

Regis Antonio Saraiva Albuquerque

**Seleção Dinâmica de Comitês de Classificadores
Baseada em Diversidade e Acurácia para
Detecção de Mudança de Conceitos**

Manaus - AM

Junho de 2018

Regis Antonio Saraiva Albuquerque

**Seleção Dinâmica de Comitês de Classificadores Baseada
em Diversidade e Acurácia para Detecção de Mudança
de Conceitos**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito para a obtenção do grau de Mestre em Informática.

Universidade Federal do Amazonas – UFAM

Instituto de Computação

Programa de Pós-Graduação em Informática

Orientadora: Eulanda Miranda dos Santos, Ph.D.

Manaus - AM

Junho de 2018

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

A345s Albuquerque, Regis Antonio Saraiva
Seleção Dinâmica de Comitês de Classificadores Baseada em
Diversidade e Acurácia para Detecção de Mudança de Conceitos /
Regis Antonio Saraiva Albuquerque. 2018
71 f.: il. color; 31 cm.

Orientadora: Eulanda Miranda dos Santos
Dissertação (Mestrado em Informática) - Universidade Federal do
Amazonas.

1. Mudança de Conceito. 2. Comitês de Classificadores. 3.
Seleção Dinâmica. 4. Diversidade. I. Santos, Eulanda Miranda dos
II. Universidade Federal do Amazonas III. Título



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



UFAM

FOLHA DE APROVAÇÃO

**"Seleção Dinâmica de Comitês de Classificadores Baseada em
Diversidade e Acurácia para Detecção de Mudança de Conceitos"**

REGIS ANTONIO SARAIVA ALBUQUERQUE

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Profa. Eulanda Miranda dos Santos - PRESIDENTE

Prof. Eduardo James Pereira Souto - MEMBRO INTERNO

Prof. Alceu de Souza Britto Junior - MEMBRO EXTERNO

Manaus, 08 de Junho de 2018

*Este trabalho é dedicado aos meus pais, irmãos, sobrinho e
a todas as pessoas que lutam por um mundo melhor.*

Agradecimentos

Agradeço a Deus, que me deu a vida, a coragem e a fé para seguir o caminho cristão.

À minha família, principalmente aos meu pais, irmãos e sobrinho que são minha base de amor e alegria, que me tornam mais forte. Muitas vezes não pude estar com eles devidos aos trabalhos.

À minha orientadora que tanto contribuiu para que nosso trabalho desse certo, que me guiou sabiamente nos caminhos da pesquisa.

Aos professores da UFAM, pelo grande conhecimento que aprendi e pela paciência em me ensinar da melhor forma possível.

Agradeço a FAPEAM pelo apoio financeiro disponibilizado para a execução dessa pesquisa através do edital de Chamada Pública Número 002/2014 RH-TI Capacitação/Amazônia Turmas Fora de Sede.

A todos que direta ou indiretamente me ajudaram nesta formação, o meu muito obrigado.

*“Sábio é o ser humano que tem coragem
de ir diante do espelho da sua alma
para reconhecer seus erros e fracassos
e utilizá-los para plantar as mais belas
sementes no terreno de sua inteligência.
(Augusto Cury)*

Resumo

Muitas aplicações de aprendizado de máquina estão relacionadas com problemas de classificação em ambientes dinâmicos. Mudança de conceito figura nesse tipo de ambiente e pode prejudicar muito a acurácia de sistemas de classificação. Nesse contexto, a utilização de comitês de classificadores é interessante porque possibilita a implementação de processos de detecção e de reação à mudança mais acurados e robustos. Sistemas de classificação que utilizam comitês podem possuir três grandes fases: geração; seleção; e integração de classificadores. A etapa de seleção pode ser feita de forma dinâmica, isto é, para cada instância desconhecida, o classificador ou comitê de classificadores com maior probabilidade de acerto é escolhido para atribuir uma classe à essa instância. Neste trabalho, é proposto um método para detecção e reação à mudança de conceito que utiliza seleção dinâmica de comitês de classificadores. O método proposto escolhe o comitê especialista com base nos valores de diversidade e de acurácia de cada comitê candidato. A fim de avaliar o impacto do uso de seleção dinâmica guiada por diversidade e acurácia nas tarefas de detecção e reação a mudança de conceito, foram realizadas quatro séries de experimentos com bases sintéticas e reais. Além disso, como o método proposto é dividido em quatro fases: geração da população de comitês; seleção dinâmica do comitê especialista; detecção de mudanças; e reação à mudança, diferentes versões desse método foram investigadas em função da definição de parâmetros de cada fase. Os resultados dos experimentos mostraram que, de maneira geral, as versões estudadas são bem equivalentes em termos de acurácia média final. Adicionalmente, quando comparado a dois *baselines*: (1) DDM - que utiliza um único classificador; e (2) Leveraging Bagging - que utiliza um comitê de classificadores, o método proposto alcançou melhores taxas de acurácia, menores taxas de atraso de detecção, não deixou de detectar as mudanças conhecidas nas bases e produziu reduzidas taxas de falsa detecção, apesar de apresentar maior complexidade computacional. Portanto, o trabalho mostra que o uso de seleção dinâmica guiada por diversidade e acurácia melhora a precisão de detecção, bem como a acurácia geral de sistemas de classificação utilizados em problemas que apresentam mudança de conceitos.

Palavras-chave: mudança de conceito; comitês de classificadores; seleção dinâmica.

Abstract

Many machine learning applications have to deal with classification problems in dynamic environments. This type of environment may be affected by concept drift, which may reduce the accuracy of classification systems significantly. In this context, methods using ensemble of classifiers are interesting due to the fact that ensembles of classifiers allow the design of strategies for drift detection and reaction more accurate and robust to changes. A classification system based on ensemble of classifiers may be divided into three main phases: classifier generation; single classifier or subset of classifier selection; and classifier fusion. The selection phase may be performed as a dynamic process. In this case, for each unknown sample, the individual classifier or classifier ensemble most likely to be correct is chosen to assign a label to the sample. In this work, it is proposed a method for concept drift detection and reaction based on dynamic classifier ensemble selection. The proposed method chooses the expert classifier ensemble according to diversity and accuracy values. Focusing on evaluating the impact of dynamic ensemble selection guided by diversity and accuracy in terms of concept drift detection and reaction, four series of experiments were carried in this work using both synthetic and real datasets. In addition, since the proposed method is broken down into four phases: pool of ensemble classifiers generation; dynamic ensemble selection; drift detection; and drift reaction, different versions of the proposed method were investigated by varying the parameters of each phase. The results show that, in general, all these different versions attain very similar accuracy values. Besides, when compared to two baselines: (1) DDM - single classifier-based; and (2) Leveraging Bagging - classifier ensemble-based, our method outperforms both baselines since it achieved higher accuracy, lower detection delay and false detection rates, and it did not present missing detection. However, both baselines present lower time complexity. Therefore, this work shows that dynamic classifier ensemble selection guided by diversity and accuracy helps to improve detection precision and the general accuracy of classification systems employed in problems with concept drift.

Keywords: concept drift; ensembles of classifiers; dynamic selection.

Lista de ilustrações

Figura 1 – Tipos de mudança de conceito quando se leva em consideração a velocidade da mudança. Fonte: Zliobaite (2009)	27
Figura 2 – Classificação dos métodos que lidam com mudança de conceito. Fonte: Khamassi et al. (2013)	28
Figura 3 – Fases de um sistema de múltiplos classificadores. Fonte: Britto, Sabourin e Oliveira (2014)	30
Figura 4 – Método proposto dividido em quatro fases: a) geração de conjuntos de classificadores; b) seleção dinâmica de conjuntos; c) detecção de mudanças; d) reação a mudança. Fonte: O autor.	41
Figura 5 – Exemplo de uma fronteira de Pareto. Em azul(quadrado), as soluções da fronteira de Pareto da iteração 3155 da base Sine1. Em preto(triângulo), as soluções não pertencentes à fronteira. Fonte: O autor.	45
Figura 6 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 2000 (última instância) para os comitês heterogêneos(parte superior) e comitês homogêneos(parte inferior) para a base Line. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não à fronteira de Pareto de todas as demais iterações. Fonte: O autor.	53
Figura 7 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 10000 (última instância) para os comitês heterogêneos(parte superior) e comitês homogêneos(parte inferior) para a base Sine1. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não a fronteira de Pareto de todas as demais iterações. Fonte: O autor.	53
Figura 8 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 10000 (última instância) para os comitês heterogêneos(parte superior) e comitês homogêneos(parte inferior) para a base Gauss. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não a fronteira de Pareto de todas as demais iterações. Fonte: O autor.	54

Figura 9 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 4000 (última instância) para os comitês heterogêneos(parte superior) e comitês homogêneos(parte inferior) para a base Circle. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não a fronteira de Pareto de todas as demais iterações. Fonte: O autor.	54
Figura 10 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Line. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.	64
Figura 11 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Sine1. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.	65
Figura 12 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Gauss. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.	65
Figura 13 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Circle. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.	65
Figura 14 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Spam. Fonte: O autor.	66
Figura 15 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Elec. Fonte: O autor.	66
Figura 16 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base KDDCup99. Fonte: O autor.	67
Figura 17 – Estudo referente a base Line. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	76

Figura 18 – Estudo referente a base Sine1. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	76
Figura 19 – Estudo referente a base Gauss. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	77
Figura 20 – Estudo referente a base Circle. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	78
Figura 21 – Estudo referente a base Spam. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	78
Figura 22 – Estudo referente a base Elec. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	79
Figura 23 – Estudo referente a base KDDCup99. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.	80

Lista de tabelas

Tabela 1	– Comparação entre medidas de diversidade - conjuntos heterogêneos. . .	39
Tabela 2	– Versões do DESDD testadas nos experimentos. Essas versões são obtidas ao variarmos métodos de reação, algoritmos de aprendizado e detectores de mudança.	48
Tabela 3	– Acurácia frequencial média dos conjuntos heterogêneos C_j^* escolhidos da fronteira de Pareto - bases sintéticas.	55
Tabela 4	– Acurácia frequencial média dos conjuntos homogêneos C_j^* escolhidos da fronteira de Pareto - bases sintéticas.	55
Tabela 5	– Acurácia frequencial média dos conjuntos heterogêneos C_j^* escolhidos da fronteira de Pareto - bases reais	56
Tabela 6	– Acurácia frequencial média dos conjuntos homogêneos C_j^* escolhidos da fronteira de Pareto - bases reais	56
Tabela 7	– Comparação entre medidas de diversidade - conjuntos heterogêneos. . .	58
Tabela 8	– Comparação entre medidas de diversidades - conjuntos homogêneos. . .	58
Tabela 9	– Diferentes versões do DESDD investigadas na terceira série de experimentos.	60
Tabela 10	– Resultado dos testes com as quatro diferentes versões do método - bases sintéticas.	61
Tabela 11	– Resultado dos testes com as quatro diferentes versões do método - bases reais.	61
Tabela 12	– Comparação entre os métodos - bases sintéticas. A maior taxa de acurácia obtida para cada base está destacada em negrito. O símbolo (*) indica diferença estatística significativa para um dos métodos, enquanto o símbolo (**) indica diferença estatística significativa para dois métodos.	63
Tabela 13	– Comparação entre os métodos - bases reais. A maior taxa de acurácia obtida para cada base está destacada em negrito. O símbolo (*) indica diferença estatística significativa para um dos métodos, enquanto o símbolo (**) indica diferença estatística significativa para dois métodos.	63
Tabela 14	– Resultado dos testes de detecção para as bases sintéticas e reais. Siglas: (D) Quantidades de mudanças detectadas; (FD) Quantidade de falsa detecções; (DP) Quantidade de detecções perdidas; (TMA) Taxa Média de Atraso. $TMA = \frac{\sum Atrasos}{NumMudancas}$	66

Sumário

1	INTRODUÇÃO	21
1.1	Contextualização e Motivação	21
1.2	Hipótese de Pesquisa	22
1.3	Descrição do Problema	23
1.4	Objetivos	24
1.5	Contribuições	24
1.6	Estrutura do Documento	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Mudança de Conceito	25
2.2	Tipos de Mudança de Conceito	26
2.3	Soluções para Lidar com Mudança de Conceito	26
2.4	Conjunto de Classificadores	29
2.5	Seleção Dinâmica de Classificadores	30
2.6	Diversidade	31
2.7	Algoritmos de Aprendizado Online	33
2.8	Considerações Finais	34
3	TRABALHOS RELACIONADOS	35
3.1	Considerações Finais	38
4	MÉTODO PROPOSTO	41
4.1	Geração de População de Comitês	41
4.2	Seleção Dinâmica Baseada em Diversidades	42
4.3	Deteccção de Mudança	46
4.4	Reação à Mudança	46
4.5	Considerações Finais	48
5	EXPERIMENTOS E RESULTADOS	49
5.1	Protocolo Experimental	49
5.1.1	Bases de dados	49
5.1.2	Classificadores Base	50
5.1.3	Métricas de Avaliação	50
5.2	Comparação entre funções objetivo: Diversidade x Acurácia	51
5.2.1	Resultados em bases sintéticas	52
5.2.2	Resultados em bases reais	55

5.3	Comparação entre Medidas de Diversidades	57
5.4	Análise comparativa entre diferentes versões do método proposto .	59
5.5	Comparação com <i>baselines</i>	62
5.6	Considerações Finais	67
6	CONCLUSÕES	69
6.1	Trabalhos Futuros	70
	REFERÊNCIAS	71
	APÊNDICE A – ESTUDO - DIVERSIDADE E ACURÁCIA	75

1 Introdução

A maioria das pesquisas na área de aprendizagem de máquina é voltada para ambientes estáticos, os quais não mudam ao longo do tempo, ou possuem todos os dados representativos do problema disponíveis para o treinamento dos algoritmos. No entanto, em diversas aplicações, os algoritmos de aprendizado são executados em ambientes dinâmicos, isto é, ambientes que mudam ao longo do tempo, ou que possuem dados gerados em fluxo. Redes de sensores, monitoramento, gerenciamento de tráfego, telecomunicações e análise de logs da internet são exemplos de aplicações que possuem ambientes dinâmicos (BRZEZINSKI; STEFANOWSKI, 2014).

Nesse tipo de ambiente ocorre a mudança de conceito (do inglês *concept drift*), que se refere a mudanças nas definições de classes do problema que ocorrem no decorrer do tempo. Vale ressaltar que a mudança de conceito também pode ocorrer em ambientes estáticos, quando a base de treinamento não reflete, ou reflete de forma muito limitada, o ambiente em estudo. Mudanças de conceito possuem um impacto direto no desempenho dos sistemas de classificação (PINAGE; SANTOS, 2015).

Algoritmos projetados para lidar com mudança de conceito devem possuir estratégias de detecção e de reação às mudanças. Esses algoritmos podem ser divididos em dois grandes grupos: *online*, quando uma instância é aprendida por vez; e em lote, quando é necessário um bloco de instâncias para que o aprendizado seja realizado (ELWELL; POLIKAR, 2011). Para Minku, White e Yao (2010), algoritmos de aprendizado *online* têm sido bastante úteis para um grande número de aplicações cujos dados de treinamento são disponibilizados continuamente em fluxo de dados e que apresentam limitações de tempo e memória. Exemplos dessas aplicações são: processos de controle industrial, segurança de computadores, categorização de spams, previsões climáticas e detecção de fraudes financeiras.

1.1 Contextualização e Motivação

Um sistema de classificação deve ser capaz de atribuir classes a um grande número de dados, mesmo em ambientes mutáveis (KHAMASSI et al., 2013). O grande desafio acontece quando as distribuições que geram os dados mudam com o tempo. Outra dificuldade ocorre quando dados antigos não estão mais disponíveis, sendo necessária uma abordagem incremental, onde o aprendizado depende do sistema e dos dados correntes (ELWELL; POLIKAR, 2011). Considerando esse cenário, comitês de classificadores têm sido utilizados com sucesso para aumentar a acurácia dos sistemas de classificação que precisam lidar com mudanças (MINKU; WHITE; YAO, 2010).

No contexto de comitês de classificadores, sistemas que utilizam comitês podem possuir três fases (BRITTO; SABOURIN; OLIVEIRA, 2014): geração, seleção e integração. Na primeira, uma população de classificadores ou de comitês de classificadores é criada; na segunda, um classificador ou um subconjunto de classificadores é selecionado; na última fase é feita a predição do sistema obtida a partir da fusão das decisões do(s) membro(s) que foi(foram) selecionado(s).

Considerando especificamente a etapa de seleção, segundo Santos (2008), esta pode ser definida como uma estratégia onde cada membro do comitê de classificadores, ou cada comitê dentre uma população de comitês, é um especialista de classificação em algum lugar no espaço de características. É importante destacar que resultados interessantes foram obtidos através da seleção de classificadores específicos para cada instância de teste. Esse tipo de seleção é conhecido como seleção dinâmica de classificadores (BRITTO; SABOURIN; OLIVEIRA, 2014) .

O processo de seleção dinâmica consiste em escolher o classificador ou o comitê especialista para prever a classe de uma determinada instância individualmente. São criadas regiões de competências para os classificadores, as quais são partições no espaço de características. Caso as regiões sejam determinadas durante a fase de treinamento, diz-se que a seleção é estática. Por outro lado, quando a partição é definida durante a fase de classificação, diz-se que a seleção é dinâmica (BRITTO; SABOURIN; OLIVEIRA, 2014).

Dado um cenário de uma população de comitês de classificadores e levando-se em consideração que o processo de seleção dinâmica escolhe um comitê especialista para cada instância desconhecida, esse tipo de abordagem pode ser utilizada em problemas que envolvem algoritmos de aprendizado *online*, onde uma instância é processada por vez. Porém, a escolha do comitê especialista deve ser feita com base em algum critério. Nesse contexto, a diversidade surge como candidata natural, pois, segundo a literatura, a diversidade entre os membros dos comitês é considerada como fundamental para o sucesso desse tipo de método (BROWN; KUNCHEVA, 2010). No entanto, estratégias que selecionam classificadores com base somente na diversidade não produzem elevado desempenho devido ao fato de não haver uma relação bem definida entre diversidade e acurácia (KO; SABOURIN; BRITTO, 2008). Portanto, a diversidade não deve ser utilizada como único critério na etapa de seleção dinâmica de comitês de classificadores. Alguns trabalhos indicam que a acurácia é a melhor métrica para compor um par de critérios de seleção dinâmica ideal (SANTOS, 2008).

1.2 Hipótese de Pesquisa

Um método de detecção de mudanças de conceitos que utilize seleção dinâmica de comitês baseada em diversidade e acurácia pode aumentar a acurácia e a precisão de

detecção de métodos atuais.

1.3 Descrição do Problema

Em problemas de classificação supervisionada, as instâncias de dados são representadas como pares (x, y_i) , onde x denota uma instância descrita na forma de um vetor de valores de atributos e y_i um rótulo de classe, sendo que y_i pertence à um conjunto de classes $Y = \{y_1, y_2, \dots, y_c\}$. Um algoritmo de aprendizado de máquina constrói um classificador cuja saída deve ser a predição de uma classe y_j , com $j \in \{1, 2, \dots, c\}$, à instância dada. Em aprendizado incremental ou *online*, exemplos chegam continuamente em forma de fluxo de dados (BRZEZINSKI; STEFANOWSKI, 2014).

Geralmente, fluxos de dados podem ser processados a cada instância individualmente ou através de blocos de dados com tamanho fixo B_1, B_2, \dots, B_n , sendo que a atualização dos classificadores é feita após o processamento dos exemplos dos blocos (BRZEZINSKI; STEFANOWSKI, 2014). Em aprendizado *online*, as instâncias são processadas uma a uma. Nesse caso, assume-se que podem ocorrer mudanças de conceitos (exemplo: no tempo t_1 a probabilidade a posteriori $P(y_i|x)$ será diferente de $P(y_i|x)$ no tempo t_2) ou os exemplos de treinamento são limitados e não é possível atingir uma boa generalização com os dados de treinamento disponíveis inicialmente. Portanto, o algoritmo de aprendizado *online* deve adaptar-se às mudanças de conceitos para manter boa acurácia tanto na presença como na ausência de mudanças de conceito.

Além disso, em sistemas de classificação existe o dilema da estabilidade/plasticidade (CHEN; MA; JIANG, 2012). Para Elwell e Polikar (2009), estabilidade está relacionada com reter informações importantes, enquanto plasticidade é a capacidade de aprender novos conceitos. Comitês de classificadores representam uma abordagem promissora para detecção de mudança de conceito por serem efetivos ao prover uma boa relação entre a estabilidade e a plasticidade. Para Chen, Ma e Jiang (2012), diferentes conceitos podem ser armazenados em diferentes classificadores e podem coexistir em um sistema de classificação. Conceitos antigos podem ser armazenados para uso posterior e o sistema pode aprender novos conceitos sem esquecer os antigos.

Na literatura existem algumas abordagens de classificação voltadas para ambientes dinâmicos que utilizam comitês de classificadores para lidar com a mudança de conceito, como por exemplo, o método proposto por Minku e Yao (2012), Diversity for Dealing With Drifts (DDD), que mantém conjuntos de classificadores com diferentes níveis de diversidade. No entanto, foram feitos poucos estudos com seleção dinâmica aplicada a problemas com mudança de conceito (PINAGE; SANTOS, 2015). Diante desse contexto, e em virtude do fenômeno da mudança de conceito possuir impacto direto no desempenho dos sistemas de classificação, neste trabalho é proposto um método que utiliza seleção

dinâmica de comitês de classificadores para detectar e reagir à mudanças de conceito.

1.4 Objetivos

Para melhorar a compreensão dos objetivos da proposta, os mesmos se apresentam divididos em geral e específicos.

Geral

Desenvolver um método de detecção e reação à mudança de conceito em ambientes dinâmicos que utiliza seleção dinâmica de comitês de classificadores baseada em diversidade e acurácia e demonstrar a eficiência do método proposto por meio de comparação de desempenho.

Específicos

- Definir um método de criação de comitês de classificadores com diversidades variáveis;
- Definir um mecanismo de seleção dinâmica de comitês de classificadores baseada em diversidade e acurácia;
- Definir uma estratégia de reação à mudanças de conceitos;

1.5 Contribuições

Como contribuição principal deste trabalho tem-se que o uso da seleção dinâmica de comitês de classificadores baseada em diversidade e acurácia, juntamente com mecanismos de reação a detecção explícita de mudança de conceito, melhora o desempenho de soluções existentes.

1.6 Estrutura do Documento

Este trabalho organiza-se em seis capítulos, incluindo este introdutório, onde são abordadas as principais etapas relacionadas ao objeto da pesquisa realizada. No capítulo 2 serão apresentados os conceitos necessários para uma fundamentação teórica referente aos métodos de reação à mudança de conceito. O capítulo 3 explanará sobre alguns trabalhos relacionados que representam o estado da arte do tema proposto. O capítulo 4 mostra o método proposto, onde são definidas técnicas específicas, abordando suas relações com o resultado final da pesquisa. No capítulo 5 são descritos os experimentos e os resultados obtidos. Por fim, no capítulo 6 são apresentadas as considerações finais.

2 Fundamentação Teórica

Este capítulo apresenta os principais conceitos necessários à compreensão do tema abordado nesta pesquisa. Serão apresentados conceitos referentes à mudança de conceito, a conjuntos de classificadores e ao funcionamento de métodos de reação à mudança de conceito.

2.1 Mudança de Conceito

Sistemas de classificação que recebem continuamente instâncias de exemplo devem manter-se atualizados e preparados para mudança de conceito. Para [Chen, Ma e Jiang \(2012\)](#), esse fenômeno ocorre quando conceitos de um conjunto de dados estão sujeitos à mudanças com o decorrer do tempo, o que torna alguns exemplos do conjunto de dados desatualizados e, portanto, inúteis. O termo conceito refere-se à distribuição da probabilidade conjunta $P(X, y)$ em um determinado tempo t , onde X refere-se aos valores de atributos de entrada e y aos rótulos das classes ([KHAMASSI et al., 2013](#)).

[Elwell e Polikar \(2011\)](#) apresentam duas definições para o termo mudança de conceito (*concept drift*). A primeira, informal, diz que mudança de conceito refere-se à mudança nas definições de classes (conceitos) com o passar do tempo, e conseqüentemente, uma mudança nas distribuições a partir das quais os dados desses conceitos são extraídos. Para definir mudança de conceito formalmente, os autores utilizam a probabilidade a *posteriori* (Teorema de Bayes – Equação 2.1) de uma classe à qual uma dada instância pertence. Mudança de conceito ocorre quando a probabilidade a *posteriori* muda com o passar do tempo. Por exemplo: $P_{t+1}(y|X) \neq P_t(y|X)$.

$$P(y|X) = \frac{P(X|y) \times P(y)}{P(X)} \quad (2.1)$$

Nesse caso, $P(X)$ descreve as probabilidades das características dos dados e $P(X|y)$, a probabilidade de uma instância de dados X , dada uma classe particular. Trata-se de uma probabilidade condicionada à classe. Uma mudança na probabilidade condicional de classe pode ser vista como uma indicação de que os rótulos de classes também podem ter mudado. Por fim, $P(y)$ define a probabilidade a *priori* das classes ([ELWELL; POLIKAR, 2011](#)).

Portanto, uma mudança de conceito representa uma mudança na distribuição do problema, podendo ser uma mudança nas características (mudança somente na probabilidade incondicional da função de distribuição), uma mudança condicional (mudança somente na probabilidade a *posteriori*) ou uma mudança dupla (em ambas as probabilidades)

(MINKU; WHITE; YAO, 2010). Exemplos reais de aplicações que apresentam mudança de conceito são: classificação de spams, previsão do tempo, sistemas de monitoramento, detecção de fraudes financeiras, e evolução das preferências dos clientes (BRZEZINSKI; STEFANOWSKI, 2014).

2.2 Tipos de Mudança de Conceito

Segundo Ditzler, Rosen e Polikar (2012), mudança de conceito pode ser de dois tipos: real ou virtual, de acordo com a sua natureza. O primeiro tipo representa mudanças na probabilidade *a posteriori* $p(y|X)$, enquanto o segundo tipo é denotado por observações de mudança. Por exemplo, mudança na probabilidade condicional $p(X|y)$ – como um resultado incompleto da representação da verdadeira distribuição dos dados. Ou seja, a mudança virtual pode ocorrer em um sistema de classificação quando os dados de treinamento não são suficientes para permitir elevada generalização. Para Elwell e Polikar (2011) a diferença chave entre mudança virtual e real é que a mudança real requer substituição de aprendizado, logo, conhecimentos antigos se tornam irrelevantes. Já a mudança virtual requer aprendizado suplementar, o qual é adicionado ao conhecimento corrente.

Outra forma de diferenciar os tipos de mudança de conceito leva em consideração a velocidade da mudança. Segundo Zliobaite (2009), mudança de conceito pode ser classificada em: abrupta, gradual, incremental e contexto recorrente. Mudança abrupta ocorre quando um conceito $C1$ é repentinamente substituído por outro conceito $C2$. A gradual é referente à mudança progressiva de um conceito, ou seja, em determinado período de tempo $C1$ e $C2$ coexistem, mas com o passar do tempo, o conceito $C1$ é totalmente substituído pelo $C2$. A mudança incremental é muitas vezes confundida com a gradual, aqui, porém, para o conceito $C1$ se transformar em $C2$, $C1$ primeiramente assume formas intermediárias como $C1.1$, $C1.2$ e $C1.3$, por exemplo. Por último, o contexto recorrente ocorre quando um conceito $C1$, que já havia deixado de existir, reaparece após algum tempo. Na figura 1 os tipos de mudança de conceito, de acordo com a velocidade da mudança, são esquematizados (ZLIOBAITE, 2009).

Como o problema de mudança de conceito pode ser observado em diferentes aplicações práticas, há na literatura algumas estratégias propostas para lidar com esse problema. A próxima seção apresenta soluções genéricas existentes. Os trabalhos relacionados, porém, são descritos detalhadamente somente no próximo capítulo.

2.3 Soluções para Lidar com Mudança de Conceito

Algoritmos de classificação devem ser capazes de se adaptar adequadamente às mudanças nas fronteiras de aprendizado. Ao mesmo tempo, devem manter o conhecimento

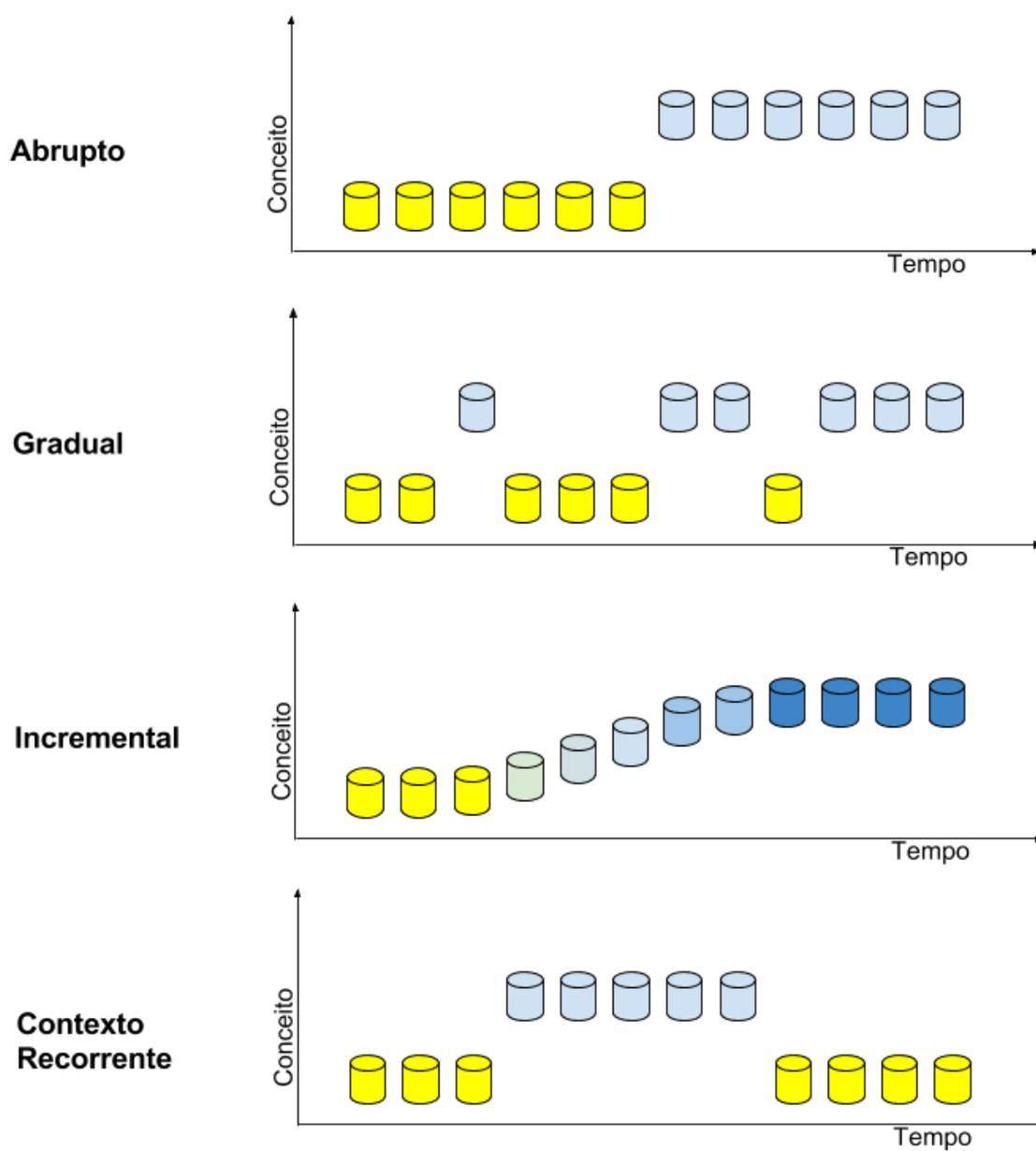


Figura 1 – Tipos de mudança de conceito quando se leva em consideração a velocidade da mudança. Fonte: Zliobaite (2009)

relevante aprendido anteriormente (ELWELL; POLIKAR, 2009). Segundo (KARNICK et al., 2008), um algoritmo que lida com mudança de conceito deve implementar um ou mais dos seguintes procedimentos: 1) detectar a existência da mudança; 2) detectar sua magnitude; 3) ajustar seus parâmetros para aprender o novo conceito; e 4) esquecer o que não é mais relevante.

Para gerenciar mudança de conceito, Khamassi et al. (2013) definem três passos: monitoramento, atualização e diagnóstico. Métodos de monitoramento, chamados de métodos informados, são usados para prover descrições sobre tamanho, localização e o grau da mudança. Esses métodos monitoram o desempenho e outros indicadores de aprendizado. No passo de atualização, as estratégias de atualização são diferenciadas de acordo com o método utilizado para lidar com a mudança de conceito. Métodos cegos adaptam o aprendiz para o conceito corrente de tempos em tempos sem utilizar detectores explícitos de mudanças. Por outro lado, métodos informados atualizam o modelo quando uma mudança é detectada. O passo de diagnóstico objetiva interpretar alterações detectadas em conceitos ou nos parâmetros e estrutura do aprendiz. Na figura 2 é apresentada uma taxonomia dos métodos que lidam com mudança de conceito proposta por (KHAMASSI et al., 2013).

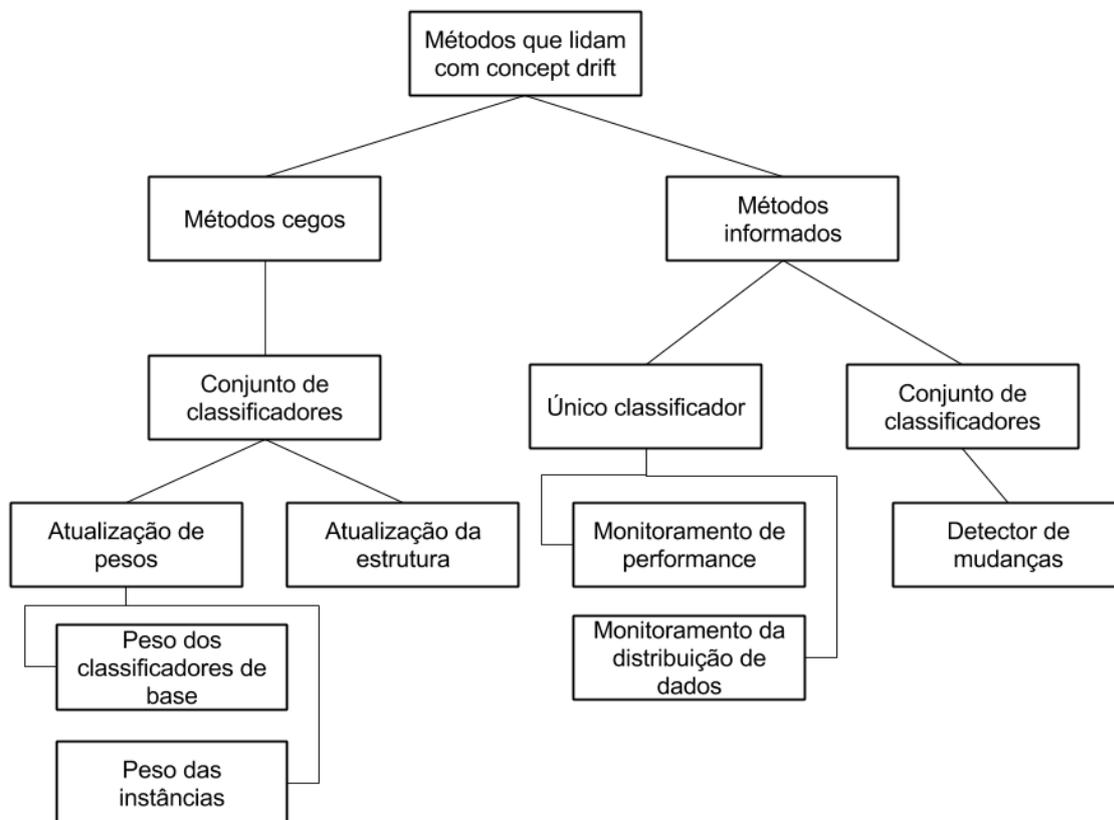


Figura 2 – Classificação dos métodos que lidam com mudança de conceito. Fonte: Khamassi et al. (2013)

Nessa taxonomia, os métodos podem ser divididos em métodos informados e métodos

cegos. Os métodos informados possuem detector de mudança e fazem o monitoramento de alguns indicadores. Os métodos cegos não possuem detector de mudança, pois lidam com esse fenômeno de forma implícita.

É importante mencionar que os métodos informados podem funcionar baseados no monitoramento e diagnóstico das mudanças. Portanto, a acurácia dos classificadores pode ser conservada independentemente da natureza da mudança encontrada. Esses métodos podem usar um único classificador, baseado em monitoramento de desempenho, como acurácia e precisão por exemplo, ou que mede diferenças entre blocos de dados consecutivos; ou ainda usar um conjunto de classificadores com detector de mudança.

Os métodos cegos também utilizam conjuntos de classificadores, porém, sem empregar ferramentas de detecção de mudanças. A detecção é implícita ao processo de atualização. Esses métodos podem atualizar a estrutura do conjunto ou a estrutura dos pesos. No segundo caso, a estrutura dos conjuntos é fixa, mas os pesos dos classificadores, ou das instâncias, podem ser atualizados (KHAMASSI et al., 2013).

Com base nesse contexto, o uso de conjuntos de classificadores no contexto de detecção de mudança pode ocorrer basicamente de duas formas: com detector de mudança (métodos informados) e sem detecção explícita de mudança (métodos cegos). O primeiro combina a flexibilidade de conjuntos de classificadores para detectar diferentes tipos de mudanças e prover descrições de localização, tamanho e severidade. O segundo grupo se adapta às mudanças sem fazer uso de ferramentas de detecção explícita e divide-se em: conjunto de classificadores com atualização da estrutura - por exemplo, um classificador com baixo desempenho é substituído por um novo classificador; e conjuntos de classificadores com atualização dos pesos - quando se mantém uma estrutura fixa de classificadores, por exemplo, os pesos dos classificadores são atualizados de acordo com os novos dados (CHEN; MA; JIANG, 2012).

2.4 Conjunto de Classificadores

Conjuntos de classificadores têm sido usados com sucesso para aumentar a acurácia de classificadores únicos em sistemas de aprendizado incremental (MINKU; WHITE; YAO, 2010). A modularidade dos conjuntos de classificadores provê uma maneira natural de se adaptar a mudanças (BRZEZINSKI; STEFANOWSKI, 2014).

Quando se usa conjuntos de classificadores para lidar com mudança de conceito, primeiro deve-se determinar como escolher os subconjuntos de treino para cada classificador. Então, uma regra de fusão deve ser determinada para combinar as respostas dos classificadores. E por último, um mecanismo deve estabelecer como lidar com a mudança de conceito e esquecer as instâncias menos relevantes (ERDEM, 2007).

Dados podem ser gerados continuamente e disponibilizados em pequenos blocos em intervalos de tempo. Novos conceitos podem ser introduzidos nesses blocos. Por isso, um sistema de classificação existente deve adquirir novos conhecimentos sem esquecer conhecimentos anteriores. Essa habilidade é chamada de aprendizado incremental ([ERDEM, 2007](#)).

Um caso particular de aprendizado incremental são os algoritmos de aprendizado online, que processam cada exemplo de treinamento uma vez durante a “sua chegada” ([MINKU; WHITE; YAO, 2010](#)). Algoritmos de aprendizado online que aprendem uma instância por vez, em vez de aprender blocos de instâncias (*batches*), possuem melhor plasticidade, mas pior estabilidade do que o aprendizado com blocos. Estes se beneficiam da disponibilidade de um maior número de dados, mas podem ser muito ineficientes se o tamanho do bloco for muito pequeno ([ELWELL; POLIKAR, 2011](#)). Um exemplo de algoritmo de aprendizado online é o *OnLine Bagging* ([MINKU; YAO, 2012](#)), utilizado neste trabalho e descrito na seção 2.7.

Conforme mencionado anteriormente, sistemas que utilizam múltiplos classificadores podem apresentar três fases: geração, seleção e integração (representadas na Figura 4). Na primeira, uma população de classificadores é gerada; na segunda, um classificador ou um subconjunto de classificadores é selecionado; na última fase é feita a predição do sistema com base no(s) classificador(es) que foi(foram) selecionado(s). A fase de seleção e integração são facultativas ([BRITTO; SABOURIN; OLIVEIRA, 2014](#)). Uma das estratégias possíveis de seleção é a seleção dinâmica, descrita na próxima seção.



Figura 3 – Fases de um sistema de múltiplos classificadores. Fonte: [Britto, Sabourin e Oliveira \(2014\)](#)

2.5 Seleção Dinâmica de Classificadores

Quando se gera múltiplos classificadores é interessante selecionar classificadores específicos ou subconjuntos de classificadores para fazer a predição final do sistema. Esta fase da seleção que pode ser de dois tipos: estática ou dinâmica. Tanto o esquema estático como o dinâmico tem como função principal prover um classificador único ou um conjunto,

a partir da população de classificadores gerados (BRITTO; SABOURIN; OLIVEIRA, 2014).

A seleção é feita geralmente pela estimativa da competência dos classificadores disponíveis na população. Um processo de particionamento é usado durante o treinamento ou na fase de testes do método. Geralmente, as diferentes partições são associadas com os classificadores mais capazes. A diferença entre os dois tipos de seleção está na forma de particionar as regiões. Na seleção estática, as partições são criadas durante a fase de treinamento, enquanto na seleção dinâmica, as regiões de competência são definidas durante a fase de classificação. No entanto, muitos métodos de seleção dinâmica reportados na literatura definem regiões de competência durante a fase de treinamento (SANTOS, 2008).

Muitas medidas de competência estão presentes na literatura. Existem medidas que consideram classificadores individualmente e em grupo. Medidas que consideram o classificador individualmente, na maioria das vezes levam em conta a acurácia dos classificadores. Podem ser divididos em medidas por: acurácia, ranqueamento, probabilidade, comportamento e oráculo. Medidas que consideram os classificadores em grupo podem ser classificadas por: diversidade e manipulação de dados (BRITTO; SABOURIN; OLIVEIRA, 2014). Como o nosso método emprega seleção dinâmica de classificadores guiada por diversidade, a próxima seção discute essas medidas.

2.6 Diversidade

A diversidade entre membros de conjuntos de classificadores é considerada na literatura como propriedade fundamental para o sucesso de conjuntos (BROWN; KUNCHEVA, 2010). Pode ser alcançada quando os classificadores são construídos variando diferentes fatores, como: algoritmos de classificação, dados de treinamentos e configurações de parâmetros. Existem diversas medidas de diversidade propostas na literatura (LUSTOSAFILHO; CANUTO; XAVIER, 2015).

Dada a importância da diversidade, a relação entre essa propriedade e acurácia já foi bastante investigada no contexto de problemas que não consideram mudança de conceito, como por exemplo (LUSTOSAFILHO; CANUTO; XAVIER, 2015), (BROWN; KUNCHEVA, 2010) e (SANTOS; SABOURIN; MAUPIN, 2008). No contexto de mudança de conceito, entretanto, há poucos estudos que investigam em detalhes a relação entre diversidade e acurácia. Em (MINKU; WHITE; YAO, 2010), por exemplo, é apresentada uma análise de diversidade na presença de diferentes tipos de mudanças. Os autores concluem que é melhor usar conjuntos de classificadores com baixa diversidade na maior parte do tempo, porém, logo após a mudança, conjuntos com alta diversidade apresentam melhores taxas, principalmente para mudanças com alta severidade.

O principal desafio para pesquisas com comitês de classificadores é entender e medir a diversidade a fim de estabelecer o perfeito compromisso entre diversidade e acurácia (SANTOS; SABOURIN; MAUPIN, 2008). A seguir são apresentadas algumas medidas de diversidade estudadas neste trabalho. Onde C_j é o comitê de classificadores candidato, X o conjunto de dados, e c e n seus tamanhos, respectivamente. Em termos de medidas pareadas, N^{ab} denota o número de exemplos de X classificados, onde a se refere ao primeiro classificador e b ao segundo, a e b podem assumir 1 quando o classificador está correto e 0 quando está incorreto. As medidas pareadas são calculadas para dois classificadores c_i e c_k , já as não-pareadas para todo o comitê C_j . As duas primeiras medidas descritas a seguir são não-pareadas, enquanto as demais são pareadas.

Ambiguidade

$$a_i(x) = \begin{cases} 0, & \text{se } y_i = w_k \\ 1, & \text{outros casos} \end{cases} \quad (2.2)$$

a_i é a ambiguidade e y_i é a saída do classificador c_i para a instância x , e w_k é a saída do comitê. A ambiguidade do comitê é medida por:

$$\gamma = \frac{1}{n \times c} \sum_{i \in C_j} \sum_{x \in X} a_i(x) \quad (2.3)$$

Margem

$$\mu(x) = \frac{v(mv|x) - \max_{k \neq mv} v(w_k|x)}{|C_j|} \quad (2.4)$$

Onde mv é o voto majoritário.

Qstatics

$$Q = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (2.5)$$

DoubleFault

$$DF = \frac{N^{00}}{N^{11} + N^{00} + N^{01} + N^{10}} \quad (2.6)$$

Correlation

$$C = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11}N^{00}) + (N^{01}N^{10}) + (N^{11}N^{00}) + (N^{10}N^{00})}} \quad (2.7)$$

Disagreement

$$DF = \frac{N^{01} + N^{10}}{N^{11} + N^{00} + N^{01} + N^{10}} \quad (2.8)$$

Para haver seleção baseada em diversidade é necessário que o sistema possua comitês com diferentes níveis de diversidade. Para isso, deve-se usar algoritmos de aprendizado onde se possa controlar o nível de diversidade do comitê. Na próxima seção são apresentados os algoritmos de aprendizado *online* utilizados neste trabalho.

2.7 Algoritmos de Aprendizado Online

Algoritmos de aprendizado *online* têm sido bastante úteis para um grande número de aplicações cujos dados de treinamento são disponibilizados continuamente em forma de fluxo de dados e apresentam limitações de tempo e memória. Como exemplo: processos de controle industrial, segurança de computadores, categorização de *spams*, previsões climáticas e detecção de fraudes financeiras (MINKU; WHITE; YAO, 2010). Neste trabalho foram utilizados os algoritmos OnLine Bagging e Leveraging Bagging.

O primeiro algoritmo de aprendizado *online* (algoritmo 1) utilizado foi uma versão modificada do OnLine Bagging (OZA; RUSSELL, 2001), proposta em Minku e Yao (2012). O algoritmo foi modificado para incluir o parâmetro λ na distribuição $Poisson(\lambda)$, em vez de forçar $\lambda = 1$. Dessa forma é possível variar o nível de diversidade do comitê.

Algorithm 1 Online Bagging Modificado

INPUT: lambda λ ; qtde de classificadores n ; baselearner BL; instância x ;

```

1: for ( $i \leftarrow 0$ ;  $i < n$ ;  $i++$ ) do
2:    $K \leftarrow Poisson(\lambda)$ 
3:   while ( $K > 0$ ) do
4:      $h_i \leftarrow BL.aprender(h_i, x)$ 
5:      $K \leftarrow K - 1$ 
6:   end while
7: end for
8: output: comitê atualizado  $h$ 

```

No algoritmo 1 para cada classificador h_i uma instância de treino x é apresentada K vezes, onde K é definido pela função de $Poisson(\lambda)$. No final, tem-se como retorno todos os classificadores atualizados aleatoriamente com a instância x de acordo com o parâmetro λ na função de $Poisson(\lambda)$.

O segundo algoritmo, também é uma versão modificada do Online Bagging. Neste caso, além de adicionar o parâmetro λ , o Leveraging Bagging também adiciona mais aleatoriedade aos comitês através do uso de *random output codes*. Ou seja, é feita uma codificação binária relacionando as classes do problema para cada classificador do comitê,

de forma aleatória. O algoritmo é proposto por [Bifet, Holmes e Pfahringer \(2010\)](#) para lidar com mudanças de conceito.

Algorithm 2 Leveraging Bagging

```

1: Inicializa base models  $h_m$  for all  $m \in 1, 2, \dots, M$ 
2: Calcula  $\mu_m(y)$ 
3: for todos os exemplos de treino  $(x, y)$  do
4:   for  $m = 1, 2, \dots, M$  do
5:      $w \leftarrow Poisson(\lambda)$ 
6:     Atualiza  $h_m$  com o exemplo corrente com peso  $w$  e classe  $\mu_m(y)$ 
7:   end for
8: end for
9: if ADWIN detectou mudança em pelo menos um classificador then
10:  Substitua o classificador com maior erro por um novo
11: end if
12: anytime output:
13: Return  $h_{fin}(x) = argmax_{y \in Y} \sum_{t=1}^T I(h_t(x) = \mu_t(y))$ 

```

No algoritmo 2, primeiro é construída uma matriz de valores μ para cada classificador e classe. Para cada nova instância que chega, é atribuído um peso aleatório de $Poisson(\lambda)$. O classificador é treinado com este peso, e quando uma mudança é detectada, o pior classificador do conjunto é removido e um novo classificador é adicionado. Para fazer a predição de uma classe para uma instância desconhecida, primeiro é calculada para cada classe c a soma dos votos de $\mu(c)$ para todos os classificadores do comitê, e a predição é dada pela classe mais votada ([BIFET; HOLMES; PFAHRINGER, 2010](#)).

Neste trabalho, é apresentado um método baseado na utilização de acurácia e diversidade como funções objetivo de seleção dinâmica de conjuntos de classificadores online em problemas com mudança de conceito. São analisados tanto conjuntos homogêneos (mesmo algoritmo de classificação) quanto conjuntos heterogêneos (diversos algoritmos de classificação) em bases de dados sintéticas e reais com diversos tipos de mudanças de conceitos.

2.8 Considerações Finais

Neste capítulo foram discutidos vários conceitos de interesse deste trabalho: mudança de conceito, conjunto de classificadores e seleção dinâmica. Esses conceitos foram definidos para possibilitar a identificação de algumas características existentes nos métodos de reação à mudança de conceito que utilizam conjunto de classificadores. O capítulo 3 apresenta trabalhos relacionados encontrados na literatura.

3 Trabalhos Relacionados

Neste capítulo são descritos os principais trabalhos relacionados ao tema desta pesquisa. Tais trabalhos têm como foco a criação ou aperfeiçoamento de métodos de detecção e de reação à mudanças de conceitos utilizando conjunto de classificadores. Um dos métodos mais famosos para detecção de mudança de conceito é o DDM, que se baseia em uma teoria que assume que se a distribuição dos dados é estacionária, então a taxa de erro do sistema decairá. Quando a distribuição muda, o erro do classificador aumentará. Para o DDM, quando a taxa de erros aumenta atingindo um valor de alerta, um novo contexto é declarado. Caso a taxa de erros aumente ainda mais, atingindo um valor de mudança, o algoritmo aprende o novo contexto usando somente os exemplos armazenados durante a fase de alerta (GAMA et al., 2004).

É consenso na literatura que para se obter uma boa classificação utilizando conjuntos de classificadores é necessário gerar classificadores com diversidade. Existem alguns métodos que combinam classificadores para lidar com mudanças de conceitos, no entanto, como informado na introdução, poucos estudos foram feitos com seleção dinâmica de classificadores utilizando diversidade para problemas de aprendizado online com mudanças de conceitos. Nesta seção, são descritos trabalhos que utilizam diversidade para selecionar conjuntos de classificadores com ou sem a presença de mudanças de conceitos e também estudos de análise de diversidade relacionada ao uso de conjuntos de classificadores no contexto de mudanças de conceitos.

Proposto por Brzezinski e Stefanowski (2014), o AUE2 (Accuracy Updated Ensemble) combina um mecanismo de atribuição de pesos baseado em acurácia do modelo de um conjunto formado incrementalmente por classificadores *Hoeffding Trees*. Para melhorar a reação dos classificadores para os diversos tipos de mudanças, foi utilizada atualização incremental dos classificadores. O AUE2 mantém uma população de classificadores com pesos e prediz a classe dos exemplos de entrada por meio da fusão das predições dos classificadores utilizando a regra de votação ponderada. Após a chegada de cada exemplo de treinamento, um novo classificador é criado, o qual substitui os classificadores que possuem menor desempenho. Após a substituição, os classificadores restantes são atualizados. Embora os autores tenham utilizado *Hoeffding Trees* como classificadores, pode ser utilizado outro algoritmo de aprendizado online como classificadores base.

O algoritmo CSHT (Classifiers ensemble Selected for integration based on Hypothesis Test) de Chen, Ma e Jiang (2012) utiliza teste de hipótese para fazer a seleção e integração dos classificadores. Para cada novo conjunto de dados, são realizados testes nos classificadores de formação com o intuito de detectar mudanças. Se não forem detectadas

mudanças, o classificador é considerado útil e usado para a predição final. O processo de atualização do CSHT funciona da seguinte forma: a) um novo classificador é criado e colocado na lista de classificadores para integração final; b) cada classificador anterior deve tentar detectar mudanças de conceitos. Se não forem detectadas mudanças, o classificador é considerado útil e colocado na lista de classificadores para integração final; c) os classificadores da lista são integrados por uma maioria ponderada para a previsão final.

Proposto por GOMES et al. (2013), Coll-Stream é uma abordagem que utiliza conjuntos de classificadores e seleção dinâmica. O aprendizado é incremental e feito de forma colaborativa. Dispositivos móveis podem colaborar usando o conhecimento disponível nesses dispositivos de forma similar ao conceito local. Coll-Stream particiona o espaço de características em um conjunto de regiões. Quando o sistema é ativado para rotular um novo registro, o melhor modelo de predição da região é usado.

No trabalho de Han, Hui e Sun (2010), foi proposto um framework baseado em conjunto de classificadores para predição de crises financeiras utilizando janelas deslizantes e seleção dinâmica. O método seleciona dinamicamente a janela apropriada para lidar com mudanças de conceitos, e então um método de seleção dinâmica é usado para construir um modelo. Experimentos em bases de dados financeiras indicaram que janelas deslizantes e múltiplos SVMs podem eficientemente adaptar-se a mudanças de conceitos.

Bifet, Holmes e Pfahringer (2010) propuseram uma nova variante do *Bagging* chamada *Leveraging Bagging*. É um método *online* que combina a simplicidade do *bagging* com a adição de mais aleatoriedade: aumentando o *resampling* e manipulando as saídas dos classificadores através do uso de *random output codes*. O método melhorou o desempenho do *Bagging* focando na aleatoriedade como ferramenta para aumentar a acurácia e a diversidade na construção de conjuntos de classificadores. *Leveraging Bagging* utiliza detecção explícita de mudanças através do detector ADWIN. Quando uma mudança é detectada, o pior classificador do conjunto é removido e um novo classificador é adicionado.

O método *Diversity for Dealing with Drifts* (DDD), proposto em (MINKU; YAO, 2012) mantém conjuntos com diferentes níveis de diversidade e lida com detecção de mudanças de maneira explícita no contexto de aprendizado online. O DDD funciona em dois modos: antes da detecção e depois da detecção da mudança de conceito. Antes da detecção, o sistema é composto por dois conjuntos de classificadores: um com baixa diversidade e outro com alta. Os dois são treinados com os exemplos de entrada, porém apenas o de baixa diversidade é usado para predição. Após a detecção, dois novos conjuntos de baixa e alta diversidade são criados. Os conjuntos existentes antes da mudança de conceito são mantidos e denominados antigos conjuntos de baixa e alta diversidade. O DDD foi projetado para usar informações aprendidas do antigo conceito com o intuito de chegar ao novo conceito. É importante destacar que a seleção de classificadores realizada é estática, dado que o sistema sempre escolherá o conjunto com baixa diversidade para

classificar as instâncias desconhecidas.

Diferentemente do DDD, o DYNSE (Dynamic Selection Based Drift Handler) utiliza seleção dinâmica de classificadores para lidar com mudanças de conceitos. Porém, o processo de aprendizado ocorre em *batch*. Os classificadores são treinados a cada novo *batch* disponível e adicionados à população de classificadores. Em vez de manter os últimos conceitos (classificadores), como é feito em muitos trabalhos, o DYNSE mantém muitos classificadores na população. O diferencial do método é a seleção de um subconjunto de classificadores para cada instância desconhecida durante a fase de classificação. Os autores do método afirmam que a seleção do conjunto mais apropriado para a instância a ser classificada é equivalente a classificá-la com base no conceito mais corrente.

A seleção utiliza a vizinhança da instância a ser predita em um conjunto de validação. Portanto, instâncias rotuladas devem estar disponíveis no decorrer do tempo para treinar e selecionar classificadores dinamicamente. Ao receber um novo *batch* com dados rotulados, uma janela de acurácia é estimada e atualizada. Essa janela é usada pelo *framework* para estimar a competência dos classificadores nas regiões locais de cada instância a ser classificada. Por fim, é utilizada uma regra de fusão para combinar os resultados dos classificadores do conjunto selecionado (ALMEIDA et al., 2016). É possível perceber que diversidade não é utilizada no processo de seleção dos conjuntos e sim, competência local medida a partir de taxa de erro.

Além dos métodos que fazem seleção de classificadores usando ou não diversidade, existem vários estudos a respeito da análise da diversidade, porém, poucos deles estão associados com a seleção dinâmica de classificadores. Como exemplo, tem-se o trabalho de LustosaFilho, Canuto e Xavier (2015), no qual é apresentada uma análise de diferentes medidas de diversidade considerando dois métodos de seleção dinâmica. Entretanto, esses trabalhos são ainda mais raros se o aspecto da mudança de conceito também for incluído na análise.

Como exemplo de trabalho neste sentido, temos a análise de diversidade na presença de diferentes tipos de mudanças descrita em Minku, White e Yao (2010). O principal objetivo desse trabalho era entender quando, porque e como aprendizado online com comitês pode ajudar a lidar com mudanças de conceitos. Por não haver estudos a respeito do papel da diversidade na presença de mudanças de conceitos e pelo fato da diversidade ser um fator com papel importante no uso de conjuntos de classificadores, os autores conduziram um estudo sobre a influência da diversidade no contexto de problemas com mudanças de conceitos.

Os autores chegaram à seguinte conclusão: a diversidade exerce um papel importante, o qual varia de acordo com a mudança de conceito e com o tempo de ocorrência da mudança, isto é, antes, logo após, ou muito após a ocorrência da mudança. Antes da mudança, conjuntos com baixa diversidade obtêm as melhores respostas independentemente do

tipo de mudança. Logo após a mudança, conjuntos com mais diversidade apresentam melhor desempenho e a diferença é mais significativa em relação a conjuntos com menos diversidade quando testados com mudanças com alta severidade. Os autores concluem que é uma boa estratégia manter alta diversidade para obter melhores resultados logo após a mudança de conceito, independentemente do tipo de mudança. Porém, muitos *time steps* (iterações) após a mudança, a importância da alta diversidade é reduzida para todos os tipos de mudanças. Portanto, mudanças com diferentes severidades requerem diferentes diversidades. Alta diversidade é mais importante para mudanças mais severas, enquanto o efeito da diversidade em mudanças com velocidades diferentes é muito pequeno.

Entretanto, a análise apresentada em [Minku, White e Yao \(2010\)](#) não considera seleção dinâmica de classificadores. Neste trabalho, o seguinte cenário é definido para que a relação entre diversidade e acurácia seja analisada: dada uma população de n conjuntos de classificadores online, para cada nova instância desconhecida x , se um conjunto é selecionado de forma dinâmica para prever a classe dessa instância com base na diversidade de cada conjunto, qual o impacto das diferentes possibilidades de escolha na acurácia do sistema?

3.1 Considerações Finais

É consenso entre os autores que para se obter uma boa classificação utilizando conjuntos de classificadores é necessário gerar classificadores com diversidade. Porém, somente a geração desses conjuntos não é suficiente. É necessário combinar ou selecionar classificadores. Existem muitos métodos na literatura que combinam classificadores para lidar com mudança de conceito: AUE2, CSHT, Leveraging Bagging e DDD. No entanto, poucos estudos foram feitos com seleção dinâmica e mudanças de conceitos.

DYNSE é o método que mais se aproxima do objetivo deste trabalho, pois é um método que realiza seleção dinâmica de classificadores para lidar com mudanças de conceitos, embora seja de forma implícita (sem o uso de detectores). Apesar do método não utilizar a diversidade como fator de seleção dinâmica, em seu estudo comparativo é indicado que métodos que geram população de classificadores com mais diversidade alcançam bons resultados. O Coll-Stream e o *framework* proposto por [Han, Hui e Sun \(2010\)](#) são exemplos de abordagens que utilizam seleção dinâmica para lidar com mudanças de conceitos. O primeiro foi comparado a outros métodos e mostrou elevada acurácia. No entanto, embora seja uma abordagem voltada para colaboração entre dispositivos móveis, não foram realizados testes relacionados a custos de comunicação e protocolos para compartilhar os modelos de classificação entre os dispositivos. O segundo usa janelas deslizantes e seleção dinâmica para fazer suas previsões. Embora os experimentos indiquem que as técnicas utilizadas podem eficientemente adaptar-se a mudanças de conceitos, o método não foi comparado com outros.

A tabela 1 apresenta uma síntese de comparação entre os trabalhos descritos neste capítulo. São mencionadas características como o tipo de abordagem utilizada, a estratégia de geração dos classificadores empregada, se é realizada seleção e a função de combinação aplicada. O último método descrito, é o método proposto neste trabalho, DESDD. O DESDD foi incluído na tabela para destacar suas diferenças quando comparado aos métodos existentes na literatura.

Método	Abordagem	Geração	Seleção	Fusão
DDM	Informada e Online	-	Não	Único classificador
AUE2	Cega e Bloco	No treinamento cria classificador para substituir o de pior desempenho	Não	Votação ponderada
CSHT	Informada e Bloco	Criação de um classificador a cada bloco	Seleciona utilizando o detector de mudanças	Votação por maioria ponderada dos classificadores selecionados como úteis
Coll-Stream	Cega e Incremental	Classificadores criados em dispositivos móveis separados. Usa janela deslizante de tamanho fixo	A partir de regiões do espaço de características, é selecionado o modelo com a melhor acurácia para a região que contém a instância x	Selecionado apenas um classificador
Framework FDP	Cega e Bloco	Bagging (SVM)	Baseado na estimativa da acurácia de cada classificador. Usa distância euclidiana.	Selecionado apenas um classificador
LB	Informada e Online	Usa online bagging com $\lambda > 1$. Utiliza função de mapeamento binário. Possui um detector ADWIN para cada classificador. Caso ocorra uma mudança, troca o pior classificador.	Não	Voto majoritário. Usa output codes (mapeamento binário)
DDD	Informada e Online	Online bagging. Inicialmente cria 2 comitês: (baixa e alta diversidade). Após o drift: 4 comitês (antigos com baixa e alta diversidade e novos com baixa e alta diversidade)	Sim (estático): Modo antes drift: comitê com baixa diversidade Modo depois drift: mantém novo comitê com baixa diversidade e antigos com alta e com baixa diversidade	Antes Drift: Voto majoritário Depois Drift: Votação ponderada
DYNSE	Cega e Bloco	A cada bloco supervisionado é gerado um novo classificador e colocado na população	Usa a vizinhança para estimar a competência de cada classificador da população. As competências dos classificadores são usadas para selecionar um conjunto.	Voto majoritário, porém, o método pode ser configurado para usar outras funções de combinação.
DESDD	Informada e Online	Algoritmo Online (Ex. Online Bagging ou Leveraging Bagging)	Seleção dinâmica via fronteira de Pareto tendo diversidade e acurácia como funções objetivo. É selecionado o especialista pertencente à fronteira	Voto Majoritário

Tabela 1 – Comparação entre medidas de diversidade - conjuntos heterogêneos.

Neste capítulo foi apresentada uma revisão da literatura de métodos de detecção e de reação à mudança de conceito que utilizam conjuntos de classificadores. No próximo capítulo será apresentado o método proposto neste trabalho.

4 Método Proposto

Neste capítulo é apresentado o método proposto de detecção e reação à mudanças de conceitos que utiliza conjuntos de classificadores, seleção dinâmica e detecção explícita de mudança. Conforme mencionado anteriormente, esse método trabalha de forma *online*, ou seja, as instâncias são classificadas e treinadas uma por vez.

O DESDD (*Dynamic Ensemble Selection for Drift Detection*) possui 4 fases: geração; seleção dinâmica de conjuntos de classificadores; detecção de mudanças; e reação à mudança (Figura 4). Inicialmente é gerada de forma *online* uma população $P = \{C_1, C_2, \dots, C_c\}$ composta por c conjuntos com n classificadores cada e com diferentes taxas de diversidade. Em seguida, para cada nova instância x_i a ser classificada, um dos conjuntos (C_j^*) é selecionado como especialista para atribuir uma classe y_i^* para x_i . Esse processo de seleção dinâmica é realizado com base em valores de diversidade e acurácia. Tão logo a classe real y_i de x_i esteja disponível, o método passa para a fase de detecção de mudança. Caso não haja mudança, o sistema prossegue na geração online da população de comitês. Por outro lado, se houver uma detecção de mudança, o sistema reage para adaptar-se ao novo contexto e um novo processo envolvendo geração da população e seleção dinâmica é iniciado. O método proposto é descrito com detalhes nas próximas seções.

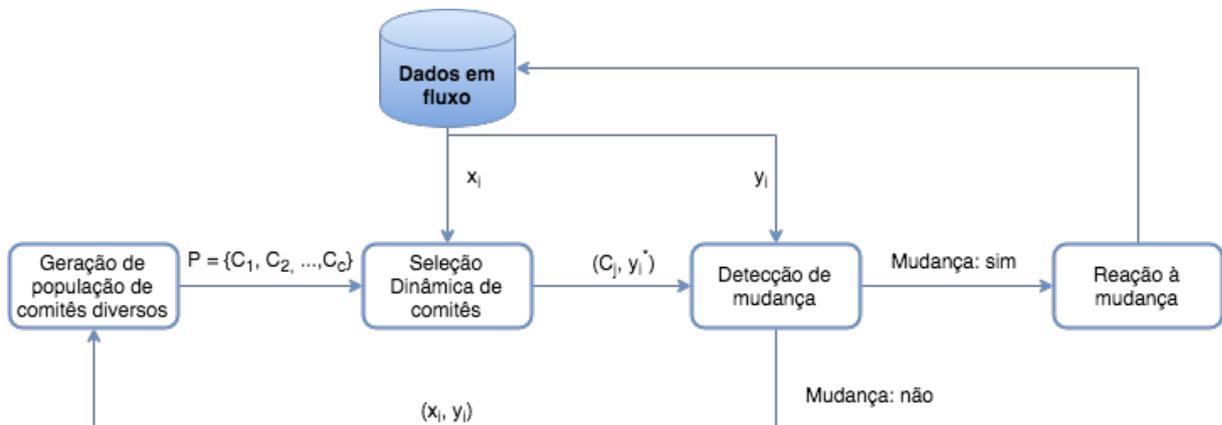


Figura 4 – Método proposto dividido em quatro fases: a) geração de conjuntos de classificadores; b) seleção dinâmica de conjuntos; c) detecção de mudanças; d) reação a mudança. Fonte: O autor.

4.1 Geração de População de Comitês

Primeiramente, é gerada uma população de comitês de classificadores, sendo que esses comitês devem apresentar vários níveis de diversidade. A forma e o nível de inserção de diversidade nos comitês de classificadores dependem do algoritmo de aprendizado *online*

utilizado. Por exemplo, o algoritmo Online Bagging Modificado, descrito no [Capítulo 2](#), treina os classificadores do conjunto de acordo com a função de $Poisson(\lambda)$. Portanto, ao variar-se o valor de λ , cria-se conjuntos com diferentes níveis de diversidade. Nesse caso, quanto menor o valor de λ , maior é a diversidade do conjunto.

Portanto, essa fase do método proposto pode ser realizada utilizando-se qualquer algoritmo de aprendizado *online* que insere explicitamente diversidade nos comitês gerados. Neste trabalho, os dois algoritmos descritos no [Capítulo 2](#) são utilizados, precisamente: Online Bagging Modificado e Leveraging Bagging. Os dois algoritmos recebem como argumento o parâmetro λ , que serve para definir o nível de diversidade do comitê a ser criado. O tipo de algoritmo de aprendizado *online* é escolhido no início da utilização do método, na rotina de configuração do algoritmo.

Além do tipo de algoritmo de aprendizado *online* utilizado, outros parâmetros devem ser definidos para o DESDD. É necessário informar os seguintes parâmetros: 1) quantidade c de comitês a serem criados; 2) valores de λ para cada comitê; 3) algoritmo de aprendizado online a ser utilizado durante o treinamento do método; 4) quantidade n de classificadores membros dos comitês; 5) tipo de classificadores base BL de cada comitê - os quais poderão ser homogêneos ou heterogêneos; e 6) métrica a ser utilizada para calcular a diversidade dos comitês. Além disso, neste trabalho foram investigadas diferentes estratégias de seleção dinâmica de comitês, diferentes detectores de mudança de conceito, e quatro estratégias de reação à mudança.

É importante destacar que nós utilizamos nos experimentos a técnica de *Test Than Train*, ou seja, primeiramente testa-se a instância x_i e depois treina-se a população de comitês com x_i e seu rótulo y_i . Porém, é possível também realizar-se o treinamento com um lote inicial de instâncias rotuladas, para então, seguir-se o processo de aprendizado *Test Than Train*. Dessa forma, o método iniciará com um conhecimento mais abrangente sobre o conceito corrente. Além disso, em termos de detecção de mudança, duas possibilidades são investigadas neste trabalho: (1) geração de um detector para toda a população de comitês; e 2) geração de um detector para cada classificador membro de cada comitê. A primeira abordagem segue a ideia geral do fluxo de processamento do DESDD. Já a segunda estratégia é inspirada no método Leveraging Bagging, o qual também foi proposto para detectar mudança e usa um detector por classificador membro do comitê. Esse processo será descrito com mais detalhes na seção que trata da fase de reação à mudança.

4.2 Seleção Dinâmica Baseada em Diversidades

Após a criação da população de comitês e com a chegada de uma instância desconhecida x_i , deve-se encontrar o melhor conjunto C_j^* para classificar essa instância. A fase de seleção do DESDD, aqui apresentada, utiliza análise de Pareto com duas

Algorithm 3 DESDD-TreinamentoComDetectorPorClassificador()

INPUT: vetor de ensembles *Ensembles*; acurácias frequenciais *ACC*, instância *x*; e janela de cálculo *J*;

- 1: $ACC \leftarrow atualiza(ACC, Ensembles, x)$
- 2: **if** $J > 0$ **then** ▷ Usa janela deslizante
- 3: $JANELA \leftarrow desliza(JANELA, J, x)$
- 4: **for** ($i \leftarrow 0$; $i < N$; $i++$) **do**
- 5: $DIV[i] \leftarrow MD.calculaPorJanela(Ensembles[i], JANELA)$
- 6: **end for**
- 7: **end if**
- 8: $AO.treinarEnsembles(Ensembles, x)$
- 9: **if** pelo menos 1 ensemble apresentou drift **then**
- 10: $ER.executaReacao(Ensembles, x)$
- 11: **end if**
- 12: Return $Ensembles, ACC, DIV$

funções objetivo: acurácia frequencial e diversidade, para encontrar C_j^* que apresente o melhor compromisso entre diversidade e acurácia frequencial. Como pode ser visto no próximo capítulo, foi investigada a possibilidade de seleção dinâmica mono-objetivo, isto é, apenas guiada por diversidade. Porém, os resultados obtidos indicaram que a abordagem bi-objetivo, via fronteira de Pareto, apresenta melhores taxas de acurácia.

De fato, de acordo com a literatura, para garantir elevado desempenho (alta acurácia e classificação confiável), os classificadores membros de um comitê devem ser acurados e diferentes uns dos outros. Portanto, o principal desafio para comitês de classificadores é entender e medir diversidade de forma que seja estabelecido um compromisso entre diversidade e acurácia. Segundo Gu e Jin (2015), diversidade e acurácia são objetivos conflitantes. Logo, se ranquearmos as soluções não dominadas, via fronteira de Pareto em ordem ascendente de acordo com uma função objetivo, um ranqueamento em ordem decrescente será obtido de acordo com a outra função objetivo.

Poranto, para cada instância a ser classificada, inicialmente calcula-se a diversidade e a acurácia frequencial de todos os comitês. Em seguida, encontra-se as soluções não dominadas por meio do conceito de fronteira de Pareto. Por fim, dentre as soluções não dominadas encontradas, seleciona-se a melhor solução C_j^* com base em um critério definido pelo usuário. O cálculo da diversidade pode ser realizado utilizando-se qualquer uma das fórmulas descritas no Capítulo 2. Em nossos experimentos, todas essas medidas foram testadas. Vale a pena também ressaltar que dois critérios de seleção da melhor solução não dominada C_j^* foram investigados: maior ou menor diversidade. Os resultados dos experimentos mostram que, dentre os comitês pertencentes à fronteira de Pareto, os que apresentam maior diversidade também apresentam maior acurácia média final. Portanto, por padrão, o DESDD seleciona como C_j^* a solução não dominada que apresenta a maior diversidade. Porém, o método pode ser configurado de forma a selecionar a solução com

menor diversidade também.

O conceito de fronteira de Pareto é descrito formalmente da seguinte maneira. Uma solução C_j no espaço de variável representa um comitê e para cada C_j existe um ponto $f(C_j) = z = (div(C_j), acc(C_j))$ no espaço objetivo, sendo $div(C_j)$ a primeira função objetivo, que representa a diversidade do comitê, e $acc(C_j)$ a segunda função objetivo, que representa a acurácia prequencial. No processo de seleção bi-objetivo, procura-se maximizar a acurácia prequencial e minimizar a diversidade. Os experimentos iniciais realizados na tentativa de utilizar diversidade como única função objetivo mostraram que melhores valores de acurácia são obtidos quando busca-se minimizar a diversidade. Por isso, a ordenação bi-objetivo das soluções via fronteira de Pareto é feita maximizando-se a acurácia e minimizando-se a diversidade. Essa configuração de funções objetivo também segue um padrão encontrado na literatura de seleção de comitês de classificadores, pois, há vários trabalhos que indicam que essa é a melhor forma de combinar diversidade e acurácia como funções objetivo (SANTOS; SABOURIN; MAUPIN, 2008), (LUSTOSAFILHO; CANUTO; XAVIER, 2015).

A solução ótima em uma otimização multi-objetivo pode ser definida a partir de um conceito matemático de ordenação parcial. No contexto de otimização multi-objetivo, o termo dominação é usado para esse propósito. Diz-se que uma solução C_j domina outra solução C_l , se as duas condições seguintes são verdadeiras (DEB, 2011):

- A solução C_j não é pior que C_l para todas as funções objetivos.
- A solução C_j é estritamente melhor que C_l em pelo menos um objetivo ($div(C_j)$ ou $acc(C_j)$).

Todos os pontos que não são dominados por outros definem uma fronteira, quando visualizados juntos em um espaço objetivo, e representam a "fronteira não dominada" ou "fronteira de Pareto" (DEB, 2011). A figura 5 mostra a fronteira de Pareto obtida ao utilizarmos as duas funções objetivo definidas acima na base de dados sintética Sine1. As soluções avaliadas na iteração 3155 são todas exibidas no espaço objetivo, sendo que as soluções não dominadas estão destacadas em azul (quadrado).

O algoritmo 4 resume a fase de seleção dinâmica do método DESDD. O algoritmo recebe os seguintes parâmetros de entrada: a população de comitês P criada na fase anterior; a quantidade de comitês c ; as acurácias prequenciais dos comitês ACC calculadas até a instância x_{i-1} , a instância a ser classificada x_i e a medida de diversidade MD a ser utilizada. Em seguida, para cada comitê é calculado o valor de diversidade para a instância x_i (linhas 2-5). É importante destacar que, caso o DESDD tenha sido configurado para empregar uma medida de diversidade pareada, como essas medidas são supervisionadas, o valor de diversidade dos comitês é calculado em janela deslizante composta pelas últimas

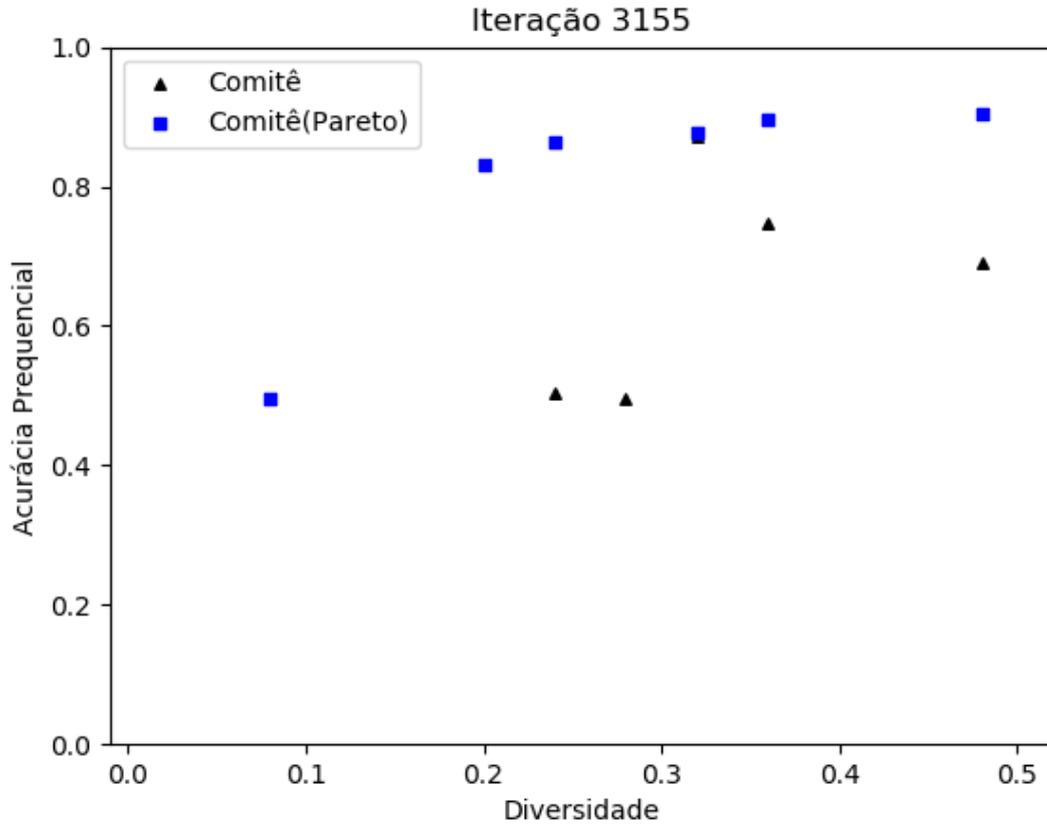


Figura 5 – Exemplo de uma fronteira de Pareto. Em azul(quadrado), as soluções da fronteira de Pareto da iteração 3155 da base Sine1. Em preto(triângulo), as soluções não pertencentes à fronteira. Fonte: O autor.

Algorithm 4 DESDD - Seleção Dinâmica baseada em Diversidade x Acurácia via fronteira de Pareto

INPUT: população de comitês (*ensembles*); quantidade de comitês c ; medida de diversidade MD; acurácias prequenciais ACC ; instância desconhecida x_i ; e janela de cálculo J .

- 1: **if** $J = 0$ **then** ▷ Não usa janela deslizando
 - 2: **for** ($j \leftarrow 0$; $j < c$; $j++$) **do**
 - 3: $DIV(C_j) \leftarrow MD.calcula(C_j, x_i)$
 - 4: **end for**
 - 5: **else** ▷ Calculado com janela deslizando
 - 6: **for** ($j \leftarrow 0$; $j < c$; $j++$) **do**
 - 7: $DIV(C_j) \leftarrow MD.calcula(C_j)$
 - 8: **end for**
 - 9: **end if**
 - 10: $PARETO \leftarrow EncontraParetoFront(Ensemble, DIV, ACC, c)$
 - 11: $C_j^* \leftarrow ES.seleciona(PARETO)$
 - 12: **Return** $votoMajoritario(C_j^*, x_i)$
-

J instâncias de treinamento (linhas 6-8). Na sequência, a função `EncontrarParetoFront` é invocada (linha 10) para encontrar as soluções não dominadas. Caso haja empate, é seguida a ordem em que os comitês foram inseridos durante a configuração do método. A função `EncontrarParetoFront` retorna as soluções da fronteira de Pareto ordenadas pelo valor da diversidade. Então, o comitê com maior diversidade é escolhido como a melhor solução C_j^* (linha 11). Por fim, o rótulo de x_i é definido por voto majoritário entre os classificadores membros do comitê C_j^* . Uma vez definida a classe a ser atribuída à instância x_i , segue-se para a fase de detecção de mudança.

4.3 Detecção de Mudança

Como o DESDD é um método de aprendizado *online*, espera-se que o rótulo da instância x_i torne-se disponível logo após a sua classificação feita na fase de seleção dinâmica. Tão logo o rótulo y_i da instância x_i seja disponibilizado, o método proposto entrará na fase de detecção de mudança. Nesse modo, será passada ao detector de mudança a predição do método e a indicação se o método fez uma predição correta ou não. A partir dessas informações, o detector de mudança indicará se houve mudança de conceito ou não. Quando uma mudança é detectada, o DESDD entra na fase de reação à mudança. Caso contrário, x_i e y_i são usados para atualizar a população de comitês por meio de treinamento *online*.

Portanto, qualquer detector de mudança supervisionado pode ser utilizado nessa terceira fase do DESDD. Neste trabalho, foram testados dois detectores: DDM e ADWIN. No entanto, outros detectores como EDDM e AUE2 podem ser empregados. Além disso, como o DESDD foi desenvolvido utilizando o framework MOA, pode-se utilizar qualquer detector disponível nesse framework, e até mesmo pode-se implementar novos detectores utilizando os recursos do framework.

A fase de detecção de mudança do DESDD foi ajustada para operar de duas formas. Caso o detector entre em nível de alerta, o método começa a armazenar instâncias para a etapa de reação à mudança, como pode ser observado no algoritmo 5, nas linhas 3-5. Essa estratégia de armazenamento é baseada no funcionamento do DDM. Caso seja utilizado outro método de detecção que não implemente a fase de alerta, como o ADWIN, as linhas 3-5 serão ignoradas.

4.4 Reação à Mudança

Por fim, caso uma mudança seja detectada, o DESDD executará a fase de reação. Novamente, nesta fase há várias possibilidades de se proceder com a reação. Conforme mencionado anteriormente, de acordo com o fluxo natural do processo do DESDD, apenas

Algorithm 5 DESDD-Detecção de Mudança()

INPUT: população de comitês (ensembles); acurácias frequenciais ACC; algoritmo de aprendizado online AO; detector de drift D; estratégia de reação ER; instância x_i ; rótulo y_i ; predição y_i^* de C_j^* ;

- 1: $predicao \leftarrow acertou(y_i^*, y_i)$
- 2: $retornoDetector \leftarrow D.submete(predicao)$
- 3: **if** $retornoDetector = "alerta"$ **then**
- 4: $armazenaNoBuffer(x_i)$
- 5: $AO.treinarEnsembles(ensembles, x_i)$
- 6: **else**
- 7: **if** $retornoDetector = "mudanca"$ **then**
- 8: $ER.executaReacao(ensembles, x_i)$
- 9: $limpaBuffer()$
- 10: **else**
- 11: $limpaBuffer()$
- 12: $ACC \leftarrow atualiza(ACC, ensembles, x_i)$
- 13: $AO.treinarEnsembles(ensembles, x_i)$
- 14: **end if**
- 15: **end if**
- 16: Return $ensembles, ACC$

um detector é gerado para toda a população de comitês. Porém, com base nos *baselines* investigados neste trabalho, foi testada também a possibilidade de gerarmos um detector por classificador membro de cada comitê. Esses dois diferentes cenários, considerando ainda os dois tipos de detectores investigados (DDM e ADWIN), bem como os dois algoritmos de geração *online* de população de comitês possibilitam inúmeras variações do método. Em termos de estratégias de reação, são testadas neste trabalho quatro diferentes abordagens.

- **Simple Reinicio(SR):** Caso haja um detector por classificador membro de cada comitê, se ocorrer uma mudança em qualquer detector, o pior classificador do comitê é reiniciado, juntamente com seu detector.
- **Reinicia Conjunto(RC):** Caso haja um detector por comitê, se houver uma mudança em qualquer detector, o comitê todo é reiniciado.
- **Reinicia Sistema(RS):** Se ocorrer uma detecção de mudança, todos os comitês são reiniciados.
- **Reinicia e Treina com Instâncias de Alerta(RB):** Caso haja mudança, todo o sistema é reiniciado e treinado com dados armazenados durante a fase de alerta do detector.

Dado que o Leveraging Bagging demanda um detector por classificador, o DESDD foi testado com as configurações dispostas resumidamente na tabela 9. No próximo serão discutidos os resultados obtidos nos experimentos comparativos entre essas versões.

Versão	Algoritmo	Detector	Reações
1	Leveraging Bagging	1 ADWIN p/ Class.	SR, RC e RS
2	Online Bagging	1 DDM p/ Class.	SR, RC e RS
3	Online Bagging	1 ADWIN	RS
4	Online Bagging	1 DDM	RS e RB

Tabela 2 – Versões do DESDD testadas nos experimentos. Essas versões são obtidas ao variarmos métodos de reação, algoritmos de aprendizado e detectores de mudança.

4.5 Considerações Finais

O método proposto neste capítulo utiliza uma população de comitês de classificadores e aplica seleção dinâmica via fronteira de Pareto, tendo diversidade e acurácia prequencial como funções objetivo. O DESDD apresenta como vantagem o fato de gerar uma população com elevada diversidade entre os classificadores, pois o método pode ser configurado para produzir vários comitês com taxa de diversidade variável. No entanto, ao lidar com populações de comitês a complexidade do DESDD é maior do que os métodos que utilizam apenas um comitê ou apenas um classificador.

Neste capítulo mostrou-se o método proposto, o qual envolve população de comitês de classificadores, seleção dinâmica com base em diversidade e acurácia prequencial e detecção de mudança de conceito de forma explícita e anotada. Os detalhes de cada etapa do método DESDD são descritos. No próximo capítulo são apresentados e discutidos os resultados obtidos nos experimentos.

5 Experimentos e Resultados

Este capítulo descreve os experimentos realizados neste trabalho, os quais estão divididos em quatro (04) séries. A primeira série investiga o uso da combinação de diversidade com acurácia como par de critérios de seleção dinâmica por meio do ranqueamento via fronteira de Pareto. Apenas uma medida de diversidade foi utilizada. A segunda série de experimentos compara diferentes medidas de diversidade no contexto da seleção dinâmica bi-objetivo (diversidade e acurácia prequencial). Em seguida, a terceira série investiga diferentes versões do DESDD, as quais são obtidas ao variar o método de geração da população de comitês, o algoritmo de detecção e a estratégia de reação à mudança de conceito. Por fim, a última série apresenta resultados de comparação entre a melhor versão do método proposto e *baselines*.

O DESDD foi implementado em Java usando o *Massive Online Analysis* (MOA) *framework* (BIFET et al., 2010), que possui a implementação dos principais métodos considerados estado-da-arte. Antes de apresentarmos os resultados, algumas informações sobre o protocolo experimental são discutidas na seção seguinte.

5.1 Protocolo Experimental

Nesta seção são detalhadas todas as bases utilizadas, os classificadores base investigados e as métricas de avaliação. Os experimentos foram realizados em um computador Intel Core i5-4690 3.50GHz com 8 GB de memória RAM.

5.1.1 Bases de dados

Foram realizados experimentos com quatro bases sintéticas, todas apresentando mudança de conceito, variando-se a presença de ruídos e a velocidade da mudança, e três bases reais. As bases sintéticas investigadas são: Sine1, base com mudança abrupta em que a classificação é invertida após a mudança de conceito, ou seja, o que caracterizava uma classe após a inversão caracterizará a outra; Circle, apresenta mudança gradual; Gauss, apresenta mudança abrupta com ruídos, sendo que, após cada mudança de conceito, a classificação é invertida; e Line, que apresenta mudança abrupta sem ruído. Todas essas bases são balanceadas e possuem duas classes. Cada bloco de 1000 exemplos representa um conceito (PINAGE; SANTOS, 2015).

Em termos de dados reais, as seguintes bases foram investigadas. Spam, que usa dados da coleção SpamAssasin, consiste em 9324 exemplos com 40000 atributos e acredita-se que essa base apresenta mudança de conceito gradual (KATAKIS; TSOUMAKAS;

VLAHAVAS, 2010). Elec, cujos dados foram coletados da *The Australian New South Wales Electricity Market*. Consiste no período de 2 anos de coleta com 45312 instâncias registradas a cada meia hora, 6 características de entrada e duas classes: UP ou DOWN. Por fim, KDD Cup 1999, uma base de dados relacionada à detecção de intrusões em rede. Essa base contém uma extensa variedade de intrusões simuladas em um ambiente de rede militar. A base de dados contém 489844 exemplos. Cada exemplo corresponde a uma conexão e possui 41 atributos de entrada, como tamanho da conexão, tipo de protocolo, serviço de rede no destino, etc. A classe alvo identifica quando a conexão é um ataque ou uma conexão normal.

5.1.2 Classificadores Base

Foram utilizados nos experimentos tanto conjuntos homogêneos quanto heterogêneos. Os conjuntos homogêneos foram formados pelo classificador Hoeffding Tree, o qual é um método de aprendizado online de grandes volumes de dados em fluxo proposto por Domingos e Hulten (2000). Já os conjuntos heterogêneos foram formados por: HoeffdingTree, kNN, RandomHoeffdingTree, Perceptron e NaiveBayes. Esses algoritmos foram escolhidos para formar conjuntos heterogêneos com classificadores de diversos tipos: árvores de decisão, estatísticos, bayesianos e redes neurais. Com isso, busca-se uma diversidade maior em comparação com os conjuntos homogêneos. Em todas as séries de experimentos a decisão de cada comitê é feita por voto majoritário.

5.1.3 Métricas de Avaliação

As medidas de avaliação utilizadas neste trabalho são as seguintes: acurácia, acurácia média e acurácia prequencial.

Acurácia:

$$Acc = \frac{Acertos}{Acertos + Erros} \times 100 \quad (5.1)$$

Acurácia média:

$$Acc_{media} = \frac{\sum_{i=1}^t Acc_i}{t} \quad (5.2)$$

Acurácia prequencial: A regra usada para obter a acurácia prequencial na iteração t é apresentada:

$$Acc_{media} = \begin{cases} acc_{ex}(t), & \text{se } t = f \\ acc(t-1) + \frac{acc_{ex}(t) - acc(t-1)}{t-f+1}, & \text{caso contrário} \end{cases} \quad (5.3)$$

Sendo que acc_{ex} é 0 se a predição do exemplo corrente ex antes do aprendizado estiver errada e 1 se estiver correta; e f é a primeira iteração usada no cálculo.

5.2 Comparação entre funções objetivo: Diversidade x Acurácia

Inicialmente foi feito um estudo para análise da possibilidade de uso de diversidade como único objetivo da seleção dinâmica de comitês de classificadores, bem como o impacto desse processo na acurácia de conjuntos formados por classificadores homogêneos e por classificadores heterogêneos. Para gerar conjuntos com diversos níveis de diversidade foi utilizado o método Online Bagging modificado por [Minku e Yao \(2012\)](#), descrito no [Capítulo 2](#). Tal algoritmo consiste em incluir o parâmetro λ , que é utilizado para variar a diversidade do conjunto gerado. Os valores de λ inicialmente estudados para cada base de dados foram: 5000, 1000, 500, 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005. Essa escala de valores foi ampliada do estudo presente em [Minku e Yao \(2012\)](#). O cálculo da diversidade foi feito por meio da ambiguidade, Equação 2.2 descrita no [Capítulo 2](#). A ambiguidade foi utilizada por ser uma medida de diversidade não dependente do rótulo das instâncias, ou seja, trata-se de uma medida totalmente não supervisionada. Esse estudo está detalhado no Apêndice [Estudo - Diversidade e Acurácia](#).

Os resultados desse estudo mostram que o valor da acurácia reduz bastante quando são selecionados tanto o comitê de classificadores que apresenta a maior diversidade quanto o que apresenta a menor diversidade, independentemente do tipo de mudança existente. Portanto, não é possível especificar um limiar fixo de diversidade que garanta elevada acurácia. No entanto, um ponto de partida para a seleção dinâmica do comitê que melhor classificará uma instância é escolher comitês cujo valor de diversidade seja próximo da menor diversidade e excluir comitês que possuem valor de diversidade próximo da máxima diversidade.

Além disso, esse estudo inicial mostrou que o uso de conjuntos heterogêneos, apesar de aumentar a taxa de diversidade, não produz aumento significativo da acurácia. Na prática, os conjuntos heterogêneos permitem que a variação de diversidade ocorra em maior escala, sem afetar diretamente a acurácia do sistema. Dessa forma, podemos concluir que, embora o uso da diversidade possa beneficiar a seleção de comitês de classificadores, combinar diversidade com um segundo critério, como a acurácia calculada através de janela deslizante ou a acurácia prequencial, pode tornar o processo de seleção dinâmica mais preciso. Diante desse resultado, nós decidimos combinar diversidade e acurácia como funções objetivos. Por tratar-se de um problema bi-objetivo, foi utilizada a teoria de fronteira de Pareto, conforme visto no capítulo anterior.

Nesta primeira série de experimentos, o tamanho da população de comitês foi igual a 11, ou seja, foi executado um experimento com 11 conjuntos heterogêneos e um

experimento com 11 conjuntos homogêneos. O tamanho de cada comitê foi fixado em 25 classificadores. Os valores de λ estudados para cada base de dados foram: 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001. Foram removidos os conjuntos que apresentaram pior desempenho no estudo discutido no Apêndice [Estudo - Diversidade e Acurácia](#). Por fim, como o objetivo desta série de experimentos é definir como escolher C_j^* dentre as soluções não dominadas encontradas, nas bases sintéticas foi utilizado um detector de mudanças do tipo oráculo - quando uma mudança ocorre, o sistema é totalmente reiniciado; e para as bases reais foi empregado o detector DDM. Os resultados obtidos são discutidos a seguir.

5.2.1 Resultados em bases sintéticas

Todas as bases apresentaram resultados muito similares. Em linhas gerais, se escolhermos como C_j^* o comitê com a menor diversidade ou com a maior diversidade (extremos) no início das iterações do método, o desempenho de classificação é baixo, principalmente para os conjuntos heterogêneos. Esse comportamento é ilustrado nas figuras 6, 7, 8 e 9, que mostram exemplos da fronteira de Pareto obtida para cada uma das quatro bases sintéticas investigadas em iterações relativamente iniciais e iterações mais avançadas. Por exemplo, para a base Line, a figura 6 mostra que nas iterações 50 e 1050 (heterogêneos e homogêneos), o comitê com menor diversidade e o comitê com maior diversidade (extremos) apresentam valores baixos de acurácia.

Se avaliarmos apenas as soluções pertencentes à fronteira de Pareto, observa-se que, no início do aprendizado de um conceito, a solução de maior diversidade apresenta a melhor acurácia média final. No entanto, com a consolidação do conceito (muitas iterações após uma mudança de conceito), essa diferença entre a acurácia dos dois comitês diminui bastante, como pode ser observado no exemplo da figura 6 (iteração 950 e 2000 - heterogêneos e homogêneos). Portanto, quando o conceito aprendido está bem consolidado, as soluções de menor e maior diversidade da fronteira de Pareto apresentam desempenhos similares. Esse mesmo comportamento foi apresentado em todas as demais bases testadas, principalmente para os comitês heterogêneos.

Portanto, de maneira geral, selecionar como C_j^* a solução pertencente à fronteira de Pareto que possui maior diversidade se mostra melhor do que escolher a de menor diversidade. Isso pode ser observado, por exemplo, nas iterações 950, 1050 e 10000 (conjuntos heterogêneos) e 50, 1050 e 10000 (conjuntos homogêneos) da base Gauss exibidas na figura 8.

As tabelas 3 e 4 mostram a média de acurácia geral alcançada para cada base. Com base nessas tabelas podemos concluir o seguinte:

- Line: a solução com maior diversidade possui também maior acurácia para os conjuntos heterogêneos, já para os conjuntos homogêneos, embora a diferença seja

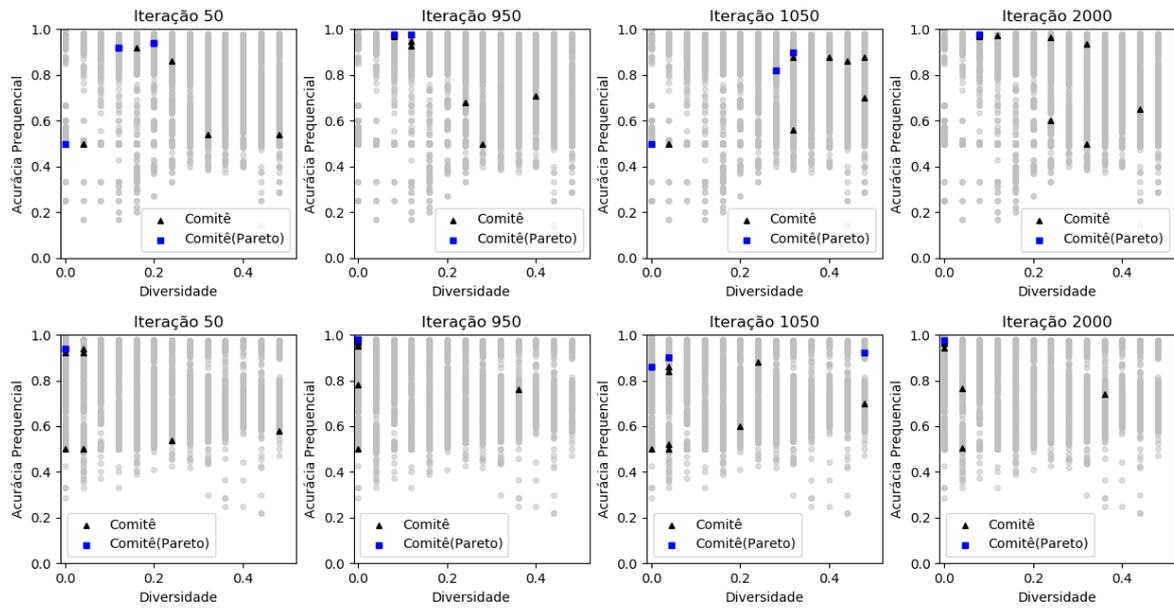


Figura 6 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 2000 (última instância) para os comitês heterogêneos (parte superior) e comitês homogêneos (parte inferior) para a base Line. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não à fronteira de Pareto de todas as demais iterações. Fonte: O autor.

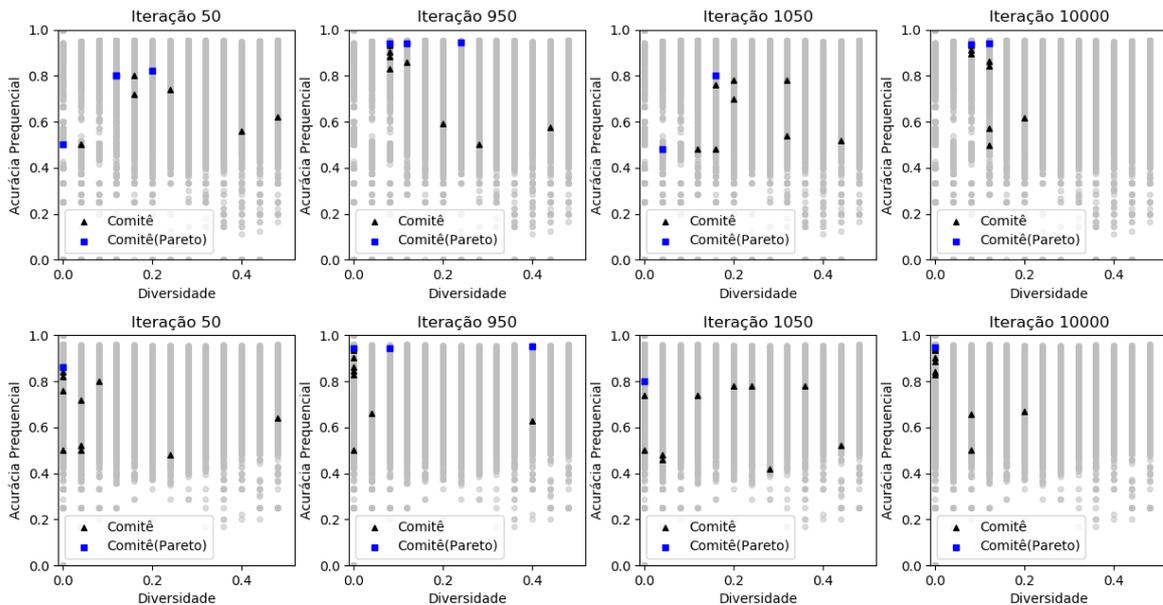


Figura 7 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 10000 (última instância) para os comitês heterogêneos (parte superior) e comitês homogêneos (parte inferior) para a base Sine1. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não à fronteira de Pareto de todas as demais iterações. Fonte: O autor.

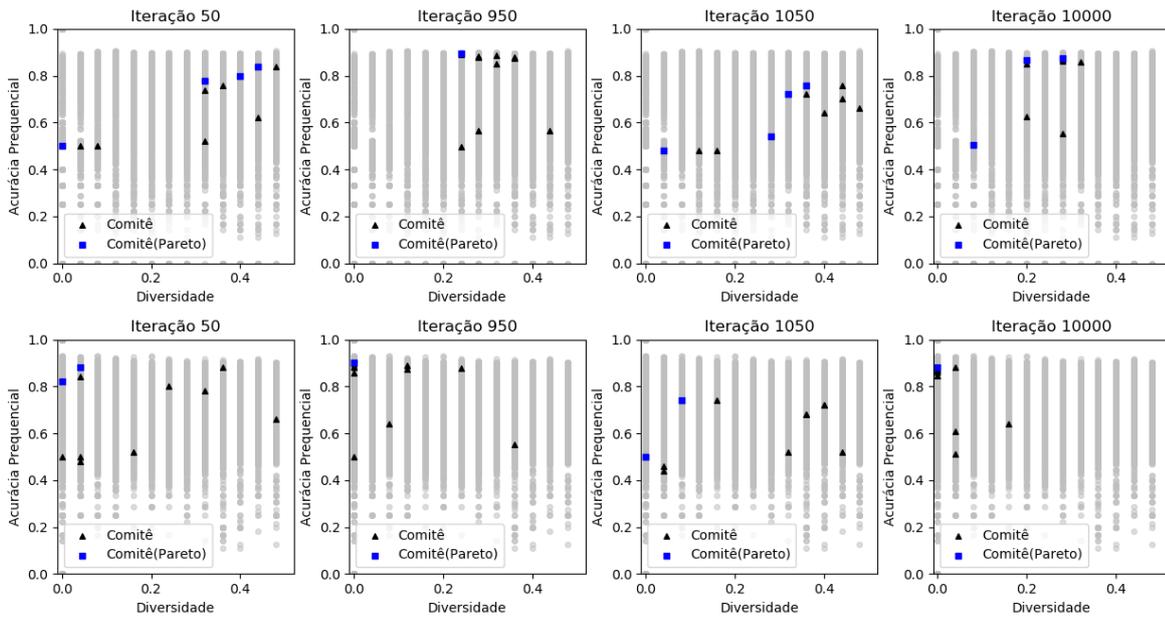


Figura 8 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 10000 (última instância) para os comitês heterogêneos (parte superior) e comitês homogêneos (parte inferior) para a base Gauss. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não a fronteira de Pareto de todas as demais iterações. Fonte: O autor.

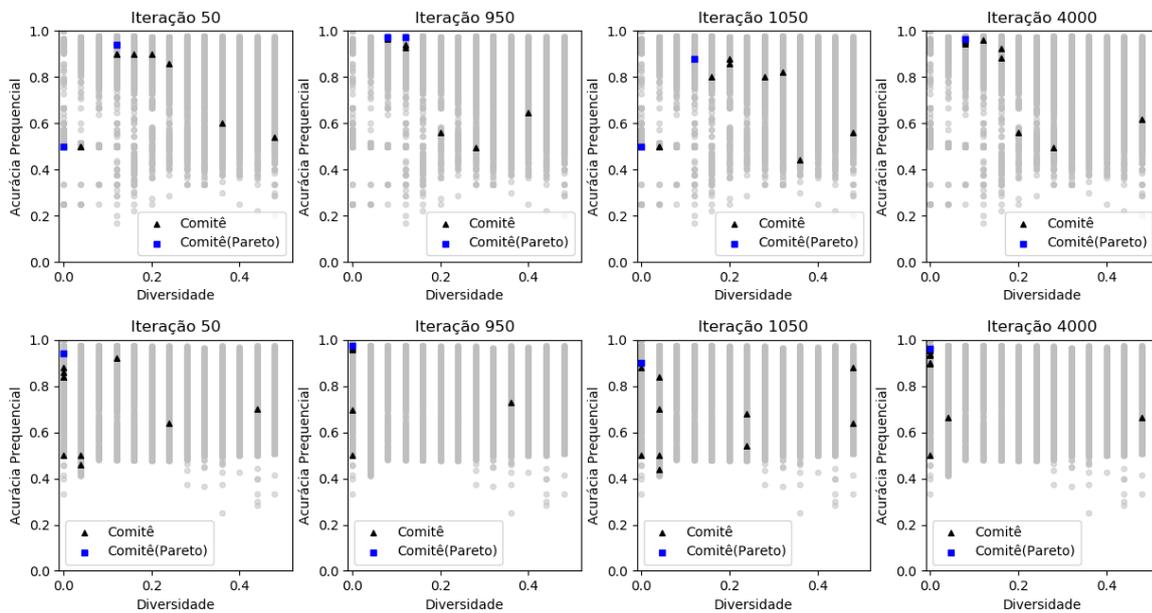


Figura 9 – Espaço de funções objetivo nas iterações: 50, 950, 1050 e 4000 (última instância) para os comitês heterogêneos (parte superior) e comitês homogêneos (parte inferior) para a base Circle. Em azul (quadrados) os comitês pertencentes à fronteira de Pareto na iteração. Em preto (triângulos) os comitês não pertencentes a fronteira de Pareto na iteração. Em cinza (círculos) os comitês pertencentes ou não a fronteira de Pareto de todas as demais iterações. Fonte: O autor.

pequena, a solução de menor diversidade apresenta a melhor acurácia média final. Portanto, para a base Line, dentre as soluções não dominadas, pode-se escolher como C_j^* o comitê que possui a maior diversidade.

- Sine1: a solução com maior diversidade possui também a maior acurácia, tanto para conjuntos heterogêneos quanto para os homogêneos. Portanto, para a base Sine1, nas soluções não dominadas deve-se escolher como C_j^* a solução que possui a maior diversidade.
- Gauss: em termos de média geral, a solução com maior diversidade possui maior acurácia. Portanto, para a base Gauss, também deve-se escolher como C_j^* a solução que possui maior diversidade dentre as soluções não dominadas.
- Circle: A solução com maior diversidade apresenta maior acurácia para os conjuntos heterogêneos. Para os conjuntos homogêneos, o comitê de menor diversidade apresentou melhor desempenho, embora tenha sido uma diferença muito pequena. Portanto, as conclusões alcançadas para as bases anteriores também se aplicam para a base Circle.

Solução	Line	Sine1	Gauss	Circle
Menor Diversidade	83,45	80,31	74,25	81,18
Maior Diversidade	97,65	94,67	88,78	96,65

Tabela 3 – Acurácia prequencial média dos conjuntos heterogêneos C_j^* escolhidos da fronteira de Pareto - bases sintéticas.

Solução	Line	Sine1	Gauss	Circle
Menor Diversidade	97,90	93,67	88,44	96,70
Maior Diversidade	97,85	95,34	89,05	96,50

Tabela 4 – Acurácia prequencial média dos conjuntos homogêneos C_j^* escolhidos da fronteira de Pareto - bases sintéticas.

Por fim, é possível observar que os comitês homogêneos alcançaram médias de acurácia equivalentes às médias obtidas pelos comitês heterogêneos em todas as quatro bases investigadas. Esses resultados indicam que a seleção dinâmica utilizando diversidade e acurácia via ranqueamento de fronteira de Pareto pode ser utilizada com ambos os tipos de comitês.

5.2.2 Resultados em bases reais

Com relação às bases reais, por não saber-se o exato momento da mudança de conceito, não foi possível fazer o *reset* dos conjuntos de classificadores no exato momento

da mudança (detector oráculo). Nesse caso, o *reset* foi feito com base na detecção do DDM. Para as bases reais, foi avaliado o resultado geral ao final da execução total, calculando-se média de diversidade e média de acurácia dos respectivos comitês com menor e maior diversidade que compunham a fronteira de Pareto, como mostram as tabelas 5 e 6. Com base nos resultados exibidos nessas tabelas, as principais conclusões são as seguintes:

- Spam: Da mesma forma como ocorreu nas bases sintéticas, o comitê com maior diversidade apresentou a melhor acurácia. Portanto, para a base Spam, dentre as soluções não dominadas deve-se escolher como C_j^* a solução que possui a maior diversidade. Em termos de diferença entre as taxas médias obtidas pelas soluções, a diferença foi ligeiramente menos acentuada do que nas bases sintéticas.
- Elec: Assim como ocorreu na base Spam, os resultados da base Elec mostram que a diferença entre as acurácias médias de ambos os comitês, com menor e maior diversidade, da fronteira de Pareto foi menos acentuada do que nas bases sintéticas. Também sobressaiu-se a solução de maior diversidade.
- KDDCup99: As acurácias médias da solução de menor diversidade e da solução de maior diversidade sobre a fronteira de Pareto foram similares. A diferença entre a acurácia das duas soluções foi a menor entre todos os testes realizados. Portanto, para a base KDDCup99, dentre as soluções não dominadas pode-se também escolher como C_j^* a solução com a maior diversidade.

Solução	Spam	Elec	KDDCup99
Menor diversidade	93,76	80,94	99,89
Maior Diversidade	97,32	87,52	99,95

Tabela 5 – Acurácia frequencial média dos conjuntos heterogêneos C_j^* escolhidos da fronteira de Pareto - bases reais

Solução	Spam	Elec	KDDCup99
Menor diversidade	96,24	88,01	99,93
Maior Diversidade	97,36	90,18	99,92

Tabela 6 – Acurácia frequencial média dos conjuntos homogêneos C_j^* escolhidos da fronteira de Pareto - bases reais

Como conclusões gerais dessa primeira série de experimentos, podemos destacar o seguinte. A seleção dinâmica de comitês de classificadores não deve ser conduzida a selecionar o comitê com menor diversidade, pertencente ou não à fronteira de Pareto. Caso contrário, será produzida uma taxa de classificação baixa. De acordo com as análises

realizadas ao longo das iterações, os comitês pertencentes ao conjunto das soluções não dominadas e que possuem maior diversidade, são também as soluções com maior acurácia frequencial, e, portanto, mostram-se mais eficazes para serem selecionadas. Isso é bem notável principalmente logo após a ocorrência de uma mudança de conceito. Significativamente após a mudança, os conjuntos com baixa diversidade tornam-se mais acurados. No entanto, na maioria dos casos não superam os conjuntos com maior diversidade pertencentes à fronteira de Pareto.

Como os experimentos apresentados nesta seção foram realizados apenas com ambiguidade como medida de diversidade, na segunda série de experimentos, diferentes medidas de diversidade são combinadas à acurácia frequencial para compor pares de função objetivo, a fim de selecionar comitês de classificadores de acordo com o ranqueamento provido pela fronteira de Pareto. Esses resultados são discutidos na próxima seção.

5.3 Comparação entre Medidas de Diversidades

Nesta seção são apresentados os resultados dos testes do método proposto DESDD variando-se as medidas de diversidade. Embora o módulo de detecção de mudança de conceito do DESDD possa ser composto por diferentes detectores, nos experimentos realizados nesta segunda série, foi utilizado apenas o detector DDM, dado que o objetivo aqui é identificar a melhor medida de diversidade. Além disso, no módulo de reação foi utilizada a mesma estratégia do DDM (armazenamento de instâncias na fase de alerta para treinamento após confirmação da mudança). Por sua vez, no módulo de geração de população de comitês foi utilizado o método Online Bagging.

Assim como na série de experimentos anterior, foram geradas duas populações de comitês: 1) heterogêneos; e 2) homogêneos. Cada população também foi composta por 11 comitês. Entretanto, como experimentos adicionais que avaliaram o impacto da quantidade de membros de cada comitê mostraram que entre 10 e 25 membros não há diferença significativa em termos de acurácia frequencial, optou-se por utilizar comitês menores. Portanto, a quantidade de membros de cada comitê foi definida como 10. Os valores de λ estudados para cada base de dados foram: 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, ou seja, foram os mesmos valores testados na seção anterior.

Com base nas conclusões apresentadas na série de experimentos anterior, à cada nova instância, o comitê com maior valor de diversidade, dentre as soluções não dominadas, foi selecionado para atribuir o rótulo à instância desconhecida x_i . As seguintes medidas de diversidade foram investigadas: ambiguidade, margem, *disagreement*, *double fault*, *gstatistics* e *correlation*. Em virtude da necessidade de uma base de instâncias para o cálculo das medidas de diversidade supervisionadas, foi utilizada janela deslizante $J = 100$ (representando 10% de um conceito de 1000 instâncias, por exemplo) para todas as medidas.

As tabelas 7 e 8 resumem os valores médios de acurácia obtidos ao compararmos as seis medidas de diversidade. A maior taxa de acurácia alcançada em cada base está destacada em negrito.

Base	Ambiguidade	Margem	QStat.	Correlation	Disag.	DF
Line	96,39	96,39	96,64	96,36	96,56	96,94
Sine1	92,63	92,63	92,82	92,45	92,60	92,32
Gauss	83,96	83,96	84,20	86,56	84,02	85,92
Circle	95,56	95,56	95,54	95,47	95,56	95,25
Spam	95,83	95,83	95,60	95,67	95,83	95,52
Elec	87,27	87,27	88,01	88,13	87,27	87,90
KDDCup99	99,96	99,96	99,97	99,97	99,97	99,97

Tabela 7 – Comparação entre medidas de diversidade - conjuntos heterogêneos.

Base	Ambiguidade	Margem	QStat.	Correlation	Disag.	DF
Line	96,70	96,70	96,12	96,53	96,70	96,11
Sine1	93,09	93,09	93,09	93,08	93,10	93,13
Gauss	86,85	86,85	86,73	86,87	86,86	86,86
Circle	95,56	95,56	95,54	95,47	95,56	95,25
Spam	95,83	95,83	95,60	95,67	95,83	95,52
Elec	90,25	90,25	90,57	90,47	90,22	89,97
KDDCup99	99,96	99,96	99,96	99,96	99,96	99,96

Tabela 8 – Comparação entre medidas de diversidades - conjuntos homogêneos.

Esses resultados mostram que, embora algumas medidas de diversidade tenham se sobressaído de forma discreta sobre as demais, não houve diferenças significativas entre as taxas de acurácia produzidas. Para as bases sintéticas, as seguintes medidas de diversidade foram ligeiramente melhores: ambiguidade, margem e *disagreement*, as quais apresentaram empate nas bases Line (conjuntos homogêneos) e Circle (conjuntos heterogêneos e homogêneos). Para as bases reais, a medida de diversidade que apresentou maior acurácia foi *disagreement*, que foi superior nas bases Spam (conjuntos heterogêneos e homogêneos) e KDDCup99 (conjuntos heterogêneos e homogêneos). Porém, essas diferenças não são consideradas significativas. Como consequência, o módulo de seleção do DESDD pode ser realizado com qualquer uma das seis medidas testadas. É importante destacar que esses resultados confirmam a literatura, pois as medidas de diversidade são consideradas extremamente correlacionadas entre si (BROWN; KUNCHEVA, 2010) e (SANTOS; SABOURIN; MAUPIN, 2008).

Por outro lado, uma vantagem de se utilizar medidas de diversidades não supervisionadas é a não necessidade de se manter uma base de instâncias rotuladas para realizar o seu cálculo. Portanto, a conclusão geral desta segunda série de experimentos é que, tanto a

medida ambiguidade quanto a medida margem devem ser utilizadas como função objetivo no módulo de seleção dinâmica do método proposto. Uma vez definido esse parâmetro, o DESDD tem ainda três parâmetros a definir: algoritmo de geração da população de comitês, detector de mudança de conceito e estratégia de reação à mudança. Esses parâmetros são estudados na próxima seção.

5.4 Análise comparativa entre diferentes versões do método proposto

O módulo de geração de população de comitês do DESDD pode ser implementado por qualquer método de geração *online* de comitês. Entretanto, essa escolha pode determinar a forma como os módulos de detecção e de reação à mudanças funcionarão. Dois algoritmos de geração de comitês são testados nesta terceira série de experimentos: 1) Leveraging Bagging; e 2) Online Bagging. O algoritmo Leveraging Bagging opera associado ao detector ADWIN, dessa forma, para o módulo de detecção são testados também dois algoritmos: 1) ADWIN; e 2) DDM. Por fim, quatro estratégias de reação foram testadas no módulo de reação à mudança:

- **Simples Reinício (SR):** Essa estratégia segue o procedimento padrão do Leveraging Bagging, ou seja, é utilizada apenas quando cada classificador membro dos comitês possui um detector associado. Nesse caso, quando uma mudança é indicada por qualquer detector de um comitê, o pior classificador desse comitê é retreinado, juntamente com seu detector.
- **Reinicia Comitê (RC):** Também utilizado quando há um detector associado a cada classificador membro dos comitês. Caso haja mudança em qualquer detector dos membros de um comitê, todos os classificadores membros desse comitê são reiniciados.
- **Reinicia Sistema (RS):** Caso haja mudança detectada, todos os comitês são reiniciados.
- **Reinicia e Treina com Instâncias de Alerta (RB):** Essa estratégia segue o procedimento utilizado pelo detector DDM. Nesse caso, se uma mudança é detectada, todo o sistema é reiniciado e treinado com dados armazenados durante a fase de alerta do detector.

Considerando as variações possíveis entre algoritmos de geração, detecção e reação, além do fato de serem criadas populações de 11 comitês, sendo cada comitê composto por 10 classificadores, criou-se as versões do DESDD apresentadas na tabela 9.

Os valores de λ estudados para cada base de dados também foram: 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001. Por outro lado, nesta e na próxima série de experimentos

Versão	Algoritmo	Detector	Reações
1	Leveraging Bagging	110 ADWIN (1 p/ classificador)	SR, RC e RS
2	Online Bagging	110 DDM (1 p/ classificador)	SR, RC e RS
3	Online Bagging	1 ADWIN	RS
4	Online Bagging	1 DDM	RS e RB

Tabela 9 – Diferentes versões do DESDD investigadas na terceira série de experimentos.

foram usados apenas conjuntos homogêneos, compostos por membros base do tipo Hoeffding Tree, uma vez que os *baselines* investigados na quarta série de experimentos utilizam apenas classificadores Hoeffding Tree.

Devido à natureza do método Leverage Bagging, cada classificador membro do comitê deve ter seu detector, mais precisamente, o detector ADWIN. Portanto, para poder utilizar o Leverage Bagging como método de geração de conjuntos de comitês, o DESDD deverá ter um detector por classificador. Consequentemente, para comparar o Leverage Bagging com o Online Bagging, a segunda versão do DESDD utiliza Online Bagging e um detector por classificador, no entanto, utilizou-se nessa versão o detector DDM. Para seguir o processo de fluxo natural da arquitetura do método DESDD, e também reduzir a complexidade do método proposto, a terceira e a quarta versão do DESDD consistem em usar apenas um detector para todo o sistema, sendo que, ADWIN é utilizado na versão três, enquanto DDM é testado na versão quatro. Nessas duas últimas versões, apenas o Online Bagging é empregado como gerador de comitês por não precisar criar um detector por classificador membro.

Considerando esse contexto, para as versões três e quatro não é possível aplicar as estratégias SR e RE, pois essas estratégias necessitam de um detector para cada classificador. Também não é possível utilizar a estratégia RB nas versões 1, 2 e 3, pois essa estratégia demanda apenas um detector que possua fase de alerta (como o DDM). Em virtude da medida de ambiguidade ter alcançado bom desempenho nos resultados da série de experimentos dois, e por ser uma medida de diversidade não supervisionada, essa foi a medida de diversidade escolhida para os experimentos a partir desta seção.

As taxas de acurácias e os tempos de execuções obtidos na terceira série de experimentos são descritas nas tabelas 10 e 11, para as bases sintéticas e reais, respectivamente. O melhor resultado está destacado em negrito. Nas colunas das acurácias (*Acc*) estão sendo exibidas apenas as acurácias das reações que tiveram melhores resultados para as versões do método. As siglas do melhores resultados estão entre parênteses. Da mesma forma, nas colunas de tempos (*Tem.*) estão relacionados os tempos referentes as melhores reações das respectivas versões.

As principais conclusões obtidas a partir desses resultados são as seguintes:

Versão	Line		Sine1		Gauss		Circle	
	Acc	Tem.	Acc	Tem.	Acc	Tem.	Acc	Tem.
1	96,23(RS)	0,69	92,12(RS)	4,44	86,94(RS)	5,74	95,43(SR)	2,92
2	96,54(SR)	1,93	92,65(SR)	7,75	85,89(RC)	10,70	95,69(SR)	3,27
3	96,19(RS)	1,64	90,61(RS)	9,49	86,10(RS)	10,12	95,35(RS)	3,18
4	96,70(RB)	1,62	93,09(RB)	8,72	86,85(RB)	9,98	95,56(RB)	3,09

Tabela 10 – Resultado dos testes com as quatro diferentes versões do método - bases sintéticas.

Versão	Spam		Elec		KDDCup99	
	Acc	Tempo	Acc	Tempo	Acc	Tempo
1	96,61(SR)	238,05	90,71(SR)	56,15	99,97(SR)	546,06
2	96,79(SR)	229,87	92,06(SR)	67,24	99,97(SR)	820,18
3	94,99(RS)	234,92	90,19(RS)	71,83	99,96(RS)	780,43
4	96,29(RB)	216,87	90,64(RS)	72,17	99,96(RB)	843,98

Tabela 11 – Resultado dos testes com as quatro diferentes versões do método - bases reais.

- Para a base Line (mudança abrupta sem ruídos), não houve uma grande diferença entre as médias de acurácia obtidas pelas 4 versões investigadas. Porém, a versão 4: Online Bagging + DDM, em conjunto com a estratégia de reação que retreina os comitês com as instâncias armazenadas durante a fase de alerta do detector de mudança, produziu a maior taxa de acurácia.
- A mesma configuração do DESDD que obteve maior taxa de acurácia para a base Line também foi superior para a base Sine1 (mudança abrupta onde a classificação é invertida após a mudança). Neste caso, as diferenças entre as taxas de acurácia alcançadas por cada versão foram mais elevadas.
- Para a base Gauss (mudança abrupta com ruídos), a versão 1, que combina Leveraging Bagging + ADWIN, juntamente com a estratégia que reinicia todo o sistema do zero, foi ligeiramente superior às demais. Essa também foi a estratégia que apresentou menor tempo de execução.
- No caso da base Circle (mudança gradual), todas as quatro versões obtiveram taxas de acurácia muito similares. A maior taxa foi alcançada pela versão 2 do DESDD, combinada à estratégia de reação que reinicia apenas o pior classificador do comitê que detectar a mudança de conceito.
- Para as bases reais, as diferenças entre as quatro versões do método foram mais significativas, exceto pela base KDDCup99. Porém, a versão vencedora foi a mesma para as 3 bases, isto é, versão 2 combinada ao reinício do pior classificador em cada comitê que apresentar mudanças de conceito.

Portanto, os resultados obtidos pelas diferentes versões do método proposto são, de maneira geral, bem equivalentes em termos de acurácia média. Por outro lado, as diferenças são mais significativas em termos de tempo de execução. No intuito de encontrar o melhor compromisso entre acurácia, tempo de execução e simplicidade, a versão 4 do DESDD, combinada com a estratégia RB, será utilizada para a comparação com os *baselines* na próxima série de experimentos. Com isso, os três parâmetros investigados nesta seção foram definidos da seguinte forma: método de geração de população de comitês: Online Bagging; detector de mudança de conceito: DDM; estratégia de reação à mudança: reiniciar todo o sistema, retreinando a nova população de comitês com as instâncias armazenadas durante a fase de alerta do DDM.

5.5 Comparação com *baselines*

Nesta última série de experimentos, o método proposto foi comparado a dois *baselines*. Foram calculados a acurácia e o tempo de execução de cada método nas bases sintéticas e reais apresentadas anteriormente. A metodologia usada tanto no DESDD quanto nos demais métodos foi *Test Than Train*.

O DESDD é comparado aos seguintes *baselines*:

- DDM: Drift Detection Method
- LB: Leveraging Bagging Method

O objetivo desses experimentos é verificar o impacto do uso da seleção dinâmica de comitês de classificadores guiada por diversidade combinada com acurácia prequencial no intuito de detectar mudanças de conceito. Por isso, os *baselines* escolhidos são os mesmos utilizados como componentes do método proposto, porém, em suas versões originais. Todos os métodos foram configurados com o classificador Hoeffding Tree como algoritmo de aprendizado base. Para o DESDD foram utilizados 11 conjuntos homogêneos, todos com 10 classificadores. Os valores de λ dos conjuntos são: 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001. Em relação aos *baselines*, estes foram empregados em suas versões padrão, ou seja, Leveraging Bagging: 10 classificadores Hoeffding Tree, $\lambda = 6$ e cada classificador com um ADWIN; DDM: 1 classificador Hoeffding Tree.

Para prover uma comparação mais robusta, foi realizado o teste estatístico Wilcoxon (DEMŠAR, 2006) entre os pares: (DESDD, DDM), (DESDD e Leveraging Bagging) e (DDM e Leveraging Bagging). Para cada base de dados e métodos foram extraídos 30 amostras após dividir os resultados em 30 partições iguais. Foi utilizado um p -value de 0,05. As tabelas 12 e 13 relatam as acurácias médias e os tempos totais de execução dos métodos estudados para as bases sintéticas e reais, respectivamente. Acurácia é medida

pela porcentagem final de exemplos corretamente classificados através do teste do método, exemplo por exemplo, enquanto o tempo é medido em segundos.

Método	Line		Sine1		Gauss		Circle	
	Acc	Tem.	Acc	Tem.	Acc	Tem.	Acc	Tem.
DESDD	96,70**	2,60	93,09**	9,99	86,85**	11,21	95,56	4,64
DDM	96,19	0,01	89,15	0,04	80,74	0,04	95,82*	0,02
Lev. Bagging	95,93	0,05	84,75	0,33	76,54	0,32	93,92	0,12

Tabela 12 – Comparação entre os métodos - bases sintéticas. A maior taxa de acurácia obtida para cada base está destacada em negrito. O símbolo (*) indica diferença estatística significativa para um dos métodos, enquanto o símbolo (**) indica diferença estatística significativa para dois métodos.

Método	Spam		Elec		KDDCup99	
	Acc	Tempo	Acc	Tempo	Acc	Tempo
DESDD	96,29**	183,93	90,25*	81,47	99,9637**	668,48
DDM	91,00	3,09	85,97	0,43	99,9564	8,06
Lev. Bagging	94,35	14,40	90,25*	2,67	99,9630	45,99

Tabela 13 – Comparação entre os métodos - bases reais. A maior taxa de acurácia obtida para cada base está destacada em negrito. O símbolo (*) indica diferença estatística significativa para um dos métodos, enquanto o símbolo (**) indica diferença estatística significativa para dois métodos.

Essas tabelas mostram que, em termos de acurácia, o DESDD é estatisticamente superior aos *baselines* em três, dentre as quatro bases sintéticas investigadas. É importante destacar que essas três bases apresentam mudança abrupta. Portanto, pode-se afirmar que o uso do DESDD é mais indicado quando se tem mudanças abruptas. Além disso, o DESDD ajuda a tornar o sistema mais robusto a ruídos, dado que a maior diferença de acurácia foi obtida na base Gauss, a qual apresenta ruídos. Nas bases reais, as diferenças entre as taxas de acurácia obtidas pelos três métodos comparados foi menor. Porém, novamente o DESDD superou os *baselines* em duas, dentre as três bases reais testadas. Quando se compara DDM com o Leveraging Bagging, nossos resultados confirmam a literatura, pois o DDM é melhor nas bases sintéticas, enquanto o Leveraging Bagging é superior nas bases reais.

Quanto à comparação de complexidade computacional, a ordem entre os três métodos empregados, considerando custo crescente, é: DDM, Leveraging Bagging e DESDD. Esse resultado era esperado, dado que DDM utiliza apenas um classificador, enquanto o Leveraging Bagging emprega um comitê com 10 classificadores, e o DESDD, uma população com 11 comitês, sendo cada comitê formado por 10 classificadores. Por outro lado, o aumento significativo de acurácia justifica esse aumento de complexidade.

Para identificar e analisar com detalhes os fatores que levam o método proposto a alcançar taxa de acurácia superior às taxas obtidas pelos *baselines*, as figuras 10, 11, 12, 13, mostram as curvas de aprendizado dos modelos testados para nas bases sintéticas. Como as mudanças são conhecidas nessas bases, cada figura também mostra o momento em que ocorre uma mudança (linha azul na vertical) e o momento em que cada método detecta a mudança. Como pode ser observado, o DESDD apresenta melhor desempenho em termos de detecção também, pois, detecta praticamente todas as mudanças e com menor atraso.

A tabela 14 resume os resultados dos testes de detecção em detalhes, tanto para as bases sintéticas quanto para as reais. São exibidas quatro métricas: quantidade de mudanças detectadas; quantidade de falsas detecções; quantidade de detecções perdidas; e taxa média de atraso. Esses resultados mostram uma superioridade indiscutível do método proposto. O DESDD produziu a menor taxa de falsa detecção. Essa métrica é importante porque uma falsa detecção leva à reinicialização desnecessária do sistema. Por outro lado, o Leveraging Bagging apresentou a mais elevada taxa de falsa detecção. Em termos de detecções perdidas, nenhum dos três métodos deixou de detectar as mudanças. Porém, o DESDD alcançou a menor taxa média de atraso de detecção, ou seja, foi o método que detectou a mudança de forma mais precisa. Esse resultado é importante, pois permite que o sistema reaja rapidamente à mudança, sem que esta cause grande perda de acurácia.

Diante desses resultados, é possível perceber que o uso de população de comitês de classificadores em combinação com seleção dinâmica de comitês, permite que métodos de detecção de mudanças reduzam atrasos de detecção e taxas de falsas detecções. Dessa forma, a acurácia geral desses métodos aumenta, bem como sua robustez a ruído (que pode levar à falsa detecção), e uma reação ocorre somente quando de fato há uma mudança de conceito.

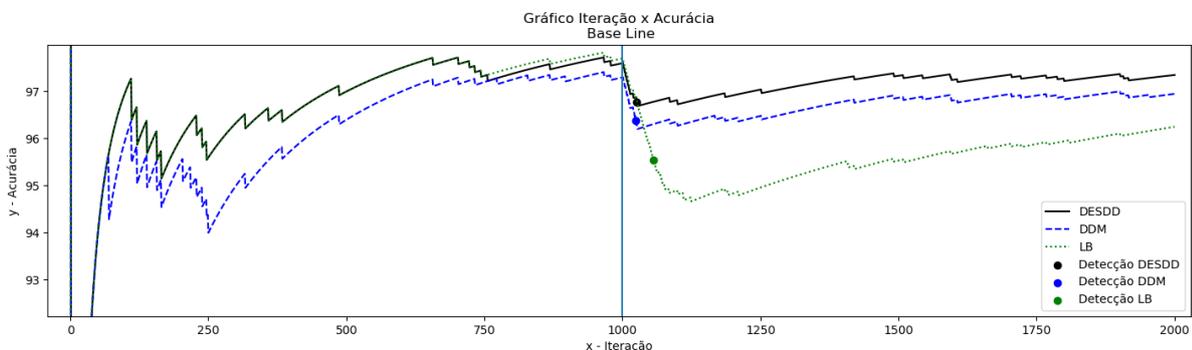


Figura 10 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Line. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.

Em relação às bases reais, não é possível identificar o momento exato de ocorrência

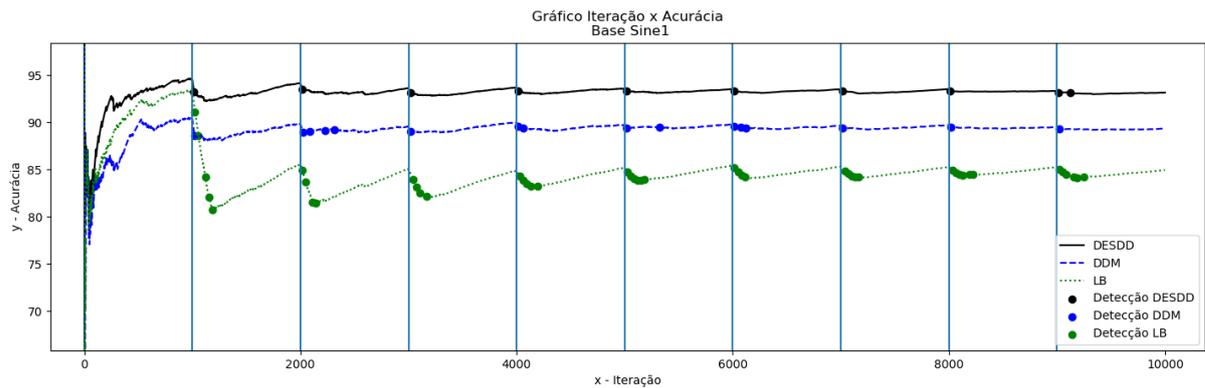


Figura 11 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Sine1. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.

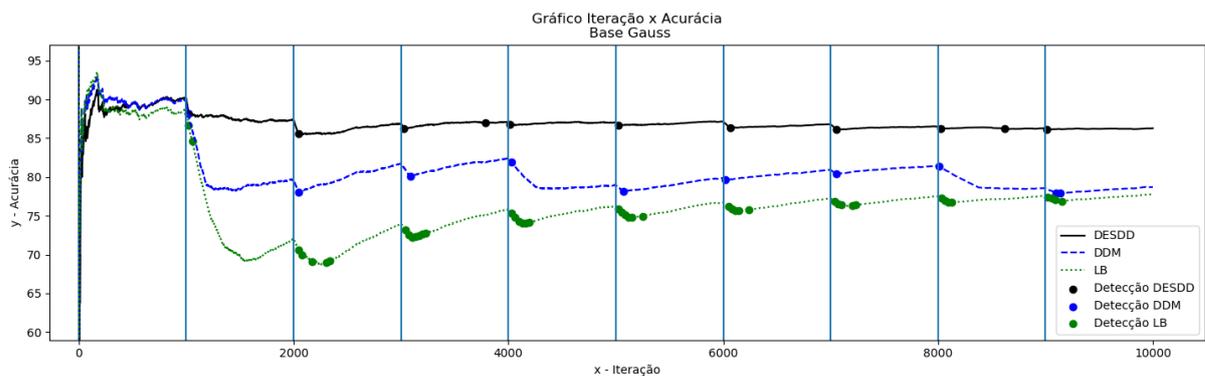


Figura 12 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Gauss. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.

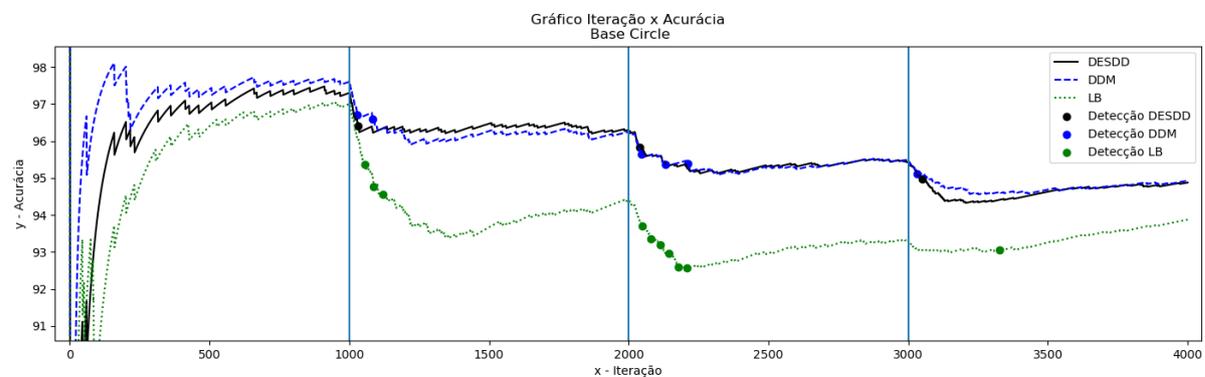


Figura 13 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Circle. O círculo nos gráficos indica o momento de detecção de mudança realizada pelos métodos e a linha azul vertical indica o momento exato da ocorrência de uma mudança. Fonte: O autor.

Base	DESDD				DDM				LB			
	D	FD	DP	TMA	D	FD	DP	TMA	D	FD	DP	TMA
Line	1	0	0	26,00	1	0	0	24,00	1	0	0	56,00
Sine1	10	1	0	15,22	16	7	0	19,00	45	36	0	27,56
Gauss	11	2	0	32,89	10	1	0	47,56	44	35	0	34,67
Circle	3	0	0	39,33	6	3	0	35,67	10	7	0	144,00
Spam	4	-	-	-	40	-	-	-	10	-	-	-
Elec	34	-	-	-	187	-	-	-	204	-	-	-
KDD	46	-	-	-	49	-	-	-	45	-	-	-
Acumulado	109	3	0	113,44	309	11	0	126,23	359	78	0	262,23

Tabela 14 – Resultado dos testes de detecção para as bases sintéticas e reais. Siglas: (D) Quantidades de mudanças detectadas; (FD) Quantidade de falsa detecções; (DP) Quantidade de detecções perdidas; (TMA) Taxa Média de Atraso.

$$TMA = \frac{\sum Atrasos}{Num.Mudanças}.$$

de uma mudança. Apesar disso, as figuras 14, 15 e 16 mostram a curva de acurácia obtida ao longo do tempo por cada um dos métodos investigados. Como pode ser observado nessas figuras, o DESDD supera os dois outros métodos na base Spam, e apresenta resultados similares ao Leveraging Bagging nas demais bases reais.

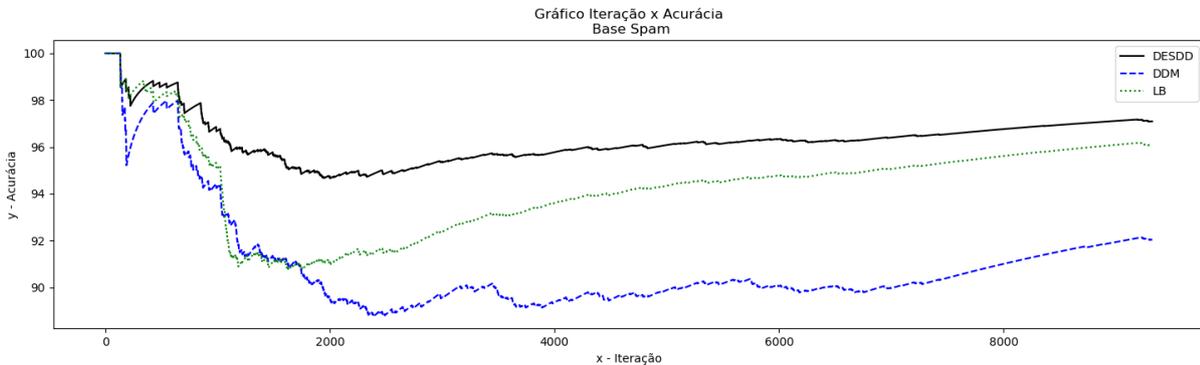


Figura 14 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Spam. Fonte: O autor.

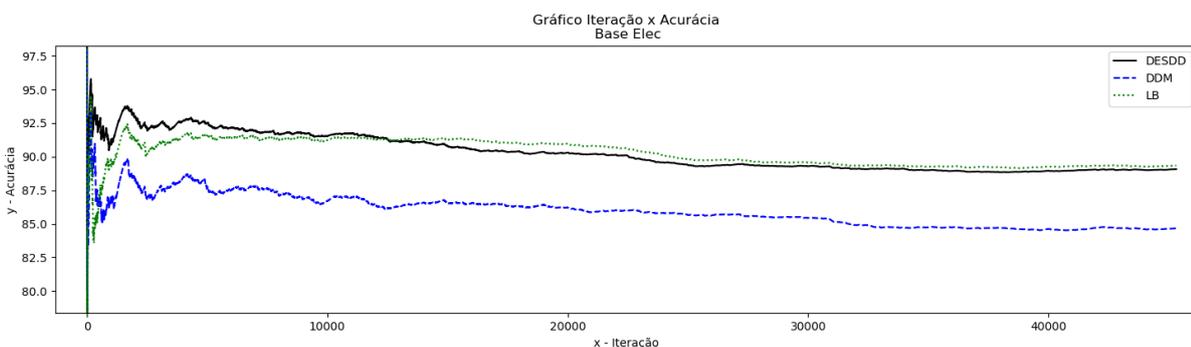


Figura 15 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base Elec. Fonte: O autor.

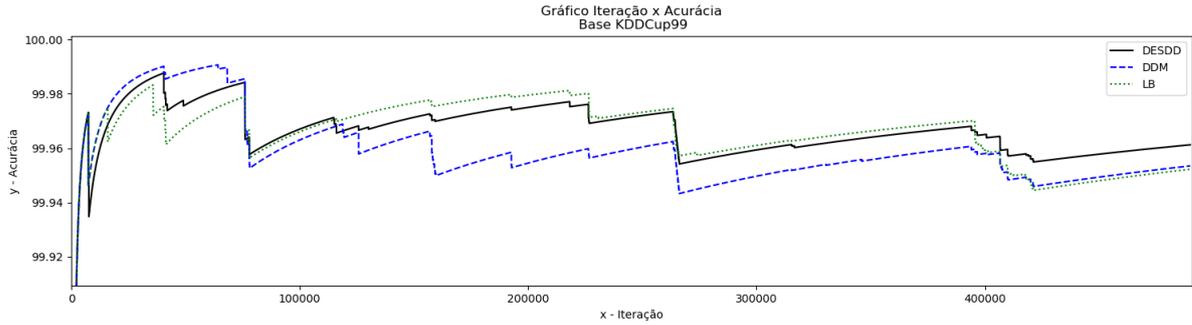


Figura 16 – Gráficos de iteração x acurácia dos métodos DESDD, DDM e LB para a base KDDCup99. Fonte: O autor.

De maneira geral, os resultados obtidos nesta última série de experimentos mostram que o método proposto foi superior aos *baselines* na maioria dos testes. Apesar de exigir maior tempo de execução, por utilizar um número maior de classificadores e mecanismos necessários para realizar a seleção dinâmica, o DESDD apresenta um ganho significativo de acurácia em relação ao DDM e ao Leveraging Bagging, além de ser superior em detecção de mudanças correta e realizada no momento certo.

5.6 Considerações Finais

Neste capítulo mostrou-se os experimentos e resultados obtidos. Foram realizadas quatro séries de experimentos em quatro bases sintéticas e três bases reais: análise de Pareto (diversidade e acurácia); emprego de diferentes medidas de diversidade, análise comparativa entre as diferentes versões do método proposto; e comparação do método com os *baselines*. Os resultados mostraram que o uso de seleção dinâmica de comitês de classificadores torna o DESDD melhor do que os *baselines*.

6 Conclusões

Neste trabalho foi proposto o método *online* DESDD para detecção e reação à mudança de conceito, o qual utiliza seleção dinâmica de comitês de classificadores guiada por diversidade e acurácia. O método mantém uma população de comitês de classificadores construída por meio de algoritmo de geração *online* de comitês com diversidade variável. A cada instância a ser classificada, o método utiliza seleção dinâmica para escolher o comitê que classificará a instância individualmente, de acordo com a diversidade e a acurácia de cada comitê. Os comitês são ranqueados por meio da estratégia de fronteira de Pareto. O método proposto também realiza detecção explícita de mudança. Quando uma mudança é detectada, é ativado um mecanismo de reação à mudança com o intuito de adequar os comitês a aprender o novo conceito.

A fim de definir os parâmetros do método proposto, estudou-se uma forma de selecionar dinamicamente o melhor comitê de classificadores dentre os membros de uma população de comitês com níveis de diversidades variáveis. Primeiramente, concluiu-se com base em experimentos que não é possível especificar um limiar fixo de diversidade que garanta elevada acurácia. Esse resultado levou à conclusão de que é necessário combinar diversidade com pelo menos outro critério de busca para tornar o processo de seleção dinâmica mais preciso. Em vista disso, diversidade e acurácia foram combinadas como função objetivo por meio de ranqueamento feito por fronteira de Pareto. Essa estratégia garantiu melhores resultados de seleção na maioria das bases investigadas. Adicionalmente, algumas medidas de diversidades foram testadas, incluindo medidas supervisionadas e não supervisionadas. Os experimentos mostraram pouca diferença nos resultados alcançados por cada medida investigada. Porém, deve-se destacar que medidas não supervisionadas podem ser consideradas uma melhor opção, porque não necessitam de uma base de instâncias rotuladas para a realização de seu cálculo. Por fim, dentre os comitês pertencentes à fronteira de Pareto (soluções não dominadas), este trabalho indica que é melhor escolher o comitê com maior diversidade como *expert* para classificar a instância corrente, pois essa estratégia apresenta melhores desempenhos em relação à acurácia média final.

Como o método proposto é dividido em quatro fases: geração da população de comitês; seleção dinâmica do comitê *expert*; detecção de mudanças; e reação à mudança, diferentes versões desse método podem ser obtidas em função da definição de parâmetros de cada fase. Com isso, foram feitas análises comparativas entre diferentes versões do DESDD, variando-se os métodos de geração *online* de população de comitês, detectores de mudança de conceito e estratégias de reação à mudança. Os resultados dos experimentos mostraram que, de maneira geral, as versões estudadas são bem equivalentes em termos de acurácia média final. Porém, a versão que apresentou melhor desempenho tem a seguinte

configuração: Online Bagging na primeira fase, o detector DDM na terceira fase e, na quarta fase, a estratégia de reação que reinicia toda a população de comitês por meio do treinamento desses comitês com instâncias armazenadas durante a estapa de alerta do DDM.

A melhor versão do DESDD foi então comparada a dois *baselines*: DDM - que utiliza um único classificador; e Leveraging Bagging - que utiliza um comitê de classificadores. Com base nos resultados dos experimentos de comparação com os *baselines*, pode-se observar que o DESDD alcançou melhores taxas de acurácia, menores taxas de atraso, não deixou de detectar as mudanças conhecidas nas bases e produziu reduzidas taxas de falsa detecção. Portanto, o método proposto superou os *baselines*. Por outro lado, conforme já esperado, o DESDD apresenta maior complexidade computacional. Porém, nós podemos assumir que os ganhos nas demais métricas superam o aumento em complexidade, dado o contexto de detecção explícita de mudança para obter-se uma rápida reação à mudança de conceitos.

6.1 Trabalhos Futuros

Embora a ambiguidade e a margem não necessitem dos rótulos para o cálculo da função objetivo diversidade, a segunda função objetivo, acurácia frequencial, necessita de dados rotulados para ser calculada. Esse fator torna o módulo de seleção dinâmica do DESDD parcialmente dependente de algumas instâncias rotuladas para funcionar adequadamente. Portanto, um dos trabalhos futuros envolve a pesquisa referente à combinação da função objetivo diversidade com outras funções objetivo que não sejam dependentes dos rótulos.

Outra questão que deve ser investigada é aprimorar os métodos de geração de comitês com diversidades variadas. Nesse sentido, é possível testar diferentes formas de introdução de diversidade durante o processo de geração da população de comitês. Por fim, a criação de uma medida de diversidade frequencial, ou seja, que acumule os valores passados, da mesma forma que a acurácia frequencial funciona, pode ser muito útil no contexto de uso de diversidade em algoritmos de aprendizado *online*.

Referências

- ALMEIDA, P. R. L. D.; OLIVEIRA, L. S.; SOUZA, A. D.; JR, B.; SABOURIN, R. Handling Concept Drifts Using Dynamic Selection of Classifiers. n. October, 2016. Citado na página 37.
- BIFET, A.; HOLMES, G.; PFAHRINGER, B. Leveraging bagging for evolving data streams. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 6321 LNAI, n. PART 1, p. 135–150, 2010. Citado 2 vezes nas páginas 34 e 36.
- BIFET, A.; HOLMES, G.; PFAHRINGER, B.; KRANEN, P.; KREMER, H.; JANSEN, T.; SEIDL, T. MOA: Massive Online Analysis, a Framework for Stream Classification and Clustering. *HaCDAIS 2010*, v. 11, n. November 2016, p. 3, 2010. Citado na página 49.
- BRITTO, A. S.; SABOURIN, R.; OLIVEIRA, L. E. S. Dynamic selection of classifiers - A comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665–3680, 2014. Citado 4 vezes nas páginas 13, 22, 30 e 31.
- BROWN, G.; KUNCHEVA, L. I. “Good” and “Bad” Diversity in Majority Vote Ensembles. p. 124–133, 2010. Citado 3 vezes nas páginas 22, 31 e 58.
- BRZEZINSKI, D.; STEFANOWSKI, J. Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, v. 25, n. 1, p. 81–94, 2014. Citado 5 vezes nas páginas 21, 23, 26, 29 e 35.
- CHEN, H.; MA, S.; JIANG, K. Detecting and adapting to drifting concepts. *Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012*, n. Fskd, p. 775–779, 2012. Citado 4 vezes nas páginas 23, 25, 29 e 35.
- DEB, K. Multi-objective optimization using evolutionary algorithms: an introduction. *Multi-objective evolutionary optimisation for product design and manufacturing*, p. 1–24, 2011. Citado na página 44.
- DEMŠAR, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006. Citado na página 62.
- DITZLER, G.; ROSEN, G.; POLIKAR, R. A Transductive Learning Algorithm for Nonstationary Environments. *International Joint Conference on Neural Networks*, p. 945–952, 2012. Citado na página 26.
- DOMINGOS, P.; HULTEN, G. Mining high-speed data streams. *Kdd*, p. 71–80, 2000. Citado na página 50.
- ELWELL, R.; POLIKAR, R. Incremental Learning in Nonstationary Environments with Controlled Forgetting. p. 771–778, 2009. Citado 2 vezes nas páginas 23 e 28.
- ELWELL, R.; POLIKAR, R. Incremental Learning of Concept Drift in Nonstationary Environments. *IEEE Transactions on Neural Networks*, v. 22, n. 10, p. 1517–1531, 2011. Citado 4 vezes nas páginas 21, 25, 26 e 30.

- ERDEM, Z. Ensemble Based Incremental Svm Classifiers for Changing Environments. *Simulation*, 2007. Citado 2 vezes nas páginas 29 e 30.
- GAMA, J.; MEDAS, P.; CASTILLO, G.; RODRIGUES, P. Learning with drift detection. *Advances in Artificial Intelligence–SBIA 2004*, p. 286–295, 2004. Citado na página 35.
- GOMES, J. B.; GABER, M. M.; SOUSA, P. A. C.; MENASALVAS, E. Collaborative Data Stream Mining in Ubiquitous Environments Using Dynamic Classifier Selection. *International Journal of Information Technology & Decision Making*, v. 12, n. 6, p. 1287–1308, 2013. Citado na página 36.
- GU, S.; JIN, Y. Generating diverse and accurate classifier ensembles using multi-objective optimization. *IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - MCDM 2014: 2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making, Proceedings*, p. 9–15, 2015. Citado na página 43.
- HAN, J. G.; HUI, X. F.; SUN, J. Dynamic financial distress prediction modeling based on slip time window and multiple classifiers. *2010 International Conference on Management Science and Engineering, ICMSE 2010*, p. 148–155, 2010. Citado 2 vezes nas páginas 36 e 38.
- KARNICK, M.; AHISKALI, M.; MUHLBAIER, M. D.; POLIKAR, R. Learning concept drift in nonstationary environments using an ensemble of classifiers based approach. *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, p. 3455–3462, 2008. Citado na página 28.
- KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Tracking recurring contexts using ensemble classifiers: An application to email filtering. *Knowledge and Information Systems*, v. 22, n. 3, p. 371–391, 2010. Citado na página 50.
- KHAMASSI, I.; SAYED-MOUCHAWEH, M.; HAMMAMI, M.; GHÉDIRA, K. Ensemble classifiers for drift detection and monitoring in dynamical environments. *Phm*, p. 1–14, 2013. Citado 5 vezes nas páginas 13, 21, 25, 28 e 29.
- KO, A. H. R.; SABOURIN, R.; BRITTO, A. S. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, v. 41, n. 5, p. 1735–1748, 2008. Citado na página 22.
- LUSTOSAFILHO, J. A. S.; CANUTO, A. M. P.; XAVIER, J. C. An analysis of diversity measures for the dynamic design of ensemble of classifiers. *Proceedings of the International Joint Conference on Neural Networks*, v. 2015-Septe, 2015. Citado 3 vezes nas páginas 31, 37 e 44.
- MINKU, L. L.; WHITE, A. P.; YAO, X. The impact of diversity on online ensemble learning in the presence of concept drift. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 5, p. 730–742, 2010. Citado 8 vezes nas páginas 21, 26, 29, 30, 31, 33, 37 e 38.
- MINKU, L. L.; YAO, X. DDD: A new ensemble approach for dealing with concept drift. *IEEE Transactions on Knowledge and Data Engineering*, v. 24, n. 4, p. 619–633, 2012. Citado 6 vezes nas páginas 23, 30, 33, 36, 51 e 75.

OZA, N. C.; RUSSELL, S. Experimental comparisons of online and batch versions of bagging and boosting. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, p. 359–364, 2001. ISSN 158113391X. Disponível em: <<http://portal.acm.org/citation.cfm?doid=502512.502565>>. Citado na página 33.

PINAGE, F. A.; SANTOS, E. M. d. A Dissimilarity-Based Drift Detection Method. *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, p. 1069–1076, 2015. Citado 3 vezes nas páginas 21, 23 e 49.

SANTOS, E. M. Static and Dynamic Overproduction and Selection of Classifier Ensembles with Genetic Algorithms. 2008. Citado 2 vezes nas páginas 22 e 31.

SANTOS, E. M. d.; SABOURIN, R.; MAUPIN, P. Pareto analysis for the selection of classifier ensembles. *GECCO'08: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation 2008*, p. 681–688, 2008. Citado 4 vezes nas páginas 31, 32, 44 e 58.

ZLIOBAITE, I. Learning under concept drift: an overview. *Training*, abs/1010.4, p. 1–36, 2009. Citado 3 vezes nas páginas 13, 26 e 27.

APÊNDICE A – Estudo - Diversidade e Acurácia

Neste estudo, dois grupos de conjuntos de classificadores são investigados: 1) conjuntos homogêneos; e 2) conjuntos heterogêneos. O intuito de estudar esses dois conjuntos é analisar o impacto da diversidade e acurácia para conjuntos com apenas um tipo de classificador e para conjuntos com mais de um tipo de classificador. Neste último caso, espera-se que a diversidade apresente valores maiores. Foram gerados conjuntos com 25 classificadores cada. Os conjuntos homogêneos foram formados por HoeffdingTree, enquanto os classificadores heterogêneos foram formados por: 5 HoeffdingTree, 5 kNN, 5 RandomHoeffdingTree, 5 Perceptron e 5 NaiveBayes. O conjunto heterogeneo possui classificadores de diversos tipos: Árvores de decisão, estatísticos, bayesianos e redes neurais. Com isso, obteve-se uma diversidade maior em relação aos conjuntos homogêneos.

Para gerar conjuntos com diversos níveis de diversidade foi utilizado o método Online Bagging modificado por [Minku e Yao \(2012\)](#), descrito no [Capítulo 2](#). Tal algoritmo consiste em incluir o parâmetro λ , que é utilizado para variar a diversidade do conjunto gerado. O valor de diversidade aumenta à medida que o valor de λ diminui.

Os valores de λ estudados para cada base de dados foram: 5000, 1000, 500, 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005. Essa escala de valores foi ampliada do estudo presente em [Minku e Yao \(2012\)](#). O cálculo da diversidade foi feito utilizando o cálculo da ambiguidade, descrito no [Capítulo 2](#). A ambiguidade foi utilizada por ser uma medida de diversidade não dependente do rótulo das instâncias, ou seja, trata-se de uma medida totalmente não supervisionada. Dessa forma, esse tipo de medida pode ser utilizada diretamente na seleção dinâmica de conjuntos de classificadores.

Os resultados da primeira série de experimentos são descritos nesta seção. Para a base line, conforme pode ser observado na [figura 17](#), as melhores acurácias médias foram encontradas para $\lambda = 100$ (conjuntos heterogêneos) e $\lambda = 5$ (conjuntos homogêneos), ou seja, com baixos valores de diversidade, porém esse não foi o menor valor testado. O uso de classificadores heterogêneos provocou uma maior diversidade. Esse comportamento pode ser observado na [Figura 17](#). Os conjuntos heterogêneos mostram-se mais estáveis em relação à acurácia.

Para a base Sine1 ([figura 18](#)), valores elevados de diversidade ($\lambda = 0,05$) também levam à queda na acurácia dos conjuntos. As melhores acurácias médias foram encontradas com $\lambda = 100$, ou seja, baixos valores de diversidade. Porém, novamente, esses valores de diversidade não correspondem ao menor valor testado. Quando a diversidade é muito

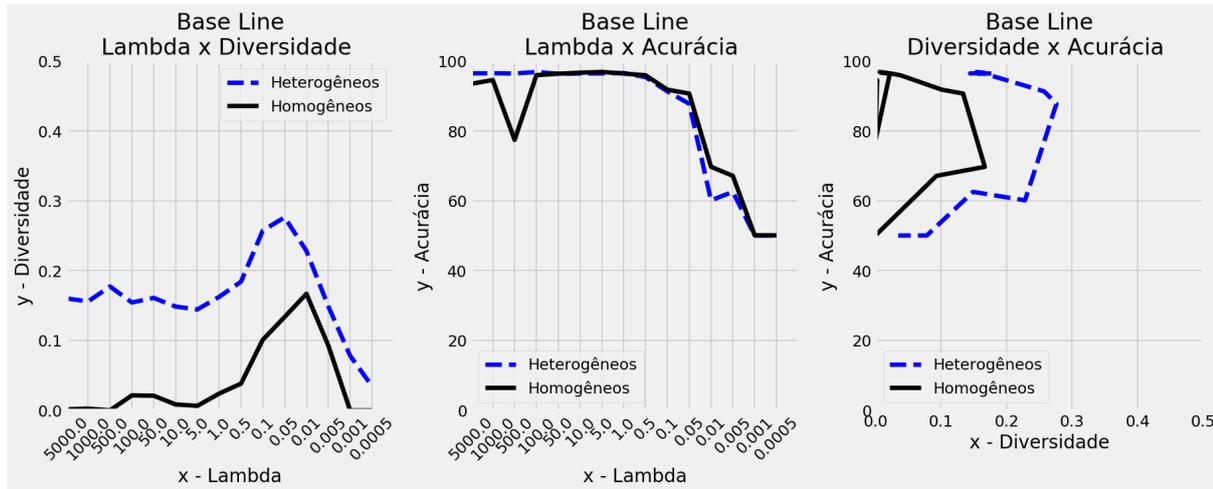


Figura 17 – Estudo referente a base Line. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

baixa, a acurácia também reduz. O uso de classificadores heterogêneos também aumentou a diversidade na base Sine1, provavelmente, esse comportamento faz com que a taxa de acurácia obtida por conjuntos heterogêneos permaneça estável à medida que a diversidade aumenta. Essa taxa cai somente com valores muito elevados de diversidade.

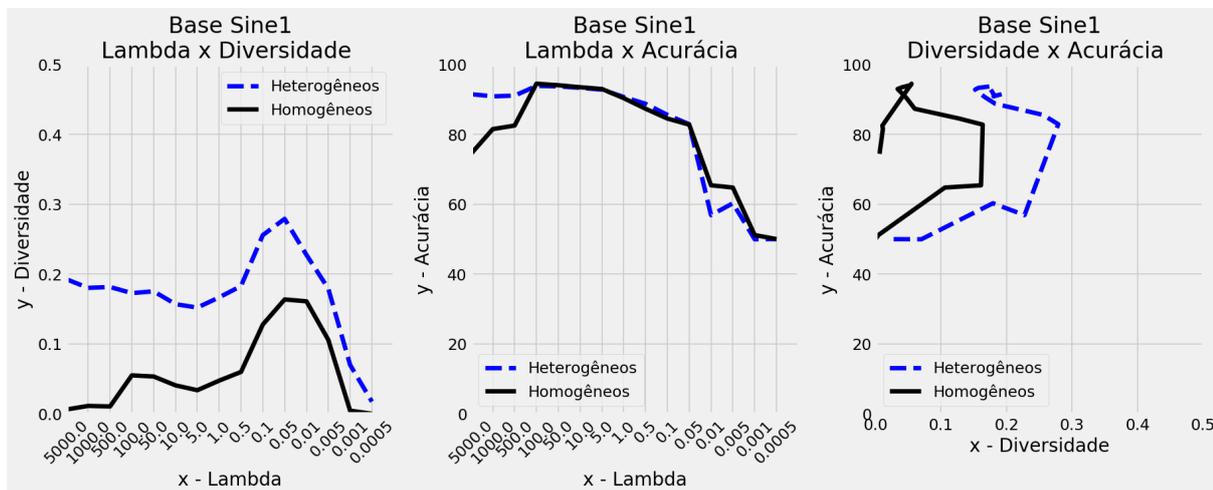


Figura 18 – Estudo referente a base Sine1. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

Já para a base Gauss (figura 19), os resultados também são parecidos, ou seja, os melhores valores de acurácia ocorrem com valores médios de diversidade. Novamente os conjuntos de classificadores heterogêneos conseguem manter a acurácia estável à medida

que a diversidade aumenta. Enquanto os homogêneos reduzem a taxa de acurácia tanto para valores muito baixos quanto para valores muito elevados de diversidade. As maiores acurácias foram encontradas para $\lambda = 1$ (conjuntos heterogêneos) e $\lambda = 0,5$ (conjuntos homogêneos), ou seja, baixos valores de diversidade, porém não o menor.

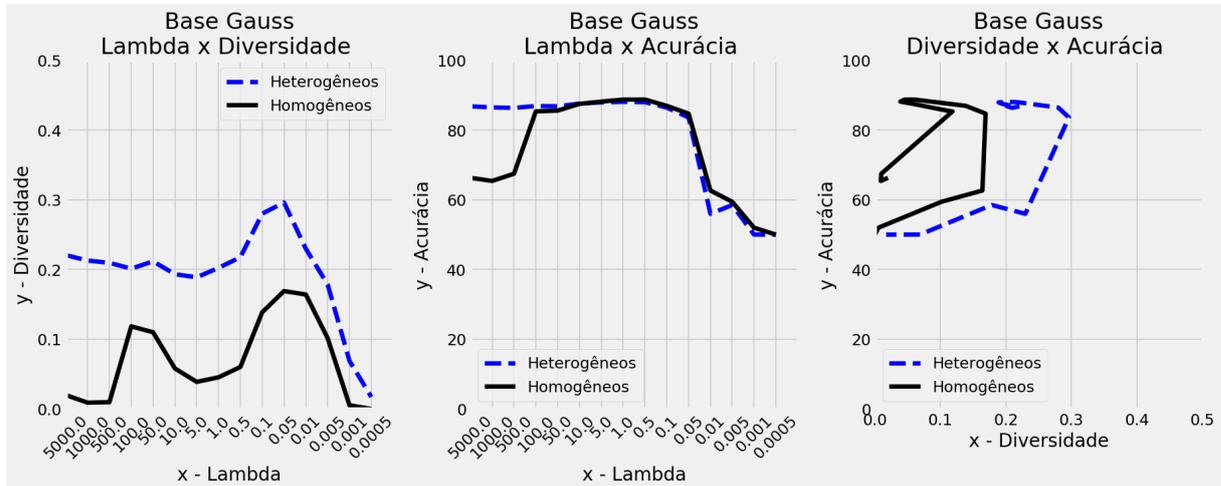


Figura 19 – Estudo referente a base Gauss. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

O mesmo comportamento observado para as bases anteriores também é observado para a base Circle (figura 20). As maiores acurácias médias foram encontradas com valores baixos de diversidade, porém não o menor, $\lambda = 50$ (conjuntos heterogêneos) e $\lambda = 1$ (conjuntos homogêneos). Novamente os conjuntos heterogêneos mostram-se mais estáveis em relação à acurácia.

Resumidamente, em relação aos resultados obtidos nas bases sintéticas, podemos concluir que independentemente do tipo de mudança, se selecionarmos conjuntos de classificadores com base em valores baixos de diversidade, melhores taxas de acurácia serão produzidas pelos classificadores. Porém, é importante destacar que os valores de diversidade não devem ser extremamente baixos. Além disso, se conjuntos de classificadores homogêneos forem utilizados, valores muito baixos de diversidade devem ser totalmente evitados.

Em relação às bases reais, os resultados são ligeiramente diferentes. Para a base Spam (figura 21), as taxas de acurácia reduzem à medida que a diversidade dos conjuntos aumenta, tanto com conjuntos heterogêneos quanto homogêneos. As melhores acurácias médias foram encontradas para $\lambda = 50$, isto é, com baixos valores de diversidade, porém não o menor. Conjuntos homogêneos e heterogêneos apresentaram comportamento similar, no entanto, conjuntos heterogêneos mostram-se mais estáveis em relação à acurácia.

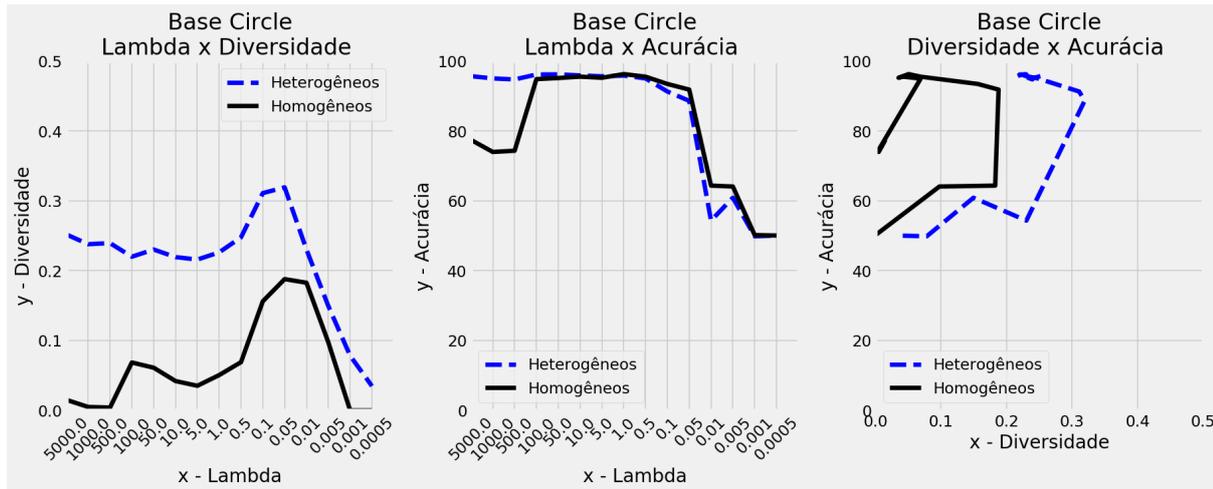


Figura 20 – Estudo referente a base Circle. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

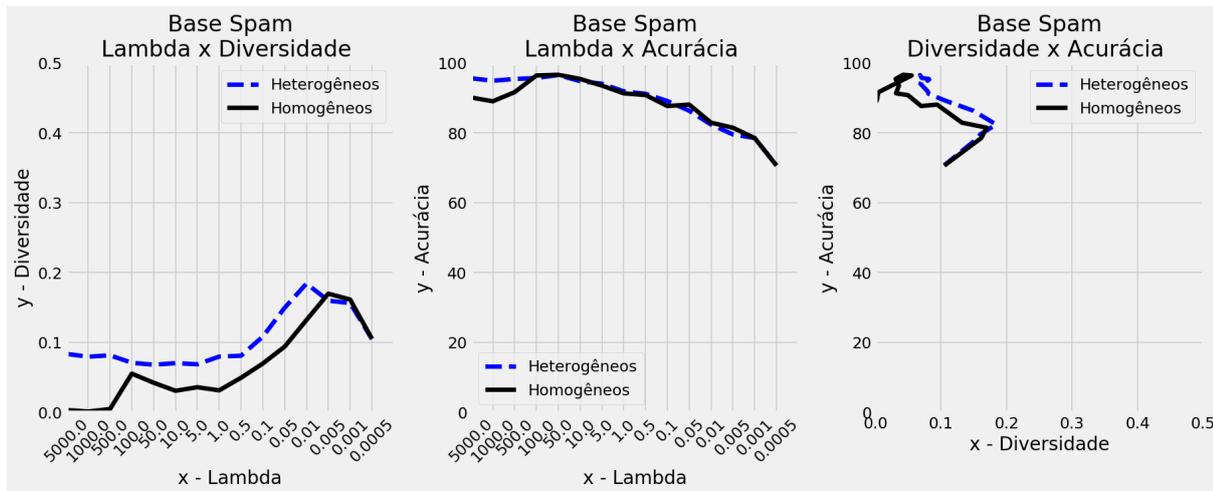


Figura 21 – Estudo referente a base Spam. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

Em relação à base Elec, conforme pode ser observado na figura 22, a relação diversidade x acurácia mostra um comportamento muito semelhante ao comportamento observado nas bases sintéticas, isto é, baixas taxas de acurácia são obtidas com valores muito baixos ou muito elevados de diversidade. As maiores acurácias médias foram encontradas para $\lambda = 10$ (conjuntos homogêneos) e $\lambda = 100$ (conjuntos heterogêneos). Diferentemente dos resultados anteriores, nessa base, as melhores acurácias foram obtidas com conjuntos

homogêneos, embora o uso de classificadores heterogêneos tenha inserido maior diversidade.

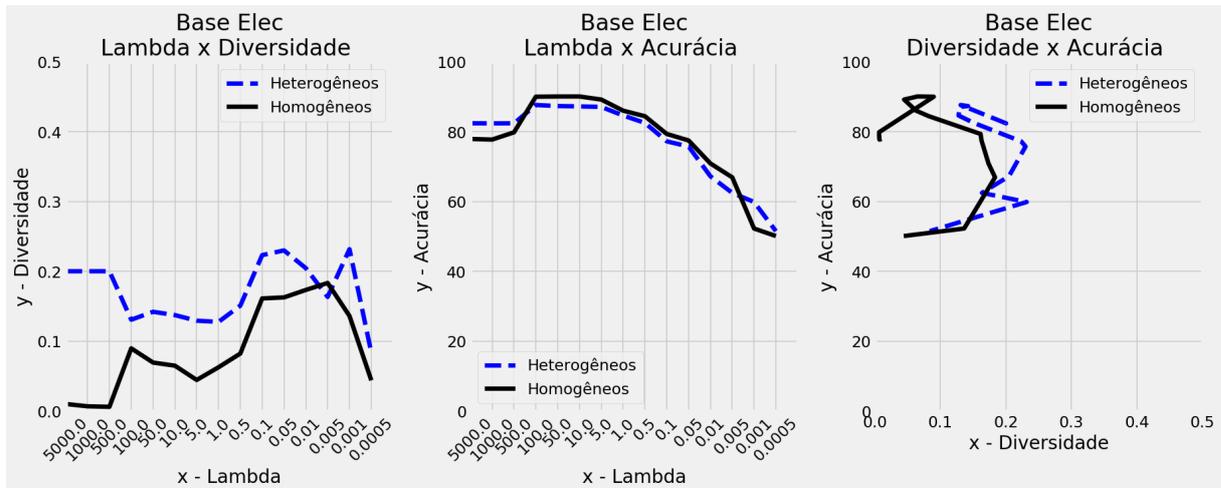


Figura 22 – Estudo referente a base Elec. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

Para a base KDDCup99, da mesma forma que as anteriores, os conjuntos heterogêneos apresentaram maior diversidade. Porém, os melhores valores de acurácia foram registrados para a menor diversidade $\lambda = 5000$. Por outro lado, da mesma forma como ocorreu nas demais bases, altos valores de diversidade promoveram a queda da acurácia, conforme pode ser observado na figura 23.

Pode-se observar que para todas as bases, ou seja, independentemente do tipo de mudança apresentado, nos entornos da alta diversidade o valor da acurácia decai bastante, mostrando que altos valores de diversidade produzem uma baixa acurácia média. A partir do momento em que se atinge o máximo da diversidade, a acurácia apresenta seus piores valores.

Portanto, para as bases analisadas, não há como especificar um limiar fixo de diversidade que garanta elevada acurácia. No entanto, um ponto de partida para fazer a seleção dinâmica do conjunto que melhor classificará uma instância seria utilizar diversidades próximas da menor diversidade e excluir conjuntos que possuem valor de diversidade próximo da máxima diversidade.

O uso de conjuntos heterogêneos, apesar de aumentar a curva de diversidade, não impactou significativamente no aumento da acurácia. Porém, conjuntos heterogêneos permitem que a variação de valor de diversidade ocorra em maior escala, sem afetar diretamente a acurácia do sistema. Dessa forma, embora o uso da diversidade possa trazer benefícios para a seleção do conjunto de classificadores, talvez utilizar um segundo critério, como a acurácia calculada através de janela deslizante ou a acurácia prequencial, possa

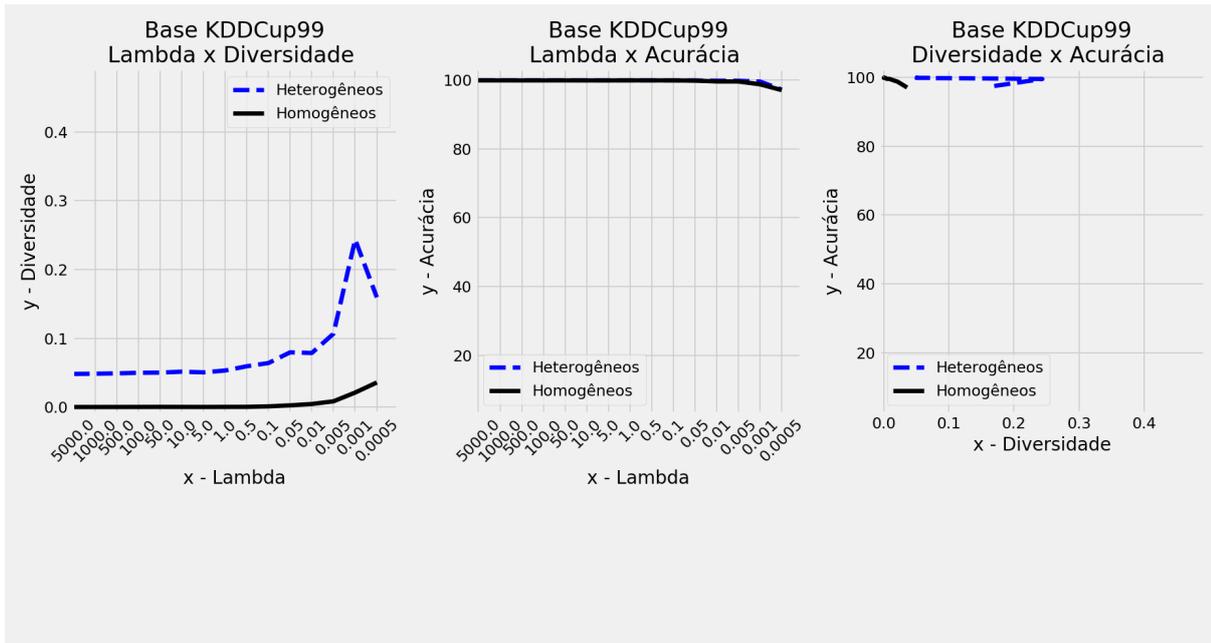


Figura 23 – Estudo referente a base KDDCup99. Média final das diversidades de todos os conjuntos homogêneos e heterogêneos estudados(Esquerda). Média final das acurácias de todos os conjuntos homogêneos e heterogêneos estudados(Centro). Curva de relação entre as diversidades médias finais e acurácia médias finais de todos os conjuntos homogêneos e heterogêneos estudados(Direita). Fonte: O autor.

tornar mais precisa a escolha do conjunto.