Uso de Redes de Função de Base Radial e Cadeias de Markov para detecção online de mudanças de conceito em fluxos contínuos de dados

Discente: Ruivaldo Neto
Orientador: Ricardo Rios

Universidade Federal da Bahia
Departamento de Ciência da Computação
Programa de Pós-Graduação em Ciência da Computação

Contato: rneto@rneto.dev

16 de Dezembro de 2019

Roteiro

1. Introdução
2. Fundamentação Teórica
3. RBFChain
4. Experimentos
5. Conclusões e Trabalhos Futuros

Introdução

Introdução

- Avanços tecnológicos recentes contribuiram para um aumento exponencial no volume de dados produzidos por sistemas computacionais [16].
- Parte significativa desses dados é produzida atráves de Fluxos Contínuos de Dados (FCDs): sequências ininterruptas e potencialmente infinitas de eventos [2].
- FCDs estão presentes em diversos domínios de aplicação:
  - Análise do Mercado Financeiro;
  - Gestão de redes de telecomunicação;
  - Detecção de intrusos;
  - Monitoramento de tráfico, etc.

Introdução

- Técnicas de Aprendizado de Máquina (AM) têm sido aplicadas para extrair informações úteis desses grandes conjuntos de dados.
- Cenários com FCDs limitam a aplicação de técnicas de AM, pois impõem restrições de tempo de resposta, de uso dos recursos computacionais e apresentam comportamento não estacionário.
- Em cenários não estacionários, o contexto do processo gerador e/ou a distribuição dos dados podem sofrer alterações (mudanças de conceito) ao longo do tempo.
- A ocorrência de mudanças de conceito (*concept drifts*) pode impactar negativamente a acurácia das técnicas aplicadas.

Introdução

- A atualização periódica de modelos, apesar de computacionalmente ineficiente, foi utilizada inicialmente como estratégia para mitigar a perda de acurácia causada por mudanças de conceito.
- A fim de obter maior eficiência e precisão, pesquisadores propuseram novos métodos de detecção de mudança de conceito baseados em monitoramento.

Introdução

- Entretanto, esses métodos ainda apresentam limitações ao serem aplicados em cenários com FCDs [2]:
  - Necessidade de rotulação;
  - Eficiência computacional (tempo de resposta e uso de recursos).

Introdução

- Visando superar essas limitações, este trabalho propõe um novo método de detecção de mudanças de conceito baseado em Redes de Função de Base Radial (redes RBF) e Cadeias de Markov, denominado **RBFChain**.
- O método proposto se diferencia por detectar mudanças em tempo de execução, de forma computacionalmente eficiente e independente de rótulos.

Fundamentação Teórica

Fluxos Contínuos de Dados e Aprendizado de Máquina

- Fluxos Contínuos de Dados (FCDs) são sequências ininterruptas e potencialmente infinitas de eventos [2].
- Não podem ser armazenados em sua totalidade e, por serem de alta frequência, devem ser analisados em tempo real.
- Algoritmos supervisionados [7, 5, 14, 4, 9] e não-supervisionados [3, 1, 12] da área de AM foram adaptados para atenderem a essas restrições.
- Contudo, essas especializações não tratam a ocorrência de mudanças de conceito.

Mudança de Conceito

- A Teoria Bayesiana de Decisão [8] é comumente utilizada para descrever a tarefa de classificação e pode ser utilizada para formalizar a noção de mudança de conceito.
- Considerando que p_{t_0} e p_{t_1} denotam as distribuições de probabilidades conjuntas nos instantes t_0 e t_1, é possível afirmar que há mudança de conceito entre os instantes t_0 e t_1 se:

$$\exists X : p_{t_0}(X, c) \neq p_{t_1}(X, c) \tag{1}$$

- Ou seja: um conjunto de dados possui resultados esperados legítimos em t_0, mas este mesmo conjunto passa a ter resultados esperados diferentes, também legítimos, em t_1 [11].

Mudança de Conceito

- As mudanças de conceito podem ser categorizadas como Virtuais ou Reais [10]:
  - Mudanças Virtuais são causadas por alterações na probabilidade a priori das classes, $P(c)$, e não alteram os conceitos-alvo.
  - Mudanças Reais surgem a partir de alterações na probabilidade a posteriori, $p(c|X)$, e modificam os resultados esperados.

Mudança de Conceito

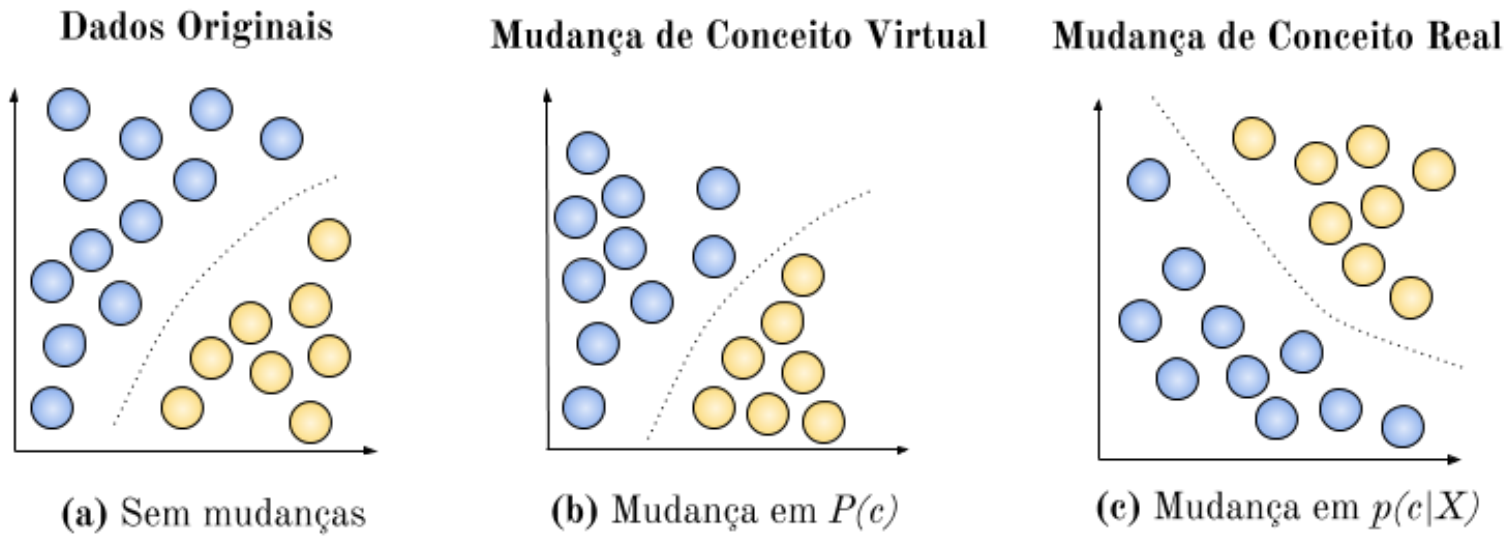


Figura 1: Mudança de Conceito Virtual vs. Mudança de Conceito Real

Mudança de Conceito

- As mudanças de conceito podem ocorrer de forma abrupta, gradual, incremental ou recorrente [15].

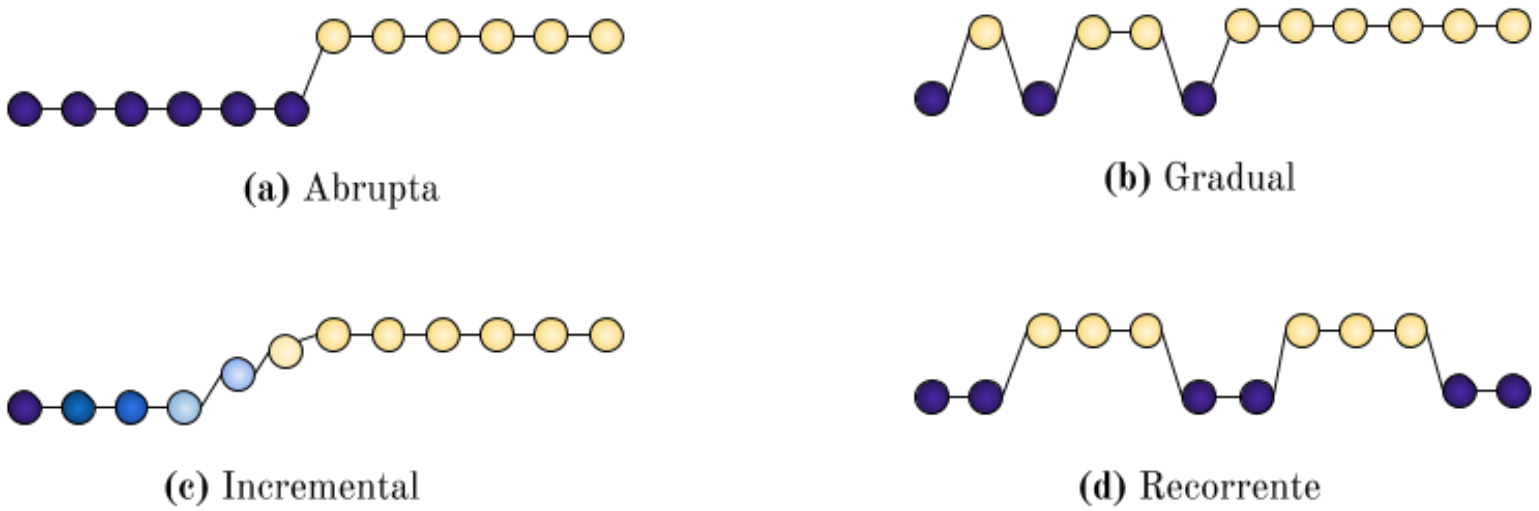


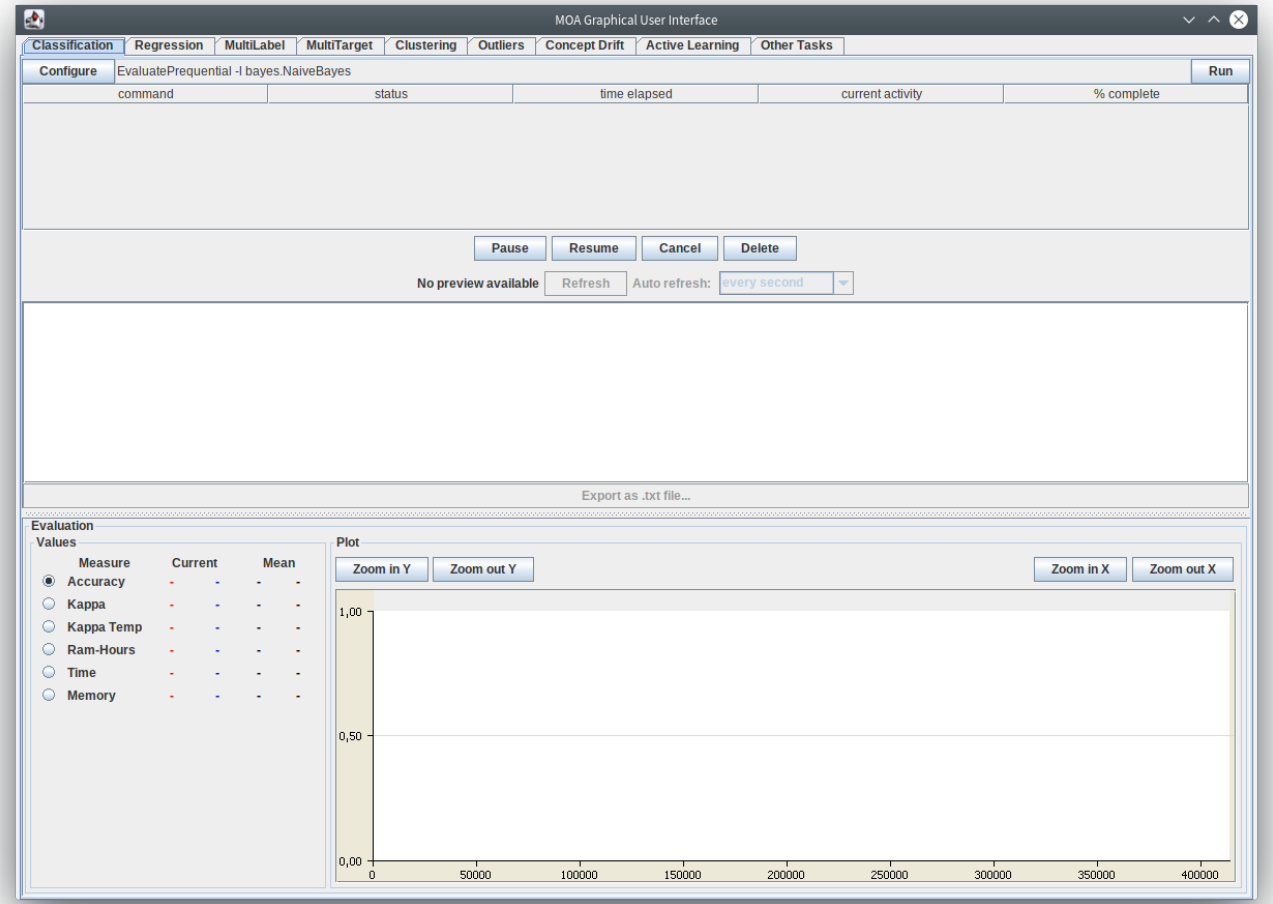
Figura 2: Padrões de ocorrência de Mudanças de Conceito

Algoritmos para Detecção de Mudança de Conceito

- Os algoritmos para Detecção de Mudanças de Conceito se dividem em duas categorias, conforme a necessidade de rotulação dos dados [15]:
  - Explícitos/Supervisionados: **Dependem** da rotulação dos dados. Realizam a detecção a partir do monitoramento de medidas de performance como taxa de erro e acurácia.
  - Implícitos/Não Supervisionados: **Independem** da rotulação dos dados. Realizam a detecção através do monitoramento de características dos próprios dados ou de indicadores produzidos pelas técnicas de aprendizado aplicadas.

Ferramenta: MOA

- Principal framework para mineração de dados em fluxos contínuos.
- Permite implementar e validar novos métodos de detecção de mudança de conceito de forma trivial.



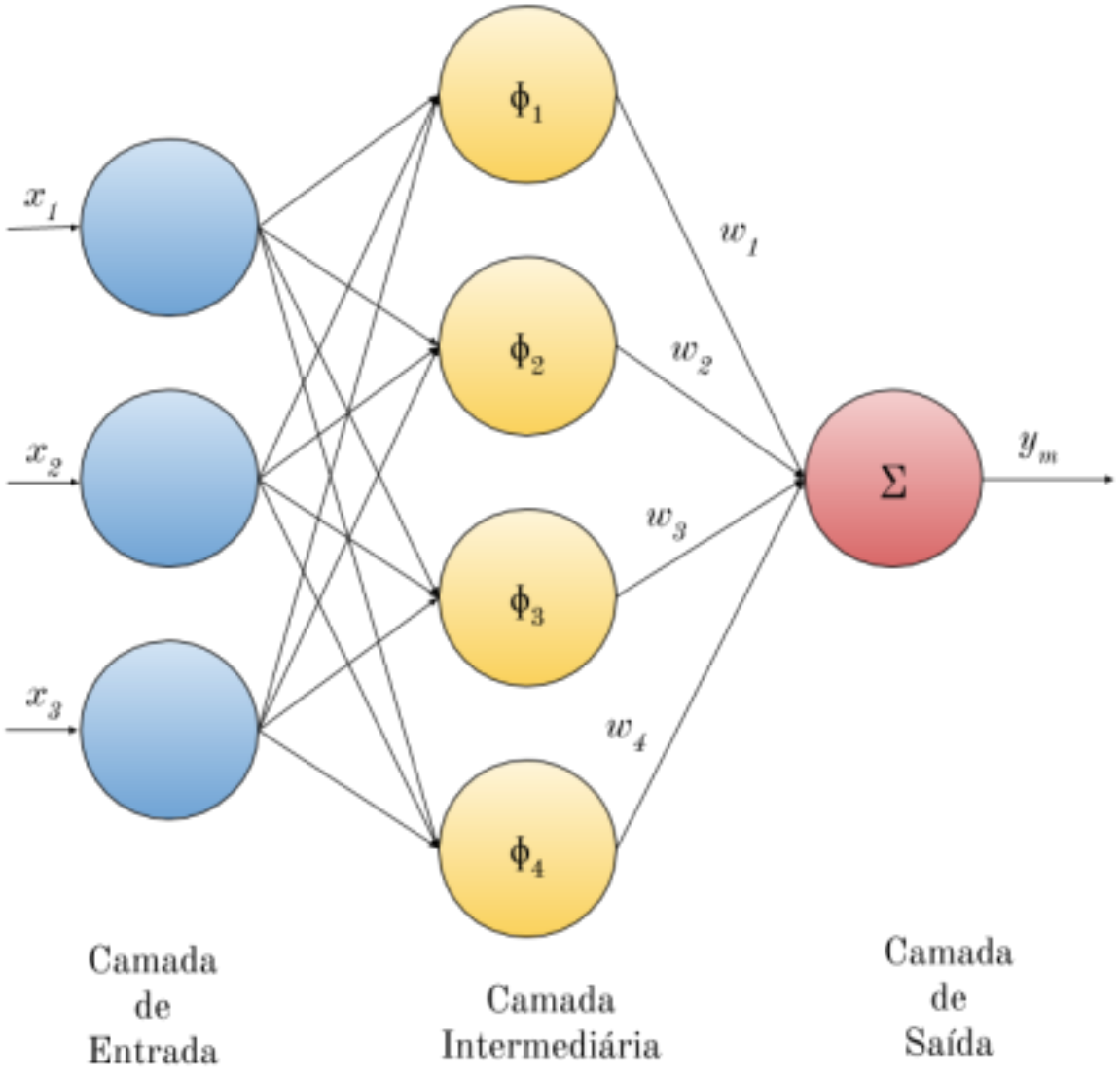
Mudança de Conceito - RBFChain

- O método proposto neste trabalho é capaz de identificar mudanças sob qualquer padrão de ocorrência.
- Por ser independente de rótulos, considera todas mudanças identificadas como mudanças reais.
- Implementado e validado através da plataforma MOA.

Redes de Função de Base Radial

- Redes de Função de Base Radial são redes neurais cujo principal diferencial é a forma de ativação: pois é realizada através do cálculo da distância entre o evento e um centro definido [6].
- A arquitetura de uma rede RBF, em sua forma mais básica, envolve três camadas:
  - Entrada: Recepciona os dados e encaminha para camada intermediária.
  - Intermediária: Composta por funções de ativação de base radial que atuam como neurônios.
  - Saída: Pondera os resultados da camada intermediária, agregando-os linearmente para compor a resposta final da rede.
- Na literatura, as funções Gaussianas são as funções de ativação mais usuais em redes RBF.

Redes de Função de Base Radial



Redes de Função de Base Radial

- O RBFChain utiliza uma rede RBF adaptada, composta apenas pelas camadas inicial e intermediária.

Redes de Função de Base Radial

- O RBFChain utiliza uma rede RBF adaptada, composta apenas pelas camadas inicial e intermediária.
- O processo de ativação realizado na camada intermediária produz, implicitamente, grupos a partir das observações recebidas ao longo do tempo.

Redes de Função de Base Radial

- O RBFChain utiliza uma rede RBF adaptada, composta apenas pelas camadas inicial e intermediária.
- O processo de ativação realizado na camada intermediária produz, implicitamente, grupos a partir das observações recebidas ao longo do tempo.
- Mudanças de conceito são identificadas quando o grupo ativo deste agrupamento é alterado.

Cadeias de Markov

- Equações determinísticas não podem ser utilizadas para descrever sistemas com múltiplos caminhos evolutivos;
- Nestes casos, processos estocásticos são utilizados [13];
- Um processo estocástico é uma coleção de variáveis aleatórias indexadas no tempo: $\{X_t : t \in T\}$;
- Considerando que o processo estocástico esteja no estado s_i e no tempo $t-1$, a probabilidade do processo estar no estado s_j no tempo t é dada pela Equação 2:

$$\mathbb{P}\left(X_t = s_j | X_{t-1} = s_i, \ldots, X_0 = s_0\right) \tag{2}$$

Cadeias de Markov

- Uma Cadeia de Markov, ou Processo de Markov, é um processo estocástico no qual a probabilidade do estado em um dado período de tempo depende apenas do estado no período imediatamente anterior:

$$\mathbb{P}\left(X_t = s_j | X_{t-1} = s_i, \ldots, X_0 = s_0\right) = \mathbb{P}\left(X_t = s_j | X_{t-1} = s_i\right) = p_{ij} \tag{3}$$

Cadeias de Markov

- Um processo de Markov pode assumir os estados $a_1, a_2, \ldots, a_r$, de tal modo que a probabilidade de transição de um estado a_i para um estado a_j seja P_{ij} (um valor dependente apenas de i e j);
- Portanto, é viável elaborar uma matriz com as probabilidades de todas transições (matriz estocástica) - Equação 4:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1r} \\ P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{bmatrix} \tag{4}$$

Cadeias de Markov

- O RBFChain utiliza uma Cadeia de Markov para modelar o agrupamento criado na rede RBF.

Cadeias de Markov

- O RBFChain utiliza uma Cadeia de Markov para modelar o agrupamento criado na rede RBF.
- Os grupos formados representam os estados do modelo markoviano e as mudanças (ativações) entre estes grupos, as transições.

Cadeias de Markov

- O RBFChain utiliza uma Cadeia de Markov para modelar o agrupamento criado na rede RBF.
- Os grupos formados representam os estados do modelo markoviano e as mudanças (ativações) entre estes grupos, as transições.
- Estas mudanças são refletidas no modelo através do aumento da probabilidade correspondente e a diminuição proporcional das outras transições. Estas alterações são realizadas respeitando a condição $0 \leq P_{ij} \leq 1$.

Trabalhos Relacionados

- Pesquisa na literatura em busca de trabalhos que propõem métodos para identificação de mudanças de conceito em fluxos contínuos de dados, de forma online e independente de rótulos.
- Também foram estudadas técnicas que pudessem subsidiar o desenvolvimento de novos algoritmos que atendam a esses requisitos.

Trabalhos Relacionados

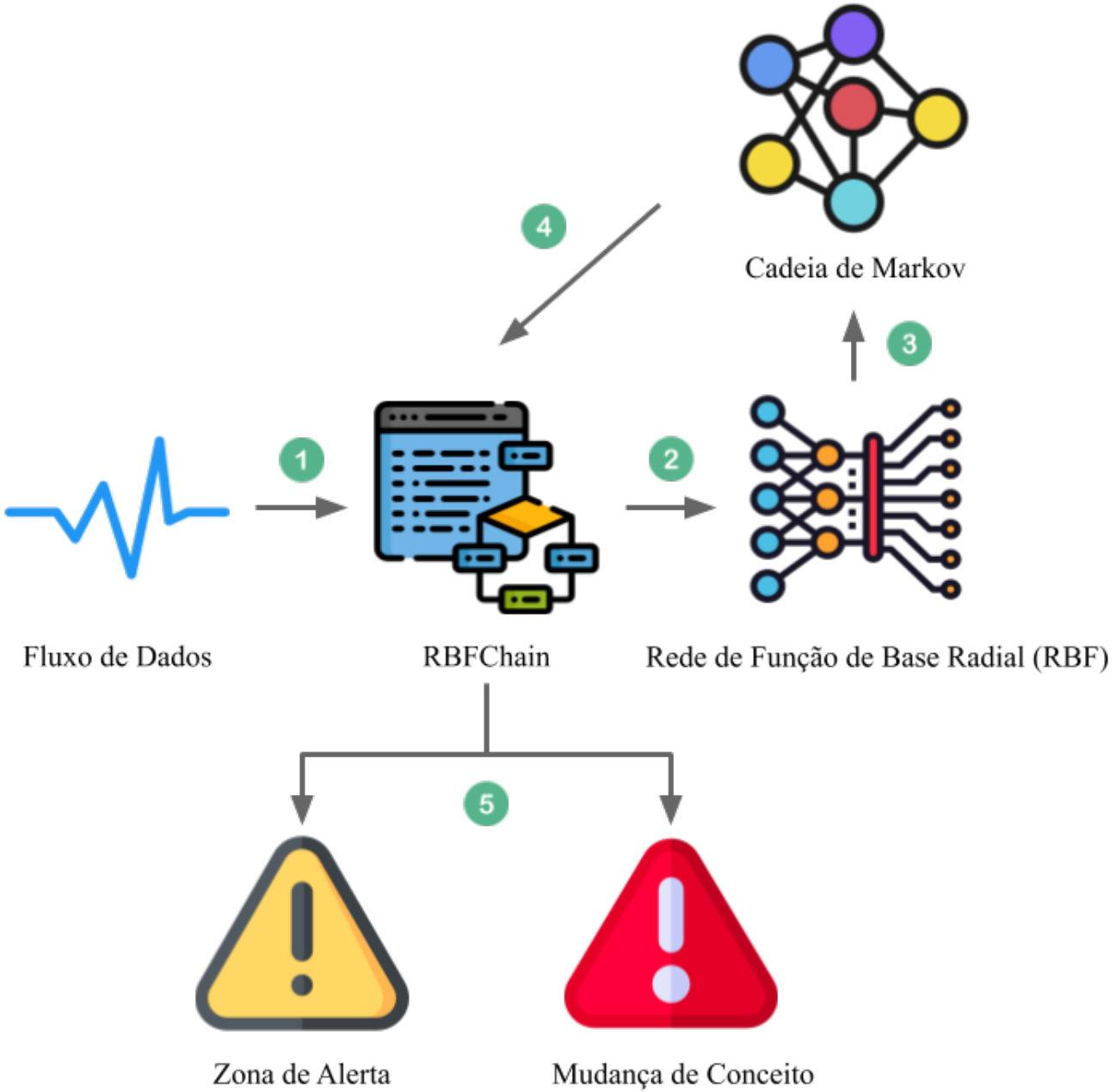
- Análise dos algoritmos Implícitos/Não Supervisionados da subcategoria Detecção de Novidade / Métodos de Agrupamento.
- Análise dos métodos para detecção de *Change Points* em séries temporais que atuam de forma online:
  - Modelos autoregressivos;
  - Séries com autosimilaridade e periodicidade.
- Análise da aplicação de algoritmos de agrupamento estáveis.
- Identificação de lacuna de pesquisa.

RBFChain

Visão Geral

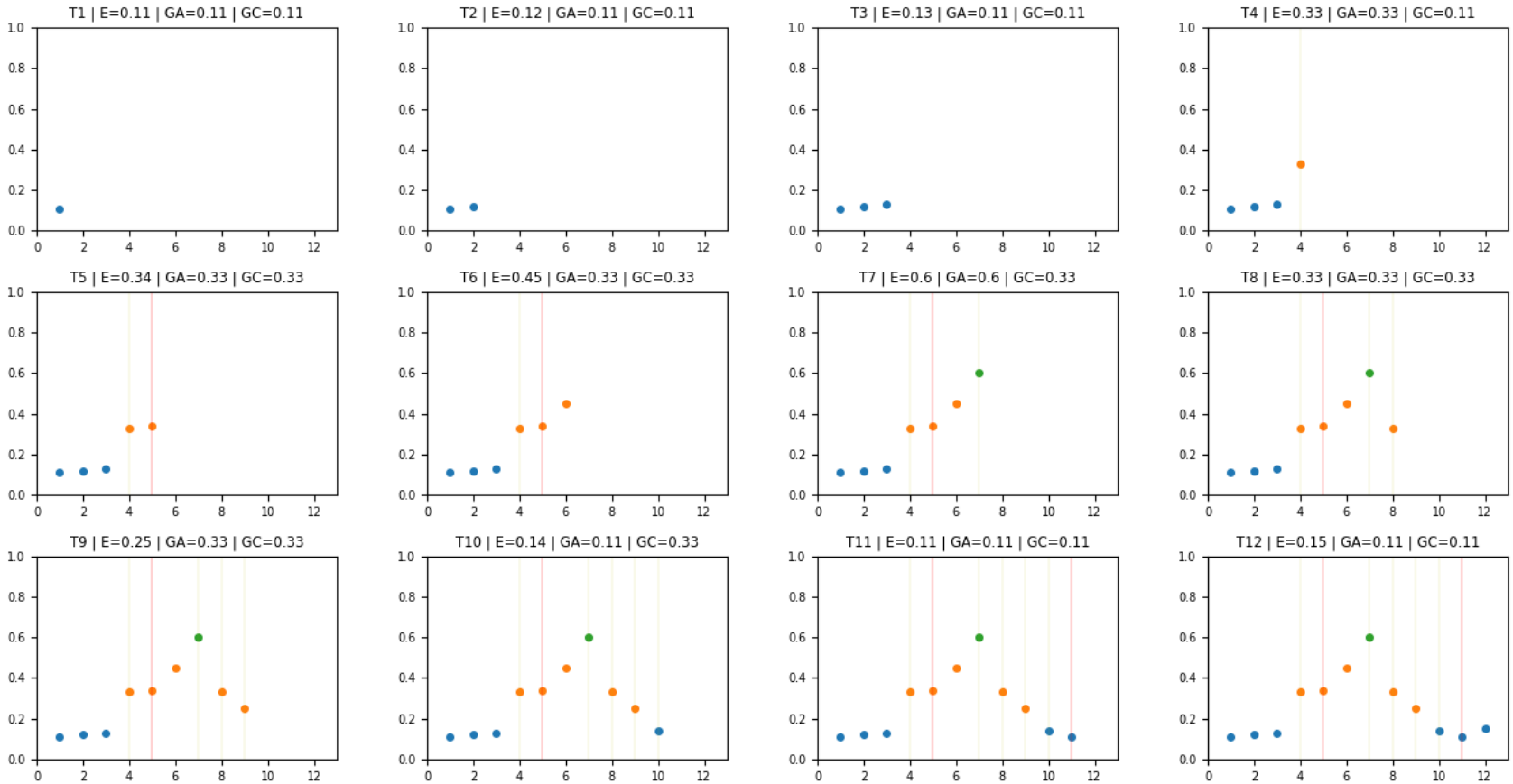
- O RBFChain atua diretamente sobre o fluxo de dados e é composto por dois componentes principais: uma Rede de Função de Base Radial (RBF) adaptada e uma Cadeia de Markov.

Visão Geral


Cadeia de Markov
4
3
1
2
Fluxo de Dados
RBFChain
Rede de Função de Base Radial (RBF)
5
Zona de Alerta
Mudança de Conceito

Execução de exemplo

- $S = 0.11, 0.12, 0.13, 0.33, 0.34, 0.45, 0.6, 0.33, 0.25, 0.14, 0.11, 0.15$
- $\sigma = 3, \lambda = 0.8, \alpha = 0.25, \delta = 0.5$

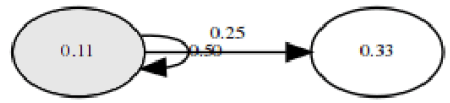


Execução de exemplo

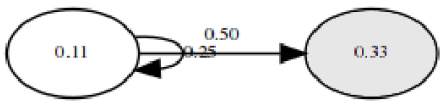
T1

T2

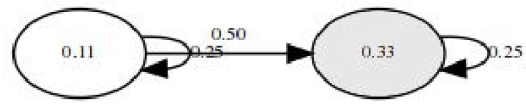
T3



T4



T5



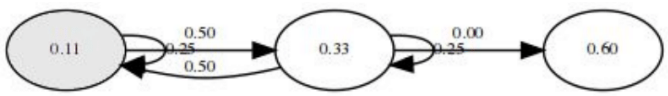
T6

T7

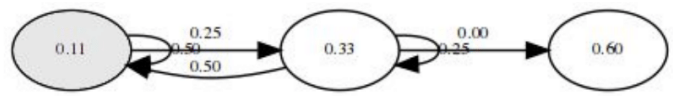
T8

T9

T10



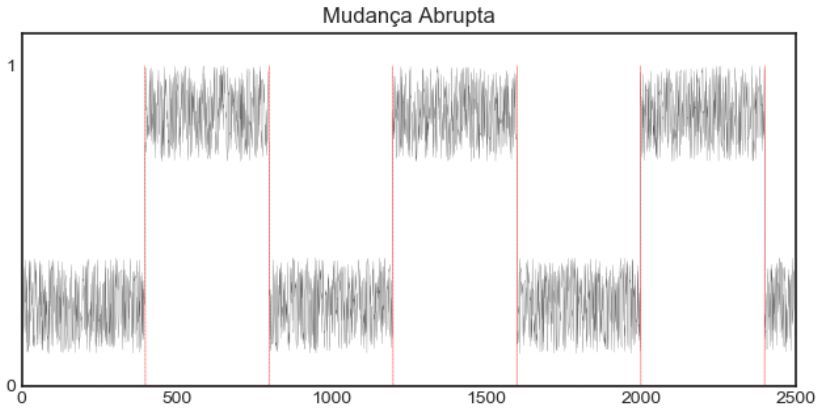
T11

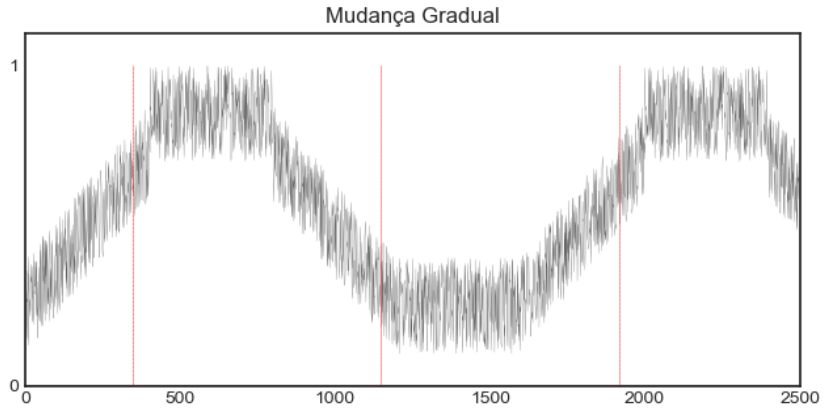


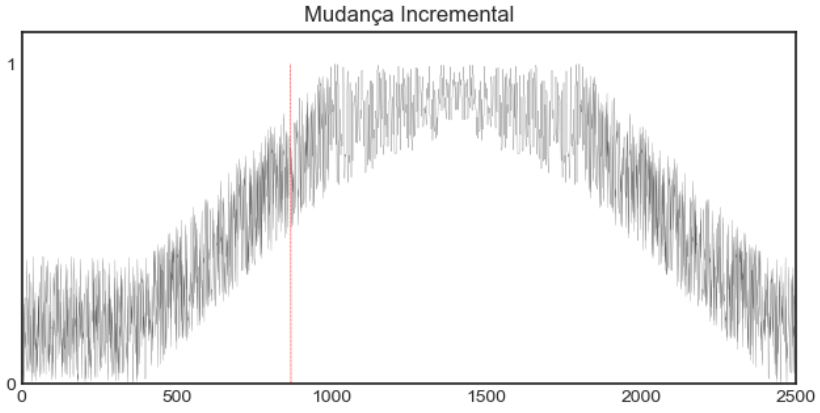
T12

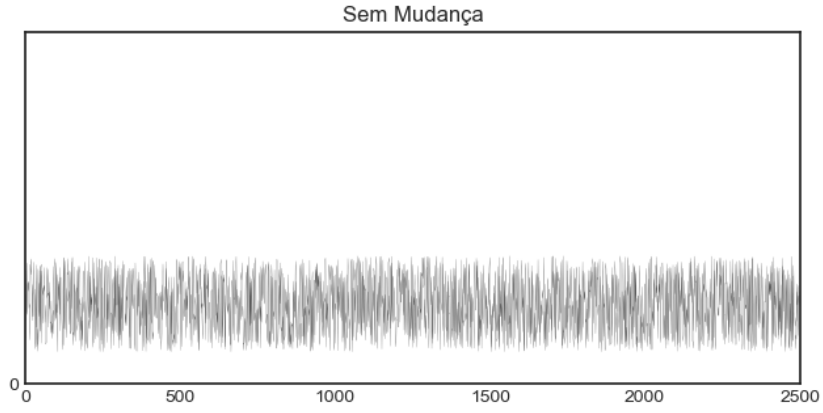
Experimentos

Dados Sintéticos









Dados Sintéticos - Critérios de Avaliação

Indicador	Observação
TP	**Tempo de Processamento** por instância (média em seg.).
MR	**Mudanças Reais** existentes no conjunto de dados.
VP	**Verdadeiro Positivo**. Quantidade de detecções corretas.
FP	**Falso Positivo**. Quantidade de detecções errôneas.
ATR	**Atraso de detecção**. Quantidade média de instâncias até a detecção.

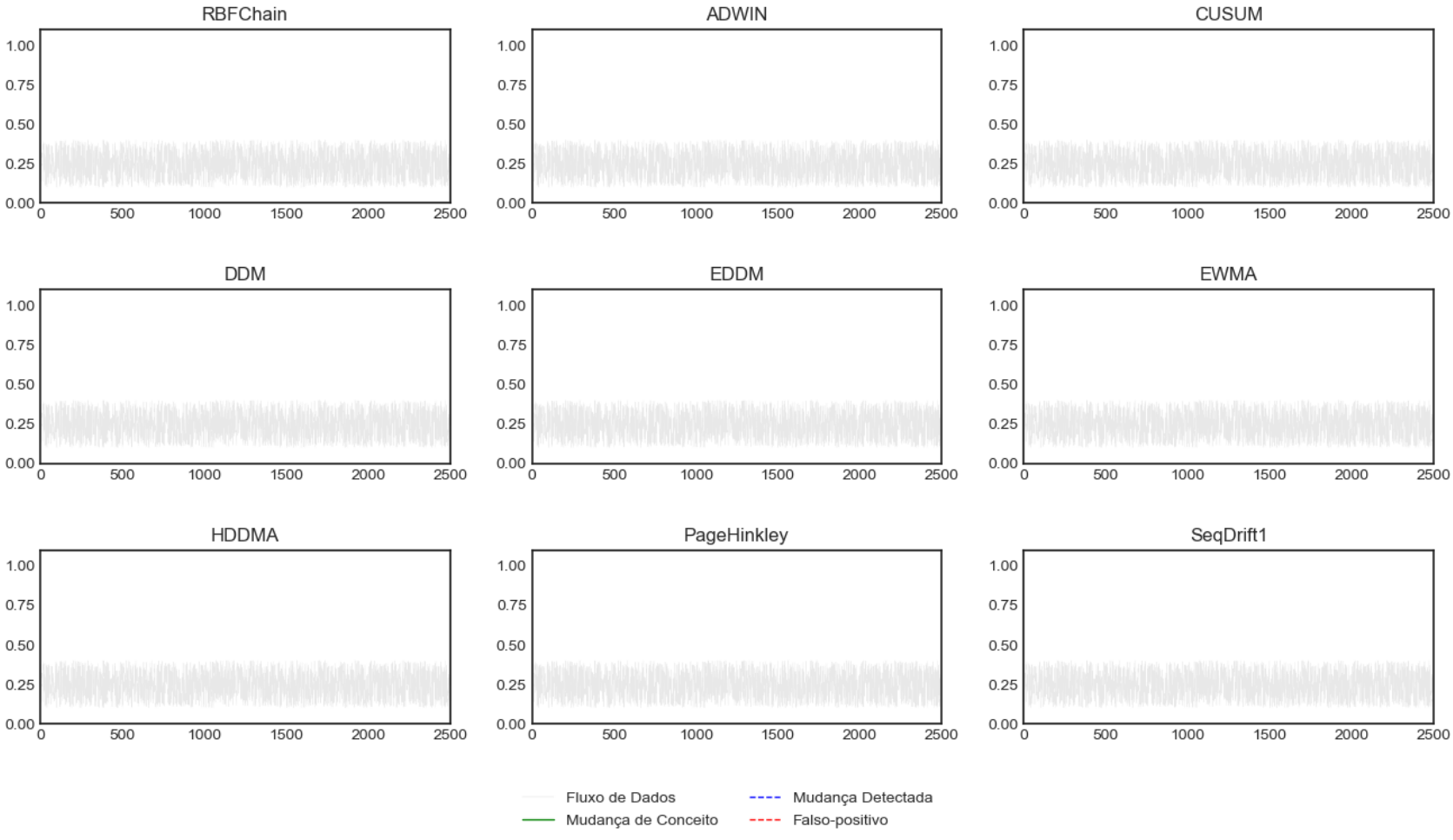
Dados Sintéticos - Resultado: Sem mudanças de conceito

Algoritmo	TP	MR	VP	FP	ATR
RBFChain	0.013	0	0	0	—
ADWIN	0.025	0	0	0	—
CUSUM	0.016	0	0	0	—
DDM	0.014	0	0	0	—
EDDM	0.013	0	0	0	—
EWMA	0.014	0	0	0	—
HDDMA	0.017	0	0	0	—
PageHinkley	0.015	0	0	0	—
SeqDrift1	0.017	0	0	0	—

Dados Sintéticos - Resultado: Sem mudanças de conceito

- Todos algoritmos testados demonstraram tolerância a ruídos e não indicaram nenhum falso positivo.
- RBFChain obteve a melhor média em tempo de processamento, juntamente com o EDDM.

Dados Sintéticos - Resultado: Sem mudanças de conceito



Conclusões e Trabalhos Futuros

Lorem Ipsum

- A:

Lorem Ipsum

- A:
  - A1;
  - A2.

Lorem Ipsum

- A:
  - A1;
  - A2.
- B

Referências i


M. R. Ackermann, M. Märtens, C. Raupach, K. Swierkot, C. Lammersen, and C. Sohler.
Streamkm++: A clustering algorithm for data streams.
J. Exp. Algorithmics, 17:2.4:2.1–2.4:2.30, May 2012.


C. C. Aggarwal.
Data Streams: Models and Algorithms (Advances in Database Systems).
Springer-Verlag, Berlin, Heidelberg, 2006.


C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu.
A framework for clustering evolving data streams.
In *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29*, VLDB '03, pages 81–92. VLDB Endowment, 2003.

Referências ii


C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu.
On demand classification of data streams.
In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 503–508, New York, NY, USA, 2004. ACM.


A. Bifet, B. Pfahringer, J. Read, and G. Holmes.
Efficient data stream classification via probabilistic adaptive windows.
In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 801–806, New York, NY, USA, 2013. ACM.


A. Braga, A. C. Carvalho, and T. B. Ludermir.
Redes Neurais Artificiais: Teoria e aplicações, volume 2.
LTC Editora, 2007.

Referências iii


P. Domingos and G. Hulten.
Mining high-speed data streams.
In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00, pages 71–80, New York, NY, USA, 2000. ACM.


R. O. Duda, P. E. Hart, and D. G. Stork.
Pattern Classification (2Nd Edition).
Wiley-Interscience, New York, NY, USA, 2000.


J. a. Gama, R. Rocha, and P. Medas.
Accurate decision trees for mining high-speed data streams.
In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 523–528, New York, NY, USA, 2003. ACM.

Referências iv


J. a. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia.
A survey on concept drift adaptation.
ACM Comput. Surv., 46(4):44:1–44:37, Mar. 2014.


J. Z. Kolter and M. A. Maloof.
Dynamic weighted majority: An ensemble method for drifting concepts.
J. Mach. Learn. Res., 8:2755–2790, Dec. 2007.


P. Kranen, I. Assent, C. Baldauf, and T. Seidl.
The clustree: Indexing micro-clusters for anytime stream mining.
Knowl. Inf. Syst., 29(2):249–272, Nov. 2011.

Referências v


H. Taylor, S. Karlin, and H. Taylor.
An Introduction to Stochastic Modeling.
Elsevier Science, 1998.


H. Wang, W. Fan, P. S. Yu, and J. Han.
Mining concept-drifting data streams using ensemble classifiers.
In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 226–235, New York, NY, USA, 2003. ACM.


I. Zliobaite.
Learning under concept drift: an overview.
CoRR, abs/1010.4784, 2010.

Referências vi


M. Zwolenski and L. Weatherill.
The digital universe rich data and the increasing value of the internet of things.
Australian Journal of Telecommunications and the Digital Economy, 2, 10 2014.