

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

LUIZ FERNANDO GONÇALVES

**DESENVOLVIMENTO DE UM SISTEMA DE
MANUTENÇÃO INTELIGENTE EMBARCADO**

Porto Alegre

2011

LUIZ FERNANDO GONÇALVES

**DESENVOLVIMENTO DE UM SISTEMA DE
MANUTENÇÃO INTELIGENTE EMBARCADO**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: microeletrônica.

ORIENTADOR: Prof. Dr. Marcelo Soares
Lubaszewski

Porto Alegre
2011

LUIZ FERNANDO GONÇALVES

**DESENVOLVIMENTO DE UM SISTEMA DE
MANUTENÇÃO INTELIGENTE EMBARCADO**

Esta dissertação foi julgada adequada para a obtenção do título de Doutor em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: _____

Prof. Dr. Marcelo Soares Lubaszewski, UFRGS

Doutor pelo Instituto Politécnico Nacional - Grenoble, França

Banca Examinadora:

Prof. Dr. Altamiro Amadeu Susin, UFRGS

Doutor pelo Instituto Politécnico Nacional - Grenoble, França

Prof. Dr. Carlos Eduardo Pereira, UFRGS

Doutor pela Universidade de Stuttgart - Stuttgart, Alemanha

Prof. Dr. Gilson Inácio Wirth, UFRGS

Doutor pela Universidade de Dortmund - Dortmund, Alemanha

Prof. Dr. João Maurício Rosário, UNICAMP

Doutor pela Escola Central de Artes e Manufaturas - Paris, França

Prof. Dr. Roberto Kawakami Harrop Galvão - ITA

Doutor pelo Instituto Tecnológico de Aeronáutica - São José dos Campos, Brasil

Coordenador do PPGEE: _____

Prof. Dr. Alexandre Sanfelice Bazanella

Porto Alegre, abril de 2011

“O valor do próprio saber está no fato de poder compartilhá-lo.”

Provérbio Chinês.

AGRADECIMENTOS

Gostaria de agradecer, inicialmente, a orientação acadêmica prestada pelo Prof. Dr. Marcelo Soares Lubaszewski, ao longo destes anos, para a realização deste trabalho. O professor Marcelo, desde o início, sempre esteve à disposição, mostrando-se interessado e empenhado em relação às atividades pertinentes ao doutorado. Com o passar do tempo, presenteou-me com sua amizade, sabedoria e respeito.

Não poderia deixar de agradecer o Prof. Dr. Altamiro Amadeu Susin e o Prof. Dr. Gilson Inácio Wirth pelas suas contribuições prestadas para a melhoria do presente texto, pelo coleguismo e amizade. Agradeço também a todos os colegas do Laboratório de Processamento de Sinais e Imagens, em especial, a Prof. Dra. Leticia Vieira Guimarães, o Dr. Marcelo Negreiros, o Dr. André Borin Soares e o M.Sc. Alessandro Cristovão Bonatto as inúmeras dicas e sugestões fornecidas para realizar e melhorar este trabalho.

Sou muito grato também a todos os amigos que fiz no Laboratório de Prototipação e Teste: o Prof. Dr. Tiago Roberto Balen, o M.Sc. Franco Ripoll Leite, o M.Sc. Marcos Barcelos Hervé, o M.Sc. Lucas Brusamarello, o Engenheiro Gabriel de Moraes Borges e os bolsistas de iniciação científica Thiago Hanna Both, Bruno Canal, Otto Aureliano Roloff e Gabriel Cunha Marimon.

Meus sinceros agradecimentos a todos os colegas do Instituto de Informática, do Programa de Pós-Graduação em Computação e do Programa de Pós-Graduação e Microeletrônica e, em especial, ao amigo M.Sc. Jefferson Luiz Bosa que muito ajudou na elaboração desta tese, fornecendo inúmeras dicas, sugestões e contribuições.

Agradeço ao Prof. Dr. Renato Ventura Bayan Henriques, pela sua colaboração, e aos demais professores do Programa de Pós-Graduação em Engenharia Elétrica; em especial, o Prof. Dr. Carlos Eduardo Pereira.

Além desses, gostaria de ressaltar que a realização deste trabalho foi possível graças à atenção e a dedicação apresentada pelo pessoal da empresa Coester Automação S.A., da Petrobrás S.A e do *Industry and University Cooperative Research Center on Intelligent Maintenance Systems* em diversos momentos.

Agradeço, por fim, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, a todos os colegas de Pós-Graduação e a todos os professores e funcionários da Universidade Federal do Rio Grande do Sul e, em particular, do Departamento de Engenharia Elétrica.

RESUMO

A evolução tecnológica dos sensores, da eletrônica e dos sistemas embarcados melhorou o desempenho, a confiabilidade e a robustez dos sistemas assim como as atividades de manutenção, em especial, as de manutenção proativa. Estes avanços tecnológicos possibilitaram uma nova visão sobre as práticas de manutenção existentes. A expansão das áreas de processamento de sinais e inteligência artificial proporcionou novas abordagens aos sistemas de controle, promovendo a criação de novos modelos de confiabilidade e disponibilidade de equipamentos e sistemas. Além disso, aumentou a precisão no reconhecimento de padrões de falhas, ampliou a avaliação e o diagnóstico de danos em equipamentos e sistemas, e adicionou inteligência aos sistemas de manutenção existentes. Diversas técnicas de processamento de sinais (tais como a transformada de Fourier), de inteligência artificial (as redes neurais artificiais e a lógica nebulosa, por exemplo) e de filtragem adaptativa (os filtros adaptativos, como exemplo) já são utilizadas com sucesso para detectar e prevenir falhas em vários tipos de equipamentos. Os sistemas de manutenção que fazem uso das técnicas de processamento de sinais e inteligência artificial, em conjunto, por exemplo, são conhecidos como sistemas de manutenção inteligente. Através desses sistemas, é possível monitorar as condições físicas, tomar decisões, efetuar ações de manutenção e fornecer diagnósticos precisos de falhas. Este trabalho aborda a implementação de um sistema de manutenção inteligente embarcado que usa a transformada *wavelet packet* e os mapas auto-organizáveis ou os filtros adaptativos para detectar, classificar e prever falhas em atuadores elétricos. A idéia principal deste trabalho é determinar qual destas ferramentas, mapas auto-organizáveis ou filtros adaptativos, é a mais adequada para o embarque. Espera-se com a implantação embarcada desse sistema de manutenção, por exemplo, evitar falhas nos atuadores e promover uma maior reutilização de peças.

Palavras-chave: manutenção, processamento de sinais, inteligência artificial, filtragem adaptativa, sistemas embarcados, FPGA.

ABSTRACT

The technological evolution of sensors, electronics, and embedded systems has improved the performance, reliability and robustness of systems as well the maintenance activities, especially the proactive maintenance. These technological advances have provided a new view about the existing maintenance practices. The expansion of signal processing and artificial intelligence has provided new approaches in industrial control systems leading to the proposal of new reliability and availability models for equipments and systems. Moreover, it has increased the precision in failure pattern recognition, has extended the assessment and diagnosis of damages in equipments and systems, and has added intelligence to existing maintenance systems. Several techniques for signal processing (such as Fourier transform), artificial intelligence (artificial neural networks, for example) and adaptive filtering (adaptive filters, as an example) are already used successfully to detect and prevent failures in several kinds of equipments. The maintenance systems that use, for example, the techniques for signal processing and artificial intelligence together are known as intelligent maintenance systems. It is possible to control the physical conditions, make decisions, perform maintenance activities and do accurate diagnosis of failures using those systems. This work presents the implementation of an embedded intelligent maintenance system using wavelet packet analysis and self organizing maps or adaptive filters for detection, classification, and prediction of failures in electrical actuators. The main idea is to determine which of these tools, self-organizing maps or adaptive filters, is the most suitable for the implementation in embedded systems. It is expected that with the implementation of this maintenance system, failures in actuators are avoided, and that a greater reuse of parts is achieved.

Keywords: maintenance, signal processing, artificial intelligence, adaptive filtering, embedded systems, FPGA.

SUMÁRIO

LISTA DE FIGURAS	12
LISTA DE TABELAS	15
LISTA DE ABREVIATURAS E SIGLAS	16
LISTA DE SÍMBOLOS	18
1 INTRODUÇÃO	24
1.1 OPERAÇÃO	30
1.2 MOTIVAÇÃO	31
1.3 OBJETIVOS	32
1.4 METODOLOGIA	32
1.5 CONTRIBUIÇÕES	33
1.6 ESTRUTURA	33
2 MANUTENÇÃO	34
2.1 TERMOS EMPREGADOS NA MANUTENÇÃO	34
2.1.1 Degradação	34
2.1.2 Defeito	35
2.1.3 Falha	35
2.1.4 Confiabilidade	35
2.1.5 Disponibilidade	36
2.1.6 Tempo Médio Entre Falhas	36
2.1.7 Tempo Médio Para Reparo	36
2.2 DEFINIÇÃO DE MANUTENÇÃO	36
2.3 ESTRATÉGIAS DE MANUTENÇÃO	38
2.3.1 Manutenção Corretiva	41
2.3.2 Manutenção Preventiva	42
2.3.3 Manutenção Preditiva	43
2.3.4 Manutenção Proativa	45
2.4 ESTADO DA ARTE NA ÁREA DE MANUTENÇÃO	46
2.4.1 Sistemas de Manutenção Inteligentes	48
2.4.2 Sistemas de Manutenção Inteligentes Embarcados	50
2.4.3 Metodologia de Desenvolvimento de Sistemas de Manutenção Inteligentes	51
2.5 CONCLUSÕES DO CAPÍTULO	53
3 DESENVOLVIMENTO DO MODELO DO ATUADOR	54
3.1 ESTUDO DE CASO	55
3.1.1 Operação do Atuador Elétrico	56
3.1.2 Forças Envolvidas na Operação do Atuador Elétrico	57
3.2 MODELO DO ESTUDO DE CASO	57
3.2.1 Definição do Modelo do Estudo de Caso	58
3.2.2 Modelo da Máquina Assíncrona	58

3.2.3 Modelo do Conjunto de Engrenagens	63
3.2.4 Modelo da Válvula e Tubulação	66
3.2.5 Conjunto de Equações Algébricas e Diferenciais do Modelo	68
3.2.6 Solução do Modelo do Atuador Elétrico	70
3.3 INJEÇÃO DE FALHAS NO MODELO	72
3.4 CONCLUSÕES DO CAPÍTULO	76
4 ANÁLISE DAS FERRAMENTAS MATEMÁTICAS	77
4.1 ANÁLISE DAS FERRAMENTAS DE PROCESSAMENTO DE SINAIS	77
4.2 ANÁLISE DA ENERGIA DOS SINAIS	81
4.3 ANÁLISE DOS MAPAS AUTO-ORGANIZÁVEIS	82
4.3.1 Inicialização dos Mapas Auto-organizáveis	83
4.3.2 Treinamento dos Mapas Auto-organizáveis	83
4.3.3 Teste dos Mapas Auto-organizáveis	85
4.4 ANÁLISE DOS FILTROS ADAPTATIVOS	92
4.4.1 Inicialização dos Filtros Adaptativos	94
4.4.2 Treinamento dos Filtros Adaptativos	94
4.4.3 Teste dos Filtros Adaptativos	95
4.5 DETERMINAÇÃO DO TEMPO DE VIDA ÚTIL DO ATUADOR	102
4.5.1 Tempo de Vida Útil Obtido Através dos Mapas Temporais de Kohonen	104
4.5.2 Tempo de Vida Útil Obtido Através dos Filtros Adaptativos	108
4.5.3 Comparação dos Resultados de Predição	113
4.6 CONCLUSÕES DO CAPÍTULO	113
5 DESENVOLVIMENTO DO SISTEMA EMBARCADO	115
5.1 EMBARQUE DO SISTEMA DE MANUTENÇÃO INTELIGENTE	116
5.1.1 Embarque dos Mapas Auto-organizáveis	117
5.1.2 Embarque dos Filtros Adaptativos	127
5.2 PLATAFORMA DE EMBARQUE	134
5.2.1 Especificação dos Mapas Auto-organizáveis para Embarque	137
5.2.2 Especificação dos Filtros Adaptativos para Embarque	137
5.3 RESULTADOS DE EMBARQUE	138
5.3.1 Área do Circuitos	138
5.3.2 Desempenho dos Circuitos	139
5.3.3 Tempo de Processamento dos Circuitos	139
5.4 CONCLUSÕES DO CAPÍTULO	141
6 CONCLUSÃO	142
7 ANEXO A: PROCESSAMENTO DE SINAIS	147
7.1 TÉCNICAS DE PROCESSAMENTO DE SINAIS	147
7.2 TRANSFORMADA DE FOURIER	148
7.2.1 Tipos de Análises de Fourier	149
7.2.2 Transformada Contínua de Fourier	149
7.2.3 Transformada de Fourier de Tempo Discreto	150
7.2.4 Transformada Discreta de Fourier	150
7.2.5 Vantagens da Transformada de Fourier	151
7.2.6 Desvantagens da Transformada de Fourier	151
7.3 TRANSFORMADA DE FOURIER COM JANELAMENTO	152
7.3.1 Transformada Contínua de Fourier com Janelamento	152
7.3.2 Transformada Discreta de Fourier com Janelamento	153

7.3.3 Utilidade da Transformada de Fourier com Janelamento	154
7.3.4 Desvantagens da Transformada de Fourier com Janelamento	154
7.4 TRANSFORMADA WAVELET	156
7.4.1 Transformada <i>Wavelet</i> Contínua	157
7.4.2 Transformada <i>Wavelet</i> Discreta	158
7.4.3 Benefícios da Transformada <i>Wavelet</i>	162
7.4.4 Desvantagens da Transformada <i>Wavelet</i>	162
7.5 TRANSFORMADA <i>Wavelet Packet</i>	163
7.5.1 Transformada <i>Wavelet Packet</i> Discreta	163
7.5.2 Situações de Aplicação da Transformada <i>Wavelet Packet</i>	166
7.6 COMPARAÇÃO DAS FERRAMENTAS DE PROCESSAMENTO DE SINAIS	166
7.7 ENERGIA DE SINAIS CONTÍNUOS E DISCRETOS	167
7.8 APLICAÇÃO DE PROCESSAMENTO DE SINAIS	168
7.9 CONCLUSÕES DO CAPÍTULO	169
8 ANEXO B: INTELIGÊNCIA ARTIFICIAL	170
8.1 PRINCIPAIS CARACTERÍSTICAS	170
8.2 TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL	171
8.3 REDES NEURAIS ARTIFICIAIS	174
8.3.1 Propriedades das Redes Neurais Artificiais	174
8.3.2 Modelo de Neurônio Artificial	176
8.3.3 Tipos de Redes Neurais Artificiais	176
8.3.4 Arquiteturas das Redes Neurais Artificiais	177
8.3.5 Paradigmas de Aprendizagem	178
8.3.6 Redes de Funções de Base Radial	182
8.3.7 Redes <i>Perceptron</i> de Múltiplas Camadas	183
8.3.8 Tipos de Funções de Ativação	184
8.3.9 Mapas Auto-organizáveis	185
8.3.10 Mapas Auto-organizáveis Temporais	196
8.3.11 Outros tipos de Mapas Auto-organizáveis	199
8.3.12 Vantagens das Redes Neurais Artificiais	200
8.3.13 Desvantagens das Redes Neurais Artificiais	200
8.4 CONCLUSÕES DO CAPÍTULO	201
9 ANEXO C: FILTRAGEM ADAPTATIVA	202
9.1 TIPOS DE FILTROS	202
9.1.1 Filtros Determinísticos	204
9.1.2 Filtros Ótimos	204
9.2 FILTROS DE WIENER	205
9.2.1 Método Probabilístico ou Solução de Wiener	206
9.2.2 Método dos Mínimos Quadrados	207
9.3 FILTROS ADAPTATIVOS	208
9.3.1 Estrutura dos Filtros Adaptativos	209
9.3.2 Características dos Filtros Adaptativos	210
9.3.3 Classes de Aplicação de Filtros Adaptativos	211
9.3.4 Classificação dos Filtros Adaptativos	212
9.3.5 Método da Descida Mais Íngreme	214
9.3.6 Algoritmo dos Mínimos Quadrados Médio	216
9.3.7 Etapas dos Filtros Adaptativos	219
9.3.8 Outros Algoritmos de Filtragem Adaptativa	220

9.4 VANTAGENS DOS FILTROS ADAPTATIVOS	220
9.5 DESVANTAGENS DOS FILTROS ADAPTATIVOS	221
9.6 APLICAÇÃO DOS FILTROS ADAPTATIVOS	221
9.7 CONCLUSÕES DO CAPÍTULO	221
REFERÊNCIAS	223
10 APÊNDICE I: LISTA DE PUBLICAÇÕES	233

LISTA DE FIGURAS

Figura 1:	Esquema do sistema de manutenção inteligente embarcado.	29
Figura 2:	Operação do sistema: a) fase de treinamento e b) fase de teste.	31
Figura 3:	Dois exemplos de falhas: a) queima e b) curto no estator (WEG, 2002).	37
Figura 4:	Estratégias de manutenção baseadas na ocorrência de falhas (NUNES, 2001).	39
Figura 5:	Estratégias de manutenção baseadas no funcionamento dos sistemas.	40
Figura 6:	Classificação de manutenção adotada.	41
Figura 7:	Exemplo de diagrama de causa-efeito.	51
Figura 8:	Exemplo de gráfico de frequência de falha versus tempo desperdiçado.	52
Figura 9:	Estudo de caso: atuador elétrico, válvula e tubulação.	55
Figura 10:	Principais partes do atuador elétrico.	56
Figura 11:	Sistema de coordenadas.	59
Figura 12:	Engrenagens de: a) dentes retos b) coroa e sem-fim e c) cônicas.	64
Figura 13:	Janelas da interface do simulador do modelo do atuador elétrico.	71
Figura 14:	Curvas do modelo considerando as falhas do tipo 1: a) torque e b) posição.	73
Figura 15:	Curvas do modelo considerando as falhas do tipo 2: a) torque e b) posição.	73
Figura 16:	Curvas do modelo considerando as falhas do tipo 3: a) torque e b) posição.	74
Figura 17:	Representação da etapa de injeção de falhas.	75
Figura 18:	Valor de confiança para a falha do tipo 1 (K_h).	79
Figura 19:	Valor de confiança para a falha do tipo 2 (K_m).	80
Figura 20:	Valor de confiança para a falha do tipo 3 (C_a).	80
Figura 21:	Curvas de energia dos sinais de: a) torque e b) posição.	81
Figura 22:	Janelas da interface de simulação dos mapas auto-organizáveis.	83
Figura 23:	Etapa de treinamento dos MAO para a tarefa de detecção.	84
Figura 24:	Etapa de treinamento dos MAO para as tarefas de classificação e predição.	85
Figura 25:	Etapa de teste dos MAO para a tarefa de detecção e classificação.	86
Figura 26:	Erro de quantização para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).	87
Figura 27:	Resultados de classificação para as falhas do tipo: 1 (K_h), 2 (K_m) e 3 (C_a).	88
Figura 28:	Resultado de classificação para a falha do tipo 1 (K_h).	88
Figura 29:	Resultado de classificação para a falha do tipo 2 (K_m).	89
Figura 30:	Resultado de classificação para a falha do tipo 3 (C_a).	89
Figura 31:	Etapa de teste dos mapas temporais de Kohonen para a tarefa de predição.	90
Figura 32:	Resultado de predição para falha do tipo 1 (K_h).	90
Figura 33:	Resultado de predição para falha do tipo 2 (K_m).	91
Figura 34:	Resultado de predição para falha do tipo 3 (C_a).	91
Figura 35:	Gráficos de predição para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).	92
Figura 36:	Janelas da interface de simulação dos filtros adaptativos.	94
Figura 37:	Etapa de treinamento dos filtros adaptativos para a tarefa de detecção.	95
Figura 38:	Etapa de treinamento dos filtros adaptativos para a tarefa de classificação.	96

Figura 39:	Etapa de teste dos FA para a tarefa de detecção e classificação.	97
Figura 40:	Erro de detecção para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).	97
Figura 41:	Erro de classificação para a falha do tipo 1 (K_h).	98
Figura 42:	Erro de classificação para a falha do tipo 2 (K_m).	99
Figura 43:	Erro de classificação para a falha do tipo 3 (C_a).	99
Figura 44:	Etapa de teste dos filtros adaptativos para a tarefa de predição.	100
Figura 45:	Gráfico de predição para a falha do tipo 1 (K_h).	101
Figura 46:	Gráfico de predição para a falha do tipo 2 (K_m).	101
Figura 47:	Gráfico de predição para a falha do tipo 3 (C_a).	102
Figura 48:	Principais janelas da ferramenta <i>Curve Fitting Tool</i>	104
Figura 49:	Gráficos da: a) primeira e b) segunda derivada de $\mathbf{W}(t)$ para falha do tipo 1. 105	
Figura 50:	Gráfico de $R(t)$ para a falha do tipo 1 (K_h).	106
Figura 51:	Gráficos da: a) primeira e b) segunda derivada de $\mathbf{W}(t)$ para falha do tipo 2. 106	
Figura 52:	Gráfico de $R(t)$ para a falha do tipo 2 (K_m).	107
Figura 53:	Gráficos da: a) primeira e b) segunda derivada de $\mathbf{W}(t)$ para falha do tipo 3. 107	
Figura 54:	Gráfico de $R(t)$ para a falha do tipo 3 (C_a).	109
Figura 55:	Gráficos da: a) primeira e b) segunda derivada de \mathbf{y} para a falha do tipo 1. 109	
Figura 56:	Gráfico de $R(t)$ para a falha do tipo 1 (K_h).	110
Figura 57:	Gráficos da: a) primeira e b) segunda derivada de $\mathbf{y}(t)$ para a falha do tipo 2. 110	
Figura 58:	Gráfico de $R(t)$ para a falha do tipo 2 (K_m).	111
Figura 59:	Gráficos da: a) primeira e b) segunda derivada de $\mathbf{y}(t)$ para a falha do tipo 3. 112	
Figura 60:	Gráfico de $R(t)$ para a falha do tipo 3 (C_a).	112
Figura 61:	Arquitetura do circuito dos MAO: a) treinamento e b) teste.	116
Figura 62:	Arquitetura do circuito dos FA: a) treinamento e b) teste.	117
Figura 63:	Arquitetura do circuito de detecção dos mapas auto-organizáveis.	118
Figura 64:	Máquina de estados do circuito de detecção dos mapas auto-organizáveis. 119	
Figura 65:	Arquitetura do circuito de classificação dos mapas auto-organizáveis.	121
Figura 66:	Máquina de estados do circuito de classificação dos MAO.	122
Figura 67:	Arquitetura do circuito de ativação dos mapas temporais de Kohonen.	123
Figura 68:	Máquina de estados do circuito de ativação dos MTK.	124
Figura 69:	Arquitetura do circuito de predição dos mapas temporais de Kohonen.	125
Figura 70:	Máquina de estados do circuito de predição do MTK.	126
Figura 71:	Arquitetura do circuito do filtro adaptativo FIR.	127
Figura 72:	Máquina de estados do circuito do filtro adaptativo FIR.	129
Figura 73:	Arquitetura do circuito de detecção dos filtros adaptativos.	130
Figura 74:	Máquina de estados do circuito de detecção dos filtros adaptativos.	131
Figura 75:	Arquitetura do circuito de atualização dos filtros adaptativos.	132
Figura 76:	Máquina de estados do circuito de atualização dos filtros adaptativos.	133
Figura 77:	Arquitetura do circuito de predição dos filtros adaptativos.	134
Figura 78:	Máquina de estados do circuito de predição dos filtros adaptativos.	135
Figura 79:	Estratégia de envio/recepção de dados do sistema embarcado.	136
Figura 80:	Exemplo do uso da TF: a) sinal e b) conteúdo de frequência.	152
Figura 81:	Exemplo do uso da TDFJ: a) sinal e b) espectrograma.	154
Figura 82:	Representação das células de resolução Δt e Δf para a TFJ.	155
Figura 83:	Varredura do espectro de frequências através da TDFJ.	156
Figura 84:	Decomposição da TWD usando a análise multiresolução.	160
Figura 85:	Exemplo de árvore de decomposição da TWD.	161

Figura 86:	Exemplo do uso da TWD: a) sinal e b) coeficientes de aproximação (HHH).	161
Figura 87:	Representação das células de resolução Δt e Δf para TW.	162
Figura 88:	Varredura do espectro de frequências através da TW.	163
Figura 89:	Exemplo de árvore de decomposição da TWP.	165
Figura 90:	Exemplo do uso da TWP: a) sinal e b) coeficientes de aproximação (HHH).	166
Figura 91:	Sinais de a) força e b) de energia de um transdutor.	169
Figura 92:	Neurônio biológico (WILEY, 2010).	174
Figura 93:	Modelo de neurônio artificial.	176
Figura 94:	RN não realimentadas de: a) uma única camada e b) múltiplas camadas.	177
Figura 95:	Rede neural com realimentação.	178
Figura 96:	Aprendizado supervisionado.	179
Figura 97:	Aprendizado não-supervisionado.	180
Figura 98:	Aprendizado por reforço.	181
Figura 99:	Rede neural do tipo radial.	182
Figura 100:	Rede neural do tipo <i>perceptron</i> com uma única camada escondida.	183
Figura 101:	Arquitetura dos mapas auto-organizáveis.	186
Figura 102:	Exemplo de matriz de distâncias unificadas.	192
Figura 103:	Exemplo de classificação de falhas obtido a partir dos MAO.	193
Figura 104:	Exemplo de gráfico de erro de quantização.	196
Figura 105:	Exemplo de mapa auto-organizável temporal.	199
Figura 106:	Representação do problema de filtragem ótima linear.	205
Figura 107:	Filtro causal de estrutura transversal.	206
Figura 108:	Estrutura geral de um filtro adaptativo.	210
Figura 109:	Configurações básicas de filtros adaptativos.	211
Figura 110:	Filtro adaptativo FIR.	214
Figura 111:	Diagrama de blocos do sistema de filtragem adaptativa via MQM.	217
Figura 112:	Etapas dos filtros adaptativos: a) treinamento e b) teste.	219

LISTA DE TABELAS

Tabela 1:	Principais características das estratégias clássicas de manutenção.	38
Tabela 2:	Tipos de simulações de injeção de falhas.	73
Tabela 3:	Faixa e taxa de variação de K_h , K_m e C_a	76
Tabela 4:	Quadro comparativo do T_u obtido com os MAO e FA.	113
Tabela 5:	Especificação do mapa auto-organizável para embarque.	137
Tabela 6:	Especificação do filtro adaptativo para embarque.	137
Tabela 7:	Área do circuitos obtida após o roteamento.	139
Tabela 8:	Desempenho dos circuitos obtido após a síntese.	140
Tabela 9:	Tempo de processamento dos circuitos obtido após roteamento.	140
Tabela 10:	Diferentes tipos de análise de Fourier.	149
Tabela 11:	Quadro comparativo das ferramentas de processamento de sinais.	167
Tabela 12:	Funções de ativação de redes <i>perceptron</i> de múltiplas camadas.	184
Tabela 13:	Funções de ativação de redes de função de base radial.	185
Tabela 14:	Principais classes de algoritmos de filtros adaptativos.	212

LISTA DE ABREVIATURAS E SIGLAS

ABNT:	Associação Brasileira de Normas Técnicas
ASIC:	<i>Application Specific Integrated Circuit</i>
CBM:	<i>Condition-Based Maintenance</i>
CC:	<i>Corrente contínua</i>
DGE:	Dispositivos Genéricos Embarcados
FBR:	Funções de Base Radial
FPGA:	<i>Field Programmable Gate Array</i>
FA:	Filtros Adaptativos
FMEA:	<i>Failure Mode and Effects Analysis</i>
FMECA:	<i>Failure Mode, Effects, and Criticality Analysis</i>
HARF:	<i>Hyperstable Adaptive Recursive Filter</i>
IA:	Inteligência Artificial
ICAS:	<i>Integrated Condition Assessment System</i>
IEEE:	<i>Institute of Electrical and Electronics Engineers</i>
IIR:	<i>Infinite-duration Impulse Response</i>
IMS:	<i>Intelligent Maintenance Systems</i>
IOB:	<i>Input Output Block</i>
LN:	Lógica Nebulosa
LUT:	<i>Look-Up Table</i>
LMS:	<i>Least Mean Square</i>
MAO:	Mapas Auto-organizáveis
MDMI:	Método da Descida Mais Íngreme
MDU:	Matriz de Distâncias Unificada
MGA:	Malha de Gradiente Adaptativo
MQ:	Mínimos Quadrados
MQM:	Mínimos Quadrados Médio
MQMN:	Mínimos Quadrados Médio Normalizado

MQR:	Mínimos Quadrados Recursivo
MMQ:	Malha de Mínimos Quadrados
MTK:	Mapa Temporal de Kohonen
NBR:	Norma Brasileira Regulatória
OSA:	<i>Open Systems Architecture</i>
PMC:	<i>Perceptron</i> de Múltiplas Camadas
PPGEE:	Programa de Pós-Graduação em Engenharia Elétrica
PROTEUS:	<i>Plateforme Reconfigurable pour l'Observation, pour les Télécommunications et les Usages Scientifiques</i>
RN:	Redes Neurais
RNA:	Redes Neurais Artificiais
SHARF:	<i>Simplified Hyperstable Adaptive Recursive Filter</i>
SMI:	Sistema de Manutenção Inteligente
SRAM:	Static Random Access Memory
TDF:	Transformada Discreta de Fourier
TDFJ:	Transformada Discreta de Fourier com Janelamento
TELMA:	<i>TELé-MAintenance</i>
TF:	Transformada de Fourier
TFJ:	Transformada de Fourier com Janelamento
TMEF:	Tempo Médio Entre Falhas
TMR:	Tempo Médio para Reparo
TW:	Transformada <i>Wavelet</i>
TWD:	Transformada <i>Wavelet</i> Discreta
TWP:	Transformada <i>Wavelet Packet</i>
UFRGS:	Universidade Federal do Rio Grande do Sul
VHDL:	<i>VHSIC Hardware Description Language</i>
VHSIC:	<i>Very High Speed Integrated Circuit</i>

LISTA DE SÍMBOLOS

Parâmetros Referentes ao Processamento de Sinais

A	Coefficientes de aproximação
C_ψ	Constante espectral
D	Coefficientes de detalhamento
\mathcal{D}	Transformada discreta de Fourier com Janelamento
\mathcal{E}	Energia do sinal contínuo $x(t)$
\mathbf{E}	Vetor de energia do sinal $y[n]$
f	Frequência
f_0	Frequência fundamental de um sinal
f_h	Componente de mais alta frequência do sinal $x(t)$
\mathcal{F}	Transformada contínua de Fourier
g	Filtro de quadratura (passa-altas)
\mathcal{G}	Transformada discreta de Fourier
h	Filtro de quadratura (passa-baixas)
k	Índice de amostragem
K	Número de detalhes
L	Limite de um quadro da TW
\mathcal{L}	Energia do sinal discreto $x[n]$
n	Índice de amostragem
N	Número de amostras
\mathcal{P}	Transformada <i>Wavelet Packet</i>
\mathcal{S}	Transformada contínua de Fourier com Janelamento
t	Instante de tempo
\mathcal{T}	Transformada de Fourier de Tempo Discreto
\mathcal{V}	Transformada <i>Wavelet</i> discreta
\mathcal{W}	Transformada <i>Wavelet</i> contínua
ω	Frequência angular
$x(t)$	Sinal contínuo no domínio tempo

$x[n]$	Sinal discreto no domínio tempo
$X(\omega)$	Sinal contínuo no domínio frequência
$X[n]$	Sinal discreto no domínio frequência
α	Parâmetro de dilatação
τ	Parâmetro de translação
β	Variação da dilatação
θ	Função de janelamento
ϕ	Função de escalonamento
Φ	Transformada de Fourier da função de janelamento
Θ	Transformada de Fourier da função de escalonamento
ψ	Função <i>Wavelet</i> mãe
Ψ	Transformada de Fourier da função <i>Wavelet</i> mãe $\psi(t)$
Δt	Resolução temporal
Δf	Resolução espectral
κ	Variação da escala
η	Parâmetro de frequência
v	Coefficientes da TWP
μ_E	Média do vetor \mathbf{E}
σ_E	Variância do vetor \mathbf{E}

Parâmetros Referentes à Inteligência Artificial

A	Ativação temporal do neurônio
C_m	Número de colunas do mapa auto-organizável
\mathbf{d}	Vetor de resposta desejada
D	Distância Euclidiana
\mathbf{e}	Vetor de erro
E_q	Erro quadrático
h	Função de vizinhança
L_m	Número de linhas do mapa auto-organizável
\mathcal{M}	Mapa auto-organizável
M	Número de elementos do vetor de entrada da rede neural
P	Número de neurônios da camada de escondida
Q	Número de neurônios da camada de saída
\mathbf{W}	Vetor de pesos sinápticos do mapa auto-organizável
μ_W	Média do vetor de pesos sinápticos

ζ	Vetor de saídas da rede neural artificial
$\tilde{\xi}$	Espaço contínuo de dados de entrada
ξ	Vetor de entradas da rede neural artificial
δ	Vetor de ativação do neurônio artificial
χ	Vetor de pesos sinápticos
$\tilde{\chi}$	Espaço contínuo de dados de saída
$\Delta\chi$	Vetor de ajuste dos pesos sinápticos
ε	Erro de quantização
φ	Função de ativação do neurônio artificial
υ	Variável da função de ativação
σ	Largura da vizinhança
λ	Tamanho da memória
α	Taxa de aprendizado
τ	Constante de tempo

Parâmetros Referentes à Filtragem Adaptativa

c	Vetor de erro de cálculo dos coeficientes
d	Vetor de resposta desejada
e	Vetor de erro de estimação
J	Função de custo
k	Índice de amostragem
M	Número de coeficientes do filtro adaptativo
n	Índice de amostragem
p	Vetor de correlação cruzada
Q	Matriz de autovetores
R	Matriz de correlação
u	Vetor de entrada do filtro adaptativo
v	Vetor de aprendizagem
y	Vetor de saída do filtro adaptativo
w	Vetor de coeficientes do filtro adaptativo
w₀	Vetor de coeficientes ótimos do filtro adaptativo
ε	Valor esperado de uma variável
σ_d	Variância do sinal desejado
λ	Autovalores da matriz de correlação

τ	Constante de tempo
μ	Tamanho do passo

Parâmetros Referentes à Modelagem do Atuador

a	Posição de haste
a_a	Aceleração da haste
A_v	Área do obturador da válvula
C_a	Constante de atrito
C_v	Coefficiente de vazão
d_1	Distância entre o eixo de rotação do sem fim e o ponto de aplicação de F_1
d_2	Distância entre o eixo de rotação da coroa e o ponto de aplicação de F_2
d_h	Distância entre o acoplamento da válvula e a sua gaveta
$f()$	Função vetorial não linear (equações diferenciais)
$g()$	Função vetorial não linear (equações algébricas)
F_1	Força aplicada no sem-fim
F_2	Força aplicada na coroa
F_a	Força de atrito
F_f	Força do fluido
F_h	Força do atuador
F_m	Força da mola
F_p	Fator de geometria da tubulação
F_y	Fator de fluxo crítico
H_g	Constante de inércia
\mathbf{i}	Vetor de corrente
I_{ds}	Corrente do estator no eixo direto
I_{qs}	Corrente do estator no eixo em quadratura
K_m	Coefficiente de hook da mola
K_R	Coefficiente de redução cinemática geral
$K_{1,2,3}$	Coefficientes de redução cinemática
L_r	Indutância do rotor
L_s	Indutância do estator
L_m	Indutância de magnetização
N_r	Número de Reynolds
N_m	Fator de ajuste numérico
M_h	Massa da haste

p	Vetor de constantes
P_s	Passo do sem-fim
R_r	Resistência do rotor
R_s	Resistência do estator
s	Coefficiente de escorregamento
T_e	Torque eletromagnético
T_h	Torque transmitido para a manga de saída
T_m	Torque mecânico fornecido pelo rotor
T_{mb}	Torque mecânico de base do motor
u	Vetor de entradas
v_a	Velocidade da haste
V_{ds}	Tensão do estator no eixo direto
V_{qs}	Tensão do estator no eixo em quadratura
V'_d	Tensão interna transitória no eixo direto
V'_q	Tensão interna transitória no eixo em quadratura
V_f	Vazão do fluido
v	Tensão
v	Velocidade tangencial da coroa,
ψ	Fluxos magnético
X	Reatância
x	Vetor de variáveis de estado
z	Vetor de variáveis algébricas
ω_b	Frequência de base
ω_r	Frequência do rotor
ω_s	Frequência síncrona do sistema
ρ	Massa específica do fluido
θ	Ângulo entre a força do atuador e a haste
ΔP	Diferença de pressão

Índices Inferiores

m	Índice dos neurônios da camada de entrada da RNA
p	Índice dos neurônios da camada escondida da rede PMC
q	Índice dos neurônios de camada de saída da rede PMC
l	Índice dos neurônios da camada de saída da RN de FBR
d	Grandezas relacionadas ao eixo direto

q	Grandezas relacionadas ao eixo de quadratura
s	Grandezas relacionadas ao estator
r	Grandezas relacionadas ao rotor
min	Valor mínimo
$máx$	Valor máximo

Índices Superiores

T	Transposta da matriz
H	Transposta e conjugado da matriz
-1	Transformada inversa
$*$	Conjugado
j^*	Neuônio vencedor
j	Índice do vetor de entrada dos MAO
k	Índice do vetor de coeficientes dos MAO
\cdot	Primeira derivada
$\ddot{}$	Segunda derivada

Outros Símbolos

j	Número imaginário
I	Matriz identidade
Z	Atraso unitário
$\ \cdot \ $	Norma
$ \cdot $	Módulo
$\langle \cdot \rangle$	Produto interno
$*$	Convolução
$\hat{}$	Variância
d	Operador diferencial
\int	Integral da grandeza
\forall	Para todos, para qualquer
Σ	Somatório
$\{ \cdot \}$	Conjunto de valores
\mathbb{N}	Conjunto dos números naturais
\mathbb{Z}	Conjunto dos números inteiros.
\mathbb{R}	Conjunto dos números reais
\mathbb{I}	Conjunto dos números imaginários

1 INTRODUÇÃO

Os equipamentos industriais, dispositivos ou máquinas, à medida que são utilizados, podem sofrer desgastes físicos e químicos, tais como: corrosões, rachaduras, acúmulo de sujeiras, infiltrações, entre outros. Algumas medidas podem ser tomadas com o objetivo de restabelecer a funcionalidade desses equipamentos e evitar falhas.

A manutenção consiste de uma série de medidas, técnicas ou ações, para manter os equipamentos ou processos em funcionamento, restabelecendo os sistemas danificados, corrigindo os danos e estendendo o ciclo de vida.

O início da manutenção se deu durante a revolução industrial, época em que um conjunto de mudanças tecnológicas provocou profundas alterações nos processos produtivos em níveis econômicos e sociais. As atividades de manutenção e os recursos destinados a novos projetos de recuperação de falhas eram quase inexistentes ou ficavam em segundo plano.

Naquela época, o processo de industrialização dos meios de produção estava em seu início e havia pouca preocupação com a produtividade que não era considerada relevante, não se creditando qualquer importância à prevenção das falhas.

A estratégia de manutenção, desta época, consistia de um simples conserto dos equipamentos após alguma avaria. Ao mesmo tempo, a maioria das máquinas era simples e muitas delas super dimensionadas, apresentando facilidade de reparo e alguma confiabilidade.

Estes consertos e substituições de peças correspondem à primeira noção de manutenção que se tem conhecimento: a manutenção corretiva. Essa estratégia pode ser entendida como as atividades de manutenção realizadas após a ocorrência de alguma falha nos sistemas, visando restabelecê-los às suas funções originais (MARÇAL, 2000; ENDRENYI et al., 2001).

Neste panorama, não havia necessidade de manutenção sistematizada, nem de pessoal especializado para a sua execução. Contudo, este contexto se modificou com o crescente aumento da mecanização ocorrida durante o início do século XX.

A partir da década de 50, as indústrias e a sociedade passaram a depender cada vez mais das máquinas, que começaram a se multiplicar e modificar em tipo e complexidade. Deste modo, prevenir uma eventual paralisação das máquinas tornou-se cada vez mais relevante.

Nos anos 60, a manutenção consistia de revisões gerais dos equipamentos executados em intervalos pré-determinados. Nesse período, começou a ser estruturado o conceito de manutenção preventiva (NUNES, 2001).

Na manutenção preventiva as ações de manutenção são planejadas e realizadas periodicamente visando diminuir ou evitar paradas no sistema produtivo provocadas por falhas inesperadas (SHIKARI; SADIWALA; DWIVEDI, 2004; DJURDJANOVIC et al., 2003).

A quantidade de capital investido em reparos, já no início da década de 70, e o custo do próprio capital direcionaram as ações para o aumento da vida útil dos equipamentos e das instalações. Preocupação que se transformou na base para o desenvolvimento de novas técnicas de manutenção, nesta época, suportadas por computadores de grande porte.

A partir de então, a necessidade das máquinas operarem continuamente sem falhas vem sendo cada vez maior e qualquer interrupção devido às falhas passou a tomar proporções cada vez mais críticas.

A manutenção preditiva, que surgiu nesta época, tem como objetivos principais aumentar a confiança e a disponibilidade, reduzir o número de falhas e aumentar a vida útil de equipamentos e instalações através da supervisão e controle dos equipamentos.

O uso de ferramentas de manutenção preditiva, em especial, com a intenção de antecipar-se às falhas e sinalizar o momento adequado de se realizar um serviço de manutenção tem aumentado. Principalmente, porque o custo de uma simples parada não atendida que perdure por certo intervalo de tempo pode comprometer o perfil de produção ou atendimento, além de superar inúmeras vezes o custo de projeto e instalação de novos meios de controle e comando.

A evolução histórica da manutenção pode ser dividida de modo simplificado em três gerações distintas (ALKAIM, 2003):

1. A primeira geração apresentava uma maior ênfase no conserto dos equipamentos após a ocorrência de falhas e quebra dos equipamentos.
2. A segunda geração, por sua vez, esteve associada ao surgimento de maiores exigências com relação à disponibilidade operacional e à vida útil dos equipamentos a custos menores.
3. Já a terceira geração, que se refere aos tempos atuais, diz respeito ao cumprimento de certos requisitos característicos, tais como: maior disponibilidade, confiabilidade, segurança e vida útil, com ausência de danos ao meio-ambiente e ações de manutenções mais eficazes, aliados à redução dos custos envolvidos.

Ainda, diversas são as estratégias e classificações encontradas na literatura a respeito da manutenção. Geralmente, a manutenção é dividida de acordo com as três modalidades clássicas: preditiva, preventiva, e corretiva.

Contudo, recentemente, um novo enfoque de manutenção está em destaque: a manutenção proativa. A manutenção proativa pode ser vista como um avanço em relação as estratégias clássicas; pois essa, além de monitorar o equipamento, realiza diagnósticos, quantifica a perda de desempenho e atua sobre o sistema em tempo real.

No caso de detecção de uma falha, ao se utilizar um esquema de manutenção proativa, um sistema de controle ou mesmo um sistema embarcado, pode automaticamente atuar no equipamento de modo que ele continue operando, mesmo com certo nível de degradação, até a realização da manutenção.

A degradação do desempenho de equipamentos, em especial, é vista como resultado do uso, exposição ao ambiente, passagem do tempo e desgaste de peças. Estes fatores diminuem a confiabilidade das máquinas e aumentam a probabilidade de falhas.

A quantificação do nível de degradação pode ser usada para prever um desempenho inaceitável nos equipamentos e sinalizar a ocorrência de falhas, por exemplo. A quantificação da degradação permite sinalizar o momento adequado para se realizar as atividades de manutenção e a aumentar a reutilização de peças ou componentes, por exemplo.

Assim, a detecção do comportamento de sistemas é útil para avaliar os efeitos provocados pela degradação, sustentar o correto funcionamento dos equipamentos e diminuir a ocorrência de defeitos e falhas (DJURDJANOVIC et al., 2003).

Um defeito permite que o sistema esteja disponível para uso. Uma falha, ao contrário, já impede o sistema de desempenhar a função requerida. Assim, prevenir falhas e danos nos equipamentos, enquanto estes ainda estão apresentando um defeito, de forma que as paradas no processo produtivo e o tempo de manutenção de uma máquina possam ser reduzidos ao mínimo, pode trazer inúmeros benefícios, tais como um aumento da confiabilidade e da disponibilidade, redução de custos e uma maior proteção dos equipamentos.

Aliado a isso, a complexidade cada vez maior dos sistemas, com a aceleração da automação, transformou a confiabilidade e a disponibilidade em fatores primordiais para um bom desempenho operacional. Os padrões de qualidade, tanto nos serviços quanto nos produtos, passaram a ser mais rigorosos e a análise das falhas e de suas conseqüências para a segurança e o meio-ambiente, representam, em muitos casos, a garantia de sobrevivência das empresas.

Atualmente, as grandes transformações experimentadas pelo setor industrial permitem realizar um diagnóstico e análise muito maior das origens das falhas e dos seus efeitos. Estas transformações vêm influenciando diretamente as atividades de manutenção atuais em especial a proativa.

Esta revolução está sendo dirigida principalmente pela renovação tecnológica e pelos métodos de detecção, classificação e predição de falhas (LEE et al., 2006; YAN; GAO, 2005). Neste sentido, o aspecto tecnológico, além do econômico, sempre presente na vida das organizações e das atividades de manutenção, passa ser mais importante.

Deste modo, novas técnicas e tecnologias voltadas para a manutenção estão sendo implementadas com a intenção de aumentar a confiabilidade das máquinas, manter os serviços de alta qualidade e minimizar tempo de parada de produção.

Assim, os serviços de manutenção estão se tornando tecnologias que favorecem a competitividade para sustentar e customizar a produção nas indústrias de modo a satisfazer o mercado global o máximo possível. Em linhas gerais, dentre os fatores que vem transformando a área da manutenção em um segmento estratégico, pode-se citar:

- A evolução tecnológica dos equipamentos, processos e técnicas de manutenção;
- A necessidade de um controle de equipamentos cada vez mais eficiente;
- O desenvolvimento de ferramentas de apoio à decisão;
- O avanço de estudos relativos ao desgaste das peças e suas conseqüências.

Dentre as ferramentas matemáticas atualmente utilizadas para o desenvolvimento de sistemas de manutenção (que possam identificar as possíveis falhas, além de gerenciar suas conseqüências), citam-se as da área de processamento de sinais, de inteligência artificial (IA) e de filtragem adaptativa.

O processamento de sinais compreende a análise de sinais de modo a capturar informações destes e deixá-los mais adequados a alguma aplicação. Dentre as principais técnicas de processamento de sinais, cita-se a transformada de Fourier (TF), a transformada *wavelet* (TW) e a transformada *wavelet packet* (TWP), por exemplo (LATHI, 1998; SARKAR; SU, 1998).

Estas técnicas podem ser empregadas, por exemplo, na decomposição do sinal de sensores, instalados em máquinas, no domínio tempo e frequência, localizando o espectro de frequência destes sinais (PARRAGA, 2002); podem também ser usadas para o reconhecimento de padrões (QUISPE, 2005); ou na detecção de falhas. (DJURDJANOVIC; LEE; NI, 2003).

As ferramentas de processamento de sinais são muito utilizadas na extração de características, no reconhecimento de padrões e na detecção e predição de falhas, por exemplo, geralmente em conjunto com técnicas estatísticas, de IA ou de filtragem adaptativa.

A inteligência artificial preocupa-se em desvendar a natureza do raciocínio humano com o intuito de elaborá-lo em sistemas computacionais, provendo, assim, certa capacidade de aprendizado e de auto-ajuste para esses sistemas. Como exemplo de ferramentas de IA, cita-se: as redes neurais artificiais (RNA), a lógica nebulosa (LN) e as cadeias de Markov.

As aplicações, encontradas na literatura, que fazem uso das técnicas de IA, em especial das RN, são muitas, cita-se: detecção de falhas em robôs (TINÓS, 1999), diagnóstico de defeitos em sistemas mecânicos (FUJIMOTO, 2005) e análise da confiabilidade e disponibilidade em sistemas elétricos de potência (JENS, 2006).

Um outro exemplo de trabalho que utiliza as redes neurais pode ser visto em (HUANG et al., 2005). Nesse trabalho, foram usados métodos de avaliação e predição da vida útil de enrolamentos baseados em um tipo de rede neural específico, conhecido como mapas auto-organizáveis (MAO) ou mapas de Kohonen (KOHONEN et al., 1995).

Quando estes métodos de raciocínio utilizados pelo ser humano foram modelados e desenvolvidos no computador, através do uso da inteligência artificial em conjunto com ferramentas de processamento de sinais, ocorreu uma expansão destas tecnologias para a manutenção, tornando-a mais inteligente.

Na grande maioria das aplicações, as redes neurais artificiais (em especial os mapas auto-organizáveis) têm sido empregadas como classificadores baseados nas variáveis medidas do sistema. Entretanto, há pouca pesquisa científica centrada na investigação de como utilizar os mapas auto-organizáveis para detectar e classificar falhas e para prever o tempo de vida de atuadores elétricos e válvulas, por exemplo.

Dentre as ferramentas de filtragem adaptativa, tais como: os filtros de Wiener, filtros adaptativos ou os filtros de Kalman, destacam-se os filtros adaptativos (FA) (HAYKIN, 2002; NEGREIROS, 1994; WIDROW; STEARNS, 1985).

Os filtros adaptativos são filtros com coeficientes que variam ao longo do tempo, adaptando-se automaticamente aos sinais de entrada. Geralmente, o processo de adaptação é baseado em um algoritmo de minimização do erro.

A filtragem adaptativa é desejável em muitas situações em engenharia, tais como a remoção de ruídos, a detecção e a predição de falhas. Em (BEZERRA, 2004), por exemplo, foi realizado um estudo comparativo entre diferentes técnicas de detecção de falhas em rolamentos por análise de vibração. Os resultados alcançados mostraram que das técnicas usadas o FA foi a mais eficiente, sendo capaz de detectar novas falhas nos rolamentos.

De modo semelhante aos MAO, ainda há pouca pesquisa científica voltada para a investigação de como fazer uso dos FA para a detecção, classificação e predição de falhas de atuadores elétricos, por exemplo.

Os sistemas de manutenção (definidos como sistemas de manutenção inteligentes) que utilizam estas ferramentas (estatísticas, de processamento de sinais, de IA e filtragem adaptativa) requerem, basicamente, sensores instalados nos equipamentos e um acesso aos sinais gerados por estes que aliados a aplicativos e simuladores, permitem a análise e predição de eventos.

A manutenção inteligente, em sua plenitude, é uma aplicação direta do conceito de adaptabilidade¹ em nível de macro-sistema. Um sistema de manutenção inteligente (SMI), também conhecido como proativo, permite estimar a “saúde” de um equipamento ao longo de todo o seu ciclo de vida útil. Isto é feito através da utilização de técnicas de monitoração, avaliação do nível de degradação e de atuação no caso de necessidade.

Além destes aspectos, um sistema de manutenção que implementa os procedimentos de monitoração, avaliação e adaptação precisa ser confiável para garantir que, mesmo quando afetado por uma falha, as informações que ele presta e ações que toma levem o macro-sistema a um estado seguro de operação. Esta confiabilidade pode ser garantida com base no conceito da adaptabilidade, aplicado neste caso de maneira recursiva até o nível da aplicação.

Além disso, diversos sistemas de planejamento e controle, desenvolvidos em conjunto, vêm sendo utilizados para obter-se um melhor controle das ações de manutenção e aumentar a eficácia das atividades de manutenção. Os sistemas informatizados e, posteriormente, os sistemas embarcados vêm ampliando ainda mais o desempenho das atividades de manutenção.

Um sistema de manutenção embarcado é um sistema projetado para coletar informações de sensores, processar dados e atuar sobre o equipamento a ele conectado. Estes sistemas podem ser usados para controlar as variáveis e o comportamento de um equipamento, realizar predições e diagnósticos em tempo real das possíveis falhas (ALTUG; CHOW; TRUSSELL, 1999; SU et al., 2005).

Neste trabalho, pretende-se embarcar em *hardware* um conjunto de ferramentas para a implementação de um sistema de manutenção inteligente. Para isto, serão utilizadas plataformas que possuam elementos programáveis.

¹É a disposição de um sistema para se integrar ao ambiente e dar respostas adequadas a novas situações.

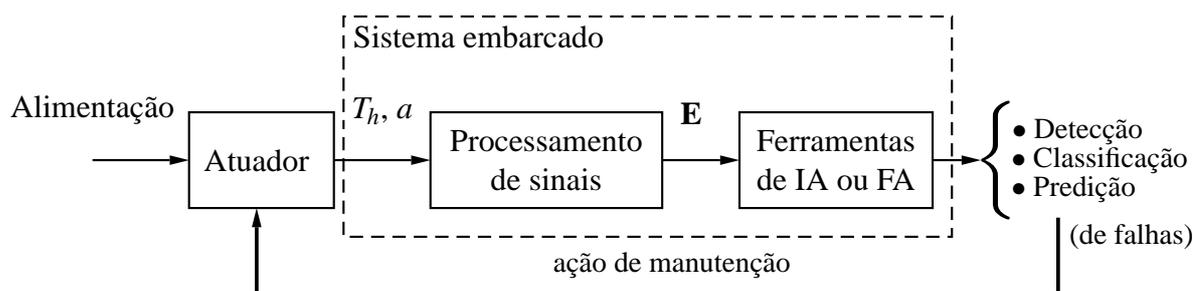


Figura 1: Esquema do sistema de manutenção inteligente embarcado.

A idéia básica é embarcar as ferramentas matemáticas, além de outras funcionalidades, em FPGA (*Field Programmable Gate Array*) e usá-las em uma placa para auxiliar na manutenção de um atuador elétrico, como visto na Fig. 1.

Atuadores elétricos são equipamentos que, acoplados a uma válvula, permitem controlar o fluxo de um líquido (água ou petróleo, por exemplo) que passa por uma tubulação ou por uma comporta, por exemplo (CAMPOS; TEIXEIRA, 2006).

Neste trabalho, o estudo de caso analisado é um sistema formado por atuador elétrico, válvula gaveta e tubulação. Este sistema é usado para controlar o fluxo de petróleo em uma rede de distribuição da empresa Transpetro S.A.

O sistema embarcado, irá realizar com base nos valores de torque (T_h) e do percentual de abertura da válvula (a), as etapas de detecção, classificação e predição de falhas, fornecendo informações sobre o estado atual e futuro do sistema.

O embarque se faz necessário visto que, por meio dos sistemas embarcados, é possível realizar tarefas específicas, em especial, em tempo-real. Além disso, é possível alterar rapidamente as especificações do sistema embarcado, aumentar a confiabilidade e reduzir custos com o uso dos sistemas embarcados.

Sendo assim, foi desenvolvido em VHDL um sistema embarcado, voltado para a detecção, classificação e predição de falhas, baseado em ferramentas de processamento de sinais, filtragem adaptativa e inteligência artificial.

As ferramentas de processamento de sinais analisadas foram: a transformada de Fourier, a transformada de Fourier com janelamento e a transformada *wavelet packet*. Através do uso destas técnicas de processamento de sinais é possível extrair as informações características dos sinais do atuador (como a energia do sinal, E , por exemplo) para cada tipo de comportamento (normal, degradação ou falha) ou estado, tornando possível detectar ou prever a ocorrência de alguma falha, como observado na Fig. 2.

Além disto, o uso correto destas ferramentas de processamento de sinais permite aumentar a confiabilidade e disponibilidade do sistema; tornar mais robusto as atividades de detecção, classificação e predição de falhas e, em especial, reduzir a frequência das inspeções e das atividades de manutenção e aumentar a vida útil do atuador elétrico.

Com relação às ferramentas de inteligência artificial, as RNA foram escolhidas. Mais especificamente, os mapas auto-organizáveis e os mapas temporais de Kohonen.

Estas ferramentas foram analisadas e utilizadas para realizar as etapas de detecção, e classificação (MAO), e predição (MTK) de falhas, respectivamente.

As RNA foram escolhidas, pois podem aprender sobre a condição normal de funcionamento do atuador elétrico e determinar se os sinais estão significativamente diferentes desta condição; são úteis quando há uma grande quantidade de dados disponíveis e podem se adaptar e aprender sobre processos dinâmicos e não-lineares.

Os filtros adaptativos também foram analisados e usados para realizar as etapas de detecção, classificação e predição de falhas no atuador elétrico, de modo semelhante às RNA.

A análise e uso dos filtros adaptativos se justifica devido ao baixo número de equações, à sua facilidade de implementação em *hardware* e a sua capacidade de realização das tarefas de detecção, classificação e predição de falhas.

Assim, este trabalho apresenta as principais etapas para o desenvolvimento de um sistema de manutenção inteligente embarcado (SMI) voltado para a detecção, classificação (ou diagnóstico) e predição (ou prognóstico) de falhas em atuadores elétricos².

Através do uso de um conjunto de ferramentas de processamento de sinais, de inteligência artificial ou de filtragem adaptativa (mais especificamente a transformada *wavelet packet*, os mapas auto-organizáveis ou os filtros adaptativos) embarcadas, espera-se realizar uma detecção e classificação das falhas e ainda efetuar uma predição do tempo de vida útil do atuador. Assim, espera-se também auxiliar na programação e realização das operações de manutenção desse equipamento.

1.1 OPERAÇÃO

A operação do sistema de manutenção inteligente embarcado é dividida em duas fases: uma fase de treinamento (geração de dados do modelo, cálculo da energia dos sinais de torque e posição e treinamento da ferramenta de inteligência artificial ou de filtragem adaptativa) e uma de teste (realização das etapas de detecção, classificação ou predição de falhas), como visto na Fig. 2.

Na fase de treinamento (realizada em um computador) os valores de torque da haste (T_h) e de percentual de abertura da válvula (a) (dados que representam o comportamento normal, de degradação e de falha do estudo de caso) foram gerados por um modelo matemático (composto por um conjunto de equações) do atuador elétrico. Estes valores foram repassados para a ferramenta de processamento de sinais a seguir. A energia dos sinais de torque da haste e de percentual de abertura (E) obtidos a partir desta ferramenta foram então repassados para a ferramenta de inteligência artificial ou de filtragem adaptativa. Por fim, os valores dos pesos sinápticos obtidos com a ferramenta de inteligência artificial (os mapas auto-organizáveis, mais especificamente) ou dos coeficientes do filtro (filtro adaptativo) foram armazenados na memória do sistema embarcado.

²Neste trabalho serão adotados os termos classificação e predição ao invés de diagnóstico e prognóstico.

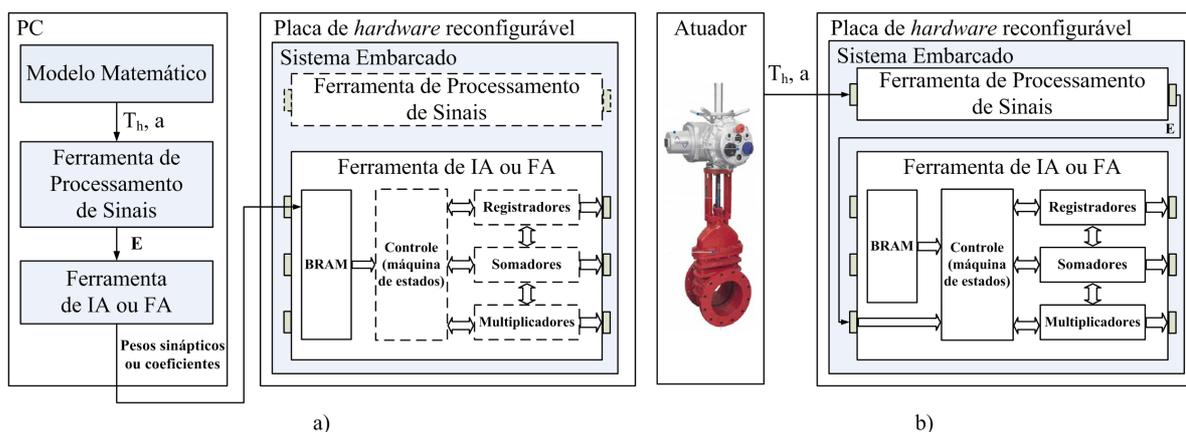


Figura 2: Operação do sistema: a) fase de treinamento e b) fase de teste.

Na fase de teste (realizada em uma placa) os valores de torque da haste e de percentual de abertura da válvula (dados que representam um comportamento real ou visam simular uma situação real de utilização do atuador ao longo do tempo) são novamente repassados para a ferramenta de processamento de sinais embarcada. A energia dos sinais de torque da haste e de percentual de abertura (**E**) obtidos a partir desta ferramenta são então mais uma vez repassados para a ferramenta de inteligência artificial ou de filtragem adaptativa. Estas ferramentas, embarcadas, são responsáveis, por fim, pela realização das etapas de detecção, classificação e predição de falhas do atuador elétrico.

Para tal, as etapas de teste (ou recuperação), relativa aos mapas auto-organizáveis e mapas temporais de Kohonen e aos filtros adaptativos, mais especificamente, foram programadas em VHDL e embarcadas uma placa da Xilinx. A ferramenta de processamento de sinais encontra-se embarcada, sendo atualmente programada em linguagem C.

Os circuitos do sistema embarcado, programados em VHDL, são compostos por distintos blocos: memórias (BRAM), registradores, somadores, subtratores, comparadores e multiplicadores, por exemplo, entre outros.

1.2 MOTIVAÇÃO

Atualmente, grande parte das atividades de manutenção dos atuadores, válvulas, e tubulações, além das tarefas de controle destes equipamentos, é realizada por empresas como a Petrobrás, Transpetro e outras.

Deste modo, uma previsão de falhas, além de um diagnóstico preciso das causas que contribuíram para o aparecimento destas falhas, melhoraria estas atividades.

Assim, a manutenção proativa e o desenvolvimento de sistemas de manutenção inteligente embarcados vêm se tornando especialmente interessante para a indústria petrolífera nacional, pois através dessa é possível realizar diagnósticos e previsões de possíveis falhas que possam ocorrer nos sistemas de produção e extração de petróleo.

1.3 OBJETIVOS

Dentre os principais objetivos do desenvolvimento deste sistema de manutenção inteligente embarcado voltado para atuadores elétricos, cita-se:

- Quantificar e qualificar a degradação do comportamento do atuador elétrico;
- Realizar uma predição do tempo de vida útil do atuador elétrico;
- Sinalizar o momento mais adequado para se efetuar a manutenção;
- Facilitar a desmontagem e a reutilização de peças;
- Reduzir custos de manutenção e de reposição de peças.

Assim, espera-se com a implantação embarcada desse sistema: auxiliar nas tarefas de manutenção, evitar danos nos atuadores em tempo real, reduzir os custos provenientes de falhas inesperadas, e promover uma maior sustentabilidade do sistema.

Um outro objetivo deste trabalho é determinar qual das ferramentas de processamento de sinais (TF, TFJ e TWP) e quais das ferramentas usadas nas etapas de detecção, classificação e predição de falhas (MAO ou FA) são as mais adequadas (em termos de área, desempenho e tempo de processamento) para serem embarcadas em uma placa de circuito impresso.

1.4 METODOLOGIA

A metodologia utilizada para realizar o desenvolvimento deste sistema de manutenção inteligente segue o padrão *Open Systems Architecture for Condition-Based Maintenance*, ou simplesmente OSA-CBM, para o desenvolvimento de sistemas de manutenção baseado na condição dos sistemas em análise (MATHEW et al., 2006).

Esta metodologia pressupõe a finalização das seguintes etapas:

1. Monitoração de parâmetros vitais;
2. Processamento de sinais;
3. Avaliação da degradação;
4. Predição do comportamento;
5. Diagnóstico do sistema, e;
6. Adaptação e/ou tomada de decisão.

O padrão OSA-CBM é um padrão de comunicação aberto para sistemas de monitoração e diagnóstico de máquinas que divide um sistema de manutenção em diferentes módulos ou níveis (BENGTSSON, 2003; MULLER; MARQUEZ; IUNG, 2008).

1.5 CONTRIBUIÇÕES

Dentre as principais contribuições deste trabalho, destaca-se:

1. Desenvolvimento de um modelo matemático (formado por um sistema de equações diferenciais e algébricas) que representa o comportamento do sistema formado pelo atuador elétrico, válvula e tubulação;
2. Simulação (através de métodos numéricos) das principais falhas que ocorrem no conjunto, utilizando o modelo matemático, tais como: a obstrução do obturador ou a degradação do sistema de engrenagens;
3. Determinação (através do método da regressão logística), da ferramenta de processamento de sinais mais adequada para o estudo de caso;
4. Desenvolvimento, embarque e prototipação de um sistema de detecção, classificação e predição de falhas do atuador através dos mapas auto-organizáveis ou filtros adaptativos voltado para a manutenção proativa;
5. Determinação (com base na área ocupada, no tempo de processamento e no desempenho dos circuitos), da ferramenta (MAO ou FA) mais adequada para o embarque.

1.6 ESTRUTURA

Esta tese de doutorado está estruturada da seguinte forma: no capítulo 2, a definição de manutenção, incluindo as suas principais estratégias e o estado da arte da manutenção (manutenção inteligente), entre outros tópicos, são abordados.

O desenvolvimento do modelo do atuador elétrico (atuador, válvula e tubulação), estudo de caso deste trabalho, é visto no capítulo 3.

A análise e os resultados obtidos a partir das ferramentas matemáticas (MAO e FA, por exemplo) são apresentados no capítulo 4.

A etapa de embarque das ferramentas matemáticas, incluindo uma breve descrição das arquiteturas, máquinas de estado e processos, é vista no capítulo 5.

No capítulo 6, são apresentadas as conclusões e perspectivas de continuidade do presente trabalho.

As principais técnicas de processamento de sinais, incluindo a TF, TFJ, TW e TWP, são analisadas a seguir no anexo A.

No anexo B, uma descrição sobre as seguintes técnicas de inteligência artificial: redes neurais artificiais, mapas auto-organizáveis e temporais de Kohonen, será apresentada.

Por fim, uma descrição sobre os filtros adaptativos, incluindo os principais algoritmos e a análise da estabilidade, é realizada no anexo C.

2 MANUTENÇÃO

Os sistemas ¹ à medida que são utilizados, ficam sujeitos a vários tipos de deterioração: desgaste, sujeira, corrosão, rachaduras e outras anomalias. Caso não sejam tomadas algumas medidas preventivas com a intenção de restaurar esses sistemas, os mesmos começarão a apresentar algum defeito. Ocorrendo o defeito, não sendo realizada nenhuma ação preventiva, os sistemas poderão falhar, ficando indisponíveis para desempenharem suas funções.

A manutenção consiste de uma série de medidas de prevenção (para manter os sistemas em funcionamento), correção (com o intuito de restabelecer os sistemas danificados), e predição (visando monitorar o estado de funcionamento do sistema, detectar alguma falha incipiente e fornecer subsídios para o planejamento de ações de intervenção, sejam estas preventivas ou corretivas) praticadas com a intenção de corrigir os danos provocados pela degradação e manter os sistemas funcionando.

Neste capítulo, é apresentada, inicialmente, na seção 2.1, a definição básica de alguns termos empregados na manutenção. Uma definição básica sobre a manutenção é apresentada a seguir na seção 2.2. As distintas estratégias de manutenção são abordadas na seção 2.3. O estado da arte da manutenção é visto na seção 2.4, bem como uma breve descrição sobre sistemas embarcados de manutenção. Por fim, as principais conclusões desse capítulo são apresentadas na seção 2.5.

2.1 TERMOS EMPREGADOS NA MANUTENÇÃO

Confiabilidade, falha e defeito são exemplos de termos que dizem respeito à manutenção cujas definições devem ser apresentadas para uma correta contextualização do tema.

2.1.1 Degradação

Degradação é o processo pelo qual a taxa de falha aumenta devido ao uso, exposição ao ambiente e passagem do tempo, conforme (ENDRENYI et al., 2001). A degradação do desempenho de sistemas é visto como resultado do envelhecimento e desgaste de peças que diminui a confiabilidade no desempenho e aumenta a probabilidade de ocorrência de falhas.

Um ambiente hostil, uma má operação e uso dos sistemas (a falta de lubrificação e a não substituição de peças desgastadas, por exemplo), defeitos de fabricação, aliados ao processo natural de envelhecimento e degradação, freqüentemente provocam um aumento da ocorrência de defeitos e falhas (TAN; HUO, 2005).

Assim, a degradação do desempenho é vista como um sinalizador de falhas nos sistemas e pode ser usada para predizer um desempenho inaceitável antes de esse ocorrer (DJURDJANOVIC; LEE; NI, 2003).

¹Por sistema entende-se: processos industriais, equipamentos, dispositivos, máquinas ou componentes.

Para modelar o processo de degradação de um sistema, é crucial determinar um indicador adequado que capture a evolução da degradação, a partir dos sinais desse sistema, tais como a evolução do valor RMS, ou da covariância ou do valor de pico-a-pico.

Mancais, rolamentos e engrenagens são exemplos de peças de máquinas que geralmente apresentam algum tipo de degradação e que já tiveram seus processos de degradação modelados, (HUANG et al., 2005), por exemplo. Geralmente, essas peças sofrem muitos desgastes mecânicos, atritos, excessos de carga e aumento de temperatura ao longo da sua vida útil.

2.1.2 Defeito

Toda alteração física ou química no estado de funcionamento de um sistema que não o impede de desempenhar sua função requerida (conjunto de ações de funcionamento para o qual o sistema foi projetado, fabricado ou instalado), podendo o mesmo operar com restrições, é considerado como defeito (NUNES, 2001).

Defeitos podem ocorrer provocando uma falha que pode ser catastrófica em alguns processos considerados críticos. Assim, os sistemas industriais devem apresentar características de tolerância às falhas, principalmente em aplicações desse tipo.

2.1.3 Falha

Falha é o estado de funcionamento de um sistema que impede o desempenho de sua função requerida e o leva invariavelmente à indisponibilidade (ENDRENYI et al., 2001).

Uma falha em um sistema dinâmico pode ser entendida como qualquer tipo de mau funcionamento em seus componentes que conduz a um desempenho não aceitável na realização de determinadas tarefas.

Falhas podem ocorrer abruptamente, provocando a parada das instalações, ou lentamente. Essas últimas, ocasionam uma perda gradual de desempenho no sistema e fadiga nos mecanismos sobrecarregados, ou seja, uma degradação.

Assim, uma falha é o término da capacidade de um sistema desempenhar a função requerida. Já um defeito permite que o sistema esteja disponível para uso.

Logo, caso sejam tomadas algumas medidas de manutenção enquanto o sistema está apenas apresentando algum defeito pode-se aumentar a vida útil desses, promover uma maior sustentabilidade, reutilizar peças e reduzir custos operacionais, antes da provável ocorrência de alguma falha.

2.1.4 Confiabilidade

Confiabilidade é a habilidade de um sistema de realizar sua função, sob condições específicas, de forma adequada, por certo período de tempo (IEEE-1, 1990). A confiabilidade também está associada à garantia de execução de funcionalidades sistêmicas para atender requisitos não-funcionais.

Segundo a Norma Brasileira Regulatória (NBR) 5462, de 1994, a confiabilidade de um item é a probabilidade de que esse item desempenhe a função requerida, por um intervalo de tempo estabelecido, sob condições definidas de uso.

Sellito (SELLITTO, 2005) comenta que a confiabilidade é a probabilidade que um produto ou serviço opere adequadamente e sem falhas sob as condições de projeto que foram estipuladas, durante um tempo especificado.

Desse modo, uma estratégia de manutenção centrada na confiabilidade é uma metodologia estruturada empenhada em permitir que um equipamento preencha as funções e os padrões de desempenho pretendidos no seu contexto operacional atual (ALKAIM, 2003).

2.1.5 Disponibilidade

O conceito de disponibilidade está relacionado ao quanto um sistema estará operacional e pronto para uso quando seu uso for requisitado (IEEE-1, 1990; CUNHA, 2002).

Um sistema com alta disponibilidade é capaz de manter os serviços para os quais foi programado durante o máximo tempo possível.

2.1.6 Tempo Médio Entre Falhas

O Tempo Médio Entre Falhas (TMEF) para um dado período na vida útil de um sistema é o valor médio dos intervalos de tempo entre falhas consecutivas sob determinadas condições.

O TMEF é usado para mensurar a confiabilidade de um sistema. Assim, quanto mais elevado o valor do TMEF, mais confiável será o sistema (IEEE-2, 1995).

2.1.7 Tempo Médio Para Reparo

O Tempo Médio para Reparo (TMR) para um determinado período de vida útil de um sistema é o tempo médio necessário para se realizar uma manutenção corretiva.

Esse intervalo de tempo é visto como uma medida da complexidade e da modularidade de um sistema. Logo, quanto maior o TMR, mais complexo será o sistema.

2.2 DEFINIÇÃO DE MANUTENÇÃO

Na literatura que trata da manutenção, encontram-se diversas definições e conceitos apresentados para o termo manutenção. A Associação Brasileira de Normas Técnicas (ABNT), em sua norma TB-116 de 1975, estabelece que a manutenção é o conjunto de todas as ações necessárias para que um item ou equipamento seja conservado ou restaurado de modo a permanecer de acordo com uma condição especificada.

Já a NBR 5462, de 1994, afirma que manutenção é a combinação de todas as ações técnicas e administrativas, incluindo as de supervisão, destinadas a manter ou recolocar um item em um estado no qual possa desempenhar uma função requerida.

A Associação Francesa de Normatização apresenta a manutenção como o conjunto de ações que permite conservar ou restabelecer um bem, dentro de um estado específico ou na medida para assegurar um determinado serviço (NUNES, 2001).

Segundo Marçal e Susin (MARÇAL; SUSIN, 2005) o processo de manutenção inclui todas as atividades técnicas e organizacionais que garantam que as máquinas e equipamentos em geral operem dentro da confiabilidade esperada.

A manutenção ainda pode ser definida como um conjunto de atividades realizadas em um sistema para preservar ou sustentar sua capacidade de fornecer serviço de um modo eficiente (KOTHAMASU et al., 2004).

Um dos propósitos da manutenção é estender o ciclo de vida do equipamento ou, pelo menos, aumentar o tempo médio para a ocorrência da próxima falha (TMEF) cuja reparação poderá ser menos onerosa (ENDRENYI et al., 2001; LI; BROWN, 2004).

No sentido literal da palavra, manter significa continuar em um estado existente; ou seja, a manutenção é o conjunto de técnicas de atuação para que os sistemas (itens, equipamentos, dispositivos e instalações, por exemplo) cumpram ou preservem suas funções específicas, mantendo o desempenho desejável, sem apresentar defeitos ou falhas.

Dois exemplos de falhas no estator de máquinas síncronas provocados por falta de manutenção e degradação de peças são vistos na Fig. 3. Na Fig. 3 a), a falha ocorrida foi a queima do estator, provocada pelo excesso de sujeira nos canais de ventilação, em virtude da falta de manutenção preventiva. Já na Fig. 3 b), a falha foi um curto entre as fases do estator, devido à deterioração do isolamento com o tempo e, também, ao excesso de sujeira.



Figura 3: Dois exemplos de falhas: a) queima e b) curto no estator (WEG, 2002).

De forma mais abrangente, o termo manutenção engloba os conceitos de prevenção (visando manter os sistemas em bom estado), correção (com a intenção de restabelecer os sistemas ao seu estado normal de funcionamento) e predição (para realizar uma predeterminação e quantificação dos níveis de desempenho e degradação).

Na pesquisa realizada, algumas das publicações, artigos e trabalhos relacionados à manutenção, concentram-se na definição do sistema de manutenção propriamente dito, analisando as variáveis intrinsecamente ligadas ao desempenho dos sistemas e as atividades de manutenção a serem executadas.

Inúmeras publicações encontradas na literatura, (YAN; LEE., 2005; TAN; HUO, 2005; TINÓS, 1999), abordam a inspeção, detecção, avaliação, predição e diagnóstico de falhas com a intenção de alcançar os melhores índices de disponibilidade e confiabilidade, por exemplo.

Altug, Chow e Trussell (ALTUG; CHOW; TRUSSELL, 1999) discutiram a importância da detecção e diagnóstico de falhas, oriundas da degradação gradual, de motores voltados para aplicações industriais, na manutenção da confiabilidade, por exemplo.

Huang (HUANG et al., 2005) analisou os defeitos presentes em rolamentos de máquinas rotativas usando métodos estatísticos para prever o tempo de vida desses equipamentos.

Outras publicações estão focadas nas políticas de manutenção, considerando a probabilidade de falha dos sistemas, com o objetivo de determinar o tempo ótimo de manutenção, aumentar a confiabilidade e disponibilidade dos sistemas e reduzir custos. Além disso, são abordados os aspectos de prevenção do estado de funcionamento e a recuperação das instalações, no caso da ocorrência de falhas ou defeitos, em (LEE et al., 2006), por exemplo.

2.3 ESTRATÉGIAS DE MANUTENÇÃO

Na literatura, pode-se encontrar distintas definições para as estratégias de manutenção. Conforme a ocorrência ou não de uma falha/defeito ou com base no funcionamento dos sistemas são definidas distintas estratégias de manutenção, por exemplo.

Contudo, as estratégias de manutenção tradicionalmente encontradas na literatura são a manutenção corretiva, preventiva, e preditiva.

Marçal (2000) apresentou uma tabela, reproduzida na Tab. 1, onde se pode observar um resumo das principais características das três estratégias clássicas de manutenção.

Tabela 1: Principais características das estratégias clássicas de manutenção.

Características	Estratégias		
	Manutenção Corretiva	Manutenção Preventiva	Manutenção Preditiva
Estado de operação da máquina/equipamento	Fora de serviço	Fora de serviço	Operando ou fora de serviço
Razão da interferência	Falha	Inspeção programada	Controle programado ou contínuo
Tarefas a serem executadas na máquina/equipamento	Reposição de peças/componentes	Desligamento para inspeção/reposição	Monitoramento
Objetivo da intervenção	Retorno ao trabalho	Manter o funcionamento	Predizer ou detectar

Contudo, pode-se encontrar algumas variações oriundas das modalidades clássicas de manutenção citadas anteriormente, além de outros tipos de classificação, vistos a seguir.

Para ilustrar as tradicionais estratégias de manutenção encontradas na literatura, (NUNES, 2001), por exemplo, apresentou um diagrama estruturado como um fluxo, onde com base na ocorrência ou não de uma falha, no conhecimento ou não da degradação e na periodicidade da supervisão, se define a forma de manutenção a ser adotada, conforme a Fig. 4.

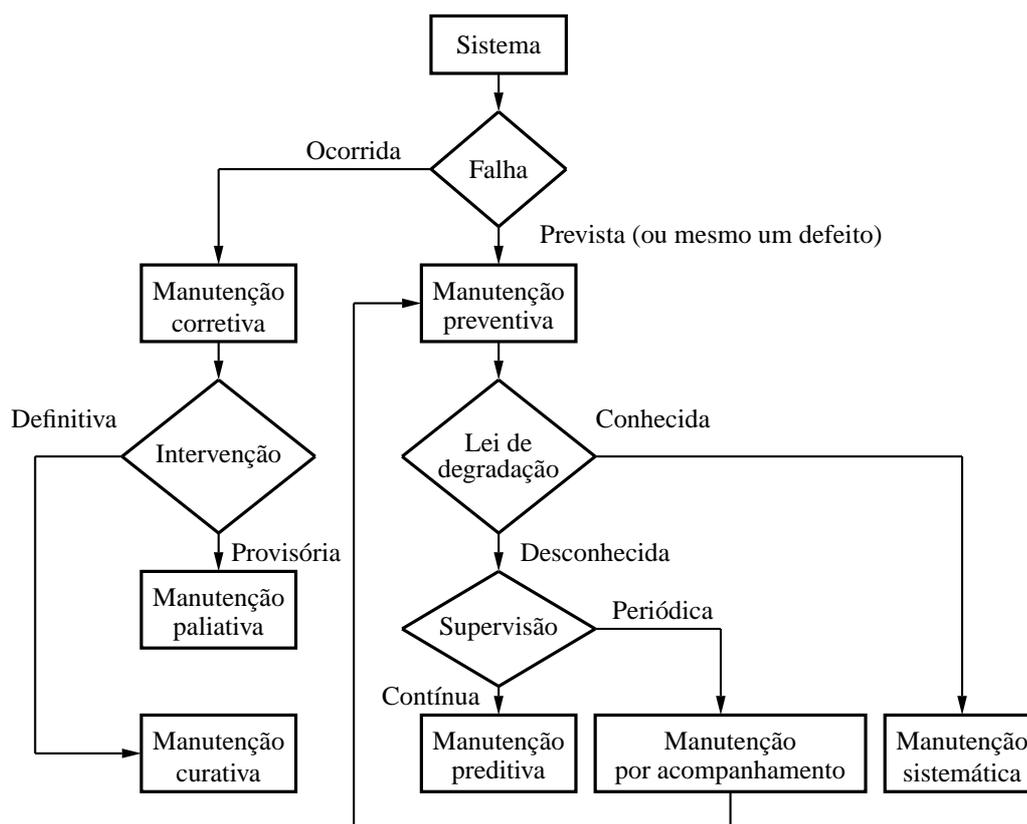


Figura 4: Estratégias de manutenção baseadas na ocorrência de falhas (NUNES, 2001).

Pode-se observar na Fig. 4 as estratégias clássicas de manutenção (corretiva, preventiva e preditiva) além de outras estratégias, como a preditiva e a sistemática (NUNES, 2001).

Pode-se observar, também, que a manutenção corretiva, que inclui ações paliativas (manutenção paliativa) e curativas (manutenção curativa), está relacionada com as situações de estado de falha onde o sistema, ou processo, já sofreu a falha.

Já a manutenção preventiva inclui ações de supervisão contínua (manutenção preditiva) e periódica (manutenção por acompanhamento) caso a lei de degradação seja desconhecida e ações sistemáticas (manutenção sistemática) antes da ocorrência de uma falha.

A manutenção sistemática é aquela que se aplica quando se conhece a lei de degradação dos sistemas. Caracteriza-se pela substituição das peças ou equipamentos dentro do tempo de durabilidade prevista pelo fabricante.

Quando a lei de degradação é desconhecida, a manutenção por acompanhamento é aquela realizada quando há uma supervisão periódica dos sistemas.

Por fim, a manutenção preditiva ocorre quando se aplica uma supervisão e controle contínuo dos sistemas com a intenção de reduzir a probabilidade de falhas e a degradação.

Uma outra classificação das estratégias de manutenção, onde essa é vista como uma parte do controle de ativos pode ser encontrada por (ENDRENYI et al., 2001). Nessa classificação, as quatro principais estratégias de manutenção são: a preditiva, a programada, aquela especificada pelo fabricante e aquela que consiste apenas de substituições de peças e equipamentos. Aparecem também, nessa classificação, a manutenção baseada em métodos empíricos, modelos matemáticos, na idade dos equipamentos e a manutenção centrada na confiabilidade.

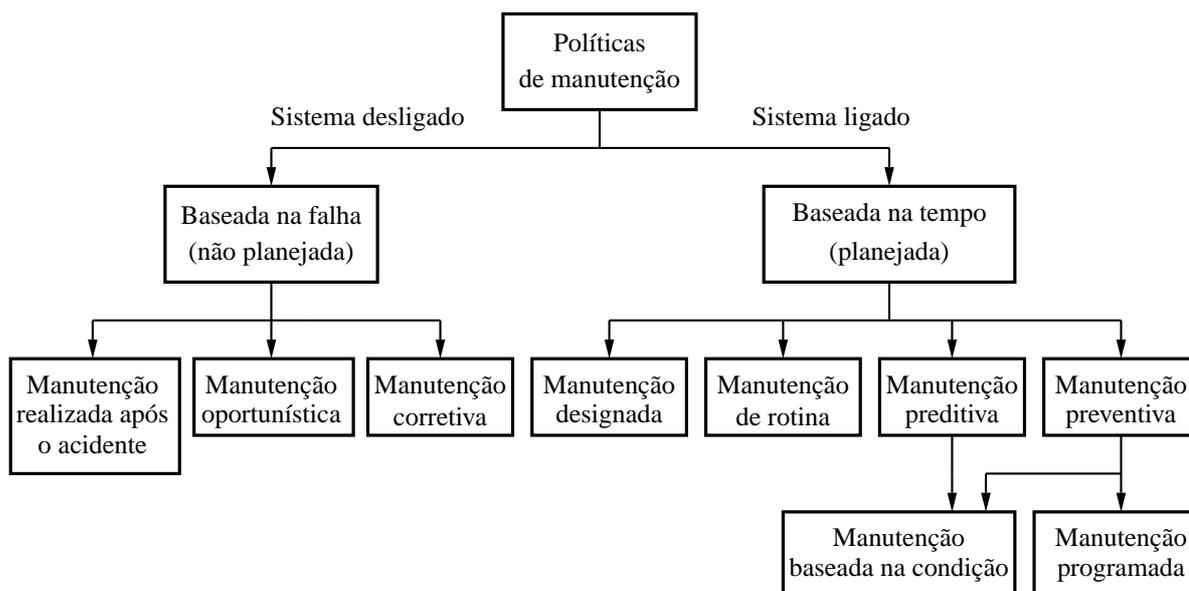


Figura 5: Estratégias de manutenção baseadas no funcionamento dos sistemas.

Um outro tipo de classificação para a manutenção foi apresentado por (SHIKARI; SADIWALA; DWIVEDI, 2004), onde as estratégias de manutenção foram divididas com base no funcionamento (ou não) dos sistemas, como visto na Fig. 5.

Assim, nessa classificação, as políticas de manutenção aparecem divididas em dois grupos principais: as baseadas na falha (não planejada) e as baseadas no tempo (planejadas).

No primeiro grupo, destaca-se a estratégia de manutenção realizada após o acidente, a oportunística e a corretiva.

Já no segundo grupo, a manutenção designada, a de rotina, a preditiva e a preventiva, são as principais estratégias de manutenção.

Na manutenção realizada após o acidente, apenas alguns ajustes, lubrificações, e trocas de componentes são feitos no sistema, por exemplo. Essa estratégia de manutenção é aplicada basicamente quando há poucos equipamentos e quando a falha não provocará riscos ambientais, graves prejuízos financeiros ou mesmo falhas em outros sistemas, por exemplo.

Na manutenção oportunística, as tarefas de manutenção são realizadas, na maioria das vezes, aproveitando a disponibilidade operativa do equipamento.

A manutenção de rotina, em especial, é a forma mais simples de manutenção planejada, como o nome sugere. Essa é realizada em intervalos regulares e envolve certas tarefas, tais como a limpeza, lubrificação e inspeção, por exemplo.

A classificação de manutenção adotada neste trabalho, vista na Fig. 6, comporta as três estratégias clássicas de manutenção: corretiva, preditiva e preventiva e mais a proativa.

Neste trabalho, estar-se-á particularmente interessado nas três estratégias clássicas de manutenção e, em especial, na estratégia proativa.

Em função dos recentes avanços da eletrônica, da computação e dos sistemas embarcados a manutenção proativa, em especial, também conhecida como manutenção inteligente, vem ganhando força em todo o mundo.

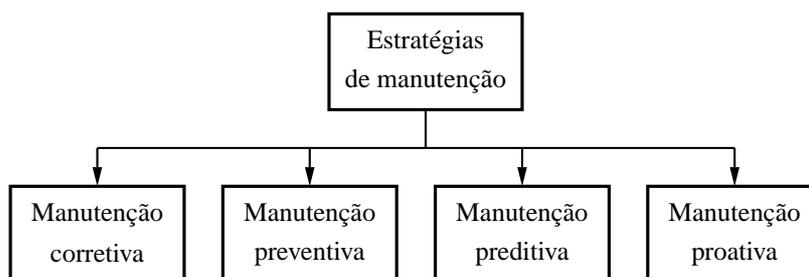


Figura 6: Classificação de manutenção adotada.

As grandes transformações experimentadas recentemente pelo setor eletrônico, em especial os sistemas embarcados, permitem um diagnóstico e análise muito maior das origens das falhas e dos seus efeitos, influenciando diretamente as atividades de manutenção, em especial, as preditivas e proativas (DJURDJANOVIC; LEE; NI, 2003).

Assim, uma maior explanação sobre as estratégias clássicas (corretiva, preventiva e preditiva) e sobre a manutenção proativa, será realizada nas sessões a seguir.

2.3.1 Manutenção Corretiva

Manutenção Corretiva (ou de reparação) corresponde à primeira aproximação de manutenção. Está relacionada com as ações praticadas depois da ocorrência e descoberta de um desarranjo, uma disfunção ou uma falha nos sistemas (MARÇAL, 2000).

Pode ser entendida como o trabalho de manutenção realizado após a falha do sistema, visando restabelecê-lo a sua função requerida, eliminando o estado de falha (NUNES, 2001).

A manutenção corretiva consiste da substituição de um, ou mais componentes, por um novo componente (ou tão bom como novo) que irá gerar um custo e resultar em uma limitada melhoria das condições do sistema (ENDRENYI et al., 2001; YAM et al., 2001).

Os tipos de manutenção corretiva são:

- Paliativa: compreende as intervenções corretivas executadas provisoriamente a fim de colocar o equipamento em funcionamento para, a seguir, efetuar o reparo definitivo.
- Curativa: compreende as ações típicas de reparo em caráter definitivo, a fim de restabelecer os itens à função requerida.

A principal vantagem ao se adotar esse tipo de manutenção é que essa não exige a realização de acompanhamentos ou inspeções nas máquinas.

Como exemplos de desvantagens, pode-se citar: apresenta custos elevados e conduz a paradas no sistema produtivo e a necessidade de um estoque de peças para reparos. Ainda, não é possível realizar um planejamento de custos ao se adotar essa política de manutenção, porque as falhas dos equipamentos e suas conseqüências não são previstas. Além disso, danos em outros equipamentos e paradas inaceitáveis no processo produtivo também poderão ocorrer.

Entretanto, quando o equipamento ou processo em questão não forem essenciais, ou mesmo quando o dano no equipamento é de fácil reparo, a manutenção corretiva é a mais indicada, devido ao seu baixo custo operacional (YAM et al., 2001).

Por fim, a manutenção corretiva geralmente é aplicada como complemento à manutenção preventiva, pois qualquer que seja a natureza ou nível de prevenção executada, sempre existirá um grupo de falhas que necessariamente irão exigir uma ação corretiva.

2.3.2 Manutenção Preventiva

Nas situações em que não se caracterizou um estado de falha nos sistemas a manutenção preventiva é aquela praticada com a intenção de reduzir a probabilidade de ocorrência de falhas ou danos.

Segundo (SHIKARI; SADIWALA; DWIVEDI, 2004), na manutenção preventiva as ações de manutenção são planejadas. Através de inspeções realizadas periodicamente se minimiza as paradas e a taxa de degradação, restaurando o sistema a sua função original.

Assim, essa estratégia de manutenção corresponde às ações previstas, preparadas ou programadas antes do provável aparecimento da falha, intervindo antes dos danos acontecerem.

Essas ações correspondem a um conjunto de práticas, tais como: lubrificação, limpeza, ajustes, e troca de peças ou itens que visam evitar as falhas.

A manutenção preventiva tem como intenção principal eliminar as quebras dos sistemas e as paradas no processo de produção provocadas por operações regulares de manutenção corretiva (DJURDJANOVIC et al., 2003).

A atividade de manutenção preventiva sistemática é aplicada quando a lei de degradação é conhecida. Essa lei diz respeito ao conhecimento sobre a evolução do desgaste do equipamento, na medida em que esse é utilizado.

Contudo, um inconveniente encontrado nesse tipo de estratégia são as constantes intervenções, muitas vezes desnecessárias. Em função das variações normalmente encontradas nos materiais e peças utilizadas, dois equipamentos dificilmente vão apresentar o mesmo comportamento ou o mesmo tempo de funcionamento antes da ocorrência de uma falha, ainda que eles sejam nominalmente idênticos.

Ainda, como a degradação depende do uso do equipamento (frequência de trabalho e aplicação, por exemplo), é possível que ocorra a substituição de uma peça que ainda teria muito tempo de vida útil.

Isso pode vir a reduzir a produtividade, diminuir a disponibilidade e elevar o custo total do processo em virtude das paradas excessivas.

Além disso, essa prática de manutenção provoca gastos desnecessários, porque mesmo a manutenção sendo realizada em intervalos periódicos, mantendo o sistema em um nível aceitável de funcionamento, paradas inesperadas provocadas por falhas súbitas podem acontecer antes de a manutenção ser executada.

Assim, essa estratégia de manutenção é realmente eficiente para os sistemas que sofrem algum tipo de degradação em um ritmo uniforme e para as quais os custos de uma falha sejam altos quando comparados aos custos da sua manutenção (LIMA; MARCORIN, 2003).

Hoje, essa estratégia é usada principalmente em conjunto com as estratégias de manutenção corretiva e de manutenção baseada na condição.

Nas distribuidoras elétricas, por exemplo, a prática de manutenção mais frequentemente é aquela realizada em intervalos fixos (isto é, manutenção preventiva) e eventualmente algumas correções adicionais são realizadas (ENDRENYI et al., 2001).

Caso a lei de degradação seja desconhecida, segundo (NUNES, 2001), a manutenção se subdivide em dois tipos: preditiva e por acompanhamento.

Normalmente, a manutenção por acompanhamento é baseada em datas predeterminadas (manutenção periódica) pela empresa, efetuando-se inspeções ou rondas periódicas.

Já a manutenção preditiva, é realizada com base nos parâmetros de valores de desempenho dos sistemas, executada no momento adequado e antes que se processe a falha do componente.

Na literatura, encontra-se tanto a denominação de técnicas preditivas, e não de manutenção preditiva, por se entender que essa forma de atuação estaria englobada pela manutenção preventiva; bem como a denominação de manutenção preditiva, sendo considerada com um outro tipo de manutenção.

Neste trabalho, a manutenção preditiva, será considerada como uma estratégia distinta de manutenção não englobada dentro da manutenção preventiva.

2.3.3 Manutenção Preditiva

A manutenção preditiva ocorre quando se aplica uma supervisão contínua dos parâmetros de controle e desempenho. Essa estratégia de manutenção tem como intenção estabelecer parâmetros para análise probabilística do comportamento dos sistemas e das leis de degradação.

Também são realizadas análises completas no caso de fracasso intermitente ou falha, fixando assim, metodologias para o uso de técnicas e tecnologias de medida e monitoração como também métodos avançados de análise sistêmica, confiabilidade e dados.

A manutenção preditiva tem como intenção aumentar a confiabilidade e a disponibilidade dos sistemas para reduzir o número de falhas e para prevenir paradas súbitas e inesperadas. Isto é particularmente interessante, visto que uma parada inesperada geralmente custa mais caro do que uma parada planejada.

A manutenção preditiva pode ser aperfeiçoada através do uso de indicadores satisfatórios de predição e disfunção das peças e equipamentos, e de uma análise de tendências ou de avaliações probabilísticas do estado de degradação de equipamentos, reduzindo assim a probabilidade de falhas e a degradação de componentes. Com base nesses indicadores estatísticos é possível prever a necessidade da manutenção antes de uma falha ocorrer e programar a parada dos equipamentos, ou linhas de produção, num momento adequado.

Desse modo, pode-se também prolongar a vida útil de equipamentos e instalações, realizar uma programação para aquisição de peças de reposição e reduzir os gastos com estoques e manutenção.

O custo de implantação dessa política de manutenção, normalmente, é bastante elevado, uma vez que implica na instalação da infra-estrutura necessária à monitoração dos equipamentos ou sistemas.

Contudo, ela se torna mais vantajosa quando aplicada nos equipamentos ou sistemas mais críticos do processo, para os quais o tempo para manutenção e/ou o custo de sua indisponibilidade são altos. Normalmente, faz-se um estudo inicial dos pontos críticos do processo a fim de determinar os dispositivos que deverão ser monitorados.

Quando se compara a manutenção preditiva com a manutenção corretiva, essa apresenta muitas vantagens, econômicas em especial, e vem se tornando a estratégia de manutenção mais recomendada para sistemas que necessitam de um alto grau de confiabilidade, como abordou (MATHEW et al., 2006).

Assim, nos últimos anos, a manutenção realizada em intervalos fixos tem sido substituída por esquemas de manutenção (preditiva) mais flexíveis baseado na análise das necessidades e prioridades, ou no estudo da informação obtida através do monitoramento, contínuo ou periódico, das condições dos equipamentos (ENDRENYI et al., 2001).

Geralmente, a manutenção preditiva é implantada juntamente com a manutenção preventiva, pois a monitoração ou acompanhamento periódico do desempenho e/ou deterioração das partes da máquina, são efetuados pela equipe de manutenção utilizando instrumentos portáteis, tais como sistemas de aquisição de dados.

Usar as estratégias de manutenção preditiva em conjunto com as preventivas tem se mostrado eficaz no reconhecimento de sintomas de falha eminente em máquinas. A finalidade dessa estratégia é aplicar a manutenção somente quando e se houver necessidade, através da determinação de um ponto ótimo para se executar a manutenção preventiva no equipamento.

Usar em conjunto essas duas estratégias de manutenção pode trazer algumas vantagens, como por exemplo:

- A vida útil das máquinas/equipamentos é ampliada;
- Previne-se de falhas e paradas no processo produtivo;
- Reduzem-se os estoques de peças para reparo;
- As reformas e substituições de peças são realizadas somente quando essas estão realmente comprometidas;
- Uma maior continuidade do funcionamento das máquinas é assegurada, pois os consertos são realizados em horários previamente programados.

Pode-se citar como características e requisitos para a operacionalização das modalidades de manutenção preventiva e preditiva:

- A necessidade de um plano de manutenção bem definido;
- A existência de uma equipe de profissionais especializados e treinados;
- A equipe de manutenção deve estar atualizada com as novas tecnologias;
- O acompanhamento e inspeções periódicas com uso de equipamentos adequados.

Ainda, diversas plataformas voltadas para a manutenção preditiva, tais como: *Integrated Condition Assessment System (ICAS)*, *Telé-Maintenance (TELMA)* e *Plateforme Reconfigurable pour l'Observation, pour les Télécommunications et les Usages Scientifiques (PROTEUS)* (que visam a integração dos sistemas de controle, monitoração e envio das informações) vêm sendo desenvolvidas (MULLER; MARQUEZ; IUNG, 2008).

2.3.4 Manutenção Proativa

Nos últimos anos, um novo paradigma de manutenção está emergindo. A estratégia tradicional de manutenção preventiva está sendo substituída pela manutenção proativa.

As práticas de manutenção proativa (ou o desenvolvimento de sistemas de manutenção inteligente) estão transformando as tradicionais técnicas focadas no conserto da falha para uma manutenção mais focada na prevenção e predição das falhas (YAN; LEE., 2005).

Na prática de manutenção preventiva as máquinas são mantidas em funcionamento graças a práticas rotineiras de trocas de peças, limpeza, entre outras atividades. Já na estratégia de manutenção preditiva são realizadas análises sistemáticas para prevenir, antecipar e diagnosticar a ocorrência de falhas.

A manutenção proativa é uma avanço em relação as estratégias de manutenção preventiva e preditiva, pois além de se aplicar todos os conceitos referentes a manutenção preditiva, como o monitoramento do equipamento em função do desgaste ocorrido, pode-se diagnosticar, e quantificar a perda de desempenho e atuar (de forma automática) sobre o sistema.

Assim, a manutenção proativa é aquela que envolve tarefas de monitoração do desgaste dos equipamentos ou processos, geração de diagnósticos, quantificação da perda de desempenho e atuação sobre o sistema, fornecendo informações importantes e úteis para a equipe de manutenção e alterando o padrão de comportamento dos equipamentos (quando detectada uma provável falha) até a realização da manutenção.

No caso de um falha ser detectada, o sistema pode automaticamente reconfigurar (atuar) no sistema de forma a continuar operando, ainda que de maneira degradada, até a ocorrência da manutenção.

Essa característica de reconfiguração permite, por exemplo, limitar o torque ou a corrente, de um motor com a intenção de reduzir o esforço sobre os componentes e assim prolongar o tempo de vida útil dos sistemas (LEE, 1996).

Como exemplo de possíveis vantagens alcançadas, ao se usar essa estratégia de manutenção, pode-se citar:

- Redução da equipe de manutenção;
- Diminuição das paradas do processo produtivo;
- Previsões precisas da vida útil do equipamento e das falhas.

Já como exemplo de possíveis desvantagens da manutenção proativa, cita-se:

- É necessário um plano de manutenção bem definido e rigoroso;
- Requer uma equipe de manutenção treinada com as novas tecnologias;
- Cada esquema (ou sistema) de manutenção desenvolvido é único e específico;
- A implantação requer um maior investimento de recursos e de tecnologia, quando comparado com as outras estratégias.

Ao se adotar a estratégia de manutenção proativa, a monitoração, o diagnóstico e prognóstico são extremamente necessários para se alcançar um desempenho onde o número de paradas tende a ser o mais próximo de zero possível (GEREMIA, 2001).

Essa monitoração, diagnóstico e predição de falhas podem ser alcançadas através do uso de sistemas embarcados. Através desses, é possível realizar análises, predições e diagnósticos de falhas em tempo real.

2.4 ESTADO DA ARTE NA ÁREA DE MANUTENÇÃO

No início do desenvolvimento da manutenção, essa era realizada somente quando a máquina enguiçava. Contudo, essa estratégia de manutenção tradicional conhecida como manutenção corretiva resulta em atrasos de operação, ociosidade do sistema e custos adicionais (materiais, mão de obra e serviços).

Uma série de estudos mostrou que o custo de manutenção poderia ser minimizado pela modificação do paradigma da substituição de peças de equipamentos após um defeito ou falha, por inspeções efetuadas em intervalos periódicos.

Além disso, a necessidade de tornar mais eficientes as ações de manutenção tem sido comentada em vários artigos científicos (LEE et al., 2006), (DJURDJANOVIC et al., 2003), (GRALL et al., 2002), por exemplo.

Vários estudos têm mostrado também que as técnicas tradicionais de manutenção corretiva e preventiva são onerosas, sendo que a maiorias dos custos são devidos as perdas de produtividade, e que necessitam de uma revisão (BENGTSSON, 2003).

Contudo, a maioria das técnicas de manutenção aplicadas hoje em dia na prática das indústrias ainda é reativa (consertando ou substituindo os equipamentos após as falhas) ou programada (assumindo-se certo nível de degradação do desempenho, realizando-se o serviço de manutenção, como uma rotina programada, sendo necessário ou não).

As estratégias preditiva e proativa, por sua vez, são aplicadas em menor número, em comparação a corretiva. Isso se deve, principalmente, ao fato de que estas estratégias são mais onerosas e voltadas para os equipamentos mais críticos dentro de um processo industrial.

Ambos os cenários de manutenção (corretiva e preventiva) são extremamente dispendiosos. Deve-se observar que as máquinas podem falhar de modo repentino, antes da ocorrência da operação de manutenção programada, colocando em risco os equipamentos e seres humanos envolvidos com o processo de produção.

Outro inconveniente encontrado na estratégia preventiva diz respeito as constantes intervenções, muitas vezes necessárias, em virtude das variações presentes nos materiais, peças, e frequência de uso dos equipamentos, reduzindo a produtividade e elevando o custo de produção e manutenção.

Aliado a isso, mesmo as atividades de manutenção preventiva que se repetem com baixa frequência (a cada 3 ou 5 anos, por exemplo), podem apresentar um alto impacto em custos, receitas, eficiência e capacidade de atendimento dos compromissos com clientes e fornecedores. Além disso, não devem ser tratadas como acontecimentos isolados e dissociados dos resultados globais, principalmente em sistemas produtivos complexos com altos níveis de integração e dependência entre os subsistemas que os compõem, (ROCHA, 2002).

Também, não há ainda, na maioria dos casos, medidores do processo de degradação implantados nas indústrias, que poderiam vir a sinalizar a possível ocorrência de uma falha (DJURDJANOVIC et al., 2003).

Além disso, a crescente complexidade dos processos industriais e de outros tipos de sistemas de engenharia tem modificado o âmbito de interesse da área de engenharia de controle. Anteriormente, essa área estava mais centrada no desenvolvimento de controladores baseados em critérios específicos, tais como a estabilidade e precisão.

Hoje, os modernos sistemas industriais exigem uma maior demanda de confiabilidade e segurança e um regime de operação com baixos custos (LO; WONG; RAD, 2006).

Porém, graças à evolução tecnológica, a informação sobre a degradação do desempenho e uma previsão de falhas em sistemas está mais acessível, tornando as atividades de manutenção mais inteligentes, confiáveis e menos dispendiosas.

Atualmente, sensores sofisticados e sistemas informatizados são capazes de fornecer dados importantes, avaliar com rapidez e precisão os indicadores de desempenho e prever o comportamento atual e futuro de máquinas e equipamentos.

Contudo, o monitoramento, a identificação e o diagnóstico de desgaste e falhas, bem como da evolução dos defeitos operacionais, exigem o desenvolvimento e a automatização de metodologias precisas, robustas, de resposta rápida e com um baixo custo operacional.

Recentemente, a evolução dos sistemas embarcados, o desenvolvimento dos sistemas de processamento de sinais, o avanço tecnológico dos sensores e algoritmos de simulação e a adição de inteligência artificial às técnicas de manutenção, estão levando a manutenção a passar por mudanças, sem precedentes, em sua filosofia; promovendo a criação e o desenvolvimento de sistemas de manutenção inteligentes.

Muitas ferramentas e métodos de predição vêm sendo desenvolvidos com bons resultados. Técnicas de fusão de sensores também estão sendo utilizadas devido à superioridade inerente em compartilhar informações de múltiplos sensores (LEE et al., 2006).

Uma grande variedade de técnicas de processamento de sinais e IA, que vem sendo aplicadas na manutenção, com a capacidade de diagnosticar uma anormalidade e calcular o tempo de vida útil de peças e equipamentos, tem sido relatada na literatura, a maioria deles desenvolvidos para aplicações específicas e não genéricas.

A transformada de Fourier e a transformada *wavelet*, por exemplo, têm sido utilizadas no processamento de sinais e na extração das características em conjunto com técnicas de inteligência artificial, como as redes neurais, (YAN; GAO, 2005).

A lógica nebulosa e as redes neurais vêm sendo usadas na predição e diagnóstico do desempenho de máquinas e equipamentos. Essas ferramentas permitem responder, por exemplo, quando uma máquina necessita de manutenção ou reparo.

Assim, a avaliação, diagnóstico e predição do desempenho de sistemas, realizadas através de algoritmos sofisticados e técnicas de IA possibilitam uma mudança no paradigma das tradicionais práticas de manutenção reativa, ou preventiva, para a prática proativa.

Esta adição de inteligência na manutenção torna possível a evolução das técnicas tradicionais de manutenção reativa, focadas no ajuste dos dispositivos, para as metodologias de predição e prevenção, mais centradas na precisão da informação (LEE et al., 2006).

2.4.1 Sistemas de Manutenção Inteligentes

As atividades de manutenção geralmente envolvem a determinação do que precisa ser reparado em um sistema e de quando será necessária essa reparação. Essa determinação baseia-se numa combinação de experiência e inteligência, ambas necessárias para determinar e interpretar as informações disponíveis (SHIKARI; SADIWALA; DWIVEDI, 2004).

Um diagnóstico específico de qualquer falha é uma tarefa crítica e particularmente difícil em ambiente repleto de máquinas e complexos processos. Um técnico de manutenção qualificado usará sua inteligência e experiência (além de dados de fabricantes e manuais) para realizar um diagnóstico na ocorrência de alguma falha.

A maioria dos sistemas de manutenção utiliza pacotes gráficos de rotinas que são responsáveis pela emissão de ordens de trabalho. Outros sistemas de manutenção são capazes de detectar a existência de um defeito, mas não são capazes de determinar o que é necessário fazer para corrigi-lo.

Esses fatores evidenciam que é necessário captar o conhecimento e experiência de especialistas de um determinado sistema ou área, e torná-los disponíveis aos funcionários, usuários, ou sistemas de manutenção para a realização de um diagnóstico correto e de uma tomada de ação no caso da ocorrência de algum defeito.

Sistemas de manutenção baseados no conhecimento, ou sistemas de manutenção inteligentes, são sistemas concebidos precisamente para se atingir esses objetivos. Esses sistemas fornecem as ferramentas e técnicas para preservar as competências e experiências dos técnicos ou especialistas em programas de computador.

A base da construção de sistemas inteligentes consiste na captura do conhecimento sob a forma de regras. O conjunto de todas as regras irá formar um sistema, mais complexo, baseado nessas regras.

Os sistemas inteligentes são formados por muitas regras. Teoricamente, essas regras podem ser escritas em qualquer linguagem de programação que fará uso delas para alcançar algum resultado e realizar alguma ação. É claro, é necessário ter cuidado na determinação de quais serão as regras e na associação com as falhas.

Outro aspecto fundamental dos sistemas inteligentes é que eles expressam o conhecimento também de um modo qualitativo, e não somente quantitativo tornando mais fácil a leitura e a compreensão.

Um sistema inteligente completo é composto, além do conjunto de regras que descrevem o conhecimento, de um motor de inferência (parte do programa do computador que manipula o conjunto de regras). O motor de inferência compreende o processo onde as regras são utilizadas para se alcançar uma possível conclusão.

Um sistema inteligente é capaz de analisar o padrão de vibração de uma máquina de indução, de um atuador ou de uma prensa, como exemplo, realizar um diagnóstico automático do motivo da falha e determinar se a falha em questão foi nos rolamentos, ou se foi devido a desalinhamentos, por exemplo, (SHIKARI; SADIWALA; DWIVEDI, 2004; SINGH; KAZZAZ, 2003).

Como exemplo, cita-se o desenvolvimento de um sistema inteligente de supervisão, controle e diagnóstico de falhas de um sistema composto por tanques e válvulas baseado em lógica nebulosa e algoritmos genéticos apresentado por (LO; WONG; RAD, 2006). Os resultados obtidos nesse trabalho demonstraram que é possível controlar o sistema de tanques e válvulas, usando o sistema proposto, com êxito.

Ainda, a implantação de sistemas de manutenção baseados em microprocessadores tem aumentado, pois esses permitem que se realize uma análise contínua das máquinas para prever possíveis condições de falhas (ALTUG; CHOW; TRUSSELL, 1999).

Dentre as principais vantagens dos sistemas inteligentes, pode-se citar o fato de que esses são capazes de monitorar continuamente as condições de operação de um sistema, tomar decisões, efetuar ações de manutenção e fornecer diagnósticos precisos sobre as condições atuais e futuras dos equipamentos, tais como uma classificação de falhas, por exemplo.

Já como exemplo de aspectos negativos dos sistemas de manutenção inteligentes que devem ser considerados, cita-se:

- O processo de modelagem dos sistemas pode não ser completo, não cobrindo todas as regras lógicas do sistema;
- Esses sistemas, normalmente, não incorporam análises econômicas na etapa de tomada de decisões;
- Uma vez que é relativamente simples fazer alterações na base de conhecimento, também fica fácil a introdução de possíveis erros nesses sistemas, tais como a definição de regras ou associação com as falhas de modo incorreto.
- As falhas a serem consideradas já devem ter sido observadas e analisadas previamente. Contudo, falhas que não foram observadas poderão ocorrer.

Assim, há ainda uma série de questões por resolver quanto a elaboração de sistemas inteligentes de manutenção, por exemplo: uma questão existente na construção de sistemas inteligentes eficazes e robustos é como aprender a questionar os técnicos para se obter a melhor resposta para as diversas situações que irão surgir ao longo do projeto.

2.4.2 Sistemas de Manutenção Inteligentes Embarcados

Tipicamente, sistemas embarcados são construídos a partir de componentes do tipo processador, memória, dispositivos programáveis, eletrônica de condicionamento de sinal, sensores, atuadores e, mais recentemente, de blocos criptográficos para agregar segurança às informações trocadas através da rede em sistemas distribuídos (CARRO, 2001).

Os sistemas embarcados são sistemas computacionais complexos, por envolver conceitos até agora pouco analisados pela computação de propósitos gerais, tais como as questões da portabilidade e do limite de consumo de potência sem perda de desempenho, a baixa capacidade de memória e a necessidade de confiabilidade, por exemplo.

Esses sistemas computacionais estão presentes em vários equipamentos eletrônicos, como por exemplo, nos telefones celulares com máquina fotográfica e agenda, nos sistemas de controle dos carros e ônibus e nos computadores portáteis.

Um dispositivo com um sistema de manutenção embarcado é um dispositivo projetado com um sistema embutido de tecnologia e equipamentos capazes de recolher informações, processar dados e atuar sobre o equipamento a ele conectado.

Desse modo, os sistemas de manutenção embarcados podem ser usados para uma análise regular das variáveis e comportamento das máquinas, sendo possível assim realizar uma predição e um diagnóstico em tempo real das possíveis falhas (ALTUG; CHOW; TRUSSELL, 1999).

Devido a essas características específicas dos sistemas embarcados, tais como a capacidade de trabalhar em tempo real, muitas empresas que desenvolvem sistemas de gestão de manutenção estão implantando seus sistemas de modo embarcado (SU et al., 2005).

Os autores em (SU et al., 2005) apresentaram um sistema de prognóstico inteligente composto de Dispositivos Genéricos Embarcados (DGE) usados para implantar esquemas de manutenção preditiva e de prognóstico da qualidade, além de um servidor remoto. Nesse trabalho, os DGE foram projetados para considerar todas as interfaces necessárias para diagnóstico em uma indústria de semicondutores e apresentavam uma interface aberta, permitindo a conexão com qualquer aplicação.

2.4.3 Metodologia de Desenvolvimento de Sistemas de Manutenção Inteligentes

A metodologia aqui utilizada para se projetar um sistema de manutenção inteligente (SMI) pode ser aplicada no âmbito de processos, sistemas e equipamentos.

Inicialmente, deve-se verificar se existe disponível um histórico com dados de manutenção. Caso exista, pode-se aplicar alguma técnica quantitativa para determinação dos componentes mais críticos (plotar gráfico que correlaciona a frequência de falhas com o custo de falhas, por exemplo). Caso não exista, é necessário o uso de métodos qualitativos: diagramas de causa-efeito e tabelas FMEA/FMECA, Fig. 7, por exemplo.

A FMEA (do inglês *Failure Mode and Effects Analysis*) é uma metodologia desenvolvida para identificar possíveis modos de falha em produtos, processos ou sistemas, associando a esses modos de falha um risco. A partir da priorização desses riscos em uma tabela, ações corretivas de manutenção são definidas e implantadas (VAZ et al., 2006; RIGONI; PELOW; SILVEIRA, 2004; VOLLERTT, 1996).

A FMECA (*Failure Mode, Effects, and Criticality Analysis*) é uma metodologia semelhante à FMEA, mas que acrescenta ao método uma análise de criticidade da falha, em termos de severidade, dificuldade de detecção e probabilidade de ocorrência da falha.

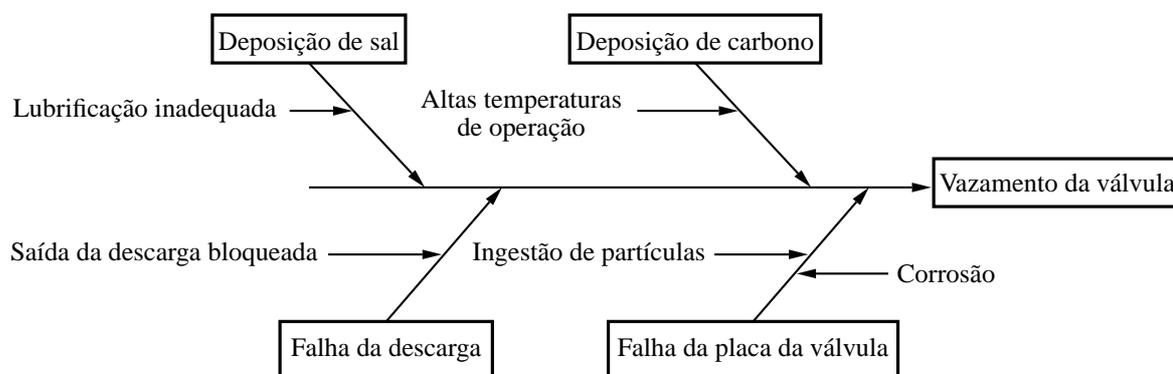


Figura 7: Exemplo de diagrama de causa-efeito.

A FMEA e FMECA são técnicas de análise de falhas que foram desenvolvidas para serem aplicadas principalmente aos componentes de hardware com a intenção de analisar as maneiras pelas quais esses componentes podem vir a falhar e quais os efeitos dessas falhas.

Assim, para se desenvolver um SMI, na abrangência dos processos (indústrias, usinas, represas, entre outros) deve-se:

1. Levantar o histórico de manutenção;
2. Identificar os sistemas mais críticos no processo industrial;
3. Determinar a frequência de ocorrência de falhas;
4. Obter o tempo desperdiçado para cada sistema na ocorrência de uma falha;
5. Selecionar os sistemas que mais falham e os que desperdiçam mais tempo;
6. Analisar os modos de falha do sistema;
7. Determinar as variáveis e sinais de interesse.

Após a identificação dos sistemas mais críticos, parte-se para a análise de quais são os equipamentos que mais falham. Com isso, a partir do gráfico de frequência de falhas versus o tempo desperdiçado nas operações de manutenção, pode-se identificar quais os equipamentos que mais necessitam de manutenção.

A Fig. 8 exibe um exemplo deste tipo de gráfico, cedido pelo pessoal do *Intelligent Maintenance Systems* (IMS, 2010).

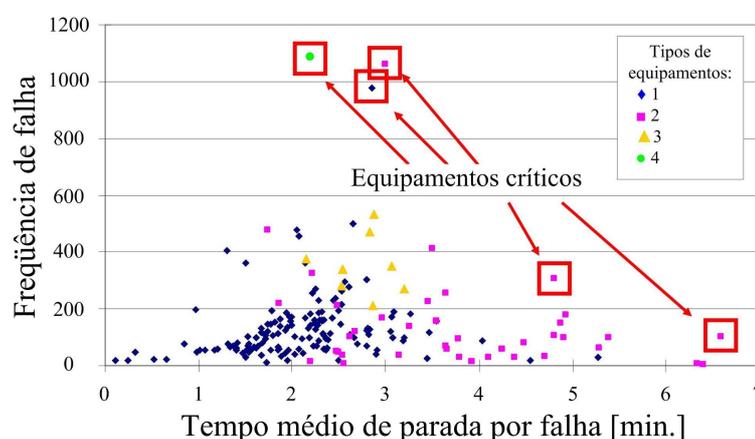


Figura 8: Exemplo de gráfico de frequência de falha versus tempo desperdiçado.

Os equipamentos mais críticos, assinalados na Fig. 8, são aqueles que possuem uma frequência de falha muito alta (isto é, falham muito) ou apresentam um tempo médio de parada por falha muito elevado (ou seja, é desperdiçado muito tempo para realizar a manutenção nesses equipamentos).

Assim, a metodologia para se projetar um sistema de manutenção inteligente aplicado no âmbito de sistemas (grupo gerador, robô, linha de produção, etc.) inclui os seguintes procedimentos:

1. Determinar, para os sistemas mais críticos, os equipamentos/máquinas mais críticos;
2. Levantar a frequência de ocorrência de falhas e o tempo desperdiçado para cada equipamento na ocorrência de alguma falha;
3. Selecionar os equipamentos que apresentam a maior frequência de falhas e o maior tempo desperdiçado na manutenção caso ocorra uma falha.

Tendo identificado quais são os equipamentos mais críticos, segue-se para a determinação de quais são as falhas que mais ocorrem. Assim, no âmbito dos equipamentos (gerador síncrono, pinça, esteira, por exemplo), pode-se, para cada equipamento: levantar quais são as principais falhas e determinar os sinais que devem/podem ser extraídos do equipamento.

A seguir, deve-se determinar quais são os sensores que serão instalados nos equipamentos, quais condicionadores de sinais serão usados, qual o tempo de amostragem necessário, entre outras questões.

Por fim, as ferramentas matemáticas (ferramentas de processamento de sinais, compressão de dados, ou de inteligência artificial, por exemplo) e computacionais (sistemas supervisórios, ambientes operacionais, ferramentas matemáticas, etc.) necessárias para a implantação do SMI devem ser escolhidas.

É nesta última etapa da metodologia do desenvolvimento de SMI que o trabalho aqui apresentado estará focado. Ao longo dos próximos capítulos as principais ferramentas de processamento de sinais, de inteligência artificial e filtragem adaptativa serão apresentadas.

Essa metodologia permite responder, por exemplo: quando um processo ou equipamento observado irá apresentar uma falha, quando a degradação irá atingir um nível inaceitável, por que ocorreu redução do desempenho, qual é a causa da degradação ou qual é o processo ou equipamento mais crítico que mais necessita de manutenção.

2.5 CONCLUSÕES DO CAPÍTULO

Ao longo deste capítulo a definição de manutenção e as suas diferentes estratégias (preventiva, preditiva ou corretiva), de acordo com a ocorrência ou não de uma falha ou defeito, foram abordadas.

Ressaltou-se que as técnicas de manutenção preditiva e proativa, estão se tornando importantes devido à crescente globalização da economia e evolução tecnológica. O avanço das técnicas de manutenção preditiva e proativa, que utilizam técnicas de processamento de sinais e inteligência artificial, está sendo observado. Essas tecnologias e ferramentas permitem analisar as condições de operação dos equipamentos de modo a determinar se esses estão funcionando em condições normais de operação ou não, por exemplo.

Por fim, o estado da arte da manutenção e os sistemas de manutenção inteligentes, que tornam possível a avaliação, diagnóstico e predição do desempenho de máquinas e equipamentos enquanto esses estão em funcionamento, também foram discutidos.

3 DESENVOLVIMENTO DO MODELO DO ATUADOR

Um dos objetivos deste trabalho é efetuar uma classificação das principais falhas que ocorrem em um atuador elétrico por meio de ferramentas de inteligência artificial ou filtragem adaptativa. Para tal, é necessário obter um conjunto de dados que represente as condições de operação do atuador. Estes dados serão utilizados para treinar e testar as ferramentas de inteligência artificial e filtragem adaptativa.

No conjunto de dados, deverão estar contidos os dados que representam as condições de operação normal, de degradação e de falha (para cada falha que se deseja classificar) do atuador elétrico.

Há dois modos de se obter estes conjuntos de dados, através de ensaios ou de simulações. Os ensaios, bem como as simulações, procuram reproduzir as mais diversas situações de comportamento do sistema.

Entretanto, em alguns tipos de sistemas, nem todas as situações de falhas podem ser ensaiadas sem que haja um dano ou quebra de peças. Além disso, geralmente o tempo e custo necessários para a realização destes ensaios são inviáveis, impossibilitando a sua execução. Nestes casos, a simulação do sistema, através de um modelo matemático, pode ser a melhor escolha.

De modo simplificado, o processo de construção de um modelo de um sistema, ou de um fenômeno físico, pode ser dividido em três etapas principais: a definição, a solução e a validação do modelo (HELLING; SCHOOR; HELBERG, 2004).

Assim, a modelagem matemática de sistemas é vista como a atividade de construir modelos que expliquem, expressem ou representem as características ou o comportamento de um sistema e de obter, através de métodos matemáticos e computacionais, a solução e a validação desses modelos.

Neste trabalho, optou-se por desenvolver um modelo matemático do atuador elétrico para simular corretamente o comportamento desse atuador nas mais variadas situações. Para tal, foi necessário conhecer o comportamento de cada componente do conjunto formado pelo atuador, válvula e tubulação. Após, foi realizada a representação do sistema, como um todo, em um sistema de equações algébricas e diferenciais não-lineares. Por fim, a solução do sistema de equações não-lineares (através de métodos numéricos) e a etapa de injeção de falhas no modelo foram realizadas.

Assim, neste capítulo, uma descrição do estudo de caso analisado neste trabalho, incluindo as principais partes e forças envolvidas na operação do atuador elétrico, é apresentada inicialmente na seção 3.1. O desenvolvimento do modelo matemático do estudo de caso (composto por uma máquina assíncrona, sistema de engrenagens, válvula e tubulação) é visto na seção 3.2. A etapa de injeção de falhas (tais como o aumento do atrito ou o desgaste das engrenagens) no modelo do atuador é descrita na seção 3.3. Por fim, na seção 3.4 são apresentadas as principais conclusões deste capítulo.

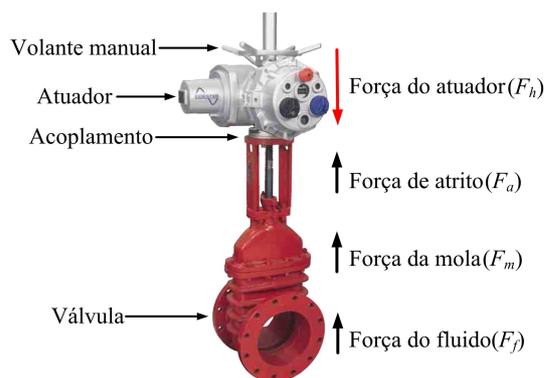


Figura 9: Estudo de caso: atuador elétrico, válvula e tubulação.

3.1 ESTUDO DE CASO

Um modelo de atuador (CSR25) da empresa Coester Automação S.A. e uma válvula tipo gaveta estão sendo utilizados como estudo de caso neste trabalho. Na Fig. 9 pode-se observar as principais partes do estudo de caso: atuador e válvula.

Atuadores, elétricos ou pneumáticos, são equipamentos que permitem a motorização de válvulas, *dampers*, comportas e outros equipamentos similares. O atuador pode ser acoplado à válvula, por exemplo, através de unidades de adaptação ou redutores e sua função é movimentar a haste de uma válvula, ou de um comporta, abrindo-a ou fechando-a.

As válvulas, em especial, são equipamentos cuja finalidade principal é o controle do fluxo de água, esgoto, ou petróleo, entre outros fluidos. Existem diversos tipos de válvulas: gaveta, esfera, ou globo, por exemplo, cada uma indicada para um tipo de aplicação em particular (CAMPOS; TEIXEIRA, 2006; HELLING; SCHOOR; HELBERG, 2004).

Neste trabalho, o conjunto formado pelo atuador elétrico, válvula gaveta e tubulação é utilizado para controlar o fluxo de petróleo nas redes de distribuição da empresa Petrobrás.

O conjunto apresenta-se instrumentado, sendo possível monitorar o torque mecânico exercido pelo atuador nas engrenagens que comandam a abertura e fechamento da válvula e a posição em que se encontra o obturador da válvula. Estas são as duas principais variáveis usadas, até o momento, para a implantação do sistema de manutenção inteligente embarcado.

Os dados de torque, simulados, visam imitar os dados oriundos de uma célula de carga que mede o contra torque exercido pelo atuador. Já os dados de posição de abertura e fechamento, reproduzidos pelo modelo, correspondem aos valores de um potenciômetro que mede o percentual de abertura ou fechamento da válvula conectada ao atuador.

Através do sistema de engrenagens do atuador, existe a transferência do esforço sofrido sobre todo conjunto, para a célula de carga. Por sua vez, a célula ao se deformar, envia um sinal analógico proporcional ao esforço convertido à placa controladora que processa os dados enviados. O programa de controle avalia o valor do esforço e verifica se deve desligar o motor e gerar um alarme, por exemplo. Os dados de esforço e posição durante o movimento ficam registrados em uma memória para posterior análise e visualização.

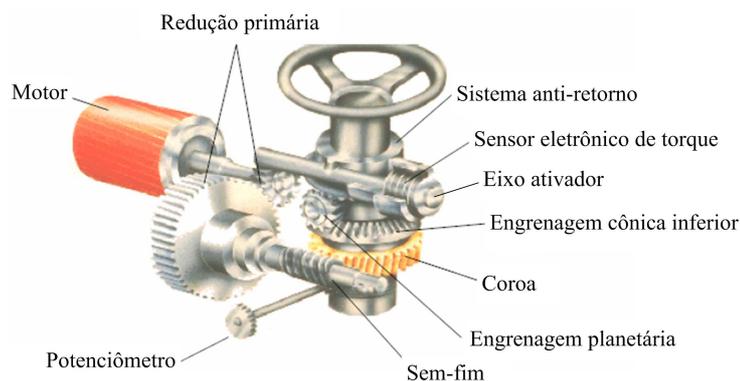


Figura 10: Principais partes do atuador elétrico.

O atuador é composto por um conjunto elétrico e mecânico. Dentre as principais partes desse conjunto, visto na Fig. 10, pode se destacar:

- Motor elétrico: motor de indução trifásico com gaiola de esquilo;
- Cadeia mecânica: sistema de transmissão mecânica;
- Acoplamento de saída: manga com chaveta para hastes rotativas;
- Sensor eletrônico de torque: célula de carga;
- Sensor eletrônico de posição e movimento: potenciômetro de alta precisão.

O sistema de transmissão mecânica apresenta três tipos de reduções cinemáticas, formadas por quatro tipos de engrenagens: uma de dentes retos, um par coroa sem-fim e um conjunto de engrenagens helicoidais e planetárias (formando um sistema de transmissão diferencial).

3.1.1 Operação do Atuador Elétrico

O motor de indução, responsável pelo movimento da gaveta da válvula, aciona o primeiro par de engrenagens (de dentes retos), nas quais se processa o que se chama de redução primária, ver Fig. 10.

O movimento continua no par coroa/sem-fim que aciona uma engrenagem cônica inferior, responsável pelo movimento das engrenagens planetárias ligadas à manga de saída.

O sistema de engrenagens diferenciais permite giro independente das rodas cônicas maiores, o que, associado ao sistema anti-retorno, impede que o volante seja movimentado através do motor, mas permite que o mesmo esteja sempre disponível para movimento partindo dele.

O movimento da manga de saída, ligado à válvula, sente maior ou menor resistência, transferindo uma carga correspondente para o mecanismo anti-retorno, ao qual está associado o eixo de comando do sensor eletrônico de torque, conhecido como eixo ativador.

O eixo ativador trabalha sempre sob compressão, deformando os *strain gages* ligados a ele, informando para a placa controladora o valor da deformação, que é processado como valor correspondente de torque.

Além disso, geralmente está associada à manga de saída uma tomada de movimento que, devidamente reduzida, produz o giro do eixo do sistema do sensor de posição (potenciômetro).

3.1.2 Forças Envolvidas na Operação do Atuador Elétrico

Sobre o comportamento do atuador, válvula e tubulação pode-se notar que existe uma parte estática que está relacionada com a determinação da abertura (ou posição da gaveta) associada com a vazão para essa abertura e outra dinâmica que corresponde à transmissão do torque para que ocorra o movimento do obturador da válvula até a posição determinada e por conseguinte a vazão para o processo.

No conjunto formado pelo atuador, válvula e tubulação há uma série de forças que atuam para que ocorra o movimento de abertura ou fechamento do obturador da válvula, como visto na Fig.9, onde a força do atuador (F_h) é aquela transmitida pelo motor através do sistema de engrenagens. Estas forças foram levadas em consideração na hora da modelagem do conjunto.

Além da força do atuador, há a força de atrito (F_a), que corresponde ao atrito existente entre a haste e a gaxeta (peça usada para impedir vazamentos), a força (F_m) da mola (a mola serve para fazer com que a gaveta da válvula retorne a sua posição inicial, quando na ausência de alimentação) e a força do fluido (F_f) que corre pela tubulação.

3.2 MODELO DO ESTUDO DE CASO

Um dos principais objetivos deste trabalho é efetuar uma classificação das principais falhas que ocorrem no atuador elétrico. Para se realizar essa classificação, um conjunto de dados representativo do sistema faz-se necessário.

Nesse conjunto de dados deverão estar contidos os dados que correspondem às condições normais, de degradação e de falhas. Estes dados podem ser coletados, por exemplo, após a realização de uma série de ensaios onde se procura imitar as situações normais e de falha do sistema. Outra opção é simular as diversas situações de operação do sistema a partir de modelos formados por equações matemáticas.

Em alguns sistemas, nem todas as situações de falhas podem ser ensaiadas sem que haja uma degradação ou quebra das peças. Além disso, geralmente o tempo e custo para a realização destes ensaios são grandes, inviabilizando a sua execução. Neste caso, um modelo matemático é a melhor escolha (opção adotada neste trabalho).

O processo de modelagem computacional de um sistema, ou de um fenômeno físico, geralmente, é dividido em três etapas fundamentais: a definição, a solução e a validação do modelo (HELLING; SCHOOR; HELBERG, 2004).

A primeira etapa consiste na identificação dos fatores que influenciam de maneira relevante o sistema. Isto implica na escolha adequada dos princípios físicos, das considerações e simplificações tomadas e da definição das variáveis que descrevem o sistema, resultando em um modelo matemático formado por equações diferenciais e algébricas não-lineares.

A segunda etapa do processo de modelagem consiste em obter a solução do modelo matemático. Esta tarefa é geralmente atribuída aos métodos numéricos, tais como o método de Runge Kutta ou de Newton.

Por fim, na etapa de validação, os resultados do modelo são comparados com os valores obtidos a partir do sistema real sob as mesmas condições de operação; ou seja, é realizado um cruzamento dos resultados de simulação com os resultados experimentais.

3.2.1 Definição do Modelo do Estudo de Caso

Optou-se por desenvolver um modelo matemático do sistema para simular corretamente o comportamento do atuador elétrico e da válvula nas mais variadas situações. Para tal, foi necessário conhecer o comportamento de cada componente do conjunto formado pelo atuador, válvula e tubulação.

A representação do sistema, como um todo, foi obtida a partir das equações oriundas da modelagem isolada das partes do sistema. Isso foi realizado conectando as equações correspondentes a cada um dos componentes; ou seja, identificando as variáveis comuns a mais de um componente e representando esta inter-relação em um único sistema de equações algébricas e diferenciais.

Deste modo, as principais partes consideradas para a construção deste modelo foram:

- O motor elétrico;
- O sistema de engrenagens;
- O fluxo de líquido pela válvula e tubulação;
- A posição, velocidade e aceleração da haste;
- As forças envolvidas no movimento do motor, das engrenagens e da gaveta da válvula.

O desenvolvimento do modelo do motor elétrico (máquina assíncrona), da válvula e tubulação e do sistema de engrenagens do atuador elétrico, que é o estudo de caso considerado neste trabalho, fez-se necessário também na realização da etapa de injeção de falhas.

3.2.2 Modelo da Máquina Assíncrona

A máquina assíncrona, ou simplesmente motor de indução, é o tipo de motor de corrente alternada mais comum de todos os motores. Possui uma construção simples e resistente e apresenta boas características de funcionamento. O tipo mais importante de motor de indução polifásico é o motor trifásico (KUNDUR, 1994; FITZGERALD, 1990; NASAR, 1984).

No processo de modelagem da máquina assíncrona, inicialmente, obtém-se as equações de tensão e fluxo de rotor e do estator tendo em mente as referências formadas pelos eixos magnéticos das fases do estator e do rotor.

Usualmente, denomina-se o conjunto formado por estes eixos como sistemas de coordenadas abc e o modelo de máquina síncrona obtida nesse sistema como modelo de uma máquina síncrona no sistema de coordenadas abc ou, simplesmente, modelo abc , como visto na Fig. 11.

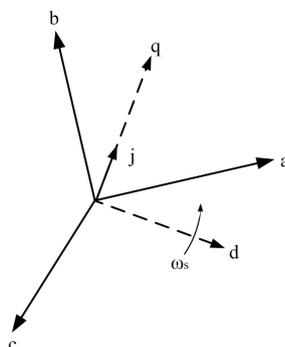


Figura 11: Sistema de coordenadas.

Após a definição das equações da máquina assíncrona no sistema de coordenadas abc , é realizada uma mudança de variáveis nas equações do estator e rotor para um outro sistema de coordenadas formado pelos eixos direto (d) e em quadratura (q), ou sistema $dq0$, obtendo-se assim as equações do modelo $dq0$ do motor assíncrono. Destas equações, adquire-se distintos modelos, que se diferenciam a partir da consideração ou não de determinadas constantes de tempo, reatâncias e tensões, específicas para cada período de tempo.

3.2.2.1 Considerações para a Máquina Assíncrona

No desenvolvimento das equações da máquina assíncrona é usual considerar-se hipóteses simplificadoras. A partir destas considerações desenvolvem-se as equações de tensão e fluxo de estator e do rotor e faz-se o levantamento das equações mecânicas e elétricas.

Dentre as considerações típicas da análise de máquinas assíncronas, destaca-se:

- O circuito magnético é linear;
- As perdas magnéticas e mecânicas são desprezadas;
- O fluxo de potência elétrica está entrando na máquina assíncrona;
- O campo magnético no entreferro do motor é distribuído senoidalmente;
- Há uma simetria elétrica e espacial entre as fases dos enrolamentos do estator e rotor.

3.2.2.2 Equações Mecânicas da Máquina Assíncrona

Semelhante à máquina de corrente contínua (CC), a máquina assíncrona é composta de duas partes principais: um estator (parte estacionária) e um rotor (parte rotativa).

Quando o enrolamento do estator é energizado através de uma alimentação trifásica, cria-se um campo magnético rotativo dentro do estator.

À medida que o campo varre os condutores do rotor, é induzida uma força eletro-motriz nos condutores ocasionando o aparecimento de um fluxo de corrente nos condutores. Os condutores do rotor transportando corrente no campo do estator possuem um torque exercido sobre eles que fazem o rotor girar.

Assim, na máquina assíncrona de rotor em gaiola a corrente de estator cria um campo eletromagnético girante com velocidade síncrona ω_s que induz as correntes nos circuitos do rotor. A velocidade síncrona do campo girante induzida pelas correntes de estator se relaciona com a frequência da tensão elétrica aplicada ao estator através da seguinte equação:

$$\omega_s = 2\pi f_s \quad (1)$$

onde: f_s é a frequência elétrica das tensões de estator da máquina assíncrona em Hz.

O eixo do rotor da máquina assíncrona gira a uma velocidade diferente da velocidade síncrona (para a máquina síncrona funcionando como motor, inferior, e superior no caso da máquina síncrona estar funcionando como gerador). A essa diferença de velocidade entre o eixo do rotor da máquina e o fluxo magnético gerado pelas tensões de estator se dá o nome de escorregamento definido por:

$$s = \frac{\omega_s - \omega_r}{\omega_s} \quad (2)$$

onde: ω_r é a velocidade de rotação do eixo do rotor em rad/s e ω_s a frequência do campo magnético em rad/s. O escorregamento (s) é uma medida adimensional.

Para operação como gerador a velocidade de rotação do eixo do rotor deve ser superior à velocidade de rotação síncrona do campo eletromagnético, ou seja, a máquina assíncrona de rotor em gaiola opera como gerador com escorregamento negativo. Já para operação como motor a velocidade de rotação do eixo do rotor é inferior à velocidade de rotação síncrona do campo eletromagnético e o escorregamento fica positivo.

Quanto maior o torque mecânico aplicado ao eixo do rotor, maior a velocidade de rotação do mesmo e, conseqüentemente, maior o módulo do escorregamento. A partir da equação que relaciona os torques que estão agindo no gerador obtém-se a seguinte expressão para a velocidade de rotação:

$$\dot{\omega}_r = \frac{1}{2H_g}(T_e - T_m) \quad (3)$$

sendo o torque elétrico (T_e), em pu¹, obtido através de:

$$T_e = X_m \Im(\mathbf{i}_s^* \mathbf{i}_r) \quad (4)$$

onde: X_m é a reatância de magnetização da máquina assíncrona, H_g é a constante de inércia, T_m é o torque mecânico e \mathbf{i}_s e \mathbf{i}_r são, respectivamente, as correntes do estator e rotor².

¹pu é a sigla para: por unidade.

² \Im representa parte imaginária do produto $\mathbf{i}_s^* \mathbf{i}_r$.

3.2.2.3 Equações Elétricas da Máquina Assíncrona

As tensões nos terminais do estator (\mathbf{v}_s) e do rotor (\mathbf{v}_r) podem ser expressas em função das correntes (\mathbf{i}) e fluxos magnéticos (ψ) da máquina assíncrona como variáveis complexas no sistema por unidade (pu) da seguinte maneira (KUNDUR, 1994):

$$\mathbf{v}_s = -R_s \mathbf{i}_s - j\omega_s \psi_s + \frac{1}{\omega_b} \dot{\psi}_s \quad (5a)$$

$$\mathbf{v}_r = -R_r \mathbf{i}_r - j(\omega_s - \omega_r) \psi_r + \frac{1}{\omega_b} \dot{\psi}_r \quad (5b)$$

e os fluxos concatenados (do estator, ψ_s , e do rotor, ψ_r) dados por:

$$\psi_s = X_s \mathbf{i}_s + X_m \mathbf{i}_r \quad (6a)$$

$$\psi_r = X_m \mathbf{i}_s + X_r \mathbf{i}_r \quad (6b)$$

onde: X_r e X_s são, respectivamente, as reatâncias do rotor e estator e ω_b é a frequência base.

3.2.2.4 Modelo de Quinta Ordem da Máquina Assíncrona

As variáveis complexas, vistas na Eq. (5) e (6), são então referenciadas no novo sistema de coordenadas ortogonais dado pelos eixos direto (d) e em quadratura (q), fixo no referencial síncrono como visto na Fig. 11. Deste modo, obtêm-se as seguintes equações:

$$\mathbf{v}_s = \mathbf{v}_{ds} + j \mathbf{v}_{qs} \quad (7a)$$

$$\mathbf{v}_r = \mathbf{v}_{dr} + j \mathbf{v}_{qr} \quad (7b)$$

$$\mathbf{i}_s = \mathbf{i}_{ds} + j \mathbf{i}_{qs} \quad (7c)$$

$$\mathbf{i}_r = \mathbf{i}_{dr} + j \mathbf{i}_{qr} \quad (7d)$$

$$\psi_s = \psi_{ds} + j \psi_{qs} \quad (7e)$$

$$\psi_r = \psi_{dr} + j \psi_{qr} \quad (7f)$$

Substituindo as Eq. (7) nas Eq. (3), (4), e (5) obtêm-se o modelo clássico de quinta (5^a) ordem da máquina assíncrona formado pelas seguintes equações:

$$\mathbf{v}_{ds} = -R_s \mathbf{i}_{ds} - \omega_s \psi_{qs} + \frac{1}{\omega_b} \dot{\psi}_{ds} \quad (8a)$$

$$\mathbf{v}_{qs} = -R_s \mathbf{i}_{qs} - \omega_s \psi_{ds} + \frac{1}{\omega_b} \dot{\psi}_{qs} \quad (8b)$$

$$\mathbf{v}_{dr} = -R_r \mathbf{i}_{dr} - (\omega_s - \omega_r) \psi_{qr} + \frac{1}{\omega_b} \dot{\psi}_{dr} \quad (8c)$$

$$\mathbf{v}_{qr} = -R_r \mathbf{i}_{qr} - (\omega_s - \omega_r) \psi_{dr} + \frac{1}{\omega_b} \dot{\psi}_{qr} \quad (8d)$$

$$\dot{\omega}_r = \frac{1}{2H_g} (T_e - T_m) \quad (8e)$$

$$T_e = X_m (\mathbf{i}_{ds} \mathbf{i}_{qr} - \mathbf{i}_{qs} \mathbf{i}_{dr}) \quad (8f)$$

onde: R_r é a resistência do rotor e R_s é a resistência do estator, ambas em pu.

onde:

$$\begin{bmatrix} \Psi_{ds} \\ \Psi_{qs} \\ \Psi_{dr} \\ \Psi_{qr} \end{bmatrix} = \begin{bmatrix} X_s & 0 & X_m & 0 \\ 0 & X_s & 0 & X_m \\ X_m & 0 & X_r & 0 \\ 0 & X_m & 0 & X_r \end{bmatrix} \cdot \begin{bmatrix} \mathbf{i}_{ds} \\ \mathbf{i}_{qs} \\ \mathbf{i}_{dr} \\ \mathbf{i}_{qr} \end{bmatrix} \quad (9a)$$

O modelo da máquina assíncrona pode ser representado por variáveis de estado em termos da velocidade de rotação, das correntes ou dos fluxos magnéticos, sendo que ao se utilizar as correntes como variáveis de estado ocorre um aumento considerável da complexidade do modelo dinâmico, ao invés dos fluxos magnéticos.

3.2.2.5 Modelo de Terceira Ordem da Máquina Assíncrona

Uma redução da ordem do modelo dinâmico formado pelas Eq. (8) e (9) pode ser realizada omitindo-se os transitórios do estator; isto é, considerando $\Psi_{ds} = 0$ e $\Psi_{qs} = 0$. Além disso, para se obter a representação simplificada da máquina assíncrona, modelo de 3ª ordem, são definidas as tensões internas do eixo direto (V'_d) e do eixo em quadratura (V'_q) como novas variáveis de estado do seguinte modo:

$$V'_d = \omega_s \frac{X_m}{X_r} \Psi_{qr} \quad (10a)$$

$$V'_q = -\omega_s \frac{X_m}{X_r} \Psi_{dr} \quad (10b)$$

Desconsiderando-se a dinâmica do estator, a respectiva frequência elétrica ω_s pode ser considerada constante e igual à frequência fundamental da rede em pu, ou seja:

$$\omega_s = 1 \quad (11)$$

Assim, o escorregamento, Eq. (2), fica definido como:

$$s = 1 - \omega_r \quad (12)$$

Esta simplificação não compromete a precisão dos resultados e diminui a necessidade dos recursos computacionais para as simulações numéricas (KUNDUR, 1994). Logo, considerando as simplificações realizadas para os fluxos e frequência do estator, substituindo as Eq. (10) e (12) em (8) e (9), e rearranjando, obtém-se o modelo de 3ª ordem da máquina assíncrona, cujas equações diferenciais são dadas por:

$$\dot{s} = \frac{1}{2H_g} (T_e - T_m) \quad (13a)$$

$$\dot{V}'_d = \frac{-1}{T_0} \left[V'_d - I_{qs}(X_s - X'_s) \right] + s \omega_s V'_q \quad (13b)$$

$$\dot{V}'_q = \frac{-1}{T_0} \left[V'_q + I_{ds}(X_s - X'_s) \right] - s \omega_s V'_d \quad (13c)$$

onde: I_{ds} e I_{qs} são as correntes do estator no eixo direto e em quadratura, respectivamente.

As equações algébricas são dadas por:

$$V_{ds} = V'_d - R_s I_{ds} + X'_s I_{qs} \quad (14a)$$

$$V_{qs} = V'_q - R_s I_{qs} + X'_s I_{ds} \quad (14b)$$

$$T_e = V'_d I_{ds} + V'_q I_{qs} \quad (14c)$$

sendo as constantes T_0 e X'_s obtidas através das seguintes expressões:

$$T_0 = L_{rr} / (\omega_s R_r) \quad (15a)$$

$$X'_s = L_{ss} - L_m^2 / L_{rr} \quad (15b)$$

onde: L_r é a indutância do rotor, com $L_{rr} = L_r + L_m$; L_s a indutância do estator ($L_{ss} = L_s + L_m$), e L_m a indutância de magnetização (REGINATTO, 1993; TARNOWSKI, 2006).

3.2.2.6 Estratégia de Controle da Máquina Assíncrona

A estratégia de controle adotada para o motor é a do tipo liga/desliga, isto é:

$$V_{ds} = \begin{cases} 0 & : T_h \geq 110\% \\ 1 & : 0 \leq T_h < 110\% \end{cases} \quad (16)$$

Assim, o motor irá parar de funcionar caso o valor de torque alcance o sobretorque máximo admissível (110%) definido neste trabalho.

3.2.3 Modelo do Conjunto de Engrenagens

As engrenagens são elementos mecânicos formadas, geralmente, por um par de rodas dentadas que se ligam a um eixo rotativo, ao qual imprimem movimento. A função básica de uma engrenagem é transmitir velocidade e torque.

As engrenagens apresentam tamanhos e formatos variados, alterando assim o tipo de transmissão de movimento, como visto na Fig. 12. Deste modo, pode-se classificar as engrenagens de acordo com os seguintes tipos:

1. Engrenagens cilíndricas retas;
2. Engrenagem coroa e parafuso sem-fim;
3. Engrenagens cilíndricas helicoidais;
4. Engrenagens cilíndricas cônicas.

No desenvolvimento do modelo do sistema de engrenagens do atuador, Fig. 10, foi levando em conta estes tipos de engrenagens: cilíndrica reta, sem-fim e cônica.

O motor de indução aciona o primeiro par de engrenagens de dentes retos (primeiro sistema de redução cinemática). No eixo central da segunda engrenagem está acoplado o parafuso sem-fim que transmite movimento a coroa (segundo sistema de redução cinemática).

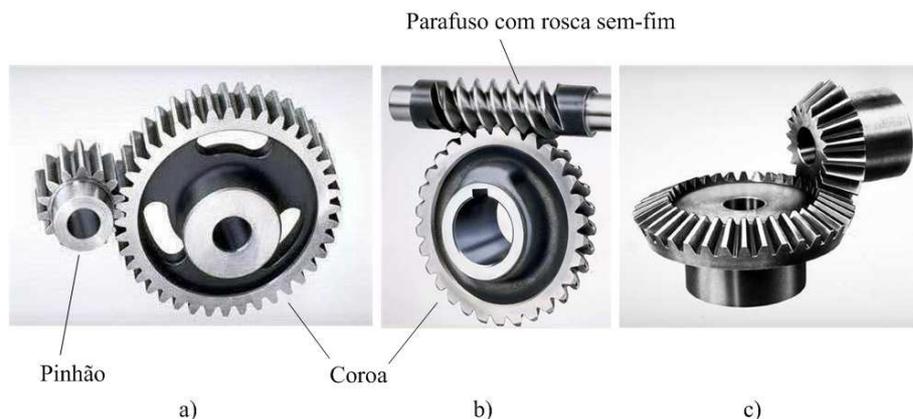


Figura 12: Engrenagens de: a) dentes retos b) coroa e sem-fim e c) cônicas.

O movimento do par coroa/sem-fim aciona uma engrenagem cônica inferior, que é responsável pelo movimento das engrenagens planetárias ligadas à manga de saída (terceiro sistema de redução cinemática). A manga de saída, acoplada a um eixo (a haste da válvula), é responsável final então pelo movimento da gaveta da válvula, abrindo-a ou fechando-a.

3.2.3.1 Primeiro Sistema de Redução Cinemática

O primeiro sistema de redução cinemática é formado por um par de engrenagens retas. Estas engrenagens são chamadas assim devido ao fato de possuírem um par de engrenagens cilíndricas de dentes retos. São normalmente empregadas para transmitir o movimento entre eixos paralelos, como visto na Fig.12 a).

A relação de transmissão de velocidade (ω) e torque (T) entre o primeiro par de engrenagens (primeiro sistema de redução cinemática, ou redução primária) é dada por:

$$K_1 = \frac{\omega_r}{\omega_1} = \frac{T_1}{T_m} \quad (17)$$

onde: ω_r é velocidade de giro do eixo do rotor (acoplada ao primeiro pinhão), ω_1 é a velocidade de giro do parafuso com rosca sem-fim (acoplada à coroa), T_m é o torque mecânico transmitido pelo motor, T_1 é o torque transmitido ao parafuso com rosca sem-fim.

Assim, o torque T_1 , em Nm, pode ser escrito como:

$$T_1 = T_m K_1 \quad (18)$$

e obtido (em pu) do seguinte modo:

$$T_1 = -T_m T_{mb} K_1 \quad (19)$$

onde: T_{mb} é o torque mecânico de base em Nm e K_1 é o coeficiente de redução cinemática.

A relação entre a força (F), em Newtons, e o torque aplicado em um eixo é dada por:

$$T = \mathbf{d} \times \mathbf{F} = dF \text{sen}(\theta) \quad (20)$$

onde: d é a distância (em metros) entre o eixo de rotação e ponto de aplicação da força e θ é o ângulo entre o braço do torque e a força.

Substituindo a Eq. (19) na Eq. (20) obtém-se a expressão da força aplicada no sem-fim:

$$F_1 = \frac{T_1}{d_1 \text{sen}(\theta_1)} = \frac{-T_m T_{mb} K_1}{d_1 \text{sen}(\theta_1)} \quad (21)$$

onde: $\theta_1 \cong 90^\circ$ e d_1 é a distância entre o eixo de rotação e ponto de aplicação da força F_1 .

3.2.3.2 Segundo Sistema de Redução Cinemática

O segundo sistema de redução cinemática é formado por um parafuso com rosca sem-fim e uma engrenagem do tipo coroa, como mostra a Fig.12 b).

A relação de transmissão de velocidade para o segundo par de engrenagens é dada por:

$$K_2 = \frac{\omega_1}{v_2} = \frac{2\pi}{P_s} = \frac{F_2}{T_1} \quad (22)$$

onde: v_2 é velocidade tangencial da coroa, P_s é o passo do sem-fim, K_2 é o segundo coeficiente de redução cinemática e ω_1 é a frequência de giro do sem-fim.

Assim, a relação entre o torque do sem-fim (T_1) e a força aplicada na coroa (F_2) é:

$$F_2 = T_1 K_2 \quad (23)$$

Ainda, o torque sentido pela coroa é:

$$T_2 = F_2 d_2 \text{sen}(\theta_2) \quad (24)$$

onde: d_2 é a distância entre o eixo de rotação da coroa e o ponto de aplicação de F_2 e $\theta_2 \cong 90^\circ$.

3.2.3.3 Terceiro Sistema de Redução Cinemática

O terceiro sistema de redução cinemática é formado por um conjunto de engrenagens cilíndricas helicoidais e cônicas, conhecido como sistema diferencial, como na Fig.12 c).

Esse sistema foi modelado, levando em conta uma relação linear entre os torques da coroa, T_2 , e da haste, T_h , isto é:

$$T_h = T_2 K_3 \quad (25)$$

Agora, considerando a Eq. (20), obtém-se a expressão para a força na haste, F_h :

$$F_h = \frac{T_2 K_3}{d_h \text{sen}(\theta_3)} \quad (26)$$

onde: d_h é a distância entre o eixo de rotação do sistema diferencial e o ponto de aplicação de F_h ; isto é, o comprimento do eixo acoplado a manga de saída, e $\theta_3 \cong 90^\circ$.

Substituindo as Eq. (24), (23), e (19) na equação (26) adquire-se a expressão para a força aplicada na haste:

$$F_h = \frac{-T_m T_{mb} K_1 K_2 K_3 d_2}{d_1 d_h \text{sen}(\theta_3)} \quad (27)$$

que, simplificando, fica:

$$F_h = \frac{-T_m K_h T_{mb}}{d_h} \quad (28)$$

com:

$$K_h = \frac{K_1 K_2 K_3 d_2}{d_1 d_h} \quad (29)$$

onde: K_h é definido como o coeficiente de redução cinemática geral do modelo e K_3 é o terceiro coeficiente de redução cinemática.

Utilizando a Eq. (22), a Eq. (28) pode ser re-escrita como a seguir, e assim pode-se obter a expressão final para a F_h :

$$F_h = \frac{T_h}{d_h \text{sen}(\theta)} \quad (30)$$

com: $\theta_3 = \theta$.

3.2.4 Modelo da Válvula e Tubulação

As equações do modelo da válvula (e tubulação) levam em conta as equações de dimensionamento da válvula e do balanço de forças que existe dentro do trecho da tubulação (CAMPOS; TEIXEIRA, 2006).

A relação que expressa a vazão de um fluido pela área ocupada pelo obturador da válvula e os fatores geométricos do corpo da válvula é da pela seguinte equação:

$$V_f = \frac{a}{100} N_m F_p F_y N_R C_v \sqrt{\Delta P \rho} \quad (31)$$

onde: V_f é a vazão do fluido (dada em kg/s), a é a abertura percentual da válvula, N_m é uma constante numérica que depende das unidades de medida, F_p é o fator de geometria da tubulação adjacente, F_y é o fator de correção devido ao fluxo crítico (expressa a relação entre a pressão diferencial máxima e efetiva na produção de vazão), N_R é o número de Reynolds (adimensional, é usado para caracterizar o tipo de fluxo que caracteriza a turbulência: laminar ou turbulento), C_v [m³/s] é o coeficiente de vazão, ΔP [N/m²] é a diferença de pressão do fluido existente entre os dois lados do obturador e ρ [kg/m³] é a densidade do fluido.

Quando o fluxo é laminar ($N_R < 2100$) o líquido tende a se mover em camadas com velocidades consideradas baixas. Já quando o fluxo é turbulento ($N_R > 4000$) as velocidades são maiores e ocorre a formação de redemoinhos e de movimentos caóticos que dificultam a vazão:

A Eq. (31) pode ser escrita do seguinte modo:

$$V_f = N_R f(a) C_v \sqrt{\Delta P \rho} \quad (32)$$

sendo:

$$f(a) = \frac{a}{100} N_m F_p F_y \quad (33)$$

onde: $f(a)$ expressa a característica de vazão inerente da válvula.

Essa característica de vazão está relacionada à geometria da válvula; i.e. ao tipo de válvula que está sendo analisada: gaveta, borboleta, ou globo, por exemplo (SHARIF; GROSVENOR, 1998). Dependendo do tipo de válvula, distintos tipos de vazões (assim como distintas equações para a vazão) podem ser obtidos, tais como:

$$f(a) = a \quad (34a)$$

$$f(a) = R^{a-1} \quad (34b)$$

$$f(a) = \frac{a}{\sqrt{3-2a^2}} \quad (34c)$$

cuja característica é linear, exponencial, ou parabólica, respectivamente.

A característica de vazão escolhida para este trabalho foi a linear ³:

$$f(a) = \begin{cases} 100 - a & : \text{ para fechamento} \\ a & : \text{ para abertura} \end{cases} \quad (35)$$

Substituindo a Eq. (35) (para fechamento) em (31) e isolando ΔP , obtém-se a expressão para a diferença de pressão do fluido:

$$\Delta P = \frac{V_f^2}{\rho N_R^2 (100 - a)^2 C_v^2} \quad (36)$$

A força exercida pelo fluido em uma certa área de seção transversal (A_v) é dada por:

$$F_f = \Delta P A_v \quad (37)$$

Substituindo a Eq. (36) em (37) obtém-se:

$$F_f = \frac{V_f^2 A_v}{\rho N_R^2 (100 - a^2) C_v^2} \quad (38)$$

As forças dinâmicas consideradas no desenvolvimento do modelo do conjunto atuador, válvula e tubulação foram: a força da haste, ou do atuador, (F_h), a força de atrito (F_a), a força da mola (F_m) e a força do fluido (F_f), como visto na Fig. 9. Realizando o balanço dinâmico destas forças obtém-se:

$$F_h - \Delta P_f A_v - F_a - F_m = M_h \frac{d^2 a}{dt^2} \quad (39)$$

Substituindo (37) em (39) e reescrevendo a equação obtém-se a equação para a aceleração da haste, a_a :

$$a_a = \frac{d^2 a}{dt^2} = \frac{1}{M_h} (F_h - F_f - F_a - F_m) \quad (40)$$

onde: M_h é a massa do conjunto haste e obturador; F_h é a força da haste, F_a é a força de atrito e F_m é a força da mola.

³Foi simulada somente a situação de fechamento.

As expressões para a força de atrito (F_a), considerando somente o atrito viscoso, e para a força da mola (F_m) são dadas, respectivamente, por:

$$F_a = C_a v_a \quad (41a)$$

$$F_m = K_m a \quad (41b)$$

onde: C_a é a constante de atrito (atrito entre a haste e a gaxeta da válvula gaveta), K_m é o coeficiente de Hook da mola e v_a é a velocidade da haste, que é dada por:

$$v_a = \dot{a} = \frac{da}{dt} \quad (42)$$

A expressões para a força da haste (F_h) e para a força do fluido (F_f) são dadas pelas Eq. (30) e Eq. (38), respectivamente.

3.2.5 Conjunto de Equações Algébricas e Diferenciais do Modelo

A dinâmica dos principais componentes do modelo (atuador, válvula e tubulação), bem como das principais forças envolvidas na abertura e fechamento da válvula, foi representada por uma série de equações algébricas e diferenciais: Eq. (43), (44) e (45). Uma série de considerações físicas foram realizadas na modelagem para simplificar a representação e conseqüentemente o esforço computacional exigido. Por sua vez, as simplificações foram limitadas pela exatidão requerida nas análises.

Assim, durante o processo de modelagem, buscou-se manter um compromisso entre a exatidão e a simplicidade de cálculo exigidas, para que uma análise razoável pudesse ser efetuada num intervalo de tempo admissível para o problema em questão.

Na literatura a respeito, pode-se encontrar diversos modelos para a máquina assíncrona, desde os mais simples (1^a ordem) até os mais complexos (5^a ordem), bem como quanto a estratégia de controle, destas máquinas, utilizada, como visto na subseção 3.2.2.

O modelo de máquina assíncrona (motor de indução) escolhido, em especial, foi o modelo de 3^a ordem; pois esse, possui um bom compromisso entre a exatidão e o tempo desejados (KUNDUR, 1994; FITZGERALD, 1990; NASAR, 1984; PEREIRA, 2007; TARNOWSKI, 2006; REGINATTO, 1993).

O modelo de válvula e tubulação proposto é baseado nas equações de dimensionamento da válvula e no balanço de forças que existe dentro do trecho da tubulação ocupado pela válvula (CAMPOS; TEIXEIRA, 2006). A característica de vazão escolhida para este trabalho foi a linear, pois ela representa perfeitamente a dinâmica de abertura ou fechamento da válvula gaveta (modelo de válvula estudado).

O sistema de engrenagens foi modelado levando em conta os acoplamentos existentes entre os três sistemas de engrenagens do atuador: um par de engrenagens cilíndricas retas, um par formado por um parafuso sem fim e uma engrenagem coroa, e um sistema formado por engrenagens cilíndricas helicoidais e cônicas (BUCKINGHAM, 1935; HENRIOT, 1967).

Assim, as equações diferenciais ⁴ que descrevem o motor assíncrono em termos do seu escorregamento (s), das tensões internas no eixo direto e no eixo em quadratura (V'_q e V'_d) e da posição (a) e velocidade da gaveta são dadas por:

$$\dot{s} = \frac{1}{2H_g}(T_e - T_m) \quad (43a)$$

$$\dot{V}'_d = \frac{-1}{T_0} \left[V'_d - I_{qs}(X_s - X'_s) \right] + s \omega_s V'_q \quad (43b)$$

$$\dot{V}'_q = \frac{-1}{T_0} \left[V'_q + I_{ds}(X_s - X'_s) \right] - s \omega_s V'_d \quad (43c)$$

$$\dot{a} = v_a \quad (43d)$$

$$\ddot{a} = a_a \quad (43e)$$

As equações algébricas que descrevem as tensões do estator do motor de indução no eixo direto e no eixo em quadratura (V_{ds} e V_{qs} , respectivamente), o torque elétrico do motor assíncrono (T_e), a força (F_h) e torque (T_h) ⁵ transmitidos pela haste para a gaveta da válvula, a força da mola (F_m), a força de atrito (F_a), a força exercida pelo fluido (F_f) (CAMPOS; TEIXEIRA, 2006) e a aceleração da gaveta (a_a) são:

$$V_{ds} = V'_d - R_s I_{ds} + X'_s I_{qs} \quad (44a)$$

$$V_{qs} = V'_q - R_s I_{qs} + X'_s I_{ds} \quad (44b)$$

$$T_e = V'_d I_{ds} + V'_q I_{qs} \quad (44c)$$

$$T_h = -T_m K_h T_{mb} \quad (44d)$$

$$F_h = \frac{T_h}{d_h \text{sen}(\theta)} \quad (44e)$$

$$F_m = K_m a \quad (44f)$$

$$F_a = C_a v_a \quad (44g)$$

$$F_f = \frac{V_f^2 A_v}{\rho N_R^2 (100 - a)^2 C_v^2} \quad (44h)$$

$$a_a = \frac{1}{M_h} (F_h - F_f - F_a - F_m) \quad (44i)$$

Ainda, há uma série de equações auxiliares dadas por:

$$L_{rr} = L_r + L_m \quad (45a)$$

$$X_s = L_{ss} = L_s + L_m \quad (45b)$$

$$T_0 = \frac{L_{rr}}{\omega_s R_r} \quad (45c)$$

$$X'_s = L_{ss} - \frac{L_m^2}{L_{rr}} \quad (45d)$$

$$K_h = \frac{K_1 K_2 K_3 d_2}{d_1 d_h} \quad (45e)$$

onde: L_r é a indutância do rotor, L_s a indutância do estator e L_m a indutância de magnetização.

⁴Modelo de 5ª ordem, $V_{qs} = 0$, e $\theta = 90^\circ$.

⁵O sinal negativo se deve a convenção adotada para o sinal do T_m : (-) motor (+) gerador.

Por fim, cabe salientar que a modelagem deste sistema, composto por uma máquina de indução trifásica, engrenagens, válvula e tubulação, não é única. Diversos outros modelos, mais complexos ou mais simples, levando em conta ou desconsiderando outros parâmetros, para o motor de indução, o sistema de engrenagens, as forças envolvidas, a válvula e a tubulação podem ser desenvolvidos e aplicados.

3.2.6 Solução do Modelo do Atuador Elétrico

Ao se pesquisar sobre a possível ferramenta de solução de sistemas não lineares a ser utilizada para simular o modelo, fez-se necessário escolher uma ferramenta que fosse aberta e que permitisse incluir distintos modelos de válvulas e atuadores.

As rotinas dessa ferramenta deveriam apresentar certa flexibilidade na estrutura e na modelagem do sistema a ser simulado; permitindo assim, a simulação de qualquer configuração do sistema e possibilitando a análise de vários modelos de atuadores e válvulas, bem como de diversas situações que pudessem ocorrer na tubulação.

Tendo em vista esta situação, utilizou-se um simulador de dinâmica de sistemas não-lineares, cujas janelas principais podem ser vistas na Fig. 13, adaptado de (GONÇALVES, 2004), para simular o modelo do atuador elétrico. Esse simulador foi desenvolvido com base no Matlab, pois apresenta as características mencionadas anteriormente.

Após a definição do modelo e do simulador, foi feita a inclusão das Eq. (43), (44) e (45) em um sistema de equações diferenciais e algébricas não-lineares, também conhecido como sistema singular ou descritor, conforme abaixo:

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{p}) \quad (46a)$$

$$0 = g(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{p}) \quad (46b)$$

onde:

- \mathbf{x} é um vetor de variáveis de estado;
- \mathbf{z} é um vetor de variáveis algébricas;
- \mathbf{u} é um vetor de entradas;
- \mathbf{p} é um vetor de parâmetros constantes;
- $f()$ e $g()$ são funções vetoriais não lineares formadas pelas Eq. (43), (44) e (45).

A representação algébrico-diferencial do sistema, mais especificamente a descrição estrutural das funções $f()$ e $g()$ (formada pelas Eq. (43), (44) e (45), respectivamente) constitui a informação básica de entrada para o simulador.

O procedimento usado pelo simulador trata de, a cada instante de tempo, calcular simultaneamente os valores dos vetores \mathbf{x} e \mathbf{z} ; por isto, também chamado de solução instantânea.

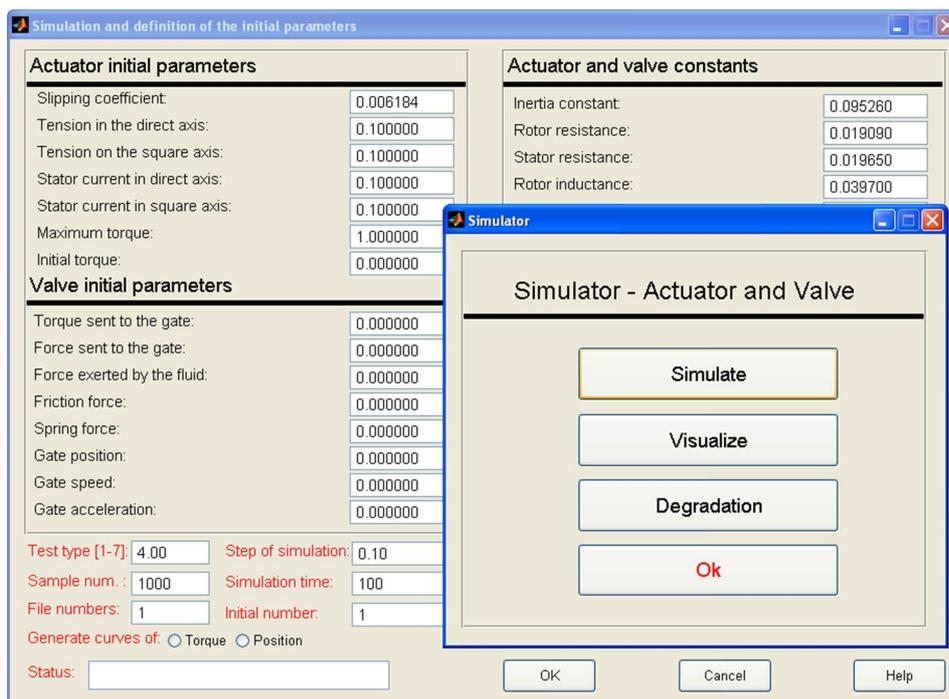


Figura 13: Janelas da interface do simulador do modelo do atuador elétrico.

Os vetores \mathbf{u} , \mathbf{x} , \mathbf{z} e \mathbf{p} foram definidos como a seguir:

$$\mathbf{u} = \begin{bmatrix} \omega_s & \dot{V}_{ds} & T_m \end{bmatrix} \quad (47a)$$

$$\mathbf{x} = \begin{bmatrix} \dot{s} & \dot{V}'_d & \dot{V}'_q & \dot{a} & \ddot{a} \end{bmatrix} \quad (47b)$$

$$\mathbf{z} = \begin{bmatrix} I_{ds} & I_{qs} & T_e & T_h & F_h & F_m & F_a & F_f & a_a \end{bmatrix} \quad (47c)$$

$$\mathbf{p} = \begin{bmatrix} H_g & R_s & X_s & X'_s & \dots & K_h & K_m & C_a & A_v & M_h \end{bmatrix} \quad (47d)$$

O procedimento de solução instantânea, utilizado para calcular os valores dos vetores \mathbf{x} e \mathbf{z} , inclui o uso de dois métodos numéricos: um para obter a solução das equações diferenciais (e do fluxo de carga) e outro para resolver o sistema de equações como um todo.

O método de integração usado para resolver as equações foi a regra do trapézio ou regra de Euler Modificada. A idéia central desse método é aproximar a derivada de uma função, $f(t)$, no intervalo de tempo entre t e $t - \Delta t$, onde Δt é o passo de integração, por uma constante cujo valor é tomado como a média entre os valores da derivada nos extremos do intervalo, como visto a seguir:

$$\int_{t-\Delta t}^t f(t) dt \cong \frac{1}{2} [f(t - \Delta t)\Delta t + f(t)\Delta t] \quad (48)$$

O método utilizado para se obter a solução do sistema de equações não-lineares, a cada instante de tempo foi o conhecido método de Newton-Raphson (KUNDUR, 1994). A partir de uma representação das equações diferenciais por um sistema singular, feita através da regra do Trapézio, pode-se resolver o sistema de equações do sistema, a partir de uma aproximação inicial x_0 (obtida através da solução do fluxo de carga).

3.3 INJEÇÃO DE FALHAS NO MODELO

O modelo proposto ainda está em fase de validação. Estar-se-á aguardando da empresa Coester, que realiza a montagem do atuador elétrico, e da empresa WEG, que fabrica os motores, os valores específicos de parâmetros do motor e alguns detalhes técnicos construtivos e operacionais, além de outros valores, como a vazão do fluido e os coeficientes de atrito ou vazão, que deverão ser obtidos junto a empresa Petrobrás.

Após a coleta de todos os parâmetros do modelo, adquiridos a partir dos fabricantes dos componentes (válvula, motor e engrenagens), ou a partir de experimentos ou ensaios, e o cruzamento dos sinais obtidos do modelo (ajustado com os valores reais) com os sinais do atuador o modelo poderá ser validado.

Para tal, já se encontra em processo de montagem uma bancada de experimentos com um atuador elétrico da Coester, onde serão realizados uma série de ensaios e experimentos a fim de validar o modelo e ensaiar processos de degradação real dos componentes.

Atualmente, está se utilizando os parâmetros de um motor de indução trifásico com a mesma ordem de grandeza (5HP) e com o mesmo valor de torque máximo de saída (250Nm), coletados de (KUNDUR, 1994). Já é possível realizar simulações de comportamento normal e das principais falhas observadas no conjunto utilizando o modelo do atuador elétrico e visualizar as suas conseqüências; tais como: um aumento gradual do torque, ou um mau fechamento da gaveta da válvula, devido a alterações nos parâmetros do modelo.

Para fins de validação da técnica proposta neste trabalho para o desenvolvimento de um sistema de manutenção inteligente, foram efetuadas uma série de simulações de injeção de falhas usando o modelo do atuador elétrico descrito na seção 3.2.

Nestas simulações, foram geradas, inicialmente, uma série de curvas de torque e posição, por exemplo, para o modelo de atuador CSR25 (com torque máximo igual a 250Nm ou 1pu e 100% de abertura/fechamento) cujo comportamento foi definido como normal. Em seguida, efetuaram-se outras simulações, de injeção de falhas, onde foram alterados gradualmente certos parâmetros de algumas equações do modelo. Como exemplo de conseqüências destas alterações, cita-se: o valor de torque para abrir ou fechar a válvula atingiu o valor de 275Nm (10% de sobretorque ou 1.1 pu, que é o valor de sobretorque máximo admitido); a gaveta da válvula não alcançou a sua posição de abertura ou de fechamento esperada; ou seja, não atingiu o seu fim de curso (80% de abertura ou fechamento, por exemplo); ou a gaveta da válvula demorou mais do que o esperado para efetuar o movimento de abertura/fechamento.

Os três principais tipos de simulações de injeção de falhas realizados, atualmente, através do modelo, os parâmetros alterados e as principais conseqüências nas curvas de torque e posição em virtude da alteração destes parâmetros, estão listados na Tab. 2.

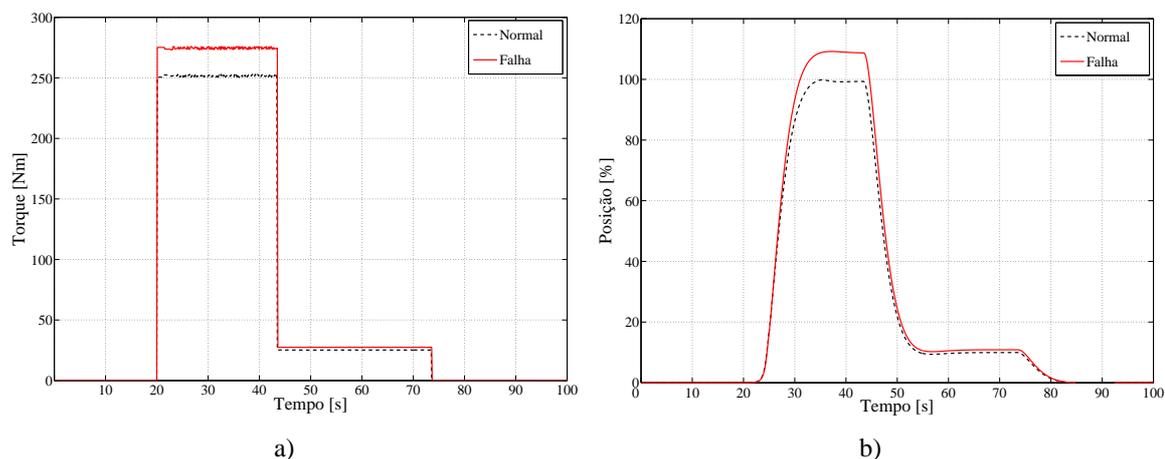
A falha do tipo 1 (parâmetro K_h) corresponde a falha observada no segundo sistema de redução cinemática. Mais especificamente, trata-se da degradação do sem-fim (peça mais frágil do sistema de engrenagens) que ocorre devido a esforços mecânicos adicionais.

Tabela 2: Tipos de simulações de injeção de falhas.

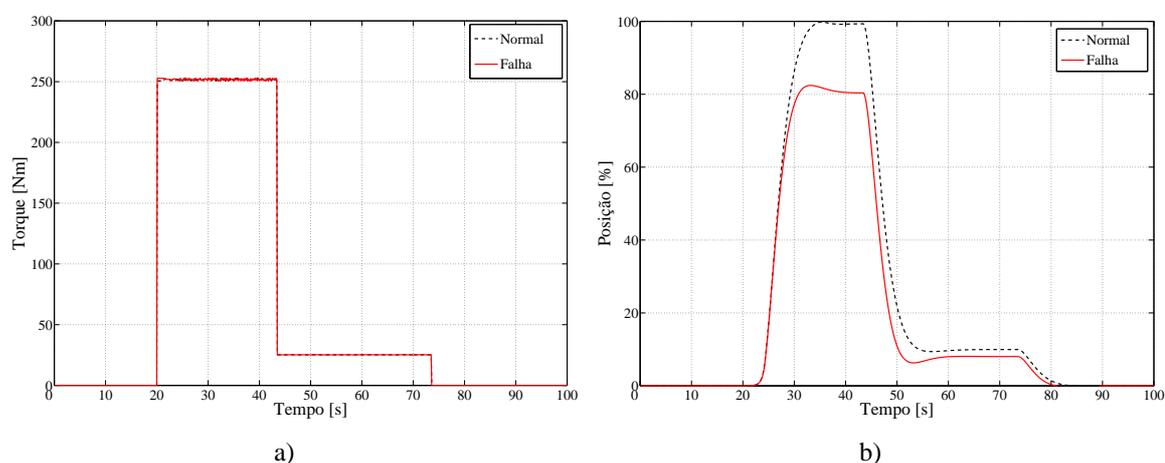
Tipo	Parâmetro	Conseqüências
1	K_h	Alteração do torque e posição da gaveta
2	K_m	Modificação da posição da gaveta da válvula
3	C_a	Alteração do tempo de resposta abertura/fechamento

Pode-se observar na Fig. 14 um exemplo de curvas de torque e posição, gerados pelo modelo, para as situações de comportamento normal e de falha considerando a injeção de falhas do tipo 1 (K_h). Nota-se, nessa figura, que a curva de torque, na falha, alcançou o valor máximo permitido (275Nm) e a posição da gaveta ultrapassou o curso previsto ($\cong 110\%$).

Esta situação de extrapolação da posição da gaveta na prática não existe, somente em nível de simulação; pois, há o limite físico construtivo da válvula, que impede o deslocamento da gaveta acima dos 100% e que não poderia ser ultrapassado.

**Figura 14: Curvas do modelo considerando as falhas do tipo 1: a) torque e b) posição.**

A falha do tipo 2 (K_m) corresponde a degradação da mola. A mola, ao longo do tempo, vai perdendo a sua elasticidade; alterando a sua ação de compressão e descompressão.

**Figura 15: Curvas do modelo considerando as falhas do tipo 2: a) torque e b) posição.**

Um exemplo de curvas de torque e posição, gerados pelo modelo, para as situações de comportamento normal e de falha considerando agora a injeção de falhas do tipo 2 é visto na Fig. 15. Observa-se, nessa figura, que a curva de torque alcança, na situação de falha e de comportamento normal, o mesmo valor máximo (250Nm). Já com relação a posição da gaveta, essa não atingiu o curso previsto ($\cong 80\%$) na situação de falha.

Já a falha do tipo 3 (C_a), equivale a um aumento do atrito existente entre a haste e a gaxeta. Eventualmente, pode ocorrer um vazamento de algum fluido para a região de acoplamento da haste com a válvula. Esse fluido com o tempo irá se cristalizar, alterando a constante de atrito.

Considerando a injeção de falhas do tipo 3, pode-se observar um exemplo de curvas de torque e posição, gerados pelo modelo, para as situações de comportamento normal e de falha nas Fig. 16 a) e b), respectivamente.

Nota-se, nessas figuras, que a curva de torque, semelhante à falha do tipo 3, alcança, na situação de falha e de comportamento normal, o mesmo valor máximo (250Nm). Com relação à posição da gaveta, na situação de falha, essa atingiu o curso previsto ($\cong 100\%$). Contudo, o tempo para que a gaveta alcançasse a posição final durante a falha é maior do que na situação de comportamento normal.

Destaca-se que os tipos de simulações de injeção de falhas (K_h , K_m e C_a) que foram analisados visam reproduzir situações normalmente verificadas no atuador e válvula, em campo, no decorrer do seu uso. Além disto, essas simulações são baseadas em dados experimentais e na experiência dos técnicos de manutenção.

Uma situação, normalmente, observada em aplicações de controle de fluxo de água e esgoto (reportada pelos técnicos de manutenção da Coester) é o acúmulo de sedimentos (ou resíduos) na sede da válvula. Estes resíduos podem ser grãos de areia ou até pequenas partículas (como galhos).

Nestes casos, para que o obturador se acomode na sede da válvula e ocorra uma perfeita vedação, é necessário que o atuador exerça um torque maior (muitas vezes atingindo o limite de torque e degradando o sem-fim e outras peças).

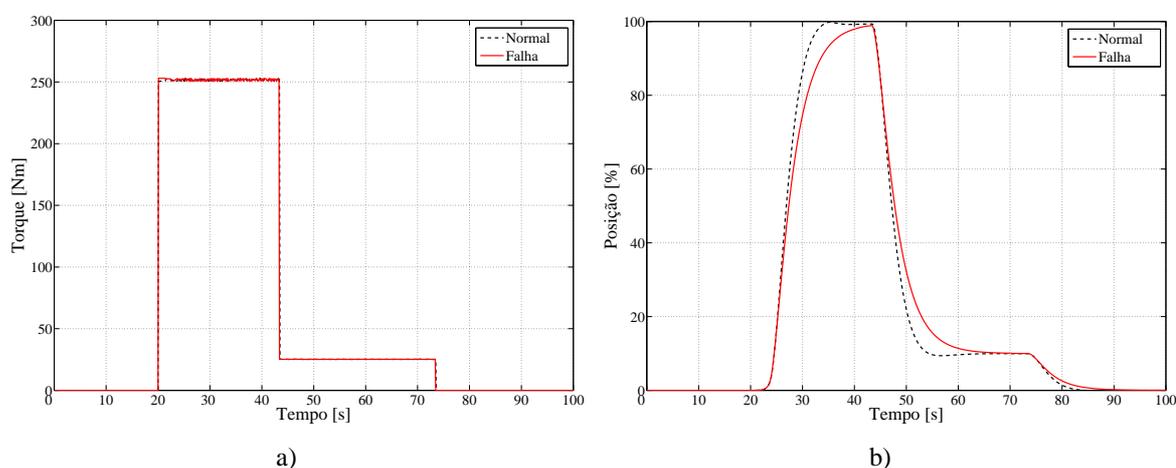


Figura 16: Curvas do modelo considerando as falhas do tipo 3: a) torque e b) posição.

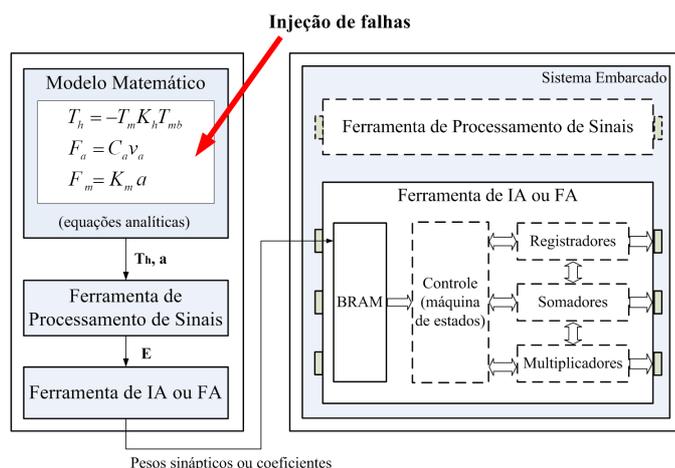


Figura 17: Representação da etapa de injeção de falhas.

Esta situação de degradação do sem-fim, em função do acúmulo de sedimentos ou mesmo em função da degradação natural das engrenagens e outros componentes, ocorre com frequência no uso e operação do atuador.

Na Fig. 17 pode-se observar como foi realizada a etapa de injeção de falhas. Através da alteração gradual de certos parâmetros de algumas das equações do modelo simulou-se a injeção de falhas no modelo do atuador. Em seguida, estes dados (torque, T_h , e posição, a) foram repassados para a ferramenta de processamento de sinais onde se analisou o conteúdo de energia (**E**) destes sinais. Por fim, estes dados de energia foram utilizados pelas ferramentas de IA ou FA nas etapas de inicialização, treinamento e teste.

As equações, onde se processou a injeção de falhas, através da alteração gradual dos parâmetros, foram reescritas abaixo:

$$T_h = -T_m K_h T_{mb} \quad (49a)$$

$$F_m = K_m a \quad (49b)$$

$$F_a = C_a v_a \quad (49c)$$

onde: K_h é o coeficiente de redução cinemática, K_m é o coeficiente da mola e C_a é a constante de atrito da haste.

A faixa de variação dos parâmetros K_h , K_m e C_a utilizada para a injeção de falhas, vista na Tab. 3, foi escolhida de modo a atingir o limite de sobretorque do atuador e alterar de modo significativo a posição final ou o tempo de abertura/fechamento da gaveta da válvula. A taxa de variação dos parâmetros, por sua vez, foi escolhida de modo a alcançar o limite de 100 (arquivos) amostras⁶ por faixa de variação.

Essa variação gradual dos parâmetros K_h , K_m e C_a do modelo visa imitar uma situação de degradação dos componentes (engrenagens, mola e haste da válvula) até a ocorrência da falha com base nos valores de torque e posição atingidos.

⁶Uma amostra, ou ciclo, corresponde a geração (ou a tomada) de um conjunto de dados para uma dada situação. Neste caso, a cada ciclo de abertura/fechamento da válvula.

Tabela 3: Faixa e taxa de variação de K_h , K_m e C_a .

Parâmetro	Faixa	Taxa
K_h	11.00 - 12.00	0.01
K_m	4.215 - 5.215	0.01
C_a	16.00 - 21.00	0.05

Assim, os dados de simulação (100 amostras acrescidas de ruído branco para cada tipo de comportamento) gerados a partir do modelo podem ser divididos em três categorias de comportamento:

1. Normal: dados que representam o comportamento normal ($T_h = 250\text{Nm}$ e $a = 100\%$) do atuador elétrico, com $K_h = 11.00$, $K_m = 4.215$ e $C_a = 16.00$.
2. Degradação: dados que simulam a degradação ($250 \leq T_h \leq 275\text{Nm}$ e $a \neq 100\%$) do sistema, com $11.00 \leq K_h \leq 12.00$, $4.215 \leq K_m \leq 5.215$ e $16.00 \leq C_a \leq 21.00$.
3. Falha: dados que expressam a situação de falha ($T_h = 275\text{Nm}$ e $a \neq 100\%$) já estabelecida, $K_h = 12.00$, $K_m = 5.215$ e $C_a = 21.00$.

Deste modo, foram geradas, no total, 300 amostras para cada tipo de falha: 100 amostras iniciais que representam a região de comportamento normal do atuador, seguidas de 100 amostras que simulam a degradação, finalizando com mais 100 amostras para a situação de falha já estabelecida ⁷.

3.4 CONCLUSÕES DO CAPÍTULO

Ao longo deste capítulo, uma descrição do estudo de caso, um atuador elétrico, incluindo as principais partes do atuador (motor, engrenagens e válvula) e forças envolvidas na operação do atuador elétrico (força do atuador, mola, atrito e fluido), foi apresentada.

O desenvolvimento do modelo matemático do estudo de caso (composto por uma máquina assíncrona, um sistema de engrenagens, uma válvula gaveta e tubulação) foi visto a seguir. O modelo do atuador elétrico permite a realização de uma série de simulações de comportamento normal, de degradação e de falhas.

Por fim, a etapa de injeção de falhas no modelo do atuador (como, por exemplo, um aumento do atrito ou um desgaste do sistema de engrenagens) foi descrita. Para realizar a injeção de falhas, foram alterados gradualmente certos parâmetros de algumas equações do modelo. A alteração destes parâmetros produz uma série de consequências nas curvas de torque e posição, em especial, do modelo; simulando situações de degradação e falha nos componentes do atuador.

⁷O número de 100 amostras foi escolhido de modo aleatório.

4 ANÁLISE DAS FERRAMENTAS MATEMÁTICAS

A metodologia de desenvolvimento de sistemas de manutenção inteligentes, adotada neste trabalho, inclui, entre outras tarefas, a definição das ferramentas matemáticas (ferramentas de processamento de sinais, de inteligência artificial ou de filtragem, por exemplo) e computacionais (ambientes operacionais, ferramentas matemáticas, etc.) necessárias e adequadas para o estudo de caso aqui analisado.

Assim, neste trabalho, a determinação da ferramenta de processamento de sinais mais apropriada (para realizar a etapa de análise da energia do sinal) e a definição de qual ferramenta (MAO ou FA) é a mais adequada, em termos qualitativos e quantitativos, para realizar as tarefas de detecção, classificação e predição de falhas fez-se necessária.

Neste capítulo, é apresentada inicialmente, na seção 4.1, uma análise das ferramentas de processamento de sinais (TF, TFJ e TWP) através do método da regressão logística. A seguir, os resultados de energia dos sinais de torque e posição, obtidos através da ferramenta de processamento de sinais, são vistos na seção 4.2. A análise das etapas (de inicialização, treinamento e teste) e dos resultados (de detecção, classificação e predição de falhas) obtidos através dos mapas auto-organizáveis e filtros adaptativos, são exibidos nas seções 4.3 e 4.4, respectivamente. A determinação do tempo de vida útil, por meio dos mapas auto-organizáveis e filtros adaptativos, é abordada na seção 4.5. Por fim, as principais conclusões deste capítulo são vistas na seção 4.6.

4.1 ANÁLISE DAS FERRAMENTAS DE PROCESSAMENTO DE SINAIS

A determinação da ferramenta de processamento de sinais mais adequada para um determinado estudo de caso, ou para determinados tipos de sinais, leva em consideração uma série de fatores, tais como: a frequência do sinal, o número de amostras a serem computadas, se o sinal é estacionário ou não, entre outros aspectos. Assim, em virtude do tipo de sinal ou sistema analisado, haverá um tipo de ferramenta de processamento de sinais que será a mais apropriada para ser utilizada.

Deste modo, uma das etapas deste trabalho é realizar uma análise comparativa dentre as ferramentas de processamento de sinais existentes para definir qual seria a ferramenta mais adequada para o estudo de caso aqui analisado: o atuador elétrico. Uma descrição mais detalhada das principais ferramentas de processamento de sinais é apresentada no anexo A. Dentre as ferramentas de processamento de sinais, vistas no anexo A, foram examinadas as transformadas: de Fourier, de Fourier com janelamento e *wavelet packet*.

O método utilizado neste trabalho para determinar qual é a ferramenta de processamento de sinais mais apropriada para o estudo de caso foi o método da regressão logística (RL). O método da RL faz parte de uma categoria de modelos estatísticos chamados de modelos generalizados lineares (JOHNSON et al., 2006; YAN; LEE., 2005).

Este método permite obter uma saída discreta, como uma detecção, a partir de um conjunto de dados que pode ser contínuo, discreto, ou binário, por exemplo. Geralmente a resposta possui dois estados como presença/ausência, sucesso/falha, normal/falha, por exemplo. O método da regressão logística pode então ser utilizado na avaliação do comportamento de sistemas ou na detecção de falhas, por exemplo.

O *Center for Intelligent Maintenance Systems* (IMS Center), em especial, está desenvolvendo uma solução embarcada para avaliação, predição, e diagnóstico de falhas chamada *Watchdog Agent* (um PC industrial), além de uma interface implementada para o *software* Matlab, conhecida como *Watchdog Agent Toolbox* (JOHNSON et al., 2006).

As ferramentas de processamento de sinais (transformadas de Fourier, Fourier com janelamento e *wavelet packet*) e extração das características, bem como de avaliação do desempenho (método da regressão logística e de reconhecimento estatístico de padrões, por exemplo) foram desenvolvidas pelo pessoal do IMS Center nessa interface e usadas neste trabalho.

Para tal, inicialmente, foram geradas uma série de amostras a partir do atuador que visam imitar uma situação de evolução da degradação do comportamento do atuador, partindo de um comportamento normal, até a ocorrência da falha.

As etapas de injeção de falhas e de geração dos dados de torque e posição (obtidos a partir do modelo do atuador) que representam o comportamento normal, de degradação e de falha foram realizadas como descrito na seção 3.3.

O sinal obtido a partir das equações do modelo do atuador, analisado pelas ferramentas de processamento de sinais, é composto pelo torque da válvula, T_h , e pela posição do obturador da válvula gaveta, a , como visto a seguir:

$$y[n] = \begin{bmatrix} T_h[n] & a[n] \end{bmatrix} \quad (50)$$

onde: n é o número de amostras.

Os valores de energia, divididos em bandas, determinados a partir das transformadas de Fourier, Fourier com janelamento e *wavelet packet* foram então obtidos a seguir. O vetor de energia, \mathbf{E} , obtido a partir das ferramentas de processamento de sinais, é definido como:

$$\mathbf{E} = \sum_{n=1}^N |y[n]|^2 \quad (51)$$

Os valores de energia, \mathbf{E} , Eq. (51), foram utilizados como entradas para o algoritmo do método de regressão logística que foi usado para realizar a avaliação das ferramentas de processamento de sinais.

O gráfico que exprime o estado do sistema, gerado a partir da RL, conhecido como valor de confiança (VC), é determinado através de indicadores estatísticos que ressaltam as semelhanças entre o comportamento normal (e de falha) e o comportamento recentemente observado. Conforme a degradação altera o comportamento do sistema, os sinais atuais de desempenho do equipamento vão se diferenciando dos sinais de comportamento normal, reduzindo o VC.

O valor de confiança, V_c , quando se usa o método de regressão logística, é determinado a partir da seguinte expressão:

$$V_c(\mathbf{E}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 E_1 + \dots + \beta_k E_k)}} \quad (52)$$

onde: β_0, \dots, β_k , são parâmetros do método.

O VC varia de zero a um, onde um valor mais elevado indica que o desempenho está mais perto do normal e um valor mais próximo do zero revela uma maior proximidade da ocorrência de algum tipo de falha, como visto na Fig. 18.

Na Fig. 18, pode-se observar o resultado do valor de confiança para as três ferramentas de processamento de sinais analisadas (TF, TFJ e TWP) considerando a falha do tipo 1 (alteração do coeficiente de redução cinemática, K_h). Nota-se que o valor de confiança obtido com as TF e TWP (e RL) retrataram fielmente as condições (e regiões) de comportamento normal, degradação e falha que foram simuladas através do modelo. Já o valor de confiança alcançado com a TFJ não reproduziu as condições que foram simuladas.

O resultado do valor de confiança para as três ferramentas de processamento de sinais (TF, TFJ e TWP) levando em conta a falha do tipo 2 (alteração do coeficiente da mola, K_m) pode ser visto na Fig. 19. Observa-se que o valor de confiança (obtido com as TFJ e TWP) reproduziu corretamente as condições de comportamento normal, degradação e falha que foram simuladas. Contudo, o valor de confiança obtido com a TF não retratou adequadamente estas condições simuladas a partir do modelo do atuador.

Por fim, pode-se observar na Fig. 20, o resultado do valor de confiança para as três ferramentas de processamento de sinais considerando a falha do tipo 3 (alteração de C_a). O valor de confiança, neste caso, obtido com as TF, TFJ e TWP retratou fielmente as condições de comportamento simuladas. Entretanto, o valor de confiança obtido com a TWP apresentou os melhores resultados, ressaltando melhor a evolução da degradação (amostras 100 a 200).

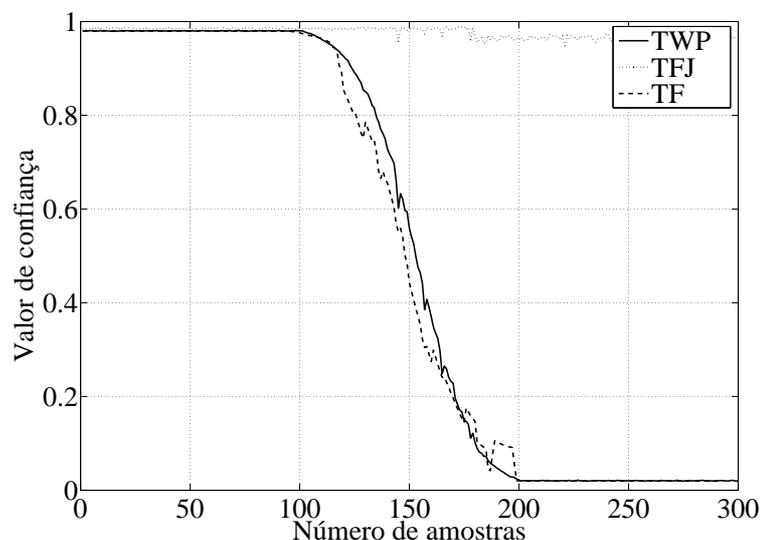


Figura 18: Valor de confiança para a falha do tipo 1 (K_h).

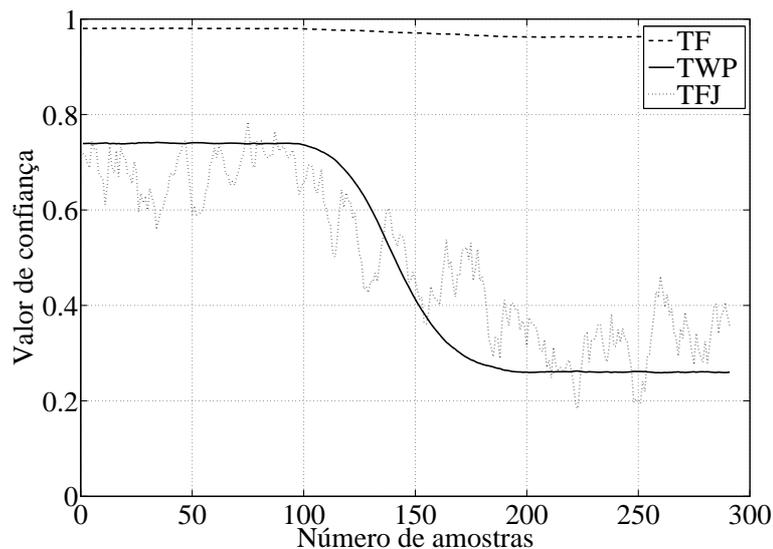


Figura 19: Valor de confiança para a falha do tipo 2 (K_m).

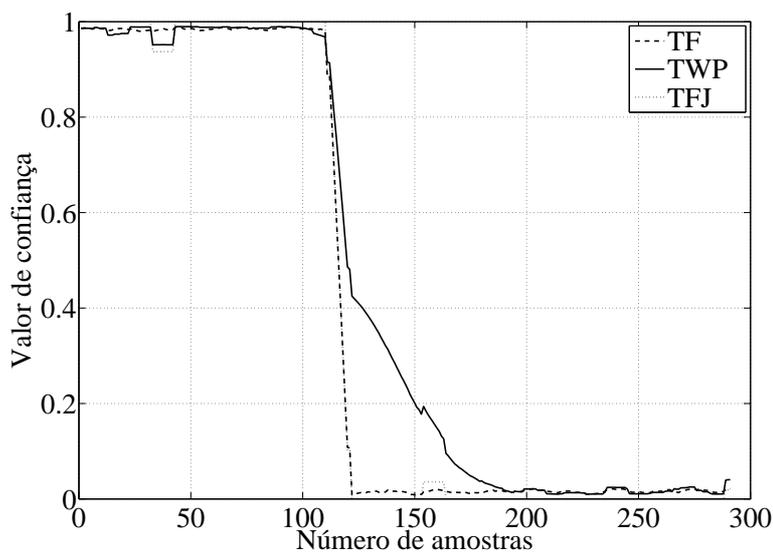


Figura 20: Valor de confiança para a falha do tipo 3 (C_a).

A partir da observação dos resultados do valor de confiança, obtidos com a regressão logística para as transformadas de Fourier, transformadas de Fourier com janelamento e transformada *wavelet packet*, visto nas Fig. 18, 19 e 20, para três tipos de falhas, nota-se que a TWP apresentou os melhores resultados.

Ainda, devido a certas características específicas dos sinais do estudo de caso, um atuador elétrico, (tais como um grande número de amostras e a necessidade de uma alta velocidade computacional) e devido aos objetivos a serem alcançados neste trabalho (tais como uma detecção e previsão de falhas) a transformada *wavelet packet*, dentre as outras ferramentas de processamento de sinais, transformada de Fourier e a transformada de Fourier com janelamento, mostrou-se a ferramenta de processamento de sinais mais recomendada para este estudo de caso.

4.2 ANÁLISE DA ENERGIA DOS SINAIS

Após a determinação da transformada *wavelet packet* como a ferramenta de processamento de sinais mais adequada para o estudo, passou-se para a geração e análise da energia dos sinais de torque e posição do modelo obtido a partir da TWP. A função *wavelet* mãe usada foi uma *Daubechies* de ordem 6, neste caso.

A energia, \mathbf{E} , contida no sinal \mathbf{y} foi particionada em termos dos coeficientes de detalhamento (D) e aproximação (A) da TWP para cada amostra do sinal:

$$\mathbf{E} = \sum_{n=1}^N |y[n]|^2 = \sum_{n=1}^N |A_k[n]|^2 + \sum_{n=1}^N |D_k[n]|^2 \quad (53)$$

Foram geradas 100 amostras, para cada categoria de comportamento (normal, degradação e falha), considerando os três tipos de injeção de falhas, como visto na seção 3.3, Tab. 3.

Os sinais de energia (obtidos a partir da TWP) também foram geradas através do pacote de ferramentas, desenvolvido pelo IMS Center, implementado para o *software* Matlab, conhecido como *Watchdog Agent Toolbox* (DJURDJANOVIC; LEE; NI, 2003).

Um exemplo dos gráficos dos sinais de energia espectral (torque e posição separados) para as situações de comportamento normal e de falha (falha do tipo 1, K_h), divididos em bandas de frequência, são vistos na Fig. 21. Pode-se observar nestes gráficos que os valores de energia nas situações de comportamento normal e de falha são distintos.

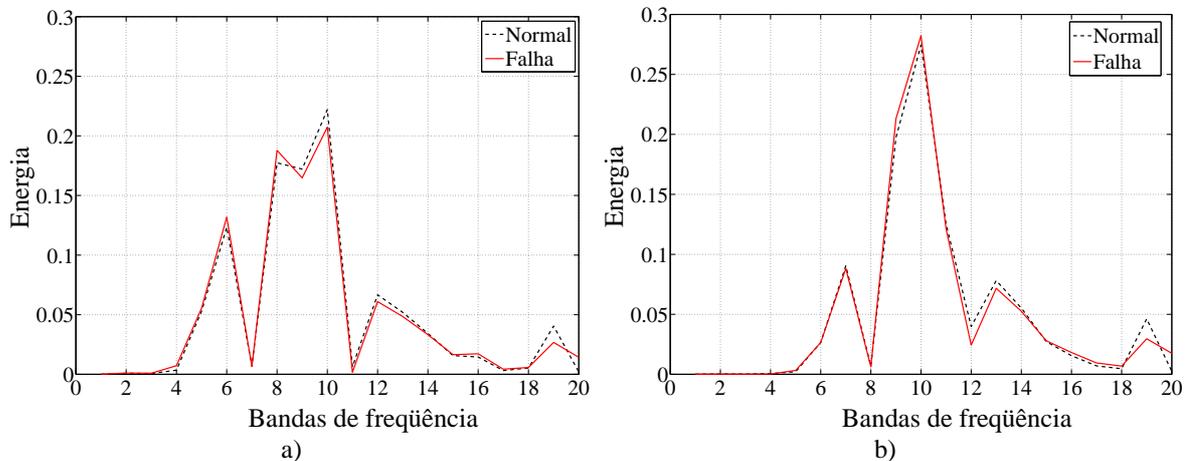


Figura 21: Curvas de energia dos sinais de: a) torque e b) posição.

A seguir, se extraiu a variância (σ_E^2) dos sinais de energia (\mathbf{E}) como visto a seguir:

$$\sigma_E^2 = \frac{1}{N} \sum_{n=0}^N |\mathbf{E} - \mu_E|^2 \quad (54)$$

onde: μ_E é a média do sinal \mathbf{E} e N é o número de amostras.

Estes valores de variância do sinal de energia foram utilizados nas etapas de treinamento e teste dos filtros adaptativos para a realização das tarefas de detecção, classificação e predição de falhas. Para fins de simplificação, a variância do sinal de energia espectral será redefinida como $\sigma_E^2 \triangleq \hat{\mathbf{E}}$.

4.3 ANÁLISE DOS MAPAS AUTO-ORGANIZÁVEIS

As próximas etapas efetuadas neste trabalho, após o cálculo da energia dos sinais de torque e posição, foram a inicialização, treinamento e teste dos mapas auto-organizáveis de Kohonen e mapas temporais de Kohonen para a realização das tarefas de detecção, classificação e predição de falhas.

Os mapas auto-organizáveis pertencem a uma classe de rede neural que apresenta um paradigma de aprendizagem não supervisionado. São representados por estruturas matriciais, onde cada elemento do mapa é denominado de neurônio (KOHONEN et al., 1995). Uma descrição mais detalhada sobre as redes neurais e MAO é apresentada no anexo B.

Os mapas temporais de Kohonen também fazem parte de uma classe de rede neural com algoritmo de aprendizado não-supervisionado, voltada para a predição de séries temporais, derivada dos mapas auto-organizáveis (CHAPPEL; TAYLOR, 1993).

Destaca-se que os mapas auto-organizáveis de Kohonen, mais especificamente, foram utilizados neste trabalho na realização das tarefas de detecção e classificação e os mapas temporais de Kohonen na realização da tarefa de predição de falhas.

A etapa de injeção de falhas (e geração dos conjuntos de dados que representam o comportamento normal, de degradação e de falha, para cada uma das falhas, do atuador) foi executada como descrito na seção 3.3. A etapa de injeção de falhas é útil e necessária para avaliar os mapas auto-organizáveis na realização das tarefas de detecção, classificação e predição de falhas (simuladas a partir do modelo matemático do atuador).

Os resultados de energia do sinal (**E**) foram obtidos a seguir pela transformada *wavelet packet* e repassados como dados de entrada para os mapas auto-organizáveis.

As etapas de inicialização, treinamento e teste dos mapas auto-organizáveis e temporais de Kohonen diferem basicamente quanto aos tipos de conjuntos de dados que são repassados como entrada.

Assim, o conjunto de dados de energia (**E**) do sinal $y[n]$ repassados aos MAO e MTK (durante as etapas de inicialização, treinamento e teste) em função do tipo da tarefa que se deseja implantar (detecção, classificação, ou previsão de falhas), foi dividido em três grupos:

1. **Mapa para detecção:** o MAO é inicializado e treinado somente com dados normais. Na etapa de teste, foram utilizados dados que representam o comportamento normal, de degradação e de falha do atuador;
2. **Mapa para classificação:** foram usados dados normais, de degradação e de falha nas etapas de inicialização, treinamento e teste dos MAO;
3. **Mapa para previsão:** semelhante à classificação, foram utilizados os três tipos de dados em todas as etapas dos MTK (normais, degradação e falha). Contudo, esses dados são repassados não um de cada vez, mas sim em lote, como uma série de dados.

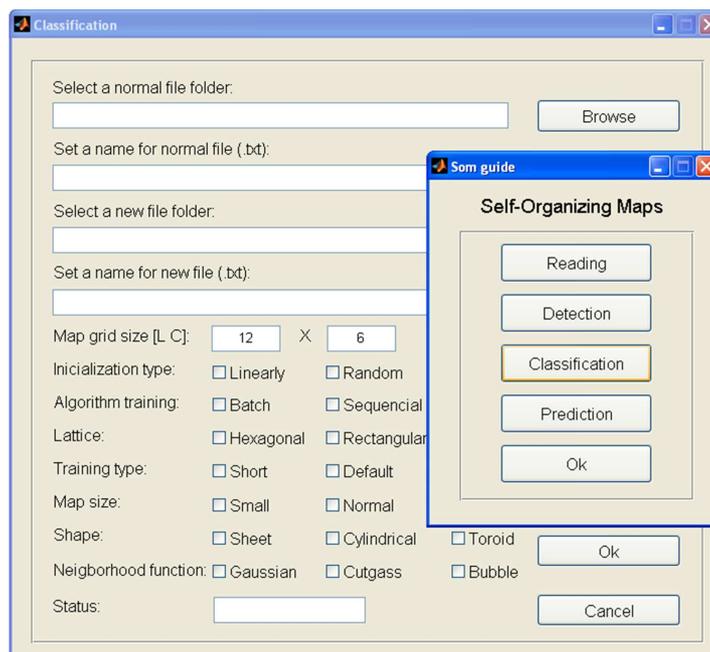


Figura 22: Janelas da interface de simulação dos mapas auto-organizáveis.

Os algoritmos dos mapas auto-organizáveis usados nas etapas de detecção, classificação, e predição de falhas foram adaptados de (VESANTO et al., 2000) e (VESANTO, 1997) e implementados para o Matlab. A janela principal e a interface de entrada de dados para a etapa de classificação dos algoritmos dos MAO podem ser vistas na Fig. 22.

Através da interface, vista Fig. 22, podem ser escolhidos os diretórios onde os arquivos (normais, de degradação e falhas) usados na etapa de classificação se encontram; além de ser possível definir as dimensões do mapa, o tipo de treinamento e a função de vizinhança, por exemplo.

4.3.1 Inicialização dos Mapas Auto-organizáveis

Na etapa de inicialização, os vetores de pesos sinápticos dos MAO e MTK, χ , aqui definidos como \mathbf{W} , isto é: $\chi \triangleq \mathbf{W}$, foram inicializados linearmente com base no cálculo dos autovalores e autovetores dos dados de energia repassados como entradas aos mapas. Um mapa estrutural foi criado inicialmente sem nenhuma organização prévia imposta ao mapa.

O conjunto de dados de entrada, ξ , aqui definidos como sendo os valores de energia, \mathbf{E} , ou seja: $\xi \triangleq \mathbf{E}$, repassados aos MAO e MTK durante a etapa de inicialização foram divididos em função do tipo da tarefa que se deseja realizar, como descrito anteriormente.

4.3.2 Treinamento dos Mapas Auto-organizáveis

Após uma inicialização e treinamento, as características específicas dos MAO e MTK permitem a realização, em especial, de três tarefas: detecção, classificação e predição de falhas, vistas a seguir.

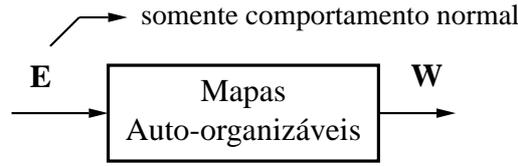


Figura 23: Etapa de treinamento dos MAO para a tarefa de detecção.

4.3.2.1 Treinamento dos Mapas Auto-organizáveis para a Tarefa de Detecção

O treinamento dos MAO para a etapa de detecção foi realizado repassando como entrada ao mapa os dados de energia espectral, \mathbf{E} , dados que representam somente o comportamento do tipo normal, como visto na Fig. 23.

As fases de competição, cooperação e adaptação para a etapa de treinamento dos MAO foram realizadas com a análise da similaridade sendo efetuada com base no cálculo da distância Euclidiana, D , ver anexo B para maiores detalhes, através da seguinte expressão:

$$D^{kj} = \|\mathbf{E}^k - \mathbf{W}^j\| = \sqrt{\sum_{n=1}^N (E_n^k - W_n^j)^2} \quad (55)$$

onde: \mathbf{E} é o vetor de entradas do mapa, \mathbf{W} é o vetor de pesos sinápticos J é o número total de neurônios no mapa, K é o número de ciclos de amostragem e N o número de elementos dos vetores \mathbf{E} e \mathbf{W} , sendo n, j, J, k e $N \in \mathbb{N}$ com $1 \leq n \leq N$, $1 \leq j \leq J$ e $1 \leq k \leq K$.

O neurônio que apresentar o menor valor para o cálculo da distância Euclidiana é definido como neurônio vencedor, segundo:

$$\mathbf{W}^{j^*} = \underset{j}{\operatorname{argmin}} \|\mathbf{E}^k - \mathbf{W}^j\| \quad (56)$$

onde: j^* é o índice do neurônio vencedor \mathbf{W} do MAO e “arg” significa o “índice”.

O vetor de pesos sinápticos \mathbf{W}^j do neurônio j , foi atualizado de forma a se aproximar mais do vetor \mathbf{E}^k , de acordo com a seguinte equação:

$$\mathbf{W}^j(n+1) = \mathbf{W}^j(n) + \alpha(n) h(n) [\mathbf{E}^k(n) - \mathbf{W}^j(n)] \quad (57)$$

onde: $\alpha(n) = 1$ e $h(n)$ é função uma gaussiana.

4.3.2.2 Treinamento dos Mapas Auto-organizáveis para a Tarefa de Classificação

O treinamento dos MAO para a etapa de classificação foi feito repassando como entrada para o mapa os dados de energia, \mathbf{E} . Contudo, neste caso, diferente da tarefa de detecção, os dados representam o comportamento do tipo normal, de degradação e falha (para cada tipo de falha que se deseja classificar), como visto na Fig. 24.

Isto foi necessário para que fosse realizada a atribuição dos rótulos aos agrupamentos (um função dos conjuntos de dados) formados pelo mapa. Assim, o mapa foi dividido em regiões distintas e cada grupo de neurônios foi representado por rótulos diferentes.

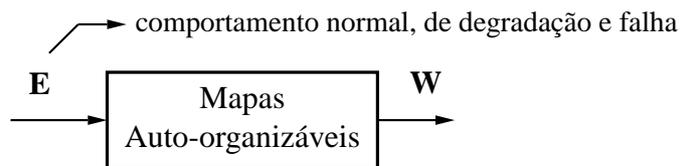


Figura 24: Etapa de treinamento dos MAO para as tarefas de classificação e predição.

O treinamento dos mapas auto-organizáveis, para a tarefa de classificação, foi realizado do mesmo modo como descrito para a tarefa de detecção, usando as Eq. (55), (56) e (57) para treinar o mapa e formar os agrupamentos.

4.3.2.3 *Treinamento dos Mapas Temporais de Kohonen para a Tarefa de Predição*

Os mapas temporais de Kohonen, semelhante aos MAO, utilizam um treinamento de aprendizado não-supervisionado voltados para a predição (ver anexo B).

O treinamento dos MTK para a etapa de predição foi realizado repassando como entrada para o mapa dados de energia (**E**) de modo semelhante ao efetuado para os MAO na tarefa de classificação, como visto na Fig. 24. Isto é, foram utilizados no treinamento dados que representam o comportamento do tipo normal, de degradação e falha (para cada tipo de falha).

As etapas de treinamento, em especial, de inicialização e teste dos mapas temporais de Kohonen (a não ser pelo cálculo dos neurônios vencedores) são as mesmas etapas dos algoritmos dos mapas auto-organizáveis, descritas anteriormente. Nos MTK, o neurônio escolhido como vencedor é aquele que apresentar o maior valor para o cálculo da ativação temporal, A^{kj} . O neurônio vencedor, ou a posição desse no mapa, é obtido a partir de:

$$\mathbf{W}^{j^*} = \underset{j}{\operatorname{argmax}} (A^{kj}) \quad (58)$$

onde: j^* é o índice do neurônio vencedor **W** do MTK e “arg” significa o “índice”.

A expressão para o cálculo da ativação temporal é dada por:

$$A^{kj} = \lambda A^{k-1,j} - \frac{1}{2} (D^{kj})^2 \quad (59)$$

onde: D^{kj} é a distância Euclidiana e λ é o tamanho da memória.

4.3.3 **Teste dos Mapas Auto-organizáveis**

Após o treinamento dos MAO e MTK foi realizada a etapa de teste. Nessa etapa, os resultados de teste apresentados, a seguir, diferem quanto à tarefa que se pretende executar: detecção, classificação ou predição de falhas.

4.3.3.1 *Resultados de detecção*

Na etapa de teste, novos dados de energia foram gerados para serem avaliados. Contudo, agora os dados repassados para os MAO representam os comportamentos: normal, de degradação e falha, como visto na Fig. 25.

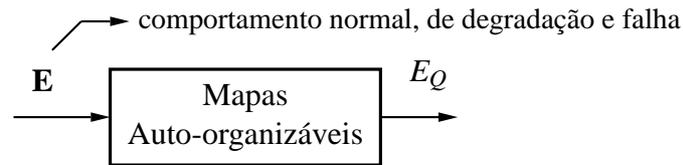


Figura 25: Etapa de teste dos MAO para a tarefa de detecção e classificação.

A tarefa de detecção de falhas visa sinalizar ao usuário a ocorrência de falhas no atuador. Quando um limite pré-estabelecido para o erro de quantização, $\varepsilon_q \triangleq E_Q$, Eq. (139), é atingido, ocorre a detecção e a sinalização de uma falha.

O erro de quantização corresponde a raiz quadrada da soma dos erros quadráticos, que é a própria distância Euclidiana, ou seja:

$$E_Q(k) = \sqrt{\sum_{n=1}^N (e_n^k)^2} = \sqrt{\sum_{n=1}^N (E_n^k - W_n^{j^*})^2} \quad (60)$$

onde: o erro (e_n^k) corresponde à diferença entre a componente n do vetor de pesos sinápticos do neurônio vencedor (\mathbf{W}^{j^*}) e do novo dado de entrada (\mathbf{E}^k), j^* é índice do neurônio vencedor no mapa e K é o número de amostras.

O gráfico do erro de quantização, usado para detecção de falhas, para os três tipos de falhas simuladas (K_h , K_m e C_a), obtidos a partir dos mapas auto-organizáveis, é visto na Fig. 26. Pode se observar na Fig. 26 que o mapa auto-organizável detecta corretamente as regiões de comportamento normal, degradação e falha que foram simuladas pelo modelo para cada tipo de falha.

Na Fig. 26 pode ser observado que o erro de quantização, para os três tipos de falhas analisadas, sofre um aumento gradual entre as amostras de número 100 e 200 (que corresponde a região de alteração gradual dos parâmetros). A partir da amostra de número 200 o limiar de detecção da falha é alcançado e um alarme pode ser acionado, por exemplo.

Além disso, o erro de quantização apresenta níveis diferentes, para as diferentes falhas. A falha do tipo 1 (K_2) apresentou um erro de quantização na faixa de 0.01 ($E_Q \cong 0.0095$), já para a falha do tipo 2 (K_m), o erro de quantização ficou na faixa de 0.05 ($E_Q \cong 0.048$) e para a falha do tipo 3 (C_a) o erro ficou em torno de 0.06 ($E_Q \cong 0.062$).

4.3.3.2 Resultados de classificação

Na etapa de teste, voltada para a classificação, novos dados de energia espectral foram gerados pelo modelo para serem classificados. Os novos dados repassados para os mapas auto-organizáveis representam os comportamentos normal, de degradação e falha, como visto na Fig. 25.

A tarefa de classificação é útil para mostrar ao usuário qual é o tipo de falha específica que ocorreu (ou está ocorrendo) no atuador. Assim, quando uma falha é detectada, a classificação é útil para determinar o tipo de falha específico.

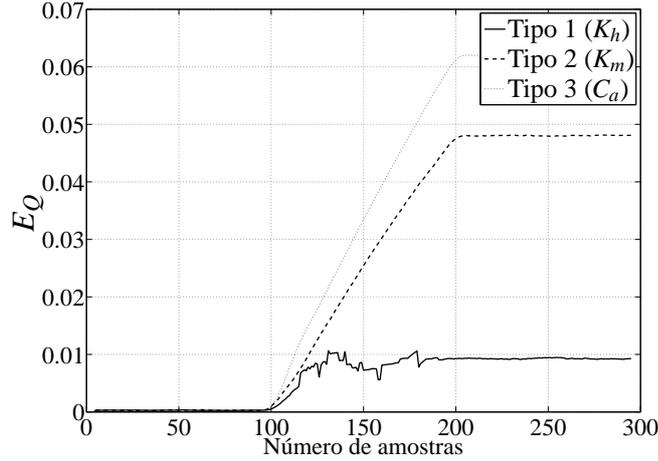


Figura 26: Erro de quantização para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).

A classificação dos três tipos de falhas, neste caso, foi feita através da determinação dos neurônios vencedores, Eq. (56), e da matriz de distâncias unificadas (MDU), como visto na sessão 8.3.9. A matriz de distâncias unificadas, M_D , é definida como a seguir:

$$M_D = \begin{bmatrix} \mathbf{W}^1 & \mathbf{W}^{(L_m+1)} & \mathbf{W}^{(2L_m+1)} & \dots & \mathbf{W}^{[L_m(C_m-1)+1]} \\ \mathbf{W}^2 & \mathbf{W}^{(L_m+2)} & \mathbf{W}^{(2L_m+2)} & \dots & \mathbf{W}^{[L_m(C_m-1)+2]} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{W}^{L_m} & \mathbf{W}^{2L_m} & \mathbf{W}^{3L_m} & \dots & \mathbf{W}^{L_m C_m} \end{bmatrix} \quad (61)$$

onde: \mathbf{W} são os pesos sinápticos, Eq (57), L_m é o número de linhas da matriz, C_m é o número de colunas, sendo L_m e $C_m \in \mathbb{N}$.

A Fig. 27 mostra os resultados de classificação com a MDU sendo obtida para os três tipos de falhas analisados: 1 (K_h), 2 (K_m) e 3 (C_a). Na Fig. 27 os neurônios aparecem agrupados em torno das condições de operação: normal (N), degradação do tipo 1 (Dh), falha do tipo 1 (Kh), degradação do tipo 2 (Dm), falha do tipo 2 (Km), degradação do tipo 3 (Da) e falha do tipo 3 (Ca), totalizando sete agrupamentos. Cada agrupamento, correspondendo a diferentes condições de operação, foi assinalado com uma cor diferente.

Para obter a MDU vista na Fig. 27 foram repassados, inicialmente, durante o treinamento, um conjunto de dados contendo os seguintes tipos de dados: normais, de degradação e de falha (para cada tipo de falha: K_h , K_m and C_a). Assim, o mapa topológico foi montado e os rótulos e agrupamentos localizados no mapa ¹.

Em seguida, na etapa de teste, um novo conjunto de dados gerados pelo modelo e depois repassados a TWP para se obter a energia, contendo dados normais, de degradação e de falha para cada tipo de falha foi novamente apresentado ao MAO, gerando a MDU vista na Fig. 27.

Supondo agora que, durante a etapa de teste, um novo conjunto de dados contendo dados normais, de degradação e de falha, para somente um tipo de falha, a falha do tipo 1 (K_h) por exemplo, seja agora apresentado ao mapa auto-organizável. O resultado deste procedimento pode ser visto na Fig. 28.

¹ $L_m = 20$, $C_m = 10$, e $h(n)$ é uma gaussiana.

Pode se observar na Fig. 28 que os dados normais (N), de degradação (Dh) e de falha (falha do tipo 1, Kh) foram mapeados corretamente no mapa. Isto é, o mapa auto-organizável classificou fielmente as condições de comportamento normal, degradação, e falha, para o tipo de injeção de falha que foi simulada, agrupando estes dados nas suas respectivas regiões.

Resultado semelhante pode ser obtido considerando agora que, durante a etapa de teste, um novo conjunto de dados gerado pelo modelo, contendo dados normais, de degradação e de falha, neste caso somente para a falha do tipo 2 (K_m), sejam apresentados ao mapa. O resultado desta classificação é visto na Fig. 29.

Procedendo do mesmo modo, durante a etapa de teste, agora considerando somente a falha do tipo 3 (alteração da constante de atrito C_a), obtém-se o mapa voltado para a classificação desse tipo de falha, visto na Fig. 30.

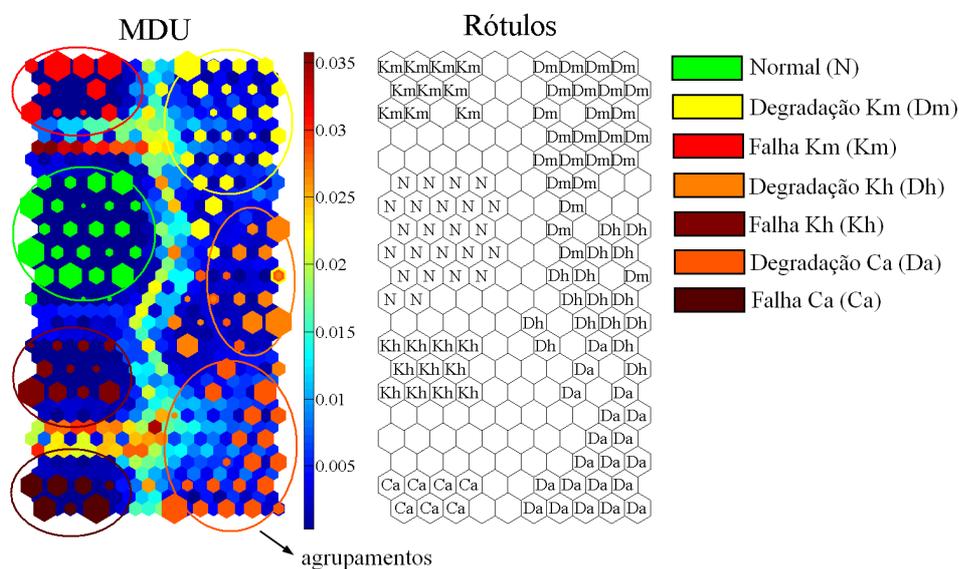


Figura 27: Resultados de classificação para as falhas do tipo: 1 (K_h), 2 (K_m) e 3 (C_a).

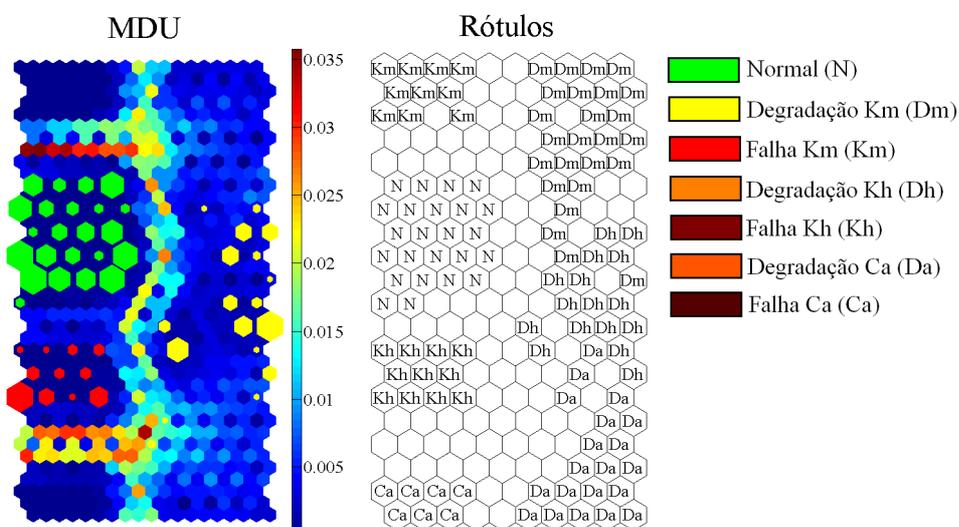


Figura 28: Resultado de classificação para a falha do tipo 1 (K_h).

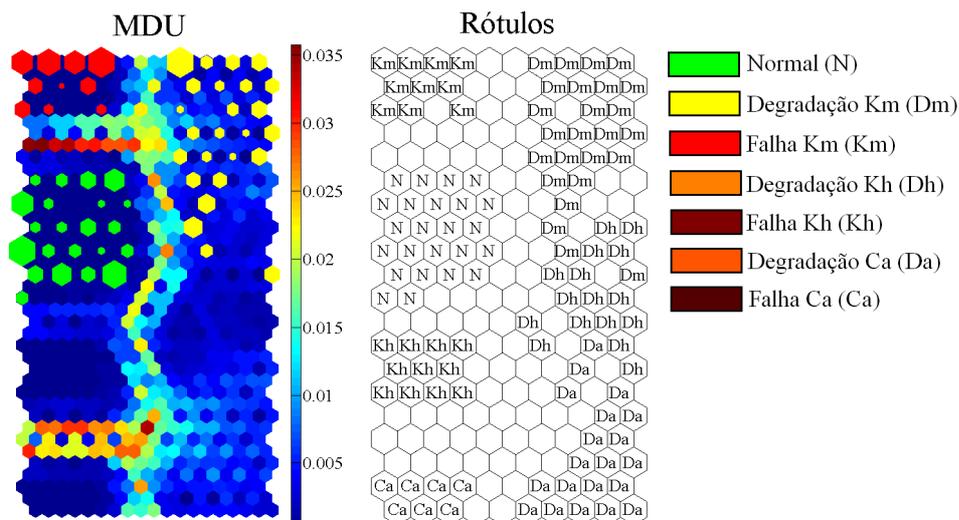


Figura 29: Resultado de classificação para a falha do tipo 2 (K_m).

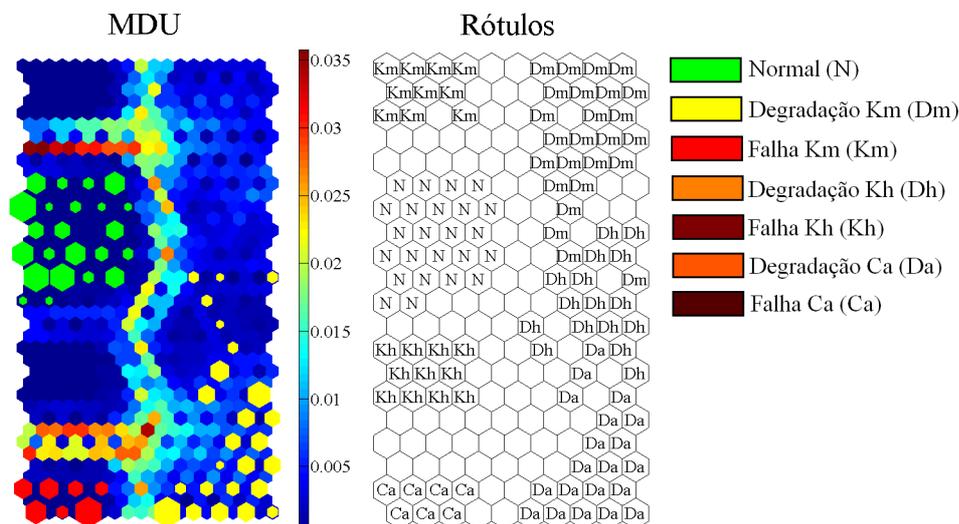


Figura 30: Resultado de classificação para a falha do tipo 3 (C_a).

4.3.3.3 Resultados de predição

Novos dados de energia (E) foram gerados para serem classificados durante a etapa de teste, voltada para a predição. Esses novos dados repassados para os MTK representam os comportamentos normal, de degradação e falha, como visto na Fig. 31.

A tarefa de predição é útil para informar ao usuário sobre o início de uma degradação e sobre a sua evolução, através de uma trajetória traçada no mapa, até a ocorrência da falha bem como para obter o tempo de vida útil do atuador, como visto a seguir, por exemplo.

A trajetória traçada nos MTK foi obtida com base da determinação dos neurônios vencedores (W^{j*}) e na ativação (A^{kj}) para cada série temporal.

Na condição de comportamento normal, os neurônios vencedores seguem um caminho dentro da região mapeada para o comportamento do tipo normal.

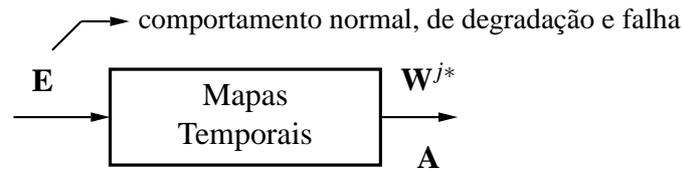


Figura 31: Etapa de teste dos mapas temporais de Kohonen para a tarefa de predição.

Durante a evolução da degradação até a ocorrência da falha, o neurônio vencedor escolhido desvia-se da região mapeada como normal para a região mapeada como falha. A amplitude e a velocidade do desvio dependem do tipo e gravidade da falha em questão.

A Fig. 32 exhibe os resultados de predição, a trajetória dos neurônios vencedores, obtidos a partir dos mapas temporais de Kohonen para a falha do tipo 1 (K_h), considerando o comportamento do tipo normal, de degradação e falha do atuador.

Inicialmente, para se obter a matriz de distâncias unificadas, vista na Fig. 32, foram repassados ao mapa temporal de Kohonen, semelhante a etapa de classificação, durante o treinamento, um conjunto de dados contendo os seguintes tipos de dados: normais, de degradação e de falha (para cada tipo de falha: K_m , K_h and C_a). Assim, o mapa topológico foi montado e os rótulos e agrupamentos localizados no mapa ².

Depois, na etapa de teste, um novo conjunto de dados de energia contendo dados normais, de degradação e de falha somente para o tipo 1 de falha, nesta seqüência, foi apresentado ao mapa, gerando por fim a trajetória vista na Fig. 32.

Pode-se observar nessa figura que a trajetória traçada no mapa começa a partir da região cujos neurônios foram mapeados e rotulados como normais (N), passando através da região dos neurônios classificados como de degradação (Dh), chegando a região que representa o tipo de falha que foi simulado e analisado pelo MTK, neste caso o tipo 1 (K_h). Resultados semelhantes foram obtidos para os outros dois tipos de falhas.

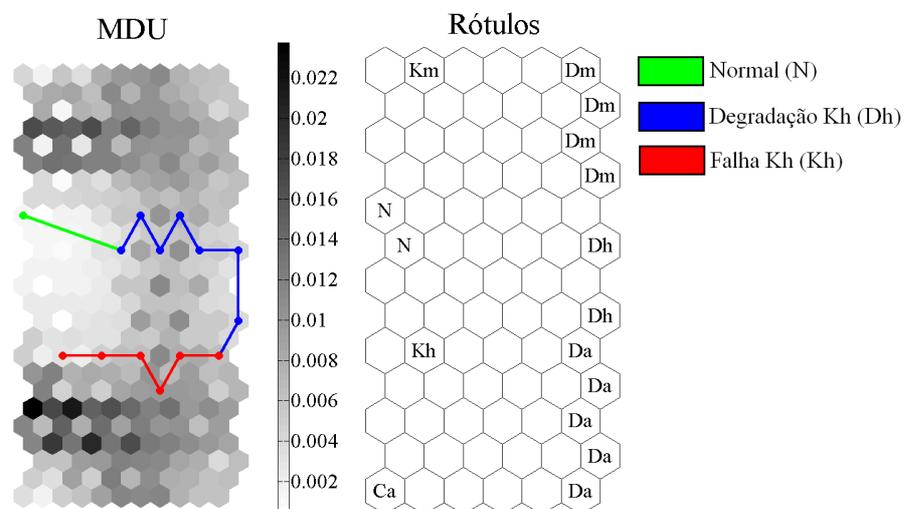


Figura 32: Resultado de predição para falha do tipo 1 (K_h).

² $L_m = 20$, $C_m = 10$, e $h(n)$ é uma gaussiana.

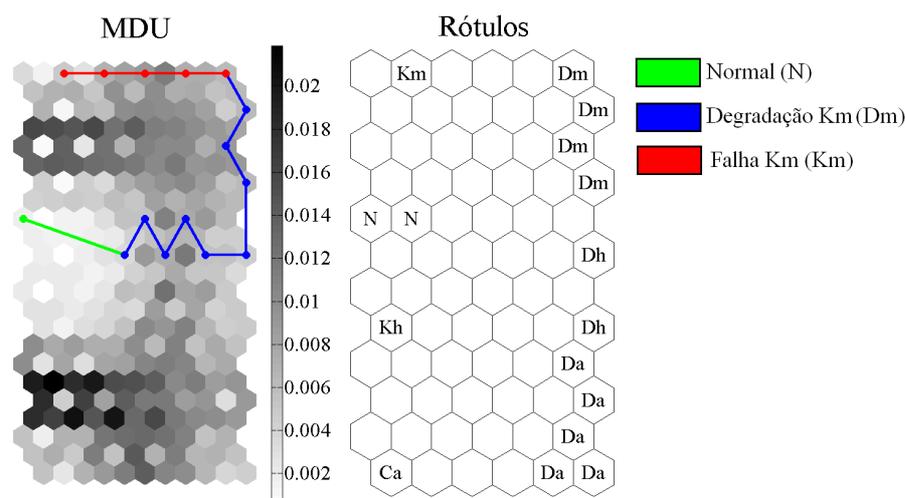


Figura 33: Resultado de predição para falha do tipo 2 (K_m).

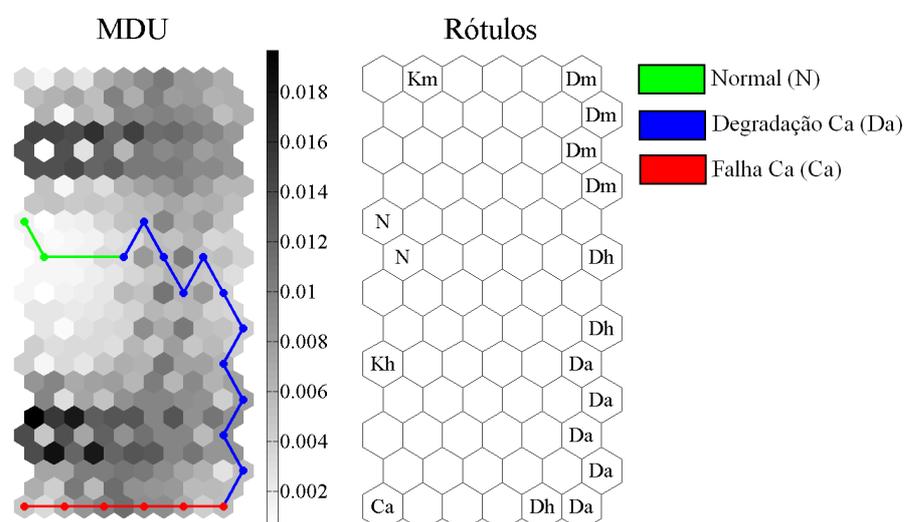


Figura 34: Resultado de predição para falha do tipo 3 (C_a).

Considerando agora que durante a etapa de teste, um novo conjunto de dados, contendo dados normais, de degradação e de falha, neste caso somente para a falha do tipo 2 (K_m), sejam agora apresentados nesta seqüência ao mapa. O resultado de predição, para este caso, falha do tipo 2, é visto na Fig. 33.

Pode-se observar nessa figura, semelhante à falha do tipo 1, que a trajetória traçada no mapa começa a partir da região cujos neurônios foram mapeados e rotulados como normais (N), passando através da região dos neurônios classificados como de degradação (Dm), chegando a região do tipo de falha analisado, neste caso o tipo 2 (K_m).

Supondo agora que durante a etapa de teste, um novo conjunto de dados, contendo dados (E) normais, de degradação e de falha, somente para a falha do tipo 3 (C_a), sejam apresentados ao mapa. O resultado de predição, neste caso, é visto na Fig. 34. Pode-se observar nessa figura que a trajetória traçada no mapa começa a partir da região mapeada como normal (N), passando através da região de degradação (Da), chegando a região de falha (Ca).

O próximo passo realizado, na etapa de predição, foi extrair a variância (\hat{W}) dos neurônios vencedores (\mathbf{W}^{j^*}), cujas trajetórias podem ser visualizadas nas Fig. 33, 32 e Fig. 34, segundo a Eq (54), aqui re-escrita:

$$\hat{W} = \frac{1}{N} \sum_{n=0}^N |W^{j^*}[n] - \mu_W|^2 \quad (62)$$

onde: μ_W é a média de \mathbf{W}^{j^*} e N é o número de amostras.

A Fig. 35 exibe os resultados de predição, a variância dos neurônios vencedores, obtidos a partir dos mapas temporais de Kohonen para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a) considerando o comportamento do tipo normal, de degradação e falha (para cada tipo de falha) do atuador.

Para se obter as curvas, vistas na Fig. 35, foram obtidos do mapa temporal de Kohonen, durante o teste, na etapa de predição, o vetor dos neurônios vencedores \mathbf{W}^{j^*} , para os três tipos de falhas. A seguir foi extraído a variância desse vetor, como visto na Eq. (62).

Esses resultados de variância, e os resultados da primeira e segunda derivadas de \hat{W} , serão utilizados, como visto a seguir, para se efetuar a predição do tempo de vida útil do atuador.

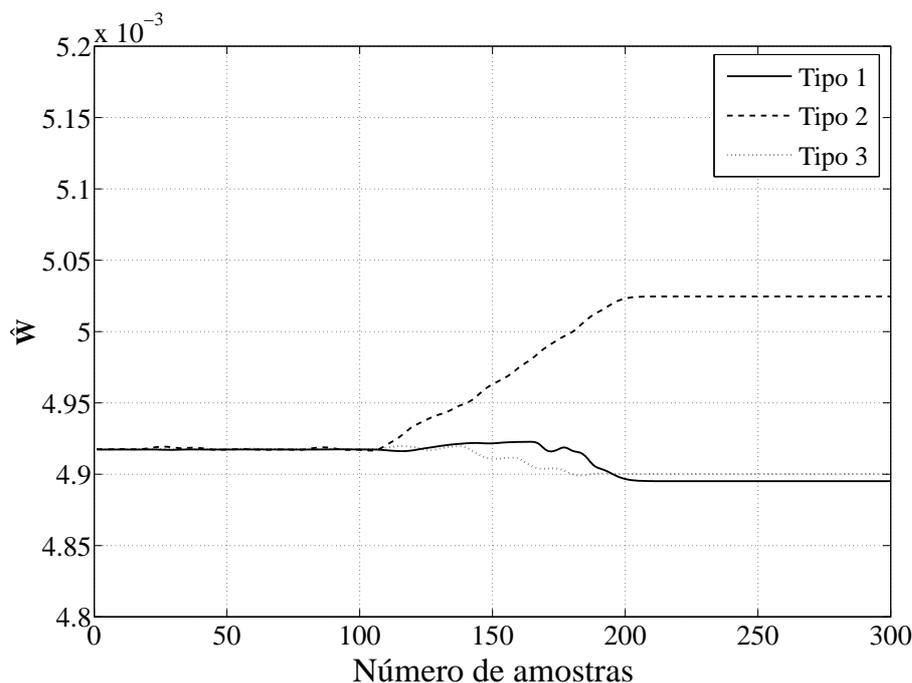


Figura 35: Gráficos de predição para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).

4.4 ANÁLISE DOS FILTROS ADAPTATIVOS

De modo semelhante ao realizado para os mapas auto-organizáveis e mapas temporais de Kohonen, os filtros adaptativos também foram utilizados na realização das tarefas de detecção, classificação e predição de falhas.

Um sistema adaptativo (um FA, por exemplo) é aquele cuja estrutura é alterável (através do ajuste dos seus coeficientes, por exemplo) de tal modo que seu comportamento se ajuste, de acordo com algum critério de desempenho previamente estabelecido, ao ambiente ao qual ele será inserido. Uma descrição mais detalhada sobre os FA é vista no anexo C.

Filtros adaptativos são filtros com coeficientes variáveis que se adaptam ao longo do tempo, automaticamente, aos sinais de entrada. O processo de adaptação aqui utilizado é baseado no algoritmo dos mínimos quadrados médios (HAYKIN, 2002), como descrito no anexo C, seção 9.3.

A geração dos conjuntos de dados que representam o comportamento normal, de degradação e de falha, para cada uma das falhas do atuador, bem com a etapa de injeção de falhas, foi realizada como descrito na seção 3.3.

A etapa de injeção de falhas é necessária para avaliar a robustez dos filtros adaptativos na realização das tarefas de detecção, classificação e predição de falhas, simuladas a partir do modelo do atuador elétrico.

Os resultados de covariância do sinal de energia espectral, $\hat{\mathbf{E}}$, Eq. (54), foram obtidos a seguir pela transformada *wavelet packet* e repassados como dados de entrada para os filtros adaptativos.

As etapas de treinamento e teste dos FA diferem basicamente quanto aos tipos de conjuntos de dados desejados (\mathbf{d}) que são repassados aos FA, de modo semelhante aos MAO.

Assim, o conjunto de dados de energia ($\hat{\mathbf{E}}$) do sinal $y[n]$, Eq 50, repassados aos FA (durante as etapas de treinamento e teste) em função do tipo da tarefa que se deseja realizar (detecção, classificação, ou previsão), foi dividido em três grupos:

1. **Filtro para detecção:** o FA foi treinado somente com dados normais. Na etapa de teste, são utilizados dados que representam o comportamento normal, de degradação e de falha do atuador;
2. **Filtro para classificação:** foram usados somente dados de falha, durante o treinamento. Na etapa de teste são usados dados normais, de degradação e de falha;
3. **Filtro para previsão:** não há uma etapa de treinamento. Foram utilizados os três tipos de dados na etapa de teste dos FA (normais, degradação e falha). Entretanto, estes dados, são repassados como uma série de dados.

Os algoritmos dos filtros adaptativos usados nas etapas de detecção, classificação, e predição de falhas foram adaptados de (NEGREIROS, 1994) e programados no Matlab.

A janela principal e a interface de entrada de dados para a etapa de detecção dos algoritmos dos MAO podem ser vistas na Fig. 36. Através da interface, vista Fig. 36, pode-se escolher os diretórios onde os arquivos (normais, de degradação e falhas) usados na etapa de detecção se encontram, além de ser possível também definir o número de amostras, e o tamanho do passo, por exemplo.

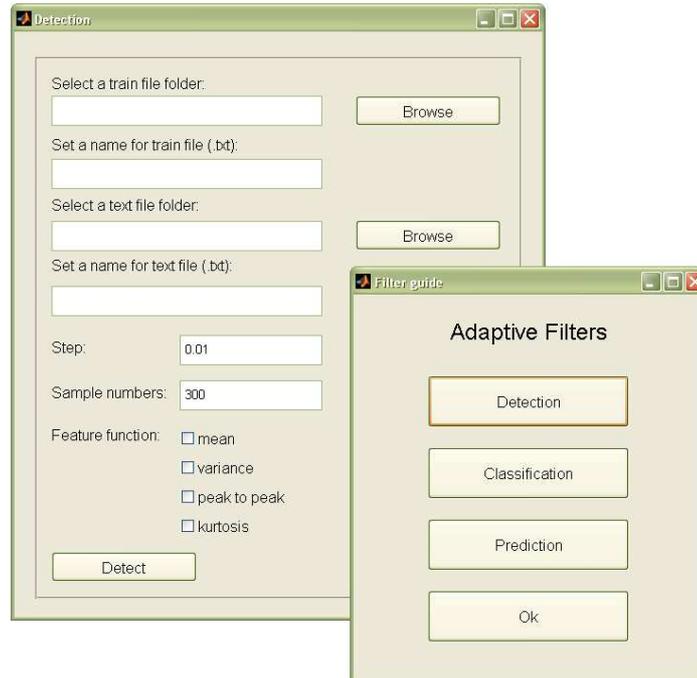


Figura 36: Janelas da interface de simulação dos filtros adaptativos.

4.4.1 Inicialização dos Filtros Adaptativos

Na etapa de inicialização, os vetores de pesos sinápticos do filtro adaptativo, \mathbf{w} , foram inicializados de modo semelhante para as tarefas de detecção, classificação e predição de falhas, como visto na Eq. (182), aqui re-escrita:

$$\mathbf{w}[0] = 0 \quad (63)$$

4.4.2 Treinamento dos Filtros Adaptativos

Após uma inicialização, e treinamento, as características específicas dos filtros adaptativos permitem a realização, em especial, de três tarefas: detecção, classificação e predição (para a tarefa de predição não é necessário realizar um treinamento prévio) de falhas, vistas a seguir.

4.4.2.1 Treinamento dos Filtros Adaptativos para a Tarefa de Detecção

O treinamento dos FA, para a etapa de detecção, foi feito repassando como saída do sistema livre de falhas (sinal desejado) os dados de variância de energia espectral, \hat{E} (somente dados normais), obtidos a partir do modelo, Eq. (54), como na Fig. 37.

Assim, o treinamento dos FA para a etapa de detecção, visto na Fig. 37, foi realizado como descrito no anexo C, seção 9.3, com a análise da similaridade realizada com base no cálculo do erro da saída do filtro, \mathbf{e} , através da seguinte expressão:

$$e[n] = \hat{E}[n] - y[n] \quad (64)$$

onde: $\hat{E}[n]$ é o valor desejado, $y[n]$ é a saída do filtro e n é o número da amostra.

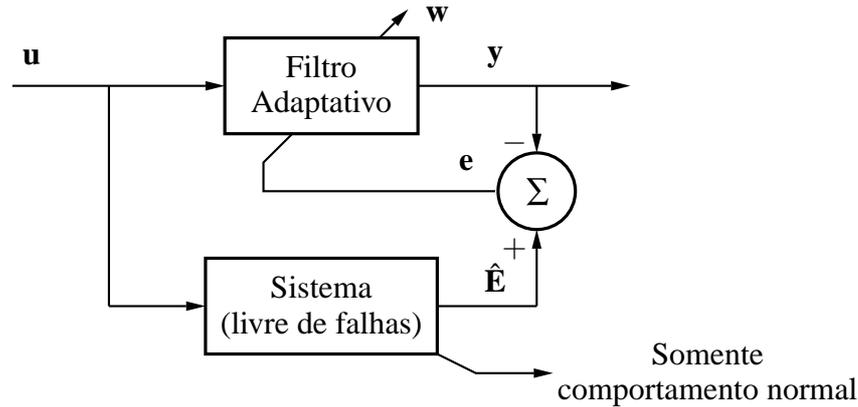


Figura 37: Etapa de treinamento dos filtros adaptativos para a tarefa de detecção.

A expressão para se obter a saída do filtro é dada por:

$$y[n] = \mathbf{w}^T [n] \mathbf{u}[n] \quad (65)$$

onde: k é o número da amostra e o vetor $u[n] \triangleq -1 + 0.001 * (\text{rand}([1]) - 0.5)^3$.

A atualização do vetor de coeficientes é realizada utilizando a expressão vista na Eq. (185), aqui re-escrita sem o símbolo do complexo conjugado para fins de simplificação:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu \mathbf{u}[n] e[n] \quad (66)$$

Estes passos, vistos nas Eq. (64), (65) e (66) são repetidos até que o valor de erro atinja um valor próximo de zero, $e[k] \cong 0$; isto é, quando o filtro apresentar um comportamento aproximadamente igual ao sistema livre de falhas.

4.4.2.2 Treinamento dos Filtros Adaptativo para a Tarefa de Classificação

Na tarefa de classificação, o treinamento dos FA foi executado repassando como saída do sistema com falhas (sinal desejado) os dados de variância de energia espectral $\hat{\mathbf{E}}$ (neste caso, dados que representam somente o comportamento de falha que se deseja classificar), como visto na Fig. 38.

Deste modo, o ajuste dos coeficientes (\mathbf{w}), Eq. (66), realizado durante o treinamento dos FA para a etapa de classificação, visto na Fig. 38, foi feito como descrito para a tarefa de detecção, com a análise da similaridade efetuada com base no cálculo da saída do filtro (\mathbf{y}) e do erro (\mathbf{e}), através das Eq. (64) e (65), respectivamente.

De modo semelhante, os passos vistos nas Eq. (64), (65) e (66) são repetidos até que o valor de erro atinja um valor próximo de zero e o filtro emule o sistema com falha.

4.4.3 Teste dos Filtros Adaptativos

Após o treinamento dos FA foi realizada a etapa de teste. Nessa etapa, de modo semelhante aos MAO, os resultados de teste, vistos a seguir, diferem basicamente quanto ao tipo de tarefa que se deseja realizar: detecção, classificação ou predição de falhas.

³ $\text{randn}(\cdot)$ é o ruído branco.

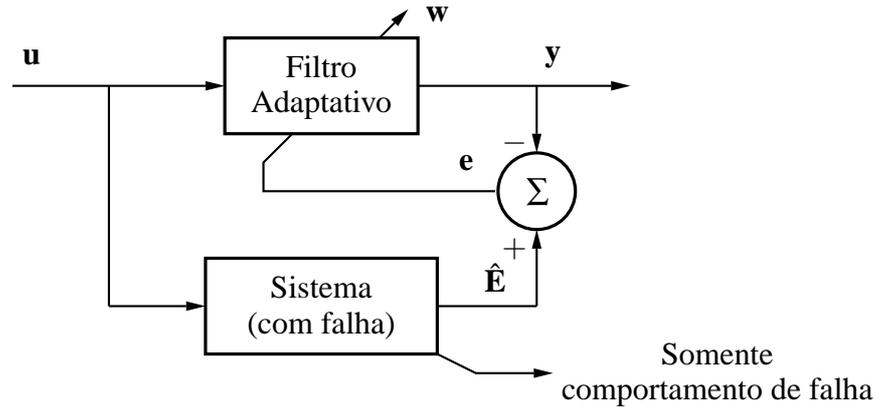


Figura 38: Etapa de treinamento dos filtros adaptativos para a tarefa de classificação.

4.4.3.1 Resultados de detecção

Na etapa de teste, voltada para detecção, novos dados de variância de energia espectral (\hat{E}) foram gerados pelo modelo para serem avaliados. Contudo, agora os dados repassados para os FA representam os comportamentos normal, de degradação e falha (para cada tipo de falha que se deseja detectar), como visto na Fig. 39.

Destaca-se que os coeficientes obtidos durante o treinamento dos FA, w , foram agora usados durante a etapa de teste. Ou seja, o FA emula o comportamento normal do sistema. A saída do FA é então comparada com a saída do sistema em teste, gerando o sinal de erro, e :

$$e[n] = \hat{E}[n] - y[n] \quad (67)$$

A tarefa de detecção de falhas é útil para avisar ao operador a ocorrência de alguma falha no atuador. Quando um limite pré-estabelecido para o erro, $e[k]$, Eq. (67), é atingido, ocorre a detecção e a sinalização de uma falha.

O gráfico do valor absoluto do erro, usado para detecção de falhas, para os três tipos de falhas simuladas (K_h , K_m e C_a), alcançado a partir dos filtros adaptativos, é visto na Fig. 40. Pode se observar na Fig. 40 que o FA realiza a detecção corretamente, detectando as três falhas simuladas pelo modelo.

Para obter o gráfico do erro visto na Fig. 40 foram repassados ao FA, durante a etapa de teste, um conjunto de dados contendo dados que representam o comportamento normal (100 amostras), de degradação (100 amostras) e de falha (100 amostras), nesta ordem, para cada tipo de falha, como visto na seção 3.3.

Nota-se, na Fig. 40, que o sinal de erro, para os três tipos de falhas analisadas, sofre um aumento entre as amostras de número 100 e 200 (que corresponde a região de alteração gradual dos parâmetros). A partir da amostra de número 200, o nível de detecção das falhas é alcançado e um alarme pode ser acionado, por exemplo.

O erro apresentou dois níveis diferentes para as três distintas falhas. A falha do tipo 2 (K_m) apresentou um erro na faixa de 1.2×10^{-4} . Já para as falhas do tipo 1 (K_h) e 3 (C_a), o erro ficou em torno de 0.2×10^{-4} ($\cong 0.18 \times 10^{-4}$ e 0.23×10^{-4} , respectivamente).

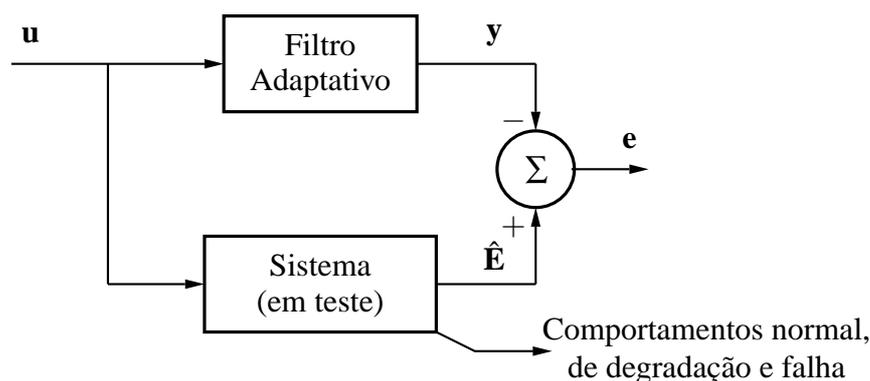


Figura 39: Etapa de teste dos FA para a tarefa de detecção e classificação.

4.4.3.2 Resultados de Classificação

Na etapa de teste, voltada para a classificação de falhas, novos dados de variância da energia espectral (\hat{E}) foram gerados para serem classificados. Os dados repassados para os FA representam os comportamentos normal, de degradação e falha (para cada tipo de falha que se deseja classificar), como visto na Fig. 39.

Os coeficientes obtidos durante o treinamento dos FA, w , agora são usados durante a etapa de teste. Ou seja, o FA emula o comportamento de falha do sistema. A saída do FA é então comparada com a saída do sistema em teste, gerando um sinal de erro, e , conforme a Eq. (67).

Contudo, diferentemente da tarefa de detecção onde ocorre a detecção de falhas quando um limite pré-estabelecido para o erro é atingido, aqui a classificação de falhas é realizada quando o valor do erro atinge um valor próximo de zero.

O gráfico do valor absoluto do erro, usado na classificação de falhas, para a falha do tipo 1 (K_h), obtido a partir dos FA, é visto na Fig. 41.

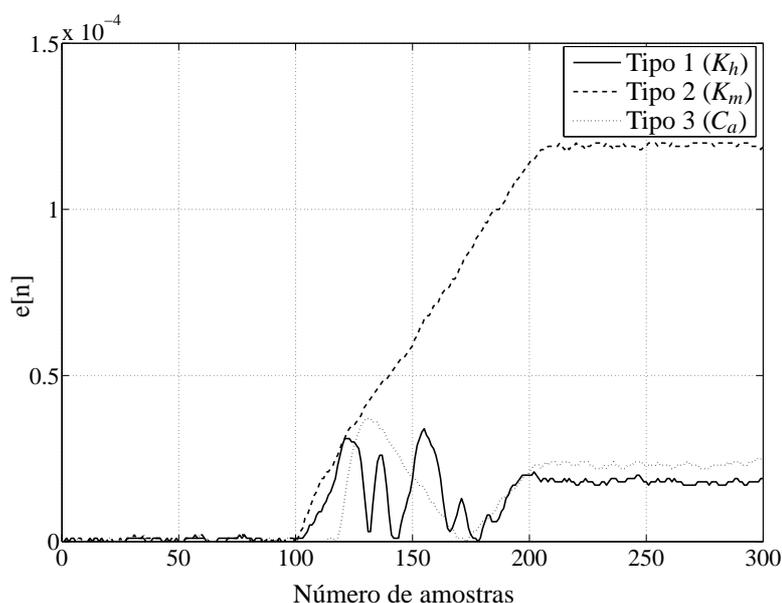


Figura 40: Erro de detecção para as falhas do tipo 1 (K_h), 2 (K_m) e 3 (C_a).

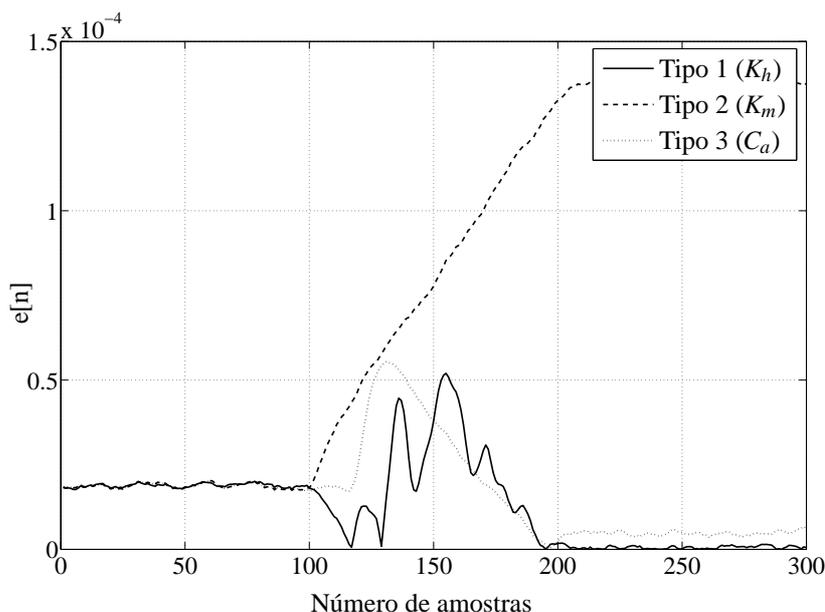


Figura 41: Erro de classificação para a falha do tipo 1 (K_h).

Foi repassado ao FA, durante a etapa de teste, um conjunto de dados contendo dados normais (100 amostras), de degradação (100 amostras) e de falha (100 amostras), nesta ordem, para cada tipo de falha: 1 (K_h), 2 (K_m) e 3 (C_a).

Pode se observar na Fig. 41 que o valor do erro, para a falha do tipo 1, a partir da amostra de número 200, atinge um valor próximo do zero; ou seja, a falha foi classificada corretamente pelo filtro adaptativo. Para as falhas do tipo 2 e 3, o valor do erro ficou em torno de 1.4×10^{-4} e 0.05×10^{-4} , respectivamente. Um valor tão próximo de zero, obtido para a falha do tipo 3, poderia resultar em uma classificação errônea desta falha.

Para a falha do tipo 2 (K_m), o gráfico do valor absoluto do erro é visto na Fig. 42. Foi apresentado ao FA, também, durante a etapa de teste, um conjunto de dados contendo dados normais (100 amostras), de degradação (100 amostras) e de falha (100 amostras), nesta ordem, para os três tipos de falhas.

O valor do erro, para a falha do tipo 2, a partir da amostra de número 200, atinge um valor próximo do zero; isto é, o FA classificou corretamente essa falha, como visto na Fig. 42. Para as falhas do tipo 1 e 3, o valor do erro ficou na faixa de 1.33×10^{-4} e 1.38×10^{-4} , respectivamente.

Por fim, para a falha do tipo 3 (C_a), o gráfico do valor absoluto do erro é visto na Fig. 43. Foi repassado ao FA, também, durante a etapa de teste, neste caso, um conjunto de dados (normal, degradação e falha) de modo semelhante ao realizado para as falhas do tipo 1 e 2.

Nota-se na Fig. 43, que o valor do erro, para a falha do tipo 3, a partir da amostra de número 200, alcança um valor próximo do zero; ou seja, a falha foi classificada corretamente pelo filtro adaptativo. Para as falhas do tipo 1 e 2, o valor absoluto do erro ficou em torno de 0.05×10^{-4} e 1.42×10^{-4} , respectivamente. O valor baixo de erro obtido para a falha do tipo 1, poderia acarretar em uma classificação errônea para esse tipo de falha.

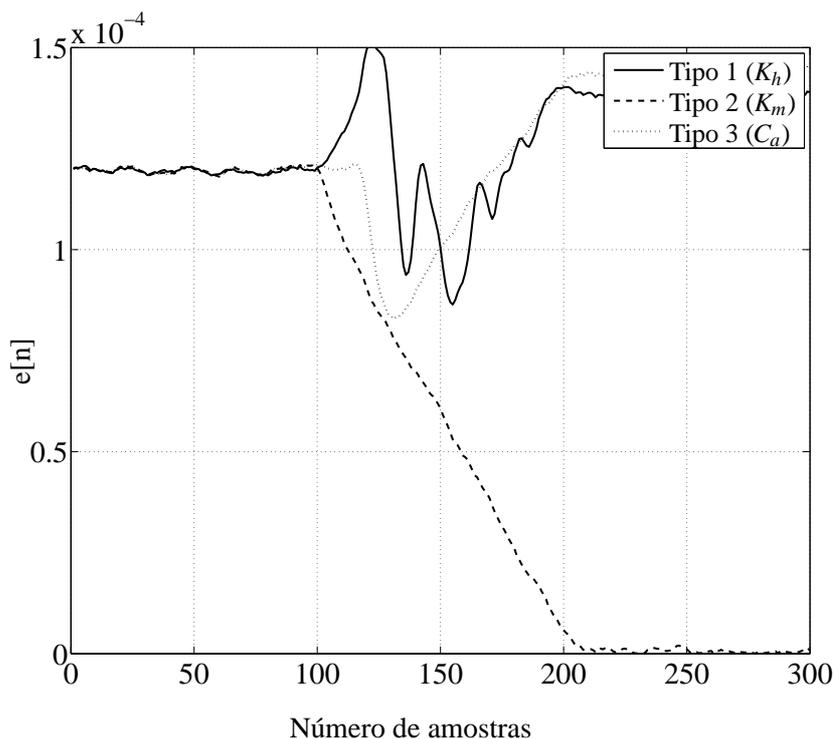


Figura 42: Erro de classificação para a falha do tipo 2 (K_m).

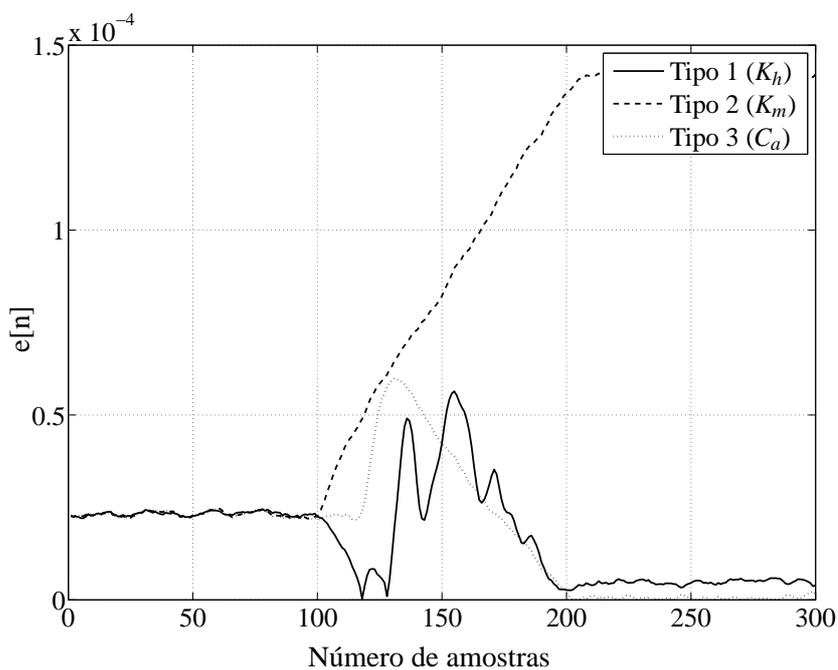


Figura 43: Erro de classificação para a falha do tipo 3 (C_a).

4.4.3.3 Resultados de predição

Para realizar a tarefa de predição não é necessário um treinamento prévio dos filtros adaptativos propriamente dito, para gerar os coeficientes do filtro, \mathbf{w} , como realizado anteriormente para as tarefas de detecção e classificação. Neste caso, somente é necessário efetuar a etapa de teste.

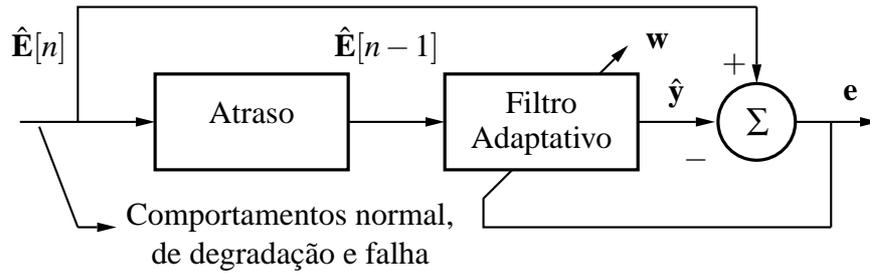


Figura 44: Etapa de teste dos filtros adaptativos para a tarefa de predição.

Na tarefa de predição, foram repassados como entrada do filtro adaptativo os dados de variância de energia espectral (neste caso, dados que representam o comportamento do tipo normal, de degradação e falha, para cada tipo de falha que se deseja prever), obtidos a partir do modelo, como visto na Fig. 44.

Os dados de entrada, neste caso os dados de variância de energia ($\hat{\mathbf{E}}[n]$), atrasados de uma amostra ($\hat{\mathbf{E}}[n-1]$), foram agora apresentados ao filtro adaptativo, como observado na Fig. 44. A saída do filtro adaptativo ($\hat{\mathbf{y}}[n]$) foi então comparada com a amostra de entrada ($\hat{\mathbf{E}}[n]$), gerando um sinal de erro ($\mathbf{e}[n]$) que é utilizado para atualizar os coeficientes.

Inicialmente, é então apresentado ao filtro o sinal de entrada ($\hat{\mathbf{E}}[n]$) que é atrasado de uma amostra. Após, é efetuado o cálculo do sinal de saída do filtro (principal variável de interesse neste caso), segundo a equação (183), aqui re-escrita:

$$\hat{\mathbf{y}}[n] = \mathbf{w}^T [n] \hat{\mathbf{E}}[n-1] \quad (68)$$

A seguir, o erro ($\mathbf{e}[n]$) é determinado a partir de:

$$\mathbf{e}[n] = \hat{\mathbf{y}}[n] - \hat{\mathbf{E}}[n] \quad (69)$$

O vetor de coeficientes é atualizado utilizando a seguinte expressão:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu \hat{\mathbf{E}}[n-1] \mathbf{e}[n] \quad (70)$$

A partir de certo número de amostras, o algoritmo converge para um ponto onde o erro atinge um valor próximo de zero e a amostra de entrada passa a ser aproximadamente igual a saída do filtro.

O gráfico do sinal de entrada e da saída do filtro adaptativo, usado na predição de falhas, para a falha do tipo 1 (K_h), é visto na Fig. 45. Foi repassado ao FA, durante a etapa de teste, um conjunto de dados contendo dados normais (100 amostras), de degradação (100 amostras) e de falha (100 amostras), para a falha do tipo 1 (K_h), nesta ordem.

Pode se observar na Fig. 45 que o valor da saída do filtro, $\hat{\mathbf{y}}$, a partir da amostra de número 30, que corresponde à fase inicial de ajuste dos coeficientes do filtro, passa a ser aproximadamente igual a entrada.

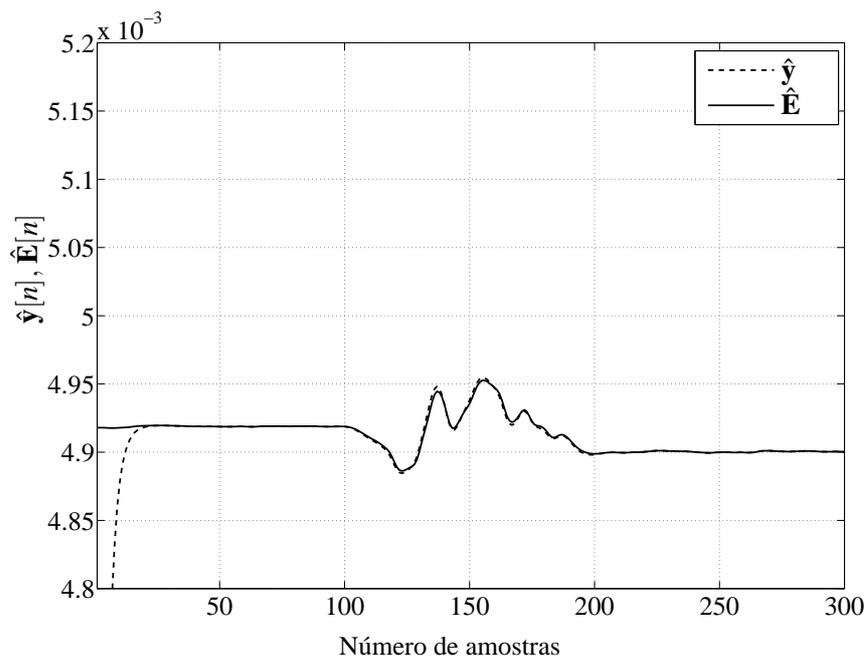


Figura 45: Gráfico de predição para a falha do tipo 1 (K_h).

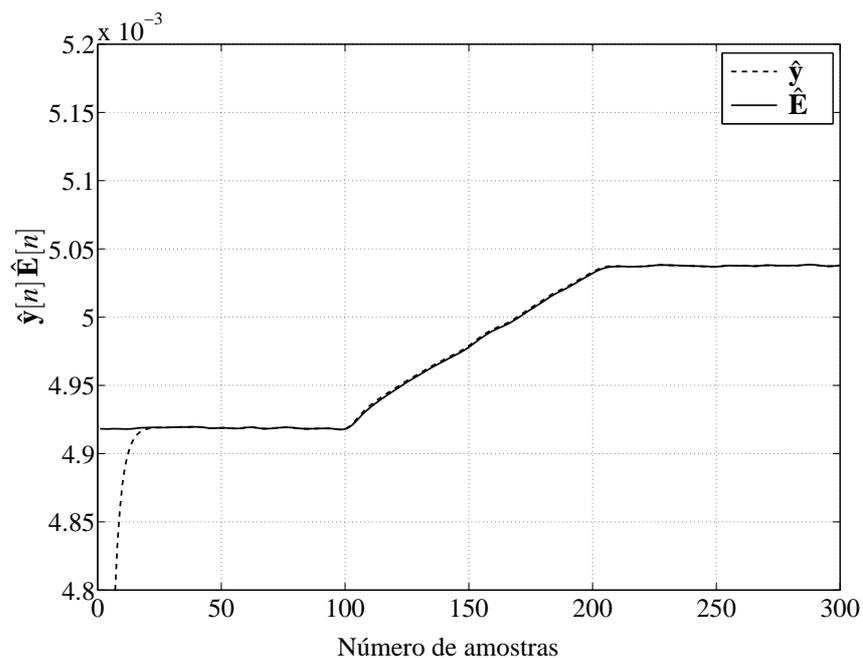


Figura 46: Gráfico de predição para a falha do tipo 2 (K_m).

Para a falha do tipo 2 (K_m), o gráfico do sinal de entrada e da saída do filtro adaptativo usado para a predição de falhas, é visto na Fig. 46. Foi repassado ao filtro adaptativo, durante esta etapa, um conjunto de dados (normal, degradação e falha) de modo semelhante ao realizado anteriormente para a falha do tipo 1, neste caso para a falha do tipo 2 (K_m).

Pode se observar na Fig. 46 que o valor da saída do filtro adaptativo, $\hat{\mathbf{y}}$, a partir da fase inicial de ajuste dos coeficientes do filtro adaptativo, aproximadamente a partir da amostra de número 30, passa a ser aproximadamente igual ao sinal de entrada.

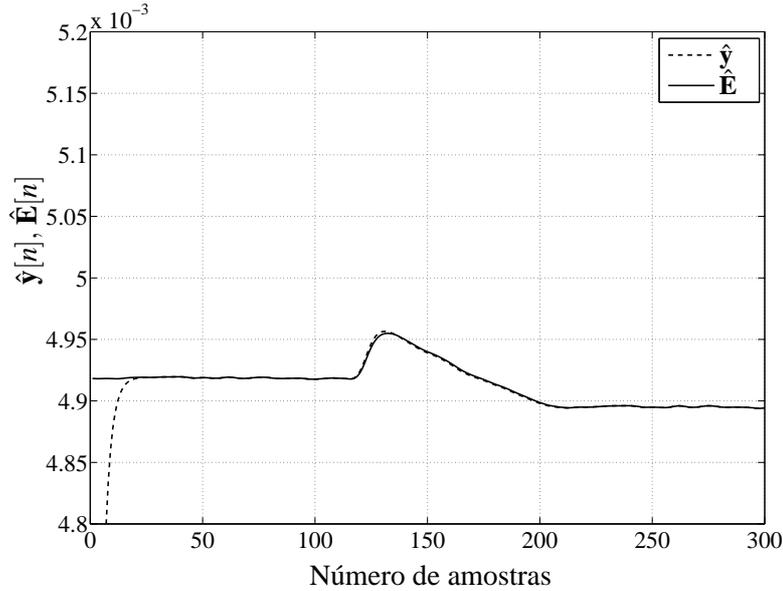


Figura 47: Gráfico de predição para a falha do tipo 3 (C_a).

Resultado semelhante foi obtido para a falha do tipo 3 (C_a). O gráfico da saída do filtro adaptativo usado para a predição do sinal de entrada, é visto na Fig. 47. Foi repassado ao filtro adaptativo, durante esta etapa, um conjunto de dados (normal, degradação e falha) do mesmo modo ao feito para as falhas do tipo 1 e 2.

Pode se observar na Fig. 47 que o valor da saída do filtro, \hat{y} , a partir a fase inicial de ajuste dos coeficientes do filtro, passa a ser aproximadamente igual a entrada, \hat{E} ; isto é, o filtro seguiu corretamente o sinal de referência.

4.5 DETERMINAÇÃO DO TEMPO DE VIDA ÚTIL DO ATUADOR

A próxima etapa realizada neste trabalho, após a etapa de predição (efetuada através dos MAO e FA), é determinar o tempo de vida útil do atuador usando os resultados de predição de falhas obtidos com os mapas auto-organizáveis e filtros adaptativos.

A metodologia utilizada para realizar a previsão do tempo de vida útil do atuador consiste basicamente em aplicar diferentes funções de interpolação, inicialmente, e extrapolação para prever os futuros valores de uma equação de transição de estados, vista a seguir:

$$R(t) = \sum_{n=1}^N f(n) + \dot{f}(n)t(n) + \frac{1}{2}\ddot{f}(n)t(n)^2 \quad (71)$$

onde: f é a variável de estado, \dot{f} e \ddot{f} as suas respectivas primeira e segunda derivadas, N é o número total de amostras, t é o período de tempo entre as amostras e R é definida como a variável característica de previsão do tempo de vida útil (SWANSON, 2002).

Os instantes de tempo, $t(n)$, dado em horas, foram definidos do seguinte modo:

$$t(n+1) = t(n) + 1 \quad (72)$$

onde: n é o número da amostra.

A variável de estado, f , no caso dos mapas auto-organizáveis, foi definida como sendo a variância dos neurônios vencedores ($\hat{\mathbf{W}}$), segundo a Eq (62), cujo gráfico, para os três tipos de falhas, pode ser visto na Fig. 35.

Já no caso dos filtros adaptativos, a variável de estado, f , é a própria saída do filtro adaptativo, $\hat{\mathbf{y}}$, Eq (68), cujos gráficos, para cada tipo de falha, são vistos na Fig. 45, 46 e 47.

A idéia do método aqui utilizado, consiste em analisar os valores das variáveis de estado f ($\hat{\mathbf{W}}$ no caso dos MAO e $\hat{\mathbf{y}}$ para os FA, aqui redefinidos como \mathbf{W} e \mathbf{y} para fins de simplificação) e das suas primeira ($\dot{\mathbf{W}}$ e $\dot{\mathbf{y}}$) e segunda derivadas ($\ddot{\mathbf{W}}$ e $\ddot{\mathbf{y}}$) para determinar o instante de tempo (t) onde os valores das derivadas se alteram de modo significativo, isto é:

$$\dot{f}(t) \gg 0 \text{ e } \ddot{f}(t) \gg 0 \quad (73)$$

O tempo de vida útil do atuador, T_u , considerando o instante de tempo t , é definido como:

$$T_u = \hat{t} - t \quad (74)$$

onde: \hat{t} corresponde ao instante de tempo onde $R(t)$ atinge um valor limite, referente a um tipo de falha específico, como visto a seguir:

$$R(\hat{t}) \geq R_l \quad (75)$$

onde: R_l é o valor limite para $R(t)$ para uma determinada falha.

Assim, são utilizadas as primeiras amostras de R , tomadas até o instante de tempo t , para efetuar a extrapolação de R até atingir o valor R_l e com isso obter \hat{t} e T_u . Diferentes funções podem ser usadas para a interpolação e extrapolação de R , tais como: funções racionais, polinomiais e exponenciais, por exemplo. A função que apresentar o menor T_u é escolhida.

Dentre estas funções, em especial, destaca-se aquelas utilizadas neste trabalho:

- **R1:** função racional que possui um polinômio de 4ª ordem como numerador e um polinômio de 2ª ordem como denominador:

$$R_1(x) = \frac{p_1x^4 + p_2x^3 + p_3x^2 + p_4x^1 + p_5}{x^2 + q_1x + q_2} \quad (76)$$

- **R2:** função racional que apresenta um polinômio de 5ª ordem como numerador e um polinômio de 2ª ordem como denominador:

$$R_2(x) = \frac{p_1x^5 + p_2x^4 + p_3x^3 + p_4x^2 + p_5x^1 + p_6}{x^2 + q_1x + q_2} \quad (77)$$

- **R3:** função racional que possui um polinômio de 5ª ordem como numerador e um polinômio de 3ª ordem como denominador:

$$R_3(x) = \frac{p_1x^5 + p_2x^4 + p_3x^3 + p_4x^2 + p_5x^1 + p_6}{x^3 + q_1x^2 + q_2x + q_3} \quad (78)$$

onde: p e q são os coeficientes do numerador e denominador, respectivamente e x é a variável de interesse, que neste caso são os instantes de tempo.

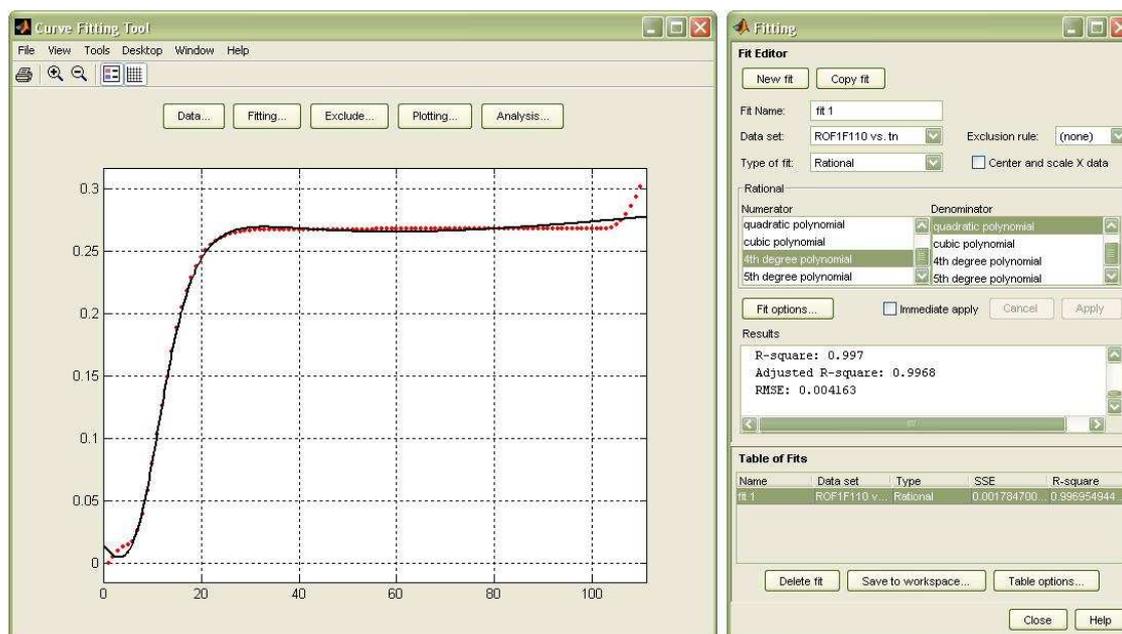


Figura 48: Principais janelas da ferramenta *Curve Fitting Tool*

A ferramenta utilizada para realizar a interpolação e extrapolação foi a *Curve Fitting Tool*, desenvolvida para o ambiente Matlab, cujas janelas principais podem ser vistas na Fig. 48. Nessa ferramenta pode-se importar ou exportar um conjunto de dados para a análise, efetuar e avaliar interpolações e extrapolações com diferentes funções e traçar os gráficos de resultados das análises dos conjuntos de dados, por exemplo.

4.5.1 Tempo de Vida Útil Obtido Através dos Mapas Temporais de Kohonen

Inicialmente, para se determinar o tempo de vida útil do atuador, através dos MAO, definiu-se, como variável de estado, a variância dos neurônios vencedores (\mathbf{W}), Eq (62). O gráfico da variância dos neurônios vencedores, para os três tipos de falhas, pode ser observado na Fig. 35.

O próximo passo, foi obter o valor da primeira ($\dot{\mathbf{W}}$) e da segunda derivada ($\ddot{\mathbf{W}}$) da variância dos neurônios vencedores (\mathbf{W}), isto é:

$$\dot{\mathbf{W}}(t) = \frac{d\mathbf{W}(t)}{dt} \quad (79a)$$

$$\ddot{\mathbf{W}}(t) = \frac{d\dot{\mathbf{W}}(t)}{dt} \quad (79b)$$

De posse dos valores e gráficos de $\dot{\mathbf{W}}(t)$ e $\ddot{\mathbf{W}}(t)$, o passo seguinte foi determinar o instante de tempo (i), para cada falha, onde os valores das derivadas se alteram significativamente.

Após, foi obtido o valor de $R(t)$ para o intervalo de tempo entre 0 e i , para as três falhas, com uma amostra sendo tomada a cada hora, como definido na Eq. 72, para a interpolação.

Por fim, utilizando as primeiras amostras de R , tomadas até o instante de tempo i , realizou-se a extrapolação de R até o final do intervalo de tempo analisado; ou seja, 300h.

Isto foi feito usando as três diferentes funções de interpolação (inicialmente) e extrapolação: R_1 , R_2 e R_3 . Com base nas extrapolações, três cenários distintos para o tempo de vida útil serão obtidos. A curva que apresentar o menor tempo de vida útil para o atuador será a escolhida.

4.5.1.1 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 1

Os gráficos dos valores absolutos da primeira (\dot{W}) e segunda (\ddot{W}) derivadas da variância dos neurônios vencedores (W), para a falha do tipo 1 (K_h), obtidos a partir dos MAO, podem ser vistos na Fig. 49.

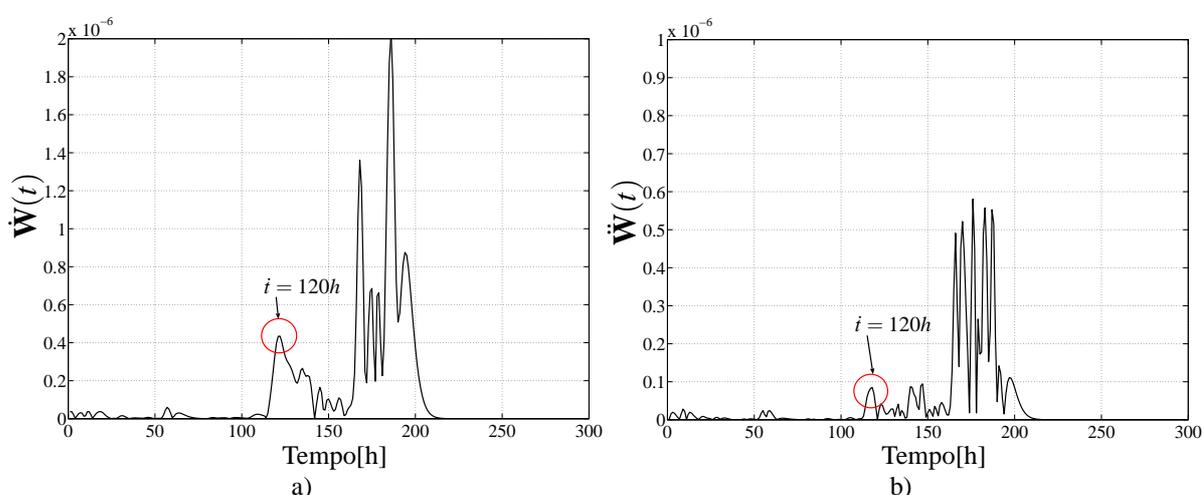


Figura 49: Gráficos da: a) primeira e b) segunda derivada de $W(t)$ para falha do tipo 1.

Pode-se observar na Fig. 49 que os valores das derivadas têm o seu valor alterado de modo significativo, $\dot{W}(t) \gg 0$ e $\ddot{W}(t) \gg 0$, a partir da amostra de número 120 (que corresponde a 120 horas); isto é, neste caso $i = 120h$. Assim, foram utilizadas as 120 primeiras amostras de R (aqui definido com R_0) para realizar a extrapolação desse sinal até a amostra de número 300, fazendo uso das três funções racionais para tal: R_1 , R_2 e R_3 .

O gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) usados para efetuar a extrapolação, e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 , para a falha do tipo 1, podem ser vistos na Fig. 50.

Pode-se notar na Fig. 50 que o valor limite (R_l) para R , neste caso, é aproximadamente igual a 0.83 ($R_l \cong 0.83$). Dentre as três curvas de extrapolação de R_0 , aquela que atinge esse limite primeiro, na amostra de número 162, é a curva R_2 ; isto é, $R_2(t = 162) \cong 0.83$. Assim, $\hat{t} = 162h$ e $T_u = 42h$. Deste modo, pode-se dizer que o atuador ainda possui, pelo menos, 42h de vida útil até a provável ocorrência da falha do tipo 1. Caso fosse utilizado a curva 1, pior cenário, para prever R , ter-se-ia $\hat{t} = 213h$ e $T_u = 93h$; isto é, 13 horas a mais do que o simulado ($\hat{t} = 200h$, $T_u = 80h$).

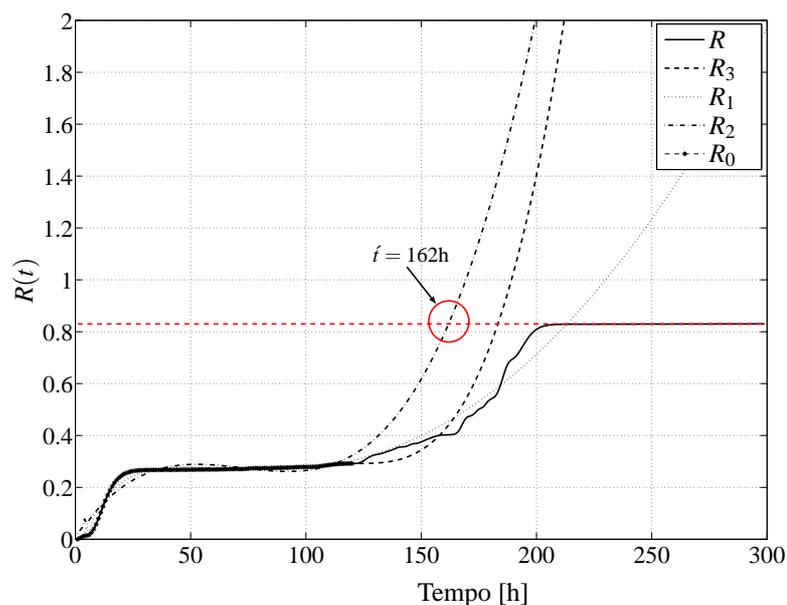


Figura 50: Gráfico de $R(t)$ para a falha do tipo 1 (K_h).

4.5.1.2 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 2

Considerando agora a falha do tipo 2 (K_m), os gráficos da primeira e segunda derivadas da variância dos neurônios vencedores (\mathbf{W}), podem ser vistos na Fig. 51.

Os valores das derivadas têm o seu valor alterado de modo significativo, $\dot{\mathbf{W}}(t) \gg 0$ e $\ddot{\mathbf{W}}(t) \gg 0$, a partir da amostra de número 110 (que corresponde a 110 horas); isto é, neste caso $i = 110$ h, como observado na Fig. 51. Deste modo, foram usadas as 110 primeiras amostras de R (neste caso, R_0) para efetuar a extrapolação desse sinal até a amostra de número 300.

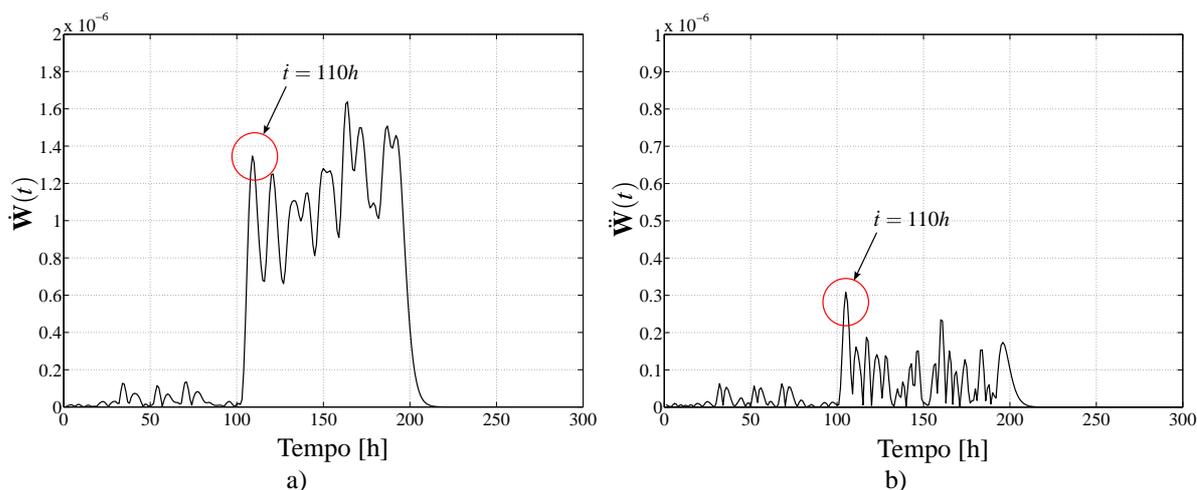


Figura 51: Gráficos da: a) primeira e b) segunda derivada de $\mathbf{W}(t)$ para falha do tipo 2.

Na Fig. 52 podem ser observados o gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) utilizados para realizar a extrapolação, e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 para a falha do tipo 2.

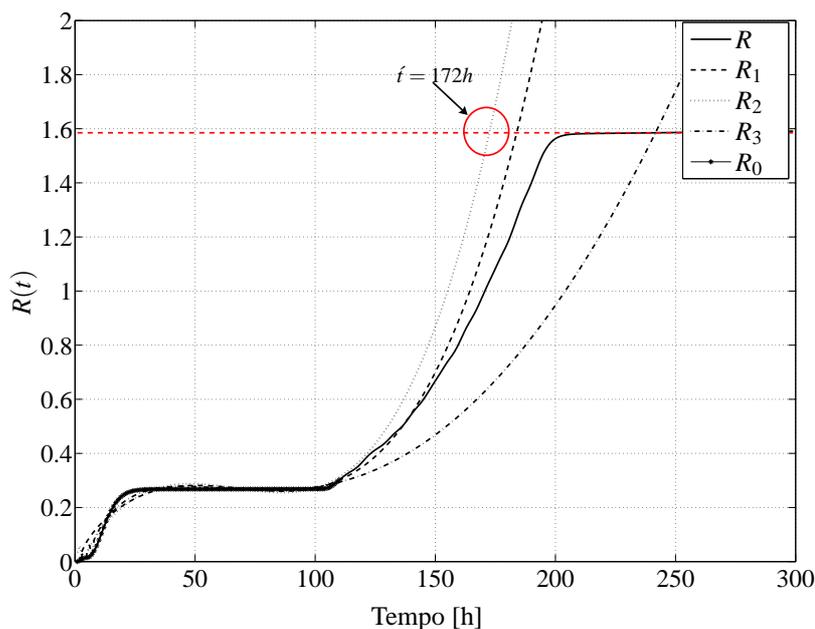


Figura 52: Gráfico de $R(t)$ para a falha do tipo 2 (K_m).

O valor limite (R_l) para R , neste caso, é aproximadamente igual a 1.585 ($R_l \cong 1.585$), como visto na Fig. 52. Dentre as três curvas de extrapolação de R_0 , aquela que alcança esse limite primeiro, na amostra de número 172, é a curva R_2 ; ou seja, $R_2(t = 172) \cong 1.585$. Assim, $\hat{t} = 172h$ e $T_u = 62h$. Deste modo, pode-se dizer que o atuador ainda possui, pelo menos, 62h de vida útil até a provável ocorrência da falha do tipo 2. No pior caso, neste caso para a curva R_3 , obtém-se $\hat{t} = 242$ e $T_u = 132h$; isto é, 42 horas a mais do que foi simulado ($T_u=90h$ e $\hat{t} = 200h$).

4.5.1.3 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 3

Na Fig. 53, pode-se observar os gráficos da primeira e segunda derivadas da variância dos neurônios vencedores (\mathbf{W}), para a falha do tipo 3 (C_a).

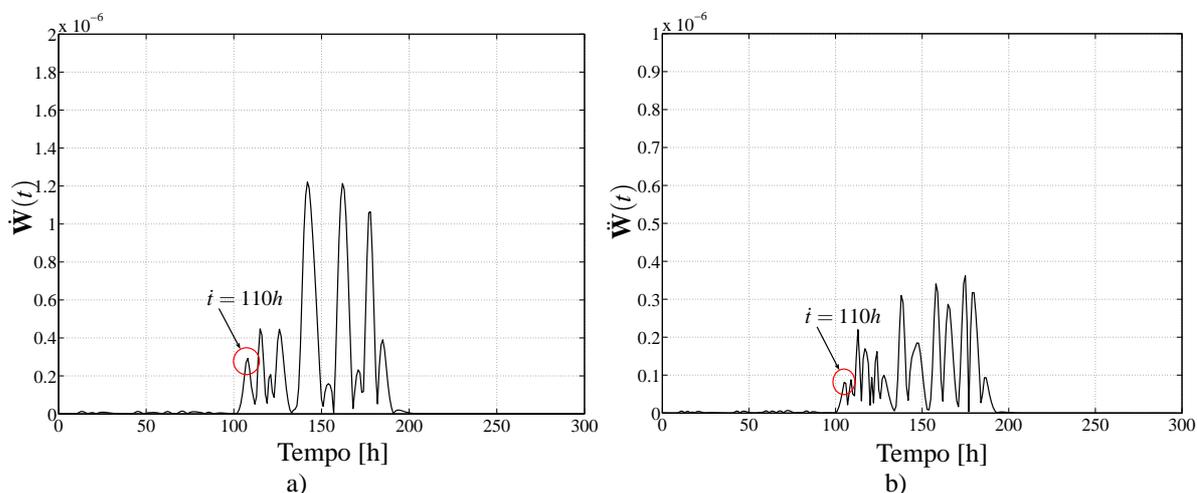


Figura 53: Gráficos da: a) primeira e b) segunda derivada de $\mathbf{W}(t)$ para falha do tipo 3.

Pode-se observar na Fig. 53 que os valores das derivadas têm o seu valor alterado de modo significativo, $\dot{W}(t) \gg 0$ e $\ddot{W}(t) \gg 0$, a partir da amostra de número 110 (isto é, 110 horas); isto é, neste caso $\hat{t} = 110\text{h}$. Assim, foram usadas as 110 primeiras amostras de R , para efetuar a extrapolação desse sinal até a amostra de número 300, fazendo uso de R_1 , R_2 e R_3 .

O gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) utilizados para realizar a extrapolação, e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 , para a falha do tipo 3, podem ser vistos na Fig. 54.

Pode-se notar na Fig. 54 que o valor limite (R_l) para R , neste caso, é aproximadamente igual a 0.61 ($R_l \cong 0.61$). Dentre as três curvas de extrapolação de R_0 , aquela que atinge esse limite primeiro, na amostra de número 152, é a curva R_2 ; isto é, $R_2(t = 152) \cong 0.61$. Assim, $\hat{t} = 152\text{h}$ e $T_u = 42\text{h}$. Deste modo, pode-se dizer que o atuador ainda possui, pelo menos, 42h de vida útil até a provável ocorrência da falha do tipo 3. Caso fosse utilizado a curva R_1 , pior cenário, para prever R , ter-se-ia $\hat{t} = 184\text{h}$ e $T_u = 74\text{h}$; isto é, 16 horas a menos do que o esperado ($\hat{t} = 200\text{h}$ e $T_u = 90\text{h}$).

4.5.2 Tempo de Vida Útil Obtido Através dos Filtros Adaptativos

Para se obter o tempo de vida útil do atuador, através dos FA, foi definido como variável de estado a saída do filtro adaptativo, y , Eq (68). Os gráficos da variância dos neurônios vencedores, para cada tipo de falha aqui analisado, podem ser vistos nas Fig. 45, 46 e 47.

Após isso, determinou-se o valor da primeira (\dot{y}) e da segunda derivada (\ddot{y}) da saída do filtro adaptativo, ou seja:

$$\dot{y}(t) = \frac{dy(t)}{dt} \quad (80a)$$

$$\ddot{y}(t) = \frac{d\dot{y}(t)}{dt} \quad (80b)$$

O próximo passo foi determinar o instante de tempo (\hat{t}), para cada falha, onde os valores das derivadas (\dot{y} e \ddot{y}) se alteram significativamente, semelhante ao feito para os MAO.

A seguir, obteve-se o valor de $R(t)$ para o intervalo de tempo entre 0 e \hat{t} , para os três tipos de falhas, com um amostra sendo tomada a cada hora, para se efetuar a interpolação.

Finalmente, fazendo uso das primeiras amostras de R , tomadas até o instante de tempo \hat{t} , efetuou-se a extrapolação de R até o final do intervalo de tempo analisado; ou seja, 300 horas, de modo semelhante ao realizado para os MAO. Foram usadas três diferentes funções para a interpolação e extrapolação de R_0 . A função que apresentar o menor tempo de vida útil para o atuador será a escolhida.

4.5.2.1 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 1

Considerando inicialmente a falha do tipo 1 (K_h), os gráficos da primeira (\dot{y}) e segunda (\ddot{y}) derivada da saída do filtro adaptativo, y , podem ser vistos na Fig. 55.

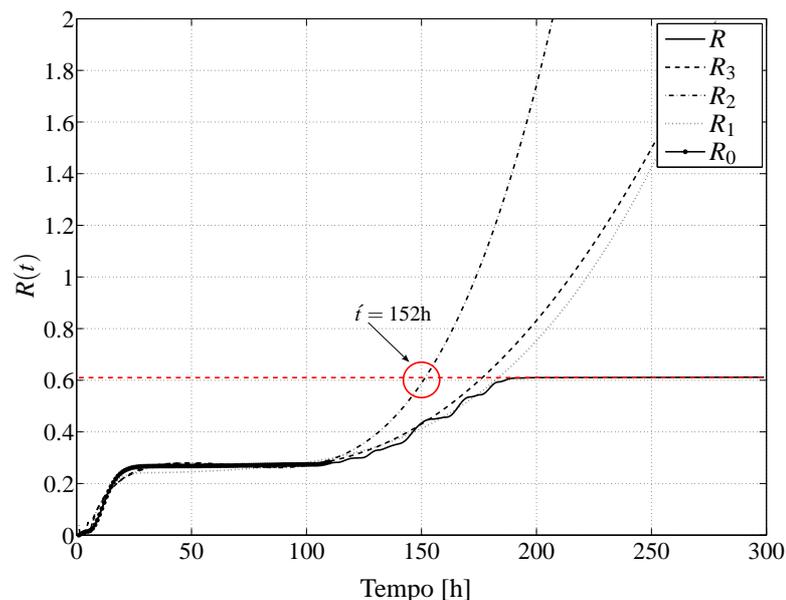


Figura 54: Gráfico de $R(t)$ para a falha do tipo 3 (C_a).

O valor da primeira derivada teve o seu valor alterado de modo significativo, $\dot{y}(t) \gg 0$, a partir da amostra de número 113 (que corresponde a 113 horas). Já para a segunda derivada, essa teve o seu valor modificado de modo considerável, $\ddot{y}(t) \gg 0$, a partir da amostra de número 108, como observado na Fig. 55. Neste caso, em virtude do número da amostra onde a primeira e segunda derivada tiveram seu valor alterado significativamente serem diferentes, foi escolhido um valor intermediário; isto é, $i = 110$ h. Deste modo, foram usadas as 110 primeiras amostras de R (neste caso, R_0) para realizar a extrapolação desse sinal até a amostra de número 300.

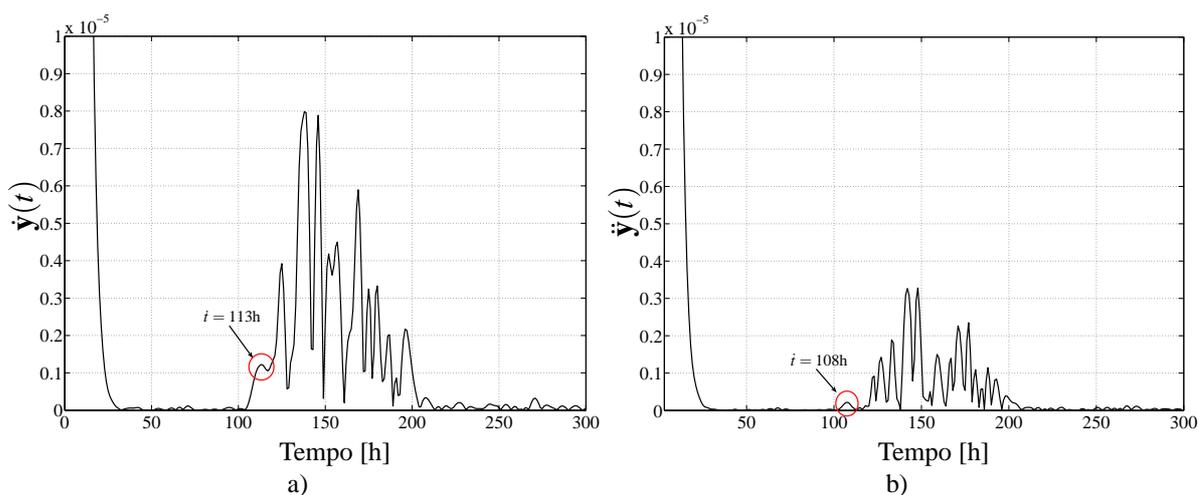


Figura 55: Gráficos da: a) primeira e b) segunda derivada de y para a falha do tipo 1.

Na Fig. 56 podem ser observados o gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) utilizados para efetuar a extrapolação de R , e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 para a falha do tipo 1.

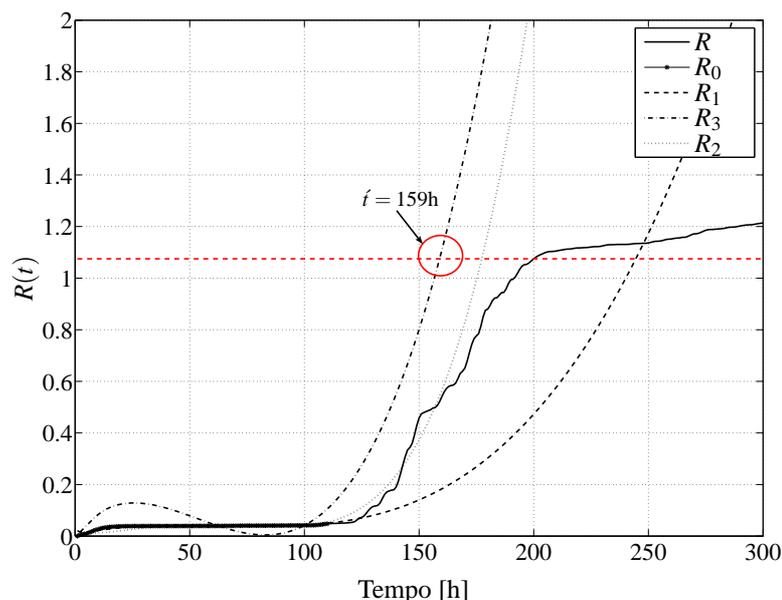


Figura 56: Gráfico de $R(t)$ para a falha do tipo 1 (K_h).

O valor limite (R_l) para R , neste caso, é aproximadamente igual a 1.075 ($R_l \cong 1.075$), como visto na Fig. 56. Dentre as três curvas de extrapolação de R_0 , aquela que alcança esse limite primeiro, na amostra de número 159, é a curva R_3 ; ou seja, $R_3(t = 159) \cong 1.075$. Assim, $\hat{t} = 159h$ e $T_u = 49h$. Deste modo, pode-se dizer que o atuador ainda possui, pelo menos, 49h de vida útil até a provável ocorrência da falha do tipo 1. No pior caso, neste caso para a curva R_1 , ter-se-ia $\hat{t} = 244$ e $T_u = 134h$; isto é, 44 horas a mais do que o simulado ($T_u=90h$ e $\hat{t} = 200h$).

4.5.2.2 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 2

Os gráficos dos valores absolutos da primeira (\dot{y}) e segunda (\ddot{y}) derivadas da variância dos neurônios vencedores (y), para a falha do tipo 2 (K_m), obtidos a partir dos FA, podem ser vistos na Fig. 57.

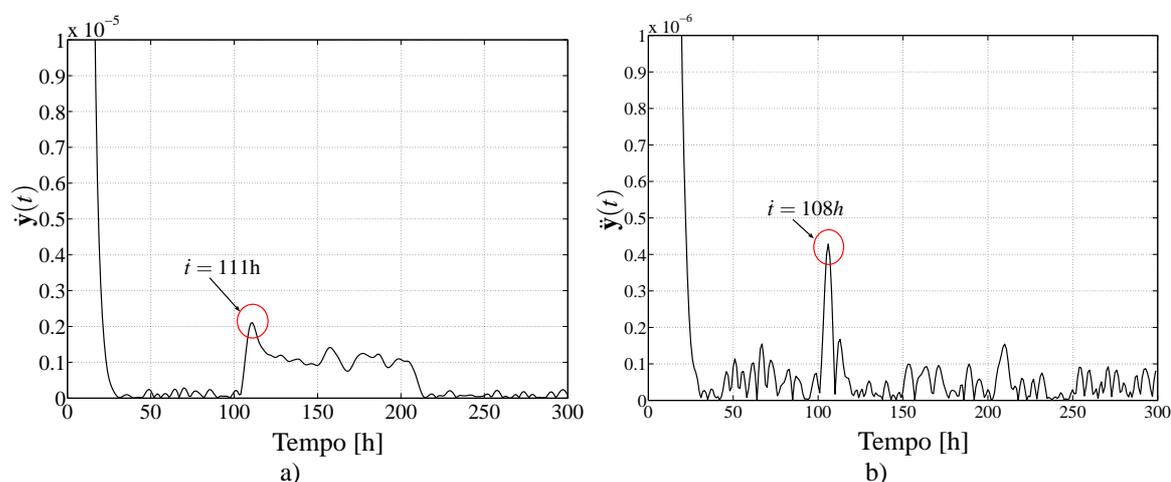


Figura 57: Gráficos da: a) primeira e b) segunda derivada de $y(t)$ para a falha do tipo 2.

Pode-se notar na Fig. 57 que o valor da primeira derivada teve o seu valor alterado de modo significativo, $\dot{y}(t) \gg 0$, a partir da amostra de número 111 (que corresponde a 111 horas). A segunda derivada teve o seu valor modificado consideravelmente, $\ddot{y}(t) \gg 0$, a partir da amostra de número 108.

Assim, em função do número da amostra onde a primeira e segunda derivada tiveram seu valor alterado significativamente serem distintos, foi escolhido um valor intermediário; isto é, $\hat{t} = 110\text{h}$. Deste modo, foram usadas as 110 primeiras amostras de R (neste caso, R_0) para realizar a extrapolação desse sinal, como visto na Fig. 57.

O gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) usados para realizar a extrapolação, e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 , para a falha do tipo 2, podem ser vistos na Fig. 58.

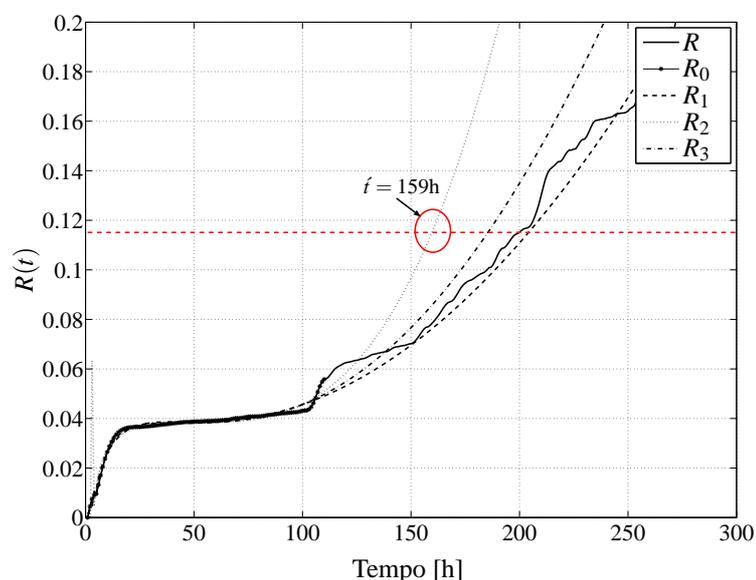


Figura 58: Gráfico de $R(t)$ para a falha do tipo 2 (K_m).

Pode-se observar na Fig. 58 que o valor limite (R_l) para R , neste caso, é aproximadamente igual a 0.115 ($R_l \cong 0.115$). Dentre as três curvas de extrapolação de R_0 , aquela que atinge esse limite primeiro, na amostra de número 159, é a curva R_2 ; isto é, $R_2(t = 159) \cong 0.115$. Assim, $\hat{t} = 159\text{h}$ e $T_u = 49\text{h}$. Deste modo, pode-se dizer que o atuador ainda possui, pelo menos, 49h de vida útil até a provável ocorrência da falha do tipo 2. Caso fosse utilizado a curva R_1 , pior cenário, para prever R , ter-se-ia $\hat{t} = 203\text{h}$ e $T_u = 93\text{h}$; isto é, 3 horas a mais do que o simulado ($\hat{t} = 200\text{h}$, $T_u = 90\text{h}$).

4.5.2.3 Predição do Tempo de Vida útil do Atuador para a Falha do Tipo 3

Considerando por fim a falha do tipo 3 (C_a), os gráficos dos valores absolutos da primeira (\dot{y}) e segunda (\ddot{y}) derivadas da variância dos neurônios vencedores (y), obtidos a partir dos FA, podem ser vistos na Fig. 59.

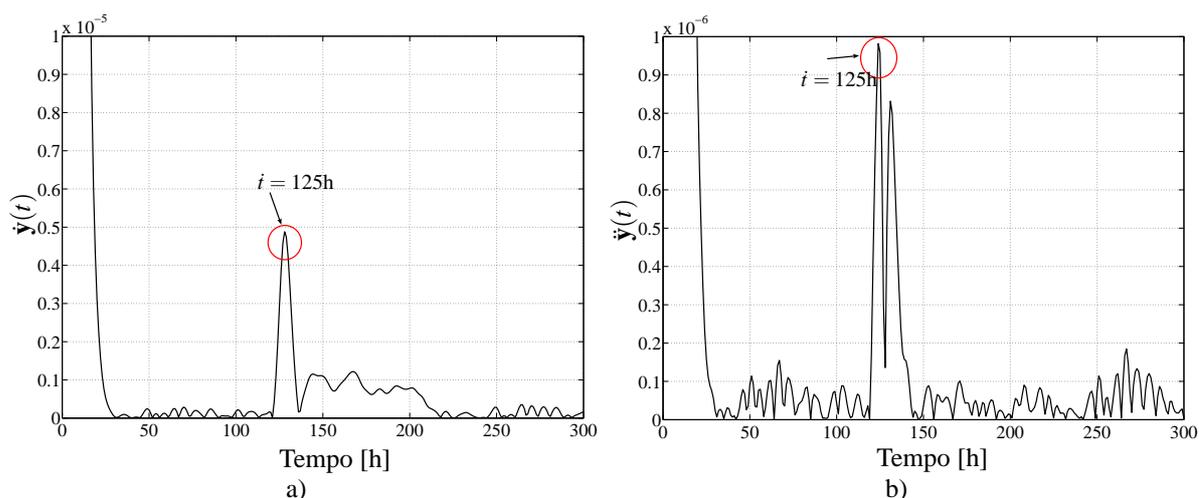


Figura 59: Gráficos da: a) primeira e b) segunda derivada de $y(t)$ para a falha do tipo 3.

Os valores da primeira e segunda derivadas tiveram o seu valor alterado de modo significativo, $\dot{y}(t)$ e $\ddot{y}(t) \gg 0$, a partir da amostra de número 125 (ou 125 horas), como visto na Fig. 59; isto é, neste caso $i = 125h$. Deste modo, foram usadas as 125 primeiras amostras de R , para efetuar a extrapolação desse sinal até a amostra de número 300, fazendo uso das funções racionais R_1 , R_2 e R_3 .

Na Fig. 60 podem ser observados o gráfico da variável característica de previsão do tempo de vida útil do atuador (R), o conjunto inicial de valores (R_0) utilizados para realizar a extrapolação de R , e as três curvas de extrapolação para três funções racionais (R_1 , R_2 e R_3) obtidas a partir de R_0 para a falha do tipo 3, bem como o valor limite (R_l) para R , que neste caso é aproximadamente igual a 0.175 ($R_l \cong 0.175$).

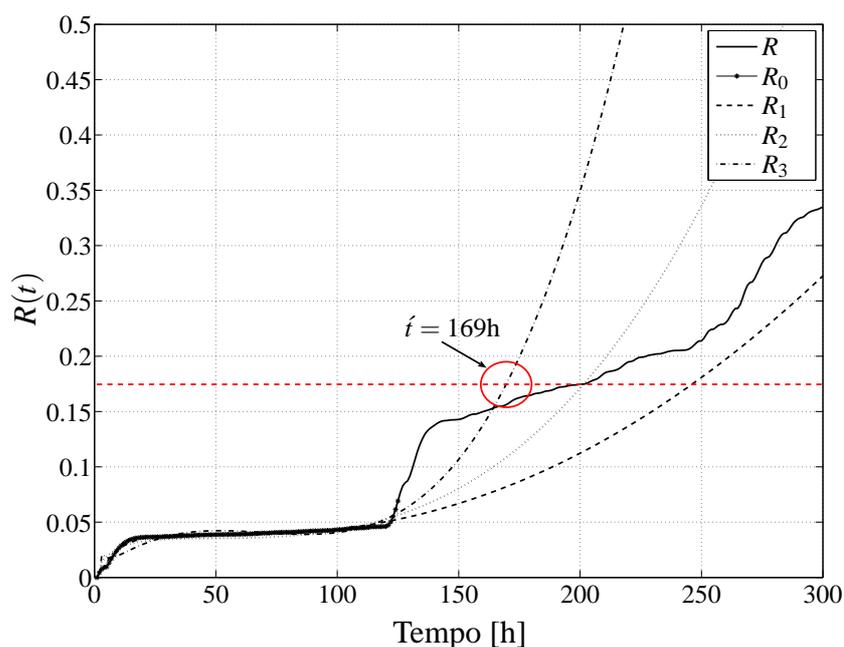


Figura 60: Gráfico de $R(t)$ para a falha do tipo 3 (C_a).

Dentre as três curvas de extrapolação de R_0 , aquela que alcança esse limite primeiro, na amostra de número 169, é a curva R_3 ; ou seja, $R_3(t = 169, \mathbf{y}) \cong 0.175$. Assim, $\hat{t} = 169\text{h}$ e $T_u = 44\text{h}$. Deste modo, pode-se dizer que o atuador ainda possui, ao menos, 44h de vida útil até a provável ocorrência da falha do tipo 3. No pior caso, neste caso para a curva R_1 , ter-se-ia $\hat{t} = 246$ e $T_u = 121\text{h}$; isto é, 46 horas a mais do que o simulado ($T_u=75\text{h}$, $\hat{t} = 200\text{h}$).

4.5.3 Comparação dos Resultados de Predição

A Tab. 4 exibe um quadro comparativo dos valores (em horas) e dos erros (percentuais) para o tempo de vida útil obtido através dos MAO e FA.

Tabela 4: Quadro comparativo do T_u obtido com os MAO e FA.

Tipo de falha	Função racional	MAO		FA	
		T_u	Erro[%]	T_u	Erro [%]
1 (K_h)	R_1	93h	+16.25	134h	+48.88
	R_2	42h	-47.50	66h	-26.67
	R_3	63h	-21.25	59h	-34.44
2 (K_m)	R_1	74h	-17.77	93h	+3.330
	R_2	62h	-31.11	49h	-45.55
	R_3	132h	+46.67	76h	-15.55
3 (C_a)	R_1	74h	-17.71	121h	+61.33
	R_2	42h	-53.33	76h	+1.330
	R_3	67h	-25.55	44h	-41.33

Pode-se observar nessa tabela que, para os MAO, a função racional que apresentou o menor erro percentual, para os três tipos de falhas, foi a função R_1 . Já para os FA, a função racional que exibiu o menor erro percentual, para os três tipos de falhas, foi a função R_3 .

Contudo, em função do critério adotado neste trabalho (a função que apresentar o menor tempo de vida útil para o atuador é a escolhida) a função R_2 , para os MAO é a melhor escolha. Para os FA, a função R_3 foi a que apresentou os melhores resultados.

4.6 CONCLUSÕES DO CAPÍTULO

Este capítulo, inicialmente, apresentou uma análise das principais ferramentas de processamento de sinais (transformada de Fourier, transformada de Fourier com janelamento e transformada *wavelet packet*) através do método da regressão logística para determinar qual destas ferramentas é a mais adequada para o estudo de caso aqui analisado, um atuador.

A partir dos gráficos do valor de confiança, obtidos com o método da regressão logística, observou-se que a transformada *wavelet packet* apresentou os melhores resultados, identificando corretamente o comportamento normal, de degradação e falha, para três tipos de falhas. Assim, constatou-se que a ferramenta de processamento de sinais mais recomendada para este estudo de caso é a transformada *wavelet packet*.

A seguir, os resultados de energia espectral, dos sinais de torque e posição do modelo, obtidos através da transformada *wavelet packet*, foram vistos.

Uma descrição das etapas de inicialização, treinamento e teste dos mapas auto-organizáveis e filtros adaptativos, bem como os resultados alcançados na realização das tarefas de detecção, classificação e predição de falhas, com estas ferramentas, foram vistos a seguir.

Através dos mapas auto-organizáveis e dos filtros adaptativos foi possível realizar uma correta detecção, classificação e predição de falhas. Para essas tarefas, pode-se dizer que ambas as ferramentas, mapas auto-organizáveis e filtros adaptativos, mostraram resultados satisfatórios e aproximadamente iguais em termos qualitativos. Com relação a tarefa de classificação de falhas, pode-se notar que os mapas auto-organizáveis apresentaram resultados ligeiramente superiores aos obtidos com os filtros adaptativos. Isso se deve às características específicas dos mapas auto-organizáveis, tais como a aprendizagem e generalização, que permitem obter melhores resultados para a classificação de falhas.

Por fim, a determinação do tempo de vida útil, por meio dos mapas auto-organizáveis e filtros adaptativos, através da extrapolação da curva característica de previsão do tempo de vida útil do atuador, foi abordada.

Considerando o critério adotado neste trabalho (a função racional que apresentar o menor tempo de vida útil para o atuador é a função escolhida) a função racional R_2 é a melhor escolha para os mapas auto-organizáveis. Para os filtros adaptativos, a função racional R_3 foi a função que apresentou os melhores resultados de predição.

5 DESENVOLVIMENTO DO SISTEMA EMBARCADO

Uma das principais etapas deste trabalho é o desenvolvimento, embarque e prototipação de um sistema de detecção, classificação e predição de falhas do atuador, voltado para a manutenção proativa. Para tal, é necessário determinar a ferramenta (MAO ou FA) mais adequada para o embarque. Assim, optou-se por embarcar, em uma placa de circuito impresso, o sistema em um *chip* para a realização e avaliação destas tarefas em tempo real.

Atualmente, a grande maioria dos *chips* encontrados nos circuitos que acompanham as televisões, celulares, e máquinas fotográficas por exemplo, já vêm todos pré-programados; isto é, com todas as suas funcionalidades definidas no ato de sua fabricação. Esses *chips* são conhecidos como *Application Specific Integrated Circuit* (ASIC).

O ASIC é um processador construído para executar uma tarefa específica, ou seja, um processador dedicado. Outros exemplos de processadores dedicados são os coprocessadores aritméticos e os *chipsets* de vídeo (CARRO, 2001).

Recentemente, surgiu uma categoria nova de *hardwares* reconfiguráveis, os quais têm as suas funcionalidades programadas exclusivamente pelos usuários e não pelos fabricantes. Dentre uma grande gama de *hardwares* reconfiguráveis pode-se destacar o *Field Programmable Gate Array* (FPGA).

Um FPGA é um dispositivo semicondutor largamente utilizado para o processamento de informações digitais. Foi apresentado pela Xilinx em 1985 como um dispositivo que poderia ser programado de acordo com as aplicações do usuário (ou programador).

Atualmente, esses dispositivos estão presentes em diversas áreas (espacial, médica e entretenimento, por exemplo) voltados para os mais distintos tipos de aplicações, tais como: automotivas, industriais ou espaciais.

O FPGA suporta a implementação de circuitos lógicos programáveis. Consiste, basicamente, de um grande arranjo de células lógicas ou blocos lógicos configuráveis contidos em um único circuito integrado. Este dispositivo faz parte de uma classe de dispositivos digitais que são definidos como plataformas físicas configuráveis por *software*; ou seja, utiliza-se um programa para descrever a configuração que implementa um determinado circuito.

Para programar um FPGA (ou um ASIC) são usados basicamente dois tipos de linguagens: VHDL e Verilog. Neste trabalho, estar-se-á particularmente interessado na linguagem VHDL. Basicamente, para se projetar circuitos digitais em FPGAs (ou ASICs) utilizando a linguagem VHDL é necessário definir, inicialmente, a máquina de estados, os processos e por fim a arquitetura do circuito que será programado.

A máquina de estados corresponde à modelagem do comportamento do circuito. É composta por estados, transições e ações. O processo é um modo formal e estruturado de se fazer uma lista seqüencial de operações que corresponde ao comportamento, ou funcionalidade, esperada do sistema para se obter um determinado resultado. Já a arquitetura corresponde as interfaces, entidades (ou blocos) e sinais (internos e externos) do circuito em questão.

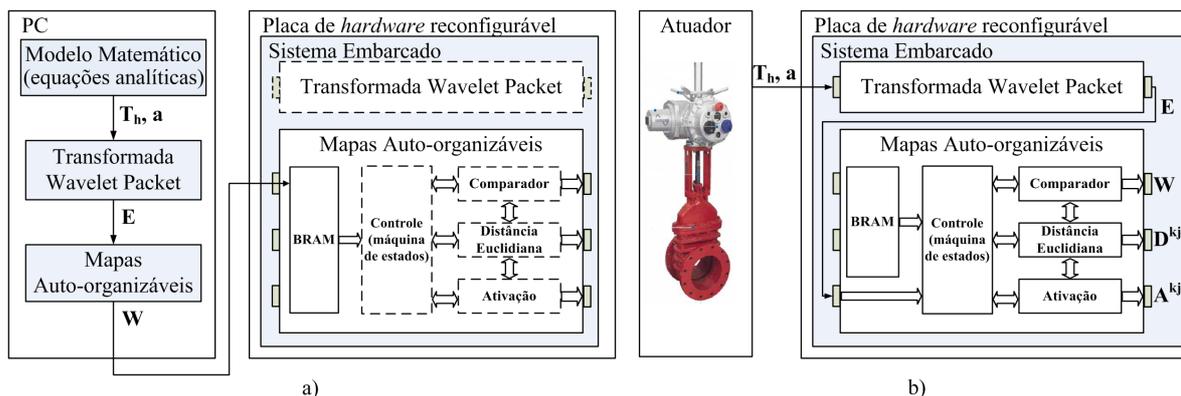


Figura 61: Arquitetura do circuito dos MAO: a) treinamento e b) teste.

Assim, neste capítulo, na seção 5.1, o embarque do SMI (circuitos de detecção, classificação e previsão de falhas, desenvolvidos a partir dos MAO e FA), incluindo uma descrição dos processos, das arquiteturas e das máquinas de estados, é apresentado. A seguir, na seção 5.2, uma descrição da plataforma de embarque, incluindo a especificação dos MAO e FA para embarque, é realizada. Os resultados de embarque, em termos de área, tempo e desempenho, são exibidos na seção 5.3. Por fim, as conclusões do capítulo são abordadas na seção 5.4.

5.1 EMBARQUE DO SISTEMA DE MANUTENÇÃO INTELIGENTE

A etapa de embarque do SMI, semelhante aos resultados de previsão do tempo de vida útil do atuador, se divide em duas partes: o embarque dos MAO e o embarque dos FA.

As etapas de inicialização e treinamento dos MAO foram realizadas em *software*. A etapa de teste, incluindo a determinação dos neurônios vencedores (W^{j*}), Eq. (56), da distância Euclidiana (D^{kj}), Eq. (55), para os MAO e da trajetória dos neurônios vencedores (W^{j*}), Eq. (58), e da ativação temporal (A^{kj}), Eq. (59), para os MTK, foi programada em VHDL e realizada em *hardware*, como visto na Fig. 61. A arquitetura simplificada do circuito para cálculo da distância Euclidiana, do neurônio vencedor e da ativação temporal, como visto na Fig. 61, é composta de um bloco comparador, um bloco para a determinação da distância Euclidiana, uma BRAM para armazenar um mapa já treinado, um bloco para cálculo da ativação e uma máquina de estados para controlar o processo, além de registradores e multiplexadores.

De modo semelhante, as etapas de inicialização e treinamento dos FA foram realizadas em *software*. A etapa de teste, que compreende o cálculo do erro de detecção e de classificação (e), Eq. (67), da saída da rede neural (y), Eq. (68), do erro de previsão (e), Eq. (69), e da atualização dos coeficientes do filtro adaptativo (w), Eq. (70), também foram programadas em VHDL, como observado na Fig. 62. Na Fig. 62 pode-se notar que a arquitetura simplificada do circuito para cálculo da saída do FA, do erro e da atualização dos coeficientes é composta de um bloco subtrator, um bloco para o cálculo da saída do filtro, uma BRAM para armazenar os coeficientes do filtro (obtidos no treinamento) e uma máquina de estados, além de registradores e multiplexadores (não exibidos na figura).

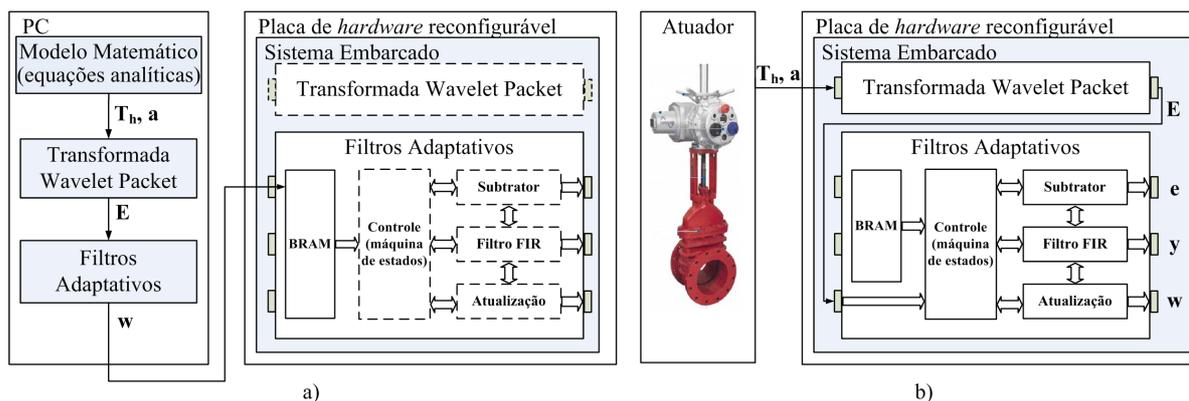


Figura 62: Arquitetura do circuito dos FA: a) treinamento e b) teste.

Ressalta-se que o cálculo das grandezas, quer seja para os filtros adaptativos ou mapas auto-organizáveis, é útil para a realização das tarefas de detecção, classificação e predição de falhas, como descrito nas seções 4.3 e 4.4. Assim, dentre as principais etapas necessárias para a implantação deste sistema embarcado voltado para a detecção, classificação e predição de falhas, em VHDL, cita-se:

1. Especificação das arquiteturas dos circuitos;
2. Elaboração das máquinas de estados;
3. Programação dos algoritmos (através de processos) em VHDL;
4. Realização de simulações;
5. Implementação dos circuitos em *hardware*.

Cita-se ainda que a etapa de implementação dos circuitos em *hardware* inclui a síntese (onde ocorre a conversão do algoritmo descrito em VHDL para estruturas de dados representando as conexões, blocos e portas lógicas) e a implementação (quando é definido em qual tipo de tecnologia o circuito será implementado; inclui as etapas de tradução, mapeamento e roteamento), além de testes e simulações. Destaca-se também: todas as máquinas de estado são do tipo Mealy.

5.1.1 Embarque dos Mapas Auto-organizáveis

Os circuitos (arquiteturas, máquinas de estados e processos) programados em VHDL voltados para o embarque dos MAO (para a realização das tarefas de detecção, classificação e predição de falhas dos atuadores) serão simplesmente definidos como circuito de detecção, classificação e predição, respectivamente. Vale ressaltar que estes blocos não executam as tarefas de detecção, classificação e predição sozinhos, sempre há a necessidade de uso do bloco da BRAM e da máquina de estados, além de registradores, na execução destas tarefas.

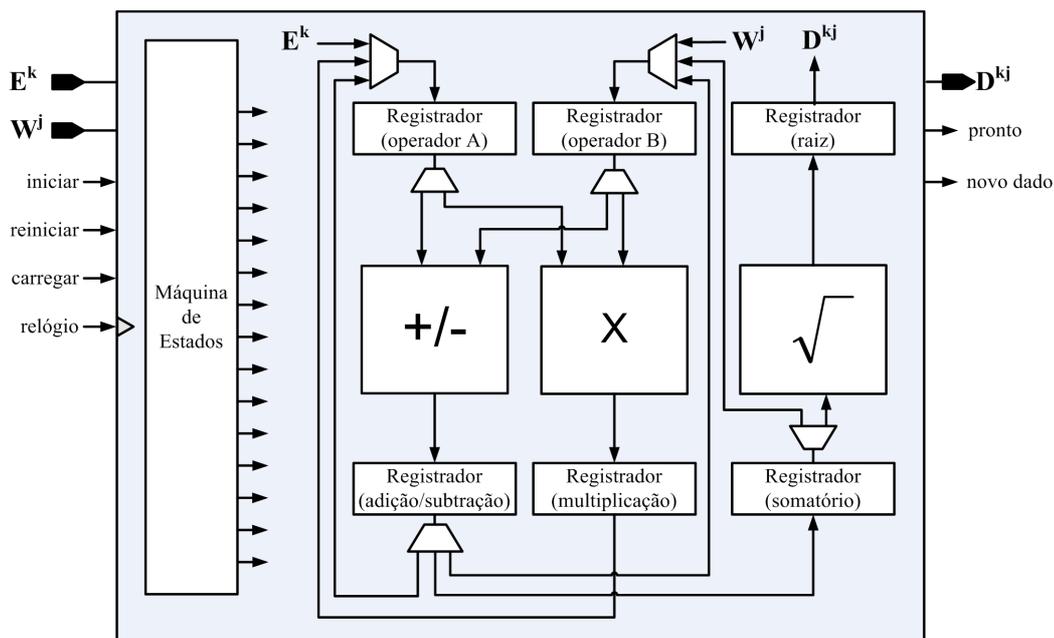


Figura 63: Arquitetura do circuito de detecção dos mapas auto-organizáveis.

5.1.1.1 Circuito de Detecção

A detecção de falhas, através dos MAO, é realizada com base no cálculo do erro de quantização, (E_Q), Eq. (60). Assim, partindo da Eq. (60) constatou-se que para implementar o cálculo de E_Q seria necessário usar um bloco somador, um subtrator, um multiplicador, e um extrator de raiz quadrada; além, é claro, de alguns registradores e multiplexadores.

a) Arquitetura do Circuito de Detecção

Deste modo, programou-se, em VHDL, o circuito visto na Fig. 63. Através da máquina de estados os valores de E^k e W^j são repassados através dos blocos de subtração, adição, multiplicação, e radiciação, além dos registradores. Inicialmente, os valores de E^k e W^j são remetidos aos registradores operadores e então, a seguir, enviados ao bloco somador/subtrator para ser efetuada a subtração. Após, o resultado da subtração é endereçado, passando pelos registradores, ao bloco multiplicador para que seja feita a multiplicação.

O próximo passo é realizar o somatório (usando o bloco somador/subtrator). Por fim, finalizado o somatório, é extraída a raiz quadrada desse somatório, que corresponde a distância Euclidiana, D^{kj} , entre os vetores E^k e W^j .

Os principais blocos do circuito de detecção: somador/subtrator, multiplicador e raiz quadrada foram gerados a partir da biblioteca da ferramenta da Xilinx, bastando instanciá-los na rotina principal (XILINX, 2010).

b) Máquina de Estados do Circuito de Detecção

A máquina de estados, simplificada, do circuito de detecção, responsável pelo controle do circuito de detecção visto na Fig. 63, possui 15 estados, como visto na Fig. 64.

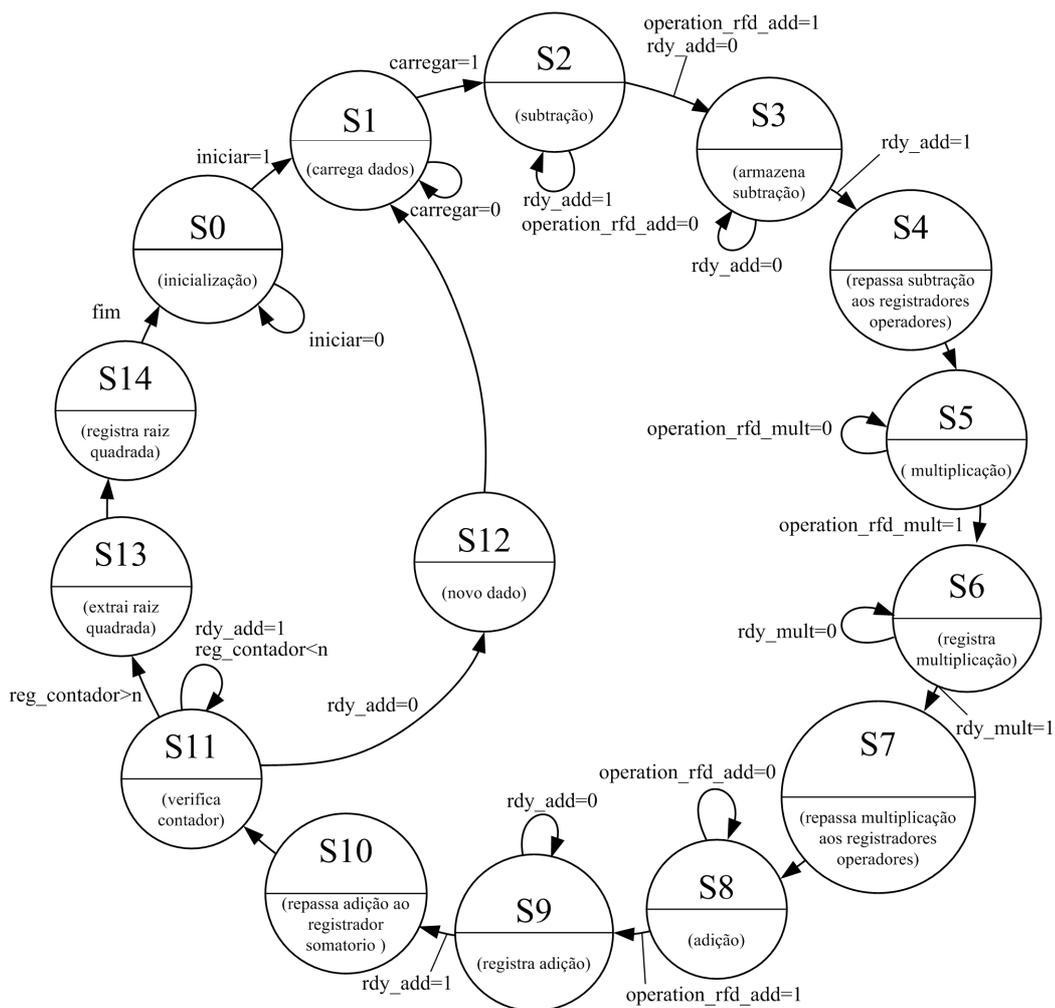


Figura 64: Máquina de estados do circuito de detecção dos mapas auto-organizáveis.

Os principais estados do circuito de detecção, vistos na Fig. 64, são resumidos a seguir:

- S1:** carrega um novo dado;
- S2:** a subtração é realizada;
- S5:** a multiplicação é efetuada;
- S9:** registra o resultado da adição;
- S11:** o somatório é finalizado ou um novo ciclo é iniciado;
- S13:** a raiz quadrada do somatório é extraída.

Os sinais rdy_add , rdy_mult , $operation_rfd_add$ e $operation_rfd_mult$, vistos na Fig. 64, são sinais de confirmação de chegada e saída de dados dos blocos de subtração/adição e multiplicação, respectivamente. Os sinais $iniciar$, $reinicar$, $carregar$, $pronto$, e $novo\ dado$, vistos na Fig. 63, são sinais externos responsáveis pela operação do circuito.

c) Processos do Circuito de Detecção

Os principais processos programados para implementar o circuito de detecção foram: os processos dos registradores (operadores, adição, subtração, multiplicação, somatório e raiz quadrada), processos para dirigir a parte operativa e a parte de controle da máquina de estados e o processo para controlar a lógica de transição entre os estados. A declaração do bloco subtrator/somador, com as suas entradas e saídas é apresentado, como exemplo, abaixo:

```
component float_add_sub
port (
  a,b: IN std_logic_vector(31 downto 0);
  operation: IN std_logic_vector(5 downto 0);
  operation_nd: IN std_logic;
  clk,sclr: IN std_logic;
  result: OUT std_logic_vector(31 downto 0);
  rdy: OUT std_logic);
end component;
```

onde: a, b, result, etc. são sinais de entrada (IN) e saída (OUT) do bloco; `std_logic` e `std_logic_vector` são tipos de dados.

5.1.1.2 Circuito de Classificação

A classificação de falhas, através dos mapas auto-organizáveis, é realizada com base no cálculo do neurônio vencedor, (\mathbf{W}^{j*}), Eq. (56). Deste modo, notou-se que para por em prática o circuito desejado seria necessário usar um bloco comparador, um bloco para a determinação da distância Euclidiana, uma BRAM para armazenar um mapa já treinado e uma máquina de estados (parte operativa e de controle) para controlar os processos, além de registradores e multiplexadores.

Alguns dos blocos necessários para projetar o circuito em questão em VHDL (memória e comparador, por exemplo) se encontram disponíveis e programados na biblioteca da ferramenta da Xilinx, bastando instanciá-los na rotina principal. Já outros blocos, tais como o de cálculo da distância Euclidiana, registradores e máquina de estados, foram programados, como visto na Fig. 63, como descrito anteriormente para o circuito de detecção.

a) Arquitetura do Circuito de Classificação

Sendo assim, após a definição dos blocos necessários, programou-se, em VHDL, o circuito de classificação cuja arquitetura simplificada pode ser vista na Fig. 65. Inicialmente, o registrador de endereços armazena a posição de memória BRAM que será acessada. Nesse endereço, está armazenado o valor de \mathbf{W}^j para cada iteração. Esta posição de memória também será repassada ao registrador vencedor, caso este seja o neurônio vencedor. O valor de \mathbf{W}^j é armazenado em um registrador (dado da BRAM) a cada iteração.

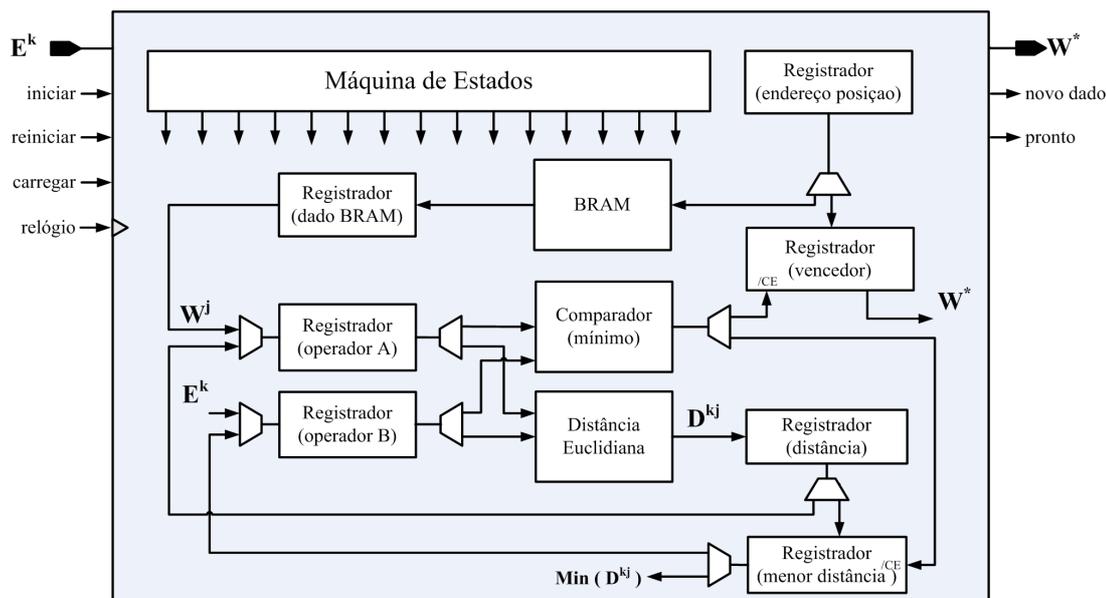


Figura 65: Arquitetura do circuito de classificação dos mapas auto-organizáveis.

Após, o valor de W^j e E^k são enviados aos registradores operador A e operador B e repassados ao bloco da distância Euclidiana. Após, o valor de D^{kj} é comparado com o último valor armazenado no registrador de menor distância. Caso a comparação (de menor valor) seja verdadeira, a posição de memória de W^j é repassada ao registrador vencedor. Os sinais externos (iniciar e pronto, por exemplo) vistos na Fig. 65 são sinais de operação do circuito.

Máquina de Estados do Circuito de Classificação

Após a definição da arquitetura do circuito de classificação, efetuou-se a programação da máquina de estados, Fig. 66, composta por 14 estados, cujos principais são descritos a seguir:

- S0:** inicializando registradores;
- S2:** os contadores são reinicializados;
- S3:** carrega e armazena os dados da memória;
- S6:** realiza o cálculo da distância euclidiana;
- S8:** atualiza e incrementa o endereço da memória;
- S12:** armazena o menor valor.

Os sinais *dist_done*, *dist_new_data*, *comp_rdy_add* e *comp_operator_fwd*, vistos na Fig. 66, são sinais de confirmação de chegada e finalização de cálculos dos blocos distância Euclidiana e comparador, respectivamente. Os sinais *reg_contador* e *reg_neuronios* são sinais de registradores que armazenam o número de iterações e de neurônios do mapa, respectivamente. O sinal *n_clocks* representa o número de pulsos de relógio.

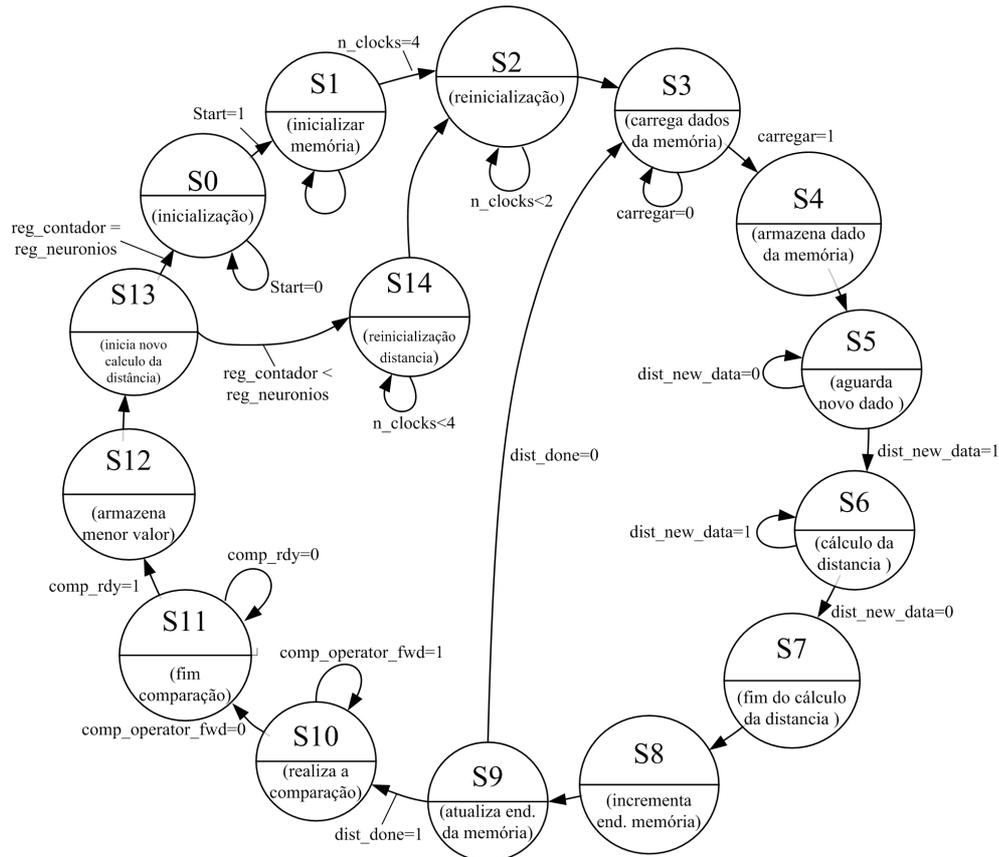


Figura 66: Máquina de estados do circuito de classificação dos MAO.

c) Processos do Circuito de Classificação

Com base na arquitetura e na máquina de estados do circuito de classificação, foram programados os processos em VHDL, dentre os quais se destaca: processo para controlar a lógica de próximo estado, processo para controlar os registradores (endereço da memória, dado da BRAM, operadores, distância euclidiana, menor erro e vencedor, por exemplo), processo para controlar a máquina de estados (parte operativa e de controle).

5.1.1.3 Circuito de Predição

A predição de falhas, através dos MTK, é baseada na determinação do neurônio vencedor (\mathbf{W}^{j*}), Eq. (58) e da ativação temporal (A^{kj}), Eq. (59). Para implementar essas equações em VHDL, observou-se que seria necessário programar, inicialmente, um bloco para obter o valor da ativação temporal (definido como circuito de ativação) e um para o neurônio vencedor (definido como circuito de predição).

a) Arquitetura do Circuito de Ativação

Programou-se, inicialmente, em VHDL, o circuito visto na Fig. 67 para determinação da ativação temporal. Por meio da máquina de estados os valores de A^{kj} e D^{kj} são enviados aos blocos de subtração e multiplicação, passando pelos seus respectivos registradores.

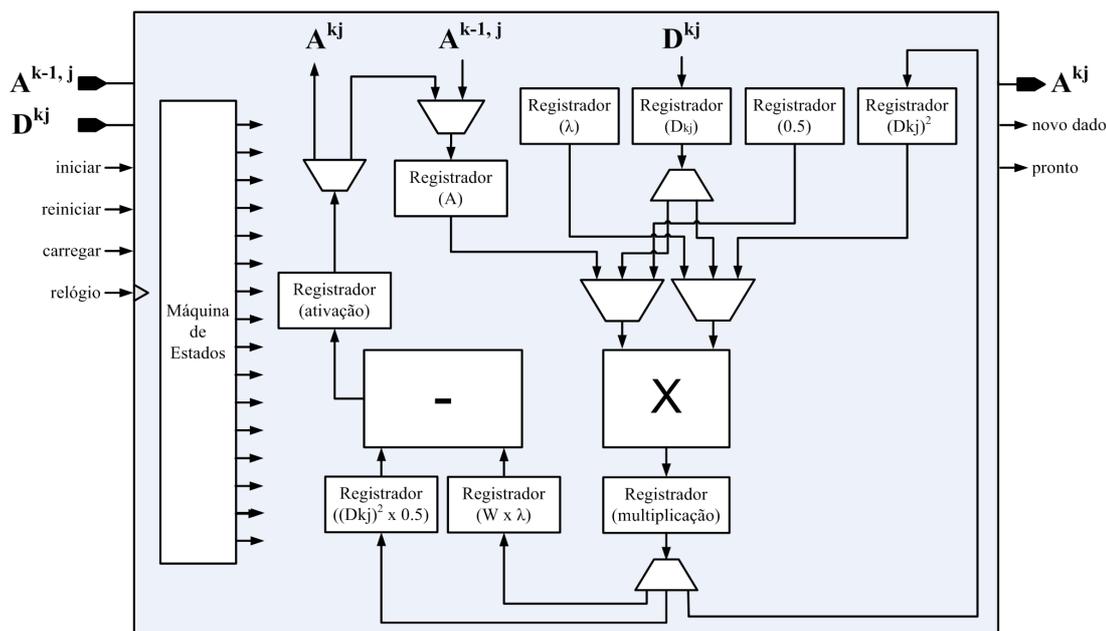


Figura 67: Arquitetura do circuito de ativação dos mapas temporais de Kohonen.

Inicialmente, o valor de D^{kj} é repassado ao multiplicador (em ambas as entradas) para que seja feita a obtenção do quadrado de D^{kj} . Após, é efetuada a multiplicação de $(D^{kj})^2$ por $1/2$. A seguir, é realizada a multiplicação de λ pela ativação $A^{k-1,j}$. Por fim, é realizada a subtração e o valor da ativação A^{kj} é colocado na saída do circuito. Cabe ressaltar, que durante estas operações, os valores vão sendo armazenados nos registradores vistos na Fig. 67.

Os sinais iniciar, reiniciar, carregar, pronto, e novo dado, vistos na Fig. 67, são sinais externos responsáveis pela operação do circuito de ativação. O sinal relógio é responsável pelo envio dos pulsos de relógio ao circuito de ativação.

b) Máquina de Estados do Circuito de Ativação

A máquina de estados simplificada do circuito de predição possui 16 estados, como observado na Fig. 68. Uma breve descrição dos principais estados é vista a seguir.

S0: os registradores são inicializados;

S1: carrega os novos dados;

S3: efetua a multiplicação;

S8: registra a multiplicação;

S12: realiza a subtração;

S14: armazena os resultados nos registradores e coloca o resultado final na saída.

A transição entre os estados foi efetuada em função de uma série de sinais (`rdy_sub`, `operation_rfd_mult` e `rdy_mult`, por exemplo), dos blocos de multiplicação e subtração.

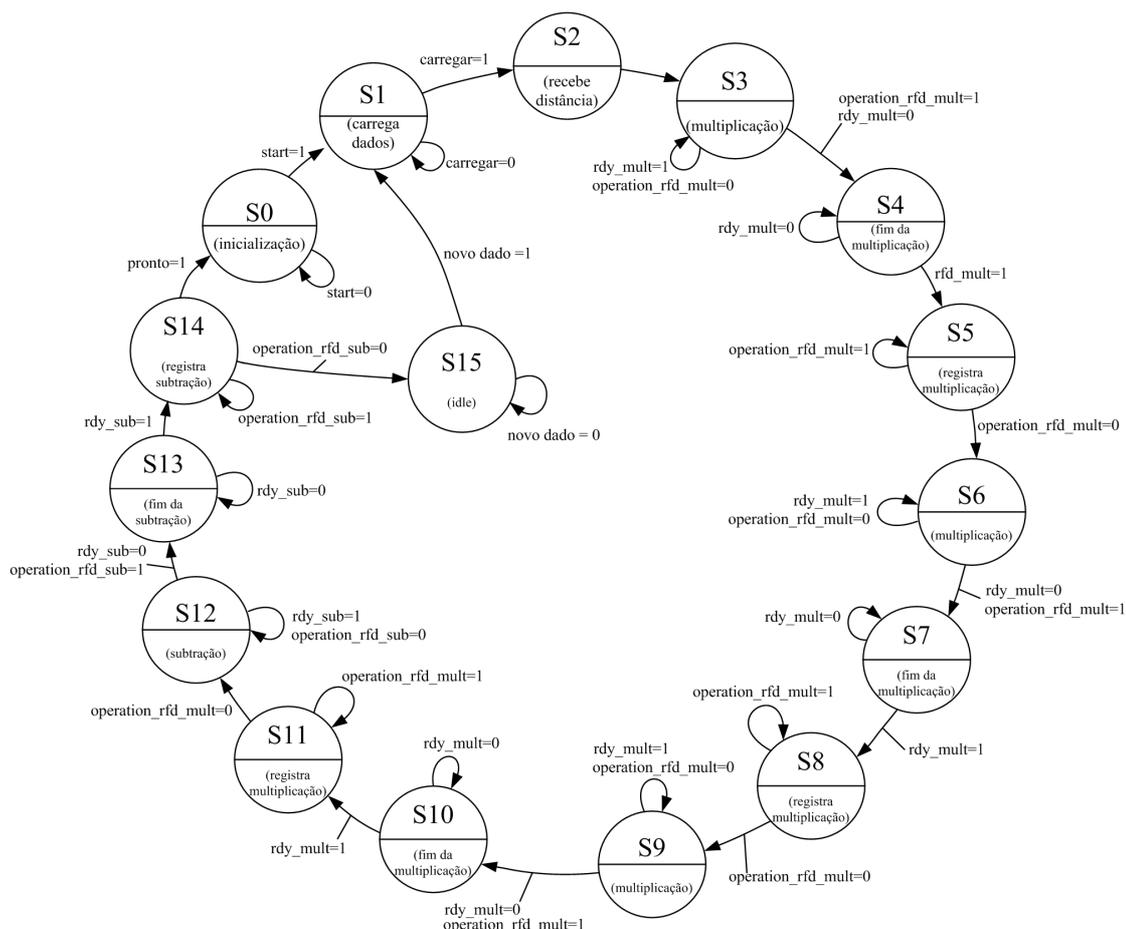


Figura 68: Máquina de estados do circuito de ativação dos MTK.

c) Processos do Circuito de Ativação

Os processos, escritos em VHDL, para o circuito de ativação foram programados com base na arquitetura e na máquina de estados do circuito de ativação, vistos anteriormente. Dentre os processos destaca-se: processo para controlar a lógica de próximo estado, processo para dirigir os registradores (ativação e multiplicação, por exemplo), processo para controlar a máquina de estados (parte operativa e de controle).

d) Arquitetura do Circuito de Predição

Para implementar o circuito de predição e determinar o neurônio vencedor (\mathbf{W}^*) constatou-se que seria necessário usar um bloco comparador, um bloco para a determinação da ativação temporal, um bloco para obter a distância Euclidiana, e BRAMs de leitura e escrita, além da máquina de estados para controlar os processos envolvidos.

Assim, após a definição dos blocos necessários, programou-se, em VHDL, o circuito de predição cuja arquitetura simplificada pode ser vista na Fig. 69. Inicialmente, o registrador de endereços armazena a posição do vetor na memória BRAM (leitura) que será acessada.

Os valores relativos ao tamanho da série temporal, ao número de elementos dos vetores e ao número de neurônios no mapa são repassados para a máquina de estados, a seguir.

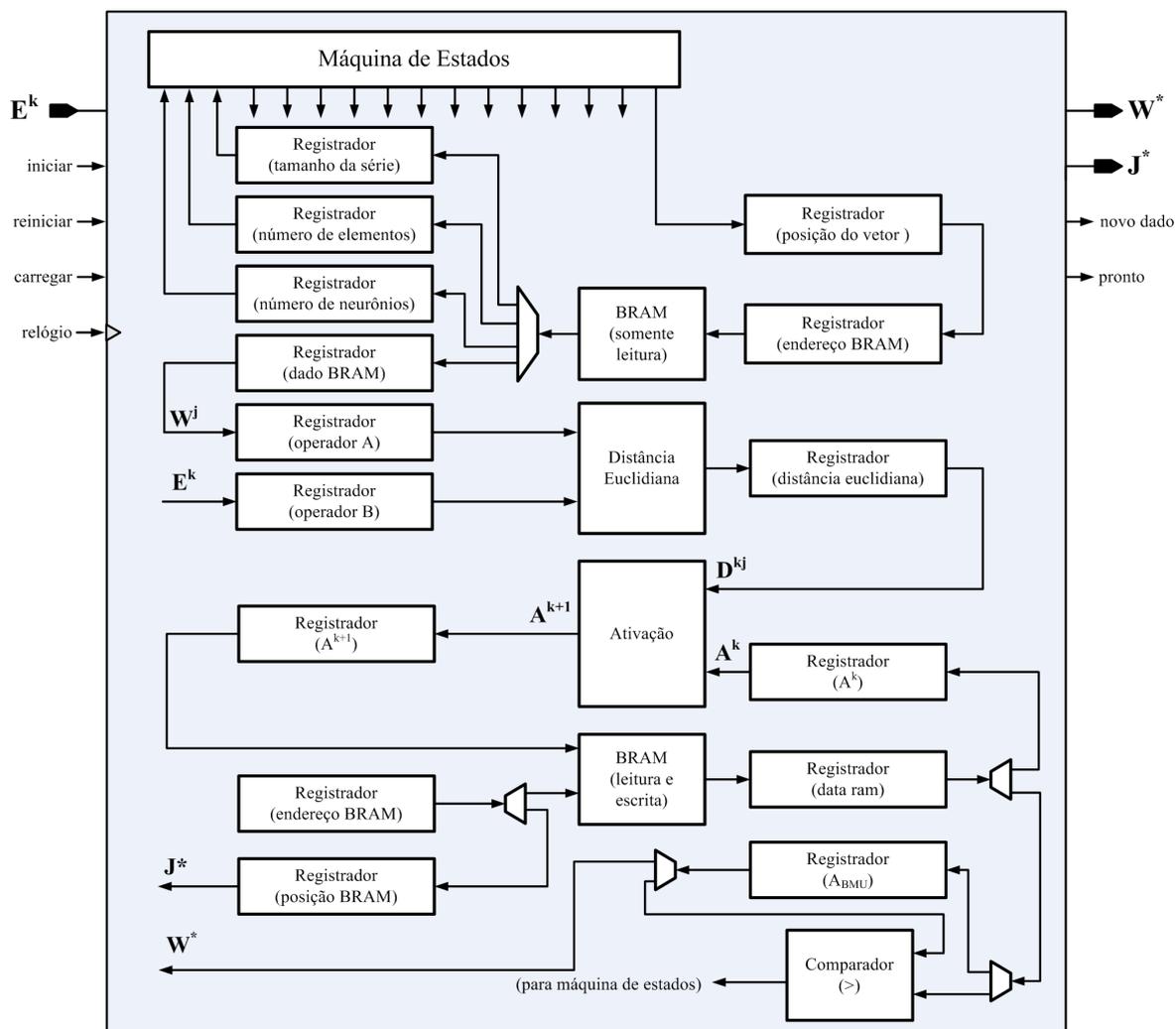


Figura 69: Arquitetura do circuito de predição dos mapas temporais de Kohonen.

Após, os valores dos pesos sinápticos (W^j) e de entrada (E^k) são enviados ao bloco distância Euclidiana para a determinação de D^{kj} . A seguir, o valor de ativação armazenado na BRAM (leitura e escrita) e valor da distância Euclidiana são repassados ao bloco Ativação para o cálculo de A^{kj} . Por fim, através do comparador é determinado o valor do neurônio vencedor (W^{j^*}) e a posição (J^*) desse no mapa.

Os sinais iniciar, reiniciar, carregar, pronto, e novo dado, vistos na Fig. 69, de modo semelhante aos circuitos de detecção e classificação, são sinais externos responsáveis pela operação do circuito de predição. O sinal relógio é encarregado do envio dos pulsos de relógio ao circuito de predição.

e) Máquina de Estados do Circuito de Predição

A máquina de estados, simplificada, do circuito de predição, responsável pelo controle do circuito visto na Fig. 69, possui 15 estados, como visto na Fig. 70, cujos principais são resumidos a seguir:

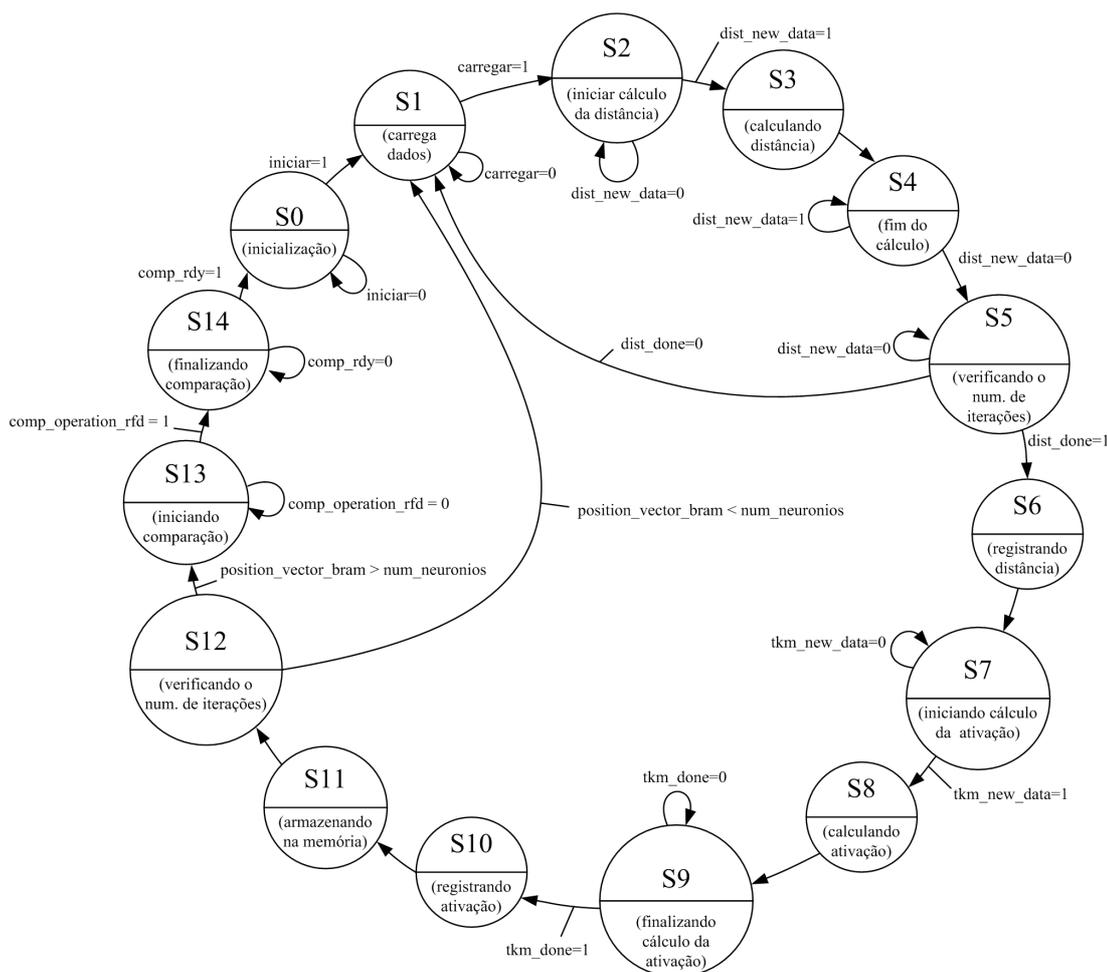


Figura 70: Máquina de estados do circuito de predição do MTK.

S0: inicializando registradores e demais blocos;

S1: carregando novos dados;

S3: a distância euclidiana é calculada;

S8: o cálculo da ativação é efetuado;

S10: registrando o valor da ativação;

S13: é realizada a comparação.

Os sinais `dist_new_data`, `tkm_done`, `comp_rdy`, por exemplo, vistos na Fig. 70, são sinais de confirmação de chegada e saída de dados dos blocos de determinação da distância Euclidiana, da ativação temporal e de comparação dos valores de ativação, respectivamente.

f) Processos do Circuito de Predição

Em função da arquitetura e da máquina de estados do circuito de predição, vistos nas Fig. 69 e 70, foram programados diversos processos em VHDL, dentre os quais destaca-se:

processo para controlar a lógica de próximo estado, processos para os registradores (endereço da memória, dado da BRAM, operadores, distância euclidiana e ativação, por exemplo) e os processos envolvidos no controle e operação da máquina de estados.

5.1.2 Embarque dos Filtros Adaptativos

No embarque dos FA, os circuitos programados em VHDL serão definidos como circuitos de detecção, classificação e predição, semelhante ao efetuado para os MAO.

5.1.2.1 Circuito de Detecção

A detecção de falhas, efetuada através dos FA, é realizada com base no cálculo do erro de detecção, (e), Eq. (67). Assim, com base na Eq. (67) constatou-se que para implementar o cálculo do erro seria necessário usar um bloco para se obter a saída (y) do filtro FIR (definido como circuito FIR) e um alguns outros blocos para realizar a subtração entre \hat{E} e y (circuito de detecção).

a) Arquitetura do Circuito FIR

Para implementar o cálculo da saída do filtro FIR, como visto na Fig. 110, observou-se que seria necessário usar um bloco somador, um bloco multiplicador e uma memória para armazenar os coeficientes do FA obtidos durante o treinamento, além de registradores e multiplexadores. Após esta definição dos blocos necessários, programou-se, em VHDL, o circuito do filtro FIR cuja arquitetura simplificada pode ser vista na Fig. 71.

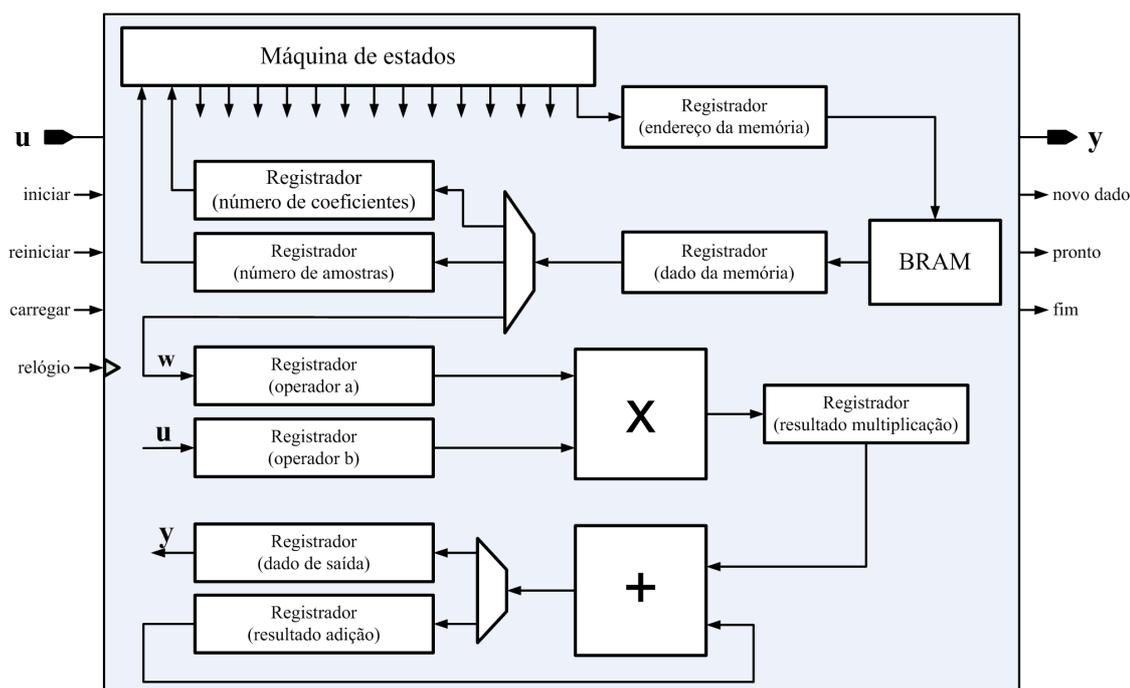


Figura 71: Arquitetura do circuito do filtro adaptativo FIR.

Inicialmente, são repassados da memória para os registradores alguns parâmetros importantes para a operação do circuito, tais como: o número de coeficientes do filtro e o número de amostras do vetor de entrada. Após, os valores de \mathbf{w} (oriundos da memória) e \mathbf{u} (valores de entrada) são enviados aos registradores operador a e operador b e a seguir repassados ao bloco de multiplicação. A seguir, é efetuada a adição pelo bloco somador para se obter o valor da saída do filtro \mathbf{y} .

Os sinais externos ao bloco (iniciar, carregar e pronto, por exemplo), vistos na Fig. 71, são responsáveis pela operação do circuito. O sinal relógio é responsável pelo envio dos pulsos de relógio ao circuito.

b) Máquina de Estados do Circuito FIR

A máquina de estados simplificada do circuito FIR, responsável pela operação e controle das operações do circuito, possui 15 estados, como observado na Fig. 72. Uma breve descrição de alguns estados é apresentada a seguir.

S0: inicializando registradores e demais blocos;

S2: carregando novos dados;

S3: iniciando multiplicação;

S7: finalizando adição;

S9: verificando o número de iterações;

S14: finalizando o cálculo da saída do filtro.

Os sinais `rdy_sum`, `operation_rfd_sum`, `rdy_mult`, `operation_rfd_mult`, por exemplo, vistos na Fig. 72, são sinais de confirmação de saída e chegada de dados dos blocos somador e multiplicador, respectivamente.

c) Processos do Circuito FIR

Com base na arquitetura e na máquina de estados do circuito FIR, foram programados os processos em VHDL para esse circuito, dentre estes se destaca: processos para a parte operativa e de controle da máquina de estados, processo para dirigir a lógica de próximo estado, processos para controlar os registradores (endereço da memória, número de coeficientes e operadores, como exemplo).

d) Arquitetura do Circuito de Detecção

Para desenvolver a tarefa de detecção em VHDL, através dos FA, como visto na Fig. 39, constatou-se que seria necessário usar um bloco para se obter a saída do filtro FIR, visto na Fig. 71, e um subtrator, além de registradores. Após a definição dos blocos, programou-se, em VHDL, o circuito de detecção cuja arquitetura simplificada pode ser vista na Fig. 73.

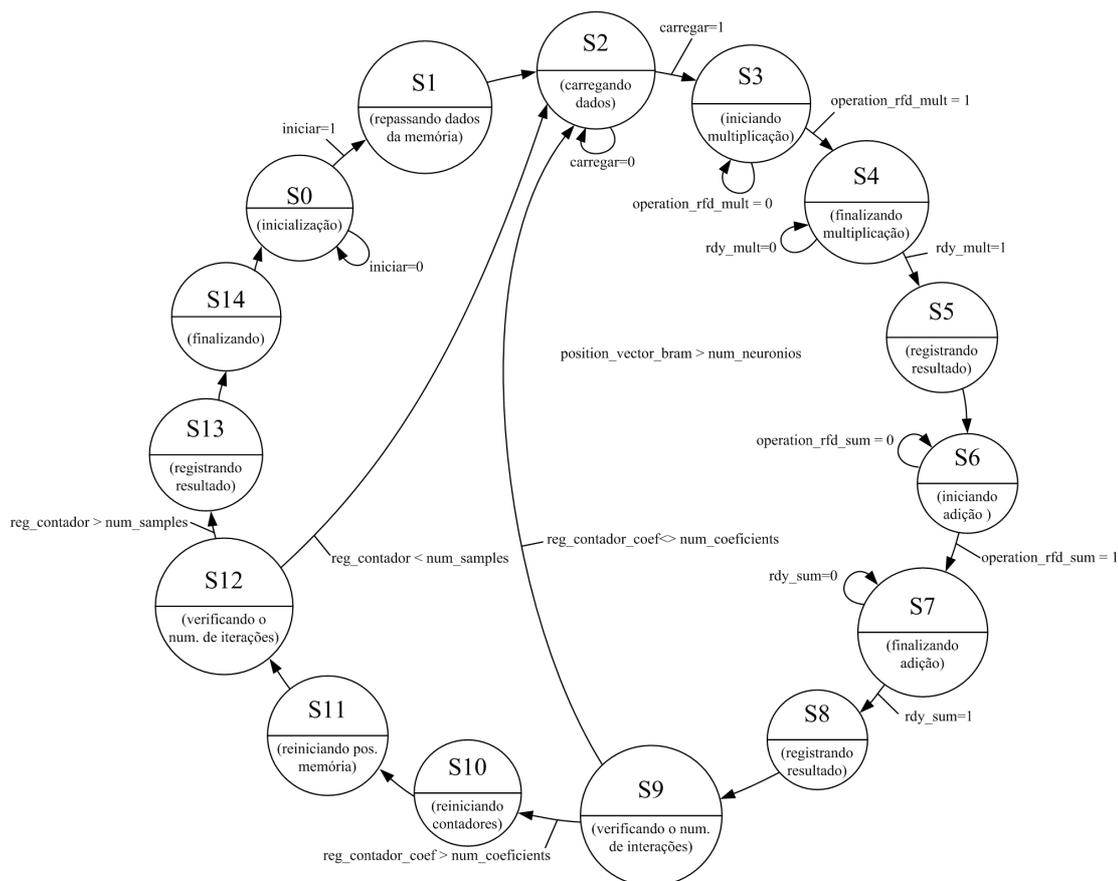


Figura 72: Máquina de estados do circuito do filtro adaptativo FIR.

Inicialmente, os dados de entrada ($\hat{\mathbf{E}}$ e \mathbf{u}) são repassados aos seus respectivos registradores. Após, o bloco Filtro FIR calcula a saída do filtro (\mathbf{y}). Por fim, o erro (\mathbf{e}) é obtido através do bloco subtrator. Os sinais externos ao bloco (*iniciar*, *pronto*, etc.) apresentam as mesmas funcionalidades descritas anteriormente.

e) Máquina de Estados do Circuito de Detecção

A máquina de estados, simplificada, do circuito de detecção, responsável pelo controle e operação do circuito visto na Fig. 73, possui 12 estados, como pode ser observado na Fig. 74, cujos principais são descritos, resumidamente, a seguir:

S0: inicializando registradores e demais blocos;

S2: carregando novos dados;

S3: calculando a saída do filtro FIR;

S6: iniciando subtração;

S11: finalizando o cálculo da saída do erro.

Os sinais *rdy_sub* e *operation_rfd_sum*, são sinais de confirmação de saída e chegada de dados do blocos subtrator. O sinal *done_filter* sinaliza o fim do cálculo de \mathbf{y} .

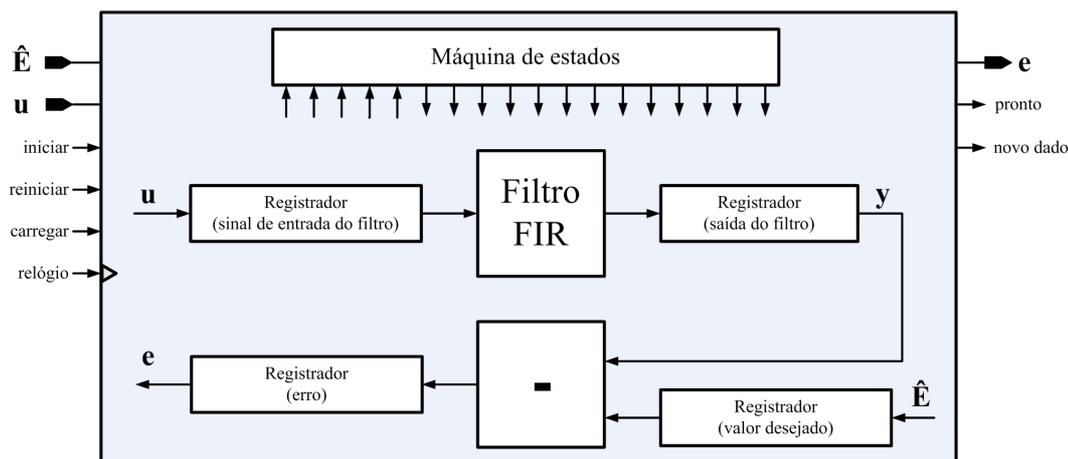


Figura 73: Arquitetura do circuito de detecção dos filtros adaptativos.

f) Processos do Circuito de Detecção

Em função da arquitetura e da máquina de estados do circuito de detecção, vistos nas Fig. 73 e 74, foram programados os seguintes processos em VHDL: processo para controlar a lógica de próximo estado, processos para controlar os registradores (entrada, erro, saída do filtro e valor desejado) e os processos necessários para a máquina de estados.

5.1.2.2 Circuito de Classificação

A classificação de falhas, através dos FA, é efetuada com base no cálculo do erro de classificação, (e), Eq. (67). Deste modo, observou-se que para implementar o cálculo do erro seria necessário usar os mesmos blocos (máquinas de estados e processos também) utilizados para a tarefa de detecção: um bloco para se obter a saída (y) do filtro FIR, definido como circuito FIR, Fig. 71, e um bloco para realizar a subtração entre \hat{E} e y , como na Fig. 73.

5.1.2.3 Circuito de Predição

A tarefa de predição de falhas, por meio dos FA, é realizada com base no cálculo da saída do filtro FIR (y), Eq. (68), na determinação do erro (e), Eq. (69), e na atualização dos coeficientes do filtro (w), Eq. (68). Assim, para desenvolver o cálculo da predição embarcado foi necessário programar um bloco para se adquirir a saída do filtro FIR (circuito FIR, descrito anteriormente), um bloco para se obter o valor do erro (um subtrator, propriamente dito) e um bloco para se determinar a atualização dos coeficientes do filtro adaptativo (definido como circuito de atualização), além de registradores, multiplexadores e memórias (BRAM).

a) Arquitetura do Circuito de Atualização

Para implementar o cálculo da atualização dos coeficientes do FA, Eq. (68), como visto na Fig. 44, observou-se que seria necessário usar um bloco somador e um bloco multiplicador, além de multiplexadores e registradores.

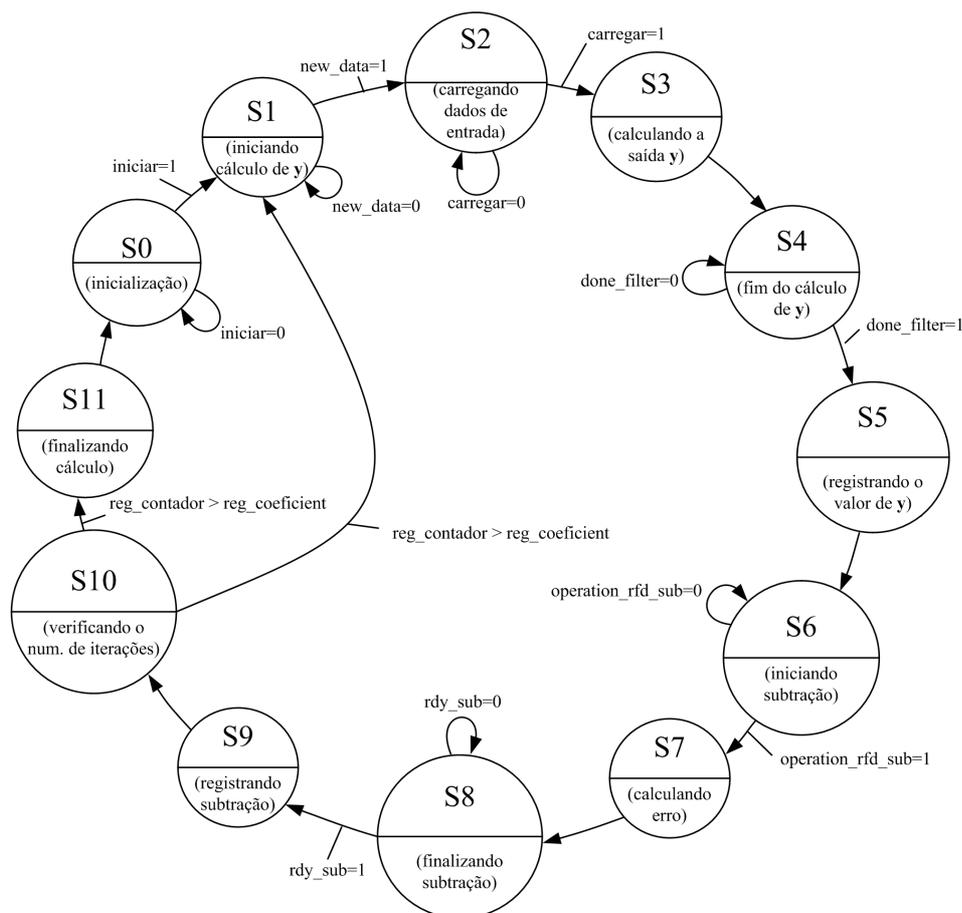


Figura 74: Máquina de estados do circuito de detecção dos filtros adaptativos.

Assim, programou-se, em VHDL, o circuito de atualização cuja arquitetura simplificada pode ser vista na Fig. 75. Os dados de entrada (\mathbf{w} , $\hat{\mathbf{E}}$, \mathbf{e} e μ), inicialmente, foram repassados aos seus respectivos registradores. Após, o bloco multiplicador realiza a multiplicação entre o erro e a taxa de atualização (μ). A seguir, a multiplicação entre o sinal de entrada $\hat{\mathbf{E}}$ e o resultado da multiplicação anterior ($\mathbf{e}\mu$) é efetuada. Por fim, é realizada a adição entre o vetor de coeficientes e o resultado da última multiplicação. Assim, o valor do vetor \mathbf{w} é atualizado.

b) Máquina de Estados do Circuito de Atualização

A máquina de estados, simplificada, do circuito de atualização, visto na Fig. 75, possui 13 estados, como pode ser observado na Fig. 76. Três estados são descritos, resumidamente, a seguir:

S2: carregando novos dados de entrada;

S3: iniciando multiplicação;

S12: finalizando o cálculo da atualização.

Os sinais rdy_mult , rdy_add , $\text{operation_rfd_mult}$, operation_rfd_add , são sinais de confirmação de saída e chegada de dados do blocos subtrator.

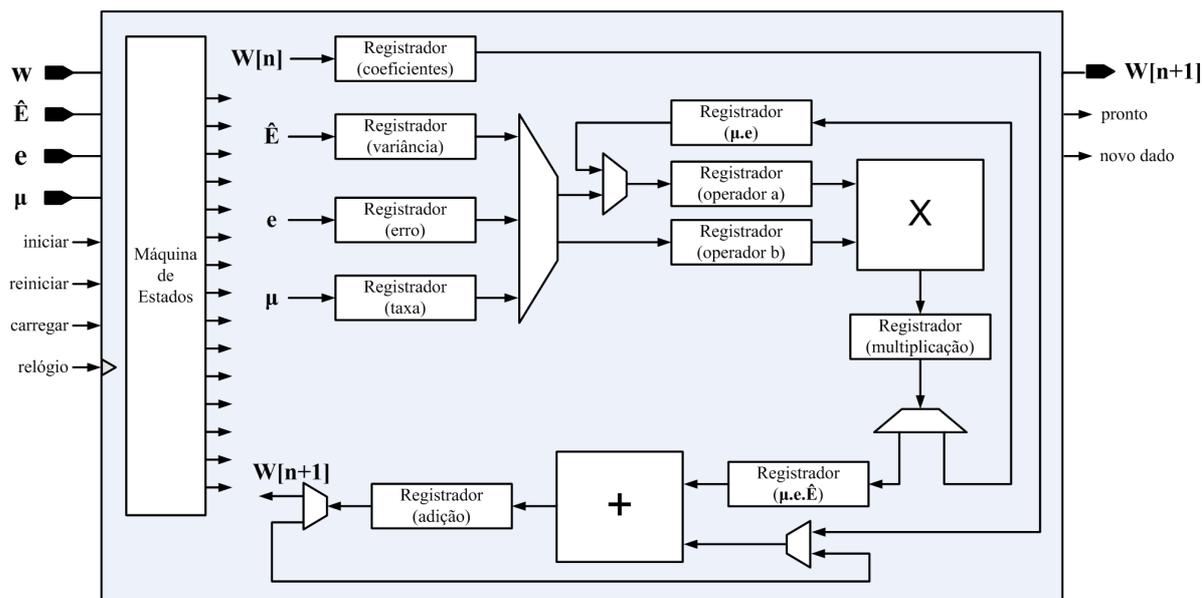


Figura 75: Arquitetura do circuito de atualização dos filtros adaptativos.

c) Processos do Circuito de Atualização

Após a definição da arquitetura e da máquina de estados do circuito de atualização, vistos nas Fig. 75 e 76, foram programados os seguintes processos em VHDL: processo para controlar a lógica de próximo estado, processos para controlar os registradores (coeficientes, variância, taxa, entre outros) e os processos necessários para a operação e controle da máquina de estados.

d) Arquitetura do Circuito de Predição

A arquitetura simplificada do circuito voltado à tarefa de predição de falhas através dos FA é vista na Fig. 77. Nessa figura pode-se observar que foram utilizados dois blocos de memória BRAM (um para provocar o atraso nas amostras de entrada, \hat{E} , e outro para armazenar os coeficientes do filtro, w , após a sua atualização), um bloco para calcular a saída do filtro FIR (descrito anteriormente, Fig. 71), um bloco subtrator para se obter o valor do erro, e , e um para se determinar a atualização dos coeficientes (circuito de atualização, visto na Fig. 75).

Inicialmente, o dado de entrada (\hat{E} é repassado, M amostras) ao seu respectivo registrador e depois enviado para a BRAM 1 onde, através de processos, ocorrerá um atraso temporal de uma unidade. Após, estes dados em conjunto com os coeficientes do FA (w) são enviados ao bloco Filtro FIR (sem a necessidade do bloco de memória BRAM, neste caso) para o cálculo de y . Depois, o erro (e) entre a saída do filtro e o dado de entrada é obtido. Por fim, os M valores dos coeficientes do filtro são atualizados e armazenados na BRAM 2.

Os sinais externos ao circuito de predição (iniciar, reiniciar, carregar, relógio, pronto e novo dado) apresentam as mesmas funcionalidades descritas anteriormente para os outros blocos.

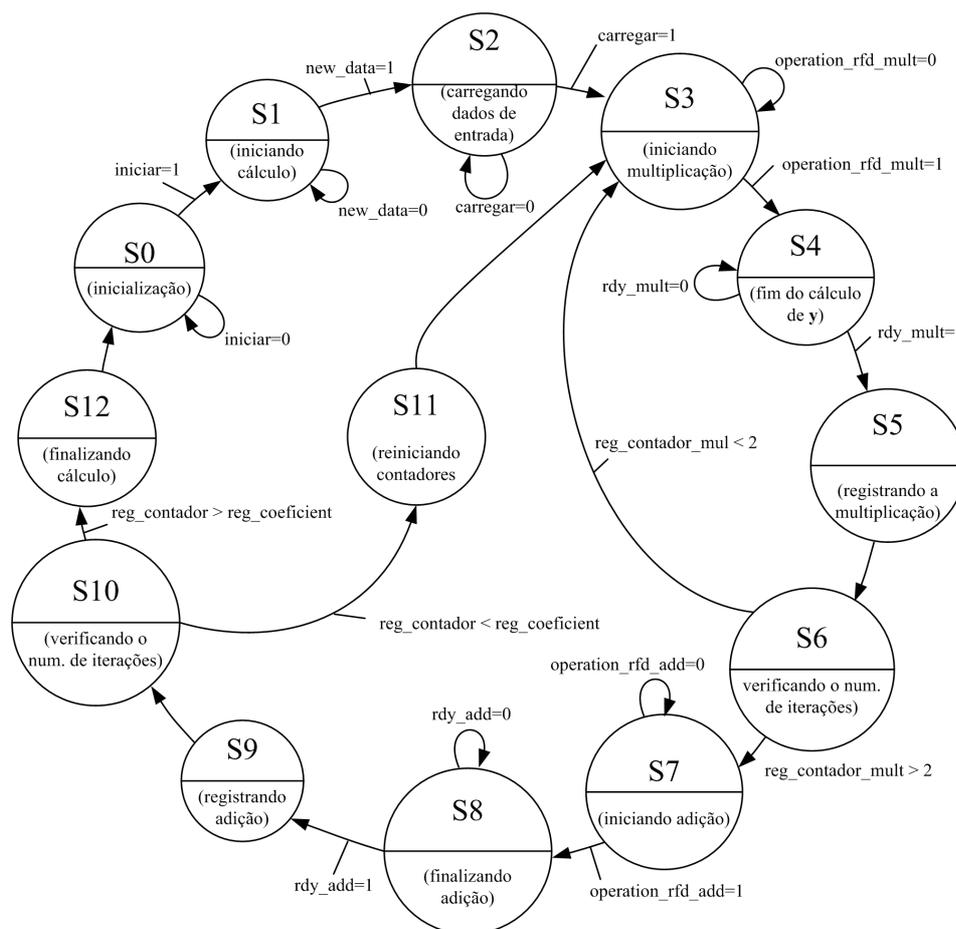


Figura 76: Máquina de estados do circuito de atualização dos filtros adaptativos.

e) Máquina de Estados do Circuito de Predição

A máquina de estados, simplificada, do circuito de predição, visto na Fig. 77, possui 16 estados, como pode ser observado na Fig. 78. Alguns destes estados são descritos, resumidamente, a seguir:

- S0:** inicializando registradores e demais blocos;
- S2:** atrasando a amostra de entrada;
- S8:** calculando a saída do filtro FIR;
- S12:** obtendo o valor do erro;
- S14:** finalizando o cálculo da atualização.

Os sinais `new_data_fp`, `new_data_af`, `done_fp` e `done_af`, são sinais de confirmação de chegada e saída de dados dos blocos filtro FIR e atualização. Os sinais `reg_samples_number` e `reg_coef_number` são sinais que representam o número de amostras e de coeficientes do filtro, respectivamente.

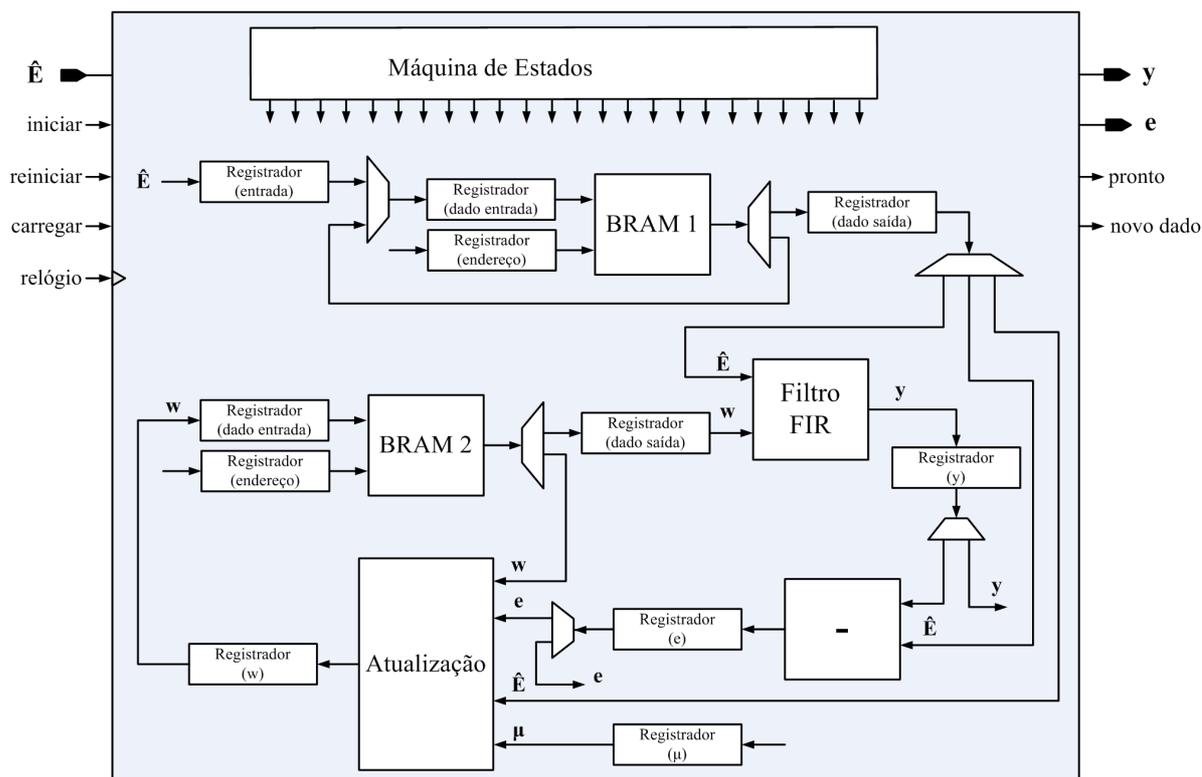


Figura 77: Arquitetura do circuito de predição dos filtros adaptativos.

Os sinais `reg_contador` e `reg_position_bram` são usados, respectivamente, para armazenar o número de interações e a posição de endereçamento da memória (BRAM) nos seus respectivos registradores.

f) Processos do Circuito de Predição

A partir da arquitetura e da máquina de estados do circuito de predição, vistos nas Fig. 77 e 78, respectivamente, foram programados os seguintes processos em VHDL: processo para controlar a lógica de próximo estado, processos para controlar os registradores (dados de entrada e saída das memórias, erro, saída do filtro e coeficientes, por exemplo) e os processos necessários para o controle e operação da máquina de estados.

5.2 PLATAFORMA DE EMBARQUE

Dentre as plataformas de embarque que utilizam os mais diversos tipos de dispositivos semicondutores voltados para o processamento de informações digitais, tais como: ASIC (*Application Specific Integrated Circuit*) e FPGA (*Field Programmable Gate Array*), por exemplo, optou-se por uma plataforma de embarque que utilizasse FPGA.

Esta escolha se deve a facilidade e flexibilidade de programação por *software*, depuração, análise dos sinais e embarque dos circuitos, escritos em VHDL, em FPGA que este tipo de plataforma apresenta.

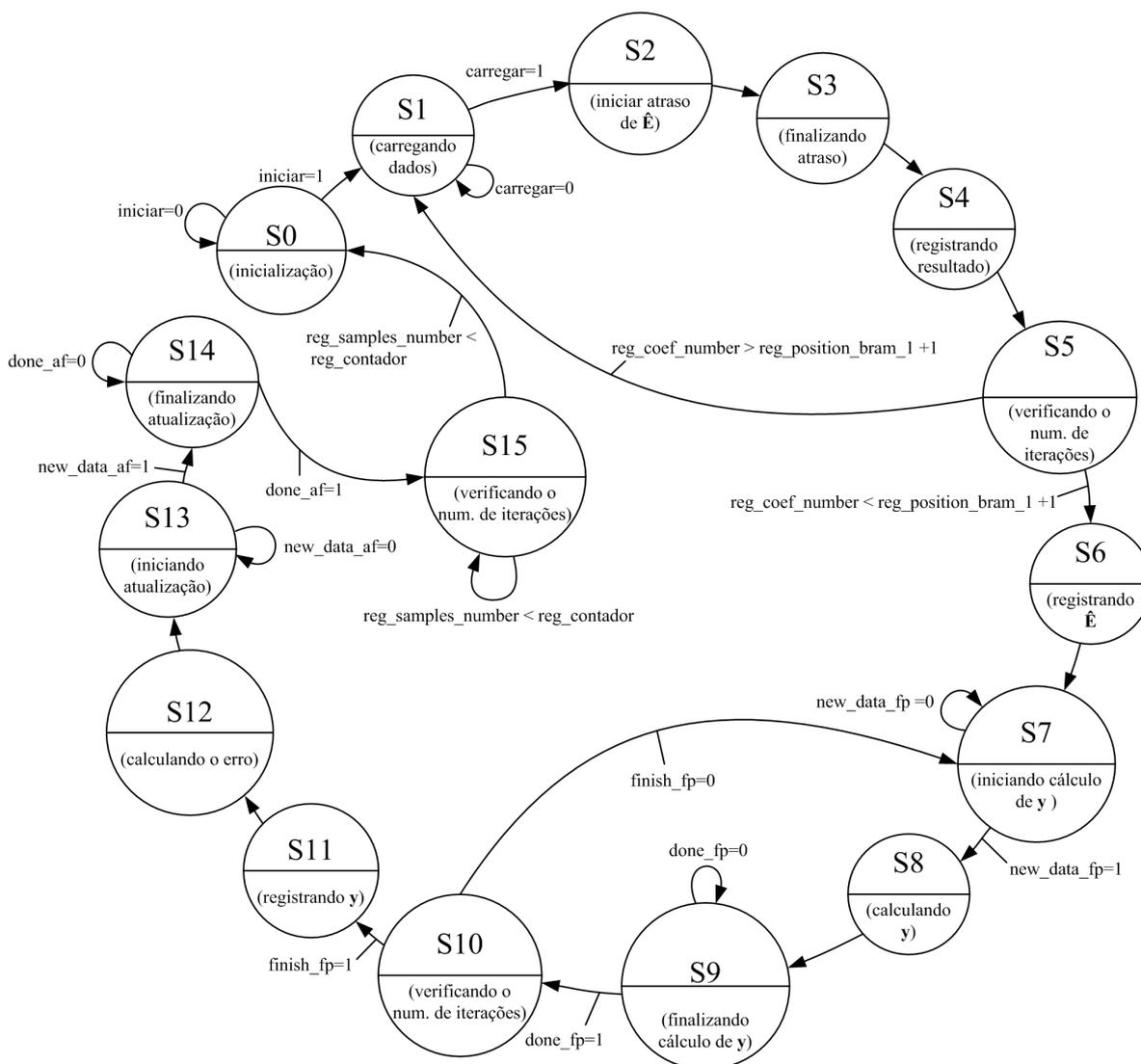


Figura 78: Máquina de estados do circuito de predição dos filtros adaptativos.

Sendo assim, o protótipo dos circuitos do SMI foi implementado em uma matriz de FPGA fabricada pela empresa Xilinx. Essa matriz é acoplada a uma plataforma de desenvolvimento, uma placa de circuito impresso, fabricada pela empresa Digilent modelo XUP Virtex-2 PRO (XILINX, 2010).

Essa placa possui uma arquitetura adequada para projetos de sistemas embarcados, pois é totalmente configurável, faz uso de periféricos padronizados e difundidos no mercado, tais como: portas USB, interface serial RS-232, conversores A/D e D/A e interface para memória RAM DDR, por exemplo. O chip do FPGA embarcado na placa é do modelo Virtex-2 PRO (XC2VP30) e possui blocos lógicos programáveis, multiplicadores, memória BRAM e dois processadores *Power PC* integrados.

Inicialmente, o sistema embarcado foi projetado (somente para os MAO) usando a arquitetura de processamento *Microblaze* acoplando periféricos ao barramento de comunicação de dados (XILINX, 2010).

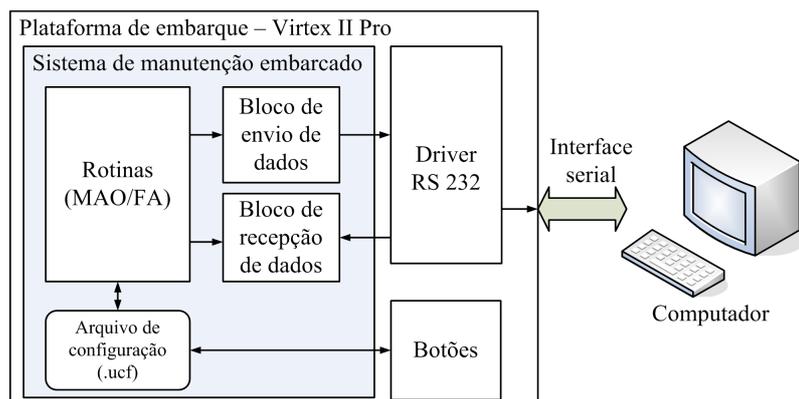


Figura 79: Estratégia de envio/recepção de dados do sistema embarcado.

Esse sistema possui um ambiente computacional completo, com 512Mb de memória RAM, *flash* e entrada/saída serial. Também foi instalado na memória *flash* o Petalinux, um sistema operacional voltado para aplicações embarcadas embutido como o Unix. O Petalinux é o responsável pelas configurações do sistema operacional e pela integração do sistema embarcado em VHDL com a placa.

Para isso, foi desenvolvido, em linguagem C, uma série de rotinas de cálculo, envio e recepção de dados para o sistema escrito em VHDL. Estas rotinas também são responsáveis pela conversão dos sinais do sistema em VHDL para o padrão de comunicação do barramento usado na placa. Através destas rotinas foi possível fazer um conjunto de experimentos de detecção, classificação, e predição de falhas em tempo real. Mais detalhes e resultados a respeito desta parte do embarque podem ser encontrados em (BOSA, 2009)

Atualmente, foi implementado um outro tipo de estratégia de envio e recepção de dados com a plataforma e de comunicação com o computador sem a utilização de nenhum sistema operacional. Nesta nova estratégia, o sistema embarcado na placa realiza a comunicação com o computador através de uma interface serial de envio e recepção de dados (blocos de envio e recepção de dados) embarcada e programada junto com o SMI embarcado, como visto na Fig. 79. Assim, a transmissão e recepção de dados, entre computador e plataforma de embarque, se dão através da interface serial (padrão RS232).

O controle (inicialização e finalização dos cálculos e envios dos dados ao computador, por exemplo) do sistema embarcado se dá através de botões presentes na placa. Um arquivo de configuração é o responsável pela definição das funções que cada botão na placa irá executar (iniciar cálculos, enviar dados e finalizar operações, por exemplo) e dos sinais de envio e recebimento dos dados (relógio, iniciar e finalizar, como exemplo) dos blocos de detecção, classificação e predição de falhas do SMI. Esse arquivo de configuração é anexado junto as rotinas do sistema em VHDL e utilizado somente durante a síntese das rotinas.

Com esta nova estratégia reduz-se a área ocupada (o aumento de área ocorre em virtude da adição de rotinas extras para integrar o sistema embarcado ao barramento da placa) e o tempo de envio e recepção de dados.

5.2.1 Especificação dos Mapas Auto-organizáveis para Embarque

No projeto dos circuitos dos MAO, foram assumidos certos limites máximos para o tamanho dos dados de processamento: o número máximo de neurônios é 512, a dimensão máxima dos vetores de pesos sinápticos é de 32 bits, cada elemento do vetor de entrada também tem 32 bits, e os números foram representados no formato de ponto flutuante (IEEE-1, 1990; IEEE-2, 1995; MUÑOZ et al., 2010).

O principal limitador observado no embarque deste projeto foi o tamanho da memória necessária para armazenar o mapa já treinado. Neste caso, o tamanho da memória é proporcional ao número de neurônios (J) do mapa; isto é, às dimensões do mapa auto-organizável ($J = L_m C_m$), onde: L_m e C_m são o número de linhas e de colunas do MAO, respectivamente.

As dimensões do MAO, previamente gerado e treinado em *software*, gravado na BRAM da placa, bem como as dimensões dos vetores, como descrito na Tab. 5, foram escolhidas para fornecerem uma razoável precisão, em termos qualitativos e quantitativos dos resultados, e para não extrapolar capacidade física do FPGA.

Tabela 5: Especificação do mapa auto-organizável para embarque.

Parâmetro	Valor
Dimensões do mapa	15 linhas X 6 colunas
Número de neurônios	90 neurônios
Dimensão dos vetores	40 elementos
Dimensão de cada elemento do vetor	32 bits

5.2.2 Especificação dos Filtros Adaptativos para Embarque

De modo semelhante aos MAO, no projeto dos circuitos dos FA também foram definidos certos limites máximos para o tamanho dos dados de processamento: o número de coeficientes e a dimensão dos vetores são iguais a 6, a dimensão de cada elemento do vetor é de 32 bits e os números foram representados no formato de ponto flutuante, como visto na Tab. 6.

Tabela 6: Especificação do filtro adaptativo para embarque.

Parâmetro	Valor
Taxa de atualização	2100
Número de coeficientes	6 coeficientes
Dimensão dos vetores	6 elementos
Dimensão de cada elemento do vetor	32 bits

O tamanho da memória, neste caso, não foi visto como um limitador do projeto dos circuitos dos FA em VHDL, ao contrário dos MAO. Neste caso, o tamanho da memória é proporcional ao número de coeficientes do mapa do mapa já treinado.

De modo semelhante aos MAO, as dimensões dos vetores, do número de coeficientes do FA, previamente gerado e treinado, gravado na BRAM da placa, bem como a dimensão de cada elemento foram definidas para fornecerem uma precisão qualitativa e quantitativa razoável dos resultados e para não extrapolar capacidade física do FPGA.

5.3 RESULTADOS DE EMBARQUE

Dentre os possíveis resultados gerados pela ferramenta da Xilinx (ISE e SimPower), que poderiam ser colocados como resultados de síntese e mapeamento do FPGA, necessários para se medir o desempenho dos circuitos desenvolvidos neste trabalho, destaca-se:

- Área: refere-se ao espaço físico ocupado pelos blocos funcionais, ou de entrada e saída, dentro do FPGA, necessário para o mapeamento dos circuitos lógicos;
- Desempenho: está relacionado ao tempo necessário para que certo conjunto de instruções seja processado e executado pelo sistema numa certa frequência de operação;
- Tempo de processamento: tempo necessário para a execução de determinada tarefa;
- Potência: refere-se ao consumo de corrente do circuito mapeado.

A principal idéia ao se programar um FPGA é sempre obter um circuito que ocupe a menor área, nunca ultrapassando a área máxima permitida pelo fabricante do FPGA, e opere com a maior velocidade possível. Contudo, a otimização de um destes parâmetros, área ou desempenho, geralmente irá afetar a performance do outro. Destaca-se que a potência não é um fator importante a ser levado em conta neste trabalho; pois, o atuador encontra-se alimentado com tensão e corrente suficiente para suprir a placa após o seu embarque.

Assim, serão apresentados apenas os resultados de área, desempenho e tempo de processamento dos circuitos de detecção, classificação e predição de falhas para os MAO e FA, apresentados na seção 5.1.

5.3.1 Área do Circuitos

A Tab. 7 apresenta os resultados relativos às áreas dos circuitos de detecção, classificação, e predição, para os mapas auto-organizáveis e filtros adaptativos, obtidos após o roteamento dos mesmos em termos do número de *Slices*, *Flip Flops*, *Luts*, *IOBs* e BRAMs utilizados.

Pode-se observar na Tab. 7 que o circuito de detecção baseado nos FA ocupa aproximadamente a mesma área no *chip* que o circuito de detecção obtido a partir dos MAO. Com relação ao circuito de classificação, o circuito feito a partir dos FA preenche uma área menor (exceto com relação ao número de *IOBs*); em especial, o número de BRAMs é bem menor ($\approx 20\%$). Já com relação ao circuito de predição, o circuito desenvolvido a partir dos MAO ocupa uma área um pouco menor que o circuito dos FA, exceto pelo número de células de memória que é cerca de 32% maior.

Tabela 7: Área do circuitos obtida após o roteamento.

Circuitos	Chip (Virtex-2 PRO)		Mapa auto-organizável		Filtro adaptativo	
	Parâmetro	Disponível	Usado	Utilização [%]	Usado	Utilização [%]
Circuito de detecção	<i>Slices</i>	13696	1694	12.36%	1880	13.72%
	<i>Flip Flops</i>	27392	2155	7.867%	2043	7.458%
	<i>Luts</i>	27392	1859	6.786%	2365	8.633%
	<i>IOBs</i>	556	102	18.34%	103	18.52%
	BRAMs	136	0	0%	1	0.731%
Circuito de classificação	<i>Slices</i>	13696	2158	15.75%	1880	13.72%
	<i>Flip Flops</i>	27392	2597	9.481%	2043	7.458%
	<i>Luts</i>	27392	2664	9.725%	2365	8.633%
	<i>IOBs</i>	556	79	14.21%	103	18.52%
	BRAMs	136	29	21.32%	1	0.731%
Circuito de predição	<i>Slices</i>	13696	3137	22.90%	3833	27.98%
	<i>Flip Flops</i>	27392	3770	13.76%	3934	14.36%
	<i>Luts</i>	27392	3594	13.12%	4485	16.37%
	<i>IOBs</i>	556	84	15.10%	135	24.28%
	BRAMs	136	46	33.82%	2	1.471%

No geral, pode-se notar que os circuitos desenvolvidos a partir dos FA ocupam uma área menor no *chip* do que os circuitos baseados nos MAO. Isso se deve, em grande parte, à necessidade de uma área de memória maior no *chip* para armazenar os neurônios do MAO do que a área necessária para armazenar os coeficientes do FA.

5.3.2 Desempenho dos Circuitos

Os resultados relativos ao desempenho dos circuitos de detecção, classificação e predição desenvolvidos com base nos MAO e FA, obtidos após a sua síntese, incluindo a frequência máxima de operação e os atrasos máximo de chegada e mínimo de saída do pulso de relógio (*clock*) nos registradores são apresentados na Tab. 8.

Pode-se observar na Tab. 8 que a frequência de operação máxima (F_{max}) dos circuitos de detecção, classificação e predição de falhas, quer sejam eles baseados nos MAO ou FA, ultrapassou o valor de 115.2MHz, que é a frequência de transmissão de dados da interface serial utilizada. Assim, é possível operar os circuitos na mesma frequência de comunicação entre a placa e o computador. Nota-se, também, que os tempos máximos e mínimos de saída e chegada dos dados nos registradores não ultrapassou o período mínimo de operação do circuito em nenhum dos casos.

5.3.3 Tempo de Processamento dos Circuitos

Outro requisito importante, relacionado ao embarque dos circuitos, é o desempenho dos circuitos embarcados, medido aqui pelo tempo de processamento das amostras.

Tabela 8: Desempenho dos circuitos obtido após a síntese.

Circuitos	Parâmetros	Mapa auto-organizável	Filtro adaptativo
		Valor	Valor
Circuito de detecção	Frequência de operação máxima	168.5MHz	168.5MHz
	Tempo mínimo de chegada do sinal de entrada do registrador antes do sinal de <i>clock</i>	2.618ns	2.615ns
	Tempo máximo de envio do sinal de saída do registrador necessário depois do sinal de <i>clock</i>	3.293ns	3.293ns
Circuito de classificação	Frequência de operação máxima	152.5MHz	168.5MHz
	Tempo mínimo de chegada do sinal de entrada do registrador antes do sinal de <i>clock</i>	2.520ns	2.615ns
	Tempo máximo de envio do sinal de saída do registrador necessário depois do sinal de <i>clock</i>	3.636ns	3.293ns
Circuito de predição	Frequência de operação máxima	133.2MHz	168.5MHz
	Tempo mínimo de chegada do sinal de entrada do registrador antes do sinal de <i>clock</i>	1.837ns	2.391ns
	Tempo máximo de envio do sinal de saída do registrador necessário depois do sinal de <i>clock</i>	3.293ns	3.293 ns

A válvula leva em torno de 100s para realizar um movimento completo de abertura ou fechamento. As amostras, durante o movimento de abertura ou fechamento, de torque e posição são tomadas a cada 5s; isto é, o movimento completo é dividido em 20 etapas, onde cada etapa tem a duração de 5s, aproximadamente. Portanto, nesta aplicação, pode-se dizer que 5s é o prazo máximo para a realização de todos os cálculos.

A comparação do tempo de processamento entre os circuitos de detecção, classificação e predição de falhas obtidos com os MAO e FA é apresentada na Tab. 9.

Tabela 9: Tempo de processamento dos circuitos obtido após roteamento.

Circuitos	Ferramenta	Tempo [μ s]
Circuito de detecção	MAO	8.750
	FA	2.280
Circuito de classificação	MAO	19.01
	FA	2.280
Circuito de predição	MAO	18.34
	FA	7.020

Destaca-se que tempo de processamento dos circuitos, visto na Tab. 9, foi alcançando considerando o cálculo de uma amostra somente. O tempo de processamento das amostras nunca ultrapassou o limite máximo permitido, que é de 5s. Nota-se também, que o tempo de processamento dos circuitos obtidos a partir dos filtros adaptativos é mais rápido do que os circuitos baseados nos MAO.

5.4 CONCLUSÕES DO CAPÍTULO

Neste capítulo, a etapa de embarque do sistema de manutenção inteligente foi descrita. O embarque dos circuitos de detecção, classificação e predição de falhas desenvolvidos a partir dos mapas auto-organizáveis e filtros adaptativos, incluindo uma descrição dos processos, das arquiteturas e da máquinas de estados, programados em VHDL, foi apresentado.

A idéia principal deste capítulo é determinar (com base na área ocupada no *chip*, no tempo de processamento e no desempenho dos circuitos), qual é a ferramenta (MAO ou FA) mais adequada para o embarque.

A seguir, uma descrição da plataforma de embarque, uma placa de circuito impresso, fabricada pela empresa Digilent modelo XUP Virtex-2 PRO, incluindo a especificação dos mapas auto-organizáveis e filtros adaptativos para embarque, foi realizada.

Por fim, os resultados de embarque, em termos de área ocupada dentro do chip, tempo de processamento dos circuitos e desempenho, também foram exibidos.

No geral, os circuitos obtidos com os filtros adaptativos apresentaram um menor consumo de área do *chip* do que os circuitos desenvolvidos a partir dos mapas auto-organizáveis. Os resultados de desempenho foram satisfatórios e aproximadamente iguais para os mapas auto-organizáveis e filtros adaptativos. Com relação ao tempo de processamento dos dados dos circuitos, os filtros adaptativos apresentaram resultados melhores do que os mapas auto-organizáveis.

Os resultados de embarque alcançados (em termos da área ocupada no *chip*, do tempo de processamento e do desempenho dos circuitos), quer sejam com os MAO ou FA, foram satisfatórios e ressaltam a possibilidade da real aplicabilidade do sistema embarcado na manutenção destes atuadores elétricos.

6 CONCLUSÃO

Certos tipos de equipamentos, tais como robôs ou válvulas, por exemplo, ao falharem podem provocar sérios danos materiais e colocar em risco os usuários ou funcionários presentes no ambiente de trabalho.

Com a crescente automatização de processos industriais, mecanismos e técnicas que garantam maior segurança e confiabilidade aos equipamentos estão cada vez mais sendo exigidos. É crescente a busca pela minimização das perdas econômicas e de paradas nos processos de produção. Além disso, entre outras exigências, deseja-se também que os equipamentos tenham um bom desempenho e sejam livres de falhas.

Neste contexto, as estratégias de manutenção tradicionais (corretiva e preventiva) não são as mais indicadas, sendo necessário fazer uso de estratégias mais modernas como a preditiva e proativa para aumentar a produtividade e a disponibilidade de sistemas, por exemplo.

As grandes transformações tecnológicas do setor eletrônico, em especial os sistemas embarcados, permitem realizar um diagnóstico muito maior das origens das falhas e dos seus efeitos. Estes avanços científicos têm influenciando diretamente as atividades de manutenção, em especial a manutenção proativa (YAN; LEE., 2005; DJURDJANOVIC; LEE; NI, 2003).

Sob essas circunstâncias, a manutenção proativa, que é baseada na medição e avaliação do estado atual dos sistemas, surge como a estratégia mais apropriada e eficiente para se alcançar um bom desempenho sem paradas, através de uma redução significativa das falhas.

Sistemas de detecção, diagnóstico e previsão de falhas que possuam ferramentas de análise de dados e de apoio à tomada de decisão, voltadas para a manutenção proativa, estão se tornando cada vez mais frequentes e importantes.

O desenvolvimento das técnicas de processamento de sinais, inteligência artificial e filtragem adaptativa também proporcionaram novas abordagens aos sistemas de controle e melhorou as atividades de manutenção proativa, promovendo uma maior confiabilidade e disponibilidade de equipamentos e sistemas (SINGH; KAZAZ, 2003; QUISPE, 2005).

As ferramentas de processamento de sinais, tais como a transformada *wavelet packet*, vêm sendo utilizadas em conjunto com técnicas de IA (os mapas auto-organizáveis, em especial) e técnicas de filtragem adaptativa (os filtros adaptativos, em destaque) em muitas aplicações de detecção, classificação, ou mesmo previsão de falhas (NÁCUL et al., 2002).

Os sistemas de manutenção que fazem uso dessas ferramentas (estatísticas, de processamento de sinais, de inteligência artificial ou filtragem adaptativa, por exemplo) são conhecidos como sistemas de manutenção inteligentes.

Assim, este trabalho, buscou apresentar as etapas necessárias para o desenvolvimento de um sistema de manutenção inteligente embarcado voltado para a detecção, classificação e previsão de falhas em atuadores elétricos. Este sistema de manutenção inteligente foi desenvolvido através do uso da transformada *wavelet packet* e dos mapas auto-organizáveis, ou filtros adaptativos, embarcados em uma placa de circuito impresso (XUP Virtex-2 PRO).

Na maioria das publicações encontradas na literatura a respeito dos mapas auto-organizáveis e dos filtros adaptativos, podem ser achadas inúmeras aplicações para estas ferramentas (tais como a detecção de falhas em circuitos digitais de rádio frequência mistos, a classificação de falhas em enrolamentos de mancais, a detecção de falhas em barras de rotores ou a previsão da vida útil de enrolamentos, por exemplo) (HUANG et al., 2005; TAN; HUO, 2005; QIU et al., 2006; BEZERRA, 2004).

Entretanto, ainda há pouca pesquisa científica centrada na investigação de como se utilizar os mapas auto-organizáveis e filtros adaptativos para detectar, classificar e prever falhas em atuadores elétricos e válvulas, em especial.

Iniciamente, foi apresentada uma visão geral, os principais objetivos, a metodologia utilizada e as principais etapas necessárias para o desenvolvimento deste sistema de manutenção inteligente embarcado.

A seguir, foi exibida uma descrição do estudo de caso, um sistema formado por um atuador elétrico (composto por uma máquina assíncrona e um conjunto de engrenagens), uma válvula gaveta e tubulação (KUNDUR, 1994; CAMPOS; TEIXEIRA, 2006; HENRIOT, 1967). Este atuador elétrico é usado, por exemplo, para controlar o fluxo de petróleo nas refinarias da empresa Petrobrás S.A.

Este trabalho também apresentou o desenvolvimento de um modelo matemático composto por um conjunto de equações diferenciais e algébricas não lineares que expressa o comportamento do sistema formado pelo atuador, válvula e tubulação em diversos tipos de situações (normais e de falhas).

Uma previsão das falhas e um diagnóstico dos possíveis motivos que provocaram a ocorrência de uma falha nestes equipamentos são particularmente interessantes para a manutenção, visto que isto pode representar uma maior segurança, economia e determinação do grau de responsabilidade dos envolvidos nas tarefas de manutenção.

Assim, foram realizadas uma série de simulações de injeção de falhas que ocorrem no conjunto, tais como o desgaste natural da mola ou um aumento do atrito da haste da válvula durante o movimento de abertura ou fechamento. Para realizar essas simulações fez-se uso do modelo matemático do atuador.

A partir desses dados de simulação, foram extraídos os respectivos valores de energia espectral dos sinais de torque (coletado a partir de uma célula de carga que mede o contra torque exercido pelo atuador) e posição (percentual de abertura e fechamento da válvula gaveta coletada de potenciômetro).

Em seguida, uma análise das principais ferramentas de processamento de sinais (transformada de Fourier, transformada de Fourier com janelamento e transformada *wavelet packet*) através do método da regressão logística foi apresentado.

O método da regressão logística foi usado para se determinar qual das ferramentas de processamento de sinais é a mais adequada para o estudo de caso aqui analisado, o atuador elétrico.

A partir dos gráficos do valor de confiança, obtidos com o método da regressão logística, observou-se que a transformada *wavelet packet* apresentou os melhores resultados, identificando corretamente o comportamento normal, de degradação e falha, para três tipos de falhas. Assim, constatou-se que a ferramenta de processamento de sinais mais recomendada para este estudo de caso é a transformada *wavelet packet*.

Após a definição da ferramenta de processamento de sinais mais adequada para este estudo de caso (a transformada *wavelet packet*, neste caso), os dados de energia e de sua variância (dados que representam o comportamento normal, de degradação e de falha) foram então repassados aos mapas auto-organizáveis e filtros adaptativos para a realização das etapas de detecção, classificação, predição de falhas e determinação do tempo de vida útil do atuador elétrico.

Além destas contribuições e resultados, também foi apresentado o desenvolvimento dos circuitos (arquiteturas, máquinas de estados e processos) de detecção, classificação e predição de falhas (através dos FA e MAO), os resultados de embarque destas ferramentas em VHDL e uma descrição da plataforma de embarque.

A idéia principal em se analisar e comparar (qualitativamente e quantitativamente) os resultados de simulação e embarque dos filtros adaptativos e mapas auto-organizáveis é determinar qual destas ferramentas é a mais adequada ao embarque no atuador elétrico.

Com relação aos resultados obtidos, destaca-se que foi possível efetuar uma correta detecção, classificação e predição de falhas quer seja através dos mapas auto-organizáveis ou dos filtros adaptativos.

Para a realização das tarefas de detecção e predição de falhas, ambas as ferramentas, MAO e FA, mostraram resultados satisfatórios e aproximadamente iguais em termos qualitativos. Com relação a tarefa de classificação de falhas, pode-se observar que os MAO apresentaram resultados ligeiramente superiores aos alcançados com os FA. Isto se deve às características específicas dos mapas auto-organizáveis que tornam possíveis a obtenção de melhores resultados de identificação de padrões.

A determinação do tempo de vida útil, por meio dos mapas auto-organizáveis e filtros adaptativos, através da extrapolação da curva característica de previsão do tempo de vida útil do atuador, foi abordada a seguir.

Considerando o critério adotado neste trabalho (a função racional que apresentar o menor tempo de vida útil para o atuador é a função escolhida) a função racional R_2 foi a melhor escolha para os mapas auto-organizáveis. Para os filtros adaptativos, a função racional R_3 foi a função que apresentou os melhores resultados de predição.

Por fim, na etapa de embarque dos circuitos, os filtros adaptativos apresentaram um menor consumo de área do *chip* do que os circuitos desenvolvidos com base nos mapas auto-organizáveis. Os resultados de desempenho foram satisfatórios e aproximadamente iguais para os mapas auto-organizáveis e filtros adaptativos. O tempo de processamento dos dados dos circuitos dos FA foi ligeiramente menor do que os circuitos dos MAO.

Os resultados alcançados até o momento foram satisfatórios e ressaltaram a aplicabilidade do sistema embarcado na manutenção de atuadores elétricos. Com o embarque desse sistema no atuador (incluindo todas as ferramentas matemáticas) evitam-se as falhas, reduzem-se os custos de manutenção e aumentam-se os índices de confiabilidade e disponibilidade, por exemplo.

Durante o desenvolvimento deste trabalho, surgiram inúmeras possibilidades de ampliar o estudo aqui apresentado. Assim, como exemplos de possíveis trabalhos a serem desenvolvidos no futuro para se ampliar as análises e resultados, destacam-se:

1. Simular e treinar os algoritmos dos filtros adaptativos e mapas auto-organizáveis com outros tipos de falhas e sinais (tais como sinais de ruído, vibração ou corrente elétrica);
2. Validar o modelo do atuador elétrico, válvula e tubulação com dados reais de operação;
3. Integrar todos os circuitos dos mapas auto-organizáveis e filtros adaptativos em um único circuito, alterando a máquina de estados e inserindo novas funcionalidades, blocos e processos, caso necessário. Enfim; finalizar o embarque de todas as ferramentas, inclusive as de processamento de sinais e estatísticas;
4. Treinar e testar os filtros adaptativos e mapas auto-organizáveis com dados reais do atuador elétrico, coletados a partir de ensaios realizados em bancada de experimentos;
5. Fazer uma análise e comparação dos resultados obtidos com os MAO fazendo uso de outras funções de ativação, tais como as funções do tipo sigmoideal ou multi-quadrática;
6. Estudar e examinar outros tipos de redes neurais (tais como as redes de base radial ou *perceptron* de múltiplas camadas), comparando os resultados de detecção, classificação e predição de falhas alcançados com essas redes com os resultados obtidos a partir dos mapas auto-organizáveis;
7. Examinar e comparar os resultados de detecção, classificação e predição, adquiridos com a TWP, usando outras funções *wavelets*, tais como as *wavelets* de Haar ou as *wavelets symlet*, por exemplo;
8. Analisar e comparar os resultados de simulação e embarque obtidos com os MAO fazendo uso de outras métricas para obter a distância entre vetores, além da distância Euclidiana, tais como a distância de Hamming ou a distância de Manhattan (YAMAKAWA; HORIO; HIRATSUKA, 2002; PINO et al., 1999);
9. Pesquisar e verificar os resultados alcançados com os filtros adaptativos através dos algoritmos dos MQM com os resultados alcançados com outros tipos de algoritmos, como por exemplo, o MQM normalizado ou o MQR;

10. Verificar os resultados de detecção, classificação e predição obtidos com os filtros adaptativos utilizando outras ferramentas estatísticas, tais como a média ou a covariância, além da variância;
11. Analisar e comparar os mapas auto-organizáveis e filtros adaptativos com outras ferramentas de inteligência artificial ou filtragem adaptativa, tais como os filtros de Kalman, os algoritmos genéticos ou a lógica nebulosa;
12. Aplicar e verificar a proposta de desenvolvimento de sistemas de manutenção inteligentes em outros estudos de caso, como por exemplo moto-bombas, rolamentos e robôs manipuladores.

Em suma, as possibilidades de aplicação e ampliação do estudo e pesquisa científica relacionada ao desenvolvimento de sistemas de manutenção inteligentes embarcados são muitas e promissoras e não se resumem somente aos aspectos e tópicos apresentados neste trabalho.

7 ANEXO A: PROCESSAMENTO DE SINAIS

Os processos industriais, como exemplo, geralmente produzem saídas observáveis que podem ser caracterizados como sinais. Os sinais podem ser de natureza discreta (sinais de acesso à memória de um computador, por exemplo) ou contínua (medidas de temperatura, pressão, volume, etc.). Os sinais podem ser estacionários (suas propriedades estatísticas não variam ao longo do tempo) ou não. Ainda, os sinais podem ser oriundos somente de uma única fonte ou corrompidos por outra fonte de sinal (ruído), por distorção na transmissão, reverberação, entre outros.

O processamento de sinais consiste na análise ou modificação de sinais, de forma analógica ou digital, para extrair informações dos mesmos e torná-los mais apropriados para alguma aplicação específica como, por exemplo, a determinação ou predição do comportamento de sistemas.

Nas últimas décadas, o rápido desenvolvimento da capacidade computacional de computadores e sistemas inteligentes tem permitindo o avanço dos métodos de processamento de sinais, e de reconhecimento de padrões, e ampliado o seu uso em diversas aplicações.

As técnicas de processamento de sinais estão presentes em várias áreas, incluindo entretenimento, comunicações, exploração espacial, entre outras. Algoritmos sofisticados de processamento de sinal têm sido implantados em muitos sistemas, quer seja para uso militar, industrial ou educacional.

Os métodos baseados em transformadas são técnicas amplamente utilizadas no processamento de sinais, entre as transformadas mais difundidas estão as transformadas de Fourier, Fourier com janelamento, *wavelet* e *wavelet packet*.

Neste capítulo, as principais técnicas de processamento de sinais são apresentadas na seção 7.1. A seguir, na seção 7.2, a transformada, discreta e contínua, de Fourier é mostrada. Após, na seção 7.3, a transformada de Fourier com janelamento é abordada. Na seção 7.4 a transformada *wavelet*, contínua e discreta, é discutida. A transformada *wavelet packet* é analisada na seção 7.5. Um quadro comparativo das principais transformadas é visto na seção 7.6. A energia de sinais e discretos e contínuos é tratada na seção 7.7. Por fim, as aplicações de processamento de sinais e as conclusões são apresentadas nas seções 7.8 e 7.9.

7.1 TÉCNICAS DE PROCESSAMENTO DE SINAIS

As diversas técnicas de processamento de sinais podem ser de muita utilidade no controle e análise de sistemas físicos de interesse dos mais diversos pesquisadores, não só engenheiros, como médicos, químicos e físicos. Atualmente, estas técnicas têm atraído a atenção de profissionais de diversas áreas como a militar, comercial e industrial. Além disso, é possível encontrar uma infinidade de aplicações para as técnicas de processamento de sinais: compressão e aquisição de dados, diagnóstico de imagens e prospecção mineral, por exemplo.

Dentre as principais técnicas utilizadas no processamento de sinais pode-se citar:

- Transformada de Fourier;
- Transformada de Fourier com janelamento;
- Transformada *wavelet*;
- Transformada *wavelet packet*.

Através do uso destas diferentes técnicas de processamento de sinais é possível obter informações de diagnóstico vitais, a partir de sinais de vibração, por exemplo, que permitem prever a ocorrência de falhas (SINGH; KAZZAZ, 2003).

Há inúmeras publicações abordando o uso das técnicas de processamento de sinais. Uma aplicação da transformada *wavelet packet* para a classificação automática de vozes patológicas ou normais foi apresentada por (PARRAGA, 2002), por exemplo.

Em (QUISPE, 2005), foram estudados e aplicados vários métodos para reconhecimento de padrões e processamento de sinais (Fourier e *wavelet*), utilizando dados obtidos a partir de diferentes montagens experimentais de um Nariz Eletrônico, onde os sinais gerados por um conjunto de sensores foram analisados. Quispe obteve, neste trabalho, um conjunto de padrões que lhe permitiram identificar substâncias químicas específicas.

Um outro exemplo do uso de técnicas de processamento de sinais foi apresentado por (DJURDJANOVIC; LEE; NI, 2003). Segundo os autores, por meio das transformadas de Fourier e *wavelet* é possível realizar uma detecção e quantificação da degradação do desempenho em algumas máquinas (prensas, portas eletrônicas) através da descrição quantitativa da correspondente mudança nos sinais dos sensores instalados nestas máquinas.

7.2 TRANSFORMADA DE FOURIER

Em 1807, O Barão Jean Baptiste Joseph Fourier (1768-1830), afirmou que qualquer forma de onda pode ser representada como uma soma infinita de ondas senoidais de diferentes frequências. Na época existiam sérias dúvidas sobre a exatidão dos argumentos de Fourier por isso seu trabalho só foi publicado 15 anos depois.

Hoje, a análise de Fourier está presente em muitas áreas de pesquisa e em muitas disciplinas científicas: Física, Teoria das Probabilidades, Estatística, Criptografia, Acústica, Óptica, Geometria e outras.

O método de análise de frequências, baseado no teorema de Fourier, estabelece que qualquer função periódica possa ser decomposta por uma série de ondas senoidais puras com frequências distintas e múltiplas da frequência fundamental. Estas componentes constituem o espectro de frequência (BRIGHAM, 1974; MARÇAL; SUSIN, 2005).

A transformada de Fourier se constitui então na decomposição de um sinal $x(t)$ num somatório de funções periódicas que se estendem por todo o domínio temporal, revelando o espectro de frequências que compõe o sinal (LATHI, 1998; OPPENHEIN; GEHAFFER, 1989).

7.2.1 Tipos de Análises de Fourier

As duas principais classes de análises de Fourier são a série de Fourier e a transformada de Fourier. A série de Fourier é usada na análise de sinais periódicos enquanto a transformada de Fourier é tipicamente usada para sinais não periódicos.

Ainda, a análise de Fourier pode ser dividida, de acordo com certas características do sinal no domínio tempo, em quatro tipos como visto na Tab. 10.

Tabela 10: Diferentes tipos de análise de Fourier.

Propriedade	Sinal	
	Periódico	Não-periódico
Contínua	série de Fourier	transformada de Fourier
Discreta	série de Fourier de tempo discreto	transformada de Fourier de tempo discreto

Estar-se-á particularmente interessado na análise de sinais não-periódicos aleatórios discretos ou contínuos (tais como os sinais de tensão e corrente). Assim, a transformada de Fourier (TF) contínua e de tempo discreto serão analisadas nesta sessão. Ainda, as versões contínua e discreta das transformadas de Fourier com janelamento, transformada *wavelet* e transformada *wavelet packet* também serão analisadas nas sessões a seguir.

7.2.2 Transformada Contínua de Fourier

Geralmente, a denominação “transformada de Fourier” refere-se à TF para funções contínuas. Através da TF pode-se representar qualquer função integrável $x(t)$ como a soma de exponenciais complexas com frequência angular ω ¹ e amplitude complexa $X(\omega)$.

Matematicamente, a transformada direta de Fourier (\mathcal{F}) e a transformada inversa de Fourier, (\mathcal{F}^{-1}) são expressas como (LATHI, 1998; BRIGHAM; MORROW, 1967):

$$X(\omega) = \mathcal{F}\{x(t)\} \text{ e } x(t) = \mathcal{F}^{-1}\{X(\omega)\} \quad (81)$$

ou:

$$x(t) \xleftrightarrow{\mathcal{F}} X(\omega) \quad (82)$$

onde:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (83)$$

e:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega. \quad (84)$$

¹ $\omega = 2\pi f$.

7.2.3 Transformada de Fourier de Tempo Discreto

Para uso em computadores, seja para aplicações científicas ou em processamento digital de sinais, é preciso ter valores discretos de $x(t)$. Para isso, existe a versão da transformada de Fourier para tempo discreto.

Considere um sinal $x(t)$ discretizado ao longo do tempo:

$$x(t) \rightarrow x[n] \quad (85)$$

onde: n é o índice da amostragem.

A relação matemática entre o sinal $x[n]$ e a sua TF de tempo discreto (\mathcal{T}), dada por $X[\omega]$, é expressa como:

$$x[n] \xleftrightarrow{\mathcal{T}} X[\omega] \quad (86)$$

Assim, amostrando o sinal $x(t)$ em intervalos uniformes, obtém-se a versão da transformada de Fourier de tempo discreto:

$$\mathcal{T}\{x[n]\} = X[\omega] = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \quad (87)$$

e a sua inversa (\mathcal{T}^{-1}):

$$\mathcal{T}^{-1}\{X[\omega]\} = x[n] = \frac{1}{2\pi} \int_{2\pi} X[\omega] e^{j\omega n} d\omega \quad (88)$$

com: $n \in \mathbb{Z}$ (para se assegurar de que as amostras dos dados ocorram somente em intervalos discretos) e $-\infty \leq n \leq \infty$.

7.2.4 Transformada Discreta de Fourier

Para certos tipos de sinais, um sinal exponencial, por exemplo ², a periodicidade de $X[\omega]$ sugere que não há a necessidade de se avaliar a frequência ω de $-\infty$ a ∞ . Ainda, a natureza contínua de um sinal $X[\omega]$ é uma característica indesejável. Caso a capacidade computacional disponível (memória, velocidade do processador, entre outros) seja baixa, pode ser inviável avaliar todo o espectro de frequência.

Para resolver essa restrição, basta usar o mesmo método usado para analisar um sinal analógico com um processador digital. Matematicamente, basta derivar uma amostra equivalente da transformada de Fourier de tempo discreto. Esta técnica é conhecida como transformada discreta de Fourier (TDF).

Para obter a TDF, a função $X[\omega]$ é amostrada somente durante um período igual 2π . Durante este período, N amostras equidistantes são amostradas. Deste modo, a resolução de frequência é igual a $2\pi/N$.

² $x[n] = \alpha^n u[n] \leftrightarrow 1/(1 - (\alpha e^{-j\omega}))$

Assim, substituindo ω por $2\pi/N$ na Eq. (87) e usando a variável discreta k , obtém-se a expressão da TDF:

$$\mathcal{G}\{x[n]\} = X\left[\frac{2\pi}{N}k\right] = \sum_{n=-\infty}^{\infty} x[n]e^{-j\frac{2\pi}{N}nk} \quad (89)$$

Desde que $X[\omega]$ seja periódica, não é necessário avaliar k de $-\infty$ até ∞ . A TDF de $x[n]$ de duração finita (com N amostras) pode ser expressa como:

$$\mathcal{G}\{x[n]\} = X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}nk} \quad (90)$$

e a sua inversa (\mathcal{G}^{-1}):

$$\mathcal{G}^{-1}\{X[k]\} = x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k]e^{j\frac{2\pi}{N}nk} \quad (91)$$

onde: N é o número de amostras e k e $n \in \mathbb{N}$ são índices de amostragem. Assim, $X(\omega)$ é computada em intervalos regulares de $2\pi/N$ rad (BRIGHAM, 1974; LATHI, 1998).

7.2.5 Vantagens da Transformada de Fourier

Em processamento de sinais, a TF é utilizada para isolar as componentes de frequência de um sinal. Para sistemas de monitoramento a TF é uma ferramenta útil para a extração das características de um sinal e remoção de ruídos, por exemplo.

Toda vez que uma peça ou mecanismo tem suas características mecânicas alteradas seja por desgaste ou trinca, por exemplo, seu espectro de frequências de vibração é modificado.

Folgas, defeitos ou desalinhamentos de rolamentos ou mancais de máquinas rotativas refletem-se na alteração de frequências ou no surgimento de novas frequências como destacou (MASCARENHAS, 2006). Deste modo, através da TF, pode-se observar uma alteração mecânica, um defeito, que poderá possivelmente provocar uma falha.

Na Fig. 80 vê-se, como exemplo do uso da TF, obtido a partir do MatLab, o gráfico de um sinal ³ $x(t)$ e da sua estimativa da energia do sinal obtida através da TF. Cabe destacar aqui que: ambas as frequências que aparecem na Fig. 80 possuem o mesmo conteúdo espectral de potência e energia. Contudo, na segunda frequência, o conteúdo está mais concentrado, em uma pequena faixa, do que na primeira frequência que aparece mais disperso. Isto ocorre devido aos distintos períodos das duas senóides presentes no sinal $x(t)$.

7.2.6 Desvantagens da Transformada de Fourier

Após o sinal ter sido transformado, a informação no domínio tempo é agora representada no domínio frequência (espectro). Isto torna possível identificar as componentes de frequência. Contudo, a transformada de Fourier não preserva a informação temporal.

³ $x(t) = \text{sen}(2\pi 50t) + \text{sen}(2\pi 250t) + 2\text{randn}(\text{size}(t))$; onde: $\text{randn}(\cdot)$ e $\text{size}(\cdot)$ são funções do Matlab onde se obtém o ruído branco e a dimensão do vetor de instantes de tempo t , respectivamente.

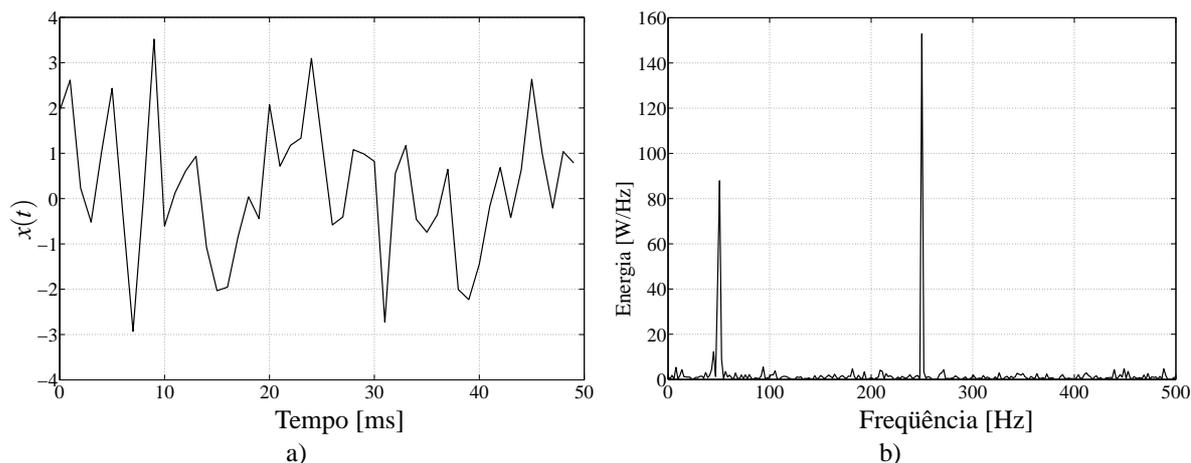


Figura 80: Exemplo do uso da TF: a) sinal e b) conteúdo de frequência.

Uma análise de frequências por Fourier, mesmo fornecendo uma informação sobre a composição espectral do sinal, não mostra os instantes onde estas frequências ocorrem.

Portanto, a TF é ideal para uma análise de sinais cujo espectro não varie no tempo ou sinais estacionários como as ondas sonoras e luminosas; porém, não é eficaz para o estudo de fenômenos transitórios como os sinais biológicos, pois os mesmos, muitas vezes, possuem características não estacionárias (SCHUCK, 1998; SILVA, 2005).

Para revelar a ocorrência de um evento num certo instante de tempo e as mudanças nas componentes das frequências ao longo do tempo é necessário realizar uma análise temporal e espectral em conjunto.

Uma análise temporal e espectral é uma maneira de se obter ao mesmo tempo informações sobre o conteúdo espectral e temporal de um sinal. As formas de se realizar uma análise temporal e espectral (ou em outras palavras, obter uma representação tempo-frequência) mais conhecidas são as transformada de Fourier com janelamento, a transformada *wavelet* e a transformada *wavelet packet*, vistas a seguir.

7.3 TRANSFORMADA DE FOURIER COM JANELAMENTO

Em 1946, Dennis Gabor adaptou a TF com uma técnica chamada de janelamento do sinal. A adaptação de Gabor, conhecida como transformada de Fourier com janelamento (TFJ), coloca o sinal em uma função de duas dimensões: tempo e frequência (SARKAR; SU, 1998).

A transformada de Fourier com janelamento, semelhante à transformada de Fourier, possui uma versão contínua e discreta, que serão vistas a seguir.

7.3.1 Transformada Contínua de Fourier com Janelamento

A TFJ é uma extensão da definição da TF. Na TF, a análise é realizada ao longo de todo sinal, visto que as frequências são assumidas fixas para todo o período avaliado. Já na TFJ, o sinal é dividido em regiões e aplica-se a TF a cada uma destas regiões.

Para o caso de sinais não-estacionários, a TFJ simplesmente usa a TF para cada segmento de tempo janelado de modo apropriado. A TFJ calcula primeiro a TF dos segmentos de tempo individualmente e então combina linearmente os resultados.

A TFJ (\mathcal{S}), na sua versão contínua, é obtida a partir da seguinte equação (HLAWATSCH; BOURDREAUX-BARTELS, 1992):

$$\mathcal{S}\left\{x(t') * \phi(t' - t)\right\} = X(t, f) = \int_{-\infty}^{\infty} x(t') \theta^*(t' - t) e^{-j2\pi f t'} dt' \quad (92)$$

e a sua inversa (\mathcal{S}^{-1}):

$$\mathcal{S}^{-1}\left\{X(t, f)\right\} = x(t') \phi(t' - t) = e^{-2\pi f t} \int_{-\infty}^{\infty} X(f') \Phi^*(f' - f) e^{j2\pi f' t} df' \quad (93)$$

onde: θ é uma função de janelamento e Φ é a sua respectiva TF. Pode-se observar que a função de janelamento está deslocada de t unidades de tempo.

A função $\mathcal{S}(\cdot)$ pode ser vista como a TF do sinal $x(t')$ previamente limitada por uma função de janelamento $\phi(t')$ ao longo do tempo t . Este método assume que o sinal $x(t')$ é estacionário dentro da janela limitada $\phi(t')$. Uma função de janelamento muito utilizada é a função gaussiana.

Dado um $x(t)$ real, a TFJ geralmente é um sinal complexo. A definição do espectro da TFJ é basicamente a mesma vista para a transformada de Fourier, exceto pela presença da função de janelamento $\phi(t)$.

Pela definição da TFJ, vê-se que ela é altamente influenciada pelas características da função de janelamento. Para se obter uma boa resolução temporal, é necessário que a janela $\phi(t)$ seja estreita em torno do instante de interesse. Contudo, procedendo assim, se obtém pouca resolução espectral.

Em contrapartida, para se obter uma boa resolução espectral, é preciso que a janela seja estreita no domínio freqüência. Como consequência, a resolução temporal fica prejudicada. Assim, não é possível se ter, ao mesmo tempo, uma alta resolução espectral e temporal, pois não existe uma janela que seja ao mesmo tempo estreita no domínio do tempo e da freqüência.

7.3.2 Transformada Discreta de Fourier com Janelamento

A transformada discreta de Fourier com janelamento (\mathcal{D}) é dada pela seguinte equação:

$$\mathcal{D}\left\{x[n] * \phi[n - \beta]\right\} = X\left[\frac{2\pi}{N}k\right] = \sum_{n=-\infty}^{\infty} x[n] \phi[n - \beta] e^{-j\frac{2\pi}{N}nk} \quad (94)$$

e a sua inversa (\mathcal{D}^{-1}):

$$\mathcal{D}^{-1}\left\{X\left[\frac{2\pi}{N}k\right]\right\} = x[n] \phi[n - \beta] = \frac{1}{2\pi} \int_0^{2\pi} X\left[\frac{2\pi}{N}k\right] e^{j\frac{2\pi}{N}nk} d\omega \quad (95)$$

onde: N é o número de amostras e k, β e $n \in \mathbb{N}$ são os índices de amostragem, com k, n e $\beta = 0, 1, 2, \dots, N - 1$. Assim, a transformada discreta de Fourier com janelamento também é computada em intervalos regulares.

A magnitude da transformada discreta de Fourier com janelamento (TDFJ) representa as amplitudes das componentes espectrais nas frequências discretizadas do sinal de entrada $x[n]$.

A TDFJ mapeia o sinal em duas dimensões: tempo e frequência e apresenta um compromisso temporal e espectral do sinal analisado.

Os gráficos obtidos pela TDFJ costumam ter representada a variável tempo nas abscissas, a variável frequência nas ordenadas, e amplitude do módulo ou fase como uma escala de cinza ou cores, conhecido como espectrograma (LATHI, 1998).

A Fig. 81 apresenta um exemplo (obtido a partir do Matlab) do uso da transformada de Fourier com janelamento onde se pode observar um sinal senoidal, $x(t)$ ⁴, e o seu respectivo espectrograma.

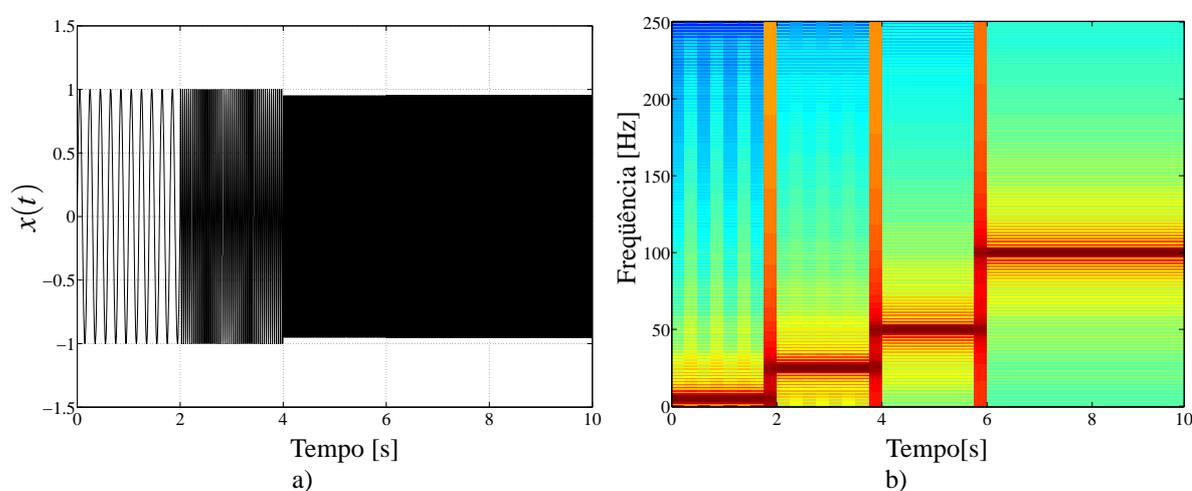


Figura 81: Exemplo do uso da TDFJ: a) sinal e b) espectrograma.

7.3.3 Utilidade da Transformada de Fourier com Janelamento

Visto que a TFJ fornece uma informação temporal e espectral, esta análise é aplicável para sinais altamente não estacionários que podem envolver diversas situações e comportamentos transitórios, como um torno mecânico, uma prensa, ou um motor (LEE et al., 2006).

7.3.4 Desvantagens da Transformada de Fourier com Janelamento

Em certas circunstâncias, quando um sinal apresenta diversas componentes de frequência em um pequeno intervalo de tempo, faz-se necessário uma maior definição temporal.

Por outro lado, quando se observa mudanças infinitesimais de frequência em um sinal, como os observados nos sistemas de controle, ou aqueles vistos na área médica, uma maior resolução da frequência é desejada.

A transformada de Fourier com janelamento apresenta um compromisso entre a resolução temporal e a resolução espectral, fornecendo a informação de quando e em qual frequência ocorre um determinado evento no sinal.

⁴ $x(t) = \text{sen}(2\pi ft)$, com $f=5\text{Hz} \forall 0 \leq t < 2\text{s}$, $25\text{Hz} \forall 2 \leq t < 4\text{s}$, $50\text{Hz} \forall 4 \leq t < 6\text{s}$ e $f=100\text{Hz} \forall t \geq 6\text{s}$.

Este compromisso existente entre a resolução em frequência e a duração temporal da janela pode ser descrito pelo princípio da incerteza, ou equação de Heisenberg aplicado à análise de sinais. A equação de Heisenberg é dada por:

$$\Delta t \Delta f \geq \frac{1}{4\pi} \quad (96)$$

onde: Δt é a duração da janela e Δf é a resolução em frequência (KAISER, 1994).

A informação sobre a frequência tem precisão limitada pelo tamanho da janela temporal. Uma vez escolhido o tamanho da janela Δt , este será o mesmo para todas as frequências. Um exemplo da divisão do sinal em células de resolução é visto na Fig. 82.

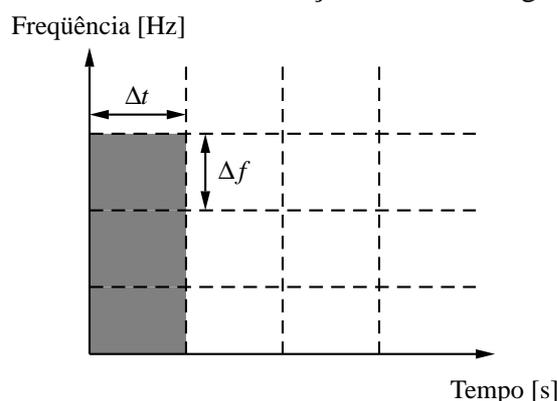


Figura 82: Representação das células de resolução Δt e Δf para a TFJ.

Além disso, a resolução temporal e a resolução espectral são dependentes do tamanho e da forma da janela utilizada $\phi[n - \beta]$. Ambas variáveis são constantes durante a análise da TFJ, logo, não é possível alternar a precisão temporal e/ou espectral da transformada. Isto torna a varredura do espectro de frequências uniforme, como visto na Fig. 83, onde f_0 é a frequência fundamental de um sinal.

Uma janela muito estreita resulta em uma boa identificação temporal de eventos transitórios, mas ao mesmo tempo tem associado um espectro de baixa resolução, o que reduz a resolução entre as frequências próximas (CROVATO, 2004).

Já com uma janela larga, se obtém uma baixa identificação temporal de eventos e uma melhor identificação de frequências adjacentes.

Assim, a análise de transitórios eletromagnéticos em alguns sistemas, por exemplo, com esta configuração de janela, usando este método, não é recomendada.

Uma solução usualmente utilizada para resolver esta questão é escrever um sinal $x(t)$ como uma soma de funções que podem sofrer dilatação tanto no tempo quanto em frequência que permite avaliar a evolução temporal das frequências de baixa e alta duração. Esta técnica é conhecida como transformada *wavelet* e será vista a seguir.

Por fim, cabe dizer que há várias formas de se programar a TDFJ. O IMS Center, por exemplo, programa a TDFJ de um modo um pouco diferente: ao invés de se utilizar os coeficientes resultantes da TDFJ como base para obter a assinatura característica do sinal, utiliza uma distribuição da energia do sinal no plano tempo-frequência.

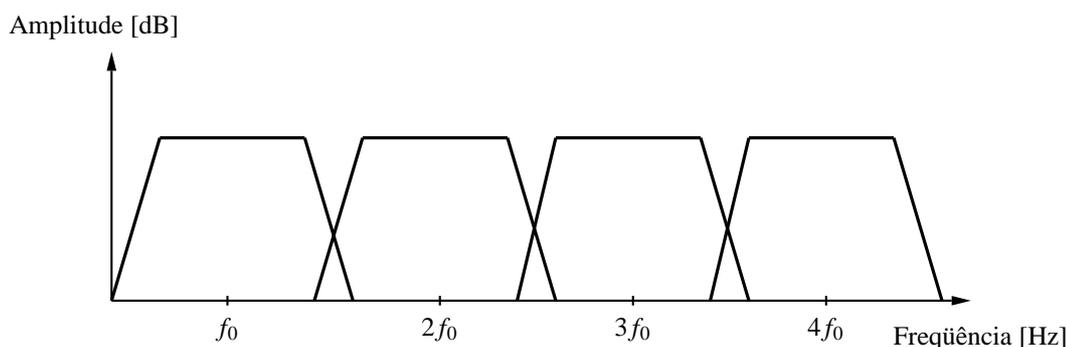


Figura 83: Varredura do espectro de freqüências através da TDFJ.

Há diversas distribuições deste tipo (Wigner-Ville e Cohen, por exemplo), cada uma com suas propriedades únicas. O IMS Center, por exemplo, opta por utilizar a classe de distribuições conhecida como Cohen. A distribuição Cohen, por sua característica de covariância no tempo e na freqüência, permite deslocar a distribuição no tempo e na freqüência de modo proporcional. Caso o sinal tenha um atraso e seja modulado, a distribuição será deslocada no tempo e na freqüência proporcionalmente ao atraso (AUGER et al., 1996; COHEN, 1995).

7.4 TRANSFORMADA WAVELET

As bases matemáticas das *wavelets* são recentes, se comparadas com as de Fourier e começaram a ser aplicadas na análise de sinais somente na década de 80 (OPPENHEIN; GEHAFFER, 1989) como lembrado em (HELENA, 2004) e (SALIM, 2006).

O termo *wavelet* foi introduzido inicialmente pelo geofísico francês Jean Morlet. Morlet analisava dados sísmicos que apresentavam uma grande variação de freqüência ao longo do tempo, quando percebeu que TF não era adequada para este caso, onde uma maior resolução temporal para eventos de alta freqüência faz-se necessária.

A análise com a transformada *wavelet* ajudou a superar a limitação de janela fixa que existe na TFJ através de uma técnica conhecida como técnica de “janelamento variável”.

A técnica de “janelamento variável” permite o uso de uma janela de tempo maior para analisar informações de baixa freqüência, de forma mais precisa, e de uma pequena janela de tempo para avaliar informações de alta freqüência (VALINS, 2005).

Assim, por meio da transformada *wavelet* (TW) é possível localizar fenômenos com duração temporal muito pequena. Desta forma, a transformada *wavelet* serve como ferramenta de análise de fenômenos transitórios, não estacionários, ou variantes no tempo (SCHUCK, 1998). Deste modo, a TW pode ser vista como uma extensão da TFJ (YAN; GAO, 2005).

Assim como a transformada de Fourier, a transformada *wavelet* pode ser encontrada em duas versões diferentes: a transformada *wavelet* contínua (TW) que faz o mapeamento de um sinal original unidimensional no domínio do tempo, para uma nova função, em um espaço bidimensional e a transformada *wavelet* discreta que realiza a decomposição de um sinal amostrando-o em diferentes níveis de resolução de freqüência.

7.4.1 Transformada *Wavelet* Contínua

A transformada *wavelet* contínua emprega um protótipo de função, $\psi(t)$, chamado *wavelet* mãe. A função *wavelet* mãe deve ter média zero e parte central oscilante; ou seja, deve decair para zero em ambos os lados de sua trajetória, como visto na seguinte equação:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (97)$$

Isto é garantido desde que a TF da função *wavelet* mãe, $\Psi(\omega)$, tenha zero na componente de frequência $\omega = 0$, ou seja:

$$|\Psi(\omega)|_{\omega=0}^2 = 0 \quad (98)$$

Além disso, a função *wavelet* mãe deve possuir norma quadrática finita, isto é:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < +\infty \quad (99)$$

A função *wavelet* mãe, na TW, é dilatada através de um parâmetro escalar α e transladada pelo parâmetro τ ⁵ do seguinte modo:

$$\psi_{\alpha,\tau}(t) = \frac{1}{\sqrt{|\alpha|}} \psi\left(\frac{t-\tau}{\alpha}\right) \quad (100)$$

A variação do escalonamento e da translação cria um conjunto de funções que será usada na representação da função a ser analisada. Deste modo, a base é composta por *wavelets* que são cópias escalonadas e transladadas da *wavelet* mãe (OPPENHEIN; GEHAFFER, 1989).

Matematicamente, a transformada *wavelet* contínua (\mathcal{W}) de um dado sinal $x(t)$ é definida como um produto interno na seguinte forma:

$$\mathcal{W}\{x, \psi\} = \frac{1}{\sqrt{|\alpha|}} \left\langle x(t), \psi_{\alpha,\tau}(t) \right\rangle = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{\infty} x(t) \cdot \psi^*\left(\frac{t-\tau}{\alpha}\right) dt \quad (101)$$

onde: α é o fator de escala⁶, τ é o fator de translação e $1/\sqrt{|\alpha|}$ é um fator par anormalização da energia através das diferentes escalas.

A inversa da TW contínua (\mathcal{W}^{-1}) é dada por:

$$\mathcal{W}^{-1}\left\{\left\langle x(t), \psi_{\alpha,\tau}(t) \right\rangle\right\} = x(t) = \frac{1}{C_{\psi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}\{x, \psi\} \frac{1}{\sqrt{|\alpha|}} \psi^*\left(\frac{t-\tau}{\alpha}\right) \frac{d\alpha d\tau}{\alpha^2} \quad (102)$$

Na Eq. (102), C_{ψ} é uma constante (definida como constante espectral) dada por:

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega \quad (103)$$

onde: $\Psi(\omega)$ é a transformada de Fourier da função *wavelet* mãe $\psi(t)$.

⁵ α e $\tau \in \mathbb{R}$, $\alpha \neq 0$.

⁶ A função *wavelet* mãe é dilatada para $\alpha > 1$ e contraída para $\alpha < 1$.

A transformada *wavelet* pode ser vista como a decomposição de uma função $x(t)$ em uma combinação linear de funções *wavelets*. Os coeficientes da transformada *wavelet* indicam o quão semelhante a função $x(t)$ é em relação a uma função de decomposição da *wavelet* mãe.

Visto que uma certa função *wavelet* mãe pode ser visualizada como uma janela, Eq. (101), verifica-se que o fator de escala e o tamanho da janela são interdependentes: um fator de escala menor implica em uma janela menor, por exemplo.

Assim, para se analisar componentes de altas frequências presentes no sinal deve-se utilizar um fator de escala pequeno; já para as componentes de baixa frequência, usa-se um fator de escala maior.

Por fim, cita-se que a inversa da transformada *wavelet* exibe como a função original $x(t)$ pode ser sintetizada pela soma das componentes (projeções) do sinal em *wavelets*. Contudo, dentre o conjunto de funções *wavelets*, estas não precisam ser necessariamente ortogonais a função original nem tampouco entre si.

7.4.2 Transformada *Wavelet* Discreta

O cálculo dos coeficientes de todas as possíveis escalas da transformada *wavelet* contínua, Eq. (101), requer um enorme esforço computacional e gera uma informação redundante. Uma alternativa a TW contínua é a utilização de parâmetros de escalonamento e translação discretos. A este tipo de análise se dá o nome de transformada *wavelet* discreta (TWD).

A transformada *wavelet* discreta (\mathcal{V}) de um sinal $x(t)$ é expressa como:

$$\mathcal{V}\{x, \Psi\} = \langle x(t), \Psi_{\kappa\beta} \rangle = \int_{-\infty}^{\infty} x(t) \cdot \Psi_{\kappa\beta}(t) dt \quad (104)$$

A inversa da transformada *wavelet* discreta (\mathcal{V}^{-1}) é dada por:

$$\mathcal{V}^{-1}\left\{\langle x(t), \Psi_{\kappa\beta} \rangle\right\} = x(t) = \sum_{\kappa=-\infty}^{\infty} \sum_{\beta=-\infty}^{\infty} \mathcal{V}\{x, \Psi\} \Psi_{\kappa\beta}(t) \quad (105)$$

A função $\Psi_{\kappa\beta}$ forma uma base de funções *wavelet*, criadas a partir da função *wavelet* mãe mediante parâmetros de escala, ou dilatação, e translação discretos.

A função $\Psi_{\kappa\beta}$ é obtida a partir de:

$$\Psi_{\kappa\beta}(t) = \frac{1}{\sqrt{\alpha^\kappa}} \Psi\left(\frac{t - \beta \alpha^\kappa \tau}{\alpha^\kappa}\right) \quad (106)$$

onde: α é a variação da dilatação, τ é o passo de translação e κ e β são constantes $\in \mathbb{Z}$.

Uma possível escolha, geralmente adotada, feita para as constantes α e τ é $\alpha = 2$ e $\tau = 1$. Deste modo, ter-se-á uma transformada *wavelet* amostrada a um valor de 2^κ :

$$\Psi_{\kappa\beta} = \frac{1}{\sqrt{2^\kappa}} \Psi\left(\frac{t - \beta 2^\kappa}{2^\kappa}\right) \quad (107)$$

Deste modo, obtém-se uma escala de dilatação como uma potência de dois e passos de translação de um passo de escala de dilatação (PARRAGA, 2002).

Pode-se observar também na Eq. (107) que o tempo permanece como uma variável contínua e os parâmetros κ e β é que fornecem a dilatação e translação, respectivamente. Isto é, as constantes κ e β é que determinam a variação da dilatação e do passo de translação, respectivamente.

A escolha das constantes κ e β também irá definir se o conjunto de funções *wavelets* será, ou não, ortogonal. Pode-se dizer que duas funções, ψ , são ortogonais entre si, caso:

$$\langle \psi_{\kappa\beta}, \psi_{\kappa k} \rangle = \begin{cases} 1 & : \beta = k \\ 0 & : \beta \neq k \end{cases} \quad (108)$$

onde:

$$\langle \psi_{\kappa\beta}, \psi_{\kappa k} \rangle = \int \psi_{\kappa\beta}^* \psi_{\kappa k} dt \quad (109)$$

Para reconstruir um sinal $x(t)$ a partir dos coeficientes da *wavelet* é necessário escolher adequadamente a função *wavelet* mãe, $\psi(t)$. Segundo Daubechies, (DAUBECHIES, 1990), a condição necessária e suficiente para reconstruir um sinal a partir dos coeficientes da *wavelet* é que a energia da transformada *wavelet* discreta, \mathcal{V} , esteja entre dois limites positivos:

$$L_1 \|x(t)\|^2 \leq \sum_{\kappa\beta} \left| \langle x(t), \psi_{\kappa\beta} \rangle \right|^2 \leq L_2 \|x(t)\|^2 \quad (110)$$

onde: $\|x(t)\|^2$ é a densidade espectral do sinal $x(t)$ e $L_1 > 0$ e $L_2 < \infty$ são os limites.

Caso a condição da Eq. (110) seja satisfeita, $\psi_{\kappa\beta}$, com $\kappa \in \mathbb{Z}$ e $\beta \in \mathbb{Z}$, é um quadro, e L_1 e L_2 são os limites deste quadro. A energia da transformada *wavelet* será proporcional à energia espectral do sinal e ter-se-á um “*tight frame*” caso os limites L_1 e L_2 sejam iguais (DAUBECHIES, 1990).

Deste modo, a reconstrução do sinal será possível a partir da seguinte equação:

$$x(t) = \frac{2}{L_1 + L_2} \sum_{\kappa} \sum_{\beta} C_{\kappa\beta} \psi_{\kappa\beta}(t) \quad (111)$$

onde: $C_{\kappa\beta} = \langle x(t), \Theta_{\kappa\beta} \rangle$.

Um modo eficiente de aplicar a transformada *wavelet* é através da utilização de filtros. Esta técnica, conhecida como análise multiresolução, foi inicialmente desenvolvida por (MALLAT, 1989) e permite a análise de sinais em múltiplas bandas de frequências.

Na transformada *wavelet* discreta, utilizando a análise multiresolução, o sinal é dividido em coeficientes de aproximação (obtidos após a passagem do sinal por um filtro passa-baixas) e detalhamento (obtidos a partir de um filtro passa-altas), como visto na Fig. 84.

A função *wavelet* mãe, $\psi(t)$, produz os coeficientes de detalhamento da decomposição *wavelet*. Para gerar os coeficientes de aproximação faz-se uso de uma função adicional conhecida como função de escalonamento, definida como ϕ , contínua e geralmente real, não satisfaz a condição de admissibilidade, isto é:

$$|\Theta(\omega)|_{\omega=0}^2 \neq 0 \quad (112)$$

onde: Θ é a TF da função de escalonamento ϕ .

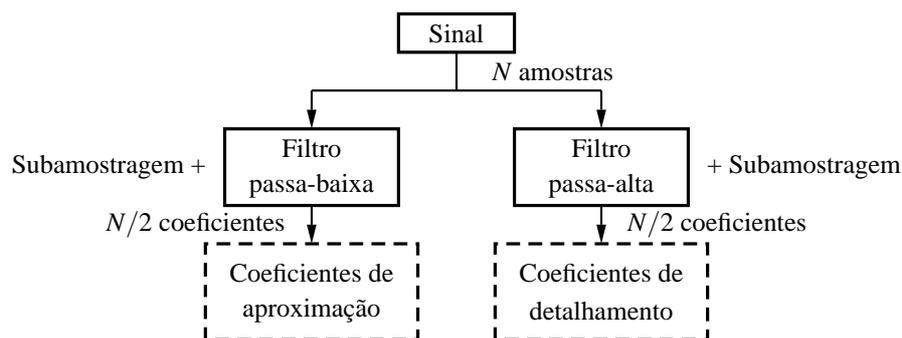


Figura 84: Decomposição da TWD usando a análise multiresolução.

Essa função de escalonamento é então dilatada por um fator igual a 2, por exemplo:

$$\phi_{\kappa\beta} = \frac{1}{\sqrt{2^\kappa}} \phi\left(\frac{t - \beta 2^\kappa}{2^\kappa}\right) \quad (113)$$

Os coeficientes de aproximação, A , de um sinal $x(t)$, para um nível de decomposição, são obtidos a partir de:

$$A_{\kappa\beta} = \langle x(t), \phi_{\kappa\beta} \rangle \quad (114)$$

onde: κ e β indicam o nível de dilatação e de translação, respectivamente.

Os filtros utilizados na transformada *wavelet* discreta multiresolução (passa-baixas e passa-altas) são conhecidos como filtros de quadratura. Esses filtros são usados para filtrar e decimar um sinal, respectivamente. O processo de decimação consiste em eliminar os dados ou pontos intercalados, diminuindo a taxa de amostragem do sinal. Ao decimar um sinal pelo fator 2, por exemplo, estar-se-á eliminando um ponto a cada dois pontos e obtendo somente a metade do sinal original, como visto na Fig. 84.

Os coeficientes terão uma dimensão igual a metade da dimensão do sinal original, neste caso. Assim, na operação de decimação, aumenta-se o tamanho do período de amostragem do sinal e diminui-se a taxa de amostragem.

Na transformada *wavelet* discreta apenas os coeficientes de aproximação (filtro passa-baixas) são novamente divididos em coeficientes de aproximação e detalhamento novamente, chamado de segundo nível de decomposição; e assim pode ser feito sucessivamente, n vezes (PARRAGA, 2002), obtendo-se a árvore de decomposição da *wavelet*.

Deste modo, a transformada *wavelet* discreta é equivalente à implementação de um banco de filtros não uniformes com uma melhor resolução em frequência em baixas frequências e melhor resolução em tempo para as altas frequências (SARKAR; SU, 1998).

Um sinal pode ser decomposto, por exemplo, em 4 grupos em uma transformada *wavelet* discreta de 3 níveis: um grupo de aproximação da informação (HHH) e 3 grupos de detalhamento da informação (HHG, HG, G), como visto na Fig. 85.

Assim, através dos filtros de quadratura é possível obter uma cobertura do espectro de frequências do sinal original.

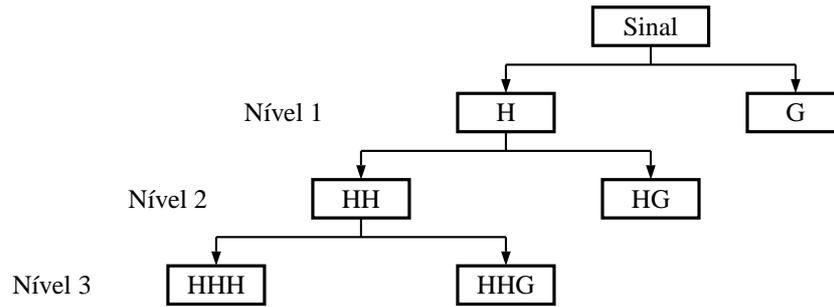


Figura 85: Exemplo de árvore de decomposição da TWD.

A função de escalonamento e as funções *wavelets* estão associadas aos filtros de quadratura, $h(\cdot)$ e $g(\cdot)$, através das seguintes equações:

$$h(\beta - 2k) = \langle \phi_{\kappa-1, \beta}, \phi_{\kappa k} \rangle = 1/\sqrt{2} \int \phi(t - \beta) \cdot \phi(t/2 - k) dt \quad (115a)$$

$$g(\beta - 2k) = \langle \phi_{\kappa-1, \beta}, \psi_{\kappa k} \rangle = 1/\sqrt{2} \int \phi(t - \beta) \cdot \psi(t/2 - k) dt \quad (115b)$$

onde: k é um índice de amostragem, com $k \in \mathbb{Z}$.

Deste modo, é possível calcular os coeficientes de um nível de resolução para outro fazendo uso dos filtros digitais e de algoritmos recursivos. As relações vistas nas Eq. (115a) e (115b) são constantes em qualquer nível de dilatação, não sendo necessário calcular estas novamente para um novo nível de dilatação para se obter, considerando k fixo, os novos coeficientes de aproximação, A_{κ} , e detalhamento, D_{κ} , segundo as seguintes expressões:

$$A_{\kappa} = \sum_{\beta} A_{\kappa-1, \beta} h(\beta - 2k) \quad (116a)$$

$$D_{\kappa} = \sum_{\beta} D_{\kappa-1, \beta} g(\beta - 2k) \quad (116b)$$

Um exemplo de uso da TWD pode ser visto na Fig. 86. Nesse exemplo, obtido a partir do Matlab, pode-se observar um sinal ⁷ e o gráfico dos seus respectivos coeficientes do terceiro nível de aproximação (HHH).

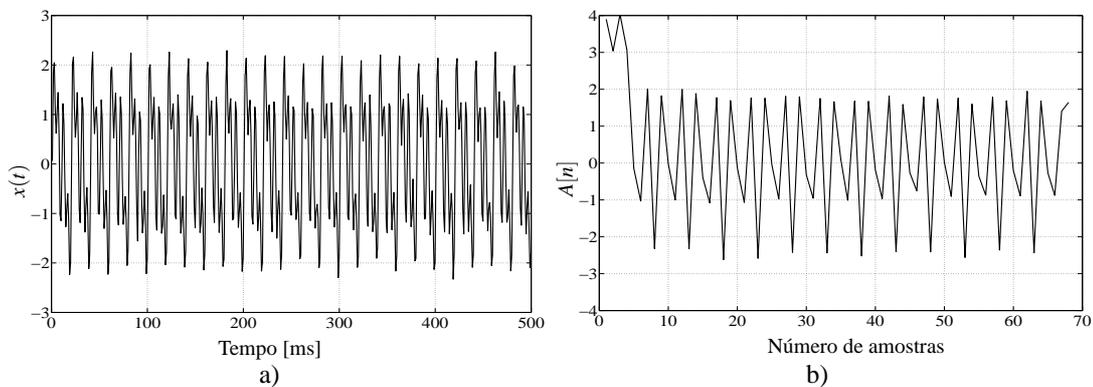


Figura 86: Exemplo do uso da TWD: a) sinal e b) coeficientes de aproximação (HHH).

⁷ $x(t) = \text{sen}(2\pi 50t) + \text{sen}(2\pi 100t) + \text{sen}(2\pi 250t) + 0.1 \text{randn}(\text{size}(t))$.

7.4.3 Benefícios da Transformada *Wavelet*

A TW, semelhante a TFJ, também apresenta um compromisso entre a resolução temporal e a resolução espectral. Na TW as resoluções temporal e espectral variam no plano do tempo e da frequência. Uma representação da divisão do sinal em células de resolução temporal e espectral para a TW pode ser vista na Fig. 87.

Visto que a TWD é equivalente à implementação de um banco de filtros, a resolução temporal é proporcional à frequência central destes filtros. Já a resolução em domínio frequência é definida como:

$$\frac{\Delta f}{f} = c \quad (117)$$

onde: c é uma constante.

Deste modo, o banco de filtros é composto por uma série de filtros passa-faixa, com uma banda de passagem constante. Assim, o banco de filtros possui uma resposta em frequência igualmente espaçada na escala logarítmica de frequência, como visto na Fig. 88.

Quando a Eq.(117) é satisfeita, verifica-se que Δf varia proporcionalmente e Δt inversamente proporcional, respectivamente, a variação da frequência central do filtro analisado.

O princípio da incerteza de Heisenberg, Eq. (96), ainda é satisfeito. Contudo, a resolução temporal aumenta nas frequências altas, enquanto a resolução do espectro de frequências é maior nas baixas frequências.

Assim, a TW serve principalmente como ferramenta de análise de sinais transitórios, não estacionários ao longo do tempo. Deste modo, pode-se encontrar inúmeros exemplos de aplicação da TW na estudo de máquinas rotativas, na análise da vibração de rotores, no diagnóstico e previsão de falhas em rolamentos, e na análise da voz, por exemplo (YAN; GAO, 2005; PRABHAKAR; MOHANTY; SEKHAR, 2002; PARRAGA, 2002).

7.4.4 Desvantagens da Transformada *Wavelet*

Devido à fraca decomposição dos detalhes das componentes de alta frequência de um sinal, geralmente torna-se difícil extrair as características dinâmicas provocadas por defeitos estruturais nos sistemas, tais como uma rachaduras ou desgastes (YAN; GAO, 2005).

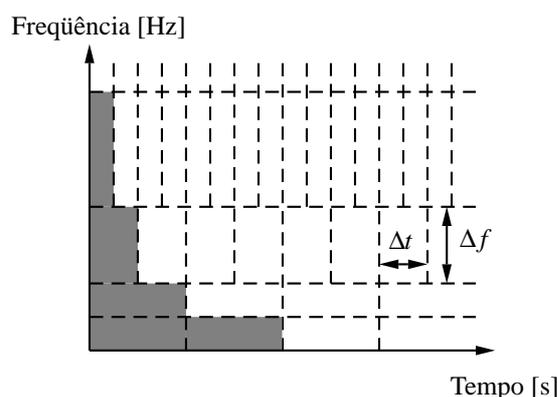


Figura 87: Representação das células de resolução Δt e Δf para TW.

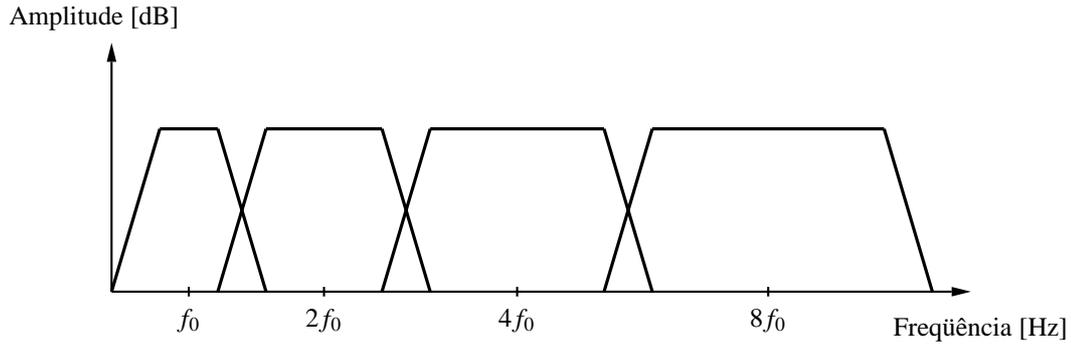


Figura 88: Varredura do espectro de frequências através da TW.

A transformada *wavelet packet*, vista a seguir, em comparação à TW, realiza não somente uma decomposição dos coeficientes de aproximação, mas também dos coeficientes de detalhamento do sinal.

7.5 TRANSFORMADA *Wavelet Packet*

A transformada *wavelet packet* (TWP) pode ser vista como uma generalização do conceito da transformada *wavelet* discreta, onde a resolução temporal e espectral pode ser escolhida conforme o sinal.

Na TWP os coeficientes de aproximação e detalhamento são obtidos de modo semelhante à TWD. Contudo, no primeiro nível de decomposição da *wavelet packet*, os coeficientes de detalhamento também são divididos em coeficientes de aproximação e detalhamento, decompondo o sinal em forma de árvore binária, chamada *wavelet packet*, como visto na Fig. 89.

Este processo, também, é feito respeitando os limites do princípio de incerteza de Heisenberg, como visto na Eq. (96).

Deste modo, o sinal pode ser representado em bandas de frequências com diferentes resoluções. Esta decomposição dos sinais em distintas bandas de frequências melhora o desempenho dos classificadores ao se extrair as características relativas de um sinal. Esta técnica também pode ser usada como uma ferramenta para se obter as características relevantes de sinais de uma ou duas dimensões.

A transformada *wavelet packet*, ao contrário das TF, TFJ e TW que apresentam uma versão contínua e discreta, possui uma versão discreta, que será vista a seguir.

7.5.1 Transformada *Wavelet Packet* Discreta

O cálculo numérico dos coeficientes da transformada *wavelet packet* (\mathcal{P}) é a seqüência do produto interno de um sinal $x(t)$ com as funções *wavelet packet*, $\psi_{\kappa\beta\eta}(t)$, como a seguir:

$$\mathcal{P}\{x, \psi\} = \langle x(t), \psi_{\kappa\beta\eta} \rangle = \triangleq v_{\kappa\eta}(\beta) = \frac{1}{\sqrt{\alpha^k}} \int_{-\infty}^{\infty} x(t) \cdot \psi_{\kappa\beta\eta} dt \quad (118)$$

onde: $\psi_{\kappa\beta\eta}$ é a função *wavelet packet* que forma uma base de funções *wavelet*, criadas a partir da função *wavelet* mãe mediante parâmetros discretos κ , β e η .

A função *wavelet packet* é obtida a partir de:

$$\Psi_{\kappa\beta\eta} = \frac{1}{\sqrt{\alpha^\kappa}} \Psi_\eta \left(\frac{t - \beta \alpha^\kappa \tau}{\alpha^\kappa} \right) \quad (119)$$

De modo semelhante ao realizado para a TWD, é possível escolher as constantes α e τ iguais 2 e 1, respectivamente. Deste modo, também, ter-se-á a *wavelet packet* amostrada a uma taxa de 2^κ :

$$\Psi_{\kappa\beta\eta} = \frac{1}{\sqrt{2^\kappa}} \Psi_\eta \left(\frac{t - \beta 2^\kappa}{2^\kappa} \right) \quad (120)$$

onde: $\eta \in \mathbb{N}$ é o índice de frequência.

Pode-se observar, através da Eq. (120), que, ao contrário da função ψ usada na TWD, que é uma função somente de duas variáveis, a função ψ utilizada na transformada *wavelet packet* é uma função de três variáveis, incluindo a variável adicional η que corresponde ao índice de frequência da *wavelet packet*.

A biblioteca de funções *wavelet packet* é construída com um algoritmo recursivo e as funções ψ são obtidas, considerando uma escala fixa e definindo $\Psi_{\kappa\beta\eta} \triangleq \Psi_\eta$ para simplificar as expressões, a partir das seguintes equações:

$$\int_{\mathbb{R}} \Psi_0(t) dt = 1 \quad (121a)$$

$$\Psi_{2\eta}(t) = \sqrt{2} \sum_{\beta} h(\beta) \cdot \Psi_\eta(2t - \beta) \quad (121b)$$

$$\Psi_{2\eta+1}(t) = \sqrt{2} \sum_{\beta} g(\beta) \cdot \Psi_\eta(2t - \beta) \quad (121c)$$

onde: $h(\cdot)$ e $g(\cdot)$ são filtros de quadratura (WICKERHAUSER, 1994; PARRAGA, 2002).

Definindo ψ como uma série em função de β e considerando t e η constantes, obtém-se:

$$\psi(\beta) = \sqrt{2} \Psi_\eta(2t - \beta) \quad (122)$$

pode-se escrever as Eq. (121b) e (121c), como a seguir:

$$\Psi_{2\eta}(t+k) = \sqrt{2} \sum_{\beta} h(2k - \beta) \cdot \Psi_\eta(2t + \beta) \quad (123a)$$

$$\Psi_{2\eta+1}(t+k) = \sqrt{2} \sum_{\beta} g(2k - \beta) \cdot \Psi_\eta(2t + \beta) \quad (123b)$$

onde: $\Psi_{2\eta}(t+k) = A_k$ e $\Psi_{2\eta+1}(t+k) = D_k$, sendo A_k e D_k os coeficientes de aproximação e detalhamento, respectivamente.

A coleção destas funções, associada aos filtros $h(\cdot)$ e $g(\cdot)$, é chamada de *wavelet packet*. A decomposição *wavelet packet* pode ser obtida a partir das seguintes equações:

$$v_{\kappa+1, 2\eta}(\beta) = \sum_k h(k) v_{\kappa\eta}(2\beta - k) \quad (124a)$$

$$v_{\kappa+1, 2\eta+1}(\beta) = \sum_k g(k) v_{\kappa\eta}(2\beta - k) \quad (124b)$$

onde: k e $\beta \in \mathbb{Z}$.

Lembrando que, caso $v_{\kappa,2\eta}$ e $v_{\kappa,2\eta+1}$ sejam dois 2 subespaços vetoriais ortogonais, $v_{\kappa+1,\eta}$ é a soma direta dos dois subespaços vetoriais, conforme a seguinte equação:

$$v_{\kappa+1,\eta} = v_{\kappa,2\eta} \oplus v_{\kappa,2\eta+1} \quad (125)$$

O produto interno de todo o sinal, na TWP, não precisa ser calculado novamente, ao se utilizar as equações (124a) e (124b). É necessário, apenas, aplicar recursivamente os filtros H e G nos coeficientes do nível de dilatação anterior.

Aplicando sucessivamente um certo número de filtros passa-baixa, h , seguidos de um único filtro passa-alta, g , ao sinal $x(t)$, obtém-se a TWD conforme a seguinte equação:

$$\frac{1}{\sqrt{2^\kappa}} \int_{\mathbb{R}} x(t) \cdot \psi_1 \left(\frac{t - \beta 2^\kappa}{2^\kappa} \right) dt = g h^{\kappa-1} x[n] \quad (126)$$

onde: $x[n]$ são os coeficientes da expansão de x no espaço gerado pela função de escala.

Na transformada *wavelet packet* o sinal também é dividido em coeficientes de aproximação (filtro passa-baixas) e detalhamento (um filtro passa-altas), como visto nas Eq. (124a) e (124b). Entretanto, tanto os coeficientes de aproximação, quanto os coeficientes de detalhamento, são novamente divididos em coeficientes de aproximação e detalhamento, e assim sucessivamente.

Deste modo, a transformada *wavelet packet*, semelhante a transformada *wavelet* discreta, também equivale à implementação de um banco de filtros não uniformes. Contudo, apresenta uma resolução maior do que a transformada *wavelet* discreta nas altas frequências.

Um sinal pode ser decomposto, por exemplo, em 8 grupos em uma transformada *wavelet packet* de 3 níveis: um grupo de aproximação da informação (HHH), um grupo de detalhamento (GGG) e 6 grupos de aproximação e detalhamento da informação (HHG, HGH, HGG, GHH, GHG, GGH) como visto na Fig. 89.

Os blocos retangulares marcadas com asterísco, vistos na Fig. 89, constituem um conjunto de bases particular para a representação do sinal $x(t)$. Vale ressaltar que esta representação não é única, há outros conjuntos de bases disponíveis que também podem representar a decomposição completa do sinal original em suas respectivas sub-bandas.

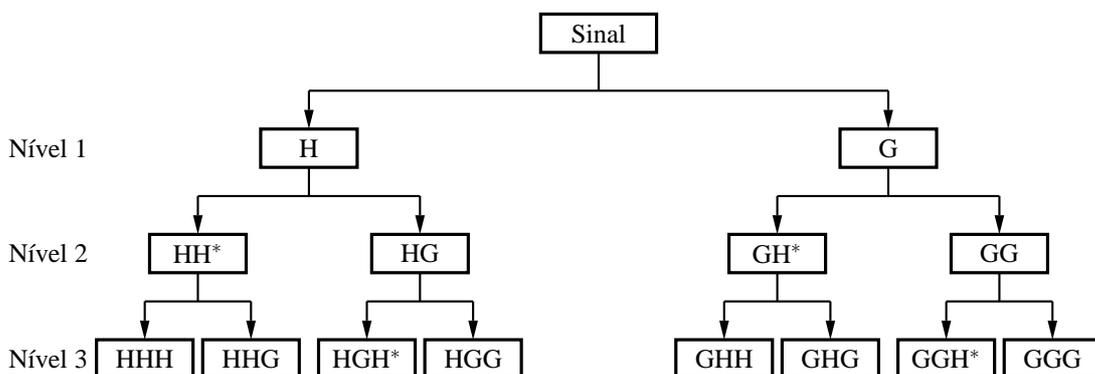


Figura 89: Exemplo de árvore de decomposição da TWP.

Na Fig. 90, pode-se observar um exemplo de uso da TWP. Nesse exemplo, obtido a partir do Matlab, um sinal ⁸ e o gráfico dos seus respectivos coeficientes do terceiro nível de aproximação podem ser visualizados.

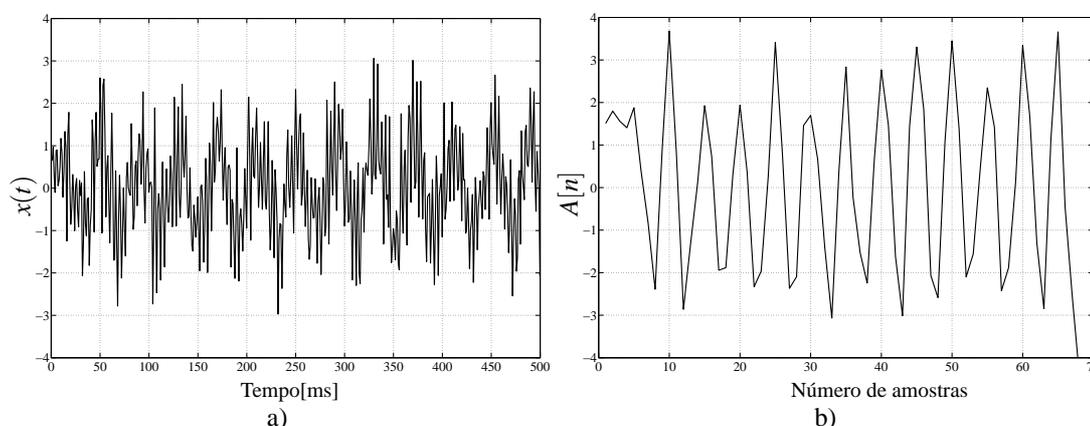


Figura 90: Exemplo do uso da TWP: a) sinal e b) coeficientes de aproximação (HHH).

7.5.2 Situações de Aplicação da Transformada *Wavelet Packet*

Visto que a transformada *wavelet packet* preserva a informação temporal e espectral e possui uma resolução espectral maior do que a transformada *wavelet* discreta para altas frequências, esta ferramenta pode ser a mais indicada para análise de sinais que variam ao longo do tempo; isto é, sinais não estacionários.

Possíveis aplicações podem incluir processos como impactos não periódicos como os que utilizam prensas, atuadores, válvulas, além dos processos médicos como a análise de sinais de voz, eletrocardiogramas, tomografias, entre outros (CROVATO; SCHUCK, 2007).

A capacidade de decompor o sinal nas regiões de altas frequências também torna a TWP uma ferramenta atrativa para analisar texturas de imagens (LAINE; FAN, 1993), decompor sinais e remover ruídos (RAMCHANDRAN; VETTERLI, 1993), e analisar sinais impulsivos de engrenagens, rolamentos, e outros equipamentos (YEN; LIN, 2000; YAN; LEE., 2005; EREN; DEVANEY, 2004).

Ainda, visto que o cálculo da transformada *wavelet packet*, em função do algoritmo usado, pode ser muito mais rápido do que o da transformada de Fourier e transformada de Fourier com janelamento, quando a velocidade computacional exigida é alta e a quantidade de dados envolvidos é grande, a transformada *wavelet packet* pode ser a melhor escolha.

7.6 COMPARAÇÃO DAS FERRAMENTAS DE PROCESSAMENTO DE SINAIS

Um quadro comparativo das ferramentas de processamento de sinais (TF, TFJ, TW e TWP), abordadas anteriormente, apresentando as principais vantagens e desvantagens de cada transformada, é visto na Tab. 11.

⁸ $x(t) = \text{sen}(2\pi 25t) + \text{sen}(2\pi 250t) + 0.5 \text{randn}(\text{size}(t))$.

Tabela 11: Quadro comparativo das ferramentas de processamento de sinais.

Ferramenta	Vantagens	Desvantagens
Transformada de Fourier	Boa resolução espectral Representação compacta do sinal Facilmente computável	Perda de informação temporal Não aplicável para sinais não estacionários
Transformada de Fourier com janelamento	Baseada na decomposição do sinal Razoável resolução espectral e temporal Aplicável para sinais não estacionários	Necessidade de bons computadores Cálculo demorado quando comparado com a TF e TWP (em função do algoritmo) Incapacidade de alteração da resolução do sinal
Transformada <i>wavelet</i>	Capacidade de alteração da resolução espectral e temporal do sinal Aplicável para sinais não estacionários	Baixa resolução espectral nas componentes de alta frequência
Transformada <i>wavelet packet</i>	Capacidade de alteração da resolução espectral e temporal do sinal Aplicável para sinais não estacionários Facilmente computável	A escolha da <i>wavelet</i> mãe é experimental e depende do sinal que está sendo processado

O estudo de caso apresentado neste trabalho, Capítulo 5, é composto por um atuador elétrico (que inclui um sistema de redução cinemática formado por um conjunto de engrenagens), válvula e tubulação. Este sistema é encontrado em diversos processos de diversas refinarias de petróleo e usinas de tratamento de água. Nestes locais, o conjunto é utilizado algumas vezes ao longo do dia com uma frequência de uso aleatória.

Devido as características dos sinais do estudo de caso (tais como um grande número de amostras) e devido aos objetivos a serem alcançados neste trabalho, tais como uma detecção e previsão de falhas desenvolvidas em *hardware*, a transformada *wavelet packet* mostrou-se, inicialmente, a ferramenta de processamento de sinais mais adequada para este estudo de caso.

7.7 ENERGIA DE SINAIS CONTÍNUOS E DISCRETOS

A representação no domínio da frequência de uma função (ou sinal), Eq. (83), é conhecido como seu espectro. A integral do produto entre o espectro e o seu conjugado é conhecido como energia de um sinal, ou simplesmente energia.

A energia de um sinal $x(t)$ contínuo, definida como \mathcal{E} , é obtida a partir da seguinte expressão (BRIGHAM, 1974; ROBERTS, 2001):

$$\mathcal{E}\{x(t)\} = \int_{-\infty}^{\infty} x(t)x^*(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)X^*(\omega)d\omega \quad (127)$$

Já para sinais discretos, a energia, \mathcal{L} , de um sinal discreto $x[n]$, é obtida a partir de:

$$\mathcal{L}\{x[n]\} = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (128)$$

A energia de um sinal (ES) também é conhecida como a assinatura característica deste sinal, pois revela o conteúdo de frequência do sinal; ou seja, suas características fundamentais.

Um outro método usado para obter a energia de um sinal, consiste em dividir o espectro em diversas bandas de frequência, previamente determinadas, e em seguida calcular a energia contida em cada banda.

A partir do Teorema de Parseval, pode-se chegar à expressão útil para a energia do sinal contida em cada banda:

$$\mathcal{E}\{x(\omega)\} = \int_{\omega_1}^{\omega_2} |X(\omega)|^2 d\omega \quad (129)$$

onde: ω_1 e ω_2 são, respectivamente, os limites inferior e superior da banda de frequência angular analisada.

O Teorema de Parseval também diz que caso as funções de escalonamento e as *wavelets* formem bases ortogonais, a energia do sinal $x(t)$ pode ser relacionada com a ES existente em cada uma das componentes e dos seus respectivos coeficientes *wavelets*.

Ou seja, a energia contida no sinal é igual a soma das energias contidas nos diferentes níveis de resolução da sua respectiva TWP. Isto é, a energia do sinal pode ser particionada em termos de seus coeficientes de detalhamento e aproximação do seguinte modo:

$$\mathcal{L}\{x[n]\} = \sum_{n=1}^N |x[n]|^2 = \sum_{n=1}^N |A_k[n]|^2 + \sum_{n=1}^N |D_k[n]|^2 \quad (130)$$

onde: A_k e D_k são os coeficientes de aproximação e detalhamento, vistos em (124a) e (124b), respectivamente, N é o número de amostras do sinal e $n \in \mathbb{N}$.

Com o sinal sendo decomposto em um número finito de sub-bandas, as características fundamentais de um sinal podem ser extraídas dos coeficientes da TWP em cada sub-banda para fornecer informações das condições do sistema que está sendo monitorado.

A Fig. 91 apresenta um exemplo do uso da ES, obtida a partir da TF, na análise de falhas, fornecido pelo IMS Center (QIU et al., 2006). A Fig. 91 a) exhibe os sinais de força de um transdutor acoplado a uma ferramenta quando nova e quando desgastada. A Fig. 91 b) exhibe o gráfico de energia destes sinais. Pode-se observar nitidamente, na Fig. 91 b) a diferença entre os sinais de energia quando a ferramenta esta nova e quando esta desgastada.

O conteúdo da energia de cada sub-banda do sinal está diretamente relacionado ao grau de severidade do defeito, isto pode ser usado como indicador efetivo das características chaves das condições das máquinas. Esta informação, sendo repassada para algum sistema com um certo grau de inteligência artificial, pode ser útil na detecção e predição de falhas, por exemplo.

7.8 APLICAÇÃO DE PROCESSAMENTO DE SINAIS

O uso correto das ferramentas de processamento de sinais permite aumentar a confiabilidade e disponibilidade de sistemas, torna mais robusto as atividades de detecção e predição de falhas, e em especial, reduz as inspeções e as atividades de manutenção e aumenta a vida útil de equipamentos.

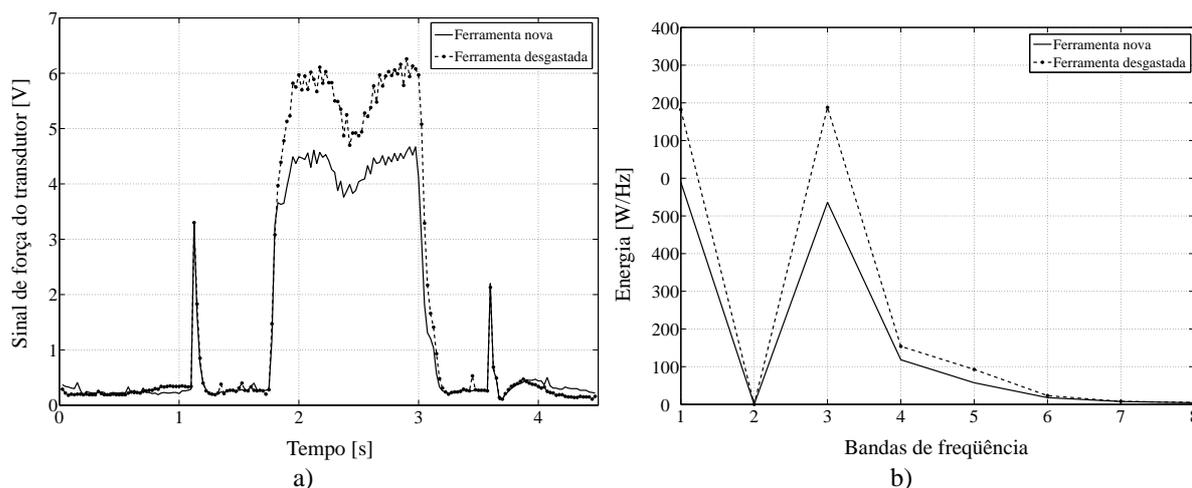


Figura 91: Sinais de a) força e b) de energia de um transdutor.

Na literatura, há muitos exemplos de uso das ferramentas de processamento de sinais voltados para atividades de manutenção em diversas áreas, tais como: aeroespacial, médica, ou mecânica, por exemplo, (STASZEWSKI; BOLLER; TOMLINSON, 2004; LIN et al., 2004).

Uma análise de falhas em rolamentos do estator de um motor de indução através do processamento de sinais foi realizada por (EREN; DEVANEY, 2004) e (NANDI; TOLIYAT, 1999). Os autores comentam que a análise do sinal de corrente do estator dos motores através da TWP (ou da TF) diminui as atividades e os custos envolvidos na manutenção.

A detecção de falhas têm sido uma preocupação constante na concepção, uso e manutenção de máquinas e equipamentos, como abordado por (DENG; WANG; GIURGIUTIU, 1999). Através de atividades de manutenção não intrusivas (tais como radiografias e análise de vibrações) e do processamento de sinais é possível detectar falhas em seu estado inicial de desenvolvimento.

7.9 CONCLUSÕES DO CAPÍTULO

O uso das técnicas de processamento de sinais, analógico ou digital, voltado para a análise e extração de alguma informação de sistemas, tem sido ampliado; sendo possível encontrá-las atualmente em distintas aplicações, tais como a compressão de imagens, a análise de séries temporais, ou a análise de sistemas de transmissão de energia.

Neste capítulo, as principais técnicas de processamento de sinais: transformada de Fourier, transformada de Fourier com janelamento, transformada *wavelet* e a transformada *wavelet packet*, nas versões analógica e digital, foram analisadas.

Ressaltou-se também, que as ferramentas de processamento de sinais são muito utilizadas na extração de características, no reconhecimento de padrões, e na detecção e predição de falhas, por exemplo. Aplicações destas transformadas, em conjunto com técnicas de IA, são encontradas na classificação de sinais de diversos tipos, tais como sinais de vibração de máquinas, sinais eletrocardiogramas, texturas de imagens, entre outras.

8 ANEXO B: INTELIGÊNCIA ARTIFICIAL

A fim de projetar e produzir sistemas que simulem uma atuação cada vez mais próxima daquela desempenhada pela inteligência humana foi criada uma disciplina, ou área, científica específica: a inteligência artificial (IA).

Uma das grandes preocupações dessa ciência é desvendar a natureza do pensamento humano para decifrar o processo criativo em suas múltiplas formas e nos mínimos detalhes. A elaboração do processo de raciocínio em rotinas que possam compreender e assimilar algum tipo de conhecimento é um dos principais objetivos da IA.

Neste capítulo, inicialmente, na seção 8.1, são apresentadas as principais características dos sistemas de inteligência artificial. Após, uma breve descrição das principais técnicas de inteligência artificial é vista na seção 8.2. As redes neurais artificiais, em especial os mapas auto-organizáveis e os mapas temporais de Kohonen, são analisadas na seção 8.3. Por fim, na seção 8.4, são apresentadas as conclusões desse capítulo.

8.1 PRINCIPAIS CARACTERÍSTICAS

No início dos anos 50, publicou-se o primeiro artigo, (TURING, 1950), que tratava a inteligência artificial como um novo conceito (REIS; PATI, 2000). Contudo, foi somente em 1956 que a área conhecida como inteligência artificial foi formalmente criada.

Nesse ano, nos Estados Unidos, John McCarthy reuniu, em uma conferência, na universidade de *New Hampshire*, vários pesquisadores de renome para estudar o que foi denominado de inteligência artificial, expressão usada para designar um tipo de inteligência construída pelo homem para dotar as máquinas de comportamentos inteligentes (FILHO, 2003).

Desde então, os sistemas de inteligência artificial que incluem os procedimentos, equipamentos, programas, dados e o conhecimento necessário para desenvolver sistemas e máquinas que demonstrem certas características de inteligência foram sendo aperfeiçoados.

Dentre as principais características dos sistemas de inteligência artificial, destaca-se:

- Capacidade de aprender com a experiência e de aplicar o conhecimento adquirido;
- Desenvoltura para resolver problemas quando informações importantes são perdidas;
- Possibilidade de reagir rápida e adequadamente frente a uma nova situação;
- Manipular e processar símbolos, sons e imagens;
- Possuir certo grau de criatividade e/ou imaginação.

Segundo (REIS; PATI, 2000), os sistemas de inteligência artificial são basicamente sistemas computacionais que executam certas funções dos seres humanos. A IA tenta imitar os métodos de aprendizagem e resolução de problemas, feito pelos seres humanos, através da aquisição do conhecimento sobre certos procedimentos.

8.2 TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Com a abertura e estruturação da inteligência artificial, esse novo campo do conhecimento começou a ser pesquisado de forma variada e intensa por diversos pesquisadores. Ainda, quando métodos de raciocínio utilizados pelo ser humano foram modelados e implementados no computador através do uso de diversas técnicas de IA, ocorreu uma expansão dessa para diversas áreas de pesquisa, tais como a robótica e a manutenção.

Isto se tornou possível em grande parte graças ao desenvolvimento de uma série de técnicas de inteligência artificial. Dentre estas, pode-se destacar:

- Algoritmos genéticos: são modelos matemáticos que servem para fornecer certo nível de aprendizado para sistemas. Esses foram, inicialmente, criados por Jonh Holland, (HOLLAND, 1975), com o objetivo de simular operadores genéticos específicos (como cruzamento, mutação e reprodução) da mesma forma como são observados na natureza;
- Programação evolutiva: proposto inicialmente em 1966 por Lawrence J. Fogel, (FOGEL; OWENS; WALSH, 1966), assemelha-se aos algoritmos genéticos, dando maior ênfase na relação comportamental entre os parentes e seus descendentes. As soluções para os problemas são obtidas e transmitidas para a nova população por meio de sucessivas tentativas, simuladas em rotinas;
- Lógica nebulosa: metodologia que serve para representar manipular e modelar informações incertas; foram estruturados por Lotfi Zadeh no ano de 1965 (ZADEH, 1965);
- Raciocínio baseado em casos: faz uso de uma grande biblioteca de casos para consulta e resolução de problemas. Os problemas são resolvidos através da recuperação e consulta de casos já solucionados e da conseqüente adaptação das soluções encontradas;
- Redes neurais artificiais: são redes massivamente paralelas e interconectadas, formadas por elementos simples com uma organização hierárquica. Estes elementos interagem de modo semelhante ao sistema nervoso biológico (MCCULLOCH; PITTS, 1943; HEBB, 1949; ROSENBLATT, 1958);
- Cadeias de Markov: técnica que permite determinar a probabilidade de um sistema se encontrar em cada um de seus estados em um determinado instante de tempo, por exemplo, (MARKOV, 1943).

Destaca-se que outras técnicas, tais como os filtros de Kalman e adaptativos, também possuem certas características de inteligência artificial, tal como a capacidade de aprender e de aplicar o conhecimento adquirido. Assim, estas ferramentas também podem ser utilizadas para armazenar conhecimento e auxiliar na realização de tarefas, tais como a detecção, classificação e previsão de falhas (BYINGTON; ROEMER; GALIE, 2002; JARDINE; LIN; BANJEVIC, 2006; LEE et al., 2006).

Destaca-se que as técnicas de inteligência artificial, resumidamente, visam elaborar e manipular entidades inteligentes tendo como último objetivo copiar o pensamento humano e tomar ações de modo inteligente.

Os pesquisadores da área de inteligência artificial usam essas diversas técnicas, tais como os sistemas especialistas, as redes neurais artificiais e a lógica nebulosa, por exemplo, que de uma forma ou de outra tentam imitar a lógica e o processo de raciocínio dos humanos, de modo a ajudar na tomada de decisões ou na realização de ações (CHOW, 1997; TINÓS, 1999; MARÇAL; SUSIN, 2005).

Em oposição às técnicas tradicionais de detecção, classificação e predição de falhas, onde equipamentos caros ou modelos matemáticos refinados são necessários, as técnicas de IA, como a lógica nebulosa (LN) ou as redes neurais (RN), podem ser usadas para fornecer um esquema de detecção, classificação e predição de falhas mais barato e eficiente (ALTUG; CHOW; TRUSSELL, 1999; SINGH; KAZZAZ, 2003).

Em especial, as técnicas de inteligência artificial podem ser utilizadas em sistemas de detecção, classificação e predição de falhas para três funções distintas: produção de classificadores baseados nas variáveis medidas do processo ou sistema; estimação dos parâmetros de um sistema; e reprodução do comportamento dinâmico do sistema.

Inúmeras são as referências encontradas a respeito de aplicações que fazem uso das técnicas de inteligência artificial. Um sistema de classificação de voz disfônica empregando a transformada *wavelet packet*, um algoritmo redutor de dimensionalidade e seis redes neurais artificiais atuando como um conjunto de sistemas especialistas foi apresentado por (CROVATO, 2004), por exemplo.

Tinós (TINÓS, 1999) apresentou um novo enfoque para detecção e diagnóstico de falhas em robôs utilizando duas redes neurais artificiais e empregando quatro algoritmos diferentes para treinar estas redes.

Garga (GARGA et al., 2002) introduziu um razoável método de predição híbrido, que integra diferentes ramos de conhecimento e dados de máquinas. Nesse trabalho, uma rede neural foi construída para gerar uma representação explícita do conhecimento.

Dois metodologias baseadas em lógica nebulosa, usadas para automatizar o diagnóstico de defeito em equipamentos mecânicos, além de uma comparação do desempenho utilizando casos experimentais, foram descritas por Fujimoto (FUJIMOTO, 2005).

Berni (BERNI, 2004) apresentou uma implementação de um módulo de equipamentos genérico e autônomo (um sensor virtual) que pode ser empregado em diversos processos complexos, que faz uso da lógica nebulosa.

As cadeias de Markov foram empregadas por Jens (JENS, 2006) para realizar cálculos dos índices de confiabilidade e disponibilidade em todos os pontos de um sistema elétrico de potência. Nesse trabalho, o autor fez uso de dados históricos de falhas como parâmetros de entrada para construir um modelo de confiabilidade e disponibilidade de um sistema elétrico de potência real.

Dois sistemas híbridos compostos por redes neurais e lógica nebulosa, usados para detecção e diagnóstico de falhas, oriundas da degradação gradual, em motores, foram apresentados por (ALTUG; CHOW; TRUSSELL, 1999). Ambos os sistemas forneceram bons resultados na detecção e diagnóstico de falhas.

Lee (LEE et al., 2006) apresentou um dispositivo que possui uma gama de ferramentas de IA que possibilitam responder, por exemplo, as seguintes questões: quando o equipamento observado irá falhar, qual é o motivo da degradação do desempenho da máquina, ou qual é a peça mais crítica no sistema que necessita de manutenção ou reparo.

Uma estratégia para detectar falhas em barras de rotores, que é o tipo mais comum de falha que pode ocorrer em motores de indução do tipo gaiola de esquilo trifásicos, baseada em RN e lógica nebulosa foi abordada por (TAN; HUO, 2005).

A metodologia de detecção de falhas proposta por (TAN; HUO, 2005) é inovadora e utiliza um modelo genérico e híbrido composto de redes neurais e lógica nebulosa para prever o comportamento de máquinas de mesmo modelo. O modelo híbrido (definido como *Neurofuzzy*) que captura as características do motor assíncrono apresentado é a peça chave da metodologia proposta nesse trabalho.

Um modelo *Neurofuzzy* é um sistema nebuloso definido de acordo com a estrutura de uma rede neural, de modo que a rede de raciocínio lingüístico ou sistema nebuloso possa ser fundido com a capacidade de aprendizagem das RN.

Um outro trabalho que abordou algumas questões que envolvem a previsão da vida útil de enrolamentos foi desenvolvido por (HUANG et al., 2005). Nesse trabalho, foram utilizados métodos de avaliação e previsão da degradação baseados em um tipo de rede neural conhecido como mapas auto-organizáveis (MAO).

Ainda, uma nova técnica de previsão de séries temporais não-lineares através de redes neurais, foi apresentada em (CASTRO, 2001).

Um novo algoritmo híbrido para a formação de RN foi apresentado em (NASR; CHTOUROU, 2009). Nesse trabalho, uma nova abordagem que utiliza RN e a LN, para treinar os pesos das camadas ocultas, e o método de descida do gradiente, para treinar os pesos da camada de saída, foi proposta. Foram realizadas também uma série de experimentos para avaliar o desempenho do método proposto e compará-la com outros algoritmos conhecidos.

Em (BEZERRA, 2004), foi feito um estudo comparativo entre diversas técnicas de detecção de falhas, tais como a análise da densidade espectral de potência e os filtros adaptativos (FA), em rolamentos por análise de vibração. Os resultados obtidos mostraram que, de todas as técnicas utilizadas, o FA foi a mais eficiente, sendo capaz de detectar novas falhas.

Em suma, as técnicas de inteligência artificial empregadas em conjunto com outras ferramentas para extração das características, tais como a TF e a TW, já são usadas com significativa eficiência substituindo ou auxiliando o homem em tarefas de aprendizagem, diagnóstico, previsão de padrões, automação de processos industriais e controle de qualidade (LEE et al., 2006; HELLING; SCHOOR; HELBERG, 2004).

8.3 REDES NEURAIS ARTIFICIAIS

O cérebro humano é formado por aproximadamente 100 bilhões de neurônios. Estes bilhões de neurônios se encontram interligados por uma intrincada rede de ramificações, através da qual eles se comunicam (por meio de sinapses), armazenam e elaboram informações (HERCULANO-HOUZEL; LENT, 2005).

Cada neurônio, como visto na Fig. 92, representa o ponto de chegada (dentritos) e de partida (axônios) de numerosos sinais enviados por outros neurônios mais próximos. Nas regiões onde os axônios entram em contato com os dentritos ocorrem as sinapses e troca de informações processadas pelo corpo celular.

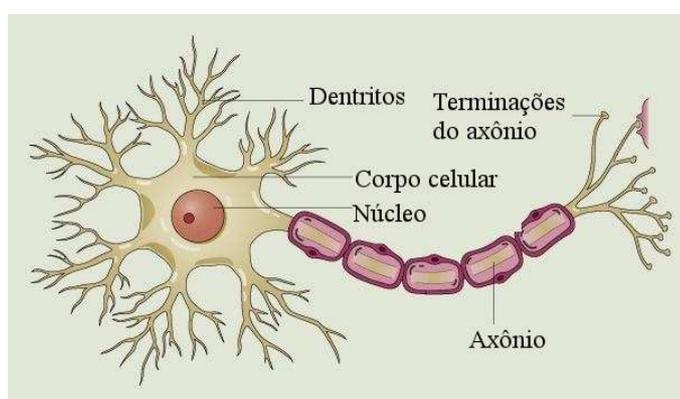


Figura 92: Neurônio biológico (WILEY, 2010).

Devido a essa complexa organização, o cérebro apresenta condições de executar variadas operações simultâneas e uma elevada capacidade de raciocínio (processamento de informações) em paralelo.

As redes neurais são sistemas computacionais inspirados nas características do sistema nervoso, mas especificamente, no funcionamento básico dos neurônios biológicos, como abordado por (TINÓS, 1999). O adjetivo “neural” é usado, obviamente, porque muito da inspiração de tais sistemas vem da Neurociência.

As RNA, semelhantes ao sistema nervoso, realizam uma representação distribuída da informação, na forma de conexões entre um grande número de neurônios artificiais que apresentam a mesma função: executar a soma ponderada de suas entradas e efetuar uma transformação linear, ou não-linear, sobre essa quantia, conforme (HUANG et al., 2005).

Deste modo, as RNA podem ser vistas como um modelo matemático dos neurônios biológicos e das suas interconexões de redes (MORETO, 2005).

8.3.1 Propriedades das Redes Neurais Artificiais

As redes neurais artificiais têm a capacidade de aprender padrões quando o processo regente dos dados é desconhecido, ou não estacionário, como o são a maior parte dos processos encontrados no mundo real (CASTRO, 2001).

Diferentemente da análise estatística tradicional, as redes neurais não requerem prévio conhecimento sobre a distribuição dos dados, para analisá-los.

Desde que haja uma relação entre os dados, mesmo que desconhecida sua representação analítica ou estatística, as redes neurais artificiais podem vir a apresentar um desempenho melhor do que os métodos estatísticos tradicionais.

Assim, as redes neurais artificiais são ferramentas extremamente flexíveis em um ambiente dinâmico. Elas têm a capacidade de compreender rapidamente padrões complexos e tendências presentes nos dados e de se adaptar rapidamente às mudanças, características essas que são extremamente desejáveis em se tratando de predição de séries temporais, por exemplo.

Dentre algumas características relevantes das redes neurais artificiais, que justificam e fundamentam a adequação da aplicação de redes neurais artificiais ao processamento de sinais e a aprendizagem de processos estocásticos, (MORETO, 2005; KROSE; SMAGT, 1996), pode-se citar:

- Não-Linearidade: é a possibilidade de considerar o comportamento não-linear dos fenômenos físicos responsáveis pela geração dos dados de entrada;
- Aprendizagem: é capacidade de aprender acerca de seu ambiente, o que é atingido através de uma sessão de treinamento com exemplos entrada/saída que sejam representativos do ambiente;
- Generalização: é capacidade de obter uma saída adequada a uma entrada não presente no processo de aprendizagem, o que torna possível apresentar um desempenho satisfatório em resposta a dados não pertencentes ao conjunto de treino;
- Tolerância à falhas: o conhecimento é distribuído pela RNA, o que permite à rede continuar a exibir resultados aceitáveis no caso de falha de alguns neurônios;
- Adaptabilidade: é a possibilidade da implementação em outros ambientes, através de um novo treinamento.

Assim, as RNA podem ser aplicadas para a realização de inúmeras tarefas, como por exemplo: aproximação de funções (especialmente as não-lineares), associação de padrões, classificação de padrões, processamento de imagens e predição de séries temporais.

Em sistemas de detecção e diagnóstico de falhas, as redes neurais artificiais têm sido empregadas nos últimos anos especialmente em sistemas estáticos e menos intensivamente em sistemas dinâmicos.

Na maioria das aplicações, as redes neurais artificiais têm sido utilizadas como classificadores baseados nas variáveis medidas do processo. Contudo, há pouca pesquisa científica centrada na investigação de como utilizar as redes neurais para prever o tempo de vida de atuadores elétricos e válvulas, em especial.

8.3.2 Modelo de Neurônio Artificial

O primeiro modelo matemático de RN foi apresentado em (MCCULLOCH; PITTS, 1943). O modelo geral adaptado, do modelo McCulloch & Pitts, que emprega como unidade de processamento fundamental o neurônio artificial, pode ser visto na Fig. 93.

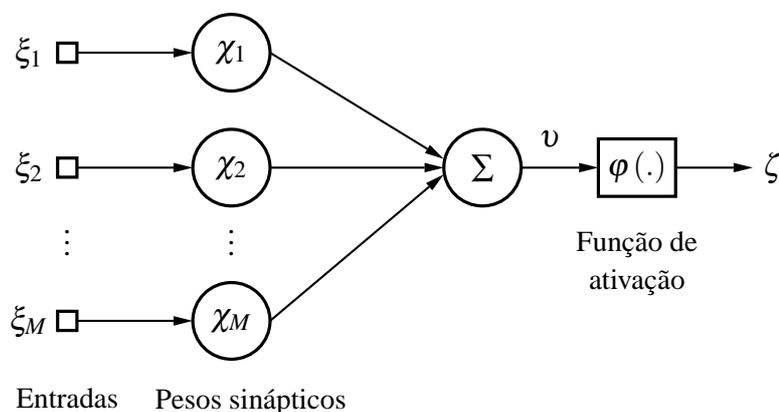


Figura 93: Modelo de neurônio artificial.

A operação desse neurônio artificial pode ser descrita do seguinte modo:

1. Os sinais são apresentados às entradas ξ ;
2. Cada sinal é multiplicado por um número χ que indica o seu peso na saída do neurônio;
3. É realizada a soma ponderada dos sinais produzindo um nível de atividade v ;
4. Caso esse nível de atividade exceda certo limite, a unidade ativa sua saída ζ .

O funcionamento do neurônio artificial pode ser descrito pela seguinte equação:

$$\zeta = \varphi(v) = \varphi\left(\sum_{m=1}^M \chi_m \xi_m\right) \quad (131)$$

onde: ξ são as entradas, χ são os pesos sinápticos dos neurônios, e a ativação da saída, ζ , é definida por uma função de ativação $\varphi(\cdot)$ específica, m e $M \in \mathbb{N}$, com $1 \leq m \leq M$.

8.3.3 Tipos de Redes Neurais Artificiais

Existe uma grande variedade de tipos de redes neurais artificiais que diferem, basicamente, em suas arquiteturas, formas de aprendizado e funções de ativação (SOUZA, 2005).

Entre as diferentes arquiteturas, existem as com e sem realimentação (de uma única camada ou de múltiplas camadas). Entre as formas de aprendizado sobressaem a supervisionada e a não-supervisionada (KROSE; SMAGT, 1996). Já entre as funções de ativação, destacam-se as funções de limiar e as gaussianas.

Os dois tipos de RNA supervisionadas mais conhecidas são as redes *perceptron* de múltiplas camadas (PMC) e as redes de funções de base radial (FBR) (TINÓS, 1999). Ambas as redes são aproximadoras universais ¹, a principal diferença entre elas é que as redes de funções de base radial tendem a produzir aproximações locais, enquanto que as redes *perceptron* de múltiplas camadas tendem a resultar em aproximações globais.

Quando se trata de aprendizado continuado, por exemplo, como no caso da predição de séries temporais, as redes de FBR se mostram mais adequadas. Isto se deve ao fato de que o custo computacional de treino de uma rede PMC é muito superior ao de uma rede de funções de base radial, o que impossibilita a operação de forma dinâmica (CASTRO, 2001).

Há ainda um tipo específico de RN, não-supervisionada, que pode ser realimentada ou não, com uma arquitetura bidimensional (ou tridimensional), conhecido como mapas auto-organizáveis, visto em mais detalhes na subseção 8.3.9.

8.3.4 Arquiteturas das Redes Neurais Artificiais

O modo como os neurônios se encontram interligados dentro de uma RN, a dimensão dos vetores e o número de camadas existentes definem as distintas arquiteturas de RNA existentes. Entre as diferentes arquiteturas, existem as sem realimentação e as com realimentação, com somente uma camada ou de múltiplas camadas, ambas vistas a seguir.

8.3.4.1 Redes Neurais Não Realimentadas

Nesta arquitetura de RNA, os neurônios se encontram organizados somente em uma camada, Fig.94 a), ou em múltiplas camadas, Fig. 94 b).

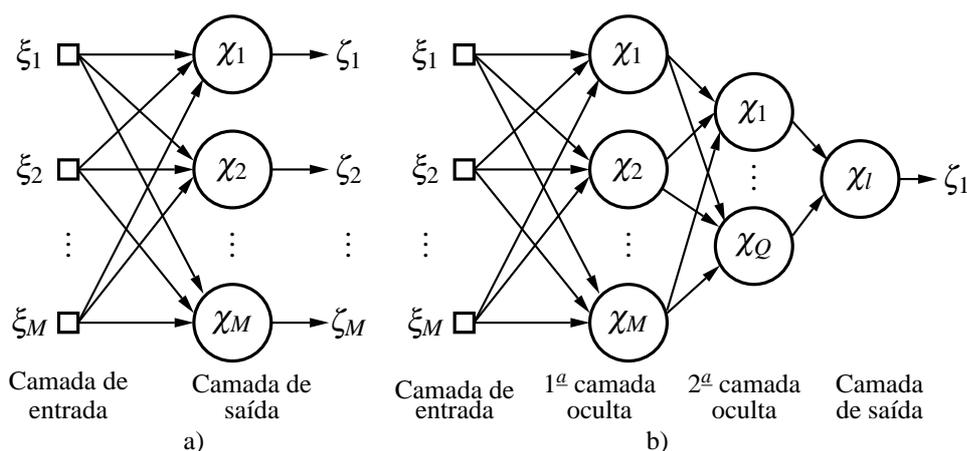


Figura 94: RN não realimentadas de: a) uma única camada e b) múltiplas camadas.

Na arquitetura com uma única camada, os pesos sinápticos estão implícitos nas conexões entre a saída e entrada. Já na arquitetura com múltiplas camadas, há uma ou mais camadas entre a saída e a entrada da rede.

¹Isto é, qualquer função pode ser aproximada por estas redes neurais com precisão arbitrária.

8.3.4.2 Redes Neurais Realimentadas

Neste tipo de arquitetura de RNA, os neurônios se encontram realimentados por pelo menos um laço de realimentação, como visto na Fig. 95. Os neurônios que são utilizados no laço de realimentação são definidos como neurônios ocultos e servem basicamente para melhorar a capacidade de aprendizado e desempenho da RN.

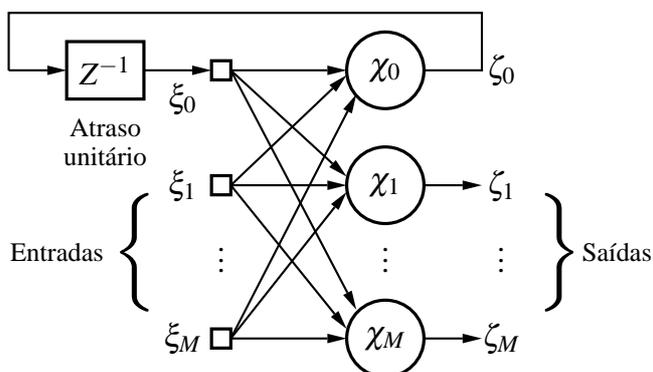


Figura 95: Rede neural com realimentação.

8.3.5 Paradigmas de Aprendizagem

A capacidade de aprender sobre o seu ambiente e com isso melhorar o seu desempenho é uma das mais importantes propriedades das RN. Isso pode ser realizado por meio de um processo iterativo de ajustes aplicados aos pesos sinápticos da rede, chamado de treinamento.

Assim, após a definição da arquitetura e da função de ativação de uma rede neural, é necessário realizar um treinamento, para que essa possa aprender sobre o sistema ou ambiente que se deseja analisar. Para tal, é necessário determinar uma lei matemática, uma regra de aprendizagem, que defina o processo de alteração, ou treinamento, dos pesos sinápticos.

O treinamento de uma rede neural em geral consiste em reunir uma amostra de dados do sistema em questão e submetê-los sequencialmente a rede, segundo o algoritmo visto a seguir:

1. Submeter as amostras individualmente através do vetor de entrada;
2. Realizar uma ponderação através do(s) vetor(es) de peso(s) sináptico(s);
3. Aplicar a função de ativação;
4. Verificar a saída, caso o processo de aprendizado seja supervisionado;
5. Atualizar os pesos sinápticos segundo a regra de aprendizagem;
6. Retornar ao item 1, caso o critério de interrupção não tenha sido alcançado.

O aprendizado ocorre quando a rede neural alcança uma solução generalizada para uma determinada classe de problemas. Após o treinamento a rede neural é capaz de desempenhar algumas tarefas, tais como a detecção e o diagnóstico de padrões.

Geralmente, os modelos de RNA apresentam alguma regra de treinamento, segundo a qual os pesos de suas conexões sinápticas são ajustados conforme os padrões apresentados; isto é, a rede aprende por meio de exemplos provenientes de casos reais já conhecidos, extraindo, assim, as regras básicas a partir dos exemplos (HUANG et al., 2005).

O processo de aprendizado das RNA pode ser dividido, simplificadaamente, em duas categorias: o aprendizado realizado com professor e aquele realizado sem professor. O professor é um sistema, uma outra rede neural, por exemplo, capaz de fornecer uma resposta desejada para um vetor de treinamento.

Dentre estas duas categorias, os três principais paradigmas de aprendizagem existentes são: o aprendizado supervisionado (com professor), o não-supervisionado e o aprendizado por reforço (ambos sem professor).

8.3.5.1 *Aprendizado Supervisionado*

O modelo de aprendizado supervisionado de RNA, também conhecido como aprendizado com um professor, faz uso de um bloco (definido como professor) para treinar a RNA, como visto na Fig. 96.

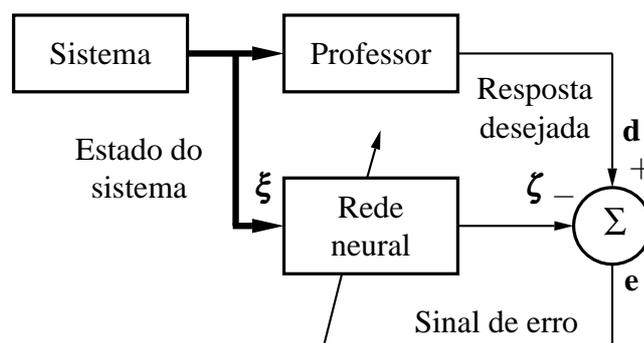


Figura 96: Aprendizado supervisionado.

No paradigma de aprendizagem supervisionado, o estado do sistema é representado por um vetor que é aplicado à entrada da rede neural artificial (ξ) e ao professor. O professor possui conhecimento técnico sobre o sistema, fornecendo assim a resposta desejada (d). A saída da rede neural pode ser descrita pela Eq. (131).

Dentre os principais algoritmos de aprendizado supervisionado, pode-se citar: o de retropropagação de erro. O erro quadrático médio, ou a soma dos erros quadráticos, pode ser utilizado, por esse e outros algoritmos, como medida de desempenho do treinamento da RNA.

8.3.5.2 *Aprendizado Não-supervisionado*

Em algumas aplicações não se conhece o comportamento das variáveis (resposta desejada). Nesse caso, devem ser empregadas redes neurais artificiais com aprendizado não supervisionado (ou auto-organizável) (TINÓS, 1999; KROSE; SMAGT, 1996).

Como o nome mesmo sugere, nessa metodologia de aprendizagem, não há um professor responsável pelo fornecimento da resposta desejada; isto é, não há exemplos da função a ser aprendida pela rede. O treinamento é realizado com base em uma medida de qualidade da representação, independente da tarefa que a rede necessita realizar, e os parâmetros da rede são atualizados com relação a essa medida.

Para o treinamento desse tipo de rede são usados somente os estados do sistema, como visto na Fig. 97. Assim, o aprendizado se dá pela própria redundância das entradas.

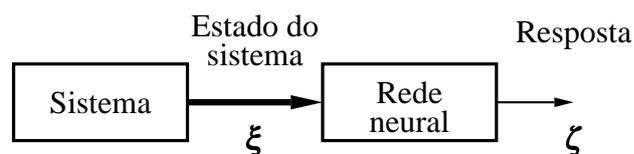


Figura 97: Aprendizado não-supervisionado.

No aprendizado não supervisionado, a rede faz uso dos neurônios como classificadores das entradas. Para isso, utiliza processos de competição e cooperação entre os neurônios vizinhos da rede neural.

Uma vez que a rede está treinada de acordo com as propriedades estatísticas dos dados de entrada, ela apresenta a capacidade de formar representações internas para codificar as características dos dados de entrada e assim criar novas classes automaticamente.

Dentre os diferentes algoritmos de aprendizado não supervisionado, pode-se citar: o método da máxima verossimilhança, o método dos mínimos quadrados e o método da máxima entropia (CRAMER, 1986; SALOMONI et al., 2007).

8.3.5.3 *Aprendizado por Reforço*

A aprendizagem por reforço é um paradigma computacional de aprendizagem em que um algoritmo busca maximizar uma medida de desempenho baseado nos reforços (recompensas ou punições) que recebe ao interagir com um sistema desconhecido (RIBEIRO; KOERICH; ENEMBRECK, 2006; MONTEIRO; RIBEIRO, 2004).

No aprendizado por reforço, o treinamento é realizado através de uma interação contínua com o sistema de modo a minimizar algum índice ou critério de desempenho pré-estabelecido.

A Fig. 98 exhibe os principais blocos de um exemplo de sistema de aprendizado por reforço. O bloco definido como crítico converte um sinal de reforço primário, recebido do sistema, em um sinal de melhor qualidade chamado sinal de reforço heurístico².

Nesse tipo de aprendizado, a rede neural é treinada através da liberação atrasada do reforço, o que significa que a rede recebe uma seqüência temporal de estímulos (vindos também do sistema), que acabam resultando na geração do sinal de reforço heurístico.

²Uma heurística pode ser definida como uma técnica que melhora, na média, a eficiência na solução de um problema (RUSSELL; NORVIG, 2004).

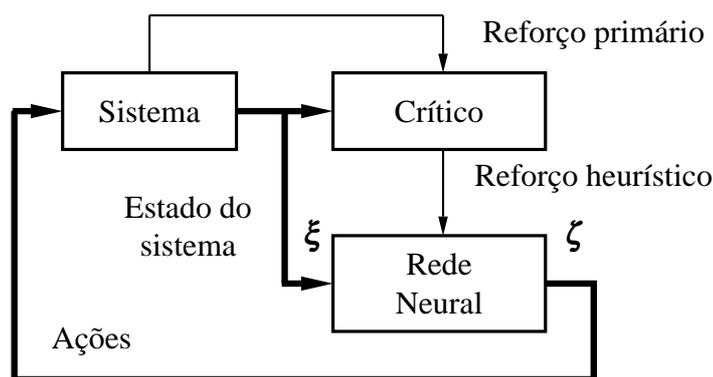


Figura 98: Aprendizado por reforço.

O objetivo da aprendizagem por reforço é minimizar uma função de custo. Essa função de custo é definida como a expectativa do custo acumulado das ações realizadas em uma seqüência de etapas (KROSE; SMAGT, 1996).

Pode acontecer que algumas das ações tomadas anteriormente, em uma determinada seqüência de passos, são de fato as melhores determinantes do comportamento global do sistema. A função da rede neural é descobrir essas ações para então realimentá-las de volta ao sistema.

As duas principais dificuldades encontradas ao se implementar o aprendizado por reforço retardada são:

- Não há um professor para fornecer uma resposta desejada a cada passo do processo de treinamento;
- O atraso na geração do sinal de reforço primário implica que o treinamento da rede deve levar em conta a questão de alocação temporal dos pesos sinápticos e ações. Isto é, a rede deve ser capaz de avaliar o peso de cada ação ao longo de todo o tempo necessário para gerar a resposta final.

Apesar destas dificuldades, a aprendizagem por reforço é muito interessante. Ele fornece a base para que um sistema interaja com seu ambiente, desenvolvendo assim a capacidade de aprender e executar uma tarefa estipulada unicamente com base nos resultados da sua experiência que resulta do treinamento.

Quando não existem modelos do sistema em estudo disponíveis, o aprendizado por reforço é uma técnica muito útil para resolver uma série de questões de controle e planejamento, por exemplo. Isso se deve ao fato de que os algoritmos apresentam uma convergência, para um ponto de equilíbrio, garantida.

Dentre os principais algoritmos voltados para o aprendizado por reforço, cita-se: o algoritmo *Q-learning* e o *K-Nearest-Neighbor*.

8.3.5.4 Tarefas de Aprendizagem

A escolha de um determinado algoritmo de aprendizagem é influenciada pela tarefa de aprendizado que a rede neural terá que executar. Nesse contexto, algumas tarefas de aprendizagem em especial merecem destaque, tais como:

- Controle;
- Filtragem;
- Aproximação de funções;
- Associação e identificação de padrões.

8.3.6 Redes de Funções de Base Radial

As redes neurais artificiais de base radial são inspiradas em neurônios que têm ativações localmente sintonizadas ou neurônios seletivos, que respondem somente para determinadas faixas de sinais de entrada.

A arquitetura das redes neurais de funções de base radial é tal que apresenta uma camada escondida definida por um conjunto de funções de base radial, das quais a rede deriva seu nome, como visto na Fig. 99. As funções de base radial são, simplesmente, uma das classes de funções existentes (CASTRO, 2001).

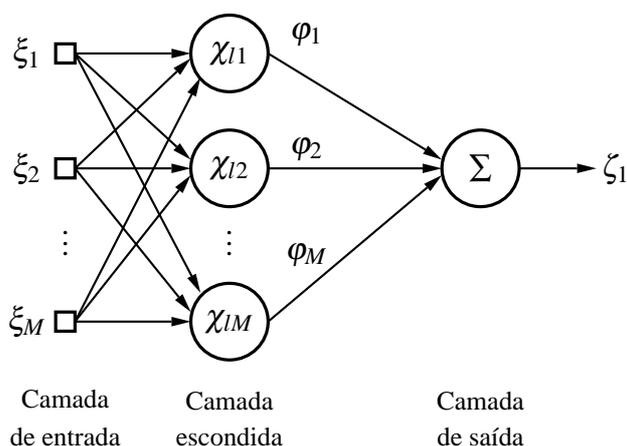


Figura 99: Rede neural do tipo radial.

De um modo geral, o aprendizado de uma rede neural de FBR equivale ao ajuste de uma superfície não-linear a um conjunto de dados, em um espaço multidimensional, de acordo com algum critério estatístico pré-determinado.

O processo de generalização consiste em usar essa superfície multidimensional para interpolar outros pontos que não pertençam ao conjunto de treino, mas que estejam em torno de sua vizinhança (CASTRO, 2001).

Os neurônios da camada escondida de uma rede neural de FBR podem ser vistos com um conjunto de funções que forma uma base no espaço onde os dados de entrada podem ser representados. Cada uma destas funções de base radial está centrada em uma coordenada específica do espaço multidimensional. Estas coordenadas, por sua vez, irão definir os centros de região de maior aglomeração de pontos do espaço de dados de entrada.

Para implementar uma rede neural de FBR é necessário obter, através de um processo de aprendizagem, os valores adequados dos parâmetros da RN. Estes parâmetros incluem as variâncias, os vencedores e os pesos sinápticos.

O treinamento compreende a determinação destes parâmetros de modo que, dado um conjunto de entradas ξ , as saídas ζ estejam o mais próximo possível dos valores desejados (CASTRO, 2001).

Para realizar o ajuste dos parâmetros, diferentes algoritmos podem ser usados, tais como: o algoritmo *k-means*, o método do gradiente estocástico e o algoritmo de Moore-Penrose. Estes algoritmos podem ser utilizados na inicialização ou atualização dos vencedores e dos pesos sinápticos da rede neural. A combinação destes diferentes algoritmos irá resultar em diferentes modos de treinamento.

8.3.7 Redes *Perceptron* de Múltiplas Camadas

No final da década de 1950, Frank Rosenblatt (ROSENBLATT, 1958) deu prosseguimento às idéias de McCulloch e Pitts, propondo uma rede de múltiplos neurônios com múltiplas camadas baseada na unidade de limiar linear, conhecida como *Perceptron* (KROSE; SMAGT, 1996).

Uma rede *perceptron* com apenas uma camada escondida, com a soma ponderada e a função de pertinência omitidas, pode ser vista na Fig. 100.

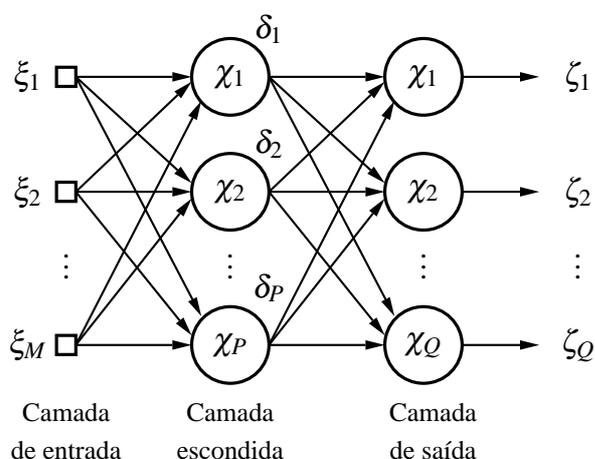


Figura 100: Rede neural do tipo *perceptron* com uma única camada escondida.

A relação entrada/saída de uma rede *perceptron* define um mapeamento de um espaço de entrada Euclidiano de dimensão M para um espaço de saída Euclidiano de dimensão Q .

O treinamento supervisionado de uma RNA *perceptron*, conforme visto na seção 8.3.5.1, consiste em ajustar os pesos sinápticos dos neurônios da rede de modo que o sinal de erro resultante da comparação com a saída desejada apresente o menor valor possível.

As redes *perceptron* de uma ou de múltiplas camadas com aprendizado supervisionado são, atualmente, as redes neurais artificiais mais utilizadas na prática, devido a sua grande capacidade de reconhecer padrões e realizar mapeamentos (VIEIRA; BAUCHSPIESS, 1999).

Isso se deve principalmente a relativa facilidade de implementação do algoritmo de treinamento supervisionado desenvolvido por Rumelhart, Hilton e Williams em 1986 (RUMELHART; HINTON; WILLIAMS, 1986) chamado de algoritmo de retropropagação³.

O algoritmo de retropropagação é uma generalização do algoritmo dos mínimos quadrados médios que utiliza a técnica do gradiente descendente de forma iterativa, para minimizar uma função de custo igual à diferença média quadrática entre a saída desejada e a saída da RNA.

8.3.8 Tipos de Funções de Ativação

A função de ativação é dada pela expressão que determina a lei matemática empregada na ativação ou disparo de um neurônio durante o treinamento ou teste de uma rede neural (SOUZA, 2005; RIBEIRO; KOERICH; ENEMBRECK, 2006).

As funções de ativação geralmente empregadas nas redes *perceptron* de múltiplas camadas são as funções de limiar, linear por partes, ou sigmóides, vistas na Tab. 12.

Tabela 12: Funções de ativação de redes *perceptron* de múltiplas camadas.

Função de Limiar	$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases}$
Função linear por partes (ou função de saturação)	$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 1/2 \\ v - 1/2, & \text{se } -1/2 < v < 1/2 \\ 0, & \text{se } v \leq -1/2 \end{cases}$
Sigmóide	$\varphi(v) = \frac{1}{1 + e^{(-v)}}$

Já dentre as funções de ativação voltadas para as redes de função de base radial, destacam-se as lâminas *spline* finas, as multi-quadráticas, e as gaussianas. Alguns tipos de funções de ativação de redes de função de base radial são vistas na Tab. 13.

Na Tab. 13, para as funções Gaussiana e lâmina *spline* fina, o parâmetro ρ determina o raio de influência de cada função; isto é, ρ define a taxa de decaimento do valor da função de base radial.

³O algoritmo de retropropagação também é conhecido, em inglês, por *back-propagation*.

Tabela 13: Funções de ativação de redes de função de base radial.

Lâmina <i>spline</i> fina	$\varphi(v) = \frac{v}{\rho^2} \log\left(\frac{v}{\rho}\right)$
Multi-Quadrática	$\varphi(v) = \sqrt{(v^2 + \rho^2)}$
Multi-Quadrática Inversa	$\varphi(v) = \frac{1}{\sqrt{(v^2 + \rho^2)}}$
Gaussiana	$\varphi(v) = e^{\left(\frac{-v^2}{2\rho^2}\right)}$

No caso da função Gaussiana, função mais utilizada na prática, ρ é o seu próprio desvio-padrão. Deste modo, ρ define a distância Euclidiana média que mede o espalhamento dos dados representados pela função de base radial em torno de seu centro, por exemplo.

8.3.9 Mapas Auto-organizáveis

Uma classe de rede neural, em especial, que possui um paradigma de aprendizagem não supervisionado é conhecida como mapas auto-organizáveis ⁴. Esse tipo de rede neural foi proposto inicialmente por Kohonen em 1982 (KOHONEN et al., 1995; KROSE; SMAGT, 1996; CHAPPEL; TAYLOR, 1993).

Segundo (BASTOS, 2007), um mapa topológico auto-organizável é definido por um arranjo de neurônios representando diferentes elementos processadores, ou filtros, operando em paralelo sobre sinais portadores de informação.

Um mapa auto-organizável é caracterizado pela formação de um mapa topográfico dos estímulos de entrada, em que as localizações espaciais (coordenadas) dos neurônios, na rede neural, são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada (BOSA, 2009).

Os mapas auto-organizáveis, ou mapas de Kohonen, são representados por estruturas matriciais, onde cada elemento, semelhante as RN, também é denominado de neurônio. Contudo, em um MAO os neurônios são alocados em nós formando uma rede, ou grade, bidimensional, como visto na Fig. 101.

Definindo $\tilde{\xi}$ como um espaço contínuo de dados de entrada, possuindo uma topologia determinada por certas relações métricas entre um conjunto de vetores ξ , com $\xi \in \tilde{\xi}$. Supondo que, do espaço $\tilde{\xi}$, sejam conhecidas somente um conjunto finito de amostras e que estas estejam organizadas conforme uma densidade de probabilidade $p(\xi)$.

⁴Os mapas auto-organizáveis também são conhecidos como SOM, do inglês *Self Organizing Maps*.

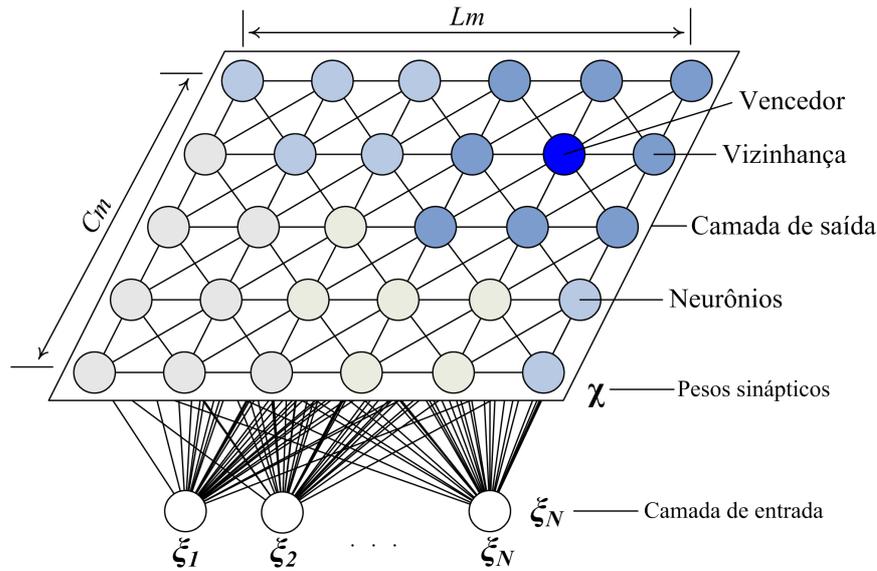


Figura 101: Arquitetura dos mapas auto-organizáveis.

Determinando agora um espaço contínuo de dados de saída $\tilde{\chi}$ que apresente uma topologia definida por um arranjo geométrico de um conjunto de J neurônios.

Deste modo, pode-se definir então o mapeamento do espaço de entrada $\tilde{\xi}$ para o espaço de saída $\tilde{\chi}$ a partir de uma transformação não linear \mathcal{M} , do seguinte modo:

$$\mathcal{M} : \tilde{\xi} \rightarrow \tilde{\chi} \quad (132)$$

Sendo N a dimensão do vetor do espaço contínuo de entrada $\tilde{\xi}$. Um vetor característico de entrada $\xi \in \tilde{\xi} \subset \mathbb{R}^N$, escolhido aleatoriamente, é definido como: $\xi = [\xi_1 \xi_2 \dots \xi_N]^T$. Cada neurônio do mapa auto-organizável apresenta um vetor de pesos sinápticos χ com a mesma dimensão do vetor de entrada ξ . A representação do vetor de pesos sinápticos χ de um neurônio j é dada por: $\chi^j = [\chi_1^j \chi_2^j \dots \chi_N^j]^T$.

Durante o treinamento, o MAO aprende a partir de exemplos, realizando um mapeamento de um espaço de entrada multidimensional $\tilde{\xi}$ em um espaço bidimensional (ou tridimensional) $\tilde{\chi}$ de J neurônios ($J = C_m L_m$); isto é, o mapa $\mathcal{M}(\chi)$, definido pela matriz χ , atribui ao vetor de entrada atual, ξ , um índice de neurônio no mapa.

Os mapas auto-organizáveis são redes neurais que utilizam processos de aprendizagem competitiva durante o treinamento. Supondo que um padrão de entrada ξ de dimensão N seja apresentado, à camada competitiva, durante o processo de aprendizagem, no instante de tempo t . Os neurônios de saída da rede, dentro de uma certa vizinhança, competem e interagem entre si para serem ativados ou disparados, sendo que apenas um neurônio de saída que vence a competição é definido como neurônio vencedor.

A vizinhança é então novamente determinada pelos neurônios mais próximos ao neurônio vencedor. Somente o vencedor e a sua vizinhança têm os pesos sinápticos alterados novamente de modo a ficar ainda mais próximo do padrão de entrada.

Deste modo, os neurônios podem se tornar seletivamente sintonizados a vários padrões de entrada (estímulos) ou classes de padrões de entrada no decorrer de um processo de aprendizagem (KROSE; SMAGT, 1996).

A similaridade entre padrões, geralmente, é obtida com base em uma medida de distância vetorial (distância Euclidiana, por exemplo). Quando um padrão de entrada é avaliado como não similar em relação aos agrupamentos anteriormente formados, um novo agrupamento é estabelecido. Usualmente, a topologia, ou estrutura do mapa, utilizada para conectar os neurônios aos neurônios adjacentes no mapa é retangular ou hexagonal.

Com base nas medidas de similaridade entre os agrupamentos formados, são eleitos neurônios vencedores para cada padrão de entrada. Segundo Kohonen (KOHONEN et al., 1995), as localizações dos neurônios vencedores se tornam ordenadas entre si de forma que um sistema de coordenadas significativo para diferentes características, ou padrões, de entrada seja criado sobre o mapa.

Uma organização estrutural semelhante ao mapas auto-organizáveis também é encontrada no cérebro humano e de animais, mais especificamente na região do córtex cerebral. Nele, regiões fisicamente distintas são responsáveis pelo processamento de diferentes informações, como a fala, a audição e a visão, por exemplo. Em particular, as entradas sensoriais como a tátil, a visual e acústica são mapeadas para diferentes áreas do córtex cerebral de uma maneira topologicamente ordenada.

O desenvolvimento de mapas auto-organizáveis como modelo neural é motivado por essa característica distinta do cérebro humano: organizar as informações e sensações em distintas regiões do cérebro. Assim, de modo semelhante ao cérebro, os mapas auto-organizáveis se constituem num método de mapeamento ordenado de informações em distintas regiões de uma rede neural artificial.

8.3.9.1 Características Fundamentais

Uma característica importante dos mapas auto-organizáveis é que ele utiliza um treinamento não-supervisionado. Nesse, a rede busca encontrar similaridades baseando-se apenas nos padrões de entrada, sem nenhum conhecimento prévio dos dados. Esse aspecto dos mapas auto-organizáveis pode ser útil para detectar características importantes do sistema em estudo quando há pouco conhecimento sobre tal.

Ao contrário de redes baseadas em aprendizado supervisionado, que exige que os valores alvo sejam vetores conhecidos, o mapa auto-organizável pode ser usado para classificar dados sem saber a que grupo os dados de entrada pertencem a priori. Ainda, em contraste com os métodos tradicionais, como a análise das componentes principais, os mapas auto-organizáveis também podem ser criados a partir de dados de sistemas não-lineares, agrupando os dados de entrada que são semelhantes entre si, formando classes ou grupos (HUANG et al., 2005).

8.3.9.2 Principais Objetivos

O principal objetivo de um MAO é transformar padrões de entrada, sinais de dimensão arbitrária, em um mapa discreto de uma ou duas dimensões. Essa transformação é feita de uma maneira adaptativa, preservando a mesma ordem topológica dos dados de entrada (KROSE; SMAGT, 1996).

Outro objetivo é a formação de padrões de dados com características distintas. Isso permite, após o treinamento, efetuar uma classificação no mapa, de um dado recente, com base nos padrões já existentes e já mapeados.

8.3.9.3 Etapas dos Mapas Auto-organizáveis

O processamento interno dos algoritmos dos MAO pode ser dividido, simplificadamente, em três etapas, ou fases, distintas (ASAMER; DIN; WERNER, 2007):

1. Inicialização: os pesos sinápticos da grade são inicializados;
2. Treinamento: é realizada a aquisição do conhecimento pelo mapa;
3. Teste ou recuperação: os dados de entrada são classificados no mapa.

8.3.9.4 Etapa de Inicialização

Na etapa de inicialização os vetores de pesos sinápticos do MAO podem ser inicializados aleatoriamente ou linearmente. Em ambos os tipos de inicialização, um mapa estrutural é criado inicialmente e nenhuma organização prévia é imposta ao mapa.

Quando a inicialização do MAO é aleatória, os valores são distribuídos aleatoriamente entre os valores mínimo e máximo de χ , para cada componente do vetor de entrada ξ .

Na inicialização do tipo linear, a inicialização dos pesos sinápticos é feita distribuindo os valores entre os valores mínimo e máximo de χ de modo linear ou seqüencial.

8.3.9.5 Etapa de Treinamento

Durante a etapa de treinamento, ocorre a construção do MAO propriamente dito; isto é, a rede neural se auto-organiza. Usando distintos processos, são calculadas as ativações (pesos sinápticos) dos neurônios da camada competitiva. Define-se, inicialmente, o número de épocas, as dimensões do mapa, entre outros parâmetros. Além disso, os vetores de entrada também são normalizados.

O melhor ajuste das dimensões do MAO e do vetor de entrada depende do estudo de caso analisado. Para grandes volumes de dados, geralmente, é mais indicado se utilizar mapas maiores. No entanto, mapas grandes podem comprometer o desempenho do algoritmo. Já mapas muito pequenos poderão comprometer a formação topológica da rede.

Na literatura, existem algumas propostas para se determinar a melhor dimensão dos mapas auto-organizáveis. Em (VESANTO; ALHONIEMI, 2000), por exemplo, foram propostas técnicas de análise e divisão dos dados de entrada em grupos para se encontrar uma dimensão para os MAO que minimizasse o custo de memória, tempo de processamento e ainda apresentasse resultados de mapeamento satisfatórios.

A etapa de treinamento dos MAO pode ser dividida, simplificada, em três fases ou processos principais (HUANG et al., 2005):

1. Competição: nessa fase procura-se encontrar o melhor casamento (minimizando a distância) entre os vetores de entrada e os vetores de pesos sinápticos do mapa (neurônios). O neurônio que apresentar o menor valor para o cálculo da distância é definido como neurônio vencedor;
2. Cooperação: o neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo assim a base para a cooperação entre os neurônios vizinhos;
3. Adaptação: os vetores de pesos sinápticos são atualizados de forma a se aproximar mais do vetor de entrada de acordo com alguma regra de atualização.

O processo de competição, dentre as três fases do treinamento, é a que apresenta o maior custo computacional; pois, nessa fase, é realizada uma busca seqüencial em todo o mapa para se localizar quais são os neurônios vencedores.

Fase de competição

Na fase de competição, procura-se encontrar o melhor casamento entre os vetores de entrada e de pesos sinápticos do mapa. Isso é realizado através da seleção do neurônio j que apresentar o maior valor para o produto interno $\boldsymbol{\chi}^T \boldsymbol{\xi}$. Assim, o local onde a vizinhança topológica dos neurônios excitados estará centrada é então determinado.

O melhor critério de correspondência, com base na maximização do produto interno, é matematicamente equivalente a minimizar a distância Euclidiana, D , se $\boldsymbol{\xi}$ e $\boldsymbol{\chi}$ estiverem ambos normalizados, através da seguinte expressão (KROSE; SMAGT, 1996):

$$D^{kj} = \|\boldsymbol{\xi}^k - \boldsymbol{\chi}^j\| = \sqrt{\sum_{n=1}^N (\xi_n^k - \chi_n^j)^2} \quad (133)$$

onde: $\boldsymbol{\xi}$ é o vetor de entradas do mapa; $\boldsymbol{\chi}$ é o vetor de pesos sinápticos; n, j, k e $N \in \mathbb{N}$ com $1 \leq n \leq N$, $1 \leq j \leq J$ e $1 \leq k \leq K$, onde $J \in \mathbb{N}$ é o número total de neurônios no mapa ($J = C_m L_m$) e K é o número de ciclos de amostragem.

O neurônio que apresentar o menor valor para o cálculo de D_{kj} é definido como neurônio vencedor, segundo a seguinte expressão:

$$\boldsymbol{\chi}^{j^*} = \underset{\forall j}{\operatorname{argmin}} \|\boldsymbol{\xi}^k - \boldsymbol{\chi}^j\| \quad (134)$$

onde: j^* é o índice do neurônio vencedor $\boldsymbol{\chi}$, “arg” significa o “índice” e \forall “para todos”.

A Eq. (134) resume a essência do processo de competição entre os neurônios no mapa. Dependendo da aplicação de interesse, a resposta da rede pode ser tanto o índice do neurônio vencedor (ou seja, a sua posição no mapa), quanto o vetor de pesos sinápticos do próprio neurônio vencedor.

Fase de cooperação

Na fase de cooperação, ocorre a atualização de uma vizinhança de neurônios em torno do neurônio vencedor. O neurônio vencedor irá determinar o centro de uma vizinhança topológica de neurônios mais próximos ao padrão de entrada que irão interagir entre si.

A atualização da vizinhança é realizada por meio de uma função de ativação conhecida como função de vizinhança, $h(n)$. Essa atualização é feita através de uma interação entre um agrupamento de neurônios excitados. O propósito de uma função de vizinhança é essencialmente correlacionar as atualizações dos pesos sinápticos de uma vizinhança de neurônios do mapa.

Deve-se fazer com que a excitação da vizinhança topológica, em torno do neurônio vencedor, decaia proporcionalmente à distância Euclidiana. Uma escolha típica, para a função que satisfaça tal exigência, é a função gaussiana:

$$h(j) = e^{\left(-\frac{(D^{kj^*})^2}{2\sigma^2}\right)} \quad (135)$$

onde: σ é a largura efetiva da vizinhança topológica e D^{kj^*} é a distância Euclidiana entre o vetor de entrada ξ^k e o vetor de pesos sinápticos χ^{j^*} do neurônio vencedor j^* .

Por meio da função de vizinhança $h(n)$, há uma interação entre o neurônio vencedor e a sua vizinhança. Quanto mais perto um determinado neurônio estiver do neurônio vencedor maior será essa interação. Deste modo, o MAO consegue preservar a topologia dos dados de entrada (LIU; WEISBERG; SHAY, 2007).

Durante o processo de aprendizagem, o valor do parâmetro σ , decai gradativamente. A redução do número de vizinhos é crucial para um correto ordenamento e convergência do mapa. Uma possível escolha para $\sigma(n)$ é dada por:

$$\sigma(n) = \sigma_0 e^{\left(\frac{-n}{\tau_1}\right)} \quad (136)$$

onde: τ_1 é uma constante de tempo associada ao decaimento de σ e σ_0 é o valor inicial σ .

Fase de adaptação

Após o neurônio vencedor ser determinado e a sua função de vizinhança ser atualizada, é preciso realizar um ajuste nos pesos sinápticos do neurônio vencedor assim como nos pesos dos neurônios vizinhos, de modo a orientar e atualizar os pesos sinápticos de todos os neurônios em direção ao vetor de entrada.

Após a apresentação dos dados de treinamento, os vetores de pesos sinápticos dos neurônios do mapa tendem a seguir a distribuição dos vetores de entrada, devido a atualização da sua vizinhança (KROSE; SMAGT, 1996).

Nesta fase, portanto, há uma ordenação topológica do mapa de modo a preservar as características dos dados de entrada, no sentido de que os neurônios adjacentes na rede tendem a ter pesos sinápticos semelhantes.

Assim, nessa fase, o vetor de pesos sinápticos $\boldsymbol{\chi}^j$ do neurônio j , é atualizado de forma a se aproximar mais do vetor $\boldsymbol{\xi}^k$, de acordo com a seguinte equação:

$$\boldsymbol{\chi}^{j(n+1)} = \boldsymbol{\chi}^{j(n)} + \alpha(n) h(j) \left[\boldsymbol{\xi}^k(n) - \boldsymbol{\chi}^{j(n)} \right] \quad (137)$$

onde: α é a taxa de aprendizado, $\alpha \in \mathbb{R}$, com $0 < \alpha < 1$ e h é a função de vizinhança (ou de ativação) que decresce de acordo com a distância, entre o neurônio vencedor e o vetor de entrada, e o tempo de treinamento (KOHONEN et al., 1995; RÜPING; RÜCKERT; GOSER, 1993; BARRETO, 2007).

A Eq. (137) é aplicada a todos os neurônios do mapa que se encontram dentro da vizinhança topológica do neurônio vencedor $\boldsymbol{\chi}^{j^*}$. Além disso, a taxa de aprendizagem, α , pode variar ao longo do tempo, como indicado na Eq. (137). Em especial, α pode iniciar com um valor inicial α_0 e ir gradualmente decaindo ao longo do tempo.

Assim, a taxa de aprendizado pode ser obtida a partir da seguinte expressão, por exemplo:

$$\alpha(n) = \alpha_0 e^{\left(\frac{-n}{\tau_2}\right)} \quad (138)$$

onde: α_0 é o valor inicial de α e τ_2 é uma constante de tempo associada ao decaimento de α .

Ao final da etapa de treinamento, os pesos sinápticos dos neurônios formam agrupamentos de neurônios em função da distância Euclidiana entre eles e do espaço de entrada. A idéia principal é usar os neurônios vencedores para visualizar a estrutura dos agrupamentos do MAO e as distâncias vetoriais entre os limites de cada agrupamento.

8.3.9.6 Etapa de Teste

Na etapa de teste ou recuperação, um novo vetor de entrada deve ser localizado no mapa, sendo automaticamente classificado e rotulado. Existe somente um neurônio vencedor para cada agrupamento de dados, cujo vetor de pesos sinápticos é o que mais se aproxima do vetor de entrada.

Inicialmente, dependendo de como cada vetor de entrada excita um neurônio específico no mapa, é realizada a atribuição dos rótulos aos agrupamentos formados pelo mapa. Como resultado, os neurônios no mapa são divididos em um número de regiões ou agrupamentos coerentes e distintos. Cada grupo distinto de neurônios é representado no mapa por rótulos diferentes.

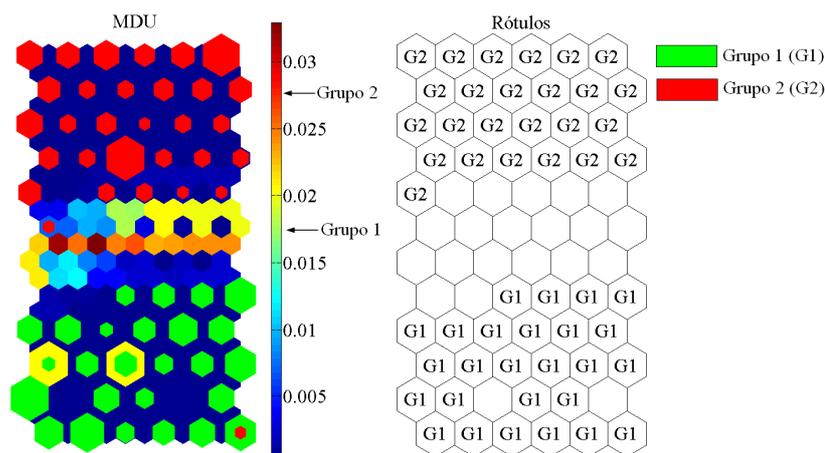


Figura 102: Exemplo de matriz de distâncias unificadas.

Após, semelhante à fase de competição, busca-se avaliar o melhor casamento entre o vetor de entrada e os vetores do mapa já treinado e rotulado com base na distância Euclidiana, como visto nas Eq. (133) e (134), determinando qual é o neurônio vencedor no mapa a cada novo vetor de entrada.

O resultado é a determinação da matriz de distâncias (Euclidianas) unificada (MDU). O resultado gerado a partir da MDU sobre o mapa é uma imagem em duas (ou três dimensões), onde o nível de intensidade de cada *pixel* é proporcional a uma distância Euclidiana previamente calculada, como visto na Fig. 102.

Para realizar o treinamento e gerar o MAO⁵ visto na Fig. 102, foram apresentados, inicialmente, um conjunto de dados com dois grupos distintos: G1 e G2. O primeiro resultado do MAO é um mapa contendo duas regiões com os rótulos dos respectivos grupos sendo determinados no mapa. Após, durante a etapa de teste, um novo conjunto de dados, contendo dados dos grupos 1 e 2, é apresentado novamente ao mapa. O segundo resultado é a MDU que exibe os neurônios do mapa classificados de acordo com o seu grupo e coloridos com diferentes cores, sendo cada cor referente ao seu respectivo agrupamento.

Na imagem, a coloração dos *pixels* varia de acordo com a intensidade de cada componente da MDU. Regiões que apresentam baixos valores (vales) agrupam neurônios com padrões similares (grupos G1 e G2). As regiões com valores altos (picos) correspondem a fronteiras entre os agrupamentos. A intensidade das componentes da MDU é obtida com base no cálculo da distância Euclidiana entre os dados de entrada e os dados do mapa já treinado.

O método de visualização dos MAO é uma ferramenta útil para descobrir estruturas gerais de um espaço de estados desconhecido e para verificar e monitorar as condições de comportamento de diferentes sistemas. As condições de comportamento normal, de degradação e de falha de um sistema podem ser visualizadas em distintos agrupamentos, por exemplo. Deste modo, é possível realizar uma detecção e uma classificação das distintas falhas de um sistema.

⁵Inicializado linearmente, com uma topologia hexagonal e uma gaussiana como função de vizinhança.

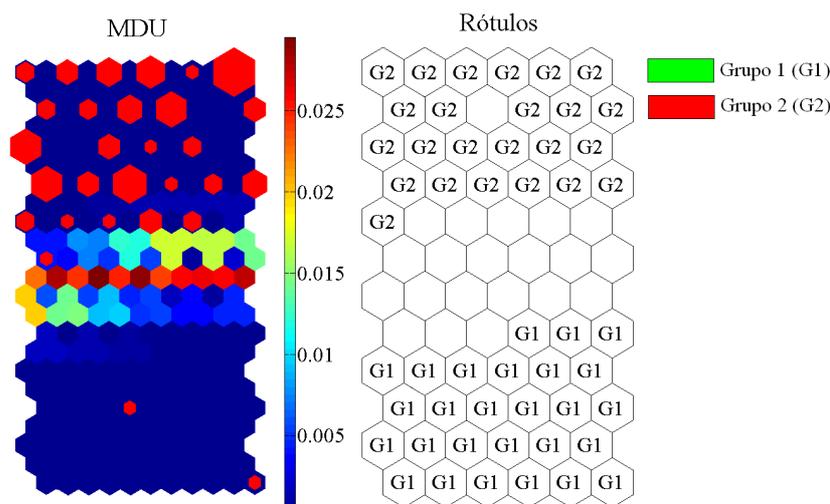


Figura 103: Exemplo de classificação de falhas obtido a partir dos MAO.

Supondo, por exemplo, que os conjuntos de dados G1 e G2 vistos anteriormente na Fig. 102 representem as condições normais e de falha (somente uma falha) de um sistema qualquer, respectivamente.

Para realizar uma classificação de falhas para esse sistema, utilizando os MAO, supondo somente uma falha neste caso, por exemplo, é necessário apresentar, inicialmente, dois conjuntos de dados ao MAO durante as etapas de inicialização e treinamento: um conjunto de dados que represente o comportamento normal (conjunto G1, neste caso) e um conjunto de dados que represente a respectiva falha (para este caso, o conjunto G2).

Supondo agora que durante a etapa de teste, por exemplo, sejam apresentados ao MAO somente novos dados de falha, dados pertencentes ao conjunto G2. O resultado da classificação de falhas (somente uma falha neste caso) pode ser visto na Fig. 103. Nessa figura, pode-se observar que estes novos dados de falha ocuparam somente (com algumas poucas exceções) a região mapeada anteriormente para o conjunto de dados G2, dados de falha desse sistema.

Supondo agora que um sistema qualquer possua mais de uma falha conhecida, n falhas neste caso. Para realizar uma classificação destas n falhas, usando os MAO, por exemplo, seria necessário apresentar ao mapa auto-organizável, durante as etapas de inicialização e treinamento, $n + 1$ conjuntos de dados: um conjunto de dados que represente o comportamento normal e n conjuntos de dados, cada um representando as respectivas n falhas.

O resultado, após as etapas de inicialização e treinamento, para esse novo conjunto de dados (com n falhas), seria um mapa auto-organizável com $n + 1$ regiões distintas, com os novos dados, durante a etapa de teste, sendo mapeados nestas diferentes regiões.

8.3.9.7 Avaliação do Mapa

Após o treinamento é importante avaliar a qualidade do mapa gerado. Um bom parâmetro para avaliar a qualidade do mapa gerado é o erro de quantização, definido como ϵ_q .

O erro de quantização corresponde a raiz quadrada da soma dos erros quadráticos, ou seja:

$$\varepsilon_q(k) = \sqrt{\sum_{n=1}^N (e_n^k)^2} \quad (139)$$

O erro (e_n^k) corresponde à diferença entre o vetor de pesos sinápticos do neurônio vencedor (χ^{j^*}) e novo dado de entrada (ξ^k), isto é:

$$e_n^k = \xi_n^k - \chi_n^{j^*} \quad (140)$$

onde: j^* é índice do neurônio vencedor no mapa.

A medida do erro de quantização revela a similaridade dos dados de entrada em relação ao conhecimento adquirido pelo mapa durante a etapa de treinamento. Quando se utiliza apenas um padrão de dados de entrada para treinar o mapa, por exemplo, o erro de quantização revela, durante a etapa de teste, a similaridade de novos dados de entrada em relação a esse mapa já treinado.

Um valor baixo para o erro de quantização significa que há uma grande similaridade dos dados de entrada com os dados treinados. Já um valor alto, representa um desvio das características que foram armazenadas, revelando que os dados novos não correspondem ao padrão de dados armazenado ou que o treinamento não foi satisfatório.

Isto é particularmente útil para se detectar falhas. Na medida em que um mapa é treinado somente com dados normais, ao se determinar o erro de quantização em relação a uma série de dados novos, uma degradação do comportamento, além da falha, poderá ser facilmente detectada.

Supondo novamente, por exemplo, que os conjuntos de dados G1 e G2 vistos na Fig. 102 representem as condições normais e de falha (somente uma falha) de um sistema, respectivamente. Para realizar a detecção de falhas durante a inicialização e treinamento deve ser apresentado ao MAO somente os dados que representem o comportamento normal; isto é, somente o conjunto de dados G1.

Agora, supondo que seja apresentado um novo conjunto de dados: 100 amostras que representam um comportamento normal, conjunto G1, e mais 100 amostras que representam o comportamento de falha, conjunto G2, durante a etapa de teste, nesta ordem, totalizando 200 amostras.

O resultado da detecção de falhas, nesse caso, e do erro de quantização, é visto na Fig. 104. Pode-se observar nesse gráfico que durante as 100 primeiras amostras, que representam o comportamento normal, o erro é praticamente nulo. A partir da amostra de número 100 o erro de quantização aumenta consideravelmente, e a falha é detectada. Deste modo, o erro de quantização, visto como exemplo na Fig. 104, pode ser usado para detectar as falhas desse suposto sistema.

Dentre as vantagens do uso do erro de quantização, pode citar: falhas podem ser detectadas em sua fase inicial, além de ser de fácil computação, por exemplo, (HUANG et al., 2005).

8.3.9.8 Resumo do algoritmo dos MAO

Os parâmetros essenciais para o uso do algoritmo dos mapas auto-organizáveis são os seguintes:

- Um espaço de dados de entrada contínuo gerado de acordo com uma determinada distribuição de probabilidades;
- Uma topologia de mapa auto-organizável na forma de uma rede de neurônios, que irá definir um espaço de saída discreto;
- Uma função de vizinhança, $h(n)$, variante ao longo do tempo definida em torno de um neurônio vencedor \boldsymbol{x}^{j^*} ;
- Uma taxa de aprendizagem $\alpha(n)$ que inicia de um valor α_0 e, em seguida, diminui gradualmente ao longo do tempo n sem, contudo, atingir um valor igual a zero.

Há três passos básicos envolvidos na aplicação do algoritmo dos mapas auto-organizáveis após a sua inicialização: amostragem, análise da similaridade e atualização (que correspondem as fases de competição, cooperação e adaptação realizadas durante a etapa de treinamento). Estes três passos são repetidos sucessivamente até que a formação do mapa esteja concluída (KROSE; SMAGT, 1996).

O resumo do algoritmo dos mapas auto-organizáveis é visto a seguir:

1. Inicialização: escolher valores aleatórios para os vetores de pesos sinápticos iniciais $\boldsymbol{x}^j(0)$. A única restrição é que os valores iniciais dos pesos sinápticos $\boldsymbol{x}^j(0)$ (com $j = 1, 2, \dots, J$) sejam diferentes.
2. Amostragem: escolher uma amostra, um vetor $\boldsymbol{\xi}$, dos dados de entrada. O vetor $\boldsymbol{\xi}$ deve representar o padrão de dados que será apresentado ao mapa;
3. Análise da similaridade: determinar o neurônio vencedor no instante de tempo n usando o menor valor para a distância Euclidiana como critério:

$$\boldsymbol{x}^{j^*} = \underset{j}{\operatorname{argmin}} \left\| \boldsymbol{\xi}^k - \boldsymbol{x}^j \right\| \quad (141)$$

Outras medidas de distância vetorial, tais como a distância de Hamming ou a distância de Manhattan (YAMAKAWA; HORIO; HIRATSUKA, 2002; PINO et al., 1999), também pode ser utilizadas.

4. Atualização: ajustar os pesos sinápticos de todos os neurônios:

$$\boldsymbol{x}^j(n+1) = \boldsymbol{x}^j(n) + \alpha(n) h(n) \left[\boldsymbol{\xi}^k(n) - \boldsymbol{x}^j(n) \right] \quad (142)$$

5. Continuação: retornar ao passo 2 até que nenhuma mudança topológica no mapa auto-organizável seja observada.

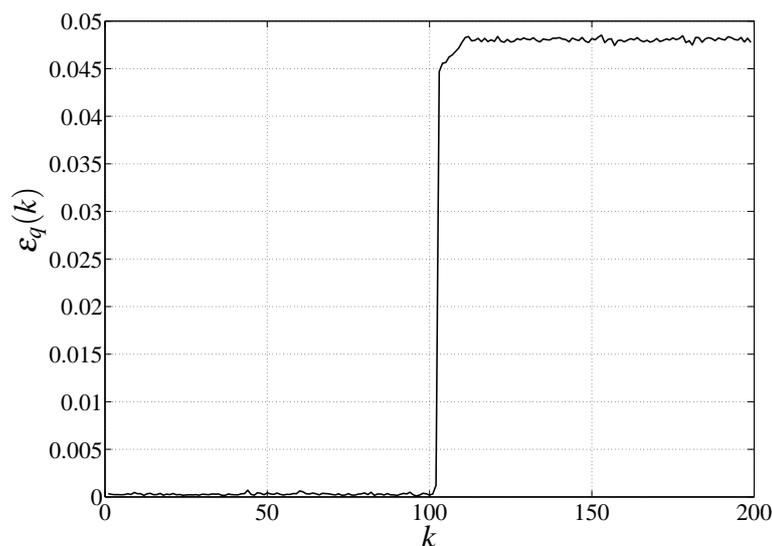


Figura 104: Exemplo de gráfico de erro de quantização.

8.3.10 Mapas Auto-organizáveis Temporais

Os mapas auto-organizáveis são voltados principalmente para as tarefas de classificação e detecção de certas características dos sistemas, tais como as suas falhas, como visto anteriormente na seção 8.3.9.

O conceito original dos mapas auto-organizáveis é baseado somente no ajuste estático de padrões, e a convergência não é garantida a não ser que as relações topológicas entre os diferentes padrões de entrada sejam estáveis.

Contudo, os padrões de entrada podem, além de estarem relacionados espacialmente, ocorrer sequencialmente. Neste caso, a evolução temporal do comportamento do sistema também pode ser analisada.

O comportamento dinâmico de um processo pode ser introduzido no mapa através do seu histórico de medições. Desse modo, o instante de tempo em que são observados os padrões desempenha um papel mais importante do que o grupo ao qual eles pertencem e pode ser levado em conta pelo MAO.

Chappell e Taylor (CHAPPEL; TAYLOR, 1993) foram os primeiros a modificar o MAO original usando um mecanismo de retenção conhecido como memória de curto prazo. Eles propuseram um modelo, chamado de mapa temporal de Kohonen (MTK) para a classificação de uma seqüência de dados.

O mapa temporal de Kohonen é uma rede neural, com um algoritmo de aprendizado não-supervisionado, voltada para a predição de séries temporais, derivada dos mapas auto-organizáveis.

Nos mapas temporais de Kohonen, ou mapas auto-organizáveis temporais, a mudança entre os estados de operação pode ser descrita por uma trajetória determinada em função dos neurônios vencedores.

Em um modo de operação normal, os vencedores devem seguir uma trajetória dentro da região do modo normal. Quando ocorrer uma falha, o vencedor deverá desviar da região normal. O desvio dependerá do tipo e da gravidade da falha.

A migração da trajetória da falha possui informações importantes sobre o modo de falha. Assim, usando dados normais, de degradação e de falha para treinar o mapa temporal, não só o início da degradação pode ser detectado, mas também uma predição dos possíveis modos de falha pode ser realizada, com base nessa trajetória.

Desse modo, os mapas temporais de Kohonen podem ser utilizados para modelar uma seqüência de estados de um sistema em funcionamento (uma série temporal) e assim prever o estado futuro desse sistema.

8.3.10.1 Características Fundamentais

Os mapas temporais de Kohonen, de modo semelhante aos mapas auto-organizáveis, utilizam um treinamento de aprendizado não-supervisionado. A rede neural procurar levantar características similares levando em conta apenas os padrões de entrada, sem possuir algum conhecimento, à priori, dos dados que estão sendo repassados.

8.3.10.2 Objetivos

O principal objetivo dos MTK é transformar um padrão de entrada, um sinal de dimensão arbitrária, em um mapa discreto de uma ou duas dimensões, de modo semelhante aos MAO, preservando a ordem topológica e cronológica dos dados de entrada.

Outro objetivo é a formação de trajetórias em função dos neurônios vencedores. Isso permite, após o treinamento, gerar uma visualização da trajetória do comportamento dinâmico do sistema ao longo do tempo.

8.3.10.3 Etapas

As etapas de inicialização, treinamento e teste dos mapas temporais de Kohonen (exceto pela determinação dos neurônios vencedores, na fase de competição) são as mesmas etapas dos algoritmos dos mapas auto-organizáveis tradicionais, apresentadas anteriormente.

Nos mapas temporais de Kohonen, um histórico das ativações (durante a etapa de teste) de cada neurônio é armazenado por meio de uma variável definida como ativação temporal do neurônio, A^{kj} . Supondo o vetor de ativação da amostra de número k sendo definido como $\mathbf{A}^k = [A^{k1} A^{k2} \dots A^{kJ}]$, cada componente deste vetor é dada por:

$$A^{kj} = \lambda A^{k-1,j} - \frac{1}{2} \left(D^{kj} \right)^2 \quad (143)$$

onde: D^{kj} é a distância Euclidiana, Eq. (133), λ é o tamanho da memória, $1 \leq j \leq J$ e $1 \leq k \leq K$, sendo $J \in \mathbb{N}$ o número de neurônios no mapa e $K \in \mathbb{N}$ o número de ciclos de amostragem.

Nos MAO, os neurônios vencedores são determinados através do cálculo da distância Euclidiana, Eq. (133): o neurônio que apresentar o menor valor para a distância é escolhido como vencedor, como visto na Eq. (134).

Já nos MTK, o neurônio que possuir o maior valor para o cálculo da ativação temporal, é escolhido como neurônio vencedor. O neurônio vencedor, ou a posição desse no mapa, neste caso, é dado de acordo com a seguinte expressão:

$$\chi^{j^*} = \operatorname{argmax}_{\forall j} (A^{kj}) \quad (144)$$

onde: j^* é o índice do neurônio vencedor χ do MTK e “arg” significa o “índice”.

A Eq. (143) preserva um histórico das ativações anteriores como uma soma ponderada. Os valores de ativação são calculados através de uma somatório recursivo (também baseado no cálculo da distância Euclidiana). Os neurônios vencedores, obtidos para cada nova amostra k , irão determinar a trajetória do comportamento do sistema no mapa ao longo do tempo.

Durante a etapa de teste, um novo vetor de entrada (ou uma série de novos vetores) é apresentado ao MTK. Um novo neurônio vencedor deve ser escolhido para esse vetor, ou para essa série de vetores. A posição desse neurônio vencedor, cujo vetor de pesos sinápticos é o que mais se aproxima do vetor de entrada, é então localizada e assinalada, no mapa. Como resultado, a trajetória dos neurônios vencedores ao longo do tempo vai sendo traçada no mapa, como visto na Fig. 105.

Para realizar o treinamento e gerar o MTK ⁶ visto na Fig. 105, foram apresentados ao mapa, inicialmente, três grupos de dados distintos definidos como: normal, degradação e falha. O primeiro resultado do MTK (semelhante aos MAO) é um mapa contendo três regiões com os rótulos dos respectivos grupos sendo determinados nesse mapa.

Após, durante a etapa de teste, um novo conjunto de dados, contendo dados dos três grupos, é apresentado sequencialmente ao mapa (inicialmente, dados normais; em seguida, dados de degradação; e finalmente, os dados de falha) novamente ao mapa. O segundo resultado é a MDU que exhibe os neurônios do mapa, coloridos com diferentes escalas de cinza. A trajetória dos neurônios vencedores (com três cores distintas, uma para cada grupo) é então traçada no mapa na medida em que os novos dados vão sendo apresentados ao mapa.

Assim, a evolução do comportamento de um sistema pode ser visualizada no mapa temporal de Kohonen. Desse modo, é possível realizar também uma predição de falhas de um sistema. Vale ressaltar, que os MTK são fundamentalmente utilizados como ferramenta de visualização. Os estados do processo podem ser visualizados como uma trajetória no mapa, sendo possível acompanhar a dinâmica do processo ao longo do tempo.

8.3.10.4 *Resumo do algoritmo dos MTK*

Os parâmetros essenciais para o uso do algoritmo dos mapas temporais de Kohonen são os mesmos utilizados no algoritmo dos mapas auto-organizáveis, como visto na subseção 8.3.9.8.

⁶Inicializado linearmente, com uma topologia hexagonal e uma gaussiana como função de vizinhança.

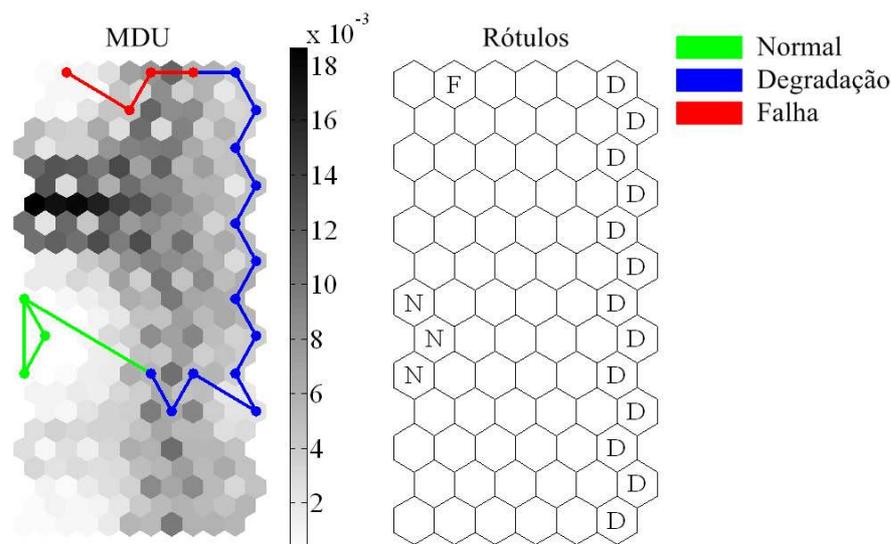


Figura 105: Exemplo de mapa auto-organizável temporal.

Os três passos básicos envolvidos na aplicação do algoritmo dos mapas temporais de Kohonen são os mesmos dos mapas auto-organizáveis após a sua inicialização: amostragem, análise da similaridade e atualização. Estes três passos são repetidos de modo sucessivo até que a formação do mapa esteja concluída:

O resumo do algoritmo dos mapas temporais de Kohonen é visto a seguir:

1. Inicialização: definir valores aleatórios para os vetores de pesos sinápticos.
2. Amostragem: escolher uma amostra, um vetor ξ , dos dados de entrada. O vetor ξ deve representar o padrão de dados que será apresentado ao mapa;
3. Análise da similaridade: determinar o neurônio vencedor usando o maior valor para a ativação temporal, A^{kj} , Eq. (143), como critério:

$$\chi^{j*} = \operatorname{argmax}_{\chi^j} (A^{kj}) \quad (145)$$

4. Atualização: ajustar os pesos sinápticos de todos os neurônios:

$$\chi^{j(n+1)} = \chi^{j(n)} + \alpha(n) h(n) [\xi^{k(n)} - \chi^{j(n)}] \quad (146)$$

5. Continuação: retornar ao passo 2 até que nenhuma mudança topológica seja observada.

8.3.11 Outros tipos de Mapas Auto-organizáveis

Além do mapa auto-organizável e do mapa temporal de Kohonen, existem outros tipos de mapas auto-organizáveis, tais como: *Time-Delayed Self-Organizing Map*, *Recurrent Self-Organizing Map* e *Contextual Self-Organizing Map* (BASTOS, 2007), com diferentes algoritmos, voltados para diversas aplicações, tais como: a classificação de padrões ou a predição de séries temporais, por exemplo.

8.3.12 Vantagens das Redes Neurais Artificiais

Pode-se citar como exemplos de vantagens das redes neurais artificiais os seguintes aspectos: podem aprender sobre a condição normal de funcionamento de um sistema e determinar se os sinais estão significativamente diferentes dessa condição; são úteis quando há uma grande quantidade de dados disponíveis; podem se adaptar e aprender sobre processos dinâmicos e não-lineares.

8.3.12.1 Vantagens dos Mapas Auto-organizáveis

Resumidamente, cita-se como exemplos de vantagens dos MAO os seguintes aspectos: a capacidade de detecção, classificação e predição de falhas de sistemas dinâmicos e não-lineares. Assim, dentre as principais aplicações dos MAO, pode-se destacar a análise de dados, o reconhecimento e classificação de padrões, e a predição de falhas, especialmente aquelas que são desenvolvidas para sistemas em tempo real.

Com relação aos modelos de previsão de séries temporais, baseados nos MAO, que possuem um paradigma de aprendizagem não supervisionado, estes pertencem a uma classe de modelos capazes de executar a aproximação de uma função localmente ⁷. Um modelo de previsão de séries temporais global, como as redes PMC por exemplo, faz uso de representações do espaço de entrada altamente distribuídas, o que torna difícil a interpretação dos resultados.

Deste modo, os modelos de predição baseados no MAO, os MTK por exemplo, permitem ao usuário compreender melhor a dinâmica do processo que gera as séries temporais, ao menos dentro da região local usada para calcular a saída do modelo atual. Essa natureza do MTK aliada com a preservação da topologia é útil, por exemplo, na segmentação ou visualização de séries temporais (BARRETO, 2007).

8.3.13 Desvantagens das Redes Neurais Artificiais

Como desvantagem das redes neurais artificiais, pode-se citar: os dados de treinamento devem ser representativos, de modo que possa haver sucesso na fase de treinamento e o treinamento, em alguns casos, pode ser demorado.

8.3.13.1 Desvantagens dos Mapas Auto-organizáveis

Como exemplo de desvantagem dos MAO, pode-se citar: os dados de treinamento devem ser escolhidos corretamente, de modo a representar adequadamente os padrões de comportamento que se deseja identificar, por exemplo. Além disso, novos dados de treinamento precisam ser acrescentados ao mapa, quando um novo tipo de falha ocorre, e o mapa precisa ser novamente treinado utilizando o novo conjunto de dados (com os novos dados acrescentados).

⁷Por local, denotam-se os modelos cujas saídas são determinadas através de operações matemáticas agindo em regiões localizadas do espaço de entrada.

Embora os sinais de algumas falhas possam ser obtidos, é muito difícil compor um conjunto completo de dados de treinamento que represente as características de todas as falhas do sistema; pois, sempre haverá uma falha que ainda não foi detectada e mapeada.

O treinamento também, em função do número de amostras, dimensão do vetor de entradas, dimensões do mapa, e outros fatores, pode ser lento. Uma boa escolha destes parâmetros irá resultar em um menor custo computacional.

8.4 CONCLUSÕES DO CAPÍTULO

Neste capítulo, uma descrição sucinta sobre a inteligência artificial e sobre as suas principais técnicas (tais como: os algoritmos genéticos, a programação evolutiva e as cadeias de Markov) foi apresentada.

As redes neurais artificiais e os mapas auto-organizáveis, duas técnicas de inteligência artificial em especial, foram abordados em mais detalhes. As suas principais etapas, características, modelos e aplicações, por exemplo, foram descritas.

Constatou-se, ao longo da pesquisa bibliográfica, que as técnicas de inteligência artificial estão ampliando os sistemas de controle, detecção e diagnóstico de falhas e tornando as atividades e os sistemas de manutenção mais inteligentes.

Os mapas auto-organizáveis, em específico, vêm sendo largamente aplicados na análise de dados, no reconhecimento e classificação de padrões e na predição de falhas, por exemplo.

9 ANEXO C: FILTRAGEM ADAPTATIVA

O termo filtro é geralmente usado para se referir a um sistema que é projetado para separar, extrair ou prever certas informações sobre uma quantidade prescrita ou um sinal de interesse que geralmente possui um ruído inerente (HAYKIN, 2002).

Filtros são utilizados, basicamente, para separar elementos distintos. Quando esses elementos são sinais elétricos, por exemplo, os filtros são responsáveis pela alteração do conteúdo de frequências destes sinais. Os filtros são usados, neste caso, para restringir a passagem de certas bandas de frequência específicas ou mesmo para remover ruídos ou interferências do sinal original. Os filtros também podem ser utilizados para realizar a interpolação e até mesmo a identificação e a predição de sinais, por exemplo.

Uma das ferramentas de estimação, ou predição, mais utilizadas, em inúmeras aplicações voltadas para aplicações em tempo real, em especial, são os filtros adaptativos (FA). Os filtros adaptativos são estimadores ótimos recursivos para modelos lineares estocásticos.

Um filtro adaptativo é um tipo de sistema adaptativo que ajusta o seu comportamento, isto é, a sua resposta ao impulso, automaticamente. Este ajuste das características do filtro é feito com base num critério de desempenho pré-estabelecido pelo projetista. Este critério de desempenho define uma função de custo que será minimizada através de um processo de ajuste de parâmetros para alcançar o melhor desempenho possível.

Os filtros adaptativos são voltados para situações onde não se conhece ou não se pode prever as características do ambiente de operação do filtro. Através do algoritmo de adaptação mais adequado, o filtro se ajusta automaticamente ao ambiente, reconhecendo as suas características. A capacidade de acompanhamento das mudanças do meio através do filtro é definida pelo algoritmo e processos relacionados. O tipo de estrutura do filtro e seu algoritmo são determinados basicamente pelo tipo de aplicação em que este será usado.

Os FA lineares, construídos em torno de um combinador linear (que pode ser visto como um único neurônio operando em seu modo linear), são projetados para realizar uma aprendizagem contínua. Em virtude da sua estrutura simples, os FA são usados amplamente em aplicações tão distintas como as voltadas para a medicina ou economia (HAYKIN, 2002).

Inicialmente, neste capítulo, uma breve descrição dos principais tipos de filtros (determinísticos e ótimos) é apresentada na seção 9.1. A seguir, na seção 9.2, os filtros de Wiener são abordados. Os FA são analisados na seção 9.3. Por fim, as principais vantagens, desvantagens e aplicações dos FA e as conclusões são vistas nas seções 9.4, 9.5, 9.6 e 9.7, respectivamente.

9.1 TIPOS DE FILTROS

Os sinais elétricos, mecânicos ou sonoros, por exemplo, estão sujeitos a inúmeros tipos de ruídos que prejudicam a sua correta interpretação e utilização. Esses ruídos podem ser atenuados ou eliminados através do uso de filtros, melhorando a qualidade do sinal.

Segundo (STEMMER et al., 2000), um filtro é uma rede que transforma um sinal de entrada, de um determinado modo, no sinal de saída desejado. Os sinais podem ser considerados no domínio temporal ou no domínio espectral, de maneira similar, os requisitos de saída do filtro podem ser gerados em termos temporais ou espectrais.

Já segundo (HAYKIN, 2002), filtro é um dispositivo que é aplicado a um conjunto de dados com ruído a fim de extrair alguma informação sobre uma dada quantidade de interesse.

Existem três operações, ou problemas de extração, básicas que podem ser realizadas com os filtros adaptativos:

1. Filtragem: é a extração da informação sobre a quantidade de interesse no instante t , usando a informação obtida até o instante t , inclusive;
2. Interpolação: é um método que permite construir um novo conjunto de dados a partir de um conjunto de dados previamente conhecidos. A informação sobre a quantidade de interesse não precisa estar disponível no instante t , e os dados obtidos após este instante podem ser utilizados para obter esta informação;
3. Predição: é a determinação aproximada da quantidade de interesse num instante $t + \tau$ no futuro, para $\tau > 0$, usando a informação medida até o instante t , inclusive.

Os filtros podem ser divididos, simplificadamente, em duas categorias: filtros determinísticos (tais como os filtros passa-altas e passa-baixas, por exemplo) e filtros ótimos (como os filtros de Wiener ou como os filtros adaptativos, como exemplo).

Os filtros determinísticos são usados para selecionar ou rejeitar uma ou várias bandas de frequência de um determinado sinal determinístico ¹.

Já os filtros ótimos também podem ser usados, por exemplo, em outras tarefas tais como a detecção ou predição de sinais aleatórios ².

Neste trabalho estar-se-á interessado principalmente na análise dos filtros ótimos, sendo apresentado somente uma breve descrição dos principais filtros determinísticos. Dentre os filtros ótimos, destacam-se os filtros adaptativos.

Os filtros adaptativos podem ser encontrados em diversos tipos de aplicações: remoção de ruídos, identificação de sistemas, cancelamento de eco, entre outros exemplos.

Em (NASCIMENTO; PROVASE, 2003), por exemplo, foi investigado o uso de filtros adaptativos na redução de distorções harmônicas não-lineares. Neste trabalho foi apresentado um sistema para corrigir não-linearidades em amplificadores transistorizados.

¹Um sinal é dito determinístico se ele é exatamente previsível para o intervalo de tempo de interesse e pode ser descrito por uma função matemática. Contudo, o conhecimento da função matemática não é necessário para que o sinal seja dito como determinístico; basta que se saiba conceitualmente que tal função existe.

²Um sinal aleatório sempre possui um elemento de incerteza associado e, portanto, não é previsível no sentido determinístico e é descrito em termos probabilísticos.

9.1.1 Filtros Determinísticos

Os filtros determinísticos, ou filtros seletivos de frequência, são conhecidos como filtros fixos; isto é, seus coeficientes não variam ao longo do tempo.

Dentre as principais características desses filtros, pode-se citar: os filtros são lineares e invariantes no tempo; durante o projeto do filtro são utilizadas informações sobre a banda de frequência em que o filtro irá atuar; e, os filtros apresentam um melhor desempenho quando as várias componentes espectrais do sinal ocupam bandas não sobrepostas.

Os principais tipos de filtros determinísticos são:

- Passa-baixas: rejeita a passagem das componentes de um sinal a partir de certo valor;
- Passa-altas: permite a passagem das componentes de um sinal a partir de um valor;
- Passa-banda: permite a passagem de componentes de uma faixa de frequências;
- Rejeita-banda: rejeita as componentes de uma faixa de frequências.

Aplicações que envolvem sensores em circuitos, como as aplicações médicas, circuitos de áudio, sensores industriais e telecomunicações podem exigir em certos momentos o uso de filtros determinísticos quer seja para rejeitar ou permitir a passagem de bandas de frequência.

9.1.2 Filtros Ótimos

Quando as características estatísticas (tais como a média, variância ou covariância) de um sinal a ser processado conferem com a informação prévia utilizada para projetar o filtro, o filtro é dito ótimo.

Ao se projetar um filtro ótimo, um filtro de Wiener, por exemplo, é necessário ter um conhecimento prévio das características do sinal a ser processado. Na maioria dos casos, não é possível projetar o filtro de Wiener, ou o projeto pode não ser robusto, quando estas informações são conhecidas parcialmente. Nesta situação, pode-se usar uma outra abordagem onde uma estimativa dos parâmetros estatísticos é utilizada.

Diz-se que um filtro é linear, caso a quantidade de interesse, presente na saída do filtro, seja uma função linear do sinal aplicado na entrada do filtro. Para simplificar a análise dos filtros ótimos, o filtro deve ser linear e causal para que seja possível a sua implementação.

O principal objetivo, ao se projetar um filtro ótimo, é que este forneça uma saída, \mathbf{y} , o mais próximo possível da resposta desejada, \mathbf{d} , quando sujeito a uma entrada \mathbf{u} , como visto na Fig. 106. Para tal, busca-se minimizar uma função que possui o erro de estimação, \mathbf{e} , como parâmetro, conhecida como função de custo.

Esta função de custo (ou superfície de performance), que é função do erro (diferença entre o sinal desejado e a saída atual do filtro) da estimativa feita pelo filtro, também pode ser usada como critério de avaliação do filtro. Esta função de custo pode ser definida a partir de métodos probabilísticos ou do método dos mínimos quadrados.

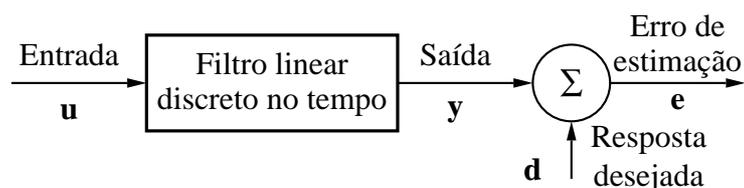


Figura 106: Representação do problema de filtragem ótima linear.

Na solução pelo método dos mínimos quadrados, busca-se encontrar a melhor estimativa para a resposta desejada, considerando como função de custo a soma dos quadrados do sinal de erro e considerando que é conhecido um número finito de amostras do sinal de entrada e dos demais sinais envolvidos no processo.

Já na solução por métodos probabilísticos, ou solução de Wiener, a função de custo é assumida como a média do quadrado do sinal de erro e as propriedades estatísticas dos sinais são conhecidas para todo o tempo e não somente em um intervalo finito.

Ainda, a questão da estimação linear, em relação ao sinal e ao ruído, do ponto de vista probabilístico, pode ser vista de dois modos: a sua covariância é dada sobre sinal e ruído (filtros de Wiener) ou o sinal e o ruído são descritos por modelos de espaço de estados (filtros de Kalman).

O filtro de Wiener é mais adequado para trabalhar com casos onde os sinais de entrada são estacionários. Já nas situações onde a não-estacionaridade do sinal de interesse ou do ruído é intrínseca ao problema, o filtro de Kalman é o mais sugerido. O filtro de Kalman é um estimador ótimo recursivo para modelos lineares estocásticos em espaço de estados (KALMAN, 1960). Neste trabalho, somente o filtro de Wiener será apresentado.

9.2 FILTROS DE WIENER

Os estudos sobre a estimação da mínima média dos quadrados em processos estocásticos iniciaram nos anos 30 e 40 com contribuições feitas por Komolgorov, Krein e Wiener. Wiener equacionou a questão da predição linear para tempo contínuo e obteve uma equação para o preditor ótimo; além de analisar a estimação de um processo corrompido por um processo ruidoso aditivo. Para se obter a equação da estimativa ótima é necessário resolver uma equação integral conhecida como *Equação de Wiener-Hopf*. A questão da filtragem de Wiener em tempo discreto e formulação matricial da *Equação de Wiener-Hopf* foram propostas por Levinson em 1947.

Ao se trabalhar com a questão da filtragem ótima linear, Fig 106, algumas restrições iniciais são necessárias para facilitar a implementação digital do filtro de Wiener, tais como o uso de um filtro causal de estrutura transversal e de resposta ao impulso de duração finita. Na Fig. 107 pode-se observar a estrutura de filtro transversal onde as entradas, u , e os coeficientes, w , são utilizados para se obter a saída, y , do filtro.

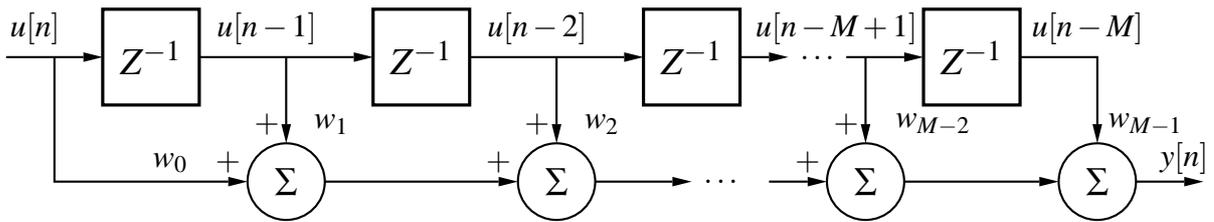


Figura 107: Filtro causal de estrutura transversal.

Também é necessário definir que a entrada do filtro e a resposta desejada são realizações únicas de processos estocásticos e discretos no tempo, ambos com média zero.

A idéia principal é obter o menor valor possível para o erro de estimação levando em conta um critério estatístico, que neste caso é igual à média do quadrado do erro de estimação.

Então, a principal questão ao se utilizar o filtro de Wiener é: dado um conjunto de amostras de entrada, como projetar um filtro linear causal discreto no tempo, de estrutura transversal, cuja saída forneça uma estimativa da resposta desejada, tal que o valor da média do quadrado do erro de estimação (que é a diferença entre a resposta desejada e a resposta atual) seja minimizado.

9.2.1 Método Probabilístico ou Solução de Wiener

A saída do filtro de *Wiener*, assumindo que a série temporal e que os M coeficientes do filtro podem ter valores complexos, é obtida a partir de:

$$y[n] = \sum_{k=0}^{M-1} w_k u[n-k] \quad (147)$$

onde: n, k e $M \in \mathbb{N}$.

A função de custo, J , definida como o valor médio do quadrado do erro, $e[n]$, vista a seguir, onde ε é o valor esperado, é minimizada para otimizar o projeto do filtro:

$$J = \varepsilon \{ |e[n]|^2 \} \quad (148)$$

Na expressão anterior, o erro de estimação é dado por:

$$e[n] = d[n] - y[n] \quad (149)$$

Definindo então o vetor de coeficientes do filtro transversal, \mathbf{w}^3 , como:

$$\mathbf{w} = [w_0 \ w_1 \ \dots \ w_{M-1}]^T \quad (150)$$

Definindo também o sinal de entrada, \mathbf{u} , como sendo:

$$\mathbf{u} = [u[n] \ u[n-1] \ \dots \ u[n-M+1]]^T \quad (151)$$

³ T representa a operação de transposição do vetor.

A expressão do sinal de saída do filtro na sua formulação matricial pode ser então obtida:

$$y[n] = \mathbf{w}^T \mathbf{u} \quad (152)$$

Definindo agora a matriz de correlação do sinal de entrada do filtro transversal:

$$R = \varepsilon \{ \mathbf{u} \mathbf{u}^H \} \quad (153)$$

e definindo, também, o vetor de correlação cruzada, \mathbf{p} , como:

$$\mathbf{p} = \varepsilon \{ \mathbf{u} d^* \} \quad (154)$$

onde: H representa as operações de transposição e complexo conjugado.

Pode-se então obter a expressão que determina os coeficientes ótimos do filtro transversal, \mathbf{w}_O^* , assumindo-se que a matriz de autocorrelação, R , é não singular e conhecida, como visto a seguir:

$$\mathbf{w}_O^* = R^{-1} \mathbf{p} \quad (155)$$

e pode-se obter também a expressão para o mínimo valor da função de custo, J , como:

$$J_{min} = \sigma_d^2 - \mathbf{p}^H R^{-1} \mathbf{p} \quad (156)$$

onde: σ_d é a variância do sinal desejado.

A Eq. (148) pode então ser reescrita na sua forma matricial (HAYKIN, 2002):

$$J = J_{min} + (\mathbf{w} - \mathbf{w}_O)^T R (\mathbf{w} - \mathbf{w}_O)^* \quad (157)$$

ou:

$$J = \sigma_d^2 - \mathbf{w}^T \mathbf{p} - \mathbf{p}^H \mathbf{w}^* + \mathbf{w}^T R \mathbf{w}^* \quad (158)$$

onde: o asterisco (*) representa o complexo conjugado do vetor.

9.2.2 Método dos Mínimos Quadrados

Para solucionar a questão de filtragem linear ótima sem fazer o uso de ferramentas estatísticas, utiliza-se o método dos mínimos quadrados (MQ). Este método pode ser visto como uma alternativa para a solução do problema de filtragem linear ótima. Neste método, as médias temporais são utilizadas, ao contrário do filtro de Wiener que faz uso de médias probabilísticas.

Supondo, inicialmente, que foram realizadas um conjunto de N medidas para o vetor de entrada, \mathbf{u}_n , nos respectivos instantes de tempo, t_n . O método dos mínimos quadrados consiste em obter uma curva que chegue o mais próximo possível dos dados de entrada. Quando a soma dos quadrados da diferença entre os valores estimados e valores medidos ao longo de todo o tempo é minimizada, o método atinge o seu ponto ótimo.

A Fig. 107 exhibe o modelo de filtro transversal usado no método dos MQ. A estrutura é a mesma utilizada para a solução do problema de filtragem ótima linear via filtro de Wiener, vista anteriormente.

Contudo, os sinais, neste caso, não fazem parte de um processo estocástico. O sinal $\mathbf{d}[n]$ é assumido como sendo obtido por um modelo de regressão linear múltipla, com um erro (um ruído branco de variância e média iguais a zero) não observável; isto é, um processo estocástico discreto no tempo (HAYKIN, 2002).

A idéia básica do filtro é estimar os parâmetros, \mathbf{w} , do modelo com base no conjunto de variáveis observáveis, $u[n]$ e $d[n]$, para $n \in \mathbb{N}^*$. O erro de estimação, $e[n]$, ou resíduo neste caso é dado por:

$$e[n] = d[n] - y[n] \quad (159)$$

onde: y é a saída do FA.

A saída do filtro é dada por:

$$y[n] = \sum_{k=0}^{M-1} w_k u[n-k] \quad (160)$$

e a função de custo, neste caso, a ser minimizada é a soma dos quadrados dos erros:

$$J = \sum_{n=n_1}^{n_2} |e[n]|^2 \quad (161)$$

onde: n_1 e n_2 definem os limites onde a minimização do erro ocorre.

Existem dois modos para descrever a condição ótima do filtro de mínimos quadrados: o princípio da ortogonalidade e o sistema de equações normais. Uma descrição mais detalhada destes dois métodos pode ser encontrada em (HAYKIN, 2002).

9.3 FILTROS ADAPTATIVOS

Outra solução, ou abordagem, para a questão de filtragem linear é fazer uso de um filtro adaptativo. O filtro adaptativo é um dispositivo que ajusta sua própria resposta ao impulso. Os filtros adaptativos não necessitam de qualquer informação estatística a respeito dos dados a serem processados; isto é, o funcionamento dos filtros adaptativos é baseado em algoritmos recursivos. Isto torna possível que o filtro tenha um desempenho satisfatório em um ambiente onde as características dos sinais não estão disponíveis.

Um sistema adaptativo (um filtro adaptativo, por exemplo) é aquele cuja estrutura é alterável (através do ajuste dos seus coeficientes, por exemplo) de tal modo que seu comportamento se ajuste, de acordo com algum critério de desempenho previamente estabelecido, ao ambiente ao qual ele será inserido.

O ajuste do filtro adaptativo, neste caso, é realizado através da programação de algoritmos adaptativos específicos, devidamente escolhidos, que têm como principal objetivo atender aos requisitos do sistema.

A partir de um conjunto de condições iniciais predeterminadas o algoritmo é inicializado. Após sucessivas iterações do algoritmo, num ambiente estacionário, ocorre a convergência para a solução ótima de Wiener. Já, em um ambiente não-estacionário, é possível seguir as variações do comportamento estatístico do sinal de entrada, desde que estas variações sejam lentas o suficiente (HAYKIN, 2002).

Assim, quando há exigências para processar sinais resultantes de operações em ambientes estatísticos desconhecidos, ou quando o processo é não estacionário e apresenta pequenas e lentas variações do comportamento, o uso dos FA é uma solução atrativa, bem como apresenta um melhor desempenho, quando comparado aos métodos que usam filtros determinísticos.

A questão básica ao se estudar os filtros adaptativos, que é diferente do problema dos filtros determinísticos, é como determinar a estimativa do sinal de um sistema corrompido por um ruído. Isto é, deseja-se projetar um filtro específico para recuperar um sinal ao qual foi adicionado um ruído.

Simplificadamente, a operação de um algoritmo de filtragem adaptativo pode ser dividida em dois processos: filtragem (destinado a produzir uma saída em resposta a uma seqüência de dados de entrada) e adaptação (cujo objetivo é realizar um controle adaptativo de um conjunto de parâmetros ajustáveis utilizados no processo de filtragem). Estes dois processos trabalham de modo iterativo um com o outro. Cabe destacar que a estrutura do processo de filtragem altera significativamente a operação e desempenho do algoritmo.

Não há uma única solução para o problema de filtragem adaptativa, assim como há diversos algoritmos, com distintas características e diferentes desempenhos, tais como os mínimos quadrados médio e o mínimos quadrados recursivo, por exemplo (WIDROW; STEARNS, 1985; HAYKIN, 2002).

Neste trabalho, estar-se-á particularmente interessado na aplicação e descrição do método da descida mais íngreme e do algoritmo dos mínimos quadrados médio. Há outros algoritmos que também podem ser citados, tais como: *random search*, mínimos quadrados recursivo, que não serão analisados.

Alguns conceitos importantes para a análise de algoritmos para filtros adaptativos, relacionados à análise da estabilidade, tais como: curva de aprendizagem e tempo de estabilização do algoritmo serão analisados. Contudo, há outras questões relativas aos filtros adaptativos, tais como: a análise para ambientes não-estacionários e outros algoritmos relacionados que não serão abordados neste trabalho.

9.3.1 Estrutura dos Filtros Adaptativos

Um diagrama genérico de como um filtro adaptativo é estruturado é visto na Fig. 108. Geralmente, o bloco “Qualidade da Resposta do Filtro” é formado pela simples subtração entre o sinal de saída do filtro, \mathbf{y} , e o valor desejado, \mathbf{d} , que dependendo do algoritmo pode ser dispensável, tendo o sinal de erro, \mathbf{e} , como resultado.

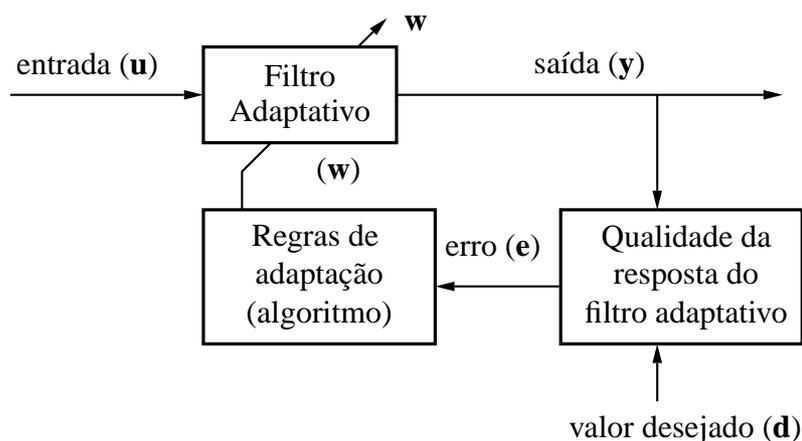


Figura 108: Estrutura geral de um filtro adaptativo.

O filtro adaptativo é o responsável por processar o sinal de entrada, u , fazendo uso de regras de adaptação, ajustando os parâmetros do filtro, w , e levando em conta o erro obtido, e .

A partir da Fig. 108, pode-se observar que o objetivo final do FA é estimar seu valor de saída, aproximando ao máximo este resultado com o valor desejado, até que o valor do erro seja nulo ou próximo de zero. A realimentação do sinal de erro conduz o sistema a esta adaptação.

9.3.2 Características dos Filtros Adaptativos

Dentre as principais características dos filtros adaptativos, pode-se citar (WIDROW; WALLACH, 1995; HAYKIN, 2002):

1. Adaptação automática: o filtro se adapta à medida que ocorre alguma modificação do ambiente e/ou do sistema;
2. Aprendizagem: possui a capacidade de aprender sobre uma tarefa específica de filtragem ou decisão através de um processo de treinamento;
3. Adaptabilidade: podem extrapolar o espaço de conhecimento e lidar com novas situações após um novo treinamento;
4. Baixa complexidade: possuem baixa complexidade matemática (não há a inversão de matrizes, por exemplo);
5. Tolerância à falhas: até certo ponto podem reparar a si mesmos, mesmo quando sujeitos a certos tipos de defeitos ou limitações;

Geralmente, os filtros adaptativos são mais complexos e difíceis de serem analisados que os filtros determinísticos. Contudo, apresentam um desempenho substancialmente melhor quando as características do sistema são desconhecidas ou variantes no tempo.

9.3.3 Classes de Aplicação de Filtros Adaptativos

As classes de aplicação, ou configurações básicas, de filtros adaptativos, vistas na Fig. 109, podem ser divididas em 4 esquemas principais:

- predição;
- identificação;
- modelagem inversa, e;
- cancelamento de interferência.

A predição é uma tarefa de aproximação de uma função cujo objetivo é estimar valores futuros de uma seqüência de observações baseado nos valores atuais e passados desta seqüência. Assim, na predição, o principal objetivo é prever o valor do sinal de entrada.

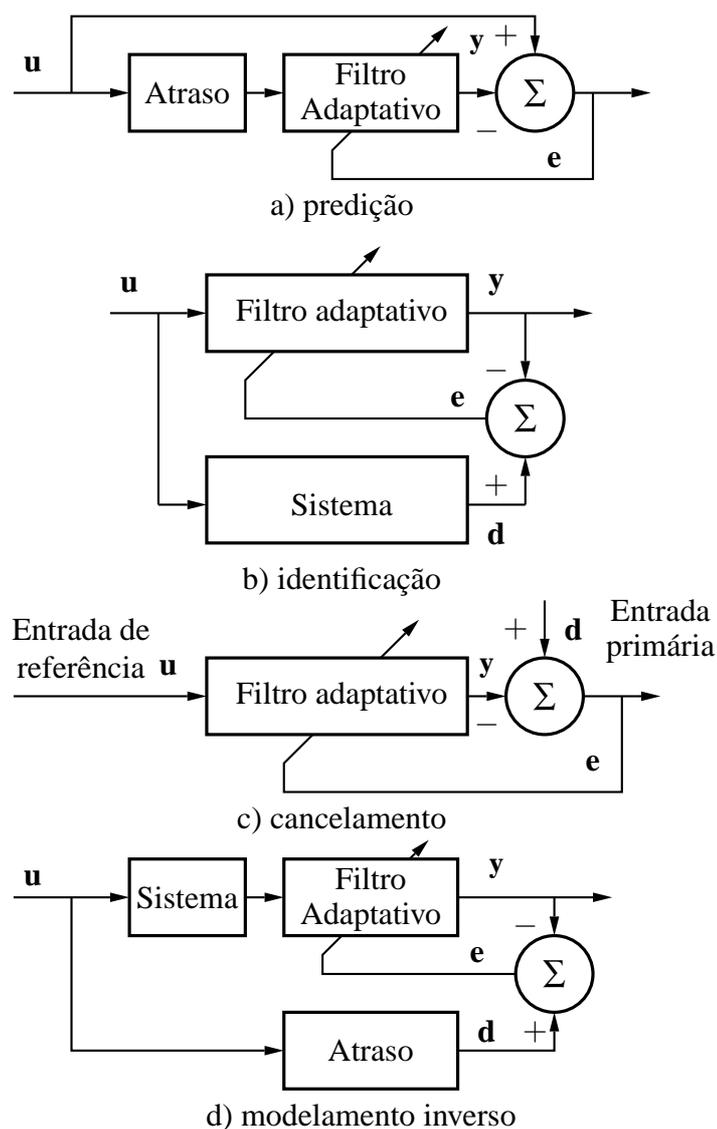


Figura 109: Configurações básicas de filtros adaptativos.

Na aplicação de identificação, espera-se identificar o sistema, fazendo uso do filtro como modelo. Ao se minimizar o erro, a função de transferência do filtro fica o mais próximo possível da função de transferência do sistema, por exemplo. É possível encontrar este tipo de aplicação em sistemas de controle e de comunicações, por exemplo.

No cancelamento de interferência, o principal objetivo é anular a interferência presente na entrada primária, a partir da estimativa da interferência fornecida pela entrada de referência. O cancelamento de eco e a formação adaptativa do diagrama de recepção de antenas são exemplos de aplicação do cancelamento de interferência.

O objetivo na aplicação da modelagem inversa é descobrir a função de transferência inversa (modelo inverso) da planta, ao contrário de descobrir a própria função de transferência. Este tipo de configuração é usado, por exemplo, em sistemas de equalização adaptativa.

9.3.4 Classificação dos Filtros Adaptativos

A classificação dos FA, em geral, não é trivial devido ao elevado número de algoritmos existentes. Contudo, de modo simplificado, os principais algoritmos podem pertencer a três classes: mínimos quadrados, descida mais íngreme e hiper-estabilidade, como visto na Tab. 14. Todos algoritmos procuram alcançar o ponto mínimo da função de custo J .

Esta, pode ser vista, em relação aos coeficientes do filtro, como uma superfície côncava em forma de um vale. Ela possui uma concavidade voltada para cima e seu valor mínimo deve ser maior ou igual a zero devido ao fato de ser função do quadrado do erro. Essa superfície é caracterizada por um único mínimo global, onde o gradiente de J vale zero.

Os algoritmos que fazem a busca do ponto mínimo de J através do gradiente possuem como fundamento o método da descida mais íngreme⁴. Com base em algum parâmetro de convergência, responsável por controlar as correções, este método busca alcançar o ponto mínimo da superfície de performance através de sucessivas iterações.

Tabela 14: Principais classes de algoritmos de filtros adaptativos.

Modelo de Sistema	Estrutura	Classe de algoritmo	Representante
FIR	Direta	Descida mais íngreme	MQM
		Mínimos quadrados	MQR MQR rápido
	Malha	Descida mais íngreme	MGA
		Mínimos quadrados	MMQ
IIR	Direta	Descida mais íngreme	IIR, MQM
		Hiper-estabilidade	HARF, SHARF

Os algoritmos pertencentes a classe descida mais íngreme, tais como o algoritmo dos mínimos quadrados médio (MQM) ou o algoritmo de malha de gradiente adaptativo (MGA), usam uma estimativa para o gradiente baseada nos dados de entrada.

⁴O método da descida mais íngreme também é conhecido como *steepest descent*.

Dependendo de como é realizada a estimativa, diferentes algoritmos podem ser derivados. Estes algoritmos, em geral, são considerados os mais simples em termos de estrutura do algoritmo e de requisitos computacionais exigidos, em virtude da estimativa do gradiente e do seu caráter recursivo.

Os algoritmos que buscam o ponto mínimo através dos mínimos quadrados, tais como o algoritmo de mínimos quadrados recursivo (MQR) ou o de malha de mínimos quadrados (MMQ) por exemplo, tem como base a resolução direta das suas equações. Deste modo, a solução a cada iteração é a ótima, a não ser pela qualidade das médias temporais (estimativas) utilizadas.

Contudo, em relação à classe de algoritmos derivados do gradiente, estes algoritmos apresentam uma estrutura mais complexa e a carga computacional é bastante elevada. Isto se deve ao grande número de amostras necessárias para se obter as estimativas e a necessidade de se calcular uma nova solução, com novos dados, a cada nova iteração. Em compensação, possuem uma taxa de convergência mais elevada.

Os algoritmos pertencentes à classe hiper-estabilidade usam uma estimativa simplificada, em vez da média, para o cálculo do gradiente (como o algoritmo dos MQM, por exemplo) ou fazem uso de coeficientes para fornecer uma versão filtrada do erro (como por exemplo os algoritmos HARF e SHARF ⁵).

Com relação a estrutura, os filtros possuem duas estruturas básicas: direta e de malha, como visto na Tab. 14. A estrutura direta apresenta os resultados mais simples. Contudo, possui uma alta sensibilidade ao mau condicionamento da matriz de correlação que interfere na velocidade de convergência. Já a estrutura de malha é mais robusta. Entretanto, é mais complexa. Em geral os algoritmos com estrutura na forma de malha apresentam algumas vantagens, em relação aos algoritmos de estrutura direta, tais como: uma maior velocidade de convergência e uma maior imunidade ao mau condicionamento dos dados, entre outras.

Quanto ao modelo de sistema, pode-se encontrar os tipos: filtro de resposta ao impulso de duração finita (FIR) ⁶ e o filtro de resposta ao impulso de duração infinita (IIR) ⁷.

Os filtros de resposta ao impulso de duração finita, filtros FIR, são intrinsecamente estáveis e apresentam uma função de custo que possui um único mínimo global. Isto, ajudou a disseminar a implementação deste tipo de FA.

Os filtros IIR apresentam um custo computacional menor, em relação aos filtros FIR, na síntese de funções de transferência de maior duração. Contudo, são mais instáveis do que os filtros FIR. Frequentemente há a presença de mínimos locais na função de custo e a velocidade de convergência é menor, por exemplo.

Com relação às estruturas de implementação de filtros adaptativos, para sistemas FIR, a forma direta apresenta os melhores resultados.

⁵*Hyperstable Adaptive Recursive Filter e Simplified HARF*, respectivamente.

⁶Do inglês *Finite-duration Impulse Response*, ou simplesmente FIR.

⁷Derivado da expressão em inglês *Infinite-duration Impulse Response*, ou apenas IIR.

Contudo, a sensibilidade ao mau condicionamento da matriz de correlação é maior, afetando a velocidade de convergência. A estrutura de malha possui uma sensibilidade menor, mas a sua carga computacional e a complexidade dos algoritmos são maiores.

Destaca-se que ao longo deste trabalho, somente os métodos (descida mais íngreme e mínimos quadrados) relacionados aos filtros FIR com estrutura direta serão analisados.

9.3.5 Método da Descida Mais Íngreme

O método da descida mais íngreme (MDMI) é um método recursivo no sentido que, partindo de um dado ponto na superfície de performance, na medida que o número de aproximações cresce, melhor fica a estimativa dos coeficientes do filtro em direção ao filtro de Wiener.

O filtro de Wiener é um ótimo dispositivo de extração de sinal que minimiza o erro médio quadrático ao longo de todo o processo de filtragem, maximizando a relação sinal-ruído. Seu comportamento é determinado pelas características do sinal, como descrito na seção 9.2.

Do método da descida mais íngreme se originou uma classe de algoritmos que são baseados na busca do gradiente, tais como o dos mínimos quadrados médio e o de malha de gradiente adaptativo, ver Tab. 14. O método da descida mais íngreme descreve um sistema de controle multivariável, realimentado e determinístico que encontra o ponto mínimo da superfície de performance J sem o conhecimento da mesma.

O processo de filtragem, a partir do método da descida mais íngreme, envolve o cálculo da saída de um filtro FIR (\mathbf{y}) em resposta a um sinal de entrada (\mathbf{u}) e a geração de um erro de estimação (\mathbf{e}) que é obtido através da comparação da saída do filtro com uma resposta desejada (\mathbf{d}), como visto na Fig. 110, onde M é o número de coeficientes do filtro e n é o número da amostra do sinal.

Após, no processo de adaptação, ocorre o ajuste automático dos parâmetros do filtro (\mathbf{w}) de acordo com o erro de estimação e da função de custo, calculados anteriormente durante o processo de filtragem.

Os processos de filtragem e adaptação do método da descida mais íngreme são repetidos até o valor do erro seja aproximadamente igual a zero.

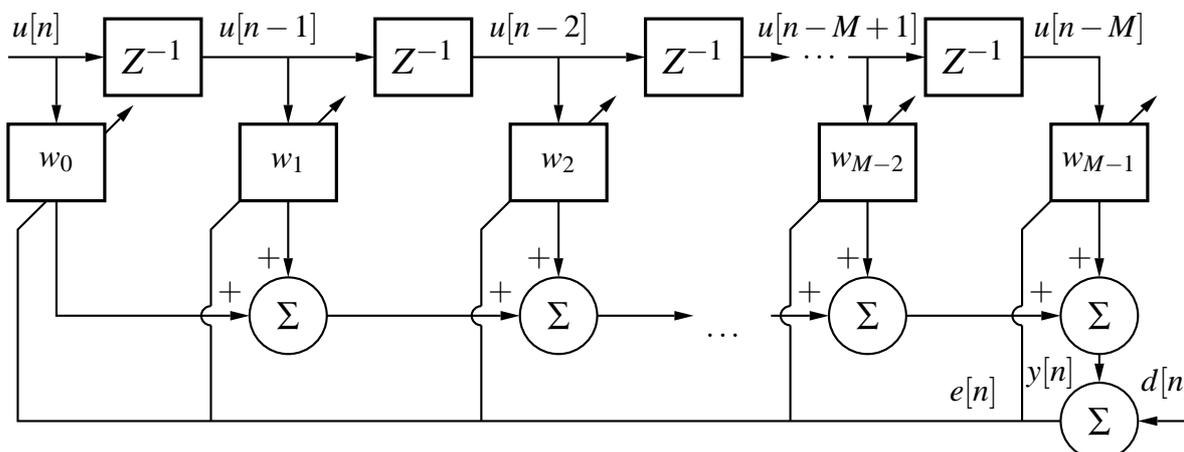


Figura 110: Filtro adaptativo FIR.

9.3.5.1 Derivação do Algoritmo da Descida Mais Íngreme

A cada iteração, o filtro adaptativo FIR (que possui uma estrutura direta com os sinais sendo realizações de processos estocásticos estacionários), como visto na Fig. 110, muda sua resposta ao impulso $\mathbf{w}[n]$.

A resposta ao impulso ótima, \mathbf{w}_O^* , é o filtro de Wiener, obtido a partir de:

$$\mathbf{w}_O^* = R^{-1} \mathbf{p} \quad (162)$$

onde: \mathbf{p} é o vetor de correlação cruzada, R é a matriz de correlação do sinal, e \mathbf{w}_O^* é o vetor de coeficientes ótimos conjugados do filtro, sendo \mathbf{w}^* :

$$\mathbf{w}^* = [w_0^* \ w_1^* \ \dots \ w_{M-1}^*]^T \quad (163)$$

e o vetor do sinal de entrada \mathbf{u} :

$$\mathbf{u}[n] = [u[n] \ u[n-1] \ \dots \ u[n-M+1]]^T \quad (164)$$

onde: M é o número de coeficientes.

A saída do filtro, neste caso, é dada por:

$$y[n] = \mathbf{w}^T[n] \mathbf{u}[n] \quad (165)$$

O sinal de erro é definido como:

$$e[n] = d[n] - y[n] \quad (166)$$

A função de custo é dada por:

$$J[n] = \sigma_d^2 - \mathbf{w}^T[n] \mathbf{p} - \mathbf{p}^H \mathbf{w}^*[n] + \mathbf{w}^T[n] R \mathbf{w}^*[n] \quad (167)$$

Ressalta-se que o processo do erro de estimação $e[n]$ é não-estacionário, visto que o erro muda com a variação dos coeficientes. Entretanto, mesmo $J[n]$ sendo variável, o seu mínimo é dado por:

$$J[n] = \sigma_d^2 - \mathbf{p}^H \mathbf{w}_O^* \quad (168)$$

Assim, a correção aplicada ao vetor de coeficientes é dada por:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + \mu(-J[n]) \quad (169)$$

onde: μ é uma constante positiva conhecida como tamanho do passo.

A função de custo, $J[n]$, é dada por:

$$J[n] = -2\mathbf{p} + 2R\mathbf{w}^*[n] \quad (170)$$

Assumindo que R e \mathbf{p} são conhecidos, obtém-se a equação final do algoritmo do MDMI:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + 2\mu(\mathbf{p} - R\mathbf{w}^*[n]) \quad (171)$$

Pode-se observar que o tamanho do passo, μ , é responsável pelo tamanho da correção aplicada ao vetor de coeficientes \mathbf{w} a cada iteração.

9.3.5.2 Resumo do Algoritmo do Método da Descida Mais Íngreme

Os parâmetros necessários para o uso do algoritmo do método da descida mais íngreme são os seguintes: o parâmetro de adaptação μ , o vetor de correlação cruzada \mathbf{p} e a matriz de correlação do sinal R .

Há dois passos básicos envolvidos na utilização do algoritmo do método da descida mais íngreme após a sua inicialização: o cálculo da função de custo (J) e a atualização do vetor de coeficientes (\mathbf{w}). Estes dois passos são repetidos até que o erro atinja um valor mínimo pré-estabelecido (HAYKIN, 2002).

O resumo do algoritmo do método da descida mais íngreme, considerando o esquema de identificação, é apresentado a seguir:

1. Inicialização: inicializar os valores do vetor de coeficientes, \mathbf{w} :

$$\mathbf{w}^*[0] = 0 \quad (172)$$

2. Cálculo da função de custo: determinar o valor da função de custo (J) a partir de:

$$J[n] = -2\mathbf{p} + 2R\mathbf{w}^*[n] \quad (173)$$

3. Atualização do vetor de coeficientes: ajustar o valor das componentes do vetor de coeficientes (\mathbf{w}^*) segundo a seguinte equação:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + \mu(-J[n]) \quad (174)$$

4. Continuação: retornar ao passo 2 até que o erro seja aproximadamente igual a zero:

$$e[n] = d[n] - y[n] \quad (175)$$

9.3.6 Algoritmo dos Mínimos Quadrados Médio

O algoritmo dos mínimos quadrados médio ⁸ (MQM) é um membro importante da família de algoritmos dos filtros adaptativos que são baseados no gradiente (HAYKIN, 2002; NEGREIROS, 1994; WIDROW; STEARNS, 1985).

Certas características deste algoritmo, tais como a sua simplicidade matemática (não apresenta nenhuma inversão de matrizes, por exemplo) e a sua facilidade de desenvolvimento aliadas a um bom desempenho, fazem dele o algoritmo padrão usado no processo de filtragem adaptativa (em comparação aos demais algoritmos de filtros adaptativos).

O algoritmo dos MQM pode ser visto como sendo derivado do método da descida mais íngreme através do uso de uma estimativa simplificada para o gradiente.

⁸O algoritmo dos mínimos quadrados médio também é conhecido como *least mean squares* ou LMS.

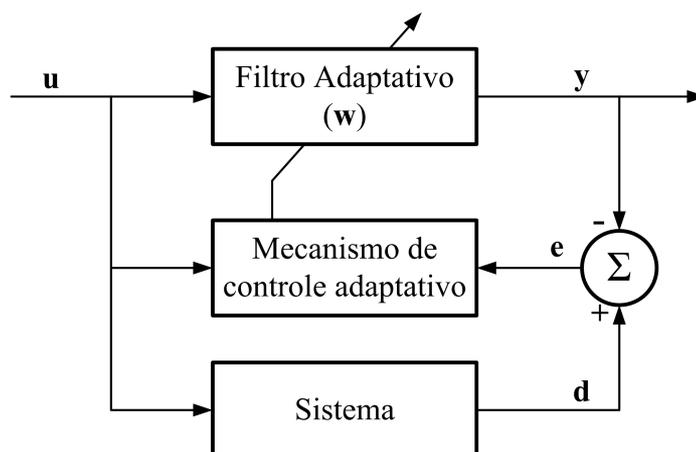


Figura 111: Diagrama de blocos do sistema de filtragem adaptativa via MQM.

Como o método da descida mais íngreme, o algoritmo dos mínimos quadrados médio também representa um sistema de controle realimentado, formado pelos processos de filtragem e adaptação dos coeficientes, como visto na Fig. 111.

De modo semelhante ao método da descida mais íngreme, o processo de filtragem, a partir dos mínimos quadrados médio, envolve o cálculo da saída de um filtro linear (y) em resposta a um sinal de entrada (u) e a geração de um erro de estimação (e) que é obtido através da comparação da saída do filtro com uma resposta desejada (d). No processo de adaptação, ocorre o ajuste automático dos parâmetros do filtro (w) de acordo com o erro de estimação.

A estimativa inicial para o gradiente usada pelo algoritmo dos mínimos quadrados médio é baseada nos dados de entrada do filtro. A simplificação consiste em usar o produto $u[n-k]e[k]$ como estimativa para a componente do gradiente. Esta simplificação, a cada iteração, provoca um erro na estimativa do gradiente.

Neste caso, diz-se então que o algoritmo dos mínimos quadrados médio possui um gradiente ruidoso. Esse gradiente ruidoso faz com que o comportamento do algoritmo dos mínimos quadrados médio seja diferente do método da descida mais íngreme na busca do ponto mínimo da superfície de performance (ou solução de Wiener).

O algoritmo dos mínimos quadrados médio executa um movimento aleatório na região em torno do ponto mínimo, em vez de finalizar na solução ótima, . Isso faz com que o erro mínimo obtido pelo algoritmo dos mínimos quadrados médio em regime permanente seja maior que o erro mínimo do método da descida mais íngreme.

Segundo Haykin (HAYKIN, 2002), esta diferença representa o preço pago por usar um mecanismo adaptativo estocástico de controle dos coeficientes, em lugar de uma abordagem determinística como no método da descida mais íngreme.

Destaca-se que o fator de convergência μ , parâmetro definido pelo projetista, pode ser alterado de modo a controlar a taxa de convergência e o desajustamento do algoritmo, se estiver dentro dos limites de estabilidade.

9.3.6.1 Derivação do Algoritmo dos Mínimos Quadrados Médio

Do método da descida mais íngreme, Eq. (174), vem:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + 2\mu(-J[n]) \quad (176)$$

ou, re-escrevendo:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + 2\mu(\mathbf{p} - R\mathbf{w}^*[n]) \quad (177)$$

A determinação de \mathbf{p} e R , respectivamente, é feita através do seguinte modo:

$$\hat{R}[n] = \mathbf{u}[n]\mathbf{u}^H[n] \quad (178a)$$

$$\hat{\mathbf{p}}[n] = \mathbf{u}[n]d^*[n] \quad (178b)$$

Substituindo a Eq. (178) em (177) obtém-se a seguinte expressão para o gradiente:

$$J[n] = -2\mathbf{u}[n]d^*[n] + 2\mathbf{u}[n]\mathbf{u}^H[n]\mathbf{w}^*[n] \quad (179)$$

onde: H representa as operações de transposição e complexo conjugado (*).

Substituindo esta expressão, Eq. (179), na equação de atualização do vetor de coeficientes para o método da descida mais íngreme, Eq. (174), obtém-se:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + \mu\mathbf{u}[n](d^*[n] - \mathbf{u}^H[n]\mathbf{w}^*[n]) \quad (180)$$

ou, simplificando:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + \mu\mathbf{u}[n]e^*[n] \quad (181)$$

que é a expressão de atualização do vetor de coeficientes, \mathbf{w} , para o algoritmo dos MQM.

9.3.6.2 Resumo do Algoritmo dos Mínimos Quadrados Médio

Os parâmetros necessários para o uso do algoritmo dos mínimos quadrados médio são os seguintes: o parâmetro μ , o vetor de entrada \mathbf{u} e o vetor da resposta desejada \mathbf{d} .

Há três passos básicos na execução no algoritmo do algoritmo dos mínimos quadrados médio após a sua inicialização: o cálculo da saída do filtro (\mathbf{y}), a determinação do erro (\mathbf{e}) e a atualização do vetor de coeficientes (\mathbf{w}^*). Estes passos são repetidos até que o erro atinja um valor mínimo.

O resumo do algoritmo dos mínimos quadrados médio, para o esquema de identificação, é visto a seguir:

1. Inicialização: inicializar os valores do vetor de coeficientes, \mathbf{w} :

$$\mathbf{w}^*[0] = 0 \quad (182)$$

2. Cálculo da saída do filtro: obter o valor da saída do filtro, $y[n]$, a partir de:

$$y[n] = \mathbf{w}^T[n]\mathbf{u}[n] \quad (183)$$

3. Determinação do erro: obter o valor do erro, \mathbf{e} , usando a seguinte expressão:

$$e[n] = d[n] - y[n] \quad (184)$$

4. Atualização do vetor de coeficientes: ajustar o valor das componentes do vetor de coeficientes, \mathbf{w}^* , segundo a seguinte equação:

$$\mathbf{w}^*[n+1] = \mathbf{w}^*[n] + \mu \mathbf{u}[n] e^*[n] \quad (185)$$

5. Continuação: retornar ao passo 2 até que o erro, $e[n]$, seja aproximadamente igual a zero.

9.3.6.3 Análise da Estabilidade do Algoritmo dos Mínimos Quadrados Médio

9.3.7 Etapas dos Filtros Adaptativos

O processamento interno dos algoritmos dos FA, quando voltados para as tarefas de detecção, classificação e predição de falhas, pode ser dividido, simplificadamente, em três etapas, ou fases, distintas:

1. Inicialização: os valores dos vetores de coeficientes são inicializados;
2. Treinamento: é realizado o cálculo da saída do FA (\mathbf{y}), do erro (\mathbf{e}) e do vetor de coeficientes (\mathbf{w}), considerando o sistema livre de falhas, até que o erro seja aproximadamente igual a zero, como visto na Fig. 112 a), ou seja quando o FA apresentar um comportamento aproximadamente idêntico ao sistema em treinamento;
3. Teste: é realizado o cálculo da saída do FA (\mathbf{y}) e do erro (\mathbf{e}), usado para detectar ou classificar um determinada condição de operação. O valor da saída do FA é comparado como a resposta desejada (\mathbf{d}), gerando o valor do erro, como observado na Fig. 112 b).

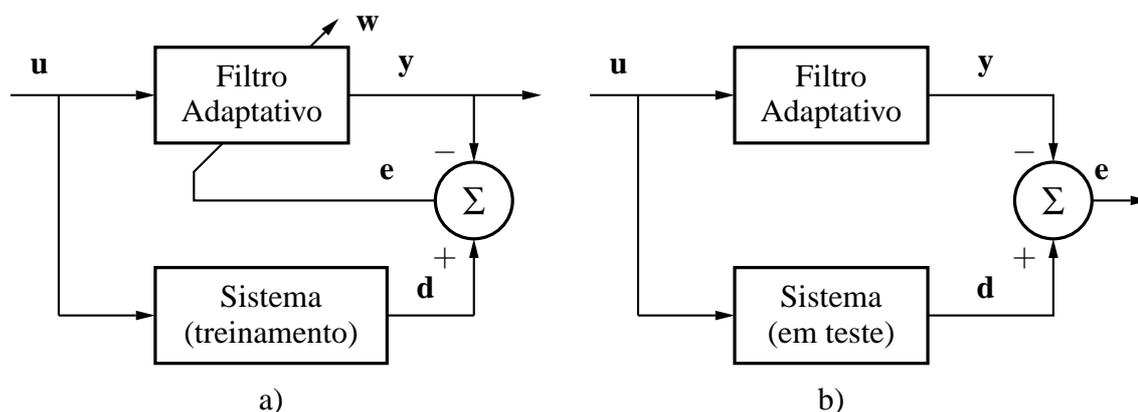


Figura 112: Etapas dos filtros adaptativos: a) treinamento e b) teste.

As etapas de treinamento e teste dos FA diferem basicamente quanto aos tipos de dados que são repassados como valores desejados, de modo semelhante aos MAO, como visto na Fig. 112. Assim, o conjunto de dados desejados repassados aos FA (durante as etapas de treinamento e teste) em função do tipo da tarefa que se deseja realizar (detecção, classificação, ou previsão), foi dividido em três grupos:

1. **Filtro para detecção:** o FA é treinado somente com dados normais. Na etapa de teste, são utilizados dados que representam o comportamento normal, de degradação e de falha do sistema;
2. **Filtro para classificação:** são utilizados somente dados de falha, durante o treinamento. Na etapa de teste são usados dados normais, de degradação e de falha;
3. **Filtro para previsão:** não há uma etapa de treinamento. São utilizados os três tipos de dados na etapa de teste dos FA (normais, degradação e falha), como na Fig. 109 a). Entretanto, estes dados, são repassados como uma série de dados.

9.3.8 Outros Algoritmos de Filtragem Adaptativa

Muitos algoritmos são derivados do algoritmo do MQM. Alguns destes algoritmos podem ser vistos como extensões do MQM e outros como simplificações deste. Um exemplo é o algoritmo dos MQM normalizado (MQMN). No algoritmo dos MQMN, a correção do gradiente é normalizada em relação à norma Euclidiana de $\mathbf{u}[n]$ e não proporcional ao termo $\mathbf{u}[n] \mathbf{e}^*[n]$, como no MQM.

Outros algoritmos, tais como: *sing-error*, *sing-data*, ou *sing-sing* LMS, visam diminuir a carga computacional do algoritmo dos MQM usando somente o sinal dos termos $\mathbf{u}[n]$, $\mathbf{e}[n]$ ou de ambos. A idéia principal é trocar as multiplicações por um conjunto de somas de deslocamentos binários. Contudo, estas modificações alteram o comportamento do algoritmo, diminuindo o seu desempenho e convergência.

Há também os algoritmos dos MQR (onde os coeficientes do filtro ótimo são obtidos através da resolução das equações normais) e o MQR rápido (que faz uso de técnicas específicas para reduzir a carga computacional).

Outros algoritmos, baseados no algoritmo dos MQM, realizam a adaptação de apenas alguns coeficientes, selecionados, a cada iteração. Assim, a carga computacional do algoritmo fica reduzida. Maiores detalhes sobre estes, podem ser encontrados em (HAYKIN, 2002).

9.4 VANTAGENS DOS FILTROS ADAPTATIVOS

Como exemplo de vantagem do uso dos FA, pode-se citar: possui baixa complexidade matemática, pode ser desenvolvido em *hardware* de forma eficiente e alguns algoritmos apresentam elevada robustez numérica e alta velocidade de convergência, por exemplo.

9.5 DESVANTAGENS DOS FILTROS ADAPTATIVOS

As desvantagens encontradas na bibliografia a respeito dos filtros adaptativos sempre buscam relacionar as diferenças entre os filtros FIR e IIR (tais como: os filtros FIR apresentam uma maior complexidade computacional quando comparados aos filtros IIR e alguns casos não podem ser desenvolvidos com filtros FIR, somente com filtros IIR) e não as desvantagens específicas dos FA.

9.6 APLICAÇÃO DOS FILTROS ADAPTATIVOS

Na literatura à respeito, pode-se encontrar diversas publicações abordando o uso dos FA em diferentes aplicações: remoção de ruídos, identificação de sistemas, cancelamento de eco, teste de circuitos integrados, por exemplo (STEMMER et al., 2000; NÁCUL et al., 2002).

Em (ROSA, 2005), foi elaborada uma ferramenta para gerar uma descrição VHDL de filtros FIR paralelos com coeficientes constantes a partir das especificações do filtro, tais como a curva de ganho. O principal objetivo desse trabalho é obter a menor área ocupada, atraso temporal ou consumo de potência (ou qualquer combinação destes). Para tal, três técnicas de otimização em foram desenvolvidas.

Um outro exemplo do uso de técnicas de filtragem adaptativa foi apresentado por (NÁCUL et al., 2002). Neste trabalho, os autores descreveram uma técnica para detectar falhas em circuitos digitais de rádio frequência mistos fazendo o uso de filtros adaptativos. Os resultados apontaram que a técnica de teste é promissora, pois é precisa, apresenta baixo custo computacional e permite uma cobertura de falhas completa com um tempo de teste relativamente pequeno.

Em (BARBOSA et al., 2008), foi apresentado um método para a estimação da frequência em sistemas elétricos de potência fazendo uso de filtros adaptativos baseados no método dos mínimos quadrados. A análise do sistema de potência foi efetuada por meio dos sinais de tensão trifásica, normalizados e enviados ao filtro adaptativo. O método se mostrou fácil de ser desenvolvido apresentado baixo custo computacional.

Entretanto, apesar de ser possível encontrar inúmeras publicações à respeito dos FA, ainda há pouca pesquisa científica voltada para a área da manutenção, ou mesmo centrada na investigação de como utilizar os filtros adaptativos para detectar falhas ou prever o tempo de vida útil de atuadores elétricos e válvulas, em especial.

9.7 CONCLUSÕES DO CAPÍTULO

Ao longo deste capítulo, foi apresentada uma descrição dos FA. Os FA são filtros com coeficientes que variam ao longo do tempo, adaptando-se automaticamente aos sinais de entrada. O processo de adaptação geralmente é baseado em algoritmos de minimização do erro.

Em virtude da sua estrutura simples, os filtros adaptativos vem sendo utilizados largamente em distintas aplicações relacionadas a área das telecomunicações, engenharia, sismologia e radiologia, por exemplo, para a realização de diferentes tarefas, tais como a detecção e a predição de falhas.

Também foram analisados neste trabalho dois dos principais algoritmos dos filtros adaptativos: o da descida mais íngreme e o dos mínimos quadrados médio, incluindo a uma descrição dos algoritmos e análise da estabilidade.

Cabe ressaltar por fim, que há outras classes de algoritmos que também podem ser citadas, tais como: *random search* e hiper-estabilidade que não foram abordados neste trabalho.

REFERÊNCIAS

- ALKAIM, J. L. **Metodologia para incorporar conhecimento intensivo às tarefas de manutenção centrada na confiabilidade aplicada em ativos de sistemas elétricos**. 2003. 239 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2003.
- ALTUG, S.; CHOW, M. Y.; TRUSSELL, H. J. Fuzzy inference systems implemented on neural architectures for motor fault detection and diagnosis. **IEEE Transactions on Industrial Electronics**, Raleigh, v.46, n.6, p.1069–1079, Dec. 1999.
- ASAMER, J.; DIN, K.; WERNER, T. Self organizing maps for traffic prediction. In: INTERNATIONAL MULTI-CONFERENCE OF ARTIFICIAL INTELLIGENCE AND APPLICATIONS, 25., 2007, Innsbruck. **Proceedings...** Anaheim: Acta-Press, 2007. p.24–29.
- AUGER, F. et al. **Time-frequency toolbox for use with matlab**. Paris: CNRS/Rice University, 1996.
- BARBOSA, D. et al. Filtragem adaptativa para a estimação da frequência em sistemas elétricos de potência. **Revista Controle e Automação**, Natal, v.19, n.2, p.226–234, abr. 2008.
- BARRETO, G. A. Time series prediction with the self-organizing map : a review. In: HAMMER B.; HITZLER, P. (Ed.). **Perspectives of neural-symbolic integration**. Berlin: Springer, 2007. p.135–158. (Studies in computational intelligence, v.77).
- BASTOS, E. N. F. **Uma rede neural auto-organizável construtiva para aprendizado perpétuo de padrões espaço-temporais**. 2007. 137 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.
- BENGTSSON, M. Standardization issues in condition based maintenance. In: INTERNATIONAL CONGRESS OF CONDITION MONITORING AND DIAGNOSTIC ENGINEERING MANAGEMENT, 16., 2003, Växjö. **Proceedings...** Växjö: University Press, 2003. p.651–660.
- BERNI, C. C. **Implementação em hardware/firmware de um sensor virtual utilizando algoritmo de identificação nebulosa**. 2004. 179 f. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo, 2004.
- BEZERRA, R. A. **Detecção de falhas em rolamentos por análise de vibração**. 2004. 152 f. Tese (Doutorado em Engenharia Mecânica) - Universidade Estadual de Campinas, Campinas, 2004.

- BOSA, J. L. **Sistema embarcado para a manutenção inteligente de atuadores elétricos**. 2009. 169 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.
- BRIGHAM, E. O. **The fast fourier transform**. New Jersey: Prentice Hall, 1974.
- BRIGHAM, E. O.; MORROW, R. E. The fast fourier transform. **IEEE Spectrum Magazine**, New York, v.4, n.12, p.63–70, Dec. 1967.
- BUCKINGHAM, E. **Manual of gear design**. 3rd ed. New York: Industrial Press, 1935.
- BYINGTON, C. S.; ROEMER, M. J.; GALIE, T. Prognostic enhancements to diagnostic systems for improved condition-based maintenance. In: AEROSPACE CONFERENCE, 2002, Big Sky. **Proceedings...** Big Sky: IEEE, 2002. v.6, p.2815–2824.
- CAMPOS, M. C. M. M. de; TEIXEIRA, H. C. G. **Controles típicos de equipamentos e processos industriais**. Rio de Janeiro: Edgard Blucher, 2006.
- CARRO, L. **Projeto e prototipação de sistemas digitais**. Porto Alegre: Editora da Universidade Federal do Rio Grande do Sul, 2001.
- CASTRO, M. C. F. de. **Predição não-linear de séries temporais usando redes neurais RBF por decomposição em componentes principais**. 2001. 192 f. Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Campinas, Campinas, 2001.
- CHAPPEL, G. J.; TAYLOR, J. G. The temporal kohonen maps. **Neural Networks**, Oxford, v.6, n.3, p.441–445, June 1993.
- CHOW, M. Y. **Methodologies of using neural network and fuzzy logic technologies for motor incipient fault detection**. River Edge: World Scientific Publishing, 1997.
- COHEN, L. **Time-frequency analysis**. New Jersey: Prentice-Hall, 1995.
- CRAMER, J. S. **Econometric applications of maximum likelihood methods**. Cambridge: Cambridge University Press, 1986.
- CROVATO, C. D. P. **Classificação de sinais de voz utilizando a transformada wavelet packet e redes neurais artificiais**. 2004. 143 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.
- CROVATO, C. D. P.; SCHUCK, A. The use of wavelet packet transform and artificial neural networks in analysis and classification of dysphonic voices. **IEEE Transactions on Biomedical Engineering**, Gainesville, v.54, n.10, p.1898–1900, Oct. 2007.
- CUNHA, J. F. **Raciocínio baseado em casos : uma aplicação na manutenção de máquinas e equipamentos**. 2002. 116 f. Dissertação (Mestrado em Engenharia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- DAUBECHIES, I. The wavelet transform : time frequency localization and signal analysis. **IEEE Transactions on Information Theory**, Zürich, v.36, n.5, p.961–1005, Aug. 1990.
- DENG, X.; WANG, Q.; GIURGIUTIU, V. Structural health monitoring using active sensors and wavelet transforms. In: ANNUAL INTERNATIONAL SYMPOSIUM ON SMART STRUCTURES AND MATERIALS, 6., 1999, Newport Beach. **Proceedings...** Bellingham: SPIE, 1999. p.1–8.

DJURDJANOVIC, D. et al. Web-enabled remote spindle monitoring and prognostics. In: INTERNATIONAL CIRP CONFERENCE ON RECONFIGURABLE SYSTEMS, 2., 2003, Ann Arbor. **Proceedings...** Ann Arbor: IMS, 2003. v.1, p.21–22.

DJURDJANOVIC, D.; LEE, J.; NI, J. Watchdog agent: an infotronics-based prognostic approach for product performance degradation assessment and prediction. **Advanced Engineering Informatics**, Amsterdam, v.17, n.3, p.109–125, July 2003.

ENDRENYI, J. et al. The present status of maintenance strategies and the impact of maintenance on reliability. **IEEE Transactions on Power Systems**, Ciudad Real, v.16, n.4, p.638–646, Nov. 2001.

EREN, L.; DEVANEY, M. J. Bearing damage detection via wavelet packet decomposition of the stator current. **IEEE Transactions on Instrumentation and Measurement**, Toronto, v.53, n.2, p.431–436, Apr. 2004.

FILHO, A. M. V. **Um modelo para implementação de consciência em robôs móveis**. 2003. 140 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2003.

FITZGERALD, A. E. **Electric machinery**. New York: McGraw-Hill, 1990.

FOGEL, L. J.; OWENS, A. J.; WALSH, M. J. **Artificial intelligence through simulated evolution**. New York: John Wiley & Sons, 1966.

FUJIMOTO, R. Y. **Diagnóstico automático de defeitos em rolamentos baseados em lógica fuzzy**. 2005. 158 f. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo, 2005.

GARGA, A. K. et al. Hybrid reasoning for prognostic learning in CBM systems. In: AEROSPACE CONFERENCE, 2002, Big Sky. **Proceedings...** Big Sky: IEEE, 2002. v.6, p.2815–2824.

GEREMIA, C. F. **Desenvolvimento de programa de gestão voltado à manutenção das máquinas e equipamentos e ao melhoramento dos processos de manufatura fundamentado nos princípios básicos do total productive maintenance (TPM)**. 2001. 211 f. Dissertação (Mestrado em Engenharia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

GONÇALVES, L. F. **Contribuições para o estudo teórico e experimental de sistemas de geração distribuída**. 2004. 151 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

GRALL, A. et al. Continuous-time predictive-maintenance scheduling for a deteriorating system. **IEEE Transactions on Reliability**, Knoxville, v.51, n.2, p.141–150, June 2002.

HAYKIN, S. **Adaptive filter theory**. 4th ed. Upper Saddle River: Prentice Hall, 2002.

HEBB, D. O. **The organization of behavior**. New York: John Wiley & Sons, 1949.

HELENA, E. L. S. **Detecção de fusos de sono em adultos jovens por meio de transformadas wavelet**. 2004. 120 f. Dissertação (Mestrado em Medicina) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

HELLING, A.; SCHOOR, G.; HELBERG, A. S. J. Data acquisition system for capturing dynamic transients in a control valve. In: AFRICON CONFERENCE IN AFRICA, 7., 2004, Gaborone. **Proceedings...** Gaborone: IEEE, 2004. v.2, p.1085–1090.

HENRIOT, G. **Manual practico de engranajes**. Barcelona: Marciombo, 1967.

HERCULANO-HOUZEL, S.; LENT, R. Isotropic fractionator : a simple and rapid method for the quantification of total cell and neuron numbers in the brain. **The Journal of Neuroscience**, Washington, v.25, n.10, p.2518–2521, Mar. 2005.

HLAWATSCH, F.; BOURDREAUX-BARTELS, G. F. Linear and quadratic time-frequency signal representations. **IEEE Signal Processing Magazine**, Redmond, v.9, n.2, p.21–67, May 1992.

HOLLAND, J. H. **Adaptation in natural and artificial systems**. Massachusetts: MIT Press, 1975.

HUANG, R. et al. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. **Mechanical Systems and Signal Processing**, Amsterdam, v.21, n.1, p.193–207, Oct. 2005.

IEEE-1. **IEEE standard computer dictionary**. Washington: Institute of Electrical & Electronics Engineers, 1990.

IEEE-2. **IEEE standard glossary of computer hardware terminology**. Washington: Institute of Electrical & Electronics Engineers, 1995.

IMS. **Center for Intelligent Maintenance Systems**. Disponível em: <<http://www.imscenter.net>>. Acesso em: 13 out. 2010.

JARDINE, A. K. S.; LIN, D.; BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. **Mechanical Systems and Signal Processing**, Amsterdam, v.20, n.7, p.1483–1510, Oct. 2006.

JENS, R. D. **Modelo de monitoramento e avaliação da confiabilidade e disponibilidade de sistemas de distribuição de energia elétrica com base nas condições de uso de transformadores**. 2006. 134 f. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo, 2006.

JOHNSON, K. et al. **Watchdog toolbox**: integration of multisensor performance assessment tools. Disponível em: <<http://www.imscenter.net>>. Acesso em: 7 mar. 2011.

KAISER, G. **A friendly guide to wavelets**. New York: Springer, 1994.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, New York, v.82, n.1, p.45–35, Nov. 1960.

KOHONEN, T. et al. **The self-organizing map program package**. 3rd ed. Rakentajanaukio: Springer, 1995.

KOTHAMASU, R. et al. Comparison of selected model evaluation criteria for maintenance applications. **Structural Health Monitoring**, Cincinnati, v.3, n.3, p.213–224, Sept. 2004.

- KROSE, B.; SMAGT, P. **An introduction to neural networks**. 8th ed. Amsterdam: The University of Amsterdam, 1996.
- KUNDUR, P. **Power system stability and control**. New York: McGraw-Hill, 1994.
- LAINE, A.; FAN, J. Texture classification by wavelet packet signatures. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, v.15, n.11, p.1186–1191, Nov. 1993.
- LATHI, B. P. **Modern digital and analog communication systems**. 3rd ed. New York: Oxford University Press, 1998.
- LEE, J. Measurement of machine performance degradation using a neural network model. **Computers in Industry**, Amsterdam, v.30, n.3, p.193–209, Oct. 1996.
- LEE, J. et al. Intelligent prognostics tools and e-maintenance. **Computers in Industry**, Amsterdam, v.57, n.6, p.476–489, Aug. 2006.
- LI, F.; BROWN, R. E. A cost-effective approach of prioritizing distribution maintenance based on system reliability. **IEEE Transactions on Power Delivery**, Toronto, v.19, n.1, p.439–441, Jan. 2004.
- LIMA, C. R. C. de; MARCORIN, W. R. Análise dos custos de manutenção e de não manutenção em equipamentos produtivos. **Revista de Ciência e Tecnologia**, Piracicaba, v.11, n.22, p.35–42, jul. 2003.
- LIN, D. et al. An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure. **Mechanical Systems and Signal Processing**, Amsterdam, v.18, n.5, p.993–1007, Sept. 2004.
- LIU, Y.; WEISBERG, R. H.; SHAY, L. K. Current patterns on the west florida shelf from joint self-organizing map analyses of HF radar and ADCP data. **Journal of Atmospheric AND Oceanic Technology**, Boston, v.24, n.4, p.702–712, Apr. 2007.
- LO, C. H.; WONG, Y. K.; RAD, A. B. Intelligent system for process supervision and fault diagnosis in dynamic physical systems. **IEEE Transactions on Industrial Electronics**, Raleigh, v.53, n.2, p.581–592, Apr. 2006.
- MALLAT, S. G. A theory for multiresolution signal decomposition : the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, v.11, n.7, p.679–693, July 1989.
- MARÇAL, R. F. M. **Um método para detectar falhas incipientes em máquinas rotativas baseado em análise de vibração e lógica fuzzy**. 2000. 111 f. Tese (Doutorado em Engenharia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.
- MARÇAL, R. F. M.; SUSIN, A. A. Detectando falhas incipientes em máquinas rotativas. **Revista Gestão Industrial**, Ponta Grossa, v.1, n.21, p.87–96, jun. 2005.
- MARKOV, A. A. The theory of algorithms. **Academic Science of URSS**, Moscow, v.42, n.1, p.3–375, May 1943.

- MASCARENHAS, A. W. **Dispositivo baseado em acelerômetros capacitivos para monitoração de máquinas rotativas**. 2006. 149 f. Tese (Doutorado em Engenharia) - Universidade Federal de Campina Grande, Capina Grande, 2006.
- MATHEW, A. D. et al. Reducing maintenance cost through effective prediction analysis and process integration. **Advances in Vibration Engineering**, Chennai, v.5, n.2, p.87–96, Aug. 2006.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Journal Bulletin of Mathematical Biology**, New York, v.5, p.115–133, Dec. 1943.
- MONTEIRO, S. T.; RIBEIRO, C. H. C. Desempenho de algoritmos de aprendizagem por reforço sob condições de ambiguidade sensorial em robótica móvel. **Revista Controle e Automação**, Natal, v.15, n.3, p.320–338, 2004.
- MORETO, M. **Localização de faltas de alta impedância em sistemas de distribuição de energia** : uma metodologia baseada em redes neurais artificiais. 2005. 126 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005
- MULLER, A.; MARQUEZ, A. C.; IUNG, B. On the concept of e-maintenance : review and current research. **Reliability Engineering and System Safety**, Amsterdam, v.93, n.8, p.1165–1187, Aug. 2008.
- MUÑOZ, D. M. et al. Tradeoff of FPGA design of a floating-point library for arithmetic operators. **Journal Integrated Circuits and Systems**, São Paulo, v.5, n.1, p.42–52, Sept. 2010.
- NANDI, S.; TOLIYAT, H. A. Condition monitoring and fault diagnosis of electrical machines - a review. In: INDUSTRY APPLICATIONS CONFERENCE, 34., 1999, Phoenix. **Proceedings...** Phoenix: IEEE, 1999. p.197–204.
- NASAR, S. A. **Máquinas elétricas**. São Paulo: McGraw-Hill, 1984.
- NASCIMENTO, V. H.; PROVASE, Y. S. Redução de distorção harmônica em amplificadores por meio de filtros adaptativos analógicos. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 20., 2003, Rio de Janeiro. **Anais...** Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2003. p.1–6.
- NASR, M. B.; CHTOUROU, M. A fuzzy neighborhood-based training algorithm for feedforward neural networks. **Neural Computation and Application**, New York, v.18, n.2, p.127–133, Feb. 2009.
- NÁCUL, A. et al. Testing of RF mixers with adaptive filters. **Microelectronics Journal**, Amsterdam, v.33, n.10, p.847–853, Oct. 2002.
- NEGREIROS, M. **Filtros adaptativos aplicados à separação de sinais**. 1994. 162 f. Dissertação (Mestrado em Engenharia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994.
- NUNES, E. L. **Manutenção centrada em confiabilidade (MCC)** : análise da implantação em uma sistemática de manutenção preventiva consolidada. 2001. 146 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, 2001.

OPPENHEIN, A. V.; GEHAFFER, R. W. **Discrete-time signal processing**. New Jersey: Prentice Hall, 1989.

PARRAGA, A. **Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas**. 2002. 163 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

PEREIRA, D. O. F. **Análise da estabilidade de sistemas de geração eólica com aerogeradores de indução com rotor de gaiola**. 2007. 98 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

PINO, B. et al. Design and evaluation of a reconfigurable digital architecture for self-organizing maps. In: INTERNATIONAL CONFERENCE ON MICROELECTRONICS FOR NEURAL, FUZZY AND BIO-INSPIRED SYSTEMS, 7., 1999, Granada. **Proceedings...** Granada: IEEE, 1999. v.1, p.395–402.

PRABHAKAR, S.; MOHANTY, A. R.; SEKHAR, A. S. Application of discrete wavelet transform for detection of ball bearing race faults. **Tribology International**, Amsterdam, v.35, n.12, p.793–800, Dec. 2002.

QIU, H. et al. Wavelet filter based weak signature detection method and its application on rolling element bearing prognostics. **Journal of Sound and Vibration**, Amsterdam, v.289, n.4, p.1066–1090, Feb. 2006.

QUISPE, G. C. S. **Reconhecimento de padrões em sensores integrados**. 2005. 111 f. Tese (Doutorado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo, 2005.

RAMCHANDRAN, K.; VETTERLI, M. Best wavelet packet bases in a rate-distortion sense. **IEEE Transactions on Image Processing**, Evanston, v.2, n.2, p.160–175, Apr. 1993.

REGINATTO, R. **Controle por campo orientado do motor de indução com adaptação de parâmetros via MRAC**. 1993. 146 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, 1993.

REIS, D.; PATI, N. Applications of artificial intelligence to condition-based maintenance. **Revista de Administração de Empresas**, São Paulo, v.40, n.2, p.102–107, abr. 2000.

RIBEIRO, R.; KOERICH, A. L.; ENEMBRECK, F. Uma nova metodologia para avaliação de desempenho de algoritmos baseados em aprendizagem por reforço. In: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE, 33., 2006, Campo Grande. **Anais...** Campo Grande: SBC, 2006. p.433–446.

RIGONI, E.; PELOW, L. A.; SILVEIRA, P. R. Sistema Especialista de Apoio à Manutenção de Sistemas Automatizados. In: CONGRESSO BRASILEIRO DE MANUTENÇÃO, 19., 2004. **Anais...** Curitiba: Abraman, 2004.

ROBERTS, M. J. **Signals and Systems**. New York: McGraw-Hill, 2001.

ROCHA, D. X. **Otimização do planejamento da manutenção preventiva em sistemas complexos, com foco na cadeia de suprimento**. 2002. 94 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, 2002.

ROSA, V. S. da. **Uma ferramenta para a geração otimizada de filtros FIR paralelos com coeficientes constantes**. 2005. 119 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

ROSENBLATT, F. The perceptron : a probabilistic model for information storage and organization in the brain. **Psychological Review**, Washington, v.65, n.6, p.386–408, Nov. 1958.

RÜPING, S.; RÜCKERT, U.; GOSER, K. Hardware design for self organizing feature maps with binary input vectors. **Lecture Notes in Computer Science**, New York, v.21, n.686, p.488–493, Dec. 1993.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Parallel distributed processing**. Massachusetts: MIT Press, 1986.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. 2ed. Rio de Janeiro: Campus, 2004.

SALIM, R. H. **Deteção de faltas em sistemas de distribuição utilizando a transformada wavelet**. 2006. 97 f. (Projeto de Diplomação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

SALOMONI, S. et al. Algoritmo de máxima verossimilhança para a estimação da velocidade de condução média de sinais eletromiográficos de superfície. In: CONGRESSO LATINO AMERICANO DE ENGENHARIA BIOMÉDICA, 4., 2007, Margarita Island. **Anais...** Margarita Island: IFMBE, 2007. v.18, p.1049–1053.

SARKAR, T. K.; SU, C. A Tutorial on wavelets from electrical engineering perspective : the continuous case. **IEEE Antennas and Propagation Magazine**, San Diego, v.40, n.6, p.36–37, Dec. 1998.

SCHUCK, A. **Análise e processamento de sinais de voz disfônica através da transformada wavelet discreta**. 1998. 189 f. Tese (Doutorado em Engenharia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1998.

SELLITTO, M. A. Formulação estratégica da manutenção industrial com base na confiabilidade dos equipamentos. **Revista Produção**, São Paulo, v.15, n.1, p.44–59, jan. 2005.

SHARIF, M. A.; GROSVENOR, R. I. Fault diagnosis in industrial control valves and actuators. In: INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 98., 1998, Minnesota. **Proceedings...** Minnesota: IEEE, 1998. v.2, p.770–778.

SHIKARI, B.; SADIWALA, C. M.; DWIVEDI, R. K. **Automation in condition based maintenance using vibration analysis**. Disponível em: <http://www.maintenanceworld.com/articles/shikarib/international_paper.pdf>. Acesso em: 9 mar. 2004.

SILVA, S. V. da. **Projeto de uma arquitetura dedicada à compressão de imagens no padrão JPEG2000**. 2005. 125 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

SINGH, G. K.; KAZAZ, S. A. S. A. Induction machine drive condition monitoring and diagnostic research : a survey. **Electric Power Systems Research**, Amsterdam, v.64, n.2, p.145–158, Feb. 2003.

SOUZA, F. R. **Representações reduzidas por redes neurais com aprendizado local não-supervisionado sobre campos receptivos restritos**. 2005. 110 f. Dissertação (Mestrado em Física) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

STASZEWSKI, W. J.; BOLLER, C.; TOMLINSON, G. R. **Health monitoring of aerospace structures**. Munich: John Wiley & Sons, 2004.

STEMMER, G. et al. Implementações de filtros adaptativos para cancelamento de eco. In: WORKSHOP IBERCHIP, 6., 2000. **Anais...** São Paulo: Iberchip, 2000. p.140–148.

SU, Y. C. et al. Design and implementation of an intelligent prognostics system. **IEEE Transactions on Semiconductor Manufacturing**, Cambridge, v.1, n.2, p.195–207, Aug. 2005.

SWANSON, D. C. A general prognostic tracking algorithm for predictive maintenance. In: AEROSPACE CONFERENCE, 2002, Big Sky. **Proceedings...** Big Sky: IEEE, 2002. v.6, p.2971–2977.

TAN, W. W.; HUO, H. A generic neurofuzzy model-based approach for detecting faults in induction motors. **IEEE Transactions on Industrial Electronics**, Raleigh, v.52, n.5, p.1420–1427, Oct. 2005.

TARNOWSKI, G. C. **Metodologia de regulação da potência ativa para operação de sistemas de geração eólica com aerogeradores de velocidade variável**. 2006. 123 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

TINÓS, R. **Deteção e diagnóstico de falhas em robôs manipuladores via redes neurais artificiais**. 1999. 117 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo, 1999.

TURING, A. M. Computing machinery and intelligence. **Mind**, Oxford, v.59, p.433–460, 1950.

VALINS, T. F. **Relé digital de distância baseado na teoria de ondas viajantes e transformada wavelet**. 2005. 143 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo, 2005.

VAZ, O. R. et al. Análise de causas de falhas e desempenho de sistemas de distribuição de energia elétrica. In: SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 17., 2006, Campina Grande. **Anais...** Campina Grande: SBSE, 2006. p.137–143.

VESANTO, J. Using the SOM and local models in time-series prediction. In: WORKSHOP ON SELF-ORGANIZING MAPS, 7., 1997, Espoo. **Proceedings...** Espoo: Helsinki University of Technology, 1997. v.1, p.209–214.

VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. **IEEE Transactions on Neural Networks**, Beijing, v.11, n.3, p.586–600, May 2000.

- VESANTO, J. et al. **SOM toolbox for matlab 5**. Disponível em: <<http://www.cis.hut.fi/projects/som-toolbox/2000>>. Acesso em: 12 mar. 2011.
- VIEIRA, Z. P.; BAUCHSPIESS, A. Implementação do servocontrole auto-sintonizado em tempo-real utilizando rede perceptron multicamadas. In: CONGRESSO BRASILEIRO DE REDES NEURAIS, 4., 1999, São Paulo. **Anais...** São Paulo: Sociedade Brasileira de Redes Neurais, 1999. p.308–313.
- VOLLERTT, J. R. **Confiabilidade e falhas de campo** : um estudo de caso para melhoria da confiabilidade de um produto e do reparo através de um procedimento sistemático de coleta de dados. 1996. 128 f. Dissertação (Mestrado em Engenharia) - Universidade Federal de Santa Catarina, Florianópolis, 1996.
- WEG. **Weg indústrias**. Disponível em: <<http://www.weg.net/br>>. Acesso em: 05 jan. 2002.
- WICKERHAUSER, M. V. **Adapted wavelet analysis from theory to software**. Wellesley: AK Peters, 1994.
- WIDROW, B.; STEARNS, S. D. **Adaptive signal processing**. Upper Saddle River: Prentice Hall, 1985.
- WIDROW, B.; WALACH, E. **Adaptive inverse control**. 4th ed. Upper Saddle River: Prentice Hall, 1995.
- WILEY, J. **Editora John Wiley & Sons**. Disponível em: <<http://www.wiley.com/WileyCDA>>. Acesso em: 12 out. 2010.
- XILINX. **Embedded system tools reference manual**. Disponível em: <<http://www.xilinx.com>>. Acesso em: 21 dez. 2010.
- YAM, R. C. M. et al. Intelligent predictive decision support system for condition-based maintenance. **International Journal of Advanced Manufacturing Technology**, Amsterdam, v.17, n.5, p.383–391, Feb. 2001.
- YAMAKAWA, T.; HORIO, K.; HIRATSUKA, T. Advanced self-organizing maps using binary weight vector and its digital hardware design. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING, 9., 2002, Yishun. **Proceedings...** Yishun: IEEE, 2002. v.3, p.1330–1335.
- YAN, J.; LEE., J. Degradation assessment and fault modes classification using logistic regression. **Journal of Manufacturing Science and Engineering**, Evanston, v.127, n.4, p.912–914, Nov. 2005.
- YAN, R.; GAO, R. X. An efficient approach to machine health diagnosis based on harmonic wavelet packet transform. **Robotics and Computer-Integrated Manufacturing**, Amsterdam, v.21, n.4, p.291–301, Aug. 2005.
- YEN, G. G.; LIN, K. C. Wavelet packet feature extraction for vibration monitoring. **IEEE Transactions on Industrial Electronics**, Raleigh, v.47, n.3, p.650–667, June 2000.
- ZADEH, L. A. Fuzzy sets. **Information and Control**, Amsterdam, v.8, p.338–353, June 1965.

10 APÊNDICE I: LISTA DE PUBLICAÇÕES

Ao longo do período de duração do doutorado, foram escritos sete artigos; sendo autor principal em cinco publicações e co-autor em outras três, como visto a seguir:

GONÇALVES, L. F. et al. Fault detection, diagnosis and prediction in electrical valves using self-organizing maps. **Journal of Electronic Testing**, New York, v.27, n.2, p.1-14, May 2011.

BORGES, G. M. et al. Evaluating the effectiveness of a mixed-signal TMR scheme based on design diversity. In: ANNUAL SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEMS DESIGN: CHIP IN SAMPA, 23., 2010, São Paulo. **Proceedings...** New York:ACM, 2010, p.134-139, Sept. 2010.

BORGES, G. M. et al. Diversity TMR: proof of concept in a mixed-signal case. In: LATIN-AMERICAN TEST WORKSHOP, 11., 2010, Punta Del Este. **Proceedings...** New York: IEEE, p.237-242, Mar. 2010.

GONÇALVES, L. F. et al. Fault prediction in electrical valves using temporal kohonen maps. In: LATIN-AMERICAN TEST WORKSHOP, 11., 2010, Punta Del Este. **Proceedings...** New York: IEEE, p.237-242, Mar. 2010.

GONÇALVES, L. F. et al. Design of an embedded system for the proactive maintenance of electrical valves. In: ANNUAL SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEMS DESIGN: CHIP ON THE DUNES, 22., 2009, Natal. **Proceedings...** New York:ACM, 2009, p.1-6, Sept. 2009.

GONÇALVES, L. F. et al. Um método de classificação de falhas em atuadores elétricos. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 17., 2008, Juiz de Fora. **Anais...** Campinas:SBA, set. 2008.

GONÇALVES, L. F. et al. Development of an intelligent maintenance system for electronic valves. In: International Congress of Mechanical Engineering, 19., 2007, Brasília. **Proceedings...** New York:ASME, Nov. 2007.