanation*, págs. 133–141, Boston, MA, 1990.

[79] M. J. Druzdzel y M. Henrion. Efficient reasoning in qualitative probabilistic networks. En: *Proceedings of the 11th National Conference on AI (AAAI-93)*, Washington D.C., 1993.

[80] M. J. Druzdzel y H. A. Simon. Causality in Bayesian belief networks. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 3–11, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[81] R. O. Duda, J. Gaschnig y P. E. Hart. Model design in the PROSPECTOR consultant program for mineral exploration. En: D. Michie (ed.), *Expert Systems in the Microelectronic Age*, págs. 153–167. Edinburgh University Press, Edinburgh, 1979.

[82] R. O. Duda, P. E. Hart y N. Nilsson. Subjective Bayesian methods for rule-based inference systems. *Proceedings of the 1976 National Computer Conference (AFIPS)*, **45**:1075–1082, 1976.

[83] D. M. Eddy. Probabilistic reasoning in clinical medicine: Problems and opportunities. En: D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgement under uncertainty: Heuristics and biases*, cap. 18, págs. 249–267. Cambridge University Press, Cambridge, UK, 1982.

[84] C. Elsaesser. Explanation of probabilistic inference. En: L.N. Kanal, T.S. Levitt y J.F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 3*, págs. 387–400. Elsevier Science Publishers, Amsterdam, 1989.

[85] C. Elsaesser y M. Henrion. Verbal expressions for probability updates: How much more probable is "much more probable"? En: M. Henrion, R. D. Shachter, L.N. Kanal y J.F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 5*, págs. 319–328. Elsevier Science Publishers, Amsterdam, 1990.

[86] L. M. Fagan, J. C. Kunz, E. A Feigenbaum y J. J. Osborn. Extensions to the rule-based formalism for a monitoring task. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 22, págs. 397–423. Addison-Wesley, Reading, MA, 1984.

[87] E. A. Feigenbaum. The art of artificial intelligence: Themes and case studies of knowledge engineering. En: *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI–77)*, págs. 1014–1029, Cambridge, MA, 1977.

[88] G. Fernández. Panorama de los sistemas expertos. En: J. Cuena (ed.), *Inteligencia Artificial: Sistemas Expertos*, págs. 23–52. Alianza Editorial, Madrid, 1986.

[89] I. Ferreira Montero (ed.). *Cardiología*. Ediciones CEA, Madrid, 1987.

[90] B. Fishhoff. Debiasing. En: D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgement under uncertainty: Heuristics and biases*, cap. 31, págs. 249–267. Cambridge University Press, Cambridge, UK, 1982.

[91] Franz Inc. *Common Lisp. The Reference*. Addison-Wesley, Reading, MA, 1988.

[92] J. Gaschning, P. Klahr, H. Pople, E. Shortliffe y A. Terry. Evaluation of expert systems: Issues and case studies. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (eds.), *Building Expert Systems*, cap. 8, págs. 241–280. Addison-Wesley, Reading, MA, 1983.

[93] M. R. Genesereth. The use of design descriptions in automated diagnosis. *Artificial Intelligence*, **24**:411–436, 1984.

[94] R. Giles. Semantics for fuzzy reasoning. *International Journal of Man-Machine Studies*, **17**:401–415, 1982.

[95] G. A. Gorry. Computer-assisted clinical decision making. *Methods of Information in Medicine*, **12**:45–51, 1973.

[96] G. A. Gorry y G. O. Barnett. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, **1**:490–507, 1968.

[97] D. W. Hasling, W. J. Clancey y G. Rennels. Strategic explanations for a diagnostic consultation system. *International Journal of Man-Machine Studies*, **20**:3–19, 1984.

[98] F. Hayes-Roth, D. A. Waterman y D. B. Lenat (eds.). *Building Expert Systems*. Addison-Wesley, Reading, MA, 1983.

[99] D. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. En: L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, págs. 167–196. Elsevier Science Publishers, Amsterdam, 1986.

[100] D. Heckerman. An empirical comparison of three inference methods. En: R. D. Shachter, T.S. Levitt, L.N. Kanal y J.F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 4*, págs. 283–302. Elsevier Science Publishers, Amsterdam, 1990.

[101] D. Heckerman. Causal independence for knowledge acquisition and inference. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 122–127, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[102] D. Heckerman y A. Mamdani (eds.). *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, Washington, D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[103] D. E. Heckerman. *Probabilitic Similarity Networks*. Tesis doctoral, Dept. of Computer Science, Stanford University, STAN–CS–90–1316, 1990.

[104] D. E. Heckerman. *Probabilitic Similarity Networks*. MIT Press, Cambridge, MA, 1991.

[105] D. E. Heckerman y E. J. Horvitz. On the expresiveness of rule-based systems for reasoning with uncertainty. En: *Proceedings of the 6th National Conference on AI (AAAI-87)*, págs. 121–126, Seattle, WA, 1987.

[106] D. E. Heckerman y E. J. Horvitz. The myth of modularity in rule-based systems for reasoning with uncertainty. En: J. F. Lemmer y L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence 2*, págs. 23–34. Elsevier Science Publishers, Amsterdam, 1988.

[107] D. E. Heckerman, E. J. Horvitz y B. N. Nathwani. Toward normative expert systems: Part I — The Pathfinder Project. *Methods of Information in Medicine*, **31**:90–105, 1992.

[108] D. E. Heckerman y B. N. Nathwani. Toward normative expert systems: Part II — Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine*, **31**:106–116, 1992.

[109] M. Henrion. Propagation of uncertainty by logic sampling in Bayes' networks. En: J. F. Lemmer y L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence 2*, págs. 149–164. Elsevier Science Publishers, Amsterdam, 1988.

[110] M. Henrion. Some practical issues in constructing belief networks. En: L. N. Kanal, T. S. Levitt y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 3*, págs. 161–173. Elsevier Science Publishers, Amsterdam, 1989.

[111] M. Henrion y D. R. Cooley. An experimental comparison of knowledge engineering for expert systems and for decision analysis. En: *Proceedings of the 6th National Conference on AI (AAAI-87)*, págs. 471–476, Seattle, WA, 1987.

[112] M. Henrion y M. J. Druzdzel. Qualitative propagation and scenario-based schemes for explaining probabilistic reasoning. En: P. P. Bonissone, M. Henrion, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 6*, págs. 17–32. Elsevier Science Publishers, Amsterdam, 1991.

[113] E. Herskovits. *Computer-based probabilistic-network construction.* Tesis doctoral, Dept. of Computer Science, Stanford University, STAN–CS–90–1367, 1990.

[114] R. R. Hoffman. A survey of methods for eliciting the knowledge of experts. *SIGART Newsletter*, **108**:19–27, 1989.

[115] J. E. Hopcroft y J. D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, Reading, MA, 1979.

[116] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. En: *Proceedings of the National Academy of Sciences USA*, págs. 2554–2558, 1982.

[117] E. Horvitz y D. Heckerman. The inconsistent use of measures of certainty in Artificial Intelligence research. En: L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, págs. 137–151. Elsevier Science Publishers, Amsterdam, 1986.

[118] R. A. Howard y J. E. Matheson. Influence diagrams. En: R. A. Howard y J. E. Matheson (eds.), *Readings on the Principles and Applications of Decision Analysis*, págs. 719–762. Strategic Decisions Group, Menlo Park, CA, 1984.

[119] R. A. Howard, J. E. Matheson y K. L. Miller. *Readings in Decision Analisys.* Stanford Research Institute, Menlo Park, CA, 1976.

[120] T. Hrycej. Common features of neural-network models of high and low level human information processing. En: T. Kohonen, K. Mäkisara, O. Simula y J. Kangas (eds.), *Artificial Neural Networks*, págs. 861–866. Elsevier Science Publishers, Amsterdam, 1991.

[121] F. V. Jensen, B. Chamberlain, T. Nordahl y F. Jensen. Analysis in HUGIN of data conflict. En: P. P. Bonissone, M. Henrion, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 6*, págs. 519–528. Elsevier Science Publishers, Amsterdam, 1991.

[122] F. V. Jensen, K. G. Olesen y S. K. Andersen. An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, **20**:637–660, 1990.

[123] D. Kahneman, P. Slovic y A. Tversky (eds.). *Judgement under uncertainty: Heuristics and biases.* Cambridge University Press, Cambridge, UK, 1982.

[124] J. Kalagnanam y M. Henrion. A comparison of decision analysis and expert rules for sequential diagnosis. En: R. D. Shachter, T.S. Levitt, L.N. Kanal y J.F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 4*, págs. 271–281. Elsevier Science Publishers, Amsterdam, 1990.

[125] J. H. Kim. *CONVINCE: A conversation inference consolidation engine*. Tesis doctoral, Dept. of Computer Science, University of California, Los Angeles, 1983.

[126] J. H. Kim y J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. En: *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83)*, págs. 190–193, Karlsruhe, Germany, 1983.

[127] U. Kjærulff. Graph triangulation — algorithms giving small total state space. Technical Report R–90–09, University of Aalborg, Denmark, 1990.

[128] U. Kjærulff. A computational scheme for reasoning in dynamic probabilistic networks. En: *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, págs. 121–129, Stanford University, 1992. Morgan Kaufmann, San Mateo, CA.

[129] U. Kjærulff. Optimal decomposition of probabilistic networks by simulated annealing. *Statistics and computing*, **2**:7–17, 1992.

[130] A. L. Komaroff. The variability and inaccuracy of medical data. *Proceedings of the IEEE*, **67**:1196–1207, 1979.

[131] A. Kong, G. O. Barnett, F. Mosteller y C. Youtz. How medical professionals evaluate expressions of probability. *New England Journal of Medicine*, **315**:740–744, 1986.

[132] B. Kosko. Fuzziness vs. probability. *Internacional Journal of General Systems*, **17**:211–240, 1990.

[133] C. A. Kulikowski y S. M. Weiss. Representation of expert knowledge for consultation: The CASNET and EXPERT projects. En: P. Szolovits (ed.), *Artificial Intelligence in Medicine*, págs. 21–55. Westview Press, Boulder, CO, 1982.

[134] K. B. Laskey. Adapting connectionist learning to Bayes networks. *International Journal of Approximate Reasoning*, **4**:261–282, 1990.

[135] S. L. Lauritzen y D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, **50**:157–224, 1988.

[136] R. S. Ledley y L. B. Lusted. Reasoning foundations of medical diagnosis. *Science*, **130**:9–21, 1959.

[137] D. V. Lindley. The probability approach to the treatment of uncertainty in artificial intelligence. *Statistical Science*, **2**:17–24, 1987.

[138] R. Marín. *Un sistema experto para el diagnóstico y tratamiento anteparto del estado materno-fetal.* Tesis doctoral, Universidad de Santiago, Santiago de Compostela, 1987.

[139] R. Marín, M. Taboada, J. Mira, A. Barreiro y R. P. Otero. Design and integration of a graphic interface for an expert system in oncology. *International Journal of Biomedical Computing*, **33**:25–43, 1993.

[140] D. Marr. *Vision.* Freeman, New York, 1982.

[141] J. McCarthy. Recursive functions of symbolic expressions and their computation by machine. *Communications of the ACM*, **3**:184–195, 1960.

[142] J. McCarthy. History of LISP. *SIGPLAN Notices*, **13**:217–223, 1978.

[143] W. S. McCulloch y W. H. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, **5**:115–133, 1943.

[144] J. McDermott. R1: A rule-base configurer of computer systems. *Artificial Intelligence*, **19**:39–88, 1982.

[145] D. Michie. Turing's Test and conscious thought. *Artificial Intelligence*, **60**:1–22, 1993.

[146] R. A. Miller, H. E. Pople Jr. y J. D. Myers. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, **307**:468–476, 1982.

[147] R. A. Miller y F. E. Masarie. The demise of the "greek oracle" model for medical diagnostic systems. *Methods of Information in Medicine*, **29**:1–2, 1990.

[148] M. Minsky. *Computation: Finite and Infinite Machines.* Prentice-Hall, Englewood Cliffs, NJ, 1967.

[149] M. Minsky. A framework for representing knowledge. En: P. H. Winston (ed.), *The Psychology of Computer Vision*, págs. 211–277. McGraw-Hill, New York, 1975.

[150] M. L. Minsky y S. Papert. *Perceptrons, An Essay on Computational Geometry.* MIT Press, Cambridge, MA, 1969.

[151] J. Mira, A. E. Delgado, M. Santos, A. P. de Madrid y J. R. Alvarez. Local learning in networks of universal analogic neurons. *Cybernetics and Systems*, **25**:259–273, 1994.

[152] M. J. Monaghan. *Practical Echocardiography and Doppler.* John Wiley & Sons, Chichester, U.K., 1990.

[153] J. D. Moore y W. R. Swartout. Pointing: A way toward explanation dialogue. En: *Proceedings of the 8th National Conference on AI (AAAI-90)*, págs. 457–464, Boston, MA, 1990.

[154] M. A. Nakao y S. Axelrod. Numbers are better than words: Verbal specifications of frequency have no place in medicine. *American Journal of Medicine*, **74**:1061–1065, 1983.

[155] R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley-Interscience, New York, 1990.

[156] A. Newell, J. C. Shaw y H. A. Simon. Empirical explorations with the Logic Theory Machine. En: E. A. Feigenbaum y J. Feldman (eds.), *Computers and Thought*, págs. 109–133. McGraw-Hill, New York, 1963.

[157] A. Newell y H. A. Simon. GPS, a program that simulates human thought. En: E. A. Feigenbaum y J. Feldman (eds.), *Computers and Thought*, págs. 279–293. McGraw-Hill, New York, 1963.

[158] K. C. Ng y B. Abramson. Uncertainty management in expert systems. *IEEE Expert*, págs. 29–48, Abril 1990.

[159] A. E. Nicholson y J. M. Brady. Sensor validation unsing dynamic belief networks. En: *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, págs. 207–214, Stanford University, 1992. Morgan Kaufmann, San Mateo, CA.

[160] N. J. Nilsson. Logic and artificial intelligence. *Artificial intelligence*, **47**:31–56, 1991.

[161] S. W. Norton. An explanation mechanism for Bayesian inference systems. En: J. F. Lemmer y L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence 2*, págs. 165–174. Elsevier Science Publishers, Amsterdam, 1988.

[162] R. M. O'Keefe, O. Balci y E. P. Smith. Validating expert system performance. *IEEE Expert*, **2**:81–90, 1987.

[163] K. G. Olesen, S. L. Lauritzen y F. V. Jensen. aHUGIN: A system creating adaptive causal probabilistic networks. En: *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, págs. 223–229, Stanford University, 1992. Morgan Kaufmann, San Mateo, CA.

[164] R. P. Otero. *Medtool: Una herramienta para el desarrollo de sistemas expertos*. Tesis doctoral, Universidad de Santiago, Santiago de Compostela, 1991.

[165] R. P. Otero y J. Mira. Medtool: A teachable medical expert system development tool. En: *Proceedings of the Third International Symposium on Knowledge Engineering*, págs. 191–200, Madrid, 1988.

[166] C. L. Paris. Combining discourse strategies to generate descriptions to users along a naive/expert spectrum. En: *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI–87)*, Milan, 1987.

[167] R. S. Patil, P. Szolovits y P. Schwartz. Causal understanding of patient illness in medical diagnosis. En: *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI–81)*, págs. 893–899, Vancouver, 1981.

[168] S. G. Pauker, G. A. Gorry, J. P. Kassirer y W. B. Schwartz. Towards the simulation of clinical cognition. Taking a present illness by computer. *American Journal of Medicine*, **60**:981–996, 1976.

[169] S. G. Pauker y J. P. Kassirer. Decision analysis. *New England Journal of Medicine*, **316**:250–258, 1987.

[170] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. En: *Proceedings of the 2nd National Conference on AI (AAAI-82)*, págs. 133–136, Pittsburgh, Pennsylvania, 1982.

[171] J. Pearl. A constraint-propagation approach to probabilistic reasoning. En: L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, págs. 357–370. Elsevier Science Publishers, Amsterdam, 1986.

[172] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, **29**:241–288, 1986.

[173] J. Pearl. Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence*, **32**:245–257, 1987.

[174] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988. Reimpreso con correcciones en 1991.

[175] J. Pearl. Belief networks revisited. *Artificial Intelligence*, **59**:49–56, 1993.

[176] J. Pearl. From conditional oughts to qualitative decision theory. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 12–20, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[177] J. Pearl, D. Geiger y T. Verma. Conditional independence and its representations. *Kybernetika*, **25**:33–44, 1989.

[178] J. Pearl y T. S. Verma. A theory of inferred causation. En: J. A. Allen, R. Fikes y E. Sandewall (eds.), *Principles of Knowledge Representation: Proceedings of the Second International Conference*, págs. 441–452, San Mateo, CA, 1991. Morgan Kaufmann.

[179] J. Pearl y T. S. Verma. A statistical semantics for causation. *Statistics and Computing*, **2**:91–95, 1992.

[180] Y. Peng y J. A. Reggia. Plausibility of diagnostic hypotheses. En: *Proceedings of the 5th National Conference on AI (AAAI-86)*, págs. 140–145, Philadelphia, 1986.

[181] M. A. Peot y R. D. Shachter. Fusion and propagation with multiple observations in belief networks. *Artificial Intelligence*, **48**:299–318, 1991.

[182] A. D. Preece. Towards a methodology for evaluating expert systems. *Expert Systems*, **7**:215–293, 1990.

[183] G. M. Provan. Tradeoffs in constructing and evaluating temporal influence diagrams. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 40–47, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[184] M. R. Quillian. Semantic memory. En: M. Minsky (ed.), *Semantic information processing*, págs. 227–270. MIT Press, Cambridge, MA, 1968.

[185] G. Rebane y J. Pearl. The recovery of causal poly-trees from statistical data. En: L. N. Kanal, T. S. Levitt y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 3*, págs. 175–182. Elsevier Science Publishers, Amsterdam, 1989.

[186] J. A. Reggia y B. T. Perricone. Answer justification in medical decision support systems based on Bayesian classification. *Computers in Biology and Medicine*, **15**:161–167, 1985.

[187] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, **13**:81–132, 1980.

[188] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, **32**:57–95, 1987.

[189] G. A. Ringland y D. A. Duce. *Approaches to Knowledge Representation: An Introduction*. John Wiley & Sons, New York, 1986.

[190] A. Rosenblueth, N. Wiener y J. Bigelow. Behavior, purpose and teleology. *Philosophy of Science*, **10**:18–24, 1943.

[191] D. E. Rumelhart. Toward an interactive model of reading. Technical Report #CHIP-56, University of California, La Jolla, CA, 1976.

[192] D. E. Rumelhart y J. L. McClelland (eds.). *Parallel Distributed Processing*. MIT Press, Cambridge, MA, 1986.

[193] J. C. Schlimmer y P. Langley. Machine learning. En: S. C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence*, págs. 785–805. John Wiley & Sons, New York, segunda edición, 1992.

[194] S. M. Schwartz, J. Baron y J. R. Clarke. A causal Bayesian model for the diagnosis of appendicitis. En: J. F. Lemmer y L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence 2*, págs. 423–434. Elsevier Science Publishers, Amsterdam, 1988.

[195] W. B. Schwartz. Decision analysis: A look at the chief complains. *New England Journal of Medicine*, **300**:556–559, 1979.

[196] P. Sember y I. Zukerman. On the generation of explanations for Bayesian belief networks. En: *Proceedings of the 12th Australian Computer Science Conference*, Febrero 1989.

[197] P. Sember y I. Zukerman. Strategies for generating micro explanations for Bayesian belief networks. En: *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, págs. 295–302, Windsor, Ontario, 1989.

[198] R. D. Shachter. Evaluating influence diagrams. *Operations Research*, **34**:871–882, 1986.

[199] R. D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, **36**:589–605, 1988.

[200] R. D. Shachter, S. T. Andersen y P. Szolovits. The equivalence of exact methods for probabilistic inference on belief networks. Enviado a *Artificial Intelligence*, Octubre 1991.

[201] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[202] G. Shafer. Probability judgment in artificial intelligence and expert systems. *Statistical Science*, **2**:3–44, 1987.

[203] G. Shafer y J. Pearl. *Readings in Uncertain Reasoning*. Morgan Kaufmann, San Mateo, CA, 1990.

[204] R. C. Shank. Conceptual Dependency: A theory of natural language understanding. *Cognitive Psychology*, **3**:552–631, 1972.

[205] C. E. Shannon. Automatic chess player. *Scientific American*, **182**, 1950.

[206] S. C. Shapiro (ed.). *Encyclopedia of Artificial Intelligence*. John Wiley & Sons, New York, segunda edición, 1992.

[207] E. H. Sherman. Evaluation of user interfaces. Technical Report KSL–90–55, Knowledge Systems Laboratory, Stanford University, 1990.

[208] B. Shneiderman. *Designing the Human Interface*. Addison-Wesley, New York, 1987.

[209] E. H. Shortliffe y B. G. Buchanan. A model of inexact reasoning in medicine. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 11, págs. 233–262. Addison-Wesley, Reading, MA, 1984.

[210] E. H. Shortliffe, B. G. Buchanan y E. A. Feigenbaum. Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proceedings of the IEEE*, **67**:1207–1224, 1979.

[211] E. H. Shortliffe, A. C. Scott, M. B. Bischoff, A. B. Campbell, W. van Melle y C. D. Jacobs. An expert system for oncology protocol management. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 35, págs. 653–665. Addison-Wesley, Reading, MA, 1984.

[212] M. Sokolow y M. B. McIlroy. *Clinical Cardiology*. Lange Medical Publications, Los Altos, CA, 1986.

[213] J. F. Sowa. *Conceptual Structures: Information Processing in Minds and Machines*. Addison-Wesley, Reading, MA, 1984.

[214] J. F. Sowa. Semantic networks. En: S. C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence*, págs. 1493–1511. John Wiley & Sons, New York, segunda edición, 1992.

[215] D. J. Spiegelhalter, R. C. G. Frankling y K. Bull. Assessment, criticism and improvement of imprecise subjective probabilities. En: M. Henrion, R. D. Shachter, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 5*, págs. 285–294. Elsevier Science Publishers, Amsterdam, 1990.

[216] D. J. Spiegelhalter y R. P. Knill-Jones. Statistical and knowledge-based approaches to clinical decision support systems, with an application to gastroenterology. *Journal of the Royal Statistical Society, Series A*, **147**:35–77, 1984.

[217] D. J. Spiegelhalter y S. L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, **20**:579–605, 1990.

[218] P. Spirtes, C. Glymour y R. Scheines. *Causation, Prediction and Search.* Springer-Verlag, New York, 1993.

[219] S. Srinivas. A generalization of the noisy-Or model. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, págs. 208–215, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[220] S. Srinivas y J. S. Breese. IDEAL: A software package for analysis of influence diagrams. En: *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, págs. 212–219, Cambridge, MA, 1990.

[221] W. Stallings. Fuzzy set theory versus Bayesian statistics. *IEEE Transactions on Systems, Man and Cybernetics*, págs. 216–219, Marzo 1977.

[222] M. Stefik, J. Aikins, R. Balzer, J. Benoit, L. Birnbaum, F. Hayes-Roth y E. Sacerdoti. Basic concepts for building expert systems. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (eds.), *Building Expert Systems*, cap. 3, págs. 59–86. Addison-Wesley, Reading, MA, 1983.

[223] J. Stillman. On heuristics for finding loop cutsets in multiply connected belief networks. En: P. P. Bonissone, M. Henrion, L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 6*, págs. 233–243. Elsevier Science Publishers, Amsterdam, 1991.

[224] T. M. Strat. The generation of explanations within evidential reasoning systems. En: *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI–87)*, págs. 1097–1104, Milan, 1987.

[225] T. M. Strat y J. D. Lowrance. Explaining evidential analyses. *International Journal of Approximate Reasoning*, **3**:299–353, 1989.

[226] L. E. Sucar, D. F. Gillies y D. A. Gillies. Objective probabilities in expert systems. *Artificial Intelligence*, **61**:187–208, 1993.

[227] C. Suen, P. Grogono, R. Shinghal y F. Coallier. Verifying, validating and measuring performance of expert systems. *Expert System Applications*, **2**:93–102, 1990.

[228] H. J. Suermondt. *Explanation in Bayesian Belief Networks.* Tesis doctoral, Dept. of Computer Science, Stanford University, STAN–CS–92–1417, 1992.

[229] H. J. Suermondt y G. F. Cooper. Probabilistic inference in multiply connected belief networks using loop cutsets. *International Journal of Approximate Reasoning*, **4**:283–306, 1990.

[230] H. J. Suermondt y G. F. Cooper. A combination of exact algorithms for inference on Bayesian belief networks. *International Journal of Approximate Reasoning*, **5**:521–542, 1991.

[231] H. J. Suermondt y G. F. Cooper. Initialization for the method of conditioning in Bayesian belief networks. *Artificial Intelligence*, **50**:83–94, 1991.

[232] W. R. Swartout. XPLAIN: A system for creating and explaining expert consulting programs. *Artificial Intelligence*, **21**:285–325, 1983.

[233] B. K. Sy. Reasoning MPE to multiply connected belief networks using message passing. En: *Proceedings of the 10th National Conference on AI (AAAI-92)*, págs. 570–576, San Jose, CA, 1992.

[234] P. Szolovits y S. G. Pauker. Categorical and probabilistic reasoning in medicine. *Artificial Intelligence*, **11**:115–144, 1978.

[235] M. J. Taboada. *Un sistema para el diseño, desarrollo e implementación de interfases para sistemas expertos.* Tesis doctoral, Universidad de Santiago, Santiago de Compostela, 1993.

[236] R. E. Tarjan y M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclity of hypergraphs and selectively reduce acyclic hypergraphs. *SIAM Journal of Computing*, **13**:566–579, 1984.

[237] R. L. Teach y E. H. Shortliffe. An analysis of physician attitudes regarding computer-based clinical consultation systems. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 34, págs. 635–652. Addison-Wesley, Reading, MA, 1984.

[238] A. M. Turing. Computing machinery and intelligence. *Mind*, **59**:433–460, 1950.

[239] A. Tversky y D. Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, **185**:1124–1131, 1974.

[240] W. van Melle. The structure of the MYCIN system. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 4, págs. 67–77. Addison-Wesley, Reading, MA, 1984.

[241] W. van Melle. A domain-independent production-rule system for consultation programs. En: *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, págs. 923–925, Detroit, MI, 1989.

[242] W. van Melle, E. H. Shortliffe y B. G. Buchanan. EMYCIN: A knowledge engineer's workbench. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 15, págs. 302–313. Addison-Wesley, Reading, MA, 1984.

[243] J. von Neumann. The general and logical theory of automata. En: L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, págs. 1–41. Wiley, New York, 1951.

[244] J. W. Wallis y E. H. Shortliffe. Customized explanations using causal knowledge. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 20, págs. 371–388. Addison-Wesley, Reading, MA, 1984.

[245] H. R. Warner, A. F. Toronto y L. G. Veasy. Experience with Bayes' theorem for computer diagnosis of congenital heart disease. *Annals of the New York Academy of Sciences*, **115**:558–567, 1964.

[246] D. A. Waterman. *A Guide to Expert Systems*. Addison-Wesley, Reading, MA, 1986.

[247] J. L. Weiner. BLAH, a system which explains its reasoning. *Artificial Intelligence*, **15**:19–48, 1980.

[248] S. M. Weiss, C. A. Kulikowski, S. Amarel y A. Safir. A model-based method for computer-aided medical decission-making. *Artificial Intelligence*, **11**:145–172, 1978.

[249] M. P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, **44**:257–303, 1990.

[250] M. P. Wellman. Graphical inference in qualitative probabilistic networks

[251] S. Wright. Correlation and causation. *Journal of Agricultural Research*, **20**:557–585, 1921.

[252] J. Wyatt y D. Spiegelhalter. Evaluating medical expert systems: What to test, and how? En: J. Talmon y J. Fox (eds.), *Knowledge Based Systems in Medicine*, págs. 274–289. Springer-Verlag, Berlin, 1989.

[253] M. Yannakakis. Computing the minimal fill-in is NP-complete. *SIAM Journal of Algebraic Discrete Methods*, **2**:77–79, 1981.

[254] V. L. Yu, L. M. Fagan, S. M. Wraith, W. J. Clancey, A. C. Scott, J. F. Hannigan, R. L. Blum, B. G. Buchanan y S. N. Cohen. An evaluation of MYCIN's advice. En: B. G. Buchanan y E. H. Shortliffe (eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, cap. 31, págs. 589–596. Addison-Wesley, Reading, MA, 1984.

[255] L. A. Zadeh. Fuzzy sets. *Information and Control*, **8**:338–353, 1965.

[256] L. A. Zadeh. Fuzzy sets as a basis for the theory of possibility. *Fuzzy Sets and Systems*, **1**:3–28, 1978.

[257] L. A. Zadeh. A computational approach to fuzzy quantifiers in natural language. *Computers and Mathematics*, **9**:149–184, 1983.

[258] L. A. Zadeh. Linguistic variables, approximate reasoning and dispositions. *Medical Informatics*, **8**:173–186, 1983.

[271] L. A. Zadeh. Is probability theory sufficient for dealing with uncertainty in AI: A negative view. En: L. N. Kanal y J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, págs. 103–116. Elsevier Science Publishers, Amsterdam, 1986.

[272] L. A. Zadeh. The calculus of fuzzy if/then rules. *AI Expert*, págs. 23–27, Marzo 1992.