



XXXIX

Congresso da Sociedade Brasileira de Computação

14 a 18 de Julho | Centro de Convenções da Amazônia | Belém - PA

Realizado por:



Sociedade Brasileira
de Computação

8° Brazilian Workshop on Social Network Analysis and Mining

Não-estacionariedade: Quando tempo e espaço interferem



CEFET/RJ

Eduardo Ogasawara

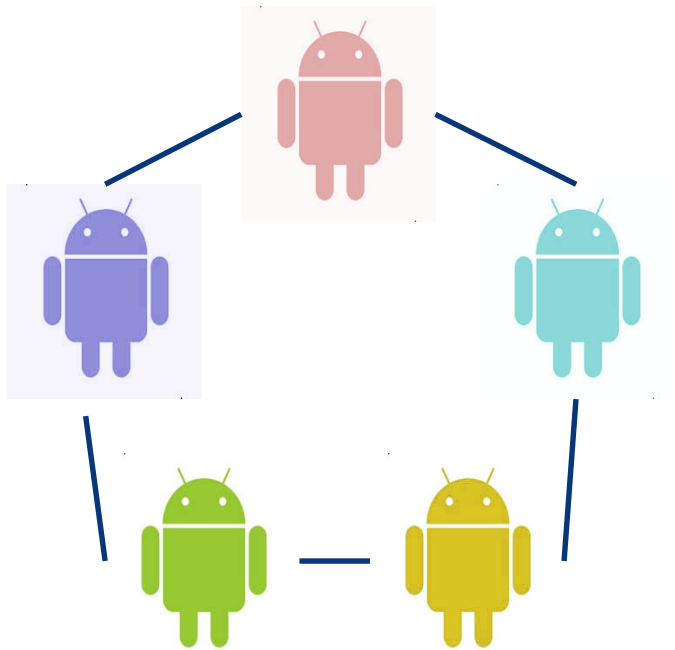
eogasawara@ieee.org

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

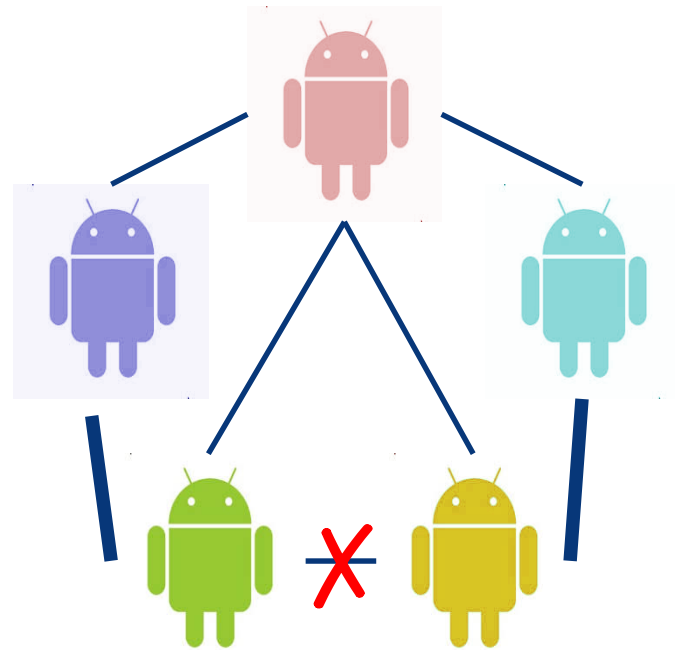
Não-estacionariedade

- Estacionariedade
 - Dataset D
 - Amostras D_s
 - Propriedades estatísticas em D_s não variam com o tempo
 - Séries temporais: média, variância e covariância
- Não-estacionariedade
- Métodos de *data analytics*
 - A grande maioria dos métodos assumem “implicitamente” estacionariedade

Dinâmica evolutiva das redes sociais

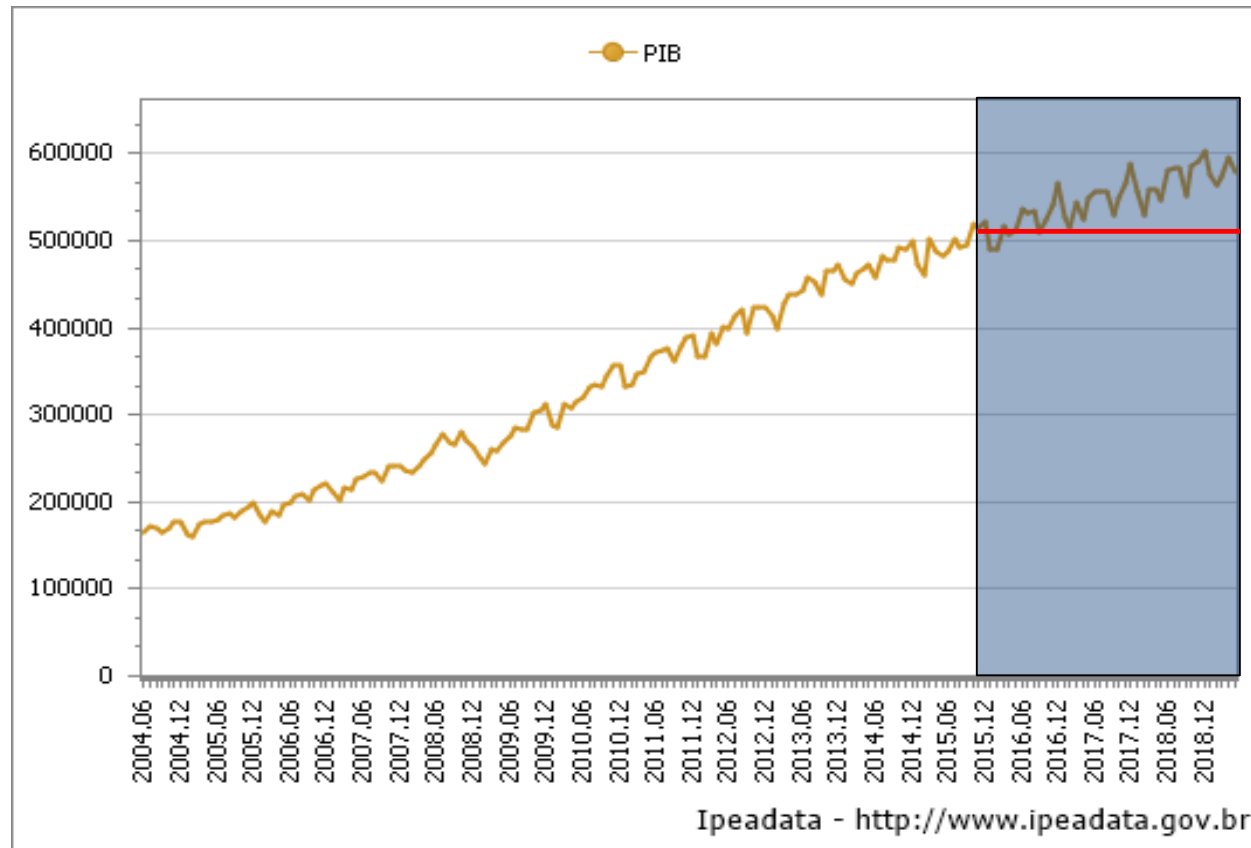


t



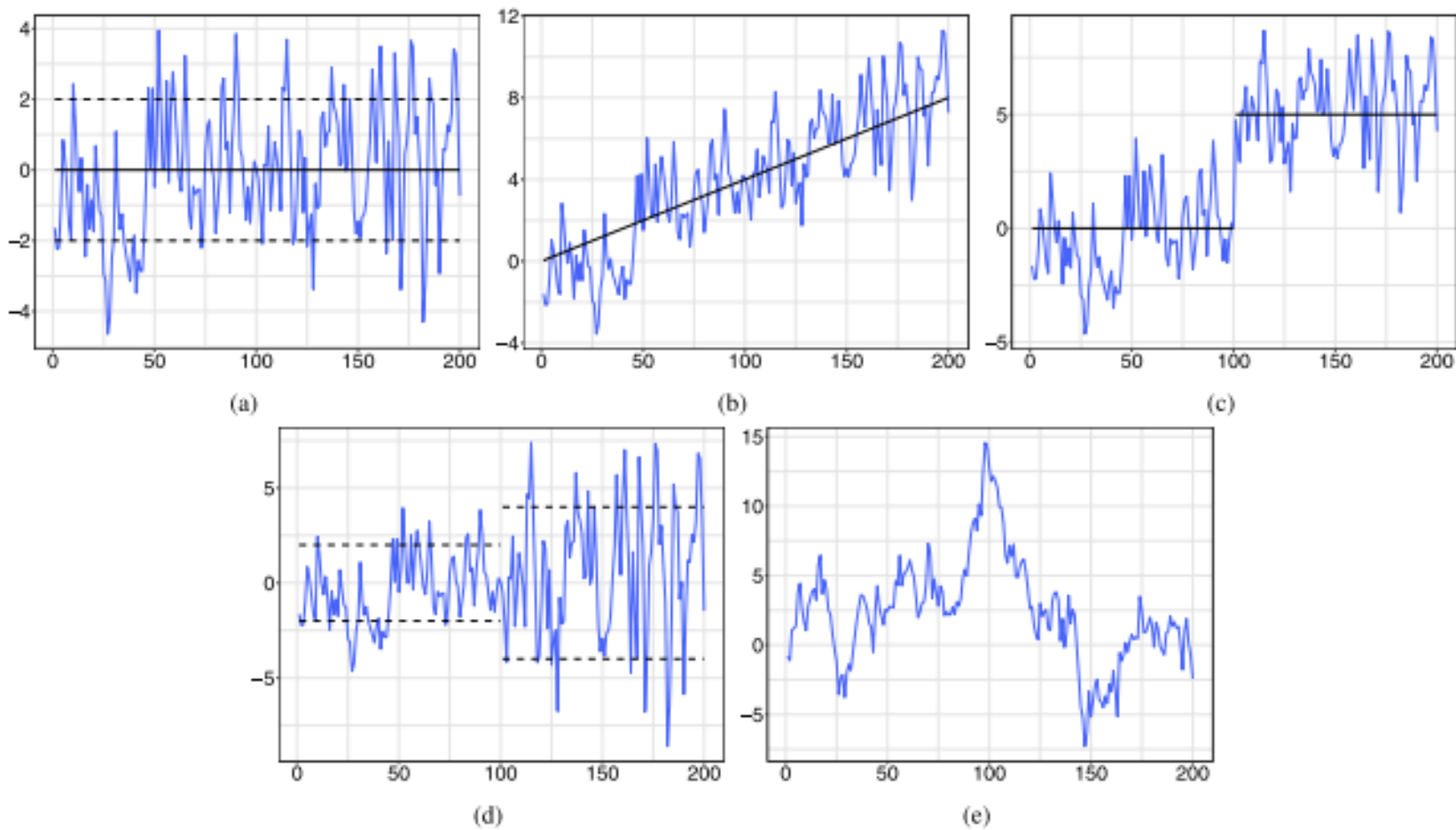
u

Predição de séries temporais



Normalização (min/max)

Tipos de não-estacionariedade



As múltiplas faces da não-estacionariedade

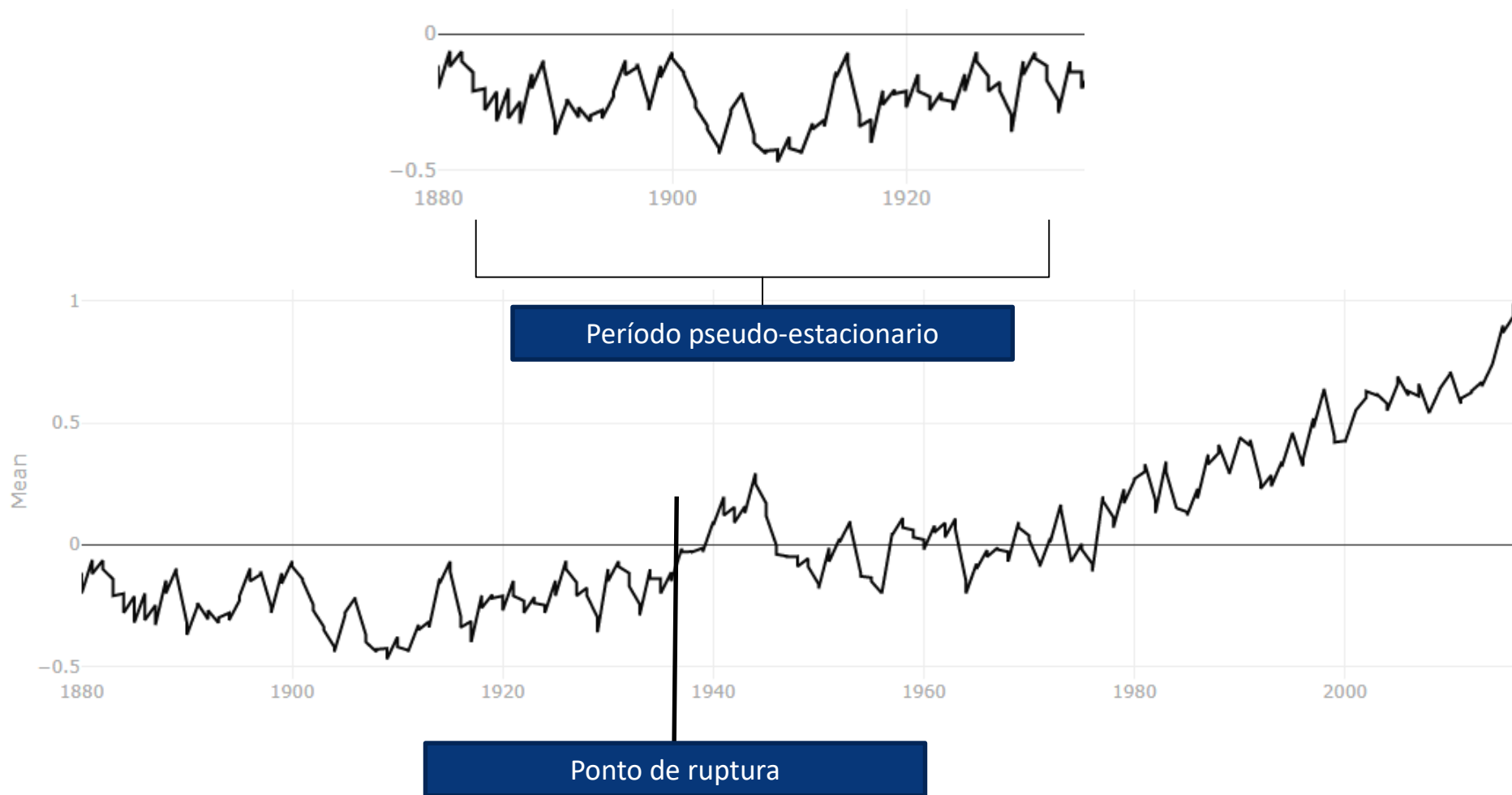
- Crítica de Lucas
 - visão econômica/estatística
- Dilema da Plasticidade e Estabilidade
 - visão de aprendizado de máquina
- Mudança de conceito (*Concept Drift*)
 - visão de Mineração de Dados
- Padrões emergentes (*Emerging patterns*)
 - visão de Banco de Dados

Crítica de Lucas

- “Dado que a estrutura de um modelo econométrico consiste em regras de decisão ótimas dos agentes econômicos, e que as regras de decisão ótimas variam sistematicamente com as mudanças na estrutura das séries relevantes para o decisor, conclui-se que qualquer mudança na política sistematicamente irá alterar a estrutura dos modelos econométricos”

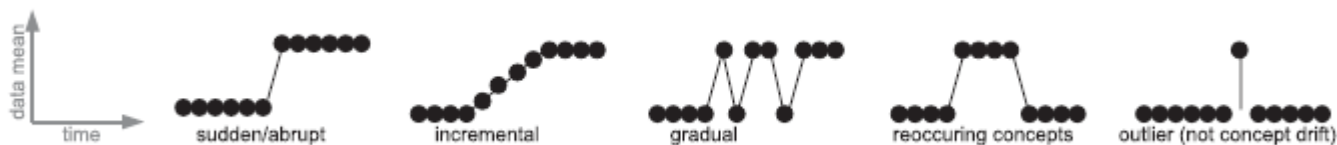
Pseudo-estacionariedade

Monthly mean temperature anomalies in degrees Celsius relative to a base period



Dilema da Plasticidade e Estabilidade

- Redes neurais são conhecidas pela adaptabilidade
 - Capacidade de atualizar os pesos em função de alterações no ambiente
 - Treinamento incremental
 - Alteração dos pesos sinápticos
- Sistemas adaptativos visam abordar não-estacionariedade
 - Buscando-se robustez, adota-se adaptabilidade
 - Maior adaptabilidade, mais suscetível a situações espúrias, menor robustez
 - Dilema: encontrar o tempo certo para se adaptar



[1] S.O. Haykin, 2008, *Neural Networks and Learning Machines*. 3 ed. New York, Prentice Hall.

[2] Grossberg, S., 1988. *Neural Networks and Natural Intelligence*, Cambridge, MA: MIT Press.

[3] G. Ditzler, M. Roveri, C. Alippi, e R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.

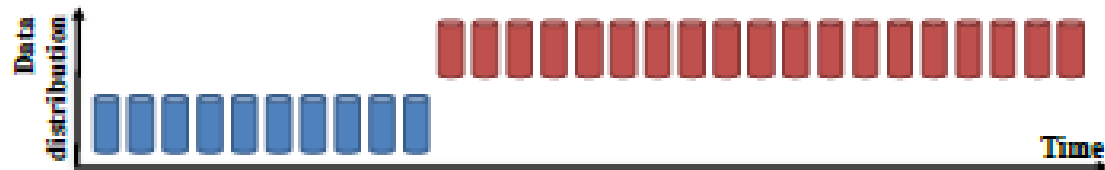
Concept Drift

- Aprendizado no contexto de distribuições não-estacionárias
 - Aprendizado é feito em lotes (*batches*)
 - *Data streams* (objetos com *timestamps*)
 - Definições
 - $P(Y)$ probabilidade da variável dependente (rótulo)
 - $P(X)$ probabilidade das variáveis independentes (objetos)
 - $P(X, Y)$ probabilidade conjunta dos objetos e rótulo
 - $P(Y|X)$ distribuição provável do rótulo para objeto
 - Concept = $P(X, Y) = P(\chi)$
 - Drift = $P_t(\chi) \neq P_u(\chi)$

Tipos de Concept Drift

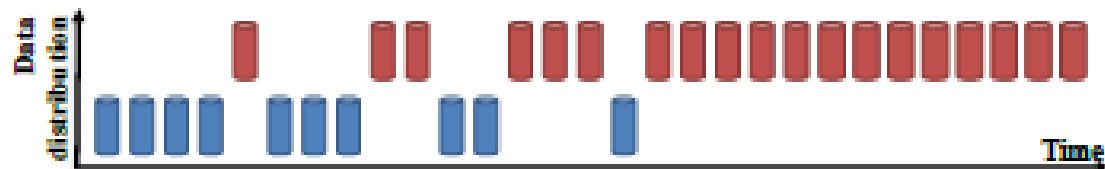
Sudden Drift:

A new concept occurs within a short time.



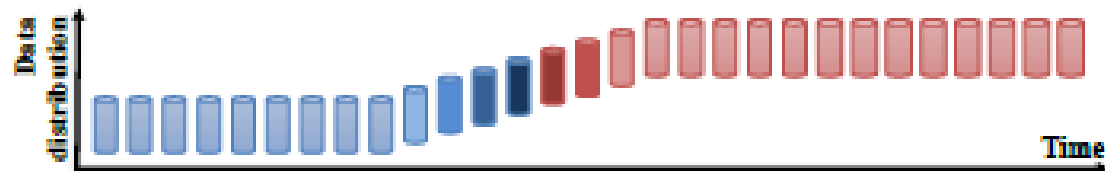
Gradual Drift:

A new concept gradually replaces an old one over a period of time.



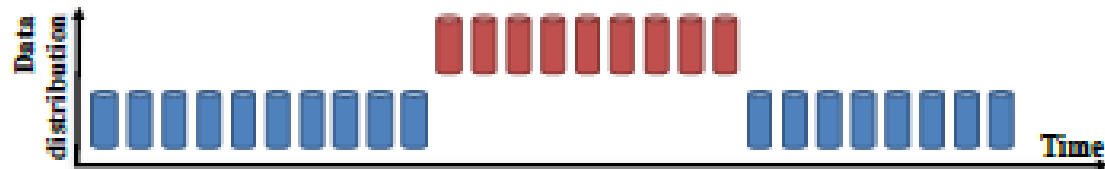
Incremental Drift:

An old concept incrementally changes to a new concept over a period of time.



Reoccurring Concepts:

An old concept may reoccur after some time.



Taxonomia de não-estacionariedade



Magnitude e Real/Virtual Concept Drift

- Magnitude do *Concept Drift*: $D(t, u)$
- *Real Concept Drift*
 - $P_t(Y|X) \neq P_u(Y|X)$ e $P_t(X) = P_u(X)$
- *Virtual Concept Drift*
 - $P_t(Y|X) = P_u(Y|X)$ e $P_t(X) \neq P_u(X)$

Emerging patterns

- Padrões emergentes são coleções de itens cuja frequência muda de um *dataset (batch)* para outro
- *Datasets* D_t (anterior) e D_u (próximo)
 - Crescimento para itens χ : $\rho(\chi)$

$$\rho(\chi) = \begin{cases} \infty, & \text{support}_t(i) = 0 \\ 0, & \text{support}_t(\chi) = \text{support}_u(\chi) = 0 \\ \frac{\text{support}_u(\chi)}{\text{support}_t(\chi)}, & \text{otherwise} \end{cases}$$

- Dado um limite σ , um padrão χ é emergente se $\rho(\chi) \geq \sigma$

[1] G. Dong e J. Li, 1999, Efficient Mining of Emerging Patterns: Discovering Trends and Differences, In: *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 43–52

[2] K. Ramamohanarao e J. Bailey, 2004, Emerging Patterns: Mining and Applications, In: *Proceedings of International Conference on Intelligent Sensing and Information Processing, ICISIP 2004*, p. 409–414.

Abordagens para não-estacionariedade

- Gerência da memória
- Adaptabilidade
- Transformações

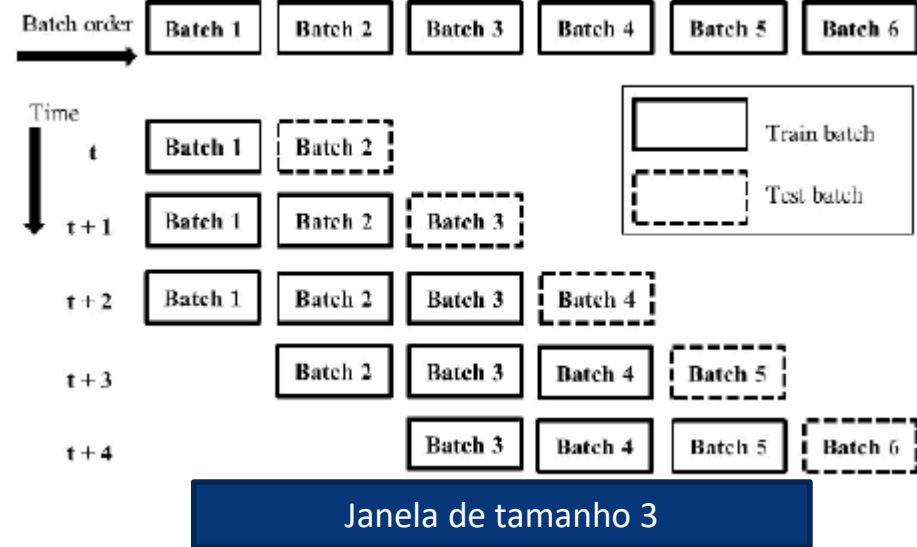
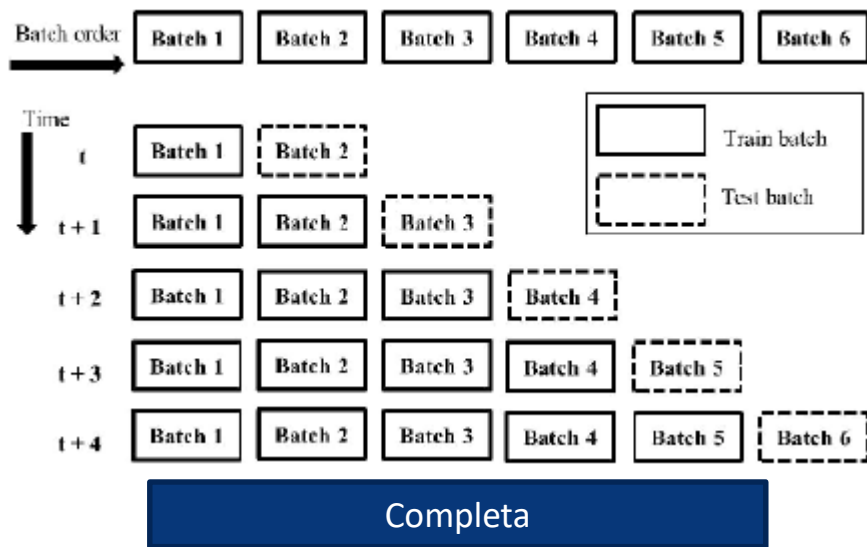
[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

[3] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, e E. Ogasawara, "Nonstationary time series transformation methods: An experimental review", *Knowledge-Based Systems*, nov. 2018.

Memória

- Processo
 - Testa-se no último batch (previsão)
 - Incorpora-se último batch no treino
- Memória
 - Completa
 - Sem memória
 - Janelas deslizantes

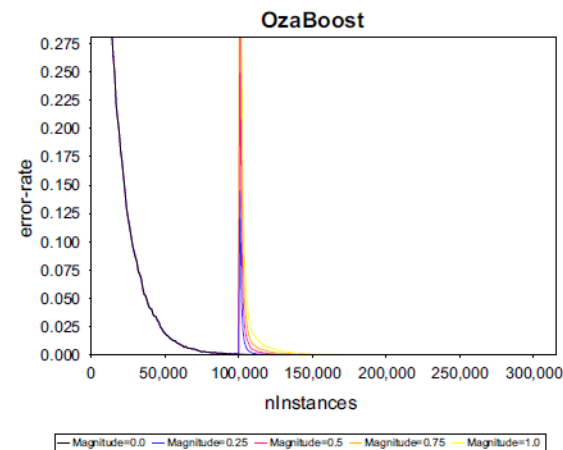
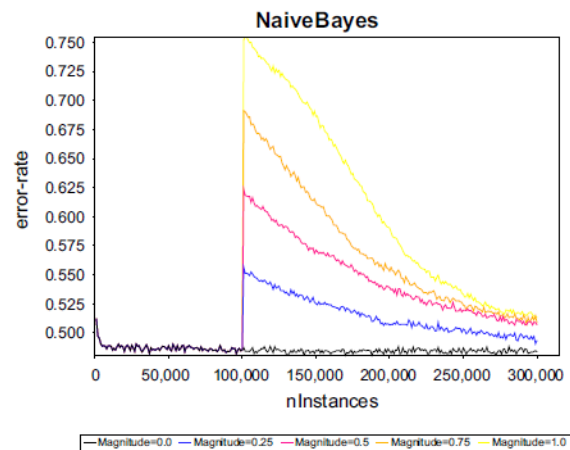


[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

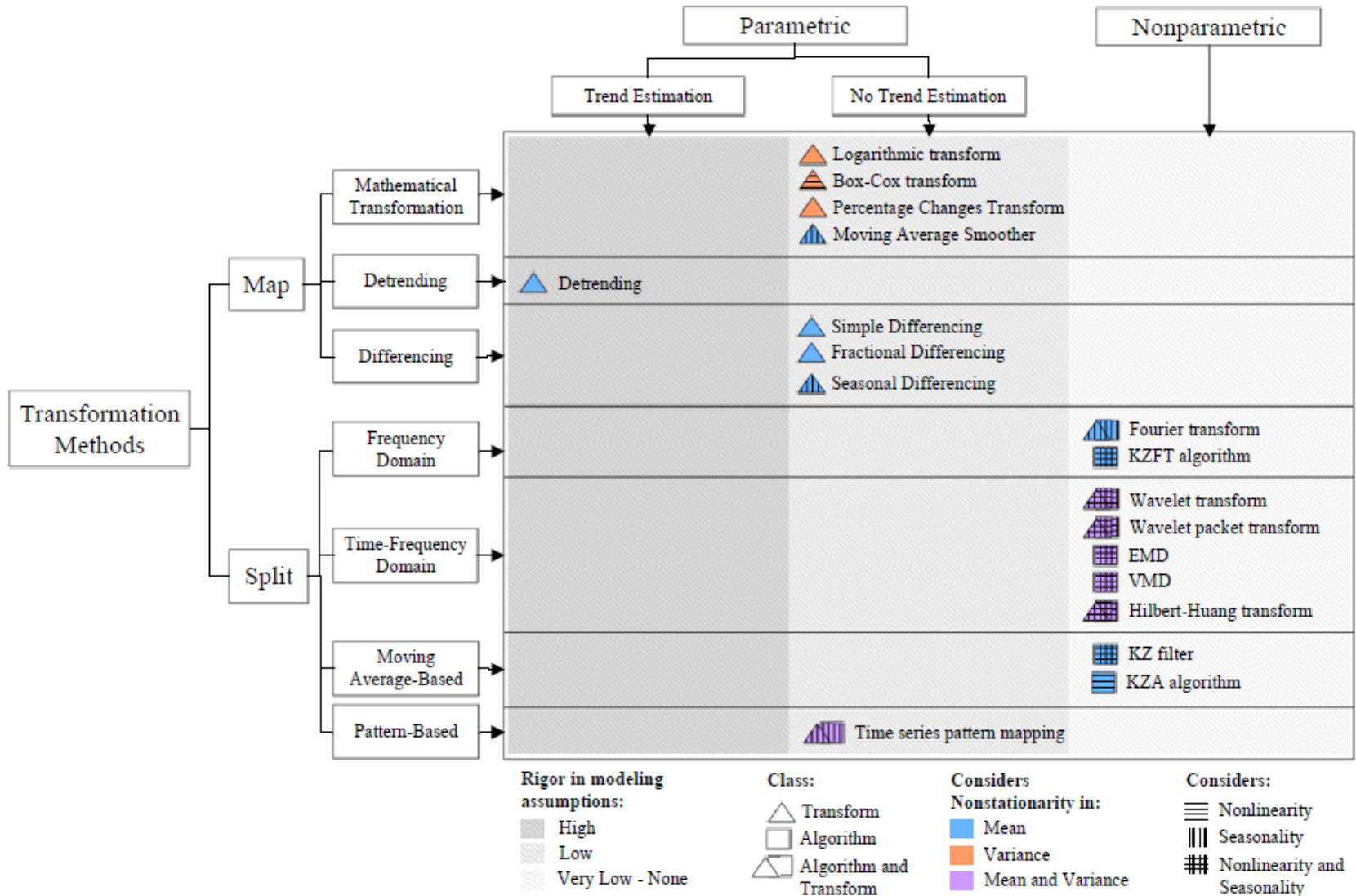
Adaptabilidade

- Detecção de *drift*
 - Ativa
 - Passiva
- Aprendizado
 - Incremental
 - Não-incremental
- Modelos
 - Único
 - Ensemble

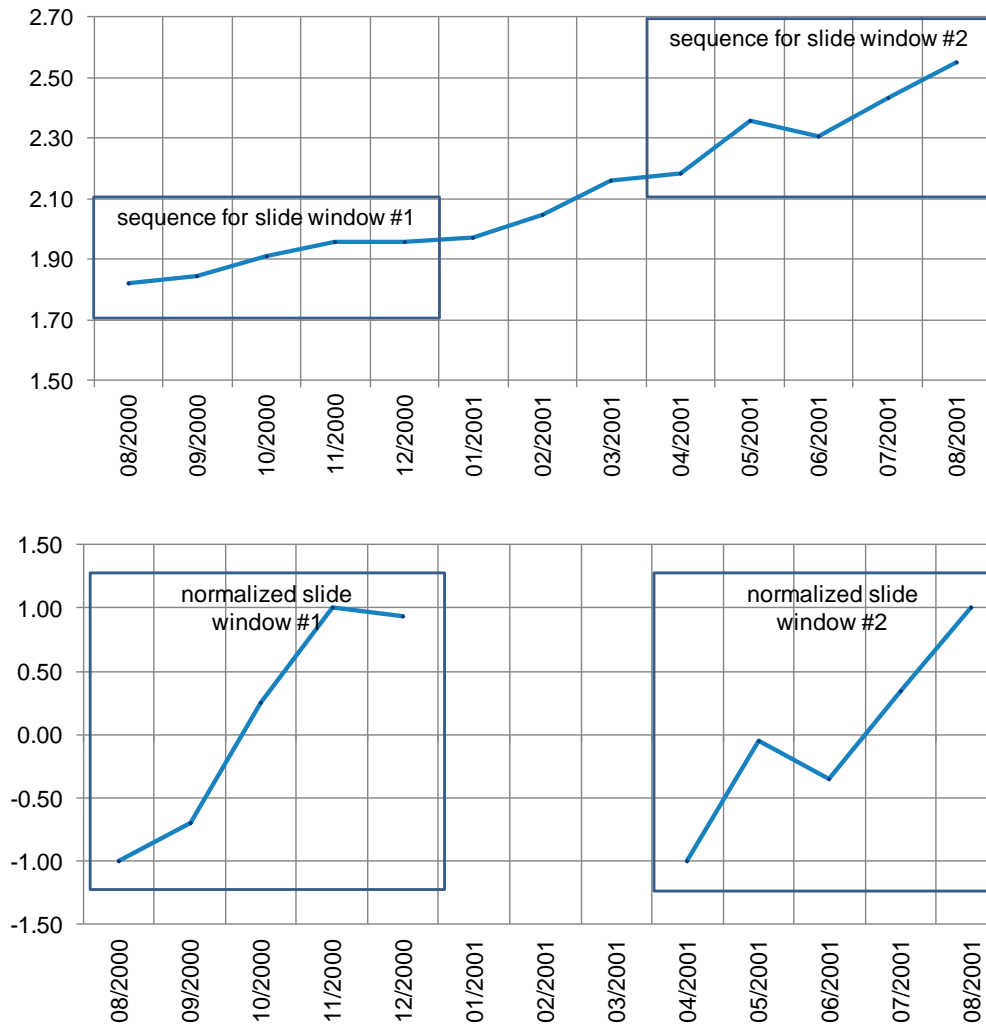


- [1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4
- [2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1
- [3] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, e F. Petitjean, 2016, Characterizing concept drift, *Data Mining and Knowledge Discovery*, v. 30, n. 4, p. 964–994.

Transformações



Problemas de normalização usando janelas deslizantes



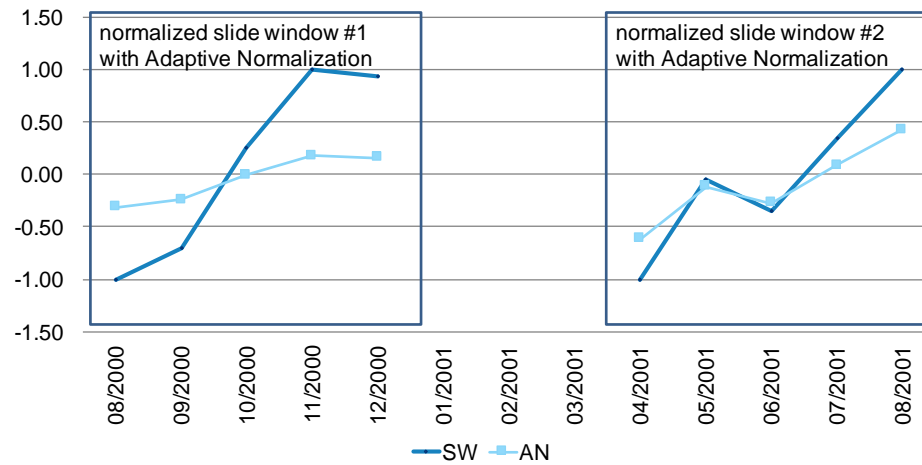
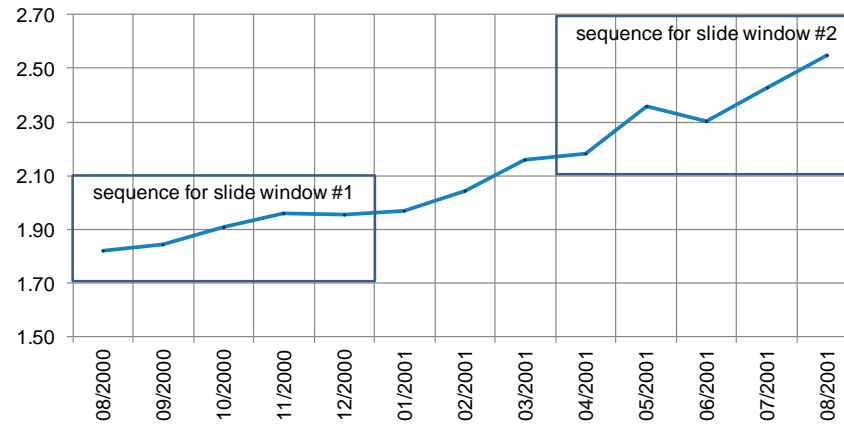
Monthly average exchange rate of U.S. Dollar to Brazilian Real
normalized by sliding window technique from aug/2000 to dec/2000 and from apr/2001 to aug/2001

Funções inerciais

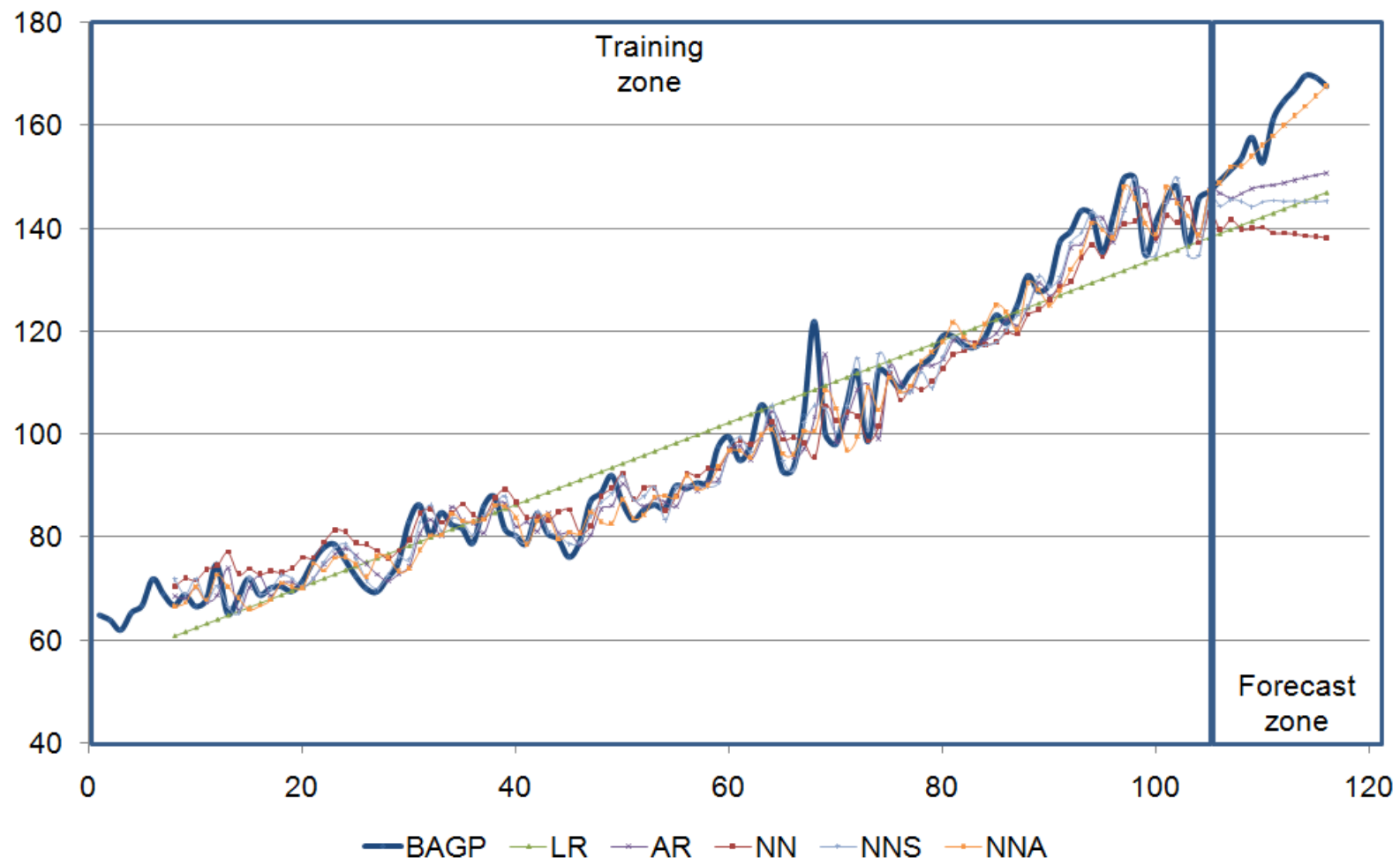
Normalização adaptativa

- Transformação
 - Conversão da série em janelas deslizantes
 - Cálculo de inércia da janela
 - Diferenciação das observações pela inércia
 - Análise da distribuição e remoção de outliers
 - Normalização
- Desconversão
 - Predição
 - Desnormalização
 - Adição da inércia

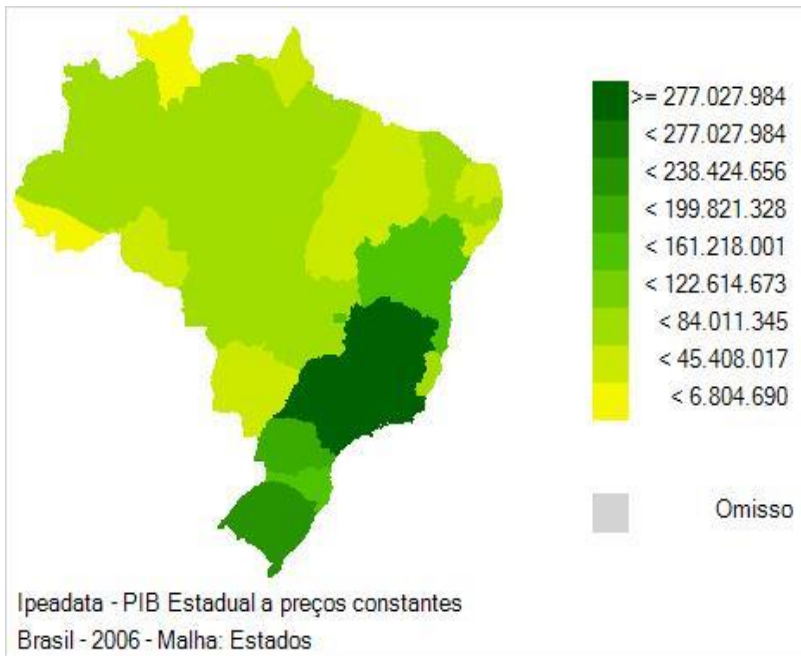
Intuição



Predição de séries temporais usando aprendizado de máquina



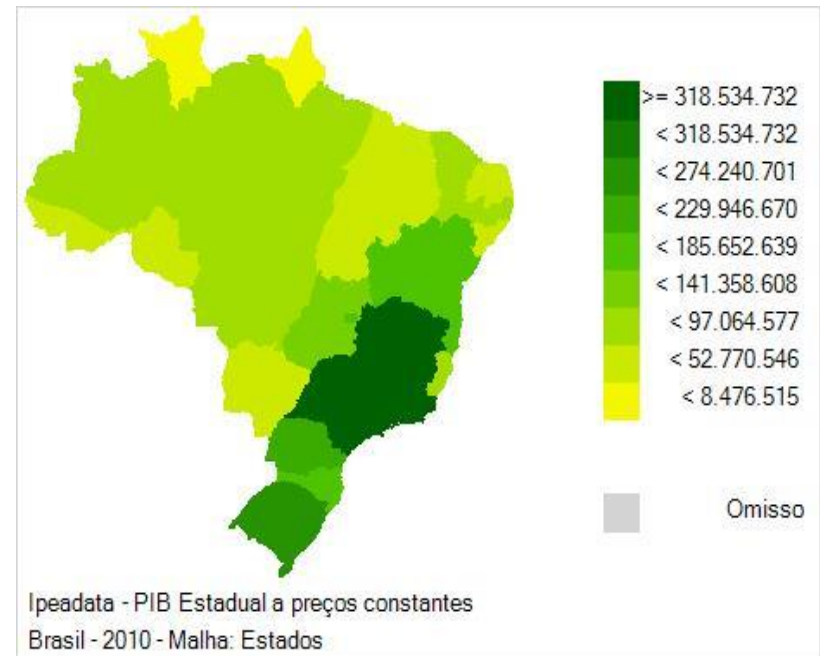
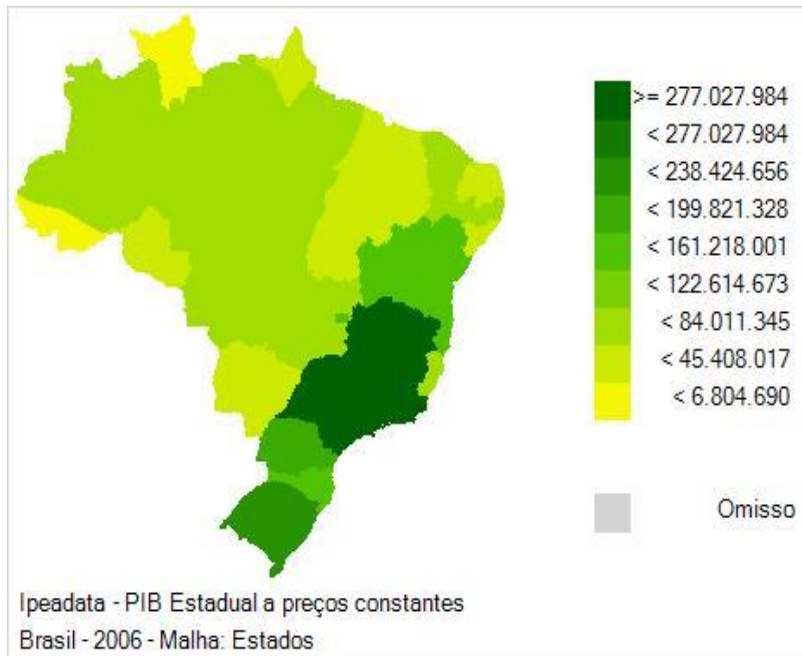
Aspectos espaço-temporais



Sigla	Estado	2006	2010
AC	Acre	6.633.867	8.476.515
AL	Alagoas	20.802.615	24.574.808
AM	Amazonas	50.816.007	59.779.292
AP	Amapá	6.804.690	8.265.965
BA	Bahia	131.479.024	154.340.458
CE	Ceará	64.306.577	77.865.415
DF	Distrito Federal	125.765.530	149.906.319
ES	Espírito Santo	66.563.030	82.121.834
GO	Goiás	78.044.303	97.575.930
MA	Maranhão	37.195.271	45.255.942
MG	Minas Gerais	302.431.433	351.380.905
MS	Mato Grosso do Sul	34.311.309	43.514.207
MT	Mato Grosso	46.453.960	59.599.990
PA	Pará	69.415.228	77.847.597
PB	Paraíba	26.429.318	31.947.059
PE	Pernambuco	77.462.625	95.186.714
PI	Piauí	17.958.480	22.060.161
PR	Paraná	179.844.892	217.289.677
RJ	Rio de Janeiro	354.234.639	407.122.794
RN	Rio Grande do Norte	28.261.660	32.338.895
RO	Rondônia	17.978.571	23.560.644
RR	Roraima	5.007.053	6.340.601
RS	Rio Grande do Sul	217.001.407	252.482.597
SC	Santa Catarina	132.634.660	152.482.338
SE	Sergipe	19.962.748	23.932.155
SP	São Paulo	1.024.208.900	1.247.595.927

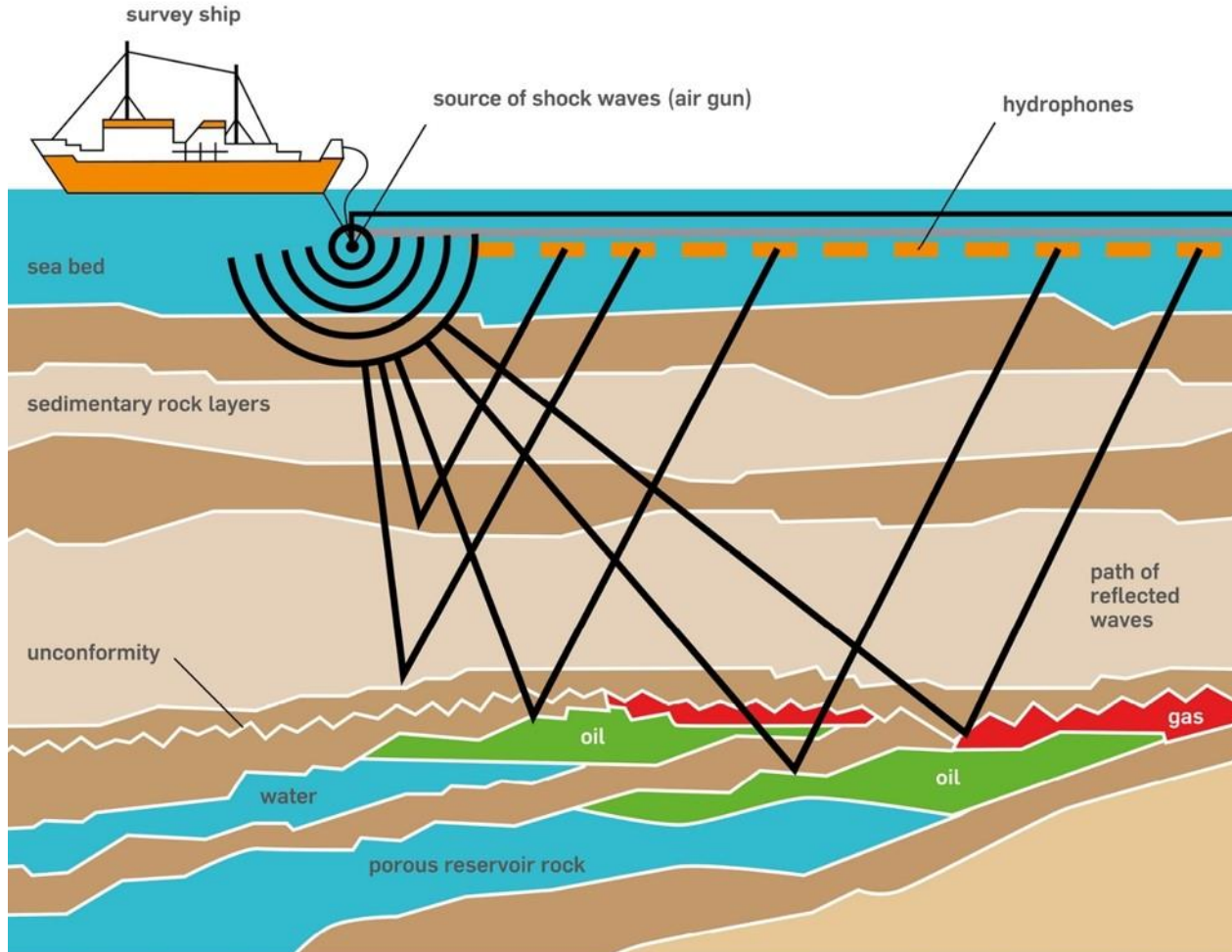


Aspectos espaço-temporais (funções inerciais)



Análise na sísmica

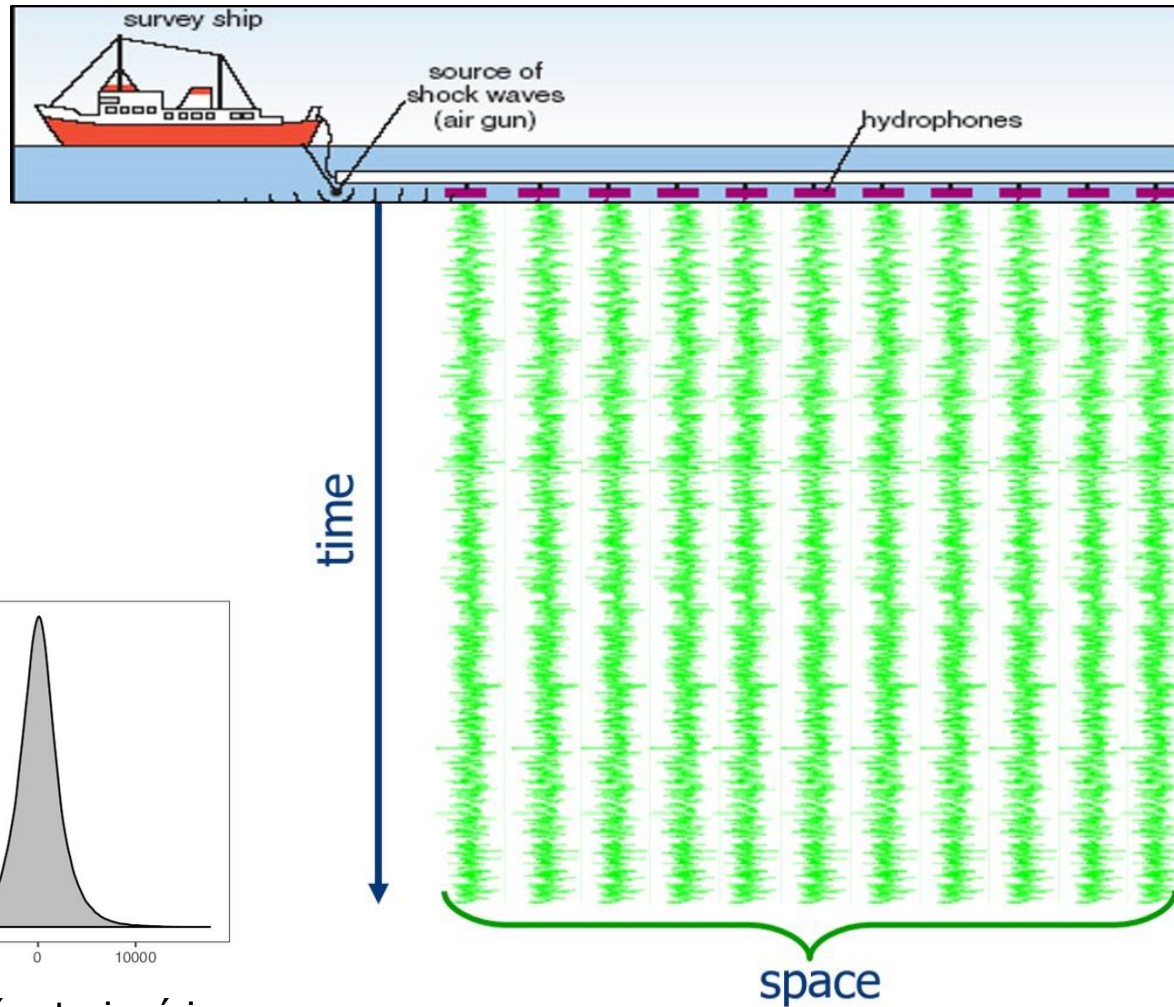
Séries espaço-temporais têm uma posição associadas a sensores



Source: <https://krisenergy.com/company/about-oil-and-gas/exploration/>

Análise na sísmica

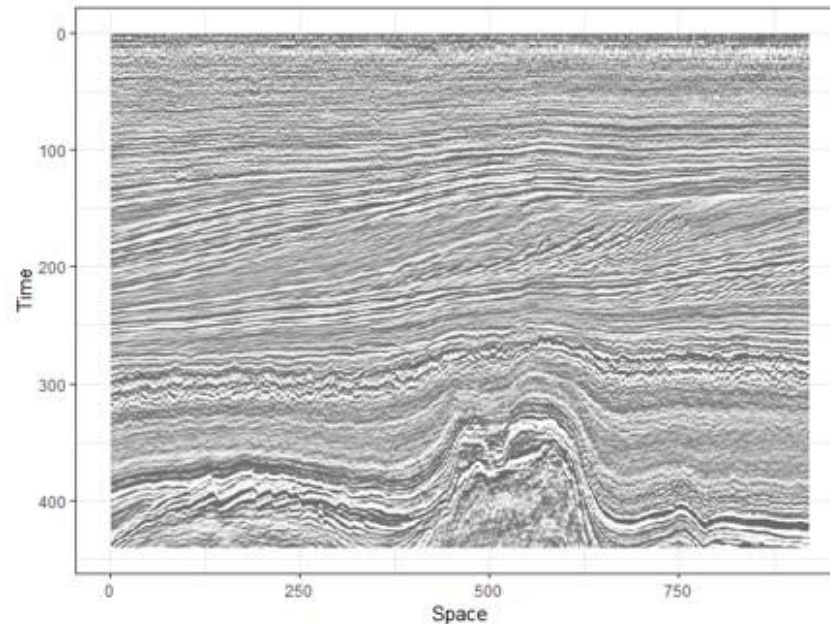
Cada sensor está associada a uma série espaço-temporal



Cada série é estacionária

Não-estacionariedade no espaço-tempo

- Probabilidades diferentes no tempo-espaço
- Modelos especializados para regiões



Desafios

- Aprendizado semi-supervisionado /não-supervisionado
- Estruturas complexas/heterogêneas
- Streaming
- Drifts transientes
- Classes desbalanceadas
- Big Data e Small Data
- Frameworks teóricos

[1] G. Ditzler, M. Roveri, C. Alippi, e R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.

[2]. Lu, A. Liu, F. Dong, F. Gu, J. Gama, e G. Zhang, 2018, Learning under Concept Drift: A Review, *IEEE Transactions on Knowledge and Data Engineering*.



XXXIX

Congresso da Sociedade Brasileira de Computação

14 a 18 de Julho | Centro de Convenções da Amazônia | Belém - PA

Realizado por:



Sociedade Brasileira
de Computação

8° Brazilian Workshop on Social Network Analysis and Mining

Não-estacionariedade: Quando tempo e espaço interferem



CEFET/RJ

Eduardo Ogasawara

eogasawara@ieee.org

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