



Universidade Federal da Bahia
Instituto de Matemática

Programa de Pós-Graduação em Ciência da Computação

**APLICANDO REDES DE FUNÇÃO DE BASE
RADIAL PARA DETECÇÃO DE MUDANÇAS
DE CONCEITO EM FLUXOS CONTÍNUOS
DE DADOS**

Ruivaldo Azevedo Lobão Neto

QUALIFICAÇÃO DE MESTRADO

Salvador
03 de Abril de 2019

RUIVALDO AZEVEDO LOBÃO NETO

**APLICANDO REDES DE FUNÇÃO DE BASE RADIAL PARA
DETECÇÃO DE MUDANÇAS DE CONCEITO EM FLUXOS
CONTÍNUOS DE DADOS**

Esta Qualificação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Ricardo Araújo Rios

Salvador
03 de Abril de 2019

TERMO DE APROVAÇÃO

RUIVALDO AZEVEDO LOBÃO NETO

APLICANDO REDES DE FUNÇÃO DE BASE RADIAL PARA DETECÇÃO DE MUDANÇAS DE CONCEITO EM FLUXOS CONTÍNUOS DE DADOS

Esta Qualificação de Mestrado foi julgada adequada à obtenção do título de Mestre em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia.

Salvador, 03 de Abril de 2019

Prof. Dr. Ricardo Araújo Rios
UFBA

RESUMO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Palavras-chave: Aprendizado de Máquina, Fluxos Contínuos de Dados, Mudanças de Conceito, Redes de Função de Base Radial, Não supervisionado

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Keywords: Machine Learning, Data Streams, Concept Drift, Radial Basis Function Networks, Unlabeled

SUMÁRIO

Capítulo 1—Introdução	1
1.1 Contexto e Motivação	1
1.2 Hipótese e Objetivo	3
Capítulo 2—Revisão Bibliográfica	5
2.1 Considerações Iniciais	5
2.2 Fluxos Contínuos de Dados	5
2.3 Mudança de Conceito	7
2.3.1 Algoritmos para Detecção de Mudança de Conceito	8
2.3.2 Ferramentas	8
2.4 Redes de Função de Base Radial	8
2.5 Trabalhos Relacionados	8
2.6 Considerações Finais	9
Capítulo 3—Plano de Pesquisa	11
3.1 Considerações Iniciais	11
3.2 Descrição do Problema	11
3.2.1 Atividades de Pesquisa	11
3.3 Considerações Finais	12
Capítulo 4—Experimentos Iniciais	13
4.1 Considerações Iniciais	13
4.2 Configuração dos Experimentos	13
4.3 Método de Pettitt	13
4.4 Redes de Função de Base Radial	14
4.5 Considerações Finais	14

LISTA DE FIGURAS

LISTA DE TABELAS

INTRODUÇÃO

1.1 CONTEXTO E MOTIVAÇÃO

A quantidade de dados produzidos por sistemas computacionais tem crescido exponencialmente. Relatório do IDC (*International Data Corporation*) estimava que em 2014 seriam produzidos 4,4 zettabytes (trilhões de gigabytes) de dados (ZWOLENSKI; WEATHERILL, 2014). O mesmo relatório, prevê que em 2020 essa quantidade será de 44 zettabytes.

Grande parte desses sistemas produz dados através de sequências de eventos. Estas sequências são geradas de forma contínua, ordenada, em alta frequência e potencialmente infinita (FEIGENBAUM et al., 2003). Na literatura, sequências com essas características são denominadas Fluxos Contínuos de Dados (FCDs). Exemplos de sistemas que produzem fluxos dessa natureza, incluem: monitoramento de dados de sensores (Lee; Wang; Ryu, 2007), tráfico TCP/IP, histórico de compra de clientes, filtragem de SPAM em mensagens de e-mail (KATAKIS; TSOUMAKAS; VLAHAVAS, 2010), detecção de intrusos (LANE; BRODLEY, 1998), análise de sentimentos (SMAILOVIĆ et al., 2014), etc.

Nos últimos anos, técnicas de Aprendizado de Máquina (AM) aplicadas a fluxos contínuos de dados têm se tornado um importante tema de pesquisa. Nestes cenários, os algoritmos de aprendizagem devem atender a severas restrições de tempo e processamento (BIFET, 2009). Por exemplo, um classificador deve fornecer um rótulo para um evento antes que o próximo ocorra.

Outra dificuldade encontrada nesses ambientes é a mudança na distribuição dos dados. Isto é, o contexto do processo gerador sofre alterações que impactam no fluxo produzido e, por conseguinte, as previsões esperadas. Este problema é conhecido como **mudança de conceito** (GAMA, 2010) e sua ocorrência impacta a acurácia do modelo criado. Podendo, inclusive, torná-lo obsoleto. Estas mudanças são comumente classificadas conforme a velocidade com que ocorrem. Mudanças de conceito **abruptas** identificam transições rápidas entre conceitos. Transições mais lentas, são ditas **graduais** (GAMA et al., 2014).

Como exemplificação, suponhamos que o histórico de transações via cartão de crédito de determinado cliente seja armazenado. Este cliente, ao longo dos anos, tem utilizado

o cartão apenas para comprar alimentos, sempre em uma mesma região da cidade. A partir desse conjunto de dados, um modelo de aprendizagem é criado. Contudo, não é coerente considerar que o comportamento do cliente permanecerá inalterado.

Diante deste contexto, devemos considerar as seguintes hipóteses: 1) Abruptamente, esse cliente compra vários eletrônicos em outro país. Neste caso, compete aos métodos de detecção identificar se houve fraude ou se o cliente está apenas viajando e aproveitando ofertas. 2) Gradualmente, o cliente passa a utilizar o cartão para compras de passagens aéreas e diminui paulatinamente a utilização para compra de alimentos. Após determinado período, o perfil de compra estará completamente renovado. Assim, cabe novamente aos métodos de detecção identificar a mudança de comportamento e emitir um alerta, para que o modelo construído seja atualizado.

Diversos algoritmos para detecção de mudanças de conceito foram propostos na literatura. Cada uma dessas técnicas possui características e parâmetros diferentes que visam aumentar sua eficiência, conforme a natureza dos dados e do tipo de mudança que se deseja otimizar.

Parte dos algoritmos detecta a mudança de conceito através do monitoramento da taxa de erro do modelo em relação a uma determinada métrica. Alguns algoritmos que adotam essa abordagem são: DDM (GAMA et al., 2004), EDDM (BAENA-GARCÍA et al., 2006), ADWIN (BIFET; GAVALDÀ, 2007), ECDD (ROSS et al., 2012), PL (Bach; Maloof, 2008), FCWM (SEBASTIÃO et al., 2010), STEP (NISHIDA; YAMAUCHI, 2007), DOF (SOBHANI; BEIGY, 2011) e SCCDD/CRCDD (COSTA; RIOS; MELLO, 2016).

Outra metodologia bastante comum para lidar com mudanças de conceito é a aplicação de comitês de classificadores (*Ensemble Classifiers*). Esta técnica consiste na aplicação simultânea de múltiplos classificadores. A combinação das predições individuais é utilizada para compor uma predição global (GAMA et al., 2014). Os métodos para atualização do arranjo (criação e exclusão de classificadores) e a forma de integração das predições individuais variam para cada algoritmo.

Exemplos de comitês de classificadores são DWM (KOLTER; MALOOF, 2007), AUE (BRZEZIŃSKI; STEFANOWSKI, 2011), WMA (BLUM, 1997), DDD (MINKU; YAO, 2012), ADOB (SANTOS et al., 2014), dentre outros.

As abordagens mencionadas dependem da disponibilidade do rótulo correto após determinado período de tempo. Contudo, na maioria das aplicações reais, isto não é viável. Assim, foram desenvolvidos algoritmos para detecção de mudanças de conceito que não dependem dessa disponibilidade de rótulos.

Os métodos independentes de rótulos se baseiam na identificação de eventos que não se enquadram na atual estrutura dos dados (SPINOSA; CARVALHO; GAMA, 2007). Técnicas de agrupamento e detecção de *outliers* são comumente utilizadas nas implementações. Além destas abordagens, estratégias de sumarização dos dados e aplicação de medidas de dissimilaridade (RYU et al., 2012) também são utilizadas para identificar novos conceitos.

Entretanto, as técnicas propostas na literatura apresentam limitações. Algoritmos que dependem da correta rotulação são de aplicação limitada e apresentam dificuldade para lidar com mudanças abruptas. Metodologias inspiradas em algoritmos não super-

visionados envolvem maior custo computacional e tempo de resposta. Visando resolver essas limitações, este projeto de mestrado discute um método para detecção de mudanças de conceito baseado em Redes de Função de Base Radial.

1.2 HIPÓTESE E OBJETIVO

Com base nas observações citadas anteriormente, a seguinte hipótese foi formulada:

“A análise de Fluxos Contínuos de Dados através de Redes de Função de Base Radial permite a detecção de mudanças de conceito, sem requerer a manutenção de estados prévios, de forma ágil e com precisão satisfatória em relação ao estado da arte.”

O objetivo deste trabalho será o desenvolvimento e validação desta hipótese. Para atingir este objetivo, será desenvolvido um algoritmo para detecção de mudanças de conceito baseado em Redes de Função de Base Radial. Este algoritmo diferencia-se por realizar a escolha dos centros de forma *online*, conforme novos eventos são recepcionados, e por apresentar um raio dinâmico. A ativação de novos centros é usada como indicador para possíveis mudanças de conceito.

O algoritmo implementado será comparado com o estado da arte. Os fluxos utilizados nos experimentos serão divididos em dois conjuntos. Um conjunto formado por fluxos sintéticos, que permitirão uma análise detalhada da abordagem, uma vez que as características e os comportamentos são conhecidos. O outro conjunto será composto por *datasets* obtidos a partir de sistemas computacionais utilizados na indústria, visando apresentar uma aplicação prática para a solução proposta neste trabalho.

Este trabalho está organizado conforme a seguinte estrutura: O **Capítulo 2** possui uma revisão bibliográfica dos principais conceitos utilizados neste trabalho como, por exemplo, fluxos contínuos de dados, mudança de conceito e redes de função de base radial; No **Capítulo 3** o plano de pesquisa definido é detalhado, identificando a metodologia que será aplicada e o cronograma de atividades. Por fim, o **Capítulo 4** apresenta um conjunto de experimentos preliminares e a análise dos resultados obtidos em relação ao estado da arte.

REVISÃO BIBLIOGRÁFICA

2.1 CONSIDERAÇÕES INICIAIS

Este capítulo discute conceitos importantes dentro do contexto deste trabalho de pesquisa. Em primeiro momento, Fluxos Contínuos de Dados serão descritos. Em seguida, será abordada a ideia de Mudança de Conceito, seus tipos e características, principais algoritmos de detecção e ferramentas. Posteriormente, serão detalhadas as Redes de Função de Base Radial. Ao fim do capítulo, elencamos os trabalhos relacionados encontrados na literatura.

2.2 FLUXOS CONTÍNUOS DE DADOS

Os avanços em hardware e software dos últimos anos possibilitaram a aquisição de dados em larga escala (AGGARWAL et al., 2003). Como resultado, o interesse de pesquisa no gerenciamento de grandes volumes de dados tem crescido. Parte relevante desses dados é adquirido através de sequências contínuas, potencialmente infinitas e de alta frequência. A literatura denomina séries com essas características como Fluxos Contínuos de Dados (FCDs) (AGGARWAL et al., 2003; GAMA, 2010; GAMA et al., 2004). Entretanto, lidar com volumes de dados tão grandes impõe desafios aos pesquisadores, sobretudo limitações de tempo de execução e recursos computacionais.

Conforme (BABCOCK et al., 2002) as principais diferenças de FCDs para os dados convencionais são:

- Tamanho ilimitado;
- Dados recepcionados de forma contínua e em alta frequência;
- Não há controle sobre a ordem de chegada dos eventos;
- Eventos processados devem ser descartados.

Diversas aplicações práticas da indústria e academia produzem ou consomem dados em fluxos contínuos. São exemplos de aplicações:

Finanças detecção de fraudes em transações, acompanhamento da bolsa de valores;

Energia monitoramento, previsão e planejamento do consumo;

Redes e telecomunicações monitoramento de rede, detecção de intrusos.

A área de Aprendizado de Máquina (AM) tem como objetivo estudar algoritmos que melhoram o seu desempenho na realização de determinada tarefa conforme ganham experiência (MITCHELL, 1997). Atualmente, diversas técnicas de AM estão disponíveis, sendo aplicadas a um amplo conjunto de problemas.

Segundo (GAMA, 2010), a área de AM dedicou-se durante muitos anos em resolver problemas de aprendizado não-incrementais (*batch*), nos quais um conjunto de dados pré-definido é aplicado na fase de aprendizado. Após o processamento desses dados, um modelo de decisão é gerado. Este modelo é estático e produzirá previsões para novos exemplos. A metodologia descrita é adequada para ambientes cuja distribuição dos dados é estacionária, pois o modelo preditivo permanecerá inalterado.

Contudo, os principais cenários da indústria e pesquisa que envolvem FCDs são dinâmicos. Nestes cenários, os dados fluem de forma contínua, mas o contexto do processo que os produz pode sofrer alterações com o passar do tempo. Isto é, são fluxos cuja distribuição dos dados não é estacionária. Conforme (GAMA; RODRIGUES, 2009), no contexto de classificação, o principal problema no aprendizado a partir de FCDs é a manutenção de um modelo de decisão com boa acurácia. Para que esta continuidade seja possível, o algoritmo de aprendizado deve permitir a atualização do modelo de decisão conforme novos dados são analisados. Esta metodologia é denominada Aprendizado Incremental (GAMA et al., 2014).

Conforme (FISHER, 1987), a ideia de aprendizado incremental tem origem na própria observação do mundo real, no qual o conhecimento sofre atualizações constantemente e uma base contínua é mantida para lidar com novos estímulos. O paradigma incremental acredita que o contexto do processo pode mudar ao longo do tempo e, portanto, o conhecimento adquirido deve ser constantemente atualizado.

Segundo (REIMANN; SPADA, 1995), um algoritmo *A* pode ser dito incremental se *A* recebe um evento de treinamento por vez, não reprocessa dados do passado e mantém somente uma estrutura de conhecimento em memória. Além destas, o trabalho também elenca as seguintes características dos algoritmos incrementais:

- Tempo de processamento constante;
- Dados analisados não são mantidos em memória;
- A fase de treinamento pode ser interrompida a qualquer momento para realização de previsões.

Os algoritmos incrementais apresentam as seguintes vantagens (PINTO; GAMA; PORTO, 2005):

- Otimização do tempo de processamento;

- Menor utilização de memória;
- Menor tempo de resposta;
- Aproximam-se da forma de aprendizado dos humanos.

Todavia, os algoritmos incrementais são sensíveis à ordem dos exemplos fornecidos. A depender da ordem de recepção dos exemplos, modelos diferentes podem ser produzidos.

Por suas características, algoritmos incrementais mostram-se bastante interessantes para aplicação em cenários com FCDs. Contudo, (GAMA et al., 2004) ressalta que o aprendizado a partir de FCDs exige que os algoritmos além de incrementais, também tratem mudanças de conceito envolvidas no processo. Nesses casos, é preciso descartar exemplos antigos, que não mais condizem com o cenário atual e adaptar o modelo de decisão aos novos dados.

Algoritmos de aprendizado de máquina clássicos utilizam técnicas como agrupamento (*clustering*) e classificação. A partir dessas técnicas, novos algoritmos foram desenvolvidos para trabalhar com FCDs.

A organização de dados em grupos é uma das formas mais fundamentais de compreensão e aprendizagem (JAIN; DUBES, 1988). Agrupamento é a tarefa de associar objetos conforme sua similaridade. A técnica tem como objetivo formar grupos com alta similaridade entre seus objetos e baixa similaridade para objetos de outros grupos. Técnicas de agrupamento permitem explorar a estrutura dos dados e não exigem os pressupostos comuns à maioria dos métodos estatísticos. A representação criada permite verificar se os grupos estão de acordo com determinados parâmetros ou ainda sugerir novos experimentos (JAIN; DUBES, 1988).

Os métodos de agrupamento têm sido utilizados em diversas áreas do conhecimento, o que ensejou o desenvolvimento de uma grande variedade de algoritmos. Dentre os principais algoritmos para cenários *batch*, destacam-se: K-Means (LLOYD, 2006), DBSCAN (ESTER et al., 1996), PAM (KAUFMAN; ROUSSEAU, 1990) e OPTICS (ANKERST et al., 1999).

Segundo (GAMA, 2010), a principal dificuldade na aplicação de técnicas de agrupamentos em FCDs é definida como: manter a qualidade do agrupamento e a consistência em relação à sequência de dados observados, utilizando pouca memória e tempo. Estas dificuldades são impostas pela chegada contínua de dados e a necessidade de produzir respostas em tempo hábil.

2.3 MUDANÇA DE CONCEITO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit

blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.3.1 Algoritmos para Detecção de Mudança de Conceito

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.3.2 Ferramentas

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.4 REDES DE FUNÇÃO DE BASE RADIAL

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.5 TRABALHOS RELACIONADOS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut

porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.6 CONSIDERAÇÕES FINAIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

PLANO DE PESQUISA

3.1 CONSIDERAÇÕES INICIAIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.2 DESCRIÇÃO DO PROBLEMA

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.2.1 Atividades de Pesquisa

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut

porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.3 CONSIDERAÇÕES FINAIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

EXPERIMENTOS INICIAIS

4.1 CONSIDERAÇÕES INICIAIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.2 CONFIGURAÇÃO DOS EXPERIMENTOS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.3 MÉTODO DE PETTITT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut

porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.4 REDES DE FUNÇÃO DE BASE RADIAL

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.5 CONSIDERAÇÕES FINAIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, C. C. et al. A framework for clustering evolving data streams. In: *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29*. VLDB Endowment, 2003. (VLDB '03), p. 81–92. ISBN 0-12-722442-4. Disponível em: <http://dl.acm.org/citation.cfm?id=1315451.1315460>.
- ANKERST, M. et al. Optics: Ordering points to identify the clustering structure. *SIGMOD Rec.*, ACM, New York, NY, USA, v. 28, n. 2, p. 49–60, jun. 1999. ISSN 0163-5808. Disponível em: <http://doi.acm.org/10.1145/304181.304187>.
- BABCOCK, B. et al. Models and issues in data stream systems. In: *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. New York, NY, USA: ACM, 2002. (PODS '02), p. 1–16. ISBN 1-58113-507-6. Disponível em: <http://doi.acm.org/10.1145/543613.543615>.
- Bach, S. H.; Maloof, M. A. Paired learners for concept drift. In: *2008 Eighth IEEE International Conference on Data Mining*. [S.l.: s.n.], 2008. p. 23–32. ISSN 1550-4786.
- BAENA-GARCÍA, M. et al. Early drift detection method. In: *In Fourth International Workshop on Knowledge Discovery from Data Streams*. [S.l.: s.n.], 2006.
- BIFET, A. Adaptive learning and mining for data streams and frequent patterns. *SIGKDD Explor. Newsl.*, ACM, New York, NY, USA, v. 11, n. 1, p. 55–56, nov. 2009. ISSN 1931-0145. Disponível em: <http://doi.acm.org/10.1145/1656274.1656287>.
- BIFET, A.; GAVALDÀ, R. Learning from time-changing data with adaptive windowing. In: *SDM*. SIAM, 2007. p. 443–448. ISBN 978-1-61197-277-1. Disponível em: <http://dblp.uni-trier.de/db/conf/sdm/sdm2007.html#BifetG07>.
- BLUM, A. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. *Machine Learning*, v. 26, n. 1, p. 5–23, Jan 1997. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1007335615132>.
- BRZEZIŃSKI, D.; STEFANOWSKI, J. Accuracy updated ensemble for data streams with concept drift. In: CORCHADO, E.; KURZYŃSKI, M.; WOŹNIAK, M. (Ed.). *Hybrid Artificial Intelligent Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 155–163. ISBN 978-3-642-21222-2.
- COSTA, F. da; RIOS, R.; MELLO, R. de. Using dynamical systems tools to detect concept drift in data streams. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 60, n. C, p. 39–50, out. 2016. ISSN 0957-4174. Disponível em: <https://doi.org/10.1016/j.eswa.2016.04.026>.

ESTER, M. et al. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1996. (KDD'96), p. 226–231. Disponível em: <http://dl.acm.org/citation.cfm?id=3001460.3001507>.

FEIGENBAUM, J. et al. An approximate l1-difference algorithm for massive data streams. *SIAM J. Comput.*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, v. 32, n. 1, p. 131–151, jan. 2003. ISSN 0097-5397. Disponível em: <https://doi.org/10.1137/S0097539799361701>.

FISHER, D. H. Knowledge acquisition via incremental conceptual clustering. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 2, n. 2, p. 139–172, set. 1987. ISSN 0885-6125. Disponível em: <https://doi.org/10.1023/A:1022852608280>.

GAMA, J. *Knowledge Discovery from Data Streams*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2010. ISBN 1439826110, 9781439826119.

GAMA, J. et al. Learning with drift detection. In: BAZZAN, A. L. C.; LABIDI, S. (Ed.). *SBIA*. Springer, 2004. (Lecture Notes in Computer Science, v. 3171), p. 286–295. ISBN 3-540-23237-0. Disponível em: <http://dblp.uni-trier.de/db/conf/sbia/sbia2004.html/#GamaMCR04>.

GAMA, J.; RODRIGUES, P. An overview on mining data streams. In: _____. [S.l.: s.n.], 2009. v. 206, p. 29–45.

GAMA, J. a. et al. A survey on concept drift adaptation. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 46, n. 4, p. 44:1–44:37, mar. 2014. ISSN 0360-0300. Disponível em: <http://doi.acm.org/10.1145/2523813>.

JAIN, A. K.; DUBES, R. C. *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988. ISBN 0-13-022278-X.

KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Tracking recurring contexts using ensemble classifiers: An application to email filtering. *Knowl. Inf. Syst.*, Springer-Verlag, Berlin, Heidelberg, v. 22, n. 3, p. 371–391, mar. 2010. ISSN 0219-1377. Disponível em: <https://doi.org/10.1007/s10115-009-0206-2>.

KAUFMAN, L.; ROUSSEEUW, P. *Finding Groups in Data: an introduction to cluster analysis*. [S.l.]: Wiley, 1990.

KOLTER, J. Z.; MALOOF, M. A. Dynamic weighted majority: An ensemble method for drifting concepts. *J. Mach. Learn. Res.*, JMLR.org, v. 8, p. 2755–2790, dez. 2007. ISSN 1532-4435. Disponível em: <http://dl.acm.org/citation.cfm?id=1314498.1390333>.

LANE, T.; BRODLEY, C. E. Approaches to online learning and concept drift for user identification in computer security. In: *Proceedings of the Fourth International Conference*

on *Knowledge Discovery and Data Mining*. AAAI Press, 1998. (KDD'98), p. 259–263. Disponível em: <http://dl.acm.org/citation.cfm?id=3000292.3000339>.

Lee, Y. K.; Wang, L.; Ryu, K. H. A system architecture for monitoring sensor data stream. In: *7th IEEE International Conference on Computer and Information Technology (CIT 2007)*. [S.l.: s.n.], 2007. p. 1026–1031.

LLOYD, S. Least squares quantization in pcm. *IEEE Trans. Inf. Theor.*, IEEE Press, Piscataway, NJ, USA, v. 28, n. 2, p. 129–137, set. 2006. ISSN 0018-9448. Disponível em: <http://dx.doi.org/10.1109/TIT.1982.1056489>.

MINKU, L. L.; YAO, X. Ddd: A new ensemble approach for dealing with concept drift. *IEEE Trans. on Knowl. and Data Eng.*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 24, n. 4, p. 619–633, abr. 2012. ISSN 1041-4347. Disponível em: <http://dx.doi.org/10.1109/TKDE.2011.58>.

MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072.

NISHIDA, K.; YAMAUCHI, K. Detecting concept drift using statistical testing. In: COR-RUBLE, V.; TAKEDA, M.; SUZUKI, E. (Ed.). *Discovery Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 264–269. ISBN 978-3-540-75488-6.

PINTO, C. M. S.; GAMA, J.; PORTO, U. do. Algoritmos incrementais para aprendizagem bayesiana. *Faculdade de Economia da Universidade do Porto*, 2005.

REIMANN, P.; SPADA, H. *Learning in Humans and Machines: Towards an Interdisciplinary Learning Science*. Emerald Group Publishing Limited, 1995. ISBN 9780080425696. Disponível em: <https://books.google.com.br/books?id=YkdyQgAACAAJ>.

ROSS, G. J. et al. Exponentially weighted moving average charts for detecting concept drift. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 33, n. 2, p. 191–198, jan. 2012. ISSN 0167-8655. Disponível em: <http://dx.doi.org/10.1016/j.patrec.2011.08.019>.

RYU, J. W. et al. An efficient method of building an ensemble of classifiers in streaming data. In: SRINIVASA, S.; BHATNAGAR, V. (Ed.). *Big Data Analytics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 122–133. ISBN 978-3-642-35542-4.

SANTOS, S. G. T. de C. et al. Speeding up recovery from concept drifts. In: *Proceedings of the 2014th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part III*. Berlin, Heidelberg: Springer-Verlag, 2014. (ECMLPKDD'14), p. 179–194. ISBN 978-3-662-44844-1. Disponível em: https://doi.org/10.1007/978-3-662-44845-8_12.

SEBASTIÃO, R. et al. Monitoring incremental histogram distribution for change detection in data streams. In: GABER, M. M. et al. (Ed.). *Knowledge Discovery from Sensor*

Data. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 25–42. ISBN 978-3-642-12519-5.

SMAILOVIĆ, J. et al. Stream-based active learning for sentiment analysis in the financial domain. *Inf. Sci.*, Elsevier Science Inc., New York, NY, USA, v. 285, n. C, p. 181–203, nov. 2014. ISSN 0020-0255. Disponível em: <https://doi.org/10.1016/j.ins.2014.04.034>.

SOBHANI, P.; BEIGY, H. New drift detection method for data streams. In: *Proceedings of the Second International Conference on Adaptive and Intelligent Systems*. Berlin, Heidelberg: Springer-Verlag, 2011. (ICAIS'11), p. 88–97. ISBN 978-3-642-23856-7. Disponível em: <http://dl.acm.org/citation.cfm?id=2045295.2045309>.

SPINOSA, E. J.; CARVALHO, A. P. de Leon F. de; GAMA, J. a. Olindda: A cluster-based approach for detecting novelty and concept drift in data streams. In: *Proceedings of the 2007 ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2007. (SAC '07), p. 448–452. ISBN 1-59593-480-4. Disponível em: <http://doi.acm.org/10.1145/1244002.1244107>.

ZWOLENSKI, M.; WEATHERILL, L. The digital universe rich data and the increasing value of the internet of things. *Australian Journal of Telecommunications and the Digital Economy*, v. 2, 10 2014.