

# Seleção Estatística de Dados para Classificadores Neurais de Sinais de Sonar Passivo

João Baptista de Oliveira e Souza Filho,  
José Manoel de Seixas

Laboratório de Processamento de Sinais - LPS - COPPE/EP  
Universidade Federal do Rio de Janeiro, Caixa Postal 68509, Rio de Janeiro, Brasil.  
Tel. (fax): 2562-8205  
Emails: {nash,seixas}@lps.ufrj.br

**Abstract**—Neste trabalho, discute-se a seleção de dados para o projeto de classificadores em problemas sujeitos a restrições estatísticas do banco de dados. Considerando a classificação de espectros de sonar passivo, provenientes de 25 navios pertencentes a 8 classes distintas, é mostrado que a seleção dos conjuntos de projeto e teste, por meio de técnicas de análise de agrupamentos, permite, com um menor custo computacional, obter classificadores de maior eficiência média e melhor desempenho para as classes mais críticas. Utilizando esta técnica, um classificador com uma eficiência média de 94,6% foi obtido, o que é significativo para o problema em questão.

## I. INTRODUÇÃO

A vigilância realizada pelas forças armadas utiliza equipamentos variados, com o objetivo de buscar alvos que podem ser ameaças ou objetos de interesse. Neste contexto, os submarinos de guerra realizam, comumente, operações de patrulha e ataque. Desta maneira, a identificação, em tempo hábil, de embarcações eventualmente presentes em sua vizinhança, referidas como contatos, é fundamental.

O principal sistema utilizado nos submarinos para a detecção e identificação de contatos é o sistema de sonar passivo. Operadores especialmente treinados, analisando a sonoridade e o conteúdo de frequência do sinal de sonar, definem o contato mais provável, já que cada embarcação produz um ruído característico, conhecido como assinatura, que está relacionado à sua cavitação, ao ruído do hélice e das máquinas em operação no seu interior [1]. A identificação de contatos não é uma tarefa trivial, dada a complexidade do ruído captado, que reúne sinais provenientes de múltiplas fontes, entre elas: o contato, os ruídos biológicos, assim como as múltiplas reflexões destes ruídos no assoalho submarino.

Desta forma, sistemas automáticos de classificação constituem uma ferramenta bastante útil ao operador, podendo ser utilizados como instrumento de apoio à decisão, o que reduz o volume de trabalho e aumenta o nível de confiabilidade e a rapidez do processo de decisão pelo operador. Para implementação de um tal sistema, a técnica de redes neurais artificiais [2] é especialmente adequada, dada sua habilidade em extrair informação relevante, mesmo em ambientes de alta-dimensionalidade, produzindo classificadores compactos e de elevada eficiência.

Para o desenvolvimento de sistemas de classificação de contatos, a existência de uma base de dados que reflita as

condições reais de operação das diferentes classes é fundamental. Assim, o procedimento comumente adotado no ambiente de sonar consiste em realizar corridas de navios pertencentes a variadas classes, segundo diferentes condições operativas, numa raia acústica, promovendo a aquisição dos sinais produzidos no interior da raia, de forma a simular uma situação real. Este procedimento é um processo complexo e caro, sujeito a variadas restrições, inclusive em relação ao número de navios disponíveis à caracterização de cada classe, o que, normalmente, restringe a estatística dos dados disponíveis.

Neste trabalho discutimos a seleção de dados para o projeto de um sistema automático de classificação de contatos baseado em redes neurais artificiais, visando reconhecer 8 classes de navios distintas. Foi verificado, através de alguns ensaios utilizando a técnica de validação cruzada [3], uma flutuação significativa das eficiências de generalização, o que motivou a aplicação de técnicas de agrupamento para a seleção dos conjuntos de desenvolvimento e teste do classificador.

Para a produção dos agrupamentos, realizou-se a compactação dos sinais, uma vez que as técnicas de agrupamento costumam apresentar melhor desempenho em espaços de dimensão reduzida. Para esta compactação, duas técnicas foram avaliadas: a análise de componentes principais (PCA) [4] e a análise de componentes principais para discriminação (PCD) [5].

Utilizando os agrupamentos produzidos para cada modalidade de compactação, foram definidos vários conjuntos de projeto e teste, cada um baseado numa granularidade e número de componentes particular, visando analisar a dependência da técnica de seleção proposta com respeito, também, a estes parâmetros. Resultados significativos foram obtidos, mostrando que a técnica proposta permitiu, a um baixo custo computacional, uma seleção de dados que resultou na produção de um classificador de maior eficiência média e com melhor desempenho para as classes de detecção mais crítica.

A estrutura do trabalho é a seguinte: inicialmente, descreveremos as técnicas utilizadas para a compactação dos dados, visando produzir agrupamentos mais eficientes em dimensão reduzida. Em seguida, a metodologia para a produção dos agrupamentos e a base de dados utilizada são apresentados. Finalmente, os resultados são apresentados e são retiradas as

conclusões do trabalho.

## II. COMPACTAÇÃO DE SINAIS

Para um grande número de aplicações, a seleção de variáveis relevantes é vital, sem a qual a solução de um problema particular pode ser prejudicada, seja pelo volume de informações irrelevantes, ou mesmo, inviabilizada, dado o volume de processamento exigido. Estas restrições são bastante comuns em sistemas que visam operação em tempo real, tais como o sonar passivo.

Verifica-se ainda que as técnicas de agrupamento apresentam melhores resultados se a dimensionalidade dos dados é reduzida. Deste modo, a aplicação destas técnicas à ambientes de alta-dimensionalidade, tais como o de sonar passivo, exigem a utilização de técnicas de extração de características visando promover a compactação ou redução da dimensão dos dados envolvidos.

A análise de componentes principais (PCA - *Principal Component Analysis*) [4] é baseada na série de Karhunen-Löve, sendo um método de extração de características linear, consagrado, aplicado em variadas áreas, entre elas: detecção, estimação e no reconhecimento de padrões. Para dados com dimensão  $N$ , esta análise obtém um conjunto de  $M$  direções ( $M \ll N$ ) que, quando utilizadas para a compactação dos dados, reduzem sua dimensão para  $M$ , produzindo uma perda mínima no sentido do erro médio quadrático (energia). Trata-se, portanto, de uma metodologia que produz direções privilegiadas no sentido da representação dos dados. Variados métodos podem ser utilizados para a obtenção destas componentes, entre eles: métodos algébricos [6] e baseados em redes neurais [7].

Outra análise, cuja aplicação à compactação dos dados é proposta neste trabalho, é a análise de componentes principais de discriminação (PCD - *Principal Components for Discrimination*). O objetivo da análise PCD é extrair direções de projeção dos dados tal que o máximo de informação relativa à sua discriminação em diferentes classes seja obtida, em contraposição à análise PCA, a qual busca uma melhor "representação conjunta" das classes.

Uma forma possível de extrair componentes PCD é utilizar um classificador neural do tipo *perceptron* multi-camadas [2] de duas camadas, formado, inicialmente, por  $N$  nós de entrada, 1 neurônio na camada escondida e  $K$  neurônios na camada de saída, onde  $K$  corresponde ao número de classes <sup>1</sup>. Atingida convergência, o vetor de pesos que conecta o único neurônio da camada intermediária aos nós de entrada é considerado como a primeira componente de discriminação, por fornecer a direção que melhor promove a separação dos dados nas diferentes classes, através da atuação conjunta com a camada de saída. Este vetor de pesos será mantido congelado durante o restante do treinamento, o qual é reiniciado, sendo inserido mais um neurônio à camada intermediária da rede, o qual fornecerá a segunda componente de discriminação ao fim desta

<sup>1</sup>A codificação das classes nos vetores-alvo e na saída da rede é realizada de forma maximamente esparsa [2].

etapa de treinamento. Este processo é repetido por um total de  $M$  vezes, produzindo  $M$  neurônios na camada escondida da rede, sendo  $M$  o número de componentes de discriminação a serem extraídas.

Dada a inserção gradativa de neurônios na estrutura da rede, esta metodologia é também conhecida como extração PCD construtiva. Por outro lado, como esta técnica obtém direções privilegiadas quanto à discriminação, é esperado que a técnica retenha, em geral, num menor número de componentes, um maior volume de informação útil para fins de classificação quando comparada à análise PCA. Assim, é esperado que classificadores baseados em dados projetados em direções fornecidas pela análise PCD possuam maior eficiência que os obtidos através de procedimento análogo utilizando a análise PCA, para o mesmo número reduzido de componentes.

## III. AGRUPAMENTO DE DADOS

Para a produção dos agrupamentos, foi utilizado um algoritmo inspirado no modelo neural auto-organizável ART-2 [8], aqui referido como ART-M, o qual permite o reconhecimento estável, em tempo real, dos padrões apresentados sequencialmente a sua entrada. Segundo a técnica ART-M, uma nuvem de dados produz um conjunto de agrupamentos, cada um formado por uma hipersfera caracterizada por um neurônio que define o seu centro e raio de vigilância [9].

O algoritmo da técnica ART-M é o seguinte: os eventos são apresentados em ordem aleatória e a cada evento ( $\mathbf{x}$ ) é produzida uma medida de similaridade com respeito aos neurônios existentes na rede. Esta medida é determinada pela relação  $s_i = r_i^2 - \|\mathbf{x} - \mathbf{c}_i\|^2$ , ou seja, o raio de vigilância ( $r_i$ ) e a distância do evento ao centro ( $\mathbf{c}_i$ ) de cada neurônio são considerados. Posteriormente, o neurônio que apresentar maior similaridade é definido como vencedor, sendo determinada a distância (euclidianda) deste padrão ao seu centro. Caso esta distância seja inferior ao raio de vigilância a ele associado, as coordenadas do seu centro são atualizadas pela fórmula  $\mathbf{c}_i = \mathbf{c}_i + \eta \cdot \|\mathbf{x} - \mathbf{c}_i\|$ , onde  $0 < \eta < 1$ . A constante  $\eta$  é reduzida, gradativamente, ao longo do treinamento. Em caso contrário, ou seja, para eventos no exterior da hipersfera definida pelo neurônio vencedor, é criado um novo neurônio, cujo centro corresponde ao próprio evento. Repete-se este processo até que nenhum neurônio seja inserido ou que os centros apresentem deslocamentos inferiores a um dado limiar.

Nesta técnica, o raio de vigilância determina a granularidade do agrupamento, tanto maior quanto menor for o valor do raio. Raios menores tendem a produzir maior número de neurônios. Neste trabalho, todos os neurônios utilizados possuem o mesmo raio de vigilância, escolhido com base na distância (euclidianda) mais freqüente (moda) dos eventos.

## IV. BASE DE DADOS

A base de dados utilizada é formada pelo ruído irradiado por 25 navios pertencentes a 8 classes, sendo que cada classe possui, pelo menos, 2 navios distintos. Para obtenção deste ruído, cada navio realizou uma corrida, deslocando-se, segundo condições de maquinário e operação fixas, pela raia

acústica da Marinha do Brasil, situada em Arraial do Cabo. Esta raia possui uma profundidade de 45 m (aproximadamente), havendo um hidrofone ao fundo. Um total de 263 corridas foram digitalizadas segundo uma frequência de amostragem de 22,05 kHz, utilizando 16 bits de resolução para as amplitudes.

Os sinais digitalizados foram submetidos a uma cadeia de pré-processamento, com o objetivo de extrair o conteúdo espectral do sinal na faixa de frequência de 0 a 3kHz, a qual contém informação discriminante, pois retém informação relativa à cavitação, ao hélice e às máquinas em operação no interior dos navios [1]. É também papel desta cadeia extrair o ruído de fundo do ambiente de medição, o qual não contribui para a classificação. Na Figura 1, a cadeia utilizada é apresentada.

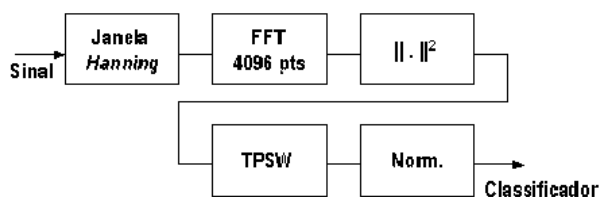


Fig. 1. Cadeia de pré-processamento

Pelo pré-processamento, o sinal é dividido, sem superposição, em janelas de 4096 amostras, o que totaliza trechos de, aproximadamente, 186 ms de duração. É suposto que neste intervalo, o processo aleatório associado ao ruído irradiado pelos navios possa ser considerado como estacionário no sentido amplo [10]. Cada janela é multiplicada por uma janela de *Hanning*, sendo seu espectro calculado pela transformada rápida de *Fourier* (FFT) [11]. Das 2048 amostras distintas produzidas, apenas 557 são consideradas, as quais cobrem a faixa de interesse (0 a  $\approx$  3kHz). Procede-se o cálculo do módulo e estima-se o ruído de fundo do ambiente de medição segundo o algoritmo *Two-Pass Split Window* [12], que é utilizado para normalizar o sinal. Por fim, realiza-se uma segunda normalização, na qual o sinal passa a possuir energia unitária, a fim de que o classificador seja invariante à distância do contato ao hidrofone.

## V. RESULTADOS

Os classificadores neurais utilizados foram do tipo MLP, com a topologia de 557 nós de entrada, um número variável de neurônios na camada escondida e 8 neurônios na camada de saída. A função de ativação utilizada para todos os neurônios da rede foi a tangente hiperbólica. A codificação das classes utilizou o critério de máxima esparsidade, isto é, a cada neurônio é atribuída uma classe, apresentando o valor alvo de +1,0 para os eventos à ela pertencentes; e de -1,0, em caso contrário. Para definição da classe vencedora, utilizou-se o critério de máxima probabilidade [2], ou seja, a classe vencedora é definida pelo neurônio de maior valor de saída. Para o treinamento, os dados foram divididos em conjuntos de treino

e teste, e o critério de parada antecipada [13] foi adotado. Este critério foi baseado na eficiência média do classificador, estimada pela média aritmética ponderada das eficiências de cada classe, assumindo, como fatores de ponderação, o número de eventos disponíveis por classe. O treinamento foi realizado em batelada, ou seja, só há atualização dos parâmetros após todo conjunto de treinamento ter sido utilizado para o cálculo do gradiente da superfície de erro. Por fim, a inicialização dos parâmetros (pesos e limiares) foi realizada segundo uma distribuição uniforme, compreendendo valores na faixa de -0,5 a +0,5.

Dada a complexidade do classificador, determinada pelo número de classes, pela elevada dimensão e número de eventos disponíveis, foi necessário investigar uma técnica para o ajuste de parâmetros que fosse, computacionalmente, mais adequada ao problema. Realizaram-se alguns ensaios envolvendo as técnicas de *backpropagation* [14], gradiente conjugado escalado (SCG) [15] e *resilient backpropagation* (RPROP) [16]. Foi verificado que o algoritmo RPROP, utilizando os parâmetros sugeridos no seu artigo descritivo, apresentou um desempenho significativamente melhor que os demais, tanto em termos das eficiências de classificação e generalização, quanto em relação ao custo computacional, sendo definido como o algoritmo padrão de treinamento aplicado à solução do problema.

O treinamento dos classificadores, o qual considerou os dados em sua dimensão original, em contraposição à produção dos agrupamentos, foi realizado de forma construtiva e cooperativa, seguindo procedimento análogo à extração de componentes PCD. Este processo costuma produzir classificadores com desempenho similar aos treinados pela forma usual, porém a um custo computacional inferior, dado o menor número de parâmetros a serem ajustados, em especial, nos casos de múltiplos ensaios onde o número de neurônios na camada escondida é crescente [17]. Visando ainda minimizar a susceptibilidade do treinamento a mínimos locais, para a inserção de um novo neurônio, foi realizado um total de 5 tentativas, considerando-se, para cada uma, diferentes valores iniciais para os parâmetros. A tentativa que apresentou a maior eficiência de generalização definiu os pesos associados ao novo neurônio.

### A. Validação cruzada

Visando avaliar o impacto das restrições estatísticas do conjunto de dados no desempenho do classificador, foi utilizada uma variante da técnica de validação cruzada, conhecida como técnica de *holdout* [18]. Foram considerados 8 ensaios com eventos selecionados por amostragem uniforme, com reposição, para cada classe. Dada uma seleção, os eventos foram divididos, igualmente, na formação dos conjuntos de treino e teste de cada ensaio. Para cada ensaio, uma rede foi treinada segundo a metodologia anteriormente descrita, compreendendo de 1 até 35 neurônios na camada intermediária.

No treinamento de classificadores, em geral, procura-se maximizar a detecção média. Considerando a aplicação de sonar, também é relevante a detecção de cada classe individual, a qual determina a probabilidade de perda, ou seja, de que um

evento de uma dada classe não seja identificado como a ela pertencente. Assim, outro parâmetro relevante na detecção de múltiplas classes é a pior eficiência individual, aqui referida como eficiência da pior classe, a qual estabelece um limite superior para a probabilidade de perda.

Para cada ensaio, foram levantadas a eficiência média e da pior classe, estimadas para cada neurônio inserido na camada intermediária, visando permitir uma análise do impacto da topologia da rede e do efeito da inicialização dos parâmetros em relação à eficiência de generalização do classificador. Os valores de eficiência média e da pior classe foram caracterizados através do valor médio, estimado por média aritmética simples, e pela dispersão média, estimada pelo valor RMS, das cinco tentativas realizadas por neurônio inserido. Vale ainda observar que o cálculo da eficiência média de cada ensaio utilizou o mesmo procedimento que o adotado para o cálculo da eficiência envolvida na parada do treinamento.

Por fim, de posse dos valores médios para cada neurônio e ensaio, determinou-se um valor médio e dispersão associada, considerando todos ensaios realizados, de forma a totalizar as flutuações relacionadas às diferentes seleções possíveis do conjunto de projeto do classificador. Esta análise mostrou que a dispersão da eficiência da pior classe era maior que a apresentada pela eficiência média, situando-se, no máximo, em 2,3% e 0,4%, respectivamente, para os classificadores com um número superior a 15 neurônios na camada intermediária. Foi verificado ainda que os valores de eficiência tendem a crescer com o aumento do número de componentes, tendência que é mais acentuada na faixa de 1 a 10 e se estabiliza no intervalo entre 10 a 20 neurônios, conforme esperado. Deste modo, concluiu-se que um classificador com 25 neurônios na camada intermediária apresenta um bom compromisso entre eficiência e custo computacional. Observa-se, no entanto, uma significativa flutuação entre os valores apresentados pelos diferentes ensaios.

Na Tabela I são apresentados alguns valores de eficiência média e da pior classe, com suas respectivas incertezas, considerando todos ensaios. É destacada ainda a diferença entre a média da eficiência média e a da pior classe em pontos percentuais (p.p). Da análise dos resultados apresentados, verificamos que a diferença de desempenho nos dois casos é significativa, tendendo a se reduzir com o aumento do número de neurônios da camada escondida (NN), conforme esperado. As diferenças são de 6 pontos percentuais para 15 componentes (máximo) e de 4 pontos para 30 componentes (mínimo).

As flutuações observadas nos ensaios de validação cruzada, assim como as diferenças verificadas entre as eficiências média e da pior classe, apontaram a complexidade do problema e a existência de restrições estatísticas no conjunto de dados. Desta maneira, uma análise de agrupamentos foi realizada, com o objetivo de isolar conjuntos de estatística mais significativa, os quais permitirão uma seleção mais adequada dos conjuntos de projeto e teste do classificador.

TABELA I. Eficiências média e para pior classe (%)

Eficiências para múltiplos ensaios			
NN	Pior classe	Média	Diferença (p.p)
15	86,0 ± 2,3	92,0 ± 0,3	6,0
20	88,5 ± 1,2	93,1 ± 0,4	4,6
25	89,5 ± 1,5	93,7 ± 0,4	4,2
30	90,0 ± 1,2	94,0 ± 0,3	4,0

### B. Compactação dos sinais

A compactação dos sinais consistiu na projeção dos eventos nas direções fornecidas pelas análises PCA e PCD. Para a escolha do número de componentes utilizadas nesta compactação, avaliou-se a eficiência de generalização da rede extratora PCD ao final da extração de cada componente. Dada a estrutura e a forma como esta rede é treinada, estas eficiências fornecem uma rápida estimativa da ordem das eficiências de um classificador neural baseado em eventos compactados com base nestas direções. Através desta análise, verificou-se que um valor apropriado para compactação é de 10 componentes, atingindo uma eficiência média em torno de 95%.

Com o objetivo de comparar quanto de informação relevante à classificação foi retido por cada tipo de análise, três ensaios foram realizados, dois deles compreendendo a compactação dos eventos em 10 componentes, fornecidas pelas análises PCA e PCD; e um terceiro, que considerou apenas 5 direções PCD. Para cada ensaio, foi treinada uma rede neural segundo o procedimento construtivo e cooperativo mencionado anteriormente, utilizando, de forma semelhante ao processo de extração, um conjunto de treinamento formado por todos eventos e classes disponíveis.

Na Tabela II resume-se as eficiências médias obtidas em função do número de neurônios da camada intermediária do classificador, juntamente com as máximas dispersões (MD), ou seja, o pior valor RMS observado para cada técnica de compactação. Como era esperado, verifica-se que as componentes PCD retêm um maior volume de informação discriminatória que as componentes PCA, permitindo a obtenção de classificadores com eficiências superiores em torno de 10 pontos percentuais para redes com 25 neurônios na camada intermediária. Quando comparadas as compactações PCD-5C e PCA-10C, ainda que metade do número de componentes seja utilizado, a extração PCD produz classificadores, no mínimo, 2 pontos percentuais mais eficientes que a extração PCA.

TABELA II. Eficiências obtidas para diferentes técnicas de compactação

NN	PCA-10C	PCD-5C	PCD-10C
15	81,7	86,9	94,3
20	83,2	87,0	94,8
25	84,8	87,2	95,2
30	85,8	87,5	95,5
<b>MD</b>	0,2	0,02	0,05

### C. Análise dos agrupamentos

Um agrupamento foi produzido para cada classe, utilizando como raio inicial valores proporcionais à moda da distância dos seus eventos constituintes. Dois valores foram selecionados, um correspondente a uma vez a moda, aqui referido como ensaio m; e outro correspondente a 3 vezes este valor, referido como ensaio M. Esta escolha foi heurística e utilizou valores de raio inicial que visavam produzir agrupamentos com granularidades razoavelmente distintas. Vale observar que uma granularidade menor é obtida para um valor de raio inicial maior em relação a moda da distância dos eventos.

Uma primeira análise dos agrupamentos produzidos pela técnica ART-M consiste em avaliar o número de grupos produzidos em cada ensaio. Suponha que um dado conjunto de dados seja compactado por duas técnicas distintas para um mesmo número de dimensões. Considere ainda que a mesma granularidade seja utilizada em ambos os casos. Espera-se que o agrupamento que apresentar um maior número de grupos reflita uma estrutura de dados mais complexa.

Para esta análise, esboçou-se o número de agrupamentos produzidos para cada ensaio (m,M), considerando as três modalidades de compactação (PCA-10C, PCD-10C e PCD-5C), o que resultou na Figura 2<sup>2</sup>. Neste gráfico, no eixo vertical tem-se o número de grupos por classe do contato (eixo horizontal), para cada técnica de compactação (padrão de hachura das barras). É possível observar que a compactação PCD-5C produziu agrupamentos, para a grande maioria das classes, de complexidade equivalente, senão superior aos produzidos pelas demais técnicas de compactação, exceto para as classes D e G. É esperado portanto que esta compactação tenha retido maior volume de informação dos dados, o que deverá se refletir em maiores eficiências de classificação. Para os ensaios de 10 componentes, verifica-se que a técnica PCD tende a apresentar um maior número de grupos que a técnica PCA para o ensaio m; enquanto, para o ensaio M, reverte-se esta tendência. De forma geral, verifica-se um comportamento similar de ambos ensaios, ocorrendo uma redução em torno de 3 vezes do número de agrupamentos para a maioria das classes, o que é coerente com as resoluções escolhidas.

Para cada ensaio e tipo de compactação, um classificador foi treinado. Os conjuntos de projeto e teste de cada classificador foram formados pela divisão igualitária dos eventos pertencentes a um mesmo agrupamento. De forma análoga aos ensaios de validação cruzada, foi verificado que a dispersão associada à pior classe era superior àquela obtida para a eficiência média, situando-se em, no máximo, 1,0% e 0,1%, respectivamente. Comparando estes valores com os obtidos anteriormente, verificamos que estas flutuações são menores.

Em relação às piores eficiências, verifica-se, para uma mesma metodologia de compactação, que os classificadores obtidos para o Ensaio M (raio correspondente à três vezes a moda) são sempre superiores que aos obtidos para o Ensaio

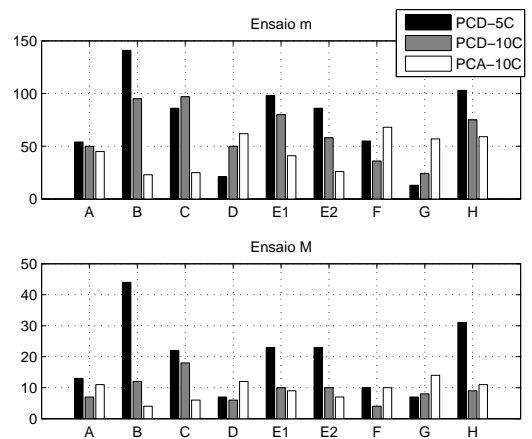


Fig. 2. Número de agrupamentos por técnica de compactação e granularidade, de acordo com cada classe de contato.

m, quando de 15 a 35 neurônios são utilizados na camada intermediária, exceto no ensaio PCD-5C, para valores de 15 e 16 neurônios. Conclui-se que é desejável agrupamentos de menor granularidade. Neste caso, o melhor desempenho é obtido para o ensaio PCD-5C, seguido pelo PCD-10C, o qual apresenta desempenho superior ou equivalente ao ensaio PCA-10C, exceto para as componentes 17 a 20, 29 e 30. Na Tabela III resume-se as eficiências para pior classe e média, para o Ensaio M, quando 15, 20, 25 e 30 neurônios são utilizados na camada intermediária do classificador. Para 20 e 25 componentes, o melhor desempenho é obtido pela técnica PCD-5C (em negrito), enquanto que, no caso de 15, a diferença em relação a técnica de melhor desempenho (PCD-10C) não foi significativa (0,2%). Quando estes resultados são comparados com os obtidos por validação cruzada (Tabela I), verifica-se que as eficiências obtidas são superiores, assim como as diferenças entre a eficiência média e da pior classe são menores.

TABELA III. Eficiências obtidas para as diferentes técnicas de compactação considerando a seleção por agrupamentos - Ensaio M (%)

PCD-5C	Pior Classe	Média	Diferença
15	88,3	92,7	4,4
<b>20</b>	<b>90,7</b>	<b>94,1</b>	<b>3,4</b>
<b>25</b>	<b>91,6</b>	<b>94,6</b>	<b>3,0</b>
30	91,8	94,6	2,8
PCD-10C	Pior Classe	Média	Diferença
<b>15</b>	<b>88,5</b>	<b>92,7</b>	<b>4,2</b>
20	89,9	93,2	3,3
25	91,6	93,9	2,3
30	91,5	94,0	2,5
PCA-10C	Pior Classe	Média	Diferença
15	87,1	92,6	5,5
20	90,8	93,8	3,0
25	91,0	94,4	3,4
<b>30</b>	<b>91,9</b>	<b>94,8</b>	<b>2,9</b>

<sup>2</sup>Dado o elevado número de eventos disponíveis na classe E (em torno de 7000), esta classe foi ainda subdividida em duas, referidas como E1 e E2, contendo, aproximadamente, quantidades iguais de eventos.

Finalmente, as eficiências classe-a-classe para o classificador treinado com o melhor conjunto obtido por validação cruzada (C-V) foram comparadas com as resultantes da seleção PCD-5C, ensaio M, para um classificador de 25 neurônios. Os resultados obtidos são apresentados na Tabela IV. Podemos perceber que o classificador baseado na seleção PCD-5C possui um desempenho superior para a maioria das classes, especialmente em relação à classe H, onde se verifica uma diferença de 2,8% pontos percentuais. Quanto a eficiência média, temos 94,6% para o método PCD-5C e 93,8% para a validação cruzada.

TABELA IV. Comparação classe-a-classe entre as eficiências dos melhores classificadores por agrupamento e validação cruzada (%) para um classificador de 25 PCDs

	A	B	C	D
<b>PCD-5C</b>	92,4	<b>94,5</b>	91,5	<b>94,7</b>
<b>C-V</b>	<b>92,7</b>	92,8	<b>91,8</b>	94,0
	E	F	G	H
<b>PCD-5C</b>	<b>98,0</b>	<b>96,0</b>	<b>94,7</b>	<b>95,0</b>
<b>C-V</b>	97,1	95,2	94,5	92,2

## VI. CONCLUSÕES

Neste trabalho, tendo como base a classificação de sinais de sonar passivo, foi discutida a seleção apropriada de conjuntos de projeto e teste, visando permitir o desenvolvimento de classificadores de elevada eficiência de generalização, em especial, para problemas que envolvam conjunto de dados com restrição estatística e razoável número de classes, tal qual o apresentado.

Considerando um conjunto de dados formado por 25 navios, 263 corridas e 8 classes distintas, foi verificado que a seleção do conjunto de dados para desenvolvimento e teste por técnicas de agrupamento permitiu a obtenção de um classificador com maior valor de eficiência média, assim como de melhor desempenho em relação às classes de identificação mais críticas. A seleção por agrupamento também se mostrou mais conveniente do ponto de vista computacional que a por validação cruzada, uma vez que o custo envolvido no treinamento da rede ARTM é significativamente menor que o da rede classificatória. Assim, conjuntos de elevada dimensionalidade e número significativo de eventos e classes podem exigir, na seleção por validação cruzada, um grande número de ensaios, o que pode ser crítico ou mesmo vir a inviabilizar o processo de seletivo.

Na produção dos agrupamentos, a adoção de componentes principais de discriminação (PCD) para a compactação dos dados se mostrou extremamente promissora, produzindo um agrupamento mais complexo quando comparado a compactação por componentes principais, mas que permitiu a obtenção de um classificador de maior capacidade de generalização, ainda que fossem utilizadas apenas a metade do número de componentes, somente cinco no total, numa dimensão original de 557 componentes.

Foi verificado que, para os agrupamentos produzidos por ambas técnicas de compactação (análises PCA e PCD), aqueles

de granularidade mais larga tenderam a produzir melhores classificadores, o que nos motivou a iniciar a pesquisa de uma metodologia para definição da granularidade ótima, que está em andamento.

Como trabalhos futuros, buscar-se-á aplicar outras técnicas para a produção dos agrupamentos, assim como avaliar-se-á a utilização de outros índices e métricas para o seu dimensionamento. Pretende-se ainda, por meio dos agrupamentos produzidos, extrair outras informações das corridas e dos navios, como, por exemplo, suas condições operativas e as características de seu maquinário.

## VII. AGRADECIMENTOS

Nossos agradecimentos ao CNPq, CAPES e FAPERJ, pelo apoio concedido ao projeto, e ao Instituto de Pesquisas da Marinha (IPqM), pela disponibilização do banco de dados e pela cooperação técnica, fundamentais para o desenvolvimento deste trabalho.

## REFERENCES

- [1] W. Soares Filho, "Classificação do ruído irradiado por navios usando redes neurais," Tese de doutorado, COPPE/UFRJ, 2001.
- [2] S. Haykin, *Neural Networks: a Comprehensive Foundation*. Prentice-Hall, 1999.
- [3] M. Stone, "Cross-validatory choice and assessment of statistical predictions," vol. Journal of the Royal Statistical Society, no. B36, pp. 111–147, 1974.
- [4] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.
- [5] L. Calôba, F. S. Pereira, and J. Seixas, "Neural discriminating analysis for a second-level trigger system," in *International Conference on Computing in High Energy Physics*, Rio de Janeiro, RJ, Brasil, 1995.
- [6] G. Golub and C. Loan, *Matrix Computations*, T. J. H. U. Press, Ed. The Johns Hopkins University Press, 1989.
- [7] K. Diamantaras and S. Kung, *Principal Components Neural Networks - Theory and Applications*. John Wiley and Sons, 1996.
- [8] G. Carpenter and S. Grossberg, "Art2: Self-organization on stable category recognition codes for analog input patterns," *Applied Optics*, vol. 26, pp. 4919–4930, 1987.
- [9] M. Vassali, J. M. Seixas, and L. Calôba, "A neural particle discriminator based on a modified art architecture," in *IEEE International Symposium on Circuits and Systems*, vol. II, 2002, pp. 121–124.
- [10] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1991.
- [11] A. Oppenheim, *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [12] R. O. Nielsen, *Sonar Signal Processing*. Artech House, 1991.
- [13] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. John Wiley & Sons, 2001.
- [14] D. Rumelhart, G. Hinton, and R. Williams, *Learning internal representations by error propagation*, parallel distributed processing: exploration in the microstructure of cognition ed. MIT Press, 1986, vol. 1, pp. 318–362.
- [15] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525–533, 1993.
- [16] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in *Proc. of the IEEE Intl. Conf. on Neural Networks*, 1993, pp. 586–591.
- [17] W. Soares Filho, J. M. Seixas, and L. Calôba, "Principal component analysis for classifying passive sonar signals," in *IEEE International Symposium on Circuits and Systems*, 2001, pp. 1–4.
- [18] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.