

3 Modelos de Markov Ocultos

3.1. Processos Estocásticos

Um processo estocástico é definido como uma família de variáveis aleatórias $X(t)$, sendo t geralmente a variável tempo. $X(t)$ representa uma característica mensurável de interesse no tempo t . Exemplificando, $X(t)$ pode representar o nível de estoque de um produto no fim do dia t .

Processos estocásticos são usados para descrever o comportamento ao longo de um período de tempo de um sistema onde a incerteza é significativa. Em termos mais formais, a variável aleatória $X(t)$ representa o estado do sistema no instante t .

Os processos estocásticos podem ser classificados:

Em relação ao estado:

- Estado discreto: $X(t)$ é definido sobre um conjunto enumerável ou finito,
- Estado contínuo: caso contrário.

Em relação ao tempo:

- Tempo discreto: t é finito ou enumerável,
- Tempo contínuo: caso contrário.

3.2. Processos Markovianos

Um processo estocástico é dito markoviano se a probabilidade do processo estar em um dado estado, em um instante qualquer no futuro, condicionada à seqüência completa de estados até o presente permanece a mesma quando condicionada apenas ao estado presente.

Usando a notação q_t para designar a variável aleatória $X(t)$, as probabilidades condicionais $P(q_{t+1} = S_j | q_t = S_i)$ são denominadas probabilidades

de transição e representam, a probabilidade do estado q_{t+1} ser S_j no instante $t+1$ dado que o estado q_t é S_i no instante t .

3.3. Cadeias de Markov

Um processo markoviano é uma cadeia de Markov quando as variáveis aleatórias $X(t)$ estão definidas em um espaço de estados discreto.

Consideram-se sistemas que podem ser descritos como estando em qualquer instante de tempo em um estado de um conjunto de N estados distintos (S_1, S_2, \dots, S_N) como ilustrado na figura 1 (onde $N=3$ para simplificar a explicação).

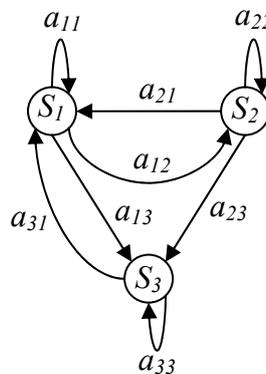


Figura 1 - Cadeia de Markov com 3 estados (S_1 a S_3), onde as setas indicam as transições possíveis entre os estados.

Para instantes de tempo igualmente espaçados, o sistema passa por uma mudança de estado (podendo retornar ao mesmo estado) de acordo com um conjunto de probabilidades de transição. Cada instante de tempo, onde ocorrem as mudanças de estado, é descrito por $t = 1, 2, \dots$, e o estado no tempo t é denotado com o símbolo q_t , considerando um espaço de tempo discreto.

Uma descrição probabilística completa do sistema requer a especificação do estado presente (no tempo t) e também dos estados anteriores. Para o caso especial de uma cadeia de Markov de primeira ordem, essa descrição pode ser truncada apenas no estado presente e no seu antecessor:

$$P(q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots) = P(q_t = S_j | q_{t-1} = S_i) \quad \text{Equação 1}$$

Considerando processos em que o lado direito da equação 1 é independente do tempo, tem-se um conjunto de probabilidades de transição de estado a_{ij} :

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i), \quad 1 \leq i, j \leq N. \quad \text{Equação 2}$$

As probabilidades de transição de estado têm as seguintes propriedades:

$$a_{ij} \geq 0 \quad \text{Equação 3a}$$

$$\sum_{i=1}^N a_{ij} = 1 \quad \text{Equação 3b}$$

Quando as probabilidades de transição não se alteram ao longo do tempo, diz-se que o processo é estacionário.

O processo estocástico mencionado acima poderia ser chamado de modelo de Markov observável uma vez que a saída do processo é um conjunto de estados em cada instante de tempo, onde cada estado corresponde a um evento físico observável.

3.4. Modelos de Markov Ocultos

Nos modelos de Markov descritos anteriormente, cada estado corresponde a um evento físico observável. Essa característica torna esses modelos muito restritos para solução de diversos problemas de interesse. Para estes casos aplica-se uma extensão chamada de modelos de Markov ocultos (do inglês: *Hidden Markov Models* - HMM), onde o estado não é diretamente observável, mas uma observação é uma função probabilística do estado.

Um HMM é um processo duplamente estocástico composto por um processo latente que não é observável (*hidden*), mas que se manifesta através de um outro processo estocástico o qual produz uma sequência de símbolos observáveis (Bunke & Caelli, 2001).

3.4.1. Elementos de um Modelo de Markov Oculto

Em um HMM, as observações (v_i) são símbolos emitidos por estados não observáveis (S_i) de acordo com determinadas funções probabilísticas, sendo que cada seqüência de estados é uma cadeia de Markov de primeira ordem. A figura 2 ilustra um HMM, sendo N o número de estados do modelo (o conjunto dos estados individuais é representado por $\mathbf{S} = \{S_1, \dots, S_N\}$ e o estado no tempo t pelo símbolo q_t) e M o número de símbolos distintos observáveis (o conjunto dos símbolos individuais é representado por $\mathbf{V} = \{v_1, \dots, v_M\}$).

Um HMM básico é constituído de 3 conjuntos de parâmetros principais:

1. as probabilidades de emissão de símbolo $B = \{b_{jk}\}$ – a probabilidade do símbolo v_k ser emitido pelo estado S_j , i.e.

$$b_{jk} = P(v_k \text{ em } t | q_t = S_j), \quad 1 \leq j \leq N \text{ e } 1 \leq k \leq M \quad \text{Equação 4}$$

2. as probabilidades de transição de estado $A = \{a_{ij}\}$ – a probabilidade de estar no estado S_j no instante de tempo subsequente dado que o estado atual é S_i , i.e.

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \quad 1 \leq i, j \leq N \quad \text{Equação 5}$$

3. A distribuição de probabilidade a priori $\boldsymbol{\pi} = \{\pi_i\}$ do sistema estar em um dado estado S_i no instante inicial de tempo (não mostrado na Figura 2), i.e.

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N \quad \text{Equação 6}$$

Com respeito à figura 2, se o sistema pode passar de um estado S_i para outro S_j , $a_{ij} > 0$ e se esses dois estados não são conectados, $a_{ij} = 0$. Apesar dos estados serem escondidos, em muitas aplicações práticas, estes possuem significado físico relevante para o problema em questão. Este é o caso da aplicação alvo do presente trabalho, em que os estados estão associados aos estágios fenológicos de culturas agrícolas.

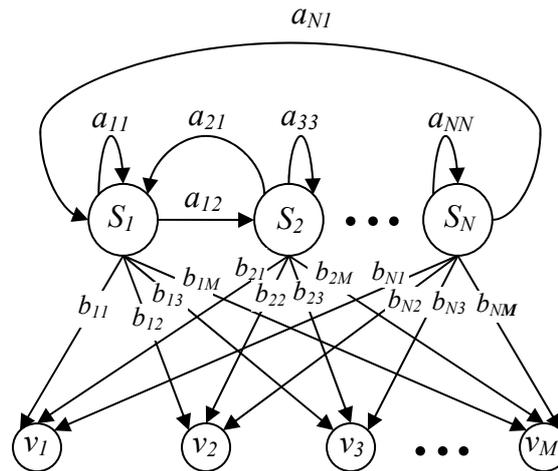


Figura 2 - Exemplo de um modelo de Markov Oculto ($S_i \rightarrow$ estados, $v_k \rightarrow$ símbolos observáveis, $a_{ij} \rightarrow$ probabilidade de transição de estado, $b_{jk} \rightarrow$ probabilidade de emissão de símbolo).

3.4.2. Problemas básicos associados a HMMs

A maioria das aplicações de HMMs envolve a solução de até três problemas principais:

- **Problema 1:** Considere uma seqüência de observações $O = \{O_1, \dots, O_M\}$. Dado o modelo $\lambda = (A, B, \pi)$, calcular $P(O | \lambda)$ (a probabilidade de que o modelo λ gere a seqüência de observações O).
- **Problema 2:** Para uma seqüência de observações $O = \{O_1, \dots, O_M\}$, dado o modelo $\lambda = (A, B, \pi)$, determinar a seqüência dos estados $Q = \{q_1, \dots, q_N\}$, de modo que $P(S, O | \lambda)$ (distribuição de probabilidade conjunta das seqüências de observações, O , e de estados S , dado o modelo λ) seja máxima.
- **Problema 3:** Estimar os parâmetros do modelo $\lambda = (A, B, \pi)$ de modo que $P(O | \lambda)$ ou $P(S, O | \lambda)$ seja maximizada.

O problema de classificação de culturas agrícolas, que é o tema do presente trabalho, corresponde ao problema 1. O problema 2 consiste em revelar a parte oculta do sistema, na aplicação alvo, determinar o estágio fenológico em que se encontra cada cultura em cada data. Ambos os problemas são, portanto relevantes neste trabalho.

O problema 3, em que se estimam os valores dos parâmetros do modelo, pressupõe a inexistência de dados de treinamento rotulados. Para a aplicação alvo deste trabalho está disponível um conjunto de dados em que o modelo (cultura), os estados (estágios fenológicos) e os símbolos emitidos (respostas espectrais) são conhecidos. A partir desta informação, podem-se estimar diretamente os parâmetros do modelo, como será descrito na Seção 4.2. Por isso o texto que se segue não apresenta a solução do problema 3. Detalhes sobre métodos de solução do problema 3 podem ser encontrados, por exemplo, em (Rabiner, 1989).

3.4.3.

Soluções para os problemas 1 e 2 associados a HMMs

3.4.3.1.

Solução do problema 1

Deseja-se calcular $P(O|\lambda)$, ou seja, a probabilidade de que o modelo dado λ produza a seqüência de observações, $O = \{O_1 O_2 \dots O_T\}$, sendo T o número de observações na seqüência. Esta probabilidade pode ser calculada através de um procedimento que consiste de duas etapas, denominadas direta e reversa.

A) A etapa direta:

Seja α_t a probabilidade (chamada usualmente na literatura de direta) da seqüência parcial de observações, $O = \{O_1 O_2 \dots O_t\}$, até o tempo t e de que o sistema esteja no estado S_i no tempo t , dado o modelo λ , ou seja,

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda) \quad \text{Equação 7}$$

O valor de $\alpha_t(i)$ pode ser calculado indutivamente, pelos seguintes três passos:

1. Inicialização:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad \text{Equação 8}$$

2. Indução:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq i \leq T-1 \text{ e } 1 \leq j \leq N \quad \text{Equação 9}$$

3. Terminação:

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i). \quad \text{Equação 10}$$

No primeiro passo a probabilidade direta é iniciada como a probabilidade conjunta do estado S_i e da observação inicial O_1 . O passo de indução, a etapa mais importante do cálculo, é ilustrada pela figura 3. Esta figura mostra que o estado S_j pode ser alcançado no instante $t+1$ partindo-se de qualquer um dos N estados possíveis, S_i , $1 \leq i \leq N$, no tempo t . Como $\alpha_t(i)$ é a probabilidade de $O_1 O_2 \dots O_t$ ser observada e o estado no tempo t ser S_i , o produto $\alpha_t(i) a_{ij}$ é então a probabilidade de $O_1 O_2 \dots O_t$ ser observada e o estado S_j ser alcançado no tempo $t+1$ via o estado S_i no tempo t . Somando este produto para todos os N estados possíveis S_i , $1 \leq i \leq N$, no tempo t obtém-se a probabilidade de estar em S_j no tempo $t+1$ com todas as observações parciais anteriores.

Tendo sido calculado este valor e conhecendo-se o estado S_j , é fácil notar que $\alpha_{t+1}(j)$ é obtido agregando-se a informação sobre a observação O_{t+1} à informação sobre o estado S_j , ou seja, multiplicando-se o valor do somatório pela probabilidade $b_j(O_{t+1})$. O cálculo da equação 9 é feito para todos os estados S_j , $1 \leq j \leq N$, e para $t = 1, 2, \dots, T-1$.

O cálculo de $P(O | \lambda)$ é então a soma das variáveis *diretas* finais $\alpha_T(i)$ já que $\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i | \lambda)$. O cálculo da probabilidade *diretas* é ilustrado pela estrutura mostrada na figura 4.

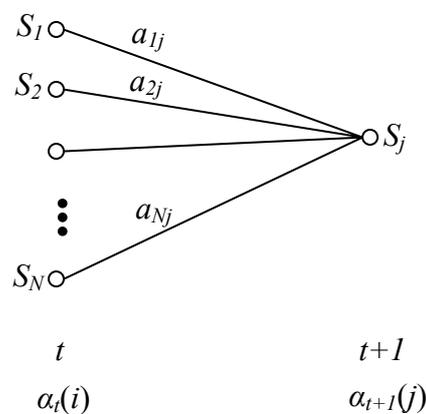


Figura 3 - Ilustração das seqüências de operações necessárias para o cálculo da variável direta $\alpha_{t+1}(j)$.

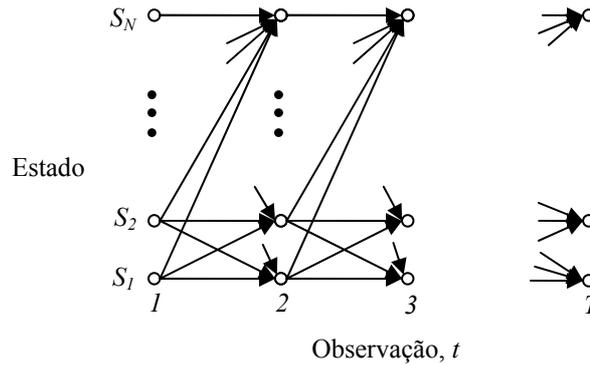


Figura 4 - Procedimento de cálculo de $\alpha_{t+1}(j)$ em função dos instantes das observações t e dos estados S_i

B) A etapa reversa:

De maneira similar ao procedimento descrito anteriormente, seja a variável $\beta_t(i)$ (chamada usualmente na literatura de reversa) definida como

$$\beta_t(i) = P(O_{t+1}O_{t+2}\dots O_T \mid q_t = S_i, \lambda), \quad \text{Equação 11}$$

i.e., $\beta_t(i)$ é a probabilidade da seqüência parcial de observações de $t+1$ até o fim, $\{O_{t+1}O_{t+2}\dots O_T\}$ dados o estado S_i no tempo t e o modelo λ . O valor de $\beta_t(i)$ pode ser calculado indutivamente, segundo os seguintes 3 passos:

1. Inicialização:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad \text{Equação 12}$$

2. Indução:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \text{ e } 1 \leq i \leq N \quad \text{Equação 13}$$

A etapa de inicialização define $\beta_T(i)$ arbitrariamente como 1 para todos os valores de i . A etapa de indução, ilustrada na figura 5, mostra que, para estar no estado S_i no tempo t tendo em vista a seqüência de observações do tempo $t+1$ em diante, é preciso considerar a possibilidade de que o sistema modelado estará em qualquer um dos possíveis estados S_j no tempo $t+1$. Isto é feito levando em conta a transição de S_i para S_j (o termo a_{ij}) para $1 \leq j \leq N$, assim como a observação O_{t+1} no estado S_j – o termo $b_j(O_{t+1})$ – e a seqüência parcial de observações restantes a

partir do estado S_j (o termo $\beta_{t+1}(j)$). $\beta_t(i)$ é calculado para $1 \leq t \leq T$ e $1 \leq i \leq N$. O cálculo da probabilidade *reversa* é ilustrado por uma estrutura similar à mostrada na figura 4.

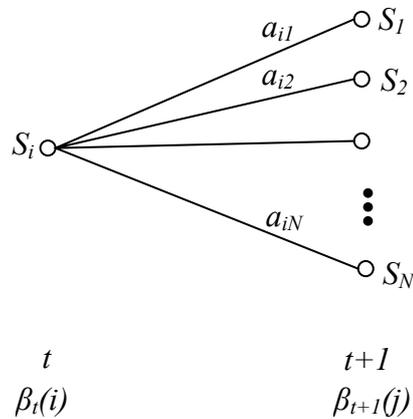


Figura 5 - Ilustração das seqüências de operações necessárias para o cálculo da variável reversa $\beta_{t+1}(j)$.

Neste trabalho foi escolhido o procedimento direto apesar da resposta dos dois procedimentos ser semelhante.

3.4.3.2. Solução do problema 2

Diferentemente do problema 1 para o qual uma solução exata pode ser obtida, existem diversas maneiras possíveis de resolver o problema 2. Uma técnica formal para achar a melhor seqüência de estados existente é o algoritmo de Viterbi.

Algoritmo de Viterbi:

Para achar a seqüência ótima de estados, $Q = \{q_1 q_2 \dots q_T\}$, tendo em vista uma dada seqüência de observações $O = \{O_1 O_2 \dots O_T\}$, define-se a variável $\delta_t(i)$:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1 q_2 \dots q_t = S_i, O_1 O_2 \dots O_t | \lambda) \quad \text{Equação 14}$$

i.e., $\delta_t(i)$ é a probabilidade do caminho (seqüência de estados) mais provável que leva ao estado S_i no tempo t , considerando as primeiras t observações. Para indução tem-se

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] b_j(O_{t+1}) \quad \text{Equação 15}$$

Para recuperar a seqüência de estados é necessário armazenar a informação sobre o argumento que maximizou a equação 15, para cada t e j . Isto é feito através da variável $\psi_t(j)$. O procedimento completo de obtenção da seqüência de estados pode ser resumido da seguinte forma:

1. Inicialização:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad \text{Equação 16}$$

$$\psi_1(i) = 0. \quad \text{Equação 17}$$

2. Recursão:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T \text{ e } 1 \leq j \leq N \quad \text{Equação 18}$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \text{ e } 1 \leq j \leq N \quad \text{Equação 19}$$

3. Terminação:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad \text{Equação 20}$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad \text{Equação 21}$$

4. Levantamento da seqüência de estados:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad \text{Equação 22}$$

É importante destacar que a implementação do algoritmo de Viterbi é similar (exceto pela etapa de levantamento da seqüência de estados) ao cálculo das equações 8 a 10 (procedimento *direto*). A maior diferença é a maximização

(equação 18) dos estados anteriores que é usada no lugar do procedimento de soma na equação 9. Cabe ressaltar que a estrutura mostrada na figura 4 também ilustra o algoritmo de Viterbi.

Mais detalhes sobre soluções dos problemas relacionados aos HMMs e aos HMMs em si encontram-se em (Rabiner, 1989).