CAPITULO 1

RAZONAMIENTO CATEGÓRICO Y CORRECCIÓN BAYESIANA

- **Interpretación Diferencial**
- **Elementos del Razonamiento Categórico**
- **Un Procedimiento Sistemático para el Razonamiento Categórico**
- **La Corrección Bayesiana**
- **Resumen**
- **Textos básicos**

1. RAZONAMIENTO CATEGÓRICO Y CORRECCIÓN BAYESIANA

La inteligencia artificial no sólo se ocupa de mecanismos generales relacionados con la búsqueda de soluciones en un espacio dado, o de cómo representar y utilizar el conocimiento de un determinado dominio de discurso. Otro aspecto, hasta ahora sólo esbozado en el capítulo anterior, es el que corresponde a los mecanismos y/o procesos inferenciales, que consideraremos como el punto de partida de los llamados *modelos de razonamiento*.

En cualquier dominio, la propagación del conocimiento[44] por medio de programas de IA se efectúa siempre siguiendo un modelo de razonamiento bien definido. Estos modelos de razonamiento forman parte del motor de inferencias, si hablamos de sistemas de producción, o de las estructuras de control del conocimiento, si hablamos de cualquier otro tipo de sistemas de IA, y contribuyen de manera decisiva a organizar correctamente la búsqueda de soluciones.

Normalmente, las características del dominio, y las características de los problemas que deben resolverse, condicionan el tipo de modelo de razonamiento que debemos emplear. Así:

- Hay dominios de naturaleza marcadamente simbólica, en los que las soluciones pueden establecerse “con total seguridad”. En estos casos emplearemos modelos categóricos de razonamiento.

- Por otra parte, hay dominios de naturaleza estadística, en los que las soluciones no pueden ser unívocamente obtenidas y en los que, además, tendremos que averiguar cuál de las posibles soluciones encontradas es la más probable. En estos casos preferiremos razonar con modelos de naturaleza estadística, de los cuales, dadas las peculiaridades de los procesos inferenciales que trata la IA, el esquema bayesiano es el más utilizado.

- Hay otros dominios en los que aparece el concepto de *incertidumbre*, que puede ser inherente a los datos del problema y a los hechos del dominio, o a los propios mecanismos inferenciales. En estos casos elegiremos modelos de razonamiento que sean capaces de manipular correctamente dicha incertidumbre.

- Por último[45], hay dominios en los que los elementos inferenciales incluyen matices de carácter lingüístico, entre los que pueden establecerse jerarquías y

---

[44] O lo que es lo mismo, el establecimiento de circuitos inferenciales apropiados.

[45] Aunque hay más tipos de dominios diferentes, los modelos de razonamiento que desarrollaremos bastarán para poder abordar gran cantidad de problemas interesantes. Más aún si tenemos en cuenta el carácter introductorio de este texto.

clasificaciones. En estos casos es conveniente emplear modelos de razonamiento basados en *conjuntos difusos*[46].

Como es obvio, la clasificación que acabamos de establecer es sólo eso, una clasificación, y habrá dominios que participen de varias de las características mencionadas. En estos casos podrá optarse por la combinación de diferentes modelos, o por la implementación del modelo de razonamiento que más se ajuste a las características del dominio.

1.1. Interpretación Diferencial

Una de las grandes cuestiones en la resolución de problemas de inteligencia artificial es cómo utilizar los datos y las verdades demostradas, según un procedimiento encadenado y lógico, al objeto de poder discriminar entre las posibles "soluciones", inicialmente candidatas, hasta encontrar la verdadera respuesta del problema planteado.

Cuando el dominio es de naturaleza simbólica, ya se ha comentado que el proceso de razonamiento adecuado debe seguir una aproximación categórica. Uno de tales procedimientos categóricos es el de la *interpretación diferencial*. Supongamos que a un experto de un universo de discurso concreto le preguntamos:

-¿Cómo interpretaría usted estos datos y esta información en este contexto?-

Una posible respuesta del experto a nuestra pregunta podría ser la siguiente:

"En primer lugar analizo los datos disponibles tratando de evaluar la importancia relativa de los mismos. Parte de la información puede ser de vital importancia. El resto de la información puede ser simplemente una consecuencia menor del problema principal, o puede estar relacionada con el problema, pero ser... digamos, información de segundo orden. A continuación, una vez clasificada la información de partida, trato de establecer un conjunto de posibles interpretaciones que sean compatibles con los datos iniciales. Una vez establecido este conjunto inicial de hipótesis realizo un análisis más profundo de los datos, y busco información complementaria que me permita ir descartando una a una las hipótesis iniciales. Este proceso lo continúo hasta que finalmente encuentro la solución. Pero a veces mi conjunto inicial de hipótesis está mal planteado, o mi conocimiento sobre el caso a veces no es completo, lo que se suele traducir en que no soy capaz de resolver el problema; es decir, me quedo sin hipótesis. Otras veces, por el contrario, el problema está en los datos, que no son suficientes. En estos casos suelo encontrarme con más de una hipótesis que es perfectamente compatible con la información de partida y con mi conocimiento sobre el tema. En otras palabras... Soy capaz de "acotar" la solución, pero al no ser única, mi interpretación puede no bastar, puede no decidir...-"

Nuestro hipotético experto, que es de una modestia más que notable, nos acaba de describir un caso típico de interpretación diferencial que nos va a servir para ilustrar

---

[46] O borrosos - *fuzzy* en inglés-.

el modelo de razonamiento categórico, sus ventajas, sus inconvenientes, y el verdadero alcance de este procedimiento en el contexto de los procesos inferenciales que trata la IA.

Tratemos de formalizar este complejo proceso de razonamiento que nos ha descrito nuestro hipotético experto. En primer lugar, nuestro experto ha tenido que reunir todo un conjunto de información *relevante*. Luego ha efectuado una ponderación de la importancia relativa de dicha información. A continuación ha tratado de relacionar la información disponible con un conjunto de interpretaciones inicialmente posibles. Finalmente ha ido eliminando de su "lista" inicial de hipótesis todas aquellas que no se podían concatenar lógicamente con los datos. Por último, tras el proceso de eliminación de hipótesis potencialmente relevantes, nuestro experto se ha dado cuenta de que: (a) ha encontrado la solución del problema, o (b) el conjunto de hipótesis formulado inicialmente no es consistente con los datos, o (c) hay varias hipótesis que se corresponden con los datos, entre las cuales no es posible discriminar. Así, el proceso global podría seguir el siguiente esquema:

- Recopilación de información.

- Análisis de la importancia relativa de las manifestaciones del problema.

- Análisis de las posibles causas del problema tras considerar, conjunta y razonablemente, todas las manifestaciones del problema. Ello implica el establecimiento tentativo de relaciones causa-efecto.

- Exclusión una a una de todas aquellas interpretaciones que no pueden ser explicadas completa y razonablemente por los datos.

- Fin del proceso con alguno de los siguientes resultados:

    - Existe una única solución

    - No hay ninguna solución

    - Hay varias soluciones posibles entre las que no se puede discriminar.

Este proceso de razonamiento -sistemático pero complejo- puede simplificarse en función del verdadero grado de experiencia del experto. Así, muchos pasos pueden ser claramente innecesarios para un profesional consolidado, que aplicará su "sentido común" y su conocimiento heurístico para restringir al máximo su conjunto inicial de hipótesis. También, su intuición, matizada y depurada a través de largos años de profesión, llevan al experto a efectuar el proceso de establecimiento de relaciones causa-efecto de una manera eficaz y eficiente.

El novato (o simplemente, el "menos experto"), tiende a aplicar metodologías completas que le permitan resolver adecuadamente el problema. El resultado final puede ser el mismo, incluso puede ser correcto, pero el verdadero experto suele optimizar mucho más sus recursos que el novato. En resumen, nosotros asumiremos el papel de novatos y trataremos de definir un modelo de razonamiento que nos permita resolver problemas de naturaleza categórica. Exigiremos además que nuestro modelo:

- Sea sistemático
- Utilice algún sucedáneo del "sentido común"
- Incorpore algo parecido a la "intuición"

1.2. Elementos del Razonamiento Categórico

Puesto que una de las tareas que hay que realizar en un proceso de razonamiento categórico es la de establecer un conjunto de relaciones causales[47], comenzaremos describiendo nuestro dominio de discurso a partir de dos entidades diferentes, entre las cuales deberemos ser capaces de establecer relaciones. Estas entidades son:

- las manifestaciones posibles en el dominio de discurso
- las interpretaciones posibles en el dominio de discurso

Cualquier dominio estará perfectamente descrito cuando hayamos especificado:

- Todas las posibles manifestaciones de los problemas que puedan darse en el dominio.
- Todos los posibles problemas del dominio.
- Todas las relaciones causales que puedan establecerse entre problemas y manifestaciones.

Dicho de este modo la cuestión puede parecer algo abstracta. Trataremos ahora de clarificar todo este embrollo con un ejemplo: Supongamos que queremos construir un sistema inteligente para efectuar predicciones meteorológicas. En este caso, el dominio será el de las PREDICCIONES METEOROLOGICAS. El conjunto de manifestaciones podría estar formado por una serie de observaciones como *el color del cielo, la presencia de nubes, el grado de humedad,...* El conjunto de interpretaciones podría estar formado por hipótesis como *posibilidad de lluvia, posibilidad de tormentas, posibilidad de vientos fuertes,...* Finalmente, las relaciones causales serán las que nos permitan inferir interpretaciones a partir de manifestaciones. Por ejemplo: *Si el crepúsculo es de color rojizo, y no hay nubes en el cielo, entonces el pronóstico es de buen tiempo.*

Formalmente, para construir el modelo necesitamos definir una serie de funciones de carácter booleano[48] que nos sirvan para describir el dominio. Si $x_1,..., x_n$ es el conjunto completo de todas las manifestaciones posibles del universo de discurso, entonces la función $f(x_1,..., x_n)$, de carácter booleano, le asignará el valor "1" a la manifestación x_i, si la manifestación x_i está presente en nuestro problema concreto, o le asignará el valor "0", si dicha manifestación está ausente.

---

[47] Es decir, relaciones causa-efecto.

[48] Las funciones son de carácter booleano puesto que el modelo es categórico. En tales modelos, algo está presente o está ausente, algo es posible o es imposible.

Del mismo modo, si $y_1,..., y_m$ representa el conjunto completo de todas las posibles interpretaciones que se pueden dar a los problemas del dominio, para un problema concreto la función $g(y_1,..., y_m)$ le asignará el valor "1" a la interpretación y_j si la interpretación y_j es posible[49], o le asignará el valor "0" si dicha interpretación es imposible.

Necesitaremos una tercera función, que denominaremos función E que representa el conjunto de todas las posibles relaciones causales que se pueden establecer en nuestro dominio de discurso, entre manifestaciones e interpretaciones. La función E es la *función de conocimiento*. Con estos tres elementos nuestro problema se reduce a encontrar, ante un conjunto de manifestaciones relacionadas con un problema, en un dominio, la función g que satisface:

$E: (f \rightarrow g)$

Expresión que podemos leer del siguiente modo: "... encontrar el conjunto de interpretaciones que es compatible con las observaciones y datos de que disponemos, tras la aplicación de nuestro conocimiento sobre el dominio de discurso..."

Claramente, puesto que E es la función de conocimiento que nos permite establecer relaciones de tipo causal entre manifestaciones e interpretaciones, su expresión formal será del tipo:

$E\ (x_1, ..., x_n, y_1, ..., y_m)$

Para ilustrar cómo podemos utilizar estas tres funciones en un modelo categórico de razonamiento trataremos de resolver un ejemplo sencillo:

Sea un dominio D en el que podemos identificar dos posibles manifestaciones que están relacionadas con dos posibles interpretaciones. Llamaremos:

- D = Dominio de discurso
- M = Conjunto de manifestaciones = [m(1), m(2)]
- I = Conjunto de interpretaciones = [i(1), i(2)]

Supongamos que, tras estudiar el dominio, hemos sido capaces de identificar el siguiente conocimiento:

- Para que la interpretación i(2) sea cierta, la manifestación m(1) debe estar presente.

- Para que la interpretación i(1) sea cierta y, al mismo tiempo, podamos descartar la interpretación i(2), entonces la manifestación m(2) debe estar presente.

- Para que la interpretación i(2) sea cierta y, al mismo tiempo, podamos descartar la interpretación i(1), entonces la manifestación m(2) no debe estar presente.

---

[49] ... a la vista de las manifestaciones del problema.

- Si alguna de las manifestaciones está presente es porque alguna de las interpretaciones puede establecerse.

Con las declaraciones anteriores, que expresan en lenguaje natural nuestro conocimiento sobre el dominio, vamos a construir la función E:

$$E = \{ \begin{array}{lcll} [\, i(2) & \rightarrow & m(1)] & \text{and} \\ [\, i(1) * \neg i(2) & \rightarrow & m(2)] & \text{and} \\ [\, \neg i(1) * i(2) & \rightarrow & \neg m(2)] & \text{and} \\ [\, m(1) + m(2) & \rightarrow & i(1) + i(2)] & \end{array} \}$$

En esta expresión, los símbolos "*" y "+" deben leerse, respectivamente como "and" y como "or inclusivo".

Una vez hemos sido capaces de formalizar nuestro conocimiento, trataremos de resolver un problema que presenta las siguientes manifestaciones:

- la manifestación $m(2)$ está presente en el problema
- hay evidencia total de que la manifestación $m(1)$ no está presente

Con esta información la función booleana f será:

$$f = \neg m(1) * m(2)$$

Nótese que, según nuestro planteamiento, la función f debe incluir toda la evidencia positiva y toda la evidencia negativa. Así, dicha función sería diferente si no dispusiéramos de información acerca de una manifestación dada[50].

Según hemos mencionado ya, nuestro problema de razonamiento es encontrar la función g que verifique:

$$E: (f \rightarrow g)$$

En lenguaje natural, la pregunta que deberíamos ser capaces de responder es la siguiente: ¿Qué "conjunto" de interpretaciones verifica simultáneamente la ausencia de la manifestación $m(1)$ y la presencia de la manifestación $m(2)$?

En este sencillo ejemplo es fácil ver que la solución es[51]:

$$g = i(1) * \neg i(2)$$

En este caso la aplicación de los métodos y procedimientos lógicos habituales conducen rápidamente a la solución. La situación, no obstante, cambia drásticamente si

[50] Más concretamente, si cambiamos la declaración (b) de nuestro ejemplo por esta otra declaración: -No tenemos información acerca de $m(1)$-, entonces la función f sería: $f = m(2)$

[51] Los autores animan al lector a que traten de resolver por los procedimientos "clásicos" el problema planteado.

consideramos -pongamos por caso- un dominio en el que fuesen posibles 30 manifestaciones, relacionadas con 600 posibles interpretaciones a través de una función E en la que se resumen 125 relaciones causales[52]. En este contraejemplo, los procedimientos lógicos clasicos, la complejidad de los procesos de resolución, y la necesidad de efectuar muchas sustituciones, podrían hacer inviable la resolución del problema. Aparece así la necesidad de encontrar alternativas mejores, conceptualmente correctas, y computacionalmente eficaces. Una de tales alternativas es la que se describe a continuación.

1.3. Un Procedimiento Sistemático para el Razonamiento Categórico

Vamos a plantearnos la cuestión anterior desde una perspectiva algo diferente. Decíamos que, para un dominio concreto, habíamos sido capaces de encontrar todas las posibles manifestaciones de los problemas que pudieran aparecer, y que estas manifestaciones estaban relacionadas con un conjunto de interpretaciones. Por otra parte, las relaciones entre manifestaciones e interpretaciones se establecían a través de nuestro conocimiento sobre el dominio. En este contexto, el procedimiento sistemático que proponemos para razonar categóricamente consta de las siguientes fases:

- construcción del conjunto de todas las combinaciones que se pueden establecer entre las manifestaciones del dominio. A este conjunto lo denominamos *conjunto de complejos de manifestaciones*.

- construcción del conjunto de todas las combinaciones que se pueden establecer entre las interpretaciones del dominio. A este conjunto lo denominamos *conjunto de complejos de interpretaciones*.

- construcción del conjunto completo de todas las combinaciones posibles entre complejos de manifestaciones y complejos de interpretaciones. A este conjunto lo denominamos *conjunto de complejos manifestación-interpretación*.

De este modo, si en el dominio considerado hemos identificado n manifestaciones (i.e., $m(1)$, ..., $m(n)$) y t interpretaciones (i.e., $i(1)$, ..., $i(t)$), entonces el número de complejos de manifestaciones será 2^n, el número de complejos de interpretaciones será 2^t, y el número de complejos manifestación-interpretación será $2^{(n+t)}$.

Dado el carácter exhaustivo del procedimiento, los tres conjuntos que acabamos de definir tienen las siguientes propiedades:

- Sus elementos son mutuamente excluyentes
- Son conjuntos completos

El conjunto de complejos manifestación-interpretación representa el total de situaciones idealmente posibles en nuestro universo de discurso, pero es evidente que

---

[52] En cualquier caso, un dominio relativamente reducido.

no todas ellas van a poder darse en la realidad. Es más, muchas de ellas serán claramente absurdas. El papel del conocimiento será el de restringir el conjunto total de situaciones idealmente posibles a un conjunto de situaciones realmente posibles (i.e., permitidas por el conocimiento disponible).

Para ilustrar mejor el procedimiento, aplicaremos este esquema al problema que nos ha servido de ejemplo en la sección anterior, y en el que habíamos identificado dos posibles manifestaciones, $m(1)$ y $m(2)$, posiblemente relacionadas con dos interpretaciones, $i(1)$ e $i(2)$, a través del siguiente conocimiento:

(a)	$i(2)$	$\rightarrow$	$m(1)$
(b)	$i(1) * \neg i(2)$	$\rightarrow$	$m(2)$
(c)	$\neg i(1) * i(2)$	$\rightarrow$	$\neg m(2)$
(d)	$m(1) + m(2)$	$\rightarrow$	$i(1) + i(2)$

de forma que E = [(a) and (b) and (c) and (d)]

Si M es el conjunto de complejos de manifestaciones, entonces, en función de los valores lógicos de cada manifestación particular, y de acuerdo con el criterio que se muestra a continuación:

$m(1)$	0	0	1	1
$m(2)$	0	1	0	1
	m_1	m_2	m_3	m_4

el conjunto M estará formado por los siguientes elementos:

$M = [m_1, m_2, m_3, m_4]$, donde:

$m_1 = \neg m(1) * \neg m(2)$ -primera columna-
$m_2 = \neg m(1) * m(2)$ -segunda columna-
$m_3 = m(1) * \neg m(2)$ -tercera columna-
$m_4 = m(1) * m(2)$ -cuarta columna-

Nótese que los valores de los complejos de M dependen del criterio elegido para asignar valores lógicos a las manifestaciones individuales. El resultado hubiese sido distinto si decidiésemos emplear el siguiente criterio:

$m(1)$	0	1	0	1
$m(2)$	0	0	1	1
	m_1	m_2	m_3	m_4

Obsérvese también que las manifestaciones de un problema concreto del dominio vendrán representadas exclusivamente por uno cualquiera de los complejos de M[53].

---

[53] Evidentemente, siempre que todos los datos sean conocidos.

Análogamente, si I es el conjunto de complejos de interpretaciones, entonces, en función de los valores lógicos de cada interpretación particular, y de acuerdo con el criterio que se muestra a continuación:

$i(1)$	0	0	1	1
$i(2)$	0	1	0	1
	i_1	i_2	i_3	i_4

el conjunto I estará formado por los siguientes elementos:

$I = [i_1, i_2, i_3, i_4]$, donde:

$i_1 = \neg i(1) * \neg i(2)$ -primera columna-
$i_2 = \neg i(1) * i(2)$ -segunda columna-
$i_3 = i(1) * \neg i(2)$ -tercera columna-
$i_4 = i(1) * i(2)$ -cuarta columna-

A continuación debemos construir el conjunto completo de todas las posibles combinaciones entre complejos de manifestaciones y complejos de interpretaciones:

	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
$m(1)$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
$m(2)$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$i(1)$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
$i(2)$	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i_1				i_2				i_3				i_4			

El resultado es una lista exhaustiva de complejos manifestación-interpretación, que se denomina *base lógica expandida*, y que -en este caso- estará formada por 16 complejos siguientes[54]:

$$\text{Base Lógica Expandida} = \text{BLE} = [\begin{array}{llll} m_1i_1, & m_2i_1, & m_3i_1, & m_4i_1, \\ m_1i_2, & m_2i_2, & m_3i_2, & m_4i_2, \\ m_1i_3, & m_2i_3, & m_3i_3, & m_4i_3, \\ m_1i_4, & m_2i_4, & m_3i_4, & m_4i_4, \end{array}]$$

Pero no todos los complejos manifestación-interpretación de la base lógica expandida van a ser “realmente” posibles en el dominio de discurso. Por el contrario, muchos de tales complejos van a estar prohibidos por el conocimiento y, por lo tanto, habrá que descartarlos. Así, la aplicación de la función de conocimiento E sobre la base lógica expandida nos genera la llamada “Base Lógica Reducida” -BLR-, en la que sólo figuran aquellos complejos manifestación-interpretación que son compatibles con el conocimiento que se tiene sobre el dominio en cuestión:

E: BLE $\rightarrow$ BLR

[54] Recuerde que para n manifestaciones y t interpretaciones, el número máximo de complejos manifestación-interpretación es $2^{(n+t)}$. Aquí tenemos 2 manifestaciones y 2 interpretaciones, por lo que hay 16 complejos.

De este modo, la declaración (a) de la función E del ejemplo que estamos considerando, elimina de la BLE los complejos:

- m_1i_2
- m_1i_4
- m_2i_2
- m_2i_4

Esta eliminación se explica considerando que, con independencia de lo que pueda pasar con $i(1)$, para que $i(2)$ sea cierta (i.e., en términos de complejos: i_2 ó i_4), la manifestación $m(1)$ debe estar presente. De ello se deduce inmediatamente que m_1 y m_2 no tienen sentido en el contexto de i_2 y de i_4.

Siguiendo el mismo razonamiento, la declaración (b) del ejemplo elimina los complejos:

- m_1i_3
- m_3i_3

Análogamente, la declaración (c) elimina los complejos:

- m_2i_2 (ya eliminado en un paso anterior)
- m_4i_2

Por último, la declaración (d) elimina los complejos:

- m_2i_1
- m_3i_1
- m_4i_1

Al final, la aplicación del conocimiento del dominio sobre la BLE se traduce en la siguiente BLR:

$$\text{BLR} = [m_1i_1, m_3i_2, m_2i_3, m_4i_3, m_3i_4, m_4i_4]$$

Si nuestro conocimiento sobre el dominio es completo, y si el dominio está descrito correctamente, la solución a cualquier problema que podamos plantearnos está en BLR.

Volvamos al problema que nos sirve de ejemplo, y en el cual la cuestión planteada es encontrar una solución para el problema representado por la función:

$f = \neg m(1) * m(2)$

Nótese que, de acuerdo con el criterio elegido para el establecimiento de los correspondientes valores lógicos:

$f = \neg m(1) * m(2) = m_2$

La cuestión hay que formularla ahora en estos términos: ¿Qué complejos de BLR contienen al complejo de manifestaciones m_2?

En este caso sólo hay uno: m_2i_3. Por lo tanto la solución al problema planteado es:

$g = i_3 = i(1) * \neg i(2)$

Esta solución es, precisamente, la que habríamos encontrado si hubiésemos resuelto el problema siguiendo un procedimiento "lógico tradicional".

Ahora somos capaces de asegurar que la interpretación $i(1)$ es absolutamente cierta, y que la interpretación $i(2)$ es falsa o imposible, conclusión a la que hemos llegado tras la aplicación sistemática de un procedimiento que es consistente con la lógica.

Supongamos ahora que revisamos nuestro conocimiento, y encontramos una nueva declaración que es cierta:

(e) $\neg i(1) * \neg i(2) \rightarrow m(1) + m(2)$

Esta nueva declaración debería ser añadida a E con lo que obtendríamos una nueva función de conocimiento: [$E' = E$ and (e)], y su aplicación sobre la BLE nos generaría la siguiente BLR:

BLR' = [m_3i_2, m_2i_3, m_4i_3, m_3i_4, m_4i_4]

ya que la declaración (e) elimina al complejo m_1i_1. ¿Qué ocurriría ahora si tuviésemos que resolver un problema con las manifestaciones: $f' = \neg m(1)$ and $\neg m(2)$? En este caso, $f' = m_1$, y no hay ningún complejo en BLR que contenga a dicho complejo de manifestaciones. Esta situación puede darse por una cualquiera (o todas) de las siguientes razones:

- Las manifestaciones no son realmente esas.
- El conocimiento no es correcto. Hay algún error en la función E'.
- El dominio no está bien construido.

El primer problema se resuelve efectuando una nueva recogida de datos, al objeto de comprobar que no hemos cometido errores al construir la función f'. Si el conjunto de manifestaciones sigue siendo el mismo, entonces hay que sospechar que el error está, bien en la función de conocimiento (que puede ser incompleta, demasiado restrictiva, o simplemente falsa), bien en la construcción del dominio, en el pudiera haber manifestaciones y/o posibles interpretaciones que no hayan sido tenidas en cuenta.

En este ejemplo, el lector atento se habrá dado cuenta de que la declaración (e) es contradictoria con la declaración (d), por lo que la función E' es incorrecta. Además, aún considerando que el dominio está bien construido, el nuevo problema planteado, f', indica que no hay manifestaciones, y si no hay manifestaciones no hay problemas, y si no hay problemas... ¿para qué queremos interpretar algo que no existe?[55]

Analicemos a continuación otra posibilidad que puede surgir. Supongamos ahora que el problema se manifiesta del siguiente modo:

$f'' = m(1) * \neg m(2) = m_3$

En este caso aparecen en BLR dos complejos manifestación-interpretación:

- $m_3 i_2$
- $m_3 i_4$

Por lo tanto, de la BLR derivamos que:

$g'' = i_2 + i_4 = \neg i(1) * i(2) + i(1) * i(2)$

En otras palabras, la aplicación del método nos permite afirmar que la interpretación $i(2)$ es cierta; sin embargo, no nos permite afirmar nada acerca de $i(1)$. Estamos ante uno de los problemas que nos comentaba nuestro hipotético experto: hemos podido “acotar” la solución, pero no hemos podido resolver la cuestión completamente. Como regla general, esta situación no es aceptable, y constituye uno de los problemas más serios del modelo que acabamos de desarrollar.

Otro problema importante del esquema categórico es el de la casi siempre inevitable explosión combinatoria. En efecto, con sólo 7 manifestaciones y 24 posibles interpretaciones, nuestra BLE estaría constituida por 2.147.483.600 complejos manifestación-interpretación.

Estas, y otras deficiencias más sutiles, aconsejan la puesta a punto de un modelo alternativo.

1.4. La Corrección Bayesiana

Las interpretaciones categóricas son más bien infrecuentes en el mundo real. Además:

- ¿Podemos afirmar siempre, y sin equivocarnos, que todos los problemas se manifiestan? Ya hemos comentado que no. Al contrario, hay problemas que no se manifiestan nunca, y otros que tardan mucho en manifestarse.

- ¿La presencia de una manifestación dada es siempre indicativa de algún problema? En principio, parece que si hay algún tipo de manifestación es porque hay algo que

[55] Los autores son conscientes de que esta última afirmación es peligrosa. En efecto, por ejemplo en medicina existen muchas enfermedades, algunas de ellas muy serias, que son asintomáticas, y por lo tanto no se manifiestan, o tardan mucho en manifestarse. En tales casos, este modelo podría no ser apropiado.

hace que tal manifestación se produzca, pero no siempre una manifestación dada es indicativa de lo que podríamos llamar "un problema importante"

Con estas consideraciones vamos a replantearnos la cuestión desde otro punto de vista. Para ello trataremos de encontrar una respuesta a la pregunta siguiente:

- Dado un universo y un conjunto de atributos... ¿cual es la probabilidad de que un determinado elemento del universo presente ciertos atributos del conjunto total?

En términos estadísticos, si N es una determinada población, y $x_1, x_2,..., x_n$ es el conjunto de todos los atributos posibles, la función booleana $f(x_1,..., x_n)$ genera un subconjunto de atributos. De este modo, si $N(f)$ es el subconjunto de los elementos de N que presentan tales atributos, la probabilidad total de f será:

$$P(f) = \frac{N(f)}{N}$$

El concepto de probabilidad total, sin embargo, no es suficiente para construir un modelo de razonamiento. Necesitamos introducir el concepto de *probabilidad condicional*, que la excelente enciclopedia Espasa define como "... la probabilidad de las causas".

En la probabilidad condicional aparecen involucrados dos sucesos, de forma que la ocurrencia del segundo depende de la ocurrencia del primero. Veamos un ejemplo:

Para tratar de resolver un problema concreto sabemos que podemos ejecutar dos acciones potencialmente eficaces, A ó B. Después de ejecutar una cualquiera de tales acciones, sabemos que nuestro problema puede evolucionar positivamente (E = éxito), o puede evolucionar mal (F = fracaso). En cualquier caso sabemos que la evolución del problema depende de la acción[56] y, por lo tanto, el problema no evoluciona solo. Elijamos la acción por el método de la moneda[57]. Así:

- $p(A) = 0.5$
- $p(B) = 0.5$

Sabemos además que, para este problema concreto, y después de muchos ensayos, la evolución del problema tras ejecutar la acción A fue positiva el 20% de las veces. En el caso de la acción B se obtuvo una evolución positiva el 60% de las veces. Según este planteamiento, la probabilidad "a priori" de resolver el problema es:

$$p(E) = p(E/A)p(A) + p(E/B)p(B) = (0.2 \times 0.5) + (0.6 \times 0.5) = 0.4$$

en donde:

- $p(E)$ = probabilidad total de éxito "a priori"

[56] Es decir, hay una relación causal más o menos evidente.
[57] Un método como otro cualquiera, en ocasiones muy utilizado.

- $p(A)$ = probabilidad de ejecutar la acción *A*
- $p(B)$ = probabilidad de ejecutar la acción *B*
- $p(E/A)$ = probabilidad de éxito si ejecutamos la acción *A*
- $p(E/B)$ = probabilidad de éxito si ejecutamos la acción *B*

Claramente, $p(E/A)$ y $p(E/B)$ son probabilidades condicionales "a priori", puesto que proceden de estudios previos.

Este planteamiento se acerca algo más a los aspectos relativos al *razonamiento*. Sin embargo, todavía nos falta algo. En realidad lo que a nosotros nos interesa, en principio, es interpretar cosas que ya han pasado, para lo cual necesitamos introducir algún mecanismo para razonar "a posteriori". Para ello, vamos a variar algo las premisas del ejemplo:

Sabemos que hemos tenido un problema concreto para cuya resolución podíamos ejecutar dos acciones, *A* ó *B*, cada una de las cuales llevaba asociada una determinada probabilidad de éxito. Sabemos que, efectivamente, una de tales acciones fue ejecutada, a consecuencia de lo cual el problema evolucionó positivamente. ¿Cuál es la probabilidad de que la acción ejecutada haya sido la acción *A*?

Este fue, precisamente, el planteamiento del reverendo Bayes, que finalmente se tradujo en el establecimiento de su famoso *teorema*[58]. En concreto, en nuestro caso, el problema planteado es encontrar $p(A/E)$. La solución que se deriva del teorema de Bayes, y que desarrollaremos a continuación, es la siguiente:

$$p(A/E) = \frac{p(E/A)\,p(A)}{p(E)}$$

En esta expresión, como cabía esperar, la probabilidad condicional "a posteriori" se obtiene a partir de la probabilidad condicional "a priori" y de las probabilidades totales.

Para ilustrar la obtención de la ecuación del teorema de Bayes que acabamos de utilizar, consideraremos una población sobre parte de cuyos elementos ha sido ejecutada la acción *A*, de entre un número arbitrario de acciones posibles. Sobre todos los elementos de dicha población se ha ejecutado una acción (*A* u otra cualquiera de las posibles), a consecuencia de lo cual ha habido una determinada respuesta (*E*, *F*, *G*,...) Supongamos también que nos interesa investigar la relación causal *A*-*E*. De este modo:

- N = población
- Acciones posibles = $A, \neg A$
- Respuestas posibles = $E, \neg E$
- $n(A)$ = nº de elementos de *N* sobre los que se ejecutó *A*
- $n(\neg A)$ = nº de elementos de *N* sobre los que no se ejecutó *A*
- $n(E)$ = nº de elementos de *N* cuya respuesta fue *E*

[58] Nombre quizás demasiado pretencioso para lo que en realidad es una simple ecuación.

- $n(\neg E)$ = nº de elementos de N cuya respuesta no fue E
- $n(A \cap E)$ = nº de elementos de N sobre los que se ejecutó A y se obtuvo como respuesta E

Por la definición de probabilidad condicional:

$$p(E/A) = \frac{n(A \cap E)}{n(A)}$$

Si dividimos numerador y denominador por N obtenemos:

$$p(E/A) = \frac{\frac{n(A \cap E)}{N}}{\frac{n(A)}{N}} = \frac{p(A \cap E)}{p(A)}$$

De donde:

$$p(A \cap E) = p(E/A)\,p(A)$$

Análogamente,

$$p(A/E) = \frac{\frac{n(A \cap E)}{N}}{\frac{n(E)}{N}} = \frac{p(A \cap E)}{p(E)}$$

De donde:

$$p(A \cap E) = p(A/E)\,p(E)$$

Igualando términos:

$$p(E/A)\,p(A) = p(A/E)\,p(E)$$

Expresión que representa la ecuación del teorema de Bayes utilizada en el ejemplo anterior.

Esta ecuación debe poder generalizarse para el análisis de problemas más complicados. La forma de obtener una expresión generalizada para el teorema de Bayes se ilustra a continuación.

Sea una característica cualquiera x y denotemos por A al conjunto de individuos de una población estadísticamente significativa en los que la característica x está presente. Obviamente, $\neg A$ es el conjunto de individuos de la misma población estadística en los que x está ausente. Sea P una prueba potencialmente resolutiva para investigar la característica x, de forma que E es el conjunto de elementos de la población para los cuales P ha dado resultados positivos, y $\neg E$ es el conjunto de elementos de la población para los cuales *P* ha dado resultados negativos. Sean ahora los datos representados en la Tabla 1.1, donde las letras minúsculas representan números.

	A	**¬A**	**Totales**
E	*a*	*b*	*a* + *b*
¬*E*	*c*	*d*	*c* + *d*
Totales	*a* + *c*	*b* + *d*	*N*

Tabla 1.1

Del análisis de la Tabla 1.1, fácilmente se observa que:

- *a* = nº de positivos reales
- *b* = nº de falsos positivos
- *c* = nº de falsos negativos
- *d* = nº de negativos reales

A partir de los datos de la tabla se pueden establecer las siguientes probabilidades totales o *prevalencias*:

$$p(E) = \frac{(a+b)}{N}$$

$$p(\neg E) = \frac{(c+d)}{N}$$

$$p(A) = \frac{(a+c)}{N}$$

$$p(\neg A) = \frac{(b+d)}{N}$$

También a partir de los datos de la tabla podemos establecer las siguientes probabilidades condicionales (que numeramos para referenciarlas en el curso de la demostración):

(1) $p(E/A) = \dfrac{a}{a+c}$ (5) $p(A/E) = \dfrac{a}{a+b}$

(2) $p(E/\neg A) = \dfrac{b}{b+d}$ (6) $p(A/\neg E) = \dfrac{c}{c+d}$

(3) $p(\neg E/A) = \dfrac{c}{a+c}$ (7) $p(\neg A/E) = \dfrac{b}{a+b}$

(4) $p(\neg E/\neg A) = \dfrac{d}{b+d}$ (8) $p(\neg A/\neg E) = \dfrac{d}{c+d}$

Con esta información ¿cómo podríamos conocer la probabilidad condicional “a posteriori” de la característica *x* dado un resultado positivo de la prueba *P*?

Según el teorema de Bayes que acabamos de demostrar:

$$p(A/E) = \frac{p(E/A)\,p(A)}{p(E)}$$

pero,

$$p(E) = \frac{(a+b)}{N}$$

de la expresión (1) se deduce que:

$$a = (a+c)\,p(E/A)$$

y de la expresión (2) se deduce que:

$$b = (b+d)\,p(E/\neg A)$$

por otra parte,

$$p(A) = \frac{(a+c)}{N}$$
$$p(\neg A) = \frac{(b+d)}{N}$$

por lo que,

$$(a+c) = N \cdot p(A)$$
$$(b+d) = N \cdot p(\neg A)$$

y efectuando las sustituciones oportunas,

$$a = N \cdot p(A) \cdot p(E/A)$$
$$b = N \cdot p(\neg A) \cdot p(E/\neg A)$$

Si sustituimos ambas expresiones en la ecuación de $p(E)$ resulta:

$$p(E) = p(A)\,p(E/A) + p(\neg A)\,p(E/\neg A)$$

Llevando este resultado a la expresión original del teorema de Bayes, obtenemos que:

$$p(A/E) = \frac{p(E/A)\,p(A)}{p(E/A)\,p(A) + p(E/\neg A)\,p(\neg A)}$$

Esta última ecuación es directamente generalizable. Efectivamente, si contemplamos más de dos posibilidades (en lugar de simplemente A y ¬A), resulta:

$$p(A_0/E)=\frac{p(E/A_0)p(A_0)}{\sum_i p(E/A_i)p(A_i)}$$

que es la expresión generalizada del teorema de Bayes que emplearemos en este texto[59].

Esta ecuación podemos tratar de aplicarla al ejemplo que habíamos utilizado para ilustrar el razonamiento categórico, y para el cual, recordemos, habíamos obtenido la siguiente BLR:

BLR = [m_1i_1, m_3i_2, m_2i_3, m_4i_3, m_3i_4, m_4i_4]

La cuestión que entonces se nos planteaba era la siguiente:

Dado el conjunto de manifestaciones $f = m(1) \times \neg m(2) = m_3$, ¿cuál es la función g de las interpretaciones permitidas?

Para este problema, directamente de BLR, concluíamos que $g = i_2 + i_4$ o, lo que es lo mismo: $g = \neg i(1)\ i(2) + i(1)\ i(2)$, lo cual traducíamos del siguiente modo: -la interpretación $i(2)$ es segura, pero de $i(1)$ no podemos asegurar nada-.

Según la corrección bayesiana que acabamos de sugerir, el problema se reduce a resolver las siguientes expresiones:

- ¿$p(i_2/m_3)$?
- ¿$p(i_4/m_3)$?

Para ello, y como ya hemos demostrado:

$$p(i_2/m_3)=\frac{p(m_3/i_2)p(i_2)}{p(m_3/i_1)p(i_1)+p(m_3/i_2)p(i_2)+p(m_3/i_3)p(i_3)+p(m_3/i_4)p(i_4)}$$

y, análogamente:

$$p(i_4/m_3)=\frac{p(m_3/i_4)p(i_4)}{p(m_3/i_1)p(i_1)+p(m_3/i_2)p(i_2)+p(m_3/i_3)p(i_3)+p(m_3/i_4)p(i_4)}$$

Evidentemente, si queremos conocer los valores absolutos de tales probabilidades condicionales desde una perspectiva totalmente general, necesitaremos conocer los valores de:

$p(i_1)$,	$p(i_2)$,	$p(i_3)$,	$p(i_4)$
$p(m_1/i_1)$,	$p(m_1/i_2)$,	$p(m_1/i_3)$,	$p(m_1/i_4)$
$p(m_2/i_1)$,	$p(m_2/i_2)$,	$p(m_2/i_3)$,	$p(m_2/i_4)$
$p(m_3/i_1)$,	$p(m_3/i_2)$,	$p(m_3/i_3)$,	$p(m_3/i_4)$
$p(m_4/i_1)$,	$p(m_4/i_2)$,	$p(m_4/i_3)$,	$p(m_4/i_4)$

[59] Hay expresiones todavía más generales.

Sin embargo, si nuestro tratamiento lógico del problema es correcto, si los conjunto de manifestaciones e interpretaciones son completos, si las manifestaciones y las interpretaciones cumplen el requisito de independencia exigido por el teorema de Bayes[60], si la función de conocimiento está bien construida, y si el tratamiento estadístico efectuado es correcto, entonces sólo las siguientes probabilidades condicionales tendrán valores distintos de cero:

$$p(m_1/i_1),\ p(m_3/i_2),\ p(m_2/i_3),\ p(m_4/i_3),\ p(m_3/i_4),\ p(m_4/i_4)$$

ya que son las únicas probabilidades condicionales relativas a complejos manifestación-interpretación que aparecen en BLR. Dicho de otro modo: si después de un análisis estadístico encontramos valores distintos de cero, para las probabilidades condicionales de complejos manifestación-interpretación que no aparecen en BLR (e.g., $p[m_2/i_2] \neq 0$), entonces tendremos que pensar en alguna (o todas) de las siguientes deficiencias:

- nuestro planteamiento lógico no es correcto
- nuestra función de conocimiento no es correcta, bien porque sea incompleta o, simplemente, porque esté mal construida
- la estadística no ha sido bien realizada

Además, para asegurar la consistencia matemática del modelo se tiene que cumplir que:

$$\sum_i P(m_i/i_0) = 1$$

Asumamos para nuestro ejemplo los siguientes valores de las prevalencias y de las probabilidades condicionales:

$$p(i_1) = \frac{910}{1000} \qquad p(i_2) = \frac{50}{1000} \qquad p(i_3) = \frac{25}{1000} \qquad p(i_4) = \frac{15}{1000}$$

$$p(m_1/i_1) = 1 \qquad p(m_3/i_2) = \frac{3}{5} \qquad p(m_2/i_3) = 1 \qquad p(m_4/i_3) = \frac{2}{3} \qquad p(m_3/i_4) = \frac{2}{5} \qquad p(m_4/i_4) = \frac{1}{3}$$

Con estos valores:

---

[60] Volveremos sobre esta cuestión un poco más adelante.

$$p(i_2 / m_3) = \frac{\frac{3}{5} \times \frac{50}{1000}}{\left(\frac{3}{5} \times \frac{50}{1000}\right) + \left(\frac{2}{5} \times \frac{15}{1000}\right)}$$

$$p(i_4 / m_3) = \frac{\frac{2}{5} \times \frac{15}{1000}}{\left(\frac{3}{5} \times \frac{50}{1000}\right) + \left(\frac{2}{5} \times \frac{15}{1000}\right)}$$

Nótese que éstos son valores absolutos de probabilidad condicional. Si sólo hubiésemos deseado averiguar cuál de ambos complejos de interpretaciones es más probable[61], dado que el denominador es el mismo, simplemente dividiendo ambas expresiones entre sí obtendríamos:

$$\frac{p(i_2 / m_3)}{p(i_4 / m_3)} = \frac{\frac{3}{5} \times \frac{50}{1000}}{\frac{2}{5} \times \frac{15}{1000}} = \frac{5}{1}$$

lo que indica que el complejo de interpretaciones $i_2 = \neg i(1) \times i(2)$ es cinco veces más probable que el complejo de interpretaciones $i_4 = i(1) \times i(2)$.

A pesar de lo que acabamos de comentar, la corrección bayesiana también presenta problemas. Así, si manifestaciones e interpretaciones no son independientes, el modelo bayesiano fracasa. No pretendemos abordar una demostración rigurosa y formal; sin embargo, sí trataremos de ilustrar este hecho por medio de un ejemplo.

Sean tres posibles interpretaciones para tres manifestaciones posibles en un universo de discurso dado. Manifestaciones e interpretaciones están relacionadas según los datos de la tabla que se muestra a continuación, y en la que no se cumple el criterio de independencia:

	INTERPRETACIONES				
MANIFESTACIONES	**sólo A_1**	**A_1 y A_2**	**sólo A_2**	**A_3**	**totales**
sólo m_1	$b1$	$c1$	$d1$	$e1$	$n1$
m_1 y m_2	$b2$	$c2$	$d2$	$e2$	$n2$
sólo m_2	$b3$	$c3$	$d3$	$e3$	$n3$
m_3	$b4$	$c4$	$d4$	$e4$	$n4$
totales	B	C	D	E	N

Tabla 1.2

Trataremos de encontrar $p(A_1/m_1)$, con los datos de la tabla, de dos formas diferentes:

---

[61] Dadas las manifestaciones del problema, claro está.

(a) utilizando el teorema de Bayes
(b) directamente a partir del concepto de probabilidad condicional

Para calcular $p(A_1/m_1)$ a partir del teorema de Bayes necesitamos:

$$p(A_1) = \frac{B+C}{N}$$
$$p(A_2) = \frac{C+D}{N}$$
$$p(A_3) = \frac{E}{N}$$
$$p(m_1 / A_1) = \frac{b1+b2+c1+c2}{B+C}$$
$$p(m_1 / A_2) = \frac{c1+c2+d1+d2}{C+D}$$
$$p(m_1 / A_3) = \frac{e1+e2}{E}$$

Aplicando directamente la ecuación del teorema de Bayes:

$$p(A_1 / m_1) = \frac{p(m_1 / A_1)\, p(A_1)}{p(m_1 / A_1)\, p(A_1) + p(m_1 / A_2)\, p(A_2) + p(m_1 / A_3)\, p(A_3)}$$

Sustituyendo valores:

$$p(m_1 / A_1)\, p(A_1) = \frac{b1+b2+c1+c2}{N}$$
$$p(m_1 / A_2)\, p(A_2) = \frac{c1+c2+d1+d2}{N}$$
$$p(m_1 / A_3)\, p(A_3) = \frac{e1+e2}{N}$$

y, en consecuencia:

$$p(A_1 / m_1) = \frac{\dfrac{b1+b2+c1+c2}{N}}{\dfrac{b1+b2+c1+c2}{N} + \dfrac{c1+c2+d1+d2}{N} + \dfrac{e1+e2}{N}}$$

... operando:

$$p(A_1 / m_1) = \frac{b1+b2+c1+c2}{b1+b2+2c1+2c2+d1+d2+e1+e2} = \frac{b1+b2+c1+c2}{n1+n2+c1+c2}$$

Este resultado es manifiestamente falso. Aplicando el concepto de probabilidad condicional directamente sobre los datos de la tabla obtenemos que:

$$p(A_1 / m_1) = \frac{b1 + b2 + c1 + c2}{n1 + n2}$$

Esta limitación del modelo plantea problemas cuando se pretende su aplicación en dominios del mundo real, en los que los requisitos de independencia casi nunca se cumplen.

Pero ésta no es la única deficiencia del modelo bayesiano. En los problemas interesantes para la aplicación de técnicas de inteligencia artificial, la información suele ir apareciendo progresivamente, secuencialmente y, generalmente, de forma poco ordenada. En estos casos, adecuar la aproximación bayesiana a la interpretación secuencial supone considerar que la información factual aparece incrementalmente y, por lo tanto, habrá que adaptar las ecuaciones correspondientes:

Sea *E*1 el conjunto de toda la información disponible en un momento dado, y sea *S*1 un nuevo dato (i.e., un nuevo elemento de información que acaba de aparecer), entonces *E* será el nuevo conjunto formado por la información de *E*1 y el nuevo dato *S*1. Con estas consideraciones, la ecuación del teorema de Bayes debe reescribirse del siguiente modo:

$$p(I_i / E) = \frac{p(S1 / I_i \text{ y } E1)\, p(I_i / E1)}{\sum_j p(S1 / I_j \text{ y } E1)\, p(I_j / E1)}$$

En estas condiciones, y dado que la aplicación del modelo bayesiano tiene que estar avalada por un estudio estadístico correcto y fiable, la reformulación de la ecuación de Bayes para su utilización en la interpretación secuencial complica, aún más, la estadística asociada.

Otro problema frecuente de este modelo procede de su aplicación poco cuidadosa. Ilustraremos este tipo de situaciones con un ejemplo, muy exagerado pero ilustrativo:

Es claro que existe una relación entre el cáncer de pulmón y la hemoptisis[62]. Supongamos que a un determinado hospital llega un paciente con hemoptisis. El médico que lo atiende, que es de la escuela bayesiana, trata de cuantificar la probabilidad de que dicha manifestación sea consecuencia de un cáncer de pulmón. Para ello se dirige a los archivos del hospital y observa que de los 15315 pacientes ingresados durante los últimos tres años, 320 padecían cáncer de pulmón. Por otra parte, de esos 15315 pacientes, 231 ingresaron a causa de una hemoptisis. Por último, 150 pacientes de los 320 diagnosticados de cáncer de pulmón manifestaban hemoptisis. Con estos datos, nuestro médico bayesiano realiza el siguiente análisis:

---

[62] Efectivamente, muchos pacientes que padecen esta terrible enfermedad ingresan en el hospital escupiendo sangre por la boca.

$$p(\text{hemoptisis}) = \frac{231}{15315}$$

$$p(\text{cancer de pulmón}) = \frac{320}{15315}$$

$$p(\text{hemoptisis / cancer de pulmón}) = \frac{150}{320}$$

Aplicando Bayes, y dado que el paciente presenta una hemoptisis:

$$p(\text{cancer de pulmón / hemoptisis}) = \frac{\frac{150}{320} \times \frac{320}{15315}}{\frac{231}{15315}} = 0.67$$

El médico, apesadumbrado, le anuncia al paciente:

"- Tengo malas noticias para usted, tiene un 67% de probabilidades de padecer un cáncer de pulmón."

En esto se abre la puerta de la sala y entra una enfermera que le pregunta al médico:

"- Doctor, ¿le sacamos ya al paciente la viga de hierro que lleva incrustada en el pecho?"

Evidentemente, este ejemplo no es más que un esperpento de lo que teóricamente podría suceder cuando un modelo, en este caso el bayesiano, se aplica mal.

El último gran problema del modelo bayesiano es consecuencia de su consistencia matemática. Al respecto, y por definición, siempre se tiene que cumplir que:

$$p(A) + p(\neg A) = 1$$

Sin embargo, cuando tratamos con problemas del mundo real, los expertos difícilmente asumen esta consistencia[63]. Así, si estamos investigando una hipótesis *H*, hasta el momento avalada, por ejemplo, por las observaciones *O*1, *O*2 y *O*3, la aplicación de un esquema bayesiano podría traducirse del siguiente modo:

$$p(H/O1 \text{ y } O2 \text{ y } O3) = x; \quad 0 \leq x \leq 1$$

lo que inmediatamente nos conduciría a que:

$$p(\neg H/O1 \text{ y } O2 \text{ y } O3) = 1 - x$$

[63] ... y no olvidemos que estamos tratando de diseñar programas que "razonen" como expertos.

Es decir, que un mismo conjunto de evidencias apoya simultáneamente (aunque en distinto grado), a una hipótesis y a su negación. Este es uno de los puntos más débiles de los modelos estadísticos[64] cuando tratamos con “conocimiento” en lugar de tratar con “datos”. En problemas del mundo real, dada una proposición basada en la experiencia, la consistencia matemática no tiene por qué mantenerse. Necesitamos la puesta a punto de un nuevo modelo.

1.5. Resumen

En este tema abordamos por primera vez los problemas de razonamiento desde la perspectiva de la inteligencia artificial. Introducimos el tema identificando distintos tipos de dominios para las cuales unos esquemas de razonamiento son más apropiados que otros. Planteamos a continuación el problema de la interpretación diferencial como uno de los métodos disponibles para formalizar el razonamiento categórico. Según este modelo, el proceso consiste en construir todas las posibles combinaciones entre todas las evidencias posibles y todas las interpretaciones. Ello da lugar a la llamada Base Lógica Expandida que está formada por todos los complejos manifestación-interpretación idealmente posibles. Sobre esta Base Lógica Expandida tendremos que aplicar el conocimiento del dominio para eliminar relaciones entre manifestaciones e interpretaciones que sean absurdas o imposibles. El resultado es una Base Lógica Reducida en la que, previsiblemente, estará la solución de cualquier problema que pudiéramos plantear en el dominio. No obstante, nuestro conocimiento puede no ser completo, o simplemente, nuestras observaciones y evidencias pueden señalar a más de una interpretación. En este último caso, la propia naturaleza del modelo no nos permite discriminar. Esta razón, y la inevitable explosión combinatoria, nos llevan a considerar a los modelos estadísticos como una alternativa potencialmente resolutiva en lo que a los esquemas de razonamiento se refiere. Así, planteamos el esquema bayesiano que utiliza las llamadas probabilidades condicionales -en las que aparecen involucrados dos sucesos, de forma que la ocurrencia del segundo depende de la ocurrencia del primero-. Después de analizar algunas propiedades del modelo bayesiano, aplicamos dicho esquema en la resolución -con éxito- de un problema no resuelto mediante el uso de la aproximación categórica. No obstante, los requisitos de independencia exigidos por el modelo bayesiano, la necesidad de efectuar previamente estudios estadísticos amplios, y el hecho de que, de acuerdo con la consistencia matemática del esquema una misma observación apoye simultáneamente a una hipótesis y a su negación, nos llevan a buscar nuevas soluciones alternativas.

1.6. Textos básicos

- Dude, Hart, Nilsson, “Subjective Bayesian Methods for Rule-Based Inference Systems”, National Computer Conference, 1976.

---

[64] No sólo de los modelos bayesianos.

- Grosof, “Non-monotonicity in Probabilistic Reasoning”, Uncertainty in Artificial Intelligence, vol.2, 1988.

- Ledley, Lusted, “Reasoning Foundations in Medical Diagnosis”, Science, vol.130, 1959.