

Universidade Federal do Rio Grande do Norte
Centro de Ciências Exatas e da Terra
Departamento de Informática e Matemática Aplicada
Programa de Pós-Graduação em Sistemas e Computação

Dilema da Diversidade-Acurácia: Um estudo empírico no contexto de
multiclassificadores

Diogo Fagundes de Oliveira

Natal
Setembro de 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Universidade Federal do Rio Grande do Norte
Centro de Ciências Exatas e da Terra
Departamento de Informática e Matemática Aplicada
Programa de Pós-Graduação em Sistemas e Computação

Dilema da Diversidade-Acurácia: Um estudo empírico no contexto de
multiclassificadores

Dissertação de mestrado submetida ao Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte como parte dos requisitos para a obtenção do grau de Mestre em Sistemas e Computação (MSc.).

Diogo Fagundes de Oliveira

Prof. Dr. Marcílio Carlos Pereira de Souto
Orientador

Profa. Dra. Anne Magály de Paula Canuto
Orientadora

Natal, Setembro de 2008

Dilema da Diversidade-Acurácia: Um estudo empírico no contexto de multiclassificadores

Diogo Fagundes de Oliveira

Esta dissertação de mestrado foi avaliada e considerada aprovada pelo Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte.

Prof. Dr. Marcílio Carlos Pereira de Souto
Orientador

Profª. Dra. Anne Magály de Paula Canuto
Orientadora

Profª. Dra. Thaís de Vasconcelos Batista
Coordenadora do Programa de Pós-Graduação

Banca Examinadora:

Prof. Dr. Marcílio Carlos Pereira de Souto

Profª. Dra. Anne Magály de Paula Canuto

Prof. Dr. Adrião Duarte Doria Neto

Prof. Dr. Ivan Gesteira Costa Filho

Agradecimentos

Agradeço a Deus por ter me dado a oportunidade de iniciar e concluir o mestrado. Sou imensamente grato por todas as oportunidades que me foram dadas de crescimento intelectual e moral.

À minha família agradeço por todo auxílio prestado, pela compreensão e apoio em todos os momentos.

Agradeço aos meus orientadores, Marcílio Carlos Pereira Souto e Anne Magály de Paula Canuto, pelos conhecimentos transmitidos.

A minha namorada, Íria Caline Saraiva Cosme, que esteve comigo durante a escrita desta dissertação, me dando forças e sempre acreditando em mim.

Aos meus amigos que me incentivaram nos momentos difíceis.

A Lucas de Melo Oliveira e Raul Benites Paradedda por ter cedido o Weka modificado para que pudéssemos realizar os experimentos. A Antonio Jesus Nebro Urbaneja por ter tirado todas as dúvidas sobre o jMetal.

Agradeço ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) por financiar uma bolsa de desenvolvimento técnico-científico (DTI) no âmbito do Projeto Desenvolvimento de Sistemas Híbridos Inteligentes, coordenado pela professora Teresa Ludermir. Essa bolsa indispensável para a conclusão desta dissertação.

Resumo

Sistemas Multiclassificadores, também conhecidos como comitês de classificadores, têm sido amplamente utilizados para resolver os mais variados problemas, pois em geral têm melhores desempenhos que os classificadores base que formam esses sistemas. Para que isso ocorra, porém, é necessário que os classificadores base sejam tão acurados quanto diversos entre si – isso é conhecido como dilema da diversidade-acurácia. Dado a sua importância, alguns trabalhos sobre o estudo do comportamento dos comitês no contexto desse dilema foram propostos. Entretanto, a maioria dos trabalhos estudou tal problema para comitês homogêneos, ou seja, comitês formados apenas por classificadores do mesmo tipo. Sendo assim, motivado por esta limitação, esta dissertação, usando algoritmos genéticos, efetua um estudo mais detalhado sobre o dilema da diversidade-acurácia em comitês heterogêneos.

Área de Concentração: Inteligência Computacional.

Palavras-chave: Algoritmo Genético Multiobjetivo; Sistemas de Combinação de Classificadores; Comitês de Classificadores.

Abstract

Multi-classifier systems, also known as ensembles, have been widely used to solve several problems, because they, often, present better performance than the individual classifiers that form these systems. But, in order to do so, it's necessary that the base classifiers to be as accurate as diverse among themselves – this is also known as diversity/accuracy dilemma. Given its importance, some works have investigate the ensembles behavior in context of this dilemma. However, the majority of them address homogenous ensemble, i.e., ensembles composed only of the same type of classifiers. Thus, motivated by this limitation, this thesis, using genetic algorithms, performs a detailed study on the dilemma diversity/accuracy for heterogeneous ensembles.

Area of Concentration: Computational Intelligence

Keywords: Multi-objective Genetic Algorithm, Classifier Combination Systems; Ensembles;

Sumário

Introdução.....	13
1.1 Contexto Geral.....	13
1.2 Contexto Específico.....	16
1.3 Motivação e Objetivo.....	18
1.4 Organização do Trabalho.....	19
Construção de Multiclassificadores.....	21
2.1 Classificadores base.....	22
2.1.1 k -Vizinhos mais próximos.....	22
2.1.2 Rede Neural <i>Perceptron</i> de Multicamada com retropropagação.....	24
2.1.3 Máquinas de Vetores Suporte.....	26
2.2 Medidas de Diversidade.....	28
2.2.1 Estatística- Q.....	29
2.2.2 Coeficiente de Correlação ρ	29
2.2.3 Medida de Concordância k	29
2.3 Métodos de Escolha dos Classificadores Base.....	30
2.4 Métodos de Combinação.....	31
2.4.1 Soma.....	32
2.4.2 Classificador <i>Naive Bayes</i>	33
Abordagem Proposta.....	35
3.1 Algoritmos Genéticos.....	35
3.2 Trabalhos Relacionados.....	40
3.2.1 Algoritmos Genéticos Mono e Multi Objetivos.....	40
3.2.2 DIVACE-II.....	43
3.3 Nossa Abordagem.....	49
3.4 Considerações Finais.....	52
Experimentos.....	56
4.1 Base de Dados.....	56
4.1.1 <i>SatImage</i>	58
4.1.1.1 Splice.....	59
4.1.2. SCOP.....	60

4.2	Metodologia de desenvolvimento dos experimentos.....	61
	Resultados e Discussão.....	63
5.1	Classificadores base.....	63
5.2	Comitês.....	64
5.2.1	<i>SatImage</i>	65
5.2.2	<i>Splice</i>	71
5.2.3	SCOP.....	76
	Considerações Finais.....	82
6.1.	Trabalhos Futuros.....	83
	Bibliografia.....	85

Notação e Acrônimos

AG	Algoritmo Genético
AGMO	Algoritmos Genético Multiobjetivo
DIVACE	<i>Diverse and Accurate Ensemble Learning Algorithm</i>
k -NN	k -vizinhos mais próximos (do inglês, <i>k-nearest neighbour</i>)
MLP	<i>Perceptron</i> de Multicamada (do inglês, <i>multilayer neural networks</i>)
NSGA-II	<i>Non-dominated Sorting Genetic Algorithm II</i>
RBF	Função de Base Radial (do inglês, <i>radial basis function</i>)
SVM	Máquinas de Vetores Suporte (do inglês, <i>Support Vector Machines</i>)
f	função desejada
\mathcal{H}	espaço de hipóteses
D_i	classificador
\mathbf{x}	vetor de características (instância) de forma $\langle \quad , \quad , \dots , \quad \rangle$
x_{ij}	refere-se a j -ésima características de \mathbf{x}_i
Ω	conjunto de rótulos
	rótulo da classe

Lista de Figuras

Figura 1.1. Fundamento estatístico da razão de um multiclassificador poder ter melhor desempenho do que um classificador único.	14
Figura 1.2. Fundamento computacional do porque de um multiclassificador poder ter melhor desempenho do que um único classificador.	15
Figura 1.3. Fundamento representacional do porque de um multiclassificador poder ter melhor desempenho do que um único classificador.	15
Figura 2.1. Esquema de classificação pelo Método k -NN.	23
Figura 2.2. Exemplo de uma rede MLP.	24
Figura 2.3. Exemplo classificação de promotores por meio de uma SVM linear.	27
Figura 3.1. Possíveis tipos de soluções encontradas em um AGMO.	36
Figura 3.2. <i>Framework</i> DIVACE-II.	45
Figura 3.3. Fluxograma do DIVACE.	46
Figura 3.4. Representação genotípica de um indivíduo do DIVACE-II.	48
Figura 3.5. Representação cromossômica da nossa abordagem.	50
Figura 3.6. Ilustração dos operadores genéticos utilizados.	51

Lista de Tabelas

Tabela 4.1. Resumo das bases de dados.	57
Tabela 4. 2. Distribuição dos padrões da base de dados <i>SatImage</i>	58
Tabela 4. 3. Resumo do problema da base de dados <i>SatImage</i>	59
Tabela 4. 4. Distribuição dos padrões da base de dados <i>Splice</i>	59
Tabela 4.5. Resumo do problema da base de dados <i>Splice</i>	60
Tabela 4.6. Distribuição dos padrões da base de dados SCOP.	60
Tabela 4.7. Resumo do problema da base de dados <i>Splice</i>	61
Tabela 5.1. Acurácias dos classificadores base utilizado na <i>SatImage</i>	63
Tabela 5.2. Acurácias dos classificadores base utilizado na <i>Splice</i>	63
Tabela 5.3. Acurácias dos classificadores base utilizado na SCOP.	63
Tabela 5.4. Acurácias obtidas com a base de dados <i>SatImage</i>	66
Tabela 5.5. Diversidades obtidas com a base de dados <i>SatImage</i>	67
Tabela 5.6. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base <i>SatImage</i> , levando-se em consideração aos valores da acurácia.	68
Tabela 5.7. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base <i>SatImage</i> , levando-se em consideração aos valores da diversidade.	68
Tabela 5.8. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base <i>SatImage</i> , levando-se em consideração aos valores da acurácia.	69
Tabela 5.9. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base <i>SatImage</i> , levando-se em consideração aos valores da diversidade.	70
Tabela 5.10. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados <i>SatImage</i>	70
Tabela 5.11. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados <i>SatImage</i>	70
Tabela 5.12. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados <i>SatImage</i>	71
Tabela 5.13. Acurácias obtidas com a base de dados <i>Splice</i>	72
Tabela 5.14. Diversidades obtidas com a base de dados <i>Splice</i>	72

Tabela 5.15. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base <i>Splice</i> , levando-se em consideração a acurácia.	73
Tabela 5.16. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base <i>Splice</i> , levando-se em consideração a diversidade.	73
Tabela 5.17. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base <i>Splice</i> , levando-se em consideração a acurácia.	74
Tabela 5.18. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base <i>Splice</i> levando-se em consideração a diversidade.	75
Tabela 5.19. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados <i>Splice</i>	75
Tabela 5.20. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados <i>Splice</i>	75
Tabela 5.21. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados <i>Splice</i>	76
Tabela 5.22. Acurácias obtidas com a base de dados SCOP.	77
Tabela 5.23. Diversidades obtidas com a base de dados SCOP.	77
Tabela 5.24. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base SCOP, levando-se em consideração a acurácia.	78
Tabela 5.25. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base SCOP, levando-se em consideração a diversidade.	78
Tabela 5.26. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base SCOP, levando-se em consideração os valores da acurácia.	79
Tabela 5.27. <i>P-valores</i> encontrados ao analisar a relação entre os melhores e os demais da base SCOP, levando-se em consideração os valores da diversidade.	80
Tabela 5.28. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados SCOP.	80
Tabela 5.29. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados SCOP.	80

Tabela 5.30. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados SCOP.....	81
---	----

Capítulo 1

Introdução

1.1 Contexto Geral

A área de Aprendizado de Máquina é responsável pelo desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de adquirir conhecimento de forma automática (Monard & Baranauskas, 2000). De acordo com o conhecimento prévio sobre a classe das instâncias de treinamento, duas metodologias são abordadas para a realização do aprendizado – o aprendizado supervisionado e o aprendizado não-supervisionado (Duda, Hart, & Stork, 2001).

No contexto supervisionado, a entrada para o algoritmo de aprendizado (indutor) consiste em um conjunto de instâncias de treinamento da forma $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ para uma função desconhecida $y = f(\mathbf{x})$. Os valores \mathbf{x}_i , geralmente, são vetores de forma $\langle x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in} \rangle$ cujos componentes (características) assumem valores discretos ou reais, como altura, peso, cor, idade, sexo, entrou outros. A notação x_{ij} refere-se a j -ésima características de \mathbf{x}_i . Dado o conjunto de treinamento, o algoritmo induz uma hipótese h que deve aproximar a função desejada f , de tal modo que dado uma nova instância x , h prediz o valor y correspondente. Chama-se regressão no caso dos valores y pertencerem a um conjunto real. No caso dos valores y pertencerem a um domínio discreto, chama-se de classificação (Dietterich, 2000a). A tarefa de classificação é tratada nesta dissertação.

Dito de outra forma, um algoritmo de aprendizado pode ser visto como uma busca no espaço de hipóteses \mathcal{H} para identificar a melhor hipótese no espaço. O desempenho das hipóteses geradas, também chamadas de classificadores, depende de alguns fatores, como a quantidade de instâncias do conjunto de treinamento e o domínio no qual está sendo aplicado o indutor.

Quando a quantidade de dados de treinamento possuída é muito pequena em relação ao tamanho do espaço de hipótese, o algoritmo de aprendizado pode achar muitas hipóteses diferentes em \mathcal{H} , sendo todas com a mesma acurácia nos dados de treinamento. Escolher apenas um desses classificadores, podem nos levar a uma decisão inadequada. Muitos dos

algoritmos realizam buscas locais que podem ficar estagnadas em um ótimo local. Mesmo nos casos onde existam dados de treinamento suficientes, ainda assim, é muito difícil um indutor encontrar a melhor hipótese. Em muitas aplicações dos algoritmos de aprendizado, a função desejada f não pode ser representada no espaço de hipóteses \mathcal{H} . Por exemplo, suponha um conjunto de dados não-linear. Se restringirmos o nosso espaço de hipóteses somente para hipóteses lineares, então a hipótese desejada para o problema não se encontrará no espaço procurado.

Uma forma de lidar com esses problemas é por meio do uso de multiclassificadores. Nesse contexto, em lugar de um classificador único, a tarefa de classificação é desempenhada por um conjunto de classificadores (Dietterich, 2000a; Kuncheva, 2004). Construindo-se um multiclassificador baseado nos classificadores induzidos (classificadores base), o algoritmo pode combinar as suas saídas e reduzir o risco de escolher erroneamente um classificador. A Figura 1.1 ilustra essa situação. A curva externa denota o espaço de hipótese \mathcal{H} . A curva interna ilustra o conjunto de hipóteses que obtêm uma boa acurácia no conjunto de treinamento. O ponto rotulado como f é a hipótese desejada. Podemos notar que a média das acurácias das hipóteses encontradas é uma boa aproximação de f .

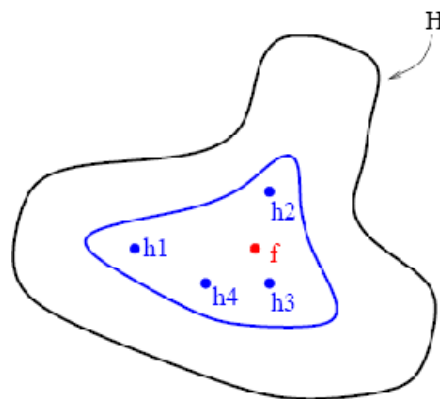


Figura 1.1. Fundamento estatístico da razão de um multiclassificador poder ter melhor desempenho do que um classificador único.¹

A construção de um multiclassificador através da execução de uma pesquisa local de diferentes pontos iniciais pode prover uma melhor aproximação da função desejada do que qualquer um dos seus classificadores individuais, como mostra a Figura 1.2. Dessa forma,

¹ Figura Extraída de (Dietterich, 2000a).

um multiclassificador pode reduzir o risco dos classificadores que o compõem ficarem presos em um ponto ótimo local.

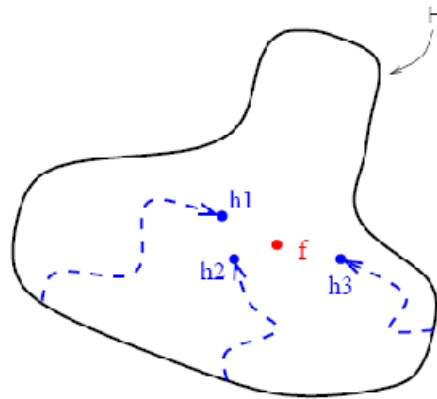


Figura 1.2. Fundamento computacional do porque de um multiclassificador poder ter melhor desempenho do que um único classificador.²

Suponha, novamente, um conjunto de dados não-lineares e um espaço de hipótese restrito a apenas hipóteses lineares. Então a hipótese desejada f para o problema não se encontra no espaço \mathcal{H} . A Figura 1.3 ilustra a capacidade de um multiclassificador composto apenas de classificadores lineares poder aproximar a função desejada f .

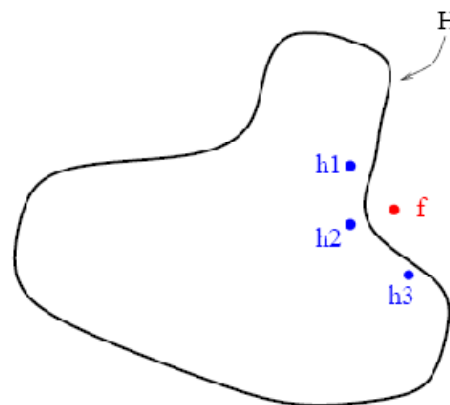


Figura 1.3. Fundamento representacional do porque de um multiclassificador poder ter melhor desempenho do que um único classificador.²

Devido a essas razões, expressas inicialmente por Dietterich (2000a), nos últimos anos, tem-se dado muita atenção a multiclassificador para resolver diversas tarefas de classificação, como reconhecimento de voz (Mukhopadhyay, Peng, Raje, Palakal, & Mostafa, 2003), reconhecimento de biométricas (Shou-Jue & Xu, 2003; Lemieux & Parizeau, 2003; Wang, Chen, & Ke, 2007; Liu, Lin, & Chen, 2008), reconhecimento de caracteres (Heutte, Nosary, & Paquet, 2004), reconhecimento de seqüências de proteínas

² Figura Extraída de (Dietterich, 2000a).

(Mohamed, Rubin, & Marwala, 2007), reconhecimento de células do cólon (Kruk, Osowski, & Koktysz, 2007) e reconhecimento de tumores (Liu, Zhou, & Chen, Ensemble Classification for Cancer Data, 2008). Na literatura, o termo multiclassificador também é conhecido como comitê de classificadores e mistura de especialistas (Kuncheva, 2004).

1.2 Contexto Específico

Multiclassificadores são baseados na idéia que um *pool* de diferentes classificadores, referidos individualmente como especialistas ou classificadores base, pode fornecer informações complementares sobre instâncias a serem classificadas, melhorando a eficácia do processo de classificação.

Nesse contexto, um aspecto que tem sido reconhecido de grande importância é a diversidade do sistema (Kuncheva & Whitaker, 2003; Banfield, Hall, Bowyer, & Kegelmeyer, 2003; Tsymbal, Pechenizkiy, & Cunningham, 2005; Windeatt, 2005; Tang, Suganthan, & Yao, 2006). Diversidade é vista como o grau de desacordo entre os erros realizados pelos classificadores. Por exemplo, não há nenhum ganho de acurácia na construção de multiclassificadores a partir de um conjunto de classificadores base idênticos. De fato, a situação ideal seria um conjunto de classificadores base – com erros não correlacionados e inversos – combinados de tal forma que minimizassem o efeito de seus erros. Em outras palavras, os classificadores base devem ser diversos entre si.

Embora diversos, se os classificadores base não forem tão acurados, a combinação deles, geralmente, apresentará um desempenho insatisfatório, pois a combinação será dominada por muitos componentes não acurados e, na maioria das vezes, errará, levando, assim, a uma generalização pobre. Se os especialistas forem muito acurados, será difícil encontrar diversidade entre eles e a suas combinações, freqüentemente, levam a uma melhora insignificante do desempenho do sistema (Xuchun, Zhu, & Sung, 2008). Devido a esses problemas os classificadores base devem ser tão acurados quanto diversos entre si (Hansen & Salamon, 1990; Xuchun, Zhu, & Sung, 2008; Kuncheva & Whitaker, 2003). A esse conflito chamamos de Dilema Diversidade–Acurácia (Dietterich, 2000b).

Uma das possíveis formas de se alcançar diversidade é utilizar diferentes tipos de algoritmos de aprendizado para construirmos os classificadores base (comitês heterogêneos). Apesar disso, poucos estudos utilizaram essa abordagem (Yates &

Partridge, 1996; Wang, Jones, & Partridge, 2000; Wang, Partridge, & Etherington, 2001; Langdon, Barrett, & Buxton, 2002; Woods, Kegelmeyer, & Bowyer, 1997; Chandra, 2004). Dessa forma, esta dissertação concentra-se no estudo empírico do comportamento de multiclassificadores heterogêneos mediante o Dilema Diversidade–Acurácia.

Além da diversidade, outro aspecto importante é que, em algumas aplicações, o número de classificadores base necessário para formar o multiclassificador com uma acurácia considerada razoável pode ser extremamente grande (Aksela, 2003; Valentino, 2005). Neste caso, com o intuito de diminuir o número de componentes do multiclassificador e, ao mesmo tempo, manter a diversidade entre eles, utilizam-se diferentes métodos para escolha dos classificadores base. A maioria desses métodos é baseada em algoritmos evolucionários, em que o tamanho do multiclassificador pode ser determinado automaticamente (Liu, Yao, & Higuchi, 2000; Islam, Yao, & Murase, 2003; Chandra & Yao, 2004; Chandra, 2004; Santos, Sabourin, & Maupin, 2006; Sylvester & Chawla, 2006). Existem vários métodos para selecionar os classificadores base que formarão o multiclassificador. Entre esses trabalhos, os mais relacionados com a nossa abordagem são os encontrados em (Chandra, 2004) e em (dos Santos, Sabourin, & Maupin, 2006). As diferenças entre a nossa abordagem e a desses trabalhos são explicadas no Capítulo 3.

O uso de algoritmos genéticos permite não apenas a seleção automática de classificadores base que formarão os comitês, como também no caso algoritmos genéticos multiobjetivo tem-se uma maneira direta de se investigar o impacto do uso simultâneo das duas medidas na escolha desses classificadores, facilitando o estudo do dilema diversidade-acurácia. Nesse contexto, com o intuito de desenvolvermos nossa investigação, duas versões diferentes de algoritmos genéticos foram utilizadas: uma versão mono-objetivo e outra multiobjetivo. A primeira (mono-objetivo) é utilizada para escolha dos classificadores base, quando utilizamos apenas um único objetivo – somente a acurácia ou apenas a diversidade. A segunda (multiobjetivo), quando utilizamos dois objetivos – acurácia e diversidade, conjuntamente. Em ambas as versões, os multiclassificadores são formados dinamicamente de acordo com o *pool* de classificadores base. Nesta dissertação, realizamos uma análise do dilema diversidade-acurácia no contexto de multiclassificadores heterogêneos: (1) impacto de selecionarmos os classificadores base usando diferentes objetivos; (2) implicação de selecionarmos os classificadores base usando diferentes

métodos de combinação; (3) o relacionamento entre as acurácias dos melhores comitês encontrados e a média de todos os comitês encontrados; e (4) a relação entre os desempenhos dos melhores e piores comitês encontrados.

1.3 Motivação e Objetivo

No contexto de criação de comitês, houve poucos estudos que levasse em consideração os tipos de algoritmos utilizados para a construção de classificadores base. Yates e Partridge (1996) usaram redes neurais *perceptron* de multicamada e redes neurais de funções de base radial juntas em um comitê e mostraram que diferentes tipos resultam em maior diversidade e alcançam um melhor desempenho. Árvores de decisão e redes neurais foram combinadas na criação de comitês heterogêneos em (Wang, Jones, & Partridge, 2000; Wang, Partridge, & Etherington, 2001; Langdon, Barrett, & Buxton, 2002). Nesses estudos, diferentemente desta dissertação, não foram utilizadas nenhuma medida de diversidade para guiar o processo criação dos comitês.

Woods et al (1997) combinaram classificadores lineares *bayes*, classificadores quadráticos *bayes*, redes neurais *perceptron* de multicamada e *k*-vizinhos mais próximos e árvores de decisão na formação de diferentes tamanhos de multiclassificadores. Embora, esse trabalho sugira uma metodologia para determinar a melhor combinação de classificadores base, esse estudo tinha como principal objetivo a investigação de um novo método de combinação. Nos resultados encontrados, os comitês usando quatro tipos de classificadores se mostraram melhores que a construção usando todos os cinco tipos, o que sugere que a escolha de arbitrária dos tipos de algoritmos de aprendizado utilizados, embora resultem em bons multiclassificadores, podem não ser uma escolha ótima.

A seleção de tipos de classificadores, dificilmente pode ser evitada devido à natureza do campo de aprendizado de máquina. Há inúmeros algoritmos de aprendizado de máquina, além disso, novos algoritmos vêm surgindo. Isso sugere que o espaço de busca, de acordo com os tipos de classificadores, está em constante expansão e um ótimo nesse caso não pode ser encontrado nesse momento e nem em um futuro próximo. Dessa forma, no contexto de criação de comitês, é necessário termos não somente bons classificadores base em relação aos demais membros, mas também em relação a outras hipóteses presentes em seu espaço de hipótese. Portanto, escolher os tipos de classificadores é melhor do que uma decisão arbitrária (Chandra, 2004). Nesse contexto, Chandra (2004) analisou a criação

de comitês híbridos utilizando redes neurais, máquinas de vetores suporte e redes neurais de função de base radial. No entanto, não é analisado o impacto das medidas de diversidade no processo de criação dos comitês, sua maior preocupação é em verificar se o desempenho dos comitês gerados são superiores a comitês criados a partir de outros algoritmos encontrados na literatura.

Dessa forma, faltam estudos detalhados sobre o dilema diversidade–acurácia em comitês heterogêneos. Portanto, como mencionado anteriormente, o objetivo do presente trabalho é efetuar um estudo detalhado sobre o comportamento de multiclassificadores heterogêneos perante o dilema diversidade–acurácia. Com esse intuito, analisamos os resultados obtidos a partir das seguintes perspectivas:

- Impacto de selecionarmos os classificadores base usando diferentes objetivos. Efetuamos essa análise com o intuito de verificarmos se existe um conjunto de objetivos mais indicado para o processo de criação de comitê.
- Impacto de selecionarmos os classificadores base usando diferentes métodos de combinação. Desenvolveremos esse estudo com o intuito de verificarmos se existe um método de combinação superior a outro.
- Relação dos desempenhos dos melhores comitês encontrados e de todos os obtidos, bem como a relação entre os melhores e os piores comitês encontrados. Essa análise tem o objetivo de verificarmos um panorama mais amplo do comportamento dos algoritmos genéticos na geração dos comitês.

1.4 Organização do Trabalho

O restante desta dissertação é organizado da seguinte forma:

- O Capítulo 2 apresenta o referencial teórico relativo à construção de multiclassificadores.
- O Capítulo 3 descreve a nossa abordagem para a investigação do dilema diversidade–acurácia no contexto de multiclassificadores, fazendo uma breve revisão dos trabalhos relacionados ao nosso, bem como uma comparação entre as abordagens deles e a nossa.
- A metodologia de desenvolvimento e de avaliação dos experimentos é apresentada no Capítulo 4.

- No Capítulo 5, as análises dos experimentos são introduzidas, focando-se na discussão dos resultados obtidos.
- Capítulo 6 apresenta as considerações finais desta dissertação e os possíveis trabalhos futuros da investigação do dilema diversidade–acurácia.

Capítulo 2

Construção de Multiclassificadores

Durante a construção de multiclassificadores, os dois principais problemas são: os componentes do comitê e o método de combinação utilizado (Kuncheva, 2004). Em relação ao primeiro problema, a escolha correta dos classificadores base é fundamental para o desempenho do sistema. Como mencionado anteriormente, a situação ideal seria um conjunto de especialistas com erros não correlacionados, combinados de tal modo que o efeito desses erros fosse minimizado. Dito de outra forma, os classificadores base devem ser diversos entre si.

Diversidade pode ser alcançada de diferentes formas, por exemplo, usando diferentes tipos de algoritmos de aprendizado na indução de classificadores, variando-se os parâmetros dos classificadores base e utilizando diferentes instâncias no conjunto de treinamento. Devido a sua importância na construção de comitês, diferentes índices para medir a diversidade foram propostos (Liu & Yao, 1999; Banfield, Hall, Bowyer, & Kegelmeyer, 2003; Kuncheva & Whitaker, 2003; Windeatt, 2005).

Uma vez que o *pool* de classificadores base tenha sido criado e um método de escolha dos componentes do comitê aplicado, o próximo passo é escolher uma forma eficiente de associar as saídas desses membros. Na literatura há um grande número desse tipo de método. Eles podem, em geral, ser classificados de acordo com suas funcionalidades em três estratégias principais: método baseado em combinação, método baseado em seleção e método híbrido (Kuncheva, 2004).

As próximas seções desse capítulo serão dedicadas a cada um dos principais aspectos considerados na criação de multiclassificadores: construção dos componentes do comitê, medidas de diversidade, métodos de escolha dos classificadores base e métodos de combinação.

2.1 Classificadores base

Nesta dissertação, usamos três tipos diferentes de algoritmos de aprendizado para construir os classificadores base: k -vizinhos mais próximos (k -NN, do inglês *k-nearest neighbour*), rede neural *perceptron* de multicamada com *backpropagation* (MLP com *backpropagation*, do inglês *Multilayer Neural Networks with backpropagation*) e máquinas de vetores suporte (SVM, do inglês *Support Vector Machines*). Esses três algoritmos são amplamente utilizados na literatura de aprendizado de máquina (Mitchell, 1997; Duda, Hart, & Stork, 2001). Além disso, eles representam diferentes paradigmas: k -NN faz parte do chamado *lazy learning*, pois não há um treinamento explícito, cada instância de treinamento é adicionada em uma lista de treinamento, o SVM e a MLP com *backpropagation* estão, respectivamente, no contexto de aprendizado estatístico e conexionista. Todos os algoritmos de aprendizados utilizados foram exportados através de uma adaptação realizada na ferramenta WEKA (Witten & Frank, 2005).

2.1.1 k -Vizinhos mais próximos

O k -vizinhos mais próximo (k -NN, do inglês *k-nearest neighbor*) é uma das técnicas mais simples e fáceis de implementar (Mitchell, 1997; Duda, Hart, & Stork, 2001). Classificadores gerados por esse tipo de modelo baseiam suas predições em instâncias previamente conhecidas, fazendo comparações entre as instâncias de entrada (instâncias desconhecidas) com as instâncias previamente conhecidas. Estas instâncias conhecidas foram armazenadas, durante a fase de treinamento, em uma lista (lista de instâncias de treinamento), ou seja, cada instância de treinamento amostrada é adicionada nessa lista. Uma de suas principais vantagens é o pequeno tempo de treinamento, pois se resume a isto: adicionar instâncias na lista. A tarefa de classificação é descrita no Algoritmo 2.1.

Algoritmo 2.1. Algoritmo de classificação do k -NN.

- 1 **Para cada** instância \mathbf{x}_i do treinamento e entrada \mathbf{x}_j **faça**
 - 2 Calcular a distância entre \mathbf{x}_i e \mathbf{x}_j ;
 - 3 Selecionar as k instâncias mais próximas;
 - 4 Selecionar a classe mais votada entre as instâncias selecionadas.
Em caso de empate selecione aleatoriamente uma dentre elas.
-

Em geral, a distância Euclidiana é utilizada para se calcular os vizinhos mais próximos. Dados duas instâncias com m atributos (características), $\mathbf{x}_i = \langle \dots, \dots, \dots \rangle$ e $\mathbf{x}_j = \langle \dots, \dots, \dots \rangle$, a distância euclidiana é calculada como segue:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (2.1)$$

Como podemos observar, a classificação do k -NN é bastante simples e intuitiva. Para classificar uma instância, primeiro encontram-se as k instâncias mais próximas dessa instância e dentro desse conjunto, a classe mais representativa. A Figura 2.1 ilustra uma situação de classificação do k -NN, em que existem amostras distribuídas em três classes e deseja-se determinar as classes de dois pontos (instâncias) desconhecidos. Nesse exemplo, os sete vizinhos próximos, de cada ponto desconhecido, são questionando sobre o rótulo da instância de entrada. Nesse contexto, o ponto desconhecido 1 é rotulado como classe B e o ponto desconhecido 2, como A.

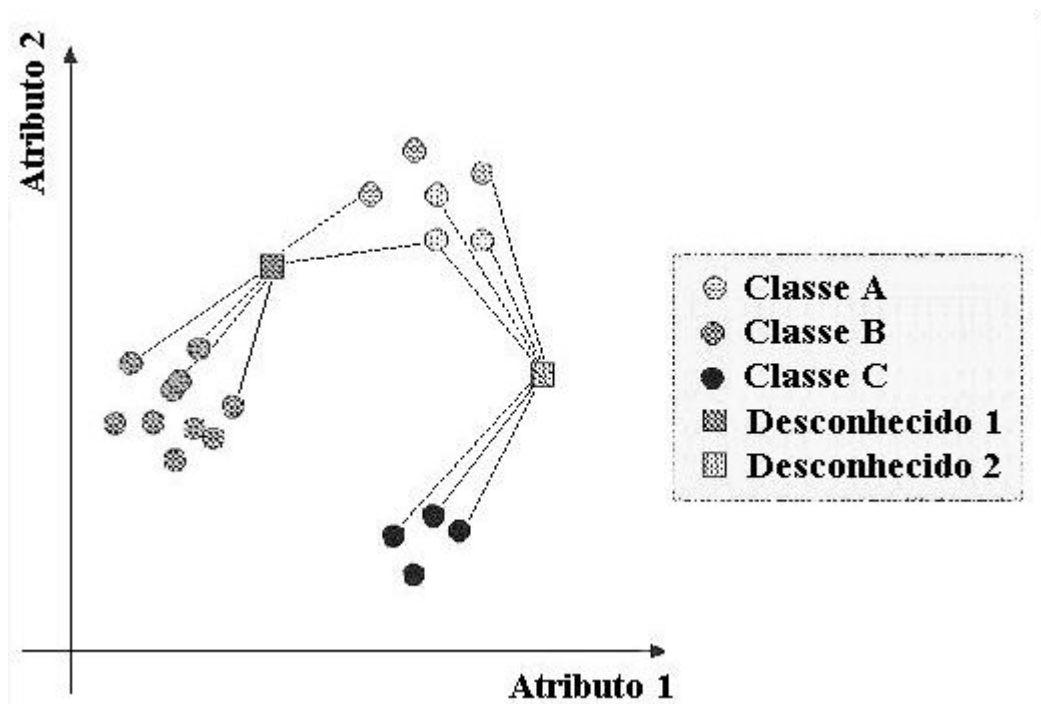


Figura 2.1. Esquema de classificação pelo Método k -NN.³

Um refinamento possível é punir as instâncias com distâncias mais longínquas, ou seja, ponderar a contribuição de cada instância dando maior peso aos vizinhos mais

³ Figura extraída de (Gnecco, Moraes, Machado, & Cabral, 2001).

próximos. É importante observar que dependendo do número de instâncias na lista de treinamento e do número de atributos de uma instância, a predição pode ser bastante onerosa devido aos excessivos cálculos matemáticos envolvidos. Pode-se destacar entre outras desvantagens a sensibilidade a ruídos, a intolerância à redundância de dados, a falta de mecanismos para lidar com atributos faltosos e a limitação de processar atributos nominais. (Aha, Kibler, & Albert, 1991).

2.1.2 Rede Neural *Perceptron* de Multicamada com retropropagação

As Redes Neurais Artificiais (RNAs) foram desenvolvidas inspiradas no funcionamento do cérebro, em que neurônios, conectados em forma de rede, são capazes de aprender e generalizar conhecimento (Mitchell, 1997; Duda, Hart, & Stork, 2001). Um exemplo de RNAs é a rede neural *perceptron* de multicamadas (MLP do inglês, *multilayer neural networks*), que consiste em um conjunto de unidades sensoriais (nós ou neurônios) conectados. O conjunto de neurônios pode ser dividido nas seguintes categorias: uma camada de entrada, uma ou mais camadas intermediárias (também conhecidas como camadas ocultas ou escondidas) e uma camada de saída. Cada conexão está associada a um peso. A Figura 2.2 ilustra uma rede MLP.

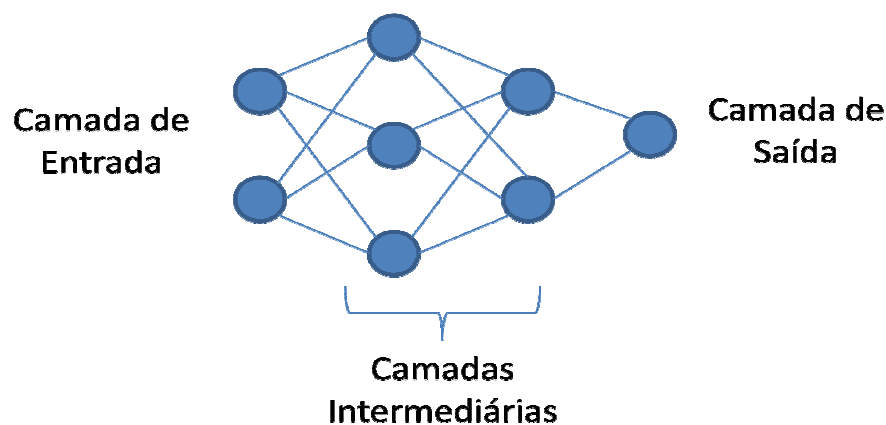


Figura 2.2. Exemplo de uma rede MLP.

O aprendizado em RNAs ocorre por meio da apresentação de um conjunto de instâncias à rede. Estas instâncias, que representam um determinado problema, são utilizadas por um algoritmo de treinamento para ajustar os pesos (sinapses) das conexões entre os neurônios. O objetivo do processo de treinamento é extrair o conhecimento necessário para a resolução do problema em questão. A camada de entrada tem a única função de propagar os valores inseridos (instâncias) em seus nós para os nós da camada

subseqüente – camada intermediária. As camadas escondidas são responsáveis por extrair características das instâncias observadas, codificando e representando essas características em seus pesos.

Na maioria das vezes, em redes MLP o algoritmo de treinamento utilizado é o *backpropagation* de erro (Duda, Hart, & Stork, 2001). Este é o algoritmo utilizado nesta dissertação. Basicamente, a aprendizagem por *backpropagation* de erro consiste em dois passos através das diferentes camadas da rede: um passo para frente – a propagação – e um passo para trás – a retropropagação. Os principais passos desse processo são demonstrados no Algoritmo 2.2.

Algoritmo 2.2. Algoritmo de *backpropagation*.⁴

1. Inicializar pesos da rede com valores aleatórios;
 2. **Repita**
 3. erro_total = 0;
 4. **para todo** cada instância de treinamento **faça**
 5. **para todo** cada camada i da rede, $i = 1, 2, \dots, n$ **faça**
 6. **para todo** cada nodo n_{ij} da i -ésima camada **faça**
 7. Calcular saída_produzida do nodo;
 8. **fim-para**
 9. **fim-para**
 10. erro = saída_desejada – saída_produzida;
 11. **para todo** cada camada i da rede, $i = n, n - 1, \dots, 1$ **faça**
 12. **para todo** cada nodo n_{ij} da i -ésima camada **faça**
 13. Ajustar pesos do nodo;
 14. **fim-para**
 15. **fim-para**
 16. erro_total = erro_total + erro;
 17. **fim-para**
 18. **até que** erro_total > valor_desejado
-

Dessa forma, no passo para frente, uma instância (vetor de entrada) é aplicada aos nós da camada de entrada, e seu efeito se propaga através da rede, camada a camada, até que um conjunto de saídas é produzido como resposta da rede. Durante o passo de

⁴ Extraído de (de Souto, Lorena, Delbem, & de Carvalho, 2003).

propagação, os pesos da rede são todos fixos. No passo para trás, esses pesos são todos ajustados de acordo com uma regra de correção de erro. Especificamente, a resposta da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Este sinal de erro é então propagado para trás através da rede, direção oposta das conexões. Os pesos são ajustados para fazer com que a resposta da rede se mova para mais próximo da saída desejada.

Dessa forma, devido às camadas ocultas, a rede pode extrair progressivamente as características mais significantes das instâncias de entrada, até mesmo para tarefas complexas. Outras vantagens que podemos citar são tolerância a presença de ruídos nos dados e habilidade de aproximar qualquer função (Mitchell, 1997; Duda, Hart, & Stork, 2001). Em contrapartida, entre os principais problemas são a dificuldade de definição de seus parâmetros (como o número de nós nas camadas escondidas e a taxa de aprendizado) e dificuldade de compreensão dos conceitos aprendidos pela rede.

2.1.3 Máquinas de Vetores Suporte

Nos últimos anos, uma técnica de aprendizado que vem recebendo grande atenção é as Máquinas de Vetores Suporte (SVMs, do inglês *Support Vector Machines*) (Hearst, Schölkopf, Dumais, Osuna, & Platt, 1998). Isso se deve, principalmente, a sua boa capacidade de generalização e robustez diante de dados de grande dimensão.

As SVMs baseiam-se no princípio da Teoria de Aprendizado Estatístico (Vapnik, 2000), que apresenta diversos limites na capacidade de generalização de um classificador linear. Dado um conjunto de treinamento composto por n instâncias das classes $\{-1,+1\}$, ele procura um hiperplano $\mathbf{w} \cdot \mathbf{x} + b = 0$ capaz de separar as instâncias dessas duas classes, onde \mathbf{w} é um vetor de pesos ortogonal ao hiperplano, \mathbf{x} é a instância a ser classificada e b um fator de deslocamento (do inglês, *offset*). De acordo com esse princípio, o hiperplano ótimo é aquele que maximiza a margem de separação (ρ) entre as classes, podendo ser obtido resolvendo o seguinte problema de otimização:

$$\text{Minimize: } \|\mathbf{w}\| + \quad (2.2)$$

$$\text{Sob as condições: } \begin{matrix} \geq 0, & = 1, \dots, \\ (\mathbf{w} \cdot \mathbf{x} + & \geq 1 - \end{matrix})$$

em que C é um parâmetro que impõe um compromisso entre o erro de treinamento e a generalização e ρ representa a variável responsável pela imposição do relaxamento. As restrições impostas tendem a nos garantir que nenhuma instância de treinamento ficará nos limites e são abrandadas com o intuito de evitar um superajuste (do inglês *overfitting*) na presença de dados ruidosos. A minimização do somatório é responsável por controlar o número de erros durante a fase de treinamento, bem como permitir instâncias nos limites.

Dessa forma, informalmente, podemos definir o processo de treinamento como dado um conjunto de treinamento que contém instâncias de duas classes, uma SVM separa as amostras de treinamento positivas das negativas por meio de um hiperplano. Os hiperplanos são determinados por pontos (vetores suporte) presentes nesse conjunto. Dessa forma, o hiperplano que provê a maior separação (maior margem possível entre as amostras das classes) é selecionado. A Figura 2.3 ilustra a classificação de um conjunto de dados através de uma SVM.

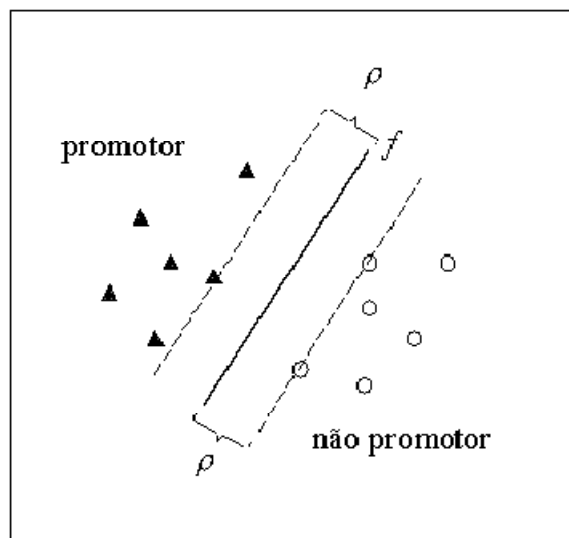


Figura 2.3. Exemplo classificação de promotores por meio de uma SVM linear.⁵

A prévia formulação ainda apresenta utilização limitada a problemas lineares, embora possa ser generalizada para problemas não linearmente separáveis através de funções de núcleo (do inglês, *kernel*) não-lineares, baseando-se na idéia, explicada no Capítulo 1, que um multiclassificador composto somente por classificadores lineares pode aproximar uma função desejada f , mesmo que ela não se encontre no espaço de hipótese procurado, somente SVMs lineares foram empregadas nesta dissertação.

⁵ Figura extraída de (de Souto, Lorena, Delbem, & de Carvalho, 2003).

A escolha da função de núcleo e seus parâmetros, bem como do parâmetro C influenciam diretamente no desempenho do classificador gerado por uma SVM (Muller, Mika, Ratsch, Tsuda, & Scholkopf, 2001). Essa sensibilidade a escolha dos parâmetros representa uma das dificuldades das SVMs. Outra desvantagem é a dificuldade de interpretação o modelo gerado, similar as redes neurais artificiais.

2.2 Medidas de Diversidade

A diversidade é vista como o grau de desacordo entre os erros realizados pelos classificadores, dessa forma os multiclassificadores baseia-se na idéia que um *pool* de diferentes classificadores pode fornecer informações complementares sobre instâncias a serem classificadas, melhorando a eficácia do processo de classificação. Nesse contexto, a diversidade tem sido reconhecida como de grande importância para o desempenho do sistema (Kuncheva & Whitaker, 2003; Banfield, Hall, Bowyer, & Kegelmeyer, 2003; Tsymbal, Pechenizkiy, & Cunningham, 2005; Windeatt, 2005; Tang, Suganthan, & Yao, 2006).

Em geral, a diversidade pode ser alcançada de diferentes maneiras (Brown, Wyatt, Harris, & Yao, 2005). Nesta dissertação, fizemos com que os componentes classificadores iniciassem o processo de aprendizagem em diferentes condições, ou seja, variamos os parâmetros iniciais e de aprendizagem do modelo. Utilizamos diferentes arquiteturas ou algoritmos de aprendizado para treinar os classificadores base. Como dito anteriormente, utilizou-se os seguintes algoritmos: k -NN, SVMs e MLP com *backpropagation*.

Podemos classificar as medidas de diversidade em duas categorias: com paridade e sem paridade (Kuncheva & Whitaker, 2003). Na primeira, a diversidade do comitê é calculada usando-se os componentes dois a dois. Na outra, a análise é feita individualmente, utilizando-se a idéia de entropia e correlação de saídas individuais ou baseando-se na dificuldade de distribuição dos pontos dos dados trabalhados.

De fato, medir a diversidade não é uma tarefa tão simples de se realizar, pois não há uma definição formal genérica aceita. No contexto mais experimental, Kuncheva e Whitaker (2003) nos traz um estudo detalhado sobre dez medidas de diversidades encontradas na literatura. Baseado nesse estudo empírico, os referidos autores sugerem utilizar três medidas das analisadas. A razão disso é de que elas são as de mais fáceis

interpretações, uma vez que, o valor zero significa independência entre os classificadores, os valores positivos significam que eles tendem a acertar ou errar os mesmos padrões e os valores negativos que eles tendem a acertar ou errar padrões diferentes.

Assim, nesse contexto, podemos afirmar que os classificadores são considerados mais diversos quando possuem valores negativos na sua medida de diversidade e menos diversos quando têm valores positivos. Em outras palavras, queremos que nossas medidas assumam valores negativos. A seguir as três medidas serão apresentadas, todos os valores das medidas encontram-se no intervalo $[-1,1]$.

2.2.1 Estatística- Q

Estatística-Q é uma medida que utiliza paridade, sendo dada por:

$$Q = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (2.3)$$

em que N^{00} e N^{11} são, respectivamente, o número de padrões que ambos os classificadores erram e acertam; N^{01} é a quantidade de exemplos que o primeiro erra, enquanto o segundo acerta; de forma análoga, N^{10} nos informa o número de padrões errados do primeiro classificador e o número de corretos do outro.

2.2.2 Coeficiente de Correlação ρ

Coeficiente de Correlação ρ é uma medida de paridade dada por:

$$\rho = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11} + N^{10}) + (N^{01} + N^{00}) + (N^{11} + N^{01}) + (N^{10} + N^{00})}} \quad (2.4)$$

em que N^{ab} tem o mesmo significado dos apresentados na medida anterior.

2.2.3 Medida de Concordância k

A Medida de Concordância k é uma medida estatística sem paridade, calculada de acordo com a seguinte fórmula:

$$k = \frac{2(N^{11}N^{00} - N^{01}N^{10})}{(N^{11} + N^{10}) + (N^{01} + N^{00}) + (N^{11} + N^{01}) + (N^{10} + N^{00})} \quad (2.5)$$

em que, N^{ab} tem a mesma interpretação das medidas com paridade.

2.3 Métodos de Escolha dos Classificadores Base

Há um amplo estudo sobre o treinamento de classificadores base e a escolha de um subconjunto deles, baseando-se em algum critério, para geração de comitê (Margineantu & Dietterich, 1997; Lazarevic & Obradovic, 2001; Giacinto & Roli, 2001; Aksela, 2003; Banfield, Hall, Bowyer, & Kegelmeyer, 2003; Wu & Chen, 2004; Chandra, 2004; Valentino, 2005; dos Santos, Sabourin, & Maupin, 2006). Esses procedimentos de escolha, geralmente, baseiam-se na acurácia (Margineantu & Dietterich, 1997; Lazarevic & Obradovic, 2001; Wu & Chen, 2004) ou na diversidade (Giacinto & Roli, 2001; Banfield, Hall, Bowyer, & Kegelmeyer, 2003; Valentino, 2005).

No contexto de aprendizado não-supervisionado, por exemplo, os classificadores base podem ser agrupados de acordo com suas diversidades (Giacinto & Roli, 2001). Nesse caso, os especialistas que pertencem ao mesmo grupo tendem a fazer erros correlacionados. Assim, um membro de cada grupo é selecionado para compor o comitê. Banfield, Hall, Bowyer, & Kegelmeyer (2003) introduziram um procedimento baseado em pontos desconhecidos, esses pontos são responsáveis por causar diversidade para os multiclassificadores. Dessa forma, os membros que tem a maior acurácia são escolhidos para formarem o comitê.

Na escolha de componentes baseado na acurácia, Wu & Chen (2004) propuseram um novo algoritmo baseado no *Bagging* (Breiman, 1996) e em algoritmo genético. Lazarevic & Obradovic (2001), um método para poda de comitês. Em (Aksela, 2003), no contexto de reconhecimento de caracteres, encontramos um estudo empírico de seis critérios de avaliação para a seleção de classificadores. Entretanto, somente um desses critérios utilizou a acurácia e diversidade na escolha dos membros dos comitês – a contagem do erro.

A combinação da acurácia com a diversidade como critério de escolha dos membros de um comitê mostra-se, em geral, superior do que essas medidas separadas (Kuncheva & Whitaker, 2003; Santana, Soares, Souto, & Canuto, 2008). Dessa forma, algoritmos evolucionários são atrativos na busca de classificadores que comporão o comitê (Sirlantiz, Fairhurst, & Guest, 2002). Nesse contexto, alguns trabalhos envolvendo algoritmos evolucionários foram desenvolvidos (dos Santos, Sabourin, & Maupin, 2006; Chandra & Yao, 2004; Chandra, 2004; Sylvester & Chawla, 2006). Esta dissertação também se

encontra inserida nesse contexto, tendo como trabalhos mais relacionados os de Chandra (2004) e dos Santos, Sabourin, & Maupin (2006).

2.4 Métodos de Combinação

Embora a construção de comitês baseia-se no conceito de que diferentes classificadores possam fornecer informações complementares, a associação das saídas pode ser problemática. Desse modo, as vantagens e desvantagens de cada classificador escolhido devem ser levadas em considerações nas estratégias de associação (Ho, Hull, & Srihari, 1994).

Muitas dessas estratégias foram propostas na literatura. Como dito anteriormente, elas podem ser classificadas, de acordo com suas funcionalidades, em três estratégias principais: método baseado em combinação, método baseado em seleção e método híbrido (Kuncheva, 2004). Devido à falta de termos mais adequados emprega-se o termo métodos de combinação para designar o método genérico de informar a saída de um comitê, mesmo para quando um classificador é selecionado para dar a saída do mutliclassificador, sem precisar se combinar com outro. Tornando a terminologia ainda mais confusa, encontramos o termo métodos baseados em combinação, designando os métodos que de fato combinam as saídas dos classificadores que compõem um comitê. Para diminuir esse problema, a partir desse ponto, chamaremos os métodos baseado em combinação de métodos de fusão, outro sinônimo que encontramos na literatura (Kuncheva, 2004).

Os métodos de fusão assumem que os classificadores base possuem experiências equivalentes em todo o espaço de busca e, portanto, a opinião de todos é considerada na tarefa de classificação. Na literatura, há um amplo número de diferentes métodos de fusão. De acordo com suas características, podem ser classificados como linear ou não-linear.

No caso dos métodos baseados em seleção – também chamados de métodos de seleção –, durante o processo de classificação, somente um único classificador base é utilizado para classificar a instância de entrada. Para isso, durante a fase operacional, é aplicado um procedimento para selecionar o especialista que será utilizado. De modo geral, o classificador base é um perito em uma determinada região do espaço de busca, sendo escolhido pela vizinhança da instância de entrada. Dessa forma, utilizam-se métodos de seleção, quando se treina classificadores base explicitamente para diferentes partes do

espaço busca. Utilizam-se métodos de fusão, quando se treina classificadores base utilizando todo o espaço de busca (abordagem usada nesta dissertação), mas sem se preocupar muito com o desempenho individual deles.

Com o intuito de provê uma saída mais adequada para uma instância de entrada, os métodos híbridos usam características das técnicas de fusão e de seleção. Nesse contexto, em termos gerais, há um processo criterioso para decidir se será utilizado método de seleção ou de fusão, aumentando assim a complexidade de escolha. A idéia principal é usar o método de seleção se, e somente se, o melhor classificador for estatisticamente superior ao segundo melhor para rotular a instância de entrada, senão o método baseado em combinação é utilizado (Giacinto & Roli, 2001).

Nesta dissertação, com o intuito de diminuirmos a complexidade dos experimentos e devido à forma de treinamento aplicada nos classificadores base (explicada no Capítulo 3), usamos métodos de fusão. Mais especificamente, usaremos a Soma (linear) e o classificador *Naive Bayes* (não-linear), explicados a seguir.

2.4.1 Soma

Soma (Kittler & Alkoot, 2003) é um combinador dado pela equação:

$$(\mathbf{x}) = \frac{1}{C} \sum_{c=1}^C z_c, \quad c = 1, 2, \dots, N \quad (2.6)$$

em que C é o número de classificadores, N é o número de classes, \mathbf{x} é a instância que se quer classificar e o vetor \mathbf{z} é formado pelas saídas dos classificadores base (Canuto, Howells, & Fairhurst, 2000).

Dito de outra forma, dado um padrão de entrada, cada classificador informa as suas confiabilidades (valores de confiança) para cada classe ao módulo combinador. Este efetua um somatório por classes das confiabilidades de todos os classificadores. A classe que obtiver o maior somatório das confiabilidades será a escolhida para representar a saída do comitê.

Por exemplo, suponha que para uma instância x de entrada, as confiabilidades fornecidas para as classes $\Omega = \{c_1, c_2, c_3\}$ pelos cinco classificadores que formam um comitê sejam:

$$C(x) = \begin{bmatrix} 0,1 & 0,5 & 0,4 \\ 0,0 & 0,0 & 1,0 \\ 0,4 & 0,3 & 0,4 \\ 0,2 & 0,7 & 0,1 \\ 0,1 & 0,8 & 0,2 \end{bmatrix}$$

em que cada coluna representa a classe na ordem em que ela aparece em N e a linha os classificadores presentes no multiclassificador. Dessa forma, como temos o somatório das confiabilidades de cada classe sendo

$$= 0,1 + 0,0 + 0,4 + 0,2 + 0,1 = 0,8$$

$$= 0,5 + 0,0 + 0,3 + 0,7 + 0,8 = 2,3$$

$$= 0,4 + 1,0 + 0,4 + 0,1 + 0,2 = 2,1$$

então a classe escolhida como saída do multiclassificador é c_2 .

2.4.2 Classificador Naive Bayes

O *Naive Bayes* (NB) é um algoritmo de aprendizado de máquina simples que se baseia na aplicação do teorema de Bayes (Duda, Hart, & Stork, 2001) e na suposição de que os atributos das instâncias são independentes dadas às classes.

Apesar da suposição, o NB mostra-se eficiente em várias tarefas de classificação, mesmo em casos de atributos apresentarem uma dependência em relação à classe (Mitchell, 1997; Witten & Frank, 2005).

Para classificarmos uma instância x , precisa-se calcular a probabilidade de x pertencer à classe c_k . Para isso usa-se a seguinte equação:

$$P(c_k | x) = \frac{P(c_k) \prod_{i=1}^n P(x_i | c_k)}{\sum_{j=1}^K P(c_j) \prod_{i=1}^n P(x_i | c_j)} \quad (2.7)$$

em que n é a quantidade de atributos de uma instância.

No caso de multiclassificadores – contexto desta dissertação –, devido à suposição de independência dos atributos em relação à classe, uma matriz de confusão precisa ser criada. Dessa forma, a probabilidade da classe ser classificada é dada por

$$P_k = \frac{1}{N_k} \sum_{i=1}^m mc_{ki} S_i$$

em que N_k é a quantidade de instâncias da classe k no conjunto de treinamento, mc é a matriz de confusão dos classificadores utilizados e S_i é o rótulo dado pelo classificador D_i para a instância de entrada \mathbf{x} , onde $S_i \in \Omega$. A classe que possuir a maior probabilidade será responsável por representar a saída do comitê.

Por exemplo, suponha que há vinte instâncias, divididas em três classes: $N_1 = 8$, $N_2 = 9$ e $N_3 = 3$. Ao classificarmos uma instância de entrada \mathbf{x} utilizando-se dois classificadores, D_1 associou \mathbf{x} a classe 1 e D_2 a classe 2. As matrizes de confusão, respectivamente, são

$$mc_1 = \begin{bmatrix} 6 & 2 & 0 \\ 1 & 8 & 0 \\ 1 & 0 & 2 \end{bmatrix} \text{ e } mc_2 = \begin{bmatrix} 4 & 3 & 1 \\ 3 & 5 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

Dessa forma, como temos

$$P_1 = \frac{1}{8} \times 2 \times 4 = 1$$

$$P_2 = \frac{1}{9} \times 8 \times 3 = 2,67$$

$$P_3 = \frac{1}{3} \times 0 \times 0 = 0$$

então a saída do comitê para a instância entrada é a classe 1. Desse modo, utilizando o classificador *Naive Bayes* para combinar as saídas dos especialistas, efetua-se uma espécie de votação ponderada.

Capítulo 3

Abordagem Proposta

Este capítulo descreve a nossa abordagem usando algoritmos genéticos para auxiliar a investigação do dilema diversidade-accurácia no contexto de multiclassificadores. No entanto, antes de fazer isso, a fim de colocar nosso estudo em perspectiva revisaremos os trabalhos relacionados ao nosso. Para facilitar o entendimento da discussão, apresentaremos, a seguir, uma breve introdução a algoritmos genéticos.

3.1 Algoritmos Genéticos

Algoritmos Genéticos (AGs) foram desenvolvidos inicialmente por Holland em 1960 (Holland, 1992). AGs são métodos estocásticos de otimização global inspirados nos mecanismos biológicos, como evolução e hereditariedade. Nos últimos anos, têm sido muito utilizados em várias tarefas diferentes, como, por exemplo, otimização numérica (Zhong, Liu, Xue, & Jiao, 2004), otimização de redes neurais artificiais (Iba & Takefuji, 1998), agentes evolucionários para tarefas de classificação (Oliveira, Canuto, & Campos, 2006) entre outros.

Nos AGs as possíveis soluções são codificadas como cromossomos (indivíduo), o conjunto desses indivíduos é conhecido como população. Os indivíduos da população são avaliados constantemente através de uma função chamada de função de aptidão. Esta função nos informa quão bom é o indivíduo avaliado para resolver o problema. Novos indivíduos são gerados usando operadores genéticos como, por exemplo, recombinação e mutação (Fogel & Ghozeil, 1997). A recombinação é responsável pela procura da solução local, enquanto a mutação proporciona saltos os quais fazem com que a busca da solução passe a ser global. De modo geral, podemos resumir as etapas de qualquer AG como descrito no Algoritmo 3.1.

Esse tipo de AG funciona para problemas que têm apenas um objetivo na função de aptidão. Por outro lado, muitos problemas reais de otimização lidam com vários objetivos os quais devem ser otimizados simultaneamente para que a solução seja considerada satisfatória (dos Santos, Sabourin, & Maupin, 2006). Sendo assim, esse algoritmo foi

estendido para comportar a solução desses problemas, surgindo assim os Algoritmos Genéticos Multiobjetivos (AGMO). No caso de AGMO, o uso de vários objetivos conflitantes muda a definição de ótimo: em lugar de encontrarmos uma única solução ótima (usando um objetivo), geralmente, encontraremos um conjunto de ótimas.

Algoritmo 3.1. Etapas genéricas de um algoritmo genético.

1. **Iniciar** a população com N indivíduos;
 2. **Enquanto** uma das condições não é alcançada **faça**
 3. **Selecionar** N indivíduos;
 4. **Aplicar** operadores genéticos nos indivíduos selecionados;
 5. **Aplicar** um modelo populacional para encontrar os indivíduos que formarão a população da próxima iteração;
-

No conjunto de soluções encontrado pelo AGMO, nenhuma das soluções encontradas é melhor que as demais em pelo menos um objetivo. Essas soluções são chamadas de soluções não-dominadas, conhecidas também como soluções de Pareto (Abbass, 2003). Além dessas soluções, existem também soluções dominadas as quais todos os objetivos encontrados são inferiores a pelo menos outra solução encontrada. A Figura 3.1 mostra um exemplo desses tipos de soluções.

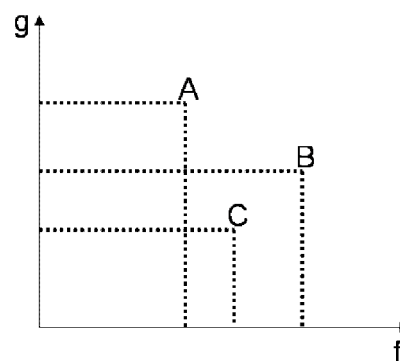


Figura 3.1. Possíveis tipos de soluções encontradas em um AGMO.

Suponha um AGMO cujo objetivo é minimizar as funções f e g simultaneamente, e em determinada etapa esse algoritmo encontra as soluções A, B e C conforme mostradas na Figura 3.1. Como podemos observar em relação à solução B podemos afirmar que ela não é melhor que nenhuma das outras possíveis soluções em nenhum aspecto, uma vez que possui valores maiores para ambas as funções. Por causa disso, chamamos B de solução

dominada; como A é melhor do que a C em relação à f , mas não é melhor em relação à g , ou seja, não é melhor em todas as funções em relação a C, chamamos A de solução não-dominada; de forma análoga temos C também como solução não-dominada.

Como podemos observar as soluções dominadas não são melhores em nenhum aspecto que as não-dominadas, sendo assim deverão ser descartadas. É importante ressaltar que ao contrário de um AG, em que temos uma única solução ótima no final da execução, no AGMO teremos várias soluções ótimas (soluções não-dominadas) para o problema. A partir disso surge o problema: qual solução utilizar para resolver o problema? Em geral, este questionamento cabe ao um especialista na área do problema em questão responder (Linden, 2006).

No caso específico desta dissertação e do trabalho em (dos Santos, Sabourin, & Maupin, 2006) foi utilizado o *Nondominated Sorting Genetic Algorithm II* – NSGA-II (Deb, Pratap, Agarwal, & Meyarivan, 2002) para os experimentos realizados com multiobjetivos.

O NSGA-II baseia-se na aplicação de operadores genéticos (recombinação e mutação) para obter uma nova população a partir da original. Após isso, os indivíduos das duas populações são ordenados rapidamente com base na não dominação e emprega-se o conceito de *crowding distance* – estimativa da densidade de soluções ao redor de uma solução particular – para manter a diversidade da população e compor um operador de comparação (*crowded comparison*). O NSGA-II funciona como descrito no Algoritmo 3.2.

 Algoritmo 3.2. Algoritmo do NSGA-II.⁶

1. **Gerar** aleatoriamente n^P indivíduos para compor a população inicial P_0 ;
 2. **Ordenar** P_0 de acordo com a não dominação;
 3. **Determinar** a adequabilidade para cada solução p . Nessa primeira geração, a adequabilidade de uma solução é igual ao seu nível de não dominação, $rank(p)$. O melhor nível é 1 (adequabilidade deverá ser minimizada);
 4. **Usar** seleção por torneio binário, recombinação e mutação para **criar** uma população filha Q_0 , também de tamanho n^P . Para a seleção, nessa primeira geração, somente o valor da adequabilidade é considerado: dois indivíduos são sorteados da população P_0 e aquele com o menor valor de aptidão é selecionado;
 5. **Inicializar** o número da geração: $t = 0$;
 6. **Formar** uma população combinada $R_t = P_t \cup Q_t$, de tamanho $2 n^P$;
 7. **Ordenar** R_t de acordo com a não dominação;
 8. **Selecionar** os indivíduos para a nova população P_{t+1} da seguinte maneira:
 9. Enquanto $|P_{t+1}| + |F_i| \leq n$
 10. **Calcular** a *crowding distance* para as soluções do fronte F_i ;
 11. **Incluir** os indivíduos do fronte F_i , na população P_{t+1} ($P_{t+1} = P_{t+1} \cup F_i$);
 12. $i = i + 1$;
 13. **Ordenar** em ordem decrescente o fronte F_i ;
 14. **Escolher** os primeiros $n^P - |P_{t+1}|$ elementos de F_i e os inclui na população P_{t+1} ;
 15. **Usar** seleção por torneio binário, recombinação e mutação para criar a população filha Q_{t+1} de tamanho n^P . Para criar cada indivíduo da população Q_{t+1} :
 16. **Selecionar** dois indivíduos da população P_{t+1} , utilizando torneio binário;
 17. **Aplicar** os operadores de recombinação e mutação;
-

O procedimento de ordenação, mencionado no Algoritmo 3.2, é dado pelo Algoritmo 3.3. Com isso, cada indivíduo terá um *rank*, que corresponde ao nível de não dominação em que ele se encontra – *rank* de não dominação, $rank(p)$.

⁶ Algoritmo Extraído de (Faceli, 2006).

 Algoritmo 3.3. Procedimento de Ordenação do NSGA-II.⁷

1. Para cada indivíduo $p \in P$ é feita uma contagem do número de soluções que dominam a solução p , chamada de contagem de dominação, cd_p , e determinado o conjunto de soluções que p domina, S_p ;
 2. Em seguida, os indivíduos são distribuídos em frentes de não dominação, F_i , em vários níveis i . As soluções não dominadas do primeiro frente, F_1 , têm a contagem de dominação igual a zero;
 3. Para cada solução p com $cd_p = 0$, cada $q \in S_p$ é visitado e tem seu valor decrementado em um;
 4. Cada q para o qual a contagem de diminuição foi zerada é colocado em uma lista separada, Q . Os indivíduos de Q compõem o segundo frente não dominado, F_2 ;
 5. O procedimento é repetido até que todos os frentes não dominados sejam identificados (todos os indivíduos estejam associados a um nível de não dominação).
-

O operador de comparação (*crowded comparison*), \prec , usado no processo de seleção para guiar o algoritmo em direção a um frente de Pareto uniformemente distribuído. Considera que cada indivíduo p na população tem dois atributos: o *rank* de não dominação ($rank(p)$) e a *crowding distance* (crd , o procedimento do seu cálculo é descrito no Algoritmo 3.4). Dessa forma, ao compararmos dois indivíduos p_1 e p_2 , teremos $p_1 \prec p_2$ se $rank(p_1) < rank(p_2)$ ou ($rank(p_1) = rank(p_2)$ e $crd(p_1) < crd(p_2)$), ou seja, entre duas soluções que possuem diferentes *ranks* de dominação, a melhor é aquela com menor *rank* e nos casos em que estes são iguais (estão no mesmo frente), a melhor solução é aquela que está na região menos povoada.

⁷ Algoritmo extraído de (Faceli, 2006)

 Algoritmo 3.4. Procedimento para calcular o *crowding distance*.⁸

1. **Para cada** função objetivo z_j :
2. **Ordenar** as soluções de F_i em ordem crescente de z_j ;
3. **Encontrar** as soluções limites p_1 , com menor valor de z_j (), e p_l , com maior valor de z_j ();
4. **Definir** a *crowding distance* em relação à z_j dessas soluções como sendo ∞ , ou seja, $crd(z_j, p_1) = crd(z_j, p_l) = \infty$;
5. **Para as demais** soluções p_w , com $w=2, \dots, l-1$, **calcular** a *crowding distance* em relação a z_j dada por:

$$crd(z_j, p_w) = \frac{z_j(p_{w-1}) - z_j(p_{w+1})}{z_j(p_w) - z_j(p_{w-1})}$$

6. A *crowding distance* de uma solução p é dada pela soma das *crowding distances* em relação aos m objetivos: $cdr(p) = \sum_{z_j} crd(z_j, p)$.
-

3.2 Trabalhos Relacionados

Como mencionado anteriormente, com o intuito de colocarmos o nosso estudo em perspectiva essa seção revisa os trabalhos relacionados ao nosso. Todos aqueles trabalhos que envolvam o estudo do dilema da acurácia-diversidade, podem ser citados como trabalhos relacionados ao nosso. Entre eles, destacam-se dois, descritos a seguir.

3.2.1 Algoritmos Genéticos Mono e Multi Objetivos

Em (dos Santos, Sabourin, & Maupin, 2006) uma investigação empírica sobre a formação de multiclassificadores foi realizada. Nesse contexto, combinações entre a taxa de erro, a cardinalidade (número de classificadores) e a diversidade foram responsáveis por guiar o processo de otimização, pois através da minimização da taxa de erro alcança-se o

⁸ Algoritmo Extraído de (Faceli, 2006).

principal objetivo do reconhecimento de instâncias – encontrar indutores de alto desempenho (alta acurácia); inspirando-se na idéia, introduzida por métodos de seleção de características, de que a redução do número de características, pode melhorar o processo de classificação, a minimização de número de componentes aparenta ser uma boa função objetiva; e baseando-se na idéia de que a diversidade é reconhecida como uma importante característica na combinação dos classificadores. Nesse estudo empírico, os referidos autores utilizaram doze medidas de diversidades – dez analisadas em (Kuncheva & Whitaker, 2003), uma definida em (Ruta & Gabrys, 2005) e uma analisada em (Zenobi & Cunningham, 2001).

Devido à observação de Ruta e Gabrys (2005) – algoritmos genéticos são bons algoritmos para problemas de classificação, devido à possibilidade de se lidar com um conjunto de soluções ao invés de uma única solução –, os algoritmos genéticos foram utilizados como método de busca de classificadores que comporão o comitê. Dessa forma, foi utilizado o NSGA-II e um AG. Ambos os algoritmos utilizam a representação cromossômica binária, onde um indivíduo equivale a um comitê. Os operadores genéticos utilizados foram à recombinação de um ponto e a mutação por alternância de *bit* (do inglês, *bit flip mutation*) cujas taxas foram, respectivamente, 0,8 e 0,01. Esses valores foram definidos através de experimentos. Através de uma investigação empírica, também foram definidos o tamanho da população (128) e o número de gerações (1000).

O *pool* de classificadores base é formado por 100 especialistas *k*-NN. Dessa forma, o vetor de representação é formado por 100 posições (genes), onde cada gene (*bit*) representa se o classificador está presente (1) ou não (0) no comitê. A saída de cada comitê é dada pelo Voto Majoritário (Kuncheva, 2004).

Com o intuito de alcançar altas taxas de acurácia e de evitar o problema da alta dimensionalidade dos dados – um dos principais problemas do algoritmo k-NN –, utilizaram-se do *Random Subspace Method* (Ho T. K., 1998) na criação dos indivíduos presentes *pool* de classificadores.

Na análise dos resultados, mesmo para o AGMO, foi considerado apenas o melhor comitê encontrado, considerando como critério a acurácia. A partir dessa análise, os referidos autores ratificaram a idéia de que a diversidade sozinha não é melhor do que a acurácia como função objetiva para encontrar os comitês mais acurados (Kuncheva & Whitaker, 2003; Ruta & Gabrys, 2005). Ao usarem conjuntamente a acurácia e a diversidade como objetivos do AGMO, os comitês encontrados (96,30% de acurácia) mostraram-se superiores aos obtidos somente com a diversidade (93,50% de acurácia). Entretanto, os desempenhos dos comitês obtidos usando a acurácia combinada com a diversidade (96,30% de acurácia) ainda permaneceu pior que os desempenhos dos comitês obtidos usando apenas a acurácia (96,50% de acurácia). Ao contrário do que se esperava, a acurácia conjuntamente com a cardinalidade não reduziu a complexidade dos comitês (em torno de 12 classificadores), enquanto aumentava o seu desempenho (96,33% de acurácia). De fato, em termos gerais, os melhores compromissos entre desempenho (96,33% de acurácia) e cardinalidade (em torno de 5 classificadores) foram encontrados usando-se a diversidade em conjunto com a acurácia. Desse modo, os referidos autores concluíram que a vantagem de usar, nesse contexto, a diversidade mais a taxa de erro é encontrar pequenos multiclassificadores com desempenhos razoáveis, como, por exemplo, entre 96,30% a 96,35% de acurácia para a base de dados estudada (NIST *Special Database 19*).

3.2.2 DIVACE-II

Como explicado anteriormente, o dilema da diversidade-accurácia nos informa que para uma melhora significativa do processo de classificação, os classificadores base devem ser tão acurados quanto diversos entre si. Nesse contexto, Chandra e Yao (2004) desenvolveram um algoritmo de aprendizado de comitês, chamado de DIVACE (*Diverse and Accurate Ensemble Learning Algorithm*), que fossem tão acurados quanto diversos. Na investigação empírica realizada em (Chandra & Yao, 2004), os resultados obtidos confirmaram a idéia de que a diversidade pode ser combinada com a accurácia para gerar comitês com melhores desempenhos. Nessa investigação foi utilizado apenas um tipo de algoritmo de aprendizado na criação dos classificadores base.

Nesse contexto, o DIVACE foi generalizado para um *framework* hierárquico que pode ser usado para sintetizar novos algoritmos de aprendizado de comitês (Chandra, 2004; Chandra & Yao, 2005). A Figura 3.2 ilustra o *framework* proposto. Nele existem três níveis de evolução. O primeiro, chamado de Evolução da Mistura (do inglês, *Evolution of the Mix*), consiste em utilizar vários tipos de algoritmos de aprendizado na criação dos classificadores base. O segundo, Evolução baseada no conjunto de treinamento (do inglês, *Evolution based on training set*), considera a evolução do comitê baseada na estruturação do conjunto de treinamento. O terceiro nível, Evolução baseada nas características da arquitetura ou do indutor (do inglês, *Evolution based on architectures/predictor intricacies*), corresponde a um processo similar ao DIVACE – ilustrado na Figura 3.3 –, por isso também é conhecido como estágio DIVACE. Nesse estágio, a recombinação depende de fatores como arquitetura nos casos das redes neurais RBFs (*radial base function*), arquitetura ou pesos nos casos de redes neurais MLPs e a função de núcleo no caso das SVMs.

A instanciação desse *framework* é conhecida como DIVACE-II cujo algoritmo está descrita em Algoritmo 3.5. No DIVACE-II, os referidos autores incorporaram os dois primeiros níveis de diversidade presentes no *framework* em questão. Basendo-se na idéia de que os classificadores base devem não ser somente bons com respeito a outros componentes, mas também em relação a outras hipóteses (determinado pelo algoritmo de treinamento), utilizaram-se de MLPs com retropropagação, SVMs e RBFs para alcançar o primeiro nível de diversidade. Durante a fase de treinamento utilizaram-se do *Bagging* (Bauer & Kohavi, 1999) e de uma variante do *AdaBoost* (Bauer & Kohavi, 1999), alcançando assim o nível 2.

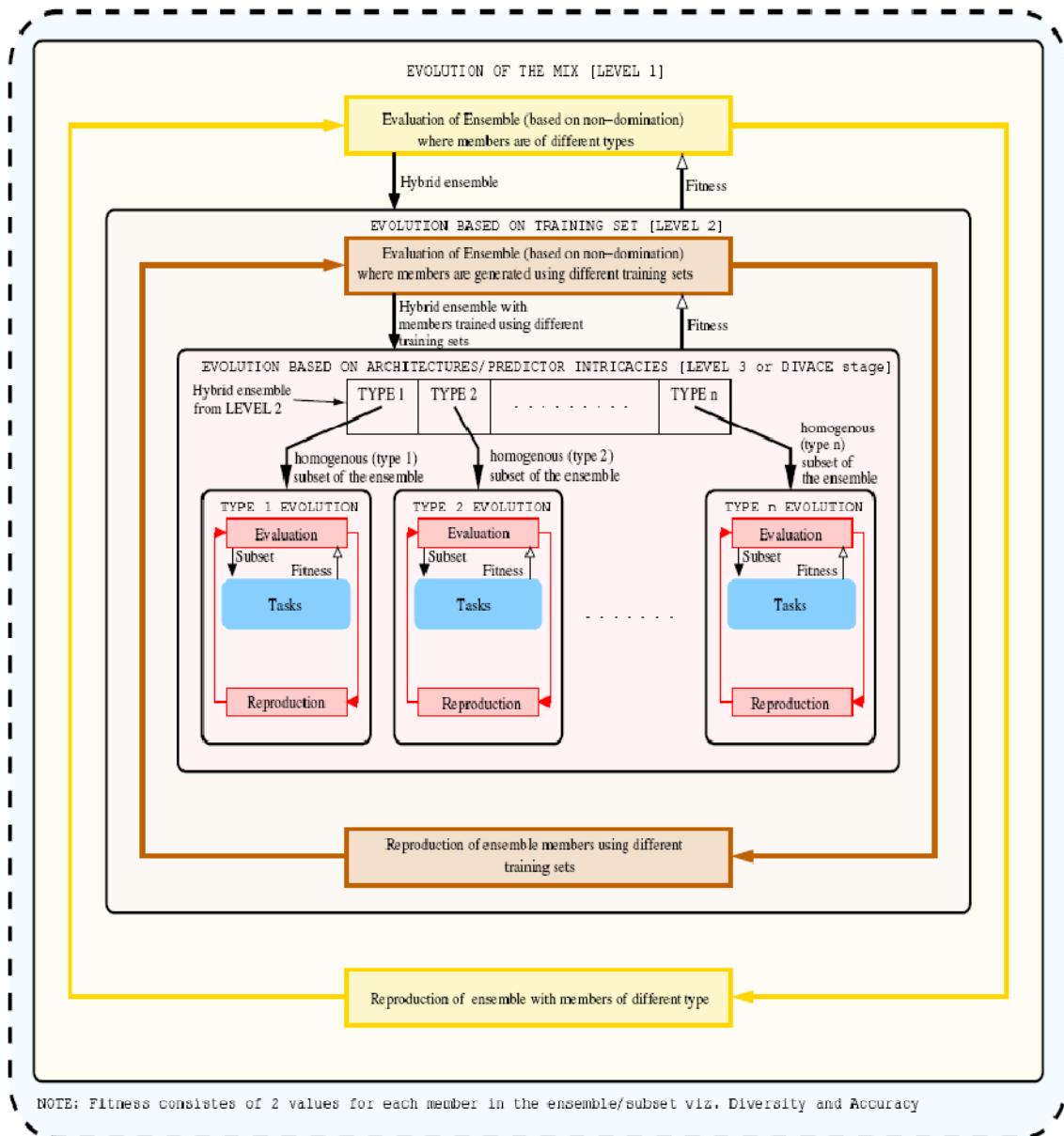


Figura 3.2. Framework DIVACE-II.⁹

⁹ Figura Extraída de (Chandra, 2004).

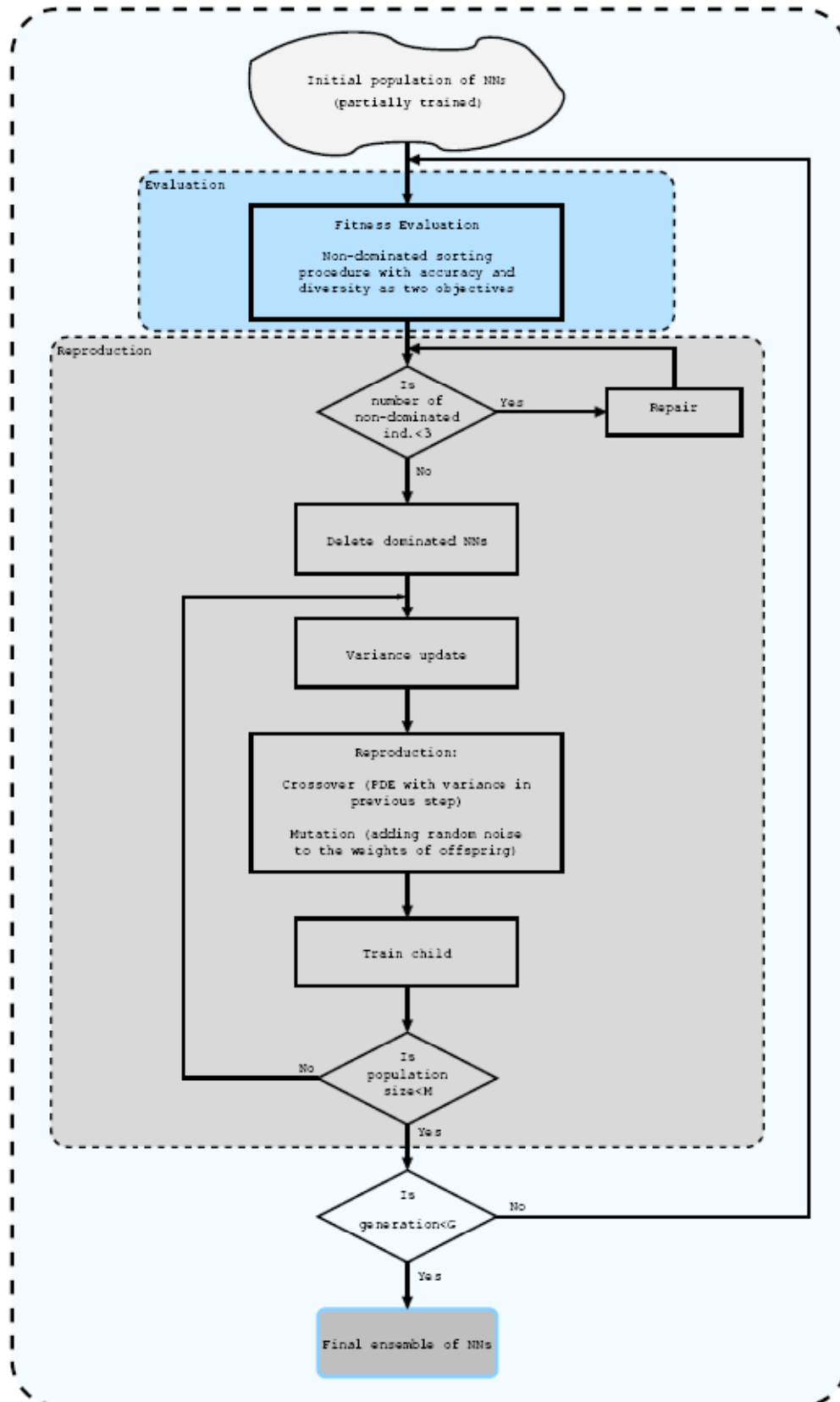


Figura 3.3. Fluxograma do DIVACE.¹⁰

¹⁰ Figura Extraída de (Chandra, 2004).

 Algoritmo 3.5. Principais passos do Algoritmo DIVACE-II.

- I. **Inicializar** a população de classificadores usando *Bagging* e a variante *AdaBoost*;
 - II. **Efetuar** o agrupamento *k*-médias (Faber, 1994) usando distância Euclidiana para criar *m* grupos e **selecionar** o melhor classificador de cada grupo formando o comitê inicial;
 - III. **Repetir até que** (número de gerações seja alcançado)
 4. **Armazenar** comitê atual se dominar o melhor comitê;
 5. **Avaliar** os indivíduos de acordo com a acurácia e a diversidade e **rotular** os não-dominados;
 6. **Selecionar** todos os indivíduos não-dominados como pais. **Gerar** o novo conjunto de treinamento, aumentando as probabilidades do aparecimento de exemplos classificador erroneamente. Onde $\frac{1}{4}$ das instâncias são geradas similar ao *AdaBoost* e as restantes escolhidas aleatoriamente do conjunto de dados original.
 7. **Se** (indivíduos não-dominados = total de indivíduos no comitê atual) **vá para 8**
 - **Gerar** um *pool* de filhos usando o conjunto de treinamento do passo anterior;
 - **Agrupar** os filhos usando *k*-médias, onde *k* é igual ao número de indivíduos dominados;
 - **Substitua** os indivíduos dominados com o melhor individual de cada grupo, somente os indivíduos do mesmo tipo podem ser substituídos. Garantindo assim, representações equivalentes para cada tipo de indutor;
 - **Vá para 9**;
 8. **Gerar** um *pool* de filhos usando o conjunto de treinamento gerado no passo 6, onde o tamanho é o número de tipos de indutores usados. Substituir o pior indivíduo (da população e um que não esteja no comitê) com o melhor do mesmo tipo desse *pool*.
 9. **Reagrupar** a população, onde o número de grupo é igual ao tamanho do comitê. Isso é feito para garantir que há somente um membro de cada grupo presente no comitê em todas as iterações. O melhor membro será incluído no novo comitê.
 10. (Opcional) Nível 3 de evolução. **Executar** o DIVACE para vários subconjuntos do novo comitê que conseqüentemente gerar um novo comitê;
 - IV. **Usar** o comitê encontrado, como comitê híbrido final;
-

Como podemos observar pelo Algoritmo 3.5, os indivíduos no DIVACE-II são classificadores. Em seus experimentos, os autores, configuraram a população inicial como composta por 60 indivíduos, sendo 20 de cada tipo utilizado – 20 MLPs, 20 RBFs e 20 SVMs. Os *pool* de comitês criados no passo 7 e no passo 8, foram respectivamente, 45 (15

de cada tipo) e 3 (1 de cada tipo). A representação genotípica de um indivíduo encontra-se na Figura 3.4. A informação armazenada pelo indivíduo é dividida em duas partes. A primeira parte, *book-keeping information*, armazena informações essenciais para o processo evolucionário, como, por exemplo, acurácia, diversidade, *flag* de dominância e grupo no qual o indivíduo pertence. A outra parte, *predictor specific information*, contém informações detalhadas sobre as características da estrutura e da função dos indutores.

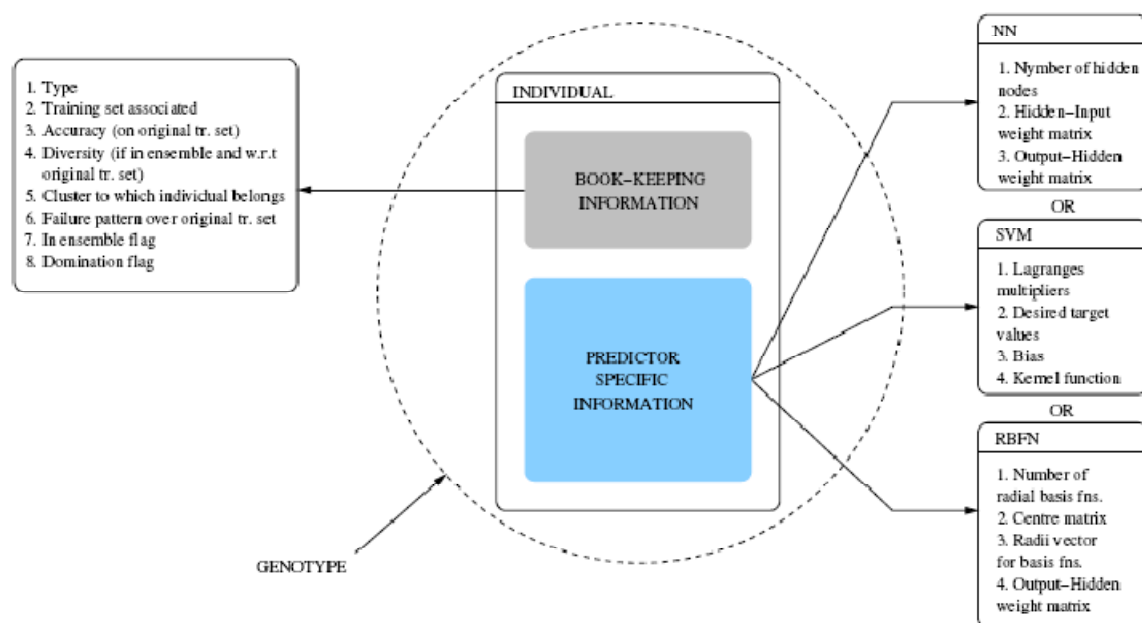


Figura 3.4. Representação genotípica de um indivíduo do DIVACE-II.¹¹

A recombinação foi o único operador genético utilizado, provavelmente, devido à dificuldade de se aplicar a mutação na representação genotípica escolhida. Entretanto, esse operador foi aplicado somente no conjunto de treinamento durante o estágio evolutivo – correspondendo à variante do *Adaboost* (Bauer & Kohavi, 1999), mencionada anteriormente, utilizada para promover a diversidade.

As medidas de diversidades utilizadas foram a *Negative Correlation Learning* (Liu & Yao, 1999) e a *Pairwise Failure Crediting* (Chandra, 2004). A *Pairwise Failure*

¹¹ Figura Extraída de (Chandra, 2004).

Crediting (PCF) pode ser utilizada tanto nos comitês (PCF para comitês) quanto individualmente nos membros (PCF para os indivíduos). A combinação da saída do comitê foi dada pelo Voto Majoritário.

Chandra e Yao (2005) compararam o desempenho do DIVACE-II com duas variações do *Memetic Pareto Artificial Neural Networks* (Abass, 2001), com o DIVACE e com o *Evolutionary ensembles with negative correlation learning* (Liu, Yao, & Higuchi, 2000) e verificaram que essa instanciação do *framework* proposto gerou comitês mais acurados do que os algoritmos analisados. De fato, a meta dos autores era verificar a robustez do DIVACE-II, comparando-o com outros algoritmos presentes na literatura. A análise do impacto da diversidade foi realizada pelo DIVACE, mas, como dito anteriormente, apenas em comitês homogêneos.

3.3 Nossa Abordagem

Como mencionado anteriormente, em nossa abordagem investigaremos dois aspectos. O primeiro aborda um único objetivo (ou apenas a acurácia ou somente a diversidade), para isso usamos um algoritmo genético mono-objetivo, enquanto que no segundo caso usamos um multiobjetivo, pois o problema abordado usa dois objetivos (acurácia e diversidade em conjunto).

Os passos genéricos dos algoritmos genéticos utilizados são os mesmos descritos, respectivamente, em Algoritmo 3.1 e Algoritmo 3.2 (NSGA-II). Em termos de implementação, tanto no caso do AG quanto do AGMO, reutilizamos os procedimentos encontrados no *framework Metaheuristic Algorithms in Java* (jMetal) (Durillo, Nebro, Luna, Dorronsoro, & Alba, 2006).

Como explicado na seção 2.1, em nossos experimentos utilizamos três tipos de algoritmos de aprendizado na formação de nosso *pool* de classificadores. Criamos 10

classificadores de cada tipo, dessa forma o nosso *pool* de classificadores é formado por 30 indutores (10 MLPs, 10 k -NNs e 10 SVMs). Com o objetivo de garantir chances iguais a todos os classificadores base de serem escolhidos durante o processo evolucionário, a população inicial foi definida como sendo formada por todos os especialistas que compõe o esse *pool*.

A representação de um indivíduo da população foi feita utilizando a representação cromossômica binária. Desse modo, temos um vetor binário de 30 posições, onde cada uma dessas posições indica se um classificador base participará (1) ou não (0) na formação do comitê. A Figura 3.5 ilustra a representação de um indivíduo da população. É importante ressaltar que não existirá nenhum indivíduo cujos todos os valores do vetor assumam zero, isto é, não existirá um comitê formado por zero especialista.

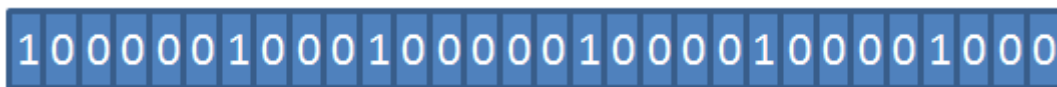


Figura 3.5. Representação cromossômica da nossa abordagem.

Dito de outra forma, ao contrário da maioria dos algoritmos genéticos, onde a população é escolhida aleatoriamente, na abordagem proposta, ela é formada sempre pelos mesmos N indivíduos que compõe o *pool* de especialistas, onde cada um dos indivíduos corresponde a uma posição como 1 (atribuindo ao classificador ao comitê) e as demais sendo 0 (não atribuindo).

Utilizamos em nossos algoritmos a recombinação de dois pontos e a mutação por alternância de *bit* (do inglês, *bit flipping*) como operadores genéticos cujas taxas foram de, respectivamente, 0,9 e 0,05. O número de iterações para a base SatImage (explicada no próximo Capítulo) utilizado foi 25.000, que é o valor padrão do jMetal. Já para as demais bases o uso desse valor padrão não levou a uma população final cujos multiclassificadores apresentassem um desempenho superior àqueles dos classificadores base. Por causa disto,

executamos uma série de experimentos em que identificamos que, para demais bases, um número de iterações de 50.000 superava o problema.

A recombinação de dois pontos funciona da seguinte forma: dados dois indivíduos da população, escolhemos aleatoriamente dois pontos quaisquer, em seguida efetuamos a troca do material genético presente entre esses pontos dos indivíduos. A Figura 3.6a ilustra esse processo com um cromossomo binário de seis posições. Devido ao tamanho do cromossomo, escolhemos um processo de recombinação de dois pontos, ao invés do tradicional (apenas um ponto). Desse modo, aumentaremos a probabilidade de diversidade genética na nossa população. A mutação escolhida foi a tradicional, a qual escolhe aleatoriamente um ponto e inverte o seu valor binário, conforme ilustrado na Figura 3.6b.

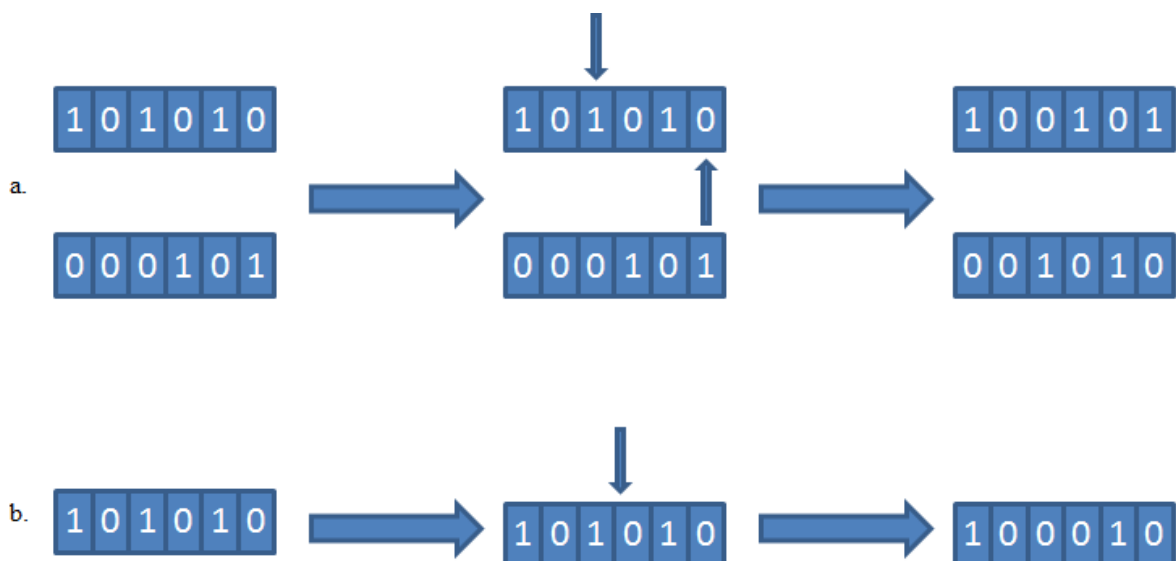


Figura 3.6. Ilustração dos operadores genéticos utilizados.

Através dessa abordagem, esperamos analisar se existe algum conjunto de objetivos mais indicado para o processo de criação de comitês, se existe algum método de combinação superior a outro e encontrar uma diretriz para escolha de comitês gerados por AGMO. Como métodos de combinação, utilizamos o método Soma (seção 2.4.1) e o classificador *Naive Bayes* (seção 2.4.2). As medidas de diversidades analisadas foram a

Estatística-Q (seção 2.2.1), o coeficiente de correlação (seção 2.2.2) e a concordância (seção 2.2.3).

3.4 Considerações Finais

Em relação aos 2 trabalhos relacionados ao nosso apresentados neste Capítulo, o trabalho apresentado em (dos Santos, Sabourin, & Maupin, 2006), em contraste ao nosso, utiliza apenas comitês homogêneos. Já o estudo apresentado em (Chandra & Yao, 2004), utilizando o DIVACE-II foi desenvolvido no contexto de comitês heterogêneos. No entanto, diferentemente da nossa análise, não foi investigado o dilema diversidade-precisão: o objetivo dos autores foi de comparar o desempenho do DIVACE-II com outras abordagens existentes na literatura.

Em termos do algoritmo genético usado em nossa abordagem em relação aos trabalhos descritos nos parágrafos anteriores, Santos, Sabourin e Maupin (2006) utilizaram a representação binária para descrever um comitê. De forma análoga, utilizamos esse tipo de representação para descrever um multiclassificador. No entanto, enquanto utilizamos um vetor de 30 posições para representar um comitê (indivíduo) – em que cada posição representa se um classificador está ativo (1) ou não (0) –, Santos, Sabourin e Maupin (2006) utilizaram um vetor de 100 posições. Isso reflete o fato de que utilizamos 30 classificadores base, enquanto eles utilizaram 100 para compor o *pool* de classificadores base. Diferentemente dessas abordagens, a representação usada por Chandra e Yao (2005) foi a objeto para descrever um classificador.

Utilizamos 30 indivíduos na nossa população inicial, em que cada indivíduo desses é formado por apenas um classificador presente no nosso *pool* de classificadores. Dessa forma, temos 10 indivíduos formados por *k*-NNs, 10 indivíduos constituídos por SVMs e 10 indivíduos por redes neurais do tipo MLP. Santos, Sabourin e Maupin (2006) usaram

128 indivíduos cujo número de classificadores presentes em cada comitê é encontrado aleatoriamente. Chandra e Yao (2005) utilizaram 60 indivíduos, sendo 20 classificadores SVMs, 20 classificadores redes neurais do tipo MLP e 20 classificadores redes neurais do tipo RBF.

Apesar de um cromossomo representar um multiclassificador, a saída final, mesmo para o AGMO é um único indivíduo, aquele dito ser o melhor em relação à acurácia. Enquanto que na nossa abordagem, a saída é uma coleção de conjunto de classificadores. No DIVACE-II, os indivíduos são classificadores, sendo sua saída final um multiclassificador.

Em relação aos operadores genéticos, a abordagem de Santos, Sabourin e Maupin (2006) utiliza a mutação e recombinação semelhante a nossa abordagem. A diferença entre essas abordagens encontra-se no operador de recombinação utilizado. Enquanto escolhemos utilizar a recombinação de dois pontos, em (dos Santos, Sabourin, & Maupin, 2006) a recombinação de um ponto foi escolhida, mesmo trabalhando com indivíduos grandes (100 posições). Durante o processo de geração de classificadores acurados, essa recombinação escolhida pode ter causado uma parada prematura nos algoritmos genéticos, o que explicaria o porquê da diversidade não contribuir tão positivamente, como era esperado. No DIVACE-II, o único operador genético utilizado foi à recombinação, aplicada no conjunto de treinamento com o intuito de gerar um novo subconjunto de treinamento a ser utilizado na construção de um novo classificador base. Sendo assim, o classificador base é treinado em tempo de execução. Dessa forma, o nosso algoritmo se tornar mais simples, uma vez que não é necessário o treinamento de novos classificadores à medida que avançamos nas iterações, pois a aplicação dos nossos operadores genéticos corresponde a apenas alterações na representação cromossômica, ou seja, a inserção,

remoção ou troca dos classificadores que compõe o indivíduo. A Tabela 3.1 apresenta a síntese das características dessas abordagens.

Características	Nossa Abordagem	dos Santos et al	DIVACE-II
Tipo de Análise	Estudo do Dilema diversidade- acurácia em comitês heterogêneos	Estudo do Dilema diversidade- acurácia em comitês homogêneos	Robustez do algoritmo frente a outros presentes na literatura
Algoritmos de Aprendizado utilizados	MLP, SVM e k -NN	Somente k -NN	MLP, SVM e RBF
Tamanho da população inicial	30 (10 de cada tipo)	128	60 (20 de cada tipo)
Representação Cromossômica	Binária	Binária	Objeto
Recombinação Aplicada	Recombinação de dois pontos	Recombinação de um ponto	Variante do <i>Adaboost</i> no conjunto de treinamento juntamente com escolhas de instâncias aleatórias
Taxa de Recombinação	0,9	0,8	Não se Aplica
Mutação Aplicada	Alternância de <i>bit</i>	Alternância de <i>bit</i>	Não se Aplica
Taxa de Mutação	0,05	0,01	Não se Aplica
Saída dos Algoritmos	Conjunto de comitês não dominados	O melhor comitê, de acordo com a acurácia	Um comitê
Número de iterações	25000 (SatImage) e 50000 (Splice e SCOP)	1000	Não informado
Métodos de combinação	Soma e Classificador <i>Naive Bayes</i>	Voto Majoritário	Voto Majoritário
Medidas de diversidade	Estatística-Q, Correlação e Concordância	As dez analisadas em (Kuncheva & Whitaker, 2003), 1 definida em (Ruta & Gabrys, 2005) e 1 analisada em (Zenobi & Cunningham, 2001)	NCL e PCF (para comitê e para indivíduos)

Tabela 3.1. Síntese das diferenças entre a nossa abordagem, a dos Santos et al e o DIVACE-II.

Capítulo 4

Experimentos

Com o intuito de validar nossa abordagem, empregamos o procedimento *10-fold-cross-validation* nos classificadores base (Mitchell, 1997). Desse modo, todos os resultados apresentados nesta dissertação referem-se a média sobre 10 diferentes conjuntos de teste. A significância estatística desses resultados é verificada por meio de teste de hipótese (teste-t). Nesse contexto, o nível de confiança (α) utilizado foi 0,05.

Nas próximas seções, apresentaremos as três bases de dados utilizadas nas execuções dos experimentos, bem como a metodologia de desenvolvimento dos experimentos utilizada.

4.1 Base de Dados

Utilizamos três bases de dados retiradas do repositório do UCI (Blake & Merz, 1998): *Landsat Satellite* (chamada nesta dissertação apenas de *SatImage*), Sequência Genética de Junção de *Splice* (referenciada nesta dissertação apenas de *Splice*) e Classificação Estrutural de Proteínas (referenciada nesta dissertação como SCOP). A partir dessas bases de dados analisamos diferentes características, pois a *SatImage* possui grande quantidade de exemplos utilizando atributos numéricos distribuídos em várias classes, mas utilizando uma pequena dimensão; a *Splice* possui uma grande quantidade de instâncias usando atributos categóricos em poucas classes, mas em uma alta dimensão; e *SCOP* possui uma pequena quantidade de padrões de alta dimensão em várias classes, usando valores numéricos. A Tabela 4.1 sintetiza essas informações.

Base de Dados	Número de Instâncias	Número de Classes	Distribuição de instâncias por classe	Dimensionalidade (sem considerar a classe)
<i>SatImage</i>	5850	6	red soil (1394), cotton crop (638), grey soil (1234), damp grey soil (570), soil with vegetation stubble (643), very damp grey soil (1371)	4 atributos numéricos
<i>Splice</i>	2900	3	Sem junção (1505) exon/intron (697) intron/exon (698)	60 atributos nominais
<i>SCOP</i>	530	5	all- α (100), all- β (161), α/β (189), $\alpha+\beta$ (40), Small (40)	126 atributos numéricos

Tabela 4.1. Resumo das bases de dados.

4.1.1 *SatImage*

Essa base de dados foi gerada a partir de dados de imagens tiradas pelo satélite *Landsat Multi-Spectral Scanner (Landsat MSS)*. Um único quadro do Landsat MSS consiste de quatro imagens da mesma cena, mas em faixas espectrais diferentes. Duas dessas faixas são regiões visíveis (correspondendo a aproximadamente as regiões verde e vermelha) e as outras duas estão perto da região de infra-vermelho. Cada pixel encontra-se no intervalo de $[0,255]$, onde o limite inferior corresponde a cor preta e o superior a cor branca. A imagem original contém 2340×3380 pixels, um pixel corresponde a aproximadamente $80m \times 80m$ de resolução espacial. No entanto, a base de dados utilizada consiste de apenas 82×100 pixels, ou seja, apenas uma pequena área da imagem, onde cada linha de dado corresponde a uma vizinhança de 3×3 pixels dessa imagem.

A *SatImage* possui 5 classes distribuídas em 5850 padrões distribuídos conforme mostrado na Tabela 4. 2.

<i>Classe</i>	<i>Quantidade de exemplos</i>	<i>Proporção (%)</i>
red soil	1394	23,82
cotton crop	638	10,92
grey soil	1234	21,1
damp grey soil	570	9,73
soil with vegetation stubble	643	10,99
very damp grey soil	1371	23,43

Tabela 4. 2. Distribuição dos padrões da base de dados *SatImage*.

Há duas variações da base dados – uma com 37 atributos e a outra com 5 atributos. Um desses atributos é a classe, restando assim, respectivamente, 36 e 4 atributos para cada variação. Na segunda, os 4 atributos correspondem a quatro faixas de valores espectrais, onde cada espectro é calculado a partir 9 pixels de vizinhança, 9 pixels esses presentes explicitamente na primeira variação. Portanto, as duas variações são equivalentes.

Através de um estudo empírico, usando os algoritmos de aprendizados escolhidos para os experimentos deste trabalho, verificamos que o comportamento dos classificadores base se assemelha. Entretanto, como o tempo de treinamento levado para a primeira variação é bem superior a segunda, por questão de otimização, a segunda variação foi escolhida para ser utilizada em nossos experimentos. Para esta variação, os atributos encontram-se normalizados entre $[-1,1]$. A Tabela 4. 3 exibe o resumo desse problema.

Problema	Identificação de terrenos presentes em uma imagem.
Dado	Conjunto de valores espectrais dos pixels presentes na imagem.
Faça	Gerar um indutor capaz de determinar o tipo de terreno presente em uma parte da imagem analisada.

Tabela 4. 3. Resumo do problema da base de dados *SatImage*.

4.1.1. Splice

Em organismos eucariotos, o reconhecimento completo dos genes envolve também a identificação de suas regiões que codificam proteínas (os exons) e de porções de DNA que intermediam os exons e não produzem proteínas (os introns). Uma vez que se faz necessário a demarcação precisa dos segmentos de DNA que são efetivamente traduzidos em proteínas daqueles que não são, o reconhecimento das fronteiras entre esses elementos é de grande importância.

Uma terminologia utilizada neste domínio referencia as bordas exon/introns como regiões “doadoras” e as intron/exon como regiões “receptoras” (Alberts, Johnson, & Walter, 2004). Na identificação de doadores e receptores, comumente são utilizadas evidências reportadas em estudos da Biologia. Ambas as regiões caracterizam-se por possuir, em geral, um par de nucleotídeos que se conservam no intron da junção, sendo guanina e timina para doadores e adenina e guanina para receptores. Porém, esses mesmos pares podem está presentes nos exons, dificultando assim a sua utilização para o processo de a classificação dessas regiões.

A *Splice* consiste em 2900 padrões distribuídos desbalanceadamente em 3 classes, conforme especifica a Tabela 4. 4.

<i>Classe</i>	<i>Quantidade de exemplos</i>	<i>Proporção (%)</i>
Sem junção	1505	51,88
<i>exon/intron</i>	697	24,04
<i>intron/exon</i>	698	24,08

Tabela 4. 4. Distribuição dos padrões da base de dados *Splice*.

Cada padrão é constituído por 60 atributos nominais, que representam os possíveis nucleotídeos (guanina, citosina, adenina e timina) do DNA usados no processo da criação

protéica, e um atributo numérico – representando a classe. A Tabela 4.5 ilustra o resumo desse problema.

Problema	Identificação de sítios de <i>splicing</i> .
Dado	Conjunto de seqüências de DNA (de organismos eucariotos) de tamanho fixo com fronteiras do tipo intron/exon, exon/intron, e sem nenhum desses sítios.
Faça	Gerar um classificador capaz de determinar se uma janela de tamanho fixo de uma seqüência de DNA possui uma fronteira intron/exon, exon/intron, ou nenhuma delas.

Tabela 4.5. Resumo do problema da base de dados *Splice*.

4.1.2. SCOP

Esta base contém a classificação de estruturas conhecidas de proteínas de acordo com os seus relacionamentos evolucionários e estruturais. Nesse contexto, uma proteína é representada por um conjunto de vetores baseados em sete propriedades físico-químicas e estruturais dos aminoácidos ao longo da seqüência: hidrofobicidade, polaridade, polarizabilidade, predição da estrutura secundária, volume de Van der Waals normalizado, composição de aminoácidos da seqüência da proteína e comprimento da proteína. Cada propriedade é representada por 18 atributos.

No contexto da base de dados estudada, as proteínas podem ser classificadas em cinco classes: all- α , all- β , α/β , $\alpha+\beta$ e *small*. Sendo assim, uma instância representando uma proteína é formada por 127 atributos: 126 características e uma classe. Ao todo a base de dados possui 584 instâncias normalizadas entre [0,1] e distribuídos desbalanceadamente conforme a Tabela 4.6.

<i>Classe</i>	<i>Quantidade de exemplos</i>	<i>Proporção (%)</i>
all- α	100	18,83
all- β	160	30,32
α/β	189	35,61
$\alpha+\beta$	41	7,54
Small	40	7,70

Tabela 4.6. Distribuição dos padrões da base de dados SCOP.

Dessas classes citadas, as técnicas de aprendizado de máquina apresentam a maior dificuldade de reconhecimentos de instâncias da classe $\alpha+\beta$, pois um número elevado de instâncias é erroneamente classificado nas classes $\text{all-}\alpha$ e $\text{all-}\beta$ (Chinasamy, Sung, & Mittal, 2004). A Tabela 4.7 ilustra o resumo do problema associado a essa base de dados.

Problema	Identificação de estrutura de proteínas de acordo com os seus relacionamentos evolucionários e estruturais.
Dado	Conjunto de vetores baseado em sete propriedades físico-químicas e estruturais dos aminoácidos ao longo da seqüência, onde cada propriedade é representada por 18 atributos.
Faça	Gerar um classificador capaz de determinar a estrutura da proteína.

Tabela 4.7. Resumo do problema da base de dados *Splice*.

4.2 Metodologia de desenvolvimento dos experimentos

Como mencionado anteriormente, o objetivo deste trabalho é investigar o comportamento de comitês heterogêneos mediante o dilema diversidade-accurácia. Para isso, investigamos o impacto de se escolher classificadores base utilizando-se: apenas a accurácia; apenas a diversidade; e, conjuntamente, a accurácia e a diversidade. A cada experimento que envolvia a diversidade, este foi executado três vezes, alterando-se em sua configuração as medidas de diversidade usadas para uma das seguintes: Estatística-Q, Correlação ou Concordância. Além disso, os mesmos experimentos foram executados utilizando-se os métodos de combinação Soma e o classificador *Naive Bayes*.

Utilizamos os algoritmos de aprendizado: k -NN, SVM e MLP com retro-propagação. Para cada um desses algoritmos, foram treinados 10 (dez) classificadores base utilizando o método de amostragem *10-fold-cross-validation*. Ou seja, o nosso *pool* de classificadores base é formado por 30 classificadores por *fold*, 10 de cada tipo, e cabe o algoritmo genético selecionar – de acordo com a função objetivo – os classificadores que formarão o comitê.

Durante o processo de treinamento, para todas as bases de dados, tentamos construir classificadores base que apresentassem um desempenho, em termos de accurácia, apenas um pouco melhor do que classificadores cujas predições sejam aleatórias – por exemplo, uma taxa de acerto ligeiramente superior a 60%. A idéia é observar se, mesmo com

classificadores base apresentando uma baixa acurácia (entre 60% e 70%), os multiclassificador gerados superam essa deficiência.

Como dito anteriormente, utilizamos dois algoritmos genéticos (um mono-objetivo e um multiobjetivo) na escolha de componentes (classificadores base) que comporão o comitê. O mono-objetivo quando investigamos o impacto de se escolher os classificadores base utilizando-se: ou apenas a acurácia ou somente a diversidade. Já o multiobjetivo, quando investigamos o impacto de se escolher esses componentes utilizando conjuntamente essas duas funções objetivas. Para cada um desses algoritmos a população inicial possuía 30 indivíduos, o número de iterações utilizado foi 25.000 para a base de dados *SatImage* e 50.000 para as bases de dados *Splice* e *SCOP* e as taxas de recombinação e mutação foram, respectivamente, 0,9 e 0,05. A explicação do por que desses valores encontra-se na seção 3.3.

Devido aos algoritmos genéticos usados serem não-determinísticos, ao executarmos um algoritmo genético com os mesmos parâmetros de configuração, podemos obter como resultado dois indivíduos diferentes. Dessa forma, com o intuito de validar a nossa abordagem, efetuamos a execução de cada experimento dez vezes.

Capítulo 5

Resultados e Discussão

Com o intuito de colocarmos nossos resultados em perspectiva em relação aos dos classificadores base utilizados, a próxima seção, seção 5.1, ilustra para cada conjunto de dados (explicados na seção 4.1), a melhor acurácia encontrada para esses classificadores, bem como a pior acurácia e a média das acurácias. Na seção 5.2 encontram-se os resultados obtidos pelos comitês, bem como uma discussão sobre esses resultados encontrados em termos de acurácia.

5.1 Classificadores base

A Tabela 5.1 ilustra os melhores resultados, bem como os piores e as médias de cada tipo de classificador para a base *SatImage*. De forma análoga, temos na Tabela 5.2 e na Tabela 5.3, respectivamente, os resultados para as bases *Splice* e SCOP.

<i>Tipo</i>	<i>Acurácia</i>		
	Melhor Classificador	Pior Classificador	Média dos Classificadores
k-NN	85,24	82,28	84,09
SVM	68,34	63,88	67,28
MLP	83,85	83,67	83,76

Tabela 5.1. Acurácias dos classificadores base utilizado na *SatImage*.

<i>Tipo</i>	<i>Acurácia</i>		
	Melhor Classificador	Pior Classificador	Média dos Classificadores
k-NN	79,33	72,94	75,03
SVM	88,13	82,80	85,13
MLP	89,71	87,61	88,93

Tabela 5.2. Acurácias dos classificadores base utilizado na *Splice*.

<i>Tipo</i>	<i>Acurácia</i>		
	Melhor Classificador	Pior Classificador	Média dos Classificadores
k-NN	68,41	61,87	64,77
SVM	69,35	67,01	67,95
MLP	67,94	61,12	63,34

Tabela 5.3. Acurácias dos classificadores base utilizado na SCOP.

Como mencionado na seção 4.2, durante o processo de treinamento, para todas as bases tentamos construir classificadores base que apresentassem um desempenho, em

termos de acurácia, apenas um pouco melhor do que classificadores cujas predições sejam aleatórias – por exemplo, uma taxa de acerto entre 60% e 70%. Para a base SCOP, variando-se os parâmetros dos algoritmos, conseguimos obter classificadores cujas acurácias média estão no intervalo [60%,70%], como desejado. Em contraste, para a base *Splice*, mesmo variando-se os parâmetros dos algoritmos, as médias dos diferentes tipos de classificadores gerados (k -NN, redes neurais do tipo MLP e SVM) foram superiores a 70%. No caso da base *SatImage*, apenas as SVMs apresentaram um desempenho dentro do limite inferior estabelecido.

Como podemos observar, a base de dados SCOP foi a única base que conseguimos obter todos os classificadores no intervalo desejado. Isso, possivelmente, se deve ao fato dessa base ser desbalanceada e, como dito anteriormente, a classe $\alpha+\beta$, ser confundida com as classes all- α e all- β , para um número elevado de instâncias.

5.2 Comitês

Para termos um panorama mais amplo do comportamento dos algoritmos genéticos na geração dos comitês, apresentamos nessa seção os melhores e os piores resultados encontrados. Nesse contexto, se as médias dos melhores forem próximas das dos piores, isso significa que a população como todo convergiu para um ponto comum (mesma solução). Nesse caso, as médias também deveriam estar em torno de um valor comum. Por outro lado, grandes diferenças entre esses valores, juntamente com a análise das médias, é um indício que o algoritmo não levou à população, como um todo, para um ponto ótimo. Dessa forma, a população final é mais heterogênea. Em alguns casos, isso pode refletir em uma falta de robustez do algoritmo ou, talvez, em um número de iterações insuficiente.

É importante ressaltar que, como mencionado na seção 4.2, cada um dos classificadores base utilizados como membros de um multiclassificador dos nossos algoritmos genéticos, corresponde a uma média de 10 classificadores devido a utilização do *10-fold-cross-validation* na criação desses classificadores. Sendo assim, cada resultado apresentado nas Tabelas Tabela 5.4, Tabela 5.13 e Tabela 5.22 como melhor acurácia, pior acurácia e acurácia média correspondem, respectivamente a média das 10 execuções do genético em relação a: acurácia dos melhores comitês envolvidos, acurácia dos piores multiclassificadores envolvidos e a média das acurácias dos comitês envolvidos. De forma

análoga, temos nas Tabelas Tabela 5.5, Tabela 5.14 e Tabela 5.23 os resultados apresentados como melhor diversidade, pior diversidade e média das diversidades.

Nesse contexto, os testes-t foram realizados levando-se em consideração os valores da diversidade e acurácia da seguinte forma:

- Entre os comitês obtidos com diferentes objetivos. Nesse caso, efetuamos três abordagens. Na primeira, comparamos os resultados encontrados utilizando apenas a diversidade com os obtidos usando apenas a acurácia, no intuito de verificarmos se apenas a uma medida de diversidade pode substituir a acurácia como função objetivo. Com o intuito de verificarmos se a diversidade ajuda no processo de formação de comitês acurados, na segunda abordagem, efetuamos um estudo comparativo entre os resultados encontrados nos multiobjetivos com os obtidos com apenas a acurácia. A última abordagem, objetivando encontrar uma medida de diversidade que se destaque das demais, realiza um estudo comparativo, entre elas. Nesse caso, efetuando a comparação dos resultados utilizando apenas a diversidade e dos utilizando a diversidade mais acurácia.
- Entre os comitês encontrados utilizando-se os métodos de combinação Soma e classificador *Naive Bayes* (representados apenas como *Naive*). Isso foi feito para verificarmos se existia algum método de combinação mais adequado para lidar com o dilema diversidade-acurácia no comportamento da criação de comitês.
- Entre os melhores e os piores comitês com os mesmos objetivos e entre os melhores e as médias de todos os comitês para termos um panorama mais amplo do comportamento dos algoritmos genéticos na geração dos comitês.

5.2.1 *SatImage*

As acurácias encontradas nos experimentos executados com a base de dados *SatImage* são mostradas pela Tabela 5.4. Efetuando-se a análise estatística dos resultados encontrados. De modo geral, os resultados das diversidades combinadas com os da acurácia superam aqueles obtidos apenas com a acurácia. Nesse contexto, por exemplo, a acurácia combinada com a Estatística-Q, utilizando-se do classificador Naive como

método de combinação, mostrou-se superior no Melhor ($90,12 \pm 2,08$; $p\text{-valor}^{12} \approx 0,0017$), no Pior ($86,80 \pm 3,65$; $p\text{-valor} \approx 0,0315$) e na Média ($88,21 \pm 2,00$; $p\text{-valor} \approx 0,0066$) dos resultados. Usando o combinador Soma, mostrou-se superior para o Melhor ($89,70 \pm 3,48$; $p\text{-valor} \approx 0,0205$), o Pior ($85,99 \pm 3,09$; $p\text{-valor} \approx 0,0097$) e a Média ($87,01 \pm 2,06$; $p\text{-valor} \approx 0,0035$). Para a acurácia combinada com a Concordância, apenas o valor encontrado no Pior ($86,16 \pm 3,02$; $p\text{-valor} \approx 0,0488$) quando utilizamos o classificador Naive, como método de combinação, mostrou-se estatisticamente superior a apenas acurácia. No caso da acurácia combinada com a Correlação, obtivemos superioridade estatística em relação a apenas a acurácia quando usamos o classificador Naive apenas para o Melhor ($89,30 \pm 1,13$; $p\text{-valor} \approx 0,0031$), já utilizando o Soma obtivemos essa superioridade para o Melhor ($89,00 \pm 3,04$; $p\text{-valor} \approx 0,0340$) e a Média ($87,07 \pm 2,05$; $p\text{-valor} \approx 0,0029$).

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	86,00±3,02	86,12±2,73	82,00±3,08	83,02±3,60	84,12±1,78	85,05±2,57
Correlação	85,19±2,00	85,07±1,98	81,89±2,16	82,07±3,18	83,06±3,28	84,44±1,42
Concordância	84,96±3,77	85,00±1,08	81,12±1,97	82,19±2,97	83,12±3,46	84,35±1,02
Estatística-Q	85,02±4,08	85,26±2,00	82,16±3,74	83,05±3,01	82,98±3,70	84,95±2,22
Acurácia e Correlação	89,00±3,04	89,30±1,13	85,10±3,70	85,07±2,50	87,07±2,05	87,12±3,14
Acurácia e Concordância	88,17±3,80	88,07±3,01	85,04±4,00	86,16±3,02	86,21±4,08	87,17±4,08
Acurácia e Estatística-Q	89,70±3,48	90,12±2,08	85,99±3,09	86,80±3,65	87,01±2,06	88,21±2,00

Tabela 5.4. Acurácias obtidas com a base de dados *SatImage*.

Para os demais Objetivos (Estatística-Q, Concordância e Correlação) em relação à Acurácia, não foram encontradas evidências estatísticas suficientes para afirmar que os comitês criados são diferentes daqueles encontrados usando apenas a acurácia. Desse modo, não podemos ratificar os resultados obtidos por Kuncheva & Whitaker (2003), que nos informa que a diversidade é uma medida inadequada para substituir a acurácia no processo de criação de comitês. No Capítulo 6, é dada uma a possível razão para essa diferença ter acontecido.

Os testes de hipóteses entre os comitês que usam diferentes medidas de diversidade (Correlação com Concordância; Correlação com Estatística-Q; Concordância com

¹² É a probabilidade, calculada supondo-se que H_0 é verdadeira, de que a estatística de teste assumisse um valor tão ou mais extremo do que o valor realmente observado. Um $P\text{-valor}$ pequeno é uma evidência contrária à H_0 , pois afirmar que o resultado observado é improvável de ocorrer quando H_0 é verdadeira (Moore, 2005).

Estatística-Q; Acurácia e Correlação com Acurácia e Concordância; Acurácia e Correlação com Acurácia e Estatística-Q; e Acurácia e Concordância com Acurácia e Estatística-Q) nos mostram que não há evidências estatísticas suficientes para afirmar que uma medida de diversidade é superior as outras, mesmo quando combinadas com a acurácia. Chegamos a essa mesma conclusão, ao efetuarmos os mesmos testes estatísticos nos comitês utilizando-se os valores da diversidade, ao invés dos da acurácia. Esses resultados corroboram aqueles apresentados em (Kuncheva & Whitaker, 2003), pois não podemos afirmar que, entre as medidas de diversidade analisadas, existe uma medida melhor do que a outra para a criação de comitês tanto no aspecto de diversidade quanto no de acurácia.

As diversidades encontradas nos experimentos executados com a base de dados *SatImage* são mostradas pela Tabela 5.5. É importante ressaltarmos que para os resultados com dois objetivos – acurácia mais uma medida de diversidade – o melhor dos comitês em relação à diversidade corresponde ao pior em relação à acurácia. Da mesma forma, o pior dos comitês em relação à diversidade, corresponde ao melhor em relação à acurácia. Isso deve ao fato de trabalharmos apenas com soluções não dominadas (Capítulo 3). No nosso caso, temos um comitê superior ao outro em apenas um objetivo, mas inferior em relação ao outro objetivo.

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA	NA	NA
Correlação	0,67±0,18	0,66±0,21	0,77±0,15	0,76±0,17	0,73±0,14	0,70±0,10
Concordância	0,68±0,11	0,65±0,18	0,76±0,19	0,75±0,09	0,74±0,18	0,69±0,11
Estatística-Q	0,66±0,21	0,66±0,19	0,77±0,14	0,76±0,10	0,75±0,16	0,70±0,15
Acurácia e Correlação	0,65±0,13	0,64±0,15	0,78±0,20	0,79±0,09	0,74±0,13	0,76±0,12
Acurácia e Concordância	0,64±0,19	0,63±0,12	0,77±0,17	0,77±0,17	0,74±0,20	0,74±0,15
Acurácia e Estatística-Q	0,65±0,21	0,64±0,15	0,76±0,15	0,77±0,21	0,72±0,14	0,75±0,17

Tabela 5.5. Diversidades obtidas com a base de dados *SatImage*.

As Tabela 5.6 e Tabela 5.13 mostram o número de vezes que cada função objetivo mostrou-se estatisticamente diferente em relação às demais funções, analisando-se, respectivamente, os valores da acurácia e da diversidade. AC, COR, CON, Q, AC + COR, AC + CON e AC + Q significam, respectivamente, Acurácia; Correlação; Concordância; Acurácia e Correlação; Acurácia e Concordância; e Acurácia e Estatística-Q. Os valores *a/b*

significam a quantidade de vezes que a hipótese nula foi recusada (*a*) e o número de vezes que a hipótese nula não foi rejeitada (*b*).

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	0/6	0/6	0/6	3/3	1/5	6/0
COR	0/6	–	0/6	0/6	NA	NA	NA
CON	0/6	0/6	–	0/6	NA	NA	NA
Q	0/6	0/6	0/6	–	NA	NA	NA
AC + COR	3/3	NA	NA	NA	–	0/6	0/6
AC + CON	1/5	NA	NA	NA	0/6	–	0/6
AC + Q	6/0	NA	NA	NA	0/6	0/6	–

Tabela 5.6. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base *SatImage*, levando-se em consideração aos valores da acurácia.

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	NA	NA	NA	NA	NA	NA
COR	NA	–	0/6	0/6	NA	NA	NA
CON	NA	0/6	–	0/6	NA	NA	NA
Q	NA	0/6	0/6	–	NA	NA	NA
AC + COR	NA	NA	NA	NA	–	0/6	0/6
AC + CON	NA	NA	NA	NA	0/6	–	0/6
AC + Q	NA	NA	NA	NA	0/6	0/6	–

Tabela 5.7. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base *SatImage*, levando-se em consideração aos valores da diversidade.

Ao efetuarmos o teste-t entre os métodos de combinação (Soma e classificador Naive), não encontramos evidências estatísticas suficientes para refutarmos a hipótese nula. Dito de outra forma, não temos evidências estatísticas suficientes para diferenciarmos os comitês encontrados usando o método de combinação Soma daqueles encontrados utilizando o classificador Naive. Nesse contexto, levando-se em consideração a acurácia, o menor *p-valor* (*p-valor* $\approx 0,1333$) foi encontrado quando comparamos a Média dos desempenhos dos comitês encontrados usando apenas a Estatística-Q. Levando-se em consideração a diversidade, o menor *p-valor* (*p-valor* $\approx 0,4632$) foi encontrado quando comparamos a Média dos desempenhos dos comitês encontrados usando apenas *Correlação*.

Analisando o comportamento dos melhores multiclassificadores em relação aos piores e a média, levando-se em consideração a acurácia, vimos que, de forma geral, os melhores são estatisticamente superiores aos piores, mas não o são em relação à média dos comitês envolvidos. A Tabela 5.8 sintetiza os *p-valores* encontrados ao efetuarmos o teste-t

entre os melhores com as médias, *Melhor-Média*, e entre os melhores com os piores (*Melhor-Pior*). Os valores em destacados são aqueles cujos *valores-p* foram menores que o coeficiente α . Entre aqueles que são maiores, chamamos a atenção para os valores encontrados usando a *Estatística-Q*, no caso do *Melhor-Média* (*p-valor* $\approx 0,0690$) e, no caso, do *Melhor-Pior*, chamamos atenção para os encontrados usando em *Acurácia e Estatística-Q* (*p-valor* $\approx 0,0507$) e *Acurácia e Correlação* (*p-valor* $\approx 0,0535$), pois apesar da hipótese nula não ter sido rejeitada, o valor de *p-valor* encontra-se muito perto da margem de erro estabelecida.

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	0,1071	0,3787	0,0089	0,0436
Correlação	0,0965	0,4242	0,0023	0,0208
Concordância	0,2704	0,2643	0,0105	0,0115
Estatística-Q	0,2567	0,6941	0,1196	0,0690
Acurácia e Correlação	0,1133	0,0535	0,0190	0,0001
Acurácia e Concordância	0,2809	0,5814	0,0896	0,1736
Acurácia e Estatística-Q	0,0497	0,0507	0,0213	0,0223

Tabela 5.8. *P-valores* encontrados ao analisar a relação entre os melhores e os demais da base *SatImage*, levando-se em consideração aos valores da acurácia.

Ao fazermos a mesma análise, levando-se em consideração a diversidade, vimos que não encontramos evidências estatísticas suficientes para refutarmos a hipótese nula tanto em relação aos melhores com as médias (*Melhor-Média*) quanto em relação aos melhores com os piores (*Melhor-Pior*). A Tabela 5.9 sintetiza os *p-valores* encontrados ao efetuarmos o *teste-t* entre os melhores com as médias, *Melhor-Média*, e entre os melhores com os piores (*Melhor-Pior*). Os valores em destacados são aqueles cujos *p-valores* foram menores que o coeficiente α .

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA
Correlação	0,5452	0,3487	0,3487	0,2570
Concordância	0,8117	0,1985	0,1985	0,2570
Estatística-Q	0,7695	0,3065	0,3065	0,1580
Acurácia e Correlação	0,6024	0,5350	0,5350	0,0142
Acurácia e Concordância	0,7219	0,6805	0,6805	0,0474
Acurácia e Estatística-Q	0,5452	0,8175	0,8175	0,1285

Tabela 5.9. *P-valores* encontrados ao analisar a relação entre os melhores e os demais da base *SatImage*, levando-se em consideração aos valores da diversidade.

As Tabelas Tabela 5.10, Tabela 5.11 e Tabela 5.12 ilustram a quantidade média de tipos de classificador escolhidos para formarem os comitês. Em alguns casos, esse número encontra-se arredondados para uma melhor análise dos resultados.

Objetivos	Melhor em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	1	1	2	3	2	1
Correlação	1	2	1	2	1	1
Concordância	2	2	2	3	3	1
Estatística-Q	1	1	3	2	2	1
Acurácia e Correlação	1	0	2	2	2	2
Acurácia e Concordância	2	1	2	3	1	1
Acurácia e Estatística-Q	1	1	3	2	1	1

Tabela 5.10. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados *SatImage*.

Objetivos	Pior em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	3	2	1	2	2	1
Correlação	2	2	0	2	1	0
Concordância	3	2	1	3	2	4
Estatística-Q	4	4	2	3	1	2
Acurácia e Correlação	3	4	0	2	0	2
Acurácia e Concordância	3	3	1	0	1	2
Acurácia e Estatística-Q	4	3	0	2	1	3

Tabela 5.11. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados *SatImage*.

Objetivos	Média					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	3	2	2	2	2	2
Correlação	2	2	2	2	1	1
Concordância	3	2	3	2	2	2
Estatística-Q	2	3	1	0	2	2
Acurácia e Correlação	2	3	1	1	1	2
Acurácia e Concordância	2	2	1	1	2	2
Acurácia e Estatística-Q	3	3	1	1	2	1

Tabela 5.12. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados *SatImage*.

Como podemos observar, o classificador *Naive* tende a selecionar menos classificadores para compor o comitê. Provavelmente, isso se deve ao fato de utilizar uma soma ponderada ao invés da soma simples (usado pelo *Soma*). Apesar do *k-NN* ser um algoritmo de aprendizado simples para classificação de padrões, se comparado aos outros dois utilizados, notamos sua considerável participação na composição do comitê. Isso se deve ao fato de que para a base de dados *SatImage* o *k-NN* possui individualmente a maior acurácia, conforme mostra a Tabela 5.1.

5.2.2 *Splice*

Os desempenhos dos multiclassificadores encontrados pelos AGs utilizando a base de dados *Splice* encontra-se na Tabela 5.13. A partir da análise desses resultados, vemos que, de modo geral, a acurácia combinada com alguma medida de diversidade, encontra comitês estatisticamente superiores aos obtidos usando somente a acurácia. Por exemplo, utilizando-se o classificador *Naive* obtivemos esses resultados com a Estatística-Q para o Melhor ($89,07 \pm 2,15$; $p\text{-valor} \approx 0,0324$), o Pior ($87,17 \pm 2,23$; $p\text{-valor} \approx 0,0001$) e a Média ($89,07 \pm 2,15$; $p\text{-valor} \approx 0,0188$); com a Concordância para o Pior ($87,17 \pm 2,23$; $p\text{-valor} \approx 0,0001$) e a Média ($89,12 \pm 1,17$; $p\text{-valor} \approx 0,0118$); e com a Correlação para o Pior ($87,12 \pm 3,06$; $p\text{-valor} \approx 0,0007$) e a Média ($89,31 \pm 3,45$; $p\text{-valor} \approx 0,0272$). Usando-se o combinador *Soma* encontramos esses resultados com a Estatística-Q para o Melhor ($92,77 \pm 4,08$; $p\text{-valor} \approx 0,0366$), o Pior ($86,04 \pm 3,00$; $p\text{-valor} \approx 0,0396$) e a Média ($88,32 \pm 3,21$; $p\text{-valor} \approx 0,0185$); com a Concordância para o Pior ($86,01 \pm 3,00$; $p\text{-valor} \approx$

0,0412) e a Média (89,89±3,62; p -valor \approx 0,0033); e com a Correlação para o Pior (86,00±3,00; p -valor \approx 0,0417) e a Média (89,12±3,06; p -valor \approx 0,0057).

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	89,09±3,15	89,77±3,20	82,67±3,75	82,15±2,40	84,08±4,06	85,00±4,50
Correlação	87,79±2,89	88,01±4,51	81,07±4,08	83,02±3,77	83,08±3,15	84,32±3,80
Concordância	87,87±2,86	88,33±3,87	82,17±3,80	83,15±4,02	83,61±4,51	84,87±1,98
Estatística-Q	87,62±2,65	87,63±2,89	81,17±3,05	82,72±3,53	84,78±4,09	84,77±2,79
Acurácia e Correlação	90,78±2,83	91,12±3,80	86,00±3,00	87,12±3,06	89,12±3,06	89,31±3,45
Acurácia e Concordância	91,01±2,58	91,82±1,89	86,01±3,00	87,17±2,23	89,89±3,62	89,12±1,17
Acurácia e Estatística-Q	92,77±4,08	93,00±3,03	86,04±3,00	87,17±2,23	88,32±3,21	89,07±2,15

Tabela 5.13. Acurácias obtidas com a base de dados *Splice*.

Outra análise realizada foi à comparação entre as medidas de diversidades – utilizando tanto as medidas sozinhas como combinadas com a acurácia. Através dessa análise, ratificamos a idéia de que entre as medidas de diversidade analisadas, não existe uma medida melhor do que a outra para a criação de comitês, pois em nenhuma das comparações a hipótese nula foi rejeitada. Ao efetuarmos o teste estatístico levando-se em consideração a acurácia, o menor p -valor (p -valor \approx 0,2212) foi encontrado ao compararmos o Acurácia e Correlação com a Acurácia e Concordância no caso do Melhor usando o método *Soma*. Já ao levarmos em consideração a diversidade, o menor p -valor (p -valor \approx 0,4683) foi encontrado ao compararmos a *Correlação* com a *Concordância* no caso do *Melhor* usando o método *Soma*. A Tabela 5.14 ilustra os valores de diversidade encontrados nos comitês.

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA	NA	NA
Correlação	0,78±0,13	0,78±0,10	0,67±0,12	0,68±0,19	0,73±0,16	0,73±0,12
Concordância	0,79±0,08	0,76±0,14	0,68±0,20	0,68±0,11	0,74±0,14	0,73±0,14
Estatística-Q	0,76±0,10	0,75±0,11	0,70±0,12	0,69±0,09	0,72±0,19	0,71±0,10
Acurácia e Correlação	0,77±0,12	0,76±0,13	0,69±0,15	0,71±0,12	0,74±0,12	0,74±0,15
Acurácia e Concordância	0,76±0,17	0,78±0,10	0,68±0,15	0,69±0,14	0,73±0,11	0,75±0,14
Acurácia e Estatística-Q	0,74±0,19	0,76±0,12	0,69±0,10	0,72±0,11	0,72±0,13	0,74±0,12

Tabela 5.14. Diversidades obtidas com a base de dados *Splice*.

A Tabela 5.15 mostra o número de vezes que cada função objetivo mostrou-se estatisticamente diferente em relação às demais funções, levando-se em consideração os

valores da acurácia. A Tabela 5.16 mostra esse mesmo número, levando-se em consideração os da diversidade. AC, COR, CON, Q, AC + COR, AC + CON e AC + Q significam, respectivamente, Acurácia; Correlação; Concordância; Acurácia e Correlação; Acurácia e Concordância; e Acurácia e Estatística-Q. Os valores a/b significam o número de vezes que a hipótese nula foi recusada (a) e o número de vezes que a hipótese nula não foi rejeitada (b).

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	0/6	0/6	0/6	4/2	4/2	6/0
COR	0/6	–	0/6	0/6	NA	NA	NA
CON	0/6	0/6	–	0/6	NA	NA	NA
Q	0/6	0/6	0/6	–	NA	NA	NA
AC + COR	4/2	NA	NA	NA	–	0/6	0/6
AC + CON	4/2	NA	NA	NA	0/6	–	0/6
AC + Q	6/0	NA	NA	NA	0/6	0/6	–

Tabela 5.15. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base *Splice*, levando-se em consideração a acurácia.

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	NA	NA	NA	NA	NA	NA
COR	NA	–	0/6	0/6	NA	NA	NA
CON	NA	0/6	–	0/6	NA	NA	NA
Q	NA	0/6	0/6	–	NA	NA	NA
AC + COR	NA	NA	NA	NA	–	0/6	0/6
AC + CON	NA	NA	NA	NA	0/6	–	0/6
AC + Q	NA	NA	NA	NA	0/6	0/6	–

Tabela 5.16. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base *Splice*, levando-se em consideração a diversidade.

De forma análoga, efetuamos a análise em relação aos métodos de combinação e também não encontramos evidências suficientes para afirmar que há diferenças nos desempenhos dos comitês criados ao se utilizar os diferentes métodos de combinação analisados, tanto levando em consideração a acurácia quanto a diversidade. No caso da acurácia, o menor p -valor (p -valor $\approx 0,1333$) foi encontrado quando comparamos o *Pior* dos desempenhos dos comitês encontrados usando apenas a Estatística-Q. Enquanto na diversidade, o menor p -valor (p -valor $\approx 0,5314$) foi encontrado quando comparamos o *Pior* dos desempenhos dos comitês encontrados usando a *Acurácia e Estatística-Q*.

Analisando o comportamento dos melhores multiclassificadores em relação aos piores e a média, levando em consideração a acurácia, vimos que os melhores são

estatisticamente superiores aos piores e, de forma geral, também o são em relação à média dos comitês envolvidos. A Tabela 5.17 sintetiza os *p-valores* encontrados ao efetuarmos o teste-t entre os melhores com as médias (Melhor-Média) e entre os melhores com os piores (Melhor-Pior). Os valores em destacados são aqueles cujos *p-valores* foram menores que o coeficiente α . Entre esses valores, no caso Melhor-Pior, chamamos atenção para aqueles encontrados na estatística-Q (*p-valor* $\approx 8,36 \times 10^{-5}$) utilizando-se o método de combinação Soma e, utilizando-se do classificador Naive, para somente a acurácia (*p-valor* $\approx 1,07 \times 10^{-5}$) e quando é combinada com a Concordância (*p-valor* $\approx 8,69 \times 10^{-5}$), pois obtiveram valores muito próximos de zero. Entre aqueles que são maiores, chamamos a atenção para os valores encontrados usando a Correlação (*p-valor* $\approx 0,0634$) e o classificador Naive, pois seu valor está muito perto da margem de erro adotada.

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	0,0064	0,0137	0,0006	$1,07 \times 10^{-5}$
Correlação	0,0026	0,0634	0,0005	0,0151
Concordância	0,0212	0,0215	0,0013	0,0088
Estatística-Q	0,0819	0,0371	$8,36 \times 10^{-5}$	0,0031
Acurácia e Correlação	0,2239	0,2794	0,0018	0,0184
Acurácia e Concordância	0,4360	0,0012	0,0008	$8,69 \times 10^{-5}$
Acurácia e Estatística-Q	0,0143	0,0036	0,0005	0,0001

Tabela 5.17. *P-valores* encontrados ao analisar a relação entre os melhores e os demais da base *Splice*, levando-se em consideração a acurácia.

Ao efetuarmos a mesma análise levando-se em consideração a diversidade, não se encontram evidências suficientes para afirmar que há diferenças na diversidade dos melhores comitês encontrados com a média dos comitês, bem como dos melhores comitês com os piores. A Tabela 5.18Tabela 5.17 sintetiza os *p-valores* encontrados ao efetuarmos o teste-t entre os melhores com as médias (Melhor-Média) e entre os melhores com os piores (Melhor-Pior).

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA
Correlação	0,4530	0,3248	0,06489	0,1580
Concordância	0,3397	0,6375	0,12373	0,1724
Estatística-Q	0,5630	0,4060	0,24018	0,1985
Acurácia e Correlação	0,5830	0,7536	0,20436	0,3832
Acurácia e Concordância	0,6450	0,5881	0,27915	0,1154
Acurácia e Estatística-Q	0,7866	0,7137	0,47097	0,4472

Tabela 5.18. *P*-valores encontrados ao analisar a relação entre os melhores e os demais da base *Splice* levando-se em consideração a diversidade.

As Tabelas Tabela 5.19, Tabela 5.20 e Tabela 5.21 ilustram a quantidade média de tipos de classificador escolhidos para formarem os comitês. Em alguns casos, esse número encontra-se arredondados para um melhor entendimento dos resultados.

Objetivos	Melhor em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	1	2	2	2	3	3
Correlação	3	2	2	2	2	1
Concordância	4	3	0	1	1	2
Estatística-Q	2	1	2	1	2	2
Acurácia e Correlação	2	1	1	0	2	2
Acurácia e Concordância	1	2	1	0	2	1
Acurácia e Estatística-Q	2	1	2	2	2	2

Tabela 5.19. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados *Splice*.

Objetivos	Pior em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	2	2	2	2	2	1
Correlação	2	2	2	2	2	2
Concordância	2	3	2	1	2	3
Estatística-Q	2	2	1	1	2	2
Acurácia e Correlação	2	2	2	2	3	2
Acurácia e Concordância	1	2	2	2	2	2
Acurácia e Estatística-Q	2	2	2	2	2	2

Tabela 5.20. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados *Splice*.

Objetivos	Média					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	3	2	2	2	2	2
Correlação	2	2	2	2	1	1
Concordância	3	2	3	2	2	2
Estatística-Q	2	3	1	0	2	2
Acurácia e Correlação	2	3	1	1	1	2
Acurácia e Concordância	2	2	1	1	2	2
Acurácia e Estatística-Q	3	3	1	1	2	1

Tabela 5.21. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados *Splice*.

Diferentemente da base anterior, o classificador *Naive* não tendeu a selecionar menos classificadores para compor o comitê do que o método *Soma*. Nessa base de dados, podemos ver um equilíbrio no número de tipos de classificadores presente em cada comitês, isso se dar ao fato dessa base de dados ser facilmente classificada tanto por k-NN, quanto por SVM e MLP.

5.2.3 SCOP

As acurácias e as diversidades encontradas pelos nossos experimentos para a base de dados SCOP encontram-se, respectivamente, nas Tabelas Tabela 5.22 e Tabela 5.23. Em termos gerais, a medida de diversidade combinada com a acurácia não ajudou significativamente no processo de criação de comitês tão acurados quanto diversos. Aquelas que conseguiram ajudar a acurácia nesse processo nas Médias foram a Estatística-Q ($76,01 \pm 3,79$; $p\text{-valor} \approx 0,0144$), a Concordância ($74,87 \pm 2,20$; $p\text{-valor} \approx 0,0152$) e a Correlação ($75,83 \pm 4,07$; $p\text{-valor} \approx 0,0259$) quando o combinador foi o classificador Naive. Utilizando o mesmo combinador, no caso dos melhores, a única medida de diversidade que ajudou ao processo foi à Estatística-Q ($78,39 \pm 3,75$; $p\text{-valor} \approx 0,0196$). Para as demais funções objetivas, a hipótese nula não foi rejeitada. Entretanto, alguns desses valores obtiveram a margem de erro próxima da escolhida ($\alpha = 0,05$), a Correlação: usando o Naive ($77,07 \pm 2,72$; $p\text{-valor} \approx 0,0527$) e o Soma ($76,01 \pm 3,02$; $p\text{-valor} \approx 0,0598$) nos Melhores; e usando a Soma ($74,33 \pm 2,16$; $p\text{-valor} \approx 0,0620$) na Média.

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	72,45±4,72	73,40±4,89	69,18±4,18	71,34±3,54	71,48±3,98	72,32±2,05
Correlação	68,12±3,17	69,10±1,78	63,10±2,68	64,15±2,68	66,34±2,98	68,00±3,15
Concordância	71,01±2,87	70,98±2,72	64,98±2,32	65,04±3,14	69,52±3,05	67,68±4,78
Estatística-Q	70,54±3,08	74,12±3,45	69,17±4,05	69,30±2,15	68,16±1,87	73,12±4,08
Acurácia e Correlação	76,01±3,02	77,07±2,72	68,56±2,50	71,42±4,16	74,33±2,16	75,83±4,07
Acurácia e Concordância	75,00±2,05	75,00±2,05	70,11±2,09	72,19±2,19	73,91±3,40	74,87±2,20
Acurácia e Estatística-Q	77,13±3,17	78,39±3,75	71,02±3,71	74,19±3,45	74,89±4,15	76,01±3,79

Tabela 5.22. Acurácias obtidas com a base de dados SCOP.

Objetivos	Melhor		Pior		Média	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA	NA	NA
Correlação	0,60±0,12	0,61±0,13	0,42±0,12	0,43±0,11	0,53±0,10	0,52±0,14
Concordância	0,62±0,13	0,62±0,12	0,41±0,13	0,44±0,14	0,54±0,13	0,56±0,13
Estatística-Q	0,61±0,11	0,60±0,15	0,40±0,11	0,46±0,18	0,55±0,12	0,55±0,12
Acurácia e Correlação	0,63±0,09	0,64±0,12	0,42±0,17	0,43±0,22	0,58±0,16	0,57±0,19
Acurácia e Concordância	0,64±0,13	0,63±0,11	0,43±0,18	0,44±0,18	0,55±0,19	0,54±0,20
Acurácia e Estatística-Q	0,63±0,11	0,64±0,15	0,44±0,19	0,46±0,10	0,57±0,17	0,55±0,19

Tabela 5.23. Diversidades obtidas com a base de dados SCOP.

Diferentemente, dos resultados das outras bases de dados. De modo geral, em relação à acurácia, nesta base os resultados dos comitês encontrados apenas com a acurácia mostraram-se superiores aos encontrados utilizando apenas as medidas de diversidades. Para essa base, nossos resultados corroboram com aqueles encontrados em (Kuncheva & Whitaker, 2003). No Capítulo 6 é dada uma possível razão de porque isso ter acontecido. Em alguns resultados, não foram encontradas evidências suficientes para refutar a hipótese nula. Por exemplo, utilizando o classificador Naive mais: a Estatística-Q para o Pior ($69,30 \pm 2,15$; $p\text{-valor} \approx 0,1367$), o Melhor ($74,12 \pm 3,45$; $p\text{-valor} \approx 0,7080$) e a Média ($73,12 \pm 4,08$; $p\text{-valor} \approx 0,5863$); a Concordância ($70,98 \pm 2,72$; $p\text{-valor} \approx 0,1882$) para o Melhor. Utilizando-se o método Soma mais: a Concordância, no Melhor ($71,01 \pm 2,87$; $p\text{-valor} \approx 0,4205$) e na Média ($69,52 \pm 3,05$; $p\text{-valor} \approx 0,2323$); mais Estatística-Q, no Pior ($69,17 \pm 4,05$; $p\text{-valor} \approx 0,9957$) e no Melhor ($70,54 \pm 3,08$; $p\text{-valor} \approx 0,2980$).

De modo geral, a Estatística-Q mostrou-se superior a Correlação, por exemplo, usando-se o classificador Naive no Melhor ($p\text{-valor} \approx 0,0007$), no Pior ($p\text{-valor} \approx 0,0001$) e na Média ($p\text{-valor} \approx 0,0056$); usando-se o Soma no Pior ($p\text{-valor} \approx 0,0009$). A Estatística-

Q, em termos gerais, mostrou-se superior a Concordância, por exemplo, usando o Naive no Melhor (p -valor $\approx 0,0364$), no Pior (p -valor $\approx 0,0001$) e na Média (p -valor $\approx 0,0056$); usando o Soma no Pior (p -valor $\approx 0,0109$). Em termos gerais, a Concordância e a Correlação não mostraram superioridade uma à outra, exceto a Concordância em relação à Correlação, ao usarmos o método Soma no Melhor (p -valor $\approx 0,0465$) e na Média (p -valor $\approx 0,0298$).

As Tabelas Tabela 5. 24 e Tabela 5. 25 mostram a quantidade de vezes que cada função objetivo mostrou-se estatisticamente diferente em relação às demais funções levando-se em consideração, respectivamente, à acurácia e à diversidade. AC, COR, CON, Q, AC + COR, AC + CON e AC + Q significam, respectivamente, Acurácia; Correlação; Concordância; Acurácia e Correlação; Acurácia e Concordância; e Acurácia e Estatística-Q. Os valores a/b significam o número de vezes que a hipótese nula foi recusada (a) e o número de vezes que a hipótese nula não foi rejeitada (b).

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	6/0	3/3	1/5	1/5	1/5	3/5
COR	6/0	–	2/3	4/2	NA	NA	NA
CON	3/3	2/3	–	4/2	NA	NA	NA
Q	1/5	4/2	4/2	–	NA	NA	NA
AC + COR	1/5	NA	NA	NA	–	0/6	1/5
AC + CON	1/5	NA	NA	NA	0/6	–	0/6
AC + Q	3/5	NA	NA	NA	0/6	1/5	–

Tabela 5. 24. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base SCOP, levando-se em consideração a acurácia.

	AC	COR	CON	Q	AC + COR	AC + CON	AC + Q
AC	–	NA	NA	NA	NA	NA	NA
COR	NA	–	0/6	0/6	NA	NA	NA
CON	NA	0/6	–	0/6	NA	NA	NA
Q	NA	0/6	0/6	–	NA	NA	NA
AC + COR	NA	NA	NA	NA	–	0/6	0/6
AC + CON	NA	NA	NA	NA	0/6	–	0/6
AC + Q	NA	NA	NA	NA	0/6	0/6	–

Tabela 5. 25. Resumo do número de vezes em que cada função objetivo mostrou-se estatisticamente diferente das outras funções para base SCOP, levando-se em consideração a diversidade.

Em termos gerais – tanto levando em consideração a acurácia quanto a diversidade –, não houve evidências suficientes para diferenciar os comitês criados com o método de combinação Soma ou com o classificador Naive, exceto em três casos em que o Naive

mostrou superior ao método Soma. Todos esses casos encontrados considerando-se os valores encontrados na acurácia (Tabela 5.22). Dois desses casos foram utilizando apenas a Estatística-Q, no Melhor ($74,12 \pm 3,45$; $p\text{-valor} \approx 0,0248$) e na Média ($73,12 \pm 4,08$; $p\text{-valor} \approx 0,0029$). O outro foi quando a acurácia é ajudada pela Concordância ($72,19 \pm 2,19$; $p\text{-valor} \approx 0,0248$). Entre os resultados que não refutaram a hipótese nula, destacamos os Piores comitês ($74,19 \pm 3,45$; $p\text{-valor} \approx 0,0633$) encontrados usando-se o Naive, onde os objetivos foram a Acurácia e a Estatística-Q.

De modo geral, os Melhores comitês encontrados foram estatisticamente superiores aos Piores, mas não foram superiores em relação às Médias de todos os comitês encontrados. A Tabela 5.26 exibe todos os *valores-p* encontrados quando se utilizou os valores da acurácia dos comitês. Os valores em destaque são os encontrados abaixo do valor α escolhido (0,05). Entre esses resultados, chamamos atenção para o encontrado utilizando-se a acurácia e correlação quando o método de combinação foi o Soma, pois o seu *p-valor* foi o mais o próximo de zero ($p\text{-valor} \approx 9,02 \times 10^{-7}$).

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	0,2948	0,5276	0,1183	0,2948
Correlação	0,0001	0,3491	0,0007	0,0001
Concordância	0,0002	0,0739	0,0001	0,0002
Estatística-Q	0,0014	0,5613	0,4024	0,0014
Acurácia e Correlação	0,0021	0,4335	$9,02 \times 10^{-7}$	0,0021
Acurácia e Concordância	0,0083	0,8928	0,0007	0,0083
Acurácia e Estatística-Q	0,0178	0,1751	0,0013	0,0178

Tabela 5.26. *P-valores* encontrados ao analisar a relação entre os melhores e os demais da base SCOP, levando-se em consideração os valores da acurácia.

A Tabela 5.27 exibe todos os valores-p encontrados quando se utilizou os valores da diversidade dos comitês. Os valores em destaque são os encontrados abaixo do valor α escolhido (0,05).

Objetivos	Melhor-Média		Melhor-Pior	
	Soma	Naive	Soma	Naive
Acurácia	NA	NA	NA	NA
Correlação	0,1735	0,1536	0,0035	0,0036
Concordância	0,1856	0,2976	0,0019	0,0063
Estatística-Q	0,2590	0,4212	0,0004	0,0750
Acurácia e Correlação	0,4004	0,3376	0,0028	0,0162
Acurácia e Concordância	0,2322	0,2284	0,0078	0,0106
Acurácia e Estatística-Q	0,3611	0,2550	0,0135	0,0054

Tabela 5.27. *P*-valores encontrados ao analisar a relação entre os melhores e os demais da base SCOP, levando-se em consideração os valores da diversidade.

As Tabelas Tabela 5.28, Tabela 5.29 e Tabela 5.30 Tabela 5.21 ilustram a quantidade média de tipos de classificador escolhidos para formarem os comitês. Em alguns casos, esse número encontra-se arredondados para um melhor entendimento dos resultados.

Objetivos	Melhor em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	2	2	2	2	3	3
Correlação	3	3	3	3	4	3
Concordância	2	2	4	3	4	3
Estatística-Q	1	1	3	3	2	2
Acurácia e Correlação	2	2	2	2	3	2
Acurácia e Concordância	2	2	3	2	2	2
Acurácia e Estatística-Q	2	1	2	2	2	2

Tabela 5.28. Quantidade média dos tipos de classificadores encontrados nos melhores comitês para a base de dados SCOP.

Objetivos	Pior em relação à Acurácia					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	1	2	1	3	3	3
Correlação	2	3	2	2	3	3
Concordância	2	2	1	2	3	3
Estatística-Q	1	2	0	2	2	2
Acurácia e Correlação	2	2	2	1	2	2
Acurácia e Concordância	2	1	1	2	2	2
Acurácia e Estatística-Q	2	2	2	3	2	2

Tabela 5.29. Quantidade média dos tipos de classificadores encontrados nos piores comitês para a base de dados SCOP.

Objetivos	Média					
	K-NN		SVM		MLP	
	Soma	Naive	Soma	Naive	Soma	Naive
Acurácia	1	1	2	2	1	1
Correlação	2	2	2	2	2	2
Concordância	2	1	2	2	3	3
Estatística-Q	1	2	2	2	2	1
Acurácia e Correlação	2	1	1	1	2	2
Acurácia e Concordância	1	1	2	2	1	2
Acurácia e Estatística-Q	2	1	2	2	1	2

Tabela 5.30. Quantidade média dos tipos de classificadores encontrados nos comitês para a base de dados SCOP.

De maneira similar a base de dados *Splice*, na base de dados SCOP o classificador *Naive* não tendeu a selecionar menos classificadores para compor o comitê do que o método *Soma*. Da mesma forma, houve um equilíbrio por parte dos tipos de algoritmos utilizados para a formação dos multiclassificadores.

Capítulo 6

Considerações Finais

Neste trabalho foi realizado um estudo sobre o comportamento de sistemas multiclassificadores heterogêneos mediante o dilema da diversidade-accurácia. Através da análise dos resultados vimos que mesmo utilizando diferentes tipos de base de dados, chegamos a resultados semelhantes.

Diferente dos resultados encontrados em (Kuncheva & Whitaker, 2003), que nos informar que a diversidade não é uma boa medida, por si só, para substituir a accurácia, nós não encontramos evidências para refutarmos a idéia que a diversidade não pode substituir a accurácia. No caso dos referidos autores foram usados *weak classifiers*¹³. Nesse contexto, é possível que a diversidade sozinha não tenha sido eficaz em guiar o processo de busca realizado. Já no nosso caso, em termos gerais, a accurácia dos classificadores base foi ligeiramente alta (superior a 70%). Devido a isso, a diversidade sozinha pode guiar o processo de otimização.

Por outro lado, nossos resultados estão de acordo com aqueles encontrados por Kuncheva e Whitaker (2003), de que combinar a diversidade com a accurácia pode nos guiar a criação de comitês mais acurados do que os gerados com essas medidas isoladamente. Portanto, a diversidade possui um papel importante na geração de multiclassificadores. De fato, queremos que esses multiclassificadores sejam tão acurados quanto diversos. Além disso, nossos resultados estão de acordo com os resultados de Kucheve e Whitaker (2003) de que não há uma medida ideal.

A partir da análise das Tabelas Tabela 5.8, Tabela 5.17 e Tabela 5.26, vimos que em, em termos gerais, os Melhores comitês encontrados mostraram-se superiores as Médias. Para a base de dados SCOP, provavelmente, isso aconteceu porque os Piores comitês encontrados eram os mais diversos, mas não eram tão acurados, como podemos observar nas Tabelas Tabela 5.26 e Tabela 5.27. Reduzindo, dessa forma, a accurácia média e,

¹³ Classificadores cujos desempenhos em relação à accurácia são um pouco melhor que uma escolha aleatória, entre]50%, 60%].

conseqüentemente, contribuindo para aumentar a diferença entre a acurácia média e a acurácia do melhor comitê. Para as bases de dados *SatImage* e *Splice* não obtivemos dados estatísticos suficientes para identificarmos a razão desse comportamento. De forma análoga, tivemos o mesmo comportamento para o algoritmo genético mono-objetivo. No entanto, nesse caso, o esperado seria que a população final convergisse para um único ponto (ponto ótimo). De fato, isso não aconteceu, pois, possivelmente, o número de iterações não foi suficiente para que a população final tendesse ao ponto ótimo.

Em duas das três bases de dados analisadas (*Splice* e SCOP), encontramos que alguns comitês com *Naive Bayes* foram estatisticamente superiores aos do combinador Soma. Isso pode ser devido ao fato de que a base *SatImage* pode ser mais “fácil” de classificar do que a *Splice* e a SCOP. Neste caso, o uso de “ponderação” (*Naive Bayes*) não fez diferença. Por outro lado, em bases “difíceis”, por exemplo, onde há misturas de classes como a SCOP, a ponderação pode nos ajudar no processo de classificação.

Como mencionado anteriormente, um dos problemas encontrados ao usarmos um algoritmo genético multiobjetivo é saber quais indivíduos do conjunto de soluções devemos utilizar, pois todos eles são soluções ótimas de acordo com os critérios utilizados. Para um problema cuja principal meta seja melhorar o desempenho do processo de classificação em relação à acurácia, então mesmo utilizando-se de um AGMO para encontrar os comitês, podemos escolher o comitê mais acurado dentre as soluções obtidas como saída que representará o algoritmo genético. Uma vez que nesse tipo de problema de classificação, o objetivo maior é a melhoria do processo de classificação em relação à acurácia, a utilização das demais funções objetivas têm o papel de auxiliar no processo de otimização desejado.

6.1. Trabalhos Futuros

O estudo realizado neste trabalho ainda pode ser expandido, alguns trabalhos futuros possíveis são:

- Alterar a forma de geração das partições base
 - Utilizar *weak classifiers* (Duin, 2002).

- Utilizar outros tipos de algoritmos de aprendizado de máquina, tais como, árvore de decisão, classificador *Naive Bayes*, redes neurais do tipo *radial basis function* e do tipo *Fuzzy MLP*.
- Usar diferentes instâncias da base de dados para treinar diferentes classificadores. Isto pode ser alcançado por meio do uso de técnicas de seleção de atributos (Santana L. E., 2008).
- Utilizar outros tipos de métodos de combinação como, Voto Majoritário, classificador MLP e Seleção Dinâmica de Classificadores (Kuncheva, 2004);
- Adicionar outra medida como função objetivo como, por exemplo, *Minimum Description Length* (Grünwald, 2007) – princípio que provê uma solução genérica para o problema de seleção de modelo.

Bibliografia

Abass, H. A. (2001). A Memetic Pareto Evolutionary Approach to Artificial Neural Networks. In *Lecture Notes in Computer Science* (pp. 113-152). Springer Berlin / Heidelberg.

Abbass, H. A. (2003). Pareto neuro-evolution: Construcção ensemble of neural networks using multi-objective optimization. *The IEEE 2003 Conference on Evolutionary Computation*, 3, pp. 2074-2080. IEEE Press.

Aha, D., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6, pp. 37-66.

Aksela, M. (2003). Comparison of classifier selection methods for improving committee performance. *Multiple Classifier Systems*, 2709 in *Lecture Notes in Computer Science*, pp. 84-93.

Alberts, B., Johnson, A., & Walter, P. (2004). *Biologia Molecular da Célula* (4ª Edição ed.). Artmed.

Banfield, R. E., Hall, L. O., Bowyer, K. W., & Kegelmeyer, W. P. (2003). A new ensemble diversity measure applied to thinning ensembles. *4th International Workshop on Multiple Classifier Systems*, (pp. 306-316).

Bauer, E., & Kohavi, R. (1999, Julho). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36 (1-2), pp. 105-139.

Blake, C., & Merz, C. (1998). *UCI Machine Learning Repository*. Retrieved 07 01, 2008, from <http://www.ics.uci.edu/~mllearn/MLRepository.html>

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, pp. 123-140.

Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Journal of Informatin Fusion*, 6 (1), pp. 5-20.

Canuto, A., Howells, G., & Fairhurst, M. (2000, Dezembro). The use of confidence measures to enhance combination strategies in multi-network neuro-fuzzy systems. *Connection Science*, 12 (3-4), pp. 315-331.

Chandra, A. (2004). *Evolutionary framework for the creation of diverse hybrid ensembles for better generalisation*. Dissertação de Mestrado, University of Birmingham, School of Computer Science.

Chandra, A., & Yao, X. (2004). DIVACE: Diverse and Accurate Ensemble Learning Algorithm. In *Intelligent Data Engineering and Automated Learning* (Vol. 3177, pp. 619-625). Springer Berlin.

Chandra, A., & Yao, X. (2005). Evolutionary Framework for the Construction of Diverse Hybrid Ensembles. *13th European Symposium on Artificial Neural Networks*, (pp. 253-258). Bruges (Bélgica).

- Chinasamy, A., Sung, W., & Mittal, A. (2004). Protein structure and fold prediction using tree-augmented bayesian classifier. *Pacific Symposium on Biocomputing*, (pp. 387-398). Hawaii, EUA.
- de Souto, M. C., Lorena, A. C., Delbem, A. C., & de Carvalho, A. C. (2003). Técnicas de Aprendizado de Máquina para Problemas de Biologia Molecular. In *III Jornada de Mini-Curso de Inteligência Artificial* (pp. 103-152). SBC.
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6 (2), pp. 182-197.
- Dietterich, T. G. (2000b). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40 (2), pp. 139-157.
- Dietterich, T. G. (2000a). Ensemble Methods in Machine Learning. In J. Kittler, & F. Roli, *Multiple Classifier Systems* (Vol. 1857, pp. 1-15). New York: Springer Verlag.
- dos Santos, E. M., Sabourin, R., & Maupin, P. (2006). Single and Multi-Objective Genetic Algorithms for the Selection of Ensemble of Classifiers., (pp. 3070-307).
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (Second Edition ed.). New York: Wiley-Interscience.
- Duin, R. P. (2002). The combining classifier: to train or not to train? *Proceedings of the 16th International Conference on Pattern Recognition*, (pp. 765-700).
- Durillo, J. J., Nebro, A. J., Luna, F., Dorronsoro, B., & Alba, E. (2006). *JMetal - A Java Framework for Developing Multi-Objective Optimization Metaheuristics*. Relatório Técnico, University of Málaga, Departamento de Lenguajes y Ciencias de la Computación.
- Faber, V. (1994). Clustering and the continuous k-means algorithm. *Los Alamos Science: High Performance Computing*, 22, pp. 138-144.
- Faceli, K. (2006). *Um framework para análise de agrupamento baseado na combinação multi-objetivo de algoritmos de agrupamento*. Tese de Doutorado, Universidade de São Paulo - São Carlos, Instituto de Ciências Matemáticas e de Computação, São Carlos - SP.
- Fogel, D. B., & Ghozeil, A. (1997). A Note on Representations and Variation Operators. *IEEE Transactions on Evolutionary Computation*, 1 (2), pp. 159-161.
- German, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, pp. 1-5.
- Giacinto, G., & Roli, F. (2001). An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 22 (1), pp. 25-33.
- Gnecco, B. B., Moraes, R. M., Machado, L. S., & Cabral, M. C. (2001). Um Sistema de Visualização Imersivo e Interativo de Apoio ao Ensino de Classificação de Imagens. *Symposium on Virtual Reality*, (pp. 291-301). Florianópolis, SC.
- Grünwald, P. D. (2007). *The Minimum Description Length Principle*. MIT Press.

- Hansen, L., & Salamon, P. (1990). Neural network ensembles. *IEEE Trans Pattern Analysis and Machine Intelligence*, 12, pp. 993-1001.
- Hearst, M. A., Schölkopf, B., Dumais, S., Osuna, E., & Platt, J. (1998). Trends and controversies - support vector machines. *IEEE Intelligent Systems*, 13 (4), pp. 18-28.
- Heutte, L., Nosary, A., & Paquet, T. (2004). A multiple agent architecture for handwritten text recognition. *Pattern Recognition*, 37, pp. 665–674.
- Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (8), pp. 832-844.
- Ho, T. K., Hull, J. J., & Srihari, S. N. (1994, Janeiro). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (1), pp. 66-75.
- Holland, J. H. (1992). *Adaptation in Nature and Artificial System*. Cambridge: MIT Press.
- Iba, T., & Takefuji, Y. (1998). Adaptation of Neural Agent in Dynamic Environment: Hybrid System of Genetic Algorithm and Neural Network. *IEEE International Conference on Knowledge-Based Intelligent Electronic Systems*, (pp. 21-23).
- Islam, M. M., Yao, X., & Murase, K. (2003). A constructive algorithm for training cooperative neural network ensembles. *IEEE Transactions on Neural Networks*, 14 (4), pp. 820-834.
- Kittler, J., & Alkoot, F. M. (2003). Sum versus vote fusion in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, pp. 110-115.
- Kruk, M., Osowski, S., & Koktysz, R. (2007, Agosto 12-17). Recognition of Colon Cells Using Ensemble of Classifiers. *International Joint Conference on Neural Networks*, pp. 288-293.
- Kuncheva, L. I. (2004). *Combining Pattern Classifiers*. New Jersey: Wiley.
- Kuncheva, L. I., & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51 (2), pp. 181-207.
- Langdon, W. B., Barrett, S. J., & Buxton, B. F. (2002). Combining decision trees and neural networks for drug discovery. *Proceedings of the 5th European Conference on Genetic Programming* (pp. 60-70). Ireland.
- Lazarevic, A., & Obradovic, Z. (2001). Effective pruning of neural network classifier ensembles. *Proceedings of the IEEE International Conference on Neural Networks* (pp. 796-801). IEEE.
- Lemieux, A., & Parizeau, M. (2003). Flexible multi-classifier architecture for face recognition. *Vision Interface*.
- Linden, R. (2006). *Algoritmos Genéticos - Uma importante ferramenta da Inteligência Computacional*. Brasport.

- Liu, Y., & Yao, X. (1999). Ensemble learning via negative correlation. *Neural Networks*, 12 (10), pp. 1399-1404.
- Liu, Y., & Yao, X. (1999). Learning and evolution by minimization of mutual information. *Neural Networks*, 12 (10), pp. 1399-1404.
- Liu, Y., Lin, Y., & Chen, Y. (2008). Ensemble Classification Based on ICA for Face Recognition. *The 2008 International Congress on Image and Signal Processing*, (pp. 144-148). Hainan.
- Liu, Y., Yao, X., & Higuchi, T. (2000). Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, 4 (4), p. 380.
- Liu, Y., Yao, X., & Higuchi, T. (2000, Novembro). Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, p. 380.
- Liu, Y., Zhou, J., & Chen, Y. (2008, Maio 27-30). Ensemble Classification for Cancer Data. *International BioMedical Engineering and Informatics*, pp. 269-273.
- Margineantu, D. D., & Dietterich, T. G. (1997). Pruning adaptive boosting. *Proceedings of the 14th International Conference on Machine Learning*, (pp. 211-218).
- Mitchell, T. M. (1997). *Machine Learning*. (E. M. Munson, Ed.) WCB/McGraw-Hill.
- Mohamed, S., Rubin, D., & Marwala, T. (2007, Agosto 12-17). Incremental Learning for Classification of Protein Sequences. *International Joint Conference on Neural Networks*, pp. 19-24.
- Monard, M., & Baranauskas, J. (2000). Conceitos sobre aprendizado de máquina. In S. Rezende, *Sistemas inteligentes: Fundamentos e aplicações* (pp. 89-114). São Paulo: Manole.
- Moore, D. S. (2005). *A Estatística Básica e Sua Prática*. (C. F. Pessoa, Trans.) Rio de Janeiro: Editora LTC.
- Mukhopadhyay, S., Peng, S., Raje, R., Palakal, M., & Mostafa, J. (2003). Multi-agent information classification using dynamic acquaintance lists. *Journal of the American Society for Information Science and Technology*, 54, pp. 966-976.
- Muller, K. -R., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12 (2), pp. 181-201.
- Oliveira, D. F., Canuto, A., & Campos, A. (2006). GNeurAge: An Evolutionary Agent-Based System for Classification Tasks. *Sixth International Conference on Hybrid Intelligent Systems*, (pp. 11-11). Rio de Janeiro.
- Ruta, D., & Gabrys, B. (2005). Classifier Selection for Majority Voting. *Information Fusion*, 6, pp. 163-168.
- Santana, A., Soares, R., Souto, M. C., & Canuto, A. M. (2008). Empirical Comparison of Dynamic Classifier Selection Methods Based on Diversity and Accuracy for Building Ensembles. *International Joint Conference on Neural Networks*. Hong Kong.

- Santana, L. E. (2008). *Uma Análise Comparativa entre Sistemas de Combinação de Classificadores com Distribuição Vertical dos Dados*. Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte, Departamento de Matemática e Informática Aplicada, Natal.
- Santos, E. M., Sabourin, R., & Maupin, P. (2006). Single and Multi-Objective Genetic Algorithms for the Selection of Ensemble of Classifiers. *International Joint Conference on Neural Networks (IJCNN'06)*, (pp. 3070-3077). Vancouver.
- Shou-Jue, W., & Xu, C. (2003). Biomimetic (topological) pattern recognition - a new model of pattern recognition theory and its application. *Proceedings of the International Joint Conference on Neural Networks (IJCNN'03)*, 3, pp. 2258–2262.
- Sirlantiz, K., Fairhurst, M. C., & Guest, R. M. (2002). An evolutionary algorithm for classifier and combination rule selection in multiple classifier system. *Proceedings of the 16th International Conference Pattern Recognition, 2002.*, 2, pp. 771-774.
- Sylvester, J., & Chawla, N. V. (2006). Evolutionary Ensemble Creation and Thinning. *International Joint Conference on Neural Networks (IJCNN'06)*, (pp. 5148-5155).
- Tang, E. K., Suganthan, P. N., & Yao, X. (2006). An analysis of diversity measures. *Machine Learning*, 65 (1), pp. 247-271.
- Tsymbal, A., Pechenizkiy, M., & Cunningham, P. (2005). Diversity in search strategies for ensemble feature selection. *Information Fusion*, 6 (1), pp. 83-98.
- Valentino, G. (2005). An experimental bias-variance analysis of SVM ensembles based on resampling techniques. *IEEE Transactions on System, Man and Cybernetics - Part B*, 35 (5), pp. 1252-1271.
- Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory* (2nd Edition ed., Vol. XIX). Springer-Verlag.
- Wang, S., Chen, & Ke. (2007, Agosto 12-17). Ensemble Learning with Active Data Selection for Semi-Supervised Pattern Classification. *International Joint Conference on Neural Networks*, pp. 355-360.
- Wang, W., Jones, P., & Partridge, D. (2000). Diversity between neural networks and decision trees for building multiple classifier systems. In F. Roli, J. Kittler, & T. Windeatt (Eds.), *Multiple Classifier Systems* (Vol. XII, pp. 240-249). Calgiari, Itália: Springer.
- Wang, W., Partridge, D., & Etherington, J. (2001). Hybrid ensembles and coincident-failure diversity. *Proceedings of the International Joint Conference on Neural Networks*. 4, pp. 2376-2381. Washington: IEEE Press.
- Windeatt, T. (2005). Diversity measures for multiple classifier system analysis and design. *Information Fusion*, 6 (1), pp. 21-36.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations* (2nd Edition ed.). Elsevier.

- Woods, K., Kegelmeyer, W. P., & Bowyer, K. (1997). Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 19, pp. 405-410.
- Wu, X., & Chen, Z. (2004). Recognition of exon/intron boundaries using dynamic ensembles. *Proceedings of the International IEEE Computer Society Computational Systems Bioinformatics Conference* (pp. 485-486). IEEE Computer Society.
- Xuchun, L., Zhu, Y., & Sung, E. (2008). A New Diversity Augmentation Methodology for Handling Diversity/Accuracy Dilemma in Boosting. *submetido ao IEEE Transactions on Neural Networks* .
- Yates, W., & Partridge, D. (1996). Use of methodological diversity to improve neural network generalization. *Neural Computing and Applications* , 4 (2), pp. 114-128.
- Zenobi, G., & Cunningham, P. (2001). Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error. *Proceedings of the European Conference on Machine Learning*, (pp. 576-587).
- Zhong, W., Liu, J., Xue, M., & Jiao, L. (2004). A multiagent genetic algorithm for global numerical optimization. *IEEE Transactions on Systems, Man, and Cybernetics - Part B* , 34 (2), pp. 1128- 1141.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)