

Aprendizagem Espectral em Modelos Ocultos de Markov

Jonas Rocha Lima Amaro

Instituto de Matemática Pura e Aplicada

Resumo

Modelos Ocultos de Markov são modelos versáteis para representar processos estocásticos e com isso, muitas perguntas podem ser feitas a cerca do processo partindo dos parâmetros do modelo. Neste trabalho, o leitor será introduzido à abordagem discreta de Modelos Ocultos de Markov e também à uma forma eficiente de resolver dois problemas: Calcular probabilidades de sequências; e calcular probabilidade da próxima observação dada a sequência de observações que o antecede. Finalmente, apresentaremos estudos comparativos que trazem em média uma redução de XX% no tempo a execução de rotinas se comparada à implementação mais comum.

Introdução

Cadeia de Markov

Cadeia de Markov é um processo estocástico no qual a probabilidade do próximo estado depende apenas do estado anterior. Se o processo ainda tiver uma quantidade finita de estados e o tempo for discreto, é possível representar a cadeia como grafo direcionado completo como representado na Figura 1. Cada estado corresponde a um nó do grafo, partindo de cada nó o estado seguinte é uma variável aleatória com as probabilidades representadas nos pesos de cada aresta. Por tanto, no caso ilustrado na Figura 1

$$\begin{aligned} \text{Para todo } e, e' \in E = \{A, B, C\} \\ P(e'|e) \in [0, 1] \text{ e} \\ P(A|e) + P(B|e) + P(C|e) = 1 \end{aligned}$$

Como se espera de uma função probabilidade com domínio discreto. A partir do grafo é bastante direto representar os parâmetros de forma matricial

$$T = \begin{pmatrix} P(A|A) & P(B|A) & P(C|A) \\ P(A|B) & P(B|B) & P(C|B) \\ P(A|C) & P(B|C) & P(C|C) \end{pmatrix}$$

Por outro lado, há redundâncias na matriz, pois cada linha soma 1, por isso é suficiente 6 parâmetros para determinar uma cadeia de 3 estados. Além disso, existe uma distribuição de probabilidade para o valor inicial da sequência que será representado como $\pi_0(e)$ que é equivalente ao vetor

$$\Pi_0 = \begin{pmatrix} \pi_0(A) \\ \pi_0(B) \\ \pi_0(C) \end{pmatrix}$$

De forma geral, em uma *Cadeia de Markov* com n estados, são necessários $n^2 - n$ parâmetros e

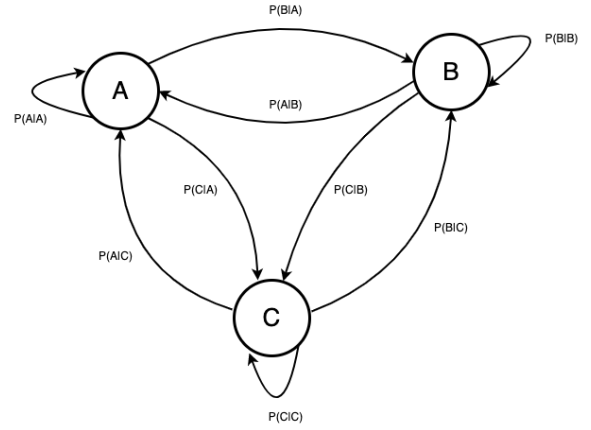


Figura 1: Exemplo de diagrama de estados de uma Cadeia de Markov com 3

cada probabilidade condicional é representada numa matriz $T \in \mathbb{R}^{n \times n}$ com entradas não negativas e que

$$T \mathbf{1}_n = \mathbf{1}_n$$

Isto é, cada linha soma 1. Além disso, também vetor $\Pi_0 \in \mathbb{R}^n$ com a distribuição de probabilidades para o estado inicial que também somam 1:

$$\Pi_0^T \mathbf{1}_n = 1$$

E da mesma forma que a matriz transição, à redundância no vetor Π_0 , pois é suficiente determinar apenas as probabilidades de $n - 1$ estados.

Concluí-se que uma *Cadeia de Markov* de n estados é definida por $n^2 - 1$ parâmetros.

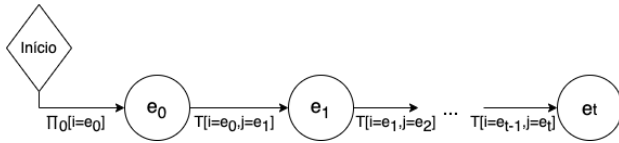


Figura 2: Visualização sequencial da Cadeia de Markov de distribuição de probabilidade inicial Π_0 e matriz transição T

Estimativa de parâmetros

Considere o seguinte conjunto sequências

ABCABACCACAB
BACABACAAC
BBABAAAACACABABABAC

Assumindo a premissa de que tais sequências foram geradas por um processo modelável como *Cadeia de Markov*, os parâmetros de tal cadeia serão estimados. É urgente se observar que as sequências apresentam apenas 3 estados. Por tanto, a matriz transição $T \in \mathbb{R}^{3 \times 3}$ e $\Pi_0 \in \mathbb{R}^3$.

No total, há 20 ocorrências da letra *A* seguida de algum estado, 10 da *B*, e 9 da *C*. Ao se contar cada ocorrência das tuplas *AA*, *AB*, *AC*, e assim por diante, dividindo pela ocorrência do respectivo estado antecessor. Assim a seguinte matriz transição é obtida

$$T = \begin{pmatrix} \frac{4}{20} & \frac{8}{20} & \frac{8}{20} \\ \frac{8}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{7}{9} & \frac{0}{9} & \frac{2}{9} \end{pmatrix}$$

Já para a estimativa do vetor probabilidade de estado inicial Π_0 , o processo é imediato, é a frequência pela quantidade de sequências

$$\Pi_0 = \left(\frac{1}{3} \quad \frac{2}{3} \quad 0 \right)$$

Probabilidade de uma sequência

O interesse nesta subseção é de calcular, partindo dos parâmetros que descrevem uma *Cadeia de Markov*, a probabilidade de cada sequência de estados. Seja $(e_i)_{i=1}^t$ um processo de Markov que gere uma sequência de estados de uma *Cadeia de Markov* que tenha probabilidade inicial Π_0 e uma matriz transferência de estados T , como ilustrado na figura 2. Como Bishop [1] apresenta, a transferência de estado só depende do estado anterior, que simplifica bastante o cálculo da probabilidade conjunta a partir das probabilidades condicionais

$$P((e_k)_{k=1}^t) = P(e_1) \prod_{k=2}^t P(e_k | e_{k-1})$$

$$\Rightarrow P((e_k)_{k=1}^t) = \Pi_0[i = e_1] \prod_{k=2}^t T[i = e_{k-1}, j = e_k]$$

No caso do modelo da seção anterior, a sequência *BACABBAC* tem probabilidade

$$P(BACABBAC) = P(B)P(A|B)P(C|A)P(A|C)P(B|A)P(B|B)P(A|B)P(C|A)$$

$$= 2/38/108/207/98/201/108/108/20 \approx 2.212 * 10^{-3}$$

O leitor mais atento observou que a medida que as sequências crescem elas se tornam mais improváveis, por isso deve haver uma precaução na manipulação dessas probabilidades e evitar comparar sequências de tamanhos diferentes.

Modelo Oculto de Markov

Aprendizagem Espectral

Experimentos

Discussões

Referências

- [1] Christopher M. Bishop. *Pattern recognition and machine learning*. Textbook for graduates. New York: Springer, c2006. Pp. 607–610. URL: <http://www.loc.gov/catdir/enhancements/fy0818/2006922522-t.html>.