

# Discords em Séries Temporais

Padrões raros e isolados que se destacam na série temporal.

Eduardo Ogasawara

[eduardo.ogasawara@cefet-rj.br](mailto:eduardo.ogasawara@cefet-rj.br)

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

# Definição de Discord

Um discord é uma subsequência "isolada" que se destaca das demais na série temporal. Diferente de uma anomalia pontual, o discord é definido no espaço de subsequências, não em um único ponto.

Considere a série X. Um discord é uma subsequência cuja forma é muito diferente das outras subsequências extraídas da mesma série. Ele mede raridade comparando janelas temporais entre si, revelando padrões temporais excepcionais.

# Anomalias Pontuais vs Discords

## Anomalia Pontual

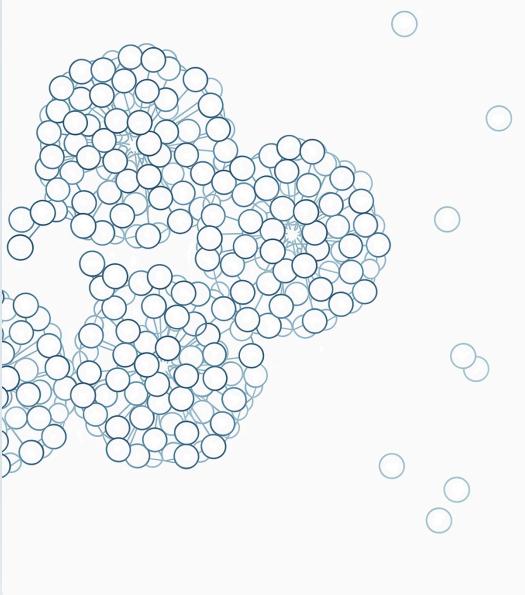
Desvio grande em um único instante. Detectada quando o resíduo  $|r_t|$  excede um limiar  $\tau$ .

## Discord

Desvio grande de uma janela inteira em relação a outras janelas. Envolve distância  $d(X_{i:i+w}, X_{j:j+w})$  entre subsequências.

A diferença fundamental está na unidade do evento: anomalia pontual olha para um ponto; discord olha para um trecho completo. Discords podem ser vistos como anomalias coletivas, pois o padrão inteiro da janela é raro.

# Dualidade: Motifs e Discords



## 1 Motifs

Padrões frequentes com muitas janelas próximas. Definidos por pequenas distâncias:  
 $d(s_i, s_j) < \epsilon$

## 2 Discords

Padrões raros com janelas isoladas. A subsequência fica o mais distante possível do restante.

Motifs e discords são conceitos opostos no mesmo espaço: motifs são regiões densas de padrões repetidos; discords são pontos isolados, com pouca ou nenhuma vizinhança semelhante. Discords vivem onde motifs não vivem.

# Discords vs Pontos de Mudança



Change points são definidos por mudança de parâmetros e segmentam a série em regimes. Discords são definidos por raridade no espaço de subsequências e destacam exceções dentro dos dados.

Um change point é uma afirmação sobre o processo: "o regime mudou aqui". Já um discord é uma afirmação sobre padrões: "este trecho é muito diferente dos outros". Discords podem aparecer dentro de regimes estáveis como exceções.

# Subsequências como Base

## Construindo o Espaço $S$

Recortamos janelas de tamanho  $w$  da série temporal. Cada janela  $X_{t:t+w-1} = (x_t, x_{t+1}, \dots, x_{t+w-1})$  vira um objeto  $s_t$  no conjunto  $S$ .

O conjunto  $S$  contém todas as subsequências extraídas, com  $t$  variando de 1 até  $T - w + 1$ . Discords são definidos comparando objetos dentro deste espaço.

# Medindo Isolamento: Distância Mínima ( $D(s_i)$ )

Para cada subsequência  $s_i$ , procuramos o vizinho mais próximo. A distância mínima  $D(s_i)$  mede o isolamento local dessa janela.

01

## Calcular $D(s_i)$

$$D(s_i) = \min_{s_j \in S, j \neq i} d(s_i, s_j)$$

02

## Interpretar o Valor

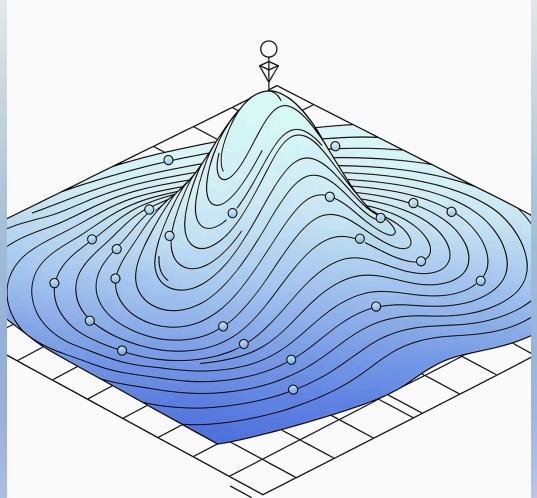
Grande distância mínima indica candidato a discord

03

## Identificar Padrões Raros

Se nem o melhor vizinho é parecido,  $s_i$  está isolada

A ideia é elegante: em vez de comparar  $s_i$  com todas as outras, perguntamos "qual é a coisa mais parecida com  $s_i$ ?". Se a resposta for "nada muito parecido", temos um padrão raro.



# Definição Matemática de Discord

O discord principal é a subsequência que maximiza a distância mínima:  $s^* = \operatorname{argmax}_{s_i \in S} D(s_i)$

1

## Discord Principal

Subsequência mais isolada, com maior  $D(s_i)$

2

## Top-k Discords

Múltiplos discords ordenados por  $D(s_i)$  decrescente

3

## Critério de Raridade

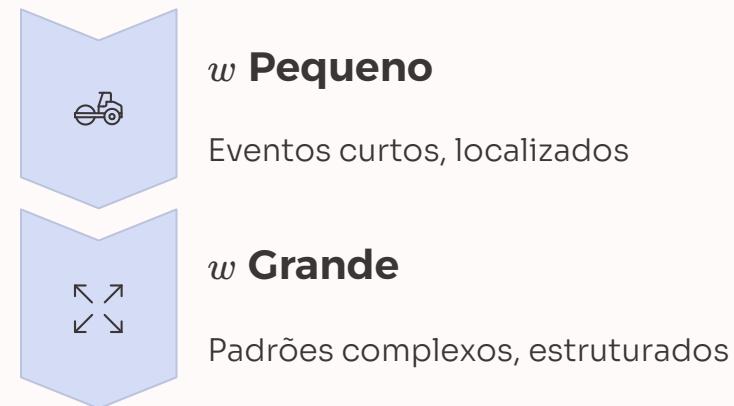
Distância ao vizinho mais próximo define isolamento

# O Papel Crucial do Parâmetro $w$

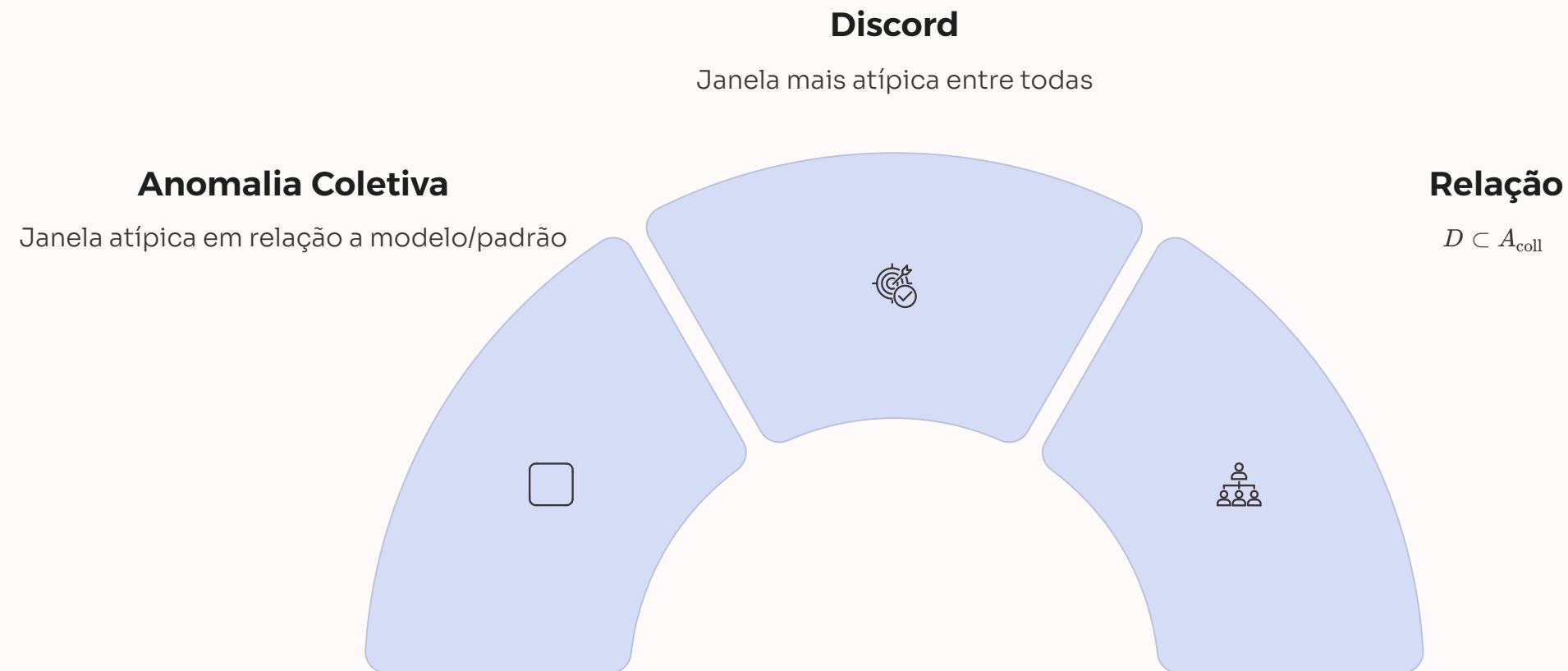
## Escala do Padrão Raro

O tamanho da janela  $w$  define a escala temporal do "padrão raro" que buscamos. Janelas pequenas aproximam discords de eventos curtos e localizados.

Janelas grandes capturam padrões raros mais complexos e estruturados. A escolha de  $w$  muda fundamentalmente o conjunto de discords:  $D(w) \neq D(w')$ .

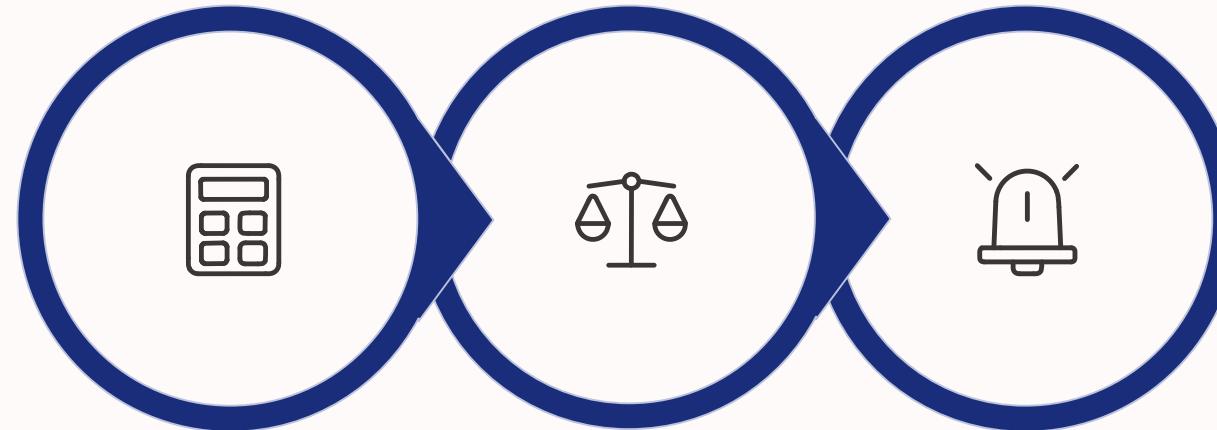


# Discords como Anomalias Coletivas



Discord é um caso extremo de atipicidade em subsequências. Uma anomalia coletiva já é um trecho estranho como padrão; o discord é o mais isolado entre esses padrões. Por isso faz sentido enxergar discords como "anomalias coletivas extremas".

# Detecção Baseada em Distância



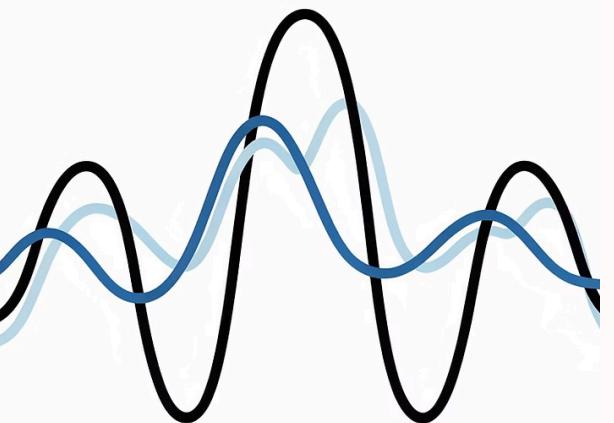
**Calcular  
 $g(s_i)$**

**Comparar  
com  $\tau$**

**Declarar  
discord**

O score de discord é definido como  $g(s_i) = \min_{s_j \in S, j \neq i} d(s_i, s_j)$ . A regra de decisão é simples: declaramos discord se o score passar de um limiar  $\tau$ .

Esta é uma visão geométrica: tratamos cada janela como um ponto no espaço de subsequências, e discords são outliers - pontos sem vizinhos próximos.



# Matrix Profile: Estrutura Unificadora

O Matrix Profile guarda, para cada subsequência, a menor distância para qualquer outra:

$$MP(i) = \min_{j \neq i} d(s_i, s_j)$$

## Maiores Valores

Indicam discords - janelas sem vizinhos próximos

## Menores Valores

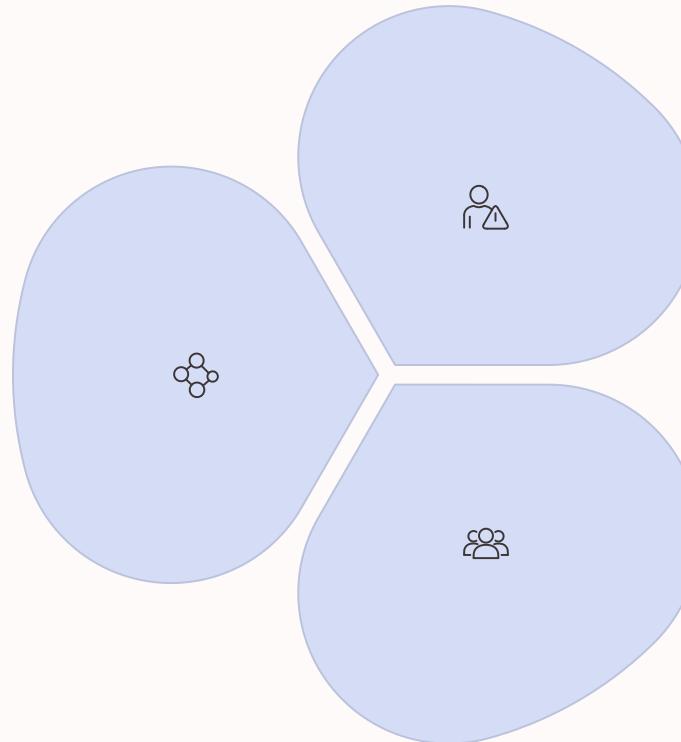
Indicam motifs - janelas com muitas cópias similares

O Matrix Profile unifica os dois conceitos com o mesmo objeto matemático, resumindo para cada janela quão parecida ela é com sua melhor "cópia" na série.

# Abordagem por Clustering

## Regiões Densas

Clusters  $C_k$  representam motifs



## Pontos Isolados

$s_i \notin \cup_k C_k$  são discords

## Agrupamento

Aplicado ao conjunto  $S$

Tratamos subsequências como pontos e procuramos grupos. Discords aparecem como "outliers" de clustering: o que sobra fora dos grupos densos tende a ser raro e isolado. É a mesma ideia de distância, mas com uma camada de estrutura.

# Abordagem Probabilística

---

## Representação

$$Z_t = R(X_{t:t+w})$$

---

## Modelo de Densidade

Ajustar  $p(Z_t)$

---

## Detecção

Discord se  $p(Z_t) < \varepsilon$

## Raridade Probabilística

Em vez de medir distância, medimos "quão improvável" é o padrão. A janela vira um vetor  $Z_t$  e o modelo aprende onde estão as regiões típicas.

Discords aparecem onde o modelo atribui baixa probabilidade - padrões raros segundo a dinâmica aprendida.

# Complexidade Computacional

  $O(T^2)$ 

## Abordagem Ingênua

Comparar todas as janelas entre si

O número de janelas cresce com  $T$  e  $w$ . Comparar todas contra todas é computacionalmente caro, com custo quadrático no pior caso. Por isso a literatura enfatiza otimizações e estruturas que tornam o problema mais escalável.

  $O(T \log(T))$ 

## Métodos Otimizados

Estruturas eficientes como Matrix Profile

# Avaliação Formal de Discords

Comparamos discords reais  $D$  com estimados  $\hat{D}$  usando métricas padrão, incluindo tolerância temporal  $\epsilon$  para casar janelas.

## Precisão

$$|\hat{D} \cap D| / |\hat{D}|$$

Quantos detectados eram realmente discords

## Revocação

$$|\hat{D} \cap D| / |D|$$

Quantos discords reais foram encontrados

## F1-Score

$$2 \cdot \frac{(\text{Precisão} \cdot \text{Revocação})}{(\text{Precisão} + \text{Revocação})}$$

Média harmônica balanceada

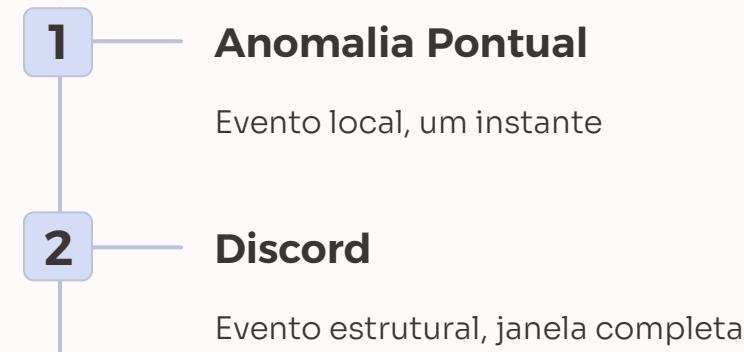
Como discords são intervalos, a avaliação precisa de uma regra de "casamento" temporal: uma janela estimada conta como acerto se estiver suficientemente alinhada com uma janela verdadeira, dentro de  $\epsilon$ .

# Discords como Anomalias Estruturais

## Generalização do Conceito

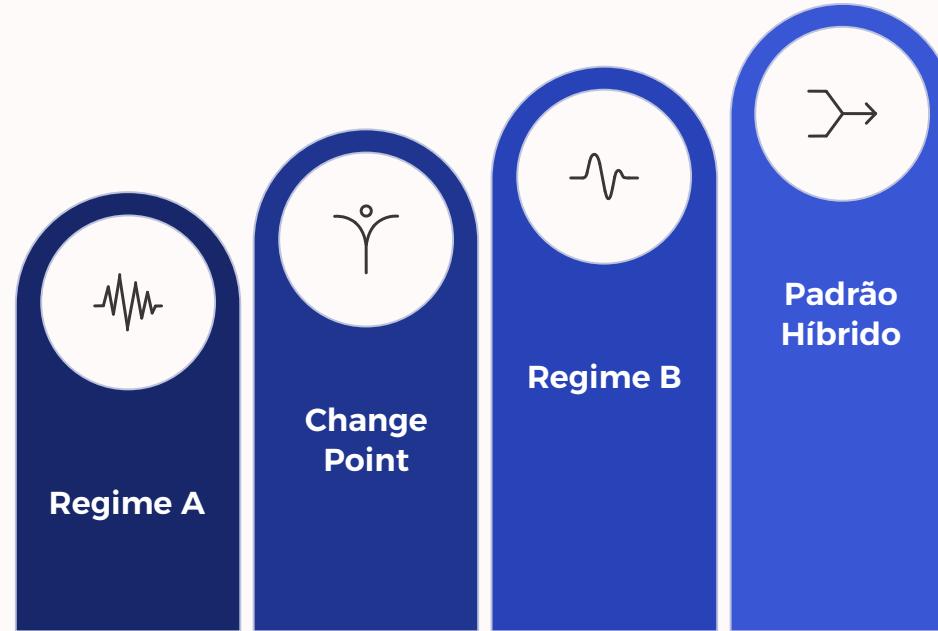
Anomalia pontual é um evento local:  $|r_t| > \tau$ . Discord é um evento estrutural em janela:  $D(s_i)$  grande.

Discords generalizam anomalia para subsequências, formando o subconjunto  $D \subset A_{\text{coll}}$  das anomalias coletivas.



Se anomalia é "fora do típico", discords são "fora do típico" quando o objeto é uma janela. O que é raro é o formato do trecho, não apenas um valor isolado.

# Interação com Rupturas Estruturais



Janelas que atravessam um change point misturam regimes, o que pode aumentar a dissimilaridade dessas janelas. Se  $\tau \in [t, t + w]$ , então  $D(X_{t:t+w})$  tende a aumentar.

Perto de um change point, algumas janelas "pegam um pedaço de cada regime", produzindo um padrão híbrido que não se parece nem com janelas do regime anterior nem do posterior. Isso faz a janela aparecer como discord - um efeito colateral de mudanças estruturais.

# 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.