

Estudo de padrões ordinais com grafos de transição e Cadeias de Markov

Eduarda T. C. Chagas¹, Heitor S. Ramos¹, Alejandro C. Frery, Osvaldo A. Rosso

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

eduarda.chagas@dcc.ufmg.br

Definição do problema

Uma das limitações das técnicas de análise de séries temporais baseadas em padrões ordinais é não permitirem fazer previsões (*forecast*) nem simulações. Este trabalho explora essa possibilidade.

Sejam $\mathbf{Z} = (Z_1, \dots, Z_n)$ uma série temporal de valores reais, $\boldsymbol{\pi}^{(D,\tau)} = (\pi_1^{(D,\tau)}, \dots, \pi_{N-(D-1)\tau}^{(D,\tau)})$ a série de padrões a ela associados (calculados com palavras de dimensão D e atraso τ), e $G = (V, E)$ o grafo de transições obtido a partir de $\boldsymbol{\pi}^{(D,\tau)}$.

A proposta consiste em fazer a junção das evidências de \mathbf{Z} , $\boldsymbol{\pi}^{(D,\tau)}$, e G para fazer simulações e previsões a respeito da série.

O primeiro passo consiste em analisar \mathbf{Z} e $\boldsymbol{\pi}^{(D,\tau)}$. Para cada padrão observado em $\boldsymbol{\pi}^{(D,\tau)}$, serão coletados os dados que o originaram em \mathbf{Z} . Consideremos, por exemplo, o caso do padrão $\pi_1^{(3,1)} = b_j b_{j+1} b_{j+2}$, e suponhamos que ele corresponde a todas as palavras que satisfazem $z_j < z_{j+1} < z_{j+2}$. Todas essas palavras serão coletadas e analisadas para obter:

- uma estimativa da distribuição três-variada, ou
- estimativas do valor central e estimativas de uma medida de dispersão de cada um dos três valores, por exemplo a média e o desvio padrão.

Teremos, assim, associados ao padrão $\pi_1^{(3,1)}$,

- um modelo $\widehat{\mathcal{D}}(\pi_1^{(3,1)})$, ou
- três médias $\widehat{\mu}_{b_j}, \widehat{\mu}_{b_{j+1}}, \widehat{\mu}_{b_{j+2}}$ e três desvios padrão $s_{b_j}, s_{b_{j+1}}, s_{b_{j+2}}$.

O segundo passo consiste em formar a matriz de transições do grafo G , digamos M . Por construção, a cadeia é irredutível, e basta com que haja uma única transição entre estados iguais para que a cadeia seja aperiódica. Com estas propriedades, há uma única distribuição de equilíbrio Π , que é a solução de $\Pi M = \Pi$.

Dada a série temporal $\mathbf{Z} = (Z_1, \dots, Z_n)$, associada à sequência $\boldsymbol{\pi}^{(D,\tau)} = (\pi_1^{(D,\tau)}, \dots, \pi_{N-(D-1)\tau}^{(D,\tau)})$ de padrões, simularemos o evento Z_{n+1} com dois elementos:

- o padrão $\pi_1^{(D,\tau)}, \dots, \pi_{N-(D-1)\tau+1}^{(D,\tau)}$ que possui probabilidade máxima de ocorrência em Π após o último padrão, e
- uma observação do modelo \mathcal{D} correspondente a esse padrão. Note-se que será necessário obter apenas uma amostra da distribuição marginal de \mathcal{D} dadas as observações já presentes nos últimos estágios da série.

A nossa previsão da observação que sucede Z_n será o estado de equilíbrio mais plausível que segue ao último padrão que inclui Z_n , e em cada posição do padrão colocaremos a estimativa de centralidade, munida da sua estimativa de precisão. De forma mais sofisticada, poderemos usar o algoritmo EM (?).