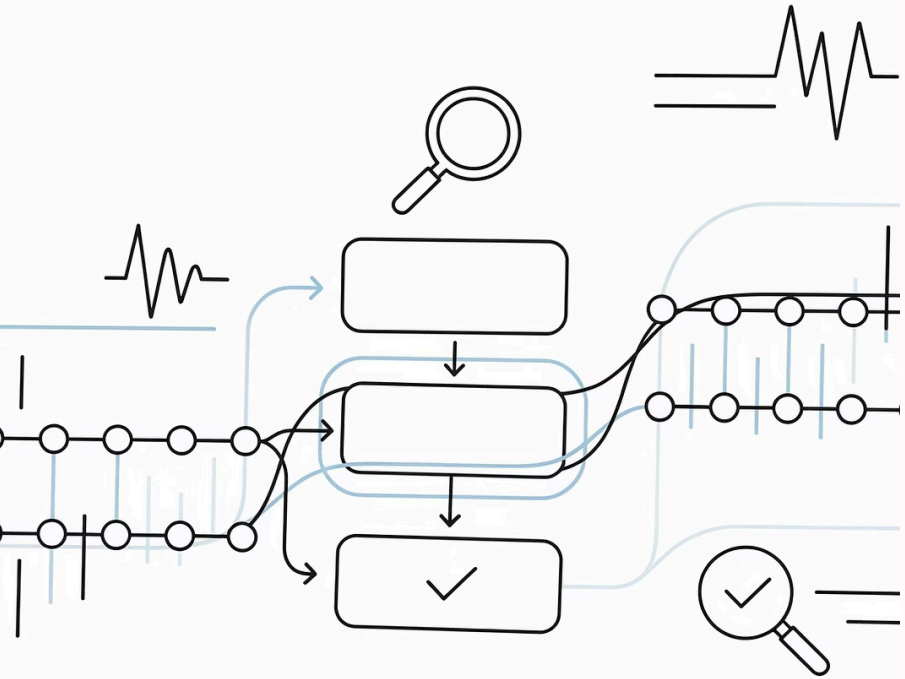


CEFET/RJ

Formulação do Problema de Detecção

Transformando séries temporais em decisões estruturadas sobre anomalias.

Eduardo Ogasawara

eduardo.ogasawara@cefet-rj.br

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

Conceito Formal de Detecção

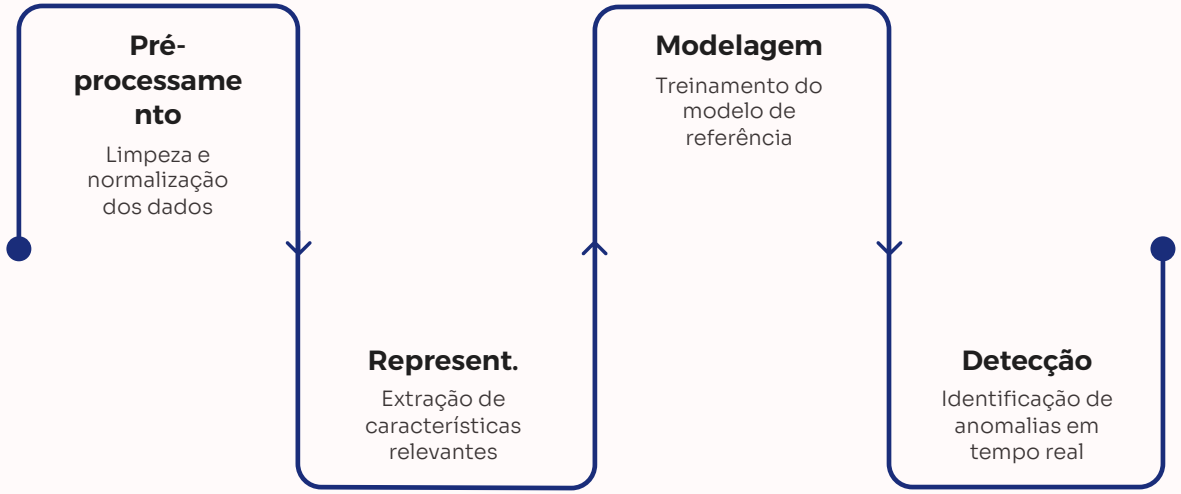
A detecção de anomalias em séries temporais é formulada como um problema de mapeamento estruturado. Uma série temporal X = {x_t}_{t=1}^{T} serve como entrada, onde a representação temporal define o espaço de trabalho para análise.

O modelo estabelece o conceito de tipicidade nesse espaço representado, enquanto o detector mapeia a representação em um conjunto de anomalias identificadas. Este framework formaliza que anomalias não são detectadas diretamente nos dados brutos, mas em uma representação transformada.

$$\mathcal{D} : \mathcal{R}(X) \rightarrow \mathcal{A}$$

PIPELINE

Arquitetura do Pipeline de Detecção



O pipeline transforma dados brutos em decisões de anomalia através de etapas bem definidas, cada uma com papel específico no processo.

A série temporal passa por transformações $\mathcal{T}$ no pré-processamento, gerando $X_t^{(prep)}$. A função de representação $\mathcal{R}$ produz variáveis Z_t adequadas para modelagem.

O modelo $\mathcal{M}$ gera o comportamento esperado $\hat{Z}_t$, e finalmente o detector $\mathcal{D}$ aplica regras de decisão para identificar o conjunto $\mathcal{A}$ de anomalias.

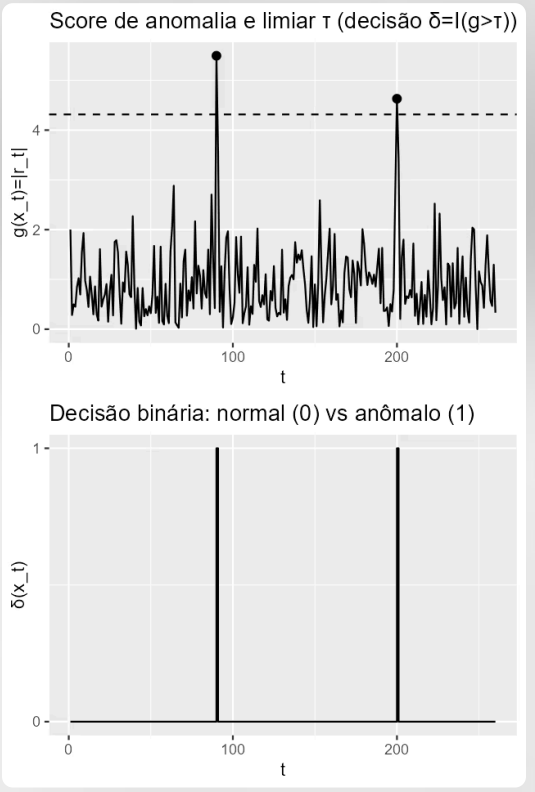
$$X_t \xrightarrow{\mathcal{T}} X_t^{(prep)} \xrightarrow{\mathcal{R}} Z_t \xrightarrow{\mathcal{M}} \hat{Z}_t \xrightarrow{\mathcal{D}} \mathcal{A}$$

Hipótese de Tipicidade

A tipicidade é formulada como uma hipótese nula H_0, onde $x_t \sim F_{\theta_0}$ representa o comportamento esperado sob condições típicas. A hipótese alternativa H_1 representa violações deste modelo de tipicidade.

Detectar anomalias equivale a rejeitar H_0 em instantes específicos quando há evidência suficiente de desvio. Esta formulação aproxima a detecção de anomalias dos testes estatísticos clássicos, fornecendo base teórica rigorosa para decisões.

$$H_0 : x_t \sim F_{\theta_0}, \quad H_1 : x_t \not\sim F_{\theta_0}$$



Espaço de Decisão do Detector

Decisão Binária

O detector produz uma decisão binária: anômalo (1) ou normal (0) para cada instante temporal.

Score de Anomalia

A função $g(x_t)$ quantifica o grau de desvio, convertendo observações em valores numéricos comparáveis.

Limiar de Controle

O parâmetro τ controla a sensibilidade do sistema, definindo o ponto de corte para classificação.

A regra de decisão é formalizada através da função indicadora: $\delta(x_t) = \mathbb{I}(g(x_t) > \tau)$. Esta separação entre score e limiar permite ajustar a sensibilidade sem modificar o modelo subjacente.

Dimensões da Detecção de Anomalias

Todo método de detecção precisa especificar escolhas em múltiplas dimensões que definem seu comportamento operacional. A representação $\mathcal{R}$ determina qual "sinal" é considerado relevante para análise.

01

Representação

Define o espaço de trabalho e quais características são extraídas dos dados brutos

02

Modelo

Estabelece o conceito de tipicidade no espaço representado

03

Score

Quantifica a medida de desvio em relação à tipicidade esperada

04

Limiar

Define a regra operacional para classificação binária

05

Cenário Temporal

Impõe restrições de processamento: offline, online ou preditivo

ESTATÍSTICO

Detecção Estatística de Anomalias

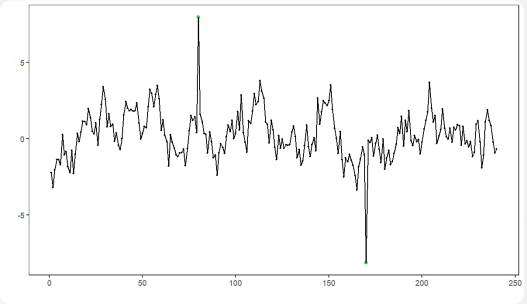
No paradigma estatístico, a normalidade é modelada através de uma distribuição condicional F_θ, conhecida ou estimada a partir dos dados históricos.

A estatística $S_t = \phi(x_t, F_\theta)$ resume a evidência de desvio em relação ao modelo probabilístico. A decisão é tomada aplicando um limiar sobre esta estatística.

Exemplos incluem resíduos padronizados, CUSUM, e razão de verossimilhança. Este framework formaliza "improvável" como critério operacional mensurável.

$$x_t | \mathcal{F}_{t-1} \sim F_\theta$$

$$S_t = \phi(x_t, F_\theta), \quad \delta(x_t) = \mathbb{I}(S_t > \tau)$$



BASEADO EM MODELOS

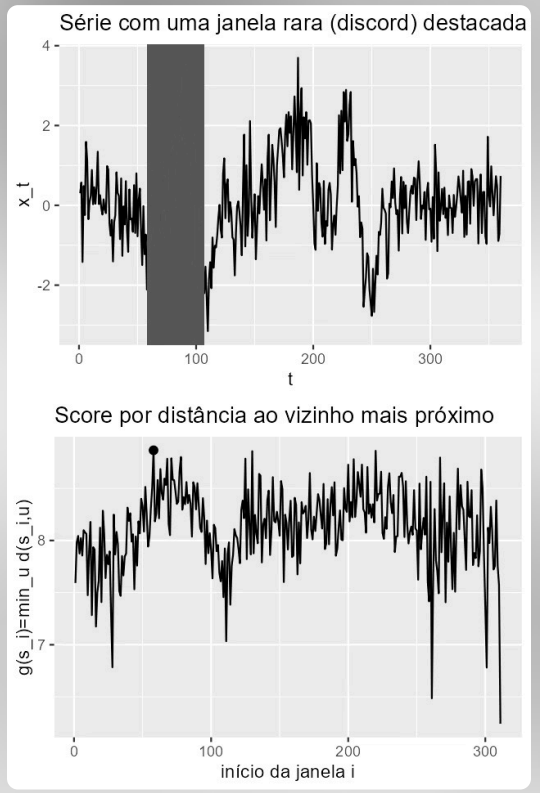
Detecção Baseada em Modelos Temporais

Neste paradigma, um modelo temporal explícito captura a dependência no tempo através da função $f(x_{t-1}, \ldots, x_{t-p})$, onde p representa o número de defasagens consideradas.

A previsão $\hat{x}_t$ define o comportamento esperado sob normalidade. O score de anomalia é calculado como o erro de previsão: $g_t = |x_t - \hat{x}_t|$. Uma anomalia é sinalizada quando este erro excede o limiar estabelecido.

Este approach conecta diretamente detecção com modelagem preditiva: quanto melhor o modelo captura a dinâmica normal, mais informativo tende a ser o erro como indicador de anomalia.

$$x_t = f(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_t$$



BASEADO EM DISTÂNCIA

Detecção Baseada em Distância

Este paradigma trabalha com subsequências de tamanho w, formando o conjunto $\mathcal{S} = \{X_{t:t+w-1}\}$. A normalidade não é definida por um modelo probabilístico, mas pela existência de padrões semelhantes na própria série.

O score é calculado como a distância à subsequência mais parecida: $g(s) = \min_{u \in \mathcal{S}, u \neq s} d(s, u)$. Uma subsequência "isolada" no espaço de padrões, sem vizinhos próximos segundo a métrica $d(\cdot, \cdot)$, é candidata a anomalia.

Este enquadramento é natural para anomalias coletivas e conecta diretamente com o conceito de discords em séries temporais.

APRENDIZADO DE MÁQUINA

Detecção com Aprendizado de Máquina

A representação $Z_t = \mathcal{R}(x_t)$ define variáveis adequadas para aprendizado automático. O modelo f_θ aprende uma representação implícita de normalidade diretamente dos dados, sem necessidade de especificação explícita.

O score de anomalia é derivado do erro de reconstrução ou predição no espaço transformado: $g_t = |Z_t - \hat{Z}_t|$. Quando uma observação não é bem reconstruída ou prevista, isso indica desvio do padrão aprendido.

O modelo não precisa ser interpretável como distribuição ou equação simples - sua força está na capacidade de capturar padrões complexos e não-lineares nos dados históricos.

$$Z_t \approx f_\theta(Z_{t-1}, \ldots, Z_{t-p})$$

Síntese dos Paradigmas de Detecção

Estatístico

Tipicidade definida por distribuição probabilística conhecida ou estimada

Baseado em Modelos

Tipicidade expressa através de previsão temporal explícita

Baseado em Distância

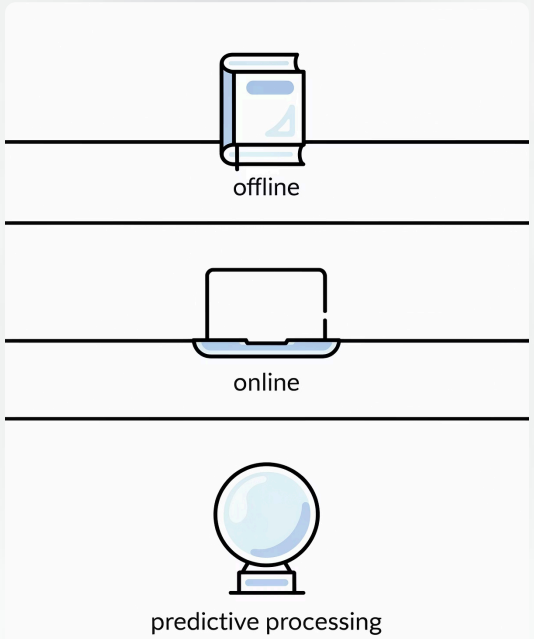
Tipicidade determinada por similaridade de padrões na série

Aprendizado

Tipicidade aprendida implicitamente via representação e erro

Todos os paradigmas executam a mesma operação em alto nível - produzir score e aplicar limiar - mas diferem fundamentalmente em duas decisões: qual representação $\mathcal{R}$ utilizar e como definir tipicidade através de $\mathcal{M}$.

$$\mathcal{D} = f(\mathcal{R}, \mathcal{M}, g, \tau)$$



Cenários Temporais de Operação

Offline

Série completa $X_{1:T}$ disponível. Permite calibração global de modelos e limiares com toda informação.

Online

Dados chegam sequencialmente $X_t, t \to \infty$. Decisões tomadas à medida que observações aparecem.

Preditivo

Estima probabilidade de anomalia futura $\mathbb{P}(a_{t+h} = 1 | \mathcal{F}_t)$ com horizonte h.

A mesma definição de anomalia exige estratégias diferentes dependendo do cenário temporal. Cada modo impõe restrições distintas de memória, latência e capacidade de adaptação.

ARQUITETURA

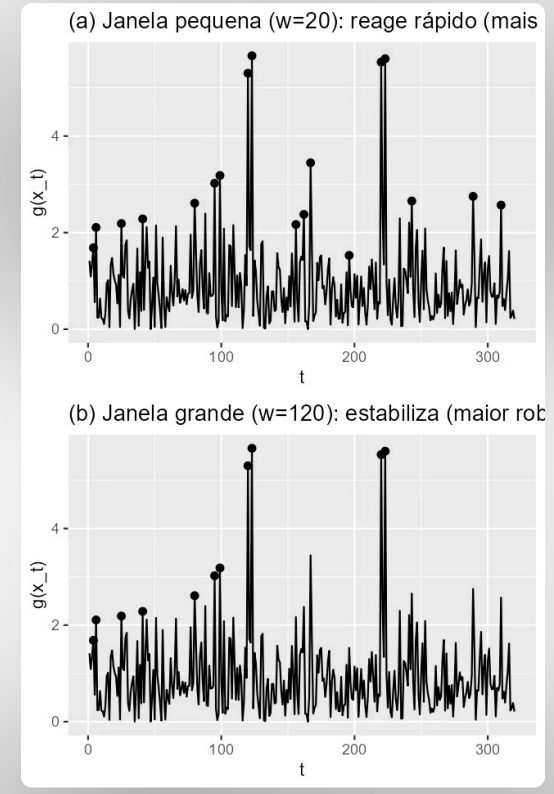
Arquiteturas Computacionais

A implementação pode ser batch, streaming ou híbrida, cada uma com características operacionais distintas. A arquitetura engloba escolhas de representação, modelo, detector e estratégia de janela temporal.

A janela $\mathcal{W}$ controla como o tempo é incorporado no algoritmo, afetando atualização de parâmetros, estabilidade e atraso de detecção. Esta decisão determina custo computacional e latência do sistema.

$$\mathcal{A}rch = (\mathcal{R}, \mathcal{M}, \mathcal{D}, \mathcal{W})$$

Separar "método" de "implementação" é crucial: o mesmo detector conceitual pode rodar em diferentes arquiteturas, cada uma adequada a contextos operacionais específicos.



Papel das Janelas Temporais

A janela de tamanho w define o "contexto operacional" sobre o qual o detector opera. Em vez de processar a série completa, o sistema analisa a subsequência recente $X_{t-w+1:t}$.

1

Janela Pequena

Reação rápida, mas pode confundir ruído com anomalia

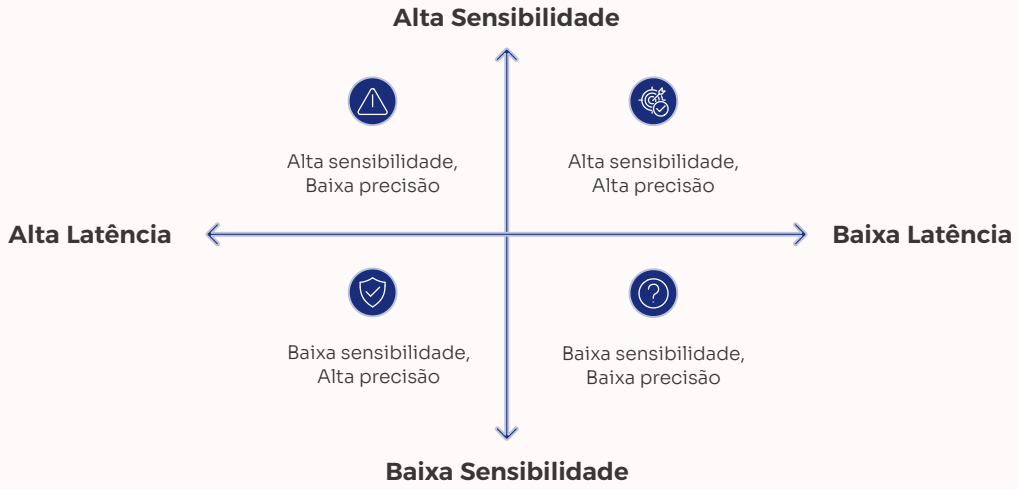
2

Janela Grande

Decisão estabilizada, mas aumenta latência e pode diluir desvios curtos

A escolha de w altera fundamentalmente a sensibilidade e granularidade do detector, mudando o próprio conjunto de anomalias identificadas: $\mathcal{A}(w) \neq \mathcal{A}(w')$. Portanto, o tamanho da janela não é apenas um parâmetro técnico, mas parte da definição operacional de anomalia.

Trade-offs na Detecção de Anomalias



A função de custo formaliza os compromissos operacionais do detector através de pesos que refletem prioridades da aplicação.

Não existe detector ótimo universal - cada aplicação requer um ponto de operação específico. A sensibilidade compete com precisão (falsos negativos vs falsos positivos), latência compete com robustez, e complexidade compete com interpretabilidade.

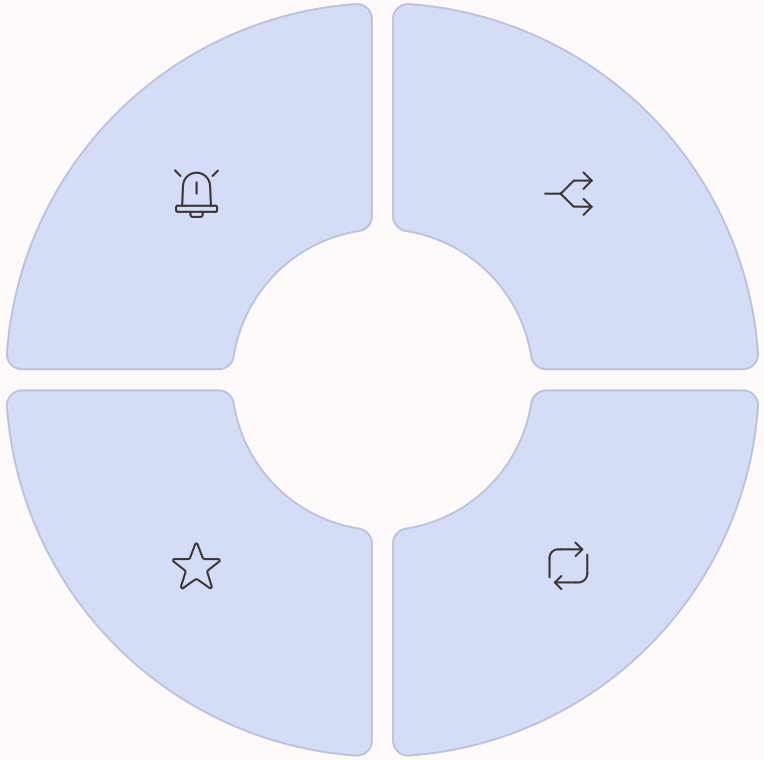
A função $\mathcal{L}(\mathcal{D}) = \alpha \cdot FP + \beta \cdot FN + \gamma \cdot \Delta t$ penaliza falsos positivos, falsos negativos e atraso de detecção. Ajustar os pesos α, β, γ muda fundamentalmente o que significa "bom detector" para o contexto específico.

TAXONOMIA

Anomalias na Taxonomia de Eventos

Anomalias

Desvios locais que representam o nível inicial de detecção de eventos



Change Points

Rupturas persistentes que marcam transições entre regimes distintos

Discords

Padrões raros que não possuem vizinhos similares no espaço de subsequências

Motifs

Padrões recorrentes que aparecem múltiplas vezes na série temporal

A união formal $\mathcal{E} = \mathcal{A} \cup \mathcal{C} \cup \mathcal{M} \cup \mathcal{D}$ organiza as classes no guarda-chuva conceitual de "detecção de eventos". Anomalias servem como porta de entrada por serem a forma mais direta de evento.

AVALIAÇÃO

Avaliação Formal de Detectores

A avaliação é naturalmente baseada em conjuntos: comparamos anomalias detectadas $\hat{\mathcal{A}}$ com anomalias reais $\mathcal{A}$. A precisão mede a proporção de detecções corretas, enquanto a revocação mede a cobertura das anomalias verdadeiras.

$$\text{Precision} = \frac{|\hat{\mathcal{A}} \cap \mathcal{A}|}{|\hat{\mathcal{A}}|}$$

$$\text{Recall} = \frac{|\hat{\mathcal{A}} \cap \mathcal{A}|}{|\mathcal{A}|}$$

A métrica F1 combina precisão e revocação em uma única medida balanceada, fornecendo resumo do desempenho geral.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Métricas Temporais de Qualidade

Em séries temporais, detectar "certo" tarde demais pode ser operacionalmente inútil. O desempenho não é apenas acerto, mas também pontualidade. O atraso $\Delta t = t_{det} - t_{real}$ mede a latência entre o evento real e sua detecção.

01

Calcular Atraso

Diferença temporal entre detecção e ocorrência real

02

Aplicar Penalização

Função $\mathcal{P}(\Delta t) = e^{-\lambda \Delta t}$ reduz pontuação com atraso crescente

03

Combinar Métricas

Qualidade final $Q = F_1 \cdot \mathcal{P}(\Delta t)$ integra acurácia e tempo

O parâmetro $\lambda > 0$ controla a taxa de penalização temporal. Esta formulação reconhece que desempenho é acurácia multiplicada por pontualidade.

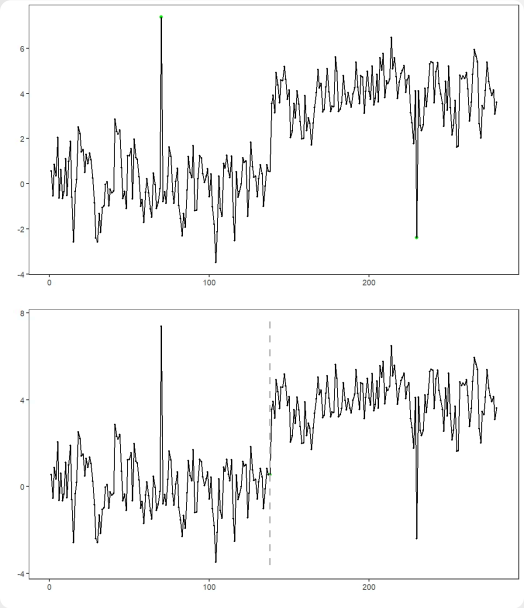
Anomalias versus Pontos de Mudança

Anomalias representam desvios locais, tipicamente manifestados em resíduos: $|r_t| > \tau$. São eventos pontuais que não alteram a estrutura subjacente da série.

Change points marcam mudanças persistentes de parâmetros: $\theta_1 \neq \theta_2$. São transições estruturais que definem regimes distintos antes e depois do ponto de mudança.

A interseção pode existir: $\mathcal{A} \cap \mathcal{C} \neq \varnothing$. Mudanças de regime frequentemente produzem resíduos grandes perto da transição, então um detector pode sinalizar anomalias na vizinhança de um change point.

Porém, os conceitos não são equivalentes: $\mathcal{A} \neq \mathcal{C}$. O alvo conceitual difere fundamentalmente.



Anomalias, Motifs e Discords

Motifs

Padrões recorrentes com alta similaridade: $s_i \approx s_j$. Representam comportamentos que se repetem ao longo da série.

Discords

Padrões raros com alta distância: $s^* = \arg\max_{s \in \mathcal{S}} \min_{u \neq s} d(s, u)$. Subsequências isoladas no espaço de padrões.

Anomalias Pontuais

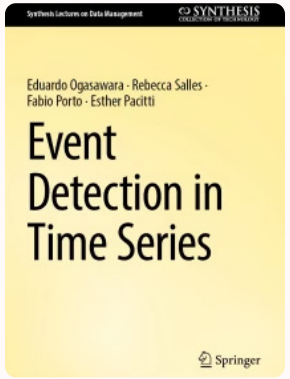
Desvios locais em valores ou resíduos em instantes específicos, sem considerar contexto de janela.

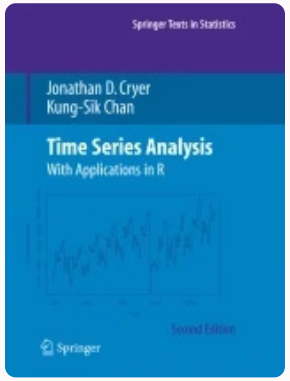
Quando a "anomalia" não é um ponto isolado, mas um padrão raro ao longo de uma janela, ela se aproxima do conceito de discord. Isso justifica a leitura de discords como anomalias coletivas: $\mathcal{A}_{coll} \approx \mathcal{D}$.

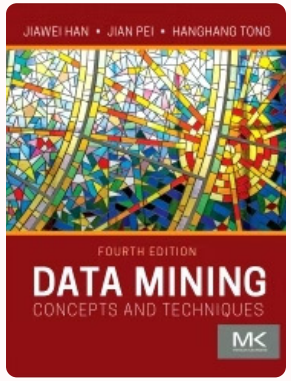
RECURSOS ACADÊMICOS

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