Sistemas Expertos, Redes Bayesianas y sus aplicaciones

Semana ESIDE, Abril 2005

Álvaro Marín Illera
alvaro@rigel.deusto.es

Temas a tratar

1. Introducción a la IA y a los Sistemas Expertos
2. Conceptos básicos estadísticos
3. Redes Bayesianas
4. Cadenas de Markov
5. Proyecto ESIDE-DEPIAN
6. Aplicación: HUGIN
7. Programación: openPNL

Inteligencia Artificial

Inteligencia Artificial: parte de la Ciencia que se ocupa del diseño de sistemas de computación inteligentes.

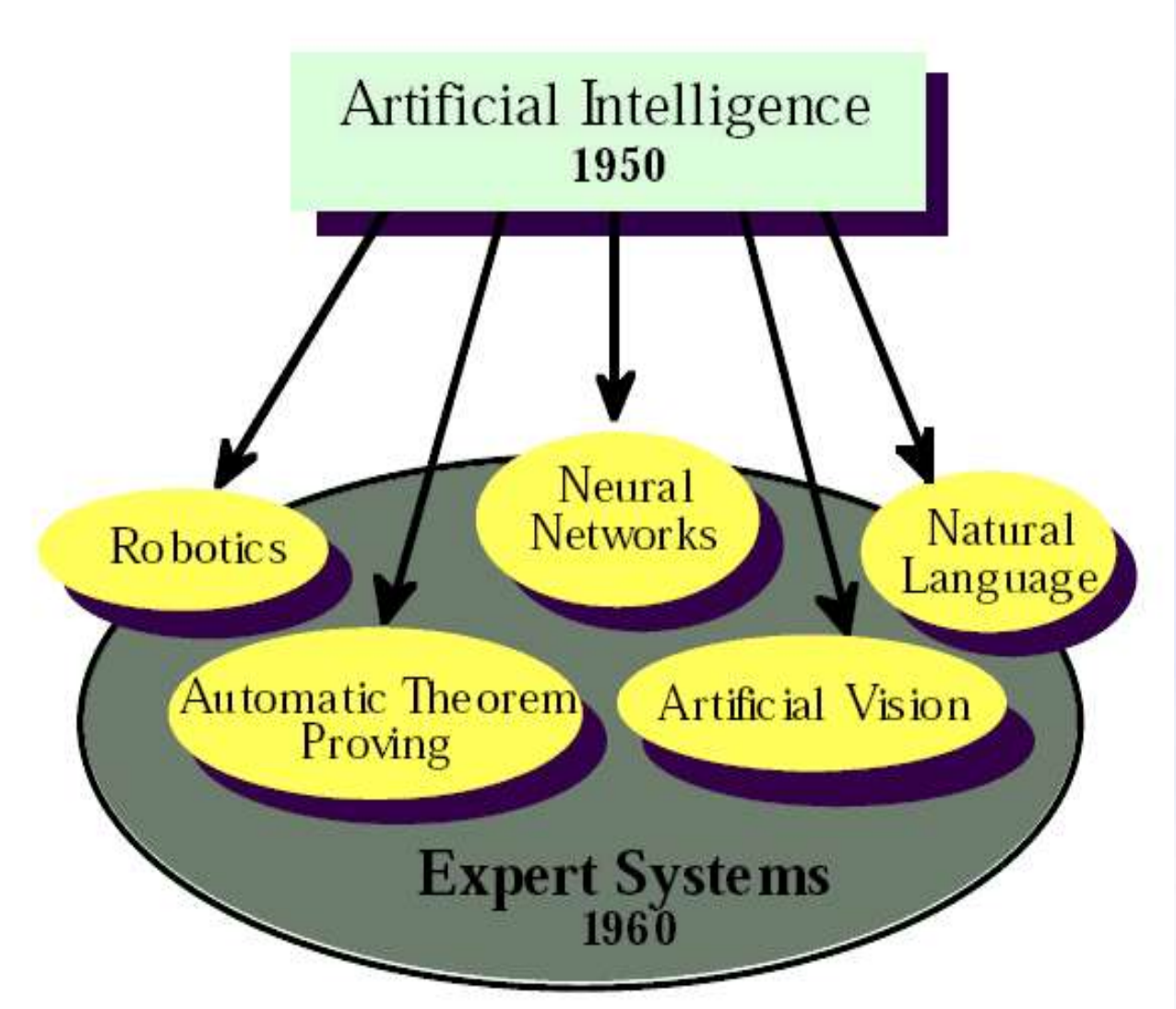
Pero...¿qué entendemos por inteligencia?

- Comprensión del lenguaje
- Aprendizaje
- Razonamiento
- Resolución de problemas
- ...

Inteligencia Artificial

1943-1956
Estudios centrados en Redes Neuronales.
Demostración de Teoremas y Ajedrez.
1952-1969:
Creación de sistemas que resuelvan cualquier problema.
Avances limitados por los recursos computacionales.
1966-1974:
Algoritmos genéticos.
Problemas en la representación del conocimiento.
1969-1979:
DENDRAL, MYCIN...
1980-1988:
Las empresas se interesan por la IA. Control industrial y robótica.
1988-... :
Resolución de problemas del mundo real.
Sistemas especializados que cooperan ¿remember UNIX?

Inteligencia Artificial



Sistemas Expertos

Sistema Experto: sistema informático que simula a los expertos humanos en un área específica dada.

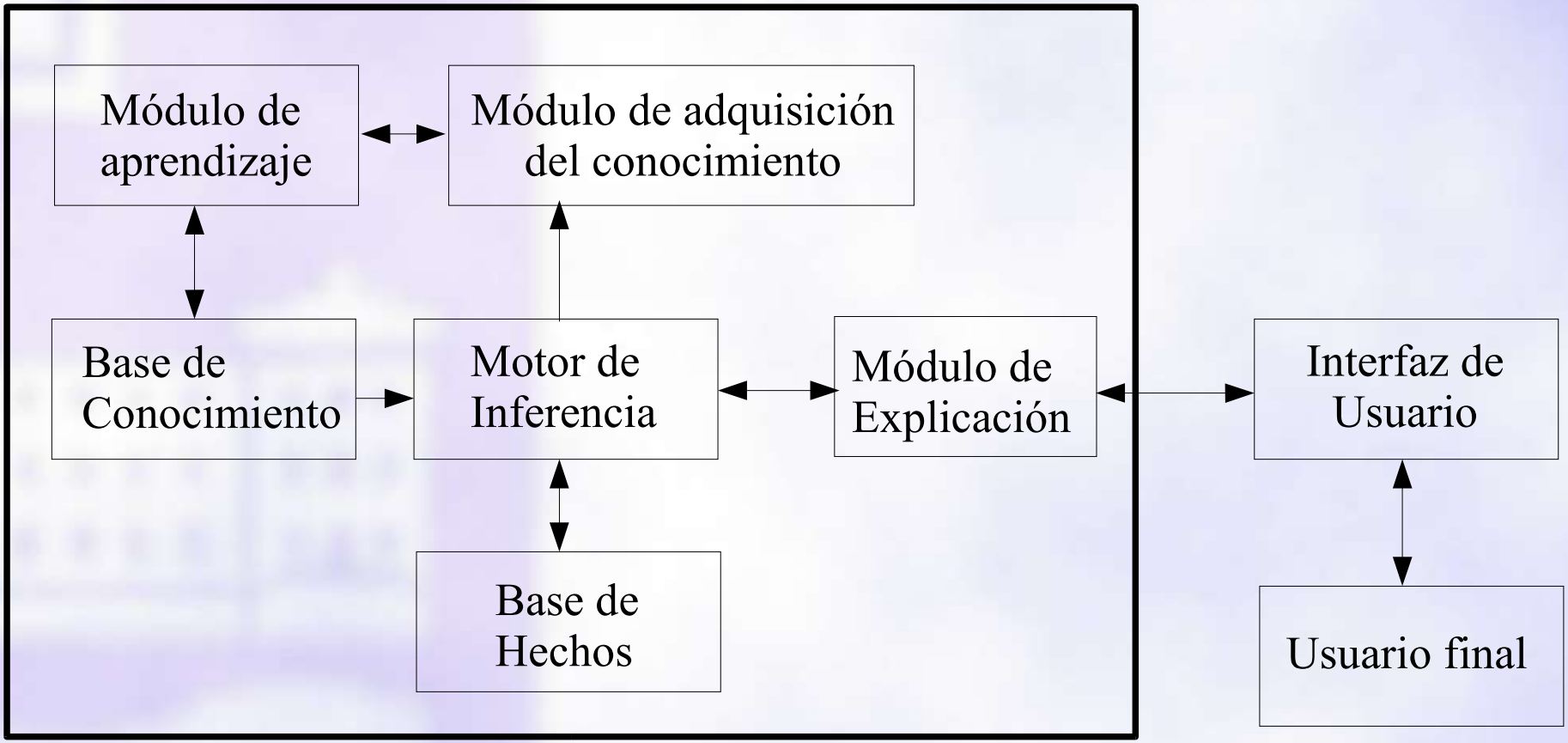
Debería ser capaz de:

- Procesar y memorizar información
- Aprender y razonar en determinadas situaciones
- Comunicación con el experto u otros sistemas
- Toma de decisiones
- ...

Sistemas Expertos

- Meteorología.
- Transacciones bancarias.
- Control de tráfico en una ciudad.
- Diagnóstico médico.
- Enfoque automático de imágenes fotográficas.
- Diagnóstico de problemas en automóviles.
- Interfaces de ordenador en lenguaje hablado.
- Gestión distribuida de redes de ordenador.
- Urbanismo y Gestión del territorio.
- Administración local.
- Navegación terrestre y marítima.

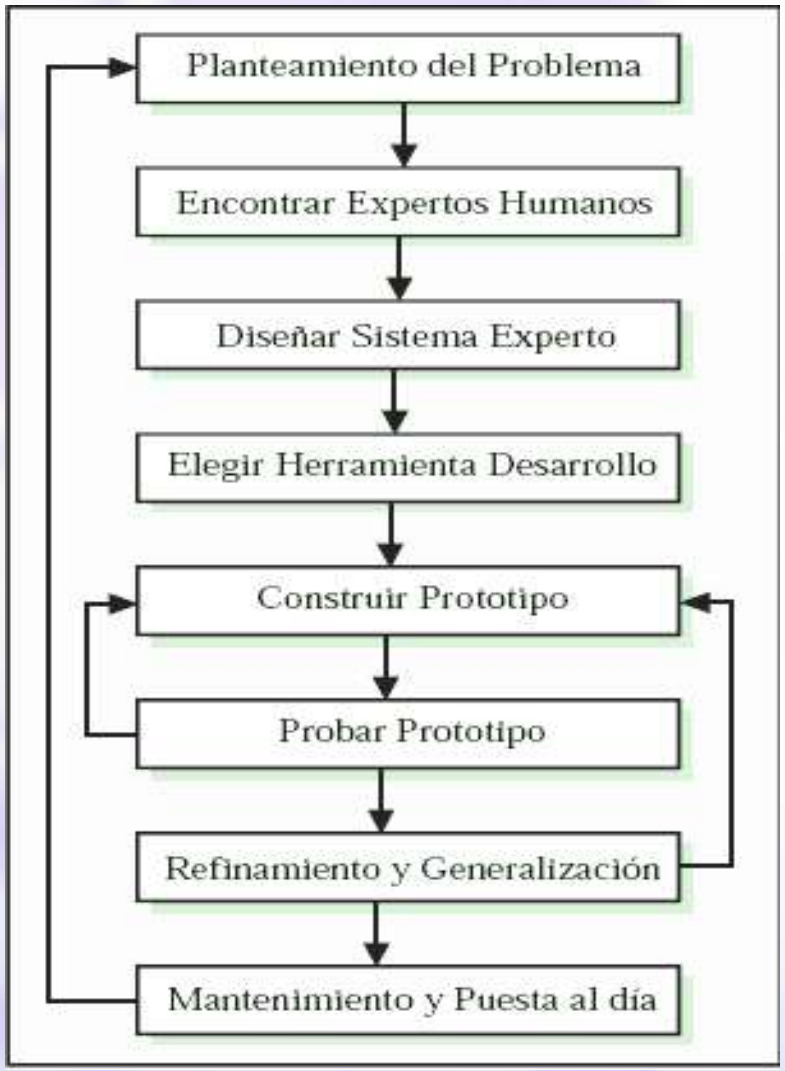
Sistemas Expertos



Sistemas Expertos

Base de Conocimiento
Conocimientos del experto humano codificado (estático).
Base de Hechos
Memoria temporal de trabajo (dinámico).
Motor de Inferencia
Combina BC y BH para deducir nuevos hechos => resolver problema.
Interfaz de Usuario
Comunicación entre el SE y el usuario final.
Módulo de Explicación
Justificación y explicación de los resultados obtenidos.
Módulo de Adquisición de Conocimiento
Añadir nuevo conocimiento a la BC.
Módulo de Aprendizaje
Aprender a partir de la resolución de problemas.

Sistemas Expertos



Fases del desarrollo de un Sistema Experto, por Weiss y KuliKowski en 1984.

Sistemas Expertos

Mediados de los sesenta (IA): Alan Newell y Herbert Simon desarrollan **GPS** (General Problem Solver).

Centrarse en áreas de conocimientos muy concretos => surgen los SE.

DENDRAL (Ledeberg y Feigenbaum): considerado el primer SE (1967), deducía estructuras químicas a partir de su análisis espectográfico.

MYCIN (Univ. Stanford): diagnóstico de enfermedades infecciosas (1970-80). Separa la base de conocimiento del motor de inferencia.

IF the infection is pimary-bacteremia
AND the site of the culture is one of the sterile sites
AND the suspected portal of entry is the gastrointestinal tract
THEN there is suggestive evidence (0.7) that infection is bacteroid.

Sistemas Expertos

MYCIN da lugar a otros sistemas expertos como:

- **EMYCIN** : contiene el sistema de manejo de la BC y la inferencia del MYCIN, para su uso en otros SE.
- **SACON** : generación de estructuras de ingeniería.
- **PUFF** : estudio médico de enfermedades pulmonares.
- **GUIDON**: elección de tratamientos terapeúticos.

HEARSAY: identificación de la palabra hablada.
PROSPECTOR: búsqueda de yacimientos minerales (Bayes).

A partir de 1980 se “ponen de moda” y se extiende su uso en industrias y empresas.

Sistemas Expertos

Tipos de sistemas expertos (según naturaleza del problema):

- Deterministas => el estado actual depende del estado anterior y las acciones sobre el entorno. Son los Sistemas Expertos basados en reglas, que usan un mecanismo de *razonamiento lógico* para sacar sus conclusiones.

- Estocásticos => sistemas en los que existe incertidumbre, por lo que necesita ser tratada. Son los Sistemas Expertos Probabilísticos y la estrategia de razonamiento usada es el *razonamiento probabilístico*.

Sistemas Expertos

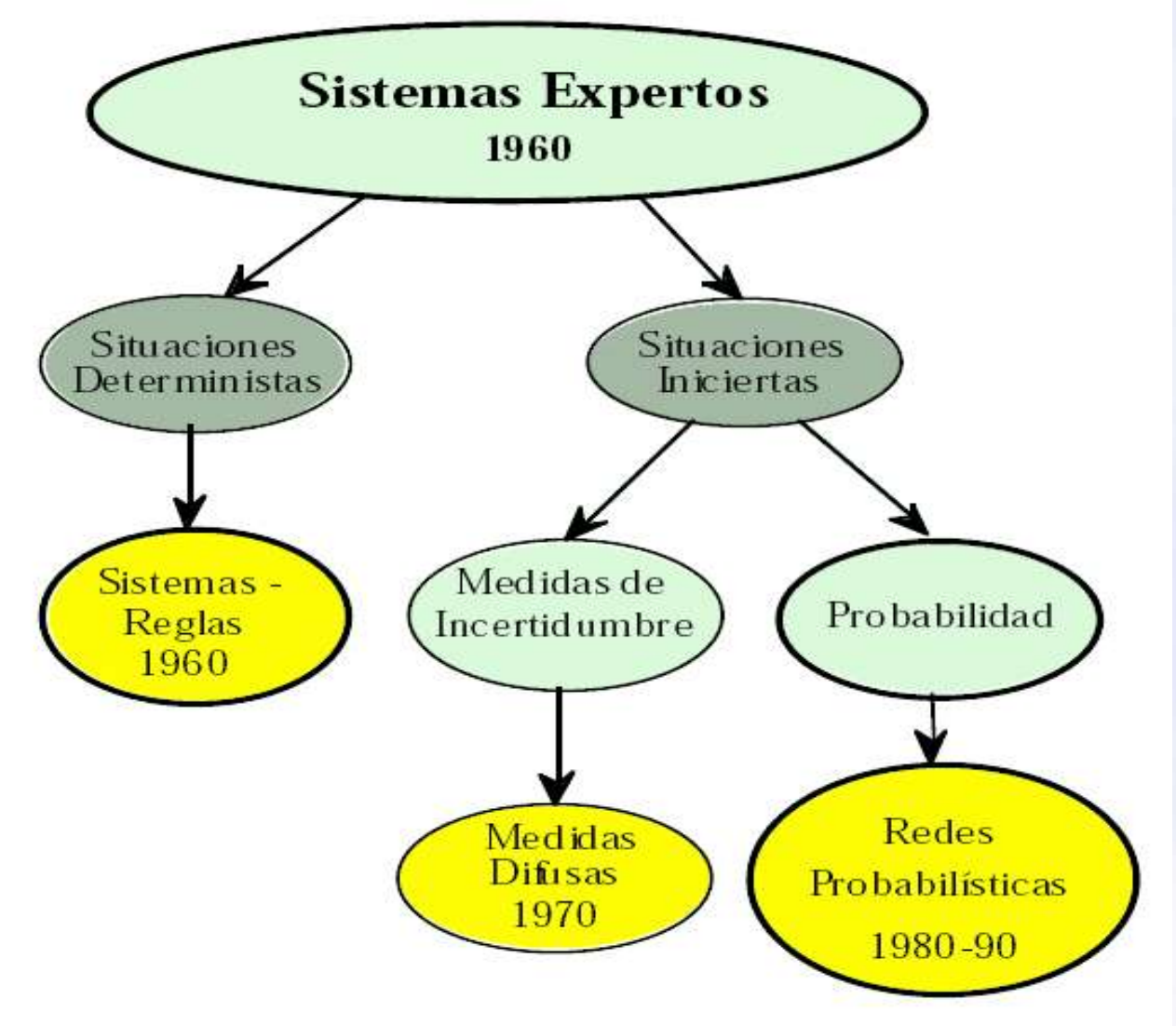
Pero...¿Qué es exactamente la incertidumbre?

Se define como la falta de certidumbre o certeza, siendo certeza el conocimiento seguro y claro de algo.

¿En qué situaciones se da incertidumbre?

- Cuando los hechos o datos pueden no ser conocidos con exactitud (por ej, un paciente puede no estar seguro de haber tenido fiebre la noche pasada) => subjetividad, imprecisión, errores, datos ausentes...
- Cuando el conocimiento no es determinista. Por ej, las relaciones entre enfermedades y síntomas; un mismo conjunto de síntomas puede estar asociado a varias enfermedades.

Sistemas Expertos



Sistemas Expertos

En los primeros Sistemas Expertos, se usaba la probabilidad para tratar la incertidumbre, pero al encontrarse algunos problemas por el uso incorrecto de algunas hipótesis, se desechó.

Con la aparición de redes probabilísticas (Redes Bayesianas y Cadenas de Markov, principalmente) el uso de la probabilidad para el tratamiento de la incertidumbre ha vuelto a ser aceptado y hoy en día es la forma más usada.

Sistemas Expertos

Sistemas Expertos basados en reglas

- Una regla es una afirmación lógica que relaciona información conocida con otra que puede ser inferida o se sabe que es cierta.

- Una regla se compone de la premisa y el consecuente.
  Premisa: condiciones para que la regla se ejecute.
  Consecuente: conclusiones deducidas.

Ejemplo de regla:

IF TarjetaNoValida
THEN PagoNoAutorizado
ELSE PagoAutorizado

Sistemas Expertos

El Motor de Inferencia

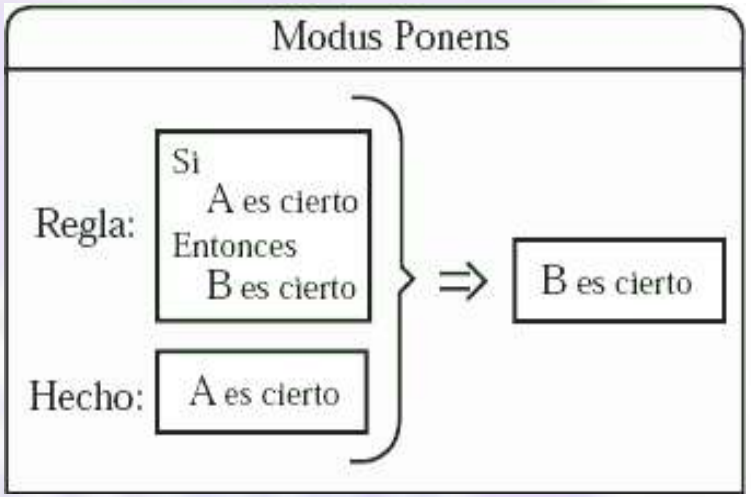
La Inferencia permite deducir nuevo conocimiento a partir de conocimiento que se sabe que es cierto.

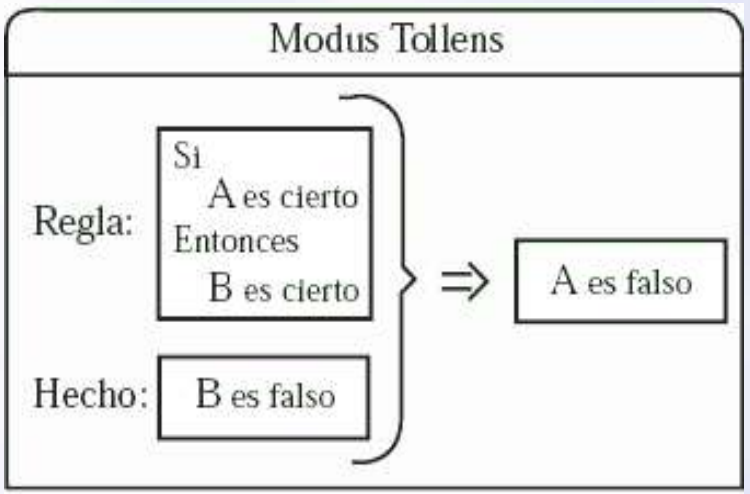
Usa la Base de Hechos y el Conocimiento Base para obtener nuevas conclusiones o hechos.

Existen diferentes reglas de inferencia (Modus Ponens, Modus Tollens) y diferentes estrategias de inferencia (Encadenamiento de reglas hacia delante y hacia atrás).

Sistemas Expertos

Reglas de inferencia





Se examina la premisa y si es cierta, la conclusión pasa a formar parte del conocimiento.

Se examina la conclusión y si es falsa, se concluye que la premisa también es falsa.

Sistemas Expertos

Encaminamiento (de reglas) hacia delante

- Obtiene nuevos hechos a partir de la evaluación de reglas.
- Comienza insertando unos hechos iniciales en la BH.
- Se exploran las reglas de la BC y se añaden nuevos hechos a la BH. Termina cuando no se cumple ninguna regla.
- El objetivo es *deducir* todo el conocimiento posible.

Encaminamiento (de reglas) hacia atrás

- Deducir el conocimiento necesario para demostrar un hecho.
- Comienza fijando un hecho o meta a demostrar.
- Se busca la regla que contiene dicho hecho como consecuente y se demuestran los hechos del antecedente de la regla.
- El objetivo es *demostrar* una meta.

Conceptos básicos estadísticos

- Haremos un breve repaso al método probabilista clásico.

Variable aleatoria: aquella que toma valores, que a priori, no conocemos con certeza.

Por ej: cogemos 2 personas al azar. Su edad y sexo, serán 2 variables aleatorias.

Los valores de una variable van a ser siempre (est. clásica):

- Exclusivos: son incompatibles entre sí.
- Exhaustivos: cubren todas las posibilidades.

Por ej: La edad de una persona: < 18, 18-65, >65

Conceptos básicos estadísticos

Tipos de variables aleatorias:

- **Discretas**: el número de valores es finito.
  Ejemplos:
    Número de puerto TCP origen.
    Dirección IP origen.
- **Continuas**: puede asumir todos los valores posibles en cierto intervalo *a*-*b*. Ejemplos:
    Temperatura ambiente.
    Tiempo de fallo de un dispositivo.
    Distancia del robot a la pared.

Conceptos básicos estadísticos

▷ **Probabilidad conjunta.** Dado un conjunto de variables discretas $\bar{X} = \{X_1, \ldots, X_n\}$, definimos la *probabilidad conjunta* como una aplicación que a cada n-tupla $\bar{x} = (x_1, \ldots, x_n)$ le asigna un número real no negativo de modo que

$$\sum_{\bar{x}} P(\bar{x}) = \sum_{x_1} \cdots \sum_{x_n} P(x_1, \ldots, x_n) = 1 \tag{2.1}$$

▷ **Probabilidad marginal.** Dada una distribución de probabilidad conjunta $P(x_1, \ldots, x_n)$, la *probabilidad marginal* para un subconjunto de variables $\bar{X}' = \{X'_1, \ldots, X'_{n'}\} \subset \bar{X}$ viene dada por

$$P(\bar{x}') = P(x'_1, \ldots, x'_{n'}) = \sum_{x_i \,|\, X_i \notin \bar{X}'} P(x_1, \ldots, x_n) \tag{2.2}$$

▷ **Probabilidad condicional.** Dados dos subconjuntos disjuntos de variables, $\bar{X} = \{X_1, \ldots, X_n\}$ e $\bar{Y} = \{Y_1, \ldots, Y_m\}$, y una tupla $\bar{x}$ (es decir, una asignación de valores para las variables de $\bar{X}$) tal que $P(\bar{x}) > 0$, la *probabilidad condicional* de $\bar{y}$ dado $\bar{x}$, $P(\bar{y}|\bar{x})$, se define como

$$P(\bar{y} \,|\, \bar{x}) = \frac{P(\bar{x}, \bar{y})}{P(\bar{x})} \tag{2.6}$$

Conceptos básicos estadísticos

Ej: Supongamos una población de 500personas, cuya distribución por edades(X1) y sexos(X2) es:

N	Varón	Mujer	TOTAL
<18	67	68	135
18-65	122	126	248
>65	57	60	117
TOTAL	246	254	500

Probabilidades marginales:

Pjoven=P(x1j)=135/500=0'270

Pmujer=P(x2m)=254/500=0'508

Probabilidades conjutas:

P(x1j,x2m)=68/500=0'136

P(x1a,x2m)=126/500=0'252

P	Varón	Mujer	TOTAL
<18	$P(x_1^j, x_2^v) = 0'134$	$P(x_1^j, x_2^m) = 0'136$	$P(x_1^j) = 0'270$
18-65	$P(x_1^a, x_2^v) = 0'244$	$P(x_1^a, x_2^m) = 0'252$	$P(x_1^a) = 0'496$
>65	$P(x_1^t, x_2^v) = 0'114$	$P(x_1^t, x_2^m) = 0'120$	$P(x_1^t) = 0'234$
TOTAL	$P(x_2^v) = 0'492$	$P(x_2^m) = 0'508$	$1'000$

Probabilidades condicionales:

P(x1t|x2v)=P(x1t,x2v)/P(x2v)

= 0'114/0'492 = 0'23171

Conceptos básicos estadísticos

▷ **Valores independientes.** Dos valores x e y de dos variables X e Y, respectivamente, son independientes sii $P(x, y) = P(x) \cdot P(y)$.

▷ **Valores correlacionados.** Dos valores x e y de dos variables X e Y, respectivamente, están correlacionados sii no son independientes, es decir, sii $P(x, y) \neq P(x) \cdot P(y)$. Cuando $P(x, y) > P(x) \cdot P(y)$, se dice que hay correlación positiva. Cuando $P(x, y) < P(x) \cdot P(y)$, se dice que hay correlación negativa.

De los conceptos de independencia y correlación **entre valores** podemos pasar a los de independencia y correlación **entre variables**.

▷ **Variables independientes.** Dos variables X e Y son independientes sii todos los pares de valores x e y son independientes, es decir, sii

$$\forall x, \forall y, \quad P(x, y) = P(x) \cdot P(y) \tag{2.15}$$

▷ **Variables correlacionadas.** Dos variables X e Y están correlacionadas sii no son independientes, es decir, sii

$$\exists x, \exists y, \quad P(x, y) \neq P(x) \cdot P(y) \tag{2.16}$$

Conceptos básicos estadísticos

- **Valores condicionalmente independientes.** Sean tres valores x, y y z de las variables X, Y y Z, respectivamente, tales que $P(z) > 0$; x e y son condicionalmente independientes dado z sii $P(x, y|z) = P(x|z) \cdot P(y|z)$.

- **Variables condicionalmente independientes.** Las variables X e Y son condicionalmente independientes dada una tercera variable Z sii todo par de valores x e y es condicionalmente independiente para cada z tal que $P(z) > 0$; es decir, sii

$$\forall x, \forall y, \forall z, \quad P(z) > 0 \implies P(x, y \,|\, z) = P(x \,|\, z) \cdot P(y \,|\, z) \tag{2.17}$$

- **Separación.** La variable Z *separa* las variables X e Y sii éstas dos últimas son condicionalmente independientes dada Z.

Estas definiciones son igualmente válidas para conjuntos de variables $\bar{X}$, $\bar{Y}$ y $\bar{Z}$.

Conceptos básicos estadísticos

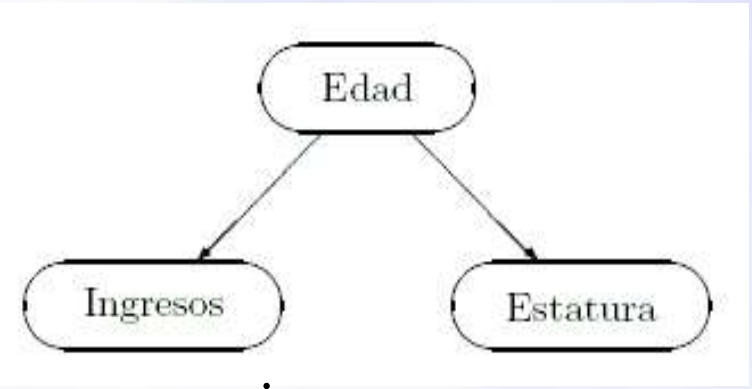
Representación gráfica de dependencias/independencias

Cuando una variable ejerce *influencia causal* sobre otra, se traza una flecha entre ambas:

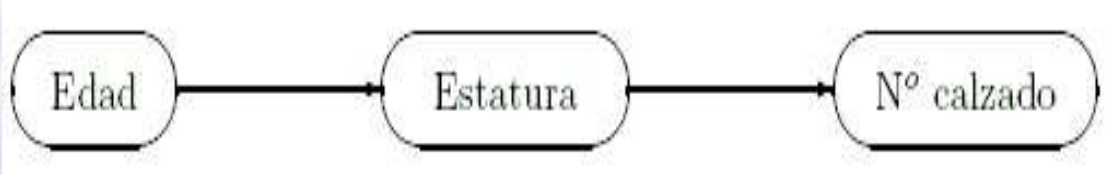
Figura 2.1: Dos variables independientes.

Figura 2.2: Dependencia causal entre dos variables.

Conceptos básicos estadísticos



La edad influye en la estatura
La edad influye en los ingresos
Hay correlación (a priori) entre estatura e ingresos
Cuando se conoce la edad, desparece dicha correlación (indep)



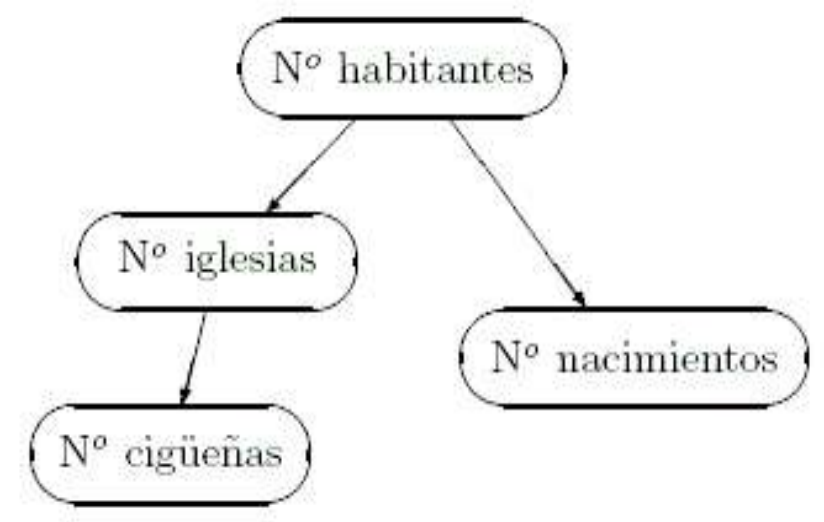
La edad influye en la estatura
La estatura influye en el nº de calzado
Hay correlación (a priori) entre edad y nº calzado
Cuando se conoce la estatura, desaparece dicha correlación (indep)

Conceptos básicos estadísticos

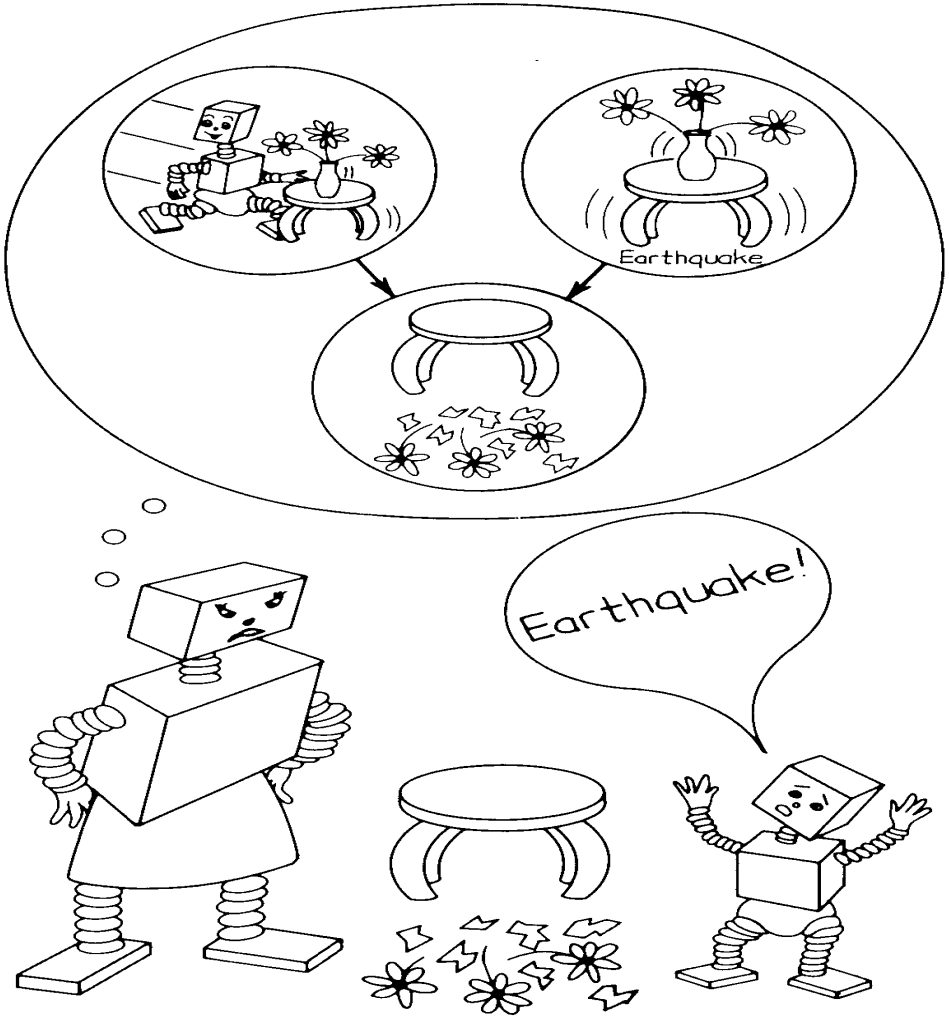
Causalidad **Vs** Correlación

Causalidad implica correlación pero no a la inversa.

Un estudio demostró que había una fuerte correlación entre el nº de cigüeñas de una ciudad y el nº de nacimientos de niños. ¿Traen las cigüeñas a los niños? ¿O es acaso la presencia de niños quien atrae a las cigüeñas?



Redes Bayesianas

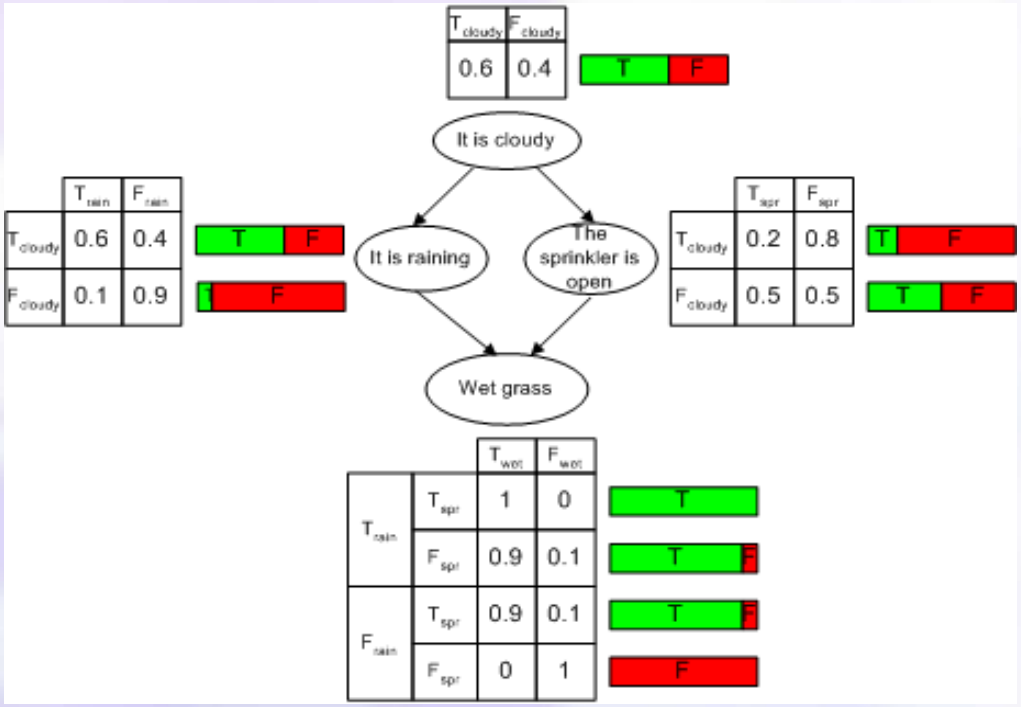


Probabilistic Reasoning in a Causal Network

[Neapolitan 90]

Redes Bayesianas

Definición: Es un grafo dirigido acíclico conexo más una distribución de probabilidad sobre sus variables.



DAG + CPD

Redes Bayesianas

Existen distintos tipos de Redes Bayesianas:

- Naive Bayes = bayes “ingenuo” o Idiot's Bayes
  Forma de “**V**” => 2^n estados en el nodo inferior
- DBNs = Redes Bayesianas Dinámicas
  Cambian con el tiempo (t, t+1, t+2...)
  Lo pasado en t, tiene relación con lo que suceda en t+1
- Redes Gaussianas = distribución gaussiana
  Para nodos con variables contínuas
- Cadenas de Markov = subconjunto de las RB

Ejemplos: clippo, meteorología, aire acondicionado...

Redes Bayesianas

Teorema de Bayes:

Teorema 2.17 (**Teorema de Bayes**) Dadas dos variables X e Y, tales que $P(x) > 0$ para todo x y $P(y) > 0$ para todo y, se cumple

$$P(x \mid y) = \frac{P(x) \cdot P(y \mid x)}{\sum_{x'} P(x') \cdot P(y \mid x')} \qquad (2.21)$$

En la práctica, se utiliza para conocer la probabilidad *a posteriori* de cierta variable de interés dado un conjunto de hallazgos (ya no es condicional).

Redes Bayesianas

Hallazgo: determinación del valor de una variable, a partir de un dato (una observación, una medida...).

Evidencia: conjunto de todos los hallazgos disponibles en un determinado momento.

Probabilidad a priori: es la probabilidad de una variable o subconjunto de variables cuando no hay ningún hallazgo. Coincide con la probabilidad marginal $P(x)$.

Probabilidad a posteriori: es la probabilidad de una variable o subconjunto de variables dada la evidencia e. Se trata de la probabilidad condicional $P(x|e)$.

Redes Bayesianas

Ejemplo: congreso con 50 personas de 3 universidades (23,18,9).

1ª:30% Ciencias, 40% de Ing, 25% Humanidades y 5% Economía.

2ª:25% Ciencias, 35% de Ing, 30% Humanidades y 10% Economía

3ª:20% Ciencias, 50% de Ing, 10% Humanidades y 20% Economía.

A la salida, nos encontramos un profesor *a)*¿Probabilidad de que sea de la tercera universidad? *b)*Y si nos enteramos de que es de Economía ¿Cuál sería?

Solución:

a) Probabilidad a priori : *P(x)=9/50=0'18*=**18%**

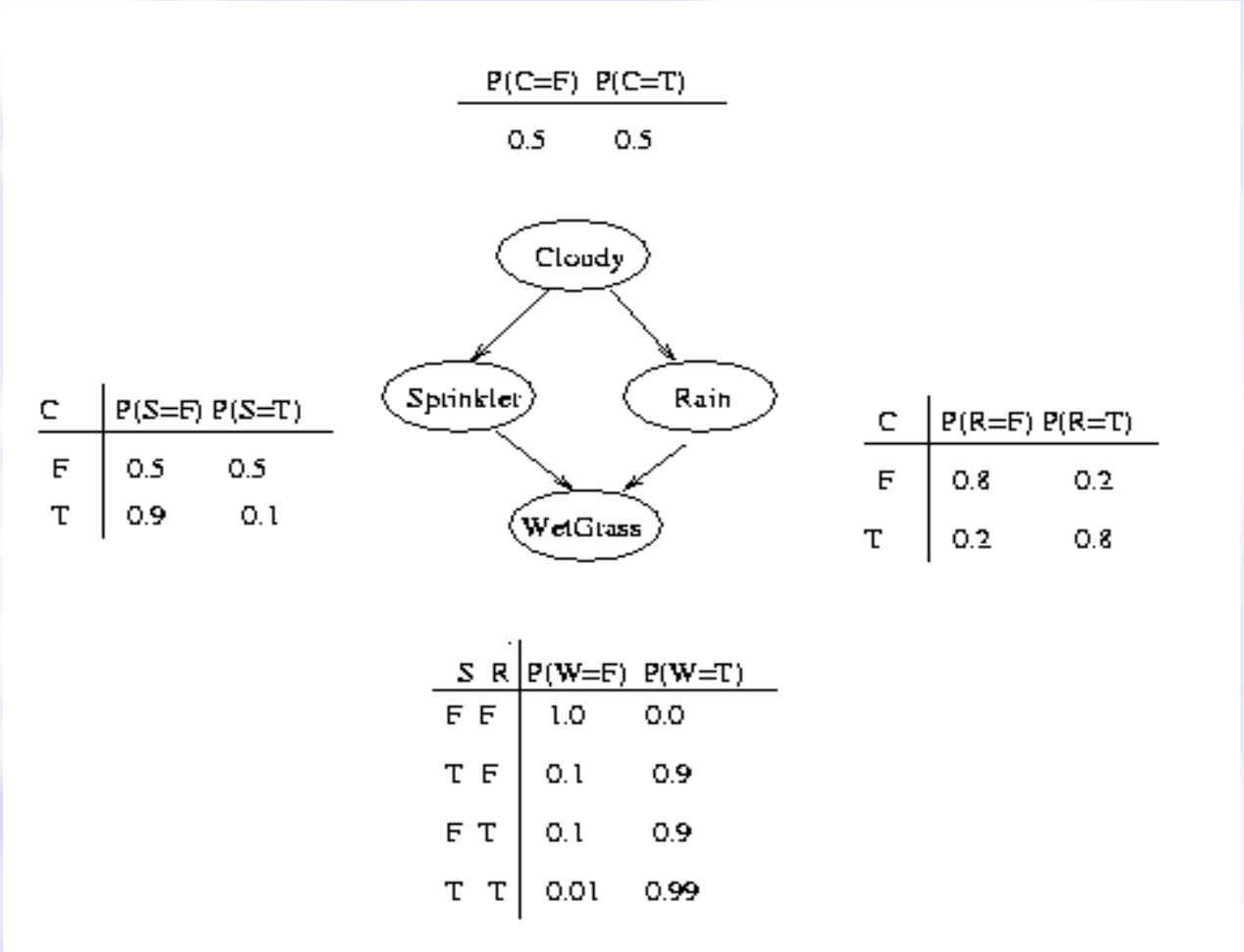
b) Para esta respuesta, hacemos la siguiente tabla(x=uni|y=especialidad):

$P(y \mid x)$	x^1	x^2	x^3
y^c	0'30	0'25	0'20
y^i	0'40	0'35	0'50
y^h	0'25	0'30	0'10
y^e	0'05	0'10	0'20

Aplicando Bayes:

$$P^*(x^3) = P(x^3 \mid y^e) = \frac{P(x^3) \cdot P(y^e \mid x^3)}{\sum_x P(x) \cdot P(y^e \mid x)} = \frac{0'18 \cdot 0'20}{0'46 \cdot 0'05 + 0'36 \cdot 0'10 + 0'18 \cdot 0'20} = 0'379 = \mathbf{37'9\%}$$

Redes Bayesianas



HUGIN

$P(C=true)=0'5$

$P(S=true|C=false)=0'5$

$P(S=false|C=true)=0'9$

$P(R=true|C=true)=0'8$

$P(R=false|C=true)=0'2$

$P(W=true|S=t,R=f)=0'9$

$P(W=false|S=t,R=t)=0'01$

Redes Bayesianas

- Aprendizaje paramétrico

  Aprende las probabilidades de la red en base a casos dados, por ej un archivo pasado con los valores de cada variable.

  Existen distintos algoritmos de aprendizaje, entre ellos:

  - **EM** (Expansión-Maximización):

    No necesita datos completos para el aprendizaje.

    Contiene 2 fases:

    1º Expansión: calculo de todas las probabilidades posibles por toda la red.

    2º Maximización: se escoge la mayor probabilidad.

  - **ML** (Maximum Likelihood):

    Necesita de datos completos para poder aprender.

    Es parecido al EM, pero sin la primera fase (E).

Redes Bayesianas

- Aprendizaje estructural

  Estos algoritmos son capaces de aprender enlaces.

  Existen 2 tipos de aprendizaje de estructural:

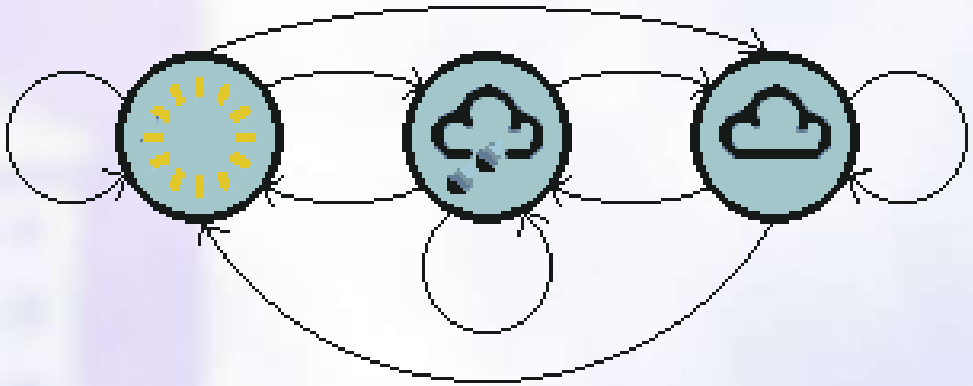
  - Basados en tests de independencia (algoritmos PC,NPC...)
  - Puntuación y búsqueda (Score & Search)

    Para puntuar, se va penalizando para conseguir cuál es el grafo más óptimo (AIC,BIC).

    Para la búsqueda, algoritmos como K2, LK2,Montecarlo, B...

    Algunos algoritmos, no reconsideran los enlaces ya existentes en la red (puede ser interesante).

Cadenas de Markov



Cadenas de Markov

Encontrar un patrón que aparece durante un espacio de tiempo

- Secuencia de comandos de un usuario ante un PC
- Determinadas palabras en un texto (spam ;))

Un sistema determinista, como por ejemplo un semáforo, es fácilmente comprensible y el siguiente estado solo depende del estado anterior: luz roja -> luz verde -> luz ambar -> luz roja...

El problema llega con sistemas no deterministas => axioma de Markov (el siguiente estado solo depende del anterior).

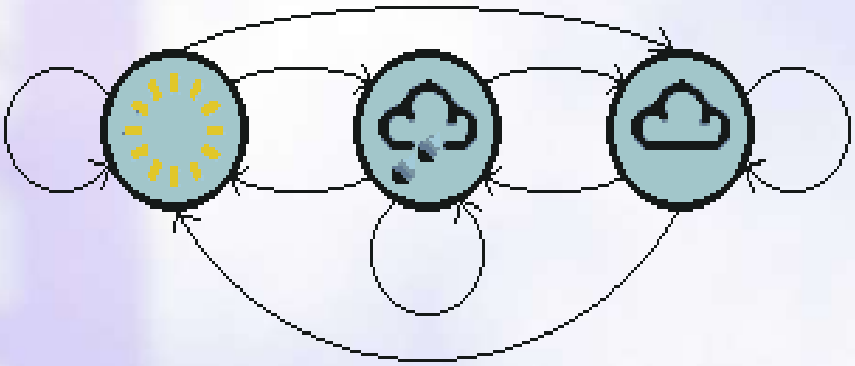
Cadenas de Markov

Un proceso de Markov es un proceso que se va moviendo de estado en estado dependiendo exclusivamente de los N estados anteriores.

Dependiendo del número de estados N que se tengan en cuenta:

n-gramas: unigram, bigram, trigram...

La diferencia de un proceso de orden 1 con un proceso determinista es que la elección del estado final se hace de forma probabilística, no determinista.



Cadenas de Markov

En un proceso de Markov aparte de los estados, está la matriz de transiciones entre estados, que indica la probabilidad de pasar de un estado a otro o a sí mismo.

weather yesterday \ weather today	Sun	Cloud	Rain
Sun	0.5	0.25	0.25
Cloud	0.375	0.125	0.375
Rain	0.125	0.625	0.375

Además hace falta tener una matriz que indique como se encuentra el sistema al inicio, vector de probabilidades iniciales.

Sun	Cloud	Rain
1.0	0.0	0.0

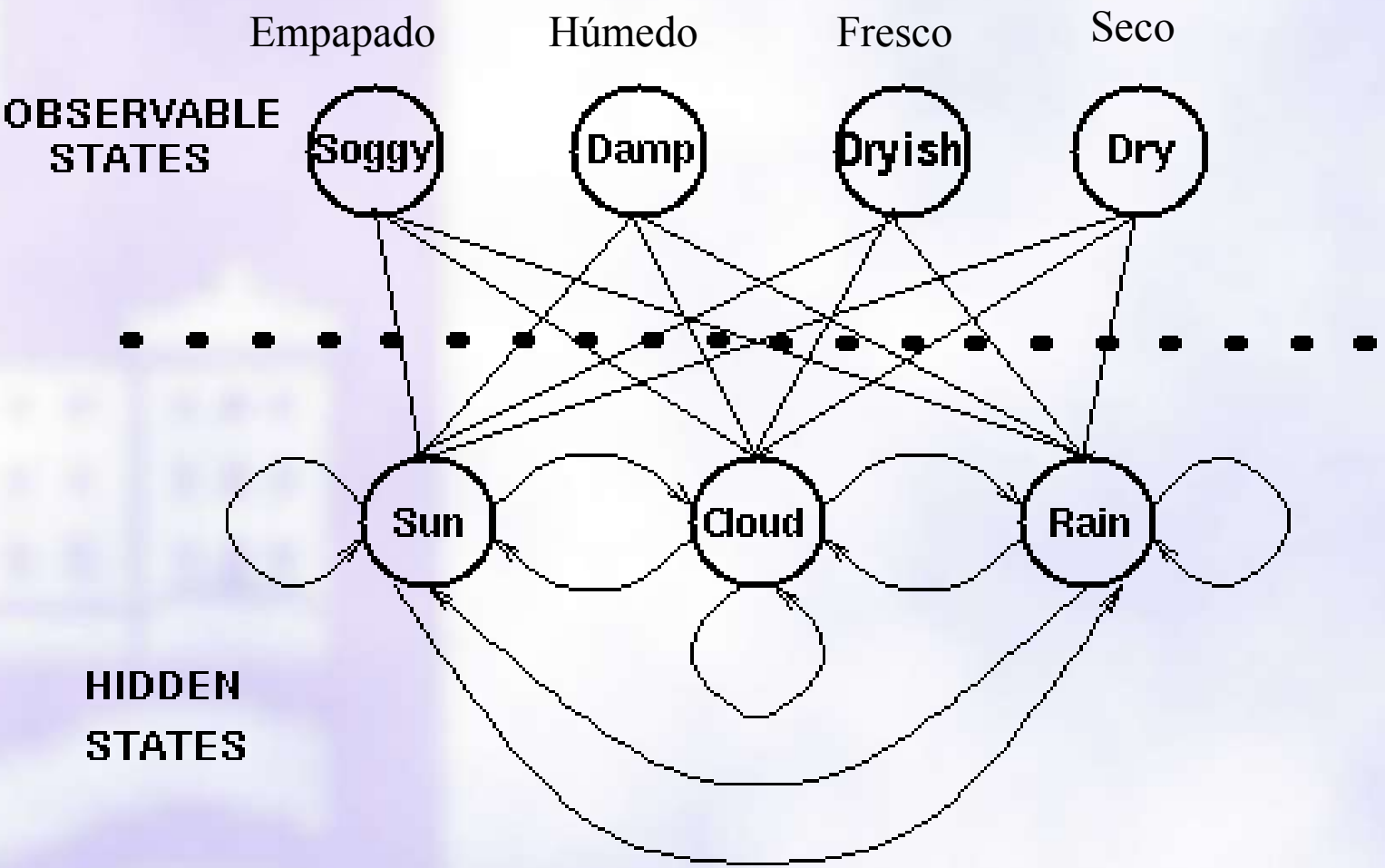
Cadenas de Markov

Cadenas de Markov Ocultas(HMM): tenemos dos partes diferenciadas y no vale para modelar simplemente con una cadena de Markov anterior. Por ej, en el caso del reconocimiento de palabras.

Sonido que oímos = cuerdas vocales+lengua+labios...

Este tipo de sistemas trabajan considerando que la producción interna de fonemas es una secuencia de **estados ocultos** (no observables) y que los sonidos resultantes son una secuencia observable generada por los estados ocultos. El número de estados ocultos puede ser diferente del número de estados observables.

Cadenas de Markov



Cadenas de Markov

La conexión entre los estados ocultos y las variables observables, símbolos, representa la probabilidad de generar un símbolo en particular desde un determinado estado oculto.

$$P(Obs|Sun)+P(Obs|Cloud)+P(Obs|Rain) = 1$$

Para representar estas probabilidades se tiene la matriz de dispersión/confusión del sistema:

		Seaweed: Dry	Dryish	Damp	Soggy
weather	Sun	0.60	0.20	0.15	0.05
	Cloud	0.25	0.25	0.25	0.25
	Rain	0.05	0.10	0.35	0.50

Cadenas de Markov

Soluciones a varios problemas:

- Encontrar la probabilidad de una determinada secuencia de símbolos (evaluación)
- Encontrar la secuencia de estados ocultos que más probablemente ha generado una secuencia de símbolos (decodificación)
- Generar un HMM dado una secuencia de observaciones (aprendizaje)

Cadenas de Markov

Evaluación: Se quiere saber qué HMM ha generado una determinada secuencia más probablemente. Se usa el **algoritmo Forward** para calcular la probabilidad de cada secuencia en cada sistema HMM y se obtiene el más probable de generar dicha secuencia.

Decodificación: Encontrar los estados ocultos que generaron la salida observada. Se usa el **algoritmo de Viterbi** para determinar la secuencia de estados ocultos más probable para una observación. Esto se emplea ampliamente en el procesado del lenguaje natural, para clasificar palabras en su clase sintáctica (adjetivo, sustantivo, verbo etc).

Aprendizaje: Generar un modelo HMM a partir de una secuencia de observaciones. Se usa el **algoritmo Forward-Backward**.

ESIDE-DEPIAN

**Entorno de Seguridad Informática de la universidad de DEusto
para la DEtección y Prevención de Intrusiones de red basado en ANomalías**

ESIDE-DEPIAN

- Proyecto del Tecnológico de Deusto
- Construir un Motor de Análisis Híbrido que tenga los puntos fuertes de:
  - Sistemas de Detección de Patrones de Uso Indebido
    - ante los ataques ya conocidos
  - Sistemas de Detección de Anomalías
    - ante ataques desconocidos o variantes de ataques ya documentados.

ESIDE-DEPIAN

- Objetivos principales:
  - Aprendizaje no supervisado
  - Inferencia de conclusiones
- Alcanzando estos dos objetivos, se consigue:
  - Ir mas allá de la detección de intrusiones.
  - Detección y prevención ataques.
  - Requisitos de administración mínimos

ESIDE-DEPIAN

Técnicas y tecnologías usadas por ESIDE-DEPIAN:

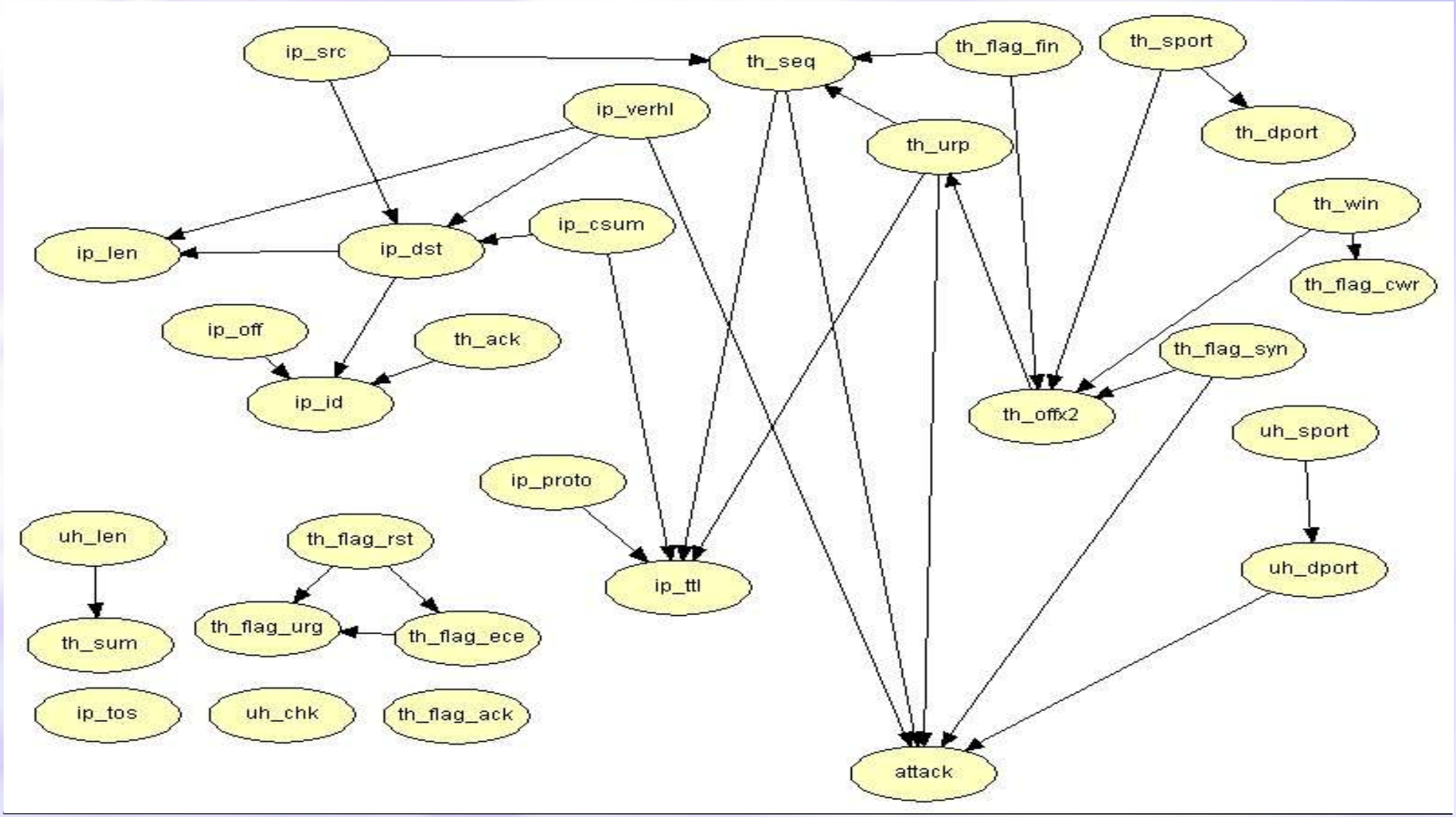
- Detección de intrusiones

  Detección basada en patrones o en firmas: **Snort**

  Detección de anomalías: **Redes Bayesianas**

- Inducción de gramáticas: **Cadenas de Markov**

ESIDE-DEPIAN



ESIDE-DEPIAN

Arquitectura:

- Maestro o Supervisores (Master)
  - Supervisan el aprendizaje de los agentes analizadores (esclavos).
  - Relación **n:m**
  - Snort modificado para el envío de información.
- Esclavos o Analizadores (Slave)
  - Reciben paquetes etiquetados del/los maestros
  - Analizan en dos fases: Inferencia y Aprendizaje
- Comunicación cifrada entre maestros y esclavos

ESIDE-DEPIAN

- Reacción temprana a ataques documentados
  - Gracias al esquema maestro-esclavo
- Capacidad de reacción ante ataques no documentados y de adquisición de nuevo conocimiento
  - Gracias al “poder” de las redes bayesianas
- Uso de un estándar “de facto”
  - El uso de Snort como maestro proporciona una base sólida para el aprendizaje del sistema.

openPNL

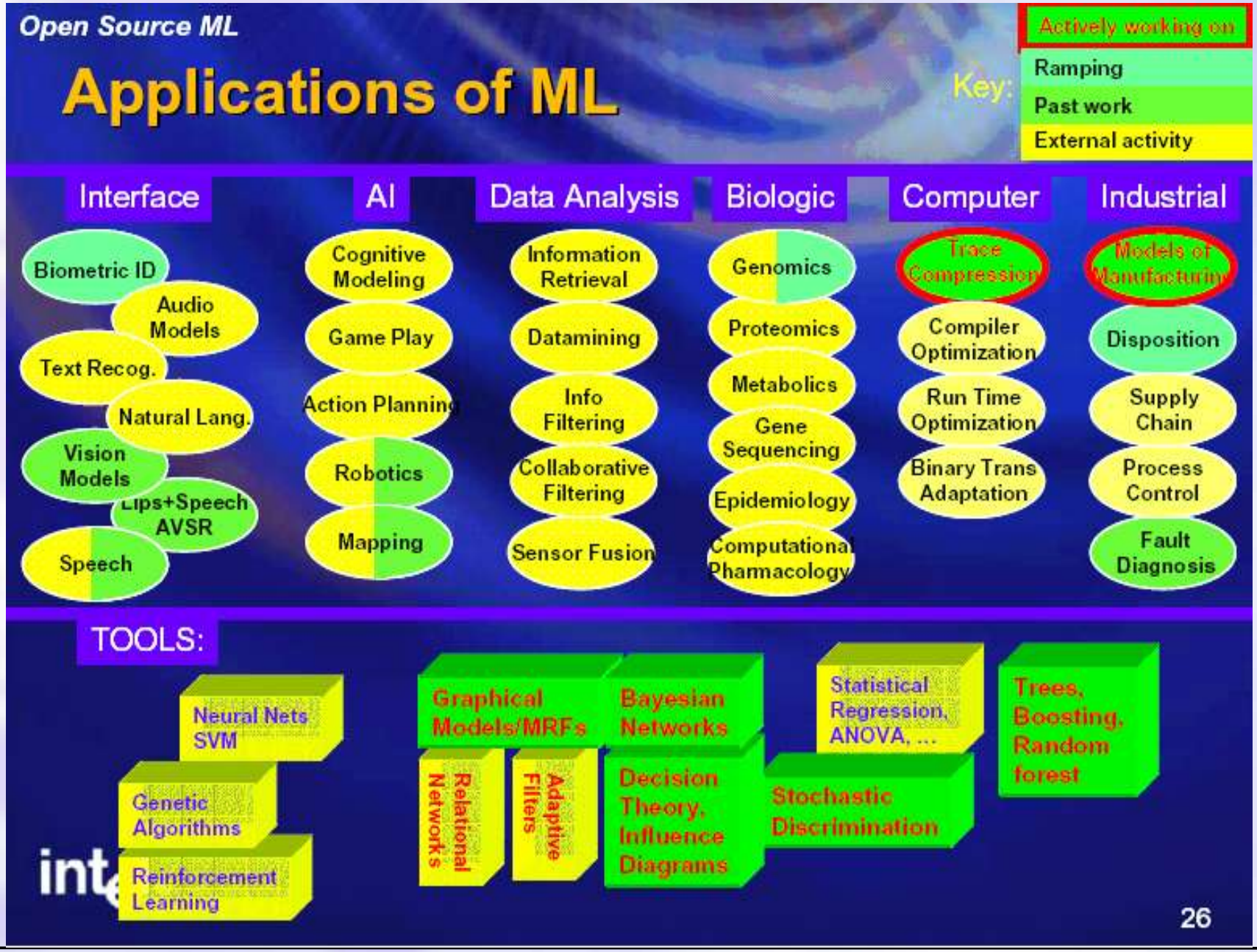
- Librerías OpenSourceMachineLearning de Intel:
  - openCV : Computer Vision
  - openAVSR: Audio Visual Speech Recognition
  - openPNL : Probabilistic Network Library

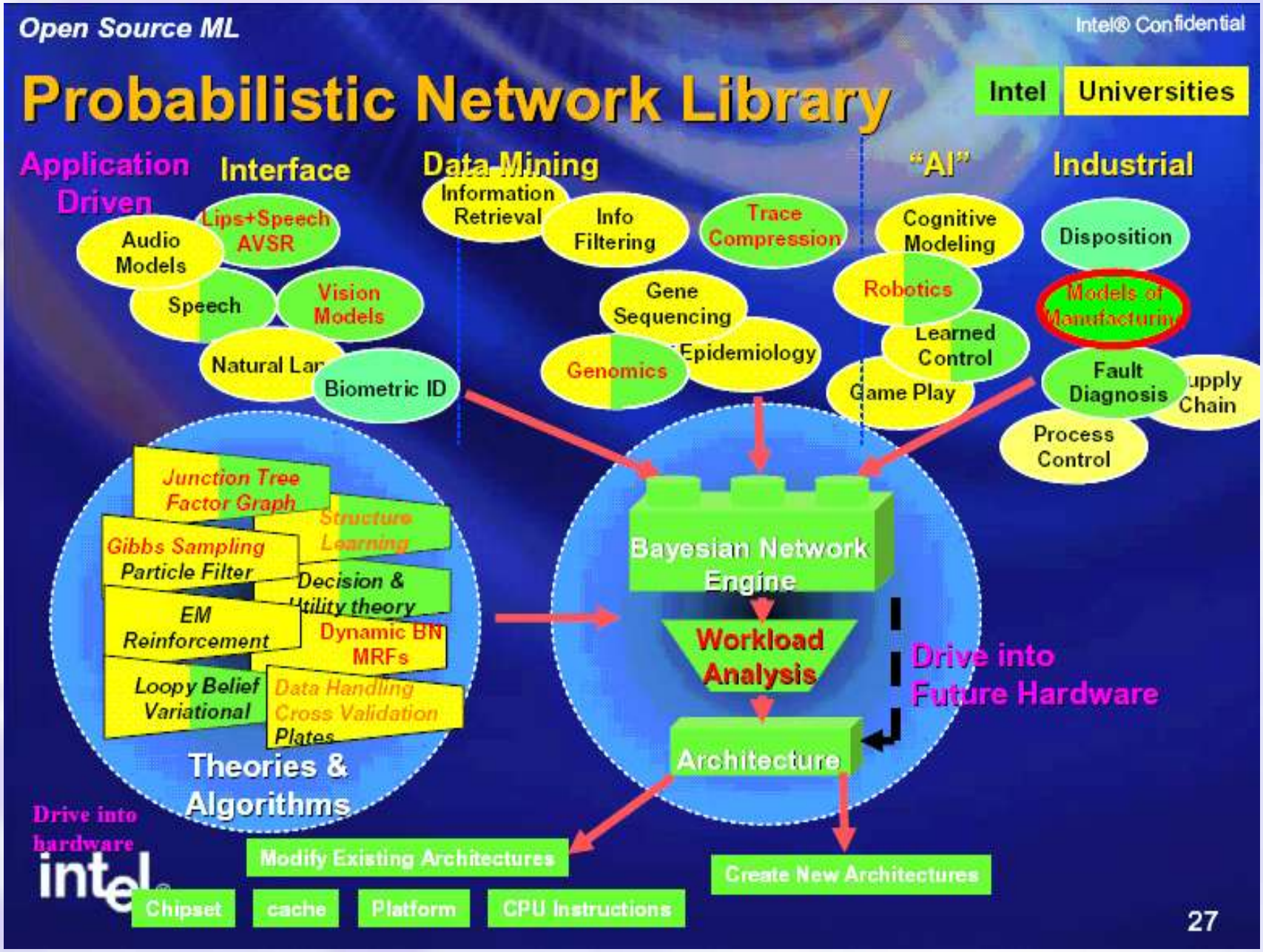
- Librería desarrollada en C/C++. Licencia estilo BSD.
- API disponible en C++ y MatLab (swig).

http://www.intel.com/research/mrl/pnl/
http://www.sourceforge.net/projects/openpnl/

openPNL

Intel Labs: USA, Rusia y China


Open Source ML
Applications of ML
Key:
Actively working on
Ramping
Past work
External activity
Interface
AI
Data Analysis
Biologic
Computer
Industrial
Biometric ID
Audio Models
Text Recog.
Natural Lang.
Vision Models
Lips+Speech AVSR
Speech
Cognitive Modeling
Game Play
Action Planning
Robotics
Mapping
Information Retrieval
Datamining
Info Filtering
Collaborative Filtering
Sensor Fusion
Genomics
Proteomics
Metabolics
Gene Sequencing
Epidemiology
Computational Pharmacology
Trace Compression
Compiler Optimization
Run Time Optimization
Binary Trans Adaptation
Models of Manufacturing
Disposition
Supply Chain
Process Control
Fault Diagnosis
TOOLS:
Neural Nets SVM
Genetic Algorithms
Reinforcement Learning
Graphical Models/MRFs
Bayesian Networks
Relational Networks
Adaptive Filters
Decision Theory, Influence Diagrams
Statistical Regression, ANOVA, ...
Stochastic Discrimination
Trees, Boosting, Random forest
26


Open Source ML
Intel® Confidential
Probabilistic Network Library
Intel
Universities
Application Driven
Interface
Data Mining
"AI"
Industrial
Audio Models
Lips+Speech AVSR
Speech
Vision Models
Natural Lan
Biometric ID
Information Retrieval
Info Filtering
Trace Compression
Gene Sequencing
Genomics
Epidemiology
Cognitive Modeling
Robotics
Learned Control
Game Play
Disposition
Models of Manufacturing
Fault Diagnosis
upply Chain
Process Control
Junction Tree
Factor Graph
Structure Learning
Gibbs Sampling
Particle Filter
Decision & Utility theory
EM
Reinforcement
Dynamic BN
MRFs
Loopy Belief
Variational
Data Handling
Cross Validation
Plates
Theories & Algorithms
Bayesian Network Engine
Workload Analysis
Drive into Future Hardware
Architecture
Drive into hardware
intel
Modify Existing Architectures
Create New Architectures
Chipset
cache
Platform
CPU Instructions
27

Intel's OpenPNL Beta Release Functionality Matrix

Representations →	Bayesian Networks				DBNs				MRFs			Factor Graphs		
Parameter Types → / Algorithms ↓	Discrete	Gaussian	Mixed	Tree CPD[1]	Discrete	Gaussian	Mixed	Tree CPD[1]	Discrete	Gaussian	Mixed	Discrete	Gaussian	Mixed
Junction Tree Inference	Y	N	N	N	N	N	N	N	n/a	n/a	n/a	n/a	n/a	n/a
1.5 Junction Tree Inference	n/a	n/a	n/a	n/a	Y	N	N	N	n/a	n/a	n/a	n/a	n/a	n/a
Sum/Max Product	Y	Y	N	N	N	N	N	N	Y	Y	N	Y	Y	N
BK Inference	n/a	n/a	n/a	n/a	Y	N	N	N	n/a	n/a	n/a	n/a	n/a	n/a
Gibbs Sampling	Y	Y	Y	N	N	N	N	N	Y	Y	N	N	N	N
Gibbs Samp w/ Annealing	Y	Y	Y	N	N	N	N	N	Y	Y	N	N	N	N
ML Param Learning (Complete Data)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
EM Param Learning	Y	Y	Y[3]	N	Y	Y	Y	N	N	N	N	N	N	N
Greedy Structure Learning (BIC)	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	N	N	N	N	N	N
Exhaustive Structure Learning (BIC)	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	Y[2]	N	N	N	N	N	N

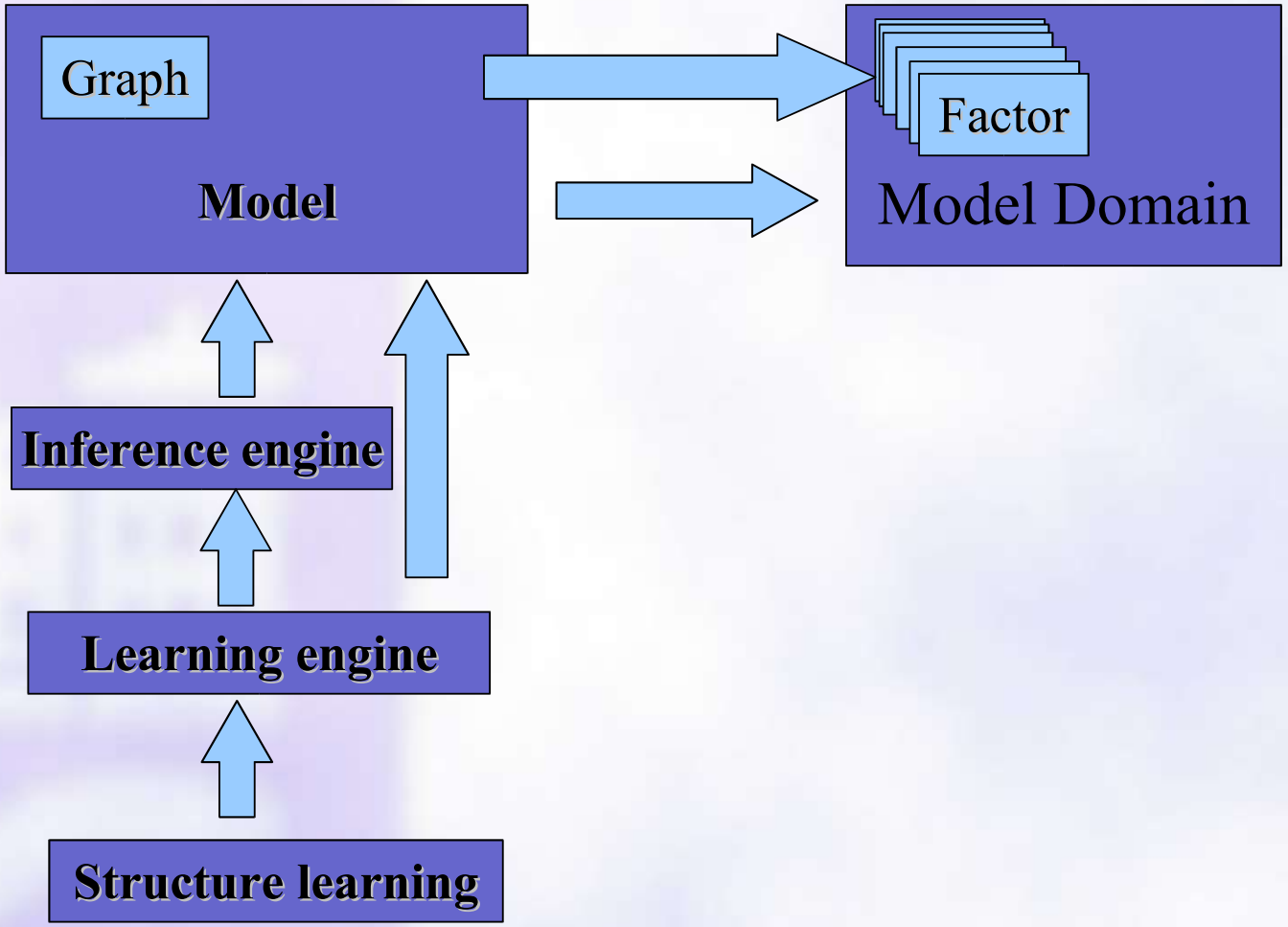
Footnotes:

1. AKA "Context-specific independence". See Friedman and Goldszmidt, UAI-96.
2. Complete data only, BIC criterion. No priors implemented yet.
3. Continuous vars must be observed.

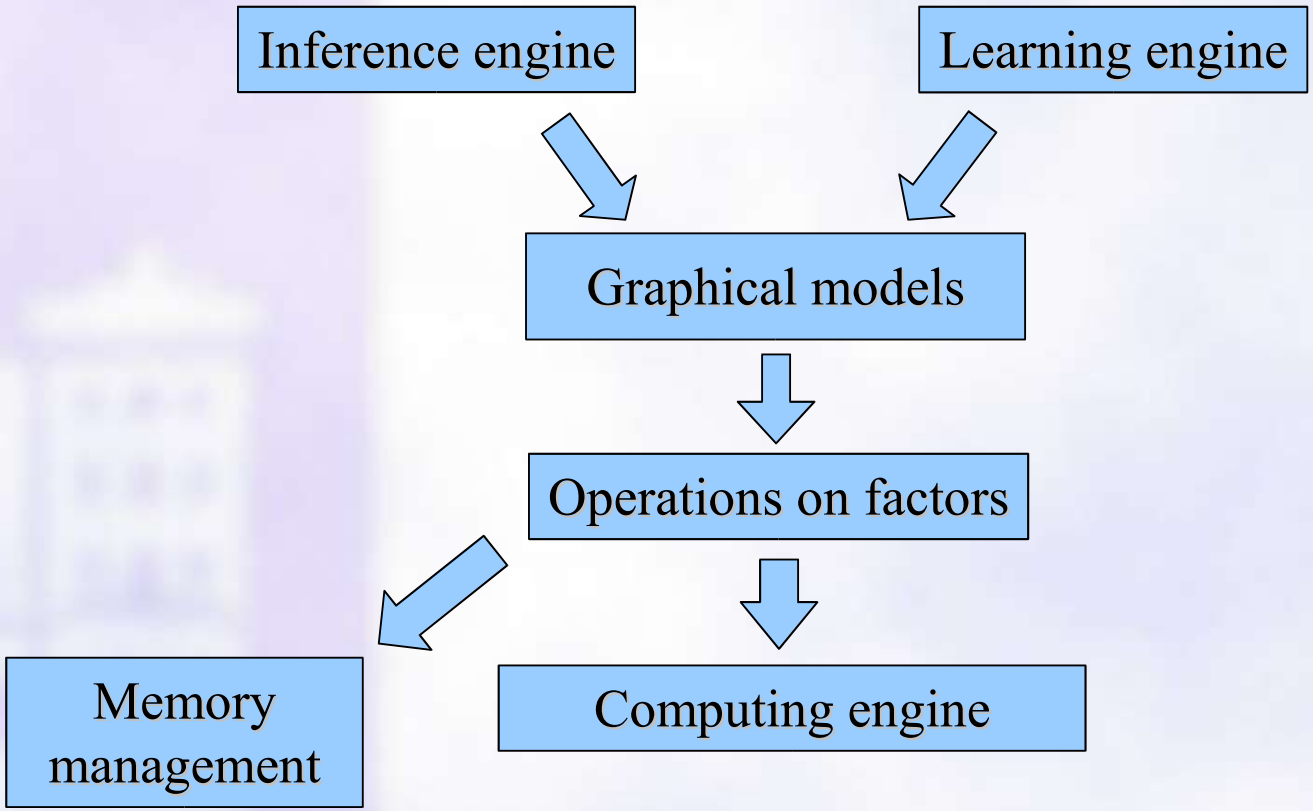
Y	Implemented
Y	Implemented (conditionally)
N	Not Implemented Yet
n/a	Not Applicable

Last Modified: 19 April, 2004

openPNL - Arquitectura



openPNL – Arquitectura computacional



P(Fin=true)=1

- José M. Gutiérrez (Univ. Cantabria):
  http://personales.unican.es/gutierjm/
- Francisco Javier Díez Vegas (UNED):
  http://www.ia.uned.es/~fjdiez/
- Asignatura de Inteligencia Artificial en ESIDE:
  http://asignaturas.deusto.es/ia/

- Hugin: http://www.hugin.com
- Elvira: http://www.ia.uned.es/~elvira/

- ESIDE-DEPIAN: eside-ids@deusto.es