

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
DE PRODUÇÃO**

**UM PROCEDIMENTO PARA AVALIAÇÃO DA
SAÚDE FINANCEIRA DE PEQUENAS EMPRESAS:
ESTUDO DE UM CASO USANDO REDES
NEURONAIAS ARTIFICIAIS**

GERTRUDES APARECIDA DANDOLINI

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de produção da Universidade Federal de Santa Catarina para obtenção do Graus de Mestre em Engenharia, com especialidade em Engenharia de Produção.



0.273.753-2

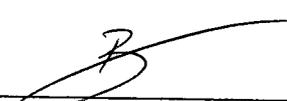


*Florianópolis
Novembro/1997*

GERTRUDES APARECIDA DANDOLINI

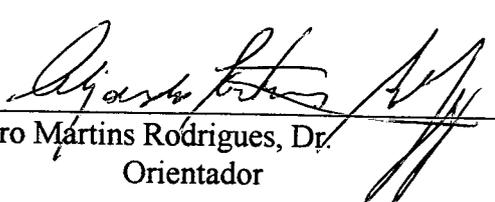
**UM PROCEDIMENTO PARA AVALIAÇÃO DA SAÚDE
FINANCEIRA DE PEQUENAS EMPRESAS: ESTUDO DE UM
CASO USANDO REDES NEURONASIA ARTIFICIAIS**

Esta dissertação foi julgada para obtenção do Título de Mestre, Especialidade em Engenharia de Produção e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

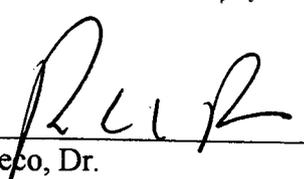


Ricardo Miranda Barcia, Ph.D.
Coordenador

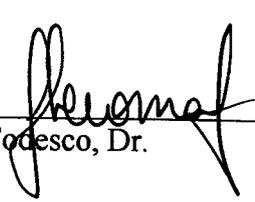
Banca Examinadora



Alejandro Martins Rodrigues, Dr.
Orientador



Roberto C. S. Pacheco, Dr.



José Leomar Todesco, Dr.

João Artur e Heitor

AGRADECIMENTOS

Agradecer a todos que contribuíram para realização deste trabalho, constitui-se numa tarefa bastante difícil, pois foram inúmeras as contribuições que recebi.

Manifestarei aqui meus agradecimentos a algumas pessoas e instituições, sem os quais não teria realizado este trabalho.

À CAPES pelo apoio financeiro.

À Universidade Federal de Pelotas, pela liberação de minhas atividades acadêmicas para realização deste Mestrado, principalmente aos companheiros de trabalho que me apoiaram nesta liberação.

À Universidade Federal de Santa Catarina, em especial ao Programa de Pós-Graduação em Engenharia de Produção em nome do Prof. Ricardo Miranda Barcia, por ter me recebido de braços abertos.

A Sérgio Farraco (Presidente Regional de Contabilidade) pelo fornecimento dos dados usados nesta dissertação, sem os quais o trabalho não cumpriria com os objetivos.

Aos Profs. Roberto Pacheco e José Leomar Todesco pelo incentivo e orientação durante o estágio inicial deste aprendizado.

Ao professor e Orientador Alejandro Martins Rogriguez, pelo acompanhamento, orientação e apoio desde os passos iniciais até a conclusão deste trabalho.

A todos os amigos, em especial ao Roberto Raitz, pela colaboração, apoio e amizade.

Aos meus pais, Romeu e Adulce, dos quais nunca faltaram muito amor e incentivo, durante todo o meu desenvolvimento pessoal e profissional.

É um agradecimento todo especial as duas pessoas que acompanharam cada passo desta longa trajetória. Por aquele rostinho sorridente que chegava de mansinho “*Mamãe! você quer um chazinho?*” nas horas de concentração (*Não dava para recusar...*) ou pelo “*Você quer uma ajuda?*” *Ou Apenas um abraço*, nas horas do desânimo ...

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Tema e Contexto	1
1.2 Justificativas	2
1.3 Proposta de trabalho	3
1.4 Objetivos do trabalho	3
1.5 Metodologia	4
1.6 Estrutura do trabalho	4
2. ANÁLISE DE ÍNDICES FINANCEIROS	6
2.1 Introdução	6
2.2 Demonstrações Financeiras	6
2.2.1 Introdução	6
2.2.2 Principais Declarações Financeiras	7
2.3 Análise das Demonstrações Financeiras	13
2.4 Análise de Índices Financeiros	15
2.4.1 Introdução	15
2.4.2 Histórico sobre o uso dos Índices Financeiros	16
2.4.3 Aspectos a Serem Considerados	20
2.4.4 Principais Índices Financeiros	26
2.4.5 Outros Aspectos	28
2.5 Aplicações dos Índices Financeiros	33
2.6 Conclusão	35
3. TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL	37
3.1 Introdução	37
3.2 Técnicas de Inteligência Artificial	38
3.2.1 Introdução	38
3.2.2 Sistemas especialistas	40
3.2.3 Redes Neurais Artificiais	44
3.2.4 Sistemas Difusos	50
3.2.5 Algoritmos Genéticos	61
3.2.6 Raciocínio Baseado em Casos	66
3.3 Sistemas Híbridos	72
3.4 Conclusões	76
4. REDES NEURONAIS ARTIFICIAIS	78
4.1 Introdução	78
4.2 Algumas Características das RNAs	78
4.2.1 Introdução	78
4.2.2 Histórico sobre as pesquisas sobre as RNAs	79
4.2.3 Arquiteturas	84
4.2.4 Aprendizagem	85
4.2.5 Aplicações e Principais Áreas de Atuação	88
4.2.6 Taxinomia das Redes Neurais	90

4.3	Caracterização das Principais Redes Neurais	92
4.3.1	Introdução	92
4.3.2	Rede de Kohonen.....	92
4.3.3	Rede LVQ (Learning Vector Quantization).....	97
4.3.4	Rede ART (<i>Adaptative Resonance Theory</i>).....	100
4.3.5	Rede de Hopfield	105
4.3.6	Perceptron Multicamadas.....	110
4.3.7	Função de Base Radial (Radial Base Function - RBF).....	116
4.4	Um método híbrido: FAN (Free Associative Neurons).....	120
4.4.1	Introdução	120
4.4.2	FAN.....	121
4.5	Conclusões.....	123
5.	<i>APLICAÇÃO: CLASSIFICAÇÃO DA SAÚDE FINANCEIRA DE PEQUENAS EMPRESAS CATARINENSES</i>.....	125
5.1	Introdução.....	125
5.2	Aplicação.....	125
5.3	Diagnóstico através de RN.....	130
5.3.1	MLP com Backpropagation.....	131
5.3.2	RBF.....	133
5.3.3	LVQ.....	135
5.3.4	FAN.....	136
5.4	Comparação dos Resultados.....	137
5.5	Conclusão.....	138
6.	<i>CONCLUSÕES E RECOMENDAÇÕES</i>.....	140
6.1	Conclusões.....	140
6.2	Recomendações.....	142
	<i>REFERENCIAS BIBLIOGRÁFICAS</i>.....	144
	<i>APÊNDICE A: DADOS DE TREINAMENTO E TESTE</i>.....	165
	<i>APÊNDICE B: HISTOGRAMA DO ÍNDICES FINANCEIROS</i>.....	167
	<i>APÊNDICE C: RESULTADOS DA BACKPROPAGATION</i>.....	170
	<i>APÊNDICE D: RESULTADOS DA RBF</i>.....	171
	<i>APÊNDICE E: RESULTADOS DA LVQ</i>.....	181
	<i>APÊNDICE F: RESULTADOS DO FAN</i>.....	185

LISTA DE FIGURAS

Figura 2.1 Estrutura do Balanço Patrimonial.....	8
Figura 2.2: Estrutura da Demonstração do Resultado.....	10
Figura 2.3: Estrutura da Demonstração das Origens e Aplicações de Recursos.....	11
Figura 2.4: Estrutura da Demonstração do Fluxo de Caixa.....	12
Figura 2.5: Classificação de Foster das Técnicas de Análise de Declarações Financeiras [Foster, 1986].....	14
Figura 2.6: Triângulo Du Pont [Matarazzo, 1887].	24
Figura 2.7: Curva da relação quantidade de informação versus quantidade de índices [Matarazzo, 1987].....	28
Figura 3.1: Áreas de Aplicação de Sistema Inteligentes [Goonatilake & Treleaven, 1995a].	39
Figura 3.2 Estrutura genérica de um Sistema Especialista	41
Figura 3.3: Modelo de um neurônio artificial.....	45
Figura 3.4: Perceptron de múltiplas camadas com uma camada escondida e um neurônio de saída.....	45
Figura 3.5: Processo de aprendizagem de uma RNA.....	48
Figura 3.6: Interpretação para barato e caro na Lógica tradicional.....	51
Figura 3.7: Interpretação para barato e caro na Lógica Difusa.	51
Figura 3.8: Derivação de $y=b$ a partir de $x=a$ e $y=f(x)$, onde a e b são pontos e $f(x)$ é uma curva [Jang & Sun, 1995].	55
Figura 3.9: Derivação de $y=b$ a partir de $x=a$ e $y=f(x)$, onde a e b são intervalos e $f(x)$ é uma função intervalar [Jang & Sun, 1995].	55
Figura 3.10: Extensão cilíndrica A.....	55
Figura 3.11: Relação difusa F sobre x e y	55
Figura 3.12: Operação min.....	56
Figura 3.13: Projeção sobre o eixo y	56
Figura 3.14: (A) Crossover entre dois cromossomos. (B) Mutação no alelo 3.....	63
Figura 4.1: Eventos significantes no desenvolvimento de Redes Neurais [Patterson, 1995].	84
Figura 4.2: Uma taxinomia de arquiteturas de redes feed-forward e recorrentes/feedback.....	85
Figura 4.3 : Arquitetura da rede Kohonen, caso de duas dimensões [Pandya & Macy, 1995].	93
Figura 4.4: Topologia quadrada [Pandya & Macy, 1995].	94
Figura 4.5: Feedbaak lateral em uma camada de Kohonen de duas dimensões [Pandya & Macy, 1995].	95
Figura 4.6 Classificação de padrões em dois estágios.....	96
Figura 4.7 Arquitetura da LVQ [Pandya & Macy, 1995].	98
Figura 4.8 Treinamento da LVQ [Pandya & Macy, 1995].	99
Figura 4.9: Estrutura básica da ART1 [Fauset, 1994].	103
Figura 4.10 :Rede de Hopfield Discreta [Fauset, 1994].	107
Figura 4.11 : Atratores de pontos fixos e bacias de atração [Patterson, 1995].	108
Figura 4.12 Apresentação esquemática da correção dos pesos por backpropagation.....	112
Figura 4.13: Estrutura das redes MLP e RBF [Pandya & Macy, 1995].	119
Figura 4.14: Arquitetura do FAN.....	122
Figura 5.1 : (a) Convergência da Backpropagation usando 8 neurônios, taxa da aprendizagem 0.08, momentum 0.8 e 1000 épocas. Com esta arquitetura obtemos 81.25% de acerto nos dados de teste; (b) Convergência da Backpropagation usando 8 neurônios, taxa da aprendizagem 0.08, momentum 0.8 e 2000 épocas. Com esta arquitetura obtemos 68.75% de acerto nos dados de teste.....	133
Figura 5.2: (a) Convergência da rede RBF com 15 centros, 1.3 de raio e 14 épocas. Com esta arquitetura rede acertou 77.5% no treinamento e 81.25% teste; (b) Convergência da rede RBF com 15 centros, 1.3 de raio e 34 épocas. Com esta arquitetura rede acertou 100% no treinamento e 37.50% teste.....	135
Figura 5.3: (a) Convergência da rede LVQ, trabalhando com os dados normalizados no intervalo $[-1,1]$ com taxa de aprendizagem 0.2 decrescendo não linearmente e 45 épocas. Com esta arquitetura rede acertou 72.5% no treinamento e 62.25% no teste; (b) Convergência da rede LVQ, trabalhando com os dados normalizados no intervalo $[-1,1]$ com taxa de aprendizagem 0.2 decrescendo não	

<i>linearmente e 200 épocas. Com esta arquitetura rede acertou 70.00% no treinamento e 62.25% no teste.</i>	<i>136</i>
<i>Figura 6.1 : Arquitetura híbrida para análise da situação financeira.</i>	<i>143</i>

LISTA DE TABELAS

<i>Tabela 2.1 Classificação dos Índices Financeiros segundo alguns pesquisadores.</i>	26
<i>Tabela 3.1: Funções de pertinências mais comuns.</i>	52
<i>Tabela 3.1 : Comparação das Técnicas Inteligentes [Goonatilake e Treleaven, 1995^a -modificado].</i>	72
<i>Tabela 3.2: Modelos para integrar sistemas inteligentes [Medsker & Bailey, 1992].</i>	74
<i>Tabela 3.3: Três classes híbridas proposta [Goonatilake & Khebbal, 1995b].</i>	75
<i>Tabela 4.1: Algoritmos de Aprendizagem mais conhecidos.</i>	91
<i>Tabela 5.1: Importância de cada índice financeiro.</i>	129
<i>Tabela 5.2 : Índices Financeiros de duas empresas e a categoria o qual eles pertencem.</i>	129
<i>Tabela 5.1 : Funções principais para a rede backpropagation do MATLAB.</i>	131
<i>Tabela 5.2 : Resultados da MLP com backpropagation usando os dados normalizados no intervalo [-1,1].</i>	132
<i>Tabela 5.1 : Funções principais para a rede RBF do MATLAB.</i>	134
<i>Tabela 5.2 . Resultados da RBF trabalhado com os dados normalizados no intervalo [-1,1].</i>	135
<i>Tabela 5.1 : Resultado da LVQ com dados normalizados no intervalo [1,1].</i>	136
<i>Tabela 5.1 : Resultado do FAN (Porcentagem de acerto).</i>	137
<i>Tabela 5.1 : Comparação dos resultados.</i>	138
<i>Tabela 6.1 : Percentual de acerto de cada modelo.</i>	142
<i>Tabela A-1 : Dados de treinamento.</i>	166
<i>Tabela A-2 : Dados de Teste.</i>	167
<i>Tabela C-1: Resultado da Backpropagation usando função de ativação sigmoide na camada escondida e linear na última camada.</i>	172
<i>Tabela C-2: Resultado da Backpropagation usando função de ativação tangente hiperbólica na camada escondida e linear na última camada.</i>	173
<i>Tabela C-3 : Resultado da Backpropagation usando função de ativação sigmoideal na camada escondida e linear na última camada.</i>	174
<i>Tabela C-4 : Resultado da Backpropagation usando função de ativação não linear (sigmoideal e tangente hiperbólica) na camada escondida e na camada de saída</i>	174
<i>Tabela D-1 : Resultados do desempenho da rede RBF usando os dados normalizados no intervalo [-1,1].</i>	176
<i>Tabela D-2 : Resultados do desempenho da rede RBF usando os dados normalizados no intervalo [0,1].</i>	180
<i>Tabela E-1 : Resultados da rede LVQ usando os dados não normalizados e a taxa de aprendizagem decresce linearmente em função do número de épocas.</i>	181
<i>Tabela E-2: Resultados da rede LVQ usando os dados não normalizados e a taxa de aprendizagem decresce não linearmente em função do número de épocas.</i>	182
<i>Tabela E-3 : Resultados da rede LVQ usando os dados normalizados entre [-1,1] e a taxa de aprendizagem decresce linearmente em função do número de épocas.</i>	183
<i>Tabela E-4 : Resultados da rede LVQ usando os dados normalizados entre [-1,1] e a taxa de aprendizagem decresce não linearmente em função do número de épocas.</i>	184
<i>Tabela F-1: Resultados da FAN com H=1 (grau de combinação das variáveis).</i>	185

LISTA DE QUADROS

Quadro 2.1: Quadro Resumo dos Índices Financeiros..... 31

RESUMO

Tendo em vista a importância das pequenas empresas na economia nacional, esta dissertação tem como principal objetivo o auxílio das pequenas empresas na determinação de sua saúde financeira (diagnóstico) através de modelos computacionais que usam técnicas de Inteligência Artificial. Estes modelos visam fornecer um “especialista” às pequenas empresas com menor custo.

Faz-se, inicialmente, uma revisão sobre análise de índices financeiros, uma das principais ferramentas para diagnóstico de empresas. Em seguida, revisa-se algumas técnicas de Inteligência Artificial que são usadas frequentemente na solução de problemas da área financeira.

Objetivando um diagnóstico mais apurado, analisou-se diferentes arquiteturas de redes neurais (Backpropagation, RBF e LVQ) e uma rede híbrida (FAN) considerando alguns requisitos, aprendizagem dos dados, poder de generalização, rapidez, necessários para uma melhor classificação (diagnóstico).

Os dados trabalhados foram índices calculados de demonstrativos financeiros (Balanço Patrimonial e Demonstrativos de Resultados) de 56 Micro Empresas Catarinenses do setor de comércio.

Os modelos usados apresentaram um bom desempenho apesar da pequena quantidade de dados, mostrando o poder de aprendizado e classificação (diagnóstico) que as redes neurais possuem.

Deve-se enfatizar que a utilização correta e a difusão do uso de tal diagnóstico poderá auxiliar a pequena empresa no seu gerenciamento, representando uma economia de recursos para o país.

ABSTRACT

The motivation of this work was based on the necessity of developing intelligent systems to give support to the Brazilian small and medium firms; these firms are essential to the Brazilian economy.

The main objective of this work is to aid small and medium firms to determine their financial health through the use of computational models. The referred models use artificial intelligence techniques; they have the potential to provide a virtual "specialist" to the firms at a lower cost than traditional financial health diagnosis techniques.

A review of financial ratio analysis is presented at the beginning of the work. It follows a revision of Artificial Intelligence techniques that are frequently used in problem solving in the financial area.

In order to reach a precise diagnosis of firms' financial health, different neural networks architectures (Backpropagation, RBF and LVQ) and a hybrid net (FAN) were analyzed. The experiments have taken into account specific requirements of the real world data; they are: learning capacity, generalization power and speed.

The data used consisted in financial ratios that were evaluated from the main financial reports (e.g., Balance Sheet and Income Statement), corresponding to fifty six (56) small firms from the retail sector of the Santa Catarina State.

The models presented a satisfactory performance in spite of the small amount of data; this fact demonstrated the learning and classification power of artificial neural networks models.

Finally, it is worthy to emphasize that the use and diffusion of such diagnosis could boost the financial management of the small and medium Brazilian firms, representing an important economy to the country.

1. INTRODUÇÃO

1.1 Tema e Contexto

A habilidade de uma empresa em usar as informações e conhecimentos especializados tem se tornado um aspecto vital que pode determinar o sucesso ou a falência de uma organização. Isto é particularmente verdade para as chamadas “pequenas empresas” devido a seu tamanho e expansão crescente, principalmente no Brasil.

Neste contexto, muito dos pequenos negócios tem passado por dificuldades em seu gerenciamento devido a falta de conhecimento especialista em algumas das principais áreas de gerenciamento. De acordo com os dados do SEBRAE¹ [Souza, 1995], as micro empresas representam cerca de 98,3% de todas as empresas registradas, e as suas atividades são responsável por 20,40% do Produto Interno Bruto (PIB) e por 54,40% da mão de obra ocupada no país. Em particular na região de Santa Catarina, uma pesquisa feita pelo Departamento de Engenharia de Produção [Batalha & Demori, 1990], relatou que o gerenciamento financeiro é considerado uma das áreas mais problemáticas para as pequenas indústrias. A consequência direta de tais dificuldades é transformado em problemas de **liquidez**, baixo capital de **giro**, alto risco devido à posição dos **débitos** e crescimento **desestruturado** entre outros casos. Estes aspectos posteriormente afetam os custos e rendas operacionais da empresa, e causam, dependendo do tempo de uma correção, a falência da empresa.

A performance de uma firma depende do equilíbrio entre liquidez e lucratividade. O monitoramento de problemas financeiros deve ser uma tarefa contínua. Mesmo uma empresa muito rentável pode ocultar fatores que determinem

¹ SEBRAE: Serviço de Apoio a Micro e Pequenas Empresas.

sua falência. A tarefa de monitoramento identifica fatores que podem ser ajustados para evitar a falência.

Uma técnica central para endereçar situações financeiras problemáticas é a Análise de Declarações Financeiras, um processo no qual os especialistas reorganizam as informações da firma e outras origens, criam variáveis auxiliares (como os índices financeiros), e fazem uma comparação com padrões a fim de identificar e entender os desvios. Com o objetivo de encontrar uma conclusão sem perdas em tal processo complexo, o especialista procura por sintomas financeiros baseados em sua experiência. Este processo é essencialmente desestruturado. Em uma análise mais detalhada, o especialista focaliza um aspecto particular a fim de verificar suas causas principais, entender os desvios e finalmente conseguir uma conclusão. Este processo é muito mais estruturado que o primeiro, pois neste ponto o especialista tem mais sintomas específicos para analisar.

Existem também na literatura técnicas de Inteligência Artificial (IA) dedicadas à análise da saúde financeira: previsão de falência [Wilson & Hoff, 1994; Odom & Shardo, 1990; Lacher et al, 1991], análise de crédito [Klymasauskas, 1991; Barker, 1990], auditoria [Blocher, 1990; [Mui et al., 1990], diagnóstico e solução [Martins, 1996; Pacheco, 1996].

1.2 Justificativas

A Análise de Demonstrativos Financeiros tradicional inclui várias técnicas tais como *cross sectional* (análise de índices financeiros e declarações de tamanho comum), series de tempo e a combinação de informações de declarações financeiras com outros tipos de dados. O processo geral de extração de informações importantes para suportar uma decisão não é simples. Ele depende do setor econômico, do tamanho da empresa, do padrão de comparação e além disto é dinâmico. Portanto, a Análise de Declarações Financeiras é um processo que requer um especialista com grande experiência em análise financeira.

Usualmente uma empresa tem duas alternativas quanto às atividades de monitoramento e resolução de problemas: a própria empresa fazer ou contratar um consultor financeiro. Enquanto a primeira escolha requer recursos humanos, a

segunda só pode ser uma solução se a empresa dispor de recursos financeiros para a contratação de um especialista. Geralmente as pequenas empresas não podem contar com nenhuma destas soluções, pois não tem nenhum especialista que trabalhe na empresa e também não dispõem de recursos financeiros para a contratação do mesmo. Além disto, quando uma destas alternativas é possível, a solução pode chegar após um processo de interpretação muito lento e portanto não ser mais útil.

O monitoramento da saúde financeira das empresas é um fator crítico para seu sucesso. Uma das soluções para este problema é o desenvolvimento de sistemas inteligentes que não somente analisam os problemas mas também sugerem soluções. A agregação de ambas as tarefas envolve raciocínio dedutivo e indutivo, o que justifica mais de uma técnica de Inteligência Artificial (IA).

Sistemas baseados em técnicas de IA podem tornar o conhecimento de um especialista financeiro acessível a um grande número de empresas. Se corretamente implementado, tais sistemas podem salvar milhões de dólares para um país devido o uso mais eficiente de todos os instrumentos financeiros e decisões internas.

1.3 Proposta de trabalho

Neste trabalho pretende-se fazer uma análise dos diferentes modelos de redes existentes na literatura, buscando os que mais se adaptam ao problema de diagnóstico de empresas. Feita esta análise, usar-se-á as redes selecionadas para fazer o diagnóstico das pequenas empresas catarinenses.

1.4 Objetivos do trabalho

O objetivo geral deste trabalho é auxiliar as pequenas empresas na determinação da saúde financeira, possibilitando acesso aos especialistas com menor custo.

Os objetivos específicos são:

1. Apreciar a capacidade do sistema inteligente como elemento de avaliação de saúde financeira das empresas;

2. Analisar o uso de diferentes algoritmos de Redes Neurais e o sistema híbrido FAN (*Free Associative Neuron*) quando aplicado em um domínio.

A fim de satisfazer os objetivos específicos, os sistemas implementados neste trabalho seguirão os seguintes requisitos:

- trabalhar com dados de empresas comerciais;
- usar o menor volume de dados mínimo em relação ao algoritmo;
- facilidade de adaptar aos novos dados;
- flexibilidade a fim de ser adaptado para diferentes setores econômicos;
- apresentar algumas ferramentas que auxiliam o gerenciamento financeiro das pequenas empresas;

1.5 Metodologia

A primeira parte do trabalho, consiste em um levantamento bibliográfico (através de livros, artigos, internet), acerca de conceitos, desenvolvimentos, e aplicações sobre Análise de Declarações Financeiras e sobre algumas técnicas de IA.

A segunda parte utiliza dados (Demonstrativos Financeiros) de pequenas empresas catarinenses. Analisa-se estes dados e determina-se quais os índices financeiros que serão utilizados como entrada para os sistemas inteligentes.

A terceira parte do trabalho refere-se à implementação dos Sistemas Inteligentes. Obtém-se nesta fase o diagnóstico do(s) problema(s) financeiros das empresas analisadas.

Finalmente na última etapa do trabalho tem-se as conclusões e possíveis sugestões para outros trabalhos.

1.6 Estrutura do trabalho

A fim de melhor apresentar este trabalho, o mesmo divide-se em seis capítulos, como seguem:

No Capítulo 2 - **“Análise de Índices Financeiros”**- discute-se o uso das informações das declarações financeiras, com ênfase aos índices financeiros, como indicadores financeiros para um diagnóstico. Também apresenta-se neste capítulo os conceitos das categorias de índices de liquidez, lucratividade, capital de giro,

No Capítulo 3 - **“Técnicas de Inteligência Artificial”** - descreve-se os fundamentos de algumas técnicas de IA (Sistema Especialistas, Conjuntos Difusos, Redes Neurais Artificiais, Algoritmos Genéticos e Raciocínio Baseado em Casos). Também faz-se uma análise das principais vantagens e desvantagens destas técnicas, e discute-se a necessidade de combinação das referidas técnicas (hibridização) a fim de se alcançar um melhor desempenho.

No Capítulo 4 - **“Redes Neurais Artificiais”** - analisa-se diversos modelos de redes neurais artificiais, inclusive a rede híbrida FAN, buscando identificar as mais adequadas para o problema de diagnóstico.

No Capítulo 5 - **“Aplicação: Saúde Financeira de Pequenas Empresas”** - discute-se a definição e relevância do problema na área de finanças. Explica-se a base da escolha das variáveis financeiras dos dados reais envolvidos no problema. A discussão é baseada nas descobertas de Martins sobre as ferramentas mais adequadas em Finanças para diagnóstico e indicação de soluções para os problemas financeiros [Martins, 1996]. Após a análise do problema sobre o ponto de vista financeiro, discute-se os resultados obtidos pelos sistemas inteligentes implementados.

Finalmente, o Capítulo 6 refere-se as conclusões e recomendações do trabalho.

2. ANÁLISE DE ÍNDICES FINANCEIROS

2.1 Introdução

Uma empresa é uma inter-relação de decisões, fluxos físicos e informações. Existe uma multiplicidade de fatores que determinam a lucratividade de um negócio e suas chances de sobrevivência. Um destes fatores de interesse particular da empresa (do gerente, especialmente) é a eficácia da tomada de decisão. As informações das demonstrações financeiras (Balanço Patrimonial, Demonstrativo de Resultados, Demonstração de Fluxo de Caixa, e Demonstrativo das Origens e Aplicações de Recursos) da empresa servem como um sinal de que decisões são necessárias ou fornecem informações que podem ser usadas na decisão. Uma das técnicas principais aceitas como ferramenta da Análise de Declarações Financeira é a análise de índices financeiros. A análise destes índices, além de auxiliar no gerenciamento da empresa, auxilia na previsão estatística de falência, análise da saúde financeira, e planejamento de estratégias.

Neste capítulo apresentar-se-á inicialmente um resumo sobre as demonstrações financeiras mais usadas na Análise de Demonstrações Financeiras. Devido a sua grande utilidade, falar-se-á em seguida sobre a Análise de Índices Financeiros: histórico, aspectos teóricos, principais índices e algumas aplicações.

2.2 Demonstrações Financeiras

2.2.1 Introdução

Um dos elementos mais importantes na tomada de decisões relacionadas a uma empresa é a análise das suas demonstrações financeiras. As **Demonstrações**

Financeiras relatam o que acontece na empresa em termos de vendas de bens, despesas, lucros, dividendos, etc. Por exemplo, nas empresas de Sociedades Anônimas, esta é uma das informações que os investidores fornecem e a comunidade dos investidores usam para formar expectativas sobre como a magnitude, risco de retorno da empresa, o preço de mercado das ações serão afetados. O entendimento das demonstrações financeiras são, portanto, importantes para investidores e para o gerente da empresa.

As principais declarações financeiras baseadas na contabilidade são o Balanço Patrimonial, Demonstração do Resultado do Exercício, Demonstração de Origens e Aplicações de Recursos, e Demonstração de Fluxo de Caixa.

2.2.2 Principais Declarações Financeiras

As demonstrações financeiras fornecem uma série de dados sobre a empresa, de acordo com regras contábeis. A **Análise de Demonstrações Financeiras** transforma esses dados em informações e será tanto mais eficiente quanto melhores informações produzir. Estas informações auxiliam o tomador de decisão na avaliação da posição financeira passada ou corrente da empresa assim como na previsão do futuro da empresa. O escopo da análise financeira depende de seu objetivo, variando de uma análise total dos pontos fortes e fracos da empresa a uma análise mais simplificada de sua liquidez a curto prazo.

O **Balanço Patrimonial** (Figura 2.1) é a demonstração que apresenta todos os bens e direitos da empresa - **Ativo** - assim como as obrigações - **Passivo Exigível** - em determinada data e o **Patrimônio Líquido**. No ativo relacionam-se todas as aplicações de recursos efetuadas pela empresa. Estes recursos poderão estar distribuídos em ativos circulantes, assim denominado por apresentarem uma maior rotação em relação aos chamados “ativos fixos”, como: valores em caixa, valores a receber a curto prazo, etc.; em ativos realizáveis a longo prazo; e em ativos classificados como permanentes, como: prédios, terrenos,... os quais servem a vários ciclos operacionais. Por outro lado, o Passivo Exigível identifica a origem de todos os recursos de terceiros levantados pela empresa. A diferença entre o Ativo e o Passivo Exigível é chamado Patrimônio Líquido e representa o capital investido pelos

proprietários da empresa, quer através de recursos trazidos de fora da empresa, quer gerados por esta em suas operações internamente.

BALANÇO PATRIMONIAL	
ATIVO	PASSIVO E PATRIMÔNIO LÍQUIDO
CIRCULANTE	CIRCULANTE
Disponibilidades	Fornecedores.....XXXX
Caixa e Bancos.....XXXX	Imposto a Pagar.....XXXX
Aplic. de liquidez Imediata.....XXXX	Dividendos a Pagar.....XXXX
Cientes	Contas a Pagar.....XXXX
Duplicatas a Receber	TOTAL DO PASSIVO CIRCULANTEXXXX
Provisão p/ Devedores Duvidosos.....XXXX	EXIGÍVEL A LONGO PRAZO
Duplicatas Descontadas.....XXXX	Empréstimos Bancários.....XXXX
Estoques	TOTAL EXIGÍVEL A LONGO PRAZO.....XXXX
Produtos Acabados.....XXXX	PATRIMÔNIO LÍQUIDO
Produtos em Processo.....XXXX	Capital Social
Matéria-Prima.....XXXX	Subscrito.....XXXX
TOTAL DA ATIVO CIRCULANTE.....XXXX	A realizar.....XXXX
PERMANENTE	Reservas de Capital
Investimentos	Correção Monetária do Capital.....XXXX
Ações de outras empresas.....XXXX	Reservas de Lucro
Incentivos Fiscais.....XXXX	Reserva Legal.....XXXX
Imobilizado	Lucros Acumulados.....XXXX
Imóveis.....XXXX	TOTAL DO PATRIMÔNIO LÍQUIDO.....XXXX
Veículos.....XXXX	TOTAL DO PASSIVO.....XXXX
Máquinas e Equipamentos.....XXXX	
Depreciação Acumulada.....XXXX	
Diferido	
Despesas Pré-Operacionais.....XXXX	
TOTAL DO ATIVO PERMANENTE.....XXXX	
TOTAL DO ATIVO.....XXXX	

Figura 2.1 Estrutura do Balanço Patrimonial.

Porém o balanço apresenta a posição de uma empresa em um dado momento. A informação que ele fornece é totalmente estática, e muito provavelmente, a sua estrutura se apresentará relativamente diferente algum tempo após o seu encerramento. O Balanço fornece uma visão da empresa que exige do analista ou gerente muita experiência para poder entender a situação financeira da empresa.

Entretanto, o Balanço servirá como elemento de partida para o conhecimento retrospectivo da situação econômica e financeira de uma empresa, através das informações contidas nos seus vários grupos de contas. E, em função do comportamento verificado nos últimos exercícios, poder-se-á identificar determinada posição futura da empresa.

A **Demonstração do Resultado do Exercício** (Figura 2.2) é uma demonstração dos aumentos e reduções causados no Patrimônio Líquido pelas operações da empresa. As receitas representam normalmente aumento no Passivo, através do ingresso de novos elementos, como duplicata a receber ou dinheiro proveniente de transações. Aumentando o Ativo, aumenta o Patrimônio Líquido. As despesas representam redução do Patrimônio Líquido, através de dois caminhos possíveis: redução do Ativo e Aumento do Passivo exigível.

A Demonstração do Resultado registra então o fluxo de receitas e relata as despesas de uma empresa entre duas datas, geralmente um ano. Por isto é classicamente chamada de Demonstração do Resultado de Fluxo de Renda [Matarazzo, 1987].

A Demonstração do Resultado retrata somente o fluxo econômico e não o fluxo monetário. Para a Demonstração do Resultado não importa se uma receita ou despesa tem reflexos em dinheiro, basta apenas que afete o Patrimônio Líquido. Como as modificações no Patrimônio Líquido produzidas por receitas e despesas afetam a riqueza dos proprietários, elas são retratadas na Demonstração do Resultado que é uma peça de caráter eminentemente econômico (relacionada à riqueza) e não financeiro (relacionado à dinheiro).

DEMONSTRAÇÃO DO RESULTADO	
RECEITA OPERACIONAL BRUTA	
	Vendas Brutas
(-)	Deduções e descontos concedidos
(-)	Devoluções
(-)	Impostos Pagos
RECEITA LÍQUIDA OPERACIONAL:	
(-)	Custo de produção das mercadorias vendidas ou Custo das mercadorias vendidas
LUCRO BRUTO:	
(-)	Despesas Operacionais
	Despesas de Administração
	Despesas de Vendas
	Despesas Gerais
LUCRO/PREJUÍZO OPERACIONAL:	
(±)	Receitas e Despesas Financeiras
LUCRO/PREJUÍZO OPERACIONAL LÍQUIDO:	
(+)	Receitas não operacionais
(-)	Despesas não operacionais
(±)	Saldo da conta de correção monetária
LUCRO/PREJUÍZO LÍQUIDO ANTES DO IR:	
(-)	Provisão para Imposta de Renda
LUCRO/PREJUÍZO LÍQUIDO:	
(-)	Participações
(-)	Contribuições
LUCRO (OU PREJUÍZO) LÍQUIDO DO EXERCÍCIO	

Figura 2.2: Estrutura da Demonstração do Resultado.

A **Demonstração das Origens e Aplicações de Recursos (DOAR)** (Figura 2.3) acrescenta um enorme conjunto de informações sobre a dinâmica financeira da empresa. A DOAR é uma comparação de dois Balanços consecutivos, os quais identificam as variações ocorridas na estrutura financeira da empresa durante o período considerado, permitindo, por conseguinte, melhores critérios para a análise financeira de empresa.

Dado que o Balanço é uma informação totalmente estática, a DOAR apresenta-se como um dos instrumentos mais importantes para que se conheçam as mutações verificadas na posição financeira da empresa, evidenciando-se, de uma maneira abrangente, os financiamentos (origens de recursos) e os investimentos (aplicações de recursos) verificados no período [Neto, 1983].

DEMONSTRAÇÃO DAS ORIGENS E APLICAÇÕES DE RECURSOS	
ORIGENS DE RECURSOS	
• Das Operações:	
■ Lucro Líquido do ano	
■ Mais: Depreciações e Amortizações	
Resultado da Correção Monetária do Exercício (saldo devedor)	
Variações Monetárias de Empréstimos e Financiamentos a Longo Prazo	
■ Menos: Participação nos Resultados de Controladas e Coligadas	
Resultado da Correção Monetária (saldo credor)	
Resultado na Alienação de Bens do Imobilizado	
Total das Operações	
• Dos Acionista:	
■ Integralização de Capital	
• De Terceiros:	
■ Ingresso de Empréstimos a Longo Prazo	
■ Alienação de Itens do Imobilizado (valor de venda)	
■ Resgate de Investimentos Temporários a Longo Prazo	
Total das Origens	
APLICAÇÕES DE RECURSOS	
• Aquisições de direitos do Imobilizado (ao custo)	
• Adições ao Custo no Ativo Diferido	
• Aplicações em Investimentos Permanentes em outras Sociedades	
• Aplicações em investimentos Temporários a Longo Prazo	
• Transferência para Curto prazo de Empréstimos e financiamentos a Longo Prazo	
• Dividendos Propostos	
Total das Aplicações	
ACRÉSCIMO (DECRÉSCIMO) NO CAPITAL CIRCULANTE LÍQUIDO	

Figura 2.3: Estrutura da Demonstração das Origens e Aplicações de Recursos.

A **Demonstração do Fluxo de Caixa** (Figura 2.4), apesar de ser uma das demonstrações mais úteis, não é divulgada pelas empresas. Desta forma, o analista externo precisa elaborá-la baseado nas demais demonstrações, como o DOAR.

DEMONSTRAÇÃO DO FLUXO DE CAIXA	
Exercício de ----/----/----	
LUCRO LÍQUIDO DO EXERCÍCIO	
+	Despesas não desembolsáveis
	Depreciação
	Saldo Da Correção Monetária

-	Receitas não Embolsáveis
	Resultado da Equivalência Patrimonial

GERAÇÃO BRUTA DE CAIXA	
+	Acréscimo de Fontes Operacionais
	Fornecedores
	Obrigações de Funcionamento
-	Acréscimo de Aplicações Operacionais
	Duplicatas a Receber
	Estoques
	Outras Contas
=	GERAÇÃO OPERACIONAL DE CAIXA
+	Fontes não Operacionais
	Realização de Capital
	Venda de Imobilizado
	Venda de Investimentos
	Aumento de Exigível Longo Prazo
	Diminuição de realizável Longo Prazo
	Acréscimo de Empréstimos de Curto Prazo
-	Aplicações não Operacionais
	Dividendos
	Aquisição de Investimentos
	Aquisição de Imobilizado
	Aumento de Ativo Diferido
	Aumento de Realizável Longo Prazo
	Diminuição de Exigível Longo Prazo
	Diminuição de Empréstimos Curto Prazo
=	GERAÇÃO LÍQUIDA DE CAIXA
+	Saldo Inicial de caixa
=	Saldo Final de caixa

Figura 2.4: Estrutura da Demonstração do Fluxo de Caixa.

Muitas vezes os problemas de insolvência ou liquidez ocorrem por falta de uma administração adequada do fluxo de caixa, daí a importância de sua análise.

As principais vantagens da Demonstração de Fluxo de Caixa são: - avaliar alternativas de investimentos; - avaliar e controlar ao longo prazo as decisões importantes que são tomadas na empresa, com reflexos monetários; - avaliar a situação presente e futura do caixa da empresa, posicionando-a para que não chegue a

situação de liquidez; - certificar que os excessos momentâneos de caixa estão sendo devidamente aplicados [Matarazzo, 1987].

As quatro Demonstrações financeiras citadas acima representam informações diferentes e complementares para o analista financeiro. Além das informações, o analista financeiro tem que identificar a melhor abordagem para o problema da análise da saúde financeira da empresa. Foster [Foster, 1987] classifica os diferentes esquemas de acordo com o objetivo da análise. Uma visão da classificação de Foster é mostrada na Figura 2.5.

2.3 Análise das Demonstrações Financeiras

A Análise Financeira de empresas é tarefa bastante complexa e de fundamental importância numa sociedade moderna [Silva, 1988]. Como o próprio nome sugere, a Análise Financeira é a seleção, avaliação e interpretação de dados financeiros e outros dados pertinentes. O analista financeiro deve determinar quais informações são necessárias e como usá-las. Da mesma forma, um investigador que pretende adquirir ações de determinada companhia também deve conhecer a referida organização.

O primeiro objetivo da Análise das Declarações Financeiras é a análise *Cross-Sectional*, que é, o estudo de empresas em um setor econômico específico. As declarações de tamanho comum e análise de índices financeiros são duas técnicas para comparar empresas. Uma das formas mais simples e direta de analisar mudanças no tempo é calcular declarações de tamanho-comum. O Balanço de tamanho-comum tenta reduzir o efeito do tamanho da empresa quando se compara as declarações financeiras das empresas. A estratégia é dividir todas as contas da declaração por uma destas contas para colocá-las em uma base de porcentagem comum. Por exemplo, um demonstrativo de resultado de tamanho-comum é construído dividindo-se os vários componentes pela renda líquida.

O método mais usado e muitas vezes confundido com a análise de declarações financeiras é a análise de índices financeiros. Estes índices comparam variáveis financeiras que são extraídas, geralmente, do Balanço Patrimonial e do demonstrativo de resultado. Esta técnica reduz a quantidade de informações e enfatiza as relações

entre elementos financeiros ao invés de seus valores individuais. Na próxima seção faz-se um estudo mais minucioso sobre a análise dos índices financeiros.

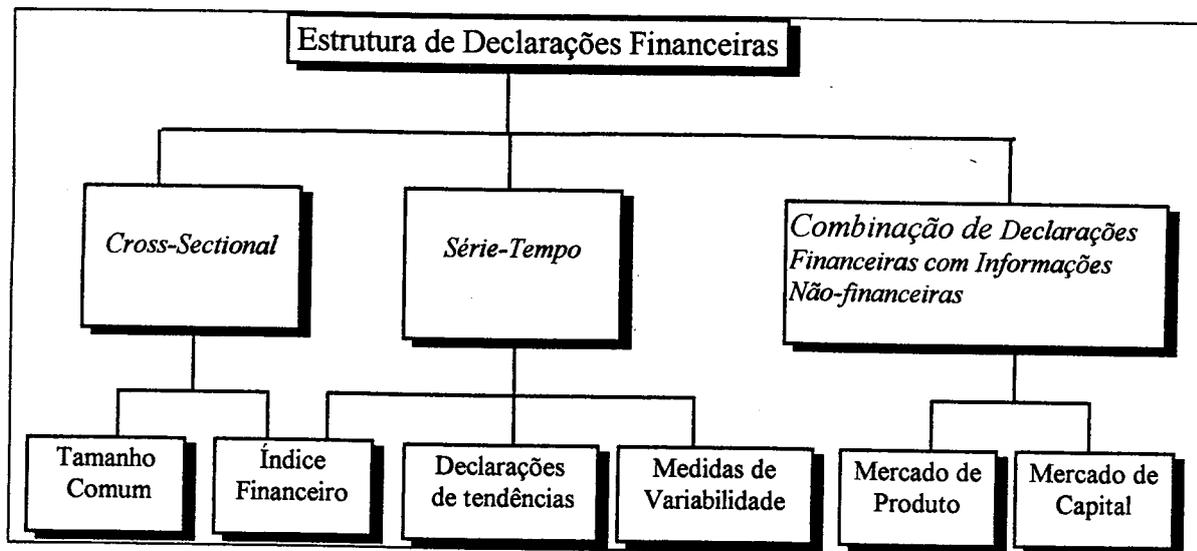


Figura 2.5: Classificação de Foster das Técnicas de Análise de Declarações Financeiras [Foster, 1986].

Um outro objetivo da análise de declarações financeiras é a análise de Série-Tempo. Esta análise consiste do estudo da performance da empresa no tempo afim de prever sua saúde financeira, baseado em informações passadas e presentes. A Série-Tempo usa informações sobre a empresa para prever seu desenvolvimento futuro. A análise de Série-tempo é acompanhada também pela análise de índices e declarações das tendências. As declarações das tendências são elaboradas fixando-se um período base e expressando-se os elementos financeiros dos períodos subsequentes por seus valores relativos no período base. A análise da Série-Tempo pode também ser identificada pelo estudo de índices financeiros no tempo. A terceira abordagem para série-tempo é a análise da variabilidade onde os índices e as outras variáveis são medidos no tempo. Contudo, nesta abordagem o objetivo é definir as relações entre os valores extremos (máximo e mínimo) e a média sobre o período [Foster, 1986].

O terceiro objetivo da Análise Financeira é estudar a performance afim de estabelecer as estratégias de investimentos no mercado de capitais. Isto envolve a combinação de declarações financeiras e informações não financeiras como mercado de produtos e mercado de capitais.

Particularmente para o problema considerado aqui, a análise de declarações tem uma utilidade dupla:

- 1) possível diagnóstico de problemas financeiros de pequenas empresas baseado em seus índices financeiros (e tendência) e no valor médio correspondente do setor econômico; e
- 2) suportar a estratégia dedutiva para verificar as causas e oferecer soluções para os problemas financeiros.

Portanto, de acordo com a Figura 2.5, a Análise de Declarações Financeiras é usada neste trabalho como uma técnica de análise cross-sectional e de série-tempo baseada em índices financeiros. Então, na próxima seção direcionar-se-á um estudo sobre a análise de índices financeiros.

2.4 Análise de Índices Financeiros

2.4.1 Introdução

Índice financeiro é a relação (x/y) entre contas ou grupos de contas das Demonstrações Financeiras, que visa evidenciar determinado aspecto da situação econômica ou financeira de uma empresa. Se ambos x e y vêm do Balanço Patrimonial o índice pode ser chamado estático e se pelo menos um vem do Demonstrativo de Resultado pode ser chamado de dinâmico [Salmi & Martikainen, 1994]. A Análise de Índices Financeiros é usada com vários objetivos por profissionais como investidores, gerentes, financiadores assim como por acadêmicos. Na prática os índices são usados, por exemplo, para prever o sucesso da empresa, enquanto que o principal objetivo dos pesquisadores é desenvolver modelos que exploram estes índices.

Qualquer sistema de análise financeira que é baseado no uso de índices financeiros deve levar em consideração muitos outros indicadores, tais como situação econômica, política da empresa, setor do negócio e princípios contábeis que estão afetando a empresa. Este conjunto de requerimentos determina a validação dos dados de entrada e a confiança na análise.

2.4.2 Histórico sobre o uso dos Índices Financeiros

A adoção de índices como uma ferramenta de análise financeira é relativamente recente. A Análise de Declarações financeiras surgiu e desenvolveu-se dentro do sistema bancário que é até hoje seu principal usuário.

As principais causas para o desenvolvimento de análise de índices financeiros foram a maturidade industrial na metade do século XIX e principalmente a análise de crédito. Ambos ocorreram nos USA. Porém, a abordagem de análise de crédito dominou o desenvolvimento dos índices financeiros. Seu início remonta ao final do século passado, quando os banqueiros americanos passaram a solicitar Balanços às empresas tomadoras de empréstimos. Durante o ano de 1890 o volume e fluxo de informações financeiras crescia muito e começou-se a análise de itens correntes e não correntes. Então, o uso de índices financeiros, na análise de declarações financeiras, pode ser dito ter iniciado com o advento dos índices correntes. A medida ganhou aceitação ampla quando, em 9 de fevereiro de 1895, o Conselho Executivo da Associação dos Bancos do estado de New York resolveu recomendar ao seus membros que solicitassem aos tomadores de empréstimos declarações escritas e assinadas de seus ativos e passivos. Em 1900, esta mesma associação divulgou um formulário de proposta de crédito que incluía o espaço para o Balanço. É bem provável que, nesta época, os Balanços apresentassem dados que eram examinados apenas superficialmente, sem nenhuma técnica analítica ou tentativa de medição quantitativa.

A literatura contábil do começo do século atual menciona a importância de comparações de dados das demonstrações financeiras, porém as idéias eram, via de regra, vagas em relação ao que comparar. Com o decorrer dos anos e o sucessivo recebimento de balanços foi-se desenvolvendo a noção de comparação de diversos itens, sendo o mais comum a do Ativo Circulante com o Passivo Circulante.

No período anterior e durante a I Guerra Mundial (1900 - 1919) ocorreram três grandes desenvolvimentos: - foram definidos um grande número de índices; - critério de índice absoluto; - alguns analistas começaram a reconhecer a necessidade de análise entre empresas, e conseqüentemente a necessidade de índices relativos. Porém, ainda poucos analistas usavam índices, e aqueles que usavam, utilizavam apenas um índice, o índice corrente.

Por volta de 1913, chamava-se a atenção para outros índices como: depósitos bancários em relação ao exigível, percentual de contas a receber em relação aos demais itens do ativo, percentual de estoques em relação a vendas anuais.

Entretanto, as demonstrações financeiras na época não eram preparadas adequadamente para os fins a que se destinavam. Não havia uniformidade nas disposições e na terminologia, nem na classificação das rubricas.

Importante contribuição para a melhoria da forma de apresentação das demonstrações financeiras foi feita pelo Federal Reserve Board, em 1918, quando padronizou a forma dos Balanços e das Demonstrações de Lucros e Perdas.

Em 1919, Alexander Wall, considerado o pai da Análise de Balanços, apresentou um modelo de Análise de Balanço através de índices. Ele também mostrou e tornou popular a necessidade de considerar outras relações (índices) além de Ativo circulante contra passivo circulante.

Paralelamente, mas na área de gerenciamento, o uso de margem de lucro e turnovers, já estava bem desenvolvida.

Alexander Wall desenvolveu, posteriormente, em parceria com outros autores, fórmulas matemáticas de avaliação de empresas, ponderando diversos índices de Balanço.

Tendo adotado o método de computar vários coeficientes, os autores sentiam a necessidade de padrões de referências que os auxiliassem em suas avaliações.

Em mais ou menos 1919, a companhia Du Ponte começou a usar um sistema de triângulo nas avaliações de seus resultados operantes. No topo do triângulo estava um índice de investimento sobre o retorno (lucro/total de assets) e a base consistia do índice de margem de lucro (lucro/vendas) e um índice de capital turnover. Este sistema fornece estrutura onde os índices podem ser desenvolvidos em um estilo lógico.

A década de 1920 foi um período de grande entusiasmo para as possibilidades de usar índices financeiros como ferramenta de análise.

Em 1925, Stephen Gilman fez algumas críticas à análise de índices, como por exemplo, “suas mudanças no tempo não podem ser interpretadas pois o numerador e o denominador variam”. Ele propôs que a construção de índices encadeados que

indicassem as variações havidas nos principais ítems em relação a um ano (iniciando o que chamamos de análise vertical).

Na próxima década existiram dois desenvolvimentos significantes relacionados diretamente com os índices financeiros. O primeiro foram as discussões sobre a determinação do grupo de índices mais eficaz. Aqui pode-se destacar o trabalho de Roy A. Foulke. A abordagem de Foulke pode ser chamada de “empírica pragmática”, e provavelmente foi suficiente para as necessidades dos que usavam a análise de índices, mas deixa o sujeito da análise de índices livre de qualquer teoria testável. Contudo, o segundo desenvolvimento significativo desta época foi o estudo do uso de índices financeiros para previsão de dificuldades financeiras. Winakor e Smith iniciaram este movimento. Trabalharam também nesta linha Fitzpatrick, Ranser e Foster [Horrigan, 1968]. Eles representaram um evento extremamente significativo no desenvolvimento de índices financeiros pois foram os primeiros que cuidadosamente se esforçaram em usar métodos científicos para determinar a utilidade dos índices.

Na década de 30 também surgiu, dentro da empresa Du Pont, um modelo de análise da rentabilidade que decompunha a taxa de retorno em taxas de margem de lucro e giro dos negócios, chamado análise de ROI (*return on Investment*).

Desde o início dos anos 40, o desenvolvimento da análise de índices financeiros continuou por vários caminhos. Um aspecto importante deste período foi a crescente ênfase dada ao papel dos índices financeiros nas operações de pequenos negócios. A administração de pequenos negócios, em particular, gerou muito interesse na utilidade de índices financeiros como um instrumento de gerenciamento. Outro aspecto importante, foi o interesse pela análise de índices financeiros por outros países como Austrália, Inglaterra, França, Índia, Japão, Canadá e outros [Matarazzo, 1987; Horrigan, 1968].

No Brasil, até 1968, a Análise de Balanço era um instrumento ainda pouco utilizado na prática. Nesse ano, foi criada a SERASA, empresa que passou a operar como central de Análise de Balanços de bancos comerciais. Atualmente, novas pesquisas estão sendo efetuadas pela DIEM - Diagnósticos Empresariais Computadorizados S/C Ltda. - cujo Know-how nada deve aos centros internacionais mais avançados [Matarazzo, 1987].

Os índices financeiros são usados como entrada para modelos estatísticos para prever vários tipos de negócios e identificar características financeiras. Horrigan em 1966 usou a análise de correlação [Horrigan, 1965], Pinches e Ming em 1973 usaram a Análise Discriminante Múltipla [Pinches & Mingo, 1973] enquanto Ingram e Copeland em 1984 utilizaram a Análise de Regressão [Ingram & Copeland, 1984]. Mas o principal foco foi sobre o teste (principalmente multivariado) de modelos estatísticos que usaram índices financeiros para prever a falência de empresas. Estes são baseados no trabalho de Beaver (1966) e Altman (1968) [Beaver, 1966; Altman, 1968].

Existe uma multiplicidade de fatores que determinam a lucratividade de um negócio e suas chances de sobrevivência. Um destes fatores, de interesse particular aos indivíduos no gerenciamento dos negócios, é a eficácia da tomada de decisão do gerenciamento. As informações dos registros contábeis da empresa servem como sinal que decisões são necessárias ou fornecem dados que podem ser usados na decisão.

A razão para o uso da análise de índices financeiros é que ela expressa as diversas figuras das declarações financeiras como índices e a informação revelará o que está faltando quando números individuais são observados. A teoria é que gerentes podem então usar esta informação para melhorar a eficiência e a lucratividade de suas operações. Associado com esta teoria está a suposição implícita que a informação da análise de índices, especialmente a tendência dos índices, capacita o gerenciamento prever e possivelmente evita a falência dos negócios.

Beaver, em 1966, deu uma contribuição substancial para suportar parte da informação em estudo de índices financeiros diferenciando entre o sucesso e o não sucesso da empresa. O trabalho de Beaver é um marco para as pesquisas de índices financeiros. Mais tarde estudos forneceram evidências de que os índices financeiros podem prever a falência de corporações, pelo menos estatisticamente, por desenvolver e testar funções discriminantes preditivas [Altman, 1968; Deakin, 1976]. Contudo, somente Edmister [Edmister, 1972] usou uma amostra de pequenas empresas para desenvolver funções discriminantes preditivas. Embora alguns conceitos metodológicos terem sido expressados com respeito à previsão

retrospectiva da falência do negócio [Eisenbeis, 1977], os estudos mencionados acima têm fornecido evidências do conteúdo de informações dos índices financeiros.

Além do fato de que, o uso da análise de índices financeiros por pequenas empresas ser muito importante e essencial na averiguação de como a firma está operando (eficientemente ou não) [Brigham, 1979], existem algumas evidências empíricas sobre o melhoramento da lucratividade e sobrevivência de pequenas empresas que usam índices financeiros, quando comparado com aquelas que não usam [Thomas & Evanson, 1986].

Muito estudos foram e estão sendo feitos sobre a análise de índices financeiros. Entre eles estão os estudos que usaram técnicas de Inteligência Artificial como Redes Neurais, Sistemas Especialistas e Conjuntos Difusos em Análise de Índices Financeiros para o diagnóstico da saúde financeira de empresas [Whalen & Schott, 1985, Pacheco, 1986].

2.4.3 Aspectos a Serem Considerados

Outros aspectos importantes a serem considerados são os seguintes:

- a) não é necessário considerar um grande número de índices financeiros afim de analisar uma situação específica devido ao fato que muitos índices têm variáveis em comum;
- b) a relevância dos índices diferentes depende completamente da comparação dos índices com dados similares da mesma atividade econômica [Bernstein, 1989].

As principais áreas de pesquisa de índices financeiros são [Salmi & Martikainen, 1994]:

- a forma de funcionalidade dos índices financeiros;
- propriedades estatísticas através do tempo;
- categorização dos problemas financeiros.

Forma Funcional dos Índices Financeiros

Em relação a forma de distribuição dos índices e seus comportamentos estatísticos, não existe ainda consenso na literatura. As diferentes formas de

distribuição dos índices dependem basicamente do tipo de índice e atividade econômica [Barnes, 1987; Ezzamel & Mar-Molinero, 1990; Foster, 1987]. Em relação ao comportamento do índice financeiro ao longo do tempo, na maioria dos casos os índices não têm um comportamento randômico e seu ajustamento depende das condições ambientais do setor, dos objetivos estratégico e da informação desejada. [Ezzamel & Mar-Molinero, 1990; [Chu, 1992; Lee & Wu, 1988].

Tradicionalmente o maior objetivo do uso de dados financeiros na forma de índices é tornar os resultados comparáveis através de empresas e através do tempo para controlar o tamanho. O requerimento determinado usualmente no controle por tamanho é que o numerador e o denominador sejam proporcionais. O trabalho precursor que está neste campo é de Lev and Sunder [Lev & Sander, 1979]. Com o objetivo de avaliar a hipótese da proporcionalidade, deve ser assumido alguma suposição em relação ao tamanho da empresa e o setor econômico. A forma funcional do denominador do índice é muito importante para a validade da hipótese.

Alguns pesquisadores [Barnes, 1982] mostraram que a não normalidade de vários índices pode ser deduzida da análise de proporcionalidade dos índices. Horrigan [Horrigan, 1983] coloca que a pesquisa de análise financeira deveria estar mais interessada no papel dos índices financeiros por si só ao invés da “natureza dos componentes dos índices ou no papel incidental dos índices como o desvio do tamanho dos dados”. A validade dos índices financeiros deveria ser determinada pelo seu sucesso no processo de tomada de decisão das diferentes partes interessadas (donos, gerentes, empregados) [Salmi & Marttikainen, 1994]. Para ilustrar, considera-se o impacto potencial da economia de escala. Para avaliar a eficiência do gerenciamento uma comparação direta dos índices financeiros de grandes e pequenas empresas teria que ser ajustada para o efeito do tamanho. Por outro lado, um investidor avaliando diferentes investimentos pode estar mais interessado no nível de lucratividade sem considerar se é ou não o resultado do efeito do tamanho.

McDonald e Morris [McDonald & Morris, 1984; McDonald & Morris, 1985] apresentaram o primeiro estudo empírico extensivo da validade estatística do método de índice. Os autores estudaram a hipótese da heterocedasticidade do modelo Y/X onde $Y(i) = a + bX(i) + e(i)$. O modelo acima é a idéia central nesta área. Eles encontram suporte para a Análise de Índices Financeiros para comparações dentro de

ramos da indústria, porém em comparações inter-indústrias a proporcionalidade dos índices financeiros não é suportada.

Alguns autores como Berry e Nix [Berry & Nix, 1991], contudo, lançaram dúvidas em relação a generalidade do trabalho de McDonald e Morris. Pela comparação de valores e índices agregados com pesos iguais McLeay e Fieldsend [McLeay & Fieldsend, 1987] concluíram que o comportamento não proporcional de alguns índices financeiros varia com cada índice, tamanho da empresa e setor econômico diferentes. Suposto ser convertida em caixa durante o ciclo operacional, o termo bens fixos não é correto e responde a um aspecto histórico. Mais precisamente, bens fixos são compostos por bens que afirma reserva em diversos períodos, a fim de manter o negócio. As despesas correntes são compostas e débito a curto prazo que na maioria dos casos seus pagamentos estão relacionados com a conversão de bens correntes em caixa. Os débitos a longo prazo são compostos por débitos cuja maturidade dos dados estão a mais de um ano formando os dados do Balanço Patrimonial corrente. A categoria das ações representa quantos destes bens estão financiados pelos donos da empresa.

É digno mencionar que o desvio da proporcionalidade dos índices financeiros está na verdade relacionado à hipótese de distribuição. Por exemplo, Fieldsend, Longford e McLeay [Fieldsend et al., 1987] notaram que alguns índices são esperados ser uma distribuição log-normal devido ao limite inferior. Conseqüentemente eles testaram empiricamente um modelo de regressão log-normal $\ln Y(ij) = a + b \ln X(ij) + g(j) + e(ij)$ onde o efeito da indústria $g(j)$ é especificado explicitamente pelo modelo. Outros trabalhos estão em correspondência com as pesquisas mencionadas: a hipótese da proporcionalidade tem suporte teórico somente se os efeitos da atividade econômica específica são considerados.

Outro aspecto importante em relação a forma de funcionalidade dos índices financeiros é a existência de índices baseado no mercado; estes índices não serão considerados neste trabalho devido à suposição em relação ao tamanho da empresa (trabalha-se com pequenas empresa). Mas por outro lado é importante mencionar que este tipo de índices também apresentam desvios da hipótese da proporcionalidade

como foi mostrado na análise do índice E/P^2 por Booth, Martikainen, Perttunen e Yli-Olli [Booth et al., 1994].

Características de Distribuição dos Índices Financeiros

A motivação para olhar para as propriedades da distribuição normal dos índices financeiros é porque geralmente na análise de índices financeiros assume-se que os índices tenham uma distribuição normal. Isto é devido à importância dos testes nos métodos paramétricos que prevalecem nas pesquisas de análise de índices financeiros. Podemos citar, por exemplo, a análise de regressão e a análise discriminante que confiam na suposição da normalidade.

Um dos trabalhos pioneiros neste área foi feito por Mecimore. Usando medidas estatísticas descritivas (desvio padrão, média) ele encontrou a não normalidade cross-seccional e assimetria positiva na amostra de 500 empresas Fortum selecionadas randomicamente [Mecimore, 1968].

O trabalho mais citado na literatura, e também o trabalho pioneiro neste campo foi feito por Deakin [Deakin, 1976]. Suas descobertas qui-quadradas rejeitam (com uma exceção) a normalidade de onze índices numa amostra de 1114 companhias. Desvio menores foram observados quando transformações logarítmicas e raiz-quadrada foram aplicadas, mas a normalidade não foi ainda suportada.

Outros pesquisadores posteriores [Bird & McHugh, 1977] também confirmaram a hipótese da não normalidade. Estas evidências guiaram muitos pesquisadores a procurar alguns métodos que pudessem restituir a hipótese da normalidade. Os métodos mais comuns usados foram transformações funcionais e remover os pontos discrepantes.

Finalmente, é bom mencionar que a pesquisa conduzida por Watson [Watson, 1990] que examinou as propriedades das distribuições de quatro índices de quatrocentas amostras de empresas. O principal resultado foi que a distribuição normal multivariada é rejeitada se os pontos discrepantes não são removidos.

² E/P = Earnings/price (Lucro/Preço)

Categorização dos Índices Financeiros

Uma questão principal em Análise de Índices Financeiros e na prática é determinar um conjunto de índices financeiros que cubra as atividades da empresa. Existem na literatura quatro abordagens principais em relação a classificação dos índices financeiros. A descrição básica é mostrada abaixo:

- a) **Abordagem Pragmática:** Baseada na experiência prática ou nas visões dos autores [Lev, 1974; Foster, 1978; Bernstein, 1989; White et al., 1994].
- b) **Abordagem Dedutiva:** Baseada no sistema do triângulo Du Pont da Figura 2.6 ((lucro/ bens totais), (lucro/vendas) e (vendas/bens totais)) [Courtis, 1978; Laitinen, 1983].
- c) **Abordagem Indutiva:** Baseada principalmente nos dados e métodos estatísticos [Pinches & Mingo, 1973; Chen & Shimerda, 1981; Aho, 1980; Yli-Olli & Virtanen, 1986; Yli-Olli & Virtanen, 1989; Yli-Olli & Virtanen, 1990].
- d) **Abordagem Confirmatória:** Supõe uma classificação a priori e então tenta confirmar a classificação com evidências empíricas [Laurent, 1979; Luoma & Ruuhela, 1991; Kandel & Langholz, 1992].

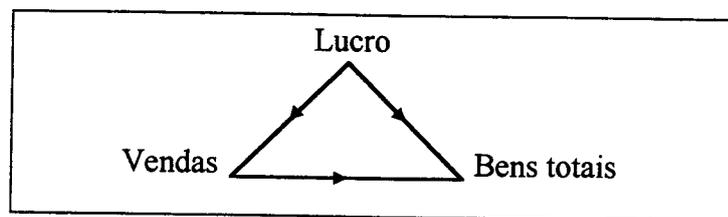


Figura 2.6: Triângulo Du Pont [Matarazo, 1887].

Classificação dos índices

Na literatura, duas espécies de padrões de classificação de índices financeiros tem sido apresentado. Tradicionalmente, a fim de ilustrar a dimensão chave da empresa os índices são classificados de acordo com sua interpretação conceitual convencional. O padrão de classificação tradicional apresentado por LEV divide índices financeiros em 4 categorias [Lev, 1974]:

Índice de Lucratividade: São construídos para avaliar a performance operacional da empresa

Índice de liquidez: Foi definida para definir para indicar a habilidade da empresa saldar suas obrigações financeiras a curto prazo.

Índice de alavancagem financeira (solvência a longo prazo): Indica como a empresa pode saldar o juro e o capital das obrigações a longo prazo.

Índice de eficiência (giro): São designados para medir a eficiência operacional da empresa.

A classificação de LEV é mais popular, mas existem também, por exemplo, a classificação de Horrigan, Courtis e de Tamari [Horrigan, 1965; Courtis, 1978; Tamari, 1978].

A segunda forma de criar padrões de classificação de índices financeiros é selecionar uma faixa ampla de índices financeiros diferentes e estudar como estes índices estão classificados empiricamente. O principal objetivo nestes estudos tem sido expressar a quantidade máxima de informação nos índices financeiros originais por um conjunto reduzido de fatores. O estudo básico desta área foi feito por Pinches e Mingo em 1973 que classificaram 48 índices de 221 empresas em 7 categorias empíricas [Pinches & Mingo, 1973]:

- retorno sobre investimentos
- grau de utilização do capital
- grau de utilização do estoque
- alavancagem financeira
- grau de dependência das vendas a prazo
- liquidez a curto prazo
- posição do caixa

Johnson, usando análise de componentes principais e o método de rotação ortogonal, reportou 9 fatores empíricos para 41 índices [Johnson, 1978]. Outros autores também classificaram os índices em categorias [Laurent, 1979; Aho, 1980; Gombola & Ketz, 1983; Yli-Olli & Virtanen, 1986] - Tabela 2.1).

Classificação dos Índices em categorias	
Lev (1974)	lucratividade, liquidez, alavancagem financeira, eficiência (giro)
Pimches, Mingo e Carruthes (1973)	retorno sobre investimentos, grau de utilização do capital, grau de utilização do estoque, alavancagem financeira, grau de dependência das vendas a prazo, liquidez a curto prazo, posição do caixa
Aho (1980)	estrutura financeira, lucratividade, liquidez, índices de giro e índice de capital de giro, e oportunidades financeiras para investimento
Yli-Olli (1986)	lucratividade, fluxo de caixa de operações e estrutura de gastos, liquidez a curto prazo, estrutura do capital, <i>receivable intensiveness</i> .

Tabela 2.1 Classificação dos Índices Financeiros segundo alguns pesquisadores.

Quando se interpreta os resultados concernentes aos padrões de classificação empíricos de índices financeiros deve-se levar em conta que a classificação depende muito dos índices selecionados, e a utilidade destas categorias em problemas práticos é relativamente baixa se os fatores empíricos não puderem ser interpretados claramente.

2.4.4 Principais Índices Financeiros

A análise através de índices é certamente a mais conhecida, chegando mesmo a ser confundida com a análise de Balanço. Ao longo do tempo, diversos estudos têm demonstrado a validade dos índices financeiros, como “ferramenta” que propicia a medida de desempenho e a solidez das empresas [Silva, 1988].

No Quadro 2.1 que resume os principais índices que são utilizados na análise de Balanço. Para facilitar, dividimos estes índices em 6 categorias:

- **Índices de estrutura de capitais** : Estes índices são aqueles que relacionam a composição de capitais (próprios e de terceiros), que medem os níveis de imobilização de recursos e que buscam diversas relações na estrutura da dívida da empresa. De certa forma, estes índices estão ligados às decisões financeiras de financiamento e investimento.

- **Índices de Estrutura do Ativo Circulante:** Estes índices fornecem elementos para análise e controle das várias contas do circulante.
- **Índices de Capital de Giro:** Estes fornecem valores para uma melhor avaliação da participação e liquidez dos recursos próprios a curto prazo.
- **Índices de Liquidez:** Estes índices visam fornecer uma medida, ou melhor, um indicador da capacidade da empresa de pagar suas dívidas a partir da comparação entre os direitos realizáveis (a curto e a longo prazo) e a exigibilidade (também de a curto e a longo prazo).
- **Índices de Rotação:** os índices de rotação constituem-se em categorias de elevada importância para o analista. O balanço da empresa representa sua situação patrimonial em determinado momento, isto é, como se fosse um fato que mostra algo de forma estática, sem refletir sua mobilidade, seu dinamismo. A empresa, em suas operações, compra, fabrica, estoca, vende e recebe num processo dinâmico e contínuo. Os índices de rotação têm grande contribuição na interpretação da liquidez e da rentabilidade da empresa, à medida que servem de indicadores dos prazos médios de rotação de estoques, recebimentos das vendas e pagamento das compras.
- **Índices de Retorno:** Os índices de retorno, também conhecidos por índices de lucratividade ou mesmo rentabilidade, indicam qual o retorno que o empreendimento está propiciando aos seus acionistas ou proprietários. Podem-se obter, através da análise das comparações financeiras, os indicadores de retorno sobre o investimento, retorno sobre vendas e retorno sobre capital próprio, entre outros.

É bom lembrar aqui que esta classificação não é muito fácil, pois os índices estão inter-relacionados e são interdependentes entre si (servindo este agrupamento por categorias como uma forma didática de apresentá-los). Um mesmo índice pode ser usado como indicador de mais de um problema (embora, eles não forneçam o mesmo nível de informação em cada caso). Além disto, a análise de índices é mais complexa na medida que a conclusão final depende da análise conjunta de um conjunto de índices que refletem a empresa.

2.4.5 Outros Aspectos

Embora autores e profissionais de Análise de Balanço tenham alguns pontos em comum quanto aos principais índices de que se valem, existem algumas diferenças em suas análises. Certos índices como Participação de Terceiros, Liquidez Corrente e Rentabilidade do patrimônio Líquido, são usados por praticamente todos os analistas. Outros porém, como Composição do Endividamento, Liquidez Seca, Rentabilidade do Ativo, Margem Líquida de Lucro, nem sempre fazem parte dos modelos de análise.

O importante não é um grande número de índices, mas de um conjunto de índices que permita conhecer a situação da empresa, segundo o grau de profundidade desejada da análise.

Entretanto, a análise de índices é do tipo que começa muito bem e vai perdendo fôlego à medida que se acrescentam novos índices [Matarazzo, 1987]. Quando, por exemplo, se dobra o número de índices, não se consegue dobrar a quantidade de informações. Isto pode ser ilustrado pela Figura 2.7.

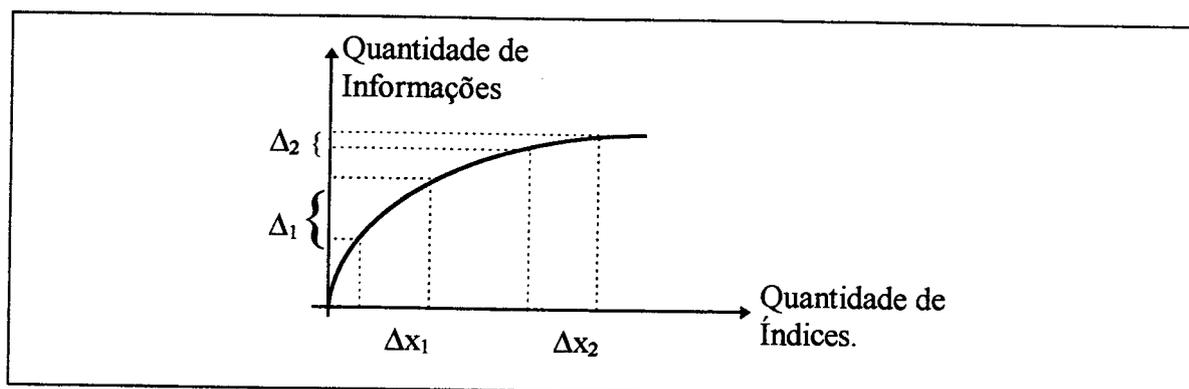


Figura 2.7: Curva da relação quantidade de informação versus quantidade de índices [Matarazzo, 1987].

É importante saber quais índices devem ser selecionados como referencial de avaliação da empresa e quais terão função complementar no auxílio ao analista, em sua interpretação da tendência da empresa. Também é necessário conhecer a importância relativa de cada índice no contexto geral, no sentido de se determinar se cada índice receberá um peso específico ou se serão atribuídos pesos iguais a todos os índices.

O número de índices a ser utilizado deve levar em consideração os seguintes pontos:

- a) **Utilidade dos índices:** Os índices devem ter uma relação direta com o que deseja-se analisar. Eles devem ser abrangentes e um pouco sensíveis em relação à saúde financeira da empresa (bom, médio e ruim); mas não tão sensíveis em relação aos aspectos de tamanho da empresa.
- b) **Contribuição:** como pode ser visto na Figura 2.7, deve-se levar em conta relação quantidade de índices versus o benefício adicional na avaliação da empresa.
- c) **Praticidade:** um número muito grande de índices tende a confundir o analista (principalmente se o analista ainda não tem muita experiência).
- d) **Segurança:** o número de índices utilizados deve propiciar ao analista razoável grau de tranquilidade quanto a eficácia da avaliação do risco.

Além da escolha dos índices, é importante atribuir um peso (um conceito) a cada um dos índices, de modo que se tenha uma avaliação final da empresa, em face dos padrões adotados. Tais padrões devem ser constituídos com base em empresas de mesma atividade, da mesma região geográfica e, se possível, do mesmo porte.

Existem três tipos básicos de avaliações de um índice [Matarazzo, 1987]:

1. pelo significado intrínseco;
2. pela comparação ao longo de vários exercícios;
3. pela comparação com índices de outras empresas - índices-padrão.

Na *avaliação intrínseca de um índice* é possível, de maneira grosseira, avaliar índices pelo seu significado intrínseco. Por exemplo, pode-se tentar qualificar a situação financeira de uma empresa com base no índice de Liquidez Corrente: se apresentar valor de 1,5, sabe-se que para cada unidade monetária de dívida a curto prazo existem 1,5 unidades monetárias de investimento a curto prazo. A folga de 50%

pode ser considerada suficiente como margem de segurança à empresa, essa conclusão é feita ou por intuição ou por experiência do analista.

Mas, a análise do valor intrínseco de um índice é limitada e só deve ser usada quando não dispõe de índices-padrões proporcionados pela análise de um conjunto de empresas.

Na *comparação dos índices no tempo*, a comparação dos índices de uma empresa com os valores observados nos anos anteriores revela-se bastante útil por mostrar tendências seguidas pela empresa. Por exemplo, se por um lado uma empresa pode endividar-se mais a cada exercício e, simultaneamente, apresentar aumento de sua rentabilidade, por outro lado pode ocorrer redução dos índices de liquidez. Essas informações permitem ao analista formar uma opinião a respeito de diversas políticas seguidas pela empresa, bem como das tendências que estão sendo registradas. É fundamental em qualquer avaliação que os índices sejam analisados conjuntamente.

Na *comparação com padrões*, a avaliação de um índice e a sua conceituação como ótimo, bom, satisfatório, razoável ou deficiente só pode ser feita através da comparação com padrões. Assim é preciso definir um conjunto (universo) e, em seguida, comparar um elemento com os demais do conjunto para atribuir-lhe determinada qualificação. Este é um processo natural do raciocínio humano onde todas as avaliações são feitas por comparações, ainda que dificilmente tabuladas metodologicamente (por isso as divergências de opiniões).

No Brasil, a Serasa é certamente quem possui um dos maiores arquivos de empresas analisadas, as quais são utilizadas para elaboração de índices-padrões, compreendendo praticamente todo o território nacional e todos os segmentos e atividades empresariais. Tais padrões são usados pela própria Serasa para atribuição de avaliação às empresas por ela analisadas.

O processo geral de extração de informações importantes para suportar decisões numa empresa não é simples. Este processo depende do setor econômico, do tamanho da empresa, do padrão de comparação e além disso é dinâmico. Portanto, a Análise Financeira é um processo que requer especialistas com muita experiência. A primeira questão em Análise Financeira é a categorização do problema. Porém, é importante notar que os limites entre cada uma das categorias não é rígida. A

principal razão é devido ao fato que as categorias sobrepõem-se no tempo; isto é, um problema de débito pode transformar-se em um problema de lucratividade no futuro e vice versa. Isto mostra a que uma figura financeira (por exemplo, índices) não pertença estritamente a uma categoria. O que sem dúvida está presente são graus diferentes de relações entre cada uma das categorias.

A principal conclusão extraída do trabalho de Salmi e Martikainen [Salmi & Marttikainen, 1994] é que a melhor abordagem para a classificação dos índices depende do problema, isto é, depende do objetivo da análise de índices e o meio da aplicação. Particularmente para o diagnóstico da saúde financeira de pequenas empresas, através do raciocínio indutivo, uma abordagem adequada identificada por Martins [Martins, 1996] é a metodologia empírica-pragmática de LEV citada na seção 1.4.3.3.

Símbolo	Índice	Fórmula	Indica
	Estrutura de Capital		
1. CT/PL	<ul style="list-style-type: none"> Participação de capitais de terceiros (endividamento) 	Capital de terceiros/ Patrimônio Líquido	O percentual de capitais de terceiros em relação ao Patrimônio Líquido, retratando a dependência da empresa em relação aos recursos externos.
2. PC/CT	<ul style="list-style-type: none"> Composição do endividamento 	Passivo Circulante/ Capitais de terceiros	Indica quanto da dívida total da empresa deverá ser pago a curto prazo, isto é, compõe o ativo circulante.
3. AP/PL	<ul style="list-style-type: none"> Imobilização do Patrimônio Líquido 	Ativo Permanente/ Patrimônio Líquido	Quanto do Patrimônio Líquido da empresa está aplicado no ativo permanente.
4. AP/ PL+ELP	<ul style="list-style-type: none"> Imobilização dos Recursos não correntes 	Ativo Permanente/ (Patrimônio Líquido + Exigível a Longo Prazo)	Quanto dos recursos não concorrentes (Patrimônio Líquido e Exigível a longo prazo) foi destinado ao Ativo Permanente.
	Estrutura do Passivo		
1. PL/PC	<ul style="list-style-type: none"> Participação do exigível a curto prazo 	Patrimônio Líquido/ Passivo Circulante	Indica a proporção de recursos próprios em relação aos recursos de terceiros aplicados a curto prazo. O inverso (Passivo Circulante/PL) fornece o nível de endividamento a curto prazo da empresa.
2. PL/ELP	<ul style="list-style-type: none"> Participação do exigível a longo prazo 	Patrimônio Líquido/ Exigível a longo Prazo	Indica a proporção de recursos próprios em relação aos recursos de terceiros a longo prazo. O inverso (Exigível a LP/PL) fornece o nível de endividamento a longo prazo da empresa.
3. PL/ET	<ul style="list-style-type: none"> Participação do exigível total 	Patrimônio Líquido/ Exigível Total	Indica a proporção de recursos próprios em relação ao total de recursos de terceiros existentes na empresa.
4. PL/PT	<ul style="list-style-type: none"> Participação do passivo total 	Patrimônio Líquido/ Passivo Total	Indica a proporção de recursos próprios em relação ao total de recursos (próprios e de terceiros) existentes na empresa.

Estrutura da Ativo Circulante			
1. Disp/AC	<ul style="list-style-type: none"> Participação do Disponível 	Disponível/ Ativo Circulante	Indica a porcentagem de participação das disponibilidades no ativo circulante.
2. RLP/AC	<ul style="list-style-type: none"> Participação do Realizável 	Realizável a curto prazo/ Ativo Circulante	Indica a porcentagem de participação dos direitos realizáveis a curto prazo no ativo circulante.
3. E/AC	<ul style="list-style-type: none"> Participação dos Estoques 	Estoques/ Ativo Circulante	Indica a porcentagem de participação dos estoques no ativo circulante.
4. AC/AT	<ul style="list-style-type: none"> Participação do Circulante Total 	Ativo Circulante/ Ativo Total	Indica a participação relativa do ativo circulante no total do Ativo.
5. Emp/AC	<ul style="list-style-type: none"> Participação de recursos de terceiros 	Empréstimos/ Ativo Circulante	Indica a participação relativa dos empréstimos no ativo circulante.
6. Forn/AC	<ul style="list-style-type: none"> Participação de credores 	Fornecedores/ Ativo Circulante	Indica a participação relativa de fornecedores no ativo circulante.
Capital de Giro			
1. CG/AC	<ul style="list-style-type: none"> Volume de recursos próprios aplicados a curto prazo 	Capital de Giro/ Ativo Circulante	Indica a porcentagem dos recursos próprios da empresa aplicados no ativo circulante.
2. CG/PL	<ul style="list-style-type: none"> Liquidez dos recursos próprios 	Capital de Giro/ Patrimônio Líquido	Indica a porcentagem de liquidez dos recursos próprios da empresa; ou seja, do total do seu patrimônio líquido, qual o volume aplicado em itens de rápida conversão.
Liquidez			
1. LG	<ul style="list-style-type: none"> Liquidez Geral 	(Ativo Cir. + Realiz. a Longo Prazo)/ (Passivo Cir. + Exig. a Longo Prazo)	Indica quanto a empresa possuem em dinheiro, bens e direitos realizáveis a curto e longo prazo, para fazer face às suas dívidas.
2. LC	<ul style="list-style-type: none"> Liquidez Corrente 	Ativo Circulante/ Passivo Circulante	Indica quanto a empresa possui em dinheiro mais bens e direitos realizáveis no próximo exercício, comparado com suas dívidas a serem pagas no mesmo período.
3. LS	<ul style="list-style-type: none"> Liquidez Seca 	(Disponível+Títulos a Rec. + Outros Ativos de Rápida Conversibilidade)/ Passivo Circulante	Indica quanto a empresa possui em dinheiro, mais aplicações financeiras, mais Títulos a Receber e Outros Ativos de Rápida Conversibilidade, para fazer face ao seu Passivo Circulante.
4. LA	<ul style="list-style-type: none"> Liquidez Absoluta 	Disponível / Passivo Circulante	Indica a porcentagem das dívidas a curto prazo em condições de serem salgadas imediatamente.
5. LE	<ul style="list-style-type: none"> Liquidez Estimada 	(Disponível +Previsão de entradas de caixa a curto prazo)/ Previsão de saídas de caixa a curto prazo	Indica a liquidez a curto prazo através de uma previsão de entradas e saídas de caixa.
Rentabilidade			
1. LB/VL	<ul style="list-style-type: none"> Margem Bruta 	Lucro Bruto/ Vendas Líquidas	Indica o desempenho dos custos de produção. Dado que o lucro bruto é obtido pela diferença entre vendas e o custo de produção de mercadorias ou dos produtos vendidos, um aumento deste índice denotará melhor eficiência produtiva da empresa, no que se refere aos seus custos fabris.
2. V/AT	<ul style="list-style-type: none"> Giro do Ativo 	Vendas Líquidas/	Estabelece relação entre as vendas efetuadas

		Ativo Médio	no período e os investimentos totais da empresa.
3. LL/V	• Margem Líquida	Lucro Líquido/ Vendas Líquidas	Compara o lucro líquido em relação às vendas líquidas do período, fornecendo percentual de lucro que a empresa está obtendo em relação ao seu faturamento.
4. LL/AT	• Rentabilidade do Ativo	Lucro Líquido/ Ativo Total	Indica a lucratividade que a empresa propicia em relação aos investimentos totais.
5. LL/PL	• Rentabilidade do Patrimônio Líquido	Lucro Líquido/ Patrimônio Líquido - Lucro Líquido	Indica o retorno dos recursos próprios investidos na empresa. Este índice, espesso pela relação entre os recursos líquidos obtidos em determinado período e o capital próprio empregado, é de grande importância para os acionistas ou proprietários da empresa. Exerce, inclusive, decisiva influência a médio e longo prazo sobre o valor de mercado das ações ou da própria empresa.
Rotação			
1. PMRE	• Prazo médio de rotação de Estoques	Estoque Médio * DP ³ / Custo da mercadoria vendido	Indica quantos dias, em média, os produtos ficam armazenados na empresa antes de serem vendidos ⁴ .
2. PMRV	• Prazo médio de recebimentos das vendas	Duplicatas a Receber * DP / Vendas	Indica quantos dias, em média, a empresa leva para receber suas vendas.
3. PMPC	• Prazo médio de pagamento das compras	Fornecedores * DP/ Compras	Indica quantos dias, em média, a empresa demora para pagar seus fornecedores.
Índices de dependência e independência financeira			
1. PL/AT	• Independência Financeira	Patrimônio Líquido / Ativo Total	Indica a independência financeira da empresa, ou seja, do montante investido pela empresa no seu ativo, qual a participação dos recursos próprios. A empresa alcança uma independência absoluta quando este índice for igual a 1.
2. ET/AT	• Dependência Financeira	Exigível Total/ Ativo Total	Indica a dependência financeira da empresa. É um complemento do índice anterior. Indica a participação do capital de terceiros nos investimentos efetuados no ativo.

Quadro 2.1: Quadro Resumo dos Índices Financeiros.

2.5 Aplicações dos Índices Financeiros

Os Índices Financeiros tem sido usados como entradas para modelos estatísticos avançados para prever muitos tipos de negócios e para identificar

³ DP - Dias do período considerado.

características financeiras [Barnes, 1987]. Notáveis estudos incluem Ingram e Copeland [Ingram & Copeland, 1984] que usaram Análise de regressão para medir a relação entre as diferenças nos Índices Financeiros através das municipalidades e os riscos em seus bônus (e.g. *Junk Bonds*).

Horrigan usou análise de correlação [Horrigan, 1966] enquanto Pinches e Mingo usaram Análise Discriminante Multivariada (ADM) para prever a cotação dos bônus da empresa por meio de índices individuais [Pinches & Mingo, 1973]. Rege e Belkaoui realizaram estudos onde usaram índices Financeiros para identificar características de *takeover targets* [Rege, 1984], [Belkaoui, 1978].

O principal foco tem sido os modelos estatísticos (principalmente multivariada) que usaram Índices Financeiros para prever a falência de negócios. Estes são baseados no trabalho original de Beaver e Altman [Beaver, 1966], [Altman, 1968].

Beaver comparou uma amostra de firmas falidas com uma amostra de firmas não falidas e estudou seus índices financeiros durante mais ou menos 5 anos antes da falência e descobriu que eles tem grande habilidade de previsão [Beaver, 1966]. A técnica usada por Beaver é conhecida como a análise de classificação e é essencialmente univariada.

Altman usou a técnica bem conhecida em estatística, ADM. Ele mostrou que os presságios dos perigos financeiros são discrimináveis usando uma combinação de índices financeiros. Os índices específicos são determinados usando ADM [Altman, 1968]. O modelo de Altman popularizou-se com o nome de Z-score e foi comercializado com sucesso para análise de crédito, análise de investimentos e avaliação para ver se existe ou não problemas de concordata ou falência na empresa.

Desde 1968, ADM tem sido a principal abordagem para identificar os problemas financeiros em uma grande variedade de áreas em negócios. Contudo, a técnica de ADM tem sido criticada pois a validade de seus resultados depende de suposições restritivas sobre a separabilidade linear, na normalidade multivariada e independe das variáveis preditivas [Ohlson, 1980; Odom & Shardo, 1990; Karels & Prakash, 1987]. Infelizmente, índices financeiros, em sua grande maioria, violam as

⁴ A Administração dos estoques é tão importante que, por si só, representa complexo tema da administração

suposições da ADM. As restrições da ADM são incompatíveis com a natureza complexa, os limites e a inter-relação dos índices. Então os poderes da ADM para índices financeiros está comprometida [Lacher et al, 1991].

Mas, apesar dos muito estudos, os pesquisadores não descobriam nada melhor que a ADM até os recentes desenvolvimentos das Redes Neurais Artificiais.

A primeira comparação de análise discriminante com Redes Neurais Artificiais na área de previsão de falência foi apresentada por [Odom e Shardo, 1990]. Neste trabalho Redes Neurais Artificiais mostram-se superiores à análise discriminante. Já no trabalho apresentados por Erxleben e Baetge em 1991 a análise discriminante mostrou ser superior, enquanto no trabalho apresentado em 1992, a análise discriminante e as redes neuronais apresentaram resultados semelhantes [Refenes, 1995]. Não é possível determinar se as redes Neurais são superiores (existem muitos tipos de redes neuronais) [Refenes, 1995] mas já se pode afirmar que elas apresentam uma performance tão boa quanto a análise discriminante.

Muitos outros pesquisadores estão usando redes Neurais, como Lacher et al. que usaram uma Rede Neuronal Artificial para a classificação da saúde financeira de uma firma [Lacher et al, 1991]. Whalen que usou uma Rede Neuronal Difusa para diagnóstico de empresas [Whalen & Schott, 1985]; Pacheco em sua tese de doutorado também usou uma Rede Neuronal Difusa para diagnósticos de pequenas empresas [Pacheco, 1996].

2.6 Conclusão

A Análise de Declarações Financeiras é tarefa bastante complexa e de fundamental importância numa sociedade moderna. O método mais usado e muitas vezes confundido com a análise de declarações financeiras é a análise de índices financeiros. Estes índices comparam variáveis financeiras e são extraídos, geralmente, do Balanço Patrimonial e do demonstrativo de resultado. Esta técnica reduz a quantidade de informações e enfatiza as relações entre elementos financeiros ao invés de seus valores individuais.

A análise destes índices, além de auxiliarem no gerenciamento da empresa, auxilia na previsão estatística de falência, na análise da saúde financeira, e no planejamento de estratégias.

As pesquisas sobre a forma funcional e as propriedades da distribuição dos índices financeiros apontam uma relação dos mesmos com o tamanho e atividade da empresa, requerendo períodos de análise pequenos. Quanto à categorização dos índices, esta depende do objetivo da análise e do meio da aplicação e também do tempo.

Portanto, o processo geral de extração de informações importantes para suportar decisões numa empresa não é simples. Depende do setor econômico, do tamanho da empresa, do padrão de comparação e além disso é dinâmico.

3. TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

3.1 Introdução

O termo Inteligência Artificial (IA) foi proposto em 1956 por Jonh MacCarthy, um dos pioneiros, da Universidade de Stantford, nos Estados Unidos.

IA é um termo que abrange muitas definições [Turban, 1992]. Mas a maioria dos especialistas concordam que a IA está baseada em duas idéias básicas: primeiro envolve o estudo do processo do pensamento humano (para entender o que é inteligência); segundo, trata com a representação destes processos via máquina (computador, robôs,...).

Nos últimos anos tem-se presenciado um crescente interesse pela interdisciplinaridade em aplicações de Inteligência Artificial (IA). O tópico está sendo abordado em muitas publicações. Muitos periódicos⁵ têm publicado e dedicado muita atenção sobre IA. Dezenas de livros sobre IA tem aparecido no mercado. Muitos artigos novos em IA estão sendo publicados regularmente, e conferências e convenções neste tópico estão sendo realizados mundialmente. E além disto, as aplicações comerciais estão projetadas para alcançar milhões de dólares até o ano 2000 [Turban, 1995].

Este desenvolvimento pode ter um impacto significativo em muitas organizações públicas e privadas, e no modo que elas são administradas.

Atualmente, dentro do campo da IA, existem diversas linhas de pesquisa e técnicas diferentes em estudo, como Sistemas Especialistas, Redes Neurais Artificiais, Sistemas Difusos, Algoritmos Genéticos, entre outras. Essas técnicas têm

⁵ Muitos periódicos tem publicado artigos na área de IA como por exemplo, IEEE Transaction in Fuzzy Systems, Fuzzy Sets and Systems, Intelligent Decision Analysis.

mudado constantemente a cada nova adaptação, a cada novo problema, tornando-se cada vez mais elaboradas, mais complexas e naturalmente mais “inteligentes”. Ao mesmo tempo, encontra-se com mais frequência a combinação destas técnicas para a solução de problemas mais complexos. Essa combinação é chamada na literatura como Sistemas Híbridos.

O surgimento dos Sistemas Híbridos foi muito natural, afinal o próprio homem constitui-se de muitos elementos como evolução, adaptação, sinapses, memória e processamento de muitas outras facetas que ainda se desconhece ou não se consegue entender. O corpo humano é um grande Sistema Híbrido e extremamente complexo. Seguindo nesta linha, tem-se um campo de pesquisa extremamente vasto.

Neste capítulo apresenta-se, de uma forma simplificada, as bases de cinco técnicas inteligentes cuja aplicação no presente trabalho tem relevância: Sistemas Especialistas, Redes Neurais Artificiais, Sistemas Difusos, Algoritmos Genéticos e Raciocínio Baseado em Casos, assim como, sistemas inteligentes híbridos, uma técnica emergente na área de IA. Todas estas técnicas que serão vistas, estão hoje sendo aplicadas para resolver problemas na área de finanças, indo desde análise de pequenos negócios até previsão de bolsa de valores.

3.2 Técnicas de Inteligência Artificial

3.2.1 Introdução

As companhias estão sempre reunindo uma grande quantidade de dados como os balanços das empresas, índices de bolsas de valores, vendas de ações. Detalhes são registrados em uma base de dados. As companhias inovadoras estão considerando a grande quantidade de dados que elas mantêm como um tesouro em potencial que pode ser “explorado” para encontrar padrões e relações importantes, talvez fundamentais, e que podem trazer soluções para muitos problemas.

Sistemas Inteligentes são uma categoria de inventos computacionais que podem encontrar padrões e descobrir relações numa quantidade grande de dados. Estes sistemas cercam um repertório de técnicas incluindo Redes Neurais, Sistemas Difusos, Algoritmos Genéticos e Sistemas Especialistas.

O domínio de aplicação está crescendo rapidamente (Figura 3.1). Countrywide Funding, a maior criador de hipotecas dos Estados Unidos, usa um sistema baseado em regras adaptativo para subscrever suas hipotecas. A Internacional Visa tem um sistema de rede Neuronal para detectar o uso fraudulento de cartões de créditos e o Banco Fiji usa sistemas Difusos para *bond trading*.

Em todas estas companhias o fator de motivação para usar sistemas inteligentes é similar - o aumento da qualidade de seus serviços e redução de custos.

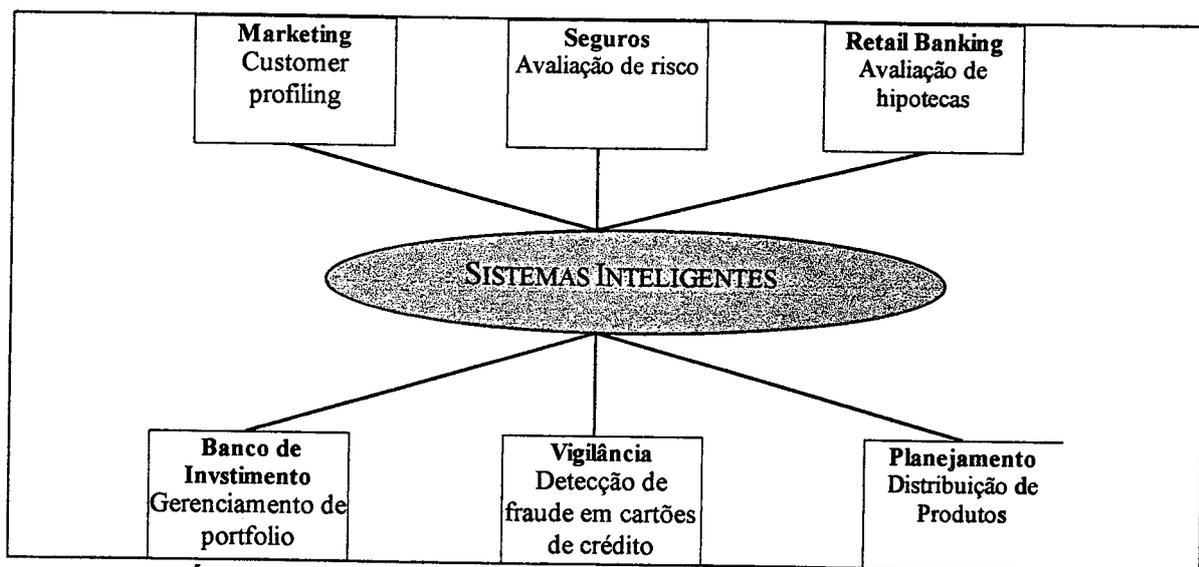


Figura 3.1: Áreas de Aplicação de Sistema Inteligentes [Goonatilake & Treleaven, 1995a].

Um tema comum em sistemas inteligentes é a sua imitação da natureza (humana). Redes Neurais, por exemplo, são inspiradas na funcionalidade das células nervosas no cérebro. Semelhante a humanos, as Redes Neurais podem aprender a reconhecer padrões pela exposição repetida de exemplos diferentes. Elas podem ser usadas para reconhecer padrões ou salientar características, quer eles sejam caracteres manuscritos, empréstimos ou decisões de bons negócios. Assim como os humanos tem a capacidade de reconhecer um caracter manuscrito produzido por pessoas diferentes que possuem estilos distintos, as Redes Neurais podem também reconhecer padrões em dados que são inexatos e incompletos.

Na seqüência, faz-se uma revisão das técnicas de IA: Sistema Especialistas, Redes Neurais, Sistemas Difusos, Algoritmos Genéticos e Raciocínio Baseado em

Casos. Para cada técnica introduz-se as operações básicas, seus pontos fortes e fracos e algumas aplicações. Também apresenta-se a integração destas técnicas (Sistemas Híbridos).

3.2.2 Sistemas especialistas

Sistemas Especialistas representam os primeiros e mais estabelecidos sistemas inteligentes. Existem centenas de Sistemas Especialistas (SE) operacionais em domínios abrangendo desde diagnósticos imprecisos até comércio de mercadorias. Um SE segundo Feigenbaum, é um programa inteligente de computador que soluciona problemas difíceis os quais o ser humano não dá conta de resolvê-los [Feigenbaum, 1977]. Como o nome sugere, os SE tentam incorporar o conhecimento em um domínio específico (limitado) de um especialista humano em um programa de computador. O processo de aquisição de conhecimento de um especialista - elicitacão do conhecimento - tipicamente envolve uma série de entrevistas e o registro cuidadoso das observações quando o especialista está realizando as tarefas.

Uma vez que o conhecimento está adquirido ele é representado em uma forma que possa ser manipulada pelo computador. Existem muitas pesquisas sobre como encontrar bons esquemas de representação do conhecimento que são eficientes e fáceis de usar. Entre os esquemas de representação do conhecimento mais usados tem-se as regras de produção, *frames*, e redes semânticas.

Uma regra de produção tem a seguinte forma:

SE (*condição-1*) **E** (*condição-2*) **ENTÃO** (*ação-1*) **E** (*ação-2*)

Um SE geralmente constitui-se dos seguintes componentes básicos (Figura 3.2):

- ◆ *base de conhecimento*, que contém conhecimentos necessários para entender, formular e solucionar os problemas (são os fatos conhecidos pelo sistema);
- ◆ *máquina de inferência*, que é o mecanismo de controle para guiar o processo de raciocínio do sistema. Também conhecido como estrutura de controle ou interpretador de regras (em SE baseados em regras);
- ◆ *memória de trabalho (Blackboard)*, que é a porção da memória reservada para guardar a trilha das entradas, conclusões intermediárias e a saída;

- ◆ *Interface com o usuário*, que é designada para fornecer um meio conveniente de comunicação entre o usuário e o engenho de inferência;
- ◆ *Mecanismo de aquisição do conhecimento*, que é usado para adquirir experiência humana e transformar em base de conhecimento.

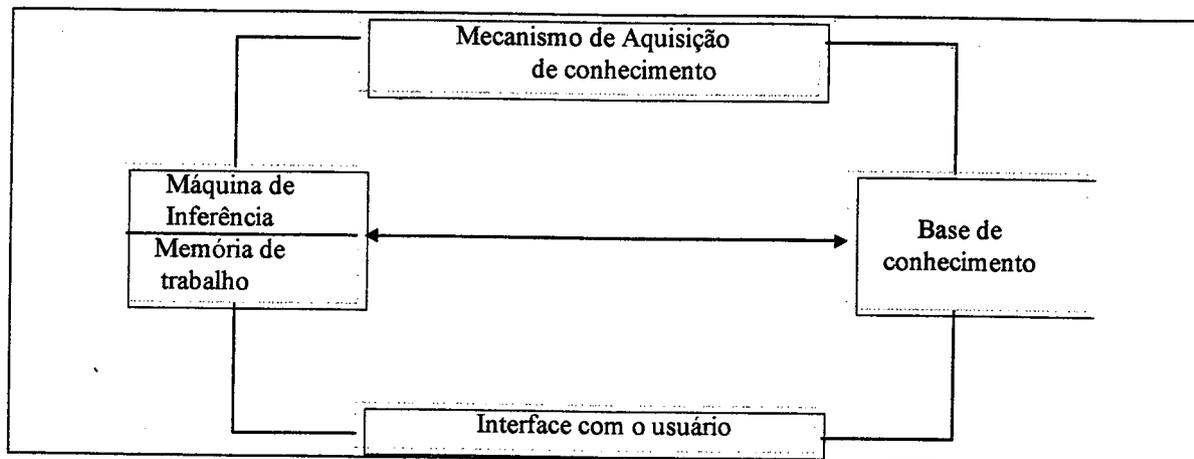


Figura 3.2 Estrutura genérica de um Sistema Especialista .

Os SEs representam uma transição revolucionária do processamento de dados tradicional para o processamento do conhecimento. Eles oferecem um meio para incorporar a capacidade dos humanos e o poder dos computadores.

Vantagens e Limitações dos Sistemas Especialistas

Uma grande vantagem dos SE é a representação explícita do conhecimento, de modo que o conhecimento contido nos programas seja relativamente fácil de ler e entender (são capazes de tratar com informação simbólica). Os SEs podem gerar explicações de como ele chegou a uma conclusão particular e também podem interagir com o usuário melhorando sua produtividade.

Como uma limitação particular dos SEs, uma das principais desvantagens, é que eles não têm mecanismos de aprendizagem automática das regras. Além disto, eles não se adaptam ou aprendem com as mudanças no meio no qual eles operam.

Em algumas áreas, especialistas descrevem suas ações como sendo difusas e intuitivas e têm dificuldades para transformar suas experiências em regras de sistemas especialistas, que usualmente tem definições de lógica binária. Desta forma,

esquemas de representação de conhecimento, tal como Lógica Difusa, são mais flexíveis e necessárias.

Aplicações

As aplicações dos SEs são muita amplas atingindo desde jogos até sistemas de controle. As aplicações em finanças e negócios podem ser divididas nas áreas de corporação financeira, mercado financeiro e instituições financeiras de seguros . Dentre as aplicações em negócios que tiveram sucesso pode-se destacar:

- ◆ controle de fraude de cartões de créditos [Newquist, 1987] ;
- ◆ análise de índices financeiros [Pau, 1991];
- ◆ liberação de crédito [Enrado, 1991; Weber, 1993];
- ◆ planejamento financeiro [Brown et al., 1990];
- ◆ contabilidade [Brown et al., 1990];
- ◆ avaliação de risco [Radding, 1991];
- ◆ análise de *portfolio* [Ram, 1990].

Os dois sistemas especialistas que são dirigidos para análise financeira são *Answers* [Blocher, 1990] e *Financial Statement Analyzer (FSA)* [Mui et al., 1990].

Answers é um SE que constitui-se de dois módulos, um que realiza uma análise financeira e o outro uma análise de projeções. A função da análise de índices é provocar comentários das variações observadas em índices financeiros *chaves* usados pelo sistema. O objetivo destes comentários é sugerir aos gerentes questões a serem endereçadas ou tópicos a serem analisados. Originalmente, este sistema foi concebido para auditoria, mas mais tarde foi usado para treinamento e recomendações.

O FSA é um SE comissionado pelos *U.S. Securies e Commision Exchange State* (SEC) para Arthur Andersen a fim de extrair informações valiosas das declarações financeiras correntes. Originalmente, o FSA tinha por objetivo processar automaticamente o cálculo dos índices. O sistema FSA calcula índices padrões das declarações financeiras das empresas.

Histórico

Distingui-se três períodos no desenvolvimento de SE:

- ◆ **1960-1969** - Durante este período, os SE foram desenvolvidos em domínio científico. Uma das primeiras aplicações (em Química) foi o DENTRAL, designado para interpretar dados de espectrômetros de massas e determinar a estrutura de moléculas. Um SE para entender a fala HEARSAY, e o MACSYMA para solução de problemas matemáticos, foi também desenvolvidos neste período.
- ◆ **1970-1979** - O início dos anos setenta foi marcado pelo desenvolvimento de diversos SE no campo de Medicina. MYCIN foi designado para ajudar em diagnósticos de meningite e infecções no sangue. CASNET e INTERNIST também foram projetados neste período. No final dos anos setenta, o PROSPESTOR foi projetado para prever depósitos de minerais. Foi também desenvolvido o primeiro SE cujo sujeito de domínio era a educação, SOPHIE. Este SE ensinava a resolver problemas difíceis em circuitos eletrônicos.
- ◆ **1980 - presente** - O início dos anos 80 trouxe grande popularidade e interesse comercial para os SE seguindo o sucesso dos SE desenvolvidos no período anterior. Este período é caracterizado pela comercialização de ferramentas desenvolvidas em SE conhecidas como *expert system shells* (ESS).

No período de 1960 a 1979 os SE eram *hand-crafted*, isto é, todos os componentes de software, assim como a base de conhecimento, eram codificadas pelas pessoas que o tinham desenvolvido. O desenvolvimento destes SE envolveu muitas pessoas e recursos computacionais. Em seus esforços para reduzir os custos de desenvolvimento, pesquisadores discutiram a idéia do ESSs. Um ESS é simplesmente um esqueleto de um sistema especialista com todos os componentes de um sistema exceto o conhecimento de uma dada área específica.

Por exemplo, a shell EMYCIN é o SE MYCIN sem seu conhecimento. O EMYCIN foi usado com sucesso no desenvolvimento do SE PUFF que fornece consultas diagnósticas no domínio de doença nos pulmões. Foi desenvolvido simplesmente pelo povoamento da base de conhecimento vazia do EMYCIN com

conhecimentos em doenças nos pulmões e usando as funções do SE disponíveis do EMYCIN.

3.2.3 Redes Neurais Artificiais

Uma Rede Neuronal Artificial (RNA) é um modelo que emula uma rede neuronal biológica. Porém, uma RNA usa um conjunto muito limitado dos conceitos do sistema biológico. Os conceitos são usados para implementar softwares de processos massivamente paralelos que envolve elementos de processamentos (também chamados de neurônios artificiais) interconectados em uma arquitetura de rede. Os neurônios artificiais recebem entradas que são análogas aos impulsos que os detritos dos neurônios biológicos recebem dos outros neurônios. A saída dos neurônios artificiais corresponde aos sinais enviados de um neurônio biológico através de seu axônio. Estes sinais artificiais podem ser mudados similarmente às mudanças ocorridas nas sinapses.

O estado da arte nas RNAs baseia-se sobre o entendimento das redes neuronais biológicas. Contudo, ainda se esta longe de ter uma máquina semelhante ao cérebro. Apesar dos estudos extensivos em neurobiologia e psicologia, questões importantes permanecem sobre como o cérebro e a mente trabalham. Entretanto, pesquisas e desenvolvimentos nesta área de RNAs estão produzindo sistemas úteis e interessantes que usam algumas características do sistema biológico.

Uma RNA é composta de elementos de processamento, organizados em diferentes modos para formar a estrutura da rede:

- ◆ **Elementos de processamento:** Uma RNA é composta de neurônios artificiais (que são referidos como neurônios, nós ou nodos); estes são os elementos de processamentos. Cada um destes neurônios recebem entradas, processa as entradas, e emite uma única saída. Este processo é mostrado na Figura 3.3. Uma entrada pode ser a entrada inicial ou a saída de outro neurônio.
- ◆ **Rede:** Cada RNA é composta de uma coleção de neurônios que estão agrupados em camadas. Uma estrutura típica é mostrada na Figura 3.4. Nesta figura tem-se três camadas: entrada, intermediária (chamada camada escondida), e a saída, mas diversas camadas podem ser colocadas entre a camada de entrada e de saída.

- ◆ **Estrutura da rede:** Similar as redes biológicas, uma RNA pode ser organizada de vários modos diferentes (topologias); isto é, os neurônios podem ser conectados de vários modos distintos. Portanto, as RNAs aparecem em muitas configurações. No processamento da informação, muitos dos elementos de processamento executam seus cálculos ao mesmo tempo. Este processamento paralelo assemelha-se ao modo que o cérebro trabalha, e difere do processamento serial dos cálculos tradicionais.

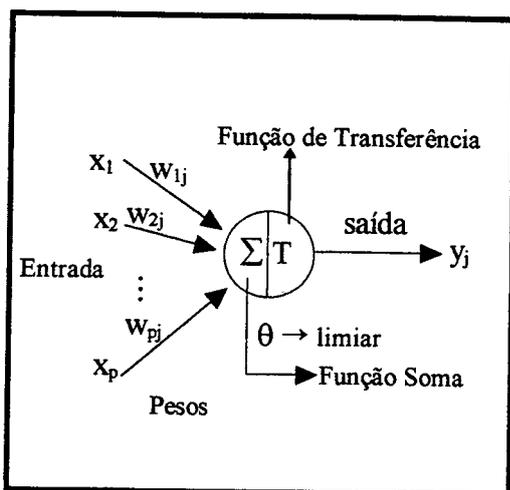


Figura 3.3: Modelo de um neurônio artificial.

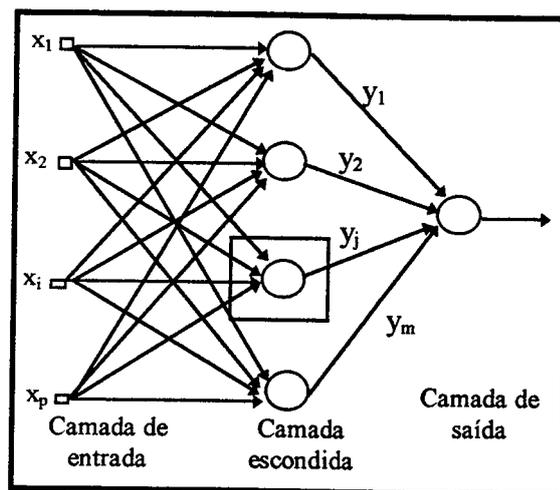


Figura 3.4: Perceptron de múltiplas camadas com uma camada escondida e um neurônio de saída.

Uma vez que a estrutura da rede está determinada, informações podem ser processadas. Os principais conceitos relacionados ao processamento (Figura 3.3) são:

- ◆ **entradas:** Cada entrada corresponde a um simples atributo. Por exemplo, se o problema é decidir sobre a aprovação ou desaprovação de um empréstimo, um atributo pode ser o nível de renda ou a idade. O valor numérico de um atributo é a entrada da rede. Diversos tipos de dados podem ser usados como entradas (valores dos pixels de caracteres ou outros gráficos, padrões digitalizados de imagens ou voz, dados codificados,...). Contudo, o pré-processamento muitas vezes é necessário.
- ◆ **Saídas:** A saída da rede é a solução do problema. Por exemplo, no caso do empréstimo, a saída pode ser “sim” ou “não”. A RNA atribui valores numéricos,

por exemplo, 1 para “sim” e 0 para “não”. O objetivo da rede é computar os valores da saída.

- ◆ **Pesos:** Um elemento chave em uma RNA é o **peso**. Os pesos expressam potências relativas (ou valor matemático) dos dados de entradas ou das várias conexões que transferem dados de uma camada para outra. Em outras palavras, os pesos expressam a importância relativa de cada entrada para um elemento de processamento. Os pesos são cruciais e é através do ajustamento repetido dos pesos que a rede “aprende”.
- ◆ **Função Soma:** A função soma obtém a soma dos pesos de todos os elementos que entram em cada elemento de processamento. Esta função soma fornece o somatório de todos os produtos de cada valor de entrada (x_i) pelo seus pesos (w_i). A fórmula para p entradas no k -ésimo elemento de processamento (neurônios) é:

$$v_k = \sum_{j=1}^p x_j w_{kj} \quad (3.1)$$

- ◆ **Função de Ativação (ou Função de Transferência):** A função soma calcula a estimulação interna, ou o nível de ativação, de um neurônio. Baseado neste nível, o neurônio pode ou não produzir uma saída. A relação entre o nível de ativação interno e a saída pode ser linear ou não linear. Tais relações são expressadas pela **Função de Ativação**. Existem diversos tipos diferentes de funções de transferência que são utilizadas e uma das mais utilizadas é a chamada **função sigmoide**:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (3.2)$$

onde a é um parâmetro.

O objetivo desta transformação é modificar o nível de saída para valores razoáveis (por exemplo, entre 0 e 1). Esta transformação é feita para limitar a amplitude da saída de um neurônio (este valor poderia ser muito grande

principalmente quando se tem muitas entradas). Algumas vezes, ao invés da função transferência, é usado um valor limiar. Por exemplo, qualquer valor maior ou igual a 0,5 muda para 1 e abaixo de 0,5 muda-se para 0.

Uma transformação pode ocorrer na saída de cada elemento de processamento, ou pode ser feito na saída final da rede.

Uma RNA aprende de suas experiências. O processo usual de aprendizagem envolve três tarefas (Figura 3.5) :

1. Calcula as saídas (y_k).
2. Compara as saídas com a resposta desejada (d_k).
3. Ajusta os pesos e repete o processo.

O processo de aprendizagem começa pela declaração dos pesos através de alguma regra ou randomicamente. A diferença entre a saída desejada (d_k) e a saída da rede (y_k) é chamada de erro e_k ($e_k = d_k - y_k$). O objetivo da rede é minimizar este erro. A redução do erro é feita através de mudanças nos pesos. A questão é mudar os pesos na direção correta, isto é, fazer mudanças que promovam a redução do erro. Durante o estágio de aprendizagem, os pesos das interconexões mudam em resposta ao conjunto de treinamento apresentado ao sistema.

Existem diversas formas de calcular o erro, dependendo do **algoritmo de aprendizagem** utilizado. Existem muitos algoritmos de aprendizagem, mas pode-se dividi-los em duas categorias: Aprendizagem supervisionada e não supervisionada. Na aprendizagem supervisionada é fornecido à rede a resposta desejada enquanto na aprendizagem não supervisionada não se conhece a resposta desejada. Neste último só fornece-se a rede os dados de entrada. Um algoritmo de aprendizado muito usado é o Algoritmo *Back-Propagation*. Um padrão de entrada é apresentado na camada de entrada e é propagado através das camadas intermediárias, passando por todos os processos, para produzir uma saída. Esta saída é então comparada com a saída objetivo e o erro é propagado para trás (da última camada para a camada de entrada) através das camadas da rede. O erro propagado é usado para ajustar os pesos das conexões. Este processo de treinamento é então repetido com um novo par de entrada, e um novo erro é propagado para trás. Este processo é repetido muitas vezes com

muitos pares de amostras de padrões até que o erro seja minimizado. Terminado este processo a rede esta treinada.

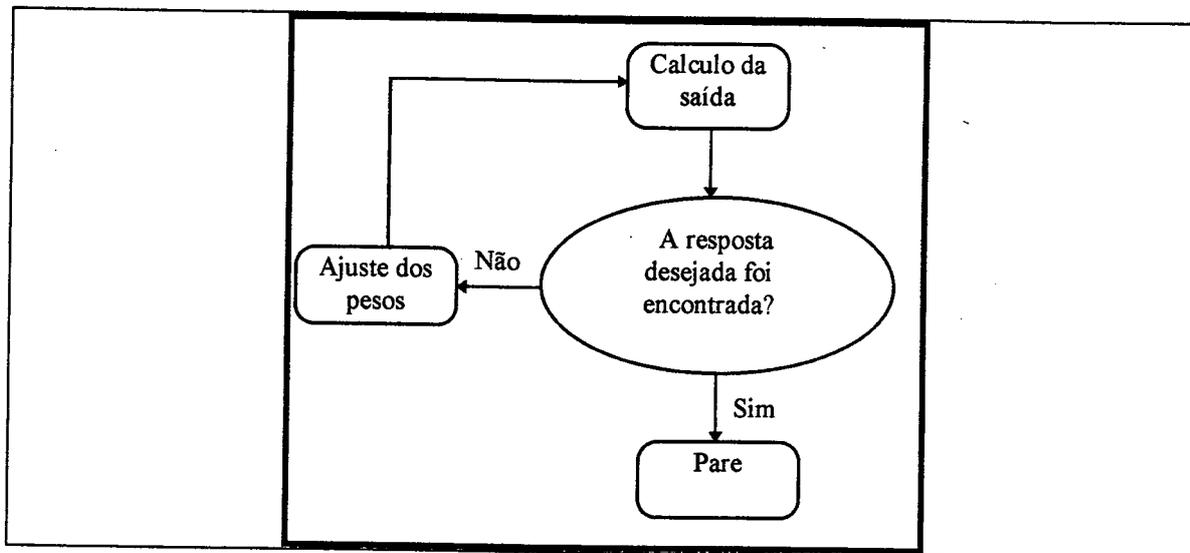


Figura 3.5: Processo de aprendizagem de uma RNA.

Contudo, se quando qualquer técnica de aprendizado, deve-se ter muito cuidado para que a rede não aprenda padrões muito específicos em seu treinamento (super-treinamento). As relações de aprendizagem deveriam ser verdadeiramente representativas das tarefas e não meramente refletir as propriedades contidas nos dados de treinamento que podem ser estatisticamente não representativos. Se uma RNA for submetida a um super-treinamento, ela somente seria capaz de aprender a reconhecer os padrões no conjunto de dados - ela não será capaz de reconhecer padrões fora do conjunto de treinamento. Isto significa que ela não teria flexibilidade ou capacidade de generalizações que os problemas exigem. A fim de evitar esta situação todas as RNA deveriam validar completamente os dados que estão fora do conjunto de treinamento. Existem diversos métodos para determinar quando um sistema de aprendizado tem um nível "correto" de treinamento [Haykin, 1994].

Vantagens e Limitações das RNA

As RNA fornecem uma forma relativamente fácil para modelar e prever sistemas não lineares. Esta é portanto uma vantagem sobre muitos métodos estatísticos que são comumente usados em negócios e finanças que são primariamente

lineares. Elas também são muito eficientes na aprendizagem de padrões em dados com ruído, incompletos e que podem conter dados contraditórios. A habilidade de aprender e a capacidade para tratar com dados imprecisos torna as RNA muito úteis e eficientes no processamento de informações de negócios.

A principal limitação das RNA é que lhes faltam a capacidade de explicação. Elas não fornecem ao usuário detalhes de como as mesmas raciocinaram com os dados para chegar as conclusões. As RNAs são portanto melhores para aplicações que requerem reconhecimento de padrões com ruído, dados incompletos e para tarefas onde especialistas são ineficazes ou onde regras claras não podem ser facilmente formuláveis. Elas não são adequadas para aplicações onde a explicação do raciocínio é crucial.

Aplicações de Redes Neurais Artificiais

As RNAs podem ser úteis em modelos estruturados de estatística, problemas de otimização e gerenciamento. Elas podem ajudar a resolver muitos problemas difíceis de otimização, de locação que não são solúveis com modelos padrões. Os exemplos incluem:

Reconhecimento de Padrões (gerar fala, transmissão de dados, detecção de bombas, reconhecimento de caracteres, sistema de reconhecimento de voz, ...), interpretação de dados (análise financeira, avaliação de empréstimos, diagnósticos médicos,...) e otimização.

Existem hoje muitos sistemas de RNA em setores de finanças. Aqui descreve-se alguns exemplos representativos:

- ◆ aprovação de crédito [Klymasauskas, 1991];
- ◆ previsão de falência [Odom & Shardo, 1993];
- ◆ predição do mercado de ações [Kimoto et al., 1990];
- ◆ previsão de índices da corporate bonds [Dutta & Shekhar, 1988].

Breve Histórico sobre as RNA

As pesquisas em RNA passou por três períodos de extensa atividade. O primeiro pico nos anos quarenta foi devido a McCulloch e Pitts [McCulloch & Pitts,

1943], os pioneiros neste trabalho. O segundo ocorreu nos anos sessenta com o teorema da convergência do perceptron de Rosenblatt [Rosenblatt, 1962] e com o trabalho de Minsky e Paper [Minsky e Papert, 1969] mostrando as limitações do perceptron simples. Os resultados de Minsky e Papert abalaram o entusiasmo da maioria dos pesquisadores, especialmente os da comunidade das ciências computacionais. O resultado acalmou as pesquisas sobre RNA por quase 20 anos. Desde o início dos anos oitenta, as RNA têm recebido considerável interesse. O maior desenvolvimento deste ressurgimento inclui a abordagem de energia de Hopfield [Hopfield, 1982] em 1982 e o algoritmo de aprendizagem back-propagation para o perceptron multicamadas, primeiro proposto por Werbos [Werbos, 1982], reinventado diversas vezes, e então popularizado por Rumelhart et al. em 1986 [Rumelhart et al., 1986].

3.2.4 Sistemas Difusos

Os sistemas difusos são baseados na teoria dos conjuntos Difusos iniciada por Lolfi Zadeh em 1965 [Zadeh, 1965]. Um objetivo desta abordagem é imitar os aspectos da cognição humana que pode ser chamado de raciocínio aproximado. Os Sistemas Difusos podem ser menos precisos que os sistemas convencionais mas são mais semelhantes às experiências do dia-a-dia, como a tomada de decisão humana. A tendência é falar em termos difusos como “alto” e “raramente”. Estes termos não são precisos, mas eles são significativos e permitem que as pessoas descrevam situações do mundo real e raciocinem sobre eles.

A Lógica Difusa [Cox, 1994; Zadeh, 1984; Kosko, 1992] é designada para manusear com conceitos lingüísticos imprecisos tal como pequeno, grande, alto, velho ou jovem. Os sistemas baseados em lógica difusa exibem uma flexibilidade inerente e têm provado ser úteis em uma grande variedade de tarefas de controle industrial e tarefas de reconhecimento de padrões, abrangendo desde reconhecimento de manuscritos à avaliação de créditos. Existem hoje diversos produtos incluindo máquinas de lavar, fornos micro-ondas e câmaras com foco automático que usam a lógica difusa no controle de seus mecanismos. O Japão é o país que está conduzindo as aplicações de técnicas de lógica difusa.

A idéia central da flexibilidade que a lógica difusa fornece é a noção de um conjunto difuso. Na teoria dos conjuntos convencionais um elemento x pertence ou não a um determinado conjunto. A noção de pertinência é dicotômica. Um valor de pertinência é 0 ou 1. Por exemplo, em uma aplicação de mercado pode-se classificar o preço de um produto dizendo que o preço de uma ação é barato ou caro.

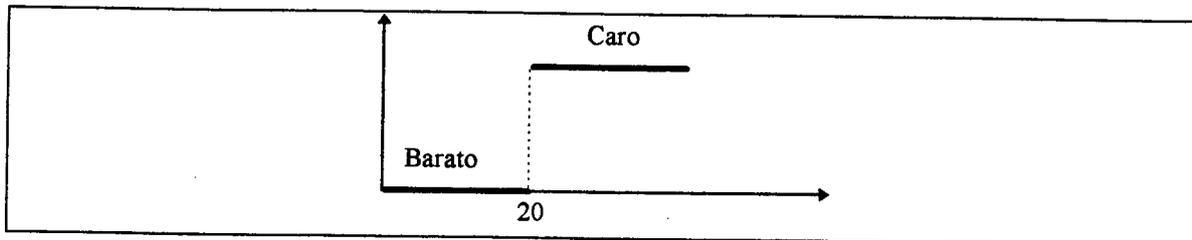


Figura 3.6: Interpretação para barato e caro na Lógica tradicional.

Como pode ser visto na Figura 3.6 existe uma mudança brusca da condição barato para a condição caro quando o preço é de 20. Em outras palavras, quando o preço for 19,5 é considerado barato e quando for 20,5 é considerado caro, embora exista uma diferença de apenas 1 nos dois preços. Contudo, tal corte não tem correspondência com a vida diária, do qual considera-se ser barato ou caro. Se, considera-se 19 ser barato, então 21 pode ser barato também, mas com grau de "barateza" menor. É esta variação no grau de medida da linguagem tal como "barato", "caro", "baixo", "alto" que a lógica difusa tenta capturar.

Uma representação mais realística do fenômeno dos preços é de um conjunto difuso onde existe um declínio gradual no grau do conjunto barato e um aumento gradual no conjunto caro quando os preços aumentam. Isto é representado na diagrama de funções de pertinências mostrado na Figura 3.7.

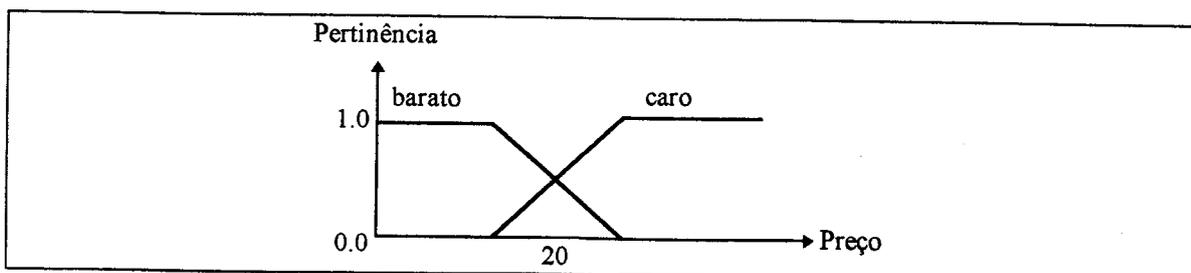


Figura 3.7: Interpretação para barato e caro na Lógica Difusa.

O intervalo dos valores de uma variável (preço da ação) representado ao longo do eixo x na Figura 3.7 é referido como o universo de discurso X . O valor da função de pertinência, que é o grau para qual um preço particular pertence aos dois conjuntos difusos é representado no eixo vertical e é representado por $\mu_X(x)$. O intervalo dos valores de pertinência geralmente é $[0, 1]$ (intervalo normalizado).

Genericamente tem-se: Dado um espaço de objetos (elementos) X , um conjunto A difuso é um conjunto de pares ordenados definidos por :

$$A = \{ (x, \mu_A(x)) \mid x \in X \wedge \mu_A(x) \in [0,1] \} \quad (3.3)$$

onde $\mu_A(x)$ é chamado o grau de pertinência de x pertencer a X .

Todo conjunto difuso estabelece um mapeamento de um universo de discurso X ao intervalo $[0,1]$, chamada função de pertinência (μ_A). Este mapeamento representa a noção de elemento pertencer parcialmente em uma classe.

Os dados que foram convertidos pelas funções de pertinências são referidos como *fuzificados*.

Usualmente as formas e o universo da função de pertinência são definidas por um especialista humano. Na Tabela 3.1 apresenta-se algumas funções de pertinências usadas pelos especialista.

Função	Fórmula
triangular	$\mu_A(x) = \max[\min\{(x-a)/(b-a);(c-x)/(c-b)\};0]$
trapezoidal	$\mu_A(x) = \max[\min\{(x-a)/(b-a);(d-x)/(d-c)\};0]$
sigmoidal	$\mu_A(x) = (1 - \exp(-a(x-c)))^{-1}$

Tabela 3.1: Funções de pertinências mais comuns.

As operações sobre conjuntos difusos são importantes para sistemas especialistas pois o processo de inferência é baseado no processamento de conectivos lógicos que têm quantificadores equivalentes à Teoria dos Conjuntos Clássicos. Por exemplo, dado a regra: Se $(A$ e $B)$ então C , onde A , B e C são conjuntos difusos, o valor verdade de C é o valor verdade da conjunção $(A$ e $B)$, ou seja, $(A \cap B)$. Pela

mesma razão, operações sobre conjuntos difusos são cruciais para inferências em sistemas especialistas difusos. As três operações básicas união, intersecção e complemento difusos foram originalmente definidos por Zadeh [Zadeh, 1965]:

$$\text{Uni\~{a}o: } A \text{ ou } B \Leftrightarrow A \cup B = \{ x, \max(\mu_A(x), \mu_B(x)) \} \quad (3.4)$$

$$\text{Intersec\~{c}o\~{a}: } A \text{ e } B \Leftrightarrow A \cap B = \{ x, \min(\mu_A(x), \mu_B(x)) \} \quad (3.5)$$

$$\text{Complemento: } \text{n\~{a}o}(A) \Leftrightarrow \neg A = \{ x, 1 - \mu_A(x) \} \quad (3.6)$$

Estas opera\~{c}o\~{e}s s\~{a}o chamadas de *opera\~{c}o\~{e}s difusas padr\~{o}es* [Klir & Yuan, 1995]. Elas representam precisamente as opera\~{c}o\~{e}s correspondentes para conjuntos cl\~{a}ssicos quando o intervalo dos graus de pertin\~{e}ncia \u00e9 o conjunto $\{0, 1\}$. Isto \u00e9, estas opera\~{c}o\~{e}s s\~{a}o generaliza\~{c}o\~{e}s correspondentes \u00e0s opera\~{c}o\~{e}s sobre conjuntos cl\~{a}ssicos. No entanto, as opera\~{c}o\~{e}s de uni\~{a}o, intersec\~{c}o\~{a} e complemento n\~{a}o s\~{a}o \u00fanicas como nos conjuntos cl\~{a}ssicos, e existem outras classes de opera\~{c}o\~{e}s [Klir & Yuan, 1995] que tamb\~{e}m s\~{a}o generaliza\~{c}o\~{e}s das opera\~{c}o\~{e}s cl\~{a}ssicas. As opera\~{c}o\~{e}s de uni\~{a}o e intersec\~{c}o\~{a} difusas s\~{a}o usualmente referidas na literatura como t-norma e t-conorma, respectivamente. J\~{a} que, estas opera\~{c}o\~{e}s difusas n\~{a}o s\~{a}o \u00fanicas, diferentes fun\~{c}o\~{e}s podem ser apropriadas para representar estas opera\~{c}o\~{e}s em diferentes contextos. A capacidade de determinar as fun\~{c}o\~{e}s de pertin\~{e}ncias apropriadas e tamb\~{e}m as opera\~{c}o\~{e}s em um contexto particular \u00e9 crucial para o sucesso da Teoria dos Conjuntos Difusos. Na grande maioria das aplica\~{c}o\~{e}s os autores usam as opera\~{c}o\~{e}s padr\~{o}es.

Regras Difusas

Uma regra difusa SE - ENT\~{A}O (ou implica\~{c}o\~{a}o difusa) assume a forma:

$$\text{“SE } x \text{ \u00e9 } A \text{ ent\~{a}o } y \text{ \u00e9 } B\text{”} \quad (3.7)$$

onde A e B são valores lingüísticos definidos por conjuntos difusos sobre um universo de discurso X e Y, respectivamente. Frequentemente, “x é A” é chamado de premissa enquanto “y é B” é chamado de conseqüente (ou conclusão) da regra.

Antes de poder empregar as regras difusas SE - ENTÃO para modelar e analisar um sistema, primeiro tem-se que formalizar qual o significado da expressão “SE x é A então y é B”, que algumas vezes é abreviada por $A \rightarrow B$. Em essência, a expressão descreve uma relação entre duas variáveis x e y; isto sugere que a regra difusa SE - ENTÃO seja definida como uma relação binária difusa R sobre o produto cartesiano no espaço $X \times Y$. Note que R é uma extensão do produto cartesiano clássico, onde cada elemento $(x, y) \in X \times Y$ está associado com o grau de pertinência denotado por $\mu_R(x, y)$. Alternativamente, uma relação difusa binária R pode ser vista como um conjunto difuso com universo de discurso $X \times Y$, e seu conjunto difuso é caracterizado pela função de pertinência bi-dimensional

Geralmente, para calcular $\mu_R(x, y)$ usa-se a fórmula do Modus Ponens Generalizado:

$$\mu_R(x, y) = \sup\{c / \mu_A(x) * c \leq \mu_B(y) \text{ e } 0 \leq c \leq 1\} \quad (3.8)$$

onde $R = A \rightarrow B$, * é um operador t-norma.

Raciocínio Difuso (ou Raciocínio Aproximado)

O Raciocínio Difuso também conhecido como Raciocínio Aproximado é um procedimento de inferência usado para derivar conclusões de um conjunto de regras difusas SE - ENTÃO em uma ou mais condições.

O raciocínio difuso de inferência é uma generalização da seguinte noção familiar. Suponha que tem-se uma curva $y = f(x)$ que relaciona x com y. Quando, dado $x = a$ então de $y = f(x)$ pode-se inferir que $y = b = f(a)$, ver Figura 3.8.

Uma generalização do processo acima permite que ‘a’ seja um intervalo e $f(x)$ uma função intervalar, como mostra a Figura 3.9. Para encontrar o intervalo resultante $y=b$ correspondente ao intervalo $x=a$, primeiro constrói-se uma extensão cilíndrica de

'a' (isto é, estende-se o domínio de 'a' para $X \times Y$) e então encontra-se sua intersecção I com a curva intervalar. A projeção de I sobre o eixo y produz o intervalo $y = b$.

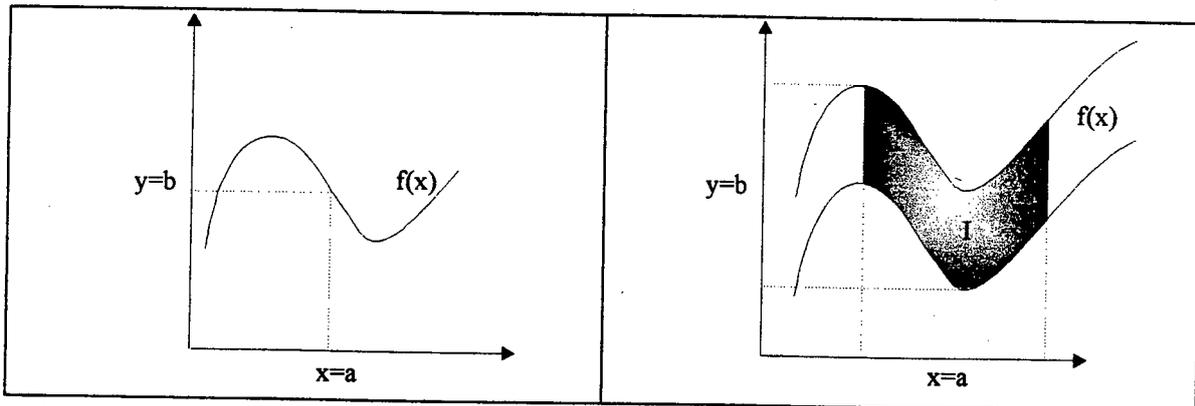


Figura 3.8: Derivação de $y=b$ a partir de $x=a$ e $y=f(x)$, onde a e b são pontos e $f(x)$ é uma curva [Jang & Sun, 1995].

Figura 3.9: Derivação de $y=b$ a partir de $x=a$ e $y=f(x)$, onde a e b são intervalos e $f(x)$ é uma função intervalar [Jang & Sun, 1995].

Seguindo um passo a mais na generalização, assumi-se que A é um conjunto difuso de X e F é um conjunto difuso da relação $X \times Y$, como mostrado na Figura 3.10 e na Figura 3.11. Para encontrar o conjunto resultante B , novamente constrói-se a extensão cilíndrica $cilin(A)$ com base A (isto é, expande-se o domínio de A de X para $X \times Y$ para obter $cilin(A)$). A intersecção de $cilin(A)$ e F (Figura 3.12) forma uma região de intersecção análoga a região I da Figura 3.9. Pela projeção de $cilin(A) \cap F$ sobre o eixo y , infere-se y como um conjunto difuso B sobre o eixo y , como mostra a Figura 3.13.

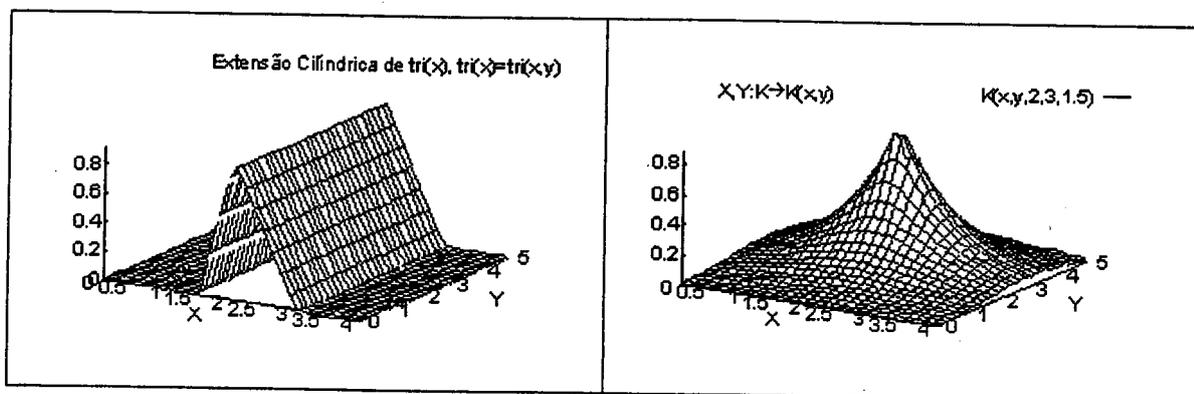


Figura 3.10: Extensão cilíndrica A .

Figura 3.11: Relação difusa F sobre x e y .

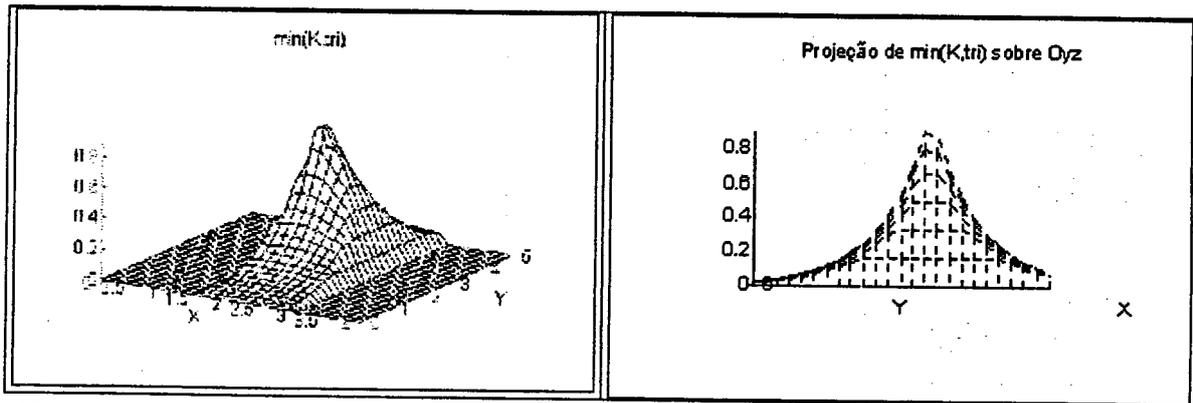


Figura 3.12: Operação min.

Figura 3.13: Projeção sobre o eixo y.

Especificamente, seja μ_A , μ_{Cilin} , μ_B e μ_F as funções de pertinências de A, *cilin*(A), B e F, respectivamente, onde $\mu_{Cilin}(A)$ está relacionado com μ_A através

$$\mu_{Cilin}(A)(x,y) = \mu_A(x). \tag{3.9}$$

Então:

$$\begin{aligned} \mu_{Cilin}(A) \cap F(x,y) &= \min[\mu_{Cilin}(A)(x,y), \mu_F(x,y)] \\ &= \min[\mu_A(x), \mu_F(x,y)]. \end{aligned} \tag{3.10}$$

Pela projeção *cilin*(A) \cap F sobre o eixo y, tem-se:

$$\mu_B(y) = \max \min[\mu_A(x), \mu_F(x,y)]. \tag{3.11}$$

Esta fórmula é referida na literatura como composição max min e B é representado como:

$$B = A \circ F \tag{3.12}$$

onde \circ denota o operador de composição. Se for escolhido o produto para t-norma (E) e max para t-conorma (OU), então tem-se a composição max-produto e $\mu_B(y)$ será dado por:

$$\mu_B(y) = \max [\mu_A(x) \mu_F(x,y)]. \quad (3.13)$$

Usando a regra composicional de inferência, pode-se formalizar um procedimento de inferência, chamado Raciocínio Difuso, sobre um conjunto de regras difusas. A regra básica de inferência na lógica tradicional (bi-valorada) é o Modus Ponens, de acordo com que se pode inferir a verdade da proposição B sabendo que A e a implicação $A \rightarrow B$ são verdades. Este conceito é ilustrado como:

$$\begin{array}{ll} \text{premissa 1 (fato):} & x \text{ é } A \\ \text{premissa 2 (regra):} & \underline{\text{se } x \text{ é } A \text{ então } y \text{ é } B} \\ \text{conseqüente (conclusão):} & y \text{ é } B \end{array} \quad (3.14)$$

Contudo, na maioria dos raciocínios humanos, o Modus Ponens é empregado de uma maneira aproximada. Por exemplo, tendo-se a regra de implicação “Se a criança está pálida então a criança está com anemia” e sabe-se que a “criança está quase pálida” então pode-se inferir que a “criança está quase com anemia”. Isto é escrito como :

$$\begin{array}{ll} \text{premissa 1 (fato):} & x \text{ é } A' \\ \text{premissa 2 (regra):} & \underline{\text{se } x \text{ é } A \text{ então } y \text{ é } B} \\ \text{conseqüente (conclusão):} & y \text{ é } B' \end{array} \quad (3.15)$$

onde A' está próximo de A e B' está próximo de B . Quando A , A' , B e B' são conjuntos difusos de universos apropriados, o procedimento de inferência acima é chamado de Raciocínio Difuso ou Raciocínio Aproximado (também chamado de Modus Ponens generalizado desde que se tenha o Modus Ponens como caso especial).

Usando a regra de composição de inferência introduzido inicialmente pode-se formular o processo de inferência do Raciocínio Difuso com a seguinte definição.

Definição: *Raciocínio Difuso Baseado sobre a composição Max-Min:* Seja A , A' e B conjuntos Difusos de X , X' e Y respectivamente. Assume-se que a implicação difusa

$A \rightarrow B$ seja expressada como uma relação difusa R sobre $X \times Y$. Então o conjunto Difuso B' induzido por “ x é A' ” e a regra difusa “se x é A então y é B ” é definido por

$$\mu_{B'}(y) = \max \min[\mu_{A'}(x), \mu_R(x,y)] \quad (3.16)$$

ou equivalentemente

$$B' = A' \circ R = A' \circ (A \rightarrow B) \quad (3.17)$$

Relembrando que (3.17) é uma expressão genérica para o Raciocínio Difuso, enquanto (3.16) é um exemplo de Raciocínio Difuso onde o max e min são os operadores difusos para a t-conorma e t-norma, respectivamente.

Agora pode-se usar o processo de inferência do Modus Ponens Generalizado para derivar as conclusões, desde que a implicação difusa $A \rightarrow B$ seja definida como uma relação difusa binária apropriada.

Vantagens e Limitações dos Sistemas Difusos

Um dos principais pontos fortes da lógica difusa comparado com outros métodos é a forma de tratar dados imprecisos. Sua base de conhecimento, que é em forma de regras, é facilmente examinada e entendida. O formato das regras também torna fácil a atualização e manutenção da base de conhecimento.

Como limitação da lógica difusa e principal deficiência é que as funções de pertinência e as regras tem que ser especificadas manualmente. A determinação das funções de pertinências pode consumir algum tempo num processo de tentativa e erro. Além disto, a elicitação das regras de especialistas humanos pode ser um processo caro e propenso a erro. E mais, elas não podem adaptar-se automaticamente para as mudanças no meio em que se está operando. Novas regras tem que ser manualmente alteradas se as condições do problema mudarem.

Aplicações

As áreas nas quais a lógica difusa foi aplicada com êxito são freqüentemente bem concretas. A primeira aplicação comercial importante foi na área de controle de fornos de cimento. É uma operação que requer que um operador controle quatro estados internos do forno, controle quatro jogos de operações, e dinamicamente administre 40 ou 50 "regras thumb" sobre sua inter-relação, tudo com a meta de controlar um conjunto altamente complexo de interações químicas. Um relato completo deste sistema, muito bem sucedido, pode ser encontrado em Umbers e King [Umbers e King, 1980].

A lógica difusa pode ser aplicada em muitas áreas como classificação, tomadas de decisões e reconhecimento de padrões. Mas, a lógica difusa tem um valor particular nas aplicações de controle onde é difícil ou impossível desenvolver um controlador tradicional. No momento, o Japão está sendo o líder no desenvolvimento de sistemas de controle lógico-difuso para diversas aplicações como ar condicionados, câmaras, máquinas de lavar, micro-ondas e sistemas de estradas de ferro [Self, 1990] e [Waller, 1989].

A MIRC (Market Intelligence Research Corporation) espera que a renda total para produtos baseados em lógica difusa combinados com redes neuronais cresça para 10 bilhões até 1998 [Kandel, 1993]. Maiers e Sherif [Maiers e Sherif, 1985] forneceram 450 referências de aplicações de lógica difusa e teoria. Levy [Levy et al., 1991] descreve uma avaliação de aplicações para empréstimos comerciais.

Os sistemas especialistas foram os recipientes bem óbvios dos benefícios de lógicas difusas, desde que seu domínio é freqüentemente inerentemente difuso. Exemplos de sistemas especialistas com lógica difusa são os sistemas de apoio à decisão, planejadores financeiros, sistemas de diagnósticos para determinar patologia de soja, e um sistema especialista meteorológico na China para determinar áreas que devem plantar pomares de seringueiras [Zadeh, 1984]. Outra área de aplicação, semelhante aos sistemas especialistas, é a recuperação de informação [Radecki, 1982].

Histórico

A precisão matemática deve seu êxito em grande parte aos esforços de Aristote e aos filósofos que o precederam que se esforçaram para inventar uma teoria concisa de lógica. A matemática, chamada "Leis do Pensamento" foram descritas em [Korner, 1967]. Uma destas, a "Lei do Meio Excluído", declara que qualquer proposição deve ser Verdadeira ou Falsa. Até Parmínedes propor a primeira versão desta lei (em torno de 400 A.C.) haviam objeções imediatas e fortes: por exemplo, Heraclitus propôs que coisas podiam ser simultaneamente Verdadeiras e não Verdadeiras.

Foi Plato que colocou a fundamentação para o que teria tornado-se lógica difusa, indicando que havia uma terceira região (além de Verdade e Falso) onde estas contradições deixam de valer. Outros filósofos mais modernos ecoaram seus sentimentos, destacando notavelmente Hegel, Marx, e Engels. Mas foi Lukasiewicz que primeiro propôs uma alternativa sistemática à lógica bi-valorada de Aristotle [Lejewsk, 1967].

No início dos anos de 1900, Lukasiewicz descreveu uma lógica tri-valorada. O terceiro valor proposto pode melhor ser traduzido como o termo "possível", e ele designou um valor numérico entre Verdade e Falso. Eventualmente, propôs uma notação completa e um sistema axiomático do qual ele esperava derivar a matemática moderna.

Mais tarde, ele explorou lógicas de quatro valores, lógicas de cinco valores, e então declarou que em princípio não havia nada para impedir a derivação de uma lógica infinito-valorada. Lukasiewicz percebeu que lógicas de três e infinitos valores eram mais intrigantes, mas ele finalmente determinou uma lógica quadro-valorada porque pareceu ser mais fácil para adaptar à lógica Aristoteliana.

Knuth propôs uma lógica tri-valorada semelhante a de Lukasiewicz, do qual ele especulou que aquela matemática tornar-se-ia mais elegante que a lógica bi-valorada tradicional. Sua introspecção, aparentemente perdida por Lukasiewicz, era usar o conjunto $\{1, 0, -1\}$ ao invés de $\{0, 1, 2\}$. Não obstante, esta alternativa não conseguiu ganhar aceitação e passou em obscuridade relativa.

Foi só muito recentemente que a noção de uma lógica de infinitos-valores tornou-se segura. Em 1965 Lotfi A. Zadeh publicou seu trabalho seminal "Conjuntos Difusos" [Zadeh, 1965; Zadeh, 1968] onde descreveu a matemática da teoria dos conjuntos difusos, e por extensão a lógica difusa. Esta teoria propôs uma função de pertinência (ou os valores Falso e Verdadeiro) que opera sobre o intervalo de números dos reais $[0, 1]$. Novas operações para o cálculo de lógica foram propostas, e foram mostradas ser em princípio pelo menos uma generalização de lógica clássica.

3.2.5 Algoritmos Genéticos

Os Algoritmos Genéticos [Goldberg, 1989] são basicamente técnicas de busca, que foi primeiro sugeridas por John Holland [Holland, 1975], e que se baseiam nos processos observados em evolução natural. Estes algoritmos têm sido muito usados na solução de problemas de otimização. Eles são capazes de performar uma busca robusta, altamente eficiente e rapidamente localizar soluções muito boas para problemas de buscas difíceis.

Em geral, os AGs performam buscas randômicas diretas através de um conjunto de dados de alternativas com o objetivo de encontrar a melhor alternativa com respeito a um dado critério. Estes critérios devem ser expressados em termos de uma função objetivo, que é usualmente referida como uma função de ajustamento.

O primeiro passo para aplicação de AGs a um problema é representar (codificar) soluções possíveis no espaço de busca como uma seqüência de símbolos gerados a partir de um dado alfabeto finito A . Geralmente se usa o alfabeto binário $A=\{0,1\}$. Cada seqüência corresponde a um cromossomo, e cada elemento da seqüência é equivalente a um gene. O conjunto de cromossomos é chamado população.

Por exemplo, dado um conjunto de pontos (x,y) pode-se querer ajustá-los a uma reta. Para isto, codifica-se o conjunto de parâmetros para uma reta $y = P_1x + P_2$, para criar seqüências de símbolos independentes para as duas constantes desconhecidas P_1 e P_2 e então concatena-se (concatenação de seqüências). As seqüências de símbolos geralmente são combinações de 0's e 1's, que representam o valor de um número na forma binária. Uma cadeia de n -bits pode acomodar todos os

inteiros até 2^n-1 . Por exemplo, o número 7 requer uma cadeia de 3 bit ($2^3-1=7$) e o número 10 (1010_2) de um cadeia de 4 bits. Esta cadeia de bits pode ser mapeada pelos valores de um parâmetro dado por:

$$P_i = P_{\min} + \frac{b}{2^L} (P_{\max} - P_{\min}) \quad (3.18)$$

onde b é o número decimal que está sendo representado em binário, L é o comprimento da cadeia de bits, P_{\min} e P_{\max} são constantes que dependem do problema e representam os valores máximo e mínimo que P_i pode assumir.

Após definir a codificação para os parâmetros, uma população inicial de n seqüências (cromossomos) de comprimento L é criada (de forma randômica). Cada uma das seqüências é decodificada dentro de um conjunto de parâmetros que o representa. Este conjunto de parâmetros é passado através de um modelo numérico do espaço do problema. Os modelos numéricos dão uma solução baseada no conjunto de entrada dos parâmetros. Baseado na qualidade desta solução cada cromossomo recebe um valor de ajustamento. Com estes valores de ajustamento, os operadores genéticos reprodução, crossover e mutação, são usados para gerar uma nova população de soluções, que se espera ser melhor que as gerações prévias (melhor valor de ajustamento). O novo conjunto de cromossomos é novamente decodificado e avaliado, e uma nova população é criada usando os operadores básicos dos AGs. Este processo continua até que um critério de parada seja atingido (convergência).

Entre os três operadores, a *reprodução* copia um indivíduo de uma geração para próxima; *crossover* (Figura 3.14 (A)) combina características de dois ou mais pais para produzir um ou mais filhos, *mutação* (Figura 3.14(B)) faz pequenas mudanças locais. A reprodução e o crossover de ajuste individual fornecem um impulso para melhorar, enquanto mutação (menos utilizada) mantém a população diversificada e permite uma exploração mais ampla.

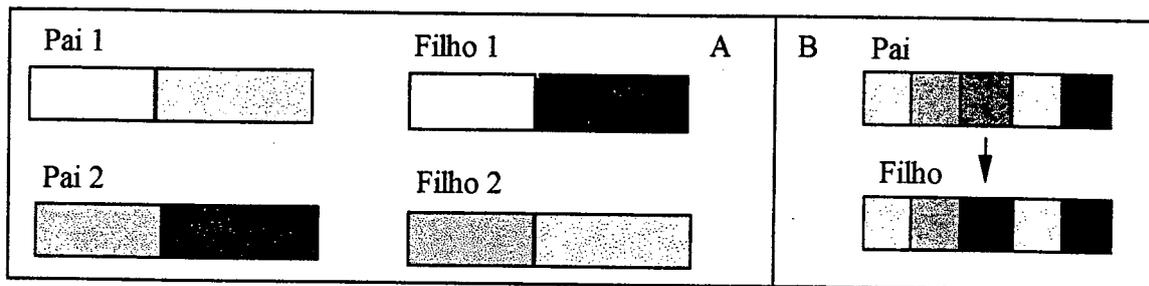


Figura 3.14: (A) Crossover entre dois cromossomos. (B) Mutação no alelo 3.

O desenvolvimento de um AG para solucionar um problema particular envolve dois tipos de decisão. A primeira concerne na forma no qual o problema será modelado dentro das exigências do AG, o que inclui a definição do espaço de soluções possíveis, a forma da função e ajustamento, e o caminho no qual os indivíduos são representados como seqüência de símbolos (cromossomos). A segunda refere-se aos parâmetros dos AGs por si só e inclui as proporções da população para ser reproduzida como um resultado de reprodução, crossover e mutação, o procedimento de seleção, o tamanho da população, o número de gerações e um número de outras decisões que dizem respeito as variantes deste algoritmo básico.

Vantagens e Limitações dos AGs

Devido à sua estrutura paralela os AGs provaram ser muito eficientes nas buscas em conjuntos de dados muito grandes [Goldberg, 1989]. Este processo de busca tem também outra vantagem sendo altamente desejável para implementação em computadores paralelos. Eles tem tido um sucesso particular em problemas de otimização incluindo *job-shop scheduling*, escala de horários e otimização de portfólio [Davis, 1991; Deboeck, 1994]. Os AGs podem aprender relações complexas, conjunto de dados incompletos e podem ser como as ferramentas “*data mining*” para descobrir padrões desconhecidos previamente. Eles podem adaptar mudanças no meio onde estão operando e podem fornecer explicações das decisões que produzem em um formato que os humanos possam entender.

Uma limitação dos AGs é que a declaração dos parâmetros assim como as taxas de mutação e crossover é dependente do problema e consome muito tempo no processo de tentativa-e-erro. Além disto, a performance do AG é afetada pelo

esquema de representação que é empregado. A seleção de um esquema apropriado pode requerer experimentos extensos.

Aplicações

Os AGs podem ser vistos como um tipo de máquina de aprendizagem para automaticamente resolver problemas complexos. Eles fornecem um conjunto de heurísticas de busca, independente do domínio e eficientes para um amplo espectro de aplicações. Austin [Austin, 1990] tem indicado algumas áreas de aplicações gerais como controle de processo dinâmico, indução de otimização de regras, descobrimento de novas topologias conexionistas, simulação de modelos biológicos de comportamento e evolução, projeto complexo de estruturas de engenharia e reconhecimento de padrões.

O AG é uma ferramenta muito interessante. Ela rejeita soluções inferiores e acumula as boas. Também, os AGs são desejáveis para processamento paralelo [Austin, 1990]. Finalmente, os AGs são também usados em lógica difusa [Karr, 1991].

As aplicações em negócios está crescendo rapidamente em aplicações com sucesso em comércio financeiro, avaliação de crédito e detecção de fraude. Eles estão também sendo usados para prever falência de corporações [Kingdon & Feldman, 1995]. Quando concede-se um empréstimo ou crédito a uma companhia, uma avaliação de sua capacidade de pagar seus débitos é uma consideração fundamental. Se existe a possibilidade da companhia estar indo a falência, então a organização que fornece o crédito pode não conceder o empréstimo ou mudar para um nível muito alto de juro para compensar o tal risco. Correntemente o método primário de predição de falência envolve técnicas de modelagem linear, principalmente análise discriminante multivariada.

Kingdon e Feldman usaram os AGs para inferir regras para previsão de falência, usando índices financeiros extraídos dos balanços das companhias. Exemplos de índices financeiros, usados no processo de aprendizagem das regras, são vendas/ativo total, ativo circulante/passivo circulante e receita/vendas. Em média o método AG foi 15% melhor em acurácia que o modelo de análise discriminante

múltipla, comumente usado, e aproximadamente comparável com a performance das RNA usada com os mesmos dados.

Packard [Packard, 1990] usa os AGs para auxiliar o governo na alocação de recursos.

Outros exemplos de aplicação são os seguintes trabalhos:

- Reconhecimento de Suspeitos Criminosos [Caldwell & Johnston, 1991];
- Composição de Música [Horner & Goldberg, 1989];
- *Scheduling* [Gabbert et al., 1989; Syswerda & Palmucci, 1991];
- Otimização de data de base *query* [Bennett et al., 1989];
- Projeto de aviões [Bramlette & Bouchard, 1991];
- Aposta em corridas de cavalos [de la Maza, 1989];

Histórico

Embora Charles Darwin tenha formulado a Teoria da Evolução no final do século passado [Darwin, 1985], foi só recentemente que se tentou idealizar um modelo matemático do processo evolutivo. Nos anos 60, John Holland, da Universidade de Michigan, começou a definir as bases de algoritmos de otimização de inspiração genética. Seu trabalho culminou na publicação do livro *Adaptação em Sistemas Naturais e Artificiais* em 1975 [Holland, 1975]. O livro, que é hoje muito citado, mas pouquíssimo lido, foi pouco divulgado na época, em grande parte ao estilo pesado com notação muito complexa. Felizmente Holland e seus muitos discípulos, quase todos seus alunos de pós graduação, continuaram sua linha de investigação, publicando resultados com alguma timidez mas com perseverança.

A grande popularidade que os AGs atingiram recentemente se deve a dois importantes fatores: a publicação de um livro tutorial sobre AGs [Goldberg, 1989] por um dos alunos de doutorado de Holland, David Goldberg, um pesquisador extremamente ativo e com excelente potencial didático, e às conferências Internacionais sobre Algoritmos Genéticos (IC-GAs), que têm se realizado a cada dois anos nos Estados Unidos, desde 1985.

3.2.6 Raciocínio Baseado em Casos

O Raciocínio Baseado em Casos (RBC) é uma abordagem recente da IA que tem recebido muita atenção nos últimos anos. Seu objetivo principal é buscar uma solução semelhante, para um problema atual, através do estabelecimento de graus de similaridade, com uma experiência passada, armazenada na base de casos [Kolodner, 1992].

O RBC é um conjunto de métodos e técnicas usadas para representar a forma como os homens raciocinam. O RBC modela a habilidade humana de lembrar e adaptar um fato passado para resolver um problema novo, semelhante de alguma forma a esta experiência passada.

O RBC é um paradigma para resolver problemas que em muitos aspectos são fundamentalmente diferente das outras técnicas de IA. Ao invés de contar somente com o conhecimento geral de um problema domínio, ou fazer associações junto com relações generalizadas entre o problema descrito e conclusões, o RBC é capaz de usar o conhecimento específico de experiências prévias, situações de problemas concretos (casos). Um novo problema é resolvido encontrando um caso passado mais similar, e reusando na situação do novo problema. Uma segunda diferença importante é que o RBC é uma abordagem com aprendizagem com incremento contínuo, desde que uma experiência nova pode ser retida na base de casos cada vez que um problema foi resolvido, tornando imediatamente disponível para problemas futuros.

Mas o que é um caso? De uma forma mais simples, é a representação de um fato ou de uma experiência através de suas características (atributos). Estes atributos devem descrever não apenas o conteúdo da experiência, como também, o contexto em que esta se passou. Alguns exemplos incluem as informações sobre uma forma de empréstimo-pessoal e se o empréstimo foi concedido ou não, e o histórico de uma empresa com o diagnóstico associado. Os casos formam a base de casos.

Um novo problema é resolvido seguindo os seguintes passos:

- **recuperação** de um ou mais casos previamente armazenado na base de casos;
- **seleção** de um ou mais casos dentre os casos recuperados;
- **revisão** deste(s) caso(s) para determinar a necessidade de adaptação;

- **reutilização** do caso adaptado para resolver o problema de entrada;
- **avaliação** da solução do problema de entrada;
- **inclusão** do caso adaptado na base de casos (aprendizagem).

O desenvolvimento do CBR é uma tarefa complexa que está formulada em quatro etapas distintas [Weber-Lee, 1996] e [Aamodt, 1991].:

- **representação** dos casos
- **recuperação** do(s) caso(s) mais similares
- **adaptação** dos casos
- **reutilização, avaliação e aprendizagem;**

Alguns autores apresentam estas etapas com diferentes enfoques. Algumas etapas são difíceis de serem apresentadas isoladamente, daí a razão do agrupamento de algumas delas, em face dos tipos de sistemas implementados.

Representação dos Casos

O problema da representação dos casos representa um dos primeiros tópicos a serem cuidadosamente estudados, na medida em que, uma ou outra orientação na abstração do caso, pode definir o sucesso ou fracasso do sistema a ser implementado.

A representação da base de conhecimento num sistema de CBR consiste em modelar os casos e definir o estilo de memória que organize estes casos. A execução destas etapas estão fortemente ligadas entre si e estas, por sua vez, dependem fortemente das outras etapas do sistema. As outras etapas do sistema, recuperação, adaptação e aprendizagem devem ser projetadas simultaneamente à representação.

O conhecimento no CBR é representado através da memória de casos. A representação do conhecimento trata de como descrever os casos e, de como organizá-los na memória. Na representação dos casos existem dois componentes básicos: a descrição do problema e da solução, como um algoritmo. Kolodner [Kolodner, 1983^a] inclui um terceiro componente, o resultado da aplicação da solução do problema. Entretanto, o resultado da solução é descritivo.

A indexação é uma questão que pode ser bastante importante conforme a estrutura e o conteúdo da memória. A memória deve ser indexada para proporcionar uma recuperação e reutilização eficientes. As características dos casos tornam-se índices que caracterizam um evento.

Recuperação dos Casos

A partir de um problema a ser resolvido [problema de entrada], a etapa de recuperação parte da identificação das características deste problema, após faz uma busca na memória de casos e, então seleciona a melhor solução, através de algoritmos que estabelecem as similaridades. As subtarefas seguem a seguinte orientação: identificação das características, *matches* iniciais, pesquisa e seleção.

A **identificação** das tarefas básicas começa num conjunto relevante de identificação, que orienta a busca da melhor solução inicial, através de uma similaridade que selecionará a melhor solução final. A tarefa de identificação é somente necessária em domínios de aplicação, onde as características não sejam diretas ou não estejam claras.

A **recuperação** é a etapa em que uma função é utilizada para recuperar os casos mais similares. Esta função pode utilizar uma métrica, ou pode ser orientada por metas ou restrições. Métodos de classificação também podem ser utilizados. Esta recuperação requer que um limiar seja estabelecido, para definir limites de casos que podem ser recuperados. Em recuperações orientadas por restrições, alguns índices podem ser mais ou menos importantes podendo até excluir casos.

A **similaridade** é a primeira questão a ser estudada na etapa de recuperação. O que faz um caso ser similar ao outro, depende do domínio do conhecimento da aplicação. Quando a recuperação é do tipo que busca a similaridade diretamente comparando os índices, uma métrica de similaridade é usada para este cálculo. A Similaridade é a essência do CBR. É em razão de haver uma experiência similar à atual na memória, que o sistema é validado.

O caso escolhido como solução para o problema de entrada é tratado na literatura como *Best Match*. Normalmente, esta seleção é uma etapa mais elaborada

do que a busca pelo conjunto de casos mais similares. A importância desta etapa reside no fato de que seu resultado é exatamente a saída do sistema.

Etapa de Adaptação

Assim que o *best match* é escolhido, o próximo passo é revisar a solução, para verificar a necessidade de adaptação em relação ao problema de entrada.

A **Adaptação** no contexto de CBR, significa modificar um caso para solucionar um problema de entrada. A adaptação avalia as diferenças entre o problema escolhido e, o problema de entrada. Várias técnicas de IA podem ser consideradas para desempenhar a função de adaptação. O estabelecimento de regras representa uma maneira simples e satisfatória de ajuste.

Etapa de Reutilização, Avaliação e Aprendizagem

Reutilizar um caso, significa gerar a possibilidade de utilizá-lo para resolver o problema de entrada numa outra consulta.

Os objetivos da **avaliação** são mensurar a qualidade da solução adaptada ao problema de entrada, para definir se esta tem condições de ser adicionada à memória. Alguns sistemas sugerem que a solução que não foi bem avaliada deve possuir uma observação que permita o sistema agregar este conhecimento, para após repetidas situações, ser retirado da memória.

A **aprendizagem** consiste no caso que foi avaliado, revisado e armazenado na memória, sem restrições. A etapa de aprendizagem constitui-se num dos diferenciais do CBR em relação aos outros sistemas inteligentes, por armazenar na memória, casos que podem ser recuperados em situações futuras através de algoritmos.

Aplicações

Os problemas mais comuns para o desenvolvimento de aplicações em CBR são: interpretação, diagnóstico, análise e elaboração de projetos, formação de preços, desenvolvimento de propostas, planejamento, configuração, *scheduling*, monitoramento, *debugging*, consertos, análise situacional, classificação, instrução, aprendizagem e controle.

Na área de finanças tem-se ainda poucas aplicações, dentre elas pode-se citar:

- Previsão e Fluxo de caixa [Weeber-Lee et al., 1995];
- Análise da Saúde Financeira de Empresas [Martins et al., 1996].

Vantagens e Limitações de RBC

Uma das principais vantagens da técnica RBC, é a sua capacidade de explicar seus resultados. As justificativas são sempre consistentes com as soluções por serem as próprias experiências, representando mais um aspecto de proximidade ao comportamento humano do sistema. Além disso, as justificativas podem avisar sobre possíveis riscos que o uso de determinada abordagem pode implicar. Outra vantagem do RBC é a forma de representação do conhecimento que resume-se em escolher o tipo de estrutura da base de casos.

Uma dos principais problema de RBC é determinar quais os casos da base são “similares” ao caso corrente. A relevância das soluções da base necessitam ser organizadas na memória de modo que a descrição dos problemas de entradas possam ser recuperada eficientemente.

Histórico

As raízes do CBR na Inteligência Aplicada são oriundas dos trabalhos de Schank [Schank, 1982], sobre memória dinâmica e recuperação de situações passadas. A memória dinâmica usa uma estrutura hierárquica denominada de Pacotes de Organização de Memória [**MOP** - Memory Organization Packtes], que agrupam um conjunto de casos com características similares. Nesta estrutura, os casos são caracterizados pelos episódios aos quais estão associados e, seus atributos não são apenas nomes próprios, mas atributos das abstrações que juntas modelam o contexto do caso.

Um MOP é caracterizado por ser mutável durante sua utilização e, possuir regras através das quais, os casos possam ser indexados, através de nomes, valores e índices que diferenciam os casos indexados ao mesmo MOP.

O primeiro sistema de raciocínio baseado em casos, chamava-se CYRUS. Este sistema foi desenvolvido por Kolodner [Kolodner, 1983b], na Universidade de Yale,

no grupo de pesquisa de Schank [Schank, 1982]. Cyrus era fundamentado no modelo de memória dinâmica e, na teoria de solução de problemas e aprendizagem MOP. O sistema funcionava basicamente, com pergunta-resposta. Sua base de conhecimento era a agenda de viagens/reuniões do Secretário de Estado Norte-Americano Cyrus Vance.

O modelo de memória de casos desenvolvida para este sistema, foi utilizado mais tarde para muitos outros sistemas de CBR, como o MEDIATOR, desenvolvido por Simpson [Simpson, 1985] que tinha por objetivos mediar disputas, num sistema de planejamento e diagnóstico, PERSUADER [Sycara, 1985] utilizado para mediar negociações sindicais, CHEF [Hammond, 1989] para planejamento, JULIA [Hinrichs, 1992] para desenvolvimento de projetos e CASEY [Katon, 1989] utilizado para diagnosticar a causa e soluções para problemas cardíacos.

Porter [Porter et al, 1986] na Universidade do Texas, desenvolveu um modelo para resolver problemas de aprendizagem com máquinas, através de classificação de tarefas. Este sistema, originou o sistema PROTOS [Bareiss, 1988] que utilizava classificação heurística para diagnóstico e, enfatizava a integração geral de domínios do conhecimento, com conhecimento de casos específicos unificados numa estrutura de representação.

A combinação de casos com estruturas de domínios de conhecimento, orientou a criação do sistema GREBE [Branting, 1991], que é utilizado no domínio do Direito.

Na Europa os resultados da pesquisa sobre CBR iniciaram-se com os trabalhos de Richter [Richter, 1993] da Universidade de Kaiserslauten, que desenvolveu o sistema MOLTKE. Mais tarde ele aprimorou este sistema criando o sistema PATDEX. Lopez e Plaza [Lopez & Plaza, 1993] desenvolveram um sistema de aprendizagem baseado em casos para diagnóstico médico.

Em outros países as pesquisas sobre CBR são pontuais, conforme pode-se verificar na Índia [Ventakamaran, 1993] e no Japão [Kitano, 1993].

3.3 Sistemas Híbridos

Enquanto sistemas inteligentes estão sendo aplicados com sucesso numa variedade de tarefas, certos problemas mais complexos não podem ser resolvidos usando uma única técnica de IA. Cada técnica de IA tem seus pontos fortes e limitações particulares que as tornam mais adequadas para aplicações particulares e não para outras. Como mencionou-se previamente, enquanto as RNA são boas para Reconhecimento de Padrões, elas não são adequadas para explicar como elas alcançam seus resultados. Sistemas de Lógica Difusa explicam suas decisões mas não podem automaticamente adquirir regras que usam para tomar decisões. Estas limitações tem sido o principal motivo para a criação de Sistemas Inteligentes Híbridos onde duas ou mais técnicas são combinadas para superar as limitações de cada técnica individual [Goonatilake & Treleaven, 1995^a].

Por exemplo, quando usar um sistema difuso para avaliar um empréstimo, um especialista em avaliações de empréstimos tem que especificar todas as regras necessárias para o sistema tomar a decisão. As RNA com a sua capacidade de aprendizagem podem ser usadas para aprender estas regras de decisão difusas automaticamente, então criando um sistema híbrido que supera os limites de um sistema difuso.

A Tabela 3.1 fornece uma avaliação de algumas técnicas de IA com respeito as 5 propriedades desejáveis (aprendizagem, adaptação, flexibilidade, explicação e descobertas). Como se pode ver na Tabela 3.1 um sistema híbrido neuro-difuso pode ser visto como uma combinação de uma técnica particular (sistema difuso) com baixa capacidade em uma propriedade particular (aprendizagem) com uma técnica (RNA) com alta capacidade nessa propriedade. Outros exemplos similares de sistemas híbrido são os que usam AG para encontrar as funções de pertinência difusas [Karr, 1995], ou usam RNA como padrões de *matchers* em sistemas especialistas [Tirri, 1995] ou ainda usam AG para encontrar os pesos de uma RNA [Montana, 1995].

Tecnologia	Aprendizagem	Flexibilidade	Adaptação	Explicação	Descobertas
Redes Neurais	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓	✓	✓✓
Algoritmos Genéticos	✓✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓	✓✓✓✓✓
Sistemas Difusos	✓	✓✓✓✓✓	✓	✓✓✓	✓
Sistemas Especialistas	✓	✓	✓	✓✓✓✓✓	✓
RBC	✓✓✓✓	✓✓✓	✓✓✓	✓✓✓✓✓	✓✓

Tabela 3.1 : Comparação das Técnicas Inteligentes [Goonatilake e Treleaven, 1995^a - modificado].

Os Sistemas Híbridos são também importantes quando se considera a natureza variada das aplicações. A maioria dos domínios complexos tem muitos problemas com componentes diferentes, cada um deles podendo requerer processamentos diferentes. Por exemplo, se um problema complexo (como o problema de orçamento e planejamento) tem três componentes principais - uma tarefa de previsão, uma de otimização e outra de raciocínio em série - então uma RNA, um AG e um SE podem ser usados para resolver estes problemas componentes respectivamente. Os diferentes sistemas inteligentes componentes comunicam seus resultados entre eles para produzir o resultado final. Os Sistemas Híbridos deste tipo são referidos por Goonatilake e Khebbal [Goonatilake & Khebbal, 1995b] como híbridos por intercomunicação.

Existem diversas classificações dos sistemas híbridos. A Tabela 3.2 mostra cinco diferentes estratégias híbridas de desenvolvimento que tem sido identificadas segundo [Medsker & Bailey, 1992]: modelos independentes, transformacional, acoplamento livre, acoplamento rígido e totalmente integrado. Na Tabela 3.3 mostra-se o esquema de classificação proposto por Goonatilake e Khebbal [Goonatilake & Khebbal, 1995b], onde as três classes, substituição de funções, intercomunicação e polimorfismo, estão de acordo com a funcionalidade, arquitetura de processamento e requerimentos de comunicação.

O esquema de Medsker diferencia sistemas sobre uma base de questões de implementação, enquanto o esquema Goonatilake e Khebbal é um esquema mais genérico que diferencia sistema sobre uma base de funcionalidade [Goonatilake & Khebbal, 1992], Tabela 3.3.

Modelo	Descrição	Exemplo
Independentes 	São compostos por módulos independentes não existindo nenhuma interação entre os módulos. Não são uma proposta de integração propriamente, mas eles são propícios para comparar o desempenho de duas técnicas para uma aplicação específica.	Comparação de diferentes diagnósticos em consertos de computador.
Transformacional 	São similares ao modelo independente não havendo uma interação um com o outro. A diferença é que ele começa com um modelo e termina com o outro, um exemplo é começar com um sistema especialista e em seguida partir para uma rede neuronal ou vice versa.	Sistemas de ajuda na tomada de decisão, onde por exemplo uma rede neuronal é usada para identificar tendências e relações, e em seguida os resultados são usados em um sistema especialista para assistir a tomada de decisão.
Acoplamento Livre 	É a primeira forma verdadeira de integração. O problema é decomposto em sistemas separados que se comunicam via arquivos de dados. Entre as variações dos modelos de acoplamento livre estão pré-processadores, pos-processadores, coprocessadores, e interface de usuários.	Modelo para prever a utilização de força de trabalho. Usa-se uma rede neuronal que prediz a carga de trabalho, através dos dados. A previsão é repassada para um arquivo de dados, e então um sistema especialista usa a carga de trabalho para determinar a utilização da força de trabalho.
Rígido 	Assemelha-se ao anterior, onde a diferença recai em como os dados são transferidos de um sistema para outro. Neste caso com dados residentes na memória. Logo, são mais adequados para problemas embutidos.	Usados como modelos de cooperação, são aplicados para monitoramento e controle, tomada de decisão, etc..
Totalmente integrado 	Estes modelos compartilham estruturas de dados e representação do conhecimento. A comunicação entre os diferentes componentes é completa via a natureza dual das estruturas. O raciocínio é feito por cooperação ou por componentes designados como controladores. Existe uma grande variedade de sistemas deste tipo.	Sistemas para identificação de objetos, uma rede neuronal recebe os dados de sensores e dados do ambiente do sistema especialista.

Tabela 3.2: Modelos para integrar sistemas inteligentes [Medsker & Bailey, 1992].

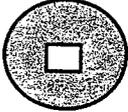
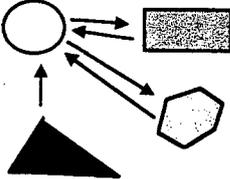
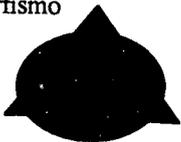
Classe	Descrição	Exemplo
Substituindo Funções 	Refere-se a composição funcional de uma técnica inteligente única. A função principal de uma dada técnica é substituída por uma outra técnica de processo inteligente.	Ajuste de pesos em uma rede neuronal via algoritmos genéticos [Montana, 1995], ajuste das funções de pertinência em sistema difuso via algoritmo genético [Karr, 1995].
Intercomunicação 	Módulos de processo inteligente independentes que permutam informação. Se um problema pode ser subdividido em varias subtarefas distintas, então técnicas independentes podem ser aplicada para resolver cada parte.	Se um problema inclui reconhecimento de padrões, raciocínio serial e otimização, então uma rede neuronal, um sistema especialista e um algoritmo genético podem executar estas respectivas subtarefas.
Polimorfismo 	São sistemas que usam uma arquitetura de processamento única para realizar a funcionalidade de técnicas de processos inteligentes diferentes.	Um exemplo classe polimorfismo são redes neuronais que funcionam como se elas estivessem fazendo procura genética.

Tabela 3.3: Três classes híbridas proposta [Goonatilake & Khebbal, 1995b].

Os sistemas híbridos inteligentes não inclui em somente a combinação de diferentes técnicas inteligentes mas também a integração de técnicas inteligentes com sistemas computacionais convencionais tal como a base de dados [Khebbal & Shamhong, 1995]. Para sistemas inteligentes aumentarem a capacidade de decisões organizacionais devem ser capazes de extrair e usar informações de origens muito variadas. Além disto, as decisões ou resultados produzidos por sistemas inteligentes deveriam ser propagados para aplicações existentes ou outros sistemas além do processamento. Por estas razões é vital que existam métodos e protocolos para integrar sistemas inteligentes com outros sistemas computacionais convencionais. Um destes métodos é a programação orientada a objetos [Wiener & Pinson, 1988] que é uma metodologia de engenharia de software que pode fornecer a “cola” para unir diferentes técnicas de processamento. Isto forma um modelo natural para os sistemas inteligentes híbridos pois técnicas individuais podem ser definidas como objetos que interagem enviando um conjunto comum de mensagens [Khebbal & Shamhong, 1995].

Sistemas híbridos inteligentes são uma classe muito poderosa de métodos computacionais que podem fornecer soluções para problemas que não podem ser solucionados com uma única técnica. Porém, como os desenvolvimentos e aplicações

são ainda relativamente novos, não existe a mesma disponibilidade de ferramentas e ambientes de desenvolvimento como nas técnicas estabelecidas como RNA e SE.

Um problema associado com o desenvolvimento de sistemas híbridos inteligentes é “educacional”. Para desenvolver sistemas com sucesso, necessita-se estar bem informado dos detalhes de diversas técnicas inteligentes, diferente de quando se trabalha com uma só técnica. E hoje existem poucos pesquisadores que têm experiência e treinamento suficiente no desenvolvimento e aplicações de mais de uma técnica. Para obter benefícios de sistemas híbridos deve-se ter uma visão pragmática de várias técnicas.

Já existem muitas aplicações usando sistemas inteligentes híbridos, como:

- Diagnóstico Médico [Gallant, 1988]
- Pré-clasificação de amostras de DNA em estudos do HIV [Benachenhou et al., 1990]
- Controle de Qualidade [Rosen & Silverman, 1992]
- Diagnóstico de falhas e realização de controle em sistemas de telecomunicações [Senjen et al., 1993]
- Previsão do preço de ações [Yonn et al., 1994]
- Controle para sistemas caóticos adaptativos (satélite) [Dracopoulos & Jones, 1995]
- Predição da tendência global de curto prazo do mercado de ações [Neto, 1995]
- Controle (adaptativo) de veículo [Spooner & Passino, 1996]
- Diagnóstico e solução de problemas financeiros de pequenas empresas [Pacheco, 1996]

3.4 Conclusões

Em poucos anos os sistemas inteligentes cresceram muito, representando um papel importante na solução de muitos problemas. Agora, com o desenvolvimento de técnicas inteligentes híbridas, problemas mais complexos poderão ser solucionados.

Estas técnicas fornecem uma alternativa para os métodos de análise de dados comumente usados como análise de regressão e para métodos de busca como programação linear. Além da capacidade de precisão algumas técnicas inteligentes

têm a capacidade de produzir modelos de decisão que são facilmente entendidas pelo usuário, outras tem a capacidade de aprendizagem e generalização, outras mecanismo de raciocínio semelhante ao do homem, conhecimento ambíguo e impreciso.

Uma tendência emergente é embutir as técnicas inteligentes em outras áreas de aplicações. As Redes Neurais estão agora sendo incorporadas em ferramentas de previsão sobre dados financeiros e sistemas difusos em produtos de recuperação de dados.

Uma área significativa de crescimento provável é o uso de sistemas inteligentes híbridos. A maioria dos problemas aparentemente intratáveis se beneficiarão do uso das múltiplas técnicas inteligentes. Uma das áreas que certamente será beneficiada será a previsão de séries temporais, como a previsão de mercado futuro.

Para problemas de classificação (como diagnóstico de empresas) uma das técnicas mais usadas hoje na área de IA são as redes neurais. Porém, como existem várias arquiteturas distintas, dedicar-se-á o próximo capítulo à análise de algumas arquiteturas, buscando as que mais se adaptam ao problema de diagnóstico.

4. REDES NEURONAIS ARTIFICIAIS

4.1 Introdução

Como foi visto no capítulo anterior as redes neuronais artificiais (RNAs) representam uma tecnologia emergente enraizada em muitas disciplinas. Elas são dotadas de alguns atributos como: aproximador universal (mapeamento de entrada/saída), classificador, habilidade de aprender do meio e adaptar-se ao meio, habilidade de trabalhar dados com ruído.

Sendo que se tem um problema de classificação (diagnóstico de empresas) e sabendo das capacidades das RNA como classificadores, este capítulo trará um estudo um pouco mais profundo sobre as redes neuronais (arquiteturas).

Então, este capítulo tem como um dos objetivos explicar e ilustrar os princípios básicos de algumas redes neuronais, tal como, Kohonen, LVQ, Hopfield, ARTs, MLPs e RBF.

Ainda considerando o problema de diagnóstico, também estudar-se-á um sistema híbrido, FAN (Free Association Neuron). O FAN é um sistema híbrido que está sendo desenvolvido por Roberto Raittz, em seu mestrado neste Programa de Pós-Graduação em Engenharia de Produção.

4.2 Algumas Características das RNAs

4.2.1 Introdução

As Redes neuronais artificiais (RNAs) são sistemas computacionais, de implementação em hardware ou software, que imitam as habilidades computacionais do sistema nervoso biológico, usando neurônios (também chamados de elementos de processamento ou nós) artificiais interconectados em arquiteturas específicas.

Os neurônios artificiais são emulações simples dos neurônios biológicos. Eles recebem informações de sensores ou de outros neurônios artificiais, produzindo operações simples sobre estes dados, e passam os resultados para outros neurônios. As redes funcionam através de seus neurônios, que processam seus dados usando paralelismo lógico combinado com operações seriais (quando a informação de uma camada é transferida para neurônios de outra camada). Existem três características principais que descrevem uma rede neuronal, e que contribuem para sua habilidade funcional: estrutura (arquitetura), dinâmica e aprendizado.

As RNAs são também chamadas de modelos conexionistas devido à importância das conexões entre os neurônios. Os pesos das conexões são a “memória” do sistema.

Esta seção descreve um histórico sobre as principais pesquisas sobre RNAs. Isto é seguido pela apresentação dos principais tipos de arquiteturas, algoritmos de aprendizagem e áreas de atuação. Finalmente, uma taxinomia sobre as RNAs.

4.2.2 Histórico sobre as pesquisas sobre as RNAs

O interesse em redes neuronais data do início da década de 1940, com o trabalho pioneiro de McCulloch e Pitts [McCulloch & Pitts, 1943]. Warren McCulloch foi um psiquiatra e neuroanatomista que estudou por 20 anos a representação de um evento no sistema nervoso. Walter Pitts foi um prodígio matemático, que juntou-se a McCulloch em 1942 e em 1943 escreveram um artigo que se tornou clássico, e recebeu muita atenção da comunidade que estudava o modelo do neurônio. Um grupo da Universidade de Chicago, sob a liderança de Rashevsky, vinha estudando o neurônio pelo menos 5 anos antes da publicação do artigo [Haykin, 1994].

Outro resultado foi publicado por Hebb [Hebb, 1949], que foi um dos primeiros a sugerir um processo plausível para rede neuronal aprender. Até hoje, muitos dos modelos de aprendizagem usados por pesquisadores são de alguma forma baseados na aprendizagem Hebbiana.

Provavelmente a primeira simulação computacional de RNAs foi relatada por Rochester e seus colegas (1956) na conferência de verão de Dartmouth onde

reconheceram como data do início oficial da Inteligência Artificial. Conduzindo simulações do modelo de Hebb, este grupo descobriu que algumas mudanças eram essenciais para que o modelo exibisse certas propriedades preditas pelo modelo. Eles generalizaram o modelo, incluindo inibição, de modo que células ativas pudessem inibir outros. Eles também introduziram normalização dos pesos para prevenir o crescimento ilimitado de alguns pesos. Embora este trabalho não fosse muito conclusivo, foi importante como precursor de muitos estudos de simulação.

Um das realizações mais sensacionais da pesquisa neste período foi o trabalho de Rosenblatt sobre os perceptrons [Rosenblatt, 1958]. Rosenblatt era um psicólogo que acreditou que o cérebro funcionava como um associador de aprendizagem que classifica em resposta a estímulos. Ele desenvolveu diversas variações de redes que ele chamou de perceptrons e estudou formas diferentes de aprendizagem. A rede básica do perceptron era uma unidade lógica limiar composta de três camadas: uma camada de entrada (foto) sensória que foi randomicamente ligada a uma camada de associação que foi, por sua vez conectada a uma camada de saída ou camada de classificação. Se as entradas cumulativas da camada sensória excederem algum limiar, aquela unidade é ativada e passa um impulso à camada de resposta. A camada de resposta então produz uma saída de +1 (classe 1) se a entrada cumulativa exceder algum limiar e, uma saída de 0 se não (não classe 1). Mas, é importante notar aqui que muitas das idéias e conceitos estudados por Rosenblatt são ainda objetos de pesquisas hoje.

Bernard Widrow foi um dos primeiros pesquisadores a desenvolver aplicações práticas de RNAs. Ele desenvolveu um modelo de neurônio simples semelhante ao Perceptron chamado ADALINE (ADaptive Linear Neuron), e redes de ADALINES que ele chamou MADALINE (Múltiple ADALINES). Estes tipos de unidades estão em uso ainda hoje para redução de eco para circuitos telefônicos de longa distância e para redução de ruídos em MODEMS de alta velocidade. Widrow e seu colegas são também responsáveis pelo desenvolvimento de um procedimento de aprendizagem supervisionado conhecido como mínimos quadrados (*Least Mean Square -LMS*) ou método de aprendizagem de Widrow-Hoff [Widrow & Hoff, 1960] usado para aprender associações de pares de entrada-saída. O LMS foi importante não só pelo

método em si, mas também porque serviu como um precursor do método de aprendizagem *backpropagation* usado nas redes multicamadas *feedforward*. A principal diferença entre o “perceptron”, apresentado por Roseblatt, e o ADALINE de Widrow situa-se no procedimento de treinamento. Após a apresentação do perceptron em 1960, acreditava-se que redes neuronais (perceptrons) poderiam fazer qualquer coisa, ou seja, resolver qualquer problema.

As possibilidades de IA introduzido pelos perceptrons produziram uma grande quantidade de pesquisas durante os anos sessenta. Então, em direção ao fim da década, Minsky e Papert publicaram seu tratado crítico sobre Perceptrons [Minsky & Papert, 1969]. Este livro era uma análise matemática elegante das capacidades computacionais e limitações dos perceptrons. Ele essencialmente mostrou quais funções lógicas simples o perceptrons podia e não podia calcular

Após a publicação do livro de Minsky e Papert, das limitações quanto a computadores e estações de trabalhos para conduzir experimentos, e também, por não haver suporte financeiro para conduzir projetos nesta área, as pesquisas com redes neuronais ficaram esquecidas pelo menos até o início de 1980. Só alguns pesquisadores valentes continuaram seus trabalhos, James Anderson, Teuvo Kohonen, Stephen Grossberg, Bernard Widrow, Chr. Von der Malsburg, Amari e alguns poucos outros.

Durante o início dos anos setenta, um número de investigações foram conduzidas sobre as memórias associativas por pesquisadores tal como Kohonen [Kohonen, 1972], Anderson [Anderson, 1972] e outros. Os modelos usados nestes trabalhos foram os associadores de camada linear única. Eles usaram uma forma de aprendizagem Hebbiana ou Correlativa. O trabalho destes pesquisadores propiciaram muitos *insight* sobre comportamento de associadores lineares e fenômenos relacionados tal como o “*crosstalk*” que está relacionado à armazenagem de muitos padrões não ortogonais.

James Anderson da Universidade Brown e seus companheiros de trabalho também desenvolveram modelos de memórias associativas endereçado por conteúdo baseado na aprendizagem de Hebb [Patterson, 1995]. Os padrões eram armazenados como *superimposition* (a soma) de matrizes e recuperados quando entradas

suficientemente semelhantes fossem apresentadas para a rede. Anderson foi também responsável por uma extensão do modelo associador linear chamado de *Brain-State-in-a-Box* (BSB). Neste modelo, a saída é truncada para prevenir crescimento ilimitado como o modelo iterativo para achar uma solução. Os valores truncados definem um hipercubo a qual a saída está contida nele.

Stephen Grossberg, fundador e diretor do Centro para Sistemas Adaptivos e um professor de matemática, psicologia, e engenharia biomédica na universidade de Boston, foi um pesquisador ativo em processamento de informação psicológica e biológica e no uso de redes neuronais artificiais para modelar a percepção e cognição humana desde o início dos anos sessenta. O trabalho inicial de Grossberg focalizou-se sobre sistemas de aprendizagem cooperativo-competitivo. O trabalho posterior feito por Grossberg e seus colegas focalizou sobre propriedades dinâmicas matemáticas das RNAs. Este trabalho levou a um importante teorema sobre a convergência global das redes dinâmicas. Grossberg é talvez mais conhecido pelo alto sucesso das redes ARTs (Adaptive Resonance Theory), que ele inventou. Ele e seus colegas, particularmente Gail Carpenter, estudaram, generalizaram e caracterizaram as redes ART extensivamente [Carpenter & Grossberg, 1987].

Em 1982, John Hopfield apresentou um artigo na Academia Nacional de Ciência descrevendo como uma análise de pontos estáveis poderia ser realizada por redes recorrentes simétricas. A análise foi baseada no uso de uma função de energia de Lyapunov para equações não lineares. Ele mostrou que a função energia dissipava e convergia para um mínimo e permanecia ali. Assim, padrões são armazenados na memória como atratores dinamicamente estáveis. As redes, que agora são chamadas Hopfield, podem ser usadas como redes de memória associativa ou para encontrar soluções de problemas com restrições tal como “o problema do cacheiro viajante”. Hopfield, que foi um ganhador de um prêmio nobel em física, recebeu o crédito por ressuscitar o interesse das pesquisas em redes neuronais no início dos anos oitenta.

Um dos desenvolvimentos mais importantes das recentes pesquisas em RNAs foi o descobrimento de um algoritmo de aprendizagem para ajustar os pesos em redes multi camadas *feedforward* (também referidas como perceptrons multi camadas). O algoritmo é conhecido como *backpropagation* já que os pesos são ajustados da

camada de saída para trás camada-a-camada para reduzir os erros de saída. O método foi descoberto em diferentes tempos por Werbos [Werbos, 1974], Parker [Parker, 1985], Rumelhart, Hinton e Williams do Grupo de Processamento Distribuído Paralelo [Rumelhart et al., 1986]. Este desenvolvimento abriu caminho para elaboração de redes neuronais mais genéricas superando as limitações dos perceptrons de camada simples. Tais redes são capazes de resolver problemas não lineares tal como a função lógica XOR.

O pesquisador japonês Kunihiko Fukushima é o fundador das redes Cognitron e Neocognitron [Fukushima, 1988; Fukushima et al., 1983]. A rede mais recente, o Neocognitron, é uma rede *feedforward* hierárquica que aprende através de métodos supervisionados ou não. O Fukushima e seus colegas publicaram resultados mostrando que o neocognitron é capaz de reconhecer caracteres manuscritos, independente de escala, posição e alguma deformação nos caracteres. Um dos aspectos raros dos Neocognitrons é a conectividade das camadas da rede. As camadas são conectadas de tal modo que características de baixo-nível são reconhecidas e sucessivamente combinadas em uma forma completa coerente para identificação de objetos.

Em 1988, Broomhead e Lowe descreveram um procedimento para o projeto de uma rede neuronal (feedforward) usando funções de base radial, conhecida na literatura como “radial basis function” (RBF), que proporcionou um perceptron alternativo [Broomhead & Lowe, 1988]. Bashkirov, Braverman e Muchnick em 1964 desenvolveram um método de funções potenciais. A idéia básica da RBF lembra bastante este método [Haykin, 1994]. Um trabalho suplementar apareceu em 1990 com Poggio e Girosi, que enriqueceram a teoria de redes RBF aplicando a teoria da regularização de Tikhonov [Poggio & Girosi, 1990].

Os acontecimentos precedentes são resumidos na Figura 4.1.

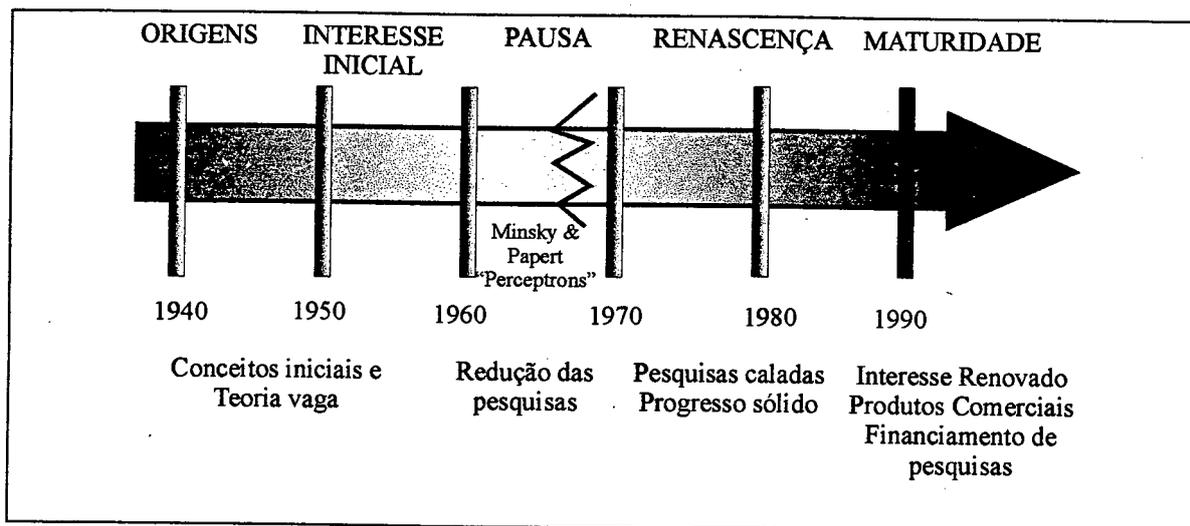


Figura 4.1: Eventos significantes no desenvolvimento de Redes Neurais [Patterson, 1995].

4.2.3 Arquiteturas

As RNAs podem ser vistas como grafos direcionados com pesos na qual os neurônios artificiais são os nós e as extremidades (com pesos) são as conexões entre neurônios de entrada e de saída.

Baseado nos padrões de conexões (arquitetura), as RNAs podem ser agrupadas em duas categorias:

- *redes feed forward*, nas quais os grafos não têm loop (canto superior esquerdo da Figura 4.1), e
- *redes recorrentes (ou feedback)*, nas quais os loops ocorrem devido as conexões feedback (canto superior direito da Figura 4.1).

Na família mais comum de redes feed-forward, redes Multicamadas, os neurônios são organizados nas camadas que têm conexões unilaterais entre eles. A Figura 4.2 também mostra redes típicas para cada categoria.

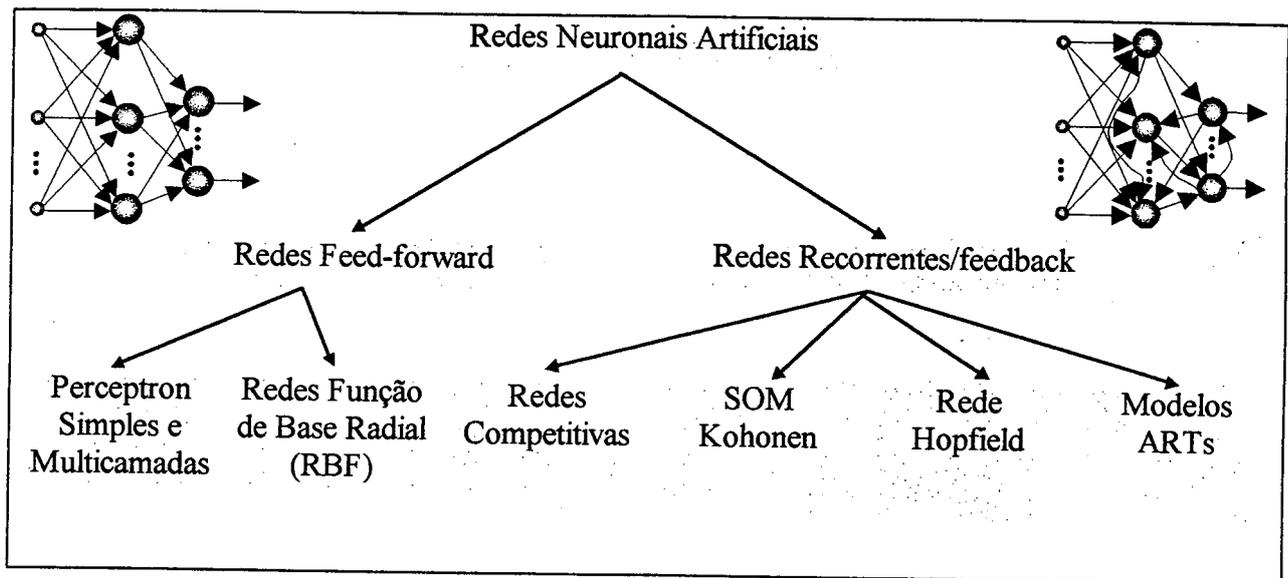


Figura 4.2: Uma taxinomia de arquiteturas de redes feed-forward e recorrentes/feedback.

4.2.4 Aprendizagem

A habilidade para aprender é uma característica fundamental da inteligência. Embora uma definição precisa de aprendizagem seja difícil de formular, um processo de aprendizagem no contexto de RNAs pode ser visto como um problema da arquitetura da rede e dos pesos entre as conexões adaptarem-se de modo que a rede possa eficientemente realizar uma tarefa específica.

Para entender o processo de aprendizagem, primeiro se deve ter um modelo do meio no qual uma rede neuronal opera, isto é, deve-se conhecer quais informações estão disponíveis para a rede. Este modelo é referido na literatura como um paradigma da aprendizagem [Haykin, 1994]. Segundo, deve-se entender como os pesos das redes são adaptados, isto é, que regras de aprendizagem que governam o processo de adaptação. Um algoritmo de aprendizagem refere-se ao procedimento no qual regras de aprendizagem são usadas para ajustar os pesos.

Existem três principais paradigmas de aprendizagem: supervisionada, não supervisionada e híbrida:

- Na *aprendizagem Supervisionada*, ou aprendizagem com “professor”, o ajuste ocorre quando o sistema diretamente compara a saída y da rede com a resposta correta conhecida d . Os pesos são ajustados de modo a permitir que a rede produza

saídas mais próximas da resposta desejada (conhecida). A *aprendizagem com reforço* é uma variante da aprendizagem supervisionada, onde o sistema conhece somente se a saída da rede é correta ou não, mas não conhece a saída correta. Neste caso, a rede tenta aprender através de tentativa e erro com objetivo de maximizar um índice de performance chamado de sinal de reforço.

- Na *aprendizagem Não Supervisionada*, ou aprendizagem "sem professor", a rede não requer uma resposta correta associada com cada padrão de entrada no conjunto dos dados de treinamento. A rede explora a estrutura implícita ou a correlação entre os padrões de entradas nos dados, e organiza estes padrões em categorias em função desta correlação.
- A *aprendizagem Híbrida* combina aprendizagem supervisionada e não supervisionada. Partes dos pesos são usualmente determinados através de aprendizagem supervisionadas, enquanto outros são obtidos através da aprendizagem não supervisionada.

A teoria da aprendizagem deve levar em consideração três questões fundamentais e práticas associadas com a aprendizagem através das amostras: capacidade, complexidade da amostra e complexidade computacional:

- *Capacidade*, concerne em como muitos padrões podem ser armazenados, e quais funções e limites de decisões uma rede pode formar.
- *Complexidade da amostra*, determina qual o número de padrões de treinamento são necessários para treinar uma rede de modo a garantir uma generalização válida.
- *Complexidade Computacional*, refere-se ao tempo requerido para um algoritmo de aprendizagem estimar uma solução dos padrões de treinamento. Muitos algoritmos de aprendizagem tem alta complexidade computacional. Designar algoritmos eficientes para a aprendizagem de uma rede neuronal ainda é um tópico ativo de muitas pesquisas.

Existem quatro tipos básicos de regras de aprendizagem: aprendizagem Hebbiana, de correção do erro, estocástica, e competitiva.

APRENDIZAGEM DE HEBB: A regra de aprendizagem mais antiga é o *postulado de Hebb de aprendizagem* [Hebb, 1949]. Hebb baseou-se na seguinte observação de experimentos neurobiológico: se neurônios em ambos os lados de uma sinapse são ativados sincronicamente e repetidamente então o comprimento da sinapse é seletivamente aumentado.

Matematicamente, a regra de Hebb pode ser descrita como

$$w_{ij}(t+1) = w_{ij}(t) + \eta y_j(t)x_i(t) \quad (4.1)$$

onde x_i e y_j são os valores de saídas dos neurônio i e j , respectivamente, os quais são conectados pela sinapse w_{ij} , e η é a taxa de aprendizagem. Note que, x_i é a saída para a sinapse.

Uma importante propriedade desta regra é que a aprendizagem é feita localmente, isto é, a mudança nos pesos depende somente da ativação dos dois neurônios conectados pelo peso. Isto simplifica significativamente a complexidade da aprendizagem.

CORREÇÃO DO ERRO: No paradigma da aprendizagem supervisionada, a rede recebe uma saída desejada para cada padrão de entrada. Durante o processo de aprendizagem, a saída y gerada pela rede pode não ser igual a saída desejada d . O princípio básico da regra de aprendizagem correção do erro é usar o sinal de erro ($d - y$) para modificar os pesos das conexões para gradualmente reduzir este erro.

A regra de aprendizagem do Perceptron e o algoritmo de aprendizagem backpropagation são baseados neste princípio de correção de erro.

ESTOCÁSTICA: A regra de aprendizagem estocástica é acompanhada pelo ajuste dos pesos de uma maneira probabilística. Exemplos de aprendizagem estocástica são encontradas nas máquinas de Boltzmann e de Cauch, onde os estados de todos os neurônios são determinados por uma distribuição de probabilidade. Durante a fase de aprendizagem, o sistema é operado em dois modos: *clamped*, no qual neurônios

visíveis (entrada/saída) são *clamped* em estados específicos determinados pelo meio; e *free-running*, aos quais, neurônios visíveis e escondidos, é permitido operar livremente. Permite-se que a rede opere em ambos os modos até que um equilíbrio “térmico” seja alcançado. Neste momento, os pesos w_{ij} nas conexões entre os nós i e j são então ajustados com base na diferença entre as duas propabilidades $p_{ij}(C)$ e $p_{ij}(U)$. Aqui $p_{ij}(C)$ é a probabilidade do i -ésimo e o j -ésimo elementos estarem ambos no modo *clamped*, enquanto $p_{ij}(U)$ é a probabilidade de ambos estarem no modo *free-running*. Como a rede é simétrica é possível definir uma função de energia para o sistema como:

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} s_i s_j \quad (4.2)$$

onde s_i é o estado binário $\{0,1\}$ ou estado bipolar $\{-1,1\}$ da i -ésima unidade. O equilíbrio é alcançado quando a função de energia alcança um mínimo.

COMPETITIVA: Ao contrário da regra de Hebb, na aprendizagem competitiva as unidades de saída competem entre elas para ativação. Como um resultado, somente uma unidade de saída é ativada de cada vez. Este fenômeno é conhecido como *winner-take-all*. A aprendizagem competitiva é encontrada nas redes neurais biológicas.

Este tipo de aprendizagem frequentemente categoriza os dados de entradas. Padrões similares são agrupados pela rede e representados por uma única unidade. Este agrupamento é feito automaticamente baseado na correlação dos dados.

O exemplo mais conhecido de aprendizagem competitiva é a *quantização de vetor* para compressão de dados.

4.2.5 Aplicações e Principais Áreas de Atuação

A técnica de RNA pode tratar de diversos problemas, que se pode dividir em quatro categorias principais [Folgelman-Souliè, 1995]:

1. *Classificação e diagnóstico*: Um padrão é apresentado para a rede, que tem que decidir em qual das classes pré-definidas aquele padrão pertence. Esta situação tipicamente ocorre em problemas de Reconhecimento de Padrões.
2. *Aproximação de Função*: A rede deve associar ao padrão de entrada um valor próximo ao valor da função desconhecida que representa um fenômeno “real”. Este problema será encontrado em previsão de séries temporais, identificação e processos de controle.
3. *Compressão, extração de características e quantização*: Na maioria das aplicações, dados brutos, tal como aqueles vindo de sensores, tem alta dimensionalidade e ruídos. As Redes Neurais podem ser usadas para reduzir a dimensão e o ruído, tanto por compressão, como através de extração das características relevantes. Quantização, supervisionada e não-supervisionada podem também ser usadas para representar os dados através de um conjunto limitado de referências.
4. *Otimização*: RNAs podem ser definidas de modo que seus pesos possam representar uma função objetivo a ser otimizada, e as restrições nas variáveis. Uma solução (usualmente aproximada) é então obtida como um estado limite da dinâmica da rede.

A maioria das aplicações de RNAs desenvolvidas usam um conjunto muito restrito de algoritmos: Redes Multicamadas (principalmente Perceptron Multicamadas (MLP) com *backpropagation* e *Radial Base Function* (RBF)), *Learning Vector Quantization* (LVQ) e os mapas topológicos de Kohonen (SOFM) (que usam mecanismo “*winner-take-all*”). Aproximadamente 90% de todas as aplicações são usadas para classificação, identificação, diagnóstico e predição. Embora os primeiros desenvolvimentos de aplicações com sucesso datam nos anos 50, a introdução das RNAs na indústria em larga escala tem seu início nos meados dos anos 80 com as redes MLP.

Os principais domínio de aplicações onde as RNAs estão sendo usadas são os seguintes:

- *Setor Militar*: Processamento de sinais para identificação de alvos ou origens, fusão de dados, análise de imagens para localização e identificação do alvo.
- *Processamento de Imagens*: Reconhecimento de objetos, análise de cenas, análise de imagens de satélites e teledetecção.
- *Telecomunicações*: controle, codificação, filtros, canal de transmissão/equalização.
- *Manufatura*: previsão do consumo de energia (água, eletricidade), identificação e controle de processos industriais.
- *Robótica*: veículos autônomos, controle do braço de um robô.
- *Biologia e Medicina*: análise de imagens, processamento de sinais, diagnóstico automático, análise de genoma.
- *Setor Terciário*: leitura automática de textos (digitado e manuscrito), autenticação de assinaturas, detecção de fraudes.
- *Finanças*: previsão do preço de ações e de índices econômicos, análise de riscos, diagnóstico de empresas.

4.2.6 Taxinomia das Redes Neurais

A Tabela 4.1 resume vários algoritmos de aprendizagem e as principais arquiteturas associadas. Os paradigmas de aprendizagem supervisionado e não supervisionado empregam regras de aprendizagem baseadas nas aprendizagens Correção do Erro, Hebbiana e Competitiva. Regras de aprendizagem baseadas na Correção do Erro podem ser usadas para redes de treinamento *feed-forward*, enquanto a aprendizagem de Hebb tem sido usada para todos os tipos de arquiteturas. Contudo, cada algoritmo de aprendizagem é projetado para treinamento de uma arquitetura específica. Portanto, quando discuti-se um algoritmo de aprendizagem, uma arquitetura particular está subentendida. Cada algoritmo pode realizar bem somente umas poucas tarefas. A última coluna da Tabela 4.1 lista tarefas que cada algoritmo pode realizar.

Paradigma de Aprendizagem	Regra de Aprendizagem	Tipo de Arquitetura	Rede Neuronal	Tipo de Aplicações
Supervisionado	Correção-Erro	Feed-forward Simples ou Multicamadas	Perceptron, PMC com Back, Adaline e Madaline	Classificação de Padrões, Aproximação de Funções, Predição e Controle
	Estocástica	Recorrente	Máquina de Boltzmann e de Cauchy	Classificação de padrões e Otimização
	Hebbiana	Feed-forward com uma camada	Hopfield	Memória Associativa
	Competitiva	Feed-forward com uma camada	LVQ	Compressão de dados
Não Supervisionada	Hebbiana	Recorrente	Hopfield	Reconhecimento de Padrões e Classificação
	Competitiva	Feed-forward Multicamadas	Neocognitron	Otimização
Híbrida	Correção-Erro / Competitiva	Recorrente	ARTs	Reconhecimento de Padrões
		Feed-forward com uma camada	SOFM	Memória Associativa, Classificação e Reconhecimento de Padrões
		Feed-forward com uma camada	RBF	Otimização, Classificação, Reconhecimento de Padrões e Predição

Tabela 4.1: Algoritmos de Aprendizagem mais conhecidos.

4.3 Caracterização das Principais Redes Neurais

4.3.1 Introdução

Nesta seção apresenta-se a estrutura básica das redes de Kohonen, LVQ, Hopfield, ART, MLP e RBF assim como suas limitações, vantagens e algumas aplicações.

4.3.2 Rede de Kohonen

Os Mapas de Características Auto-Organizáveis (Self-Organizing Feature Maps – SOFM) ou a rede de Kohonen, introduzido por Teuvo Kohonen [Kohonen, 1982; Kohonen, 1988; Kohonen, 1990] (pesquisador da Universidade de Helsinki), é um dos modelos de redes neurais mais usados, principalmente para quantização de vetor e análise de dados, mas também aplicável em quase todas as tarefas onde redes neurais tiveram sucesso.

O modelo de Kohonen é do tipo *feed forward*, de treinamento não supervisionado com competição. O esquema básico do modelo têm a propriedade de modificar a si próprio. Assim, os neurônios próximos ao modificado respondem similarmente. Os neurônios da camada competem entre si para serem os vencedores em cada modificação. Os neurônios cujo vetor dos pesos gerar a menor distância Euclidiana (ou outra distância) com o vetor de entrada é o vencedor, mas também os pesos dos seus vizinhos (dentro de senso físico) são ajustados.

Este comportamento tem por objetivo fazer com que a rede simule uma atividade cerebral. Este paradigma é baseado na teoria de que as células nervosas corticais estão arranjas anatomicamente em relação aos estímulos que recebem dos sensores às quais estão ligadas. Alguns cientistas, rastreando a atividade cerebral, notaram que, apesar de todas as células nervosas estarem ligadas e coligadas entre si, existem centros da atividades mais intensificados conforme a atividade exercitada no momento. Quando assiste-se um filme, por exemplo, ou quando se está efetuando um raciocínio matemático, existem áreas do cérebro diferentes com maior atividade. Este é o paradigma topológico.

Teuvo Kohonen introduziu um conceito muito interessante de mapas de características topologicamente auto-organizados, que são mapas que preservam a topologia de uma representação multidimensional dentro de uma matriz de neurônios de 1, 2 ou mais dimensões.

O conceito de topologia (ou melhor, o conceito de preservação topológica) tornou-se a característica essencial da abordagem de Kohonen na pesquisa de RNA.

Arquitetura

O SOFM é provavelmente a RNA mais próxima entre todas as arquiteturas e esquemas de aprendizagem de RNA da rede neuronal biológica. Como uma regra, a rede Kohonen é baseada em uma única camada de neurônios arranjados em um plano bi-dimensional, por exemplo, tendo uma topologia bem definida (Figura 4.3).

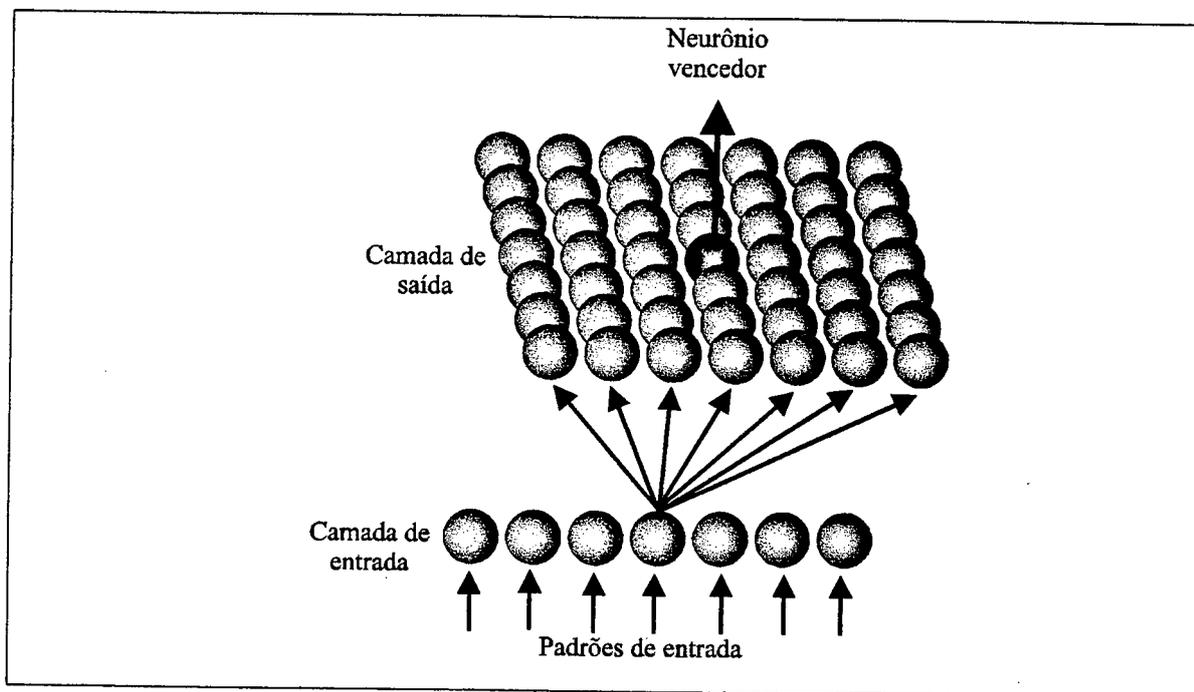


Figura 4.3 : Arquitetura da rede Kohonen, caso de duas dimensões [Pandya & Macy, 1995].

Uma topologia definida significa que cada neurônio tem um número de neurônios como vizinhos mais próximos, segundo vizinhos mais próximos, etc.. A vizinhança de um neurônio é usualmente arranjada em quadrados ou hexágonos, o que significa que cada neurônio tem 4 ou 6 vizinhos mais próximos. O conceito de

vizinhos mais próximos necessita de alguma elaboração quando se fala em topologia. Por exemplo, a vizinhança topológica quadrada é frequentemente considerada como tendo 8 e não quadro vizinhos mais próximos (Figura 4.4). Os pontos dos cantos estão mais distantes do centro que os verdadeiros vizinhos mais próximos, mas o interesse é na topologia, isto é, nas conexões e não na verdadeira distância. Na concepção de Kohonen sobre rede neuronais, o sinal de similaridade está relacionado com a relação espacial (topológica) entre os neurônios na rede.

O SOFM de Kohonen tenta mapear as entradas de maneira que sinais similares excitem neurônios que estão próximos (em termos de distância espacial). Esta relação distância-por-similaridade deve ser generalizada para incluir todas as relações de similaridade entre diferentes sinais também. A aprendizagem de Kohonen representa uma tentativa de ajustar o espaço de sinais dentro de uma rede neuronal por uma espécie de procedimento suave. O objetivo da aprendizagem de Kohonen é mapear sinais similares em direção às posições de neurônios similares.

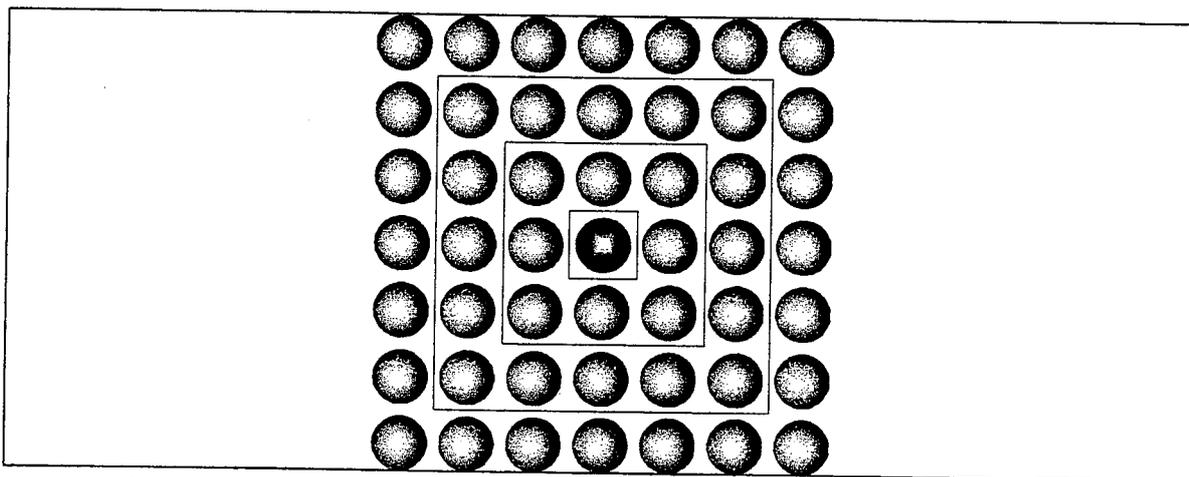


Figura 4.4: Topologia quadrada [Pandya & Macy, 1995].

Em uma rede de Kohonen, pode-se falar de duas camadas: entrada e saída. Somente a camada de saída ativa e usualmente arranjada como uma grade de neurônios (mas pode também ser arranjada como um vetor).

Pode-se trabalhar também com mapas mais complexos, como espaços de 3 ou 4 dimensões. Neste caso, os mapeamentos topológicos podem ser usados para transformar a superfície de 3 dimensões, por exemplo, em uma superfície plana.

Todos os neurônios na camada ativa recebem a mesma entrada. O aspecto mais característico da rede de Kohonen é a existência de somente um *feedback* local, isto é, a saída de cada neurônio não está conectada para todos os outros neurônios no plano (como na rede de Hopfield), mas somente para um número pequeno que estão topologicamente próximos a ele. Portanto, os neurônios topologicamente próximos comportam-se de forma similar quando sinais similares são apresentados (Figura 4.5).

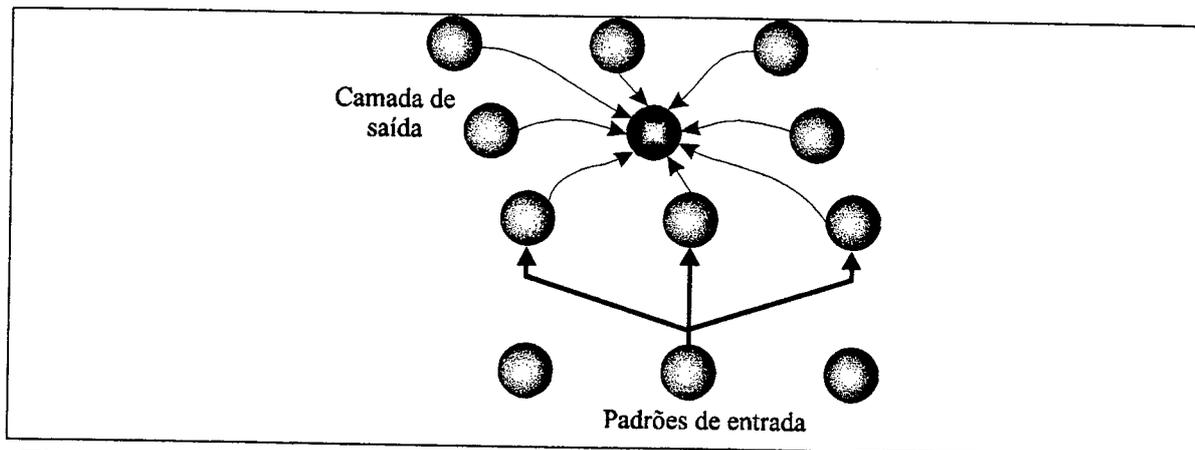


Figura 4.5: Feedback lateral em uma camada de Kohonen de duas dimensões [Pandya & Macy, 1995].

Na aprendizagem competitiva, somente um neurônio é ativo numa camada após cada entrada. A rede de Kohonen seleciona o neurônio vencedor “c” como sendo aquele que está mais próximo do vetor de entrada. Após encontrar o neurônio vencedor deve-se corrigir os pesos do neurônio vencedor e de sua vizinhança.

Vantagens e Limitações

O SOFM tem uma série de desvantagens computacionais que afetam a performance de aplicações de grande porte executadas em computadores paralelos (mas não em série). A fim de encontrar o neurônio vencedor (e vizinhança), o programa deve verificar todos os n neurônios. Esta é uma restrição séria quando redes grandes estão sendo treinadas. Até em computadores paralelos, isto envolve $n/2$ comparações paralelas, que requerem pelo menos $\log_2 n$ passos. Em máquinas comuns, todos os cálculos devem ser feitos sequencialmente, e guardando a maior saída não afeta muito o desempenho total.

A rede de Kohonen ou uma camada de Kohonen pode ser construída em uma rede complexa como uma de suas camadas [Wasserman & Schwartz, 1988] ou implementada como em combinação com outras técnicas [Ritter et al., 1990].

Aplicações

A rede de Kohonen tem encontrado aplicações nas áreas de quantização de vetores para compressão de dados, correção de erros e a geração de códigos de palavras. Além disto, podem ser efetivamente usadas para otimização, controle, reconhecimento de voz e mapeamentos sensoriais.

Teuvo Kohonen [Kohonen, 1988] e seus colegas desenvolveram um “*phonetic typewriter*” no início dos anos oitentas. O *phonetic typewriter* é um sistema de reconhecimento que pode transcrever discursos em textos ortograficamente corretos.

A rede de Kohonen unidimensional foi usada com sucesso para resolver o problema de otimização do Caixeiro Viajante [Angeniol et al., 1988].

Para ilustrar a habilidade de uma rede de Kohonen na realização de mapeamentos não lineares pode-se citar também a aplicação: controle do braço de um robô [Ritter et al., 1992].

Em classificação de padrões, o objetivo é classificar os sinais de entrada em um número finito de classes tal que a probabilidade média do erro seja minimizada. Em tais tarefas, a questão de particular importância é delinear os limites das classes onde as decisões são feitas. Para encontrar melhores resultados para classificação de padrões, o uso dos SOFMs deveria ser acompanhado por um esquema de aprendizado supervisionado. Uma possibilidade é o uso de abordagem híbrida como mostrado na Figura 4.6, que envolve uma combinação de SOFM e o LVQ (próxima seção) para classificação de padrões.

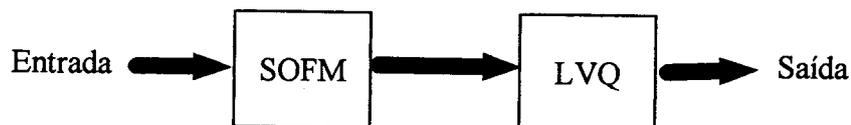


Figura 4.6 Classificação de padrões em dois estágios.

4.3.3 Rede LVQ (Learning Vector Quantization)

Um caso particular dos SOFM são as redes VQ (*vector quantization*). Ao contrário das redes SOFM, nas redes VQ não existe o feedback lateral. A unidade vencedora é a única beneficiada no momento da aprendizagem. Ou seja,

$$\begin{cases} w_c = w_c + a(t)(x - w_c) \\ w_c = w_c \end{cases} \quad (4.3)$$

onde c é a unidade vencedora, w é o vetor de pesos, x é a entrada e $a(t)$ é uma função para a taxa de aprendizagem.

Estas redes são capazes de aprender uma forma efetiva de quantização⁶ de vetor. As redes VQs aprendem a quantizar ou comprimir padrões de entrada de algum meio, um processo que também é aprendido em redes biológicas. Em aplicações tal como reconhecimento de fala e processamento de imagens, uma grande quantidade de dados deve ser armazenada, processada e possivelmente transmitidas. Tipicamente existe muita redundância nos dados deste tipo e existe a necessidade de comprimí-los. Um método de compressão é a rede VQ. Nestas redes os vetores de entrada x de dimensão n são transformados em um número finito de classes onde cada classe está representada por um protótipo w_i ($i=1, \dots, m$). Esta aproximação causa alguma distorção na representação, mas esta desvantagem é compensada pela considerável economia na armazenagem e redução na complexidade computacional do processo de compressão de dados.

Existem também as formas supervisionadas de VQ que são chamadas LVQ (Learning Vector Quantization). Estas redes permitem a especificação das categorias no qual as entradas estarão classificadas. A designação das categorias para o conjunto de treinamento são conhecidos a priori e fazem parte do conjunto de treinamento (como em qualquer método supervisionado). A arquitetura da LVQ é exatamente a mesma que a SOFM com a única exceção que cada neurônio na camada de saída é designado a pertencer a uma das categorias de classificação. Isto é ilustrado na Figura

⁶ Quantização é o processo de transformar variáveis de valores contínuos em variáveis discretas.

4.7. Em geral, diversos neurônios são atribuídos para cada classe. Como antes, o vetor peso (chamado vetor de referência ou protótipo para a classe que ele representa), para uma dada unidade de saída representa um exemplar dos vetores da entrada para o qual ele responde mais fortemente.

Durante o treinamento, como mostra a Figura 4.8, o neurônio vencedor c é premiado se ele pertence à categoria correta, movendo-se em direção ao vetor de entrada. Contrariamente, se o neurônio vencedor não pertence à categoria correta, ele é punido, sendo forçado a mover-se para longe do vetor de entrada. Isto é,

$$\begin{cases} w_c^{novo} = w_c^{velho} + a(x - w_c^{velho}), & \text{se } c \text{ é a classe correta} \\ w_c^{novo} = w_c^{velho} - a(x - w_c^{velho}), & \text{caso contrário.} \end{cases} \quad (4.4)$$

Depois do treinamento, uma rede LVQ classifica um vetor de entrada designando-o à mesma classe cujo vetor de peso da unidade de saída (vetor referência) é o mais próximo do vetor de entrada.

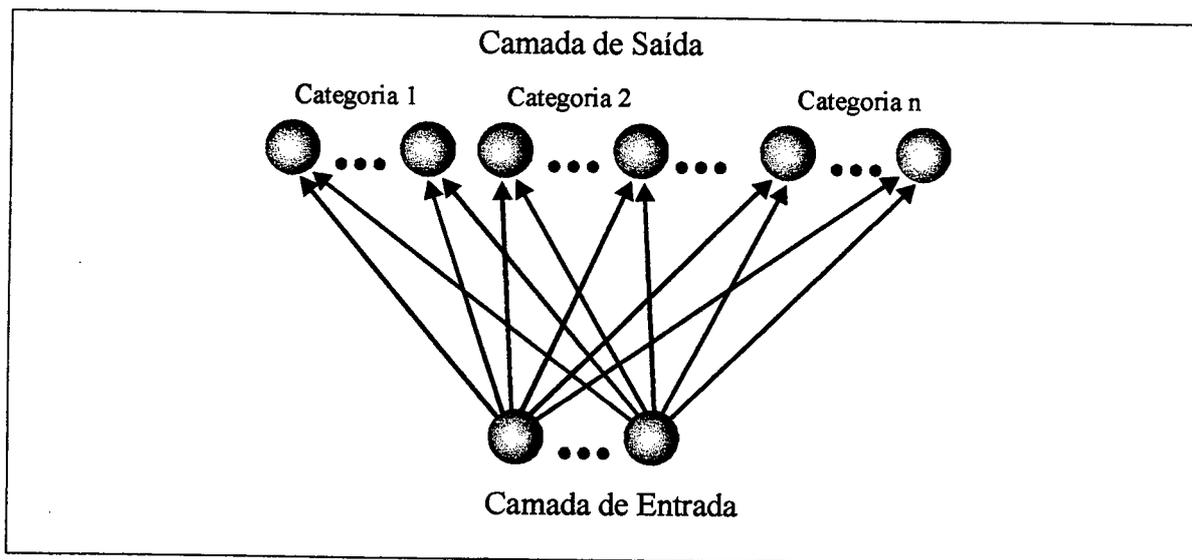


Figura 4.7 Arquitetura da LVQ [Pandya & Macy, 1995].

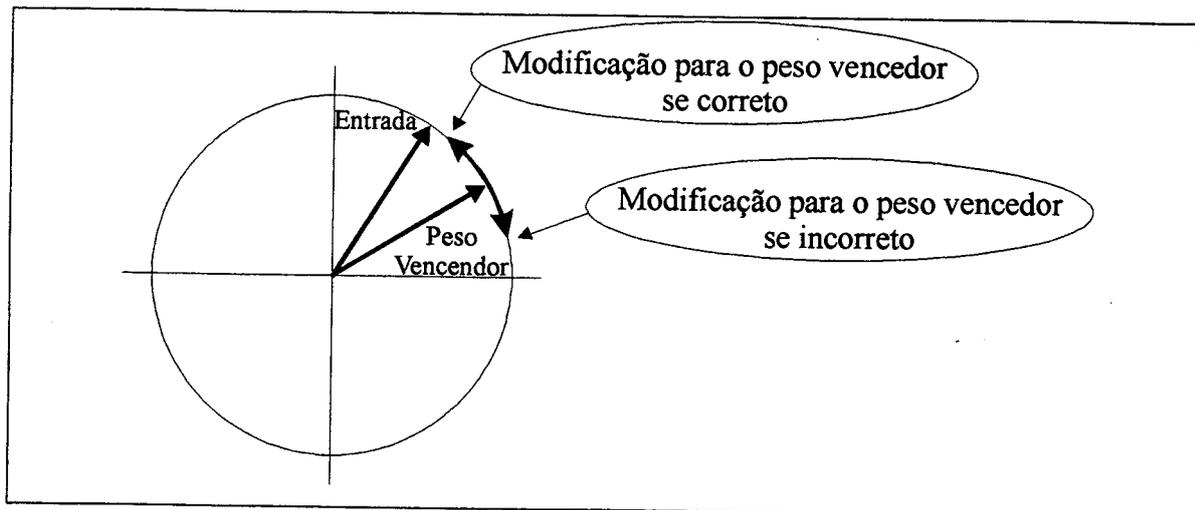


Figura 4.8 Treinamento da LVQ [Pandya & Macy, 1995].

Foram desenvolvidas variações do algoritmo básico LVQ (**LVQ 2**, **LVQ 2.1** e **LVQ 3**). No LVQ original - também chamado de LVQ1 - somente o vetor de referência que está mais próximo do vetor de entrada é atualizado. A direção que este vetor se move depende do vetor de referência pertence à mesma classe do vetor de entrada ou não. Nos algoritmos aperfeiçoados, dois vetores (o vencedor e o segundo) aprendem, se várias condições são satisfeitas. A idéia é a de que se a entrada está mais ou menos a mesma distância tanto do vencedor quanto do segundo, então ambos devem aprender.

Vantagens e Limitações

Estes classificadores tipicamente apresentam taxas de erros que são semelhantes ao *backpropagation*, mas freqüentemente o treinamento é mais rápido, e requer mais memória e tempo de computação durante a classificação. Também geralmente fornecem taxas de erros menores quando comparado com o SOFM, especialmente quando o número de nós é pequeno.

Aplicações

As redes LVQ são usadas principalmente para classificação, reconhecimento de padrões e compressão de imagens.

Baykal e Yalabki [Baykal & Yalabki, 1992] usaram uma rede LVQ em conjunto com uma rede *feed forward* no reconhecimento de caracteres de múltiplas

fontes. Eles registraram uma taxa de reconhecimento de 87% até mesmo com caracteres distorcidos, alterados ou rotacionados.

O classificador LVQ tem sido comparado com o classificador MLP com *backpropagation* usando problemas artificiais, problemas de reconhecimento de fonemas [McDermott & Katagiri, 1988] e para previsão de falência [Refenes, 1995]. Nestas comparações verificou-se a equivalência entre as duas redes.

4.3.4 Rede ART (*Adaptative Resonance Theory*)

Outra importante classe de redes recorrentes são as redes ARTs. Estas redes foram desenvolvidas nos meados dos anos setenta e oitenta por Stephen Grossberg em parceria com membros do Centro de Sistemas Adaptativos de Boston, principalmente com Gail Carpenter, sua esposa. A principal preocupação de Grossberg era desenvolver uma rede que tivesse plasticidade (discriminação) e estabilidade (generalidade).

No desenvolvimento de uma RNA espera-se que algumas das propriedades básicas do cérebro exibidas pela RNA sejam comparáveis com a rede biológica. Em particular, deseja-se que a RNA tenha habilidade de continuamente adaptar-se quando mudanças no meio ocorrem. Isto significa, ser capaz de reter fatos e informações úteis na memória enquanto, ao mesmo tempo, aprender fatos novos e importantes. Os fatos novos aprendidos não deveriam “apagar” informações antigas e úteis. Ao mesmo tempo, deseja-se que o modelo fosse capaz de ignorar fatos irrelevantes e até mesmo esquecer informações antiquadas ou não importantes. Em outras palavras, deseja-se que as redes exibissem um alto grau de estabilidade na aprendizagem adaptativa de novos conceitos ou categorias. Por outro lado, ela não deveria esquecer (ou perder) os fatos úteis previamente armazenados na memória para acomodar o novo conhecimento aprendido. Naturalmente, tal rede deveria ser capaz de discriminar entre as informações úteis e as irrelevantes. Este conflito, estabilidade e plasticidade, são chamados por Grossberg de *dilema da estabilidade/plasticidade* [Carpenter & Grossberg, 1987].

As RNA convencionais têm falhado na resolução do dilema estabilidade/plasticidade. Muito freqüentemente, em outros modelos de redes como

feed forward com aprendizagem *backpropagation*, ou na rede *Counterpropagation*, o aprendizado de um novo padrão apaga ou modifica os padrões treinados anteriormente. Caso uma rede completamente treinada necessite aprender um novo padrão de treinamento, ela pode modificar tanto os pesos que um outro treinamento é necessário.

Em alguns casos isto não tem importância, como na situação em que o conjunto de treinamento permanece fixo. Mas, em muitos casos do mundo real, a rede estará exposta a uma constante mudança ambiental; com isto pode-se não ter nunca o mesmo conjunto de treinamento duas vezes. Sobre tais circunstâncias, uma rede com aprendizagem *backpropagation*, por exemplo, freqüentemente não aprenderá nada; isto é, irá continuamente modificar seus pesos sem proveito, nunca chegando a um conjunto satisfatório.

A Teoria da Ressonância Adaptativa foi desenvolvida como uma extensão aos sistemas de aprendizagem competitivo/cooperativo. Ela foi desenvolvida na tentativa de superar o problema da estabilidade/plasticidade e outras características da aprendizagem instável associada com redes competitivas. A rede resultante (ART) teve um número importante de características que outras arquiteturas não têm, incluindo: aprendizagem em tempo-real (on-line); capacidade de auto-organização (aprendizagem não supervisionada); memória auto-estável em relação aos muitos padrões de entradas arbitrários; busca adaptativa rápida para melhor *match* de padrões armazenados; habilidade de aprendizagem rápida (ou lenta); rejeição de padrões de entradas não familiares quando a capacidade de memória for alcançada; critério de erro variável que permite um regulamento variável de agrupamentos em categorias e retenção bem sucedida das características de estabilidade e plasticidade

As redes ARTs mapeiam padrões de entradas n-dimensional em categorias ou classes de saídas baseadas nas características dos padrões de entradas. Os padrões de entradas similares (vizinhos mais próximos) são agrupados dentro da mesma classe e os padrões dissimilares em classes distintas. O grau de similaridade requerido para grupos de padrões é ajustável de modo que muitos grupos de classe de padrões altamente semelhantes sejam criados quando um limiar de semelhança é posto a um nível alto. No outro extremo, menos classes são criadas quando o limiar é posto a um

valor baixo. Neste último caso, os membros de classe podem possuir um grau de similaridade menor. A definição do limiar pode ser ajustado manualmente ou dinamicamente durante operações da rede dependendo da aplicação específica. Isto permite que uma rede ART torne-se mais ou menos seletiva enquanto continua a adaptar.

A aprendizagem nas redes ARTs ocorre naturalmente em tempo real durante operação normal da rede. Esta é uma forma de aprendizagem contínua e adaptativa não supervisionada onde uma nova categoria é automaticamente formada sempre que um padrão de entrada novo (dissimilar aos já apresentados) é apresentado a rede. Novas categorias continuarão sendo formadas da nova entrada até que a rede esgote seu limite de neurônios de categorias de saídas neutros (neurônios não categorizados). Neste momento rejeitará qualquer novo padrão de entrada. Padrões de entradas que são similares a categorias já estabelecidas são prontamente “reconhecidos” produzindo uma saída alta no neurônio da categoria selecionada. Entradas que assemelham-se as categorias existentes também iniciam algum grau de aprendizagem para a categoria selecionada, sem ao mesmo tempo perturbar a estabilidade das categorias aprendidas.

A arquitetura básica inclui três grupos de neurônios (Figura 4.9): um campo de processamento das entradas (camada F_1), as unidades de cluster (categorização) (camada F_2), e um mecanismo para controlar o grau de similaridade de padrões colocados no mesmo cluster (um mecanismo de reajuste). Denota-se a parte da entrada da camada F_1 como $F_1(a)$ e a parte de interface como $F_1(b)$.

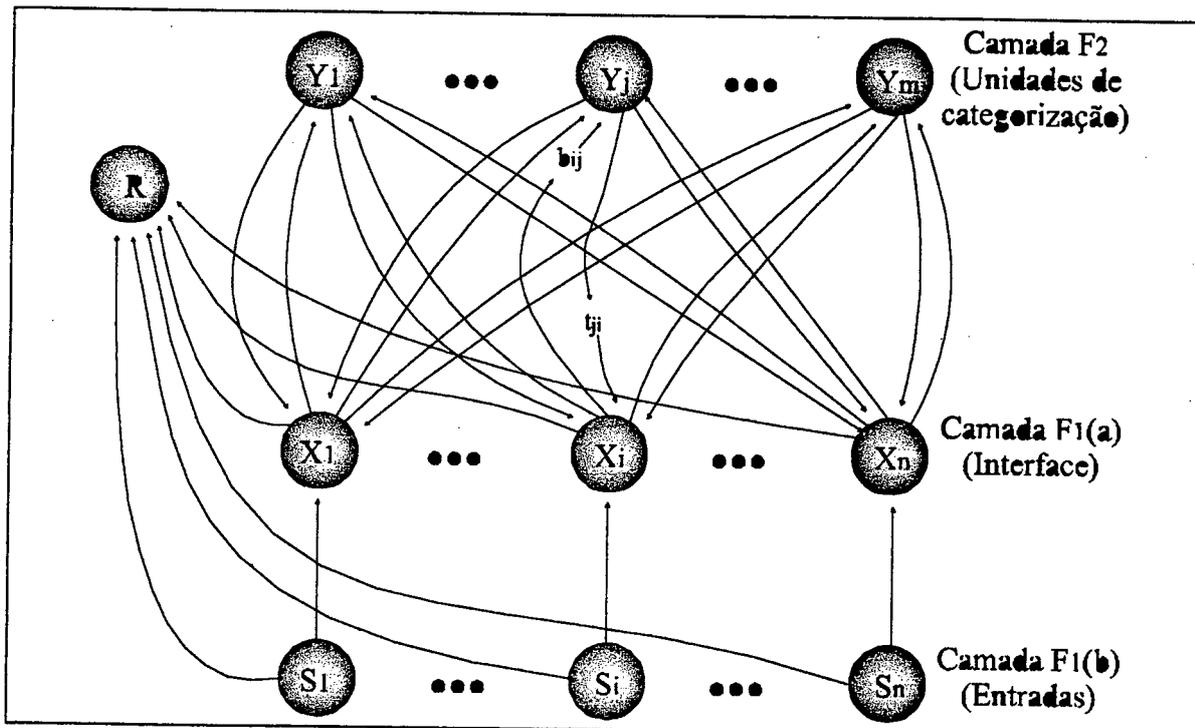


Figura 4.9: Estrutura básica da ART1 [Fauset, 1994].

Existem dois conjuntos de conexões entre cada unidade na parte de interface para o campo de entrada e cada unidade de *cluster*: b_{ij} conexão da i -ésima unidade F_1 para a j -ésima unidade F_2 (pesos *bottom-up*); t_{ji} conexão da j -ésima unidade F_1 para a i -ésima unidade F_1 (pesos *top-down*).

A camada F_2 é uma camada competitiva. A unidade de interface combina informações das unidades de entrada e unidades de categorização. Se a unidade de *cluster* pode aprender ou não, o padrão de entrada, depende da unidade de reajuste.

Uma tentativa de aprendizagem em ART consiste da apresentação de um padrão de entrada. A atualização é feita pela: ativação, inativação e/ou inibição de todas as unidades. Assim que um padrão é apresentado, ele continua enviando seus sinais de entrada até que a tentativa de aprendizagem seja completada. O grau de similaridade para padrões em um mesmo cluster é controlado por um parâmetro especificado pelo usuário, conhecido como parâmetro de vigilância [Fauset, 1994].

Vantagens e limitações

Algumas das vantagens das redes ARTs são:

1. Teoremas [Carpenter & Grossberg, 1987] mostram que as redes ARTs exibem considerável estabilidade e não são perturbadas por dados espúrios.
2. A rede adapta-se para refletir os tipos de padrões mais frequentemente observados no meio, atualizando os protótipos das categorias de forma adequada.
3. A arquitetura da ART pode facilmente ser integrada com outras teorias hierárquicas de cognição.
4. Teoria da ressonância adaptativa.

A arquitetura da ART é suposta manter a plasticidade requerida para novos padrões, enquanto previne a modificação de padrões que foram apresentados à rede anteriormente. Contudo, para a rede ART1 o problema da estabilidade não está completamente resolvido. Se a rede recebe um número de variações sobre um padrão de entrada armazenado, a rede pode gradativamente mudar em uma dada direção no espaço dos padrões. Cada variação pode assemelhar-se ao protótipo da categoria armazenada previamente, aproximando-se o suficiente para ser substituída dentro da mesma categoria. Isto resultaria em uma mudança suficientemente grande, de modo que a rede não possa mais reconhecer o padrão original. Outro problema da rede ART é a sensibilidade à ordem com o qual os padrões são apresentados a rede [Kung, 1993].

O grau que a ART usa para discriminar classes distintas de padrões de entrada pode ser variado por uma escolha adequada do valor do parâmetro de vigilância. Então a granularidade com o qual os padrões de entradas são classificados é determinado pelo fator de vigilância. Um valor grande causaria uma discriminação mais fina entre as classes para um dado conjunto de padrões de entrada (ou seja, o número de categorias formadas é maior). Por outro lado, um valor menor deste parâmetro permitiria que padrões com ruídos fossem classificados dentro da mesma categoria.

Aplicações

Redes ARTs estão sendo usadas em grande número de aplicações importantes, incluindo diagnóstico, processamento de voz e imagens, controle, avaliação de riscos. Para exemplificar pode-se citar:

- Classificação [Carpenter & Grossberg, 1987]
- Diagnóstico de falhas [Kalkunte et al, 1992]
- Análise do mercado de ações [Szu et al., 1992]
- Sistema de recuperação de Informações [Patterson, 1995]
- Processo de Monitoramento e Controle [Whiteley & Davis, 1993].
- Reconhecimento de Caracteres [Gan & Lua , 1992].

4.3.5 Rede de Hopfield

Em 1982, o físico americano J.J. Hopfield trouxe as Redes Neurais Artificiais (RNA) de volta da aversão do qual elas tinham sofrido nos anos setenta e início dos anos oitenta. Seu artigo “*Neural Networks and Physical Systems with Emergent Collective Computational Abilities*” [Hopfield, 1982] foi um marco para emergir a nova era das pesquisas de RNA, introduzindo a função de transferência não linear para a avaliação da saída final dos neurônios.

A rede de Hopfield performa uma das tarefas mais interessantes que o cérebro é capaz de fazer: auto-associação, por meio do qual uma imagem armazenada (ou qualquer outra informação armazenável através de um vetor ou uma matriz multidimensional) é regenerada de dados parciais ou corruptos.

Hopfield dedicou sua atenção a duas propriedades da interconexão das células de sistemas não lineares simples: primeiro, que tais sistemas têm estados estáveis que sempre serão alcançados se a rede for iniciada em um estado similar e, segundo, o fato que tais estados podem ser criados pela mudança do comprimento das interconexões entre as células.

A análise é feita baseada em uma definição de “energia” na rede e a prova que a rede opera para minimizar esta energia quando instalado dentro do padrão estável. Para alcançar um mínimo, a função de energia deve diminuir sempre que um neurônio muda de estado.

Hopfield ilustrou uma aplicação clara na Física (modelo de spin-glass) na minimização da superfície de energia para encontrar soluções estáveis no padrão da rede de ativação. Esta aplicação contribuiu para uma compreensão formal matemática que seria usada para estabilizar uma base de trabalho renovada das RNAs.

As redes de Hopfield são redes recorrentes de camada única com matriz de pesos simétrica no qual os elementos da diagonal principal são todos nulos. Um exemplo de uma rede de Hopfield é dado na Figura 4.10.

Estas redes armazenam algum número P de padrões protótipos $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^P$ chamados atratores de ponto-fixo. A localização dos atratores são determinadas pela matriz de pesos W . Os padrões armazenados podem ser especificados calculando-os diretamente tal como na aprendizagem de Hebb ou podem ser aprendidas através de algum esquema de atualização gradiente descendente tal como a Regra Delta.

Uma vez que a rede aprendeu os P padrões protótipos, eles podem ser usados para um lembrar associativo. Para recordar um padrão \mathbf{x}^k , a rede opera recursivamente pela alimentação dos sinais de saídas da rede de volta às entradas repetidamente em cada tempo t até a rede finalmente estabilizar. Iniciando em algum estado inicial arbitrário, um sinal de entrada $\mathbf{x}(0)$ é “clamped” dentro das conexões da rede no tempo 0, e as saídas são então calculadas pelas unidades. Para sistemas de tempo discreto, as saídas são determinadas pela equação:

$$x_i(t+1) = \text{sgn}\left(\sum_{j=1}^n w_{ij} x_j(t) - \theta\right) \quad (4.5)$$

para $i = 1, 2, \dots, n$, onde o limiar θ é uma constante não-negativa e onde usa-se a função de ativação bipolar, $x_i \in \{-1, 1\}$ com

$$\text{sgn}(x) = \begin{cases} +1 & \text{se } x > 0 \\ -1 & \text{se } x < 0 \end{cases} \quad (4.6)$$

e por convenção

$$x_i(t+1) = x_i(t) \quad \text{se} \quad x = 0. \quad (4.7)$$

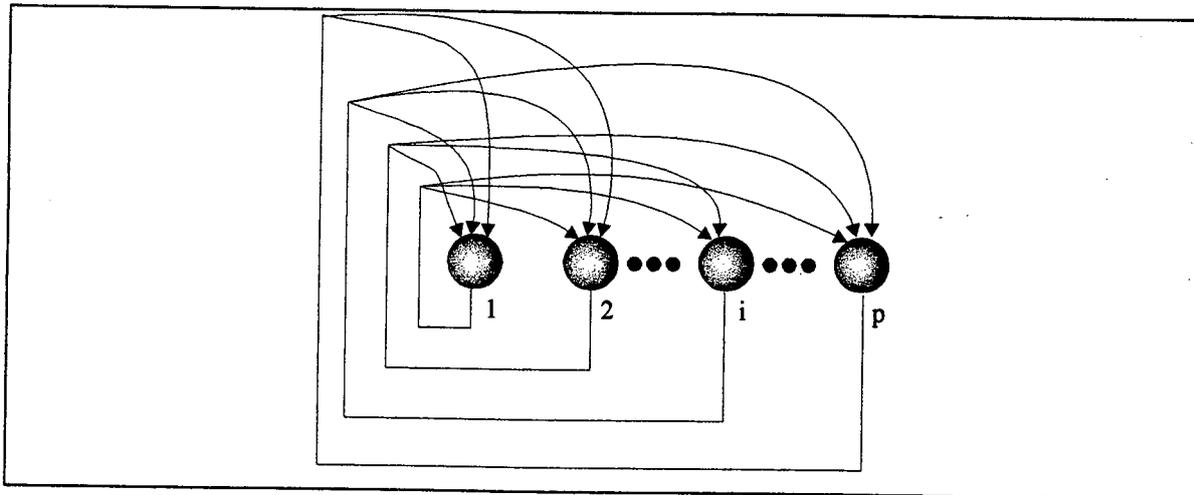


Figura 4.10 : Rede de Hopfield Discreta [Fauset, 1994].

Começando com um vetor x como entrada, as saídas são calculadas de acordo com a equação (4.5) e usadas como entradas (*feed back*) através de algum esquema de recursão/adaptação. Novas saídas são então calculadas e novamente usadas como às entradas no próximo tempo. Este processo é repetido recursivamente que até a rede estabilize em um ponto fixo, correspondendo a um padrão aprendido.

Durante o relembrar, uma rede Hopfield genérica pode alcançar um dos dois estados: (1) um ciclo, no qual para algum t suficientemente grande e um período fixo $T > 1$, $x(t+T) = x(t)$ ou (2) um ponto fixo definido por $x(t+1) = x(t)$ para t suficientemente grande.

Para funcionar como uma memória associativa, a rede deveria convergir para algum ponto fixo próximo do vetor de entrada $x(0)$ após algum número finito de iterações. Este seria o caso para uma matriz de pesos W simétrica.

Caracterização da Função energia

Um dos aspectos mais interessantes das Redes Hopfield é a caracterização do estado da rede com uma função energia. Devido ao fato que a matriz dos pesos nesta rede ser simétrica é possível definir uma função energia E , onde

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j \quad (4.8)$$

Esta função é similar a função energia que caracteriza materiais magnéticos em física, usando um simples modelo “spin” onde os átomos do material podem assumir uma das duas orientações + ou - (para cima ou para baixo). Esta analogia física pode ser usada para provar várias propriedades das Redes Hopfield já que existe um isomorfismo entre o modelo “spin” e as redes.

Hopfield foi capaz de provar que, como os sistemas de RNA evoluem de acordo com suas dinâmicas, a energia deve eventualmente alcançar um estado estável, pois a definição da função energia E não pode crescer após cada adaptação [Hopfield, 1982]. Ela deve decrescer ou pelo menos permanecer a mesma. Devido ao fato que existem um número finito de estados, a rede deve eventualmente convergir para um mínimo local. A energia mínima corresponde a um atrator de ponto-fixa. O estado do sistema na convergência determina o padrão de saída. Ele, na verdade, depende do estado inicial da rede e da matriz de pesos W . Para W fixo, todos os estados iniciais dentro de uma certa distância de um ponto atrator formam a chamada “bacia de atração” como ilustrado na Figura 4.11. De qualquer estado inicial determinado por um padrão de entrada, o sistema evolui movimentando-se abaixo da superfície de energia até um mínimo local a ser alcançado.

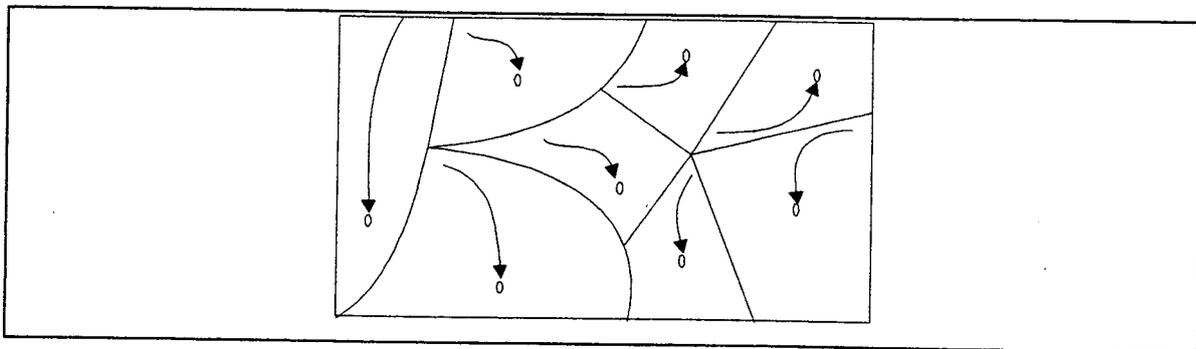


Figura 4.11 : Atratores de pontos fixos e bacias de atração [Patterson, 1995].

A discussão acima aplica-se a Rede de Hopfield discreta. A rede foi generalizada para operar em tempo contínuo e produzir saídas com valores contínuos.

A versão contínua da Rede Hopfield [Hopfield, 1984] é uma generalização direta da discreta através do uso de funções de ativação contínuas no lugar da função bipolar. Tipicamente a função de ativação sigmóide ou tangente hiperbólica é usada. A dinâmica da rede é modelada em tempo contínuo e a função de energia é dada por:

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_j x_i + \sum_{i=1}^n \int_0^{x_i} f^{-1}(x) dx \quad (4.9)$$

Vantagens e Limitações

Em cada uso das redes de Hopfield para memória associativa e otimização os problemas de aplicação são resolvidos quando a rede alcança um estado de equilíbrio em um mínimo de energia.

A rápida capacidade computacional da rede é sua maior vantagem. Ela surge como consequência da natureza paralela de seu processamento de convergência.

A rede apresenta também uma série de limitações:

- *Capacidade de memória:* Um “esquecimento” catastrófico pode ocorrer se tentarmos memorizar mais padrões do que a rede pode suportar. O número m de padrões de treinamento deve ser aproximadamente o número de neurônios n , ou menor. Isto significa que a capacidade de memória de uma rede Hopfield é severamente limitada.
- *Limitações de discrepância:* O novo padrão a ser reconhecido como os dos padrões de treinamento não deveria diferir de qualquer padrão de treinamento por mais que 25%.
- *Ortogonalidade entre os padrões:* Quanto maior a ortogonalidade (dissimilaridade) entre os padrões de treinamento melhor o reconhecimento.
- *Pesos simétricos:* A matriz de pesos tem de ser simétrica a fim da rede alcançar um equilíbrio. Os pesos sinápticos simétricos não são todos plausíveis biologicamente, mas são uma limitação útil aqui.
- *Problema de mínimo local.* Um desvantagem maior na rede de Hopfield é que ela pode repousar em um mínimo local em vez de um mínimo global do estado de energia, assim associa-se um novo padrão de entrada com um estado espúrio.

Aplicações

Pode-se citar aplicações desta rede em reconhecimento de padrões, processamento de palavras, restauração de base de dados e processamento de imagens.

A versão contínua das redes de Hopfield foi usada primariamente em aplicações de otimização tal como o problema do caixeiro viajante, *scheduling* e otimização de funções. Outras aplicações incluem processamento de imagens [Bilbro et al., 1988], e controle [Patterson, 1995].

Quando cria-se uma rede para resolver um problema de otimização, os pesos e o bias são usados de modo a expressar as restrições do problema. Encontrar um bom conjunto de equações para as restrições é a chave para encontrar uma solução com sucesso. Geralmente, a solução da rede não será ótima, somente próxima do ótimo, mas útil para muitas situações.

4.3.6 Perceptron Multicamadas

O conceito de perceptron [Roseblatt, 1958] foi um dos desenvolvimentos mais empolgantes durante o início das redes neuronais artificiais. É uma rede de processadores elementares (arranjados de um modo que recorda as redes neuronais biológicas) que podem aprender a reconhecer e classificar padrões automaticamente. Os processadores são elementos simples arranjados em uma camada. Neste perceptron clássico de uma única camada, dada duas classes de padrões, tenta encontrar um limite de decisão linear que separe as duas classes. Se os dois conjuntos de padrões são linearmente separáveis, o algoritmo perceptron com certeza encontra um hiperplano que separa as duas classes em um número finito de passos. Contudo, se o espaço dos padrões não é linearmente separáveis, o perceptron falha.

A demonstração das limitações de uma rede neuronal com uma única camada foi um fator significativo no declínio do interesse por redes neuronais nos anos setentas. O descobrimento (por diversos autores independentemente) e propagação de um método efetivo de treinamento para uma rede neuronal multicamadas [Rumelhart et al., 1986] teve um papel importante no ressurgimento das redes neuronais como uma ferramenta para resolver uma ampla variedade de problemas. Neste tópico

discuti-se uma das arquiteturas mais populares e versáteis, as redes *feed forward* multicamadas com aprendizado *backpropagation*. Este tipo de rede é também chamada de Perceptron Multicamadas ou simplesmente *backpropagation*. Este método de treinamento, conhecido como *backpropagation* ou *regra delta generalizada* é um método gradiente descendente (correção de erro) simples para minimizar o erro quadrado total da saída da rede.

Uma rede MLP consiste de uma camada de entrada, pelo menos uma camada intermediária ou escondida, e uma camada de saída, os neurônios de cada camada estão completamente interconectados (em algumas aplicações particulares, parcialmente interconectados) com todos os neurônios da próxima camada (Figura 4.13 - lado esquerdo).

As MLPs foram colocadas em prática somente quando algoritmos de aprendizagem foram desenvolvidos para elas, um deles foi o tão conhecido algoritmo *backpropagation* [Werbos, 1990; Rumelhart et al., 1986; e outros].

O termo *backpropagation* surgiu após 1985. Contudo, a idéia básica do *backpropagation* foi primeiro descrita por Werbos em sua tese de Ph.D. [Werbos, 1974]. Subseqüentemente, foi redescoberto por Rumelhart, Hinton, e Williams em 1986. Uma generalização similar do algoritmo foi desenvolvido por Parker em 1985, e um algoritmo de aprendizagem similar, foi também estudado por LeCun [Lecun, 1985].

O desenvolvimento do algoritmo *backpropagation* representa “*landmark*” em RNA em que fornece um método computacionalmente eficiente para o treinamento de perceptrons multicamadas. Embora não se possa dizer que o algoritmo *backpropagation* resolva todos os problemas (que tem solução), é bom ressaltar que ele acabou com o pessimismo sobre redes neurais, que pode ter sido inferido pelo livro de Minsky e Papert [Minsky & Papert, 1969].

De fato, tem se tornado tão popular que para muitos autores o termo “redes neurais” simplesmente significa o método *backpropagation*. A atratividade pelo método vem do conjunto de equações bem-definidas e explícitas para correção dos pesos. Estas equações são aplicadas através das camadas, começando com a correção

dos pesos da última camada (camada de saída), e então continuando para trás (por isto o nome) em direção da primeira camada (Figura 4.12).

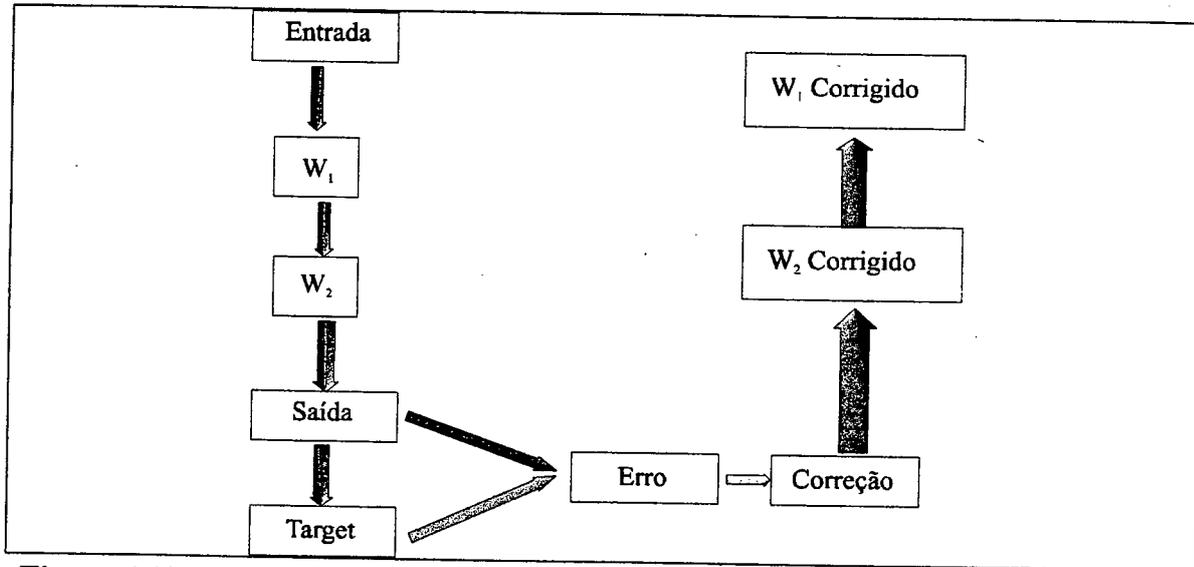


Figura 4.12 Apresentação esquemática da correção dos pesos por backpropagation.

Os neurônios na rede MLP tem valores de entrada e saídas contínuos, e a função de ativação não linear (geralmente usa-se a função sigmode). Uma regra de gradiente descendente pode ser usada para encontrar os pesos das conexões ótimas w_{ij} que minimizam o erro global E . Uma mudança em um peso Δw_{ij} em um ciclo $(t+1)$ ocorre na direção do gradiente negativo do erro E :

$$\Delta w_{ij}(t+1) = -\eta(\partial E / \partial w_{ij}(t)) \tag{4.10}$$

onde η é a taxa de aprendizagem. A regra do gradiente garante que após um número de ciclos, o erro E alcança um valor mínimo, ou o menor “platô”, se o erro E é representado como uma superfície no espaço dos pesos. Um erro global para o treinamento de uma entrada pode ser calculado como segue:

$$E = \sum_{(p)} \sum_{(j)} Err_j(p), \tag{4.11}$$

onde o erro pode ser dado, por exemplo, por $Err_j(p) = (y_j^{(p)} - d_j^{(p)})^2 / 2$.

A regra do gradiente descendente para mudar um peso entre os neurônios i e j pode ser expresso pela regra delta:

$$\Delta w_{ij}(t+1) = \eta Err_j y_i \quad (4.12)$$

ou alternativamente pela regra delta generalizada:

$$\Delta w_{ij}(t+1) = \eta Err_j g'(u_j) y_i, \quad (4.13)$$

onde Err_j é o erro entre o valor de saída desejado d_j e o valor y_j produzido pelo neurônio j que pode ser simplesmente expressado por $Err_j = |y_j - d_j|$. O valor de $g'(u_j)$ é a derivada $\partial g / \partial u$ da função de ativação g para o net de entrada u , para um valor particular de u_j ; e y_i é o valor de saída para o neurônio i . Quando a função de ativação g é uma função logística a derivada $g'(u_j)$ é expressada como $y_j(1-y_j)$. A fórmula acima pode ser simplificada como segue:

$$\Delta w_{ij}(t+1) = \eta Err_j d_j (1 - d_j) d_i \quad (4.14)$$

Cada ciclo de aprendizagem (um ciclo, também chamado época, é definido como o processo de propagação através da rede usando o conjunto de treinamento e o cálculo do erro E) constitui-se de dois passos: (1) um passo *forward*, quando as entradas são fornecidas e propagadas através das camadas intermediárias para a camada de saída; (2) um passo *backward*, quando um erro calculado na saída e propagado para trás para calcular as mudanças dos pesos. Esta é a maior caracterização deste algoritmo. Durante o passo *backward*, um erro Err_i para um nó interno i é calculado pela multiplicação dos erros Err_j de todos os neurônios j para os quais o neurônio está conectado pelos pesos correspondentes w_{ij} . Este erro é então usado *backward* para ajustar os pesos dos neurônios de uma camada anterior, conectada ao neurônio i . O processo de treinamento é repetido por muitas épocas com a mesma amostra de treinamento até que o erro global E seja suficientemente pequeno.

O cálculo do erro e as mudanças dos pesos podem também ser feitos de um modo *batch* (uma agregação ou um erro médio é calculado para todos ou algumas amostras de treinamento) ou em um modo individual (o erro é calculado e os pesos são alterados após cada amostra de treinamento).

Cada peso pode ter sua taxa de aprendizagem individual (regra *delta bar delta*) [Jacobs, 1988]. Se os pesos mudam alternativamente em sinal a taxa de aprendizagem deveria ser diminuída. Se as mudanças nos pesos é estável, a taxa de aprendizagem deve ser aumentada.

Existem muitas modificações e melhoramentos do algoritmo backpropagation, mas estes não serão explorados aqui [Haykin, 1994]. Estas modificações diferem nos seguintes pontos: cálculo do erro, função de ativação, forma de atualização dos pesos, número de épocas para adaptação dos pesos, e outros parâmetros.

Independente do algoritmo de treinamento usado para uma rede MLP, existem algumas características comuns da arquitetura MLP. Algumas delas são:

- As redes MLPs são aproximadores universais [Hornik, et al. 1989; e outros]
- MLPs são modelos de regressão não linear multivariada.
- MLPs podem aprender probabilidades condicionais.

Vantagens e limitações

As redes MLPs são capazes, pelo menos em teoria, de aproximar funções arbitrárias com acurácia desde que se forneça um número suficiente de camadas escondidas. Tais redes também são bons aproximadores de regressão não linear não paramétricas e são capazes de aprender mapeamentos desejados se um número suficiente de amostras de treinamento estão disponíveis.

As redes MLPs são provavelmente as redes mais usadas. Contudo existem diversos problemas. Alguns deles são citados abaixo:

- A questão de como *escolher no número de neurônios* nas camadas intermediárias, e uma questão mais geral, como *escolher a estrutura da rede*. Existem algumas heurísticas que tentam solucionar este problema [Haykin, 1996].

- *Esquecimento catastrófico*, fenômeno que representa a habilidade de uma rede esquecer o que ela aprendeu de amostras anteriores quando elas não são mais apresentadas, mas outras são. Isto é devido a mudança dos pesos de acordo com o novo padrão apresentado [Kasabov, 1996].
- *Overfitting*, fenômeno que indica que a RN aprendeu muito bem os dados de treinamento (decorou), que podem conter ruídos. Em tais casos a rede pode não generalizar os novos dados. Existem alguns caminhos para superar este problema: cessar o treinamento antes da rede chegar ao erro mínimo ou usar menos neurônios escondidos (mais neurônios escondidos leva a melhores aproximações) [Kasabov, 1996].
- A taxa de convergência do algoritmo de aprendizagem *backpropagation* tende a ser muito lento (ordem de convergência linear), o que o torna computacionalmente caro [Haykin, 1996].
- A superfície de erro pode influenciar na performance da rede quando existem mínimos locais além do mínimo global. A rede, neste caso pode convergir para um mínimo local.
- Existem problemas na escolha do número ótimo de camadas escondidas e de nós.

Aplicações

As redes MLP têm provado ser eficiente como ferramentas de mapeamento para uma ampla variedade de problemas, e conseqüentemente, têm sido usadas extensivamente e com sucesso pelos pesquisadores em vários domínios de aplicações, incluindo engenharia, direito, ciência da computação, controle, estatística, medicina, manufatura, transportes, finanças, telecomunicações, e muito mais. Elas podem competir com outras arquiteturas para a maioria das áreas de aplicações com possível exceção de categorias que dependem de uma aprendizagem não supervisionada. A ampla popularidade das redes MLP para tal gama de aplicações, naturalmente, vem de sua habilidade de fazer mapeamentos. Se tem-se um bom conjunto de dados de treinamento para alguma aplicação, é provável que uma rede MLP com uma ou duas camadas escondidas pode aprender os padrões para uma tarefa determinada. Embora sua performance ser tão boa quanto outras redes e outras técnicas (não neural) na

solução de diversos problemas, pode ser difícil treiná-las. Mas, em geral, quando existe dúvida em relação a qual arquitetura usar para uma dada aplicação, as redes MLP com *backpropagation* podem sempre ser consideradas. Isto se um conjunto de treinamento apropriado estiver disponível.

Devido a variedade de aplicações das redes MLP, são citados alguns exemplos onde estas redes foram usadas com sucesso:

Classificação e Diagnóstico: Classificação de células para Diagnóstico de Câncer [Moallemi, 1991] e Identificação de Falhas no Sistema Telefônico [Sone, 1993].

Controle e Otimização: Autonomously Driven Land Vehicle [Thorpe et al., 1991] e [Kanade et al., 1994] e Controlador Inteligente para Fabricação de Aço [Widrow et al., 1994].

Previsão e Predição: Previsão de Série Temporais Financeiras [Patterson et al., 1993], Previsão dos Retornos de Ações [Barr & Mani, 1994], Previsão de Séries Temporais Caóticas [Lapedes & Farber, 1988] e Predicting Creditworthiness for Loan Applications [Morose, 1990; Morose, 1993].

Reconhecimento de Padrões: Reconhecimento de Caracteres Manuscritos [Yanicoglu & Sandon, 1993], Detecção de Ataque Epilético [Hamilton & Hufnagel, 1992] e Identificação Automática de Indivíduos [Colombi et al., 1993].

4.3.7 Função de Base Radial (Radial Base Function - RBF)

As redes MLP com seus algoritmos de aprendizagem foram um passo gigante na área de RNAs. Elas influenciaram muito no desenvolvimento de outros modelos de redes que usaram algumas de suas idéias, mas que também geraram novas idéias. Uma destas redes é a Função de Base Radial.

As perspectivas (combinada com simplicidade) é a atração principal das RBFs no contexto das pesquisas de RNAs. De várias maneiras as redes RBFs agem como ponte, ligando uma variedade de áreas da ciência aparentemente diferentes.

A Função de Base Radial foi usada como um o artifício de cálculo desde o último século em termos da interação eletromagnética entre partículas carregadas. Esta analogia foi explorada pela Corporação NESTOR no início dos anos oitenta na

introdução da rede *Reduced Coulomb Energy* - uma arquitetura originalmente não supervisionada e com categorização que pode ser considerada como a precursora da rede RBF. Alguns dos princípios básicos das redes neuronais incorporado na Rede RBF tem emergido nas últimas duas décadas da neurobiologia na forma de modelos de memória (como o SOFM de Kohonen).

A incorporação da Função de Base Radial dentro da área de redes neuronais foi originalmente performada a fim de trazer algum tipo de interpretação do que significa “aprendizagem” e “generalização” em redes neuronais artificiais. A motivação primária foi dentro da teoria de aproximação de funções por um processo de interpolação, principalmente por Powell [Powell, 1985] e Micchelli [Micchelli, 1986].

Uma arquitetura geral das redes RBFs é mostrada no lado direito da Figura 4.13. Uma rede RBF consiste de 3 camadas. A primeira é a camada das p entradas. Elas são completamente interconectadas aos neurônios da segunda camada. Um nó escondido tem uma função de base radial (por isto o nome da rede) como uma função de ativação. Esta função é simétrica radialmente (por exemplo, função gaussiana):

$$f(x) = \exp\left[-(x - \mu)^2 / 2\sigma^2\right] \quad (4.15)$$

onde μ e σ são dois parâmetros significando a média e o desvio padrão da variável de entrada x . Para um nó intermediário particular i , sua RBF_i está centrada em c_i e é representada pelo vetor de pesos ($w_{1i}, w_{2i}, \dots, w_{pi}$) entre os p nós de entradas e o nó escondido i . O desvio padrão para este cluster define o range para a RBF_i . Uma função de base radial é não monotônica em constraste com a função sigmoideal. A segunda camada é conectada com a camada de saída. Os nós na saída realizam uma função soma simples com uma função de ativação de linear linear.

O treinamento de uma rede RBF consiste de duas fases: (1) ajustar as funções de base radial dos neurônios escondidos aplicando um método de clusterização estatística; isto representa a fase de aprendizagem não supervisionada; (2) aplicando um algoritmo gradiente descendente (como backpropagation) ou um algoritmo de

regressão não linear para ajustar as conexões da segunda camada; está é uma fase de aprendizagem supervisionada.

Durante o treinamento os seguintes parâmetros das redes RBFs são ajustados:

- A posição n-dimensional dos centros c_i das RBF_i. Isto pode ser encontrado usando o algoritmo de clusterização K-means; o algoritmo encontra k (número de neurônios escondidos) centros que minimizam a distância média entre as amostras de treinamentos e os centros mais próximos;
- o parâmetro de desvio σ_i para cada RBF_i; é definido usando distância média aos centros dos m-clusters mais próximos;

$$\sigma_i = \left(\frac{\sum_{q=1, \dots, m} \text{abs}(c_i - c_{iq})}{m} \right)^{1/2} \quad (4.16)$$

onde c_{iq} é o centro do p-ésimo cluster mais próximo do cluster i ;

- os pesos da segunda camada de conexões.

A Figura 4.13 contrasta as redes MLP e RBF. Primeiramente note a heterogeneidade dos neurônios da RBF (neurônios lineares e não lineares). A diferença primária entre as duas redes está na natureza de suas funções de ativação. Os nós nas camadas escondidas em uma MLP formam uma função de base sigmoideal que é diferente de zero sobre uma região infinitamente grande da espaço de entrada, enquanto a função de base em uma RBF cobre somente regiões localizadas pequenas. Enquanto alguns problemas podem ser resolvidos mais eficientemente com função de base sigmoideal outros são mais eficientes para função de base localizadas. Porém RBFs têm maior poder de clusterização local que o perceptron convencional (como a MLP).

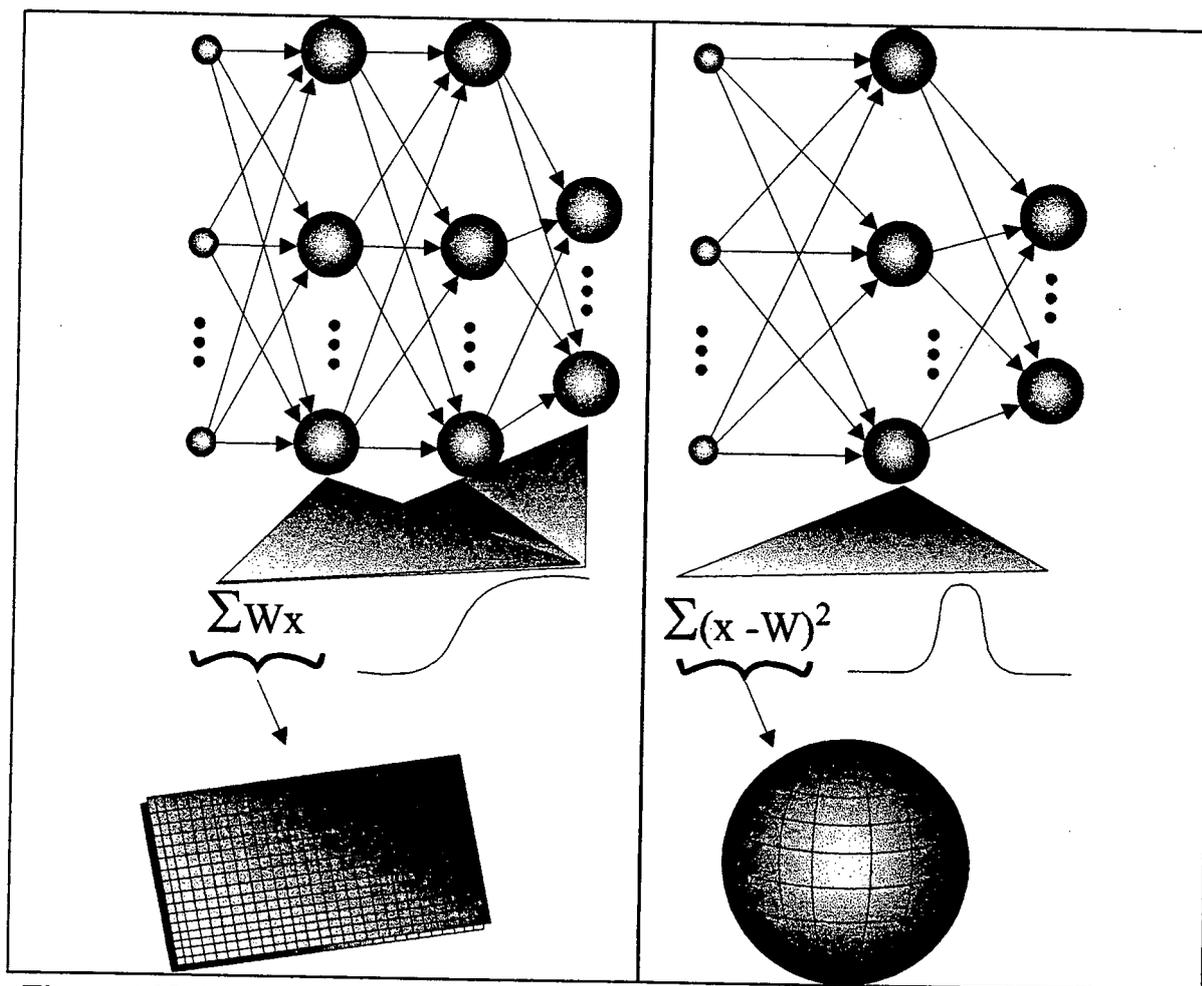


Figura 4.13: Estrutura das redes MLP e RBF [Pandya & Macy, 1995].

Vantagens e Limitações

Tanto a RBF quanto a MLP são exemplos de redes neuronais (feedforward) não-lineares. Ambas são aproximadores universais, sendo sempre possível uma RBF ser capaz de precisamente imitar uma específica MLP, ou vice-versa (Haykin, 1990). Contudo, as seguintes vantagens das redes RBFs sob as redes MLP com backpropagation tem sido experimentalmente e teoricamente provados:

1. O treinamento nas Redes RBFs são mais rápidos que o treinamento em um MLP com *backpropagation* de tamanho comparável;
2. Uma melhor generalização é encontrada nas redes RBFs;
3. As redes RBFs tem convergência muito mais rápida que as redes multicamadas convencionais com função e ativação sigmoidal, pois qualquer função pode ser

aproximadamente uma combinação linear de funções de bases fatorizáveis ajustada localmente.

4. Não existe o problema do mínimo local.
5. A rede RBF pode ser interpretada como um modelo conexionista difuso pois as funções de base radial podem ser consideradas como as funções de pertinências.
6. A camada escondida têm uma interpretação muito mais clara que as redes MLP com *backpropagation*. É mais fácil explicar o que uma rede RBF aprende que uma MLP com *backpropagation*.

Existem também as desvantagens de usar uma RBF, uma delas é encontrar o número apropriado de neurônios escondidos (centros). Uma aprendizagem não supervisionada pode ser necessária para encontrar o número de clusters. O número de neurônios escondidos é então setado igual a este número.

Aplicações

O range de aplicações das redes RBFs cobertos na literatura é muito amplo, como ilustrado pela seguinte lista representativa:

- Processamento de Imagens :[Saha et al., 1991]; [Poggio e Edelman, 1990]
- Reconhecimento de Voz [Ng e Lippmann, 1991]; [Niranjan e Fallside, 1990]
- Análise de Séries Temporais [He & Lapedes, 1991]; [Moody & Darken, 1989]; [Broomhead & Lowe, 1988]
- Equalização Adaptativa [Chen et al., 1992]; [Kassam & Cha, 1993]
- Diagnóstico Médico [Lowe & Webb, 1990].

4.4 Um método híbrido: FAN (Free Associative Neurons)

4.4.1 Introdução

Free Associative Neurons (FAN) é um exemplo de sistema híbrido. A base do FAN são as redes conexionistas, modelagem difusa e representação de padrões. FAN é baseado fracamente na noção do aprendizado por adaptação neuronal; utiliza técnicas de conjuntos difusos para promover a granularidade de informações; e

permite incluir diferentes métodos de associação de padrões para melhorar a sua capacidade de aprendizado [Raitz, et al. 1997].

4.4.2 FAN

A motivação para o desenvolvimento de FAN foi elaborar um método capaz de simular abstrações. Devido a esta capacidade cognitiva, os seres humanos compreendem um cenário completo baseado apenas nas partes que o compõem. De fato, na linguagem natural as pessoas resumem a informação, classificando objetos em estruturas de categorias de similares (palavras), evitando a sobrecarga de dados [Zadeh, 1996]. Em FAN, a noção de classificação de padrões é feita por três processos: (a) expansão do espaço de entrada; (b) projeção da vizinhança do padrão resultante; (c) quantificação e aprendizado do grau de similaridade entre as classes FAN e as projeções dos padrões.

Em FAN, cada padrão de entrada é expandido em uma vizinhança difusa. Cada conjunto suporte desta vizinhança é a combinação dos valores das características próximos dos originais. O grau de similaridade entre a vizinhança difusa e o padrão original de entrada é feita através das técnicas utilizadas na teoria dos conjuntos difusos [Klir & Yuan, 1995]. O aprendizado acontece através da projeção de toda a vizinhança difusa no espaço FAN. Existe uma unidade FAN para cada classe do domínio do problema. Cada unidade é um grid composto por todas as combinações de características observadas em sua classe correspondente (Figura 4.14). Durante o treinamento, cada combinação é representada por uma célula difusa que contém um peso correspondente à sua frequência de ocorrência e grau de pertinência. O treinamento é baseado no reforço na célula (se a classificação foi correta) ou em esquecimento (se houve uma classificação incorreta).

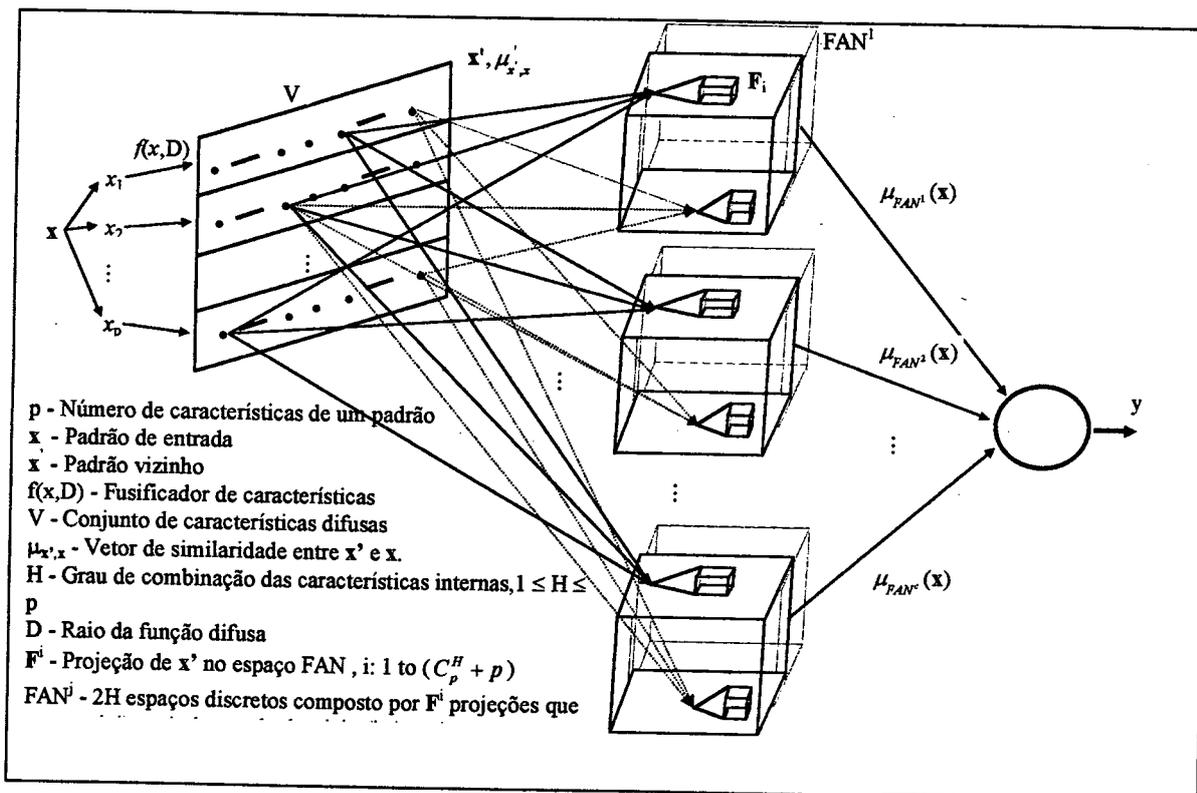


Figura 4.14: Arquitetura do FAN.

Aplicações

FAN foi aplicado em problemas práticos. Resultados excelentes foram obtidos em três problemas de classificação [Raittz et al., 1997]: (a) Os dados da IRIS de Fisher; (b) Classificação de cromossomos, segundo o grupo de Denver, com os dados de Copenhagen; e (c) Diagnóstico financeiro de empresas baseados em testes financeiros.

Características

A principal característica observada nestas aplicações é a capacidade de descobrir gradualmente as classes. FAN define as classes durante a primeira época de treinamento. Ou seja, ele reconhece uma classe antes mesmo de processar todo o conjunto de treinamento. Este comportamento parece indicar que o treinamento em FAN ocorre com crescente capacidade de generalização. As unidades FAN acumulam conhecimento durante o processo de treinamento de forma que as modificações nos pesos se tornam progressivamente menores. Em grandes conjuntos de treinamento, os

últimos padrões são vistos mais como dados de teste, causando modificações muito pequenas na rede FAN.

Os resultados obtidos usando FAN são comparáveis aos métodos tradicionais de redes neuronais, em termos de classificação e superior em termos de estabilidade no aprendizado e capacidade de generalização

4.5 Conclusões

Diversos modelos conexionistas foram introduzidos neste capítulo. Eles são caracterizados pelos tipos de neurônios usados, a organização de sua estrutura conexionista e pelos métodos de aprendizagem. Por esta razão, eles apresentam diferentes propriedades que podem ser usadas para muitos problemas distintos. Porém, todos os modelos são caracterizados por algumas características em comum: aprendizagem, generalidade, robustez e processamento massivamente paralelo.

Uma destas características, o processamento paralelo, torna as RNAs muito atrativas para realização de *hardwares*. Este processamento massivamente paralelo torna possível a construção de máquinas que são mais rápidas que o cérebro humano para resolver problemas difíceis.

Depois de mais de 10 anos em que as pesquisas sobre RNAs permaneceram em laboratórios, a década de oitenta viu uma explosão de interesse pelo domínio.

Hoje, aplicações vão desde reconhecimento de padrões até previsões financeiras. Todas estas aplicações são baseadas nas habilidades das RNs aprenderem de dados ou regras e generalizar sobre dados não vistos anteriormente. As redes fornecem uma aproximação de uma função objetivo que mapeia o espaço do domínio dentro de um espaço solução. As RNAs facilitam o mapeamento de dados aproximados, o que significa que elas podem trabalhar com dados ausentes, dados incompletos e dados corrompidos (com ruídos). As saídas das redes podem ter diversos significados, isto é, probabilidades, certezas, valores reais, conceitos simbólicos, categorias, etc..

Reconhecimento e classificação de padrões foi uma das primeiras aplicações das RNAs. Com efeito, o perceptron foi concebido principalmente como instrumento capaz de reconhecer letras. A principal razão é que reconhecimento e classificação de

padrões é uma tarefa geralmente desempenhada muito melhor usando a capacidade cognitiva do homem do que a execução de um algoritmo.

Os paradigmas mais comuns de aprendizagem no caso de reconhecimento de padrões é o supervisionado associado com uma rede multi-camadas (MLP e RBF). Devido a sua disponibilidade, a regra back-propagation é freqüentemente usada, assim como suas variantes. Entretanto bons resultados são obtidos também com aprendizagem competitiva de Kohonen. Este último é bastante usado quando não se sabe quantas classes possíveis existem a identificar. Também tem se usado, mas com menos freqüência, uma variação supervisionada da rede de Kohonen, a LVQ.

Neste capítulo foi apresentado um método de aprendizagem híbrida, o qual representa uma nova forma de reconhecimento de padrões. FAN é baseado no aprendizado conexionista e modelagem difusa. FAN é constituído por unidades independentes com capacidade autônoma de aprendizagem. O poder de aprendizagem de FAN é baseado na associação entre suas unidades e o uso de granularidade para representar informação. Os primeiros resultados obtidos usando FAN são comparáveis aos métodos tradicionais de redes neurais, em termos de classificação e superior em termos de estabilidade no aprendizado e capacidade de generalização.

5. APLICAÇÃO: CLASSIFICAÇÃO DA SAÚDE FINANCEIRA DE PEQUENAS EMPRESAS CATARINENSES

5.1 Introdução

Neste capítulo apresenta-se a aplicação desta dissertação: a classificação de pequenas empresas quanto a saúde financeira usando Redes Neurais Artificiais (Backpropagation, RBF e LVQ) e um Sistema Híbrido (FAN).

Estes modelos visam fornecer um “especialista” às pequenas empresas com menor custo. A relevância deste trabalho se deve a importância das pequenas empresas na economia nacional

Inicialmente, apresenta-se a definição e relevância do problema no campo financeiro e a escolha das variáveis financeiras (índices financeiros).

Após analisar o problema do ponto de vista financeiro, serão expostos os resultados obtidos nos treinamentos e nos testes das três rede neuronais e do FAN.

O estudo observa a eficácia de técnicas de IA que usa padrões para classificar a saúde financeira das pequena empresas e também faz uma comparação nos desempenhos dos quatro sistemas usados, levantando em consideração pontos fortes e fracos de cada um.

5.2 Aplicação

A fim de superar a ausência de um especialista nas pequenas empresas, podemos incluir o conhecimento especialista em um modelo computacional. Para isto, implementou-se três redes neuronais (MLP, RBF e LVQ) e um sistema híbrido (FAN). Estes sistemas modelam o primeiro estágio da análise da saúde financeira no

qual o especialista procura por desvios financeiros (diagnóstico). O segundo estágio, que poderia ser feito por um Sistema Especialista Difuso, modela o processo do especialista de verificação, entendimento e obtenção de uma solução para o problema diagnosticado. Neste trabalho explorou-se apenas o primeiro estágio.

Como o objetivo é a classificação de pequenas empresas quanto a situação da saúde financeira, a primeira questão a ser endereçada é a escolha de um conjunto adequado de índices que cubra suas atividades. Com estes índices pode-se chegar a classificação dos problemas financeiros a serem diagnosticados.

Dados das Empresas

Os dados financeiros empregados neste trabalho foram obtidos junto ao contador Sérgio Farraco (Presidente do Conselho Regional de Contabilidade). Os índices calculados foram extraídos do Balanço Patrimonial e do Demonstrativos de Resultados de 56 pequenas empresas catarinenses da região da Grande Florianópolis, dos anos de 1994 e 1995. Dentre estas empresas temos, postos, transportadores, distribuidoras, agências de turismo, entre outras.

No primeiro capítulo viu-se quatro abordagens para a classificação dos problemas financeiros: pragmática, dedutiva, indutiva e confirmatória. Porém, os índices usados para classificação depende do problema [Salmi & Martikainen, 1994], isto é, depende do objetivo da análise dos índices e o domínio onde será aplicado. Neste presente trabalho utilizou-se a abordagem prática de Lev [Lev, 1974], também considerada no trabalho de [Martins, 1996].

Das quatro categorias identificadas por Lev, foram escolhidos três como grupos representativo dos problemas financeiros que provavelmente são advertidos por relações financeiras no caso de pequenas empresas. As categorias escolhidas são: rentabilidade, endividamento e giro.

Índices de Rentabilidade (Retorno): O lucro é o principal estímulo do empresário e uma das formas de avaliação de êxito de um empreendimento. Os índices de retorno também chamados de índices de lucratividade ou de rentabilidade indicam qual o

retorno sobre o investimento, retorno sobre as vendas, retorno sobre o capital próprio, entre outros.

Aqui trabalhar-se-á com dois índices de retorno:

$$1. C_Ope = \frac{\text{Fluxo de Caixa Oper.}}{\text{Receita Bruta}} = \frac{\text{Lucro Líquido} + \text{IR} + \text{Depreciação}}{\text{Receita Bruta}}$$

- Este índice mede a capacidade da empresa gerar caixa.

$$2. LLiq = \frac{\text{Lucro Líquido}}{\text{Receita Bruta}}$$

- Indicador mais fiel da lucratividade.

Estes índices são calculados a partir de dados do Demonstrativo de Resultado de cada empresa.

Índices de Giro (rotação): Os índices de rotação constituem-se em categorias de elevada importância para o analista. Como vimos no capítulo um, o balanço da empresa mostra sua situação patrimonial em determinado momento, isto é, mostra a empresa de forma estática, sem refletir sua mobilidade, seu dinamismo. A empresa, em suas operações de compra, estocagem, vende e recebe num processo dinâmico e contínuo. Os índices de rotação têm grande contribuição na interpretação da liquidez e da rentabilidade da empresa, à medida que servem de indicadores dos prazos médios e rotação de estoques, recebimentos das vendas e pagamento das compras.

Será utilizado um índice de rotação:

$$3. CF = \text{Ciclo Financeiro} = \frac{NLCG}{\text{Receita Bruta}} \times 365^7$$

⁷ NLCG (Necessidade Líquida de capital de Giro) = (Ativo Circulante - Caixa) - (Passivo Circulante - Dívida)

Este índice é calculado a partir do Demonstrativo de Resultado e do Balanço Patrimonial.

Índices de Endividamento: Fornecem informações sobre o grau de financiamento e obrigações fixas da empresa e sua habilidade para satisfazer estas obrigações financeiras. São índices de muita importância pois indicam a relação de dependência da empresa com relação a capital de terceiros.

Serão usados dois índices de endividamento:

$$4. TIE = \frac{\text{Lajir}}{\text{Despesas Operacionais Financeiras}}^8$$

- Indica quantas vezes o lucro corresponde às despesas operacionais.

$$5. Endiv = \frac{\text{Dívida}}{\text{Capital Próprio}}$$

- Este índice deve manter um equilíbrio, pois um valor muito alto indica muita dívida e valor baixo indica pouco uso de alavancagem.

Estes índices são calculados a partir do Demonstrativo de Resultado e do Balanço Patrimonial.

Na Tabela 5.1, tem-se os cinco índices acima e o que eles indicam para uma empresa. No apêndice B encontram-se o histograma de cada um destes índices, onde pode-se notar claramente que estes índices não apresentam distribuições normais.

Das 56 empresas, 40 foram usadas para o treinamento dos modelos e 16 para o teste (Apêndice A). As empresas foram classificadas por Alejandro Martins em 7

⁸ Lajir (Lucro antes de juros e impostos) = Lucro Líquido + IR + Despesas Operacionais Financeiras.

categorias, de acordo com o(s) problemas apresentados pelos índices: rentabilidade(1), giro(2), endividamento(3), rentabilidade e giro(4), rentabilidade e endividamento(5), giro e endividamento(6), e nenhum dos anteriores(7).

Índice	Relevância	Problema relacionado
C_Ope	Poderoso indicador da capacidade de geração de fluxo de caixa pois mede a quantidade relativa de fundos originado por vendas e/ou produtividade.	Rentabilidade
LLiq	Indica a rentabilidade das vendas. Particularmente no Brasil, a margem é mais restrita. Um aumento excessivo da margem pode levar a perda do mercado enquanto reduções intensas podem levar a problemas financeiros.	Rentabilidade
CF	Indica (a) quantos dias do faturamento são necessários para financiar as necessidades de capital de giro (b) quão sensível está o fluxo de caixa a mudanças nas vendas ou lucros.	Rotatividade
TIE	Mede a capacidade de lucro da empresa em correspondência com sua posição de alavancagem financeira.	Endividamento
Endiv	Retrata o posicionamento da empresa com relação aos capitais de terceiros. Indica a estrutura do capital da empresa.	Endividamento

Tabela 5.1: Importância de cada índice financeiro.

Na Tabela 5.2 encontram-se os índices financeiros de duas empresas A e B e o(s) problema(s) que estes índices detectam.

A empresa A, por exemplo, tem capacidade de gerar caixa, refletindo em lucro, giro baixo e pouco endividamento (pouco uso de alavancagem financeira).

Empresa	C_Ope (%)	LLiq (%)	CF (dias)	TIE (u)	Endiv (u)	Categoria
A	53.0	50.2	48.6	141	24	3
B	1.2	0.9	38.1	231	109	5

Tabela 5.2 : Índices Financeiros de duas empresas e a categoria a qual eles pertencem.

Já, na empresa B temos um exemplo um pouco mais complexo. A empresa tem pouca geração de caixa, pouco lucro líquido, pouco giro e baixo endividamento.

Como pode-se perceber nos exemplo, nem sempre a classificação é um problema simples. Muito pelo contrário, é uma tarefa bastante complexa, pois a

interpretação dos índices estão relacionados entre si, ou seja, a interpretação de um índice depende dos outros índices.

5.3 Diagnóstico através de RN

Desde 1968, a principal abordagem usada, para identificar problemas financeiros nas empresas, é a técnica ADM (Análise Discriminante Múltipla). Contudo esta técnica tem sido agudamente criticada pois a validade de seus resultados dependem de suposições restritivas sobre normalidade, separabilidade, normalidade multivariada, e independência das variáveis preditivas [Ohlson, 1980; Odom & Shardo, 1990; Karels & Prakash, 1987]. Índices financeiros, na grande maioria, violam as suposições da ADM. As restrições de ADM são incompatíveis com a natureza complexa, limitada e o inter-relacionamento dos índices financeiros. Então o poder da ADM para análise de índices financeiros está comprometido. Porém, nenhuma técnica melhor foi desenvolvida até o desenvolvimento das Redes Neurais Artificiais.

Este estudo foi motivado pelo fato que uma análise dos mesmos índices financeiros usado por Altman, para o mesmo objetivo, é possível sem qualquer uma das suposições da ADM [Lacher et al, 1993]. Uma rede neuronal é uma forma de processar conjunto de informações de entradas e suas saídas associadas (ou respostas) de tal maneira a convergir num padrão que satisfatoriamente reflete o relacionamento de entrada-saída.

O uso de RNA para este problema segue também de sua habilidade de analisar e modelar dados de negócios. A habilidade de generalizar dos dados de treinamento capacita a rede prever os novos dados subsequentemente apresentado ao sistema. As redes são também apropriada para este problema devido a falta de entendimento dos princípios e regras sobre previsão de falência e: a rede é não-paramétrica e faz suposições mais delicadas a respeito da distribuição dos dados de entrada do que métodos estatísticos frequentemente utilizados; uma RNA é capaz de formar fronteiras de decisão altamente não-lineares no espaço de características [Haykin, 1994].

Além disto, apresentam todas as características vistas no capítulo anterior, como tolerância com dados ruidosos e faltosos.

Como viu-se no capítulo 4, existem várias arquiteturas de redes. Cada uma é mais adequada para determinado tipo de tarefa. Como o problema é de classificação (de padrões complexos - padrões não linearmente separáveis), usou-se 3 das redes mais usadas para o problema de classificação: MLP com backpropagation, RBF, LVQ e o sistema híbrido FAN, analisando a performance de cada um. Estas redes foram escolhidas em virtude da capacidade de trabalhar com padrões complexos.

5.3.1 MLP com Backpropagation

A elaboração desta rede foi baseada na toolbox sobre redes neuronais do MATLAB [Demuth & Beale, 1994]. Foram usadas as funções **INITFF**, **TRAINBPX** e **SIMUFF** (ver Tabela 5.1).

Função	O que realiza?
INITFF	Inicia os pesos e bias para a rede backpropagation.
TRAINBPX	Treina a rede usando momentum e taxa de aprendizagem adaptativa. Treinamento mais rápido do backpropagation.
SIMUFF	Simula a rede (teste).

Tabela 5.1 : Funções principais para a rede backpropagation do MATLAB.

A rede apresenta uma única camada escondida, com 5 neurônios na camada de entrada e 1 na camada de saída. Os dados de entradas foram normalizados.

Foram realizados inúmeros testes variando os parâmetros (taxa de aprendizagem e momentum⁹), a função de ativação, o número de neurônios na camada escondida e o número de épocas. Os dados de entradas também foram trabalhados de duas formas distintas: - normalizados no intervalo [-1, 1] e - normalizados no intervalo [0, 1]. Porém, como poder ser visto no apêndice C, a rede

⁹ O momentum pode ser visto como um supervisor na aprendizagem: acelera a convergência para interações que conservam sinal (sinal do ajuste anterior e sinal do gradiente corrente); diminui a oscilação quando os sinais são diferentes; previne mínimos locais.

apresentou uma performance melhor quando os dados de entrada foram normalizados no intervalo $[-1,1]$.

No apêndice C encontram-se as tabelas C-1, C-2, C-3 e C-4 que mostram vários resultados obtidos com a rede backpropagation.

Para ilustrar a performance da rede, alguns resultados são apresentados na Tabela 5.2.

Como pode-se observar nesta tabela, existem várias combinações de parâmetros que fornecem os mesmos resultados. Apesar dos dados serem bastante reduzidos, a rede backpropagation conseguiu alcançar uma boa generalização. Um fator essencial para obtenção dos resultados melhores foi a utilização da combinação de funções de ativação não linear na camada escondida e linear na camada de saída.

Neurônios	Épocas	Taxa de Aprendizagem	Momentum	Fç. De Ativação	Treinamento (%)	Teste (%)	Tempo (s)
7	2000	0,100	0,90	sig/lin	87,50	68,75	27,90
7	2500	0,100	0,90	sig/lin	87,50	75,00	34,76
7	3000	0,100	0,90	sig/lin	90,00	81,25	41,52
7	3000	0,001	0,90	tan/lin	85,00	81,25	40,26
6	3000	0,001	0,90	tan/lin	85,00	75,00	38,55
9	3000	0,001	0,80	tan/lin	87,50	75,00	42,13
7	1000	0,010	0,90	tan/sig	35,00	31,25	12,96
20	1000	0,010	0,90	tan/sig	35,00	31,25	18,62
7	1000	0,900	0,10	tan/sig	35,00	31,25	12,97

Tabela 5.2 : Resultados da MLP com backpropagation usando os dados normalizados no intervalo $[-1,1]$.

A taxa de aprendizagem e *momentum* também têm uma contribuição especial. Pôde-se observar durante os teste realizados que: - quando a taxa de aprendizagem é pequena (por exemplo, 0.01) e o *momentum* é grande (0.9) a aprendizagem é rápida e a generalização é boa; - quando a taxa de aprendizagem é grande (por exemplo, 0.9) e o *momentum* é pequeno (0.1) a aprendizagem é mais lenta mas a generalização também é boa; - agora, quando as duas taxas estão próximas a generalização é razoável e a aprendizagem da rede é oscilante. O melhor resultado (81,25%) foi obtido quando a taxa de aprendizagem era pequena (0.001) e o *momentum* grande (0.9).

Quanto ao número de neurônios na camada escondida para obter bons resultados varia de 7 a 11 neurônios. Por isto a maioria dos teste foram feitos usando 7 neurônios, visando maior velocidade.

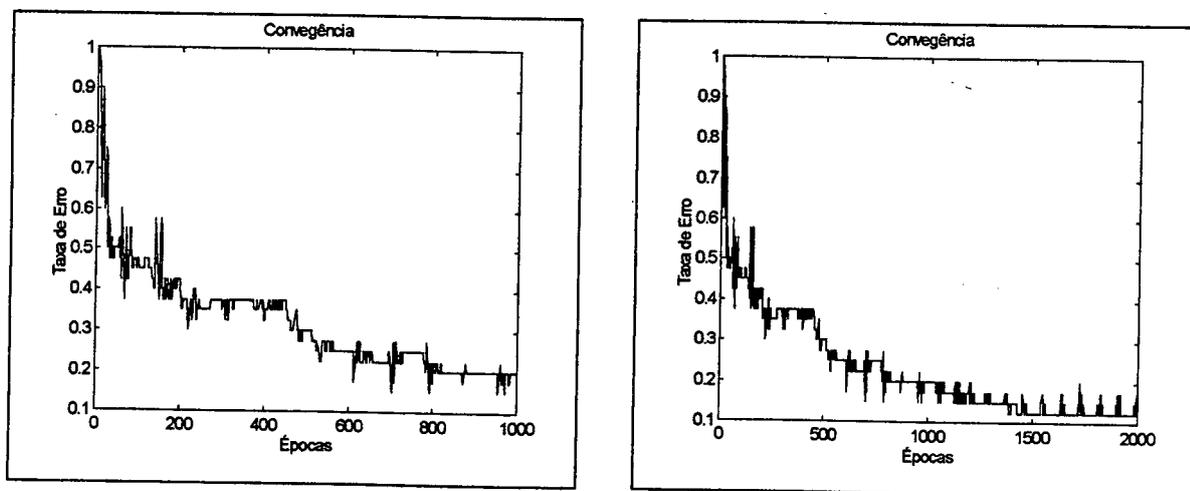


Figura 5.1 : (a) Convergência da Backpropagation usando 8 neurônios, taxa da aprendizagem 0.08, momentum 0.8 e 1000 épocas. Com esta arquitetura obtemos 81.25% de acerto nos dados de teste; (b) Convergência da Backpropagation usando 8 neurônios, taxa da aprendizagem 0.08, momentum 0.8 e 2000 épocas. Com esta arquitetura obtemos 68.75% de acerto nos dados de teste.

Um dos problemas da Backpropagation é a falta de estabilidade em relação a generalização. As Figuras 5.1 (a) e (b) mostram os resultados da convergência da rede com 1000 e 2000 épocas. Como pode ser visto a taxa de erro é decrescente, porém, a rede generalizou melhor com 1000 épocas (81.25%) do que com 2000 épocas (68.75%). Isto mostra o problema de *overtraining*, mencionado no capítulo anterior.

Os resultados também dependem dos pesos iniciais. Tem-se, por exemplo, uma classificação de 75% de acertos com 400 épocas. Mas o melhor resultado 81.25% foi obtido com 3000 épocas (mais comum) e até com apenas 1000 épocas (devido a inicialização dos pesos).

5.3.2 RBF

A elaboração desta rede também foi baseada na *toolbox* sobre redes neuronais do MATLAB [Demuth & Beale, 1994]. Foram usadas as funções **SOLVERB** e **SIMURB** (Tabela 5.1).

Função	O que realiza?
SOLBERB	Encontra uma RBF com centros suficiente para ajustar uma função dentro da meta de erro (0.02) e treina esta rede.
SIMURB	Simula a RBF.

Tabela 5.1 : Funções principais para a rede RBF do MATLAB.

Da mesma forma que a backpropagation, temos uma rede RBF com cinco neurônios na camada de entrada e um na camada de saída. A função de ativação usada na camada escondida foi a gaussiana.

A performance da rede depende dos parâmetros (centro e largura da função gaussiana) e é bastante sensível ao número de épocas. No apêndice D encontra-se os resultados obtidos combinando os parâmetros de formas distintas.

Como pode ser observado nas Figura 5.2 (a) e (b), se o número de épocas for maior que o número de centros, a rede aprende muito melhor, porém perde a generalidade, apresentando o mesmo problema de *overtraining* da Backpropagation. Mas na Backpropagation, como depende dos pesos iniciais, é muito difícil, senão impossível, determinar o número de épocas ideal. Isto já não acontece com a RBF.

Outro fator importante que influencia na performance da rede é a forma de normalização dos dados. Usando os dados normalizados entre [-1,1] a rede consegue generalizar melhor que quando normalizados entre [0,1] (ver Apêndice D).

Dentre os teste realizados, os resultados melhores foram obtidos usando 14 ou 15 centros com raio de 1.3, e trabalhando com os dados normalizados no intervalo [-1,1].

Se o número de centros for muito grande (30, por exemplo) a rede aprende 100% porém perde a generalidade.

Centros	Raio	Épocas	Treinamento (%)	Teste (%)	Tempo(s)
13	1,30	12	65,00	68,75	1,38
14	1,30	13	77,50	81,25	1,32
15	1,30	14	77,50	81,25	1,38
16	1,30	15	82,50	75,00	1,38
15	1,30	34	100,00	37,50	3.19

Tabela 5.2 . Resultados da RBF trabalhado com os dados normalizados no intervalo [-1,1].

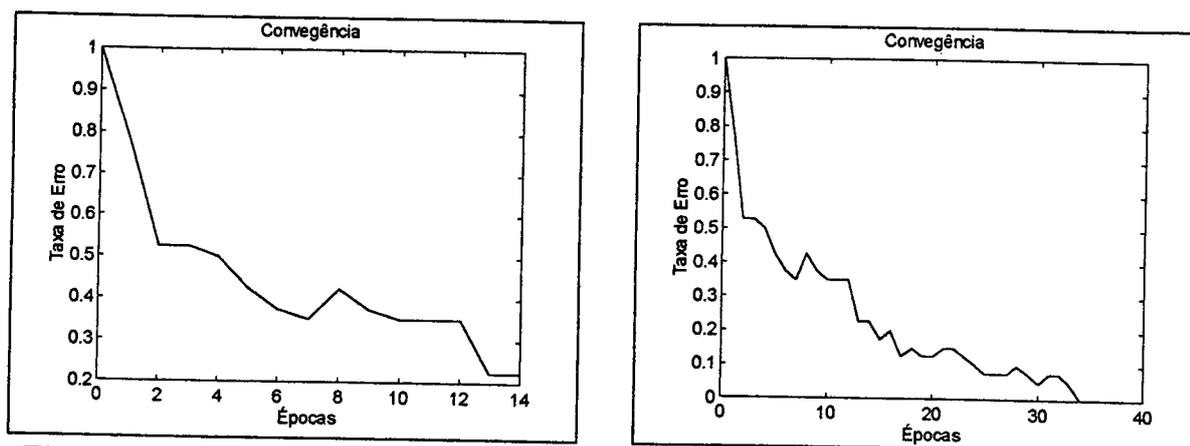


Figura 5.2: (a) Convergência da rede RBF com 15 centros, 1.3 de raio e 14 épocas. Com esta arquitetura rede acertou 77.5% no treinamento e 81.25% teste; (b) Convergência da rede RBF com 15 centros, 1.3 de raio e 34 épocas. Com esta arquitetura rede acertou 100% no treinamento e 37.50% teste.

5.3.3 LVQ

O algoritmo da LVQ foi implementado em MATLAB baseado no algoritmo encontrado no livro da Fausett [Fausett, 1994].

O pesos foram iniciados com os primeiros dados do conjunto de treinamento. Foram feitos inúmeros testes usando os dados normalizados no intervalo [-1,1] e não normalizados. Como pode ser visto na apêndice E, com os dados normalizados a rede apresentou uma performance um pouco melhor.

Testou-se duas formas distintas de variar (diminuir) a taxa de aprendizagem: linear e não linearmente em função do número de épocas. Com as duas formas obtém-se bons resultados (62.50%). Porém, quando usou-se a forma não linear, a rede

aprende e generaliza seu limite e não perde a generalidade mesmo quando muitas épocas são realizadas (ver Figura 5.3 (a) e (b)).

Taxa de Aprendizagem	Varição da Taxa da Apreñ	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
0,10	Não linear	45	62,50	50,00	24,450
0,20	Não linear	45	72,50	62,50	18,950
0,20	Não linear	80	72,50	62,50	33,170
0,20	Não linear	200	70,00	62,50	87,280
0,15	Não linear	40	60,00	50,00	6,480
0,20	linear	45	80,00	62,50	23,78
0,20	linear	100	15,00	31,25	53,82
0,10	linear	50	72,50	62,50	25,98

Tabela 5.1 : Resultado da LVQ com dados normalizados no intervalo [1,1].

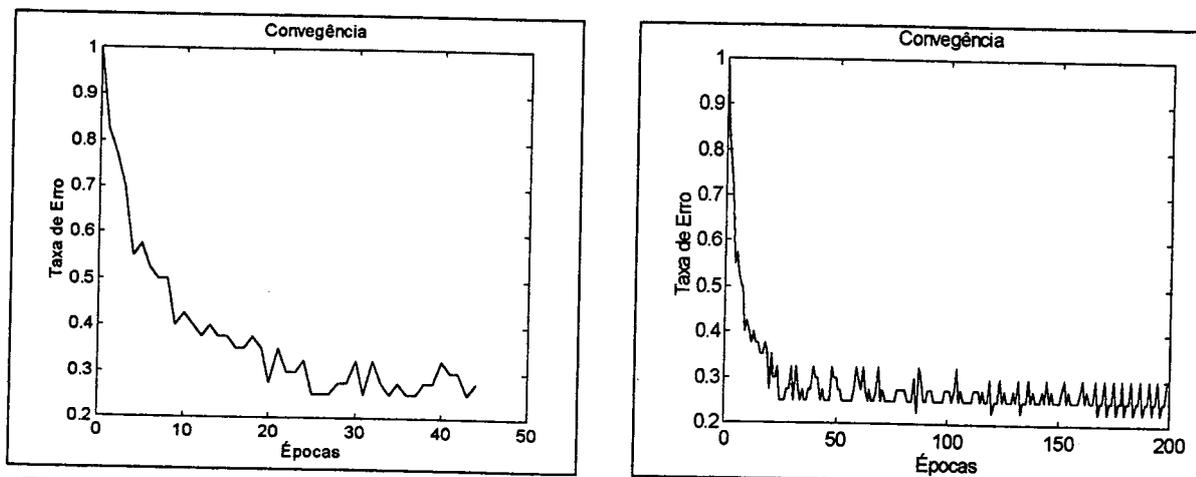


Figura 5.3: (a) Convergência da rede LVQ, trabalhando com os dados normalizados no intervalo [-1,1] com taxa de aprendizagem 0.2 decrescendo não linearmente e 45 épocas. Com esta arquitetura rede acertou 72.5% no treinamento e 62.25% no teste; (b) Convergência da rede LVQ, trabalhando com os dados normalizados no intervalo [-1,1] com taxa de aprendizagem 0.2 decrescendo não linearmente e 200 épocas. Com esta arquitetura rede acertou 70.00% no treinamento e 62.25% no teste.

5.3.4 FAN

O algoritmo do FAN foi implementado por Raittz [Raittz et al., 1997] usando clipper. Este algoritmo híbrido ainda está em fase de testes.

O diagnóstico realizado com o sistema híbrido apresenta também suas sensibilidade com alguns parâmetros, normalização das variáveis (range), raio de

decomposição (raio da função de fusificação - D) e grau de combinação das variáveis (H). Nos teste realizados a grau de combinação usado foi 1.

Alguns dos resultados obtidos usando o modelo FAN são apresentados na Tabela 5.1. Com raio de decomposição 2 a rede conseguiu alcançar uma taxa de acerto de 75% nos dados de teste com apenas uma época de treinamento. Isto mostra a capacidade do FAN generalizar com pouca aprendizagem. Esta caracteriza uma das vantagens do FAN. Uma outra característica do modelo é que a partir de uma entrada ele já consegue prever a qual classe o próximo padrão estaria, ou seja, um modelo antecipatório.

Função de Pertinência	Raio	Range	Épocas	Treinamento (%)	Teste (%)
Gaussiana	1	19	5	85.00	75.00
Triangular	1	21	5	92.50	75.00
Triangular	2	20	1	55.00	75.00

Tabela 5.1 : Resultado do FAN (Porcentagem de acerto).

5.4 Comparação dos Resultados

Apresentou-se na Tabela 5.1 os resultados obtidos usando quatro sistemas distintos para classificação de empresas. Como pôde-se notar, as redes RBF e backpropagation apresentaram o mesmo desempenho (81.25% de acerto nos dados de teste). Mas, como já esperávamos, a RBF é muito mais rápida que a rede backpropagation (a RBF leva aproximadamente 1 segundo para fornecer este resultado enquanto a backpropagation leva aproximadamente 40s¹⁰).

O FAN, apesar de ser um sistema ainda em testes, apresentou uma boa generalização (75%). Mas o FAN conseguiu esta performance com apenas 1 iteração quando o raio da função de pertinência era 2 e levou 6 épocas quando o raio era 1. O FAN mantém este resultado se mais iterações forem feitas, ou seja ele não perde a generalização como acontece com a RBF e a Backpropagation . Porém, se houver uma superaprendizagem (por exemplo, 100% quando o range de normalização usado

¹⁰ Todos os resultados foram obtido num Pentim 166.

foi 100) ele perde a generalidade, assim como acontece com a RBF quando usamos 30 centros.

A rede LVQ não conseguiu uma performance tão boa quanto Backpropagation e a RBF, mas esta rede apresenta uma maior estabilidade na sua aprendizagem. Cada peso na LVQ representa uma das classes, então no momento do treinar a rede seria interessante termos um número de amostras semelhantes para cada classe, o que não acontece com nossos dados (Apêndice A e B). Talvez, por este motivo, a rede não apresentou um desempenho tão bom quanto as outras.

A tabela abaixo apresenta as redes em ordem quanto aos melhores desempenhos em relação a tempo de execução, número de épocas, aprendizagem e generalização. Podemos então concluir que a rede RBF foi a mais rápida em questão de tempo, o FAN foi mais rápido em relação ao número de épocas e aprendeu melhor os dados de treinamento, e a RBF e a backpropagation generalizaram melhor.

	1 ^o	2 ^o	3 ^o	4 ^o
Tempo		LVQ	Back	FAN
N ^o de épocas		RBF	LVQ	Back
Treinamento		Back 90%	RBF 77.5%	LVQ 72.5%
Teste		FAN 75%	LVQ 62.5%	

Tabela 5.1 : Comparação dos resultados.

5.5 Conclusão

Os resultados vistos acima comprovam a eficiência de técnicas de IA aplicadas à problemas de classificação de empresas. Mesmo trabalhando com tão poucos dados os sistemas apresentaram boa performance. Este tipo de trabalho sendo realizado em um sistema pode ser muito útil para as empresas que podem ter uma visão de como anda a saúde de sua empresa, principalmente para pequenas empresas que não podem ter acesso a um especialista para executar esta tarefa. Com isto, a

empresa pode detectar que tipos de problemas sua empresa apresenta, podendo desta forma tomar medidas que possam evitar falência ou melhorar a sua lucratividade.

Quanto ao desempenho das redes e do sistema híbrido, não é apropriado dizer qual é o melhor ou mais adequado para o problema de classificação de empresas. O que pode-se concluir é que para estes 56 dados com os quais trabalhou-se, a RBF destacou-se pelo bom desempenho e rapidez em termos de tempo de execução. A backpropagation teve um bom desempenho, porém é muito lenta. O FAN destacou-se pela sua habilidade de generalização com pouco treinamento, mas também foi lento. A LVQ, apesar de não apresentar um desempenho tão bom, mostrou-se bastante estável em relação a aprendizagem, ou seja, ela não perde a generalidade se muitas épocas forem executadas.

Concluindo, para estes dados a RBF foi a rede que apresentou uma melhor performance de um modo geral. Mas vale ressaltar aqui que os modelos trabalhados tiveram um bom desempenho, levando em consideração a pequena quantidade de dados disponíveis para o treinamento.

6. CONCLUSÕES E RECOMENDAÇÕES

6.1 Conclusões

O presente trabalho foi motivado pela necessidade de um maior conhecimento de como avaliar em termos empíricos a saúde financeira das pequenas empresa (catarinenses). O monitoramento da saúde financeira das empresas é um fator crítico para seu sucesso. Este representa mais um passo para dar suporte estratégico aos micro-empresários a verificar a situação financeira da empresa, sem precisar dispendir muitos recursos. Para tal fim, torna-se imprescindível o desenvolvimento de sistemas inteligentes

Para análise de empresas existem muitos sistemas inteligentes, porém estes eram voltados mais para análise de crédito e previsão de falência que são problemas diferentes de determinar a saúde financeira. No caso da análise de crédito, estuda-se as condições financeiras de uma empresa com objetivo exclusivo de verificar se um empréstimo é seguro; já no caso de previsão de falência prevê a “vida” ou falência da empresa. A determinação da saúde financeira é um problema mais complexo pois deve analisar vários aspectos da empresa (giro, endividamento, liquidez, lucratividade).

Uma técnica central para endereçar situações financeiras problemáticas é a Análise de Declarações Financeiras, um processo no qual os especialistas reorganizam as informações da firma e outras origens, criam variáveis auxiliares (como os índices financeiros), e fazem uma comparação com padrões a fim de identificar e entender os desvios.

Para elaboração deste estudo, iniciou-se analisando as declarações financeiras (Balanço Patrimonial e Demonstrativos de resultados) de 56 pequenas empresas

catarinenses, extraíndo deles cinco (5) índices financeiros (Caixa Operacional, Lucro Líquido, CG_V, TIE e Endividamento) com os quais classificou-se as empresas em sete (7) categorias quanto aos problemas de rentabilidade, giro e endividamento.

O raciocínio indutivo foi modelado por quatro (4) algoritmos alternativos que usam técnicas de Inteligência Artificial: três (3) redes neurais distintas (Backpropagation, RBF e LVQ) e um sistema híbrido (FAN). O objetivo era comparar os desempenhos buscando os que mais se adaptam ao problema de diagnóstico de empresas

Dos 56 padrões de entradas, usou-se quarenta (40) para treinar os sistemas e o restante (16) usou-se para a validação. Apesar de dispor-se de poucos dados os resultados encontrados de um modo geral foram muito bons, mostrando a capacidade de aprendizagem das redes com poucos dados. Como pode-se observar na Tabela 6.1, em relação a validação, as redes Backpropagation e RBF apresentaram o melhor resultado, porém, como já esperávamos, a RBF é muito mais rápida que a Backpropagation. O FAN também conseguiu um resultado bom, porém ele foi relativamente mais lento que a Backpropagation. Mas devemos lembrar que o FAN foi programado em Cliper e o Backpropagation em Matlab. A LVQ foi a que apresentou o desempenho mais baixo, mas vale lembrar que usamos a versão mais simples das LVQs. Existem outros modelos de LVQ, como LVQ1 e LVQ2, que são um pouco mais elaboradas. Além disto, estamos trabalhando com poucos dados. Acredita-se que com mais dados os resultados podem melhorar, principalmente, se houver uma número semelhante de padrões para cada cluster. Isto não só para a LVQ, mas para as outras redes também.

Mas sem levantar mais hipóteses, analisando resultados que foram obtidos com estes modelos, sejam eles os mais elaborados ou não, nesta ou naquela linguagem, pode-se concluir que a rede RBF foi a melhor em termos de generalização e rapidez e o FAN foi o melhor em termos de aprendizagem.

Para problemas de classificação com características semelhantes ao problema que foi trabalhado nesta dissertação, recomenda-se o uso da rede RBF pois apresenta as seguintes características:

- Rapidez para treinamento

- Facilidade para determinar os parâmetros que fornecem o melhor desempenho.
- Boa capacidade de generalização.
- Capacidade de classificar dados complexos (não linearmente separáveis).

Modelo	Treinamento (%)	Teste (%)
Backpropagation	90.00	81.25
LVO	72.50	62.50
FAN	92.50	75.00
RBF	77.50	81.25

Tabela 6.1 : Percentual de acerto de cada modelo.

6.2 Recomendações

Um das principais limitações dos modelos utilizados é que eles não fornecem uma explicação dos resultados obtidos, eles simplesmente fornecem um indicador numérico. E os usuários deste tipo de sistemas necessitam de uma explicação sobre as causas do diagnóstico apresentado. Tendo em vista esta limitação propõe-se os seguintes trabalhos futuros:

- 1) Construção de um sistema híbrido, que além do diagnóstico, pudesse detectar as causas e sugerir soluções aos problemas diagnosticados. Já existe trabalhos nesta linha como o sistema híbrido inteligente desenvolvido por Pacheco [Pacheco, 1996], porém não aplicado a empresas brasileiras. Este sistema usa Redes Neurais Artificiais e um Sistema Especialista Difuso. A sugestão é trabalhar com um sistema híbrido que trabalhe com RNA, Conjuntos Difusos e Raciocínio Baseado em Casos. Neste sistema, uma RNA seria usada para fazer o diagnóstico, este diagnóstico seria *fuzificado* e o resultado seria um dos atributos de cada caso. O RBC, forneceria as causas e possíveis soluções para o problema diagnóstico pela RNA. A Figura 6.1 abaixo mostra o esquema básico do modelo sugerido.

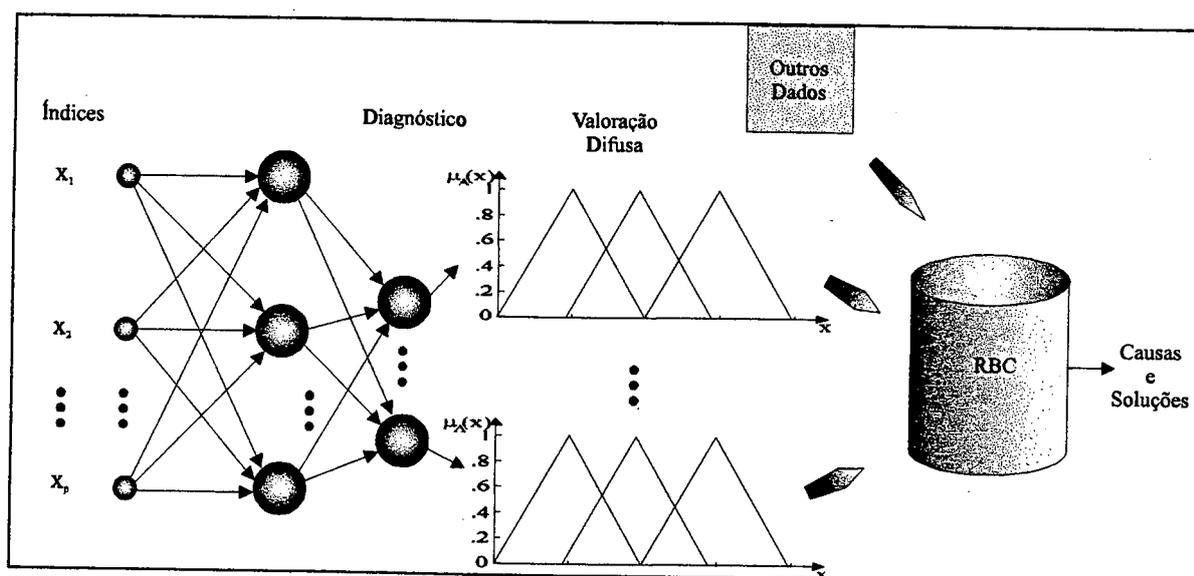


Figura 6.1 : Arquitetura híbrida para análise da situação financeira.

- 2) Uma outra forma de melhorar os modelos é a aplicação de algoritmos de extração de conhecimento (regras) da rede treinada. Já existem alguns trabalhos neste linha como [Andrews, et al. 1995; Fu, 1994].
- 3) Uma outra rede que poderia ser usada para o diagnóstico de empresas é a rede de Kohonen (SOFM - *Self-Organizing Features Maps*). Os mapas de Kohonen não resolveriam o problema de explicar os resultados obtidos, mas fornecem informações (visuais) mais completas sobre a situação financeira das empresas [Martin-del-Brio & Serrano-Cinca, 1995].

Além disto, propõe-se a implementação destes algoritmos para uso através da rede Internet.

REFERÊNCIAS BIBLIOGRÁFICAS

- [Aamodt, 1991] AAMODT, A. *Case-based Reasoning: Foundational Issues: Methodological Variations and Systems Approaches*. AICOM, v7, n.1, 1991.
- [Aho, 1980] AHO, T. Empirical classification of financial ratios. *Management Science in Finland Proceedings*, ed. C. Carlsson, 1980.
- [Altman, 1968] ALTMAN, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, September, pp. 589-609, 1968.
- [Anderson, 1972] ANDERSON, J. A. "A Simple Neural Network Generating an Interactive Memory", *Mathematical Biosciences* 14, pp. 197-220, 1972.
- [Andrews, et al. 1995] ANDREWS, R.; DIEDERICH, J. e TICKLE, A. B. A Survey and Critique of Technics for Rule Extracting Rules from Trained Artificial Neural Networks, *Neural Computer Research Centre*, Queensland University of Technoloy, Autralia, Technical Report, 1995.
- [Angeniol et al., 1988] ANGENIOL, B., G. DE LA CRAIX V. e LE TEXIER J.-Y. Self-Organizing Feature Maps and the Traveling Salesman Problem, *Neural Networks*, Vol. 1, pp. 289-293, 1988.
- [Austin, 1990] AUSTIN, S. An Introduction to Genetic Algorithms. *AI Expert*, 1990.
- [Bareiss, 1988] BAREISS, R. *PROTOS A unified approach to concept representation, classification and learning*. Ph.D. Dissertation. University of Texas at Austin, Dep. of Computer Sciences Technical Report AI, pp. 88-83, 1988.
- [Barker, 1990] BARKER, D., Analysing Financial Health: Integrating Neural Networks and Expert Systems, *PC AI*, May/June, 1990.
- [Barnes, 1982] BARNES, P., Methodological implications of non-normally distributed financial ratios, *Journal of Business Finance and Accounting*, pp.51-62, 1982.
- [Barnes, 1987] BARNES, P. The Analysis and Use of Financial Ratios, *Journal of Business Finance and Accounting*, Vol. 14, N^o 4, pp. 449-461, 1987.

- [Barr & Mani, 1994] BARR, D. S. e MANI, G. Using Neural Nets to Manage Investments. *AI Expert*, Vol.9, Nº 2, pp. 16-21, 1994.
- [Batalha & Demori, 1990] BATALHA, M., O., & DEMORI, F. "A Pequena e Média Indústria em Santa Catarina", Editora da Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, 1990.
- [Baykal & Yalabki, 1992] BAYKAL, N. & YALABKI, N. Object Orientation Detection and Character Recognition using Optimal Feed-forward Network and Kohone's Feature Map, *SPIE*, Vol. 1709, pp. 292-303, 1992.
- [Beaver, 1966] BEAVER, W. H. Financial Ratios as Predictors of Failure, *Journal of Accounting Research*, pp. 77 - 111, 1966.
- [Belkaoui, 1978] BELKAOUI, A. Financial Ratios as Predictors of Canadian takeovers, *Journal of Business Finance and Accounting*, pp. 93-67,1978.
- [Benachenhou et al., 1990] BENACHENHOU, D.; CADER, M.; SZU, H.; MEDSKER, L.; WITTEWER, C. & GARLING, D. Neural Networks for Computing Invariant Clustering of a Large Open Set of DNA-PCR Primers Generated by a feature-knowledge based system. *Proceedings of IJCNN-90*, San Diego, CA, Vol. II, pp. 83-89, 1990.
- [Bennett et al., 1989] BENNETT, K.; FERRIS, M. C. e IOANNIDIS, Y. E., A Genetic Algorithm for Database Query Optimization, *In Proceedings of the Fourth International Conference on Genetic Algorithms*, R. K. Belew e L. B. Booker (eds.), San Mateo, 1991.
- [Bernstein, 1989] BERNSTEIN, L., *Financial Statement Analysis, Theory, Application and Interpretation*. Richard D. Irwin, Inc., Homewood, 1989.
- [Berry & Nix, 1991] BERRY, R.H., and NIX, S. Regression analysis v. ratios in the cross-section analysis of financial statements, *Accounting and Business Research* 21/82, pp 107-117, 1991.
- [Bilbro et al., 1988] BILBRO, G. et al. Optimization by Mean Field Annealing. *In Advance In Neural Information Processing systems*, I. D. S. Touretzky (Ed.), Morgan Kalfmann, San Mateo, CA, pp. 91-98, 1988.

- [Bird & McHugh, 1977] BIRD, R.G., and MCHUGH, A.J. Financial ratios - an empirical study, *Journal of Business Finance and Accounting* 4/1, 29-45, 1977.
- [Blocher, 1990] BLOCHER, E. ANSWERS: An Expert System for Financial Analysis. *Em [Liebowitz, 1990]*, (1990) pp. 101-125.
- [Booth et al., 1994] BOOTH, G., MARTIKAINEN, T., PERTTUNEN, J., and YLI-OLLI, P. , On the functional form of earnings and stock prices: international evidence and implications for the E/P anomaly, *Journal of Business Finance and Accounting*, 21/3, pp.395-408, 1994.
- [Bramlette & Bouchard, 1991] BRAMLETTE, M. F. & BOUCHARD, E. E. Genetic Algorithm in Parametric Design of Aircraft, *In Handbook of Genetic Algorithm*, Edited by L. Davis. New York: Van Nostrand Reinhold, 1991.
- [Branting, 1991] BRANTING, K. Exploiting the complementary of rules and precedents with reciprocity and fairness. In: *Proceedings from the Case-Based Reasoning Workshop*, Washington DC, May 1991. Sponsored by DARPA. Morgan Kaufmann, pp. 39-50, 1991.
- [Brigham, 1979] BRIGHAM, E. F., *Financial Management: Theory and Practice*. Dryden Press, p. 769, 1979.
- [Broomhead & Lowe, 1988] BROOMHEAD, D. S. & LOWE, D. Multivariable Functional Interpolation and Adaptive Networks, *Complex Systems* 2, pp. 321-355, 1988.
- [Brown et al., 1990] BROWN, C. & PHILLIPS, M. E. *Expert Systems for Management Accounts*. *Management Accounting*, pp.18-23, January, 1990.
- [Caldwell & Johnston, 1991] CALDWELL, C. & JOHNSTON, V. S., Tracing a Criminal Suspect Through 'Face-Space' with a Genetic Algorithm, *In Proceedings of the Fourth International Conference on Genetic Algorithms*, R. K. Belew e L. B. Booker (eds.), San Mateo, 1991.
- [Carpenter & Grossberg, 1987] CARPENTER, G. A. & GROSSBERG, S. A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine, *Computer, Vision Graphics and Image Processing*, 37, pp. 54-115, 1987.

- [Chen & Shimerda, 1981] CHEN, K. H. & SHIMERDA, T. A. An Empirical Analysis of Useful Financial Ratios. *Financial Management*, pp. 51-60, 1981.
- [Chen et al., 1992] CHEN, S; MULGREW, B.; McLAUGHLIN, S. and GRANT, P. M. Adaptive Bayesian Equalizer with Feedback for Mobile Radio Channels. *Workshop on Adaptive Algorithms in Communications*, Bordeaux, France, 1992.
- [Chu, 1992] CHU, Pai-Cheng. *Applying Object-Oriented Concepts To Developing Financial Systems*, Journal of Systems Management, Vol. 43, No. 5, May 1992.
- [Colombi et al., 1993] COLOMBI, J. M. et al. Auditory Model Representation for Speaker Recognition. *Proceedings of the SPIE Applications of Artificial Neural Networks IV, Orland, FL*, pp. 9-14, 1993.
- [Curtis, 1978] COURTIS, J.K. Modeling a financial ratios categorical framework. *Journal of Business Finance and Accounting* 5/4, 371-386, 1978.
- [Cox, 1994] COX, E. "The Handbook of Fuzzy Systems". Academic Press, New York, 1994.
- [Darwin, 1985] DARWIN, C. R. "On the Origins of Species by Means of Natural Selection". Penguin Classics, 1985.
- [Davis, 1991] DAVIS, L. "Handbook of Genetic Algorithms". Van Nostrand Reinhold, New York, 1991.
- [de la Maza, 1989] DE LA MAZA, M, A SEAGUL Visits the Race Tracking. In *Proceedings of the Third International Conference on Genetic Algorithms*, Edited by J. D. Schaffer, San Mateo, 1989.
- [Deakin, 1976] DEAKIN, E.B Distributions of financial accounting ratios: some empirical evidence. *Accounting Review*, Janeiro, pp. 90-96, 1976.
- [Deboeck, 1994] DEBOECK, G., (ed.), *Trading on the Edge: Neural, Genetic, Fuzzy Systems for Chaotic Financial Marketes*, John Wiley, Chichester, 1994.
- [Demuth & Beale, 1994] DEMUTH, H. e BEALE, M. *Neural Network Toolbox User's Guide*, the MathWorks, 1994.

- [Dracopoulos & Jones, 1995] DRACOPOULOS, D. C. e JONES, A. J. Adaptive Neuro-Genetic Control of Chaos applied to the Attitude Control Problem. *IEEE Transactions on systems man and Cybernetics*, Vol. XX, Nº Y, pp. 1-12 1995.
- [Dutta & Shekhar, 1988] DUTTA, S. e SHEKHAR, S. Bond Rating: A non-conservative Application of Neural Networks, *Proceeding of IEEE International Conference on Neural Networks*, San Diego, 1988.
- [Edmister, 1972] EDMISTER, R. O. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis*, Março, pp. 1477-1493, 1982.
- [Eisenbeis, 1977] EISENBLES, R. A. Pitfalls in the Applications of discriminant Analysis in Business Finance, and economics. *Journal of Finance*, Junho, pp. 875-899, 1977.
- [Enrado, 1991] ENRADO, P. Application Watch, *AI Expert*, September, p. 64, 1991.
- [Ezzamel & Mar-Molinero, 1990] EZZAMEL, M. and MAR-MOLINERO, C. The Distributional Properties of Financial Ratios in UK Manufacturing Companies. *Journal of Business Finance and Accounting*, v17, n1, pp. 1-29, 1990.
- [Fauset, 1994] FAUSET, L. *Fundamentals of Neural Networks - Architectures, Algorithms, and applications*. Prentice Hall International, Inc., 1994.
- [Feigenbaum, 1977] FEIGENBAUM, E. The art of artificial intelligent : Themes and case studies of knowledge engineering. *Proc, Fifth Int. Joint Conf. On Artificial Intelligent*, Morgan Kaufmann, CA, pp. 1014-1029, 1977.
- [Fieldsend et al., 1987] FIELDSEND, S., LONGFORD, N., and MCLEAY, S. Industry effects and the proportionality assumption in ratio analysis: a variance component analysis. *Journal of Business Finance and Accounting* 14/4, 497-517, 1987.
- [Folgelman-Souliè, 1995] FOLGELMAN-SOULIÈ, F. Applications of Neural Networks. *In the Handbook of Brain Theory and Neuro Network*, Cambridge, pp. 94-98, 1995.
- [Foster, 1978] FOSTER, G. *Financial Statement Analysis*. Prentice-Hall, Primeira ed., 1978.

- [Foster, 1987] _____ *Financial Statement Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1987.
- [Foster, 1986] _____ *Financial Statement Analysis*. Prentice-Hall, 2nd ed., 1986.
- [Fu, 1994] FU, L. M. Rule Generation from Neural Networks. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 28, No. 8, pp. 1114 - 1124, 1994.
- [Fukushima et al., 1983] FUKUSHIMA, K.; MIYAKE, S.; ITO, T., Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition. *IEEE Transaction on Systems, Man, and Cybernetics SMC-13*, pp. 826-834, 1983.
- [Fukushima, 1988] FUKUSHIMA, K, Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition. *Neural Networks 1*, pp. 119-130, 1988.
- [Gabbert et al., 1989] GABBERT, P. S. et al. A system for Learning Routes and Schedules whit Genetic Algorithm. *In Proceedings of the Fourth International Conference on Genetic Algorithms*, R. K. Belew e L. B. Booker (eds.), San Mateo, 1991.
- [Gallant, 1988] GALLANT, S., Connectionist Expert Systems. *Communications of the ACM*, Vol. 31, No 2 February, pp. 114-136, 1988.
- [Goldberg, 1989] GOLDBERG, D. E. *Genetic Algorithmis in Search, Optimization, and Machine Learning*. Addison-Wesley, p. 412, 1989.
- [Gombola & Ketz, 1983] GOMBOLA, M.J., and KETZ, J.E. *A note on cash flow and classification patterns of financial ratios"*, *Accounting Review* 63/1, 105-114, 1983.
- [Goonatilake & Khebbal, 1995a] GOONATILAKE, S. e KHEBBAL, S. *Intelligent Hybrid Systems*. Jonh Wiley &sons Ltd., 1995.
- [Goonatilake & Khebbal, 1995b] _____ *Intelligent Hybrid Systems: Issues, Classifications, and Future Directions*. in [Goonatilake & Khebbal, 1995a].
- [Goonatilake & Treleaven, 1995] GOONATILAKE, S. e TRELEAVEN, P. *Intelligent System for Finance and Busines"s*. Jonh Wiley &sons Ltd., 1995.

- [Hamilton & Hufnagel, 1992] HAMILTON, A. e HUFNAGEL, Early Detection of Epileptic Attacks . *In Applications in Neural Networks*, H. G. Schuster (ed.), Weinheim, pp.173-8, 1992.
- [Hammond, 1989] HAMMOND, C. *Case-based planning*. Academic Press, 1989.
- [Haykin, 1994] HAYKIN, S. *Neural Networks: A comprehensive Foundation*. Macmillan College Publish Company, p. 895, New York, 1994.
- [He & Lapedes, 1991] HE, X. e LAPEDES, A. *Nonlinear Modeling and Prediction by Successive Approximation Using Radial Basis Function*. Technical Report LA-UR-91-1375. Los Alamos National Laboratori, Los Alamos, NM, 1991.
- [Hebb, 1949] HEBB, D. O. *The Organization of Behavior*, John Wiley & Sons, New York, 1949.
- [Hinrichs, 1992] HINRICHS, T. *Problem solving in open worlds*. Lawrence Erlbaum Associates, 1992.
- [Holland, 1975] HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. Univ. of Michigan Press, 1975.
- [Hopfield, 1982] HOPFIELD, J. J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl. Acad. Science. USA* 79, pp. 2555 - 2558, 1982.
- [Hopfield, 1984] HOPFIELD, J. J. Neurons with Graded Responde Have Collective Computational Properties like those of two-state Neurons. *Proceedings of the National Academy of Science*, Vol. 81, pp. 3080-3092, 1984.
- [Horner & Goldberg, 1989] HORNER, A. & GOLDBERG, D. E. Genetic Algorithms and Computer-Assisted Music Composition. *In Proceedings of the Fourth International Conference on Genetic Algorithms*, R. K. Belew e L. B. Booker (eds.), San Mateo, 1991.
- [Hornik, et al. 1989] HORNİK, K, et al.. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 , pp.359-366, 1989.
- [Horrigan, 1965] HORRIGAN, J.O. Some empirical bases of financial ratio analysis. *Accounting Review*, Julho, pp. 558-568, 1965.

- [Horrigan, 1966] _____ The Determination of Long-Term Credit Standing With Financial ratio. *Empirical Research in Accounting: Selected Studies, Supplement to Journal of Accounting Research* 4, pp. 44-62, 1966.
- [Horrigan, 1968] _____ A short History of Financial ratio analysis. *The Accounting Review*, Abril, pp. 284-294, 1968.
- [Horrigan, 1983] _____ Methodological implications of non-normally distributed financial ratios: a comment. *Journal of Business Finance and Accounting* 10/4, pp. 683-689, 1983.
- [Ingram & Copeland, 1984] INGRAM R. W. & COPELAND R.M. The Association Between Municipal Accounting Numbers and Credit Risk and Returns. *In Advances in Accounting*, Vol. 1, pp. 19-40, 1984.
- [Jacobs, 1988] JACOBS , R. Increased rates of convergence through learning rate adaptation. *Neural Networks* 1, pp. 295-307, 1988.
- [Jang & Sun, 1995] JANG, J. S. R. & SUN, C. T. Neuro-Fuzzy Modeling and Control. *Proceedings of the IEEE*, Vol. 83, No. 3, pp. 378-406, 1995.
- [Jonhson, 1978] JOHNSON, W.B. The cross-sectional stability of financial patterns. *Journal of Business Finance and Accounting* 5/2, pp. 207-214, 1978.
- [Kalkunte et al, 1992] Kalkunte, S. S., KIMAR, M, e PATNAIK, L. M. A neural network approach for high resolution fault Diagnosis in Digital Circuits. *Proceedings of the IJCNN-92*, Beijing, Vol. I, pp. I-83, I-88, 1992.
- [Kanade et al., 1994] KANADE, T.; REED, M. L. e WEISS, L. E. *New Tecnologies and applications in Robotics, Communications of the ACM*, Vol. 37, Nº 3, pp. 58-67, 1994.
- [Kandel & Langholz, 1992] KANDEL, A. e LANGHOLZ, G. *Hybrid Architectures for Intelligent Systems*. Ed. CRC Press. Inc., 1992.
- [Kandel, 1993] KANDEL, A. Interview. *PC AI*, pp. 40-41, Mar./Abr, 1993.
- [Karels & Prakash, 1987] KARELS, G. V., e PRAKASH, A. Multivariate Normality and Forecasting of Business Bankruptcy. *Journal of Business Finance and Accounting*, Winter, 1987.

- [Karr, 1991] KARR, C. Applying Genetics to Fuzzy Logic. *AI Expert*, 1991.
- [Karr, 1995] Karr, C. Genetic Algorithms and Fuzzy Logic for adaptive control. in *[Goonatilake & Khebbal, 1995a]*.
- [Kasabov, 1996] Kasabov, K. *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*. MIT Press, 1996.
- [Kassam & Cha, 1993] KASSAM, S. A. & Cha, I. Radial Basis Functions Networks in Nonlinear Signal Processing Applications. *27th Annual Asilomar Conference on Signals, Systems, and Computers Pacific Grove, CA*, 1993.
- [Katon, 1989] KATON, P. *Using experience in learning and problem solving*. Massachusetts Institute of Technology, Laboratory of Computer Science (Ph. D. Diss, October, 1988). MIT/LCS/TR-441. 1989.
- [Khebbal & Shamhong, 1995] KHEBBAL, S. e SHAMHONG, D. *Tools and Environments for Hybrid Systems*. in *[Goonatilake & Khebbal, 1995a]*.
- [Kimoto et al., 1990] KIMOTO, T. et. Al. Stock Market Prediction System with Modular Neural Networks. In *proceeding of the IEEE International Joint Conference on Neural Network*, IEEE, pp. 11 - 16, San Diego, CA, 1990.
- [Kingdon & Feldman, 1995] KINGDON, J. & FELDMAN, K. Genetic Algorithms for Bankruptcy Prediction. *SearchSpace Research Report 01*, SearchSpace Ltda, London, 1995.
- [Kitano, 1993] KITANO, K. Challenges for massive parallelism. *IJCAI-93, Proceedings of the Thirteenth International Conference on Artificial Intelligence*, Chambéry, France. Morgan Kaufmann .pp. 813-834, 1993.
- [Klir & Yuan, 1995] KLIR, G. J. & YUAN, B. *Fuzzy sets and fuzzy logic - theory and applications*. Prentice Hall, p.574, 1995.
- [Klymasauskas, 1991] KLYMASAUSKAS, C. C. Using ANNs for Credit Approval. *PC AI*, Jan/Fev, 1991, pp. 30 - 33 and Mar/Apr, pp. 27 - 34, 1991.
- [Kohonen, 1972] KOHONEN, T. Correlation Matrix Memories. *IEEE Transaction on Computers C-21*, pp. 353-359, 1972.

- [Kohonen, 1982] _____ Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, v. 43, p. 59-69, 1982.
- [Kohonen, 1988] _____ The "Neural" Phonetic Typewriter. *IEEE Computer*, Vol. 21, Nº 3, pp.11-22, 1988.
- [Kohonen, 1990] _____ The self-organizing map. *Proceedings of the IEEE*, 78:1464-1497, 1990.
- [Kolodner, 1983a] KOLODNER, J. Maintaining organization in a dynamic long-term memory. *Cognitive Science*, Vol.7, s.243-280, 1983.
- [Kolodner, 1983b] _____ Reconstructive memory, a computer model. *Cognitive Science*, Vol.7. s.281-328, 1983.
- [Kolodner, 1992] _____ *An Introduction to case-based Reasoning*. Artificial Intelligence Review 6 (1), pp.3.34, 1992.
- [Korner, 1967] KORNER, S. Laws of thought. *Encyclopedia of Philosophy*, Vol. 4, MacMillan, NY., pp. 414-417, 1967.
- [Kosko, 1992] KOSKO, B. *Neural Networks and Fuzzy Systems*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- [Kung, 1993] KUNG, S. Y. *Digital Neural Networks*. Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [Lacher et al, 1993] LACHER, R. C.; COATS, P. K.; SHARMA, S. C. e FANT, L. F. *A neural Network for Classifying the Financial Health of a Firm*. Technical Report of the Florida State University, Dept. of Computer Science, 1993.
- [Laitinen, 1983] LAITINEN, E.K. A multivariate model of the financial relationship in the firm. *Finnish Journal of Business Economics* 32/4, 317-333, 1983.
- [Lapedes & Farber, 1988] LAPEDES, A. e FARBER, R. How Neural Nets Work. *In Neural Information Processing Systems*, D. Z. Anderson (ed.), American Institute Of Physics, New York, 1988.
- [Laurent, 1979] LAURENT, C.R. Improving the Efficiency and Effectiveness of Financial Ratio Analysis. *Journal of Business Finance & Accounting* Vol. 6, No. 3, pp.401-413, 1979.

- [Lecun, 1985] LECUN, Y. Une Procedure d'apprendissage pour Receau a seuil Assymetrique, *Cognitiva* 85, pp. 599-604, 1985.
- [Lee & Wu, 1988] LEE, C. F.; WU, C. Expeptation Formation and Financial Ratio Adjustment Process. *The Accounting Review*, v53, n2, 1988.
- [Lejewsk, 1967] LEJEWSK, C. I. Jan Lukasiwicz. *Encyclopedia of Philosophy*, Vol. 5, MacMillan, NY:, pp. 104-107, 1967.
- [Lev & Sander, 1979] LEV, B., and SUNDER, S. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics* 1/3, pp.187-210, 1979.
- [Lev, 1974] LEV, B. *Financial Statement Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [Levy et al., 1991] LEVY J., et al. A Fuzzy Logic Evaluation System for Commercial Loan Analysis. *Omega*, Vol. 19, Nº6, 1991.
- [Liebowitz, 1990] LIEBOWITZ, J. (editor), *Expert System for Business and Managemen*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
- [Lopez & Plaza, 1993] LOPEZ, B, PLAZA, E. Case-based planning for medical diagnosis, In: J. Komorowski, Z.W. Ras (Eds.) *Methodologies for Intelligent Systems: 7th International Symposium*, p. 96-105. Lecture Notes in Artificial Intelligence 689. Springer Verlag, 1993.
- [Lowe & Webb, 1990] LOWE, D. e WEBB, A. R. Exploiting prior knowledge in network optimization: An Illustration from medical prognosis. *Network* 1, pp. 299-323, 1990.
- [Luoma & Ruuhela, 1991] LUOMA, M., and RUUHELA, R. Consistency and Comovement of Financial Ratios: a Firm-Specific Approach. *Finnish Journal of Business Economics* 1, pp. 39-49, 1991.
- [Maiers e Sherif, 1985] MAIERS, J. e SHERIF, Y. S. Applications of Fuzzy Set Theory. *IEEE Trans. Systems, Man and Cybernetics*, vol. SMC-15, Nº 1, pp. 175-189, Fev. 1985.
- [Martin-del-Brio & Serrano-Cinca, 1995] MARTIN-DEL-BRIO, B. & SERRANO-CINCA, C. Self Organizing Neural Networks: The Finantial State of Spanish Copanies. Universidad the Zaragoza, Spain, in [Refenes, 1995].

- [Martins, 1996] MARTINS, A. R., “*Modelling Financial Statements Analysis for Indicating Correstive Actions to Financial Problems*”, Dissertação de Doutorado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1996.
- [Martins et al., 1996] MARTINS, A., PACHECO, R., WEBER-LEE, R., BARCIA, R. M., Integrating Expert Networks and CBR in a Hybrid Architecture to Analyze the Financial Health of a Firm. *Brazil 2nd International Congress of Industrial Engineering and 16th National Congress of Production Engineering - ENEGEP*, October, 7 - 10, Piracicaba, São Paulo, Brasil, 1996.
- [Matarazzo, 1987] MATARAZZO, D. C. *Análise Financeira de Balanços - Abordagem Básica*, 2ª Edição, São Paulo, Ed. Atlas, 1987.
- [McCulloch & Pitts, 1943] MCCULLOCH, W. S. & PITTS, W. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics* 5, pp. 115-133, 1943.
- [McDonald & Morris, 1984] MCDONALD, B., and MORRIS, M.H. The statistical validity of the ratio method in financial analysis: an empirical examination. *Journal of Business Finance and Accounting* 11/1, pp. 89-97, 1984.
- [McDonald & Morris, 1985] _____ The functional specification of financial ratios: an empirical examination. *Accounting and Business Research* 15/59, (1985), pp. 223-228.
- [McLeay & Fieldsend, 1987] MCLEAY, S., and FIELDSEND, S. Sector and size effects in ratio analysis: an indirect tests of a ratio proportionality. *Accounting and Business Research* 17/66, pp. 133-140, 1987.
- [Mecimore, 1968] MECIMORE, C.D. Some empirical distributions of financial ratios. *Management Accounting* 50/1, pp. 13-16, 1968.
- [Medsker & Bailey, 1992] MEDSKER, L. e BAILEY, D. Models and Guidelines for Integrating Expert Systems and Neural Networks. *In Hybrid Architectures for Intelligent Systems*, Ed. by Abraham Kandel and Gideon Langholz, CRC Press, Boca Raton, Florida, USA, 1992.

- [Micchelli, 1986] MICCHELLI, C. A. Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Function. *Constructive Approximation* 2, pp. 11-22, 1986.
- [Minsky & Papert, 1969] MINSKY, M. L. & PAPERT, S. A. *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- [Moallemi, 1991] MOALLEMI, C. Classifying Cells for Cancer Diagnosis Using Neural Networks. *IEEE Expert, December*, pp. 8-12, 1991.
- [Montana, 1995] MONTANA, D. Neural Network weight selection using genetic algorithms, in [Goonatilake & Kheball, 1995a].
- [Moody & Darken, 1989] MOODY, J. E. & DARKEN, C. J. Fast Learning in Networks of Locally-tuned Processing Unit. *Neural Computation* 1, pp. 281-294, 1989.
- [Morose, 1990] MOROSE, R. A. A Financial Neural-Network Application. *AI Expert*, Maio, pp. 50-53, 1990.
- [Morose, 1993] MOROSE, R. A. A Financial Neural-Network Application. *Neural Networks in Finance and Investing*, R. R. Trippi e E. Turban (eds.), Probus Publishing, Chicago, pp. 75-83, 1993.
- [Mui et al., 1990] MUI, C. K.; HAFELL, C. e COURTIS, L. The Financial Statements Analyser. In [Liebowitz, 1990] pp.127-142, 1990.
- [Neto, 1983] NETO, A. A. *Estrutura e Análise de Balanço - Um enfoque Econômico-financeiro*. 2^a Edição, São Paulo, Ed. Atlas, 1983.
- [Neto, 1995] NETO, J. S. Uma estrutura Fuzzy-Neural-Genética-Caótica para predição da tendência global de curto prazo do mercado de ações. *II Congresso Brasileiro de Redes Neurais - III Escola de Redes Neurais*, Curitiba, Novembro, 1995.
- [Newquist, 1987] NEWQUIST III, Harvey P. In Practice: American Express and AI: Don't Leave Home Without Them. *AI Expert*, pp. 63-65, April 1987.
- [Ng e Lippmann, 1991] NG, K. e LIPPMAN, R. P. Practical Characteristics of Neural Network and Conventional Pattern Classifiers. In *Advance In Neural Information*

- Processing systems 3*, R. P. Lippmann, J. E. Moody and D. S. Touretzky (Eds.), Morgan Kaufmann, San Mateo, CA, pp. 970-976, 1991.
- [Niranjan e Fallside, 1990] NIRANJAN, M. e FALLSIDE, J. Neural Networks and Radial Basis Functions in Classifying Static Speech Patterns. *Computer Speech and Language* 4, pp. 275-289, 1990.
- [Odom & Shardo, 1990] ODOM, M. & SHARDO, R. A Neural Network Model for Bankruptcy Prediction. *In Proceedings of the International Joint Conference on Neural Networks*, 1990.
- [Odom & Shardo, 1993] _____ A Neural Network Model for Bankruptcy Prediction. *Decision Support Systems*, Vol. 9, N. 6, 1993.
- [Ohlson, 1980] OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, Spring, pp 109-131, 1980.
- [Pacheco, 1996] PACHECO, R. C. S. *A Hybrid Intelligent System for Diagnosing and Solving Financial Problems of Small Retail Firms*. Ph.D. Dissertation, Universidade Federal de Santa Catarina UFSC, Programa de Pós-graduação em Engenharia de Produção, Brasil, 1996.
- [Packard, 1990] PACKARD, N. H. A Genetic Learning Algorithm for the Analysis of Complex Data. *Complex Systems* 4, pp. 543-572, 1990.
- [Pandya & Macy, 1995] PANDYA, A. & MACY, R.B. *Pattern Recognition with Neural Networks in C++*. CRC Press, 1995.
- [Parker, 1985] PARKER, D. B. *Learning-Logic: Casting the Cortex of the Human Brain in Silicon*. Technical Report TR-47. Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA, 1985.
- [Patterson et al., 1993] PATTERSON, D. W.; CHAN, K. H. e TAN, C. M. Time Series Forecasting with Neural Networks: A Comparative Study. *Proceedings of the International Conference on Neural Network Applications to signal Processings (NNASP-93)*, Singapore, pp. 269-74, 1993.
- [Patterson, 1995] PATTERSON, D. W. *Artificial Neural Networks - Theory and Applications*. Prentice Hall, 1995.

- [Pau, 1991] PAU, L. F. *Artificial Intelligence and Financial Services. IEEE Transactions on Knowledge and Data Engineering*, Vol. 3, No. 2 , pp. 137-148, June 1991.
- [Pinches & Mingo, 1973] PINCHES, G. E., & MINGO, K.A. A multivariate analysis of industrial bond ratings. *Journal of Finance*, 28/1, 1-18, 1973.
- [Poggio & Girosi, 1990] POGGIO, T. , GIROSI, F. *Networks for approximation and learning. Proceedings of the IEEE*, v. 78, p. 1481-1497, 1990.
- [Poggio e Edelman, 1990] POGGIO, T. e EDELMAN, S. A Network that Learns to Recognize three-dimensional Objects. *Nature (London)* 343, pp. 263-266, 1990.
- [Porter et al., 1986] PORTER, B.; BAREISS, R. e PROTOS, D. An experiment in knowledge acquisition for heuristic classification cases. III. *Proceedings of the First International Meeting on Advances in Learning (IMAL)*, Les Arcs, France, pp. 159-174, 1986.
- [Powell, 1985] POWELL, M. J. D. Radial Bases Function for Multivariable Interpolation: A Review. *In IMA Conference on Algorithms for the Approximation of Functions and Data.*, RMCS, Shrivenham, pp.143-167, 1985.
- [Radding, 1991] RADDING, A. C. Loan Experts in the Struggle to Limit Risk. *Bank Management*, pp. 48-50, July 1991. (also reprinted in [Zahed, 1993] pp. 102-104)
- [Radecki, 1982] RADECKI, T. An Evaluation of the Fuzzy Set Theory Approach to Information Retrieval. in R. Trappl, N. V. Findler, and W. Horn, *Progress in Cybernetics and System Research*, Vol. 11: Proceedings of a Symposium Organized by the Austrian Society for Cybernetic Studies, Hemisphere Publ. Co.,NY:1982.
- [Raitz et al., 1997] RAITZ, R.; SOUZA, J.A.; DANDOLINI, G. A.; PACHECO, R.; MARTINS, A., GAUTHIER, F & BARCIA, R.. Learning by Means of Free Associative Neurons. *NAFIPS 97: Annual Meeting Of The North American Fuzzy Information Processing Society*, September 21-24, Syracuse, New York, 1997.
- [Ram, 1990] RAM, S. e RAM, S. Screening Financial Innovations: An Expert System Approach. *IEEE Expert*, Vol. 5, No. 9, pp. 20-27, August - 1990.

- [Refenes, 1995] REFENES, A. (Editor) *Neural Networks in the Capital Markets*. Wiley Finance Edition, 1995.
- [Rege, 1984] REGE, U. P. Accounting Ratios to Locate takeover Targets. *Journal of Business Finance and Accounting*, pp. 391-411, outono, 1984.
- [Richter, 1993] RICHTER, MM. Similarity, uncertainty and case-based reasoning in PATDEX. In R.S. Boyer (ed.): *Automated reasoning, essays in honour of Woody Bledsoe*. Kluwer, pp. 249-265, 1993.
- [Ritter et al., 1992] RITTER, H. J., MARTINETZ, T. e SCHULTEN, K. J. *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, Reading, MA, 1992.
- [Rosen & Silverman, 1992] ROSEN, E. M. e SILVERMAN, D. C. Corrosion Prediction from Polarization Scans Using an Artificial Neural Network Integrated with an Expert System Corrosion, September, pp. 734-744, 1992.
- [Rosenblatt, 1958] ROSENBLATT, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65, 386-408, 1958.
- [Rosenblatt, 1962] _____ *Principles of Neurodynamics*. Washington, DC: Spartan Books, 1962.
- [Rosenfeld, 1993] ROSENFELD, E. Intelligence , (a periodical). To subscribe, call 800-NEURALS, ou send e-mail to ier@aol.com.
- [Rumelhart et al., 1986] RUMELHART, D. E.; HINTON, G. E. e WILLIAMS, R. J. *Learning internal representation by error propagation*. In Rumelhart DE., McClelland JL (eds.), *Parallel Distributed Processings Exploration in the Microstructure of Cognition*, vol. 1:Foundations. Cambridge, Mass, MIT Press, 1986.
- [Saha et al., 1991] SAHA, A. et al. Oriented non-radial basis functions for Image Coding and analysis. In *Advance In Neural Information Processing systems* 3, R. P. Lippmann, J. E. Moody and D. S. Touretzky (Eds.), Morgan Kaufmann, San Mateo, CA, pp. 728-134, 1991.

- [Salmi & Martikainen, 1994] SALMI, T., & MARTIKAINEN, T. A Review of the Theoretical and Empirical Basis of Financial Ratio Analysis. *Finish Journal of Business Economics* 4(94), 1994.
- [Schank, 1982] SCHANK, R. *Dynamic Memory: a theory of reminding and learning in computers and people*. Cambridge University Press, 1982.
- [Self, 1990] SELF, K. Designing with Fuzzy Logic. *IEEE Spectrum*, pp. 42-44, Nov. 1990.
- [Senjen et al., 1993] SENJEN, R.; BELER, M.; LECKIE, C. e ROWLES, C. Hybrid Expert Systems for Monitoring and Fault Diagnosis. *Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications*, Orlando, FL, March 1-5, pp. 235-241, 1993.
- [Silva, 1988] SILVA, J. P. *Análise Financeira das Empresas.*, Ed. Atlas, 1988.
- [Simpson, 1985] SIMPSON. L. A computed model of case-based reasoning in problem solving. An investigation in the domain of dispute mediation. *Technical Report GIT-ICS-85/18*, Georgia Institute of Technology, 1985.
- [Sone, 1993] SONE, T. Using Distributed Neural Networks to Identify Faults in Switching Systems, *Proceedings of the International Workshop on applications of Neural Networks to Telecommunications*, J. Alspector, R. Goodman e T. X. Brown (eds.), Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
- [Souza, 1995] SOUZA, M. C. A. F.de. *Pequenas e Médias Empresas na Reestruturação Industrial*, Sebrae, 1995.
- [Spooner & Passino, 1996] SPOONER, J.T. e PASSINO, K. M. Stable Adaptive Control Using Fuzzy Systems and Neural Networks, *IEEE Transactions on Fuzzy*, Vol. 4, Nº 3, pp. 339-359, Agosto, 1996.
- [Sycara, 1988] SYCARA, K. Using case-based reasoning for plan adaptation and repair. *Proceedings Case-Based Reasoning Workshop*, DARPA. Clearwater beach, Florida, Morgan Kaufman, pp. 425-434, 1988.
- [Syswerda & Palmucci, 1991] SYSWERDA, G. e PALMUCCI, J. The Application of Genetic Algorithm to Resource Scheduling. *In Proceedings of the Fourth*

- International Conference on Genetic Algorithms*, R. K. Belew e L. B. Booker (eds.), San Mateo, 1991.
- [Szu et al., 1992] SZU, W. K.; LIU, W.; CHAO, C. C.; LIN, H. F.; HSU, T. e MEDSKER, L. *Proceedings of the IJCNN-92*, Beijing, pp. I-333, I-339, 1992.
- [Tamari, 1978] TAMARI, M. *Financial Ratios. Analysis and Prediction*. Paul Elek Ltd., London, 1978.
- [Thomas & Evanson, 1986] THOMAS, J. & EVANSON, R. V. *An Empirical Investigation of Association Between Financial Ratio use and Small Business Success*, pp. 555-565, 1986.
- [Thorpe et al., 1991] THORPE, C.; KANADE, T. e SHAFER, S. Toward Autonomous Driving: The CMU NavLab, *IEEE Expert*, Agosto, pp. 31-42, 1991.
- [Tirri, 1995] TIRRI, H. Replacing the pattern matcher of na expert system with a neural network. *Em [Goonatilake & Kheball, 1995]*.
- [Turban, 1992] TURBAN, E. *Expert Systems and Applied Artificial Intelligence*. New York: Macmillan, 1992.
- [Turban, 1995] _____ *Decision Support and Expert Systems - Management Support Systems*, 4ª Edição, Ed. Prentice Hall International, 1995.
- [Umbers e King, 1980] UMBERS, I.G. and KING, P.J. An analysis of human decision-making in cement kiln control and the implications for automation. *Int. Jrnl. of Man-Mach. Stud.*, Vol. 12, pp. 11-23, 1980.
- [Ventamakaran, 1983] VENTAMAKARAN, S. A rule-rule-case based system for image analysis. In: *First European Workshop on Case-based Reasoning, Posters and Presentations*, 1-5 November 1993. Vol.II. University of Kaiserslautern, pp. 410-415. 1993.
- [Waller, 1989] WALLER, L. Fuzzy Logic: It's Comprehensible, It's Plactical - And It's Commerical. *Electronics*, pp. 102-103, 1989.
- [Wasserman & Schwartz, 1988] WASSERMAN, P. D. e SCHWARTZ, T. Neural Networks, Part 2: What are They and Why is Everybody so Interested in Them Now? *IEEE Expert*, Spring, pp.10-12, 1988.

- [Watson, 1990] WATSON, C. J. *Multivariate Distributional Properties, Outliers, and Transformation of Financial Ratios*?. Accounting Review 65/3, 682-695, 1990.
- [Weber, 1993] WEBER, R.O. *Sistema Especialista Difuso Para Análise De Crédito*. Master Thesis, Universidade Federal de Santa Catarina UFSC, Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, 1993.
- [Weber-Lee, 1996] WEBER-LEE, R.. Case-Based Reasoning. *Home page. EPS-UFSC.BR*, 1996.
- [Weber-Lee et al., 1995] WEBER-LEE, R; BARCIA, R. M., KHATOR, S., Case-Based Reasoning for Cash Flow Forecasting using Fuzzy Retrieval, in M. Veloso, and A. Aamodt, (Eds.), *Case-Based Reasoning Research and Development*, First International Conference, ICCBR-95, Sesimbra, Portugal, October 23 - 26, 1995.
- [Werbos, 1974] WERBOS, P. J. *Beyond Regression: New tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. Thesis, Harvard University, Cambridge, MA, 1974.
- [Werbos, 1982] _____ Application of Advances in Nonlinear Sensitivity Analysis. Em *System Modeling and Optimization: Proc. Of the Int. Federation for Information Processes*, R. Drenick & F. Kozin (eds.), pp. 762-770, 1982.
- [Werbos, 1990] _____ Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 87:10, 1990.
- [Whalen & Schott, 1985] WHALEN, T. & SCHOTT, B. Generalized Network Modeling and Diagnosis Usilg Financial Ratios. *Information Sciences* 37, 1985, pp 179-192.
- [White et al., 1994] WHITE, G.I., SOHNDI, A. C. & FRIED, D. *The Analysis and Use of Financial Statements*. John Wiley & Sons, Inc., New York, 1994.
- [Whiteley & Davis, 1993] WHITELEY, J. R. & DAVIS, J. F. Qualitative Interpretation of Sensor Patterns. *IEEE Expert*, Abril, pp. 54-63, 1993.
- [Widrow & Hoff, 1960] WIDROW, B. & Hoff, M. E. Adaptative Switching Circuits. *IRE WESCON Convention Record*, pp. 56-104, 1960.

- [Widrow et al., 1994] WIDROW, B.; RUMELHART, D., e LEHR, M. Neural Networks: Applications in Industry, Business and Science. *Communications of the ACM*, Vol. 37, No 3, pp. 490-501, 1994.
- [Wiener & Pinson, 1988] WIENER, R. S. & PINSON, J. *An Introduction to Object-Oriented Programming and C++*. Addison-Wesley, Reading, MA, 1988.
- [Wilson & Hoff, 1994] WILSON, R. L. & HOFF, R. Bankruptcy Prediction Using Neural Networks. *Decision Support Systems*, Vol. 11, Nº 5, pp. 545-557, 1994.
- [Yanicoglu & Sandon, 1993] YANICOGLU, B. A., e SANDON, P. A. Off Line Cursive Handwriting Recognition Using Neural Networks. *Proceedings of the SPIE Applications of Artificial Neural Networks IV*, Orland, FL, pp. 102-6, 1993.
- [Yli-Olli & Virtanen, 1986] YLI-OLLI, P., AND VIRTANEN, I. Classification pattern of financial ratios. A comparative analysis between US and Finnish firms on the aggregate level. *Finnish Journal of Business Economics* 2, 112-132, 1986.
- [Yli-Olli & Virtanen, 1989] YLI-OLLI, P., & VIRTANEN, I., "On the long-term stability and cross-country invariance of financial ratio patterns". *European Journal of Operational Research* 39/1, 40-53, 1989.
- [Yli-Olli & Virtanen, 1990] YLI-OLLI, P., & VIRTANEN, I. Transformation analysis applied to long-term stability and structural invariance of financial ratio patterns: U.S. vs. Finnish firms. *American Journal of Mathematical and Management Sciences* 10/1-2, pp. 73-125, 1990.
- [Yonn et al., 1994] YONN, Y., GUIMARÃES, T. & SWALES, G. Integrating Artificial Neural Networks with Rule-Based Expert Systems. *Decision Support Systems*, Vol. 11, pp. 497-507, 1994.
- [Zadeh, 1965] ZADEH, L.A. *Fuzzy sets*. *Information and Control*, Vol. 8, 1965, pp. 338-353.
- [Zadeh, 1968] _____ *Fuzzy algorithms*. *Info. & Ctl.*, Vol. 12, 1968, pp. 94-102.
- [Zadeh, 1984] _____ Making computers think like people. *IEEE Spectrum* 8, pp. 26-32, 1984.

[Zahed, 1993] _____ *Intelligence Systems for Busines. Expert Sytems with Neural Networks*, Belmont California, 1993.

APÊNDICE A

DADOS DE TREINAMENTO E TESTE

Nº	ANO	C. Ope	Liq	CG V	TIE	Endiv	SITUAÇÃO
1	95	1.7	1.4	3.2	376	1222	6
2	94	-0.7	-0.7	5.3	-88	390	2
3	95	26.9	25.6	75.0	9	13	3
4	94	-9.0	-9.0	259.4	-1171	120	5
5	95	3.5	3.1	25.0	2	75	4
6	94	0.8	0.5	28.5	1	75	6
7	95	1.7	1.4	6.5	177	58	2
8	94	-0.2	-0.2	3.8	-49	64	2
9	95	-2.4	-2.4	63.1	0	32	5
10	94	13.5	12.1	47.8	6	12	3
11	95	35.4	27.5	0.1	464	6	1
12	94	46.9	40.5	8.5	22253	14	1
13	95	15.5	13.0	1.5	67	39	1
14	94	27.4	27.4	0.5	265	27	1
15	95	0.1	0.1	82.4	1	231	7
16	93	7.4	5.2	265.9	18	62	3
17	95	3.7	2.8	104.2	2	28	3
18	94	29.4	28.0	68.2	13	12	3
19	95	36.5	28.8	3.1	102	10	1
20	95	9.5	8.1	116.1	10	76	3
21	94	6.3	4.7	67.2	23	26	3
22	95	-3.4	-3.4	20.6	1	20	2
23	93	87.3	79.9	62.3	196	15	3
24	95	3.7	2.1	3.2	95	8	1
25	94	6.0	6.0	3.9	284	30	1
26	95	68.0	65.5	1.6	682	11	1
27	95	5.8	4.4	43.0	117	30	3
28	94	1.2	0.9	38.1	231	109	5
29	95	56.0	53.6	1.4	2681	3	1
30	94	31.0	29.6	0.2	222	5	1
31	95	1.9	1.4	55.2	4	494	7
32	93	0.1	0.1	11.3	1	900	6
33	95	3.4	2.5	3.6	803	4	1
34	94	1.7	1.2	5.4	159	1	2
35	95	18.8	17.1	2.5	66	16	1
36	94	21.4	14.6	5.5	1003	72	1
37	95	2.8	2.2	5.7	2	171	2
38	94	0.8	0.8	5.5	2	125	2
39	95	53.2	50.7	1.8	324	1	1
40	94	-55.7	-55.7	0.8	1	110	2

Tabela A-1 : Dados de treinamento.

Nº	ANO	C. Ope	Liq	CG V	TIE	Endiv	SITUAÇÃO
40	94	-55.7	-55.7	0.8	1	110	2
41	95	55.0	52.6	-7.6	477	1	1
42	94	10.8	8.1	6.0	147	30	1
43	95	1.2	0.9	73.9	2	1178	7
44	94	-0.2	-0.2	6.9	1	743	6
45	95	24.7	17.3	5.2	156	70	1
46	94	-3.6	-3.6	18.2	-7	179	6
47	95	1.3	1.0	7.4	817	144	2
48	95	-5.5	-5.5	8.9	-2	600	6
49	95	91.8	84.3	10.7	191	6	1
50	95	-23.7	-24.9	115.9	-33	238	7
51	94	1.6	1.4	97.8	4	792	7
52	95	-0.1	-0.1	295.2	0	1198	7
53	95	-1.6	-1.6	12.5	0	13	2
54	95	14.0	10.5	1.5	528	1	1
55	95	47.8	47.8	219.8	107	62	3
56	93	53.0	50.2	48.6	141	24	3

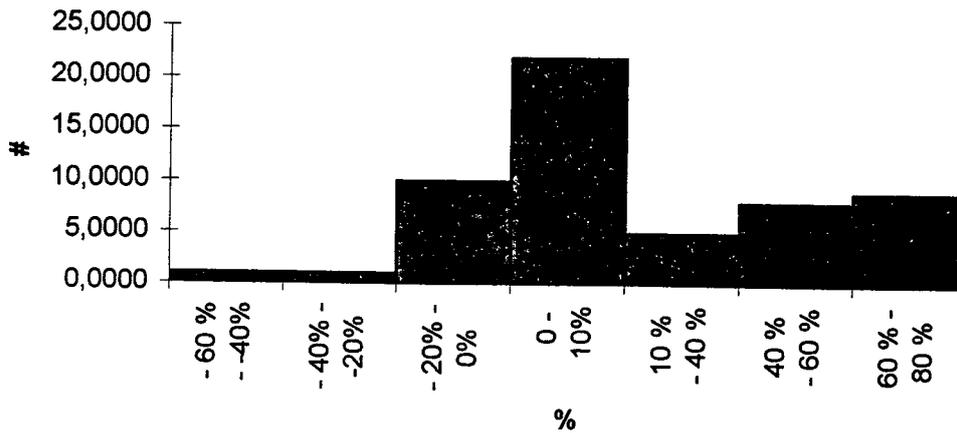
Tabela A-2 : Dados de Teste.

- * Observação: A classificação da última coluna indica:
- (1) rentabilidade
 - (2) giro
 - (3) endividamento
 - (4) rentabilidade e giro
 - (5) rentabilidade e endividamento
 - (6) giro e endividamento
 - (7) nenhum dos anteriores

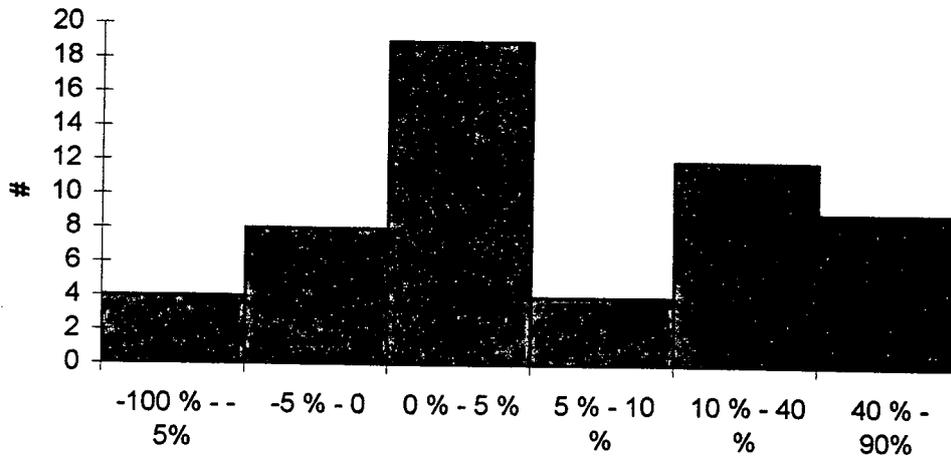
APÊNDICE B

HISTOGRAMAS DOS ÍNDICES FINANCEIROS

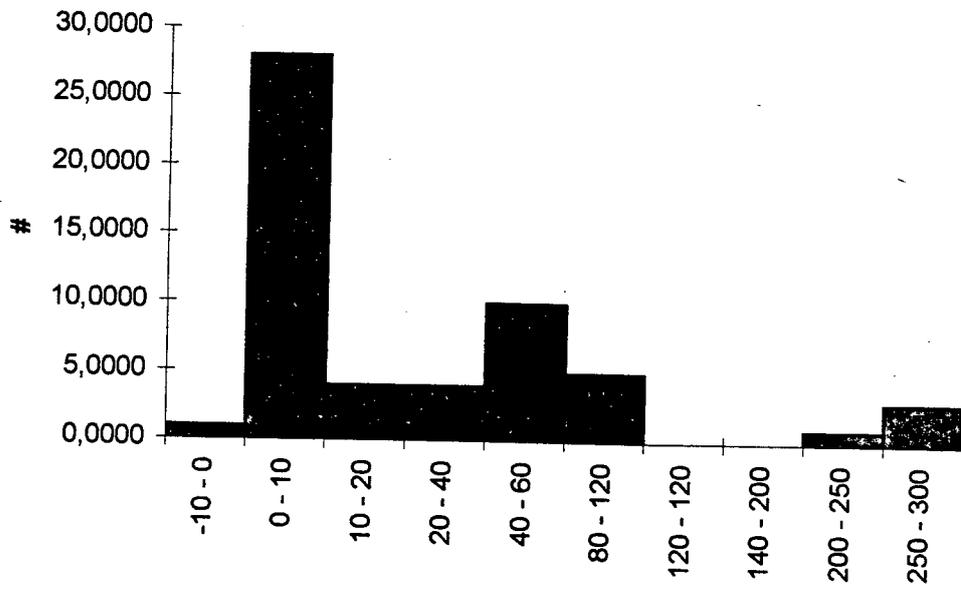
Caixa Op



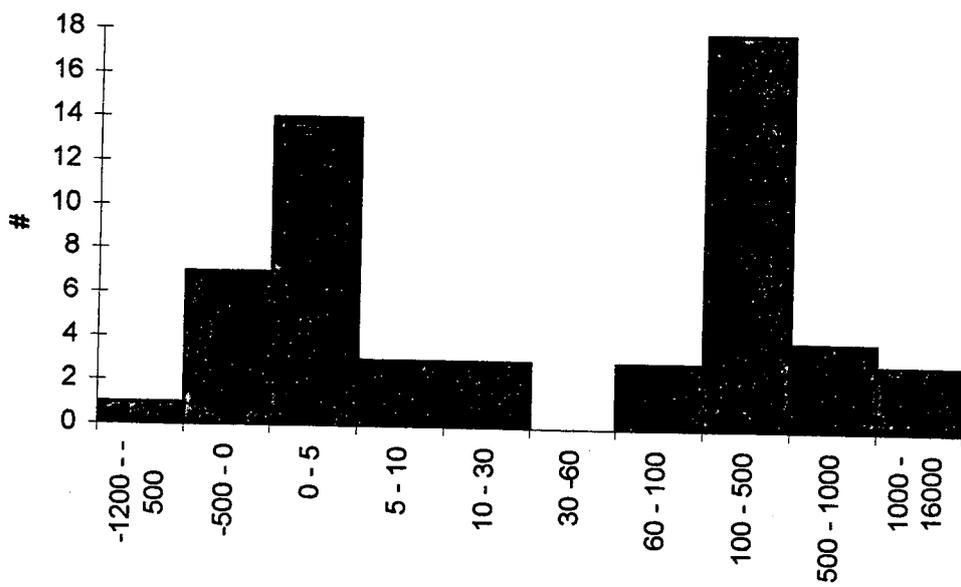
LLiquido



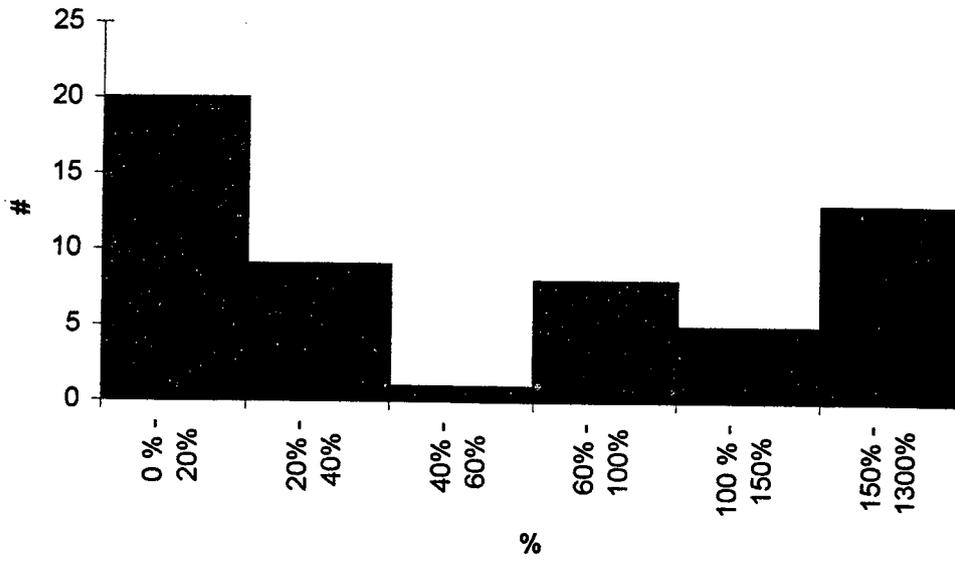
CG d de V



TIE



Endividamento



APÊNDICE C

RESULTADOS DA BACKPROPAGATION

BACKPROPAGATION COM DADOS NORMALIZADOS NO INTERVALO [-1,1] E FUNÇÃO DE ATIVAÇÃO SIGMOIDAL						
Neurônios	Épocas	Taxa de Aprendizagem	Momentum	Treinamento (%)	Teste (%)	Tempo (s)
1	1000	0,010	0,90	60,00	50,00	11,20
3	1000	0,010	0,90	72,50	68,75	11,75
5	1000	0,010	0,90	75,00	56,25	12,53
7	1000	0,010	0,90	82,50	62,50	13,35
9	1000	0,010	0,90	80,00	62,50	14,01
11	1000	0,010	0,90	87,50	75,00	14,77
13	1000	0,010	0,90	80,00	50,00	15,60
15	1000	0,010	0,90	85,00	68,75	16,21
17	1000	0,010	0,90	82,50	68,75	16,81
19	1000	0,010	0,90	87,50	75,00	17,68
21	1000	0,010	0,90	82,50	68,75	18,29
23	1000	0,010	0,90	77,50	68,75	19,22
30	1000	0,010	0,90	85,00	62,50	22,24
7	3000	0,001	0,90	90,00	68,75	39,76
7	2000	0,001	0,90	82,50	68,75	29,44
11	2000	0,010	0,90	87,50	68,75	29,28
11	2000	0,100	0,80	82,50	68,75	29,06
9	2000	0,700	0,10	65,00	56,25	27,95
7	3000	0,800	0,10	75,00	56,25	39,93
9	3000	0,010	0,90	92,50	56,25	41,74
9	1500	0,010	0,90	85,00	62,50	21,04
7	3000	0,001	0,90	87,50	68,75	39,93
7	2000	0,010	0,80	80,00	56,25	26,15
11	1500	0,600	0,01	72,50	68,75	22,24
13	1000	0,200	0,90	77,50	56,25	15,44
8	2000	0,001	0,95	90,00	75,00	26,86
5	200	0,010	0,90	57,50	56,25	2,80
5	400	0,010	0,90	67,50	56,25	5,22
5	600	0,010	0,90	57,50	50,00	7,58
5	700	0,010	0,90	72,50	68,75	9,01
5	900	0,010	0,90	65,00	50,00	11,26
5	1000	0,010	0,90	75,00	68,75	12,52
10	2000	0,020	0,90	82,50	68,75	17,41
10	1000	0,200	0,80	87,50	68,75	29,16
10	2000	0,200	0,80	75,00	75,00	14,06
10	2000	0,200	0,80	80,00	56,25	28,02
7	2000	0,700	0,40	87,50	62,50	25,59
7	2000	0,001	0,90	87,50	56,25	25,76
7	2000	0,900	0,10	77,50	56,25	26,26
9	3000	0,0010	0,80	82,50	75,00	41,14
7	3000	0,010	0,90	90,00	75,00	39,10
9	2500	0,001	0,90	85,00	68,75	34,49

11	2500	0,600	0,20	80,00	68,75	36,03
6	3000	0,800	0,10	82,50	68,75	38,29
6	3000	0,800	0,30	82,50	56,25	36,64
7	3000	0,200	0,90	87,50	75,00	39,05
7	5000	0,900	0,01	87,50	56,25	66,62
13	1000	0,700	0,50	72,50	56,25	15,32
13	1000	0,010	0,90	87,50	68,75	15,43
11	2000	0,100	0,80	82,50	62,50	28,94
7	3000	0,001	0,90	90,00	75,00	38,95
7	2500	0,001	0,90	90,00	75,00	32,41
7	2000	0,001	0,90	82,50	68,75	25,98
7	2000	0,001	0,90	80,00	68,75	25,87
7	3000	0,001	0,93	90,00	75,00	38,94
9	3000	0,050	0,92	85,00	56,25	41,30
7	3000	0,050	0,90	87,50	68,75	38,95
7	3000	0,0010	0,90	85,00	62,50	38,77
5	3000	0,020	0,92	87,50	56,25	36,69
10	500	0,010	0,80	67,50	50,00	7,30
10	1000	0,010	0,80	70,00	62,50	14,11
11	2000	0,300	0,80	87,50	62,50	31,80
7	2500	0,700	0,20	72,50	56,25	34,05
15	1000	0,010	0,80	82,50	68,75	16,75
7	1200	0,100	0,90	85,00	68,75	16,81
7	2000	0,100	0,90	87,50	68,75	27,90
7	2500	0,100	0,90	87,50	75,00	34,76
7	1000	0,100	0,90	85,00	62,50	14,18
7	2000	0,100	0,90	87,50	62,50	27,96
7	2500	0,100	0,90	87,50	62,50	35,04
7	3000	0,010	0,90	87,50	62,50	42,13
7	500	0,010	0,80	60,00	56,25	7,25
7	2000	0,010	0,90	77,50	50,00	27,68
7	3000	0,100	0,90	82,50	56,25	41,47
9	1000	0,010	0,90	77,50	50,00	13,85
9	500	0,200	0,80	77,50	68,75	7,58
9	800	0,020	0,80	77,50	68,75	11,91
9	2000	0,020	0,80	90,00	68,75	29,55
1	3000	0,030	0,80	87,50	75,00	44,21
9	5000	0,001	0,80	87,50	50,00	74,15
9	2500	0,001	0,90	87,50	75,00	37,13
9	2800	0,001	0,90	87,50	75,00	41,30
9	3200	0,001	0,90	87,50	62,50	47,34
9	4000	0,800	0,20	90,00	68,75	58,99
9	2000	0,900	0,10	77,50	68,75	29,77

Tabela C-1: Resultado da Backpropagation usando função de ativação sigmoide na camada escondida e linear na última camada.

BACKPROPAGATION COM DADOS NORMALIZADOS NO INTERVALO [-1,1] E FUNÇÃO DE ATIVAÇÃO TANGENTE HIPERBÓLICA.						
Neurônios	Épocas	Taxa de Aprendizagem	Momentum	Treinamento (%)	Teste (%)	Tempo (s)
13	2000	0,100	0,90	87,50	68,75	31,20
9	2000	0,200	0,90	90,00	75,00	28,23
7	2000	0,010	0,90	85,00	68,75	26,47
7	3000	0,010	0,90	85,00	62,50	40,10
7	2000	0,100	0,90	72,50	68,75	26,36
9	1000	0,100	0,90	82,50	56,25	14,17
5	2000	0,200	0,95	82,50	62,50	25,05
5	2000	0,500	0,80	80,00	68,75	24,71
7	1000	0,600	0,10	75,00	68,75	13,46
7	1000	0,600	0,01	67,50	56,25	13,51
7	1000	0,600	0,20	75,00	68,75	13,23
7	2000	0,600	0,01	65,00	56,25	26,86
9	1000	0,600	0,10	72,50	68,75	14,23
6	1000	0,500	0,50	75,00	62,50	12,74
6	1000	0,800	0,40	70,00	68,75	12,75
7	2000	0,010	0,90	80,00	68,75	26,47
6	3000	0,001	0,90	85,00	75,00	38,55
9	3000	0,001	0,80	87,50	75,00	42,13
8	2000	0,800	0,10	62,50	62,50	27,46
9	200	0,100	0,90	77,50	68,75	3,13
9	400	0,100	0,90	77,50	75,00	5,88
9	500	0,900	0,20	70,00	62,50	7,36
9	600	0,900	0,20	70,00	68,75	8,51
9	800	0,900	0,20	77,50	68,75	11,20
9	1000	0,900	0,10	72,50	68,75	14,06
9	1500	0,900	0,20	82,50	75,00	20,98
9	2000	0,900	0,20	52,50	43,75	27,58
9	2000	0,900	0,20	82,50	56,25	27,96
9	2000	0,200	0,90	85,00	75,00	28,23
9	200	0,200	0,90	70,00	43,75	3,13
9	400	0,200	0,90	75,00	62,50	5,82
9	500	0,200	0,90	72,50	68,75	7,30
7	1000	0,300	0,80	75,00	75,00	14,01
7	1500	0,100	0,90	80,00	62,50	20,99
7	2000	0,200	0,80	85,00	62,50	27,85
9	3000	0,020	0,90	90,00	68,75	45,04
7	3000	0,010	0,80	87,50	56,25	41,86

Tabela C-2: Resultado da Backpropagation usando função de ativação tangente hiperbólica na camada escondida e linear na última camada.

BACKPROPAGATION COM DADOS NORMALIZADOS NO INTERVALO [0,1] E FUNÇÃO DE ATIVAÇÃO SIGMOIDAL						
Neurônios	Épocas	Taxa de Aprendizagem	Momentum	Treinamento (%)	Teste (%)	Tempo (s)
9	4000	0,100	0,90	82,50	37,50	54,98
7	2000	0,010	0,90	82,50	56,25	26,09
15	1000	0,800	0,20	62,50	37,50	15,93
13	800	0,400	0,10	60,00	56,25	13,24
6	2000	0,010	0,90	87,50	43,75	25,15
11	2000	0,100	0,90	82,50	56,25	28,84
11	3000	0,010	0,90	77,50	50,00	43,06
11	5000	0,001	0,90	87,50	43,75	72,88
20	1000	0,010	0,90	75,00	50,00	17,63
20	1000	0,200	0,90	82,50	43,75	17,63
15	2000	0,700	0,10	72,50	31,25	32,35

Tabela C-3 : Resultado da Backpropagation usando função de ativação sigmoideal na camada escondida e linear na última camada.

BACKPROPAGATION COM DADOS NORMALIZADOS NO INTERVALO [-1,1] E FUNÇÃO DE ATIVAÇÃO NÃO LINEAR NAS DUAS CAMADAS						
Neurônios	Épocas	Taxa de Aprendizagem	Momentum	Treinamento (%)	Teste (%)	Tempo (s)
1	500	0,700	0,30	35,00	31,25	5,66
5	1000	0,800	0,10	35,00	31,25	12,41
1	500	0,010	0,90	35,00	31,25	5,71
10	1000	0,700	0,20	35,00	31,25	14,39
5	1000	0,010	0,90	35,00	31,25	12,36
6	1000	0,010	0,90	35,00	31,25	12,63
7	1000	0,010	0,90	35,00	31,25	12,96
20	1000	0,010	0,90	35,00	31,25	18,62
7	1000	0,900	0,10	35,00	31,25	12,97

Tabela C-4 : Resultado da Backpropagation usando função de ativação não linear (sigmoideal e tangente hiperbólica) na camada escondida e na camada de saída.

APÊNDICE D

RESULTADOS DA RBF

RBF COM DADOS NORMALIZADOS NO INTERVALO [-1,1]					
Centros	Raio	Épocas	Treinamento (%)	Teste (%)	Tempo(s)
20	1,10	20	87,50	56,25	15,00
10	1,50	10	70,00	68,75	6,31
6	1,30	6	57,50	56,25	13,40
9	1,80	9	62,50	50,00	8,68
12	1,20	12	60,00	75,00	10,05
11	1,30	11	65,00	56,25	7,03
12	1,50	12	67,50	68,75	5,11
11	1,50	11	62,50	75,00	15,98
10	1,00	10	60,00	50,00	12,96
17	1,00	17	87,50	50,00	10,33
16	1,00	16	85,00	56,25	5,99
15	1,00	15	75,00	62,50	3,35
18	1,00	18	85,00	31,25	5,49
15	1,30	15	77,50	81,25	6,21
12	1,60	12	72,50	50,00	20,05
30	1,20	30	90,00	43,75	7,36
30	1,50	30	90,00	31,25	12,25
39	1,00	40	100,00	31,25	12,14
10	1,80	10	65,00	56,25	13,57
35	1,20	34	100,00	37,50	15,32
20	1,20	20	85,00	68,75	7,25
20	1,60	20	87,50	56,25	9,28
20	1,40	20	87,50	68,75	9,34
20	1,50	20	90,00	62,50	10,88
20	1,70	20	85,00	68,75	6,16
20	1,50	20	90,00	62,50	13,13
30	1,20	30	90,00	43,75	8,13
30	1,20	30	90,00	43,75	16,37
30	2,00	30	90,00	25,00	13,56
30	1,50	30	90,00	31,25	6,92
30	1,40	30	90,00	31,25	7,80
30	1,20	30	90,00	43,75	10,21
3	1,20	2	50,00	37,50	9,28
7	1,00	6	65,00	56,25	4,07
11	1,40	10	62,50	56,25	6,43
7	1,00	6	65,00	56,25	8,19
7	1,00	6	65,00	56,25	0,77
8	1,00	7	57,50	56,25	0,88
9	1,00	8	57,50	56,25	0,94
10	1,00	9	60,00	50,00	0,99

11	1,00	10	60,00	50,00	0,99
12	1,00	11	65,00	50,00	1,15
13	1,00	12	62,50	62,50	1,21
14	1,00	13	75,00	56,25	1,26
15	1,00	14	75,00	62,50	1,32
16	1,00	15	85,00	56,25	1,37
7	1,10	6	60,00	50,00	0,77
8	1,10	7	60,00	56,25	0,93
9	1,10	8	62,50	56,25	0,93
10	1,10	9	65,00	62,50	0,93
11	1,10	10	65,00	68,75	1,04
12	1,10	11	65,00	75,00	1,21
13	1,10	12	67,50	62,50	1,16
14	1,10	13	70,00	62,50	1,27
15	1,10	14	80,00	62,50	1,32
16	1,10	15	80,00	68,75	1,54
7	1,20	6	65,00	50,00	0,77
8	1,20	7	62,50	56,25	0,88
9	1,20	8	70,00	62,50	0,94
10	1,20	9	62,50	50,00	0,99
11	1,20	10	65,00	75,00	0,99
12	1,20	11	60,00	75,00	1,15
13	1,20	12	72,50	68,75	1,21
14	1,20	13	80,00	75,00	1,43
15	1,20	14	80,00	68,75	1,37
16	1,20	15	82,50	68,75	1,38
7	1,30	6	62,50	50,00	0,83
8	1,30	7	65,00	56,25	0,93
9	1,30	8	57,50	50,00	0,99
10	1,30	9	62,50	56,25	1,05
11	1,30	10	65,00	56,25	1,04
12	1,30	11	65,00	43,75	1,21
13	1,30	12	65,00	68,75	1,38
14	1,30	13	77,50	81,25	1,32
15	1,30	14	77,50	81,25	1,38
16	1,30	15	82,50	75,00	1,38
7	1,40	6	65,00	56,25	0,83
8	1,40	7	65,00	56,25	0,94
9	1,40	8	62,50	50,00	0,94
10	1,40	9	62,50	56,25	0,99

Tabela D-1 : Resultados do desempenho da rede RBF usando os dados normalizados no intervalo [-1,1].

RBF COM DADOS NORMALIZADOS NO INTERVALO [0,1]					
Centro	Raio	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
15,00	1,30	14,00	77,50	50,00	7,85
10,00	1,00	9,00	60,00	25,00	4,45
15,00	1,30	14,00	77,50	50,00	5,60
16,00	0,80	15,00	77,50	31,25	10,32
16,00	1,20	15,00	77,50	43,75	7,52
16,00	1,30	15,00	75,00	50,00	8,79
16,00	1,50	15,00	77,50	50,00	4,88
13,00	1,20	12,00	70,00	50,00	8,02
11,00	1,30	10,00	65,00	43,75	5,49
7,00	1,50	6,00	65,00	56,25	5,71
7,00	1,20	6,00	62,50	62,50	4,67
7,00	1,30	6,00	65,00	56,25	6,26
7,00	1,00	6,00	65,00	25,00	3,35
7,00	1,00	6,00	65,00	25,00	4,18
9,00	1,30	8,00	67,50	50,00	6,15
20,00	1,00	19,00	87,50	31,25	16,81
19,00	1,20	18,00	85,00	43,75	8,35
12,00	2,00	11,00	65,00	50,00	19,45
12,00	1,50	11,00	65,00	62,50	7,85
12,00	1,40	11,00	65,00	62,50	4,88
12,00	1,30	11,00	65,00	50,00	5,66
12,00	1,20	11,00	65,00	62,50	6,20
12,00	1,10	11,00	65,00	43,75	5,44
18,00	1,20	17,00	77,50	50,00	12,20
5,00	1,50	4,00	47,50	43,75	5,66
5,00	1,00	4,00	50,00	37,50	8,35
8,00	1,30	7,00	65,00	43,75	5,66
9,00	0,80	8,00	62,50	50,00	6,92
9,00	0,50	8,00	62,50	31,25	6,86
10,00	0,80	9,00	67,50	50,00	6,20
7,00	0,50	6,00	62,50	31,25	9,78
7,00	0,90	6,00	65,00	50,00	8,85
7,00	1,00	6,00	65,00	25,00	0,93
5,00	1,00	4,00	50,00	37,50	0,77
30,00	1,00	29,00	90,00	6,25	2,42
15,00	1,60	14,00	77,50	50,00	1,37
25,00	0,70	24,00	90,00	6,25	2,14
11,00	0,90	10,00	67,50	56,25	1,10
11,00	0,70	10,00	67,50	31,25	1,10
14,00	1,10	13,00	77,50	37,50	1,32
7,00	1,00	6,00	65,00	25,00	0,88
8,00	1,00	7,00	67,50	31,25	0,93
9,00	1,00	8,00	57,50	25,00	1,15
10,00	1,00	9,00	60,00	25,00	1,04
11,00	1,00	10,00	72,50	50,00	1,10
12,00	1,00	11,00	70,00	50,00	1,26

13,00	1,00	12,00	80,00	56,25	1,32
14,00	1,00	13,00	80,00	56,25	1,37
15,00	1,00	14,00	77,50	56,25	1,43
16,00	1,00	15,00	77,50	43,75	1,60
7,00	0,90	6,00	65,00	50,00	0,82
8,00	0,90	7,00	65,00	56,25	0,94
9,00	0,90	8,00	62,50	62,50	0,94
10,00	0,90	9,00	72,50	50,00	0,99
11,00	0,90	10,00	67,50	56,25	1,04
12,00	0,90	11,00	67,50	50,00	1,21
13,00	0,90	12,00	70,00	43,75	1,26
14,00	0,90	13,00	77,50	50,00	1,32
15,00	0,90	14,00	80,00	50,00	1,43
16,00	0,90	15,00	80,00	56,25	1,43
7,00	0,80	6,00	65,00	50,00	0,82
8,00	0,80	7,00	62,50	50,00	0,88
9,00	0,80	8,00	62,50	50,00	0,99
10,00	0,80	9,00	67,50	50,00	1,04
11,00	0,80	10,00	67,50	50,00	1,10
12,00	0,80	11,00	77,50	56,25	1,21
13,00	0,80	12,00	72,50	43,75	1,37
14,00	0,80	13,00	75,00	62,50	1,32
15,00	0,80	14,00	77,50	50,00	1,38
16,00	0,80	15,00	77,50	31,25	1,37
7,00	0,70	6,00	57,50	31,25	0,83
8,00	0,70	7,00	60,00	31,25	0,94
9,00	0,70	8,00	60,00	43,75	0,99
10,00	0,70	9,00	60,00	37,50	1,04
11,00	0,70	10,00	67,50	31,25	1,05
12,00	0,70	11,00	80,00	43,75	1,38
13,00	0,70	12,00	82,50	43,75	1,26
14,00	0,70	13,00	80,00	43,75	1,37
15,00	0,70	14,00	80,00	25,00	1,37
16,00	0,70	15,00	85,00	25,00	1,49
7,00	0,50	6,00	62,50	31,25	0,82
8,00	0,50	7,00	60,00	31,25	0,93
9,00	0,50	8,00	62,50	31,25	0,98
10,00	0,50	9,00	60,00	31,25	1,15
11,00	0,50	10,00	65,00	37,50	1,10
12,00	0,50	11,00	70,00	31,25	1,21
13,00	0,50	12,00	82,50	25,00	1,27
14,00	0,50	13,00	80,00	25,00	1,32
15,00	0,50	14,00	77,50	18,75	1,42
16,00	0,50	15,00	85,00	25,00	1,43
7,00	0,30	6,00	50,00	31,25	0,82
8,00	0,30	7,00	52,50	31,25	1,04
9,00	0,30	8,00	65,00	31,25	0,93
10,00	0,30	9,00	72,50	31,25	0,99
11,00	0,30	10,00	75,00	31,25	1,04

12,00	0,30	11,00	85,00	31,25	1,21
13,00	0,30	12,00	82,50	31,25	1,27
14,00	0,30	13,00	82,50	31,25	1,26
15,00	0,30	14,00	80,00	31,25	1,37
16,00	0,30	15,00	75,00	31,25	1,53
7,00	0,10	6,00	62,50	31,25	0,82
8,00	0,10	7,00	72,50	31,25	0,88
9,00	0,10	8,00	72,50	31,25	0,99
10,00	0,10	9,00	75,00	31,25	1,04
11,00	0,10	10,00	75,00	31,25	1,10
12,00	0,10	11,00	77,50	31,25	1,21
13,00	0,10	12,00	80,00	31,25	1,31
14,00	0,10	13,00	77,50	31,25	1,32
15,00	0,10	14,00	80,00	31,25	1,59
16,00	0,10	15,00	80,00	31,25	1,43
7,00	1,10	6,00	62,50	62,50	0,82
8,00	1,10	7,00	65,00	31,25	0,93
9,00	1,10	8,00	65,00	50,00	0,99
10,00	1,10	9,00	65,00	56,25	1,04
11,00	1,10	10,00	62,50	43,75	1,05
12,00	1,10	11,00	65,00	43,75	1,26
13,00	1,10	12,00	70,00	31,25	1,43
14,00	1,10	13,00	77,50	37,50	1,32
15,00	1,10	14,00	75,00	50,00	1,43
16,00	1,10	15,00	75,00	56,25	1,43
7,00	1,20	6,00	62,50	62,50	0,83
8,00	1,20	7,00	62,50	37,50	0,94
9,00	1,20	8,00	65,00	50,00	0,99
10,00	1,20	9,00	65,00	62,50	1,04
11,00	1,20	10,00	65,00	43,75	1,20
12,00	1,20	11,00	65,00	62,50	1,27
13,00	1,20	12,00	70,00	50,00	1,32
14,00	1,20	13,00	75,00	43,75	1,31
15,00	1,20	14,00	75,00	43,75	1,38
16,00	1,20	15,00	77,50	43,75	1,43
7,00	1,30	6,00	65,00	56,25	0,88
8,00	1,30	7,00	65,00	43,75	0,94
9,00	1,30	8,00	67,50	50,00	1,16
10,00	1,30	9,00	65,00	62,50	1,10
11,00	1,30	10,00	65,00	43,75	1,10
12,00	1,30	11,00	65,00	50,00	1,21
13,00	1,30	12,00	72,50	43,75	1,26
14,00	1,30	13,00	70,00	37,50	1,32
15,00	1,30	14,00	77,50	50,00	1,37
16,00	1,30	15,00	75,00	50,00	1,43
7,00	1,50	6,00	65,00	56,25	1,05
8,00	1,50	7,00	65,00	43,75	0,93
9,00	1,50	8,00	67,50	56,25	0,99
10,00	1,50	9,00	65,00	62,50	0,99

11,00	1,50	10,00	65,00	43,75	1,04
12,00	1,50	11,00	65,00	62,50	1,21
13,00	1,50	12,00	72,50	50,00	1,27
14,00	1,50	13,00	67,50	43,75	1,32
15,00	1,50	14,00	77,50	50,00	1,37
16,00	1,50	15,00	77,50	50,00	1,59

Tabela D-2 : Resultados do desempenho da rede RBF usando os dados normalizados no intervalo [0,1].

APÊNDICE E

RESULTADOS DA LVQ

Resultados da LVQ com os Dados Não Normalizados				
Taxa de Aprendizagem	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
0,10	23,00	37,50	50,00	4,17
0,10	20,00	32,50	25,00	3,57
0,10	21,00	50,00	50,00	3,79
0,10	22,00	37,50	43,75	3,95
0,10	24,00	40,00	50,00	4,45
0,10	2,00	27,50	0,00	0,33
0,10	25,00	50,00	43,75	4,45
0,10	26,00	45,00	43,75	4,66
0,10	19,00	37,50	12,50	3,52
0,10	27,00	37,50	43,75	4,83
0,10	28,00	40,00	50,00	4,99
0,20	23,00	17,50	31,25	3,68
0,10	25,00	50,00	31,25	3,96
0,01	20,00	27,50	0,00	3,19
0,01	50,00	37,50	0,00	7,96
0,40	40,00	20,00	31,25	6,43
0,30	20,00	17,50	31,25	3,19
0,90	20,00	25,00	0,00	3,52
0,80	40,00	5,00	18,75	6,21
0,60	50,00	20,00	0,00	7,86
0,50	25,00	32,50	31,25	3,95
0,40	25,00	12,50	31,25	3,90
0,40	40,00	20,00	31,25	6,32
0,20	25,00	22,50	31,25	3,90
0,20	4,00	12,50	0,00	0,66
0,10	20,00	32,50	31,25	3,24
0,01	25,00	30,00	12,50	3,95
0,01	30,00	30,00	12,50	4,84
0,01	15,00	30,00	0,00	2,47
0,10	23,00	37,50	31,25	3,68
0,10	24,00	40,00	31,25	3,85
0,10	21,00	50,00	31,25	3,35

Tabela E-1 : Resultados da rede LVQ usando os dados não normalizados e a taxa de aprendizagem decresce linearmente em função do número de épocas.

Resultados da LVQ com os Dados Não Normalizados				
Taxa de Aprendizagem	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
0,100	20	20,00	37,50	3,57
0,100	10	32,50	25,00	1,86
0,100	15	20,00	18,75	2,64
0,100	21	17,50	12,50	3,74
0,100	22	30,00	12,50	3,96
0,100	23	27,50	12,50	4,23
0,100	24	32,50	12,50	4,34
0,100	30	20,00	31,25	5,38
0,100	32	22,50	12,50	5,93
0,700	13	35,00	31,25	2,37
0,600	20	25,00	12,50	3,63
2,000	30	12,50	31,25	5,43
0,600	20	25,00	12,50	3,73
0,400	20	20,00	12,50	3,62
0,300	21	15,00	12,50	3,74
0,300	23	17,50	0,00	4,07
0,100	50	40,00	31,25	8,41
0,200	25	25,00	31,25	4,01
0,100	20	20,00	31,25	3,18
0,200	20	25,00	31,25	3,18
0,600	20	25,00	0,00	3,13
0,010	20	7,50	62,50	3,13
0,050	20	12,50	31,25	3,18
0,020	20	7,50	50,00	3,19
0,001	20	7,50	62,50	3,18
0,010	50	5,00	31,25	7,85
0,100	10	32,50	0,00	1,60
0,100	40	45,00	31,25	6,37
0,200	50	7,50	31,25	7,96
0,010	25	5,00	62,50	4,01
0,010	15	7,50	31,25	2,42
0,010	25	5,00	62,50	4,06
0,010	30	5,00	43,75	4,83
0,001	20	7,50	62,50	3,24
0,001	10	10,00	31,25	1,65
0,050	20	12,50	31,25	3,19
0,003	20	7,50	62,50	3,18

Tabela E-2: Resultados da rede LVQ usando os dados não normalizados e a taxa de aprendizagem decresce não linearmente em função do número de épocas.

Resultados da LVQ com os Dados Normalizados				
Taxa de Aprendizagem	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
0,10	23	62,50	50,00	12,80
0,10	22	57,50	50,00	11,75
0,10	24	55,00	43,75	12,96
0,10	10	52,50	43,75	5,27
0,10	15	57,50	43,75	8,02
0,20	10	50,00	50,00	5,21
0,20	30	60,00	56,25	15,81
0,20	40	77,50	62,50	20,76
0,20	45	80,00	62,50	23,78
0,10	50	72,50	62,50	25,98
0,20	60	62,50	50,00	31,64
0,10	50	72,50	62,50	28,56
0,30	20	27,50	31,25	10,98
0,30	40	27,50	31,25	21,81
0,05	20	65,00	43,75	11,26
0,05	40	62,50	62,50	21,91
0,05	50	65,00	62,50	27,24
0,10	60	70,00	56,25	32,35
0,10	50	72,50	62,50	26,47
0,10	50	72,50	62,50	27,57
0,20	40	77,50	62,50	23,34
0,20	34	72,50	56,25	19,17
0,60	20	15,00	31,25	11,53
0,50	20	27,50	31,25	10,66
0,60	60	27,50	31,25	32,35
0,20	100	15,00	31,25	53,82
0,20	19	65,00	56,25	10,38
0,10	19	62,50	50,00	10,54
0,05	19	57,50	31,25	10,71
0,20	100	15,00	31,25	52,40
0,20	45	80,00	62,50	24,45
0,30	20	27,50	31,25	11,48
0,15	50	77,50	62,50	27,74
0,15	25	62,50	56,25	13,62
0,15	30	70,00	62,50	16,53
0,18	30	62,50	56,25	16,04
0,17	40	47,50	37,50	21,86
0,20	200	20,00	31,25	43,06

Tabela E-3 : Resultados da rede LVQ usando os dados normalizados entre [-1,1] e a taxa de aprendizagem decresce linearmente em função do número de épocas.

Resultados da LVQ com os Dados Normalizados				
Taxa de Aprendizagem	Épocas	Treinamento (%)	Teste (%)	Tempo (s)
0,200	20	65,00	50,00	11,150
0,200	40	72,50	62,50	21,750
0,300	40	25,00	31,25	22,910
0,100	45	62,50	50,00	24,440
0,100	20	60,00	50,00	10,820
0,100	23	57,50	43,75	12,420
0,100	40	62,50	50,00	21,640
0,100	45	62,50	50,00	24,450
0,200	45	72,50	62,50	18,950
0,200	50	70,00	62,50	20,870
0,200	70	67,50	62,50	29,060
0,200	80	72,50	62,50	33,170
0,200	100	72,50	62,50	43,830
0,200	200	70,00	62,50	87,280
0,900	30	35,00	31,25	4,730
0,800	20	35,00	31,25	3,240
0,700	25	35,00	31,25	4,010
0,600	15	25,00	12,50	2,480
0,500	30	17,50	25,00	4,880
0,500	50	27,50	12,50	7,910
0,400	30	10,00	12,50	4,890
0,300	30	15,00	0,00	4,720
0,100	13	52,50	43,75	2,090
0,200	50	70,00	62,50	7,910
0,200	20	65,00	50,00	3,240
0,200	25	67,50	62,50	4,010
0,150	25	52,50	50,00	4,010
0,150	40	60,00	50,00	6,480

Tabela E-4 : Resultados da rede LVQ usando os dados normalizados entre [-1,1] e a taxa de aprendizagem decresce não linearmente em função do número de épocas.

APÊNDICE F

RESULTADOS DO FAN

FAN COM DADOS NORMALIZADOS NO INTERVALO [0,1]					
Função de Pertinência	Raio	Range	Épocas	Treinamento (%)	Teste (%)
Triangular	1	18	2	82.50	50.00
Triangular	1	18	3	87.50	50.00
Triangular	1	18	4	90.00	50.00
Triangular	1	18	5	87.50	50.0
Gaussiana	1	19	1	50.00	56.25
Gaussiana	1	19	2	80.00	62.50
Gaussiana	1	19	3	82.50	62.50
Gaussiana	1	19	4	85.00	68.75
Gaussiana	1	19	5	85.00	75.00
Gaussiana	1	19	6	82.50	75.00
Gaussiana	1	13	1	42.50	56.25
Gaussiana	1	13	2	75.00	56.25
Gaussiana	1	13	3	82.50	56.25
Triangular	1	21	1	47.50	50.00
Triangular	1	21	2	80.00	62.50
Triangular	1	21	3	90.00	68.75
Triangular	1	21	4	95.00	68.75
Triangular	1	21	5	92.50	75.00
Triangular	1	21	6	92.50	75.00
Triangular	2	20	1	55.00	75.00
Triangular	2	20	2	77.50	75.00
Triangular	2	20	3	82.50	68.75
Triangular	1	100	1	40.00	62.50
Triangular	1	100	2	97.00	31.25
Triangular	1	100	3	100.00	31.25
Triangular	1	30	1	42.50	56.25
Triangular	1	30	2	85.00	62.50
Triangular	1	30	3	90.00	62.50
Triangular	1	30	4	90.00	62.50

Tabela F-1: Resultados da FAN com H=1 (grau de combinação das variáveis).