Hybrid Swarm Enhanced Classifier Ensembles

José Matheus Lacerda Barbosa[1], **Adriano Marabuco de Albuquerque Lima**[1], **Paulo Salgado Gomes de Mattos Neto**[1], **Adriano Lorena Inácio de Oliveira**[1]

[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brasil

{jmlb,amal,psgmn,alio}@cin.ufpe.br

Abstract. *Multiple classifier systems (MCSs) constitute one of the most competitive paradigms for obtaining accurate predictions in the field of machine learning. This article aims to assess if the application of hybrid swarm intelligence algorithms can improve the performance of MCSs by weigths optimization in combinations using weighted majority voting. The proposed methodology generated competitive results in 25 references datasets. Accuracy was adopted as objective function to be maximized by the following metaheuristics: particle swarm optimization (PSO), artificial bee colony (ABC), and the hybrid alternative of formers using dynamic multi swarm technique (DM-PSO-ABC).*

Resumo. *Os Sistemas de Multi-Classificadores (MCSs) constituem um dos paradigmas mais competitivos para a obtenção de classificações precisas no campo do aprendizado de máquina. Este artigo busca avaliar se a utilização de algoritmos híbridos de enxames pode melhorar a performance dos MCSs por meio da otimização de pesos em combinações por voto majoritário ponderado. A metodologia proposta rendeu resultados competitivos em 25 conjuntos de dados de referência. Adotou-se a acurácia como função objetivo a ser maximizada pelas seguintes meta-heurísticas: otimização do exame de partículas (PSO), a colônia artificial de abelhas (ABC), e a alternativa híbrida das anteriores usando a técnica de multi enxames dinâmicos (DM-PSO-ABC).*

1. Introdução

As tarefas de reconhecimento de padrões e classificação tem recebido bastante atenção da comunidade científica, o que tem possibilitado o desenvolvimento de modelos de aprendizado de máquina capazes de lidar com problemas cada vez mais complexos. Um desses ramos de pesquisa envolve o desenvolvimento de Sistemas de Multi-Classificadores (MCSs). Esses sistemas têm alcançado alcançado resultados mais acurados quando comparados com um único classificador [Mohammed et al. 2020].

MCSs tem obtido alta acurácia, uma vez que combinam classificadores diferentes em termos de vieses indutivos [Geman et al. 1992]. A abordagem clássica de focar em um classificador altamente otimizado tem vários desafios. Por exemplo, em termos do tamanho dos dados, o número de classes, a dimensionalidade do espaço de recursos, a sobreposição entre instâncias, o equilíbrio entre as categorias de classes e a complexidade não-linear das verdadeiras hipóteses desconhecidas [Zhou 2012]. Nesse contexto, usar um grupo (comitê) de classificadores surge como uma alternativa promissora.

Uma etapa importante na implementação do comitê é a função usada para combinar a saída dos classificadores [Mohammed et al. 2020]. Uma dessas formas é a votação

ponderada que atribui pesos para as diferentes previsões dos classificadores afim de obter a decisão final do comitê. No entanto, os seguintes desafios são encontrados na aplicação de uma regra de combinação baseada em peso:

1. Se a quantidade de dados aumenta ou o número de classificadores torna-se grande, o espaço de busca torna-se mais complexo;
2. O espaço de busca pode ser plano ou multimodal, ou seja, existem várias soluções (pesos) que fornecem aproximadamente a mesma precisão.

Um ramo das meta-heurísticas baseadas em população para otimização é a inteligência de enxames (IE, do inglês, *SI - Swarm Intelligence*). Os algoritmos de IE são definidos como "algoritmos inspirados na natureza que dizem respeito ao comportamento emergente coletivo de múltiplos agentes interagindo entre si e que seguem algumas regras simples"[Hassanien and Emary 2018, Albuquerque Filho et al. 2013]. A inteligência de enxame (IE) pode ser usada para identificar um padrão ou até mesmo para ajustar parâmetros de fusão de multi-classificadores em problemas mais complexos. Algoritmos baseados em população podem ser promissores no problema de busca de pesos dos classificadores, uma vez que aplicam de forma estocástica a exploração para evitar a estagnação nas soluções locais.

Nesta pesquisa busca-se avaliar a utilização de algoritmos híbridos de enxames na melhoria da performance dos MCSs por meio da otimização de pesos em combinações por voto majoritário ponderado. Os algoritmos de IE podem ser empregados na busca do peso para cada classificador. As seguintes hipóteses foram consideradas durante a condução desta pesquisa:

- O uso de IE para otimizar os pesos das decisões dos classificadores individuais de MCSs deve melhorar a decisão conjunta;
- O algoritmo híbrido de IE deve ser capaz de extrair os pontes fortes dos seus algoritmos de IE base e gerar resultados superiores aos algoritmos de IE tradicionais.

O restante deste artigo está organizado da seguinte forma: Na Seção 2, apresentamos os Fundamentos e Trabalhos Relacionados com a pesquisa proposta. Seção 3 apresenta a Metodologia usada na elaboração dos experimentos. Na Seção 4 expomos os Experimentos, assim como os resultados e análises estatística. E for fim, a Seção 5 apresenta a Conclusão do artigo.

2. Fundamentos e Trabalhos Relacionados

Nesta Seção são abordados os fundamentos desta pesquisa definindo o arcabouço teórico em que se baseia.

2.1. Comitês de classificadores

Um comitê é empregado para combinar a saída de diferentes classificadores que resolvem o mesmo problema com o objetivo de gerar um melhor modelo global [Ksieniewicz et al. 2018]. Tais sistemas reduzem os efeitos de erros causados por seus componentes, especialmente para amostras difíceis [Moreno-Seco et al. 2006]. O projeto de MCSs deve garantir um equilíbrio entre diversidade e compatibilidade entre os membros individuais para evitar redundância e obter robustez, respectivamente [Kuncheva 2004]. Informações suficientes e complementares sobre como esses sistemas operam podem ser encontrado em [Sagi and Rokach 2018, Woźniak et al. 2014].

A verdadeira hipótese (que é desconhecida) pode ser aproximada pesquisando em várias regiões do espaço de busca. Esse é um dos principais diferenciais do mecanismo de funcionamento dos MCSs. Cada classificador explora o espaço de busca de forma diferente usando diferentes subconjuntos de recursos, amostras de dados ou mecanismo de aprendizagem.

Em relação às aplicações, os MCSs têm sido usado para melhorar o desempenho de muitas tarefas de reconhecimento de padrões, como vigilância por vídeo [García-Martín and Martínez 2015], análise de dados [Bello-Orgaz et al. 2016], biometria multimodal [Fierrez et al. 2018], detecção de instrução [Aburomman and Ibne Reaz 2016], e pontuação de crédito [Feng et al. 2019].

Uma fase vital na construção de um sistema de multi-classificadores é o uso de uma estratégia de fusão adequada para agregar as decisões de resposta [Zhou 2012, Kuncheva and Rodríguez 2014]. As respostas dos classificadores individuais restringem o método de fusão e aumentam ou degradam a previsão do conjunto. Essas respostas podem estar no nível abstrato [Xu et al. 1992], em que cada classificador especifica um nome de classe como uma decisão. Ou as respostas podem ser classificadas em níveis [Parker 2001], onde cada classificador produz um subconjunto de rótulos de classe, ou mesmo valores de medição [Kuncheva and Rodríguez 2014, Niu et al. 2007]. Dessa forma, cada classificador especifica uma probabilidade *a posteriori* para cada classe. A precisão geral do modelo de conjunto pode ser melhorada ajustando-se adequadamente os pesos em um método de votação ponderada.

2.2. Inteligência de enxames

Um ramo das meta-heurísticas baseadas em população para otimização é a inteligência de enxames (IE, do inglês, *SI - Swarm Intelligence*). Os algoritmos de IE são definidos como "algoritmos inspirados na natureza que dizem respeito ao comportamento emergente coletivo de múltiplos agentes interagindo entre si e que seguem algumas regras simples"[Hassanien and Emary 2018]. Esses algoritmos imitam o comportamento social de enxames ou grupos de criaturas na natureza.

Em [Mirjalili et al. 2014], os autores discutiram os benefícios da IE sobre algoritmos evolutivos. Os algoritmos de IE mais populares na literatura são o Otimização por Enxame de Partículas (do inglês, *PSO - Particle Swarm Optimization*) [Kennedy and Eberhart 1995], a Otimização por Colônia de Formigas (do inglês, *ACO - Ant Colony Optimizer*) [Dorigo et al. 2006], a Colônia de Abelhas Artificiais (do inglês, *ABC - Artificial Bee Colony*) [Karaboga 2005], e o algoritmo inspirado em morcegos (do inglês, *BA - Bat Algorithm*) [Yang 2010]. Existem quatro razões principais para a disseminação e popularidade desses algoritmos: simplicidade, flexibilidade, mecanismo livre de derivadas e evasão de ótimos locais [Mirjalili et al. 2014].

A capacidade de otimização da inteligência de enxames (IE) reside nos operadores estocásticos, compartilhamento de informação, e presentação da informação sobre o espaço de busca ao longo das iterações. Algoritmos de enxames serão utilizados para aprimorar a fusão das classificações geradas pelos comitês de classificadores definindo os pesos ótimos para cada classificador. As seções seguintes apresentam os algoritmos utilizados nos experimentos.

2.2.1. Particle Swarm Intelligence - PSO

Otimização por Enxame de Partículas (do inglês, *PSO - Particle Swarm Optimization*) é uma técnica de otimização introduzida por [Kennedy and Eberhart 1995]. O PSO usa um mecanismo que simula o comportamento de bando de pássaros e cardume de peixes para guiar as partículas no espaço de soluções.

A execução do algoritmo do PSO começa com a inicialização das partículas, logo depois é calculado o valor de *fitness* de cada partícula, em seguida compara-se a posição atual e a melhor posição histórica com base no valor de *fitness* de cada, depois dessa comparação se as posições atuais forem superiores é feita a atualização das melhores posições de cada partícula e da vizinhança, e por fim as novas velocidades e posições são calculadas, e esses passos são repetidos até que uma condição de parada seja satisfeita.

O PSO tem alguns atrativos, simplicidade de implementação, poucos parâmetros, é eficaz em buscas globais, não sofre com a escalabilidade das variáveis e é facilmente paralelizável. Porém, o PSO com a vizinhança global tende a convergir com certa rapidez, o que pode levar a ótimos locais no processo de busca, o que pode ser resolvido com a redução do tamanho da vizinhança, como no PSO Local.

O algoritmo otimizador de enxames de partículas é amplamente usados em problemas do mundo real, tais como em processamento de sinais, sistemas de controle, aprendizado de máquina, processamento de imagens, dentre outros.

2.2.2. Artificial Bee Colony - ABC

Colônia de Abelhas Artificiais (do inglês, *ABC - Artificial Bee Colony*) é um algoritmo de otimização proposto por [Karaboga 2005]. Como o nome sugere, esse algoritmo é inspirado no comportamento das abelhas na busca por fontes de alimentos, também conhecido como néctar, e na forma como as abelhas compartilham informações sobre a localização e a qualidade das fontes localizadas.

A abordagem do algoritmo define os agente artificiais e os caracterizam em três tipos: *employed bee*, *onlooker bee* e *scout bee*. Cada tipo de abelhas tem funções diferentes. As *employed bees* trabalham na localização de fontes de alimento, as *onlooker bees* recebem as informações das fontes localizadas pelas *employed bees* e decidem se irão explorar as fontes ou se irão procurar novas fontes de alimento. Já as *scout bees* têm a função de buscar aleatoriamente por novas fontes de alimento e também de abandonar fontes já exploradas pelas outras abelhas.

A execução do ABC também segue uma sequência lógica de passos, são eles: 1) Inicialização da população, quando todas as fontes de alimento são definidas; 2) Fase das *employed bees*, quando as abelhas localizam as fontes, verificam sua vizinhança e é calculado o valor de *fitness*, para assim decidir qual fonte é mais promissora; 3) Fase das *onlooker bees*, quando as abelhas ficam esperando as *employed bees* compartilharem as informações de quais fontes são melhores para assim serem exploradas; 4) Fase das *scout bees*, as abelhas verificam quais fontes de alimento se esgotaram e inicializam novas fontes aleatoriamente; 5) Teste de término, depois de todo o processo de busca das abelhas, é verificado se a condição de parada foi satisfeita, se sim, a algoritmo encerra sua execução,

se não, repete-se os passos 2 ao 5, até que a condição de parada seja atingida.

Como no PSO, o ABC também é utilizado em vários problemas reais, como em projetos de engenharia, processamento de imagens, negócios, entre outros.

2.2.3. DM-PSO-ABC

Um algoritmo para extrair o melhor de cada um dos dois algoritmos anteriores foi proposto por [Wu and Zheng 2012], o DM-PSO-ABC, a sigla em inglês para *Dynamic Multistage Particle Swarm Optimization Artificial Bee Colony*. No DM-PSO-ABC o processo de otimização é dividido em três estágios.

Inicialmente a população é gerada aleatoriamente e dividida em sub-enxames. No primeiro estágio, para cada sub-enxame é feita uma busca grosseira focando num procedimento de *exploration* por melhores regiões no espaço de busca, utilizando as informações de melhores posições de cada sub-enxame separadamente por meio de um PSO Local. Devido ao comportamento de convergência rápida do PSO, os sub-enxames podem estagnar em ótimos locais, e para evitar esse problema é introduzido um reagrupamento periódico e aleatório entre as partículas dos enxames ao longo das iterações, esse processo manterá a diversidade dos enxames, possibilitará a troca de informações entre os sub-enxames e também evitará que os mesmos fiquem presos em ótimos locais. Após o reagrupamento, o PSO local é continuado, porém, com composições diferentes em cada sub-enxames.

No segundo estágio, é realizada uma etapa de ABC para cada sub-enxame após a execução de cada iteração do PSO local. Emprega-se o algoritmo ABC para fazer uma busca detalhada num procedimento de *exploitation*, a partir dos resultados do PSO Local, por sua habilidade de explorar melhor regiões específicas do espaço de busca. Assim, o ABC consegue aprimorar a busca grosseira realizada pelo PSO Local e facilitar a convergência para o ótimo global.

No terceiro e último estágio do algoritmo do DM-PSO-ABC é aplicado somente um PSO Global unificando os sub-enxames como forma de consolidar o processo de otimização e unir todas as informações coletadas pelos sub-enxames durante os estágios anteriores. Dessa forma, os sub-enxames se unem, compartilham as informações e conseguem convergir mais rapidamente em direção ao ótimo global, incrementando o procedimento de *exploitation* do algoritmo.

3. Metodologia

Nesta Seção, a metodologia dos experimentos e os dados utilizados são apresentados.

3.1. Hipótese da pesquisa e configuração

As duas principais hipóteses consideradas nos experimentos: (1) a capacidade de busca dos algoritmos de inteligência de exames podem melhorar a combinação de classificadores pelo ajuste dos pesos das previsões de cada classificador, e (2) a combinação das qualidades de algoritmos de enxames pode permitir a criação de algoritmos híbridos mais poderosos que os respectivos algoritmos constituintes. Durante a preparação dos dados, todos os conjuntos de dados foram pré-processados uniformizando as escalas de todas

variáveis entre 0 e 1 pela normalização min-max. Depois foi feita um separação de 20% dos dados para compor os conjuntos de teste e foi aplicada um validação cruzada com 10 repetições de 5 partes. Assim, foram executados 50 ciclos por base de dados. Os pesos foram obtidos a partir das previsões realizadas nos conjuntos de validação e os resultados a partir dos conjuntos de teste. A combinação foi realizada usando um função auxiliar que gera a decisão do comitê a partir da votos e pesos [Cruz et al. 2020].

Os parâmetros do enxames usados foram: PSO (iterações = 100, partículas = 10, dimensões = 5, inercia = 0.9, fator cognitivo = 0.5, fator social = 0.4), ABC (iterações = 1, fontes de alimento = 30, dimensões = 5, limite = 200) e DM-PSO-ABC(iterações = 100, partículas = 10, dimensões = 5, enxames = 3, troca entre enxames = 5, vizinhos = 3, inercia = 0.729, fator cognitivo = fator social = 1.49445, velocidade máxima = 0.6). Todos os experimentos foram realizados em máquinas virtuais utilizando a ferramenta Google Colaboratory para execução dos scripts.

3.2. Dados

O total de 25 bases de dados dos repositórios OpenML e UCI foram selecioandas para a avaliação experimental. Esses conjuntos de dados foram selecionados porque possuem características distintas [Mohammed et al. 2020]. A descrição dos dados é apresentada na Tabela 1 em que N, R, C, Cl e P/G representam o nome do conjunto, a quantidade de registros, a quantidade de características, o número de classes e a proporção entre a menor e a maior classe. O número de classes varia de 2 a 15, a maior quantidade de características é de 180 e a proporção entre classes varia de 1.9% a 100%. Todos os dados foram normalizados no intervalo $[0, 1]$.

N	**R**	**C**	**Cl**	**P/G**	**N**	**R**	**C**	**Cl**	**P/G**
Sonar	208	60	2	87%	Tae	151	5	3	94%
Heart-statlog	270	13	2	80%	Newthyroid	215	5	3	20%
Ionosphere	351	33	2	56%	Balance	625	4	3	17%
Saheart	462	9	2	53%	Dna	3186	180	3	46%
Wdbc	569	30	2	59%	Waveform	4999	21	3	97%
Wisconsin	683	9	2	54%	Vehicle	846	18	4	91%
Australian	690	14	2	80%	Heart-long-beach	200	13	5	19%
German	1000	20	2	43%	Thyroid-dis	2800	26	5	1.90%
Biodegradation	1055	41	2	51%	Dermatology	358	34	6	180%
Diabetic	1151	19	2	88%	Mfeat-zernike	2000	47	10	100%
Ringnorm	7400	20	2	98%	Mfeat-karh	2000	64	10	100%
Twonorm	7400	20	2	99%	Mfeat-fourier	2000	76	10	100%
					Movement-libras	360	90	15	100%

Tabela 1. Descrição dos conjuntos de dados utilizados nos experimentos.

3.3. Comitê de classificadores proposto

O comitê foi construído com as seguintes definições:

- Diversidade de exemplos: *bagging* será usado para obter várias amostras de treinamento distintas para cada classificador.

- Diversidade de características: 60% das características são selecionadas aleatoriamente para cada classificador.
- Modelos do comitê: 5 modelos diferentes (Árvore de decisão, Naive Bayes, KNN, Regressão logística e Dummy Classifier) representando 20% para cada classificador do comitê.

A combinação dos classificadores foi realizada por meio de métodos tradicionais e com uso dos algoritmo de enxames [Mohammed et al. 2020]. Os métodos de combinação foram: voto majoritário, voto probabilístico, voto majoritário ponderado e stack. No voto probabilístico cada classificador vota em uma classe e aquela com maior probabilidade *a posteriori* combinada prevalece. No voto majoritário ponderado cada classificação é multiplicada pelo peso da contribuição do respectivo classificador e a classe com mais votos ponderados prevalece. Neste trabalho os pesos são calculados utilizando a inteligência de enxames. No stack as previsões dos classificadores são combinadas por uma rede neural do tipo *Multilayer Perceptron* que define a classificação final.

3.4. Métricas e testes de hipóteses

Uma das características importantes para um comitê de classificadores é a diversidade dos modelos que o compõem. Isso é importante para maximizar a capacidade de generalização do comitê aumentando a quantidade de padrões que é capaz de reconhecer. Ela foi mensurada pela métrica de desacordo. O é calculado com base na frequência em que dois classificadores divergem em relação a classe predita, e um deles acerta a previsão. A métrica de acurácia foi utilizada por ser geralmente usada para avaliar o desempenho dos classificadores para este tipo de experimento, e servir parâmetro de avaliação e comparação pela literatura especializada.

Os resultados foram submetidos a análise estatística pelos seguintes de testes de hipóteses: teste friedman combinado com o teste posthoc de nemeyi, e o teste de *wilcoxon signed rank*. Esse tipo de avaliação é recomendada por [Demšar 2006] para comparação entre múltiplos classificadores, sendo o primeiro teste adequado para grupos e o segundo entre pares com o objetivo de identificar se há evidência estatística para comprovar a superioridade de desempenho entre classificadores.

4. Resultados

O comitê proposto utilizou 50 estimadores por classificador. Cada um dos 25 conjunto de dados foi submetido a 50 ciclos (10 cópias de 5 partes) de treinamento/validação/teste. A Tabela 2 apresenta a acurácia média das previsões de cada comitê.

Pelo rank médio de cada classificador, o enxame híbrido obteve a segunda melhor média, perdendo somente para o voto majoritário baseado nas probabilidades posteriori.

4.1. Diversidade

O comitê é composto por modelos de paradigmas de aprendizado distintos para se obter a necessária diversidade. Além disso o uso das técnicas de *bagging* de exemplos e características também teve esse objetivo. O resultado por conjunto de dados é exibido na Figura 1, e verifica-se que o nível de desacordo foi baixo para a maioria dos conjuntos [Aksela 2003].

	Individuais					Combinações			Enxames		
Dados	Árvore	NB	KNN	Rlog	Dummy	V. Maj.	V. Prob.	Stack	PSO	ABC	Híbrido
Australian	85.62%	82.49%	87.74%	**89.70%**	53.62%	**86.70%**	88.09%	86.57%	87.41%	87.23%	87.36%
Balance	81.23%	82.91%	83.89%	83.26%	45.60%	84.34%	**85.25%**	83.10%	**85.02%**	84.72%	84.93%
Biodegradation	**83.10%**	65.60%	**84.42%**	77.99%	62.56%	78.80%	81.06%	81.40%	82.65%	82.59%	82.60%
Dermatology	**98.64%**	97.58%	97.53%	98.03%	33.33%	**98.75%**	98.61%	90.67%	98.36%	98.25%	98.22%
Diabetic	**63.32%**	57.26%	61.52%	61.34%	51.52%	60.49%	**63.75%**	60.37%	62.84%	62.16%	62.71%
Dna	93.17%	91.57%	86.31%	94.41%	51.25%	94.48%	**94.94%**	**94.66%**	94.58%	94.58%	94.55%
German	73.98%	70.75%	71.76%	**75.61%**	70.50%	71.79%	72.24%	70.65%	**74.16%**	73.90%	73.95%
Hert-Beach	**29.15%**	22.80%	20.60%	**26.60%**	24.20%	25.65%	25.60%	21.60%	24.05%	24.25%	24.90%
Heart-Statlog	**83.33%**	78.70%	82.93%	82.70%	57.41%	82.93%	**83.93%**	72.04%	82.11%	82.07%	82.11%
Ionosphere	86.34%	60.62%	81.55%	**87.86%**	60.56%	78.28%	82.48%	68.45%	**86.51%**	86.48%	86.48%
Mf-fourier	79.25%	70.96%	82.50%	80.18%	9.73%	81.31%	**82.66%**	**82.72%**	81.96%	82.03%	81.84%
Mf-karh	89.82%	87.53%	94.90%	93.48%	9.71%	93.52%	**95.85%**	**95.54%**	94.96%	94.85%	94.89%
Mf-zernike	73.79%	70.11%	**81.41%**	78.08%	9.69%	77.69%	78.88%	**84.94%**	80.07%	80.30%	80.10%
Mov-Libras	65.64%	45.47%	65.08%	59.53%	3.69%	64.00%	**68.44%**	54.53%	67.19%	**67.58%**	67.31%
Newthyroid	92.88%	74.42%	**93.07%**	87.26%	74.42%	87.16%	90.23%	79.30%	92.93%	**93.12%**	92.84%
Ringnorm	**94.00%**	49.49%	80.90%	75.38%	49.48%	75.24%	90.51%	**95.67%**	93.41%	93.47%	93.88%
Sa-Heart	**94.00%**	49.49%	80.90%	75.38%	49.48%	87.16%	90.23%	78.56%	93.89%	93.96%	**94.00%**
Sonar	72.00%	67.29%	74.90%	**76.95%**	47.62%	72.90%	**75.19%**	63.90%	74.38%	73.81%	74.48%
Tae	**62.00%**	40.13%	57.23%	53.03%	34.13%	54.97%	**59.94%**	42.84%	57.03%	57.48%	57.74%
Thyroid-dis	69.03%	62.14%	69.25%	68.87%	58.75%	67.74%	68.14%	**69.39%**	69.30%	69.20%	**69.32%**
Twonorm	94.03%	50.64%	97.06%	**97.94%**	49.79%	95.53%	97.35%	97.50%	97.66%	97.68%	**97.71%**
Vehicle	**71.80%**	42.68%	69.95%	66.21%	21.76%	66.19%	**72.72%**	69.31%	71.59%	71.73%	71.39%
Waveform	80.89%	84.07%	83.35%	**85.57%**	33.84%	84.68%	84.93%	84.81%	85.43%	85.38%	85.42%
Wdbc	94.60%	69.21%	95.39%	95.47%	63.16%	94.42%	95.40%	91.93%	95.42%	**95.56%**	**95.49%**
Wisconsin	94.77%	84.96%	**96.48%**	**95.94%**	68.61%	94.74%	95.21%	91.28%	95.56%	95.62%	95.68%
Rank	5.32	9.6	5.44	5.32	10.8	6.84	**3.76**	6.76	4.04	4.12	**3.88**

Tabela 2. Resultados. Duas maiores acurácias em negrito.

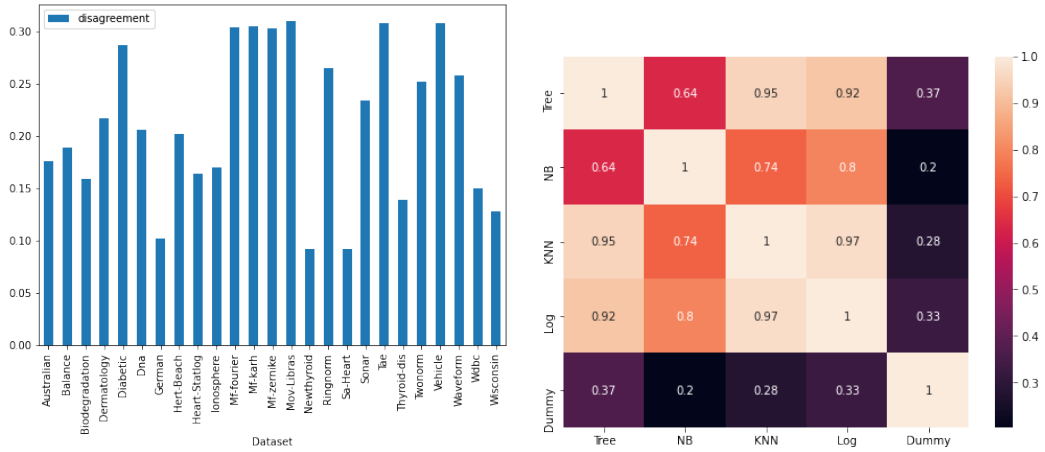
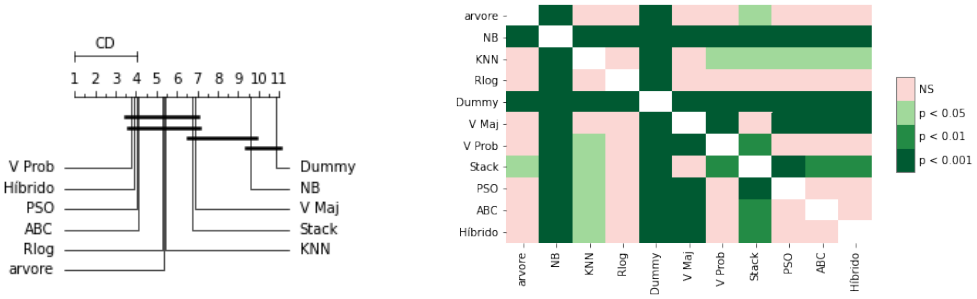


Figura 1. A métrica de desacordo e as correlações dos classificadores que compõem o comitê.

A análise da correlação entre as previsões por classificador revela que há um correlação alta entre os modelos fortes, o que explica o baixo nível de desacordo do comitê, como se vê na Figura 1. Essa condição limita a capacidade das combinações em obter resultados superiores, pois a homogeneidade das previsões oferece menos opções válidas para otimização da função de *fitness* [Lima Junior et al. 2010].

4.2. Combinação

Nesta parte avalia-se a acurácia da classificação comparando os diferentes métodos de combinação. A comparação entre o voto majoritário simples e o voto majoritário ponderado com a otimização pelo enxame híbrido mostra que nos 25 conjuntos de dados, o voto majoritário simples só teve desempenho melhor em três casos: *Dermatology*, *Heart-Beach* e *Heart-Statlog*, e por uma diferença menor que 1% nos dois primeiros casos, portanto a utilização de pesos otimizados no voto majoritário conferiu um ganho de per-



(a) Teste de Nemenyi

(b) Teste de Wilcoxon

Figura 2. Análises Estatísticas dos resultados: (a) teste de Nemenyi e (b) teste de Wilcoxon.

formance.

O comparativo com o voto probabilístico evidencia um empate com híbrido com um caso a mais a favor daquela combinação pelas probabilidades a posteriori. Além disso, esse foi o método como melhor rank médio tendo tido a melhor acurácia em 7 conjuntos de dados. Enquanto a comparação com o voto stack revela que na maioria dos casos, 19 vezes, o híbrido foi superior.

Quando se compara os enxames, nota-se que o PSO teve um desempenho ligeiramente melhor tendo sido superior ao híbrido em 14 casos, e em relação ao ABC situação se inverte com o híbrido tendo sido melhor em 16 casos. Importante destacar que devido a pequena diferença de desempenho, essa superioridade não se mostrou significativa estaticamente, o que é abordado na seção de análise estatística. De forma geral, a utilização da combinação do voto majoritário com pesos otimizados por algoritmos de enxames apresentou um desempenho agregado superior, porém não é possível afirmar que será o melhor resultado dentre os métodos testados. Também nem sempre o enxame híbrido prevalece em relação aos enxames tradicionais.

4.3. Análise Estatística

Os resultados foram submetidos a testes de hipóteses para verificar se há a rejeição da hipótese nula da ausência da melhoria em comparação aos algoritmos tradicionais. Dois testes não paramétricos foram realizados: o teste de Friedman para múltiplas comparações e e o teste de Wilcoxon para comparação de pares [Demšar 2006].

A estatística do teste de Friedman para distribuição qui-quadrado com 10 graus de liberada é 126.539 e p-valor inferior a 0,01% rejeitando a hipótese de nula da ausência de melhoria pelos algoritmos de enxame. A partir desse resultado foi executado o teste post hoc de nemeyi [Terpilowski 2019] que indicou em quais comparações foi possível rejeitar a hipótese nula. No caso do enxame híbrido com um grau de confiança de 95% rejeitou-se H_0 para o modelo naive bayes e dummy, conforme a Figura 2(a).

Em relação ao teste de Wilcoxon, rejeitou-se a hipótese nula para um maior conjunto de modelos em comparação ao enxame híbrido com uma confiança de 95%: naive bayes, knn, dummy,voto majoritário e stack, conforme Figura 2(b).

4.4. Consistência da Classificação

Um aspecto importante na avaliação dos modelos é a consistência dos classificadores. Nesse quesito os algoritmos de enxames obtiveram resultados mais consistentes que os demais, como se observa pela amplitude interquartil dos classificadores na Figura 3 .

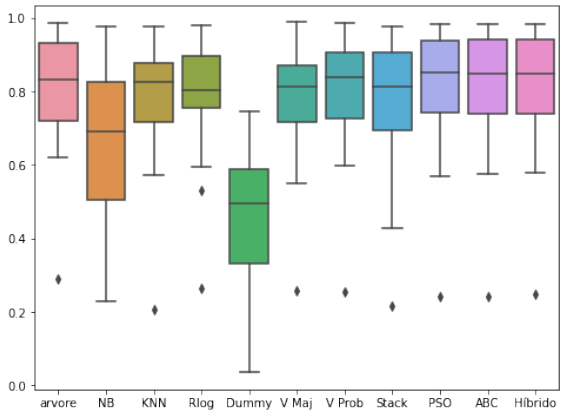


Figura 3. Boxplot da acurácia dos classificadores utilizados nos experimentos.

5. Conclusão

Algoritmos de enxames podem ser candidatos promissores para combinar saídas de comitês de classificadores. Contudo, é importante destacar que a abordagem proposta assume que existe uma solução ótima linear e estática. Essa suposição pode não ser verdadeira em conjuntos de dados sujeitos a mudanças em suas distribuições probabilidades ao longo do tempo. Além disso, o uso de algoritmos multi enxames aumenta o custo computacional da solução de otimização. Conforme o espaço de busca torna-se mais complexo, contendo muitos ótimos locais, uma solução de otimização mais robusta é necessária. Se este não é o caso, essa alternativa pode significar um custo desnecessário para o sistema de classificação.

Em relação a análise estatística, a não rejeição da hipótese nula nos testes de hipóteses entre os algoritmos de enxames deve considerar que o híbrido utilizado é uma combinação dos outros dois enxames comparados. Assim, o desempenho entre eles pode estar correlacionado dificultando a obtenção de evidências estatísticas da superioridade do algoritmo híbrido utilizado. Também é importante considerar que os conjuntos de dados ou comitê utilizado podem não ter oferecido condições para gerar diversidade nas respostas. Os resultados apresentados representam um pequeno recorte do universo de possibilidades de testes para a hipótese de pesquisa, e portanto comitês de classificadores, conjuntos de dados, hiperparâmetros ou métricas diferentes podem gerar resultados ainda mais reveladores que os expostos neste trabalho. Assim, não se pretende exaurir a hipótese de pesquisa, mas fornecer indicativos de que é uma investigação promissora a ser explorada.

A fronteira de pesquisa com comitês e algoritmos de otimização tem potencial para gerar soluções interessantes e eficazes. Nesse sentido futuros trabalhos a partir desta pesquisa podem envolver a combinação de classificadores por pesos para as diferentes classes de cada classificador baseando-se nas probabilidades *a posteriori* das classes.

Abordagens de seleção dinâmica de comitês também podem ser investigadas visando identificar as condições de melhor desempenho para cada tipo de solução.

Referências

Aburomman, A. A. and Ibne Reaz, M. B. (2016). A novel svm-knn-pso ensemble method for intrusion detection system. *Applied Soft Computing*, 38:360–372.

Aksela, M. (2003). Comparison of classifier selection methods for improving committee performance. In *Proceedings of the 4th International Conference on Multiple Classifier Systems*, MCS'03, page 84–93, Berlin, Heidelberg. Springer-Verlag.

Albuquerque Filho, F. S. d., Madeiro, F., Fernandes, S. M., de Mattos Neto, P. S., and Ferreira, T. A. (2013). Time-series forecasting of pollutant concentration levels using particle swarm optimization and artificial neural networks. *Química Nova*, 36(6):783–789.

Bello-Orgaz, G., Jung, J. J., and Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28:45–59.

Cruz, R. M. O., Hafemann, L. G., Sabourin, R., and Cavalcanti, G. D. C. (2020). Deslib: A dynamic ensemble selection library in python. *Journal of Machine Learning Research*, 21(8):1–5.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30.

Dorigo, M., Birattari, M., and Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39.

Feng, X., Xiao, Z., Zhong, B., Dong, Y., and Qiu, J. (2019). Dynamic weighted ensemble classification for credit scoring using markov chain. *Applied Intelligence*, 49(2):555–568.

Fierrez, J., Morales, A., Vera-Rodriguez, R., and Camacho, D. (2018). Multiple classifiers in biometrics. part 1: Fundamentals and review. *Information Fusion*, 44:57–64.

García-Martín, and Martínez, J. M. (2015). People detection in surveillance: classification and evaluation. *IET Computer Vision*, 9(5):779–788.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1):1–58.

Hassanien, A. E. and Emary, E. (2018). *Swarm intelligence: principles, advances, and applications*. CRC Press.

Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization. Technical report, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4.

Ksieniewicz, P., Krawczyk, B., and Woźniak, M. (2018). Ensemble of extreme learning machines with trained classifier combination and statistical features for hyperspectral data. *Neurocomputing*, 271:28–37.

Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, USA.

Kuncheva, L. I. and Rodríguez, J. J. (2014). A weighted voting framework for classifiers ensembles. *Knowledge and Information Systems*, 38(2):259–275.

Lima Junior, A. R., Silva, D. A., Mattos Neto, P. S., and Ferreira, T. A. (2010). An experimental study of fitness function and time series forecasting using artificial neural networks. In *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, pages 2015–2018.

Mirjalili, S., Mirjalili, S. M., and Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61.

Mohammed, A. M., Onieva, E., and Woźniak, M. (2020). Training set selection and swarm intelligence for enhanced integration in multiple classifier systems. *Applied Soft Computing*, 95:106568.

Moreno-Seco, F., Iñesta, J. M., de León, P. J. P., and Micó, L. (2006). Comparison of classifier fusion methods for classification in pattern recognition tasks. In Yeung, D.-Y., Kwok, J. T., Fred, A., Roli, F., and de Ridder, D., editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 705–713, Berlin, Heidelberg. Springer Berlin Heidelberg.

Niu, G., Han, T., Yang, B.-S., and Tan, A. C. C. (2007). Multi-agent decision fusion for motor fault diagnosis. *Mechanical Systems and Signal Processing*, 21(3):1285–1299.

Parker, J. (2001). Rank and response combination from confusion matrix data. *Information Fusion*, 2(2):113–120.

Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8:e1249.

Terpilowski, M. (2019). scikit-posthocs: Pairwise multiple comparison tests in python. *The Journal of Open Source Software*, 4(36):1169.

Woźniak, M., Graña, M., and Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3–17. Special Issue on Information Fusion in Hybrid Intelligent Fusion Systems.

Wu, D. and Zheng, J. (2012). A Dynamic Multistage Hybrid Swarm Intelligence Optimization Algorithm for Function Optimization. *Discrete Dynamics in Nature and Society*, 2012:578064.

Xu, L., Krzyzak, A., and Suen, C. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435.

Yang, X.-S. (2010). *A New Metaheuristic Bat-Inspired Algorithm*, pages 65–74. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman amp; Hall/CRC, 1st edition.