



**Data Science
Academy**

www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Semântica das Redes Bayesianas



O vídeo anterior descreveu o que é uma rede, mas não o que ela significa. Há duas maneiras de compreender a semântica das redes bayesianas. A primeira é ver a rede como uma representação da distribuição de probabilidade conjunta. A segunda é visualizá-la como uma codificação de uma coleção de declarações de independência condicional. As duas visões são equivalentes, mas a primeira se mostra útil na compreensão de como construir redes, enquanto a segunda é útil no projeto de procedimentos de inferência.

Visto como um pedaço de “sintaxe”, uma rede bayesiana é um grafo acíclico orientado com alguns parâmetros numéricos ligados a cada nó. Uma maneira de definir o que significa a rede — sua semântica — é definir a maneira pela qual ela representa uma distribuição conjunta específica sobre todas as variáveis. Existem basicamente dois tipos de semântica:

- **Semântica global** (ou numérica): busca entender as redes como uma representação da distribuição de probabilidade conjunta. Indica como construir uma rede.
- **Semântica local** (ou topológica): visualizá-las como uma codificação de uma coleção de declarações de independência condicional. Indica como fazer inferências com uma rede.

A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:

$$P(X_1, \dots, X_n) = \prod_i P(X_i \mid \text{parents}(X_i)) \text{ Exemplo: } P(j \wedge m \wedge a \wedge \neg r \wedge \neg t) = P(j \mid a) P(m \mid a) P(a \mid \neg r \wedge \neg t) P(\neg r) P(\neg t) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00063$$

Na semântica local (topológica): cada nó é condicionalmente independente de seus *nãodescendentes* dados seus pais: Um nó X é condicionalmente independente de seus não descendentes (Z_{ij}) dados seus pais (U_i).

A distribuição conjunta pode ser reconstruída a partir das asserções sobre a independência condicional e das tabelas de probabilidade condicional. Deste modo a semântica numérica e topológica são equivalentes.

Como construir uma Rede Bayesiana

Em termos simples, este é o processo de construção de uma rede bayesiana:

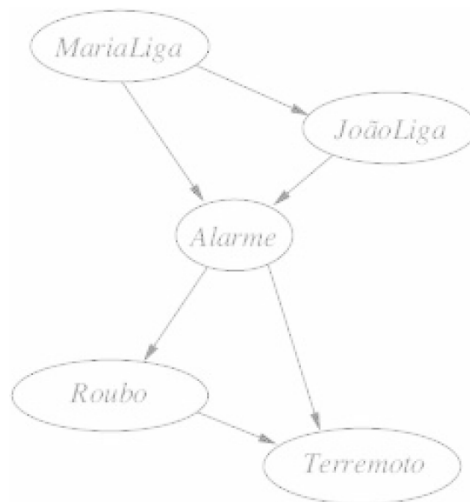
- (1) Escolhe-se o conjunto de variáveis X_i que descrevem apropriadamente o domínio.
- (2) Seleciona-se a ordem de distribuição das variáveis (Passo importante).
- (3) Enquanto ainda existirem variáveis: (a) Seleciona-se uma variável X e um nó para ela. (b) Define-se $\text{Parent}(X)$ para um conjunto mínimo de nós de forma que a independência condicional seja satisfeita. (c) Define-se a tabela de probabilidade para X .

A ordem correta em que os nós devem ser adicionados consiste em adicionar primeiro as “causas de raiz”, depois as variáveis que elas influenciam e assim por diante, até chegarmos às folhas, que não tem nenhuma influência causal direta sobre as outras variáveis. Princípio Minimalista: Quanto menor a rede, melhor ela é.

Além de ser uma representação completa e não redundante do domínio, uma rede bayesiana frequentemente pode ser muito mais compacta que a distribuição conjunta total. Essa propriedade é o que torna viável manipular domínios com muitas variáveis. A densidade das redes bayesianas é um exemplo de propriedade muito geral de sistemas localmente estruturados (também chamados sistemas esparsos). Em um sistema localmente estruturado, cada subcomponente interage diretamente apenas com um número limitado de outros componentes, não importando o número total de componentes. A estrutura local normalmente está associada a um crescimento linear, e não a um crescimento exponencial da complexidade. No caso das redes bayesianas, é razoável supor que, na maioria dos domínios, cada variável aleatória é diretamente influenciada por, no máximo, k outras, para alguma constante k . Se supusermos n variáveis booleanas por simplicidade, a quantidade de informações necessárias para especificar cada tabela de probabilidade condicional será no máximo 2^k números, e a rede completa poderá ser especificada por $n2^k$ números. Em contraste, a distribuição conjunta contém 2^n números. Para tornar esse exemplo concreto, vamos supor que tenhamos $n = 30$ nós, cada um com cinco pais ($k = 5$). Então, a rede bayesiana exigirá 960 números, mas a distribuição conjunta total exigirá mais de um bilhão.

Existem domínios em que cada variável pode ser diretamente influenciada por todas as outras, de forma que a rede seja totalmente conectada. Então, a especificação das tabelas de probabilidade condicional exige a mesma quantidade de informações que a especificação da distribuição conjunta. Em alguns domínios, existirão dependências fracas que deverão ser incluídas estritamente pela adição de novos vínculos. Porém, se essas dependências forem muito tênues, talvez não compense a complexidade adicional na rede em relação ao pequeno ganho em exatidão. Por exemplo, alguém poderia fazer uma objeção à nossa rede de alarme contra roubo afirmando que, se houvesse um terremoto, João e Maria não telefonariam mesmo que tivessem ouvido o alarme porque eles iriam supor que o terremoto fosse a causa. A decisão de adicionar o vínculo de Terremoto para JoãoLiga e para MariaLiga (e, desse modo, de ampliar as tabelas) dependerá da comparação entre a importância de obter probabilidades mais precisas e o custo de especificar as informações extras. Mesmo em um domínio

localmente estruturado, só obteremos uma rede bayesiana se ordenarmos bem para escolher o nó. O que acontecerá se escolhermos a ordem errada? Vamos considerar novamente o exemplo do alarme contra roubo. Suponha que decidimos adicionar os nós na ordem MariaLiga, JoãoLiga, Alarme, Roubo, Terremoto. Nesse caso, obtemos a rede um pouco mais complicada mostrada na figura abaixo. O processo se desenvolve assim:



A rede resultante terá dois vínculos a mais que a rede original e exigirá outras probabilidades para serem especificadas.

Alguns dos vínculos apresentam relacionamentos tênues que exigem julgamentos de probabilidade difíceis e antinaturais (probabilidade de Terremoto, dados Roubo e Alarme).

Em geral, é melhor pensar de causas para efeitos (modelo causal) e não do contrário (modelo de diagnóstico). Se nos fixarmos em um modelo causal, acabaremos tendo de especificar uma quantidade menor de números, e os números frequentemente serão mais fáceis de apresentar.

Você pode usar o Microsoft Bayesian Network Editor para construir redes bayesianas:

<https://msbnx.azurewebsites.net/>