Centro de Informática

UFPE

Aprendizagem de Dados Simbólicos e/ou Numéricos

Francisco de A.T. de Carvalho

Objetos Simbólicos 1/16

- Propósito da Análise de Dados Simbólicos:
  - analisar, visualizar, classificar, resumir as informações contidas na matriz de dados simbólicos
- Problema básico:
  - encontrar os $u \in E$ satisfazendo requisitos expressos pelas variáveis simbólicas
    - $Y_2 = 5.0$; $Y_3 \leq 40$; $Y_4 \in [-2.5, 3.5]$
    - $Y_5 \subseteq [-2.5, 3.5]$; $Y_6 \subseteq \{verde, branco\}$

Objetos Simbólicos 2/16

☛ Exemplo 1: Matriz de dados binários

Cidade(u)	Aço(Y_1)	Banco(Y_2)	Textil(Y_3)
A	1	1	0
B	0	1	0
C	0	1	1

- N = 3 cidades, p = 3 setores comercial ou industrial $O_j = \{0,1\}$, $Y = (Y_1, Y_2, Y_3)$, $x_u = Y(u) \in \{0,1\}^3$

Objetos Simbólicos 3/16

- Encontrar todas as cidades $u \in E$
  - que não tem industria de aço ($Y_1 = 0$) e
  - que tem bancos ($Y_2 = 1$)
    - query (consulta)
      $q = [Y_1 = 0] \wedge [Y_2 = 1]$
      (asserção no contexto SODAS)
    - query (consulta)
      $D = D_1 \times D_2 \times D_3 = \{0\} \times \{1\} \times O_3 \in \{0,1\}^3$
      $q = [Y \in D]$

Objetos Simbólicos 4/16

- query q induz em E
  - uma função q(•) tal que

$$
\begin{aligned}
q(u) \quad &:= \quad [Y(u) \in D] = [Y_1(u) = 0] \wedge [Y_2(u) = 1] \\
&= \quad \begin{cases} 1, \text{ se } Y(u) \in D \\ 0, \text{ senão} \end{cases}
\end{aligned}
$$

- "todas as cidades que satisfazem q" é uma entidade formalizada por um objeto simbólico
- Os elementos que satisfazem esse conceito: $Q = \{u \in E \ / \ Y(u) \in D\} = \{u \in E \ / \ q(u) = 1\} = \{B, C\}$

Objetos Simbólicos 5/16

- Exemplo 2: Matriz de dados heterogênea

Pessoa(u)	Altura(Y_1)	Peso(Y_2)	Matéria(Y_3)	Sexo(Y_4)
Ana	1.70	65.9	mat	1
Claudia	1.65	60.1	fis	1
Daniel	1.75	77.3	mat	0
Fred	1.68	67.0	econ	0

Objetos Simbólicos 6/ 16

- N = 4 estudantes
- p = 4 variáveis:
  - Y_1 = altura (cm), $O_1 = \Re_+$
  - Y_2 = Peso (Kg), $O_1 = O_2 = \Re_+$
  - Y_3 = matéria, O_3 = {mat, fis, econ},
  - Y_4 = Sexo, O_4 = {masc, fem} = {0, 1}

Objetos Simbólicos 7/16

- query (asserção)
  - $q_1 = [Y_1 \geq 1.70] \wedge [Y_2 \leq 65]$
- Objeto simbólico "estudante alto em forma"
- q_1 compara
  - as variáveis Y_1, Y_2 com $z_1 = 1.70$, $z_2 = 65$
  - usando as relações R_1 = '$\geq$' e R_2 = '$\leq$'

Objetos Simbólicos 8/16

- Esse objeto simbólico q_1 é caracterizado
  - pelo vetor de descrição $z^{(1)} = (z_1, z_2)$ ou
  - pelo conjunto de descrição
    - $D^{(1)} = \{1.70, \propto] \times [0.0, 65.0] \times O_3 \times O_4$
    $\in O_1 \times O_2 \times O_3 \times O_4$
- A extensão de q_1 é $Q_1 = \varnothing$

Objetos Simbólicos 9/16

- Exemplo 3: Matriz de dados simbólicos

País(u)	Pppulação(Y_1)	PNB(Y_2)	Industrias(Y_3)	Votos(Y_4)
Be	80	[500,600]	{C, Co, E, I}	(0.2,0.3,0.4,0.1)
Ci	45	[200,250]	{A, C, E}	(0.4,0.4,0.1,0.1
DO	60	[350,400]	{E, En, I}	(0.3,0.4,0.2,0.1)
Fu	35	[100,140]	{C, E, En, I	(0.2,0.2,0.5,0.1)

Objetos Simbólicos 10/16

- N = 4 países
- p = 4 variáveis:
  - Y_1 = população (milhões),
    - $O_1 = \Re_+$ (variável usual)
  - Y_2 = PNB (10^9 US\$),
    - $O_2 = \Re_+$ e B_2 = {todos os intervalos em $\Re_+$ }

Objetos Simbólicos 11/16

- Y_3 = Industrias,
  - O_3 = {Agrícola (A), Química (C), Comercio (Co), Engenharia (E), Energia (En), Informática (I)} e $B_3 = P(O_3)$
- Y4 = Voto,
  - O_4 = {Conservador(C), Democrata (D), Socialista (S), Outro (E)}
  - $Y_4(u) = \pi(u) = (p_c(u), p_d(u), p_s(u), p_e(u))$
  - $B_4 = M(O_4)$: distribuições definidas em O_4

Objetos Simbólicos 12/16

- query q_1 :
  - todos os países u populosos, grande PNB, sem industrias de energia e informática

- $q_1 = [Y_1 \geq 30] \wedge [Y_2 \subseteq [200,800]] \wedge [Y_3 \subseteq \{A,C,E,Co\}]$ (asserção: objeto simbólico)

Objetos Simbólicos 13/16

- Conjunto de descrição:
  - $D^{(1)} = [30, \propto] \times [200,800] \times \{A,C,E,Co\} \times O_4 \subseteq \times_j O_j$
- A extensão de q_1:
  - $Q_1 = \{u \in E \ / \ q_1(u) = 1\} = \{u \in E \ / \ Y_1(u) = D^{(1)}\} = \{Ci\}$

Objetos Simbólicos 14/16

- query q_2 :
  - todos os países u populosos, grande PNB, com pelo menos industrias de engenharia e informática
- $q_2 = [Y_1 \geq 30] \wedge [Y_2 \subseteq [200,800]] \wedge [Y_3 \supseteq \{E, I\}]$
- Conjunto de descrição: $D^{(2)}$
- A extensão de q_2:
  - $Q_2 = \{u \in E \ / \ q_2(u) = 1\} = \{Be, Do\}$

Objetos Simbólicos 15/16

- query q_3 :
  - todos os países u populosos, grande PNB, com pelo menos industrias de engenharia e informática e perfil político mais conservador ou democrata
- $Y_4(u) = \pi(u) = (p_c(u), p_d(u), p_s(u), p_e(u))$
  - satisfaz $Y_4(u)(\{c,d\}) = p_c(u) + p_d(u) = 0.6$

Objetos Simbólicos 16/16

- $q_3 = [Y_1 \geq 30] \wedge [Y_2 \subseteq [200,800]] \wedge [Y_3 \supseteq \{E, I\}] \wedge [p_c + p_d \geq 0.6]$

- A extensão de q_3:
  - $Q_3 = \{u \in E \;/\; q_3(u) = 1\} = \{Do\}$

Relações e Descrições 1/ 11

- Relações
  - A e B são dois conjuntos e $A \times B$ é o produto Cartesiano que compreende
    - os pares ordenados (a,b) com $a \in A$ e $b \in B$
  - Uma relação R em $A \times B$ é uma propriedade
    - definida para todos os pares $(a, b) \in A \times B$
      - que pode ser verdadeira ou falsa para algum deles

Relações e Descrições 2/ 11

- Relações
  - Matematicamente, uma relação R em $A \times B$
    - é uma função binária $\Phi(a,b) := [a\ R\ b]$
      - que assume valores em $\{0,\ 1\}$.
  - Interpretação de $\Phi(a,b) = 1$
    - "R é verdadeira para (a, b)"
  - Interpretação de $\Phi(a,b) = 0$
    - "R é falsa para (a, b)"

Relações e Descrições 3/ 11

- Relações
  - Exemplo 1: A e B subconjuntos de $\Re$ e R = '$\subset$'
    - $\Phi(A,B) = 1$ ou $[A\ R\ B] = V$
      - se $A \subset B$ é verdadeiro
    - $\Phi(A,B) = 0$ ou $[A\ R\ B] = F$
      - se $A \subset B$ é falso

Relações e Descrições 4/ 11

- Exemplo 2: $u \in E$ / $Y_j(u)$ em $Y(u) = (Y_1(u), \ldots, Y_p(u))'$ satisfaz uma ou várias condições:
  - (1) $[Y_j(u) \leq 1.3]$ (2) $[Y_j(u) > z_j]$ (3) $[Y_j(u) = z_j]$
  - (4) $[\,|Y_j(u) - z_j| \leq 0.1]$ (5) $[Y_j(u) \in \{\text{verde, branco}\}]$
  - (6) $[Y_j(u) \in D_j]$
  - (7) $[Y_j(u) \subseteq [4,6]]$ (8) $[Y_j(u) \subseteq \{\text{verde, branco}\}]$
  - (9) $[Y_j(u) \subseteq D_j]$ (10) $[Y_j(u) \supset D_j]$
  - (11) $[Y_j(u) \cap D_j \neq \emptyset]$ (12) $[Y_j(u)(D_j) \geq 0.5]$

Relações e Descrições 5/ 11

- Relações
  - Em (1)-(4) temos
    - uma relação R_j sobre $O_j \times O_j$
      - do tipo $[y_j \ R_j \ z_j]$ com $y_j, z_j \in O_j$
        - com $R_j \in \{\leq, >, =, \text{“proximo de”}\}$
  - Em (6) temos
    - uma relação R_j sobre $O_j \times P(O_j)$
      - do tipo $[y_j \ R_j \ D_j]$ com $y_j \in O_j$, $D_j \in P(O_j)$
        - e com R_j = ‘$\in$’

Relações e Descrições 6/ 11

- Relações
  - Em (7)-(11) temos
    - uma relação R_j sobre $P(O_j) \times P(O_j)$
      - do tipo $[D'_j\ R_j\ D_j]$ com D'_j, $D_j \in P(O_j)$
        - com $R_j \in \{\subseteq, \supset, \text{“overlap”}\}$
  - Em (12) temos
    - uma relação R_j sobre $M(O_j) \times P(O_j)$
      - do tipo $[y_j\ R_j\ D_j]$ com $y_j \in M(O_j)$, $D_j \in P(O_j)$; $y_j = Y_j(u)$ = peso de u
        - e R_j = ‘peso ≥ 0.5’

Relações e Descrições 7/ 11

- Descrições
  - Seja O_j o domínio da variável Y_j, $j = 1, \ldots, p$
  - Vetor de descrição:
    - $z = (z_1,\ldots,z_p) \in O_1 \times\ldots\times O_p$
  - Sistema de descrição:
    - cada $(D_1, \ldots, D_p)$ com $D_j \subseteq O_j$

Relações e Descrições 8/ 11

- Descrições
  - Conjunto de descrição cartesiano:
    - quando $D = D_1 \times \ldots \times D_p$
      - é o produto Cartesiano dos conjuntos de um sistema de descrição $(D_1, \ldots, D_p)$
  - Descrição:
    - qualquer combinação de p elementos $z_j \in O_j$ e conjuntos $D_j \subseteq O_j$
  - Espaço de descrição D:
    - totalidade das descrições consideradas

Relações e Descrições 9/ 11

- Relação Produto
  - Considere uma coleção $(R_1, \ldots, R_p)$ de relações R_j
    - R_j é definida sobre o produto cartesiano $A_j \times B_j$ tal que
      - $[a_j\ R_j\ b_j]$ é uma função binária definida para todos os pares $(a_j, b_j) \in A_j \times B_j$

Relações e Descrições 10/ 11

- Relação Produto
  - Denote
    - $A = A_1 \times \ldots \times A_p$ e $B = B_1 \times \ldots \times B_p$
      - os correspondentes produtos Cartesianos com
        - $a = (a_1, \ldots, a_p) \in A$ e
        - $b = (b_1, \ldots, b_p) \in B$

Relações e Descrições 11/11

- Relação Produto
  - A Relação Produto $R = R_1 \times \ldots \times R_p$ é definida sobre $A \times B$ por

$$[a\ R\ b] \;:=\; \bigwedge_{j=1}^{p} [a_j\ R_j\ b_j] \;=\; [a_1\ R_1\ b_1] \wedge \ldots \wedge [a_p\ R_p\ b_p]$$

  - Se Φ e Φ_j são as correspondentes funções binárias,

$$\Phi(a,b) = \prod_{j=1}^{p} \Phi_j(a_j, b_j)$$

Evento

- Uma condição do tipo
  - $[Y_j\ R_j\ z_j]$ ([tamanho = 100]) ou do tipo
  - $[Y_j\ R_j\ D_j]$ ([cor $\in$ {verde, branco}])
    - é chamada um evento
- Um evento $[Y_j\ R_j\ z_j]$ (ou $[Y_j\ R_j\ D_j]$) corresponde a uma função binária $E \to \{0,1\}$ dada por

$$[Y_j(u)R_j z_j] := \begin{cases} 1, \text{se } Y_j(u)R_j z_j \text{ é verdadeira} \\ 0, \text{senão} \end{cases}$$

Objetos Asserção 1/ 18

- Considere uma seqüência arbitrária de índices
  - $j_1, \ldots, j_r \in \{1, \ldots, p\}$ com relação as p variáveis
    - $Y_1, \ldots, Y_p$ (possivelmente com repetições, ex, $j_1 = j_2$)
- Denote $z_1, \ldots, z_r$ e $D_1, \ldots, D_r$ r elementos ou conjuntos “padrões”,
  - $z_\nu \in O_{j\nu}$ e $D_\nu \subseteq O_{j\nu}$ $(\nu = 1, \ldots, r\}$

Objetos Asserção 2/ 18

- Um objeto simbólico de tipo asserção
  - é uma conjunção dos correspondentes eventos

$$q = \bigwedge_{\nu=1}^{p} [Y_{j\nu} R_{j\nu} z_{\nu}] \text{ ou } q = \bigwedge_{\nu=1}^{p} [Y_{j\nu} R_{j\nu} D_{\nu}]$$

  - Exemplo
    - q = [Tamanho ≤ 100] ∧ [Tamanho ≥ 50] ∧ [Cor ∈ {verde, branco}]

Objetos Asserção 3/ 18

- A função extensão do objeto asserção
  - é a função binária $a_q : E \to \{0,1\}$ definida por

$$
\begin{aligned}
a_q(u) &:= \bigwedge_{\nu=1}^{r} [Y_{j\nu}(u) R_{j\nu} z_\nu] \\
&= [Y_{1\nu}(u) R_{1\nu} z_\nu] \wedge \ldots \wedge [Y_{p\nu}(u) R_{p\nu} z_\nu]
\end{aligned}
\qquad u \in E
$$

(da mesma forma, trocando z_j por D_j)

Objetos Asserção 4/ 18

- O conjunto dos objetos $u \in E$ que satisfaz q

$$
\begin{aligned}
\mathrm{Ext}(q) &:= \mathrm{Ext}(q \mid E) := a_q^{-1}(1) \\
&= \{u \in E \mid [Y_{j\nu}(u) R_{j\nu} z_\nu] = 1, \nu = 1, \ldots, p\}
\end{aligned}
$$

é chamado de extensão do objeto asserção q

Objetos Asserção 5/ 18

- A expressão
  - $q = [Y_1 \in (3,7]] \wedge [Y_3 \in \{verde, branco\}] \wedge [Y_5 \in [3.4, 4.0] \cup [7.8, 9.0]]$

especifica uma asserção com

- r = 3 eventos
- os subconjuntos $D_1 = (3,7]$, $D_3 = \{verde, branco\}$, $D_5 = [3.4, 4.0] \cup [7.8, 9.0]$
- as relações R_j = '$\in$', j = 1, 3, 5

Objetos Asserção 6/ 18

- Se Y_1, Y_3, Y_5 forem variáveis simbólicas multivaloradas
  - $q = [Y_1 \subseteq (3,7]] \wedge [Y_3 \subseteq \{verde, branco\}] \wedge [Y_5 \subseteq [3.4,4.0] \cup [7.8,9.0]]$
  - com relações R_j = '$\subseteq$', j = 1, 3, 5

Objetos Asserção 7/ 18

- Casos Especiais
  - O objeto asserção q é a conjunção de
    - exatamente p eventos do tipo $[Y_j\ R_j\ z_j]$ ou $[Y_j\ R_j\ D_j]$ onde
      - cada uma das p variáveis é usada exatamente uma vez

Objetos Asserção 8/ 18

- Casos Especiais
  - Nesse caso a asserção pode ser expressa pela correspondente Relação Produto

$$q = \bigwedge_{j=1}^{p} [Y_j R_j z_j] = [YRz] \quad \text{ou} \quad q = \bigwedge_{j=1}^{p} [Y_j R_j D_j] = [YRD]$$

    - onde $z = (z_1, \ldots, z_p)$ um vetor de descrição e
    - $D = (D_1, \ldots, D_p)$ um conjunto de descrição Cartesiano

Objetos Asserção 9/ 18

- Casos Especiais
  - Um objeto simbólico individual $q_{(z)}$ é uma asserção do tipo

$$q = \bigwedge_{j=1}^{p} [Y_j = z_j] = [Y = z]$$

gerada por um vetor de descrição $z = (z_1, \ldots, z_p)$

  - Essa definição se generaliza naturalmente quando alguns dos $z_j \in O_j$ são substituídos por $D_j \subset O_j$

Objetos Asserção 10/ 18

- Casos Especiais
  - A extensão
    - $Q = Ext(q_{(z)}) = \{u \in E \mid Y_j(u) = z_j, j=1, \ldots, p\}$
      - compreende todos os $u \in E$ que realizam o mesmo vetor de dados $Y(u) = z$

Objetos Asserção 11/ 18

- Ex 1: Considere a matriz de dados simbólicos

$$X = \begin{pmatrix} (3,7] & \{verm, azul\} & \{aço, estanho\} & 37.5 & N(9,4) \\ (6,8] & \{azul\} & \{estanho\} & 38,2 & N(8,5) \\ (2,6] & \{verde, azul\} & \{zinco\} & 36.4 & N(7,1) \\ (1,3] & \{verm\} & \{aço, zinco\} & 35.2 & N(9,3) \end{pmatrix}$$

  - com N = 4 objetos e
  - p = 5 variáveis simbólicas
  - Y_1 = comprimento, Y_2 = cor, Y_3 = material, Y_4 = largura, Y_5 = temperatura

Objetos Asserção 12/ 18

- A cada linha u corresponde uma descrição
  - $x_u = Y(u) = (\xi_{u1}, \ldots, \xi_{up})'$
- A tabela de dados simbólicos pode ser interpretada como
  - uma coleção de descrições de objetos simbólicos do tipo asserção a_1, a_2, a_3, a_4

Objetos Asserção 13/ 18

- Essas asserções podem ser descritas por
  - $a_u = [Y_1\ R_1\ \xi_{u1}] \wedge \ldots \wedge [Y_p\ R_p\ \xi_{up}]$ u = 1, …,4 desde que sejam especificadas as relações R_j
- Para Y_1
  - ou $[Y_1 \in (3,7]]$ (Y_1 é mono valorada) ou
  - $[Y_1 \subseteq (3,7]]$ (Y_1 é multivalorada de tipo intervalo)

Objetos Asserção 14/ 18

- Exemplo
  - E: carros usados
  - Variáveis
    - Y_1 = idade, $O_1 = \Re_+$;
    - Y_2 = potencia, $O_2 = \Re_+$;
    - Y_3 = consumo, $O_1 = \Re_+$
  - consulta (query): carros com no máximo 12 meses, com pelo menos 80 kW e consumo de no máximo 9 l/100 km

Objetos Asserção 15/ 18

- Exemplo
  - Objeto asserção
    - $a = [Y_1 \in [0,12]] \wedge [Y_2 \in [80,\infty)] \wedge [Y_1 \in (0,9]]$
  - Conjunto de descrição cartesiano
    - $D = D_1 \times D_2 \times D_3 = [0,12] \times [80,\infty) \times (0,9] \in \Re^3{}_+$
  - Relações: $R_1 = R_2 = R_3 = $ '$\in$'

Objetos Asserção 16/ 18

- Exemplo
  - O objeto asserção a pode ser escrito

$$q \quad = \quad [YRD] = \bigwedge_{j=1}^{3} [Y_j R_j D_j]$$

Objetos Asserção 17/ 18

- Exemplo
  - Ele é especificado
  - pelas variáveis envolvidas Y_1, Y_2, Y_3
  - pela descrição d = ([0,12], [80, ∝), (0,9))
  - pelos operadores de comparação $R_1 = R_2 = R_3$ = '∈' e a relação produto R
  - pela função de extensão a(u)

Objetos Asserção 18/ 18

- Isso sugere a definição de um objeto simbólico
  - pela quádrupla (a, Y, R, d) ou
  - pela tripla (a, R, d)
- quando a especificação de Y é obvia

Objetos Simbólicos Booleanos 1/15

- Objetos simbólicos de tipo asserção
  - a = [Y R z] (ou a = [Y R D])

combinam p eventos independentes

  - [Y_j R_j z_j] (ou [Y_j R_j D_j])

em uma expressão global

- Muitos casos práticos não se enquadram exatamente nesse esquema

Objetos Simbólicos Booleanos 2/15

- Combinação disjuntiva
  - $a = [\ cor \in \{verde, branco\}] \vee [peso > 70]$
- Combinação de conjunções, disjunções e regras
  - $a = [fumante = sim] \vee [[sexo = mas] \wedge [peso \geq 90]] \vee [[sexo = fem] \wedge [se\ [grávida = não]\ então\ [peso \leq 80]\ senão\ [peso \geq 95]]$

Objetos Simbólicos Booleanos 3/15

- Um objeto simbólico boleano (BSO) é caracterizado
  - por um vetor de dados Y
  - por uma descrição d
  - por uma relação R que compara Y e d
- O resultado dessa comparação é descrito por
  - uma função binária (boleana) $a : E \rightarrow \{0,1\}$
    - com $a(u) = 1$ se $Y(u)$ satisfaz os requerimentos e $a(u) = 0$ senão

Objetos Simbólicos Booleanos 4/15

- Formalmente, um objeto simbólico boleano
  - é definido como uma tripla s = (a, R, d)
- A função a é a
  - função de extensão do objeto simbólico s
- A extensão de s em E é dada por
  - $Ext_E(s) = Ext(s|E) := \{u \in E \mid a(u) = 1\} = a^{-1}(1)$

Objetos Simbólicos Booleanos 5/15

- Os objetos asserção são um caso especial considerando-se
  - $a(u) := [Y(u)\ R\ d]\ ,\ u \in E$
- R e d fornecem
  - uma descrição em compreensão do objeto simbólico
    - (especificação das propriedades desejadas)

Objetos Simbólicos Booleanos 6/15

- A função extensão α e a extensão A = Ext(s|E) $\subseteq$ E
  - fornecem uma descrição em extensão de s
    - (indica os elementos de E que satisfazem essas propriedades)

Objetos Simbólicos Booleanos 7/15

- Objetos simbólicos de tipo asserção com seleção de variáveis
  - As vezes ao procurar elementos de uma base de dados com propriedades específicas,
    - essas não dizem respeito a todas as variáveis

Objetos Simbólicos Booleanos 8/ 15

- Objetos simbólicos de tipo asserção com seleção de variáveis
  - De fato, apenas um subconjunto de $Y_1, \ldots, Y_p$ é de interesse e corresponde ao conjunto de índices $J = \{j_1, \ldots, j_t\} \subset \{1, \ldots, p\}$ (vamos supor $j_1 < \ldots < j_t$)

Objetos Simbólicos Booleanos 9/15

- Objetos simbólicos de tipo asserção com seleção de variáveis
  - Define-se um operador de filtragem h
    - que seleciona exatamente as variáveis desejadas Y_j, $j \in J$ de $Y = (Y_1, \ldots, Y_p)'$ e
      - as correspondentes descrições d_j, $j \in J$ de $d = (d_1, \ldots, d_p)$

Objetos Simbólicos Booleanos 10/ 15

- Formalmente, h pode ser escrito como:
  - $h_J(y) = h(y) = h(y_1, \ldots, y_p) := (y_j \mid j \in J) = (y_{j1}, \ldots, y_{jt})$
- Exemplo: $h_{\{1,3,6\}}(y_1, \ldots, y_p) = (y_1, y_3, y_6)$

Objetos Simbólicos Booleanos 11/ 15

- O objeto simbólico correspondente (a,R,d) será caracterizado pela função de extensão

$$a(u) = [Y_{j_1} R_{j_1} d_{j_1}] \wedge \ldots \wedge [Y_{j_t} R_{j_t} d_{j_t}] = [Y(u) R_J d]$$

  - onde a relação $h_J(R) = R_J$
    - é definida por $[y\ R_J\ d] = \wedge_{j \in J}\ [y_j\ R_j\ d_j]$
    - e refere-se apenas aos componentes selecionados
- Esse tipo de objeto simbólico é encontrado na construção de árvores de indução, por exemplo

Objetos Simbólicos Booleanos 12/ 15

- Objetos simbólicos com filtro
  - Nesse caso, a função de extensão de s é dada por
    - a(u) := [Y(u) R h(d)], onde
      - h(•) é uma transformação adequada aplicada ao vetor de descrição d
  - Esse tipo de objeto simbólico é útil quando os valores das variáveis sofrem restrições via regras ou quando do tratamento de dados confidenciais

Objetos Simbólicos Booleanos 13/ 15

- Objetos simbólicos com descrições modais
  - Esse tipo de objeto simbólico diz respeito ao caso de variáveis modais Y_j onde
    - o "valor de $Y_j(u)$ é uma distribuição do conjunto $M(O_j)$ de todas as distribuições de probabilidade do domínio O_j

Objetos Simbólicos Booleanos 13/ 15

- Objetos simbólicos com descrições modais
  - A descrição simbólica d é
    - um vetor $d = (d_1, \ldots, d_p) = (Q_1, \ldots, Q_p)$ de
      - p distribuições especificadas $Q_j \in M(O_j)$ $(j = 1, \ldots, p)$

Objetos Simbólicos Booleanos 14/ 15

- Objetos simbólicos com descrições modais
  - As relações R_j são definidas por

$$[Y_j R_j Q_j]) = \begin{cases} 1, \text{se } d(Y_j(u), Q_j) \leq \varepsilon \\ 0, \text{se } d(Y_j(u), Q_j) > \varepsilon \end{cases}$$

    - com um limiar $\varepsilon > 0$, onde
      - $d(P,Q) \geq 0$ é uma função de dissimilaridade entre duas distribuições de probabilidade P e Q $\in$ $M(O_j)$

Objetos Simbólicos Booleanos 15/ 15

- Objetos simbólicos com descrições modais
  - Com essa interpretação o objeto simbólico pode ser escrito como

$$a(u) = [Y_1(u) R_1 Q_1] \wedge \ldots \wedge [Y_p R_p Q_p], \ u \in E$$

Objetos Simbólicos Modais 1/ 11

- Definição mais geral de um objeto simbólico
- Trata-se de trocar a Relação hard R por uma relação soft (fuzzy) Φ
- Uma relação hard como
  - y R z, D' R D, Y(u) R z

assume apenas os valores 1 ou 0 (distingue V e F)

  - $y \leq z$ de $y > z$; $D' \subseteq D$ de $D' \not\subseteq D$

Objetos Simbólicos Modais 2/ 11

- Uma relação fuzzy permite gradações
  - "verdadeiro, provável, possível, improvável, falso"
  - "sempre verdadeiro, quase sempre verdadeiro, meio a meio, quase sempre falso, falso"
- Existem situações onde o grau de pertinência de Y(u) a D deve ser medido em uma escala contínua

Objetos Simbólicos Modais 3/ 11

- Uma relação fuzzy permite gradações
- Nesses casos, trata-se de trocar
  - a função binária $\Phi(y,z) := [y \; R \; z] \in \{0,1\}$ por
    - uma função real com valores no intervalo $[0,1]$ tal que $0 \leq \Phi(y,z) \leq 1$

Objetos Simbólicos Modais 4/ 11

- Nesse caso $\Phi(y,z)$ indica
  - o grau de concordância entre dois vetores de descrição y e z
    (da mesma forma para $\Phi(y,D)$, $\Phi(D',D)$)
- Φ é chamada de relação fuzzy

Objetos Simbólicos Modais 5/ 11

- Um objeto simbólico é uma tripla s = (a, Φ, d)
  - onde d é uma descrição de um espaço de descrições *D*,
  - Φ é uma relação fuzzy entre descrições e
  - a é uma função de E no intervalo [0,1] definida por a(u) := Φ(Y(u), d), u $\in$ E

Objetos Simbólicos Modais 6/ 11

- Para cada $u \in E$,
  - $a(u)$ mede o quanto u satisfaz a descrição d.
  - $a(\bullet)$ é chamada de função de extensão de s
- Para um dado limiar $\alpha \in [0,1]$,
  - a extensão ao nível α de um objeto simbólico $s = (a, \Phi, d)$ em E é definido

$$
\begin{aligned}
Ext_\alpha(s) \quad &:= \quad Ext_\alpha(s \mid E) := \{u \in E \mid a(u) \geq \alpha\} \\
&= \quad a^{-1}([\alpha,1]) = \{u \in E \mid a_s(u) = \Phi(Y(u), z) \geq \alpha\}
\end{aligned}
$$

Objetos Simbólicos Modais 7/ 11

- O conjunto Q = $Ext_{\alpha}(s|E)$ fornece a resposta para a consulta (query) que procura todos os u $\in$ E tal que
  - o vetor Y(u) e a descrição d satisfazem a relação fuzzy Φ pelo menos ao nível α
- Um objeto simbólico s = (a, Φ, d) é chamado de objeto simbólico modal
  - se Φ é uma relação fuzzy, isto é, se a função extensão *a* toma valores a(u) entre 0 e 1

Objetos Simbólicos Modais 8/ 11

- Exemplo
  - E : todas as espécies de flores do mundo (tulipas, rosas, ...)
    $u \in E$ é um objeto de segunda ordem (grupos de flores)
  - Uma espécie pode ser caracterizada pela cor (Y_1) e pelo período de floração (Y_2)
  - Y_1 e Y_2 são multivaloradas categóricas

Objetos Simbólicos Modais 9/ 11

- Exemplo
  - B_1 de Y_1 contém todos os subconjuntos de O_1 = {verde, branco}
  - B_2 de Y_2 contém todos os subconjuntos de O_2 = {Janeiro, ..., Dezembro}
  - Espécies vermelhas ou amarelas com floração em maio ou junho
  - Conjunto de descrição
    - $D = D_1 \times D_2$ = {amarelo, vermelho} $\times$ {maio, junho}

Objetos Simbólicos Modais 10/ 11

- Exemplo
  - Ω: conjunto de todas as espécies do meu jardim e o vetor de dados simbólicos $Y(u) = (Y_1(u), Y_2(u)), u \in \Omega$
  - Pode-se definir o objeto simbólico modal $s = (a,\Phi,D)$ com função de extensão $a(u) := \Phi(Y(u), D)$

Objetos Simbólicos Modais 11/ 11

- Exemplo
  - $a(u)$ indica o nível em que u = tulipas do meu jardim e a sua descrição $D' = Y(u)$ matches a descrição $D = \{amarelo, vermelho\} \times \{maio, junho\}$
  - A relação fuzzy $\Phi(D', D)$ mede a semelhança entre D' e D com valores em [0,1]
  - Essa medida pode ser
    - $\Phi(D', D) = |D' \cap D| \; / \; |D' \cup D|$