Parte III: IA e Incerteza.

Prof. Fabiano Araujo Soares, Dr. / FGA 0221 - Inteligência Artificial

Universidade de Brasília

2025

Como podemos ensinar uma máquina a mudar de ideia quando novas evidências surgem — e por que isso é essencial para a inteligência artificial do futuro?

Como podemos ensinar uma máquina a mudar de ideia quando novas evidências surgem — e por que isso é essencial para a inteligência artificial do futuro?

Nos métodos bayesianos, a incerteza não é um obstáculo — é a matéria-prima da inteligência.

Relembrando o modelo das variáveis odontológicas

- Vamos retornar ao modelo probabilístico para dor de dente (Toothache), cárie (Cavity) e sonda presa no dente (Catch).
- Agora, adicionando a variável 'Tempo atimosférico':

$$P(Toothache, Catch, Cavity, Weather)$$

- Qual a relação entre essas quatro variáveis?
- Podemos usar a regra do produto:

$$P(Toothache, Catch, Cavity, Weather) =$$
$$= P(Cloud \mid Toothache, Catch, Cavity)P(Toothache, Catch, Cavity)$$

Independência entre variáveis

- A variável "tempo atmosfético"é independente das demais.
- Isso nos leva a simplificações importantes na fórmula da distribuição conjunta.
- Definição de independência: $P(a \wedge b) = P(a)P(b)$.
- Para variáveis, independência: $P(X, Y) = P(X)P(Y)$.
- Proposições de independência ajudam a reduzir a quantidade de informação a ser especificada.
- Podem facilitar a representação e a inferência em domínios complexos.

- Nesse caso teremos:

$$P(Cloud \mid Toothache, Catch, Cavity) = P(Cloud)$$

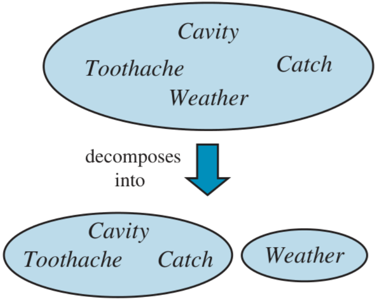
- Essa equação nos leva a:

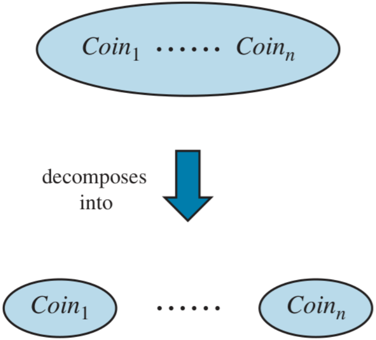
$$P(Toothache, Catch, Cavity, Weather) = P(Toothache, Catch, Cavity)P(Weather)$$

Independência entre variáveis

- Determinar a independência de variáveis (ou proposições) pode reduzir dramaticamente a quantidade de informação para especificar a distribuição conjunta completa;
- Se o conjunto completo de variáveis pode ser dividido em subconjuntos independentes, então a distribuição conjunta completa pode ser fatorada em conjuntos separados;
- Quando estão disponíveis, proposições independentes podem ajudar a reduzir o tamanho da representação do domínio e a complexidade do problema de inferência.

Independência entre variáveis





Regra do Produto revisitada

- A regra do produto pode ser escrita de duas maneiras:

$$P(a \wedge b) = P(a)P(b|a)$$
$$P(a \wedge b) = P(b)P(a|b)$$

- Igualando e dividindo pelo lado direito leva à famosa Regra de Bayes.

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

- Esta é a base para inferência probabilística em IA.

Regra de Bayes: Formulação Geral

- A formulação clássica para a equação de Bayes é

Equação de Bayes

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

- Formulação alternativa, considerando múltiplas evidências e:

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

Aplicação prática da equação de Bayes

- Muitas vezes, percebemos como evidência o efeito de alguma causa desconhecida e gostaríamos de determinar essa causa.
- reformulando a equação de Bayes para atender essa demanda temos:

Equação de Bayes – Causa e Efeito

$$P(Causa|Efeito) = \frac{P(Efeito|Causa)P(Causa)}{P(Efeito)}$$

- A probabilidade condicional $P(Efeito|Causa)$ quantifica a relação na direção **causal**, enquanto $P(Causa|Efeito)$ descreve a direção **diagnóstica**.
- Em tarefas como **diagnóstico**, normalmente temos probabilidades condicionais de relações causais. O médico sabe $P(Sintomas|Doena)$ e deseja obter um **diagnóstico**: $P(Doena|Sintomas)$.

Exemplo Clássico: Classificador de Spam

- Imagine que queremos construir um sistema que determine se um e-mail é spam ou não com base em suas palavras.
- O teorema de Bayes nos permite calcular a probabilidade de um e-mail ser spam dado que ele contém certas palavras, como "**ganhe dinheiro rápido**".

- Exemplo passo a passo — Suponha que:
  - A probabilidade de qualquer e-mail ser spam é $P(spam) = 0.3$ (30% dos e-mails são spam).
  - A probabilidade de um e-mail conter a frase "ganhe dinheiro rápido" dado que é spam é $P(ganhe\ dinheiro\ rápido \mid spam) = 0.8$.
  - A probabilidade de um e-mail conter essa frase mesmo sendo legítimo é $P(ganhe\ dinheiro\ rápido \mid \neg spam) = 0.1$.

Exemplo Clássico: Classificador de Spam

- Usando o teorema de Bayes:

$$P(spam \mid ganhe\ dinheiro\ rápido) = \frac{P(ganhe\ dinheiro\ rápido \mid spam)P(spam)}{P(ganhe\ dinheiro\ rápido)}$$

- O denominador pode ser calculado com a regra da probabilidade total:

$$\begin{aligned} P(ganhe\ dinheiro\ rápido) =& P(ganhe\ dinheiro\ rápido \mid spam)P(spam) + \\ & + P(ganhe\ dinheiro\ rápido \mid \neg spam)P(\neg spam) = \\ & (0.8)(0.3) + (0.1)(0.7) = 0.24 + 0.07 = 0.31 \end{aligned}$$

- Agora aplicamos:

$$P(spam \mid ganhe\ dinheiro\ rápido) = \frac{0.8 \times 0.3}{0.31} = \frac{0.24}{0.31} \approx 0.774$$

Exemplo Clássico: Classificador de Spam

- Ou seja, há cerca de 77,4% de chance do e-mail ser spam se contiver essa frase.
- Esse raciocínio é a base do algoritmo *Naive Bayes*, amplamente usado em processamento de linguagem natural.
- O algoritmo *Naive Bayes* estende esse cálculo para múltiplas palavras, assumindo independência entre elas (daí o termo "**ingênuo**") para simplificar os cálculos.

- **Em resumo**: o **teorema de Bayes** permite que máquinas atualizem crenças com base em **evidências**, tornando-se essencial em IA para lidar com incerteza.

Exemplo médico: Diagnóstico

- Sintomas (efeito) $\rightarrow$ Doença (causa)
- Exemplo: **meningite** causa **torcicolo** em 70% dos casos.
- Probabilidade de torcicolo em quem tem meningite: 0,7.
- Probabilidade de qualquer paciente ter meningite: $\approx$1 em 50.000.
- Probabilidade de qualquer paciente ter torcicolo: 1%.
- Seja s a proposição de que o paciente tenha torcicolo e m a probabilidade de o paciente tem meningite, temos

$$P(s \mid m) = 0.7$$
$$P(m) = \frac{1}{50000}$$
$$P(s) = 0.01$$

UnB
UnB - FCTE
Faculdade de Ciências e Tecnologias em Engenharia
conhecimento em movimento
sociedade em transformação

Exemplo médico: Diagnóstico

- Teremos então:

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 0.00002}{0.01} = 0.0014$$

- Ou seja, temos uma expectativa de que 0,14% dos pacientes com torcicolo tenham meningite.
- Note que, embora um torcicolo seja comum em pacientes com meningite (probabilidade de 0,7), a probabilidade de meningite em pacientes com torcicolo permanece pequena;
- Isso porque a probabilidade a priori de torcicolo (por qualquer causa) é muito maior do que a probabilidade *a priori* para meningite.

Regra de Bayes com normalização

- Podemos evitar a probabilidade *a priori* da evidência (($P(s)$ nesse caso), utilizando probabilidade *a posteriori* para cada variável em análise (aqui m e $\neg m$) e então, normalizar os resultados:

$$P(M|s) = \alpha\langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m)\rangle$$

- Para isso temos que estimar $P(s|\neg m)$ ao invés de $P(s)$. As vezes isso é mais fácil, as vezes é mais difícil.
- A forma geral da equação de Bayes com normalização é

$$\mathbf{P}(Y \mid X) = \alpha \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

- onde α é a constante de normalização necessária para fazer $P(Y|X)$ somar 1.

Aplicando a regra de Bayes — Combinando evidências

- O que acontece quando temos mais de uma eviência?
- Por exemplo, o que um dentista pode concluir quando o paciente tem uma dor de dente e a sonda dele prende no dente do paciente?
- Da tabela de distribuições temos:

$$P(Cavity \mid toothache \wedge catch) = \alpha\langle 0.108, 0.016\rangle \approx \langle 0.871, 0.129\rangle$$

- Podemos usar a regra de Bayes para reformular o problema:

$$P(Cavity \mid toothache \wedge catch) = \alpha P(toothache \wedge catch) P(Cavity)$$

Aplicando a regra de Bayes — Combinando evidências

- Para essa reformulação funcionar, precisamos saber a probabilidade condicional da conjunção $toothache \wedge catch$ para cada valor de $Cavity$.
- Se houver n possíveis variáveis de evidência (raios X, dieta, higiene oral, etc.), então há $O(2^n)$ combinações possíveis de valores observados para os quais precisaríamos conhecer probabilidades condicionais.
- Para progredir, precisamos encontrar algumas afirmações adicionais sobre o domínio que permitam simplificar as expressões.

Aplicando a regra de Bayes — Combinando evidências

- Seria bom se *Toothache* e *Catch* fossem independentes, mas eles não são: Se a sonda prender no dente, é possível que exista uma cárie e que essa cárie cause a dor de dente;
- Essas variáveis são independentes, no entanto, dado a presença ou ausência de uma cárie;
- Cada um é causado diretamente pela cárie, mas não tem efeito direto um no outro: a dor de dente depende do estado dos nervos no dente, e a precisão da sonda depende principalmente da habilidade do dentista, ao qual a dor de dente é irrelevante.

Aplicando a regra de Bayes — Combinando evidências

- Matematicamente temos:

$$P(toothache \wedge catch \mid Cavity) = P(toothache \mid Cavity)P(catch \mid Cavity)$$

- Esta equação expressa a independência condicional da dor de dente e da sonda prender na cavidade da cárie.
- Podemos então chegar a:

$$P(Cavity \mid toothache \wedge catch) = P(toothache \mid Cavity)P(catch \mid Cavity)P(Cavity)$$

- Agora, os requisitos das informações são os mesmos da inferência, usando cada pedaço de evidência separadamente.

Aplicando a regra de Bayes — Combinando evidências

- A definição geral de independência condicional de duas variáveis X e Y, dada uma terceira variável Z, é

Independência Condicional

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

- no domínio odontológico teríamos:

$$P(toothache, catch \mid Cavity) = P(toothache \mid Cavity)P(catch \mid Cavity)$$

Modelo de Bayes Ingênuo (*Naive Bayes*)

- O exemplo odontológico ilustra um padrão de ocorrência comum em que uma única causa influencia diretamente uma série de efeitos, todos os quais são condicionalmente independentes, dado a causa (Causa única influencia múltiplos efeitos independentes condicionalmente).
- A distribuição conjunta completa pode ser escrita como

$$P(\mathrm{Causa} \mid \mathrm{Efeito}_1, \mathrm{Efeito}_2, \ldots) \propto P(\mathrm{Causa}) \prod_i P(\mathrm{Efeito}_i \mid \mathrm{Causa})$$

- Tal distribuição de probabilidade é chamada de "ingênuo" porque muitas vezes é usada como uma **suposição simplificadora**;
- nos casos em que as variáveis de "efeito" não são estritamente independente dada a variável causa.

UnB

UnB - FCTE Faculdade de Ciências e Tecnologias em Engenharia

conhecimento em movimento sociedade em transformação

- Na prática, sistemas ingênuos de Bayes geralmente funcionam muito bem, mesmo quando as condições de independência condicional não são estritamente verdadeiras.
- Para usar um modelo ingênuo de Bayes, podemos aplicar a equação anterior para obter a probabilidade de causa dado alguns efeitos observados.
- Chamaremos os efeitos observados de $E = e$, enquanto o restante das variáveis de efeito Y não são observadas.
- Dessa forma, temos

$$P(\mathrm{Causa} \mid \mathsf{e}) = \alpha \sum_{y} P(\mathrm{Causa}, \mathsf{e}, \mathsf{y})$$

Modelo de Bayes Ingênuo (*Naive Bayes*)

- O que nos leva a

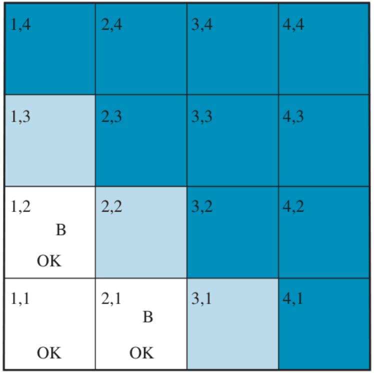
$$
\begin{aligned}
P(Causa \mid \mathrm{e}) &= \alpha \sum_y P(Causa) P(y \mid Causa) \left(\prod_j P(\mathrm{e}_j \mid Causa) \right) \\
&= \alpha P(Causa) \left(\prod_j P(\mathrm{e}_j \mid Causa) \right) \sum_y P(y \mid Causa) \\
&= \alpha P(Causa) \prod_j P(\mathrm{e}_j \mid Causa)
\end{aligned}
$$

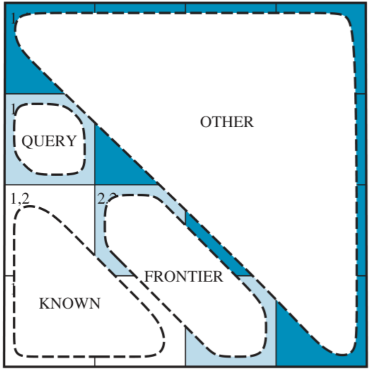
- Ou seja, para cada causa possível, multiplique a probabilidade a priori da causa pelo produto das probabilidades condicionais dos efeitos observados dada a causa e então normalize o resultado.

Wumpus World Revisado

- Vamos retornar ao mundo de Wumpus mas agora vamos simplifica-lo e considerar apenas os poços (ignorar o Wumpus e o ouro);
- Vamos considerar um caso particular onde o agente visitou os quadrados [1,1], [1,2] e [2,1];
- O agente sentiu uma brisa nos quadrados [1,2] e [2,1], ou seja não há mais quadrados seguros para entrar. O que fazer?
- O agente sabe que [1,1] não contem poço e que a probabilidade de poço nos demais quadrados é 0,2.

Wumpus World Revisado





Wumpus World Revisado

- Queremos uma variável booleana $P_{i,j}$ que é verdadeira se o quadrado [i,j] conter um poço;
- Também temos a variável booleana $B_{i,j}$ se o quadrado [i,j] possuir brisa;
- O próximo passo é especificar a distribuição conjunta completa, $P(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$. Aplicando a regra do produto, temos

$$P(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1}) = P(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})P(P_{1,1}, \ldots, P_{4,4})$$

- e temos também

$$\mathrm{P}(P_{1,1}, \ldots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} \mathrm{P}(P_{i,j})$$

Wumpus World Revisado

- Como visitamos os quadrados [1,1], [1,2] e [2,1] e sentimos brisa nos quadrados [1,2] e [2,1] podemos criar duas proposições:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$

- e também

$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

- Desejamos responder questões como $P(P_{1,3} \mid known, b)$ ou seja, qual a probabilidade de poço no quadrado [1,3]?

Wumpus World Revisado

- Para responder essa questão, usamos a equação $P(X \mid \mathrm{e}) = \alpha P(X, \mathrm{e}) = \alpha \sum_y P(X, \mathrm{e}, \mathrm{y})$,
- Logo, temos:

$$P(P_{1,3} \mid known, b) = \alpha \sum_{unknown} P(P_{1,3}, known, b, unknown)$$

- Vamos considerar os quadrados desconhecidos com *unknown*, os quadrados da *frontier* aqueles vizinhos aos visitados e os quadrados que são desconhecidos e não fazem parte da fronteira como *other*.

Wumpus World Revisado

- Temos então:

$$
\begin{aligned}
&P(P_{1,3} \mid known, b) \\
&= \alpha \sum_{unknown} P(P_{1,3}, known, b, unknown) \\
&= \alpha \sum_{unknown} P(b \mid P_{1,3}, known, unknown) P(P_{1,3}, known, unknown) \\
&= \alpha \sum_{frontier} \sum_{other} P(b \mid known, P_{1,3}, frontier, other) P(P_{1,3}, known, frontier, other) \\
&= \alpha \sum_{frontier} \sum_{other} P(b \mid known, P_{1,3}, frontier) P(P_{1,3}, known, frontier, other),
\end{aligned}
$$

Wumpus World Revisado

- Usando independência, o primeiro termo da expressão não depende de Other podemos mover o somatório para dentro:

$$P(P_{1,3} \mid known, b) = \alpha \sum_{frontier} P(b \mid known, P_{1,3}, frontier) \sum_{other} P(P_{1,3}, known, frontier, other)$$

- Por independência, o termo a direita pode ser fatorado e então reorganizado

$$\begin{aligned}
&P(P_{1,3} \mid known, b) \\
&= \alpha \sum_{frontier} P(b \mid known, P_{1,3}, frontier) \sum_{other} P(known)P(frontier)P(other) \\
&= \alpha P(known)P(P_{1,3}) \sum_{frontier} P(b \mid known, P_{1,3}, frontier)P(frontier) \sum_{other} P(other) \\
&= \alpha' P(P_{1,3}) \sum_{frontier} P(b \mid known, P_{1,3}, frontier)P(frontier),
\end{aligned}$$

Wumpus World Revisado

- Onde $P(known)$ passou para a constante de normalização e $\sum_{other} P(other) = 1$.
- Observe que $P(b \mid known, P_{1,3}, frontier)$ é 1 se brisa é consistente com as variáveis e 0 caso contrário;
- Temos então:

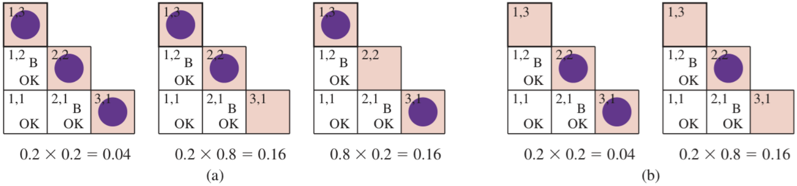


Figure 12.6 Consistent models for the frontier variables, $P_{2,2}$ and $P_{3,1}$, showing $P(frontier)$ for each model: (a) three models with $P_{1,3} = true$ showing two or three pits, and (b) two models with $P_{1,3} = false$ showing one or two pits.

Wumpus World Revisado

- O que nos leva a

$$P(P_{1,3} \mid known, b) = \alpha' \langle 0.2(0,04 + 0.16 + 0.16), 0.8(0.4 + 0.16) \rangle \approx \langle 0.31, 0.69 \rangle.$$

- O seja, [1,3] contem um poço com probabilidade 0,31;
- De forma similar podemos chegar ao fato que o quadrado [2,2] tem um poço com probabilidade 0,86.

Resumo e Aplicações

- **Conceitos principais**:
  - independência;
  - produto;
  - equação de Bayes,
  - independência condicional,
  - Modelo Ingênuo de Bayes (Naive Bayes).
- Aplicações práticas em diagnóstico, processamento de linguagem natural, visão robótica, agentes inteligentes, inferência probabilística.
- Limitações dos modelos ingênuos e avanços atuais.

Discussão / Perguntas

- Em quais contextos a independência condicional é razoável?
- Como modelos probabilísticos podem melhorar diagnósticos médicos?
- Exemplos no nosso cotidiano onde usamos inferência bayesiana sem perceber.

Referências

- Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach.* 4ª ed.
- Slides oficiais: *aima.cs.berkeley.edu*
- Outros: materiais complementares em IA.

Obrigado!

E-mail: fabianosoares@unb.br

LinkedIn: https://www.linkedin.com/in/fabiano-soares-06b6a821a/

Site do curso: https://www.fabianosoares.eng.br/fga0221-inteligência-artificial