

## 1. INTRODUCCIÓN

1.1. Probablemente fruto de la imaginación, encontramos la leyenda del Golem, que puede considerarse como uno de los primeros paradigmas de la IA. El Golem era una estatuilla de arcilla roja, a la que un rabino judío de Praga dio vida a través de un ritual de magia negra. El Golem no hacía nada por iniciativa propia, y su único cometido era el de actuar como un fiel esclavo de su rabino. Tras una vida pesada y aburrida, se fueron despertando en el Golem instintos que hasta entonces habían permanecido ocultos y, poco a poco, fue liberándose de la tutela de su amo, convirtiéndose en un personaje terrorífico y maléfico. Finalmente el Golem fue destruido por su creador, aunque la leyenda asegura que reaparece en Praga cada 33 años.

1.2. Antigua Grecia → En esta época es en donde podemos buscar los principios generales de los autómatas, según los cuales:

- Los mecanismos de un autómata actúan en virtud de su propia estructura interna.
- La acción procede de una adecuada organización de fuerzas motrices, naturales y artificiales.
- La movilidad de los autómatas afecta a todo el conjunto, y no sólo a ciertas partes.

1.3. El método Iluliano consiste en realizar un ensayo exhaustivo y sistemático de todos los procedimientos y principios básicos que pudieran ser útiles para resolver problemas concretos. El método propuesto requiere:

- Adquirir los principios fundamentales conocidos, y admitidos por todos, en un dominio de aplicación concreto.
- Agotar todas las posibles combinaciones de dichos principios.

Inconvenientes de este método:

- Determinar cuáles son realmente los principios básicos de un dominio.
- La inevitable explosión combinatoria.

1.4. Edgar Allan Poe, conocedor de los prodigios que se le atribuían al Malzel Chess Automaton (autómata capaz de jugar al ajedrez que resultó ser un fraude), construyó una prueba lógica según la cual el autómata en cuestión no podía ser auténtico. La argumentación de Poe se basaba en dos pilares fundamentales:

- Ninguna máquina puede cambiar su estrategia durante un desarrollo intelectual. Las mismas fuentes intelectuales no tienen por qué producir las mismas respuestas ante los mismos estímulos.
- No hay ninguna máquina capaz de utilizar conocimiento derivado de la experiencia. Los procesos inductivos y de aprendizaje no son propios de ingenios mecánicos.

Lo que Poe no sabía es que su argumentación, que no era más que una traducción del llamado régimen Lovelace (según el cual una máquina sólo puede hacer lo que se le ordena), es falsa. Hoy sabemos que las máquinas pueden, y en ocasiones deben, modificar sus estrategias de trabajo durante la resolución de ciertos problemas. Además, las máquinas incorporan

conocimiento derivado de la experiencia y lo utilizan para inferir nuevos hechos y para aprender de su propia experiencia.

- 1.5. En 1956, un grupo de investigadores se reúne en el Darmouth Collage para discutir sobre la posibilidad de construir máquinas inteligentes. Entre los investigadores que allí se dieron cita estaban Samuel, que había desarrollado un programa para jugar a las damas; McCarthy, que se dedicaba a la construcción de sistemas en los que se llevasen a cabo razonamientos de sentido común; y Minsky, que trabajaba sobre un problema de geometría plana con la esperanza de conseguir que el ordenador emplease razonamiento analógico sobre figuras. Junto a éstos, Newell, Shaw y Simon, fueron los verdaderos promotores de la inteligencia artificial, término que fue acuñado en su día por McCarthy.

Tras esta reunión se formaron diversos grupos de científicos que siguieron trabajando de forma independiente. Así, Newell y Simon formaron un equipo con la idea de desarrollar modelos de comportamiento humano. Por su parte, McCarthy y Minsky formaron otro equipo dedicaron a la construcción de máquinas inteligentes, sin preocuparse del comportamiento humano.

El primer planteamiento supone la emulación de la actividad cerebral y, en la medida de lo posible, la réplica de sus estructura. El segundo planteamiento supone construir sistemas en los que los procedimientos empleados para resolver problemas sean de tal naturaleza que, de ser empleados por un ser vivo, éste sería considerado inteligente.

- 1.6. Los sistemas inteligentes se pueden clasificar en tres niveles distintos: programas de IA, sistemas basados en conocimiento y sistemas expertos.

- 1.6.1. Los programas de IA exhiben cierto comportamiento inteligente fruto de la aplicación hábil de heurísticas, entendiendo como heurística un tipo de conocimiento difícilmente formalizable, que se establece implícitamente para tratar de encontrar respuestas más o menos correctas, pero siempre válidas, a un problema concreto. La utilización de conocimiento heurístico no garantiza encontrar soluciones óptimas, pero sí permite garantizar hallazgo de soluciones aceptables, si existen, a través de los denominados procesos inferenciales.

NOTA: Podemos definir inferencia como el proceso que permite la comprensión de un significado en función de cierta información relacionada.

- 1.6.2. Sistemas basados en conocimiento. En ellos los conocimientos del dominio concreto y las estructuras de control que se utilizan para manipular dicho conocimiento, se encuentran físicamente separados. Ello va a requerir la definición e implementación de arquitecturas diferentes a las que estamos habituados y en las que “conocimientos” y “estructuras de control” puedan ser desarrolladas independientemente las unas de las otras, de forma que una misma estructura de control pueda ser utilizada en muchas bases de conocimiento diferentes.

1.6.3. Los sistemas expertos pueden considerarse como especializaciones de los sistemas basados en conocimiento que utilizan conocimiento particular de un dominio de aplicación concreto para tratar de resolver problemas del mundo real, limitados en tamaño, pero de gran complejidad.

La construcción de sistemas expertos va a requerir el empleo de técnicas desarrolladas para construir programas de IA y la utilización de las arquitecturas definidas para el desarrollo de sistemas basados en conocimiento. Va a ser imprescindible la realización de esfuerzos en aspectos como la adquisición del conocimiento y el aprendizaje.

1.7. El estudio del razonamiento nos permite su clasificación en, al menos, tres modos diferentes: razonamiento deductivo, razonamiento inductivo y razonamiento abductivo. Para ilustrar los tres conceptos consideramos tres aserciones A1, A2 y A3, con las siguientes características:

- A1 = aserción de carácter general
- A2 = aserción de tipo particular referida a la primera parte de A1
- A3 = aserción de tipo particular referida a la segunda parte de A1.

1.7.1. El razonamiento deductivo parte de A1 y de A2, y trata de demostrar A3. Su empleo no genera conocimiento nuevo. Simplemente aplicamos conocimiento dado sobre situaciones particulares para obtener conclusiones válidas.

Ejemplo:

A1 “La gripe produce fiebre”  
A2 “Luís tiene gripe” } → A3 “Luís tiene fiebre”

1.7.2. El razonamiento inductivo parte de A2 y A3, tratando de obtener A1. Los procesos inductivos de razonamiento están ligados a la experimentación. Este tipo de razonamiento sí genera conocimiento nuevo. Para ello se basa en la observación de eventos y en las posibles relaciones que pueden encontrarse. Los procesos de razonamiento inductivo pueden dar lugar a errores importantes, que se minimizan cuanto más cuidadoso sea el análisis previo de las premisas particulares que se pretenden relacionar.

Ejemplo:

A2 “Estos objetos caen”  
A3 “Estos objetos tienen masa” } → A1 “Los objetos con masa caen”

1.7.3. El razonamiento abductivo trabaja sobre la plausibilidad de las conclusiones. Trata de relacionar aserciones de tipo A1 con aserciones de tipo A3, para concluir aserciones de tipo A2.

El modo de razonamiento abductivo trata de concluir algo sobre la posible relación entre dos aserciones que se sabe que son ciertas. Sin embargo, la relación establecida puede ser cierta o no.

Características del razonamiento abductivo:

- Está ligado al concepto de incertidumbre.
- Genera conocimiento nuevo.
- Es el modo de razonamiento típico de la IA.

Ejemplo:

A1 “Los cuadros de Goya presentan estas características” }  
A3 “Este cuadro presenta estas características” } →

→ A2 “Este cuadro es de Goya”

1.8. Podemos establecer distintas categorías de conocimiento, en función del origen y procedencia del conocimiento con respecto al experto humano de quien lo extraemos:

- Conocimiento público. Puede obtenerse directamente a partir de fuentes típicas, libros y manuales. Es comúnmente aceptado y universalmente reconocido.
- Conocimiento semipúblico. Es explícito, pero no está universalmente reconocido ni es comúnmente aceptado. Suele ser utilizado de forma casi exclusiva por grupos de especialistas en áreas concretas.
- Conocimiento privado. No es explícito, no está universalmente reconocido ni es comúnmente aceptado. Es de marcado carácter heurístico y responde muy bien a la frase “cada maestrillo tiene su librillo”.

## 2. RESOLUCIÓN DE PROBLEMAS.

2.1. Dirección del proceso de búsqueda. Existen dos direcciones fundamentales que podemos definir a la hora de configurar un proceso de búsqueda determinado:

- Desde los estados iniciales hacia los estados meta, mediante la generación de estados intermedios obtenidos tras la aplicación sucesiva de operadores.
- Desde los estados meta hacia los estados iniciales, investigando qué estados previos al estado (o estados) meta, y qué operadores aplicables nos producen una transición deseada.

En cierto tipo de sistemas, la primera dirección definida configura un razonamiento progresivo o dirigido por los datos, mientras la segunda configura un razonamiento regresivo o dirigido por los objetivos.

La elección sobre la dirección de búsqueda más conveniente debe considerar los siguientes aspectos:

- Tamaño relativo de los conjuntos I y M.
- Factor de ramificación.
- Inclusión de estructuras explicativas como requisito inicial en el diseño de nuestro sistema inteligente.

El tamaño relativo de los conjuntos I y M es fundamental a la hora de decidir qué dirección debemos seguir. Así, es preferible explorar el espacio de estados de forma que progrese desde un conjunto inicialmente pequeño de información de partida, hacia un conjunto mayor de estados.

El factor de ramificación es un parámetro que influye en la eficacia del proceso de búsqueda, de forma que siempre trataremos de explorar el espacio de estados según la dirección del menor factor de ramificación, ya que cuanto menor sea el factor de ramificación menor será el número de alternativas posibles que debemos explorar.

Si el sistema debe ser capaz de “explicar” un proceso dado de razonamiento, es conveniente que el razonamiento del sistema se produzca en la dirección que concuerde más aproximadamente con la forma de razonar del usuario humano.

2.2. Funciones heurísticas. Se conoce como función heurística a aquella función de carácter numérico que nos permite cuantificar el beneficio de una transición efectuada en el espacio de estados del dominio del problema a resolver. Las funciones heurísticas resultan muy útiles a la hora de optimizar los procesos de búsqueda. Para ello intentan guiar la exploración del espacio de estados en la dirección más provechosa, sugiriendo el “mejor” camino a seguir cuando disponemos de varias alternativas.

2.3. Tipos de búsqueda:

- En anchura.
- En profundidad.
- Mixta profundidad-anchura.

(Ver tema “Resolución de problemas” – Página 74 transparencias: “Estrategias de búsqueda: Comparativa”)

2.4. Tanto la ascensión a colinas básica como la máxima pendiente pueden no encontrar una solución si caen en un estado del que no es posible generar nuevos estados mejores que él. Esto ocurre cuando el proceso de búsqueda se encuentra con un máximo local, una meseta o una cresta. Considerando el espacio de estados como una superficie n-dimensional en la que cada punto está definido por los valores que definen ese estado y el valor de la función heurística de ese estado, se define:

- Un máximo local como un estado puntual mejor que cualquiera de sus vecinos, pero peor que otros estados más lejanos. Llegados a un punto de estas características, cualquier operación que hagamos, cualquier movimiento que intentemos, nos llevará a un estado aparentemente peor, aunque en realidad nos estemos aproximando a la solución. Cuando los máximos locales aparecen cerca de la solución final se denominan estribaciones.
- Una meseta como un área plana del espacio de estados en la que todos los estados individuales tienen un mismo valor de la función heurística. En las mesetas no es posible determinar cuál es la mejor dirección para continuar la búsqueda.
- Una cresta como un tipo especial de máximo local. Es un área del espacio de estados, que tiene estados con mejores valores de la función heurística que las regiones colindantes y, además, posee una inclinación, pero la orientación de esta región alta hace que sea imposible atravesar la cresta mediante transiciones simples.

Estos inconvenientes tienen difícil solución. Como norma general cuando aparece alguna de estas situaciones se pueden intentar las siguientes estrategias:

- Regresar a un nodo previo e intentar una dirección diferente puede resolver el problema de los máximos locales.
- Realizar un gran salto en el espacio de búsqueda, para tratar de alcanzar una región diferente del espacio de estados, puede resolver el problema de las mesetas.
- Aplicar más de un operador antes de comparar los valores de la función heurística puede resolver el problema de las crestas.

### 3. REPRESENTACIONES FORMALES DEL CONOCIMIENTO

3.1. Cualquier procedimiento para representar conocimiento tiene que reunir un conjunto mínimo de condiciones, entre las que destacamos:

- Transparencia. Hace referencia a si podemos o no identificar fácilmente el conocimiento representado.
- Naturalidad. Hace referencia a si podemos o no representar el conocimiento en su forma original.
- Claridad. Hace referencia a si podemos o no representar directamente el conocimiento.
- Eficiencia. Facilidad relativa con la cual se puede acceder a conocimientos específicos durante la ejecución.
- Adecuación. Capacidad del esquema de representación elegido para representar todos los conocimientos y tipos de conocimientos que requiere el sistema.
- Modularidad. Capacidad del esquema de representación para fragmentar distintos tipos de conocimiento del sistema.

3.2. Fórmulas bien definidas = FBDs, Fórmulas bien formadas = FBFs.

3.3. Componentes básicos de un esquema de representación del conocimiento basado en lógica de predicados:

- Alfabeto.
- Lenguaje formal.
- Conjunto de enunciados básicos o axiomas.
- Reglas inferenciales.

(NOTA: los axiomas describen fragmentos de conocimiento y las reglas inferenciales se aplican a los axiomas para tratar de deducir nuevos enunciados verdaderos.)

3.4. Alfabeto. Conjunto de símbolos a partir de los que se construyen los enunciados. En lógica de predicados el alfabeto está constituido por los siguiente elementos:

- Predicados.
- Variables.
- Funciones.
- Constantes.
- Juntores.
- Cuantificadores.
- Delimitadores.

3.5. El lenguaje formal asociado a la lógica de predicados es el conjunto de todas las FBDs que se pueden construir legalmente a partir del alfabeto.

Se puede definir inductivamente una FBD del siguiente modo:

- Cualquier fórmula atómica es FBD.
- Si F y G son FBDs, entonces también son FBDs las siguientes:  
not F          F and G          F or G          F → G

- Si “x” es una variable y F es una FBD, entonces también son FBDs las siguientes:  
 $(\text{para todo } x) F$        $(\text{existe } x) F$

El conjunto de FBDs que seamos capaces de construir sobre un dominio concreto constituye el lenguaje formal asociado.

Existen expresiones sencillas que, de acuerdo con la definición inductiva dada anteriormente, no pueden ser consideradas FBDs. Algunas de tales expresiones son:

- not f(A)
- (para todo P) P(A)
- (existe f) f(A)

Estas expresiones no son FBDs en la lógica de predicados. Las lógicas que no permiten cuantificación sobre predicados o funciones se denominan lógicas de primer orden.

3.6. Reglas de inferencia. La inferencia en lógica formal es el proceso de generar nuevas FBDs a partir de FBDs ya existentes mediante la aplicación de las llamadas reglas de inferencia. De tales reglas la más común es la llamada “modus ponens”, que se puede expresar así:  $[P1 \text{ and } (P1 \rightarrow P2)] \rightarrow P2$ , expresión que viene a decir que “si sabemos que P1 implica a P2, y sabemos que P1 es verdad, podemos inferir que P2 es verdad”.

Otra regla de inferencia común es la “especialización universal”, que se utiliza para generar FBD “f (INDIVIDUO)” a partir de la FBD (para todo x)[f(x)]. La especialización universal puede expresarse así:

$$\text{INDIVIDUO and (para todo } x)[f(x)] \rightarrow f(\text{INDIVIDUO})$$

3.7. Fases para la representación del conocimiento basado en lógica formal:

- Comprensión e identificación del conocimiento relevante del dominio.
- Formalización de los enunciados correspondientes.
- Análisis o fragmentación de los enunciados en sus partes constituyentes.
- Establecimiento de la simbología adecuada para representar elementos y relaciones.
- Construcción de las FBDs.

Estos cinco pasos (denominados respectivamente identificación, formalización, descomposición, traducción y recomposición), configuran la fase de codificación mediante la cual pretendemos obtener representaciones del conocimiento del dominio manejables desde una perspectiva computacional.

#### 4. MÉTODOS ESTRUCTURADOS DE REPRESENTACIÓN DEL CONOCIMIENTO.

4.1. En IA se utilizan esquemas no formales de representación del conocimiento. Tales esquemas son fundamentalmente métodos estructurados de representación y tienen que verificar las siguientes propiedades:

- Adecuación representacional. El esquema elegido debe ser capaz de representar las distintas clases del conocimiento del dominio.
- Adecuación inferencial. El esquema elegido debe permitir la manipulación del conocimiento para obtener conocimiento nuevo.
- Eficiencia inferencial. El esquema elegido debe ser versátil utilizando aquella información que permita optimizar el proceso inferencial.
- Eficacia adquisicional. El esquema elegido debe suministrar vías que permitan la incorporación de información y conocimientos nuevos.

4.2. Ventajas de los métodos de representación del conocimiento procedimental (o esquemas procedimentales):

- Al dar prioridad a los procedimientos hacemos mayor énfasis en las capacidades inferenciales del sistema.
- Permiten explorar distintos modelos y técnicas de razonamiento.
- Permiten trabajar con falta de información y con datos de carácter probabilístico.
- Incorporan de forma natural conocimiento de tipo heurístico.

4.3. Métodos declarativos:

- Redes semánticas. Permiten describir simultáneamente acontecimientos y objetos.
- Modelos de dependencia conceptual. Estructuras especializadas que proporcionan mecanismos para representar relaciones entre los componentes de una acción.
- Frames. Estructuras genéricas que permiten representar objetos complejos desde diferentes puntos de vista.
- Guiones. Estructuras especializadas que derivan de los frames y que son útiles para representar secuencias comunes de acontecimientos.

4.4. Tipos de relaciones en redes semánticas:

- Ocurrencia. Cuando se relaciona un miembro de una categoría general con la categoría a la que pertenece (se suele etiquetar “pertenece”).
- Generalización. Cuando se relaciona una entidad con otra de carácter más general (ES\_UN).
- Agregación. Cuando se relacionan componentes de un objeto con el objeto propiamente dicho (ES\_PARTE\_DE).
- Acción. Cuando se establecen vínculos dinámicos entre diferentes objetos.
- Propiedades. Relaciones entre objetos y características de los objetos.

#### 4.5. Modos de razonamiento en redes semánticas:

- Mediante rastreo. El modelo permite obtener asociaciones simplemente rastreando los enlaces del sistema. Ejemplo: 12 “mayor que” 7 “menor que” 3, el rastreo nos permite concluir que “12 es mayor que 3”.

No obstante, ninguna regla semántica rigurosa guía el proceso. En redes semánticas, las relaciones que se representan pueden no ser totalmente rigurosas y, por lo tanto, las inferencias obtenidas por rastreo no ser válidas.

- Mediante emparejamiento. Consiste en construir fragmentos de red, algunos de cuyos nodos tienen valores definidos pero otros no. En este caso los nodos sin valores se representan por variables. El sistema debe entonces tratar de encontrar un fragmento de la red semántica original que encaje perfectamente en el fragmento de red semántica que hemos construido para representar nuestro problema.

#### 4.6. La dependencia conceptual intenta representar el significado de frases en lenguaje natural de forma que:

- Se posibilite la derivación de conclusiones.
- La representación sea independiente del lenguaje utilizado en las declaraciones originales.

#### 4.7. Las frames pueden describirse como redes semánticas complejas que tratan el problema de la representación desde la óptica del razonamiento por semejanza. Describen clases de objetos y pueden definirse como representaciones estructuradas de conocimiento estereotipado.

Estructuralmente, una frame consta de una cabecera, que le da nombre a la frame, y de un conjunto de slots, cada uno de los cuales representa una propiedad o atributo del elemento genérico representado por la frame. Cada slot puede tener distintos slots anidados y sin limitación de profundidad.

Cada una de las indentaciones de los slots representa un nivel de conocimiento y su contenido es una especialización del nivel anterior.

Un sistema en el que el conocimiento esté representado por medio de frames utiliza el concepto de herencia. Para ello se emplean slots de tipo ES\_UN, que permiten la entrada de información a una frame, en un nivel de conocimiento determinado, y a partir de la cual la información de la clase correspondiente pasa al objeto considerado.

## 6. SISTEMAS DE PRODUCCIÓN.

6.1. Un sistema de producción está constituido por tres elementos fundamentales:

- La base de conocimientos.
- La memoria activa.
- El motor de inferencias.

Estos tres elementos se relacionan como se muestra en la figura 6.1, página 140 del libro.

6.2. La memoria activa es la estructura que contiene toda la información de naturaleza estática necesaria para resolver un problema concreto. Esta información incluye:

- Datos iniciales del problema.
- Datos incorporados con posterioridad.
- Hechos establecidos durante los procesos inferenciales.
- Hipótesis de trabajo, metas o submetas que todavía no han sido establecidas.

La memoria activa representa siempre nuestro estado actual.

6.3. El motor de inferencias consta, fundamentalmente, de dos entidades: un intérprete y una estrategia de control, y contiene los mecanismos necesarios para:

- Examinar la memoria activa y determinar qué reglas deben ejecutarse. Este proceso de selección se realiza en función de la estrategia de búsqueda elegida y de los modelos implementados para la resolución de los conflictos que pudiesen aparecer.
- Controlar y organizar el proceso de ejecución de las reglas seleccionadas en el paso anterior.
- Actualizar la memoria activa cuando sea preciso.
- Asegurar que el sistema tiene autoconocimiento. Ello implica saber en todo momento qué reglas han sido activadas, cuáles han sido ejecutadas, cuál ha sido el último hecho incorporado a la memoria activa, etc.

6.4. Tareas y fases del ciclo básico de un sistema de producción:

- Fase de decisión o selección de reglas.
- Fase de acción o ejecución de las reglas seleccionadas en el paso anterior.

6.4.1. La fase de decisión o selección de reglas es responsable de las siguientes tareas:

- Restricción.
- Equiparación.
- Resolución de conflictos.

La restricción trata de simplificar el proceso de equiparación eliminando del foco de atención del motor de inferencias aquellas reglas que

claramente no tienen nada que ver con el estado actual representado en la memoria activa del sistema.

La equiparación, o emparejamiento, trata de identificar qué reglas son potencialmente relevantes en el contexto del problema que queremos resolver. Si el proceso sigue un encadenamiento progresivo la equiparación consistirá en seleccionar aquellas reglas cuyos antecedentes estén representados en hechos y/o datos de la memoria activa. Si estamos ante un encadenamiento regresivo, trataremos de encontrar aquellas reglas que concluyan algo sobre hipótesis presentes en la memoria activa. En este caso podrán generarse subhipótesis que darán lugar a nuevos ciclos.

El resultado final tras la etapa de emparejamiento es la obtención del denominado conjunto conflicto que incluye todas las reglas potencialmente útiles en la resolución de nuestro problema.

Finalmente, en la etapa de resolución de conflictos, será cuando se decida qué regla (o reglas) deberemos aplicar

6.4.2. Una vez finalizada la fase de decisión o selección de reglas el sistema está ya preparado para disparar físicamente las reglas seleccionadas. Esta es la fase de acción o ejecución, cuyo resultado es:

- La actualización de la memoria activa, con nuevos hechos, con nuevas hipótesis, o con ambas a la vez.
- El marcaje de las estructuras utilizadas.
- La verificación de si debemos o no continuar el proceso cíclico.

Este ciclo básico deberá proseguir hasta que, o bien hayamos encontrado la solución del problema planteado, o bien no puedan utilizarse más reglas.

## 7. RAZONAMIENTO CATEGÓRICO Y CORRECCIÓN BAYESIANA.

### 7.1. Esquema de la interpretación diferencial:

- 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.
- 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.

7.2. Función de conocimiento E. Representa el conjunto de todas las posibles relaciones causales que se pueden establecer en nuestro dominio de discurso entre manifestaciones e interpretaciones. 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..."

7.3. Como resultado de construir el conjunto completo de todas las posibles combinaciones entre complejos de manifestaciones y complejos de interpretaciones se obtiene una lista exhaustiva de complejos manifestación-interpretación, que se denomina base lógica expandida (BLE).

$$\begin{aligned} \text{n}^\circ \text{ de manifestaciones} = m &\rightarrow \text{n}^\circ \text{ de complejos de manifestaciones} = 2^m \\ \text{n}^\circ \text{ de interpretaciones} = i &\rightarrow \text{n}^\circ \text{ de complejos de interpretaciones} = 2^i \end{aligned}$$

$$\text{n}^\circ \text{ de complejos de manifestación-interpretación} = 2^{(m+i)}$$

Pero no todos los complejos manifestación-interpretación de la BLE 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 BLE 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:  $E: (f \rightarrow g)$

7.4. La forma de obtener una expresión generalizada para el teorema de Bayes se ilustra en las páginas 179, 180 y 181 del libro.

7.5. Si después de un análisis estadístico encontramos valores distintos de cero para las probabilidades condicionales de complejos manifestación-interpretación que

no aparecen en la BLR, 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:  $\sum_i P(m_i/i_0) = 1$

## 9. FACTORES DE CERTIDUMBRE

9.1. Modelo de Shortliffe y Buchanan. Es de naturaleza “ad hoc” y por consiguiente carece de una base teórica fuerte. No obstante, fue inmediatamente aceptado debido a su fácil comprensión y a la calidad de los resultados obtenidos tras su aplicación. Las ideas básicas del modelo pueden resumirse en los siguientes puntos:

- Dada una hipótesis que está siendo considerada, la potencia evidencial de una declaración se debe representar a través de dos medidas diferentes: la medida de confianza creciente MB y la medida de desconfianza creciente MD.
- MB y MD son, en realidad, índices dinámicos que representan incrementos asociados a evidencias nuevas.
- Si  $h$  es una hipótesis y  $e$  una evidencia, la misma evidencia  $e$  no puede, simultáneamente, incrementar la confianza en  $h$  y disminuir la confianza en  $h$ .
- $MB(h,e)$  representa el incremento de la confianza en  $h$  dada la evidencia  $e$ .
- $MD(h,e)$  representa el incremento de la desconfianza en  $h$  dada la evidencia  $e$ .

Con estas premisas podemos establecer una serie de casos particulares que se explican detalladamente en las páginas 218, 219 y 220 del libro.

9.2. Shortliffe y Buchanan proponen una primera aproximación para la combinación entre pares de evidencias que se refieren a la misma hipótesis. Esta primera aproximación, en términos de factores de certidumbre, puede expresarse tal y como se muestra en las páginas 223 y 224 del libro.

Deficiencias de la primera aproximación:

- El orden en que aparecen las evidencias modifica considerablemente el resultado final.
- La falta de asociatividad. Tiene que ver con la gran sensibilidad de la formulación ante la aparición de evidencias contradictorias en estados avanzados del proceso de razonamiento.

Ni Shortliffe ni Buchanan consideraron posible esta situación y fueron ellos mismos quienes propusieron una segunda aproximación que paliaba estas “deficiencias”. La nueva formulación que ambos propusieron aparece en la página 225 del libro.

## 10. TEORÍA EVIDENCIAL.

10.1. La función básica de asignación de verosimilitud:  $m$ , toma valores en el intervalo cerrado  $[0,1]$ . Sea  $A$  un subconjunto cualquiera del marco de discernimiento,  $m(A)$  representa el grado en que  $A$  se ve favorecida ante la aparición de una determinada evidencia  $e$ .

Dempster y Shafer definen dos condiciones que ha de satisfacer  $m$  (ver fórmula en página 239).

Ambas condiciones son consecuencia de las restricciones impuestas al marco de discernimiento (que debe de ser un conjunto completo de hipótesis mutuamente excluyentes).

10.2. Combinación de evidencias en la teoría de Dempster y Shafer. Ver páginas 241 y 242 del libro.

10.3. La teoría evidencial permite el seguimiento de la evolución dinámica de la confianza depositada en los subconjuntos del marco de discernimiento a medida que aparecen nuevas evidencias. Para ello se definen dos nuevas medidas: la credibilidad y la plausibilidad que, respectivamente, son un indicador de la mínima y máxima confianza que podemos depositar en un elemento focal dado.

10.3.1. Credibilidad (ver fórmula en página 245 del libro). Sea  $A$  el elemento focal considerado, subconjunto del marco de discernimiento.  $Cr(A)$  es la credibilidad de  $A$ , y es una medida de las contribuciones que todos los subconjuntos de  $A$  ejercen sobre el propio  $A$ .

10.3.2. Plausibilidad (ver fórmula en página 247 del libro). Es un indicador de la máxima confianza que podemos depositar en un elemento focal dado. Se calcula considerando también las contribuciones de otros subconjuntos con intersección no nula. ( $Pl(A)$ , siendo  $A$  el elemento focal considerado, subconjunto del marco de discernimiento). La plausibilidad no sólo tiene en cuenta los subconjuntos del propio elemento focal, sino también todas las contribuciones de todos aquellos subconjuntos que tienen algo que ver con dicho elemento focal.

10.4. Intervalo de confianza. Se construye, para cada elemento focal, a partir de la credibilidad y de la plausibilidad. Así, en cada nivel del proceso de razonamiento, el intervalo de confianza es el segmento del espacio numérico  $[0,1]$  que tiene como valor mínimo el valor de la credibilidad del elemento focal y como valor máximo el correspondiente valor de la plausibilidad.

Conceptualmente, el intervalo de confianza representa la incertidumbre asociada al elemento focal considerado.

10.5. Casos particulares de la teoría evidencial. La teoría evidencial contiene, bajo ciertos supuestos y en ciertas situaciones, al modelo de los factores de certidumbre de Shortliffe y Buchanan. Ver análisis del caso de dos evidencias

que apoyan al mismo elemento focal A de un determinado marco de discernimiento, página 248 del libro.

## 11. CONJUNTOS DIFUSOS.

11.1. Cualquier conjunto, sea cual sea su naturaleza, es “fuzzyfiable”; es decir, se puede establecer una graduación entre los niveles de pertenencia de distintos elementos de un referencial con respecto al conjunto considerado.

11.2. Estructura algebraica de los conjuntos difusos. Propiedades en las páginas de la 260 a 265 del libro.

11.3. Operaciones algebraicas con conjuntos difusos. Ver páginas de la 265 a la 268 del libro.

11.4. Modos de razonamiento:

- Razonamiento categórico. Utiliza declaraciones difusas, pero no emplea ni cuantificadores difusos ni probabilidades difusas.

Ejemplo:

María es una chica delgada. }  
María es una chica muy inteligente. } →

→ María es una chica delgada y muy inteligente.

En este ejemplo, las premisas “delgada” y “muy inteligente” deben interpretarse como predicados difusos. Por otra parte, el predicado difuso de la conclusión es la conjunción de las premisas anteriores.

- Razonamiento silogístico. Produce inferencias con premisas que incorporan cuantificadores difusos.

Ejemplo:

La mayoría de los suecos son rubios. }  
La mayoría de los suecos rubios son altos. } →

→ (La mayoría)<sup>2</sup> de los suecos son rubios y altos.

En este caso, el cuantificador difuso “la mayoría” debe interpretarse como una proporción difusa, y “(la mayoría)<sup>2</sup>” es el cuadrado de “la mayoría” en aritmética difusa.

- Razonamiento disposicional. Las premisas son disposiciones. La conclusión obtenida es una máxima que debe interpretarse como un mandato disposicional.

Ejemplo:

Fumar mucho suele ser causa de abundante tos. → Para evitar tos abundante evite fumar mucho.

- Razonamiento cualitativo. Se define como un modo de razonamiento en el cual las relaciones entrada/salida de un sistema se representan por medio de una colección de reglas difusas, de tipo IF-THEN, en las que los antecedentes y los consecuentes incluyen variables lingüísticas. Es empleado habitualmente en las aplicaciones de la lógica difusa al análisis de sistemas y al control de procesos.

## 12. INGENIERÍA DEL CONOCIMIENTO.

12.1. Análisis de la viabilidad de un sistema experto. Para ello se usa la metodología de diseño de Waterman, basada en el estudio de cuatro características esenciales: justificación, posibilidad, adecuación y éxito.

### 12.1.1. Viabilidad.

- IF: (1) La construcción de un sistema experto está justificada.  
and: (2) La construcción de un sistema experto es posible.  
and: (3) La construcción de un sistema experto es adecuada.  
and: (4) Hay ciertas garantías de éxito tras la construcción de un sistema experto.

THEN: La construcción de un sistema experto es viable.

12.1.2. Posibilidad. El criterio de posibilidad hace referencia a algunos de los aspectos más pragmáticos en el desarrollo de un sistema experto.

- IF: (1) Existen varios expertos cooperativos.  
and: (2) Los expertos logran ponerse de acuerdo.  
and: (3) Los expertos son capaces de articular sus métodos y procedimientos de trabajo.  
and: (4) Las tareas no son excesivamente difíciles.  
and: (5) Las tareas están suficientemente estructuradas.  
and: (6) El sentido común no es determinante.  
and: (7) Se dispone de un número suficiente de casos relevantes para la verificación y posterior validación del producto.  
and: (8) Las tareas no son exclusivamente de carácter teórico.

THEN: El desarrollo de un sistema experto es posible.

### 12.2. Fases de adquisición del conocimiento.

- Conceptualización.
- Formalización.
- Elicitación.

12.3. Fase de elicitación. Es la primera “bajada a la arena” en la adquisición del conocimiento. Es un proceso de extracción del conocimiento de los expertos humanos, que se realiza de una forma estructurada y consistente con el proceso de conceptualización. La elicitación supone, en primer lugar, diseñar algún tipo de estrategia estructurada de interlocución con los expertos.

Esto puede obligarnos a identificar previamente distintas categorías de información y establecer unos formatos adecuados a la lógica de los procesos con los que nos enfrentamos. A continuación, procede el diseño y materialización de alguna estructura física con la que llevar a cabo la elicitación de conocimiento.

Esta fase concluye con la traducción de la información adquirida en una representación, sobre el papel, estructurada o formal.

#### 12.4. Técnicas de extracción del conocimiento:

- Observación directa. Técnica de extracción del conocimiento, previa al diseño real del sistema inteligente. Consiste en la simple observación pasiva del modo en que un experto se enfrenta con los problemas del dominio de aplicación. Objetivo: tratar de familiarizar al ingeniero de conocimiento con el dominio de aplicación y con el entorno.
- Disección del problema. Tras la observación directa procede la discusión informal con los expertos sobre un conjunto de problemas representativos del dominio. Objetivo: tratar de averiguar de qué manera los expertos tienden a organizar sus conocimientos, cómo representan mentalmente sus conceptos y sus ideas, y cómo analizan la información inconsistente, inexacta o imprecisa. Resultado: aparición de nuevos términos, de nuevos conceptos y de nuevas relaciones.
- Descripción del problema. Supone describir problemas típicos relacionados con cada categoría importante de respuestas. Se trata de descubrir estrategias y enfoques básicos, de carácter general, para tratar de establecer una organización jerárquica del conocimiento de los expertos.
- Análisis. Para aplicar esta técnica se requiere que los expertos resuelvan, en presencia del ingeniero de conocimiento, un conjunto de problemas del dominio. Los problemas planteados deben ser realistas, y el experto comentará todos y cada uno de sus procedimientos de resolución. Las explicaciones deberán ser pormenorizadas, exhaustivas y detalladas. El ingeniero de conocimiento deberá cuestionar cada paso de la resolución efectuada por el experto y tratará de generalizar las conclusiones y metodologías aplicadas, sin que ello suponga pérdida de la calidad inferencial.
- Refinamiento. Proceso de análisis invertido, en el que el experto plantea problemas al ingeniero de conocimiento para que éste los resuelva. Los problemas planteados deben de ser de dificultad creciente, y el experto debe supervisar al ingeniero del conocimiento mientras trata de resolver los problemas. Objetivo: crítica de planteamientos, búsqueda de soluciones alternativas y optimización de los procesos de resolución.
- Examen. Supone la revisión microscópica del conocimiento del sistema. Aquí es el experto quien debe analizar todas y cada una de las reglas del sistema, supervisar las estrategias de resolución y, en definitiva, dar el visto bueno al prototipo construido. Resultado: incremento y mejora de las posibilidades de explicación y justificación del sistema experto.
- Validación. Proceso completo mediante el cual se pretende comprobar que el sistema experto funciona correctamente frente a problemas reales y en entornos reales.

Todas estas técnicas de extracción del conocimiento se apoyan en una serie de herramientas de ayuda como son:

- Entrevistas.
- Análisis de tareas y protocolos.
- Clasificación de conceptos.

## 17. VERIFICACIÓN Y VALIDACIÓN DE SISTEMAS INTELIGENTES.

17.1. Métodos cuantitativos de validación. Están basados en análisis estadísticos que tratan de comparar las conclusiones del sistema inteligente con las producidas por los expertos del dominio. Se subdividen en tres grupos:

- Medidas de pares.
- Medidas de grupo.
- Ratios de acuerdo.

17.2. Medidas de pares. Consiste en comparar las interpretaciones del sistema con las de un único experto humano. Estas comparaciones se realizan a través de medidas que involucran a pares de expertos y cuyo proceso de realización es: a partir de los datos incluidos en la base de datos de validación se desarrolla una tabla de contingencia para cada uno de los posibles pares que se puedan formar entre los expertos involucrados en la validación (incluyendo al sistema inteligente). Después se extrae de la tabla de contingencia la medida de pares determinada.

Las medidas de pares se dividen en dos grupos:

- Medidas de acuerdo. Nos da un índice que cuantifica las coincidencias entre las interpretaciones de dos expertos. Podemos destacar:
  - Índice de acuerdo.
  - Índice de acuerdo dentro de uno.
  - Kappa.
  - Kappa ponderada.
- Medidas de asociación.

17.3. Índice de acuerdo. Cociente entre el número de observaciones de acuerdo y el número de observaciones totales (ver ecuación 17.1, página 446 del libro).  $N$  es el número total de casos,  $k$  es el número total de categorías semánticas y  $n_{ij}$  el número total de casos en la celda  $ij$  de la tabla de contingencia.

Ventaja: sencillez de su interpretación, que ha hecho que su uso se extienda en distintos campos y aplicaciones.

Inconveniente: no diferencia los desacuerdos según su importancia y no tiene en cuenta la casualidad.

17.4. Índice de acuerdo dentro de uno. Similar al anterior pero considera acuerdos parciales aquellos acuerdos que se diferencian en una única categoría semántica (ver ecuación 17.2 de la página 446 del libro).

Ventaja: elimina los problemas asociados a las categorías semánticas ordinales cuyos límites no pueden establecerse con claridad. Además, puede permitir el análisis de tendencias en las interpretaciones si distinguimos entre los acuerdos dentro de uno “optimistas” (que se producen por encima de la diagonal de acuerdo) o “pesimistas” (que se producen por debajo de la diagonal de acuerdo.).

Inconvenientes: en caso de que los acuerdos sean elevados o que las categorías semánticas sean pocas, el índice presenta una tendencia a adoptar valores cercanos a la unidad. Por otro lado, al igual que el índice de acuerdo simple, no presenta ningún ajuste para corregir aquellos acuerdos debidos a la casualidad.

17.5. Kappa. Está basada en dos cantidades:  $p_o$  (proporción de acuerdo observado) y  $p_c$  (proporción de acuerdo esperado debido a la casualidad). De esta forma,  $1-p_c$  representa el máximo acuerdo posible una vez que se ha eliminado la casualidad, y  $p_o-p_c$  representa el acuerdo obtenido una vez que se ha eliminado la casualidad. Ver ecuación 17.3, página 447 del libro. (Ver también ecuación 17.4).

Problema: trata todos los desacuerdos de la misma forma, dándoles la misma importancia. Por este motivo se modificó kappa para obtener kappa ponderada.

17.6. Kappa ponderada ( $k_w$ ). Medida de acuerdo que corrige aquellos acuerdos debido a la casualidad y pondera de forma distinta los desacuerdos encontrados. La ponderación se hace a partir de una matriz de pesos en la que, para cada posible par de categorías  $ij$ , se define un peso  $v_{ij}$ , que cuantifica el desacuerdo existente. A las celdas pertenecientes a la diagonal principal (que representan el acuerdo perfecto) se les suele asignar el valor 0 indicando que no existe ningún desacuerdo. El mayor valor de desacuerdo  $v_{max}$  es fijado por el investigador. Para cualquier conjunto de pesos, kappa ponderada es invariable ante transformaciones multiplicativas positivas, es decir, que kappa ponderada no cambiará de valor si sus pesos se multiplican por un valor mayor que cero.

Ver fórmula 17.5, página 448 del libro.  $p_{oij}$  es la proporción de acuerdo observada para la casilla  $ij$ ,  $p_{cij}$  es la proporción de acuerdo debido a la casualidad correspondiente a la casilla  $ij$ ,  $v_{ij}$  es el peso correspondiente a la casilla  $ij$ , y  $k$  es el número de categorías.

17.7. Las medidas de pares son útiles cuando el número de expertos es reducido. Sin embargo, si la validación involucra a un grupo amplio de expertos, la información que proporcionan las medidas de pares puede resultar difícil de interpretar.

Las medidas de pares pueden servir de base para otro tipo de medidas: las medidas de grupo, que tienen como objetivo analizar conjuntamente las interpretaciones de los expertos y tratar de buscar estructuras de representación que permitan una interpretación más sencilla dentro del contexto de la validación.

Dentro de las medidas de grupos están:

- Índice de Williams.
- Análisis cluster.
- Escalamiento multidimensional.
- Medidas de dispersión y tendencia.

17.8. El índice de Williams se representa en la ecuación 17.10, ver página 451 del libro.  $P_o$  representa el acuerdo existente entre un experto aislado en relación a un grupo de expertos de referencia, mientras que  $P_n$  representa el acuerdo existente dentro de dicho grupo de referencia.  $P_o$  se define en la ecuación 17.11 y  $P_n$  en la 17.12 (ver de nuevo página 451 del libro), donde  $n$  representa el número de expertos de referencia y  $P_{(a, b)}$  una medida de pares que relaciona las interpretaciones de los expertos  $a$  y  $b$ .

Interpretación de  $I_n$ :

- Si  $I_n < 1$ , entonces el acuerdo entre el experto aislado y el grupo de expertos es menor que el acuerdo entre los propios miembros del grupo.
- Si  $I_n = 1$ , entonces el experto aislado coincide con el grupo al mismo nivel que los miembros del grupo coinciden entre sí.
- Si  $I_n > 1$ , entonces el experto aislado coincide con el consenso del grupo de expertos.

Inconveniente: pueden ser mal interpretadas si dentro del grupo de referencia existe un experto claramente en desacuerdo con los demás, o si el acuerdo dentro del grupo de referencia es escaso. En tales casos se debería de estudiar si el experto en desacuerdo debería ser apartado del grupo de referencia o tratar de utilizar técnicas de desarrollo de un consenso dentro del grupo.