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

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

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

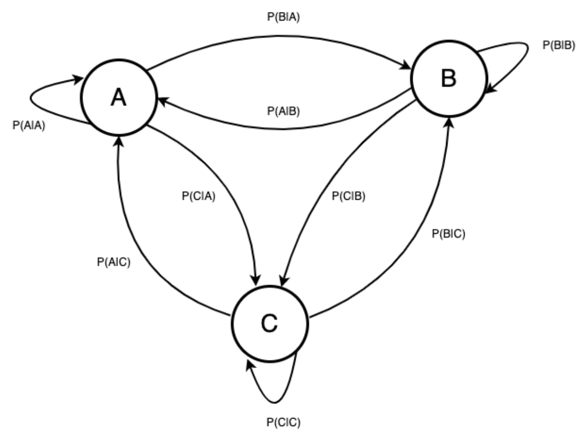

Figura 1: *Diagrama de estados.*

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:

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

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.

Exemplo

Para ilustrar a estimativa dos parâmetros de uma *Cadeia de Markov*, considere o seguinte conjunto sequências

$$ABCABACCACAB$$
$$BACABACAAC$$
$$BBABAAAACCACABABABAC$$

Para estimar a matriz transição T, observe que, há no total que 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

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

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

$$\Pi_0 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix}$$

Probabilidade Estacionária

Agora suponha que algum processo é caracterizado como *Cadeia de Markov* tenha matriz transição T, então note

$$(T - I)\mathbf{1} = T\mathbf{1} - I\mathbf{1} = \mathbf{1} - \mathbf{1} = O$$

Isto é, tal matriz é $T - I$ não é linearmente independente e por tanto, tem determinante 0. Isto implica que 1 é autovalor da matriz T. Por tanto existe um vetor linha V tal que

$$VT = V$$

Como todas entradas de T são não-negativas, V só pode ter entradas não-negativas. Além disso, podemos multiplicar a igualdade por qualquer escalar em ambos os lados que a igualdade se mantém, por isso podemos forçar que $V\mathbf{1} = 1$. Chama-se V de *Probabilidade Estacionária*, isto é, a distribuição de probabilidade dos estados converge para V

Modelos Alternativos

Há também a possibilidade de expandir o modelo adicionando mais estados na dependência do estado seguinte, como Bishop[1] discute, mas isso vem com o custo de mais variáveis para se calcular. A medida que se adiciona 1 estado na dependência, se multiplica a quantidade de parâmetros por n, a quantidade de estados.

Este nível de detalhamento pode tornar o modelo impraticável quando esta expansão se torna muito grande. Além disso, as estimativas dos parâmetros desse modelo passa depender de mais observações. Observe que não há ocorrências do par CB no exemplo visto. Isto pode ter sido causado pelo fato de que o processo não permita essa transição de estado, mas também é possível que essa transição só não tenha sido observada na amostra apresentada.

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`.