

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

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

Introdução

- Avanços tecnológicos recentes contribuíram para um aumento exponencial no volume de dados produzidos por sistemas computacionais [16].

- Avanços tecnológicos recentes contribuíram para um aumento exponencial no volume de dados produzidos por sistemas computacionais [16].
- Parte significativa dos dados é produzida através de Fluxos Contínuos de Dados (FCDs): sequências ininterruptas e potencialmente infinitas de eventos [2].

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

- Técnicas de **Aprendizado de Máquina (AM)** têm sido aplicadas para extrair informações úteis de grandes conjuntos de dados.

- Técnicas de **Aprendizado de Máquina (AM)** têm sido aplicadas para extrair informações úteis de 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**.

- Técnicas de **Aprendizado de Máquina (AM)** têm sido aplicadas para extrair informações úteis de 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.

- Técnicas de **Aprendizado de Máquina (AM)** têm sido aplicadas para extrair informações úteis de 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 a acurácia da técnica aplicada.

- A atualização periódica de modelos, apesar de computacionalmente ineficiente, foi utilizada como estratégia para mitigar a perda de acurácia causada por tais mudanças.

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

- Entretanto, os métodos disponíveis na literatura ainda apresentam limitações ao serem aplicados em cenários com FCDs [2]:

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

- 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 **RBFCChain**;

- 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 **RBFCChain**;
- 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 (FCDs) são sequências ininterruptas e potencialmente infinitas de eventos [2].

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.

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.

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**.

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:

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

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:

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

- 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]:

- 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.

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

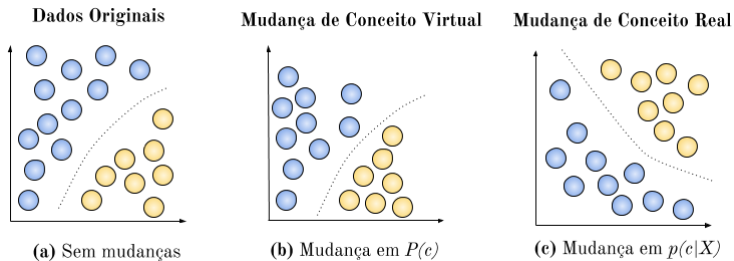


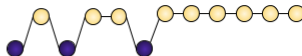
Figura 1: Mudança de Conceito Virtual vs. Mudança de Conceito Real

Mudança de Conceito

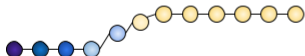
- As mudanças de conceito podem ocorrer de forma **abrupta**, **gradual**, **incremental** ou **recorrente** [15].



(a) Abrupta



(b) Gradual



(c) Incremental



(d) Recorrente

Figura 2: Padrões de ocorrência de Mudanças de Conceito

- Os algoritmos para Detecção de Mudanças de Conceito se dividem em duas categorias, conforme a necessidade de rotulação dos dados [15]:

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, pois realizam a detecção a partir do monitoramento de medidas de performance como taxa de erro e acurácia.

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, pois 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, realizando 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.

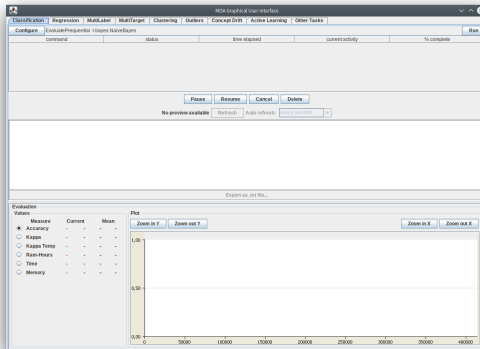


Figura 3: MOA - Tela Inicial

- O método proposto neste trabalho é capaz de identificar mudanças sob qualquer padrão de ocorrência;

- 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**;

- 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, realizada através do cálculo da distância entre o dado e um centro definido [6].

Redes de Função de Base Radial

- **Redes de Função de Base Radial** são redes neurais cujo principal diferencial é a forma de ativação, realizada através do cálculo da distância entre o dado e um centro definido [6].
- A arquitetura de uma rede RBF, em sua forma mais básica, envolve três camadas:

Redes de Função de Base Radial

- **Redes de Função de Base Radial** são redes neurais cujo principal diferencial é a forma de ativação, realizada através do cálculo da distância entre o dado 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.

Redes de Função de Base Radial

- **Redes de Função de Base Radial** são redes neurais cujo principal diferencial é a forma de ativação, realizada através do cálculo da distância entre o dado 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.

Redes de Função de Base Radial

- **Redes de Função de Base Radial** são redes neurais cujo principal diferencial é a forma de ativação, realizada através do cálculo da distância entre o dado 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.

Redes de Função de Base Radial

- **Redes de Função de Base Radial** são redes neurais cujo principal diferencial é a forma de ativação, realizada através do cálculo da distância entre o dado 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

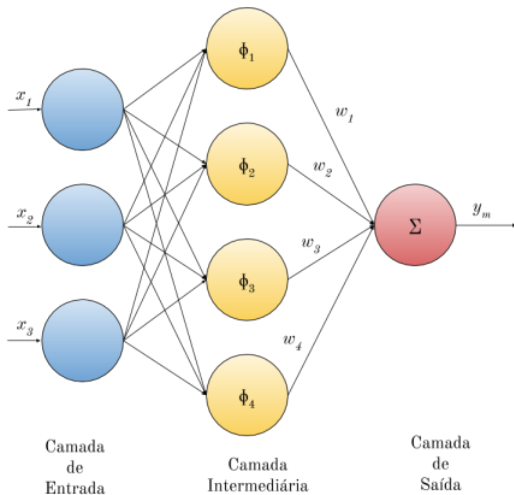


Figura 4: Arquitetura RBF

- O RBFCChain utiliza uma rede RBF adaptada, composta apenas pelas camadas inicial e intermediária;

- O RBFCChain 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;

- O RBFCChain 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.

- Equações determinísticas não podem ser utilizadas para descrever sistemas com múltiplos caminhos evolutivos;

- 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: $\{X_t : t \in T\}$;

- 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:

$$\mathbb{P}(X_t = s_j | X_{t-1} = s_i, \dots, X_0 = s_0) \quad (2)$$

- 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:

$$\mathbb{P}(X_t = s_j | X_{t-1} = s_i, \dots, X_0 = s_0) = \mathbb{P}(X_t = s_j | X_{t-1} = s_i) = p_{ij} \quad (3)$$

- Um processo de Markov pode assumir os estados a_1, a_2, \dots, 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);

- Um processo de Markov pode assumir os estados a_1, a_2, \dots, 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:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1r} \\ P_{21} & P_{22} & \dots & P_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ P_{r1} & P_{r2} & \dots & P_{rr} \end{bmatrix} \quad (4)$$

RBFCChain

- A:

- A:
 - A1;
 - A2.

- A:
 - A1;
 - A2.
- B

Experimentos

- A:

- A:
 - A1;
 - A2.

- A:
 - A1;
 - A2.
- B

Conclusões e Trabalhos Futuros

- A:

- A:
 - A1;
 - A2.

- A:
 - A1;
 - A2.
- B



M. R. Ackermann, M. Mörtens, 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.



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.



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.



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.



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.



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.