

Renato José Sassi

**Uma Arquitetura Híbrida para Descoberta de
Conhecimento em Bases de Dados:
Teoria dos *Rough Sets* e Redes Neurais Artificiais
Mapas Auto-Organizáveis**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Engenharia
Elétrica.

São Paulo
2006

Renato José Sassi

**Uma Arquitetura Híbrida para Descoberta de
Conhecimento em Bases de Dados:
Teoria dos *Rough Sets* e Redes Neurais Artificiais
Mapas Auto-Organizáveis**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Engenharia
Elétrica.

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Livre Docente
Emilio Del Moral Hernandez

São Paulo
2006

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, de dezembro de 2006.

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Sassi, Renato José

Uma arquitetura híbrida para descoberta de conhecimento em bases de dados: teoria dos rough sets e redes neurais artificiais mapas auto-organizáveis / R.J. Sassi. -- São Paulo, 2006. 169 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1.Sistemas híbridos 2.Redes neurais 3.Descoberta de conhecimento I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

Agradecimentos

À minha esposa, Valéria, por todo apoio e incentivo que tem me dado, e aos meus pais, Renato e Diva, que sempre me apoiaram em todas as etapas de minha vida.

Ao Prof. Dr. Livre Docente Emilio Del Moral Hernandez, pelo seu exemplo de incentivo, de sugestões e orientações para a elaboração deste trabalho.

Ao amigo Leandro Augusto da Silva, que contribuiu muito para a realização deste trabalho, sempre disposto a discutir, colaborar e incentivar.

Resumo

Sassi,R.J. Uma Arquitetura Híbrida para Descoberta de Conhecimento em Bases de Dados: Teoria dos *Rough Sets* e Redes Neurais Artificiais Mapas Auto-Organizáveis. São Paulo, 2006, 169p.

As bases de dados do mundo real contêm grandes volumes de dados, e entre eles escondem-se diversas relações difíceis de descobrir através de métodos tradicionais como planilhas de cálculo e relatórios informativos operacionais.

Desta forma, os sistemas de descoberta de conhecimento (*Knowledge Discovery in Data Bases* - KDD) surgem como uma possível solução para dessas relações extrair conhecimento que possa ser aplicado na tomada de decisão em organizações.

Mesmo utilizando um KDD, tal atividade pode continuar sendo extremamente difícil devido à grande quantidade de dados que deve ser processada. Assim, nem todos os dados que compõem essas bases servem para um sistema descobrir conhecimento.

Em geral, costuma-se pré-processar os dados antes de serem apresentados ao KDD, buscando reduzir a sua quantidade e também selecionar os dados mais relevantes que serão utilizados pelo sistema.

Este trabalho propõe o desenvolvimento, aplicação e análise de uma Arquitetura Híbrida formada pela combinação da Teoria dos *Rough Sets* (Teoria dos Conjuntos Aproximados) com uma arquitetura de rede neural artificial denominada Mapas Auto-Organizáveis ou *Self-Organizing Maps* (SOM) para descoberta de conhecimento.

O objetivo é verificar o desempenho da Arquitetura Híbrida proposta na geração de *clusters* (agrupamentos) em bases de dados. Em particular, alguns dos experimentos significativos foram feitos para apoiar a tomada de decisão em organizações.

Palavras-chave: Sistemas Híbridos, Redes Neurais, Descoberta de Conhecimento.

Abstract

Sassi, R.J., An Hybrid Architecture for the Knowledge Discovery in Databases: Rough Sets Theory and artificial neural nets Self-Organizing Maps. São Paulo, 2006, 169p.

Databases of the real world contain a huge amount of data within which several relations are hidden. These relations are difficult to discover by means of traditional methods such as worksheets and operational informative reports.

Therefore, the knowledge discovery systems (KDD) appear as a possible solution to extract, from such relations, knowledge to be applied in decision taking.

Even using a KDD system, such activity may still continue to be extremely difficult due to the huge amount of data to be processed. Thus, not all data which are part of this base will be useful for a system to discover knowledge.

In general, data are usually previously processed before being presented to a knowledge discovery system in order to reduce their quantity and also to select the most relevant data to be used by the system.

This research presents the development, application and analysis of an hybrid architecture formed by the combination of the Rough Sets Theory with an artificial neural net architecture named Self-Organizing Maps (SOM) to discover knowledge.

The objective is to verify the performance of the hybrid architecture proposed in the generation of clusters in databases. In particular, some of the important performed experiments targeted the decision taking in organizations.

Keywords: Hybrids Systems, Neural Network, Knowledge Discovery Systems.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

1	Introdução	1
1.1	Redução de Dados.....	2
1.2	Teoria dos <i>Rough Sets</i> (Teoria dos Conjuntos Aproximados).....	3
1.3	Redes Neurais Artificiais (RNAs).....	5
1.4	Rede SOM (<i>Self-Organizing Maps</i> – Mapas Auto-Organizáveis).....	7
1.5	A Arquitetura Híbrida.....	8
1.6	Motivação.....	9
1.7	Objetivo e Contribuições.....	9
1.8	Critérios de Avaliação.....	10
1.8.1	Visualização do Mapa Gerado pela Rede SOM.....	11
1.9	Metodologia.....	12
1.10	Organização do Trabalho.....	12
2	A Organização na Era da Informação	14
2.1	A Necessidade da Informação.....	14
2.1.1	A Organização e o Meio Ambiente.....	16
2.2	A Tecnologia da Informação (TI) como Ferramenta Organizacional.....	17
2.3	Comparação entre Economia da Informação e Economia Industrial.....	19
2.4	Do Mercado de Massa ao Nicho de Informação.....	22
2.5	Da Economia da Informação para a Economia do Conhecimento.....	26
2.6	Resumo do Capítulo.....	26
3	Descoberta de Conhecimento em Bases de Dados (<i>Knowledge Discovery in Databases</i> - KDD)	28
3.1	Descoberta de Conhecimento em Bases de Dados (KDD).....	28

3.1.1	Fases do KDD.....	29
3.1.1.1	Seleção de Dados.....	31
3.1.1.2	Pré-processamento dos Dados.....	31
3.1.1.3	Transformação de Dados.....	33
3.1.1.4	<i>Data Mining</i> (Mineração de Dados).....	34
3.1.1.5	Interpretação/Avaliação do Conhecimento.....	36
3.2	Tarefas do KDD.....	37
3.2.1	Associação.....	37
3.2.2	Classificação.....	38
3.2.3	Clusterização (Agrupamento).....	39
3.2.4	Visualização.....	40
3.3	Áreas de Aplicação.....	40
3.4	Resumo do Capítulo.....	42
4	Teoria dos <i>Rough Sets</i> (Teoria dos Conjuntos Aproximados)	43
4.1	Introdução.....	43
4.2	Principais Conceitos.....	44
4.2.1	Espaços Aproximados.....	45
4.2.1.1	Indiscernibilidade.....	46
4.2.2	Aproximação de Conjuntos.....	48
4.2.3	Qualidade das Aproximações.....	53
4.3	Dependência de Atributos.....	55
4.4	Significância de Atributos.....	56
4.5	Redução de Atributos.....	57
4.6	Matriz de Discernibilidade.....	59
4.7	Função de Discernibilidade.....	60
4.8	Regras de Decisão.....	61
4.9	Resumo do Capítulo.....	62
5	Rede SOM (<i>Self-Organizing Maps</i> – Mapas Auto-Organizáveis)	63
5.1	Introdução.....	63
5.2	Um Breve Histórico sobre Redes Neurais Artificiais (RNAs).....	64

5.3	Uma Breve Comparação entre o Neurônio Biológico e o Neurônio Artificial.....	66
5.4	Aplicações das Redes Neurais Artificiais.....	67
5.4.1	Aplicação das Redes Neurais Artificiais nos Negócios.....	67
5.5	Tarefas Básicas das Redes Neurais Artificiais em KDD.....	69
5.5.1	Classificação.....	69
5.5.2	Clusterização (Agrupamento).....	69
5.5.3	Associação.....	70
5.6	Treinamento (Aprendizado) em uma Rede Neural Artificial (RNA).....	71
5.6.1	Aprendizado Supervisionado.....	72
5.6.2	Aprendizado Não-supervisionado.....	73
5.7	Características da Rede SOM.....	74
5.7.1	Estrutura da Rede SOM.....	75
5.7.2	O Algoritmo de Aprendizado da Rede SOM.....	77
5.7.2.1	Fases da Rede SOM.....	77
5.8	Medidas de Qualidade.....	85
5.9	Visualização do Mapa Gerado pela Rede SOM.....	87
5.9.1	A Matriz-U.....	87
5.9.2	Mapa por Similaridade de Cor.....	88
5.10	Resumo do Capítulo.....	90
6.	Experimentos Comparativos com a Arquitetura Híbrida (RS com rede SOM)	91
6.1	Metodologia e Bases de Dados utilizadas nos Experimentos.....	91
6.1.1	Ferramentas utilizadas e Plataformas de Ensaio.....	94
6.1.2	Parâmetros da Rede SOM adotados nos Experimentos.....	95
6.2	Objetivo dos Experimentos Contrastivos.....	96
6.3	Realização dos Experimentos.....	97
6.3.1	Experimento 1: Realizado Com uma Base de Dados Artificial.....	98
6.3.2	Experimento 2: Com a Base de Dados Íris (Base Pública).....	103

6.3.3	Experimento 3: Com a Base de Dados Pública ZOO.....	108
6.3.4	Experimento 4: Com a Base de Dados Consumidor.....	114
6.3.4.1	Descoberta de Conhecimento no Experimento 4.....	118
6.3.5	Experimento 5: Com a Base de Dados <i>The Insurance Company</i>	122
6.4	Análise Final dos Resultados.....	126
7.	Conclusão	129
7.1	Sumário das Contribuições.....	130
7.2	Trabalhos Futuros.....	133
	Referências Bibliográficas	135
	Bibliografia Consultada	158
	Anexo A - Roteiro de Execução dos Experimentos Comparativos com a Arquitetura Híbrida	166

Lista de Figuras

2.1 A Organização formando a sua Base de Conhecimento.....	15
2.2 A Organização se relacionando com o Meio Ambiente [LAU99].....	17
2.3 Breve Cronologia da TI [TUR03].....	22
3.1 Fases do processo de KDD. Adaptado de Fayyad [FAY96a].....	30
3.2 Esforço requerido para cada fase do processo de KDD.....	30
3.3 Formas de pré-processar dados.....	33
3.4 Esquema simplificado de <i>Data Mining</i>	35
4.1 Espaço Aproximado $A = (U,R)$ e $X \subseteq U$	45
4.2 Aproximação Inferior de X	50
4.3 Aproximação Superior de X	51
4.4 Todas as regiões de X em A	52
5.1 Comparação entre o neurônio biológico (letra a) e o neurônio artificial (letra b).....	66
5.2 Arquitetura de uma Rede Neural Artificial <i>MultiLayer Perceptron</i> (MLP).....	71
5.3 Arranjo hexagonal para uma rede SOM com grade bidimensional. Adaptado de Kohonen [KOH97a].....	76
5.4 Rede SOM com Aprendizado Competitivo [KOH97a].....	77
5.5 Processo de escolha do neurônio vencedor	79
5.6 Alteração do vetor de pesos do neurônio vencedor	80
5.7 Convergência dos pesos da rede SOM para os centros de massa dos agrupamentos de padrões de entrada.....	81
5.8 Matriz-U (matriz de distâncias unificadas).....	87
5.9 Mapa por Similaridade de Cor evidenciando a localização de agrupamentos..	88
5.10 Mapa por Similaridade de Cor onde cada neurônio foi rotulado com todos os tipos de dados para os quais foi vencedor (BMU).....	89
6.1 Primeira fase dos Experimentos em que as bases de dados são apresentadas à rede SOM.....	92
6.2 Segunda fase dos Experimentos em que as bases de dados são apresentadas à	93

Arquitetura Híbrida (rede SOM com redutos).....	
6.3 Ilustrações do Experimento 1 apresentando a Distribuição Inicial dos dados e as Distribuições sem uma das Classes.....	99
6.4 Ilustrações do Experimento 1 apresentando a Distribuição das Classes com apenas um atributo cada.....	100
6.5 Visualização dos Mapas gerados no Experimento 1 pela rede SOM sem redutos (Distribuição Inicial) e pela rede SOM com o reduto X2 e com o reduto X3 (Arquitetura Híbrida).....	102
6.6 Mapa gerado pela rede SOM sem redutos após a apresentação da base de dados Iris (primeira fase do experimento 2).....	104
6.7 Os quatro Mapas gerados no Experimento 2 pela rede SOM com redutos (Arquitetura Híbrida).....	107
6.8 Mapa gerado pela primeira fase do Experimento 3 onde os <i>clusters</i> podem ser identificados através de cores e as classes identificadas através das amostras rotuladas.....	110
6.9 Mapa gerado pela segunda fase do Experimento 3, onde os <i>clusters</i> podem ser identificados através de cores e as classes identificadas através das amostras rotuladas.....	112
6.10 Visualização dos 14 <i>clusters</i> gerados pela rede SOM sem redutos rotulados de acordo com as classes da base de dados Consumidor (primeira fase do Experimento 4).....	116
6.11 Visualização dos 9 <i>clusters</i> gerados pela rede SOM com redutos (Arquitetura Híbrida) rotulados de acordo com as classe da base de dados Consumidor (segunda fase do Experimento 4).....	117
6.12 Visualização dos neurônios localizados no mapa gerado pela Arquitetura Híbrida (Experimento 4).....	119
6.13 Visualização dos <i>clusters</i> gerados pela rede SOM sem redutos (primeira fase do Experimento 5).....	123
6.14 Visualização dos <i>clusters</i> gerados pela rede SOM com redutos (segunda	

fase do Experimento 5).....	124
A.1 Ilustração da tela do Rosetta com a base de dados carregada.....	165
A.2 Ilustração da tela do Rosetta na escolha do algoritmo e dos parâmetros de redução.....	166
A.3 Ilustração da tela do Rosetta com os redutos gerados pelo <i>Rough Sets</i> utilizando a base de dados Iris (Experimento 2).....	167

Lista de Tabelas

4.1	Sistema de Informação S.....	46
4.2	Sistema de Decisão (Sistema de Informação com o atributo de decisão Retorno).....	47
4.3	Sistema de Decisão com os elementos e2 e e3 indiscerníveis.....	48
4.4	Sistema de Informação S sem o atributo Boa Localização.....	58
4.5	Sistema de Informação S sem o atributo Qualidade do Produto.....	59
4.6	Matriz de Discernibilidade.....	60
5.1	Aplicações de RNAs na Indústria.....	68
5.2	Parâmetros de Aprendizado (treinamento) para arquiteturas de rede neural artificial.....	74
6.1	Redutos gerados pelo RS no Experimento 1 utilizando uma base de dados artificial.....	100
6.2	Os 4 redutos gerados pelo RS utilizando a base de dados Íris (Experimento 2).....	105
6.3	Características da base de dados ZOO (Experimento 3).....	108
6.4	Resultados da rede SOM sem redutos (primeira fase do Experimento3).....	109
6.5	Resultados da rede SOM com redutos (Arquitetura Híbrida) referentes ao Experimento 3.....	111
6.6	Comparação entre a rede SOM e a Arquitetura Híbrida proposta	112

(Experimento 3).....	
6.7 Números do Experimento 4 comparando a rede SOM sem redutos com a rede SOM com redutos (Arquitetura Híbrida).....	118
6.8 Informações dos atributos (reduto) consideradas pela rede SOM para agrupar os neurônios.....	120
6.9 Números Comparativos da rede SOM sem redutos com a rede SOM com redutos (Arquitetura Híbrida) referentes ao Experimento 5.....	125

Lista de Abreviaturas

AI	Aproximação Inferior
AS	Aproximação Superior
BMU	(<i>Best Match Unit</i>) Neurônio Vencedor
C	Atributos Condicionais
CAM	(<i>Computer Aided Manufacturing</i>) Manufatura Assistida por Computador
FC	Fator de Credibilidade
D	Atributo de Decisão
DW	(<i>Data Warehouse</i>) Armazém de Dados
EQ	Erro de Quantização
ET	Erro Topográfico
ERP	(<i>Enterprise Resource Planning</i>) Sistemas de Gestão Integrada
F_s	Função de Discernibilidade
GB	(<i>GigaByte</i>) Bilhão de <i>Bytes</i>
IA	Inteligência Artificial
IND	Relação de Indiscernibilidade
KDD	(<i>Knowledge Discovery in Databases</i>) Descoberta de Conhecimento em Bases de Dados
MD	Matriz de Discernibilidade
MLP	(<i>MultiLayer Perceptron</i>) <i>Perceptrons</i> de Múltiplas Camadas
RAM	(<i>Random Access Memory</i>) Memória de Acesso Randômico
RED	Reduto
RF	Região de Fronteira
RN	Região Negativa
RNA	Rede Neural Artificial
RNAs	Redes Neurais Artificiais
RS	<i>Rough Sets</i>

S Sistema de Informação
SD Sistema de Decisão
SOM (*Self-Organizing Maps*) Mapas Auto-Organizáveis
TCA Teoria dos Conjuntos Aproximados
TB (*Terabyte*) Trilhão de *Bytes*
TI Tecnologia da Informação

1 Introdução

Os avanços na área da Tecnologia da Informação têm possibilitado o armazenamento de grandes e múltiplas bases de dados. Esses dados produzidos e armazenados em larga escala são difíceis de serem analisados, interpretados e relacionados pelos métodos tradicionais, como planilhas de cálculo e relatórios informativos operacionais [PIA91]. Nesses grandes volumes de dados escondem-se diversas relações interessantes que, ao serem descobertas, acrescem conhecimento à tomada de decisão.

Por essa razão faz-se necessário o uso de sistemas que possam extrair o conhecimento dessas bases, viabilizando a análise dos dados, denominados de KDD (*Knowledge Discovery in Databases*) ou Descoberta de Conhecimento em Bases de Dados.

O KDD pode ser definido como um processo de extração de conhecimentos válidos, novos, potencialmente úteis e compreensíveis para apoiar a tomada de decisão [FAY96a]. Para tanto, o KDD se utiliza das seguintes áreas para realizar os seus processos: métodos estatísticos, reconhecimento de padrões, visualização, banco de dados, aprendizado de máquina, Inteligência Artificial, *Data Warehouse*, dentre outras.

O KDD é um processo constituído de fases que possuem inúmeros passos, os quais envolvem um número elevado de decisões a serem tomadas pelo usuário, ou seja, é um processo iterativo. É também um processo iterativo, pois ao longo do processo de KDD, um passo será repetido tantas vezes quantas se fizerem necessárias para que se chegue a um resultado satisfatório [FAY96a].

Entretanto, nem todos os dados que compõem as bases servem para um sistema descobrir conhecimento. Assim, é necessário pré-processar os dados antes de seguir com o processo de descoberta de conhecimento.

A fase de pré-processamento dos dados é importante, pois o sucesso dos resultados obtidos no processo de KDD depende diretamente da qualidade dos dados de entrada.

1.1 Redução de Dados

A redução de dados é uma forma de pré-processamento que visa obter uma representação reduzida dos dados, mas que produz os mesmos (ou quase os mesmos) resultados analíticos.

Boas razões para a redução de dados são a ultrapassagem da capacidade de processamento dos programas de aprendizagem e o tempo muito longo para obter uma solução. Pode-se citar como vantagens da redução de dados a redução do tempo de aprendizagem dos algoritmos e a interpretação mais fácil dos conceitos aprendidos [BAT03].

A redução de dados pode ser realizada por: agregação via cubo, redução de atributos (dimensão), compressão de dados, redução de casos, discretização e construção de hierarquias [LEE01].

Neste trabalho a redução de atributos será mais explorada no capítulo 4, em função de a Arquitetura Híbrida proposta utilizar *Rough Sets* para este fim.

As técnicas de redução de atributos procuram selecionar o menor conjunto de atributos para dividir o espaço de instâncias de tal maneira que a distribuição das classes no novo espaço é tão próxima quanto possível daquela do espaço original. Considera-se que a permanência de atributos irrelevantes pode levar à descoberta de padrões de baixa qualidade pelo algoritmo responsável pela descoberta de conhecimento (*Data Mining*).

Podem-se dividir em três grupos as abordagens propostas para selecionar um subconjunto de atributos [KOH97b]:

- Embutida (*Embedded*): a seleção de atributos é realizada internamente pelo próprio algoritmo responsável pela descoberta de conhecimento (*Data Mining*) como parte do processo de criação do modelo.

- Filtro: a seleção de atributos é realizada em um processo anterior separado da aplicação do algoritmo de descoberta de conhecimento. Neste trabalho foi utilizada a abordagem Filtro.
- *Wrappers*: consiste em selecionar um subconjunto de atributos e medir a precisão do classificador sobre esse conjunto de atributos. É realizada uma busca pelo subconjunto que gera o classificador com menor erro. Essa busca avalia cada subconjunto candidato, até que o critério de parada, relacionado com a precisão do classificador, seja satisfeito.

Nas próximas seções deste capítulo serão abordadas as técnicas que compõem a Arquitetura Híbrida proposta.

1.2 Teoria dos *Rough Sets* (Teoria dos Conjuntos Aproximados)

A Teoria dos *Rough Sets* (RS), ou Teoria dos Conjuntos Aproximados, (TCA) foi proposta por Zdzislaw Pawlak [PAW82] em 1982 como um novo modelo matemático para representação do conhecimento e tratamento de incerteza, tendo sido usada, posteriormente, para o desenvolvimento de técnicas para classificação aproximada em aprendizado de máquina. Devido a estas características, tem-se utilizado esta teoria em Inteligência Artificial, especialmente nas áreas de aquisição de conhecimento; raciocínio indutivo e descoberta de conhecimento em base de dados.

Conjuntos aproximados podem ser considerados conjuntos com fronteiras nebulosas, ou seja, conjuntos que não podem ser caracterizados precisamente, utilizando-se dos atributos disponíveis [PAW91].

Entre as abordagens mais tradicionais existentes para a modelagem e tratamento de incertezas, encontram-se [UCH98]: Regra de Bayes; Fator de Certeza; Teoria de Dempster-Shafer; Teoria dos Conjuntos *Fuzzy* (*Fuzzy Sets*); Raciocínio *Default* e Teoria de *Endorsements*.

A incerteza pode se manifestar de diversas formas, como: imprecisão, incompletude, inconsistência, etc. RS trata de um tipo fundamental de incerteza, a indiscernibilidade. A indiscernibilidade surge quando não é possível distinguir elementos de um mesmo conjunto e representa a situação em que esses elementos parecem todos ser um único elemento [UCH98].

Uma das principais vantagens da Teoria dos RS é que ela não necessita de informações preliminares ou adicionais sobre os dados, tais como a distribuição de probabilidade em estatística, atribuição de probabilidades básicas na Teoria de Dempster-Shafer, ou mesmo os graus de pertinência na Teoria dos Conjuntos *Fuzzy* [UCH98]. Isto pode ser verificado em Ziarko [ZIA94b].

O conceito dos RS relaciona-se, de alguma maneira, com outras teorias matemáticas desenvolvidas para manipular incerteza, particularmente com a Teoria da Evidência de Dempster-Shafer. A principal diferença é que a teoria proposta por Dempster-Shafer utiliza a função de crença como ferramenta principal, enquanto a Teoria dos RS faz uso de Aproximações Inferior e Superior. Existe também uma relação entre a Teoria dos RS e a Teoria dos *Fuzzy Sets*, as quais são frequentemente comparadas e até mesmo confundidas, pois tratam da incerteza. Um estudo comparativo mais aprofundado entre RS e *Fuzzy Sets* pode ser encontrado em Yao [YAO98].

Szladow [SLA93], utiliza um exemplo da área de processamento de imagens para explicar a diferença entre RS e *Fuzzy Sets*. Enquanto *Fuzzy Sets* trata da existência de mais de um nível de cinza nos pixels, RS trata do tamanho desses pixels. Assim, *Fuzzy Sets* trata da relação entre intensidades de elementos dentro da mesma classe, enquanto RS trata da relação entre grupos de elementos em diferentes classes. Entretanto, a Teoria dos RS não compete com a Teoria dos *Fuzzy Sets*, mas a complementa. Adjei [ADJ01] apresenta um trabalho que combina as duas teorias. Na realidade, a Teoria dos RS e a Teoria dos *Fuzzy Sets* são duas abordagens independentes para o tratamento de conhecimento impreciso.

Os conceitos dos RS têm se mostrado muito úteis quando aplicados a problemas dos seguintes tipos: redução de atributos, descoberta de dependência entre atributos e na descoberta de padrões entre os dados [PAW96a].

A redução de atributos realizada por RS é feita através dos chamados redutos, que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais.

Devido às características comentadas, a crescente utilização de RS pode ser comprovada pelo número de aplicações e publicações científicas nas seguintes áreas: KDD e *Data Mining* [ADJ01; JIA03; LIN97,02; TSU98,00], medicina [KOM99a; TSU98,00,01a,01b], negócios [CAB04; EBE04; GOL95; IJE00; JIA03; NAR04], engenharia [BON02; LAM02; LON02; ZIA93; ZON02]; reconhecimento de padrões [SAS06, VEL01a; XIA97; ZHA04], mineração de textos [YU05], dentre outros.

1.3 Redes Neurais Artificiais (RNAs)

A Arquitetura Híbrida proposta neste trabalho utiliza além de RS as redes neurais artificiais (RNAs). As redes neurais artificiais são redes inspiradas na estrutura do cérebro, com o objetivo de apresentar características similares ao comportamento humano, tais como: aprendizado, associação, generalização e abstração.

Uma rede neural artificial é um processador maciçamente paralelo, distribuído, constituído de unidades de processamento simples, que têm capacidade para armazenar conhecimento experimental e torná-lo disponível para uso [HAY99].

Vários trabalhos científicos descrevem as características vantajosas em utilizar as RNAs. Pinheiro [PIN99b] relata a alta precisão dos modelos de predição, os quais podem ser aplicados a uma extensa gama de problemas. Baets [BAE94] relata que são robustas e tolerantes a falhas, graças ao seu paralelismo. Segundo Haykin [HAY99], as RNAs podem generalizar os resultados obtidos para dados previamente desconhecidos, ou seja, produzir respostas coerentes e apropriadas para padrões ou exemplos que não foram utilizados no seu treinamento. Zhang [ZHA94] mostra a capacidade das RNAs em lidar com estruturas de dados não lineares, o que é uma característica interessante para aplicações em *marketing* que lidam com dados como vendas e preços.

Muitos trabalhos mostram o melhor desempenho das RNAs em comparação com outras técnicas. Todesco [TOD95] ressalta que as RNAs podem ter várias entradas e várias saídas, sendo aplicáveis a sistemas com muitas variáveis; a RNA é não-paramétrica e faz suposições menos restritivas a respeito da distribuição dos dados de entrada do que métodos estatísticos tradicionais.

Os trabalhos de Almeida [ALM93, 96, 98], Arraes [ARR01], Freitas [FRE99], [PIN99a] e Portugal [POR95], voltados para a área de negócios, retratam o melhor desempenho das RNAs com relação aos métodos estatísticos. As RNAs separam grupos de maneira não linear, levando a uma separação mais refinada, o que é uma vantagem com relação às técnicas estatísticas discriminatórias e classificatórias lineares. As técnicas estatísticas requerem inferências iniciais que nem sempre são fáceis de aplicar e interpretar e são difíceis de ser aplicadas a um número muito grande de variáveis; as redes neurais artificiais automatizam os processos estatísticos, aliviando o usuário final desse grande trabalho. Em Ebecken [EBE04], os resultados também foram melhores com RNAs quando aplicadas na área de energia, e em Zaremba [ZAR99], na área médica. Pode-se verificar aplicações de redes neurais artificiais nas organizações no capítulo 5, Tabela 5.1.

Ainda em Almeida [ALM95] é feita uma comparação entre RNAs e Sistemas Especialistas. As redes neurais artificiais não precisam de um especialista para a criação de sua base de conhecimentos. Elas não trabalham com regras. Sua aquisição de conhecimento é feita automaticamente a partir de exemplos coletados em bases de dados. Arraes [ARR01] e Rodrigues [ROD01] citam em seus trabalhos vantagens das RNAs sobre as Árvores de Decisão. Em Tápia [TAP00], os experimentos mostram melhores resultados obtidos com as RNAs para previsão do que com a metodologia Box-Jenkins.

Como toda técnica, as RNAs apresentam algumas desvantagens, como: utilizar predominantemente dados numéricos, em certos casos o tempo de treinamento é longo, requerer muitos parâmetros que influenciarão fortemente os resultados e a obtenção do melhor resultado depende do ajuste dos parâmetros baseados na tentativa e erro. Berry [BER97] discute as dificuldades relacionadas ao uso das RNAs e Passari [PAS03] analisa vantagens e desvantagens do uso das RNAs.

1.4 Rede SOM (*Self-Organizing Maps* – Mapas Auto-Organizáveis)

A rede SOM (*Self-Organizing Maps*), ou Mapas Auto-Organizáveis, foi desenvolvida pelo finlandês Teuvo Kohonen em 1984.

A rede SOM é um tipo de rede neural artificial baseada em aprendizado não supervisionado, sendo capaz de mapear um conjunto de dados, de um espaço de entrada multidimensional, em um conjunto finito de neurônios organizados em um arranjo normalmente unidimensional ou bidimensional.

Além das características vantajosas das redes neurais artificiais citadas na seção 1.3, a rede SOM possui algumas outras vantagens, como ajuste local do mapa e visualização [DUT01], o que faz com que seja utilizada nas mais diversas áreas, que vão da medicina [LO97] ao marketing [RUS97].

No caso de KDD, [HSU99; JIN04, KAS97, VES97, VES00a VES00b], a rede SOM é muito utilizada na tarefa de clusterização, por possibilitar em um mapa uni ou bidimensional a formação e a visualização simples dos *clusters* (grupos) e da correlação dos dados, preservando a posição relativa dos *clusters* no hiperespaço original.

Ebecken [EBE04] mostra alguns trabalhos onde a avaliação dos resultados foi realizada através da visualização do mapa gerado.

Além das vantagens sobre as técnicas estatísticas tradicionais, um estudo comparativo e detalhado entre técnicas de clusterização em mineração de dados realizado por Ultsh [ULT95] e Vesanto [VES00a] mostrou que a rede SOM apresenta melhores resultados.

Uma das desvantagens da rede SOM é a imprecisão na definição de fronteira entre os *clusters* [DUT01]. Esta desvantagem será analisada com mais detalhes no capítulo 6 (Experimentos).

1.5 A Arquitetura Híbrida

Segundo Goldschmidt [GOL05], técnicas podem ser combinadas para gerar as chamadas Arquiteturas Híbridas. A grande vantagem desse tipo de sistema deve-se ao sinergismo obtido pela combinação de duas ou mais técnicas. Este sinergismo reflete na obtenção de um sistema mais poderoso (em termos de interpretação, de aprendizado, de estimativa de parâmetros, de generalização, dentre outros) e com menos deficiências.

Existem três formas básicas de se associarem duas técnicas para a construção de uma Arquitetura Híbrida [SOU99]:

Híbrida Sequencial: nesta forma, uma técnica atua como entrada de outra técnica. A Arquitetura Híbrida proposta neste trabalho é sequencial, pois usa o RS como pré-processador da rede SOM.

Híbrida Auxiliar: esta forma poderia ser exemplificada do seguinte modo: uma RNA invoca um Algoritmo Genético para a otimização de seus pesos ou de sua estrutura. Neste caso, tem-se um maior grau de hibridização em comparação com o híbrido sequencial.

Híbrida Incorporada: nesta forma praticamente não há separação entre as duas técnicas. Pode-se dizer que a primeira técnica possui a segunda técnica e vice-versa. Poderia ser exemplificado por um sistema *neuro-fuzzy* híbrido em que um sistema de inferência *fuzzy* é implementado segundo a estrutura de uma rede neural artificial. Aqui a hibridização é a maior possível.

Existem vários trabalhos que combinam RS com outras técnicas formando uma Arquitetura Híbrida [BAN97, JAG98; LIN96b, SAN03; ZAR99; ZHA01] Com a rede SOM, pode-se destacar o trabalho de Lingras [LIN96b], no qual o algoritmo da rede SOM foi modificado com base nos conceitos de Aproximação Inferior e Superior dos RS buscando um melhor intervalo entre os clusters. Neste trabalho a Arquitetura Híbrida proposta emprega RS pré-processando a rede SOM.

1.6 Motivação

O presente trabalho tem como seu pilar fundamental desenvolver, aplicar, analisar e avaliar o desempenho de uma Arquitetura Híbrida proposta (RS com a rede SOM) na descoberta de conhecimento em bases de dados.

1.7 Objetivo e Contribuições

Por possuírem grandes quantidades de informação, as organizações cada vez mais necessitam do auxílio de técnicas computacionais apropriadas para auxiliar o homem na análise, na interpretação e no relacionamento dessas informações em busca de conhecimento.

Em geral, as técnicas apresentam ao homem vantagens na sua utilização, mas também apresentam desvantagens, levando à conclusão de que em muitos casos existe a necessidade de combinar duas ou mais técnicas a fim de eliminar ou reduzir as deficiências.

Surgiu, assim, neste trabalho a idéia de juntar RS com a rede SOM para formar uma Arquitetura Híbrida. RS, pelas vantagens apresentadas na seção 1.2 e também por ser uma teoria relativamente nova, cujo campo de pesquisa ainda está aberto à exploração e necessita de contribuição; no caso da rede SOM, pelas vantagens descritas na seção 1.4. A combinação RS com a rede SOM pela potencial melhoria sobre a baixa definição de fronteira dos *clusters* gerados pela rede SOM.

A Arquitetura Híbrida proposta será avaliada pelo seu desempenho na definição de fronteira entre os *clusters*.

Na Arquitetura Híbrida desenvolvida neste trabalho, as duas técnicas têm as seguintes funções: a função de RS é de redução de atributos (dimensão) e a função da rede SOM é a de gerar os *clusters*.

A base de dados, com todos os seus atributos iniciais, é apresentada ao RS a fim de que ele reduza o número de atributos. Em seguida, a base de dados reduzida é apresentada à rede SOM para a geração do mapa auto-organizável.

A análise dos resultados não busca apenas verificar a redução do tempo de treinamento da rede SOM, mas analisar principalmente se a deficiência na definição de borda (fronteira) dos *clusters* obtidos foi reduzida ou até eliminada.

O objetivo principal deste trabalho é, através da análise dos resultados obtidos com a Arquitetura Híbrida proposta, procurar responder a uma pergunta fundamental: “Existe melhoria nos resultados gerados pela rede SOM quando pré-processada por RS?”.

Além do objetivo principal, vislumbra-se um outro objetivo, que é contribuir para o entendimento mais aprofundado dos RS como uma técnica de redução de atributos, analisando para tanto os conceitos de dependência e significância de atributos.

Este trabalho busca contribuir com uma Arquitetura Híbrida que combina RS com rede SOM na descoberta de conhecimento em bases de dados, eliminando ou reduzindo deficiências da rede SOM para melhorar a tomada de decisão.

1.8 Critérios de Avaliação

Ainda não está claro como se deve escolher os parâmetros do algoritmo da rede SOM de maneira a se obter um “bom mapeamento” [ZUC03]. Normalmente a rede SOM é avaliada com base na resolução do mapa e em relação à preservação da topologia dos dados de entrada [KIV95]. A escolha do “melhor mapeamento” deveria ser, portanto, por aquele que “melhor representa os dados de entrada”. Este critério normalmente é traduzido por duas medidas: o erro de quantização (EQ) e o erro topográfico (ET).

Em Vesanto [VES03], Villmann [VIL94] e Pözlbauer [PÖL04] pode-se encontrar experimentos utilizando outras medidas de qualidade, mas o erro de quantização (EQ) e o erro topográfico (ET) ainda são as medidas mais utilizadas.

Além das medidas de qualidade, serão adotados também como critérios de avaliação o tempo de treinamento da rede SOM (esforço computacional requerido), o número de *clusters* gerados e a visualização do mapa através do Mapa por Similaridade de Cor. Como os *clusters* são obtidos por intermédio da aplicação dos conceitos de similaridade e distância [JOH98], um número maior ou menor de *clusters* pode indicar o que ocorreu com as similaridades entre os dados e as diferenças de separação dos *clusters* [PAK04]. Comark [COM71] afirma que a tarefa de clusterização está pautada em duas idéias básicas, a coesão interna dos elementos e o isolamento externo entre os *clusters*. Dessa forma, um bom *cluster* é aquele que apresenta alta similaridade interna (*intra-cluster*) e baixa similaridade externa (*extra-cluster*). Em Halkidi [HAL01] pode-se encontrar um estudo detalhado de validação das técnicas de clusterização.

1.8.1 Visualização do Mapa Gerado pela Rede SOM

Um dos motivos que fazem com que a rede SOM seja muito utilizada para o KDD é a visualização.

As formas de visualização mais utilizadas para detectar *clusters* gerados pela rede SOM [VES02] são: a chamada Matriz-U ou matriz de distâncias unificadas [ULT89] e o Mapa por Similaridade de Cor, que permite a análise da estrutura dos *clusters*.

Vários trabalhos se utilizaram da visualização dos mapas gerados pela rede SOM para chegar à conclusão nos seus experimentos [EBE04, JIN04, RUS97, VES99a].

Os resultados dos mapas obtidos nos experimentos realizados no sexto capítulo desta tese serão analisados através do Mapa por Similaridade de Cor.

1.9 Metodologia

A realização da pesquisa deste trabalho está embasada em consultas a fontes bibliográficas e de referencial teórico como: artigos, livros, teses, dissertações e relatórios técnicos nos campos de redes neurais artificiais, RS, KDD e de negócios. Foi desenvolvida uma parte experimental (capítulo 6) para avaliar os resultados obtidos com a aplicação da Arquitetura Híbrida proposta.

Foram realizados cinco experimentos com bases de dados distintas comparando a Arquitetura Híbrida proposta (RS com a rede SOM) com uma rede SOM. Cada experimento foi dividido em duas fases:

Na primeira fase, denominada rede SOM sem redutos, apresentou-se à rede SOM cada uma das cinco bases de dados dos experimentos com todos os atributos e avaliaram-se os resultados.

Na segunda fase, denominada Arquitetura Híbrida (rede SOM com redutos), apresentou-se primeiro RS a cada uma das cinco bases de dados para a geração dos redutos e, em seguida, as bases de dados reduzidas foram apresentadas à rede SOM.

Os detalhes dos experimentos estão relatados no capítulo 6 deste trabalho.

1.10 Organização do Trabalho

Este trabalho foi estruturado em sete capítulos. Além desta Introdução (capítulo 1), o trabalho compõe-se das seguintes partes:

Capítulo 2 - A Organização na Era do Conhecimento: o capítulo trata das constantes mudanças organizacionais decorrentes da pressão exercida pelo meio ambiente. Para mudar corretamente é necessário conhecer este meio ambiente. É neste cenário que entram os sistemas de descoberta de conhecimento como elemento importante no processo de transformação da informação em conhecimento. A contribuição do capítulo para este trabalho reside em mostrar a necessidade das organizações em extrair conhecimento de suas bases de dados, justificando a utilização dos sistemas de descoberta de conhecimento.

Capítulo 3 - Descoberta de Conhecimento em Bases de Dados (KDD): este capítulo descreve todo o processo dos sistemas de descoberta de conhecimento, os chamados KDD, constituído de fases, tarefas e técnicas. A contribuição do capítulo está em apresentar as características dos sistemas de descoberta de conhecimento, servindo de apoio para entender o objetivo da Arquitetura Híbrida proposta.

Capítulo 4 - Teoria dos *Rough Sets*: este capítulo descreve os conceitos gerais e aplicações da Teoria dos RS, também chamada de Teoria dos Conjuntos Aproximados. O capítulo contribui principalmente com o entendimento da Teoria dos RS nos conceitos referentes à incerteza e à redução de atributos que são relevantes para este trabalho, pois a Arquitetura Híbrida proposta combina RS com a rede SOM.

Capítulo 5 - Rede SOM (*Self-Organizing Maps* – Mapas Auto-Organizáveis): neste capítulo descreve-se os conceitos gerais das redes neurais artificiais e suas aplicações. É realizado um estudo mais aprofundado da Rede SOM, pois é a rede neural artificial utilizada na Arquitetura Híbrida proposta. O capítulo contribui para o entendimento de como a rede SOM realiza a tarefa de clusterização e também discute quais são as medidas de qualidade e as formas de visualização do mapa que serão utilizadas na avaliação do desempenho da Arquitetura Híbrida.

Capítulo 6 – Experimentos Comparativos com a Arquitetura Híbrida: neste capítulo são realizados os experimentos para avaliar o desempenho da Arquitetura Híbrida proposta. Ao final de cada experimento é realizada uma avaliação dos resultados obtidos, e no final do capítulo analisam-se os resultados através de um balanço geral dos experimentos.

Capítulo 7 - Conclusão: neste capítulo conclui-se o trabalho com a discussão dos resultados obtidos e propõem-se direções para futuras investigações.

2 A Organização na Era da Informação

Este capítulo descreve a importância da informação na relação das organizações com o meio ambiente e a necessidade de transformação dessas informações em conhecimento para apoiar a tomada de decisão.

2.1 A Necessidade da Informação

Atualmente as organizações precisam reagir de modo rápido às ameaças e às oportunidades que surgem no ambiente empresarial cada vez mais competitivo e mutável [ROB00]. Este ambiente está intimamente relacionado à combinação de fatores sociais, jurídicos, econômicos, físicos e políticos, que afetam as atividades da organização.

Zuboff [ZUB88] vê a informação como um elemento tão importante para organizações, que convencionou, chamar a era atual de Era da Informação, e a economia de Economia da Informação.

Na Era da Informação a riqueza nasce de idéias inovadoras e do uso inteligente da informação. As organizações da Era da Informação concorrem em um mercado repleto de desafios, que muda rapidamente, é complexo, globalizado, supercompetitivo e voltado para o cliente.

A boa informação deve, no mínimo, ser: correta, destinada à pessoa certa e em tempo hábil. Deve também ser: precisa, completa, econômica, flexível, confiável, relevante, simples e verificável.

O conhecimento consiste em informações organizadas e processadas para transmitir discernimento, experiências, aprendizagem acumulada ou habilidade, ser aplicável a um problema ou processo empresarial [MAT82]. As informações que são processadas para extrair implicações críticas e refletir experiências e habilidades

anteriores fornecem a quem as recebe o conhecimento organizacional, que é altamente valorizado.

Uma organização forma a sua base de conhecimento através da reunião das experiências individuais, parciais, dispersas e voláteis, ou seja, a base de conhecimento é a reunião de informações transformadas em conhecimento coletivo, coerente e memorizado [TER00].

A Figura 2.1 mostra a Organização formando sua Base de Conhecimento.

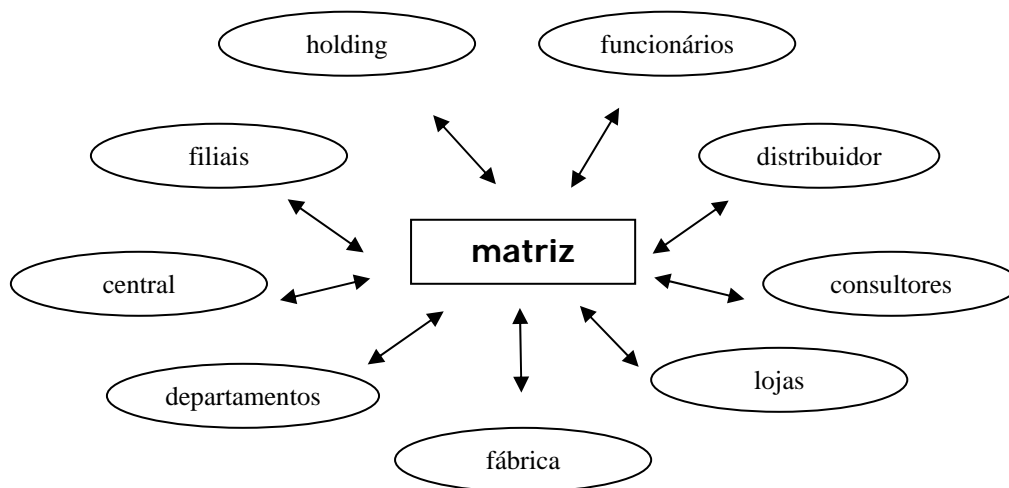


Figura 2.1: A Organização formando a sua Base de Conhecimento.

Na Figura 2.1 pode-se visualizar o processo de troca de informações entre a organização (Matriz) e suas várias partes, possibilitando a geração e armazenando do conhecimento.

As informações e o conhecimento compõem recursos estratégicos essenciais para o sucesso de uma organização. Portanto, o correto tratamento das informações pode ser uma vantagem competitiva muito importante.

A organização que trata corretamente as informações apresenta as seguintes características [LAU99]:

- Atuante: competitividade comprovada pela rentabilidade em curto prazo e sobrevida no longo prazo;

- Identidade definida: sabe quem é, o que deseja ser e têm conhecimento sobre como agir para tal;
- Aberta em seu interior (comunicante e participativa) e em seu exterior (proativa e antecipativa);
- Adaptável às mudanças de ambiente com aprendizagem inovadora;
- Aperfeiçoa os seus membros, motivando e melhorando suas competências;
- Mobiliza as energias e inteligências para atingir seus objetivos.

Através da informação a organização pode rever alguns critérios determinantes para o seu sucesso: capacidade de obter a qualidade total do serviço prestado ao cliente; rapidez de reação; capacidade de evolução; e capacidade de inovação.

A organização deve sempre recolher informações visando melhorar seu poder de adaptação nos mercados em que atua. A informação agrega ganho de capital ao produto ou serviços oferecidos.

2.1.1 A Organização e o Meio Ambiente

A organização necessita do meio ambiente (todos os elementos que, atuando fora de uma organização são relevantes para as suas operações) para retirar insumos (mão-de-obra, matéria-prima, capital, etc.) e devolve ao meio ambiente bens ou serviços. Este mesmo meio ambiente exerce pressão sobre a organização, fazendo com que ela tenha de se modificar, inovar para poder sobreviver. A Figura 2.2 ilustra a relação da organização com o meio ambiente.



Figura 2.2: A organização se relacionando com o Meio Ambiente [LAU99].

Na Figura 2.2 pode-se verificar a relação da organização com alguns dos elementos que constituem o meio ambiente onde ela está inserida. Existe uma relação de troca, ou seja, há uma dependência mútua entre a organização e os elementos que compõem este meio, como por exemplo: a organização necessita dos clientes para adquirir seus bens ou serviços e o cliente necessita da organização para que as suas necessidades sejam atendidas.

O retorno (*feedback*) da informação sobre desempenho da organização com relação ao meio ambiente é essencial para a criação de uma organização flexível onde existe um constante "aprendizado", que imediatamente implementa a estratégia de seus objetivos e reconhece a necessidade de modificar esses objetivos quando estes se tornam ineficazes [ANS83].

2.2 A Tecnologia da Informação (TI) como Ferramenta Organizacional

A relação da organização com o meio ambiente pode significar tanto uma oportunidade como uma ameaça. Assim, as organizações podem responder de forma reativa a uma pressão já existente, ou de modo pró-ativo a uma pressão esperada

através de estratégias, do foco no cliente e no serviço, de esforços contínuos de melhoria e da reengenharia de processo de negócio.

Essas reações são facilitadas pelo uso da Tecnologia da Informação (TI), a qual pode ser definida como toda forma de gerar, armazenar, veicular, processar e reproduzir a informação [FUR94].

Fundamenta-se a TI nos seguintes componentes [RES00]: *hardware* e seus dispositivos e periféricos; *software* e seus recursos; sistemas de telecomunicações; gestão de dados e informações; e gestão do conhecimento.

A TI é vital para todas as áreas funcionais de uma organização, e suas ferramentas fazem parte integral de todas essas áreas (finanças, contabilidade, *marketing*, vendas, produção, recursos humanos, etc.). Suas ferramentas são poderosos agentes de mudanças dentro da organização, pois, através da aplicação da ferramenta correta para o problema em questão, podem-se alcançar melhorias nos processos deficitários da organização, bem como diferenciais competitivos no ambiente de negócios [FUR94].

Os objetivos da TI são [TOR95]: introduzir mudanças fundamentais no trabalho; integrar funções do negócio da organização; aumentar a competitividade; criar ou manter oportunidades estratégicas de diferenciação da concorrência; possibilitar o enxugamento da estrutura organizacional.

As necessidades a serem atendidas pela TI são: necessidades organizacionais; necessidades de informação e de evolução tecnológica.

Os fatores críticos de sucesso apoiados pela TI são: o negócio da organização e seu grau de sucesso; o negócio da área de Informática e seu próprio sucesso; a percepção dos executivos da organização.

A TI permite alterar o modo como se estruturam e se administram as organizações e os processos.

Torres [TOR95] enumera diversos efeitos estratégicos ocorridos com o uso da TI. Aqui estão destacados os principais: mudanças radicais em processos operacionais com grandes ganhos de tempo e custos; melhoria da imagem da organização; informações rápidas aos clientes; agilidade nas transações; rapidez de

resposta a novas demandas do mercado; criação de laços fortes com fornecedores; criação de laços fortes com agentes comerciais ou clientes; conhecimento da concorrência; maior poder de análise para situações de risco; mudanças na estrutura e nos custos de um produto, uso de sistemas em situações impossíveis ou de grande dificuldade pelo ser humano; balanceamento de disponibilidades de recursos com demanda, rastreamento do produto desde sua fabricação até o término de sua vida útil, aumento de confiabilidade em diagnósticos e prescrições de solução de problemas; integração e aumento do potencial de vendas dos agentes de vendas.

Turban [TUR03] enumera as tendências tecnológicas da TI em função do sucesso obtido pelas organizações na sua utilização: melhoria constante na relação custo/desempenho; armazenamento e memória maiores; interfaces gráficas e amigáveis com o usuário; *Data Warehouse*; *Data Mining* (estudado com mais detalhes no capítulo 3, seção 3.1.1.4); ambiente orientado a objetos; gerenciamento de documentos eletrônicos; realidade virtual; sistemas inteligentes; computação portátil; fusão de eletro-eletrônicos; comércio eletrônico e computação doméstica integrada.

2.3 Comparação entre Economia da Informação e Economia Industrial

O uso da TI acabou por criar uma economia inteiramente nova, uma Economia de Informação, tão diferente da Economia Industrial quanto a Industrial era da Agrícola.

Atualmente vivemos na Economia da Informação e quando a fonte da riqueza das organizações se modifica, todo o meio ambiente em que ela está inserida também se modifica.

Zuboff [ZUB88] mostra as diferenças básicas entre a Economia Industrial e a Economia da Informação, onde as organizações voltadas para dentro, ou seja, apenas para os seus processos, são tidas como participantes da Economia Industrial, e as

organizações voltadas para o meio ambiente, como participantes da Economia da Informação.

Nas últimas décadas, a Revolução da Informação que gerou a Economia da Informação está mudando a fonte de riqueza. A nova fonte de riqueza não é material, é informação, conhecimento aplicado ao trabalho para criar valor. A busca de riqueza agora é, em grande medida, a busca de informação e a aplicação de capital intelectual aos meios de produção [MAS99; STE00].

Como fonte de riqueza, a informação vem de várias formas: desde os dados eletrônicos de curta validade a anos de pesquisa acumulada incorporada em memórias de computadores que operam fábricas automatizadas ao capital intelectual carregado no cérebro das pessoas.

As organizações da Economia Industrial são ineficazes na Economia da Informação, pois fortalecem a profunda especialização da mão-de-obra, a produção em massa e uma estrutura organizacional hierárquica extremamente rígida.

A TI é fundamentalmente diferente da Tecnologia Industrial no sentido de que ela pode ser programada para fazer a tarefa requerida e, se necessário, pode ser continuamente ajustada.

A maneira pela como as roupas são cortadas pode ser uma operação de elevado valor adicionado ou de perda de dinheiro. Hoje, com os cortadores controlados por computador, a máquina pode ser programada para produzir qualquer número de tamanhos diferentes em qualquer volume de peças, tornando assim economicamente viável atender a qualquer pedido do cliente [MCG95].

Os seres humanos descobriram uma maneira de aplicar ao trabalho o conhecimento rapidamente crescente que possuem para criar valor de maneiras até há bem pouco tempo desconhecidas.

Quando a TI transformou a informação no mais importante fator de produção, tornou a oportuna aquisição da melhor informação a meta número um da administração comercial.

Dentro de dados demográficos, sociológicos, econômicos e de *marketing* estão ocultas infindáveis idéias sobre produtos, serviços e estratégias de *marketing* que devem ser aproveitadas pela organização.

Hoje, mais do que nunca, ter uma estratégia comercial significa ter uma estratégia de informação, uma estratégia para reconhecer as oportunidades nas mudanças, uma estratégia para garantir que a organização tire total valor do conhecimento acumulado por seus trabalhadores em vez de permitir que o conhecimento desapareça ou se escoe para outros.

Torres [TOR95] diz que a essência da Estratégia da Era da Informação é transformar a carga de dados comerciais numa recompensa de oportunidades.

A concorrência sob a Era da Informação obriga as organizações a desenvolver qualidades como: serem globalizadas, reagirem rapidamente às novas oportunidades, serem bem informadas, mais especializadas em suas áreas do que seus concorrentes ou seus reguladores e serem capazes de explorar novas capacidades produtivas. Uma vez que a maior parte de seu capital será intelectual, elas tenderão a ser altamente flexíveis, sensíveis a seus ambientes políticos e sociais e sempre prontas a desviar as operações para países com climas mais favoráveis [ANG02].

A transformação da informação em conhecimento também é possível através da utilização da Inteligência Artificial (IA), que é o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor [RIC94].

A utilização da IA nesta transformação embasa o desenvolvimento deste trabalho na criação de uma Arquitetura Híbrida para extrair conhecimento de bases de dados presentes nas organizações. Para ilustrar a utilização da IA, a Figura 2.3 mostra uma breve cronologia da TI.

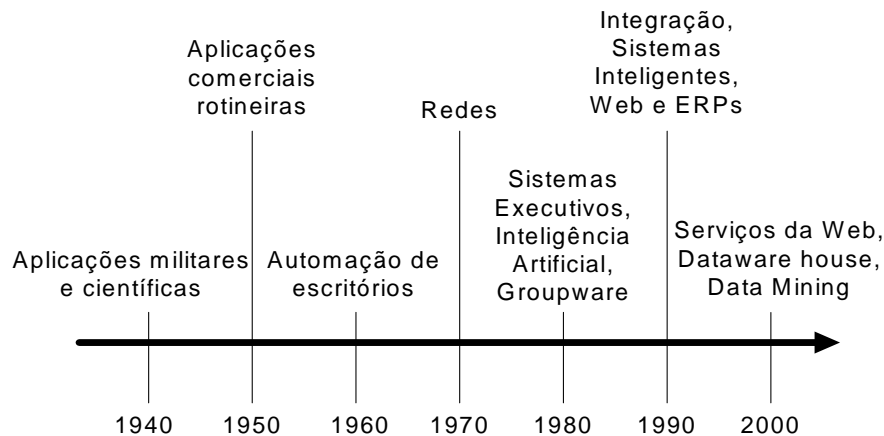


Figura 2.3: Breve cronologia da TI [TUR03]

Na Figura 2.3 pode-se observar a evolução da TI, caracterizada inicialmente por aplicações militares e comerciais rotineiras até os dias atuais (a partir do ano 2.000) com a presença do *Datawarehouse* e do *Data Mining* utilizando IA para descobrir conhecimento.

2.4 Do Mercado de Massa ao Nicho de Informação

Nas duas últimas décadas, pelo menos no mundo industrializado, assistiu-se à morte dos mercados de massa para mercadorias padronizadas.

Os clientes desejam essencialmente bens e serviços personalizados a preços de produção em massa [KOT80].

A noção de produção em massa significa essencialmente que uma organização fabrica o mesmo produto para um grande número de clientes, enquanto os produtos personalizados fornecem conjuntos de benefícios especiais que satisfazem determinadas necessidades particulares do cliente.

Torna-se cada vez mais necessário fabricar produtos altamente personalizados sem abrir mão dos benefícios de custo das técnicas de produção de massa. Essencialmente, o mercado está chegando ao ponto em que cada cliente é um nicho.

A informação está no âmago dessas tentativas de se chegar a um pluralismo de mercado.

Por exemplo, esforços para atingir a personalização em massa no processo de fabricação, através da implantação de estratégias de fabricação flexíveis, tornam-se possíveis através do uso de tecnologias onde a informação representa papel preponderante. A informação e o conhecimento embutidos em tecnologias de fabricação assistidas por computador (CAM) criam economias comparáveis às economias de escala que acionaram as estratégias de produção de massa durante a Revolução Industrial [MAS99].

As tecnologias de fabricação flexíveis utilizam a TI para embutir essas informações e esse conhecimento nas próprias máquinas, reduzindo a importância das economias de escala.

As administradoras de cartões de crédito oferecem um bom exemplo de estratégia de produto/serviço baseado em informação. Elas coletam e mantêm enormes quantidades de dados como um subproduto das suas operações de processamento de cartões de crédito. Uma vez que esses dados foram capturados, os mesmos poderão ser elaborados e revendidos como outras formas de produtos de informação. Para contas empresariais, por exemplo, os encargos podem ser resumidos e relatados por área geográfica ou tipo de gasto. Isso representa uma oportunidade adicional de extrair valor da informação que foi coletada como parte das atividades rotineiras de processamento de transações da organização.

Os requisitos estabelecidos pelo conceito de produto ótimo (quantidade maximizada, tempo de entrega minimizado e custos minimizados) somente se viabilizam com o uso intensivo da TI.

Hoje a estratégia baseia-se no domínio por meio de recursos estratégicos como o conhecimento e a informação e não mais somente no domínio dos recursos naturais [MAS99].

Essa estratégia afeta a organização como um todo e define as atividades comerciais, o modo de operar essas atividades e, particularmente, a forma de diferenciar seus produtos e serviços daqueles oferecidos pelos concorrentes. Assim,

uma estratégia abrangente deverá abordar duas questões: em primeiro lugar, as estratégias devem considerar os clientes e os segmentos de mercado aos quais a organização deseja servir e em segundo lugar, as estratégias devem considerar habilidades e recursos que a organização deverá reunir para fornecer produtos e serviços a esses mercados [ANS83]. A estratégia depende da informação, e a correta utilização da informação depende da TI.

A definição de clientes e mercados potenciais depende de informação externa sobre as necessidades e de interpretação inteligente dessas necessidades, de forma a possibilitar o bom uso dos pontos fortes da organização.

Segundo McGee [MCG95], a tradução das escolhas em relação a clientes e mercados deve levar em consideração cinco pontos:

1. A definição e o projeto de produtos e serviços a serem oferecidos.
2. O estabelecimento para a organização de objetivos de desempenho, financeiros e não financeiros.
3. A definição de processos organizacionais e operacionais que possam atender aos objetivos de desempenho, diferenciando os produtos e serviços da organização dos produtos e serviços de seus concorrentes.
4. O desenvolvimento de recursos de tal forma que maiores probabilidades sejam criadas para que os objetivos de desempenho sejam alcançados.
5. O monitoramento do desempenho organizacional e redirecionamento de recursos conforme necessário.

A questão da diferenciação é fundamental para a compreensão da estratégia, pois uma estratégia efetiva deve definir as formas pelas quais os produtos e serviços de uma organização serão superiores aos de seus concorrentes (e concorrentes em potencial) aos olhos dos clientes. Se uma estratégia não permitir que uma organização ofereça (ou pelo menos convença os consumidores de que oferece) melhor qualidade, menor custo, melhor serviço, ou alguma outra característica desejável, não será estratégia. Sem essa capacidade de claramente diferenciar seus

produtos e serviços dos oferecidos por seus concorrentes, uma organização não poderá almejar alcançar um desempenho superior.

Qual a informação que a organização precisa ter e dela se utilizar de maneira mais eficiente do que qualquer outra com relação a clientes, concorrentes e ambiente competitivo? Que informação as organizações precisam fornecer a seus clientes e fornecedores, e como estes a utilizarão (e combinarão com a sua própria informação) para fornecer subsídios relevantes para a organização?

Essas são perguntas a que uma estratégia competitiva com uso de informação deve procurar responder.

A informação desempenha papel especialmente importante no desenvolvimento de esforços para criar e manter a diferenciação.

No momento da interação entre a organização e o cliente, a informação torna-se um instrumento essencial para a individualização do serviço ao cliente num mundo anônimo. Essa personalização do serviço ao cliente evolui para a criação de um maior número de nichos de mercado, possível pela obtenção de informações mais precisas sobre esses grupos de clientes ou nichos: o limite dessa tendência é tratamento tão personalizado a ponto de cada cliente ser considerado um nicho de mercado.

A idéia de serviço individualizado ao cliente está embutida no conceito de que os consumidores cada vez mais exigem ser tratados como indivíduos e não apenas como membros de um grande grupo [KOT80].

Em suma, busca-se a experiência do comerciante do século XIX, que conhecia cada cliente pessoalmente, mas na escala global do século XXI.

2.5 Da Economia da Informação para a Economia do Conhecimento

Este cenário de plena utilização da informação e a necessidade de obtenção de conhecimento fizeram surgir o conceito de Economia do Conhecimento [TAP97], e o principal veículo desta nova economia é a *Internet*. Não basta possuir informação (as organizações, em geral, possuem muita informação), é necessário transformá-la em conhecimento.

A Gestão do Conhecimento é um dos pilares associados às novas Tecnologias da Informação como o KDD (*Knowledge Discovery in Database*) ou Descoberta de Conhecimento em Bases de Dados (capítulo 3), que transforma informação em conhecimento em benefício das organizações.

Cornella [COR97] chama a Economia do Conhecimento de Infonomia, ou seja, é a “Ciência para o ótimo aproveitamento da informação e a obtenção de conhecimento dentro das organizações”.

A chave está em aliar os negócios a TI para prever e analisar as mudanças no comportamento do cliente, as necessidades atuais e futuras deste cliente e como conseguir satisfazê-lo.

2.6 Resumo do Capítulo

O objetivo deste capítulo foi mostrar a importância da informação e de sua transformação em conhecimento para apoiar a tomada de decisão nas organizações.

O conhecimento possibilita a uma organização definir os segmentos de mercado para os quais deseja atuar e definir as estratégias necessárias para este fim.

Conhecer o meio ambiente e definir estratégias são atividades mais bem realizadas quando apoiadas na TI.

Desta forma, a Arquitetura Híbrida proposta neste trabalho foi desenvolvida para descobrir conhecimento em bases de dados apoiando a tomada de decisão organizacional.

O capítulo 3 descreve todo o processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* – KDD).

3 Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD)

Este capítulo trata da descoberta de conhecimento em bases de dados (KDD) através da análise das suas fases, das formas de interpretação dos dados (tarefas) e das aplicações nas diversas áreas do conhecimento.

3.1 Descoberta de Conhecimento em Bases de Dados (KDD)

Existem várias definições para KDD. Segundo Fayyad [FAY96a], o KDD pode ser definido como um processo de extração de conhecimentos válidos, novos, potencialmente úteis e compreensíveis, visando melhorar o entendimento de um problema. Brachman [BRA96] define KDD como um processo cujo uso de conhecimento é intensivo e estruturado em fases, consistindo de interações entre uma pessoa e um banco de dados, tendo por suporte um conjunto heterogêneo de ferramentas. Em Cabena [CAB97], descobrir conhecimento significa extrair, de grandes bases de dados, sem nenhuma formulação prévia de hipóteses, informações relevantes e desconhecidas, que podem ser utilizadas para a tomada de decisão. Para tanto, o KDD se utiliza de diversas áreas para realizar os seus processos como: métodos estatísticos, reconhecimento de padrões, visualização, banco de dados, aprendizado de máquina, Inteligência Artificial, *Data Warehouse*, dentre outras.

KDD é o termo criado pelo *Gartner Group* na década de oitenta para descrever todo o processo de extração de conhecimento dos dados devido ao crescimento vertiginoso das bases de dados das organizações.

Com toda essa informação armazenada, as organizações perceberam que não bastava apenas disponibilizá-la, era necessário interpretá-la, analisá-la e relacioná-la a fim de extrair conhecimento para apoiar a decisão [GOL05].

O KDD é um processo composto por fases que devem ser desenvolvidas para atingir o objetivo final, que é a extração de conhecimento.

3.1.1 Fases do KDD

As fases do KDD possuem numerosos passos, que envolvem um número elevado de decisões a serem tomadas, ou seja, é um processo iterativo. É também um processo iterativo, pois ao longo do processo KDD, um passo será repetido tantas vezes quantas se fizerem necessárias para que se chegue a um resultado satisfatório [FAY96a].

No processo de KDD cada fase possui uma intersecção com as demais. Desse modo, os resultados produzidos numa fase são utilizados para melhorar os resultados das próximas fases. O KDD é composto das seguintes fases: seleção dos dados, pré-processamento dos dados, transformação dos dados, mineração de dados (*Data Mining*) e interpretação/avaliação do conhecimento.

A Figura 3.1 ilustra as fases do processo do KDD, com início na apresentação dos dados até chegar à obtenção do conhecimento. A iteração entre as fases pode ser observada pelas setas tracejadas.



Figura 3.1: Fases do processo de KDD. Adaptado de Fayyad [FAY96a].

As três fases iniciais do KDD (Figura 3.1), que envolvem a seleção, o pré-processamento e a transformação, também chamadas de preparação dos dados, exigem bastante tempo, aproximadamente entre 60 e 80% do tempo utilizado em todo o processo, sendo que a maior parte desse tempo é consumida com a limpeza dos dados [PYL99].

A Figura 3.2 destaca o grande esforço requerido na preparação dos dados (seleção, pré-processamento e transformação).

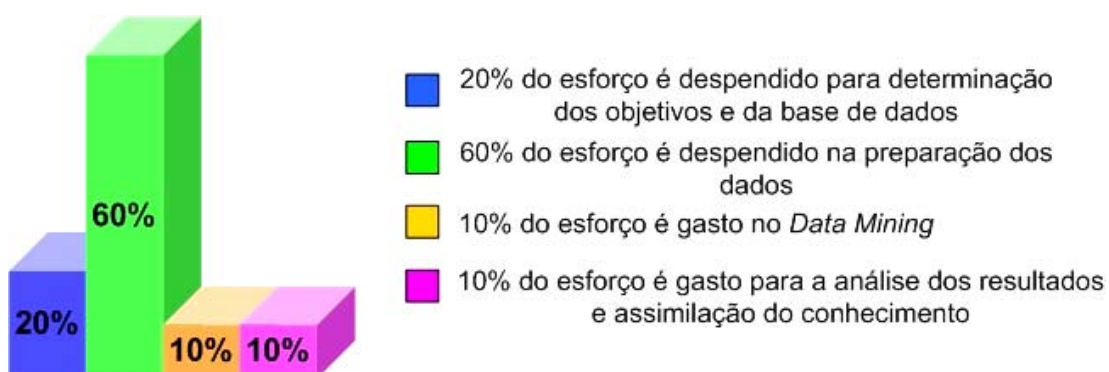


Figura 3.2: Esforço requerido para cada fase do processo de KDD. Adaptado de Cabena [CAB97].

Considerando em 100% o esforço total necessário para realizar todo o processo de KDD, pode-se verificar na Figura 3.2 que somente na fase de preparação de dados gasta-se 60% de todo o esforço.

3.1.1.1 Seleção de Dados

A seleção de dados consiste na criação de um conjunto de dados-alvo, ou dados selecionados. Nesta fase, seleciona-se um conjunto de dados ou focaliza-se um subconjunto de atributos (variáveis) em que a descoberta deverá ser efetuada. A seleção de dados vai variar de acordo com os objetivos da organização.

Na seleção dos dados escolhem-se apenas atributos relevantes do conjunto de atributos da base de dados. Em suma, a seleção de atributos consiste na escolha de um subconjunto de atributos relevantes para o objetivo da tarefa. O subconjunto selecionado é então fornecido para o algoritmo de mineração dos dados. Uma motivação para essa seleção é otimizar o tempo de processamento do algoritmo minerador, visto que ele apenas trabalhará com um subconjunto de atributos, desse modo diminuindo o seu espaço de busca.

3.1.1.2 Pré-processamento dos dados

As grandes bases de dados são altamente susceptíveis a ruídos, valores faltantes e inconsistentes. Dados limpos e compreensíveis são requisitos básicos para o sucesso da mineração dos dados [PYL99].

O pré-processamento dos dados tem por objetivo assegurar a qualidade dos dados selecionados [BAT03].

A limpeza dos dados envolve a verificação da consistência das informações, a correção de possíveis erros e o preenchimento ou a eliminação de valores nulos e redundantes. Nessa fase são identificados e removidos os dados, duplicados e/ou corrompidos. A execução dessa fase corrige a base de dados eliminando consultas

desnecessárias que seriam executadas pelo algoritmo minerador e que afetariam o seu processamento.

Os métodos de limpeza dos dados são herdados e dependentes do domínio da aplicação, assim, torna-se essencial a participação do analista de dados. Um exemplo simples de limpeza de dados seria a definição de um intervalo de possíveis valores para um determinado atributo, $\{0...10\}$. Caso surgisse qualquer valor diferente dos definidos no intervalo, esse dado seria retirado.

A limpeza dos dados consiste em resolver problemas com [BAT03]: dados com erros (valores discrepantes), registros repetidos e valores faltantes.

Os valores que são significativamente fora do esperado são denominados valores discrepantes ou *outliers*. Os *outliers* podem indicar uma notícia boa ou má. Uma boa notícia se indicarem uma nova tendência de resultados para as variáveis em questão, e uma má notícia se realmente forem dados inválidos.

Um tipo comum de *outlier* é devido a erro humano, como por exemplo, um registro de compra da ordem de milhões de reais. Esses registros devem ser corrigidos se valores razoáveis ou válidos estiverem disponíveis, caso contrário, devem ser excluídos da análise.

Como valores faltantes incluem-se aqueles que simplesmente não estão presentes no conjunto selecionado e os valores inválidos que foram eliminados durante a detecção de *outliers*. Os valores faltantes podem ocorrer devido a erros humanos, ou porque a informação não está disponível no momento do levantamento dos dados, ou quando os dados são selecionados considerando-se diferentes origens, gerando informações contraditórias. Xiaohui [XIA02b] apresenta um estudo detalhado sobre *outliers*.

A redução de dados é considerada uma técnica de pré-processamento dos dados e seu estudo tem a maior importância neste trabalho, porque a Arquitetura Híbrida proposta combina o RS como técnica de redução de atributos com as redes neurais artificiais.

A redução da quantidade de dados pode ser utilizada para reduzir o tempo de aprendizagem do algoritmo de mineração, quando o tempo para obter uma solução é

muito longo, e para tornar mais fácil a interpretação dos conceitos aprendidos [BAT03].

Na Figura 3.3 observa-se que a redução de dados pode ocorrer de duas formas: redução de atributos (colunas) e redução de casos (linhas).

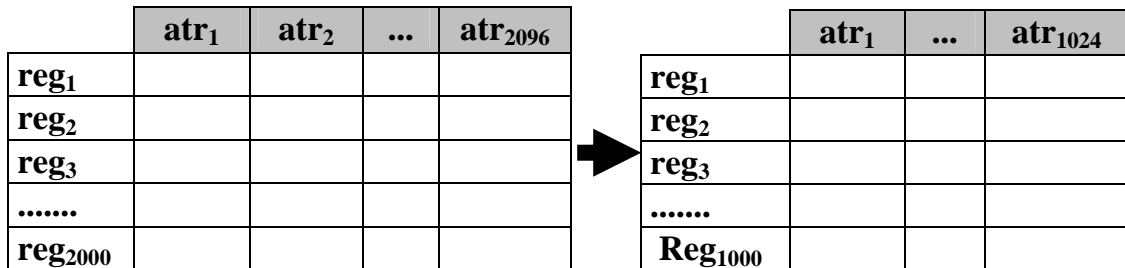


Figura 3.3: Formas de pré-processar dados. As linhas (reg) se referem a casos e as colunas (atr) se referem a atributos.

No caso de RS a redução é realizada nos atributos, também chamada de redução de dimensão.

3.1.1.3 Transformação de Dados

A fase de Transformação de Dados ou Codificação dos Dados tem como objetivo principal converter o conjunto bruto de dados em uma forma padrão de uso. Os dados de um atributo podem ser padronizados (normalizados) para cair dentro de uma faixa de valores, como por exemplo: -1,0 a 1,0 ou 0,0 a 1,0.

A transformação de dados para o atributo idade poderia ocorrer da seguinte forma: {0...18}, Faixa 1; {19...25}, Faixa 2; {26...30}, Faixa 3 e assim por diante [DOU95].

Nesse exemplo, os valores contínuos das idades foram discretizados em faixas. Em resumo, essa fase converte os dados para a forma mais adequada para a construção e interpretação do modelo.

A Transformação de Dados é potencialmente a tarefa que requer grande habilidade no processo de KDD. Tipicamente essa etapa exige a experiência do analista de dados e seu conhecimento nos dados em questão. Embora o processo de KDD possa ser executado sem essa fase, nota-se que quando efetivada os resultados obtidos são mais intuitivos e valiosos, além de, na maioria das vezes, facilitarem a construção do modelo.

As vantagens de se codificar um atributo são: melhorar a compreensão do conhecimento descoberto; reduzir o tempo de processamento para o algoritmo minerador; facilitar o algoritmo a tomar decisões globais, já que os valores dos atributos foram englobados em faixas. Como desvantagem cita-se a redução da medida de qualidade de um conhecimento descoberto, perdendo-se assim detalhes relevantes sobre as informações extraídas.

3.1.1.4 *Data Mining* (Mineração de Dados)

O termo *Data Mining* surgiu devido às semelhanças entre a procura de informação importante para o mundo dos negócios (numa base de dados) e o ato de minerar a montanha para encontrar um veio de ouro. Ambos os processos requerem ou a seleção de um imenso amontoado de material, ou um sondar inteligente desse material, para encontrar o sítio onde está o valor desejado.

A mineração de dados é considerada a etapa mais importante do processo de KDD. Caracteriza-se pela existência do algoritmo minerador (*Data Mining*), que diante da tarefa especificada será capaz de extrair de modo eficiente conhecimento implícito e útil de um banco de dados.

Segundo Berry [BER97], *Data Mining* é a exploração e análise, por meios automáticos ou semi-automáticos, de grandes quantidades de dados para descobrir modelos e regras significativas, permitindo a uma empresa aumentar, por exemplo, suas operações de *marketing*, vendas e apoio aos clientes pela melhor compreensão da clientela. Essas técnicas e ferramentas são igualmente aplicáveis em outras áreas do conhecimento.

De uma forma mais simples, *Data Mining* é produzir conhecimento novo escondido em grandes bases de dados.

O *Data Mining* usa técnicas baseadas em descobertas por meio de procura de padrões dos dados, o que é feito com o emprego de uma série de algoritmos inteligentes para encontrar relações fundamentais entre os dados [PIA93]. Tais técnicas e algoritmos caracterizam-se por serem "automáticos", isto é, são de baixa interação com o ser humano, procurando restringir a participação humana aos processos de ajuste necessários aos mecanismos de busca.

De certo modo, enquanto as técnicas de busca tradicionais procuram respostas para perguntas realizadas, as técnicas de *Data Mining* permitem avaliar como as perguntas se relacionam com as respostas (padrões e relações) encontradas. A Figura 3.4 mostra um esquema simplificado de *Data Mining*.



Figura 3.4: Esquema simplificado de *Data Mining*.

Achadas essas relações e padrões, é fornecida uma base de regras que servem de apoio aos processos de tomada de decisão. Para tal, utilizam-se técnicas de procura baseadas em Inteligência Artificial, como as Redes Neurais Artificiais, as Árvore de Decisões, a Teoria dos Conjuntos *Fuzzy*, os Algoritmos Genéticos ou, ainda, combinações entre essas técnicas gerando os chamados sistemas híbridos (Arquiteturas Híbridas) [HAN01]. Pode-se encontrar em Mitra [MIT02] um estudo sobre as técnicas de *Data Mining*.

O resultado obtido pela aplicação do *Data Mining* deve ser compacto, legível (apresentado de alguma forma simbólica), interpretável, e deve representar fielmente os dados que lhe deram origem [BAC02].

Uma questão fundamental do *Data Mining* é quanto à qualidade do conhecimento extraído, levando em consideração: a precisão, a compreensibilidade,

a surpresa e ou potencial interesse do conhecimento obtido. É necessário verificar o que foi aprendido, o que há de novo, eliminando o conhecimento “inútil” e muito óbvio. Mongiovi [MON95] exemplifica: se uma cadeia de lojas resolve buscar associações entre os itens que ela vende, certamente haverá associações entre itens que farão sentido, e outras que não farão. Faz-se necessário, então, que um especialista dos negócios da empresa avalie quais são as associações realmente relevantes para a empresa. Por último, as associações descobertas precisam ser acionáveis, isto é, é necessário que possam ser realizadas ações simples para que o conhecimento gerado seja traduzido em vantagem aos negócios da empresa.

O *Data Mining* serve para prever tendências e comportamentos futuros, permitindo a tomada de decisão baseada em fatos e não em suposições.

O *Data Mining* pode responder a questões de negócio que tradicionalmente demandariam muito tempo para resolver. Ele explora as bases de dados à procura de padrões escondidos, encontrando dados que permitem prever tendências e comportamentos futuros, que os especialistas podem não descobrir devido ao fato de essa informação sair do limite de suas expectativas e possibilitando a tomada de decisão.

3.1.1.5 Interpretação/Avaliação do Conhecimento

Esta fase também é conhecida como pós-processamento. Após a fase de mineração de dados é necessária a interpretação do conhecimento descoberto, ou algum processamento desse conhecimento. Essa interpretação deve ser incluída no algoritmo minerador, porém algumas vezes é vantajosa a implementação separadamente. Em geral, a principal meta dessa fase é melhorar a compreensão do conhecimento descoberto pelo algoritmo minerador, validando-o através de medidas da qualidade da solução e da percepção de um analista de dados. Esses conhecimentos serão consolidados em forma de relatórios demonstrativos com a documentação e explicação das informações relevantes ocorridas em cada etapa do

processo de KDD. Uma maneira genérica de obter a compreensão e interpretação dos resultados é utilizar técnicas de visualização [BIG96].

Segundo Goldschmidt [GOL05], as técnicas de visualização estimulam a percepção e a inteligência humana, aumentando a capacidade de entendimento e a associação de novos padrões.

3.2 Tarefas do KDD

Existem várias formas de interpretação dos dados pelo KDD denominadas tarefas. As tarefas mais comuns são [FAY96a]: associação, classificação, clusterização (agrupamento) e visualização.

3.2.1 Associação

Também chamada de Análise de Cestas de Mercado (*Market Basket Analysis*).

A idéia principal da associação é identificar grupos de itens tipicamente associados (o que vai com o quê?) [AGR01].

Dada uma coleção de itens e um conjunto de registros, cada qual contendo um número de itens da coleção, através de técnicas de associação realizam-se operações sobre o conjunto de registros retornando afinidades entre a coleção de itens.

O objetivo da técnica é encontrar tendências, a partir de grande número de transações que possam ser usadas para entender e explorar padrões de compra.

Regras facilmente encontradas assumiriam a forma "80% das vendas de cerveja também correspondem à venda de batatas fritas", ou "30% das vendas de fornos elétricos também correspondem à venda de luvas térmicas" ou, ainda, quando se compram batatas fritas, 65% das vezes também é adquirida coca-cola, a não ser que haja uma promoção, e nesse caso a coca-cola é campeã em 85% das vezes.

Uma vez encontradas as regras, estas poderiam ser usadas de maneiras diversas, de acordo com o tipo de atividade analisada. Poderiam ser realizadas mudanças na disposição de produtos, na determinação de políticas de formação de estoques ou de pedidos de compra, na promoção de campanhas direcionadas de vendas, no estabelecimento de práticas comuns (ou incomuns) de vendas. Na área financeira poderiam ser avaliados serviços que freqüentemente são adquiridos em conjunto.

3.2.2 Classificação

As técnicas de classificação criam automaticamente um modelo a partir de um conjunto inicial de registros, que serve de exemplo e é chamado de conjunto de treinamento. Os registros do conjunto de treinamento devem pertencer a um pequeno grupo de classes pré-definidas (pelo analista) [HAN01].

O modelo é composto de padrões, essencialmente generalizações em relação aos registros, os quais são usados para diferenciar as classes. Uma vez obtido o modelo, este é usado para classificar automaticamente os demais registros.

A classificação pode ser utilizada com êxito, por exemplo, para aplicações típicas de cartões de crédito. Dada uma base de dados dos usuários de cartões e tomando-se como exemplo aqueles cujo histórico de crédito é conhecido, atribui-se a estes um grau de risco de acordo com seu histórico (alto, médio ou baixo). Desse modo, por meio das técnicas de classificação, podem ser obtidas regras para a caracterização do grau de risco de um usuário tendo em vista dados como a faixa de renda familiar, a idade, a área de moradia (por exemplo: usuários de 30 a 50 anos, com renda familiar superior a 20.000 reais anuais, moradores da região X apresentam baixo risco).

De maneira análoga, numa aplicação de detecção de fraudes seriam obtidos registros de operações fraudulentas e não-fraudulentas. Posteriormente seriam estabelecidos os atributos desses registros, que tipificariam uma determinada operação como fraudulenta ou não, permitindo a classificação dos demais registros.

O método se ajusta a uma vasta gama de situações, tais como concessão de empréstimos, detecção de sonegação fiscal etc.

3.2.3 Clusterização (Agrupamento)

A clusterização transforma registros com grande número de atributos em conjuntos relativamente menores (segmentos).

Essa transformação é realizada, automaticamente, por meio de identificação das características que distinguem o conjunto de dados e pelo seu posterior particionamento [HAN01]. Não é necessário identificar os agrupamentos desejados nem os atributos que devem ser utilizados para a produção dos segmentos.

O objetivo nessa tarefa é maximizar similaridade *intra-cluster* e minimizar similaridade *inter-cluster* [GOL05].

Os resultados de uma operação de clusterização podem ser usados de duas diferentes maneiras. Ora para produzir um sumário da base de dados por meio das características de cada *cluster*, ora como dados de entrada para outras técnicas, por exemplo, classificação.

A clusterização pode ser usada em casos que façam uso de modelos de segmentação de população, tais como segmentação demográfica de mercados de consumidores (identificar grupos homogêneos de elementos, identificar elementos dentro do mesmo grupo maximamente semelhantes) implicando possível comparação dos hábitos de consumo de múltiplos segmentos de população, visando determinar campanhas de vendas. A clusterização será tratada com mais profundidade no capítulo 5.

3.2.4 Visualização

As ferramentas de visualização não são propriamente tarefas de *Data Mining*, mas sim meios de analisar e observar os dados de uma determinada base de dados [BER97; HAN01].

A visualização fornece meios de obter sumários visuais dos dados de uma base de dados. No caso de técnicas de clusterização, podem ser usadas ferramentas de visualização para determinar qual ou quais *clusters* criados são úteis ou interessantes para os métodos de *Data Mining*. No caso específico da rede SOM, as formas de visualização mais utilizadas são a Matriz-U [ULT89] e o Mapa por Similaridade de Cor, conforme detalhadas no capítulo 5, seção 5.9.

As ferramentas de visualização podem ainda ser usadas como um mecanismo de compreensão da informação extraída por meio das técnicas de *Data Mining*. Características difíceis de detectar pela simples observação de linhas e colunas com valores numéricos podem se tornar óbvias se forem observadas graficamente.

Por meio dessas ferramentas podem ser encontrados características ou fenômenos pouco comuns ou interessantes sem que se esteja diretamente procurando por eles [LEE95].

No capítulo 5 serão aprofundados estudos sobre as formas de visualização da rede SOM.

3.3 Áreas de Aplicação

Pode-se aplicar o *Data Mining* em diversas áreas, como [BER97; BRA96; CAR01; HAN01]:

1 - Aplicações Comerciais

É crença corrente no comércio em geral que a maior arma mercadológica é compreender as necessidades individuais de cada cliente e atender a essas

necessidades. Com essa finalidade, um número sempre crescente de companhias procura aplicar técnicas de *Data Mining* que identifiquem as preferências do consumidor e seus padrões de compra [PIA96].

Tais informações podem ser usadas para direcionar campanhas de vendas, promover oferta combinada de serviços ou produtos, avaliar o comportamento do mercado e para detectar novas tendências mercadológicas ou necessidades de consumo; também podem auxiliar na formação de estoques, estabelecer disposições mais eficientes de produtos nos balcões ou setores de um ponto de venda (visando estimular o consumo), dentre uma vasta gama de outras possibilidades. Em Ebecken [EBE04] pode-se verificar diversos trabalhos na área de tomada de decisões empresariais utilizando KDD.

2 - Aplicações Governamentais

Podem ser criados mecanismos de detecção de fraudes, descoberta de sonegação de impostos, avaliação de atividades censitárias, acompanhamento de oscilações em atividades econômicas sob a ótica de variações de políticas cambiais, de incentivo fiscal, modificação de taxas e índices e aplicações na área policial [ZHI02], dentre outras.

3 - Aplicações Científicas

Áreas com possíveis aplicações seriam pesquisa médica [GRY01], teste de novos medicamentos ou associação de medicamentos diversos, pesquisa subatômica, descoberta de galáxias e outros corpos celestes a partir de imagens capturadas por telescópios.

4 - Aplicações Financeiras

Há aplicações em áreas como prevenção e detecção de fraudes, possíveis aplicações em concessão de crédito, descoberta de áreas seguras para investimentos,

serviços adquiridos em diferentes áreas geográficas ou por diferentes segmentos populacionais [TAY02].

3.4 Resumo do Capítulo

Este capítulo procurou descrever todo o processo de KDD através da análise de suas fases, tarefas, técnicas e aplicações.

Essa descrição contribuiu para justificar o desenvolvimento deste trabalho e a finalidade da Arquitetura Híbrida proposta, que é a descoberta de conhecimento. Os experimentos realizados no capítulo 6 descrevem em contextos concretos todo o processo de KDD, com especial referência ao experimento de número 4 (seção 6.3.4).

Nos próximos dois capítulos deste trabalho inicia-se o estudo das técnicas escolhidas para compor a Arquitetura Híbrida. No capítulo 4 será estudado o RS e no capítulo 5 a rede SOM.

4 Teoria dos *Rough Sets* (Teoria dos Conjuntos Aproximados)

Neste capítulo será feita uma discussão da Teoria dos *Rough Sets* (RS), também denominada Teoria dos Conjuntos Aproximados (TCA), com o objetivo de apresentar os principais conceitos, a motivação do estudo, bem como exemplos de aplicação da Teoria.

4.1 Introdução

A Teoria dos *Rough Sets* (RS) foi proposta por Zdzislaw Pawlak [PAW82] em 1982 como um novo modelo matemático para representação do conhecimento e tratamento de incerteza, tendo sido usada, posteriormente, para o desenvolvimento de técnicas para classificação aproximada em aprendizado de máquina. Devido a essas características, tem-se utilizado essa teoria em Inteligência Artificial, especialmente nas áreas de aquisição de conhecimento, raciocínio indutivo e descoberta de conhecimento em base de dados.

Conjuntos aproximados podem ser considerados conjuntos com fronteiras nebulosas, ou seja, conjuntos que não podem ser caracterizados precisamente, utilizando-se dos atributos disponíveis [PAW91].

A incerteza pode se manifestar de diversas formas, como: imprecisão, incompletude, inconsistência, etc. RS trata de um tipo fundamental de incerteza, a indiscernibilidade. A indiscernibilidade surge quando não é possível distinguir elementos de um mesmo conjunto, e representa a situação em que esses elementos parecem todos ser um único elemento [UCH98].

Os conceitos de RS têm se mostrado muito úteis quando aplicados a problemas do tipo: redução de atributos, descoberta de dependência entre atributos e na descoberta de padrões entre os dados [PAW96a].

A redução de atributos realizada pelos RS é feita através dos chamados redutos, que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais. Este procedimento de eliminação de atributos irrelevantes é uma das características da Teoria.

Devido às características comentadas, a crescente utilização de RS pode ser comprovada pelo número de aplicações e publicações científicas nas seguintes áreas: KDD e *Data Mining* [ADJ01; JIA03; LIN97,02; TSU96,98,00, XIA96], medicina [KOM99a; TSU00,01a,01b), negócios [BEY01; CAB04; EBE04, GOL95; IJE00; JIA03; NAR04], engenharia [BON02; LAM02; LON02; ZIA93; ZON02]; reconhecimento de padrões [SAS06; XIA97; ZHA04], *Text Mining* [YU05], dentre outros.

4.2 Principais Conceitos

Os principais conceitos dos RS são: Espaços Aproximados, Aproximação Inferior (AI), Aproximação Superior (AS), Sistema de Informação (S), Sistema de Decisão (SD) e Indiscernibilidade (IND).

Este trabalho não tem como finalidade o aprofundamento no formalismo matemático dos RS, que é grande e considerado um aspecto importante da Teoria. Para isto recomenda-se os trabalhos de Uchôa [UCH 97a, 97b, 97c, 98].

Este formalismo matemático não é sistematicamente abordado, não é padronizado e tampouco muito explorado. Uma das razões para isso se deve ao fato dos RS ser uma teoria razoavelmente recente. Ainda não existe uma padronização da notação matemática, assim, em alguns casos foi necessário adotar uma notação própria, com o objetivo de torná-la mais clara, como: Aproximação Inferior (AI) \underline{B} ,

Aproximação Superior (AS) \bar{B} , Sistema de Informação (S), atributos condicionais

(C), atributo de decisão (d), Sistema de Decisão (SD), Região de Fronteira $RF(X)$, Região Negativa $RN(X)$ e Reduto (RED). Deve-se entender também que a palavra elementos é tratada como sinônimo de casos, exemplos ou registros que compõem uma base de dados.

4.2.1 Espaços Aproximados

Um espaço aproximado é um par ordenado $A = (U, R)$, onde: U é um conjunto não vazio, denominado conjunto universo, e R é uma relação de equivalência sobre U , denominada Relação de Indiscernibilidade. Uma relação binária $R \subseteq X \times X$, a qual é reflexiva (um elemento está relacionado com ele próprio xRx), simétrica (se xRy então yRx) e transitiva (se xRy e yRz então xRz), é chamada de relação de equivalência. Dados os elementos $x, y \in U$, se xRy então x e y são indiscerníveis em A , ou seja, a classe de equivalência definida por x é a mesma que a definida por y , i.e., $[x]R = [y]R$.

A classe de equivalência de um elemento $x \in X$ consiste de todos os elementos $y \in X$ para os quais xRy . Os elementos que são indiscerníveis formam conjuntos chamados de conjuntos elementares. Dessa forma, pode-se dizer que as classes de equivalência de R são os conjuntos elementares de A . Na Figura 4.1 pode-se visualizar o espaço aproximado $A = (U, R)$.

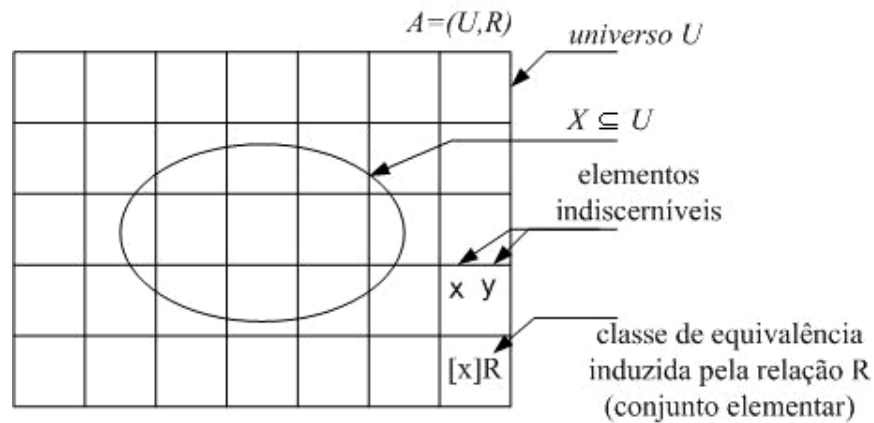


Figura 4.1: Espaço Aproximado $A = (U, R)$ e $X \subseteq U$.

A forma mais comum para representação dos dados em RS é através de um sistema de informação (S) que contém um conjunto de elementos, sendo que cada elemento tem uma quantidade de atributos condicionais. Esses atributos são os mesmos para cada um dos elementos, mas os seus valores nominais podem diferir (Tabela 4.1). Dessa forma, um sistema de informação é um par ordenado $S = (U, C)$, onde U é um conjunto finito e não-vazio de elementos chamado de universo (Figura 4.1), e C é um conjunto finito e não-vazio formado pelos atributos. Cada atributo $a \in C$ é uma função $a: U \rightarrow V_a$, onde V_a é o conjunto dos valores permitidos para o atributo a (sua faixa de valores). Na Tabela 4.1, onde é apresentado o sistema de informação S , pode-se observar os principais conceitos de RS, o espaço aproximado $A = (U, R)$, o universo U formado pelos elementos $e1$; $e2$; $e3$; $e4$; $e5$, $e6$ e os atributos (C) Experiência do Vendedor, Qualidade do Produto e Boa Localização e R a relação de equivalência sobre U .

LOJA	Experiência do Vendedor	Qualidade do Produto	Boa Localização
e1	Alta	Boa	Não
e2	Média	Boa	Não
e3	Média	Boa	Não
e4	Baixa	Média	Não
e5	Média	Média	Sim
e6	Alta	Média	Sim

Tabela 4.1: Exemplo de um Sistema de Informação (S)

4.2.1.1 Indiscernibilidade

O principal conceito envolvido em RS é a Relação de Indiscernibilidade [PAW96a], a qual normalmente está associada a um conjunto de atributos. Se tal relação existe entre dois elementos, isso significa que todos os valores nominais dos seus atributos são idênticos com respeito aos atributos considerados, portanto não podem ser discernidos (distinguidos) entre si.

Para cada subconjunto de atributos $B \subseteq C$ no sistema de informação $S = (U, C)$, uma relação de equivalência $INDs(B)$ é associada, chamada de Relação de Indiscernibilidade definida como:

$$INDs(B) = \{(x, y) \in U^2 / \forall \mathbf{a} \in B, \mathbf{a}(x) = \mathbf{a}(y)\} \quad (4.1)$$

O conjunto de todas as classes de equivalência na relação $INDs(B)$ é representado por $U/INDs(B)$, denominado quociente de U pela relação $INDs(B)$.

Em muitos casos é importante a classificação dos elementos considerando um atributo de decisão que informa a decisão a ser tomada. Dessa forma, um SI que apresenta um atributo de decisão é denominado Sistema de Decisão (SD).

Um SD pode ser representado por $SD = (U, C \cup \{d\})$, onde $d \notin C$ é o atributo de decisão. A Tabela 4.2 mostra um SD obtido a partir do sistema de informação S da Tabela 4.1, destacando os atributos condicionais (Experiência do Vendedor, Qualidade do Produto e Boa Localização) e o atributo de decisão (Retorno).

	Atributos Condicionais			Atributo de Decisão
Loja	Experiência do Vendedor	Qualidade do Produto	Boa Localização	Retorno
e1	Alta	Boa	Não	Lucro
e2	Média	Boa	Não	Prejuízo
e3	Média	Boa	Não	Lucro
e4	Baixa	Média	Não	Prejuízo
e5	Média	Média	Sim	Prejuízo
e6	Alta	Média	Sim	Lucro

Tabela 4.2 - Sistema de Decisão (Sistema de Informação com o atributo de decisão Retorno)

Os valores dos atributos são chamados de valores nominais e estão expressos como: Experiência do Vendedor {Alta, Média, Baixa}; Qualidade do Produto {Boa, Média}; Boa Localização {Não, Sim} e Retorno {Lucro, Prejuízo}. Considerando cada atributo condicional de forma independente, a relação de equivalência do sistema de informação S (Tabela 4.2) forma os seguintes conjuntos elementares:

experiência do vendedor Alta {e1, e6}; Média {e2, e3, e5}; Baixa {e4}; Qualidade do Produto: Boa {e1, e2, e3}; Média {e4, e5, e6} e Boa Localização: Não {e1, e2, e3, e4}; Sim {e5, e6}.

Ao utilizar todos os atributos condicionais do sistema de informação S da Tabela 4.2 obtêm-se os seguintes conjuntos elementares: {e1}, {e2, e3}, {e4}, {e5} e {e6}. Observando a Tabela 4.3, pode-se perceber que existem 2 elementos (casos) {e2} e {e3} iguais (destacados em negrito), no que se refere a valores de atributos condicionais.

Loja	Experiência do Vendedor	Qualidade do Produto	Boa Localização	Retorno
e1	Alta	Boa	Não	Lucro
e2	Média	Boa	Não	Prejuízo
e3	Média	Boa	Não	Lucro
e4	Baixa	Média	Não	Prejuízo
e5	Média	Média	Sim	Prejuízo
e6	Alta	Média	Sim	Lucro

Tabela 4.3 Sistema de Decisão com os elementos e2 e e3 indiscerníveis, com relação aos atributos condicionais.

Existindo a Relação de Indiscernibilidade entre os elementos {e2} e {e3} como mostrado na Tabela 4.3, significa que todos os valores nominais de seus atributos são idênticos com relação ao subconjunto de atributos B ($B \subseteq S$) considerado, ou seja, não podem ser diferenciados entre si. A redução de atributos será tratada com maior profundidade na seção 4.5 deste capítulo.

4.2.2 Aproximação de Conjuntos

A Tabela 4.3 apresenta os elementos do sistema de informação S segundo as características do atributo de decisão. Pode-se então, fazer a seguinte pergunta: quais características dos atributos condicionais definem o retorno da loja como lucro ou prejuízo? Note-se que não há uma resposta única para esta pergunta, pois as lojas

$\{e2\}$ e $\{e3\}$ apresentam as mesmas características dos atributos condicionais, mas se diferenciam no atributo de decisão.

Pode-se dizer com certeza, conforme a Tabela 4.3, que qualquer loja com características iguais às das lojas $\{e1\}$ e $\{e6\}$ terão lucro, assim como qualquer loja com características iguais às das lojas $\{e4\}$ ou $\{e5\}$ terá prejuízo, porém, nada se pode afirmar para lojas com características iguais às das lojas $\{e2\}$ e $\{e3\}$, pois, apesar de apresentarem atributos condicionais com as mesmas características, possuem atributos de decisão diferentes. São nesses casos que RS pode ser aplicado.

Um conjunto definível em S é qualquer união finita de conjuntos elementares. Para cada conceito X que é o conjunto de elementos com respeito a B , ou seja, X é obtido através das informações dos atributos de B , são computados o maior conjunto definível contido em X e o menor conjunto definível que contém X . O primeiro conjunto é chamado de Aproximação Inferior (AI) de X , enquanto o segundo conjunto é chamado de Aproximação Superior (AS) de X [PAW82].

A Aproximação Inferior $\underline{B}(X)$ e a Aproximação Superior $\bar{B}(X)$ de um conjunto de elementos $X \subseteq U$ com respeito a um conjunto de atributos $B \subseteq S$ (definindo uma relação de equivalência em U) pode ser definida em termos das classes na relação de equivalência, da seguinte forma:

$$\underline{B}(X) = \{x \in U \mid U/\text{INDs}(B) \subseteq X\} \quad (4.2)$$

$$\bar{B}(X) = \{x \in U \mid U/\text{INDs}(B) \cap X \neq \emptyset\} \quad (4.3)$$

Os elementos da Aproximação Inferior $\underline{B}(X)$ são classificados com certeza como membros de X , utilizando o conjunto de atributos B , enquanto os elementos da Aproximação Superior $\bar{B}(X)$ podem ser classificados como possíveis membros de X , utilizando o mesmo conjunto de B . Os elementos que correspondem a lucro podem ser tomados como exemplos do conceito X . Com base na Tabela 4.3 pode-se observar que existem três elementos que possuem como atributo de decisão lucro

$\{e1, e3, e6\}$, porém, existe Relação de Indiscernibilidade entre os elementos $\{e2\}$ e $\{e3\}$, impedindo que $\{e3\}$ seja considerado com certeza como membro de X . Assim, somente os elementos $\{e1, e6\}$ podem ser classificados como membros de X e elementos da Aproximação Inferior $\underline{B}(X)$. A Figura 4.2 ilustra a Aproximação Inferior $\underline{B}(X)$ destacada na cor cinza.

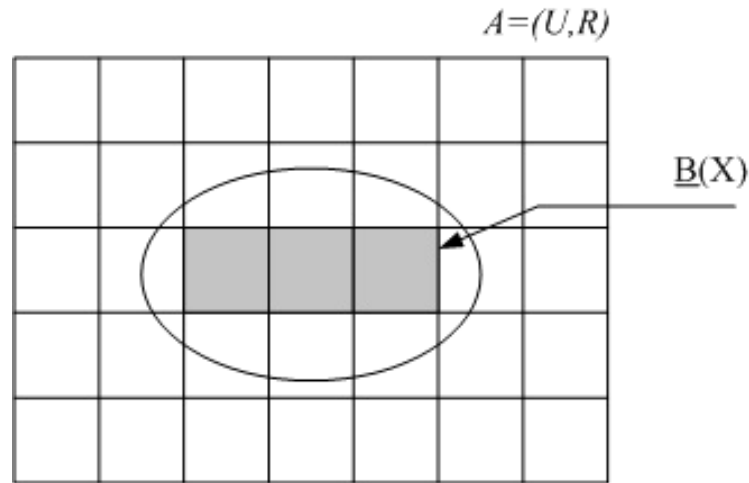


Figura 4.2: Aproximação Inferior de X .

Na Figura 4.2, nos quadrados de cor cinza, estão contidos os elementos que correspondem a Aproximação Inferior $\underline{B}(X)$, os quadrados em branco tocados pela elipse (X) estão os elementos que correspondem a Aproximação Superior $\bar{B}(X)$ e os quadrados em branco que não são tocados pela elipse correspondem com certeza ao elementos que não pertencem a $\bar{B}(X)$ (Região Negativa).

Como já foi dito, na Aproximação Superior $\bar{B}(X)$ são classificados os elementos que são possíveis membros de X . Dessa forma, a Aproximação Superior reúne os elementos com atributo de decisão igual a lucro $\{e1, e3, e6\}$ e também os elementos que possuem atributo de decisão igual a prejuízo, desde que exista uma Relação de Indiscernibilidade entre os elementos, como ocorre entre o elemento $\{e2\}$ e o elemento $\{e3\}$. A Figura 4.3 ilustra a Aproximação Superior $\bar{B}(X)$.

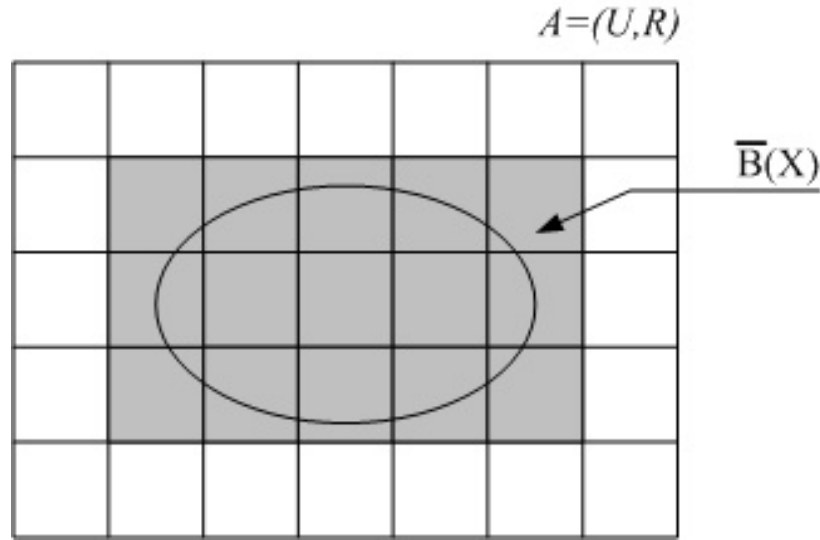


Figura 4.3 Aproximação Superior de X.

Portanto, obtém-se uma Região de Fronteira de X pela diferença de $\bar{B} X$ para $\underline{B}(X)$, representada por $RF(X)$, ou seja, $RF(X) = \bar{B}(X) - \underline{B}(X)$. A Região de Fronteira, também chamada de Duvidosa, possui somente os elementos que não podem ser classificados com certeza como pertencentes em X, utilizando o conjunto de atributos B. É a região formada pelos elementos de U que pertencem à Aproximação Superior, mas que não pertencem à Aproximação Inferior.

Um conjunto X é definido como *rough* (impreciso) se a sua Região de Fronteira é diferente do conjunto vazio ($RF(X) \neq 0$), e é definido como *crisp* (preciso) se o conjunto for vazio ($RF(X) = 0$). Na Tabela 4.3, pode-se observar que os elementos {e2, e3} fazem parte da Região de Fronteira.

A Região Negativa é dada pela diferença dos elementos de U para $\bar{B}(X)$, representada por $RN(X) = U - \bar{B}(X)$. A Região Negativa possui somente os elementos que com certeza não podem ser classificados como pertencentes à Aproximação Superior $\bar{B}(X)$, utilizando o conjunto de atributos B.

As regiões de X ficaram assim:

- Aproximação Inferior: $\underline{B}(X) \{e1, e6\}$
- Aproximação Superior: $\bar{B}(X) \{e1, e2, e3, e6\}$
- Região de Fronteira (Duvidosa): $RF(X) \{e2, e3\}$
- Região Negativa: $RN(X) \{e4, e5\}$

A Figura 4.4 ilustra todas as regiões de X em A.

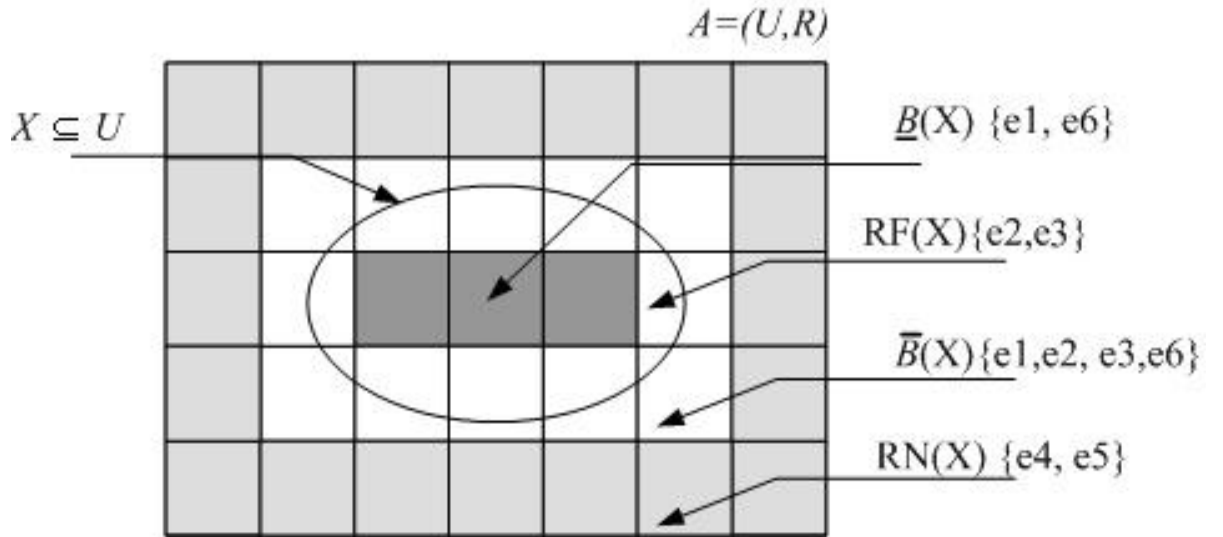


Figura 4.4: Todas as regiões de X em A.

Através dos conceitos de AS e AI, podem-se definir as quatro classes básicas de incerteza em RS [KOM99b].

Um conjunto definível em S é qualquer união finita de conjuntos elementares. O significado dessas classes, onde $-X$ denota $U - X$ (complemento) é o seguinte:

1. X é *rough* B-definível, se e somente se $\underline{B}(X) \neq \emptyset$ e $\bar{B}(X) \neq U$.

Se X é *rough* B-definível, isso significa que é possível decidir para alguns elementos de U quando eles pertencem a X e para alguns elementos de U quando eles pertencem a $-X$, utilizando B :

2. X é internamente B-indefinível, se e somente se $\underline{B}(X) = \emptyset$ e $\bar{B}(X) \neq U$.

Se X é internamente B-indefinível, isso significa que é possível decidir para alguns elementos de U quando eles pertencem a $-X$, mas não é possível decidir para nenhum elemento de U quando ele pertence a X , utilizando B ;

3. X é externamente B-indefinível, se e somente se $\underline{B}(X) \neq \emptyset$ e $\bar{B}(X) = U$.

Se X é externamente B-indefinível, isso significa que é possível decidir para alguns elementos de U quando eles pertencem a X , mas não é possível decidir para qualquer elemento de U quando ele pertence a $-X$, utilizando B ;

4. X é totalmente B-indefinível, se e somente se $\underline{B}(X) = \emptyset$ e $\bar{B}(X) = U$.

Se X é totalmente B-indefinível, isso significa que não é possível decidir para qualquer elemento de U quando ele pertence a X ou a $-X$, utilizando B .

4.2.3 Qualidade das Aproximações

A Qualidade das Aproximações obtidas pelas definições dadas previamente pode ser caracterizada numericamente a partir dos próprios elementos que as definem. O Coeficiente para medir essas Qualidades é representado por $\alpha_B(X)$, sendo X o conjunto de elementos com respeito a B , e podem ser realizadas de três formas [PAW82]:

- a) Coeficiente de Incerteza;
- b) Coeficiente de Qualidade da Aproximação Superior;
- c) Coeficiente de Qualidade da Aproximação Inferior.

As três medidas foram calculadas nos exemplos a seguir:

- a) O Coeficiente de Incerteza $\alpha_B(X)$, pode ser entendido como a Qualidade da Aproximação de X, dado por:

$$\alpha_B(X) = \frac{|\bar{B}(X)|}{|B(X)|} \quad (4.4)$$

em que $|B(X)|$ e $|\bar{B}(X)|$ denotam a cardinalidade das Aproximações Inferior e Superior, respectivamente. Obviamente $0 \leq \alpha_B \leq 1$. Se $\alpha_B(X) = 1$, X é *crisp* (preciso) em relação ao conjunto de atributos B. Se $\alpha_B(X) < 1$, X é *rough* (impreciso), em relação ao conjunto de atributos B.

O coeficiente $\alpha_B(X)$ pode ser interpretado como o percentual de todos os elementos possivelmente classificados como pertencentes a X, dado por:

$$\alpha_B(X) = \frac{|\bar{B}(X)|}{|B(X)|} = \frac{|\{e1, e6\}|}{|\{e1, e2, e3, e6\}|} = \frac{2}{4} = 0,5 \quad (4.5)$$

ou seja, com base nos elementos da Tabela 4.3, 50% de X é preciso com respeito a B.

- b) Coeficiente de Qualidade da Aproximação Superior; $\alpha_B(\bar{B}(X))$, que pode ser interpretado como sendo o percentual de todos os elementos possivelmente classificados como pertencentes a X, dado por:

$$\alpha_B(\bar{B}(X)) = \frac{|\bar{B}(X)|}{|U|} = \frac{|\{e1, e2, e3, e6\}|}{|\{e1, e2, e3, e4, e5, e6\}|} = \frac{4}{6} = 0,66 \quad (4.6)$$

ou seja, com base nos elementos da Tabela 4.3, 66% de U possivelmente pertence à X.

- c) Coeficiente de Qualidade da Aproximação Inferior $\alpha_B(\underline{B}(X))$, que pode ser interpretado como sendo o percentual de todos os elementos certamente classificados como pertencentes a X, dado por:

$$\alpha_B(B(X)) = \frac{|B(X)|}{|U|} = \frac{|\{e1, e6\}|}{|\{e1, e2, e3, e4, e5, e6\}|} = \frac{2}{6} = 0,33 \quad (4.7)$$

ou seja, com base nos elementos da Tabela 4.3, 33% de U certamente pertence à X.

4.3 Dependência de Atributos

RS pode ser utilizado para a análise de dependência entre atributos, visando principalmente a identificação e eliminação de atributos redundantes ou desnecessários. A eliminação de atributos redundantes permite que se encontre um subconjunto mínimo de atributos que possui o mesmo valor discriminatório do conjunto de atributos original [ZIA93,95].

Um conjunto de atributos D (decisão) depende totalmente de C (condicionais) ($C \Rightarrow D$), se todos os valores nominais de D forem univocamente determinados por valores nominais dos atributos de C. Em outras palavras, D depende totalmente de C se existir uma dependência funcional entre valores nominais de C e D. Se $(C,D) = 1$, diz-se que D depende totalmente de C; se $(C,D) = 0$, D não depende de C e se $0 < (C,D) < 1$, D depende parcialmente de C. [ZIA93]. Formalmente, a dependência pode ser assim definida: sejam C e D subconjuntos de S. O Grau de Dependência de D em relação a C é dado por:

$$y(C,D) = \frac{|B(X)|}{|U|} = \frac{|\{e1, e6\}|}{|\{e1, e2, e3, e4, e5, e6\}|} = \frac{2}{6} = 0,33 \quad (4.8)$$

Quando D depende parcialmente de C indica que não existe a necessidade da presença de todos os atributos condicionais de C para gerar os valores nominais do atributo de decisão D, abrindo espaço para uma redução de atributos. Do resultado de 0,33 (33%) deduz-se que dos 3 atributos condicionais pertencentes ao sistema de informação S (Experiência do Vendedor, Qualidade do Produto e Boa Localização), um deles (33%) pode ser reduzido sem que a base de dados perca a sua representatividade original e os outros dois (66%) apresentam dependência, e por

isso não podem ser reduzidos. A redução da base de dados 4.1 será apresentada na seção Redução de Atributos (4.5).

4.4 Significância de Atributos

Como foi visto na seção 4.3, alguns atributos são mais significativos (relevantes) que outros impossibilitando a sua redução.

Sejam C e D conjuntos de atributos condicionais e de decisão, respectivamente, e seja a um atributo condicional pertencente a C . A Significância de a será calculada em função da mudança do Grau de Dependência de D em relação a C (0,66) com a remoção de a , segundo a fórmula abaixo:

$$\sigma_{(P,O)}(a) = 1 - \left(\frac{(C - \{a\}, D)}{(C, D)} \right) \quad (4.9)$$

Este Coeficiente pode ser visto como o erro que ocorre na definição de D por C quando a é removido. Com base na Tabela 4.3 e utilizando os elementos com atributo de decisão igual a Lucro, temos: para $X = \{e1, e3, e6\}$, $D = \{\text{Retorno}\}$, $C = \{\text{Experiência do Vendedor, Qualidade do Produto, Boa Localização}\}$, o Grau de Significância do atributo a é dado por $1 - ((\{\text{Experiência do Vendedor, Qualidade do Produto}\})/(\{\text{Experiência do Vendedor, Qualidade do Produto, Boa Localização}\})) = 1 - ((2/3)/(2/3)) = 0$, significando que remover o atributo Boa Localização não afetará os resultados.

O mesmo resultado seria encontrado ao se remover Qualidade do Produto de $\{\text{Experiência do Vendedor, Qualidade do Produto, Boa Localização}\}$.

4.5 Redução de Atributos

A redução de atributos em RS é feita através dos chamados Redutos (RED), que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais [PAW82].

Um Reduto de B sobre um sistema de informação S é um conjunto de atributos $B' \subseteq B$ tal que todos os atributos $a \in (B - B')$ são dispensáveis. Com isso, $U/INDs(B') = U/INDs(B)$. O termo $RED(B)$ é utilizado para denotar a família de redutos de B .

O procedimento para a redução de atributos é o seguinte: comparam-se os conjuntos elementares de cada atributo individualmente e depois dois a dois, três a três, assim sucessivamente com os conjuntos elementares formados com todos os atributos $\{e1\}$; $\{e2, e3\}$; $\{e4\}$; $\{e5\}$; $\{e6\}$. Nesta comparação, quando um atributo apresentar os seus conjuntos elementares iguais aos conjuntos elementares formados com todos os atributos, este atributo pode ser reduzido.

O cálculo de reduções para gerar os redutos é um problema $n-p$ completo, e seu processamento em grandes bases de dados exige grande esforço computacional.

Desta forma, comparando cada atributo individualmente (os conjuntos elementares são formados pelo valor nominal do atributo de decisão) obtêm-se: Experiência do Vendedor: Lucro $\{e1, e3, e6\}$; Prejuízo $\{e2, e4, e5\}$; Qualidade do Produto: Lucro $\{e1, e3, e6\}$; Prejuízo $\{e2, e4, e5\}$; Boa Localização: Lucro $\{e1, e3, e6\}$; Prejuízo $\{e2, e4, e5\}$.

Nenhum atributo teve os seus conjuntos elementares iguais aos conjuntos elementares formados com todos os atributos, portanto, neste momento não houve redução.

Comparando os atributos: Experiência do Vendedor e Qualidade do Produto = Alta e Boa $\{e1\}$; Média e Boa $\{e2, e3\}$; Baixa e Média $\{e4\}$; Média e Média $\{e5\}$; Alta e Média $\{e6\}$. O resultado foi: $\{e1\}$; $\{e2, e3\}$; $\{e4\}$; $\{e5\}$; $\{e6\}$.

Os conjuntos elementares obtidos são iguais aos conjuntos elementares obtidos com todos os atributos, portanto, é possível reduzir um atributo.

Comparando os atributos: Experiência do Vendedor e Boa Localização = Alta e Não {e1}; Média e Não {e2, e3}; Baixa e Não {e4}; Média e Sim {e5}; Alta e Sim {e6}. O resultado dos conjuntos elementares foi: {e1}; {e2, e3}; {e4}; {e5}; {e6}.

Os conjuntos elementares obtidos são iguais aos conjuntos elementares obtidos com todos os atributos, portanto, existe mais de uma redução de atributo.

Comparando os atributos: Qualidade Produzida e Boa Localização = Boa e Não {e1, e2, e3}; Média e Não {e4}; Média e Sim {e5, e6}

O Resultado dos conjuntos elementares foi: {e1, e2, e3}; {e4}; {e5, e6}, mostrando-se diferente dos conjuntos elementares obtidos com todos os atributos e, portanto, não sendo possível nenhuma redução. Chega-se à conclusão de que existem dois redutos RED(B): {Experiência do Vendedor, Qualidade do Produto} e {Experiência do Vendedor, Boa Localização}, ou seja, utilizando qualquer um dos dois redutos consegue-se representar a base de dados com todos os seus atributos originais. Pode-se visualizar na Tabela 4.4, o sistema de informação S sem a presença do atributo reduzido Boa Localização.

Loja	Experiência do Vendedor	Qualidade do Produto
E1	Alta	Boa
E2	Média	Boa
E3	Média	Boa
E4	Baixa	Média
E5	Média	Média
E6	Alta	Média

Tabela 4.4: Sistema de Informação S sem o atributo Boa Localização.

Pode-se visualizar na Tabela 4.5, o sistema de informação S sem a presença do atributo reduzido Qualidade do Produto.

Loja	Experiência do Vendedor	Boa Localização
e1	Alta	Não
e2	Média	Não
e3	Média	Não
e4	Baixa	Não
e5	Média	Sim
e6	Alta	Sim

Tabela 4.5: Sistema de Informação S sem o atributo Qualidade do Produto.

4.6 Matriz de Discernibilidade

Na seção 4.5 foi mostrado como se realiza o procedimento de redução de atributos. Essa redução é feita pela função de discernibilidade, a partir da Matriz de Discernibilidade. Considerando o conjunto de atributos $B = \{\text{Experiência do Vendedor, Qualidade do Produto e Boa Localização}\}$ para o sistema de informação S, o conjunto de todas as classes de equivalência determinadas por B sobre S é dado por $U/\text{INDs}(B) = \{\{e1\}; \{e2, e3\}; \{e4\}; \{e5\}; \{e6\}\}$, que estão representadas na Tabela 4.3.

A Matriz de Discernibilidade do sistema de informação S, denotada por $MD(B)$, é uma matriz simétrica $n \times n$ com: $md(i, j) = \{a \in B \mid a(E_i) \neq a(E_j)\}$ para $i, j = 1, 2, \dots, n$, sendo $1 \leq i, j \leq n$ e $n = |U / \text{INDs}(B)|$. Logo, os elementos da matriz de discernibilidade $md(i, j)$ é o conjunto de atributos condicionais de B que diferenciam os elementos das classes com relação aos seus valores nominais.

Considerando Experiência do Vendedor (EV), Qualidade do Produto (QP) e Boa Localização (BL), com a finalidade de construir a Matriz de Discernibilidade $MD(B)$, tem-se na Tabela 4.6 a sua representação:

	e1	e2	e3	e4	e5	e6
e1	\emptyset					
e2	EV	\emptyset				
e3	EV	\emptyset	\emptyset			
e4	EV, QP	EV, QP	EV, QP	\emptyset		
e5	EV, QP, BL	QP, BL	QP, BL	EV, BL	\emptyset	
e6	QP, BL	EV, QP, BL	EV, QP, BL	EV, BL	EV	\emptyset

Tabela 4.6: Matriz de Discernibilidade.

4.7 Função de Discernibilidade

A função de discernibilidade $F_S(B)$ é uma função booleana com m variáveis, que determina o conjunto mínimo de atributos necessários para diferenciar qualquer classe de equivalência das demais, definida como:

$$F_S(\mathbf{a}_1^*, \mathbf{a}_2^*, \dots, \mathbf{a}_m^*) = \bigwedge \left\{ \bigvee m_D^*(i, j) \mid i, j = 1, 2, \dots, n, \quad m_D(i, j) \neq \emptyset \right\}.$$

Sendo

$$m_D^*(i, j) = \left\{ \mathbf{a}^* \mid \mathbf{a} \in m_D(i, j) \right\}$$

Utilizando o método de simplificação de expressões booleanas na função $F_S(B)$, obtém-se o conjunto de todos os implicantes primos dessa função, o qual determina os redutos de S . A simplificação é um processo de manipulação algébrica das funções lógicas com a finalidade de reduzir o número de variáveis e de operações necessárias para a sua realização [PAT05].

A função de discernibilidade $F_S(B)$ é obtida da seguinte forma: para os atributos contidos dentro de cada célula da Matriz de Discernibilidade $MD(B)$ (Tabela 4.6), aplica-se o operador “soma”, “or” ou “ \vee ” e, entre as células dessa matriz, utiliza-se o operador “produto”, “and” ou “ \wedge ”, resultando em uma expressão booleana de “Produto da Soma”. A $F_S(B)$ da Tabela 4.6 é representada por:

$$F_s(B) = (EV) \wedge (EV) \wedge (EV \vee QP) \wedge (EV \vee QP) \wedge (EV \vee QP) \wedge (EV \vee QP \vee BL) \wedge \\ (QP \vee BL) \wedge (QP \vee BL) \wedge (EV \vee BL) \wedge (QP \vee BL) \wedge (EV \vee QP \vee BL) \wedge (EV \vee QP \\ \vee BL) \wedge (EV \vee BL) \wedge (EV)$$

Simplificando esta expressão, utilizando teoremas, propriedades e postulados da Álgebra Booleana, obtém-se a seguinte expressão minimizada:

$F_s(B) = (EV \wedge (QP \vee BL) \wedge (EV \vee QP \vee BL))$, que ainda pode ser escrita na forma de “Soma do Produto”, ou seja $F_s(B) = (EV \wedge (QP \vee BL))$. Confirmando o que já foi verificado na seção 4.5, os redutos são $RED(B) = \{\text{Experiência do Vendedor, Qualidade do Produto}\}$ e $\{\text{Experiência do Vendedor, Boa Localização}\}$.

A função de discernibilidade determinou o termo mínimo da função, ou seja, determinou o conjunto mínimo de atributos necessários para discernir as classes formadas por todas as classes de equivalência da relação $INDs(B)$.

Estudos mais aprofundados sobre Matriz de Discernibilidade podem ser encontrados em Bozi [BOZ01], Felix [FEL99], Fujimori [FUJ98] e Slowinski [SLO92].

4.8 Regras de Decisão

Reduzir atributos é a função de RS na Arquitetura Híbrida proposta. Entretanto, RS pode gerar regras de decisão para serem usadas em classificação [SAS06]. A seguir estão descritas, em caráter ilustrativo, algumas regras geradas pelo RS a partir dos conceitos apresentados neste capítulo:

Se Experiência do Vendedor = Alta e Qualidade do Produto = Boa **então** Retorno = Lucro.

Se Experiência do Vendedor = Média e Qualidade do Produto = Boa **então** Retorno = Lucro.

Se Experiência do Vendedor = Baixa e Qualidade do Produto = Média **então** Retorno = Prejuízo.

Se Experiência do Vendedor = Alta e Qualidade do Produto = Média **então** Retorno = Lucro.

Em RS, a credibilidade de cada regra gerada é associada a um Fator de Credibilidade (FC) através de uma função de pertinência. Com isso, uma regra consistente receberá um Fator de Credibilidade igual a 1, enquanto regras não tão consistentes receberão um Fator de Credibilidade menor que 1. Quanto mais próximo de 1 é o Fator de Credibilidade, maior é a credibilidade da regra. Maior aprofundamento no assunto regras de decisão pode ser obtido em Pawlak. [PAW96b].

4.9 Resumo do Capítulo

Neste capítulo foram abordados os principais conceitos dos RS e suas aplicações no tratamento de um tipo fundamental de incerteza, a Indiscernibilidade.

O estudo da redução de dados contribui para a fundamentação e o entendimento dos RS na Arquitetura Híbrida proposta.

Desde o aparecimento de RS, muitos trabalhos vêm sendo gerados para conhecer melhor a Teoria e suas aplicações, como: utilização dos conceitos de Aproximações Superior e de Aproximação Inferior para aproximar atributos desconhecidos com base em atributos conhecidos, elaboração de relações e operações sobre conjuntos *rough* (imprecisos), similares àsquelas utilizadas em conjuntos clássicos [BON98], medidas de incerteza baseadas em teoria da informação [BEA98] e aplicação de RS em modelos de bancos de dados relacionais [CHE96, HU94, 95, LIN97].

O próximo capítulo trata das redes neurais artificiais e, mais especificamente, da rede SOM, que é a outra técnica combinada ao RS para formar a Arquitetura Híbrida proposta.

5 Rede SOM (*Self-Organizing Maps* – Mapas Auto-Organizáveis)

Este capítulo apresenta o histórico e o conceitual básico de redes neurais artificiais, enfatizando a rede SOM, que é a arquitetura de rede neural artificial utilizada neste trabalho.

5.1 Introdução

Os modelos neurais artificiais procuram aproximar o processamento dos computadores ao cérebro humano.

A partir do momento em que as máquinas começaram a evoluir, um grande desejo do homem tem sido a criação de uma máquina que possa operar independentemente do controle humano, cuja independência seja desenvolvida de acordo com seu próprio aprendizado e que tenha a capacidade de interagir com ambientes incertos (desconhecidos por ela) [CAR91]. Uma máquina que possa ser chamada de autônoma e inteligente ou cognitiva. O sucesso dessa máquina autônoma e inteligente dependeria única e exclusivamente da sua capacidade de lidar com uma variedade de eventos inesperados no ambiente em que opera. Essas máquinas teriam maior capacidade de aprender tarefas de alto nível cognitivo que não são facilmente manipuladas por máquinas atuais e continuariam a se adaptar e realizar tais tarefas gradativamente com maior eficiência, mesmo em condições de ambiente imprevisíveis. Então, seriam muito úteis naqueles casos em que a interação humana é perigosa, tediosa ou impossível, como em reatores nucleares, combate ao fogo, operações militares, exploração do espaço a distâncias em que a nave espacial estaria fora do alcance do controle na terra, porém enviando informações [SMI90].

O funcionamento dos computadores é feito de modo seqüencial, proporcionando maior eficiência na resolução de tarefas que devem ser executadas seguindo determinadas etapas. Já o cérebro humano funciona de modo paralelo, e sendo extremamente conectado é mais eficiente na resolução de tarefas que exigem a utilização de diversas variáveis.

Assim, podemos dizer que as redes neurais artificiais são redes inspiradas na estrutura do cérebro humano, com o objetivo de apresentar características similares ao comportamento humano, tais como: aprendizado, associação, generalização e abstração.

Uma rede neural artificial é um processador maciçamente paralelo, distribuído, constituído de unidades de processamento simples que têm capacidade para armazenar conhecimento experimental e torná-lo disponível para uso [HAY99].

5.2 Um Breve Histórico sobre Redes Neurais Artificiais (RNAs)

As pesquisas sobre redes neurais artificiais tiveram início em 1943 com Warren Mcculloch e Walter Pitts. A estrutura proposta abstraía a complexidade da atividade neural em sistemas neurais reais e estabeleceu a base da neurocomputação, concebendo procedimentos matemáticos análogos ao funcionamento dos neurônios biológicos [KOV96].

A contribuição de Warren Mcculloch e Walter Pitts foi puramente conceitual, buscando mais descrever um modelo artificial de um neurônio e apresentar suas capacidades computacionais do que apresentar técnicas de aprendizado, sendo tomada como base para a maioria dos modelos connexionistas desenvolvidos posteriormente.

Em 1949 Donald Hebb publica *The Organization of Behavior* e dá um passo muito importante na história das redes neurais artificiais ao propor um modelo segundo o qual se assumiu que a aprendizagem do conhecimento representado em

uma rede neural seja alcançada pelo fortalecimento das conexões entre os neurônios adjacentes (vizinhos, próximos), sempre que estes estiverem excitados.

Em 1958, Frank Rosenblatt criou o *perceptron*, um modelo cognitivo que consistia de unidades sensoriais conectadas a uma única camada de neurônios de Mcculloch e Pitts. Assim, Rosenblatt demonstrou que, se acrescidas sinapses ajustáveis às redes neurais artificiais de Mcculloch e Pitts, estas poderiam ser treinadas para classificar padrões de classes linearmente separáveis.

Com base no modelo de Mcculloch e Pitts, no início da década de 1960 Widrow e Hoff publicaram um artigo sobre o neurônio artificial, denominado *Adaline* (ADaptive LINear Element). A contribuição desse trabalho está associada à regra de aprendizagem proposta, a regra Delta.

Em 1969 Minsky e Papert, na publicação *Perceptrons*, expuseram as limitações do modelo de Rosenblatt, provando que tais redes não são capazes de resolver uma ampla gama de problemas devido às restrições de representação.

Nessa época, pensava-se que para reproduzir o comportamento do cérebro humano bastaria construir uma rede neural suficientemente grande. Uma rígida análise matemática, no entanto, comprovou o pouco poder computacional dos modelos de redes neurais artificiais utilizados na época, levando as pesquisas neste campo a ficarem esquecidas de meados dos anos 60 até o início da década de 80 do século passado [BRA00].

Nos início dos anos 80, o interesse pela área retornou devido, em grande parte, ao surgimento de novos modelos neurais, como os propostos por Hopfield e por Teuvo Kohonen. Além disso, nessa mesma época, ocorreu o desenvolvimento de computadores mais rápidos e poderosos, facilitando a implementação das redes neurais artificiais.

O modelo conexionista (baseado em redes neurais) proposto por John Hopfield em 1982 permitiu, pelas suas características computacionais e estabilidade, esclarecer boa parte das dúvidas até então existentes em relação ao processo dinâmico executado por certas redes neurais artificiais. Em 1984 Teuvo Kohonen

publica um artigo descrevendo a rede neural artificial baseada em auto-organização e nas características de aprendizado adaptativo do cérebro humano.

Cerca de dois anos mais tarde, Rumelhart, Hinton e Williams aperfeiçoaram a idéia de *perceptron*, criando o algoritmo de retropropagação do erro (*error back-propagation*), levando a uma explosão de interesse em redes neurais artificiais. O sucesso deste algoritmo estimulou o desenvolvimento de muitas pesquisas em redes neurais artificiais e de uma ampla variedade de modelos cognitivos [HAY99].

5.3 Uma Breve Comparação entre o Neurônio Biológico e o Neurônio Artificial

A Figura 5.1 ilustra uma comparação entre o neurônio biológico (parte a) e o neurônio artificial (parte b). Pode-se perceber a semelhança entre as duas estruturas. Analogamente, as entradas de um neurônio artificial representam os dendritos do neurônio biológico, a junção aditiva no neurônio artificial representa o corpo celular no neurônio biológico e a saída do neurônio artificial representa o axônio no neurônio biológico.

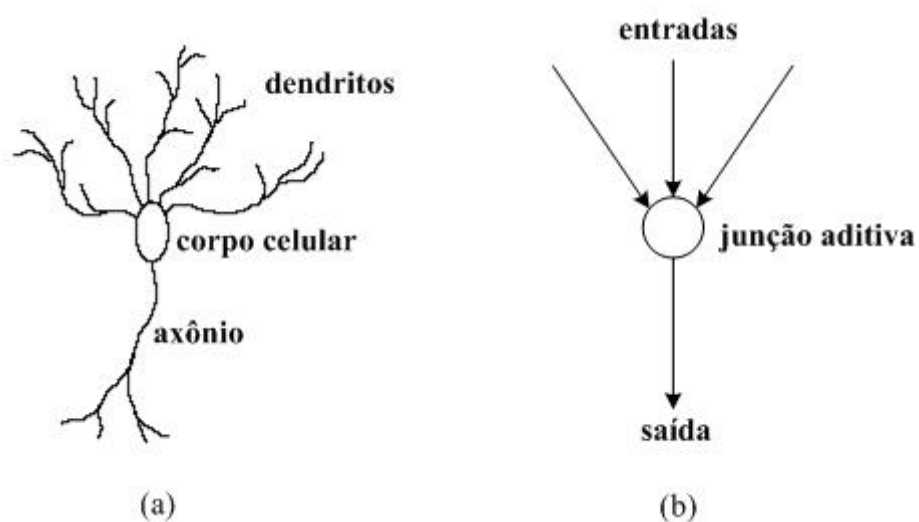


Figura 5.1 Comparação entre o neurônio biológico (parte a) e o neurônio artificial (parte b)

Em suma, uma rede neural artificial é uma tradução simplificada do funcionamento do cérebro humano e do sistema nervoso, a partir de algoritmos. As redes neurais artificiais caracterizam-se por possuírem: elevado número de elementos de processamento simples inspirados no funcionamento do neurônio biológico; conexões entre os elementos de processamento; em cada conexão um peso associado, que representa quão forte é a interação ou acoplamento entre os elementos de processamento e se a sua natureza é excitatória ou inibitória [BRA00].

5.4 Aplicações das Redes Neurais Artificiais

Por apresentar certas características vantajosas (descritas no capítulo 1), as redes neurais artificiais vêm sendo utilizadas em diversas aplicações, como: reconhecimento de padrões [TOD95]; reconhecimento e classificação de imagens (textos, assinaturas, impressões digitais e objetos) [DUD00]; previsão de séries temporais [POR95]; monitoramento e controle [AZE00]; diagnóstico médico [BAX90]; em áreas como Economia [BEL96; BEN03; CHE02; FRE99], Finanças [ALM, 93; 96; 98; ARR01; BEL96; HAW90; PAR03; TRI96], *Marketing* [BIG96; BER97; HAN01; DHA97; HAW90; ROC90]; e como técnicas do processo de KDD [ARR01; BER97; DHA97; HAN01].

5.4.1 Aplicação das Redes Neurais Artificiais nos Negócios

As RNAs são muito utilizadas na área de negócios por organizações que buscam cada vez mais conhecimento para poderem sobreviver em um ambiente extremamente complexo, onde a palavra de ordem é mudança (capítulo 2).

Essas organizações, além de relacionarem-se internamente com agentes instáveis, como funcionários e a sua própria estrutura organizacional, devem relacionar-se com consumidores, concorrentes, sistema financeiro, comunidade e sindicatos, entre outros. É grande a necessidade de buscar informações além de suas

fronteiras, para poder de certa forma reagir com mais rapidez e segurança às mudanças ambientais.

O uso da Tecnologia da Informação (capítulo 2) auxilia em muito a organização na busca dessas informações competitivas, e a aplicação de redes neurais artificiais vem ajudar na análise da grande quantidade de informações que a organização possui.

A Tabela 5.1 mostra a relação entre o tipo de organização, a área de aplicação e a tarefa executada pela RNA [BIG96]. O estudo de clusterização é alvo deste trabalho, por isso, as aplicações nas organizações que utilizam esta tarefa estão destacadas em negrito.

Área de Aplicação	Tipo de Organização	Tarefa
Banco de Dados de <i>Marketing</i>	Todas	Clusterização , Classificação e Associação
Gerenciamento de Relações de Clientes	Todas	Clusterização , Classificação e Associação
Deteção de Fraudes	Financeira e Saúde	Classificação e Associação
Reconhecimento de Caracteres Ópticos	Financeira e Varejo	Classificação
Reconhecimento de Caracteres de Caligrafia	Informática e Financeira	Clusterização e Classificação
Previsão de Vendas e Controle de Inventários	Manufatureira, Vendas e Distribuição	Clusterização
Controle de Processos	Manufatureira e Processos	Associação
Predição de Demanda de Energia Elétrica	Manufatureira	Clusterização

Tabela 5.1: Aplicações de RNAs nas Organizações. Em negrito as aplicações que utilizam clusterização (agrupamento). Adaptado de Bigus [BIG96]

As redes neurais artificiais são utilizadas nos Negócios como uma técnica para extrair conhecimento das bases de dados.

5.5 Tarefas Básicas das Redes Neurais Artificiais em KDD

As RNAs podem realizar muitas tarefas relacionadas ao KDD como: classificação, clusterização (agrupamento) e associação, dentre outras. Por isso, as RNA são muito utilizadas quando o objetivo é a descoberta de conhecimento em dados.

A seguir, serão apresentadas brevemente algumas das tarefas em que os algoritmos baseados em RNA podem ser utilizados, pois o assunto tarefas do KDD foi aprofundado no capítulo 3.

5.5.1 Classificação

Uma RNA pode ser treinada para, a partir de exemplos de uma mesma classe ou de um conjunto de classes com características em comum, separar os elementos determinando se algum registro pertence a uma classe ou não.

A RNA pode, por exemplo, classificar os itens comprados pelas pessoas. Essa classificação pode ser utilizada para verificar a relação entre itens de compra e fornecer para a empresa o perfil dos seus clientes, além de definir novas estratégias de negócios, realizar promoções, determinar a necessidade de aquisição de novos produtos para vender ou até trocar de atividade comercial. Enfim, a utilização de RNAs para fazer classificações pode ser uma ferramenta muito útil para uma empresa gerenciar o seu relacionamento com os clientes [BER97].

5.5.2 Clusterização (Agrupamento)

Uma RNA pode ser treinada para agrupar diversos objetos similares, por exemplo, itens de compras de clientes, categorias de clientes, faixa etária de clientes, entre

outros. É muito útil para uma empresa obter informações sobre a categoria de clientes que ela possui. Uma categoria pode ser entendida como classes sociais, faixas etárias e períodos de compra dos clientes. Uma empresa pode obter informações de seu interesse, como, por exemplo, que noventa por cento dos seus clientes fazem parte da classe média; ou que setenta e cinco por cento dos seus clientes são mulheres com idade entre vinte e cinco e trinta e oito anos; ou que oitenta por cento das vendas feitas pela empresa ocorrem entre o dia cinco e o dia quinze de cada mês e aos sábados à tarde. Todos estes exemplos de informações podem determinar novas estratégias de vendas de uma empresa [DAL03]. Neste trabalho, a tarefa de clusterização é de especial importância, pois, tanto a rede SOM (primeira fase dos experimentos) como a Arquitetura Híbrida proposta (segunda fase dos experimentos) têm como função a geração de *clusters*.

5.5.3 Associação

O cérebro humano é como um dispositivo de armazenamento e, ao longo de toda a sua existência, a memória humana armazena informações de forma similar a um banco de dados. As pessoas armazenam informações e fazem associações de fatos, idéias e de dados com outras memórias relacionadas.

Uma RNA pode ser treinada, por exemplo para descobrir associações entre itens de compras feitas pelos consumidores de um supermercado e, a partir desta informação, definir novas estratégias de negócios [BER97].

Uma RNA pode também realizar outras tarefas, como: Regressão, Previsão de Séries Temporais e Predição [GOL05].

5.6 Treinamento (Aprendizado) em uma Rede Neural Artificial (RNA)

Em uma rede neural artificial os neurônios estão dispostos em uma ou mais camadas e interligados por grande número de conexões, geralmente unidirecionais.

Na maioria dos modelos, essas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede [BRA00].

As redes neurais artificiais possuem a capacidade de aprender por exemplos e fazer interpolações e extrapolações do que aprenderam. Um conjunto de procedimentos bem definidos para adaptar os pesos de uma rede neural artificial para que ela possa aprender uma determinada função é chamado algoritmo de aprendizado [BIG96].

A Figura 5.2 ilustra um tipo de arquitetura de rede neural artificial muito utilizada chamada *MultiLayer Perceptron* (MLP).

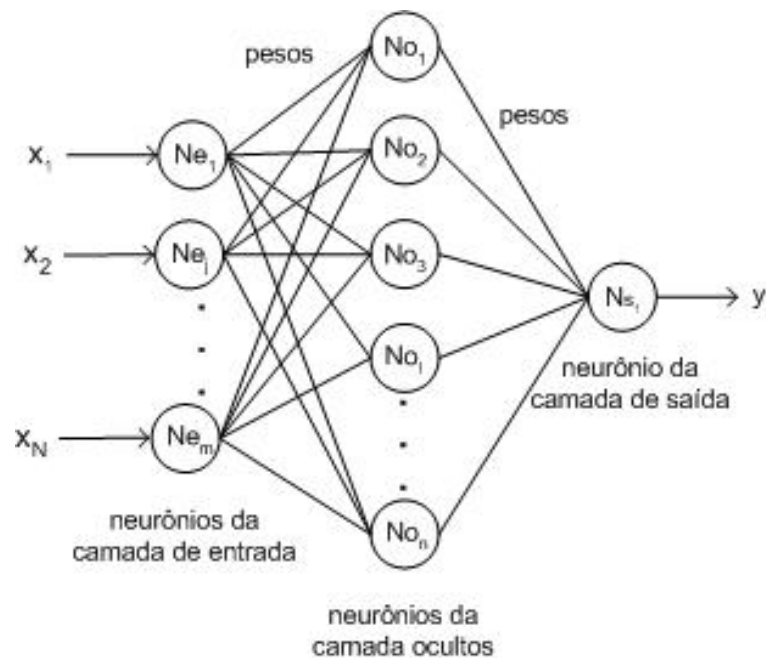


Figura 5.2: Arquitetura de uma rede neural artificial *MultiLayer Perceptron* (MLP).

Na Figura 5.2 pode-se observar os dados (vetores de dados) de entrada da rede (X_1, \dots, X_N), os neurônios da camada de entrada da rede (Ne_1, \dots, Ne_m) com seus respectivos pesos, os neurônios que formam a camada intermediária da rede (No_1, \dots, No_n) e a camada de saída (Ns_1), formada por um neurônio.

Uma definição geral do que vem a ser aprendizado em uma RNA pode ser expressa da seguinte maneira: “Aprendizado é o processo pelo qual os parâmetros de uma rede neural artificial são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando, sendo o tipo específico de aprendizagem realizada definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros” [MEN70].

Diversos métodos para aprendizado foram desenvolvidos, podendo ser agrupados em dois paradigmas principais: aprendizado supervisionado e aprendizado não-supervisionado.

5.6.1 Aprendizado Supervisionado

O treinamento supervisionado necessita de um par de vetores, composto do vetor de entrada e do vetor alvo que se deseja como saída. Juntos, estes vetores são chamados de par de treinamento ou vetor de treinamento, sendo interessante ressaltar que geralmente a rede é treinada com vários pares de treinamento.

O procedimento de treinamento funciona da seguinte forma: o vetor de entrada é aplicado, a saída da rede é calculada e comparada com o correspondente vetor alvo. O erro encontrado é então realimentado através da rede, e os pesos são atualizados de acordo com um algoritmo determinado a fim de minimizar este erro. Este processo de treinamento é repetido até que o erro para os vetores de treinamento tenha alcançado níveis bem baixos. O algoritmo mais difundido é o de retropropagação do erro (*error back-propagation*) utilizado pela rede *MultiLayer Perceptron* (Figura 5.2) [HAY99].

5.6.2 **Aprendizado Não-supervisionado**

O treinamento não-supervisionado, por sua vez, não requer vetor alvo para as saídas e, obviamente, não faz comparações para determinar a resposta ideal. O conjunto de treinamento modifica os pesos da rede de forma a produzir saídas que sejam consistentes, isto é, tanto a apresentação de um dos vetores de treinamento como a apresentação de um vetor que é suficientemente similar irão produzir o mesmo padrão nas saídas. No caso da rede SOM, o processo de treinamento extrai as propriedades estatísticas do conjunto de treinamento e agrupa os vetores similares em classes. A aplicação de um vetor de uma determinada classe à entrada da rede irá produzir um vetor de saída específico, mas não existe maneira de se determinar, antes do treinamento, qual o padrão que será produzido na saída para um vetor de entrada de uma determinada classe.

Este aprendizado não possui tempo determinado de execução. Algumas redes neurais artificiais requerem apenas uma passagem através dos dados, enquanto outras podem requerer centenas ou milhares de execuções sobre a mesma base de dados. O aprendizado de uma rede neural artificial pode ser controlado por um conjunto de parâmetros ou estes mesmos parâmetros podem servir para o ajuste dos pesos da RNA.

O aprendizado de uma RNA, na maioria dos casos, acontece com um subconjunto de exemplos (vetores de dados) que definem o chamado conjunto de treinamento, e o teste da RNA é realizado com um outro subconjunto de exemplos (vetores de dados) que definem o chamado conjunto de teste. A Tabela 5.2 apresenta alguns parâmetros adotados no aprendizado das RNAs.

Parâmetro	Arquiteturas de RNAs	Função
Taxa de Aprendizado	Todas	Controla a velocidade do ajuste dos pesos da rede neural artificial.
<i>Momentum</i>	<i>Back-Propagation</i> com MLP	Acompanha passo a passo o parâmetro taxa de aprendizado. Sua função é monitorar a possível ocorrência de oscilações grandes nos valores nas conexões de peso.
Tolerância de Erro	<i>Back-Propagation</i> com MLP	É utilizado para especificar a máxima diferença entre a saída desejada e saída gerada pela RNA. Este parâmetro especifica quão próxima a saída da RNA deve estar da saída desejada.
Função de Ativação	Todos	Seleciona a função de ativação, a qual é utilizada pela unidade de processamento da rede neural artificial. Podem ser utilizadas as funções <i>sigmóide</i> , tangente hiperbólica, função <i>gaussiana</i> , etc.
Vizinhança	Rede SOM	Na rede SOM, uma área é definida ao redor do neurônio vencedor, onde os neurônios não vencedores têm seus pesos alterados em algum grau.
Número de Épocas	Rede SOM, MLP	O número de épocas determina o número de passos para o treinamento da rede através do treinamento dos dados. Uma época ocorre ao final da apresentação de todos os itens de dados exatamente uma vez.

Tabela 5.2: Parâmetros de Aprendizado (treinamento) para arquiteturas de rede neural artificial. Adaptado de Bigus [BIG96]

As RNAs podem ser treinadas utilizando-se valores iniciais aleatórios para as conexões de pesos. Os parâmetros de aprendizado (Tabela 5.2) são inicializados e os padrões de treinamento dos vetores de dados são apresentados para a RNA. Ao longo do progresso do treinamento são ajustadas as conexões de pesos e é possível monitorar a *performance* da RNA.

5.7 Características da Rede SOM

A rede SOM (*Self-Organizing Maps*), ou Mapas Auto-Organizáveis, foi desenvolvida pelo finlandês Teuvo Kohonen em 1984.

Em vários trabalhos, esta rede é chamada de rede de Kohonen, o que pode provocar certa confusão, uma vez que Kohonen também propôs outros modelos de redes neurais artificiais [CAR91]. Neste trabalho serão chamadas simplesmente de rede SOM.

Em contraste com outros modelos, estas redes possuem uma forte inspiração neurofisiológica. Elas são baseadas no mapa topológico presente no córtex cerebral [KOH82].

O desenvolvimento da rede SOM como mapa neural foi idealizado a partir de uma característica do cérebro humano, que é organizado em muitas regiões, de tal forma que entradas sensoriais distintas são representadas por mapas computacionais topologicamente ordenados. Por exemplo, entradas sensoriais tácteis, visuais (o mapeamento é definido pelas características visuais primitivas, como intensidade de luz, orientação e curvatura de linhas) e acústicas (o mapeamento reflete as diferentes frequências sonoras) são mapeadas em diferentes áreas do córtex cerebral, de uma forma topologicamente ordenada [KOH97a].

Esses neurônios estão espacialmente ordenados dentro destas áreas e, assim, neurônios topologicamente próximos tendem a responder a padrões ou estímulos semelhantes.

5.7.1 Estrutura da Rede SOM

A rede SOM é um tipo de rede neural artificial baseada em aprendizado não-supervisionado, sendo capaz de mapear um conjunto de dados, de um espaço de entrada multidimensional, num conjunto finito de neurônios organizados em um arranjo normalmente unidimensional ou bidimensional.

Os nós da rede SOM estão localmente interconectados por uma relação de vizinhança, determinando a topologia do mapa. Num mapa bidimensional, a vizinhança pode ser hexagonal ou retangular. A Figura 5.3 exibe um arranjo hexagonal.

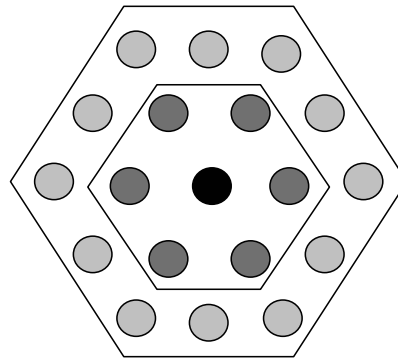


Figura 5.3: Arranjo hexagonal para uma rede SOM com grade bidimensional. Adaptado de Kohonen [KOH97a].

Na Figura 5.3 pode-se observar no centro do arranjo o neurônio vencedor (de cor preta) e sua vizinhança imediata com seis vizinhos. O formato do arranjo influencia diretamente a adaptação da rede SOM, sendo que o modelo hexagonal oferece tradicionalmente resultados melhores que o retangular [KOH97a]. Os conceitos de neurônio vencedor (competição) e adaptação serão tratados na seção 5.8.2.1.

Esta estrutura possibilita à rede SOM algumas vantagens como robustez, ajuste local do mapa e visualização [DUT01], o que faz com que seja utilizada nas mais diversas áreas, que vão da medicina [LO97] ao *marketing* [RUS97]. No caso de KDD, a rede SOM é muito utilizada [HSU99; VES00b] por possibilitar em um mapa de uma ou duas dimensões uma visualização simples dos *clusters* e da correlação dos dados, preservando a posição relativa dos *clusters* no hiperespaço original.

A idéia básica das técnicas de agrupamento é buscar regiões de alta densidade de pontos no espaço de entrada, os agrupamentos (*clusters*) de dados, e escolher os seus centros para representar os pontos das regiões (Figura 5.3 e Figura 5.7). No caso da rede SOM, a definição de relação de vizinhança não é simétrica nos extremos do mapa, fato que implica uma estimativa de densidade diferente nas regiões da borda comparadas com as unidades do centro do mapa, prejudicando a definição de borda (fronteira) entre os *clusters*, o que é uma desvantagem da rede SOM. Essa desvantagem será analisada com mais detalhes no capítulo 6 (experimentos).

5.7.2 O Algoritmo de Aprendizado da Rede SOM

O algoritmo de aprendizado da rede SOM para gerar os *clusters* é composto por três fases: competição, adaptação e cooperação.

5.7.2.1 Fases da Rede SOM

Para cada dado apresentado à rede haverá uma competição entre todos os neurônios pelo direito de representá-lo. Vence a competição o neurônio que tiver o vetor de pesos mais próximo do vetor de dados.

O algoritmo da rede SOM utiliza o chamado aprendizado competitivo. Uma rede neural com aprendizado competitivo é uma rede de uma única ou duas camadas (uni ou bidimensional) em que todos os neurônios recebem a mesma entrada. Cada neurônio computa o seu nível de ativação multiplicando o seu vetor de pesos pelo vetor de dados da entrada. O neurônio que tiver o maior nível de ativação (equação 5.1) é chamado de neurônio vencedor ou de BMU (*Best Match Unit*), ou seja, o padrão de entrada que estiver sendo apresentado à rede provocará a ativação de apenas um neurônio de saída ou um neurônio por grupo. A Figura 5.4 mostra o aprendizado competitivo em uma rede SOM bidimensional com 16 neurônios.

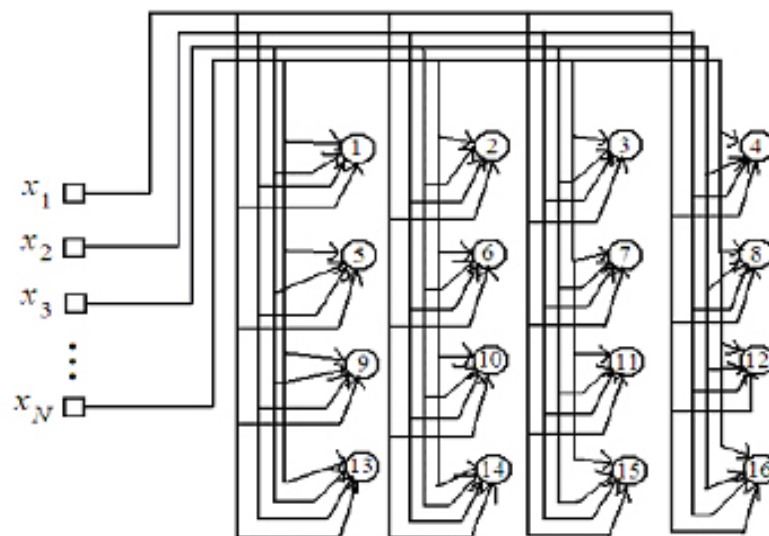


Figura 5.4: Rede SOM com aprendizado competitivo [KOH97a]

Na rede SOM da Figura 5.4, pode-se visualizar que a rede possui M neurônios organizados em uma grade bidimensional, nesse caso, $M = 16$. Os padrões de entrada são vetores N dimensionais (x_1, \dots, x_N).

Quando um dos P padrões de entrada é apresentado à rede SOM, cada um dos M neurônios recebe esse padrão e calcula o seu nível de ativação através da equação abaixo,

$$u_i = \sum_{k=1}^N w_{ik} x_k, \quad i=1, \dots, M, \quad (5.1)$$

onde x é o vetor de entrada, i é o índice que indica o neurônio e w_i é o vetor de pesos entre o padrão de entrada e o neurônio i (os pesos não estão mostrados na equação para não congestioná-la).

O neurônio é vencedor quando o padrão x que estiver sendo apresentado é aquele que tiver o maior valor de u_i , neste trabalho denominado i^* (i asterisco).

O único neurônio ativo (o BMU) em resposta ao padrão x representa o padrão de entrada. É como se os M neurônios da rede competissem entre si para determinar qual vai ficar mais ativo em resposta ao padrão de entrada e apenas o vencedor permanecesse ativo, deixando todos os perdedores inativos. Por causa disso, este tipo de rede é também chamado de rede do tipo “o vencedor fica com tudo” (*winner-takes-all*)

Uma rede do tipo o-vencedor-fica-com-tudo implementa um mapa entre um espaço N -dimensional contínuo de vetores x e um espaço discreto de M neurônios. Note que podemos ter mais de um vetor x sendo representado pelo mesmo neurônio vencedor. Neste caso, este neurônio é o representante do grupo de padrões x que o fazem ser vencedor. O neurônio vencedor é então adaptado (*winner-takes-all*). O seu vetor de pesos é alterado no sentido de representar ainda mais o dado apresentado, aumentando a probabilidade de que este mesmo neurônio volte a vencer numa próxima apresentação do mesmo dado, ou dado semelhante.

Chamando o vetor de pesos do neurônio vencedor i^* de w_{i^*} , a regra do aprendizado competitivo utilizada para a adaptação do BMU é a seguinte:

$$w_{i^*}(n+1) = w_{i^*}(n) + \eta(x(n) - w_{i^*}(n))$$

e

$$w_i(n+1) = w_i(n) \text{ para } i \neq i^*, \quad (5.2)$$

onde η é a constante de taxa de aprendizagem (um valor entre 0 e 1) que controla a rapidez com que são feitas as mudanças nos pesos.

Para entender melhor a regra do aprendizado competitivo analisa-se a mesma vetorialmente. O neurônio vencedor é aquele que tiver o maior nível de ativação u_i . Como o nível de ativação de um neurônio i é o produto entre o vetor de entrada x e o vetor de pesos do neurônio i , como demonstrado na equação 5.3,

$$u_i = w_i \cdot x \quad (5.3)$$

o critério para a escolha do neurônio vencedor é o de similaridade entre x e w_i . Este produto pode ser escrito como na equação 5.4.

$$u_i = |w_i| |x| \cos \theta, \quad (5.4)$$

onde θ é o ângulo entre os vetores x e w_i .

Se os vetores de entrada e de pesos estiverem normalizados, um valor grande de u_i indica que o vetor de entrada x está próximo do vetor de pesos w_i , ou seja, que x está nas vizinhanças de w_i . Já um valor pequeno de u_i indica que o vetor de entrada x é quase perpendicular ao vetor de pesos w_i , como pode ser visualizado na Figura 5.5.

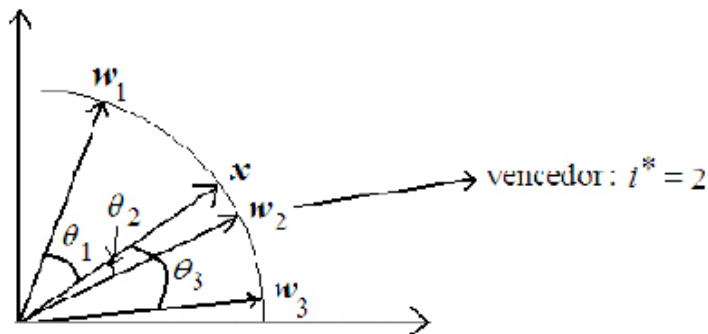


Figura 5.5 Processo de escolha do neurônio vencedor

A Figura 5.5 ilustra o processo de escolha do neurônio vencedor em um caso em que os padrões de entrada são bidimensionais, com uma rede neural com três neurônios e os vetores de peso e de entrada normalizados. O neurônio vencedor i^* é o que tiver o valor máximo de $\mathbf{w}_i \cdot \mathbf{x}$.

A normalização dos padrões de entrada provoca alterações, o que nem sempre é desejável. Para evitar isto, costuma-se usar como critério de definição do neurônio vencedor (equação 5.5) o cálculo da distância euclidiana entre o vetor \mathbf{x} e o vetor de pesos \mathbf{w}_i .

$$\|\mathbf{x} - \mathbf{w}_i\| = \sqrt{\sum_{j=1}^N (x_j - w_{ij})^2} \quad (5.5)$$

Conclui-se que maximizar o produto $\mathbf{w}_i \cdot \mathbf{x}$ é matematicamente equivalente a minimizar a distância euclidiana entre \mathbf{x} e \mathbf{w}_i . Portanto, o neurônio vencedor i^* é aquele cujo vetor de pesos tiver a menor distância euclidiana com relação ao padrão de entrada (esta equivalência vale apenas para entradas normalizadas).

Uma vez determinado o neurônio vencedor, o seu vetor de pesos é alterado pela regra do aprendizado competitivo, $\Delta \mathbf{w}_{i^*} = \eta(\mathbf{x} - \mathbf{w}_{i^*})$. Ela nos diz que o vetor de pesos do neurônio vencedor deve ser alterado por um fator η na direção de $\mathbf{x} - \mathbf{w}_{i^*}$. Geometricamente, em duas dimensões, como mostrado na Figura 5.6.

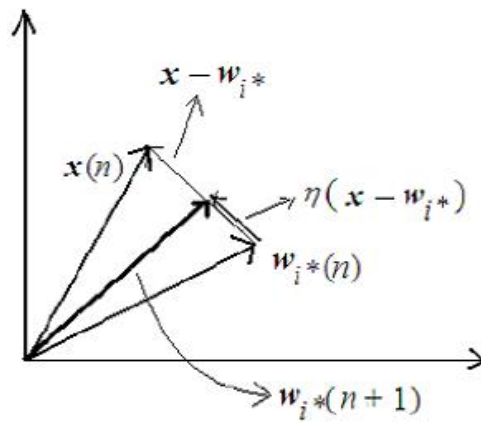


Figura 5.6 Alteração do vetor de pesos do neurônio vencedor.

Esta regra de mudança de pesos faz com que o vetor de pesos do neurônio vencedor, que já era o mais próximo do padrão de entrada, seja arrastado na direção do padrão de entrada, ficando ainda mais próximo dele.

A partir de uma população de padrões de entrada escolhe-se, a cada passo, um dos padrões, aplicando-o à entrada da rede. O neurônio vencedor é determinado e o seu peso alterado conforme a regra do aprendizado competitivo. Repetindo-se este procedimento várias vezes, os pesos acabam convergindo para uma situação de relativa estabilidade em que eles ficam nos centros de massa de agrupamentos de padrões de entrada (Figura 5.7).

A Figura 5.7 ilustra o processo de adaptação em um exemplo com 13 padrões de entrada e uma rede SOM com 3 neurônios.

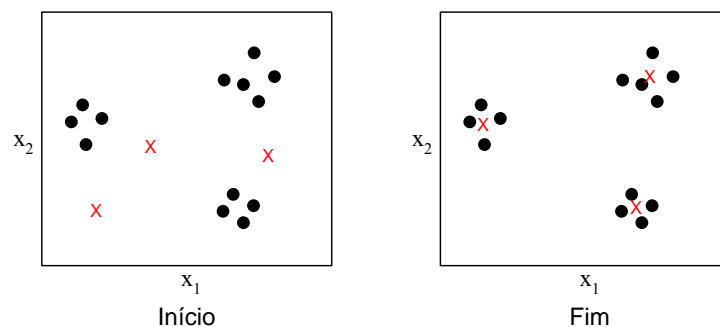


Figura 5.7: Convergência dos pesos da rede SOM para os centros de massa dos agrupamentos de padrões de entrada. No Início, os pesos da rede SOM (destacados com um X) estão distantes dos padrões de entrada (bolas pretas). No fim, a atualização dos pesos da rede SOM provoca o deslocamento do BMU para o centro de massa dos agrupamentos.

A convergência dos vetores de peso é controlada pelo valor de η . Se η for muito grande (próximo de 1) as mudanças nos pesos serão grandes e a convergência será rápida, mas os vetores de peso podem não se estabilizar. Se η for muito pequeno, a convergência pode ser muito vagarosa. Uma técnica muito adotada é fazer η variar com o passo de iteração t , começando com um valor próximo de 1 e diminuindo à medida que o aprendizado progride [VES00a].

A regra de aprendizado competitivo faz com que a rede SOM represente grupos de padrões que estejam próximos entre si por um vetor protótipo (Figura 5.7). O espaço de vetores de entrada fica dividido em regiões, ou vizinhanças, tais que

cada uma delas está associada a um vetor protótipo, que é o vetor de pesos de um dos neurônios da rede.

Os vetores protótipos são representantes de todos os possíveis vetores que estiverem na sua vizinhança.

Como já foi dito, a idéia básica das técnicas de agrupamento é buscar regiões de alta densidade de pontos no espaço de entrada, os agrupamentos (*clusters*) de dados, e escolher os seus centros para representar os pontos das regiões. Portanto, o centro de cada uma dessas regiões é o protótipo que representa todos os pontos da região. No caso da rede neural treinada por aprendizado competitivo, os protótipos são os vetores de pesos dos M neurônios da rede.

Quando um padrão \mathbf{x} é apresentado na entrada da rede, a unidade vencedora, i^* , é aquela cuja distância euclidiana entre o seu vetor de pesos \mathbf{w}_{i^*} e o padrão \mathbf{x} for a menor de todas (como no caso do aprendizado competitivo),

$$i^*(\mathbf{x}) = \min_j \|\mathbf{x} - \mathbf{w}_j\| \quad (5.6)$$

Na fase de cooperação não é só o neurônio vencedor que tem o seu vetor de pesos atualizado, mas os seus neurônios vizinhos (Figura 5.3). A regra de mudança de pesos é:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \Lambda_{i,i^*}(t)\eta(t)[\mathbf{x}(t) - \mathbf{w}_i(t)], \quad (5.7)$$

onde $\Lambda_{i,i^*}(t)$ é a chamada função de vizinhança centrada no neurônio vencedor i^* e $\eta(t)$ é a taxa de aprendizagem. Em geral, tanto $\Lambda_{i,i^*}(t)$ como $\eta(t)$ variam com o passo de aprendizagem t .

O efeito da introdução da função de vizinhança $\Lambda_{i,i^*}(t)$ é fazer com que o vetor de pesos não apenas do neurônio vencedor seja alterado na direção do padrão atual, mas que os vetores de pesos de todos os neurônios vizinhos ao neurônio vencedor também sejam alterados na direção do padrão atual; porém, a alteração do vetor de pesos vai diminuindo à medida que o neurônio correspondente vai ficando mais distante do vencedor [VES97]. Fazendo uma analogia, seria como se todos os vetores de pesos dos neurônios da rede SOM estivessem unidos por elásticos, sendo que os

elásticos mais fortes uniriam os vetores de pesos dos neurônios que fossem primeiros vizinhos, elásticos um pouco mais fracos uniriam os vetores de pesos dos neurônios que fossem segundos vizinhos, elásticos mais fracos uniriam ainda os vetores de pesos dos neurônios que fossem terceiros vizinhos e assim sucessivamente. Quando o vetor de pesos de um neurônio vencedor em um dado passo for alterado, ele irá arrastar consigo os demais vetores de peso, e mais fortemente aqueles dos neurônios mais próximos.

Em geral, usa-se uma função gaussiana para implementar a função de vizinhança [KOH97a] como a descrita na equação 5.8,

$$\Lambda_{i,i^*}(t) = \exp\left(-\frac{d_{i,i^*}^2}{2\sigma^2(t)}\right), \quad (5.8)$$

onde d_{i,i^*} é a distância entre um neurônio i e o neurônio vencedor i^* . Se a rede neural for unidimensional, essa distância é simplesmente o módulo da diferença entre os índices de i e i^* , $d_{i,i^*} = |\text{índice_de_}i - \text{índice_de_}i^*|$; se a rede for bidimensional, essa distância é dada pela distância euclidiana entre os seus vetores de posição mostrada na equação 5.9.

$$d_{i,i^*}^2 = \|\mathbf{r}_i - \mathbf{r}_{i^*}\|^2, \quad (5.9)$$

onde \mathbf{r}_i é o vetor posição do neurônio i e \mathbf{r}_{i^*} é o vetor posição do neurônio vencedor, os dois sendo medidos no espaço discreto definido pelos nós da rede neural.

O desvio padrão da função de vizinhança, $\sigma_{i,i^*}(t)$ diminui com o número de passos t . Uma maneira comum de implementar essa diminuição é por um decaimento exponencial descrito na equação 5.10,

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_1}\right), \quad (5.10)$$

onde τ_1 é uma constante temporal (determinada empiricamente).

Em geral, também se faz a taxa de aprendizagem $\eta(t)$ diminuir com o passo de iteração de uma maneira exponencial descrita na equação 5.11,

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{\tau_2}\right), \quad (5.11)$$

onde τ_2 é outra constante temporal (também determinada empiricamente).

Os padrões vizinhos no espaço de entrada são representados por neurônios vizinhos na rede neural artificial. As regiões do espaço de entrada cujos padrões x tenham maior probabilidade de ocorrer são representadas por um número maior de neurônios da rede neural artificial, isto é, elas são representadas com uma resolução maior do que as regiões cujos padrões ocorrem menos freqüentemente.

Portanto, um mapa entre o espaço contínuo de entrada e o espaço discreto da rede neural implementado pelo algoritmo SOM tende a preservar tanto a métrica como a distribuição do espaço de entrada. É importante lembrar que isto ocorre a partir apenas da informação contida nos padrões de entrada; o mapa criado pelo SOM não é supervisionado [ZUC03].

O treinamento da rede SOM ocorre em duas fases [BRA00]:

1. Fase de Auto-organização ou de Ordenamento: É nesta fase que ocorre o ordenamento topográfico dos vetores de pesos. Inicialmente, os vetores de pesos têm valores aleatórios e não possuem qualquer tipo de ordenamento. À medida que a rede vai sendo treinada, vetores de neurônios vizinhos entre si no espaço da rede neural começam a se aproximar uns dos outros, de maneira que neurônios de uma mesma área da rede neural acabam representando padrões vindos de uma mesma região do espaço de entrada. Esta fase pode levar muitas iterações para terminar, em geral mais do que 1000 .

2. Fase de Convergência: Nesta fase ocorre o refinamento do mapa, levando a uma representação mais acurada do espaço de entrada por parte da rede neural. Como regra geral, o número de iterações nesta fase deve ser de pelo menos 500 vezes o número de neurônios na rede neural. Logo, o número de iterações pode atingir dezenas de milhares de passos. Nesta fase, a taxa de aprendizagem η deve permanecer pequena, da ordem de 0,05 e a função de vizinhança deve englobar apenas o próprio neurônio vencedor e, no máximo, os seus primeiros vizinhos.

Pode-se sintetizar o algoritmo da rede SOM da seguinte forma [HAY99]:

1. Inicialização. Escolha valores aleatórios para as componentes iniciais dos vetores de pesos.
2. Escolha do padrão de entrada. Escolha um padrão \mathbf{x} da população para ser colocado na camada de entrada da rede.
3. Determinação do neurônio vencedor. Use o critério de similaridade baseado na distância euclidiana entre o vetor de entrada e os vetores de peso para determinar o neurônio vencedor i^* para o passo atual.
4. Atualização dos pesos. Modifique os vetores de pesos dos neurônios da rede.
5. Continuação. Volte para o passo 2 e continue até que não sejam observadas mudanças significativas no mapa formado.

5.8 Medidas de Qualidade

Como discutido no capítulo 1, ainda não é claro como se devem escolher os parâmetros do algoritmo da rede SOM de forma a obter um “bom mapeamento” [ZUC03]. Normalmente a rede SOM é avaliada com base na resolução do mapa e em relação à preservação da topologia [KIV95]. A escolha do “melhor mapeamento” deveria ser, portanto, por aquele que “melhor representa os dados de entrada”. Este critério normalmente é traduzido por duas medidas: o erro de quantização (EQ) e o erro topográfico (ET).

O erro de quantização (EQ) corresponde à média das distâncias entre cada vetor de dados \mathbf{x} e o correspondente vetor de pesos do neurônio BMU. A medida corresponde à resolução do mapa e está descrita na equação 5.12.

$$EQ = \frac{1}{N} \sum_{k=1}^N \|\mathbf{x} - \mathbf{w}_{i^*}\| \quad (5.12)$$

O erro topográfico (ET) quantifica a capacidade do mapa em representar a topologia dos dados de entrada. Para cada vetor de dados de entrada \mathbf{x} são calculados seu primeiro BMU e o seu segundo BMU e toda vez que eles não forem adjacentes

(vizinhos, próximos), aumenta-se o erro em uma unidade, tirando-se depois a média pelo número total de vetores [KIV95], onde $u(\mathbf{x}_k)$ igual a 1 se o primeiro e o segundo BMU não são adjacentes, caso contrário igual a 0.

$$ET = \frac{1}{N} \sum_{k=1}^N u(\mathbf{x}_k) \quad (5.13)$$

A medida EQ corresponde à acuidade, ou resolução do mapa, que é inversamente proporcional ao número de neurônios, ou seja, o erro de quantização diminui com o aumento do número de neurônios no arranjo (resolução aumenta). Se o número de neurônios for muito grande e (ou) se sofreu um processo de treinamento onde o raio de vizinhança (define a largura da região da vizinhança) torna-se menor ou igual a 1 durante muito tempo, pode ocorrer de os neurônios se posicionarem sobre os objetos a serem representados. Neste caso, $EQ = 0$, mas o arranjo pode estar tão retorcido que a capacidade de representar a topologia dos dados é perdida (ET aumenta). O comportamento de ET nesta situação dependerá também do número de neurônios disponíveis no arranjo; ET aumenta se há poucos neurônios e diminui se há muitos neurônios [ZUC03].

Quando ambos os valores de EQ e ET são muito baixos, suspeita-se de sobreajuste (*overfitting*) A rede SOM, na tentativa de representar o mais fielmente possível os dados, “dobra-se” de tal forma que acaba representando exatamente os dados, podendo perder a sua capacidade de generalização.

O fenômeno inverso, o subajuste (*underfitting*), ocorre quando um mapa é “rígido” demais. Isso pode acontecer quando há poucos neurônios para representar um número proporcionalmente grande de dados ou o raio de vizinhança final da função gaussiana for maior que 1 durante o treinamento. Neste caso, os valores de EQ podem ser mais altos (os vetores de pesos dos neurônios encontram-se, em média, menos próximos dos vetores de dados).

Em geral, valores muito baixos de ET, associados a valores mais altos de EQ, podem sugerir subajuste [KAS98; SVE98].

5.9 Visualização do Mapa Gerado pela Rede SOM

Na seção 5.8 foram descritos alguns dos motivos que fazem com que a rede SOM seja muito utilizada para o KDD, sendo que um deles é a visualização.

Nesta seção, serão apresentadas algumas formas de visualização dos *clusters* gerados pela rede SOM.

5.9.1 A Matriz-U

A Matriz-U, ou matriz de distâncias unificadas [ULT89], é a matriz de distâncias entre cada neurônio do mapa e seus vizinhos. É a técnica de visualização mais utilizada para detectar *clusters* da rede SOM [VES02]. A Figura 5.8 mostra um exemplo de Matriz-U.

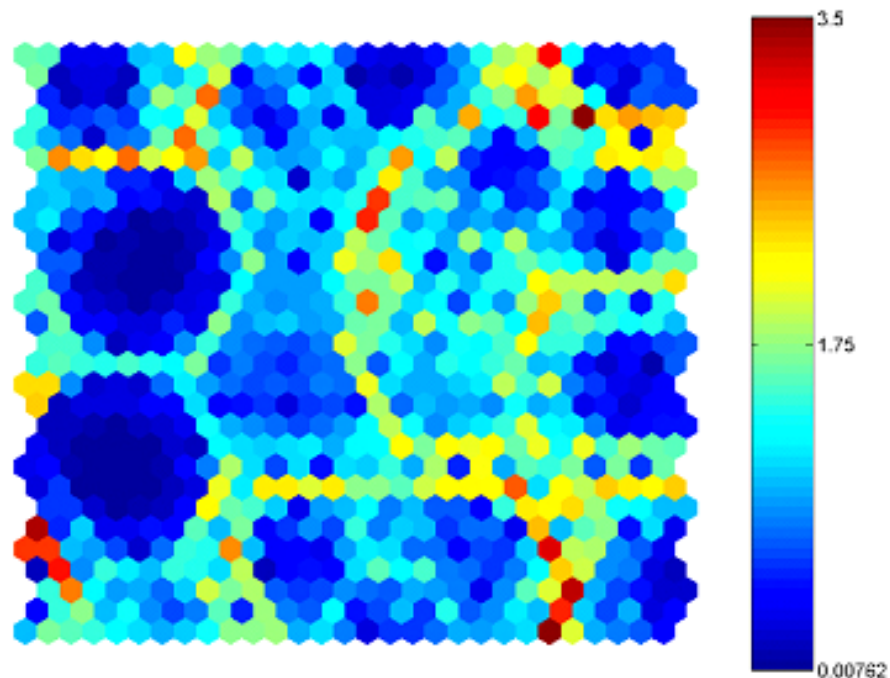


Figura 5.8: Matriz-U (matriz de distâncias unificadas).

A Matriz-U mostra a topologia do mapa e conseqüentemente a topologia dos dados de entrada. A análise da Figura 5.8 revela que os valores altos presentes na barra de cores, representados por tons escuros, correspondem a grandes distâncias no espaço de entrada, e valores baixos, representados por tons claros, indicam similaridades no espaço de entrada, caracterizando *clusters*. Assim, um dos usos mais comuns da Matriz-U é servir como ponto de partida para a clusterização.

5.9.2 Mapa por Similaridade de Cor

Esta outra forma de representação é útil tanto para a visualização da distribuição dos *clusters* no mapa como para a criação de rótulos, permitindo a análise da estrutura dos *clusters*. Na Figura 5.9 pode-se observar que foi utilizado o contraste de cores para representar os 10 *clusters* rotulados de acordo com as classes contidas na base de dados.

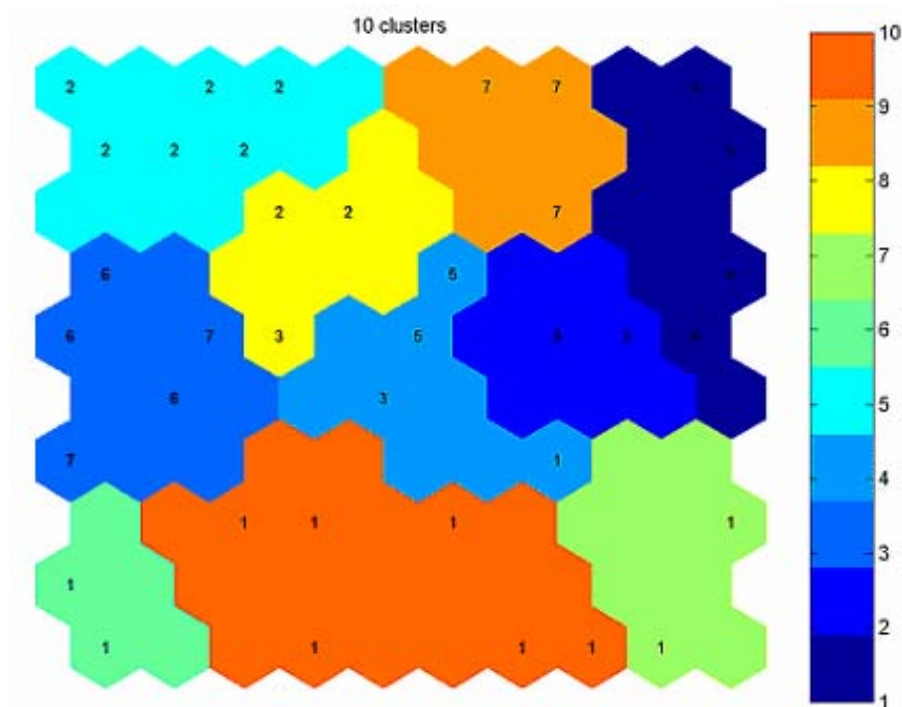


Figura 5.9: Mapa por Similaridade de Cor evidenciando a localização de agrupamentos

Na Figura 5.9, a barra ao lado do mapa representa a relação entre o número do *cluster* e a sua respectiva cor, como por exemplo: o *cluster* de número 8 corresponde no mapa à cor amarela. Estas informações são relevantes na escolha do “melhor” mapa com base nas similaridades entre os dados e nas diferenças de separação dos *clusters* e, por isso, fazem parte da análise dos experimentos realizados no capítulo 6.

A Figura 5.10 exibe um mapa por similaridade de cor onde cada neurônio foi rotulado com todos os tipos de dados para os quais foi vencedor (BMU).



Figura 5.10: Mapa por Similaridade de Cor onde cada grupo de neurônios foi rotulado com todos os tipos de dados para os quais foi vencedor (BMU).

Os resultados da qualidade visual dos mapas obtidos nos experimentos realizados no sexto capítulo serão analisados através do Mapa por Similaridade de Cor.

Sempre que se desejar descobrir conhecimento novo em uma massa de dados, deve-se pensar na possibilidade de apresentar essa massa a uma rede SOM [CAR01]. Os mapas oferecem, através das ferramentas de visualização descritas neste capítulo, uma forma geométrica simples de verificar se há algo de interessante ou organizado

na massa de dados. Muitos programas comerciais de mineração de dados possuem rotinas prontas de redes SOM [GOL05].

5.10 Resumo do Capítulo

Este capítulo apresentou de forma introdutória um panorama geral do histórico, dos principais conceitos e aplicações das redes neurais artificiais. O enfoque principal foi dado na rede SOM, onde foram abordadas as características e a estrutura da rede.

O capítulo se encerra com duas análises: das medidas de qualidade mais utilizadas para avaliar a resolução do mapa e a preservação da topologia da rede SOM: o erro de quantização (EQ) e o erro topográfico (ET) [KIV95]; e do Mapa por Similaridade de Cor, que possibilita a escolha do “melhor” mapa com base nas similaridades entre os dados e nas diferenças de separação dos *clusters*.

As análises são importantes porque a avaliação do desempenho da Arquitetura Híbrida proposta neste trabalho está embasada nos resultados do erro de quantização (EQ), do erro topográfico (ET), no número de *clusters* gerados, no Mapa por Similaridade de Cor e no tempo de treinamento da rede SOM. O sexto capítulo trata dos experimentos realizados.

6 Experimentos Comparativos com a Arquitetura Híbrida (RS com rede SOM)

Neste capítulo, realizou-se uma comparação entre a Arquitetura Híbrida proposta (RS com a rede SOM) e apenas a rede SOM analisando-se os resultados obtidos para a descoberta de conhecimento em bases de dados (KDD).

A comparação teve como objetivo principal verificar se a Arquitetura Híbrida apresentou um desempenho superior ao da rede SOM. Foram utilizadas nos experimentos cinco bases de dados diferentes, nas quais há variação de dimensionalidade, quantidade e tipos de dados disponíveis (discretos, contínuos, binários, etc) buscando uma melhor validação dos experimentos.

Neste capítulo ainda estão descritos: o motivo de escolha das bases de dados, a metodologia utilizada na realização dos experimentos, os resultados experimentais e a avaliação destes resultados com base no conhecimento descoberto em cada experimento.

6.1 Metodologia e Bases de Dados utilizadas nos Experimentos

Como dito anteriormente, foram realizadas no total cinco séries de experimentos com bases de dados distintas, comparando a Arquitetura Híbrida proposta com uma rede SOM, divididos da seguinte forma:

Experimento 1: Realizado com uma base de dados gerada artificialmente, o que possibilitou também a verificação do RS com relação aos conceitos de dependência e significância de atributos (realizado somente neste experimento).

Experimento 2: Realizado com a base de dados pública Íris [UCI98].

Experimento 3: Realizado com a base de dados pública ZOO [UCI98].

Experimento 4: Realizado com a base de dados do mundo real Consumidor [FER03, SAS05].

Experimento 5: Realizado com a base de dados do mundo real *The Insurance Company* [TIC00].

A área de *marketing* tradicionalmente tem muito interesse em KDD como por exemplo, conhecer o perfil do consumidor. Assim, as bases de dados utilizadas nos experimentos 4 e 5 são da área de *marketing*.

Vale ressaltar que o principal critério de escolha das bases de dados foi a quantidade de atributos (dimensionalidade), o que justifica a escolha de cinco bases de dados com variação no número de atributos, pois procurou-se avaliar dentro da Arquitetura Híbrida o comportamento do RS, que na geração dos redutos (capítulo 4, seção 4.5) reduz a dimensionalidade da base de dados.

Os experimentos foram realizados em duas fases distintas descritas a seguir:

Na primeira fase de cada experimento, denominada **rede SOM sem redutos**, apresentou-se à rede SOM cada uma das cinco bases de dados dos experimentos com todos os atributos e avaliaram-se os resultados.

A Figura 6.1 ilustra a primeira fase dos experimentos.

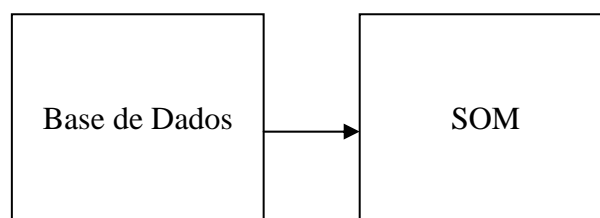


Figura 6.1: Primeira fase dos experimentos em que as bases de dados são apresentadas à rede SOM.

Na segunda fase, denominada **Arquitetura Híbrida** (rede SOM com redutos): apresentou-se primeiro ao RS cada uma das cinco bases de dados para a geração dos redutos e, em seguida, as bases de dados reduzidas foram apresentadas à rede SOM.

A Figura 6.2 ilustra a segunda fase dos experimentos.

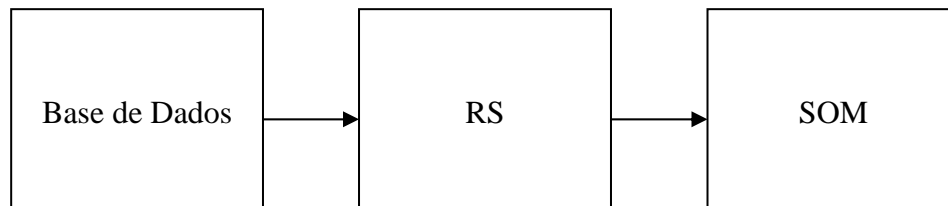


Figura 6.2 Segunda fase dos experimentos em que as bases de dados são apresentadas à Arquitetura Híbrida (rede SOM com redutos).

Nos experimentos 1 e 2, realizados com bases de dados mais simples, foram observados: os conceitos de dependência e significância de atributos na geração de redutos pelo RS, o número de *clusters* e a visualização do mapa gerado pela rede SOM utilizando o Mapa por Similaridade de Cor (capítulo 5).

Após análise dos resultados obtidos nos experimentos 1 e 2 foram realizados os experimentos 3, 4 e 5, utilizando bases de dados mais complexas, onde se observou a comparação entre a Arquitetura Híbrida proposta e a rede SOM com mais detalhes, através:

- a) do valor do Erro de Quantização (EQ)
- b) do valor do Erro Topográfico (ET)
- c) do Tempo de Treinamento da rede SOM
- d) do número de *clusters* gerados
- e) da visualização do mapa gerado utilizando o Mapa por Similaridade de Cor.

Conclui-se que o melhor desempenho é aquele que apresenta os melhores resultados para os itens descritos acima (a até e).

Note-se que nos experimentos 4 e 5, foram abordadas bases de dados que representam problemas concretos de KDD no contexto da área de aplicação discutida no capítulo 2, ou seja, organizações e o uso de mineração de dados para apoio à decisão.

6.1.1 Ferramentas utilizadas e Plataformas de Ensaio

Para realização dos experimentos com a rede SOM foi utilizada a ferramenta SOM *Toolbox* [SOM99], uma implementação do Mapa Auto-Organizável de Kohonen em Matlab. A escolha desta ferramenta baseou-se no grande número de trabalhos publicados que relatam a sua utilização [ALH00; DUT01; NUN02; VES99b; VES00b; VES00c; ZUC03]. De certa forma, esta ferramenta pode ser considerada como uma plataforma padrão que tem sido adotada em grande parte das pesquisas atuais com Mapas Auto-Organizáveis. Essa *Toolbox* foi originalmente escrita para Matlab 5.0, porém funciona também nas versões mais atuais. A *toolbox* pode ser utilizada para pré-processamento dos dados, inicialização e treinamento do Mapa Auto-Organizável em várias topologias e para visualização dos dados e do mapa após treinamento. Possui também funções de análise das propriedades dos mapas e dos dados, correlação entre variáveis e funções de clusterização. A ferramenta foi desenvolvida para conjuntos pequenos de dados (até 10.000 registros), mas é capaz de trabalhar também com conjuntos de tamanho médio (até 1.000.000 de registros). É de domínio público e pode ser encontrada para instalação no endereço <http://www.cis.hut.fi/projects/somtoolbox/> [SOM99].

Para a realização dos experimentos com RS foi utilizada a ferramenta chamada Rosetta (*A Rough Sets Toolkit for Analysis of Data*) [ROS97], de domínio público, encontrada para instalação no endereço: <http://www.idi.ntnu.no/~aleks/rosetta>.

Como acontece com a SOM *Toolbox*, a escolha da ferramenta Rosetta foi embasada no grande número de publicações que relatam a sua utilização [KOM97; 98; 02; LAZ02; ØHM99; 00a ; 00b; PIL01].

A ferramenta vem sendo continuamente aperfeiçoada desde a sua primeira versão em 1997, estando atualmente na versão 1.4.40. Foi desenvolvida numa cooperação entre o Grupo de Sistemas de Conhecimento do Departamento de Computação e Ciência da Informação da Universidade Norueguesa de Ciência e Tecnologia, Noruega e o Grupo de Lógica do Instituto de Matemática da Universidade de Warsaw, Polônia. Informações mais detalhadas sobre a ferramenta Rosetta podem ser encontradas em Pila [PIL01].

A plataforma de *hardware* utilizada nos experimentos foi um Pentium IV com 2.4 MHZ, 512 MB de memória RAM e 40 GB de disco rígido.

6.1.2 Parâmetros da Rede SOM adotados nos Experimentos

Como já foi citado no capítulo 5, a escolha dos parâmetros, tanto de inicialização quanto de treinamento do mapa, ainda não segue regras bem definidas. Mesmo lançando mão das heurísticas existentes para definição dos parâmetros, é consenso entre os pesquisadores que se devem efetuar alguns testes com diferentes configurações de mapas antes de decidir-se qual representa melhor o conjunto de dados em questão.

Em geral, são propostas heurísticas baseadas no comportamento do mapa e em medidas de qualidade como EQ e ET.

Os parâmetros que regulam a rede SOM podem ser agrupados em dois conjuntos: Parâmetros de estrutura e Parâmetros de treinamento.

Os Parâmetros de estrutura são os seguintes:

- Dimensões: tamanho do mapa, número de neurônios.
- Vizinhança: hexagonal ou retangular.
- Formato do Arranjo: Folha, Cilindro ou Toróide.

Os Parâmetros de treinamento correspondem ao número de épocas de treinamento.

Nos experimentos realizados variaram-se alguns dos parâmetros de estrutura para conhecer a sua influência nos resultados. Desta forma, todos os parâmetros descritos nesta seção foram utilizados nos experimentos tanto na primeira fase como na segunda fase, e são os seguintes:

Parâmetros de estrutura:

- Dimensões: número de neurônios para os experimentos 1 e 2 (que possuem um número menor de registros e de atributos) $5 \times 5 = 25$ neurônios.
- Número de neurônios para os experimentos 3, 4 e 5 (que possuem um número maior de registros e de atributos) $15 \times 15 = 225$ neurônios [ZUC03]
- Vizinhança: hexagonal [KOH97a]
- Formato do Arranjo: plano hexagonal.

Parâmetros de treinamento:

O treinamento de um mapa na *SOM Toolbox* é dividido em 2 etapas conforme descrito na seção 6.1.2: “*rough phase*” e “*fine tune*”. Para cada uma destas etapas são definidos diferentes números de épocas. Os valores default [VES00c] são: para a “*rough phase*” = 10 x mpd épocas, e para a “*fine tune*” = 40 x mpd épocas, onde mpd = neurônios/dados.

A taxa de aprendizado foi de 0,5 para a fase inicial e 0,05 para a fase de convergência [KAS97].

As medidas de qualidade utilizadas são o Erro Médio de Quantização (EQ) e o Erro Topográfico (ET).

6.2 Objetivo dos Experimentos Contrastivos

O objetivo principal das séries de experimentos documentadas neste capítulo é verificar qual das duas alternativas tem melhor desempenho na descoberta de conhecimento em bases de dados: a rede SOM ou a Arquitetura Híbrida?

A escolha do “melhor” mapa deve se basear em algumas características, como a representação da topologia da estrutura dos dados, as similaridades entre os dados e as diferenças de separação dos agrupamentos.

No caso da rede SOM, o critério adotado para escolher o “melhor” resultado do mapa gerado utilizando-se inclusive dos rótulos de dados é aquele que apresenta o menor erro de quantização (EQ) e o menor erro topográfico (ET). Estes indicadores expressam a capacidade do mapa em representar a topologia dos dados de entrada.

Os menores valores para esses erros, em geral, indicam uma melhor adaptação do mapa à topologia dos dados no espaço de entrada. Foram descartados os resultados que apresentaram ET igual a 0 devido à possibilidade de sobreajuste (*overfitting*) ou subajuste (*underfitting*). Neste trabalho, além do critério descrito acima, serão adotados também como critério o tempo de treinamento da rede SOM (esforço computacional requerido), o número de *clusters* gerados, o que indica as similaridades entre os dados e as diferenças de separação dos *clusters* e a visualização do mapa gerado através do Mapa por Similaridade de Cor.

6.3 Realização dos Experimentos

Para verificar o desempenho da Arquitetura Híbrida proposta neste trabalho, compararam-se os resultados obtidos na primeira fase dos experimentos com os resultados obtidos na segunda fase dos experimentos. Nos experimentos 1 e 2 avaliaram-se os conceitos de dependência e significância de atributos na geração dos redutos, o número de *clusters* gerados e a visualização do mapa através do Mapa por Similaridade de Cor. Nos experimentos 3, 4 e 5 avaliaram-se o EQ, o ET, o tempo de treinamento da rede SOM, o número de *clusters* gerados e a visualização do mapa através do Mapa por Similaridade de Cor.

O procedimento básico de realização dos experimentos utilizando as ferramentas foi o seguinte: na primeira fase (SOM sem redutos) com a utilização da

ferramenta *SOM Toolbox*, apresentou-se à rede SOM a base de dados com todos os atributos.

Na segunda fase (Arquitetura Híbrida) submeteu-se primeiro a base de dados com todos os atributos à ferramenta Rosetta para que com base em RS reduzisse o número de atributos. O reduto com o menor número de atributos foi o escolhido por apresentar o menor esforço computacional [MIT97]. Finalmente, a base de dados reduzida pelo RS foi apresentada à rede SOM.

6.3.1 Experimento 1: Realizado com uma Base de Dados Artificial

Foram geradas artificialmente três classes com diferentes distribuições normais (Figura 6.3). O primeiro conjunto (em cor vermelha), considerou para a geração dos dados os seguintes valores de média [1, 1,5 , 2] com dispersão de 0,2. No segundo conjunto (em cor verde), consideraram-se os seguintes valores de média [1, 2, 3] e dispersão 0,3 e o terceiro conjunto (em cor azul) com valores de média [1, 2,5, 3,5] e dispersão 0,1. Os valores médios foram escolhidos tendenciosamente para que cada classe (vermelho, verde e azul) tivesse dispersões diferentes e para que pelo menos um atributo fosse o responsável em forçar alguma sobreposição entre as classes.

O objetivo é utilizar o RS para investigar o atributo responsável pela sobreposição, e conseqüentemente sugerir um conjunto reduzido de atributos para reduzir essa sobreposição melhorando as fronteiras entre os *clusters*.

Considera-se como atributo cada eixo do gráfico (X1, X2, X3) ilustrado na Figura 6.3. Na Figura 6.3, temos quatro ilustrações diferentes. A primeira com a distribuição inicial dos dados, a segunda sem o atributo X1 (somente X2, X3) a terceira sem o atributo X2 (somente X1, X3) e a quarta sem o atributo X3 (somente X1, X2). O objetivo destes gráficos é visualizar a nova distribuição dos dados desprezando um atributo por vez.

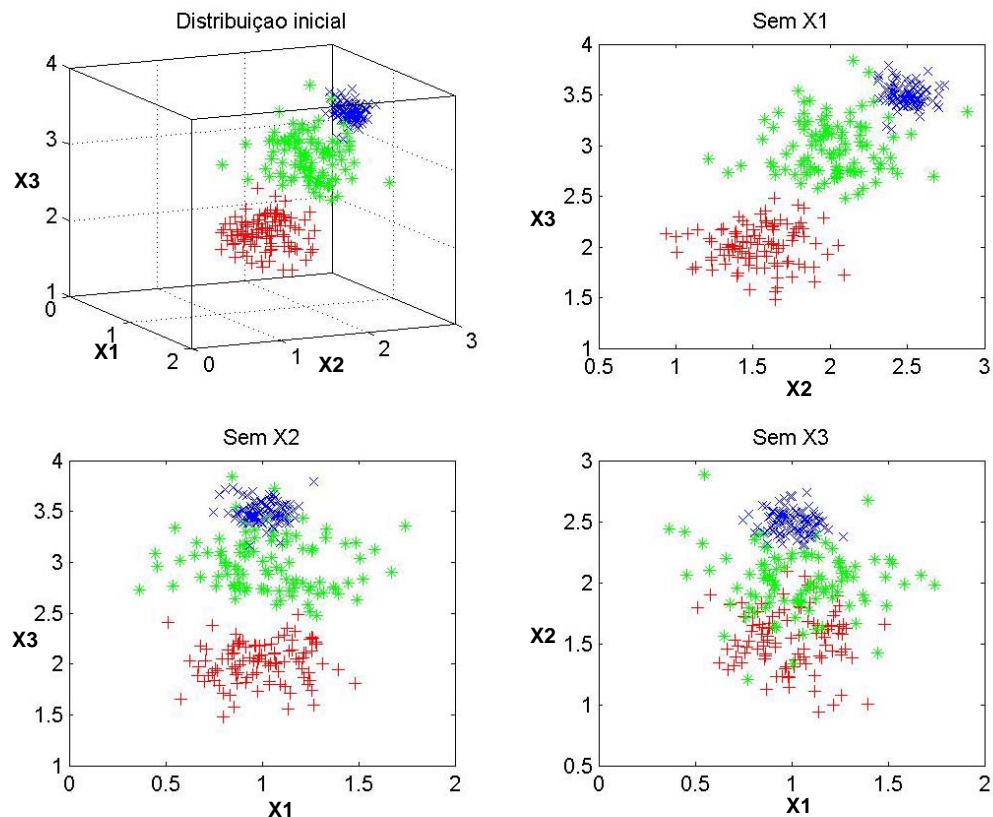


Figura 6.3: Ilustrações do Experimento 1 apresentando a Distribuição Inicial dos dados e as Distribuições sem um dos atributos. Em sentido horário: na ilustração acima, no alto, à esquerda, visualiza-se a distribuição inicial dos dados com os três atributos representados; no alto, à direita, visualiza-se a distribuição dos dados sem o atributo X1; embaixo, à esquerda, visualiza-se a distribuição dos dados sem o atributo X2; e embaixo, à direita, visualiza-se a distribuição dos dados sem o atributo X3.

Analisando a Figura 6.3, pode-se perceber que quando considera-se a presença do atributo X1 a sobreposição dos dados é maior. Espera-se então que RS gere redutos sem a presença do atributo X1.

A redução com RS sugeriu utilizar o atributo X3 ou o atributo X2, reduzindo a dimensionalidade dos dados de 3 para apenas 1 dimensão, porém sem perder a representatividade dos dados.

A Tabela 6.1 Mostra os redutos gerados pelo RS.

Número de redutos	Reduto	Suporte	Comprimento
1	{X3}	100	1
2	{X2}	100	1

Tabela 6.1: Redutos gerados pelo RS no Experimento 1 utilizando uma base de dados artificial. O suporte igual a 100 indica que os redutos representam a base de dados na sua totalidade e portanto podem gerar regras precisas no caso de uma classificação. O comprimento indica o número de atributos presentes no reduto, no caso 1 atributo.

Com o mesmo objetivo da figura anterior, apresentam-se na Figura 6.4 os gráficos de distribuição das classes com apenas um atributo cada e assim avalia-se a separabilidade entre as classes.

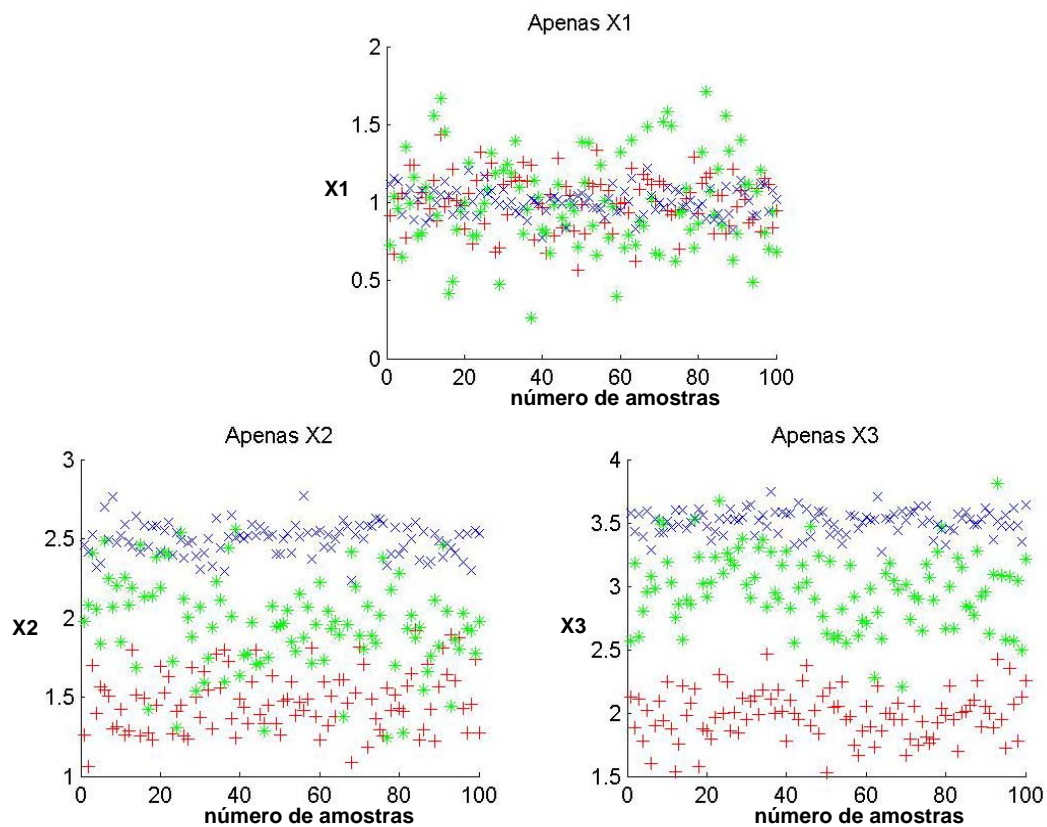
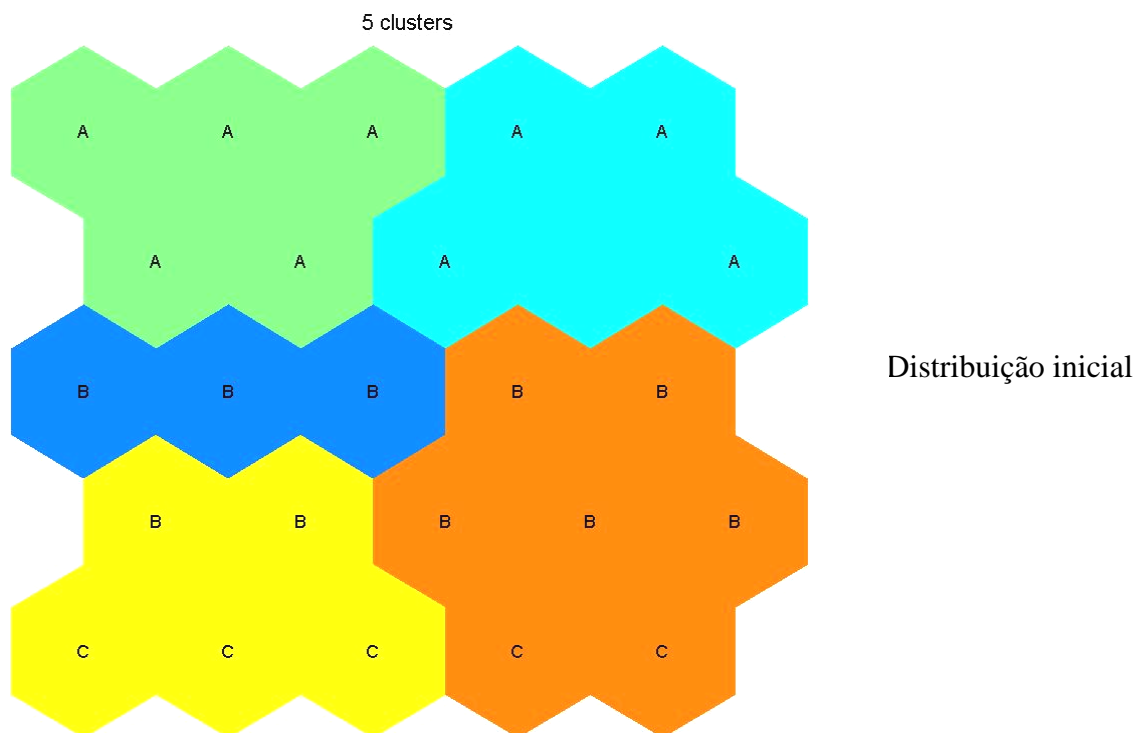


Figura 6.4: Ilustrações do Experimento 1 apresentando a Distribuição das classes com apenas um atributo cada.

Da análise da Figura 6.4, observa-se que o atributo X1 causa grande sobreposição entre as classes, o contrário do que ocorre quando analisamos os gráficos com o atributo X2 ou com o atributo X3. O atributo X3 é o que fornece

maior separação linear entre as classes. Concluiu-se então que o reduto com apenas o atributo X3 é o que apresenta melhor resultado.

Foi utilizada a rede SOM para validar os experimentos anteriores, em que se procurou explorar no mapa gerado os resultados da redução bem como procurou-se analisar a melhora entre agrupamentos. Para isso foram gerados três mapas: um com a distribuição inicial, outro utilizando o reduto com o atributo X2 e o terceiro utilizando o atributo X3, estes últimos gerados pelo RS. Os resultados são mostrados na Figura 6.5.



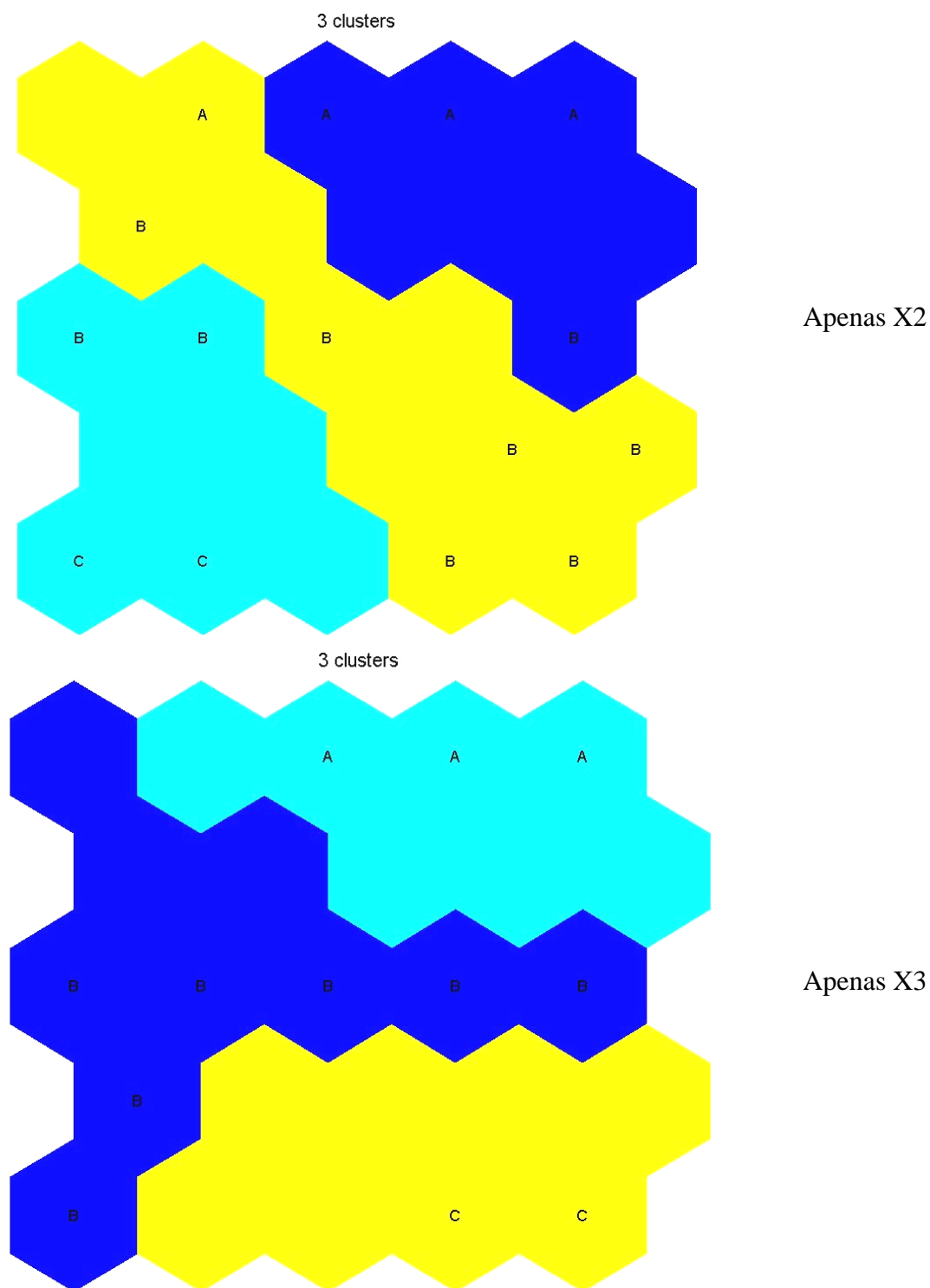


Figura 6.5: Visualização dos Mapas gerados no Experimento 1 pela rede SOM sem redutos (Distribuição inicial) e pela rede SOM com o reduto X2 e com o reduto X3 (Arquitetura Híbrida).

Na Figura 6.5, a primeira fase do experimento (rede SOM sem redutos) é mostrada através do mapa que apresenta a Distribuição inicial e a segunda fase do experimento (Arquitetura Híbrida) através dos dois mapas (ao centro e à direita) que apresentam apenas o atributo X2 e apenas o atributo X3.

Avaliação dos resultados

Ao sugerir os redutos, RS mostrou que, além de reduzir informação não necessária à rede SOM, também utilizou-se dos conceitos de dependência e significância de atributos para gerar os redutos, mantendo a representatividade inicial da base de dados.

Observou-se então na Figura 6.5 que, quando todos os atributos foram utilizados, a rede SOM apresentou um número maior de *clusters* (5) em função da sobreposição causada pelo atributo X1. Por outro lado, quando usada a Arquitetura Híbrida os resultados foram melhores, pois o RS gerou redutos sem o atributo X1, que causava a sobreposição das classes. Assim, com apenas o atributo X2 a rede SOM gerou 3 *clusters*, porém ainda com alguns exemplos confusos entre uma classe e outra. Finalmente, quando se usou apenas o atributo X3, este mostrou ser o que melhor representou as classes, formando três classes diferentes e sem nenhuma sobreposição. Isto ocorreu porque RS reduziu informação desnecessária da base de dados que era apresentada à rede SOM, melhorando a similaridade entre os dados e reduzindo a diferença de separação dos agrupamentos (de 5 para 3 *clusters*).

Neste experimento simples, conclui-se que o desempenho da Arquitetura Híbrida pode ser considerado melhor.

6.3.2 Experimento 2: Com a Base de Dados Íris (Base Pública)

A base de dados Íris talvez seja uma das bases mais conhecidas na literatura de reconhecimento de padrões [UCI98]. O conjunto de dados contém 3 classes de pétalas de 50 exemplos cada, onde cada classe refere-se a um tipo de pétala da planta Íris. Sabe-se que uma das classes é linearmente separável das outras duas e estas não

são linearmente separáveis entre si. A base de dados Íris contém 4 atributos numéricos mais a classe, descritos a seguir: 1. *sepal length* em cm; 2. *sepal width* em cm; 3. *petal length* em cm e 4. *petal width* em cm. As classes são: Íris-setosa; Íris-versicolour e Íris-virgínica. Sabe-se que a correlação entre os atributos *petal length* e *petal width* é alta (0,9565) e a Distribuição das classes é de 33,3% para cada uma das 3 classes.

O número de épocas foi calculado da seguinte forma: *rough phase* = quantidade de neurônios (225) / tamanho de dados (150) x 10 = 16,0 e a *fine tune* = quantidade de neurônios (225) / tamanho de dados (150) x 40 = 67,0. Somando as duas fases (16,0 + 67,0) obtêm-se 83 épocas.

Primeira fase do experimento: rede SOM sem redutos

Apresentada a base de dados Íris com todos os seus atributos à rede SOM, obteve-se o mapa que pode ser visualizado na Figura 6.6.

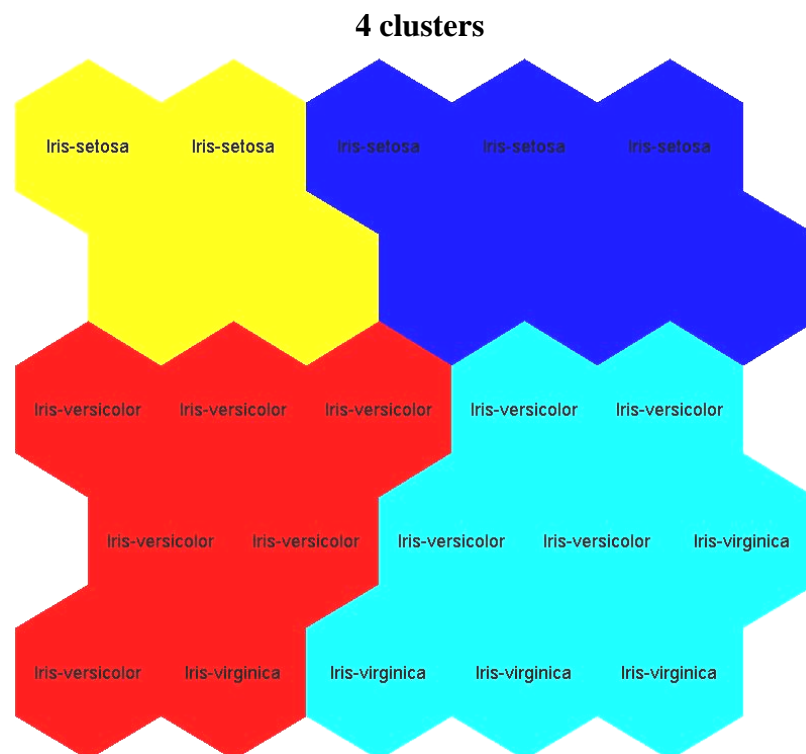


Figura 6.6: Mapa gerado pela rede SOM sem redutos após a apresentação da base de dados Íris (primeira fase do Experimento 2).

Do mapa da Figura 6.6, observou-se que a rede SOM formou 4 *clusters* com base nas três classes, sendo que duas classes apresentaram problemas de sobreposição (*cluster* na cor vermelha e *cluster* na cor azul-clara) e a terceira classe foi separada em dois *clusters* (nas cores amarela e azul-escura).

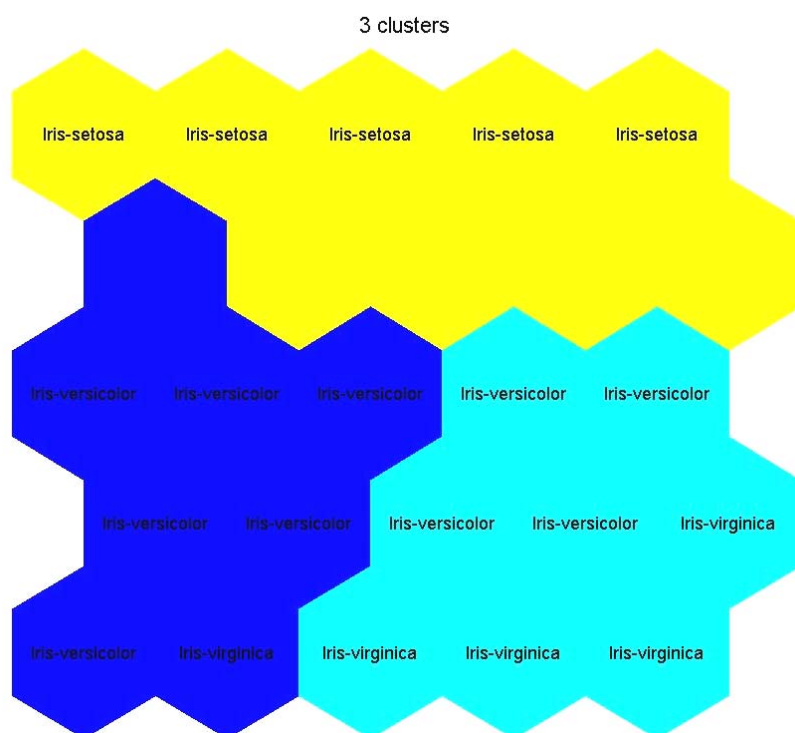
Segunda fase do experimento: rede SOM com redutos (Arquitetura Híbrida).

Ao aplicar o RS para reduzir os atributos, foram sugeridos 4 redutos, capazes de representar a base de dados Íris com todos os seus atributos. A Tabela 6.2 mostra os redutos gerados pelo RS.

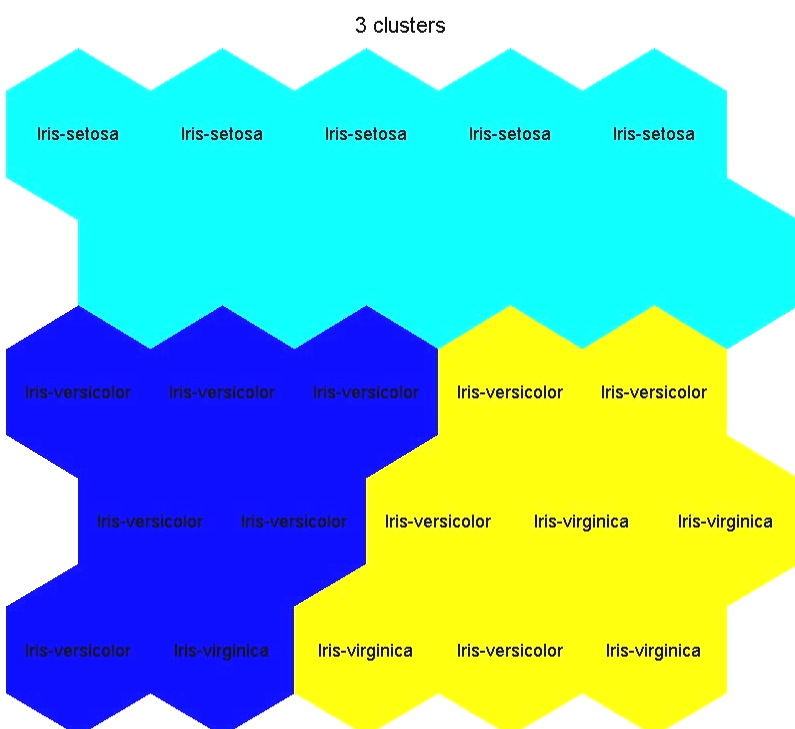
Reduto	Suporte	Comprimento
<i>{SepalLength, SepalWidth, PetalWidth}</i>	100	3
<i>{SepalLength, SepalWidth PetalLength}</i>	100	3
<i>{ SepalWidth, PetalLength, PetalWidth}</i>	100	3
<i>{ SepalLength, PetalLength,PetalWidth}</i>	100	3

Tabela 6.2: Os 4 redutos gerados pelo RS utilizando a base de dados Íris (Experimento 2)

Em seguida, a base de dados Íris foi reduzida obedecendo aos 4 redutos gerados pelo RS e apresentada à rede SOM. Os mapas gerados estão apresentados na Figura 6.7.



Reduto 1



Reduto 2

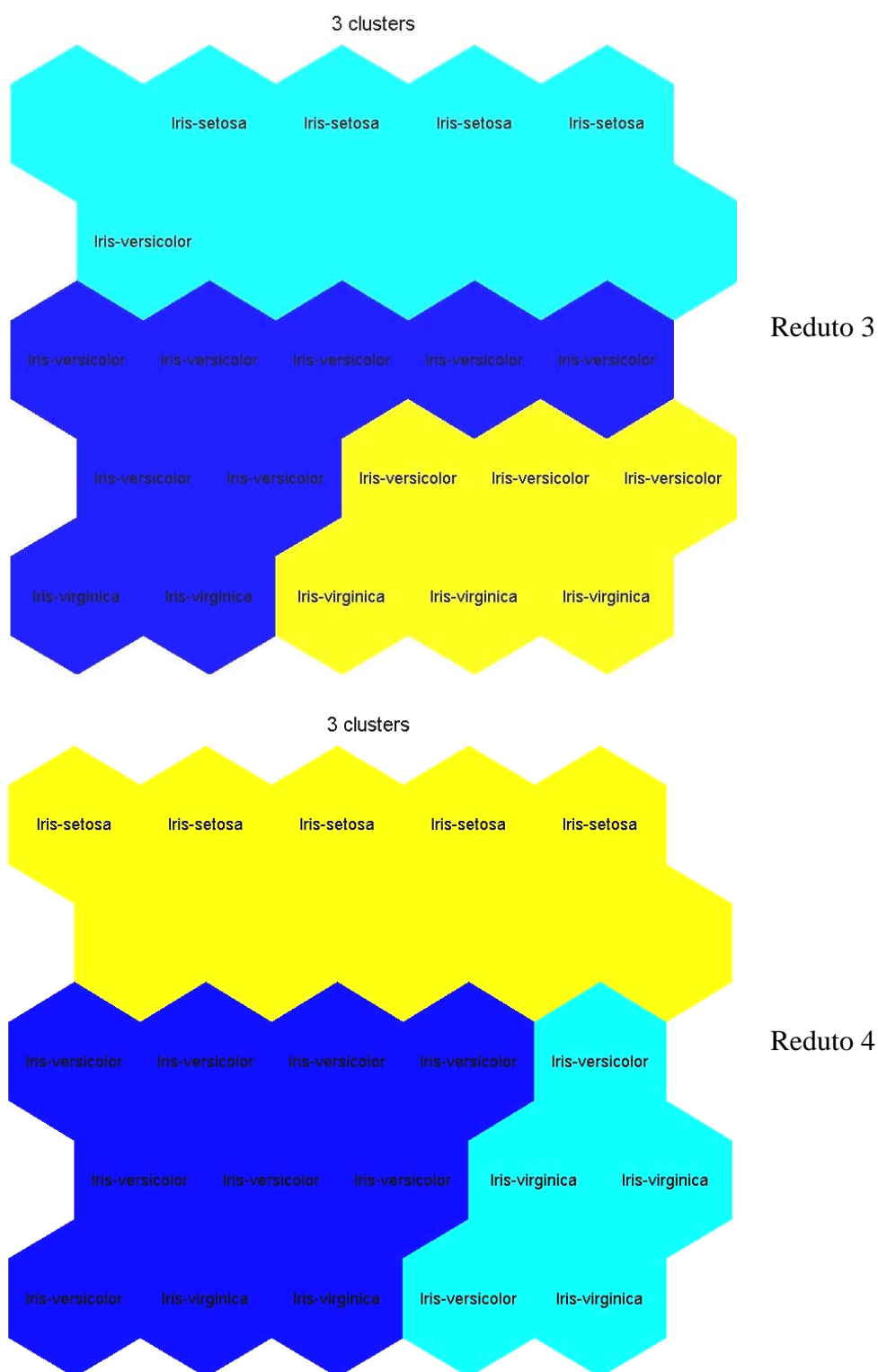


Figura 6.7: Os quatro Mapas gerados no Experimento 2 pela rede SOM com redutos (Arquitetura Híbrida).

Avaliação dos resultados

Como no experimento 1, ao sugerir os redutos, RS mostrou que não apenas reduziu informação não necessária à rede SOM como também utilizou dos conceitos de dependência e significância de atributos na geração dos redutos (capítulo 4, seções 4.3 e 4.4).

Observou-se que nos quatro mapas da Figura 6.7, foi possível defini-los com apenas 3 *clusters*, assim, a classe Íris-setosa, que foi dividida em duas na primeira fase do experimento (Figura 6.6), agora aparece unificada, e fica evidente que é ela a classe linearmente separável das outras duas. Por outro lado, as outras duas classes, Íris-versicolour e Íris-virgínica ainda, em alguns poucos casos, apresentam uma pequena sobreposição, o que pode ser avaliado como uma boa melhora, já que estas duas classes são consideradas não separáveis linearmente.

Isto ocorreu porque RS reduziu informação desnecessária da base de dados que era apresentada à rede SOM melhorando a similaridade entre os dados e reduzindo a diferença de separação dos agrupamentos (de 4 para 3 clusters). Novamente o desempenho da Arquitetura Híbrida foi melhor.

6.3.3 Experimento 3: Com a Base de Dados Pública ZOO

A base de dados ZOO [UCI98] possui um total de 101 amostras de animais com 16 atributos, sendo 1 atributo numérico (número de pernas) e 15 atributos binários (tem penas, voa, etc). O Conjunto é dividido em 7 classes, conforme a Tabela 6.3.

Classe	Descrição	Número de amostras
1	Mamíferos	41
2	Aves	20
3	Répteis	5
4	Peixes	13
5	Anfíbios	4
6	Insetos	8
7	Moluscos e Crustáceos	10

Tabela 6.3: Características da base de dados ZOO (Experimento 3).

A escolha da base de dados ZOO foi feita com base em três motivos descritos a seguir: é considerada uma base de dados “bem comportada” por possuir uma boa separação entre as classes [ZUC03], o que a torna diferente das bases de dados utilizadas nos experimentos 1 e 2. O número de classes (7) possibilita uma boa visualização e comparação de como as amostras foram alocadas no mapa na primeira fase do experimento e depois na segunda fase do experimento e, por último, com estas características, é provável que os *clusters* gerados pela rede SOM sejam de boa qualidade, surgindo então uma pergunta: “O mapa pode ser melhorado utilizando-se a Arquitetura Híbrida?”.

O número de épocas foi calculado da seguinte forma: *rough phase* = quantidade de neurônios (225) / tamanho de dados (101) x 10 = 23,0 e a *fine tune* = quantidade de neurônios (225) / tamanho de dados (101) x 40 = 90,0. Somando as duas fases (23,0 + 90,0) temos 113 épocas.

Primeira fase dos experimentos: rede SOM sem redutos

Na primeira fase do experimento foi apresentada à rede SOM a base de dados ZOO com todos os atributos (16) e verificaram-se os resultados através do EQ, do ET, do tempo de treinamento da rede e da visualização do mapa.

A Tabela 6.4 mostra os resultados obtidos após a apresentação da base de dados ZOO à rede SOM sem redutos.

Número de <i>clusters</i>	EQ	ET	Tempo
12	0.890	0.10	33s

Tabela 6.4: Resultados da rede SOM sem redutos (primeira fase do Experimento 3)

Na Figura 6.8 podem-se visualizar os 12 *clusters* gerados pela rede SOM sem redutos rotulados de acordo com as classes da base de dados ZOO.



Figura 6.8: Mapa gerado pela primeira fase do Experimento 3 onde os *clusters* podem ser identificados através de cores e as classes identificadas através das amostras rotuladas.

Segunda fase dos experimentos: rede SOM com redutos (Arquitetura Híbrida)

Na segunda fase do experimento submeteu-se a base de dados ZOO ao RS, gerando 5 redutos com 11 atributos cada e 1 reduto com 10 atributos.

Cada reduto foi de forma independente apresentado à rede SOM e verificou-se o seu desempenho através do EQ, do ET, do tempo de treinamento da rede e o número de *clusters* gerados (os dados estão na Tabela 7.5). Este procedimento foi necessário para verificar se os 6 redutos apresentavam desempenho aproximado.

A Tabela 6.5 mostra os resultados obtidos de cada reduto apresentado à rede SOM.

Reduto	Suporte	Comprimento	Número de clusters	EQ	ET	Tempo (segundos)
{ <i>nhair, eggs, airborne, aquatic, predator, backbone, venomous, legs, tail, domestic, catsize</i> }	100	11	10	0,762	0,040	28s
{ <i>nhair, eggs, airborne, aquatic, predator, breathes, venomous, legs, tail, domestic, catsize</i> }	100	11	10	0,782	0,040	31s
{ <i>nhair, eggs, airborne, aquatic, predator, toothed, venomous, legs, tail, domestic, catsize</i> }	100	11	10	0,771	0,020	32s.
{ <i>nhair, milk, airborne, aquatic, predator, toothed, venomous, legs, tail, domestic, catsize</i> }	100	11	10	0,769	0,000	32 s
{ <i>nhair, airborne, aquatic, predator, toothed, breathes, venomous, legs, tail, domestic, catsize</i> }	100	11	10	0,765	0,020	34 s
{ <i>nhair, eggs, airborne, aquatic, predator, venomous, legs, tail, Domestic, catsize</i> }	100	10	10	0,742	0,050	31 s

Tabela 6.5: Resultados da rede SOM com redutos (Arquitetura Híbrida) referentes ao Experimento 3.

Observando os resultados da Tabela 6.5, pode-se concluir que os seis redutos são iguais em número de *clusters* e se equivalem em tempo de treinamento da rede, no EQ e no ET. O reduto de número 4 foi desconsiderado em função de o erro topográfico ser zero, o que sugere *overfitting/underfitting*. Por apresentar o menor número de atributos o sexto reduto foi o escolhido [MIT97]. Na Figura 6.9 pode-se visualizar os 10 *clusters* rotulados de acordo com as classes da base de dados ZOO gerados pela rede SOM com redutos (Arquitetura Híbrida) utilizando o sexto reduto. Na Figura 6.9 pode-se visualizar os 10 *clusters* rotulados de acordo com as classes da base de dados ZOO gerados pela rede SOM com redutos (Arquitetura Híbrida) utilizando o sexto reduto.

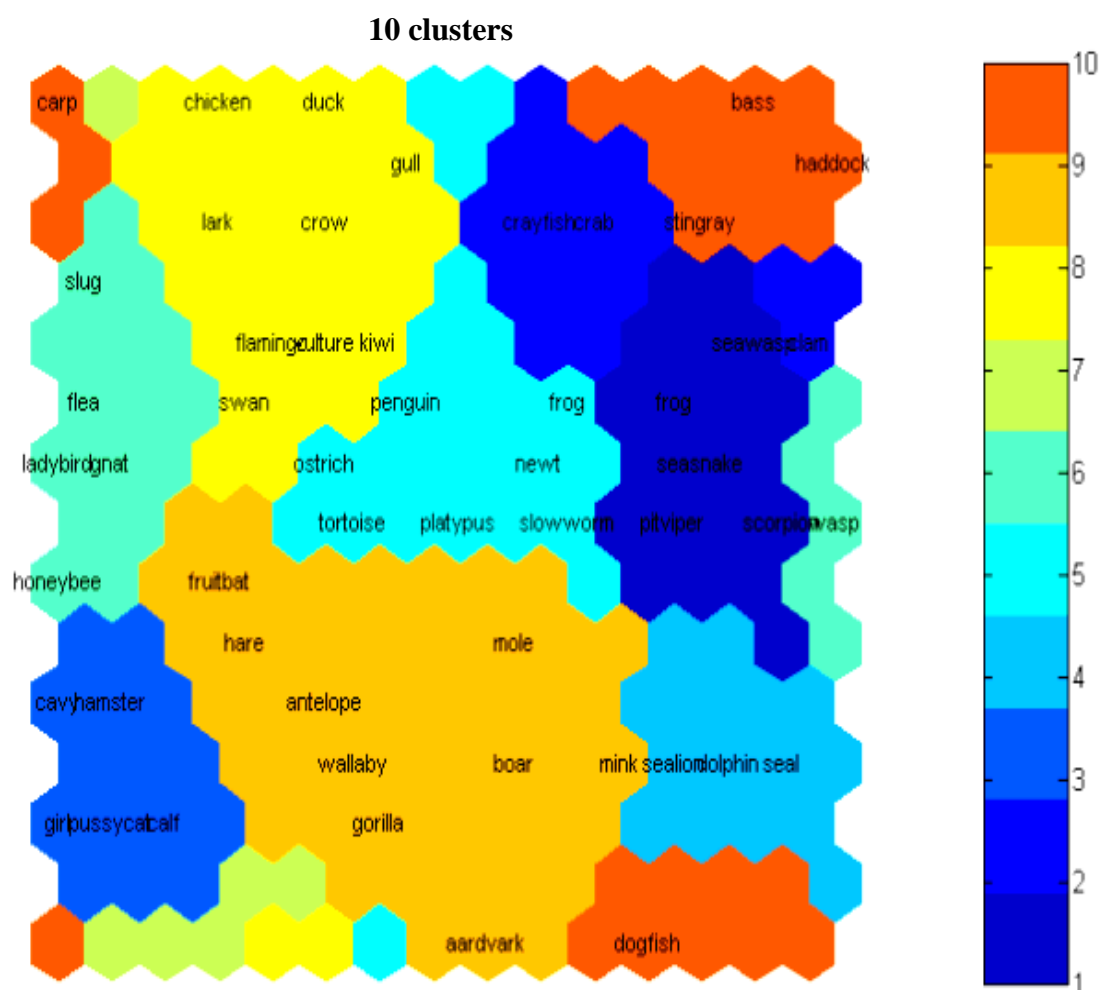


Figura 6.9: Mapa gerado pela segunda fase do Experimento 3, onde os *clusters* podem ser identificados através de cores e as classes identificadas através das amostras rotuladas.

Na Tabela 6.6 é feita a comparação dos resultados entre a rede SOM sem redutos com a Arquitetura Híbrida (escolhido o sexto reduto).

Rede	Número de <i>clusters</i>	Erro de Quantização	Erro Topográfico	Tempo
Rede SOM sem redutos	12	0,890	0,10	33s
Rede SOM com redutos (sexto reduto)	10	0,742	0,050	31 s

Tabela 6.6: Comparação entre a rede SOM e a Arquitetura Híbrida proposta (Experimento 3).

Analisando os números da Tabela 6.6 fica evidente o melhor desempenho da Arquitetura Híbrida sobre a rede SOM sem redutos.

Avaliação dos resultados

Em resposta à pergunta: “O Mapa Auto-Organizável pode ser melhorado utilizando a Arquitetura Híbrida?”, pode-se afirmar que sim, apesar de ser uma base de dados “bem comportada”. Os motivos estão relacionados a seguir:

a) Os valores de EQ e ET foram menores em todos os redutos se comparados aos resultados da rede SOM sem redutos, o que indica uma melhor representação da topologia da estrutura dos dados.

b) O tempo de treinamento foi ligeiramente menor.

c) O número de *clusters* gerados em todos os redutos foi menor, se comparado aos resultados da rede SOM sem redutos, indicando uma melhor similaridade tanto entre os dados quanto entre as diferenças de separação dos agrupamentos. Desta forma, os *clusters* apresentaram uma fronteira mais bem definida, melhorando a visualização. Isto ocorreu devido ao fato de a Arquitetura Híbrida (através do RS) ter reduzido informação não relevante que antes era apresentada à rede SOM (primeira fase do experimento) dificultando a formação das fronteiras entre os *clusters*. Dentre vários exemplos (Figuras 6.8 e 6.9) pode-se verificar que a Arquitetura Híbrida realizou uma melhor separação entre aves e mamíferos utilizando cores diferentes. Isto pode ser observado nas amostras *swan* (cisne) e *fruitbat* (morcego). Pode-se visualizar também que a rede SOM sem redutos teve dificuldades em separar os peixes colorindo os *clusters* com azul muito próximo de aves e mamíferos. Isto pode ser observado na amostra *carp* (carpa). No caso da Arquitetura Híbrida foi utilizada outra cor (marrom) para colorir e diferenciar o *cluster* dos peixes.

Como nos experimentos anteriores, a Arquitetura Híbrida proposta apresentou melhores resultados.

6.3.4 Experimento 4: Com a Base de Dados Consumidor

Nos experimentos anteriores, as bases de dados eram conhecidas em termos do número de classes e do número de amostras por classe. No caso da base de dados Consumidor disponível em SAS [SAS05], não se conhece o número de amostras por classe e devido ao número de atributos (48), fica difícil reconhecer o perfil dos consumidores através de métodos tradicionais, o que justifica a utilização do KDD. Assim, a escolha da base de dados Consumidor foi oportuna, pois todas as fases do KDD (capítulo 3, seção 3.1.1) puderam ser realizadas para a descoberta de conhecimento.

A base de dados Consumidor [FER03, SAS05] é uma base de dados que contém informações do consumo de 1.968 consumidores (registros) com 48 atributos, sendo 47 atributos condicionais e um atributo de decisão dividido em duas classes (M = masculino; F = feminino). No KDD é a chamada fase de seleção de dados.

Foi necessário pré-processar a base de dados eliminando o atributo que identificava o número da conta do cliente, o que poderia prejudicar a formação dos *clusters* pela rede SOM e a geração de redutos pelo RS. No KDD é fase de pré-processamento da base de dados. O atributo prefixo do nome é nominal e portanto foi transformado em numérico para poder ser processado pela rede SOM. A transformação foi realizada da seguinte forma: Ms = 0; Mr = 1; Mrs = 2 e None = 3. No KDD é a fase de transformação dos dados.

Na primeira fase do experimento apresentou-se à rede SOM a base de dados com todos os atributos condicionais (47) e os resultados foram avaliados através do EQ, do ET, do tempo de treinamento da rede SOM e da visualização do mapa.

Na segunda fase do experimento submeteu-se primeiramente a base de dados com todos os atributos condicionais (47) ao RS para reduzir o número de atributos gerando os redutos e em seguida a base de dados Consumidor reduzida foi submetida à rede SOM.

Nesta parte do experimento fica claro que o processo de KDD é iterativo e iterativo (capítulo 3), pois, apesar de a fase de transformação já ter sido realizada, retornou-se à fase anterior de pré-processamento para o RS reduzir os atributos da base de dados antes de apresentá-los à rede SOM.

Foram gerados 87 redutos, sendo 16 redutos com 3 atributos, 63 redutos com 4 atributos e 8 redutos com 5 atributos.

A escolha do reduto baseou-se no menor esforço computacional [MIT97]. Escolheram-se então, os 16 redutos que apresentaram um número menor de atributos (3), em detrimento dos outros redutos que apresentaram 4 ou 5 atributos. Dentre os 16 menores redutos a escolha foi aleatória já que apresentaram resultados semelhantes. O reduto escolhido foi o formado pelos seguintes atributos: valor da casa do consumidor, frequência de pedidos e idade do consumidor.

O número de épocas foi calculado da seguinte forma: *rough phase* = quantidade de neurônios (225) / tamanho de dados (1968) x 10 = 12,0 e a *fine tune* = quantidade de neurônios (225) / tamanho de dados (1968) x 40 = 58,0. Somando as duas fases (12,0 + 46,0) temos 58 épocas.

Primeira fase dos experimentos: rede SOM sem redutos

A primeira fase do experimento foi realizada com a apresentação da base de dados à rede SOM sem redutos. No processo de KDD é a fase chamada de *Data Mining*. Na Figura 6.10 podem-se visualizar os 14 *clusters* gerados pela primeira fase do experimento.

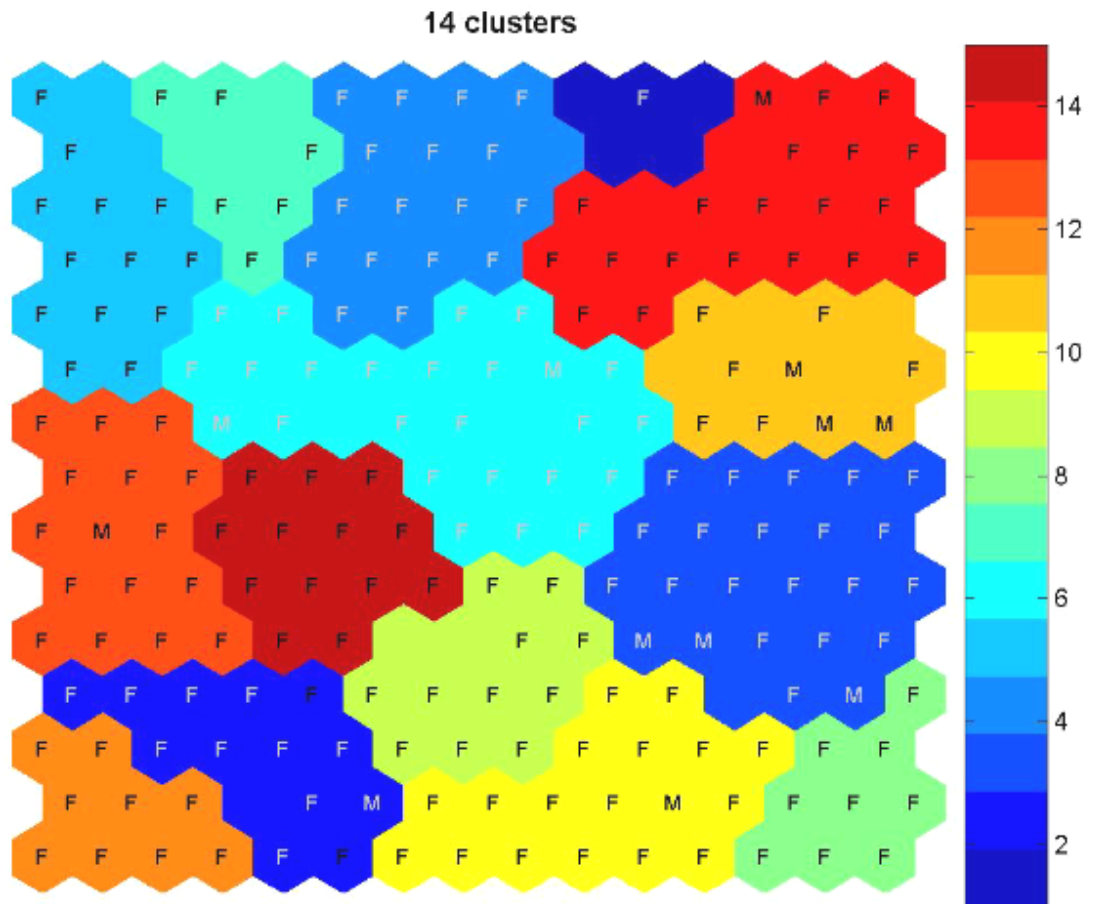


Figura 6.10: Visualização dos 14 *clusters* gerados pela rede SOM sem redutos, rotulados de acordo com as classes da base de dados Consumidor (primeira fase do Experimento 4).

Segunda fase dos experimentos: rede SOM com redutos (Arquitetura Híbrida)

Na segunda fase do experimento a base de dados Consumidor foi reduzida pelo RS e depois apresentada à rede SOM (Arquitetura Híbrida). No processo de KDD é a fase chamada de *Data Mining*. Na Figura 6.11 pode-se visualizar os 9 *clusters* gerados pela segunda fase do experimento.

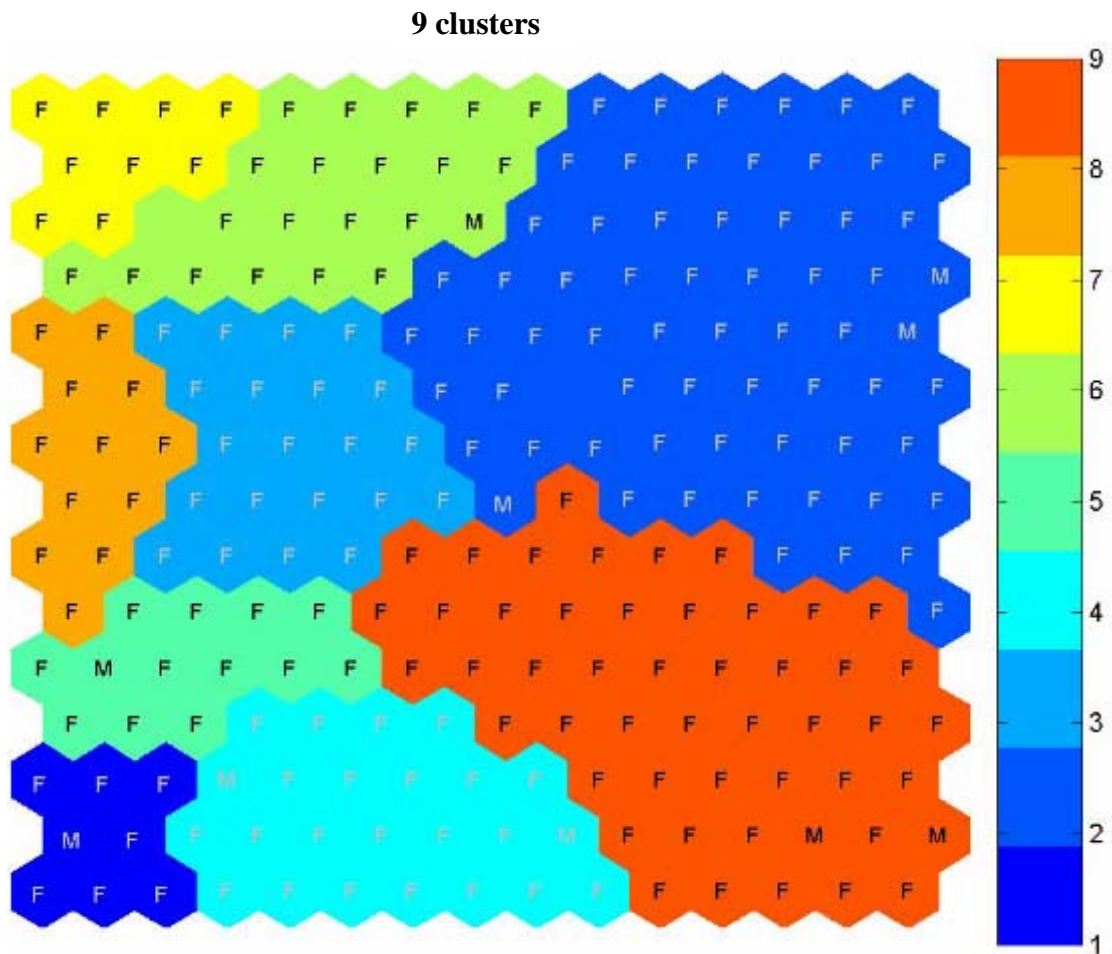


Figura 6.11: Visualização dos 9 *clusters* gerados pela rede SOM com redutos (Arquitetura Híbrida) rotulados de acordo com as classe da base de dados Consumidor (segunda fase do Experimento 4).

Uma forma de interpretar e avaliar os resultados extraídos pelo *Data Mining* é utilizar técnicas de visualização [BIG96; HAN01].

As Figuras 6.10, 6.11 e 6.12 são mapas que mostram as características dos *clusters* gerados pela rede SOM, auxiliando na fase de interpretação e avaliação dos resultados desta fase do KDD.

A Tabela 6.7 mostra a comparação entre as duas fases do experimento 4.

	Número de <i>clusters</i>	EQ	ET	Tempo
SOM sem redutos	14	4,688	0,055	4s
SOM com redutos (Arquitetura Híbrida)	9	0,251	0,039	3s

Tabela 6.7: Números do Experimento 4 comparando a rede SOM sem redutos com a rede SOM com redutos (Arquitetura Híbrida).

6.3.4.1 Descoberta de Conhecimento no Experimento 4

Não é a finalidade deste trabalho avaliar o conteúdo dos *clusters* gerados pela rede SOM e sim a qualidade de formação desses *clusters* pela Arquitetura Híbrida comparada com a rede SOM. Essa investigação poderia ficar a cargo de um classificador como por exemplo: uma outra arquitetura de rede neural artificial [VAN97], o próprio RS [SAS06], Algoritmos Genéticos [CAR99] ou utilizar-se de ferramentas para descobrir associações [GOL05].

Entretanto, em caráter didático procurou-se mostrar na Figura 6.12 que a Arquitetura Híbrida por intermédio da rede SOM agrupou os registros da base de dados Consumidor obedecendo aos critérios de igualdade ou de semelhança entre os registros. A rede SOM utilizada no experimento é uma rede neural artificial 15 X 15, ou seja, ela possui 225 neurônios mostrados na Figura 6.12, agrupados em nove *clusters* contendo os registros dos consumidores (F ou M), por igualdade ou semelhança. Foram destacados aleatoriamente 4 neurônios (1, 15, 90 e 225) para verificar se realmente a rede SOM agrupou segundo esses critérios.

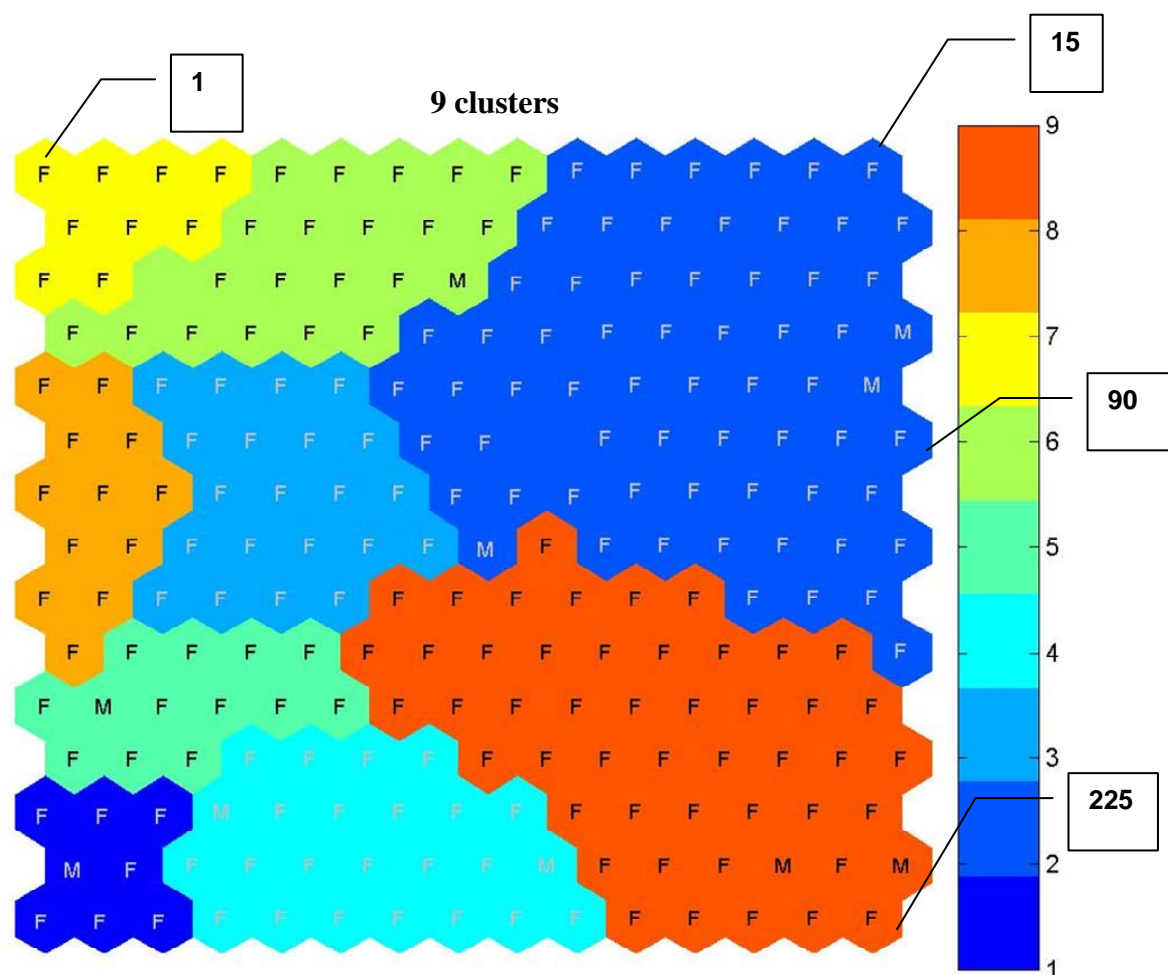


Figura 6.12: Visualização dos neurônios localizados no mapa gerado pela Arquitetura Híbrida (Experimento 4).

A Tabela 6.8 mostra como RS com a rede SOM (Arquitetura Híbrida) agrupou os registros da base de dados Consumidor de acordo com os valores de cada atributo: valor da casa do consumidor, a frequência de pedidos e a idade do consumidor (reduzido utilizado).

Neurônio	Valor da casa	Frequência de pedidos	Idade	sexo
1	98700	4	43	F
15	46300	3	44	F
90	55500	1	33	F
225	142900	2	29	F

Tabela 6.8: Informações dos atributos (reduzido) consideradas pela rede SOM para agrupar os neurônios. Pode-se verificar na tabela que o neurônio 15 e o neurônio 90 (destacados em negrito) têm atributos com valores semelhantes e por isso pertencem ao mesmo *cluster*.

Pode-se concluir que a rede SOM agrupou os registros obedecendo aos critérios de igualdade e semelhança de atributos (Figura 6.12). O neurônio 1 pertencente ao *cluster* de cor amarela mostra que este é um *cluster* de consumidores que possuem residência com valor intermediário, é um cliente com um número mais alto de pedidos e de meia-idade. Os neurônios 15 e 90 pertencentes ao mesmo *cluster* de cor azul mostram que este é um *cluster* de consumidores que possuem um valor de residência mais baixo, frequência de pedidos num intervalo médio e de meia-idade. O neurônio 225 pertencente ao *cluster* de cor vermelha mostra que o *cluster* é de consumidores que possuem casa com valor mais alto, frequência de pedidos média e são mais jovens. As conclusões a respeito de alto, médio e baixo valor da casa, frequência de pedidos alta, média e baixa, e meia-idade ou jovem foram feitas com base na Tabela 6.8.

Percebe-se que a base de dados Consumidor possui um número muito maior de casos do sexo feminino do que do sexo masculino. Isto poderia ser facilmente percebido em bases de dados pequenas, porém, a dificuldade seria maior em relação a bases de dados com milhões de registros.

Este fato justifica os *clusters* apresentarem um número maior de letras F (feminino) do que M (masculino) e a colocação no mesmo *cluster* dos dois sexos sugere perfis de consumo iguais ou muito semelhantes, que podem ser verificados com uma maior profundidade utilizando um classificador ou ferramentas de associação, como já foi comentado. No processo de KDD esta fase também é conhecida como pós-processamento. Após a fase de mineração de dados é necessária

a interpretação do conhecimento descoberto, ou algum processamento desse conhecimento.

Pode-se verificar neste experimento uma típica aplicação do KDD em *marketing* buscando descobrir segmentos de clientes na base de dados (capítulo 2, seção 2.4).

Avaliação dos resultados

Neste experimento todas as fases do processo de KDD foram realizadas.

A aplicação da Arquitetura Híbrida na base de dados Consumidor reduziu os valores de EQ e ET, indicando uma melhor representação da topologia da estrutura dos dados.

O tempo de treinamento da rede SOM também foi menor.

Houve também uma redução no número de *clusters*, indicando uma melhor similaridade tanto entre os dados quanto entre as diferenças de separação dos agrupamentos.

Os resultados levaram a concluir que o RS reduziu informação que era apresentada à rede SOM melhorando a formação dos *clusters*. Assim, houve melhora na visualização do mapa em função de melhor definição das fronteiras.

A melhora na formação dos *clusters* possibilita um maior entendimento da base de dados da organização a fim de realizar atividades como: direcionamento nas campanhas de vendas, promoção de ofertas combinadas de serviços ou produtos, avaliação do comportamento do mercado e detecção de novas tendências mercadológicas ou necessidades de consumo.

O experimento possibilitou também maior conhecimento da base de dados: sabe-se que o número de registros do sexo feminino é muito maior do que do sexo masculino e existem perfis de consumo semelhantes entre os dois sexos.

Pode-se concluir novamente que a Arquitetura Híbrida apresentou melhores resultados do que a rede SOM sem redutos.

6.3.5 Experimento 5: Com a Base de Dados *The Insurance Company*

Esta base de dados foi usada para a Competição de *Data Mining* ocorrida no ano de 2000 denominada “*The COIL 2000 Data Mining Competition*” [TIC00].

O “Desafio COIL 2000” foi organizado por um conjunto de redes de excelência custeadas pela União Européia, denominada COIL (*Computational Intelligence and Learning*), congregando esforços nas áreas de Redes Neurais Artificiais, Sistemas *Fuzzy*, Computação Evolucionária e Aprendizado de Máquina para extrair conhecimento da base de dados *The Insurance Company*. O conhecimento extraído possibilitaria entender melhor as características da base de dados auxiliando no desenvolvimento de estratégias de *marketing* mais eficazes.

O objetivo proposto na competição foi predizer quem seriam os potenciais consumidores interessados em seguros para *trailers*.

Foram gerados 29 artigos utilizando a base e que podem ser acessados no seguinte endereço eletrônico: <http://www.wi.leidenuniv.nl/~putten/library/cc2000/>.

É uma base de dados real da área de *marketing* composta por dados de consumo e dados sócio-demográficos de 5.822 consumidores com 86 atributos numéricos [PUT00].

A empresa *Dutch Data Mining Company Sentient Machine Research*, proprietária da base de dados, solicita o uso apenas para pesquisas não comerciais e para propósitos educacionais. Mais informações podem ser obtidas no seguinte endereço eletrônico: <http://www.liacs.nl/~putten/library/cc2000/> [PUT00].

A base de dados *The Insurance Company* foi incorporada em UCI (*Repository of Machine Learning Databases*). Irvine, CA: University of California, Department of information and Computer Science.

Os 5.822 registros contêm informações sobre 10 tipos de consumidores descritos a seguir: 1- consumidores bem-sucedidos, 2- produtores de hortifrutigranjeiros, 3- consumidores de classe média, 4- consumidores solitários, 5- consumidores felizes com a sua condição social, 6- consumidores da terceira idade

que gostam de viajar, 7- consumidores aposentados, 8-famílias sem crianças, 9-famílias conservadoras, 10- fazendeiros. Não se conhece a representatividade de cada tipo de consumidor dentro da base de dados. Por este motivo e pelo número de atributos (86) a base de dados *The Insurance Company* foi escolhida.

Primeira fase dos experimentos: rede SOM sem redutos

Na primeira fase do experimento apresentou-se à rede SOM a base de dados com todos os atributos (85) e os resultados foram avaliados através do EQ, ET, do tempo de treinamento da rede, do número de *clusters* gerados e da visualização do mapa. A Figura 6.13 mostra os 15 *clusters* gerados pela rede SOM sem redutos.

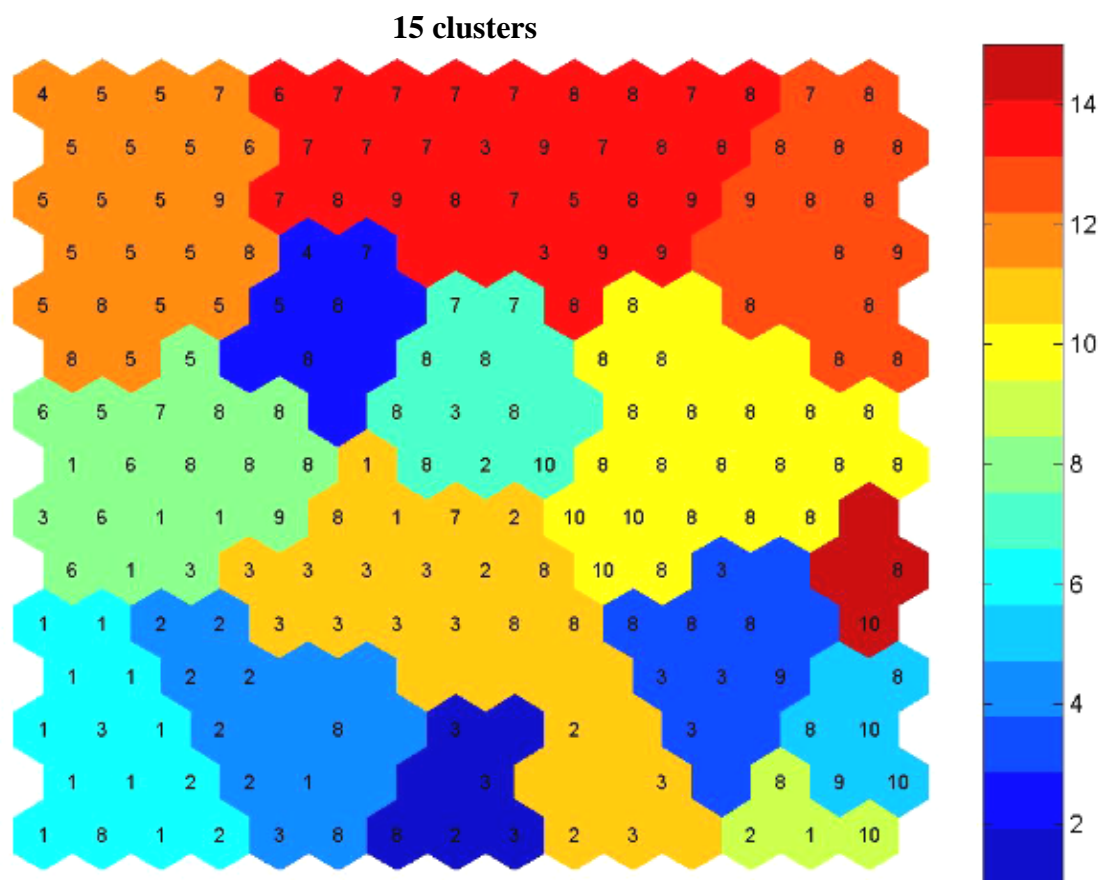


Figura 6.13: Visualização dos *clusters* gerados pela rede SOM sem redutos (primeira fase do Experimento 5).

Vale ressaltar que nas Figuras 6.13 e 6.14 os *clusters* estão rotulados de acordo com as classes da base de dados *The Insurance Company* e os números na barra ao lado do mapa representam apenas a quantidade de *clusters* presentes no mapa.

Segunda fase dos experimentos: rede SOM com redutos (Arquitetura Híbrida)

Na segunda fase do experimento submeteu-se a base de dados ao RS para reduzir o número de atributos e gerar os redutos; a base de dados reduzida foi submetida à rede SOM.

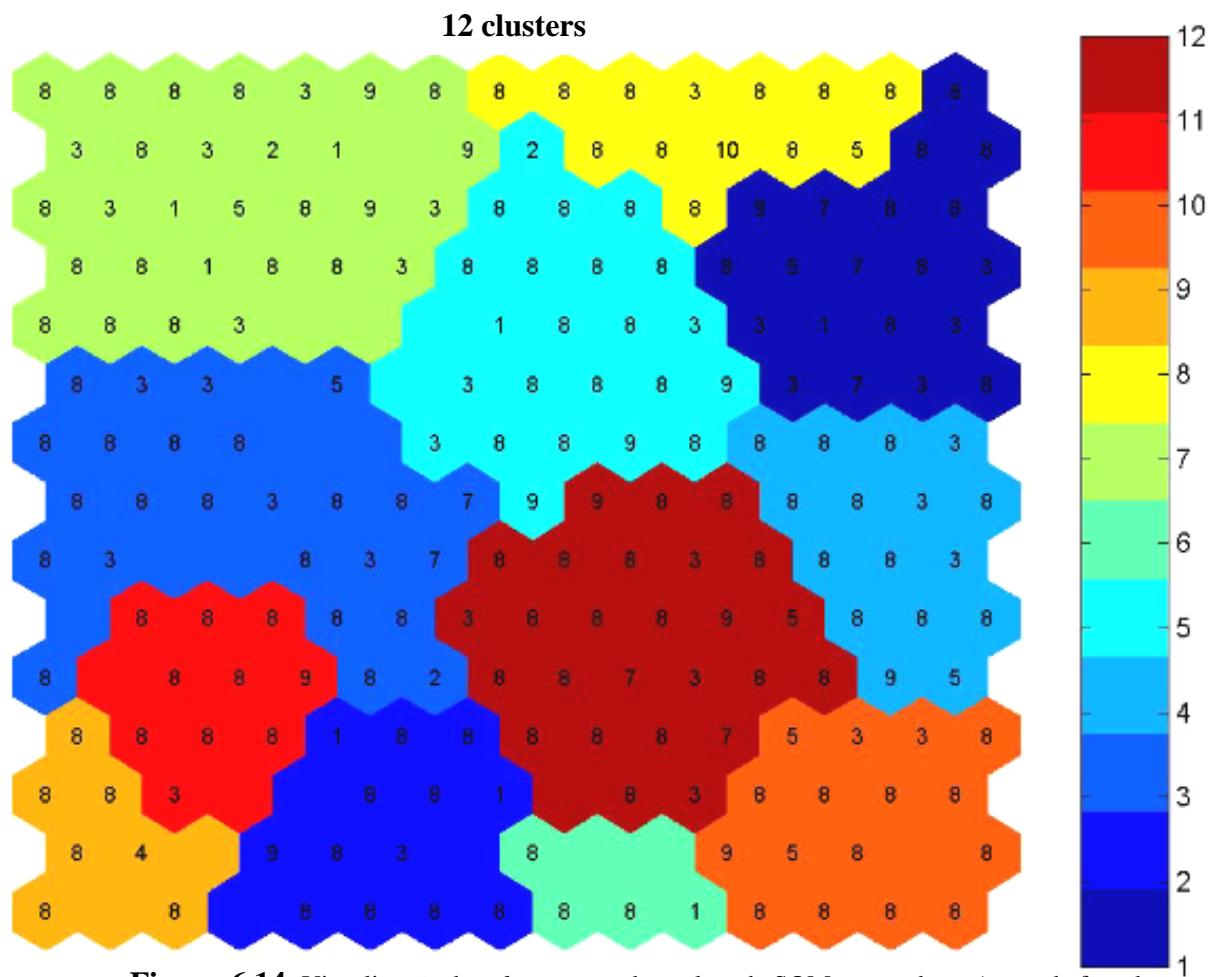


Figura 6.14: Visualização dos *clusters* gerados pela rede SOM com redutos (segunda fase do Experimento 5)

Foram gerados 153 redutos, sendo o menor reduto com 51 atributos e o maior com 67 atributos. Escolheu-se o reduto com menor número de atributos (51) pelo menor esforço computacional requerido [MIT97].

Na Figura 6.14 pode-se visualizar os 12 *clusters* gerados pela rede SOM com redutos (Arquitetura Híbrida).

Os resultados comparando as duas fases do experimento são apresentados na Tabela 6.9.

	Número de <i>clusters</i>	EQ	ET	Tempo
SOM sem redutos	15	6,174	0,064	9s
SOM com redutos (Arquitetura Híbrida)	12	5,025	0,060	6s

Tabela 6.9: Números Comparativos da rede SOM sem redutos com a rede SOM com redutos (Arquitetura Híbrida) referentes ao Experimento 5.

O número de épocas foi calculado da seguinte forma: *rough phase* = quantidade de neurônios (225) / tamanho de dados (5.822) x 10 = 4,0 e a *fine tune* = quantidade de neurônios (225) / tamanho de dados (5.822) x 40 = 16,0. Somando as duas fases (16,0 + 4,0) temos 20 épocas.

Avaliação dos resultados

Como nos experimentos anteriores, a aplicação da Arquitetura Híbrida na base de dados *The Insurance Company* reduziu os valores de EQ, de ET, o tempo de treinamento da rede SOM, o número de *clusters* gerados e a visualização do mapa, caracterizando novamente uma melhoria na representação da topologia da estrutura dos dados, na similaridade entre os dados e na diferença de separação dos agrupamentos.

Os resultados levaram a concluir que o RS reduziu informação que era apresentada à rede SOM melhorando a formação dos *clusters*. Assim, houve melhora na visualização do mapa em função de melhor definição das fronteiras.

No caso da rede SOM sem redutos (Figura 6.13), os *clusters* de maneira geral, ficaram espalhados no mapa, indicando dificuldades na separação dos agrupamentos e na similaridade entre os dados.

Com a Arquitetura Híbrida (Figura 6.14), o número de *clusters* é menor, o que indica um bom resultado.

A realização do experimento possibilitou também um melhor conhecimento da base de dados; sabe-se que o número de famílias sem crianças (tipo de consumidor número 8) é muito alto e provavelmente com características muito diversas fazendo-se presente em todos os *clusters*. Este fato pode ser justificado, pois entendem-se como famílias sem crianças as famílias sem filhos, as famílias formadas somente por adultos (filhos crescidos) e aquelas com filhos adolescentes.

Como no experimento 4, pode-se dizer que a melhora na formação dos *clusters* possibilita um maior entendimento da base de dados da organização a fim de realizar atividades como: direcionamento nas campanhas de vendas, promoção de ofertas combinadas de serviços ou produtos, avaliação do comportamento do mercado e detecção de novas tendências mercadológicas ou necessidades de consumo

Pode-se concluir que pré-processar a rede SOM com RS gera *clusters* mais definidos e de melhor visualização, o que possibilitou um melhor desempenho da Arquitetura Híbrida.

6.4 Análise Final dos Resultados

Nos experimentos 1 e 2 foi visto que RS leva em consideração os conceitos de dependência e significância de atributos discutidos no capítulo 4 para gerar os redutos. Isto leva a concluir que, apesar da redução de informação, os atributos selecionados que compõem os redutos conseguem representar a base de dados original (com todos os atributos).

A continuação dos experimentos 1 e 2 compara o desempenho da rede SOM sem redutos com a Arquitetura Híbrida proposta (RS com a rede SOM). A avaliação

visual dos mapas gerados mostrou um melhor desempenho da Arquitetura Híbrida, que reduziu a sobreposição das classes gerando fronteiras mais bem definidas entre os *clusters*.

Estes dois experimentos iniciais foram realizados com bases de dados simples e tiveram uma importância vital para o desenvolvimento deste trabalho, pois, se RS não gerasse redutos preservando a representatividade de todos os atributos da base de dados inicial ou o mapa gerado pela Arquitetura Híbrida fosse inferior ao mapa gerado pela rede SOM sem redutos, este trabalho não teria continuidade.

No experimento 3 a base ZOO foi escolhida por possuir todos os atributos numéricos, não ter dados faltantes e ser composta por várias classes. Com estas características provavelmente a rede SOM sem redutos deveria gerar um mapa com *clusters* de boa qualidade. Assim, abordou-se a questão de se a Arquitetura Híbrida conseguiria melhorar o mapa.

Comparadas as duas fases do experimento 3, a Arquitetura Híbrida apresentou os melhores resultados de EQ, ET, do tempo de treinamento da rede SOM e da visualização do mapa.

O experimento 4 teve como característica o não-conhecimento do número de amostras por classe e um número maior de atributos com relação aos experimentos 1, 2 e 3, portanto, sendo difícil conhecer o perfil dos consumidores. As duas fases foram comparadas analisando-se os resultados do EQ, do ET, do tempo de treinamento da rede SOM e da visualização do mapa. Novamente a Arquitetura Híbrida apresentou os melhores resultados.

No experimento 5 não se conhece a representatividade de cada tipo de consumidor (10) dentro da base de dados e o número de atributos é alto (86). Também neste experimento os resultados da Arquitetura Híbrida foram melhores, comparados aos da rede SOM sem redutos.

Em geral quando se fala de redução de dados sempre vem à mente um ganho em tempo de processamento, o que nos experimentos realizados neste trabalho ocorreu sensivelmente. Entretanto, a redução realizada pelo RS fez com que informação considerada desnecessária não fosse apresentada à rede SOM,

melhorando as fronteiras entre os *clusters* gerados pela rede SOM através de uma melhor separação dos agrupamentos, uma melhor similaridade entre os dados (número menor de *clusters*) e uma melhor representação da topologia da estrutura dos dados (EQ e ET menores).

Deve-se considerar que uma melhor definição das fronteiras dos *clusters* após a redução realizada pelo RS reduz a incerteza sobre as informações, tornando os conceitos de Aproximação Inferior, Aproximação Superior, Região de Borda e Região Negativa mais precisos melhorando o desempenho de um classificador.

A Arquitetura Híbrida proposta teve um desempenho superior ao apresentado pela rede SOM sem redutos em todos os cinco experimentos.

7 Conclusão

Os avanços da Tecnologia da Informação têm possibilitado o armazenamento da ordem de terabytes (TB) de informação.

Sem o auxílio de ferramentas computacionais apropriadas, a análise destas grandes quantidades de informação é inviável. Surge então uma questão: “Como aproveitar essa grande quantidade de informações armazenadas em benefício das organizações e no apoio à decisão de uma forma geral?”.

Por essa razão, faz-se necessário o uso de sistemas que possam extrair o conhecimento dessas bases, viabilizem a análise dessas informações, denominados KDD (*Knowledge Discovery Databases*) ou Descoberta de Conhecimento em Bases de Dados.

Motivado por este cenário, desenvolveu-se neste trabalho uma Arquitetura Híbrida que combina duas técnicas: a Teoria dos *Rough Sets* (Teoria dos Conjuntos Aproximados) e as redes SOM (*Self-Organizing Maps* – Mapas Auto-Organizáveis), a fim de descobrir conhecimento em bases de dados. Na Arquitetura Híbrida, a função dos RS é a de redução de atributos (dimensão) e a função da rede SOM é a de gerar *clusters*.

Assim, no primeiro capítulo deste trabalho realizou-se uma introdução conceitual dos principais assuntos abordados, além da justificativa e do objetivo que levaram ao seu desenvolvimento. No segundo capítulo foi abordada a importância da informação para as organizações e a necessidade de descobrir conhecimento para apoiar a tomada de decisão.

No terceiro capítulo, estudou-se o processo de descoberta de conhecimento em bases de dados, denominado KDD, a fim de embasar o desenvolvimento da Arquitetura Híbrida proposta.

No quarto capítulo foi apresentado o conceitual e o embasamento matemático que envolvem a Teoria dos *Rough Sets*, e no quinto capítulo estudaram-se os principais conceitos das redes neurais artificiais com ênfase na rede SOM.

No sexto capítulo, a Arquitetura Híbrida proposta foi aplicada em cinco bases de dados diferentes, e o seu desempenho avaliado através do erro de quantização (EQ), do erro topográfico (ET), do tempo de treinamento da rede SOM, do número de *clusters* e do mapa gerado utilizando o Mapa por Similaridade de Cor.

7.1 Sumário das Contribuições

Inicialmente, por se tratar de uma temática emergente, esforços foram feitos nas revisões de literatura e em particular na formulação do texto do capítulo 4 que aborda RS, sendo necessária a padronização da notação matemática objetivando maior clareza.

Em um segundo momento, o trabalho foi desenvolvido no sentido de conseguir combinar RS com a rede SOM, numa abordagem voltada à mineração de dados em que uma das formas de visualização do mapa gerado, denominada Mapa por Similaridade de Cor, foi também utilizada como um dos indicadores do desempenho da Arquitetura Híbrida na descoberta de conhecimento.

Como citado em Goldschmidt [GOL05], técnicas podem ser combinadas para gerar as chamadas Arquiteturas Híbridas. A grande vantagem desse tipo de sistema deve-se ao sinergismo obtido pela combinação de duas ou mais técnicas. Este sinergismo resulta na obtenção de um sistema mais poderoso (em termos de interpretação, de aprendizado, de estimativa de parâmetros, de generalização, dentre outros) e com menos deficiências.

Foi o que ocorreu na comparação do desempenho entre a rede SOM sem redutos (primeira fase de cada experimento) com a Arquitetura Híbrida proposta (rede SOM com redutos, segunda fase de cada experimento). O desempenho foi maior na Arquitetura Híbrida e pode ser verificado no capítulo 6 na avaliação dos resultados (seções 6.3.1, 6.3.2, 6.3.3, 6.3.4 e 6.3.5) e na análise final dos resultados (seção 6.4), onde a Arquitetura Híbrida apresentou, em todos os cinco experimentos, menor erro de

quantização (EQ), menor erro topográfico (ET), menor tempo de treinamento da rede SOM, menor número de *clusters* e melhor visualização do mapa gerado.

Os menores valores de EQ e ET indicam que a Arquitetura Híbrida conseguiu representar a topologia dos dados de entrada melhor do que a rede SOM sem redutos.

Como os *clusters* são obtidos por intermédio da aplicação dos conceitos de similaridade e distância [JOH98], um número menor de *clusters* indica melhor similaridade entre os dados e maior diferença de separação dos *clusters*, ou seja, alta similaridade intraclasse e baixa similaridade interclasses.

A análise visual do Mapa por Similaridade de Cor também indicou melhor visualização dos *clusters* gerados pela Arquitetura Híbrida.

Em geral quando se fala de redução de dados sempre vem à mente um ganho em tempo de processamento, o que nos experimentos realizados neste trabalho ocorreu de maneira sensível (menor tempo de treinamento da rede SOM). Entretanto, a redução de atributos realizada pelo RS ao pré-processar a rede SOM fez com que informação considerada desnecessária não fosse apresentada à rede, melhorando as fronteiras entre os *clusters* gerados. Esta informação desnecessária, quando submetida à rede SOM sem redutos, gerou incerteza, ocasionando em certos *clusters* uma definição de fronteira ruim, prejudicando a separação dos agrupamentos. Isto pode ser observado nos experimentos realizados com a rede SOM sem redutos no capítulo 6, seção 6.3.1 (Figura 6.5), seção 6.3.2 (Figura 6.6), seção 6.3.3 (Figura 6.8), seção 6.3.4 (Figura 6.10) e seção 6.14 (Figura 6.13).

A combinação de RS com a rede SOM em uma Arquitetura Híbrida fez com que uma das principais deficiências da rede SOM (a definição de fronteira entre os *clusters*) fosse reduzida, levando à conclusão de que em muitos casos é necessária a combinação de duas ou mais técnicas a fim de eliminar ou reduzir certas deficiências individuais de cada técnica.

O objetivo principal deste trabalho foi, através da análise dos resultados obtidos com a Arquitetura Híbrida proposta, procurar responder a uma questão fundamental:

“Existe melhoria nos resultados gerados pela rede SOM quando pré-processada por RS?”.

Assim, com base nos experimentos realizados e nos resultados apresentados, pode-se concluir que a Arquitetura Híbrida teve um desempenho superior ao da rede SOM sem redutos e responder **sim** à questão fundamental do trabalho, ou seja, existe melhoria nos resultados gerados pela rede SOM quando pré-processada por RS.

Acredita-se que, além do objetivo principal, o trabalho conseguiu atingir um outro objetivo, que foi o de possibilitar um maior conhecimento sobre RS como técnica de redução de atributos, analisando para tanto os conceitos principais e o formalismo matemático apresentados no capítulo 4.

A redução da incerteza e a conseqüente melhora na geração dos *clusters* obtida com a Arquitetura Híbrida possibilitam a formação de *clusters* mais confiáveis, pois os elementos da base de dados que estavam na região de fronteira dos *clusters* agora são agrupados melhor. Isto pode resultar numa geração de regras mais confiáveis por parte de um classificador ao ser utilizado após a submissão da base de dados à Arquitetura Híbrida.

Resumindo, as contribuições do presente trabalho foram as seguintes:

- Desenvolvimento de uma Arquitetura Híbrida (RS com rede SOM) que pode ser utilizada com mais vantagem em relação à rede Som sem redutos (capítulo 6) na descoberta de conhecimento em bases de dados.

- Maior conhecimento da Teoria dos *Rough Sets* (capítulo 4).

- A verificação do desenvolvimento de todo o processo de KDD realizado no Experimento 4 com a Base de Dados Consumidor (capítulo 6, seção 6.3.4).
- A verificação da fase de mineração de dados, onde a rede SOM se utiliza de certos critérios para agrupar os elementos e formar os *clusters* (capítulo 6, seção 6.3.4.1).
- A ilustração detalhada de como a Teoria dos *Rough Sets* gera os redutos (capítulo 4, seção 4.5).

Finalmente, com base nos resultados obtidos, pode-se considerar que a aplicação da Arquitetura Híbrida pode ser vantajosa em diversas áreas alvo de KDD, como: *marketing* (detecção de perfil do consumidor, de necessidades de consumo, de clientes que podem abandonar a fidelidade aos produtos ou serviços de uma organização, de tendências mercadológicas), medicina (imagens, diagnósticos), governamental (detecção de fraudes, perfil do sonegador de impostos), imagens espaciais (imagens de satélite, sistemas de informação geográfica, posicionamento por satélites), financeiras (concessão de crédito, descoberta de áreas seguras para investimentos), dentre outras.

7.2 Trabalhos Futuros

A aplicação da Arquitetura Híbrida pode ser realizada nos estudos citados no capítulo 4 (seção 4.9), bem como pode vir a gerar interessantes frutos no contexto dos seguintes trabalhos:

- Lingras [LIN96b], em que o algoritmo da rede SOM foi modificado com base nos conceitos de Aproximação Inferior e Superior dos RS buscando um melhor intervalo entre os *clusters*. Neste caso, a Arquitetura Híbrida reduziria informação desnecessária, a fim de melhorar a definição das regiões de borda dos *clusters*.

- Sankar [SAN03], que combina *Rough-Fuzzy* MLP em uma Arquitetura Híbrida. Aqui, poderia ser desenvolvida uma Arquitetura Híbrida *Rough* SOM-*Fuzzy* MLP.

- Lingras [LIN02], que utiliza RS com Algoritmos Genéticos para gerar *clusters* em mineração de dados na *Internet*. Neste caso, poderíamos substituir os Algoritmos Genéticos pela rede SOM.

- Pawlak [PAW98], aplicação da Arquitetura Híbrida em computação granular.

- Skowron [SKO98], aplicação da Arquitetura Híbrida em mereologia.

- Dar continuidade aos estudos da Arquitetura Híbrida proposta neste trabalho, incluindo RS como um classificador dos *clusters* gerados pela rede SOM, formando uma nova Arquitetura Híbrida RS-SOM-RS. O objetivo é verificar se haverá melhora na definição dos conceitos de Aproximação Inferior, Aproximação Superior, Região de Borda e Região Negativa.

- Verificar o desempenho da Arquitetura Híbrida proposta utilizando bases de dados que apresentem um número muito maior de registros do que as utilizadas nos experimentos deste trabalho.

- Otimização do número de neurônios da rede SOM.

Os estudos aqui realizados não têm a pretensão de esgotar o assunto, pelo contrário, buscou-se realizar uma contribuição com o desenvolvimento da Arquitetura Híbrida na descoberta de conhecimento em bases de dados. Sabe-se que existe uma clara demanda por estudos sistematizados que possam estabelecer outros domínios de aplicação ainda mais adequados para a Arquitetura Híbrida proposta. Este cenário oferece portanto amplo espaço para trabalhos de continuidade.

Referências Bibliográficas

- [ADJ01] ADJEI, O.; CHEN, L.; HENG-DA, C.; COOLEY, D.H.; CHENG, R. J.; TWOMBLY, X. A fuzzy search method for Rough Sets in Data mining IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th, volume 2, pages: 980–985.
- [AGR01] AGRAWAL R.; IMIELINSKI T.; SWAMI A. Mining association rules between Sets of itens in large databases. In: Proced, ACM SIGMOD Conference on Management of Data. Washington, DC, 2001.
- [ALH00] ALHONIEMI, E.; HIMBER, J.; PARHANKANGAS, J.; VESANTO, J. Som Toolbox V2 Beta. Helsinki University of Technology, Finland, 2000.
- [ALM93] ALMEIDA, F.C. L’Evaluation des risques de défaillance des entreprises à partir des réseaux de neurones insérés dans les systèmes d’aide à la décision – Grenoble: Universidade de Grenoble, Ecole Supérieure des Affaires, 1993. (Tese de doutorado em Ciências de Administração).
- [ALM95] ALMEIDA, F. C. Desvendando o Uso de Redes Neurais em Problemas de Administração de Empresas. Revista de Administração de Empresas EASP/FGV, volume 35, nº 1, páginas: 46-55, jan./fev. 1995.
- [ALM96] ALMEIDA, F. C. & DUMONTIER, P. O uso de Redes Neurais em Avaliação de Riscos de Inadimplência. Revista de Administração da USP, volume 31, nº 1, Jan./Mar. 1996.
- [ALM98] ALMEIDA, F. C.; ZANETI, L. A. Exploração do Uso de Redes Neurais na Previsão do Comportamento de Ativos Financeiros. III

- SEMEAD. Faculdade de Economia Administração e Contabilidade. USP. 21-23, Out. 1998.
- [ANG02] ANGELONI, M. T. Organizações do Conhecimento: infra-estrutura, pessoas e tecnologia. São Paulo: Saraiva, 2002.
- [ANS83] ANSOFF, H. I. Administração Estratégica. São Paulo. Atlas, 1983.
- [ARR01] ARRAES, D. Arquiteturas de Redes Neurais Aplicadas a Data Mining no Mercado Financeiro. Uma aplicação para a geração de Credit Ratings. V Congresso Brasileiro de Redes Neurais. Rio de Janeiro: páginas: 115-120, Abril 2-5, 2001.
- [AZE00] AZEVEDO, F. M.; BRASIL, L. M.; OLIVEIRA, R. C. L. Redes Neurais com Aplicações em Controle e em Sistemas Especialistas. Florianópolis: Visual Books, 2000.
- [BAC02] BACK, T. Adaptative Business Intelligence based on evolution strategies: some application examples of self-adaptative software. Information Sciences, volume 148, nº 1-4, pages: 113-121, may. 2002.
- [BAE94] BAETS, W.,R.,J.; VENUGOPAL, V. Neural Networks and Statistical Techniques in Marketing Research: A Conceptual Comparison. Marketing Intelligence & Planning, MCB University Press, v.12, número 7, pgs. 30-38.
- [BAN97] BANERJEE, M.; MITRA, S.; PAL, S. K.; Knowledge-based Fuzzy MLP with Rough Sets. Neural Networks, 1997, International Conference on, volume 1, 9-12 June 1997, pages: 499-504.
- [BAT03] BATISTA, A. P. A .E. G. Pré-processamento de Dados em Aprendizado de Máquina Supervisionado. Tese de Doutorado. Instituto de Ciências Matemáticas e de Computação, USP, São Carlos, 2003.
- [BAX90] BAXT, W. G. Use of an artificial neural network for data analysis in clinical decision-marking: The Diagnosis of Acute Coronary Occlusion. Neural Computing, volume 2, pages: 480-489, 1990.

- [BEA98] BEAUBOUEF, T.; LANG, R. A Rough Sets Techniques for uncertainly management in automated story generation. *Communications of the ACM*, 4:326-331.
- [BEL96] BELTRATTI, A.; MARGARITA, S. *Neural Networks for Economie and Financial Modelling*. Computer Press, 1996.
- [BEN03] BENITE, M. *Aplicação de modelos de redes neurais na elaboração e análise de Cenários Macroeconômicos*. Dissertação de Mestrado. ICMC – São Carlos, 2003.
- [BER97] BERRY, M. J. A.; LINOFF, G. *Data Mining Techniques: For Marketing, Sales and Customer Support*. John Wiley & Sons, 1997.
- [BEY01] BEYNON, M.; CURRY, B.; MORGAN, P. Knowledge Discovery in Marketing: An approach Rough Set theory, *European Journal of Marketing*, volume 35, nº 7/8, pages: 915-935, 2001.
- [BIG96] BIGUS, J. P. *Data Mining with Neural Network: Solving Business Problems from Applications Development to Decision Support*. Mcgraw-Hill, 1996.
- [BLA98] BLAKE, C. L. & MERZ, C. J. *UCI Repository of Machine Learning Databases*. Irvine, University of California, Department of Information and Computer Science. Disponível em <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [BON98] BONIKOWSKI, Z. Extensions and intentions in the Rough Set Theory. *Journal of Information Sciences*, pages: 149-167, 1998.
- [BON02] BONALDI, E. L.; DA SILVA, L. E. B.; LAMBERT-TORRES, G.; OLIVEIRA, L. E. L.; ASSUNCO, F.O. Using Rough Sets techniques as a fault diagnosis classifier for induction motors. *IECON 02 [Industrial Electronics Society]*, IEEE 2002 28th Annual Conference of the, volume: 4 , 5-8 Nov. 2002, pages: 3383–3388.
- [BOZ01] BOZI, K. Core searching on Rough Sets. *Information Technology Interfaces*, 2001. ITI 2001. Proceedings of the 23rd International Conference on , June 19-22, 2001, pages: 217–222.

- [BRA96] BRACHMAN, R. J.; KHABAZA, T.; KLOESGEN, W.; SIMOUDIS, E.; PIATETSKY, S. Mining Business Databases, Communications ACM, Special Issue on Data Mining volume 39, nº 11, Novembro 1996.
- [BRA00] BRAGA, A .P.; LUDERMIR, T. B.; CARVALHO, A .C. P. L. F. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro. LTC, 2000
- [CAB97] CABENA, P. et al. Discovering Data Mining: from concept to implementation. New Jersey : Prentice Hall, 1997.
- [CAB04] CABRAL, J. E.; GONTIJO, E. M. Fraud detection in electrical energy consumers using Rough Sets. Systems, Man and Cybernetics, 2004 IEEE International Conference on volume 4, 10-13 Oct. 2004, pages: 3625–3629.
- [CAR91] CARVALHO, L. A. V. Redes Neurais Artificiais: a volta do cérebro eletrônico? In: Revista Ciência Hoje. São Paulo, volume 12, nº 70, p.12-21, janeiro/ Fevereiro 1991.
- [CAR99] CARVALHO, D. R. Data Mining através de Indução de Regras e Algoritmos Genéticos, Dissertação de Mestrado em Informática Aplicada, PUCPR, PR, 1999.
- [CAR01] CARVALHO, L. A. V. Data Mining: A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração. Ed. Érica, São Paulo, 2001.
- [CHE96] CHEN, R.; LIN, T. Y. Supporting Rough set theory in very large databases using oracle RDBMS. Fuzzy Systems Symposium, 1996. Soft Computing in Intelligent Systems and Information Processing, Proceedings of the 1996 Asian 11-14 Dec. 1996, pages: 332–337.
- [CHE02] CHENG-YUAN, L.; YEN-TING, K. Economic states on neuron maps. Neural Information Processing, 2002. ICONIP 02. Proceedings of the 9th International Conference on, volume: 2, 18-22 Nov. 2002, pages: 787–791.

- [CHU03] CHUN-Yan, Y.; MING-Hui, W.; MING, W. Combining Rough Set Theory with Neural Network Theory for Pattern Recognition. Proceedings of the 2003 IEEE. International Conference on Robotics, Intelligent Systems and Signal Processing. Changsha, China, Oct 2003.
- [COR71] CORMACK, R. M. A Review of Classifications. Journal of Royal Statistical Society, Series A, 134, 1971, pages: 321-353.
- [COR97] CORNELLA, A. E. Es la Infonomía Y somos infonomistas. Extra! Net, 8 de noviembre de 1997.
- [COS99] COSTA, J. A. F. Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis. Tese de Doutorado. Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Elétrica e de Computação, 1999.
- [DAL03] DALAI, P. R.; JOSHI-INAMDAR, V. Clustering of automobile information using self organizing maps. TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, volume: 4 , Oct. 15-17, 2003, pages: 1267–1271.
- [DHA97] DHAR, V.; STEIN, R. Seven Methods for transforming Corporate Data into Business Intelligence. Upper Saddle River, NJ. Prentice Hall.1997.
- [DOU95] DOUGHERTY, J.; KOHAVI, R.; SAHAMI, M. Supervised and Unsupervised Discretization of Continuos Features. Proc. 12 th Int. Conf. Machine Learning, 194-202, 1995.
- [DUD00] DUDA, R.; HART P.; STORK, D. Pattern Classification and Scene Analysis. New York: John Wiley Profession, 2000.
- [DUT01] DUTRA,R.G. Aplicação de Métodos de Inteligência Artificial em Inteligência de Negócios. Dissertação de Mestrado. Escola Politécnica da Universidade de São Paulo. São Paulo, 2001.

- [EBE04] EBECKEN, N.F.F.E.; BREBBIA, C. A . ZANASI,A. Data Mining IV Fourth International Conference on Data Mining. WIT PRESS, USA, 2004.
- [FAY96a] FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, volume 39, pages: 27-34, 1996.
- [FAY96b] FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. UTHURUSAMY R. Advances in Knowledge discovery and Data Mining. MIT Press, 1996.
- [FEL99] FELIX, R.; USHIO, T. Rough Sets-based machine learning using a binary discernibility matrix. Intelligent Processing and Manufacturing of Materials, 1999. IPMM '99. Proceedings of the Second International Conference on, volume 1 , 10-15 July 1999, pages: 299–305.
- [FER03] FERNANDEZ, G. Data Mining using SAS Applications. New York. Chapman & Hall/CRC, 2003.
- [FRE99] FREITAS, A. A.; SILVA, I. N. Análise Técnica de Títulos Financeiros através de Redes Neurais Artificiais. Proceedings of the IV Brazilian Conference on Neural Networks – IV Congresso Brasileiro de Redes Neurais, pages: 67-71, july 20-22, 1999 – ITA, São José dos Campos, SP, Brasil.
- [FUJ98] FUJIMORI, S., KAIYA, T.; INOUE, T. Analysis of Discharge Currents with Discernibility Matrices. Proceedings of 1998 International Symposium on Electrical Insulating Materials. Japan, Sept. 27 – 30, 1998.
- [FUR94] FURLAN, J. D. Reengenharia da Informação, São Paulo. Ed. Makron Books, 1994.
- [GOL95] GOLAN, R. H.; ZIARKO, W. A Methodology for Stock Market analysis utilizing Rough Set Theory. Computational Intelligence for

- Financial Engineering, 1995. Proceedings of the IEEE/IAFE 1995, 9-11 April 1995, pages: 32–40.
- [GOL05] GOLDSCHMIDT, R.; PASSOS, E. Data Mining um guia prático. Conceitos, Técnicas, Ferramentas, Orientações e Aplicações. Ed. Campus, Rio de Janeiro, 2005.
- [GRZ01] GRZYMALA-BUSSE, P.; GRZYMALA-BUSSE, J. W.; HIPPE, Z. S. Melanoma prediction using data mining system LERS. Computer Software and Applications Conference, 2001. COMPSAC 2001. 25th Annual International , 8-12 Oct. 2001, pages: 615–620.
- [HAL01] HALKIDI, M., BATISTAKIS, Y., VAZIRGIANNIS, M. On Clustering Validation Techniques. Intelligent Information Systems Journal, Kluwer Publishers.
- [HAN01] HAN, J.; KAMBER, M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. San Francisco, Califórnia, 2001.
- [HAY99] HAYKIN, S. Neural Networks: A Comprehensive Foundation. New York: Willey & Sons, 1994.
- [HAW90] HAWLEY, D.; JOHNSON, J. D.; RAINA, D. Artificial Neural Systems: A New Tool for financial Decision-Marketing – Financial Analyst Journal, pages: 63-72, nov./dez., 1990.
- [HET99] HETTICH, S.; BAY, S. D. “The UCI KDD Archive”. Irvine, CA: University of California, Department of Information and Computer Science.
- [HSU99] HSU, W. H.; ANVIL, L. S.; POTTENGER, W. M.; TCHENG, D.; WELGE, M. Self-organizing Systems for Knowledge Discovery in Large Databases. IJCNN 1999. International Joint Conference on Neural Networks, 1999, volume 4 , 10-16 July 1999, pages: 2480–2485.
- [HU94] HU, X.; CERCONE, N. Discovery of Decision Rules in Relational Databases: A Rough Set Approach. CIKM 1994, page.9.

- [HU95] HU, X. Knowledge Discovery in DataBases: An Attribute-Oriented Rough Set Approach. Phd Thesis, University of Regina, 1995.
- [IJE00] I-JEN, C.; LIN, T. Y. Using Rough Sets to build-up Web-based one to one customer services. Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International , 25-27 Oct. 2000.
- [JAG98] JAGIELSKA, I. Hybrid rough Sets/neural network approach to the development of a decision support system. Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on volume 1, 4-9, May 1998, pages: 24–28.
- [JAI99] JAIN, A., MURTY, M., FLYNN, P. Data Clustering: A Review. ACM Computing Surveys 31(3): 264-323, 1999
- [JIA03] JIAJIN, H.; CHUNNIAN, L.; CHUANGXIN, O.; YAO, Y. Y.; ZHONG, N. Attribute reduction of Rough Sets in Mining Market value functions. WI 2003. Proceedings. IEEE/WIC International Conference on Web Intelligence, 2003. 13-17 Oct. 2003, pages: 470–473.
- [JIN04] JIN, H.; SHUM, W. H.; LEUNG, K. S.; WONG, M. L. Expanding self-organizing map for data visualization and cluster analysis, Information Sciences, 163, 2004, p. 157-173.
- [JOH94] JOHN, G.; KOHAVI, R. ; PFLEGER, K. Irrelevant Features and the Subset Selection Problem. In Morgan Kaufmann. Proceedings of the Eleventh International Conference on Machine Learning, San Francisco, CA, pp. 167 – 173.
- [JOH98] JOHNSON, R.A; WICHERN D.W., Applied Multivariate Statistical Analysis, 4^a ed, Prentice Hall, 1998.
- [KAS97] KASKI, S.; KOHONEM, T. Winner-Takes-All Networks. Triennial Report 1994 – 1996, Neural Networks Research Centre & Laboratory

- of Computer and Information Science, Helsinki University of Technology, Finland, 1997.
- [KAS98] KASKI, S.; NIKKILÄ, J.; KOHONEN, T. Methods for Interpreting a Self-Organized Map in Data Analysis. In: Michel Verleysen (editor), Proceedings of the 6th European Symposium on Artificial Neural Networks (ESANN'98), Bruges, April 22-24, D-Facto, Brussels, Belgium, pages: 185-190, 1998.
- [KAS99] KASKI, S.; VENNA, J.; KOHONEN, T. Coloring that Reveals High-Dimensional Structures in Data In Proceedings of ICONIP1999, 1999, to appear.
- [KIV95] KIVILUOTO, K. Topology Preservation in Self-Organizing Maps. Technical Report A 29, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1995.
- [KOH82] KOHONEN, T. Self-organized formation of topologically correct feature maps, Biological Cybernetics, 43, pages: 59-69, 1982.
- [KOH94] KOHAVI, R., FRASCA, B. Useful Feature SubSets and Rough Sets Reducts. Third International Workshop on Rough Sets and Soft Computing (RSSC 94).
- [KOH97a] KOHONEN, T. Exploration of very large databases by self-organizing maps International Conference on Neural Networks, 1997, volume 1, 9-12 June 1997, pages: PL1 - PL6.
- [KOH97b] KOHAVI, R. Wrappers for Feature Subset Selection. Artificial Intelligence 97, 273 – 324.
- [KOM97] KOMOROWSKI, J.; ØHRN, A. ROSETTA: A Rough Set Toolkit for Analysis of Data, Proc. Third International Joint Conference on Information Sciences, Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97), Durham, NC, USA, March 1-5, Vol. 3, pp. 403-407.

- [KOM98] KOMOROWSKI, J.; ØHRN, A.; SKOWRON, P. The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets: The ROSETTA System, In *Rough Sets in Knowledge Discovery: Methodology and Applications. Studies in Fuzziness and Soft Computing*, Vol. 18, Chapter 19, pp. 376-399, Physica-Verlag.
- [KOM99a] KOMOROVSKI, J; Ohrn, A. Modelling prognostic power of cardiac tests using Rough Sets. *Artificial Intelligence in Medicine*, pages: 167–191.
- [KOM99b] KOMOROWSKI, J.; PAWLAK, Z.; POLKOWSKI, L.; SKOWRON, A. *Rough Sets: A Tutorial Technical Report*. Warsaw University, december 1999.
- [KOM02] KOMOROWSKI, J.; ØHRN, A.; SKOWRON, A. The ROSETTA Rough Set Software System, In *Handbook of Data Mining and Knowledge Discovery*, W. Klösgen and J. Zytkow (eds.), ch. D.2.3, Oxford University Press.
- [KOT80] KOTLER, P. *Marketing*. São Paulo, Atlas, 1980.
- [KOV96] KOVÁCS, Z. L. *Redes Neurais Artificiais*. Segunda edição, editora Collegium Cognitio, 1996.
- [LAM02] LAMBERT-TORRES, G. Application of Rough Sets in power system control center data mining *Power Engineering Society Winter Meeting*, 2002. IEEE, volume 1, 27-31, Jan. 2002, pages: 627–631.
- [LAU99] LAUDON, K.; LAUDON, J. P. *Sistemas de Informação*. 4 ed. Rio de Janeiro: LTC, 1999.
- [LAZ02] LAZAR, A.; SETHI, I.K. *Decision Rule Extraction from Trained Neural Networks using Rough Sets*. Wayne State University. Department of Computer Science
- [LEE95] LEE, H. Y.; ONG, H. L. Exploring visualization in knowledge discovery. *Proc. 1st Int. Conf. Knowledge Discovery and Data Mining (KDD-95)*, 198-203. AAAI, 1995.

- [LEE01] LEE, D. H. Seleção e Construção de Features Relevantes para o Aprendizado de Máquina. Dissertação de Mestrado, Instituto de Ciências Matemáticas e da Computação. Universidade de São Paulo, São Carlos, 2001.
- [LEN99] LENSU, A.; KOIKKALAINEN, P. Similar document detection using Self-Organizing Maps Third International Conference on Knowledge-Based Intelligent Information Engineering Systems, 1999. 31 Aug.-1 Sept. 1999.
- [LIN96a] LINGRAS, P. Rough Neural Networks. Sixth International Conferences Information Processing and Management of Uncertainty in Knowledge-Based Systems, Proceedings (IPMU96), volume II, july 1-5. Grenada, pages: 1445-1450, 1996.
- [LIN96b] LINGRAS, P. Unsupervised learning using Rough Kohonem Neural Networks Classifiers, Proceedings Of Symposium. On Modeling, Analysis and Simulation, CESA 96. Lille France, pages:.753-757, 1996.
- [LIN97] LIN, T. Y.; CERCONE, N. Rough Sets and Data Mining: analysis of Imprecise Data. Kluwer Academic Publishers, 1997.
- [LIN02] LINGRAS, P. Rough set clustering for Web mining FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, 2002, volume 2, 12-17 May 2002, pages: 1039–1044.
- [LO97] LO, J. Y.; FLOYD, C. E. Self-organizing maps for analyzing mammographic findings. International Conference on Neural Networks, 1997, volume 4 , 9-12 June 1997, pages: 2472–2474.
- [LON02] LONGHAN, C.; CHANGXIU, C.; ZHEN, G.; JIANYONG, L. The research of fault diagnosis for fuel injection system of diesel engine with ANN based on Rough Sets theory Proceedings of the 4th World Congress on Intelligent Control and Automation, 2002, volume 1, 10-14 June 2002, pages: 410–414.
- [MAS99] MASI DE, D. A Sociedade Pós-Industrial. São Paulo: Senac, 1999.

- [MAT82] MATTOS, J. M. A Sociedade do Conhecimento. Editora da Universidade de Brasília, 1982.
- [MCG95] MCGEE, J.; PRUSAK, L. Gerenciamento Estratégico da Informação. Rio de Janeiro, Campus, 1995.
- [MEN70] MENDEL, J. M. ; MCLAREN, R. W. Reinforcement-learning control and pattern recognition systems. Adaptive, Learning and Pattern Recognition Systems, capítulo 8, pages: 287–318. Academic Press, New York e London, 1970.
- [MIT97] MITCHELL, T. Machine Learning. Mcgraw-Hill, 1997.
- [MIT02] MITRA, S.; SANKAR, P. Data Mining in Soft Computing Framework: A Survey. IEEE Transactions on Neural Networks, v. 13, pages: 3-14, 2002.
- [MON95] MONGIOVI, G. Uso de Relevância Semântica na Melhoria da Qualidade dos Resultados Gerados pelos Métodos Indutivos de Aquisição de Conhecimento a partir de Exemplos: Tese de Doutorado, Universidade Federal da Paraíba, 1995.
- [NAL00] NALIATO, F. C. Aplicação de técnicas de Mineração de Dados: Estudo de caso em Marketing Direto, Dissertação de Mestrado em Sistema e Computação, IME, RJ, 2000.
- [NAR04] NARAKESARI, N.; ZAK, J. Application of Rough Sets Methodology for Data Analysis in Business Marketing. Harvard Business School. Boston, 2004.
- [NUN02] NUNES, R. S. Mineração de Dados aplicada ao Planejamento da operação de Sistemas Hidrotérmicos. Dissertação de Mestrado. Universidade Federal de Goiás, 2002.
- [ØHM99] ØHM, A. Discernibility and Rough Sets in Medicine: Tools and Applications, PhD thesis, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. NTNU report 1999.

- [ØHM00a] ØHM, A. ROSETTA Technical Reference Manual, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. 66 pages.
- [ØHM00b] ØHM, A. The ROSETTA C++ Library: Overview of Files and Classes, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. 45 pages.
- [PAK04] PAKHIRA, M. K., BANDYOPADHYAY S., MAULIK U. Validity index for crisp and fuzzy clusters. Pattern Recognition, June 2004.
- [PAR03] PARENTE, V.; HERNANDEZ, E. del M. ; ZANINI, A ; CASTRO, R. S. Utilização de Redes Neurais Como Ferramenta Qualitativa na Regulação Tarifária do Setor Elétrico. Congresso ABAR 2003 – Associação Brasileira das Agências de Regulação, 2003, Gramado, Brasil.
- [PAS03] PASSARI, A . F. L. Exploração de Dados Atomizados para Previsão de Vendas no Varejo utilizando Redes Neurais. Dissertação de Mestrado. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2003.
- [PAT05] PATRÍCIO, C. M. M. M., PINTO, J. O. P., SOUZA, C. C. Rough Sets: Técnica de Redução de Atributos e Geração de Regras para Classificação de Dados. XXVIII – CNMAC, Congresso Nacional de Matemática Aplicada e Computacional, 2005.
- [PAW82] PAWLAK, Z. Rough Sets, International Journal of Computer and information Sciences, pages: 341-356, 1982.
- [PAW91] PAWLAK, Z. Rough Sets: theoretical aspects of reasoning about data. London, Kluwer, 1991.
- [PAW96a] PAWLAK, L. Why Rough Sets? Fuzzy Systems. Proceedings of the Fifth IEEE International Conference on, volume 2, 8-11, sept.1996, pages: 738-743.

- [PAW96b] PAWLAK, Z. Rough Sets, Rough relations and Rough functions. *Fundamenta Informaticae*, 27, pages: 103-108, 1996.
- [PAW96c] PAWLAK, Z. Rough Sets and data analysis. *Fuzzy Systems Symposium*, 1996. 'Soft Computing in Intelligent Systems and Information Processing', Proceedings of the 1996 Asian 11-14 Dec. 1996 pages: 1-6.
- [PAW98] PAWLAK, Z. Granularity of knowledge, indiscernibility and Rough Sets *Fuzzy Systems Proceedings*, 1998. *IEEE World Congress on Computational Intelligence.*, The 1998 IEEE International Conference on, volume 1, 4-9 May 1998, pages: 106-110.
- [PIA91] PIATETSKY-SHAPIO, G. Knowledge Discovery in Real Databases: A report on the IJCAI-89 Workshop. *AI Magazine*, volume 11, nº 5, jan.1991, Special issue: 68-70.
- [PIA93] PIATETSKY-SHAPIO, G.; MATHEUS, C. J.; CHAN, P. K. Systems for Knowledge Discovery in Data bases. *IEEE*, 1993, pages: 903-912.
- [PIA96] PIATETSKY-SHAPIO, G. Mining Business Databases. *Communications of the ACM*, Nov./1996, volume 39, nº 11.
- [PIL01] PILA, A .D. Seleção de Atributos Relevantes para Aprendizado de Máquina utilizando a Abordagem de Rough Sets, maio 2001. Dissertação de Mestrado, ICMC-USP.
- [PIN99a] PINHEIRO, C. A. R. Data Mining: Obtendo Vantagens com seu Data Warehouse. *Developers. Magazine*, São Paulo, nº 35, pages: 38-40, Jul. 1999.
- [PIN99b] PINHEIRO, C. A. R. Redes Neurais como Técnica de Predição em Data Mining *Developers. Magazine*, São Paulo, nº 39, pages: 46-49, Nov.1999.
- [PÖL04] PÖLZLBAUER, G. Survey and Comparison of Quality Measures for Self-Organizing Maps. Department of Software Technology. Vienna University of Technology.

- [POR95] PORTUGAL, M. S. Redes Neurais Artificiais e Previsão de Séries Temporais: Uma Introdução. Revista Brasileira de Economia, volume 49, nº 4, pages: .611-629.
- [PYL99] PYLE, D. Data Preparation for Data Mining., San Francisco: Morgan Kaufmann Publishers, 1999.
- [PUT00] PUTTEN, van der P.; SOMEREN, M. van. COIL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-09. June 22, 2000.
- [RES00] REZENDE, A. D.; ABREU, F. A. Tecnologia da Informação: Aplicada a Sistemas de Informação Empresariais. Ed. Atlas. São Paulo, 2000.
- [RIC94] RICH, E.; KNIGHT, K. Inteligência Artificial. 2ª. ed. São Paulo: Makron Books, 1994.
- [ROB00] ROBBINS, S. Administração: Mudanças e Perspectivas. Ed. Saraiva. São Paulo, 2000.
- [ROC90] ROCHESTER, J. B. – New Business Use for Neurocomputing – I/S Analyser, volume 28, nº 2, pages: 1-14, fev., 1990.
- [ROD01] RODRIGUES, J. A . F. Data Mining: conceitos, técnicas e aplicação. Dissertação de Mestrado. Escola Politécnica da Universidade de São Paulo (EPUSP), 2001.
- [ROS97] ROSETTA. A Rough Sets Toolkit for Analysis of Data Disponível em <http://www.idi.ntnu.no/~aleks/rosetta>.
- [RUS97] RUSHMEIER, H.; LAWRENCE, R.; ALMASI, G. Case study: visualizing customer segmentations produced by self organizing maps Visualization '97, Proceedings, 19-24, Oct. 1997, pages: 463-466, 578.
- [SAN03] SANKAR, K. P., SUSHMITA, M. Rough-Fuzzy MLP: Modular Evolution, Rule Generation and Evaluation. IEEE Transactions on

- Knowledge and Data Engineering, volume 15, nº 1, Janeiro/Fevereiro 2003.
- [SAS05] STATISTICAL ANALISYS SYSTEM (SAS). Disponível em <http://www.sas.com>
- [SAS06] SASSI, R. J.; SILVA, L. A.; HERNANDEZ, E. del M. Neural Networks and Rough Sets: A comparative study on data classification. The 2006 International Conference on Artificial Intelligence (ICAI'06). Junho, 26 a 29, Las Vegas, USA.
- [SKO98] SKOWRON, A .; POLKOWSKI, L. Rough Mereological Foundations for Design, Analysis, Synthesis and Control in Distributive Systems. In:Proceedings The Second Joint Annual Conference on Information Sciences. 104(1998), pages: 129-156.
- [SLA93] SZLADOW, A .; ZIARKO, W. Rough Sets: working with imperfect data. AI Expert, July, pgs 36-41, 1993.
- [SLO92] SLOWINSKI, R. Rough Sets with Strict and Weak Indiscernibility Relations. IEEE International Conference on Fuzzy Systems, 1992. 8-12 March 1992.
- [SMI90] SMITH, J. C. A Neural Network-Could it Work for You? Financial Executive, nº 6, 1990, pages: 26-30, mai./jun. 1990.
- [SOM99] SOMTOOLBOX. Disponível em <http://www.cis.hut.fi/projects/somtoolbox/>.
- [SOU99] SOUZA, F.J. Modelos Neuro-Fuzzy Hierárquicos. Tese de Doutorado. Departamento de Engenharia Elétrica (PUCRIO), 1999.
- [STE00] STEWART, T. A. Capital Intelectual: a nova vantagem competitiva das empresas. Rio de Janeiro: Campus, 2000.
- [SVE98] SVENSEN, J.F.M. GTM: The Generative Topographic Mapping. PhD Thesis, Aston University, April 1998.
- [TAN98] TANAKA, H.; LEE, H.; GUO, P. Possibility data analysis with Rough Sets concept Proceedings, 1998. IEEE World Congress on Fuzzy

- Systems. Computational Intelligence., The 1998 IEEE International Conference on, volume 1, 4-9 May 1998, pages: 117–122.
- [TAP97] TAPSCOTT, D. Economia Digital: Promessa e perigo na era da inteligência em rede, São Paulo: Makron Books, 1997.
- [TAP00] TÁPIA, M. Aplicação de Redes Neurais na Previsão de Preços de Ovos. Congresso e Mostra de Agroinformática. 18-20 Out./2000. Ponta Grossa, Paraná.
- [TAY02] TAY, F. E. H., LIXIANG, S. Economic and Financial prediction using Rough Sets Theory Model. European Journal of Operational Research, 2002.
- [TER00] TERRA, J. C. C. Gestão do conhecimento: O grande desafio empresarial. Ed. Negócio. São Paulo, 2000.
- [TIC00] TIC 2000. Disponível em <http://www.wi.leidenuniv.nl/~putten/library/cc2000/>.
- [TOD95] TODESCO, J. L. Reconhecimento de Padrões usando Rede Neuronal Artificial com uma Função de Base Radial: Uma Aplicação na Classificação de Cromossomos Humanos. Tese de Doutorado. UFSC. 1995.
- [TOR95] TORRES, N. A. Competitividade empresarial com a Tecnologia da Informação. São Paulo: Makron Books, 1995.
- [TRI96] TRIPPI, R.; TURBAN, E. Neural Networks in Finance and Investments; Using Intelligence Artificial to Improved Real World. Chicago, 1996.
- [TSU96] TSUMOTO, S.; TANAKA, H. Extraction of Domain Knowledge from Databases Based on Rough Set Theory. Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, v. 2, pages 8-11, September. 1996.
- [TSU98] TSUMOTO, S. Knowledge Discovery in Medical Databases based on Rough Sets and attribute-oriented generalization IEEE World Congress on Computational Intelligence., The 1998 IEEE

- International Conference on Fuzzy Systems Proceedings, 1998, volume 2, 4-9 May 1998, pages: 1296-1301.
- [TSU00] TSUMOTO, S. Problems with Mining Medical Data Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International Conference. 25-27 Oct. 2000, pages: 467–468.
- [TSU01a] TSUMOTO, S. Medical diagnostic rules as upper approximation of Rough Sets The 10th IEEE International Conference on Fuzzy Systems, 2001 , volume 3, 2-5 Dec. 2001, pages: 1551–1554.
- [TSU01b] TSUMOTO, S. Rule induction system based on characterization of medical diagnosis using Rough Sets IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th, volume 4, 25-28 July 2001.
- [TUR03] TURBAN, E.; RAINER, R. K.; POTTER, R. E. Administração de Tecnologia da Informação. Rio de Janeiro: Campus, 2003.
- [UCH97a] UCHÔA, J. Q.; NICOLETTI, M. C. Elementos da Teoria de Conjuntos Aproximados. Relatório Técnico. Universidade Federal de São Carlos. Março 1997.
- [UCH97b] UCHÔA, J. Q.; NICOLETTI, M. C. O uso de funções de pertinência na caracterização dos principais conceitos da Teoria de Conjuntos Aproximados. Relatório Técnico. Universidade Federal de São Carlos. Março 1997.
- [UCH97c] UCHÔA, J. Q.; NICOLETTI, M. C. Relações e funções aproximadas: uma abordagem baseada na Teoria de Conjuntos Aproximados. Relatório Técnico. Universidade Federal de São Carlos. Março 1997.
- [UCH98] UCHÔA, J. Q. Representação e Indução de Conhecimento usando Teoria de Conjuntos Aproximados. Dissertação de Mestrado. Universidade Federal de São Carlos, 1998.

- [UCI98] UCI Repository of machine learning databases. Irvine, CA: University of California, Department of information and Computer Science. Disponível em <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [ULT89] ULTSCH, A.; SIEMON, H.P. Kohonen's Self Organizing Feature Maps for exploratory Data Analysis. Proceedings of the International Neural Networks Conference. Dordrecht. Netherlands, p. 305-308; 1989.
- [ULT95] ULTSCH, A. Self-Organizing Neural Networks perform different from statistical K-means Clustering. In Proc. GfKI, Basel, Swiss, 1995.
- [VAN97] VAN GILS, M.; JANSEN, H.; NIEMINEN, K.; SUMMERS, R.; WELLER, P.R. Using artificial neural networks for classifying ICU patient states Engineering in Medicine and Biology Magazine, IEEE, volume 16, Issue: 6, Nov.-Dec. 1997, pages: 41–47.
- [VEL01a] VELLASCO, M.; AZEVEDO, H.; PASSOS, E. Mineração de Dados Aplicada a CRM em uma Base de Clientes de Telefonia de Longa Distância. Proceedings of the V Brazilian Conference on Neural Networks, pages: 397-402, April 02-05, 2001.
- [VEL01b] VELLASCO, M; PACHECO, M. A. C.; CID, D. J. A . Classificação de Dados através de Rough Sets. Proceedings of the V Brazilian Conference on Neural Networks, pages: 319-324, April 02-05, 2001.
- [VES97] VESANTO, J. Data Mining Techniques Based on the Self-Organizing Map MSc Thesis. Helsinki University of Technology, Espoo, Finland, 1997.
- [VES99a] VESANTO, J. SOM-Based Data Visualization Methods. Intelligent Data Analysis, volume 3, nº 2, pages: 111-126, 1999.
- [VES99b] VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. Self-Organizing Map in Matlab: the SOM Toolbox. In: Proceedings of the Matlab DSP Conference, Espoo, Finland, pages 35-40, 1999.

- [VES00a] VESANTO, J.; ALHONIEMI, E. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, volume 11, nº 2, pages: 586-600, 2000.
- [VES00b] VESANTO, J. Using SOM in Data Mining. *Licenciate Thesis*. Helsinki University of Tecnology, Espoo, Finland, 2000.
- [VES00c] VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. SOM Toolbox for Matlab 5. Technical report A 57. Helsinki University of Technology, Finland, 2000.
- [VES02] VESANTO, J. Data Exploration Process Based on the Self-Organizing Map. *PhD thesis*, Helsinki University of Technology, 2002.
- [VES03] VESANTO, J.; SULKAVA, M.; HOLLMÉN, J. On the Decomposition of the Self-Organizing Map Distortion Measure. *Proceedings of the Workshop on Self-Organizing Maps (WSOM'03)*, pp 11-16, Hibikino, Kitakyushu, Japan, September 2003
- [VIL94] VILLMANN, T., DER, R., MARTINEZ, T. A new quantitative measure of topology preservation in Kohonen's feature maps, *Proceedings of the IEEE International Conference on Neural Networks 94*, Orlando, Florida, USA, 645-648, June 1994
- [WEI03] WEI, M. J. Rough Set based Approach to Selection of Node. *International Journal of Computational Cognition*, volume 1, nº 2, pages: 25-40, June 2003.
- [XIA96] XIAOHUA, H.; CERCONE, N. Mining Knowledge Rules from Databases: A Rough Set Approach. *Proceedings of the Twelfth International Conference on Data Engineering*. 1996, pages: 96–105.
- [XIA97] XIAOHUA H.; CERCONE, N. Learning maximal generalized decision rules via discretization, generalization and Rough set feature selection *Proceedings., Nineteen IEEE International Conference on Tools with Artificial Intelligence*, 1997, 3-8 Nov. 1997, pages: 548–556.

- [XIA02a] XIAO-YE, W.; ZHENG, O. W. Stock Market Time series Data Mining based on regularized Neural Network and Rough set Proceedings. 2002 International Conference on Machine Learning and Cybernetics, 2002, volume 1, 4-5 Nov. 2002, pages: 315-318.
- [XIA02b] XIAOHUI, L.; GONGXIAN, C.; WU, J. X. Analyzing Outliers Cautiously. IEEE Transactions on Knowledge and Data Engineering, v. 14 , Issue 2, 2002, pages: 432–437.
- [YAH97] YAHIA, M. E.; MAHMOD, R. Rough Set decision making for uncertain data. Department of Computer Science, University Putra Malaysia, Technical Report number TR018, 1997.
- [YAO98] YAO, Y.Y. A Comparative Study of Fuzzy Sets and Rough Sets. Journal of Information Sciences, pgs 227 – 242, 1998.
- [YU03] YU, C.; WU, M., W, M. Combining Rough Set Theory with Neural Network Theory for Pattern Recognition. Proceedings of the 2003 IEEE. International Conference on Robotics, Intelligent Systems and Signal Processing. Changsha, China – October 2003, pages: 880-885.
- [YU05] YU, L; SHOUYANG, W.; LAI, K. K. A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting. International Journal of Knowledge and Systems Sciences, volume 2, nº 1, March 2005.
- [ZAR99] ZAREMBA, M. B.; WIELGOSZ, A. Predicting outcome for hospitalized cardiac patients using a combined neural network and Rough set approach IJCNN '99. International Joint Conference on Neural Networks, 1999, volume 5, 10-16 July 1999, pages: 3665-3668.
- [ZHA94] ZHANG, X. Time series analysis and prediction by neural ntworks. Optimization Methods and Software, v.4, pgs 151 – 170.
- [ZHA01] ZHAOCONG, W. Research on remote sensing image classification using neural network based on Rough Sets. Info-tech and Info-net,

2001. Proceedings. ICII 2001 - Beijing. 2001 International Conferences on volume 1, 29 Oct.-1 Nov. 2001, pages: 279–284.
- [ZHA04] ZHANG, M.; YAO, J. T. A Rough Sets based approach to feature selection Processing. NAFIPS '04. IEEE Annual Meeting of the Fuzzy Information, 2004, volume 1, 27-30 June 2004, pages: 434-439.
- [ZHI02] ZHILIANG, W.; WENBO, M.; XUEJING, G.; JIANPING, Y.; NAGAI, M.; GUOWEI, C. The Research of The Police GIS Spatial Data Classification Technology Based on Rough Set. Proceedings of the 4th World Congress on Intelligent Control and Automation, v. 1, June, 2002.
- [ZIA93] ZIARKO, W.; KATZBERG, J. D. Rough Sets approach to system modelling and control algorithm acquisition WESCANEX 93. 'Communications, Computers and Power in the Modern Environment.' Conference Proceedings., IEEE , 17-18 May 1993, pages: 154–164.
- [ZIA94a] ZIARKO, W.; NING, S. Machine learning: Rough Sets perspective. Proceedings of International Conference on Expert Systems for Development, 1994, 28-31 March 1994, pages: 114–118.
- [ZIA94b] ZIARKO, W. Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer-Verlag, 1994.
- [ZIA95] ZIARKO, W.; NING, S. Discovering Attribute Relationships, Dependencies And Rules By Using Rough Sets. International Conference on Proceedings of the Twenty-Eighth Hawaii, v. 3, 1995, pages 293-299.
- [ZON02] ZONGHAI, S.; YOUXIAN, S. Application of Rough Sets to fault diagnoses Proceedings of the 4th World Congress on Intelligent Control and Automation, 2002, volume 4, 10-14 June 2002, pages: 2742-2745.
- [ZUB88] ZUBOFF, S. In the Age of the Smart Machine: The Future of Work and Power. Basic Books. USA, 1988.

- [ZUC03] ZUCHINI, M. H. Aplicações de Mapas Auto-Organizáveis em Mineração de Dados e Recuperação de Informação. Dissertação de Mestrado. Universidade de Campinas Unicamp. Setembro, 2003.

Bibliografia Consultada

- AIJING, H.; YAYAO, Z.; MAZLACK, L. J. Data Discovery using Rough Set based reductive partitioning: some experiments. IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th, volume 1, 25-28, July 2001, pages: 203–208.
- BACK, B.; SERE, K.; VANHARANTA, H. Analyzing Financial performance with Organizing Maps. Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on, volume 1, 4-9 May 1998 pages: 266-270.
- BARRETO, J. M. Introdução às Redes Neurais Artificiais. Anais V Escola Regional de Informática da SBC Regional Sul, 5 a 10 de maio de 1997, páginas: 41-71.
- BEAUBOUF, T.; PETRY, F. E. Vague regions and spatial relationships: a Rough Set Approach .Computational Intelligence and Multimedia Applications, 2001. ICCIMA 2001. Proceedings. Fourth International Conference on, 30 Oct.-1 Nov. 2001, pages: 313-317.
- BEAUBOUF, T.; PETRY, F. E. A Rough Set foundation for Spatial Data Mining involving Vague Regions. Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on, volume 1, 12-17 May 2002.
- BENZHU, X.; XUEGANG, H.; WANG, H. A general Rough set model based on tolerance. Computer Supported Cooperative Work in Design, 2004. Proceedings. The 8th International Conference on, volume 1, 26-28 May 2004, pages: 770–774.
- BERRY, M. J. A.; LINOFF, G. Mastering Data Mining: The Art and Science of Customer Relationship Management. Wiley Computer Pub, NY. 2000.
- BEYNON, M. J.; PEEL, M. J. Variable Precision Rough Set Theory and data Discretization: An Application to Corporate Failure Prediction. Omega, v. 29, pages: 561-576, 2001.

- BISPO, C .A.; CAZARINI, E. W. Transformando Dados em Informações Via Data Mining. *Developers Magazine*, nº 29, páginas: 36-38, Jan.1999.
- CARVALHO, D. R. Data Mining através de Indução de Regras e Algoritmos Genéticos. Dissertação de Mestrado em Informática Aplicada. PUCPR, 1999.
- CHAU-YUN, H.; HWAI-EN, W. An improved algorithm for Kohonen's Self-Organizing Feature Maps. *Circuits and Systems*, 1992. ISCAS 92. Proceedings, 1992 IEEE International Symposium on, volume: 1, 3-6 May 1992, pages: 328–331.
- CHEN, P.; TOYOTA, T. Multi-valued Neural Network and the Knowledge Acquisition Method by the Rough Sets for ambiguous recognition problem. *Systems, Man, and Cybernetics*. IEEE International Conference, v. 1, 1996, pages: 736–740.
- CHUN-GUANG, C.; DING-WEI, W.; KUN-YUAN, H.; ZHI, T. Study on Rough Set based Reduction Technique for Case Attributes. *Machine Learning and Cybernetics*, 2004. Proceedings of 2004 International Conference on volume 4, 26-29 Aug. 2004, pages: 2122–2127.
- COPPOCK, S.; MAZLACK, L. Rough Sets used in the Measurement of Similarity of Mixed Mode Data. *North American Fuzzy Information Processing Society, NAFIPS 2003*, 22nd International Conference of the 24-26 July 2003, pages: 197–201.
- COSTA, J. A. F. Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis. Tese de Doutorado. Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Elétrica e de Computação, 1999.
- FAYYAD, U.; SHAPIRO, G. P. Data Mining and Knowledge Discovery In Databases: an Overview. *Communications ACM*, Special Issue on Data Mining, v. 39, nº 11, november 1996.
- FELDENS, M. A. Engenharia da Descoberta de Conhecimento em Bases de Dados: Estudo e Aplicação na Área de Saúde. Tese de Mestrado, CPGCC, UFRGS, 90p., abril 1997.

- FILHO, E. C. B. C. Modelagem, Aplicações e Implementações de Redes Neurais. Anais da IV Escola Regional de Informática da SBC Regional Sul, 21 a 27 de abril de 1996, páginas: 36-53.
- GEDIGA, G.; DUNTSCHE, I. Rough Approximation Quality Revisited. Artificial Intelligence, nº 132, 2001, pages: 219-234.
- GENNARI, M.C. Minidicionário de Informática. Ed. Saraiva. São Paulo, 1999.
- HECHT-NIELSEN, R. Self-Organizing Maps [Book Reviews]. Neural Networks, IEEE Transactions, v. 7, Issue 6, Nov. 1996.
- HEINRICHS, J. H.; LIM, J. S. Integrating Web-based Data Mining Tools with Business Models for Knowledge Management. Decision Support Systems, v. 35, pages: 103-112, 2003.
- HIRANO, S.; XIAO GUANG, S.; TSUMOTO, S. Analysis of Time-Series Medical Databases using Multiscale Structure Matching and Rough Sets-based Clustering Technique.. The 10th IEEE International Conference on Fuzzy Systems, 2001, v. 3, 2-5 Dec. 2001, pages: 1547–1550.
- HIRANO, S.; TSUMOTO, S. A. Knowledge-Oriented Clustering Technique based on Rough Sets. Computer Software and Applications Conference, 2001. COMPSAC 2001, 25th Annual International, 8-12 Oct. 2001, pages: 632–637.
- JAGIELSKA, I. A Neural Network Model for Sales Forecasting Artificial Neural Networks and Expert Systems, 1993. Proceedings First New Zealand International Two-Stream Conference on 24-26 Nov. 1993, pages :284–287.
- KIVILUOTO, K.; OJA, E. S-Map: A Network with a Simple Self-Organization Algorithm for Generative Topographic Mapping. Advances In Neural Processing Systems 10, MIT Press, pages: 549-555, 1997.
- KOCZKODAJ W. W., ORLOWSKI, M.; MAREK, V. W. Myths about Rough Set Theory. Communications of the ACM, pages: 102-103, 1998.
- KOHONEN, T. Clustering Taxonomy and Topological Maps of Patterns. Proceedings Of the Sixth International Conference on Pattern Recognition, pages.114-128, 1982.

- KOHONEN, T; RITTER, H. The Self-Organizing Semantic Map. *Biological Cybernetics*, 61, 241 – 254, 1989.
- KOHONEN, T. The Self-Organizing Map. *Proceedings of the IEEE*, v. 78, Issue 9, Sept. 1990, pages 1464–1480.
- KOHONEN, T. New Developments And Applications Of Self-Organizing Maps. *Neural Networks for Identification. Proceedings International Workshop on Robotics, and Signal/Image Processing*. 21-23 Aug. 1996.
- KROPAS, H. C. V. Data Reduction via Auto-Associative Neural Networks . *IPMM 1999. Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials*, 1999, v. 2, 10-15 July 1999.
- KUSIAK, A. Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing. *IEEE Transactions on Eletronics Packaging Manufacturing*, v. 24, 2001.
- LAROUSSE Cultural. *Dicionário da Língua Portuguesa*. Ed. Universo. São Paulo, 1992.
- LIANG, J.; ZONGBEN, X. Uncertainty Measures of Roughness of Knowledge and Rough Sets in Incomplete Information Systems. *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, 2000, v. 4, 2000, pages: 2526–2529.
- LINGRAS, P. Combination of Evidence in Rough Set Theory. *Proceedings ICCI '93., Fifth International Conference on Computing and Information*, 1993, 27-29 May 1993, pages: 289–293.
- LINGRAS, P. Comparasion of Neofuzzy and Rough Neural Networks. *Information Sciences*, pages: 207-215, 1998.
- LIPPMAN, R. P. An Introduction to Computing with Neural Nets. *IEEE Acoustics, Speach and Signal Processing*, v. 6, pages: 4-22, 1987.
- LOESCH, C.; SARI, S. *Redes Neurais Artificiais : fundamentos e modelos*. Blumenau: IPS/FURB, 1995.
- LOJKINE, J. A. *Revolução Informacional*. São Paulo. Ed. Cortez, 1995.

- MAGNIER, J. P. Utilisation des reseaux de neurones pour le developpement de systemes d'aide à la décision. Montpellier: Intitut d'Administratio d'Entreprises, 1991.
- MAOFU, L.; YANXIANG, H.; HUIJUN, H.; DANDAN, Y. Dimension Reduction Based on Rough Set. In Image Mining CIT '04. The Fourth International Conference on Computer and Information Technology, 2004. 14-16 Sept. 2004, pages: 39–44.
- MARQUES, V. M. Analisando os Dados do Programa de Melhoramento Genético da Raça Nelore com Data Warehousing e Data Mining. Dissertação de Mestrado. USP-São Carlos, setembro, 2002.
- NIMER, F. Obtendo Vantagem Competitiva com o Uso de Data Mining. Developers Magazine, fevereiro de 1998.
- NURNBERGER, A.; DETYNIECKI, M. Visualizing Changes in Data Collections Using Growing Self-Organizing Maps. IJCNN '02. Proceedings of the 2002 International Joint Conference on Neural Networks, 2002., v. 2 , 12-17 May 2002, pages: 1912–1917.
- OLIVEIRA, P. J.; RODRIGUES, F.; HENRIQUES, P. R. Limpeza de Dados: Uma visão Geral. Relatório Técnico. Universidade do Minho, Portugal.
- PAL, S. K.; MITRA, S.; MITRA, P. Rough-Fuzzy MLP: Modular Evolution, Rule Generation and Evaluation Knowledge and Data Engineering. IEEE Transactions on, v. 15 , Issue 1, Jan.-Feb. 2003, pages: 14–25.
- PAWLAK, Z.; GRZYMALA-BUSSE, J.; SLOWINSKI, R.; ZIARKO, W. Rough Sets. Communications of the ACM, pages: 89-95, 1995.
- PAWLAK, Z. Rough Sets and Data Analysis. Fuzzy Systems Symposium, 1996. Soft Computing in Intelligent Systems and Information Processing. Proceedings of the 1996. Asian, 11-14 December, 1996 pages 1–6.
- PAWLAK, Z. Theorize with Data using Rough Sets. Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International, 26-29 Aug. 2002, pages: 1125–1128.
- PEDRY, C. Z. W.; SUCCI, G.; REFORMAT, M.; MUSILEK, P.; BAI, X. Self Organizing Maps as a Tool For Software Analysis. Canadian

- Conference on Electrical and Computer Engineering, 2001, v. 1, pages 13-16, May 2001.
- PETERS, J. F.; PAWLAK, Z.; SKOWRON, A. A Rough Set Approach to Measuring Information Granules. Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International 26-29 Aug. 2002, pages 1135–1139.
- PILA, A. D. Rough Sets Aplicado à Data Mining: Algumas Limitações. II LAPTEC, SENAC. 2000.
- SHIM, J. P.; WARKENTIN, M.; COURTNEY, J.; POWER, D. J.; SHARDA, R.; CARLSSON, C. Past, Present, And Future Of Decision Support Technology. Decision Support System, v. 33, pages: 111-126, 2002.
- SHIXING, F.; ZENGXIANG, L.; HAIMING, L.; YANDA, L. Rough Set Theory and its Practice in Knowledge Discovery. Proceedings of the 3rd World Congress on Intelligent Control and Automation, 2000, v. 4, 28 June - 2 July 2000, pages: 2535–2539.
- STAIR, R. M. Princípios de Sistemas de Informação: uma abordagem gerencial. Rio de Janeiro. Ed. LTC, 1996.
- SUGIHARA, K.; ISHII, H.; TANAKA, H. Conjoint Analysis Based on Rough Aproximations by Dominance Relations Using Interval Regression Analysis. Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, 2002, v. 1, 12-17 May 2002, pages 763–766.
- SZLADOW, A.; ZIARKO, W. Rough Sets: Working with Imperfect Data. AI Expert, July, pages: 36-41, 1993.
- TANAKA, H.; KOYAMA, K.; MAEDA, Y. A Method for Reducing Information Systems with Binary Data by Rough Sets. Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, 1996, v. 2, pages 8-11 September, 1996.
- TANAKA, H.; FURUYA, S.; MAEDA, Y. On Reducing Information Systems by Accuracy Measures in Rough Sets. Soft Computing in Intelligent Systems and Information Processing. Proceedings of Fuzzy Systems Symposium, 1996. Asian , 11-14 December, 1996, pages 139–144.

- TSANG, E. C. C.; DEGANG, C.; LEE, J. W. T.; YEUNG, D. S. On the Upper Approximations of Covering Generalized Rough Sets. *Proceedings of International Conference on Machine Learning and Cybernetics*, 2004, v. 7, 26-29 August, 2004, pages 4200-4203.
- TSAU Y. L. Neighborhood Systems: Mathematical Models of Information Granulations. *IEEE International Conference on Systems, Man and Cybernetics*, 2003, v. 4, October, 2003, pages: 3188-3193.
- VAN HULLE, M. M. Faithful Representation and Topographic Maps; From Distortion to Information-Based Self-Organization. John Wiley & Sons, 2000.
- VERMASVUORI, M.; ENDEN, P.; HAAVISTO, S.; JAMSA-JOUNELA, S. L. The use of Kohonen Self-Organizing Maps in Process Monitoring. *Proceedings. Intelligent Systems, 2002 First International IEEE Symposium*, v. 3, September. 2002, pages: 2-7.
- VIANA, R. Mineração de Dados: Introdução e Aplicações. *SQL Magazine*. Edição 10, ano 1, páginas 16-25, Abril 2004.
- WANG, G.; ZHAO, J.; NA, J.; WU, Y. Theoretical Study on Attribute Reduction of Rough Set Theory: Comparison of Algebra and Information Views Cognitive Informatics. *Proceedings of the Third IEEE International Conference on 16-17 August, 2004*, pages 148–155.
- WESOLKOWSKI, S. Clustering with a Mixture of Self-Organizing Maps. *IJCNN '02. Proceedings of the 2002 International Joint Conference on Neural Networks*, 2002, v. 3, 12-17, Mayo, 2002, pages: 2363–2368.
- XIANMING, H.; JIKAI, Y. Vague Rough set and its Application To Knowledge Acquisition. *Proceedings. 2003 International Conference on Natural Language Processing and Knowledge Engineering*, 2003, 26-29 October, 2003.
- XIAOHUA H.; CERCONE, N. Learning Maximal Generalized Decision Rules via Discretization, Generalization and Rough Set Feature Selection. *Proceedings of Nineteen IEEE International Conference on Tools with Artificial Intelligence*, 1997, 3-8 November, 1997, pages 548–556.

- XIAOHUA, H. Using Rough Sets Theory and Database Operations To Construct A Good Ensemble Of Classifiers For Data Mining Applications. Proceedings IEEE International Conference on Data Mining. 2001, pages 233–240.
- YAO, Y.; Rough Sets, Neighborhood Systems And Granular Computing. IEEE Canadian Conference on Electrical and Computer Engineering, v. 3, 1999, pages 1553-1558.
- YAO, Y. Y.; YAOHUA, C. Rough Set Approximations In Formal Concept Analysis. IEEE Annual Meeting of the Fuzzy Information, v. 1, 2004, pages 73-78.
- YAWEN, W.; CHANG-N, Z. A Rough Neural Network For Material. International Conference on Communications, Circuits and Systems and West, v. 2, 2002, pages 1189-1193.
- ZADEH, L. A. Fuzzy Sets, Information and Control, v. 8, nº 3, 1965, pages 338-353.
- ZHANG, J.; SHUQING, W. Neuro-Fuzzy Network Based on Rough Sets and its Applications. IEEE International Conference on Systems, Man and Cybernetics, v. 3, 2003, pages 2803-2808.
- ZHUANG, Z.; YUN-QUAN, H. Applying AI technology and Rough Set Theory to Mine Association Rules for Supporting Knowledge Management. International Conference on Machine Learning and Cybernetics, v. 3, 2003, pages 1820-1825.

Anexo A – Roteiro de Execução dos Experimentos Comparativos com a Arquitetura Híbrida

Este anexo tem como objetivo ilustrar os passos necessários para a reprodução dos experimentos utilizando as ferramentas de ensaio. Para maiores detalhes sobre as bases de dados e parâmetros utilizados, veja o Capítulo 6.

Para a redução de dados com *Rough Sets* usou-se a ferramenta chamada Rosetta [ROS97]. O processo de redução pode ser dividido em duas etapas principais. A primeira é a escolha da base de dados e abertura do arquivo pela ferramenta (Figura A.1).

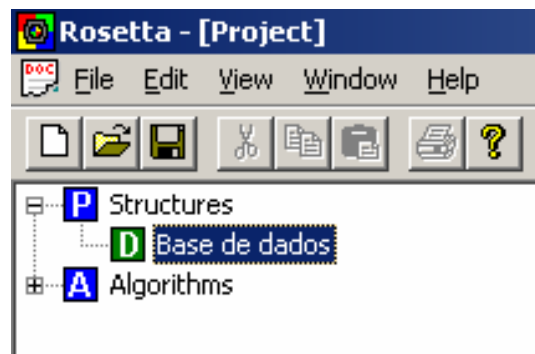


Figura A.1: Ilustração da tela do Rosetta com a base de dados carregada.

A segunda etapa é a redução dos atributos da base de dados. Para isto deve-se escolher o algoritmo de redução, que no exemplo da Figura A.2 é o algoritmo genético (*SAVGeneticReducer (Genetic Algorithm)*).

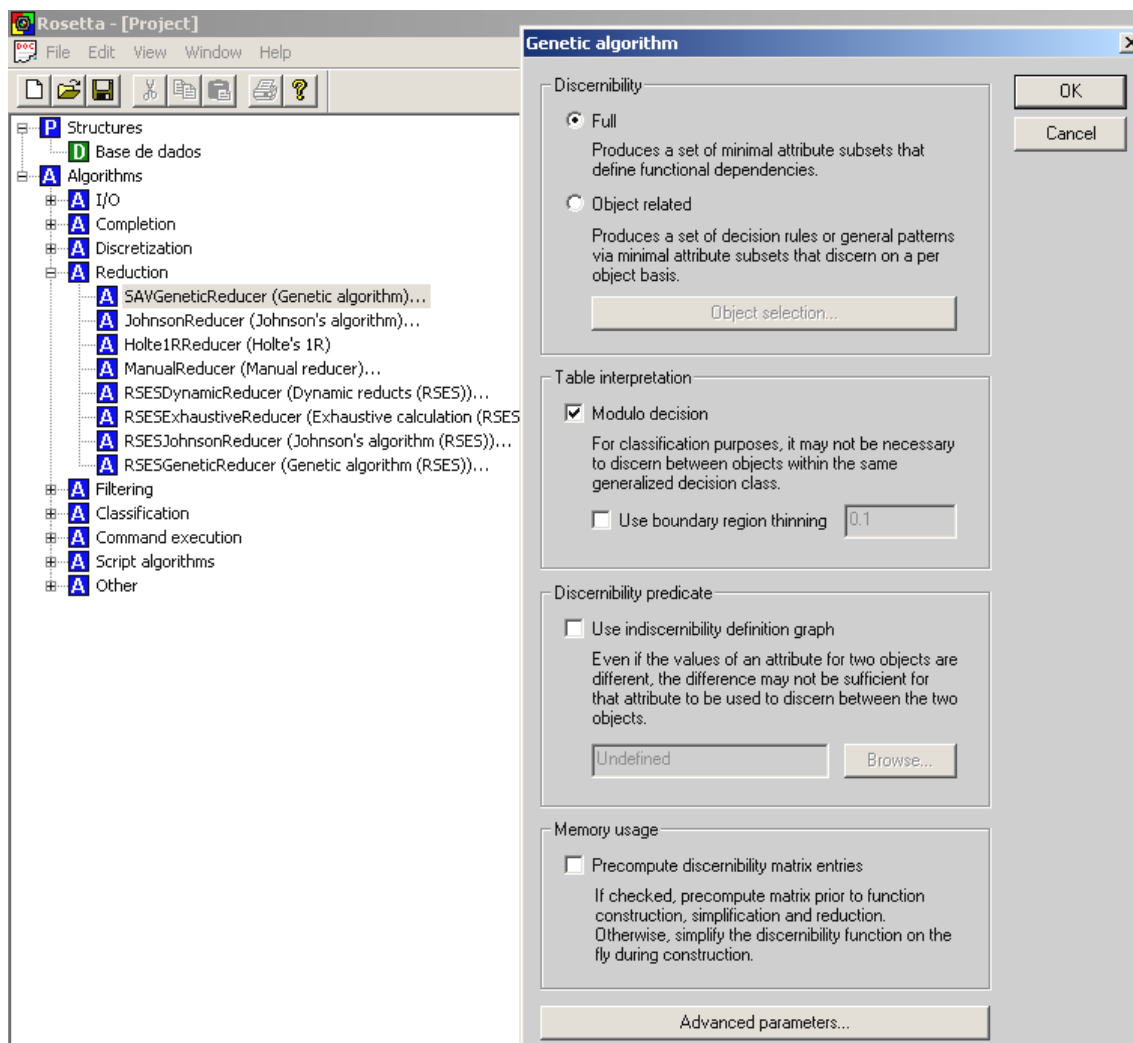


Figura A.2: Ilustração da tela do Rosetta na escolha do algoritmo e dos parâmetros de redução.

A seguir, a Figura A.3 mostra os redutos gerados pelo *Rough Sets*.

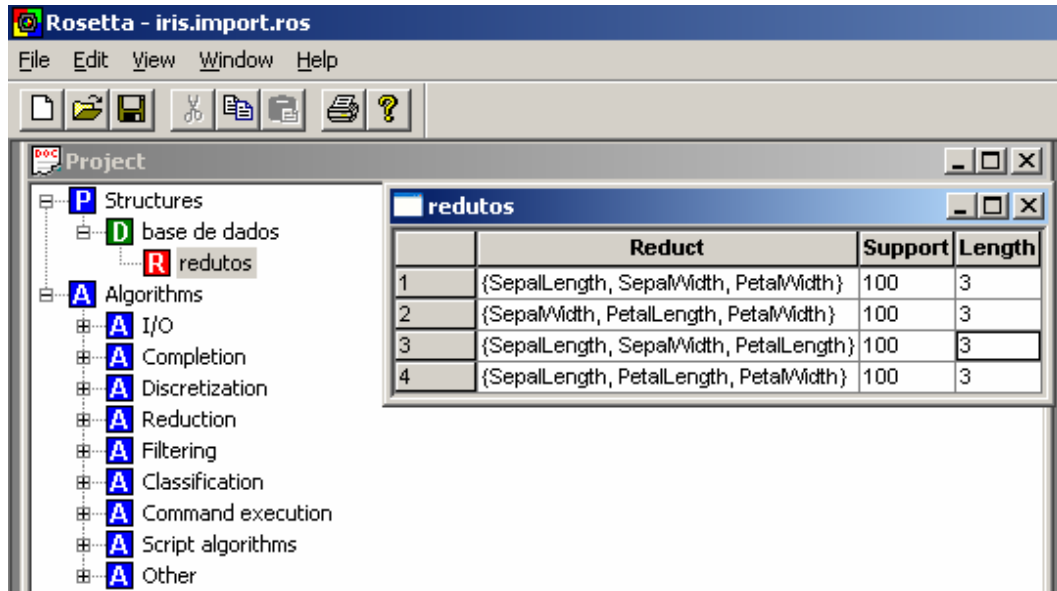


Figura A.3: Ilustração da tela do Rosetta com os redutos gerados pelo *Rough Sets* utilizando a base de dados Iris (Experimento 2).

Informações mais detalhadas sobre a ferramenta Rosetta podem ser encontradas em Pila [PIL01].

O experimento tem sequência apresentando a base de dados reduzida à rede SOM.

Para a realização dos experimentos com a rede SOM, usou-se o *SOMToolbox* descrito na Seção 6.1.1. O código fonte de chamada das funções do *SOMToolbox* está descrito a seguir:

```
% Leitura da base de dados
d=som_read_data('diretorio_da_base_de_dados');

% normaliza os dados
sD = som_normalize(d,'var');

% Inicialização e treinamento do mapa

% define o tamanho do mapa
msize=[ux uy];
Inicializa o treinamento
sMap=som_make(variavel,'forma_inicializacao','metodo_treinamento','tamanho_mapa','vizinhanca','metodo_atualizacao');
```

```
% Visualização

% Mostra o grid e os BMUs rotulados
figure,
% rotula o mapa com os labels dos dados
sMap = som_autolabel(sMap,sD);
som_show(sMap,'empty','expl');
som_show_add('label',sMap,'Textsize',12,'TextColor','k');

% Análise dos resultados

% Clusterização do mapa por similaridade de cor
[c, p, err, ind] = kmeans_clusters(sMap);
[dummy,i] = min(ind);
figure(5);

som_show(sMap,'color',{p{i}},sprintf('%d clusters',i));
colormap(jet(i)), som_recolorbar;
som_show_add('label',sMap,'Textsize',8,'TextColor','k')
```