

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

Automatização da Identificação de Padrões em Gráficos de Controle Estatístico de
Processos (CEP) Através de Redes Neurais com Lógica Difusa

Por : GÉRSO TONTINI

TRABALHO DE TESE SUBMETIDO COMO REQUISITO PARCIAL PARA OBTENÇÃO DO
GRAU DE DOUTOR EM ENGENHARIA MECÂNICA

ORIENTADOR: PROF. ABELARDO ALVES QUEIROZ, PhD

Data : Outubro/1995

**AUTOMATIZAÇÃO DA IDENTIFICAÇÃO DE PADRÕES EM GRÁFICOS DE
CONTROLE ESTATÍSTICO DE PROCESSOS (CEP) ATRAVÉS DE REDES
NEURAS COM LÓGICA DIFUSA**

GÉRSO N TONTINI

ESTA TESE FOI JULGADA PARA OBTENÇÃO DO
TÍTULO DE DOUTOR EM ENGENHARIA MECÂNICA
E APROVADA EM SUA FORMA FINAL PELO CURSO DE PÓS-GRADUAÇÃO EM
ENGENHARIA MECÂNICA



Prof. Abelardo Alves de Queiroz, Ph.D.
Orientador

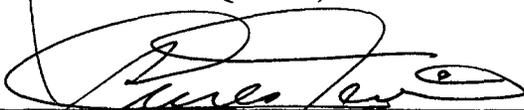


Prof. Antônio Fábio Carvalho da Silva, Dr. Eng. Mec.
Coordenador do Curso

BANCA EXAMINADORA



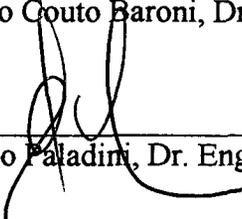
Prof. Abelardo Alves de Queiroz, Ph.D.



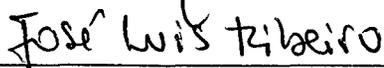
Prof. Aureo Campos Ferreira, Ph.D.



Prof. Dante Augusto Couto Baroni, Dr.



Prof. Edson Pacheco Paladini, Dr. Eng. Prod.



Prof. José Luis Duarte Ribeiro, Ph.D.

"Dedico este trabalho a minha esposa Simone
e minha filha Ana Clara, que está por nascer "

AGRADECIMENTOS

Agradeço a colega Terezinha Guedes pela sua grande colaboração no desenvolvimento do trabalho, ao Prof. Abelardo pela sua orientação e incentivo, e em particular a minha esposa Simone, pela compreensão e apoio recebido. Agradeço ainda a todos que colaboraram direta ou indiretamente para a realização deste trabalho.

ABSTRACT

Each day costumers are demanding more diversified and higher quality products, together with shorter delivery times. It has led to the adoption of Automated Manufacturing Systems (AMS). The quality control of the manufacturing process in AMS is a critical factor that needs to be addressed, requiring flexible and intelligent quality control systems.

Quality Control Windows (QCW) form an adequate framework for development of Automated Quality Control Systems. Quality Control Windows are formed by five steps : Observation, Evaluation, Diagnostic, Decision and Implementation. The most important step is Evaluation, where quality control charts are used to indicated out of control situations and possible underlying causes based on observed patterns.

The other works about automation of pattern identification in quality control charts have been limited in scope and aplicability. These limitations are due a lack in standard simulation procedures and flexibility in learning new patterns.

This work studied the performance of Fuzzy Neural Networks in the Evaluation step of the QCW. The performance of three Fuzzy Neural Networks in the identification of six different "patterns" in quality control charts were evaluated. The studied neural networks were : Radial Basis Functions Network (RBF), Fuzzy Artmap, and a new network, RBF Fuzzy Artmap. The new RBF Fuzzy Artmap Network had the best performance of classification (78.8%), while allowing on-line incremental learning. Besides, it is not sensible to the presentation order of the training "patterns".

RESUMO

A cada dia os consumidores estão exigindo produtos mais diversificados, com maior qualidade, e com prazo de entrega menor. Isto tem levado a adoção de Sistemas de Manufatura Automatizados. O controle da qualidade do processo em Sistemas de Manufatura Automatizados exige sistemas de controle de qualidade flexíveis e inteligentes.

As Janelas de Controle da Qualidade formam uma estrutura adequada para o desenvolvimento de sistemas automatizados de controle de qualidade e são formadas por cinco etapas: Observação, Avaliação, Diagnóstico, Decisão e Implementação. A etapa mais importante é a de Avaliação, onde gráficos de controle da qualidade são usados para verificar se o processo está fora de controle, e indicar as possíveis causas baseado nos padrões observados. Os trabalhos anteriores nesta área tem sido limitados em escopo e aplicabilidade, não utilizando procedimentos padrões para simulação, e não possibilitando o aprendizado de novos padrões.

Este trabalho estuda o desempenho de redes neurais difusas na etapa de avaliação das Janelas de Controle da Qualidade. Foi avaliado o desempenho de três redes neurais na identificação de seis diferentes tipos de padrões. As rede neurais estudadas foram : Rede de Funções de Base Radial (RBF), Rede Fuzzy-Artmap, e uma nova rede aqui apresentada, Rede RBF-Fuzzy Artmap. A nova rede teve o melhor desempenho de classificação (78.8%), permitindo aprendizado incremental em tempo real, além de não ser sensível a ordem de apresentação dos padrões de treinamento.

SUMÁRIO

1-Introdução.....	9
1.1 - Considerações Iniciais.....	9
1.2 - Tema e Justificativa.....	10
1.3 - Objetivos do Trabalho.....	16
1.4 - Metodologia.....	16
1.5 - Padrões a Serem Identificados.....	17
1.6 - Restrições e Características do Processo Simulado.....	19
1.7 - Estrutura do Trabalho.....	20
2 - Modelamento de Sistemas Dinâmicos.....	21
2.1 - Modelamento Determinístico.....	21
2.2 - Modelamento Estocástico.....	21
2.3 - Método GMDH.....	25
2.4 - Redes Neurais.....	26
3 - Gráficos de Controle Estatístico de Processos (CEP).....	34
4 - Automatização do Controle Estatístico de Processos.....	38
5 - Redes Neurais Estudadas.....	48
5.1 - Rede de Funções com Base Radial (RBF).....	49
5.1.1 - Treinamento da Camada Intermediária.....	51
5.1.2 - Vantagens e Desvantagens.....	56
5.2 - Teoria de Ressonância Adaptativa.....	58
5.2.1 - Rede Fuzzy-ART.....	58
5.2.2 - Rede Fuzzy-Artmap.....	61
5.2.3 - Vantagens e Desvantagens.....	68
5.3 - Nova Rede Neural Proposta (Rede RBF-Fuzzy Artmap).....	70

5.3.1 - Descrição.....	70
5.3.2 - Vantagens e Desvantagens.....	73
6 - Treinamento e Otimização das Redes Estudadas.....	74
6.1 - Entradas das Redes.....	74
6.2 - Padrões de Treinamento e Teste.....	78
6.3 - Definição dos Parâmetros Ótimos.....	83
6.4 - Comparação da Sensibilidade a Ordem de Apresentação dos Padrões.....	92
7 - Experimentos e Resultados da Comparação.....	95
7.1 - O Sistema de Simulação e Comparação.....	95
7.2 - Definição dos Níveis de Corte I e II.....	99
7.2.1 - Normas para Determinação dos Níveis de Corte.....	99
7.2.2 - Identificação Preliminar.....	101
7.2.3 - Otimização dos Valores de Corte.....	106
7.3 - Experimentos para Comparação do Desempenho.....	109
7.3.1 - Análise do Desempenho.....	112
7.3.2 - Análise do ARLIDIXc.....	114
7.3.3 - Análise do ARL para um Alarme Falso.....	117
7.4 - Conclusões.....	119
8 - Conclusão.....	120
9 - Sugestões para Continuidade da Pesquisa.....	125
9.1 - Na Área de Controle Estatístico de Processos.....	125
9.2 - Na Área de Redes Neurais.....	126
10 - Bibliografia.....	127
Apêndice 1 - Geração dos Padrões.....	136

1 - INTRODUÇÃO

1.1 - CONSIDERAÇÕES INICIAIS

O controle estatístico de processos (CEP) iniciou com os trabalhos de Shewhart na década de 30. Ele surgiu da necessidade de se controlar a qualidade do processo de fabricação através da identificação de causas assinaláveis de variação. Apesar de sua grande potencialidade, o controle estatístico de processos ficou esquecido pelas grandes potências econômicas até o início da década de 70. Neste período, o controle da qualidade foi dominado pelos métodos de inspeção por amostragem, que não impedem a fabricação de produtos defeituosos, mas apenas servem para detectar se um lote é bom ou ruim.

O CEP voltou a chamar a atenção com o aumento da concorrência, ocasionada principalmente pelo aparecimento do Japão no cenário econômico mundial, como um fornecedor de produtos de alta qualidade a preços baixos.

A globalização da economia, considerada uma tendência irreversível, coloca novos desafios ao controle da qualidade de processos de fabricação. A competição acirrada entre os fabricantes, e um grau de exigência cada vez maior por parte dos consumidores, tem levado a uma diminuição dramática do tamanho dos lotes e a uma procura por níveis de qualidade cada vez maiores. Isto tem levado ao desenvolvimento de sistemas de fabricação cada vez mais complexos e automatizados, com as decisões e ações de controle sendo tomadas em um espaço de tempo cada vez menor. Isto leva à necessidade de adoção de novos paradigmas no controle da qualidade, com a automatização desempenhando papel central neste novo contexto.

1.2 - TEMA E JUSTIFICATIVA

A grande quantidade de informações em sistemas de fabricação automatizados, aliado a necessidade de ações rápidas, leva a procura por uma automatização das etapas do controle da qualidade. Esta automatização não tem por objetivo substituir, mas sim auxiliar o homem na análise e diagnóstico de problemas da qualidade e nas decisões sobre as ações de controle. Os sistemas automatizados de controle da qualidade em processos de fabricação complexos devem ser capazes de:

- a - Tratar simultaneamente diversos tipos de informações, provenientes de diversas fontes, analisando-as em tempo real.
- b - Adaptar-se a mudanças no processo e "aprender" com os problemas e exemplos ocorridos durante sua operação, mesmo que esses exemplos não tenham sido estruturados.
- c - Estabelecer correlações entre um grande número de variáveis, mesmo que essas relações não sejam bem definidas e que haja interação entre as variáveis.
- d - Tratar situações novas, generalizando através da experiência passada.
- e - Tratar relações não lineares.
- f - Tratar problemas difusos, comuns de se encontrar na vida real.
- h - Auxiliar o analista/operador no diagnóstico de problemas da qualidade e no controle do processo.
- i - Integrar o controle das causas comuns ao controle e eliminação das causas assinaláveis de variação;

A satisfação dessas necessidades é fundamental para o desenvolvimento de um sistema automatizado de controle da qualidade que possa encontrar aplicação em sistemas de fabricação reais.

As técnicas tradicionais de controle estatístico de processos (CEP) tem aplicação limitada em ambientes de manufatura flexível automatizada ou de produção em massa de componentes de precisão. Devido ao pequeno tamanho dos lotes e a precisão requerida, o intervalo entre as amostras deve ser reduzido, chegando ao limite de 1 peça no caso da inspeção a 100%. Com a diminuição do intervalo entre as amostras sua autocorrelação se acentua, limitando a aplicabilidade dos gráficos de controle estatístico de processo tradicionais.

Este problema pode ser solucionado se os gráficos de controle forem utilizados em conjunto com o modelamento de sistemas (fig. 1.1). Neste caso os gráficos de CEP são usados para monitorar os resíduos de previsão de um modelo do processo.

Desta maneira, os gráficos de controle podem monitorar tanto causas assinaláveis de variação provenientes de fontes externas, como mudanças dinâmicas no sistema, isto é, desvios entre o modelo e o processo real.

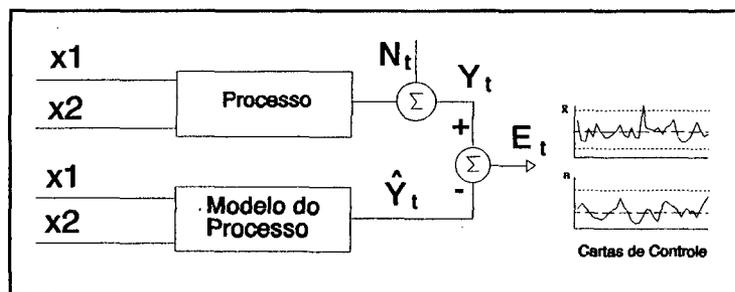


Fig.1.1 - Aplicação Conjunta do Modelamento de Sistemas e Gráficos de CEP

O mesmo problema de autocorrelação existe também na indústria de processos (ex: indústrias químicas), onde muitas vezes o processo tem uma variação previsível devido a temperatura ambiente, tempo de operação, aumento ou diminuição da concentração de reagentes, etc. Estas causas, chamadas de causas comuns de variação, são normalmente controladas, na indústria de processos, por controladores do tipo PID. Esses controladores muitas vezes não são

capazes de compensar todas as variações do processo, estando também sujeitos a defeitos. Neste caso a aplicação de gráficos de CEP é importante para o controle e determinação das causas assinaláveis de variação, com os controladores PID controlando as causas comuns.

Assim, a automatização do CEP é um fator crítico a ser desenvolvido para se atingir um sistema de fabricação confiável totalmente automatizado, tanto na indústria metal mecânica (Sistemas Flexíveis de Manufatura, e Sistemas de Fabricação em Massa de Componentes de Precisão) como na indústria de processos (Química).

Dessouki et al [60] e Dooley et al [61] propõem que esta automatização, integrando o modelamento de sistemas e/ou o controle de processos ao CEP, seja efetuada através de "Janelas de Controle da Qualidade" (Fig. 1.2). A Janela de Controle da Qualidade é uma ferramenta para controle da qualidade de processos de fabricação, composta por cinco etapas: Observação, Avaliação, Diagnóstico, Decisão e Implementação.

Observação - Nesta etapa as entradas (características de qualidade das matérias primas), as saídas (erro entre o valor real das saídas e o previsto pelo modelo) e os parâmetros que influenciam o processo, são observados através de dispositivos adequados, com seus valores passados a etapa seguinte.

Avaliação - Através de testes estatísticos e gráficos de CEP, os erros (ϵ) coletados na etapa de observação são filtrados e analisados, avaliando-se a presença de causas assinaláveis de variação (anomalias).

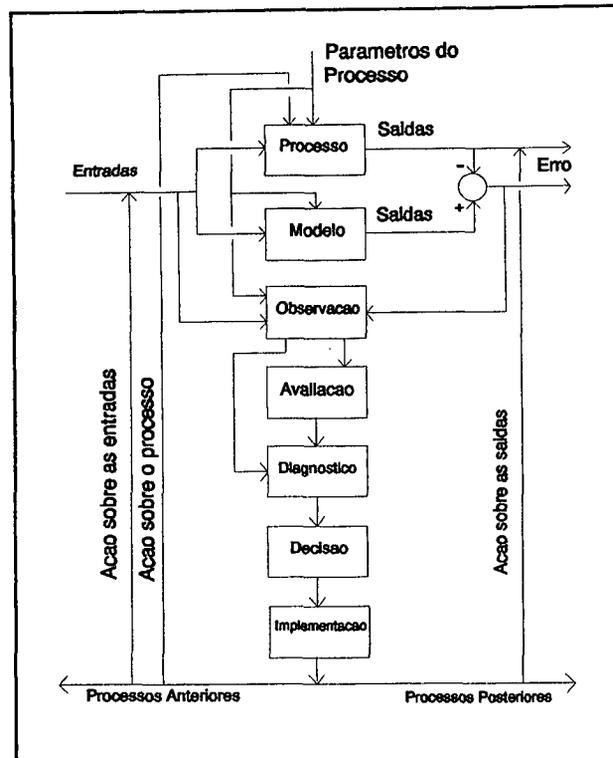


Fig. 1.2 - Janela de Controle da Qualidade

Diagnóstico - Caso exista alguma causa assinalável de variação, os dados das entradas, dos parâmetros do processo, dos erros das saídas e de experiências anteriores, são utilizados para determinar a causa ou as causas fundamentais. Aqui podem ser utilizadas ferramentas de análise e solução de problemas como: diagramas de causa e efeito, correlação, histogramas, gráficos de controle, etc [40].

Decisão - Ações corretivas para ajustar o processo são decididas após encontrar-se a causa fundamental.

Implementação - As ações corretivas decididas na etapa anterior podem ser efetuadas:

- Nos processos anteriores, caso seja um problema de matéria prima.
- Nas próprias matérias primas, separando as que causam o problema.
- Nos parâmetros do processo para compensar a ação da causa fundamental ou para

eliminá-la.

- Nas saídas, separando ou refugando as peças através de inspeção a 100%.
- Nos processos posteriores, para compensá-los devido aos desvios no processo atual.

A Janela de Controle da Qualidade divide o controle da qualidade do processo de fabricação em cinco etapas racionais bem estruturadas. Essas etapas podem ser abordadas, analisadas e automatizadas individualmente sem perda da possibilidade de integração. Esta característica abre uma nova perspectiva para a automatização do controle da qualidade em sistemas flexíveis de manufatura, possibilitando às Janelas de Controle da Qualidade as condições adequadas para a satisfação das necessidades de um sistema automatizado de controle da qualidade.

A automatização da etapa de avaliação, e consequentemente dos gráficos de CEP, é de grande importância para a automatização da Janela de Controle da Qualidade. A detecção rápida de padrões anormais no processo de fabricação permite que o processo seja parado antes de produzir muitas peças defeituosas. A identificação do tipo de padrão permite mais facilidade no diagnóstico da causa fundamental, levando a maior rapidez na correção do problema. Os trabalhos anteriores sobre automatização do CEP [45][47][54][55][56][57][58] apresentam uma série de limitações, sendo que as principais são:

- Não tratam padrões com pontos fora dos limites de controle;
- Tratam apenas da média, assumindo que o gráfico de dispersão está sob controle;
- Não permitem a incorporação de novos padrões com facilidade.

Além das limitações acima, Loucy-Bouler [59] constata que, em sua maioria, os trabalhos

não utilizam simulações comuns, testes e índices padrões para medir o desempenho dos sistemas propostos (como ARL). Além disso, eles não estudam a ocorrência de alarmes falsos quando o processo está sob controle (normal).

A superação das limitações dos trabalhos anteriores, e a satisfação das necessidades de um sistema de controle da qualidade automatizado, é fundamental para o desenvolvimento de sistemas de CEP automatizados que possam ter ampla aplicação em problemas reais.

Redes neurais artificiais com lógica difusa, como as redes RBF [64] e Fuzzy-Artmap [74], pelas suas características, abrem grandes perspectivas para a satisfação das necessidades de um sistema automatizado de controle da qualidade. Além de tratarem as informações de forma mais natural, redes neurais com lógica difusa como as redes RBF [64] e Fuzzy-Artmap[74], apresentam velocidade de treinamento muito superior às redes de Backpropagation [28] tradicionais.

A rede de RBF é funcionalmente equivalente a um sistema de lógica difusa [63]. Porém, devido a necessidade de treinamento "off-line", ela não satisfaz a necessidade "b - adaptar-se a mudanças no processo e aprender com os problemas ocorridos durante sua operação on-line", de um sistema automatizado de CEP.

A rede Fuzzy-Artmap satisfaz as necessidades de um sistema automatizado de CEP listadas anteriormente, porém porém ela apresenta grande sensibilidade a ordem de apresentação dos padrões de treinamento (Cap 5.2), o que pode afetar seu desempenho.

As características e limitações das redes RBF e Fuzzy-Artmap levou ao desenvolvimento da rede RBF-Fuzzy Artmap, apresentada pela primeira vez neste trabalho (Cap. 5.3). A rede RBF-Fuzzy Artmap atende às necessidades de um sistema automatizado de CEP, além de eliminar as limitações das redes RBF e Fuzzy-Artmap.

1.3 - OBJETIVOS DO TRABALHO

O objetivo do trabalho é comparar a aplicação de três redes neurais com lógica difusa (Rede de Funções com Base Radial (RBF), Rede Fuzzy-Artmap, e uma nova rede aqui proposta, chamada rede RBF-Fuzzy-Artmap) na automatização da etapa de avaliação da Janela de Controle da Qualidade.

O estudo comparativo das três redes pretende demonstrar a superioridade da rede RBF-Fuzzy Artmap sobre as duas outras na identificação de padrões. Formada pela fusão das redes RBF e Fuzzy-Artmap, pretende-se demonstrar que a nova rede supera as limitações da rede RBF (Não permitir treinamento "on-line") e da rede Fuzzy-Artmap (grande sensibilidade a ordem de apresentação dos padrões de treinamento).

A aplicação a gráficos de controle estatístico de processos (CEP) procura satisfazer as necessidades de um sistema automatizado de controle da qualidade, apresentadas no capítulo 6, superando as limitações dos trabalhos anteriores a respeito da identificação de padrões em gráficos de CEP.

1.4 - METODOLOGIA

Para fazer a comparação entre as três redes neurais estudadas, decidiu-se pela utilização de um processo simulado, onde os padrões são gerados por algoritmos conhecidos (Apêndice 1) utilizados em trabalhos semelhantes anteriores [45][56][58]. Embora a realização dos objetivos desta tese através de aplicações práticas seria desejável para validar na prática as hipóteses estabelecidas, a utilização da simulação apresenta as seguintes vantagens:

- Utiliza métodos comuns aos utilizados pelos trabalhos anteriores, servindo para uma eventual base de comparação [59].
- Usa técnicas de controle estatístico de processos comuns [59].
- Permite a simulação de uma ampla gama de situações de teste.
- A comparação é feita em ambiente controlado, com as mesmas situações aplicadas a todas as redes.
- As simulações permitem a realização do trabalho no tempo destinado a execução desta Tese.

A utilização em uma aplicação real, mesmo que apenas para fins de ilustração, não foi realizada principalmente pelo fator tempo. O programa computacional desenvolvido em C++ foi preparado apenas para simulação. Uma aplicação real necessitaria o desenvolvimento de uma interface para utilização pelo usuário final, escolha de um processo real, e implementação. Como o tempo necessário para isto é relativamente longo, decidiu-se deixar esta aplicação para uma fase posterior de desenvolvimento do trabalho.

1.5 - PADRÕES A SEREM IDENTIFICADOS

Existem seis tipos de padrões anormais mais comuns de serem encontrados em gráficos de CEP [42](Cap. 4) : Estratificação, Sistemático, Mistura, Ciclo, Tendência e Desvio repentino da média . Neste trabalho, as redes deverão identificar os seguintes padrões :

- ciclos, com amplitude de 1, 1,5 e 2σ (onde σ é o desvio padrão do processo sob condição normal) e período 20;
- padrões sistemáticos com amplitude de 1, 1,5 e 2σ ;

- misturas com amplitude de 1, 1,5 e 2σ ;
- desvios positivos e negativos da média, com amplitude de 1, 1,5 e 2σ ;
- diminuição da variabilidade com amplitude de 0,5, 0,6 e $0,7\sigma$.

As amplitudes dos padrões (1, 1,5 e 2σ) foram escolhidos de tal forma a cobrirem um espectro de amplitudes que ocorram na vida real, e sejam passíveis de identificação. A menor amplitude (1σ) foi escolhida de tal forma a permitir a identificação pelo sistema (Amplitudes menores tornariam a identificação dos padrões muito difíceis, pois elas não se distinguiriam do processo sob condição normal). A maior amplitude (2σ) foi escolhida de modo a permitir que o padrão esteja presente no processo durante algum tempo, antes de um ponto sair dos limites de controle, permitindo assim a sua identificação. Como as amostras são de tamanho um, a estratificação é equivalente a diminuição da variabilidade.

Devido ao ruído do processo e a possibilidade de pontos fora dos limites de controle logo após o início do distúrbio, padrões de tendências são muito difíceis de serem diferenciados dos desvios da média. Isto ocasiona muitas classificações incorretas, com sistema de identificação de padrões indicando desvios da média quando poderia ser tendência, ou o contrário. Como as pessoas são levadas a acreditar nas identificações do sistema, essas classificações errôneas afetariam negativamente a procura pela causa do desvio. Assim, é melhor que o sistema avise que há um desvio da média (incluindo desvios da média permanentes e tendências), deixando para o operador fazer o diagnóstico, do que dar uma indicação errada de qual tipo de padrão está ocorrendo. Desta maneira, os padrões de tendência não serão incluídos no trabalho.

1.6 - RESTRIÇÕES E CARACTERÍSTICAS DO PROCESSO SIMULADO

O trabalho consiste de uma comparação entre as redes neurais de Funções com Base Radial (RBF), Fuzzy-Artmap e RBF-Fuzzy Artmap, aplicadas ao monitoramento de um processo simulado com as seguintes características:

- As informações provenientes do processo são pré-processadas para uma distribuição normal padrão ($\mu = 0$, $\sigma^2 = 1$). Isto significa que elas podem representar os resíduos de um modelo do processo, dentro da Janela de Controle de Qualidade.
- O processo é medido a 100%, isto é, o sistema monitora e avalia o processo a cada nova peça produzida. Desta maneira os padrões são identificados no gráfico de controle dos pontos individuais [40]. Esta restrição é condizente com a necessidade de diminuição do intervalo de amostragem devido a precisão requerida e ao tamanho de lote nos sistemas flexíveis de manufatura.
- O processo é monitorado através de uma janela deslizante contendo as últimas sessenta peças produzidas. O valor dessas peças é utilizado para cálculo das estatísticas de entrada das redes.
- O sistema está sempre em estado de controle no início do monitoramento. O estado de controle é longo suficiente para conter toda a janela de monitoramento.
- O processo pode ter pontos fora dos limites de controle de $\pm 3\sigma$ no gráfico de pontos individuais.

1.7 - ESTRUTURA DO TRABALHO

Este trabalho está estruturado da seguinte forma:

Os capítulos 2 e 3 fazem uma revisão sobre modelamento de sistemas e gráficos de controle estatístico de processos.

O capítulo 4 apresenta em detalhes os tipos de padrões mais comuns de serem encontrados em gráficos de CEP, e faz uma revisão dos trabalhos anteriores sobre a automatização da identificação desses padrões.

O capítulo 5 apresenta as três redes neurais RBF e Fuzzy-Artmap, e apresenta a rede RBF-Fuzzy Artmap como uma evolução das outras duas. Neste capítulo são estudadas as vantagens e desvantagens da três redes.

O capítulo 6 discute a otimização das entradas das redes e de seus parâmetros internos.

O capítulo 7 faz a comparação do desempenho das três redes na identificação dos padrões, através da simulação de situações parecidas com a realidade.

Os capítulos 8 e 9 apresentam as conclusões e sugestões para trabalhos futuros.

O capítulo 10 apresenta a bibliografia.

2 - MODELAMENTO DE SISTEMAS DINÂMICOS

O modelamento de sistemas através de equações matemáticas é uma ferramenta fundamental para a compreensão e o controle de processos de fabricação.

Vários são os métodos utilizados, sendo que mais especificamente no modelamento de processos e sistemas de fabricação pode-se citar:

2.1 - MODELAMENTO DETERMINÍSTICO

No modelamento determinístico uma equação descreve fisicamente as relações de entrada, saída e o relacionamento interno do sistema [1].

Como está baseado em um entendimento das leis físicas internas do sistema, este tipo de modelamento permite que se façam simulações e extrapolações sobre o comportamento do sistema/processo sob estudo. Além disso, ele pode ser perfeitamente utilizado em sistemas não lineares.

Apesar dessas vantagens, o modelamento determinístico necessita de um entendimento profundo do sistema, crescendo muito em dificuldade com o aumento da complexidade do sistema/processo modelado. Sua aplicação em processos de fabricação fica restrito a partes específicas, com pequeno número de variáveis.

2.2 - MODELAMENTO ESTOCÁSTICO

No modelamento estocástico uma equação empírica, sem significado físico, descreve as relações de entrada e saída do sistema. Esta equação não procura modelar todas as variáveis, mas

apenas as mais importantes ou de interesse, deixando as outras para o que se chama "resíduo".

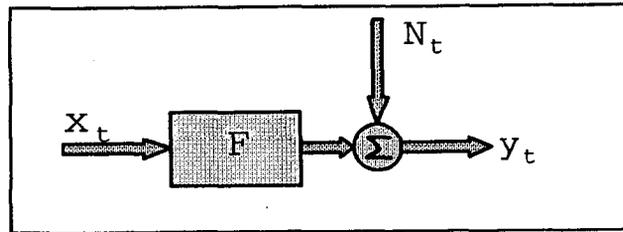


Fig. 2.1 - Sistema de Uma entrada - Uma saída

Desta maneira, um sistema com uma entrada e uma saída importante (Fig. 2.1) ou de interesse é representado por :

$$\omega B Y_t = \theta B X_t + \varepsilon_t \quad \text{Eq. 2.1}$$

onde

$$\omega B Y_t = (1 - \omega_1 Y_{t-1} - \omega_2 Y_{t-2} - \dots - \omega_p Y_{t-p})$$

$$\theta B X_t = (1 - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \dots - \theta_q X_{t-q})$$

ε_t é o resíduo.

"p" e "q" são os "lags" ou atrazo da saída e entrada respectivamente.

Como ε_t é o resultado da ação de todas as outras variáveis que influenciam o processo, ele geralmente pode ser modelado por um modelo MA [2]:

$$\varepsilon_t = (\Phi_1 + \Phi_2 B + \Phi_3 B^2 + \dots + \Phi_r B^{r-1}) t \quad \text{Eq. 2.2}$$

onde

B = operador para o passado $B a_t = a_{t-1}$

a_{t-1} = resíduo branco, ou seja, é independente e tem distribuição normal com média $\mu =$

0 e variância $\sigma^2 = 1$.

Este tipo de modelamento, proposto por Box & Jenkins [2], é largamente aplicado na previsão de séries e sistemas econômicos [2][3][4], como também abriu um grande leque de

oportunidades de aplicação no controle de processos de fabricação complexos [5][6][7].

Apesar de sua grande potencialidade, o método de Box & Jenkins tem sua aplicação restrita a sistemas econômicos, pois a necessidade de grande interação com o analista restringe sua aplicação em sistemas dinâmicos, cujos parâmetros mudam com o tempo. Este é o caso da maioria dos processos de fabricação, principalmente nos processos de usinagem.

Wu [8][9] propõe uma metodologia chamada "Dynamic Data System", ou DDS, para determinação de modelos estocásticos. O método é baseado na determinação sucessiva de modelos ARMA (Auto Regressivo de Médias Móveis) [2], ou ARMAV (ARMA com vetores externos), de ordem cada vez maior através de uma rotina de mínimos quadrados não linear, até que a redução do valor da estatística do teste F (F statistic) seja insignificante.

A vantagem da metodologia DDS é que, pelo fato de não necessitar interação com o analista, a rotina para cálculo dos modelos pode ser automatizada e aplicada a sistemas variantes com o tempo. Outra vantagem da metodologia é que os modelos por ela determinados podem ser relacionados com parâmetros físicos do sistema (amortecimento, rigidez, e frequência natural).

O advento de computadores cada vez mais velozes abriu um grande leque de aplicações para a metodologia DDS no controle e modelamento em tempo real de sistemas variantes:

- Como os modelos estão relacionados com as constantes físicas do sistema modelado, eles podem ser utilizados na análise dinâmica de estruturas [10][11]. A metodologia DDS tem várias vantagens sobre o método baseado em algoritmos de Transformada Rápida de Fourier (FFT): o esforço computacional é menor mesmo com um grande número de entradas, e apresenta grande precisão de resultados [11].
- A metodologia DDS pode ser utilizada com sucesso na identificação do comportamento dinâmico do processo de corte em usinagens [12][13][14][15], sendo assim uma ferramenta com grande potencialidade para a previsão do resultado do processo de

corte e na monitoração e controle do desgaste de ferramentas.

- O monitoramento do desgaste e a detecção de quebras de ferramentas em operações de usinagem pode ser efetuado de diversas maneiras, tais como emissão acústica, medição de vibrações, torque da máquina, medição de forças, etc. O método de monitoramento das forças de corte pode ser aperfeiçoado através do uso de filtros determinados pela metodologia DDS[16]. Esses filtros são atualizados em tempo real, de tal forma a se adaptarem as variações nas condições de usinagem. Outras aplicações da metodologia DDS nesta área podem ser vistas em [17] e [18].
- A crescente automação dos processos de usinagem através de máquinas CNC e Sistemas Flexíveis de Fabricação tem incentivado pesquisas e desenvolvimentos no controle compensatório dos erros nas máquinas ferramenta. Esses erros provêm de diversas fontes: deformações térmicas, vibrações, rigidez da máquina, etc. A sua compensação através da medição "on-line" permite um grande aumento na qualidade do trabalho. Todavia, a medição "on-line" direta dos erros introduzidos na peça usinada é muito difícil, ou até mesmo impossível, devido a presença da ferramenta de corte [19]. A metodologia DDS permite o modelamento em tempo real dos erros de usinagem. Isto é feito através de modelos estocásticos que, através de medição indireta, prevêm os erros que estão sendo introduzidos naquele instante na aresta de corte. Desta forma, atuadores adequados podem compensar o erro através da movimentação da ferramenta de corte. Kim et al [20][21] conseguiram, utilizando a metodologia DDS para previsão em tempo real dos erros de usinagem, uma redução de 56 a 64% no erro de cilindricidade de uma peça mandrilhada, e uma redução na ordem de 80% na ovalicidade de uma peça retificada externamente [22].

A metodologia DDS permitiu um grande avanço na utilização de modelos estocásticos em processos de fabricação, principalmente pela viabilidade de ajuste dos modelos em tempo real. Apesar disto, sua aplicação está restrita a sistemas lineares. A aplicação dos modelos estocásticos, determinados pela metodologia de Box & Jenkins ou DDS, em sistemas não lineares implica na linearização prévia das relações de entrada e saída através de transformações matemáticas. Esta linearização implica em um profundo conhecimento do sistema sob estudo. Além disto, a metodologia DDS não permite a determinação de efeitos interativos entre os fatores.

2.3 - MÉTODO GMDH

O método GMDH (Group Method of Data Handling) [23][24][25] consiste no ajuste sucessivo de polinômios contendo termos interativos da entrada do processo. Esta interação é calculada para cada duas variáveis, ficando o polinômio final (para um sistema com duas entradas) da forma:

$$Y = A + BX_1 + CX_2 + DX_1^2 + EX_2^2 + FX_1 X_2 \quad \text{Eq. 2.3}$$

sendo que os parâmetros A, B, C, D, E, F podem ser iguais a zero.

As referências [24] e [25] mostram bons resultados conseguidos no ajuste de uma equação para previsão da taxa de usinagem de um processo de retífica.

Este método mostra boa aplicabilidade, pois além de considerar as interações entre as variáveis, pode modelar sistemas com relações não lineares de entrada e saída. Porém, o método apresenta as seguintes dificuldades:

- O método GMDH necessita de um cuidadoso delineamento de experimentos [26][27] para a correta identificação das relações entre os parâmetros, o que limita sua aplicação em sistemas de fabricação onde variações experimentais do processo são impossíveis ou

anti-econômicas.

- Sua aplicação em sistemas variantes com o tempo fica limitada devido a necessidade de experimentos.
- Este método só permite interações dois a dois. Interações entre três ou mais variáveis não são consideradas.
- O método não trata sistemas que apresentam "lags" de tempo entre as entradas e a saída.
- Os sistemas podem apresentar apenas uma saída, necessitando um modelo diferente para cada variável de saída.

2.4 - REDES NEURAIS

Há muitos anos os pesquisadores vem tentando entender como o cérebro humano funciona na esperança de que esse conhecimento possa ser usado na implementação de máquinas inteligentes. As redes neurais artificiais são o primeiro passo consistente nesta direção.

Os primeiros estudos no desenvolvimento de redes neurais começaram nos anos 40. Todavia, o grande potencial e interesse despertado pelas redes neurais nos dias de hoje tem seguido a contribuição de muitos pesquisadores, em diferentes áreas [28].

Vários são os tipos de redes existentes, sendo que cada tipo tem sua área de aplicação específica (Tabela 2.1). Estas redes podem ser classificadas de acordo com sua "meso-estrutura" (fig. 2.2), sendo que cada meso-estrutura diferente tem também uma dinâmica diferente de funcionamento e aprendizado, o que a habilita mais a um tipo de aplicação que a outro.

REDE	ANO	APLICAÇÕES PRIMÁRIAS	VANTAGENS	DESVANTAGENS
ADALINE / MADALINE	1960	Filtro Adaptativo de Sinal, Equalização Adaptativa	Rapidez, Fácil Implementação	É Linear, classificando apenas espaços linearmente separáveis
ADAPTATIVE RESSONANCE THEORY	1983	Reconhecimento de Padrões	Auto-aprendizado e aprendizado "on-line"	A natureza das categorias pode mudar com o tempo
BACKPROPAGATION	1974/1986	Reconhecimento de Padrões, Filtro de Sinais, Remoção de Ruído, Classificação, Modelamento de Sistemas	Rapidez, bom na formação de representação interna de características das entradas, bem estudado	Longo tempo de aprendizado, não permitindo aprendizado "on-line"
RECORRENTE	1987	Controle de Robôs, Reconhecimento de Fala	Classifica e mapeia informação variante no tempo	Complexidade da rede pode dificultar treinamento
TIME-DELAY	1987	Reconhecimento de Voz	Desempenho equivalente e mais rápido que métodos tradicionais	Janela de Tempo Fixa.
FUNCTIONAL-LINK NETWORK	1988	Classificação e Mapeamento	Somente duas camadas, com treinamento mais rápido	Não há definição clara de como definir as função de ligação
RADIAL BASIS FUNCTIONS	1987/1988	Classificação e Mapeamento	Treinamento mais rápido e mesmo desempenho do Backpropagation	Ainda não conhecidas
BACKPROPAGATION OF UTILITY FUNCTION THROUGH TIME	1974	Neurocontrol	Metodologia neural mais abrangente para modelamento e/ou controle	Necessita de um modelo diferenciável exato do sistema a ser controlado
BIDIRECTIONAL ASSOCIATIVE MEMORY - BAM	1987	Memória Heteroassociativa de conteúdo endereçável	Simple, regra de aprendizado clara, estabilidade dinâmica	Pobre capacidade de armazenagem e precisão de recuperação de informações
BOLTZMANN / CAUCHY MACHINE	1984 / 1986	Reconhecimento de Padrões, Otimização	Forma representação ótima de características nos padrões	Longo tempo de aprendizado
BOUNDARY CONTOUR SYSTEM	1985	Processamento de imagens de baixo nível	Baseado biologicamente para segmentação excelente	Arquitetura complexa, com várias camadas
BRAIN-STATE-IN-A-BOX	1977	Recuperação Autoassociativa (completa padrões parciais)	Possivelmente melhor desempenho que rede de Hopfield	Aplicações potenciais e desempenho não exploradas
HOPFIELD	1982	Recuperação Autoassociativa, Otimização	Conceito simples, estabilidade dinâmica	Pouca capacidade de memória, com muitas respostas espúrias
LEARNING VECTOR QUANTIZATION	1981	Recuperação Autoassociativa, Compressão de Dados	Auto-organização dos vetores, execução rápida após treinamento	Longo tempo de treinamento, número de vetores para treinamento indefinido
NEOCOGNITRON	1975 / 1982	Reconhecimento de escrita manual	Reconhece padrões com escala, rotação e translação diferentes	Estrutura complexa, requer muitos neurônios e camadas
SELF-ORGANIZING TOPOLOGY PRESERVING MAP	1981	Mapeamento complexo, Compressão de dados, otimização	Auto organiza as classes com significado físico entre elas	Longo tempo de treinamento, número de exemplos e tempo para treinamento indefinido

Tabela 2.1 - Aplicação de Redes Neurais [28]

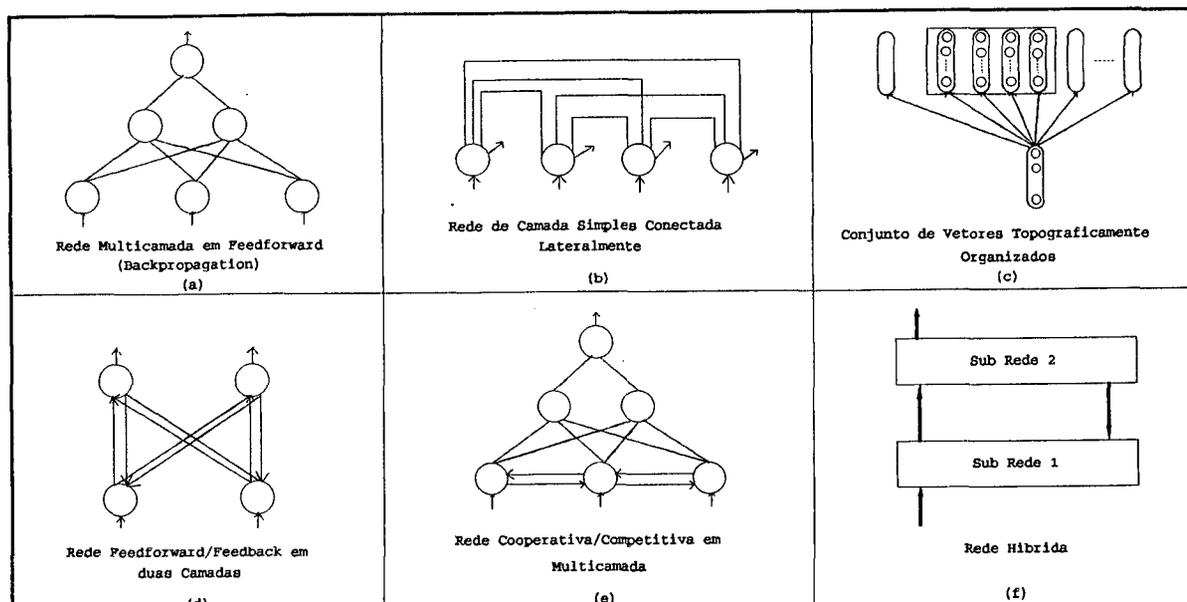


Fig. 2.2 - Seis Principais Tipos de Meso-Estrutura de Redes Neurais [28]

As redes neurais tem uma série de características que as tornam atrativas a diversas aplicações, como o modelamento e controle de sistemas dinâmicos. Entre suas características estão [28] :

- Capacidade de Aprendizado : As redes neurais tem a habilidade de aprender com base em dados apresentados para treinamento ou em experiências passadas.
- Auto-Organização : Uma rede neural pode criar sua própria organização ou representação das informações que recebe durante o treinamento e/ou operação.
- Degeneração Suave : Dados parciais, contaminados com ruído, ou uma destruição parcial da rede causa uma degeneração gradativa, diminuindo seu desempenho, mas ainda guardando alguma capacidade de associação ou operação.
- Operação em Tempo Real : Como o processamento é feito em paralelo, com cada nó da rede (em uma mesma camada) processando ao mesmo tempo as informações, os cálculos e operações são feitas muito mais rapidamente que em sistemas comuns

(desde que a rede esteja implementada em "hardware").

- É Essencialmente não Linear : Como as funções de transferência em cada nó são não lineares, a rede tem a capacidade de modelar sistemas não lineares.

A rede mais desenvolvida e aplicada ao modelamento de sistemas dinâmicos é a rede de multicamadas em "feedforward" (Multilayer Feedforward Network), também chamada de "Multilayer Perceptron" ou "BackPropagation Network" (Fig 2.3). Neste texto esta rede será referenciada como rede de "backpropagation".

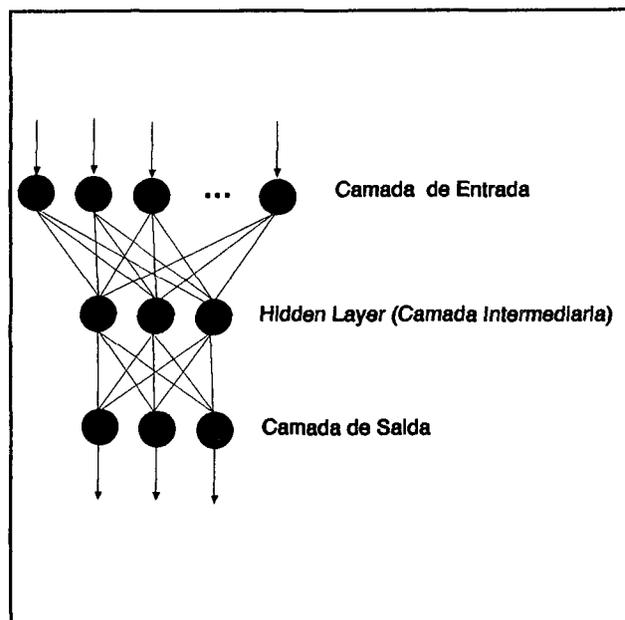


Fig. 2.3 - Rede de "Backpropagation"

A rede de "Backpropagation" consiste de uma rede com duas ou mais camadas de "neurônios" (nós). A primeira camada, que recebe as informações do exterior, é chamada de camada de entrada (input layer). A camada que entrega as informações processadas pela rede é chamada de camada de saída (output layer). As camadas entre a entrada e a saída são chamadas camadas intermediárias (hidden layers). Cada camada tem um ou mais nós, que são conectados a outros nós (geralmente todos) da camada seguinte. Todas as entradas de um neurônio são

multiplicadas por um determinado peso e depois somadas. O resultado desta soma é processada por uma função de ativação não linear (Fig. 2.4).

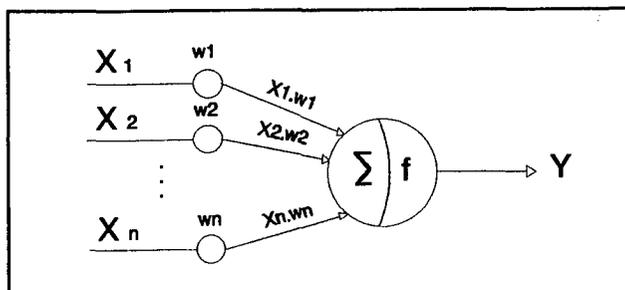


Fig. 2.4 - Neurônio

A função de ativação é geralmente do tipo sigmóide (fig. 2.5), que permite uma diferenciação da saída em relação a suas entradas em todo o seu domínio.

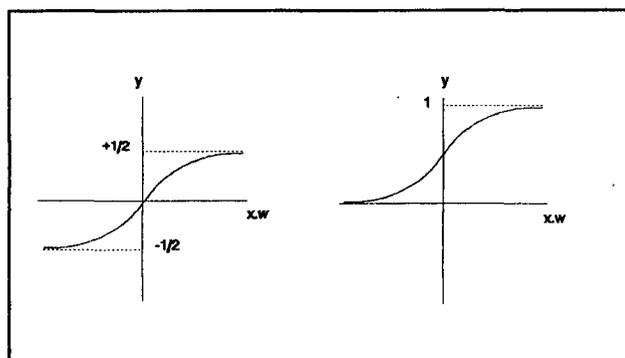


Fig. 2.5 - Função de Ativação Sigmóide

Esta diferenciação é necessária ao método de aprendizado "backpropagation" [28]. Uma função sigmóide comum é do tipo :

$$Y_j = \frac{1}{1 + e^{\sum (\omega_{ij} X_{ij}) + \theta_j}} \quad \text{Eq. 2.4}$$

onde, Y_j é a saída do neurônio j ;

ω_{ij} é o peso da i -ésima entrada do neurônio j ;

X_{ij} é a i -ésima entrada do neurônio j vinda da camada anterior;

θ_j é o Bias, que funciona como limiar de ativação do neurônio j .

O valor dos pesos (ω_{ij}) e do limiar de ativação (θ_j) de cada neurônio é que determinam o desempenho da rede. Seus valores podem ser determinados usando-se o método de "backpropagation" [28]. Esse método tem 3 fases :

- 1- O valor de Y_j de cada neurônio é calculado depois que uma entrada é apresentada e propagada pela rede.
- 2- A saída da rede é comparada com um valor desejado, e um sinal de erro é calculado (Valor Desejado - Valor Obtido).
- 3- O erro é propagado através da rede, da saída para a entrada, e os pesos e limiar de ativação de cada neurônio são ajustados para diminuir este erro.

O usuário deve decidir sobre o número de camadas intermediárias e de neurônios em cada camada. Esta seleção continua sendo uma das tarefas mais delicadas na construção de redes neurais e, geralmente, é realizada de maneira experimental. A inclusão de poucos neurônios nas camadas intermediárias não permite à rede o modelamento das informações. A inclusão de muitos neurônios não forçam a generalização das regras, e a rede não funcionará bem quando uma nova gama de dados for apresentada [28].

A capacidade que as redes neurais tem de modelar relações não lineares, de operar com grande número de variáveis e de modelar relações não bem conhecidas, tem permitido sua aplicação ao modelamento de um grande número de sistemas dinâmicos.

Chen & Billings [29] mostram a aplicação de três redes neurais diferentes na identificação de sistemas discretos temporais não lineares. Estes sistemas são da forma:

$$Y_t = f(Y_{t-1}, \dots, Y_{t-n}, U_{t-1}, \dots, U_{t-m}) + E_t \quad \text{Eq. 2.5}$$

onde

$$Y_t = [Y_1(t), \dots, Y_k(t)], \quad U_t = [u_1(t), \dots, u_p(t)], \quad E_t = [e_1(t), \dots, e_k(t)]$$

são os vetores de saída, entrada e ruído, respectivamente. "n" e "m" são os "lags" das saídas e

entradas; e $f(\cdot)$ é alguma função não linear.

Segundo Hornik [30], a rede de "backpropagation" com uma camada intermediária e um adequado número de nós é capaz de representar qualquer função contínua. Os resultados conseguidos por Chen & Billings [29], abrem grandes perspectivas para utilização das redes neurais no modelamento e previsão de processos de fabricação mecânica.

Um sistema variante com o tempo pode ser modelado como um sistema fixo se os fatores que o tornam variante puderem ser levados em consideração no modelo. Este é o caso do modelamento das forças de corte em processos de usinagem que, através da metodologia DDS [14], tinha que ter os parâmetros do modelo recalculados para cada nova condição de corte (material, velocidade de corte, etc). A utilização de redes neurais permite que as condições de corte sejam levadas em consideração, possibilitando que o modelo seja utilizado para várias velocidades de corte diferentes [31].

Fang & Yao [32] modelaram com sucesso o processo de acabamento em operações de torneamento através de redes neurais, quantificando o interrelacionamento entre a mudança na forma do cavaco, o estado de desgaste da ferramenta (incluindo desgaste de flanco e cratera) e o acabamento superficial. Sathyanarayanan et al [33] aplicaram redes neurais no modelamento de um processo de retificação. Neste estudo a potência, a força normal de usinagem e o acabamento superficial são previstos pelo tipo de rebolo, pelo avanço e pela profundidade de corte. Os resultados obtidos pelo modelo foram utilizados para achar as condições ótimas de corte através de um processo de otimização múltipla. Cook & Shannon [34] utilizaram redes neurais no modelamento de um processo de produção de chapas de aglomerado para fabricação de móveis. Neste caso, modelos de regressão linear foram capazes de explicar apenas 25% da variação no produto final. Quando analisados por séries temporais, os dados se mostraram não estacionários na sua variabilidade, e as transformações realizadas não foram capazes de atingir

estacionariedade. As redes neurais aplicadas foram capazes de prever ~~corretamente, em 70%~~ dos casos, o estado de controle de parâmetros específicos do processo de fabricação.

Apesar das inúmeras vantagens apresentadas, ainda persistem dificuldades na aplicação de redes neurais de "backpropagation" no modelamento de sistemas, entre elas: o longo tempo de treinamento, a determinação do número de neurônios e camadas intermediárias e a necessidade de repassar todos os exemplos de treinamento anteriores quando se quer ensinar novos conhecimentos.

3 - GRÁFICOS DE CONTROLE ESTATÍSTICO DE PROCESSOS (CEP)

O controle de processos efetuado através de controladores PID e atuadores, tem como objetivo reduzir a variabilidade do processo no curto prazo, compensando tanto causas comuns como assinaláveis de variação [38]. Porém, este controle não elimina as causas da variação, não levando a uma melhoria do processo no longo prazo.

O controle de longo prazo, que procura identificar e eliminar as causas que desviam o processo de seu comportamento normal (causas assinaláveis de variação), é executado pelos gráficos de controle estatístico de processos.

Os gráficos de controle estatístico de processos (CEP) tiveram sua origem nos trabalhos de Shewhart [39]. Entre os gráficos de controle mais comumente usadas hoje em dia estão os gráficos \bar{x} -R. Os gráficos \bar{x} -R são dois gráficos com valores de amostras de um processo plotados em sequência (Fig. 3.1). O gráfico \bar{x} é utilizado para monitorar a variabilidade entre as amostras. O gráfico R é utilizado para monitorar a variabilidade dentro de cada amostra.

Dois limites de controle, baseados na variabilidade das amostras, são calculados para cada gráfico. Se um ponto cair fora dos limites de controle, ou apresentar um comportamento anormal [40], detecta-se que o processo contém uma causa assinalável de variação, e uma ação corretiva pode (e deve) ser executada. Os gráficos \bar{x} -R são mais sensíveis a grandes alterações na média do processo, sendo ineficientes quando estas alterações são pequenas ($\pm 1\sigma$) [41].

O comportamento anormal dos pontos em um gráfico de controle é detectado visualmente e pela ajuda de regras auxiliares chamadas "run rules", como as da Western Electric [42] e Nelson [43][44] (Tabela 3.1). Essas regras permitem melhorar o desempenho dos gráficos \bar{x} -R para pequenas variações no processo, mas também aumenta em muito o número de alarmes falsos quando o processo está sob condição normal [45]. Um estudo realizado por Champ & Woodwal

[46] revela que as regras da Western Electric estão sujeitas a um alarme falso a cada 91.75 amostras (ARL para alarme falso). Além disto, essas regras não dão uma indicação segura do tipo de padrão anormal que está presente no processo, com uma regra se sobrepondo a outra [47].

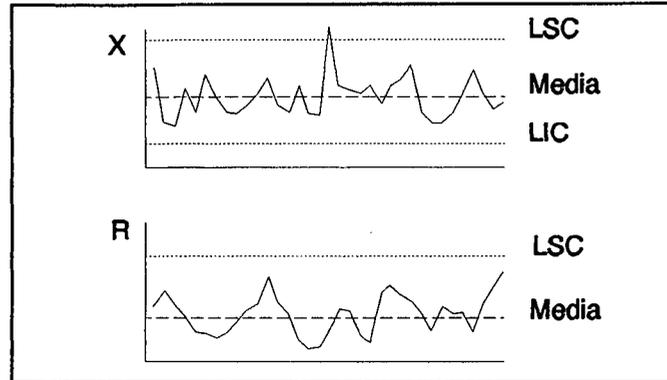


Fig. 3.1 - Gráficos \bar{x} -R (LSC - Linha Superior de Controle, LIC - Linha Inferior de Controle)

Regra 1	Um ponto fora dos limites de controle de $\pm 3\sigma$
Regra 2	Dois de três pontos consecutivos entre os limites de 2σ e 3σ (positivo ou negativo)
Regra 3	Quatro de cinco pontos consecutivos entre os limites de 1σ e 3σ (positivo ou negativo)
Regra 4	Oito pontos consecutivos acima ou abaixo da média

Tabela 3.1 - "Runs Rules" - Western Electric [42]

Pequenas variações podem ser melhor detectadas por gráficos de soma cumulativa, ou CUSUM [41][48]. Estes gráficos plotam a soma das diferenças entre a média desejada e a média real de cada amostra, comparando o resultado com limites de controle adequados. Apesar de sua sensibilidade a pequenas variações, os gráficos de CUSUM não tem vantagem sobre os gráficos \bar{x} -R quando as variações no processo são grandes ($\pm 3\sigma$) [49]. Uma maneira computacionalmente efetiva de implementar os gráficos de CUSUM é dada na referência [50], onde as somas negativas e positivas são calculadas em separado pelas fórmulas (assumindo média

zero e variância unitária):

$$S_0(+)=S_0(-)=0$$

$$S_t(+)=\text{Max}(0, S_{t-1}(+) + y_t - K)$$

$$S_t(-)=\text{Max}(0, S_{t-1}(-) + y_t + K)$$

onde, y_t é o valor da amostra;

K é um parâmetro relacionado com a sensibilidade do CUSUM, significando metade da menor alteração na média do processo considerada importante suficiente para ser detectada rapidamente.

Uma mudança da média é detectada quando $S_t(+)$ > h ou $S_t(-)$ < $-h$, onde h é um valor limite e está diretamente relacionado ao tempo para detecção. " h " é selecionado como uma solução de compromisso entre um tempo aceitavelmente longo até a ocorrência de um alarme falso quando o processo está normal, e um tempo aceitavelmente curto até a detecção quando a média do processo desviou $2K$ unidades do objetivo. Apesar de existirem tabelas para a determinação de K em condições ideais, na prática ele é determinado por tentativa e erro [78]. Um valor comumente encontrado na literatura é $h = 6$ [57][58].

Uma outra opção para detectar pequenos desvios na média do processo são os gráficos de médias móveis exponencialmente ponderadas (EWMA - Exponentially Weighted Moving Average) [51]. As médias móveis ponderadas exponencialmente calculam o valor futuro da saída do processo através da média ponderada dos valores passados, com pesos cada vez menores quanto mais antigos os dados. Assim, um gráfico de EWMA pode ter uma memória longa, provendo uma carta similar os gráficos de CUSUM, ou ter uma memória curta, tendo um comportamento similar aos gráficos \bar{x} -R.

O gráfico de EWMA é facilmente construído e tem uma forma semelhante aos gráficos

\bar{x} -R, porém dá a previsão para o resultado da próxima amostra.

Apesar de ser pouco conhecido, o gráfico de EWMA tem a vantagem de poder ser combinado com o gráfico de \bar{x} -R em uma só carta, monitorando assim tanto pequenas quanto grandes variações no processo [52].

De uma maneira geral, os gráficos de controle são construídos sobre a hipótese de que as amostras são independentes (não autocorrelacionadas), e são normalmente distribuídas em torno de uma média. Esta hipótese é frequentemente violada, principalmente em processos de fabricação onde os intervalos de amostragem são pequenos, como nos processos de usinagem em pequenos lotes, e fabricação em massa de componentes de precisão, ou na indústria de processos.

Essa correlação entre as amostras pode ser compensada pela monitoração dos resíduos de ajuste de um modelo do processo (Fig. 3.2). Se o modelo representar bem o processo, os resíduos deverão ser independentes e ter distribuição normal, sendo então possível utilizar os gráficos de CEP.

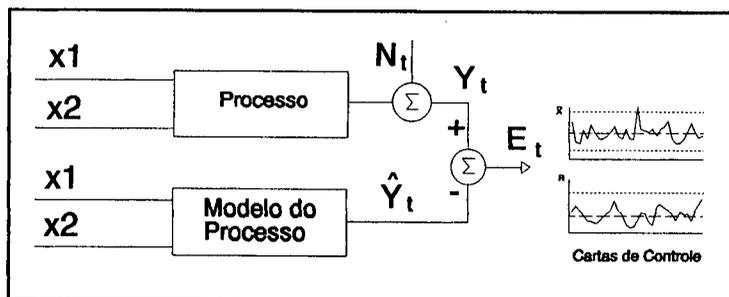


Fig. 3.2 - Aplicação Conjunta do Modelamento de Sistemas e Gráficos de CEP

4 - AUTOMATIZAÇÃO DO CONTROLE ESTATÍSTICO DE PROCESSOS

Os gráficos de CEP não servem apenas para indicar a presença de uma anormalidade no processo, mas também para auxiliar na determinação do tipo e da forma (padrão) da anormalidade que está presente. A identificação correta do tipo de anormalidade é fundamental para a determinação das causas do problema, e portanto para a melhoria do processo.

Existem seis tipos de padrões anormais mais comuns de serem encontrados em gráficos de CEP [42] : Estratificação, Sistemático, Mistura, Ciclo, Tendência e Desvio repentino da média (Figura 4.1).

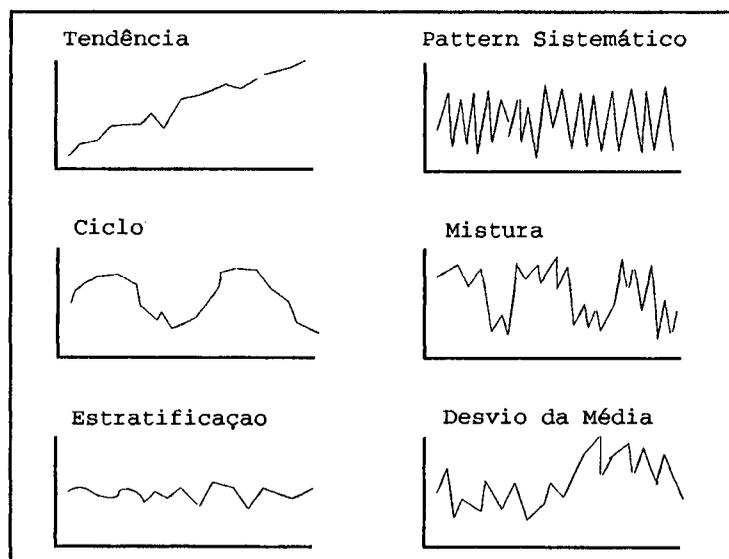


Fig. 4.1 - Tipos de Padrões Anormais em Gráficos de CEP

Estratificação - é caracterizada por uma aglomeração dos pontos em torno da média do processo. Este tipo de padrão acontece quando as amostras provêm de duas populações com distribuição de probabilidades distintas. As peças em cada amostra podem provir tanto de uma como outra população, fazendo com que a média da

amostra esteja próxima a média do processo. Segundo a Western Electric [42] as causas da estratificação podem ser por exemplo: diferentes lotes de materiais misturados na linha, diferença em padrões e equipamentos de medição e/ou pulos e instabilidade frequente no controle automático.

Sistemático - padrão sistemático é caracterizado por um comportamento previsível do processo (Por exemplo, um ponto acima e outro abaixo da média, a assim sucessivamente). As causas dos padrões sistemáticos podem ser por exemplo: divisão sistemática dos dados, uma grande diferença entre mudanças de turno e mudanças periódicas nas fontes de materiais.

Mistura - é caracterizada por um conjunto de pontos acima da média e outro abaixo, sem periodicidade definida. A mistura não apresenta uma proporção adequada de pontos próximo a média do processo. Neste caso, as amostras provêm de duas populações diferentes, porém as peças dentro de cada amostra provêm geralmente de apenas uma população. As causas da mistura podem ser por exemplo: dispositivos de fixação não confiáveis, mudanças no torque de fixação, mistura de material, fadiga do operador, amostras provenientes de processos diferentes, programa de manutenção inadequado, erro no cálculo e marcação das amostras no gráfico e equipamento de medição não confiável.

Ciclo - caracteriza-se por um comportamento previsível e periódico do processo. As causas dos ciclos podem ser por exemplo: fadiga do operador, rotação de fixações e equipamento de medição, desgaste da ferramenta e diferenças regulares entre os turnos.

Tendência - caracteriza-se por uma mudança gradual no nível (ou dispersão) do processo. Muitas vezes um ciclo, quando detectado no início, é interpretado como uma

tendência. As causas das tendências podem ser por exemplo: fadiga do operador, curva de aprendizado do operador, produtos se tornando sistematicamente mais ou menos homogêneos e desgaste de ferramenta.

Mudança da média - uma mudança brusca da média é normalmente causada por um fator novo a influenciar o processo. As causas mais comuns são operador ou equipamento novo, mudança na motivação dos operadores, mudança nos materiais ou fornecedores e mudança no intervalo de manutenção.

Vários trabalhos já foram realizados no desenvolvimento e estudo de sistemas automatizados para identificação automatizada de padrões nos gráficos de CEP. Porém todos esses trabalhos apresentam limitações quanto ao escopo e aplicabilidade.

Swift [54] desenvolveu um sistema especialista para identificação de desvios da média, tendência, sistemático, mistura, ciclos e estratificação. O sistema utiliza as regras da Western Electric [42] para sinalizar um padrão anormal. A partir dos pontos sinalizados o sistema determina o início e o fim do padrão anormal, determinando assim uma janela de análise. O sistema deduz o tipo de padrão presente através de um conjunto hierárquico de testes de hipótese sobre a média e a dispersão do processo, e por um conjunto de regras heurísticas baseadas na probabilidade de ocorrência de algumas características dos padrões. O primeiro teste estatístico conduzido é um teste para verificar a média do processo. A média dos últimos cinco pontos é comparada com a média histórica do processo para verificar o desvio na média. Se o desvio é significativo, então o padrão é considerado desvio da média ou tendência. Se o desvio não é significativo, um teste sobre a variância indica se o padrão está sob controle ou não. Se o teste sobre a variância é significativo as seguintes análises são efetuadas:

- se mais que 75% dos pontos na janela de análise está dentro dos limites de 1σ , o padrão é considerado estratificação, senão passa a etapa seguinte;
- conta o número de vezes que o gráfico tem um movimento para baixo seguido de um movimento para cima. Se é mais que 84%, o padrão é considerado sistemático, senão passa a etapa seguinte;
- faz o ajuste de um ciclo. Se a amplitude é significativa, aceita como ciclo, caso contrário uma mistura está presente; e
- um teste para identificação de tendência separa os padrões com tendência dos desvios da média.

O trabalho de Swift não faz nenhuma referência ao número de alarmes falsos para quando o processo está sob controle, e apresenta uma performance variada de acordo com a intensidade do distúrbio e ao comprimento ou duração do padrão anormal. O sistema de apresenta as seguintes limitações:

- A utilização de uma sequência de testes de hipótese aumenta muito a probabilidade de alarmes falsos [45].
- O sistema utiliza uma sequência de pontos em controle, seguido por uma sequência sujeita a uma anormalidade, seguido novamente por uma sequência de pontos sob controle, o que é uma situação improvável de acontecer em um processo real.
- Não permite a incorporação de novos padrões.

Cheng [55] desenvolveu um sistema que procura identificar os seis tipos de padrões apresentados anteriormente. Os padrões são separados em dois grupos: determinísticos e não determinísticos. Os padrões determinísticos (tendência, ciclo, sistemático e estratificação) são identificados pelo reconhecimento e combinação de três elementos formadores de suas primitivas.

Segmento entre dois pontos consecutivos a) com inclinação positiva, b) segmento com inclinação negativa, c) segmento sem inclinação. A primitiva de um padrão é formado pela combinação desses três elementos. Os padrões não determinísticos (mudanças na média e mistura) são determinados por regras sobre a distância entre pontos consecutivos. A regra para mudança da média procura por uma diferença significativa entre dois pontos consecutivos, que seja precedida e seguida de diferenças não significantes. A regra para misturas, procura um número consecutivo de diferenças significantes.

O sistema foi bem sucedido para padrões com uma forma bem definida, como ciclos e padrões sistemáticos. Porém, para padrões sem forma definida, como mudanças abruptas na média, misturas e tendências, o sistema apresentou muitos alarmes falsos e um tempo para reconhecimento (Average Run Length - ARL) maior que o esperado. Além disto, o sistema não apresenta a oportunidade de inclusão de novos padrões a serem identificados.

Hwang [45] [47] [56] desenvolveu um sistema para identificação de padrões em gráficos de CEP baseado em uma rede neural de "backpropagation". O sistema identifica padrões apenas no gráfico da média, considerando o gráfico da dispersão sob controle.

A rede neural utiliza como entrada uma janela com os últimos 8 pontos do gráfico da média. O gráfico é dividido em 7 regiões entre os limites de $\pm 3\sigma$, sendo que cada ponto é digitalizado conforme o seu valor está enquadrado em uma dessas regiões. Assim, por exemplo, um ponto que cair próximo ao limite superior de controle terá seu valor digitalizado para 1000000 (Fig. 4.2). Desta maneira, a camada de entrada da rede tem $8 \times 7 = 56$ nós. A camada intermediária apresenta 30 nós e a camada de saída um nó para cada tipo de padrão a ser identificado.

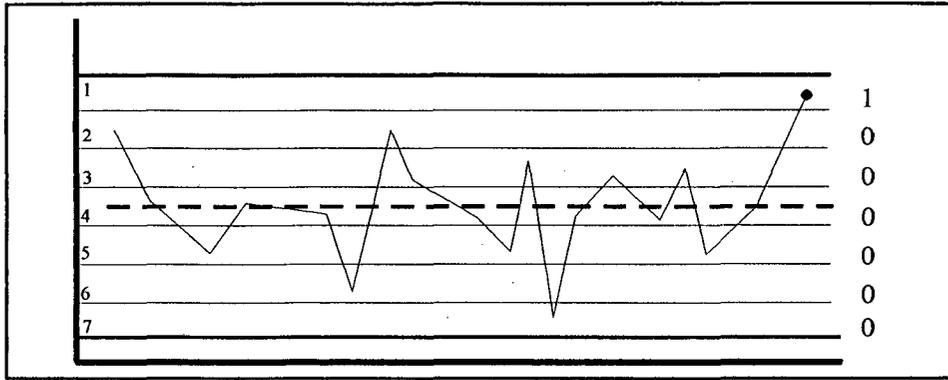


Fig. 4.2 - Digitalização das Entradas do Sistema de Hwarng

A rede foi treinada por centenas de exemplos de padrões de ciclo, sistemático, mistura, desvio positivo da média, desvio negativo da média, tendência positiva, tendência negativa e estratificação, com diversas intensidades de distúrbio. A rede dá o alarme para a existência de um determinado tipo de padrão quando o nó de saída, que o representa, ultrapassar o valor de corte (na faixa de 0.90). Os padrões foram divididos em dois tipos: repetitivos (ciclo, tendência, estratificação, mistura e sistemático) e não repetitivos (desvio na média do processo).

Para os padrões do tipo repetitivo Hwarng propõe, como medida de desempenho, a utilização de uma estatística chamada ARLDIX (Eq. 4.1), que é a razão entre o ARL para identificação correta dos padrões sobre a percentagem de classificações corretas. Para o desempenho dos padrões não repetitivos Hwarng propõe uma estatística chamada ROT, que é a percentagem de classificações corretas.

$$ARLDIX = \frac{ARL}{\% \text{ Classificações corretas}} \quad \text{Eq. 4.1}$$

Para avaliação do desempenho da rede, cada padrão repetitivo foi apresentado continuamente através de uma janela deslizante, até completar 50 observações ou ser identificado

pela rede. Os primeiros 8 pontos já continham integralmente o distúrbio, não havendo nenhuma sequência de pontos sob condição normal. Para os padrões não repetitivos, a rede foi apresentada com os padrões contendo o desvio exatamente no centro da janela de entrada. Os resultados variaram conforme o tipo de padrão, intensidade do distúrbio, nível de ruído e nível de corte.

O trabalho de Hwang apresenta as seguintes limitações, que restringem sua aplicação prática e não permitem que se possa extrapolar os resultados para o desempenho do sistema sob condição real:

- O sistema trata apenas de padrões que não apresentem pontos fora dos limites de controle, não estudando o desempenho da rede quando, por exemplo, um ciclo levar a um ponto fora de controle.
- O sistema considera que a rede é apenas apresentada a padrões onde o processo está fora de controle, não estudando a transição entre um processo sob controle e um fora de controle.
- O sistema considera que o gráfico da dispersão está sob controle, avaliando apenas o gráfico da média
- Apresenta dificuldade na inclusão de novos padrões, pois exige que a fase de treinamento seja repetida com os exemplos anteriores e os novos.

Guo & Dooley [57] utilizaram uma rede neural de "backpropagation" para identificação de aumentos na média e variância de um processo simulado. O sistema utiliza os resíduos de ajuste de um modelo do processo de fabricação para identificar as mudanças da média, da variância e da validade do modelo. A entrada da rede é formada pelos resultados de um gráfico de CUSUM computacional [50] (S_t^+ , S_t^-) e uma estatística (SSQ_t) calculada por (assumindo média zero e

variância unitária):

$$SSQ_t = \sum_{t=t_1}^{t_2} y_t^2 \quad Eq. 4.2$$

onde $t_2 - t_1 + 1 = 25$ é constante, com t_2 sendo o último resíduo do modelo do processo.

Também foram utilizados como entrada os últimos cinco resíduos ($y_t, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}$), a média e a variância de quatro janelas de resíduos ($\{t-5, t-9\}$, $\{t-10, t-14\}$, $\{t-15, t-19\}$ e $\{t-20, t-25\}$ respectivamente, onde t é o índice de tempo para a última amostra). Assim, a rede possui 16 entradas (duas para a estatística de CUSUM). A rede possui duas camadas intermediárias, a primeira com 48 nós, a segunda com 4 nós e apenas um nó na camada de saída. Se a saída for próxima de zero, corresponde a um aumento na média, se a saída for próxima de um, corresponde a um aumento na variância. A rede é acionada apenas para identificação do tipo de distúrbio presente, com o alarme de anormalidade sendo dado pelo teste estatístico de CUSUM e do Qui-quadrado (SSQ_t). O trabalho de Guo & Dooley apresenta as seguintes limitações :

- O sistema trata apenas de problemas que não apresentem pontos fora dos limites de controle, não estudando o desempenho da rede quando a anomalia levar a um ponto fora de controle.
- Identifica apenas desvios da média e variabilidade, não estudando outros tipos de padrões.

Lucy-Bouler [58] desenvolveu um sistema especialista baseado no uso de autocorrelações e autocorrelações parciais das amostras para identificação de padrões anormais no gráfico da média, considerando o gráfico da dispersão sob controle.

O sistema tem uma estrutura hierárquica onde as autocorrelações e autocorrelações parciais são primeiramente analisadas para o reconhecimento de ciclos e padrões sistemáticos. Então, as regras da Western Electric [42] são utilizadas para verificar a existência de misturas ou estratificação. Em seguida, o gráfico de CUSUM, em conjunto com regras eurísticas é utilizado para detecção de desvios da média e tendência. Caso nenhum dos estágios anteriores indique a presença de um padrão anormal, o processo é considerado sob controle. O trabalho de Lucy-Bouler é mais completo que os anteriores, pois ele não apenas estuda o desempenho do sistema proposto na identificação dos padrões anormais, mas também estuda o desempenho do sistema quanto ao ARL para a ocorrência de um alarme falso, quando o processo está sob controle. O sistema pode ser adaptado para situações reais mais facilmente que os sistemas apresentados anteriormente. Apesar do avanço sobre os trabalhos anteriores, o sistema apresenta as seguintes limitações:

- Não considerou situações onde o processo sai fora dos limites de controle de $\pm 3\sigma$ em um gráfico de controle.
- Não permite a inclusão de novos padrões para identificação.

A tabela 4.1 mostra as limitações dos diversos sistemas apresentados acima.

Swift [54]	Cheng [55]	Hwang [45]	Guo & Dooley [57]	Lucy-Bouler [58]
<ul style="list-style-type: none"> - Risco de alarmes falsos - Situação improvável na vida real - Não permite a incorporação de novos padrões 	<ul style="list-style-type: none"> - ARL grande - Não permite a inclusão de novos padrões 	<ul style="list-style-type: none"> - Não trata padrões com pontos fora de controle - Padrões anormais não são precedidos por padrões normais - Considera que o gráfico da dispersão está sob controle - Apresenta dificuldade para inclusão de novos padrões 	<ul style="list-style-type: none"> - Não trata padrões com pontos fora de controle - Só identifica desvios da média e variabilidade - Não permite a inclusão de novos padrões 	<ul style="list-style-type: none"> - Não trata padrões com pontos fora de controle - Não permite inclusão de novos padrões

Tabela 4.1 - Limitações dos Trabalhos Anteriores

A análise das limitações dos trabalhos acima revela que são três as limitações mais comuns :

- Não tratam padrões com pontos fora dos limites de controle.
- Tratam apenas da média, assumindo que o gráfico de dispersão está sob controle.
- Não permitem a incorporação de novos padrões com facilidade.

Além das limitações acima, Loucy-Bouler [59] constata que, em sua maioria, os trabalhos não utilizam simulações, testes e índices comuns para medir o desempenho dos sistemas propostos (como ARL). Além disso, eles não estudam a ocorrência de alarmes falsos quando o processo está sob controle (normal).

A superação dessas limitações é fundamental para o desenvolvimento de sistemas de de CEP automatizados que possam ter ampla aplicação em problemas reais.

5 - REDES NEURAIIS ESTUDADAS

A lógica difusa, proposta por Zadeh [62], tem encontrado cada vez mais aplicações em problemas reais. A lógica difusa é mais adequada que a lógica clássica para representação de problemas reais, onde as informações são mal definidas e imprecisas.

Na lógica clássica, um objeto ou pertence a uma classe ou conjunto, ou não pertence. Porém, a maioria dos problemas do mundo real não possuem limites perfeitamente definidos. O que ocorre é que um objeto pode pertencer ao mesmo tempo a duas ou mais classes distintas, com diferentes graus de pertinência a cada uma.

Este é o caso típico de padrões em gráficos de CEP. Por exemplo, um determinado padrão pode ter características de desvio da média, e também de diminuição da variabilidade.

As redes neurais apresentam uma série de vantagens para aplicação na identificação de padrões, tais como aprendizado e auto-organização. A introdução de lógica difusa dentro das redes neurais permite que elas tratem problemas difusos de maneira mais natural.

Neste trabalho serão analisadas três redes com lógica difusa :

- Rede de Funções com Base Radial (RBF), que é funcionalmente equivalente a um sistema com lógica difusa [63];
- Rede Fuzzy-Artmap, que utiliza operações difusas para processar e armazenar informações;
- Rede RBF-Fuzzy Artmap, uma nova rede aqui proposta, formada pela fusão das duas redes anteriores.

5.1 - REDE DE FUNÇÕES COM BASE RADIAL (RBF)

O longo tempo de treinamento e a dificuldade de determinação do número de camadas intermediárias e do número de neurônios nessas camadas, apresentada pela rede neural de "backpropagation", causam inúmeras dificuldades para sua aplicação na automatização das Janelas de Controle de Qualidade. A rede de Funções com Base Radial (Radial Basis Function Network - RBF) é uma alternativa, pois o tempo de treinamento é em geral uma ordem de grandeza menor [64], além de não ter o problema de determinação do número de camadas e neurônios intermediários. Outro fator relevante é o de ser equivalente a um sistema especialista com lógica difusa [63], podendo trabalhar melhor com dados incertos. As redes RBF podem ser utilizadas para classificação, aproximação de funções, modelamento de sistemas, etc.

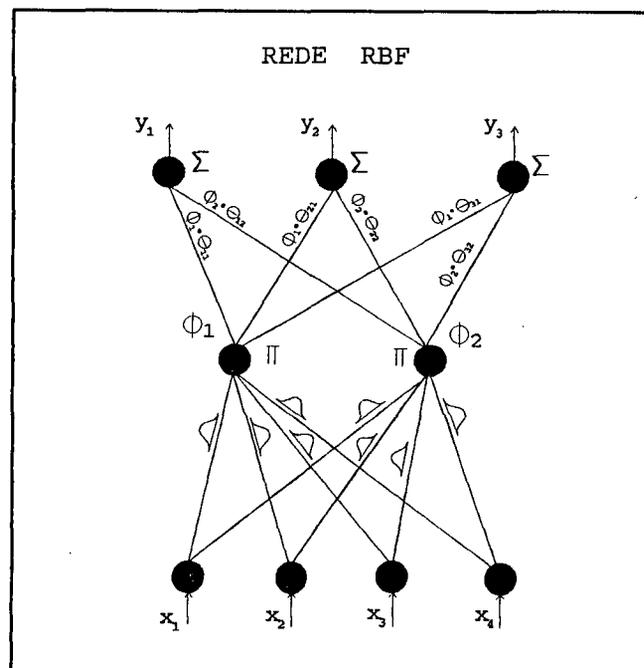


Fig. 5.1 - Rede de Funções com Base Radial - RBF

A rede RBF é formada por três camadas : camada de entrada, camada intermediária (hidden Layer) e camada de saída (Fig . 5.1) .

A camada de entrada distribui os padrões de entrada para os diversos nós da camada intermediária. A saída da camada intermediária é calculada pela função gaussiana:

$$\phi_h(x_i) = e^{\left(\frac{-\Sigma(x_i-c_j)^2}{\delta_h^2}\right)} \quad Eq. 5.1$$

para $h = 1, 2, \dots, H$; e $i = 1, 2, \dots, N$,

onde, x_i é o valor do padrão de entrada na dimensão "i",

c_i é o centro da função gaussiana na dimensão "i",

δ_h é um parâmetro que define a "largura" da função gaussiana para o nó "h",

H é o número de nós na camada intermediária,

N é a dimensão do padrão de entrada.

Os valores na camada de saída são calculados por:

$$T_j = \sum_{h=1}^H \theta_{hj} \cdot \phi_h \quad Eq. 5.2$$

onde θ_{hj} são os pesos da camada de saída, para $j = 1, 2, \dots, J$ e "J" = número de saídas.

O treinamento da rede RBF é realizado em duas etapas :

- a) inicialmente a camada intermediária é treinada por um método como o "k-means clustering" [65] , ou por outro algoritmo como a Aproximação Sucessiva[66];
- b) após o treinamento da camada intermediária, a camada de saída é treinada por um método linear, como mínimos quadrados, ou não linear, como "gradient descent".

O desempenho da rede RBF está diretamente relacionado com a eficiência no treinamento de sua camada intermediária, nominalmente, com a localização dos centros e a largura das funções gaussianas.

5.1.1 - Treinamento da Camada Intermediária

"K-Means Klustering"

O algoritmo de treinamento chamado "k-means clustering" [64] encontra o centro de agrupamentos dos padrões de treinamento e faz a partição destes em subconjuntos de tal maneira a serem associados ao grupo com centro mais próximo. Cada centro de grupo é associado a um dos nós da camada intermediária. O algoritmo encontra um mínimo local no quadrado da distância euclidiana total ($E_{k\text{-means}}$) entre os pontos de treinamento designados a um grupo e os centros de todos os grupos (denominados \hat{x}_h):

$$E_{k\text{-means}} = \sum_{h=1}^H \sum_{k=1}^K B_{hk} \|\hat{x}_h - x_k\|^2 \quad \text{Eq. 5.3}$$

onde, B_{hk} é a função de pertinência, que é uma matriz de ordem $H \times K$. H é o número de grupos (ou nós) na camada intermediária definido a priori e K é o número de padrões de treinamento. Cada coluna de B_{hk} representa um padrão de treinamento (k) e cada linha representa um grupo (classe) (h). Cada coluna tem um único valor "1" na linha correspondente ao grupo mais próximo aquele padrão de treinamento, e zero em todas as outras linhas.

No início do treinamento, o centro de cada grupo é inicializado com um valor igual a um padrão de treinamento escolhido aleatoriamente entre o conjunto de todos eles. Em seguida, cada padrão de treinamento é associado ao grupo mais próximo. Quando todos os padrões tiverem sido associados, a posição média entre eles e seus respectivos grupos é encontrada, e o centro movido para aquela posição.

Uma vez que os centros tenham sido estabelecidos, a largura " δ_h " para cada nó, em cada dimensão pode ser determinada. O objetivo é estabelecer valores de δ que cubram todos os padrões de treinamento, permitindo um ajuste suave pela camada de saída. Para conseguir isto, cada padrão de treinamento deve ativar não apenas o próprio nó, como também, em algum grau, pelo menos um outro nó próximo. Assim, a largura " δ " dos nós deve ser maior que a maior distância entre os padrões de treinamento e o centro do grupo a que pertencem. Porém, o valor de " δ " deve ser mantido a um valor mínimo para manter a influência do nó apenas nas redondezas, evitando que ele tenha ativação alta em regiões distantes daquela que ele representa. Uma largura apropriada é determinada pela eurística dos "P" vizinhos mais próximos ("P Nearest Neighbor") :

$$\delta_h = \sqrt{\frac{1}{P} \sum_{j=1}^P \|\hat{x}_h - \hat{x}_j\|^2} \quad \text{Eq. 5.4}$$

onde \hat{x}_j é o centro dos "P" grupos vizinhos mais próximos ao nó com centro em \hat{x}_h . "P" deve ser estabelecido segundo cada problema, mas uma regra usual é fazer $P = 2$ [65].

A figura 5.2 mostra um exemplo dos agrupamentos e o formato das funções de base radial para um padrão com dimensão = 2.

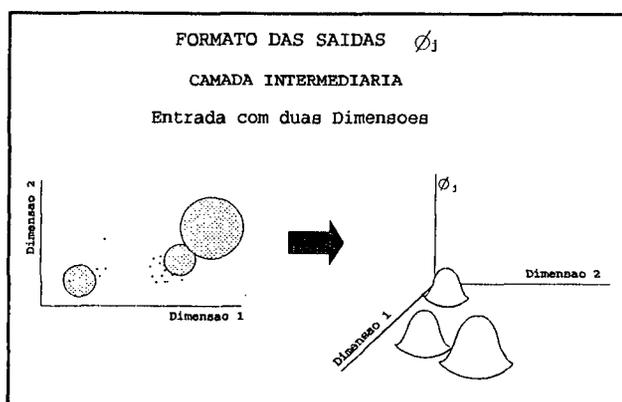


Fig. 5.2 - Treinamento Camada Intermediária - "K-Means Clustering"

O treinamento da rede RBF pelo método "k-means cluster" é tipicamente 10 vezes mais

rápido que a rede de "backpropagation" , além de não ser sensível a ordem de apresentação dos padrões de treinamento. Porém, ainda é necessário que o número de nós na camada intermediária seja definido a priori. Esta necessidade não é muito crítica, uma vez que a largura das funções de base radial (largura dos nós) é calculada pela distância entre elas, compensando em parte um possível excesso ou falta de nós.

O método "k-means clustering" ainda necessita de várias interações para o treinamento da camada intermediária.

Aproximação Sucessiva

No método de aproximação sucessiva, apresentado por Linkens & Nie [66] , os padrões são apresentados para treinamento apenas uma vez, com a camada intermediária se atualizando totalmente a cada apresentação. Este método aumenta em muito a velocidade de treinamento da rede, pois a necessidade de interação será apenas para o treinamento da camada de saída (se treinada por "gradient descent"), ou pode ser realizado em apenas duas apresentações dos padrões de treinamento (se treinada por mínimos quadrados).

Neste método, quando um padrão é apresentado para treinamento, o algoritmo :

- a) calcula o valor de saída de todos os nós ϕ_h da camada intermediária;
- b) acha o valor ϕ_h que tenha resposta máxima ($\phi_{h_{max}} = \max(\phi_h)$, para $h = 1, \dots, H$);
- c) compara o nó vencedor a um valor de corte $0 < \phi_0 < 1$:
 - se $\phi_{h_{max}} \geq \phi_0$, $\phi_{h_{max}}$ é o nó vencedor
 - se $\phi_{h_{max}} < \phi_0$, abre um novo nó
- d) se o nó $\phi_{h_{max}}$ é vencedor, modifica os parametros do nó segundo as equações :

$$n^{h_{max}} = n^{h_{max}} + 1 ; \quad \alpha^{h_{max}} = 1/n^{h_{max}} ;$$

$$c_i^{hmax} = c_i^{hmax} + \alpha^{hmax} \cdot [x_i - c_i^{hmax}] ; \text{ para } i = 1, \dots, N$$

onde "n" é o número de vezes que o nó hmax foi treinado; e

e) se um novo nó é criado, então $H = H+1$, $n_{H+1} = 1$, $c_i = x_i$, para $i = 1, \dots, N$.

A dinâmica de aprendizado será melhor explicada através do exemplo a seguir. Para este exemplo, o nível de corte " ϕ_0 " será estabelecido pequeno, de tal maneira a não haver a criação de nenhum outro nó após o primeiro padrão de treinamento ter sido apresentado. Para melhor visualização, apenas o treinamento na dimensão 8 (Fig 5.3) será ilustrado.

Inicialmente, o padrão $X = \{0,5; 0,35; 0,37; 0,48; 0,57; 0,65; 0,58; 0,52\}$ é apresentado para treinamento. Como a rede ainda não possui nós na camada intermediária, ela abre um nó centrado sobre este padrão (Fig. 5.3).

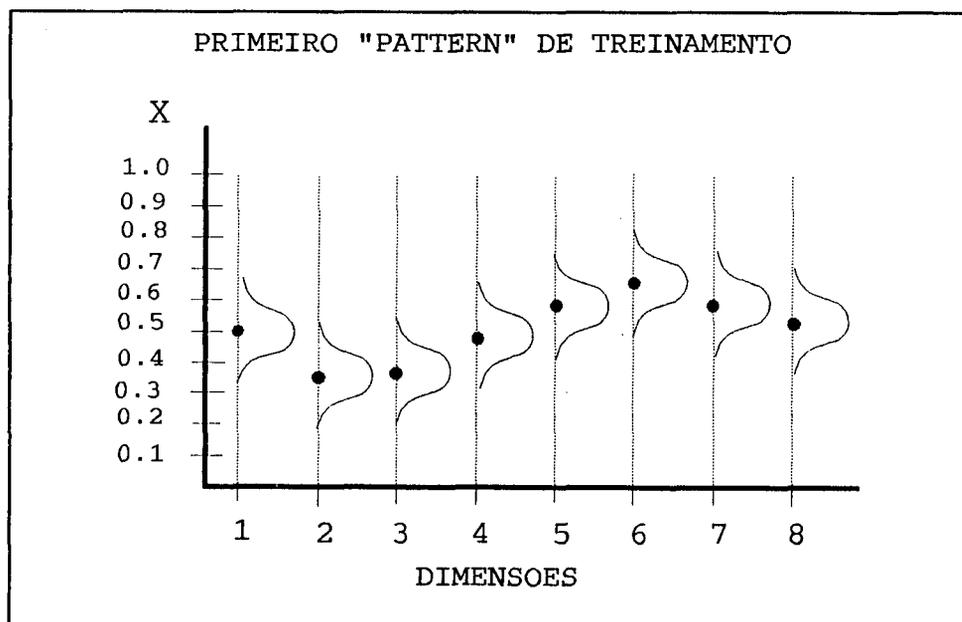
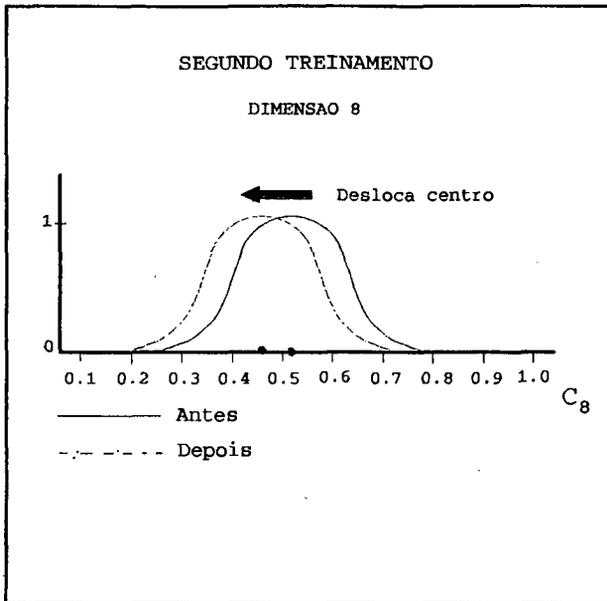


Fig. 5.3 - Posição do Primeiro Nó Aberto com o Primeiro padrão de Treinamento

Um segundo padrão, com valor $x_8 = 0,40$ é apresentado para treinamento. Assim, $n = 2$ e $\alpha = 1/2$, com a posição do centro do nó deslocada para $c_8 = 0,52 + (0,5 \cdot (0,40 - 0,52)) = 0,46$ (Fig. 5.4).



5.4 - Segundo Treinamento

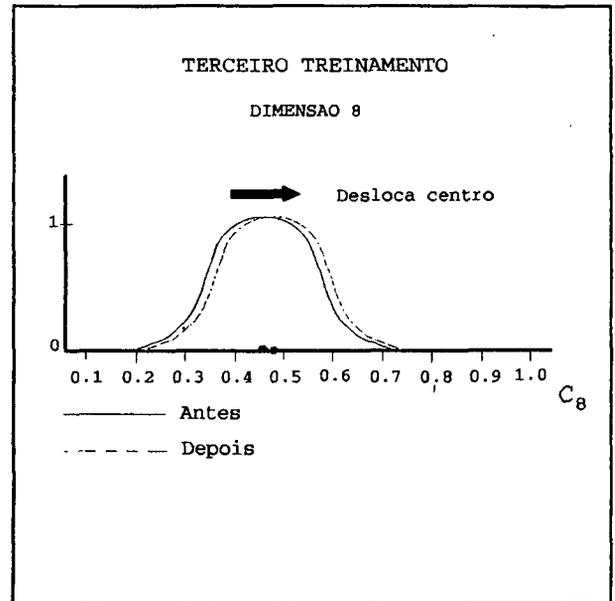


Fig. 5.5 - Terceiro Treinamento

Um terceiro padrão, com valor $x_8 = 0,52$ é apresentado. Neste caso $n = 3$, $\alpha = 1/3$ e a posição do nó é deslocada para $c_8 = 0,46 + (0,333 (0,52 - 0,46)) = 0,48$ (Fig. 5.5).

Nota-se que o centro do nó se desloca uma distância inversamente proporcional ao número de vezes que o nó foi treinado anteriormente, fazendo com que quanto maior o número de padrões de treinamento, mais insensível a "outliers" se torna a rede. Por exemplo, se um padrão

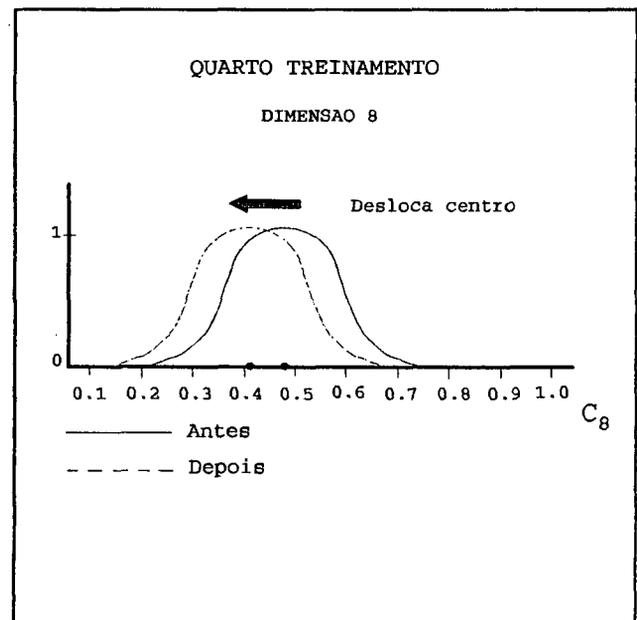


Fig. 5.6 - Quarto Treinamento

com $x_8 = 0,2$ for apresentado, $n = 4$, $\alpha = 0,25$, e o nó se deslocará para $c_8 = 0,48 + (0,25 (0,20 - 0,48)) = 0,41$. O nó se posiciona no centro de massa dos padrões de treinamento (Fig. 5.6).

Se a ordem dos padrões no exemplo acima fosse alterada para $x_8 = 0,20 \ 0,52 \ 0,40 \ 0,52$ o treinamento se comportaria da seguinte maneira :

- 1) $x_8 = 0,20$, $n = 1$, $\alpha = 1$, $c_8 = 0,20$;
- 2) $x_8 = 0,52$, $n = 2$, $\alpha = 0,5$, $c_8 = 0,36$;
- 3) $x_8 = 0,40$, $n = 3$, $\alpha = 0,333$, $c_8 = 0,373$;
- 4) $x_8 = 0,52$, $n = 4$, $\alpha = 0,25$, $c_8 = 0,41$.

Verifica-se que a posição final do nó é a mesma, mostrando que o método de aproximação sucessiva não apenas é insensível a "outliers", como também é insensível a ordem de apresentação dos padrões. Isto é uma qualidade muito importante, não necessitando cuidados extremos durante o treinamento da rede.

O método de aproximação sucessiva necessita que a largura dos nós (δ) seja definida a priori. Esta definição não é muito crítica, pois o treinamento compensará diferenças na largura com um diferente número de nós e pelo treinamento da camada de saída.

A adoção do método de aproximação sucessiva permite acelerar mais ainda o aprendizado da rede de RBF, pois o treinamento da camada intermediária é efetuado com apenas uma passada.

5.1.2 - Vantagens e Desvantagens

Vantagens

- A velocidade de treinamento é em geral uma ordem de grandeza mais rápida [64].
- Tem 3 camadas, não tendo o problema da rede de "backpropagation" para a determinação do número de camadas intermediárias.
- Menos sensível a determinação dos parâmetros da rede.

- É funcionalmente equivalente a um sistema com lógica difusa [63], podendo ser aplicada com mais naturalidade aos problemas reais.

Desvantagens

- Sua aplicação a casos reais não está ainda bem estudada, principalmente no que diz respeito a problemas da qualidade.
- Apesar da rapidez de aprendizado, mesmo com o método de aproximação sucessiva, a rede de RBF necessita de um aprendizado "off-line". Se um novo problema ou tipo de padrão tiver que ser aprendido pela rede após seu treinamento, é necessário repassar todos os exemplos de treinamento junto com os novos padrões para evitar com segurança que os padrões anteriores não sejam esquecidos.

Neste trabalho, a rede RBF terá sua camada intermediária treinada pelo método de aproximação sucessiva e a camada de saída pelo método "gradient descent".

5.2 - TEORIA DE RESSONÂNCIA ADAPTATIVA

A teoria de Ressonância Adaptativa, ou ART, foi introduzida como uma teoria sobre o processamento cognitivo de informações no cérebro humano [67]. Essa teoria levou ao desenvolvimento de uma série de modelos de redes neurais capazes de um aprendizado não supervisionado para classificação de padrões em tempo real. Os modelos nessa família compreendem: a rede ART-1 [68], que pode aprender a categorizar padrões de entrada binários apresentados em ordem arbitrária; a rede ART-2 [69], que pode aprender a categorizar padrões de entrada analógicos ou binários; e a rede ART-3 [70][71], que pode fazer uma busca paralela, ou teste de hipóteses, em códigos com reconhecimento distribuído.

A rede ART-1 foi usada no projeto de uma rede com arquitetura hierárquica chamada ARTMAP [72] que tem aprendizado instantâneo, estabelecendo uma associação entre padrões binários de entrada com padrões binários de saída.

A rede Fuzzy-ART [73] é uma generalização da rede ART-1 para torná-la capaz de reconhecer padrões binários e analógicos, sendo muito mais simples que a rede ART-2.

O desenvolvimento da rede Fuzzy-ART permitiu a adaptação da rede ARTMAP para a utilização de padrões analógicos tanto na entrada como na saída. Esta nova versão é chamada Fuzzy-ARTMAP [74] e será discutida posteriormente.

5.2.1 - Rede Fuzzy-ART

A rede Fuzzy-ART [73] é formada por duas camadas (Fig.5.7): Uma camada de entrada, com número de nós "2p" igual ao dobro da dimensão dos padrões de entrada. Uma camada de saída, formada de tantos nós "j" quantos necessários para classificar os padrões de entrada.

A camada de saída inicia com um número "q" de nós não comprometidos (não treinados).

A ligação entre as camadas de entrada e saída é realizada por um conjunto de pesos W_{ij} que tem valor 1 quando o nó "j" não está comprometido. Assim, inicialmente $W_{ij} = 1$ para $i = 1, 2, \dots, 2p$ e $j = 1, 2, \dots, q$.

Quando um padrão de entrada "X" é apresentado à camada de entrada (F1), os valores dos nós da camada de saída (F2) são calculados pela equação :

$$T_j = \frac{\sum_{i=1}^{2p} \min(x_i, W_{ij})}{\alpha + \sum_{i=1}^{2p} W_{ij}} \quad \text{Eq. 5.5}$$

onde: x_i , $i = 1, 2, \dots, 2p$ são as dimensões dos padrões de entrada normalizados por uma técnica chamada codificação complementar, com $0 \leq x_i \leq 1$;

w_{ij} são os pesos conectando o nó "i" da camada de entrada com o nó "j" da camada de saída;

α é um valor pequeno > 0 ;

T_j é o valor dos nós "j" na camada de saída, com $j = 1, 2, \dots, N$; e

N é o número de nós comprometidos na camada de saída;

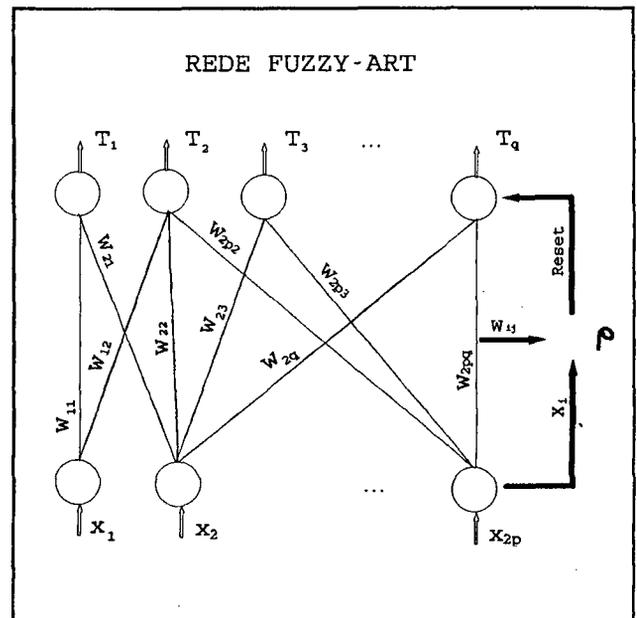


Fig. 5.7 - Rede Fuzzy-ART

A rede escolhe o nó "T_j" com o maior valor de saída e aplica o critério de comparação:

$$A = \frac{\sum_{i=1}^{2p} \min(x_i, w_{ij})}{\sum_{i=1}^{2p} x_i} \geq \rho \quad \text{Eq. 5.6}$$

onde ρ é o nível de vigilância da rede, fixado pelo usuário ($0 \leq \rho \leq 1$).

Se o critério de comparação não for válido ($A < \rho$), a rede desativa o nó vencedor "J" e escolhe outro nó. O processo se repete até que um nó satisfaça o critério de comparação. Se nenhum nó comprometido satisfizer o critério de comparação, a rede escolherá um nó não comprometido e abrirá uma outra classe na saída, fazendo $N = N+1$.

O nível de vigilância (ρ) determina quão discriminatória é a rede. Pequenos valores de ρ permitirão que um grande número de padrões sejam assimilados pelo mesmo nó de saída. Um valor de ρ grande tornará a rede mais discriminatória, aumentando também o número de nós.

Após um nó "J" adequado tenha sido escolhido, os pesos w_{ij} são alterados de acordo com a equação:

$$w_{ij} = \beta \cdot \min(x_i, w_{ij}) + (1 - \beta) w_{ij} \quad \text{Eq. 5.7}$$

Para um aprendizado mais eficiente em situações onde as entradas são afetadas por ruído, é usual estabelecer $\beta = 1$ quando o nó ainda não está comprometido, fazendo $w_{ij} = x_i$ e $\beta < 1$ nos treinamentos seguintes [73]. A rede opera em modo de aprendizado rápido quando $\beta = 1$.

Para evitar um problema de proliferação de classes [73], é necessário que as entradas sejam normalizadas. Uma regra de normalização chamada codificação complementar [73] resolve este problema enquanto preserva a informação sobre a amplitude dos padrões. Esta codificação é efetuada dobrando-se o número de entradas, fazendo $x_{p+i} = 1 - x_i$, para $i = 1, 2, \dots, p$.

5.2.2- Rede Fuzzy-Artmap

A rede Fuzzy-Artmap [74] é uma generalização da rede binária Artmap [72]. Ela é capaz de um aprendizado supervisionado incremental, atualizando-se durante a operação sem "esquecer" o que já aprendeu anteriormente. A rede Fuzzy-Artmap pode ser empregada para classificação e/ou associação de padrões binários ou analógicos de entrada e saída com dimensão arbitrária.

A rede Fuzzy-Artmap incorpora dois módulos Fuzzy-ART: ARTa e ARTb, conectados através de um módulo inter-ART, F^{ab} , chamado Módulo de Ligação (Fig. 5.8).

Os módulos ARTa e ARTb funcionam da mesma maneira que a rede Fuzzy-ART, exceto pela interação com o módulo de ligação. O módulo de ligação é usado para fazer associações preditivas entre categorias nas redes ARTa e ARTb e para executar a pesquisa de similaridade, onde o nível de vigilância (ρ_a) do módulo ARTa aumenta em resposta a um erro de previsão no módulo ARTb. A pesquisa de similaridade reconhece a estrutura de uma categoria de tal maneira que o erro de previsão não será repetido em apresentações subsequentes da mesma entrada. A operação da rede Fuzzy-Artmap é apresentada a seguir.

Definições:

- As entradas das redes ARTa e ARTb são normalizadas pelo método de classificação complementar [73];
- $\mathbf{x}^a = (x_1^a \dots x_{2Pa}^a)$ representa o vetor das entradas na camada F_1^a ;
- $\mathbf{y}^a = (y_1^a \dots y_{Na}^a)$ representa o vetor das saídas da camada F_2^a ;
- $\mathbf{w}_j^a = (w_{j1}^a, w_{j2}^a, \dots, w_{j2Pa}^a)$ representa o j-ésimo vetor de pesos da rede ARTa;
- $\mathbf{x}^b = (x_1^b \dots x_{2Pb}^b)$ representa o vetor das entradas na camada F_1^b ;
- $\mathbf{y}^b = (y_1^b \dots y_{Nb}^b)$ representa o vetor das saídas da camada F_2^b ;
- $\mathbf{w}_k^b = (w_{k1}^b, w_{k2}^b, \dots, w_{k2Pb}^b)$ representa o k-ésimo vetor de pesos da rede ARTb;

- $\mathbf{x}^{ab} = (x_1^{ab} \dots x_{N_b}^{ab})$ representa as saídas do módulo de ligação para a rede ARTb;
- $\mathbf{w}_j^{ab} = (w_{j1}^{ab}, w_{j2}^{ab}, \dots, w_{jN_b}^{ab})$ representa o vetor de pesos ligando o j -ésimo nó em F_2^a à camada F_2^b ; e
- $\mathbf{x}^a, \mathbf{y}^a, \mathbf{x}^b, \mathbf{y}^b, \mathbf{x}^{ab}$ são igualados a zero no intervalo entre duas apresentações.

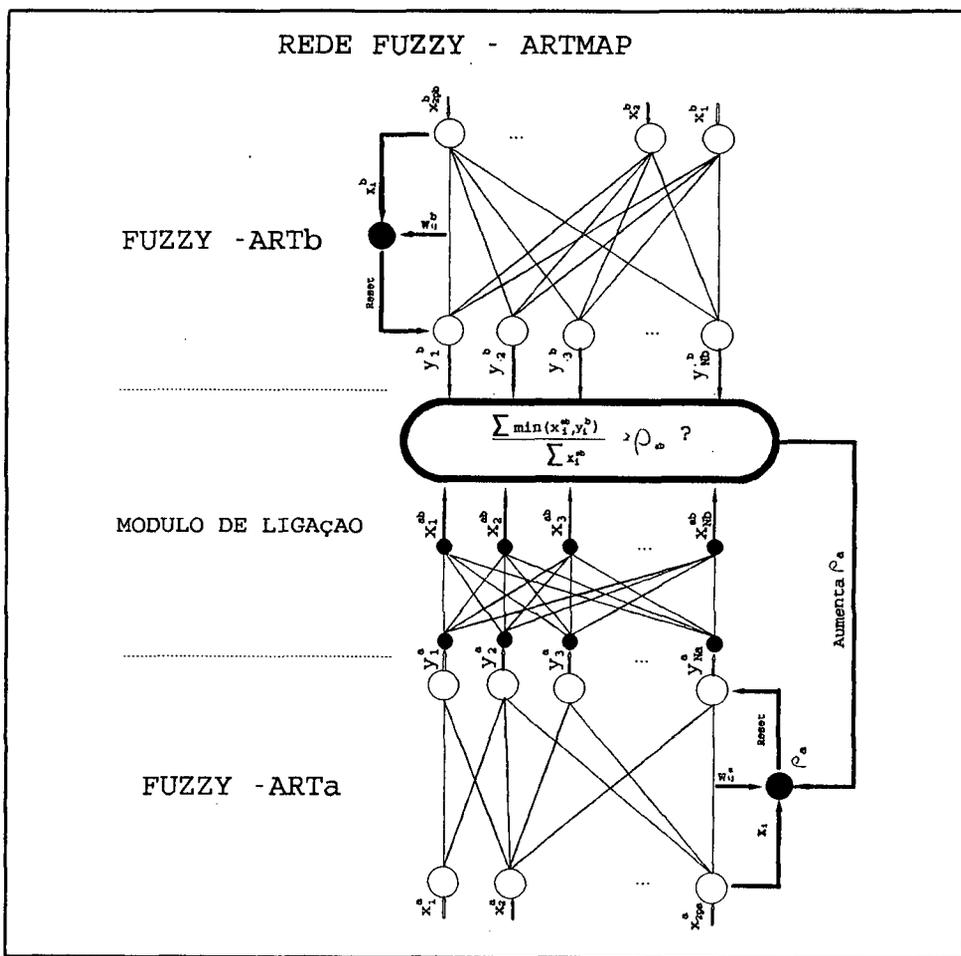


Fig. 5.8 - Rede Fuzzy-Artmap

As camadas de saída de ARTa e ARTb (F_2^a e F_2^b) estão ativas quando pelo menos um de seus nós está ativo devido a apresentação de uma entrada. O módulo de ligação F^{ab} é ativado (isto é, tem saída $\mathbf{x}^{ab} \neq 0$) quando ao menos uma das camadas de saída de ARTa e ARTb (F_2^a

e F_2^b) estiver ativa. A saída x^{ab} do módulo de ligação é dada pela seguinte regra :

$$x_i^{ab} = \begin{cases} \min(y_i^b, w_{ji}^{ab}) & \text{para } i = 1, \dots, Nb, \text{ se o } J\text{-ésimo nó da camada } F_2^a \text{ estiver ativo e } F_2^b \\ \text{também estiver ativa.} \\ w_{ji}^{ab}, & \text{para } i = 1, \dots, Nb, \text{ se o } J\text{-ésimo nó da camada } F_2^a \text{ estiver ativo e } F_2^b \\ \text{estiver inativa.} \\ y_i^b, & \text{para } i = 1, \dots, Nb, \text{ se a camada } F_2^a \text{ está inativa e } F_2^b \text{ está ativa.} \\ 0, & \text{se tanto a camada } F_2^a \text{ quanto } F_2^b \text{ estiverem inativas.} \end{cases}$$

Como nas camadas F_2^a e F_2^b só haverá um nó ativo de cada vez, x^{ab} será igual a 0 se a previsão executada por w_j^{ab} não for confirmada por y^b .

Quando um padrão é apresentado a camada F_1^a , a rede ARTa seleciona um nó vencedor "J" que melhor represente aquele padrão. A esta altura ρ_a é igual à um valor base $\bar{\rho}_a$.

Execução

Caso a rede esteja em execução, a camada F_2^b estará inativa e a saída do módulo de ligação será dado por $x^{ab} = w_j^{ab}$. Se o nó "J" de F_2^a for um nó comprometido, w_j^{ab} terá apenas um valor $w_{jk}^{ab} \neq 0$, e a previsão da rede Fuzzy-Artmap será igual a matriz $w_k^b = (w_{k1}^b, w_{k2}^b, \dots, w_{k2pb}^b)$. Se o nó "J" de F_2^a for um nó não comprometido, w_j^{ab} terá todos os valores $w_{jk}^{ab} = 1$ (para $k = 1, \dots, Nb$), significando que a rede não consegue fazer uma previsão com os casos apresentados para treinamento. Neste caso, a rede pode ser treinada imediatamente através da apresentação do padrão de saída correto na rede ARTb, ativando a camada F_2^b .

Treinamento

Caso a rede esteja sendo treinada, e a camada F_2^b esteja ativa, a saída do módulo de ligação será dado por $x_i^{ab} = \min(y_i^b, w_{ji}^{ab})$ para $i = 1, \dots, Nb$.

O módulo de ligação faz a pesquisa de similaridade, comparando se :

$$A_{ab} = \frac{\sum_{i=1}^{Nb} \min(x_i^{ab}, y_i^b)}{\sum_{i=1}^p x_i^{ab}} \geq \rho_{ab} \quad \text{Eq. 5.8}$$

onde ρ_{ab} é o nível de vigilância do módulo de ligação.

Se $A_{ab} < \rho_{ab}$, o nível de vigilância da rede ARTa (ρ_a) é aumentado para um valor um pouco maior que :

$$\rho_a > \frac{\sum_{i=1}^{2p} \min(x_i^a, w_{ij}^a)}{\sum_{i=1}^{2p} x_i^a} \quad \text{Eq. 5.9}$$

onde "J" é o índice do nó ativo em F_2^a .

Quando isto ocorre, a rede ARTa desativará o nó "J" que tinha escolhido anteriormente, procurando por outro nó que satisfaça o seu critério de similaridade. Este novo nó é novamente testado pelo módulo de ligação. Este processo se repete até que um nó que tenha uma previsão correta seja escolhido ou, se este nó não existir, a rede abrirá um novo nó que aprenderá a fazer a previsão correta. Este aprendizado é efetuado através do treinamento do módulo de ligação.

Inicialmente os pesos do módulo de ligação $w_{jk}^{ab} = 1$. Durante o tempo em que o novo nó da rede ARTa está ativo, o vetor w_j^{ab} é igualado ao vetor x^{ab} . Uma vez que o nó "J" aprende a prever a categoria "K" da rede ARTb, esta associação é permanente, isto é, $w_{JK}^{ab} = 1$ para sempre.

O valor de vigilância ρ_a é retornado para $\rho_a = \bar{\rho}_a$ no intervalo entre a apresentação de dois padrões consecutivos.

Se $A_{ab} \geq \rho_{ab}$, os pesos da rede ARTa são atualizados pela equação :

$$w_{ij} = \beta \min(x_i, w_{ij}) + (1-\beta)w_{ij} \quad \text{Eq. 5.10}$$

onde, $\beta = 1$ para o caso de aprendizado rápido.

Aprendizado a Nível de Nó

Para melhor entender a dinâmica, e consequentemente as vantagens e desvantagens da rede Fuzzy-Artmap, é necessário uma análise do mecanismo de aprendizado a nível de nó na rede ARTa. Por simplicidade, a análise será efetuada supondo que os padrões de entrada sejam unidimensionais. Desta maneira, o valor de cada nó na camada de saída da rede ARTa se reduz a equação :

$$T_j = \frac{\min(x^a, w_j^a)}{w_j^a} \quad \text{Eq. 5.11}$$

Similarmente, a avaliação de similaridade será dada pela equação:

$$A = \frac{\min(x^a, w_j^a)}{x^a} \quad \text{Eq. 5.12}$$

Supondo que um valor de entrada $x^a = 0,8$ seja inicialmente apresentado para treinamento, a rede ARTa abrirá um nó, com um peso $w_1^a = 0,8$. A função de pertinência que regula o valor da saída T_1 será, em princípio, da forma mostrada na figura 5.9.

Supondo que $\rho_a = 0,85$, o nó será desativado para qualquer valor de $x^a > 0,94$, e então a função de pertinência terá a forma da figura 5.10.

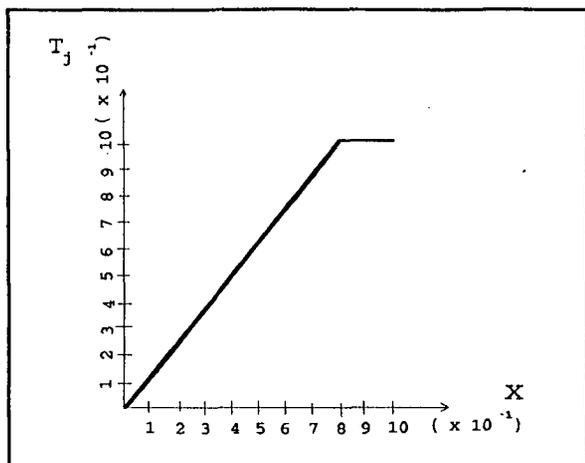


Fig. 5.9 - Função de Pertinência Parcial

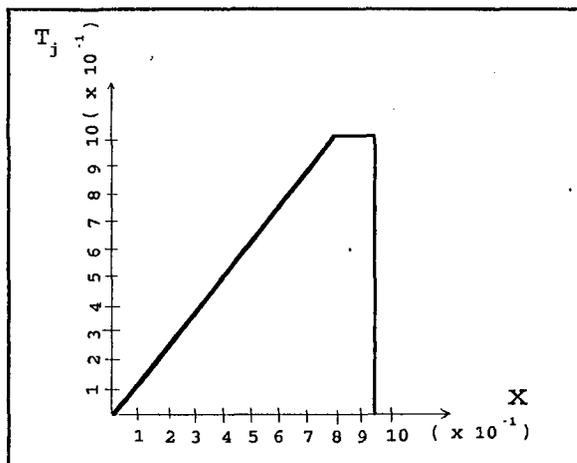


Fig. 5.10 - Função de Pertinência com Valor de Corte

Supondo que em seguida é apresentada a rede um valor de entrada $x^a = 0,6$, então o nó não será desativado, pois $A = 1$ é maior que $\rho_a = 0,85$. Assim, o peso do nó será atualizado pela equação :

$$w_j^a = (\beta * \min(x^a, w_j^a)) + ((1 - \beta) * w_j^a) \quad \text{Eq. 5.13}$$

Se $\beta = 0,8$, $w_1^a = 0,8 * 0,6 + 0,2 * 0,8 = 0,64$. Com este novo valor de w_1^a , a rede desativará o nó 1 sempre que $x^a > 0,753$ (Fig. 5.11).

A figura 5.11 demonstra que o peso do nó vai se deslocando para a esquerda, isto é, vai ficando com um valor cada vez menor e mais agudo. Isto faz com que valores menores de x^a ocasionem um valor de saída T_1 maior que anteriormente.

Isto leva a conclusão que para padrões com dimensão maior que um, a dimensão em que w_j^a mais se modifica adquire um peso maior, tornando-se mais especializada. Por outro lado, como o nó vai ficando mais agudo, ele será desativado para valores menores de x^a . Se para o exemplo acima, a entrada $x^a = 0,8$ for novamente apresentada, o nó 1 será desativado e um outro nó será aberto. Assim, um mesmo valor de entrada, que inicialmente foi classificado pelo nó 1, poderá em uma segunda apresentação ser classificado por outro nó.

Se um novo valor de entrada $x^a = 0,5$ for apresentado, o peso w_1^a será 0,528.

Para fins de exemplo, o ponto em que x^a causa a desativação do nó será definido como sendo ponto de corte do nó. Neste caso, para a última apresentação anterior, o ponto de corte será 0,621.

Se a ordem de apresentação dos pontos no exemplo anterior for alterada para

0,8 , 0,5 e 0,6 , tem-se os seguintes valores para o peso w_1^a e para o ponto de corte :

- $x^a = 0,8 \rightarrow w_1^a = 0,80$, ponto de corte = 0,94;
- $x^a = 0,5 \rightarrow w_1^a = 0,56$, ponto de corte = 0,659;
- $x^a = 0,6 \rightarrow w_1^a = 0,56$, ponto de corte = 0,659 (não altera).

Comparando o peso final anterior $w_1^a = 0,528$ com o do exemplo acima $w_1^a = 0,56$, verifica-se que a rede Fuzzy - ART é sensível há ordem de apresentação dos padrões de treinamento.

Além disso, a rede Fuzzy-ART é sensível a "outliers" com valores pequenos . O exemplo a seguir ilustra o problema.

Se $\rho_a = 0,6$, e a $x^a = (0,5; 0,8; 0,7; 0,75)$, teremos os seguintes valores para w_1^a e para o ponto de corte:

- $x^a = 0,50 \rightarrow w_1^a = 0,50$, ponto de corte = 0,833;
- $x^a = 0,80 \rightarrow w_1^a = 0,50$, ponto de corte = 0,833 (não altera);
- $x^a = 0,70 \rightarrow w_1^a = 0,50$, ponto de corte = 0,833 (não altera);
- $x^a = 0,75 \rightarrow w_1^a = 0,50$, ponto de corte = 0,833 (não altera).

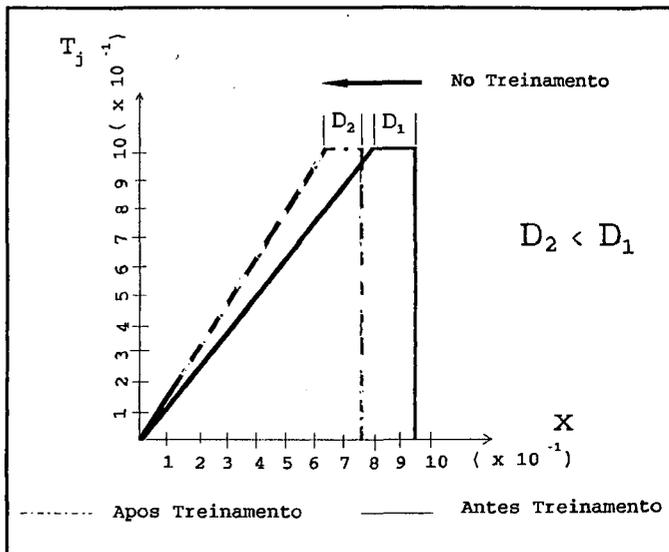


Fig. 5.11 - Evolução do Nó no Treinamento

Os valores de x^a posteriores ao primeiro não tiveram nenhuma influência no treinamento. No exemplo acima faria mais sentido se a rede abrisse um nó para classificar o valor $x^a = 0,5$, e outro para classificar os pontos $x^a = (0,80; 0,70; 0,75)$. O problema de "outliers" pode ser grave em situações onde os padrões de entrada estão sujeitos a ruído, como é o caso do controle estatístico de processos. Este problema pode ser diminuído com um acréscimo do valor de ρ_a . Se no exemplo acima $\rho_a = 0,85$, a rede abrirá dois nós :

$$x^a = 0,50 \rightarrow w_1^a = 0,50, \text{ ponto de corte} = 0,588 \quad (\text{ abre nó 1});$$

$$x^a = 0,80 \rightarrow w_1^a = 0,80, \text{ ponto de corte} = 0,941 \quad (\text{ abre nó 2});$$

$$x^a = 0,70 \rightarrow w_1^a = 0,72, \text{ ponto de corte} = 0,847 \quad (\text{ treina nó 2});$$

$$x^a = 0,75 \rightarrow w_1^a = 0,72, \text{ ponto de corte} = 0,847 \quad (\text{ não altera});$$

classificando $x^a = 0,5$ no nó 1 e os outros no nó 2.

O uso de um grande valor para ρ_a diminui o problema de "outliers", porém acarreta um aumento do número de nós. Além disso, no caso da rede Fuzzy-Artmap, o aumento do nível de vigilância (ρ_a) ocasionará um aumento do número de padrões indeterminados durante a operação da rede (rede perde capacidade de generalização).

5.2.3 - Vantagens e Desvantagens

Pela análise efetuada conclui-se que a rede Fuzzy-Artmap possui as seguintes vantagens e desvantagens:

Vantagens

- Pode aprender instantaneamente durante sua operação "on line".
- Tem capacidade de se especializar, determinando as dimensões das entradas que mais representam um tipo de padrão.

Desvantagens

- Apresenta problema de sensibilidade a ordem de apresentação dos padrões de treinamento e a "outliers" com valores pequenos.
- Ao contrário da rede RBF, a rede Fuzzy-Artmap possui uma função de pertinência gradual apenas para valores menores que o valor do nó. Valores maiores são tratados pela abertura de outros nós. Isto faz com que o processo de inferência não siga uma transição gradual entre as classes que se quer identificar.

A inclusão de funções de base radial (RBF) na rede ARTa da rede Fuzzy-Artmap, utilizando o método de treinamento por aproximação sucessiva, elimina o problema de outliers e sensibilidade a ordem de apresentação dos padrões de treinamento, além de aumentar a capacidade de generalização da rede.

Esta nova configuração, chamada Rede RBF-FUZZY-ARTMAP será apresentada no próximo capítulo.

5.3 - NOVA REDE NEURAL PROPOSTA (REDE RBF-FUZZY ARTMAP)

5.3.1 - Descrição

A rede RBF tem a vantagem de não ser muito sensível a ordem de apresentação dos padrões de treinamento. Porém, ela necessita de uma fase de aprendizado separada da fase de operação, o que limita sua aplicação em sistemas de CEP automatizados, onde há necessidade de um aprimoramento contínuo do sistema, para identificação de novos tipos de padrões.

A capacidade de aprendizado instantâneo e incremental, apresentada pela rede Fuzzy-Artmap, a torna uma forte candidata à aplicação em sistemas de CEP automatizados. Porém, sua sensibilidade a ordem de apresentação dos padrões de treinamento é uma limitação séria, que deve ser superada para torná-la útil em aplicações reais.

A rede RBF-Fuzzy Artmap, desenvolvida a seguir elimina as limitações apresentadas pelas redes RBF e Fuzzy Artmap. Ela utiliza as funções de base radial, e o mecanismo de treinamento por aproximação sucessiva, da camada intermediária da rede RBF, em substituição a rede ARTa da rede Fuzzy-Artmap. A figura 5.12 mostra a arquitetura da nova rede.

Quando um padrão é apresentado a rede RBF-Fuzzy Artmap então:

- a) as saídas ϕ_h ($h = 1, 2, \dots, N_a$) são calculadas da mesma maneira que a rede RBF;
- b) acha o valor ϕ_h que tenha resposta máxima;
- c) compara a saída do nó vencedor ao nível de vigilância ρ_a ;

- se $\phi_{h_{max}} \geq \rho_a$, $\phi_{h_{max}}$ é o nó vencedor

- se $\phi_{h_{max}} < \rho_a$, abre um outro nó com o centro nas coordenadas do padrão de entrada

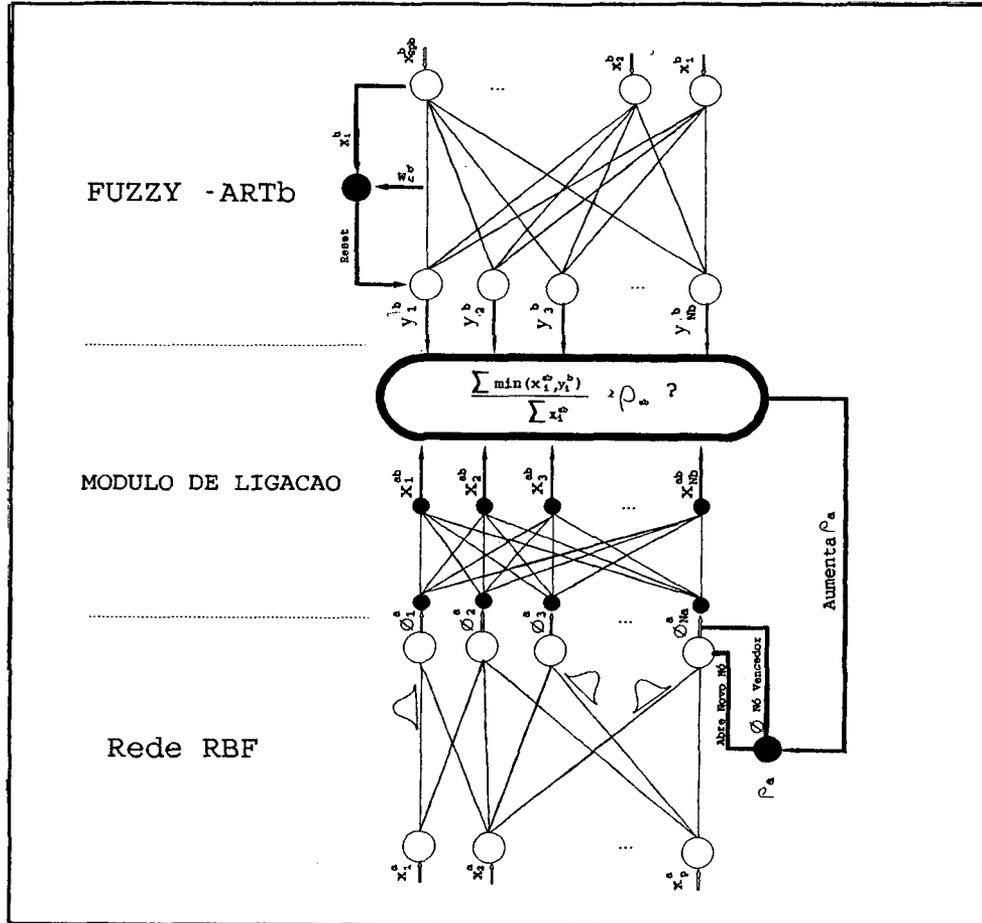


Fig. 5.12 - Rede RBF-Fuzzy Artmap

Em seguida a rede calcula a saída do módulo de ligação x^{ab} pela regra:

$$x_i^{ab} = \begin{cases} \min(y_i^b, w_{h_{\max i}^{ab}}) & \text{para } i = 1, \dots, Nb, \text{ se as camadas RBF e } F_2^b \text{ estiverem ativas.} \\ w_{h_{\max i}^{ab}} & \text{para } i = 1, \dots, Nb, \text{ se a camada RBF estiver ativa e } F_2^b \text{ estiver inativa.} \\ y_i^b & \text{para } i = 1, \dots, Nb, \text{ se a camada RBF estiver inativa e } F_2^b \text{ estiver ativa.} \\ 0 & \text{se tanto a camada RBF quanto } F_2^b \text{ estiverem inativas.} \end{cases}$$

Como nas camadas RBF e F_2^b só haverá um nó ativo de cada vez, x^{ab} será igual a 0 se a previsão executada por $w_{h_{\max}^{ab}}$ não for confirmada por y^b .

Execução

Caso a rede esteja em execução, a camada F_2^b estará inativa e a saída do módulo de ligação será dado por $x^{ab} = w_{hmax}^{ab}$. Se o nó "hmax" não for um nó novo, w_{hmax}^{ab} terá apenas um valor $w_{hmaxK}^{ab} \neq 0$, e a previsão da rede Fuzzy-Artmap será igual a matriz $w_k^b = (w_{k1}^b, w_{k2}^b, \dots, w_{k2pb}^b)$. Se o nó "hmax" for um nó novo, w^{ab} terá todos os valores $w_{hmaxk}^{ab} = 1$ (para $k = 1, \dots, Nb$), significando que a rede não consegue fazer uma previsão com os casos apresentados para treinamento. Neste caso, a rede pode ser treinada imediatamente através da apresentação do padrão de saída correto na rede ARTb, ativando a camada F_2^b .

Treinamento

Caso a rede esteja sendo treinada, e a camada F_2^b esteja ativa, a saída do módulo de ligação será dado por $x_i^{ab} = \min(y_i^b, w_{hmaxi}^{ab})$ para $i = 1, \dots, Nb$.

O módulo de ligação faz a pesquisa de similaridade, comparando se:

$$A_{ab} = \frac{\sum_{i=1}^{Nb} \min(x_i^{ab}, y_i^b)}{\sum_{i=1}^p x_i^{ab}} \geq \rho_{ab} \quad Eq. 5.14$$

onde ρ_{ab} é o nível de vigilância do módulo de ligação.

Se $A_{ab} < \rho_{ab}$, o nível de vigilância da rede ARTa (ρ_a) é aumentado a um valor um pouco maior que :

$$\rho_a > \phi_{hmax} \quad Eq. 5.15$$

onde "hmax" é o índice do nó vencedor da rede RBF.

Quando isto ocorre, um novo nó é criado com as coordenadas do padrão de entrada. Este novo nó aprenderá a fazer a previsão correta através do treinamento do módulo de ligação.

Inicialmente os pesos do módulo de ligação w_{hk}^{ab} são iguais a 1. Durante o tempo em que o novo nó da rede RBF estiver ativo, o vetor w_{hmax}^{ab} é igualado ao vetor x^{ab} . Uma vez que o nó "hmax" aprende a predizer a categoria "K" da rede ARTb, aquela associação torna-se permanente, isto é, $w_{hmaxK}^{ab} = 1$ para sempre. O valor de vigilância ρ_a é retornado para $\rho_a = \rho_a^-$ no intervalo entre apresentação de dois padrões consecutivos.

Se $A_{ab} \geq \rho_{ab}$, a camada RBF é treinada pelo método de aproximação sucessiva descrito anteriormente.

5.3.2 Vantagens e Desvantagens

Vantagens:

- Esta arquitetura permite a incorporação das qualidades da rede RBF na rede Fuzzy-Artmap, diminuindo a sensibilidade a ordem dos padrões de treinamento, ao mesmo tempo que mantém a capacidade de aprendizado instantâneo.

Desvantagens:

- A rede RBF-Fuzzy Artmap necessita de um número maior de nós, na camada RBF, do que a rede Fuzzy-Artmap necessita na rede ARTa.

6 - TREINAMENTO E OTIMIZAÇÃO DAS REDES ESTUDADAS

6.1 - ENTRADAS DAS REDES

A escolha das estatísticas e dos valores para compor o vetor de entrada das redes é de extrema importância no seu desempenho. A definição correta das estatísticas que auxiliem na identificação de padrões específicos pode aumentar muito o desempenho de classificação.

Devido a grande sensibilidade que o gráfico de CUSUM tem para os desvios da média, duas das entradas serão então os resultados computacionais deste gráfico:

$$S_0(+) = S_0(-) = 0$$

$$S_t(+) = \text{Max}(0, S_{t-1}(+) + y_t - K)$$

$$S_t(-) = \text{Max}(0, S_{t-1}(-) + y_t + K)$$

onde, y_t é o valor da última amostra medida; e

K é um parâmetro relacionado com a sensibilidade do CUSUM ($K = 0.5$).

Para auxiliar na monitoração da variabilidade do processo (importante para identificação da diminuição da variabilidade, padrões sistemáticos e misturas), utilizou-se a estatística SSQ_t utilizada por Guo & Dooley [57] :

$$SSQ_t = \sum_{i=t_1}^{t_2} y_i^2 \quad \text{Eq. 6.1}$$

Como não existe uma diretriz sobre o tamanho de $N = t_2 - t_1 + 1$, para efeitos de normalização, utilizou-se o mesmo parâmetro utilizado por Guo & Dooley [57], onde $N = t_2 - t_1 + 1 = 25$ é constante, com t_2 sendo a amostra mais recente. A estatística SSQ_t tem então uma

distribuição quiquadrado com 24 graus de liberdade. O valor de N não é um fator crucial neste momento, uma vez que a estatística SSQ_t será comparada com dois limites de controle (superior e inferior) especificados em função do valor de N, que darão as probabilidades de ocorrer um alarme falso.

Lucy-Bouler [58] mostra que o valor da autocorrelação das amostras é eficiente para identificação de padrões sistemáticos e ciclos. Assim, os "lags" de autocorrelação são calculados para entrada da rede pela fórmula:

$$Autocorr_i = \frac{\sum_{t=i}^{t=M-1} y_t \cdot y_{t-i}}{\sum_{t=i}^{t=M-1} y_{t-i}^2} \quad Eq. 6.2$$

onde, $M = 60$ (tamanho da janela de monitoração)

$i = 1, 2, \dots, L$. Onde L é no mínimo igual a metade do período do ciclo mais longo que se quer detectar. Neste caso $L = 10$, pois os padrões de ciclo têm período 20 neste trabalho.

O tamanho da janela de monitoração neste trabalho foi escolhida de maneira arbitrária. Porém, ela não pode ser muito grande, o que inviabilizaria sua utilização em pequenos lotes, nem muito pequena, não oferecendo informações suficientes para a identificação dos padrões. Assim, um tamanho de janela $M = 60$ foi considerado adequado.

Como os padrões aqui estudados podem ter intensidade grande, eles podem ocasionar um ponto fora de controle muito antes das estatísticas acima terem informação suficiente para sinalizar que tipo de padrão está ocorrendo. Assim, deve-se procurar uma composição com

valores individuais, e médias e dispersões desses valores. Seguindo o trabalho de Guo & Dooley [57], a janela de monitoramento é inicialmente dividida em 15 janelas menores, chamadas janelas estatísticas, contendo 4 pontos individuais cada uma. Para cada janela estatística calcula-se a média e o desvio padrão (fig. 6.1). Os dados individuais ajudam na identificação de padrões como mistura e sistemático.

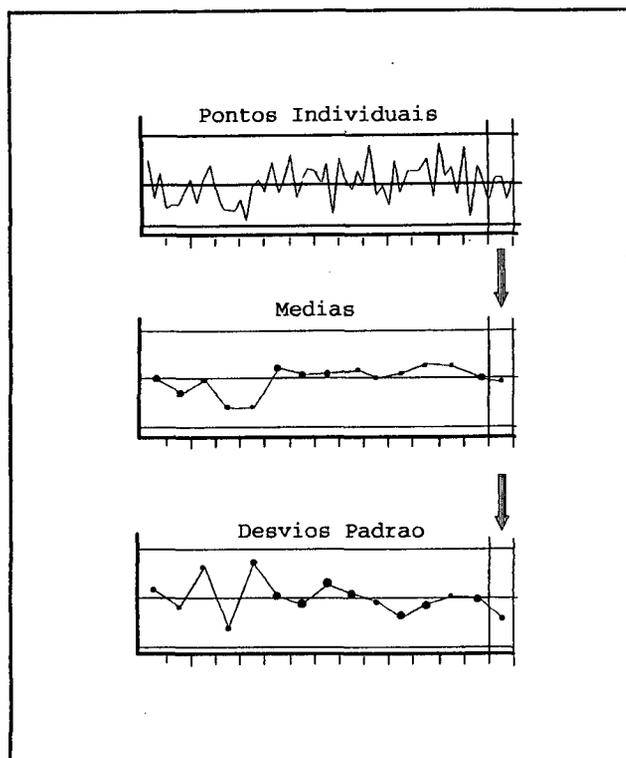


Fig. 6.1 Janelas Estatísticas

Em princípio a estatística SSQ_t deveria aumentar quando há padrões de mistura e sistemáticos presentes, e diminuir quando há diminuição da variabilidade. Na prática verifica-se uma grande sensibilidade de SSQ_t com a variação da média. Desvios da média ocasionarão um aumento no valor de SSQ_t . Por outro lado, um padrão sistemático (ou mistura) com intensidade grande, pode levar a um aumento exagerado do valor da estatística de CUSUM. Desta maneira, as duas estatísticas não ajudariam muito na discriminação entre os padrões de mistura, sistemático e desvio da média. Para diminuir a sensibilidade da estatística SSQ_t a desvios da média, propõe-se a utilização da seguinte forma de cálculo :

$$SSQ_t = \sum_{i=0}^5 \sum_{t=t_2 - ((i.n) + n)}^{t_2 - (i.n)} (y_t - media_i)^2 \quad Eq. 6.3$$

onde, $n = 4$, que é o tamanho das janelas estatísticas;

$media_i$ é a média dos pontos na janela estatística i ; e

y_{12} é o valor da amostra mais recente.

A estatística SSQ_t continuará sendo calculada pelas últimas 25 amostras, mantendo então a distribuição quiquadrado com 24 graus de liberdade. Esta fórmula de cálculo deixa a estatística SSQ_t mais robusta a desvios da média, sem tirar a sensibilidade a aumentos ou diminuições da variabilidade.

Assim, inicialmente o vetor de entrada das redes será composto pelas estatísticas de CUSUM ($S_i(+)$, $S_i(-)$), SSQ_t , os "lags" de autocorrelação $\{\text{autocorr}_1, \text{autocorr}_2, \dots, \text{autocorr}_{10}\}$, as médias e desvios padrões das janelas estatísticas $\{T, T-4\}$, $\{T-5, T-9\}$, $\{T-10, T-14\}$, $\{T-15, T-19\}$, ..., $\{T-55, T-59\}$, e os últimos 10 pontos individuais $\{y_t, y_{t-1}, y_{t-2}, \dots, y_{t-9}\}$ (Fig 6.2).

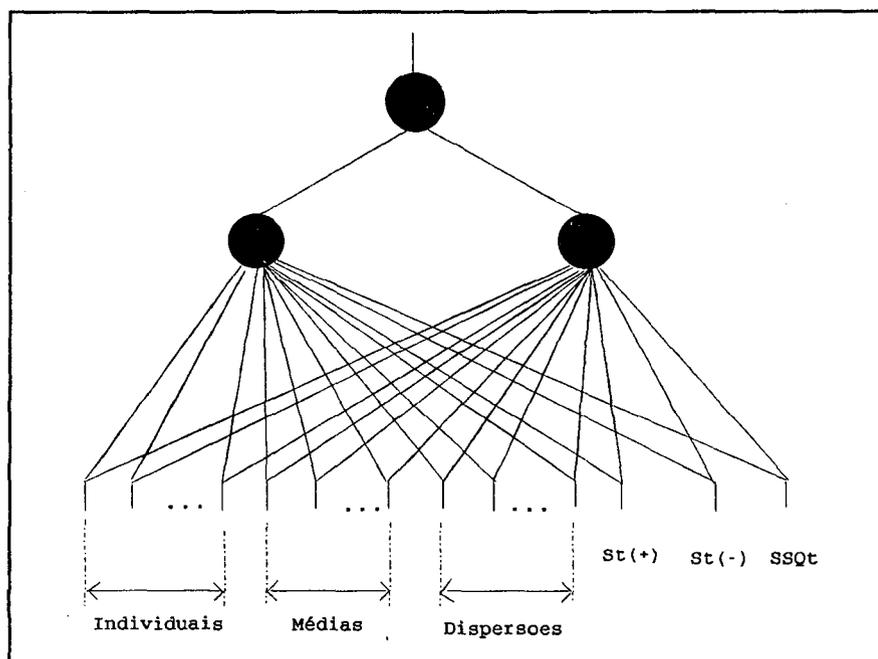


Fig. 6.2 - Vetor de entrada nas redes

6.2 - PADRÕES DE TREINAMENTO E TESTE

Hwang & Hubele [56] sugerem algumas diretrizes para selecionar os padrões para treinamento :

- **Evitar padrões ambíguos.** A definição dos padrões de treinamento deve ser efetuada de maneira a evitar ambiguidade entre as classes. Devido a variabilidade do processo, muitas vezes um padrão de uma classe pode se parecer com padrões de outra classe. Padrões conflitantes ou muito parecidos, pertencentes a classes diferentes, não devem ser incluídos no treinamento. Como regra, os padrões usados no treinamento devem ser ao menos reconhecíveis pelo usuário.
- **Os padrões de treinamento devem representar bem as classes a que pertencem.** O número de padrões deve ser suficiente para cobrir o espectro de possibilidades de variação dentro de cada classe.
- **Ter uma representação balanceada.** Alguns tipos de padrões são mais difíceis de serem reconhecidos que outros. Por exemplo, um padrão de mistura é mais difícil de ser aprendido pela rede que um desvio na média. A incorporação no treinamento de um número maior de padrões das classes mais difíceis de serem aprendidas, permite um desempenho mais equilibrado de classificação.

Como o objetivo deste trabalho é um estudo comparativo da capacidade de reconhecimento de três redes neurais, não há necessidade de se ter uma representação balanceada dos padrões de treinamento. Por outro lado, procurou-se evitar padrões ambíguos e ter um número suficiente de padrões no treinamento.

Os padrões de treinamento e teste foram obtidos por um "programa" gerador, com os padrões sendo gerados pelas fórmulas apresentadas no apêndice 1. O "programa" gerador de

padrões para treinamento funciona da seguinte maneira :

- a) Gera uma sequência de 60 pontos provenientes de uma distribuição normal $N[0,1]$. Estes pontos simulam os resultados de um processo sob controle.
- b) Os 60 pontos normais são analisados pelo programa utilizando os limites de controle de $\pm 3\sigma$ para os pontos individuais, pelo gráfico de CUSUM e comparando a estatística SSQ_t a dois valores limites (estabelecidos pela distribuição Qui-quadrada). Se um desses testes for violado, o programa gera os pontos novamente. Se nenhum dos testes for violado, o programa passa a etapa seguinte.
- c) O programa gera um ponto do padrão desejado (proveniente de um ciclo, mistura, sistemático, etc).
- d) A janela de monitoramento desloca-se também um ponto (Fig. 6.3), e o programa repete as análises executadas na etapa "b". Se um dos testes foi violado, o programa dá o alarme e pergunta ao usuário se o padrão é aceitável para treinamento. Se a resposta for sim, o vetor de entrada da rede é formado e o padrão é incluído no conjunto de treinamento e o programa retorna a etapa "a". Se a resposta for não, o programa reinicia da etapa "a". Se nenhum dos testes foi violado, o programa retorna a etapa "c". No caso dos padrões de ciclo, muitas vezes nenhuma das estatísticas dá o alarme. Neste caso, quando 60 pontos de ciclo forem gerados, o programa pergunta ao operador se o padrão é adequado. Este procedimento é repetido até que seja atingido o número de padrões desejados.

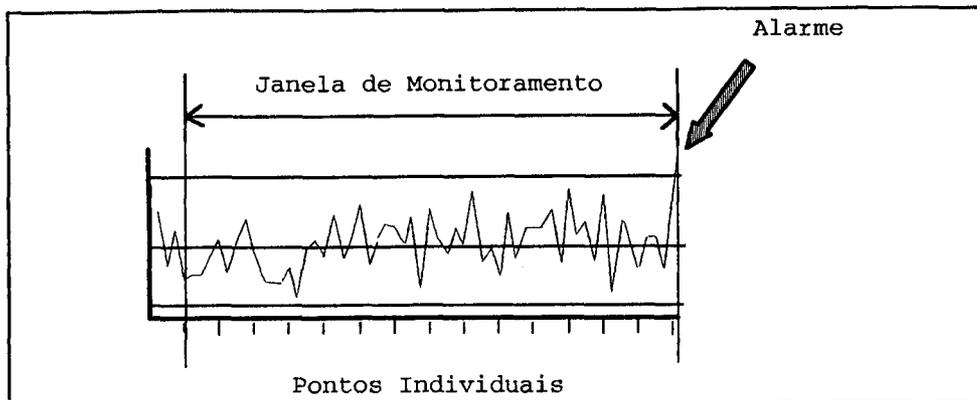


Fig. 6.3 - Janela de Monitoramento Deslizante

O conjunto de padrões de teste, utilizados no delineamento de experimentos para refinamento dos "parâmetros", é gerado de uma maneira um pouco diferente. Como se quer ter uma idéia do comportamento das redes quando em operação "on-line" e também ter padrões que sejam mais ou menos representativos das classes a que pertencem, a geração foi executada da seguinte maneira: o programa repete as etapas "a"- "c" acima, mudando a etapa "d" :

d) A janela de monitoramento desloca-se também um ponto (Fig. 6.3), e o programa repete as análises executadas na etapa "b". Se um dos testes foi violado, o programa verifica se o ARL (número de pontos desde o início do distúrbio) é maior que 4. Se o $ARL > 4$, o programa forma o vetor de entrada da rede e inclui o padrão no conjunto de teste. Se o $ARL \leq 4$ o programa reinicia da etapa "a". Se nenhum dos testes foi violado, o programa incrementa o ARL e retorna a etapa "c". No caso dos padrões de ciclo, muitas vezes nenhuma das estatísticas dá o alarme. Neste caso, quando 60 pontos de ciclo forem gerados, o programa para e inclui o padrão no conjunto de testes.

Assim, os padrões de teste são gerados automaticamente, sem interferência do usuário.

Foram gerados 1400 padrões de treinamento , 100 exemplares de cada amplitude em cada

tipo de padrão (total 200 de cada tipo):

- ciclos, com amplitude de 1 e $1,5\sigma$ (onde σ é o desvio padrão do processo) e período 20;
- padrões sistemáticos com amplitude de 1 e $1,5\sigma$;
- misturas com amplitude de 1 e $1,5\sigma$;
- desvios positivos e negativos da média, com amplitude de 1 e $1,5\sigma$;
- diminuição da variabilidade com amplitude de 0,5 e $0,7\sigma$; e
- padrões provenientes do processo sob controle (normais).

Os padrões normais foram incluídos para diminuir o erro de classificação das redes, dando a elas a oportunidade de classificarem um padrão como desconhecido ou fora dos limites de controle. Por exemplo, se um ponto cair fora dos limites de controle quando o processo está normal, e os padrões normais foram incluídos no treinamento, é mais provável que a rede classifique o padrão como normal, indicando apenas que o ponto saiu fora dos limites de controle. As redes terão um nó de saída para indicar que o processo está sob condição normal. Assim, as redes terão 7 nós de saída, um para cada classe de anomalia e um para o processo sob condição normal.

Os padrões de treinamento foram apresentados as redes em uma sequência aleatória (porém sempre a mesma para todos os treinamentos e redes) para diminuir os efeitos de uma eventual sensibilidade das redes à ordem de apresentação.

Para o conjunto de teste foram gerados 1800 padrões , 100 exemplares de cada amplitude em cada tipo de padrão (total de 300 de cada tipo) :

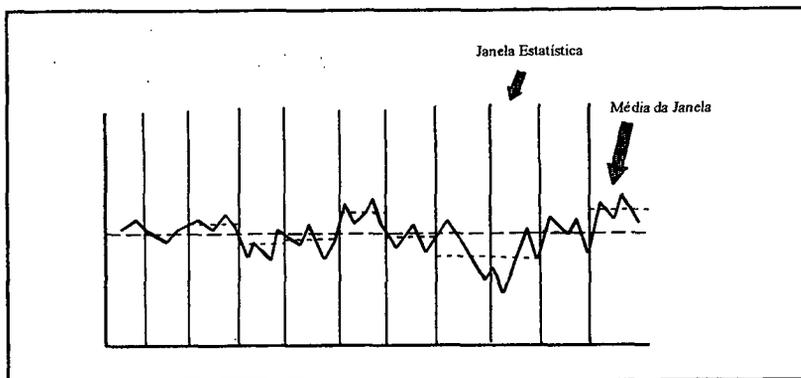
- ciclos, com amplitude de 1, 1,5 e 2σ (onde σ é o desvio padrão do processo) e período 20;

- padrões sistemáticos com amplitude de $1, 1,5$ e 2σ ;
- misturas com amplitude de $1, 1,5$ e 2σ ;
- desvio positivo da média, e desvio negativo da média, com amplitude de $1, 1,5$ e 2σ ;
- diminuição da variabilidade com amplitude de $0,5, 0,6$ e $0,7\sigma$.

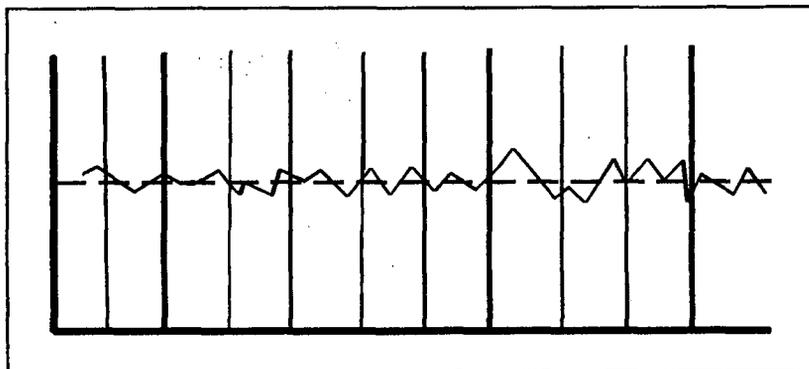
O nível de 2σ foi incluído para testar a capacidade de generalização das redes com valores extremos.

O menor desvio da média que se quer identificar neste trabalho é 1σ , o parâmetro do gráfico de CUSUM K será estabelecido igual a $0,5$ (Pag. 36). Como o que se deseja é a comparação de três redes neurais em uma mesma situação, não há necessidade de um valor exato para h , desde que ele permanece o mesmo para as três redes. Desta maneira utilizou-se para h o valor ($h = 6$) já utilizado em outros trabalhos [57][58]. Assim, os padrões de treinamento e teste foram gerados com os parâmetros do gráfico de CUSUM $K = 0,5$ e $h = 6$. Isto significa que o gráfico terá sensibilidade no mínimo para desvios da média $\geq 1\sigma$.

A estatística SSQ_t (com distribuição quiquadrado com 24 graus de liberdade) foi comparada com os limites $[6,0; 37]$. Esses limites indicam que há uma probabilidade de aproximadamente 5% dos padrões normais ultrapassarem o limite superior e quase zero de ultrapassarem o limite inferior. O limite inferior foi estabelecido bem pequeno porque, durante uma pré análise visual, verificou-se que a estatística SSQ_t , calculada pela equação 6.3, pode apresentar valores pequenos com mais frequência que a desejada. Isto ocorre porque a retirada da média na Eq. 6.3, ao mesmo tempo que dá maior robustez a desvios da média, ocasiona uma maior possibilidade de ocorrência de valores pequenos, pois em alguns casos flutuações de curta duração na média (normais em um processo sob condição normal) induzem a uma mudança da média das janelas estatísticas, e acabam sendo subtraídas (Figuras. 6.4 e 6.5).



6.4 - Flutuações da Média das Janelas em um Processo Normal



6.5 - Padrão Após Retirada das Médias - Diminuição da variabilidade

6.3 - DEFINIÇÃO DOS PARÂMETROS ÓTIMOS

Como cada rede tem uma dinâmica diferente para tratamento dos dados, elementos do vetor de entrada que podem ser importantes para um tipo de rede, podem não ser utilizadas pelas outras redes com a mesma importância. A influência de cada elemento do vetor de entrada no desempenho da rede depende de todos os elementos que compõem o vetor e dos parâmetros internos da rede. Assim, para conseguir o máximo de desempenho das redes aqui testadas,

experimentos foram projetados e executados com o objetivo de determinar o melhor conjunto de estatísticas para compor o vetor de entrada de cada rede e os valores ótimos dos parâmetros internos.

A largura das funções de base radial (δ) tem influência no número de nós criados pelas redes. Uma largura pequena, resulta em um grande número de nós, carregando o sistema. Uma largura muito grande, não permite a formação de um número suficiente de nós para que haja uma boa separação entre as classes de padrões. Assim, para as redes RBF e RBF-Fuzzy Artmap, a largura das funções de base radial é um parâmetro importante a ser pesquisado. Na rede Fuzzy-Artmap, o nível de vigilância da rede ART-a (ρ_a) regula o grau de generalização, a capacidade de aprendizado, e o número de nós da rede, sendo um parâmetro importante a ser pesquisado.

Além dos parâmetros internos da rede, os elementos que compõem o vetor de entrada devem ser otimizados. As estatísticas de CUSUM e SSQ_t não serão analisadas pelos experimentos, sendo incluídas automaticamente. O número de "lags" de autocorrelação, importantes para detecção de ciclos e padrões sistemáticos, deve ser no mínimo igual a metade do período do ciclo mais longo, permitindo uma boa representação. Como o período do ciclo que se quer identificar é 20, o número de "lags" de autocorrelação será 10. O número de pontos individuais, médias e desvios padrões no vetor de entrada não possui nenhuma regra para sua determinação. Se o número for pequeno, não haverá informação suficiente para ajudar na identificação. Se o número for muito grande, pode-se demorar mais ou atrapalhar a identificação devido a ter muitos pontos sem anomalia, sendo apenas ruído. Assim, além dos parâmetros internos das redes, foram utilizados como fatores nos experimentos: a quantidade de pontos individuais, a quantidade de médias e a quantidade de desvios padrões das janelas estatísticas. O planejamento experimental utilizou a matriz ortogonal L27 de Taguchi [27] para acomodar os fatores acima.

Delta e Delta Individuais são os parâmetros que definem a largura das funções de base radial. Delta Individuais dá a largura para as funções que processam os valores dos pontos individuais no vetor de entrada. Os níveis dos pontos individuais, médias e desvios padrão significam a quantidade dessas variáveis que estarão presentes no vetor de entrada. ρ_a é o nível de vigilância da rede ARTa na rede Fuzzy-Artmap.

Como os pontos individuais não sofrem nenhum tratamento estatístico, eles incorporam todo o ruído do processo. Um grande número de pontos individuais na entrada das redes pode induzir ruídos e dificultar a identificação pela rede. Assim, o nível máximo para o número de pontos individuais foi estabelecido arbitrariamente em 10. Os pontos individuais devem ajudar na identificação e discriminação de padrões do tipo sistemático e misturas. Assim, o nível mínimo foi estabelecido em 0 (zero), para se estudar o comportamento das redes sem a ajuda dos pontos individuais na identificação desses padrões.

O número de médias é um fator importante na identificação dos padrões, pois elas são menos sensíveis aos ruídos do processo, representando melhor as variações anormais. Assim escolheu-se o número máximo de médias móveis disponíveis (15) como o nível máximo nos experimentos. O nível mínimo foi estabelecido em 5 (cinco), pois tem-se a certeza da importância das médias na determinação dos padrões existentes, principalmente desvios da média e ciclos.

O número de desvios padrões deve ajudar na identificação de padrões como estratificação (diminuição da variabilidade), mistura e sistemáticos. Como essa ajuda não está bem clara, resolveu-se escolher os níveis cobrindo todo o espectro possível, com os níveis máximo igual a 15 (quinze), médio igual a 5 (cinco) e mínimo igual a 0 (zero).

Os níveis de Delta e Delta Individuais nas redes RBF e RBF-Fuzzy Artmap foram estabelecidos após uma análise prévia, como uma solução de compromisso entre um grande número de nós e a capacidade de generalização. O valor ótimo para esses fatores é fortemente

dependente do número e valor das entradas da rede, ainda não definidos. Assim, realizou-se vários experimentos, variando-se progressivamente os valores de Delta e Delta Individuais de valores pequenos até valores grandes, com uma combinação do número de entradas mínima (19) e outra máxima (43). Os valores ótimos ficaram entre 0,5 e 0,6 para a rede RBF, e entre 0,8 e 0,9 para a rede RBF-Fuzzy Artmap. Assim os níveis foram estabelecidos em 0,5, 0,55 e 0,6 para a rede RBF, e em 0,8, 0,85 e 0,9 para a rede RBF-Fuzzy Artmap. A mesma análise foi feita para a rede Fuzzy-Artmap, estabelecendo-se os níveis (0,6, 0,7 e 0,8).

A tabela 6.1 mostra os valores das variáveis utilizadas para cada rede:

Redes \ Níveis	Pontos Individuais			Médias			Desvios Padrão (σ)			Delta (δ)			Delta Individuais			ρ_a		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
RBF	10	5	0	15	10	5	15	5	0	0,5	0,55	0,6	0,5	0,55	0,6	-----		
Fuzzy - Artmap	10	5	0	15	10	5	15	5	0	-----			-----			0,7	0,6	0,8
RBF Fuzzy Artmap	10	5	0	15	10	5	15	5	0	0,8	0,85	0,9	0,8	0,85	0,9	-----		

Tabela 6.1 - Níveis das Variáveis Utilizadas nos Experimentos

As redes foram treinadas pelo conjunto de padrões de treinamento, e o desempenho medido pelo percentual de identificação de padrões corretos, para cada uma das classes, no conjunto de padrões de teste. Assim, utilizou-se como indicador de desempenho a média da performance de classificação entre todas as classes. As tabelas 6.2, 6.3 e 6.4 mostram os experimentos e os resultados obtidos para cada rede.

Individ.	Média	Disper.	Deta I	Delta		Sistem.	Média+	Média-	Var. -	Ciclo	Mistura	X-barra
10	15	15	0,5	0,5		81	87	91	99,67	58,33	76,33	82,22
10	10	5	0,5	0,55		74,33	85,33	88,33	99,67	58,33	85	81,33
10	5	0	0,5	0,6		73,33	83,67	85	99,67	62,33	88,33	82,06
10	10	5	0,55	0,5		72,67	80,33	88,33	99,67	56	84,33	80,22
10	5	0	0,55	0,55		75,33	87	87,33	99,67	64,67	88	83,67
10	15	15	0,55	0,6		74,67	81,33	89,67	99	57	84,33	81
10	5	0	0,6	0,5		74,33	84,67	85,33	99,67	59,67	87,33	81,83
10	15	15	0,6	0,55		73,67	80,33	89	100	57	84	80,67
10	10	5	0,6	0,6		77,33	88,33	87	100	62,33	78	82,17
5	10	0	0,5	0,5		80,67	88,67	90	99,67	62	83,67	84,11
5	5	15	0,5	0,55		78,67	82,33	87,33	99,67	63	79,33	81,72
5	15	5	0,5	0,6		80	89,33	87,67	99,67	59,67	77,33	82,28
5	5	15	0,55	0,5		79	82	87,67	99,67	60	80,67	81,5
5	15	5	0,55	0,55		77	88	90,67	99,67	59	79,33	82,28
5	10	0	0,55	0,6		85,67	86	86	100	67	75	83,28
5	15	5	0,6	0,5		77,33	81,67	89,33	99,67	56	82	81
5	10	0	0,6	0,55		83	90,67	88,33	100	64,67	50,67	79,56
5	5	15	0,6	0,6		83	81	83	100	72,33	68	81,22
0	5	5	0,5	0,5		93,33	90	85,67	100	64,67	0	72,28
0	15	0	0,5	0,55		90,67	95,33	93,33	10	64,67	0	59
0	10	15	0,5	0,6		94,33	92,67	88,67	100	66,67	0	73,72
0	15	0	0,55	0,5		90	94	90,33	99,67	60	0	72,33
0	10	15	0,55	0,55		81	92,33	88	100	66	23	75,06
0	5	5	0,55	0,6		97,33	93,67	0	99,67	0	0	48,44
0	10	15	0,6	0,5		71	91,33	94,67	100	59,67	80,67	82,89
0	5	5	0,6	0,55		95	92	77	100	56,33	0	70,06
0	15	0	0,6	0,6		92	91,67	94	12,67	67	0	59,56

Tabela 6.2 - Resultado dos Experimentos da Rede RBF

- Onde, Ind é o número de pontos individuais.
Média é o número de médias.
Disp é o número de desvios padrões.
 ρ_a é o nível de vigilância da rede ART-a na rede Fuzzy-Artmap.
Delta é a largura das funções de base radial para os pontos individuais.
Delta é a largura das funções de base radial para outras entradas que não pontos individuais.
Sist. é o desempenho da rede para identificação dos padrões sistemáticos.
Media + é o desempenho da rede para identificação de desvios positivos da média.
Media - é o desempenho da rede para identificação de desvios negativos da média.
Var - é o desempenho da rede para identificação de diminuição da variabilidade.
Ciclo é o desempenho da rede para identificação dos padrões de ciclo.
Mistura é o desempenho da rede para identificação dos padrões de mistura.
x-barra é a média dos desempenhos individuais.

Individ.	Média	Disper.	ρ_a		Sistem.	Média+	Média-	Var.-	Ciclo	Mistura	X-barra
10	15	15	0,7		65,67	43	77	46	60	10	50,28
10	10	5	0,6		80,67	93,33	95,33	99	15,33	46,67	71,72
10	5	0	0,8		78,67	90,67	60,33	95,33	37,67	33,33	66
10	10	5	0,7		10,67	72,67	84	98,67	58,67	85,67	68,39
10	5	0	0,6		63,67	68,67	86,67	96,33	48,33	63,33	71,16
10	15	15	0,8		67,67	99	96,33	19,33	27,33	14,33	54
10	5	0	0,7		58,67	65	91,67	67,67	15,67	18,67	52,89
10	15	15	0,6		94,67	19,67	53,67	52	57,67	24,67	50,39
10	10	5	0,8		56	57	83	98,67	49	39,67	63,89
5	10	0	0,7		56,67	91,67	95,33	94	39,33	33	68,33
5	5	15	0,6		41	37,33	89	99,33	49,67	60,33	62,77
5	15	5	0,8		37,33	55,33	97	99	52	79,33	70
5	5	15	0,7		83,33	62,33	71,33	92	29,67	27,0	60,94
5	15	5	0,6		49,33	47,33	94,67	59,33	31,33	64,33	57,72
5	10	0	0,8		68	81,67	89,67	96,67	48,33	62,67	74,5
5	15	5	0,7		65,33	92,67	68,67	81	39,67	14	60,22
5	10	0	0,6		68,67	86	97,67	97,33	33,67	14	66,22
5	5	15	0,8		81,33	85	96	99	29,33	43	72,28
0	5	5	0,7		63,33	42	87,33	94	47,67	56,67	65,16
0	15	0	0,6		34,33	60	8,33	16,33	40	59	36,33
0	10	15	0,8		17	91,67	81	100	27,33	79,67	66,11
0	15	0	0,7		59,67	54	87,67	61,33	32	55,33	58,33
0	10	15	0,6		21,67	71	86,67	92,33	8	14,67	49,05
0	5	5	0,8		41,33	69,33	64,33	99,33	42	67,67	64
0	10	15	0,7		30,33	79	84,33	77,67	41,67	85,67	66,45
0	5	5	0,6		62,67	59,67	88,67	83	55	58	67,83
0	15	0	0,8		64,67	65,33	84	95,33	32	71,33	68,77

Tabela 6.3 - Resultado dos Experimentos da Rede Fuzzy-Artmap

Apesar de as interações entre os fatores utilizados nos experimentos ter sido levado em consideração quando da escolha da matriz ortogonal, a correlação entre esses fatores comprometeu seriamente qualquer análise efetuada pelo Método Taguchi [27]. Além disso, o Método Taguchi apenas escolhe a melhor combinação entre os níveis utilizados, não fazendo uma procura por um ponto ótimo intermediário. Assim, utilizou-se o resultado dos experimentos para ajustar funções de regressão multivariada [1] que representem os modelos estudados e permitam analisar a influência dos diversos fatores no desempenho das redes.

Individ.	Média	Disper.	Delta I	Delta		Sistem.	Média+	Média-	Var.-	Ciclo	Mistura	X-barra
10	15	15	0,8	0,8		83	86,67	91,67	100	61	67,33	81,61
10	10	5	0,8	0,85		76,33	86,33	85,33	98,67	61,3	78,67	81,11
10	5	0	0,8	0,9		74,67	85	85,33	91,33	58,6	72	77,83
10	10	5	0,85	0,8		77,33	86	86,67	99,33	60,6	79	81,5
10	5	0	0,85	0,85		74,33	84,67	87	95,33	58	75,67	79,17
10	15	15	0,85	0,9		79	86,67	91,33	99,67	60	75,33	82
10	5	0	0,9	0,8		78	84	87,67	93,67	58,3	75	79,45
10	15	15	0,9	0,85		80,67	87,33	92	99,67	58,6	75	82,22
10	10	5	0,9	0,9		76	86,33	85,67	98	60,6	79,33	81
5	10	0	0,8	0,8		72	84,33	86,33	89,67	56,6	81	78,33
5	5	15	0,8	0,85		73,33	82,67	83,67	97	64	82,67	80,56
5	15	5	0,8	0,9		77,67	85,33	87,67	98,67	58,3	79,33	81,17
5	5	15	0,85	0,8		72,33	86	85,33	97	61,3	84	81
5	15	5	0,85	0,85		75,67	86	88	98,33	55,6	83	81,11
5	10	0	0,85	0,9		72	83,67	85,33	91,67	57,3	81,33	78,56
5	15	5	0,9	0,8		74	86,33	89	98,67	57	79,67	80,78
5	10	0	0,9	0,85		74,33	84,33	84,33	91	57,6	81,67	78,89
5	5	15	0,9	0,9		74	81,67	84,33	97	59	83	79,83
0	5	5	0,8	0,8		64	60,67	84	96	55,3	73	72,17
0	15	0	0,8	0,85		68	65,67	90	89,33	50	68,33	71,89
0	10	15	0,8	0,9		61	71,67	89,67	96,67	50,6	71,67	73,56
0	15	0	0,85	0,8		68	65,67	90	89,33	50	68,33	71,89
0	10	15	0,85	0,85		61	71,67	89,67	96,67	50,6	71,67	73,56
0	5	5	0,85	0,9		64	60,67	84	96	55,3	73	72,17
0	10	15	0,9	0,8		62,33	71,67	89,67	96,67	50,6	71	73,67
0	5	5	0,9	0,85		64	60,67	84	96	55,3	73	72,17
0	15	0	0,9	0,9		65	64,67	90	88,33	54,3	69	71,89

Tabela 6.4 - Resultado dos Experimentos da Rede RBF-Fuzzy Artmap

O programa Minitab [75] foi utilizado para o ajuste diversas equações para cada rede. As equações com o coeficiente de correlação (R^2) mais alto foram escolhidas. Essas equações foram utilizadas para determinar os níveis ótimos dos fatores para cada uma das redes, utilizando-se o programa de otimização GAMS [76]. Os resultados previstos pelos modelos foram comparados com as saídas das redes. Caso o modelo tivesse uma previsão discrepante (O valor real fosse menor que qualquer dos resultados dos experimentos), um novo modelo foi estabelecido, repetindo-se o processo acima. As equações 6.4, 6.5 e 6.6 mostram as equações finais para as redes RBF, Fuzzy-Artmap e RBF-Fuzzy Artmap respectivamente.

$$\bar{X} = 113 + 4,13\text{Ind} - 0,277\text{Ind}^2 - 0,167\text{Média}^2 + 0,019,\text{Disp}^2 - 109\text{Delta} + 5,99\text{Média}.\text{Delta} \quad \text{Eq.6.4}$$

$$R^2 = 72.2\%$$

$$\bar{X} = 50,4 + 7,22\text{Ind} + 0,992\text{Disp} - 0,212\text{Ind}^2 - 0,223\text{Média}^2 - 0,0811\text{Disp}^2 + 3,13\text{Ind}.\rho_a - 5,18\text{Média} \rho_a \quad \text{Eq.6.5}$$

$$R^2 = 59.5\%$$

$$\begin{aligned} \bar{X} = & -521 + 2,54\text{Ind} - 1,97\text{Média} - 0,61\text{Disp} + 837\text{DeltaI} + 585\text{Delta} - 0,344\text{Ind}^2 - 0,0749\text{Média}^2 - 0,0502\text{Disp}^2 - \\ & 489\text{DeltaI}^2 - 369\text{Delta}^2 + 1,57\text{Ind DeltaI} + 3,74\text{Média Delta} + 1,69\text{Disp Delta} \end{aligned} \quad \text{Eq.6.6}$$

$$R^2 = 96..6\%$$

Os valores baixos para o coeficiente de correlação das redes RBF e Fuzzy-Artmap não indica que o modelo é inválido, mas que a rede segue uma dinâmica interna complexa que é sensível as diferenças de níveis dos experimentos. Por exemplo, a rede Fuzzy-Artmap pode seguir um modelo diferente quando $\rho_a = 0,6$ e quando $\rho_a = 0,6$. Os modelos acima representam a média do comportamento das redes no intervalo dos experimentos.

As equações acima foram utilizadas para determinar os níveis ótimos dos fatores para cada uma das redes. Utilizando-se o programa de otimização GAMS [76], com a variação dos fatores restrita aos intervalos utilizados no planejamento experimental, chegou-se aos seguintes níveis ótimos para cada uma das redes (Tabela 6.5):

	Individuais	Médias	Dispersões	Delta Ind.	Delta	ρ_a
Rede RBF	05	10	00	0,50	0,50	-----
Rede Fuzzy Artmap	04	06	09	-----	-----	0,6
Rede RBF-Fuzzy Artmap	08	15	11	0,872	0,865	0,3606

Tabela 6.5 - Níveis Ótimos para Entradas e Parâmetros Internos

O parâmetro ρ_a , na rede RBF-Fuzzy Artmap foi otimizado em separado, após os outros

parâmetros terem sido otimizados. Isto foi efetuado porque a inclusão de ρ_a , devido a interação que ele têm com os outros parâmetros, aumentaria muito o número de corridas experimentais necessárias, tornando inviável o tempo necessário para execução e análise.

A tabela 6.6 mostra o desempenho de classificação do conjunto de padrões de teste para as redes com os fatores otimizados (em % de classificações corretas):

	Sistemático	Desv. Posit. da Média	Desv. Negat. da Média	Diminuição. Variabilid.	Ciclo	Mistura	Média Geral	S
Rede RBF	80,67	88,67	90,00	99,67	62,00	83,67	84,11	12,63
Rede Fuzzy Artmap	77,00	88,67	93,00	98,33	51,67	70,67	79,89	17,19
Rede RBF-Fuzzy Artmap	78,00	86,33	88,33	99,33	60,67	82,67	82,56	14,55

Tabela 6.6 - Desempenho Comparativo das 3 Redes Neurais com Parâmetros Ótimos

Assim, os vetores de entrada das redes ficam (Tabela 6.7) :

	Pontos Individuais	Médias	Desvio Padrões	Lagas de Autocorrelação	Cusum +	Cusum -	Quiquadrado
Rede RBF	05	10	00	10	$S_i(+)$	$S_i(-)$	SSQt
Rede Fuzzy Artmap	04	06	09	10	$S_i(+)$	$S_i(-)$	SSQt
Rede RBF-Fuzzy Artmap	08	15	11	10	$S_i(+)$	$S_i(-)$	SSQt

Tabela 6.7 - Vetores de Entrada das Redes

Desta forma, a rede RBF terá 28 nós na camada de entrada, a rede Fuzzy-Artmap terá 64 nós, e a rede RBF-Fuzzy Artmap terá 41 nós.

Uma análise das tabelas 6.2, 6.3 e 6.4 mostra que a rede RBF-Fuzzy Artmap é a mais robusta das três em relação a variações nos níveis dos parâmetros internos da rede e nos componentes do vetor de entrada. Isto vêm confirmar a análise teórica, efetuada no capítulo 6, de

que variações nos níveis dos parâmetros na rede RBF-Fuzzy Artmap seriam compensados pelo número de nós.

A tabela 6.6 mostra que a rede com melhor desempenho é a rede RBF. Isto era esperado, pois ela possui duas camadas para fazer a separação entre os padrões, enquanto as redes Fuzzy-Artmap e RBF-Fuzzy Artmap possuem apenas uma, com as outras camadas fazendo apenas a conexão com as classes. A rede Fuzzy-Artmap apresentou o menor desempenho. Esses resultados indicam que a rede RBF-Fuzzy Artmap aqui proposta é uma alternativa viável à rede Fuzzy-Artmap, pois é mais robusta a erros nos parâmetros (o que é muito bom para o usuário interessado em desenvolver uma nova aplicação) e apresenta melhor poder de classificação, além de ter as características de aprendizado instantâneo e incremental.

6.4 - COMPARAÇÃO DA SENSIBILIDADE A ORDEM DE APRESENTAÇÃO DOS PADRÕES

Os testes para confirmação da sensibilidade a ordem de apresentação dos padrões de treinamento, descrito teoricamente no capítulo 5, foram efetuados utilizando-se o conjunto de padrões de treinamento e teste, usados na otimização dos parâmetros das redes. Nestes testes executou-se 20 treinamentos, com os padrões de treinamento apresentados de maneira aleatória. Após cada treinamento, o conjunto de padrões de teste foi utilizado para verificar o desempenho das redes RBF-Fuzzy Artmap e Fuzzy-Artmap. A análise dos resultados (tabelas 6.8 e 6.9) mostra uma grande diferença entre as duas redes, com a rede RBF-Fuzzy Artmap tendo a variabilidade do desempenho total ($S^2 = 0,285$) 98% menor que a rede Fuzzy-Artmap ($S^2 = 13,94$).

	Sistemático	Média +	Média -	Dim.Variab.	Ciclo	Mistura	X-Barra
	79,67	87,00	92,33	99,33	59,67	80,00	83,00
	79,67	88,67	90,33	98,67	57,67	78,00	82,17
	79,00	86,33	90,67	98,67	59,00	81,67	82,56
	77,00	88,67	89,67	98,67	56,33	81,00	81,89
	80,33	85,00	89,67	99,00	58,67	77,33	81,67
	80,33	85,00	90,67	99,00	58,67	78,67	82,06
	80,33	88,67	91,00	98,67	60,00	78,33	82,83
	79,67	87,00	90,00	99,00	60,00	80,33	82,67
	81,00	87,00	90,33	99,00	62,00	79,33	83,11
	78,33	87,00	89,00	98,67	58,33	78,00	81,56
	78,00	85,33	90,00	99,33	62,67	76,33	81,94
	81,33	88,00	90,33	99,00	61,00	80,67	83,39
	80,33	85,67	90,67	99,00	57,33	80,33	82,22
	81,00	87,33	90,67	99,33	60,00	78,33	82,78
	79,33	87,00	88,33	99,00	62,00	81,33	82,83
	83,00	86,33	88,67	97,67	57,67	80,00	82,22
	75,00	88,00	87,33	99,33	59,00	81,00	81,61
	79,33	90,00	90,00	99,33	59,00	78,67	82,72
	75,33	91,33	88,00	99,00	57,00	83,00	82,28
	77,67	88,33	91,00	99,67	59,33	81,67	82,94
X-Barra	79,28	87,38	89,93	98,97	59,27	79,70	82,42
S	1,96	1,63	1,18	0,42	1,71	1,71	0,534

Tabela 6.8- Sensibilidade a Ordem de Treinamento - Rede RBF-Fuzzy Artmap

	Sistemático	Média +	Média -	Dim.Variab.	Ciclo	Mistura	X-Barra
	77,00	88,67	93,00	98,33	51,67	70,67	79,89
	42,00	68,33	89,33	94,67	49,67	84,00	71,33
	70,33	92,67	97,33	98,67	26,33	71,67	76,17
	40,00	81,67	97,33	98,67	45,67	70,67	72,33
	77,33	77,67	89,00	98,33	39,00	51,67	72,17
	49,33	86,67	98,33	99,00	30,67	76,33	73,39
	41,00	90,00	94,33	98,00	46,67	90,33	76,72
	44,33	90,00	86,33	98,33	44,00	65,67	71,44
	65,00	97,67	96,00	100,0	30,67	54,33	73,94
	58,67	88,67	98,33	100,0	29,33	54,33	71,56
	80,00	93,33	95,33	99,67	45,33	62,33	79,33
	79,33	91,00	96,33	98,33	45,33	27,00	72,89
	69,33	87,00	93,33	97,33	55,67	81,67	80,72
	80,00	84,67	97,33	99,00	28,33	54,00	73,89
	38,33	85,33	98,00	100,0	37,33	32,00	65,17
	58,00	65,33	96,67	99,33	66,33	52,67	73,06
	50,33	86,67	94,33	99,33	37,67	81,00	74,89
	70,33	93,33	97,00	100,0	50,67	21,67	72,22
	63,00	86,33	91,33	99,00	50,33	59,67	74,94
	73,33	90,67	97,67	99,33	23,67	23,67	68,06
X-Barra	61,35	86,30	94,83	98,75	41,72	59,27	73,70
S	14,96	8,00	3,46	1,21	11,17	20,45	3,73

Tabela 6.9- Sensibilidade a Ordem de Treinamento - Rede Fuzzy Artmap

Os resultados acima são apenas indicativos de como as redes se comportariam em uma aplicação real, uma vez que o conjunto de padrões de teste foi gerado com $ARL \geq 4$, além de todos terem tido um alarme estatístico quando de sua geração (exceto ciclos). Em uma aplicação real podem haver ocasiões em que um padrão ocasione um alarme estatístico antes de ter um $ARL \geq 4$. Além disso, é interessante que as redes deem um alarme por si mesmas, antes do alarme estatístico.

Para uma correta avaliação do desempenho das redes em uma aplicação real, é necessário que se realize uma simulação com os padrões sendo gerados e avaliados ponto-a-ponto. Assim pode-se avaliar não apenas o desempenho das redes em termos de classificação, mas também em termos do tempo necessário para se ter o alarme (ARL). Esta análise será apresentada no próximo capítulo.

7 - EXPERIMENTOS E RESULTADOS DA COMPARAÇÃO

7.1 - O SISTEMA DE SIMULAÇÃO E COMPARAÇÃO

Para fazer a comparação de desempenho entre as redes, simulando uma situação real, é necessário que os padrões com anomalia sejam gerados e analisados ponto a ponto. Tendo estas necessidades em vista, o sistema para simulação e comparação foi dividido em três grandes subrotinas de identificação: de geração de padrões, de análise estatística e de identificação (Fig. 7.1).

A subrotina de geração de padrões gera um ponto com anomalia e rotaciona a janela de monitoramento. A geração é efetuada através dos algoritmos apresentados no apêndice 1. Inicialmente, a subrotina gera um padrão normal com sessenta pontos (tamanho da janela de monitoramento). Os pontos dos padrões normais seguem uma distribuição normal com os limites de controle de $\pm 3\sigma$ iguais a 0,9 e 0,1. Se um ponto sai dos limites de controle, ele é igualado a 1 ou zero, conforme o limite que foi ultrapassado. Os padrões anormais são gerados da seguinte forma:

- 1) O padrão normal é remetido a subrotina de análise estatística para verificação.
- 2) Se ele não ocasionou nenhum alarme estatístico (alarme falso)¹, um ponto com anomalia é gerado para iniciar o padrão anormal.
- 3) A subrotina incorpora este novo ponto a janela de monitoramento, ao mesmo tempo em que elimina o ponto mais antigo, movendo a janela um intervalo no tempo.
- 4) O padrão de sessenta pontos é remetido a subrotina de análise estatística.

¹ Alarme Estatístico: Alarme gerado por um ponto ultrapassando os limites de controle, os limites do gráfico de CUSUM ou de SSQI.
Alarme Falso: Alarme dado pelos gráficos ou testes acima quando o padrão é normal.

A subrotina de análise estatística processa o padrão proveniente da subrotina de geração, calculando as estatísticas de CUSUM e SSQ_t e verificando se estas estatísticas, ou o último ponto do padrão, violam os respectivos limites de controle.

- 1) Se algum limite é violado, e o padrão é o primeiro padrão normal obtido pela subrotina de geração, um novo padrão normal é produzido;
- 2) senão, o vetor de entrada da rede é montado e um alarme remetido para a subrotina de identificação.
- 3) Se nenhum dos limites de controle foi violado, e o padrão não é o primeiro padrão normal, o vetor de entrada da rede é montado e remetido para a subrotina de identificação para uma avaliação adicional. Todos os componentes do vetor de entrada são normalizados para o intervalo $[0, 1]$.

O gráfico de CUSUM é estabelecido com os parâmetros $h = 6$ e $K = 0,75$, e o SSQ_t é comparado com os limites 5,5 e 37. Os padrões de treinamento foram gerados com os parâmetros de CUSUM $h = 6$ e $K = 0,5$ (Cap. 6 - pag 82), e limites de $SSQ_t = [6, 37]$ (Cap 6 - pag 82). Esta diferença nos parâmetros do gráfico de CUSUM e SSQ_t faz com que eles sejam menos sensíveis e portanto, levem mais tempo para dar um alarme estatístico quando há uma anomalia. Isto é efetuado para diminuir a probabilidade de ocorrência de alarmes falsos quando o padrão é normal, deixando para as redes, neste caso, a tarefa de identificar os padrões parecidos com aqueles do treinamento. Este procedimento também serve para avaliar a capacidade de generalização das redes.

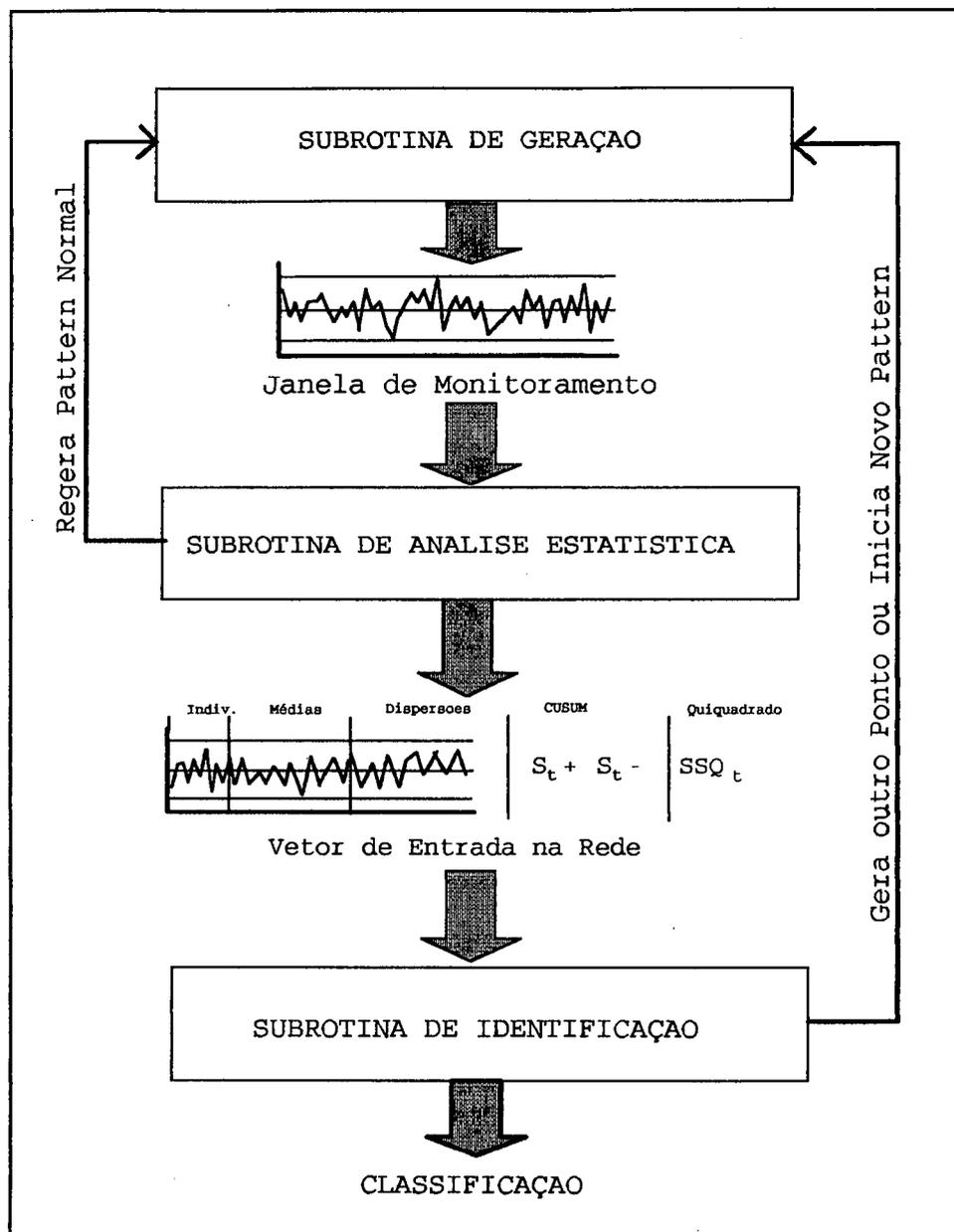


Fig. 7.1 - Sistema de Simulação e Comparação

A subrotina de identificação é composta basicamente pela rede neural (uma das três redes para cada simulação). Esta subrotina apresenta o vetor de entrada à rede neural, e analisa o vetor de saída da rede, onde as classificações ocorrem.

1) O vetor de saída é comparado com dois vetores de níveis de corte (dois níveis de corte para

cada classe de padrão) para verificar se alguma saída é significativa. Um dos vetores apresenta os níveis de corte para as saídas da rede quando não houve alarme estatístico, chamado de vetor de corte I. O outro vetor apresenta os níveis de corte para quando houve alarme estatístico, chamado de vetor de corte II.

- 2) Se um dos componentes do vetor de saída for maior que seu respectivo nível de corte, o padrão é identificado como pertencente a classe que aquele componente representa.
- 3) Se mais de um componente for maior que os respectivos níveis de corte, a classe identificada será aquela do componente com maior valor de saída.

O vetor de corte I controla a identificação dos padrões antes da ocorrência de um alarme estatístico. Se os níveis de corte forem fixados muito baixos, a rede dará muitos alarmes falsos. Se eles forem muito altos, a rede não identificará nenhum padrão anormal antes da ocorrência de um alarme estatístico, podendo levar a grandes ARLs para anomalias de pequena intensidade.

O vetor de corte II controla a identificação dos padrões após a ocorrência de um alarme estatístico. Quando há uma alarme com CUSUM ou SSQ_t , e nenhum dos componentes do vetor de saída ultrapassou seu respectivo valor de corte, a subrotina de identificação classificará o padrão como indeterminado; se um ponto saiu fora dos limites de controle, a subrotina dirá que há um ponto fora dos limites de controle. Este tipo de diagnóstico também acontece quando há um alarme estatístico e o componente vencedor do vetor de saída corresponde a classe de padrões normais. Este vetor de corte é importante porque é melhor que a rede dê uma classificação do tipo "indeterminado" ou "desconhecido", do que fazer uma identificação errônea quando um padrão é duvidoso. Caso o vetor de corte II seja estabelecido com valores muito altos, a rede dará muitos pontos como indefinidos ou fora dos limites de controle. Caso ele seja muito pequeno, o vetor não terá efeito nenhum sobre o desempenho da rede.

Assim, o vetor de corte I tem influência no ARL, e o vetor de corte II têm influência no número de padrão indeterminados e errôneos da rede, sendo diretamente conectado com o desempenho total.

Obs: Para a rede Fuzzy-Artmap e RBF-Fuzzy Artmap, a subrotina de identificação não verifica os níveis das saídas da rede na comparação com os níveis de corte. No caso da rede Fuzzy-Artmap, o que é verificado é o valor de "A" (Eq. 5.6, capítulo 5), que é utilizado para comparação com o nível de vigilância ρ_a . No caso da Rede RBF-Fuzzy Artmap, verifica-se o valor das saídas da rede RBF.

7.2 - DEFINIÇÃO DOS NÍVEIS DE CORTE I E II

7.2.1 - Norma para Estabelecimento dos Níveis de Corte

Como não existe uma norma na literatura para o estabelecimento dos níveis de corte para a saída das redes neurais, estabeleceu-se a seguinte norma para identificação e otimização:

Norma para Identificação Preliminar

- Inicialmente, os níveis de corte I e II serão estabelecidos através da análise da distribuição de frequências dos valores do vetor de saída após um alarme estatístico. Para esta análise, serão geradas 300 amostras para cada tipo de padrão, para cada rede.
- Os resultados de classificação das redes (para cada tipo de padrão) serão divididos em dois grupos : um grupo com os componentes vencedores do vetor de saída quando a rede fez classificações corretas, e outro grupo com os componentes vencedores do vetor de saída quando a rede fez classificações incorretas.
- O nível de corte II, para cada tipo de, padrão será estabelecido igual ao menor

valor ocorrido no grupo de identificações corretas. Desta maneira, qualquer valor estabelecido para os níveis do vetor de corte II na tentativa de diminuir o número de classificações incorretas após alarme estatístico, não levará também a diminuição do número de classificações corretas.

- O nível de corte I, para cada tipo de padrão, será estabelecido inicialmente próximo ao maior valor ocorrido no grupo de classificações incorretas. Este valor inicial é estabelecido como uma solução de compromisso entre um ARL grande e o desempenho da rede.

Norma para Otimização do Nível de Corte I

O objetivo aqui é reduzir a possibilidade de alarmes falsos dados pela rede quando o processo está sob condição normal.

A otimização dos valores de corte I é um problema complexo, pois necessita-se de um conjunto de valores que contribua para diminuir o ARL para detecção de um padrão anormal, ao mesmo tempo que mantenha um grande ARL até a ocorrência de alarmes falsos (identificar um padrão anormal quando só existem padrões normais). Valores pequenos para os níveis de corte I levarão não apenas a ocorrência de alarmes falsos, mas também ao aumento do número de classificações incorretas quando um padrão anormal estiver presente. Isto ocorre porque a rede identifica o padrão antes de ter informações suficientes para uma classificação correta.

Assim, cada valor de corte foi otimizado individualmente, verificando-se os valores de saída que ocorrem durante a operação normal.

- Serão realizadas 20 corridas experimentais para esta otimização. Em cada corrida será gerado uma sequência de pontos normais até a ocorrência de um alarme estatístico.

- O valor da saída será registrado para análise cada vez que as redes sinalizaram uma classe diferente da normal, mesmo que não tenha havido um alarme.
- O nível de corte I será então estabelecido igual ao valor máximo do grupo registrado para análise, ou do grupo de classificações incorretas estabelecido na fase de identificação preliminar (o que for maior).
- Os tipos de padrões que não forem identificados quando da presença de padrões normais, serão analisados quanto ao seu ARL. Caso este ARL para a menor amplitude ultrapasse 40, o valor de corte I será reduzido até que o ARL seja no máximo 40.

7.2.2 - Identificação Preliminar

Rede RBF-Fuzzy Artmap. A figura 7.2 mostra os histogramas dos valores de saída para as classificações corretas e incorretas da rede RBF-Fuzzy Artmap, para todos os tipos de padrões. Uma avaliação preliminar dos histogramas mostra que os valores das saídas para as classificações corretas e incorretas estão superpostos. Qualquer valor estabelecido para os níveis do vetor de corte II, na tentativa de diminuir o número de classificações incorretas após alarme estatístico, levará também a diminuição do número de classificações corretas, diminuindo o desempenho da rede. Os níveis do vetor de corte II foram estabelecidos iguais a 0,3, que são os menores valores das saídas tanto para os padrões classificados corretamente como para os classificados incorretamente. O vetor de corte I teve seus níveis estabelecidos próximos aos valores máximos que ocorrem nas classificações incorretas. Assim, o vetor de corte I para a rede RBF-Fuzzy Artmap ficou :

Desvio Positivo da Média	Desvio Negativo da Média	Diminuição da Variabilidade	Ciclo	Padrão Sistemático	Mistura
0,600	0,600	0,750	0,550	0,600	0,690

Tabela 7.1 - Vetor de Corte I para Rede RBF-Fuzzy Artmap

Rede Fuzzy-Artmap. Para a rede Fuzzy-Artmap, exceto para a diminuição da variabilidade, verifica-se que as classificações corretas tem uma concentração com valores 0,6. Isto é devido ao nível de vigilância ρ_a , que sempre leva a escolha de um nó com $A \geq 0,6$ (Eq. 5.6). Neste caso o vetor de corte II terá seus níveis de corte estabelecidos iguais a 0.6. A figura 7.3 mostra os histogramas para a rede Fuzzy-Artmap. Da mesma maneira que a rede RBF-Fuzzy Artmap, a rede Fuzzy-Artmap teve os valores para o vetor de corte I estabelecidos próximos aos valores máximos que ocorrem nas classificações incorretas. Assim, o vetor de corte I para a rede Fuzzy Artmap ficou :

Desvio Positivo da Média	Desvio Negativo da Média	Diminuição da Variabilidade	Ciclo	Padrão Sistemático	Mistura
0,668	0,626	0,640	0,65	0,618	0,600

Tabela 7.2 - Vetor de Corte I para Rede Fuzzy Artmap

Rede RBF. A figura 7.4 mostra os histogramas para a rede RBF. Uma análise dos histogramas revela que, exceto para a redução da variabilidade e mistura, qualquer tentativa de estabelecer os valores de corte do vetor de corte II maior que o menor valor que ocorreram durante os testes, levará a um aumento do número de padrões indeterminados em prejuízo do número de classificações corretas. Assim, os valores de corte II são estabelecidos inicialmente em 0,614 para diminuição da variabilidade, 0,36 para mistura, e 0,30 para os outros tipos de padrões. Da mesma maneira que as redes anteriores, a rede RBF teve os valores do vetor de corte I estabelecidos próximos aos valores máximos que ocorrem nas classificações incorretas. Assim, o vetor de corte I para a rede RBF ficou :

Desvio Positivo da Média	Desvio Negativo da Média	Diminuição da Variabilidade	Ciclo	Padrão Sistemático	Mistura
0,9	0,9	0,88	0,70	0,80	0,86

Tabela 7.3 - Vetor de Corte I para Rede RBF

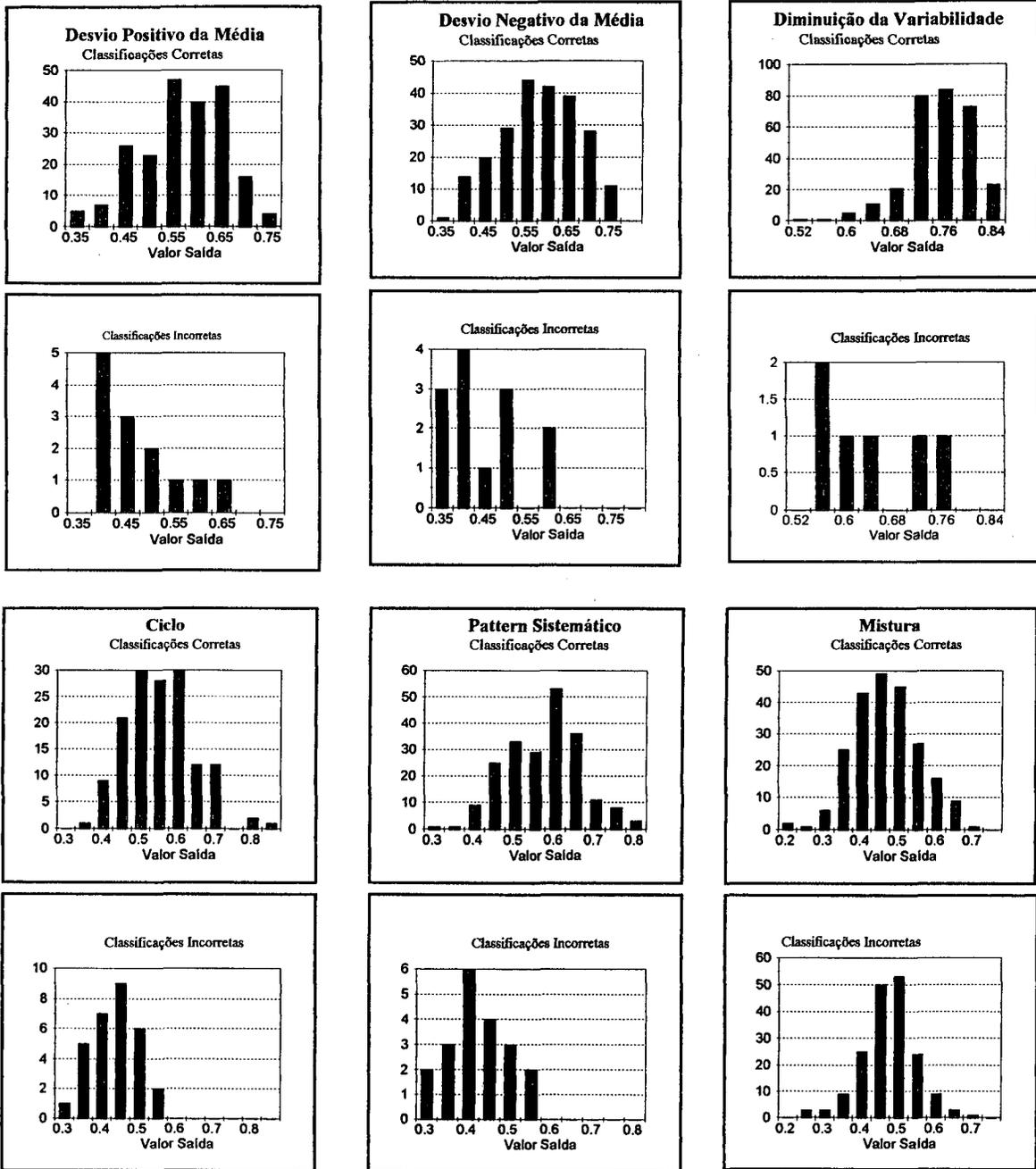


Fig. 7.2 - Distribuição de Frequências das Saídas da Rede RBF-Fuzzy Artmap

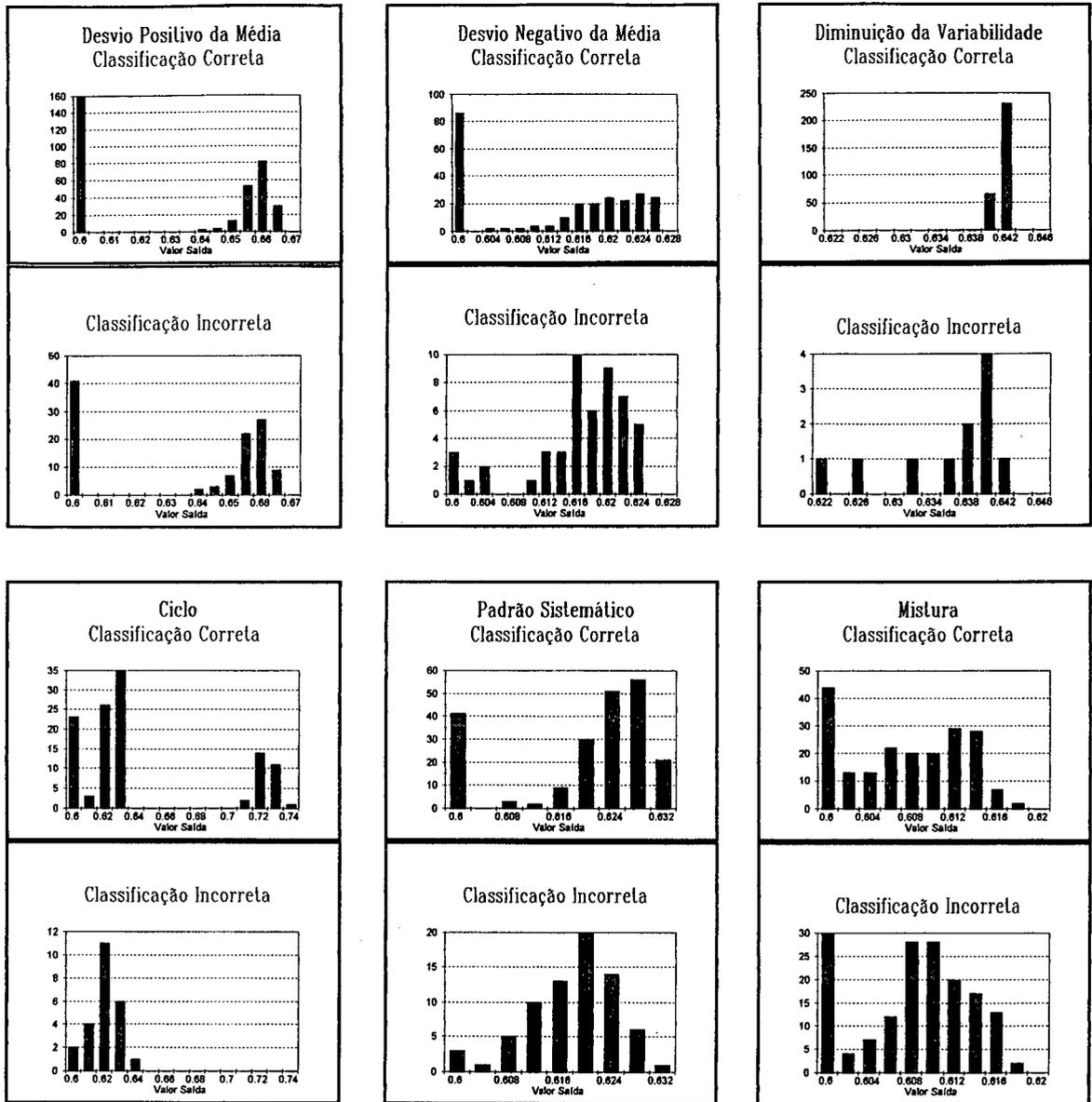


Fig. 7.3 - Distribuição de Frequências das Saídas da Rede Fuzzy-Artmap

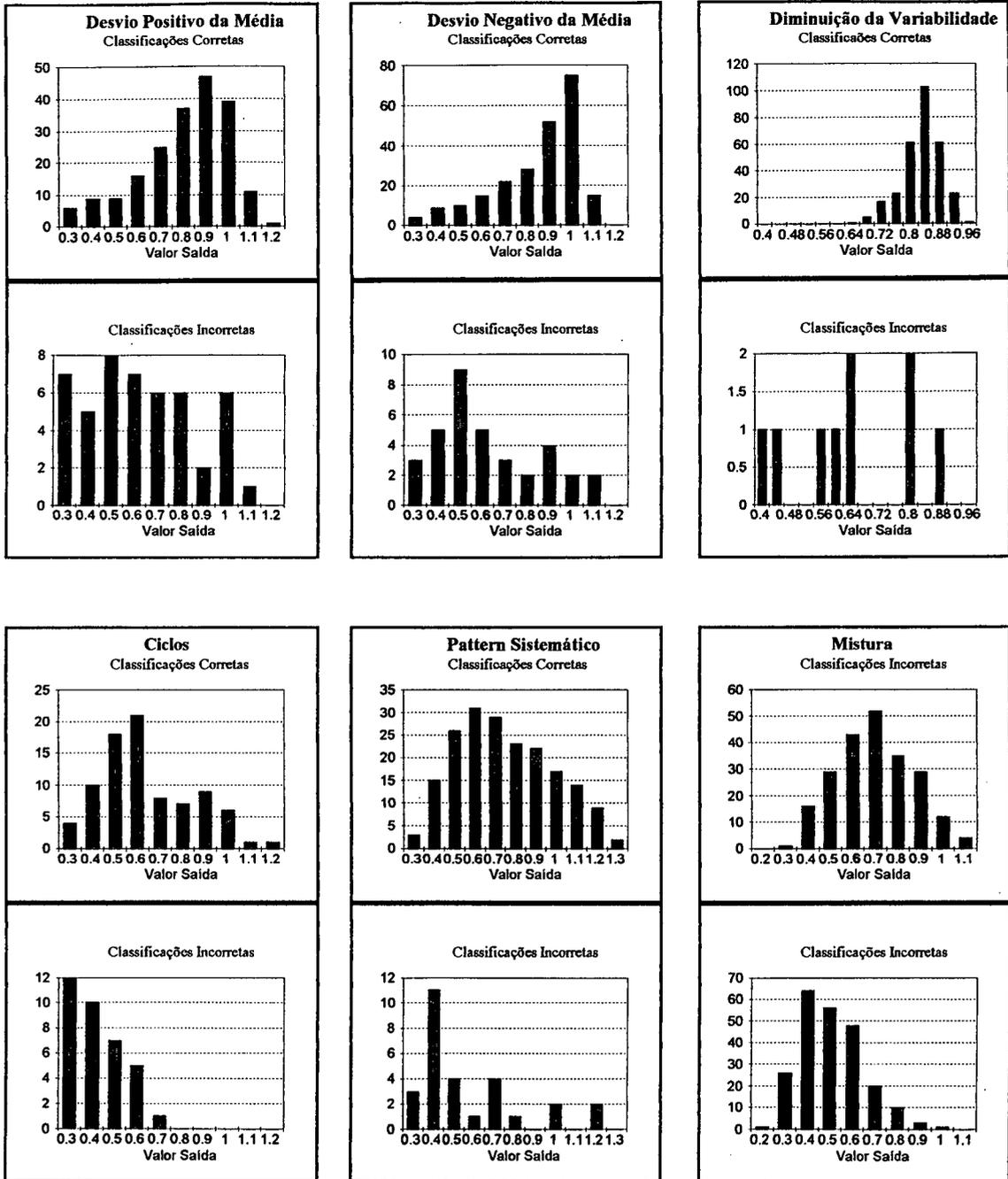


Fig. 7.4 - Distribuição de Frequências das Saídas da Rede RBF

7.2.3 - Otimização dos Valores de Corte

Rede RBF-Fuzzy Artmap. A figura 7.5 mostra os histogramas do desvio negativo da média, diminuição da variabilidade e mistura, para a rede RBF-Fuzzy Artmap durante a apresentação de padrões normais. A ausência dos outros tipos de padrão indica que a rede raramente os sinalizará quando o processo estiver sob condição normal.

O histograma para os desvios negativos da média indica que o valor previamente estabelecido (0,600) é adequado.

O corte repentino no histograma de diminuição da variabilidade, indica que a maioria dos alarmes estatísticos (sinalizados por SSQ_i) ocorreram em torno ou abaixo de 0,70. A grande quantidade de pontos um pouco menores que esse valor, sugere que deve-se estabelecer o nível de corte I em 0,7 para a diminuição da variabilidade.

O histograma para a mistura mostra que dificilmente ocorrerão alarmes falsos com padrões mistura se o valor de corte I estabelecido for 0,69 (valor inicial), podendo-se diminuí-lo para 0,65. Porém, o histograma para classificações incorretas de mistura (Fig. 7.2), sugere que deve-se continuar com o valor de corte 0,69.

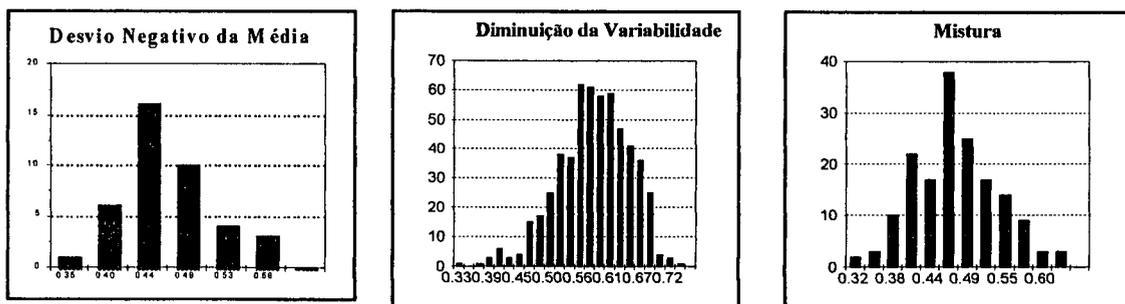


Fig. 7.5 - Distribuição de Frequências para Processo Normal - Rede RBF-Fuzzy Artmap.

O valor de corte I que foi estabelecido previamente implicou em ARLs muito grandes na identificação dos padrões de ciclo, principalmente para ciclos com amplitude pequena. Assim, o valor de corte I para os padrões de ciclo foi estabelecido procurando-se um ARL máximo de 40. O valor final para o nível de corte I foi fixado em 0,50.

Como o ARL para os desvios positivos da média e padrões sistemáticos são pequenos (< 40), e eles não aparecem quando só há padrões normais, os valores de corte para foram mantidos em 0,6.

Rede Fuzzy-Artmap. A figura 7.6 mostra os histogramas para a rede Fuzzy-Artmap. Novamente, os padrões que mais apareceram foram desvio negativo da média, diminuição da variabilidade e mistura. Uma análise desses histogramas mostra que os valores previamente estabelecidos são os menores possíveis sem aumentar o risco de alarmes falsos. Da mesma maneira que a rede RBF-Fuzzy Artmap, o valor do nível de corte I para os padrões de ciclo foi diminuído de 0,65 para 0,625. Os níveis de corte I para os outros tipos de padrão permaneceram iguais aos definidos anteriormente.

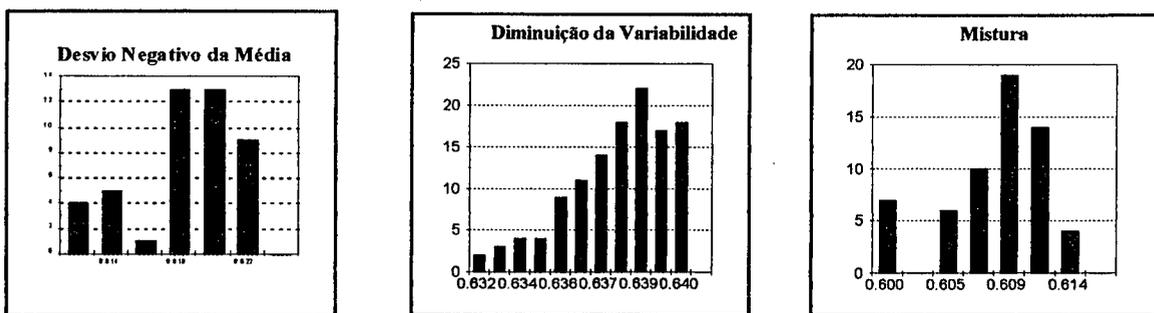


Fig. 7.6 - Distribuição de Frequências para Processo Normal - Rede Fuzzy Artmap

Rede RBF. A figura 7.7 mostra os histogramas para a rede RBF. Neste caso apenas a diminuição da variabilidade e mistura apresentam um risco considerável de provocar alarmes falsos.

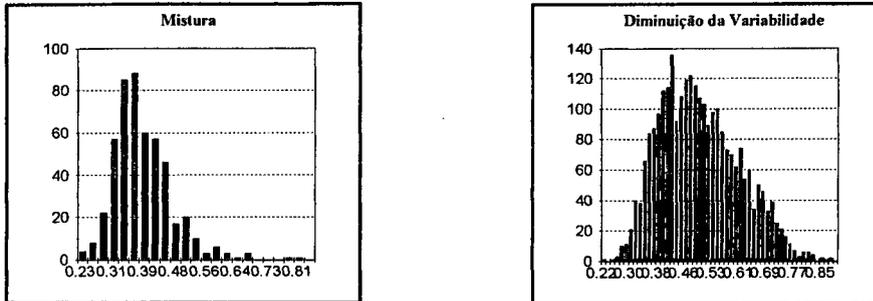


Fig. 7.7 - Distribuição de Frequências para Processo Normal - Rede RBF

O histograma para diminuição da variabilidade sugere que pode-se diminuir o valor de corte I para 0,776 sem aumentar muito o risco de alarmes falsos. O histograma da mistura mostra que o valor de corte I = 0,86, estabelecido anteriormente, é demasiado conservador em relação a proteção contra alarmes falsos. Porém, este valor não pode ser diminuído sob pena de aumentar muito o número de classificações incorretas de padrões (Fig.7.4). Os níveis de corte para os outros tipos de padrão permaneceram os mesmos.

A tabela 7.4 abaixo mostra os valores finais para os níveis de corte I e II para as redes RBF-Fuzzy Artmap, Fuzzy Artmap e RBF.

	PONTO DE CORTE I						PONTO DE CORTE II					
	Aumento Média	Dimin. Média	Dimin. Variabil.	Ciclo	Mistura	Sistem	Aumento Média	Dimin. Média	Dimin. Variabil.	Ciclo	Mistura	Sistem
RBF-Fuzzy Artmap	0,600	0,600	0,700	0,500	0,690	0,600	0	0	0	0	0	0
Fuzzy Artmap	0,668	0,626	0,640	0,625	0,618	0,631	0,600	0,600	0,600	0,600	0,600	0,600
Rbf	0,900	0,900	0,776	0,700	0,860	0,800	0,300	0,300	0,614	0,300	0,360	0,300

Tabela 7.4 - Valores de Corte Finais

7.3 - EXPERIMENTOS PARA COMPARAÇÃO DE DESEMPENHO

A comparação entre o desempenho das redes RBF, RBF-Fuzzy Artmap e Fuzzy-Artmap exige o estabelecimento de parâmetros de comparação normalizados e comuns aos usados normalmente para avaliação de gráficos de CEP.

Como as redes podem ter como resposta uma classificação indeterminada, ou fora dos limites de controle, elas não podem ser comparadas apenas pelo percentual de classificações corretas. A utilização apenas do percentual de classificações corretas, levaria a uma sub-estimação da capacidade das redes, uma vez que é desejável que elas façam a classificação como indeterminado quando não há conhecimento suficiente sobre o padrão em questão. Assim, a medida para comparação do desempenho em termos de poder de classificação será dada pelo desempenho corrigido :

$$\text{Desempenho Corrigido} = \frac{\text{Número de Classificações Corretas}}{\text{Total de Simulações} - \text{Número de Classificações Indeterminadas}} \quad \text{Eq. 7.1}$$

Além do desempenho corrigido, é importante comparar o tempo que as redes levam para dar um alarme quando há um padrão anormal presente e o tempo para dar um alarme falso quando o processo está sob controle.

O tempo para um alarme será dado pelo ARLDIX [45][47][56], calculado da seguinte maneira:

$$\text{ARLDIX}_c = \frac{\text{ARL}}{\text{Desempenho Corrigido}} \quad \text{Eq. 7.2}$$

onde o ARL é o tempo médio (número de amostras) para o alarme com classificação correta. O ARLDIX corrigido (ARLDIX_c) permite uma melhor comparação de desempenho que o ARL normal. Neste caso, uma rede que é mais rápida para dar alarme e que produz mais classificações

incorretas, pode ter um desempenho equivalente (em termos de ARLDIXc) a outra que é mais lenta mas produz mais classificações corretas.

As tabelas 7.5, 7.6, e 7.7 mostram o desempenho e o desempenho corrigido para as 20 corridas experimentais das três redes. As últimas duas linhas mostram o desempenho e o desempenho corrigido por experimento. Como os resultados de desempenho são o percentual de classificações corretas, eles tem uma distribuição binomial [77]. A distribuição binomial pode ser aproximada por uma normal quando $n \cdot \pi \geq 5$, onde n = número de amostras e π = proporção de classificações corretas [77]. Os resultados das simulações podem então ser aproximados por uma normal, uma vez que para cada padrão, em cada nível, tem-se uma amostra com $n = 1000$ e numa estimativa conservadora $\pi \geq 0,10$. Assim, $n \cdot \pi = 100 \gg 5$.

Desta forma, como as amostras são independentes, os resultados podem ser comparados pelo teste de ANOVA e "t" para a diferença entre as médias de duas populações [75][77].

Experimentos																					
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Media +	1	79,07	90,91	90,70	97,67	91,11	95,65	95,24	85,71	92,50	83,33	95,35	86,05	90,70	95,74	90,91	88,64	93,18	97,73	84,44	79,07
Media +	1,5	82,98	82,61	88,64	86,36	81,58	88,10	90,70	93,02	77,78	85,00	89,47	80,43	89,74	86,05	81,82	97,73	80,00	82,61	90,70	84,09
Media +	2,0	78,38	86,11	72,73	86,11	71,05	65,71	70,73	85,00	85,71	72,22	83,78	84,21	76,19	80,00	87,80	80,00	75,76	82,86	77,50	80,49
Media -	1	90,00	87,76	93,88	91,49	85,71	93,75	85,42	89,58	84,00	93,62	95,56	83,33	87,76	97,87	89,58	86,00	87,76	85,71	81,63	90,91
Media -	1,5	83,33	87,23	74,47	89,58	89,36	82,98	78,72	79,59	83,33	91,30	91,30	89,58	83,67	87,76	91,49	79,59	90,70	78,58	84,78	87,23
Media -	2,0	81,82	87,23	76,60	76,09	73,33	89,58	80,00	74,00	75,56	72,92	71,74	74,42	65,96	75,56	82,22	63,64	73,81	73,91	66,67	55,32
Variab -	0,5	100,0	100,0	100,0	100,0	100,0	100,0	98,00	100,0	100,0	98,00	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Variab -	0,6	100,0	97,96	100,0	100,0	100,0	100,0	100,0	98,00	100,0	100,0	100,0	100,0	98,00	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Variab -	0,7	100,0	98,00	97,96	97,96	98,00	100,0	100,0	100,0	100,0	100,0	98,00	98,00	98,00	100,0	98,00	100,0	100,0	100,0	98,00	100,0
Ciclo	1	71,74	65,12	76,09	68,89	75,00	84,78	79,59	76,19	70,45	73,91	73,17	77,78	73,33	78,26	77,78	78,57	70,45	69,77	75,00	80,43
Ciclo	1,5	63,04	53,49	38,64	64,10	35,56	47,83	50,00	50,00	45,00	62,22	37,21	54,05	58,14	51,16	57,78	43,90	60,87	61,36	63,04	36,96
Ciclo	2,0	22,22	22,50	26,83	14,29	17,95	38,48	23,68	25,71	39,47	26,83	17,07	26,67	29,27	22,50	28,26	41,03	14,29	14,63	27,27	30,23
Mistura	1	82,98	86,67	91,11	86,96	74,00	86,96	88,89	63,64	86,36	79,55	82,98	75,00	85,71	84,09	66,67	82,22	83,33	68,09	78,26	72,09
Mistura	1,5	74,47	77,78	77,08	80,00	80,43	82,93	76,74	79,55	76,60	76,19	68,89	72,92	72,34	80,00	84,78	73,33	60,47	87,80	67,44	81,40
Mistura	2,0	73,17	72,50	80,49	73,81	64,29	68,29	85,11	76,19	68,89	78,05	68,89	66,67	81,40	70,45	77,50	80,49	81,82	72,73	69,77	75,56
Sistemat	1	70,00	72,00	77,08	83,33	77,08	76,09	91,84	78,00	85,11	86,96	82,98	85,11	83,33	89,13	87,23	88,89	85,71	91,30	82,98	80,85
Sistemat	1,5	70,83	85,11	82,22	77,08	90,00	76,60	81,82	71,43	78,57	73,91	73,33	77,55	63,64	78,26	76,09	82,61	80,00	70,83	68,89	80,00
Sistemat	2,0	61,36	53,33	67,39	50,00	69,57	53,33	63,41	74,47	62,00	57,78	50,00	50,00	62,79	59,57	70,21	60,00	65,12	63,04	57,78	56,52
Perform		71,56	71,56	71,78	72,00	70,89	72,56	73,11	71,78	71,22	70,67	68,89	70,78	71,00	73,22	74,00	71,89	70,78	70,67	70,22	69,56
Perf Cor		77,40	78,92	79,17	79,70	77,33	80,42	81,03	78,68	79,14	79,19	77,21	77,40	78,40	80,66	80,73	80,07	78,74	78,42	76,89	76,53

Tabela 7.5 - Desempenho e Desempenho Corrigido para Rede RBF-Fuzzy Artmap

Experimentos																					
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Media +	1	80,00	93,75	83,67	95,74	87,23	89,39	85,42	89,36	85,42	89,58	85,71	85,71	83,67	89,36	88,64	73,91	79,59	89,13	76,09	85,42
Media +	1,5	80,00	88,89	88,89	78,72	88,89	82,61	80,43	85,42	88,89	86,96	89,58	83,33	78,26	89,58	80,43	85,11	78,72	85,42	88,37	84,78
Media +	2,0	70,83	80,00	85,71	76,60	74,47	68,09	70,21	76,09	77,08	76,09	64,58	73,91	76,74	66,67	61,70	70,83	84,78	83,33	58,33	88,64
Media -	1	90,00	81,63	79,59	84,00	89,80	83,67	83,67	94,00	94,00	90,00	92,00	92,00	86,00	87,76	89,58	80,00	86,00	86,00	87,76	80,00
Media -	1,5	82,00	82,00	80,00	87,23	83,33	81,63	87,76	89,36	88,00	87,76	80,00	80,00	83,67	90,00	76,00	93,88	74,00	74,00	80,00	93,88
Media -	2,0	77,55	79,17	79,59	75,51	86,00	82,00	73,47	82,98	81,63	79,59	71,43	72,00	73,47	73,47	76,00	75,00	75,51	81,63	71,43	74,00
Variab -	0,5	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	98,00	100,0	100,0	100,0	100,0	100,0	95,92
Variab -	0,6	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	98,00	98,00	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Variab -	0,7	100,0	100,0	100,0	100,0	98,00	98,00	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	98,00	100,0	100,0	98,00	100,0	100,0
Ciclo	1	70,21	58,33	54,00	72,92	66,67	66,67	65,31	70,21	64,44	61,70	61,22	59,18	73,91	61,70	61,22	62,50	55,10	68,00	63,04	54,35
Ciclo	1,5	57,78	50,00	51,02	51,06	63,83	44,90	51,11	48,94	53,06	44,90	55,10	58,00	63,27	51,02	62,50	43,75	48,98	56,25	51,06	48,94
Ciclo	2,0	42,00	25,00	12,50	37,21	17,78	38,30	26,67	20,83	21,74	27,08	19,57	31,25	20,83	36,73	21,74	21,74	14,58	20,00	28,26	28,57
Mistura	1	74,00	72,34	70,00	75,00	70,83	65,96	54,35	80,43	72,34	57,78	70,83	70,21	83,67	80,00	68,75	63,27	70,83	71,43	61,22	70,45
Mistura	1,5	56,25	74,47	69,39	55,32	55,10	63,83	53,06	72,34	56,00	68,75	63,27	61,70	57,14	62,50	63,27	79,59	71,43	63,04	57,45	66,00
Mistura	2,0	46,94	55,32	42,00	56,25	40,43	47,62	45,83	60,87	47,92	50,00	51,02	42,55	63,04	54,17	58,33	65,96	66,00	66,00	45,83	46,94
Sistemat	1	62,50	87,76	65,96	84,00	77,55	77,08	72,00	74,00	67,35	66,00	65,96	80,43	75,51	81,25	73,47	83,67	68,75	78,00	64,00	86,00
Sistemat	1,5	71,43	79,59	82,00	72,00	62,50	67,35	74,00	70,00	69,39	76,00	81,63	69,39	68,00	78,00	67,35	80,00	66,00	85,71	80,00	82,00
Sistemat	2,0	68,09	69,39	59,18	70,00	69,39	71,74	72,34	70,00	53,06	65,31	69,39	60,00	71,43	60,42	55,10	59,57	67,35	57,14	72,92	65,96
Perform		71,56	73,89	70,89	73,89	71,56	71,00	69,78	74,22	71,22	71,56	71,78	71,33	73,11	73,44	70,11	72,33	71,11	74,11	68,89	72,56
Perf Cor		74,02	76,70	72,42	76,61	74,36	74,22	72,43	77,14	73,59	74,02	73,58	73,37	75,63	75,72	72,53	74,74	72,73	76,05	71,68	75,23

Tabela 7.6 - Desempenho e Desempenho Corrigido para Rede Fuzzy Artmap

Experimentos																					
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Media +	1	84,78	88,89	84,40	87,80	87,23	85,71	89,58	85,11	78,72	91,30	85,11	87,23	93,33	83,72	88,37	80,00	81,25	84,44	86,67	80,00
Media +	1,5	86,96	84,78	86,96	66,67	82,22	88,89	81,25	89,13	79,59	85,37	89,36	84,78	83,33	84,09	91,30	85,42	86,36	88,89	80,43	82,61
Media +	2,0	78,57	74,42	64,44	68,29	88,37	81,40	76,74	77,27	80,95	83,33	82,22	72,73	76,74	75,00	63,83	84,09	83,33	75,61	76,60	84,09
Media -	1	85,11	90,00	89,36	91,67	89,80	77,78	91,67	85,42	75,51	86,00	89,13	81,25	89,13	81,63	89,58	92,00	83,33	86,96	91,84	86,96
Media -	1,5	81,25	82,98	78,26	91,49	82,22	89,58	80,43	75,00	68,00	89,80	85,71	77,08	96,00	85,71	85,11	87,76	79,59	86,96	81,63	81,25
Media -	2,0	74,47	89,58	74,47	62,50	64,58	72,34	66,67	77,55	70,21	73,33	57,14	61,22	88,89	64,44	67,35	78,26	76,60	68,75	78,26	77,08
Variab -	0,5	100,0	100,0	100,0	100,0	100,0	100,0	97,96	100,0	97,96	100,0	98,00	98,00	100,0	100,0	98,00	100,0	100,0	100,0	100,0	100,0
Variab -	0,6	100,0	98,00	98,00	100,0	100,0	100,0	100,0	100,0	100,0	100,0	97,96	98,00	100,0	100,0	100,0	97,92	100,0	100,0	100,0	100,0
Variab -	0,7	100,0	97,92	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	98,00	100,0	100,0	100,0	100,0
Ciclo	1	84,78	65,91	67,35	68,75	79,59	70,21	75,51	73,33	59,18	77,55	72,92	80,85	62,50	82,22	81,63	65,22	68,75	67,35	72,34	89,58
Ciclo	1,5	54,35	53,19	62,22	65,31	56,82	56,52	52,27	64,44	44,68	45,65	54,55	58,14	63,64	68,18	53,19	56,82	64,58	46,67	50,00	53,49
Ciclo	2,0	28,26	35,56	20,51	29,27	20,00	25,53	29,27	33,33	33,33	29,65	31,11	22,73	22,73	31,91	20,83	35,00	28,89	27,91	13,95	22,73
Mistura	1	79,59	87,23	88,00	87,23	85,42	83,33	81,63	89,58	86,00	88,00	87,50	74,00	81,63	83,67	87,76	89,80	82,61	95,83	77,08	88,00
Mistura	1,5	83,67	78,00	79,17	87,50	84,00	84,00	81,63	87,23	86,96	76,09	75,00	78,00	75,51	81,63	79,59	84,00	77,08	79,17	76,09	80,85
Mistura	2,0	71,43	68,75	68,75	65,22	85,42	68,09	69,57	75,00	63,83	67,35	70,45	72,00	75,51	77,55	68,09	68,00	83,33	86,96	65,91	71,11
Sistemat	1	78,00	79,59	75,00	77,08	72,92	83,33	79,17	66,67	71,43	87,50	70,21	70,83	76,00	68,75	64,00	74,47	80,85	70,00	82,98	73,47
Sistemat	1,5	65,31	72,92	80,85	69,57	66,00	70,00	74,47	83,33	75,51	74,00	74,00	76,09	73,91	67,35	73,47	68,09	81,25	81,63	56,25	79,17
Sistemat	2,0	63,27	50,00	68,75	75,51	54,17	62,50	61,22	70,21	70,00	61,22	65,96	68,09	63,83	55,10	61,36	53,06	66,67	45,83	60,42	53,19
Perform		73,89	73,89	73,22	72,67	74,33	74,22	73,33	75,44	71,89	74,78	73,22	72,44	75,11	73,22	73,11	74,11	75,11	73,22	71,33	74,00
Perf Cor		78,05	77,96	77,80	77,95	78,06	77,95	77,83	80,07	74,80	78,99	77,35	76,08	79,44	77,44	76,42	78,29	79,34	77,80	75,44	78,54

Tabela 7.7 - Desempenho e Desempenho Corrigido para Rede RBF

7.3.1 - Análise do Desempenho

Aplicando-se a análise da variância (ANOVA) para verificar a igualdade entre as médias do desempenho das três redes, verifica-se que elas são estatisticamente diferentes (com 95% de confiança) (Tabela 7.8).

Fonte	Soma dos Quadrados	Graus de Liberdade	Média Quadrada	Teste F
Entre Amostras	54,19	2	27,09	17,07
Dentro das Amostras	93,65	57	1,59	
Totais	147,84	59		

Tabela 7.8 - ANOVA para Desempenho das Três Redes

As tabelas 7.9 e 7.10 apresentam, respectivamente, os parâmetros e os valores das estatísticas de teste para o desempenho das redes. As redes Fuzzy-Artmap e RBF-Fuzzy Artmap

não têm desempenho estatisticamente diferente (com 95% de confiança). Já a rede RBF tem um desempenho claramente superior.

	N	Média	Desvio Padrão	Erro Padrão da Média
Fuzzy-Artmap	20	71,92	1,50	0,336
RBF-Fuzzy Artmap	20	71,41	1,23	0,275
RBF	20	73,63	1,08	0,241

Tabela 7.9 - Médias e Desvios Padrão dos Desempenhos para o Teste "t"

	T	P	GL	
RBF-Fuzzy Artmap X Fuzzy Artmap	1,17	0,2500	36	
RBF-Fuzzy Artmap X RBF	-6,08	0,0000	37	✓
Fuzzy-Artmap X RBF	-4,13	0,0002	34	✓

Tabela 7.10 - Teste "t" para Médias Iguais com 95% de Confiança (Desempenho)

Aplicando-se a análise da variância (ANOVA) para verificar a igualdade entre as médias do desempenho corrigido das três redes, verifica-se que elas são estatisticamente diferentes (com 95% de confiança) (Tabela 7.11).

Fonte	Soma dos Quadrados	Graus de Liberdade	Média Quadrada	Teste F
Entre Amostras	218,44	2	109,22	55,37
Dentro das Amostras	116,38	57	1,97	
Totais	334,82	59		

Tabela 7.11 - ANOVA para Desempenho Corrigido das Três Redes

As tabelas 7.12 e 7.13 mostram os parâmetros e as estatísticas de teste para as comparações do desempenho corrigido. Neste caso, pode-se afirmar que, com 95% de confiança,

a rede RBF-Fuzzy Artmap é melhor que as outras duas redes. Este resultado indica que a rede RBF-Fuzzy Artmap classifica mais padrões duvidosos como indefinidos, evitando classificações incorretas. Isto é uma grande vantagem, pois é melhor que o sistema classifique os padrões duvidosos como indefinidos do que dar uma classificação incorreta.

	N	Média	Desvio Padrão	Erro Padrão da Média
Fuzzy-Artmap	20	74,34	1,60	0,358
RBF-Fuzzy Artmap	20	78,80	1,36	0,305
RBF	20	77,78	1,31	0,292

Tabela 7.12 - Médias e Desvios Padrão dos Desempenhos Corrigidos para o Teste "t"

	T	P	GL	
RBF-Fuzzy Artmap X Fuzzy Artmap	-9,49	0,0000	37	✓
RBF-Fuzzy Artmap X RBF	2,1	0,0210	37	✓
Fuzzy-Artmap X RBF	-7,46	0,0000	36	✓

Tabela 7.13 - Teste "t" para Médias Iguais com 95% de Confiança (Desempenho Corrigido)

7.3.2 - Análise do ARLDIXc

As tabelas 7.14, 7.15 e 7.16 mostram os resultados do ARLDIXc para cada experimento, para todos os padrões e todos os níveis. A última linha e a última coluna das tabelas mostram as médias do ARLDIXc por experimento e por nível, respectivamente. O teste de hipótese para os ARLDIXc médios para cada rede mostra que apenas a rede RBF pode ser considerada diferente (melhor) que as outras redes. Este resultado indica que, embora a rede RBF-Fuzzy Artmap tenha um desempenho corrigido mais alto, ela leva mais tempo para detectar um padrão anormal. Uma análise mais detalhada das tabelas mostra que a rede RBF-Fuzzy Artmap tem ARLDIXc igual ou ligeiramente menor que a rede Fuzzy-Artmap em todos os tipos de padrões, exceto para os ciclos.

Neste caso ela tem um ARLDIXc maior para ciclos com amplitude 1 e $1,5\sigma$. Isto pode ser melhorado de duas maneiras: diminuindo o nível de corte I para ciclos, corendo-se o risco de aumentar o número de alarmes falsos, ou aumentando o número de padrões de ciclo no treinamento.

Uma análise geral dos resultados mostra que as redes apresentam o pior desempenho para ciclos, e o melhor desempenho para diminuição da variabilidade. O menor desempenho para os ciclos é decorrente da inexistência de uma estatística que os caracterizem para ajudar na identificação. Além disto, dependendo do tempo até a identificação, os ciclos podem ser confundidos com desvios da média. Isto ocorre com mais frequência quando a amplitude é grande e os ARLs são pequenos. O melhor desempenho para a diminuição da variabilidade decorre da existência de uma estatística única que a descreve (SSQ_t inferior), além de ter os pontos dos padrões mais comportados.

Experimentos																						
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	X-bar
Media +	1	21,6	16,2	18,2	20,7	23,9	18,9	21,6	21,7	18,9	21,6	18,6	19,0	20,5	17,5	17,7	20,6	19,5	20,8	27,2	23,0	20,4
Media +	1,5	11,0	11,7	9,9	9,4	10,7	9,0	11,3	9,7	12,6	11,0	11,8	11,2	11,3	9,1	10,0	8,9	10,9	10,5	9,7	11,4	10,5
Media +	2,0	7,5	7,0	8,7	7,3	7,3	10,3	8,4	7,0	6,8	8,0	6,5	6,9	8,1	6,7	6,6	7,4	7,9	7,7	7,6	7,2	7,6
Media -	1	20,7	23,2	16,8	20,9	21,8	22,8	22,1	21,9	24,6	19,5	19,8	20,6	20,9	20,1	19,7	22,8	22,4	21,6	22,4	26,5	21,5
Media -	1,5	10,1	8,3	9,9	10,1	9,0	9,4	10,1	10,7	10,9	10,0	10,1	9,5	9,4	9,9	9,2	9,6	9,3	11,5	9,7	10,6	9,9
Media -	2,0	7,1	6,8	7,6	7,9	6,9	5,8	5,9	8,0	7,4	8,1	6,6	8,3	8,4	8,0	6,0	9,0	7,5	7,0	7,6	9,9	7,5
Variab -	0,5	15,2	15,0	16,0	15,4	13,5	13,8	15,8	14,9	16,9	14,7	15,5	16,3	16,5	16,1	13,4	14,9	15,8	15,9	15,0	14,9	15,3
Variab -	0,6	16,0	19,4	20,5	20,8	21,7	16,5	19,4	20,1	18,0	18,4	19,6	18,8	22,9	16,1	17,6	16,6	18,7	16,9	17,3	21,0	18,8
Variab -	0,7	27,0	29,7	27,1	26,9	35,9	30,1	31,3	23,6	26,6	28,1	30,8	30,9	28,6	31,7	29,5	29,0	26,3	25,9	28,9	23,9	28,6
Ciclo	1	55,1	64,6	47,0	61,2	49,3	47,8	53,2	52,7	54,7	51,5	50,3	48,5	52,3	50,4	49,0	52,3	48,0	54,9	51,3	49,3	52,2
Ciclo	1,5	36,3	44,7	60,4	40,0	85,0	59,8	59,7	50,5	59,4	37,0	61,0	50,9	42,8	49,2	41,8	60,0	36,6	38,7	41,9	67,7	51,2
Ciclo	2,0	69,3	99,3	63,7	113	112	38,1	73,2	69,6	39,9	87,9	110	58,1	66,4	79,6	53,4	39,5	110	101	55,2	55,7	74,8
Mistura	1	38,1	38,2	33,9	36,6	34,4	33,0	35,3	64,2	39,3	33,3	40,2	44,3	34,6	31,7	40,5	40,6	34,0	61,2	43,0	39,1	39,8
Mistura	1,5	13,2	14,3	15,2	14,7	13,5	12,2	11,6	12,4	14,9	14,6	16,8	17,2	14,8	13,0	12,7	14,2	16,9	11,2	15,9	12,0	14,0
Mistura	2,0	7,7	8,5	8,4	9,1	9,9	7,9	7,5	7,0	9,2	6,9	7,7	8,8	6,8	7,4	6,9	7,8	7,2	8,7	8,6	8,5	8,0
Sistemat	1	25,5	30,0	31,5	23,1	25,1	25,9	20,9	24,8	21,5	18,6	22,3	23,4	24,5	22,6	19,4	25,1	20,9	22,0	24,5	24,1	23,8
Sistemat	1,5	12,9	10,8	10,4	11,0	9,4	11,5	10,8	13,3	10,6	13,2	14,0	12,3	15,0	12,5	11,9	10,9	11,3	11,8	15,8	12,0	12,1
Sistemat	2,0	9,6	11,1	10,6	12,3	8,7	12,5	9,2	7,9	9,6	10,5	12,6	13,3	10,0	10,7	8,0	11,2	9,8	8,8	9,4	10,9	10,3
Média Experim.	----	22,5	22,5	23,1	25,6	27,7	21,4	23,7	24,5	22,3	22,9	26,4	23,2	23,0	22,9	20,7	22,2	24,0	25,3	22,8	23,8	23,4

Tabela 7.14 - ARLDIXc para Rede RBF-Fuzzy Artmap

Experimentos																						
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	X-bar
Media +	1	26,1	25,9	24,6	16,8	22,3	20,4	23,3	23,7	22,3	23,6	27,7	23,3	26,7	22,9	16,5	27,5	22,5	23,1	29,0	22,5	23,3
Media +	1,5	12,0	10,3	10,4	13,4	9,8	10,9	12,5	10,3	9,2	10,6	10,3	11,6	11,9	10,1	11,6	10,4	10,4	10,1	9,3	11,0	10,8
Media +	2,0	7,9	6,3	6,3	6,7	7,1	8,0	7,0	6,4	6,9	6,7	8,9	7,4	6,8	7,7	8,9	7,8	6,7	6,6	10,0	6,0	7,3
Media -	1	18,6	17,4	19,5	21,7	20,0	19,0	19,0	18,0	20,4	19,7	15,8	16,5	19,4	20,2	17,9	22,3	21,0	22,6	18,5	19,2	19,3
Media -	1,5	9,9	10,1	9,5	10,1	7,7	9,3	9,8	9,4	8,9	10,7	10,7	10,6	9,3	10,8	10,1	7,8	12,2	11,0	8,9	9,5	9,8
Media -	2,0	6,7	6,0	7,4	8,6	4,7	5,4	6,4	5,8	6,2	6,3	6,7	6,6	6,5	6,9	5,8	6,3	6,2	6,3	7,2	5,7	6,4
Variab -	0,5	15,0	17,3	15,3	18,0	14,7	14,7	16,1	17,4	15,0	17,5	17,0	15,0	15,5	17,1	15,7	15,0	16,1	14,7	15,7	16,7	16,0
Variab -	0,6	21,3	20,2	19,2	17,5	19,7	20,7	18,0	19,2	19,2	18,2	19,0	18,0	19,1	19,2	19,0	18,9	21,5	21,2	18,8	19,4	19,4
Variab -	0,7	32,7	28,6	29,7	25,3	26,5	30,3	23,4	27,7	26,5	28,8	34,1	26,6	33,1	28,9	27,3	26,2	31,7	25,6	33,2	20,9	28,4
Ciclo	1	57,2	57,1	58,2	36,5	53,1	48,7	43,2	48,5	50,5	47,0	51,1	56,4	47,2	53,3	57,1	54,8	59,8	46,8	50,8	69,9	52,4
Ciclo	1,5	31,0	41,0	36,5	37,1	27,4	44,9	35,1	40,2	40,5	41,5	39,2	27,8	31,0	38,1	31,8	49,4	34,1	33,3	36,0	37,0	36,6
Ciclo	2,0	36,5	58,7	119	41,7	74,7	42,3	56,8	74,9	62,9	55,1	74,1	46,1	63,2	43,0	65,9	57,3	92,6	63,0	49,5	50,2	61,3
Mistura	1	46,7	51,0	37,2	55,1	44,9	39,8	60,9	30,1	34,5	63,6	41,9	33,8	37,1	38,0	42,4	54,5	40,1	38,9	60,0	43,8	44,7
Mistura	1,5	21,2	14,7	14,6	21,1	17,6	17,3	23,1	16,6	18,6	14,6	15,6	14,4	18,8	16,3	19,0	13,9	16,0	18,8	17,6	15,1	17,2
Mistura	2,0	8,4	11,8	13,8	10,6	11,3	13,4	14,3	8,9	11,0	10,2	10,4	15,2	8,3	11,4	9,9	9,1	8,7	9,1	11,9	14,1	11,1
Sistemat	1	35,7	27,0	31,9	29,2	26,7	27,2	29,6	25,9	27,4	30,3	33,2	28,5	29,2	24,4	31,3	25,6	32,8	28,3	30,7	27,1	29,1
Sistemat	1,5	12,6	11,4	11,5	13,1	16,5	12,6	10,8	13,8	12,8	10,3	11,1	14,6	11,3	12,2	16,0	10,5	14,2	10,6	11,0	10,5	12,4
Sistemat	2,0	8,0	8,3	9,6	6,6	6,8	7,6	8,3	7,4	10,6	8,3	8,4	10,4	7,9	9,8	10,3	9,8	8,6	11,7	5,6	8,8	8,7
Média Experim.	----	22,6	23,5	26,4	21,6	22,9	21,8	23,2	22,5	22,4	23,5	24,2	21,3	22,1	21,7	23,1	23,7	25,3	22,3	23,5	22,6	23,0

Tabela 7.15 - ARLDIXc para Rede Fuzzy-Artmap

Experimentos																						
Padrões	σ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	X-bar
Media +	1	22,9	22,2	26,3	20,5	23,6	24,9	21,1	24,6	25,6	21,0	23,2	23,7	19,8	18,4	21,1	22,7	22,6	25,1	21,3	21,4	22,6
Media +	1,5	10,1	11,7	10,2	13,5	11,5	10,5	11,3	10,0	12,0	11,4	10,7	9,0	10,9	10,6	10,8	10,4	10,6	9,3	9,8	11,3	10,8
Media +	2,0	6,9	7,6	9,4	8,9	6,2	6,8	8,1	7,1	8,0	7,0	7,2	8,1	8,6	7,2	9,6	6,2	7,3	7,4	7,7	6,7	7,6
Media -	1	18,0	17,5	18,1	17,9	14,9	24,2	14,8	16,1	21,7	19,3	16,6	22,1	17,6	19,5	18,1	19,6	15,3	17,1	16,9	16,8	18,1
Media -	1,5	10,7	10,4	11,1	9,3	10,1	9,2	10,7	12,3	12,9	9,6	10,3	11,0	8,8	11,5	9,7	9,6	11,5	8,5	11,5	10,6	10,5
Media -	2,0	6,7	5,6	6,8	9,1	9,0	7,0	8,3	8,0	7,6	7,1	10,1	9,4	6,5	9,1	8,2	6,6	7,1	8,5	6,6	7,5	7,7
Variab -	0,5	16,9	15,6	16,1	15,3	15,0	16,2	16,6	16,4	15,5	16,5	15,1	16,6	16,8	14,5	17,0	16,3	17,2	16,6	17,7	14,5	16,1
Variab -	0,6	20,0	22,0	20,7	19,4	20,5	21,6	18,2	21,2	19,4	20,0	17,5	20,2	20,3	17,5	20,3	18,8	18,1	18,7	17,8	16,4	19,4
Variab -	0,7	33,3	27,5	29,1	28,7	28,0	29,4	27,1	28,8	37,1	31,4	35,1	32,0	26,8	29,9	28,4	27,2	30,7	28,3	28,4	31,1	29,9
Ciclo	1	48,1	73,2	58,5	71,4	46,8	62,8	56,5	58,0	76,9	49,8	54,1	53,7	63,2	54,4	47,7	62,0	57,3	57,6	60,6	46,3	57,9
Ciclo	1,5	43,4	43,0	37,1	33,7	38,8	38,2	38,0	33,4	48,0	50,8	50,3	39,0	34,5	36,3	44,1	39,1	34,5	41,9	46,4	40,7	40,5
Ciclo	2,0	58,8	42,7	72,5	47,0	81,7	61,0	53,8	36,0	45,2	51,8	42,9	74,8	70,0	41,8	87,8	43,5	68,7	49,3	124	60,3	60,7
Mistura	1	41,4	29,9	36,4	38,5	29,7	30,9	31,9	37,7	37,9	33,6	31,7	40,2	31,9	25,0	41,2	31,0	40,7	31,9	34,9	30,5	34,3
Mistura	1,5	13,3	15,6	14,2	11,7	12,7	13,2	13,8	13,2	11,9	13,1	13,8	14,8	11,8	13,2	14,9	12,1	14,5	14,2	12,9	14,8	13,5
Mistura	2,0	7,0	8,0	7,8	8,0	6,6	8,4	7,1	7,2	10,7	8,1	7,8	7,1	7,4	5,7	5,9	8,4	6,5	5,6	7,4	7,9	7,4
Sistemat	1	25,5	23,0	25,0	27,6	25,5	26,1	20,2	32,9	25,6	20,5	28,0	29,3	25,6	27,9	33,7	25,3	24,6	29,2	24,4	30,4	26,5
Sistemat	1,5	14,4	14,9	12,1	12,1	14,3	11,8	11,0	11,6	11,8	12,5	12,8	11,3	11,4	13,3	13,8	13,8	11,9	11,6	16,0	11,5	12,7
Sistemat	2,0	8,5	12,1	8,1	6,3	10,8	9,1	9,4	8,5	8,0	10,0	8,8	9,5	9,6	10,1	9,7	11,5	8,9	12,5	9,1	11,4	9,6
Média Experim.	----	22,6	22,4	23,3	22,1	22,5	22,8	21,0	21,3	24,2	21,9	22,0	24,0	22,3	20,3	24,6	21,3	22,7	21,9	26,3	21,7	22,6

Tabela 7.16 - ARLDIXc para Rede RBF

De uma maneira geral, verifica-se que o desempenho diminui com o aumento da amplitude dos padrões. Isto ocorre porque, com o aumento da amplitude, mais facilmente ocorrerão pontos fora dos limites de controle antes de haver informação suficiente para uma classificação correta. A capacidade das redes identificarem os padrões como indefinidos torna-se importante nestes casos. O ARLDIXc diminui com o aumento da amplitude dos padrões. Este resultado reflete a grande diminuição do ARL devido a ocorrência de alarmes estatísticos ou pontos fora dos limites de controle, fazendo com que a diminuição do ARL seja maior que a diminuição do desempenho corrigido. Este comportamento só não é verdadeiro para os ciclos, onde o ARLDIXc tem uma diminuição entre as amplitudes de 1 e $1,5\sigma$, aumentando para a amplitude de 2σ . Isto ocorre devido ao grande aumento de classificações "incorretas", pois como o ARL fica pequeno, os padrões são classificados como desvio da média ou mistura.

7.3.3 - Análise do ARL para um Alarme Falso

O tempo para um alarme falso foi medido diretamente pelo ARL médio de 50 simulações com o processo sob condição normal. Neste caso, quanto maior o ARL melhor, pois mais tempo a rede operará antes de um alarme falso. A rede Fuzzy Artmap apresentou um ARL de 165,9, menor que o ARL para a rede RBF-Fuzzy Artmap (180,1) e menor que o ARL para a rede RBF (190,9). O teste "t" com 95% de confiança indicou que a rede RBF é claramente diferente da rede Fuzzy-Artmap, porém não se pode afirmar o mesmo para a rede RBF-Fuzzy Artmap. Também não se pode afirmar que a rede RBF-Fuzzy Artmap tem ARL diferente da rede Fuzzy-Artmap. Assim, a rede RBF-Fuzzy Artmap está em uma situação intermediária entre as outras duas redes.

Champ & Woodwal [46] apresentam o ARL para gráficos de X-Barra de Shewhart, com e sem as regras suplementares propostas pela western electric [42]. Uma comparação do ARL

(Tabela 7.17) para alarmes falsos das redes com os resultados apresentados em [46], demonstra que as redes são mais robustas a alarmes falsos que os gráficos que utilizam "runs rules", porém menos robustas que o gráfico de Shewhart sem regras suplementares. Por outro lado, a eficiência das redes é maior para pequenos desvios na média do processo.

Desvios da Média	0 σ	1,0 σ	1,5 σ	2,0 σ
Shewhart	370,4	43,90	16,20	6,30
Shewhart + Regras	91,75	9,22	5,00	3,13
CUSUM $k = 0,5 \quad h = 4,78$	407,4	10,09	5,99	3,84
Fuzzy Artmap	165,9	19,91	9,10	5,36
RBF-Fuzzy Artmap	180,1	18,32	9,00	5,92
RBF	190,9	19,33	9,10	5,82

Tabela 7.17 - ARL para Alarme Falso e Desvios da Média

A tabela 7.18 mostra uma comparação entre as redes neurais, gráficos de CUSUM, e gráficos de Shewhart com e sem regras suplementares.

	Robutez a Alarme Falso	Sensibilidade a Pequenos Desvios da Média	Capacidade de Aprendizado	Capacidade de Identificar Padrões
Shewhart	Grande	Pequena	Não	Nenhuma
Shewhart + Regras	Pequena	Grande	Não	Pouca
CUSUM	Grande	Grande	Não	Nenhuma
Redes	Média	Médio	Sim	Grande

Tabela 7.18 - Comparação entre Diversos Métodos de CEP

Esta tabela mostra que as redes neurais estudadas apresentam vantagens sobre as técnicas tradicionais de CEP, principalmente quando se quer identificar padrões. Além disso, devido a sua capacidade de reconhecimento, as redes neurais podem reconhecer um padrão antes da ocorrência de um alarme estatístico.

7.4 - CONCLUSÕES

Para as condições e parâmetros utilizados nesta tese, pode-se dizer que a rede RBF-Fuzzy Artmap tem melhor desempenho corrigido, maior ARL para alarme falso e muito menos sensibilidade a ordem de apresentação dos padrões que a rede Fuzzy-Artmap, enquanto não tem a necessidade de aprendizado "off-line" da rede RBF.

De uma maneira geral, pode-se dizer que a rede RBF apresenta o melhor desempenho de classificação das três redes analisadas, com um ARL menor para detecção de padrões anormais, e com o maior ARL para alarmes falsos. A presença da segunda camada aumenta seu desempenho em termos de classificações corretas, e ajuda na diminuição do ARL. Porém, não ajuda na classificação de padrões duvidosos como indefinidos. Além disso, a rede RBF necessita de treinamento "off-line", não podendo ser atualizada durante sua operação.

A rede Fuzzy-Artmap apresenta um ARL menor para os padrões de ciclo, o que faz seu ARLDIXc geral ficar igual ou um pouco melhor que para a rede RBF-Fuzzy Artmap, mesmo tendo desempenho corrigido menor. Além disso, ela tem grande sensibilidade a ordem de apresentação dos padrões para treinamento.

8 - CONCLUSÃO

A crescente demanda por produtos cada vez mais diversificados, e de maior qualidade, tem levado a uma crescente automatização dos processos de fabricação, e a uma diminuição acentuada dos tamanhos de lote e intervalo entre as amostras de controle.

A necessidade de intervalos pequenos entre as amostras, de maior precisão / qualidade dos produtos, e a crescente automatização dos processos de manufatura, tem despertado o interesse na automatização do controle da qualidade do processo. Um sistema automatizado de controle da qualidade deve satisfazer as seguintes necessidades :

- a - Tratar simultâneamente diversos tipos de informações, provenientes de diversas fontes, analisando-as em tempo real;
- b - Adaptar-se a mudanças no processo e "aprender" com os problemas e exemplos ocorridos durante sua operação, mesmo que esses exemplos não tenham sido estruturados;
- c - Estabelecer correlações entre um grande número de variáveis, mesmo que essas relações não sejam bem definidas e que haja interação entre as variáveis;
- d - Tratar situações novas, generalizando através da experiência passada;
- e - Tratar relações não lineares;
- f - Tratar problemas difusos, comuns de se encontrar na vida real;
- h - Auxiliar o analista/operador no diagnóstico de problemas da qualidade e no controle do processo;
- i - Integrar o controle das causas comuns ao controle e eliminação das causas assinaláveis de variação;

As Janelas de Controle de Qualidade (JCQ) [60][61] são uma ferramenta importante na automatização do controle da qualidade em processos de fabricação. As JCQ dividem o controle da qualidade do processo em 5 etapas: observação, avaliação, diagnóstico, decisão e implementação. Dessas etapas, a etapa de avaliação é a mais importante, pois ela é responsável por detectar desvios na normalidade do processo, evitando a fabricação de itens defeituosos. Além disso uma vez que se tenha uma avaliação adequada, o diagnóstico dos problemas e a decisão das ações corretivas são extremamente facilitadas. A etapa de avaliação das JCQ é composta por gráficos de controle estatístico de processos (CEP).

As técnicas tradicionais de controle estatístico de processos tem aplicação limitada em ambientes de manufatura flexível automatizada. Devido ao pequeno tamanho dos lotes e a precisão requerida, o intervalo entre as amostras deve ser reduzido, chegando ao limite de 1 peça. Com a diminuição do intervalo entre as amostras, a autocorrelação das amostras se acentua, limitando a aplicabilidade dos gráficos tradicionais de controle estatístico de processos.

O mesmo problema de autocorrelação existe também na indústria de processos (ex: indústrias químicas), onde muitas vezes o processo tem uma variação previsível devido a temperatura ambiente, tempo de operação, aumento ou diminuição da concentração de reagentes, etc. Estas causas, chamadas de causas comuns de variação, são normalmente controladas, na indústria de processos, por controladores do tipo PID. Esses controladores muitas vezes não são capazes de compensar todas as variações do processo, estando também sujeitos a defeitos.

Isto leva a necessidade de utilização de sistemas automatizados de CEP inteligentes, que possam levar em consideração as autocorrelações e identificar os tipos de padrão presentes.

Os trabalhos sobre a automatização dos gráficos de CEP encontrados na literatura [45][47][54][55][56][57][58], não satisfazem as necessidades de um sistema de controle da qualidade automatizado, além de apresentarem uma série de limitações quanto ao escopo e

aplicabilidade. A principal limitação dos sistemas desenvolvidos é a falta de um mecanismo de adaptação "on-line", que permita aos sistemas incorporar novos conhecimentos durante sua operação.

A utilização de redes neurais com lógica difusa permite a satisfação das necessidades apontadas acima, e a superação das limitações dos trabalhos anteriores.

Este trabalho de tese analisou três tipos de redes neurais :

- Rede de Funções com Base Radial (RBF), que é funcionalmente equivalente a um sistema especialista com lógica difusa [63];
- Rede Fuzzy-Artmap, que utiliza operações difusas para armazenar e processar as informações;
- e uma nova rede, Rede RBF-Fuzzy Artmap, formada pela união das duas redes anteriores.

O desempenho dessas três redes na identificação de padrões em gráficos de CEP foi analisado, procurando-se simular situações encontradas na realidade, superando assim uma das limitações dos trabalhos anteriores.

As redes foram treinadas e otimizadas para reconhecer seis tipos diferentes de padrões com diversos níveis de distúrbio: desvio positivo da média, desvio negativo da média, diminuição da variabilidade, ciclo, mistura e padrão sistemático. Cada rede teve seu vetor de entrada e parâmetros internos otimizados através de um delineamento experimental utilizando a matriz ortogonal L_{27} de Taguchi. Os resultados de desempenho e os fatores utilizados nos experimentos, foram utilizados para o ajuste de uma equação de regressão. Esta equação foi otimizada pelo programa "GAMS"[76], encontrando-se o nível ótimo dos fatores para cada rede.

Os desempenhos das redes otimizadas, na identificação dos padrões citados acima, foram comparados através de um programa de comparação, que simula situações encontradas na vida

real, composto de três módulos: módulo gerador de padrões, módulo de análise estatística e módulo de avaliação. O módulo gerador simula pontos provenientes de um processo que é inspecionado a 100%, com amostras de tamanho 1. A cada novo ponto, o programa gerador forma um padrão com os últimos 60 pontos, calculando: a média e o desvio padrão de 15 janelas estatísticas nesses 60 pontos, 10 "lags" de autocorrelação dos pontos individuais, duas estatísticas de soma cumulativa, e uma estatística de quiquadrado. Essas estatísticas são utilizadas para a montagem do vetor de entrada das redes neurais, e passadas para o módulo de análise estatística, que verifica se houve algum alarme estatístico através da violação dos limites de controle dos gráficos de Shewhart, CUSUM e Quiquadrado. O vetor de entrada e o resultado da análise estatística é passado ao módulo de avaliação, que contém uma das três redes neurais. O módulo de avaliação identifica o tipo de padrão presente, caso haja algum alarme dado pelo módulo de análise estatística. Caso o módulo de análise estatística não tenha dado nenhum alarme, o módulo de avaliação verifica se o processo se desviou da condição normal, dá alarme e identifica o tipo de padrão presente.

As redes foram comparadas pelo resultado do desempenho de classificação médio de 1000 identificações para cada nível e cada tipo de padrão, totalizando 18000 identificações para cada rede. Foram utilizados como critério de comparação o desempenho corrigido (percentual de classificações corretas retirando os padrões identificados como indefinidos - Eq. 7.1) e o ARLDIXc (razão entre número de amostras até a identificação do padrão e o desempenho - Eq. 7.2).

A rede Fuzzy-Artmap apresentou um desempenho de classificação médio de 74.34%, a rede RBF apresentou 77.78%, e a nova rede, RBF-Fuzzy Artmap, apresentou desempenho de 78.80%. Uma análise estatística pelo teste "t" e ANOVA indicou que as diferenças são significantes com 95% de confiança. As redes Fuzzy-Artmap e RBF-Fuzzy Artmap tiveram

ARLDIXc semelhantes (23.4 e 23), com a rede RBF apresentando um resultado um pouco melhor (22.6).

Em termos de tempo para um alarme falso quando o processo está sob condição normal, a rede Fuzzy-Artmap apresentou o menor ARL (165.9), seguido da rede RBF-Fuzzy Artmap (180.1), e pela rede RBF (190.9). Estes resultados são piores que para um gráfico de Shewhart (370.4), porém melhores que para um gráfico de Shewhart com as regras suplementares da Western Electric (91.75) [42].

Não é possível realizar-se uma comparação direta com os trabalhos anteriores, pois muitos não utilizam os mesmos índices de comparação, e/ou não fazem a medição do desempenho em situações onde é possível os pontos saírem fora dos limites de controle.

Pode-se concluir então que a rede RBF-Fuzzy Artmap é a que obteve melhor desempenho das três redes analisadas. Além disso, ela possui a capacidade de aprendizado instantâneo e incremental, o que não ocorre com a rede RBF, e é 98% menos sensível a ordem de apresentação dos padrões de treinamento que a rede Fuzzy-Artmap. A rede RBF-Fuzzy Artmap satisfaz as necessidades de um sistema automatizado de controle da qualidade, podendo formar a base para o desenvolvimento de um sistema mais completo, que incorpore a fase de diagnóstico da Janela de Controle da Qualidade. Por outro lado, a rede RBF-Fuzzy Artmap necessita de um número muito maior de nós na camada de funções radiais que a rede Fuzzy-Artmap na rede ARTa (222 contra 47 nas redes aqui analisadas).

Os resultados deste trabalho demonstram o forte potencial de aplicabilidade de redes neurais com lógica difusa na identificação de padrões em gráficos de CEP.

A inclusão dessas redes neurais dentro das janelas de controle da qualidade permite uma automatização incremental, mas integrada, das várias etapas do controle da qualidade do processo, permitindo mais um passo em direção a automatização dos sistemas de manufatura flexíveis.

9 - SUGESTÕES PARA CONTINUIDADE DA PESQUISA

9.1 - NA ÁREA DE CONTROLE ESTATÍSTICO DE PROCESSOS

Este trabalho estudou apenas uma das fases da Janela de Controle da Qualidade. O estudo da automatização das outras fases, com mais atenção sendo dada a etapa de diagnóstico, é ainda um trabalho importante a ser realizado. Devido suas características, a rede RBF-Fuzzy Artmap é também uma forte candidata a aplicação na fase de diagnóstico.

Um estudo da integração das diversas etapas da Janela de Controle de Qualidade ajudará muito no desenvolvimento de um sistema integrado e automatizado de controle da qualidade do processo. Este estudo pode utilizar um processo de fabricação simulado através de um modelo, ou ser um processo real.

Este trabalho não explorou o uso de estatísticas dedicadas nas entradas das redes para cada tipo de padrão que se quer identificar. A utilização de estatísticas dedicadas pode aumentar em muito a capacidade e velocidade de reconhecimento das redes.

A identificação de mistura de padrões é um fator importante para o diagnóstico de problemas da qualidade. Por suas características, as redes neurais aqui estudadas podem ter um desempenho significativo nesta identificação.

Por fim, para que as Janelas de Controle da Qualidade se tornem uma realidade, é necessário que os estudos futuros para o seu desenvolvimento levem em consideração as necessidades de um sistema automatizado de controle da qualidade apresentados neste trabalho, diminuindo ao máximo possível as hipóteses simplificativas e restritivas.

9.2 - NA ÁREA DE REDES NEURAIAS

Esta primeira versão da rede RBF-Fuzzy Artmap tem a largura das funções de base radial fixas a priori, com o treinamento sendo executado através do deslocamento das funções existentes ou criação de novas funções. A utilização de um mecanismo de variação da largura das funções radiais durante o treinamento, poderia aumentar ainda mais seu desempenho e diminuir o número de nós. As funções de base radial poderiam começar com uma largura pequena, e expandir durante o aprendizado. Neste caso a rede começa com um conhecimento específico e vai generalizando com o tempo. Outra opção seria começar com uma largura maior, e reduzir a largura das funções radiais nas dimensões onde a variação do padrão de entrada fosse pequena. Neste caso a rede se especialisaria com o aprendizado. A sensibilidade a ordem de apresentação dos padrões de treinamento é uma característica importante que deve ser levada em consideração.

10 - BIBLIOGRAFIA

- 1 - Box,G.E.P., Hunter,W.G., Hunter, J.S., "Statistics for Experimenters", Willey, New York, 1978, 653p;
- 2 - Box,G.E.P. , Jenkins,E.M., "Times Series Analysis: Forecasting and Control", San Francisco, Holden Day, 1970;
- 3 - Box,G.E.P., Tiao,G.C.,"Intervention Analysis With Applications to Economic and Environmental Problems", Journal of the American Statistical Association, Vol.70, 349, 1975, pp 70-79;
- 4 - Helmer,R.D., Johansson,J.K.,"An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising-Sales Relationship", Journal of Marketing Research, Vol 14, May 1977, pp 227-239;
- 5 - Phadke,M.S., Wu,S.M.,"Identification of Paper Making Process From Closed Loop Data by Bivariate Time Series Analysis", Journal of Dynamic Systems, Measurement and Control, September, 1976, pp 291-295;
- 6 - Tee,L.H., Wu,S.M., "An Application of Stochastic and Dynamic Models for the Control of a Paper Making Process", Technometrics, Vol.14, 2, 1972, pp 481-496;
- 7 - Phadke,M.S., Wu, S.M., "Identification of Multi-Input Multi-Output Transfer Function and Noise Model of a Blast Furnace from Closed Loop Data", IEEE Transactions on Automatic Control, Vol 19, 6, 1974, pp 944-951;
- 8 - Wu,S.M., "Dynamic Data System: A New Modeling Approach", Journal of Engineering for Industry, August 1977, pp 708-714;
- 9 - Pandit,S.M., Wu,S.M., "Times Series and System Analysis with Applications", John Willey and Sons Inc, 1983;

- 10- Yuan,J.X., Tang,X.J., Wu,S.M., "Digital Filtering for Machine Tool Structural Dynamic Analysis by the Dynamic Data System (DDS)", International Journal of Machine Tool Design Research, Vol 26, 3, 1986, pp 267-281;
- 11- Cho,D.W., Eman,K.F., Wu,S.M., "A New Time Domain Multiple Input Modal Analysis Method", Journal of Engineering for Industry, Vol 109, 11, 1987, pp377-384;
- 12- Ahn,T.Y., Eman,K.F., Wu,S.M., "Cutting Dynamics Identification by Dynamic Data System (DDS) Modelling Approach", Journal of Engineering for Industry, Vol 107, 2, 1985, pp 91-94;
- 13- Ahn,T.Y., Eman,K.F., Wu,S.M., "Identification of the Transfer Function of Dynamic Cutting Process - A Comparative Assessment", International Journal of Machine Tool Design Research, Vol 25, 1, 1985, pp 75-90;
- 14- Ahn,T.Y., Eman,K.F., Wu,S.M., "Determination of Inner and Outer Modulation Dynamics in Orthogonal Cutting", Journal of Engineering for Industry, Vol 109, 9, 1987, pp 275-280;
- 15- Cho, D.W., Eman, K.F., "In-Process Identification of the Milling Operation", International Journal of Machine Tools Manufacturing, Vol 30, 3, 1990, pp 325-337;
- 16- Spiewak,S., Wu,S.M., "Tool Wear Monitoring and Breakage Detection Based on Intelligent Filtering", International Journal of Machine Tools Manufacturing, Vol 28, 4, 1988, pp 483-494;
- 17- Bandyopadhyay,P., Gonzalez,E.M.,Huang,R.,Wu,S.M., "A Feasibility Study of On-Line Drill Wear Monitoring by DDS Methodology", International Journal of Machine Tool Design Research, Vol 26, 3, 1986, pp 245-257;

- 18- Fang,X.D., Yao,Y., "Modelling of Multivariate Time Series for Tool Wear Estimation in Finish-Turning", International Journal of Machine Tools Manufacturing, Vol 32, 4, 1992, pp 495-508;
- 19- Kim,K.,Eman,K.F.,Wu,S.M., "Analysis of Alignment Errors in a Laser-Based In-Process Cylindricity Measurement System", Journal of Engineering for Industry, Vol 109, 11, 1987, pp 321-329;
- 20- Kim,K.,Eman,K.F.,Wu,S.M., "In-Process Control of Cylindricity in Boring Operations", Journal of Engineering for Industry, Vol 109, 11, 1987, pp 291-296;
- 21- Kim,K., "Cylindrical Accuracy Control Based on Stochastic Modeling and Forecasting Compensation", International Journal of Machine Tools Manufacturing, Vol 28, 4, 1988, pp 495-501;
- 22- Kim,K.,Eman,K.F.,Wu,S.M., "Development of a Forecasting Compensatory Control System for Cylindrical Grinding", Journal of Engineering for Industry, Vol 109, 11, 1987, pp 385-391;
- 23- Ivakhnenko,A.G, Touda, M.M., Chukim, Yu.V., "GMDH Algorithm with Successive Determination of Polynomial Trends Using Most Essential Variables", Sovie Automatic Control, 5, 1972, pp 44-54;
- 24- Yoshida, T., Nagasaka,K., Kita,Y., Hashimoto,F., "Identification of a Grinding Wheel Wear Equation of the Abrasive Cut-Off by the Modified GMDH", International Journal of Machine Tool Design Research, Vol 26, 3, 1986, pp 283-292;
- 25- Nagasaka,K., Yoshida, T.,Kita,Y., Hashimoto,F., "Optimum Combination of Operating Parameters in Abrasive Cut-Off", International Journal of Machine Tools Manufacturing, Vol 27, 2, 1987, pp 167-179;

- 26- Box,G.E.P., Hunter,W.G., Hunter, J.S., "Statistics for Experimenters", Willey, New York, 1978;
- 27- Taguchi,G., Wu,Y., "Introduction to Off-Line Quality Control", Central Japan Quality Control Association, 1985;
- 28- Maren,J.A., Harston,C.T., Pap,R.M., "Handbook of Neural Computing Applications", Academic Press Inc., San Diego, 1990;
- 29- Chen,S., Billings,S.A., "Neural Networks for Nonlinear Dynamic System Modelling and Identification", International Journal of Control, Vol 56, 2, 1992, pp 319-346;
- 30- Hornik, K., "Approximation Capabilities of Multilayer Feedforward Networks", Neural Networks, 4, 1991, pp 251-257;
- 31- Tansel, I.N., "Modelling 3-D Cutting Dynamics with Neural Networks", International Journal of Machine Tools Manufacturing, Vol 32, 6, 1992, pp 829-853;
- 32- Yao,Y.L., Fang,X.D., "Assessment of Chip Forming Patterns with Tool Wear Progression in Machining Via Neural Networks", International Journal of Machine Tools Manufacturing, Vol 33, 1, 1993, pp 89-102;
- 33- Sathyanarayanan,G., Lin,I.J., Chen,M.K., "Neural Network Modelling and Multiobjective Optimization of Creep Feed Grinding of Superalloys", International Journal of Control,Vol 55,2, 1992;
- 34- Cook, D.F., Shannon, R.E., "A Predictive Neural Network Modelling System for Manufacturing Process Parameters", International Journal of Production Research,Vol 30, 7, 1992, pp 1537-1550;
- 35- Pao, Y., Phillips, S.M., Sobajic, D.J., "Neural-Net Computing and Intelligent Control of Systems", International Journal of Control, Vol 55, 2, 1992, pp 283-289;

- 36- Watanabe,K., Fukuda,T.,Tzafestas,S.G., "An Adaptive Control for CARMA Systems Using Linear Neural Networks", International Journal of Control, Vol 55, 2, 1992, pp 483-497;
- 37- Chen,F.C., Khalil,H.K., "Adaptative Control of Nonlinear Systems Using Neural Network", International Journal of Control, Vol 55, 6, 1992, pp 1299-1317;
- 38- Box,G., Kramer,T., "Statistical Process Monitoring and Feedback Adjustment - A Discussion", Technometrics, Vol 34, 3, 1992, pp 251-267;
- 39- Shewhart,W.A., "Economic Control of Quality of Manufactured Products", Princeton, NJ, 1931, Van Nostrand Reinhold;
- 40- Kume,H., "Statistical Methods for Quality Improvement", Tokyo, 1987, AOTS;
- 41- Montgomery,D.C., "The Use of Statistical Process Control and Design of Experiments in Product and Process Improvement", IIE Transactions, Vol 34, 5, 1992;
- 42- Western Electric Co. Inc., "Statistical Quality Control Handbook, Western Electric Co. Inc., New York, NY, 1958;
- 43- Nelson L.S., "Interpreting Shewhart X-Bar Control Charts", Journal of Quality Technology, Vol. 17, 2, April 1985;
- 44- Nelson L.S., "The Shewhart Control Chart Test for Special Causes", Journal of Quality Technology, Vol. 16, 1984, pp237-239;
- 45- Hwang H.B., Hubele N.F., "Back-Propagation Pattern Recognizer for X-Bar Control Charts: Methodology and Performance", Computers in Industrial Engineering, Vol 24, 2, 1993, pp 219-235;
- 46- Champ C.W., Woodwall W.H., "Exact Results for Shewhart Control Charts with Supplementary Runs Rules", Technometrics, Vol 29, 4, 1987, pp393-399;

- 47- Hwang H.B., "Pattern Recognition on Shewhart Control Charts Using a Neural Network Approach", Tese de Doutorado, Arizona State University, Tempe, AZ, 1992;
- 48- Lucas, J.M., "The Design and Use of V-Mask Control Schemes", *Journal of Quality Technology*, 8, 1976, pp 1-12;
- 49- Gibra, I.N., "Recent Developments in Control Charts Techniques", *Journal of Quality Technology*, Vol.7, 1975, pp 183-192;
- 50- Van Doben de Bruyn C.S., "Cumulative Sum Tests", Hafner Publications, London, 1968;
- 51- Lucas, J.M., Saccucci, M.S., "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements", *Technometrics*, Vol 32, 1, 1990, pp 23-26;
- 52- Hunter, J.S., "The Exponentially Weighted Moving Average", *Journal of Quality Technology*, Vol 18, 4, 1986, pp 203-210;
- 53- Yang, K., Hancock, W.M., "Statistical Quality Control for Correlated Samples", *International Journal of Control*, Vol.55, 6, 1992, pp 1260-1275;
- 54- Swift J.A., "Development of a Knowledge-Based Expert System for Control Chart Pattern Recognition and Analysis", Tese de Doutorado, Oklahoma State University, 1987;
- 55- Cheng C.S., "Group Technology and Expert Systems Concepts Applied to Statistical Process Control in Small-Batch Manufacturing", Tese de Doutorado, Arizona State University, Tempe, AZ, 1989;
- 56- Hwang H.B., Hubele N.F., "X-Bar Control Chart Pattern Identification Through Efficient Off-Line Neural Network Training", *IIE Transactions*, Vol 25, 3, May 1993, pp27-39;
- 57- Guo, Y., Dooley, K.J., "Identification of Change Structure in Statistical Process Control", *International Journal of Production Research*, Vol 30, 7, 1992, pp 1655-1669;

- 58- Lucy-Bouler T.L., "Using Autocorrelations, CUSUMs and Runs Rules for Control Chart Pattern Recognition : An Expert System Approach", Tese de Doutorado, University of Alabama, Tuscaloosa, AL, 1991;
- 59- Lucy-Bouler T.L., "Problems in Control Chart Pattern Recognition Systems", *International Journal of Quality & Reliability Management*, Vol 10, 8, 1993, pp5-13;
- 60- Dessouky,M., Kapoor,S., DeVor,R., "A Methodology for Integrated Quality Systems", *Journal of Engineering for Industry*, Vol 109, 3, 1987, pp 241-247;
- 61- Dooley,K., Kapoor,S., Dessouki,M., DeVor,R., "An Integrated Quality System Approach to Quality and Productivity Improvement in Continuous Manufacturing Processes", - *Journal of Engineering for Industry*, Vol 108, 4, 1986, pp 322-327;
- 62- Zadeh L.A., "Fuzzy Sets", *Information and Control*, 8, 1965, pp338-353;
- 63- Jang J.S.R., Sun C.T., "Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems", *IEEE Transactions on Neural Networks*, Vol 4, 1, January 1993, pp156-158;
- 64- Moody J., Darken C.J., "Fast Learning in Networks of Locally Tuned Processing Units", *Neural Computation*, 1, 1989, pp281-294;
- 65- Musavi M.T., Ahmed W., Chan K.H., Faris K.B., Hummels D.M., "On the Training of Radial Basis Function Classifiers", *Neural Networks*, Vol 5, 1992, pp595-603;
- 66- Linkens D.A., Nie J., "Learning Control Using Fuzzified Self-Organizing Radial Basis Function Network", *IEEE Transactions on Fuzzy Systems*, Vol 1, 4, 1993, pp280-287;
- 67- Grossberg S., "Adaptive Pattern Classification and Universal Recoding II : Feedback, Expectation, Olfaction and Illusions", *Biological Cybernetics*, 23, 1976, pp187-202;

- 68- Carpenter G.A., Grossberg S., "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine", *Computer Vision, Graphics, and Image Processing*, 37, 1987, pp54-115;
- 69- Carpenter G.A., Grossberg S., "ART 2 : Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns", *Applied Optics*, 26, 1987, pp4919-4930;
- 70- Carpenter G.A., Grossberg S., "ART 3 Hierarchical Search : Chemical Transmitters in Self-Organizing Pattern Recognition Architectures", *Proceedings of the International Joint Conference on Neural Networks*, Washington ,DC, 1990, ppII30-II33;
- 71- Carpenter G.A., Grossberg S., "ART 3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures", *Neural Networks*, 3, pp129-152, 1990;
- 72- Carpenter G.A., Grossberg S., "ARTMAP : Supervised Real-Time Learning and Classification of Non-Stationary Data by a Self-Organizing Neural Network", *Neural Networks*, 4, pp565-588, 1991;
- 73- Carpenter G.A., Grossberg S., Rosen D.B., "Fuzzy ART: An Adaptive Resonance Algorithm for Rapid, Stable Classification of Analog Patterns", *Proceedings of the International Joint Conference on Neural Networks*, Seattle, 1991, ppII411-II416;
- 74- Carpenter G.A., Grossberg S., Rosen D.B., "Fuzzy ARTMAP: An Adaptive Resonance Architecture for Incremental Learning of Analog Multidimensional Maps", *Proceedings of the International Joint Conference on Neural Networks*, Baltimore , 1992, ppIII309-III314;
- 75- Miller, R.B., "Minitab Handbook for Business and Economics", PWS-Kent Publishing Company, Boston, MA, 1987;
- 76- GAMS User Manual, 1989;

- 77- Ott L., "An Introduction to Statistical Methods and Data Analysis", 3^o Ed., PWS-KENT Publishing Company, Boston MA, 1988;
- 78- MacGregor J.F., "On-Line Statistical Process Control", Chemical Engineering Progress, October, 1988, pp 21-31..

APÊNDICE 1 - GERAÇÃO DOS PADRÕES

Os padrões do processo sob condição normal são gerados pela seguinte equação:

$$y(t) = \mu + a(t) \quad \text{Eq.A1}$$

onde, média do processo $\mu = 0$;

$a(t)$ é um ruído com média zero e variância $\sigma^2 = 1$.

Os padrões com anomalia são gerados da seguinte maneira:

Desvio da Média

$$y(t) = \mu + a(t) + (-) \delta \sigma \quad \text{Eq.A2}$$

onde, δ é o desvio da média em termos de múltiplos do desvio padrão σ . Neste trabalho δ pode ser 1, 1,5 ou 2,0.

Ciclo

$$y(t) = \mu + a(t) + \sin[2\pi(t-t_0)/\Omega] \delta \sigma \quad \text{Eq.A3}$$

onde, δ = amplitude do ciclo em termos de σ ($\delta = 1,0, 1,5$ e $2,0$);

$\Omega = 20$, período do ciclo;

t_0 = ponto em que começou o distúrbio.

Diminuição da Variabilidade

$$y(t) = \mu + \delta \sigma a(t) \quad \text{Eq.A4}$$

onde, $\delta = 0,5, 0,6$ e $0,7$.

Mistura

$$y(t) = \mu + a(t) + (-1)^w \delta \sigma \quad \text{Eq.A5}$$

onde, δ = magnitude da mistura em termos de σ ($\delta = 1,0, 1,5$ e $2,0$);

$w = 0$ se $p < P$, 1 se $p \geq P$;

p = número aleatório, $0 < p < 1$;

$P = 0,5$, probabilidade pré-especificada que determina a mudança entre as distribuições.

Padrão Sistemático

$$y(t) = \mu + a(t) + (-1)^t \delta \sigma \quad \text{Eq.A6}$$

onde, t = número da amostra, fazendo que os amostras se desviem positivamente e negativamente alternadamente;

δ = magnitude da mistura em termos de σ ($\delta = 1,0, 1,5$ e $2,0$).