conhecimento incerto - representação probabilística

Ciências ULisboa

incerteza

por fatores externos ao planeador

- ex: táxi - não controla trânsito e imprevistos

por espaço de procura demasiado vasto

por desconhecer como atuar em determinadas situações

por não conseguir determinar exatamente a situação

probabilidades

abordagem prática para lidar com incerteza

- atribuição de probabilidades a acontecimentos incertos

quantifica a (in)certeza com um valor (probabilístico)

- ⇒ entre 0 e 1
- **nota**: o máximo da incerteza não é quando a probabilidade é 0 ou 1! é quando todos os resultados são equiprováveis

  entropia

decisões

um agente necessita ter em conta as probabilidades dos resultados das ações

e

as utilidades de cada ação e seus possíveis resultados

teoria da decisão = *f*(*teoria de probabilidades*, *teoria da utilidade*)

agente racional: escolhe a ação que tem a maior utilidade esperada, em média sobre todos os seus resultados possíveis

probabilidade conjunta (tabela)

um exemplo com três variáveis

	dorDentes		¬*dorDentes*	
	sonda	¬*sonda*	*sonda*	¬*sonda*
cárie	0,108	0,012	0,072	0,008
¬*cárie*	0,016	0,064	0,144	0,576

sonda: deteção de um orifício no dente

breve recap. de probabilidades

v.a. - inicial maiúscula

valor da v.a. - minúscula

$$P(a \lor b) = P(a) + P(b) - P(a \land b)$$ probabilidade da disjunção

$$P(a|b) = \frac{P(a \land b)}{P(b)}$$ probabilidade condicional, função da conjunta

$$P(a,b) = P(a|b)P(b)$$ probabilidade conjunta, função da condicional

$$\mathbf{P}(Y) = \sum_{z \in Z} \mathbf{P}(Y, z)$$ probabilidade marginal

$$\mathbf{P}(Y) = \sum_{z} \mathbf{P}(Y|z) P(z)$$ probabilidade marginal, função da condicional

a partir da probabilidade condicionada

constante de normalização

$$\frac{1}{P(dorDentes)} = \alpha$$

$$P(cárie|dorDentes) = \frac{P(cárie, dorDentes)}{P(dorDentes)}$$

$$= \frac{0{,}108 + 0{,}012}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}6$$

$$P(\neg cárie|dorDentes) = \frac{P(\neg cárie, dorDentes)}{P(dorDentes)}$$

$$= \frac{0{,}016 + 0{,}064}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}4$$

usando a constante de normalização

$$\mathbf{P}(Cárie|dorDentes) = \alpha\mathbf{P}(Cárie, dorDentes)$$

$$= \alpha[\mathbf{P}(Cárie, dorDentes, sonda) + \mathbf{P}(Cárie, dorDentes, \neg sonda)]$$

$$= \alpha[\langle 0{,}108; 0{,}016\rangle + \langle 0{,}012; 0{,}064\rangle] = \alpha\langle 0{,}12; 0{,}08\rangle = \langle 0{,}6; 0{,}4\rangle$$

não é necessário saber o valor de α!
basta normalizar, de modo que a soma dê 1

$$\frac{0{,}12}{(0{,}12+0{,}08)} = 0{,}6$$

$$\frac{0{,}08}{(0{,}12+0{,}08)} = 0{,}4$$

inferência probabilística – caso de 3 v.a.

seja X a variável de que queremos saber as probabilidades

seja **E** a lista das variáveis de evidência (ex: dorDentes)

seja ***e*** a lista de valores observados das variáveis

e seja **Y** a lista das restantes variáveis não observadas

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

não escala bem!
para n v.a. booleanas
requer tabela $O(2^n)$ e
tempo de proc. $O(2^n)$

Ciências ULisboa

independência

suponhamos a v.a. *Tempo* ϵ {*sol, chuva, nuvens, nevoeiro*}

e a distribuição conjunta **P**{*DorDentes*, *Sonda*, *Cárie*, *Tempo*}

(com 4 vezes a tabela da pág. 5)

mas o tempo não parece ser sinal de cáries

⇒ v.a. independentes

$$\mathbf{P}(DorDentes, Sonda, Cárie, Tempo) = \mathbf{P}(DorDentes, Sonda, Cárie)\,\mathbf{P}(Tempo)$$

ou $\mathbf{P}(DorDentes \mid Tempo) = \mathbf{P}(DorDentes)$

regra de Bayes

$$P(X|Y)=\frac{P(X,Y)}{P(Y)}$$

$$P(Y|X)=\frac{P(X,Y)}{P(X)}$$

logo

$$\boxed{P(Y|X)=\frac{P(X|Y)P(Y)}{P(X)}}$$

substituição da prob. conjunta pela prob. condicionada inversa

regra de Bayes

ou, com evidências

$$\mathbf{P}(Y|X,\mathbf{e})=\frac{\mathbf{P}(X|Y,\mathbf{e})P(Y|\mathbf{e})}{P(X|\mathbf{e})}$$

formulação como causa e efeito

$$P(causa|efeito)=\frac{P(efeito|causa)P(causa)}{P(efeito)}$$

P(*efeito* | *causa*) – relação na direção causal

P(*causa* | *efeito*) – relação na direção de diagnóstico

médico conhece *P*(*sintomas* | *doença*) e pretende diagnosticar: *P*(*doença* | *sintomas*)

ex. diagnóstico

meningite causa pescoço rígido 70% das vezes:

$P(s|m) = 0{,}7$

probabilidade de paciente com meningite:

$P(m) = 1/50000$ probabilidade a priori

probabil. de paciente com pescoço rígido:

$P(s) = 0{,}01$ probabilidade a priori

donde: $$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0{,}7 \times 1/50000}{0{,}01} = 0{,}0014$$

< 1/700 pescoço rígido têm meningite apesar de doença dar sintoma em 70%, mas doença é rara e pescoço rígido não!

evidências combinadas

ex: *dorDentes* e *sonda*

sabendo a distribuição conjunta (tab. pg. 5)

$$\mathbf{P}(Cárie|dorDentes, sonda) = \alpha\langle 0{,}108; 0{,}016\rangle = \langle 0{,}871; 0{,}129\rangle$$

mas a distr. conjunta não escala com o nº de variáveis!

usando a regra de Bayes:

$$\mathbf{P}(Cárie|dorDentes, sonda) = \alpha\mathbf{P}(dorDentes, sonda|Cárie)\mathbf{P}(Cárie)$$

que também não escala bem...

hélas!

independência condicionada

na realidade, *dorDentes* e *sonda* são independentes <u>dada a presença ou ausência de cárie</u>

a cárie provoca dor de dentes e cria orifício (sonda positiva)

ou seja

$$\mathbf{P}(dorDentes, sonda \mid Cárie) = \mathbf{P}(dorDentes \mid Cárie)\,\mathbf{P}(sonda \mid Cárie)$$

donde

$$\mathbf{P}(Cárie \mid dorDentes, sonda) =$$
$$= \alpha\,\mathbf{P}(dorDentes \mid Cárie)\,\mathbf{P}(sonda \mid Cárie)\,\mathbf{P}(Cárie)$$

indep. condicionada em geral

formulação geral (mais forte, porque é com as v.a.)

$$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z)$$

no caso da dentista

$$\mathbf{P}(DorDentes, Sonda|Cárie) = \mathbf{P}(DorDentes|Cárie)\mathbf{P}(Sonda|Cárie)$$

independência condicionada escala!

em variáveis binárias passa de $O(2^n)$ para $O(n)$

conceito muito importante para a IA, em particular na Aprendizagem Automática

classificador naive Bayes

caso geral

$$\mathbf{P}(Causa, Efeito_1, \ldots, Efeito_n) = \mathbf{P}(Causa) \prod_i \mathbf{P}(Efeito_i | Causa)$$

é considerado “naive” porque nem sempre todos os efeitos são independentes dada a causa

surpreendentemente, tem frequentemente muito bons resultados, mesmo quando a independência condicional não se verifica

rede Bayesiana

é uma estrutura suficiente para representar a distrib. conjunta de todas as variáveis!

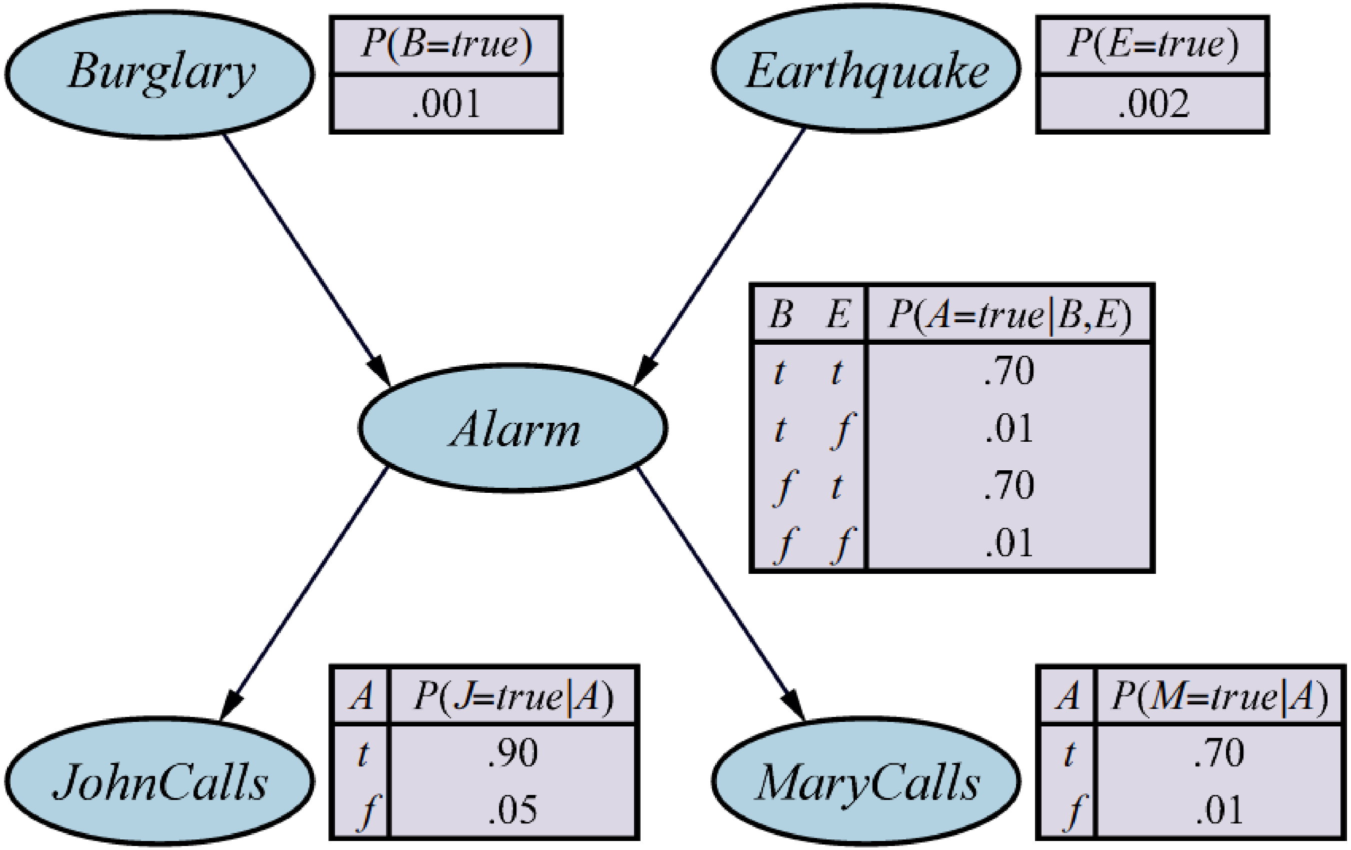
estrutura em grafo de representação de conhecimento probabilístico

- cada nó representa uma v.a.
- arco do nó *X* para o nó *Y*, refere-se como *X* é pai de *Y*
- cada nó X_i tem uma distribuição de probabilidade condicional, $\mathbf{P}(X_i \mid Pais(X_i))$, que quantifica os efeitos dos pais no nó

ideia: os pais representam as causas e os filhos representam os efeitos

modelo causal simplifica a rede, face ao modelo diagnóstico e facilita definição das probabilidades

um exemplo (clássico)



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

separação (variáveis escondidas)

- representação em rede de Bayes é muito económica
  - comparar com uma tabela com $2^5 = 32$ linhas de 5 valores cada
- a variável *A*(*larme*) “esconde” as variáveis *B* e *E* das variáveis *J* e *M*
  - tornando *J* e *M* diretamente independentes de *B* e *E*
  - *J* e *M* são idependentes entre si dada a variável *A*
- rede de Bayes não tem redundância!
  - não há o risco de falhar a lei das probabilidades

experimentar

calcular as probabilidades de:

a Mary telefonar dado que houve um tremor de terra

haver um assalto dado que o John telefonou