Data Science Academy

www.datascienceacademy.com.br

# Introdução à Inteligência Artificial

## Independência

Vamos expandir a distribuição conjunta total tabela abaixo, adicionando uma quarta variável, Tempo. A distribuição conjunta total então se torna P(DorDeDente, Boticão, Cárie, Tempo), que tem 2 × 2 × 2 × 4 = 32 entradas.

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

Ela contém quatro "edições" da tabela, uma para cada espécie de tempo. Parece natural indagar, que relacionamento essas edições mantêm umas com as outras e com a tabela original de três variáveis. Por exemplo, como P(dordedente, boticão, cárie, nublado) e P(dordedente, boticão, cárie) estão relacionadas? Podemos utilizar a regra do produto:

$$P(dordedente, boticão, cárie, nublado) \\ = P(nublado \mid dordedente, boticão, cárie)\ P(dordedente, boticão, cárie)$$

Agora, a menos que se esteja no ramo comercial de divindade, não se deve imaginar que os problemas dentários de alguém influenciam as condições do tempo. E, para a odontologia em consultório, pelo menos, parece seguro dizer que o tempo não influencia as variáveis dentárias. Portanto, a asserção a seguir parece razoável:

$$P(nublado \mid dordedente, boticão, cárie) = P(nublado)$$

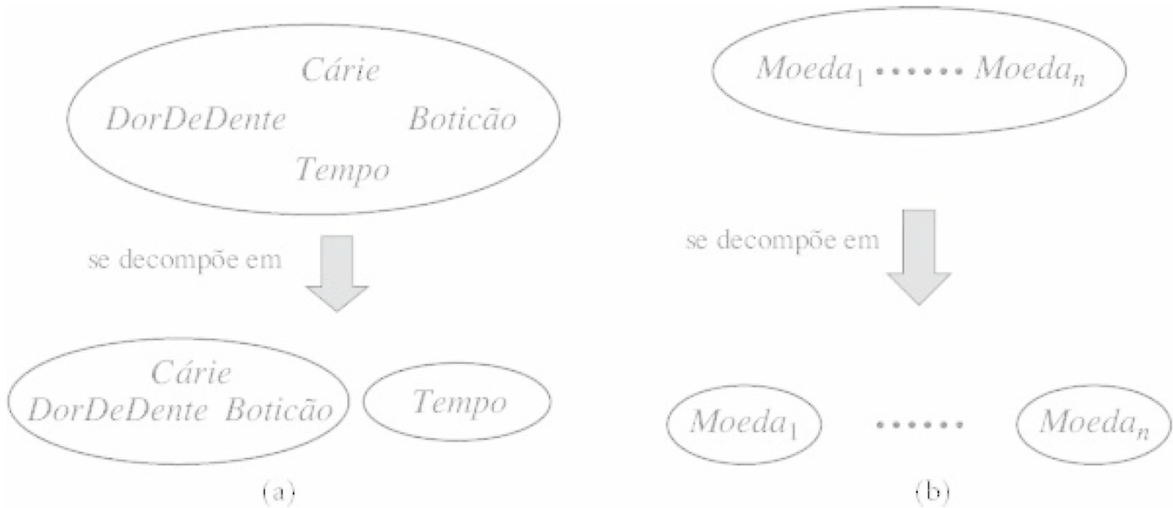
A partir disso, podemos deduzir:

$$P(dordedente, boticão, cárie, nublado) = P(nublado)\ P(dordedente, boticão, cárie)$$

Existe uma equação semelhante para toda entrada em P(DorDeDente, Boticão, Cárie, Tempo). De fato, podemos escrever a equação geral:

$$\mathbf{P}(DorDeDente, Boticão, Cárie, Tempo) = \mathbf{P}(DorDeDente, Boticão, Cárie)\mathbf{P}(Tempo)$$

Desse modo, a tabela de 32 elementos para quatro variáveis pode ser construída a partir de uma tabela de oito elementos e uma tabela de quatro elementos. Essa decomposição é ilustrada esquematicamente abaixo (item a):




Essa propriedade é chamada independência (também independência marginal e independência absoluta). Em particular, o tempo é independente dos problemas dentários de alguém. A independência entre as proposições a e b pode ser escrita como:

$$P(a \mid b) = P(a) \text{ ou } P(b \mid a) = P(b) \text{ ou } P(a \wedge b) = P(a)P(b)$$

As asserções de independência em geral se baseiam no conhecimento do domínio. Como o exemplo do tempo na dor de dente ilustra, elas podem reduzir drasticamente a quantidade de informações necessárias para especificar a distribuição conjunta total. Se o conjunto completo de variáveis puder ser dividido em subconjuntos independentes, então a distribuição conjunta total poderá ser fatorada em distribuições conjuntas separadas sobre esses subconjuntos. Por exemplo, a distribuição conjunta total sobre o resultado de n lançamentos de moedas independentes, P(C1, …, Cn), tem 2n entradas, mas pode ser representada como o produto de n distribuições de variáveis P(Ci). De modo mais prático, a independência entre a odontologia e a meteorologia é algo bom porque, do contrário, a prática da odontologia poderia exigir o conhecimento íntimo da meteorologia e vice-versa.

Portanto, quando estão disponíveis, as asserções de independência podem ajudar na redução do tamanho da representação do domínio e da complexidade do problema de inferência. Infelizmente, a separação clara e completa de conjuntos de variáveis por independência é bastante rara. Sempre que existir uma conexão, ainda que indireta entre duas variáveis, a independência deixará de ser válida. Além disso, até mesmo subconjuntos independentes podem ser bastante grandes — por exemplo, a odontologia pode envolver dezenas de doenças e centenas de sintomas, todos inter-relacionados. Para tratar de tais problemas, precisaremos de métodos mais sutis que o simples conceito de independência.

Referências:

Livro: Inteligência Artificial
Autor: Peter Norvig