

Aplicaciones de los modelos gráficos probabilistas en medicina*

F. J. Díez[†]

Abstract

La medicina tiene dos propiedades que hacen que los modelos gráficos probabilistas (MGP) encajen en ella como anillo al dedo: el conocimiento causal, correspondiente a los mecanismos patofisiológicos, y las numerosas fuentes de incertidumbre. Por ello, no es de extrañar que la mayor parte de los MGP, desde el principio hasta la actualidad, se hayan desarrollado en el campo de la medicina. En este artículo revisamos algunos de ellos y abordamos después aspectos generales, como la construcción de MGP en medicina y la explicación del razonamiento.

1 Introducción

1.1 Sistemas expertos en medicina: perspectiva histórica

El desarrollo de programas de diagnóstico basados en técnicas bayesianas comenzó en los años 60. Entre los sistemas de esa década destacan el de Warner y colaboradores [43] para el diagnóstico de cardiopatías congénitas, los de Gorry y Barnett [14, 15] y el de de Dombal [8] y colaboradores para el diagnóstico del dolor abdominal agudo. Estos sistemas aplicaban el **método probabilístico clásico**, que consiste en seleccionar una variable D , que representa los n diagnósticos posibles d_i , y m variables H_j —binarias en general— correspondientes a los posibles hallazgos, que en medicina suelen ser los síntomas y signos; para que el problema sea tratable se introducen dos hipótesis: la primera, que los diagnósticos son *exclusivos y exhaustivos* y, la segunda, la *independencia condicional*, es decir, que los hallazgos son independientes entre sí para cada diagnóstico:

$$P(h_1, \dots, h_m | d_i) = P(h_1 | d_i) \cdot \dots \cdot P(h_m | d_i), \quad \forall d_i \quad (1)$$

Con estas hipótesis, el teorema de Bayes se reduce a la siguiente expresión:

$$P(d_i | h_1, \dots, h_m) = \frac{P(h_1 | d_i) \cdot \dots \cdot P(h_m | d_i) \cdot P(d_i)}{\sum_j P(h_1 | d_j) \cdot \dots \cdot P(h_m | d_j) \cdot P(d_j)} \quad (2)$$

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

$$\frac{P(d_i | h_1, \dots, h_m)}{P(d_j | h_1, \dots, h_m)} = \frac{P(h_1 | d_i)}{P(h_1 | d_j)} \cdot \dots \cdot \frac{P(h_n | d_i)}{P(h_n | d_j)} \cdot \frac{P(d_i)}{P(d_j)} \quad (3)$$

Aunque este método sirvió de base a los sistemas de diagnóstico ya citados, con resultados satisfactorios para pequeños problemas, presenta serias deficiencias, pues ni en medicina los diagnósticos suelen ser exclusivos ni se da en general la independencia condicional, sino que los hallazgos correspondientes a cada diagnóstico suelen estar correlacionados.

*En: J. A. Gámez y J. M. Puerta (eds.). *Sistemas Expertos Probabilísticos*. Universidad de Castilla-La Mancha, Cuenca, 1998.

[†]Dpto. Inteligencia Artificial. UNED. Juan del Rosal, 16. 28040 Madrid.
Correo-e: <fjdiez@dia.uned.es>. WWW: <http://www.ia.uned.es/~fjdiez>.

Como consecuencia de las dificultades que presentaba el método probabilístico clásico, los creadores del programa **MYCIN**¹ diseñaron en los años 70 un modelo que, en vez de buscar un fundamento matemático sólido, trataba de reproducir la forma en que el ser humano combina intuitivamente distintas fuentes de información. La idea básica consistía en asignar a cada regla “Si E entonces H ” un **factor de certeza**, $CF(H, E)$. Aunque estos factores se definieron a partir de las probabilidades $P(H)$ y $P(H|E)$, en la práctica se obtenían directamente a partir de estimaciones de expertos humanos y se combinaban según reglas *ad hoc*, sin tener en cuenta los principios de la teoría de la probabilidad. A pesar del éxito que obtuvo MYCIN, cuyo índice de aciertos era comparable al de los mejores expertos humanos, pronto se comprobó —mediante razonamientos matemáticos— que contenía graves inconsistencias, por lo que fue duramente criticado (cf. [27, sec. 1.2] y [10, sec. 2.4]).

Examinando los sistemas expertos de la década de los 80, observamos que la mayor parte de ellos se basaron en la **lógica difusa** y, en menor medida, en la teoría de la posibilidad, lo cual no es de extrañar, teniendo en cuenta que una parte considerable de los conceptos médicos son difusos: presión alta, dolor agudo, fatiga leve, tumor grande, síntoma evidente, prueba muy sensible, diagnóstico complejo, pronóstico grave, terapia arriesgada, cirugía mínimamente invasiva, alta mortalidad, etc., etc. Sin embargo, al contrario de lo que ocurre otras metodologías de razonamiento aproximado, las distintas aplicaciones de la lógica difusa difieren notablemente entre sí, pues esta teoría carece de un fundamento normativo que indique cómo se debe aplicar en cada caso.

Pero fue también en la década de los 80 cuando se desarrollaron las **redes bayesianas** y los **diagramas de influencia**, desde su definición axiomática hasta el diseño de algoritmos eficientes para la computación de la evidencia, y pronto se vio que venían “como anillo al dedo” para el tratamiento de la incertidumbre en medicina. De hecho, como veremos en la sección 2, los primeros y la mayor parte de los sistemas de diagnóstico probabilístico se han construido en este campo, con un crecimiento exponencial en los últimos años.

Por otra parte, cada vez son más los argumentos tanto teóricos como empíricos² a favor del formalismo bayesiano, hasta el punto de que los propios creadores del programa MYCIN afirmaron en 1993 [7]:

En la última década, la investigación sobre tratamiento de la incertidumbre en Inteligencia Artificial 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.

1.2 Fuentes de incertidumbre en medicina

En prácticamente todas las aplicaciones de la inteligencia artificial surgen la incertidumbre y la imprecisión, fundamentalmente por tres motivos: deficiencias de la información, indeterminismo del mundo real y deficiencias de los modelos; los distintos métodos de razonamiento incierto han tratado de abordar al menos uno de estos tres tipos de incertidumbre. En medicina se pueden identificar fácilmente los siguientes:

- *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 con la información que posee, aunque sea muy limitada.

¹El proyecto MYCIN, desarrollado en la Universidad de Stanford, tenía como objetivo construir un sistema experto para el tratamiento de enfermedades infecciosas. MYCIN está considerado en la actualidad como el primer sistema experto y el “padre” de todos los sistemas basados en reglas.

²Pueden encontrarse las referencias en [10].

- *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 debe mantener siempre 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 síntomas como el dolor o la fatiga. Incluso en un método tan técnico como la ecocardiografía, por ejemplo, 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 diferente, 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 tener siempre en cuenta la probabilidad y 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, rara vez está disponible toda esta información, 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; por ejemplo, mediante la asignación de intervalos o de distribuciones de probabilidad para las probabilidades condicionales en el caso de los MGP.

De aquí se deducen dos razones recíprocas que explican por qué todos los modelos de razonamiento incierto se han centrado en alguna rama de la medicina: por un lado, la necesidad de abordar un problema médico concreto ha llevado en ocasiones a desarrollar un nuevo método, que luego se ha generalizado a distintos tipos de problemas y, por otro lado, la medicina constituye un excelente campo de pruebas para observar las cualidades y limitaciones de cualquier nuevo método que se proponga.

En los MGP se cumple claramente este principio: los primeros sistemas expertos basados en redes bayesianas tuvieron como objeto algún problema médico concreto y, de hecho, hoy en día es la medicina el campo donde se han desarrollado la mayor parte de los sistemas basados en MGP. Describimos los más importantes en la sección siguiente.

2 Ejemplos de MGP en medicina

2.1 Aplicaciones desarrolladas fuera de España

En esta sección nos vamos a centrar en los modelos normativos, es decir, en los que se ajustan a los principios de la teoría de la probabilidad y de la decisión, que, en la práctica, son aquéllos cuya base de conocimientos viene dada por una red bayesiana o por un diagrama de influencia. No vamos a describir aquí los sistemas expertos, como CASNET o el *Heart Disease Program*, de W. Long, que utilizan modelos probabilísticos aproximados. Tampoco vamos a hablar de los basados en el método probabilístico clásico, porque ya lo hicimos en la sección 1.1.

Siguiendo el orden de complejidad creciente —que no coincide con el cronológico— debemos mencionar el modelo de Schwartz, Baron y Clarke [33] para el diagnóstico de la apendicitis. Frente al

método clásico tenía la ventaja de que, al introducir causas intermedias, salvaguardaba la independencia condicional de ciertos hallazgos correlacionados respecto del diagnóstico principal. Frente a las redes bayesianas, presentaba la limitación de que cada nodo sólo podía tener un padre y, en consecuencia, tampoco admitía bucles.

La primera red bayesiana médica fue construida por Cooper [5, 6] en la Universidad de Stanford, como ejemplo para la aplicación del programa NESTOR. Entre los aspectos más avanzados de este sistema destacaban la posibilidad de definir las probabilidades condicionales mediante intervalos, y la capacidad de explicación, de la que hablaremos en la sección 4.3. La red que utilizó como ejemplo ilustrativo contenía cinco nodos: cáncer metastásico, elevación del calcio sérico, tumor cerebral, coma y jaquecas.

Otro de los primeros ejemplos de redes bayesianas médicas es la de Lauritzen y Spiegelhalter [23]; tiene 8 nodos y realiza el diagnóstico diferencial entre tuberculosis, bronquitis y cáncer de pulmón. Éstas son, con diferencia, las dos redes bayesianas más famosas, por haber sido utilizadas repetidamente para ilustrar muchos de los algoritmos que se han desarrollado desde entonces; obviamente, dos redes tan minúsculas no intentan resolver problemas reales, sino que sólo son útiles con fines ilustrativos.

En cuanto a las redes bayesianas con un conocimiento extenso, destinadas al diagnóstico clínico real, la primera y una de las más conocidas fue desarrollada por un equipo de investigadores de la Universidad de Aalborg (Dinamarca) [2, 26]; recibió el nombre de MUNIN y estaba destinada al diagnóstico de enfermedades musculares mediante electromiografía.³ Los nodos estaban agrupados en tres niveles: enfermedades, estados patofisiológicos y hallazgos. También en este grupo se desarrolló un sistema que permitía modelar el metabolismo de los carbohidratos con el fin de ajustar la dosis de insulina en pacientes diabéticos [1].

Volviendo a la Universidad de Stanford, destaca por su importancia el sistema experto PATHFINDER, de David Heckerman [18], destinado al diagnóstico de enfermedades de los ganglios linfáticos. La principal aportación del trabajo de Heckerman es la creación de las *redes de semejanza* (“*similarity networks*”), que se caracterizan por la existencia de un nodo principal, que representa los posibles diagnósticos. La limitación principal de este modelo es suponer que el paciente padece una sola enfermedad, lo cual es una hipótesis razonable en el caso de los ganglios linfáticos, pero resultaría inverosímil en otros dominios, como la cardiología, en que las enfermedades suelen estar relacionadas; a cambio, facilita la construcción del modelo (porque se centra en el diagnóstico diferencial de cada par de enfermedades), hace más eficiente la computación de la probabilidad y permite explicar el razonamiento (cf. sec. 4.3). La empresa Intellipath, que comercializa actualmente el sistema PATHFINDER, ha vendido cientos de copias, que se usan a diario en numerosos hospitales.

En la misma universidad se construyó el sistema QMR-DT [35], que es una reformulación en forma de red bayesiana del sistema experto QMR, el cual a su vez, era la versión comercial de INTERNIST-1. Las dos limitaciones principales de QMR-DT son la representación todos los diagnósticos y hallazgos mediante variables binarias y la disposición de los nodos en dos niveles, sin permitir variables intermedias; estas dos simplificaciones impiden representar correctamente las relaciones de independencia, como explican muy bien sus propios creadores. Igualmente, el programa Iliad, un tutor para medicina interna construido inicialmente mediante un modelo probabilista aproximado, ha sido reformulado posteriormente en forma de red bayesiana, con lo que se ha logrado mejorar su rendimiento [25].

En Europa, el grupo más importante dedicado a las redes bayesianas médicas —junto con el de Aalborg— es el de la Universidad de Pavía, en Italia, donde se han desarrollado redes bayesianas y diagramas de influencia para optimización de terapia en anemia urémica, monitorización, leucemia infantil, hemodiálisis, diabetes, SIDA, tratamiento de niños con trasplantes de médula ósea, nefritis, linfoma gástrico primario, trombosis idiopática en venas profundas, esplenotomía, etc.⁴

El mexicano Luis Enrique Súcar [39] ha desarrollado un sistema de visión artificial para colonoscopia; además de ser —según nuestros conocimientos— la única red bayesiana para visión que resuelve un problema real, tiene el mérito de que el sistema es capaz de obtener las probabilidades condicionales e

³Dentro de este mismo proyecto se desarrolló HUGIN, una herramienta destinada a la construcción de redes bayesianas, que es comercializada actualmente por la empresa del mismo nombre.

⁴Las referencias pueden encontrarse en <http://ipvamed9.unipv.it/lab/publications.html>.

incluso refinar la estructura de la red a partir de los datos disponibles.

Entre las redes bayesianas más utilizadas se encuentra el programa *Microsoft Pregnancy and Child Care*, que ofrece sus consejos en la red de Microsoft;⁵ la base de conocimientos fue desarrollada y comprobada por Knowledge Industries, empresa que también ha construido redes bayesianas para dermatología, alteraciones del sueño, cuidado de traumatismos, chequeo de mano y muñeca y atención sanitaria a domicilio.⁶

Otros sistemas basados en MGP son: ALARM [3], para la monitorización de pacientes durante la anestesia; VP-net [32], para monitorización e interpretación de datos en la UCI; THOMAS [24], para interpretar los resultados de experimentos clínicos aleatorizados publicados; ABDO [29], para el diagnóstico del dolor abdominal agudo; el de Haddawy y colaboradores [17], para la vesícula biliar; CPCS-BN [28], para enfermedades heptobiliares; MammoNet [21], para enfermedades de mama; etc.⁷

Una mención a parte merece el programa BANTER, de Haddawy, Jacobson y Kahn [16], cuyo objetivo no es el diagnóstico ni la toma de decisiones, sino la enseñanza de la medicina a partir de cualquier red bayesiana o diagrama de influencia cuyos nodos puedan clasificarse en hipótesis, observaciones y métodos diagnósticos. Además de calcular la probabilidad a posteriori de cada hipótesis, BANTER es capaz de seleccionar el mejor método diagnóstico para confirmar o descartar cualquier hipótesis, de examinar al usuario sobre la selección de métodos diagnósticos, y de explicar su razonamiento (utilizando el método INSITE, de Suermondt, del que hablaremos en la sec. 4.3).

2.2 Aplicaciones desarrolladas en España

DIIVAL [10, 12] es un sistema experto para el diagnóstico de enfermedades cardíacas, que considera principalmente la información ecocardiográfica, aunque teniendo en cuenta también otras fuentes de información: datos personales, síntomas y signos, hallazgos electrocardiográficos, etc. Fue desarrollado por Francisco J. Díez, de la Universidad Nacional de Educación a Distancia, en colaboración con el Hospital de la Princesa, de Madrid. El nombre se debe a que inicialmente estaba orientado al DIAGNÓSTICO de VALvulopatías. En la sección 4 hablaremos de sus aportaciones en cuanto al paso de una red bayesiana a un sistema experto completo.

DIABNET es un sistema de planificación de terapias en diabetes gestacional, construido por Elena Hernando [19] de la Facultad de Telecomunicación de la Universidad Politécnica de Madrid, en colaboración con el Servicio de Endocrinología y Nutrición del Hospital San Pau de Barcelona. Su base de conocimiento está constituida por una red bayesiana que modeliza cualitativamente el metabolismo de la glucosa. Dado que está orientado a la monitorización y al seguimiento de una enfermedad, el empleo de redes dinámicas desempeña un papel esencial.

IctNeo [31] es un sistema destinado al tratamiento de la ictericia neonatal, que está siendo desarrollado por varios investigadores del Departamento de Inteligencia Artificial de la Universidad Politécnica de Madrid, en colaboración con el Hospital Gregorio Marañón de Madrid. Además de las dificultades inherentes a la construcción del diagrama de influencia (estructura, probabilidades condicionales y funciones de utilidad), el tamaño relativamente elevado de la red —59 nodos en la última versión, con numerosos bucles— dificulta el cálculo que llevará a determinar la política terapéutica.

En la Universidad del País Vasco, Basilio Sierra y Pedro Larrañaga [36] han desarrollado un método para la construcción de redes bayesianas a partir de bases de datos mediante algoritmos genéticos, y lo han aplicado al pronóstico en dermatología, concretamente a la predicción de supervivencia de pacientes con melanoma maligno (véase el capítulo de P. Larrañaga en este mismo libro).

Por último, mencionamos dos proyectos en curso: el de Carmen Lacave y Juan Giralt, de la Universidad de Castilla-La Mancha, para el diagnóstico diferencial de enfermedades infecciosas en pediatría,

⁵Microsoft está desarrollando también una red bayesiana para cardiología (comunicación personal de Eric Horvitz y Jack Breese), aunque aún no conocemos referencias escritas.

⁶Las referencias pueden encontrarse a partir de <http://www.auai.org/auai-companies.html>.

⁷El código completo de algunas de las redes mencionadas en esta sección, como PATHFINDER, MUNIN y ALARM, puede encontrarse en <http://www-nt.cs.berkeley.edu/home/nir/public.html/Repository/>.

y el de Enrique Nell, para el diagnóstico de enfermedades del miocardio. Ambos se encuentran aún en sus comienzos.

3 Construcción de MGP en medicina

Hay básicamente dos métodos para la construcción de redes bayesianas:

- *A partir de una base de datos*, aplicando alguno de los métodos de aprendizaje de redes descritos en capítulos anteriores de este libro.
- *Con la ayuda de expertos humanos* (médicos de la especie, en nuestro caso), mediante una serie de sesiones en que el constructor del modelo interroga a los expertos y, con el conocimiento obtenido, va añadiendo nodos-variables, enlaces y probabilidades condicionales a la red.

Describimos cada uno de ellos en las dos secciones siguientes.

3.1 Construcción a partir de bases de datos

La forma más rápida de construir red bayesiana para medicina consiste en tomar una base de datos que contenga un número suficientemente grande de casos (de pacientes, generalmente) y aplicar algún algoritmo de aprendizaje. Como estos algoritmos ya se han descrito en capítulos anteriores de este libro, nos vamos a limitar a discutir aquí su aplicación en medicina.

En primer lugar, comprobamos que la mayoría de estos algoritmos suponen que tenemos una base de datos en que *el valor de cada variable está determinado con certeza* para cada caso. Sin embargo, la mayor parte de las bases de datos médicas sólo recogen unos pocos de los datos observados, junto con el diagnóstico final. En cambio, la construcción de una red bayesiana requiere especificar numerosas variables intermedias, para que tengan validez las hipótesis de independencia.

Es cierto que existen algoritmos 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 médico, con lo que la validez del modelo resultaría más que cuestionable.

Esto explica por qué, a pesar de los numerosos trabajos sobre aprendizaje automático de redes bayesianas, ninguna de las aplicaciones mencionadas en la sec. 2 —salvo las construidas con fines académicos— se hayan construido mediante estos algoritmos.

Finalmente, señalemos que tales algoritmos podrían ser útiles, a lo sumo, para construir redes bayesianas, pero no para generar diagramas de influencia, pues son incapaces de extraer de las bases de datos nodos-decisión y nodos-utilidad.

3.2 Construcción con la ayuda de expertos humanos

La construcción de un MGP puede dividirse en dos fases. La primera de ellas consiste en recopilar la *información cualitativa*, es decir, en 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. Veamos cada una de ellas por separado.

3.2.1 Obtención de la información cualitativa.

Todo MGP implica un número —generalmente elevado— de relaciones de independencia condicional, que, en principio, habría que justificar mediante análisis estadísticos; sin embargo, la falta de datos empíricos impide casi siempre realizar tal comprobación (los trabajos de Luis Enrique Súcar [40, 39] constituyen una notable excepción). La alternativa más utilizada consiste en aplicar conocimiento sobre los **mecanismos causales**, aunque rara vez los constructores de los modelos se cuestionan las hipótesis que están introduciendo (aquí, las excepciones son el trabajo de Shwe y colaboradores [35] y el de Díez

[11], que resumimos a continuación). Por ello, debemos interrogar a los expertos sobre los mecanismos causales que, a su juicio, intervienen en nuestro problema, y a partir de ellos intentar justificar las propiedades de independencia mediante la aplicación de las reglas siguientes (véase la fig. 1):

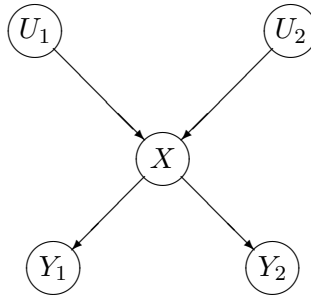


Figure 1: Independencia condicional para un nodo X con dos padres y dos hijos.

Independencia a priori. Cuando hay dos variables U_1 y U_2 tales que (1) no existe correlación conocida entre una y otra, (2) no hay ningún mecanismo causal por el que U_1 pueda producir U_2 , ni viceversa, y (3) no hay ninguna causa común de ambas, entonces podemos suponer que son a priori independientes. Por ejemplo, entre el sexo y el país de origen podemos suponer que hay independencia a priori. Cuando la correlación es pequeña (por ejemplo, entre sexo y edad), podemos considerar la posibilidad de despreciarla y tratar las variables como independientes a priori, con el fin de no complicar excesivamente la propagación de evidencia.

Independencia condicional entre varios efectos de una causa. Cuando (1) X es una causa común de Y_1 e Y_2 , (2) el mecanismo causal por el que X produce Y_1 no interactúa con el mecanismo $X \rightarrow Y_2$, (3) no hay ninguna relación causal conocida $X \rightarrow Y_1$ ni $X \rightarrow Y_2$, y (4) no hay ninguna otra causa común de Y_1 e Y_2 , entonces podemos suponer que ambas son condicionalmente independientes dado X . Por ejemplo, entre un síntoma Y_1 y una prueba de laboratorio Y_2 indicativas de una misma enfermedad X , podemos suponer casi siempre que hay independencia condicional.

Independencia condicional entre un efecto y sus “abuelos”. Cuando (1) las causas de X son U_1, \dots, U_n , (2) el mecanismo $X \rightarrow Y$ es independiente de cómo se ha producido X , y (3) no hay ningún otro mecanismo conocido $U_i \rightarrow Y$, entonces podemos suponer que U_i e Y son condicionalmente independientes dado X . Por ejemplo, la zona de origen (U_1) y el grupo sanguíneo (U_1) son dos factores de riesgo para el paludismo (X); en la práctica, podemos suponer que la probabilidad de que el test de la gota gruesa (Y) —la prueba más habitual para detectar el paludismo— dé positivo es independiente de la zona de origen y del grupo sanguíneo una vez que conozcamos con certeza si una persona padece la enfermedad o no.

Desgraciadamente, hay muchos casos en que no se conocen los mecanismos causales que llevan a una determinada alteración. Por ejemplo, un libro de cardiología puede afirmar que los principales factores riesgo del infarto agudo miocardio (IAM) son la edad, el ser varón, el ser de raza blanca, los antecedentes familiares, el tabaquismo, la obesidad, el estrés, la ingesta elevada de sodio, la hipercolesterolemia, la diabetes y la hipertensión arterial. Obviamente, estos ocho factores no son independientes entre sí, pero resulta imposible conocer en qué medida cada uno de ellos afecta a los demás, pues, que nosotros sepamos, ninguno de los numerosos estudios epidemilógicos que se han llevado a cabo sobre el IAM ha intentado estudiar la dependencia e independencia condicional entre sus factores de riesgo.

3.2.2 Aplicación de la puerta OR.

Otro de los puntos importantes, posterior a la determinación de la estructura de la red y previo a la obtención de la información cuantitativa, consiste en decidir cuáles de las familias de la red pueden ser

modeladas mediante la puerta OR. En efecto, la puerta OR requiere *muchos menos parámetros* que el modelo general, lo cual conlleva ventajas en cuanto al almacenamiento de la información, en cuanto a la *propagación de la evidencia* y, sobre todo, en cuanto a la *obtención del conocimiento*, no sólo porque necesita muchos menos parámetros, sino porque los parámetros que intervienen son mucho más significativos para un médico y más fáciles de estimar que los elementos de una tabla de probabilidad; por ejemplo, tienen mucho más sentido las preguntas como “¿cuál es la probabilidad de que U_i produzca X ?” que “¿cuál es la probabilidad de $+x$ cuando $+u_1, -u_2, +u_3$ y $+u_4$?”, con la complicación adicional de que nuestro especialista probablemente nunca ha visto un enfermo que padeciera a la vez U_1, U_3 y U_4 , con lo que le resultaría absolutamente imposible estimar dicha probabilidad.

Por último, la puerta OR presenta ventajas en cuanto a la *explicación del razonamiento*: concretamente, en presencia de un síntoma o signo S , la confirmación de una enfermedad causante de S resta credibilidad a las otras causas de S (este fenómeno se denomina en inglés “*explaining away*”); y viceversa, la exclusión de todas las causas de S excepto una, lleva a considerar ésta como el diagnóstico más probable. Este tipo de razonamiento, que en medicina se denomina *diagnóstico diferencial*, es específico de la puerta OR, y no se da en caso general.

Por tanto, es muy deseable aplicar la puerta OR siempre que sea posible, pero para ello han de darse ciertas condiciones:

1. tanto el nodo hijo como sus padres han de ser variables que indiquen el grado de presencia de una anomalía; es decir, el rango de valores debe ser “ausente/presente” o “ausente/leve/moderada/severa” o un conjunto similar [9]; esto impide la aplicación de la puerta OR cuando los padres representan otro tipo de variables, como la edad, el sexo o la raza;
2. cada uno de los padres representa una causa que puede producir el efecto (el nodo hijo) en ausencia de las demás causas;
3. no hay sinergia entre las causas; es decir, el mecanismo por el que U produce X es independiente de los mecanismos de las demás causas de X (obsérvese que estamos hablando nuevamente de **causalidad**).

Por tanto, las condiciones 2 y 3 impiden la aplicación de la puerta OR cuando los padres representan factores de riesgo, tales como el tabaquismo, la obesidad, la hipercolesterolemia, etc., ninguno de los cuales es capaz de *producir* (causar) la enfermedad (el infarto agudo de miocardio, volviendo al ejemplo anterior) en ausencia de los demás.

3.2.3 Obtención de la información cuantitativa.

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, difícilmente vamos a encontrar más que una pequeña parte de 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:

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*? Hay estudios psicológicos que pueden ofrecer una cierta ayuda a la hora de convertir las expresiones cualitativas en probabilidades numéricas, pero las variaciones en las asignaciones son tan grandes que resultan de poca utilidad.

El único número concreto que aparece, “*el 75%*”, —no sabemos si se trata de un resultado experimental o de una estimación aproximada— tampoco es de gran ayuda, porque no indica la probabilidad

de que haya un mixoma en la aurícula izquierda, sino de que, habiendo mixoma, se localice en la aurícula izquierda, lo cual no es un dato que se pueda introducir directamente en la red.

Con este sencillo ejemplo pretendemos mostrar por qué se hace necesario casi siempre recurrir a estimaciones subjetivas de expertos humanos, a pesar de que la labor es tediosa y compleja (cf. [37], [22, cap. 4]).

3.3 Funciones de utilidad en medicina

La obtención de funciones de utilidad en medicina es igualmente difícil. Algunos de los primeros trabajos utilizaban escalas subjetivas, graduadas de 0 a 100; este método fue criticado porque las unidades de medida eran arbitrarias, es decir, sin ningún significado médico objetivo, y variaban de una aplicación a otra dependiendo de cuáles fueran los extremos escogidos.

Por eso se desarrollaron otros métodos basados en datos objetivos, como la supervivencia a corto plazo (expresada en porcentajes) o la no morbilidad (el número de casos en que se curaba el paciente). La variable que con más frecuencia se ha empleado para determinar la utilidad es la esperanza de vida del paciente, medida normalmente en años, y con este criterio se han construido numerosos modelos y programas de ordenador desde la década de los 70. Sin embargo, no sólo es importante la duración de la vida, sino también la calidad, y por eso el criterio más adecuado en general es la **esperanza de vida en salud** (en inglés, “*quality-adjusted life-expectancy*”), que se define como el tiempo que va a vivir el paciente multiplicado por la calidad de vida que va a tener; más exactamente, teniendo en cuenta que la calidad de vida varía con el tiempo, $c(t)$, la **vida en salud** para un paciente se define como

$$VS = \int c(t) \cdot dt \quad (4)$$

La unidad de medida se denomina en inglés “*quality-adjusted life-year*”; nosotros hemos propuesto como traducción el término “**año-salud**” [22, pg. 66].⁸

Sin embargo, hay casos en que las funciones de utilidad universales no tienen sentido. Por ejemplo, ante la posibilidad de un embarazo, unas parejas manifiestan más interés que otras por tener un niño (en unos casos la utilidad será positiva y en otros negativa), el riesgo que cada mujer está dispuesta a asumir es distinto, cada pareja valora de forma diferente las posibles malformaciones congénitas del futuro niño, etc. Por eso se han desarrollado métodos que intentan conocer y medir las preferencias de cada paciente. (Una discusión más extensa sobre la obtención de funciones de utilidad en medicina puede encontrarse en [22, cap. 3]).

Para concluir, comentamos que, cuando se trata de decidir si el coste económico de una terapia o un plan de actuación compensa las ventajas obtenidas, éstas pueden medirse de tres modos:

Análisis coste-efectividad: Mide la utilidad en alguna *unidad médica*, como el número de vidas salvadas o el porcentaje de hipertensos controlados.

Análisis coste-beneficio: Asigna un valor económico a los estados resultantes, incluida la vida o la muerte del paciente, con el fin de medir la utilidad en alguna *unidad monetaria*.

Análisis coste-utilidad: Valora la *calidad de vida* del paciente, generalmente teniendo en cuenta sus *preferencias*, como acabamos de explicar.

Naturalmente, los diagramas de influencia son capaces de englobar estos tres tipos de análisis dentro de un mismo formalismo, pues una vez conocida la función de utilidad el tratamiento matemático es idéntico.

⁸Un *año-luz* es la distancia que la luz recorre en un año; análogamente, un *año-salud* es la “cantidad de salud” que una persona sana disfruta a lo largo de un año, o bien la salud que una persona con la calidad de vida reducida a la mitad disfruta en dos años, etc.

4 De un MGP a un sistema experto

En la sección anterior hemos hablado sobre la construcción de modelos gráficos probabilistas. Sin embargo, tales modelos no pueden considerarse por sí mismos como sistemas expertos, pues para ello necesitan, como mínimo,

- un *interfaz gráfico amigable*, de modo que el usuario pueda introducir la información de algún modo que le resulte familiar, sin tener que buscar en la red la variable correspondiente a cada hallazgo;
- un *generador de informes* que seleccione las conclusiones más relevantes, pues el mostrar en una ventana la probabilidad a posteriori de cada variable es claramente insuficiente;
- un método de *explicación del razonamiento*, que justifique el diagnóstico y las recomendaciones ofrecidas por el sistema, de modo que el usuario pueda aceptarlas o rechazarlas según su criterio.

Sin estas tres características, es seguro que incluso la red bayesiana que calcule las probabilidades más exactas o el diagrama de influencia que encuentre las mejores decisiones en cada caso, nunca llegarán a aplicarse en la práctica. Por eso vamos a describir a continuación las distintas soluciones que se han propuesto para cada uno de estos tres puntos.

4.1 Interfaz de usuario

Una limitación de los primeros sistemas expertos —basados en reglas— es que dirigían la consulta mediante una serie de preguntas, sin que el usuario pudiera tomar la iniciativa sobre la información que deseaba introducir. En cambio, en la mayor parte de los modelos gráficos probabilistas, el problema es más bien el contrario, pues suelen limitarse a ofrecer una pantalla en la que se muestra la red, de modo que el usuario debe señalar los nodos a los que desea asignar un valor en función de los hallazgos disponibles (el ejemplo más conocido es MUNIN [2]).

Una paso adelante lo constituye PATHFINDER, que agrupa los posibles hallazgos en categorías, lo cual facilita al usuario localizar el que desea introducir, e incluso sugiere cuál es el síntoma o signo que debe buscar el usuario en función del coste-efectividad [18, figs. 1.2 a 1.6].

El sistema experto DIAVAL, en cambio, implementa lo que en la terminología de los sistemas expertos se conoce como *interfaz de iniciativa mixta*, pues, por un lado, ofrece una serie de pantallas en un orden determinado, con lo que se facilita la recogida sistemática de los hallazgos ecocardiográficos, mientras que, por otro, ofrece una serie de menús que permiten acceder directamente a la ventana deseada.

4.2 Generación de diagnósticos e informes

Algunos de los sistemas de diagnóstico probabilistas se limitan a ofrecer sus conclusiones en una pantalla donde muestran la probabilidad a posteriori para cada variable [2]. Otros, como PATHFINDER [18], presentan una lista de las variables correspondientes al diagnóstico, ordenadas de mayor a menor probabilidad.

DIAVAL [10, cap. 13], en cambio, aborda el problema estableciendo dos *umbrales*, de *certeza* y de *relevancia*, y asignando a cada nodo dos factores: la *relevancia para diagnóstico positivo* (RDP) y la *relevancia para diagnóstico negativo* (RDN), en una escala subjetiva de 0 a 10;⁹ naturalmente, las enfermedades tienen factores de relevancia más altos que los estados patofisiológicos y las alteraciones intermedias. Por otro lado, los nodos están agrupados en 21 *capítulos*, cada uno de los cuales corresponde a una parte del corazón (una válvula, el pericardio, etc.) o a un tipo de información (antecedentes familiares, factores de riesgo, etc.).

Tras propagar la evidencia, el programa selecciona dentro de cada capítulo aquellos nodos que superan tanto el umbral de certeza como el de relevancia; bajando estos umbrales, el usuario puede

⁹Diagnósticos positivos son, por ejemplo, la estenosis mitral leve, moderada y severa. La ausencia de estenosis es un diagnóstico negativo.

conseguir que se muestren diagnósticos menos probables o menos importantes, respectivamente. Esto permite presentar las conclusiones de forma ordenada, primero en una pantalla, donde el médico puede introducir las correcciones que estime oportunas, y después en un informe escrito que pasa a un procesador de texto y a una impresora.

4.3 Explicación del razonamiento

Hemos mencionado ya que, para que una red bayesiana pueda considerarse como verdadero sistema experto, hace falta que pueda explicar su proceso de razonamiento.¹⁰ El objetivo principal de la capacidad de explicación es **justificar los resultados**, de modo que el usuario pueda decidir si las conclusiones que ofrece el programa son correctas o no; de hecho, es famoso el estudio de Teach y Shortliffe [42] en que se demuestra que los médicos son muy reticentes a aceptar el consejo de un ordenador si no pueden confirmar su fundamento.¹¹ Además, la capacidad de explicación es sumamente útil durante la construcción del modelo para **depurar los errores** que de otro modo serían difíciles de detectar y corregir. Y una tercera ventaja de la capacidad de explicación es la **enseñanza**; por ejemplo, en la sección 2 hemos mencionado el sistema BANTER, que permite utilizar redes bayesianas para instruir a estudiantes de medicina. En esta sección vamos a describir algunos de los métodos de explicación propuestos para sistemas probabilistas.

Como hemos dicho en la sección 1.1, los primeros sistemas bayesianos de diagnóstico se basaban en el método probabilístico clásico. Expresando la ecuación (3) en forma logarítmica,

$$\log \frac{P(d_i|h_1, \dots, h_m)}{P(d_j|h_1, \dots, h_m)} = \log \frac{P(d_i)}{P(d_j)} + \sum_{k=1}^n \frac{P(h_k|d_i)}{P(h_k|d_j)} \quad (5)$$

se observa que el primer término del segundo miembro es independiente de la evidencia encontrada, de modo que son 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 [4], en GLADYS [38] y en el sistema para la clasificación de apoplejías de Reggia y Perricone [30] y es, en esencia, el mismo que usa Heckerman en PATHFINDER [18, sec. 6.2.1]

En la misma línea, y dentro ya de las redes bayesianas, Sember y Zuckerman [34] abordan un problema diferente: cómo explicar los cambios en la probabilidad a posteriori de un nodo en un poliárbol mediante el análisis de los mensajes π y λ .

Otro trabajo interesante es el de Jensen y colaboradores [20] sobre la detección de conflictos en redes bayesianas. Para un conjunto de hallazgos \tilde{S} , definen una *medida del conflicto* como

$$\text{conf}(s_1, \dots, s_n) = \log_2 \frac{P(s_1) \cdot \dots \cdot P(s_n)}{P(s_1, \dots, s_n)} \quad (6)$$

Esta expresión refleja 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. Como el cálculo se realiza a partir de medidas locales, es posible rastrear el origen del conflicto, aunque con el inconveniente de que estas medidas locales no corresponden a la verdadera estructura de la red bayesiana, sino a la del árbol de cliques.

¹⁰En los sistemas de diagnóstico se habla a veces de *explicación* como un “conjunto de hipótesis capaz de justificar las anomalías observadas”; en cambio, aquí nos referimos a un concepto diferente: la *explicación del razonamiento* trata de mostrar cómo y por qué el sistema ha llegado a sus conclusiones.

¹¹Conviene destacar en relación con este punto la evolución que se ha producido en las dos últimas décadas en la forma de entender la inteligencia artificial. Anteriormente, el objetivo principal era construir programas que *igualaran o superaran* la capacidad de los seres humanos; por eso, en la evaluación de los sistemas expertos médicos se trataba de demostrar que diagnosticaban igual o mejor que los propios especialistas. Hoy en día, la inteligencia artificial trata ante todo de construir sistemas que *colaboren* de forma simbiótica con el ser humano, aprovechando lo mejor de lo natural y de lo artificial; en esta línea, la evaluación más positiva de un sistema experto será aquélla que demuestre que el médico ayudado por el sistema experto diagnostica más rápido y mejor que el médico solo. Y para que pueda darse esta simbiosis hombre-máquina es fundamental la explicación del razonamiento.

En una línea muy diferente, el sistema experto DIAVAL [10, cap. 8] ofrece un método de explicación que se basa en la distinción de seis tipos de enlaces, dos para el modelo general (influencia y parámetro) y cuatro para la puerta OR (causalidad, tipo, manifestación y observación), y ofrece varias opciones de explicación para cada nodo, enlace o dato cualitativo: probabilidad a priori, probabilidad a posteriori, causas, efectos, sensibilidad, especificidad, intervalos, fórmula con se calcula, etc. En particular, la utilización generalizada de la puerta OR permite explicar en la mayor parte de los casos cuál es la causa más probable de cada anomalía. Por último, el interfaz gráfico permite navegar por la red observando los diferentes nodos y enlaces.

Los métodos descritos hasta ahora intentan explicar en qué medida la distribución de probabilidad de un nodo se ve afectada por las probabilidades de sus vecinos; es lo que se conoce como *nivel micro* [34]. En cambio, el *nivel macro* rastrea las principales líneas de razonamiento a lo largo de la red.

Por ejemplo, el sistema NESTOR, de Cooper [5, 6] (véase la sec. 2), ofrecía dos posibilidades de macro-explicación. Una de ellas consistía en mostrar, en forma de texto, las cadenas de enlaces que relacionaban una hipótesis con los hallazgos. La otra mostraba numérica y gráficamente cómo se iban modificando las probabilidades de dos hipótesis seleccionadas a medida que se introducían la evidencia.

En la tesis de Suermondt [41], dirigida por Cooper, se presenta la metodología INSITE, que consiste en medir la influencia de los hallazgos sobre cada nodo de la red, con la posibilidad de examinar con más detalle ciertas cadenas de nodos. 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 [13] se basa en el concepto de *escenario*, definido como la asignación de valores para las variables (discretas) de un subconjunto. El algoritmo de explicación selecciona las variables relevantes y halla los escenarios más probables; después ordena las variables y las une mediante frases, generando así una historia causal de cómo se han producido los hechos [13, sec. 8.3.3].

El otro método que proponen Druzdzel y Henrion [13], se basa en las *redes cualitativas* de Wellman, que, en vez de utilizar información numérica, sólo consideran si la influencia de un nodo sobre otro es 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 [44]. Examinando el impacto de la propagación de evidencia entre nodos vecinos, se puede generar 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: abductiva, deductiva-predictiva o intercausal (lo explicamos en la próxima sección) .

5 Conclusión

5.1 Ventajas de los MGP

La principal ventaja de los MGP frente a métodos alternativos para el tratamiento de la incertidumbre es su **fundamento normativo**, es decir, que se basan en una teoría matemática que indica qué probabilidades son necesarias, cómo deben obtenerse y cómo han de combinarse.¹²

Por tanto, los MGP 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 métodos de inferencia difusa. La única justificación posible para estos métodos es construir un sistema y ver que funciona correctamente. Sin embargo, puede repetirse el caso de MYCIN, cuya evaluación resultó completamente satisfactoria, a pesar de que tenía graves inconsistencias, que habrían quedado en evidencia si se hubieran escogido los casos de prueba oportunos.

¹²Aunque muchos de los principales partidarios de las redes bayesianas defienden la interpretación subjetivista de la probabilidad, nuestra postura personal es que en el campo de la medicina es posible y por tanto deseable utilizar probabilidades objetivas procedentes de estudios epidemiológicos [40]. 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 intentando obtener *estimaciones subjetivas* de *magnitudes objetivas*. Aunque debate sobre la interpretación de la probabilidad es bastante complejo, afortunadamente todos los que trabajamos en el campo de los MGP estamos de acuerdo en los principios axiomáticos y en la forma de construir las redes, aunque las interpretaciones filosóficas sean diferentes.

Otra ventaja de prácticamente todos los MGP desarrollados hasta la fecha es que emplean **razonamiento causal**, lo cual permite tres tipos de razonamiento: *abductivo* (de los efectos a las causas), *deductivo-predictivo* (de las causas a los efectos) e *intercausal* (entre dos causas de un mismo efecto). Dicho de otro modo: los MGP, por su fundamento normativo, son capaces de obtener *todas y únicamente* las inferencias que están justificadas. En cambio, los sistemas basados en reglas (ya utilicen factores de certeza o lógica difusa) sólo admiten un tipo de inferencia, de los hallazgos hacia las hipótesis, sin tener en cuenta la distinción entre estos tres distintos tipos de razonamiento, lo cual puede dar lugar a serias inconsistencias (cf. [10, sec. 2.4] y las referencias que contiene).

Esta capacidad de los MGP es consecuencia directa del **tratamiento explícito de las dependencias e independencias condicionales**. Paradójicamente, el argumento que con más frecuencia se utilizó en torno a los años 70 para negar un papel relevante a los modelos probabilistas en la inteligencia artificial, especialmente en aplicaciones médicas, era que incluían hipótesis injustificadas de independencia condicional. La situación se invirtió a partir de los trabajos de Pearl, Lauritzen, Spiegelhalter y otros, que demostraron que las redes bayesianas podían representar correctamente las relaciones de dependencia e independencia, y por otro lado, los trabajos de Heckerman, Horvitz y otros, que demostraron que los métodos basados en reglas contenían hipótesis de independencia condicional más estrictas y más difíciles de justificar —por no decir que eran generalmente falsas— que las contenidas en los MGP.

5.2 Limitaciones de los MGP en medicina

Uno de los inconvenientes principales de los MGP es que los algoritmos actuales tienen **complejidad exponencial** para redes generales; de hecho, la computación de la probabilidad en redes bayesianas y, por consiguiente, la evaluación de diagramas de influencia, es un problema NP-difícil tanto para los métodos exactos como para los aproximados, según se ha comentado en capítulos anteriores. Sin embargo, algunos de los algoritmos actuales son capaces de computar modelos médicos bastante complejos en intervalos de tiempo razonables, e incluso existen aproximaciones y modelos simplificados que permiten abordar problemas de mayor tamaño.

Otro de los obstáculos para la aplicación de los MGP a problemas médicos es la dificultad para **construir las redes**. No vamos a insistir más en ello porque ya hemos comentado en la sección 3.1 tanto la escasez de bases de datos completas como las carencias de conocimiento que dificultan la construcción con ayuda de expertos humanos. Sin embargo, conviene señalar que ésta no es una deficiencia de los MGP, sino una dificultad intrínseca de los problemas que estamos abordando. Hay otros métodos que no requieren tantos parámetros, incluso hay técnicas puramente cualitativas, pero en nuestra opinión, estas alternativas no aportan ninguna ventaja, sino que simplemente prescinden de información que resulta difícil de obtener, pero que es esencial.

Por otro lado, aunque hemos criticado anteriormente los sistemas basados en reglas, debemos reconocer que tienen una ventaja importante frente a los MGP —al menos en la actualidad— que es la facilidad para **controlar el razonamiento**, fijando objetivos y generando las preguntas oportunas. En teoría, los MGP tendrían ventaja en este punto, pues la teoría de la información y de la decisión permiten determinar exactamente cuál es la relación coste-efectividad de cada nuevo hallazgo; sin embargo, los algoritmos generales son impracticables por la desorbitada cantidad de tiempo que necesitarían. Existen métodos que introducen hipótesis simplificadoras con el fin de seleccionar las preguntas adecuadas, pero aún no hay una metodología sólidamente establecida. De hecho, según nuestros conocimientos, los únicos sistemas comerciales que utilizan estos métodos son PATHFINDER [18] y los modelos de diagnóstico que Microsoft incorpora en Windows'95 y Windows'98.

Finalmente, otra limitación de los MGP es la dificultad para **explicar el razonamiento**, pues los métodos y modelos presentados en la sección 4.3 están aún lejos de ofrecer explicaciones comprensibles y satisfactorias para los médicos que pudieran utilizar los MGP desarrollados actualmente.

Nota. En la bibliografía hemos incluido solamente aquéllas referencias relativas a los MGP; el resto puede encontrarse en [10] (las relativas a las funciones de utilidad en medicina están en [22]). Muchas de ellas aparecen también en [27] o en [18].

References

- [1] 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.
- [2] 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.
- [3] I. A. Beinlich, H. J. Suermondt, R. M. Chávez y G. F. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. En: *Proceedings of the 2nd European Conference on AI and Medicine*, págs. 247–256, London, 1989. Springer-Verlag, Berlin.
- [4] 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.
- [5] G. F. Cooper. *NESTOR: A Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. Tesis doctoral, Dept. Computer Science, Stanford University, STAN-CS-84-1031, 1984.
- [6] G. F. Cooper. A diagnostic method that uses causal knowledge and linear programming in the application of Bayes' formula. *Computer Methods and Programs in Biomedicine*, **22**:223–237, 1986.
- [7] 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.
- [8] 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.
- [9] F. J. Díez. Parameter adjustment in Bayes networks. The generalized noisy OR-gate. En: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, págs. 99–105, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.
- [10] F. J. Díez. *Sistema Experto Bayesiano para Ecocardiografía*. Tesis doctoral, Dpto. Informática y Automática, UNED, Madrid, 1994.
- [11] F. J. Díez. Causality and probabilistic independence in graphical models. En: *Proceedings of EURO XV - INFORMS XXXIV*, Barcelona, 1997. Las transparencias pueden obtenerse en <ftp://ftp.dia.uned.es/pub/research/bayes-nets/slides/barcelona.ps>.
- [12] F. J. Díez, J. Mira, E. Iturralde y S. Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, **10**:59–73, 1997.
- [13] M. J. Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense*. Tesis doctoral, Dept. Engineering and Public Policy, Carnegie Mellon University, 1993.
- [14] G. A. Gorry. Computer-assisted clinical decision making. *Methods of Information in Medicine*, **12**:45–51, 1973.
- [15] G. A. Gorry y G. O. Barnett. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, **1**:490–507, 1968.

- [16] P. Haddawy, J. Jacobson y C. E. Kahn Jr. BANTER, a Bayesian network tutoring shell. *Artificial Intelligence in Medicine*, **10**:177–200, 1997.
- [17] P. Haddawy, C. E. Kahn Jr. y M. Butarbutar. A Bayesian network model for radiological diagnosis and procedure selection: Work-up of suspected gallbladder disease. *Medical Physics*, **21**:1185–1192, 1994.
- [18] D. E. Heckerman. *Probabilistic Similarity Networks*. Tesis doctoral, Dept. Computer Science, Stanford University, STAN-CS-90-1316, 1990.
- [19] M. E. Hernando, E. J. Gómez, F. del Pozo y R. Corcoy. DIABNET: A qualitative model-based advisory system for therapy planning in gestational diabetes. *Medical Informatics*, **21**:359–374, 1996.
- [20] 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.
- [21] C. E. Kahn Jr., L. M. Roberts, K. A. Shaffer y P. Haddawy. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*, **27**:19–29, 1997.
- [22] P. Juez Martel y F. J. Díez Vegas. *Probabilidad y Estadística en Medicina. Aplicaciones en la Práctica Clínica y en la Gestión Sanitaria*. Ed. Díaz de Santos, Madrid, 1996.
- [23] 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.
- [24] H. P. Lehmann y E. H. Shortliffe. THOMAS: building Bayesian statistical expert systems to aid in clinical decision making. *Computer Methods and Programs in Biomedicine*, **35**:251–260, 1991.
- [25] Y. C. Li. *Automated Probabilistic Transformation of a Large Medical Diagnostic Support System*. Tesis doctoral, Dept. of Medical Informatics, School of Medicine, University of Utah, 1995.
- [26] K. G. Olesen, U. Kjærulff, F. Jensen, F. V. Jensen, B. Falck, S. Andreassen y S. K. Andersen. A MUNIN network for the median nerve. A case study on loops. *Applied Artificial Intelligence*, **3**:385–403, 1989.
- [27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988. Reimpreso con correcciones en 1991.
- [28] G. Provan. Abstraction in belief networks: The role of intermediate states in diagnostic reasoning. En: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI'95)*, págs. 464–471, Montreal, 1995. Morgan Kaufmann, San Francisco, CA.
- [29] G. M. Provan y J. R. Clarke. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 1993.
- [30] 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.
- [31] S. Ríos-Insua, M. C. Bielza, M. Gómez, M. Fernández del Pozo, J. A. Sánchez Luna y S. Caballero. An intelligent decision system for jaundice management in newborn babies. En: F. J. Girón (ed.), *Cases Studies in Decision Analysis*. Springer-Verlag, Berlin, 1998. En prensa.
- [32] G. W. Rutledge, S. K. Andersen, J. X. Polaschek y L. M. Fagan. A belief network model for interpretation of ICU data. En: *Proceedings of the Fourteenth Annual Symposium of Computer Applications in Medical Care*, Washington, DC, 1990.

- [33] 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.
- [34] 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.
- [35] M. A. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. J. Horvitz, H. P. Lehmann y G. F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. Part I — The probabilistic model and inference algorithms. *Methods of Information in Medicine*, **30**:241–255, 1991.
- [36] B. Sierra y P. Larrañaga. Predicting the survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches. *Artificial Intelligence in Medicine*, 1998. En prensa.
- [37] 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.
- [38] 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.
- [39] L. E. Súcar y D. F. Gillies. Probabilistic reasoning in high-level vision. *Image and Vision Computing*, **12**:42–60, 1994.
- [40] L. E. Súcar, D. F. Gillies y D. A. Gillies. Objective probabilities in expert systems. *Artificial Intelligence*, **61**:187–208, 1993.
- [41] H. J. Suermondt. *Explanation in Bayesian Belief Networks*. Tesis doctoral, Dept. Computer Science, Stanford University, STAN-CS-92-1417, 1992.
- [42] 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.
- [43] H. R. Warner, A. F. Toronto, L. G. Veasy y R. Stephenson. A mathematical approach to medical diagnosis: Application to congenital heart disease. *Journal of the American Medical Association*, **177**:177–183, 1961.
- [44] M. P. Wellman y M. Henrion. Qualitative intercausal relations, or explaining “explaining away”. En: J. A. Allen, R. Fikes y E. Sandewall (eds.), *Principles of Knowledge Representation: Proceedings of the Second International Conference*, págs. 535–546, San Mateo, CA, 1991. Morgan Kaufmann.