

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

Uma Análise Comparativa entre Sistemas de Combinação de Classificadores com
Distribuição Vertical dos Dados

Laura Emmanuella Alves Santana

Natal
Fevereiro de 2008

Livros Grátis

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

Milhares de livros grátis para download.

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

Uma Análise Comparativa entre Sistemas de Combinação de Classificadores com
Distribuição Vertical dos Dados

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

Laura Emmanuella Alves Santana

Profa. Dra. Anne Magály de Paula Canuto
Orientador

Natal, Fevereiro de 2008

Catologação da Publicação na Fonte. UFRN / SISBI / Biblioteca Setorial
Especializada do Centro de Ciências Exatas e da Terra – CCET.

Santana, Laura Emmanuella Alves.

Uma análise comparativa entre sistemas de combinação de classificadores com distribuição vertical dos dados / Laura Emmanuella Alves Santana. – Natal, 2008.

61 f. : il.

Orientador: Profa. Dra. Anne Magály de Paula Canuto.

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

1. Distribuição vertical de dados - Dissertação. 2. Sistemas de Combinação de Classificadores - Dissertação. 3. Comitês de Classificadores - Dissertação. 4. Sistemas multiagentes - Dissertação. I. Canuto, Anne Magály de Paula. II. Título.

RN/UF/BSE-CCET

CDU: 004

Uma Análise Comparativa entre Sistemas de Combinação de Classificadores com Distribuição Vertical dos Dados

Laura Emmanuella Alves Santana

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

Profa. Dra. Anne Magály de Paula Canuto
Orientadora

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

Banca Examinadora:

Profa. Dra. Anne Magály de Paula Canuto

Prof. Dr. André Maurício Cunha Campos

Prof. Dr. André Carlos Ponde de Leon Ferreira de Carvalho

“O que sabemos é uma gota, o que ignoramos é um oceano”

– Isaac Newton

Agradecimentos

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

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

Agradeço à minha orientadora, Anne Magály de Paula Canuto, pelos conhecimentos transmitidos, pela atenção, pela confiança e amizade.

Não posso deixar de agradecer de coração a uma grande amiga que me ajudou desde a graduação, tirando minhas dúvidas e me apoiando, Márjory Cristiany da Costa Abreu.

Ao meu amor, George Gilson Sousa de Oliveira, que esteve comigo durante a escrita desta dissertação, me dando forças e sempre acreditando em mim.

A todos os amigos que de alguma forma, direta ou indireta, me ajudaram.

Por fim, agradeço a CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo apoio financeiro.

Resumo

Em sistemas que combinam as saídas de classificadores de padrões, sistemas de combinação, como comitês e sistemas multiagentes para classificação, um dos principais problemas é que os componentes do sistema (classificadores ou agentes) devem ser diversos entre si. Em outras palavras, não existe ganho de desempenho em sistemas formados por um conjunto de componentes idênticos. Um modo de aumentar a diversidade do sistema é distribuir os dados do padrão entre os classificadores que compõem o sistema. Neste trabalho será feita uma investigação sobre o impacto do uso de técnicas de distribuição de dados, mais especificamente distribuição de características, entre os componentes de sistemas de combinação de classificadores. Nesta investigação, diferentes técnicas de distribuição de características serão usadas e uma análise comparativa entre diferentes sistemas de combinação, usando diferentes configurações, será feita. Como resultado desta análise, espera-se detectar que sistemas de combinação são mais adequados para usar distribuição de características entre os componentes.

Área de Concentração: Inteligência Computacional.

Palavras-chave: Distribuição Vertical de Dados; Sistemas de Combinação de Classificadores; Comitês de Classificadores; Sistemas Multiagentes.

Abstract

In systems that combine the outputs of classification methods (combination systems), such as ensembles and multi-agent systems, one of the main constraints is that the base components (classifiers or agents) should be diverse among themselves. In other words, there is clearly no accuracy gain in a system that is composed of a set of identical base components. One way of increasing diversity is through the use of feature selection or data distribution methods in combination systems. In this work, an investigation of the impact of using data distribution methods among the components of combination systems will be performed. In this investigation, different methods of data distribution will be used and an analysis of the combination systems, using several different configurations, will be performed. As a result of this analysis, it is aimed to detect which combination systems are more suitable to use feature distribution among the components.

Area of Concentration: Computational Intelligence

Keywords: Vertical Data Distribution; Classifier Combination Systems; Ensembles; Multiagent Systems

Sumário

Introdução	1
1.1 Objetivos	3
1.2 Motivação	4
1.3 Organização do Trabalho	5
Classificação de Padrões	6
2.1 Paradigmas de Aprendizado	6
2.1.1 Paradigma Simbólico	7
2.1.2 Paradigma Estatístico	7
2.1.3 Paradigma Baseado em Exemplos	8
2.1.4 Paradigma Conexionista	8
2.2 Combinação de Classificadores	9
2.2.1 Métodos de Fusão	10
2.2.2 Métodos de Seleção DCS	11
2.3 Distribuição de Dados	12
2.4 A Questão da Diversidade	13
Sistemas Multiagentes	14
3.1 Sistema Multiagente para Classificação	15
Trabalhos Relacionados	18
4.1 Sistemas de Combinação de Classificadores	18
4.2 Distribuição de Dados em SCC	19
4.2.1 Distribuição Natural	19
4.2.2 Distribuição Randômica	20
4.2.3 Distribuição Não Randômica	20
Detalhamento da Investigação	22
5.1 Distribuição Vertical dos Dados	22
5.2 Experimentos	25

5.3	Investigação	26
5.4	Bases de Dados Utilizadas	26
Análise dos Resultados		28
6.1	Análise em relação ao número de componentes do sistema.....	29
6.1.1	Sistemas sem distribuição das características.....	29
6.1.2	Sistemas com distribuição parcial das características	31
6.1.3	Sistemas com distribuição total das características	33
6.1.4	Distribuição Parcial <i>versus</i> Distribuição Total.....	35
6.2	Análise em Relação à Técnica de Distribuição dos Dados.....	38
6.2.1	Sistemas com distribuição parcial das características	38
6.2.2	Sistemas com distribuição total das características	41
6.2.3	Distribuição Parcial <i>versus</i> Distribuição Total.....	43
6.3	Análise em relação ao uso de distribuição de dados.....	45
Considerações Finais		47
7.1	Trabalhos Futuros	49
Referências Bibliográficas		50
Apêndice A		54

Lista de Figuras

Figura 1: Modelo de um sistema multiclassificador paralelo tradicional.....	16
Figura 2: Modelo de um sistema multiagente para classificação, eliminando o módulo de combinação.....	16
Figura 3: Resultados dos sistemas de classificação sem distribuição para a base de imagem analisando a quantidade de componentes	30
Figura 4: Resultados dos sistemas de classificação sem distribuição para a base de proteínas analisando a quantidade de componentes	30
Figura 5: Resultados dos sistemas de classificação com distribuição parcial para a base de imagem analisando a quantidade de componentes	32
Figura 6: Resultados dos sistemas de classificação com distribuição parcial para a base de proteínas analisando a quantidade de componentes	33
Figura 7: Resultados dos sistemas de classificação com distribuição total para a base de imagem analisando a quantidade de componentes	34
Figura 8: Resultados dos sistemas de classificação com distribuição total para a base de proteínas analisando a quantidade de componentes	35
Figura 9: Variação média entre as distribuições parcial e total para a base de imagem, analisando a quantidade de componentes dos sistemas.....	36
Figura 10: Variação média entre as distribuições parcial e total para a base de proteínas, analisando a quantidade de componentes dos sistemas.....	37
Figura 11: Resultados dos sistemas de classificação com distribuição parcial para a base de imagem analisando as técnicas de distribuição	39
Figura 12: Resultados dos sistemas de classificação com distribuição parcial para a base de proteínas analisando as técnicas de distribuição.....	39
Figura 13: Comparação entre as técnicas de distribuição parcial.....	40
Figura 14: Resultados dos sistemas de classificação com distribuição total para a base de imagem analisando as técnicas de distribuição	41
Figura 15: Resultados dos sistemas de classificação com distribuição total para a base de proteínas analisando as técnicas de distribuição.....	42

Figura 16: Comparação entre as técnicas de distribuição total	43
Figura 17: Variação média entre as distribuições parcial e total para a base de imagem, analisando as técnicas de distribuição	44
Figura 18: Variação média entre as distribuições parcial e total para a base de proteínas, analisando as técnicas de distribuição	44
Figura 19: Comparação entre as distribuições.....	46

Lista de Tabelas

Tabela 1: Características das bases de dados utilizadas neste trabalho.....	27
Tabela 2: Acurácia e desvio padrão dos sistemas sem distribuição dos atributos.....	55
Tabela 3: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Correlação de Pearson	56
Tabela 4: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Entropia	57
Tabela 5: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Erro	58
Tabela 6: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica Random	59
Tabela 7: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica da Variância.....	60

Lista de Siglas

- *ClassAge: Classifier Agent System*
- *DCS: Dynamic Classifier Selection*
- *DCS-DT: Dynamic Classifier Selection – Decision Templates*
- *DCS-MCB: Dynamic Classifier Selection – Multiple Classifier Behaviour*
- *Fuzzy MLP: Fuzzy Multilayer Perceptron*
- *k-NN: k – Nearest Neighbors*
- *LCA: Local Class Accuracy*
- *MLP: Multilayer Perceptron*
- *NeurAge: Neural Agent System*
- *RBF: Radial Basis Function*
- *SCC: Sistema de Combinação de Classificadores*
- *SMA: Sistema Multiagente*
- *SMC: Sistema Multiclassificador*

Capítulo 1

Introdução

Aprendizado de máquina é uma área da inteligência artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado e construção de sistemas capazes de adquirir conhecimento de forma automática (Monard & Baranauskas, 2000). Os algoritmos de aprendizado de máquina podem ser supervisionados ou não supervisionados, de acordo com o conhecimento prévio que se tem sobre a classe dos exemplos (padrões) de treinamento.

A entrada para um algoritmo de aprendizado (indutor) supervisionado consiste usualmente de um conjunto de exemplos de treinamento rotulados com os valores y de uma função f desconhecida, onde $y = f(x)$. Dado esse conjunto de exemplos de treinamento, o algoritmo induz uma hipótese h que deve aproximar a verdadeira função f , tal que dados os valores x de um novo exemplo, h prediz o valor y correspondente (Lee et al., 2005). No caso dos valores y pertencerem a um conjunto discreto de classes, a tarefa de aprendizado é chamada de classificação, tratada neste trabalho.

O desempenho das hipóteses geradas, também denominadas de classificadores, depende de alguns fatores, como o domínio no qual está sendo aplicado o indutor, a quantidade de exemplos de treinamento e a relevância das características (atributos) desses exemplos. Esses fatores afetam de maneira diversa as hipóteses geradas por algoritmos de classificação diferentes. Devido a isso, alguns estudos (Lemieux & Parizeau, 2003) (Zhou, 2000) (Ben-Yacoub et al., 2001) mostraram que é possível fazer uso de vários classificadores, combinando-os num único classificador final, formando um sistema multiclassificador (SMC).

Os sistemas multiclassificadores possuem a tendência de melhorar o desempenho na classificação de exemplos não vistos durante o processo de aprendizado, mas isso só vai ocorrer se os classificadores componentes do sistema (classificadores base) possuírem erros não coincidentes, isto é, se não cometerem sempre os mesmos erros. Para ilustrar isso, (Dietterich, 1997) apresenta o seguinte exemplo: Dado um SMC com três

classificadores $\{h1, h2, h3\}$ e um novo exemplo a ser classificado, x . Se os três classificadores são idênticos, então quando $h1(x)$ está errado, $h2(x)$ e $h3(x)$ também estarão errados. Entretanto se os erros dos classificadores não são coincidentes, então quando $h1(x)$ está errado, $h2(x)$ e $h3(x)$ podem estar certos e a combinação de suas respostas pode classificar x corretamente.

Outro tipo de sistema que tem se mostrado eficiente na classificação de padrões são os sistemas multiagentes compostos por agentes classificadores. Esse tipo de sistema consegue em muitos casos resultados superiores aos dos sistemas multiclassificadores, pois seu processo de decisão é descentralizado, o que o torna mais flexível. Neste trabalho, serão estudados sistemas multiclassificadores e sistemas multiagentes para classificação, os dois tipos de sistemas serão referidos como sistemas de combinação de classificadores (SCC).

A diversidade na saída dos componentes de um SCC é vital para seu sucesso. Intuitivamente, espera-se que os membros do sistema sejam tão corretos quanto possível e, no caso de cometerem erros, que sejam diversos (Kuncheva, 2004). Essa diversidade pode ser promovida a partir da estrutura dos modelos (quantidade de classificadores, variedade de indutores); dos dados de treinamento (quantidade, relevância e forma de apresentação) e da arquitetura interna dos classificadores base.

Em relação aos dados de treinamento, estudos (Tsymbal et al., 2005) (Caragea et al., 2003) (Kargupta et al., 2000) (Opitz, 1999) (Provost & Hennessy, 1996) têm apontado para a distribuição. A distribuição dos dados de treinamento pode dar-se separando os diferentes exemplos ou os atributos. O primeiro caso é chamado de distribuição horizontal ou homogênea, formado por subconjuntos dos padrões com todos os atributos. A segunda forma de se distribuir os dados é chamada de distribuição vertical, sendo o objetivo deste estudo.

Na distribuição vertical dos dados, os classificadores são construídos a partir de um conjunto de treinamento particionado pelos atributos. Este conjunto é construído com o mesmo tamanho da base de dados original, já que é formado por todos os padrões originais da base, diferenciando apenas as características analisadas. Essa forma de distribuição dos dados também é chamada de distribuição heterogênea.

Além da busca pela diversidade no sistema, a distribuição de dados é usada em alguns casos por exigência da aplicação. Isso pode ser visto em duas situações principais: quando a distribuição é intrínseca ao problema e para manter a privacidade dos dados. O primeiro caso ocorre quando os dados estão armazenados em repositórios separados geograficamente. Isso é comum quando são colocados sensores para coleta de dados em locais diferentes. O segundo caso ocorre quando parte dos dados são sigilosos. Por exemplo, um sistema composto por dados de setores diferentes de uma empresa ou por dados de uma organização formada por empresas diferentes e cujos dados de uma empresa não devem ficar armazenados com os dados de outra.

Portanto, a distribuição de dados em um SCC pode ser usada como opção na busca por diversidade e conseqüente melhora de desempenho do sistema ou por necessidade da aplicação, como mostrado.

Com este trabalho, busca-se identificar quais sistemas de combinação de classificadores são mais apropriados à distribuição vertical dos dados. Além disso, pretende-se identificar que técnicas podem ser usadas com maior sucesso para separar verticalmente os dados entre os classificadores.

1.1 Objetivos

O objetivo deste trabalho é analisar a distribuição vertical de dados em sistemas de combinação de classificadores, tais como sistemas multiclassificadores paralelos (comitês) e sistemas multiagentes para classificação. Com esta análise pretende-se investigar:

- Quais sistemas de combinação são mais apropriados à distribuição vertical dos dados;
- Se esse comportamento se mantém com o aumento de componentes e se é independente da forma como os atributos foram distribuídos;
- Que técnicas podem ser usadas com maior sucesso para separar os dados entre os classificadores;
- Se há vantagens em utilizar técnicas de distribuição não-aleatória.

Para cumprir com o objetivo deste trabalho, algumas etapas serão realizadas:

- **DISTRIBUIÇÃO VERTICAL DOS DADOS:** Nesta etapa serão estudadas algumas técnicas para distribuição de características e serão aplicadas em bases de dados diferentes.
- **CONSTRUÇÃO DOS CLASSIFICADORES:** Com os dados distribuídos verticalmente, serão construídos os classificadores a partir de algoritmos de aprendizado diferentes.
- **CONSTRUÇÃO DOS SISTEMAS DE COMBINAÇÃO DE CLASSIFICADORES:** Após a construção dos classificadores, serão desenvolvidos sistemas de combinação, utilizando métodos de combinação diferentes.
- **ANÁLISE DOS RESULTADOS:** Após sua construção, os sistemas de combinação de classificadores serão testados usando diferentes configurações e tamanhos, diversificando as situações de classificação. A partir deste ponto seus desempenhos serão analisados de forma comparativa levando em consideração o tamanho do sistema, ou seja, a quantidade de classificadores usados como componentes do sistema e as técnicas aplicadas na distribuição das características.

1.2 Motivação

Como já mencionado, a distribuição dos dados em sistemas de combinação de classificadores pode ser uma exigência da aplicação (devido à distribuição intrínseca ao problema ou para manter a privacidade dos dados) ou pode ser uma opção na busca por diversidade e conseqüente melhora de desempenho do sistema.

A existência desse cenário motiva a realização de estudos que mostrem quais sistemas de combinação de classificadores são mais adequados em aplicações com distribuição vertical dos dados e que mostrem se existem técnicas mais apuradas para distribuir verticalmente os dados de uma aplicação. Estes estudos são de grande importância para a área de aprendizado de máquina, especificamente para a área de combinação de classificadores e isto motiva a realização deste trabalho.

Trabalhos têm sido desenvolvidos usando distribuição vertical dos dados, mas não há um estudo comparativo sobre os diferentes tipos de sistemas de combinação de classificadores quando submetidos à distribuição, assim como, não há um estudo comparativo sobre diferentes formas de se distribuir dados verticalmente.

1.3 Organização do Trabalho

Este trabalho está dividido em 7 capítulos, da seguinte forma:

- O Capítulo 1 apresenta uma visão geral do trabalho incluindo os objetivos e a motivação para sua realização.
- O Capítulo 2 apresenta um embasamento teórico sobre a área de classificação de padrões, mostrando os principais conceitos e algoritmos; sistemas multiclassificadores; métodos de combinação de classificadores; distribuição de dados entre classificadores e a questão da diversidade entre classificadores de padrões.
- O Capítulo 3 apresenta os principais conceitos de sistemas multiagentes e traz uma explanação sobre o *ClassAge*, um sistema multiagente para classificação de padrões.
- O Capítulo 4 mostra alguns trabalhos que utilizam distribuição de características em sistemas de combinação de classificadores.
- O Capítulo 5 apresenta o detalhamento da investigação, mostrando a metodologia dos experimentos e as técnicas utilizadas para a distribuição vertical dos dados.
- O Capítulo 6 traz os resultados encontrados e a análise dos mesmos.
- No Capítulo 7 são feitas as considerações finais do trabalho.

Capítulo 2

Classificação de Padrões

Classificação de padrões é o processo de atribuição de rótulos, também chamados de classes, a objetos (padrões), onde os objetos são descritos por um conjunto de medidas chamadas de atributos ou características (Kuncheva, 2004).

A classificação de padrões dá-se em duas fases distintas: FASE DE APRENDIZAGEM e FASE DE RECONHECIMENTO. Na primeira, é construído um modelo classificador a partir de um algoritmo de aprendizagem e de um conjunto de objetos pertencentes ao domínio que se deseja aprender. Esta fase pode ocorrer de maneira SUPERVISIONADA ou NÃO-SUPERVISIONADA. Na aprendizagem supervisionada é fornecido ao algoritmo de aprendizagem, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. Já na aprendizagem não-supervisionada, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters* (Monard & Baranauskas, 2000). Na segunda fase do processo de classificação de padrões, fase de reconhecimento, o classificador gerado na fase de aprendizagem é aplicado sobre novos objetos na determinação de suas classes.

Em outras palavras, o processo de classificação inicia-se com a construção dos classificadores, a partir de algoritmos de indução (algoritmos de aprendizado) e um conjunto de exemplos (objetos pertencentes ao domínio de conhecimento que se quer aprender). Estes classificadores serão utilizados na determinação das classes de novos exemplos.

2.1 Paradigmas de Aprendizado

Os algoritmos de aprendizagem podem pertencer a diversos paradigmas, tais como: Simbólico, Estatístico, Baseado em Exemplos e Conexionista. As próximas subseções apresentarão uma breve descrição de cada um desses paradigmas.

2.1.1 Paradigma Simbólico

Os sistemas de aprendizado simbólico buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contra-exemplos desse conceito. As representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras ou rede semântica

Uma árvore de decisão, por exemplo, é uma estrutura de dados definida recursivamente como um nó folha que corresponde a uma classe ou um nó de decisão que contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma subárvore. Em geral, o procedimento de uma árvore de decisão é similar ao dado a seguir.

1. Apresenta-se um conjunto de dados ao nó inicial (ou nó raiz) da árvore;
2. Dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos;
3. Este procedimento é repetido até que um nó terminal (nó folha) seja alcançado.

De forma resumida, em uma árvore de decisão, a classificação de um caso se inicia pela raiz da árvore, e esta árvore é percorrida até que se chegue a uma folha. Em cada nó de decisão será feito um teste que irá direcionar o padrão de teste para uma subárvore (Abreu, 2006).

2.1.2 Paradigma Estatístico

Utiliza modelos estatísticos para encontrar uma boa aproximação do conceito induzido. Entre os algoritmos estatísticos destaca-se o de aprendizado bayesiano que faz uso de fórmulas estatísticas e cálculo de probabilidades para realizar a classificação.

As principais vantagens do aprendizado estatístico, especialmente o aprendizado bayesiano, são: o fato de poder embutir nas probabilidades calculadas o conhecimento de domínio que se tem e o fato da classificação se basear em evidências fornecidas. Por outro lado, a desvantagem reside justamente no seu caráter estatístico, ou seja, muitas probabilidades devem ser calculadas e isto pode ocasionar um alto custo computacional.

2.1.3 Paradigma Baseado em Exemplos

Classifica exemplos nunca vistos por meio de exemplos similares conhecidos. Esse tipo de aprendizado é também denominado *lazy*. Sistemas *lazy* necessitam manter os exemplos na memória para classificar novos exemplos. Como exemplo deste paradigma, pode-se citar o k -NN (k – *Nearest Neighbours*) que é, juntamente com o método de Raciocínio Baseado em Casos (RBC), uma das técnicas mais conhecidas.

O k -NN gera um classificador onde o aprendizado é baseado na analogia, ou seja, classifica exemplos nunca vistos por meio de exemplos similares conhecidos. Para determinar a classe de um elemento que não pertença ao conjunto de treinamento, o classificador k -NN procura k elementos do conjunto de treinamento que estejam mais próximos deste elemento desconhecido, ou seja, que tenham a menor distância. Estes k elementos são chamados de k vizinhos mais próximos. Verifica-se quais são as classes desses k vizinhos e a classe mais freqüente será atribuída à classe do elemento desconhecido.

2.1.4 Paradigma Conexcionista

O paradigma conexionista é formado por métodos chamados de redes neurais que são construções matemáticas simplificadas inspiradas no modelo biológico do sistema nervoso. A representação de uma rede neural envolve unidades altamente interconectadas e, por esse motivo, o nome conexionismo é utilizado para descrever a área de estudo (Monard & Baranauskas, 2000). Pode-se citar as seguintes redes neurais como exemplo:

- **MULTILAYER PERCEPTRON (MLP):** As redes neurais MLP são redes de múltiplas camadas onde cada camada tem uma função específica. A camada de saída recebe os estímulos da camada intermediária e constrói o padrão que será a resposta. As camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação de características apresentadas nos padrões de entrada e permitem que a rede crie sua própria representação, mais rica e complexa, do problema.

Durante o treinamento com o algoritmo *backpropagation*, a rede opera em uma seqüência de dois passos. Primeiro, um padrão é apresentado à camada de entrada da rede. A atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, o

erro é calculado. O erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.

- **RADIAL BASIS FUNCTION (RBF):** Uma rede neural *Radial Basis Function*, ou seja, com função de ativação de base radial pode ser vista como uma rede *feedforward* de múltiplas camadas, contendo três camadas com papéis distintos. A camada de entrada é composta por unidades sensoriais que conectam a rede ao seu ambiente. A segunda camada (única camada intermediária) contém os parâmetros das funções de base radial e a terceira camada forma combinações lineares das ativações das funções de base radial para gerar a saída.
- **FUZZY MLP:** *Fuzzy MLP (Fuzzy MultiLayer Perceptron)* é uma implementação da teoria dos conjuntos *fuzzy* em uma rede neural *perceptron* de múltiplas camadas, ou seja, é o resultado da *fuzzificação* direta no nível de rede, no nível de aprendizado ou em ambos na rede MLP. Além disso, a saída desejada é calculada de forma diferente que nas redes MLP, no qual utilizam o método do *winner-takes-all*. Sendo assim, os nodos correspondentes à saída desejada são modificados durante a fase de treinamento dando origem à saída desejada *fuzzy*.

2.2 Combinação de Classificadores

Sistemas multiclassificadores são sistemas para classificação de padrões baseados na combinação das saídas de um conjunto de diferentes classificadores, chamados de classificadores base. Essa combinação tem como objetivo conseguir uma classificação mais exata, com menos erros. O aumento de desempenho tende a ocorrer se forem respeitados critérios de diversidade entre os classificadores componentes do sistema, nesse caso, a diversidade é percebida se os erros dos classificadores base não coincidem.

Um sistema multiclassificador é, portanto, formado por um conjunto de classificadores base e uma função para combinação das saídas desses classificadores. A escolha dessa função, ou mecanismo, nem sempre é trivial e influencia diretamente o desempenho final do sistema (Abreu, 2006).

De acordo com sua arquitetura interna, um sistema multiclassificador pode ser modular ou paralelo. Os primeiros dividem a tarefa final do sistema em subtarefas, criando

especialistas. Cada especialista executa uma subtarefa e o módulo combinador utiliza as soluções dessas subtarefas para determinar a saída final do sistema.

Os sistemas multiclassificadores paralelos, também conhecidos como *ensembles* ou comitês, utilizam uma combinação redundante, pois todos os classificadores base executam a mesma tarefa. Essa forma de combinação explora a idéia de que a diferença existente entre os classificadores base pode extrair informações complementares sobre os exemplos a serem classificados

Existem duas estratégias principais de combinação de classificadores: Fusão e Seleção. Estas estratégias serão apresentadas nas subseções 2.2.1 e 2.2.2, respectivamente.

2.2.1 Métodos de Fusão

A combinação de classificadores baseada em fusão utiliza a resposta de todos os classificadores componentes do sistema na construção da resposta final, ou seja, a classificação final resulta da opinião coletiva dos classificadores participantes. Alguns dos métodos de fusão mais utilizados são:

- **SOMA:** O método Soma funciona de maneira que quando apresentado um padrão de entrada para os classificadores, os valores das saídas correspondentes a cada classe de cada classificador são somados e a classe vencedora é aquela que possuir o maior valor absoluto.
- **VOTO:** No método do Voto, quando um padrão de entrada é apresentado para os classificadores, cada um deles vota na classe que ele achar que é a correta. A classe vencedora é aquela que possuir o maior número de votos.
- **NAIVE BAYES:** No *Naive Bayes*, para cada classe de decisão é calculada a probabilidade condicional de que essa classe seja correta. As probabilidades envolvidas são calculadas como frequências obtidas sobre os padrões de treinamento (Abreu, 2006). Ou seja, durante a fase de treinamento, o combinador *Naive Bayes* gera uma matriz de confusão para cada classificador base. A partir dessas matrizes, são calculadas as probabilidades de que cada uma das classes esteja correta. A classe com a maior probabilidade é dada como resposta final do sistema.
- **MLP (NN):** Este método de combinação utiliza uma rede neural *perceptron* multicamadas, já definida neste capítulo. O diferencial ao usar esse modelo como

combinador ao invés de classificador base é que suas entradas serão as saídas dos classificadores componentes do sistema.

2.2.2 Métodos de Seleção DCS

Na combinação de classificadores baseada em seleção, a resposta final do sistema é dada pelo classificador base mais capacitado para o padrão de entrada dado. Um método de seleção que tem sido bastante utilizado é o DCS, Classificador de Seleção Dinâmica.

O DCS é baseado em exatidão dinâmica, também conhecida como LCA (*Local Class Accuracy*). O DCS utiliza análise de competência para cada classificador, dado um padrão de entrada por ele classificado. O principal passo para calcular a exatidão local de uma classe (LCA) para um padrão de teste x pode ser definido da seguinte forma.

Cada classificador componente do sistema classifica o padrão de entrada, ou seja, atribui um rótulo – uma classe ao padrão. Em seguida, será verificada pelo combinador DCS a proporção de acertos de cada classificador para os k padrões mais parecidos com o padrão atual e que foram rotulados com a mesma classe pelo classificador em questão.

Por exemplo, dado um padrão de entrada x para classificação em um sistema com dois classificadores base e um módulo combinador DCS. O classificador 1 atribui x à classe A e o classificador 2 atribui x à classe B . O LCA do classificador 1 será a média de acertos dele para os 10 (por exemplo) padrões mais parecidos com x e que foram rotulados por ele como sendo da classe A . Para o classificador 2 será a média de acertos dele para os 10 padrões mais parecidos com x e que foram rotulados por ele como sendo da classe B .

A exatidão de cada classificador (LCA) será a proporção de padrões classificados corretamente. Dessa forma, o classificador que tiver o maior LCA é o mais competente segundo o DCS para classificar o padrão de entrada dado.

A partir do DCS foram desenvolvidos outros métodos de seleção. Estas derivações do DCS são chamadas de DCS-MCB e DCS-DT, definidas a seguir:

- DCS-MCB (MULTIPLE CLASSIFIER BEHAVIOUR): Neste modelo um classificador é selecionado se e somente se o maior LCA é substancialmente maior que os valores dos LCA dos demais classificadores. Para ser substancialmente melhor, o

classificador deve ter um LCA maior que um limiar previamente definido. Se não for, o padrão de teste é classificado através da aplicação da técnica do Voto.

- DCS-DT (DECISION TEMPLATES): A principal diferença do modelo anterior para este é que um teste estatístico é feito para determinar se o classificador com maior LCA é substancialmente melhor que os outros. Além disso, no caso de não haver uma diferença estatística significativa é feito um esquema baseado numa matriz de decisão (*decision template matrix*) formada a partir dos dados de treinamento. A matriz de decisão é formada da seguinte forma:

Para cada padrão de treinamento apresentado aos classificadores base, uma matriz chamada de perfil de decisão (*decision profile*) é construída, onde cada linha possui os valores entre 0 e 1 que um classificador base atribuiu para cada classe possível, significando a confiança que ele tem de que o padrão pertença aquela classe. A matriz de decisão será uma média das N matrizes de perfil de decisão, onde N é a quantidade de padrões de treinamento. Quando um padrão teste for apresentado e não houver uma superioridade estatística de um dos classificadores na sua classificação, então serão comparadas as matrizes de perfil de decisão do padrão atual e a matriz de decisão construída com os padrões de treinamento, usando uma medida de similaridade (distância euclidiana, por exemplo). O classificador mais similar classificará o padrão de entrada.

2.3 Distribuição de Dados

Assumindo que os dados são representados na forma de uma matriz, onde as linhas contêm os padrões e as colunas contêm os atributos conhecidos sobre o padrão, os dados podem ser distribuídos entre os componentes do sistema de classificação de dois modos: distribuição horizontal, onde cada componente utiliza subconjuntos de padrões com todos os atributos, e distribuição vertical, onde os componentes têm acesso a todos os padrões com subconjuntos de atributos. Partições horizontais e verticais também são chamadas de homogêneas e heterogêneas.

Os sistemas de combinação de classificadores com distribuição dos atributos (distribuição vertical) podem ser divididos em três grupos:

- **DISTRIBUIÇÃO NATURAL:** Quando as características são agrupadas naturalmente, por estarem armazenadas em repositórios distribuídos ou por serem facilmente diferenciadas.
- **DISTRIBUIÇÃO RANDÔMICA:** Escolha aleatória de subconjuntos de características.
- **DISTRIBUIÇÃO NÃO RANDÔMICA:** Usa alguma técnica para distribuir as características entre os atributos.

A distribuição dos dados busca aumentar a diversidade do sistema, gerando componentes que produzam mapeamentos (entrada – saída) diferentes para o mesmo problema e solucionar problemas de privacidade dos dados.

2.4 A Questão da Diversidade

Não é necessária a construção de um SCC, se o classificador não comete erros. Entretanto, se o classificador comete erros será preciso complementá-lo com outros classificadores que cometam erros diferentes (Kuncheva, 2004), isto quer dizer que os classificadores erram na classificação de padrões diferentes, ou seja, dado um padrão de entrada, alguns classificadores base podem não classificá-lo corretamente, enquanto outros sim. Desta forma, o resultado final após a combinação de suas respostas pode ser correto. Sendo assim, a diversidade na saída dos classificadores base é um requisito vital para o sucesso do SCC. Intuitivamente, espera-se que os componentes do SCC sejam tão corretos quanto possível e no caso de cometerem erros que ocorram em padrões diferentes.

Na etapa de construção do SCC, existem técnicas que tentam otimizar a diversidade. Essas técnicas atuam principalmente sobre os dados de treinamento, através do fornecimento de conjuntos de dados diferentes para cada um dos componentes do sistema e sobre a arquitetura de cada componente.

As técnicas mais pesquisadas para construção de um SCC diverso são as que utilizam conjuntos de treinamento diferentes para cada um dos componentes. Essas técnicas buscam gerar componentes que produzam mapeamentos (entrada – saída) diferentes para o mesmo problema.

Capítulo 3

Sistemas Multiagentes

Sistemas multiagentes (SMA) são sistemas compostos pela interação entre múltiplos elementos computacionais autônomos, conhecidos como agentes. Para o SMA, o termo autônomo designa o fato de que os agentes têm uma existência própria, independente da existência de outros agentes. Geralmente, cada agente possui um conjunto de capacidades comportamentais, um conjunto de objetivos e autonomia de utilizar suas capacidades para alcançar seus objetivos.

Dado um determinado sistema, denomina-se agente cada uma de suas entidades ditas ativas (capaz de controlar suas ações), as entidades passivas são designadas pelo termo ambiente. Um agente recebe informações e raciocina sobre o ambiente e sobre os outros agentes, decidindo quais ações deve tomar. O projetista do sistema deve definir todas as restrições aplicadas aos agentes, garantindo que cada agente desejará e realizará a resolução do problema proposto.

De forma geral, as principais capacidades que um sistema deve possuir para ser considerado um agente inteligente são: reatividade, proatividade e habilidade social (Abreu, 2006).

- Reatividade é a capacidade de perceber o ambiente e responder de forma oportuna às mudanças que ocorrem nele a fim de satisfazer algum objetivo.
- Proatividade é a capacidade de possuir comportamento dirigido diretamente a um determinado objetivo.
- Habilidade Social é a capacidade de interagir com o ambiente e com outros agentes a fim de satisfazer seus objetivos.

Denomina-se interação entre os agentes ou entre o agente e o ambiente uma troca de informações, que pode ser realizada de forma direta (comunicação explícita) ou indireta (emissão de sinais através do ambiente). Nessa troca de informações, processos mais complexos como cooperação e negociação podem ser utilizados.

A cooperação ocorre quando um agente não é capaz de executar uma determinada tarefa, porque não possui conhecimento suficiente, ou porque não conhece toda informação de entrada necessária (cooperação por partilha de tarefas). Ou ainda, quando um agente possui uma informação que sabe ser útil para um outro agente (cooperação por partilha de resultados). De qualquer forma, os agentes estão tentando atingir conjuntamente um mesmo objetivo.

O processo de negociação, por sua vez, ocorre quando os agentes divergem em relação a certos aspectos do problema a ser resolvido ou quando apesar de poderem executar suas tarefas sozinhos, negociar a execução dessas tarefas pode trazer benefícios aos agentes envolvidos. Neste último caso, a negociação pode não resultar em acordo, fazendo com que cada agente execute a tarefa inicialmente definida, já no primeiro caso, onde os agentes divergem sobre a resolução do problema, um acordo deverá ser encontrado.

A negociação pode ser vista como uma troca de mensagens baseada num dado protocolo, destinada ao estabelecimento de um acordo. Segundo (Wooldridge, 2002), este processo ocorre em uma série de *rounds*, onde em cada *round* cada agente faz uma proposta aos demais. Essas propostas são definidas pelas estratégias individuais de cada agente.

3.1 Sistema Multiagente para Classificação

Estudos têm mostrado que os sistemas multiclassificadores conseguem resultados superiores aos dos métodos individuais, porém a implementação desses sistemas não é uma tarefa fácil, já que a escolha do melhor método de combinação necessita de execuções exaustivas. Isso porque o processo de tomada de decisão é centralizado e torna o desempenho do sistema diretamente dependente da escolha do método de combinação.

Um sistema multiagente foi proposto como uma possível solução para o problema de tomada de decisão centralizado dos SMC, esse sistema foi chamado de *NeurAge*. No *NeurAge* quando uma entrada é mostrada para o sistema, todos os agentes classificadores produzem suas próprias respostas.

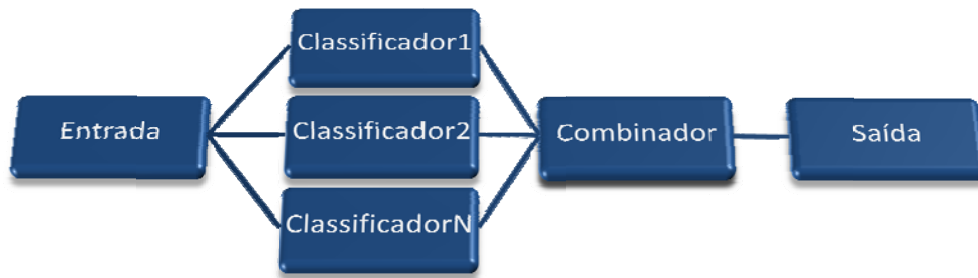


Figura 1: Modelo de um sistema multiclassificador paralelo tradicional

Ao contrário dos SMC, onde existe um método de combinação centralizado que escolhe a saída do sistema, todos os agentes se comunicarão com os demais para encontrar uma saída comum para o sistema. Dessa maneira, tem-se um caminho mais flexível para realizar a tomada de decisão.

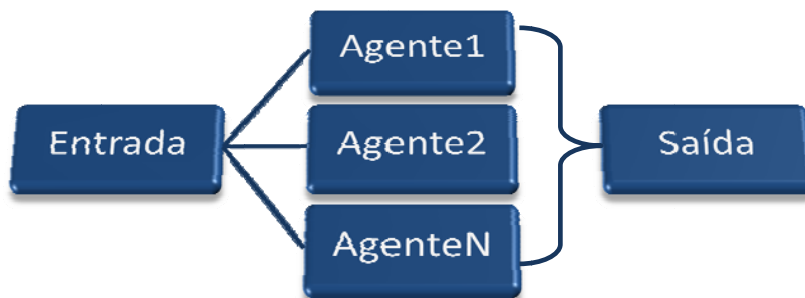


Figura 2: Modelo de um sistema multiagente para classificação, eliminando o módulo de combinação

O *NeurAge* usa agentes neurais, ou seja, os agentes classificam o padrão utilizando redes neurais. Em (Abreu, 2006), o *NeurAge* foi estendido permitindo o uso de agentes não neurais, esse novo sistema foi chamado de *ClassAge* (*Classifier Agent*). O *ClassAge* usa um método de negociação entre os agentes chamado de método da sensibilidade, descrito na próxima subseção.

3.1.1 Método de Negociação Baseado na Sensibilidade

Os métodos de classificação usados pelos agentes classificadores fornecem como resposta graus de pertinência do padrão analisado para cada uma das possíveis classes da base de dados em questão. Esses valores estão entre 0 e 1, ou seja, para cada uma das possíveis classes, um valor entre 0 e 1 é gerado indicando a certeza que o agente classificador tem sobre a pertinência do padrão àquela classe (0 % de certeza a 100% de certeza) (Santana, 2005).

O sistema *ClassAge* usa um método de negociação que tenta resolver possíveis conflitos durante a classificação dos padrões apresentados. Esse método de negociação usa os graus de pertinência como sendo a confiabilidade sobre a resposta do agente, ou seja, a certeza que o agente classificador tem de que o padrão apresentado pertence a determinada classe do domínio de conhecimento tratado.

No sistema aqui abordado, os agentes classificadores negociam um resultado comum, verificando possíveis falhas dos outros agentes e diminuindo a confiabilidade destes em suas próprias respostas. Isso é feito utilizando a análise de sensibilidade, método desenvolvido em (Canuto, 2004), que mede a suscetibilidade do agente a falhas quando os atributos de entrada são submetidos a pequenas mudanças. Isto é, se o resultado da classificação do agente sofrer variações altas quando os atributos de entrada sofrem pequenas mudanças, a confiabilidade de sua resposta deve ser revista.

Seguindo essa metodologia, os agentes sofrem uma maior ou menor diminuição em sua confiabilidade de acordo com sua maior ou menor sensibilidade às mudanças nos atributos de entrada do padrão.

Capítulo 4

Trabalhos Relacionados

Podem ser citados como trabalhos relacionados, aqueles trabalhos que envolvem sistemas de combinação de classificadores (comitês e sistemas multiagentes para classificação) e os trabalhos que envolvem distribuição de dados em sistemas de combinação de classificadores.

Na seção 4.1 serão citados alguns trabalhos que utilizam comitês e sistemas multiagentes para classificação e na seção 4.2 trabalhos que envolvem distribuição de dados em SCC.

4.1 Sistemas de Combinação de Classificadores

Em (Abreu et al., 2005) foram feitos experimentos utilizando os métodos de combinação: *Stacking*, *StackingC* e *Voto*. Como métodos de classificação foram utilizados: *Árvore de Decisão*, *k-NN*, *Naive Bayes*, *Máquinas de Vetores de Suporte* e *Redes Neurais*. Neste trabalho foi observado que os sistemas multiclassificadores têm um desempenho melhor que os classificadores.

Em (Chen et al., 2005) foi feita uma investigação sobre independência e diversidade entre classificadores de padrões participando de um esquema multiclassificador. O método de combinação utilizado é baseado em fusão, ou seja, cada classificador tenta resolver o mesmo problema e o combinador é responsável pela saída final usando as respostas de todos os classificadores base. Neste trabalho é apresentada uma definição formal da independência estatística dos classificadores.

Em (Vuurpijl & Schomaker, 1998) é descrito um framework que utiliza agentes inteligentes para reconhecimento de padrões. Com ele é implementado um reconhecedor de dígito para validação de senha. A modularidade intrínseca do uso de agentes provê uma resposta mais rápida, um desenvolvimento mais estratificado e um paralelismo para o sistema proposto.

Em (Abreu et al., 2005a) foi apresentada uma análise entre sistemas de classificação e métodos de negociação propostos no sistema *NeurAge*. Foram utilizados os métodos de negociação: Teoria dos Jogos, Leilão e Negociação Baseada na Sensibilidade. Dentre os métodos de combinação consagrados na literatura, foram utilizados a Soma e o Voto. Este trabalho mostra que o método baseado na sensibilidade possui um melhor desempenho comparado com os demais.

Em (Abreu, 2006) foi apresentado o sistema *ClassAge*, um sistema multiagente para classificação que foi baseado no sistema *NeurAge*. Este sistema foi estudado usando os métodos de negociação baseado em leilão, baseado na teoria dos jogos e o método baseado na sensibilidade.

4.2 Distribuição de Dados em SCC

Em (Villar, 2005) é apresentado um sistema multiagente com raciocínio baseado em casos para aprendizado com dados distribuídos horizontalmente e tomada de decisão descentralizada. Cada agente é capaz de aprender individualmente e resolver problemas. Cada caso (exemplo) pertence e é gerenciado por um único agente e qualquer informação só é aberta ou compartilhada se o agente decidir isso, garantindo a privacidade dos dados. O foco desse trabalho é investigar como os agentes podem colaborar para o sucesso do sistema, determinando quando o agente deve resolver o problema individualmente ou em conjunto e quais agentes devem participar do processo de decisão.

Nas próximas subseções são apresentados alguns trabalhos que envolvem sistemas de combinação de classificadores com distribuição dos atributos. Estes trabalhos podem ser divididos em três grupos: aqueles com distribuição natural, distribuição randômica e distribuição não randômica dos atributos.

4.2.1 Distribuição Natural

Em alguns problemas as características são agrupadas naturalmente. Por exemplo, em (Chen et al., 1997) apresenta-se comitês com distribuição de características e aplicações no reconhecimento de voz. A distribuição das características está relacionada ao impulso do sinal e ao espectro do discurso. Já em (Van Kreukelen, 1997), é feito reconhecimento de manuscritos, nesse caso uma imagem pode ser vista a partir de diferentes perspectivas, como por exemplo, *pixels* e características morfológicas.

4.2.2 Distribuição Randômica

A escolha randômica de subconjuntos de características é chamada de *método de subespaço aleatório*. Cada classificador no comitê é construído sobre uma escolha aleatória de um subconjunto de características com um tamanho predefinido. Em (Ho, 1998) bons resultados são obtidos para três classificadores construídos com aproximadamente metade das características, mostrando que o método de subespaço aleatório trabalha bem quando há redundância nas informações.

4.2.3 Distribuição Não Randômica

Os trabalhos desenvolvidos sob essa ótica usam técnicas diversas para distribuir as características. Em (Oza & Turner, 2001), por exemplo, é definido um algoritmo simples para selecionar os subconjuntos de características. O comitê consiste de C classificadores, onde C é o número de classes. Cada classificador tem uma classe “favorita”. Para encontrar o subconjunto de características do classificador com uma classe W favorita calcula-se a correlação entre cada característica e a variável de classe. A variável de classe tem valor 0 para todos os objetos que não são da classe W e 1 para todos os objetos que são da classe W . Dessa forma, são formados grupos correlatos em relação à classe, de modo a criar diversidade entre os classificadores do comitê.

Em (Puuronen et al., 2001) um método é proposto para construção de subconjuntos de características de forma iterativa, onde inicialmente é utilizada uma distribuição baseada na classe favorita. O procedimento consiste nos seguintes passos:

1. Gera-se um comitê inicial com C classificadores de acordo com a classe favorita baseado numa correlação simples;
2. Identifica-se o classificador cuja saída difere menos das saídas dos demais classificadores. Este classificador chamado de mediano é identificado usando alguma medida de diversidade, que podemos denotar por δ . Altos valores de δ significam alto desacordo entre os classificadores e $\delta = 0$ significa saídas idênticas. O classificador mediano é aquele com menor δ em relação aos outros componentes do comitê.

3. Executam-se alterações em cada característica do subconjunto usado para construir o classificador mediano. Construindo, desta forma, N classificadores candidatos, onde N é a quantidade de características do classificador mediano;
4. Calcula-se a acurácia do comitê substituindo o classificador mediano por cada um dos classificadores candidatos. Se houver um aumento de desempenho efetiva-se a substituição e volta ao passo 2. Se não houver aumento significativo, escolhe-se o comitê com a configuração anterior.

Em (Tsybal et al., 2003) são feitas algumas alterações no trabalho realizado em (Puuronen et al., 2001): O comitê inicial é gerado através do método de subespaço aleatório, citado anteriormente. Outra variação no procedimento do modelo iterativo é checar todos os membros do comitê e não somente o classificador mediano.

Recentemente, alguns autores têm investigado o uso de métodos de otimização como algoritmos genéticos na construção de comitês. Em (Kuncheva & Jain, 2000), por exemplo, os autores sugerem dois métodos baseados em algoritmos genéticos para distribuir as características entre os componentes do comitê, onde cada subconjunto de características é um membro da população. O primeiro usa subconjuntos disjuntos e o segundo aceita interseção de características. Nos dois casos, a função de ajuste é a acurácia do comitê. Em (Guerra-Salcedo & Whitley, 1999), os autores usam um algoritmo genético para explorar o espaço de possíveis subconjuntos de características, construindo o comitê com a população mais acurada.

Nos trabalhos mencionados, o foco principal são os métodos usados para a distribuição das características. Entretanto, a acurácia dos comitês é também fortemente afetada pelos métodos de combinação (usados para combinar as saídas dos classificadores base). Por exemplo, dois comitês compostos pelos mesmos classificadores, mas usando diferentes métodos de combinação podem ter diferentes acurácias. Diferentemente dos trabalhos citados acima, o foco deste trabalho não são os métodos de distribuição das características, mas o impacto deles em diferentes sistemas de combinação de classificadores.

Capítulo 5

Detalhamento da Investigação

Neste capítulo serão apresentadas as técnicas utilizadas na distribuição vertical dos dados de entrada do sistema, bem como as bases de dados experimentadas. Além disso, serão definidos os experimentos e suas configurações.

5.1 Distribuição Vertical dos Dados

Para investigar o comportamento de sistemas de combinação de classificadores quando submetidos à distribuição dos atributos de entrada, serão definidos três tipos de sistemas:

- **SISTEMAS SEM DISTRIBUIÇÃO DAS CARACTERÍSTICAS:** Nesse tipo de sistema, os classificadores base terão acesso a todas as características do padrão de entrada. O objetivo desse sistema neste trabalho é servir como parâmetro comparativo na análise dos sistemas com distribuição.
- **SISTEMAS COM DISTRIBUIÇÃO TOTAL DAS CARACTERÍSTICAS:** Nos sistemas com distribuição total, as características são distribuídas entre os classificadores base de forma igualitária e sem interseção, ou seja, cada atributo só pode servir de entrada para um classificador base.
- **SISTEMAS COM DISTRIBUIÇÃO PARCIAL DAS CARACTERÍSTICAS:** Neste caso, os classificadores base compartilharão 50% das características do padrão de entrada. A outra metade das características será dividida igualmente e sem interseção entre os classificadores. Por exemplo, se um padrão de entrada possui 18 atributos, 9 servirão de entrada para todos os classificadores do sistema e os outros 9 serão divididos, sem interseção, entre os classificadores.

Para os sistemas com distribuição, serão utilizadas cinco técnicas diferentes. A variedade de técnicas de distribuição serve para garantir que a comparação final entre os sistemas de combinação e entre os experimentos com distribuição e sem distribuição não foi afetada pela técnica utilizada. Além disso, busca-se identificar que técnicas podem ser usadas com maior sucesso para separar verticalmente os dados entre os classificadores.

Uma das técnicas será aleatória e quatro não aleatórias. O uso de uma técnica aleatória objetiva verificar se a forma como se distribui os atributos é significativa ou não para o resultado do sistema.

Todas as técnicas usam o mesmo procedimento geral, diferenciando apenas o critério de avaliação das características. O procedimento geral pode ser descrito como segue:

- a) Ordenam-se todos os atributos de acordo com o critério de avaliação da técnica utilizada;
- b) Divide-se esse vetor ordenado de atributos em grupos, de forma que a quantidade de atributos em cada grupo seja igual à quantidade de classificadores base;
- c) Para cada classificador base é associado um atributo de cada grupo.

Este procedimento objetiva certificar-se que os atributos importantes (baseado no critério de avaliação) serão distribuídos por todos os classificadores base de forma igualitária.

Como já mencionado, a principal diferença entre as técnicas é o critério de avaliação das características usado para ordená-las. Esses critérios são descritos a seguir:

TÉCNICA 1: DISTRIBUIÇÃO BASEADA NA ENTROPIA

Esta técnica é baseada na forma de escolha dos atributos mais significantes pelo algoritmo de árvore de decisão. A entropia mede a quantidade de informação média necessária para identificar a classe de um padrão apresentado (Santana et al., 2007). Em outras palavras, a entropia é tão maior quanto maior a dificuldade de classificação do padrão. Dessa forma, pode-se ordenar os atributos pela sua capacidade em diminuir a entropia no processo de classificação de padrões.

TÉCNICA 2: DISTRIBUIÇÃO BASEADA NO ERRO

Essa forma de avaliação define a relevância do atributo para o classificador, removendo o atributo do conjunto total e calculando a acurácia do classificador. Dessa forma, quanto maior o erro do classificador na ausência do atributo em questão, maior é a significância do atributo para a classificação dos padrões. Esse processo é feito da seguinte forma.

Para uma base com N atributos, N classificadores são treinados e testados, cada um com $N-1$ atributos, de modo que os atributos ausentes nos N classificadores sejam diferentes. Depois, os atributos são ordenados de acordo com o erro apresentado pelo classificador que não tinha este atributo no conjunto de treinamento.

TÉCNICA 3: DISTRIBUIÇÃO RANDÔMICA

Neste caso, os atributos das bases de dados foram ordenados sem a utilização de um critério específico.

TÉCNICA 4: DISTRIBUIÇÃO BASEADA NA VARIÂNCIA

Uma forma de ordenar os atributos de uma base de dados é analisando a variância desses atributos, que é uma medida que indica quão diferentes são, em geral, os valores atribuídos à variável. Essa medida pode mostrar o quão representativo é um atributo no momento da classificação de um padrão. Isto porque, se o atributo possui uma variância baixa, quer dizer que seu valor para o padrão dado é muito parecido com seu valor para os outros padrões, sendo assim, não fará muita diferença na tomada de decisão do classificador. Por outro lado, uma variância alta mostra que o atributo é mais significativo para o classificador.

TÉCNICA 5: DISTRIBUIÇÃO BASEADA NA CORRELAÇÃO DE PEARSON

O coeficiente de Correlação Pearson é um valor que indica a força de relacionamentos lineares entre as variáveis (Kuncheva, 2004). Esta medida pode ser usada, por exemplo, para definir a correlação (dependência) das características de um padrão.

Como uma técnica de distribuição de características, pode-se usar para garantir a diversidade entre os classificadores, fazendo com que possuam características pouco correlatas e internamente ao classificador, fazendo com que as características analisadas não sejam muito dependentes. Neste sentido, são calculadas as correlações dos atributos em relação aos outros, formando subconjuntos diversos.

Para esta técnica em especial, não é feita a ordenação utilizada nos modelos anteriores, e sim a formação de conjuntos com características diversas.

5.2 Experimentos

Neste trabalho, são comparados os desempenhos de sete sistemas de combinação de classificadores, sendo seis comitês e um multiagente, o *ClassAge*. Dos comitês analisados, quatro são baseados em fusão – MLP, *Naive*, Soma e Voto; e dois são comitês baseados em seleção – DCS_DT e DCS_MCB.

Seus desempenhos são analisados quando submetidos à distribuição parcial e total dos atributos de entrada. Como já mencionado, a distribuição é feita usando cinco técnicas diferentes: Correlação de Pearson; Entropia; Erro; Random e Variância.

Foram feitos também experimentos com os sistemas de combinação sem distribuição dos atributos, para servir de parâmetro na análise do comportamento dos sistemas quando submetidos à distribuição.

Nesta investigação, são analisados sistemas com quatro tamanhos diferentes, ou seja, com quantidade de componentes diferentes. São investigados sistemas com 3, 5, 7 e 9 componentes.

Os componentes dos sistemas, classificadores base, são de cinco tipos: k-NN (*Nearest Neighbor*); C4.5 (Árvore de decisão); Rede Neural RBF (*Radial Basis Function*); Rede Neural MLP (*MultiLayer Perceptron*) e Rede Neural fuzzy MLP. A variedade de classificadores base serve para garantir a diversidade do sistema. Para cada tamanho de sistema, foram feitos experimentos com várias configurações de componentes e retirada a média.

Além dos comitês e do sistema multiagente, serão analisados os resultados dos classificadores base componentes dos sistemas, ou seja, será feita a média dos resultados dos classificadores que compõem os sistemas. Estes resultados servem como parâmetro comparativo, a fim de identificar o alto ou baixo desempenho dos sistemas de combinação (comitês e *ClassAge*) para cada técnica de distribuição de dados. Como a combinação de classificadores tem a capacidade de melhorar o desempenho do sistema, se um sistema de combinação em alguma situação experimentada conseguir resultados inferiores aos dos classificadores base, aquela situação não é favorável e deve ser evitada.

5.3 Investigação

A investigação dos sistemas de combinação de classificadores com distribuição vertical dos dados será feita sob dois aspectos principais: análise dos resultados de acordo com o tamanho do sistema e de acordo com a técnica de distribuição utilizada.

Serão analisados para a distribuição parcial e para a distribuição total quais sistemas de combinação obtiveram melhores resultados e se melhoraram em relação aos classificadores base. Será investigado também se este comportamento aconteceu em todas as técnicas de distribuição e em todos os tamanhos estudados. Em relação ao tamanho do sistema, deseja-se verificar como os sistemas se comportam quando aumenta-se a quantidade de componentes e qual o melhor tamanho para o sistema. Por fim, busca-se obter resposta para a seguinte questão: Pode-se identificar uma técnica de distribuição de características que gere melhores resultados para o sistema?

Além da análise dos sistemas em relação às técnicas e ao tamanho na distribuição parcial e total, será comparado o desempenho geral da distribuição parcial com a total e de ambas com o desempenho geral sem distribuição.

Na análise dos SCC em relação à quantidade de componentes, foram feitas médias, para cada tamanho proposto, das acurácias dos sistemas nas cinco técnicas de distribuição, por exemplo, o desempenho do *ClassAge* com 3 componentes será a média dos resultados do *ClassAge* com 3 componentes em cada uma das 5 técnicas experimentadas. Já na análise em relação à técnica de distribuição, para cada técnica proposta foram feitas médias das acurácias dos SCC nos quatro tamanhos experimentados (3, 5, 7 e 9), ou seja, o desempenho do *ClassAge* na técnica de Entropia, por exemplo, será a média do desempenho do *ClassAge* com 3, 5, 7 e 9 componentes usando a técnica de Entropia para distribuir verticalmente os dados.

5.4 Bases de Dados Utilizadas

Para este trabalho, foram utilizadas duas bases de dados. Essas bases são suficientemente diferentes entre si para garantir a confiabilidade dos resultados. Uma visão geral de suas características pode ser vista na Tabela 1:

Base de dados	Qtd de Padrões	Qtd de Atributos	Qtd de Classe
Imagem	2310	18	7
Proteínas	584	126	5

Tabela 1: Características das bases de dados utilizadas neste trabalho.

BASE DE DADOS DE IMAGEM

Esta base de dados foi conseguida no UCI *repository* (Blake & Merz). A base possui 2310 padrões definidos a partir de 18 características de 7 conjuntos de imagens de outdoors. A base de Imagem é completamente balanceada, ou seja, possui a mesma quantidade de exemplos para cada classe.

BASE DE DADOS DE PROTEÍNAS

Esta base de dados representa uma classificação hierárquica de estruturas conhecidas de proteínas, organizadas de acordo com seus relacionamentos evolucionários e estruturais. Possui 584 padrões definidos por 126 características de 5 classes de proteínas. É uma base difícil de ser trabalhada por ser desbalanceada.

A base de imagem representa um ambiente de classificação menos propenso a erros, por ser balanceada em relação às classes e ter uma boa quantidade de exemplos. Já a base de proteínas representa um ambiente de difícil classificação, por ser desbalanceada, possuir poucos exemplos e atributos redundantes. Os resultados poderão mostrar quais técnicas são mais adequadas em cada caso ou se existe uma técnica eficiente nos dois casos propostos.

Capítulo 6

Análise dos Resultados

As Figuras de 3 a 19 apresentam os resultados do uso de distribuição de dados em sistemas de combinação de classificadores. Os resultados encontrados neste estudo serão analisados sob três perspectivas definidas a seguir:

- **NÚMERO DE COMPONENTES NO SISTEMA:** Esta parte da análise investigará como os sistemas de combinação de classificadores submetidos à distribuição dos dados se comportam quando a quantidade de componentes aumenta e conseqüentemente diminui-se a quantidade de informação utilizada por cada componente na classificação dos padrões. O número de componentes (classificadores ou agentes classificadores) variará entre 3, 5, 7 e 9.
- **TÉCNICAS DE DISTRIBUIÇÃO DOS DADOS:** O objetivo desta parte da análise é investigar como os sistemas de combinação se comportam quando utilizam diferentes técnicas de distribuição de características entre os componentes. Cinco técnicas serão analisadas: Correlação de Pearson, Entropia, Erro, Random e Variância.
- **USO DE DISTRIBUIÇÃO DE DADOS:** Nesta parte da análise, será comparado o desempenho geral dos sistemas de combinação sem distribuição das características, com distribuição parcial e com distribuição total a fim de verificar qual dessas abordagens apresenta os melhores resultados em cada uma das bases de dados experimentadas.

Como já mencionado, comitês com seis diferentes métodos de combinação serão usados nesta investigação, sendo quatro métodos baseados em fusão (Soma, Voto, MLP-NN e *Naive*) e dois métodos baseados em seleção (DCS_DT e DCS_MCB), além de um sistema multiagente, o *ClassAge*.

Além dos comitês e do sistema multiagente, serão analisados os resultados dos classificadores base usados na composição dos sistemas combinadores. Estes resultados serão apresentados como a média das acurácias dos classificadores base usados na composição dos sistemas de combinação para cada tamanho. Como já mencionado, estes

valores servem como parâmetro comparativo, a fim de identificar o alto ou baixo desempenho dos sistemas de combinação (comitês e *ClassAge*).

6.1 Análise em relação ao número de componentes do sistema

As Figuras de 3 a 10 apresentam os resultados dos sistemas de combinação analisando a quantidade de classificadores base usados como componentes dos sistemas. Como já mencionado, todos os sistemas de combinação foram submetidos a cinco técnicas de distribuição dos dados, descritas no capítulo anterior. Os valores apresentados nesta seção representam a média da acurácia dos sistemas nas cinco técnicas de distribuição definidas.

Esta seção está dividida em 4 subseções, na primeira são apresentados os resultados dos sistemas sem distribuição dos atributos para servir de parâmetro na análise do comportamento dos sistemas quando submetidos à distribuição. Na subseção 6.1.2 são apresentados os resultados dos sistemas com distribuição parcial das características e na subseção 6.1.3 os resultados dos sistemas com distribuição total. Na última subseção, 6.1.4, é feita uma análise comparativa dos resultados das distribuições parcial e total.

6.1.1 Sistemas sem distribuição das características

Como já mencionado, foram feitos experimentos com os sistemas de combinação sem distribuição dos atributos para servir de parâmetro na análise do comportamento dos sistemas quando submetidos à distribuição. Os resultados dos sistemas com 3, 5, 7 e 9 componentes são mostrados nas Figuras 3 e 4 para as bases de imagem e proteínas, respectivamente.

Os resultados apresentados nas figuras como sendo dos classificadores representam a média dos resultados dos classificadores base usados na composição dos sistemas para cada tamanho.

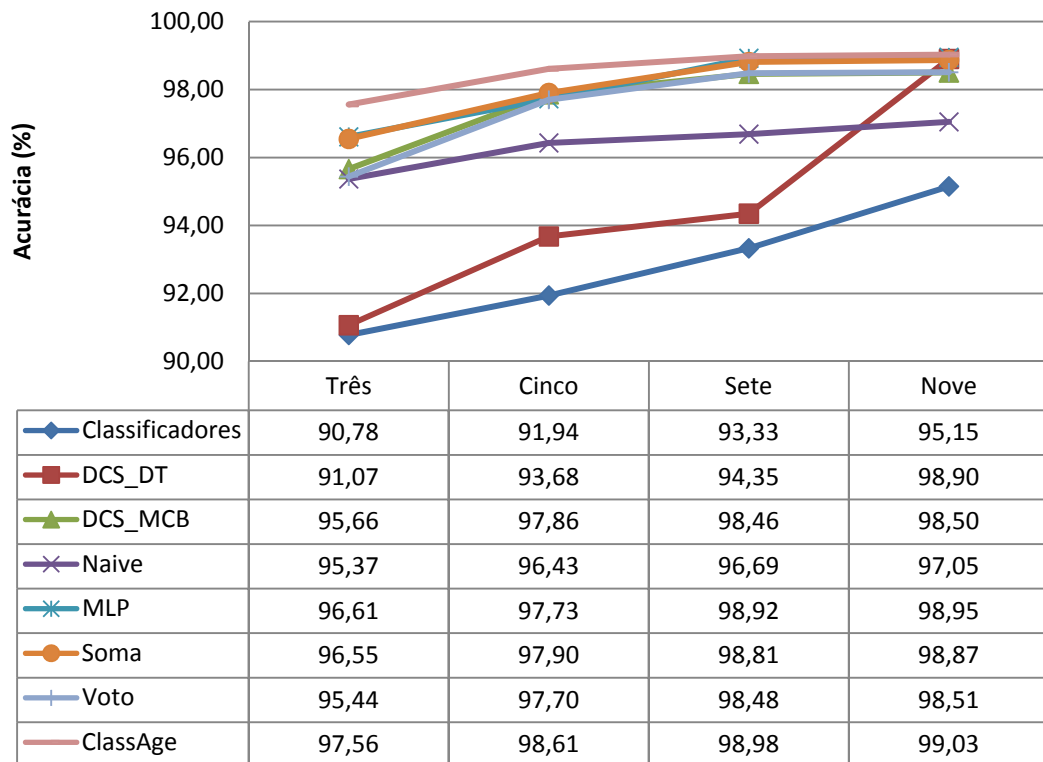


Figura 3: Resultados dos sistemas de classificação sem distribuição para a base de imagem analisando a quantidade de componentes

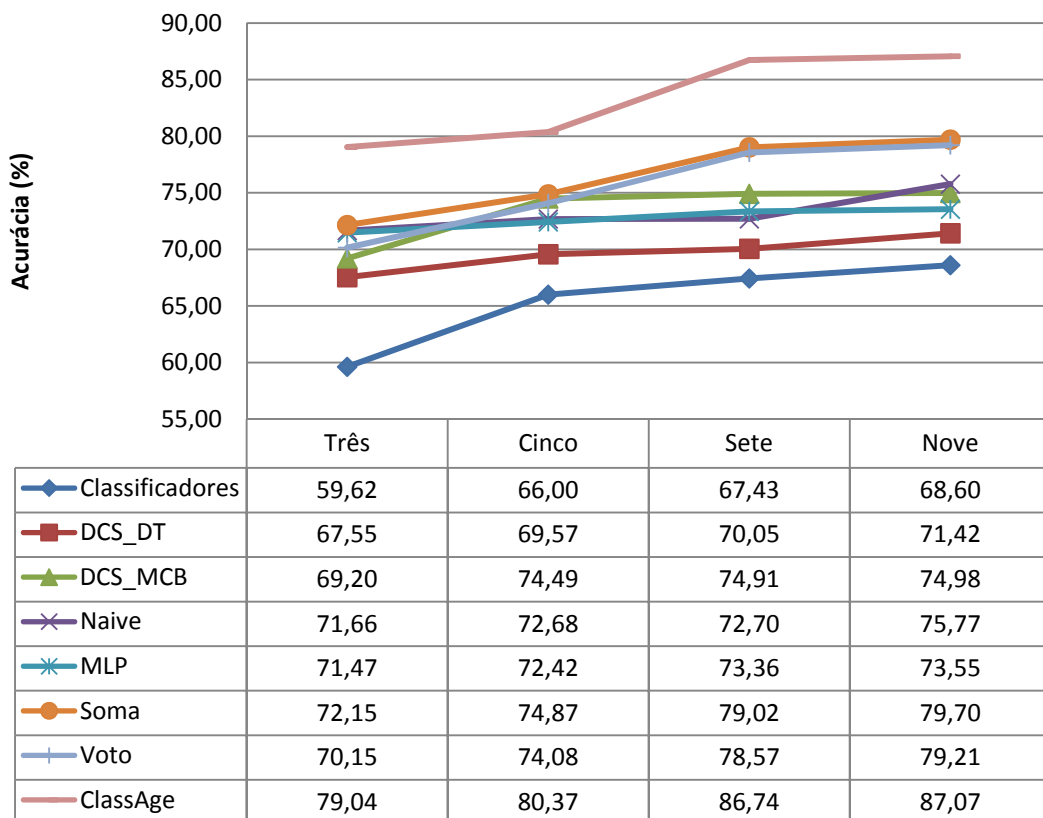


Figura 4: Resultados dos sistemas de classificação sem distribuição para a base de proteínas analisando a quantidade de componentes

Os resultados apresentados nas Figuras 3 e 4 mostram que os sistemas combinadores (comitês e *ClassAge*) conseguem um desempenho superior aos classificadores base. Este resultado é esperado, já que os sistemas de combinação de classificadores trabalham com informações complementares sobre o padrão, de acordo com o aprendizado de cada classificador.

Os sistemas que apresentaram melhores resultados foram: O *ClassAge* nas duas bases estudadas e os sistemas baseados em fusão – MLP na base de imagem e Soma na base de proteínas. O sistema com resultados mais baixos foi o sistema baseado em seleção – DCS_DT nas duas bases experimentadas.

Analisando a acurácia dos sistemas em relação à quantidade de componentes, pode ser observado que o desempenho melhora com o aumento de componentes, pois aumenta-se a diversidade no sistema.

As pesquisas na área de aprendizado de máquina mostram que estes são os resultados esperados. Pretende-se verificar se esse comportamento será mantido nos experimentos feitos com distribuição dos atributos.

6.1.2 Sistemas com distribuição parcial das características

As Figuras 5 e 6 apresentam os resultados dos sistemas de combinação com distribuição parcial das características, analisando a quantidade de componentes do sistema para as bases de imagem e proteínas, respectivamente.

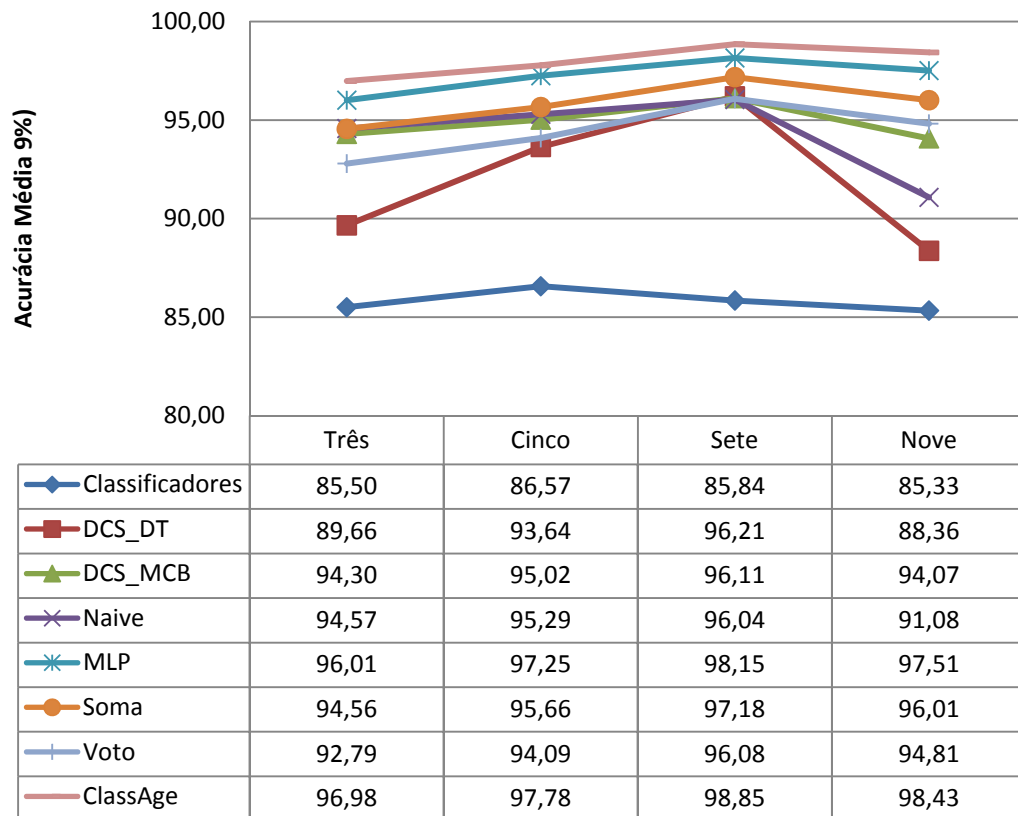


Figura 5: Resultados dos sistemas de classificação com distribuição parcial para a base de imagem analisando a quantidade de componentes

A Figura 5 mostra que os sistemas de combinação de classificadores apresentam desempenho superior aos classificadores base e que o desempenho dos sistemas tende a aumentar com o aumento na quantidade de componentes, à exceção dos sistemas com 9 componentes. A diminuição na acurácia dos sistemas com 9 componentes deve-se a pequena quantidade de informação utilizada pelos classificadores base, já que a base de imagem possui apenas 18 atributos. Quando se distribui os 18 atributos entre 9 componentes, o número de características utilizadas por cada classificador base torna-se pequeno.

Verifica-se, ainda, que os sistemas de combinação que conseguiram melhores desempenhos nas 3 configurações de tamanho foram o *ClassAge* e o sistema baseado em fusão – MLP. Já o sistema baseado em seleção DCS_DT, obteve os desempenhos mais próximos aos dos classificadores base.

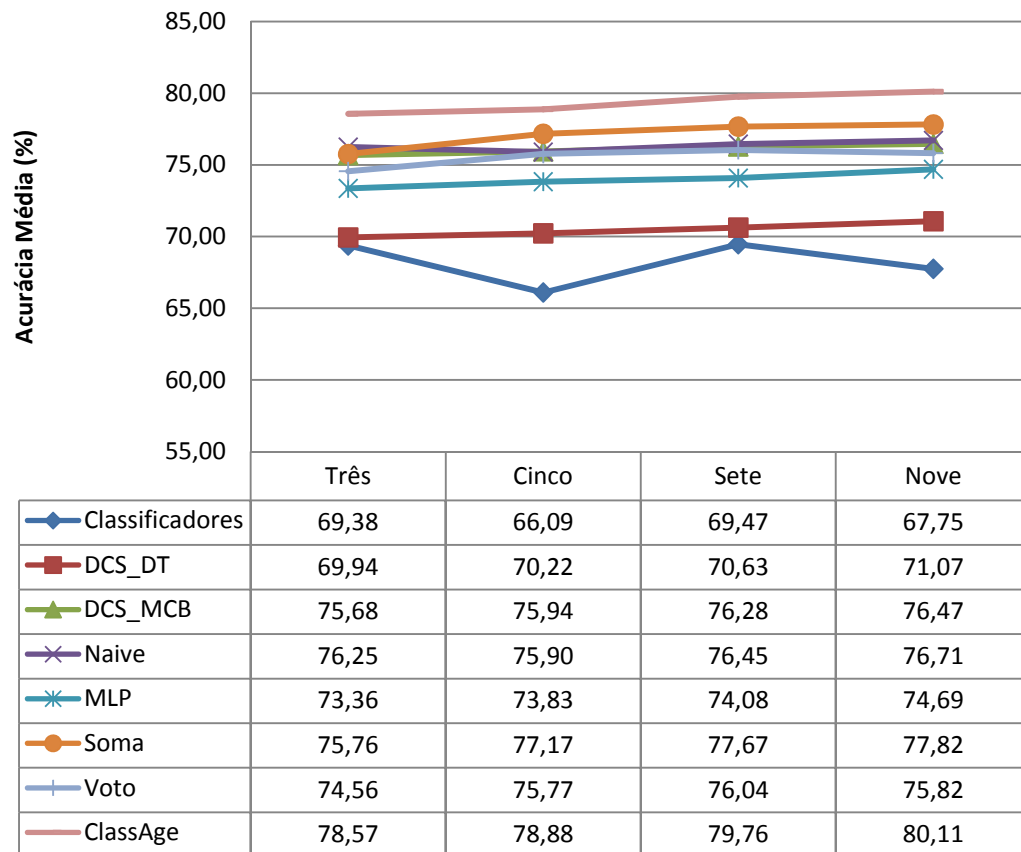


Figura 6: Resultados dos sistemas de classificação com distribuição parcial para a base de proteínas analisando a quantidade de componentes

Assim como para a base de imagem, os resultados da base de proteínas apresentados na Figura 6 mostram que os sistemas de combinação de classificadores obtêm desempenho superior aos classificadores base. Nestes resultados, pode-se verificar que os sistemas de combinação, em geral, tiveram aumento na acurácia com o aumento da quantidade de classificadores, à exceção do *Naive* e do *Voto* que tiveram um decréscimo na acurácia entre 3 e 5 e entre 7 e 9 componentes, respectivamente.

Os sistemas que conseguiram melhores desempenhos foram o *ClassAge* e o sistema baseado em fusão – *Soma*. Já o sistema *DCS_DT*, assim como para a base de imagem, foi o que obteve os resultados mais próximos aos dos classificadores base.

6.1.3 Sistemas com distribuição total das características

As Figuras 7 e 8 apresentam os resultados dos sistemas de combinação de classificadores com distribuição total das características para as bases de imagem e proteínas, respectivamente, analisando a quantidade de componentes dos sistemas.

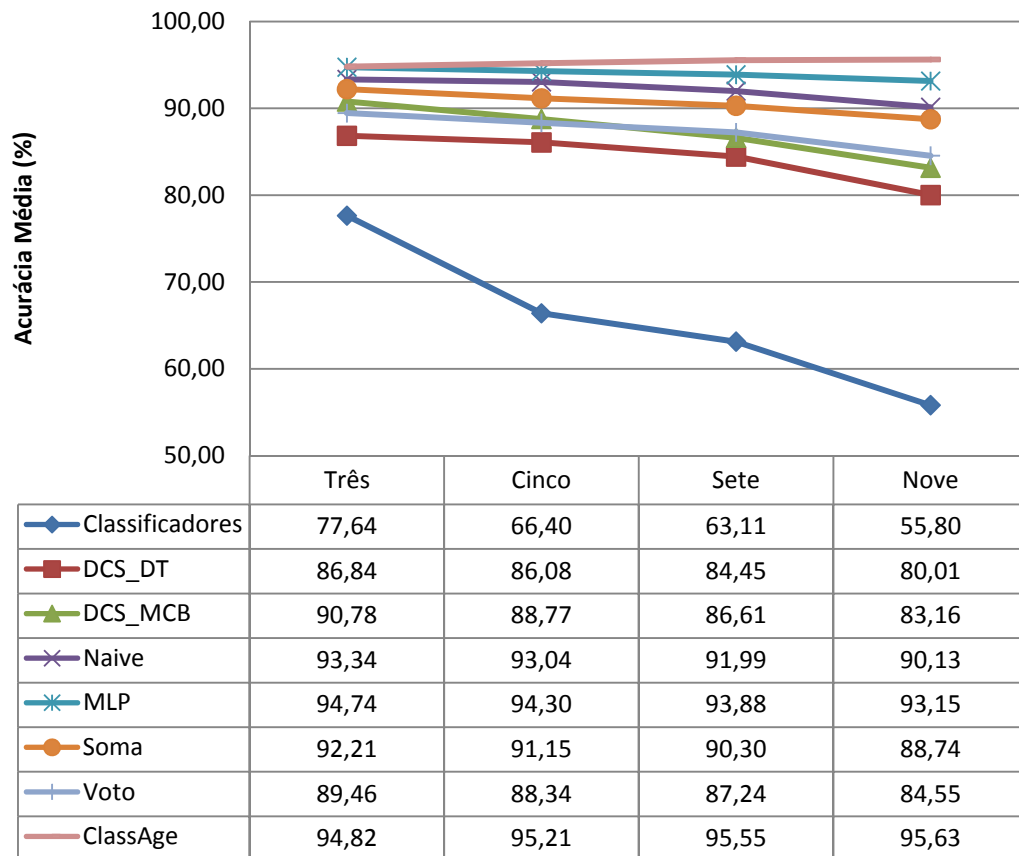


Figura 7: Resultados dos sistemas de classificação com distribuição total para a base de imagem analisando a quantidade de componentes

Os resultados apresentados na Figura 7 mostram que os sistemas de combinação conseguem desempenho superior aos classificadores base e que com o aumento de componentes, os sistemas perderam em desempenho. Ou seja, quanto maior a quantidade de componentes no sistema, menor a acurácia, à exceção do *ClassAge* que conseguiu melhorar a acurácia apesar da diminuição de informação.

Esse comportamento deve-se a pequena quantidade de informação utilizada pelos componentes para classificar os padrões, já que na distribuição total os atributos são distribuídos sem interseção entre os componentes. Além disso, a base de imagem tem o agravante de possuir uma quantidade total de atributos pequena.

Do mesmo modo que a distribuição parcial, os sistemas de combinação que apresentaram melhores desempenhos com distribuição total dos atributos na base de imagem foram o *ClassAge* e o *MLP*. O sistema *DCS_DT* apresentou os resultados mais próximos aos classificadores base.

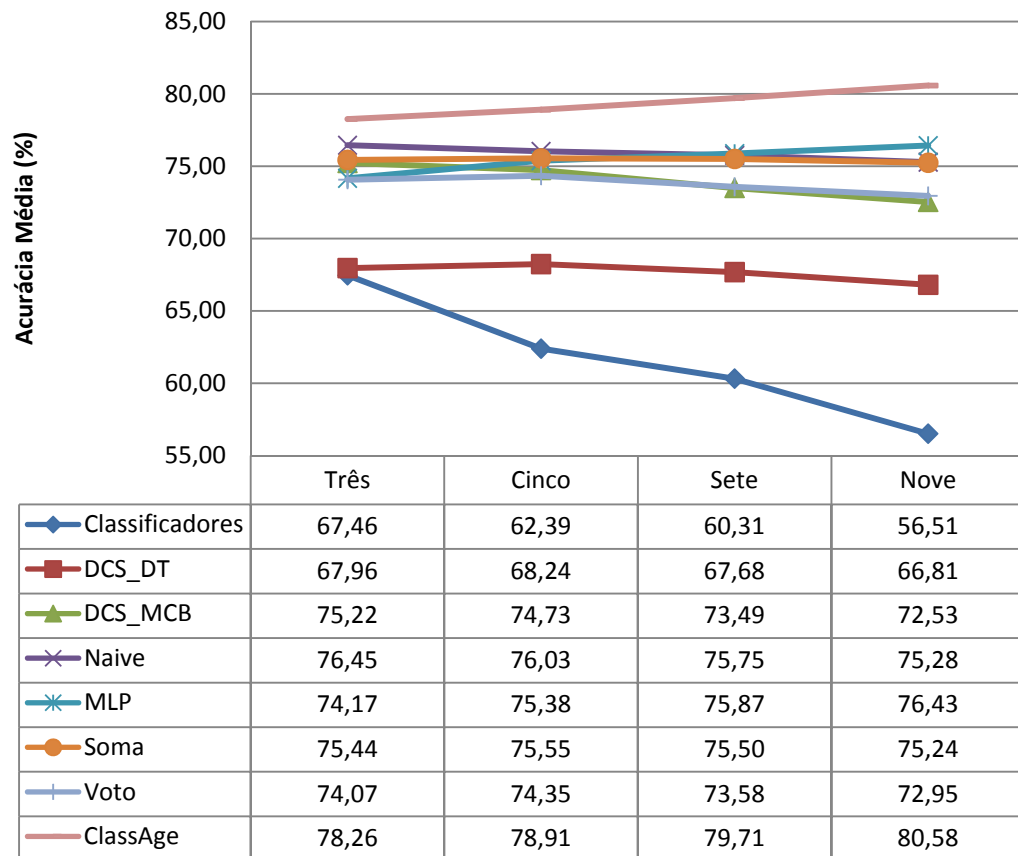


Figura 8: Resultados dos sistemas de classificação com distribuição total para a base de proteínas analisando a quantidade de componentes

Os resultados para a base de proteínas com distribuição total mostram que os sistemas de combinação conseguem desempenho superior ao dos classificadores base. Em relação à quantidade de componentes, percebe-se 3 grupos com comportamentos distintos: O primeiro, formado pelos classificadores base, pelo DCS_MCB e pelo *Naive*, diminui a acurácia com o aumento de componentes; o segundo, formado por DCS_DT, Soma e Voto, aumenta o desempenho de 3 para 5 componentes e diminui de 5 para 9; o terceiro grupo é formado pelo *ClassAge* e MLP e aumenta o desempenho com o aumento de componentes.

Pode-se verificar que o *ClassAge* consegue os melhores resultados, seguido pelos sistemas baseados em fusão. Os sistemas baseados em seleção aparecem com os resultados mais baixos entre os combinadores, principalmente com 7 e 9 componentes.

6.1.4 Distribuição Parcial versus Distribuição Total

Nesta subsecção será feita uma análise comparativa entre as distribuições parcial e total. Para isto, será analisada a variação média de desempenho dos sistemas quando passam da distribuição parcial para a total.

A variação média é calculada subtraindo-se a acurácia do sistema com distribuição parcial da acurácia do sistema com distribuição total. Essa variação mede a sensibilidade do sistema à diminuição de características, já que na distribuição total a quantidade de atributos analisados é menor. Ou seja, quanto maior o valor da variação, maior a queda de desempenho quando o sistema passa da distribuição parcial para a total. Variações negativas indicam que o desempenho do sistema foi superior com a distribuição total.

Em outras palavras, com o cálculo da variação pode-se afirmar se o sistema tem melhor desempenho com a distribuição parcial ou total e quão grande é a diferença nesse desempenho.

As Figuras 9 e 10 analisam a variação sob a ótica do tamanho do sistema.

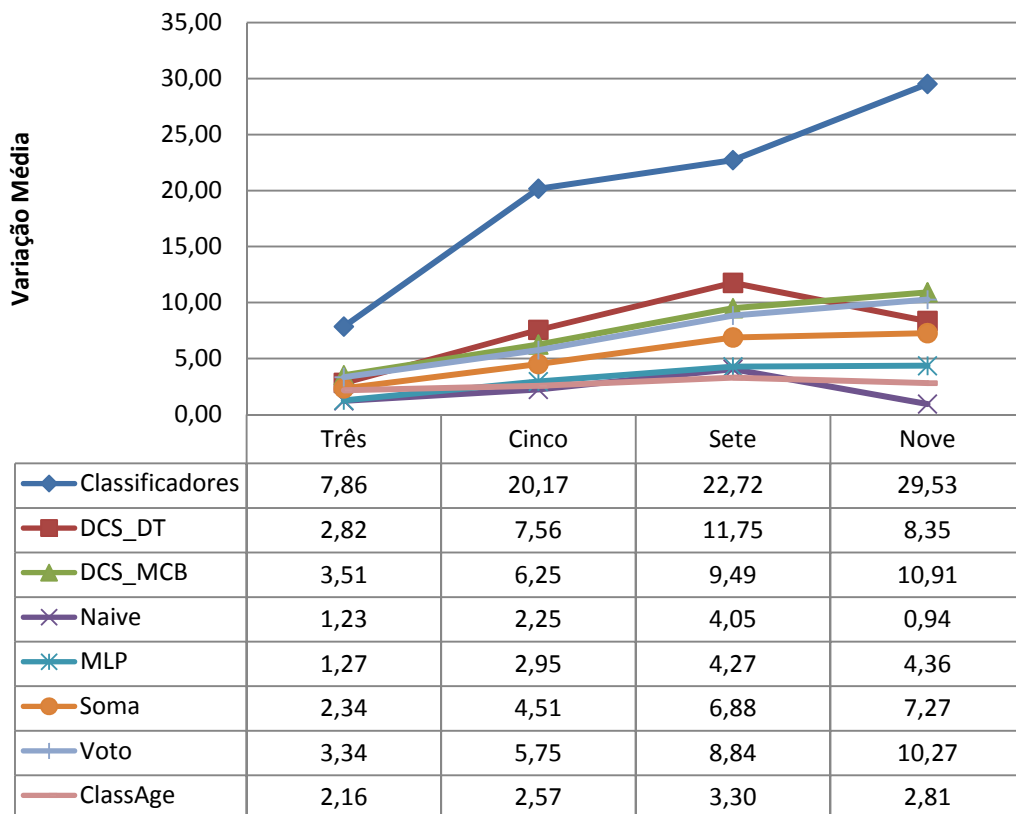


Figura 9: Variação média entre as distribuições parcial e total para a base de imagem, analisando a quantidade de componentes dos sistemas

Pode-se observar pelos resultados apresentados na Figura 9 que os classificadores base possuem uma variação alta, o que indica alta sensibilidade à quantidade de informação usada no processo de classificação. Isto é, com a distribuição total e conseqüente diminuição dos atributos analisados, seu desempenho cai consideravelmente.

Para a base de imagem, de acordo com os valores sempre positivos da variação, verifica-se que a distribuição parcial promove melhores resultados em relação à distribuição total. Os sistemas de combinação que apresentaram menor variação foram: *Naive* e *ClassAge*. Isso quer dizer que esses sistemas têm menor sensibilidade à diminuição de atributos.

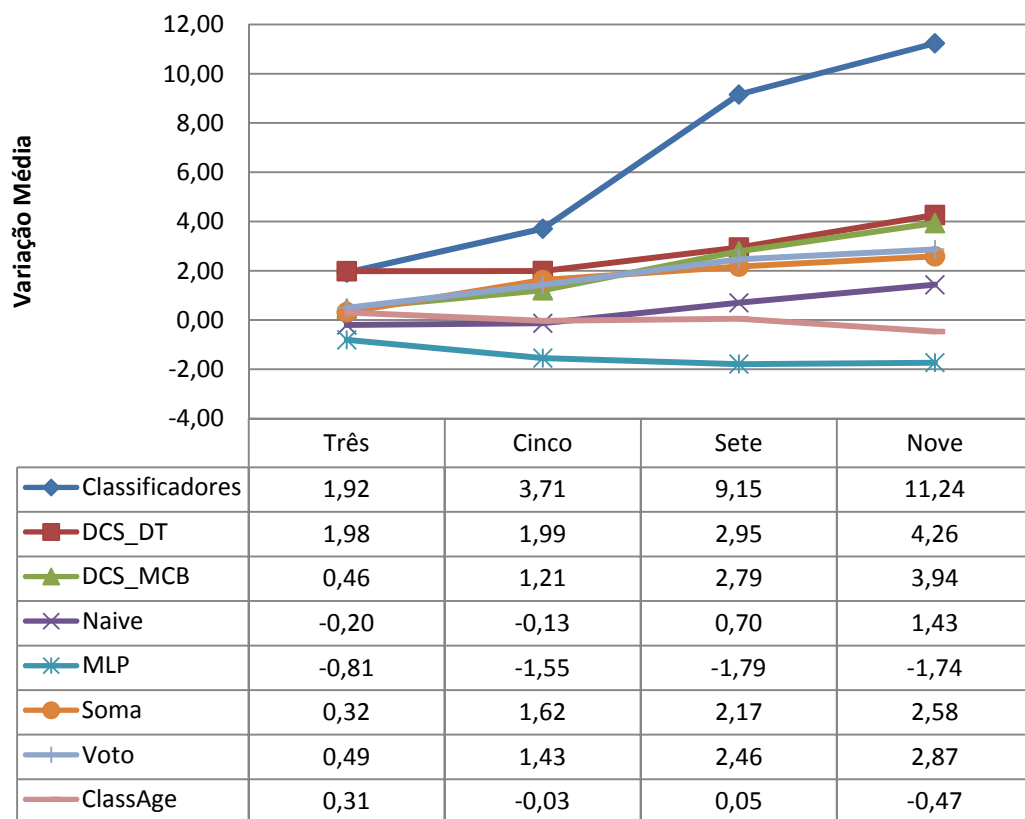


Figura 10: Variação média entre as distribuições parcial e total para a base de proteínas, analisando a quantidade de componentes dos sistemas

Já na base de proteínas, a variação negativa em alguns casos, mostra que a distribuição total dos atributos pode trazer melhores resultados. Isso porque, a base de proteínas possui muitos atributos e alguns deles redundantes, o que prejudica a classificação. Quando separados, o sistema consegue melhor desempenho.

O sistema de combinação MLP apresentou variações negativas em todos os tamanhos, mostrando que trabalha melhor, para o ambiente proposto pela base de proteínas, com uma quantidade menor de atributos e conseqüente diminuição de atributos redundantes.

É interessante observar que essa informação mostra que este sistema de combinação é mais sensível à qualidade dos dados que os outros sistemas, já que obteve melhores resultados separando-se os atributos redundantes.

Pode-se verificar ainda que na maioria dos casos a variação aumenta com o aumento de componentes, isso indica que a distribuição total deve ser evitada em sistemas com muitos componentes.

6.2 Análise em Relação à Técnica de Distribuição dos Dados

As Figuras de 11 a 18 apresentam os resultados dos sistemas de combinação analisando as técnicas de distribuição de atributos. Os valores apresentados nesta seção representam a média da acurácia dos sistemas nas 4 configurações de tamanho (3, 5, 7 e 9 componentes).

Esta seção está dividida em 3 subseções, na primeira são apresentados os resultados dos sistemas com distribuição parcial das características e na subseção 6.2.2 os resultados dos sistemas com distribuição total. Na última subseção, 6.2.3, é feita uma análise comparativa dos resultados das distribuições parcial e total.

6.2.1 Sistemas com distribuição parcial das características

As Figuras 11 e 12 apresentam os resultados dos sistemas de combinação e dos classificadores base com distribuição parcial das características analisando as técnicas de distribuição aplicadas para as bases de dados de imagem e proteínas, respectivamente.

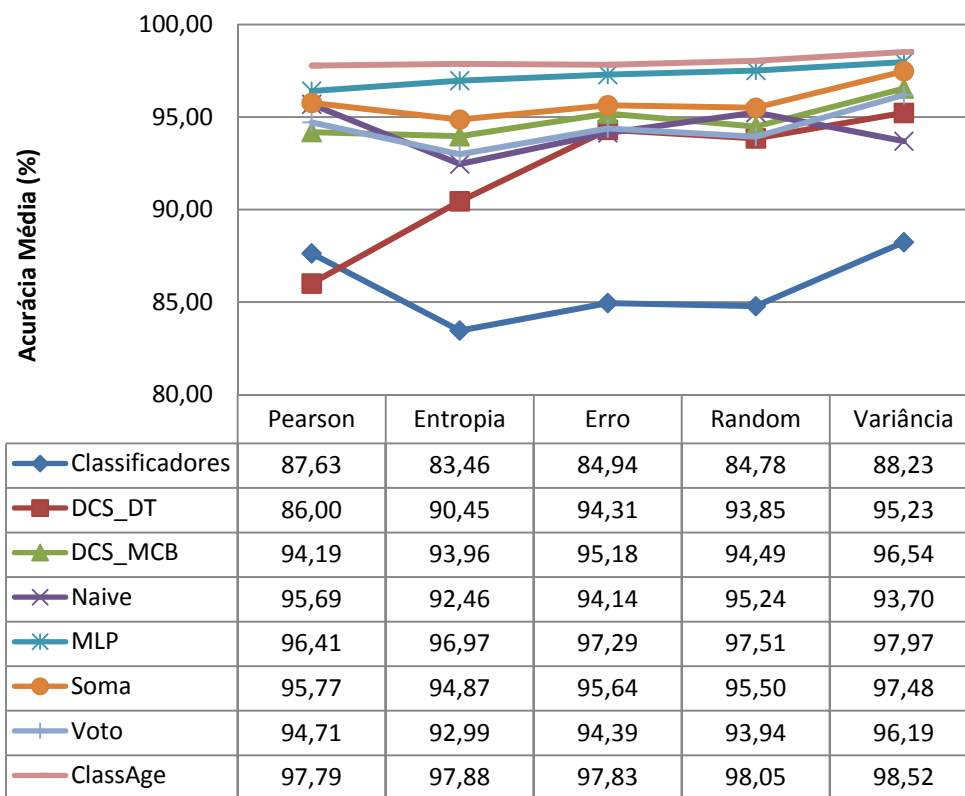


Figura 11: Resultados dos sistemas de classificação com distribuição parcial para a base de imagem analisando as técnicas de distribuição

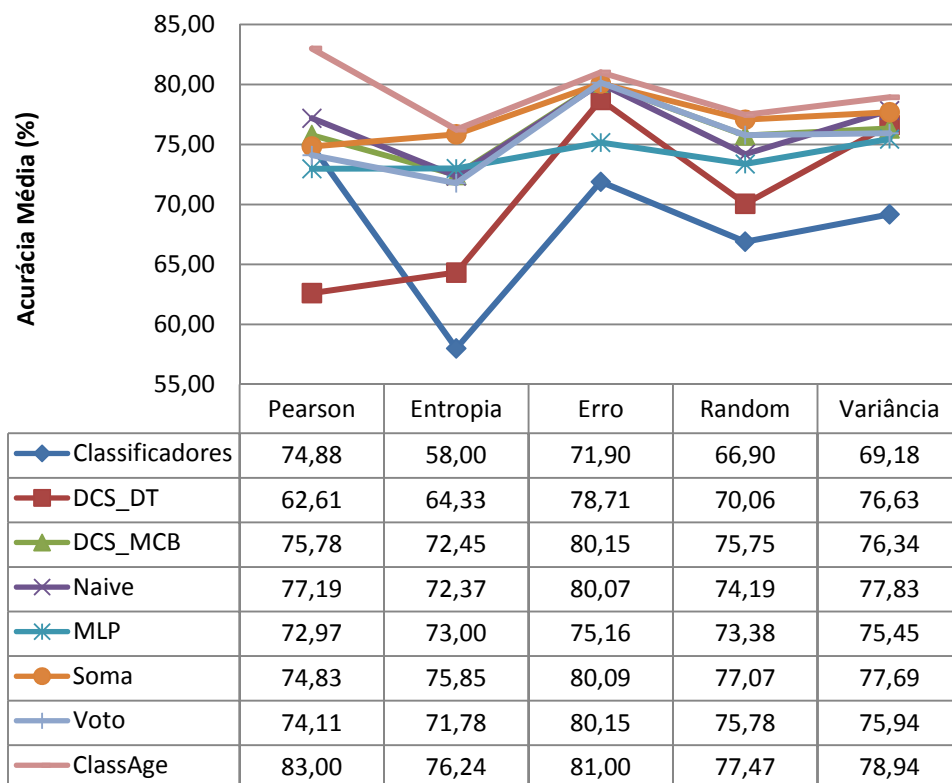


Figura 12: Resultados dos sistemas de classificação com distribuição parcial para a base de proteínas analisando as técnicas de distribuição

Dos resultados apresentados nas Figuras 11 e 12, pode-se verificar que apenas na técnica de distribuição baseada em Correlação de Pearson o comportamento não foi o esperado, já que alguns sistemas combinadores tiveram desempenho inferior ao dos classificadores base. Esses sistemas foram: DCS_DT nas duas bases experimentadas e os sistemas baseados em fusão – MLP, Soma e Voto – na base de proteínas.

Os sistemas que apresentaram melhores resultados foram: O *ClassAge* nas duas bases estudadas e os sistemas baseados em fusão – MLP na base de imagem e Soma na base de proteínas.

A Figura 13 mostra uma média geral das acurácias dos sistemas para cada técnica de distribuição parcial a fim de verificar qual técnica teve melhor desempenho.

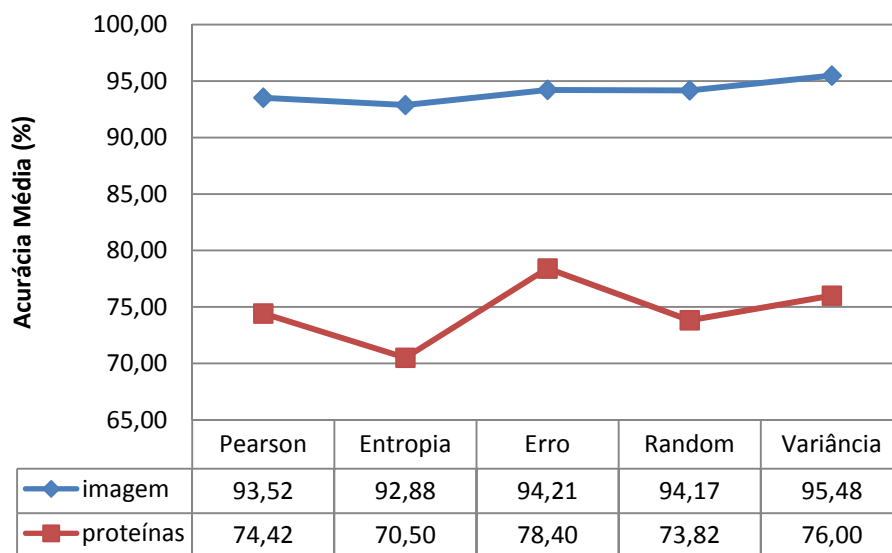


Figura 13: Comparação entre as técnicas de distribuição parcial

Verifica-se que no ambiente da base de imagem os melhores resultados, de modo geral, foram conseguidos usando a técnica baseada na Variância, enquanto que na base de proteínas a técnica que melhor se adequou ao ambiente de difícil classificação foi a técnica do Erro.

É interessante observar que as duas bases estudadas representam ambientes diferentes de classificação. Como já mencionado, a base de imagem representa um ambiente de classificação menos propenso a erros, por ser balanceada em relação às classes, por ter atributos relevantes e não redundantes e uma boa quantidade de exemplos. Já a base de proteínas representa um ambiente de difícil classificação, por ser

desbalanceada, possuir poucos exemplos e atributos redundantes e não relevantes. Os resultados mostram quais técnicas são mais adequadas em cada caso, ou seja, para situações onde o ambiente de classificação é difícil, por complicações da base de dados, a técnica do Erro mostra-se mais apropriada. Já para bases de dados mais adequadas à classificação a técnica da Variância apresenta melhores resultados.

A Figura 13 mostra que a distribuição parcial usando a técnica Random, ou seja, feita de maneira aleatória traz resultados medianos. A base de imagem apresentou uma diferença pequena entre a técnica aleatória e a melhor técnica, Variância, de 1,31. Já a base de proteínas mostrou-se mais sensível à forma como os atributos são distribuídos, sendo a diferença entre a melhor técnica, Erro, e a técnica aleatória de 2,18.

6.2.2 Sistemas com distribuição total das características

As Figuras 14 e 15 apresentam os resultados dos sistemas de combinação e dos classificadores base com distribuição total das características analisando as técnicas de distribuição aplicadas para as bases de dados de imagem e proteínas, respectivamente.

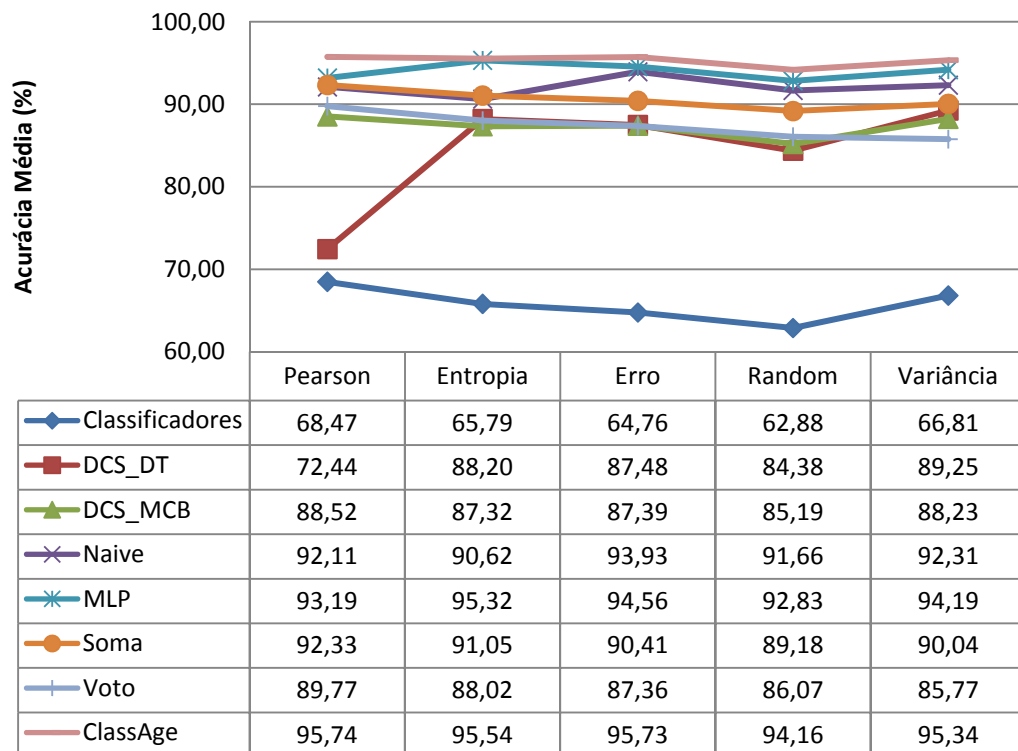


Figura 14: Resultados dos sistemas de classificação com distribuição total para a base de imagem analisando as técnicas de distribuição

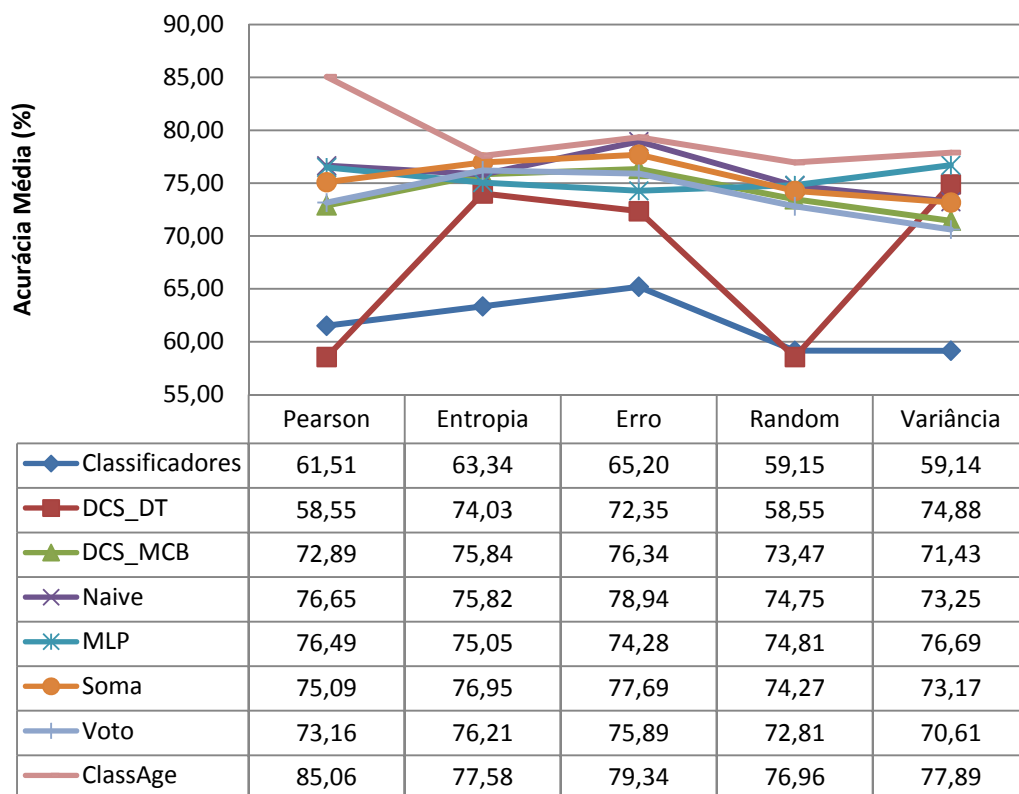


Figura 15: Resultados dos sistemas de classificação com distribuição total para a base de proteínas analisando as técnicas de distribuição

Dos resultados apresentados em relação aos sistemas de classificação, pode-se extrair as seguintes informações: Como esperado, os sistemas de combinação apresentam melhores resultados que os classificadores base, a exceção do DCS_DT que repetiu o comportamento apresentado na distribuição parcial, ou seja, não conseguiu manter-se sempre acima do desempenho dos individuais. A diferença em relação ao caso parcial, foi que esse comportamento foi constatado apenas na base de proteínas e ocorreu nas técnicas de distribuição baseadas em Correlação de Pearson e Random.

Os resultados mostram que o *ClassAge* apresenta melhor desempenho em relação aos demais, e que os sistemas baseados em fusão conseguem acurácias mais elevadas que os sistemas baseados em seleção. Este resultado confirma o comportamento apresentado na distribuição parcial.

Em relação à técnica de distribuição, a Figura 16 apresenta a média geral das acurácias dos sistemas para cada técnica de distribuição total. Os resultados confirmam que a técnica do Erro é a mais adequada para a base de proteínas, e a técnica da Variância para

a base de imagem, sendo que na distribuição total, a diferença desta técnica para as demais é pequena.

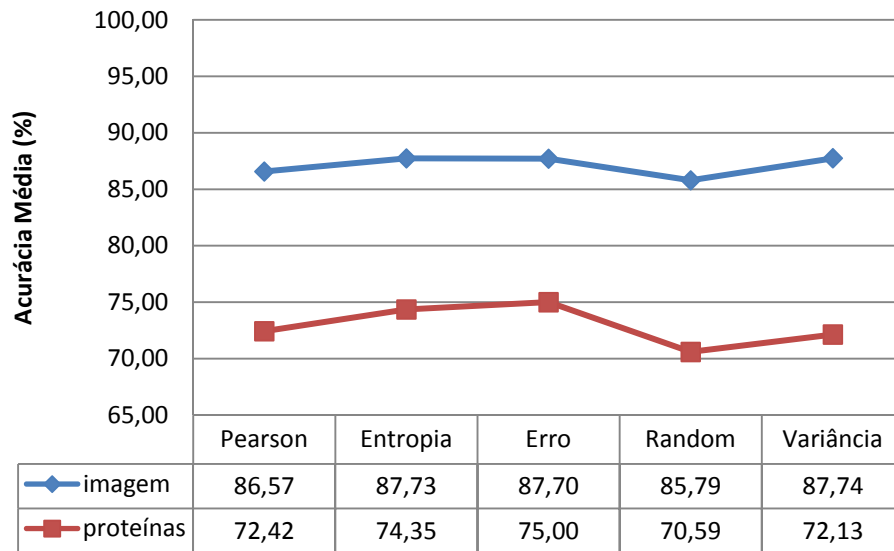


Figura 16: Comparação entre as técnicas de distribuição total

Na distribuição total, a técnica Random apresenta os resultados mais baixos, mostrando que para esse tipo de distribuição a forma de separar os atributos entre os componentes do sistema é relevante. Verifica-se uma diferença de 1,95 entre a técnica aleatória e a melhor técnica para a base de imagem e de 4,41 com a técnica Random e a melhor técnica para a base de proteínas.

6.2.3 Distribuição Parcial *versus* Distribuição Total

As Figuras 17 e 18 apresentam a variação média de desempenho dos sistemas quando passam da distribuição parcial para a total nas bases de imagem e proteínas, respectivamente, analisando as técnicas de distribuição de dados aplicadas.

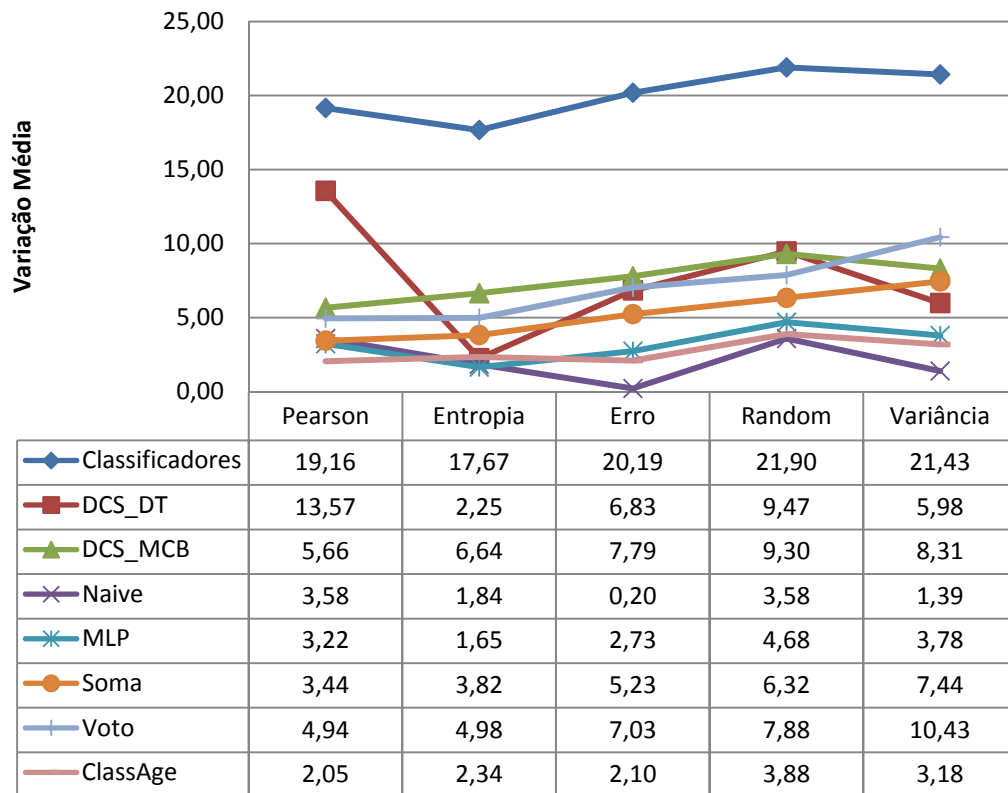


Figura 17: Variação média entre as distribuições parcial e total para a base de imagem, analisando as técnicas de distribuição

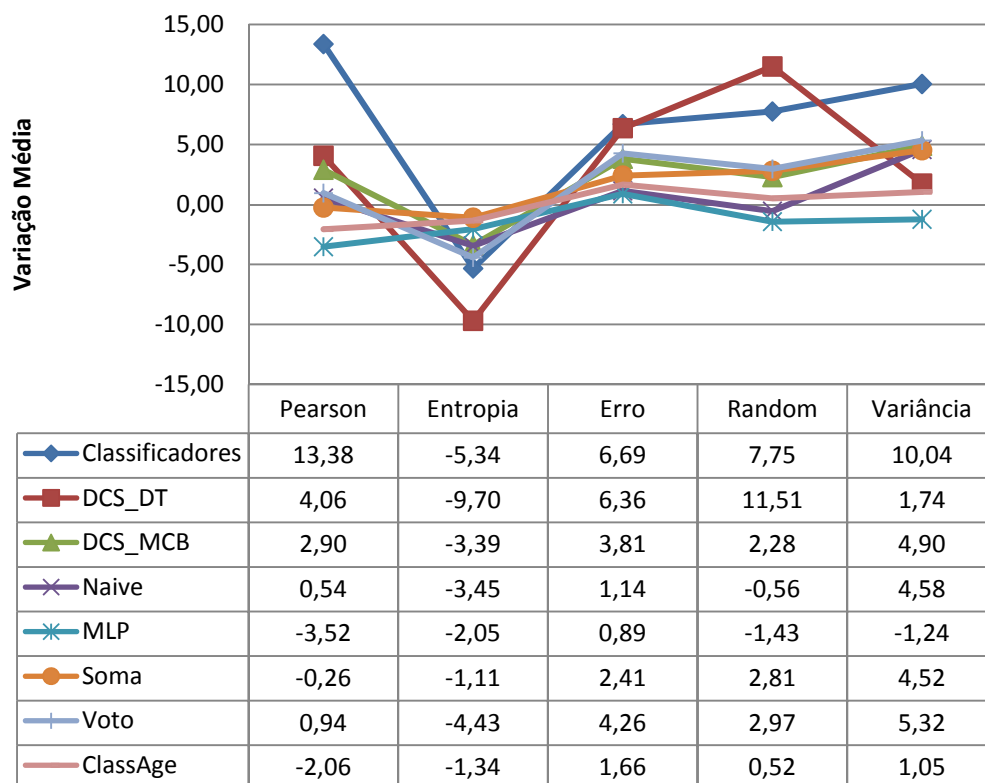


Figura 18: Variação média entre as distribuições parcial e total para a base de proteínas, analisando as técnicas de distribuição

Pode-se observar pelos resultados apresentados que os sistemas com classificadores base possuem uma variação alta, o que indica alta sensibilidade à quantidade de informação usada no processo de classificação. Isto é, com a distribuição total e conseqüente diminuição dos atributos analisados, seu desempenho cai consideravelmente.

Para a base de imagem, de acordo com os valores sempre positivos da variação, verifica-se que a distribuição parcial promove melhores resultados em relação à distribuição total. Os sistemas de combinação que apresentaram menor variação foram: *Naive* e *ClassAge*. Isso quer dizer que esses sistemas têm menor sensibilidade à diminuição de atributos.

Já na base de proteínas, a variação negativa em alguns casos, mostra que a distribuição total dos atributos pode trazer melhores resultados. Isso porque, a base de proteínas possui muitos atributos e alguns deles redundantes, o que prejudica a classificação. Quando separados, o sistema consegue melhor desempenho.

Os casos em que a distribuição total teve melhor desempenho em relação à distribuição parcial foram para todos os sistemas usando a técnica da Entropia e para o sistema de combinação MLP na maioria dos casos. Ou seja, a técnica da Entropia e o sistema de combinação MLP trabalham melhor, para o ambiente proposto pela base de proteínas, com uma quantidade menor de atributos e conseqüente diminuição de atributos redundantes.

É interessante observar que essa informação mostra que a técnica de distribuição baseada na Entropia e o sistema de combinação usando MLP são mais sensíveis à qualidade dos dados que os outros sistemas, já que obtiveram melhores resultados separando-se os atributos redundantes.

6.3 Análise em relação ao uso de distribuição de dados

Nesta seção será apresentado um resultado geral do desempenho dos sistemas sem distribuição dos atributos e com distribuição parcial e total.

A Figura 19 mostra a média geral das acurácias em cada distribuição para as duas bases de dados. A base de imagem apresenta melhores resultados com uma quantidade maior de atributos analisados, já que a acurácia média diminui com a distribuição parcial e

total. Nota-se, porém, que a diferença de desempenho sem distribuição e com distribuição parcial é consideravelmente inferior a diferença entre os sistemas sem distribuição e com distribuição total. Enquanto no primeiro caso a diferença é de 2,63 o segundo é de 9,57.

O ambiente proposto pela base de proteínas não altera-se de modo relevante com a distribuição das características, verifica-se pela Figura 19 que a diferença entre as distribuições é pequena e que a distribuição parcial pode ser uma boa alternativa para esse tipo de base de dados, com muitos atributos.

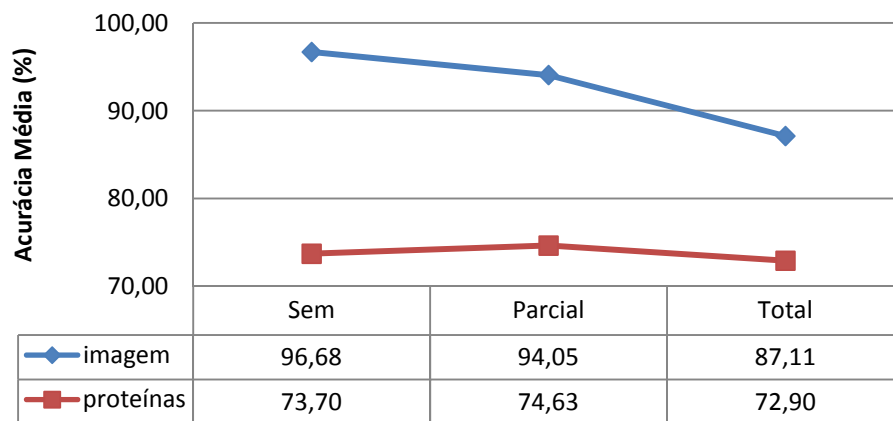


Figura 19: Comparação entre as distribuições

Capítulo 7

Considerações Finais

Neste trabalho foi realizado um estudo sobre diferentes sistemas de combinação de classificadores sujeitos à distribuição vertical dos dados.

Esse estudo foi motivado pela existência de um cenário distribuído de dados em aplicações de classificação de padrões, seja como uma opção na busca por diversidade ou por exigência da aplicação – quando a distribuição é intrínseca ao problema ou para manter a privacidade dos dados.

Foram comparados os desempenhos de sete sistemas de combinação, sendo seis comitês e um multiagente, o *ClassAge*. Seus desempenhos foram analisados quando submetidos à distribuição parcial e total dos atributos de entrada. A distribuição foi feita usando cinco técnicas diferentes: Correlação de Pearson; Entropia; Erro; Random e Variância. Nesta investigação, foram analisados sistemas com 3, 5, 7 e 9 componentes.

Foram realizados experimentos sem distribuição de características e experimentos usando apenas os classificadores base dos sistemas para servir de base no estudo sobre o comportamento dos sistemas de combinação de classificadores com distribuição dos atributos.

O objetivo deste trabalho foi investigar quais sistemas de combinação são mais apropriados à distribuição vertical dos dados; se esse comportamento se mantém com o aumento de componentes e se é independente da forma como os atributos foram distribuídos; que técnicas podem ser usadas com maior sucesso para separar os dados entre os classificadores e se há vantagens em utilizar técnicas de distribuição não-aleatória.

As seguintes respostas foram encontradas:

- De modo geral, quando submetidos à distribuição, parcial ou total, os sistemas de combinação de classificadores obtêm melhores resultados que os sistemas individuais.

- Os sistemas que tiveram melhor desempenho sem distribuição, com distribuição parcial e com distribuição total, independente da quantidade de componentes e da técnica utilizada para distribuição dos atributos, foram: *ClassAge* e o sistema baseado em fusão – MLP para a base de imagem; *ClassAge* e o sistema baseado em fusão – Soma para a base de proteínas. Na distribuição total para a base de proteínas os sistemas de combinação *Naive* e MLP tiveram desempenho semelhante ao da Soma.
- O desempenho dos sistemas sem distribuição e com distribuição parcial tende a aumentar com o aumento de componentes. A exceção a este comportamento ocorreu nos sistemas com 9 componentes para a base de imagem com distribuição parcial, devido a pouca quantidade de informação e nos sistemas baseados em fusão, *Naive* e Voto, na base de proteínas com distribuição parcial, que não apresentaram um comportamento homogêneo. Já os sistemas com distribuição total, de modo geral, têm o desempenho diminuído com o aumento de componentes, devido a diminuição de informação para classificação, à exceção do *ClassAge* nas duas bases experimentadas e do MLP na base de proteínas, que aumentaram o desempenho com o aumento de componentes.
- A técnica de distribuição baseada na Variância obteve os melhores resultados na base de imagem, enquanto que a técnica do Erro conseguiu os melhores resultados na base de proteínas. Esse comportamento foi verificado para os sistemas com distribuição parcial e total.
- Para os sistemas com distribuição parcial, o uso de uma técnica aleatória influenciou menos no desempenho que nos sistemas com distribuição total. Nestes, a técnica Random apresentou os resultados mais baixos, mostrando que a técnica utilizada na separação total das características entre os componentes é relevante.

Os resultados obtidos são importantes na construção de sistemas de combinação com distribuição vertical dos dados, pois mostra que em ambientes com muitos atributos e poucos exemplos, pode-se utilizar a distribuição parcial das características como uma opção para aumentar o desempenho do sistema. Essa distribuição parcial pode ser feita com uma quantidade maior de componentes, 7 ou 9 e apresentará bons resultados usando a técnica do Erro na distribuição dos atributos. Já em ambientes com poucas características e muitos exemplos, deve-se optar por uma quantidade menor de componentes e bons

resultados são conseguidos usando a técnica da Variância para fazer uma distribuição parcial dos atributos.

7.1 Trabalhos Futuros

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

- Utilizar outras bases de dados:
 - Bases mais complexas;
 - Bases desbalanceadas;
 - Bases com alta dimensionalidade de atributos;
- Utilizar outras técnicas de distribuição de atributos:
 - Técnicas que utilizam algoritmos genéticos para definir a melhor distribuição de atributos entre os componentes do sistema, por exemplo.
- Aumentar a quantidade de componentes dos sistemas.

Referências Bibliográficas

Abreu, M. (2006). *Analisando o desempenho do ClassAge: Um sistema multiagente para classificação de padrões*. Dissertação de mestrado, UFRN, DIMAp, Natal.

Abreu, M., Canuto, A. & Souto, M. (2005a). Uma análise comparativa de sistemas de classificação usando uma base de estrutura de proteínas. *Congresso Brasileiro de Redes Neurais*. Natal.

Abreu, M., Canuto, A., Souto, M. & Bittencourt, V. (2005). Aplicação de Multiclassificadores Heterogêneos no Reconhecimento de Classes Estruturais do Proteínas. *Congresso Brasileiro de Redes Neurais*. Natal.

Agrawal, R. & Srikant, R. (2000). Privacy-preserving data mining. *Conference on Management of Data* (pp. 439–450). Dallas: ACM Press.

Ben-Yacoub, S., Abdeljaoued, Y. & Mayoraz, E. (2001). Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Network*, 10 (5), pp. 1065-1074.

Blake, C. & Merz, C. (s.d.). *UCI Machine Learning Repository*. Acesso em 01 de 01 de 2008, disponível em <http://www.ics.uci.edu/~mlearn/MLRepository.html>

Campos, T. (2000). *Técnicas de seleção de atributos e de classificação para reconhecimento de faces*. Dissertação de mestrado, USP, DCC, São Paulo.

Canuto, A. (2004). Investigating the Use of an Agent-Based Multi-classifier System for Classification Tasks. *Lecture Notes on Computer Science*, 3316, pp. 854-859.

Caragea, D., Silvescu, A. & Honavar, V. (2003). Decision tree induction from distributed, heterogeneous, autonomous data sources. *Conference on Intelligent Systems Design and Applications* (pp. 341-350). Berlim: Springer Verlag.

Chen, D., Sirlantzis, K., Hua, D. & Ma, X. (2005). On the relation between dependence and diversity in multiple classifier systems. *International Conference on Information Technology: Coding and Computing* (pp. 134–139).

- Chen, K., Wang, L. & Chi, H. (1997). Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification. *International Journal on Pattern Recognition and Artificial Intelligence* , 11 (3), pp. 417-445.
- Dietterich, T. (1997). Machine learning research: four current directions. *AI Magazine* , 18 (1), 97-136.
- Guerra-Salcedo, C. & Whitley, D. (1999). Genetic Approach to feature selection for ensemble. *Genetic and Evolutionary Computation Conference* (pp. 236-243).
- Ho, T. (1998). The random space method for constructing decision forest. *IEEE transaction on pattern analysis and machine intelligence* , 20 (8), pp. 832-844.
- Kargupta, H., Park, B., Hershberger, D. & Johnson, E. (2000). Collective data mining: A new perspective toward distributed data mining. In: H. Kargupta & P. Chan, *Advances in Distributed and Parallel Knowledge Discovery* (pp. 133-184). Menlo Park: AAAI/ MIT Press.
- Kuncheva, L. (2004). *Combining Pattern Classifiers: methods and algorithms*. JOHN WILEY and SONS.
- Kuncheva, L. & Jain, L. (2000). Designing classifier fusion systems by genetic algorithms. *IEEE Transactions on Evolutionary Computation* , 4 (4), pp. 327-336.
- Lee, H., Monard, M. & Wu, F. (2005). Seleção de Atributos Relevantes e Não Redundantes usando a Dimensão Fractal do Conjunto de Dados. *Encontro Nacional de Inteligência Artificial* (pp. 22-29). São Leopoldo.
- Lemieux, A. & Parizeau, M. (2003). Flexible multi-classifier architecture for face recognition Systems. *International Conference on Vision Interface*.
- Modi, P. & Tae Kim, W. (2005). Classification of Examples by multiple Agents with Private Features. *International Conference on Intelligent Agent Technology* (pp. 223-229). IEEE Computer Society Press.

- Monard, M. & Baranauskas, J. (2000). Conceitos sobre aprendizado de máquina. In: S. Rezende, *Sistemas inteligentes: Fundamentos e aplicações* (pp. 89-114). São Paulo: Manole.
- Opitz, D. (1999). Feature selection for ensembles. *National Conference on Artificial Intelligence* (pp. 379–384). Menlo Park: AAAI Press.
- Oza, N. & Turner, K. (2001). Input decimation ensembles: Decorrelation through dimensionality reduction. *International Workshop on Multiple Classifier Systems* (pp. 238-247). Springer Verlag.
- Provost, F. & Hennessy, D. (1996). Scaling up: Distributed machine learning with cooperation. *National Conference on Artificial Intelligence* (pp. 74-79). Menlo Park: AAAI Press.
- Puuronen, S. Tsymbal, A., & Skrypnik, I. (2001). Correlation-based and contextual merit-based ensemble feature selection. *International Conference on Advances in Intelligent Data Analysis* (pp. 135-144). Springer Verlag.
- Santana, L. (2005). *Implementação de um sistema multiagente para classificação de padrões utilizando a plataforma JADE*. Monografia, UFRN, DIMAp, Natal.
- Santana, L., Oliveira, D., Canuto, A. & Souto, M. (2007). A Comparative Analysis of Feature Selection Methods for Ensembles with Different Combination Methods. *International Joint Conference on Neural Networks* (pp. 643-648). Orlando.
- Tsymbal, A., Pechenizkiy, M. & Cunningham, P. (2005). Diversity in search strategies for ensemble feature selection. *Information Fusion* , 6 (1), pp. 83-98.
- Tsymbal, A., Puuronen, S. & Patterson, D. (2003). Ensemble feature selection with Bayesian classification. *Information Fusion* , 4 (2), pp. 87-100.
- Vaidya, J. (2004). *Privacy Preserving Data Mining over Vertically Partitioned Data*. PhD thesis, Purdue University, West Lafayette.
- Vaidya, J. & Clifton, C. (2002). Privacy preserving association rule mining in vertically partitioned data. *Conference on Knowledge Discovery and Data Mining* (pp. 639-644). ACM Press.

- Van Kreukelen, M. (1997). Combining classifiers for the recognition of handwritten digits. *Workshop on Statistical Techniques in Pattern Recognition* (pp. 13-18).
- Villar, S. (2005). *Ensemble Case Based Learning for Multiagent Systems*. Tese de Doutorado, Universitat Autònoma de Barcelona, Institut d'Investigació en Intel·ligència Artificial, Bellaterra.
- Vuurpijl, L. & Schomaker, L. (1998). A framework for using multiple classifiers in a multiple-agent architecture. *European Workshop on Handwriting Analysis and Recognition*, (pp. 81-86).
- Wooldridge, M. (2002). *An Introduction to Multiagent Systems*. JOHN WILEY and SONS.
- Zhou, J. (2000). *Face recognition by combining several algorithms*. Dissertação de Mestrado, University of Cambridge, Department of Science Computing, Cambridge.

Apêndice A

Nas Tabelas 2, 3, 4, 5, 6 e 7 pode ser observada a acurácia e o desvio padrão de todos os sistemas (individuais; baseados em fusão; baseados em seleção e multiagente) sem distribuição e com distribuição parcial e total nas duas bases experimentadas (imagem e proteínas). Os gráficos anteriormente apresentados foram construídos com base nesses resultados.

A Tabela 2 mostra os resultados sem distribuição; a Tabela 3 apresenta os resultados das distribuições baseadas em Correlação de Pearson; a Tabela 4 mostra os resultados das distribuições baseadas em Entropia; a Tabela 5 apresenta os resultados das distribuições baseadas no Erro; a Tabela 6 mostra os resultados das distribuições aleatórias e a Tabela 7 apresenta os resultados das distribuições baseadas na Variância.

Sistemas	Base de Imagem				Base de Proteínas			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	90,78(4,21)	91,94(2,83)	93,33(3,87)	95,15(2,28)	59,62(5,40)	66,00(5,22)	67,43(4,29)	69,60(5,66)
DCS_DT	91,07(4,54)	93,68(3,48)	94,35(2,95)	98,90(0,33)	67,55(5,79)	69,57(6,33)	70,05(4,37)	71,42(5,46)
DCS_MCB	95,66(2,90)	97,86(1,59)	98,46(1,27)	98,50(1,46)	69,20(6,50)	74,49(3,16)	74,91(3,54)	74,98(2,97)
Naive	95,37(3,59)	96,43(2,95)	96,69(2,22)	97,05(1,91)	71,66(4,07)	72,68(4,69)	72,70(3,92)	75,77(3,52)
MLP	96,61(2,66)	97,73(1,86)	98,92(0,96)	98,95(0,77)	71,47(6,06)	72,42(4,94)	73,36(5,26)	73,55(5,01)
Soma	96,55(2,38)	97,90(1,84)	98,81(1,12)	98,87(1,20)	72,15(4,61)	74,87(3,58)	79,02(3,67)	79,70(3,83)
Voto	95,44(2,85)	97,70(1,63)	98,48(1,20)	98,51(1,27)	70,15(4,85)	74,08(3,02)	78,57(3,24)	79,21(3,74)
<i>ClassAge</i>	97,56(2,96)	98,61(2,10)	98,98(1,04)	99,03(2,03)	79,04(3,10)	80,37(2,56)	86,74(3,27)	87,07(2,52)

Tabela 2: Acurácia e desvio padrão dos sistemas sem distribuição dos atributos

Base de Imagem

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	87,12(6,51)	87,68(8,02)	87,84(7,73)	87,87(6,5)	79,14(16,7)	69,7(15,01)	66,13(14,8)	58,89(11,4)
DCS_DT	83,10(5,15)	84,75(5,42)	88,76(3,24)	87,4(4,37)	73,56(7,28)	73,51(7,22)	72,06(5,17)	70,61(8,56)
DCS_MCB	93,27(3,54)	94,34(3,67)	95,61(2,99)	93,52(3,39)	90,55(4,6)	89,42(3,61)	88,07(4,29)	86,05(3,86)
Naive	95,08(3,23)	96,11(2,56)	96,53(3,37)	95,00(3,60)	90,90(7,92)	92,29(6,64)	92,42(11,8)	92,82(3,61)
MLP	95,04(3,52)	97,57(2,55)	97,93(2,14)	95,09(3,32)	93,90(3,20)	93,50(2,31)	93,40(1,39)	91,94(3,24)
Soma	94,70(3,51)	96,28(3,26)	96,9(3,27)	95,18(3,47)	93,24(3,78)	93,15(2,77)	93,09(3,47)	89,84(3,92)
Voto	93,41(3,55)	94,33(3,77)	95,74(2,96)	95,37(3,14)	90,70(4,51)	90,24(3,98)	90,17(3,66)	87,97(3,34)
<i>ClassAge</i>	96,9(3,17)	97,64(2,31)	98,72(2,14)	97,88(2,95)	95,97(3,64)	95,81(2,12)	95,62(2,01)	95,54(3,09)

Base de Proteínas

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	70,21(20,3)	68,97(19,7)	80,38(9,07)	79,97(8,69)	68,77(13,7)	62,2(11,15)	59,93(12,0)	55,13(9,78)
DCS_DT	59,81(4,68)	61,36(6,01)	63,96(4,55)	65,32(7,50)	55,55(5,79)	58,42(5,82)	56,60(4,00)	60,63(5,53)
DCS_MCB	73,92(6,07)	75,83(3,86)	76,63(3,14)	76,75(3,47)	75,51(4,72)	73,96(6,83)	71,40(6,06)	70,68(6,30)
Naive	75,25(8,61)	77,08(4,21)	78,18(3,06)	78,23(8,15)	77,17(5,11)	77,03(4,92)	76,25(6,20)	76,15(5,65)
MLP	72,19(5,09)	72,68(4,77)	72,98(4,01)	74,00(4,79)	74,68(6,19)	75,96(4,63)	77,09(4,52)	78,23(5,73)
Soma	71,47(4,94)	75,57(4,06)	75,91(3,16)	76,38(4,16)	75,77(5,42)	75,65(6,45)	75,58(5,05)	73,36(6,55)
Voto	70,04(6,40)	75,4(3,75)	75,47(2,61)	75,51(3,41)	73,55(5,83)	75,02(5,62)	72,19(6,08)	71,89(6,78)
<i>ClassAge</i>	80,95(4,56)	81,97(2,30)	84,26(2,47)	84,81(3,52)	83,2(4,53)	84,04(3,75)	86,10(3,16)	86,90(5,15)

Tabela 3: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Correlação de Pearson

Base de Imagem

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	82,00(10,03)	84,13(12,7)	85,05(9,72)	82,65(10,6)	79,86(13,6)	64,05(21,2)	63,86(14,2)	55,39(17,2)
DCS_DT	84,95(4,67)	96,57(7,27)	98,16(3,22)	82,10(7,67)	91,95(4,38)	91,75(3,75)	88,78(7,47)	50,32(6,75)
DCS_MCB	93,13(3,39)	93,64(2,82)	94,89(1,60)	94,19(2,48)	92,70(2,33)	88,65(4,23)	86,10(3,77)	81,84(4,43)
Naive	93,96(4,11)	94,44(8,77)	95,20(2,68)	86,23(9,83)	94,88(3,01)	92,47(3,45)	88,89(2,64)	86,24(9,45)
MLP	95,97(2,79)	96,26(2,04)	97,76(1,96)	97,89(2,52)	96,25(2,31)	95,51(2,39)	95,05(1,99)	94,48(2,71)
Soma	93,60(3,07)	93,80(2,49)	96,34(1,87)	95,72(2,19)	94,12(2,40)	91,56(3,28)	90,07(3,19)	88,45(4,32)
Voto	90,10(3,92)	91,51(2,69)	95,45(1,87)	94,90(2,39)	91,81(2,61)	89,45(3,71)	87,10(3,50)	83,70(4,05)
<i>ClassAge</i>	96,55(3,35)	97,49(2,48)	98,95(3,64)	98,51(4,5)	95,10(2,94)	95,73(2,76)	96,46(3,26)	94,85(2,80)

Base de Proteínas

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	70,20(8,11)	50,21(28,3)	57,09(17,6)	54,51(18,9)	70,39(6,08)	63,89(7,1)	60,63(10,7)	58,46(10,7)
DCS_DT	70,08(6,56)	64,98(7,44)	62,19(6,52)	60,06(5,84)	75,51(5,07)	75,21(6,3)	73,08(6,17)	72,30(3,34)
DCS_MCB	75,13(6,09)	71,96(6,48)	71,92(4,37)	70,79(6,61)	75,92(5,37)	75,88(4,57)	75,79(4,99)	75,77(6,66)
Naive	76,30(4,80)	72,34(5,54)	70,72(6,83)	70,11(8,42)	76,98(4,78)	75,50(5,52)	75,47(5,98)	75,32(6,27)
MLP	73,89(4,55)	73,32(5,15)	72,49(3,52)	72,30(5,12)	73,21(4,60)	75,06(5,49)	75,74(4,03)	76,19(5,68)
Soma	76,11(5,21)	76,07(5,84)	76,04(4,59)	75,17(6,55)	76,30(5,18)	76,57(5,98)	77,02(4,13)	77,92(5,53)
Voto	75,77(5,20)	71,70(8,04)	70,72(4,06)	68,91(7,33)	75,92(5,22)	76,01(5,20)	76,33(4,19)	76,57(6,30)
<i>ClassAge</i>	76,87(5,29)	76,13(5,44)	76,11(5,17)	75,86(4,89)	77,04(5,35)	77,05(4,66)	78,01(4,87)	78,22(5,21)

Tabela 4: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Entropia

Base de Imagem

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	84,46(11,8)	84,66(10,2)	85,61(10,7)	85,04(10,4)	77,53(16,1)	66,96(18,7)	59,25(19,2)	55,28(18,7)
DCS_DT	92,85(4,08)	94,63(7,62)	97,73(0,75)	92,03(4,59)	89,90(3,87)	87,77(6,02)	87,38(3,33)	84,86(7,89)
DCS_MCB	94,79(3,02)	95,64(1,60)	96,31(2,29)	93,99(1,91)	91,08(3,33)	87,75(4,63)	86,69(2,91)	84,04(4,48)
Naive	94,09(4,05)	94,38(6,51)	94,84(7,08)	93,24(6,56)	93,09(3,67)	87,36(6,12)	88,22(6,58)	93,00(6,92)
MLP	96,33(2,75)	96,88(1,79)	98,55(1,43)	97,41(1,98)	95,06(2,51)	94,76(2,13)	94,50(2,02)	95,93(1,91)
Soma	94,28(3,05)	95,12(1,40)	97,65(1,33)	95,50(1,74)	92,38(2,95)	90,60(3,92)	89,45(3,81)	89,22(3,10)
Voto	92,94(3,52)	93,84(2,05)	96,48(2,30)	94,28(1,88)	89,63(3,31)	87,92(4,11)	86,94(2,96)	84,93(3,79)
<i>ClassAge</i>	97,09(3,01)	97,56(3,49)	98,94(2,99)	97,72(2,84)	95,27(2,76)	95,35(2,51)	95,67(2,92)	96,63(2,77)

Base de Proteínas

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	72,45(8,31)	72,32(7,34)	72,59(7,9)	70,22(9,74)	69,64(7,53)	65,94(6,33)	64,81(8,13)	60,41(9,04)
DCS_DT	78,43(5,83)	78,50(6,88)	78,67(7,24)	79,23(5,96)	72,49(5,91)	72,43(5,23)	72,28(7,05)	72,19(3,61)
DCS_MCB	78,68(4,32)	80,53(4,50)	80,79(4,28)	80,60(4,75)	76,91(5,23)	76,49(4,43)	76,15(3,74)	75,81(4,78)
Naive	79,28(3,78)	79,43(5,08)	80,45(4,10)	81,13(6,95)	78,57(3,39)	78,70(5,50)	79,12(4,24)	79,36(4,09)
MLP	74,08(4,92)	74,75(5,48)	75,51(5,67)	76,30(5,87)	73,02(3,42)	74,08(4,09)	74,53(4,30)	75,47(4,60)
Soma	78,53(5,05)	80,19(5,12)	80,79(5,33)	80,85(5,42)	76,19(4,68)	77,06(4,23)	78,62(4,28)	78,87(3,96)
Voto	78,53(4,32)	80,57(4,25)	80,75(4,09)	80,75(4,25)	75,13(4,67)	75,87(4,53)	76,22(4,56)	76,34(4,63)
<i>ClassAge</i>	80,22(5,92)	80,97(5,83)	81,35(6,28)	81,46(5,79)	79,06(5,56)	79,12(5,62)	79,52(6,56)	79,66(6,36)

Tabela 5: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica de Erro

Base de Imagem

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	85,5(10,97)	85,17(10,8)	84,21(10,1)	84,23(10,0)	72,44(16,8)	63,41(19,0)	61,96(18,1)	53,69(20,2)
DCS_DT	92,62(3,87)	95,65(5,27)	98,48(4,05)	88,64(5,39)	86,50(7,24)	85,49(6,02)	83,14(7,57)	82,38(8,63)
DCS_MCB	95,13(2,45)	95,22(2,15)	96,02(2,30)	91,58(2,14)	87,51(4,53)	86,11(3,33)	84,92(3,57)	82,20(4,76)
Naive	95,00(3,81)	95,12(8,89)	96,54(3,85)	94,31(4,52)	93,15(3,58)	93,03(4,34)	92,50(4,3)	87,97(8,84)
MLP	96,88(2,09)	96,97(2,35)	97,93(2,17)	98,27(2,15)	93,45(3,44)	93,42(2,37)	92,30(2,39)	94,15(2,19)
Soma	95,21(2,38)	95,41(2,10)	96,41(1,91)	94,98(1,83)	89,90(4,01)	89,32(3,14)	88,78(3,81)	88,71(3,57)
Voto	93,88(3,80)	94,66(2,43)	94,90(2,24)	92,33(2,57)	88,14(4,64)	87,50(3,66)	85,43(3,17)	83,20(4,87)
<i>ClassAge</i>	97,45(2,98)	97,50(2,68)	98,61(2,74)	98,62(3,00)	93,28(3,39)	94,08(2,88)	94,24(2,87)	95,05(2,88)

Base de Proteínas

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	65,45(14,5)	70,1(13,3)	67,22(14,9)	64,82(12,6)	63,84(12,9)	61,92(11,1)	56,89(11,6)	53,94(10,8)
DCS_DT	67,47(6,50)	69,13(7,06)	71,06(9,21)	72,57(7,35)	61,92(8,74)	60,42(7,12)	58,32(7,77)	53,55(7,32)
DCS_MCB	75,21(4,81)	75,51(4,58)	75,89(4,64)	76,38(4,45)	74,83(5,05)	74,47(5,62)	73,55(5,91)	71,02(5,47)
Naive	72,98(5,29)	73,06(7,94)	74,83(6,47)	75,89(4,13)	76,11(6,16)	75,68(5,89)	74,75(4,41)	72,45(5,20)
MLP	73,13(4,48)	73,28(5,23)	73,32(6,17)	73,77(4,73)	74,60(4,81)	74,75(4,87)	74,89(5,31)	74,98(5,06)
Soma	76,53(4,54)	76,64(3,7)	77,49(5,01)	77,62(4,52)	76,00(4,68)	75,32(4,92)	73,02(5,28)	72,72(5,85)
Voto	75,13(4,73)	75,62(4,34)	76,02(4,90)	76,34(3,77)	74,42(5,45)	74,38(5,53)	72,79(5,57)	69,66(5,78)
<i>ClassAge</i>	76,60(7,47)	76,89(7,15)	77,77(6,61)	78,62(6,94)	76,61(7,06)	76,90(6,76)	77,10(6,95)	77,21(7,20)

Tabela 6: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica Random

Base de Imagem

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	88,42(5,49)	91,19(4,92)	86,48(12,8)	86,84(11,8)	79,24(15,6)	67,86(19,4)	64,37(14,6)	55,75(16,1)
DCS_DT	94,77(4,88)	96,59(4,14)	97,90(2,03)	91,64(4,20)	92,29(5,30)	91,89(6,21)	90,90(4,75)	81,90(8,43)
DCS_MCB	95,16(2,54)	96,25(2,29)	97,70(1,66)	97,05(2,29)	92,08(4,08)	91,91(3,8)	87,29(3,59)	81,65(5,24)
Naive	94,73(2,96)	96,40(3,18)	97,09(7,17)	86,57(9,22)	93,12(4,22)	93,07(4,71)	92,41(3,17)	90,64(9,52)
MLP	95,82(3,19)	98,56(1,19)	98,60(1,30)	98,90(1,18)	95,05(3,54)	94,29(2,54)	94,16(1,11)	93,25(2,66)
Soma	94,99(3,16)	97,69(1,57)	98,59(1,22)	98,66(1,25)	91,43(4,51)	91,14(4,25)	90,11(2,62)	87,49(4,12)
Voto	93,63(3,29)	96,10(2,68)	97,84(1,91)	97,19(2,22)	87,00(4,61)	86,57(4,26)	86,55(3,18)	82,94(4,56)
<i>ClassAge</i>	96,90(3,20)	98,69(3,41)	99,04(3,03)	99,44(2,71)	94,46(2,54)	95,06(3,33)	95,75(2,78)	96,07(3,20)

Base de Proteínas

Sistemas	Distribuição Parcial				Distribuição Total			
	Três	Cinco	Sete	Nove	Três	Cinco	Sete	Nove
Individuais	68,57(7,23)	68,87(7,31)	70,06(7,27)	69,22(7,45)	64,64(9,35)	57,99(9,97)	59,31(10,3)	54,61(10,0)
DCS_DT	73,92(4,84)	77,15(5,89)	77,26(7,17)	78,17(7,10)	74,34(7,38)	74,70(6,2)	75,13(9,66)	75,36(2,57)
DCS_MCB	75,47(5,89)	75,87(4,21)	76,19(3,28)	77,81(2,83)	72,94(7,04)	72,85(5,94)	70,58(4,74)	69,36(5,62)
Naive	77,43(4,81)	77,58(3,90)	78,09(3,86)	78,21(6,32)	73,43(5,17)	73,25(4,47)	73,18(4,71)	73,13(4,16)
MLP	73,51(5,44)	75,13(4,16)	76,11(4,94)	77,06(3,13)	75,32(5,88)	77,05(5,01)	77,12(4,85)	77,28(4,23)
Soma	76,15(4,73)	77,38(4,43)	78,13(3,04)	79,09(3,09)	72,94(5,82)	73,13(5,5)	73,28(5,76)	73,32(3,84)
Voto	73,32(5,14)	75,58(4,17)	77,26(3,56)	77,58(2,83)	71,32(6,08)	70,45(5,97)	70,37(4,91)	70,30(4,99)
<i>ClassAge</i>	78,20(6,22)	78,42(5,78)	79,32(5,85)	79,81(5,77)	75,40(6,35)	77,42(6,02)	77,81(5,88)	80,91(6,37)

Tabela 7: Acurácia e desvio padrão dos sistemas com distribuição dos atributos usando a técnica da Variância

Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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