



Pós-Graduação em Ciência da Computação

**TIAGO PESSOA FERREIRA DE LIMA**

**SISTEMA HÍBRIDO INTELIGENTE PARA  
GERAÇÃO, SELEÇÃO E COMBINAÇÃO DE  
CLASSIFICADORES**



Universidade Federal de Pernambuco  
posgraduacao@cin.ufpe.br  
www.cin.ufpe.br/~posgraduacao

RECIFE  
2017

**Tiago Pessoa Ferreira de Lima**

**Sistema Híbrido Inteligente para Geração, Seleção e Combinação de Classificadores**

Este trabalho foi apresentado à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

**ORIENTADORA: Profa. Teresa Bernarda Ludermir**

RECIFE  
2017

Catálogo na fonte  
Bibliotecário Jefferson Luiz Alves Nazareno CRB4-1758

L732s Lima, Tiago Pessoa Ferreira de.  
Sistema híbrido inteligente para geração, seleção e combinação de classificadores/ Tiago Pessoa Ferreira de Lima – 2017.  
72 f.: fig., tab.

Orientadora: Teresa Bernarda Ludermir.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da computação, 2017.  
Inclui referências.

1. Inteligência artificial. 2. Geração de classificadores. 3. Sistema híbrido inteligente. I. Ludermir, Teresa Bernarda. (Orientadora). II. Título.

006.3            CDD (22. ed.)            UFPE-MEI 2018-06

**Tiago Pessoa Ferreira de Lima**

**Sistema Híbrido Inteligente para Geração, Seleção e Combinação de Classificadores**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Aprovado em: 31/08/2017.

---

**Orientadora: Profa. Dra. Teresa Bernarda Ludermir**

**BANCA EXAMINADORA**

---

Prof. Dr. George Darmiton da Cunha Cavalcanti  
Centro de Informática / UFPE

---

Prof. Dr. Ricardo Bastos Cavalcante Prudêncio  
Centro de Informática / UFPE

---

Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho  
Instituto de Ciências Matemáticas e de Computação / USP

---

Prof. Dr. Wilson Rosa de Oliveira Junior  
Departamento de Estatística e Informática / UFRPE

---

Prof. Dr. João Fausto Lorenzato de Oliveira  
Faculdade de Ciências e Tecnologia de Garanhuns Magano / UPE

*Este trabalho é dedicado à minha esposa.*

# Agradecimentos

Primeiramente à Deus, pelo dom da vida; sem Ele nada seria possível. Aos meus queridos pais, pelo amor incondicional, apoio e dedicação constantes, sobretudo por terem me ensinado que a fé é o fundamento de nossa existência. À Anésia e aos meus irmãos, por terem compartilhado comigo a satisfação de chegar até aqui. Ao casal Moisés e Luciene, por estarem comigo nos momentos de alegria e tristeza. Aos demais amigos e familiares, especialmente à minha avó Rubenita e a tia Silvia, pela força, carinho, compreensão e apoio. À minha linda e amada esposa, pelos incentivos constantes. Parceira em todos os momentos e onipresente em meus passos vindouros. À minha orientadora, por ter confiado em meu trabalho e principalmente pelo apoio dado em momentos de dificuldade. Aos demais professores do CIn, pelas aulas ministradas. Aos participantes da banca, pelas críticas e sugestões para melhoria deste trabalho. Aos meus colegas da pós, Adenilto José da Silva, Anderson Tenório Sérgio e Rodolfo Carneiro Cavalcante, pelas discussões e ensinamentos. À FACEPE, pelo apoio financeiro, sob processo IBPG-0610-1.03/12, sem o qual seria mais difícil o desenvolvimento dessa pesquisa. Ao IFPE, especialmente ao professor Willams, pelo apoio e compreensão na finalização das correções do texto. Finalmente, a todos que, direta ou indiretamente, contribuíram para a realização desta conquista. Muito obrigado!

*“O homem sábio é poderoso  
e quem tem conhecimento  
aumenta a sua força;  
quem sai à guerra  
precisa de orientação,  
e com muitos conselheiros  
se obtém a vitória.”*

*(Bíblia Sagrada - NVI, Provérbios 24:5,7)*

# Resumo

Diversos estudos em Aprendizagem de Máquina demonstram teoricamente e empiricamente que problemas de classificação geralmente apresentam melhores taxas de acerto através da combinação de múltiplos classificadores. Essa estratégia assemelha-se a natureza humana em buscar opiniões de diferentes especialistas, a fim de se obter uma decisão mais precisa. Contudo, nem sempre são selecionados os classificadores mais adequados para prever o padrão em questão, especialmente se tratando da combinação estática de classificadores. A seleção (ou combinação) dinâmica de classificadores é uma técnica que usa a ideia de regiões de competência, na qual presume-se que diferentes classificadores são mais apropriados para rotular padrões com diferentes graus de dificuldade. Embora os atuais métodos apresentem diferenças em vários aspectos, a regra que seleciona o(s) classificador(es), em geral, é genérica e fixa. Além disso, a hipótese utilizada para decidir se um classificador será selecionado é, na maioria das vezes, definida com base em apenas um único critério. Sistemas Híbridos Inteligentes são modelos que resultam da combinação de duas, ou mais técnicas distintas, visando unir vantagens para suprir deficiências individuais. Neste sentido, a presente pesquisa tem como objetivo propor e avaliar a hibridização de técnicas que orientem tanto a geração do conjunto de classificadores como a posterior combinação destes. Os resultados experimentais sugerem que o sistema proposto apresenta desempenho superior quando comparado com algumas das principais técnicas existentes na literatura.

**Palavras-chave:** Aprendizagem de Máquina. Geração de Classificadores. Seleção Dinâmica de Classificadores. Sistema Híbrido Inteligente.

# Abstract

Several studies in Machine Learning demonstrate theoretically and empirically that classification problems usually present better hit rates through the combination of multiple classifiers. This strategy resembles human nature in seeking opinions from different experts in order to obtain a more precise decision. However, the best classifiers are not always selected to predict the pattern in question, especially when dealing with the static combination of classifiers. Dynamic classifier selection (or combining) is a technique that uses the idea of regions of competence, in which different classifiers are assumed to be more appropriate for labeling patterns with different degrees of difficulty. Although the current methods present differences in several aspects, the rule that selects the classifier(s), in general, is generic and fixed. In addition, the hypothesis used to decide whether a classifier will be selected is most often defined based on only one criterion. Hybrid Intelligent Systems are models that result from the combination of two, or more distinct techniques, aiming to unite advantages to supply individual deficiencies. In this sense, the present research aims to propose and evaluate the hybridization of techniques that guide both the generation of the set of classifiers and the subsequent combination of these. The experimental results suggest that the proposed system presents superior performance when compared to some of the main techniques in the literature.

**Keywords:** Machine Learning. Classifier Generation. Dynamic Classifier Selection. Hybrid Intelligent System.

# Lista de ilustrações

Figura 1 – Probabilidade do voto majoritário estar errado com $L$ hipóteses independentes, $L = 1, \dots, 60$ , e probabilidades de erro individuais de 0.35, 0.40, e 0.45 . . . . .	18
Figura 2 – Probabilidade do voto majoritário estar errado com $L$ hipóteses independentes, $L = 1, \dots, 60$ , e probabilidades de erro individuais de 0.55, 0.60, e 0.65 . . . . .	19
Figura 3 – Princípios que justificam porque <i>ensembles</i> podem funcionar melhor do que um único classificador. Figura adaptada de (DIETTERICH, 2000a)	19
Figura 4 – Funcionamento do <i>Bagging</i> . . . . .	27
Figura 5 – Funcionamento do <i>Boosting</i> . . . . .	28
Figura 6 – Funcionamento do <i>Random Subspace</i> . . . . .	29
Figura 7 – Exemplo de um espaço de instâncias bidimensional particionado em 4 regiões de competência. Figura extraída de (KUNCHEVA, 2000). . . . .	31
Figura 8 – Topologia paralela na combinação de múltiplos classificadores . . . . .	39
Figura 9 – Topologia serial na combinação de múltiplos classificadores . . . . .	40
Figura 10 – Topologia híbrida na combinação de múltiplos classificadores . . . . .	40
Figura 11 – Exemplo dos métodos de combinação: média, máximo, mínimo e produto	43
Figura 12 – Estrutura macro do método desenvolvido . . . . .	45
Figura 13 – Representação do teste de <i>Nemenyi</i> comparando o número de componentes	53
Figura 14 – Representação do teste de <i>Nemenyi</i> comparando resultados da tabela 4	59
Figura 15 – Representação do teste de <i>Nemenyi</i> comparando resultados da tabela 5	59
Figura 16 – Representação do teste de <i>Nemenyi</i> comparando métodos da tabela 7 .	61

# Lista de tabelas

Tabela 1	– Problemas usados nos experimentos . . . . .	52
Tabela 2	– Médias da acuracidade no conjunto $\tau$ considerando <i>ensembles</i> gerados pelo Algoritmo 1 com $L$ variando de 5 à 50 . . . . .	54
Tabela 3	– Médias da acuracidade no conjunto $\tau$ considerando <i>ensembles</i> gerados pelo Algoritmo 1 com $L$ variando de 55 à 100 . . . . .	55
Tabela 4	– Médias da acuracidade considerando o conjunto de classificadores gerados pelo Algoritmo 1. O número de derrotas, empates e vitórias alcançado por cada método, de acordo com os testes de <i>Wilcoxon</i> , são apresentados no final da tabela . . . . .	57
Tabela 5	– Médias da acuracidade considerando o conjunto de classificadores gerados através do algoritmo <i>Bagging</i> . O número de derrotas, empates e vitórias alcançado por cada método, de acordo com os testes de <i>Wilcoxon</i> , são apresentados no final da tabela . . . . .	58
Tabela 6	– Médias da Correlação, Desacordo, Falta Dupla e Estatística Q entre classificadores gerados pelo Algoritmo 1 e <i>Bagging</i> . . . . .	60
Tabela 7	– Comparação da acuracidade do MP contra técnicas estáticas. O número de derrotas, empates e vitórias alcançado por cada uma dessas técnicas, de acordo com os testes de <i>Wilcoxon</i> , são apresentados no final da tabela . . . . .	62
Tabela 8	– Comparação do tempo de classificação, em nanosegundos, num processador <i>Core i7</i> de 3,6GHz do conjunto de generalização $\zeta$ . . . . .	63
Tabela 9	– Comparação do MP com Trabalhos Relacionados . . . . .	64

# Lista de abreviaturas e siglas

ADBO	<i>Adaptive Boosting</i>
AE	Algoritmo Evolucionário
CD	<i>Critical Difference</i>
CE	Combinação Estática
DES-FA	<i>Dynamic Ensemble Selection by Filter + Adaptive Distance</i>
DF	<i>Double Fault</i>
DP	<i>Decision Profile</i>
ENN	<i>Edited Nearest Neighbor</i>
KNN	<i>K Nearest Neighbors</i>
KNOP	<i>K-Nearests Output Profiles</i>
KNORA	<i>K-Nearest Oracles</i>
KNORAE	<i>K-Nearest Oracles - Eliminate</i>
KNORAU	<i>K-Nearest Oracles - Union</i>
LCA	<i>Local Class Accuracy</i>
LKC	<i>Ludmila Kuncheva Collection of Real Medical Data</i>
MCB	<i>Multiple Classifier Behavior</i>
MLP	<i>Multi-Layer Perceptron</i>
MP	Método Proposto
OCS	<i>Overproduce-and-Choose</i>
OLA	<i>Overall Local Accuracy</i>
RF	<i>Random Forest</i>
SDES	<i>Sorting-Based Dynamic Classifier Ensemble Selection</i>
SHI	Sistema Híbrido Inteligente
SVM	<i>Support Vector Machine</i>
UCI	<i>University of California, Irvine</i>

# Lista de símbolos

$H$	conjunto de classificadores
$h_i$	$i$ -ésimo classificador do conjunto $H$
$\vec{x}$	padrão de consulta
$h_i(\vec{x})$	previsão atribuída a $\vec{x}$ pelo classificador $h_i$
$L$	tamanho do conjunto de classificadores
$p$	probabilidade
$l$	número de sucessos
$\mathcal{H}$	espaço de hipóteses
$Z$	conjunto de dados
$f$	função ideal
$\tau$	conjunto de treinamento
$n$	tamanho do conjunto
$\in$	pertence
$d$	dimensionalidade ou número de características
$E$	conjunto de <i>ensembles</i>
$e_i$	$i$ -ésimo <i>ensemble</i> do conjunto $E$
$M$	tamanho do conjunto de <i>ensembles</i>
$k$	número de padrões na região de competência
$R_j$	$j$ -ésima região
$e^*$	<i>ensemble</i> com maior acuracidade
$p(e_i R_j)$	probabilidade do <i>ensemble</i> $e_i$ dada a região $R_j$
$e_{ij}$	<i>ensemble</i> $e_i$ designado para a região $R_j$
$p_c$	probabilidade de correta classificação

$p(R_j)$	probabilidade de pertencer a região $R_j$
$p_c(R_j)$	probabilidade de estar correto na região $R_j$
$\omega$	classe do padrão de consulta $\vec{x}$
$p_j(\omega \vec{x})$	probabilidade ou grau de apoio dado pelo classificador $h_j$ para a classe de $\vec{x}$
$W_i$	$i$ -ésimo peso
$x_{tes}^{\vec{}}$	padrão do conjunto de teste
$x_{tr}^{\vec{}}$	padrão do conjunto de treinamento
$N^{00}$	número de padrões que dois classificadores erram
$N^{01}$	número de padrões que o primeiro erra e o segundo acerta
$N^{10}$	número de padrões que o primeiro acerta e o segundo erra
$N^{11}$	número de padrões que dois classificadores acertam
$k_p$	índice de diversidade
$\alpha$	ambiguidade
$O$	número de classes
$\mu_i^j$	saída de $h_i$ para a classe $\omega_j$
$S$	conjunto de seletores
$s_i$	$i$ -ésimo seletor
$\nu$	conjunto de validação
$\vec{v}$	vetor de características
$\zeta$	conjunto de generalização

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Motivações</b>	<b>21</b>
1.1.1	Geração dos Componentes	21
1.1.2	Seleção e Fusão dos Componentes	21
<b>1.2</b>	<b>Objetivos</b>	<b>22</b>
<b>1.3</b>	<b>Produção Bibliográfica</b>	<b>22</b>
1.3.1	Artigos Completos Publicados em Periódicos	23
1.3.2	Trabalhos Completos Publicados em Anais de Congressos	23
<b>1.4</b>	<b>Organização do Texto</b>	<b>24</b>
<b>1.5</b>	<b>Considerações Finais</b>	<b>24</b>
<b>2</b>	<b>GERAÇÃO DOS COMPONENTES</b>	<b>25</b>
<b>2.1</b>	<b>Diversidade em Esembles</b>	<b>25</b>
<b>2.2</b>	<b>Algoritmos de Geração - Clássicos</b>	<b>26</b>
2.2.1	<i>Bagging</i>	27
2.2.2	<i>Boosting</i>	28
2.2.3	<i>Random Subspace</i>	29
<b>2.3</b>	<b>Algoritmos de Geração - Evolucionários</b>	<b>30</b>
<b>2.4</b>	<b>Considerações Finais</b>	<b>30</b>
<b>3</b>	<b>SELEÇÃO DOS COMPONENTES</b>	<b>31</b>
<b>3.1</b>	<b>Seleção Dinâmica de Classificadores</b>	<b>31</b>
3.1.1	Seleção Baseada em <i>Ranking</i>	32
3.1.2	Seleção Baseada em Acurácia	32
3.1.3	Seleção Baseada em Probabilidade	33
3.1.4	Seleção Baseada em <i>Clustering</i>	33
3.1.5	Seleção Baseada em Comportamento	34
3.1.6	Seleção Baseada em Oráculo	34
3.1.7	Seleção Baseada em Acuracidade e Diversidade	35
3.1.8	Seleção Baseada em Diversidade	35
3.1.9	Seleção Baseada em Ambiguidade	36
3.1.10	Seleção Baseada em Oráculo Randômico Linear	36
3.1.11	Seleção Baseada em <i>Overproduce-and-choose</i>	36
3.1.12	Seleção Baseada em Filtros e Distância Adaptativa	37
3.1.13	Seleção Baseada em Meta-Aprendizado	37
3.1.14	Seleção Baseada em Complexidade do Problema	38

<b>3.2</b>	<b>Considerações Finais</b>	<b>38</b>
<b>4</b>	<b>COMBINAÇÃO DOS COMPONENTES</b>	<b>39</b>
<b>4.1</b>	<b>Padrões de Arquitetura</b>	<b>39</b>
4.1.1	Topologia Paralela	39
4.1.2	Topologia Serial	40
4.1.3	Topologia Híbrida	40
<b>4.2</b>	<b>Regras de Decisão</b>	<b>41</b>
4.2.1	Regras de Decisão do Tipo <i>Hard</i>	41
4.2.2	Regras de Decisão do Tipo <i>Soft</i>	41
<b>4.3</b>	<b>Considerações Finais</b>	<b>42</b>
<b>5</b>	<b>MÉTODO PROPOSTO</b>	<b>44</b>
<b>5.1</b>	<b>Sistemas Híbridos Inteligentes</b>	<b>44</b>
<b>5.2</b>	<b>Hibridizações Propostas</b>	<b>44</b>
5.2.1	Geração dos Componentes	46
5.2.2	Geração dos Seletores	47
5.2.3	Extração das Características	48
5.2.4	Generalização	50
<b>5.3</b>	<b>Considerações Finais</b>	<b>51</b>
<b>6</b>	<b>EXPERIMENTOS E RESULTADOS</b>	<b>52</b>
<b>6.1</b>	<b>Avaliação do Número de Componentes</b>	<b>53</b>
<b>6.2</b>	<b>Comparações com Técnicas de Seleção</b>	<b>56</b>
<b>6.3</b>	<b>Comparações com Técnicas Estáticas</b>	<b>61</b>
<b>6.4</b>	<b>Comparações com Trabalhos Relacionados</b>	<b>64</b>
<b>6.5</b>	<b>Considerações Finais</b>	<b>65</b>
<b>7</b>	<b>CONCLUSÃO E TRABALHOS FUTUROS</b>	<b>66</b>
<b>7.1</b>	<b>Síntese do Trabalho</b>	<b>66</b>
<b>7.2</b>	<b>Sugestões para Continuidade</b>	<b>67</b>
	<b>REFERÊNCIAS</b>	<b>68</b>

# 1 INTRODUÇÃO

Classificação envolve a identificação de objetos ou padrões para colocá-los em determinadas categorias. Consiste basicamente em analisar um conjunto de atributos (vetor de características) do objeto a ser classificado e, segundo algum critério ou hipótese<sup>1</sup>, determinar a qual classe o mesmo pertence. Métodos responsáveis pela classificação de padrões são chamados de classificadores. Esses variam nos seus objetivos, na disponibilidade dos dados de treinamento, nas estratégias de aprendizagem e principalmente no modo que empregam para representação do conhecimento adquirido. Embora cada tipo de classificador possua suas próprias particularidades, todos eles aprendem através de buscas num espaço de hipóteses para encontrar uma solução com generalização aceitável.

Construir o classificador ideal para uma determinada tarefa é algo praticamente utópico. Uma alternativa para contornar este problema consiste em utilizar *ensembles* de classificadores, ou seja, um comitê cujas hipóteses individuais são combinadas de alguma forma para rotular o padrão em questão. Seja  $H = \{h_1, h_2, h_3\}$  um conjunto de três classificadores. Um dado padrão de consulta  $\vec{x}$  será rotulado por cada classificador  $h_i \in H$ . Seja  $h_i(\vec{x})$  a previsão atribuída a  $\vec{x}$  pelo classificador  $h_i$ . Se os três classificadores são idênticos, então quando  $h_1(\vec{x})$  está errado, conseqüentemente  $h_2(\vec{x})$  e  $h_3(\vec{x})$  também estão. Entretanto, se os erros cometidos forem não correlacionados, então quando  $h_1(\vec{x})$  está errado, é possível que  $h_2(\vec{x})$  e  $h_3(\vec{x})$  estejam corretos, de forma que a combinação das hipóteses por voto majoritário, por exemplo, pode classificar corretamente o padrão  $\vec{x}$ .

Dado um *ensemble* de tamanho  $L$ , para cada padrão  $\vec{x}$  a ser classificado, tem-se uma série de  $L$  ensaios. Considerando que cada um desses ensaios é independente e que cada ensaio é um sucesso na classificação de  $\vec{x}$  com probabilidade  $p$  ou uma falha com probabilidade  $1 - p$ , então a probabilidade do número de sucessos ser  $l$  em  $L$  ensaios é dada pela equação 1.1. Prontamente, a probabilidade do voto majoritário estar errado, ou seja, a probabilidade do *ensemble* ter mais da metade de classificadores que erram o padrão em questão é dada pela equação 1.2, que corresponde à área sob o gráfico da distribuição binomial na qual mais que a metade das hipóteses ( $\frac{L}{2}$ ) estão incorretas.

$$P(Z = l) = \binom{L}{l} p^l (1 - p)^{L-l} \quad (1.1)$$

$$P(Z > \frac{L}{2}) = 1 - \sum_{l=0}^{\frac{L}{2}} \binom{L}{l} p^l (1 - p)^{L-l} \quad (1.2)$$

<sup>1</sup> Chamaremos de hipótese cada mapeamento entrada-saída feito por um dado algoritmo. O conjunto de todos os mapeamentos possíveis de serem representados será definido como espaço de hipóteses.

Um dos principais fatores que influenciam no sucesso em se adotar *ensembles* é o método de geração dos componentes. O que se espera é que, ao final do processo de geração, classificadores diversos e parcimoniosos sejam obtidos. O desempenho da abordagem também depende fortemente de como os componentes são combinados no momento da classificação. Para ilustrar o comportamento de situações ideais, a figura 1 apresenta *ensembles* compostos por hipóteses independentes com probabilidades de erro de 0.35, 0.40 e 0.45 na predição de um dado padrão de consulta. Pode-se observar que quanto maior é o valor de  $p$ , maior é a probabilidade de erro do *ensemble*, mas ainda assim a probabilidade de erro do *ensemble* é menor que o valor de  $p$ . Por exemplo, para um *ensemble* com 31 hipóteses, cada uma delas possuindo uma probabilidade de erro de 0.35, a área de curva para 15 ou mais hipóteses simultaneamente incorretas é de aproximadamente 0.04, o qual é muito menor que a probabilidade de erro das hipóteses individuais. Observa-se também que quanto maior é o número de hipóteses independentes, menor é a probabilidade de erro do *ensemble*. Por outro lado, se as hipóteses individuais apresentarem erros não correlacionados, mas com probabilidades  $p > 0.5$ , então o erro do *ensemble* aumentará com o resultado da votação. Na figura 2 são exibidos exemplos de curvas do erro de *ensembles* compostos por hipóteses independentes com probabilidades de erro de 0.55, 0.60 e 0.65. Pode-se verificar que, nesses casos, a probabilidade de erros cometidos aumenta significativamente com o incremento do número de hipóteses.

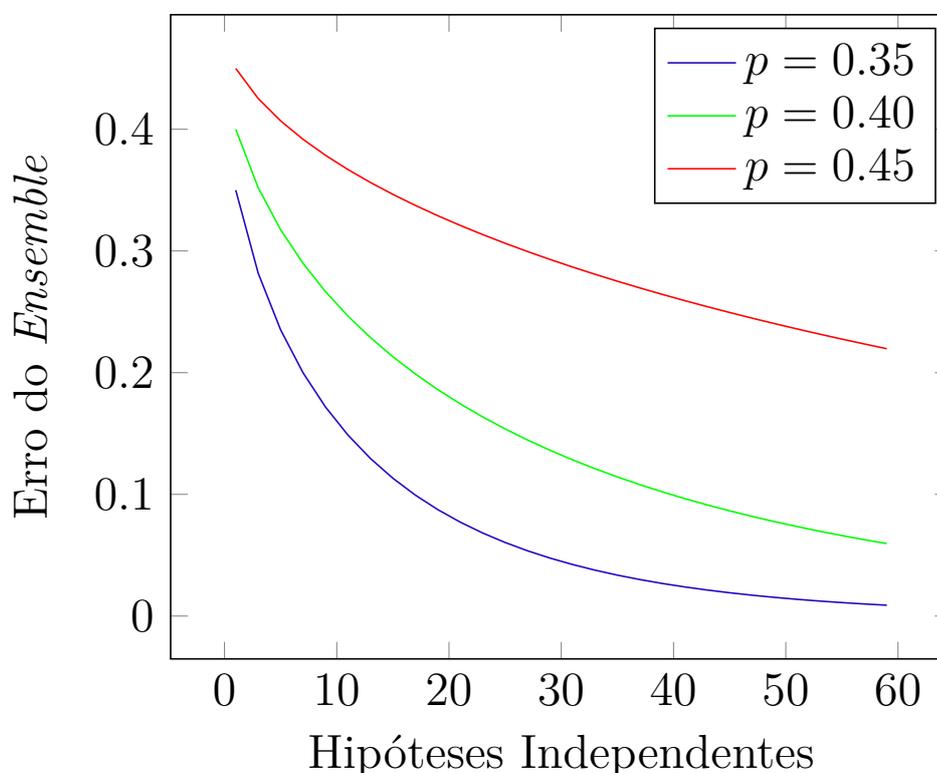


Figura 1 – Probabilidade do voto majoritário estar errado com  $L$  hipóteses independentes,  $L = 1, \dots, 60$ , e probabilidades de erro individuais de 0.35, 0.40, e 0.45

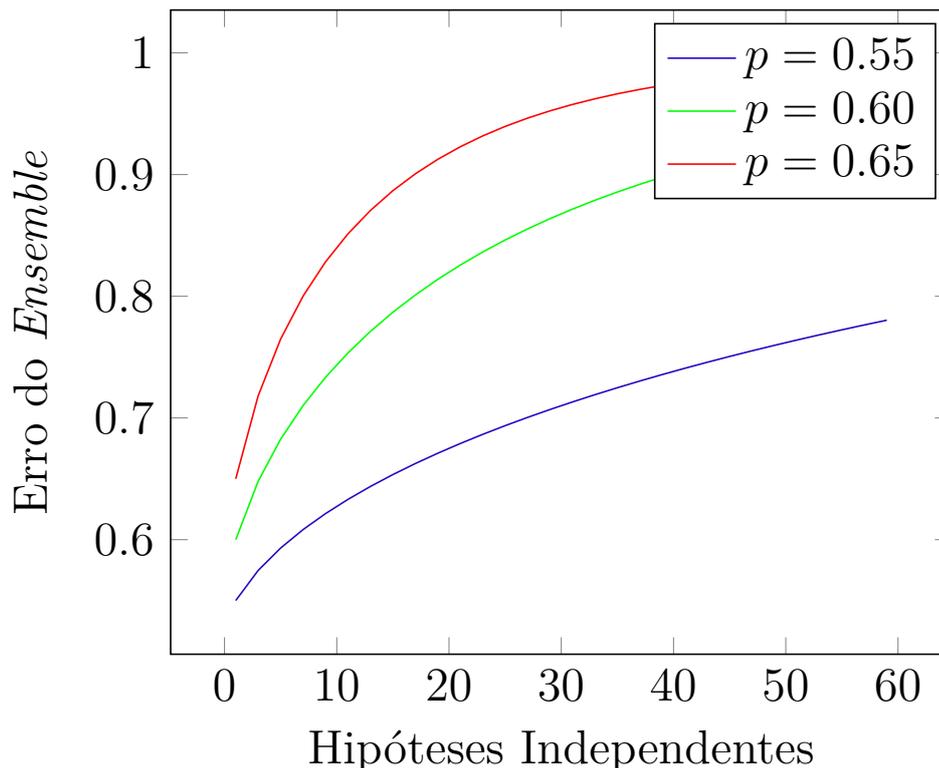


Figura 2 – Probabilidade do voto majoritário estar errado com  $L$  hipóteses independentes,  $L = 1, \dots, 60$ , e probabilidades de erro individuais de 0.55, 0.60, e 0.65

A abordagem de *ensembles* têm sido amplamente explorada como alternativa para a difícil tarefa de construir um classificador monolítico capaz de absorver toda a variabilidade de um problema, por se tratar de uma técnica bastante simples e capaz de aumentar a capacidade de generalização dos sistemas baseados em Aprendizado de Máquina. A ideia geral da combinação de múltiplos classificadores parte do conceito intuitivo de que, na grande maioria das vezes, a conjugação da opinião de vários especialistas em uma determinada área do conhecimento tende a produzir uma decisão mais precisa do que aquela gerada por um único especialista. Além disso, ainda existem três princípios fundamentais que mostram os benefícios da utilização de *ensembles* (DIETTERICH, 2000a):

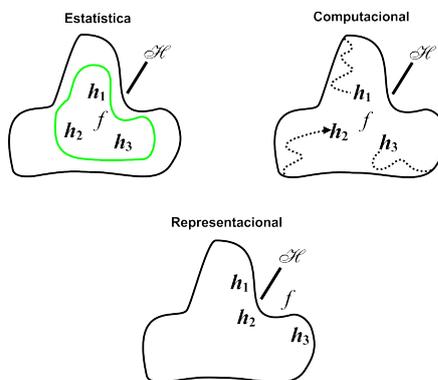


Figura 3 – Princípios que justificam porque *ensembles* podem funcionar melhor do que um único classificador. Figura adaptada de (DIETTERICH, 2000a)

1. Estatístico (ou pior caso). O algoritmo de aprendizagem de um classificador pode ser visto como um método de busca no espaço de hipóteses  $\mathcal{H}$ . A região interna na figura 3(a) delimita o conjunto das hipóteses que fornecem desempenho eficiente frente aos dados de um determinado conjunto  $Z$ . É possível escolher como solução para o problema apenas uma dessas hipóteses, correndo-se o risco de selecionar aquela com pior desempenho em termos de capacidade de generalização. Por outro lado, uma aproximação mais precisa da função ideal, representada na Figura 3 por  $f$ , pode ser alcançada através da combinação dessas hipóteses, ao invés da escolha de qualquer uma delas isoladamente. Contudo, nada garante que essa combinação venha apresentar um desempenho superior ao da melhor hipótese individual.
2. Computacional. Alguns classificadores são muito sensíveis aos algoritmos de aprendizagem, uma vez que estes, ao realizarem a busca pela hipótese  $f$ , podem ficar presos em algum ótimo local, conforme ilustra a figura 3(b). Um ótimo local ocorre quando a hipótese encontrada apresenta desempenho superior às hipóteses mais próximas, porém ainda é inferior à hipótese de melhor desempenho disponível no espaço  $\mathcal{H}$ . Ao construir um *ensemble*, com classificadores inicializados em diferentes pontos de  $\mathcal{H}$ , pode-se alcançar uma aproximação de  $f$  mais precisa do que qualquer uma das hipóteses individuais, possibilitando ainda a estabilização de sistemas formados por classificadores instáveis como redes neurais e árvores de decisão.
3. Representacional (ou melhor caso). Às vezes não é possível representar a função  $f$  pelas hipóteses existentes em  $\mathcal{H}$ . Por exemplo, para apenas classificadores lineares, quando o problema requer um classificador não linear, a função ótima  $f$  não pertencerá ao espaço  $\mathcal{H}$ . No entanto, um *ensemble* de classificadores lineares pode expandir o espaço das hipóteses representáveis (criar uma nova hipótese) e se aproximar de  $f$ , conforme mostra a figura 3(c). Existem diversas evidências experimentais de que esse fato acontece normalmente quando as hipóteses cometem diferentes erros.

Os princípios estatístico, computacional e representacional estão entre os principais motivos para os quais as abordagens de aprendizado monolíticas falham (ZHOU, 2012). Os algoritmos de aprendizagem que sofrem com os princípios estatístico ou computacional, por exemplo, geralmente apresentam uma alta variância. Por outro lado, algoritmos de aprendizagem que sofrem com o princípio representacional normalmente têm um bias elevado. Na prática, quando o bias é minimizado a variância é maximizada (vice e versa). Porém, o ideal, é que ambos valores sejam pequenos. Assim, torna-se uma tarefa muito difícil encontrar um bom e único classificador capaz de resolver um determinado problema (PONTI, 2011). Através da abordagem de *ensembles*, tanto a variância quanto o bias podem ser reduzidos. Isto foi confirmado por muitos estudos empíricos (XU; KRZYZAK; SUEN, 1992; BAUER; KOHAVI, 1999; MACLIN; OPITZ, 2011).

## 1.1 Motivações

O paradigma de *ensembles* recebeu grande destaque a partir do trabalho de *Hansen & Salamon* (HANSEN; SALAMON, 1990), onde a combinação de várias redes neurais, treinadas separadamente, pode aumentar significativamente a capacidade de generalização do sistema. Desde então, inúmeros pesquisadores têm focado seus estudos em *ensembles* e, conseqüentemente, novas soluções têm sido dedicadas para cada uma de suas possíveis fases: geração, seleção e fusão. Na primeira, um conjunto de componentes é gerado; na segunda, um subconjunto de componentes é selecionado; na última, uma decisão final é feita com base nas predições dos componentes selecionados. De forma a contribuir com o tema, as subseções a seguir apresentam os principais motivos que levaram esta tese de doutorado a recomendar o uso de alguns procedimentos, no interesse de se alcançar melhorias, na construção e utilização de *ensembles* em problemas de classificação de padrões.

### 1.1.1 Geração dos Componentes

Cada componente do *ensemble* pode ser visto como uma fronteira que separa os padrões de uma classe dos padrões da(s) outra(s). Diferenças no conjunto de treinamento, variações na ordem de apresentação dos padrões e processos internos estocásticos são capazes de fazer com que os componentes encontrem diferentes fronteiras. Uma vez que as principais estratégias para geração dos componentes não levam em consideração a otimização explícita da diversidade (BROWN et al., 2005), o *ensemble* formado pode ser composto por elementos muito similares, pois o espaço de hipóteses geralmente não é devidamente explorado. Parece razoável pensar, pelos princípios expostos na Figura 3, que com uma maior cobertura do espaço de hipóteses seria possível alcançar um maior número de padrões corretamente classificados por ao menos um dos componentes do *ensemble*.

### 1.1.2 Seleção e Fusão dos Componentes

Nenhum componente do *ensemble*, assumindo a existência da diversidade, pode ser considerado para rotular corretamente todos os padrões. Diante desse fato, recentes pesquisas têm focado na seleção dinâmica de classificadores (CRUZ; SABOURIN; CAVALCANTI, 2018). A grande maioria dos trabalhos, no entanto, geralmente se baseia numa única hipótese para realização de tal tarefa. Conforme é possível melhorar a probabilidade do sucesso de classificação com o uso de múltiplas hipóteses, como mostra o gráfico da Figura 1, também parece razoável pensar ser possível selecionar com maior grau de certeza, por meio do uso de múltiplas hipóteses, os classificadores que sejam, em teoria, considerados competentes para rotular cada padrão de consulta. Com relação à fusão, o grau de apoio dado pela competência pode ser usado na ponderação das predições de cada componente do *ensemble*, fazendo com que seja dada maior ênfase aos componentes mais competentes.

## 1.2 Objetivos

A pesquisa desta tese objetiva a construção e avaliação de um sistema híbrido inteligente capaz de superar algumas das principais limitações na geração, seleção e fusão dos componentes em *ensembles* de classificadores. Buscou-se, para tanto, propor uma abordagem capaz de gerar classificadores com foco em medidas de acuracidade e diversidade, explorando o espaço de hipóteses de forma mais eficiente. Por serem diversos, nenhum dos componentes gerados pode ser considerado competente para rotular todos os padrões existentes. Então, foi empregada uma estratégia capaz de determinar qual ou quais são os mais apropriados para rotular cada padrão em questão. Isso foi feito através de diversas medidas extraídas dos classificadores no ato da predição de cada padrão de consulta, com o intuito de se calcular a competência dos mesmos. Portanto, torna-se necessário:

- Propor e desenvolver um método para geração dos componentes de *ensembles* de classificadores, guiado por medidas de acuracidade e diversidade.
- Avaliar o comportamento do método de geração de componentes desenvolvido, considerando diferentes abordagens de seleção de classificadores.
- Propor e desenvolver uma abordagem robusta de seleção dinâmica de classificadores em que a competência é definida com base em múltiplas hipóteses.
- Avaliar o comportamento do método de seleção dinâmica desenvolvido em relação à diferentes abordagens de seleção de classificadores.
- Avaliar o desempenho final do sistema diante de soluções consagradas na literatura.

## 1.3 Produção Bibliográfica

São listados, nesta seção, trabalhos publicados pelo autor da tese em algumas das principais conferências e periódicos da área de Aprendizagem de Máquina. Essa produção contribuiu para uma melhor formação do autor na área de *ensembles*, bem como a identificação e possíveis soluções de alguns problemas existentes nas fases de geração, seleção e fusão. Toda bibliografia listada, nas seguintes subseções, trata de *ensembles* dinâmicos, sendo a maioria delas em sistemas de classificação de padrões. Em todas as publicações do autor, os componentes dos *ensembles* foram gerados por meio de um algoritmo evolucionário (variantes da Evolução Diferencial). Foram investigadas, em cada uma delas, diferentes estratégias para otimizar os parâmetros dos componentes do *ensemble*. Além disso, também foram investigadas diferentes estratégias para selecionar e/ou fundir as predições dos componentes selecionados para rotular o padrão de consulta.

### 1.3.1 Artigos Completos Publicados em Periódicos

- Anderson T. Sergio; Tiago P. F. de Lima; Teresa B. Ludermir: *Dynamic Selection of Forecast Combiners*. *Neurocomputing*, 2016. <<https://doi.org/10.1016/j.neucom.2016.08.072>>
- Tiago P. F. de Lima; Adenilton J. da Silva; Teresa B. Ludermir; Wilson R. de Oliveira: *An automatic method for construction of multi-classifier systems based on the combination of selection and fusion*. *Progress in Artificial Intelligence*, 2014. <<https://doi.org/10.1007/s13748-014-0053-6>>
- Lima, Tiago P. F; Ludermir, Teresa B: *An automatic method for construction of ensembles to time series prediction*. *International Journal of Hybrid Intelligent Systems*, 2013. <<https://doi.org/10.3233/HIS-130176>>

### 1.3.2 Trabalhos Completos Publicados em Anais de Congressos

- Tiago P. F. Lima; Teresa B. Ludermir: *Differential evolution and meta-learning for dynamic ensemble of neural network classifiers*. *International Joint Conference on Neural Networks (IJCNN)*, 2015. <<https://doi.org/10.1109/IJCNN.2015.7280553>>
- Tiago P. F. de Lima; Teresa B. Ludermir: *Ensembles of evolutionary extreme learning machines through differential evolution and fitness sharing*. *International Joint Conference on Neural Networks (IJCNN)*, 2014. <<https://doi.org/10.1109/IJCNN.2014.6889956>>
- Tiago Pessoa Ferreira de Lima; Anderson Tenorio Sergio; Teresa Bernarda Ludermir: *Improving classifiers and regions of competence in dynamic ensemble selection*. *Brazilian Conference on Intelligent Systems (BRACIS)*, 2014. <<https://doi.org/10.1109/BRACIS.2014.14>>
- Tiago Pessoa Ferreira de Lima; Teresa Bernarda Ludermir: *Optimizing dynamic ensemble selection procedure by evolutionary extreme learning machines and a noise reduction filter*. *International Conference on Tools with Artificial Intelligence (ICTAI)*, 2013. <<https://doi.org/10.1109/ICTAI.2013.87>>
- Lima, T. P. F. ; Silva, A. J. ; Ludermir, T. B.: *Selection and Fusion of Neural Networks via Differential Evolution*. *Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, 2012. <[https://dx.doi.org/10.1007/978-3-642-34654-5\\_16](https://dx.doi.org/10.1007/978-3-642-34654-5_16)>
- Lima, T. P. F. ; Silva, A. J. ; Ludermir, T. B.: *Clustering and Selection of Neural Networks Using Adaptive Differential Evolution*. *International Joint Conference on Neural Network (IJCNN)*, 2012. <<http://dx.doi.org/10.1109/IJCNN.2012.6252466>>

## 1.4 Organização do Texto

O restante deste trabalho está dividido em 6 capítulos da seguinte forma:

No capítulo 2 são abordadas as principais questões relacionadas à geração de componentes do *ensemble*, com ênfase nos algoritmos e em medidas de diversidade. No capítulo 3 é apresentada uma visão probabilística e um panorama com algumas das principais técnicas de seleção de classificadores existentes na literatura. No capítulo 4 são expostas as possíveis arquiteturas de *ensembles*, bem como algumas das mais importantes regras de fusão. No capítulo 5 são detalhadas cada uma das fases envolvidas no método proposto. No capítulo 6 são discutidos os resultados alcançados, em bases de domínio público, e realizadas comparações com recentes trabalhos sobre *ensembles*. Finalmente, no capítulo 7, são feitas as considerações finais, com destaques para as principais vantagens e desvantagens do método proposto, bem como a realização de trabalhos futuros.

## 1.5 Considerações Finais

*Ensembles* é um paradigma de Aprendizado de Máquina que explicita a combinação de módulos independentes, denominados componentes, visando uma solução única e que seja capaz de oferecer respostas com maior capacidade de generalização do que aquelas alcançadas por cada componente agindo isoladamente. Como já mencionado, a proposta desta tese é a construção de um sistema híbrido inteligente capaz de gerar, selecionar e fundir as predições dos classificadores. O presente capítulo relatou as principais motivações para realização do trabalho, apresentou os objetivos de pesquisa almejados e as principais contribuições alcançadas, justificando a importância deste estudo para a área de *ensembles* de classificadores. Finalmente, mostrou como está estruturado o restante do texto.

## 2 GERAÇÃO DOS COMPONENTES

Métodos de seleção dinâmica de classificadores tentam eleger o(s) componente(s) do *ensemble*, que seja(m), em teoria, o(s) mais apropriado(s) para rotular cada padrão de consulta. Esses algoritmos buscam, segundo alguma hipótese, avaliar a vizinhança do padrão a ser classificado e, com base nesta informação, calcular a competência. A efetividade em se adotar a seleção dinâmica, no entanto, depende de que haja diversidade, de tal forma que seja possível gerar combinações de resultados que viabilizem mudanças no processo de classificação em direção à correta predição de cada padrão de consulta. Portanto, a geração dos componentes é essencial para se alcançar melhores resultados.

### 2.1 Diversidade em Esembles

A necessidade de diversidade em *ensembles* é de certa forma intuitiva, uma vez que a combinação de hipóteses não diversas implica em acertos/erros coincidentes, fazendo com que a combinação traga apenas um aumento no custo computacional. Para que um *ensemble* possa ter boa diversidade, cada componente tem que apresentar diferentes capacidades de generalização, ou seja, os erros não devem ser correlacionados (SHARKEY; SHARKEY, 1997). Entretanto, a maior dificuldade na combinação de múltiplos classificadores é que as hipóteses tendem a ser correlacionadas (JACOBS, 1995). A diversidade em *ensembles* pode ser alcançada, segundo (KUNCHEVA, 2004a), quando os seus componentes são construídos por diferentes métodos, sendo os mais comuns citados a seguir:

- **Métodos que atuam sobre o ponto de partida no espaço de hipóteses:** os métodos inclusos nesta categoria variam os pontos iniciais da busca no espaço de hipóteses, influenciando dessa forma o ponto de convergência dos componentes.
- **Métodos que atuam sobre os dados de treinamento:** estes métodos buscam gerar diferentes mapeamentos entrada-saída, na construção de cada um dos componentes do *ensemble*, por meio do uso de diferentes conjuntos de treinamento.
- **Métodos que manipulam a arquitetura de cada componente:** através dos métodos desta categoria, varia-se a arquitetura de cada um dos componentes do *ensemble*, de tal maneira que diferentes hipóteses estejam acessíveis.
- **Métodos que atuam sobre a forma de exploração do espaço de hipóteses:** alterando a forma de exploração do espaço de hipóteses, estes métodos levam os componentes do *ensemble* a convergirem para diferentes pontos.

Existem propostas para se avaliar a diversidade entre classificadores e isso pode ajudar a escolher os componentes mais diversos na construção de um *ensemble*. Porém, nenhuma delas é aceita uniformemente, pois ainda não foi provada relação formal entre as mesmas e o erro total da combinação dos componentes. Segundo *Kuncheva & Whitaker* (KUNCHEVA; WHITAKER, 2003), essas medidas podem ser divididas em dois grupos:

- **Medidas sem paridade:** medem a diversidade considerando todos os componentes juntos, calculando diretamente um valor para o *ensemble*. Fazem parte deste grupo a entropia dos votos, índice de dificuldade, variância de *Kohavi-Wolpert*, a relação de concordância entre classificadores, a diversidade generalizada e erros coincidentes.
- **Medidas com paridade:** são calculadas para cada par de classificadores, sendo a diversidade total do *ensemble* obtida pela média desses pares. Fazem parte deste grupo as medidas de estatística Q, correlação, falta dupla e discordância.

Os algoritmos de geração dos componentes que empregam medidas de diversidade no processo construtivo são ditos explícitos, enquanto aqueles que não adotam tais medidas são chamados de implícitos (BROWN et al., 2005). Vale salientar que a grande maioria dos trabalhos, que fazem uso de algum método para geração de componentes, utiliza algoritmos implícitos, com principal ênfase aos que serão apresentados na seção seguinte.

## 2.2 Algoritmos de Geração - Clássicos

A ideia de se construir um *ensemble* é obter ganho de generalização a partir da combinação das hipóteses de seus componentes. Tal combinação é em vistas da minimização do erro individual, uma vez que os componentes do *ensemble*, quando aplicados isoladamente, produzem erro. Portanto, é extremamente necessário que haja diferença na generalização, ou seja, diversidade, como citado na seção anterior, fazendo com que os componentes não apresentem erros correlacionados, e sim uma dissimilaridade do erro.

Os componentes de um *ensemble* podem ser construídos basicamente por métodos homogêneos ou heterogêneos. Na primeira abordagem, adota-se um mesmo algoritmo de aprendizagem. Já na segunda, diferentes algoritmos são aplicados ao longo do processo de geração. Não faz parte do escopo deste trabalho a abordagem heterogênea, portanto será dado destaque aos algoritmos homogêneos. Dentre os mais utilizados, destacam-se: *Bagging* (BREIMAN, 1996), *Boosting* (SCHAPIRE, 1990) e *Random Subspace* (HO, 1998). De forma geral, esses três algoritmos têm apresentado melhoria na acuracidade de determinados tipos de classificadores quando aplicados a conjuntos de testes reais e artificiais.

### 2.2.1 Bagging

O algoritmo é baseado na ideia *Bootstrap Aggregating* (EFRON; TIBSHIRANI, 1993). Ele provê a diversidade, lançando-se mão do conceito de redistribuição aleatória dos dados. Conforme mostra a Figura 4, na construção de cada  $h_i \in H$ , para um dado conjunto de treinamento  $\tau$  de tamanho  $n$ , é gerado, por reamostragem uniforme, um subconjunto de dados (*bag*)  $\tau'$ , de mesmo tamanho. Desse modo, um padrão  $\vec{x} \in \tau$  tem probabilidade  $1 - (1 - 1/n)^n$  de ser selecionado, permitindo assim que aproximadamente 63.2% do conjunto  $\tau'$  seja composto de amostras únicas e o restante de valores duplicados. *Bagging* tem se mostrado estável quando aplicado em conjunto de dados ruidosos e bastante atrativo pelo fato da construção dos componentes poder ser realizada em paralelo. No entanto, a existência de  $L$  conjuntos de dados distintos não implica que os  $L$  componentes a serem treinados irão generalizar de forma distinta. Para sustentar esta implicação, segundo *Panov & Dzeroski* (PANOV; DZEROSKI, 2007), o algoritmo é indicado para classificadores instáveis<sup>1</sup>, os quais sofrem grande influência de pequenas variações no conjunto de treino. Caso contrário, o *ensemble* resultante será uma coleção de classificadores quase idênticos, portanto, pouco provável de melhorar o desempenho de um classificador único.

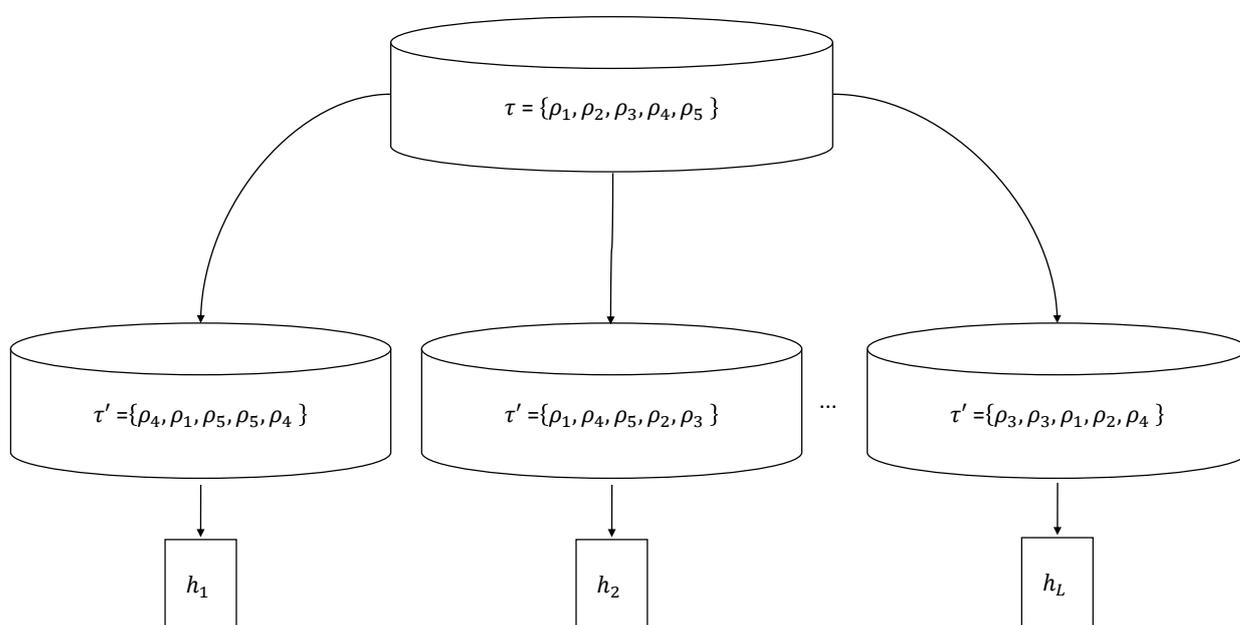


Figura 4 – Funcionamento do *Bagging*

<sup>1</sup> O uso de classificadores instáveis possibilita a criação de fronteiras de decisão diferentes, mesmo usando dados de treinamento similares. Árvores de decisão e redes neurais são exemplos clássicos.

## 2.2.2 Boosting

O algoritmo é semelhante ao *Bagging* no quesito reamostragem dos dados. No entanto, é obrigatoriamente sequencial e também não utiliza subconjuntos de dados  $\tau'$  via reamostragem aleatória uniforme. Desse modo, enquanto *Bagging* atua apenas na redução da variância<sup>2</sup>, *Boosting* é capaz de reduzir tanto a variância quanto o bias<sup>3</sup> (FRIEDMAN; HASTIE; TIBSHIRANI, 2000). Isso se deve ao fato da distribuição de probabilidade associada a cada conjunto  $\tau'$  passar a ser ajustada adaptativamente, conforme mostra a figura 5. Existem diferentes maneiras de se realizar este ajuste, mas o que há em comum entre elas é o fato de na medida em que os padrões de treinamento forem preditos incorretamente, os mesmos passam a ter maior probabilidade (peso) de serem integrados ao próximo conjunto  $\tau'$ , de forma que os últimos componentes do *ensemble* a serem treinados, geralmente, arcam com as regiões mais difíceis do espaço de instâncias.

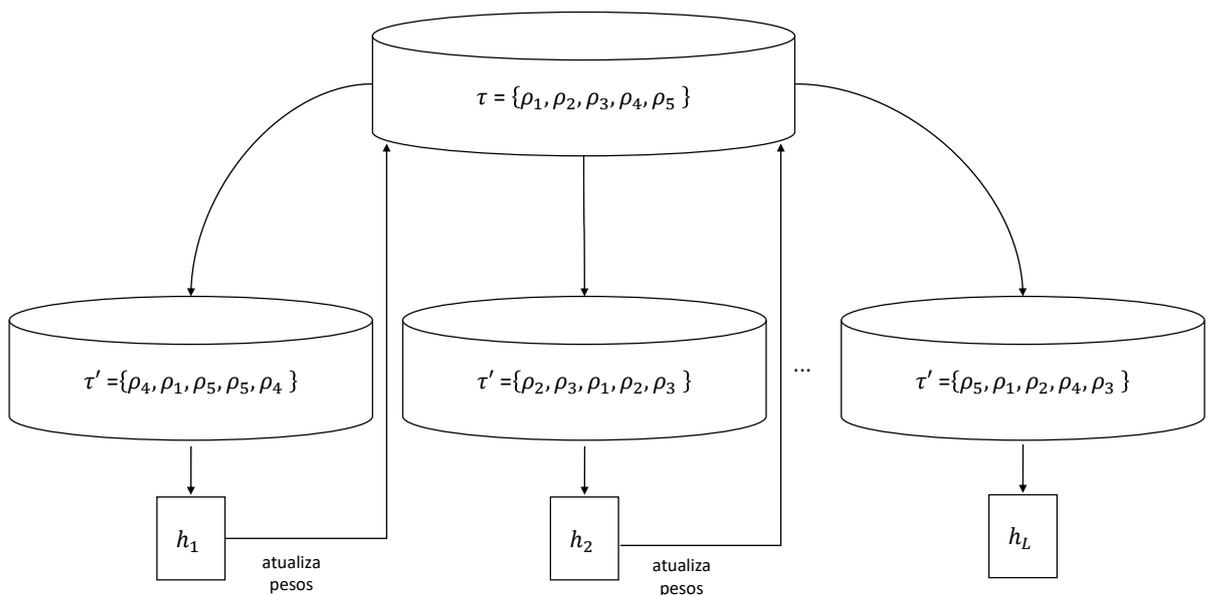


Figura 5 – Funcionamento do *Boosting*

O desempenho do *Boosting* quando aplicado a dados ruidosos geralmente não é bom (DIETTERICH, 2000b). Isso já é esperado, visto que, como os componentes são treinados hierarquicamente ao passo que são construídos, os mesmos vão se especializando em áreas do espaço de características que provavelmente irão dispor de valores corrompidos, levando à perda da capacidade de generalização. Por outro lado, geralmente apresenta, quando o problema de ruído é ausente, ganhos significativos na capacidade de generalização.

<sup>2</sup> Mede a precisão do classificador, ou seja, quanto maior ela for, menos confiável será a predição.

<sup>3</sup> Uma predição ruim pode estar relacionada a um alto bias, possivelmente causado por sobreajustamento.

### 2.2.3 *Random Subspace*

O algoritmo utiliza projeções dos padrões em subespaços como mecanismos para diversificação dos componentes. Das  $d$  características originais do problema, conforme mostra a figura 6, são selecionadas  $d'$  características aleatórias e sem repetição, sendo que  $d' < d$ . A escolha casual das características possibilita a criação de classificadores que são complementares, o que faz com que os mesmos cometam erros diferentes, que é uma fator positivo em cenários de combinação de classificadores (PONTI, 2011). A vantagem de diferentes subespaços é a possibilidade da redução do número de características irrelevantes e/ou redundantes, além de, consequentemente, possibilitar a utilização de *ensembles* em problemas de grandes dimensionalidades. Porém, um número insuficiente de características pode levar a uma fraca discriminação e consequentemente a uma precisão inferior.

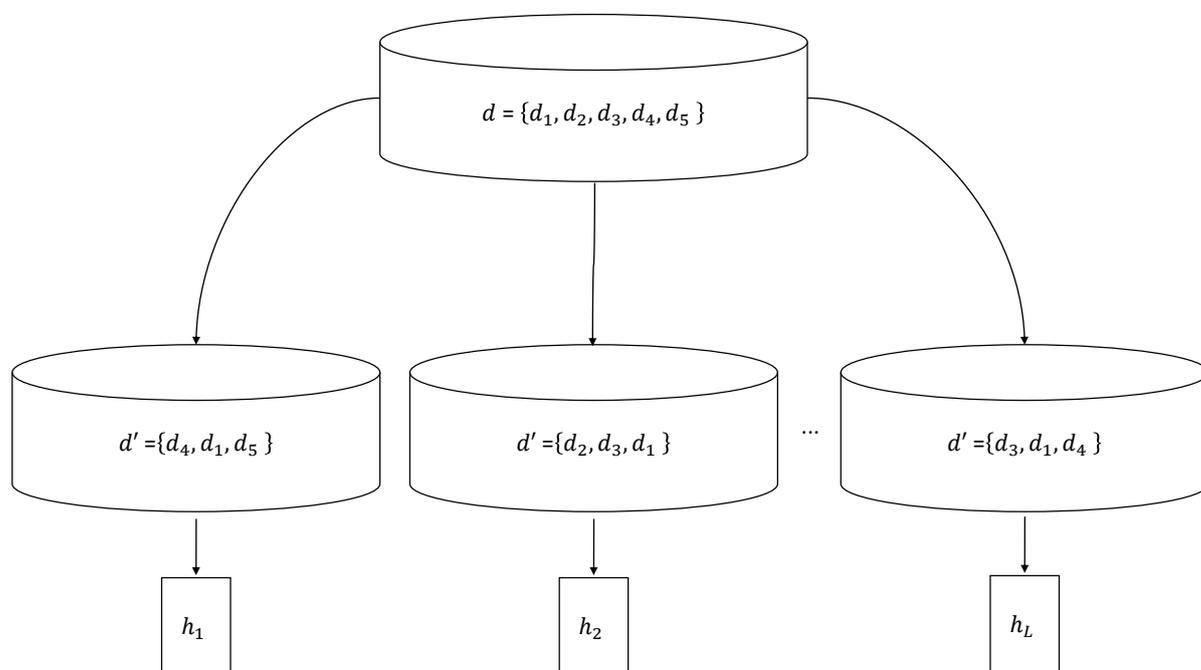


Figura 6 – Funcionamento do *Random Subspace*

Na ilustração apresentada na Figura 6, o conjunto de treinamento é composto de 5 características, das quais apenas 3 são sorteadas para a geração de cada componente. É importante destacar que não devem ser sorteadas características repetidas para a construção de um mesmo classificador, uma vez que isso não traria ganhos no momento da classificação. Todavia, classificadores distintos podem possuir características em comum.

## 2.3 Algoritmos de Geração - Evolucionários

Tem havido, nos últimos anos, um crescimento perceptível do uso da Computação Evolucionária em inúmeros problemas (EIBEN; SMITH, 2015), incluindo o projeto de geração/otimização dos componentes do *ensemble* (DUFOURQ; PILLAY, 2014; LACY; LONES; SMITH, 2015; ALMEIDA; GALVÃO, 2016; ROSALES-PEREZ et al., 2017). Algoritmos Evolucionários (AE) são métodos estocásticos, baseados em populações constituídas por indivíduos com comportamento específico e similar ao fenômeno biológico de evolução das espécies. Esses algoritmos têm a capacidade de considerar informações em diferentes pontos do espaço, realizando uma busca paralela por potenciais soluções do problema. A ideia na geração dos componentes é evoluir uma população formada por indivíduos aleatórios, através de, basicamente, manipulações na arquitetura e/ou manipulações no conjunto de treinamento, de forma que o conjunto final apresente maior acuracidade e diversidade. Por não serem baseados em informações de gradiente, esses algoritmos podem ser aplicáveis em diversos contextos, pois independem da natureza das funções objetivo/*fitness* e também das restrições do problema adotado (MARLER; ARORA, 2004). Além do mais, AEs são provavelmente otimizadores globais (GOLDBERG, 1989).

## 2.4 Considerações Finais

Sistemas de múltiplos classificadores destinam-se a soluções de problemas com maior grau de complexidade, ou seja, aqueles que geralmente apresentam um grande número de classes, dados ruidosos, padrões com pouca representatividade, e/ou que não apresentem uniformidade na distribuição dos dados. Utilizar um classificador monolítico pode ser ineficiente para se obter a solução desejada nestas situações. Entretanto, para cenários mais simples, a escolha de um classificador monolítico pode ser suficiente para solucionar o problema, obtendo altas taxas de correta classificação e principalmente baixo custo computacional. Neste capítulo foi feita uma introdução sobre a importância da diversidade na geração dos componentes de um *ensemble*, bem como breves comentários sobre algumas das principais métricas exploradas até o momento.

## 3 SELEÇÃO DOS COMPONENTES

Após a etapa de geração, dado um conjunto de  $L$  classificadores, a melhor alternativa para combinação não necessariamente implica na utilização de todos eles, mas na escolha daqueles que se demonstrarem ser complementares. Assim, o próximo passo é escolher quais componentes realmente contribuem para o processo de classificação. Esta seleção pode ser realizada de forma estática ou dinâmica (KO; SABOURIN; BRITTO, 2008). A seleção estática geralmente é realizada durante a fase de treinamento, sem considerar características dos padrões a serem classificados (AYAD; SYED-MOUCHAWEH, 2011). Neste cenário, os classificadores que se mostram mais acurados são designados para formar o *ensemble* que será empregado para rotular todas as novas instâncias. Por outro lado, na seleção dinâmica, a escolha do(s) classificador(es) leva em consideração particularidades dos classificadores na predição de cada padrão de consulta. Dessa forma, os componentes que serão combinados podem variar de acordo com a instância em questão.

### 3.1 Seleção Dinâmica de Classificadores

A adoção da seleção dinâmica de classificadores visa explorar de forma mais efetiva a diversidade dos componentes de um *ensemble*, com intuito de melhorar a acuracidade de classificação em relação à seleção estática (TSOUMAKAS; PARTALAS; VLAHAVAS, 2008). Para demonstrar como isso é possível, sejam  $H = \{h_1, h_2, \dots, h_L\}$  um conjunto de  $L$  classificadores e  $E = \{e_1, e_2, \dots, e_M\}$  ser um conjunto de  $M$  *ensembles* formados a partir de  $H$ . A seleção dinâmica pode ser vista como uma divisão do espaço de instâncias em  $K > 1$  regiões de competência, denotadas por  $R_1, R_2, \dots, R_K$ . Então, para cada região  $R_j$ ,  $j = 1, 2, \dots, K$ , é designado o *ensemble* mais competente em  $E$ . A figura 7 ilustra como pode ser feita a divisão do espaço de instâncias em regiões de competência.

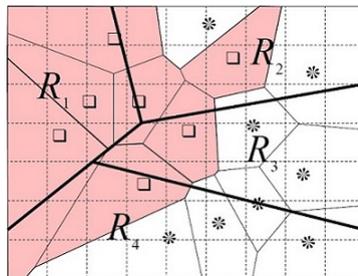


Figura 7 – Exemplo de um espaço de instâncias bidimensional particionado em 4 regiões de competência. Figura extraída de (KUNCHEVA, 2000).

Seja  $e^* \in E$  o *ensemble* com a maior acuracidade em todo o espaço de instâncias (seleção estática). Denote por  $p(e_i|R_j)$  a probabilidade de classificação correta de  $e_i$  em  $R_j$ .

Considere  $e_{ij}$  como sendo o *ensemble* designado para  $R_j$ . A probabilidade de classificação correta  $p_c$  é descrita na equação 3.1, onde  $p(R_j)$  é a probabilidade de um padrão  $\vec{x}$  pertencer a  $R_j$ . Para maximizar  $p_c$ , deve-se atribuir  $e_{ij}$  conforme a equação 3.2.

$$p_c = \sum_{j=1}^K p(R_j)p_c(R_j) = \sum_{j=1}^K p(R_j)P(e_{ij}|R_j) \quad (3.1)$$

$$p(e_{ij}|R_j) \geq p(e_t|R_j), t = 1, 2, \dots, M \quad (3.2)$$

$$\sum_{j=1}^K p(R_j)P(e_{ij}|R_j) \geq \sum_{j=1}^K p(R_j)p(e^*|R_j) \quad (3.3)$$

A partir das equações 3.1 e 3.2, tem-se na equação 3.3 que ao selecionar dinamicamente  $e_{ij}$ , sendo este considerado o mais competente *ensemble* na região, pode-se alcançar uma probabilidade de classificação correta superior ou igual ao do *ensemble*  $e^*$ , independentemente da maneira que o espaço de características tenha sido particionado. As principais técnicas existentes na literatura para se eleger  $e_{ij}$ , segundo os mais recentes levantamentos de seleção dinâmica de classificadores (BRITTO; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018), são apresentadas nas próximas subseções.

### 3.1.1 Seleção Baseada em *Ranking*

Nos métodos desta categoria, a competência (o *ranking*) de cada componente do *ensemble*  $H$  é calculada como sendo o número de padrões consecutivos da região de competência que forem corretamente classificados. O componente  $h_i \in H$  que alcançar o maior *ranking* é selecionado para classificar o padrão de consulta  $\vec{x}$ . Uma importante abordagem desta categoria foi proposta por (SABOURIN et al., 1993).

### 3.1.2 Seleção Baseada em Acurácia

A ideia deste esquema é estimar a competência de cada classificador com base no número de acertos na região de competência. As duas principais variantes dessa subcategoria foram propostas em (WOODS; KEGELMEYER W.P.; BOWYER, 1997). A primeira, conhecida por *Overall Local Accuracy* (OLA), calcula a taxa de acerto dos classificadores numa região definida em função dos  $k$  vizinhos mais próximos do padrão de consulta  $\vec{x}$ , obtidos através do conjunto de treinamento. A segunda, conhecida por *Local Class Accuracy* (LCA), difere do OLA pelo fato de limitar a vizinhança para cada classificador  $h_i \in H$ . Ou seja, dado que  $h_i$  classifique  $\vec{x}$  como sendo  $\omega$ , a vizinhança será formada considerando apenas os padrões pertencentes à classe  $\omega$ . Ambas variantes selecionam o classificador com o maior número de vizinhos corretamente classificados.

### 3.1.3 Seleção Baseada em Probabilidade

*Giacinto e Roli* (GIACINTO; ROLI, 1999) demonstraram que um classificador Bayesiano ótimo pode ser obtido pela seleção de classificadores ao se considerar o grau de apoio das suas saídas<sup>1</sup>. A partir daí tais autores passaram a assumir cálculos probabilísticos ao invés de apenas considerar o número vizinhos corretamente classificados, como nas variantes OLA e LCA propostas por (WOODS; KEGELMEYER W.P.; BOWYER, 1997). Nesse novo procedimento, foram expressas duas sistemáticas: *a priori* e *a posteriori*. Para a sistemática *a priori*, seja  $p_j(\omega|\vec{x})$  o grau de apoio dado pelo classificador  $h_j$  para a classe de  $\vec{x}$ . Então, sendo  $\vec{x}_i$  o  $i$ -ésimo padrão da vizinhança do padrão de consulta  $\vec{x}$ , a competência de  $h_j$  na região é calculada pela equação 3.4. Para a sistemática *a posteriori* são utilizados apenas os padrões cuja classe possui a mesma rotulação que o classificador  $h_j$  atribui a  $\vec{x}$ . Ambas sistemáticas selecionam o classificador que maximiza a competência na vizinhança. Com o intuito de minimizar incertezas inerentes ao tamanho das regiões de competência, *Giacinto e Roli* propuseram a aplicação de pesos conforme mostra a equação 3.5, onde  $W_i = 1/d_i$  e  $d_i$  é a distância euclidiana entre os padrões  $\vec{x}_i$  e  $\vec{x}$ .

$$Competencia(h_j|\vec{x}) = \sum_{i=1}^K p_j(\omega|\vec{x}_i) \quad (3.4)$$

$$Competencia(c_j|\vec{x}_{tes}) = \frac{\sum_{i=1}^K p_j(\omega|\vec{x}_i)W_i}{\sum_{i=1}^K W_i} \quad (3.5)$$

### 3.1.4 Seleção Baseada em *Clustering*

Nos algoritmos de agrupamento (ou *clustering*) o objetivo é dividir um conjunto de objetos em subconjuntos disjuntos, maximizando a intra-similaridade e minimizando a inter-similaridade, ou seja, os objetos que pertencem a um mesmo grupo (ou *cluster*) devem ser similares, ao passo que objetos de grupos distintos devem ser dissimilares. No contexto da seleção dinâmica de classificadores, esta estratégia recebeu grande destaque a partir da publicação de *Kuncheva* (KUNCHEVA, 2000). Neste trabalho, o espaço de instâncias foi particionado em  $k > 1$  regiões de competência. Nesse particionamento, cada região  $R_j$  é representada por um centróide e possui um classificador  $h_{ij}$  que foi previamente estabelecido por apresentar a melhor performance na  $j$ -ésima região. Nos métodos desta categoria, a seleção dos classificadores é definida durante o processo de treinamento. Por isso, são geralmente mais rápidos para rotular um dado padrão de consulta.

<sup>1</sup> Detalhes sobre respostas/saídas dos classificadores serão abordados na seção 4.2 do capítulo seguinte.

### 3.1.5 Seleção Baseada em Comportamento

A seleção dinâmica de classificadores através do conceito de comportamento foi introduzida por (GIACINTO; ROLI, 2001) com inspiração em (HUANG; SUEN, 1995). Neste método, conhecido por *Multiple Classifier Behaviour* ou *MCB*, cria-se um vetor para cada padrão  $\vec{x}$ , composto pelas hipóteses dos  $L$  classificadores, tal como na equação 3.6. Sendo  $\vec{x}_{tes}$  e  $\vec{x}_{tr}$ , respectivamente, padrões dos conjuntos de teste e treinamento, calcula-se o índice de similaridade através da equação 3.7, com  $T_i(\vec{x}_{tes}, \vec{x}_{tr})$  dado pela equação 3.8.

$$MCB(\vec{x}) = \{c_1(\vec{x}), c_2(\vec{x}), \dots, c_L(\vec{x})\} \quad (3.6)$$

$$S(\vec{x}_{tes}, \vec{x}_{tr}) = \frac{1}{L} \sum_{i=1}^L T_i(\vec{x}_{tes}, \vec{x}_{tr}) \quad (3.7)$$

$$T_i(\vec{x}_{tes}, \vec{x}_{tr}) = \begin{cases} 0 & \text{se } c_i(\vec{x}_{tes}) \neq c_i(\vec{x}_{tr}) \\ 1 & \text{se } c_i(\vec{x}_{tes}) = c_i(\vec{x}_{tr}) \end{cases} \quad (3.8)$$

A equação 3.7 define a relação de vizinhança entre os padrões  $\vec{x}_{tes}$  e  $\vec{x}_{tr}$ , havendo maior proximidade à medida que o índice de similaridade aumenta. São utilizados como vizinhos os  $k$  padrões do conjunto de treinamento mais similares ao padrão  $\vec{x}_{tes}$  ou todos os padrões do conjunto de treinamento que obtiverem um índice de similaridade superior ao limiar que foi previamente estabelecido. Por fim, seleciona-se o classificador que apresentar a melhor competência, ou seja, o maior número de vizinhos corretamente classificados.

### 3.1.6 Seleção Baseada em Oráculo

Considera-se na seleção dinâmica de classificadores a existência de um oráculo, considerado como sendo o limite superior de reconhecimento por atribuir a classe correta se ao menos um componente do *ensemble* acertar o padrão em questão. Com o intuito de se aproximar desse limite teórico, os métodos dessa subcategoria selecionam os classificadores que melhor estimam a região de competência. A principal variante é o *k*-nearest-oracles (KNORA), proposta por (KO; SABOURIN; BRITTO, 2007). Na versão KNORA-E, cada classificador que prevê corretamente todos os  $k$  vizinhos terá direito a um voto. Já para o KNORA-U, cada classificador terá um voto para cada vizinho corretamente classificado. Como em *Giacinto e Roli* (GIACINTO; ROLI, 1999), ainda existem versões em que as distâncias são utilizadas para ponderar os votos. Tais versões mantêm as definições anteriores e são denominadas KNORA-EW e KNORA-UW. Apesar dos métodos dessa subcategoria serem inspirados no oráculo, o limite superior de reconhecimento ainda está longe de ser alcançado. Um fato que pode contribuir pra isso é que, quando há padrões ruidosos na região de competência, existe probabilidade de *overfitting* e consequentemente os classificadores selecionados apresentam menor capacidade de generalização.

### 3.1.7 Seleção Baseada em Acuracidade e Diversidade

A utilização simultânea da acuracidade e diversidade como critério de escolha dos membros de um *ensemble* mostra-se, em geral, superior do que estas medidas agindo separadamente. Um método que comprova essa afirmação foi tratado por (SANTANA et al., 2006). Nesta estratégia são selecionados os  $S$  classificadores com maior acuracidade na região de competência. Dos  $S$  classificadores selecionados, são escolhidos os  $D < S$  mais diversos para compor o *ensemble*. Para o cálculo da diversidade utilizou-se a medida *Double Fault* (DF), conforme a equação 3.9 onde  $N^{00}$  é o número de padrões que ambos classificadores erram;  $N^{01}$  é o número de padrões onde o primeiro classificador erra e o segundo acerta;  $N^{10}$  é o número de padrões onde o primeiro classificador acerta e o segundo erra; e  $N^{11}$  é o número de padrões onde ambos os classificadores acertam. Por se tratar de uma medida pareada, realiza-se uma média entre todos os componentes.

$$DF = \frac{N^{00}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (3.9)$$

### 3.1.8 Seleção Baseada em Diversidade

Uma abordagem que emprega diversidade na seleção dinâmica de classificadores foi apresentada em (YAN et al., 2013). O método proposto, chamado *Sorting-Based Dynamic Classifier Ensemble Selection* (SDES), é dividido em duas etapas. Baseia-se na ideia de quanto maior for a diversidade entre os componentes selecionados, maiores serão as chances de acerto na classificação dos padrões de consulta. Na sua primeira etapa, realiza a ordenação dos componentes de acordo com uma medida pareada de diversidade empregando o índice  $k_p$ , que é definido conforme mostra a equação 3.10. Na segunda, os classificadores são selecionados segundo a ordem construída na primeira etapa até que a confiança na classificação do padrão de consulta atinja um limiar pré-estabelecido. Quando o patamar é atingido, a classe cuja confiança foi superior ao limiar é atribuída ao padrão.

$$k_p = \frac{2 \times (N^{00} \times N^{11} - N^{10} \times N^{01})}{(N^{00} + N^{10}) \times (N^{10} + N^{01}) + (N^{00} + N^{01}) \times (N^{01} + N^{11})} \quad (3.10)$$

Apesar dos pesquisadores concordarem que a diversidade é um aspecto de grande importância na combinação de múltiplos classificadores, o mesmo não acontece em relação às medidas utilizadas para quantificá-la. De fato, medir a diversidade não é uma tarefa tão simples de se realizar, dada a complexidade de sua interpretação. Além do mais, muitas das medidas existentes são ineficientes para o uso direto na construção dos componentes de um *ensembles* (TANG; SUGANTHAN; YAO, 2006).

### 3.1.9 Seleção Baseada em Ambiguidade

Os métodos nesta subcategoria priorizam o consenso entre os elementos do *ensemble* ao invés da diversidade. Conforme apresentado em (SANTOS; SABOURIN; MAUPIN, 2007), o consenso pode ser mensurado através de um simples cálculo de ambiguidade. A ambiguidade  $\alpha_i(\vec{x})$  do classificador  $h_i$  no padrão de consulta  $\vec{x}$  é representada pela equação 3.11, sendo  $e$  o *ensemble* ao qual  $h_i$  pertence. Já a ambiguidade de um dado *ensemble* para o padrão  $\vec{x}$  é definida como sendo a soma das ambiguidades de todos os classificadores que o compõe, conforme mostra a equação 3.12. O intuito dessa técnica é, dentre um conjunto de *ensembles* disponíveis, selecionar aquele em que existe a menor ambiguidade possível.

$$\alpha_i(\vec{x}) = \begin{cases} 0 & \text{se } c_i(\vec{x}) = e(\vec{x}) \\ 1 & \text{se } c_i(\vec{x}) \neq e(\vec{x}) \end{cases} \quad (3.11)$$

$$\text{Ambiguidade} = \sum_{i=1}^L \alpha_i(\vec{x}) \quad (3.12)$$

### 3.1.10 Seleção Baseada em Oráculo Randômico Linear

A abordagem desenvolvida por (KUNCHEVA; RODRIGUEZ, 2007) divide os componentes do *ensemble* em dois subclassificadores, empregando-se para tanto uma função linear aleatória. O processo consiste em sortear dois componentes e traçar uma reta entre os mesmos. Um hiperplano é então obtido perpendicularmente à reta construída. O objetivo da construção desse hiperplano não é a otimização dos conjuntos, como ocorre no classificador *Support Vector Machine* (SVM), mas incorrer em maior diversidade.

Cada um dos subclassificadores formados é responsável por um espaço distinto, cabendo ao oráculo, no instante da classificação, determinar/selecionar qual deles é o mais competente para rotular cada instância em questão. A classe atribuída é decidida por meio do uso de alguma regra de decisão, como por exemplo a votação majoritária simples. O objetivo da utilização do oráculo é dividir um problema grande, representado pelo classificador, em dois problemas menores, representados pelos subclassificadores. Segundo *Kuncheva & Rodríguez*, espera-se que, com o uso dos subclassificadores, se obtenha um desempenho melhor, ou no mínimo igual, ao do classificador monolítico.

### 3.1.11 Seleção Baseada em *Overproduce-and-choose*

Uma proposta para a seleção dinâmica de classificadores empregando a estratégia *Overproduce-and-choose* (OCS) foi desenvolvida por (SANTOS; SABOURIN; MAUPIN, 2008). O objetivo dessa abordagem foi contornar problemas apresentados pela estratégia OCS estática, que seleciona um único *ensemble* para classificar todos os padrões existentes, ou seja, todos os classificadores são tratados com a mesma importância. *Santos et al.*

propuseram uma solução que forma uma gama de *ensembles* e, de acordo com o padrão de consulta a ser classificado, seleciona aquele que melhor se adequar. Os *ensembles* são formados através de dois Algoritmos Genéticos (AG), um mono-objetivo e o outro multiobjetivo. Como critérios de otimização, foram empregadas: a minimização do erro (no AG mono-objetivo) e a maximização da diversidade (onde foram empregadas a falta dupla, ambiguidade, falha coincidente da diversidade e medida de dificuldade no AG multiobjetivo). No momento da seleção dinâmica, são aplicadas medidas de confiança (ambiguidade, margem, força e acurácia local) para que se possa determinar qual o classificador ou *ensemble* será escolhido para rotular o padrão em questão.

### 3.1.12 Seleção Baseada em Filtros e Distância Adaptativa

O trabalho desenvolvido por Cruz *et al.* (CRUZ; CAVALCANTI; REN, 2011) propõe melhorias nas regiões de competência. O objetivo é eliminar/diminuir padrões que podem incorrer em erros de classificação. O método apresentado, chamado DES-FA (*Dynamic Ensemble Selection by Filter + Adaptative Distance*) atua em duas etapas na preparação dos dados do conjunto destinado a ser utilizado na região de competência (conjunto de validação). A primeira etapa, chamada *Edited Nearest Neighbor Filter* (ENN *Filter*), trabalha eliminando dados que são considerados ruidosos, de modo a criar fronteiras de decisão mais suaves. Para tanto, o processo aplica um classificador *K-Nearest Neighbor* (KNN) sobre o conjunto de validação de modo a excluir os padrões que forem classificados incorretamente. A segunda etapa, intitulada *K-Nearest Neighbor with Adaptative Distance*, visa a aplicação de uma medida, chamada distância adaptativa, de forma que os padrões cuja vizinhança apresentam rótulos da mesma classe do padrão de consulta têm pesos maiores do que aqueles com rótulos diferentes. A classificação emprega o algoritmo KNORA-E (KO; SABOURIN; BRITTO, 2007) para seleção dos classificadores.

### 3.1.13 Seleção Baseada em Meta-Aprendizado

Com o intuito de tornar o processo de seleção dinâmica de classificadores mais robusto, Cruz *et al.* (CRUZ *et al.*, 2015) sugerem o uso de 5 meta-características para calcular a competência dos componentes do *ensemble*: i) dificuldade na classificação dos vizinhos; ii) probabilidade posteriori; iii) acurácia total local; iv) perfis de saída dos componentes; e v) confiança dos componentes. Na fase chamada de meta-treino, um conjunto de meta-treinamento, formado por essas meta-características, é extraído dos classificadores base e usado para a construção de um meta-classificador que irá atuar como seletor na fase de generalização. Portanto, para a classificação de um dado padrão de consulta  $\vec{x}$ , a seleção dinâmica é realizada pela extração das 5 meta-características, que são apresentadas ao meta-classificador. Este tem a incumbência de estimar quando um classificador base é competente suficiente para classificar a instância  $\vec{x}$ .

### 3.1.14 Seleção Baseada em Complexidade do Problema

*Brun et al.* construíram um sistema de múltiplos classificadores em (BRUN et al., 2016), denominado *Dynamic Selection based on Complexity* (DSOC), que adota informações de dificuldade do problema de classificação para orientar tanto a geração dos componentes do *ensemble* como a posterior etapa de seleção dos mesmos. A dificuldade do problema é descrita no trabalho por meio de meta-características, obtidas a partir de dados do problema usando medidas de complexidade. Na etapa de geração dos classificadores foi empregado um AE cujo objetivo foi maximizar a exploração do espaço de complexidade e ao mesmo tempo formar indutores precisos. Na etapa de seleção foram combinados três critérios: a acurácia local de cada classificador, a similaridade de sua assinatura de complexidade e a distância da instância de teste até o centroide da classe predita.

## 3.2 Considerações Finais

Em aplicações reais de reconhecimento de padrões, geralmente há dificuldades em se encontrar classificadores que satisfaçam o pressuposto dos erros independentes. Uma forma encontrada para evitar a premissa das falhas independentes é a seleção dinâmica de classificadores. Esta baseia-se no antecedente de que cada componente do *ensemble* é um especialista em alguma região do espaço de instâncias, o que permite que, dentre um conjunto de classificadores, haja pelo menos um que consiga rotular corretamente o padrão em questão. Portanto, a seleção dinâmica de classificadores representa um grande potencial para alcançar melhores taxas de reconhecimento, uma vez que a mesma tenta encontrar o(s) componente(s) do *ensemble* mais adequado(s) para cada instância a ser classificada. Nos últimos anos, *ensembles* dinâmicos de classificadores têm sido aplicados com sucesso para o reconhecimento de face (BASHBAGHI et al., 2017), classificação de gênero musical (ALMEIDA et al., 2012), análise de crédito (LESSMANN et al., 2015; XIAO; XIAO; WANG, 2016), problemas com classes desbalanceadas (GALAR et al., 2012), sistemas de recomendação (PORCEL et al., 2012), predição de erro em *software* (PANICHELLA; OLIVETO; LUCIA, 2014; NUCCI et al., 2017), e para lidar com ambientes em mudança (KRAWCZYK et al., 2017). No entanto, ainda não foram demonstrados métodos que relatassem performance próximas ao Oráculo (SOUZA et al., 2017). Assim sendo, tem-se um campo aberto para o desenvolvimento de novas pesquisas, no sentido de propor estratégias que sejam capazes de selecionar e combinar múltiplos classificadores com maiores chances de rotular corretamente o padrão em questão.

## 4 COMBINAÇÃO DOS COMPONENTES

Uma vez que os classificadores tenham sido selecionados, deve-se realizar uma decisão final com base em cada uma das hipóteses produzidas. As possíveis maneiras de se combinar (fundir) respostas dos componentes em *ensembles* dependem de qual informação é obtida. Algumas regras operam somente com informações da classe de saída, enquanto outras necessitam dos valores contínuos que são interpretados como sendo o grau de apoio dado pelo classificador para cada uma das possíveis classes. Dado que existem inúmeras revisões na literatura sobre a combinação de classificadores (KUNCHEVA, 2004a; PONTI, 2011), nas próximas subseções serão apresentadas apenas algumas das principais formas de realizá-la. Baseado nas definições anteriores, as descrições neste capítulo serão feitas com as seguintes considerações: trata-se da combinação de  $L$  classificadores,  $h_1, h_2, \dots, h_L$ , utilizados para classificação do padrão  $\vec{x}$  em  $O$  classes,  $\omega_1, \dots, \omega_O$ . A saída de  $h_i$  para a  $j$ -ésima classe será dada por  $\mu_i^j$  e a saída final do *ensemble* será denotada por  $e(\cdot)$ .

### 4.1 Padrões de Arquitetura

De acordo com a arquitetura (topologia) interna, a combinação de múltiplos classificadores, conforme apresentado em (PONTI, 2011), pode ser considerada: paralela, serial (em cascata) ou híbrida. As subseções a seguir apresentam mais detalhes.

#### 4.1.1 Topologia Paralela

Na primeira abordagem, ilustrada na Figura 8, todos os componentes do *ensemble* realizam a mesma tarefa, obtendo ao final de sua execução uma hipótese para cada padrão de consulta. As hipóteses então devem ser combinadas, segundo alguma regra, para decidir qual é a classe mais apropriada para o padrão em questão.

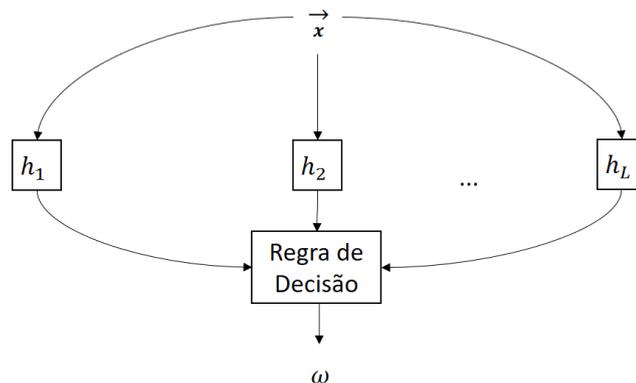


Figura 8 – Topologia paralela na combinação de múltiplos classificadores

### 4.1.2 Topologia Serial

Na segunda abordagem, apresentada na Figura 9, os componentes do *ensemble* são distribuídos em ordem crescente de complexidade, onde a instância de consulta  $\vec{x}$  é submetida inicialmente ao classificador mais simples. Caso haja uma rejeição, com base em algum critério previamente estabelecido, a instância  $\vec{x}$  é então submetida ao classificador seguinte. Este processo iterativo é realizado até que não ocorram mais rejeições ou não existam mais classificadores na sequência. O objetivo da ordenação é utilizar os classificadores mais complexos, que geralmente são computacionalmente mais caros, somente em situações em que o seu uso for estritamente necessário.

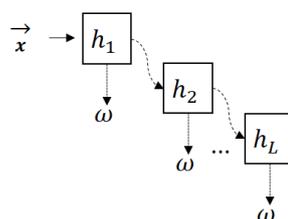


Figura 9 – Topologia serial na combinação de múltiplos classificadores<sup>1</sup>

### 4.1.3 Topologia Híbrida

Existe a possibilidade, em algumas aplicações mais específicas, de combinar/hibridizar características existentes nas topologias paralela e serial. Isso é feito basicamente com o intuito de melhorar a performance do sistema, já que o alto custo de alguns classificadores pode impossibilitar o uso do *ensemble*. A figura 10 esboça uma ideia geral dessa topologia.

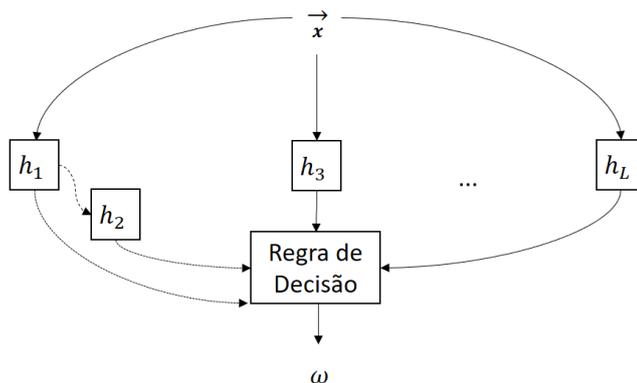


Figura 10 – Topologia híbrida na combinação de múltiplos classificadores

<sup>1</sup> As linhas que estão pontilhadas representam o fluxo de execução quando não existe a rejeição. As linhas que estão tracejadas representam o fluxo de execução quando existe a rejeição.

## 4.2 Regras de Decisão

Os classificadores de um *ensemble* podem produzir saídas do tipo *hard* ou *soft*. O primeiro tipo fornece um valor único para cada uma das amostras, condizente com o identificador da classe. No segundo tipo, tem-se o grau de apoio dado para cada classe, produzindo assim mais informações para o pós-processamento. Nas subseções a seguir, serão apresentadas algumas das principais regras de decisão para cada um desses tipos.

### 4.2.1 Regras de Decisão do Tipo *Hard*

Uma das estratégias mais intuitivas na combinação de múltiplos classificadores é a utilização do voto. No caso da votação majoritária simples, a título de exemplo, cada componente do *ensemble* tem direito a votar em apenas uma única classe, sendo escolhida aquela que obtiver o maior número de votos. De forma generalizada, seja  $\vec{\omega}_i(\vec{x}) \in \mathbb{R}^O$  o vetor de saída do componente  $h_i$  para uma dada entrada  $\vec{x}$ , em que  $O$  é o número de classes e o vetor  $\vec{\omega}_i(\vec{x})$  é binário e possui somente uma componente não nula, em geral, de valor 1. Seja ainda  $\vec{\mu}(\vec{x}) \in \mathbb{R}^O$  um vetor de votos em que cada elemento  $\mu_j(\vec{x}), j = 1, \dots, O$  possui um valor que é proporcional à frequência de resposta da  $j$ -ésima classe entre os classificadores. O vetor  $\vec{\mu}(\vec{x})$  pode então ser calculado conforme mostra a equação 4.1.

$$\vec{\mu}(\vec{x}) = \sum_{i=1}^L \omega_i(\vec{x}) = [\mu_1(\vec{x}), \dots, \mu_O(\vec{x})]^T \quad (4.1)$$

Por fim, a classe inferida pelo *ensemble* para o padrão  $\vec{x}$  é dada pela equação 4.2.

$$e(\vec{x}) = \arg \max_{j=1, \dots, O} \vec{\mu}_j(\vec{x}) \quad (4.2)$$

As regras do tipo *hard* ainda podem ser aplicadas de forma ponderada, ou seja, cada classificador recebe, antes da votação, um peso  $w$  que é proporcional ao seu nível de competência, geralmente definido pelo desempenho num dado conjunto de treinamento/validação. Nesse caso, o vetor de votos que foi apresentado na equação 4.1 passa a ser calculado pela equação 4.3, em que  $w_i$  representa o peso atribuído ao  $i$ -ésimo classificador do *ensemble*.

$$\vec{\mu}(\vec{x}) = \sum_{i=1}^L w_i \times \omega_i(\vec{x}) \quad (4.3)$$

### 4.2.2 Regras de Decisão do Tipo *Soft*

Saídas do tipo *soft* podem ter valores *fuzzy*, probabilidades *a posteriori*, graus de certeza ou confiança, dentre tantos outros. Assim, existem  $O$  valores associados à  $O$  classes, a partir dos quais, regras podem ser aplicadas para a realização de uma combinação mais efetiva. Para tanto constrói-se inicialmente o chamado *Decision Profile* (DP), conforme

exemplo ilustrado na equação 4.4. Matematicamente, o DP para uma instância  $\vec{x}$  pode ser definido como uma matriz  $L \times O : DP(\vec{x}) = [h_1(\vec{x}), h_2(\vec{x}), \dots, h_L(\vec{x})]^T$ , em que  $L$  é o número de componentes do *ensemble* e  $O$  representa o número de classes. Cada  $i$ -ésima saída *soft* é descrita por um vetor  $h_i(\vec{x}) = [\mu_i^1, \dots, \mu_i^O]$  na  $i$ -ésima linha do DP, em que  $\mu_i^j, j = 1, \dots, O$  traduz o grau de apoio dado pelo classificador  $h_i$  para a classe  $\omega_j$ . Consequentemente, em cada  $j$ -ésima coluna, tem-se a confiança dos  $L$  classificadores para a classe  $\omega_j$ .

$$DP(\vec{x}) = \begin{bmatrix} \mu_1^1 & \dots & \mu_1^j & \dots & \mu_1^O \\ \vdots & & \vdots & & \vdots \\ \mu_i^1 & \dots & \mu_i^j & \dots & \mu_i^O \\ \vdots & & \vdots & & \vdots \\ \mu_L^1 & \dots & \mu_L^j & \dots & \mu_L^O \end{bmatrix} \quad (4.4)$$

Métodos de combinação que usam uma única coluna do DP por vez, são conhecidos como *Class-Conscious Combiners* (KUNCHEVA, 2004a) ou *Consensus Theoretic Methods* (BENEDIKTSSON; SWAIN, 1992). Abaixo são listadas algumas dessas regras de fusão:

- **Média:**  $\mu_j(\vec{x}) = \frac{1}{L} \sum_{i=1}^L \{\mu_i^j\}$
- **Máximo:**  $\mu_j(\vec{x}) = \max_i \{\mu_i^j\}$
- **Mínimo:**  $\mu_j(\vec{x}) = \min_i \{\mu_i^j\}$
- **Produto:**  $\mu_j(\vec{x}) = \prod_{i=1}^L \{\mu_i^j\}$

em que, dado um padrão de consulta  $\vec{x}$ ,  $\mu_j(\vec{x})$  representa o grau de apoio dado para a classe  $\omega_j$ . Após as devidas fusões, uma atribuição do tipo *hard* pode ser obtida. Esse procedimento é feito a partir da identificação do maior valor de confiança  $\mu_j(\vec{x})$ , cujo índice  $j$  representará o rótulo a ser atribuído ao padrão  $\vec{x}$ , ou seja, a classe  $\omega_j$ . A figura 11 apresenta um exemplo didático de como essas regras de fusão funcionam na prática. No exemplo existem três classificadores que rotulam um padrão em três possíveis classes. Os valores que estão destacados em negrito representam o maior grau de apoio alcançado pela regra de fusão empregada, portanto indicam a qual classe  $\vec{x}$  pertence.

### 4.3 Considerações Finais

Neste capítulo foram apresentados conceitos básicos referentes à combinação de múltiplos classificadores. Além do detalhamento das possíveis arquiteturas de *ensembles*, foram demonstradas as principais regras de decisão. Essas regras são chamadas não-treináveis pelo motivo de não necessitarem a estimação de nenhum parâmetro, como

	$\omega_1$	$\omega_2$	$\omega_3$	
$h_1$	0.0	0.3	0.7	
$h_2$	0.0	0.4	0.6	
$h_3$	0.2	0.6	0.2	
Regra				Resultado
Média	0.067	0.44	<b>0.5</b>	$\omega_3$
Máximo	0.2	0.6	<b>0.7</b>	$\omega_3$
Mínimo	0.0	<b>0.3</b>	0.2	$\omega_2$
Produto	0.0	0.072	<b>0.084</b>	$\omega_3$

Figura 11 – Exemplo dos métodos de combinação: média, máximo, mínimo e produto

por exemplo, a competência do componente no *ensemble*. Alguns autores investigam formas de combinação não lineares, utilizando para tanto um método de aprendizagem adicional, denominado de meta-aprendizagem, tais como, por exemplo, os algoritmos de cascadeamento (GAMA; BRAZDIL, 2000) e árvores de meta-decisão (TODOROVSKI; DŽEROSKI, 2003). O próximo capítulo trata da proposta desta tese, que é inspirada na hibridização de algumas das técnicas de geração, seleção e combinação vistas até o momento. Em síntese, o método proposto é capaz de construir automaticamente um conjunto de classificadores, através de um processo evolutivo, além de selecionar e combinar dinamicamente diferentes componentes do *ensembles* para rotular cada padrão em questão.

## 5 MÉTODO PROPOSTO

O presente capítulo visa a construção de um Sistema Híbrido Inteligente (SHI), com a intenção de orientar tanto a geração do *ensemble* quanto a seleção/combinção dos seus componentes. Os resultados experimentais, apresentados no capítulo 6, sugerem que o SHI proposto é capaz de aprimorar as taxas de correta classificação, especialmente quando comparado com alguns dos principais métodos existentes na literatura.

### 5.1 Sistemas Híbridos Inteligentes

Os dois principais objetivos do uso de SHIs são: 1) a resolução de problemas complexos que só poderiam ser resolvidos se tratados com subtarefas; e 2) unir as vantagens e superar as limitações individuais de cada técnica. Desta forma, a proposta dos SHIs é a construção de sistemas mais robustos, capazes de trabalhar com vários tipos de representação de conhecimento, dar suporte a vários tipos de inferências, resolver problemas mais complexos, dentre outras vantagens. Em (GOONATILAKE; KHEBBAL, 1994), foi proposto um esquema de classificação para SHIs composto das seguintes categorias:

Tipo I *Function Replacing*. Nessa categoria, uma determinada técnica é utilizada para implementar a função de uma outra técnica. Essa forma de hibridismo não acrescenta nenhuma nova funcionalidade ao sistema inteligente, apenas tenta superar algumas das limitações da técnica principal ou otimizar a sua execução.

Tipo II *Intercommunicating Hybrids*. Essa é a categoria de SHIs usados para resolução de problemas complexos e que possam ser divididos em várias subtarefas independentes. Assim, o sistema é formado por módulos independentes, onde cada um usa uma técnica inteligente para resolver uma das subtarefas do problema principal.

Tipo III *Polymorphic Hybrids*. Nessa categoria, uma única técnica é adaptada para realizar uma tarefa inerente a outra técnica. A motivação é descobrir novas funcionalidades de uma técnica e entender como diferentes técnicas podem se relacionar.

### 5.2 Hibridizações Propostas

O sistema proposto pode ser dividido em três fases: i) Geração dos Componentes; ii) Geração dos Seletores; e iii) Generalização. A figura 12 apresenta uma visão geral de como o mesmo está estruturado. As hibridizações presentes/propostas em cada uma dessas três fases serão discutidas com maiores detalhes nas seguintes subseções.

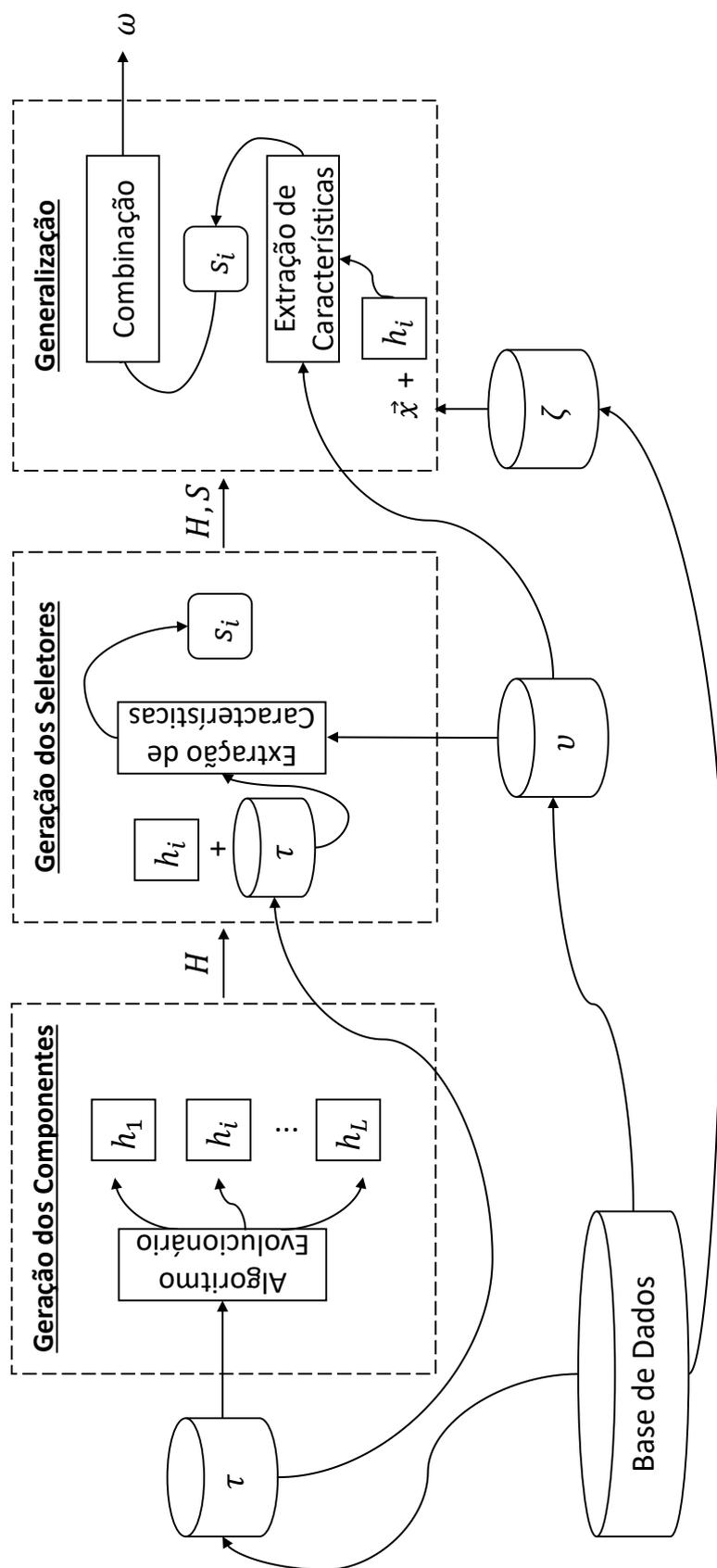


Figura 12 – Estrutura macro do método desenvolvido

Resumidamente, o sistema funciona da seguinte forma: gera-se, na primeira fase, um *ensemble*  $H = \{h_1, h_2, \dots, h_L\}$  com  $L$  componentes. A ideia é evoluir cada  $h_i \in H$  de forma que seja possível realizar diferentes mapeamentos entrada-saída, ou seja, que se alcance um maior espalhamento no espaço de hipóteses  $\mathcal{H}$ . Em seguida, tem-se o módulo da geração dos seletores. Com a conclusão dessa fase, tem-se como resultado um conjunto  $S = \{s_1, s_2, \dots, s_L\}$  de  $L$  seletores. Por fim, na fase de generalização, cada seletor  $s_i \in S$  será responsável por calcular a competência do classificador  $h_i \in H$  rotular corretamente o padrão de consulta  $\vec{x}$ . A fusão/combinção dos componentes selecionados é feita de maneira ponderada, com base na competência alcançada por cada classificador.

### 5.2.1 Geração dos Componentes

Esta fase tem como objetivo a formação de *ensembles* na qual seus componentes sejam acurazes e diversos em termos de opinião, permitindo que o processo de seleção dinâmica de classificadores possa identificar quais deles sejam considerados competentes para rotular cada padrão de consulta. Dessa forma, um algoritmo foi proposto visando a evolução/otimização de um conjunto de classificadores  $H = \{h_1, h_2, \dots, h_L\}$ . A escolha por tal estratégia baseia-se em resultados de publicações do autor, apresentadas na seção 1.3 do capítulo 1, bem como em outros trabalhos (CHANDRA; YAO, 2004; SANTOS; SABOURIN; MAUPIN, 2008; KO; SABOURIN; BRITTO, 2009; NETO; CANUTO; LUDERMIR, 2013; GU; JIN, 2014) onde a otimização explícita de alguma medida de diversidade permitiu construir *ensembles* com maior capacidade de generalização.

O método proposto para geração dos componentes é apresentado no Algoritmo 1. Cada  $h_i \in H$  é uma árvore de decisão extremamente simples, possuindo um nó interno (a raiz) que está imediatamente ligado a dois nós terminais (folhas). Não existe etapa de treinamento dos componentes, pois estes são construídos de forma completamente aleatória. Apesar disso, a evolução ocorre graças ao uso da função objetivo ou *fitness*, apresentada na equação 5.1. Vale ressaltar que a equação 5.2, necessária para o somatório da equação 5.1, mede o grau da diversidade entre o par de classificadores  $h_i$  e  $h_j$ . Os significados de  $N^{00}$ ,  $N^{01}$ ,  $N^{10}$  e  $N^{11}$ , encontrados na equação 5.2, foram apresentados na subseção 3.1.7 do capítulo 3. Como  $N^{10}$  diz respeito ao número de padrões que apenas  $h_i$  classifica corretamente, a função objetivo é responsável por guiar o processo de forma que cada componente do *ensemble* seja um especialista numa dada região do conjunto  $\tau$ . Esse fato é essencial em sistemas de seleção dinâmica de múltiplos classificadores.

$$fitness(h_i, \tau) = \frac{\sum_{j=1}^L fitness(h_i, h_j, \tau)}{L} \quad (5.1)$$

$$fitness(h_i, h_j, \tau) = \frac{N^{10}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (5.2)$$

**Algoritmo 1:** GERAÇÃO DOS COMPONENTES**Data:** conjunto  $\tau$ , número de componentes  $L$ , número de gerações  $G$ 

```

begin
  for  $i = 1$  até  $L$  do
     $h_i$  = árvore gerada aleatoriamente;
    Inserir  $h_i$  no ensemble  $H$ ;
  end
  for  $g = 1$  até  $G$  do
    for  $i = 1$  até  $L$  do
       $h_i^*$  = árvore gerada aleatoriamente;
      if  $fitness(h_i^*, \tau) > fitness(h_i, \tau)$  then
         $h_i = h_i^*$ ;
      else
         $h_i = h_i$ ;
      end
    end
  end
end
return  $H$ ;

```

## 5.2.2 Geração dos Seletores

Emprega-se, na segunda fase do método proposto, o uso de casos/situações semelhantes para auxiliar o processo da seleção dinâmica de classificadores. Ao contrário dos sistemas que são baseados em regras gerais, nos métodos baseados em casos, o conhecimento é armazenado em uma base de dados contendo várias situações possíveis do problema. Portanto, um novo caso é resolvido simplesmente a partir de soluções usadas em situações similares no passado. No contexto deste trabalho, por exemplo, cada caso armazena um conjunto de características que descrevem as situações em que os componentes do *ensemble* são considerados competentes ou incompetentes para classificar corretamente um dado padrão de consulta. Sendo assim, diante da etapa de generalização, uma pesquisa é feita na base de casos, recuperando algumas das soluções mais similares para auxiliar na resolução do novo padrão de consulta. A escolha dessa estratégia se inspira no último trabalho completo, publicado pelo autor da tese, em anais de congresso, apresentado na subseção 1.3.2 do capítulo 1, bem como nos trabalhos (WOLPERT, 1992; TODOROVSKI; DŽEROSKI, 2003; CRUZ et al., 2015). O que há em comum entre essas quatro publicações é o fato de que todas elas empregam alguma estratégia baseada em conhecimento na tentativa de melhorar as taxas de sucesso na combinação/seleção dos componentes, e, por consequência, melhorar também a performance de correta classificação do *ensemble*.

Dado que o componente  $h_i \in H$  rotule uma determinada instância  $\vec{x}$  como sendo da classe  $\omega_j$ . A proposta dessa fase consiste em preparar o sistema, no sentido de que seja possível realizar uma estimativa mais robusta de quando  $h_i$  é *competente* ou *incompetente* para a classificação de  $\vec{x}$ . No intuito de construir os elementos que serão responsáveis pela estimação dessa competência, isto é, o conjunto de seletores  $S = \{s_1, s_2, \dots, s_L\}$ , considera-se o conjunto  $\nu$  para o cálculo das regiões de competência. A partir de então, para a formação da base de casos do seletor  $s_i$ , são extraídas características de  $h_i$  em cada instância  $\vec{x} \in \tau$ . Por se tratar de um processo de aprendizado supervisionado, além do vetor de características, um rótulo indica se o classificador  $h_i$  conseguiu ou não rotular corretamente  $\vec{x}$ . A fase de geração dos seletores é apresentada de maneira resumida no Algoritmo 2 e o processo de extração das características encontra-se na subseção 5.2.3.

---

**Algoritmo 2: GERAÇÃO DOS SELETORES**


---

**Data:** conjunto  $\tau$ , conjunto  $\nu$ , ensemble  $H$

**begin**

    L = tamanho de  $H$

**for**  $i = 1$  até  $L$  **do**

**for** cada  $\vec{x} \in \tau$  **do**

$\vec{v} =$  Extração das Características( $\nu, h_i, H, \vec{x}$ );

**if**  $h_i$  classificou corretamente  $\vec{x}$  **then**

                Inserir na base de casos de  $s_i$   $\{\vec{v}, \textit{competente}\}$ ;

**else**

                Inserir na base de casos de  $s_i$   $\{\vec{v}, \textit{incompetente}\}$ ;

**end**

**end**

        Inserir  $s_i$  em  $S$ ;

**end**

**end**

**return**  $S$ ;

---

### 5.2.3 Extração das Características

Foram consideradas 5 características na construção da base de casos dos seletores, conforme mostra o Algoritmo 3. Cada uma delas corresponde a um diferente critério ou hipótese para estimar a competência dos componentes do *ensemble* na correta classificação de um dado padrão de consulta  $\vec{x}$ . A principal vantagem em usar múltiplas hipóteses, conforme demonstrado no início do capítulo 1, é o fato de mesmo que algumas delas estejam incorretas, o sistema ainda pode alcançar sucesso na classificação do padrão  $\vec{x}$ , uma vez que a maioria das hipóteses pode estar correta. Por outro lado, o simples fato do vetor de características  $\vec{v}$  ser extraído, em tempo de execução, representa um aumento considerável no custo computacional. Isso foi avaliado na seção 6.3 do capítulo 6.

**Algoritmo 3:** EXTRAÇÃO DAS CARACTERÍSTICAS

---

**Data:** conjunto  $\nu$ , classificador  $h$ , *ensemble*  $H$ , padrão  $\vec{x}$

**begin**

$\omega_j$  = classe que  $h$  atribui a  $\vec{x}$ ;

$R$  = região de competência de  $\vec{x}$  em  $\nu$ ;

$f_1$  = valor do atributo raiz de  $h$  considerando  $\vec{x}$ ;

$f_2$  = desacordo de  $h$  em  $H$  considerando  $\vec{x}$ ;

$f_3$  = *ranking* de  $h$  em  $R$ ;

$f_4$  = especificidade de  $h$  em  $R$ , considerando  $\omega_j$ ;

$f_5$  = sensibilidade de  $h$  em  $R$ , considerando  $\omega_j$ ;

$\vec{v} = \{f_1, f_2, f_3, f_4, f_5\}$

**end**

**return**  $\vec{v}$ ;

---

$R$  o tamanho da região de competência foi fixo em 7. Este valor se mostrou mais adequado em experimentos preliminares, bem como em estudos realizados anteriormente (KO; SABOURIN; BRITTO, 2008; CRUZ; CAVALCANTI; REN, 2011);

$f_1$  foi empregada com inspiração nos métodos de clusterização e seleção, como por exemplo (KUNCHEVA, 2000), que utilizam informações do espaço de instâncias para realizar o mapeamento que determina quando cada componente do *ensemble* deve ser utilizado para rotular um dado padrão de consulta;

$f_2$  é calculada pelo percentual de componentes do *ensemble* que discordam da decisão do classificador  $h$  na predição de  $\vec{x}$ . Essa característica foi empregada com inspiração nos métodos de seleção dinâmica que utilizam informações do conjunto de classificadores, como por exemplo (SANTOS; SABOURIN; MAUPIN, 2007);

$f_3$  representa o percentual de padrões consecutivos, no sentido do mais próximo para o mais distante de  $\vec{x}$ , que foi corretamente classificado na região de competência. Essa característica foi empregada com base nos métodos de seleção dinâmica que empregam o *ranking*, como por exemplo (SABOURIN et al., 1993);

$f_4$  é a proporção de verdadeiros negativos considerando a classe  $\omega_j$ , ou seja, a capacidade em prever corretamente a ausência da condição para casos que não a têm;

$f_5$  é a proporção de verdadeiros positivos considerando a classe  $\omega_j$ , ou seja, a capacidade em prever corretamente a condição para casos que realmente a têm;

## 5.2.4 Generalização

A última fase do método proposto é o processo propriamente dito da seleção dinâmica de classificadores. Nela, para rotular uma dada instância  $\vec{x} \in \zeta$ , é realizada a estimação da competência de cada classificador, baseada nos casos que foram armazenados no seu respectivo seletor. Portanto, para cada  $h_i \in H$ , o processo de extração das características, apresentado no Algoritmo 3, retorna um vetor  $\vec{v}$  que é usado como entrada do seletor  $s_i \in S$ . O classificador  $h_i$  participará do processo de combinação (fusão) se menos da metade dos  $k$  casos mais semelhantes<sup>1</sup> à  $\vec{v}$ , recuperados em  $s_i$ , forem do tipo *incompetente*. Essa estratégia de combinação foi adotada, uma vez que, conforme apresentado na Figura 1 do capítulo 1, para que haja aumento na probabilidade de sucesso num processo de combinação envolvendo múltiplas hipóteses, é preciso que a probabilidade do erro de cada uma delas seja menor do que 50%. Além do mais, para que se possa dar mais ênfase aos classificadores que alcançarem maior grau de competência, os votos são ponderados pelo número de casos que foram recuperados como sendo do tipo *competente*. O processo de generalização é apresentado de maneira resumida no Algoritmo 4.

---

### Algoritmo 4: GENERALIZAÇÃO

---

**Data:** conjunto  $\nu$ , *ensemble*  $H$ , número de classes  $O$ ,  
número de casos  $k$ , padrão  $\vec{x}$ , seletores  $S$

**begin**

votos = vetor de zeros com tamanho  $O$ ;

$L$  = tamanho de  $H$ ;

**for**  $i = 1$  até  $L$  **do**

$\vec{v}$  = Extração das Características( $\nu, h_i, H, \vec{x}$ );

casos =  $k$  ocorrências mais similares de  $\vec{v}$  em  $s_i$ ;

competência = número de casos do tipo competentes;

**if** *competência*  $> k/2$  **then**

$j$  = índice da classe que  $h_i$  atribui a  $\vec{x}$ ;

votos[ $j$ ] = votos[ $j$ ] + competência/ $k$ ;

**else**

$h_i$  não participa da votação

**end**

**end**

**end**

$j$  = índice do maior valor no vetor de votos;

**return**  $\omega_j$ ;

---

<sup>1</sup> A distância euclidiana foi empregada para o cálculo da semelhança

### 5.3 Considerações Finais

Nesse capítulo detalhou-se a estrutura do método, bem como o funcionamento de todas as etapas envolvidas no processo. Foi proposto um algoritmo capaz de evoluir os componentes do *ensemble*, de tal forma que o processo de seleção dinâmica possa identificar quais deles são considerados competentes para rotular cada padrão de consulta. Semelhantemente aos trabalhos árvores de meta-decisão (TODOROVSKI; DŽEROSKI, 2003) e Meta-DES (CRUZ et al., 2015), foi empregada uma estratégia de aprendizagem, cujo objetivo é a realização da seleção dinâmica de classificadores. A novidade aqui está no fato de usar um diferente conjunto de características e um seletor específico para cada classificador. No Meta-DES, por exemplo, as características são: dificuldade na classificação dos vizinhos; probabilidade posteriori; acurácia total local; perfis de saída dos componentes; e a confiança dos componentes. Já as árvores de meta-decisão usam atributos derivados da distribuição de probabilidade das classes preditas pelos classificadores. O fato do método usar um seletor específico para cada classificador possibilita a minimização de ambiguidades numa base única de casos, pois a extração de características de dois classificadores distintos pode apresentar os mesmos valores do vetor  $\vec{v}$ , mas com rótulos diferentes, ou seja, um sendo considerado competente e outro incompetente. Contudo, o uso de um meta-classificador único já demonstrou ser apropriado (CRUZ et al., 2015).

## 6 EXPERIMENTOS E RESULTADOS

Neste capítulo são apresentados os resultados alcançados com o intuito de validar o método proposto. Foi adotado um conjunto composto de 30 problemas amplamente conhecidos, obtidos nos repositórios: KEEL (ALCALÁ-FDEZ et al., 2011), LKC (KUNCHEVA, 2004b), STATLOG (KING; FENG; SUTHERLAND, 1995) e UCI (BACHE; LICHMAN, 2013). A tabela 1 apresenta as principais características de cada problema. São detalhados o número de instâncias, tamanho dos conjuntos de treinamento  $\tau$ , validação  $\nu$  e generalização  $\zeta$ , bem como a quantidade de atributos, o número de classes, balanceamento<sup>1</sup> e origem. Todos os ensaios foram conduzidos empregando-se 30 execuções. Em cada uma, os conjuntos  $\tau$ ,  $\nu$  e  $\zeta$  foram divididos aleatoriamente com estratificação. Todos os atributos numéricos foram normalizados entre 0 e 1 e os nominais foram convertidos para binários.

Tabela 1 – Problemas usados nos experimentos

Problema	Instâncias	$\tau$	$\nu$	$\zeta$	Atributos	Classes	Balanceamento	Origem
Abalone	4177	2089	1044	1044	8	3	1.17	UCI
Australian	690	345	173	172	17	2	1.25	STATLOG
Balance	625	313	156	156	4	3	5.88	UCI
Blood	748	374	187	187	4	2	3.20	UCI
Bupa	345	173	86	86	6	2	1.38	UCI
Cancer	699	350	175	174	9	2	1.90	UCI
Chess	3196	1598	799	799	36	2	1.09	UCI
CMC	1473	737	368	368	9	3	1.89	UCI
Contractions	98	49	25	24	27	2	1.00	LKC
Flare	1389	695	347	347	10	6	7.76	UCI
German	1000	500	250	250	20	2	2.33	STATLOG
Glass	214	107	54	53	9	6	8.44	UCI
Haberman	306	153	77	76	3	2	2.78	UCI
Heart	270	135	68	67	13	2	1.25	STATLOG
Hepatitis	155	78	39	38	19	2	3.84	UCI
Horse	368	184	92	92	22	2	1.71	UCI
ILPD	583	292	146	145	10	2	2.49	UCI
Ionosphere	351	176	88	87	34	2	1.79	UCI
Iris	150	75	38	37	4	3	1.00	UCI
Laryngeal 1	213	107	53	53	16	2	1.63	LKC
Laryngeal 3	353	177	88	88	16	3	4.11	LKC
Lymph	148	74	37	37	18	4	40.50	UCI
Mammo	961	481	240	240	5	2	1.16	KEEL
Phoneme	5404	2702	1351	1351	5	2	2.41	KEEL
Segment	2310	1155	578	577	19	7	1.00	UCI
Sonar	208	104	52	52	60	2	1.14	UCI
Thyroid	215	108	54	53	5	3	5.00	LKC
Waveforms	5000	2500	1250	1250	40	3	1.02	UCI
Weaning	302	151	76	75	17	2	1.00	LKC
Wine	178	89	45	44	13	3	1.48	UCI

<sup>1</sup> O balanceamento é calculado pela razão entre o número de instâncias da classe majoritária e minoritária

## 6.1 Avaliação do Número de Componentes

Esta seção apresenta resultados que visam avaliar o número de componentes. Para este propósito, variou-se o tamanho  $L$  do *ensemble* de 5 à 100 com intervalos de 5 (20 ensaios foram realizados ao todo para cada um dos problemas da tabela 1). Os valores das médias (com desvios padrões entre parênteses) são apresentados nas tabelas 2 e 3. Todos os *ensembles* dessa seção foram construídos por meio do Algoritmo 1, utilizando o número de gerações  $G = 100$ . Este valor foi definido após vários ensaios preliminares, onde verificou-se que a partir dele praticamente inexistente evolução dos componentes. Visando comparar os valores de  $L$ , foram realizados testes estatísticos com 95% de confiança. O teste de *Friedman* (com 19 graus de liberdade, uma vez que foram comparados resultados de 20 diferentes tamanhos de componentes) indicou que existe diferença significativa entre as taxas de amostragem. Então, efetuou-se o pós-teste de *Nemenyi* para esboçar os ranqueamentos. O resultado desse ranqueamento é apresentado na figura 13.

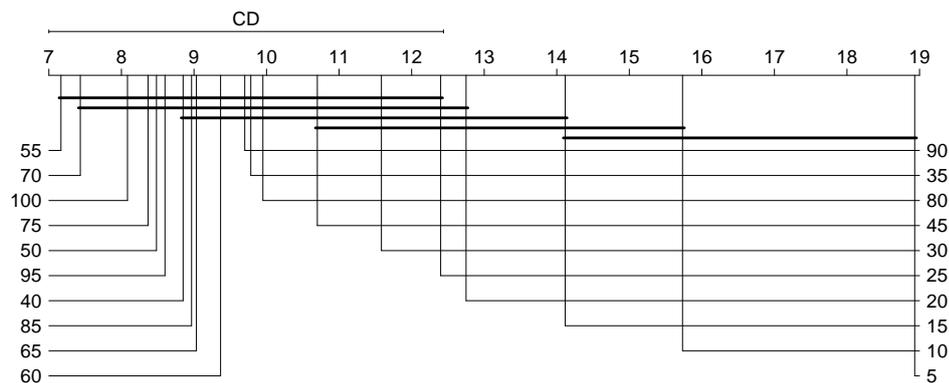


Figura 13 – Representação do teste de *Nemenyi* comparando o número de componentes

A representação gráfica do teste *Nemenyi* apresenta uma escala contendo ranqueamentos, de maneira que os números de componentes analisados são posicionados de acordo com os valores médios de suas posições no ranqueamento calculado no teste *Friedman*. Linhas horizontais conectam as abordagens cujos resultados não tiveram diferença estatística de acordo com o pós-teste. Além disso, o valor crítico, usado para verificar se duas abordagens são diferentes<sup>2</sup>, é posto acima da linha de ranqueamento. Fica evidente que não é vantajoso usar mais de 25 componentes, pois a partir deste valor os *ensembles* apresentam acuracidade estatisticamente equivalentes. Ou seja, um maior número de componentes apenas aumenta o custo computacional. Desta forma, nos experimentos das próximas seções, serão utilizados os *ensembles* gerados aqui com número de componentes  $L = 25$ .

<sup>2</sup> O desempenho é diferente quando a média do rank difere em pelo menos a diferença crítica (CD)

Tabela 2 – Médias da acuracidade no conjunto  $\tau$  considerando *ensembles* gerados pelo Algoritmo 1 com  $L$  variando de 5 à 50

Problema	L=5	L=10	L=15	L=20	L=25	L=30	L=35	L=40	L=45	L=50
Abalone	57,28 (1,71)	58,02 (1,41)	58,00 (1,18)	58,09 (1,08)	58,06 (1,49)	57,81 (1,12)	57,83 (0,82)	57,28 (1,29)	57,21 (1,14)	56,88 (1,02)
Australian	87,77 (1,23)	87,32 (1,36)	87,53 (1,33)	87,56 (1,34)	87,56 (1,28)	87,21 (1,57)	87,43 (1,49)	87,51 (1,39)	87,34 (1,49)	87,30 (1,47)
Balance	87,42 (1,22)	88,39 (1,00)	88,31 (0,88)	88,34 (0,85)	88,36 (0,81)	88,40 (0,93)	88,38 (1,00)	88,43 (0,84)	88,41 (0,95)	88,41 (0,98)
Blood	78,87 (1,96)	79,96 (1,89)	80,11 (1,98)	79,93 (1,84)	79,50 (1,93)	79,71 (1,70)	79,80 (1,87)	79,71 (2,02)	79,89 (1,71)	79,66 (2,42)
Bupa	69,29 (3,57)	73,64 (3,37)	75,38 (3,85)	76,38 (2,81)	77,03 (3,17)	76,99 (3,19)	77,48 (2,70)	77,34 (2,85)	77,71 (2,86)	78,25 (3,30)
Cancer	96,92 (0,86)	97,02 (0,69)	97,00 (0,74)	97,00 (0,76)	97,00 (0,80)	97,03 (0,78)	97,05 (0,76)	97,02 (0,78)	97,00 (0,75)	97,03 (0,78)
Chess	94,23 (1,15)	95,24 (0,97)	95,52 (0,67)	95,47 (0,80)	95,66 (0,67)	95,73 (0,56)	95,78 (0,60)	95,86 (0,54)	95,68 (0,49)	95,71 (0,55)
CMC	50,05 (3,92)	54,49 (2,65)	56,51 (2,04)	58,14 (1,65)	58,50 (1,73)	58,89 (1,67)	58,99 (1,71)	58,83 (1,66)	58,62 (1,86)	59,04 (1,87)
Contractions	89,86 (4,06)	90,48 (2,75)	90,95 (3,33)	91,29 (3,12)	90,75 (3,74)	91,63 (2,95)	91,70 (3,47)	91,09 (3,19)	91,56 (3,29)	91,77 (2,96)
Flare	70,34 (3,05)	72,12 (2,06)	72,42 (2,27)	72,74 (2,14)	72,87 (2,14)	72,73 (2,06)	72,81 (2,30)	73,29 (2,21)	73,05 (1,94)	73,30 (2,17)
German	74,25 (1,76)	77,78 (1,09)	78,31 (1,06)	78,19 (1,25)	78,26 (1,26)	78,03 (1,33)	78,07 (1,48)	78,07 (1,25)	77,87 (1,19)	78,01 (1,30)
Glass	62,74 (3,66)	64,02 (4,06)	65,08 (3,41)	64,92 (4,42)	65,14 (4,64)	65,14 (4,66)	65,42 (3,79)	65,55 (3,56)	66,60 (3,62)	65,98 (3,68)
Haberman	75,32 (2,71)	76,12 (2,68)	76,41 (2,52)	76,34 (2,81)	76,58 (2,64)	76,03 (2,58)	76,21 (2,84)	76,36 (2,56)	76,17 (2,89)	76,71 (2,72)
Heart	84,79 (2,82)	86,05 (2,89)	86,20 (2,82)	86,44 (2,78)	86,35 (2,46)	86,54 (2,48)	86,30 (2,57)	86,42 (2,83)	86,32 (2,62)	86,47 (2,67)
Hepatitis	87,69 (3,15)	89,23 (2,77)	89,53 (3,05)	90,09 (2,40)	89,70 (3,02)	90,04 (2,73)	90,34 (2,95)	90,13 (2,94)	89,66 (2,75)	90,13 (2,61)
Horse	84,47 (2,13)	85,49 (1,86)	85,69 (1,93)	85,71 (2,07)	85,54 (1,85)	85,47 (1,84)	85,82 (1,97)	85,49 (2,18)	85,49 (1,96)	85,51 (2,01)
Ilpd	72,98 (1,85)	73,84 (1,53)	73,66 (2,25)	74,19 (1,90)	74,02 (1,55)	74,32 (1,98)	74,44 (1,72)	74,11 (1,63)	74,39 (1,88)	74,75 (1,61)
Ionosphere	91,02 (2,15)	91,63 (1,55)	92,29 (1,60)	92,59 (1,82)	92,82 (1,69)	93,01 (1,64)	92,84 (1,70)	92,97 (1,79)	93,03 (1,61)	93,09 (1,46)
Iris	95,11 (2,30)	95,91 (2,04)	96,22 (1,95)	96,22 (1,89)	96,58 (1,84)	96,40 (1,76)	96,49 (1,90)	96,58 (1,74)	96,53 (1,59)	96,49 (1,66)
Laryngeal1	84,61 (3,89)	85,33 (3,45)	85,64 (2,64)	85,86 (3,01)	86,39 (2,98)	85,98 (2,99)	86,39 (3,07)	86,11 (2,56)	86,11 (2,80)	86,32 (3,19)
Laryngeal3	74,05 (2,33)	74,35 (2,28)	74,75 (2,39)	74,99 (2,84)	74,99 (2,54)	75,24 (2,63)	75,48 (2,46)	75,50 (2,40)	75,52 (2,12)	75,31 (2,45)
Lymph	81,08 (3,65)	83,06 (3,38)	83,38 (2,84)	82,39 (3,18)	83,06 (3,72)	83,33 (2,89)	83,11 (3,15)	83,78 (3,07)	83,24 (2,69)	83,33 (3,28)
Mammo	82,65 (2,84)	84,12 (1,55)	83,82 (1,74)	83,83 (1,55)	83,74 (1,68)	83,83 (1,58)	84,17 (1,38)	83,88 (1,53)	83,78 (1,74)	83,69 (1,54)
Phoneme	85,13 (0,93)	85,70 (0,49)	85,75 (0,58)	85,93 (0,63)	85,90 (0,54)	85,92 (0,61)	85,88 (0,61)	85,97 (0,45)	85,94 (0,56)	85,94 (0,59)
Segment	89,71 (4,37)	93,65 (1,06)	94,14 (0,93)	94,46 (0,76)	94,68 (0,60)	94,68 (0,73)	94,66 (0,71)	94,73 (0,61)	94,75 (0,70)	94,78 (0,72)
Sonar	80,48 (4,05)	86,28 (3,03)	87,63 (2,88)	88,24 (2,23)	88,91 (2,77)	88,91 (1,87)	89,13 (1,88)	89,36 (2,13)	89,29 (3,20)	89,81 (2,51)
Thyroid	94,35 (1,53)	94,35 (1,71)	94,91 (1,73)	94,97 (1,45)	95,31 (1,57)	95,28 (1,73)	95,28 (1,64)	95,40 (1,39)	95,25 (1,45)	95,43 (1,50)
Waveform	78,45 (0,95)	80,27 (0,98)	81,16 (0,86)	81,80 (0,57)	82,15 (0,92)	82,43 (0,70)	82,47 (0,66)	82,65 (0,71)	82,83 (0,63)	83,17 (0,73)
Weaning	86,49 (2,23)	87,95 (2,50)	87,95 (2,32)	88,61 (1,78)	88,43 (1,59)	88,63 (1,95)	88,45 (1,69)	88,45 (2,13)	88,59 (2,17)	88,59 (2,01)
Wine	96,52 (1,65)	97,04 (1,43)	97,12 (1,61)	97,45 (1,10)	97,49 (1,31)	97,60 (1,50)	97,90 (1,24)	98,01 (1,28)	98,01 (1,05)	97,87 (1,23)
Médias	81,46	82,96	83,38	83,60	83,71	83,76	83,86	83,86	83,86	83,95

Tabela 3 – Médias da acuracidade no conjunto  $\tau$  considerando *ensembles* gerados pelo Algoritmo 1 com  $L$  variando de 55 à 100

Problema	L = 55	L = 60	L = 65	L = 70	L = 75	L = 80	L = 85	L = 90	L = 95	L = 100
Abalone	57,17 (0,80)	56,68 (1,13)	56,75 (0,82)	56,70 (0,97)	56,56 (1,12)	56,47 (1,01)	56,60 (1,17)	56,35 (0,95)	56,38 (1,02)	56,10 (0,87)
Australian	87,46 (1,37)	87,36 (1,46)	87,31 (1,46)	87,22 (1,45)	87,35 (1,47)	87,25 (1,43)	87,58 (1,17)	87,11 (1,45)	87,14 (1,41)	87,09 (1,38)
Balance	88,34 (0,92)	88,37 (0,93)	88,51 (0,83)	88,42 (0,88)	88,33 (0,90)	88,40 (0,88)	88,37 (0,91)	88,29 (0,94)	88,27 (0,84)	88,41 (0,97)
Blood	79,83 (1,83)	79,62 (2,02)	79,71 (2,00)	79,71 (1,71)	79,82 (1,87)	79,57 (2,04)	79,71 (1,78)	79,71 (1,87)	79,70 (2,06)	79,56 (2,21)
Bupa	78,19 (2,91)	78,50 (2,60)	78,71 (2,75)	78,09 (2,56)	78,50 (3,15)	78,15 (2,59)	78,61 (2,62)	78,36 (2,85)	78,13 (2,62)	78,57 (2,47)
Cancer	97,03 (0,78)	97,00 (0,77)	97,02 (0,78)	97,02 (0,81)	97,03 (0,77)	96,99 (0,75)	97,07 (0,75)	96,96 (0,79)	97,00 (0,76)	97,00 (0,78)
Chess	95,85 (0,49)	95,88 (0,59)	95,77 (0,61)	95,91 (0,50)	95,84 (0,48)	95,86 (0,50)	95,88 (0,39)	95,83 (0,57)	95,89 (0,49)	95,92 (0,46)
CMC	58,72 (1,82)	58,96 (1,31)	58,81 (1,80)	58,72 (1,65)	58,38 (1,68)	58,60 (1,77)	58,48 (1,64)	58,64 (1,70)	58,07 (1,65)	58,38 (1,51)
Contractions	92,04 (2,64)	91,29 (3,47)	91,43 (2,70)	91,50 (3,04)	91,56 (3,25)	91,63 (3,36)	91,63 (2,75)	91,77 (2,91)	91,29 (3,59)	92,11 (3,29)
Flare	73,37 (2,24)	73,19 (2,08)	73,38 (2,43)	73,63 (2,15)	73,41 (1,92)	73,43 (2,18)	73,79 (2,38)	73,51 (1,92)	73,70 (2,30)	73,69 (2,16)
German	78,22 (1,21)	77,89 (1,28)	77,78 (1,19)	77,73 (1,14)	77,65 (1,23)	77,61 (1,24)	77,75 (1,11)	77,60 (1,44)	77,49 (1,17)	77,71 (1,22)
Glass	66,14 (3,58)	65,17 (4,40)	66,01 (3,54)	66,04 (4,04)	65,89 (3,95)	65,83 (3,92)	65,26 (3,62)	65,92 (4,00)	66,14 (3,42)	66,07 (3,93)
Haberman	76,12 (3,38)	76,60 (2,78)	76,23 (2,67)	76,38 (2,38)	76,47 (2,56)	76,36 (2,87)	76,06 (3,00)	76,38 (2,54)	76,23 (2,92)	76,47 (2,87)
Heart	86,42 (2,73)	86,49 (2,76)	86,32 (2,82)	86,59 (2,61)	86,27 (2,86)	86,35 (2,78)	86,42 (2,84)	86,40 (2,80)	86,52 (2,66)	86,44 (2,75)
Hepatitis	90,30 (2,56)	90,00 (2,44)	90,09 (2,73)	90,04 (3,08)	90,09 (3,08)	90,38 (2,51)	89,91 (2,65)	90,68 (2,75)	90,13 (2,92)	90,00 (2,68)
Horse	85,78 (1,68)	85,33 (2,06)	85,25 (1,69)	85,53 (2,06)	85,58 (1,67)	85,53 (1,92)	85,43 (1,73)	85,42 (1,79)	85,38 (1,84)	85,49 (1,88)
Ilpd	74,66 (1,69)	74,45 (1,86)	74,24 (1,89)	74,75 (1,98)	74,60 (1,67)	74,58 (1,61)	74,67 (1,62)	74,53 (1,93)	74,97 (1,77)	74,74 (1,64)
Ionosphere	93,14 (1,72)	93,09 (1,61)	93,37 (1,66)	93,52 (1,53)	93,33 (1,75)	93,35 (1,65)	93,20 (1,50)	93,20 (1,46)	93,26 (1,71)	93,20 (1,69)
Iris	96,49 (1,87)	96,58 (1,63)	96,62 (1,84)	96,58 (1,70)	96,49 (1,70)	96,49 (1,70)	96,40 (1,65)	96,53 (1,77)	96,76 (1,63)	96,40 (1,76)
Laryngeal1	86,36 (2,86)	86,54 (2,76)	86,36 (2,77)	86,51 (3,13)	86,39 (3,10)	86,26 (2,60)	86,48 (2,71)	86,36 (2,95)	86,51 (2,97)	86,11 (2,82)
Laryngeal3	75,57 (2,88)	75,31 (2,65)	75,39 (2,88)	75,42 (2,78)	75,46 (2,68)	75,76 (2,53)	75,46 (2,40)	75,65 (2,53)	76,10 (2,66)	75,74 (2,86)
Lymph	83,24 (3,26)	83,06 (3,15)	83,29 (3,39)	83,47 (3,40)	83,51 (3,06)	83,65 (3,26)	83,29 (3,45)	83,51 (3,12)	83,15 (3,38)	83,33 (3,39)
Mammo	83,82 (1,83)	83,79 (1,37)	83,75 (1,65)	83,52 (1,60)	83,87 (1,51)	83,73 (1,76)	83,67 (1,66)	83,80 (1,57)	83,69 (1,53)	83,81 (1,54)
Phoneme	86,00 (0,54)	85,98 (0,53)	85,94 (0,54)	85,97 (0,53)	85,94 (0,55)	85,91 (0,55)	85,90 (0,55)	85,93 (0,60)	85,96 (0,48)	85,99 (0,58)
Segment	94,85 (0,62)	94,70 (0,66)	94,81 (0,67)	94,77 (0,66)	94,79 (0,60)	94,79 (0,67)	94,85 (0,69)	94,83 (0,68)	94,87 (0,70)	94,90 (0,70)
Sonar	89,90 (2,43)	89,55 (1,94)	89,84 (2,06)	89,84 (1,93)	90,03 (1,90)	89,17 (2,51)	89,97 (1,83)	89,68 (2,38)	90,58 (1,96)	89,74 (1,86)
Thyroid	95,31 (1,57)	95,34 (1,60)	95,06 (1,71)	95,25 (1,45)	95,22 (1,50)	95,06 (1,60)	95,15 (1,57)	95,22 (1,61)	95,31 (1,63)	95,49 (1,82)
Waveform	82,82 (0,72)	82,98 (0,73)	83,28 (0,68)	83,14 (0,56)	83,40 (0,69)	83,42 (0,66)	83,54 (0,67)	83,50 (0,75)	83,54 (0,70)	83,63 (0,59)
Weaning	88,59 (1,85)	88,79 (2,02)	88,74 (2,01)	88,76 (1,85)	88,85 (1,84)	88,28 (1,95)	88,65 (1,78)	88,68 (1,95)	88,72 (1,92)	88,68 (1,92)
Wine	98,16 (1,20)	98,01 (1,21)	98,09 (1,32)	98,28 (1,05)	97,98 (1,30)	98,31 (1,17)	98,20 (1,27)	97,94 (1,11)	98,09 (1,26)	98,01 (1,28)
Médias	84,00	83,88	83,93	83,97	83,95	83,91	83,93	83,94	83,97	83,96

## 6.2 Comparações com Técnicas de Seleção

Nesta seção foram feitas comparações com algumas das principais técnicas de seleção dinâmica de classificadores: KNOP, KNORAE, KNORAU, LCA, OLA e RANK. Todos estes métodos foram apresentados na seção 3.1 do capítulo 3. Eles foram escolhidos porque são comumente utilizados na realização de comparações com novos trabalhos da área de seleção dinâmica. Foi utilizada em todos eles uma vizinhança de tamanho 7 para estimação da competência. Este valor mostrou-se mais adequado em estudos anteriores (KO; SABOURIN; BRITTO, 2008; CRUZ; CAVALCANTI; REN, 2011). Além dessas 6 técnicas dinâmicas, também foram realizadas comparações com a combinação estática (CE) considerando sempre a seleção de todos os componentes do *ensemble*.

As tabelas 4 e 5 apresentam médias da acuracidade no conjunto de generalização  $\zeta$  com desvios entre parênteses. Os valores em negrito apenas destacam a maior acuracidade alcançada para cada um dos problemas. Diferenças significativas em relação ao método proposto (MP), calculadas de acordo com teste de *Wilcoxon* com 95% de confiança, possuem o marcador asterisco. Todos os métodos da tabela 4 trabalham com o conjunto de classificadores que foi estabelecido na seção anterior. Na tabela 5 os métodos CE, KNOP, KNORAE, KNORAU, LCA, OLA e RANK trabalham com *ensembles* de 25 *Decision Stumps*<sup>3</sup> (SAMMUT; WEBB, 2010) gerados através do algoritmo *Bagging*. Nesse caso, para uma comparação justa, foram utilizadas as mesmas divisões estratificadas da seção anterior, com *bags* de 10% do conjunto  $\tau$  para a construção de cada componente. O valor de 10% foi definido com base em experimentos preliminares, onde verificou-se que para valores maiores o *ensemble* apresenta taxas menores de acuracidade e diversidade.

Para a realização de comparações com os resultados da acuracidade, apresentados nas tabelas 4 e 5, foram aplicados testes estatísticos com 95% de confiança. O teste de *Friedman* indicou que existe diferença significativa entre os métodos. Portanto, efetuou-se o pós-teste de *Nemenyi*. As representações gráficas dos ranqueamentos produzidos são exibidas nas figuras 14 e 15. É possível observar em ambos gráficos que o MP conseguiu conquistar a melhor posição geral no ranqueamento, independentemente do conjunto de classificadores que foi utilizado. Além do mais, o mesmo pode ser considerado estatisticamente superior às técnicas de seleção dinâmica que foram avaliadas, pois a diferença entre KNOP, KNORAE, KNORAU, LCA, OLA e RANK é maior do que o valor da distância crítica. Por outro lado, a CE utilizando os componentes gerados pelo Algoritmo 1 foi equivalente ao MP. Apesar disso, esta apresentou 1 vitória, 14 empates e 15 derrotas com relação ao MP segundo o teste de *Wilcoxon*. Isso pode sugerir que há mais vantagem no uso do MP na grande maioria dos problemas que foram analisados.

<sup>3</sup> Este classificador foi escolhido pela semelhança com o utilizado pelo Algoritmo 1. É uma árvore de decisão simples e binária. No entanto, produz classes de probabilidades e tem como base o ganho de informação, o qual divide os conjuntos de dados tomando como referência o atributo mais informativo.

Tabela 4 – Médias da acuracidade considerando o conjunto de classificadores gerados pelo Algoritmo 1. O número de derrotas, empates e vitórias alcançado por cada método, de acordo com os testes de *Wilcoxon*, são apresentados no final da tabela

Problema	MP	CE	KNOP	KNORAE	KNORAU	LCA	OLA	RANK
Abalone	53,03 (1,96)	<b>53,07 (1,56)</b>	51,81 (3,07)*	52,19 (1,49)*	52,81 (1,49)	49,20 (3,77)*	52,24 (1,75)	49,24 (1,27)*
Australian	84,92 (2,30)	84,98 (2,02)	84,55 (2,47)	83,10 (3,32)*	<b>85,06 (2,05)</b>	75,33 (8,65)*	82,81 (2,22)	79,21 (3,15)*
Balance	<b>88,01 (1,77)</b>	82,67 (3,09)*	81,88 (3,05)*	82,76 (2,90)*	81,82 (3,24)*	69,91 (3,39)*	83,82 (2,28)*	82,50 (2,52)*
Blood	77,17 (2,42)	76,93 (1,28)	75,51 (2,52)*	75,42 (2,82)*	<b>77,24 (1,39)</b>	68,47 (9,00)*	76,74 (2,18)	71,89 (3,66)*
Bupa	68,49 (4,65)	<b>69,07 (4,80)</b>	64,11 (5,04)*	63,02 (5,58)*	68,41 (4,75)	52,29 (5,22)*	59,38 (5,31)*	60,27 (4,63)*
Cancer	<b>96,92 (0,89)</b>	96,21 (1,03)*	95,29 (1,61)*	94,98 (1,77)*	96,11 (1,01)*	94,79 (1,94)*	94,62 (2,10)*	94,21 (2,14)*
Chess	<b>95,41 (0,99)</b>	88,60 (3,20)*	87,48 (5,46)*	87,30 (3,29)*	88,04 (4,60)*	65,47 (4,58)*	95,00 (0,89)*	94,38 (0,96)*
CMC	50,57 (3,43)	<b>50,75 (4,03)</b>	49,93 (3,50)	48,48 (3,07)*	49,67 (3,67)	40,91 (3,02)*	47,16 (2,78)*	44,78 (2,54)*
Contractions	84,31 (6,06)	<b>84,44 (7,09)</b>	83,19 (7,30)	81,81 (6,88)*	84,44 (7,25)	75,83 (8,22)*	78,75 (7,84)*	80,14 (7,56)*
Flare	70,54 (2,88)	68,59 (6,43)	68,47 (6,69)	69,21 (6,04)	67,74 (6,57)	65,59 (3,57)*	<b>72,55 (1,83)</b>	69,08 (2,17)*
German	<b>73,27 (1,63)</b>	72,75 (1,44)*	70,87 (2,15)*	70,75 (1,75)*	72,68 (1,82)	66,48 (3,99)*	69,16 (3,17)*	67,04 (3,59)*
Glass	<b>62,45 (5,70)</b>	56,16 (7,72)*	57,11 (7,88)*	57,80 (6,20)*	54,78 (8,72)*	46,92 (4,82)*	59,31 (4,31)*	61,13 (6,32)*
Haberman	72,46 (3,64)	<b>73,73 (3,50)*</b>	72,28 (4,79)	72,11 (5,06)	73,16 (2,94)*	68,03 (10,57)	69,43 (5,00)	65,09 (5,78)*
Heart	<b>82,39 (4,51)</b>	82,09 (4,43)	80,25 (3,81)*	78,16 (4,48)*	81,49 (4,30)*	72,09 (7,81)*	75,52 (5,03)*	75,12 (5,74)*
Hepatitis	<b>82,72 (5,64)</b>	82,02 (5,53)	80,61 (5,37)*	80,18 (6,44)*	81,67 (4,66)	77,54 (4,62)*	78,42 (6,08)*	77,63 (6,90)*
Horse	<b>84,13 (2,96)</b>	83,12 (2,75)*	81,96 (3,07)*	79,67 (5,02)*	83,26 (3,03)	68,30 (7,97)*	79,75 (4,16)*	77,21 (3,78)*
ILPD	70,67 (2,81)	71,17 (2,45)	69,66 (3,46)	69,59 (3,53)	<b>71,31 (2,47)</b>	63,54 (8,83)*	66,80 (2,89)*	64,92 (3,34)*
Ionosphere	<b>90,42 (2,65)</b>	84,71 (2,88)*	85,82 (1,99)*	83,64 (3,27)*	84,87 (2,80)*	75,86 (6,06)*	84,25 (3,01)*	84,10 (4,12)*
Iris	94,77 (2,74)	<b>94,86 (2,87)</b>	94,05 (3,36)	94,14 (3,02)	94,68 (2,30)	82,16 (7,89)*	93,60 (3,44)*	93,42 (4,06)*
Laryngeal1	<b>82,26 (4,74)</b>	81,95 (4,31)	81,64 (4,83)	80,63 (4,65)*	82,39 (4,94)	75,03 (6,26)*	78,24 (5,73)*	77,48 (5,71)*
Laryngeal3	<b>74,05 (3,76)</b>	71,14 (2,32)*	69,43 (3,72)*	69,62 (3,68)*	70,57 (2,45)*	63,11 (8,32)*	70,19 (4,68)*	65,11 (5,36)*
Lymph	<b>80,36 (4,65)</b>	79,10 (4,54)	77,48 (7,11)*	76,04 (5,62)*	79,10 (5,36)	63,15 (10,25)*	75,77 (6,90)*	77,12 (5,80)*
Mammo	81,88 (1,92)	<b>81,90 (2,04)</b>	81,40 (2,17)	80,65 (2,08)*	81,72 (2,03)	75,26 (5,06)*	80,33 (2,24)*	75,99 (2,51)*
Phoneme	<b>84,94 (1,13)</b>	78,46 (1,00)*	77,59 (2,65)*	81,28 (1,29)*	78,67 (1,34)*	75,80 (1,82)*	83,39 (0,94)*	84,21 (1,10)*
Segment	<b>94,12 (0,92)</b>	63,17 (7,78)*	67,02 (8,09)*	71,56 (6,26)*	65,25 (11,59)*	83,44 (2,31)*	91,68 (1,27)*	93,82 (0,82)*
Sonar	<b>79,36 (5,30)</b>	76,67 (5,64)*	73,59 (6,60)*	74,04 (5,34)*	74,68 (6,61)*	58,46 (6,65)*	71,79 (5,11)*	73,21 (6,25)*
Thyroid	<b>93,96 (3,19)</b>	85,53 (3,58)*	88,68 (4,09)*	88,11 (4,35)*	85,66 (4,28)*	81,01 (5,24)*	91,70 (4,19)*	92,39 (4,22)*
Waveform	<b>82,06 (0,89)</b>	77,69 (2,00)*	75,17 (1,61)*	72,76 (1,88)*	75,54 (3,07)*	58,77 (2,58)*	75,03 (1,23)*	72,01 (1,27)*
Weaning	<b>84,31 (3,41)</b>	83,16 (3,14)*	80,62 (3,26)*	77,47 (5,98)*	82,49 (2,82)*	66,80 (7,93)*	76,98 (4,10)*	76,36 (4,17)*
Wine	<b>96,14 (2,87)</b>	93,48 (3,09)*	91,89 (4,73)*	92,42 (4,23)*	92,12 (4,65)*	85,61 (6,76)*	94,17 (3,46)*	93,86 (3,49)*
Derrotas	-	15	21	26	14	29	25	30
Empates	-	14	9	4	15	1	4	0
Vitórias	-	1	0	0	1	0	1	0

Tabela 5 – Médias da acuracidade considerando o conjunto de classificadores gerados através do algoritmo *Bagging*. O número de derrotas, empates e vitórias alcançado por cada método, de acordo com os testes de *Wilcoxon*, são apresentados no final da tabela

Problema	MP	CE	KNOP	KNORAE	KNORAU	LCA	OLA	RANK
Abalone	53,03 (1,96)	<b>53,42 (1,00)</b>	52,81 (1,05)	53,00 (1,09)	52,93 (1,30)	52,42 (1,63)	52,82 (1,51)	51,59 (1,24)
Australian	<b>84,92 (2,30)</b>	84,98 (2,02)	84,71 (2,20)	84,55 (2,38)	84,98 (2,02)	81,63 (3,72)*	82,33 (2,34)*	80,58 (3,40)*
Balance	<b>88,01 (1,77)</b>	78,57 (3,76)*	77,80 (4,05)*	79,72 (4,53)*	77,91 (3,87)*	69,76 (3,56)*	84,55 (2,37)*	82,63 (2,18)*
Blood	<b>77,17 (2,42)</b>	76,11 (0,43)*	76,08 (1,48)*	76,33 (1,21)*	76,11 (0,43)*	70,80 (6,57)*	76,95 (2,74)	73,90 (3,07)*
Bupa	<b>68,49 (4,65)</b>	63,64 (3,47)*	61,59 (4,55)*	61,55 (5,35)*	62,02 (4,54)*	52,33 (5,58)*	59,30 (5,14)*	59,53 (4,80)*
Cancer	<b>96,92 (0,89)</b>	95,15 (1,24)*	95,25 (1,37)*	94,67 (1,24)*	95,23 (1,38)*	95,33 (1,83)*	95,17 (1,36)*	94,60 (1,38)*
Chess	<b>95,41 (0,99)</b>	66,93 (4,51)*	67,78 (5,07)*	68,76 (7,68)*	66,98 (4,78)*	68,69 (2,47)*	94,80 (1,06)*	93,45 (1,49)*
CMC	<b>50,57 (3,43)</b>	47,45 (3,22)*	46,94 (3,71)*	45,98 (2,41)*	46,67 (3,44)*	41,61 (3,54)*	47,07 (3,58)*	45,52 (3,02)*
Contractions	<b>84,31 (6,06)</b>	75,97 (8,02)*	75,28 (8,95)*	75,83 (9,63)*	75,00 (8,11)*	62,36 (16,21)*	76,81 (7,87)*	79,72 (7,24)*
Flare	70,54 (2,88)	54,95 (6,20)*	52,07 (4,24)*	54,22 (4,28)*	53,79 (4,95)*	65,41 (2,40)*	69,17 (2,96)*	69,14 (3,58)
German	<b>73,27 (1,63)</b>	70,23 (0,55)*	69,33 (2,32)*	70,24 (2,32)*	69,85 (0,79)*	68,21 (3,13)*	68,12 (3,04)*	68,12 (3,04)*
Glass	<b>62,45 (5,70)</b>	53,71 (7,18)*	48,99 (7,86)*	47,55 (7,98)*	47,99 (7,64)*	44,59 (10,95)*	58,93 (6,45)*	60,38 (6,72)
Haberman	72,46 (3,64)	<b>73,29 (2,47)</b>	72,15 (4,16)	72,63 (3,63)	73,25 (1,70)	70,31 (6,74)	71,49 (4,65)	65,70 (5,08)*
Heart	<b>82,39 (4,51)</b>	80,25 (3,73)*	76,97 (3,85)*	75,57 (5,58)*	79,40 (4,28)*	66,97 (7,15)*	75,12 (3,98)*	74,13 (4,98)*
Hepatitis	<b>82,72 (5,64)</b>	79,82 (2,79)*	80,18 (4,07)*	77,28 (4,23)*	80,35 (3,44)*	76,49 (5,44)*	80,88 (4,88)	79,56 (6,87)*
Horse	<b>84,13 (2,96)</b>	79,71 (5,40)*	78,30 (5,88)*	77,43 (5,83)*	79,24 (5,95)*	69,28 (7,68)*	79,42 (4,76)*	77,14 (4,08)*
ILPD	70,67 (2,81)	<b>72,05 (1,89)*</b>	68,55 (4,14)*	69,10 (3,11)*	71,49 (1,51)	64,05 (8,37)*	66,85 (3,24)*	65,56 (2,83)*
Ionosphere	<b>90,42 (2,65)</b>	82,80 (3,46)*	83,64 (3,22)*	82,72 (3,33)*	81,95 (3,76)*	76,48 (6,04)*	84,25 (3,11)*	83,64 (3,59)*
Iris	<b>94,77 (2,74)</b>	81,80 (7,90)*	81,17 (10,37)*	81,53 (11,69)*	77,39 (9,35)*	81,35 (7,91)*	90,72 (4,80)*	90,90 (5,66)*
Laryngeal1	82,26 (4,74)	81,64 (5,36)	80,82 (5,61)	<b>82,39 (4,89)</b>	81,76 (5,85)	75,91 (5,75)*	79,50 (5,84)*	78,49 (6,00)*
Laryngeal3	<b>74,05 (3,76)</b>	67,39 (2,86)*	67,88 (3,58)*	67,46 (3,09)*	67,12 (3,16)*	65,80 (4,22)*	67,92 (3,36)*	65,30 (4,86)*
Lymph	<b>80,36 (4,65)</b>	76,67 (7,39)*	72,61 (6,18)*	71,89 (6,33)*	77,21 (7,50)*	57,93 (9,19)*	73,60 (7,83)*	74,23 (6,46)*
Mammo	81,88 (1,92)	<b>81,94 (2,07)</b>	81,06 (2,29)*	80,79 (1,73)*	81,75 (2,07)	77,32 (2,60)*	80,56 (1,70)*	78,04 (2,88)*
Phoneme	<b>84,94 (1,13)</b>	77,25 (1,18)*	75,39 (2,12)*	77,76 (2,70)*	76,27 (0,99)*	77,17 (1,74)*	81,90 (1,54)*	82,99 (1,56)*
Segment	<b>94,12 (0,92)</b>	62,40 (6,52)*	45,67 (8,71)*	52,60 (8,23)*	48,91 (11,47)*	84,15 (2,04)*	91,98 (1,40)*	93,70 (0,93)*
Sonar	<b>79,36 (5,30)</b>	72,12 (5,86)*	68,85 (6,34)*	70,19 (7,33)*	69,74 (7,26)*	57,31 (7,88)*	70,26 (6,75)*	71,99 (7,28)*
Thyroid	<b>93,96 (3,19)</b>	78,81 (3,27)*	80,31 (2,74)*	80,00 (2,51)*	76,48 (3,70)*	77,30 (3,98)*	91,82 (3,03)*	92,89 (3,66)*
Waveform	<b>82,06 (0,89)</b>	66,55 (2,88)*	62,63 (4,53)*	61,93 (4,68)*	60,64 (2,17)*	61,07 (2,36)*	73,91 (1,50)*	70,08 (1,91)*
Weaning	<b>84,31 (3,41)</b>	80,09 (5,16)*	78,31 (4,89)*	76,84 (5,46)*	79,64 (5,51)*	66,36 (8,39)*	76,49 (3,56)*	74,62 (4,34)*
Wine	<b>96,14 (2,87)</b>	80,30 (8,72)*	76,21 (8,71)*	76,89 (8,27)*	77,65 (8,75)*	81,21 (8,35)*	89,47 (4,68)*	89,92 (4,49)*
Derrotas	-	24	26	26	24	28	25	27
Empates	-	5	4	4	6	2	4	3
Vitórias	-	1	0	0	0	0	1	0

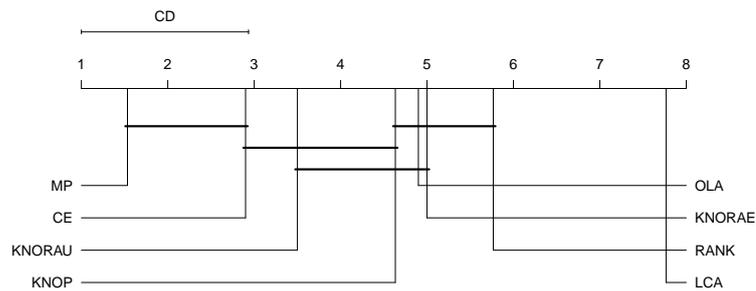


Figura 14 – Representação do teste de *Nemenyi* comparando resultados da tabela 4

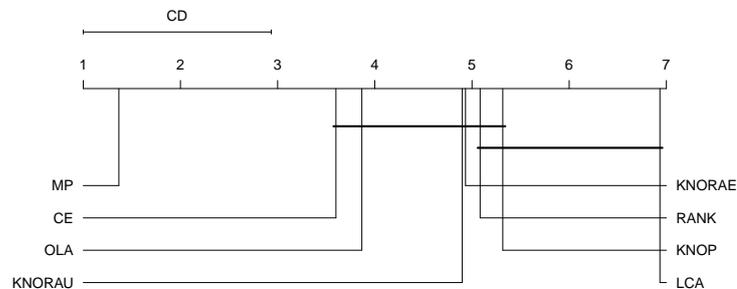


Figura 15 – Representação do teste de *Nemenyi* comparando resultados da tabela 5

Considerando as posições alcançadas nos ranqueamentos das figuras 14 e 15, nota-se que algumas técnicas puderam se beneficiar com o uso dos classificadores gerados pelo Algoritmo 1. CE, KNOP e KNORAU, por exemplo, encontram-se mais distantes de MP na figura 15 e apresentam maiores posições no ranqueamento, indicando que houve piora na acuracidade dos mesmos com o uso dos classificadores gerados por *Bagging*. Além do mais, valores de derrotas, empates e vitórias, apresentados no final das tabelas 4 e 5, evidenciam que, em geral, essas três técnicas quando utilizam os classificadores gerados pelo Algoritmo 1 alcançam melhores resultados. As técnicas KNORAE e OLA praticamente se mantiveram estáveis e as demais apresentaram pequena melhora com o *Bagging*. Para finalizar a seção, compara-se na tabela 6 medidas de diversidade no conjunto  $\zeta$  considerando os classificadores que foram gerados pelo Algoritmo 1 e *Bagging*. Semelhantemente as análises realizadas anteriormente, os valores que estão destacados negrito indicam o melhor valor para cada medida. Diferenças significativas em relação ao Algoritmo 1, calculadas de acordo com teste de *Wilcoxon* com 95% de confiança, possuem o marcador asterisco. O número de vitórias do Algoritmo 1 é consideravelmente maior. Isso já era esperado, uma vez que a diversidade é otimizada durante o processo de geração dos componentes.

Tabela 6 – Médias da Correlação, Desacordo, Falta Dupla e Estatística Q entre classificadores gerados pelo Algoritmo 1 e Bagging

Problema	Correlação		Desacordo		Falta Dupla		Estatística Q	
	Algoritmo 1	Bagging	Algoritmo 1	Bagging	Algoritmo 1	Bagging	Algoritmo 1	Bagging
Abalone	<b>0,1248 (0,0123)</b>	0,4437 (0,0806)*	<b>0,4360 (0,0060)</b>	0,2782 (0,0403)*	<b>0,3072 (0,0063)</b>	0,3462 (0,0189)*	<b>0,1099 (0,0104)</b>	0,5013 (0,1014)*
Australian	<b>0,3365 (0,0438)</b>	0,6268 (0,1250)*	<b>0,2879 (0,0223)</b>	0,1448 (0,0505)*	<b>0,1097 (0,0117)</b>	0,1198 (0,0194)*	<b>0,3719 (0,0598)</b>	0,6705 (0,1169)*
Balance	<b>0,1799 (0,0157)</b>	0,1905 (0,0263)*	<b>0,3871 (0,0105)</b>	0,3868 (0,0146)	<b>0,1870 (0,0065)</b>	0,1991 (0,0093)*	<b>0,2362 (0,0261)</b>	0,2560 (0,0392)*
Blood	<b>0,2173 (0,0233)</b>	0,6309 (0,1113)*	<b>0,3669 (0,0126)</b>	0,1599 (0,0515)*	<b>0,1531 (0,0078)</b>	0,1892 (0,0140)*	<b>0,1834 (0,0476)</b>	0,6808 (0,1261)*
Bupa	<b>0,0284 (0,0151)</b>	0,0404 (0,0303)	<b>0,4795 (0,0072)</b>	0,4765 (0,0158)	<b>0,2017 (0,0097)</b>	0,2176 (0,0124)*	<b>0,0027 (0,0255)</b>	0,0366 (0,0485)*
Cancer	<b>0,2452 (0,0384)</b>	0,3281 (0,0676)*	<b>0,1515 (0,0138)</b>	0,1255 (0,0204)*	<b>0,0345 (0,0050)</b>	0,0370 (0,0059)*	<b>0,5358 (0,0819)</b>	0,6597 (0,1019)*
Chess	<b>0,0361 (0,0103)</b>	0,5492 (0,1585)*	<b>0,4452 (0,0077)</b>	0,2021 (0,0705)*	<b>0,1370 (0,0058)</b>	0,2379 (0,0379)*	<b>-0,0834 (0,0277)</b>	0,4157 (0,2038)*
Cmc	<b>0,0702 (0,0125)</b>	0,1608 (0,0639)*	<b>0,4518 (0,0061)</b>	0,4082 (0,0297)*	<b>0,3568 (0,0117)</b>	0,3761 (0,0138)*	<b>0,0476 (0,0216)</b>	0,1719 (0,0845)*
Contractions	0,2075 (0,0581)	<b>0,1106 (0,0633)*</b>	0,3195 (0,0330)	<b>0,4297 (0,0367)*</b>	<b>0,1082 (0,0281)</b>	0,1664 (0,0309)*	0,2977 (0,1262)	<b>0,1452 (0,0929)*</b>
Flare	<b>0,1107 (0,0291)</b>	0,4434 (0,0881)*	<b>0,4278 (0,0126)</b>	0,2782 (0,0436)*	0,3767 (0,0096)	<b>0,3656 (0,0129)*</b>	<b>0,1188 (0,0442)</b>	0,5148 (0,0885)*
German	<b>0,1485 (0,0170)</b>	0,4586 (0,1294)*	<b>0,4019 (0,0104)</b>	0,2468 (0,0637)*	<b>0,1681 (0,0057)</b>	0,2068 (0,0191)*	<b>0,2005 (0,0333)</b>	0,5568 (0,1837)*
Glass	<b>0,1195 (0,0184)</b>	0,1536 (0,0367)*	<b>0,4245 (0,0089)</b>	0,3972 (0,0166)*	<b>0,3828 (0,0110)</b>	0,4254 (0,0198)*	<b>0,0551 (0,0308)</b>	0,1812 (0,0661)*
Haberman	<b>0,2181 (0,0392)</b>	0,4020 (0,0840)*	<b>0,3748 (0,0198)</b>	0,2751 (0,0422)*	<b>0,1747 (0,0131)</b>	0,1844 (0,0155)*	<b>0,2777 (0,0623)</b>	0,5288 (0,1216)*
Heart	0,1927 (0,0525)	<b>0,1496 (0,0447)*</b>	0,3416 (0,0248)	<b>0,3821 (0,0242)*</b>	<b>0,1269 (0,0165)</b>	0,1414 (0,0154)*	0,2596 (0,0886)	<b>0,2337 (0,0772)*</b>
Hepatitis	0,2951 (0,0691)	<b>0,2787 (0,0675)*</b>	0,2818 (0,0358)	<b>0,3040 (0,0374)*</b>	<b>0,1162 (0,0211)</b>	0,1257 (0,0154)*	0,4664 (0,1153)	<b>0,4212 (0,1182)</b>
Horse	0,2137 (0,0419)	<b>0,1618 (0,0630)*</b>	0,3572 (0,0230)	<b>0,3901 (0,0352)*</b>	<b>0,1284 (0,0118)</b>	0,1478 (0,0186)*	0,2362 (0,0655)	<b>0,2023 (0,0829)</b>
Ilpd	<b>0,1642 (0,0244)</b>	0,3019 (0,0708)*	<b>0,3969 (0,0145)</b>	0,3205 (0,0372)*	<b>0,1697 (0,0087)</b>	0,1769 (0,0106)*	<b>0,1932 (0,0409)</b>	0,3739 (0,0999)*
Ionosphere	<b>0,1847 (0,0327)</b>	0,2641 (0,0768)*	<b>0,3506 (0,0156)</b>	0,3087 (0,0404)*	<b>0,1114 (0,0099)</b>	0,1230 (0,0156)*	<b>0,2761 (0,0523)</b>	0,4268 (0,1206)*
Iris	<b>-0,0207 (0,0086)</b>	0,0515 (0,0299)*	<b>0,4666 (0,0032)</b>	0,4654 (0,0130)	<b>0,1198 (0,0075)</b>	0,1982 (0,0201)*	<b>-0,3067 (0,0283)</b>	-0,0379 (0,0486)*
Laryngeal1	<b>0,2824 (0,0729)</b>	0,2973 (0,0611)	<b>0,2885 (0,0311)</b>	0,2774 (0,0321)*	<b>0,1170 (0,0210)</b>	0,1206 (0,0233)	<b>0,4433 (0,1150)</b>	0,4959 (0,0930)*
Laryngeal3	<b>0,1525 (0,0271)</b>	0,4112 (0,1093)*	<b>0,4199 (0,0138)</b>	0,2886 (0,0566)*	<b>0,2089 (0,0121)</b>	0,2491 (0,0164)*	<b>0,1763 (0,0391)</b>	0,5824 (0,1409)*
Lymph	<b>0,1728 (0,0347)</b>	0,1863 (0,0488)	<b>0,3903 (0,0208)</b>	0,3894 (0,0295)*	<b>0,1667 (0,0186)</b>	0,1822 (0,0212)*	<b>0,2566 (0,0594)</b>	0,2784 (0,0648)*
Mammo	<b>0,2538 (0,0303)</b>	0,4927 (0,0954)*	<b>0,3155 (0,0191)</b>	0,1834 (0,0395)*	<b>0,1267 (0,0104)</b>	0,1289 (0,0127)	<b>0,4004 (0,0493)</b>	0,6640 (0,0853)*
Phoneme	<b>0,2088 (0,0132)</b>	0,5691 (0,0776)*	<b>0,3440 (0,0087)</b>	0,1710 (0,0322)*	<b>0,1349 (0,0031)</b>	0,1745 (0,0133)*	<b>0,2976 (0,0225)</b>	0,6664 (0,0821)*
Segment	<b>-0,0189 (0,0094)</b>	0,0130 (0,0228)*	<b>0,4076 (0,0040)</b>	0,3954 (0,0085)*	<b>0,5195 (0,0031)</b>	0,5250 (0,0042)*	<b>-0,2091 (0,0226)</b>	-0,1425 (0,0478)*
Sonar	<b>0,0712 (0,0308)</b>	0,0792 (0,0505)	0,4423 (0,0177)	<b>0,4480 (0,0289)</b>	<b>0,1652 (0,0186)</b>	0,1871 (0,0195)*	<b>0,1000 (0,0533)</b>	0,1317 (0,0872)
Thyroid	<b>0,1948 (0,0423)</b>	0,4663 (0,0759)*	<b>0,3425 (0,0155)</b>	0,2061 (0,0307)*	<b>0,1058 (0,0108)</b>	0,1445 (0,0207)*	<b>0,2304 (0,0739)</b>	0,6560 (0,1120)*
Waveform	<b>-0,0113 (0,0060)</b>	0,3964 (0,1591)*	<b>0,5055 (0,0031)</b>	0,3002 (0,0795)*	<b>0,2375 (0,0043)</b>	0,3019 (0,0348)*	<b>-0,0500 (0,0076)</b>	0,4631 (0,2174)*
Wearing	<b>0,1303 (0,0358)</b>	0,1660 (0,0662)*	<b>0,3904 (0,0218)</b>	0,3830 (0,0367)	<b>0,1295 (0,0125)</b>	0,1417 (0,0193)*	<b>0,1334 (0,0637)</b>	0,1994 (0,0988)*
Wine	<b>-0,0033 (0,0112)</b>	0,0732 (0,0601)*	<b>0,4801 (0,0065)</b>	0,4632 (0,0300)*	<b>0,1548 (0,0106)</b>	0,2296 (0,0214)*	<b>-0,1566 (0,0287)</b>	0,0469 (0,0791)*
Derrotas	-	22	-	20	-	27	-	24
Empates	-	5	-	6	-	2	-	4
Vitórias	-	3	-	4	-	1	-	2

### 6.3 Comparações com Técnicas Estáticas

O objetivo dessa seção é comparar os valores de acuracidades obtidos pelo MP com técnicas estáticas. Esse tipo de comparação é de grande importância para justificar a utilização do MP, uma vez que, em geral, sistemas dinâmicos de classificação possuem maior custo computacional do que os estáticos. Foram utilizados aqui os seguintes algoritmos: *AdaboostM1* (ADBO), *Multilayer Perceptron* (MLP), *Random Forests* (RF) e *Support Vector Machine* (SVM). Alguns desses são considerados, segundo uma recente e exaustiva avaliação (FERNÁNDEZ-DELGADO et al., 2014), como sendo os melhores classificadores da literatura. Todos os algoritmos foram obtidos a partir do Weka 3 (pode ser baixado gratuitamente a partir do link <http://www.cs.waikato.ac.nz/ml/weka>) com parâmetros configurados, para cada problema da tabela 1, conforme mostra a lista abaixo.

ADBO O número de iterações variou de 10 à 100 em intervalos de 10.

MLP O número de neurônios escondidos variou de 10 à 100 em intervalos de 10.

RF O número de árvores variou de 25 à 200 em intervalos de 25.

SVM  $\gamma$  e C são escolhidos por um *grid-search* com valores crescentes dos parâmetros.

Todos os classificadores gerados nessa seção foram treinados com o mesmo conjunto  $\tau$  que foi usado para a construção dos *ensembles* da seção 6.1. Além desses 4 algoritmos, também foi considerado o único melhor componente do *ensembles* com  $L = 25$ , chamado aqui simplesmente de *Single Best* (SB). Já que os classificadores dessa seção não requerem o uso de  $\nu$ , que é usado para calcular as regiões de competência nos algoritmos de seleção dinâmica de classificadores, este conjunto foi usado para auxiliar na definição do melhor classificador gerado. O conjunto de generalização  $\zeta$  permaneceu inalterado.

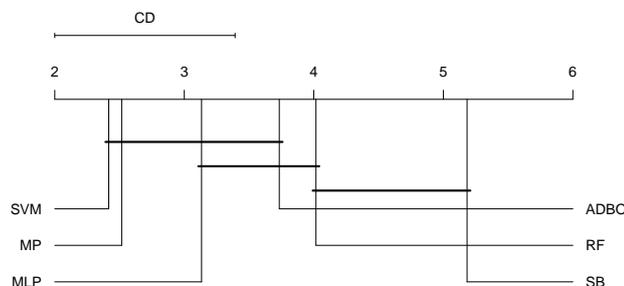


Figura 16 – Representação do teste de *Nemenyi* comparando métodos da tabela 7

A tabela 7 apresenta médias da acuracidade no conjunto  $\zeta$  com desvios entre parênteses. Os valores em negrito destacam o maior valor para cada problema. Diferenças significativas em relação ao MP, calculadas de acordo com teste de *Wilcoxon* com 95% de confiança, possuem o marcador asterisco. Para comparação de todos os métodos, também foram aplicados testes estatísticos com 95% de confiança. O teste de *Friedman* indicou que existe diferença significativa. Portanto, efetuou-se o pós-teste de *Nemenyi*. As representações gráficas dos ranqueamentos produzidos são exibidas na figura 16.

Tabela 7 – Comparação da acuracidade do MP contra técnicas estáticas. O número de derrotas, empates e vitórias alcançado por cada uma dessas técnicas, de acordo com os testes de *Wilcoxon*, são apresentados no final da tabela

Problema	MP	ADBO	MLP	RF	SB	SVM
Abalone	53,03 (1,96)	53,05 (1,22)	53,75 (1,72)	53,38 (1,06)	53,32 (1,31)	<b>54,57 (1,11)*</b>
Australian	84,92 (2,30)	84,86 (2,28)	81,80 (2,38)*	81,30 (3,26)*	<b>84,98 (2,02)</b>	83,84 (2,11)*
Balance	88,01 (1,77)	74,87 (3,69)*	<b>95,24 (2,10)*</b>	72,35 (7,85)*	61,20 (3,91)*	91,50 (1,72)*
Blood	77,17 (2,42)	<b>77,79 (2,64)</b>	77,75 (1,59)	76,06 (0,23)*	76,02 (0,95)*	76,22 (0,56)*
Bupa	68,49 (4,65)	<b>70,43 (5,42)</b>	68,76 (4,36)	61,47 (2,67)*	58,49 (6,21)*	66,43 (4,15)*
Cancer	<b>96,92 (0,89)</b>	95,59 (1,36)*	95,65 (1,48)*	95,23 (1,32)*	91,32 (1,84)*	96,78 (1,00)
Chess	95,41 (0,99)	94,34 (1,05)*	<b>99,00 (0,46)*</b>	73,49 (3,06)*	68,24 (1,31)*	97,08 (0,62)*
CMC	<b>50,57 (3,43)</b>	42,65 (0,48)*	48,19 (2,96)*	42,74 (0,22)*	44,31 (1,66)*	49,68 (3,03)
Contractions	84,31 (6,06)	83,33 (6,29)	81,67 (6,62)*	<b>85,14 (6,06)</b>	82,92 (7,45)	81,25 (7,24)*
Flare	70,54 (2,88)	52,01 (0,15)*	71,72 (1,89)	72,25 (2,24)*	51,91 (0,52)*	<b>73,89 (1,75)*</b>
German	73,27 (1,63)	73,52 (2,55)	70,48 (2,71)*	70,00 (0,00)*	69,12 (1,84)*	<b>74,03 (2,18)</b>
Glass	62,45 (5,70)	43,52 (2,37)*	<b>65,91 (5,10)*</b>	59,56 (5,38)	45,85 (3,94)*	62,58 (6,05)
Haberman	72,46 (3,64)	72,32 (4,45)	71,40 (4,85)	72,81 (2,33)	72,11 (3,55)	<b>73,25 (1,77)</b>
Heart	<b>82,39 (4,51)</b>	80,70 (4,18)	80,20 (3,50)*	80,85 (5,05)*	72,39 (5,48)*	82,39 (4,28)
Hepatitis	82,72 (5,64)	81,75 (5,12)	80,96 (7,20)	79,82 (2,22)*	79,47 (4,71)*	<b>84,47 (5,09)*</b>
Horse	<b>84,13 (2,96)</b>	80,72 (4,02)*	81,78 (3,40)*	66,16 (3,29)*	82,25 (3,43)*	81,23 (4,37)*
ILPD	70,67 (2,81)	70,44 (2,74)	68,53 (3,30)*	71,72 (0,00)	71,10 (1,45)	<b>71,72 (0,00)</b>
Ionosphere	90,42 (2,65)	<b>91,69 (2,46)*</b>	87,93 (3,43)*	84,48 (2,39)*	81,30 (3,66)*	87,89 (2,89)*
Iris	94,77 (2,74)	94,32 (2,96)	95,50 (3,97)	89,28 (11,54)*	64,68 (2,55)*	<b>96,31 (3,58)*</b>
Laryngeal1	82,26 (4,74)	80,69 (4,64)*	77,42 (4,25)*	<b>83,02 (5,00)</b>	81,32 (4,04)	81,32 (4,74)
Laryngeal3	<b>74,05 (3,76)</b>	67,92 (3,36)*	69,55 (3,97)*	70,19 (2,04)*	68,26 (3,69)*	72,61 (3,78)*
Lymph	80,36 (4,65)	74,50 (7,60)*	83,78 (5,41)*	75,59 (7,66)*	72,07 (5,95)*	<b>86,76 (4,99)*</b>
Mammo	81,88 (1,92)	<b>82,33 (2,07)</b>	80,15 (2,61)*	81,42 (1,58)	80,92 (1,95)*	82,06 (1,88)
Phoneme	<b>84,94 (1,13)</b>	79,52 (0,88)*	84,36 (1,29)*	76,86 (2,73)*	76,30 (0,96)*	77,20 (0,89)*
Segment	94,12 (0,92)	28,51 (0,16)*	<b>96,01 (0,82)*</b>	67,39 (3,65)*	28,37 (0,19)*	95,04 (0,65)*
Sonar	79,36 (5,30)	78,08 (5,24)	<b>81,09 (5,56)</b>	74,94 (5,16)*	69,23 (6,55)*	75,38 (5,50)*
Thyroid	93,96 (3,19)	93,71 (2,82)	91,13 (6,39)*	82,26 (4,77)*	79,69 (2,92)*	<b>95,41 (2,83)*</b>
Waveform	82,06 (0,89)	67,79 (4,66)*	83,70 (0,82)*	70,47 (1,38)*	56,58 (1,16)*	<b>85,86 (0,96)*</b>
Weaning	84,31 (3,41)	<b>85,91 (3,33)*</b>	81,78 (4,42)*	83,73 (3,87)	78,44 (3,38)*	80,13 (4,48)*
Wine	96,14 (2,87)	88,11 (5,05)*	<b>97,05 (2,47)</b>	93,94 (4,15)*	65,61 (4,12)*	96,89 (2,11)
Derrotas	-	14	15	21	24	10
Empates	-	14	9	8	6	10
Vitórias	-	2	6	1	0	10

Conforme se observa na figura 16, o MP foi estatisticamente equivalente ao MLP, RF e SVM. Isso sugere que pode estar entre os melhores classificadores da literatura, uma vez que estes três apresentam uma das melhores performances na análise conduzida por (FERNÁNDEZ-DELGADO et al., 2014). É importante salientar que o MP obteve vantagem em termos de derrotas, empates e vitórias quando comparado com o MLP e RF. No entanto, em geral, não demonstrou ser tão vantajoso em relação ao SVM. Contudo, foi vitorioso ou equivalente em 2/3 das bases com relação ao SVM. Finalmente, a tabela 8 apresenta o tempo gasto para rotular  $\zeta$ . Observa-se que o MP é o mais custoso, podendo isso ser um gargalo para a sua aplicação em problemas que necessitem de respostas rápidas.

Tabela 8 – Comparação do tempo de classificação, em nanosegundos, num processador Core i7 de 3,6GHz do conjunto de generalização  $\zeta$

Problema	MP	ADBO	MLP	RF	SB	SVM
Abalone	6,45E+09 (1,78E+08)	3,76E+05 (2,59E+04)	1,05E+07 (2,49E+05)	6,83E+06 (5,94E+05)	3,42E+05 (4,46E+05)	4,23E+05 (4,77E+04)
Australian	7,28E+08 (3,79E+07)	9,98E+04 (2,76E+03)	1,35E+06 (3,80E+05)	3,22E+05 (9,01E+03)	4,52E+04 (9,21E+03)	1,25E+05 (1,84E+04)
Balance	2,16E+08 (3,86E+06)	6,01E+04 (1,79E+03)	5,55E+05 (3,18E+03)	1,58E+05 (3,24E+04)	3,22E+04 (9,67E+02)	7,00E+04 (1,25E+04)
Blood	2,85E+08 (9,05E+06)	9,63E+04 (2,13E+04)	3,45E+05 (2,60E+03)	3,27E+05 (1,39E+04)	3,51E+04 (7,57E+03)	7,54E+04 (8,40E+03)
Bupa	1,06E+08 (3,30E+07)	9,72E+04 (9,24E+03)	6,27E+05 (3,19E+04)	2,25E+05 (7,30E+03)	3,15E+04 (4,94E+03)	3,57E+04 (8,52E+03)
Cancer	3,10E+08 (1,11E+07)	2,13E+05 (5,05E+04)	1,60E+06 (6,58E+04)	1,43E+06 (2,77E+05)	3,05E+04 (9,50E+02)	7,06E+04 (2,39E+03)
Chess	5,39E+10 (4,62E+08)	8,68E+05 (7,68E+04)	1,22E+08 (9,49E+06)	4,30E+06 (3,41E+05)	7,34E+05 (1,06E+05)	2,07E+06 (7,56E+05)
CMC	1,67E+09 (2,73E+07)	8,80E+04 (2,23E+04)	5,87E+06 (9,68E+04)	2,25E+06 (1,57E+05)	9,47E+04 (1,88E+04)	2,10E+05 (6,64E+03)
Contractions	3,23E+07 (1,77E+07)	1,72E+04 (1,56E+03)	4,51E+05 (2,98E+04)	4,34E+04 (1,04E+03)	7,60E+03 (3,38E+03)	1,55E+04 (6,64E+02)
Flare	1,52E+09 (3,15E+07)	8,34E+04 (1,95E+04)	4,33E+06 (2,94E+05)	3,42E+06 (3,71E+05)	9,42E+04 (9,65E+03)	2,93E+05 (3,14E+04)
German	2,12E+09 (2,13E+07)	2,99E+05 (1,37E+04)	2,05E+06 (4,10E+05)	4,67E+05 (2,15E+04)	7,58E+04 (1,59E+04)	2,03E+05 (7,90E+03)
Glass	5,35E+07 (2,35E+06)	1,21E+04 (3,76E+02)	6,34E+05 (2,25E+03)	4,45E+05 (1,34E+05)	1,46E+04 (1,91E+03)	4,19E+04 (5,57E+03)
Haberman	7,82E+07 (2,17E+07)	3,91E+04 (1,45E+03)	5,27E+05 (1,41E+04)	1,94E+05 (1,74E+04)	2,61E+04 (4,11E+03)	3,16E+04 (9,98E+03)
Heart	7,97E+07 (1,68E+06)	7,75E+04 (1,13E+04)	4,84E+05 (1,23E+04)	2,53E+05 (4,41E+03)	1,40E+04 (1,83E+03)	2,92E+04 (1,67E+03)
Hepatitis	3,95E+07 (1,43E+06)	7,21E+04 (3,05E+03)	3,55E+05 (3,27E+04)	1,20E+05 (1,06E+04)	8,85E+03 (7,04E+02)	2,05E+04 (1,42E+03)
Horse	3,85E+08 (6,08E+06)	1,77E+05 (4,23E+04)	2,87E+06 (1,48E+05)	6,76E+05 (6,83E+04)	2,98E+04 (8,46E+03)	8,03E+04 (5,57E+03)
ILPD	2,42E+08 (6,89E+06)	1,22E+05 (1,47E+04)	9,72E+05 (1,57E+05)	1,36E+05 (1,15E+04)	2,90E+04 (5,44E+03)	5,70E+04 (2,32E+03)
Ionosphere	1,87E+08 (9,50E+06)	1,10E+05 (5,53E+03)	1,17E+06 (2,08E+05)	3,27E+05 (2,32E+04)	2,18E+04 (5,92E+03)	5,31E+04 (1,02E+04)
Iris	3,17E+07 (1,49E+07)	2,51E+04 (3,67E+03)	1,12E+05 (1,37E+04)	3,67E+04 (6,97E+02)	1,39E+04 (7,65E+02)	3,50E+04 (9,58E+03)
Laryngeal1	5,88E+07 (1,02E+06)	4,67E+04 (2,24E+03)	9,94E+04 (3,62E+03)	4,43E+05 (3,30E+04)	1,24E+04 (3,75E+03)	2,55E+04 (3,07E+03)
Laryngeal3	1,36E+08 (3,77E+06)	3,75E+04 (6,38E+03)	4,02E+05 (5,04E+04)	8,93E+04 (4,79E+03)	2,05E+04 (1,96E+03)	4,57E+04 (5,79E+03)
Lymph	5,33E+07 (9,69E+05)	2,41E+04 (3,39E+03)	4,86E+05 (1,39E+04)	7,69E+04 (3,29E+03)	1,20E+04 (4,62E+03)	3,47E+04 (1,06E+03)
Mammo	4,53E+08 (1,58E+07)	1,82E+05 (1,58E+04)	1,43E+06 (1,60E+04)	1,10E+06 (9,31E+04)	4,35E+04 (7,15E+03)	9,03E+04 (1,66E+03)
Phoneme	9,70E+09 (1,80E+08)	9,94E+05 (2,53E+04)	1,14E+07 (1,84E+05)	7,90E+06 (1,15E+06)	2,51E+05 (3,13E+04)	4,49E+05 (2,26E+04)
Segment	3,24E+09 (1,20E+08)	1,29E+05 (2,18E+04)	7,93E+06 (3,29E+05)	3,47E+06 (4,01E+05)	1,68E+05 (1,75E+04)	5,65E+05 (3,14E+04)
Sonar	1,27E+08 (3,79E+06)	4,66E+04 (1,68E+03)	4,11E+05 (5,48E+04)	1,07E+05 (1,65E+04)	1,60E+04 (5,30E+03)	4,42E+04 (8,55E+02)
Thyroid	5,15E+07 (2,08E+07)	4,98E+04 (4,74E+03)	1,99E+05 (1,29E+04)	3,70E+05 (1,45E+04)	1,18E+04 (2,66E+03)	4,75E+04 (1,11E+04)
Waveform	2,84E+10 (4,40E+08)	6,39E+05 (2,17E+05)	1,44E+07 (5,73E+05)	4,96E+06 (3,26E+05)	3,95E+05 (2,30E+05)	8,26E+05 (3,21E+04)
Weaning	1,26E+08 (4,53E+07)	7,57E+04 (3,16E+04)	9,12E+05 (2,45E+04)	2,65E+05 (2,16E+05)	3,42E+04 (1,42E+04)	3,56E+04 (6,70E+03)
Wine	4,32E+07 (1,16E+07)	2,19E+04 (2,07E+03)	8,17E+04 (3,54E+03)	4,92E+04 (7,78E+03)	1,04E+04 (4,52E+02)	3,39E+04 (8,87E+03)
Média	3,69E+09	1,73E+05	6,48E+06	1,36E+06	8,85E+04	2,04E+05

## 6.4 Comparações com Trabalhos Relacionados

A tabela 9 apresenta resultados da acuracidade do MP e de alguns trabalhos relacionados: MDT (TODOROVSKI; DŽEROSKI, 2003), DIVACE (CHANDRA; YAO, 2004), DES-FA (CRUZ; CAVALCANTI; REN, 2011), GBDM (NETO; CANUTO; LUDERMIR, 2013), GDACE (GU; JIN, 2014), Meta-DES (CRUZ et al., 2015), AGLP (ALMEIDA; GALVÃO, 2016) e DSCO (BRUN et al., 2016). Este tipo de comparação deve ser feita com cautela, pois foram consideradas diferentes configurações nos experimentos e diferentes algoritmos de aprendizagem. Portanto, os valores que estão destacados em negrito apenas indicam a maior acuracidade alcançada para cada um dos problemas.

Tabela 9 – Comparação do MP com Trabalhos Relacionados

Problema	MP	MDT	DIVACE	DES-FA	GBDM	GDACE	Meta-DES	AGLP	DSOC
Australian	84,92	85,67	<b>86,20</b>	-	-	-	-	-	-
Balance	88,01	<b>92,84</b>	-	-	77,00	-	-	-	-
Blood	77,17	-	-	73,40	66,74	-	<b>79,14</b>	-	76,40
Cancer	96,92	<b>97,37</b>	-	-	-	-	-	95,99	-
Chess	95,41	<b>99,51</b>	-	-	-	-	-	-	-
German	73,27	75,06	-	74,00	75,76	<b>77,80</b>	75,55	-	72,80
Glass	62,45	68,46	-	55,32	<b>81,96</b>	-	66,87	64,96	63,10
Haberman	72,46	-	-	72,36	-	-	<b>76,71</b>	-	76,40
Heart	82,39	83,83	-	83,82	-	81,08	<b>84,80</b>	77,39	82,10
Hepatitis	82,39	<b>83,11</b>	-	-	-	-	-	-	-
Horse	84,13	-	-	-	<b>95,82</b>	-	-	62,95	-
ILPD	<b>70,67</b>	-	-	-	-	-	69,40	-	66,66
Ionosphere	90,42	90,45	-	88,63	<b>95,04</b>	91,43	89,94	-	86,90
Iris	94,77	-	-	-	-	-	-	-	-
Laryngeal 1	82,26	-	-	77,35	-	-	79,67	-	<b>82,40</b>
Laryngeal 3	<b>74,05</b>	-	-	72,03	-	-	72,65	-	67,70
Mammo	81,88	-	-	80,28	-	-	<b>84,82</b>	-	-
Phoneme	<b>84,94</b>	-	-	79,06	-	-	80,35	-	80,60
Segment	94,12	<b>97,69</b>	-	88,74	-	-	-	-	70,30
Sonar	79,36	-	-	78,52	<b>92,79</b>	-	80,55	-	71,00
Waveform	82,06	<b>86,06</b>	-	-	-	-	-	-	-
Weaning	86,80	-	-	82,89	-	-	<b>87,15</b>	-	82,90
Wine	96,14	98,02	-	95,55	<b>100,00</b>	-	99,25	-	69,40

Na grande maioria dos problemas, o MP, MDT ou Meta-DES alcançaram o melhor resultado de acuracidade. O que há em comum entre esses três trabalhos é o fato de que ambos empregam uma etapa de aprendizagem adicional. Neste sentido, há fortes indícios de que o uso das técnicas de meta-aprendizagem são bastante promissoras em sistemas de múltiplos classificadores. Isso pode ser explicado pelo fato do mecanismo seletor poder se adaptar às características de cada problema de classificação (desbalanceamento, ruído, etc). Outro trabalho que também merece grande destaque pelos resultados alcançados é o GBDM, que da mesma forma que o MP, emprega de maneira explícita medidas de diversidade para guiar o processo de construção dos componentes do *ensemble*.

## 6.5 Considerações Finais

Neste capítulo foram apresentados resultados de experimentos e análises realizadas com a intenção de validar as estratégias que foram propostas para geração, seleção e combinação de múltiplos classificadores. Inicialmente foi avaliado o tamanho do *ensemble*, que é um dos principais responsáveis pelo custo computacional. Em comparação com classificadores gerados por *Bagging*, as medidas de diversidade analisadas foram, na grande maioria das vezes, estatisticamente superior na abordagem de geração proposta. Esse fato é de grande importância para enfatizar o processo de otimização explícita da diversidade durante a geração dos componentes. Em seguida, visando avaliar a acuracidade, foram realizadas comparações com técnicas de seleção dinâmica de classificadores, técnicas estáticas e trabalhos relacionados. O desempenho do sistema apresentou performance satisfatória em todas as comparações, sendo inclusive considerado estatisticamente superior na maioria das vezes. Certamente que o uso de uma base de casos contribui para que isso fosse possível, pois resolver problemas por analogias com situações similares parece ser mais racional e promissor do que o uso de regras fixas e genéricas.

## 7 CONCLUSÃO E TRABALHOS FUTUROS

Nesta tese de doutorado foi apresentado o desenvolvimento de um sistema híbrido inteligente voltado para a geração, seleção e combinação de múltiplos classificadores. Os resultados alcançados são bastante promissores e sugerem que o mesmo é robusto, visto que consegue alcançar taxas de reconhecimento superiores à estratégias já estabelecidas na literatura. Contudo, ainda possui alguns pontos que necessitam ser levados em consideração. Uma breve síntese do sistema proposto é descrita na próxima seção, enquanto possíveis sugestões de continuidade do mesmo encontram-se na seção final do capítulo.

### 7.1 Síntese do Trabalho

Apesar da existência de inúmeras pesquisas na área de sistemas de múltiplos classificadores, pouca ênfase tem sido dada ao problema como um todo. Os principais trabalhos focam apenas em partes do processo: geração (CHANDRA; YAO, 2004; SANTOS; SABOURIN; MAUPIN, 2008; KO; SABOURIN; BRITTO, 2009; NETO; CANUTO; LUDERMIR, 2013; GU; JIN, 2014), seleção (SABOURIN et al., 1993; WOODS; KEGELMEYER W.P.; BOWYER, 1997; GIACINTO; ROLI, 2001; SANTANA et al., 2006; KUNCHEVA; RODRIGUEZ, 2007; KO; SABOURIN; BRITTO, 2007; CRUZ; CAVALCANTI; REN, 2011; CRUZ et al., 2015), combinação (WOLPERT, 1992; TODOROVSKI; DŽEROSKI, 2003).

O sistema proposto nesta tese contempla todas as fases do processo. Na etapa de geração, realizada por meio da evolução dos componentes, a simplicidade é o grande diferencial. Não existem processos complexos de cruzamento e/ou mutação, como ocorre na maioria dos Algoritmos Evolucionários. Por consequência disso, o número de parâmetros a ser configurado/otimizado é consideravelmente menor. Nas etapas de seleção e combinação, emprega-se o uso de uma base com casos para estimação da competência dos componentes do *ensemble*. O valor da competência é utilizado tanto no processo de seleção quanto na combinação das respostas. A principal vantagem dessa estratégia é a possibilidade da utilização de múltiplos critérios/características na avaliação dos componentes.

Finalmente, com a validação do método proposto, através dos experimentos e resultados que foram apresentados no capítulo 6, nota-se que o mesmo é robusto. O sistema conseguiu alcançar taxas de acuracidade superiores à estratégias estabelecidas na literatura em diversos cenários. Esse fato sugere que existe grande potencial para continuidade não apenas deste trabalho, mas também em novas pesquisas na área que tratem do uso de técnicas de aprendizagem na otimização de *ensembles*.

## 7.2 Sugestões para Continuidade

O principal problema do sistema proposto nesse trabalho é o custo computacional. Seu uso pode se tornar inviável em aplicações que possuem grandes bases de dados e que necessitem de respostas instantâneas. No maior problema analisado (Waveform), por exemplo, o sistema gasta aproximadamente 95 milissegundos para classificar um padrão utilizando um processador *Core i7* de 3,6GHz. Os classificadores monolíticos gastam menos de 1 milissegundo pra realizar a mesma tarefa. Com o intuito de contornar esse problema, a lista a seguir apresenta algumas sugestões que podem diminuir o tempo de resposta.

- Foi utilizado um conjunto fixo com 5 características para estimar a competência dos classificadores. Diferentes combinações com um número menor de características e/ou o uso de outras características podem diminuir consideravelmente o custo computacional, bem como melhorar a taxa de acerto da competência dos seletores.
- Um número menor de componentes pode ser definido através de um processo de poda, por exemplo, diminuindo conseqüentemente o tempo necessário para que o *ensemble* possa classificar o padrão. Em alguns dos problemas analisados, como se pode observar nas tabelas 2 e 3, apenas 5 componentes já são suficientes.
- Por utilizar uma topologia paralela, o emprego de *threads* tem a possibilidade de diminuir o tempo gasto para a classificação de um dado padrão de consulta. O tempo poderia ser reduzido à metade, por exemplo, com o uso de apenas duas *threads*.
- A hibridização da topologia paralela com a topologia serial, de tal forma que o *ensemble* fosse considerado apenas para rotular os padrões mais difíceis<sup>1</sup>. Os padrões fáceis poderiam ser rotulados por um classificador monolítico.

---

<sup>1</sup> São os padrões que normalmente se encontram nas fronteiras entre classes

# REFERÊNCIAS

- ALCALÁ-FDEZ, J. et al. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Multiple-Valued Logic and Soft Computing*, v. 17, n. 2-3, p. 255–287, 2011. Disponível em: <<http://dblp.uni-trier.de/db/journals/mvl/mvl17.html#Alcala-FdezFLDG11>>. 52
- ALMEIDA, L. M.; GALVÃO, P. S. Ensembles with clustering-and-selection model using evolutionary algorithms. In: *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.: s.n.], 2016. p. 444–449. 30, 64
- ALMEIDA, P. R. L. de et al. Music genre classification using dynamic selection of ensemble of classifiers. In: *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.: s.n.], 2012. p. 2700–2705. ISSN 1062-922X. 38
- AYAD, O.; SYED-MOUCHAWEH, M. Multiple classifiers approach based on dynamic selection to maximize classification performance. *International Journal of Machine Learning and Computing*, v. 1, n. 2, p. 154, 2011. Disponível em: <<https://search.proquest.com/docview/1441451399?accountid=149052>>. 31
- BACHE, K.; LICHMAN, M. *UCI Machine Learning Repository*. 2013. Disponível em: <<http://archive.ics.uci.edu/ml>>. 52
- BASHBAGHI, S. et al. Dynamic selection of exemplar-svms for watch-list screening through domain adaptation. In: *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. [S.l.: s.n.], 2017. p. 738–745. 38
- BAUER, E.; KOHAVI, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, v. 36, n. 1, p. 105–139, Jul 1999. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1023/A:1007515423169>>. 20
- BENEDIKTSSON, J. A.; SWAIN, P. H. Consensus theoretic classification methods. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 22, n. 4, p. 688–704, Jul 1992. ISSN 0018-9472. 42
- BREIMAN, L. Bagging predictors. *Machine Learning*, v. 24, n. 2, p. 123–140, 1996. 26
- BRITTO, A. S.; SABOURIN, R.; OLIVEIRA, L. E. Dynamic selection of classifiers - a comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665–3680, 2014. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320314001885>>. 32
- BROWN, G. et al. Diversity creation methods: a survey and categorisation. *Information Fusion*, v. 6, n. 1, p. 5 – 20, 2005. ISSN 1566-2535. Diversity in Multiple Classifier Systems. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1566253504000375>>. 21, 26
- BRUN, A. L. et al. Contribution of data complexity features on dynamic classifier selection. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2016. p. 4396–4403. 38, 64

- CHANDRA, A.; YAO, X. Divace: Diverse and accurate ensemble learning algorithm. In: \_\_\_\_\_. *Intelligent Data Engineering and Automated Learning – IDEAL 2004: 5th International Conference, Exeter, UK. August 25-27, 2004. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 619–625. ISBN 978-3-540-28651-6. Disponível em: <[https://doi.org/10.1007/978-3-540-28651-6\\_91](https://doi.org/10.1007/978-3-540-28651-6_91)>. 46, 64, 66
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, v. 41, n. Supplement C, p. 195 – 216, 2018. ISSN 1566-2535. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1566253517304074>>. 21, 32
- CRUZ, R. M. et al. Meta-des. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 48, n. 5, p. 1925–1935, maio 2015. ISSN 0031-3203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2014.12.003>>. 37, 47, 51, 64, 66
- CRUZ, R. M. O.; CAVALCANTI, G. D. C.; REN, T. I. A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. In: *The 2011 International Joint Conference on Neural Networks*. [S.l.: s.n.], 2011. p. 1126–1133. ISSN 2161-4393. 37, 49, 56, 64, 66
- DIETTERICH, T. G. Ensemble methods in machine learning. In: *Proceedings of the First International Workshop on Multiple Classifier Systems*. London, UK, UK: Springer-Verlag, 2000. (MCS '00), p. 1–15. ISBN 3-540-67704-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=648054.743935>>. 10, 19
- DIETTERICH, T. G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, v. 40, n. 2, p. 139–157, Aug 2000. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1007607513941>>. 28
- DUFOURQ, E.; PILLAY, N. Hybridizing evolutionary algorithms for creating classifier ensembles. In: *2014 Sixth World Congress on Nature and Biologically Inspired Computing (NaBIC 2014)*. [S.l.: s.n.], 2014. p. 84–90. 30
- EFRON, B.; TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. [S.l.]: Chapman & Hall, 1993. 27
- EIBEN, A. E.; SMITH, J. E. *Introduction to Evolutionary Computing*. 2nd. ed. [S.l.]: Springer Publishing Company, Incorporated, 2015. ISBN 3662448734, 9783662448731. 30
- FERNÁNDEZ-DELGADO, M. et al. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, v. 15, p. 3133–3181, 2014. Disponível em: <<http://jmlr.org/papers/v15/delgado14a.html>>. 61, 62
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Statist.*, The Institute of Mathematical Statistics, v. 28, n. 2, p. 337–407, 2000. Disponível em: <<http://dx.doi.org/10.1214/aos/1016218223>>. 28
- GALAR, M. et al. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, v. 42, n. 4, p. 463–484, July 2012. ISSN 1094-6977. 38

- GAMA, J.; BRAZDIL, P. Cascade generalization. *Machine Learning*, v. 41, n. 3, p. 315–343, Dec 2000. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1007652114878>>. 43
- GIACINTO, G.; ROLI, F. Methods for dynamic classifier selection. In: . [S.l.]: Lecture, 1999. p. 177–189. 33, 34
- GIACINTO, G.; ROLI, F. Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, v. 34, p. 1879–1881, 2001. 34, 66
- GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201157675. 30
- GOONATILAKE, S.; KHEBBAL, S. (Ed.). *Intelligent Hybrid Systems*. 1st. ed. New York, NY, USA: John Wiley & Sons, Inc., 1994. ISBN 0471942421. 44
- GU, S.; JIN, Y. Generating diverse and accurate classifier ensembles using multi-objective optimization. In: *2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM)*. [S.l.: s.n.], 2014. p. 9–15. 46, 64, 66
- HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 12, n. 10, p. 993–1001, out. 1990. ISSN 0162-8828. Disponível em: <<http://dx.doi.org/10.1109/34.58871>>. 21
- HO, T. K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 8, p. 832–844, 1998. ISSN 0162-8828. Disponível em: <<http://citeseer.ist.psu.edu/ho98random.html>>. 26
- HUANG, Y.; SUEN, C. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 17, n. 1, p. 90–94, Jan 1995. ISSN 0162-8828. 34
- JACOBS, R. A. Methods for combining experts' probability assessments. *Neural Comput.*, MIT Press, v. 7, n. 5, p. 867–888, set. 1995. ISSN 0899-7667. Disponível em: <<http://dx.doi.org/10.1162/neco.1995.7.5.867>>. 25
- KING, R. D.; FENG, C.; SUTHERLAND, A. *StatLog: Comparison of Classification Algorithms on Large Real-World Problems*. 1995. 52
- KO, A. H.; SABOURIN, R.; BRITTO, J. A. S. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, v. 41, n. 5, p. 1718 – 1731, 2008. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320307004499>>. 31, 49, 56
- KO, A. H.-R.; SABOURIN, R.; BRITTO, A. D. S. Compound diversity functions for ensemble selection. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 23, n. 04, p. 659–686, 2009. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S021800140900734X>>. 46, 66
- KO, A. H.-R.; SABOURIN, R.; BRITTO, A. de S. K-nearest oracle for dynamic ensemble selection. In: *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*. [S.l.: s.n.], 2007. v. 1, p. 422–426. ISSN 1520-5363. 34, 37, 66

- KRAWCZYK, B. et al. Ensemble learning for data stream analysis: A survey. *Information Fusion*, v. 37, n. Supplement C, p. 132 – 156, 2017. ISSN 1566-2535. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1566253516302329>>. 38
- KUNCHEVA, L. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004. ISBN 9780471660255. Disponível em: <<https://books.google.ca/books?id=B4TZtp7X82oC>>. 25, 39, 42
- KUNCHEVA, L. *LKC Ludmila kuncheva collection*. 2004. Disponível em: <[http://pages.bangor.ac.uk/~mas00a/activities/real\\_data.html](http://pages.bangor.ac.uk/~mas00a/activities/real_data.html)>. 52
- KUNCHEVA, L. I. Clustering-and-selection model for classifier combination. In: *Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on*. [S.l.: s.n.], 2000. v. 1, p. 185–188 vol.1. 10, 31, 33, 49
- KUNCHEVA, L. I.; RODRIGUEZ, J. J. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering*, v. 19, n. 4, p. 500–508, April 2007. ISSN 1041-4347. 36, 66
- KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 51, n. 2, p. 181–207, maio 2003. ISSN 0885-6125. Disponível em: <<http://dx.doi.org/10.1023/A:1022859003006>>. 26
- LACY, S. E.; LONES, M. A.; SMITH, S. L. Forming classifier ensembles with multimodal evolutionary algorithms. In: *2015 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2015. p. 723–729. ISSN 1089-778X. 30
- LESSMANN, S. et al. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, v. 247, n. 1, p. 124 – 136, 2015. ISSN 0377-2217. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0377221715004208>>. 38
- MACLIN, R.; OPITZ, D. W. Popular ensemble methods: An empirical study. *CoRR*, abs/1106.0257, 2011. Disponível em: <<http://arxiv.org/abs/1106.0257>>. 20
- MARLER, R.; ARORA, J. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, v. 26, p. 369–395, 2004. 30
- NETO, A. A. F.; CANUTO, A. M. P.; LUDERMIR, T. B. Using good and bad diversity measures in the design of ensemble systems: A genetic algorithm approach. In: *2013 IEEE Congress on Evolutionary Computation*. [S.l.: s.n.], 2013. p. 789–796. ISSN 1089-778X. 46, 64, 66
- NUCCI, D. D. et al. Dynamic selection of classifiers in bug prediction: An adaptive method. *IEEE Transactions on Emerging Topics in Computational Intelligence*, v. 1, n. 3, p. 202–212, June 2017. 38
- PANICHELLA, A.; OLIVETO, R.; LUCIA, A. D. Cross-project defect prediction models: L’union fait la force. In: *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*. [S.l.: s.n.], 2014. p. 164–173. 38

- PANOV, P.; DZEROSKI, S. Combining bagging and random subspaces to create better ensembles. In: *Advances in Intelligent Data Analysis VII, 7th International Symposium on Intelligent Data Analysis, IDA 2007, Ljubljana, Slovenia, September 6-8, 2007, Proceedings*. [s.n.], 2007. p. 118–129. Disponível em: <[https://doi.org/10.1007/978-3-540-74825-0\\_11](https://doi.org/10.1007/978-3-540-74825-0_11)>. 27
- PONTI, M. Combining classifiers: From the creation of ensembles to the decision fusion. In: *Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2011 24th SIBGRAPI Conference on*. [S.l.: s.n.], 2011. p. 1–10. 20, 29, 39
- PORCEL, C. et al. A hybrid recommender system for the selective dissemination of research resources in a technology transfer office. *Information Sciences*, v. 184, n. 1, p. 1 – 19, 2012. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025511004452>>. 38
- ROSALES-PEREZ, A. et al. An evolutionary multi-objective model and instance selection for support vector machines with pareto-based ensembles. *IEEE Transactions on Evolutionary Computation*, PP, n. 99, p. 1–1, 2017. ISSN 1089-778X. 30
- SABOURIN, M. et al. Classifier combination for hand-printed digit recognition. In: *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*. [S.l.: s.n.], 1993. p. 163–166. 32, 49, 66
- Decision stump. In: SAMMUT, C.; WEBB, G. I. (Ed.). *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2010. p. 262–263. ISBN 978-0-387-30164-8. Disponível em: <[https://doi.org/10.1007/978-0-387-30164-8\\_202](https://doi.org/10.1007/978-0-387-30164-8_202)>. 56
- SANTANA, A. et al. A dynamic classifier selection method to build ensembles using accuracy and diversity. In: *Neural Networks, 2006. SBRN '06. Ninth Brazilian Symposium on*. [S.l.: s.n.], 2006. p. 36–41. 35, 66
- SANTOS, E. dos; SABOURIN, R.; MAUPIN, P. Ambiguity-guided dynamic selection of ensemble of classifiers. In: *Information Fusion, 2007 10th International Conference on*. [S.l.: s.n.], 2007. p. 1–8. 36, 49
- SANTOS, E. M. D.; SABOURIN, R.; MAUPIN, P. A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition*, v. 41, n. 10, p. 2993 – 3009, 2008. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S003132030800126X>>. 36, 46, 66
- SCHAPIRE, R. E. The strength of weak learnability. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 5, n. 2, p. 197–227, jul. 1990. ISSN 0885-6125. Disponível em: <<http://dx.doi.org/10.1023/A:1022648800760>>. 26
- SHARKEY, A. J. C.; SHARKEY, N. E. Combining diverse neural nets. *The Knowledge Engineering Review*, Cambridge University Press, v. 12, n. 3, p. 231–247, 1997. 25
- SOUZA, M. A. et al. On the characterization of the oracle for dynamic classifier selection. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2017. p. 332–339. 38
- TANG, E. K.; SUGANTHAN, P. N.; YAO, X. An analysis of diversity measures. *Machine Learning*, v. 65, n. 1, p. 247–271, Oct 2006. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/s10994-006-9449-2>>. 35

- TODOROVSKI, L.; DŽEROSKI, S. Combining classifiers with meta decision trees. *Machine Learning*, v. 50, n. 3, p. 223–249, Mar 2003. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1021709817809>>. 43, 47, 51, 64, 66
- TSOUMAKAS, G.; PARTALAS, I.; VLAHAVAS, I. A taxonomy and short review of ensemble selection. In: *ECAI 08, Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications, SUEMA*. [S.l.: s.n.], 2008. 31
- WOLPERT, D. H. Stacked generalization. *Neural Networks*, v. 5, p. 241–259, 1992. 47, 66
- WOODS, K.; KEGELMEYER W.P., J.; BOWYER, K. Combination of multiple classifiers using local accuracy estimates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 19, n. 4, p. 405–410, Apr 1997. ISSN 0162-8828. 32, 33, 66
- XIAO, H.; XIAO, Z.; WANG, Y. Ensemble classification based on supervised clustering for credit scoring. *Applied Soft Computing*, v. 43, n. Supplement C, p. 73 – 86, 2016. ISSN 1568-4946. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1568494616300734>>. 38
- XU, L.; KRZYZAK, A.; SUEN, C. Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 22, n. 3, p. 418–435, maio 1992. ISSN 00189472. Disponível em: <<http://dx.doi.org/10.1109/21.155943>>. 20
- YAN, Y. et al. Sorting-based dynamic classifier ensemble selection. In: *2013 12th International Conference on Document Analysis and Recognition*. [S.l.: s.n.], 2013. p. 673–677. ISSN 1520-5363. 35
- ZHOU, Z.-H. *Ensemble Methods: Foundations and Algorithms*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2012. ISBN 1439830037, 9781439830031. 20