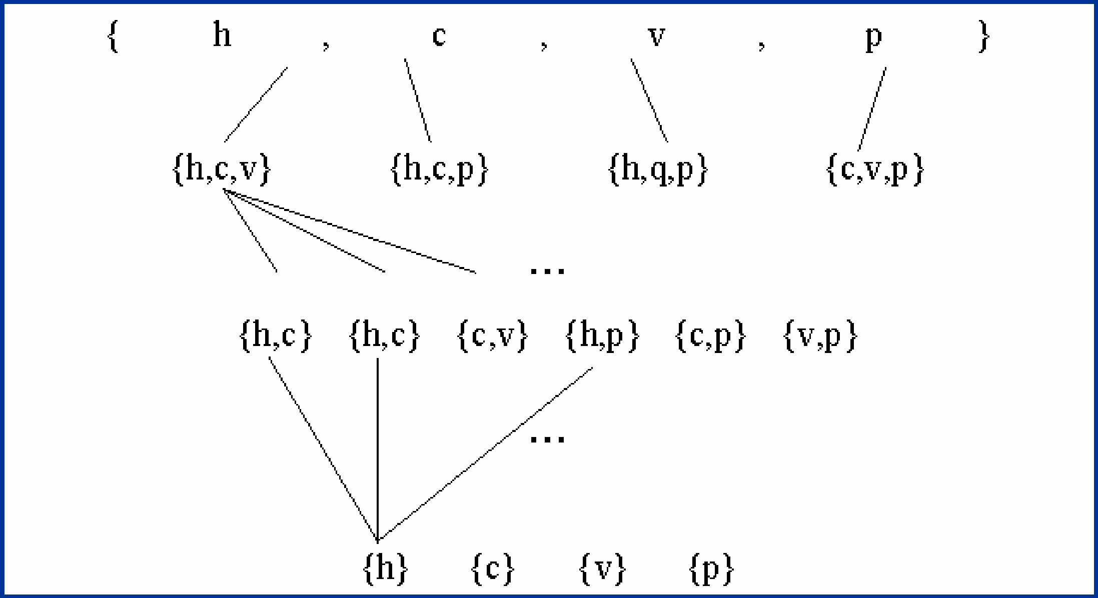
Modelos Gráficos Probabilistas

L. Enrique Sucar

INAOE

Sesión 15: Alternativas y Extensiones



Contenido

- Técnicas alternativas
  - Teoría de Dempster-Shafer
  - Ejemplo médico
- Lógica y probabilidad
  - Lógicas probabilistas
  - Modelos relacionales
  - Aplicaciones en rec. de gestos y modelado de estudiantes

Técnicas Alternativas

- Se han desarrollado algunas técnicas numéricas para manejo de incertidumbre que no siguen los axiomas de probabilidad. Entre éstas se encuentran:
  - Métodos empíricos o *ad-hoc*
  - Teoría de Dempster-Shafer
  - Lógica difusa

Técnicas Alternativas

- Algunas técnicas se pueden ver como casos especiales o extensiones de probabilidad
- Técnicas que se reducen a casos especiales de probabilidad
  - Método de factores de certeza (*MYCIN*)
  - Método de pseudo-probabilidades subjetivas (*Prospector*)
- Técnicas que *extienden* a probabilidad:
  - Teoría de Dempster-Shafer
- Técnicas basada en diferentes fundamentos:
  - Lógica Difusa

Teoría de Dempster-Shafer

Antecedentes

Teoría para representar y combinar “grados de creencia”.

Esta teoría se desarrollo básicamente como una alternativa (extensión) a teoría de probabilidad ya que los autores consideraban que ciertas situaciones no eran representadas adecuadamente con dicha teoría. En especial dos aspectos:

- Representación de "ignorancia"
- Representación de creencia NO asignada

Ejemplo

- Se tiene una moneda y dos situaciones distintas:
  1. La moneda es “normal” por lo que tiene la misma probabilidad de cada lado
  2. Se sabe que la moneda esta cargada con una mayor probabilidad de uno de los lados, pero no se sabe cual ni cuanto
- Con probabilidades ambas situaciones se representan igual – $P=0.5$, no hay forma de distinguir *ignorancia* de *igual probabilidad*

Diferencias con Probabilidad

La teoría de DS difiere en dos aspectos básicos de la teoría clásica de probabilidad:

- Los grados de creencia se asignan a subconjuntos en lugar de a elementos individuales del dominio de referencia.
- El axioma de *aditividad* no se fuerza, sino se substituye por una desigualdad.

Diferencias con Probabilidad

Estas diferencias tiene dos importantes implicaciones:

1.- La creencia en una proposición y su complemento NO necesariamente suman “1”.

2.- Se diferencia ignorancia de probabilidades iguales, dando la creencia no asignada al conjunto de todas las hipótesis.

Fundamentos Teóricos

La teoría de DS requiere de un conjunto de hipótesis exclusivas y exhaustivas:

Θ - marco de dicernimiento

2^{Θ} - conjunto de todos los subconjuntos de Θ

En base a esto se definen dos medidas:

- asignación básica de probabilidad (bpa)
- función de creencia (Bel)

Asignación básica de probabilidad (bpa)

- Representa la porción de creencia asignada exactamente a un elemento *A* (subconjunto de Θ), sin incluir la creencia asignada a sus subconjuntos.

$$\text{bpa}=m(A): 2^{\Theta} \rightarrow [0,1]$$

- Debe satisfacer las siguientes propiedades:

$$1 >= m(A) >= 0 \qquad (1)$$

$$m(\emptyset) = 0 \qquad (2)$$

$$\Sigma m(A)=1 \qquad (3)$$

Ejemplo

- Para el ejemplo de la moneda

  $\Theta = \{\text{águila, sol}\}$

  $2^{\Theta} = [\ \{\text{águila, sol}\}, \{\text{águila}\}, \{\text{sol}\}, \varnothing\ ]$
- Caso 1: igual probabilidad

  $m(\{\text{águila}\}) = 0.5,\ m(\{\text{sol}\}) = 0.5$
- Caso 2: ignorancia completa

  $m(\{\text{águila, sol}\}) = 1$

Función de creencia (Bel)

- Es la creencia total en el conjunto A, incluyendo la creencia asignada propiamente a A, así como la de todos sus subconjuntos:

$$Bel(A)=\sum m(B), \quad B \subseteq A$$

- Se puede demostrar que Bel satisface las siguientes propiedades:

$$Bel(\emptyset) = 0$$

$$Bel(\Theta) = 1$$

$$Bel(A1 \cup A2) >= Bel(A1) + Bel(A2) - Bel(A1 \cap A2)$$

Función de creencia (Bel)

- Para una hipótesis sencilla (un solo elemento) se tiene que:

  $$Bel(A)=m(A)$$

- Para una hipótesis en general se tiene que:

  $$Bel(A)>=m(A)$$

Función de creencia (Bel)

- Para el ejemplo de la moneda:
  - Caso 1:
    - Bel({águila, sol}) = 0.5 + 0.5 + 0 = 1
    - Bel({águila}) = m({águila}) = 0.5
    - Bel({sol}) = m({sol}) = 0.5
  - Caso 2:
    - Bel({águila, sol}) = 0 + 0 + 1 = 1
    - Bel({águila}) = m({águila}) = 0
    - Bel({sol}) = m({sol}) = 0

Regla de Dempster

- Para combinar distintas evidencias se calcula su suma ortogonal, aplicando lo que se conoce como la regla de Dempster, y obteniendo un nuevo grado de creencia **(*m*)** basado en la evidencia combinada:

$$[m_1 \otimes m_2](A) = \sum_{Ai \cap Bj = A} m_1(A_i) m_2(B_j)$$

- Esta fórmula la podemos interpretar de la siguiente forma:
  - La evidencia E1 asigna la creencia ml al subconjunto Al
  - La evidencia E2 asigna la creencia m2 al subconjunto B1
  - Entonces el producto de ambas (ml * m2) nos da la creencia en la intersección – A1 ∩ B1

Ejemplo

- Si hubiera dos evidencias (*expertos lanza monedas*) respecto a la moneda cargada:
  - $m1(A) = 0.7$, $m1(\Theta) = 0.3$
  - $m2(S) = 0.6$, $m2\ (\Theta) = 04$
- Entonces:

m2 \ m1	{A} 0.7	{Θ} 0.3
{S} 0.6	{∅} 0.42	{S} 0.18
{Θ} 0.4	{A} 0.28	{Θ} 0.12

Regla de Dempster

- La creencia total en A es simplemente la suma de las creencia asignadas de esta forma, es decir, la suma de la creencia de todas la intersecciones entre los conjunto A_i y B_j que den como resultado A.

- Surge un problema si alguna de las intersecciones de el conjunto vacío, ya que no se puede asignar creencia a dicho conjunto (implicaría que la suma de bpa no sea l). Para resolver este caso hay que normalizar los bpa, es decir, inflar las creencias de los demás subconjuntos en forma proporcional a la creencia asignada al conjunto vacío.

Regla de Dempster

- Entonces la regla de Dempster en su forma general es:

$$[m_1 \otimes m_2](A) = \sum_{Ai \cap Bj = A} \frac{m_1(A_i) m_2(B_j)}{1-k}, A \neq \phi$$

$$donde:$$

$$K = \sum_{Ai \cap Bj = \phi} m_1(A_i) m_2(B_j)$$

- Los nuevos valores de Bel para cada hipótesis son calculados de la misma forma, sumando los bpa's.

Ejemplo

- Normalizando:
  - $k = 0.42 \rightarrow 1\text{-}k = 0.58$
- Entonces:
  - $m1 \otimes m2(\{S\}) = 0.18 / 0.58 = 0.310$
  - $m1 \otimes m2(\{A\})\ 0.28 / 0.58 = 0.483$
  - $m1 \otimes m2(\{\Theta\})\ 0.12 / 0.58 = 0.207$

Posibilidad

- Mientras que Bel nos da la cantidad de creencia en cierta hipótesis, otra medida denominada la posibilidad (*plausibility* – Pl) indica la máxima creencia que pudiera asignarse a la hipótesis. La posibilidad se define como:

  P1(A) = 1-Bel(~A)

- *Bel* da la creencia mínima y *P1* la creencia máxima. Ambas definen un *intervalo de creencia:*

  [Bel(A), P1(A)]

- El rango dentro del cual estaría la creencia en A de acuerdo a la evidencia conocida. La diferencia entre Bel y Pl nos indica la ignorancia, es decir, la creencia que NO ha sido asignada ni a la hipótesis ni a su complemento (o demás hipótesis).

Ejemplo

- Para el caso anterior:
  - $Pl(\{A\}) = 1 - 0.310 = 0.690$
  - $Pl(\{S\}) = 1 - 0.483 = 0.517$

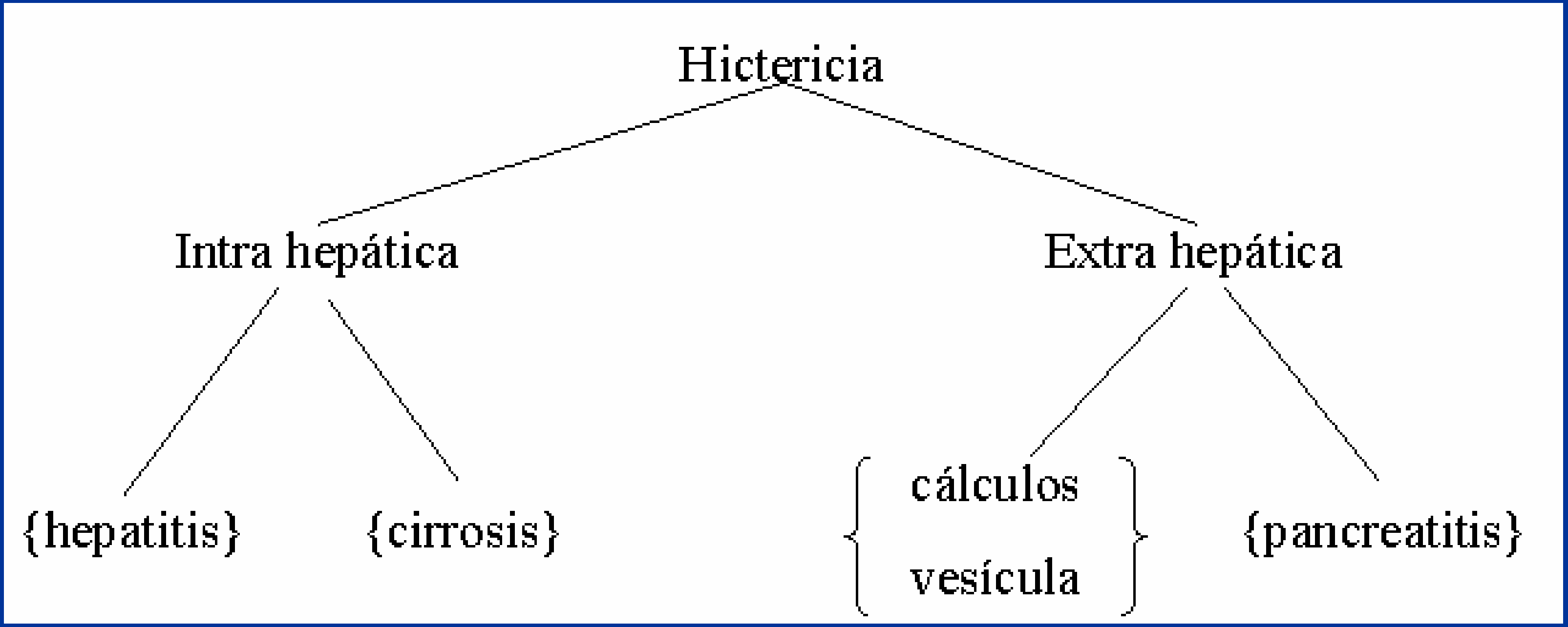
- Entonces:
  - A: [0.483   0.690]
  - S: [0.310   0.517]

Otro Ejemplo

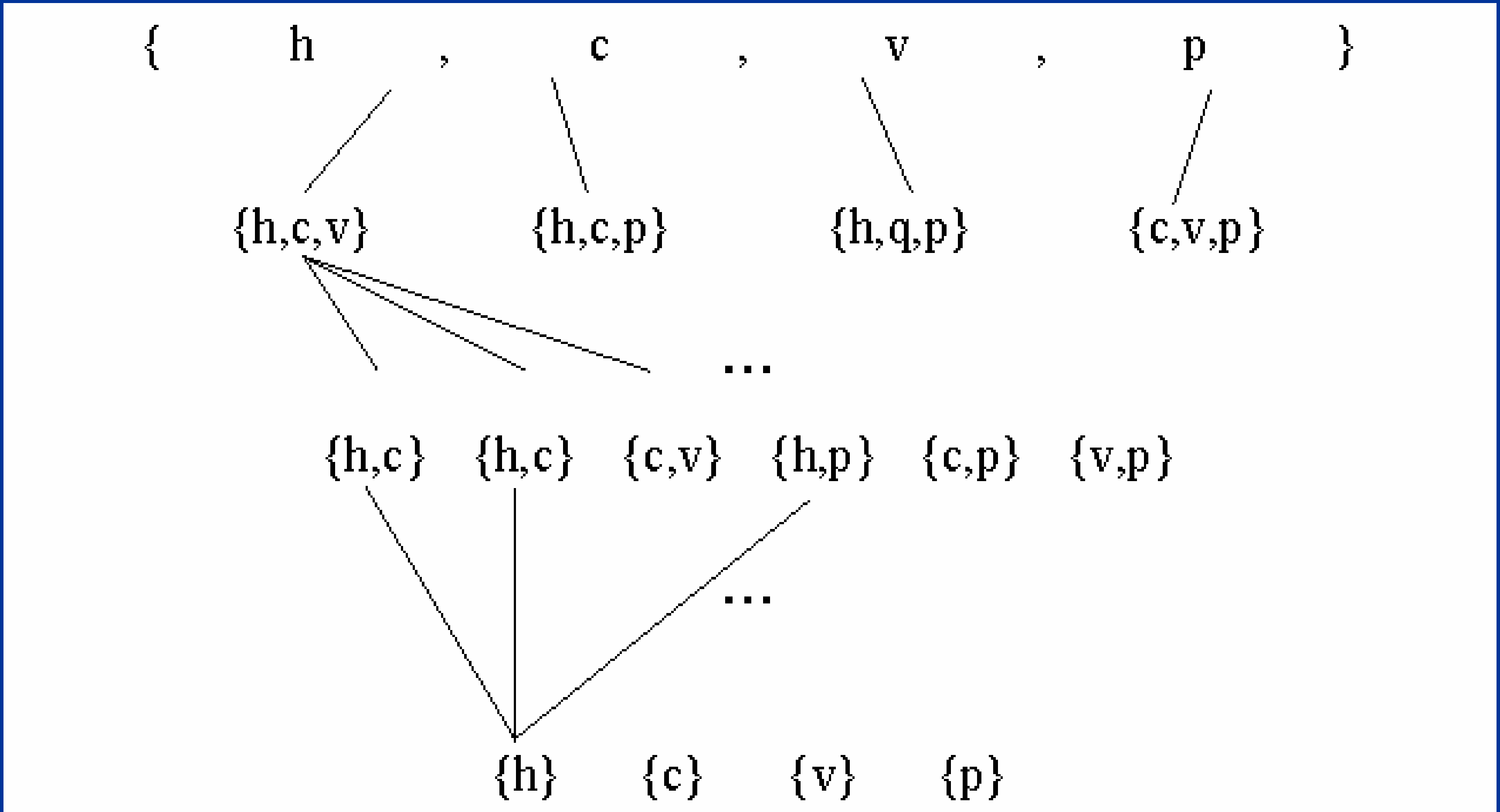
- Consideremos una aplicación médica en la que hay cuatro posibles enfermedades (hipótesis):
  - Hepatitis (h/hep)
  - Cirrosis (c/cirr)
  - Cálculos en la vesícula (v/gall)
  - Pancreatitis (p/pan)

Ejemplo Médico

- Marco de discernimiento (hipótesis) - *jerarquía*:



Ejemplo Médico - subconjuntos



Ejemplo Médico

- Evidencia 1:
  intra-hepática – 0.6
- Evidencia 2:
  no hepatitis – 0.7

		m_2	
		{cirr, gall, pan}(0.7)	θ(0.3)
m_1	{hep,cirr} (0.6)	{cirr} (0.42)	{hep, cirr} (0.18)
	θ(0.4)	{cirr, gall, pan} (0.28)	θ (0.12)

Ejemplo Médico

- A partir de las bpa se puede calcular el grado de creencia – *Bel*, por ejemplo:

Bel(intrahepática) = Bel({hep,cerr}) =

m(hep,cerr) + m(hep) + m(cerr) =

0.18 + 0 + 0.42 = 0.60

Ejemplo Médico

- Evidencia 3:
  hepatitis – 0.8

$$m_4 = m_1 \oplus m_2$$

	{cirr} (0.42)	{hep, cirr} (0.18)	{cirr,gall, pan} (0.42)	θ(0.12)
{hep} (0.6)	θ (0.42)	{hep} (0.144)	θ (0.12)	{hep} (0.096)
θ(0.2)	{cirr} (0.084)	{hep, cirr} (0.28)	{cirr, gall, pan} (0.036)	θ (0.024)

Ejemplo Médico

- Cálculo de Bel:

  $k = 0.336+0.224 = 0.56, \quad 1\text{-}k = 0.44$

  $Bel(hep) = (0.144+0.096)/0.44 = 0.545$

  $Bel(cerr) = 0.084/0.44 = 0.191$

  $Bel(hep,cerr) = 0.036/0.44 = 0.082$

  $Bel(cirr,gall,pan) = 0.056/0.44 = 0.127$

  $Bel(\Theta) = 0.024/0.44 = 0.055$

Aplicaciones

- La teoría de DS se ha llevado a diversas aplicaciones:
  - Medicina
  - Robótica
  - Visión
- Aunque los resultados son alentadores, en general es más compleja que el uso de probabilidad y la diferencia no es significativa

Ventajas

- Intervalo de creencia
- Representación de ignorancia
- Representa "la forma en que los expertos usan la evidencia"
- Modular

Desventajas

- Asume fuentes de evidencia independientes
- Interpretación de los valores finales (Bel)
- Bel no se puede interpretar como frecuencias
- Complejidad computacional (hipótesis sencillas, redes)

Referencias

- Lucas & Van Der Gaag, *Principles of Expert Systems*, Addison-Wesley, 1991 – Cap. 5
- Buchanan & Shortliffe, *Ruled-Based Expert Systems*, Addison-Weslev, 1984 - Cap 10-13.
- *Shafer, A Mathematical Theory of Evidence*, Princeton Univ. Press. 1976.

Lógica y Probabilidad

- Lógica probabilista
- Redes bayesianas con nodos lógico
- Modelos relacionales probabilistas
- Otras propuestas

Lógica y Probabilidad

- Lógica es la forma más utilizada para la representación de conocimiento
- Tiene una semántica y sintaxis bien definida, y una alta capacidad expresiva
- Pero tiene problemas para manejo de incertidumbre ...
- Una alternativa es la combinación de lógica y probabilidad

Lógica y probabilidad

- Existen varias propuestas para "integrar" lógica y probabilidad, partiendo de la lógica probabilista de Nilsson, e incluyendo entre otras:
  - Modelos relacionales probabilistas [Koller]
  - Programas lógicos – probabilistas [Haddawy]
  - Lógica de alternativas independientes [Poole]
  - Redes bayesianas con nodos lógicos [Morales y Sucar]
- Esta es todavía un área activa de investigación y aún no hay una solución definitiva

Lógica Probabilística

- Se basa en lógica de predicados y probabilidad [Nilsson 86]
- Si se tiene incertidumbre sobre el valor de verdad de una oración lógica, *S*, se considera que hay dos mundos posibles:
  - *W1*, donde *S* es verdadera, con una probabilidad p1
  - *W2*, donde *S* es falsa, con una probabilidad p2=1-p1

Lógica Probabilística

- Si se tiene L oraciones hay 2^L mundos posibles, aunque muchos son lógicamente inconsistentes ($K < 2^L$ mundos)
- La probabilidad de una oración es la suma de probabilidades de los mundos en que es verdadera:

$$\pi = P(S) = \sum_i p_i, \ \forall S = \text{verdadera en } Wi$$

Ejemplo

- Dadas las 3 oraciones:

  (1) P, (2) Q, (3) $P \wedge Q$
- Hay 4 mundos posibles:

  w1 – p=V, q=V

  w2 – p=V, q=F

  w3 – p=F, q=V

  w4 – p=F, q=F

Ejemplo

- Entonces la probabilidad de las oraciones es:

$$(1)\ \pi 1 = 1(p1)+1(p2)+0(p3)+0(p4)$$

$$(2)\ \pi 2 = 1(p1)+0(p2)+1(p3)+0(p4)$$

$$(3)\ \pi 3 = 1(p1)+0(p2)+0(p3)+0(p4)$$

- En forma matricial (V es la matriz de valores de verdad):

$$\Pi = V\ P$$

Lógica Probabilístcia

- Las ecuaciones anteriores representan un mapeo lineal de las probabilidades de los mundos posibles a las de las oraciones.
- Junto con los axiomas de probabilidad:

$$\Sigma\ pi = 0, \quad 0 < pi < 1$$

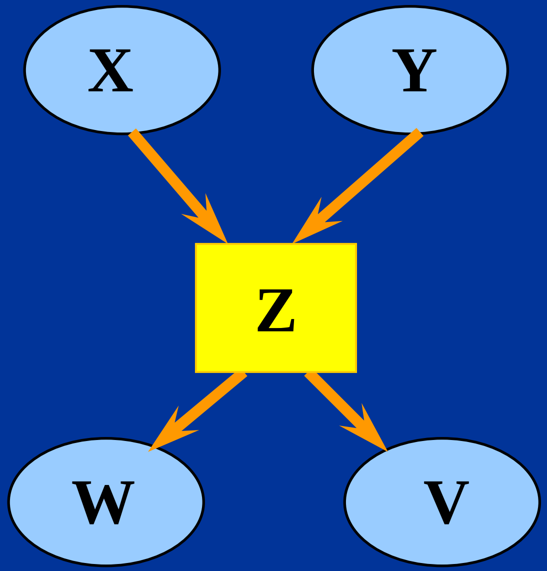
- *Restringen* los valores de probabilidad de las oraciones lógicas (pero no los determinan)

Espacio de Probabilidades

- Las probabilidades de las oraciones están dentro de un espacio convexo en el hipercubo 0—1, limitadas por los hiperplanos descritos por las ecuaciones
- Esto implica que puede haber múltiples soluciones y no es claro como seleccionar alguna de ellas

Redes lógico - probabilistas

- Nodos lógicos – programas lógicos
- Nodos probabilistas – redes bayesianas



Z:
binario-
relación (X,Y)
multivaluado -
relación (X,Y,Z)

Inferencia

- La probabilidad de *Z* depende de los valores de *X* y *Y*, y si *R* es satisfecha:

$$P(Z) = \sum \sum R(x,y)\, P(x)\, P(y)$$

- Razonamiento
  - **fuera de línea**: obtener la CPT para todos los valores de Y y Y (discretas) – determinar para el nodo lógico ***P(Z | X, Y)***
  - **en línea**: evaluar durante propagación
    - discreta: calcular la suma para todas las variables desconocidas
    - continua: técnicas de muestreo

Fuera de línea

- Ejemplo:

  $Z = Rel(X,Y) = X > Y$

  X: 1, 3 Y: 0, 2

Z=true	X=1	X=3
Y=0	1	1
Y=2	0	1

En línea

- If X= 3, Y=?

$$P(Z{=}true) = 0.3$$

- If X=1, Y=0

$$P(Z{=}true) = 1.0$$

- If X & Y unknown

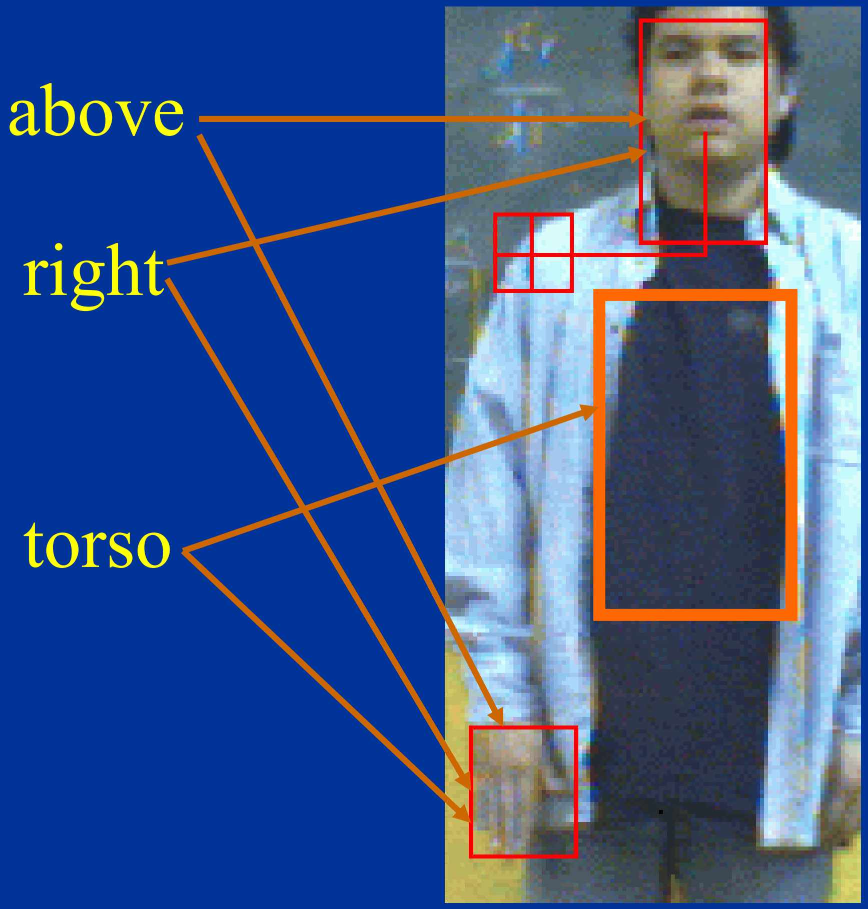
$$P(Z{=}true) = 0.58$$

Dados: P(X)=[0.7, 0.3]; P(Y)=[0.4, 0.6]

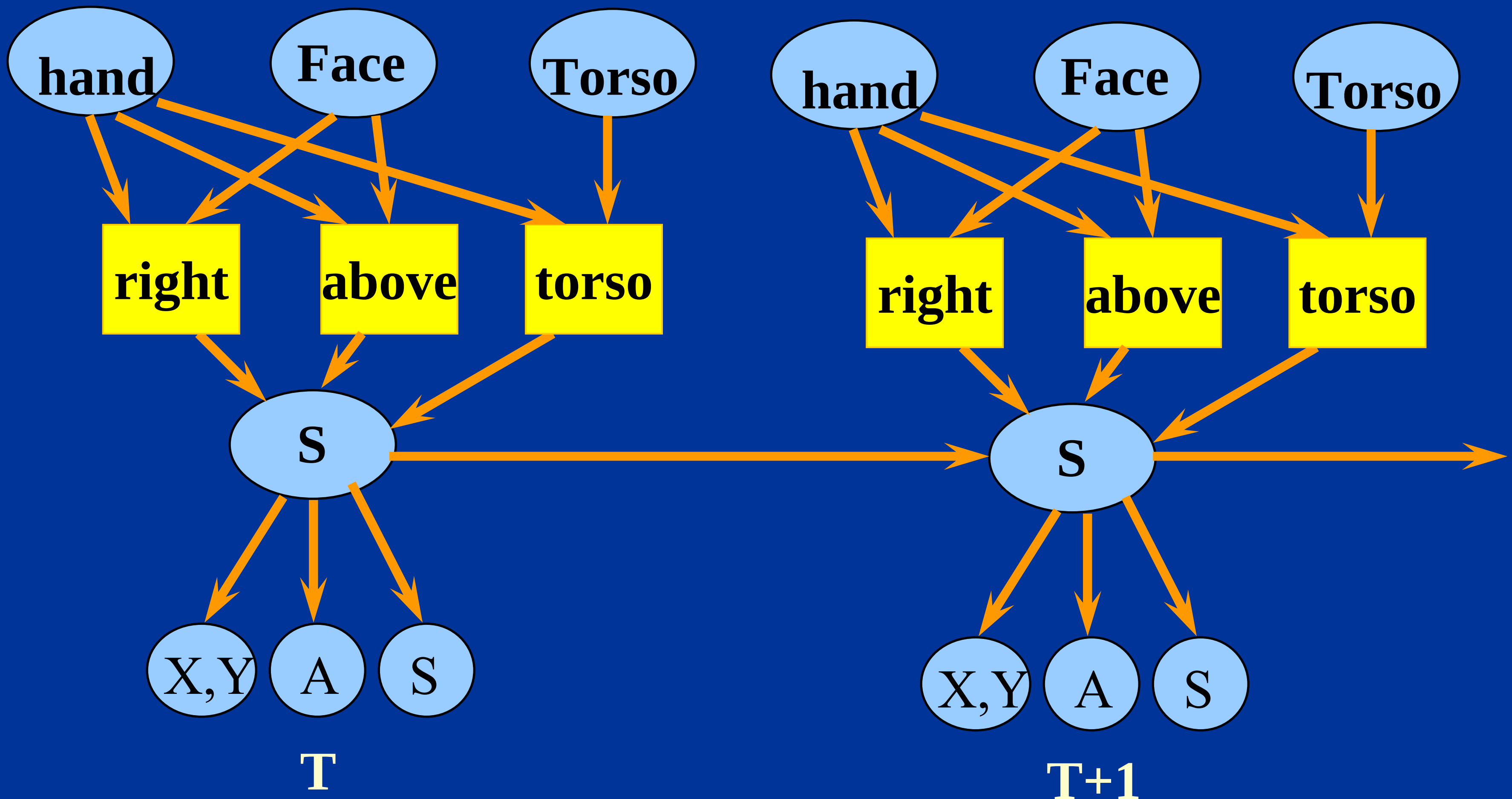
Aplicación: reconocimiento de gestos

- Basado en relaciones entre diferentes partes del cuerpo (mano, cara, torso)
- Estas relaciones están expresada como nodos lógicos en redes dinámicas lógico-probabilistas
- El modelo se usa para la obtener la probabilidad de cada gesto mediante propagación de probabilidades

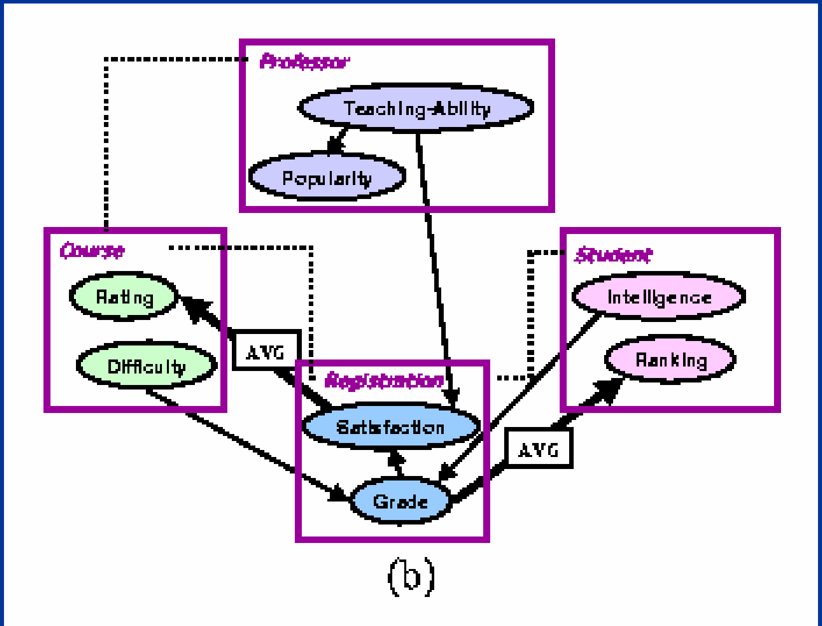
Relaciones espaciales



Modelo


hand
Face
Torso
right
above
torso
S
X,Y
A
S
T
hand
Face
Torso
right
above
torso
S
X,Y
A
S
T+1

Modelos relacionales probabilistas [Koller, 1999]



Clases:

X_1= Professor

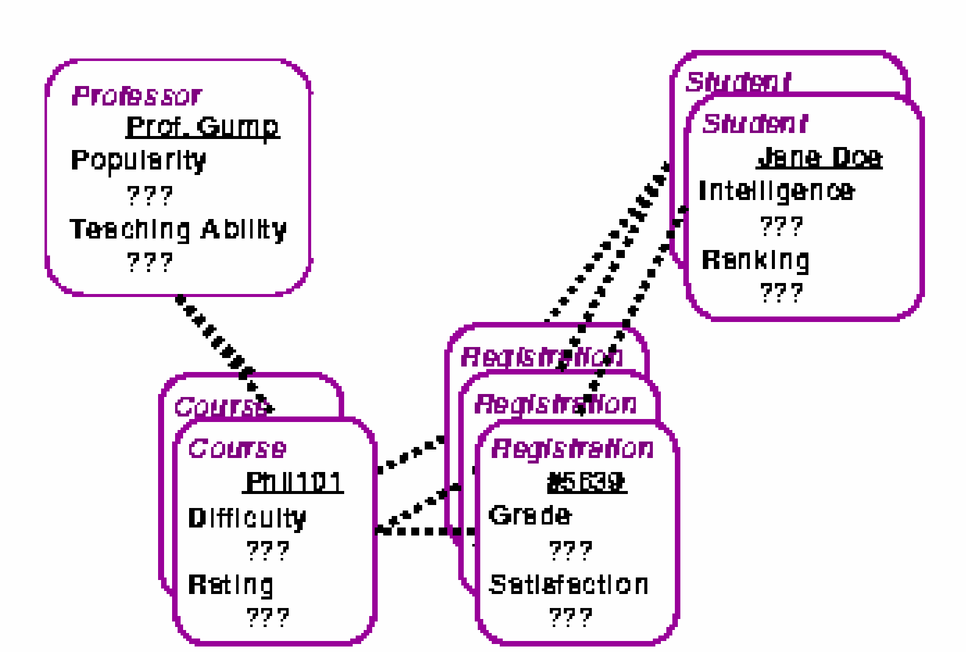
X_2= Course

X_3= Registration

X_4= Student

- Las entidades básicas son objetos del dominio.
- Los objetos del dominio son particionados en clase: X1,...,Xn.

Modelos relacionales probabilistas [Koller, 1999]



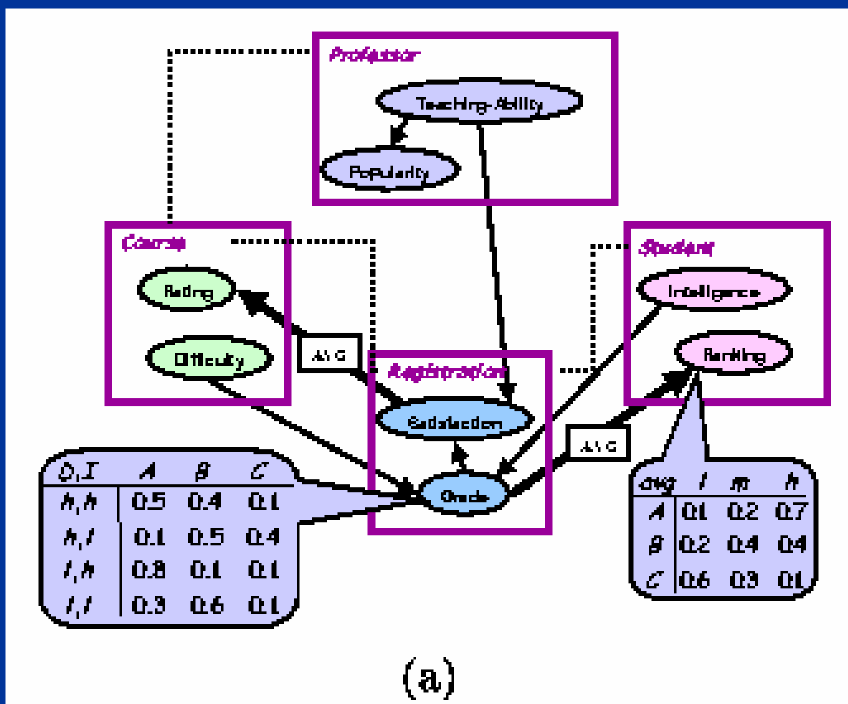
(a)

$\mathcal{A}(X_4) = \{\text{Intelligence}, \text{Ranking}\}$

X_2.Difficulty = *difficulty* attribute of *Course* class

- Cada clase es asociada con un conjunto de atributos: A(Xi).

Modelos relacionales probabilistas [Koller, 1999]



- Se usa explícitamente la estructura relacional del modelo.
- El modelo de dependencia se especifica al nivel de clases.
- Un atributo de un objeto depende de los atributos de las clases relacionadas

- Los arcos representan dependencias probabilistas:
  - Padres de la misma clase
  - Padres de diferentes clases

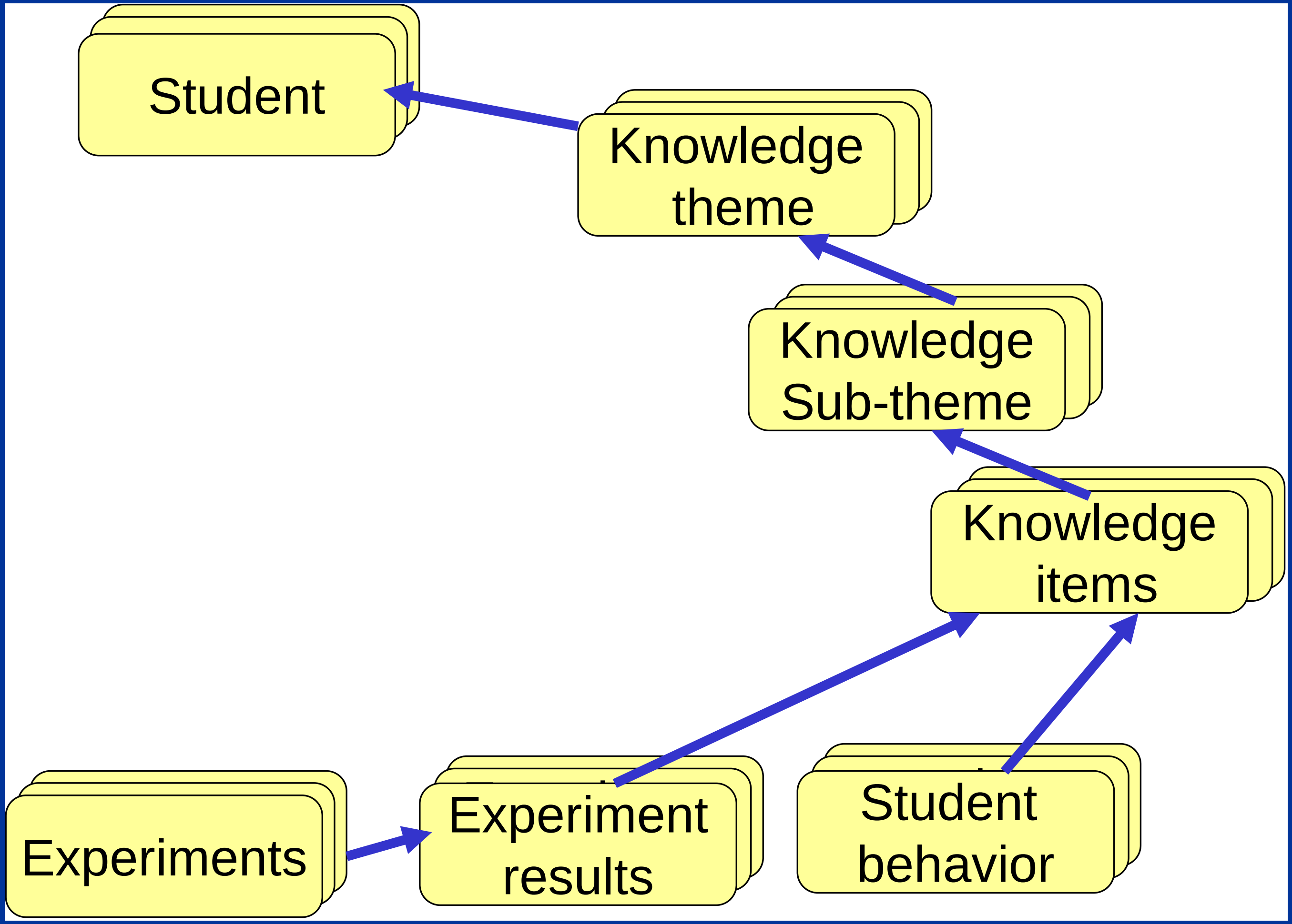
Modelos relacionales probabilistas [Koller, 1999]

- **Se basan en los mismo principios de las redes bayesianas.**
- **Permiten representar a diferentes objetos dentro del mismo modelo.**
- **Combinan las ventajas de las bases de datos relacionales con la representación de incertidumbre mediante modelos gráficos.**

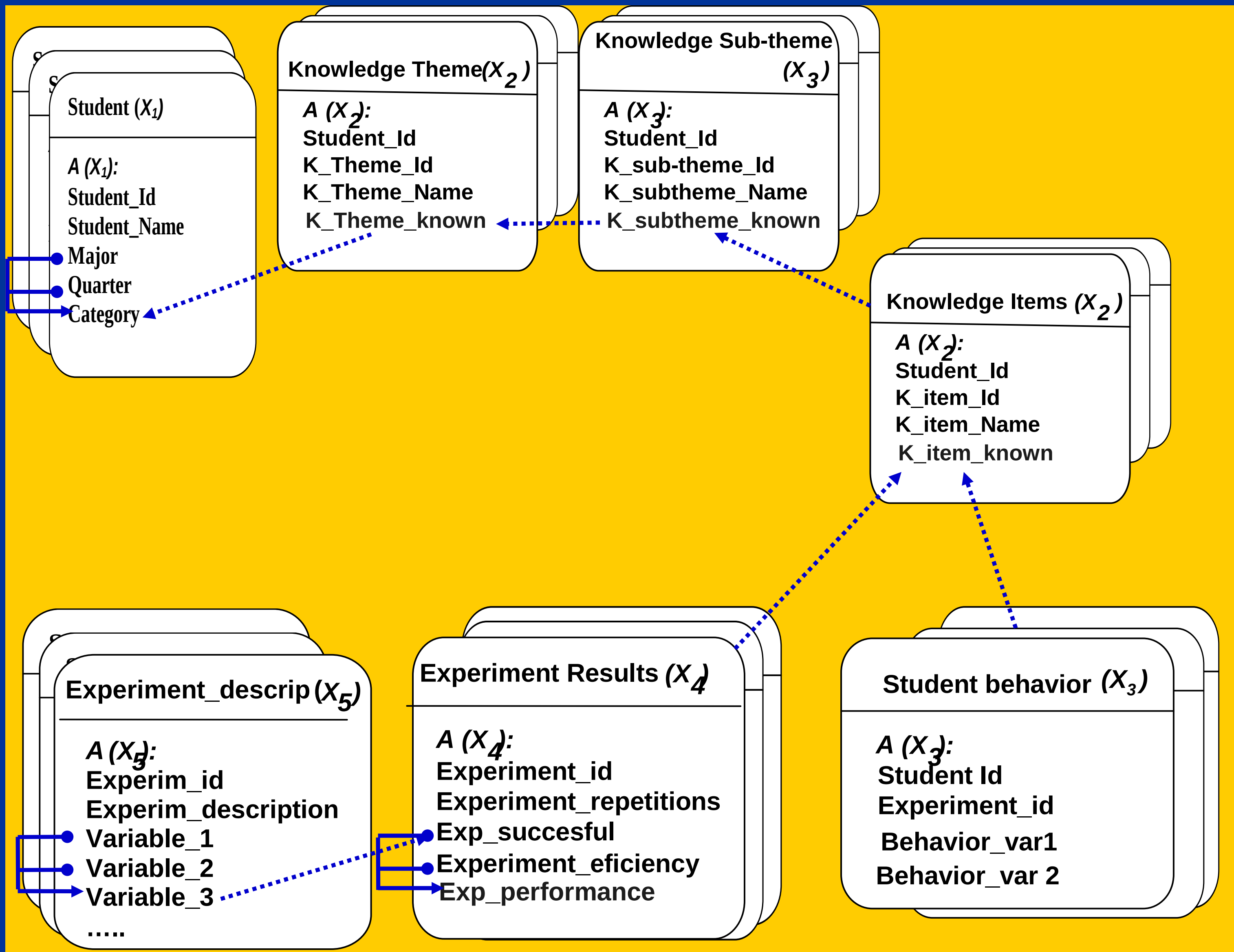
Aplicación: modelo del estudiante para un tutor inteligente

- Se representa en forma genérica el modelo del estudiante para un tutor orientado a laboratorios virtuales
- Dicho modelo se puede adaptar a diferentes experimentos y dominios
- Un vez seleccionado el dominio/experimento, el modelo se instancia a una red bayesiana sobre la cual se hace la inferencia

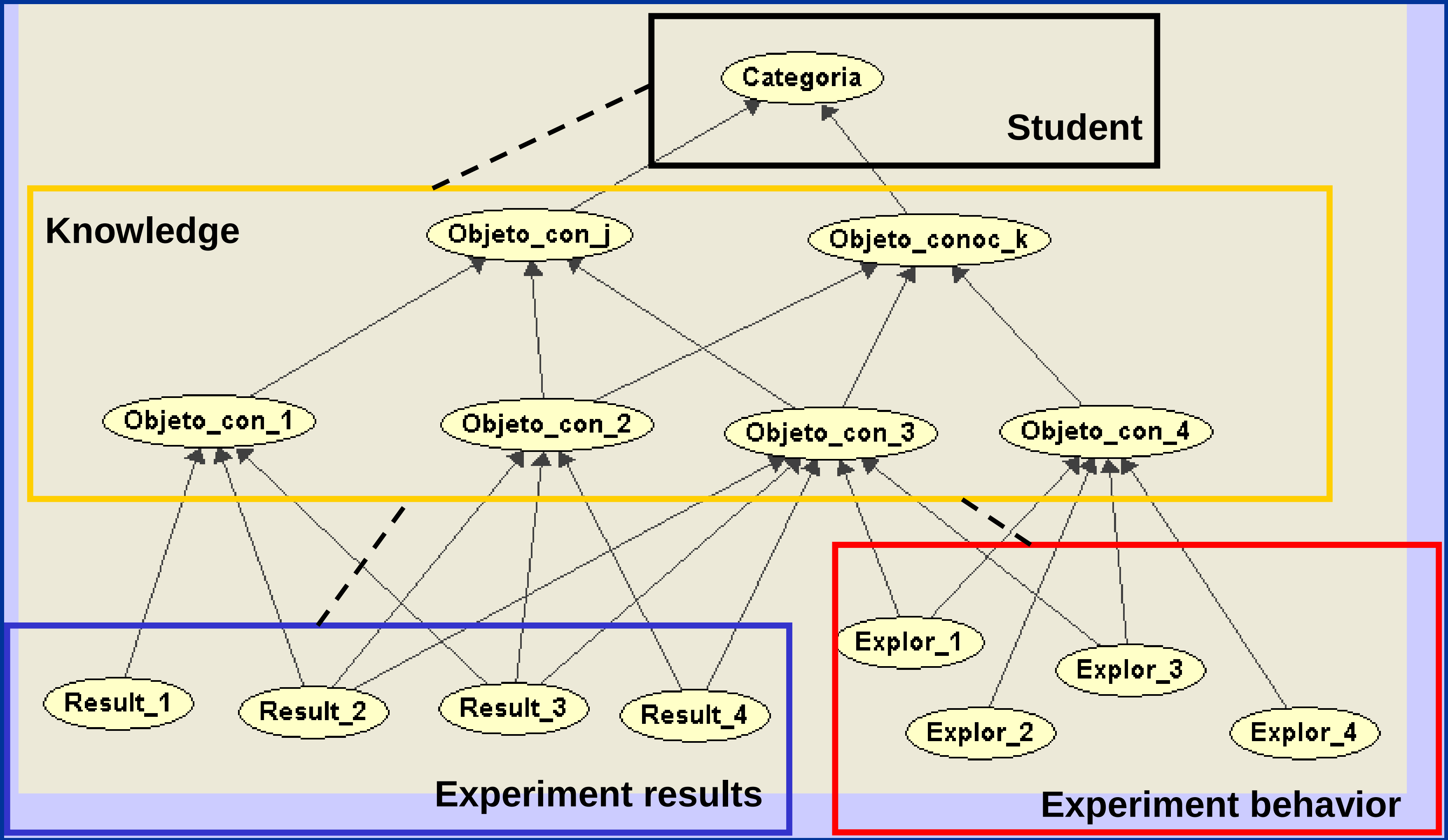
Modelo relacional del estudiante



Modelo relacional del estudiante

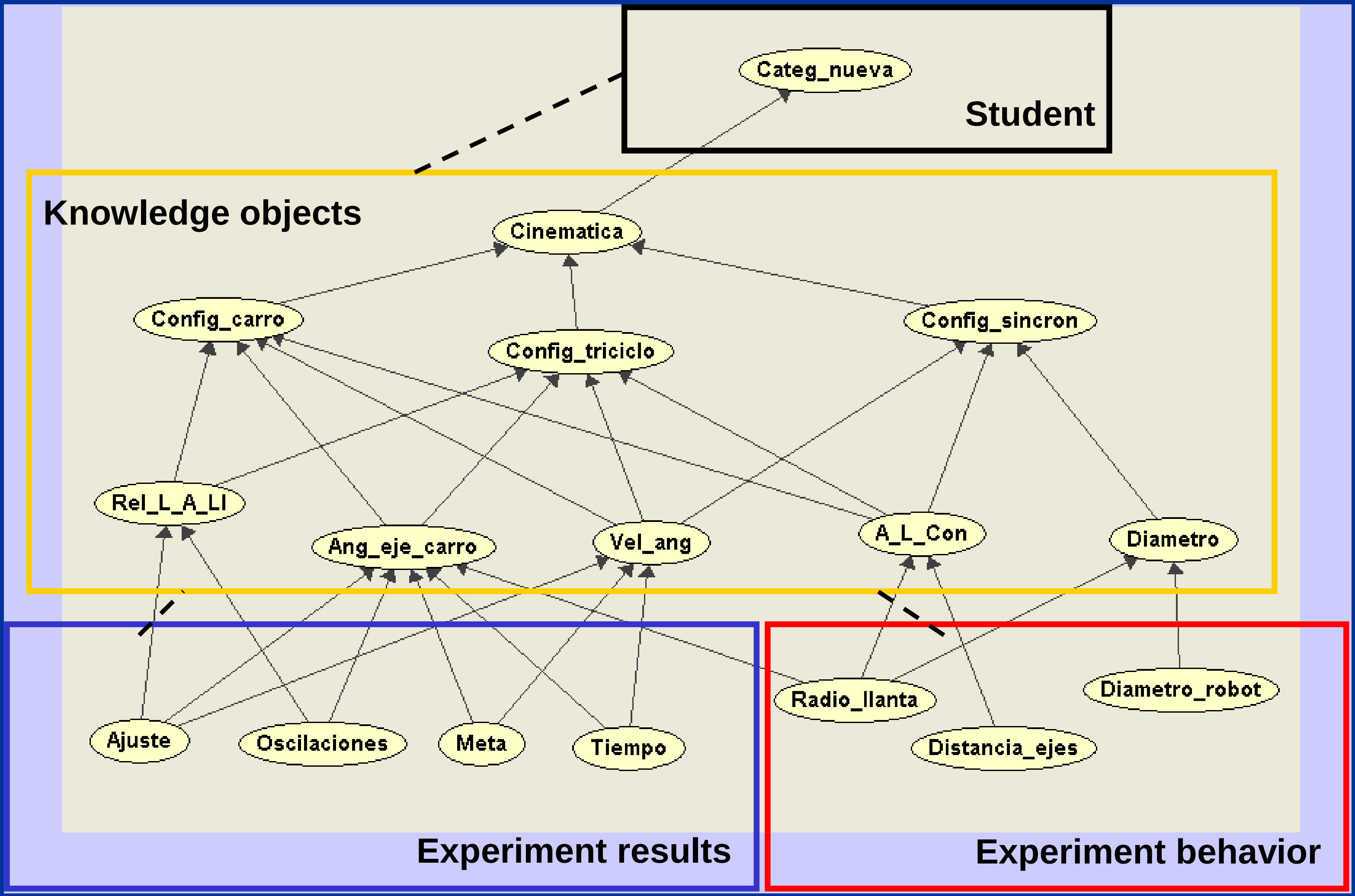


Modelo relacional del estudiante



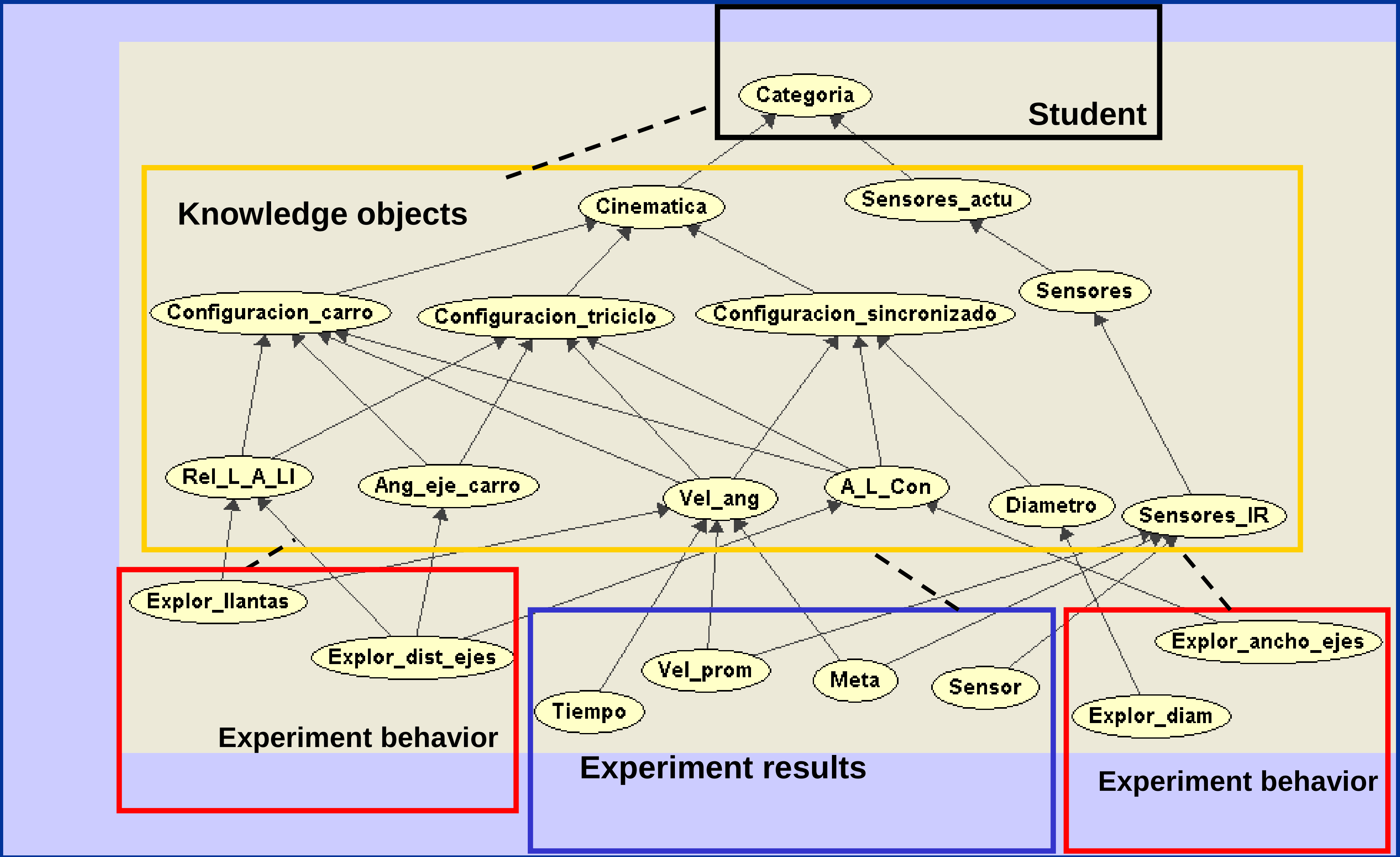
- Esqueleto extraído del modelo relacional

Modelo relacional del estudiante



- Modelo (RB) experimento particular (1)

Modelo relacional del estudiante



- Modelo para otros experimentos

Referencias

- Nilsson, Probabilistic Logic
- Koller, Halpern, AAAI 1996
- Avilés, Sucar, Mendoza, Vargas, RUR Workshop, IJCAI 2003
- Noguez, Sucar, IJEE 2006
- Poole, Artificial Intelligence 1997