www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Inferência com o Uso de Distribuições Conjuntas

Vamos descrever um método simples de inferência probabilística, isto é, a computação de probabilidades posteriores de proposições de consulta dada uma evidência observada. Utilizaremos a distribuição conjunta total como a “base de conhecimento” a partir da qual poderão ser derivadas respostas para todas as perguntas. Ao longo do caminho, também introduziremos várias técnicas úteis para manipular equações que envolvem probabilidades.

Começaremos com um exemplo muito simples: um domínio que consiste apenas nas três variáveis booleanas **DorDeDente**, **Cárie** e **Boticão** (a horrível tenaz de aço com que o dentista agarra o dente para extraí-lo). A distribuição conjunta total é uma tabela 2 × 2 × 2, como mostra a tabela abaixo:

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	0,108	0,012	0,072	0,008
¬cárie	0,016	0,064	0,144	0,576

Note que as probabilidades na distribuição conjunta têm a soma 1, conforme exigem os axiomas de probabilidade. Note também que a equação fornece um caminho direto para calcular a probabilidade de qualquer proposição, simples ou complexa: simplesmente identificamos os mundos possíveis nos quais a proposição é verdadeira e somamos suas probabilidades. Por exemplo, existem seis eventos atômicos em que cárie $\vee$ dordedente é válida:

$$P(cárie \vee dordedente) = 0{,}108 + 0{,}012 + 0{,}072 + 0{,}008 + 0{,}016 + 0{,}064 = 0{,}28$$

Uma tarefa particularmente comum é extrair a distribuição sobre algum subconjunto de variáveis ou sobre uma única variável. Por exemplo, a adição das entradas da primeira linha produz a probabilidade incondicional ou probabilidade marginal de cárie:

$$P(cárie) = 0{,}108 + 0{,}012 + 0{,}072 + 0{,}008 = 0{,}2$$

Esse processo é chamado marginalização ou totalização porque totalizamos as probabilidades para cada valor possível de outras variáveis, assim excluindo-as da equação. Podemos escrever a regra geral de marginalização a seguir para quaisquer conjuntos de variáveis Y e Z:

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

onde $\sum_{z \in Z}$ significa a soma sobre todas as combinações possíveis de valores do conjunto de variáveis Z. Por vezes, abreviamos como Σz, deixando Z implícito. Apenas utilizamos a regra como:

$$\mathbf{P}(\mathit{Cárie}) = \sum_{z \in \{\mathit{Boticão},\, \mathit{Dor\ de\ dente}\}} \mathbf{P}(\mathit{Cárie}, \mathbf{z})$$

Uma variante dessa regra envolve probabilidades condicionais em vez de probabilidades conjuntas, usando-se a regra do produto:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z})$$

Essa regra é chamada **condicionamento**. Marginalização e condicionamento se mostrarão regras úteis para todos os tipos de derivações que envolverem expressões de probabilidade. Na maioria dos casos, estaremos interessados em calcular probabilidades condicionais de algumas variáveis, dada alguma evidência sobre outras. As probabilidades condicionais podem ser descobertas usando-se primeiro a equação para obter uma expressão em termos de probabilidades não condicionais, e então avaliando-se a expressão a partir da distribuição conjunta total. Por exemplo, podemos calcular a probabilidade de uma cárie, dada a evidência de uma dor de dente, como a seguir:

$$\begin{aligned} P(\mathit{cárie} \mid \mathit{dordedente}) &= \frac{P(\mathit{cárie} \wedge \mathit{dordedente})}{P(\mathit{dordedente})} \\ &= \frac{0{,}108 + 0{,}012}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}6 \end{aligned}$$

Só para conferir, também podemos calcular a probabilidade de não haver nenhuma cárie, dada uma dor de dente:

$$\begin{aligned} P(\neg\mathit{cárie} \mid \mathit{dordedente}) &= \frac{P(\neg\mathit{cárie} \wedge \mathit{dordedente})}{P(\mathit{dordedente})} \\ &= \frac{0{,}016 + 0{,}064}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}4 \end{aligned}$$

Conforme esperado, as duas variáveis somam 1. Note que, nesses dois cálculos, a expressão 1/P(dordedente) permanece constante, não importando que valor de Cárie calculamos. De fato, ela pode ser visualizada como uma constante de normalização para a distribuição P(Cárie | dordedente), assegurando que a soma será 1. Ao longo dos capítulos que lidam com probabilidade, usaremos α para denotar tais constantes. Com essa notação, podemos transformar as duas equações precedentes em uma:

$$
\begin{aligned}
\mathbf{P}(C\acute{a}rie \mid dordedente) &= \alpha\mathbf{P}(C\acute{a}rie, dordedente) \\
&= \alpha\,[\mathbf{P}(C\acute{a}rie, dordedente, botic\tilde{a}o) + \mathbf{P}(C\acute{a}rie, dordedente, \neg botic\tilde{a}o)] \\
&= \alpha\,[\ \langle 0{,}108,\ 0{,}016\rangle \ + \ \langle 0{,}012,\ 0{,}064\rangle\ ] = \alpha\ \langle 0{,}12,\ 0{,}08\rangle \ = \ \langle 0{,}6,\ 0{,}4\rangle
\end{aligned}
$$

Em outras palavras, podemos calcular P(Cárie | dor de dente) mesmo se não soubermos o valor de P(dor de dente)! Esquecemos temporariamente o fator 1/P(dor de dente) e somamos os valores de cárie e ¬cárie, obtendo 0,12 e 0,08. Essas são as proporções relativas corretas, mas não perfazem 1, de modo que as normalizamos, dividindo cada uma por 0,12 + 0,08, ficando com as probabilidades verdadeiras de 0,6 e 0,4. A normalização acaba por ser um atalho útil em muitos cálculos de probabilidade, tanto para tornar a computação mais fácil como para permitir-nos continuar quando alguma avaliação de probabilidade (como P(dor de dente)) não estiver disponível.

Caso você tenha feito o curso de Machine Learning aqui na DSA, deve se lembrar de quantas vezes você aplicou normalização aos dados, antes de treinar algoritmos de Machine Learning.

Referências:

Livro: Inteligência Artificial
Autor: Peter Norvig