Inteligência Artificial

Prof. Luiz Antonio Ferraro Mathias

Aula 06

Redes Bayesianas

Conceitos básicos de Probabilidade

Para elucidar as Redes Bayesianas, primeiramente é necessária a compreensão da probabilidade, visto que a mesma é fundamentada nesses conceitos. Existem situações em que não é possível determinar o resultado de um evento, pois este apresenta variabilidade em sua ocorrência. Logo, para lidar com esses casos, é preciso atribuir uma medida que lide com a incerteza, a probabilidade. Segundo Faria R. (2014), há diversos conceitos possíveis para probabilidade como: a frequentista, clássica e bayesiana.

A teoria frequentista define que a probabilidade de ocorrência de um evento é a sua frequência relativa, ou seja, a quantidade de vezes que um evento aconteceu, repetido por diversas vezes até formar uma amostra, colocado sob condições similares. Com isso, é possível perceber que o resultado se aproxima de um número finito de repetições. Tomando como exemplo, um dado que é lançado 50 vezes, e observando que em 28 dos lançamentos resultaram no número 6, pode-se dizer que para esse dado, o número 6 ocorreu 0.56. No entanto, essa teoria possui alguns problemas como: a dificuldade de saber quantas vezes um experimento deve ser repetido para conseguir uma amostra significativa; definição de condições similares, pois isso poderia tirar o nível de incerteza do processo e em situações que não podem ser replicadas, não é possível obter a frequência relativa.

Conceitos básicos de Probabilidade

Já a teoria clássica estabelece que todos os possíveis resultados têm a mesma chance de ocorrência, ou seja, equiprováveis. Para isso, o cálculo é feito da seguinte forma: o número de elementos do evento sobre o número de resultados possíveis. Dando como exemplo, uma moeda e desejando-se saber qual a probabilidade de resultar em cara, sabendo que o número de caras é igual a 1 e o número de total de resultados é igual a 2 (cara ou coroa), logo cara e coroa tem 0.5 de chance de ocorrer. Essa teoria também possui uma restrição, pois não existe método para casos que não sejam equiprováveis.

Por fim, a teoria Bayesiana ou Subjetivista, inclusive o foco deste trabalho, é baseada na crença que uma pessoa tem sobre a ocorrência de um evento, o nível de crença é medido no intervalo de [0,1], no qual 0 representa a incapacidade de ocorrência e 1 o indicativo de que certamente a crença ocorrerá. Nessa teoria, não existe uma probabilidade certa, mas sim a probabilidade adquirida pelo conhecimento de um indivíduo sobre uma ocorrência. Para ilustrar esta teoria, suponha que o conhecimento de uma pessoa A sobre a chance de ocorrência de um dado seja o número 6 = 0.5 e, que outro indivíduo B que não viu o lançamento anterior, e a partir de seus conhecimentos diga que o número 6 tem a chance de ocorrência de 1. De acordo com Faria R. (2014), a teoria Bayesiana é criticada por diversos autores por alegarem que uma pessoa não está livre de preconceitos ao atribuir uma probabilidade e pelo fato de apresentar dificuldades quando conduzidas por mais de um pesquisador. No entanto são capazes de lidar com a incerteza e, também, podem atuar em conjunto com a frequência relativa e clássica.

Conceitos básicos de Probabilidade

Em resumo pode-se dizer que segundo Faria A. (2009), a Probabilidade é um campo que estuda a ocorrência de experimentos aleatórios. Experimentos aleatórios são acontecimentos que não se pode prever com certeza.

Para melhorar a compreensão dos assuntos que serão abordados, alguns conceitos de probabilidade são definidos a seguir:

a) Espaço Amostral é o conjunto de todos os resultados possíveis de um experimento aleatório. Sendo definido pela letra Ω ou S.

b) Evento é um subconjunto do espaço amostral. Sendo os eventos representados por letras maiúsculas e os elementos por minúsculas.

c) Intersecção entre eventos são os elementos que estes têm em comum.

d) Eventos disjuntos não possuem nenhum elemento em comum.

e) Eventos são mutuamente exclusivos, se não podem ocorrer simultaneamente.

Axiomas da Probabilidade

Axiomas são premissas que se admite como plenamente verdadeiras sem a necessidade de constatação. São as propriedades que dão a base para a probabilidade, portanto uma probabilidade P que associa a um evento A do Ω em um número real P(A), deve satisfazer os três axiomas:

I. Axioma: $P(A) \geq 0$: Todas as probabilidades devem ser positivas, ou seja, iguais ou maiores que 0.

II. Axioma: $P(\Omega) = 1$: A probabilidade do espaço amostral será sempre 1, isto é, a soma de todos os resultados possíveis, sempre deve resultar em 1.

III. Axioma: $A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$: Se A intersecção com B resultar no conjunto vazio, ou seja, nenhum elemento em comum, isto quer dizer que a probabilidade de A união com B é igual a probabilidade de A mais a probabilidade de B.

Probabilidade Condicional

Axiomas são premissas que se admite como plenamente verdadeiras sem a necessidade de constatação. São as propriedades que dão a base para a probabilidade, portanto uma probabilidade P que associa a um evento A do Ω em um número real P(A), deve satisfazer os três axiomas:

A probabilidade condicional define que um evento depende de outro para ocorrer. É representada por $P(B|A)$, que pode ser lida como: " a probabilidade do evento B acontecer dada a ocorrência do evento A". Logo B será calculado em função do espaço amostral reduzido de A.

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Probabilidade Condicional

A fórmula anterior é lida da seguinte maneira: a probabilidade dos eventos em comum de A e B em relação ao espaço amostral, sobre o espaço de A. Uma importante consequência da probabilidade condicional é o teorema da multiplicação (HAZZAN; IEZZI, 2004 apud GONÇALVES, 2008). Em situações, em que a probabilidade condicional é fácil de se calcular, é possível resolver a ocorrência simultânea de eventos pelo teorema da multiplicação. O teorema define que a probabilidade de dois eventos ocorrerem simultaneamente é a probabilidade do primeiro multiplicado pela probabilidade do segundo em relação ao primeiro. É representado pela fórmula a seguir:

$$P(A \cap B) = P(B|A) \,.\, P(A)$$

Considere *n* eventos $B_1 \ldots B_n \in \Omega$ e que são disjuntos, mas cuja união resultam no espaço amostral. Pode-se dizer que esses eventos formam uma partição.

Para calcular eventos diferentes que possuem partições, utilizamos o teorema da probabilidade total. Então para todo evento A ∈ Ω:

$$P(A) = \sum_{j=1}^{n} P(B_j) \,.\, P(A \mid B_j)$$

Independência

Um evento A independe de B, se a ocorrência de B não afeta a probabilidade de A.

$$P(A \cap B) = P(A) \,.\, P(B)$$

Um evento X é condicionalmente independente de Y dado Z, se a distribuição que rege Z é independente de Y dado o valor de Z, que pode ser representado por:

$$P(X \mid Y \cap Z) = P(X|Z)$$

Variáveis Aleatórias

Conforme Silva (2016, p.13), uma variável aleatória (v.a) pode ser considerada um experimento e cada possível saída para um experimento é chamada de estado. Variáveis aleatórias normalmente são representadas por letras maiúsculas (A, B) e os elementos ou estados associados a ela são denotados por letras minúsculas (a, b). Além disso, Norvig e Russel (2004 apud Valentim, 2007) ainda as classificam em três categorias: variáveis aleatórias booleanas, variáveis aleatórias discretas e variáveis aleatórias contínuas.

As variáveis aleatórias booleanas só contêm dois valores, como por exemplo: a variável *TemGaragem* pode possuir os estados {*verdadeiro* ou *falso}*. Já as variáveis aleatórias discretas, ao contrário das booleanas, permitem múltiplos valores finitos, por exemplo, a variável *EstadoCivil* pode conter os valores {*Solteiro, Casado, Divorciado, Viúvo*}. Em relação, as variáveis aleatórias contínuas, estas podem classificar seus valores em um intervalo de [0,1] ou na linha real inteira.

Distribuição da Probabilidade Conjunta

Segundo Silva (2016), a distribuição da probabilidade conjunta é a probabilidade de ocorrência de duas ou mais variáveis aleatórias $x_{1,}\, x_{2\,,\ldots}\, x_n$ ocorrerem juntas. Uma função de probabilidade conjunta necessariamente precisa satisfazer a seguinte propriedade, de acordo com Valentim (2016)

$$\sum P(X_{1,}\ X_{1,}\ \ldots\ X_n\ ) = 1$$

Na equação 1.6 diz que os somatórios de todos os valores das variáveis devem resultar em 1. Russel (2004 apud Valentim, 2007) ainda as classificam em três categorias: variáveis aleatórias booleanas, variáveis aleatórias discretas e variáveis aleatórias contínuas.

Teoria de Bayes

O Teorema de Bayes é uma extensão do teorema da probabilidade total e da probabilidade condicional. Enquanto o segundo permite calcular um resultado dentre múltiplos eventos de cada partição a priori, ou seja, antes do evento acontecer, o Bayes permite esses mesmos cálculos a partir de uma condição a posteriori, ou seja, depois que o evento a priori aconteceu. A definição formal do Teorema de Bayes definida por (LIPSCHUTZ, 1993 apud GONÇALVES, 2008):

Teorema: Suponha A_1, A_2, A_3... A_n ser uma partição de S e B, um evento qualquer. Então para qualquer i

$$P(A_i|B) = \frac{P(B\ |A_i)\ .\ P(A_i)}{P(B)} = \frac{P(B|A_i)\ .P(A_i)}{\sum i\ \ P(B\ |\ A_i)\ \ .\ P(A_i)}$$

Teoria de Bayes

Uma interpretação do Teorema de Bayes, consiste em considerar os eventos A_i como “causas” do evento B, sendo atribuído probabilidades deste evento atuar na ocorrência de B. Esta probabilidade é calculada antes da realização do experimento, sendo designada como a probabilidade *a priori* de A_i. Após a realização do experimento, é conhecido que o evento B ocorreu, então a probabilidade *a priori* é revista por meio da fórmula de Bayes e então passa a atribuir aos eventos A_i, *i* = 1,2...n as probabilidades *a posteriori* $P(A|B_i)$, *i* = 1,2...n (CRAMÉR, 1955 apud GONÇALVES, 2008)(PAULINO; TURKMAN; MURTEIRA, 2003 apud GONÇALVES, 2008).

Teoria dos Grafos

A teoria dos grafos é utilizada no estudo de Redes Bayesianas para representar visualmente as relações de causa e consequência de variáveis e probabilidades. Conforme Faria, R. (2014), um grafo é um conteúdo matemático, que tem como principal capacidade a representação de conhecimento de maneira visual, e assim é amplamente utilizado em diversas aplicações.

Um grafo pode ser definido como um conjunto de elementos que possuem conexões entre si. Os elementos são denominados vértices, nós ou variáveis e suas conexões são chamadas de arestas ou arcos, tal estrutura pode ser compreendida na Figura 17. A sua representação matemática é definida como G = (V, A), onde V são os vértices que podem ser representados por letras ou nomes dentro de um círculo, aresta ou retângulo e as arestas que são representadas por setas. A junção de nós e arcos apresentam a relação de dependência probabilística.

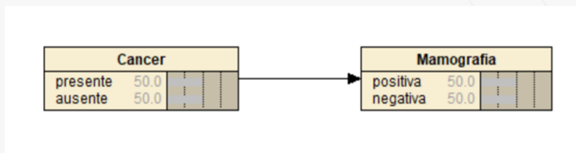
Conceito de Redes Bayesianas

As Redes Bayesianas (RB's) são uma representação qualitativa (teoria dos grafos) e quantitativa (probabilidade) de cenários que envolvam o processo decisório, prognóstico ou analítico, tendo em vista condições de incerteza, complexidade e falta de informações. De acordo com Charniak (1991), uma maneira de entender a aplicação de Redes Bayesianas é quando se tem conhecimento prévio da causa de um evento, mas ainda é insuficiente para a compreensão total do processo, e dessa forma é preciso modelar o problema probabilisticamente. Para melhor compreensão das Redes Bayesiana considere o exemplo de Yudkowsky (2003, apud Valentim, 2007, p.13).

1% das mulheres com mais de 40 anos que participam de exames de rotina são portadoras de câncer de mama. 80% das mulheres com câncer terão resultados positivos de mamografias. 9,6% das mulheres sem a doença também terão resultado positivo nas mamografias. Uma mulher dessa idade se depara com um resultado positivo de mamografia, qual a probabilidade dela portar câncer de mama?

Conceito de Redes Bayesianas

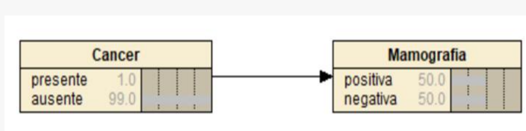
Para a construção de uma Rede Bayesiana com o exemplo dado acima, primeiramente é preciso identificar as variáveis de interesse. Como a situação foca na influência que o câncer tem sobre o resultado da mamografia, é notável que as variáveis são: *Câncer* e *Mamografia*. O próximo passo é definir as relações de causa e consequência do domínio em questão. A causa é o câncer e a consequência é a mamografia, uma vez que o fato de possuir ou não a doença é o fator que determina o resultado da mamografia, exceto para os casos em que há erros. Depois de identificado as variáveis de causa e consequência é necessário representá-las na rede como mostra abaixo:



Conceito de Redes Bayesianas

Nota-se que a estrutura de uma Rede Bayesiana é a mesma de um grafo acíclico dirigido (DAG), já que os arcos possuem direção e o nó inicial (*Câncer*) é diferente do nó final (*Mamografia*). Em relação a causalidade, esta é representada por um nó *Câncer* com um arco direcionado para o nó *Mamografia*, também sendo equivalente a dizer que *Câncer* é “pai” de *Mamografia*. Também há os estados dos nós *Câncer* e *Mamografia*, sendo respectivamente {presente, ausente} e {positiva e negativa}, no qual indica a probabilidade de ocorrência.

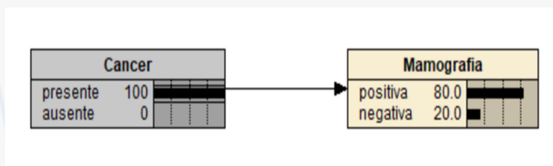
Quando a estrutura da rede é construída, a próxima etapa é a definição das probabilidades *a priori* e as probabilidades condicionais para as variáveis. As probabilidades das variáveis aleatórias *a priori* contêm os valores antes de qualquer ocorrência de outra variável, enquanto as condicionais apresentam a probabilidade de uma variável acontecer dado que outra ocorreu. A figura abaixo mostra as probabilidades *a priori* das variáveis *Câncer.*



Conceito de Redes Bayesianas

Na anterior pode-se interpretar assim como o exemplo, que a probabilidade *a priori* de uma mulher com mais de quarenta anos ter câncer é 1.0 e consequentemente, a probabilidade dela não ter é 99.0. A probabilidade *a priori* dos estados de mamografia é considerada 50.0 para resultado positivo da doença, tanto para resultado negativo, já que o exemplo não definiu uma probabilidade específica. É importante lembrar que, as probabilidades *a priori* e as condicionais, normalmente são definidas por especialistas na área, como o caso do domínio em questão que foram definidas por médicos, mas também podem ser definidas pela frequência relativa.

Assim que é finalizado a parametrização das probabilidades *a priori*, é necessário definir as probabilidades condicionais. Nesse caso, os estados da *Mamografia* {positiva, negativa}, dado que o estado de *Câncer* está presente ou ausente. A figura abaixo mostra o estado de *Câncer* estando presente e a Figura 20 apresenta a ausência da ocorrência de *Câncer*.



Conceito de Redes Bayesianas

Na figura anterior, pode-se notar que o nó *Câncer* contendo o estado presente influencia o estado do nó da *Mamografia* resultar em positivo em 80.0 dos casos e negativo em 20.0, coincidindo com as informações dadas no exemplo.

Já na figura abaixo, é perceptível que o estado da variável *Câncer* estar ausente influencia no estado da mamografia dar um resultado positivo de 9.60 e negativo de 90.4, equivalentes ao exemplo dado acima.

Conceito de Redes Bayesianas

De forma geral, uma Rede Bayesiana é uma combinação da Teoria dos Grafos, ou seja, a representação visual da Figura 1 e dos conceitos de Probabilidade e Estatística. Além disso, as redes bayesianas também podem ser chamadas de Rede de Crença, Rede Probabilística ou Causal. O objetivo da rede é apresentar as relações de causa e efeito das variáveis, juntamente com a disposição da probabilidade conjunta e, a partir disso possibilitar o procedimento de inferência. A rede também é capaz de por meio das relações de independência condicional, probabilidade condicional e independências incondicionais aliado com a representação visual das variáveis providas pelos grafos, compreender como um cenário atua e até mesmo, prever as ações de uma intervenção, a partir do cálculo de probabilidades *a priori,* ou seja, antes do evento acontecer e *a posteriori,* sendo depois do evento acontecer.

Construção de uma rede Bayesiana

Em primeiro lugar, deve-se selecionar as variáveis que são importantes no processo. Para isso, deve-se seguir alguns critérios Kjærulff e Madsen (2013 apud FARIA R, 2014): as variáveis devem conter valores mutuamente exclusivos, ou seja, que não podem ocorrer ao mesmo tempo; devem representar um conjunto único de eventos, sem a ocorrência de redundâncias; e devem ser muito bem definidas. Com as variáveis do modelo devidamente selecionadas, agora o problema é construir a estrutura da Rede Bayesiana, que é identificar entre quais variáveis terão arcos e qual direcionamento deve ser dado a cada um deles.

A construção pode ser feita através das relações de causa efeito identificadas pelo especialista da área a ser estudada, porém, caso não esteja disponível um especialista para auxiliar na construção da rede ou ele precise de um guia para ajudá-lo, é útil manter em mente os tipos de variáveis que podem surgir:

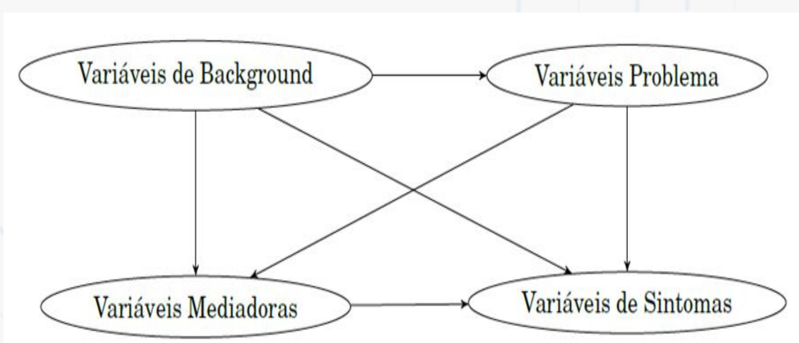
- **Variáveis problema**: São variáveis de interesse, das quais deseja-se calcular distribuições a posteriori dado os valores das variáveis de informação.
- **Variáveis de background**: São informações disponíveis antes da ocorrência do problema, com influência causal sobre as variáveis problema e variáveis de sintomas. São geralmente as variáveis ‘raízes’ das redes bayesianas.

Construção de uma rede Bayesiana

- **Variáveis de sintomas**: São as consequências do problema, sendo disponíveis somente após sua ocorrência, ou seja, variáveis de sintomas são causadas pelas variáveis problema. Geralmente, as variáveis problema e de background são pais das variáveis de sintomas.
- **Variáveis mediadoras:** São variáveis não observáveis cujas probabilidades *a posteriori* não são de interesse imediato, mas que ajudam a manter as relações de independência corretas na rede. Estas variáveis geralmente são filhas de variáveis problema e de background e pais de variáveis de sintomas.

Os tipos de variáveis acima, propostos por Kjærulff e Madsen (2013 apud ARA-SOUZA, p.41), são uma forma de auxiliar a construção de Redes Bayesianas feitas manualmente, pois, após a identificação do tipo de uma variável, é muito mais fácil conectá-la às outras, já que geralmente o fluxo causal apresentado na figura a seguir é satisfeito.

Construção de uma rede Bayesiana



Um exemplo prático – seguro de automóveis

O seguro de automóveis tem como função a responsabilização pelos riscos e também pela indenização do veículo para o cliente no caso de colisão; acidentes; roubos e entre outros, em troca de um pagamento, denominado prêmio. Para que isso seja possível, a seguradora precisa estar regulamentada pela autarquia da Superintendência de Seguros Privados (SUSEP), que é responsável pelo controle e fiscalização do mercado de seguros no Brasil.

A identificação do grupo com riscos semelhantes é realizada por um questionário de avaliação de risco do segurado. Esta avaliação de risco é feita por intermédio do corretor e se resumem basicamente em perguntas referentes a idade; tempo de habilitação; sexo; região de circulação do veículo; obtenção de garagem ou estacionamento fechado para o automóvel; utilização do veículo e dispositivos de segurança. A partir disso e de dados estatísticos, a seguradora é capaz de identificar o perfil do segurado e os possíveis riscos que irá assumir, como por exemplo, um grupo com idade entre dezoito e vinte cinco anos, do sexo masculino que apresenta muitos sinistros, ou seja, qualquer dano ao veículo, terá o valor do prêmio mais alto para todos com essas características.

Um exemplo prático – seguro de automóveis

De modo geral, existem diversos fatores que contribuem para a precificação do prêmio de seguro de automóveis: cobertura por colisão, idade do veículo, franquia, bônus e grau de periculosidade da cidade etc. Segue abaixo modelos de redes bayesianas aplicadas ao “case” citado.