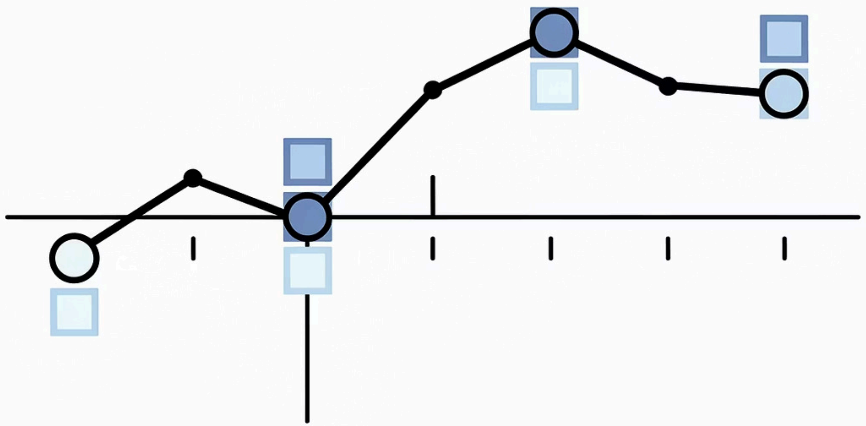


Detecção de Pontos de Mudança

Esta apresentação explora as complexidades da detecção de pontos de mudança, abordando diversos métodos de detecção.

Eduardo Ogasawara

eduardo.ogasawara@cefet-rj.br

https://eic.cefet-rj.br/~eogasawara

Definição

A detecção de pontos de mudança é um problema fundamental de estimação em séries temporais. Dada uma série observada, buscamos identificar os instantes onde a estrutura dos dados muda significativamente.

O objetivo é escolher a segmentação que melhor explica os dados, tipicamente maximizando a verossimilhança do modelo segmentado. Definimos os instantes $\tau_1, \ldots, \tau_k$ onde a estrutura muda, buscando o conjunto C que torna os dados mais prováveis segundo um modelo com segmentos.

$$\hat{\mathcal{C}} = \arg\max_{\mathcal{C}} \ \mathcal{L}(X \mid \mathcal{C})$$

Teste de Hipóteses para Mudança Estrutural

Visão Estatística

Detectar mudança equivale a rejeitar a hipótese de estabilidade estrutural através de testes formais.

A detecção pode ser formulada como um teste de hipóteses: sob a hipótese nula, os parâmetros do processo não mudam ao longo do tempo. Sob a hipótese alternativa, existe pelo menos um instante onde os parâmetros sofrem ruptura.

$$H_0 : \ \theta(t) = \theta_0 \text{ para todo } t$$

$$H_1 : \ \exists \tau \text{ tal que } \theta(\tau) \neq \theta(\tau - 1)$$

Segmentação Ótima

A detecção de pontos de mudança pode ser transformada em um problema de otimização. Toda segmentação induz um custo total que combina o ajuste dentro dos segmentos com uma penalização pela complexidade do modelo.

Custo por Segmento

Mede quão bem o modelo ajusta cada trecho da série

Penalização

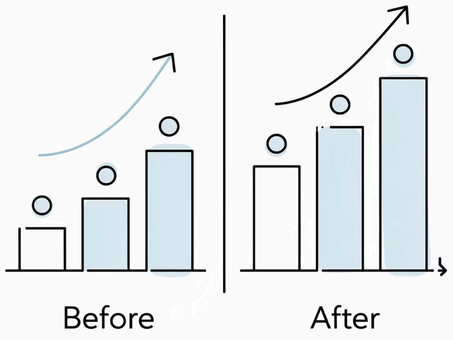
Controla o número de mudanças para evitar complexidade excessiva

Objetivo

Minimizar o custo total penalizado

Segmentações com muitos pontos ajustam melhor os dados, mas são mais complexas. A penalização β controla esse trade-off:

$$J(\mathcal{C}) = \sum_{i=0}^{k} \mathcal{L}(X_{\tau_i:\tau_{i+1}}) + \beta\, k$$



Estatísticas de Mudança

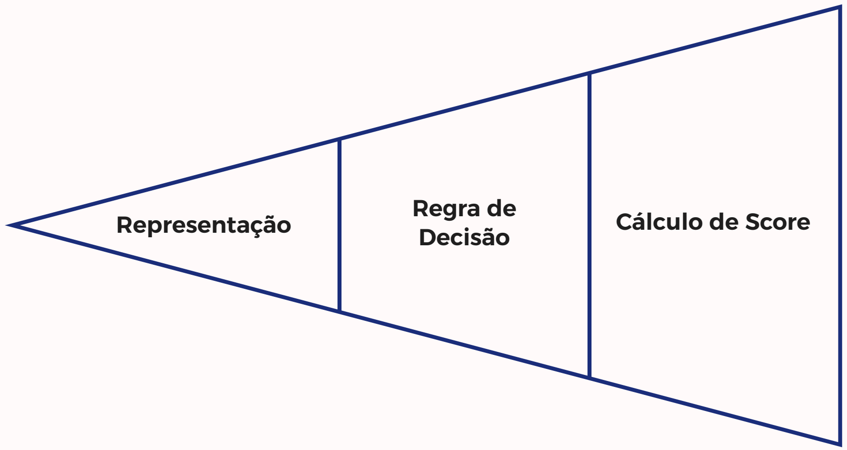
O coração de qualquer método de detecção é uma estatística que mede a discrepância entre segmentos. Construímos um "termômetro" que compara os dados antes e depois de cada tempo t.

A função ϕ pode representar diferença de médias, razão de verossimilhança ou divergência entre distribuições. A decisão ocorre quando o score ultrapassa um limiar predefinido:

$$S_t = \phi(X_{1:t}, X_{t+1:T})$$

$$S_t > \tau$$

Pipeline de Detecção



Todo método de detecção segue uma arquitetura comum com três etapas fundamentais. Primeiro transformamos a série em uma representação adequada, depois calculamos uma estatística de mudança, e finalmente aplicamos uma regra de decisão para identificar os pontos.

Essa decomposição ajuda a comparar métodos diferentes: alguns modificam a representação (trabalhando com resíduos ou embeddings), outros mudam a estatística de discrepância, e outros ajustam a regra de decisão.

Teste de Chow

Contexto

O teste de Chow é aplicado em regressão linear quando suspeitamos de uma quebra estrutural em um tempo τ específico.

Lógica

Comparamos um modelo único para todo o período versus dois modelos separados (antes e depois de τ). Se não existe mudança estrutural, ajustar dois modelos não deveria melhorar significativamente o ajuste.

Estatística F

A estatística F mede o ganho relativo de permitir dois modelos, ajustado pela quantidade de parâmetros e tamanho amostral:

$$F = \frac{(RSS_R - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)}$$

Quando existe mudança, o erro total (RSS_1+RSS_2) cai bastante em relação ao RSS_R, resultando em um valor F elevado.

CUSUM (Cumulative Sum)

O CUSUM é um detector clássico que transforma pequenas evidências repetidas em um sinal grande. Acumulamos os resíduos ao longo do tempo para detectar mudanças persistentes.

Se o processo está estável, resíduos positivos e negativos tendem a se cancelar e a soma cumulativa fica próxima de zero. Quando há mudança persistente, os resíduos passam a ter viés e a soma "derrapa" para longe do zero:

$$C_t = \sum_{i=1}^{t} r_i$$

$$|C_t| > h$$

A decisão ocorre quando a magnitude de C_t ultrapassa o limiar h.

Teste de Page-Hinkley

01

Calcular média acumulada

Computamos a média dos dados até o tempo t

02

Acumular desvios com tolerância

Somamos desvios em relação à média, descontando δ

03

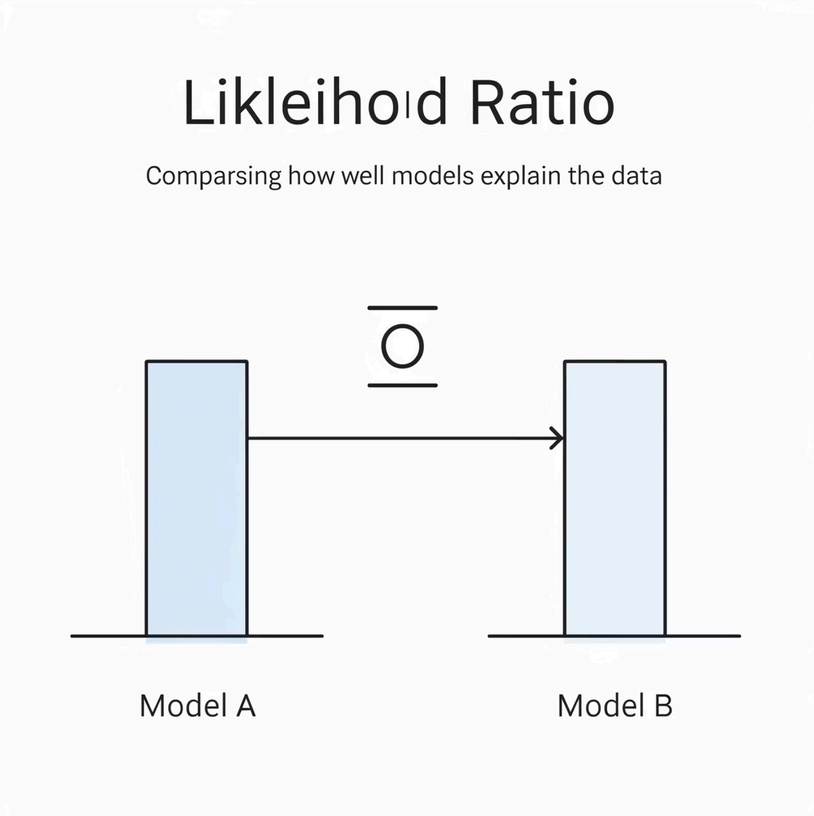
Verificar limiar

Sinalizamos mudança quando PH_t ultrapassa λ

O Page-Hinkley é uma variação sequencial focada em mudanças na média. A tolerância δ evita que ruído gere crescimento artificial da estatística. Quando a média muda de forma sustentada, os termos $(x_i - \bar{x}_t - \delta)$ ficam positivos por muito tempo:

$$PH_t = \sum_{i=1}^{t} (x_i - \bar{x}_t - \delta)$$

Teste de Razão de Verossimilhança



A razão de verossimilhança oferece uma forma geral de construir detectores. Comparamos duas "histórias" concorrentes sobre os dados.

Na história sem mudança, um único conjunto de parâmetros explica todo o período. Na história com mudança em t, permitimos parâmetros diferentes antes e depois.

$$\Lambda(t) = \frac{\sup_{\theta_1,\theta_2} \mathcal{L}(X_{1:t}, X_{t+1:T})}{\sup_{\theta} \mathcal{L}(X_{1:T})}$$

Se a melhoria de verossimilhança for grande, log Λ(t) passa do limiar e marcamos um ponto de mudança.

Programação Dinâmica

A programação dinâmica aparece porque escolher cortes "um a um" pode falhar: o melhor conjunto de mudanças é um problema de otimização global.

Primeiro definimos o custo de ajustar um modelo em qualquer trecho da série. Depois buscamos a segmentação global que minimiza a soma dos custos, com penalização por complexidade:

$$J(k) = \min_{\tau_1,...,\tau_k} \sum_{m=0}^{k} \mathcal{L}(X_{\tau_m:\tau_{m+1}}) + \beta k$$

Entre todas as formas de escolher k cortes, pegamos a que minimiza o custo total. Muitos algoritmos modernos partem exatamente dessa estrutura.

Detecção Bayesiana de Change Points

Abordagem Probabilística

Pontos de mudança são tratados como variáveis aleatórias do modelo

Combinação de Evidências

Integramos a evidência dos dados com crenças prévias sobre segmentações

Quantificação de Incerteza

A saída é uma distribuição a posteriori completa sobre C

O enfoque bayesiano calcula quão provável é cada conjunto C depois de observar os dados. A diferença fundamental é que você não precisa sair apenas com "uma resposta" - pode quantificar incerteza sobre quais mudanças são bem suportadas:

$$P(\mathcal{C} \mid X) \propto P(X \mid \mathcal{C}) \cdot P(\mathcal{C})$$

Change Points via HMM

Modelos de Markov Ocultos

Em um HMM, a série tem um estado oculto z_t que indica o regime no tempo t. Cada regime governa a distribuição das observações.

Mudança estrutural aparece como transição entre estados latentes. Isso é útil quando esperamos alternância recorrente de regimes, não apenas uma única quebra.

$$z_t \in \{1, \ldots, K\}, \qquad x_t \sim F_{\theta_{z_t}}$$

Detecção de Transições

Um ponto de mudança ocorre quando o estado latente troca:

$$z_t \neq z_{t-1}$$

O HMM formaliza a ideia de regimes alternantes, transformando detecção de mudanças em detecção de transições de estado.

Change Points via Machine Learning

A abordagem de machine learning reformula change points como "falha de previsão". Primeiro transformamos a série em uma representação Z_t, depois treinamos um modelo preditivo para o comportamento "típico".

Se o modelo aprendeu bem a dinâmica de um regime, ele prevê com erro pequeno. Quando a dinâmica muda, a previsão piora sistematicamente e o erro ultrapassa o limiar:

$$Z_t = \mathcal{R}(x_t), \qquad Z_t \approx f_\theta(Z_{t-1}, \ldots, Z_{t-p})$$

$$g_t = |Z_t - \hat{Z}_t|, \qquad g_t > \tau$$

O papel da representação Z_t é crucial: pode facilitar o aprendizado e tornar a ruptura mais visível.

Segmentação por Distância

Definir Dissimilaridade

Escolher função de distância d entre segmentos

Comparar Trechos

Medir quão diferentes são segmentos adjacentes

Aplicar Limiar

Sinalizar mudança quando $D(i,j) > \tau$

Essa abordagem não assume modelo paramétrico - mede diretamente quão diferentes são dois trechos. Se a distância entre "antes" e "depois" fica grande, isso sugere ruptura. A escolha de d define o método: pode focar em média, forma, distribuição ou padrões:

$$D(i,j) = d(X_{i:j}\,, X_{j+1:k})\,, \qquad D(i,j) > \tau$$

Detecção Offline vs Online

Modo Offline

Assume um conjunto finito e completo de dados disponível. Pode comparar passado e futuro porque todo o histórico está acessível.

Permite algoritmos mais sofisticados que consideram toda a série simultaneamente para encontrar a segmentação ótima global.

$$X_{1:T}$$

Modo Online

Dados chegam em fluxo contínuo e decisões devem ser tomadas em tempo real, sem ver o futuro.

Exige estatísticas que atualizam incrementalmente e regras que equilibram rapidez de detecção com controle de falsos alarmes.

$$X_t$$

Avaliação Formal de Change Points

Para avaliar métodos de detecção, comparamos o conjunto real C com a estimativa Ĉ. Como change points são eventos no tempo, raramente acertamos exatamente o índice correto.

P

Precisão

Quão "limpo" é o conjunto detectado

R

Revocação

Quantos verdadeiros foram encontrados

F_1

F1-Score

Equilíbrio entre precisão e revocação

Incluímos tolerância temporal ε: um ponto estimado conta como acerto se estiver "perto o suficiente" do verdadeiro. As métricas clássicas são:

$$\text{Precision} = \frac{|\hat{\mathcal{C}} \cap \mathcal{C}|}{|\hat{\mathcal{C}}|} \qquad \text{Recall} = \frac{|\hat{\mathcal{C}} \cap \mathcal{C}|}{|\mathcal{C}|}$$

Métricas Temporais

Além de acertar "se" existe mudança, importa acertar "quando" ela ocorre. Medimos o deslocamento entre tempo estimado e verdadeiro para cada evento.

O erro médio absoluto resume a precisão temporal:

$$\Delta\tau_i = \hat{\tau}_i - \tau_i \qquad MAE_\tau = \frac{1}{|\mathcal{C}|}\sum_i |\Delta\tau_i|$$

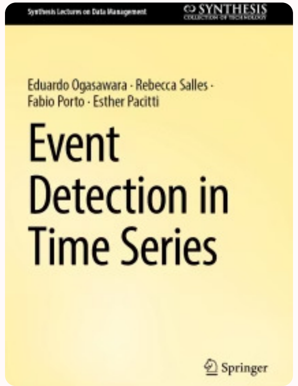
Podemos combinar acurácia com timing por meio de penalização que decai com a distância:

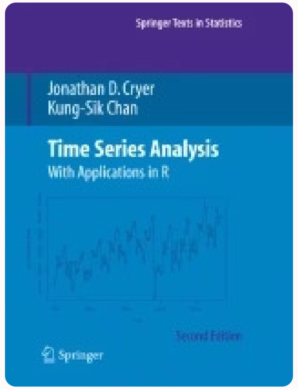
$$\mathcal{P}(\Delta\tau) = e^{-\lambda|\Delta\tau|} \qquad Q = F_1 \cdot \mathcal{P}(\Delta\tau)$$

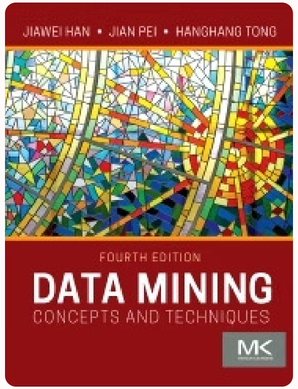
RECURSOS ACADÊMICOS

Referências Bibliográficas

Uma coleção cuidadosamente selecionada de obras fundamentais que abordam análise de séries temporais e mineração de dados.







Event Detection in Time Series

Ogasawara, E.; Salles, R.; Porto, F.; Pacitti, E. (2025). Publicação recente da Springer Nature Switzerland que explora técnicas avançadas de detecção de eventos em séries temporais.

Time Series Analysis: With Applications in R

Cryer, J. D.; Chan, K.-S. (2008). Obra clássica da Springer que combina fundamentação teórica sólida com implementações práticas.

Data Mining: Concepts and Techniques

Han, J.; Pei, J.; Tong, H. (2022). Quarta edição publicada pela Morgan Kaufmann que consolida conceitos fundamentais e técnicas avançadas de mineração de dados