ANÁLISE COMPARATIVA ENTRE MÉTODOS DE RECONHECIMENTO DE PADRÕES APLICADOS AO PROBLEMA GÁS-LIFT INTERMITENTE

Dissertação submetida à
Universidade Federal da Bahia
como parte dos requisitos exigidos
para obtenção do título de
mestre em engenharia elétrica.

por
Oswaldo Ludwig Júnior
Setembro 2004

Em memoria de Oswaldo Ludwig

Abstract

The purpose of this work is to compare statistical techniques and connectionist approaches in the recognition of fail patterns on an oil extraction process that receives the "Gas-lift"denomination. Connectionist approach is based on feedforward multilayer artificial neural network and self-organizing maps. The obtained results are justified by the VC dimension analysis. Alternative methods to features extraction problems are also applied. This work contributes to the current state of the art by the optimization of the Kohonen networks through the use of the Battacharyya distance. It is also contribution of this work the use of the entropy level in the characteristics extraction process. The obtained results are susceptible to comparison with the results from the Project SGPA II, developed by the Department of Mechanical Engineering of UFBA, with the purpose of to complement and to validate the two studies mutually.

Resumo

O propósito desta dissertação é comparar técnicas estatísticas e conexionistas na tarefa de reconhecimento de padrões de falhas no processo de elevação de óleo denominado *Gas-Lift* Intermitente. A abordagem conexionista se baseia em redes neurais artificiais *feedforward multilayer* e mapas auto-organizáveis. Os resultados obtidos são justificados por meio de análise teórica da dimensão VC dos classificadores. O trabalho também aplica métodos alternativos para o problema de extração de características. No que se refere ao atual estado da arte, este trabalho contribui otimizando o desempenho das Redes de Kohonen por meio do emprego da distância de Battacharyya na determinação do neurônio a ser ativado em resposta a um sinal de entrada. É também contribuição deste trabalho o uso do nível de entropia no processo de extração de características. Os resultados obtidos são passíveis de comparação com os resultados oriundos do Projeto SGPA II, em desenvolvido pelo Departamento de Engenharia Mecânica da UFBA, com a finalidade de complementar e validar mutuamente os dois estudos, propiciando uma significativa referência de desempenho.

Agradecimentos

Ao Professor Dr. Antonio Cezar de Castro Lima, pela orientação recebida durante o desenvolvimento deste trabalho.

Ao Professor Dr. Leizer Schnitman e a todos os professores do mestrado pelo apoio recebido.

Conteúdo

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

O reconhecimento de padrões (RP) é a área de pesquisa que tem por objetivo a classificação de um dado, ou conjunto de dados, em categorias ou classes. Assim, considerando um universo de classes e um conjunto de dados, um reconhecedor de padrões, por vezes auxiliado por pré-processamentos, extrai, seleciona características e associa o rótulo de uma classe aos anteditos dados.

Segundo Ritter e Kohonen [1], o cérebro dos animais superiores seria uma intrincada "máquina" de classificação [3]. As técnicas de classificação de padrões são, normalmente, utilizadas para classificar objetos. Contudo, nada as impede de classificar relacionamentos, regras ou estados [2]. Tal fato as eleva à posição de uma ferramenta dotada de robustez, automatismo e aplicabilidade a inúmeros domínios do conhecimento humano.

Nas últimas décadas avanços significativos foram obtidos nesta área de pesquisa que, dentre outros aspectos, possibilitaram a evolução para aplicações nas mais diversas áreas da engenharia. Dentre um universo crescente de aplicações, destacam-se as aplicações em automação de processos [12, 28, 29]. Tais aplicações requerem técnicas eficientes e robustas de reconhecimento de padrões.

Pelo sucesso em aplicações práticas e registros em referências bibliográficas atualizadas, o autor considera que as técnicas de Inteligência Artificial (IA) estão entre as principais responsáveis pelo processo de evolução das técnicas de RP [8, 11, 15].

As técnicas de inteligência artificial tem ganhado espaço nas mais diversas áreas de atuação [5, 10, 11], sendo estas protagonistas de uma verdadeira revolução nos conceitos das engenharias ocorrida nas duas últimas décadas.

Existem inúmeras abordagens ao problema de classificação de padrões. Este texto explora o RP com o uso de técnicas estatísticas, redes neurais artificiais (RNA) *feed-forward multilayer* e *self-organizer maps*. Estes métodos são comparados, no que tange ao desempenho, pela aplicação ao problema de reconhecimento de padrões de falha no processo de elevação de óleo denominado *gas-lift* intermitente [28, 29]. A operação deste mecanismo consiste, basicamente, na injeção intermitente de gás em poços de petróleo de forma a elevar até a superfície a coluna de petróleo acumulada nas tubulações.

Os resultados obtidos são justificados pelo uso de ferramentas teóricas que possibilitam mensurar a capacidade de separação de padrões dos classificadores, bem como, a complexidade da amostra necessária para o treinamento.

A principal motivação deste estudo é oferecer à industria local uma solução adequada ao problema de classificação de padrões de falha no processo *gas-lift* intermitente.

O presente texto está dividido em 6 capítulos: o Capítulo 2 introduz alguns métodos de extração de características e classificação. O Capítulo 3 descreve o problema a ser tratado. As técnicas empregadas neste estudo são descritas com maior detalhe no Capítulo 4. Os resultados dos experimentos são apresentados no Capítulo 5. Finalmente, o Capítulo 6 tece considerações sobre os resultados e sugere trabalhos futuros.

Capítulo 2

Algumas Ferramentas Empregáveis a Problemas de Reconhecimento de Padrões

2.1 Abordagem Estatística ao Reconhecimento de Padrões

Os algoritmos de RP baseados em estatística subdividem o problema de classificação em duas tarefas distintas: a extração de características e a comparação e casamento destas características com as de modelos perfeitos, livres de ruídos, representativos dos seus respectivos padrões. Estas tarefas são realizadas por dois módulos denominados de extrator de características e classificador [18, 19, 20].

A título de exemplo, a tarefa de classificar um indivíduo quanto às classes esbelto, mesomorfo e gordo pode adotar um vetor de características formado pelo valor da altura e do peso do indivíduo.

As características extraídas compõem um conjunto de valores numéricos, que devem ser suficientes para a representação adequada dos dados de entrada, no que tange à tarefa de classificação em questão. Este conjunto de valores é representado pelo vetor de características. Sendo assim, um objeto pode ser representado como um ponto em um espaço de características.

O modelo ou protótipo μ_i representativo de uma classe ou padrão i é, em geral, obtido de um conjunto $\aleph$ de exemplos de vetores de características x_i pertencentes a este mesmo padrão, por meio da estimação do vetor média

$$\mu_i = \frac{\sum_{n=1}^{N} x_i[n]}{N} \cong E(x_i) = \lim_{N\to\infty} \frac{\sum_{n=1}^{N} x_i[n]}{N} \tag{2.1}$$

sendo $N = |\aleph|$ o número de elementos do conjunto $\aleph$ e $E(.)$ o operador esperança matemática.

A Equação 2.1 é adequada ao treinamento por lote de exemplos, entretanto, algumas aplicações exigem um algoritmo recursivo. Neste caso o treinamento é seqüencial e o cálculo do vetor média μ_i é dado por:

$$\mu_i[n+1] = \frac{n}{n+1}\mu_i[n] + \frac{1}{n+1}x_i[n+1] \tag{2.2}$$

onde n é o número da iteração.

O casamento de um dado de entrada com um padrão específico se fundamenta em medidas de "distância", entre o modelo representativo deste padrão μ_i e o dado de entrada x, ou seja, uma medida de distância entre os dois pontos no espaço de características, que também pode ser interpretada como a diferença entre os dois vetores de características.

A regra de decisão adota a menor distância, ou seja, o padrão cujo modelo possui o vetor de características mais próximo ao vetor de características do dado a ser classificado é adotado como a classificação adequada a este dado.

Existem formas usuais para aferir a distância r entre dois pontos $x = [x_1, x_2, ..., x_n]^T$ e $\mu = [\mu_1, \mu_2, ..., \mu_n]^T$ do espaço de características, das quais podemos citar:

Distância euclidiana:

$$r = \|x - \mu\| = \sqrt{\sum_{n=1}^{N}(x_n - \mu_n)^2} \tag{2.3}$$

Distância manhattan:

$$r = \sum_{n=1}^{N} |x_n - \mu_n| \tag{2.4}$$

Distância de Mahalanobis:

$$r^2 = (x - \mu)^T C^{-1}(x - \mu) \tag{2.5}$$

onde C é a matriz de covariância do vetor de dados x.

Uma outra abordagem consiste em mensurar a similaridade entre dois pontos do espaço de características por meio do produto interno ou produto escalar:

$$s = \frac{x^T \mu}{\|x\| \cdot \|\mu\|} = \cos(\theta) \tag{2.6}$$

onde θ é o angulo entre os vetores x e μ.

É fácil verificar que o produto interno é máximo quando o angulo θ é zero. Neste caso, a similaridade entre os padrões é máxima.

Classificadores que se baseiam na distância de Mahalanobis têm maior capacidade de separação de padrões, pois têm fronteira de decisão na forma de superfícies quadráticas, entretanto, o custo computacional para o cálculo da inversa da matriz de covariância cresce em proporção direta ao quadrado da dimensão do vetor de características, tornando-se inviável em muitas aplicações práticas.

2.2 A Dimensão VC

A dimensão VC (i.e. Vapnik-Chervonenkis) [9] é uma medida de capacidade de classificação de uma família de funções que compõe um classificador de padrões.

Define-se a dicotomia de um conjunto $\aleph$ por uma partição de $\aleph$ em dois subconjuntos disjuntos. Assim, diz-se que um conjunto $\aleph$ de instâncias é estilhaçado pelo Espaço de Hipóteses ou família de funções $\Im$, se e só se para qualquer dicotomia de $\aleph$ existe alguma hipótese de $\Im$ consistente com esta dicotomia.

Mais precisamente, a dimVC($\Im$) é o tamanho do maior subconjunto finito de $\aleph$, estilhaçado por $\Im$. A título de exemplo, ilustra-se um classificador baseado em distância euclidiana. Neste caso, tem-se uma família de funções, descritas conforme a Equação 2.3 (i.e. uma para cada classe que formam o vetor de distâncias r).

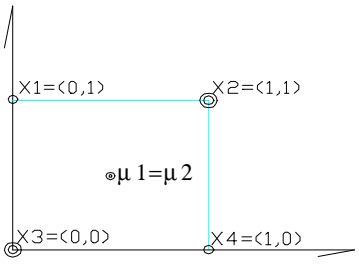

Figura 2.1: Separação de padrões do problema XOR.

Este classificador tem fronteiras de decisão (i.e. regiões que contêm todos os pontos eqüidistantes a dois protótipos) lineares (i.e. determinadas por linhas, planos ou hiperplanos). Sendo assim, sua competência se limita a padrões linearmente separáveis. Disto resultam limitações como a ilustrada na Figura 2.1, de onde é possível concluir que classificar os dados x_2 e x_3 como pertencentes ao padrão μ_1 e x_1 e x_4 como pertencentes ao padrão μ_2, com uma fronteira de decisão linear, é impossível no espaço de características $\Re^2$, visto que os vetores média dos dois padrões coincidem.

Entretanto, se houvessem apenas três pontos a serem classificados ou caso o espaço de características possuísse três dimensões, não haveriam problemas de classificação. Esta limitação é abordada no estudo da RNA *perceptron* quando é tratado o clássico problema XOR (i.e. "ou exclusivo") [9] e pode ser prevista através da análise da dimensão VC do classificador em questão que, neste caso, onde o espaço de características tem duas dimensões, é igual a três. Isto significa que este classificador só é capaz de classificar, com probabilidade de erro zero, 3 dados no espaço de características $\Re^2$.

A dimensão VC de um classificador baseado em um conjunto $\Im$ de funções de distância euclidiana ou produto interno em um espaço de características $\Re^{\mathbf{n}}$ é definida conforme segue:

$$VCdim(\Im) = n + 1 \tag{2.7}$$

onde $\boldsymbol{n}$ é o número de parâmetros ajustáveis, i.e., o número de dimensões do vetor média μ que, neste caso, é igual à dimensão do espaço de características.

A dimensão VC indica quantos dados podem ser classificados, com percentual de erro zero, pelo classificador $\Im$. Assim, se o vetor de características de um classificador do tipo mínima distância euclidiana, tiver dimensão 10, é possível assegurar a classificação adequada de, apenas, 11 dados, enquanto que classificadores como Redes Neurais Artificiais (RNA) do tipo *feedforward* de múltiplas camadas com função de transferência sigmóide nos neurônios da camada oculta e linear na camada de saída, têm dimensão VC definida por [21]:

$$VCdim = O(n^2) \tag{2.8}$$

onde $O(f(n))$ é a ordem de $f(n)$, significando que existem constantes positivas c e n_0 tal que $VC\dim \leq cf(n)$, para todo $n > n_0$.

2.3 Complexidade da Amostra

Considerando um espaço de hipóteses $\Im$, para se determinar quantos elementos são necessários para garantir a aprendizagem, é necessário a determinação da dimensão de Vapnik-Chervonenkis, conforme ilustra a Equação 2.9 [32]:

$$N \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC\dim(\Im)\log_2\frac{13}{\varepsilon}\right) \tag{2.9}$$

onde ε é o parâmetro que especifica o erro permitido no processo de aproximação do conceito alvo pelo conjunto de funções $\Im$ e δ é o parâmetro que controla a probabilidade de se construir uma boa aproximação, denominado parâmetro de crença.

A Equação 2.9 independe da distribuição de probabilidade das amostras, ou seja, da complexidade do problema, pois o método considera o pior caso. Assim, na prática, é possível obter um bom nível de treinamento com um número menor de exemplos.

2.4 Dificuldades do Processo de Extração de Características

A determinação das propriedades relevantes que irão compor o vetor de características, exige do projetista um alto nível de conhecimento a respeito do problema abordado, visto que tais características dependem intimamente do problema específico em questão. Assim, de um mesmo conjunto de dados, pode-se extrair vetores de características distintos, de acordo ao tipo de tarefa de classificação a ser implementada. Como exemplo podemos citar as tarefas de reconhecimento de fonemas e de locutor. Ambas recebem o mesmo conjunto de dados, entretanto, os vetores de características adotados podem diferir da primeira para a segunda tarefa. A extração de características via especialista é mais que ciência, é uma forma de "arte".

Uma escolha inadequada do vetor de características pode comprometer a performance de um classificador. Dentre os inúmeros transtornos oriundos de uma extração inadequada de características, pode-se citar:

·Características correlatas: neste caso as informações contidas no vetor de características são redundantes, visto que algumas de suas coordenadas podem ser expressas, com algum prejuízo, como função linear de outras. Sendo assim, a quantidade de informação contida no vetor de características pode ser insuficiente para caracterizar adequadamente o dado a ser classificado.

·Características em escalas inadequadas: quando, por exemplo, uma característica é medida em "Km" e outra em "m", a segunda poderá ter uma relevância ou peso maior no cálculo da distância ou desvio r.

·Vetores de características que exigem fronteiras de decisão não lineares para a classificação de padrões no espaço de características, ou seja, exigem classificadores muito complexos (i.e. com grande dimensão VC, a exemplo de redes neurais artificiais).

·Características inadequadas: as características são, simplesmente, irrelevantes ao problema de classificação ou insuficientes em número para diferenciar um padrão de outro.

2.5 RNA *Feedfoward Multi-Layer* em Reconhecimento de Padrões

RNAs *feedfoward multilayer* são classificadores eficientes, por serem capazes de produzir superfícies complexas para fronteiras de decisão. Tal fato é revelado pela análise da dimensão VC, a qual é proporcional ao quadrado do número de parâmetros livres da RNA [21].

Usualmente, para um problema de classificação de m classes, é utilizada uma RNA com m neurônios na camada de saída [2, 24]. Assim, a rede é treinada com valores-alvo binários, ou seja, o vetor de saída-alvo tem todas as suas coordenadas nulas, exceto aquela que indexa a classe a qual o dado de entrada pertence, cujo valor deve ser unitário. Esta abordagem permite uma melhor análise teórica, por facilitar o cálculo da dimensão VC.

É freqüente o uso de função de transferência sigmóide nos neurônios das camadas ocultas e linear na camada de saída. Esta configuração viabiliza analogias com classificadores estatísticos, conforme será elucidado mais adiante.

A regra de decisão para a "saída apropriada", mais usual, é a regra *bayesiana*, ou seja, a classe adequada é aquela indexada pela maior coordenada do vetor de saída. Assume-se precisão aritmética infinita, logo, empates não são possíveis.

Uma rede neural, conforme descrita nos parágrafos anteriores e ilustrada na Figura 2.2, é caracterizada por duas matrizes: $W_1 \in \Re^n \times \Re^j$ e $W_2 \in \Re^j \times \Re^m$. A matriz W_1 contém os pesos sinápticos entre a camada de entrada e a camada oculta, enquanto W_2 contém os pesos sinápticos entre a camada oculta e a camada de saída. A função de transferência adotada nos neurônios da camada oculta é usualmente a sigmóide.

Os *Bias* $b_1 \in \Re^j$ e $b_2 \in \Re^m$ também são levados em consideração. Neste caso, a propagação do sinal de entrada nesta RNA é dada por:

$$\mathbf{y}_h = \varphi\left(\mathbf{x}\,\mathbf{W}_1 + \mathbf{b}_1\right) \tag{2.10}$$

e

$$\mathbf{y} = \mathbf{y}_h\mathbf{W}_2 + \mathbf{b}_2 \tag{2.11}$$

onde $\varphi(.)$ é a função sigmóide e $\mathbf{y}_h$ é o vetor de saída da camada oculta.

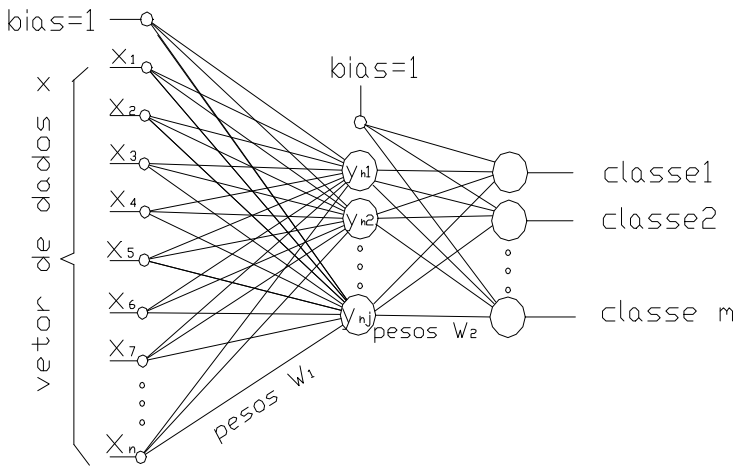

Figura 2.2: Grafo de fluxo de uma RNA *feedfoward* para RP.

O treinamento desta RNA pode ser associado a um problema de otimização numérica, onde a função objetivo a ser minimizada é o erro global, calculado conforme descrito na Equação 2.12:

$$\epsilon_{glob} = (y_d[n] - y[n])(y_d[n] - y[n])^T \tag{2.12}$$

onde $y[n]$ é o vetor de saída da rede e $y_d[n]$ é o vetor de saída alvo (i.e. saída desejada) na iteração n.

O ϵ_{glob}, assim como $y[n]$, é função das matrizes de pesos sinápticos W_1, W_2 e dos vetores de bias b_1 e b_2. Logo, é possível aproximar esta superfície de erro por meio da Série de Taylor de 1^a ordem em torno do ponto corrente (W_1, W_2, b_1, b_2). Assim, considerando $W_k = \{w_{k1}, w_{k2}, ..., w_{kj}, ..., w_{kJ}\}$ a matriz de pesos sinápticos da camada k, caso se deseje calcular o ajuste a ser aplicado ao vetor de pesos do neurônio j da camada k, representado por w_{kj} faz-se, inicialmente, a seguinte aproximação:

$$\epsilon_{glob}(w_{kj}[n] + \Delta w_{kj}[n]) \cong \epsilon_{glob}(w_{kj}[n]) + g_{kj}[n]\Delta w_{kj}[n]^T \tag{2.13}$$

onde $g_{kj}[n]$ é vetor gradiente local do neurônio j da camada k.

Observando a Equação 2.13 é possível constatar que para minimizar o erro global é necessário que a parcela:

$$g_{kj}[n]\Delta w_{kj}[n]^T \tag{2.14}$$

seja negativa e com o maior módulo possível. Para que o módulo deste produto interno seja máximo os vetores $g_{kj}[n]$ e $\Delta w_{kj}[n]$ devem ser colineares.

$$\Delta w_{kj}[n] \backslash\backslash g_{kj}[n] \tag{2.15}$$

Para que o resultado seja negativo os sinais destes vetores devem ser opostos. A regra delta atende às condições supracitadas:

$$\Delta w_{kj}[n] = -\eta g_{kj}[n] \tag{2.16}$$

onde η é o parâmetro taxa de aprendizagem.

O cálculo do gradiente local para a camada oculta, i.e. $k = 1$, é calculado por meio da regra de cadeia, conforme segue:

$$g_{1j}[n] = \left|\frac{\partial \epsilon_{glob}}{\partial w_{1j}}\right|_{w_{1j}=w_{1j}[n]} = \frac{\partial \epsilon_{glob}}{\partial y}\frac{\partial y}{\partial y_{hj}}^T\frac{\partial y_{hj}}{\partial v_{1j}}\frac{\partial v_{1j}}{\partial w_{1j}} \tag{2.17}$$

onde y_h é o vetor de saída da camada oculta, $\frac{\partial \epsilon_{glob}}{\partial y} = 2(y - y_d)$, $\frac{\partial y}{\partial y_{hj}} = W_2(j,:)$ (i.e. a linha j da matriz W_2), v_{1j} é a função de ativação do neurônio j, $\frac{\partial y_{hj}}{\partial v_{1j}} = \varphi\prime$ e $\frac{\partial v_{1j}}{\partial w_{1j}} = x$

A taxa de aprendizagem η é responsável pela velocidade com que se dá a busca no espaço de pesos, em direção aos valores que resultam em um erro global mínimo. Quanto menor a taxa de aprendizagem, mais suave e precisa é a trajetória através do espaço de pesos, entretanto, o aprendizado é lento para valores pequenos de η. Em contraposição, ao se adotar um valor muito grande para η**,** tem-se modificações muito intensas nos pesos sinápticos e, consequentemente, uma busca oscilatória, ou

seja, os valores dos pesos "passam" do ponto ótimo $w_{ki}{}^*$ e são remetidos de volta repetidamente.

Uma técnica usual para aumentar η sem comprometer a estabilidade da rede, é a inclusão do termo momento à regra delta, conforme equação abaixo:

$$\Delta w_{kj}[n] = -\eta g_{kj}[n] + \alpha \Delta w_{kj}[n-1] \tag{2.18}$$

onde α é o parâmetro constante de momento. A função deste termo é amortecer Δw_{kj} quando o processo de busca ultrapassa $w_{ki}{}^*$, gerando um efeito estabilizador e evitando oscilações no processo de busca.

O método de otimização aqui descrito é denominado *backpropagation*. Mais especificamente, trata-se do método gradiente descendente com termo momento, cujo cálculo do vetor gradiente é propagado para trás pelo uso da regra de cadeia. Este método é utilizado para o treinamento do classificador neural empregado neste estudo.

Uma abordagem interessante à análise do comportamento de uma rede neural, na tarefa de classificação de padrões, é a associação dos neurônios ocultos à extratores de características e dos neurônios da camada de saída, cuja função de transferência é linear, à classificadores estatísticos baseados no critério do maior produto interno. Assim, os neurônios ocultos extraem características salientes que caracterizam os dados de treinamento, através de uma transformação não linear, i.e. Equação 2.10, dos dados de entrada x para um novo espaço chamado de espaço oculto ou espaço de características (i.e. espaço caracterizado pelo vetor de saída da camada oculta y_h , também denominado, vetor de ativação da camada de saída), onde os padrões devem ser linearmente separáveis para classificação pelos neurônios da camada de saída, através do produto interno ilustrado na Equação 2.11.

É tentadora a possibilidade de classificação de dados brutos, a exemplo de imagens na forma matricial, diretamente em uma RNA *feedfoward* como descrita acima, visto que esta, em teoria [2], é capaz de extrair características salientes do sinal bruto,

assim como classificar o vetor de características. Entretanto, a prática não justifica esta abordagem, devido às inúmeras limitações de ordem tecnológica. Por exemplo, uma rede projetada para esta função teria um número muito grande de parâmetros livres a serem ajustados, pois, devido a grande dimensão do vetor de entrada, a matriz de pesos sinápticos W_1 seria muito grande. Tal fato implicaria em um custo computacional elevado para a tarefa de treinamento da rede. A função de custo, a ser minimizada pelo algoritmo *backpropagation*, teria um grande número de mínimos locais, implicando em uma grande probabilidade do processo de aprendizagem convergir para um destes pontos. Outra limitação relevante é a grande cardinalidade, i.e. número de amostras, necessária ao conjunto de amostras, para o ajuste de um elevado número de parâmetros livres (vide equações 2.8 e 2.9).

2.6 Pré-Processamento para Extração de Características

O presente texto sugere a utilização de algumas técnicas para refutar dados de pouca relevância, ou seja, que contêm pouca informação. Esta medida é especialmente indicada nas aplicações de RNA *feedforward multilayer*, visto que o número de parâmetros livres cresce na razão direta do número de entradas. Assim, considerando-se a Equação 2.8, a dimensão VC da RNA cresce com o quadrado do número de entradas, assim como, a complexidade da amostra (vide Equação 2.9). Caso não se reduza a dimensão do vetor de entrada da RNA, pode ser necessário um conjunto inacessível de dados pré-classificados para o treinamento supervisionado da RNA.

A idéia central é truncar o vetor de dados x. Este vetor de dados truncado, é utilizado diretamente como vetor de características disponibilizado ao classificador.

O truncamento dos dados deve conservar ao máximo as informações capazes de

caracterizar os padrões adequadamente. Sendo assim, os parâmetros sugeridos pelo autor para a rejeição de características de menor relevância são a entropia, a correlação e a relevância ao problema em estudo.

Coordenadas do vetor de dados com nível de entropia abaixo do limiar estabelecido pelo projetista ou correlatas a outras acima do limiar adotado, devem ser eliminadas do conjunto de características salientes que compõem o vetor de características.

2.6.1 Análise de Entropia

O cálculo da matriz de correlação, tem um custo computacional elevado, que cresce com o quadrado do número de coordenadas do vetor de dados, visto que a matriz de correlação é quadrada. O vetor de entropia tem a mesma dimensão do vetor de dados, sendo assim, a verificação da entropia dos dados parece ser uma abordagem computacionalmente mais econômica.

Caso as características sejam quantizadas em um número finito de níveis discretos, é possível associar cada característica a uma variável aleatória discreta. Para calcular a entropia de uma variável aleatória discreta X, é necessário, inicialmente, calcular a quantidade de informação revelada após a ocorrência do evento $X = x_i$. Esta quantidade de informação está relacionada à raridade da ocorrência deste evento, ou seja, quão mais esperado é um evento, menos informação é obtida após a sua observação, por outro lado, um evento raro é cercado de circunstâncias muito específicas e reveladoras. Define-se a quantidade de informação adquirida após a observação do evento $X = x_i$ com probabilidade p_i como:

$$I(x_i) = \log(\frac{1}{p_i}) = -\log(p_i) \tag{2.19}$$

A relação inversa com a probabilidade denota a noção de "raridade de ocorrência". A escala é logarítmica, assim se $p_i = 1$, $I(x_i) = 0$, isto é, caso o evento seja 100%

previsível, a sua ocorrência não revela nada.

A entropia $H(X)$ é o valor médio da quantidade de informação que uma variável aleatória X pode revelar, ou seja, o valor médio de I sobre os N possíveis valores x_i que X pode assumir:

$$H(X) = E\left[I(x_i)\right] = \sum_{i=1}^{N} -\log(p_i)\ p_i \tag{2.20}$$

Caso as características sejam valores contínuos, a ferramenta adequada é a entropia diferencial. Neste caso, a probabilidade do evento $X = x_i$, sendo X uma variável aleatória contínua com densidade de probabilidade $f(x)$, é:

$$P\{X = x\} = f(x)dx \tag{2.21}$$

A quantidade de informação associada a este evento é:

$$I(x) = -\log(f(x)dx) \tag{2.22}$$

Sendo assim, a entropia h é o valor médio de I sobre todos os valores que a variável aleatória contínua X pode assumir:

$$h(X) = E[I(x)] = \int_{x=-\infty}^{\infty} -\log(f(x)dx)f(x)dx \tag{2.23}$$

É evidente que este valor é infinitamente grande, pois X, sendo uma variável contínua, pode assumir infinitos valores. Sendo assim, a probabilidade de ocorrer um valor específico tende a zero, implicando em uma quantidade de informação que tende a infinito. Entretanto, o relevante são as diferenças entre a entropia das diversas características. Então, desenvolvendo a Equação 2.23:

$$h(X) = -\int_{x=-\infty}^{\infty} \log(f(x))f(x)dx - \int_{x=-\infty}^{\infty} \log(dx)f(x)dx \tag{2.24}$$

A segunda parcela é comum a todas as características, de tal forma que ao efetuar-se a diferença entre a entropia de duas variáveis aleatórias contínuas, estas parcelas se anulam. Sendo assim, define-se a entropia diferencial como:

$$h'(X) = -\int_{x=-\infty}^{\infty} \log(f(x))f(x)dx \tag{2.25}$$

Considerando-se a redução de dimensão do vetor de dados, após a rejeição dos dados que contêm menor quantidade de informação, estuda-se a viabilidade do cálculo da matriz de correlação para rejeição de dados correlatos, visto que estes podem ser determinados, com alguma imprecisão, por uma função linear de outros. Sendo assim, tais dados apresentam informação redundante.

É importante salientar que o processo de truncamento não realiza operações sobre o vetor de dados de entrada de forma a possibilitar a sua compressão sem perda de informação. Entretanto, a rejeição de dados com uma quantidade muito pequena ou nula de informação, a um baixo custo computacional, reduz a dimensão do vetor de dados e se mostra uma boa forma de pré-processamento, no sentido de viabilizar o processamento de uma RNA *feedfoward*, que por sua vez, realiza a extração de características relevantes à tarefa de classificação específica, via operações não lineares na camada oculta.

Os antecitados critérios de redução de dimensão do vetor de características independem do problema de classificação a ser tratado. Estas ferramentas caracterizam extratores de características não supervisionados. Entretanto, é possível que características com nível de entropia e grau de correlação adequados não tenham pertinência ao problema de classificação a ser estudado. Assim, em alguns casos, é necessário um extrator de características supervisionado, ou seja, que considere as saídas alvo.

2.6.2 Análise da distância de Battacharyya

A distância de Battacharyya [31] é uma medida de afastamento entre duas distribuições de probabilidades, caracterizadas por suas respectivas funções de densidade de probabilidade. Assim, sejam $f(x_n \mid c_1)$ e $f(x_n|c_2)$ as densidades de probabilidade da característica $x_n \in R$, associada às classes c_1 e c_2 respectivamente. Define-se distância de Battacharyya entre as duas distribuições por:

$$B_n = \frac{1}{\log(\rho_n)} \tag{2.26}$$

onde:

$$\rho_n = \int_R \sqrt{f(x_n \mid c_1)\ f(x_n \mid c_2)}\ dx_n \tag{2.27}$$

Observando a Figura 2.3 é possível concluir que, no caso ilustrado, $\rho \cong 0$ implicando em um grande valor para distância B. Ao passo que a Figura 2.4 ilustra um caso em que ρ é maior, implicando em uma distância B é menor.

O método de extração de características por meio da distância de Battacharyya consiste em selecionar as características que possuem a maior distância B entre as funções de densidade de probabilidade associadas às diferentes classes, dado que tais características são as que melhor possibilitam a dicotomia entre classes.

O emprego da distância de Battacharyya implica na determinação das funções de densidade de probabilidade associada de cada uma das coordenadas do vetor de características. Assim, caso hajam n características e m classes, é necessário a determinação de um número $q = nm$ funções de densidade de probabilidades (*fdp*). Este trabalho não adota funções de distribuição conjuntas. Assim, é notória a pressuposição de características x_n do vetor de características x incorrelatas, de onde resulta a simplificação:

$$P\{x = \chi\} = P\{x_1{=}\chi_1\}P\{x_2{=}\chi_2\}...P\{x_n{=}\chi_n\} \tag{2.28}$$

A determinação de uma *fdp*, normalmente, parte da suposição de que esta pertence a uma família de funções, a exemplo das gaussianas, definidas por um conjunto de parâmetros desconhecidos. O segundo passo é a estimação dos parâmetros das funções. O processo tem um elevado custo computacional e implica em um grande conjunto de amostras.

Uma alternativa é a discretização dos valores assumidos pelas características. Assim, a Equação 2.27 pode ser reescrita conforme segue:

$$\rho_n = \sum_{x \in \Omega} \sqrt{P\{x_n = \chi \mid c_1\}P\{x_n = \chi \mid c_2\}} \tag{2.29}$$

onde $P(x_n = \chi \mid c_1)$ e $P(x_n = \chi | c_2)$ são as probabilidades da característica discreta $x_n \in \Omega$ assumir o valor χ, associadas às classes c_1 e c_2 respectivamente.

Para valores discretos, a geração e normalização de histogramas é suficiente para a determinação da distância de Battacharyya.

Sabe-se que não existem "truques matemáticos" que possam suprir a falta de informação. Caso o projetista não incorpore o seu conhecimento no sistema de RP através do projeto do extrator de características, esta informação deverá estar presente em um conjunto de treinamento de grande cardinalidade.

2.6.3 Análise dos Componentes Principais

Na fase de treinamento, uma RNA supervisionada recebe um conjunto de exemplos de treinamento composto por pares de dados de entrada x e saída y_d. Estes dados são vetores cujas as coordenadas podem ser correlatas (i.e. uma coordenada pode ser expressa por meio de outras). Este fato implica em informações redundantes.

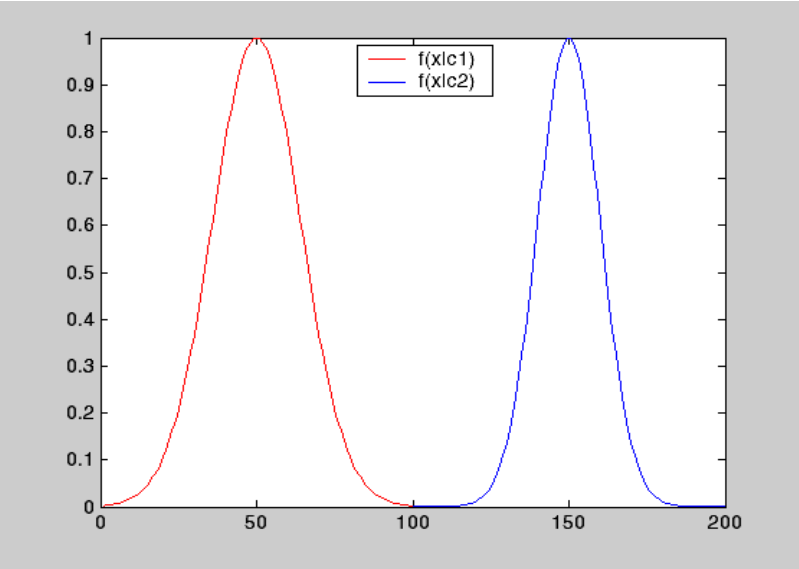

Figura 2.3: Funções de densidade de probabilidade de x associada as classes c_1 e c_2 com grande distância B.

Neste trabalho a Análise dos Componentes Principais (PCA) é aplicada aos dados de entrada x para eliminar a covariância entre suas coordenadas. O método de PCA permite a eliminação da covariância entre as coordenadas de x por meio de uma mudança de base.

Para aplicar o método PCA o primeiro passo é o cálculo da matriz de covariancia C. Em uma aproximação discreta, cada elemento desta matriz representa a covariância $s^2(x_i, x_j)$ entre as coordenadas x_i e x_j do vetor x:

$$s^2(x_i, x_j) = \sum_{n=1}^{N} x_i x_j - \frac{\sum_{n=1}^{N} x_i \sum_{n=1}^{N} x_j}{N} \tag{2.30}$$

onde N é o número de exemplos do conjunto de treinamento.

O segundo passo é a determinação da matriz diagonal λ de autovalores de C e a matriz P de autovetores de C. Os autovetores compõem as colunas da matriz P:

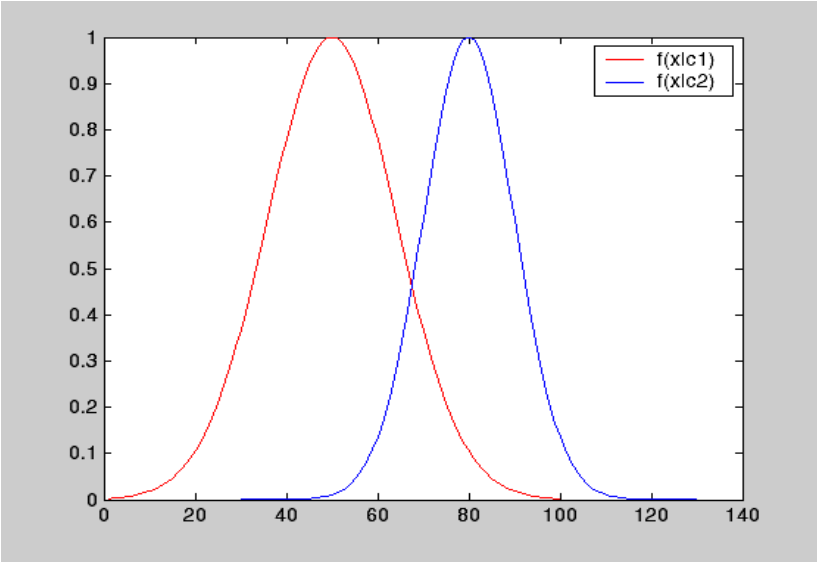

Figura 2.4: Funções de densidade de probabilidade de x associada as classes c_1 e c_2 com menor distância B.

$$P = \{v_1, v_2, ..., v_n\} \tag{2.31}$$

A matriz P pode ser empregada para mudar a base de C obtendo a matriz diagonal λ de autovalores de C:

$$CP = P\lambda \therefore P^{-1}CP = \lambda \tag{2.32}$$

Note que a matriz λ (i.e. C na base P) não apresenta covariância. Esta matriz diagonal possui apenas variâncias $s^2(x_i, x_i)$. A matriz λ é a matriz de covariância entre as coordenadas do vetor x na base P:

$$C = f(x) \therefore \lambda = P^{-1}CP = f(P^{-1}x) \tag{2.33}$$

A Equação 2.33 permite concluir que o vetor x na base P não possui covariância

entre as suas coordenadas. Em outras palavras, $P^{-1}x$ não possui informações redundantes. Assim, um menor número de coordenadas de $P^{-1}x$ é suficiente para conter grande parte da sua informação, ou seja, x na base P apresenta algumas coordenadas com grande variância enquanto outras têm variância próxima a zero.

A Figura 2.5 tem por finalidade a interpretação geométrica do método PCA. Nesta figura é possível perceber que a mudança para a base P elimina a covariância, acresce a variância da coordenada x_2 e diminui a variância de x_1.

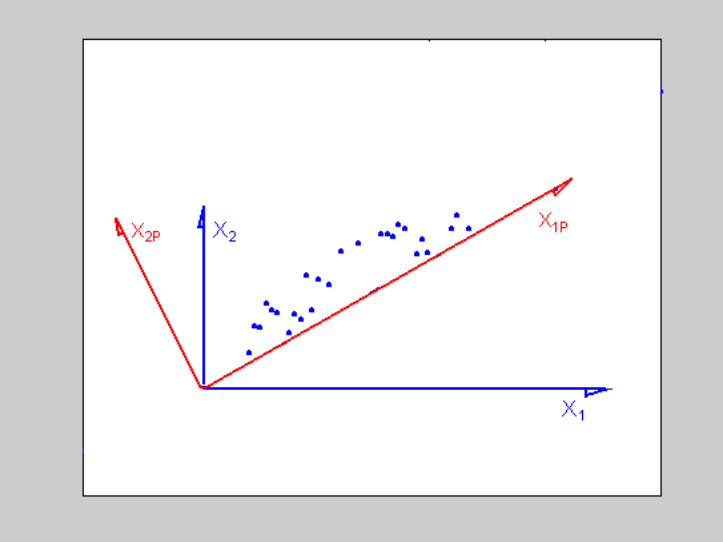

Figura 2.5: Efeitos da mudança para base P.

O objetivo do método PCA é reduzir a dimensão do vetor $P^{-1}x$ eliminando as coordenadas com baixa variância. Assim, no exemplo ilustrado na Figura 2.5 a coordenada x_{2P} seria eliminada.

2.7 Mapas Auto-Organizáveis em Reconhecimento de Padrões

O mapa auto-organizável (SOM) [7, 16] é uma rede neural de treinamento não supervisionado, cujo emprego é, usualmente, o reconhecimento de padrões. Nesta estrutura, os neurônios estão dispostos como nós de uma grade que, normalmente, é uni ou bidimensional. Cada neurônio representa uma classe e apenas um neurônio é ativado como resposta a um estímulo de entrada.

O princípio fundamental deste modelo é a aprendizagem competitiva, ou seja, ao se apresentar um sinal à rede, os neurônios competem entre si e o vencedor tem seus pesos ajustados para responder melhor ao supracitado estímulo. O algoritmo simula um processo de cooperação entre o neurônio vencedor e seus vizinhos topológicos, que também recebem ajustes. Desta forma, as características estatísticas intrínsecas contidas em um vetor de sinais de entrada, irão estimular alguma determinada localização espacial da rede, mais especificamente, aquela que contenha um grupo de neurônios sintonizados àquele estímulo. Sendo assim, podemos classificar esta rede como um paradigma topológico [6].

A motivação para a criação deste modelo neural é a teoria de que, no cérebro humano, entradas sensoriais diferentes são mapeadas em regiões específicas do córtex cerebral. A isto podemos denominar de "distribuição de probabilidade codificada por localização" [3], [6].

A Figura 2.6 ilustra um SOM com oito nós de fonte e quatro neurônios, dispostos em uma grade bidimensional quadrada.

Cada neurônio da grade está totalmente conectado com todos os nós de fonte da camada de entrada. Cada padrão de entrada apresentado à rede irá gerar um "foco" de atividade em alguma região desta grade.

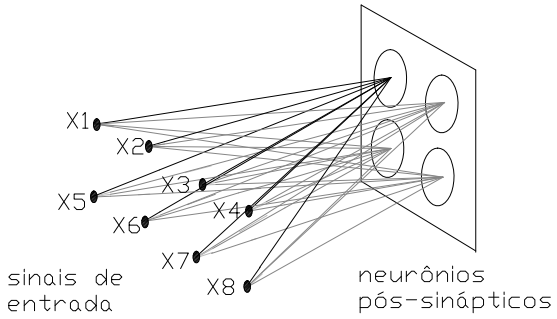

Figura 2.6: Grafo de fluxo de um SOM.

O algoritmo responsável pela organização da rede inicia o processo arbitrando pequenos valores aleatórios, em uma distribuição de probabilidades uniforme, aos pesos sinápticos, para que nenhuma organização prévia seja imposta ao mapa. Em seguida, dá-se os processos de competição, cooperação e adaptação sináptica.

O vetor de entrada, representado por $x = [x_1, x_2, ..., x_n]^T$, selecionado aleatoriamente dentre os demais exemplos de treinamento, é apresentado à rede sem que se especifique a saída desejada. Um neurônio k deverá responder melhor a esta entrada, ou seja, será o neurônio vencedor. O critério usual para a escolha do vencedor é o da menor distância euclidiana d_{kx} entre o vetor de entrada e os vetores de pesos sinápticos dos neurônios.

O neurônio vencedor indica o centro de uma vizinhança topológica de neurônios cooperativos. Estudos [3, 4] demonstram que a interação lateral entre um neurônio biológico estimulado e seus vizinhos topológicos, decresce suavemente na medida em que a distância lateral aumenta. O neurônio artificial, por analogia, tem a mesma propriedade, ou seja, o parâmetro vizinhança topológica h_{kj}, que indica o grau de interação ou cooperação entre o neurônio k e seu vizinho j, é simétrico em relação ao neurônio vencedor k e decresce monotonicamente com o aumento da distância lateral até que, no limite em que d_{kj} tende a infinito h_{kj} tende a zero. A função gaussiana atende a estas condições, sendo comum o seu uso:

$$h_{kj} = e^{(-\frac{d_{kj}^2}{2\sigma^2})} \tag{2.34}$$

O termo d_{kj}^2, no caso de uma grade bidimensional, é a distância euclidiana entre os neurônios k e j. O parâmetro σ é denominado de largura efetiva da vizinhança topológica e deve diminuir com o passar do tempo, o que implica em valores de h_{kj} menores ao longo do tempo, caracterizando uma vizinhança mais restrita e, portanto, mais especializada. O valor de σ é, normalmente, uma função exponencial:

$$\sigma\,[n] = \sigma_0\; e^{(-\frac{n}{\tau_1})} \tag{2.35}$$

onde σ_0 é o valor inicial de σ, n é o número de iterações e τ_1 é uma constante de tempo.

Da mesma forma que as demais redes neurais artificiais, o aprendizado de um mapa auto-organizável se dá pelo ajuste de seus pesos sinápticos. Ou seja, o peso sináptico w_{ij} entre o nó de entrada i e o neurônio j, sofre o ajuste Δw_{ij}, regido pela expressão:

$$\Delta w_{ij} = \eta\,[n]\;\; h_{kj}\,[n]\;\; (x_j - w_{ij}) \tag{2.36}$$

onde o termo $h_{kj}\,[n]$ é o parâmetro vizinhança topológica na iteração n, no qual o índice k se refere ao neurônio melhor classificado k.

O parâmetro taxa de aprendizagem $\eta\,[n]$ é, geralmente, definido pela expressão:

$$\eta\,[n] = \eta_0\, e^{-\frac{n}{\tau_1}} \tag{2.37}$$

onde τ_1 é uma constante de tempo e $\eta_0 \in [0, 1]$, é o valor inicial adotado.

É fácil constatar, ao observar a Equação 2.37, que a taxa de aprendizagem decresce gradualmente ao longo do tempo. A finalidade é evitar que dados novos, apresentados

após um longo treinamento, venham a comprometer seriamente o conhecimento que já está sedimentado.

É possível alterar os vetores de pesos sinápticos dos neurônios de um SOM já treinado de forma supervisionada [9], como uma forma de ajuste fino, a fim de melhorar a qualidade das regiões de decisão do classificador. Esta técnica é conhecida como quantização vetorial por aprendizagem. A idéia é mover o vetor de pesos sinápticos w do neurônio vencedor em direção ao vetor de entradas x, caso a classificação esteja correta, caso contrário, o vetor de pesos é afastado do vetor de entrada:

$$w\left[n+1\right] = w\left[n\right] + \alpha\ \left(x\left[n\right] - w\left[n\right]\right) \tag{2.38}$$

, se classificação correta.

$$w\left[n+1\right] = w\left[n\right] - \alpha\ \left(x\left[n\right] - w\left[n\right]\right) \tag{2.39}$$

, se classificação incorreta. Define-se $0<\alpha<1$ como o coeficiente de aprendizagem, cujo valor deve ser determinado pelo usuário.

Apenas o neurônio vencedor é ajustado, os demais neurônios não sofrem modificações em seus vetores de pesos.

2.8 Propriedades do Mapa Auto-Organizável em Reconhecimento de Padrões

O SOM é usualmente empregado para classificação de dados não rotulados [4]. Neste caso, o SOM apresenta propriedades interessantes tais quais:

· Aproximação do espaço de entrada $\aleph$ para o espaço discreto de saída $\wp$, onde se distribuem os neurônios. Possibilitando a compressão de dados.

· Ordenação topológica dos dados, ou seja, a localização espacial de um neurônio

corresponde a um conjunto de características em especial dos dados de entrada. Assim, neurônios vizinhos correspondem a classes semelhantes.

· Casamento aproximado de densidade de probabilidade. Regiões do espaço de entrada $\aleph$ com densidade de probabilidade maior são mapeadas em domínios maiores do espaço de saída $\wp$ (i.e. têm mais neurônios associados).

Do que foi colocado, é possível concluir que o vetor de pesos sinápticos w de cada neurônio de um SOM, corresponde, em uma analogia com a abordagem estatística, a um protótipo μ, que caracteriza uma determinada classe (vide Equação 2.1). Entretanto, diferentemente da abordagem estatística, o SOM dispensa o conhecimento das classes e de sua relação com os dados de entrada, na medida em que distribui os dados de entrada entre classes segundo os supracitados critérios de casamento de densidade e ordenação topológica.

Capítulo 3

Descrição do Método de Extração *Gas-lift* Intermitente

O método de elevação denominado "Gas Lift Intermitente" (GLI) opera em cerca de 37% dos poços de petróleo da região de produção da Bahia - E&P-BA, respondendo por 34% da produção de óleo na região.

Devido à importância deste método para a região, da quantidade de poços e da produção de óleo, justifica-se a implantação de um sistema para a automação deste método, de modo a permitir um rápido e preciso diagnóstico dos problemas operacionais, minimizando custos e maximizando a produção.

O método de reconhecimento de padrões atualmente em uso pela PETROBRAS UN-Ba é similar ao existente para o bombeio mecânico (BM), resultante de um projeto de pesquisa realizado em conjunto com a UNICAMP.

O objetivo principal é, utilizando os conceitos de RP e redes neurais, elaborar um sistema de automação inteligente que possa supervisionar, controlar, diagnosticar e propor soluções para os principais problemas que ocorrem com este método de elevação.

A supervisão consiste basicamente na aquisição de dados de pressão no revestimento (PR) e pressão na tubulação (PT) durante o ciclo de injeção de gás. O diagnóstico é baseado em RP, e visa gerar ações para o controle do tempo de ciclo, além de detectar situações para sinalizar ações preventivas.

O controle local ao poço, além de ligar e desligar o ciclo de injeção de gás em função das condições operacionais do poço, visa o ajuste do tempo de ciclo e do tempo de injeção do sistema.

3.1 Fundamentos Teóricos

A operação do GLI consiste, basicamente, na injeção intermitente de gás do revestimento para a coluna de produção, de forma a deslocar para a superfície a coluna de líquido acumulada nesta tubulação. A Figura 3.1 ilustra o processo.

O controle da injeção de gás, do revestimento para a coluna de produção, é realizado através de uma válvula operadora instalada no fundo do poço.

O controle de injeção de gás no revestimento é normalmente realizado usando um intermitor de ciclo, ou um estrangulador.

Um ciclo completo de operação do GLI compõe-se basicamente de três períodos: alimentação, elevação e redução de pressão, conforme descritos a seguir:

· Alimentação: Período durante o qual não ocorre injeção de gás para o revestimento e coluna de produção do poço (i.e. válvulas motora e operadora fechadas). O fluido da formação está se acumulando na coluna de produção, até se atingir uma coluna de fluido com extensão desejável para uma golfada de óleo.

· Elevação: Período durante o qual a golfada de líquido é lançada á superfície pela injeção de gás. Neste período ocorre injeção de gás no revestimento e coluna de produção do poço.

· Redução de Pressão: Período no qual a pressão de tubo é reduzida a um valor

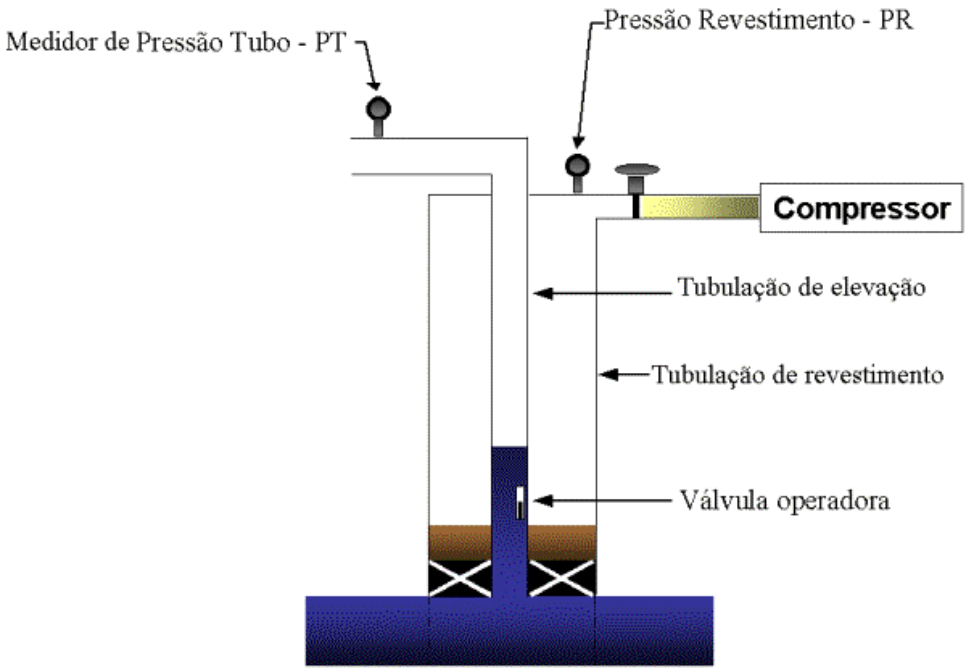

Figura 3.1: Estrutura Funcional de um Poço de Petróleo que opera com *GLI.*

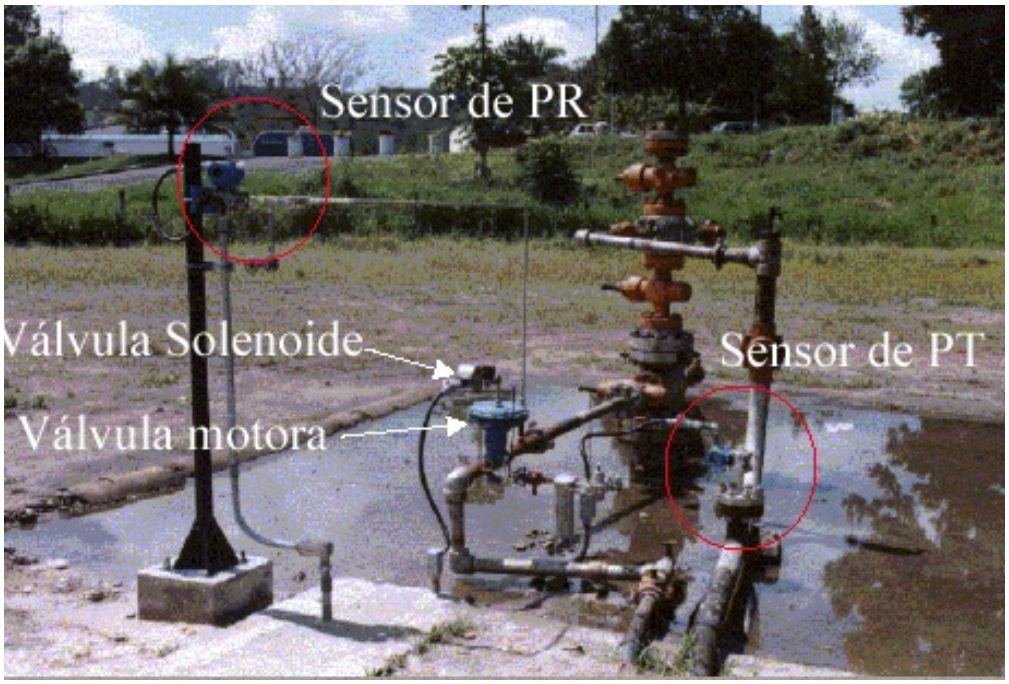

Figura 3.2: Estrutura física do Sistema *Gas Lift*

mínimo, através da liberação do gás usado para elevação da golfada.

3.2 Otimização do Sistema

A otimização do sistema é feita a partir do controle de dois parâmetros:

·***Fall-back*:** definido como a parcela de óleo localizada acima da válvula operadora no instante de sua abertura.

·***Razão gás-líquido de injeção*** (RGLI)**:** definida como a relação entre o volume de gás injetado e o volume de fluido produzido por ciclo.

Basicamente, este processo de otimização consiste em maximizar a produção de óleo, reduzindo a RGLI.

Os controles existentes no método, responsáveis pela sua maior ou menor eficiência, resumem-se na ciclagem de operação e dimensionamento da válvula operadora. Tal operação é basicamente dividida em dois tempos: **Tempo de injeção** (Ti) que é o intervalo de tempo que a válvula controladora na superfície (*motor valve*) permanece aberta, permitindo injeção de gás para o revestimento de produção e o **Tempo de ciclo** (Tc) que compreende o tempo que a válvula controladora permanece fechada, não injetando gás para o revestimento de produção, mais o tempo de injeção.

A válvula operadora é dimensionada de forma a que haja uma diferença de 200 a 300 psi entre a pressão de gás no revestimento e a pressão na coluna de produção no instante de sua abertura. Valores abaixo deste *range* podem indicar um valor alto do *"fall-back"*, implicando em perda de produção do poço. Valores acima deste *range* podem indicar um consumo adicional de gás. Este diferencial de pressão é influenciado pelo tempo de ciclo: maior tempo de ciclo, menor diferencial; menor tempo de ciclo, maior diferencial.

3.3 Acompanhamento da Operação

O acompanhamento da operação por meio da análise dos sinais de PR e PT, que fornecem um registro contínuo das pressões de tubo e revestimento na superfície, pode definir a necessidade de um simples ajuste na ciclagem de operação, assim como, a solicitação de serviço adequado para solução de problemas operacionais devido a danos em equipamentos.

3.4 O Problema de Reconhecimento de Padrões

A correta tomada de decisões para a solução de um problema no GLI passa, necessariamente, por uma boa capacidade de compreensão e a análise das cartas de PR

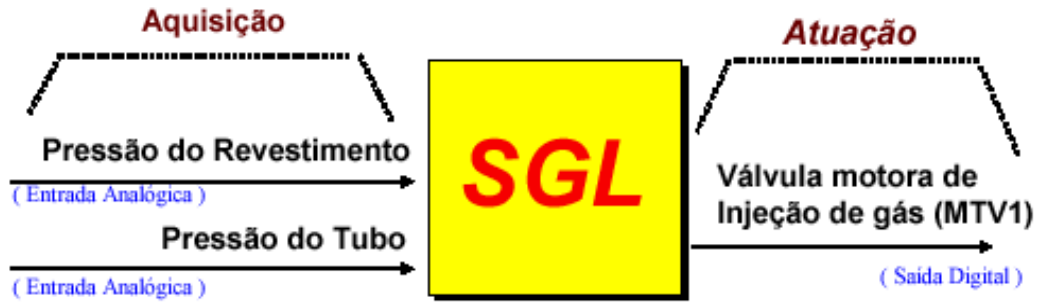

Figura 3.3: Diagrama do sistema de controle para poços de *Gas Lift* intermitente.

e PT. Por sua vez, tal compreensão passa por um conhecimento do funcionamento global do sistema. A falta de conhecimento mais profundos sobre todas as etapas do método do *gas lift* tem gerado, em alguns casos, interpretações erradas das cartas de PR e PT, levando a falsos diagnósticos sobre os problemas operacionais apresentados nos poços e, consequentemente, a solicitação de serviços desnecessários, acarretando não só perdas de produção como também aumento nos custos operacionais.

Muitas vezes, mesmo o poço estando operando sem qualquer tipo de problema operacional, ou seja, com a carta de PR e PT indicando boa operação, sua produção pode ser otimizada a partir de pequenos ajustes na sua ciclagem de operação. A ciclagem pode ser obtida, de forma aproximada, a partir de cálculos simples utilizando-se dados facilmente disponíveis.

O sistema SGL possui dois tipos de sinais adquiridos para o processo de controle: Pressão do Revestimento (PR) e Pressão do Tubo (PT). A resposta do controlador é o sinal de atuação sobre a solenóide de controle da injeção de gás, assim como, ativar avisos de falhas em componentes do sistema. O processo de RP deve determinar, a cada ciclo da operação, a que padrão de operação pertence o sinal de PR ou PT.

Neste trabalho é abordado o processo de classificação do sinal PR. O autor acredita que a metodologia aplicada ao sinal de PR poderá ser adequada ao sinal de PT em

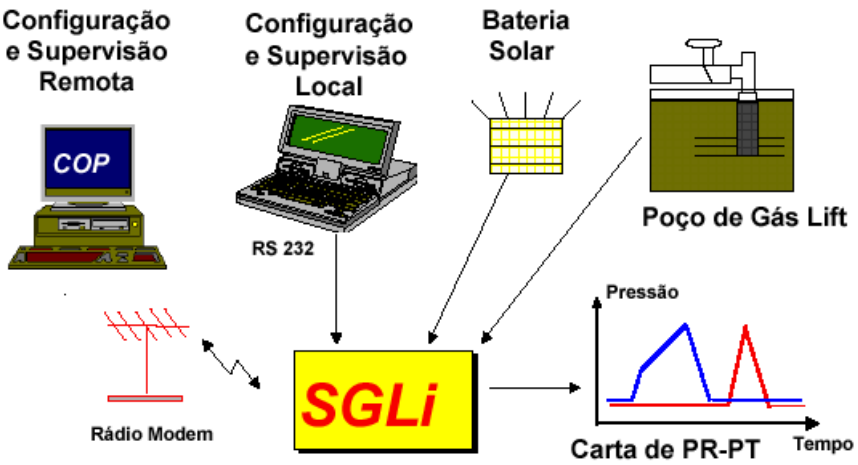

Figura 3.4: Arquitetura do Sistema GLI.

trabalhos futuros.

3.5 Arquitetura Funcional do Sistema Atual

A arquitetura atualmente prevista para a automação de poços equipados com GLi envolve a utilização de um produto desenvolvido pela PETROBRÁS dentro do projeto SICAD.

Resumidamente, o sistema envolve um microcomputador do tipo PC ligado a um rádio modem, transmitindo e recebendo sinais de diversas estações localizadas em poços. Essas estações são constituídas por dois sensores de pressão (pressão de tubo - PT, pressão de revestimento - PR), um dispositivo eletrônico PLC (Sistema controlador de GLi - SGL), válvulas solenóide e motora, e rádio modem, alimentados por bateria e placa solar.

O SGL, por meio dos sensores de pressão de revestimento e de tubo, adquire continuamente os valores das pressões em intervalos de tempo definidos pelo usuário.

É dotado de banco de dados próprio no qual são armazenadas as informações sobre as características mecânicas do poço juntamente com um conjunto de padrões para as diversas formas geométricas dos sinais de pressão.

O reconhecimento do padrão característico desses sinais se dá por meio de algoritmos desenvolvidos no projeto SICAD. Após a etapa de classificação, o sistema executa um conjunto de regras de controle e de tomadas de decisão. Regras estas fornecidas pelo usuário.

Por meio de comunicação com a solenóide responsável pelo acionamento a válvula motora, o sistema realiza alterações do tempo de ciclo (Tc) do sistema. A atualização dos dados do banco de dados é feita por meio de comunicação via rádio com o PC, o qual está localizado no Centro de Operações (COP) de engenharia ou na estação que recebe a produção dos poços.

O conjunto de regras de controle especificado pelo usuário (i.e. o algoritmo de controle do poço) é então processado e enviado do PC para o SGL via rádio modem.

O sistema SGL é Instalado no local do poço e consiste em um PLC que foi customizado para atender a esta aplicação, possuindo módulos de entradas/saídas analógico/digital, e opcionalmente uma entrada para aquisição de sinais de encoder.

O SGL possui duas saídas seriais RS422/485 de modo que, com o uso de um computador portátil é possível efetuar a supervisão e configuração local do sistema. Por meio de outra serial é possível disponibilizar a comunicação com a unidade de supervisão e controle central, localizada no COP, utilizando um rádio modem.

Capítulo 4

Técnicas de Reconhecimento de Padrões Aplicadas ao Problema *Gas-lift* Intermitente

Neste trabalho são comparados os desempenhos de cinco diferentes técnicas de RP. Em uma primeira etapa são avaliadas as combinações de duas formas distintas de extração de características (i.e. no domínio do tempo e da freqüência) com o classificador estatístico. Posteriormente, o desempenho do classificador estatístico otimizado pelo uso de entropia é avaliado com o uso de características extraídas no domínio do tempo. Finalmente, são aplicadas as técnicas conexionistas: RNA multilayer feedforward e SOM.

O problema de classificação de padrões de falhas no processo de elevação de petróleo por meio de injeção de gas não disponibiliza ao autor, até a presente data, os pares ordenados de dados para treinamento. Ou seja, informações a respeito dos sinais ou medidas da pressão P estão disponíveis, entretanto, o autor não tem informações a respeito da clasificação destes sinais por um engenheiro especialista (i.e. saída alvo para treinamento do classificador). Contudo, existem informações sobre os padrões

típicos de falhas, ou seja, existem protótipos definidos em literatura específica [28].

4.1 Descrição dos Trabalhos Precedentes

Nos trabalhos precedentes, os protótipos estão definidos por meio dos seus pontos significativos (PS) [22]. Estes pontos representam posições onde a derivada dP/dt apresenta mudanças acima de um determinado *threshold*. Isto equivale a valores de t para o qual $|\partial^2 P/\partial t^2| > \beta$, onde β é um valor limite arbitrado pelo projetista.

Atualmente, a análise dos sinais é proferida apenas no domínio do tempo por meio dos PS e requer procedimentos de pré-processamento dos dados, no intuito de resolver algumas dificuldades tais como:

· Elevado número de PS, devido a ruídos apresentados pelos dados reais. Tal fato força o pré-processamento do sinal, por meio de filtros, para eliminar o ruído e possibilitar a extração dos PS.

· O número de PS varia de acordo ao sinal. Isto resulta em vetores de características de dimensões distintas. A solução adotada em trabalhos anteriores foi a eliminação de alguns PS conforme descrito em [22].

· O procedimento atual considera uma dispersão limite em torno dos valores médios dos PS, no processo de casamento padrão×classe. Esta medida impossibilita a classificação de alguns padrões, visto que, em alguns casos, a distância euclidiana entre dois PS pertencentes a classes distintas é maior que a soma das medidas de desvio padrão limite dos dois PS. Tal fato implica na existência de regiões do espaço de características que não pertencem a nenhuma classe.

4.2 Descrição do Presente Trabalho

O presente trabalho interpola linearmente os PS e extrai amostras de pressão em 111 intervalos ao longo de um ciclo da operação de bombeio. Tais amostras compõem o vetor de características do classificador.

A extração de características no domínio da freqüência se dá por meio da *Fast Fourier Transform* (FFT).

As técnicas empregadas para avaliar a relevância de uma coordenada do vetor de característica com relação à tarefa de classificação a ser realizada são a análise da entropia da informação H e a distância de Battacharyya B. Neste trabalho os valores de pressão P assumidos por cada característica são discretizados em 10 níveis. Em seguida é gerado um histograma normalizado e finalmente aplicam-se as equações 2.20, 2.26 e 4.1 para o cálculo de $H(.)$ e $B(.)$ respectivamente.

O problema em questão contempla seis classes distintas. Neste caso, a distância de Battacharyya será dada por:

$$\rho_n = \sum_{x \in \Omega} \sqrt[6]{P\{x_n = \chi \mid c_1\} P\{x_n = \chi \mid c_2\} ... P\{x_n = \chi \mid c_6\}} \tag{4.1}$$

A Figura 4.1 ilustra a medida de entropia em ***nats*** calculadas para cada um dos coeficientes da FFT. Considerando-se a freqüência fudamental ϖ_0, vê-se que os coeficientes dos harmônicos contidos na faixa $[20\varpi_0,\ 56\varpi_0]$, têm entropia $H<1,3$; enquanto que na faixa $[0\varpi_0,\ 19\varpi_0]$ há valores $H>1,3$. A entropia é calculada sobre o conjunto de casos típicos descritos na literatura especializada [23].

A Figura 4.2 ilustra a medida de entropia em ***nats*** para cada uma das 111 amostras no domínio do tempo colhidas ao longo de um ciclo de operação, ao passo que a Figura 4.3 ilustra a distância de Battacharyya para cada uma das amostras colhidas ao longo de um ciclo de operação.

Comparando as figuras 4.2 e 4.3, observa-se que embora as amostras contidas no

intervalo [47, 80] tenham nível de entropia adequado, são pouco relevantes para a tarefa de separação dos padrões dentre as seis classes, conforme indica a Distância de Battacharyya. Na opinião do autor, a presença de ruído justifica o ocorrido. Note que, a título de exemplo, um ruído branco tem distribuição de probabilidades uniforme, assim, pode gerar elevado nível de entropia. Assim, é possível concluir que um elevado nível de entropia não implica necessáriamente em uma elevada quantidade de informação, por outo lado, um baixo nível de entropia implica em pouca informação.

No uso de classificadores estatísticos as coordenadas do vetor de características são normalizadas para que todas tenham a mesma relevância no cálculo da medida de distância entre um padrão e um protótipo. Caso contrário, uma característica com valores que variam de 0 a 100 dominaria facilmente uma característica com valores que variam de 0 a 1. Assim, o valor normalizado da coordenada j do vetor de características x, denominado $n(x_j[n])$ é dado por:

$$n(x_j[n]) = \frac{x_j[n]}{\max_n(x_j[n]) - \min_n(x_j[n])} \tag{4.2}$$

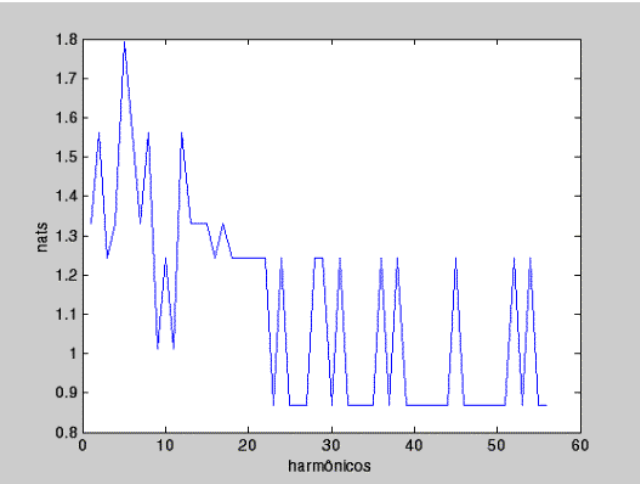

Figura 4.1: Entropia dos coeficientes da FFT.

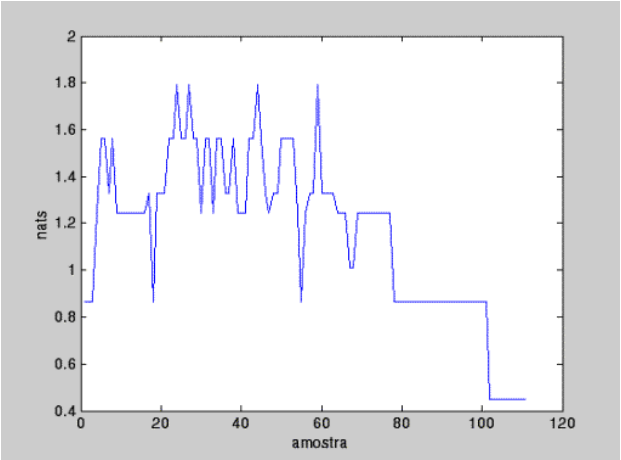

Figura 4.2: Entropia das amostras de PR (domínio do tempo).

No caso do uso de classificadores estatísticos otimizados pelo uso da entropia ou da distância de Battacharyya, aplica-se o vetor de entropia, $h = H(x)$ ou de distâncias B, para escalar as coordenadas do vetor de características x. Assim, considerando-se o uso da entropia, características com maior entropia (i.e. mais relevantes) têm maior peso no cálculo da distância entre um padrão e um protótipo. A métrica usada na medida da distância entre um dado de entrada $x = [x_1, x_2, ..., x_j]^T$ e um protótipo $\mu = [\mu_1, \mu_2, ..., \mu_j]^T$, para este classificador é:

$$\widetilde{r} = \sqrt{\sum_{J}(h_j(x_j - \mu_j))^2} \tag{4.3}$$

onde $j = 1, 2, ...J$ é o índice da coordenada do vetor de características.

O desempenho do classificador estatístico é otimizado com a introdução de pesos (i.e. medidas de entropia) para ponderar as coordenadas normalizadas do vetor de características. Desta forma, é possível implementar alterações nas escalas relativas entre as dimensões ou eixos do espaço de características. A Figuras 4.4 e 4.5 ilustram as características escaladas pela aplicação de pesos. Note que antes da aplicação dos pesos (vide Figura 4.4) o padrão x pertence μ_2. Supondo que a característica

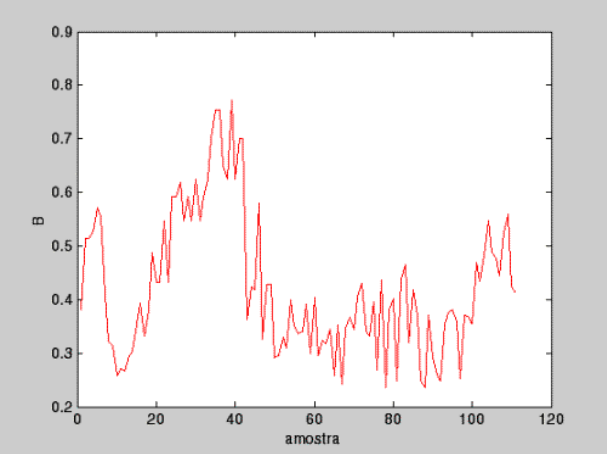

Figura 4.3: Distância de Battacharyya das características.

horizontal contém baixa entropia, por exemplo $h_1 = 0,5$ e na vertical $h_2 = 1,0$, o algoritmo iria escalar a característica horizontal por $0,5$ e a característica vertical por $1,0$. Neste caso, mudaria o desempenho do classificador e o padrão x pertenceria a μ_1.

Este trabalho também emprega a distância de Battacharyya para otimizar o desempenho do SOM. A idéia é acrescer precisão na determinação dos neurônios que melhor respondem aos padrões apresentados. Assim, a distância de Battacharyya, previamente calculada e ilustrada na Figura 4.3, é empregada para ponderar as coordenadas normalizadas do vetor de pesos sinápticos e do vetor de entrada. Esta abordagem permite alterações nas escalas relativas entre as dimensões do espaço de pesos no cálculo da menor distância euclidiana d_{ix} entre o vetor de entrada $x = [x_1, x_2, ..., x_J]^T$ e os vetores de pesos sinápticos dos neurônios $w_i = [w_{i1}, w_{i2}, ..., w_{iJ}]^T$, que neste trabalho é definida pela Equação 4.4:

$$d_{ix} = \sqrt{\sum_J (B_j(x_j - w_j))^2} \tag{4.4}$$

onde $j = 1, 2, ...J$ é o índice da coordenada do vetor de pesos sinápticos e B_j é a coordenada j do vetor de distâncias de Battacharyya.

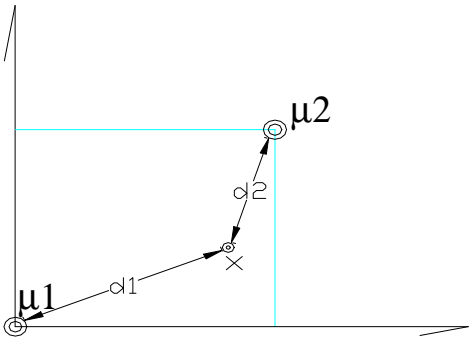

Figura 4.4: Padrão x classificado como pertencente a μ_2.

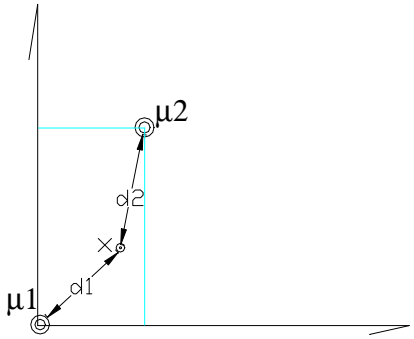

Figura 4.5: Características escaladas pela aplicação de pesos.

Os protótipos pré-definidos em [23] (i.e. valores de pressão no revestimento para seis padrões típicos de operação), são usados diretamente pelo classificador estatístico, ou seja, o classificador não é treinado por meio da Equação 2.1. Tais protótipos, caracterizados pelos seus PS, são recompostos no tempo por meio de interpolações

lineares, conforme ilustrado nas figuras 4.6 à 4.11. Em uma segunda etapa são extraidos 111 valores de pressão em intervalos de tempo regulares ao longo de um ciclo de operação. Estes valores compõem o vetor de características.

As figuras 4.12 à 4.17 ilustram os valores dos coeficientes da FFT do padrões típicos. Os valores dos coeficientes são complexos. Assim, está sendo exibido o módulo dos coeficientes.

No que tange ao emprego de técnicas estatísticas para RP, medidas de covariância podem ser usadas, em lugar das medidas de distância euclidiana, para aferir o grau de similaridade entre um padrão e um protótipo conforme será elucidado adiante.

O uso de RNAs *feedforward multilayer* para classificação de padrões pressupõe a existência de valores de saída alvo para o treinamento. Assim, foi necessário classificar visualmente um conjunto de treinamento. Esta classificação foi realizada pelo autor, que não é um especialista em técnicas de extração de petróleo. O critério adotado é a comparação com os padrões típicos encontrados em literatura específica e ilustrados pelas figuras 4.6 à 4.11 . Assim, a eficiência do classificador neural é, na hipótese mais otimista, igual a do autor.

4.3 Análise Estatística das Amostras

Este estudo faz uso do total de 1500 amostras de sinais de pressão, extraídas de três poços que operam com o GLI, para o treinamento e teste dos algoritmos empregados. As amostras se distribuem entre as classes conforme indicado na Tabela 4.1.

Tabela 4.1: Distribuição das amostras entre as classes.

Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
75,60%	5,13%	3,87%	7,67%	6,40%	1,33%

A Tabela 4.2 ilustra a dispersão dos padrões para cada classe, considerando-se como medida de dispersão a Equação 4.5, na qual N é o número total de amostras, x_n é o vetor representativo da amostra n e μ é o vetor média.

Tabela 4.2: Medidas de dispersão dos padrões por classe.

Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
654,00	538,29	543,69	565,15	465,05	643,12

$$s = \sqrt{\frac{\sum_{n=1}^{N} \|x_n - \mu\|^2}{N}} \tag{4.5}$$

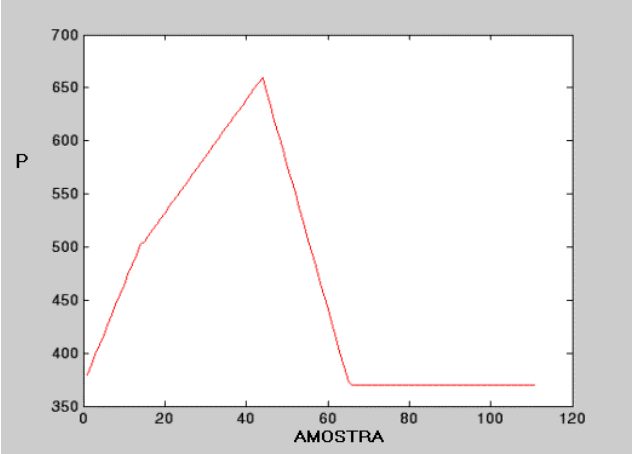

Figura 4.6: Protótipo Classe 01: "operação normal".

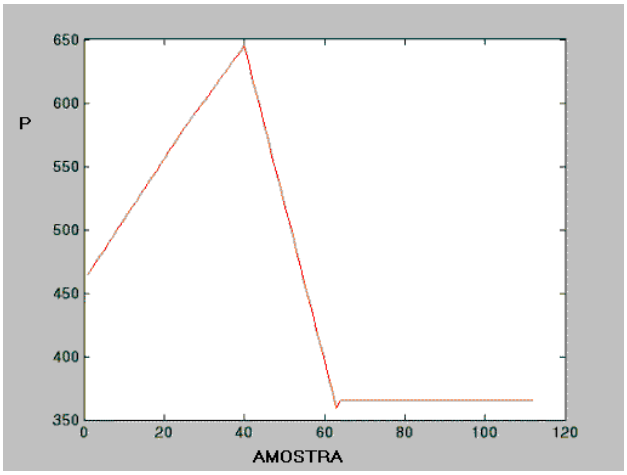

Figura 4.7: Protótipo Classe 02: "motor valve com vazamento".

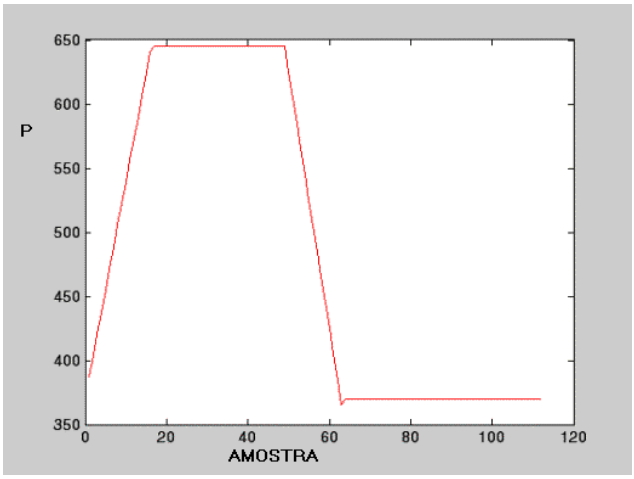

Figura 4.8: Protótipo Classe 03: "poço afogando".

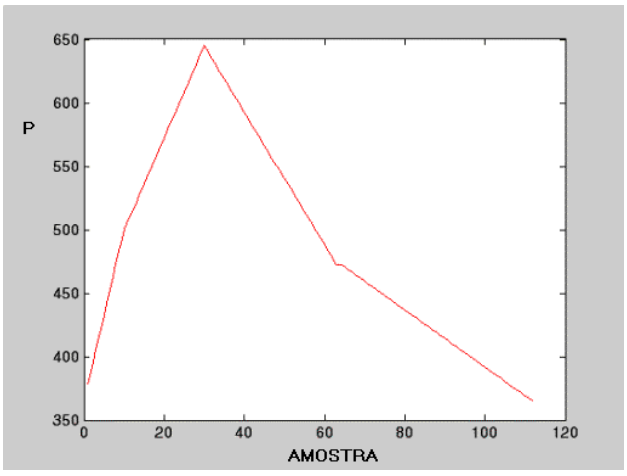

Figura 4.9: Protótipo Classe 04: "vazamento na válvula".

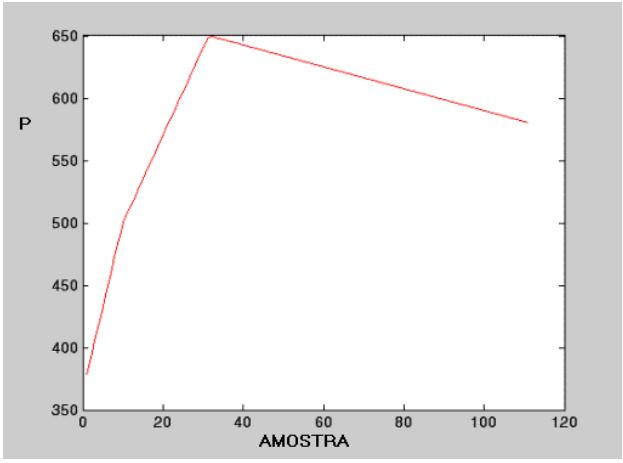

Figura 4.10: Protótipo Classe 05: "válvula operadora não abriu".

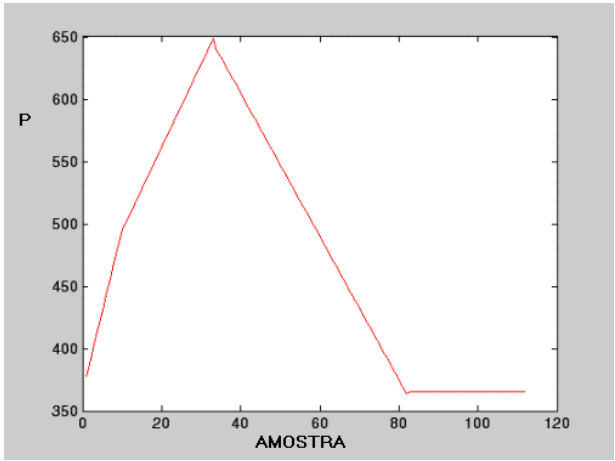

Figura 4.11: Protótipo Classe 06: "tubo parafinado".

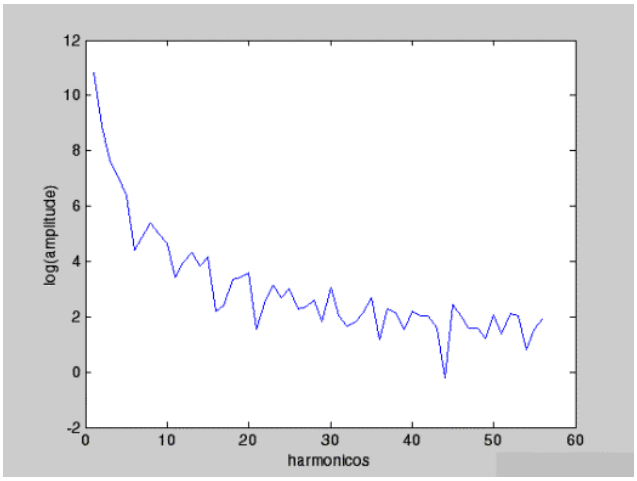

Figura 4.12: Espectro de freqüências típico "padrão normal".

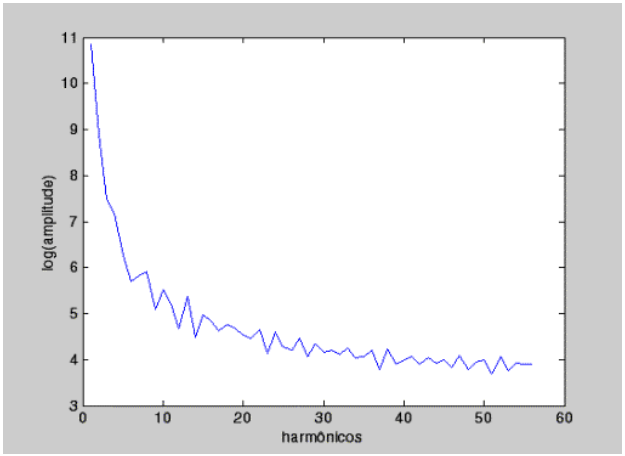

Figura 4.13: Espectro típico padrão "*motor valve* c/ vazamento".

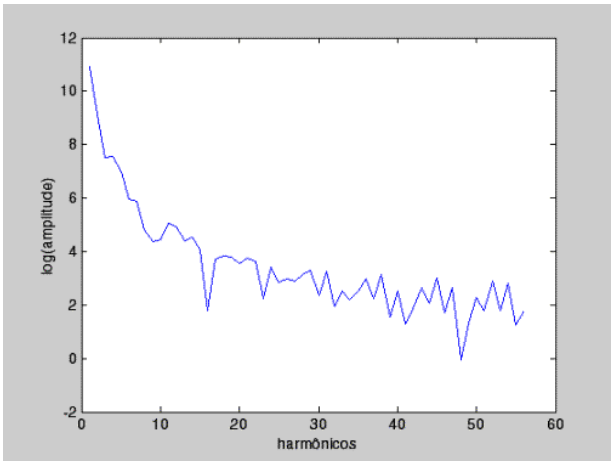

Figura 4.14: Espectro típico padrão "poço afogando".

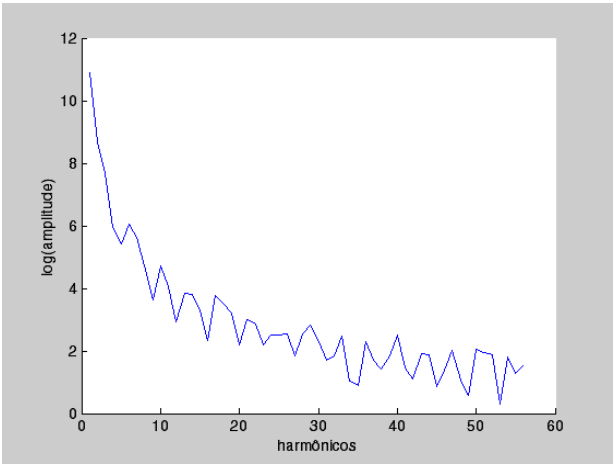

Figura 4.15: Espectro típico padrão "vazamento na válvula".

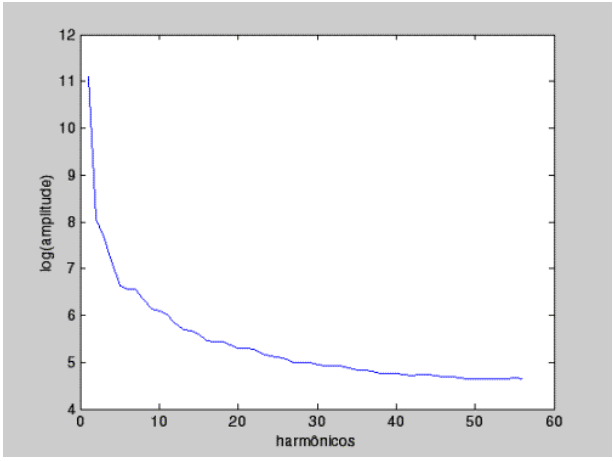

Figura 4.16: Espectro típico padrão "válvula operadora não abriu".

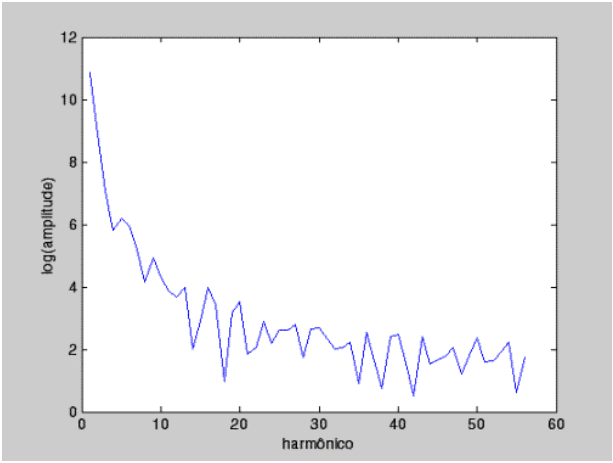

Figura 4.17: Espectro típico padrão "tubo parafinado".

Capítulo 5

Resultados Obtidos

5.1 Resultados Obtidos com o Uso de Classificadores Estatísticos

A avaliação de desempenho dos três métodos estatísticos propostos no Capítulo 4 toma como conjunto de teste 420 exemplos selecionados em três diferentes poços de extração. Cada exemplo consiste em dois vetores: o primeiro com os valores de tempo e o segundo com os respectivos valores de pressão no revestimento. Os intervalos de tempo não são regulares. Assim, o autor adota a interpolação linear dos pares ordenados (tempo, pressão) para a posterior extração de 111 amostras em intervalos de tempo regulares.

A metodologia adotada para a terefa de classificação consiste no casamento dos padrões de entrada com protótipos pre-definidos em literatura específica [28] e ilustrados nas figuras 4.6 à 4.11. Os protótipos são representados por meio de seus respectivos vetores de características, os quais contém 111 coordenadas com valores de pressão correspondentes a intervalos de tempo regulares ao longo de um ciclo de operação. O classificador adotado toma como referência para o casamento entre um

padrão e um protótipo a distância euclidiana entre os mesmos.

Os padrões são caracterizados por meio dos seus vetores de características. Tais características são extraídas no domínio do tempo e da freqüência. No domínio do tempo as coordenadas do vetor de características são 111 valores de pressão obtidos em intervalos regulares ao longo de um ciclo de operação.

Para obter o vetor de características no domínio da freqüência é aplicada a FFT às coordenadas obtidas no domínio do tempo. Assim, é gerado um vetor com 111 valores complexos de amplitude correspondentes aos 111 harmônicos da série. Entretanto, a Transformada Discreta de Fourier é uma série periódica discreta que perfaz dois períodos completos a cada período do sinal original. No caso em questão a série se repete a partir do harmônico $56\varpi_0$, onde ϖ_0 é a freqüencia fundamental. O autor adota um vetor de características com 56 coordenadas com os valores dos módulos das amplitudes dos primeiros 56 harmônicos. Os vetores de características extraídas no domínio da freqüência são comparados com os protópipos ilustrados nas figuras 4.12 à 4.17.

As coordenadas dos vetores de características são normalizadas conforme Equação 4.2, para que todas tenham a mesma relevância no cálculo da medida de distância entre um padrão e um protótipo.

Usando características extraídas no domínio do tempo, ou seja, valores de pressão no revestimento normalizados, o classificador estatístico demonstra desempenho insatisfatório, reconhecendo corretamente 66,3% dos padrões apresentados. Os resultados estão dispostos na Tabela 5.1, onde as linhas representam a saída do classificador (S.C.) e as columas representam a decisão correta (D.C.). Assim, por exemplo, a tabela indica que 7,5% do total de padrões de teste pertencentes à Classe 1 foram errôneamente classificados como pertencentes à Classe 2 e 85% dos padrões de teste pertencentes à Classe 2 foram classificados corretamente.

Tabela 5.1: Matriz de confusão do classificador estatístico.

D.C.\S.C.	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	62,0%	7,5%	0,0%	20,5%	0,0%	10,0%
Classe 2	9,0%	85,0%	0,0%	0,0%	0,0%	6,0%
Classe 3	0,0%	0,0%	76,5%	4,5%	6,0%	13,0%
Classe 4	12,5%	9,0%	0,0%	66,5%	0,0%	12,0%
Classe 5	0,0%	0,0%	12,0%	9,0%	79,0%	0,0%
Classe 6	23,0%	9,0%	0,0%	6,5%	0,0%	61,5%

As figuras 5.1 e 5.2 exemplificam a saída do classificador. Nestas figuras o protótipo em azul é contraposto ao padrão classificado em vermelho. Estas saídas apresentam alguns erros típicos cometidos pelo classificador estatístico empregado.

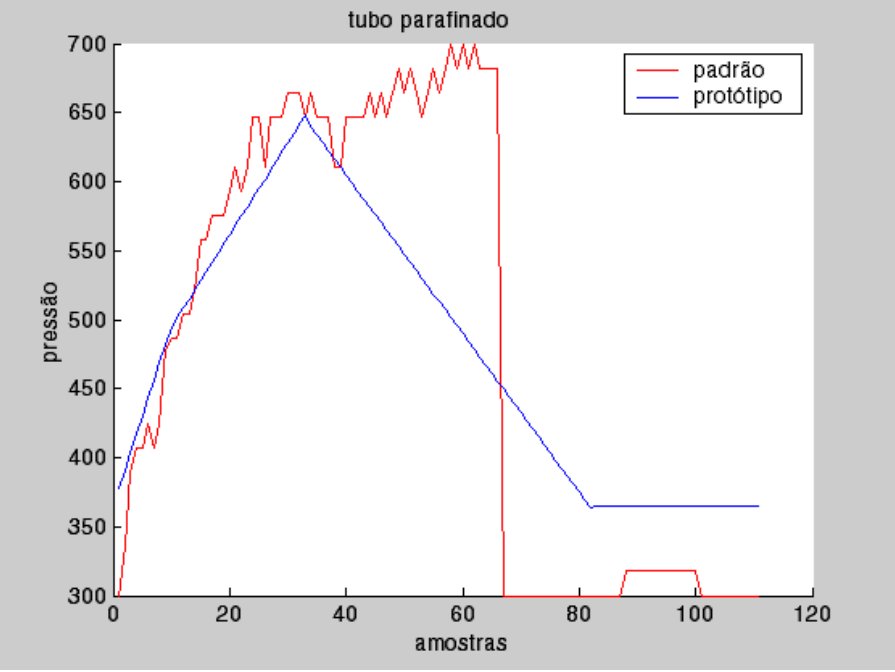

Figura 5.1: Padrão Poço afogando classificado erroneamente como Tubo parafinado.

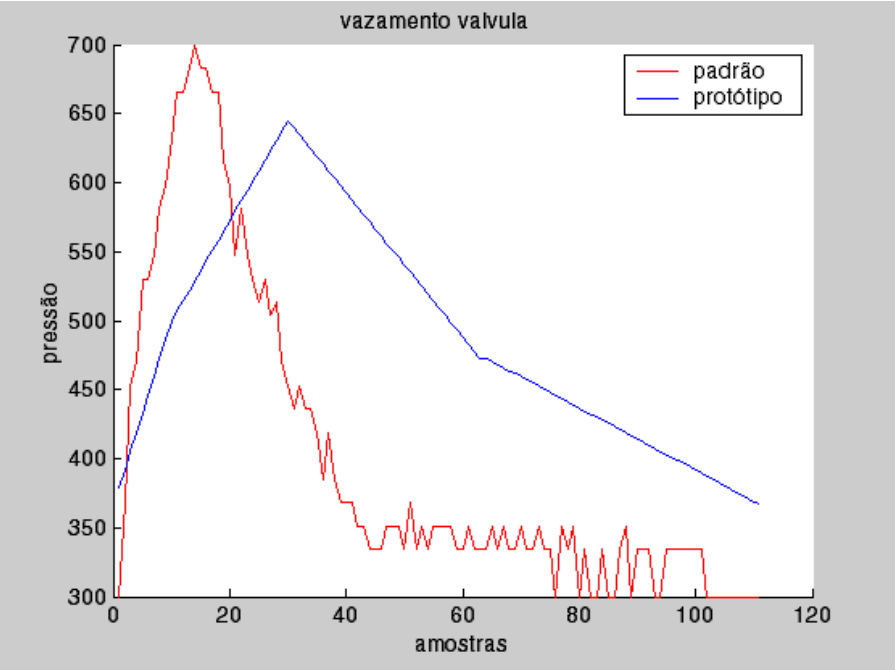

Figura 5.2: Padrão Operação normal classificado como Vazamento válvula

Em um segundo experimento, as coordenadas dos vetores de características são escaladas pela entropia. Assim, a distância euclidiana entre os padrões e os protótipos é aferida conforme Equação 4.3. O resultado revelou um acréscimo de desempenho ainda insuficiente, atingindo o reconhecimento correto de 70,2% do total de padrões apresentados. A Tabela 5.2 descreve os resultados.

Tabela 5.2: Matriz de confusão do classificador estatístico com o uso da entropia.

D.C.\S.C.	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	65,0%	7,0%	0,0%	16,0%	0,0%	12,0%
Classe 2	5,5%	90,0%	0,0%	0,0%	0,0%	4,5%
Classe 3	0,0%	0,0%	89,0%	0,0%	6,5%	4,5%
Classe 4	12,5%	9,0%	0,0%	66,5%	0,0%	12,0%
Classe 5	0,0%	0,0%	5,5%	6,5%	88,0%	0,0%
Classe 6	20,5%	9,0%	0,0%	5,5%	0,0%	65,0%

A Figura 5.3 exemplifica um caso no qual ocorre a classificação correta desta segunda abordagem, em contraposição ao que ocorre com o primeiro experimento (vide Figura 5.1). Entretanto, a dificuldade em separar adequadamente os padrões pertencentes às classes 1, 2 e 4 persiste. É importante ressaltar que, em alguns casos, o autor é incapaz de classificar com segurança estes padrões, o que implica em dúvidas quanto ao critério de avaliação da performance do algoritmo. A Figura 5.4 ilustra a dificuldade citada.

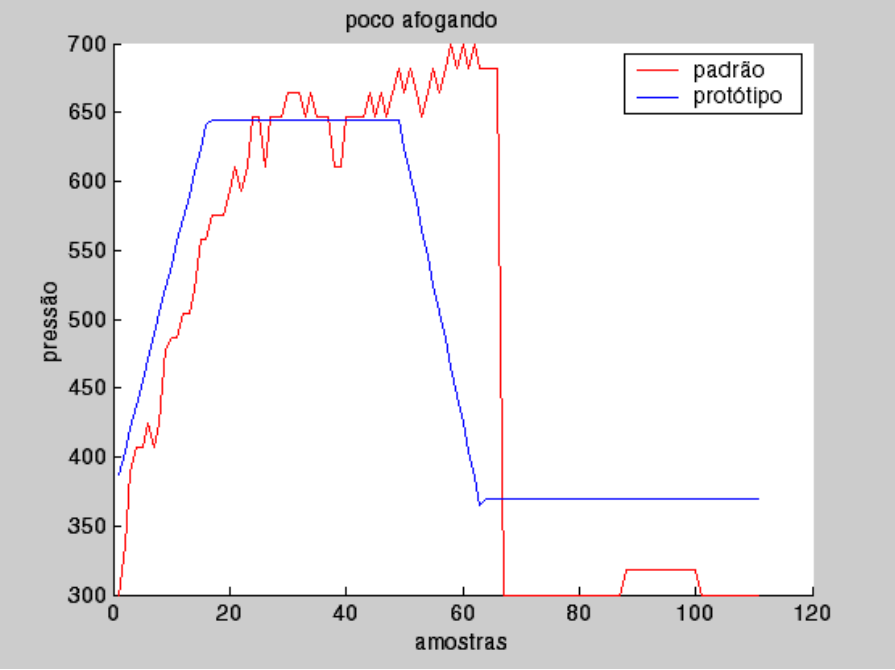

Figura 5.3: Padrão Poço afogando classificado corretamente.

É importante notar que o padrão "Tubo parafinado"difere do padrão "Vazamento válvula"devido à ausência de alteração da derivada dP/dt próximo à amostra 63, assim como, pela presença de um platô (i.e.dP/dt=0) a partir da amostra 80. Tal fato

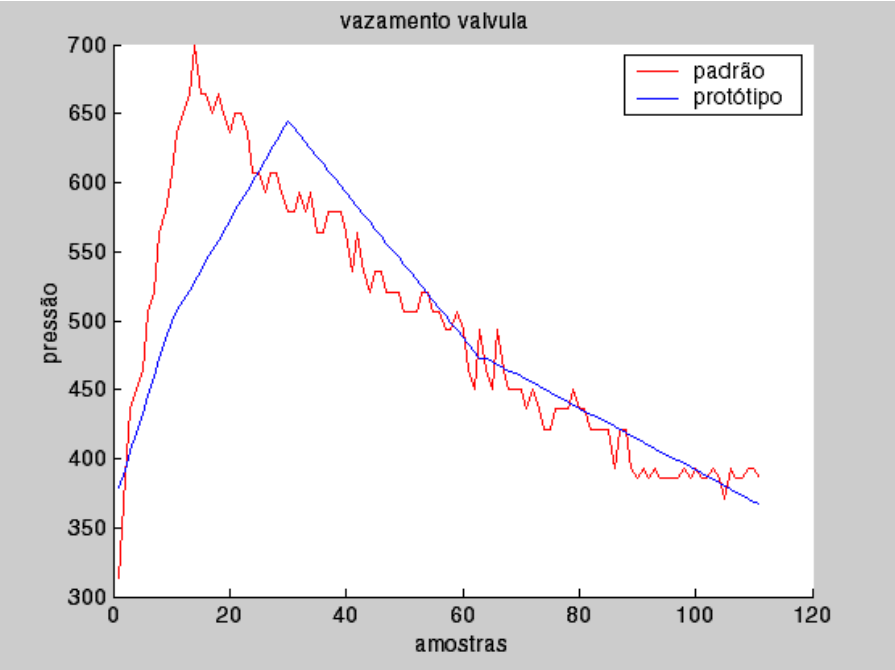

Figura 5.4: Padrão Tubo parafinado classificado como Vazamento na válvula

justifica a classificação "Tubo parafinado"adotada pelo autor para o padrão ilustrado na Figura 5.4.

O terceiro experimento do projeto consiste em extrair características no domínio da freqüência, por meio da FFT. Tais características, assim como as caracterrísticas extraídas dos protótipos, são normalizadas, escaladas pela entropia e apresentadas ao classificador estatístico para o cálculo das distâncias euclidianas entre os protótipos e o padrão. Este classificador obteve sucesso em 58% dos padrões apresentados. Os resultados são inferiores aos obtidos com as características extraídas no domínio do tempo. Entretanto, alguns problemas, a exemplo da falha ilustrada na Figura 5.4, foram resolvidos, conforme mostrado na Figura 5.5.

Medidas de covariância são usualmente adotadas para comparar sinais. Contudo, no caso em questão, o método não foi adequado. As discrepâncias ilustradas nas

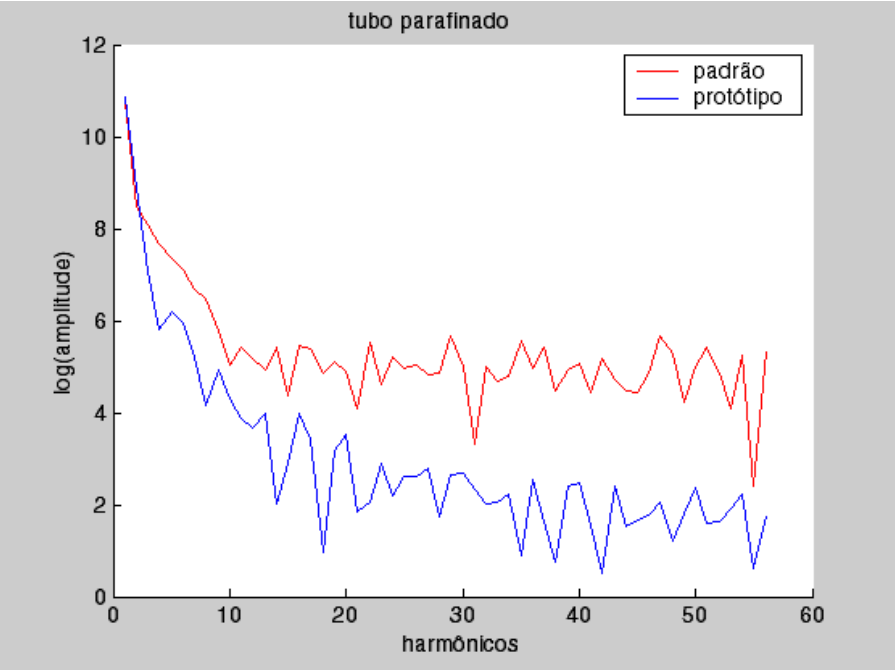

Figura 5.5: Espectro de padrão Tubo Parafinado classificado corretamente.

figuras 5.6 e 5.7 deixam claro a posição do autor.

5.2 Resultados Obtidos com o Uso de RNAs *Feed-forward Multilayer*

Diz-se que uma RNA tem boa capacidade de generalizar quando apresenta um mapeamento entrada-saída correto ou aproximadamente correto para dados não empregados na tarefa de treinamento, retirados da mesma população de onde foram extraídos os dados de treinamento [9].

Considerando-se um conjunto $\aleph$ de dados para treinamento da RNA, a capacidade de generalização desta rede é influenciada por três fatores: a cardinalidade do conjunto

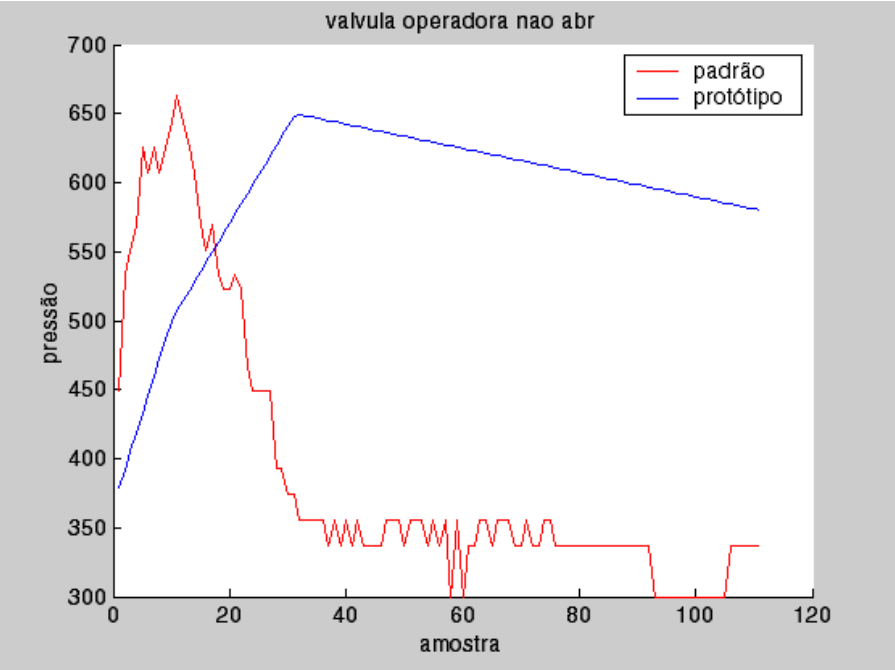

Figura 5.6: Padrão Motor valve c/ vazamento classificado como Válvula operadora não abriu.

de treinamento (i.e. $N = |\aleph|$), a arquitetura adotada para a RNA e a complexidade do problema a ser tratado. Este último, em alguns casos, é desconhecido ao projetista. Assim, resta uma relação entre os dois primeiros fatores.

Considerando a arquitetura da RNA fixa, o valor N necessário para o ajuste da RNA é formalmente caracterizado pela Equação 2.9 que é função da dimensão VC da mesma. Infelizmente, a prática demonstra que existem diferenças consideráveis entre os valores teóricos, resultantes do cálculo da dimensão VC, e os valores ótimos encontrados experimentalmente para N. Assim, o problema da complexidade da amostra é uma área de pesquisa em aberto.

Segundo Haykin [9], a regra empírica de Widrow [25] possibilita uma estimação

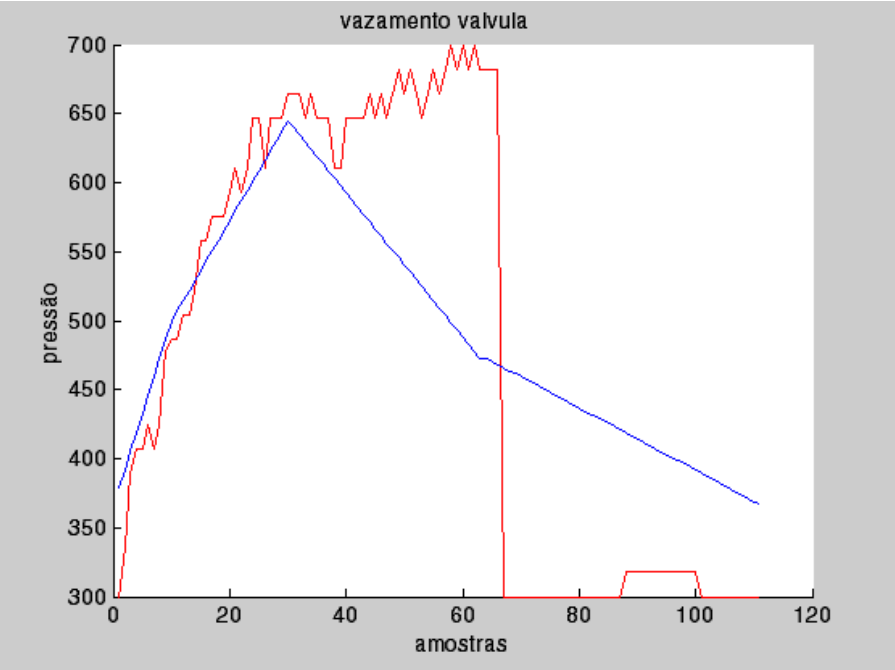

Figura 5.7: Padrão Poço afogando classificado como vazamento na válvula.

prévia mais próxima dos valores experimentais da cardinalidade do conjunto de treinamento. Esta regra é descrita conforme segue:

$$N \geq O\left(\frac{w}{\epsilon}\right) \tag{5.1}$$

onde w é o número de parâmetros livres, ϵ é o erro de classificação permitido.

Estas regras se baseiam em um limite de pior caso, independente da função de distribuição de probabilidades para as classes (i.e. da complexidade do problema a ser tratado).

Ainda que existam as supracitadas ferramentas para a determinação da relação $w(N)$, a determinação da arquitetura ótima para uma RNA *feedforward multilayer* exige um procedimento experimental. Assim, considerando-se o ajuste dos pesos e níveis de bias por meio do algoritmo de retropropagação de erro, tem-se três parâmetros a

serem ajustados no que tange à estrutura da RNA: o número de neurônios ocultos (i.e. o que implica em determinar o valor de w), o parâmetro taxa de aprendizagem η e a constante de momento α.

No presente trabalho o autor verifica experimentalmente que os valores $\eta = 0,05$ e $\alpha = 0,03$ se adequam ao problema específico em questão, ainda que a convergência seja lenta. No caso em questão, não é necessário o treinamento em tempo real, assim, velocidade não é fator decisivo. A principal questão é o adequado treinamento da RNA, visando uma boa capacidade de generalização.

Para a determinação experimental de w^* (i.e. a quantidade ótima de parâmetros livres) o autor adota quatro diferentes arquiteturas de RNAs, cujo número de parâmetros livres está relacionado na Tabela 5.3 .

Tabela 5.3: Arquitetura × parâmetros livres.

Arquitetura	w
$111 \times 10 \times 6$	6.676
$111 \times 15 \times 6$	10.011
$111 \times 20 \times 6$	13.346
$111 \times 25 \times 6$	16.681

Teoricamente, considerando um erro médio tolerável $\epsilon = 0,1$, a Equação 5.1 indica $N \geq 10^4$ para obter o desempenho esperado no caso da primeira arquitetura adotada (i.e. $111 \times 10 \times 6$). Considerando-se as demais arquitetura seriam necessários, ao menos, 10^5 exemplos de treinamento.

A obtenção, tratamento e pré-classificação de 10^5 dados para o treinamento das RNAs não é possível ao autor. Assim, este estudo adota um conjunto de 1000 dados pré-classificados pelo autor para o treinamento das RNAs. Cada dado de treinamento

consiste em um vetor contendo 111 valores de pressão obtidos em intervalos regulares durante um ciclo de operação do sistema GLI e a respectiva saída alvo.

O objetivo é alcançar desempenho satisfatório com a arquitetura mais simples possível. A Tabela 5.4 relaciona a arquitetura neural adotada com o número de épocas de treinamento e a taxa de acerto sobre um conjunto de teste composto de 500 exemplos:

Tabela 5.4: Arquitetura × épocas × tx. de acerto.

ARQUITETURA	N° DE ÉPOCAS	TX. DE ACERTO (%)
$111 \times 10 \times 6$	226	86,8
$111 \times 15 \times 6$	287	88,3
$111 \times 20 \times 6$	334	91,2
$111 \times 25 \times 6$	353	88,3

Os experimentos indicam o melhor desempenho da arquitetura $111 \times 20 \times 6$. A cardinalidade do conjunto de treinamento adotado é muito inferior ao exigido pela Equação 5.1. A situação se agrava se considerado o que é exigido pela Equação 2.9. O autor supõe que tal fato seja o responsável pela degradação do desempenho das arquiteturas com mais de 20 neurônios ocultos. Mais especificamente, o número de exemplos não é suficiente para ajustar o maior número de parâmetros livres de redes maiores, implicando em degradação da capacidade de generalização.
A Tabela 5.5 exibe os resultados obtidos com a arquitetura $11 \times 20 \times 6$.

O emprego de RNAs feedforward multilayer exige conjuntos de treinamento extensos, conforme indicam as equações 2.8 e 2.9. Assim, o número de parâmetros livres w deve ser limitado a valores que impliquem em conjuntos de treinamento acessíveis.

Tabela 5.5: Matriz de confusão da RNA feedforward multilayer.

D.C.\S.C.	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	91,0%	2,0%	0,0%	2,5%	0,0%	4,5%
Classe 2	2,0%	95,5%	0,0%	0,0%	0,0%	2,5%
Classe 3	0,0%	0,0%	95,5%	0,0%	4,5%	0,0%
Classe 4	4,5%	2,0%	0,0%	87,5%	0,0%	6,0%
Classe 5	0,0%	0,0%	2,5%	0,0%	97,5%	0,0%
Classe 6	6,0%	2,0%	0,0%	5,5%	0,0%	86,5%

Para reduzir o número de parâmetros livres w da RNA o autor aplica o método PCA, descrito no Capítulo 2, ao vetor de características x. Esta técnica possibilita a redução da dimensão do vetor de entrada x e a conseqüente redução do número de parâmetros livres da RNA.

O experimento consiste em calcular a matriz de covariância C das coordenadas do vetor x. A rotina que calcula a covariância tem como argumento o conjunto de treinamento.

Em segunda etapa, são determinadas a matriz diagonal de autovalores λ e a matriz de autovetores P. A matriz de covariância das coordenadas do vetor x na base P (i.e. $P^{-1}x$) é $P^{-1}CP = \lambda$. Esta matriz diagonal revela que a covariância entre as coordenadas de $P^{-1}x$ é nula. Assim, não há informação redundante nas coordenadas de $P^{-1}x$. Devido a este fato, um menor número de coordenadas é suficiente para conter grande parte da informação do vetor $P^{-1}x$.

A Figura 5.8 ilustra as medidas de variância das coordenadas do vetor de características antes e depois da mudança para base P. As 43 primeiras coordenadas do vetor $P^{-1}x$ exibem baixa variância, portanto, pouca informação. Assim, o vetor de entrada empregado no treinamento da RNA é constituído pelas 68 coordenadas finais do vetor $P^{-1}x$.

A arquitetura neural adotada no experimento com o método PCA é $68 \times 20 \times 6$. Após treinada, a RNA obtém taxa de acerto de $91,8\%$ sobre os dados do conjunto

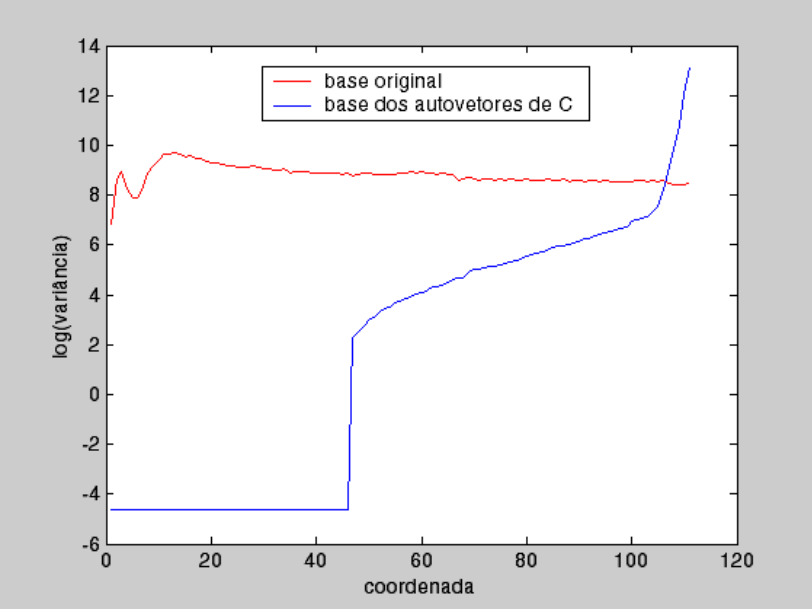

Figura 5.8: Variância das coordenadas do vetor de características antes e depois da mudança para base P.

de teste. O melhor desempenho desta configuração em relação à arquitetura $111 \times 20 \times 6$ exemplifica a importância do método PCA no emprego de RNAs feedforward multilayer.

5.3 Resultados Obtidos com o Uso do Mapa Auto-Organizável

Este experimento adota o uso do SOM associado a um mecanismo de pós-indexação dos neurônios. A idéia básica é aplicar um treinamento não supervisionado de forma a extrair características relevantes e então proceder um treinamento supervisionado.

O SOM é usado como um extrator de características devido as suas propriedades inerentes, conforme apresentado na Seção 2.7. As características dos dados de entrada

são representadas pelas coordenadas do neurônio vencedor. Assim, o SOM projetado deve ter um número maior de neurônios que o número de classes a serem reconhecidas. Tais neurônios são rotulados como subclasses das classes principais. No caso em questão existem 6 classes (i.e. estados de operação do sistema), assim, o SOM adotado tem 20 neurônios (i.e. 20 subclasses) dispostos em uma grade bidimensional com a relação 5×4. O processo pode ser entendido como uma compressão do espaço de entrada em um espaço de características ou espaço de subclasses.

Na fase do treinamento supervisionado, o vetor de saída do SOM (i.e. coordenadas do neurônio de vencedor) é classificado por um mecanismo que indexa o neurônio vencedor à saída alvo para o padrão apresentado.

Assim, cada classe é composta por um conjunto de subclasses, representadas pelos seus respectivos neurônios do SOM. As superfícies de decisão são composições de superfícies lineares (i.e. superfícies de decisão do SOM). A abordagem proposta possibilita superfícies de decisão mais complexas, conforme ilustrado nas figuras 5.9 e 5.10.

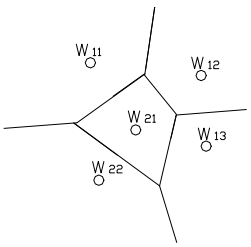

Figura 5.9: Fronteiras de decisão lineares do SOM.

Conforme descrito no Capítulo 4, a distância de Battacharyya é empregada para otimizar o desempenho do SOM neste experimento. Outra alternativa seria o emprego do nível de entropia, conforme descrito no trabalho [33].

Os dados empregados para o treinamento do SOM são os mesmos empregados no

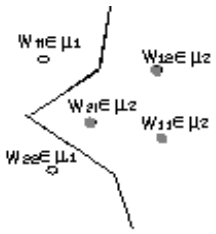

Figura 5.10: Fronteiras de decisão mais complexas do algoritmo adotado.

experimento com a RNA *feedforward multilayer*, ou seja, 1000 amostras extraídas no domínio do tempo.

O espaço de características tem 111 dimensões. Assim, o SOM em questão, por ser um classificador linear, tem *VCdim*=112, conforme indicado na Equação 2.8. Considerando-se a Equação 2.9 e, a título de exemplo, parâmetro de erro permitido $\varepsilon = 0.1$ e parâmetro de crença $\delta = 1$, o conjunto de treinamento necessário teria cardinalidade superior a $6,3 \cdot 10^4$. Tal fato indica uma possível melhoria no desempenho do SOM se adotado um conjunto de treinamento maior em trabalhos futuros.

O SOM é treinado por 100 épocas. Os parâmetros adotados são: τ_1 =10 , σ_0=0,8 e η_0=0,8.

A Figura 5.11 ilustra graficamente os valores das 111 coordenadas do vetor de pesos sinápticos de cada um dos 20 neurônios do SOM após o treinamento. A disposição dos gráficos mantém o ordenamento espacial dos neurônios na rede.

A segunda etapa do treinamento do SOM é a indexação dos neurônios às respectivas classes. Esta etapa equivale a um treinamento supervisionado no qual os padrões pertencentes ao conjunto de treinamento são apresentados ao SOM que responde ativando um único neurônio. O neurônio ativado recebe o rótulo da classe a qual o padrão pertence.

É possível que um mesmo neurônio responda a padrões pertencentes a classes

Tabela 5.6: Mapa da grade de neurônios com indexação dos neurônios.

1	1	1	1 ou 6
1 ou 2	1 ou 2	4	4
1	1	3	5
1	1	3	5
1	1		

diferentes conforme ocorre com a arquitetura adotada. A Tabela 5.6 ilustra a grade de neurônios do SOM após o treinamento. As células contêm o número da(s) classe(s) que os indexa. Nesta figura é possível perceber as propriedades do SOM, tais como, a ordenação topológica dos dados, ou seja, a localização espacial de um neurônio corresponde a um conjunto de características em especial dos dados de entrada. Assim, tem-se grupos de neurônios próximos representando uma única classe. Percebe-se ainda o casamento aproximado de densidade de probabilidade, ou seja, existem mais neurônios associados ao padrão 1 (i.e. padrão normal), visto que o mesmo ocorre com mais freqüência no conjunto de treinamento.

A questão da ambigüidade de rótulos para um mesmo neurônio pode ser resolvida por meio da adoção de uma arquitetura com maior número de neurônios. Provavelmente, a arquitetura adotada não é suficiente devido ao grande número de neurônios disponibilizados para representação da classe normal. Entretanto, simulações demonstram que uma arquitetura com 5×5 neurônios não tem êxito na separação das classes 1 e 2.

O autor adota outra solução: assume os neurônios (2,1) e (2,2) como representantes da classe 2, assim como, o neurônio (1,4) como representante da classe 6. Em seguida são aplicadas as equações 2.38 e 2.39 como forma de ajuste fino do SOM. A simulação adota α=0,03 e após cinco épocas de treinamento supervisionado as ambigüidades são eliminadas. A tabela 5.7 ilustra o resultado final.

Tabela 5.7: Mapa da grade de neurônios com indexação dos neurônios após treinamento supervisionado.

1	1	1	6
2	2	4	4
1	1	3	5
1	1	3	5
1	1		

O algoritmo ajustado teve taxa de acerto de 93,6% sobre o total de dados do conjunto de testes. A comparação com os resultados obtidos com o uso de RNAs *feedforward multilayer* (i.e. 91,8% com a aplicação do método PCA) revela o desempenho superior do SOM, ainda que a análise da dimensão VC indique uma maior capacidade de separação de padrões do primeiro classificador. A RNA *feedforward multilayer* empregada provavelmente tem seu desempenho degradado (i.e. pouca capacidade de generalizar) pelo tamanho insuficiente do conjunto de treinamento. Por outro lado, o conjunto de treinamento é mais adequado ao SOM que possui uma menor dimensão VC e, portanto, exige menor complexidade da amostra.

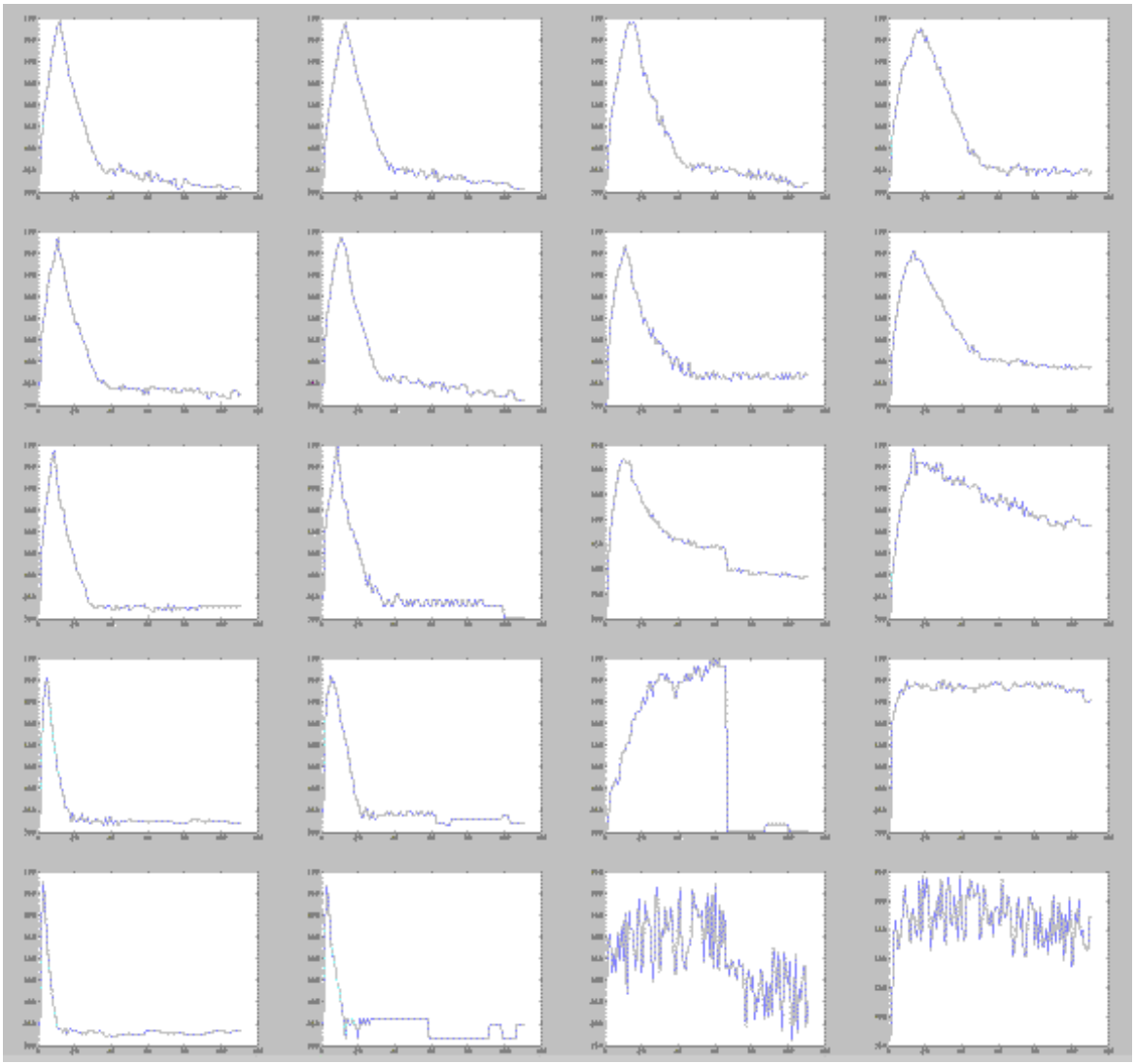

Figura 5.11: Visualização gráfica dos pesos sinápticos dos 4x5 neurônios do SOM após o treinamento.

Capítulo 6

Conclusões e Perspectivas Futuras

Os experimentos indicam a maior adequação dos mapas auto-organizáveis para a tarefa de reconhecimento de padrões implementada. Entretanto, o estudo da dimensão VC indica superfícies de decisão mais complexas para a RNA *feedforward multilayer* com arquitetura $111 \times 20 \times 6$. A Equação 2.9 indica a necessidade de um conjunto de treinamento maior para o ajuste da RNA *feedforward multilayer* se comparado ao exigido pelo SOM.

Um conjunto de dados pré-classificados com a cardinalidade exigida pela Equação 2.9 é indisponível na prática. Assim, uma possível solução é a redução da dimensão do vetor de entrada. Com uma menor dimensão na entrada a rede teria um menor número de parâmetros livres e, consequentemente, exigiria um conjunto de treinamento menor.

O autor adota neste trabalho o método PCA para a redução da dimensão do vetor de entrada. Entretanto, existem outras técnicas aplicáveis ao mesmo problema, dentre as quais é possível citar a extração de características por meio da análise da entropia ou pela análise da distância de Battacharyya. A idéia é truncar o vetor de entrada, excluindo coordenadas com baixa entropia ou pequena distância B. Em [26, 27] o autor sugere uma forma de extração de características baseada em algoritmos

genéticos.

Em futuros trabalhos é recomendável a pré-classificação dos exemplos de treinamento por um engenheiro especialista, bem como, a adoção de um maior conjunto de treinamento. A redução da dimensão do vetor de entrada é indispensável ao uso de RNAs *multilayer feedforward*.

No que tange ao estudo em questão, o autor considera a distância de Battacharyya, empregada na abordagem com o SOM, a melhor referência para refutar características irrelevantes. A justificativa se respalda na presença de ruídos que comprometem a fidelidade das medidas de entropia. Com respeito à aplicação de algoritmos genéticos, o problema a ser contornado é um espaço de busca de grande dimensão (i.e. 111 dimensões). Tal fato compromete a convergência do algoritmo.

Quanto a extração de características no domínio da freqüência, o método não se mostrou eficaz quando dos experimentos com classificadores estatísticos. Tal fato justifica a posição do autor, ou seja, o abandono desta metodologia nas abordagens conexionistas.

No que se refere ao atual estado da arte, este trabalho contribui otimizando o desempenho das Redes de Kohonen por meio do emprego da distância de Battacharyya na determinação do neurônio a ser ativado em resposta a um sinal de entrada. É também contribuição deste trabalho o uso do nível de entropia no processo de extração de características, conforme descrito em [24].

Os resultados aqui obtidos são passíveis de comparação com os resultados oriundos do Projeto SGPA II, em desenvolvimento pelo Departamento de Engenharia Mecânica da UFBA, com a finalidade de complementar e validar mutuamente os dois estudos, propiciando uma significativa referência de desempenho.

O autor acredita no aproveitamento deste trabalho na solução de problemas de engenharia da industria petroquímica, consolidando a relação já estabelecida entre a indústria e a academia.

Bibliografia

[1] Kohonen, T.; Ritter, H., "*Biological Cybernetics*", 61, 241-254, Elsevier, Amsterdam, 1989.

[2] Lippmann, R.P., "*Pattern classification using neural networks*", IEEE Communications Magazine, vol.27, 1989.

[3] Ritter, H., K. Obermayer, K. Schulten., "*Development and spatial structure of cortical feature maps: A model study*", Advances in Neural Information Processing Systems, vol.3, 1991.

[4] von der Malsburg, "*Network self-organization*", San Diego, CA: Academic Press, 1990.

[5] Kohonen, T., "*Exploration of very large databases by self-organizing maps*", International Conference on Neural Networks, vol I, 1997.

[6] Kohonen, T., "*Physiological interpretation of the self-organizing map algorithm*", Neural Networs, vol.6, 1993.

[7] Kohonen, T., "*Self organizing maps*", Berlin: Springer-Verlag, 1997.

[8] Haykin, S. et al., "*Classification of radar clutter in air traffic control environment*", Proceedings of the IEEE, vol.79, 1991.

[9] Haykin, S., "*Redes Neurais, princípios e prática*", 2^a edição, Prentice Hall, 1999.

[10] Parbhane RV, Tambe SS, Kulkarni BD., "*ANN modelling of DNA sequences: new strategies using DNA shape code*", Comput Chem 2000; 24: 699-711.

[11] Srihari, S.N., "*High-Performance Reading Machines*", Proceedings of the IEEE, 80(7), July 1992, 1120-1132.

[12] Yang, C. C., Marefat, M. M. and R. L. Kashyap, "*Automated Visual Inspecion Based on CAD Models*", Proceedings of IEEE International Conference on Robotics and Automation, San Diego, CA, May 8-13, 1994.

[13] Kaufmann, M., "*Information Visualization in Data Mining and Knowledge Discovery"*, 2001.

[14] Carreira-Perpinan, M. A., "*Compression neural networks for feature extraction: Application to human recognition from ear images*", MSc thesis, Faculty of Informatics, Technical University of Madrid, Spain, 1995.

[15] Kim, S.S.; Lee, D.J.; Kwak, K.C.; Park, J.H. and Ryu, J.W., "*Speech recognition using integra-normalizer and neuro-fuzzy method*", IEEE Conf. on Signals, Systems and Computers, v 2, p 1498-1501, 2000.

[16] Kohonen T., Ritter, H., "*Biological Cybernetics*", 61, 241-254, Elsevier, Amsterdam,1989.

[17] LeCun, Y. and Y. Bengio, "*Convolutional Networks for Images, Speech and Time Series*", MIT Press, Cambridge, 1995.

[18] Devijver P. A. and J. Kittler, "*Pattern Recognition: A Statistical Approach*", Prentice-Hall International, Englewood Cliffs, NJ, 1980.

[19] Stefik, M., "*Introduction to Knowledge Systems*", Morgan Kaufmann, San Francisco, CA, 1995.

[20] Fukunaga, K., "*Introduction to Statistical Pattern Recognition*", 2nd Ed., Academic Press, New York, 1990.

[21] Koiran, P., Sontag, E. D., "*Neural Networks with quadratic VC dimension*", Journal of Computer and System Sciences, 54(1):190

[22] Alegre, L., da Rocha, A. F. e Morroka, C. K.(1993a), "*Intelligent approach of rod pumping problems*", (SPE 26253): 249-255.

[23] Alegre, L., da Rocha, A. F. e Morroka, C. K. (1993b), "*Intelligent approach of rod pumping problems*", (SPE 26516): 97-107.

[24] Ludwig, O., Schnitman L., Souza J.A.M.F., Lepikson H., "*A Comparative Analysis Between Conventional Approaches and Connectionist Methods in Pattern Recognition Tasks*", Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, pp 639-644, Innsbruck, Austria, Feb 2004.

[25] Widrow, B., Steams, S.D., "*Adaptative Signal Processing*", Englewood Cliffs, NJ: Prentice Hall, 1985.

[26] Ludwig, O., Schnitman L., Lepikson H., "*Algoritmo Genético para o Ajuste de Classificadores Estatísticos*"I STEC, 2003.

[27] Ludwig O., Castro Lima A. C., Schnitman L., Souza J.A.M.F., "*Supervised Methods for Feature Extraction*", CONTROLO 2004, Algarve, Portugal.

[28] Santos, O. G., "*Sistema Especialista Para Diagnostico do Gas Lift Intermitente*", Campinas: Departamento de Engenharia de Petróleo da Universidade Estadual de Campinas, jul. 1995.

[29] Correa, J. F., Santos, O. G., Inazumi, P. C. M., "*Intelligent Automation for Intermittent-Gas-Lift Oil Wells*", SPE, 25-28 mar. 2001.

[30] Cerqueira, J. et al., "*Sistema de Gerenciamento de Poços de Petróleo Automatizados: SPGA*", In: XIV CONGRESSO BRASILEIRO DE AUTOMÁTICA, 2-5, 2002, Setembro, Rio Grande do Norte, Natal.

[31] Kailath T., "*The Divergence and the Battacharyya Distance Measures in Signal Selection*", IEEE Trans. Commun. Theory, COM 5, pp.52-60, 1967.

[32] Vidyasagar, M., "*A Theory of Learning and Generalization*" , Springer-Verlag, 1997.

[33] Ludwig O. Júnior; Schnitman L.; Lepikson H.; Castro Lima A. C., "*Redes de Kohonen para o Reconhecimento de Imagens*", In: XV CONGRESSO BRASILEIRO DE AUTOMÁTICA, 2004, Gramado, Rio Grande do Sul.