

Práctica 3. Modelos canónicos y su implementación en GeNIe

TUTORIAL

Hemos visto que, en el caso general, el número de parámetros necesario para especificar la probabilidad condicional de un nodo crece exponencialmente con el número de padres, lo que plantea problemas tanto de obtención y almacenamiento de datos como de tiempos de computación de los algoritmos. (por ejemplo, para un nodo binario con 10 padres binarios, la tabla de probabilidad condicional tiene $2^{10} = 1024$ parámetros).

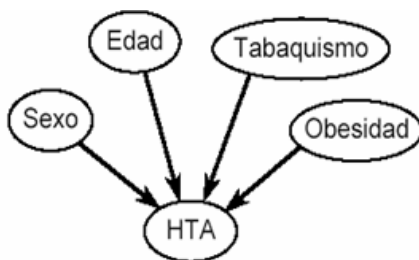
Por ello, es conveniente buscar **modelos simplificados de interacción entre variables**. Pearl los llama *modelos canónicos* por su aplicabilidad a diferentes dominios. Los más utilizados son las puertas MAX y AND ruidosas.

Analicemos primero el caso más simple: la puerta OR, que es una puerta MAX en la que todos los nodos son binarios. Para considerar una puerta OR es necesario realizar las siguientes hipótesis:

1. Cada una de las causas, por sí misma, puede producir el efecto, y basta que una de las causas produzca el efecto para que éste esté presente,
2. Cuando todas las causas están ausentes, el efecto está ausente,
3. No hay interacción entre las causas.

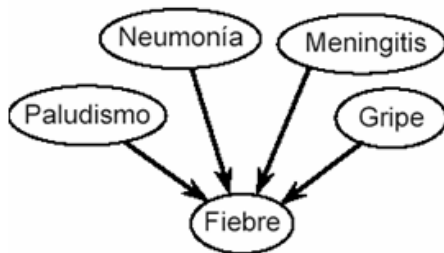
Para ilustrar situaciones en las que se dan o no se dan estas hipótesis, pensemos en los siguientes ejemplos:

Consideremos los siguientes factores, que tienen influencia causal en que una persona desarrolle hipertensión arterial (HTA):



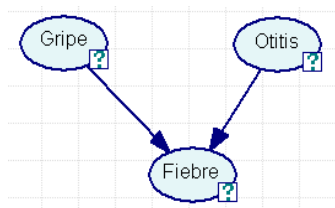
En este caso no podríamos utilizar una puerta OR, pues no se cumplen las hipótesis ya que no se cumple la hipótesis 1, ya que cada uno de los factores modelados, por sí sólo, no produce la hipertensión, sino que suponen factores de riesgo que se van acumulando.

Sin embargo, en este otro modelo, cada una de las enfermedades, por sí sola, desencadena el efecto:



En este caso sí podríamos utilizar la puerta OR, pues se cumplen las hipótesis.

Estudiaremos este modelo de interacción causal a través de un ejemplo sencillo. Supongamos la siguiente red bayesiana:



Bajo estas hipótesis, para construir las probabilidades necesarias para el modelo bastaría con dar las probabilidades de que cada una de las causas provoque el efecto por separado (que denotaremos por c_x). Sea por ejemplo:

$$c_g = P(+f/+g)=0.8$$

$$c_o = P(+f/+o)=0.6$$

Y en ese caso tendríamos que:

$$P(+f/+g, +o) = 0.8 + 0.2*0.6 = 0.92^1$$

$$P(+f/\neg g, +o) = 0.6$$

$$P(+f/+g, \neg o) = 0.8$$

$$P(+f/\neg g, \neg o) = 0$$

Supongamos que en el modelo queremos incluir que es posible que otras causas no determinadas provoquen también fiebre, introduciendo así cierto ruido en el mismo. Bajo las hipótesis anteriormente mencionadas, es posible construir las probabilidades condicionadas necesarias en el modelo a partir de unos parámetros básicos, concretamente las probabilidades de que cada causa provoque el efecto por separado y un factor de ruido r , que expresa que otras condiciones no presentes en el modelo podrían provocar la fiebre. Para continuar con el ejemplo, demos unos valores a estos parámetros:

$$c_g = P(+f/+g, \neg o, \neg r)=0.8$$

$$c_o = P(+f/\neg g, +o, \neg r)=0.6$$

$$r = P(+f/\neg g, \neg o)=0.01$$

Veamos como se calculan en este caso las probabilidades condicionadas necesarias. En primer lugar, tenemos que por la tercera hipótesis:

$$P(+f/+g, +o, \neg r) = P(+f/+g, \neg o, \neg r) + P(\neg f/+g, \neg o, \neg r) P(+f/\neg g, +o, \neg r) = 0.8 + 0.2*0.6 = 0.92^2$$

Por la segunda hipótesis:

$$P(+f/\neg g, \neg o, \neg r) = 0$$

Las otras probabilidades se calculan mediante:

$$P(+f/+g, +o) = P(+f/+g, +o, \neg r) + P(\neg f/+g, +o, \neg r) P(+f/\neg g, \neg o) = 0.92 + 0.08*0.01 = 0.9208^3$$

$$P(+f/+g, \neg o) = P(+f/+g, \neg o, \neg r) + P(\neg f/+g, \neg o, \neg r) P(+f/\neg g, \neg o) = 0.8 + 0.2*0.01 = 0.802$$

$$P(+f/\neg g, +o) = P(+f/\neg g, +o, \neg r) + P(\neg f/\neg g, +o, \neg r) P(+f/\neg g, \neg o) = 0.6 + 0.4*0.01 = 0.604$$

En general, si en un modelo tenemos un efecto X y U_1, \dots, U_n son las causas posibles, si denotamos por c_i a la probabilidad que tiene cada causa de producir el efecto por separado y por q_i a la probabilidad complementaria ($q_i = 1 - c_i$), entonces:

$$P(\neg x/u_1, \dots, u_n) = \prod_{i \in T_u} q_i$$

Donde T_u = conjunto de causas de X que están presentes. Como en el caso anterior, el modelo se puede generalizar al caso de que haya cierto ruido.

Existen otros modelos de interacción causal que también están implementados en GENIE:

- Puerta MAX: La generalización de la puerta OR al caso en que las variables son multi-valuadas. En este caso se considera que el efecto aparece a partir de cierto valor de la variable causa, y se procede de igual modo.

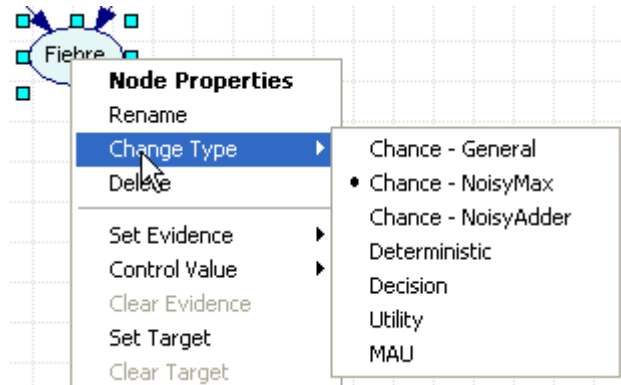
¹ el 80% tendrá fiebre porque tiene gripe, y de entre los que no tienen fiebre aunque tengan gripe, que son el 20%, un 60% tendrá fiebre (por padecer otitis).

² Es decir, la probabilidad de tener fiebre cuando tiene gripe, otitis y no actúa el ruido se puede calcular como la probabilidad de tener fiebre teniendo solo la gripe (no otitis, no ruido) + [probabilidad de no tener fiebre teniendo solo gripe (no otitis, no ruido) * probabilidad de tener fiebre por tener solo otitis (no gripe, no ruido)]

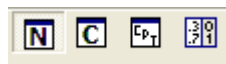
³ Probabilidad de tener fiebre dada esa combinación de causas + (probabilidad de no tener fiebre dada esa combinación * probabilidad de que actúe el ruido)

- Puerta AND: Cuando la interacción entre las variables es tal que sólo cuando dos de ellas están presentes se produce el efecto.

Para implementar una puerta OR en GENIE, sólo tenemos que definir el nodo fiebre como nodo MAX ruidoso (noisy-max), abriendo el menú contextual asociado:



Una vez abierto el menú, en la pestaña de definición del nodo tendremos nuevos botones:



El primer botón mostrará la tabla 1, que es la tabla de las probabilidades de que cada enfermedad provoque el síntoma por separado sin tener en cuenta el ruido, es decir:

| Parent State | Gripe | Otitis | LEAK |
|--------------|-------|--------|------|
| si | 0.8 | 0.6 | 0.01 |
| no | 0.2 | 0.4 | 0.99 |

El segundo botón nos muestra la tabla 2, es decir, estas mismas probabilidades pero teniendo en cuenta el ruido:

| Parent State | Gripe | Otitis | LEAK |
|--------------|-------|--------|------|
| si | 0.802 | 0.604 | 0.01 |
| no | 0.198 | 0.396 | 0.99 |

y el tercero la tabla 3, es decir, la tabla de probabilidad condicionada (Condicional Probability Table o CPT):

| Gripe | si | | no | |
|--------|--------|-------|-------|------|
| Otitis | si | no | si | no |
| si | 0.9208 | 0.802 | 0.604 | 0.01 |
| no | 0.0792 | 0.198 | 0.396 | 0.99 |

Finalmente, el cuarto botón muestra o esconde las probabilidades que no es necesario introducir por ser fijas en el modelo y que por tanto no son modificables.

TAREA Y ENTREGA

Tarea: Implementa en GeNIe una puerta OR en la que haya tres causas y un efecto. Razona la forma de construir la CPT en este caso (fórmula). Comprueba para un par de casos que los resultados coinciden con el cálculo que realiza GeNIe.

Entrega: Sube un fichero pdf en el que incluyas captura de pantalla de tu red GeNIe y de las tres tablas de probabilidad mencionadas anteriormente (tablas 1, 2 y 3) y también la demostración de que la fórmula que has aplicado te da los resultados correctos (con un par de casos es suficiente).

OPCIONAL

Si te sobra tiempo, puedes explorar en el manual de GeNIe cómo se introducen otros modelos canónicos, y añadir a tu entrega un breve informe.