Inteligência Artificial

Raciocínio sobre Conhecimento Incerto

Fontes de Incerteza

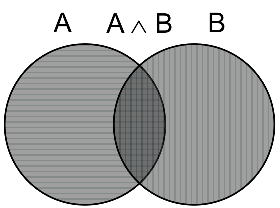
- A incerteza é a falta de informação adequada para tomar a decisão acertada
  - na medicina, pode significar a incapacidade de prescrever o melhor tratamento para o doente
  - no mundo dos negócios, pode significar perdas financeiras
  - ...

- A incerteza pode ter origem em duas fontes principais:
  - conhecimento imperfeito do domínio
    - a teoria do domínio pode ser vaga ou incompleta
    - isto leva à necessidade de utilização de conhecimento empírico, que pode obter resultados que nem sempre são correctos
  - dados de análise imperfeitos
    - os dados recolhidos podem ser imprecisos ou duvidosos
    - os dados podem ser parciais; apesar de haver formas de obter mais dados, elas podem não ser praticáveis por razões de custo, tempo ou risco

Teoria das Probabilidades

- Uma das formas de manusear conhecimento incerto é a utilização de probabilidades
  - estas atribuem um valor entre 0 e 1 aos dados (ou eventos) que não podem ser obtidos com exactidão
    - 0 indica uma crença inequívoca de que um dado é falso
    - 1 indica uma crença inequívoca de que um dado é verdadeiro
    -]0, 1[corresponde a um grau intermédio de crença quanto à verdade de um dado
  - o dado em si será verdadeiro ou falso; uma probabilidade de 0.8 indica apenas que em 80% dos casos se espera que o dado seja verdadeiro

- Axiomas
  - $0 \leq P(A) \leq 1$
  - $P(True) = 1$ $\quad\quad$ $P(False) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
    - eventos mutuamente exclusivos: $P(A \vee B) = P(A) + P(B)$
      - ex.: $P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A) = P(A) + P(\neg A) - P(False) = P(A) + P(\neg A)$
      - daqui também resulta que $P(\neg A) = 1 - P(A)$, porque $P(A \vee \neg A) = 1$



Teoria das Probabilidades (2)

- Tipos de probabilidades
  - probabilidades *a priori* (incondicionais)
    - P(Gripe) = 0.1 pode indicar, na ausência de mais informação, uma probabilidade de 10% de uma pessoa estar com gripe
  - probabilidades condicionais: calculadas com base na presença de outros eventos interdependentes
    - relevantes quando há informação sobre parte dos dados necessários à aplicação de um determinado conhecimento do domínio
    - P(Gripe | Febre) = 0.8 indicará que se um paciente tem febre, e na ausência de mais informação, a probabilidade de ter gripe é de 80%

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \wedge B) = P(A \mid B)\, P(B)$$

$$P(A \wedge B) = P(B \mid A)\, P(A)$$

    - eventos independentes: $P(A \wedge B) = P(A)\, P(B)$
      - ex.: atirar uma moeda ao ar duas vezes: $P(Cara \wedge Cara) = ½ * ½ = ¼$

Probabilidades Conjuntas

- A probabilidade condicional é definida em termos de eventos conjuntos

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

  - isto significa que para calcular P(A | B) é necessário saber a probabilidade de A e B ocorrerem simultaneamente

  - ex.: P(Gripe | Febre)
    - constrói-se uma tabela com as probabilidades conjuntas dos dois eventos:

	Febre	¬ Febre
Gripe	0.04	0.06
¬ Gripe	0.01	0.89

$$P(\text{Gripe} \mid \text{Febre}) = \frac{P(\text{Gripe} \wedge \text{Febre})}{P(\text{Febre})} = \frac{0.04}{0.04 + 0.01} = 0.80$$

  - e se houver mais do que duas variáveis a considerar?
    - para n variáveis $\Rightarrow$ 2^n entradas na tabela de probabilidades conjuntas!

Teorema de Bayes

- O Teorema de Bayes é obtido através das equações

  $P(A \wedge B) = P(A \mid B)\, P(B)$ e $P(A \wedge B) = P(B \mid A)\, P(A)$

  - igualando as duas partes da direita e dividindo por P(A), obtemos

$$\boxed{P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A)}}$$

    - P(B): probabilidade *a priori* de B, isto é, anterior à descoberta de A
    - P(B | A): probabilidade posterior, isto é, após a descoberta de A

  - o teorema também se chama fórmula de inversão, porque define P(B | A) em termos de P(A | B)
  - para que serve?
    - aparentemente, requer três termos só para calcular uma probabilidade condicional: uma probabilidade condicional e duas *a priori*
    - nalguns domínios, são conhecidas probabilidades condicionais em relações causais e pretende-se derivar um diagnóstico

Aplicação do Teorema de Bayes

- Exemplo: diagnóstico
  - um paciente tem um determinado sintoma (torcicolo): pretende-se saber se a causa é séria (meningite)
  - um médico sabe que:
    - a meningite faz com que o doente tenha torcicolos em 50% dos casos
    - a probabilidade *a priori* de um paciente ter meningite é de 1/50000, e de ter um torcicolo é de 1/20
  - seja T a proposição que afirma que o paciente tem um torcicolo, e M a proposição que afirma que o paciente tem meningite
    - P(T | M) = 0.5
    - P(M) = 1/50000
    - P(T) = 1/20

$$P(M \mid T) = \frac{P(T \mid M)\, P(M)}{P(T)} = \frac{0.5 * 1/50000}{1/20} = 0.0002$$

  - portanto, só 1 em cada 5000 pacientes com torcicolos é que têm meningite
    - apesar das pessoas com meningite costumarem ter torcicolos (50% dos casos)
    - a probabilidade *a priori* de torcicolos é muito maior do que de meningite

Aplicação do Teorema de Bayes (2)

- Porque é que não se sabe logo à partida P(M | T)?
  - o conhecimento de diagnóstico é muito mais ténue do que o conhecimento causal
  - portanto:
    - pode não haver informação sobre a probabilidade de uma pessoa com torcicolo ter meningite
      - P(M | T) é conhecimento de diagnóstico
    - mas ser possível ter uma noção consistente de quantos pacientes com meningite têm torcicolos
      - P(T | M) é conhecimento causal
  - se houver uma epidemia de meningite:
    - P(M) irá subir
    - P(M | T) deverá subir proporcionalmente a P(M)
    - a informação causal P(T | M) manter-se-á, pois reflecte o funcionamento da doença
    - o sistema é portanto robusto

Forma Genérica do Teorema de Bayes

- E se houver mais do que uma evidência (sintoma)?
  - para 2 evidências:

$$P(M \mid T \wedge S_2) = \frac{P(T \wedge S_2 \mid M)\, P(M)}{P(T \wedge S_2)}$$

    - necessário calcular $P(T \wedge S_2) = P(T \mid S_2)\, P(S_2)$
  - para n evidências: forma genérica do Teorema de Bayes

$$\boxed{P(d \mid s_1 \wedge \ldots \wedge s_n) = \frac{P(s_1 \wedge \ldots \wedge s_n \mid d)\, P(d)}{P(s_1 \wedge \ldots \wedge s_n)}}$$

    - necessário calcular $P(s_1 \wedge \ldots \wedge s_n) = P(s_1 \mid s_2 \wedge \ldots \wedge s_n)\, P(s_2 \mid s_3 \wedge \ldots \wedge s_n) \ldots P(s_n)$
    - se alguns dos sintomas forem independentes uns dos outros ($P(s_i) = P(s_i \mid s_j)$) é possível simplificar: $P(s_i \wedge s_j) = P(s_i)\, P(s_j)$
  - por vezes, podemos assumir *independência condicional* entre sintomas na presença de evidência adicional E (conhecimento do domínio):
    - $P(s_i \mid s_j, E) = P(s_i \mid E)$
      - ex.: carro com pneu furado e luzes não funcionam – 2 sintomas independentes
      - ex.: carro não pega e luzes não funcionam – 2 sintomas dependentes (bateria)

Dificuldades

- Ter em conta todas as dependências é impraticável em problemas reais
  - na prática, considera-se que a evidência é obtida aos poucos
    - o processo de *actualizações bayesianas* incorpora a evidência à medida que é obtida, actualizando a crença na hipótese considerada
      - $P(d \mid s1) \rightarrow P(d \mid s1 \wedge s2) \rightarrow P(d \mid s1 \wedge s2 \wedge s3)$
    - mesmo assim, quando não há independência condicional entre os dados é difícil e moroso recolher todas as probabilidades necessárias

- Duas hipóteses:
  - ou se assume que os dados são independentes, necessitando de menos cálculos mas obtendo resultados menos precisos (*Naïve Bayes*)
  - ou se trata das dependências entre os dados, quantificando-as e pagando o preço computacional das actualizações de crenças

- Abordagem intermédia: Redes de Crenças (Redes Bayesianas)

Redes Bayesianas

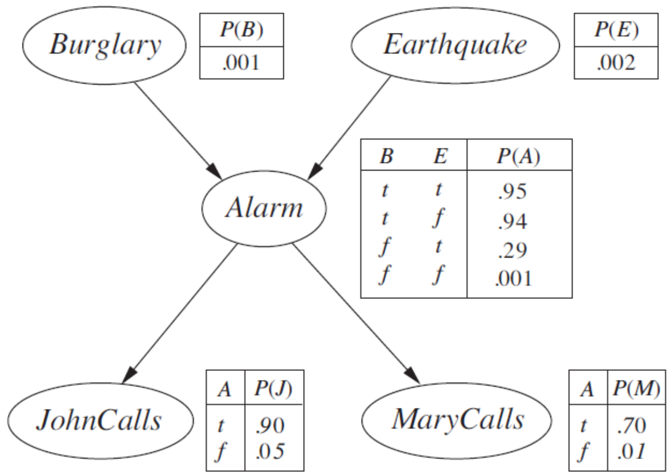
- Uma rede bayesiana é um DAG (grafo orientado acíclico)
  - os nós são variáveis aleatórias (eventos)
  - as ligações indicam dependência (influência) entre as variáveis
  - cada nó tem uma tabela de probabilidades condicionais que quantifica o efeito que os pais têm no nó

$$\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$$

$$Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$$

- A rede bayesiana pode portanto ser bastante mais compacta do que todas as distribuições conjuntas!
  - estrutura local, sistema esparso: cada uma de *n* variáveis interage com apenas $k \ll n$ outras variáveis
    - cada tabela tem apenas 2^k números, e não 2^n

Redes Bayesianas: Exemplo

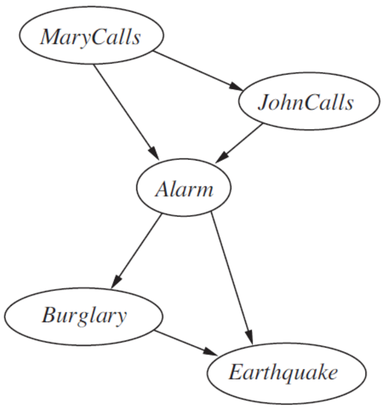


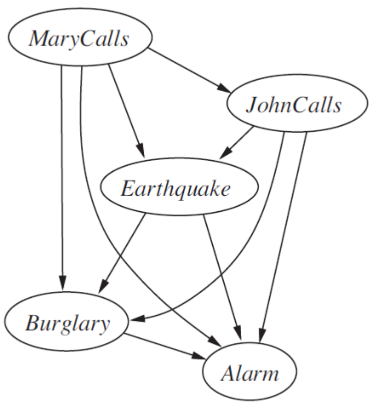
$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j \mid a) P(m \mid a) P(a \mid \neg b \wedge \neg e) P(\neg b) P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$

$$\mathbf{P}(MaryCalls \mid JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls \mid Alarm)$$

Redes Bayesianas: Construção

- Incluir apenas dependências diretas
- Dependências ténues devem ser evitadas: o ganho de precisão não compensa a complexidade adicional da rede
- Capturar relações causais (por oposição às de diagnóstico)
  - menos relações e mais fáceis de computar
  - mais independência condicional
- A ordem de consideração dos eventos é relevante

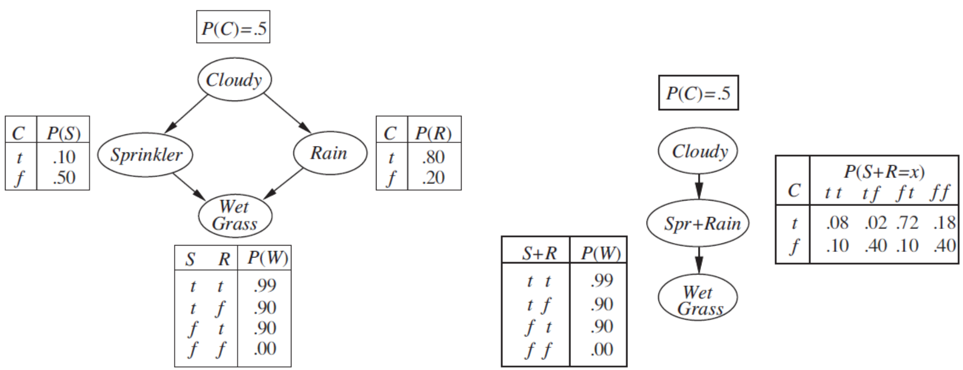




Redes Bayesianas: Complexidade

- Complexidade da inferência depende da estrutura da rede
  - redes simplesmente conectadas (*polytrees*)
    - só há um caminho entre cada 2 nós na rede
    - complexidade temporal e espacial linear no tamanho da rede
  - redes multi-conectadas
    - inferência é NP-hard (complexidade exponencial)

- Algoritmos de clustering (join tree)
  - transformar uma rede multi-conectada numa polytree



(a) (b)

Crença e Descrença

- Outro problema com a abordagem probabilística é a relação entre crença e descrença
  - à primeira vista, a crença é o oposto da descrença
    - $P(H) + P(\neg H) = 1$ $\qquad$ $P(H) = 1 - P(\neg H)$
    - $P(H \mid E) = 1 - P(\neg H \mid E)$
  - regras extraídas de peritos do domínio (p. ex., MYCIN):

    ```
    SE E1 e E2 e E3 ENTÃO H (.7)
    ```

    - na presença das três premissas E1, E2 e E3 a conclusão H tem uma probabilidade de 70%
      - $P(H \mid E1 \wedge E2 \wedge E3) = 0.7$
      - $P(\neg H \mid E1 \wedge E2 \wedge E3) = 1 - 0.7 = 0.3$ → **este valor faz sentido?**
    - a regra indica uma relação entre as premissas e H, mas não diz nada sobre a relação entre as premissas e $\neg H$
    - $P(H \mid E)$ implica uma relação causa-efeito entre E e H, mas pode não haver essa relação entre E e $\neg H$
      - $P(H \mid E) = 1 - P(\neg H \mid E)$ assume a relação entre E e $\neg H$ se houver relação entre E e H

Modelo dos Factores de Certeza

- O modelo dos factores de certeza (CF) foi inicialmente desenvolvido para o SP MYCIN

- O CF indica o grau de confirmação de uma dada conclusão:
  - CF(H | E) = MB(H, E) – MD(H, E)
    - CF é o factor de certeza (*certainty factor*) na hipótese H dada a evidência E
    - MB é a medida de crença (*measure of belief*) acrescida em H devido a E
    - MD é a medida de descrença (*measure of disbelief*) acrescida em H devido a E
  - o factor de certeza é uma forma de combinar crença e descrença
    - pode ser usado para ordenar hipóteses por ordem de importância

  - algumas características dos valores:

Gamas	Hipótese verdadeira $P(H \mid E) = 1$	Hipótese falsa $P(\neg H \mid E) = 1$	Falta de evidência
$0 \leq MB \leq 1$ $0 \leq MD \leq 1$ $-1 \leq CF \leq 1$	MB = 1 MD = 0 CF = 1	MB = 0 MD = 1 CF = -1	MB = 0 MD = 0 CF = 0

Modelo dos Factores de Certeza (2)

- $CF \in [-1; 1]$
  - valor positivo: evidência suporta a hipótese
    - $MB > MD$
  - valor negativo: a evidência favorece a negação da hipótese
    - $MB < MD$
      - há mais razões para não acreditar na hipótese do que para acreditar nela
- Relação entre as certezas em H e $\neg H$:
  - $CF(H \mid E) = -\ CF(\neg H \mid E)$ ou $CF(H \mid E) + CF(\neg H \mid E) = 0$
    - evidência suportando uma hipótese reduz o suporte à negação da hipótese numa porção equivalente
    - $MB(H, E) = MD(\neg H, E)$ $MD(H, E) = MB(\neg H, E)$
      - a crença numa hipótese corresponde à descrença na negação dessa hipótese

Modelo dos Factores de Certeza (3)

- Problema com CF = MB – MD:
  - uma porção de evidência contra uma hipótese pode controlar um conjunto de várias evidências a favor da hipótese
  - ex.:
    - MB = 0.999 obtido a partir de 10 evidências
    - MD = 0.799 obtido a partir de 1 evidência
    - CF = 0.999 – 0.799 = 0.200 é um valor demasiado baixo para as evidências existentes!

- Alteração da fórmula de cálculo do CF :

$$CF = \frac{MB - MD}{1 - \min(MB, MD)}$$

  - CF = (0.999 – 0.799) / (1 – min(0.999, 0.799)) =
    = 0.200 / (1 – 0.799) = 0.995

Factores de Certeza no MYCIN

- O SP MYCIN explorou uma alternativa à utilização do teorema de Bayes, utilizando critérios de decisão que ligam evidências a hipóteses

- O sistema utiliza regras do tipo:
  - SE $e_1 \wedge ... \wedge e_k$ ENTÃO h com CF de certeza
    - CF: quantificação subjectiva da certeza associada à conclusão da regra
    - CF = 1: temos a certeza de que a conclusão é verdadeira se as condições forem completamente satisfeitas
    - CF = -1: temos a certeza de que a conclusão é falsa se as condições forem completamente satisfeitas
    - CF > 0: as condições são evidência sugestiva em favor da conclusão h
    - CF < 0: as condições são evidência contra h

  - exemplo:

```
IF
  The stain of the organism is gramneg, and
  The morphology of the organism is rod, and
  The aerobicity of the organism is aerobic
THEN
  There is strongly suggestive evidence (.8) that the class of
  the organism is enterobacteriaceae
```

Factores de Certeza no MYCIN (2)

- Tipos de incerteza:
  - incerteza nas conclusões das regras
    - devido a um conhecimento imperfeito do domínio
  - incerteza nas premissas das regras
    - incerteza dos dados recolhidos (ou fornecidos pelo utilizador do sistema)
    - encadeamento das regras – a conclusão (incerta) de uma regra pode ser premissa de outra regra

Factores de Certeza no MYCIN (3)

- Combinação de factores de certeza:
  - CF do conjunto das premissas:
    - $CF(e_1 \wedge e_2) = \min(CF(e_1), CF(e_2))$
    - $CF(e_1 \vee e_2) = \max(CF(e_1), CF(e_2))$
    - o valor de 0.2 é usado como limite mínimo para a activação da regra
  - combinação do CF das premissas com o CF da regra:
    - CF(conclusão) = CF(premissas) * CF(regra)
  - combinação de CFs de regras que suportam a mesma conclusão:

$$CF(CF_1, CF_2) = \begin{cases} CF_1 + CF_2\,(1 - CF_1) & \text{se } CF_1, CF_2 > 0 \\ CF_1 + CF_2\,(1 + CF_2) & \text{se } CF_1, CF_2 < 0 \\ (CF_1 + CF_2) / (1 - \min(|CF_1|, |CF_2|)) & \text{senão} \end{cases}$$

Exemplo com Factores de Certeza

- Regras para diagnóstico da avaria de um automóvel:

  R1: SE motor não dá sinal E bateria má ENTÃO problema é bateria.
  R2: SE luzes fracas ENTÃO bateria má (cf 0.5).
  R3: SE rádio fraco ENTÃO bateria má (cf 0.5).

  R4: SE motor dá sinal E cheira a gasolina ENTÃO problema é encharcado (cf 0.8).

  R5: SE motor dá sinal E indicador gasolina vazio ENTÃO problema é sem gasolina (cf 0.9).
  R6: SE motor dá sinal E indicador gasolina baixo ENTÃO problema é sem gasolina (cf 0.3).

Evidências	*Problema*
o motor dá sinal; cheira a gasolina; indicador de gasolina vazio	
o motor dá sinal; cheira a gasolina cf 0.5; indicador de gasolina vazio	
o motor não dá sinal; as luzes estão fracas; o rádio está fraco	

Problemas

- Disjunção de evidências

  R_1: SE e_1 ENTÃO h (CF)
  R_2: SE e_2 ENTÃO h (CF)
  …
  R_n: SE e_n ENTÃO h (CF)

  - conduz a resultados diferentes de

  R: SE $e_1 \vee e_2 \vee \ldots \vee e_n$ ENTÃO h (CF)

Seja CF = 0.6 e cada e_i verdadeiro

R_1: h (0.6 = 0.6)
R_1+R_2: h (0.6 + 0.6(1-0.6) = 0.84*
R_1+R_2+R_3: h (0.84 + 0.6(1-0.84) = 0.936*
…

Enquanto que *R: h (0.6)*

Problemas (2)

- Encadeamento de regras
  - proposições:
    - I: irrigador funcionou esta noite
    - H: relva húmida
    - C: choveu esta noite
  - regras:
    - R_1: SE I ENTÃO H (0.9) regra causal
    - R_2: SE H ENTÃO C (0.8) regra de diagnóstico
  - facto: I (sabe-se que o irrigador funcionou esta noite)
  - inferência:
    - de I e R_1 vem H (0.9)
    - de H e R_2 vem C (0.9*0.8 = 0.72) !?!

Aplicabilidade do Modelo

- O modelo dos factores de certeza tem por base a teoria das probabilidades, mas é ao mesmo tempo parcialmente “ad hoc”

- Do modelo dos Fatores de Certeza:
  - $CF(A \wedge B) = \min(CF(A), CF(B))$

- Da Teoria das Probabilidades:
  - Eventos dependentes:
    - $P(A \wedge B) = P(A \mid B)\,P(B) \leq P(B)$
    - $P(A \wedge B) = P(B \mid A)\,P(A) \leq P(A)$
  - De onde:
    - $P(A \wedge B) \leq \min(P(A), P(B))$
  - Eventos independentes:
    - $P(A \wedge B) = P(A)\,P(B) \leq \min(P(A), P(B))$
      - $P(A \wedge B) = \min(P(A), P(B))$ sse $P(A) = 1 \vee P(B) = 1$

Aplicabilidade do Modelo (2)

- Principais vantagens:
  - cálculos simples a partir dos quais a incerteza é propagada pelo sistema
  - modelo simples e fácil de entender

- Contudo, nalgumas situações este modelo pode obter resultados contraditórios relativamente à teoria das probabilidades

- Os factores de certeza resultam melhor quando:
  - as regras são puramente de diagnóstico (MYCIN) ou puramente causais
    - p. ex., não juntar SE irrigador ENTÃO relva molhada (causal)
      SE relva molhada ENTÃO choveu (diagnóstico)
  - as evidências são introduzidas apenas nas “raízes” do conjunto das regras
    - isto é, regras no meio do encadeamento obtêm *todas* as premissas a partir de outras regras
  - o encadeamento das regras é curto

- Nestas circunstâncias, a inferência com factores de certeza é equivalente à inferência com Redes de Crenças