André Luiz Brun

Geração e Seleção de Classificadores com base na Complexidade do Problema

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Doutor em Informática.

Curitiba
2017

André Luiz Brun

Geração e Seleção de Classificadores com base na Complexidade do Problema

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Doutor em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Alceu de Souza Britto Jr.
Co-orientador: Prof. Dr. Robert Sabourin

Curitiba
2017

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

B894g 2017

Brun, André Luiz
Geração e seleção de classificadores com base na complexidade do problema / André Luiz Brun ; orientador, Alceu de Souza Britto Jr. ; co-orientador, Robert Sabourin. – 2017.
100 f. ; il. 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2017
Bibliografia: f. 92-100

1. Reconhecimento de padrões. 2. Complexidade computacional. 3. Algoritmos genéticos. 4. Informática. I. Britto Júnior, Alceu de Souza, 1966-. II. Sabourin, Robert. III. Pontifícia Universidade Católica do Paraná. Programa Pós-Graduação em Informática. IV. Título.

CDD 20. ed. – 004

Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

PUCPR
GRUPO MARISTA

ATA DE SESSÃO PÙBLICA

DEFESA DE DISSERTAÇÃO DE DOUTORADO Nº 42/2017

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA – PPGIa
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ - PUCPR

Em sessão pública realizada às 14h00 de 14 de Fevereiro de 2017, no LEPS- Bloco 8– Escola Politécnica, ocorreu a defesa da tese de doutorado intitulada **"Geração e Seleção de Classificadores com Base na Complexidade do Problema"** elaborada pelo aluno **André Luiz Brun**, como requisito parcial para a obtenção do título de **Doutor em Informática**, na área de concentração **Ciência da Computação,** perante a banca examinadora composta pelos seguintes membros:

Prof. Dr. Alceu de Souza Britto Junior (orientador) - PPGIA/PUCPR

Prof. Dr. Robert Sabourin (co-orientador) - ETS-CANADÁ

Prof. Dr. Fabrício Enembreck – PPGIA/PUCPR

Prof. Dr. Julio Cesar Nievola – PPGIA/PUCPR

Prof. Dr. Luiz Eduardo Soares de Oliveira – UFPR

Prof. Dr. George Darmiton da Cunha Cavalcanti - UFPE

Após a apresentação da tese pelo aluno e correspondente arguição, a banca examinadora emitiu o seguinte parecer sobre a tese:

Membro	**Parecer**	
Prof. Dr. Prof. Dr. Alceu de Souza Britto Junior	(X) Aprovada	( ) Reprovada
Prof. Dr. Robert Sabourin	(X) Aprovada	( ) Reprovada
Prof. Dr. Fabrício Enembreck	(X) Aprovada	( ) Reprovada
Prof. Dr. Julio Cesar Nievola	(X) Aprovada	( ) Reprovada
Prof. Dr. George Darmiton da C. Cavalcanti	(X) Aprovada	( ) Reprovada
Prof. Dr. Luiz Eduardo Soares de Oliveira	(X) Aprovada	( ) Reprovada

Portanto, conforme as normas regimentais do PPGIa e da PUCPR, a tese foi considerada:

(X) **APROVADA**

(aprovação condicionada ao atendimento integral das correções e melhorias recomendadas pela banca examinadora, conforme anexo, dentro do prazo regimental)

( ) **REPROVADA**

E, para constar, lavrou-se a presente ata que vai assinada por todos os membros da banca examinadora. Curitiba, 14 de Fevereiro de 2017.

Prof. Dr. Dr. Alceu de Souza Britto Junior

Prof. Dr. Robert Sabourin

Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

Prof. Dr. Fabrício Enembreck

Prof. Dr. Julio Cesar Nievola

Prof. Dr. George Darmiton da C. Cavalcanti

Prof. Dr. Luiz Eduardo Soares de Oliveira

Rua Imaculada Conceição, 1155 Prado Velho CEP 80215 901 Curitiba Paraná Brasil
Tel.: (41) 3271 1669 Fax.: (41) 3271 2121 www.ppgia.pucpr.br

Agradecimentos

Gostaria de agradecer ao professor Alceu de Souza Britto Jr. por todo apoio ao longo destes anos de orientação, pelos ensinamentos, pela confiança na realização desta pesquisa, paciência e motivação. Agradeço também a todos os professores que contribuíram para a realização deste trabalho: Alessandro Koerich, Bráulio Ávila, Edson Scalabrin, Júlio Nievola, Edson Justino, Andreia Malucelli, Manoel Camilo Neto, Cinthia Freitas e Sirley Filipak. Agradeço em especial aos professores Luiz Oliveira, Fabricio Enembrek e Robert Sabourin pelas ideias, pelos ensinamentos, esclarecimentos e apoio direto à construção da presente pesquisa. Deixo também um enorme agrecimento ao professor Jacques Facon que, além das disciplinas lecionadas, possibilitou meu ingresso no Programa e que, junto com o professor Alceu, confiou na minha capacidade para o cumprimento desta etapa.

Agradeço em especial à minha esposa Greicy Kiel que foi meu porto seguro, me apoiando em cada decisão tomada ao longo destes anos, me incentivando e que sempre me serviu de inspiração. Agradecimento também à família por todo o carinho e suporte.

A realização deste doutorado me permitiu, além de construir um grande aprendizado, conhecer pessoas especiais que, cada uma à sua forma, me auxiliaram no cumprimento deste desafio: Alexandre Belarmino, Alonso de Carli, Anderson Bertling, Andreia Marini, Angela Roveredo, Arlete Beuren, Bruno Souza, Cheila Cristina, Cleverton Vicentini, Denise Sato, Edenilson Silva, Eduardo Viegas, Emerson Fedechen, Elias Carvalho, Erich Malinowski, Eunelson Silva, Fabiano Utiyama, Flávio Silva, Franciele Beal, Francis Baranoski, Grasielli Zimmermann, Gregory Wanderley, Gustavo Bonacina, Heitor Gomes, Irapuru Florido, Jean-Paul Barddal, Jhonatan Geremias, Jurandir dos Santos, Kelly Wiggers, Luiz Giovanini, Marcelo Pereira, Marcelo Zacharias, Marcia Pascutti, Mariza Dosciatti, Nicolas de Paula, Patrícia Antoniolli, Priscila Santin, Rodolfo Botto, Rodrigo Siega, Ronan Assumpção Silva, Sandoval Ruppel, Sidnei Schuindt, Vilmar Abreu, Viviane Dal Molin, Voncarlos Araújo e Wendel Goes.

Deixo também meu agradecimento a todos meus amigos e colegas que, mesmo não fazendo parte do Programa do doutorado, contribuíram para que eu obtivesse este título.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

AG	*Algoritmo Genético*
ALL	*Combinação de todos os classificadores*
DCoL	*Data Complexity Library*
DES	*Dynamic Ensemble Selection*
DES-FA	*Dynamic Ensemble Selection by Filter + Adaptative Distance*
DSOC	*Dynamic Selection Over Complexity*
DWEC-CV	*Dynamic Weightinh Ensemble Classifiers based on Cross Validation*
GDES	*Dynamic Classifier Ensemble Selection based on GMDH*
GDMH	*Group Method of Data Handing*
KEEL	*Knowledge Extraction based on Evolutionary Learning*
KNN	*K Nearest Neighbors*
KNORA	*K-Nearest Oracles*
LCA	*Local Class Accuracy*
LKC	*Ludmila Kuncheva Collection of Real Medical Data*
MAJ	*Voto Majoritário*
MCB	*Multiple Classifier Behavior*
MST	*Minimal Spanning Tree*
NIST	*National Institute of Standards and Technology*
OCS	*Overproduce-and-Choose Strategy*
OLA	*Overall Local Accuracy*
RSS	*Random Subspaces*
SB	*Single Best*
SC	*Subconjunto*
SDES	*Sorting-based Dynamic Classifier Ensemble Selection*
SMC	*Sistema de Múltiplos Classificadores*
SOCS	*Selection based on Overproduce-and-Choose Strategy*
SVM	*Support Vector Machine*
UCI	*University of California, Irvine*

Resumo

O reconhecimento de padrões tem como uma de suas principais aplicações atribuir a um determinado objeto, uma classe entre várias possíveis. Este processo de rotulação recebe o nome de classificação. Sistemas baseados em múltiplos classificadores (SMCs) têm sido utilizados como alternativa para a difícil tarefa de construir um único classificador capaz de absorver toda a variabilidade de um problema. Em SMCs, a seleção de classificadores para cada instância de teste tem se mostrado uma estratégia promissora. Além disto, diversos estudos também demonstram que a análise de complexidade dos dados pode contribuir para a escolha da abordagem de classificação. A adoção de informações acerca da complexidade dos dados do problema no processo de seleção, no entanto, ainda encontra-se em estado incipiente, carecendo de pesquisas que analisem a relação de tais medidas com o processo de seleção dos classificadores. Assim sendo, a presente pesquisa teve como objetivo o desenvolvimento e avaliação de um SMC no qual a novidade é a adoção de informações de dificuldade do problema de classificação para orientar tanto a geração dos conjuntos de classificadores como a posterior seleção destes. A dificuldade do problema é descrita através de meta-características obtidas a partir dos dados do problema usando as medidas de complexidade. Para a etapa de geração dos subconjuntos foi desenvolvido um algortimo genético cujo objetivo foi maximizar a exploração do espaço de complexidade e ao mesmo tempo formar indutores precisos. Para a etapa de seleção foram combinados três critérios: a acurácia local de cada classificador, a similaridade de sua assinatura de complexidade e a distância da instância de teste até o centróide da classe predita. Visando avaliar as abordagens propostas para a geração e seleção, bem como o SMC como um todo, executou-se um protocolo experimental robusto sobre 30 problemas provindos de diferentes repositórios considerando 20 replicações, comparando-os com diversos métodos já estabelecidos na literatura. Os resultados mostram que a estratégia evolutiva construída para a geração pôde contribuir para o aumento da acurácia dos métodos da literatura, uma vez que observou-se melhora na acurácia em 126 de 180 casos (70.00%). Além disso, verificou-se que a estratégia pôde formar pools mais bem distribuídos no espaço de complexidade em 29 dos 30 problemas testados. Já a aborda-

gem de seleção dinâmica proposta suplantou as concorrentes em 82.00% dos cenários. Ao compararmos o SMC construído com os métodos da literatura verificamos uma melhora em termos de acurácia em 91.67% dos problemas estudados. Os resultados observados com a realização desta pesquisa permitiram concluir que a exploração de informações relacionadas à complexidade dos dados é uma alternativa interessante para a geração de pools, estimação da competência dos classificadores, bem como para todo o processo de classificação desempenhado pelo SMC.

Palavras-chave: Sistemas de Múltiplos Classificadores, Geração de Pools de Classificadores, Seleção Dinâmica de Classificadores, Dificuldade do Problema de Classificação

Abstract

Pattern recognition has as one of its main tasks to assign to a particular object a class from a set of possibilities. This labeling process is named classification. Multi-classifier systems (MCSs) have been used as an alternative to the difficult task of building a single classifier capable of absorbing all the variability of a classification problem. In MCSs, the selection of classifiers for each test instance has shown to be a promising strategy. In addition, several studies have shown that data complexity analysis plays an important role in the classification process. The adoption of information about the data complexity of the problem in the selection process, however, is still in an incipient state, lacking researches that analyze the relation of such measures to the process of classifiers selection. Therefore, the present research had as objective the development and evaluation of an MCS in which the novelty is the adoption of information of difficulty of the classification problem to guide both a generation of the sets of classifiers and thier later selection. The classification dificulty is described by meta-features estimated from the problem data using complexity measures. For the subsets generation stage a genetic algorithm was developed whose objective was to maximize the exploration of the complexity space and at the same time to form accurate inductors. For the selection stage, three criteria were combined: the local accuracy of each classifier, the similarity of its complexity signature, and the distance from the test instance to the predicted class centroid. Aiming to evaluate the proposed approaches to generation and selection, as well as the MCS as a whole, a robust experimental protocol was executed on 30 problems from different repositories considering 20 replications, comparing them with several methods already established in the literature. The results show that the evolutionary strategy built for the generation could contribute to the increase of accuracy of the literature methods, since there was an improvement in accuracy in 126 of 180 cases (70.00%). In addition, it was found that the strategy was able to form pools better distributed in complexity space in 29 of the 30 problems tested. The proposed dynamic solution approach supplanted rivals in 82.00% of the scenarios. When

comparing the built MCS to the methods of the literature we verified an improvement in terms of accuracy in 91.67% of the problems studied. The results obtained with this research allowed us to conclude that the exploration of information related to the data complexity is an interesting alternative for the pools generation, the estimation of the classifiers competence, as well as for the entire classification process performed by the SMC.

Keywords: Multiple Classifier Systems, Classifier Pool Generation, Dynamic Classifiers Selection, Classification Problem Difficulty

Capítulo 1

Introdução

O reconhecimento de padrões tem como uma de suas principais aplicações atribuir a um determinado objeto, uma classe entre várias possíveis. Este processo de rotulação recebe a alcunha de classificação. O procedimento consiste em analisar um conjunto de informações (vetor de características) sobre o elemento a ser classificado e, segundo critérios definidos, determinar a qual classe ele pertence.

Os métodos responsáveis por realizar a atribuição de rótulos aos elementos ainda não classificados são chamados de classificadores. Espera-se que estes, com base nas características do objeto, possam realizar a atribuição da classe de forma precisa. Para tanto, a escolha de um classificador que seja adequado ao contexto é primordial. Segundo Gunes *et al.* (GUNES et al., 2003) o critério mais adotado neste sentido consiste em aplicar o classificador que obtém a maior acurácia.

Entretanto, o classificador selecionado para uma situação pode ter desempenho inferior em outros cenários. Classificadores instáveis ou que apresentam taxas de precisão baixas são ditos "fracos" (SKURICHINA; DUIN, 2002). Uma alternativa para a melhora da eficácia dos métodos é adotar vários classificadores no processo classificatório (KITTLER et al., 1998), (JAIN; DUIN; MAO, 2000), (SKURICHINA; DUIN, 2002), (KUNCHEVA; WHITAKER, 2003) e (KO; SABOURIN; BRITTO JR., 2008). A abordagem, conhecida como Sistemas de Múltiplos Classificadores (SMC's), vale-se do fato de que classificadores distintos geralmente cometem erros diferentes em amostras distintas (KO; SABOURIN; BRITTO JR., 2008),(YU-QUAN et al., 2011). Um fator importante para que os erros cometidos sejam variados é que haja diversidade entre os classificadores selecionados.

Muitos pesquisadores têm focado seus estudos nos SMC e, consequentemente, novas soluções têm sido dedicadas para cada uma das três possíveis fases dos SMC: a) geração, b) seleção, e c) integração. Na primeira fase, um *pool* de classificadores é gerado; na segunda, um subconjunto destes classificadores é selecionado, enquanto na última fase,

uma decisão final é feita com base nas predições dos classificadores selecionados.

A etapa de geração pode ser realizada de forma homogênea ou heterogênea. Quando a construção dos *pools* envolve apenas indutores de um mesmo tipo treinados em diferentes conjuntos de dados, ela pertence à primeira estratégia. Por outro lado, pertencem à segunda abordagem, os métodos de geração que se baseiam em diferentes indutores treinados sobre o mesmo conjunto de dados.

Nos SMC's, as técnicas mais usadas para a construção dos *pools* são o Bagging (BREIMAN, 1996), Boosting (FREUND; SCHAPIRE, 1996) e RSS (HO, 1998), as quais geram grupos de classificadores buscando obter diversidade entre eles. Com exceção do Boosting que, ao gerar novos subconjuntos, considera as instâncias erroneamente classificadas anteriormente, os métodos de geração geralmente manipulam randomicamente os dados para treinar classificadores fracos e diversos, sem levar em conta informações de complexidade dos dados usados para o treinamento.

Visando descrever o nível de dificuldade do problema em análise, através de índices, diversos pesquisadores utilizam as medidas de complexidade (HO; BASU, 2000), (HO; BASU, 2002), (HO; BASU; LAW, 2006), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010). Sabendo-se que tais medidas têm relação com o comportamento dos classificadores, os autores buscam descrever através delas, quão sobrepostas estão duas classes, como se comporta a região fronteiriça entre elas ou mesmo como é a distribuição espacial de cada uma.

Uma vez que as estratégias de geração não levam em conta a dificuldade do problema de classificação, o *pool* formado pode ser composto de elementos treinados em subconjuntos muito similares em termos de complexidade, fazendo com que o espaço que representa a dificuldade do problema seja pouco explorado. Parece razoável pensar que com uma melhor exploração do espaço que representa a complexidade do problema (ou dificuldade) seria possível melhorar o desempenho do SMC em termos de acurácia. Esta ideia é inspirada em trabalhos que tentam encontrar o indutor mais promissor para um problema de classificação específico, considerando no processo a dificuldade do problema (HA; ZIMMERMANN; BUNKE, 1998).

Assim sendo, propõe-se neste trabalho uma nova abordagem de geração de *pools* de classificadores que buscam explorar de forma mais efetiva o espaço de complexidade do problema sobre o qual estiver trabalhando, gerando classificadores treinados em subconjuntos mais divesificados em termos de complexidade. Para tanto, tratou-se como metas a variedade dos valores dos índices de complexidade. Assim, foi necessário o desenvolvimento de um método de otimização para a geração dos subconjuntos das amostras de treinamento.

A seleção dos classificadores pode ser realizada de forma individual ou em *ensembles*. Na primeira, apenas o classificador mais apto é escolhido, enquanto na segunda abordagem é construído um grupo formado pelos elementos mais promissores. Além disso, a escolha dos classificadores pode se dar de forma estática ou dinâmica. Quando estes são escolhidos durante a fase de treinamento, sem considerar a instância de teste, o processo é dito estático. Neste caso, o mesmo conjunto de classificadores selecionados será usado para rotular todas as instâncias. Por outro lado, caso no momento da seleção seja levado em conta informações sobre a instância de teste, a estratégia é considerada dinâmica. Esta abordagem recebe esta alcunha pois, para cada nova instância, um conjunto distinto de classificadores pode ser selecionado. A estratégia dinâmica tem recebido especial atenção da comunidade científica.

Os métodos de seleção dinâmica encontrados na literatura buscam medir a competência dos classificadores disponíveis, visando selecionar um ou vários classificadores que sejam, em teoria, os mais apropriados para classificar cada instância. Estas abordagens buscam, segundo diversos critérios, avaliar a região vizinha à amostra a ser classificada e com base nestas informações, medir a competência dos classificadores. Tais estratégias, em sua grande maioria, consideram apenas acurácia para medida de competência dos classificadores dada uma instância de teste. Além da acurácia há alguns métodos que consideram a diversidade ou ainda consenso como forma de medir a competência em grupo, mas até onde sabemos nenhum método considera informações relacionadas ao nível de dificuldade da instância a ser classificada.

Estudos focados na seleção do melhor classificador ou melhor grupo de classificadores podem basear-se na acurácia local destes, como o *Overall Local Accuracy* (OLA) (WOODS; KEGELMEYER JR.; BOWYER, 1997), (DIDACI et al., 2005), *Local Class Accuracy (LCA)* (WOODS; KEGELMEYER JR.; BOWYER, 1997), (DIDACI et al., 2005), A Priori (GIACINTO; ROLI, 1999), (DIDACI et al., 2005), A Posteriori (GIACINTO; ROLI, 1999), (DIDACI et al., 2005), *K-Nearest Oracles* (KNORA) (KO; SABOURIN; BRITTO JR., 2008). Outros estudos baseiam-se na diversidade dos classificadores como em (SANTANA et al., 2006) e (YAN et al., 2013). As técnicas empregadas podem utilizar filtros e regiões de competência (CRUZ; CAVALCANTI; REN, 2011) ou basear-se em outras abordagens como SVM e *Fuzzy Pattern Matching* (AYAD; SYED-MOUCHAWEH, 2011) ou sobre *Multistage Organization* (CAVALIN; SABOURIN; SUEN, 2013).

A adoção de informações acerca da complexidade dos dados do problema no processo de seleção, no entanto, ainda encontra-se em estado incipiente, carecendo de pesquisas que analisem a relação de tais medidas com o processo de seleção dos classificadores. Assim sendo, nesta pesquisa, além de uma estratégia de geração de *pools*, buscou-se ava-

liar a viabilidade da adoção de medidas de complexidade em conjunto com a acurácia como critério para seleção dinâmica de classificadores.

Visto que diferentes trabalhos na literatura, como as obras de Ho e Basu (HO; BASU, 2002) e Macià *et al.* (MACIà et al., 2013), sugerem que uma boa estratégia para seleção de classificadores é compreender melhor a complexidade dos subconjuntos em que os classificadores são treinados e das vizinhanças das instâncias em avaliação, estudou-se a hipótese de que, se fosse determinado previamente o melhor classificador para cobrir regiões específicas do espaço do problema representado por medidas de complexidade, então seria possível selecionar o melhor classificador para um padrão desconhecido pertencente a uma região de complexidade similar.

Nas obras (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (OKUN; VALENTINI, 2008) e (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) os autores indicam que o desempenho dos classificadores NN e métodos de seleção dinâmica são influenciados pelas características de complexidade dos dados sobre as quais estão sendo trabalhadas, fato que reforça a ideia de que a seleção dos classificadores com base na complexidade da vizinhança do novo padrão pode ser uma alternativa viável às técnicas de seleção dinâmica conhecidas.

1.1 Hipóteses

Neste trabalho duas hipóteses foram levantadas. A primeira considera que o uso de informações relativas ao nível de dificuldade da classificação obtidas a partir dos dados do problema na geração do *pool* de classificadores de um SMC, permite gerar classificadores que juntos cobrem melhor o espaço de complexidade do problema e, consequentemente, apresentam um melhor desempenho.

A segunda hipótese considerada é de que as medidas de complexidade podem contribuir para a seleção dinâmica de classificadores. A similaridade em termos de nível de dificuldade de classificação entre a vizinhança do exemplo de teste definida na base de validação e o subconjunto de treinamento usado para criar um determinado classificador do *pool* pode ser aplicada como um indicador de competência.

Dessa forma, a principal questão da pesquisa é a seguinte: O uso da análise de complexidade dos dados em ambas as fases de um SMC (geração e seleção) pode trazer contribuição adicional? Para respondermos esta pergunta, faz-se necessário termos a resposta de algumas questões secundárias: Qual é o impacto em termos de desempenho na classificação quando a informação referente à análise da complexidade dos dados orienta a geração do *pool* de um SMC? O *pool* gerado com base nas características de dificuldade é capaz de melhor cobrir o espaço de complexidade? A complexidade das regiões locais

no espaço do problema pode ser uma medida apropriada para determinar a região de competência de um classificador de um dado *pool*?

1.2 Proposta

Buscou-se construir uma abordagem para geração de classificadores com foco no espaço de complexidade explorando de forma mais abrangente tais medidas. Além disso, avaliou-se o impacto dessa técnica sobre o processo de seleção baseado em complexidade.

Buscando avaliar a aplicabilidade das medidas de complexidade do conjunto de dados como critério de seleção de classificadores de forma dinâmica, propôs-se o desenvolvimento um *framework* capaz de determinar qual ou quais classificadores são mais apropriados para rotular cada umas das amostras de teste, usando no processo medidas de complexidade dos classificadores e das vizinhanças das instâncias em avaliação.

1.3 Objetivos

O objetivo geral do trabalho consiste em avaliar o impacto do uso de informações relativas à complexidade do problema de classificação em um SMC baseado na seleção dinâmica de classificadores, em dois momentos: a) na geração do *pool* e b) no mecanismo de seleção. Contudo, para isto torna-se necessário:

- Definir como representar a dificuldade de um problema de classificação e quais características utilizar nas etapas de geração e seleção.
- Desenvolver um novo método de geração de *pools* de classificadores dirigido por medidas de complexidade combinadas com acurácia.
- Avaliar o impacto do novo método de geração de *pools* considerando diferentes abordagens de seleção dinâmica e soluções estáticas.
- Avaliar o comportamento dos subconjuntos formados pela estratégia de geração proposta no espaço de complexidade.
- Desenvolvimento de uma nova abordagem de seleção dinâmica de classificadores em que a competência é definida com base em descritores de complexidade.
- Avaliar o impacto do novo método de seleção em relação à diferentes abordagens de seleção dinâmica.

- Avaliar o comportamento do SMC proposto diante de soluções consagradas na literatura.

1.4 Contribuições

Esperava-se por meio da realização desta pesquisa, construir conhecimento mais aprofundado da relação entre o desempenho dos métodos de classificação dinâmicos e as características de complexidades pertinente aos dados, permitindo avançar no estudo destas medidas e assim, contribuir para o progresso dos métodos de reconhecimento, geração de *pools*, seleção dinâmica de classificadores e também na utilização de descritores de complexidade dos problemas.

- Novo método de geração de *pools* de classificadores com base na exploração do espaço de complexidade do problema em estudo.
- Avaliação da contribuição dos descritores de dificuldade do problema no momento da geração dos subconjuntos para treinamento dos classificadores em relação às técnicas consagradas na literatura.
- Estudo do impacto no espaço de complexidade dos subconjuntos gerados por um método direcionado pela exploração de tal espaço.
- Definição de critérios para a estimação de forma dinâmica da competência de classificadores com base em índices de complexidade.
- Novo método de seleção dinâmica de classificadores com base na similaridade em termos de dificuldade entre classificadores e a vizinhança da instância de teste combinada com acurácia local.
- Avaliação da contribuição de critérios de complexidade no processo de seleção dinâmica em diversos problemas.
- Novo sistema de múltiplos classificadores que considera informações da dificuldade do problema na geração e estimação da competência dos conjuntos de classificadores.
- Avaliação da contribuição da adoção das medidas de complexidade nas duas principais etapas de um SMC em comparação à soluções já estabelecidas na literatura.
- Análise do comportamento de estratégias de seleção dinâmica consagradas nos *pools* gerados pelo método proposto em relação à uma solução tradicional.

Com base nas contribuições descritas foi possível derivar outras contribuições pontuais:

- A conclusão de que informações da complexidade dos dados onde os classificadores são treinados podem trazer contribuição no processo de estimação da competência dos mesmos.
- A confirmação de que gerar um conjunto de classificadores de forma a melhor explorar o espaço de complexidade pode trazer ganho em termos de acurácia em métodos de seleção dinâmica de classificadores, bem como ao *single best* e combinação dos elementos do pool.
- A confirmação de que a adoção de uma estratégia evolutiva, que busca otimizar acurácia combinada com a exploração da complexidade, consegue formar um grupo de subconjuntos que apresentam grande diversidade entre si no espaço de complexidade.
- A conclusão de que descritores da dificuldade dos dados podem ser usados com sucesso nas etapas de geração e seleção de um SMC.

1.5 Estrutura do Trabalho

Após a introdução realizada no Capítulo 1, apresenta-se na sequência, a Revisão da Literatura acerca do processo de classificação. Neste capítulo são apresentadas formas de se construir classificadores, como estes podem ser combinados e selecionados, seja estática ou dinamicamente. Dado o caráter desta pequisa, fez-se um levantamento de diversos estudos já realizados no campo da seleção de classificadores, monolíticos e *ensembles* com foco na seleção dinâmica. No Capítulo 3 são apresentadas e detalhadas diversas medidas de complexidade, as quais servem de critério para os processos aqui implementados.

No quarto capítulo é descrita a metodologia construída. São apresentadas as etapas desenvolvidas, de forma genérica, para a realização da geração e seleção dinâmica baseada nas medidas de complexidade descritas na seção anterior.

No capítulo seguinte (5), são descritos os experimentos realizados. Esta seção discorre, com detalhes, a configuração e os resultados de cada ensaio executado. O objetivo foi apresentar um cenário incremental no qual o primeiro experimento baseia-se apenas na geração, o segundo envolve a etapa de seleção e, por fim, o terceiro experimento que combina as duas estratégias, formando o SMC proposto.

No Capítulo 6 são apresentadas as conclusões formadas após a realização da pesquisa e as considerações finais do trabalho. Por fim, as referências sobre as quais se embasou esta pesquisa são apresentadas na última seção deste trabalho.

Capítulo 2

Classificação

Os métodos de reconhecimento na literatura buscam medir a competência dos classificadores disponíveis, visando selecionar aquele ou aqueles classificadores que sejam, em teoria, os mais apropriados para classificar cada instância. Essas abordagens buscam, segundo diversos critérios, avaliar a região vizinha à amostra a ser classificada e com base nestas informações, medir a competência dos classificadores.

A efetividade em se adotar vários classificadores depende, no entanto, de que os classificadores empregados apresentem diversidade entre si, cometendo erros não relacionados, de forma que padrões com características distintas possam ser classificados corretamente. Um fator que influi diretamente na características do *pool* de classificadores é o método de construção adotado. O desempenho da abordagem adotada reside também na forma em que os classificadores são selecionados e em como são combinados no momento da classificação. A obra de Britto, Sabourin e Oliveira (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) apresenta o funcionamento de um SMC dividido em três etapas, cada qual referente a um dos fatores de impacto na acurácia do método, conforme apresentado na Figura 2.1.

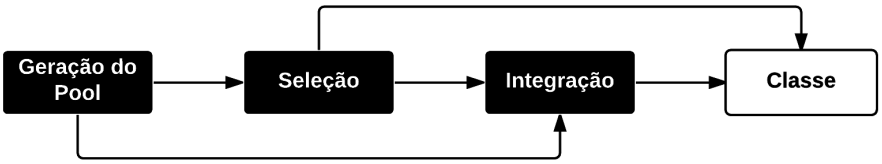

Figura 2.1: Fases de um Sistemas de Múltiplos Classificadores

Inicialmente são construídos os classificadores responsáveis pela classificação dos novos padrões. Esta etapa, que pode ocorrer de forma homogênea ou heterogênea, é apresentada na seção 2.1. Uma vez formado o grupo de classificadores, faz-se necessária

a escolha de um ou vários deles para realizar a classificação da nova instância. A ideia é selecionar o(s) classificador(es) que pode(m) ser mais preciso(s) no momento de atribuir o rótulo à amostra de teste. Este processo pode ser feito de forma estática ou dinâmica. A segunda, foco deste trabalho, realiza a escolha com base na instância de teste, podendo variar a cada iteração. Estudos focados na seleção do melhor classificador ou melhor grupo de classificadores podem basear-se na acurácia local destes, em probabilidades a priori e posteriori, em comportamento, diversidade, acurácia, regiões de competência, entre outras. Os métodos de seleção dinâmica são discutidos mais detalhadamente na Seção 2.2, onde são tratadas técnicas de seleção de classificador único (2.2.1) e de conjuntos de classificadores (2.2.2).

A terceira etapa, responsável pela combinação dos classificadores selecionados, é descrita na seção 2.2.3.

A representação, segundo (BRITTO JR.; SABOURIN; OLIVEIRA, 2014), não é única, visto que a abordagem adotada pode não conter, por exemplo, a etapa de seleção em casos onde todos os classificadores são empregados no momento da classificação. Além disso, existem cenários em que o processo de integração faz-se desnecessário. Tal fato pode ocorrer quando é selecionado apenas um classificador na segunda etapa.

2.1 Construção de Conjuntos de Classificadores

A construção de classificadores visa, com base em um conjunto de dados de um problema específico, desenvolver vários subconjuntos dos dados, de forma que, trabalhando de forma cooperada no momento da classificação, possam obter taxas de reconhecimento superiores a simples aplicação individual de um classificador.

Os classificadores podem ser construídos por métodos homogêneos ou heterogêneos. Na primeira abordagem, durante o processo de geração são adotados os métodos semelhantes de construção. Já na segunda, diferentes algoritmos são aplicados ao longo do processo. Dentre as abordagens homogêneas mais aplicadas tem-se *Bagging* (BREIMAN, 1996), *Boosting* (FREUND; SCHAPIRE, 1996) e *Random Subspaces* (HO, 1998).

2.1.1 Bagging

O método de*Bagging* fundamenta-se em sortear aleatoriamente, e com reposição, elementos do conjunto de treino para formar os classificadores. Proposto por Breiman (BREIMAN, 1996) a abordagem consiste em gerar subconjuntos de treinamento distintos, tomando como base o conjunto original de dados. A ideia é que, com a adoção do processo

casual, seja obtida certa diversidade entre os conjuntos construídos.

Conforme apresentado na Figura 2.2, representantes do conjunto original são sorteados até que o subconjunto tenha a mesma dimensão do grupo base. Dado que a aleatoriedade é aplicada com reposição, um elemento pode aparecer repetidas vezes em um mesmo subconjunto, bem como ser selecionado diversas vezes para subconjuntos distintos (STEFANOWSKI, 2005). Em decorrência da repetição de elementos, várias instâncias do bloco inicial não estarão presentes no novo conjunto. Segundo Dietterich (DIETTERICH, 2000) e Skurichina & Duin (SKURICHINA; DUIN, 2002), cada subgrupo conterá, em média, 63.2% da formação original.

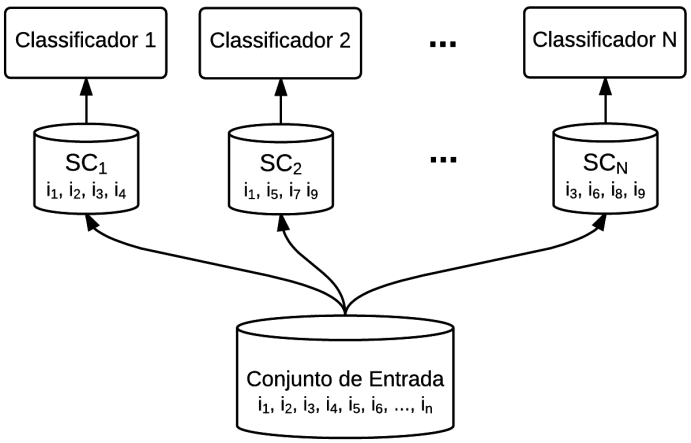

Figura 2.2: Estrutura do funcionamento do Bagging

Segundo Panov & Dzeroski (PANOV; DEROSKI, 2007) o método é indicado para algoritmos instáveis, os quais sofrem grande influência de pequenas variações no conjunto de treino.

2.1.2 Boosting

O Algoritmo de *Boosting*, assim como o *Bagging*, baseia-se na ideia de sorteio considerando-se o conjunto de treinamento. Entretanto, nesta abordagem a escolha é feita considerando pesos para cada instância. O processo consiste em sortear um conjunto de elementos aleatoriamente, onde, inicialmente, todos têm a mesma chance de serem selecionados. Então é feita a classificação das amostras sorteadas. Aquelas que forem classificadas erroneamente terão seus pesos aumentados, fazendo com que, em um sorteio seguinte, tenham mais chances de serem selecionadas a compor o novo subconjunto. As instâncias que são rotuladas indevidamente são consideradas difíceis (FREUND; SCHAPIRE, 1996).

A Figura 2.3 ilustra o funcionamento da abordagem. Verifica-se que o conjunto de treinamento de uma etapa é processado e então serve de "entrada"para a fase seguinte. Esta dependência deve-se à atualização dos pesos de cada instância. O processo iterativo, no qual gera-se um novo classificador a cada iteração, é realizado até que o número desejado de classificadores seja atingido.

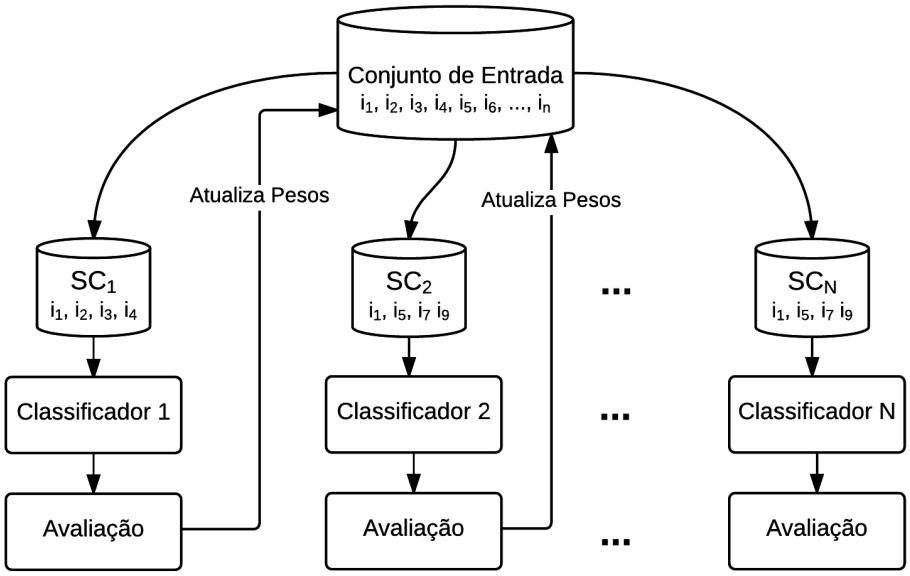

Figura 2.3: Ideia do funcionamento do Boosting

Conforme Freund & Shapire (FREUND; SCHAPIRE, 1996) e Quinlan (QUINLAN, 1996), como o método atribui pesos maiores às instâncias classificadas incorretamente, ele tende a focar nos classificadores relativamente mais fracos. Todavia, verificou-se que com a combinação dos vários classificadores fracos, consegue-se obter o equivalente a um classificador ótimo.

2.1.3 Random Subspaces (RSS)

Proposto por Ho (HO, 1998), esta técnica constrói o novo classificador por meio do sorteio de subespaços do conjunto de atributos da base de treinamento. A ideia é que, dentre um conjunto de n características para cada instância, sejam selecionados k atributos aleatoriamente (em que $k < n$) para compor cada classificador. A Figura 2.4 demostra o funcionamento do método.

Na ilustração, o conjunto inicial é composto de n características, das quais apenas 4 são sorteadas para compor os novos classificadores. É importante destacar que não devem ser sorteados atributos repetidos para formar um mesmo elemento, uma vez que tal repetição não traria ganho no momento da classificação. Todavia, classificadores distintos

podem possuir características em comum.

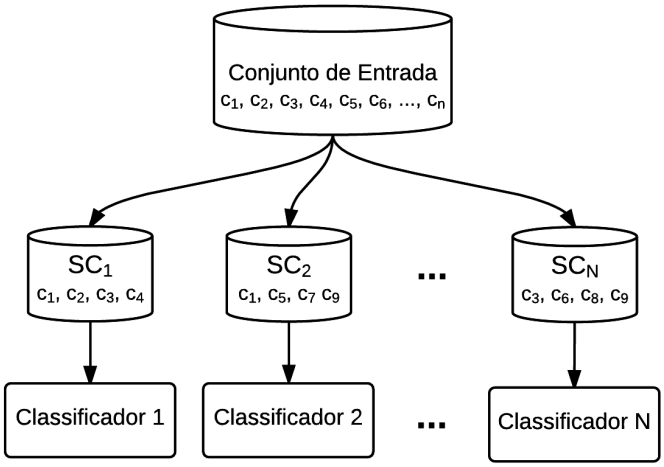

Figura 2.4: Construção de classificadores via Random Subspaces

Segundo Ponti (PONTI JR., 2011) a escolha casual dos atributos deve criar classificadores que são complementares, o que faz com que cometam erros diferentes, que é uma característica positiva em cenários de combinação de classificadores.

A aplicação do RSS é indicada para cenários em que o conjunto de dados é composto de muitos atributos e com características redundantes, visto que o método consegue evitar a maldição da dimensionalidade (HO, 1998), (KUNCHEVA et al., 2001), (PONTI JR., 2011).

2.1.4 Targeted-Complexity Problems

Uma abordagem alternativa para geração de classificadores é proposta em (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010). O foco deste método, diferente dos anteriores que buscam explorar a diversidade, reside no estudo da complexidade do problema em análise. A ideia, segundo os autores, é que os problemas reais não permitem testar minuciosamente o comportamento das regiões de fronteira por não cobrir todo o espaço de complexidade, carecendo de uma estratégia que permita um estudo mais aprofundado de tais medidas.

Após estudo sobre 264 problemas binários, os autores verificaram que mesmo com tal gama de problemas não foi possível explorar de forma efetiva a complexidade dos problemas. Concluiu-se que tal fato pode estar relacionado às amostras que formam o problema (as amostras que compõe o problema não permitem uma exploração minuciosa) ou ao fato de o problema real não possuir característica que permita tal exploração.

A técnica apresentada baseia-se em um algoritmo genético (AG) multiobjetivo,

cujas funções de otimização consistem em minimizar ou maximizar as medidas de complexidade. O algoritmo forma novas instâncias sintéticas com base nas amostras reais que formam o problema. Estas instâncias artificiais tendem a oferecer uma cobertura mais completa do estudo de complexidade do problema. A etapa de cruzamento é responsável pela geração das novas instâncias, umas vez que há a troca de "segmentos"dos vetores de características entre dois indivíduos da população.

Após um experimento executado sobre três bases reais, verificou-se a viabilidade da aplicação de um algoritmo genético na geração dos classificadores com o objetivo de alcançar um espaço de complexidade mais abrangente do que o problema original. Os autores destacam, no entanto, o custo computacional necessário, uma vez que há o processamento envolvido no AG e também do cálculo das medidas de complexidade. Há também preocupação inerente com o número de objetivos adotados, uma vez que a competência do método decresce conforme aumentam os objetivos (principalmente com mais de três alvos).

Dentre as abordagens heterogêneas destacam-se *Stacking* (WOLPERT, 1992) que consiste em realizar o processo de classificação das instâncias por diferentes algoritmos de classificação e comparar os resultados visando determinar qual é o mais confiável e *StackinC* (SEEWALD, 2003), que adota abordagem similar ao *Stacking*, porém avalia a relevância dos atributos dos dados visando eliminar os *features* de forma a reduzir a dimensionalidade do processo.

2.1.5 Diversidade entre Classificadores

A presença de diversidade em um conjunto de classificadores desempenha papel fundamental nos SMCs, permitindo que o desempenho de comitês de classificadores possa ser superior ao de abordagens individuais (SHIPP; KUNCHEVA, 2002), (KUNCHEVA; WHITAKER, 2003), (BROWN et al., 2005), (WINDEATT, 2005). Todavia, dada a complexidade de interpretação da diversidade, não há ainda consenso acerca de seu grau de influência efetiva na acurácia dos métodos.

Segundo Ponti Jr. (PONTI JR., 2011), um ponto de consenso é que, quando os classificadores cometem erros estaticamente diferentes, a combinação destes tem potencial para melhorar a performance do sistema. Uma classificação da diversidade em níveis é proposta pelo autor e por Brown *et al.* (BROWN et al., 2005):

- Para cada padrão não mais que um classificador está errado. Não há coincidência dos erros de modo que a função alvo é coberta.

- Há a ocorrência de alguns erros coincidentes, no entanto, a maioria está sempre correta. Contudo, há a necessidade de que o *ensemble* tenha dimensão superior a 4 classificadores.

- O voto da maioria nem sempre implicará em resposta correta, porém, pelo menos um classificador está certo para cada padrão.

- Todos os classificadores estão errados para alguns padrões. Neste cenário, a função alvo não é totalmente coberta.

Visando obter uma compreensão mais acurada da diversidade no processo classificatório, bem como avaliar a relação dela com a acurácia dos *ensembles*, Kuncheva & Whitaker (KUNCHEVA; WHITAKER, 2003) apresentam uma relação de dez medidas de diversidade, das quais 4 são medidas entre pares de classificadores e 6 são medidas que trabalham com conjuntos de classificadores. Fazem parte do primeiro grupo as medidas de estatística Q, correlação, falta dupla e discordância, enquanto no segundo grupo constam a entropia dos votos, índice de dificuldade, variância de Kohavi-Wolpert, a relação de concordância entre classificadores, a diversidade generalizada e a diversidade de erros coincidentes.

Além do estudo das medidas de diversidade, há pesquisas que buscam construir *ensembles* de forma a contribuir positivamente para a diversidade. Os métodos de formação de *pools* que empregam medidas de diversidade no processo construtivo são ditos explícitos (enquanto aqueles que não adotam tal fator, como *Bagging*, *Boosting* e RSS, são chamados de métodos implícitos) (KUMAR; KUMAR, 2012).

Uma abordagem que busca criar heterogeneidade entre os classificadores empregando-se dados artificiais é proposta por Melville & Mooney (MELVILLE; MOONEY, 2004). Os autores apresentam o método DECORATE (*Diverseensemble Creation by Oppositional Relabeling of Artificial Training Examples*) que, com base em meta-classificadores, pode usar classificadores robustos para construir comitês. A acurácia do método mostrou-se superior ou equivalente aos métodos de *Bagging*, *Boosting* e RSS para um conjunto composto por 15 problemas disponíveis na *UCI Machine Learning* (BACHE; LICHMAN, 2013).

Uma exploração mais detalhada das abordagens de geração de diversidade em *ensembles* é apresentada por Brown *et al.* (BROWN et al., 2005). Os autores apresentam um estudo aprofundado da interpretação da diversidade e apresentam uma ideia inicial de rotulação de métodos de criação de diversidade em explícitos e implícitos, bem como cenários em que estes podem ser aplicados.

Medidas de diversidade podem contribuir também no momento da seleção dos classificadores, conforme apresentado por Santana *et al.* (SANTANA et al., 2006). Os

autores propuseram duas abordagens de seleção dinâmica de classificadores com base na acurácia e diversidade dos mesmos. Os resultados obtidos mostram a viabilidade da adoção da diversidade como fator de escolha dos classificadores na formação dos *ensembles*. A diversidade foi adotada como critério de seleção dinâmica de classificadores também no trabalho de Yan *et al.* (YAN et al., 2013).

2.2 Seleção Dinâmica de Classificadores

Segundo Giacinto & Roli (GIACINTO; ROLI, 1999) e Ayad & Syed-Mouchaweh(AYAD; SYED-MOUCHAWEH, 2011) a maioria dos métodos de combinação assume que os classificadores envolvidos produzem diferentes erros de rotulação, tais técnicas são conhecidas como fusão (cujas topologias são apresentadas na seção 2.2.3). Entretanto, em aplicações reais de reconhecimento de padrões, geralmente há dificuldade em se encontrar classificadores que satisfaçam o pressuposto dos erros independentes.

Uma forma encontrada para evitar a premissa das falhas independentes é a seleção dinâmica de classificadores. Esta baseia-se no antecedente de que cada classificador é especialista em alguma região do espaço de características (AKSELA, 2003) e (AYAD; SYED-MOUCHAWEH, 2011) o que permite que, dentre um conjunto de classificadores, haja um ou vários que consigam rotular corretamente a instância em avaliação. O desafio reside em como determinar o elemento mais apto para classificar a instância.

A seleção do classificador ou classificadores pode ser realizada de forma estática ou dinâmica (GUNES et al., 2003),(KO; SABOURIN; BRITTO JR., 2008),(YU-QUAN et al., 2011). A Figura 2.5 ilustra três abordagens distintas para a etapa de seleção. No primeiro cenário (Figura 2.5(a)) é representada a escolha estática de um conjunto de classificadores. Nesta abordagem, o grupo escolhido é empregado na classificação de todas as amostras. As representações restantes delineiam o funcionamento da escolha dinâmica de um classificador (Figura 2.5(b)), onde é definido um único classificador para rotular cada nova instância; e seleção dinâmica de um conjunto de classificadores (Figura 2.5(c)), onde são elencados vários classificadores distintos para cada instância a ser classificada.

A seleção estática é realizada durante a fase de treinamento, sem considerar as características dos dados a serem classificados (AYAD; SYED-MOUCHAWEH, 2011). Neste cenário, os classificadores que se mostraram mais acurados são escolhidos para formar o grupo empregado para rotular todas as novas instâncias. A seleção dinâmica, entretanto, realiza a escolha do(s) classificador(es) levando em conta as particularidades de cada amostra do grupo de teste. Dessa forma, os classificadores que participam da rotulagem podem variar de acordo com a instância em foco.

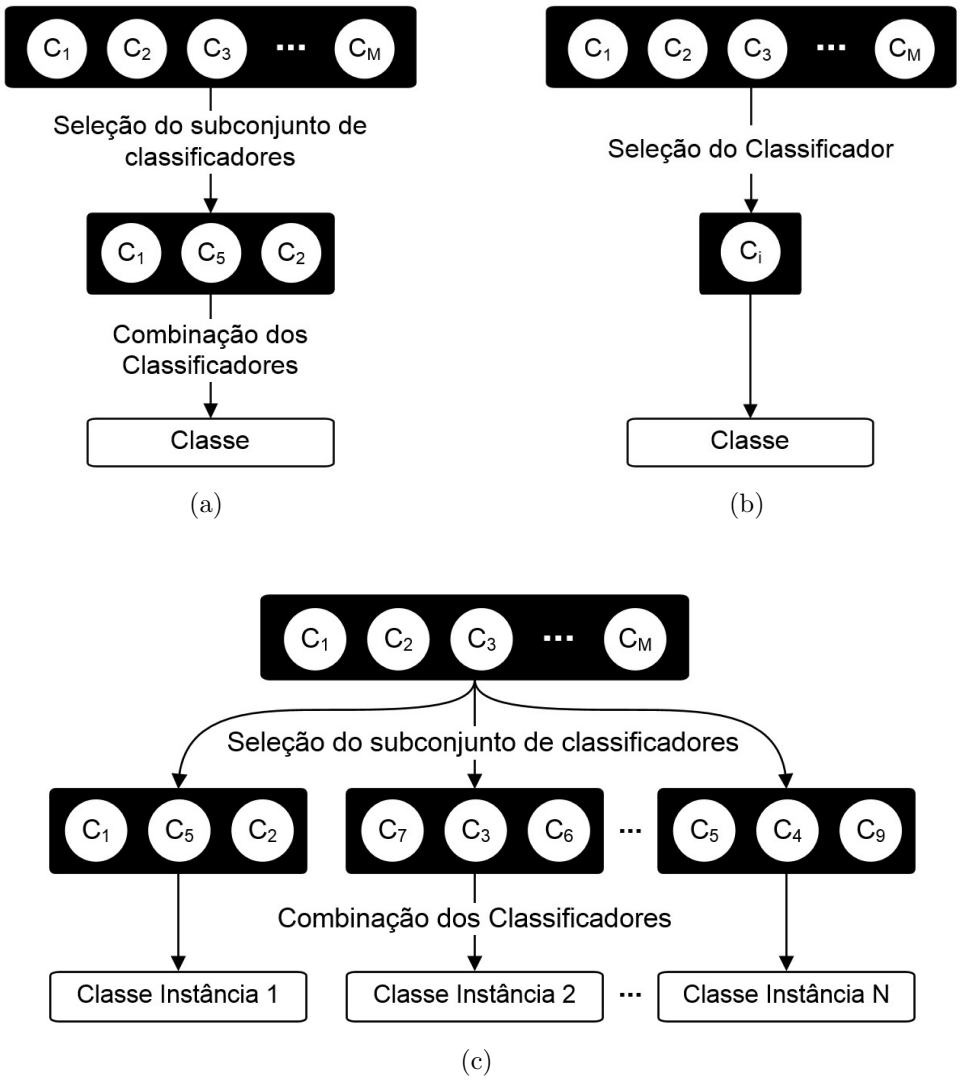

Figura 2.5: Três abordagens para seleção e combinação de classificadores (Adaptado de [(KO; SABOURIN; BRITTO JR., 2008)]): a) seleção estática de conjunto de classificadores; b) seleção dinâmica de classificador único e c) seleção dinâmica de conjunto de classificadores

A adoção da seleção dinâmica visa explorar de forma mais efetiva a variabilidade dos erros dos classificadores e a diversidade destes no intuito de melhorar a acurácia da classificação em comparação à seleção estática (TSOUMAKAS; PARTALAS; VLAHAVAS, 2008). Pesquisas apontam para esta melhoria no desempenho dos classificadores, dentre as quais destacam-se os trabalhos de Woods, Kegelmeyer & Bowyer (WOODS; KEGELMEYER JR.; BOWYER, 1997), Giacinto & Roli (GIACINTO; ROLI, 1999), Gunes *et al.* (GUNES et al., 2003), Kuncheva & Whitaker (KUNCHEVA; WHITAKER, 2003), Didaci & Giacinto (DIDACI; GIACINTO, 2004) e Didaci *et al.* (DIDACI et al., 2005).

Segundo a taxonomia proposta em (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) a seleção dinâmica de classificadores pode se basear em duas estratégias principais: aquelas fundamentadas em características individuais e aquelas que baseiam-se em informações

coletivas dos classificadores. No primeiro grupo os classificadores são selecionados com base na sua competência individual no espaço de características representado pelo conjunto de treino ou validação, ou em uma determinada região local. Fazem parte deste grupo as seleções baseadas em ranking, em acurácia, probabilísticas, em comportamento ou mesmo em oráculo.

Já no segundo grupo a competência dos classificadores é determinada pela combinação de acurácia dos classificadores base com alguma informação relacionada à interação existente entre os elementos do *pool*, tal como diversidade, ambiguidade ou complexidade. As estratégias mais comuns deste grupo são as seleções baseadas em diversidade, em ambiguidade ou na manipulação dos dados.

Uma diferente taxionomia para as técnicas de seleção dinâmica é apresentada em (CRUZ et al., 2015). Nela os autores dividem as estratégias em três grupos: 1) Acurácia local do classificador: inicialmente é definida uma pequena região no espaço de características ao redor da instância de teste, chamada região de competência. Então, avalia-se a acurácia dos classificadores sobre os elementos que compõe esta região; 2) Decision templates: nesta categoria busca-se selecionar aquelas instâncias que são parecidas com o padrão de teste. Para tanto, geralmente se cria um perfil de saída para as instâncias para avaliar a similaridade entre as instâncias; 3) Medida de consenso ou similaridade: diferente das demais, técnicas desta categoria trabalham com conjuntos de *ensembles* de classificadores onde, dada a instância de teste, o nível de competência do*ensemble* é definido pelo grau de consenso entre seus classificadores base.

O foco desta pesquisa no entanto, reside apenas sobre as estratégias de seleção dinâmica, as quais são detalhadas nas seções seguintes: a seleção dinâmica de classificador individual é tratada na Seção 2.2.1 enquanto a seleção dinâmica de *ensembles* é abordada na Seção 2.2.2.

2.2.1 Seleção Dinâmica de Classificador Único

Conforme apresentado na Figura 2.5, o processo de selecionar classificadores dinamicamente busca encontrar aquele ou aqueles que mais se ajustam à cada uma das novas instâncias. Na seleção dinâmica de um classificador único é atribuído o rótulo à nova instância com base na decisão feita pelo classificador escolhido. O sucesso desta técnica depende de quão confiável é o classificador escolhido (KUNCHEVA; WHITAKER, 2003).

Nas seções seguintes são apresentadas algumas das abordagens de seleção dinâmica individual mais comuns na literatura.

2.2.1.1 Acurácia Local Total - OLA

Esta abordagem realiza a escolha do classificador para a instância x* com base na acurácia local (WOODS; KEGELMEYER JR.; BOWYER, 1997),(DIDACI et al., 2005). Inicialmente cada classificador deve rotular os vizinhos mais próximos à instância x*. Será escolhido o classificador que conseguir classificar corretamente o maior percentual dos k vizinhos de x*, conforme a Equação 2.1.

$$C_j|LA_{j,k}(x^*) = max_i(\frac{K_{T,i}}{K}) \tag{2.1}$$

em que K corresponde ao número de vizinhos da instância em análise, enquanto $K_{T,i}$ é o número de vizinhos que classificador i classificou corretamente.

2.2.1.2 Acurácia Local da Classe - LCA

Inicialmente a instância é atribuída por um classificador à uma determinada classe ω_p, então calcula-se a razão entre o número de vizinhos (entre os k mais próximos) de x* classificados corretamente com o rótulo ω_p e o número total de vizinhos classificados como ω_p (mesmo que incorretamente). O classificador que apresentar a maior relação é o escolhido (WOODS; KEGELMEYER JR.; BOWYER, 1997)(DIDACI et al., 2005), como demonstrado na Equação 2.2.

$$LA_{j,k}(x^*) = max_i(\frac{N_{pp}}{\sum_{i=1}^{M} N_{ip}}) \tag{2.2}$$

em que N_{pp} refere-se ao número de vizinhos corretamente rotulados como ω_p, enquanto $\sum_{i=1}^{M} N_{ip}$ representa o total de vizinhos de x* classificados como ω_p pelo classificador i.

2.2.1.3 Seleção A Priori

Proposta por Didaci *et al.* (DIDACI et al., 2005) e Giacinto & Roli (GIACINTO; ROLI, 1999), esta abordagem calcula, com base na probabilidade do classificador acertar a classe dos k vizinhos mais próximos da instância x*, a acurácia de cada classificador. A Figura 2.6 ilustra a ideia do funcionamento da abordagem. Na imagem, o hexágono central representa o elemento a ser classificado, enquanto os elementos $V_1, .., V_5$ em coloração preta referem-se aos vizinhos mais próximos da instância. Já os vizinhos em coloração branca não fazem parte da vizinhança imediata de x*. As setas em vermelho correspondem à distância euclidiana até cada um dos k vizinhos mais próximos e, as regiões hachuradas, consistem nas vizinhanças individuais de $V_1, .., V_5$.

Inicialmente são encontrados os k vizinhos da instância x* a ser classificada. No exemplo, os elementos selecionados, $V_1, .., V_5$, têm seus pesos calculados (utilizando o inverso da distância euclidiana). Então, para cada um dos vizinho V_i de x*, calcula-se a proporção de seus vizinhos classificados corretamente pelo classificador. Posteriormente, a proporção dos vizinhos é multiplicada pelo peso de cada um deles e então somados. O resultado deste somatório é então dividido pelo somatório dos pesos dos vizinhos de x*. No cenário apresentado, são obtidos 5 proporções e 5 pesos, para cada um dos vizinhos $V_1, .., V_5$. O método selecionará o classificador que apresentar o maior somatório, indicando que, dentro daquela região, ele é o mais apto a definir a classe da instância de teste. A Equação 2.3 define, matematicamente, a ideia da abordagem A Priori.

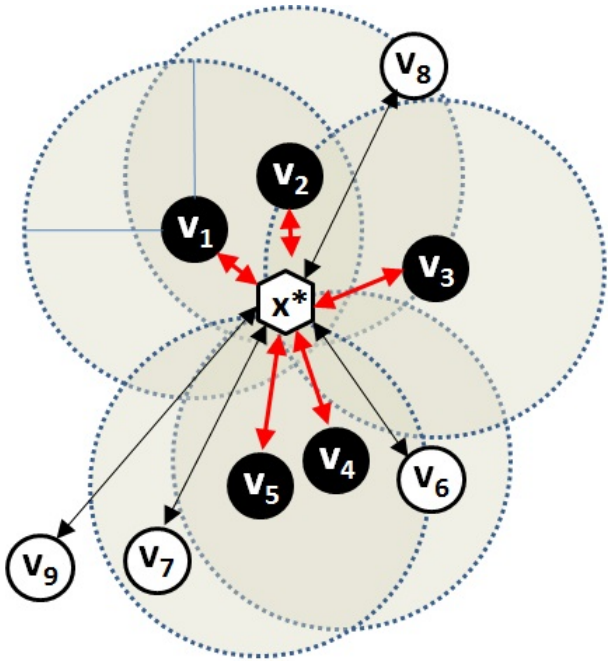

Figura 2.6: Avaliação da vizinhança da instância a ser classificada

$$C_* = arg_i max \frac{\sum_{j=1}^{N} \hat{p}(\omega_k | x_j \epsilon \omega_k, c_i) W_j}{\sum_{j=1}^{N} W_j} \tag{2.3}$$

em que N corresponde ao número de vizinhos considerados para cada um dos V_i de x*. A probabilidade do classificador acertar o rótulo de cada vizinho V_i é representada por $\hat{p}(\omega_k | x_j \epsilon \omega_k, c_i)$, enquanto W_j corresponde ao peso de cada vizinho até a instância de teste.

2.2.1.4 Seleção A Posteriori

O método proposto por Didaci *et al.* (DIDACI et al., 2005) e Giacinto & Roli (GIACINTO; ROLI, 1999) calcula a relação entre o somatório da probabilidade dos vizinhos de

x* serem classificados com a mesma classe ω_p e o somatório das probabilidades das classes a que seus k vizinhos pertencem. O passo inicial é calcular o peso W_j de cada vizinho V_i até a instância x*. Então, o classificador deve atribuir um rótulo ω_p a cada vizinho V_i. Em seguida, é calculada a proporção de vizinhos de V_i corretamente classificados como ω_p perante o total de vizinhos que receberam tal rótulo. A proporção é então multiplicada pelo peso do vizinho V_i e adicionados a um somatório, que representa o numerador da Equação 2.4. Calcula-se também o somatório da quantidade de vizinhos de V_i que receberam o rótulo ω_p multiplicados pelo peso de cada vizinho. Este segundo somatório corresponde ao denominador da equação. O classificador que apresentar a maior relação entre os acertos da classe ω_p e o total de ω_p atribuídos é escolhido para classificar a instância x*.

$$C_*(\omega_k) = arg_i max \frac{\sum_{x_j \epsilon \omega_k} \hat{p}(\omega_k|x_j, c_i)W_j}{\sum_{j=1}^{N} \hat{p}(\omega_k|x_j, c_i)W_j} \tag{2.4}$$

2.2.1.5 Seleção baseada em Comportamento - MCB

Em seu trabalho, Giacinto *et al.* (GIACINTO; ROLI; FUMERA, 2000) propuseram uma abordagem baseada no comportamento (*Multiple Classifier Behavior - MCB*) dos classificadores para a escolha do classificador mais adequado para cada padrão de teste. A ideia é avaliar o comportamento apresentado para instâncias de treino similares à instância a ser classificada e, segundo a conduta adotada, classificar o elemento na classe mais adequada.

Os autores definem comportamento como sendo o conjunto de opiniões dos classificadores para uma instância qualquer. Neste sentido, o método constrói um vetor de comportamento para cada amostra de treino onde, nesta estrutura, são armazenadas as opiniões de todos os classificadores. Assim, sabe-se exatamente a atitude tomada (classe atribuída), pelos classificadores para cada elemento individualmente.

O processo de classificação então consiste em, dada a instância a ser classificada, obter a opinião de todos os classificadores sobre ela. Em seguida, encontrar nos vetores de comportamento de treino, aqueles que apresentaram o mesmo comportamento do padrão de teste. Escolhe-se então o classificador que acertar o maior número de instâncias que possui o mesmo comportamento do elemento em avaliação. Caso não haja, no conjunto de treino, amostras com comportamento semelhante, trabalha-se com uma folga, escolhendo aqueles que se comportam mais similarmente ao objeto a ser classificado.

A Equação 2.5 apresenta o cálculo da acurácia dos classificadores. O termo $\hat{P}_j(\omega_i|X_n)$ corresponde à acurácia do classificador C_j para a instância de treino X_n, uma vez que esta instância é igual ou similar ao padrão de teste. Já W_n corresponde ao peso

de X_n em relação à X^*, calculado pelo inverso da distância euclidiana entre as amostras. O método seleciona o classificador que maximizar o valor de CA.

$$CA_j(X^*) = \frac{\sum_{x_n \epsilon \omega_i} \hat{P}_j(\omega_i|X_n) \cdot W_n}{\sum_{m=1}^{M} \sum_{x_n \epsilon \omega_m} \hat{P}_j(\omega_i|X_n) \cdot W_n} \tag{2.5}$$

Os experimentos foram realizados sobre um conjunto de três problemas disponíveis na base de dados ELENA (*Enhanced Learning for Evolutive Neural Architeture*). Os resultados mostraram que a abordagem é mais adequada do que a seleção estática visto que obteve desempenho superior à adoção do melhor classificador em todos os casos e acurácia similar ou superior à combinação pelo voto majoritário.

2.2.2 Seleção Dinâmica de Conjunto de Classificadores

A seleção dinâmica de conjunto de classificadores visa elencar um grupo de n classificadores perante as N possibilidades, buscando formular uma decisão mais subsidiada ao invés de se basear em apenas um classificador. A seleção de parte do comitê de classificadores ao invés de utilizar todos no processo de classificação pode levar a resultados mais acurados, entretanto, a escolha do subconjunto ótimo de classificadores não é uma tarefa trivial (YAN et al., 2013).

Algumas das abordagens de seleção dinâmica de comitês mais comuns na literatura são apresentadas nas seções a seguir.

2.2.2.1 K Oráculos mais Próximos - KNORA

O método KNORA (*K-Nearest-ORAcles*), proposto por Ko *et al.* (KO; SABOURIN; BRITTO JR., 2008) busca encontrar, para cada instância x*, o conjunto de classificadores que consegue classificar de forma mais precisa, os k vizinhos de x*. O pressuposto é de que os classificadores com maior acurácia na vizinhança do padrão de teste, têm, em teoria, maior competência em atribuir rótulo à instância.

Esta abordagem emprega o conceito de Oráculo, que, segundo Kuncheva & Rodriguez (KUNCHEVA; RODRIGUEZ, 2007), consiste na descoberta do classificador que é mais apto para classificar a instância em questão. Ao se compor um conjunto de classificadores com base nos mais competentes, aumenta-se a chance de sucesso na classificação das amostras.

São propostas duas abordagens: *KNORA-Eliminate* (KN-E) e *KNORA-Union* (KN-U). Na primeira são selecionados os classificadores que conseguem classificar corretamente pelo menos n dos k vizinhos (em que $n <= k$) de x*. Conforme ilustrado na

Figura 2.7 (quadro central), os classificadores que acertaram a classe de cada um dos $V_1, .., V_5$ vizinhos são selecionados para o processo de classificação de x*. O processo de combinação dos classificadores emprega o voto majoritário simples e ponderado (*KNORA-Eliminate Weighted* - KN-E-W).

Já o *KNORA-Union*, menos incisivo, escolhe os classificadores que conseguem rotular corretamente pelo menos um dos k vizinhos de x* (quadro à direita na Figura 2.7). Assim como na abordagem *eliminate*, os processo de combinação emprega o voto majoritário simples e ponderado (*KNORA-Union Weighted* - KN-U-W).

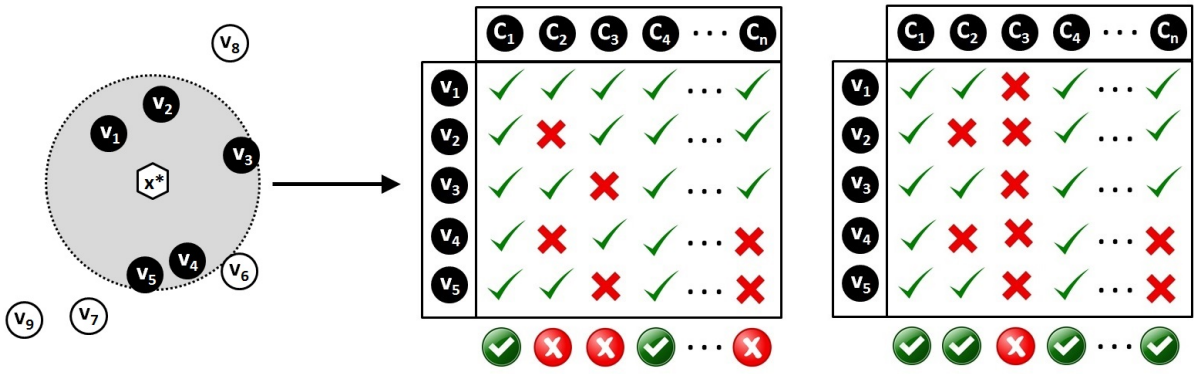

Figura 2.7: Ideia do funcionamento dos métodos KNORA-Eliminate e KNORA-Union

Os experimentos, que foram realizados sobre seis problemas provenientes do repositório da *UCI Machine Learning*, compararam várias implementações de seleção de classificadores, estáticas e dinâmicas, unitárias e de comitê, trabalhando sobre um conjunto composto por 10 classificadores construídos pelos métodos de *Bagging*, *Boosting* e *Random Subspaces*. As abordagens estáticas foram a escolha do melhor classificador e da combinação de todos os classificadores. As técnicas de seleção dinâmicas avaliadas foram OLA, LCA, A Priori, Posteriori, KN-E, KN-E-W, KN-U e KN-U-W. Os resultados mostraram que os KNORAS obtiveram desempenho superior às técnicas de seleção estáticas e ligeiramente superior às demais abordagens de seleção dinâmica.

2.2.2.2 Seleção baseada em Ranking

Em seu trabalho, (SABOURIN et al., 1993) o ranking é construído pela estimação de três parâmetros relacionado à exatidão dos classificadores do *pool*. A informação mútua destes três parâmetros é estimada aplicando-se parte dos dados de treino. Os parâmetros adotados são a distância até o vencedor, distância até o primeiro não vencedor, distância média entre o vencedor e o primeiro não vencedor. A ideia empregada no cálculo da informação mútua é avaliar o nível de incerteza na decisão relacionada a cada um dos

parâmetros de classificação. Após determinados os critérios que mais contribuem para o processo de classificação, é construído um meta-espaço que armazena os valores dos parâmetros de classificação para cada elemento.

No momento da seleção, os valores dos parâmetros dos classificadores associados à vizinhança do padrão de teste e ordenados de acordo com a acurácia e, o melhor deles é selecionado para classificar a instância em avaliação.

Os experimentos realizados sobre a base NIST mostraram que o método superou a solução monolítica, inclusive diminuindo o processamento desprendido no processo, uma vez que foi empregada a poda do conjunto de treino.

2.2.2.3 Seleção baseada em Diversidade e Acurácia

Uma proposta que adota a diversidade como critério para seleção dos classificadores de forma dinâmica para a construção de *ensembles* é apresentada em (SANTANA et al., 2006). Os autores, utilizam a acurácia do classificador em conjunto com a diversidade. O trabalho apresenta duas abordagens distintas para a formação do comitê. A primeira usa um algoritmo de agrupamento (k-means) enquanto a segunda emprega o método de vizinhos mais próximos (KNN).

Na abordagem de agrupamento os dados de validação são separados em k grupos usando-se o k-means. Então, para cada cluster, constrói-se uma lista de classificadores ordenada de forma crescente para diversidade e decrescente para acurácia. Para determinar a diversidade de cada classificador, foi adotada a medida de falta dupla (KUNCHEVA; WHITAKER, 2003). Então, no momento da classificação, cada padrão de teste é atribuído ao cluster que possuir o centróide mais próximo. Na sequência selecionam-se os N classificadores mais acurados. Deste grupo, são escolhidos os J (em que $J <= N$) elementos com maior diversidade, os quais, segundo voto majoritário, atribuem a classe à instância de teste.

A segunda estratégia encontra, para cada padrão de teste, os k vizinhos mais próximos e então, com base nessa vizinhança, constrói a lista de classificadores ordenados de forma decrescente em acurácia e crescente de diversidade. Na etapa seguinte são escolhidos os N classificadores com maior acurácia e, dentre estes, os J com maior diversidade. Tal como na abordagem anterior, a instância é classificada pelo voto majoritário dos J classificadores.

Visando avaliar ao desempenho dos métodos, foram executados experimentos com duas bases de dados, uma representando sequências de DNA e a outra estruturas de proteínas. Em paralelo às duas abordagens de seleção de comitês de forma dinâmica,

foram executadas também as seleções estática e dinâmica de um classificador. Os autores empregaram *ensembles* de tamanho 10 ($N = 6$ e $J = 3$) e 15 ($N = 15$ e $J = 10$). Os resultados apontaram que ambas abordagens de seleção dinâmica de conjuntos obtiveram acurácia superior às abordagens de seleção estática e dinâmica de um classificador. Entre as abordagens de grupamento e de vizinhança, a primeira demonstrou ligeira vantagem de desempenho.

2.2.2.4 Seleção baseada em Diversidade - SDES

Uma segunda abordagem de seleção dinâmica de comitês de classificadores que emprega como meta a diversidade é proposta em (YAN et al., 2013). O método, chamado *Sorting-Based Dynamic Classifier Ensemble Selection* (SDES), baseia-se na ideia de que quanto maior a diversidade entre os classificadores selecionados, maior a chance de acerto na classificação das instâncias. Os autores buscam contornar a necessidade de se encontrar os K vizinhos mais próximos de cada instância de teste.

O algoritmo divide-se em duas etapas. A primeira realiza a ordenação decrescente dos classificadores de acordo com sua diversidade perante os demais, empregando o índice K_p como medida de concordância. Esta medida, no entanto, considera apenas a relação entre dois classificadores. Dessa forma, para se calcular a diversidade geral do classificador, os autores realizaram o somatório entre cada classificador em comparação a todos os outros.

A segunda etapa realiza a seleção do subconjunto de classificadores para efetuar a classificação da instância. Os classificadores são selecionados segundo a ordenação construída na primeira etapa até que a confiança na classificação da instância para uma classe dentre as possíveis atinja um limiar pré-estabelecido, cujo valor geralmente é próximo de 1. Quando o patamar é atingido, a classe cuja confiança foi superior ao limiar é atribuída à instância.

Os testes foram realizados sobre um conjunto composto por 6 bases, das quais cinco pertencem ao repositório da UCI *Machine Learning Repository* e a sexta é a base NIST. O experimento comparou o desempenho do método construído frente a outras 5 abordagens, 4 estáticas (*Bagging*, *AdaBoost*, *Ordering pruning*, *Gasen*) e uma dinâmica (KNORA). Os resultados mostraram que o SDES pôde atingir taxas similares ao KNORA e superiores às demais (com exceção da base NIST, onde o método *AdaBoost* foi ligeiramente superior). No entanto, a eficiência do algoritmo em relação ao KNORA é significativamente maior, aumentando conforme o tamanho do problema em estudo.

2.2.2.5 Seleção baseada em Filtros e Distância Adaptativa - DES-FA

O trabalho desenvolvido por Cruz *et al.* (CRUZ; CAVALCANTI; REN, 2011) visa realizar a seleção dinâmica de *ensembles* baseando-se na melhora das regiões de competência. O intuito é diminuir ou eliminar instâncias que podem incorrer em erros de classificação, principalmente em métodos que consideram a vizinhança como critério na etapa de classificação.

O método apresentado, chamado DES-FA (*Dynamic Ensemble Selection by Filter + Adaptative Distance*), atua, por duas etapas, na preparação dos dados de validação. A primeira etapa, chamada *Edited Nearest Neighbor Filter* (ENN *Filter*), trabalha removendo ruídos nos dados, de forma a criar fronteira mais suaves entre as classes, eliminando amostras cuja vizinhança possui rótulo distinto. O processo aplica um classificador KNN sobre todas as instâncias do conjunto, excluindo aquelas que forem classificadas indevidamente.

A etapa seguinte, intitulada *K-Nearest Neighbor with Adaptative Distance* (ou *Adaptavive-KNN*), visa aplicar uma medida chamada distância adaptativa, de forma que instâncias cuja vizinhança pertença à mesma classe têm pesos maiores do que aquelas cuja vizinhança apresenta rótulos distintos. Esta adaptação é empregada no momento da seleção da vizinhança da instância de teste I_{te} no conjunto de treinamento. Quando o algoritmo está calculando a distância (como distância euclidiana ou manhatam) entre as instâncias de treino I_{tr} e I_{te}, ele considera também a distância de I_{tr} até seu vizinho mais próximo cujo rótulo difere do seu. Dessa forma, quanto maior a distância de I_{tr} até o primeiro vizinho divergente, mais peso ele recebe na escolha dos vizinhos de I_{te}. Assim, instâncias "cercadas"de elementos de classes distintas, têm chance menor de serem escolhidas no momento da classificação. A classificação emprega o algoritmo KNORA-E (KO; SABOURIN; BRITTO JR., 2008) para seleção dos conjuntos de classificadores.

Visando validar o método, foi realizado experimento composto de nove problemas de classificação, dos quais sete estão disponíveis na base da *UCI Machine Learning Repository* enquanto os dois restantes foram geradas artificialmente. Os autores empregaram o*Bagging* para geração dos classificadores, dos quais foram construídos *ensembles* de dimensão dez. Foram testadas abordagens do ENN usando tamanho 1, 3 e 5, em comparação ao KNORA-E, seleção estática de *ensembles* e *single best*. Os resultados mostraram que a abordagem proposta alcança resultados superiores às abordagens estáticas e também ao KNORA-E puro, fato que evidencia que a preparação das regiões de competência, através do ENN e Adaptative-KNN, permite a obtenção de resultados mais acurados.

2.2.2.6 Seleção ponderada pela Validação Cruzada - DWEC-CV

A atribuição dinâmica de pesos aos classificadores para realizar a classificação dos padrões de teste foi proposta por Yu-Quan *et al.* (YU-QUAN et al., 2011). O objetivo é explorar o fato de que os classificadores compostos por diferentes *features* conseguem classificar regiões distintas no espaço. Assim, o método desenvolvido, chamado *Dynamic Weighting Ensemble Classifiers based on Cross-Validation* (DWEC-CV), busca atribuir pesos de acordo com a acurácia em cada região.

Inicialmente são construídos os classificadores com metade dos atributos do conjunto de dados original. Sequencialmente são construídos grupos sobre os quais os classificadores terão sua acurácia medida. Esta etapa divide o conjunto de treino em M *folds* de forma estratificada, mantendo assim a proporção das classes. Então faz-se o processo de avaliação dos classificadores via validação cruzada, de cada *fold* contras as M-1 restantes. A ideia da validação cruzada visa explorar melhor bases com limitação de tamanho, fato que pode prejudicar o processo de classificação por apresentar amostras de treino insuficientes.

No momento da classificação, é feita a eliminação de "falsos vizinhos"que podem influenciar negativamente a classificação das instâncias de teste. Neste processo é adotada a abordagem baseada em comportamento proposta em (GIACINTO; ROLI; FUMERA, 2000) onde, os vizinhos com similaridade inferior a um patamar especificado são excluídos. Então, encontra-se os K elementos mais próximos da instância de teste e, avalia-se a acurácia de cada classificador sobre esta vizinhança. Caso o classificador consiga acertar todos os K elementos, ele poderá dar seu voto na classificação do padrão em análise. Os votos, no entanto, são ponderados de acordo com a acurácia do classificador naquele textitfold. Aqueles que forem mais aptos naquela região, terão maior peso no momento de atribuir o rótulo.

O método foi testado empregando-se um conjunto de 10 problemas de classificação da *UCI Machine Learning Repository* com menos de 800 registros. A escolha deste limite busca demonstrar que o DWEC-CV consegue explorar melhor conjuntos de dados de tamanhos menores. Os testes comparam o algoritmo proposto com técnicas estáticas (*Bagging*, *AdaBoost*, *Random Subspaces* e *single best*) e dinâmicas (LCA, DCS-MCB e KNORA-E). Observou-se que a abordagem proposta pelos autores consegue atingir patamares interessantes de acerto, mostrando-se equivalente às demais técnicas em alguns cenários e superiores em outros.

2.2.2.7 Seleção baseada em Ambiguidade

Em seu trabalho Santos *et al.* (DOS SANTOS; SABOURIN; MAUPIN, 2007) propuseram um mecanismo de seleção dinâmica de *ensembles* baseado na ambiguidade presente nos comitês. O intuito da técnica é, dentre um conjunto de *ensembles* disponíveis, selecionar aquele em que há a maior concordância na classificação de cada padrão de teste especificamente. Dessa forma, cada instância pode ser classificada por um conjunto distinto de classificadores.

O primeiro passo consiste na geração do *pool*, empregando-se o *Random Subspaces* para tal. Em posse do grupo total de classificadores, o algoritmo constrói um conjunto de *ensembles* que serão empregados na etapa de classificação. Para formar os *ensembles* é empregado um algoritmo genético uni e multi-objetivo, adotando como funções de maximização a ambiguidade e a diversidade das falhas coincidentes e como funções de minimização a medida de dificuldade e falta dupla.

No processo como todo, o conjunto de dados é dividido em 10 *folds*, dos quais um é utilizado como conjunto de otimização, um como conjunto de validação e outro como conjunto de teste. Os dois primeiros são empregados como parâmetros na construção dos *ensembles*, enquanto o último é aplicado na avaliação do método. Os 7 *folds* restantes formam o conjunto de treino que é usado na geração do *pool*.

No momento da seleção, todos os *ensembles* têm sua ambiguidade calculada para cada amostra de teste. Aquele que apresentar o menor valor é selecionado para atribuir rótulo à instância. Havendo empate, faz-se o voto majoritário e, caso a votação não seja conclusiva, descartam-se os *ensembles* e selecionam aquele(s) que apresentar(em) o segundo menor índice de ambiguidade.

Nos experimentos realizados foi construído um *pool* composto por 700 classificadores (gerados 100 a partir de cada *fold* de treinamento), dos quais formaram-se 21 *ensembles* compostos de 128 elementos. Nos testes, que compararam os métodos LCA, *single best* frente à nova implementação, foram utilizadas três bases de dados: dna, satimage e texture. Os resultados apresentados permitiram concluir que a técnica desenvolvida mostrou-se mais acurada que o LCA em todos os cenários. No entanto, a seleção estática do melhor classificador demonstrou a melhor acurácia em grande parte dos experimentos, dando assim margem para melhorias na abordagem proposta. Verificou-se também que a medida de dificuldade empregada no algoritmo genético, foi a que, em sua maioria, obteve as melhores taxas de acerto.

2.2.2.8 Seleção baseada em Oráculo Randômico Linear

Uma estratégia que combina fusão e seleção para realizar a seleção dinâmica de comitês de classificadores foi proposta por Kuncheva & Rodríguez (KUNCHEVA; RODRIGUEZ, 2007). A abordagem desenvolvida divide os classificadores construídos em dois subclassificadores empregando-se uma função linear aleatória. Dessa forma, cada subclassificador é responsável por um espaço distinto, cabendo ao oráculo, no momento da classificação, escolher qual dos dois é mais apto a rotular a instância de teste. A classe atribuída é decidida através de algum método de combinação, como voto simples.

O objetivo da adoção do oráculo é dividir um problema, representado pelo classificador, em dois problemas mais simples (os subclassificadores), de forma que a diversidade do conjunto todo seja aumentada. Segundo os autores, espera-se que os dois subclassificadores tenham desempenho maior, ou no mínimo igual, ao classificador monolítico e que a adoção de vários subconjuntos implique em maior diversidade do que a apresentada originalmente.

Para proceder a segmentação do classificador original, o método emprega uma função linear que funciona como um hiperplano, separando o conjunto do classificador em dois subconjuntos distintos. O processo consiste em sortear dois elementos do grupo e traçar uma linha entre eles. O hiperplano é então obtido perpendicularmente à reta construída. Realizada dessa forma, verifica-se que o objetivo da construção do hiperplano não é otimizar a divisão dos conjuntos, como ocorre no SVM mas incorrer em maior diversidade.

Visando avaliar a contribuição da abordagem para a seleção dos comitês, foram realizados dois experimentos. No primeiro utilizaram-se 35 bases de dados provenientes da UCI e no segundo um grupo composto de 7 problemas médicos reais. Durante os testes foram avaliados 20 abordagens de seleção de *ensembles*, para as quais avaliou-se o desempenho empregando-se o oráculo randômico em comparação aos cenários em que ele não era empregado. Observando os resultados constatou-se que realmente a adoção da técnica proposta apresenta ganho em acurácia para quase todos os cenários (em dez deles, o benefício alcançado foi significativo). As abordagens de seleção de *ensembles* que mais se beneficiaram foram o *Bagging* e Subespaços Aleatórios.

2.2.2.9 Seleção Adaptativa de Conjunto de Classificadores baseada em GMDH

Em seu trabalho, (XIAO; HE, 2009) apresentam uma solução para a seleção dinâmica de comitês de classificadores chamada GDES que se baseia na ideia do método GMDH

proposto por (IVAKHNENKO, 1970) onde o autor trabalha com uma rede neural multicamadas para combinar dois modelos visando formar novos protótipos empregando no processo critérios externos.

O GDES, que busca contornar a redundância dos classificadores gerada no GMDH, seleciona, para cada instância de teste, um subconjunto de classificadores apropriados do *pool* inicial, determinando os pesos da combinação entre estes classificadores em relação à vizinhança do padrão a ser classificado e selecionando aqueles em que a complexidade é maximizada.

O método foi testado sobre um conjunto de 6 bases disponíveis na UCI (as mesmas empregadas em (KO; SABOURIN; BRITTO JR., 2008) para critérios de comparação) construindo-se um *pool* composto de dez classificadores construídos por *Bagging* e adotando KNN de dimensão 1. Os resultados mostraram que a abordagem proposta pôde alcançar acurácia superior aos demais métodos de seleção dinâmica (LCA e KNORA-E) e estática (SB e MAJ) em três das bases e resultados bastante aproximados nas demais bases.

2.2.2.10 Seleção baseada em Overproduce-and-choose Dinâmica - SOCS

Uma proposta para seleção dinâmica de comitês de classificadores empregando a estratégia de *Overproduce-and-Choose Strategy* (OCS) foi desenvolvida em (DOS SANTOS; SABOURIN; MAUPIN, 2008) onde a primeira etapa busca gerar *ensembles* com alta acurácia enquanto a segunda aplica medidas de confiança para determinar qual deve ser o classificador ou comitê escolhidos.

O objetivo dos autores foi contornar os problemas apresentados pelo OCS estático que seleciona um *ensemble* único para rotular todos os novos padrões e que, no momento da seleção, trata todos os classificadores com a mesma importância, sem considerar fatores que podem formar um *pool* mais acurado. Para tanto, os autores propuseram uma solução que forma uma gama de *pools* e, de acordo com a instância a ser rotulada, seleciona aquele que melhor se adequar.

O primeiro passo do processo é construir o *pool* de classificadores. Nesta etapa foram empregados o *Bagging* e RSS. Em seguida são formados os *pools* dos classificadores através de dois algoritmos genéticos, um mono-objetivo e outro multiobjetivo. Como critérios de otimização do algoritmo genético foram empregadas a minimização do erro (adotado no AG de um objetivo) e a maximização da diversidade (onde foram empregadas a falta dupla, ambiguidade, falha coincidente da diversidade e medida de dificuldade no AG multiobjetivo). No momento da seleção dinâmica foram adotadas quatro abordagens

distintas (ambiguidade, margem, força e acurácia local).

Após a realização de testes com sete bases observou-se que o desempenho obtido pelo método SOCS foi superior à combinação de todos, ao *single best* e ao KNN individual. Verificou-se também ganho de acurácia em relação do método estático de seleção de *ensembles* via OCS, tanto para o AG mono quanto multiobjetivo.

2.2.2.11 Seleção dinâmica de ensembles baseada em Meta-Aprendizado - META-DES

De forma a tornar o processo de estimação da competência dos classificadores mais robusto, Cruz *et al.* (CRUZ et al., 2015) propõe a adoção de cinco meta-características, cada uma relacionada a um critério para efetuar a seleção dos classificadores: a dificuldade na classificação dos vizinhos; probabilidade a posteriori; acurácia total local; perfis de saída dos classificadores e a confiança dos classificadores.

Em seu trabalho, eles dividem o sistema de múltiplos classificadores proposto em três etapas. A primeira, *overproduce*, é responsável pela geração do *pool* de classificadores. Neste contexto foram gerados, através de bagging, 100 perceptrons (multi-class perceptrons para problemas com mais de duas classes) com 50% do tamanho do conjunto e treino. Na segunda, chamada meta-treino, os conjuntos de cinco meta-características são extraídas do treino e usadas para treinar o classificador que irá trabalhar como seletor na etapa seguinte, a generalização.

Na terceita fase as meta-características são extraídas da instância de teste e são passadas ao meta-classificador. Este tem a incumbência de estimar quando um classificador base é competente suficiente para rotular o novo padrão.

Para validar o método proposto os autores compararam seu desempenho com outros oito métodos de seleção dinâmica da literatura, os quais baseiam-se em apenas um critério para estimar a competência de cada classificador. Os testes foram realizados sobre um conjunto composto de 30 problemas de classificação. Os resultados mostraram que o META-DES pôde obter a maior acurácia na maioria das bases (18 de 30).

2.2.3 Combinação de Classificadores

Classificadores distintos, trabalhando sobre conjuntos de dados distintos, geralmente cometem erros diferentes. Tal fato possibilita que, através da combinação destes classificadores, se obtenha conjuntos que tomem decisões mais acuradas (KITTLER et al., 1998), (GUNES et al., 2003),(KUNCHEVA; WHITAKER, 2003) e (KO; SABOURIN; BRITTO JR.,

2008).

Quando é formado um conjunto de classificadores para realizar a rotulação de uma nova instância é necessário definir uma forma de combinar a "opinião" destes classificadores.

Segundo Ponti Jr. (PONTI JR., 2011), este processo pode ser realizado por meio da seleção, onde é escolhido um classificador do conjunto para atribuir o padrão à amostra, ou através da combinação (ou fusão) dos classificadores, em que todos os classificadores, segundo algum critério, contribuem na decisão do rótulo para a instância.

A combinação, entretanto, será mais efetiva apenas em cenários nos quais os classificadores individuais forem acurados e variados, ou seja, os classificadores selecionados precisam ter baixas taxas de erro e cometer erros independentes, chamados complementares (TUMER; GHOSH, 1996), (GUNES et al., 2003). Estes fatores podem ser satisfeitos usando-se diferentes espaços de características, diferentes conjuntos de aprendizagem ou classificadores variados (mudando-se a configuração de parâmetros ou os tipos dos classificadores).

A combinação, conforme (LU, 1996), (RANAWANA; PALADE, 2006) e (PONTI JR., 2011), pode seguir três abordagens (ou topologias): paralela, serial (ou em cascata) e híbrida. No primeiro método, apresentado na Figura 2.8, todos os classificadores realizam a mesma tarefa de classificação, trabalhando sobre o mesmo conjunto de dados e obtendo cada um, ao fim da sua execução, um rótulo para a nova instância. A "opinião" dos classificadores então deve ser combinada, segundo alguma regra, decidindo a classe mais apropriada a ser atribuída.

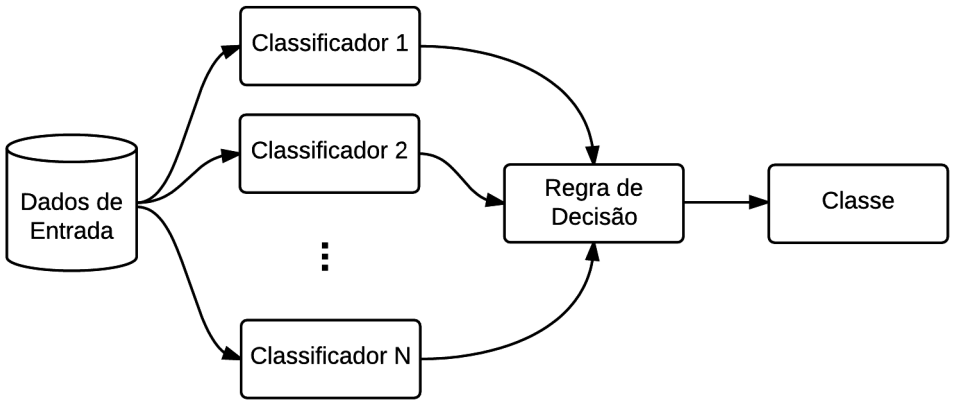

Figura 2.8: Topologia paralela

Segundo Gunes *et al.* (GUNES et al., 2003) e Kittler *et al.* (KITTLER et al., 1998), as formas mais simples de se combinar classificadores são: as regras da soma, produto, do

mínimo, do máximo, da média e da mediana. Além das abordagens mais básicas, podem ser citadas ainda o voto da maioria, voto ponderado, borda count, combinação bayesiana e métodos que empregam conceitos de inteligência computacional, como teoria das crenças (GUNES et al., 2003).

Na segunda abordagem, ilustrada na Figura 2.9, os classificadores são distribuídos em ordem crescente de complexidade, onde as instâncias são submetidas inicialmente ao classificador mais simples. Caso existam elementos rejeitados, com base em um patamar pré-estabelecido, eles são submetidos ao classificador seguinte, sendo reavaliados.

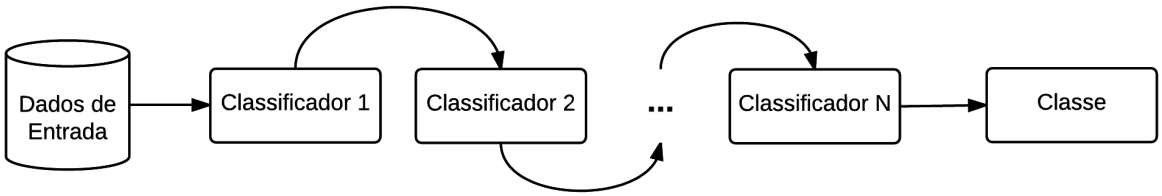

Figura 2.9: Combinação de classificadores pela abordagem serial

A cada iteração são reduzidas as instâncias e o número de classes (LAM, 2000), (RANAWANA; PALADE, 2006). O processo é realizado até que não ocorram mais rejeições ou que não haja mais classificadores na sequência. O objetivo da ordenação é utilizar classificadores mais complexos, que geralmente são mais caros computacionalmente, somente quando necessário.

Uma desvantagem apresentada pela estrutura em cascata é que classificadores subsequentes não conseguem corrigir erros cometidos pelos classificadores anteriores. Dessa forma, há a propagação dos equívocos até o classificador final (LU, 1996), (RANAWANA; PALADE, 2006).

A abordagem híbrida combina características das técnicas paralela e serial em busca de uma performance ótima. Além de oferecer a possibilidade do processamento independente e paralelo, permite a adoção de checagem de erros, de forma a evitar a propagação de equívocos entre as iterações. A Figura 2.10 esboça a ideia empregada na topologia híbrida.

2.3 Considerações Finais

Neste capítulo foi apresentada a estrutura de um sistema que emprega diversos classificadores (SMC) no processo de reconhecimento de novas instâncias na tentativa de superar as dificuldades de se empregar um classificador monolítico, desde a etapa de geração dos subconjuntos para treinamento dos classificadores, passando pelo processo de

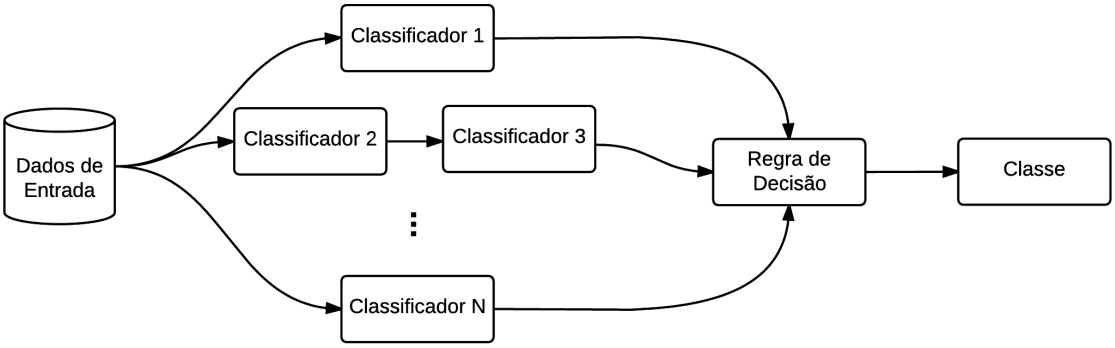

Figura 2.10: Topologia híbrida

seleção destes e a posterior combinação. Além do detalhamento de cada uma das fases, ao longo do capítulo foram apresentadas as estratégias mais comuns na literatura.

Além das etapas que compõe um SMC, outro fator que influencia o desempenho do processo classificatório é o conjunto de dados sobre o qual a análise é realizada. As pesquisas de (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (OKUN; VALENTINI, 2008) e (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) indicam que o desempenho dos classificadores NN e métodos de seleção dinâmica são influenciados pelas características de complexidade dos dados.

Haja vista tal dependência, Ho e Basu (HO; BASU, 2002), Macia *et al.* (MACIà et al., 2013) sugerem que uma boa abordagem para seleção de classificadores para um problema de classificação é compreender melhor a complexidade dos dados. Dessa forma, no capítulo seguinte são apresentadas várias medidas propostas na literatura para descrever o comportamento dos dados em termos de complexidade.

Capítulo 3

Análise da Complexidade

A complexidade de um problema é a medida do nível de dificuldade da classificação dados os atributos empregados para representá-la e ela pode ser observada através das características da base. Ou seja, a complexidade de um problema é medida com base nos dados que o compõe.

É sabido que o conjunto de dados tem influência direta no sucesso do processo de reconhecimento ou classificação, pois se a determinação e extração das características forem executadas inadequadamente, a acurácia da classificação será seriamente prejudicada (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010).

Visando estabelecer medidas para estimar o desempenho de classificadores sobre bases de dados, vários estudos foram realizados (LI; FANG, 1988), (HO; BASU, 2000), (HO; BASU, 2002), (HO; BASU; LAW, 2006), (SINGH, 2003), (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (OKUN; PRIISALU, 2009).

Os critérios levantados comumente são divididos em três categorias (HO; BASU, 2000) (HO; BASU, 2002) (SáNCHEZ; MOLLINEDA; SOTOCA, 2007): sobreposição (*overlap*) das classes (F1, F1v, F2, F3, F4), separabilidade das classes (L1, L2, N1, N2, N3) e medidas de geometria, topologia e densidade (L3, N4, T1, T2, D1, D2, D3, C1).

As medidas de sobreposição focam na efetividade de uma característica individual na identificação das classes e são discutidas na Seção 3.1. As medidas de separabilidade, foco da Seção 3.2, buscam estimar quão separáveis são as classes do problema examinando a existência e as formas das regiões de fronteira (SOTOCA; SáNCHEZ; MOLLINEDA, 2005) (CAVALCANTI; REN; VALE, 2012). Já as medidas de geometria, topologia e densidade, apresentadas na Seção 3.3, visam descrever a geometria ou a forma das variações abrangidas por cada classe visando oferecer compreensão mais superficial do relacionamento das classes (HO; BASU, 2000), (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006), (SOTOCA; MOLLINEDA; SáNCHEZ, 2006), (LUENGO; HERRERA, 2010),

(CAVALCANTI; REN; VALE, 2012).

3.1 Medidas de Sobreposição

Estas medidas têm o objetivo de mensurar quanto duas classes estão sobrepostas no espaço de características e para tanto, analisa-se o grau de justaposição dos atributos individualmente ou em conjunto.

3.1.1 Relação Máxima do Discriminante de Fischer (F1)

Esta medida exprime quão separáveis são duas classes de acordo com alguma característica específica (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (CAVALCANTI; REN; VALE, 2012).

Segundo Landeros (LANDEROS, 2008), F1 pode ser interpretado como a distância entre o centro de duas classes, de forma que, quanto maior o valor do índice, maior a separação entre as classes.

O cálculo consiste em comparar as médias e desvio-padrões das classes para cada atributo, de forma a mensurar seu nível de discrepância. A equação 3.1 apresenta como é realizado o cálculo para cada atributo especificamente. Os elementos μ_1, μ_2, σ_1 e σ_2 correspondem às médias e desvio-padrões das classes 1 e 2, respectivamente, para uma determinada característica do espaço. O valor adotado para F1 será o maior dentre todas as *features*, conforme denotado na Equação 3.2.

$$F1_i = \frac{(\mu_{1_i} - \mu_{2_i})^2}{{\sigma_{1_i}}^2 - {\sigma_{2_i}}^2} \tag{3.1}$$

$$F1^* = argmax(F1_i) \tag{3.2}$$

Entretanto, a equação 3.1 trata apenas de problemas de classificação com duas classes. Em cenários com C classes, emprega-se a equação 3.3.

$$F1 = \frac{\sum_{i=1}^{C} n_i * \delta(\mu, \mu_i)}{\sum_{i=1}^{C} \sum_{j=1}^{n_i} \delta(x_j^i, \mu_i)} \tag{3.3}$$

em que n_i denota o número de elementos da classe i, μ refere-se à média geral enquanto μ_i corresponde à média da classe i, δ é uma medida (como a distância euclidiana) e x_j^i corresponde ao elemento j da classe i.

A análise individual deste índice, segundo Landeros (LANDEROS, 2008), pode le-

var à interpretações precipitadas uma vez que ela não considera a forma das regiões de fronteira das classes. Conforme representado na Figura 3.1 a medida d_1 representa a distância entre o centro das duas classes. No primeiro cenário, à esquerda, as classes são linearmente separáveis, mostrando-se um problema simples. Entretanto, no segundo cenário, à direita, o mesmo intervalo separa o centro das duas classes, as quais tem certa sobreposição.

Ao comparar-se dois valores obtidos para F1, espera-se que o menor dentre eles indique maior sobreposição entre as classes, como ocorre nos cenários à esquerda das Figuras 3.1 e 3.2. Todavia, como ilustrado à direita da Figura 3.2, mesmo a medida d_2 sendo menor que d_1, as classes são linearmente separáveis. Isso mostra que mais de uma medida de complexidade deve ser empregada para caracterizar a região fronteiriça e a distribuição das classes.

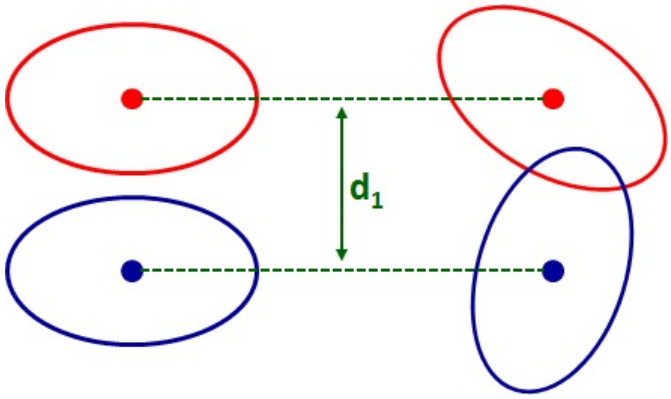

Figura 3.1: Classes com mesmo índice de discriminação (d_1) mas com relações distintas. Adaptado de (LANDEROS, 2008)

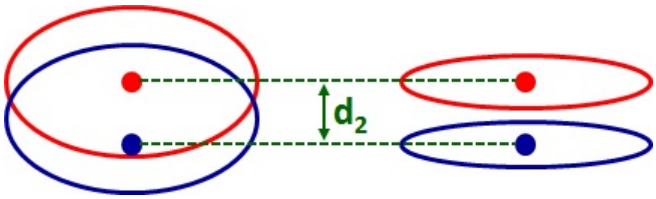

Figura 3.2: Mesmo índice de Fisher (d_2) porém com diferente relação entre as classes. Adaptado de (LANDEROS, 2008)

Dado que a proporção discriminante máxima de Fisher considera em seu cálculo as médias e desvios padrões, ela é fortemente indicada para casos em que os dados apresentem uma distribuição gaussiana. Todavia, para cenários onde tal fato não é verificado, como anéis concêntricos sem sobreposição, tal medida não consegue separar eficientemente as classes (HO; BASU; LAW, 2006; LANDEROS, 2008).

3.1.2 Sobreposição de Atributos por Classe (F2)

Segundo (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (CAVALCANTI; REN; VALE, 2012), F2 corresponde ao nível de sobreposição de uma única característica entre duas classes. Esta medida pode ser determinada encontrando-se, para cada característica, os valores máximos e mínimos para cada classe, e então calculando o comprimento da região de sobreposição normalizada pelo alcance dos valores cobertos pelas classes. Para determinar a sobreposição total de duas classes, deve-se calcular F2 para todas as *features* do conjunto e então multiplicá-las, como apresentado na Equação 3.4.

$$F2 = \prod_{i=1}^{d} \frac{MIN(max(f_i, c_1), max(f_i, c_2)) - MAX(min(f_i, c_1), min(f_i, c_2))}{MAX(max(f_i, c_1), max(f_i, c_2)) - MIN(min(f_i, c_1), min(f_i, c_2))} \tag{3.4}$$

em que i corresponde ao número da característica em análise, d indica o número de atributos, f_i refere-se à *feature* i enquanto c_i indica a classe.

A Figura 3.3 representa o cálculo realizado pela Equação 3.4. A região denotada como Min-Max corresponde à sobreposição do atributo número 1 para as classes A e B. Já a região Max-Min indica toda a extensão alcançada pelo atributo ao considerarmos os valores apresentados pelas classes em questão.

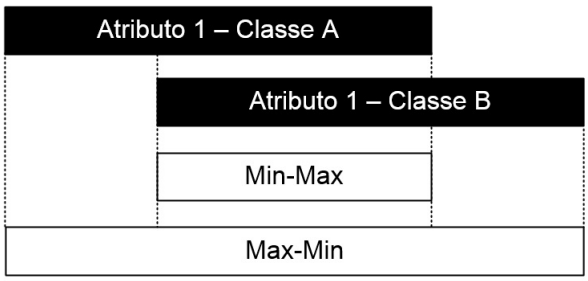

Figura 3.3: Ilustração da Equação 3.4 em que o numerador é representado por Min-Max enquanto o denominador por Max-Min

Conforme destacado por Ho & Basu (HO; BASU, 2002), basta que para um atributo não haja sobreposição para que o valor de F2 seja zero, dado o caráter do produtório. Outro fator importante é que, uma vez que o valor de sobreposição obtido é normalizado, quanto maior a dimensionalidade dos dados, menor será o valor de F2.

Em seu trabalho, Landeros (LANDEROS, 2008), propôs uma variação à equação 3.4. Visando desvincular o valor do índice ao número de atributos presentes nas classes, a autora sugeriu a escolha do menor valor de sobreposição entre os atributos das classes, o qual corresponde ao melhor cenário de *overlap* entre as classes. Se o melhor cenário

corresponde a um alto valor de justaposição, então tem-se um problema complexo. A autora destaca, no entanto, que tal abordagem é relativamente sensível à *outliers*, que podem inflar os valores dos atributos e criar uma representação inexata do fenômeno.

Uma terceira abordagem foi proposta por Lorena *et al.* (LORENA et al., 2012). Os autores propõe que ao invés de operar a multiplicação dos valores de sobreposição dos atributos seja realizada a soma dos valores normalizados. Dessa forma, evita-se que em bases com alta dimensionalidade, os valores do índice sejam muito pequenos.

Esta opção, no entanto, ainda é influenciada pelo número de atributos que compõe o conjunto dos dados pois, quanto mais atributos, maior tende a ser o índice, mesmo que trate-se de um problema simples.

A Equação 3.4 é empregada em cenários onde são analisados conjuntos compostos por duas classes. Para generalizar-se tal medida para problemas constituídos por n classes, pode se empregar a Equação 3.5 em que os resultados do produtório entre cada combinação de duas classes são somados.

$$F2 = \sum_{(c_i,c_j)} \prod_{i=1}^{d} \frac{MIN(max(f_i,c_1),max(f_i,c_2)) - MAX(min(f_i,c_1),min(f_i,c_2))}{MAX(max(f_i,c_1),max(f_i,c_2)) - MIN(min(f_i,c_1),min(f_i,c_2))} \quad (3.5)$$

3.1.2.1 Abordagens pela Média e Mediana

Em vista às desvantagens apresentadas pelas abordagens propostas em (HO; BASU, 2002), suscetível à não-sobreposição e ao número de atributos, em (LANDEROS, 2008), influenciada pelo número de atributos e em (LORENA et al., 2012) que sofre interferência de *outliers*, propõe-se uma variação das técnicas apresentas. Ao invés de adotar-se o produtório, soma simples ou mínimo das sobreposições dos *features*, uma proposta interessante é calcular a média das regiões de *overlaps* normalizadas, conforme apresentado na Equação 3.6, dividindo-se a soma total das sobreposições pelo número de atributos (d).

$$F2 = \frac{\sum_{i=1}^{d} \frac{MIN(max(f_i,c_1),max(f_i,c_2)) - MAX(min(f_i,c_1),min(f_i,c_2))}{MAX(max(f_i,c_1),max(f_i,c_2)) - MIN(min(f_i,c_1),min(f_i,c_2))}}{d} \quad (3.6)$$

A utilização da média contorna o problema da influência da dimensionalidade e também da possibilidade de não haver sobreposição de alguma característica entre as classes. No entanto, o índice ainda pode sofrer interferência de *outliers*, criando tendenciosidade na interpretação da medida. Uma solução para atenuar essa distorção seria empregar a mediana dos valores de sobreposição, como apresentado na Equação 3.7.

$$F2 = mediana \frac{MIN(max(f_i, c_1), max(f_i, c_2)) - MAX(min(f_i, c_1), min(f_i, c_2))}{MAX(max(f_i, c_1), max(f_i, c_2)) - MIN(min(f_i, c_1), min(f_i, c_2))} \quad (3.7)$$

3.1.3 Eficiência Máxima por Atributo Individual (F3)

Em problemas com alta dimensionalidade é interessante compreender como as informações discriminantes estão distribuídas entre os atributos. Neste contexto, F3 demonstra quanto cada característica contribui para a separação de duas classes. A eficiência de cada característica corresponde ao percentual de pontos que podem ser separados conforme aquela *feature* específica (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (CAVALCANTI; REN; VALE, 2012). Dessa forma, quanto maior o valor encontrado para F3, mais discriminante é o atributo. A eficiência máxima individual de um atributo será aquela que apresentar o maior valor entre todos os atributos considerados.

$$F3_i = separabilidade(f_i) \quad (3.8)$$

$$F3^* = argmax(F3_i) \quad (3.9)$$

Visando generalizar esta medida para problemas com mais de duas classes envolvidas, pode-se analisar a quantidade de instâncias que encontram-se em regiões de sobreposição. Tal processo pode ser realizado da seguinte forma: inicialmente todas as instâncias são setadas como não marcadas. Então, para cada uma das instâncias de cada uma das classes, avalia-se se ela possui algum atributo que tenha valor em uma região de cobertura de outra classe, o que pode fazer com que ela seja classificada incorretamente. Caso ela incorra em tal situação, é marcada. O valor da eficiência será a razão entre o número de instâncias marcadas e o total de instâncias (HO; BASU, 2002), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (CAVALCANTI; REN; VALE, 2012).

Esta medida, no entanto, não leva em conta a contribuição conjunta dos atributos (HO; BASU, 2002), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005).

3.1.4 Eficiência Coletiva dos Atributos (F4)

Segundo Orriols-Puig, Macià e Ho (ORRIOLS-PUIG; MACIà; HO, 2010), esta medida segue a mesma ideia que F3, entretanto, ela trata da capacidade discriminante de todos os atributos de forma conjunta.

O processo é realizado como segue: inicialmente é escolhido o atributo que consegue separar o maior número de elementos de uma classe, conforme a Equação 3.9. Esses elementos que puderam ser classificados perante a *feature* escolhida são então removidos do conjunto de dados e um novo atributo com o maior índice de segregação é escolhido e faz-se a separação dos elementos classificados. Este processo é realizado até que todas as instâncias possam ser rotuladas ou até que todos os atributos tenham sido analisados. Então, o índice corresponde à proporção de instâncias que puderam ser discriminadas.

A ideia desta abordagem é que, diferentemente de F3, sejam considerados todos os atributos no processo (ORRIOLS-PUIG; MACIà; HO, 2010).

3.2 Medidas de Separabilidade

Os descritores de separabilidade analisam a região fronteiriça entre duas classes, geralmente adotando estratégias de vizinhança das instâncias, visando descrever o quão complexo é o comportamento dos conjuntos neste setor.

3.2.1 Soma Minimizada da Distância de Erro de um Classificador Linear (L1)

Esta medida evidencia quanto os dados de treinamento são linearmente separáveis (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006), (LORENA et al., 2012), (CANO, 2013).

O processo consiste em inicialmente construir um classificador linear ótimo, de forma a minimizar os erros na separação das duas classes. Uma vez construído o classificador, L1 pode ser calculado pela soma das distâncias das amostras erroneamente classificadas até a fronteira linear construída pelo classificador (HO; BASU, 2000),(HO; BASU, 2002), (HERNáNDEZ-REYES; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2005), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006), (SOUTO et al., 2010), (CAVALCANTI; REN; VALE, 2012), (LORENA et al., 2012), conforme apresentado na Equação 3.10.

$$L1 = \sum \delta(C(x_i^-), x_i) \tag{3.10}$$

em que δ corresponde à distância euclidiana entre o classificador linear C que erra ao classificar a instância x_i ($C(x_i^-)$), conforme ilustrado na Figura 3.4 em que o classificador linear é denotado pela reta em vermelho, destacando duas instâncias classificadas de forma equivocada.

Um valor igual a zero indica que as classes são linearmente separáveis. Por outro lado, quanto maior o valor de L1, mais intrincada está a distribuição das amostras,

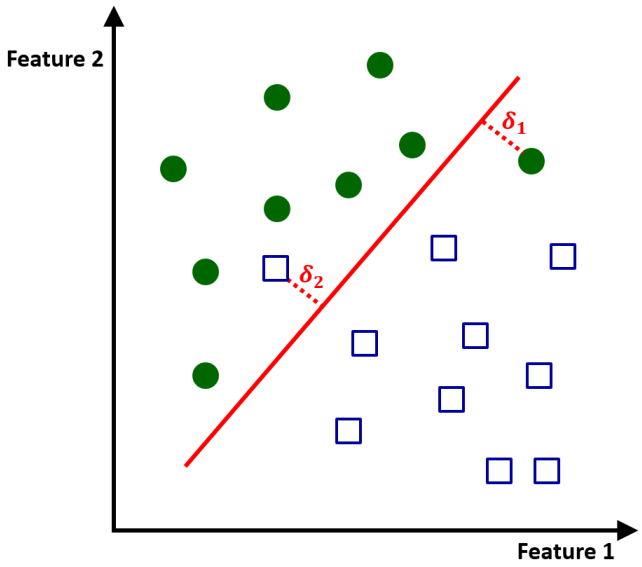

Figura 3.4: Classificador linear ótimo que erra ao classificar as duas instâncias em destaque

tornando a separação linear ineficiente.

3.2.2 Taxa de Erro de um Classificador Linear sobre o Treino (L2)

L2 exprime a taxa de erro obtida através da utilização de um classificador linear ótimo sobre os dados de treino (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006), (LORENA et al., 2012), (CANO, 2013). A ideia é empregar o mesmo classificador construído para L1 e avaliar quantas amostras estão posicionadas na região correspondente à classes diferentes da sua. O índice então é calculado dividindo-se o número de elementos interpretados incorretamente pelo número total de instâncias, como definido pela Equação 3.11.

$$L2 = \frac{contagem(C(x_i^-))}{n} \tag{3.11}$$

Caso o valor de L2 seja zero, as duas classes são linearmente separáveis. Quanto mais próximo de 1 for o valor obtido, pior é a separação linear entre as classes. No caso apresentado na Figura 3.4 há erro na classificação de duas instâncias, uma de cada classe.

3.2.3 A Fração de Pontos na Região de Fronteira (N1)

Esta medida baseia-se na construção de uma árvore de cobertura mínima (MST - *Minimal Spanning Tree*) que conecta todos os pontos do conjunto ao seu vizinho mais próximo, conforme apresentado na Figura 3.5. As ligações contínuas representam conexões

entre componentes da mesma classe, enquanto as ligações serrilhadas indicam vizinhança entre elementos de classes distintas. Os pontos que têm ligação com elementos de classes diferentes são considerados elementos de bordas das classes.

O valor de N1 é calculado pela relação entre o número de pontos conectados à instâncias de outras classes pelo total de elementos presentes no conjunto todo conforme definido pela Equação 3.12 (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006), (LILEIKYTE; TELKSNYS, 2011), (CAVALCANTI; REN; VALE, 2012). Dessa forma, quanto maior o valor do índice, mais intricada é a região de fronteira e, consequentemente, mais complexo é o problema.

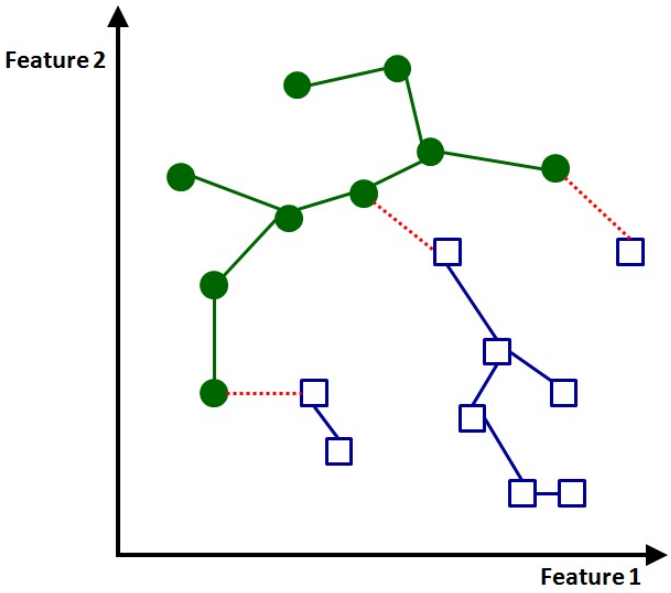

Figura 3.5: Árvore de cobertura mínima construída com base em duas classes

$$N1 = \frac{contagem(\overline{x_i \neq x_j})}{n} \tag{3.12}$$

em que $\overline{x_i \neq x_j}$ corresponde a elementos que estão conectados com instâncias de classes diferentes, enquanto n indica o número de elementos presentes no conjunto.

3.2.4 Proporção das Distâncias Intra/Inter Classes até o Vizinho Mais Próximo (N2)

A aplicação de N2 visa determinar o quanto duas classes são separáveis analisando-se a existência e a forma da fronteira entre as classes. O método consiste em comparar a distância média entre os elementos mais próximos dentro da classe com a distância dos vizinhos mais próximos fora da classe.

A ideia se caracteriza por calcular a distância euclidiana entre cada elemento do conjunto até o vizinho mais próximo dentro da mesma classe e também até o vizinho mais próximo fora da classe, conforme demonstrado na Figura 3.6. Então as distâncias entre os elementos da mesma classe são somados e divididos pela soma das distâncias entre as instâncias das classes diferentes (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HERNáNDEZ-REYES; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2005), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005), (HO; BASU; LAW, 2006), (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), (LUENGO; HERRERA, 2010), (LILEIKYTE; TELKSNYS, 2011), (CAVALCANTI; REN; VALE, 2012), conforme a Equação 3.13 em que $\delta(N_1^{=}(x_i), x_i)$ representa a distância entre a instância i e o vizinho da mesma classe que está mais próximo (representado pela linha contínua). Já $\delta(N_1^{\neq}(x_i), x_i)$ consiste na distância do elemento i até o elemento mais próximo pertencente à classe diferente da sua (destacado pela linha serrilhada).

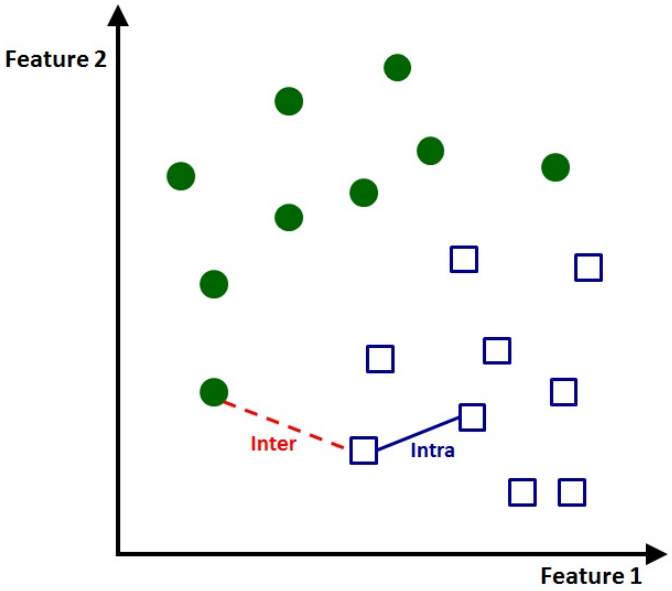

Figura 3.6: Representação da distância entre os vizinhos mais próximos intra e inter-classes

$$N2 = \frac{\sum_{i=1}^{n} \delta(N_1^{=}(x_i), x_i)}{\sum_{i=1}^{n} \delta(N_1^{\neq}(x_i), x_i)} \tag{3.13}$$

3.2.5 Taxa de Erro do Classificador KNN pela Abordagem Leave-One-Out (N3)

Esta medida consiste na taxa de erro da aplicação do classificador KNN (*K-Nearest-Neighbor*) com uma vizinhança de uma unidade sobre o próprio conjunto de treino (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HERNáNDEZ-REYES; CARRASCO-

OCHOA; MARTÍNEZ-TRINIDAD, 2005), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005), (HO; BASU; LAW, 2006), (LUENGO; HERRERA, 2010), (LILEIKYTE; TELKSNYS, 2011).

O percentual de erros, que é estimado pelo método *leave-one-out*, denota quão próximos os exemplos de diferentes classes são. Valores baixos para N3 indicam que há uma boa lacuna entre os elementos das bordas das classes, enquanto valores altos inferem na sobreposição das regiões fronteiriças.

3.3 Medidas de Geometria, Topologia e Densidade

3.3.1 Fração de Esferas de Cobertura Máxima (T1)

Seguindo a ideia de descrever a forma das classes proposta por Lebourgeois e Emptoz (LEBOURGEOIS; FRELICOT, 1998), T1 corresponde ao número de circunferências necessárias para cobrir cada uma das classes. Tais circunferências têm seu centro posicionado em cada uma das instâncias do conjunto de dados e são aumentadas até que seja tocado um elemento da outra classe (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HERNáNDEZ-REYES; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2005), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005), (HO; BASU; LAW, 2006), (LUENGO; HERRERA, 2010), (LILEIKYTE; TELKSNYS, 2011), (CAVALCANTI; REN; VALE, 2012). A Figura 3.7 ilustra a ideia da adoção das circunferências como delimitador das classes, onde os mapeamentos em cores distintas indicam o crescimento das duas classes.

Após todas as instâncias serem usadas como centro das representações periféricas crescentes, elimina-se aquelas que estão completamente abrangidas por um círculo maior. Assim, faz-se a contagem do número de esferas empregadas para cobrir cada classe. O valor de T1 então corresponderá a este valor dividido pelo total de instâncias presentes no conjunto.

Conforme Ho, Basu & Law (HO; BASU; LAW, 2006), o número e o tamanho das bolas indicam quanto os pontos tendem a agrupar-se em hiper-circunferências ou distribuir-se em estruturas menores e mais esparsas. Quando o conjunto apresenta pontos muito próximos entre as classes, o tamanho das esferas será menor e a quantidade destas empregadas para cobrir toda a classe será maior, aumentando assim o valor de T1, que indica regiões de sobreposição entre as classes.

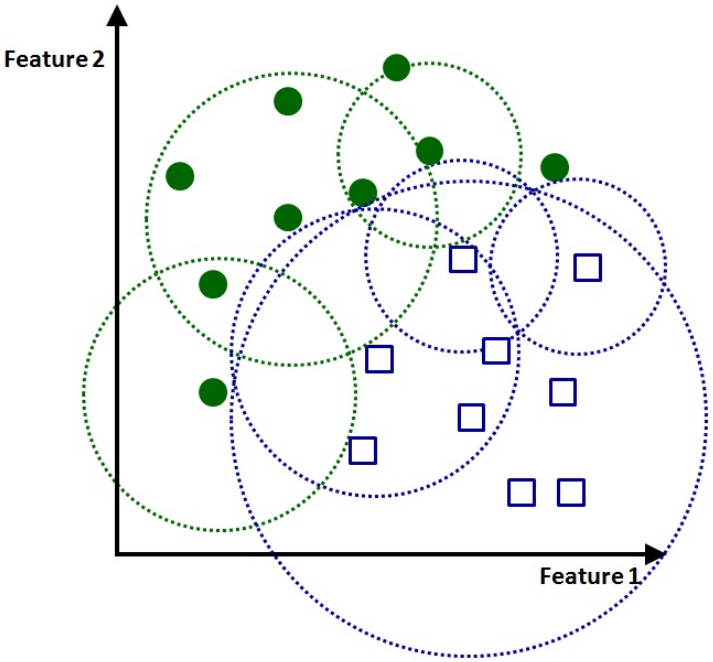

Figura 3.7: Representação da aderência por esferas para duas classes

3.3.2 Número Médio de Pontos por Dimensão (T2)

T2 descreve a densidade da distribuição espacial das amostras calculando a razão entre o número de elementos do conjunto de dados pelo número de atributos que formam a base (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HERNáNDEZ-REYES; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2005), (MOLLINEDA; SáNCHEZ; SOTOCA, 2005), (HO; BASU; LAW, 2006), (LUENGO; HERRERA, 2010), (CAVALCANTI; REN; VALE, 2012).

Esta medida, segundo Ho & Basu (HO; BASU, 2002) e Cavalcanti, Ren & Vale (CAVALCANTI; REN; VALE, 2012), é geralmente usada para investigar a influência da dimensionalidade de cada base de dados. Os autores apontam que T2 apresenta relevância ainda não tão clara quanto à separabilidade das classes com base em classificadores lineares, contudo, oferece informações pertinentes em cenários não lineares, como o de aplicação de um classificador KNN.

Uma variação desta medida foi proposta por Landeros (LANDEROS, 2008). Na abordagem proposta pela autora, o valor de T2 é obtido pela razão entre a raiz d-ésima de n (quantidade de elementos presentes na base) pelo número de atributos d, conforme a Equação 3.14.

$$D = \frac{\sqrt[d]{n}}{d} \tag{3.14}$$

3.3.3 Não-Linearidade de um Classificador Linear (L3)

Segundo Hoekstra & Duin (HOEKSTRA; DUIN, 1996), Ho & Basu (HO; BASU, 2002), Sotoca, Sánchez & Mollineda (SOTOCA; SáNCHEZ; MOLLINEDA, 2005) e Ho, Basu & Law (HO; BASU; LAW, 2006), dado um conjunto de treino, o método inicialmente cria um conjunto de teste através da interpolação linear entre pares randomicamente escolhidos dentro de uma mesma classe (do conjunto de treino) com coeficientes também randômicos. Então L3 corresponderá ao valor da taxa de erro dos dados de treino versus o conjunto de testes aplicando-se um classificador linear, tal como realizado em L1.

A Figura 3.8 ilustra o funcionamento do processo de geração do conjunto de teste. O conjunto de treino original é apresentado na Figura 3.8(a). A partir desde grupo, são então sorteados elementos dentro da mesma classe (representados pelos quadrados e esferas) e também o peso de cada elemento na formação da nova instância. Na Figura 3.8(b) os indivíduos sorteados para a formação do novo padrão são ligados por linhas e as marcações entre estes consistem na distribuição dos pesos que cada "pai"teria sobre a formação do novo elemento. Por fim, a terceira representação exibe o conjunto de teste formado, sobre o qual será calculado o valor do índice.

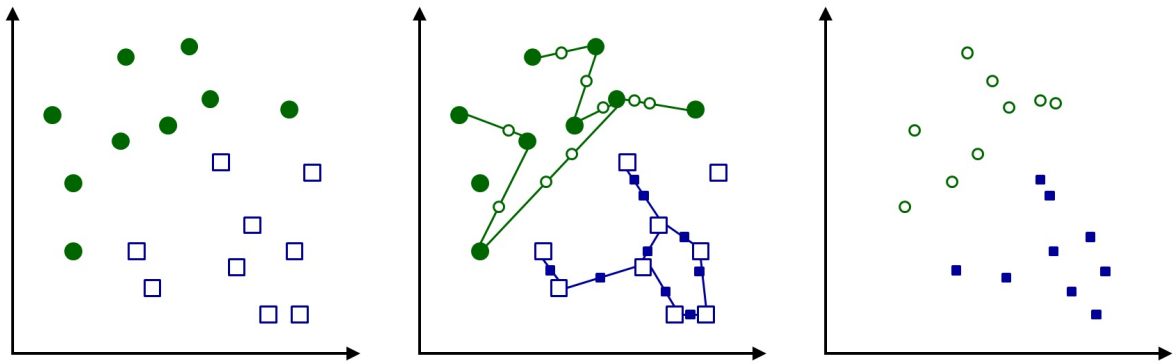

Figura 3.8: Processo de geração do conjunto de teste adotado em L3

3.3.4 Não-Linearidade de um Classificador KNN (N4)

A ideia desta medida segue o mesmo princípio de criação do conjunto de testes adotado por L3 (HO; BASU, 2002), (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), (HO; BASU; LAW, 2006). Todavia, no momento de calcular a taxa de erro sobre o conjunto de testes, ao invés de se adotar um classificador linear, é empregado o classificador KNN.

3.3.5 Densidade (D1)

A medida de densidade, segundo Sotoca, Sánchez & Mollineda (SOTOCA; SáNCHEZ; MOLLINEDA, 2005), pode ser definida como o número médio de amostras por unidade de volume onde os elementos estão distribuídos. O valor do volume é obtido pelo produto do alcance de todos os *features* de todas as classes.

Assim como ocorre com a medida F2, a densidade também é influenciada pela dimensionalidade dos dados quando estes estão normalizados.

3.3.6 Volume da Vizinhança Local (D2)

Esta medida representa o volume médio ocupado pelos k vizinhos mais próximos de cada instância de treino (SáNCHEZ; MOLLINEDA; SOTOCA, 2007). Uma vez que existe uma relação inversa entre volume e densidade, esta medida também pode ser vista como uma estimação local de densidade.

Considerando $N_k(x_i)$ como o conjunto de k vizinhos mais próximos de um dado exemplo x_i cuja classe é ω_i, então o cálculo do volume para x_i pode ser definido conforme a Equação 3.15.

$$V_i = \prod_{h=1}^{d} (max(f_h, N_k(x_i)) - min(f_h, N_k(x_i))) \tag{3.15}$$

em que $max(f_h, N_k(x_i)$ e $min(f_h, N_k(x_i)$ correspondem aos valores máximo e mínimo do atributo f_h dentro do conjunto formado pelos k vizinhos mais próximos da instância x_i.

Com base nos valores levantados para V, o valor de D2 pode ser expressado como a média dos valores de V_i referente às n instâncias de treino. A Equação 3.16 representa o cálculo realizado para determinar o valor do volume da vizinhança local.

$$D2 = \frac{1}{n} \sum_{i=1}^{n} V_i \tag{3.16}$$

3.3.7 Densidade da Classe na Região de Sobreposição (D3)

Uma vez que geralmente as regiões de sobreposição contêm os casos mais críticos para a tarefa de classificação e que estes incorrem na causa de grande parte dos erros de classificação, Sánchez, Mollineda & Sotoca (SáNCHEZ; MOLLINEDA; SOTOCA, 2007), propuseram a medida D3, que visa determinar a densidade relativa de cada classe dentro das regiões de sobreposição.

O processo consiste em, inicialmente, encontrar os k vizinhos mais próximos de

cada exemplo x_i. Então, se a maioria desses k vizinhos pertencem a uma classe diferente à de x_i, considera-se que o elemento faz parte de uma região de sobreposição. O valor de D3 para uma determinada classe ω é obtido pela relação do número de elementos na região de justaposição com o total de instâncias pertencentes à classe.

Nota-se que quanto menor o valor de D3 para uma determinada classe, menor será o número de exemplos daquela classe presentes na região de sobreposição.

3.3.8 Balanço da Classe (C1)

Visa determinar o balanceamento das classes no conjunto de dados estimando-se a entropia normalizada da distribuição dos tamanhos das classes (LORENA et al., 2012). O valor do índice é obtido pela Equação 3.17 em que $n_{k\omega}$ é o número de amostras da classe ω, c é o número de classes do conjunto e n refere-se ao número de instâncias presentes no conjunto.

$$C1 = -\frac{1}{log(C)} \sum_{k=1}^{C} \frac{n_{k\omega}}{n} log \frac{n_{k\omega}}{n} \tag{3.17}$$

A medida terá valor 0 se uma todas as amostras pertencerem à mesma classe e valor 1 se todas as classes possuírem o mesmo número de instâncias.

3.4 Considerações Finais

Este capítulo teve o objetivo de apresentar e detalhar o conceito de complexidade dos dados e como esta pode ser calculada sob diversas perspectivas, as quais são divididas em três categorias: aquelas que tentam representar matematicamente o grau de sobreposição de duas classes; as que visam descrever a dispersão da classes no espaço de características e aquelas que analisam quão separáveis duas classes são. O estudo de tais medidas fomentou o projeto do método que é descrito no capítulo seguinte, onde é apresentado o sistema de múltiplos classificadores proposto, o qual se baseia nos conceitos aqui apresentados para efetuar a geração e a seleção de classificadores para o processo de classificação.

Capítulo 4

Metodologia

Os sistemas de múltiplos classificadores foram uma alternativa encontrada para melhorar a performance de sistemas monolíticos (GUNES et al., 2003)(KUNCHEVA; WHITAKER, 2003)(KITTLER et al., 1998)(JAIN; DUIN; MAO, 2000). A ideia consiste em adotar diversos classificadores para atenuar a variância observada em sistemas individuais de classificação, o bias causado pelos dados de treinamento e também visando diminuir a dependência das particularidades de um sistema monolítico. A estratégia apresentada neste capítulo visa construir um conjunto homogêneo de classificadores com base na acurácia combinada com a exploração do espaço de complexidade dos dados usados no treino, de forma que o conjunto obtido possa superar as dificuldades em se adotar um classificador individual.

Além da geração do *pool* de classificadores, é adotada no processo uma estratégica dinâmica de seleção, baseada nas características da instância de teste. Para tanto, são usadas informações sobre a complexidade dos dados e acurácias sobre o conjunto de instâncias presentes na vizinhança do novo padrão. Interessante destacar aqui que a complexidade, ou dificuldade, não se restringe apenas ao número de instâncias, classes e número de características. Na verdade, ela engloba aspectos importantes inerentes ao problema de classificação que são estimados a partir de medidas calculadas sobre os dados envolvidos no problema. Entre vários aspectos, as medidas de complexidade geralmente tentam descrever e quantificar quão sobrepostas são duas classes, como se comportam as regiões de fronteira ou mesmo a distribuição espacial de cada classe. A Figura 4.1 apresenta uma visão geral do SMC proposto.

Na fase de geração, um *pool* contendo M classificadores é criado. Para este propósito, o conjunto de treino de um dado problema serve como entrada para um processo de amostragem que inicialmente gera M subconjuntos $(SC_1^1, SC_2^1, ..., SC_M^1)$. Cada SC_i corresponde a um indivíduo da população do algoritmo genético o qual é orientado pela

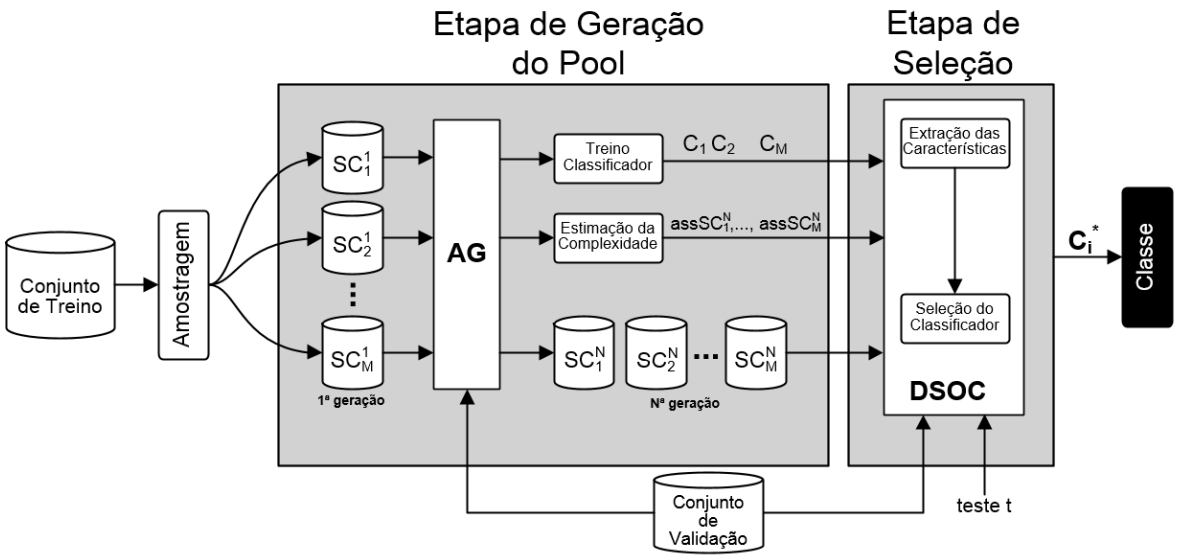

Figura 4.1: Estrutura macro do método desenvolvido, apresentando os processos de geração, seleção e classificação.

acurácia e informações de dificuldade do problema de classificação estimadas com base em algumas medidas de complexidade.

A ideia é obter subconjuntos de dados com diferentes níveis de dificuldade para treinar os classificadores que comporão o conjunto selecionado. A saída deste módulo consiste no *pool* de classificadores $(C_1, C_2, ..., C_M)$, nos subconjuntos usados para treinar cada um dos classificadores do conjunto $(SC_1^N, SC_2^N, ..., SC_M^N)$, e suas assinaturas de complexidade correspondentes $(ass_{SC_1^N}, ass_{SC_2^N}, ..., ass_{SC_M^N})$, em outras palavras, um conjunto de características que descreve quão difícil é cada subconjunto.

Em seguida, como a segunda fase do SMC, tem-se um novo esquema para a seleção dinâmica dos classificadores, no qual informações acerca da dificuldade do problema de classificação também são usadas. Dada uma instância de teste t, um vetor contendo três características ($c1_i$, $c2_i$ e $c3_i$) é estimado levando em conta a assinatura de complexidade $(assSC_i^N)$ do subconjunto SC_i usado para treinar o classificador C_i e a vizinhança de t no conjunto de validação. A similaridade entre a complexidade da vizinhança da instância de teste com a assinatura de complexidade do subconjunto de treinamento de cada classificador é combinada com informação de acurácia para estimar a competência de cada classificador.

Este capítulo busca descrever o funcionamento do método proposto, detalhando os passos envolvidos nas fases de treinamento e operacional, e como ocorre o relacionamento entre elas, desde a entrada dos dados do problema até a classificação do novo padrão. O detalhamento da primeira fase é apresentado na Seção 4.1 enquanto os pormenores do

processo de seleção são discutidos na Seção 4.2.

Visando ilustrar o funcionamento e avaliar o desempenho das soluções propostas, no capítulo seguinte são apresentados os experimentos realizados.

4.1 Geração de Classificadores

A etapa de geração tem como objetivo formar subconjuntos, aqui nomeados indivíduos, nos quais serão treinados os classificadores para a etapa de seleção. A ideia é que o conjunto construído seja acurado e diverso em termos de opinião permitindo que a etapa de selecão possa escolher aqueles classificadores mais adequados para maximizar a precisão do reconhecimento.

Nesta seção, a hipótese investigada é a de que a exploração da dispersão dos descritores de complexidade dos indivíduos no espaço de complexidade combinada com a acurácia de um classificador base treinado sobre tais indivíduos pode ser empregada na construção de um *pool* de classificadores diversos e acurados.

Para tal finalidade, um Algoritmo Genético (AG) foi proposto visando evoluir um conjunto inicial de classificadores de forma que o grupo resultante tenha uma maior cobertura do espaço de complexidade e que ao mesmo tempo, apresente acurácia. A escolha por tal estratégia baseia-se nos trabalhos (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010), (MACIà et al., 2013) onde a adoção de AG's permitiram com sucesso a exploração do espaço de complexidade. No entanto, neste trabalho o AG foi usado para gerar subconjuntos de dados a partir do conjunto de treino original. A função de fitness combina a diferença em termos de dificuldade entre os subconjuntos gerados e a acurácia dos classificadores correspondendentes treinados nestes subconjuntos. O indutor base é um parâmetro do método proposto.

A Figura 4.2 apresenta a estrutura do AG desenvolvido. Cada subconjunto consiste em um indivíduo dentro da população, como representado na ilustração. Os genes dos elementos correspondem às instâncias que compõe cada conjunto. O grupo de genes dos cromossomos correspondem às instâncias usadas para treinar cada classificador. O número de instâncias que compõe todos individuos são semelhantes e não variam ao longo da execução do método.

Para se formar o grupo inicial (primeira geração) é construído um *pool* composto de M classificadores. Este conjunto é gerado através da seleção aleatória e com reposição das instâncias, similar ao Bagging, conforme apresentado na Figura 4.1. O grupo formado pela técnica de amostragem servirá como entrada para o AG. A ideia então é evoluir este grupo inicial de forma que o desempenho final seja superior (em termos de exploração do

espaço de complexidade e de acurácia) àquele alcançado pela geração inicial. As etapas do AG são descritas no Algoritmo 1.

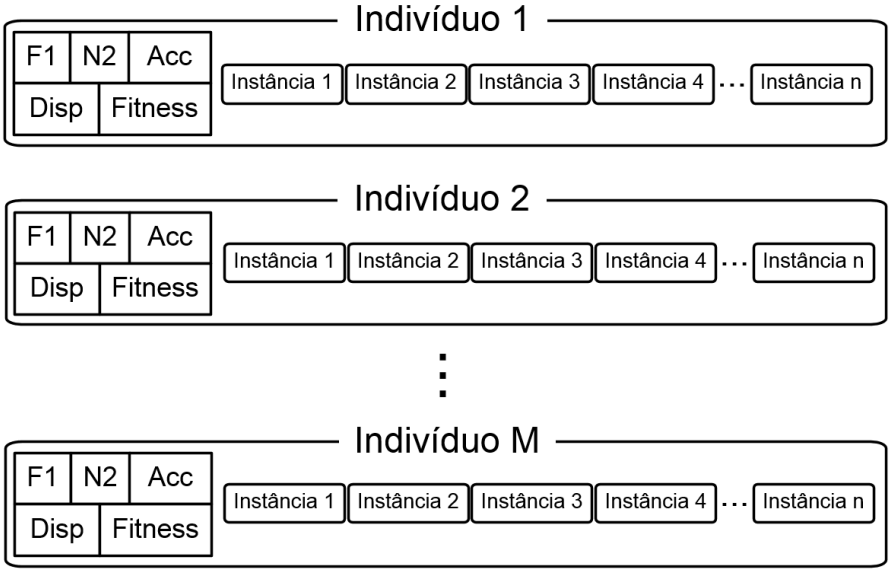

Figura 4.2: Estrutura adotada para o AG.

O processo de otimização consiste em maximizar uma função que combina as duas medidas a seguir:

- A dispersão do subconjunto onde o classificador é treinado, em relação aos demais, dentro do espaço de complexidade;
- A acurácia do classificador sobre o conjunto de validação.

O primeiro fator ("Disp" na representação gráfica) corresponde à distância média de cada subconjunto (definido nas linhas 7 e 35 do Algoritmo 1) até todos os outros elementos no espaço de complexidade. Para determinar tal valor foi necessário estimar a assinatura de complexidade (linhas 6 e 34 do algoritmo) dos subconjuntos envolvidos no processo. As medidas de complexidade provêm de cada uma das categorias descritas anteriormente: a) sobreposição das classes; b) separabilidade das classes; e c) geometria, topologia e densidade das classes. Com base nesta premissa, foram selecionados três índices: $F1$, $N2$ e $N4$. Para orientar a escolha, realizou-se um experimento com treze bases provenientes da UCI, no qual foram analisadas as correlações entre as 14 medidas de complexidade disponíveis na biblioteca DCoL (ORRIOLS-PUIG; MACIà; HO, 2010). Verificou-se que estas três medidas apresentaram baixa correlação entre si, indicando que elas podem explicar diferentes fenômenos. Entretanto, nos experimentos de geração de *pools*, observou-se que N4 não trazia contribuições e portanto foi descartada. Assim sendo,

na versão atual do método proposto a dificuldade de cada indivíduo é calculada baseada em F1 e N2.

A estimação do primeiro objetivo do AG é definida pela Equação 4.1 proposta em (CORRIVEAU et al., 2012). O processo consiste em calcular a média das distâncias euclidianas, no espaço de complexidade ($F1xN2$), entre o indivíduo i e todos os demais presentes na população. Quanto maior o valor obtido pela equação, mais afastado estará o elemento das regiões de concentração.

$$Disp_{C_i} = \frac{\sum_{j=1}^{M} \sqrt{\sum_{k=1}^{nc} (x_{i,k} - x_{j,k})^2}}{M - 1} \tag{4.1}$$

em que M corresponde ao número de classificadores presentes no *pool* enquanto nc refere-se ao número de medidas de complexidade adotadas no processo. Já $x_{i,k}$ e $x_{j,k}$ correspondem aos valores da k-ésima medida no espaço de complexidade para os indivíduos i e j, respectivamente.

Eu seu trabalho, Corriveau *et al.* (CORRIVEAU et al., 2012), analisaram o comportamento de 15 medidas distintas propostas por diversos pesquisadores, apontando vantagens e desvantagens de cada uma. Dentre as relacionadas, a escolha pela Equação 4.1 deu-se pelo fato de possibilitar analisar o comportamento de cada conjunto (aqui representado pelos indivíduos do AG) em relação aos demais em determinado espaço. Em nosso contexto, o espaço analisado é o da complexidade dos subconjuntos.

O segundo fator ("Acc", que é calculado na linha 13 do Algoritmo) consiste na acurácia de cada um dos classificadores da população sobre o conjunto de validação.

Dessa forma, classificadores que demonstraram maior acurácia sobre o conjunto de validação e cujos subconjuntos sobre os quais foram treinados mostraram-se mais distantes das áreas de concentração no espaço de complexidade apresentaram fitness maior em relação aos demais elementos. Portanto, o processo busca privilegiar a acurácia mas ao mesmo tempo tenta expandir a exploração do espaço de complexidade, como pode ser visto na Equação 4.2 (cálculo realizado na linha 14 do algoritmo).

$$Fit_{C_i} = Acc_{C_i} + Disp'_{C_i} \tag{4.2}$$

onde $Disp'_{C_i}$ corresponde à medida $Disp_{C_i}$ normalizada. Ela foi normalizada adotando-se a escala MinMax conforme denotado na Equação 4.3.

$$Disp'_{C_i} = \frac{Disp_{C_i} - Disp_{min}}{Disp_{max} - Disp_{min}} \tag{4.3}$$

Após determinada a aptidão de cada indivíduo é realizado o processo de elitismo

Algorithm 1: Geração do *pool* de classificadores com base em acurácia e informações de complexidade

Input: conjunto de treino Tr; conjunto de validação Va; número de classificadores M; tamanho dos bags Ba; número de gerações NumGe, tamanho do elitismo El; indutor base BI

Output: o *pool* final C de M classificadores; a assinatura de complexidade em todo o conjunto SC; os M bags finais SC

$SC = \{\}$;
for $i \leftarrow 1$ **to** M **do**
    Gerar bag SC_i^1 com tamanho Ba baseado em Tr;
    $SC = SC \cup SC_i^1$;
    $C_i^1 =$ Treinar BI com o bag SC_i^1;
    Calcular a assinatura de complexidade de SC_i^1;
end
for $g \leftarrow 1$ **to** $NumGe$ **do**
    $SC_{temp} = \{\}$;
    $E = \{\}$;
    for $i \leftarrow 1$ **to** M **do**
        Calcular a distância média $Disp_{SC_i^g}$;
        Estimar a acurácia de C_i^g sobre Va;
        Calcular o fitness $Fit_{C_i^g}$;
    end
    for $i \leftarrow 1$ **to** El **do**
        Select the i-th best individual $DS_i^g \ni E$;
        $E = E \cup DS_i^g$;
    end
    for $i \leftarrow 1$ **to** El **do**
        Selecionar o i-esimo melhor classificador $C_i^{g*} \ni E$;
        $E = E \cup SC_i^g$;
    end
    while *tamanho*(DS_{temp})<$(M - El)$ **do**
        Selecionar os pais SC_{p1} e SC_{p2};
        SC_{new1} = cruzamento de dois pontos de SC_{p1} e SC_{p2};
        SC_{new2} = cruzamento de dois pontos de SC_{p1} e SC_{p2};
        Aplica mutação em SC_{new1} e SC_{new2};
        $SC_{temp} = SC_{temp} \cup SC_{new1} \cup SC_{new2}$;
    end
    for *cada bag* $SC_i^g \in SC_{temp}$ **do**
        Remove os genes duplicados;
    end
    $SC = SC_{temp} \cup E$;
    for $i \leftarrow 1$ **to** M **do**
        $C_i^{g+1} =$ Treinar BI com o bag SC_i^{g+1};
        Calcular a assinatura de complexidade de SC_i^{g+1};
    end
end

(nas linhas 17-19 do algoritmo), previnindo que os melhores elementos sofram alterações durante as etapas de cruzamento e mutação. Tais elementos são conduzidos diretamente à geração seguinte.

Na sequência é então realizada a etapa de cruzamento. Para tanto, o processo de seleção (efetuado na linha 22) dos indivíduos é realizado de forma aleatória (roleta) de forma que, quanto maior o fitness do elemento, maior a chance deste ser selecionado para propagar seus genes aos novos indivíduos.

Nesta etapa adotou-se o cruzamento de dois pontos (linhas 23 e 24 do algoritmo), os quais são determinados aleatoriamente. Proposta por Macia *et al.* (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010), a ideia é que as instâncias dos conjuntos sejam trocadas entre os classificadores, mantendo, no entanto, a estratificação entre as classes. De forma a garantir que não haja mudança nas proporções, as instâncias são organizadas por classe, conforme ilustrado pelos números "1", "2" e "3" nas Figuras 4.3(a) e 4.3(b). Assim, quando se dá a troca de "segmentos" entre os dois pais, não haverá mudança no número de instâncias pertencentes à cada classe. O processo de cruzamento é ilustrado nas Figuras 4.3(a) e 4.3(b) em que os indivíduos i e j foram selecionados para propagar seus genes à geração seguinte.

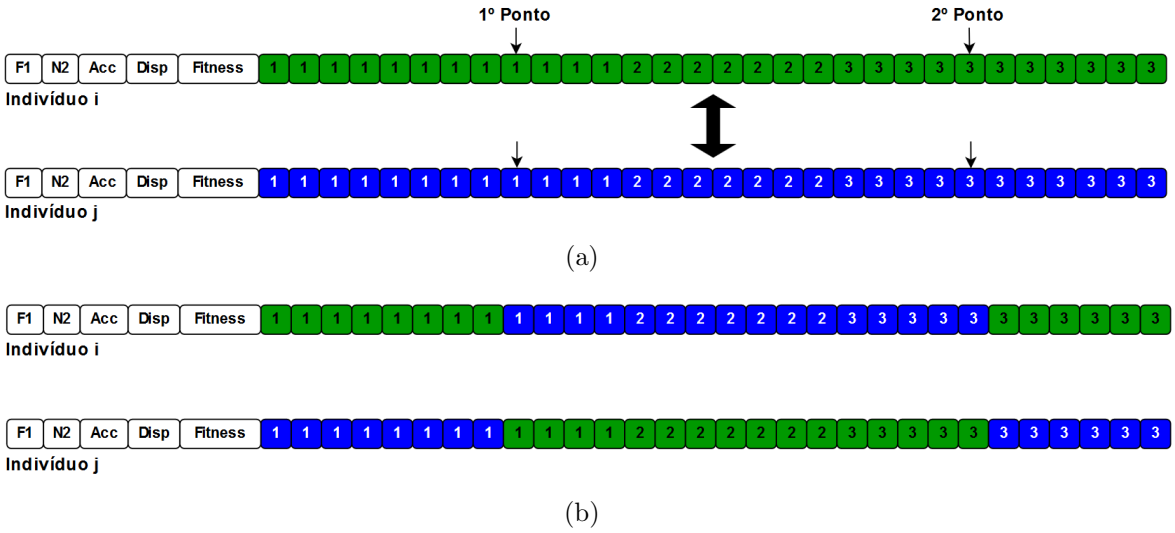

Figura 4.3: Funcionamento do processo de cruzamento implementado: a) Seleção dos dois pontos de cruzamento, posicionados necessariamente em classes distintas; b) Segmentos trocados entre os indivíduos i e j

Na abordagem proposta, apesar do caráter randômico da seleção dos dois pontos de cruzamento, restringiu-se que cada ponto estivesse posicionado em uma classe diferente. Dessa forma, o processo é mais agressivo, garantindo que elementos de diferentes classes sejam trocados. Esta decisão de projeto, definida durante a etapa de testes, foi a que se

mostrou mais adequada aos objetivos buscados.

Assim que os novos elementos são gerados pelo cruzamento, eles são submetidos à etapa de mutação (executado na linha 25). Neste estágio cada gene é submetido a uma probabilidade e, caso satisfeita, ele é mutado. No nosso cenário, cada instância corresponde a um gene, então quanto esta passa pela mutação, ela é substituída por outra da mesma classe, garantindo assim a manutenção do balanceamento. Este processo é representado na Figura 4.4. Na ilustração, a instância x do indivíduo i é substituída pelo gene y do elemento j, ambos pertencentes à classe 1.

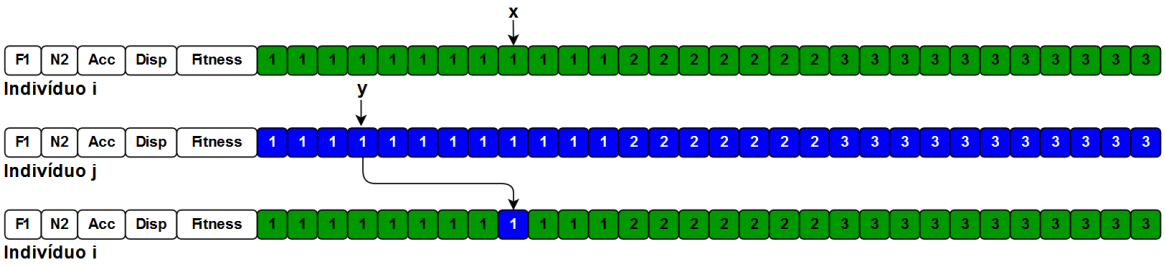

Figura 4.4: Processo de Mutação: a instância selecionada é trocada por outra randomicamente escolhida em um indivíduo diferente, necessariamente pertencente à mesma classe.

No momento da troca, a escolha pelo novo gene é feita randomicamente. Inicialmente um classificador diferente é selecionado e, dentro deste, uma instância aleatória que pertença à mesma classe do gene é escolhida. O processo é repetido para todos os genes de todos os elementos gerados pelo cruzamento.

A última fase consiste em remover as instâncias duplicadas (na linha 29 do algoritmo). O processo, executado sobre as repetições, é similar ao realizado pela mutação e é repetido até que não haja mais duplicatas no conjunto.

Este problema de otimização é submetido à quatro restrições obedecidas ao longo da execução:

- Número fixo de instâncias: os indivíduos de uma mesma população devem ter o mesmo número de instâncias, determinado pelo projetista. Este valor corresponde a um percentual do total de instâncias do conjunto de treino;

- Proporcionalidade das classes: cada indivíduo possui um conjunto de instâncias proporcional àquele observado no conjunto de treino e caso haja desbalanceamento entre as classes, ele será mantido;

- Valores das medidas de complexidade: o conjunto de instâncias que forma cada indivíduo deve possibilitar o cálculo das medidas de complexidade. Para cumprir

tal restrição garante-se que todas as classes do conjunto original estejam presentes no elemento;

- Duplicidade: a presença de repetições pode implicar em bias sobre a estimação e, além disso, influenciar no cálculo das medidas de complexidade, como em N2 que baseia-se na distância entre vizinhos dentro e fora da classe. Assim sendo, é interessante que instâncias duplicadas sejam evitadas.

4.2 Seleção de Classificadores

O sucesso de um método de seleção dinâmica depende da adoção de um bom critério para medir a competência dos classificadores em reconhecer os padrões de teste a serem classificados. Em seu trabalho, Britto Jr *et al.* (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) apresentam uma taxonomia para caracterizar os métodos de acordo com o critério usado para mensurar a aptidão dos classificadores. Segundo os autores, os métodos podem ser separados em dois grandes grupos: aqueles baseados em competência individual e aqueles que levam em conta o relacionamento entre os classificadores que compõe o *pool*. Apesar do grande número de estratégias diferentes e aspectos usados para medir a competência dos classificadores do conjunto, é possível observar o uso frequente da avaliação baseada em acurácia, que geralmente é combinada com alguma outra informação adicional.

Nesta seção buscou-se estimar a contribuição de características relacionadas ao nível de dificuldade de problemas de classificação obtidas pela análise de complexidade dos dados durante o processo de definição da competência dos classificadores de acordo com a instância de teste em análise. Tal abordagem para seleção dinâmica baseada em complexidade recebeu a alcunha de *Dynamic Selection Over Complexity* - DSOC.

A estratégia aqui apresentada é inspirada em trabalhos que visam encontrar os indutores mais promissores para um problema específico de classificação, levando em conta sua dificuldade (HO; BASU, 2002). Entretanto, a ideia foi investigar se o nível de dificuldade estimada da vizinhança do padrão de teste sobre o conjunto de validação pode contribuir para estimar a competência dos classificadores do *pool* de um SMC.

Assim sendo, a premissa é a de que a complexidade da vizinhança da instância de teste, obtida em um conjunto de validação, quando combinada com informação de acurácia possa ser usada para estimar a competência dos classificadores. Para tal propósito, a aptidão dos indutores é estimada considerando sua acurácia em uma região local do espaço de características a partir da o qual é calculada também seu nível de dificuldade. As etapas de geração e operacional são apresentadas nas Figuras 4.5 e 4.6 enquanto os passos do

método são descritos no Algoritmo 2. Uma vez que a fase de treinamento faz parte da primeira etapa do SMC, optou-se por omitir aqui alguns detalhes do processo que fomentam a fase de seleção.

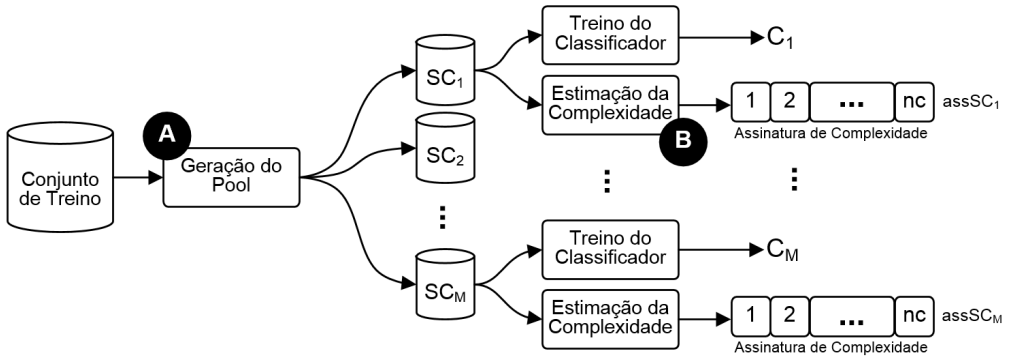

Figura 4.5: Detalhamento de parte da etapa de treinamento - Fluxo de informações que serão adotadas na etapa operacional

Na fase de treinamento (Figura 4.5), um conjunto de treino de um dado problema de classificação é empregado para gerar um *pool* de inicial composto de M classificadores, usando no processo uma técnica de formação de *ensembles* (Bagging, Boosting ou mesmo a técnica evolutiva descrita na seção anterior) para prover diversidade e acurácia (Figura 4.5 - A). Estes subconjuntos $(SC_1, SC_2, ..., SC_M)$ são utilizados para treinar o *pool* de M classificadores $(C_1, C_2, ..., C_M)$, como destacado na Figura 4.5 - B.

Em seguida, para cada subconjunto de dados gerado, um vetor composto de nc medidas de complexidade é processado (Figura 4.5 - B, nas linhas 1-3 do Algoritmo 2). Esse conjunto de características (ass_{SC_i}) é usado como uma assinatura nc-complexa para cada subconjunto (SC_i) formado.

Para compor a assinatura de complexidade adotou-se um descritor de cada uma das três categorias descritas anteriormente (sobreposição, geometria e densidade e separabilidade). Assim como no processo de geração, as medidas escolhidas foram F1, N2 e N4. Entretanto, no processo de seleção manteve-se N4, uma vez que verificou-se que esta contribuía para a abordagem.

Após a realização da fase de treinamento são levados à etapa de seleção cada um dos M subconjuntos $(SC_1, SC_2, ..., SC_M)$, o *pool* de M classificadores $(C_1, C_2, ..., C_M)$ treinados nestes subconjuntos e também as assinaturas de complexidade $(ass_{SC_1}, ..., ass_{SC_M})$ de cada um dos subconjuntos formados, conforme destacado na Figura 4.6-1. Estes elementos serão empregados no momento de definir a competência dos classificadores para decidir quem deve atribuir o rótulo à nova instância.

Durante a etapa operacional (Figura 4.6), a seleção dinâmica é realizada pela estimação da competência de cada classificador baseada em três características (Figura

4.6 - C). Para tanto, são empregados no processo elementos provenientes da etapa de treinamento, conforme destacado na Figura 4.6 - 1.

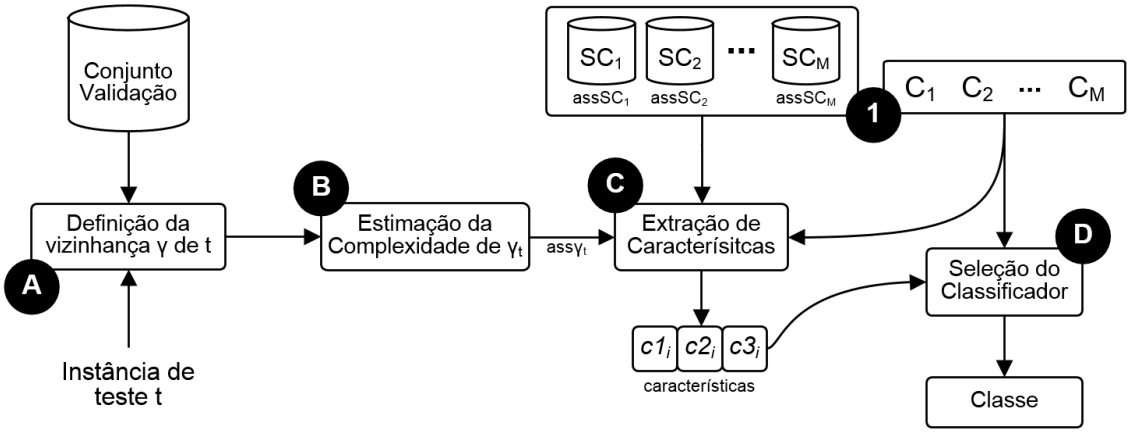

Figura 4.6: Ilustração da etapa operacional do SMC - levantamento das características e estimação da competência dos classificadores

De forma a descrever as características empregadas no processo de seleção, considere SC_i como o subconjunto usado para treinar o classificador C_i enquanto ass_{SC_i} é a assinatura nc-complexa (valores de F1, N2 e N4) calculados com base em SC_i. Além disso, γ_t como a k-vizinhança da instância de teste t, enquanto ass_{γ_t} consiste na assinatura nc-complexa calculada sobre γ_t.

A primeira característica é a similaridade em termos de complexidade ($c1_i$). Dada uma instância de teste t, o primeiro passo é definir a vizinhança γ_t no conjunto de validação (Figura 4.6 - A, realizado na linha 5). Em seguida, a assinatura ass_{γ_t} é calculada com base em γ_t (Figura 4.6 - B, linha 6 do Algoritmo 2). A similaridade entre a assinatura de complexidade ass_{γ_t} com a assinatura ass_{SC_i} de cada subconjunto usado para treinar os classificadores é calculada pela distância euclidiana conforme denotado na Equação 4.4. Dessa forma, é possível determinar qual o classificador treinado no subconjunto com complexidade mais similar àquela observada na vizinhança da instância de teste.

$$c1_i = \delta(sig_{\gamma_t}, sig_{SC_i}) \tag{4.4}$$

A segunda informação ($c2_i$) é a distância até a classe predita. Considere y_i como sendo a classe predita pelo classificador C_i para a instância de teste t, SC_i como o subconjunto usado para treinar C_i, e α_{ij} como o centróide da classe predita y_i no subconjunto de treinamento SC_i. Calcula-se a distância da instância de teste t até o centróide α_{ij} como demonstrado na Equação 4.5. A ideia é descrever melhor o espaço de complexidade, uma vez que as medidas de complexidade podem apresentar valores similares para representar

a dificuldade entre duas classes mesmo quando elas estão distribuídas de forma diferente no espaço de características.

$$c2_i = \delta(t, \alpha_{ij}) \tag{4.5}$$

A acurácia local da classe é a terceira característica levantada ($c3_i$). Basicamente consiste na acurácia local de cada classificador C_i considerando a classe predita y_i para a instância de teste t. Esta acurácia local é estimada na vizinhança γ_t como denotado na Equação 4.6.

$$c3_i = Acuracia(C_i, y_i, \gamma_t) \tag{4.6}$$

As características são calculadas para cada classificador (nas linhas 7-11 do Algoritmo). O valor final de competência do classificador C_i é obtido pela combinação das três informações levantadas. Foram estimados a combinação dos fatores usando-se a soma e mutiplicação. Ambas estratégias mostraram resultados similares. A combinação final usando a soma dos fatores é apresentada na Equação 4.7.

$$Comp_{\text{-}C_i} = (1 - c1'_i) + (1 - c2'_i) + c3_i \tag{4.7}$$

Em que $c1'_i$ e $c2'_i$ correspondem aos valores normalizados para as medidas $c1_i$ e $c2_i$, respectivamente. A normalização foi realizada através da função MinMax como apresentado na Equação 4.8 para a característica $c1_i$.

$$c1_i^{'} = \frac{c1_i - c1_{i_{min}}}{c1_{i_{max}} - c1_{i_{min}}} \tag{4.8}$$

O classificador selecionado (linha 12 do Algoritmo) será aquele que apresentar a maior competência $Comp_{\text{-}C_i}$.

4.3 Considerações Finais

Neste capítulo foi apresentado o sistema de múltiplos classificadores proposto, o qual baseia-se em medidas de complexidade e acurácia para efetuar o processo de geração dos classificadores, bem como determinar a competência de cada um no processo de classificação de uma nova instância. Detalhou-se a estrutura do sistema como um todo e também o funcionamento das etapas envolvidas no processo. Os experimentos realizados para validar as fases desenvolvidas, bem como do SMC como um todo são apresentados no capítulo seguinte. Inicialmente as etapas de geração e seleção são tratadas separadamente

Algorithm 2: DSOC - Seleção Dinâmica baseada em Complexidade

Input: o conjunto C composto de M classificadores; os conjuntos de treino, validação e teste, Tr, Va e Te; e o tamanho da vizinhança, K

Output: C^*, o classificador mais promissor para cada instância de teste t presente no conjunto Te

```
for cada instância de teste t_i ∈ Te do
    Definir γ_t como os K vizinhos mais próximos de t_i em Va;
    Estimar a assinatura de complexidade de γ_t;
    for cada classificador C_i ∈ C do
        Calcular a similaridade entre a vizinhança do teste γ_t com o
         subconjunto SC_i (c1_i);
        Calcular a distância entre os centróides de SC_i e γ_t (c2_i);
        Estimar a acurácia de C_i para a classe predita na vizinhança γ_t (c3_i);
        Normalizar c1_i and c2_i;
        Comp_{-C_i} = (1 − c1'_i) + (1 − c2'_i) + c3_i;
    end
    C* = argmax(Comp_{-C_i});
    Usar o classificador C* para classficar t_i;
end
```

para, na sequencia, serem tratadas como um sistema completo.

Capítulo 5

Resultados Experimentais

Nesta seção são apresentados os experimentos realizados com o intuito de validar as estratégias propostas para geração (Seção 5.1) e seleção dos classificadores (Seção 5.2) descritas neste trabalho. Inicialmente as abordagens são apresentadas e discutidas de forma independente para, em seguida, serem tratadas de forma combinada (Seção 5.3), envolvendo todo o processo de um sistema de múltiplos classificadores, desde a formação do *pool* de classificadores até a atribuição dos rótulos às instâncias de teste.

Visando a avaliar o desempenho das abordagens propostas (geração e seleção, bem como a combinação de ambas), foi adotado um conjunto composto de trinta bases distintas, das quais dezesseis são provenientes do repositório da UCI (BACHE; LICHMAN, 2013), quatro procedem do repositório KEEL (Knowledge Extraction based on Evolutionary Learning) (ALCALá-FDEZ et al., 2011), quatro de propriedade da LKC (Ludmila Kuncheva Collection of Real Medical Data) (KUNCHEVA, 2004), quatro oriundas do projeto STATLOG (KING; FENG; SUTHERLAND, 1995) e duas bases artificiais geradas com o *toolbox* PRTools do Matlab.

Todos os problemas apresentam apenas atributos numéricos sem valores faltantes. Além disso, eles têm sido frequentemente utilizados na literatura para mensurar o desempenho de métodos de seleção dinâmica. A Tabela 5.1 apresenta os detalhes de cada um dos trinta conjuntos. São detalhados o número de instâncias de cada base, tamanho dos conjuntos de treino, teste e validação, bem como a quantidade de atributos e classes presentes. Por fim, relata também o repositório fonte de cada conjunto.

Os ensaios foram conduzidos adotando-se 20 repetições. Para cada uma, as bases de dados foram randomicamente divididas em distribuições de 50% para o conjunto de treino, 25% para validação e 25% para o grupo de teste, mantendo, entretanto, a proporção inicial das classes.

Três conjuntos de experimentos foram conduzidos. No primeiro avaliou-se o método

Tabela 5.1: Principais características das bases usadas nos experimentos

Base	Instâncias	Treino	Teste	Validação	Atributos	Classes	Fonte
Adult	690	345	172	173	14	2	UCI
Banana	2000	1000	500	500	2	2	PRTools
Blood	748	374	187	187	4	2	UCI
CTG	2126	1063	531	532	21	3	UCI
Diabetes	766	383	192	191	8	2	UCI
Ecoli	336	168	84	84	7	8	UCI
Faults	1941	971	485	485	27	7	UCI
German	1000	500	250	250	24	2	STATLOG
Glass	214	107	53	54	9	6	UCI
Haberman	306	153	76	77	3	2	UCI
Heart	270	135	67	68	13	2	STATLOG
ILPD	583	292	145	146	10	2	UCI
Segmentation	2310	1155	577	578	19	7	UCI
Ionosphere	350	176	87	87	34	2	UCI
Laryngeal1	213	107	53	53	16	2	LKC
Laryngeal3	353	177	88	88	16	3	LKC
Lithuanian	2000	1000	500	500	2	2	PRTools
Liver	345	173	86	86	6	2	UCI
Magic	19020	9510	4755	4755	10	2	KEEL
Mammo	830	415	207	208	5	2	KEEL
Monk	432	216	108	108	6	2	KEEL
Phoneme	5404	2702	1351	1351	5	2	ELENA
Sonar	208	104	52	52	60	2	UCI
Thyroid	692	346	173	173	16	2	LKC
Vehicle	847	423	212	212	18	4	STATLOG
Vertebral	300	150	75	75	6	2	UCI
WBC	569	285	142	142	30	2	UCI
WDVG	5000	2500	1250	1250	21	3	UCI
Weaning	302	151	75	76	17	2	LKC
Wine	178	89	44	45	13	3	UCI

de geração de comitês. O *pool* gerado para cada problema é comparado àqueles gerados de forma randômica e com reposição (similar ao Bagging). Além disso, foram comparadas a acurácia de seis técnicas de seleção dinâmica já estabelecidas na literatura. Cada método foi testado utilizando-se os *pools* gerados pelo Bagging e empregando-se aqueles construídos pelo método proposto que baseia-se no AG.

No segundo conjunto de experimentos, avaliou-se o desempenho do método de seleção proposto (DSOC). Para tanto, testou-se o desempenho da estratégia perante as mesmas seis abordagens de seleção dinâmicas adotadas no primeiro conjunto de experimentos. No último grupo de experimentos foi avaliado o SMC como um todo, considerando a estrégia de geração de *pools* e de seleção propostas, ambas empregando critérios de complexidade dos problemas. A ideia foi observar quando o método proposto para a seleção, em conjunto com a geração dos *pools*, poderia trazer um ganho adicional em termos de acurácia do problema de classificação.

5.1 Experimento 1 - Geração dos Classificadores usando Complexidade

Esta seção apresenta os experimentos realizados visando avaliar o método de geração proposto. Para cada problema, um *pool* composto de 100 perceptrons foi criado através de processo de seleção randômico e com reposição. O perceptron foi adotado como classificador base por tratar-se de um indutor fraco e instável.

Tais subconjuntos, contendo 50% do tamanho do conjunto de treino, foram utilizados para formar a população inicial empregada no AG. Assim sendo, cada população do AG foi composta de 100 indivíduos, os quais foram evoluídos durante 30 gerações. A quantidade de épocas foi definida com base no trabalho de Macia *et al.* (MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010) no qual a quantidade de trinta gerações mostrou-se suficiente para a evolução do método. Além disso, a adoção de um número maior de épocas poderia levar a *overfitting*, uma vez que um dos critérios do AG é maximizar a acurácia dos indivíduos sobre o conjunto de validação.

Para o processo de cruzamento foi empregada uma taxa de 80%. Assim a evolução do algoritmo não será tão lenta e o processo de substituição dos membros antigos por novos indivíduos não será tão agressivo, atenuando a chance de que elementos com alta aptidão possam ser perdidos.

A taxa de mutação empregada foi de 5%. Tal valor pode evitar que o processo entre em estagnação ou mesmo pode fazer com que novas regiões do espaço de soluções sejam exploradas. Valores maiores poderiam tornar o processo aleatório, levando, inclusive, à perda de bons indivíduos.

Para garantir a propagação dos melhores membros de cada população à geração seguinte, empregou-se o elitismo com um total de 4 elementos por época. Tal valor foi definido empiricamente, uma vez que possibilitou a manutenção de elementos de alta aptidão sem levar a uma convergência precoce do método.

Como descrito anteriormente (Seção 4.1), no processo evolutivo do AG, foram utilizadas duas medidas de complexidade: $F1$ e $N2$.

Visando a avaliar a contribuição da utilização do AG proposto na geração de subconjuntos, neste experimento foram comparadas as acurácias de seis métodos de seleção dinâmica já estabelecidos na literatura. Cada método foi testado utilizando os conjuntos gerados de forma aleatória e os conjuntos obtidos ao término da execução do AG para o treinamento dos classificadores.

Foram implementadas estratégias baseadas em seleção individual de classificadores (LCA (WOODS; KEGELMEYER JR.; BOWYER, 1997), OLA (WOODS; KEGELMEYER JR.;

BOWYER, 1997), a Priori e a Posteriori (GIACINTO; ROLI, 1999) e (DIDACI et al., 2005)) e também abordagens baseadas na seleção de conjuntos de classificadores (KNORA-E e KNORA-U (KO; SABOURIN; BRITTO JR., 2008)). Para todas as soluções empregou-se uma vizinhança de dimensão 7. Tal valor provou-se o mais apropriado em estudos anteriores (KO; SABOURIN; BRITTO JR., 2008; CRUZ et al., 2015). Além dos métodos dinâmicos, implementou-se também a combinação de todos os classificadores e o *single best*.

No processo de combinação foi usado o voto majoritário para combinar os classificadores treinados no *pool* formado aleatoriamente, enquanto para aqueles construídos pelo AG foi empregado o voto ponderado combinado com uma função sigmóide, aplicada com o intuito de suavizar os pesos dos votos, conforme definida pela Equação 5.1, para estimar os pesos de cada classificador baseado em seus fitness.

$$f(x, a, c) = \frac{1}{1 + e^{-a(x-c)}} \tag{5.1}$$

em que x corresponde ao fitness de cada elemento, a se refere à inclinação da curva enquanto c é o ponto de flexão da curva.

O desempenho médio das duas estratégias de geração para cada uma das bases é apresentado nas Tabelas 5.2 e 5.3. A primeira detalha os valores obtidos pelos métodos de seleção dinâmica de classificador individual enquanto a segunda apresenta os resultados dos métodos de seleção de *ensembles*, *single best* e da combinação de todos os classificadores. Os valores em negrito representam a maior acurácia de cada método de seleção para cada um dos 30 problemas.

Observou-se que em 126 de 180 casos verificados (70.00%), adotar o AG no processo de geração trouxe aumento na acurácia do método de seleção dinâmica. Por outro lado, em 22.22% dos cenários (40 casos) foi mais adequado empregar apenas o sorteio com reposição para a geração dos subconjuntos. Em catorze ocasiões (7.78%) as abordagens exibiram comportamento similar. Analisando-se a combinação de todos os classificadores do *pool*, o AG se sobressaiu em 63.33% dos casos. Já adotar a seleção randômica na geração alcançou maior acurácia em 30.00% dos problemas. Além disso, houve dois casos em que ocorreu empate entre as soluções (6.67%). Para o *single best* observou-se ganho em 17 casos dos 30 testados (56.67%), verificou-se perda em 36.67% dos cenários e empate em duas situações (6.67%).

Tabela 5.2: Comparação do método de geração de *pool* proposto baseado em acurácia e exploração do espaço de complexidade com a estratégia de geração aleatório de *pools* em quatro cenários de seleção dinâmica de classificador individual: OLA, LCA, A Priori and A Posteriori. Os resultados apresentados consistem na média e desvio padrão das 20 repetições. Os melhores resultados são destacados em negrito.

	OLA		LCA		A Priori		A Posteriori	
Dataset	Randômica	AG	Randômica	AG	Randômica	AG	Randômica	AG
Adult	84.04 (2.87)	**84.77 (2.80)**	83.26 (2.52)	**83.52 (2.87)**	84.01 (2.34)	**85.73 (2.78)***	83.55 (2.60)	**85.20 (2.91)***
Banana	**85.03 (1.73)**	84.87 (1.68)	**84.88 (1.78)***	84.68 (1.79)	84.52 (1.64)	**84.57 (1.87)**	83.82 (1.69)	**83.83 (1.76)**
Blood	**76.12 (0.26)**	**76.12 (0.26)**	**75.86 (1.23)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**
CTG	**86.99 (0.89)**	86.70 (1.04)	80.94 (0.95)	**81.33 (0.82)***	86.45 (1.38)	**86.76 (0.76)**	86.28 (1.29)	**86.30 (0.85)**
Diabetes	64.92 (2.85)	**65.00 (2.27)**	62.37 (2.64)	**63.75 (2.89)***	**65.00 (2.03)**	64.51 (2.46)	64.38 (2.44)	**64.41 (2.76)**
Ecoli	**64.94 (4.74)**	62.38 (3.09)	52.56 (5.52)	**52.72 (5.91)**	64.05 (5.05)	**64.46 (4.83)**	**63.39 (4.00)**	62.38 (3.94)
Faults	**58.35 (2.31)***	56.82 (2.99)	37.96 (10.5)	**38.84 (10.1)***	**56.08 (2.15)**	55.70 (3.42)	54.63 (2.25)	**54.73 (3.66)**
German	71.90 (2.35)	**73.48 (2.42)***	64.22 (3.31)	**65.88 (2.98)***	71.74 (2.54)	**72.80 (3.08)**	69.80 (2.39)	**71.52 (2.85)**
Glass	**53.77 (4.79)**	51.89 (7.46)	24.15 (9.04)	**25.47 (10.2)**	47.92 (7.41)	**53.11 (6.09)**	24.72 (19.5)	**30.75 (16.9)**
Haberman	73.68 (0.59)	**74.14 (0.75)**	73.62 (0.88)	**74.41 (1.74)***	73.82 (0.39)	**74.41 (1.35)**	73.68 (0.59)	**73.29 (3.00)**
Heart	78.58 (4.04)	**80.37 (4.12)**	70.15 (4.58)	**72.84 (5.31)***	78.81 (4.02)	**80.67 (3.73)**	72.24 (8.64)	**77.31 (4.10)***
ILPD	**69.72 (3.38)**	68.62 (2.91)	61.76 (3.67)	**63.79 (3.82)***	69.24 (3.22)	**70.34 (2.97)**	67.52 (4.16)	**68.41 (3.68)**
Image	**42.03 (2.30)**	41.25 (2.12)	32.18 (4.11)	**32.82 (3.87)***	40.68 (1.82)	**41.33 (1.99)**	40.29 (2.35)	**41.42 (2.34)**
Ionosphere	80.91 (3.88)	**81.48 (4.56)**	69.72 (4.21)	**72.10 (4.14)***	80.63 (4.72)	**80.72 (4.14)**	77.84 (4.62)	**81.02 (4.36)***
Laryngeal1	**80.09 (4.40)**	78.87 (5.08)	70.94 (6.24)	**73.49 (5.04)***	**81.13 (5.57)**	80.00 (3.89)	**78.11 (6.92)**	77.08 (5.11)
Laryngeal3	66.02 (4.19)	**67.22 (4.90)**	55.34 (5.77)	**58.81 (6.25)***	66.02 (4.70)	**66.14 (3.23)**	62.84 (6.18)	**63.92 (4.43)**
Lithuanian	**68.30 (2.49)**	67.06 (3.19)	**70.38 (1.84)**	69.03 (2.40)	67.06 (2.42)	**67.49 (3.21)**	64.50 (3.36)	**65.15 (3.16)**
Liver	**60.99 (3.63)**	60.17 (3.14)	50.06 (5.30)	**51.34 (5.15)**	**58.90 (5.51)**	57.38 (4.96)	**55.17 (5.47)**	54.24 (8.20)
Magic	**79.22 (0.70)**	79.00 (0.56)	78.17 (0.55)	**78.23 (0.60)**	78.73 (0.81)	**78.78 (0.53)**	**78.59 (0.83)**	78.53 (0.63)
Mammo	79.78 (2.29)	**80.56 (2.96)**	76.23 (3.11)	**77.97 (3.08)***	80.10 (3.17)	**80.51 (3.44)**	79.08 (2.74)	**79.15 (3.44)**
Monk	81.99 (3.56)	**82.59 (3.22)**	69.26 (3.98)	**72.73 (3.81)***	79.58 (4.07)	**81.48 (4.86)***	66.76 (12.8)	**68.19 (14.5)**
Phoneme	77.18 (1.25)	**77.22 (0.82)**	76.54 (0.96)	**76.77 (0.86)**	76.17 (1.47)	**77.14 (0.98)***	76.08 (1.55)	**77.05 (0.96)***
Sonar	**61.73 (6.32)**	60.29 (5.38)	**46.83 (7.28)**	45.96 (5.73)	58.65 (6.10)	**61.44 (5.94)**	41.63 (22.5)	**45.96 (20.8)**
Thyroid	93.67 (1.25)	**93.73 (1.45)**	91.88 (1.97)	**91.99 (1.71)**	93.29 (1.71)	**93.41 (1.92)**	**92.51 (2.79)**	92.02 (1.97)
Vehicle	32.16 (3.87)	**32.44 (2.65)**	**24.53 (4.66)**	23.96 (4.74)	30.28 (3.92)	**32.04 (3.33)**	29.55 (3.85)	**31.11 (4.01)**
Vertebral	**81.53 (4.35)**	80.73 (4.70)	71.67 (5.13)	**73.20 (5.20)**	**82.53 (3.32)**	81.67 (5.07)	**78.00 (6.06)**	77.87 (8.13)
WBC	77.36 (14.6)	**77.57 (13.5)**	84.86 (3.70)	**84.89 (3.25)**	77.75 (14.1)	**79.79 (12.18)**	70.32 (17.3)	**79.01 (10.6)**
WDVG	79.92 (1.13)	**80.24 (0.88)**	67.81 (3.77)	**69.91 (3.41)***	79.18 (1.21)	**80.08 (1.19)***	78.37 (1.35)	**79.49 (1.33)**
Weaning	76.87 (4.14)	**77.27 (4.43)**	60.40 (5.50)	**63.33 (6.57)***	75.67 (5.21)	**78.40 (4.19)***	59.13 (16.8)	**64.93 (19.9)**
Wine	**33.52 (3.44)**	**33.52 (2.95)**	45.23 (14.1)	**43.86 (13.3)**	34.32 (4.71)	**35.68 (7.58)**	32.61 (2.41)	**33.41 (2.88)**

Tabela 5.3: Comparação do método de geração de *pool* proposto baseado em acurácia e exploração do espaço de complexidade com a estratégia de geração aleatório de *pools* em dois cenários de seleção dinâmica de *ensembles* de classificadores: KNOA-Union (KU) e KNORA-Eliminate (KE). Além disso, são apresentados também os resultados do *single best* (SB) e da combinação de todos os classificadores (ALL). Os resultados apresentados consistem na média e desvio padrão das 20 repetições. Os melhores resultados são destacados em negrito.

	SB		ALL		KNORA-U		KNORA-E	
Dataset	Randômica	AG	Randômica	AG	Randômica	AG	Randômica	AG
Adult	84.74 (2.92)	**85.29 (2.93)**	**86.83 (2.58)**	86.66 (2.67)	83.08 (2.03)	**84.24 (1.85)***	81.92 (1.85)	**83.72 (1.92)***
Banana	**84.49 (1.43)**	84.10 (1.61)	84.16 (1.59)	**84.40 (1.76)**	**87.59 (1.24)**	87.42 (1.12)	**85.06 (1.46)**	**85.06 (1.45)**
Blood	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**
CTG	86.62 (1.24)	**86.85 (1.54)**	**88.14 (1.12)**	**88.14 (1.14)**	84.67 (0.96)	**85.01 (0.93)***	83.95 (0.89)	**84.21 (0.89)***
Diabetes	64.95 (2.06)	**65.09 (2.38)**	64.48 (1.31)	**65.26 (2.54)**	64.48 (0.95)	**64.64 (0.87)**	64.56 (1.09)	**64.79 (1.11)**
Ecoli	63.63 (7.54)	**65.06 (4.65)**	53.39 (12.5)	**53.75 (13.7)**	54.40 (2.77)	**54.44 (2.77)**	52.02 (2.10)	**52.26 (1.99)**
Faults	55.76 (1.94)	**55.85 (3.11)**	**44.16 (11.7)**	43.98 (11.8)	43.65 (2.49)	**44.96 (2.80)***	37.41 (1.96)	**39.25 (2.27)***
German	**72.74 (2.99)**	72.28 (3.21)	**76.38 (2.28)**	75.70 (2.01)	71.88 (1.02)	**72.94 (1.36)***	71.52 (0.88)	**72.28 (1.15)***
Glass	**52.26 (7.98)**	50.00 (4.64)	48.77 (6.76)	**49.15 (8.23)**	51.13 (5.37)	**51.98 (5.15)**	45.28 (5.13)	**47.45 (4.71)***
Haberman	**74.01 (1.24)**	73.75 (0.65)	73.75 (0.29)	**73.95 (0.79)**	**73.68 (0.00)**	**73.68 (0.00)**	**73.68 (0.00)**	**73.68 (0.00)**
Heart	79.40 (4.54)	**79.93 (4.49)**	**84.33 (2.52)**	83.36 (3.48)	75.67 (4.77)	**77.16 (4.50)**	74.55 (4.21)	**76.19 (4.31)**
ILPD	69.48 (3.94)	**69.76 (3.88)**	**71.86 (3.89)**	71.83 (3.83)	**71.90 (0.61)**	**71.90 (0.75)**	**71.83 (0.45)**	**71.83 (0.45)**
Image	40.75 (3.10)	**41.05 (3.53)**	25.60 (7.83)	**25.76 (7.81)**	**36.39 (1.28)**	36.35 (1.52)	**35.86 (1.19)**	**35.88 (1.48)**
Ionosphere	81.59 (3.74)	**82.39 (4.03)**	82.61 (2.62)	**83.24 (3.33)**	**89.38 (3.54)**	87.73 (3.40)	**88.69 (3.83)**	87.50 (3.17)
Laryngeal1	**80.47 (4.40)**	80.19 (4.94)	**81.89 (4.60)**	81.42 (4.55)	73.58 (4.77)	**74.25 (3.79)**	72.08 (4.57)	**72.74 (4.07)**
Laryngeal3	67.05 (4.07)	**67.10 (3.42)**	**70.68 (2.80)**	69.94 (3.62)	**61.14 (3.48)**	61.02 (3.43)	58.64 (3.86)	**59.37 (3.25)**
Lithuanian	63.67 (2.97)	**64.43 (4.02)**	65.50 (2.26)	**66.39 (2.46)**	58.43 (0.98)	**59.87 (1.40)***	56.69 (0.80)	**58.04 (1.21)***
Liver	**63.02 (6.00)**	60.17 (5.25)	61.45 (4.05)	**62.56 (4.97)**	56.63 (5.59)	**57.09 (6.77)**	52.21 (3.51)	**55.17 (4.58)***
Magic	**78.74 (0.76)**	**78.74 (0.57)**	78.85 (0.58)	**78.87 (0.57)**	78.90 (0.58)	**78.95 (0.58)**	**78.80 (0.55)**	**78.80 (0.58)**
Mammo	80.58 (3.06)	**80.66 (3.57)**	81.76 (2.15)	**81.86 (2.41)**	78.50 (2.66)	**79.13 (2.93)**	77.85 (2.83)	**78.77 (2.76)***
Monk	**79.58 (3.98)**	79.21 (3.43)	79.86 (1.75)	**82.64 (5.43)***	**79.31 (3.26)**	78.84 (3.67)	78.15 (3.72)	**79.86 (3.08)**
Phoneme	77.08 (1.06)	**77.15 (0.86)**	76.59 (0.63)	**77.11 (0.80)***	74.77 (0.74)	**75.74 (0.81)***	74.07 (0.64)	**75.17 (0.74)***
Sonar	61.06 (5.40)	**61.63 (4.65)**	61.54 (5.27)	**64.62 (6.21)**	53.37 (1.03)	**55.00 (1.86)***	53.17 (0.92)	**53.85 (1.36)***
Thyroid	**93.61 (1.97)**	93.21 (2.86)	**96.42 (1.06)***	95.61 (1.32)	**89.22 (2.73)**	88.70 (4.11)	**88.79 (2.82)**	88.38 (4.05)
Vehicle	**31.30 (3.68)**	30.52 (4.08)	32.39 (5.21)	**33.65 (5.24)**	27.51 (1.69)	**28.22 (2.57)**	26.30 (1.23)	**26.75 (1.31)**
Vertebral	**81.33 (4.02)**	80.93 (5.64)	**83.53 (3.16)**	83.13 (4.22)	79.27 (4.47)	**80.47 (4.64)**	78.47 (3.80)	**79.67 (4.58)**
WBC	67.78 (20.2)	**81.73 (15.7)***	85.92 (3.06)	**90.18 (2.16)***	**89.15 (3.16)**	88.49 (2.96)	**88.59 (3.72)**	88.27 (3.47)
WDVG	**79.86 (1.14)**	79.54 (0.97)	81.50 (0.79)	**81.54 (1.02)**	71.07 (1.12)	**72.71 (1.15)***	69.66 (1.05)	**71.6 (1.18)***
Weaning	75.60 (5.45)	**76.60 (4.82)**	81.47 (4.56)	**83.18 (3.42)**	66.93 (4.98)	**71.07 (3.96)***	64.47 (4.45)	**68.47 (3.58)***
Wine	34.43 (6.32)	**36.59 (9.00)**	32.84 (1.13)	**38.30 (10.9)***	**32.84 (1.13)**	**32.84 (1.13)**	**32.84 (1.13)**	**32.84 (1.13)**

A Figura 5.1 apresenta a comparação par-a-par entre os métodos randômico e AG para todas as seis estratégias dinâmicas, *single best* e para a combinação de todos os classificadores. As colunas em vermelho representam os cenários nos quais adotar a estratégia aleatória na geração dos subconjuntos mostrou-se a melhor opção, enquanto as colunas em azul correspondem aos casos em que o AG pôde alcançar a maior acurácia. As representações em verde indicam os casos onde ocorreu empate entre as abordagens.

De forma a comparar o comportamento das duas estratégias de geração de subconjuntos foi aplicado o teste de Wilcoxon com uma significância de 5%. Os asteriscos apresentados nas Tabelas 5.2 e 5.3 destacam os casos em que foi observada diferença significativa entre as abordagens comparadas.

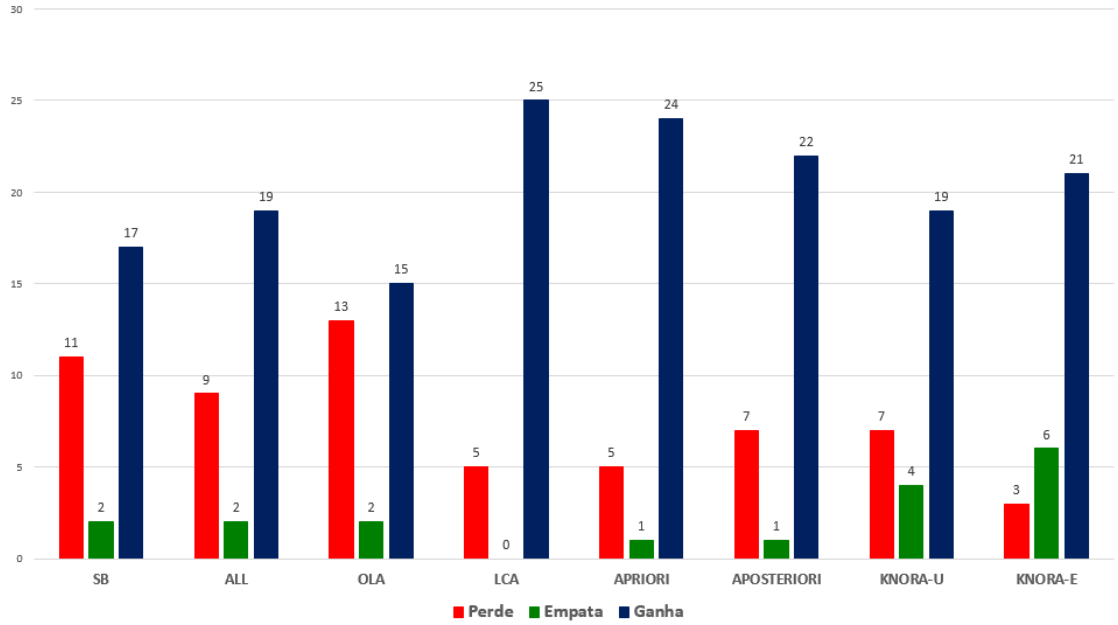

Figura 5.1: Comparação par-a-par da performance dos métodos seleção dinâmica, *single best* e combinação com base nos dois métodos de geração.

5.1.1 Análise de Dispersão

Analisado-se os valores obtidos para as acurácias das duas estratégias de geração de subconjuntos foi possível notar a interessante possibilidade em se adotar um algoritmo genético para evoluir os subconjuntos focando em acurácia unida à exploração do espaço de complexidade. Os resultados mostraram que, tanto para as estratégias estáticas, quanto para as abordagens dinâmicas (individuais e *ensembles*) os *pools* formados com foco em acurácia e exploração da complexidade possibilitaram alcançar taxas de reconhecimento maiores.

No entanto, além da acurácia dos métodos de seleção dinâmica e combinação, as dispersões dos subconjuntos no espaço de complexidade também foram analisadas, visto que um dos objetivos buscados nesta pesquisa foi gerar conjuntos de forma a melhor cobrir tal espaço.

Visando a comparação entre as estratégias, para cada um dos 30 problemas foi calculada a dispersão média de cada conjunto formado randomicamente e também pelo AG. A Equação 5.2 define a forma com que os valores da dispersão média foram calculados. A Tabela 5.4 apresenta as distribuições médias calculadas ao longo das 20 repetições.

$$\overline{Dispersao} = \frac{\sum_{j=1}^{r} \frac{\sum_{i=1}^{M} Disp_{C_i}}{M}}{r} \tag{5.2}$$

em que M corresponde ao número de classificadores presentes no *pool*, nc refere-se ao número de medidas de complexidade adotadas no processo e r representa o número de repetições executadas enquanto $x_{i,k}$ e $x_{j,k}$ correspondem aos valores da k-ésima medida no espaço de complexidade para os indivíduos i e j, respectivamente.

Os valores apresentados evidenciam o aumento na dispersão dos subconjuntos dentro do espaço de complexidade quando se adota o AG. Assumindo que tais valores representam o espaçamento médio de cada subconjunto em relação aos demais, pode-se verificar que as áreas de cobertura do espaço de complexidade foram aumentadas consideravelmente, alcançando dessa forma, um dos objetivos da pesquisa.

Os valores observados na Tabela 5.4 revelam o aumento na dispersão dos subconjuntos no espaço de complexidade quando é adotado o AG no processo de geração. Para corroborar tal afirmação aplicou-se o teste de Wilcoxon com 5% de significância para comparar os resultados dos métodos de geração analisados. Diferenças significativas apresentam o marcador "*".

De forma a retratar a ocupação do espaço, os gráficos apresentados nas Figuras 5.2, 5.3 e 5.4 ilustram a mudança ocorrida na exploração do espaço de complexidade para as dimensões $F1$ e $N2$. Nas representações, que tratam das bases Haberman, Heart e Laryngeal1, respectivamente, os círculos vermelhos correspondem ao conjunto inicial gerado aleatoriamente, enquanto os marcadores em azul referem-se ao conjunto final obtido pela execução do AG.

As ilustrações refletem um comportamento comum observado ao longo dos problemas estudados: a distribuição de complexidade do *pool* gerado pelo AG consegue explorar melhor o espaço de complexidade $F1$ x $N2$, no entanto, a variação em relação à medida $F1$ é mais evidente, enquanto a variação no eixo $N2$, mais discreta, é caracterizada por um certo "deslocamento" do conjunto. Este comportamento é fruto da busca pela cobertura

Tabela 5.4: Dispersão média dos subconjuntos gerados pela estratégia randômica e pelo AG no espaço de complexidade

	Dispersão	
Base	Randômica	AG
Adult	0.53 (0.07)	**1.26 (0.36)***
Banana	0.20 (0.02)	**0.25 (0.03)***
Blood	0.14 (0.01)	**0.22 (0.04)***
CTG	0.24 (0.05)	**0.25 (0.05)**
Diabetes	0.18 (0.03)	**0.28 (0.05)***
Ecoli	20.4 (5.02)	**21.1 (4.55)**
Faults	**0.74 (0.06)**	0.73 (0.07)
German	0.12 (0.02)	**0.31 (0.15)***
Glass	2.62 (2.16)	**7.90 (23.9)**
Haberman	0.17 (0.03)	**0.36 (0.10)***
Heart	0.40 (0.11)	**2.19 (1.68)***
ILPD	0.08 (0.01)	**0.12 (0.03)***
Image	12.7 (2.44)	**13.1 (2.43)**
Ionosphere	0.21 (0.04)	**0.47 (0.16)***
Laryngeal1	0.68 (0.12)	**1.59 (0.58)***
Laryngeal3	1.77 (0.29)	**1.93 (0.39)***
Lithuanian	0.17 (0.01)	**0.24 (0.03)***
Liver	0.10 (0.01)	**0.17 (0.03)***
Magic	0.03 (0.00)	**0.04 (0.00)**
Mammo	0.26 (0.04)	**0.30 (0.09)**
Monk	0.28 (0.04)	**0.60 (0.15)***
Phoneme	0.04 (0.01)	**0.06 (0.01)***
Sonar	0.28 (0.04)	**0.72 (0.26)***
Thyroid	0.83 (0.11)	**1.62 (0.34)***
Vehicle	0.37 (0.04)	**0.42 (0.05)***
Vertebral	0.30 (0.05)	**0.52 (0.15)***
WBC	0.55 (0.07)	**0.86 (0.12)***
WDVG	0.12 (0.01)	**0.13 (0.02)**
Weaning	0.38 (0.06)	**0.65 (0.15)***
Wine	1.90 (0.25)	**2.16 (0.59)***

do espaço de complexidade, uma vez que os conjuntos construídos apresentam centróides mais afastados no espaço de características (ocasionando o aumento de $F1$) e, ao mesmo tempo, regiões de fronteira mais intrincadas, evidenciadas pelo aumento no valor de $N2$.

Com base nos valores observados para as distribuições médias dos subconjuntos no espaço de complexidade (apresentados na Tabela 5.4), podemos concluir que um dos objetivos da pesquisa foi alcançado, pois com a execução do algoritmo genético proposto, foi possível gerar uma maior cobertura do espaço de complexidade e, além disso, prover *pools* que permitem aos métodos de seleção dinâmica e às estratégias estáticas uma melhora em sua acurácia.

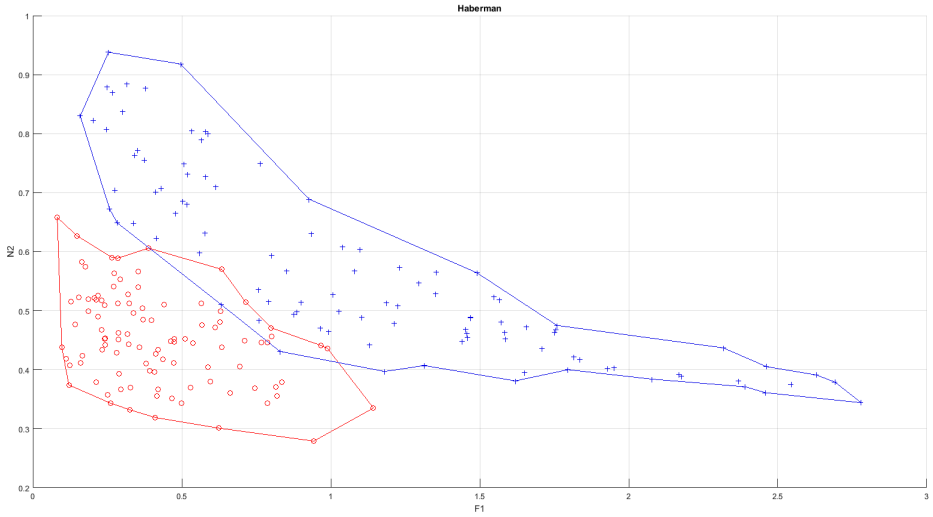

Figura 5.2: Dispersão dos classificadores gerados para a base Haberman no espaço de complexidade. Em vermelho os elementos gerados de forma randômica e, em azul, o *pool* obtido pelo AG.

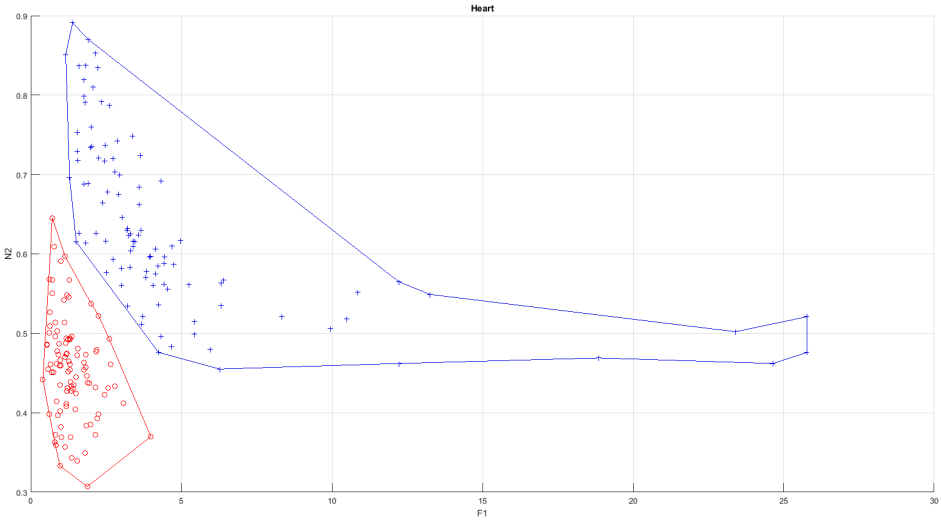

Figura 5.3: Dispersão dos classificadores gerados para a base Heart no espaço de complexidade. Em vermelho os elementos gerados de forma randômica e, em azul, o *pool* obtido pelo AG.

5.2 Experimento 2 - Seleção de Classificadores baseada Complexidade

Esta seção apresenta os experimentos realizados visando avaliar o método de seleção proposto, adotando, como critério de seleção, índices de acurácia unidos às medidas de complexidade. Neste escopo, foram empregadas as trinta bases detalhadas na Tabela 5.1.

Para cada problema construi-se um *pool* composto de 100 perceptrons gerados

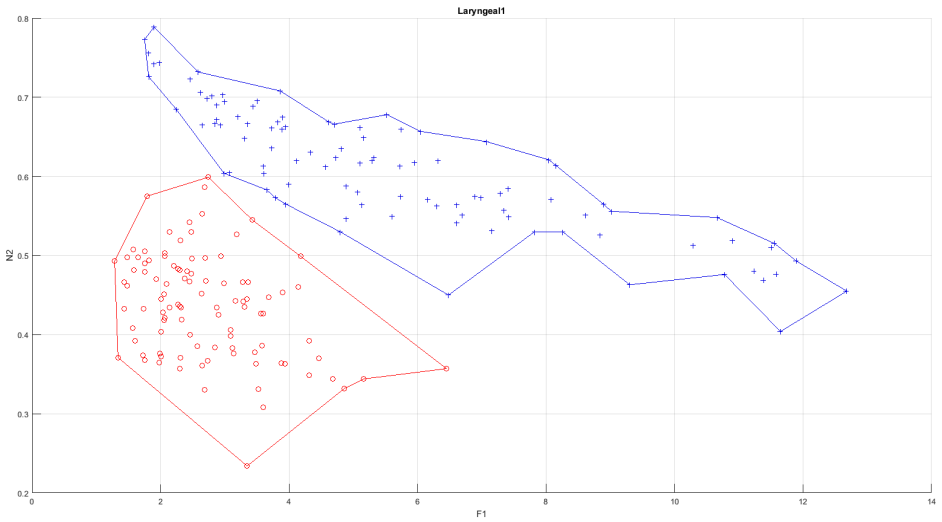

Figura 5.4: Dispersão dos classificadores gerados para a base Laryngeal1 no espaço de complexidade. Em vermelho os elementos gerados de forma randômica e, em azul, o *pool* obtido pelo AG.

através do Bagging (BREIMAN, 1996). Cada subconjunto contém 10% ou 20% da quantidade de elementos presentes no conjunto de treino, de acordo com o tamanho dos conjuntos de treino. A ideia foi formar um *pool* composto de classificadores fracos e cujos votos possuísse certa complementariedade. O percentual de 20% foi adotado para as bases menores. A escolha pelo perceptron como classificador base deu-se devido ao fato de ele caracterizar-se como um indutor fraco e instável. Além disso, classificadores fracos tendem a evidenciar as diferença de performance entre abordagens de seleção dinâmica (KO; SABOURIN; BRITTO JR., 2008).

O tamanho das vizinhanças usadas para estimar os descritores de complexidade para cada instância de teste foi definido como 30 elementos. Ao adotar tal valor assegurou-se, para as bases testadas, a presença de elementos de pelo menos duas classes distintas na vizinhança, tornando-se assim possível o cálculo das medidas de complexidade. Todavia, visando determinar o tamanho da vizinhança supracitada, foram rodados experimentos em 13 problemas da UCI, onde variou-se o tamanho da vizinhança entre 20 e 50 instâncias.

Como descrito anteriormente (Seção 4.2), foram utilizadas três medidas de complexidade: F1, N2 e N4. A ideia foi empregar um descritor de cada umas das três categorias descritas no Capítulo 3. Para orientar nossa escolha, realizou-se um estudo com treze bases do repositório da UCI (BACHE; LICHMAN, 2013), no qual foi analisada a correlação de Pearson entre todas as 14 medidas de complexidade disponíveis na biblioteca DCoL (ORRIOLS-PUIG; MACIà; HO, 2010). Verificou-se que estas três medidas apresentam baixa correlação entre si, indicando que elas podem explicar fenômenos distintos.

Visando avaliar a contribuição em se adotar medidas de complexidade no processo

de seleção, comparou-se o método proposto com outras 5 técnicas de seleçao dinâmica já estabelecidas na literatura. O desempenho foi comparado também às abordagens *single best* (SB) e combinação de todos os classificadores. Com relação às estratégias dinâmicas, foram implementadas soluções baseadas em seleção dinâmica de classficicador individual (LCA (WOODS; KEGELMEYER JR.; BOWYER, 1997), OLA (WOODS; KEGELMEYER JR.; BOWYER, 1997), e A Priori (GIACINTO; ROLI, 1999)) bem como abordagens para seleção dinâmica de *ensembles* de classificadores (KNORA-Union e KNORA-Eliminate (KO; SABOURIN; BRITTO JR., 2008)). Com o objetivo de estimar a acurácia local dos classificadores, para tais métodos foi utilizada uma vizinhança de dimensão 7. Este valor mostrou-se o mais adequado em estudos anteriores (KO; SABOURIN; BRITTO JR., 2008), (SANTANA et al., 2006). Nesta avaliação, todos os métodos trabalharam sobre pools gerados através do Bagging.

O desempenho médio ao longo das vinte repetições para cada abordagem para cada problema de classificação é apresentado na Tabela 5.5. Os valores em negrito representam a maior acurácia obtida para cada problema. Na última coluna da tabela é apresentada a acurácia do Oráculo para cada problema considerando o *pool* de classificadores construído. Tal limite superior em termos de performance é estimado considerando a suposição de que se pelo menos um classificador consegue reconhecer com sucesso um dado padrão, então o *pool* também é capaz de reconhecê-lo.

Analisando-se os valores apresentados é possível observar que o método de seleção dinâmica proposto suplantou o desempenho do melhor classificador (SB) em 28 dos 30 problemas, e que em relação à combinação de todos os classificadores, pôde obter melhor desempenho em 23 dos 30 casos analisados. Quando comparado apenas com os métodos de seleção dinâmica, a estratégia proposta venceu em 123 dos 150 cenários (82.00%). A Figura 5.5 apresenta a comparação par-a-par com todos os métodos testados. As barras em azul representam o número de casos em que há contribuição em se adotar as medidas de complexidade e as barras em vermelho correspondem ao número de problemas em que a solução proposta perde. Em uma situação há empate entre LCA e nossa abordagem (representado em verde no gráfico).

Tabela 5.5: Comparação do método de seleção baseado em complexidade proposto (DSOC) com o melhor classificador (*single best* - SB) do *pool*, com a combinação de todos os classificadores (ALL), com métodos de seleção dinâmica como OLA, LCA, A Priori, Knora-U (KU), KNORA-E (KE), e o desempenho do oráculo. Os resultados apresentados consistem na média e desvio padrão das 20 repetições. Os melhores resultados são destacados em negrito.

Dataset	SB	ALL	OLA	LCA	a Priori	KU	KE	DSOC	Oracle
Adult	83.6 (2.3)	**86.7 (2.4)**	82.4 (2.8)	82.3 (2.5)	80.6 (4.8)	76.6 (2.3)	71 (3.2)	85.6 (2.5)	99.7 (0.4)
Banana	85.3 (1.4)	84.1 (1.4)	89.2 (1.9)	**89.5 (1.9)**	86.1 (2.5)	89.2 (1.4)	84.4 (1.9)	87.4 (2.4)	89.8 (1.9)
Blood	76.4 (0.3)	**76.4 (0.2)**	74.2 (3.0)	74.2 (3.2)	69.0 (16.4)	76.4 (0.2)	76.4 (0.2)	72.7 (2.7)	100 (-)
CTG	69.8 (11.1)	86.6 (1.7)	87.9 (1.1)	88.4 (1.2)	84.1 (1.6)	85.3 (0.9)	81.3 (1.0)	**88.8 (1.1)**	99.9 (0.1)
Diabetes	66.0 (1.3)	64.5 (1.4)	69.9 (2.9)	**70.0 (2.4**)	58.6 (7.8)	65.5 (0.4)	65.1 (-)	69.4 (3.5)	92.3 (7.2)
Ecoli	63.7 (3.9)	42.1 (0.6)	77.9 (3.8)	79.9 (2.9)	55.1 (9.2)	64.0 (4.2)	42.1 (0.6)	**80.5 (3.7)**	97.1 (1.7)
Faults	31.2 (14.2)	63.5 (2.8)	64.9 (2.5)	66.4 (1.6)	51.4 (2.9)	53.6 (2.0)	36.7 (2.3)	**67.6 (1.5)**	99.2 (0.4)
German	59.5 (5.0)	**75.7 (2.5)**	68.7 (2.9)	70.0 (2.9)	66.7 (3.4)	70.1 (0.3)	70.0 (-)	72.8 (2.4)	100 (-)
Glass	56.6 (6.6)	58.0 (5.2)	59.9 (6.9)	60.7 (8.6)	46.4 (9.4)	49.3 (5.8)	33.6 (1.7)	**63.1 (6.2)**	99.8 (0.6)
Haberman	75.3 (2.9)	73.7 (-)	75.3 (3.9)	74.9 (3.8)	73.9 (1.4)	73.8 (0.3)	73.7 (-)	**76.4 (3.5)**	88.8 (5.9)
Heart	79.1 (4.9)	**83.8 (3.2)**	76.9 (3.3)	75.7 (4.3)	75.8 (6.3)	70.8 (4.1)	68.2 (3.6)	82.1 (3.4)	100 (-)
ILPD	68.1 (3.5)	70.6 (3.5)	66.9 (3.0)	67.7 (3.2)	64.6 (6.0)	**71.7 (-)**	**71.7 (-)**	66.6 (2.9)	100 (0.1)
Image	16.1 (5.6)	36.3 (1.1)	68.6 (3.0)	**70.9 (2.6)**	47.9 (3.8)	49.9 (1.9)	27.8 (1.2)	70.3 (2.4)	77.8 (2.7)
Ionosphere	78.3 (2.8)	72.0 (2.6)	80.3 (2.6)	86.1 (3.2)	72.1 (4.9)	79.5 (6.2)	56.3 (15.3)	**86.9 (3.2)**	98.2 (1.9)
Laryngeal1	80.0 (3.8)	78.6 (4.7)	79.4 (5.0)	79.8 (4.9)	76.2 (4.3)	69.2 (3.9)	66.9 (3.3)	**82.4 (5.2)**	99.9 (0.4)
Laryngeal3	66.0 (5.1)	66.5 (3.3)	65.4 (5.3)	66.2 (4.9)	61.5 (5.5)	57.1 (4.0)	50.1 (3.6)	**67.7 (4.0)**	99.6 (0.7)
Lithuanian	67.9 (6.5)	50.8 (0.5)	**95.9 (1.1)**	95.8 (1.2)	85.9 (2.7)	72.3 (3.2)	50.0 (-)	95.7 (2.5)	99.9 (0.5)
Liver	65.6 (3.4)	59.5 (2.7)	64.5 (4.5)	**66.7 (3.8)**	54.1 (6.5)	49.9 (4.6)	41.9 (-)	66.0 (3.2)	100 (-)
Magic	60.2 (9.5)	78.3 (0.6)	80.7 (0.6)	80.6 (1.7)	77.4 (0.6)	77.9 (0.5)	77.3 (0.5)	**80.9 (0.8)**	90.0 (0.5)
Mammo	64.2 (14.4)	**81.0 (2.6)**	78.9 (2.1)	78.8 (1.6)	77.5 (3.5)	75.9 (2.4)	72.6 (2.4)	80.6 (2.4)	98.3 (1.0)
Monk	78.4 (4.2)	80.5 (2.6)	86.5 (3.3)	86.5 (3.2)	77.5 (3.5)	63.8 (3.9)	55.1 (3.2)	**89.3 (3.2)**	100 (-)
Phoneme	62.2 (6.6)	76.3 (0.8)	81.6 (1.1)	**82.0 (0.9)**	76.1 (1.4)	75 (0.6)	72.9 (0.8)	80.6 (1.1)	96.5 (0.5)
Sonar	61.4 (9.0)	54.6 (1.9)	68.9 (6.8)	70.3 (6.8)	53.6 (5.9)	53.6 (1.3)	53.2 (0.9)	**71.0 (7.7)**	100 (-)
Thyroid	93.3 (1.9)	94.4 (1.3)	94.3 (1.9)	**95.6 (1.2)**	90.1 (20.8)	72 (3.5)	21.4 (4.8)	94.5 (1.4)	100 (-)
Vehicle	26.4 (3.8)	36.0 (5.7)	59.1 (3.7)	**59.4 (3.2)**	35.3 (5.9)	46.5 (3.1)	25.7 (0.2)	58.2 (7.5)	100.0 (-)
Vertebral	80.9 (2.9)	81.3 (3.6)	81.5 (4.9)	**81.8 (4.3)**	76.5 (5.3)	75.7 (2.9)	68.7 (1.3)	**81.8 (3.7)**	100 (-)
WBC	85.3 (0.3)	53.6 (0.2)	92.7 (3.0)	**93.2 (3.2)**	81.7 (16.4)	88.3 (0.2)	62.9 (0.2)	92.5 (2.5)	100 (-)
WDVG	44.6 (0.3)	**83.4 (1.1)**	80.1 (1.1)	80.4 (1.0)	77.2 (1.7)	65.2 (1.2)	61.5 (1.1)	82.4 (1.3)	99.9 (0.1)
Weaning	76.9 (5.6)	79.3 (5.0)	77.0 (3.3)	76.9 (4.1)	71.7 (4.9)	58.4 (1.7)	53.5 (2.2)	**82.9 (3.6)**	100 (-)
Wine	59.2 (8.7)	32.8 (1.1)	70.0 (5.2)	**70.2 (4.9)**	61.5 (6.0)	66.9 (4.3)	32.8 (1.2)	69.4 (6.4)	100 (-)

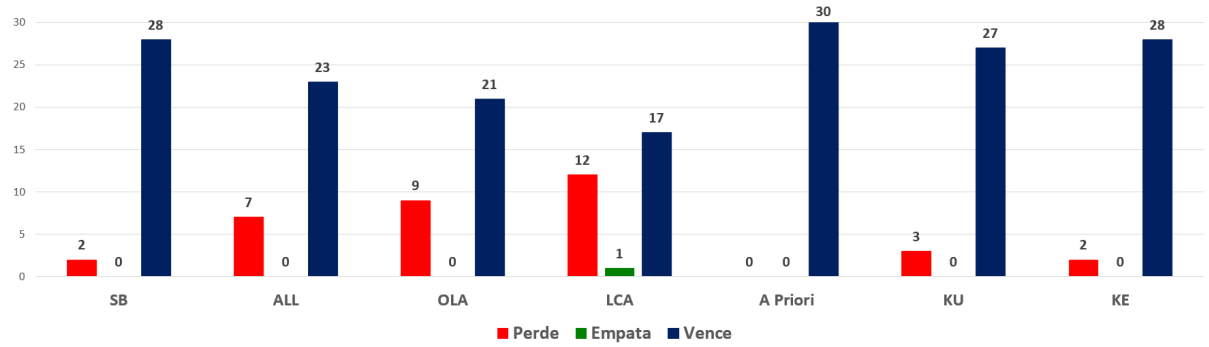

Figura 5.5: Comparação par-a-par do DSOC com todos os métodos testados. As barras em azul representam os número de problemas em que a adoção da complexidade na seleção superou o método comparado, enquanto as barras em vermelho referem-se ao número de derrotas da abordagem proposta. Os empates foram representados pelas barras na cor verde.

Visando comparar o comportamento das abordagens foi realizado o teste de Friedman com confiança de 95% e grau de liberdade igual a 7 (uma vez que foram comparados 8 métodos). Para todos os 30 problemas a hipótese nula foi rejeitada, indicando que há diferença significativa entre as acurácias das estratégias. Sendo assim, efetuou-se o teste post-hoc de Nemenyi para esboçar os rankings dos algoritmos em todos os problemas. Os resultados são apresentados na Figura 5.6. É possível notar que nossa abordagem alcançou a melhor posição geral do ranking. Entretanto, a diferença até o OLA e LCA é menor do que a distância crítica.

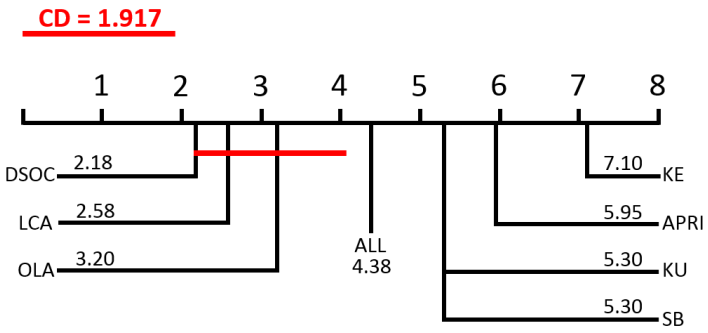

Figura 5.6: Representação gráfica do teste de Nemenyi comparando todos os métodos. Os valores apresentados próximos dos nomes dos métodos correspondem ao seu rank médio considerando os 30 problemas de classificação.

Com base nos resultados pode-se responder uma das perguntas desta pesquisa em que se indagava se a adoção de informação relacionada à análise de complexidade poderia contribuir para estimar a competência dos classificadores em um sistema multi-classificador baseado em seleção dinâmica.

Apesar dos resultados interessantes observados é importante destacar que a abor-

dagem proposta nem sempre apresentou o melhor desempenho. Deste modo, foi necessário entender porque em alguns casos há ganho na acurácia e em outros observa-se um declínio no desempenho quando foram adotados os descritores de complexidade. Focando neste objetivo, analisou-se cenários em que as duas situações ocorreram.

Considerando que a abordagem proposta leva em conta a similaridade entre a complexidade dos subconjuntos de dados sobre os quais os classificadores são treinados e da vizinhança da instância de teste, foi analisado o comportamento dos descritores $F1$, $N2$ e $N4$. Neste propósito, dividiu-se os descritores de complexidade em cem bins. Dessa forma, foi possível comparar ambas as distribuições, aquela relacionada aos subconjuntos adotados no treinamento bem como aquela estimada a partir da vizinhança das instâncias de teste projetadas no conjunto de validação.

A Figura 5.7 apresenta o comportamento observado para uma repetição das 20 realizadas para as base Monk (representada na coluna esquerda da figura) e Sonar (posicionada na coluna direita). O método de seleção dinâmica proposto obteve uma melhora em 5.5 pontos percentuais (um ganho significativo) para a base Monk em comparação com o valor observado para o segundo colocado no ranking (LCA). Por outro lado, observou-se, para a base Sonar, uma perda de 3.4 pontos percentuais (uma perda significativa) na acurácia do método quando comparada com a segunda posição do ranking (novamente o LCA). É possível observar a sobreposição entre a as distribuições de complexidade, em vermelho a distribuição estimada a partir da vizinhança da instância de teste e em azul a distribuição estimada com base nos conjuntos utilizados no treino. A distribuição do lado esquerdo, Figuras 5.7(a), 5.7(c) e 5.7(e) referem-se à medidas $F1$, $N2$ e $N4$ para a base Monk na qual o método proposto obteve resultados promissores. Da mesma forma, as Figuras 5.7(b), 5.7(d) e 5.7(f) estão relacionadas à base Sonar, na qual a adoção da abordagem proposta não é indicada.

Como pode ser observado nas representações dos bins, quando a sobreposição entre as distribuições é mais evidente, a contribuição do método de seleção dinâmica proposto é mais significativa. Tal fato motivou a investigação de estratégias para modificar o algoritmo de aprendendizagem dos *ensembles*, usando medidas de complexidade para direcionar a geração dos subconjuntos de treinamento empregados na geração do *pool* de classificadores. A ideia buscada com esta alternativa foi buscar uma melhor cobertura do espaço de complexidade do problema em mãos. Tal estratégia foi apresentada na Seção 4.1.

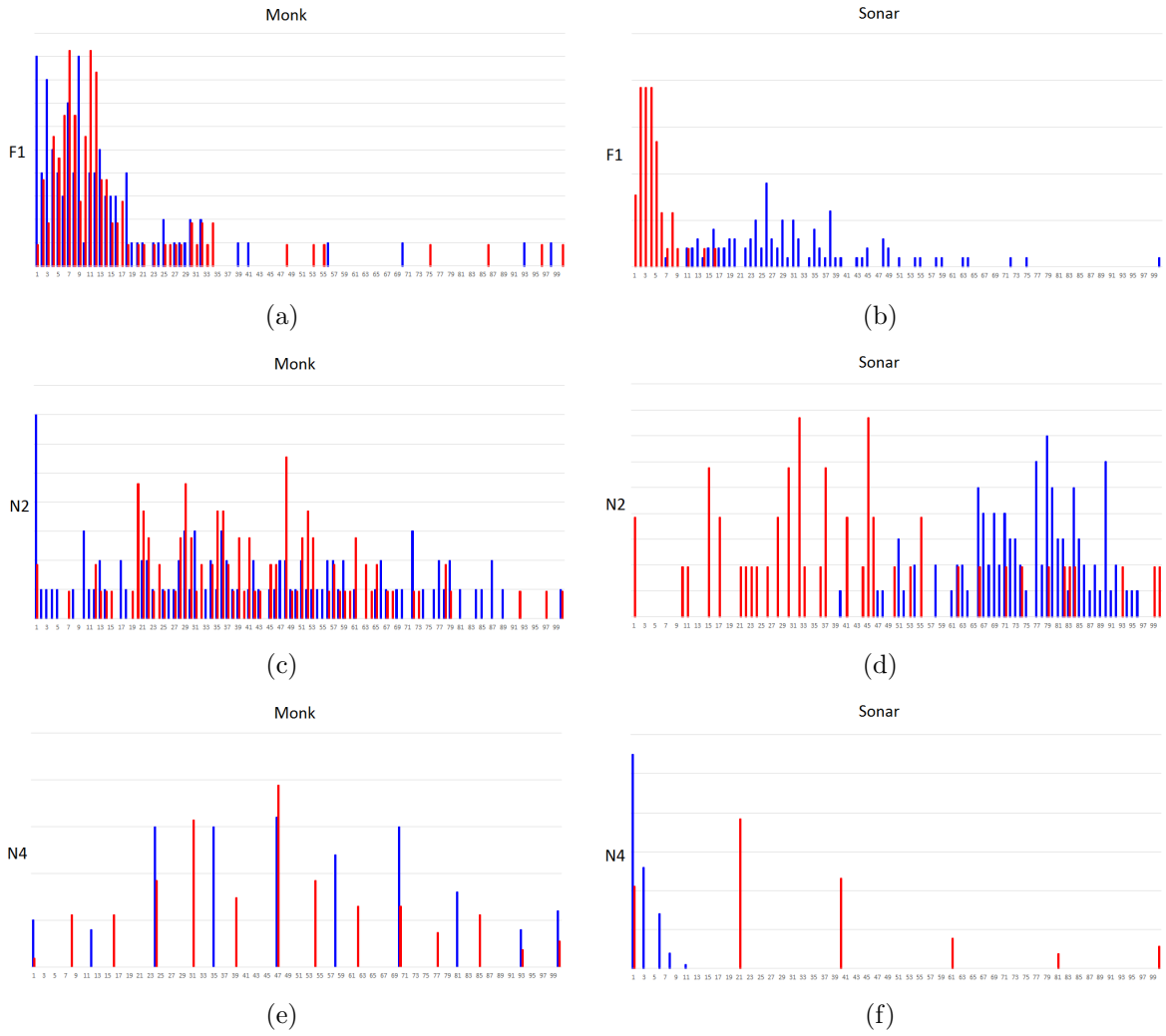

Figura 5.7: Sobreposição entre as distribuições de complexidade, em vermelho a distribuição estimada a partir das vizinhanças de cada instância e, em azul, a distribuição estimada com base nos conjuntos de treinamento: As Figuras 5.7(a), 5.7(c) e 5.7(e) referem-se às medidas F1, N2 e N4 para a base monk; similarmente as ilustrações 5.7(b), 5.7(d) e 5.7(f) representam a base sonar.

5.3 Experimento 3 - Combinando complexidade na geração e seleção dos classificadores

Com o objetivo de avaliar a contribuição da adoção de medidas de complexidade no processo de geração e seleção dos classificadores combinados, comparou-se o SMC proposto com outras seis técnicas de seleçao dinâmica já estabelecidas na literatura, para as quais, adotou-se os *pools* gerados de forma randômica, similar ao Bagging. Foram implementadas soluções baseadas em seleção dinâmica de classficicador individual (LCA (WOODS; KEGELMEYER JR.; BOWYER, 1997), OLA (WOODS; KEGELMEYER JR.; BOWYER, 1997), e A Priori e A Posteriori (GIACINTO; ROLI, 1999)) bem como abordagens para

seleção dinâmica de *ensembles* de classificadores (KNORA-Union e KNORA-Eliminate (KO; SABOURIN; BRITTO JR., 2008)). Novamente foi utilizada uma vizinhança de dimensão 7 para estimar a acurácia local dos classificadores.

O desempenho médio ao longo das vinte repetições para cada abordagem sobre todos os 30 problemas de classificação é apresentado na Tabela 5.6. Além da acurácia média de cada técnica, são apresentados também os valores dos desvios padrão observados. Os valores em negrito representam a maior acurácia obtida para cada problema.

Analisando-se os valores apresentados é possível observar que o sistema de múltiplos classificadores proposto suplantou o desempenho dos métodos de seleção dinâmica combinados com a geração aleatória de subconjuntos. A estratégia proposta venceu em 165 dos 180 cenários testados, correspondendo a 91.67% dos casos. Em dez oportunidades as técnicas de seleção dinâmica da literatura obtiveram o melhor desempenho (5.56%). Além disso, observou-se empate em 5 dos 180 testes (2.78%).

A Figura 5.8 apresenta a comparação par-a-par do SMC proposto com todos os métodos testados para os trinta problemas em estudo que foram treinados nos subconjuntos gerados aleatoriamente. As barras em azul representam o número de casos em que há contribuição em se adotar as medidas de complexidade para gerar e selecionar os classificadores, enquanto as barras em vermelho correspondem ao número de problemas em que a solução proposta perde. As barras na cor verde indicam a quantidade de empates observadas entre as estratégias.

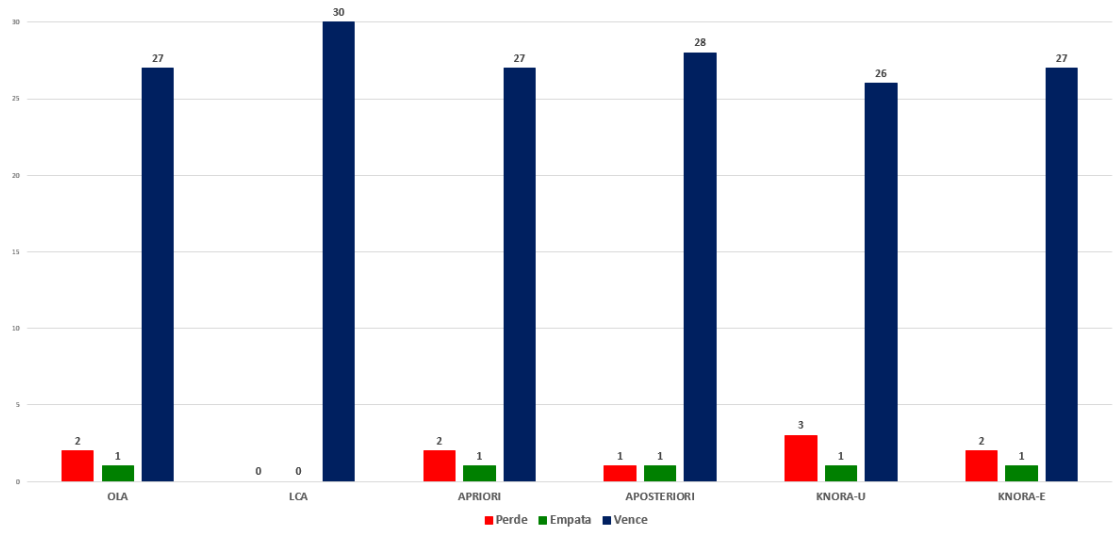

Figura 5.8: Comparação par-a-par da estratégia de SMC proposto perante todos os métodos de seleção testados, baseados na formação randômica do *pool*. As barras em azul representam os número de problemas em que a adoção da complexidade na geração e seleção superou o método comparado, enquanto as barras em vermelho referem-se ao número de derrotas da abordagem proposta. Os empates são representados pelas barras na cor verde.

Tabela 5.6: Comparação do SMC proposto com os métodos de seleção dinâmica OLA, LCA, A Priori (APRI), A Posteriori (APOS), KNORA-Union (KU), KNORA-Eliminate (KE) baseados na geração aleatória. Os resultados apresentados correspondem aos valores médios e desvios padrão das 20 repetições executadas. Os melhores valores são apresentados em negrito.

	SB	OLA	LCA	APRI	APOS	KU	KE	DSOC
	Rand	Rand	Rand	Rand	Rand	Rand	Rand	AG
Adult	84.74	84.04 (2.87)	83.26 (2.52)	84.01 (2.34)	83.55 (2.60)	83.08 (2.03)	81.92 (1.85)	**86.80 (1.85)***
Banana	84.49	85.03 (1.73)	84.88 (1.78)	84.52 (1.64)	83.82 (1.69)	**87.59 (1.24)**	85.06 (1.46)	87.20 (1.62)
Blood	76.12	**76.12 (0.26)**	75.86 (1.23)	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**
CTG	86.62	86.99 (0.89)	80.94 (0.95)	86.45 (1.38)	86.28 (1.29)	84.67 (0.96)	83.95 (0.89)	**87.50 (1.25)**
Diabetes	64.95	64.92 (2.85)	62.37 (2.64)	65.00 (2.03)	64.38 (2.44)	64.48 (0.95)	64.56 (1.09)	**68.41 (3.93)***
Ecoli	63.63	64.94 (4.74)	52.56 (5.52)	64.05 (5.05)	63.39 (4.00)	54.4 0(2.77)	52.02 (2.10)	**75.00 (2.44)***
Faults	55.76	58.35 (2.31)	37.96 (10.5)	56.08 (2.15)	54.63 (2.25)	43.65 (2.49)	37.41 (1.96)	**65.02 (1.58)***
German	72.74	71.90 (2.35)	64.22 (3.31)	71.74 (2.54)	69.80 (2.39)	71.88 (1.02)	71.52 (0.88)	**73.98 (3.28)***
Glass	52.26	53.77 (4.79)	24.15 (9.04)	47.92 (7.41)	24.72 (19.5)	51.13 (5.37)	45.28 (5.13)	**59.25 (5.22)***
Haberman	74.01	73.68 (0.59)	73.62 (0.88)	73.82 (0.39)	73.68 (0.59)	73.68 (0.00)	73.68 (0.00)	**74.61 (2.43)**
Heart	79.40	78.58 (4.04)	70.15 (4.58)	78.81 (4.02)	72.24 (8.64)	75.67 (4.77)	74.55 (4.21)	**83.58 (3.24)***
ILPD	69.48	69.72 (3.38)	61.76 (3.67)	69.24 (3.22)	67.52 (4.16)	**71.90 (0.61)**	71.83 (0.45)	66.86 (2.73)
Image	40.75	42.03 (2.30)	32.18 (4.11)	40.68 (1.82)	40.29 (2.35)	36.39 (1.28)	35.86 (1.19)	**51.03 (1.11)***
Ionosphere	81.59	80.91 (3.88)	69.72 (4.21)	80.63 (4.72)	77.84 (4.62)	**89.38 (3.54)**	88.69 (3.83)	86.53 (4.23)
Laryngeal1	80.47	80.09 (4.40)	70.94 (6.24)	81.13 (5.57)	78.11 (6.92)	73.58 (4.77)	72.08 (4.57)	**82.45 (4.51)**
Laryngeal3	67.05	66.02 (4.19)	55.34 (5.77)	66.02 (4.70)	62.84 (6.18)	61.14 (3.48)	58.64 (3.86)	**68.75 (5.04)**
Lithuanian	63.67	68.30 (2.49)	70.38 (1.84)	67.06 (2.42)	64.50 (3.36)	58.43 (0.98)	56.69 (0.80)	**82.47 (2.55)***
Liver	**63.02**	60.99 (3.63)	50.06 (5.30)	58.90 (5.51)	55.17 (5.47)	56.63 (5.59)	52.21 (3.51)	61.86 (5.39)
Magic	78.74	79.22 (0.70)	78.17 (0.55)	78.73 (0.81)	78.59 (0.83)	78.90 (0.58)	78.80 (0.55)	**79.99 (0.73)**
Mammo	80.58	79.78 (2.29)	76.23 (3.11)	80.10 (3.17)	79.08 (2.74)	78.50 (2.66)	77.85 (2.83)	**80.99 (2.23)**
Monk	79.58	81.99 (3.56)	69.26 (3.98)	79.58 (4.07)	66.76 (12.8)	79.31 (3.26)	78.15 (3.72)	**85.42 (3.44)***
Phoneme	77.08	77.18 (1.25)	76.54 (0.96)	76.17 (1.47)	76.08 (1.55)	74.77 (0.74)	74.07 (0.64)	**79.00 (1.04)***
Sonar	61.06	61.73 (6.32)	46.83 (7.28)	58.65 (6.10)	41.63 (22.5)	53.37 (1.03)	53.17 (0.92)	**68.17 (8.48)**
Thyroid	93.61	93.67 (1.25)	91.88 (1.97)	93.29 (1.71)	92.51 (2.79)	89.22 (2.73)	88.79 (2.82)	**94.02 (1.60)**
Vehicle	31.30	32.16 (3.87)	24.53 (4.66)	30.28 (3.92)	29.55 (3.85)	27.51 (1.69)	26.30 (1.23)	**35.43 (2.28)***
Vertebral	81.33	81.53 (4.35)	71.67 (5.13)	**82.53 (3.32)**	78.00 (6.06)	79.27 (4.47)	78.47 (3.80)	80.33 (3.60)
WBC	67.78	77.36 (14.6)	84.86 (3.70)	77.75 (14.1)	70.32 (17.3)	89.15 (3.16)	88.59 (3.72)	**93.13 (2.19)***
WDVG	79.86	79.92 (1.13)	67.81 (3.77)	79.18 (1.21)	78.37 (1.35)	71.07 (1.12)	69.66 (1.05)	**82.32 (1.11)**
Weaning	75.60	76.87 (4.14)	60.40 (5.50)	75.67 (5.21)	59.13 (16.8)	66.93 (4.98)	64.47 (4.45)	**81.67 (4.48)***
Wine	34.43	33.52 (3.44)	45.23 (14.1)	34.32 (4.71)	32.61 (2.41)	32.84 (1.13)	32.84 (1.13)	**55.68 (11.0)***

Visando comparar o comportamento das abordagens foi realizado o teste de Kruskal-Wallis com uma confiança de 95% e grau de liberdade igual a 6 (uma vez que foram comparados 7 métodos). Para 28 dos 30 problemas a hipótese nula foi rejeitada, indicando que há diferença significativa entre as acurácias das estratégias. As exceções foram as bases Blood e Haberman, onde o comportamento dos métodos não apresentou diferença suficiente para refutar a hipótese alternativa. Para os problemas em que rejeitou-se H_0, as diferenças significativas são destacas na Tabela 5.6 pelo símbolo "*".

Para analisar o comportamento de cada técnica perante as demais, efetuou-se o teste post-hoc de Nemenyi para esboçar os ranking dos algoritmos em todos os problemas. Os resultados são apresentados na Figura 5.9. É possível notar que a abordagem proposta alcançou a melhor posição geral do ranking. Entretanto, a diferença até o OLA é menor do que a distância crítica.

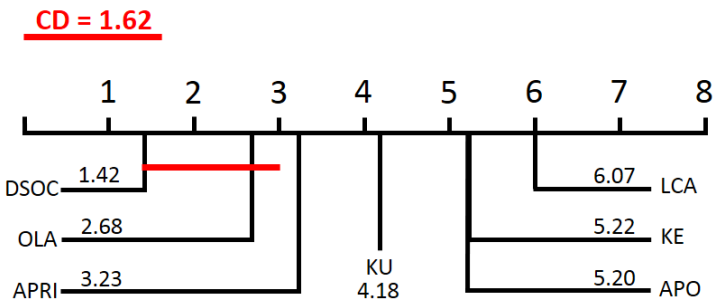

Figura 5.9: Representação gráfica do teste de Nemenyi comparando todos os métodos. Os valores apresentados próximos dos nomes dos métodos correspondem ao seu ranking médio considerando os 30 problemas de classificação.

5.4 Análise adicional dos pools formados pelo AG

Os resultados apresentados na Seção 5.2 mostraram que a adoção de medidas de complexidade unidas à acurácia consiste em uma estratégia promissora para a estimação (DSOC), efetuada de forma dinâmica, da competência dos classificadores de um *pool* de acordo com as características da instância a ser classificada. Os experimentos descritos, entretanto, basearam-se na avaliação do método sobre um *pool* criado de forma aleatória, sem um direcionamento específico, uma vez que busca-se obter diversidade entre os classificadores gerados de forma implícita. Dessa forma, uma hipótese pertinente foi avaliar se, empregando-se o método de seleção proposto sobre um conjunto de classificadores treinados em subconjuntos gerados com foco na exploração da complexidade e acurácia, o desempenho poderia atingir taxas de reconhecimento maiores.

Visando analisar esta possibilidade, foram comparados os desempenhos do DSOC trabalhando sobre *pools* formados de forma aleatória e também sobre conjuntos de classificadores treinados nos subconjuntos que melhor explorassem o espaço de complexidade (AG proposto), buscando assim, mais diversidade neste espaço entre os elementos componentes do *pool*.

Os valores apresentados na Tabela 5.4 mostraram que houve aumento na cobertura do espaço de complexidade empregando-se o AG para evoluir o conjunto gerado de forma aleatória. Os conjuntos obtidos apresentam certo aumento na diversidade no espaço de complexidade, como demonstrado nas Figuras 5.2, 5.3 e 5.4. Esperava-se então que aplicando-se o método DSOC sobre este conjunto, o desempenho obtido fosse superior ao observado quando fossem empregados *pools* gerados randomicamente.

A Tabela 5.7 apresenta os resultados desta comparação para os trinta problemas em estudo. Os valores correspondem ao desempenho médio das duas estratégias ao longo de 20 repetições. Entre parênteses são apresentados os desvios padrão calculados. Os valores em negrito indicam o melhor desempenho para cada problema.

Os valores apresentados na Tabela 5.7 evidenciam o aumento no desempenho do método de seleção quando foram empregados os classificadores treinados nos subconjuntos gerados pelo algoritmo genético proposto. Para corroborar tal afirmação aplicou-se o teste de Wilcoxon com 5% de significância comparando os resultados das estratégias. Diferenças significativas apresentam o marcador "*".

Observando-se os resultados alcançados, confirma-se a hipótese de que se aplicássemos o DSOC sobre um conjunto de classificadores treinados em conjuntos que foram construídos de forma a explorar o espaço de complexidade, o desempenho alcançado poderia ser superior à adoção de tal estratégia de seleção em um cenário onde a geração do *pool* foi realizada de forma aleatória.

As acurácias alcançadas indicam que a adoção de um *pool* gerado de forma a melhor cobrir o espaço de complexidade pôde contribuir para o aumento na acurácia do método DSOC, indicando que, com os classificadores mais dispersos no espaço de complexidade, no momento da seleção, foi possível escolher classificadores mais adequados para cada uma das instâncias, tomando como um dos critérios, a semelhança da complexidade da vizinhança do novo padrão com a do conjunto em que o classificador foi treinado.

A avaliação do SMC proposto mostrou que o desempenho alcançado com a adoção de informações de complexidade nas etapas de geração e seleção foi superior aos métodos de seleção presentes na literatura quando estes adotaram classificadores treinados em subconjuntos gerados aleatoriamente (conforme apresentado na Tabela 5.6). No entanto, para avaliar se a abordagem de seleção dinâmica proposta consegue tirar maior proveito

Tabela 5.7: Comparação do desempenho do método DSOC trabalhando sobre os *pools* obtidos pelo AG e randomicamente.

Dataset	DSOC + AG	DSOC + Randômica
Adult	**86.80 (2.24)**	86.77 (2.65)
Banana	**87.20 (1.62)***	82.17 (1.64)
Blood	**76.12 (0.26)***	73.98 (2.8)
CTG	**87.50 (1.25)***	85.68 (1.14)
Diabetes	**68.41 (3.93)***	66.69 (3.29)
Ecoli	**75.00 (2.44)***	69.29 (3.23)
Faults	**65.02 (1.58)***	48.68 (3.17)
German	**73.98 (3.28)***	71.88 (2.66)
Glass	**59.25 (5.22)***	53.11 (8.03)
Haberman	**74.61 (2.43)**	74.14 (3.18)
Heart	**83.58 (3.24)**	83.43 (2.79)
ILPD	**66.86 (2.73)**	64.97 (4.10)
Image	**51.03 (1.11)***	38.30 (1.58)
Ionosphere	**86.53 (4.23)**	86.08 (5.12)
Laryngeal1	**82.45 (4.51)**	80.66 (4.81)
Laryngeal3	**68.75 (5.04)***	65.45 (6.68)
Lithuanian	**82.47 (2.55)***	74.86 (3.00)
Liver	**61.86 (5.39)***	59.36 (5.05)
Magic	**79,99 (0,73)**	78,46 (0,61)
Mammo	80.99 (2.23)	**81.04 (2.48)**
Monk	**85.42 (3.44)***	82.69 (2.9)
Phoneme	**79.00 (1.04)***	76.61 (1.2)
Sonar	**68.17 (8.48)**	67.40 (7.44)
Thyroid	**94.02 (1.60)***	89.10 (2.73)
Vehicle	**35.43 (2.28)***	33.25 (2.21)
Vertebral	**80.33 (3.60)***	77.40 (4.14)
WBC	**93.13 (2.19)**	92.75 (1.88)
WDVG	**82.32 (1.11)***	75.54 (2.38)
Weaning	**81.67 (4.48)**	80.67 (3.83)
Wine	**55.68 (10.9)**	49.77 (12.8)

dos *pools* de classificadores treinados em subconjuntos gerados de forma a obter diversidade em termos de complexidade do que as soluções dinâmicas concorrentes, realizou-se a comparação do desempenho de todas as estratégias dinâmicas, individuais e de *ensembles*, com o DSOC em um cenário onde todos os métodos adotaram os mesmos *pools* formados pelo AG proposto. Os resultados obtidos são apresentados na Tabela 5.8. Os valores em negrito indicam o melhor desempenho para cada problema.

Tabela 5.8: Comparação do SMC proposto com os métodos de seleção dinâmica OLA, LCA, A Priori (APRI), A Posteriori (APOS), KNORA-Union (KU), KNORA-Eliminate (KE). Cenário em que todos adotaram os *pools* gerados pelo AG proposto. Os resultados apresentados correspondem aos valores médios e desvios padrão das 20 repetições executadas. Os melhores valores são apresentados em negrito.

	OLA	LCA	APRI	APOS	KU	KE	DSOC
Adult	84.77 (2.80)	83.52 (2.87)	85.73 (2.78)	85.20 (2.91)	84.24 (1.85)	83.72 (1.92)	**86.80 (2.24)**
Banana	84.87 (1.68)	84.68 (1.79)	84.57 (1.87)	83.83 (1.76)	**87.42 (1.12)**	85.06 (1.45)	87.20 (1.62)
Blood	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**	**76.12 (0.26)**
CTG	86.70 (1.04)	81.33 (0.82)	86.76 (0.76)	86.30 (0.85)	85.01 (0.93)	84.21 (0.89)	**87.50 (1.25)***
Diabetes	65.00 (2.27)	63.75 (2.89)	64.51 (2.46)	64.41 (2.76)	64.64 (0.87)	64.79 (1.11)	**68.41 (3.93)***
Ecoli	62.38 (3.09)	52.72 (5.91)	64.46 (4.83)	62.38 (3.94)	54.44 (2.77)	52.26 (1.99)	**75.00 (2.44)***
Faults	56.82 (2.99)	38.84 (10.1)	55.70 (3.42)	54.73 (3.66)	44.96 (2.80)	39.25 (2.27)	**65.02 (1.58)***
German	73.48 (2.42)	65.88 (2.98)	72.80 (3.08)	71.52 (2.85)	72.94 (1.36)	72.28 (1.15)	**73.98 (3.28)**
Glass	51.89 (7.46)	25.47 (10.2)	53.11 (6.09)	30.75 (16.9)	51.98 (5.15)	47.45 (4.71)	**59.25 (5.22)***
Haberman	74.14 (0.75)	74.41 (1.74)	74.41 (1.35)	73.29 (3.00)	73.68 (0.00)	73.68 (0.00)	**74.61 (2.43)**
Heart	80.37 (4.12)	72.84 (5.31)	80.67 (3.73)	77.31 (4.10)	77.16 (4.50)	76.19 (4.31)	**83.58 (3.24)***
ILPD	68.62 (2.91)	63.79 (3.82)	70.34 (2.97)	68.41 (3.68)	**71.9 (0.75)**	71.83 (0.45)	66.86 (2.73)
Image	41.25 (2.12)	32.82 (3.87)	41.33 (1.99)	41.42 (2.34)	36.35 (1.52)	35.88 (1.48)	**51.03 (1.11)***
Ionosphere	81.48 (4.56)	72.10 (4.14)	80.72 (4.14)	81.02 (4.36)	**87.73 (3.40)**	87.50 (3.17)	86.53 (4.23)
Laryngeal1	78.87 (5.08)	73.49 (5.04)	80.00 (3.89)	77.08 (5.11)	74.25 (3.79)	72.74 (4.07)	**82.45 (4.51)**
Laryngeal3	67.22 (4.90)	58.81 (6.25)	66.14 (3.23)	63.92 (4.43)	61.02 (3.43)	59.37 (3.25)	**68.75 (5.04)**
Lithuanian	67.06 (3.19)	69.03 (2.40)	67.49 (3.21)	65.15 (3.16)	59.87 (1.40)	58.04 (1.21)	**82.47 (2.55)***
Liver	60.17 (3.14)	51.34 (5.15)	57.38 (4.96)	54.24 (8.20)	57.09 (6.77)	55.17 (4.58)	**61.86 (5.39)**
Magic	79.00 (0.56)	78.23 (0.60)	78.78 (0.53)	78.53 (0.63)	78.92 (0.58)	78.80 (0.58)	**79.99 (0.73)***
Mammo	80.56 (2.96)	77.97 (3.08)	80.51 (3.44)	79.15 (3.44)	79.13 (2.93)	78.77 (2.76)	**80.99 (2.23)**
Monk	82.59 (3.22)	72.73 (3.81)	81.48 (4.86)	68.19 (14.5)	78.84 (3.67)	79.86 (3.08)	**85.42 (3.44)***
Phoneme	77.22 (0.82)	76.77 (0.86)	77.14 (0.98)	77.05 (0.96)	75.74 (0.81)	75.17 (0.74)	**79.00 (1.04)***
Sonar	60.29 (5.38)	45.96 (5.73)	61.44 (5.94)	45.96 (20.8)	55.00 (1.86)	53.85 (1.36)	**68.17 (8.48)***
Thyroid	93.73 (1.45)	91.99 (1.71)	93.41 (1.92)	92.02 (1.97)	88.70 (4.11)	88.38 (4.05)	**94.02 (1.60)**
Vehicle	32.44 (2.65)	23.96 (4.74)	32.04 (3.33)	31.11 (4.01)	28.22 (2.57)	26.75 (1.31)	**35.43 (2.28)***
Vertebral	80.73 (4.70)	73.20 (5.20)	**81.67 (5.07)**	77.87 (8.13)	80.47 (4.64)	79.67 (4.58)	80.33 (3.60)
WBC	77.57 (13.5)	84.89 (3.25)	79.79 (12.2)	79.01 (10.6)	88.49 (2.96)	88.27 (3.47)	**93.13 (2.19)***
WDVG	80.24 (0.88)	69.91 (3.41)	80.08 (1.19)	79.49 (1.33)	72.71 (1.15)	71.60 (1.18)	**82.32 (1.11)***
Weaning	77.27 (4.43)	63.33 (6.57)	78.40 (4.19)	64.93 (19.9)	71.07 (3.96)	68.47 (3.58)	**81.67 (4.48)**
Wine	33.52 (2.95)	43.86 (13.3)	35.68 (7.58)	33.41 (2.88)	32.84 (1.13)	32.84 (1.13)	**55.68 (10.98)***

Visando comparar o comportamento das abordagens foi realizado o teste de Kruskal-Wallis com 95% de confiança e grau de liberdade 6. Para apenas 2 dos 30 problemas a hipótese nula foi confirmada (bases Blood e Haberman), indicando que não há diferença significativa entre as acurácias das estratégias. Para todos os 28 problemas restantes, obsevou-se diferença suficiente para refutar a hipótese alternativa. Para os problemas em que rejeitou-se H_0, as diferenças significativas são destacas na Tabela 5.8 pelo símbolo "*".

Os valores apresentados mostram que mesmo quando as soluções da literatura empregaram *pools* gerados pelo método proposto o desempenho alcançado pelo SMC foi superior em 26 dos 30 problemas testados. A performance dos métodos de seleção reflete o princípio de que uma vez que o DSOC adota critérios de complexidade no momento de estimar a competência dos classificadores ele conseguiria ter um melhor desempenho em cenários em que os subconjuntos usados para treinar os classificadores fossem mais dispersos no espaço de complexidade.

Visando avaliar o comportamento de cada técnica em relação às demais, efetuou-se o teste post-hoc de Nemenyi para esboçar o ranking dos algoritmos em todos os 30 problemas. Os resultados são apresentados na Figura 5.10. É possível notar que o SMC proposto alcançou a melhor posição geral do ranking. Novamente a diferença do DSOC até o OLA é menor do que a distância crítica. No entanto, se compararmos o desempenho do método proposto nesta representação com àquela observada na Figura 5.6 (em que todas as estratégias empregaram *pools* gerados aleatoriamente) nota-se o aumento da disparidade em favor da estratégia apresentada neste trabalho. O mesmo comportamento é observado em relação à Figura 5.9 (em que nosso SMC baseou-se nos *pools* gerados pelo AG enquanto as demais técnicas nos *pools* formados aleatoriamente). Essa melhora no desempenho em relação às demais ténicas mostra que o DSOC consegue aproveitar melhor as características dos classificadores quando estes são treinados em subconjuntos mais bem distribuídos no espaço de complexidade.

5.5 Considerações Finais

Neste capítulo foram apresentados os experimentos e análises realizadas com o intuito de avaliar as soluções propostas. Inicialmente foram analisados o algoritmo genético construído para a etapa de geração e o método de seleção dinâmica proposto (DSOC) de forma individual, avaliando se atendiam aos propósitos a que se propunham. Em seguida, visando avaliar o comportamento do processo como um todo empregamos as duas estratégias propostas de forma conjunta, unindo a geração e seleção com base em critérios

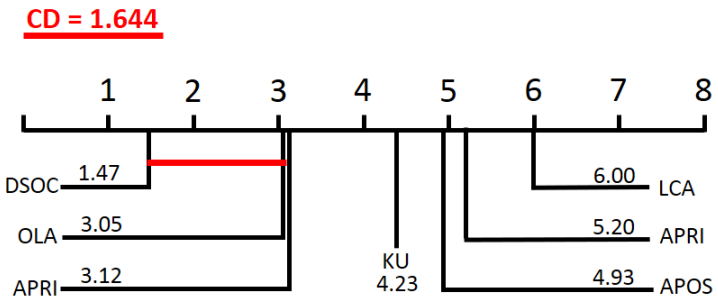

Figura 5.10: Representação gráfica do teste de Nemenyi comparando o desempenho de todos os métodos adotando-se os *pools* gerados pelo AG proposto. Os valores apresentados próximos dos nomes dos métodos correspondem ao seu ranking médio considerando os 30 problemas de classificação.

de complexidade do problema.

Respondendo a primeira pergunta da pesquisa, nós podemos dizer que o uso de características de complexidade dos dados na geração dos *pools* e na seleção dos classificadores mostrou uma interessante contribuição em termos de desempenho no processo de classificação. O método DSOC pôde obter o melhor resultado em 165 dos 180 experimentos (91.67%) quando comparados às outras 6 estratégias dinâmicas de seleção. É possível concluir que a estratégia de seleção proposta se beneficia com a melhor cobertura do espaço de complexidade do problema provida pelo método de geração baseado no AG.

Com base nos experimentos realizados para analisar a geração dos *pools*, é possível responder a segunda questão da pesquisa. Observou-se um impacto positivo no desempenho da classificação quando informações da complexidade do problema foram usadas para orientar a geração dos *pools* de classificadores. Na comparação com os métodos de seleção dinâmica, foi possível observar que em 126 dos 180 experimentos realizados (70.00%) a adoção do AG proposto para gerar o *pool* permitiu melhorar a acurácia da classificação. Além disso, foram observados ganhos quando comparadas a combinação de todos, onde a solução proposta obteve ganho em 63.33% dos casos (19 dos 30 problemas) e no *single best*, onde houve melhora na acurácia em 17 dos 30 problemas (56.67%).

Para verificar a cobertura do espaço de complexidade do problema, para cada uma das bases, foi calcula a dispersão média de complexidade de ambos os *pools*, aqueles gerados pelo método baseado no AG e naqueles gerados de forma aleatória. Os valores obtidos mostraram que houve um aumento significativo na cobertura do espaço de complexidade dos problemas, o que mostra que os subconjuntos formados pelo AG são mais bem espalhados no espaço de complexidade, gerando uma maior diversidade em termos de dificuldade para cada problema em estudo, respondendo assim a terceira pergunta desta pesquisa.

Os experimentos realizados para avaliar o método de seleção com base na complexidade do problema (DSOC) serviram para respondermos a última questão da pesquisa. Os resultados apresentados na Seção 5.2 mostraram que quando a seleção dinâmica dos classificadores leva em conta informações de complexidade as taxas de reconhecimento podem ser mais altas em relação às técnicas da literatura usadas na comparação. Tal fato indica que a adoção de descritores de complexidade do problema em estudo podem, em conjunto com acurácia, ser um bom critério para estimação da competência dos classificadores.

Capítulo 6

Conclusões

Tendo ciência da relação existente entre o comportamento dos classificadores e os conjuntos sobre os quais eles foram treinados, neste trabalho foi desenvolvido um sistema de múltiplos classificadores que considerou informações de complexidade do problema em estudo no momento da geração dos *pools* e da seleção dos classificadores. O objetivo foi avaliar se tais informações poderiam contribuir em cada uma das etapas do processo e no próprio SMC como um todo.

Para a etapa de geração dos subconjuntos destinados ao treinamento dos classificadores foi desenvolvido um algoritmo genético cujo objetivo foi otimizar uma função que combina a exploração do espaço de complexidade em conjunto com acurácia. A ideia foi construir um grupo de classificadores treinados em subconjuntos cujos descritores de complexidade estivessem bem dispersos no espaço de complexidade e, além disso, fossem acurados.

De forma a avaliar a estratégia proposta, foram executados testes com 30 problemas provenientes de diferentes repositórios. Comparou-se os desempenhos de técnicas estáticas (*single best* e combinação de todos) e de seleção dinâmica (OLA, LCA, A Priori, A Posteriori, KNORA-U e KNORA-E) usando um conjunto de classificadores gerados de forma randômica com reposição e empregando também conjuntos de classificadores construídos pela estratégia proposta.

Os experimentos mostraram que a técnica evolutiva construída foi capaz de gerar conjuntos de classificadores mais acurados para cenários estáticos (*single best* e combinação de todos) e dinâmicos, tanto para classificadores individuais (LCA, OLA, A Priori e A Posteriori) como para *ensembles* (KNORAS Union e Eliminate). Os resultados mostraram que houve ganho em termos de acurácia em 70.00% dos cenários dinâmicos, melhora na acurácia para o *single best* em 56.67% dos casos e para a combinação de todos, um aumento na taxa de acerto em 63.33% dos problemas testados.

Além da acurácia, foi avaliada a dispersão do *pool* gerado no espaço de complexidade. Observou-se que os subconjuntos gerados pelo AG proposto tiveram uma cobertura do espaço de complexidade significativamente maior do que se tivessem sido gerados de forma aleatória. Esse aumento deu-se pelo fato de priorizarmos subconjuntos mais distantes das regiões de concentração, favorecendo assim o aumento da diversidade em termos de complexidade do grupo todo. Observou-se melhora na distribuição dos subconjuntos no espaço de complexidade em 29 dos 30 problemas testados e, além disso, constatou-se um aumento estatisticamente significativo, em termos de distribuição, em 22 casos.

Os resultados observados confirmam a hipótese de que a adoção de informações de complexidade do problema em estudo no momento da geração dos *pools* poderia obter uma maior cobertura do espaço de complexidade, provendo assim um conjunto de classificadores mais dispersos neste espaço. Além disso, observou-se que a adoção destes conjuntos contribuiu com a melhora na acurácia dos métodos testados.

Para a etapa de seleção implementou-se um *framework* que emprega informações de complexidade e acurácia para estimar a competência dos classificadores presentes no *pool*. Neste contexto, foram combinados três critérios: a acurácia local de cada classificador, a similaridade de sua assinatura de complexidade (calculada com base no subconjunto onde foi treinado) e a distância da instância de teste até o centróide da classe predita. A ideia foi selecionar o classificador que houvesse treinado no conjunto mais similar à vizinhança da instância de teste e, que ao mesmo tempo, fosse acurado nesta região.

Com o objetivo de avaliar o método de seleção implementado, comparou-se seu desempenho com diversas técnicas de seleção dinâmica estabelecidas na literatura no mesmo grupo de 30 problemas apresentado na etapa de geração. A avaliação deu-se em duas oportunidades. Na primeira, todos os métodos de seleção basearam-se em classificadores gerados randomicamente, enquanto, na segunda, todas as abordagens usaram classificadores treinados em subconjuntos gerados pelo algoritmo genético, ou seja, conjuntos que possuíam maior diversidade em termos de complexidade.

Nos dois cenários testados o DSOC atingiu taxas de reconhecimento mais altas que seus concorrentes. No primeiro, a estratégia conseguiu superar as demais abordagens dinâmicas em 82.00% dos casos, o *single best* em 28 dos 30 problemas e vencer a combinação de todos em 23 dos 30 cenários. No segundo cenário, o DSOC alcançou taxas de reconhecimento maiores que os métodos concorrentes em 87.78% dos casos. Com base nos valores pode-se confirmar a segunda hipótese desta pesquisa, uma vez que foi demonstrado que a adoção de critérios de complexidade contribuiu para a estimação da competência dos classificadores de um *pool*.

Entretanto, analisando o comportamento das estratégias, observa-se que a seleção

baseada em critérios de complexidade do problema consegue uma maior margem de vantagem quando adota os *pools* gerados de forma a obter maior dispersão no espaço de complexidade. Essa melhora no desempenho do DSOC deve-se ao fato de que os *pools* gerados pelo AG apresentam uma diversidade maior de classificadores em termos de dificuldade, uma vez que os subconjuntos sobre os quais eles foram treinados estão melhor distribuídos no espaço de complexidade.

Por fim, analisou-se o desempenho do SMC completo, em que as etapas de geração e seleção consideraram informações da dificuldade do problema em estudo. De forma a validar o método, comparou-se sua performance perante 6 estratégias estabelecidas na literatura, as quais utilizaram uma estratégia randômica, sem considerar o espaço de complexidade dos dados. A avaliação foi realizada sobre o mesmo conjunto de problemas que foi usado na avaliação das etapas de geração e seleção individualmente.

Os valores alcançados demonstram superioridade do SMC proposto em grande parte dos problemas estudados (91.67%), mostrando que a adoção de informações de complexidade no processo de formação dos *pools* e no momento da seleção podem contribuir para a melhora no desempenho dos sistemas de reconhecimento, independente de aplicação.

Com a validação dos métodos propostos nota-se que o *framework* construído é robusto, visto que consegue alcançar taxas de reconhecimento superiores à estratégias já estabelecidas na literatura em muitos cenários. A técnica de geração pôde contribuir para a melhora na acurácia nos métodos estáticos ou dinâmicos, além de contribuir na exploração do espaço de complexidade. A proposta de seleção conseguiu taxas de acerto maiores trabalhando com *pools* gerados randomicamente ou orientados pelas medidas de complexidade, além de conseguir tirar maior proveito dos *pools* gerados pelo AG proposto, em relação às tecnicas comparadas. O SMC como um *framework* completo pôde gerar classificadores mais bem distribuídos no espaço de complexidade e aproveitar tal fato para tornar o processo de seleção mais eficiente, alcançando maiores taxas de acerto em 26 dos 30 problemas testados.

Apesar dos interessantes resultados observados, pode ser interessante estudar variações dos parâmetros adotados nos experimentos do SMC, como avaliar o comportamento de diferentes indutores base ou uma estratégia evolutiva distinta da adotada. Além disso, pode ser proveitoso investigar outros descritores de complexidade que possam melhor representar a complexidade do problema.

Um viés que merece atenção tange à geração dos subconjuntos usados no treinamento dos classificadores. Seria interessante analisar o comportamento de *pools* gerados por outras técnicas, tal como o Boosting e RSS em relação à estratégia proposta neste tra-

balho, ou mesmo explorar a possibilidade de se treinar os classificadores em subconjuntos que tenham variações na quantidade de instâncias que os compõe.

Referências Bibliográficas

AKSELA, M. Comparison of classifier selection methods for improving committee performance. In: WINDEATT, T.; ROLI, F. (Ed.). *Multiple Classifier Systems*. Springer Berlin Heidelberg, 2003, (Lecture Notes in Computer Science, v. 2709). p. 84–93. ISBN 978-3-540-40369-2. Disponível em: <http://dx.doi.org/10.1007/3-540-44938-8_9>.

ALCALá-FDEZ, J. et al. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, v. 17, n. 2-3, p. 255–287, 2011. Cited By 275.

AYAD, O.; SYED-MOUCHAWEH, M. Multiple classifiers approach based on dynamic selection to maximize classification performance. *International Journal of Machine Learning and Computing*, v. 1, n. 12, p. 154–162, 2011.

BACHE, K.; LICHMAN, M. *UCI Machine Learning Repository*. 2013. Disponível em: <http://archive.ics.uci.edu/ml>.

BREIMAN, L. Bagging predictors. *Machine Learning*, Kluwer Academic Publishers-Plenum Publishers, v. 24, n. 2, p. 123 – 140, 1996. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/A%3A1018054314350>.

BRITTO JR., A. S.; SABOURIN, R.; OLIVEIRA, L. E. S. Dynamic selection of classifiers - a comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665 – 3680, 2014. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320314001885>.

BROWN, G. et al. Diversity creation methods: A survey and categorisation. *Journal of Information Fusion*, v. 6, p. 5–20, 2005.

CANO, J.-R. Analysis of data complexity measures for classification. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 40, n. 12, p. 4820–4831, set. 2013. ISSN 0957-4174. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2013.02.025>.

CAVALCANTI, G.; REN, T.; VALE, B. Data complexity measures and nearest neighbor classifiers: A practical analysis for meta-learning. In: *Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference on*. [S.l.: s.n.], 2012. v. 1, p. 1065–1069. ISSN 1082-3409.

CAVALIN, P.; SABOURIN, R.; SUEN, C. Dynamic selection approaches for multiple classifier systems. *Neural Computing and Applications*, Springer-Verlag, v. 22, n. 3-4, p. 673–688, 2013. ISSN 0941-0643. Disponível em: <http://dx.doi.org/10.1007/s00521-011-0737-9>.

CORRIVEAU, G. et al. Review and study of genotypic diversity measures for real-coded representations. *Evolutionary Computation, IEEE Transactions on*, v. 16, n. 5, p. 695–710, Oct 2012. ISSN 1089-778X.

CRUZ, R.; CAVALCANTI, G. D. C.; REN, T. I. A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. In: *Neural Networks (IJCNN), The 2011 International Joint Conference on*. [S.l.: s.n.], 2011. p. 1126–1133. ISSN 2161-4393.

CRUZ, R. M. et al. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 48, n. 5, p. 1925–1935, maio 2015. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2014.12.003>.

DIDACI, L.; GIACINTO, G. Dynamic classifier selection by adaptive k-nearest-neighbourhood rule. In: ROLI, F.; KITTLER, J.; WINDEATT, T. (Ed.). *Multiple Classifier Systems*. [S.l.]: Springer Berlin Heidelberg, 2004, (Lecture Notes in Computer Science, v. 3077). p. 174–183. ISBN 978-3-540-22144-9.

DIDACI, L. et al. A study on the performances of dynamic classifier selection based on local accuracy estimation. *Pattern Recognition*, v. 38, n. 11, p. 2188 – 2191, 2005. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320305000956>.

DIETTERICH, T. G. Ensemble methods in machine learning. In: *Proceedings of the First International Workshop on Multiple Classifier Systems*. London, UK, UK: Springer-Verlag, 2000. (MCS '00), p. 1–15. ISBN 3-540-67704-6. Disponível em: <http://dl.acm.org/citation.cfm?id=648054.743935>.

DOS SANTOS, E.; SABOURIN, R.; MAUPIN, P. Ambiguity-guided dynamic selection of ensemble of classifiers. In: *Information Fusion, 2007 10th International Conference on*. [S.l.: s.n.], 2007. p. 1–8.

DOS SANTOS, E. M.; SABOURIN, R.; MAUPIN, P. A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition*, Elsevier Science Inc., New York, NY, USA, v. 41, n. 10, p. 2993–3009, out. 2008. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2008.03.027>.

FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: *Proceedings of the 13th International Conference on Machine Learning*. [S.l.: s.n.], 1996. p. 148–156.

GIACINTO, G.; ROLI, F. Methods for dynamic classifier selection. In: *Proceedings of the 10th International Conference on Image Analysis and Processing*. Washington, DC, USA: IEEE Computer Society, 1999. (ICIAP '99), p. 659–. ISBN 0-7695-0040-4. Disponível em: <http://dl.acm.org/citation.cfm?id=839281.840806>.

GIACINTO, G.; ROLI, F.; FUMERA, G. Selection of classifiers based on multiple classifier behaviour. In: FERRI, F. et al. (Ed.). *Advances in Pattern Recognition*. Springer Berlin Heidelberg, 2000, (Lecture Notes in Computer Science, v. 1876). p. 87–93. ISBN 978-3-540-67946-2. Disponível em: <http://dx.doi.org/10.1007/3-540-44522-6_9>.

GUNES, V. et al. Combination, cooperation and selection of classifiers: a state of the art. *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific Publishing Company, v. 17, n. 8, p. 1303–1324, 2003.

HA, T. M.; ZIMMERMANN, M.; BUNKE, H. Off-line handwritten numeral string recognition by combining segmentation-based and segmentation-free methods. *Pattern Recognition*, v. 31, n. 3, p. 257–272, 1998.

HERNáNDEZ-REYES, E.; CARRASCO-OCHOA, J. A.; MARTÍNEZ-TRINIDAD, J. F. Classifier selection based on data complexity measures. In: *Proceedings of the 10th Iberoamerican Congress Conference on Progress in Pattern Recognition, Image Analysis and Applications*. Berlin, Heidelberg: Springer-Verlag, 2005. (CIARP'05), p. 586–592. ISBN 3-540-29850-9, 978-3-540-29850-2. Disponível em: <http://dx.doi.org/10.1007/11578079_61>.

HO, T.; BASU, M.; LAW, M. Measures of geometrical complexity in classification problems. In: BASU, M.; HO, T. (Ed.). *Data Complexity in Pattern Recognition*. Springer

London, 2006, (Advanced Information and Knowledge Processing). p. 1–23. ISBN 978-1-84628-171-6. Disponível em: <http://dx.doi.org/10.1007/978-1-84628-172-3_1>.

HO, T. K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 20, n. 8, p. 832–844, ago. 1998. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.709601>.

HO, T. K.; BASU, M. Measuring the complexity of classification problems. In: *Pattern Recognition, 2000. Proceedings. 15th International Conference on*. [S.l.: s.n.], 2000. v. 2, p. 43–47 vol.2. ISSN 1051-4651.

HO, T. K.; BASU, M. Complexity measures of supervised classification problems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 24, n. 3, p. 289–300, Mar 2002. ISSN 0162-8828.

HOEKSTRA, A.; DUIN, R. P. W. On the nonlinearity of pattern classifiers. *Pattern Recognition, International Conference on*, IEEE Computer Society, Los Alamitos, CA, USA, v. 4, p. 271, 1996. ISSN 1051-4651.

IVAKHNENKO, A. G. Heuristic self-organization in problems of engineering cybernetics. *Automatica*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 6, n. 2, p. 207–219, march 1970. ISSN 0005-1098. Disponível em: <http://dx.doi.org/10.1016/0005-1098(70)90092-0>.

JAIN, A.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 22, n. 1, p. 4–37, 2000. ISSN 0162-8828.

KING, R. D.; FENG, C.; SUTHERLAND, A. *StatLog: Comparison of Classification Algorithms on Large Real-World Problems*. 1995.

KITTLER, J. et al. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 20, n. 3, p. 226–239, 1998. ISSN 0162-8828.

KO, A. H.; SABOURIN, R.; BRITTO JR., A. S. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, v. 41, n. 5, p. 1718 – 1731, 2008. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320307004499>.

KUMAR, G.; KUMAR, K. The use of artificial-intelligence-based ensembles for intrusion detection: A review. *Appl. Comp. Intell. Soft Comput.*, Hindawi Publishing Corp., New York, NY, United States, v. 2012, p. 21:21–21:21, jan. 2012. ISSN 1687-9724. Disponível em: <http://dx.doi.org/10.1155/2012/850160>.

KUNCHEVA, L. *StatLog: Comparison of Classification Algorithms on Large Real-World Problems*. 2004. Disponível em: <http://pages.bangor.ac.uk/ mas00a/activities/real_data.htm>.

KUNCHEVA, L.; RODRIGUEZ, J. Classifier ensembles with a random linear oracle. *Knowledge and Data Engineering, IEEE Transactions on*, v. 19, n. 4, p. 500–508, 2007. ISSN 1041-4347.

KUNCHEVA, L. et al. Complexity of data subsets generated by the random subspace method: An experimental investigation. In: *Transportation Research Board, Special Report*. [S.l.]: Springer-Verlag, 2001. p. 349–358.

KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, Kluwer Academic Publishers, v. 51, n. 2, p. 181–207, 2003. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/A%3A1022859003006>.

LAM, L. Classifier combinations: Implementations and theoretical issues. In: *Multiple Classifier Systems*. Springer Berlin Heidelberg, 2000, (Lecture Notes in Computer Science, v. 1857). p. 77–86. ISBN 978-3-540-67704-8. Disponível em: <http://dx.doi.org/10.1007/3-540-45014-9_7>.

LANDEROS, A. I. *Data complexity and classifier selection*. 2008. 180 p. Copyright - Copyright ProQuest, UMI Dissertations Publishing 2008; Última atualização em - 2014-01-20; Primeira página - n/a; M3: Ph.D. Disponível em: <http://search.proquest.com/docview/304682789?accountid=40690>.

LEBOURGEOIS, F.; FRELICOT, C. A pretopology-based supervised pattern classifier. In: *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*. [S.l.: s.n.], 1998. v. 1, p. 106–109 vol.1. ISSN 1051-4651.

LI, Z.; FANG, D. The research on speech feature representation method and distance measure method. In: *Pattern Recognition, 1988., 9th International Conference on*. [S.l.: s.n.], 1988. v. 1, p. 631–633.

LILEIKYTE, R.; TELKSNYS, L. Quality estimation methodology of speech recognition features. *Elektronika ir Elektrotechnika*, Kaunas University of Technology, Faculty of Telecommunications and Electronics, v. 110, n. 4, p. 113–116, 2011.

LORENA, A. C. et al. Analysis of complexity indices for classification problems: Cancer gene expression data. *Neurocomput.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 75, n. 1, p. 33–42, jan. 2012. ISSN 0925-2312. Disponível em: <http://dx.doi.org/10.1016/j.neucom.2011.03.054>.

LU, Y. Knowledge integration in a multiple classifier system. *Applied Intelligence*, Kluwer Academic Publishers, v. 6, n. 2, p. 75–86, 1996. ISSN 0924-669X. Disponível em: <http://dx.doi.org/10.1007/BF00117809>.

LUENGO, J.; HERRERA, F. Domains of competence of fuzzy rule based classification systems with data complexity measures: A case of study using a fuzzy hybrid genetic based machine learning method. *Fuzzy Sets Syst.*, Elsevier North-Holland, Inc., Amsterdam, The Netherlands, The Netherlands, v. 161, n. 1, p. 3–19, jan. 2010. ISSN 0165-0114. Disponível em: <http://dx.doi.org/10.1016/j.fss.2009.04.001>.

MACIà, N. et al. Learner excellence biased by data set selection: A case for data characterisation and artificial data sets. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 46, n. 3, p. 1054–1066, mar. 2013. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2012.09.022>.

MACIà, N.; ORRIOLS-PUIG, A.; BERNADó-MANSILLA, E. In search of targeted-complexity problems. In: *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*. New York, NY, USA: ACM, 2010. (GECCO '10), p. 1055–1062. ISBN 978-1-4503-0072-8. Disponível em: <http://doi.acm.org/10.1145/1830483.1830674>.

MELVILLE, P.; MOONEY, R. J. Creating diversity in ensembles using artificial data. *Journal of Information Fusion: Special Issue on Diversity in Multi Classifier Systems*, v. 6, n. 1, p. 99–111, 2004. Disponível em: <http://www.cs.utexas.edu/users/ailab/?melville:if04>.

MOLLINEDA, R. A.; SáNCHEZ, J. S.; SOTOCA, J. M. Data characterization for effective prototype selection. In: *Proceedings of the Second Iberian Conference on Pattern Recognition and Image Analysis - Volume Part II*. Berlin, Heidelberg: Springer-Verlag, 2005. (IbPRIA'05), p. 27–34. ISBN 3-540-26154-0, 978-3-540-26154-4. Disponível em: <http://dx.doi.org/10.1007/11492542_4>.

OKUN, O.; PRIISALU, H. Dataset complexity in gene expression based cancer classification using ensembles of k-nearest neighbors. *Artif. Intell. Med.*, Elsevier Science Publishers Ltd., Essex, UK, v. 45, n. 2-3, p. 151–162, fev. 2009. ISSN 0933-3657. Disponível em: <http://dx.doi.org/10.1016/j.artmed.2008.08.004>.

OKUN, O.; VALENTINI, G. Dataset complexity can help to generate accurate ensembles of k-nearest neighbors. In: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on.* [S.l.: s.n.], 2008. p. 450–457. ISSN 1098-7576.

ORRIOLS-PUIG, A.; MACIà, N.; HO, T. K. *Documentation for the Data Complexity Library in C++*. Barcelona, Spain, 2010. Disponível em: <http://dcol.sourceforge.net/>.

PANOV, P.; DEROSKI, S. Combining bagging and random subspaces to create better ensembles. In: *Advances in Intelligent Data Analysis VII.* [S.l.]: Springer Berlin Heidelberg, 2007, (Lecture Notes in Computer Science, v. 4723). p. 118–129. ISBN 978-3-540-74824-3.

PONTI JR., M. P. Combining classifiers: From the creation of ensembles to the decision fusion. In: *Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2011 24th SIBGRAPI Conference on.* [S.l.: s.n.], 2011. p. 1–10.

QUINLAN, J. R. Bagging, boosting, and c4.s. In: *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 1.* AAAI Press, 1996. (AAAI'96), p. 725–730. ISBN 0-262-51091-X. Disponível em: <http://dl.acm.org/citation.cfm?id=1892875.1892983>.

RANAWANA, R.; PALADE, V. Multi-classifier systems: Review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.*, IOS Press, Amsterdam, The Netherlands, The Netherlands, v. 3, n. 1, p. 35–61, jan. 2006. ISSN 1448-5869. Disponível em: <http://dl.acm.org/citation.cfm?id=1232855.1232859>.

SABOURIN, M. et al. Classifier combination for hand-printed digit recognition. In: *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on.* [S.l.: s.n.], 1993. p. 163–166.

SANTANA, A. et al. A dynamic classifier selection method to build ensembles using accuracy and diversity. In: *Neural Networks, 2006. SBRN '06. Ninth Brazilian Symposium on.* [S.l.: s.n.], 2006. p. 36–41.

SEEWALD, A. K. Towards a theoretical framework for ensemble classification. In: *Proceedings of the 18th international joint conference on Artificial intelligence.* San Francisco,

CA, USA: Morgan Kaufmann Publishers Inc., 2003. (IJCAI'03), p. 1443–1444. Disponível em: <http://dl.acm.org/citation.cfm?id=1630659.1630891>.

SHIPP, C. A.; KUNCHEVA, L. I. Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion*, v. 3, p. 135–148, 2002.

SINGH, S. Multiresolution estimates of classification complexity. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 25, n. 12, p. 1534–1539, dez. 2003. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/TPAMI.2003.1251146>.

SKURICHINA, M.; DUIN, R. P. W. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications*, Springer-Verlag London Limited, v. 5, n. 2, p. 121–135, 2002. ISSN 1433-7541. Disponível em: <http://dx.doi.org/10.1007/s100440200011>.

SáNCHEZ, J. S.; MOLLINEDA, R. A.; SOTOCA, J. M. An analysis of how training data complexity affects the nearest neighbor classifiers. *Pattern Anal. Appl.*, Springer-Verlag, London, UK, UK, v. 10, n. 3, p. 189–201, jul. 2007. ISSN 1433-7541. Disponível em: <http://dx.doi.org/10.1007/s10044-007-0061-2>.

SOTOCA, J. M.; MOLLINEDA, R. A.; SáNCHEZ, J. S. A meta-learning framework for pattern classification by means of data complexity measures. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, v. 10, n. 29, p. 31–38, 2006. Disponível em: <http://dblp.uni-trier.de/db/journals/aepia/aepia10.html#SotocaMS06>.

SOTOCA, J. M.; SáNCHEZ, J. S.; MOLLINEDA, R. A. A review of data complexity measures and their applicability to pattern classification problems. In: *Actas del III Taller Nacional de Minería de Dados y Aprendizaje*. [S.l.: s.n.], 2005. (TAMIDA,05), p. 77–83.

SOUTO, M. et al. Complexity measures of supervised classifications tasks: a case study for cancer gene expression data. *Neural Networks (IJCNN), The 2010 International Joint Conference on*, IEEE Computer Society, Los Alamitos, CA, USA, p. 1352–1358, 2010.

STEFANOWSKI, J. An experimental study of methods combining multiple classifiers-diversified both by feature selection and bootstrap sampling. *Issues in the Representation and Processing of Uncertain and Imprecise Information*, p. 337–354, 2005.

TSOUMAKAS, G.; PARTALAS, I.; VLAHAVAS, I. A taxonomy and short review of ensemble selection. In: *ECAI 08, Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications, SUEMA*. [S.l.: s.n.], 2008.

TUMER, K.; GHOSH, J. Error correlation and error reduction in ensemble classifiers. *Connection Science*, v. 8, n. 3-4, p. 385–403, 1996.

WINDEATT, T. Diversity measures for multiple classifier system analysis and design. *Information Fusion*, v. 6, n. 1, p. 21–36, mar. 2005. ISSN 15662535. Disponível em: <http://dx.doi.org/10.1016/j.inffus.2004.04.002>.

WOLPERT, D. H. Original contribution: Stacked generalization. *Neural Netw.*, Elsevier Science Ltd., Oxford, UK, UK, v. 5, n. 2, p. 241–259, fev. 1992. ISSN 0893-6080. Disponível em: <http://dx.doi.org/10.1016/S0893-6080(05)80023-1>.

WOODS, K.; KEGELMEYER JR., W. P.; BOWYER, K. Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 19, n. 4, p. 405–410, abr. 1997. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.588027>.

XIAO, J.; HE, C. Dynamic classifier ensemble selection based on gmdh. In: *Computational Sciences and Optimization, 2009. CSO 2009. International Joint Conference on.* [S.l.: s.n.], 2009. v. 1, p. 731–734.

YAN, Y. et al. Sorting-based dynamic classifier ensemble selection. In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on.* [S.l.: s.n.], 2013. p. 673–677. ISSN 1520-5363.

YU-QUAN, Z. et al. Dynamic weighting ensemble classifiers based on cross-validation. *Neural Computing and Applications*, Springer-Verlag, v. 20, n. 3, p. 309–317, 2011. ISSN 0941-0643. Disponível em: <http://dx.doi.org/10.1007/s00521-010-0372-x>.