



Universidade Federal de Alagoas
Centro de Ciências Exatas e da Natureza
Instituto de Computação

Pós-graduação em Modelagem Computacional de Conhecimento

Uma Interpretação Nebulosa dos Mapas de Kohonen

Andrilene Ferreira Maciel

Dissertação de Mestrado

Alagoas
12 de novembro de 2008

Universidade Federal de Alagoas
Centro de Ciências Exatas e da Natureza
Instituto de Computação

Andrilene Ferreira Maciel

Uma Interpretação Nebulosa dos Mapas de Kohonen

Trabalho apresentado ao Programa de Pós-graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas como requisito parcial para obtenção do grau de Mestre em Modelagem Computacional.

Orientador: *Prof. Dr. Luís Cláudius Coradine*
Co-orientador: *Prof. Dra. Roberta Vilhena Vieira*

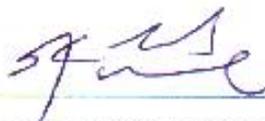
Alagoas
12 de novembro de 2008

Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária Responsável: Helena Cristina Pimentel do Vale

- M152 Maciel, Andrilene Ferreira.
 Uma interpretação nebulosa dos mapas de Kohonen / Andrilene Ferreira
 Maciel. – Maceió, 2008.
 xxiii, 130f. : il.
- Orientador: Luís Cláudius Coradine.
 Co-orientador: Roberta Vilhena Vieira
 Dissertação (mestrado em Modelagem Computacional de Conhecimento) –
 Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2008.
- Bibliografia: f. 123-130.
1. Modelagem computacional. 2. Mineração de dados – (Computação).
 3. Mapas auto-organizáveis. 4. Lógica nebulosa. 5. Diabetes mellitus. I. Título.

CDU: 004.421

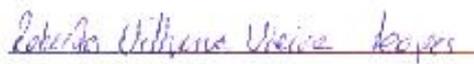
Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento pelo Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento, da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina:



Prof. Dr. Luis Cláudius Coradine

UFAL – Instituto de Computação

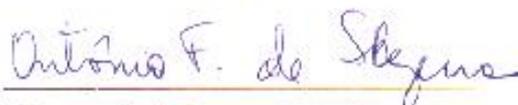
Orientador



Prof. Dra. Roberta Vilhena Vieira Lopes

UFAL – Instituto de Computação

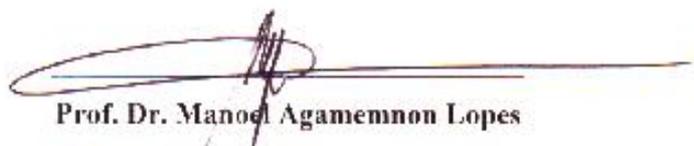
Co-orientadora



Prof. Dr. Antônio Fernando de Sousa Bezerra

UFAL – Faculdade de Medicina

Examinador



Prof. Dr. Manoel Agamemnon Lopes

UFAL – Centro de Ciências Agrárias

Examinador



Prof. Dr. Manoel Eusébio de Lima

UFPE – Centro de Informática

Examinador

Maceió, novembro de 2008.

Dedico este trabalho a minha mãe.

Agradecimentos

A conclusão deste trabalho só foi possível devido à colaboração de muitas pessoas e ao apoio de algumas instituições:

Ao professor Luís Cláudius Coradine pela orientação, confiança, apoio, incentivo, e sobretudo pela compreensão, sem os quais eu jamais teria concluído este trabalho.

A professora Roberta Vilhena Viera, da Universidade Federal de Alagoas - UFAL, pela orientação, confiança e incentivo; por ter me ensinado a ter sempre uma visão crítica sobre o meu trabalho.

Ao professor Manoel Eusébio de Lima pela orientação, confiança, incentivo, e sobretudo pela compreensão e apoio do projeto PROCAD - UFAL/UFPE.

Aos amigos da UFAL, Marcos Machado, Paulo Coutinho, Roberto Correia, Romero, Hilbert, Vitor, Jobson, Umberto e Nilda;

Aos meus amigos Rejane Ferreira, Erica Gomes da Universidade Federal de Pernambuco;

Ao Professor Tiago Alessandro Espínola da Universidade Federal Rural de Pernambuco, pelo grande incentivo na escolha da área de pesquisa;

Ao Professor Alexandre Stranford da Universidade Federal de Pernambuco, pelo apoio e incentivo;

A Professora Ana Paula Cabral Seixas da Universidade Federal de Pernambuco, pela orientação e incentivo;

Aos meus Amigos da Agência Nacional de Telecomunicações João Batista, pela compreensão e especialmente Paulo Avelar e Dilermando, pelo incentivo e influência na escolha da área acadêmica;

À Fundação de Amparo a Pesquisa de Alagoas - FAPEAL, pelo apoio financeiro;

À Companhia Energética de Alagoas, pelo apoio e fornecimento dos dados;

Ao Projeto de Cooperação Acadêmica entre o Mestrado Multidisciplinar em Modelagem Computacional de Conhecimento da UFAL e o Programa de Pós-Graduação em Ciência da Computação da UFPE, pelo apoio financeiro.

Quanto maior a dificuldade, tanto maior o mérito em superá-la.

—H.W BEECHER

Resumo

As técnicas de mineração de dados baseadas nos mapas auto organizáveis de Kohonen têm sido bastante utilizada na classificação de sinais nas mais diversas áreas de conhecimento. Geralmente, a rede SOM (Self-Organizing Maps) é usada para especificar relações de similaridade entre objetos abordando análise de agrupamentos. O custo computacional, a preparação dos dados e modelagem matemática poderá influenciar na interpretação dos resultados, entre suas limitações encontram-se aquelas provenientes da avaliação das classes. Os mapas de Kohonen não permite avaliar de forma detalhada a classe dos objetos, os quais poderão está definidos pelo limite da classe, ou seja, definir uma medida que possa relacionar quando um objeto que pertença a uma classe particular possa migrar de uma classe para outra. Para adotar essa abordagem a solução proposta nesta dissertação de mestrado têm como objetivo aplicar os mapas auto-organizáveis de Kohonen e a lógica nebulosa para gerar as vizinhanças entre as classes visando aplicação dessas técnicas em dois estudos de casos na classificação dos sinais provenientes dos sistemas elétricos de potência e sinais biomédicos adotando uma interpretação nebulosa dos mapas de Kohonen. O trabalho se divide basicamente em três etapas: na primeira, será realizada uma revisão das técnicas de mineração de dados e da lógica nebulosa mostradas na literatura; na segunda, concentra-se aplicar o algoritmo classificador utilizando redes neurais artificiais, especificamente redes neurais SOM como técnica de mineração de dados para efetuar a classificação dos sinais; na terceira etapa demonstramos a abordagem multidisciplinar da rede SOM e da lógica nebulosa como uma ferramenta alternativa aos métodos de mineração de dados.

Palavras-chave: mineração de dados, mapas auto-organizáveis, lógica nebulosa, fluxo de potência, diabetes mellitus.

Abstract

The Data Mining techniques, based on the Kohonen self-organizing maps have been largely used for classifying signals in several areas of expertise. Generally, the SOM network (Self-Organizing Maps) is used to specify similarity relationships between objects by adopting cluster analysis. The computational cost, data preparation and mathematical modeling can influence the interpretation of results, in which those from the evaluation classes are among its limitations. The Kohonen maps do not permit detailed evaluation of the class of objects, which may however be defined by the class limits, in other words defining a measure that can link when an object belonging to a particular class can migrate from one class to another . To adopt this approach the solutions proposed in this Masters dissertation are designed to implement the Kohonen self-organizing maps and the fuzzy logic to generate neighborhoods between classes aimed at applying these techniques on a two-case study for classifying signals from potential power systems and Biomedical output signals adopting an interpretation of the Kohonen nebula maps. The work is basically divided into three stages: the first which would be followed by a review of the data-mining techniques and fuzzy logic shown in literature; the second focuses on applying the classifier algorithm using artificial neural networks, specifically the usage of neural networks as SOM data mining techniques to enable the classification of signals while the third step demonstrates the SOM network fuzzy logic multidisciplinary approach as an alternative tool of the data-mining methods.

Keywords: Data Mining, Self-Organizing Maps, Fuzzy Logic, Power Flow, Diabetes Mellitus.

Sumário

1	Descoberta de Conhecimento	1
1.1	Introdução	1
1.2	Principais Tarefas do KDD	2
1.3	Motivação da Dissertação	3
1.4	Objetivos	4
1.5	Descrição da Dissertação	5
1.6	Organização da Dissertação	5
2	Mineração de Dados	7
2.1	Introdução	7
2.2	A Mineração de Dados e Análise Estatística	8
2.3	Funcionalidades da Mineração de Dados	8
2.4	Funções do Modelo	9
2.4.1	Classificação	9
2.4.2	Regressão	9
2.4.3	Análise de Associação	9
2.4.4	Análise de seqüência	10
2.4.5	Sumarização	11
2.4.6	Visualização	11
2.5	Representação do Modelo	13
2.5.1	Árvores de Decisão e Regras de Decisão	13
2.6	Análise de Agrupamento	16
2.6.1	Métodos de Agrupamento	17
2.6.2	Agrupamento Hierárquico	17
2.6.3	Método Aglomerativo	18
2.6.4	Métodos Divisivos	20
2.6.5	Métodos Particionais	21
2.7	Grafos	22
2.8	Métodos de Projeção	23
2.9	Considerações Finais	25
3	Redes Neurais	27
3.1	Introdução	27
3.1.1	O Neurônio Biológico	28
3.1.2	O Neurônio Artificial	29

3.1.3	Variações de RNAs	31
3.1.4	Redes SOM	33
3.1.5	Principais processos do algoritmo de Kohonen	35
3.1.5.1	O Processo Competitivo	36
3.1.5.2	Maximização do Produto Interno	36
3.1.5.3	O Processo Cooperativo	36
3.1.5.4	Função de Vizinhança	37
3.1.6	O Processo Adaptativo	39
3.1.7	Normalização dos Dados	39
3.1.8	O valor dos pesos no Melhoramento da Rede	40
3.1.9	Resumo do Algoritmo	40
3.1.10	Treinamento e Convergência	41
3.1.11	Interpretação do Mapa produzido pelo SOM	42
3.1.12	Matriz-U	42
3.1.13	Mapas Contextuais	43
3.1.14	Análise do Mapa	43
3.1.15	Considerações Finais	45
4	Lógica Nebulosa	47
4.1	Introdução	47
4.1.1	Conjuntos	48
4.1.2	Funções	49
4.1.3	Conjuntos Nebulosos	49
4.1.4	Propriedades	51
4.1.5	Operadores <i>Fuzzy</i>	52
4.1.5.1	Intersecção	52
4.1.5.2	União	53
4.1.6	Fuzzificação	54
4.1.7	Variáveis Lingüísticas	54
4.1.7.1	Modelagem das Variáveis Lingüísticas - Caso Fluxo de Potência	55
4.1.7.2	Modelagem das Variáveis Lingüísticas - Caso Sinais Biomédicos para pacientes portadores de Diabetes	55
4.1.8	Regras Fuzzy	56
4.1.9	Método de Mandani	56
4.1.10	Método de Takagi-Sugeno	57
4.1.11	Método de Tsukamoto	59
4.1.12	A base de Conhecimento	59
4.1.13	Mecanismo de Raciocínio	59
4.1.14	Defuzzificação	59
4.1.15	Método centro da área ou centróide	60
4.1.16	Método da Média Ponderada	60
4.1.17	Método Centro de Somas	60
4.1.18	Considerações finais	61

5	Construção do Sistema de Controle <i>Fuzzy</i> do Mapa de Kohonen	63
5.1	O Sistema de Controle	63
5.1.1	Considerações Finais	67
6	Aplicações em Mineração de Dados	
	Estudo de Casos	69
6.1	Conjunto de Dados - Caso Real: Problema do Fluxo de Potência	69
6.1.1	Introdução	69
6.1.2	Sistema Elétrico de Potência	70
6.1.3	Medidas Elétricas	72
6.1.4	Base de Dados	73
6.1.5	Análise preliminar da Base de Dados	73
6.1.5.1	Base de dados do SEP	73
6.1.6	Seleção de Variáveis	73
6.1.7	Divisão dos Dados	74
6.1.8	O Processo de Fuzzificação da Matriz de Dados pelo Método Mandani	78
6.2	Conjunto de Dados Públicos - Análise de Sinais Biomédicos	84
6.2.1	Introdução	84
6.2.2	Tipos de Diabetes	85
6.2.3	Estágios do diabetes	86
6.2.4	Complicações Clínicas	88
6.2.5	Técnicas de Mineração de Dados para Análise de Sinais Biomédicos	88
6.2.6	Análise preliminar da Base de Dados	89
6.2.6.1	Base de dados Diabetes	89
6.2.7	O Processo de Fuzzificação da Matriz	93
6.2.8	Considerações Finais	96
7	Conclusão	99
7.1	Contribuições	99
7.2	Extensões	99
A	Métodos Estatísticos utilizados em Mineração de Dados	101
A.1	Tipos de Regressão	101
A.1.1	Regressão Linear	101
A.1.2	Regressão Múltipla	101
A.1.3	Regressão Polinomial	104
B	Apontamentos Matemáticos	105
B.1	Função Gaussiana	105
C	Grandezas Elétricas	107
C.1	Fatores Típicos das Cargas	107
C.1.1	Demanda	107
C.1.2	Tensão	108
C.1.3	Fator de Potência	109

D Variabilidade da Carga - Dados Originais - Fator de Potência	113
E Rótulação dos Sinais de Glicemia	117

Lista de Figuras

1.1	Principais tarefas do processo do KDD [89]	2
1.2	Arquitetura do sistema <i>VisionFuzzy</i>	5
2.1	Mineração de dados e seu aspecto multidisciplinar [43]	8
2.2	Exemplo de uma regra de associação [19]	10
2.3	Diagrama de associação das chamadas telefônicas [19]	12
2.4	Faces de Chernoff [17]	13
2.5	Classificando dos pinos através de árvore de decisão [94]	14
2.6	Dados dos pinos - Classificação que caracteriza a variedade do pino como uma função do comprimento e do diâmetro [94]	15
2.7	Regra de decisão [19]	15
2.8	Classificação simplificada dos métodos de agrupamentos[38]	16
2.9	Classificação simplificada dos métodos de agrupamentos adaptado [38] e [100]	17
2.10	Ilustração do Critério de Ligação Simples e Completa	19
2.11	Diferentes tipos de representação de agrupamentos	20
2.12	Método <i>k-means</i> na formação de <i>clusters</i>	22
2.13	Algoritmo de caminho mínimo abrangendo a árvore de <i>clusters</i>	23
3.1	Neurônios do sistema nervoso central dos vertebrados: a) neurônio motor da medula espinhal; b) célula de Purkinje (apresenta dendritos extensamente ramificados) encontrada no cerebelo [53]	28
3.2	Modelo não-linear de um neurônio artificial [30]	29
3.3	Gráficos das Funções de Ativação empregadas na modelagem de neurônios [53]	30
3.4	RNA de multicamadas [53]	31
3.5	Modelo de Arquiteturas de RNAs [39]	31
3.6	Mapa citoarquitetural do córtex cerebral simplificado [52]	33
3.7	Arquitetura do mapa auto organizável [29]	34
3.8	Tipos de estímulos produzidos de acordo com a função chapéu de mexicano [11]	35
3.9	Exemplos de função de vizinhança de acordo com o critério de monotonicidade: a) base quadrada e b) base triangular [52] e [29]	37
3.10	Função de vizinhança gaussiana [30]	38
3.11	Ordenação topológica dos vetores de peso [52]	41
3.12	Exemplo de Matriz- U : (A) por quantidade de neurônios, (B) modelo por subdivisão de neurônios, (C) matriz de distância, (D) matriz com grade hexagonal e (E) matriz com variáveis no plano [92]	43
3.13	Exemplo de um mapa topológico com grades hexagonal e retangular [18]	43

3.14	Mapa contextual com rotulação [100]	44
4.1	Diagrama de blocos de um sistema fuzzy [23] e [82]	48
4.2	Função de pertinência da variável lingüística "Fator de Potência".	55
4.3	Função de pertinência da variável lingüística "DiagnósticoCurvaGlicêmica".	55
4.4	Função de pertinência $\mu A \cap B(x) = \min$	57
4.5	Função de pertinência $\mu A \cup B(x) = \max$	57
4.6	Representação esquemática do método de Mandani	58
4.7	Representação esquemática da saída final do método de Mandani	58
4.8	Algoritmo de identificação das regras fuzzy [13]	58
5.1	Seleção dos parâmetros e base de dados do sistema VisionFuzzy	63
5.2	Visualização do conjunto de exemplo	64
5.3	Treinamento da rede de acordo com os parâmetros	64
5.4	Treinamento da rede e visualização do mapa	65
5.5	Parâmetros do Controlador Fuzzy	65
5.6	Sistema de Controle	66
5.7	Visualização do conjunto de regras do sistema	66
5.8	Gráfico de controle para o exemplo de cargas	67
6.1	Diagrama de blocos do sistema [46]	71
6.2	Diagrama unifilar do sistema elétrico de potência [46]	72
6.3	Mapa contextual dos intervalos de carga por subestação	76
6.4	Mapa contextual por tipo de carga de acordo com as subestações	77
6.5	Funções de pertinência da variável lingüística de entrada Cza21as24CMedia.	78
6.6	Variável lingüística de saída Fator de Potência	79
6.7	Gráfico de superfície das subestações no intervalo de carga de 21 às 24h (carga média)	81
6.8	Variáveis lingüísticas de entrada e saída do controlador para o intervalo de carga de 21 às 24 horas	81
6.9	Variabilidade do fator de potência de acordo com o perfil da carga de todas as subestações - Método Modus Pones Generalizado	82
6.10	Variabilidade da carga de todas subestações após aplicação da lógica fuzzy	83
6.11	Estimativa da população com diabetes para 2030 [71]	86
6.12	Estimativa da população por idade 2030 [71]	87
6.13	Representação dos Sinais de Glicemia no Mapa de Kohonen	90
6.14	Similaridade dos Sinais de Glicemia no mapa de Kohonen - Classe I - Sinal 77 (pré-almoço 07:00 às 9:00) e Sinal 100 (pré-lanche 22:00)	90
6.15	Sinais de Glicemia no mapa de Kohonen - Classe I - Sinal 20 (pós-ceia 19:00 às 21:00) e Classe II - Sinal 90 (pré-ceia 15:00 às 17:00)	91
6.16	Similaridade dos Sinais de Glicemia no mapa de Kohonen - Classe II - Sinal 1 (pré-almoço 07:00 às 09:00) e Classe II - Sinal 89 (pré-almoço 07:00 às 09:00)	91
6.17	Mapa contextual por paciente	92
6.18	Funções de pertinência da variável lingüística de entrada Pré-almoço	93

6.19	Funções de pertinência da variável lingüística de saída Estágio de Glicemia	94
6.20	Gráfico de superfície de controle dos sinais de glicemia	95
6.21	Gráficos dos sinais das variáveis de entrada do controlador <i>fuzzy</i> , após a execução do mapa de Kohonen	96
6.22	Saída <i>Crisp</i> do controlador fuzzy proveniente do Estágio de Glicemia	96
A.1	Plotagem dos dados : embora os pontos não são abrangidos por uma linha reta, o modelo apresentado sugere uma relação linear entre x (anos de experiência) e y (salário)	103
B.1	Evolução da largura da gaussiana nas distribuições normais	105
C.1	Curva diária de demanda [46]	108
C.2	Modelo de transformação de energia elétrica em outras formas de energia [75]	109
D.1	Variabilidade da carga no intervalo de 00 às 07 - Carga Leve	113
D.2	Variabilidade da carga no intervalo de 07 às 18 - Carga Média	114
D.3	Variabilidade da carga no intervalo de 18 às 21 - Carga Pesada	114
D.4	Variabilidade da carga no intervalo de 21 às 24 - Carga Média	115

Lista de Tabelas

2.1	Seqüências de transações dos consumidores [19]	10
2.2	Características Seqüenciais com Suporte > 40% [19]	11
2.3	Representação tabular das chamadas telefônicas	12
2.4	Representação formal das Medidas de Distâncias:(A)Euclidiana,(B)Quadrado da Euclidiana,(C)Manhattan e (D)Chebychev.	18
6.1	Base de dados Originais: (A e B) período e intervalo das medições; (C,D,E)-valores de corrente das três fases (F) Fator de Potência	73
6.2	Intervalos de horários dos patamares de carga: (1)segunda/ sábado e (2) domingo/feriados (sem horário de verão); (3)segunda/ sábado e (4) domingo/feriados(com horário de verão) [70]	74
6.3	Rotulação dos dados no Mapa de Kohonen	75
6.4	Rotulação dos dados no Mapa de Kohonen: (A) refere-se ao mapeamento grosso do mapa considerado fase 1 e (B) refinamento do mapa na fase 2	77
6.5	Funções de pertinência das subestações :Cruz das Almas, Benedito Bentes, Tabuleiro, Trapiche das cargas residenciais.	79
6.6	Descrição do mapa de regras do controlador <i>fuzzy</i> de forma reduzida pelo método "Modus Ponens generalizado"[13]	80
6.7	Mapa de Regras por acesso aleatório	83
6.8	Os 10 países com a maior taxa de população diabética na faixa etária de 20-79 anos e estimativa para 2025 [5]	85
6.9	Interpretação dos resultados da glicemia de jejum e do teste de tolerância à glicose [81]	87
6.10	Amostra da base de dados Originais	89
6.11	Rotulação dos dados no Mapa de Kohonen	92
6.12	Funções de pertinência das medições glicêmia (Entrada/Saídas)	94
6.13	Mapa de Regras por acesso aleatório composto por: Estágio da Glicemia (EG), Glicemia Mínima (GM), Glicemia Normal (GN), Glicemia Alterada (GA)	95
A.1	Dados de Salário	102
A.2	Dados para o plano de regressão	102
E.1	(A) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen	118
E.2	(B) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen	119
E.3	(C) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen	120
E.4	Descrição detalhada dos rótulos (medições de glicemia)	121

Descoberta de Conhecimento

1.1 Introdução

Os sistemas de informações (SI) tradicionais construídos para apoiar o processo decisório geralmente armazenam seus dados em sistemas de banco de dados ou até mesmo em grandes repositórios como, por exemplo, o *data warehouse*. Segundo Inmon [35] e Kimbal [50], "Um data warehouse consiste de um banco de dados especializado capaz de manipular um grande volume de informações corporativas obtidas de banco de dados operacionais e de fontes de dados externos à organização, melhorando o desempenho, controle e acesso aos dados". De acordo com o tipo de informação, o processo de extração pode ser considerado complexo e poderá superar a capacidade humana de analisar essas informações, principalmente no que se refere em transformar os dados armazenados e esquecidos nesses repositórios em conhecimento.

Segundo Nonaka [68], "O conhecimento é um processo humano dinâmico de justificar a crença pessoal com relação a verdade, significa sabedoria adquirida a partir da personalidade como um todo e se torna um fator de grande importância para as organizações por ser considerado o principal caminho para o sucesso entre aquelas empresas (públicas ou privadas) que visam competitividade no mercado". De acordo com Stewart [84] "O conhecimento se tornou o ativo mais importante e indispensável, por ser considerado a principal matéria-prima com o qual todos trabalham, sendo assim, mais valioso e poderoso que qualquer outro ativo físico e financeiro".

Neste contexto, a informação sempre esteve presente nas atividades que envolvem pessoas, processos, sistemas, recursos financeiros e tecnologias e tem se tornado um grande diferencial no auxílio ao processo de tomada de decisão. Para que essas informações sejam extraídas e transformadas em conhecimento é necessária a utilização de técnicas e ferramentas que propiciem a descoberta ou mineração de padrões.

O processo de extrair conhecimento utilizável a partir de grandes volumes de dados chamado *Knowledge Discovery in Databases (KDD)* ou apenas descoberta de conhecimento em base de dados, pode ser considerado um processo de extração de informações relevantes ou de padrões nos dados contidos em grandes bancos de dados e que sejam não-triviais, implícitos, previamente desconhecidos e potencialmente úteis e tem como principal objetivo a construção de hipóteses para auxiliar processos de tomada de decisão [89].

1.2 Principais Tarefas do KDD

O processo de descoberta de conhecimento em bases de dados envolve uma seqüência de tarefas pré-estabelecidas (figura 1.1) de forma interativa e são fundamentais para a transformação das informações em conhecimento [43] e [89]. A figura 1.1, apresenta as principais tarefas do KDD, as quais serão descritas a seguir:

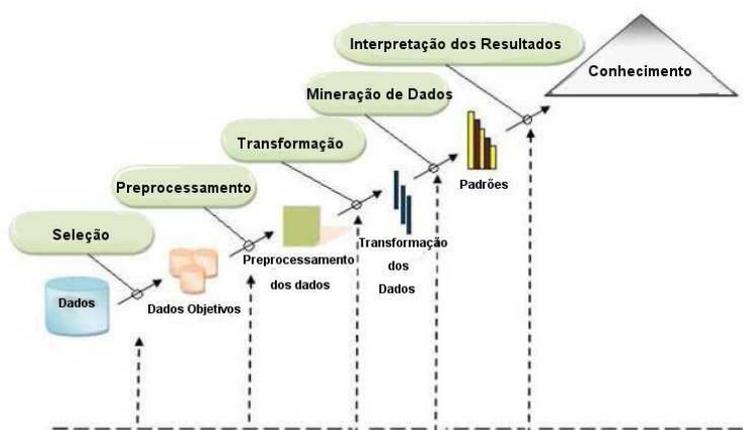


Figura 1.1 Principais tarefas do processo do KDD [89]

- **Análise de Requisitos** : corresponde ao processo de modelagem e especificação com objetivo de criar elementos necessários para o estudo e compreensão do domínio da aplicação.

- **Seleção dos Dados**: busca identificar um conjunto de dados relevantes e seus subconjuntos de variáveis e analisar a relevância entre eles, objetivando a criação de um conjunto de dados nos quais a "descoberta de conhecimento" será realizada.

- **Preprocessamento**: envolve atividades de limpeza dos dados, executando operações básicas de remoção dos ruídos, proporcionando a coleta de informações necessárias para a modelagem ao processo do KDD, decisão sobre os dados inexistentes, elaboração de esquema e mapeamento de valores desconhecidos.

- **Transformação nos dados**: utiliza métodos que proporcionam a redução ou transformação da dimensionalidade dos dados, possibilitando a redução do número de variáveis. Os métodos de redução de dados podem incluir tabulação simples, agregação (com estatística descritiva) ou técnicas mais sofisticadas como análise de agrupamentos [77].

- **Mineração de Dados** : no processo de mineração de dados ou (*data mining*) tem-se como objetivo aplicar técnicas específicas em dados pré-processados na busca por modelos de interesse numa forma particular de representação ou num conjunto de tais representações, incluindo

regras de classificação, árvores de decisão, regressão, agrupamento [19], [43] e [89].

De acordo com Diniz [19], Junglos [45] e Fayyad [89], a classificação associa um item em uma ou várias classes predefinidas, analisa um conjunto de dados de treinamento e constrói modelos para cada classe, com base nas características dos dados. Existem muitos métodos de classificação desenvolvidos no campo de aprendizagem de máquina, estatística, redes neurais tais como: o modelo de regressão que associa um item a uma ou mais variáveis de predição de valores reais; análise de associação que determina relações entre campos de um banco de dados; análise de *cluster* que associa um item comum ou vários agrupamentos determinados pelos dados, onde as classes são predeterminadas e definidas através de agrupamentos naturais dos dados, baseados em medidas de similaridades ou modelos probabilísticos; sumarização descreve de forma compacta um subconjunto de dados a partir de métodos estatísticos. As técnicas mais recentes incluem a lógica nebulosa ou *fuzzy*, [32], [49], [97] e [99] e os algoritmos genéticos [96].

- **Interpretação dos Dados :** Permitem avaliar padrões dos dados com o objetivo de determinar quais são as melhores maneiras de usar tais informações na tomada de decisão, capazes de incorporar o conhecimento ao processo ou melhorar o conhecimento anterior.

1.3 Motivação da Dissertação

Devido a necessidade das organizações (públicas ou privadas) em extrair informações relevantes das bases de dados que possa apoiar o processo de tomada de decisão a partir do conhecimento extraído dessas bases de dados, faz com que muitos métodos estatísticos de análise de dados sejam considerados de fundamental importância para a coleta e manipulação de dados. Isto se deve, principalmente quando tratamos de tarefas que permitem reunir os dados semelhantes de acordo com algum interesse particular capaz de gerar conhecimento construído a partir desses métodos. Muitas abordagens alternativas podem ser vista na literatura em relação aos métodos estatísticos tradicionais [52, 77]. Dentre elas, as redes neurais vêm sendo considerada uma abordagem atrativa quando se pensa em analisar problemas que envolvem relacionamento complexos para classificar ou reconhecer padrões associando-os a um padrão de entradas [30].

Os Mapas Auto-Organizáveis de Kohonen [52] conhecidos como redes SOM é um tipo de rede neural artificial baseada em aprendizado competitivo e não supervisionado, bastante utilizado em atividades de mineração de dados e têm como objetivo mapear um conjunto de dados permitindo ampliar a capacidade de análise de agrupamentos pertencentes a um espaço de elevada dimensão, preservando ao máximo possível a topologia do espaço original de dados de forma adaptativa [11], [25], [28], [43],[52] e [100]. Pesquisas recentes têm abordado aplicação das redes SOM em áreas multidisciplinares, conforme o que segue:

- Qualidade de sedimentos integrando características físicas e químicas, análise de *cluster* hierárquico com os métodos estatísticos tradicionais e avaliações da qualidade biológica do solo [2];

- Análise e modelagem de recursos hídricos [47];
- Identificação de componentes visuais de *clustering* [34];
- Implementação de algoritmo de quantização vetorial na extração de características dos sinais de eletrocardiograma [86];
- Comparação da rede SOM com Análise de Componentes Principais (PCA) para interpretar dados multidimensionais a partir de um processo biológico [4] entre outros.

A classificação dos dados realizada pela rede SOM trabalha especificamente nas relações de similaridades entre os neurônios e o conjunto de dados, realizando uma projeção não linear do espaço de dados de entrada, sua principal característica na preservação máxima da topologia original dos dados [11], [29], [30] e [52]. Por exemplo, um cliente que pertence a classe (adimplente) poderá ser considerado da classe (inadimplente) em um determinado período de tempo devido a um problema pessoal que proporcionou esta transformação. A interpretação dos dados obtidos pela rede SOM exibem a relação do objeto pertencente a uma classe específica. As redes SOM não são capazes de avaliar os objetos da classe que estão no limiar estabelecido pelo limite da classe, proporcionando a definição de vizinhança entre as classes ou uma medida de quando um objeto pertencente a uma determinada classe migre de uma classe para outra. Essas limitações permitem que técnicas complementares como a lógica nebulosa possam ser utilizadas para tratar as limitações proveniente da rede SOM resultante de problemas complexos, incertos, contraditórios e incompletos [21], [97], [99].

1.4 Objetivos

Esta dissertação têm como principal objetivo:

- Estruturar um processo de mineração de dados a partir da interpretação nebulosa dos mapas de Kohonen, que seja capaz de proporcionar uma visão das vizinhanças entre classes geradas pela rede SOM;
- Desenvolver um sistema que gere automaticamente a visão unificada do processo;
- Analisar o comportamento do sistema desenvolvido a ser aplicado em dois estudos de casos: o primeiro relacionado ao problema do fluxo de potência utilizando dados reais e segundo na análise e diagnóstico de doenças proveniente de pacientes com diabetes(dados públicos).

1.5 Descrição da Dissertação

Neste trabalho será apresentado o desenvolvimento de um sistema que integra a classificação proveniente do mapa de Kohonen e a lógica nebulosa, como pode ser visto na Figura 1.2.

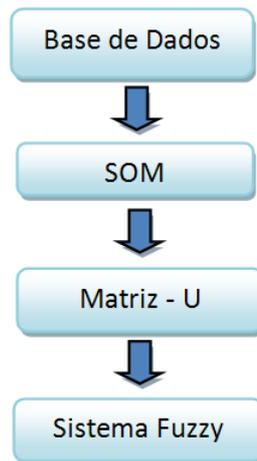


Figura 1.2 Arquitetura do sistema *VisionFuzzy*

A figura 1.2 apresenta uma arquitetura básica do sistema *VisionFuzzy*. Inicialmente, a base de dados poderá ser importada para o sistema executar as tarefas de agrupamento (*clustering*). Nesta fase, o usuário deverá inserir os parâmetros da rede. Durante o treinamento da rede SOM, não existe um padrão de entrada associado a uma saída esperada, o algoritmo explora as similaridades entre o conjunto de dados e um agrupamento específico para efetuar o processo de classificação a partir da matriz de distância unificada (Matriz-U), em seguida, o mapa contextual é gerado. Após o processo de classificação o sistema exibirá o módulo para inclusão dos parâmetros do sistema com objetivo de apresentar a visão unificada e nebulosa do mapa de Kohonen.

1.6 Organização da Dissertação

De acordo com os objetivos definidos na Seção 1.5 deste capítulo, organizou-se esta dissertação em sete capítulos, incluindo esta introdução. No capítulo 2, apresenta-se uma breve revisão da bibliográfica sobre os principais métodos de mineração de dados abordando as principais técnicas de agrupamento aplicados ao problema de classificação. No capítulo 3 apresenta-se as principais características da rede neural SOM e seus critérios de convergência a partir de sua modelagem matemática. O capítulo 4 apresenta uma introdução à modelagem matemática da lógica nebulosa. No capítulo 5, será apresentada a implementação da rede neural SOM e o processo de fuzzificação do mapa de Kohonen a partir do modelo proposto. No capítulo 6, serão apresentadas duas aplicações; a primeira no que se refere à análise de sinais proveniente dos sistemas elétricos de potência com foco no problema do fluxo de potência e a segunda na

análise de sinais biomédicos de pacientes portadores de diabetes. No capítulo 7, será apresentada a conclusão dessa dissertação e as perspectivas de trabalhos futuros.

Mineração de Dados

Este capítulo aborda uma breve revisão da literatura adotando os principais métodos existentes de análise multivariada de dados e mineração de dados com critério de classificação. Serão abordados os principais métodos de análise de agrupamentos e de projeção, sendo discutidos as principais características e suas aplicações.

2.1 Introdução

Na década de 70, muitos especialistas foram instruídos a armazenar seus dados em qualquer recurso tecnológico (disco rígidos, fitas magnéticas, banco de dados etc.), que fornecesse segurança. Com a evolução da tecnologia e o surgimento de novos métodos de armazenamento de dados e a popularização dos Sistemas de Gerenciamento de Banco de Dados (SGBD) como recursos da tecnologia da informação (TI), favoreceram a proliferação da informação. Com o surgimento dos sistemas de apoio a decisão (SAD) na década de 80 e a necessidade de reduzir o impacto das integrações entre sistemas de diversas plataformas, tanto no que se refere aos custos com a tecnologia da informação quanto ao aumento da velocidade de processamento dos sistemas de informação (SI), fez com que novas tecnologias fossem adotadas para o processo de armazenamento de dados[35].

A idéia de se criar um banco de dados (BD) para armazenar os registros do sistema, fez com que o tamanho desses bancos crescesse rapidamente [35]. A tecnologia adotada como *data warehouse* permite atender sistemas de informação capazes de produzir transações de alto desempenho com objetivo de armazenar e cruzar grande volume de dados [3], [19] e [35]. Para que esses dados sejam manipulados e posteriormente transformados em conhecimento, faz necessário a utilização de técnicas sofisticadas que propiciem a automação do comportamento inteligente a partir da inteligência artificial, a qual envolve técnicas necessárias para a compreensão da linguagem, percepção, raciocínio, aprendizagem e resolução de problemas, buscando a criação de teorias e modelos com capacidade cognitiva e a implementação de sistemas computacionais baseados nestes modelos focando as técnicas do KDD [13], [78] e [80]. Segundo Quispe [76], A mineração de dados ou *data mining* provém da análise inteligente e automática de dados para descobrir padrões ou regularidades em grandes conjuntos de dados, através de técnicas que envolvem métodos matemáticos, algoritmos baseados em conceitos biológicos, processos lingüísticos e heurísticas, os quais fazem parte do processo do KDD responsável pela busca de conhecimentos em banco de dados [3], [19], [43],[73] e [89].

2.2 A Mineração de Dados e Análise Estatística

Os processos estatísticos são muito usados na coleta, manipulação e análise de dados, conceitos como distribuição normal, variância, análise de regressão, análise de dispersão dos dados, análise discriminante, análise de agrupamento, intervalo de confiança e testes de hipótese são utilizados para realizar pesquisas nos dados, bem como analisar e descobrir relacionamentos entre eles [45].

As técnicas estatísticas como análise multivariada de dados combinadas com as técnicas de mineração de dados possuem como função extração de informação relevante em um conjunto de dados [43]. Na mineração de dados é possível utilizar métodos estatísticos tradicionais ou técnicas mais sofisticadas. Por exemplo, a inteligência artificial, pode fazer com que a mineração de dados possa ser visto como o descendente direto da estatística e surge exatamente no limite do que poderia ser encontrado e inferido por métodos tradicionais de análise de dados, tratando de questões que estão além do domínio desses procedimentos [3], [19] e [43].

2.3 Funcionalidades da Mineração de Dados

A multidisciplinaridade da mineração de dados pode ser considerada inevitável devido à integração de diversas áreas de conhecimento no processo de análise abordando áreas de pesquisas que envolvem estatística, matemática e a computação, as quais são disciplinas fundamentais para realização do processo de mineração de dados.

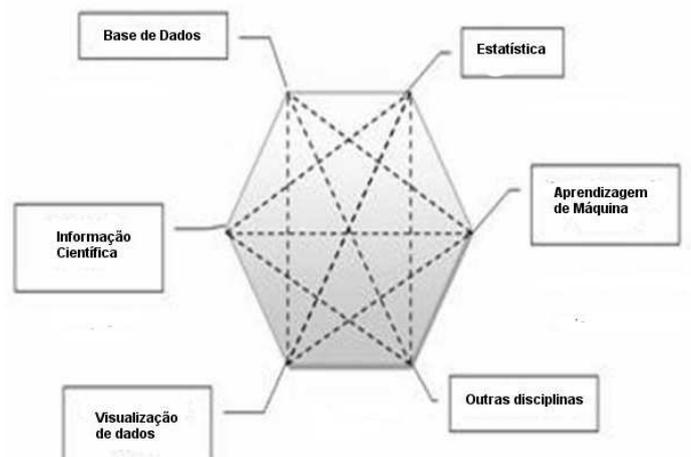


Figura 2.1 Mineração de dados e seu aspecto multidisciplinar [43]

Algumas limitações podem ser identificadas quando se refere à escolha do melhor método de mineração de dados, uma vez que, existe uma grande dificuldade dos especialistas na identificação do melhor método minerador. Geralmente, os especialistas buscam adotar o método mais adequado para a resolução do problema específico. O processo de mineração de dados pode ser dividido em componentes capazes de favorecer a identificação mais adequada dos

algoritmos de mineração a ser levado em consideração algumas informações relevantes tais como: a função do modelo e a representação do modelo.

2.4 Funções do Modelo

As funções do modelo são utilizadas para especificar o tipo de aplicação do algoritmo minerador, as mais comuns são descritas a seguir:

2.4.1 Classificação

A função de classificação tem por objetivo analisar um conjunto de dados que tenham as mesmas características. Segundo Quispe [76], "A classificação é uma função de previsão que pode ser usada para encontrar um modelo que classifique um item de dado entre várias classes previamente definidas. Uma vez que o algoritmo classificador foi desenvolvido de forma eficiente será usado de forma preditiva para classificar novos registros naquelas mesmas classes pré-definidas". Assim, um modelo de predição com classificação pode ser usado para estabelecer uma específica classe para cada registro do banco de dados. A classe deve ser de um conjunto finito de possíveis e predeterminados valores de classe [19]. Por exemplo, um sistema de classificação pode ser treinado para identificar o perfil de clientes inadimplentes na concessão de crédito, a partir das informações cadastrais armazenadas no banco de dados da empresa e pode ser usado como suporte a tomada de decisão no momento de conceder o crédito.

2.4.2 Regressão

Trata-se de um conjunto de métodos e técnicas que permite a interpretação da relação funcional entre as variáveis com boa aproximação, bem como, tentar formar idéia da existência de uma relação entre as variáveis de cada conjunto, de tal modo que essa medida possa estabelecer modelos utilizados para fins de predição. As aplicações abordando métodos de regressão podem ser utilizadas em diversas áreas de conhecimento, por exemplo: prevê a economia nacional com base em certas informações (níveis de renda, investimentos e assim por diante), quais fatores ajudam a manter a qualidade dos serviços oferecidos, viabilidade de um novo produto ou o retorno esperado de um novo empreendimento, prever as séries de tempo onde as variáveis de ingresso podem ser intervalos irregulares. A análise de regressão pode ser considerada uma ferramenta analítica poderosa para explorar todos os tipos de relações de dependência [44]. Uma técnica de regressão bastante utilizada no ponto de vista gerencial e estatístico é a análise regressão multivariada que pode ser usada para analisar a relação entre uma única variável dependente ou diversas variáveis independentes conforme os tipos de regressão [77] e [31], os tipos de regressão quais podem ser vistos com maiores detalhes no Apêndice A.

2.4.3 Análise de Associação

Análise de associação tem como objetivo elaborar uma representação explícita entre os objetos, visando determinar relacionamentos entre conjuntos de itens de associação. "Gera redes

de interações e conexões presentes nos conjuntos de dados usando as associações item a item. Entende-se que a presença de um item implica necessariamente na presença do outro item na mesma transação [19]". Uma regra de associação pode ser representada formalmente através do tipo, tanto X como Y considerado um conjunto de valores (produtos comprados por um cliente, sintomas apresentados por um paciente, etc.). A figura 2.2, exemplifica uma regra de associação voltada a identificar afinidades entre itens de um subconjunto de dados.

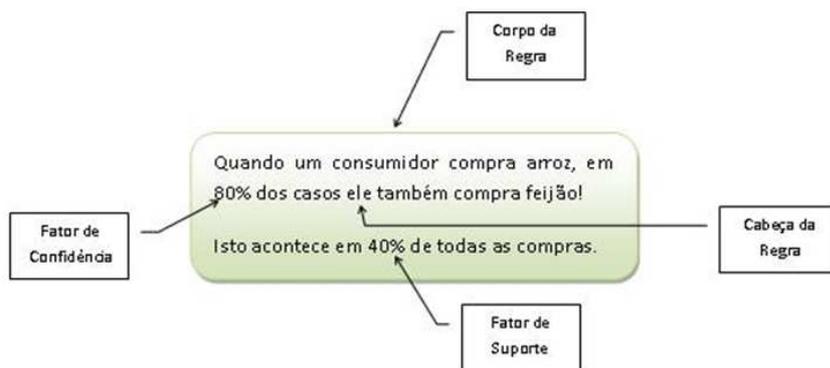


Figura 2.2 Exemplo de uma regra de associação [19]

2.4.4 Análise de seqüência

A análise de seqüência constitui-se de uma variação da análise associativa objetivando extrair e registrar desvios e tendências no tempo. As regras identificadas são usadas para reconhecer seqüências relevantes que possam ser utilizadas para prever comportamentos, modelar processos gerando uma seqüência ou relatando tendências de um processo ao longo do tempo. Assim, dado um conjunto de dados ordenado pelo sobrenome do consumidor e pelo período de transação, por exemplo, José Oliveira visitou a loja em dois consecutivos dias. Ele comprou cerveja no primeiro dia, vodka no segundo dia, a tabela 2.1, mostra as seqüências de transações dos consumidores organizadas segundo o tempo. Cada conjunto de parênteses indica uma transação que inclui um ou mais itens.

Consumidor	Seqüências dos consumidores
José Oliveira	(Cervejas) (Vodka)
João Soares	(Guaraná, Suco), (Cerveja) (Água, Licor, Vinho) , (Gin, Licor)
Pedro Tenório	(Cerveja), (Água, Gin, Vinho), (Vodka, Soda)
José Zappa	(Vodka)

Tabela 2.1 Seqüências de transações dos consumidores [19]

As técnicas de busca de características seqüencial detectam características entre transações de tal forma que a presença de um conjunto de itens é seguido por outro conjunto de itens em

um banco de dados de transações em um período de tempo [3], [19] e [89]. A técnica determina a frequência de cada combinação de transações que pode ser produzida nas seqüências de consumidores e disponibilizam as características seqüências cujas ocorrências relativas são maiores que um nível de suporte mínimo requerido. A tabela 2.2, apresenta as características seqüenciais com suporte maior que 40%. A característica seqüencial "cerveja é comprada em uma transação anterior a transação em que vodka é comprada" ocorre em dois dos quadros consumidores.

Características Seqüenciais	Consumidores de Apoio
com Fator de Sustentação > 40%	
(Cervejas), (Vodka)	(José Oliveira, Pedro Tenório)
(Cerveja), (Vinho, Água)	(João Soares, Pedro Tenório)

Tabela 2.2 Características Seqüenciais com Suporte > 40% [19]

2.4.5 Sumarização

A função de sumarização visa obter uma descrição compacta de um conjunto de dados, bastante usada na análise exploratória de dados objetivando a geração automatizada de relatórios [19], [73], [76]. Geralmente, a sumarização não é usada para a resolução de problemas, mas possibilita identificar características no conjunto de dados que possa sofrer ruídos que interfiram no processo de análise. A função de caracterização descreve as qualidades relevantes a partir de análises quantitativas, propiciando uma descrição compacta do conjunto, podendo generalizar, sumarizar e possivelmente contrastar características de dados. As funções de sumarização e caracterização tendem a ser complementares [76]. A sumarização é usada, principalmente, no pré-processamento dos dados, onde valores inválidos, no caso de variáveis quantitativas, são determinados através do cálculo de medidas estatísticas tais como mínimo, máximo, média, moda, mediana e desvio padrão amostral e, no caso de variáveis categóricas, através da distribuição de frequência dos valores [19] e [43]. Outras técnicas complementares de sumarização podem ser vista na literatura, a mais sofisticada destina-se a técnica de visualização de dados, a qual têm sido parte integrante da análise estatística tornando-se de extrema importância para se obter de um entendimento, muitas vezes indutivo do conjunto de dados [31] e [77].

2.4.6 Visualização

As técnicas de visualização podem ser consideradas uma ferramenta poderosa para se analisar grande quantidade de dados. Em muitas situações, elas são suficientes para a extração das respostas de interesse, descobrindo padrões, tendências, estruturas e relações dentro de um conjunto de dados [43]. O método de visualização escolhido para análise dependerá basicamente do tipo do conjunto de dados disponível e como esses dados podem ser modelados, por exemplo, se o conjunto de dados envolve chamadas telefônicas feitas em um intervalo de tempo específico, então uma representação visual desta informação poderia ser sumarizada através

o representará [12] e [100]. Outros métodos de visualização de dados, incluem: diagramas baseados em proporções, diagramas de dispersão, histogramas, box plots entre outros [77] e [31]. Os modelos que incluem a representação dos dados através de figuras poligonais podem ser visualizados a partir das "faces de *Chernoff*". As faces de *Chernoff*, é considerada uma técnica para ilustrar tendências em dados multidimensionais [16] e [17]. As faces representadas por este modelo ilustram características faciais para representar dados em diferentes dimensões capazes de representar tendências em termos de valores nos dados e podem ser utilizados para visualizar graficamente dados multivariados complexos. As faces apresentadas na figura 2.4, ajudam ao usuário na detecção de padrões, agrupamentos e correlações entre os dados, os quais são simplificados a partir de desenhos proveniente da face humana.

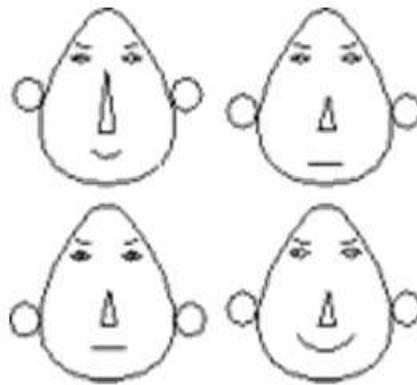


Figura 2.4 Faces de Chernoff [17]

2.5 Representação do Modelo

As funções do modelo têm um papel importante na análise e modelagem do problema. Porém, a integração da função e representação do modelo podem ser consideradas um dos fatores de grande relevância, uma vez que, os modelos representados a partir de algoritmos de mineração de dados, podem determinar a flexibilidade do mesmo em representar o conjunto de dados e a sua interpretação. Os modelos mais complexos podem ajustar melhor os dados, entretanto, ficam mais difíceis de serem interpretadas [19], representações mais tradicionais incluem árvore de decisão [3], conjunto de regras [43], métodos de agrupamento [89], modelos lineares [77] e não lineares [100] e [31], os quais são descritos a seguir.

2.5.1 Árvores de Decisão e Regras de Decisão

Quando o processo de mineração de dados é direcionado ao critério de classificação, o método de árvore de decisão pode ser conveniente quando o objetivo relaciona-se a categorização dos dados. As árvores de decisão são ferramentas poderosas e populares para classificação e diagnóstico [76]. A estrutura de uma árvore de decisão pode ser ilustrada conforme figura 2.6,

onde cada nodo interno identifica um dos atributos de previsão. Cada linha que sai desde nodo identifica um valor que poderá ser assumido por tal nodo; cada folha identifica o resultado da previsão ou objetivo.

A árvore é formada por nós e é no primeiro deles, o nó raiz, que envolve todo o conjunto de dados, onde o processo de classificação se inicia. O nó raiz testa todos os itens para o comprimento $\leq 0,75$. Itens que satisfazem este teste vão para o arco abaixo pela esquerda (verdadeiro), permanecendo em um nó (nó terminal ou folha), indicando que todos os pinos pertencem a uma classe (Quadrado) e nenhum teste será mais necessário. O arco a direita (Falso) do nó raiz recebe todos os casos que falharam no teste inicial. Esses pinos ainda não pertencem a uma só classe e, portanto, futuros testes serão necessários ao nó intermediário.

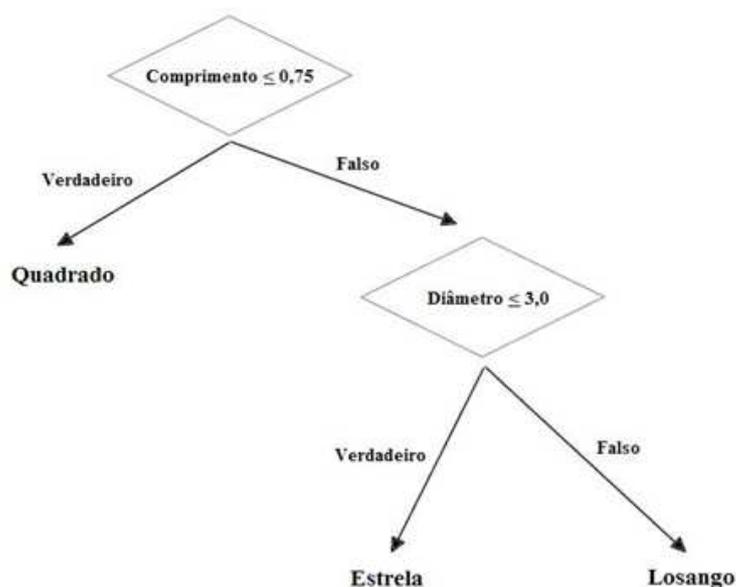


Figura 2.5 Classificando dos pinos através de árvore de decisão [94]

O teste nesse nó é para o diâmetro $\leq 3,00$. Os pinos que satisfazem o teste estão todos em uma só classe (Estrela) e aqueles que falharam também estão em uma classe (Losango), e assim ambos deste nó intermediário conduziram a nós terminais [94]. A integração dos métodos de árvore de decisão e regra de decisão pode ser considerada ferramentas fundamentais como técnicas de previsão. As regras de decisão podem ser considerados um processo para analisar uma série de dados e a partir dela gerar padrões [56]. Os modelos de predição utilizando árvores de decisão e regras de decisão deverão ser capazes de predizer corretamente situações já conhecidas antes de iniciar a predição de novas situações [43]. Um exemplo simples e hipotético apresentado por Weiss [94] e interpretado por Diniz [19], é apresentado para assimilação dessas técnicas. Suponha que estão disponíveis dados sobre o comprimento e diâmetro de uma série de pinos que podem ter o formato de quadrado, de estrela ou de losango. Uma classificação que caracteriza a variedade do pino como uma função do comprimento e do diâmetro pode ser útil

para se entender como essas variedades diferem. Os dados são ilustrados na figura 2.5, apresenta duas linhas paralelas aos eixos, uma no comprimento = 0,75 e outra no diâmetro = 3,00, que parecem particionar as três variedades em três diferentes sub-áreas. Métodos de solução por decisão por árvore fornecem automaticamente estas partições de eixos paralelos.

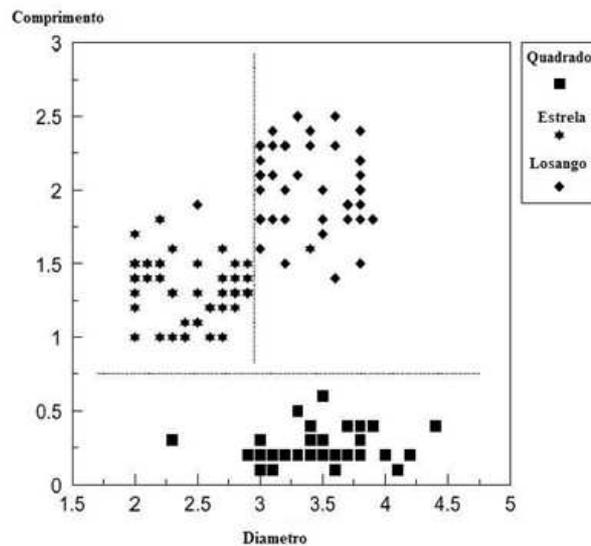


Figura 2.6 Dados dos pinos - Classificação que caracteriza a variedade do pino como uma função do comprimento e do diâmetro [94]

As regras de decisão são soluções alternativas baseadas em decisão por árvore. Uma regra pode ser construída através da formação de um conjunto de testes que ocorrem nos caminhos entre nó raiz e os nós terminais. A coleção de todas tais regras obtidas considerando cada caminho nó raiz a um nó terminal é uma solução, baseada em regras, para a classificação. No exemplo dos pinos, utilizamos para ilustrar a decisão por árvore, a solução por regra:

Se (*Comprimento* \leq 0,75)
 Então Quadrado

Se (\neg (*Comprimento* \leq 0,75)) & (*Diâmetro* \leq 3,00)
 Então Estrela

Se (\neg (*Comprimento* \leq 0,75)) & (\neg (*Diâmetro* \leq 3,00))
 Então Losango

Figura 2.7 Regra de decisão [19]

Gerada uma solução utilizando árvore de decisão ou regra de decisão, esta pode ser usada para estimar ou prever a resposta ou classe variável para um novo caso.

2.6 Análise de Agrupamento

A prática de classificar objetos de acordo com similaridades percebidas pode ser considerada a base inicial para vários aspectos da ciência, o principal objetivo da análise de agrupamento (*cluster*) está relacionado ao processo de agrupar elementos de dados mediante o particionamento de uma população heterogênea em vários subgrupos mais homogêneos. De acordo com Jain [38] "A análise de agrupamento é o estudo formal dos algoritmos e dos métodos para agrupar, ou classificar objetos". No agrupamento, não há classes pré-definidas, os elementos são agrupados de acordo com a semelhança, o que a diferencia da tarefa de classificação, buscando reunir indivíduos ou objetos em grupos tais que os objetos no mesmo grupo são mais parecidos uns com os outros do que com os objetos de outros grupos [43]. A idéia é maximizar a homogeneidade de objetos dentro dos grupos, ao mesmo tempo em que maximiza a heterogeneidade entre os grupos [44].

Os métodos de agrupamentos vêm sendo abordadas em diversas áreas de conhecimento e pode variar desde ciências biológicas (por exemplo, criar a taxonomia biológica para a classificação de vários grupos de animais) às ciências sociais (por exemplo, analisar vários perfis psiquiátricos), faz com que os métodos de agrupamentos sejam usados de forma adequada no processo de análise de dados. Os métodos de agrupamento são aqueles que buscam dividir um conjunto de objetos não rotulados em grupos (partições) de forma que os objetos de cada grupo tenham mais semelhanças entre si do que em relação aos objetos de qualquer outro grupo [100]. O principal objetivo da análise de agrupamentos é definir a estrutura dos dados colocando as observações mais parecidas em grupos. Os grupos são determinados de forma a obter-se homogeneidade dentro dos grupos e heterogeneidade entre eles. Segundo Jain [38], análise de agrupamento é um tipo especial de classificação. A figura 8, mostra uma árvore abordando diferentes métodos de agrupamento aplicados ao problema de classificação de forma simplificada, os quais são descritos a seguir:



Figura 2.8 Classificação simplificada dos métodos de agrupamentos[38]

2.6.1 Métodos de Agrupamento

A análise de agrupamentos é considerada uma técnica analítica para desenvolver subgrupos significativos de indivíduos ou objetos e têm como objetivo classificar uma amostra de entidades em um pequeno número de grupos mutuamente excludentes, com base nas similaridades entre eles [44]. Esta técnica pode ser dividida em três etapas: a primeira relaciona-se a medida de similaridade ou associação entre as entidades para determinar quantos grupos realmente existem na amostra; a segunda refere-se a busca do processo de agrupamento, nos quais entidades são particionadas em grupos, e o último passo busca em estabelecer o perfil das variáveis para determinar sua composição. Os métodos analisados são expostos na figura 2.9. O critério de classificação utilizando os métodos intrínsecos é a essência da análise de agrupamento [38]. Adaptações ao modelo simplificado aos métodos de agrupamentos podem ser visto em Alvin [77] e Zuchini [100] e Jain [38], os quais podem ser representados graficamente a partir da figura 2.9.

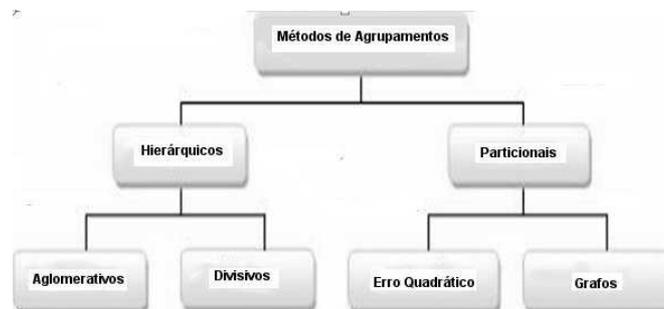


Figura 2.9 Classificação simplificada dos métodos de agrupamentos adaptado [38] e [100]

2.6.2 Agrupamento Hierárquico

A análise de agrupamentos corresponde ao grupo de técnicas multivariadas de dados cuja finalidade primária é agregar objetos com base nas características que eles possuem. Entre essas técnicas encontram-se o procedimento hierárquico, o qual opera para formar um intervalo inteiro de soluções de agrupamentos [44]. Também, pode ser considerado um método aglomerativo que objetiva fundir agrupamentos individuais (inicialmente cada grupo contém um único objeto) em partições maiores até a obtenção de uma única partição contendo todos os objetos do conjunto, onde os agrupamentos são formados pela combinação de outros já existentes. O procedimento hierárquico trata o conjunto de dados como uma estrutura de partições, cada uma correspondendo a um agrupamento, hierarquicamente organizadas segundo a similaridade entre seus objetos [38], [77] e [100]. A maioria dos métodos de análise de agrupamento requer uma medida de similaridade entre os elementos a serem agrupados, normalmente expressos como uma função distância ou métrica. A correspondência ou associação de dois objetos baseada nas variáveis da variável estatística de agrupamento pode ser contextualizado como formação da medida de similaridade. Em Joseph [44], a similaridade pode ser medida a partir de uma medida de associação, com coeficientes de correlação positivos maiores representando maior similaridade. A proximidade entre cada par de objetos pode avaliar a similaridade onde

medidas de distância ou de diferença são empregadas e as menores distâncias ou diferenças representam maior similaridade [31], [77] e [43].

Representação Formal	Medidas de Similaridade
$X=[x_1, x_2, x_3, \dots, x_n]$ e $Y = [y_1, y_2, y_3, \dots, y_n]$	Objetos
$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$	A
$d_{xy} = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 = \sum_{i=1}^p (x_i - y_i)^2$	B
$d_{xy} = x_1 - y_1 ^2 + x_2 - y_2 ^2 + \dots + x_n - y_n ^2 = \sum_{i=1}^p x_i - y_i ^2$	C
$d_{xy} = \text{maximo}(x_1 - y_1 + x_2 - y_2 + \dots + x_n - y_n)$	D

Tabela 2.4 Representação formal das Medidas de Distâncias:(A)Euclidiana,(B)Quadrado da Euclidiana,(C)Manhattan e (D)Chebychev.

A tabela 2.4, exibe a representação formal de muitas medidas de similaridade, muitas dessas medidas poderão ser adotadas na análise de agrupamentos. Porém, a métrica mais utilizada é a distância euclidiana, por ser um procedimento comum quando nenhuma outra informação prévia existente acerca dos dados de entrada e poderá afetar diretamente na quantidade de formação dos grupos encontrados pelos algoritmos de agrupamento, onde os aspectos de estrutura do espaço podem (ou não) ser levados em consideração durante o processo de acordo com a métrica escolhida [30], [38], [52], [77] e [100].

2.6.3 Método Aglomerativo

Os métodos hierárquicos envolvem a construção de uma hierarquia de uma estrutura do tipo árvore. As técnicas aglomerativas mais populares podem ser utilizada para descobrir agregado e são divididos em:

a) **Ligação Individual** : o procedimento de ligação individual ou ligação simples é baseado na distância mínima. Ele encontra os dois objetos separados pela menor distância e os coloca primeiro no agrupamento. Em seguida, a próxima distância mais curta é determinada, e um terceiro objeto se junta aos dois primeiros para formar um agregado, ou um novo agrupamento de dois membros é formado. O processo continua até que todos os objetos formem um só agregado. Esse procedimento também foi chamado de abordagem do vizinho mais próximo.

b) **Ligação Completa**: o procedimento de ligação completa é semelhante ao da ligação simples ou individual, exceto em que o critério de agrupamento se baseia na distância máxima. Por essa razão, às vezes é chamado de abordagem do vizinho mais distante ou de método do

diâmetro [44]. A distância máxima entre indivíduos em cada agregado representa a menor esfera (diâmetro mínimo) que pode incluir todos os objetos em ambos os agrupamentos. Esse método é chamado de ligação completa porque todos os objetos em um agrupamento são conectados um com o outro a alguma distância máxima ou similaridade mínima. Podemos dizer que a similaridade interna se iguala ao diâmetro do grupo. Esta técnica elimina o problema de encadeamento identificado na ligação individual.

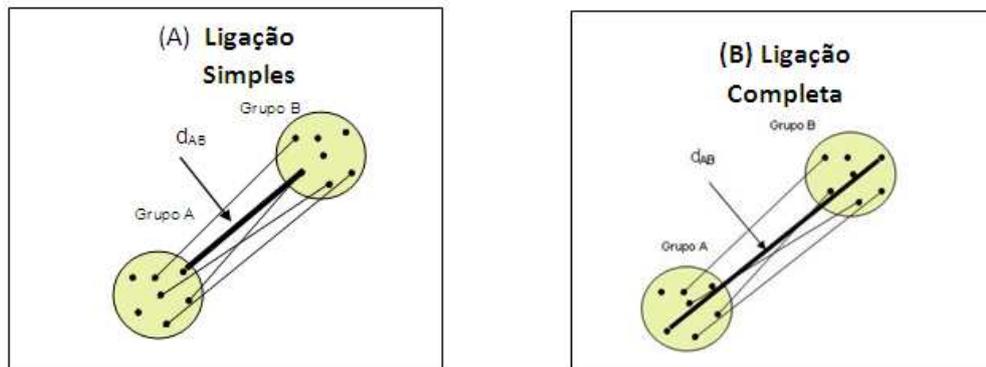


Figura 2.10 Ilustração do Critério de Ligação Simples e Completa

O critério de ligação simples define d_{AB} como a menor distância entre todos os pares (x, y) de dois objetos onde $x \in A$ e $y \in B$ (Figura 2.10), considerando o critério de ligação completa utilizando como a maior distância entre todos os pares (x, y) . Assim, depois de calculadas as distâncias entre os agrupamentos os algoritmos promovem a união dos agrupamentos com a menor distância entre si [100].

c) **Ligação Média:** o método de ligação média começa da mesma forma que a ligação individual ou completa, mas o critério de agrupamento é a distância média de todos os indivíduos em um agrupamento aos demais em um outro. Esta técnica não depende de valores extremos, como ocorre na ligação individual ou completa, e a partição é baseada em todos os elementos dos agregados, ao invés de um único par de membros extremos. Abordagens de ligação média tendem a combinar agregados com pequena variação interna. Elas também tendem a produzir agregados com aproximadamente a mesma variância.

d) **Método de Ward:** no método do ward, a distância entre dois agrupamentos é a soma dos quadrados entre os dois agrupamentos feita sobre as variáveis. Em cada estágio do procedimento de agrupamento, a soma interna de quadrados é minimizada sobre todas as partições (o conjunto completo de agrupamentos disjuntos ou separados) que podem ser obtidas pela combinação de dois agregados do estágio anterior. Este procedimento tende a combinar agrupamentos com um pequeno número de observações. Ele também tende a produzir agregados com aproximadamente o mesmo número de observações.

e) **Método Centróide:** no método centróide, a distância entre dois agrupamentos é a distância

(geralmente euclidiana quadrada ou euclidiana simples) entre seus centróides. Centróides são valores médios das observações na variável estatística de agrupamento. Neste método, toda vez que indivíduos são reunidos, um novo centróide é computado. Os centróides migram quando ocorrem fusões de agregados. Em outras palavras, existe uma mudança no centróide do agrupamento toda vez que, um novo indivíduo ou grupo de indivíduos é acrescentado a um agregado já existente [44]. Esses algoritmos diferem na forma como a distância entre os agrupamentos é computada. A saída gerada por esses algoritmos podem proporcionar a criação de gráficos demonstrando o processo de formação desses agrupamentos, conforme o que segue:

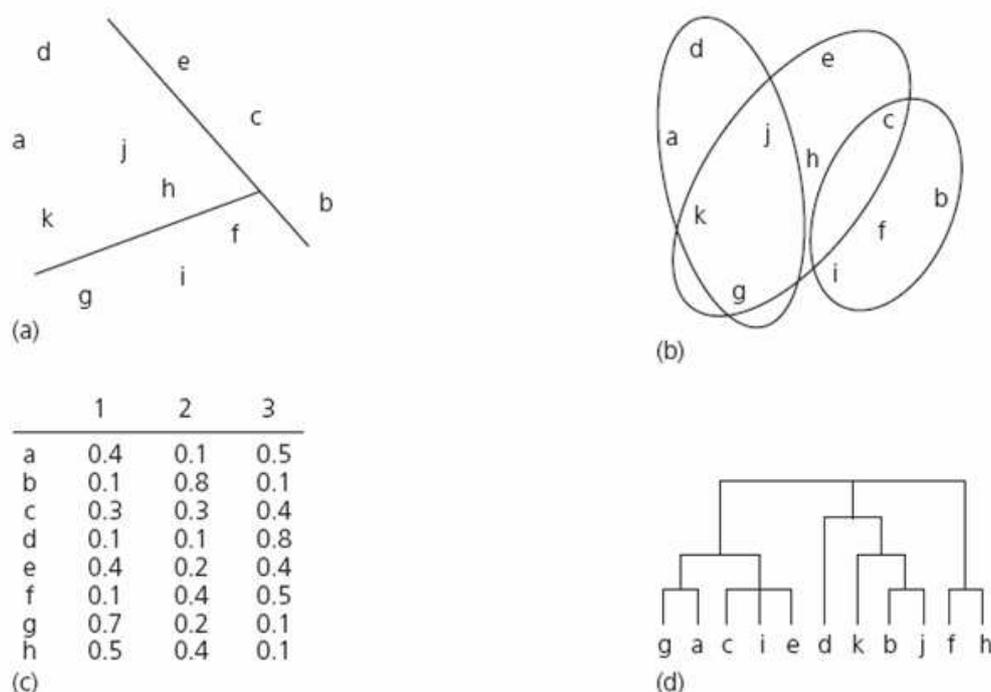


Figura 2.11 Diferentes tipos de representação de agrupamentos

Os diagramas representados na figura 2.11, exibem uma representação gráfica de diferentes tipos de *clusters*, as representações mais comuns são a do modelo dendrograma. A palavra grega *dentro* significa "uma árvore", ou seja, diagramas em árvore [95]. Os dendrogramas pode ser considerado uma espécie de grafo em árvore que representa as junções sucessivas de partições e que pode gerar agrupamentos diferentes conforme o nível que é seccionada [77] e [100].

2.6.4 Métodos Divisivos

Quando o processo de classificação utiliza métodos divisivos ocorre uma inversão com o método hierárquico aglomerativo. Pode-se observar que um conjunto contendo todos os dados é particionado a partir de um aglomerado unificado. Os métodos divisivos começam o processamento a partir de um grande agregado que contém todas as observações (objetos). Em passos

sucessivos, as observações mais diferentes entre si são separadas e transformadas em agrupamentos menores, ou seja, podemos considerar a existência de uma única partição (o próprio conjunto de dados), subdividindo esta partição em uma série de partições alinhadas [38] e [44].

Os métodos hierárquicos são bastante utilizados no processo de análise multivariada de dados. Porém, a literatura mostra que em bases de dados de grande porte esses métodos podem se tornar impraticável por ser difícil visualizar as informações contidas no dendrograma e nos demais tipos de gráficos resultados dessa classificação. Segundo Jain [38], essas técnicas são impraticáveis quando o número de objetos é elevado, fato bastante comum no processo de mineração de dados.

2.6.5 Métodos Particionais

Os métodos particionais têm como objetivo dividir um conjunto de objetos em um número pré-estabelecido de *clusters*. O método mais popular é conhecido como *k-means*. Este algoritmo divide o conjunto de dados em partes disjuntas, satisfazendo as seguintes recomendações:

- a) objetos de uma mesma parte estão próximos de acordo com um critério de dado;
- b) objetos de partes distintas que estão longe, de acordo com este mesmo critério.

A subdivisão realizada pelo algoritmo *k-means* como método particional, opera da seguinte forma: cria-se uma partição inicial aleatória de K partes e posteriormente em um processo iterativo, os elementos das partes vão sendo realocados para outras partes de modo a melhorar o particionamento a cada iteração, isto é, de modo que cada parte realmente contenha objetos que estão próximos e objetos em partes distintas estejam longe um do outro. Segundo Jain [38], Rencher [77] e Zuchini [100], os métodos particionais dividem o conjunto dos N objetos em K agrupamentos sem relacioná-los hierarquicamente entre si, como o fazem métodos hierárquicos.

Normalmente as partições são obtidas pela otimização de um critério local definido (sobre um subconjunto de objetos ou globalmente na forma de uma função objetivo). A aproximação dos objetos pode ser analisada a partir da formação de uma matriz de distância ou das similaridades de acordo com uma métrica estabelecida (no caso da distância euclidiana). Assim, podemos descrever os objetos do conjunto de dados por uma simples representação matemática, onde $X = [x_1, x_2, x_3, \dots, x_n]$ representa o conjunto de objetos de um número $K \leq X$ representando o número de *clusters* que se deseja formar. Seja $C = [C_1, C_2, C_3, \dots, C_n]$ uma partição do conjunto de dados em k *clusters* e sejam os elementos escolhidos em cada um dos *clusters*, representando os centros dos mesmos conhecidos na literatura como os centróides. Cada entidade influencia o grupo, cujo *protótipo* $= [v_1, v_2, v_3, \dots, v_k]$ os elementos escolhidos em cada um dos *clusters*, representando os centros de área ou centróides. Cada entidade $x_i (i = 1, \dots, n)$, influencia o grupo $C_k (k = 1, \dots, K)$, cujo protótipo v_k está mais próximo. Os centróides constituem valores médios dos objetos contidos no agrupamento sobre cada variável usados nas variáveis estatísticas de agrupamento ou no processo de validação. Toda vez que os objetos são reunidos, um novo centróide é computado [44], [77] e [100] e os objetos são realocados a partir de v_k :

$$v_k = \frac{1}{m} \sum_{j=1}^m x_j^k \quad (2.1)$$

Uma característica relevante do *k-means* está direcionada ao emprego de uma função objetivo conhecida como erro quadrático utilizado na aproximação das partições e pode ser representado matematicamente por:

$$e_k^2 = \sum_{j=1}^K \sum_{i=1}^N \|x_i - v_k\|^2 \quad (2.2)$$

O *k-means* recebe como entrada um número k de agrupamentos e atribui aleatoriamente um objeto como sendo o centróide inicial de cada agrupamento. Sucessivamente, cada objeto é associado ao agrupamento mais próximo e o centróide de cada agrupamento é então recalculado levando-se em conta o novo conjunto de objetos a ele pertencentes [100].

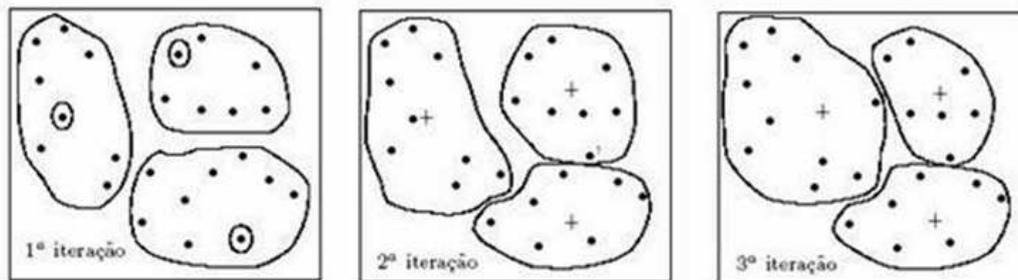


Figura 2.12 Método *k-means* na formação de *clusters*

Na figura 2.12, podemos observar as iterações do algoritmo *k-means* utilizando $k = 3$, os objetos selecionados em círculos representa a escolha aleatória dos k protótipos, nas iterações seguintes os centróides são marcados pelo sinal de $+$. Em seguida, podemos visualizar o critério de convergência do algoritmo que pode ser analisado a partir das observações referente as trocas de objetos. O algoritmo *k-means* converge quando ocorrem poucas trocas de objetos entre os grupos ou quando o valor de e_k^2 é minimizado, ou até mesmo quando v_k não se altera em duas iterações consecutivas. De acordo com Rencher [77] e Zuchini [100], o método *k-means* possui sua maior vantagem quando atua sobre um conjunto de dados com elevado número de objetos.

2.7 Grafos

Os agrupamentos de dados baseado em grafos utiliza algoritmo baseado na construção de uma árvore geradora mínima (*Minimum Spanning Tree - MST*) e tem como objetivo à geração de um grafo de modo que os objetos não possua ciclos e seja conectado por um arco, ou seja, uma árvore. Assim, os agrupamentos são obtidos a partir do primeiro arco de maior comprimento de produção dos aglomerados [41] e [100].

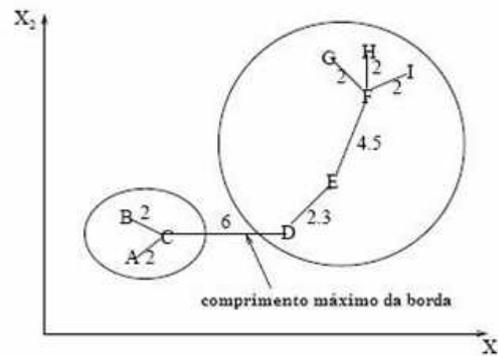


Figura 2.13 Algoritmo de caminho mínimo abrangendo a árvore de *clusters*

Os agrupamentos obtidos pelo algoritmo MST são subgrafos bastante similares ao método de agrupamento aglomerativo, especificamente ligação simples e completa. Maiores detalhes do algoritmo pode ser visto em Anil [41].

2.8 Métodos de Projeção

Os algoritmos de projeção visto em Anil [38], Jain [41] e Zuchini [100], procuram realizar um conjunto de testes a partir do mapeamento de objetos no espaço d -dimensional de n em um espaço m -dimensional, onde $m < d$. O principal objetivo da realização desses testes é permitir a análise visual dos dados utilizando técnicas que possam exibir uma estrutura do espaço original o mais fielmente possível no hiperplano de projeção, possibilitando assim uma análise de agrupamentos que pode ser realizada visualmente caso $m = 2$ ou $m = 3$ e que poderá servir para validar resultados obtidos por outros métodos de mineração de dados [100]. Quando esses testes são executados a partir de uma projeção linear podemos obter novas características de como a combinação linear das características originais de d . Segundo Jain [38], uma projeção linear pode ser representada por:

$$y_i = x_i, i = 1, \dots, n \quad (2.3)$$

Onde x é uma matriz m, d , que gera os vetores $y = [y_1, y_2, y_2, \dots, y_d] \in R^d$, podendo ser representado por uma combinação linear de suas colunas. Os algoritmos de projeção lineares são bastante simples de usar. O mais popular é a projeção de autovetores que também pode ser conhecida como método de *Karhunen – Loeve* ou Análise de Componentes Principais (PCA) [30]. Este algoritmo tem como objetivo encontrar um método para condensar dados originais obtidos a partir de um conjunto de variáveis com dimensão elevada em um conjunto menor de variáveis com uma perda mínima de informação. O PCA é um método de identificar padrões nos dados, visando expressar os mesmos de modo a salientar as similaridades e diferenças existentes [77]. Essas diferenças podem ser denominadas de processo de seleção ou extração de características [83]. De acordo Haykin [30], "a seleção de característica se refere a um processo no qual um espaço de dados é transformado em um espaço de característica que, em teoria, tem

exatamente a mesma dimensão que o espaço original de dados. Entretanto, a transformação é projetada de tal forma que o conjunto de dados pode ser representado por um número reduzido de características "efetivas" e ainda reter a maioria do conteúdo de informação intrínseco dos dados; em outras palavras, o conjunto de dados sofre uma redução de dimensionalidade". Outros autores como Haykin [30], Simar [31] e Thornhilla [87], consideram o PCA um método simples, principalmente quando se refere a escolha dos componentes principais e na formação do vetor característica considerado uma das fases relevantes do processo de redução de dimensionalidade.

O método PCA toma um conjunto $X = [x_1, x_2, x_3, \dots, x_n]$ (dados transformados com a subtração da média) representando uma realização do vetor de entrada. Para as n soluções possíveis para o vetor q , podem ser constatadas a existência de n projeções possíveis do vetor de dados x [30] a serem considerados por:

$$a_i = q_i^T x = x q_i^T, i = 1, 2, \dots, n \quad (2.4)$$

$$Q = [q_1, q_2, q_3, \dots, q_n] (\text{autovetores}) \quad (2.5)$$

Onde a_i são considerados as projeções de x sobre as direções principais representadas pelos vetores de entrada, também conhecidos como componentes principais e possuem as mesmas dimensões físicas que o vetor de dados x . A equação 2.8, pode ser vista como uma fórmula de análise. Para reconstruir exatamente o vetor de dados original x a partir das projeções a_i , considere $[a_i | i = 1, 2, 3, \dots, n]$, combinações do conjunto de projeções em um único vetor a partir de:

$$Q = Q^T x \quad (2.6)$$

Multiplicado pela matriz Q e então usando a relação de :

$$Q^T = Q^{-1} \quad (2.7)$$

Conseqüentemente, o vetor de dados original x pode ser reconstruído por:

$$X = Qa = \sum_{i=1}^m a_i q_i \quad (2.8)$$

Neste sentido, os vetores representam uma base no espaço de dados, ou seja, não é nada mais do que uma transformação de coordenadas, de acordo com a qual um ponto x no espaço de dados é transformado em um ponto a correspondente um espaço características. Desta forma, a praticidade na análise de componentes principais constitui em fornecer uma técnica efetiva para redução de dimensionalidade, reduzindo o número de características necessárias para a representação efetiva dos dados descartando aquelas combinações lineares que têm variância pequenas e retendo apenas aqueles termos de variâncias grandes [83].

2.9 Considerações Finais

Este capítulo buscou fazer uma revisão da literatura, abordando os principais métodos de mineração de dados, não tendo a intenção de ser uma obra completa. Neste contexto, outras técnicas sofisticadas podem ser vistas na literatura como forma complementar aos métodos de mineração de dados, por exemplo, a teoria dos conjuntos nebulosos por Zadeh [99] oferece subsídios para a extensão dos métodos de agrupamento *fuzzy cluster* [63] e [82]. Em Anil [41] faz uma breve revisão dos métodos de agrupamentos, inclusive naqueles baseados em algoritmos genéticos. Outros métodos de projeção podem ser vistos em Simar [31], Joseph [44], Rencher [77] e Zucchini [100], os quais incluem: análise discriminante, escalonamento multidimensional, projeção de sammon, curvas principais, entre outros.

Redes Neurais

3.1 Introdução

Desde os mais remotos tempos, o ser humano vive envolvido com o problema de lidar com o ambiente que o cerca. A ciência tem evoluído na busca de entender e prever o comportamento do universo e dos sistemas que o compõe criando modelos adequados que conciliem as observações feitas sobre esses sistemas. Com o ressurgimento das redes neurais artificiais na década de 80, muitas definições abordando sistemas de processamento paralelo e distribuídos foram abordados na literatura e conhecidas como conexionismo. Esta forma de computação abordando redes neurais trabalha com abordagem de decisões não-estruturadas e relembram a estrutura do cérebro humano por não ser baseada em regras ou em programas algoritmos e se constitui um método alternativo à computação algorítmica convencional.

As redes neurais têm sido empregadas com sucesso em várias áreas onde a abordagem convencional tem falhado em fornecer soluções satisfatórias. Essas abordagens fazem parte do conjunto de técnicas estudadas na área de inteligência computacional [80]. Segundo Haykin [30], as redes neurais artificiais se constituem em uma técnica de inteligência artificial, cujo objetivo é simular o processo de funcionamento do cérebro humano, através de um neurônio artificial e possui a capacidade de armazenar conhecimento experimental e torná-lo disponível para uso e têm apresentado desempenho satisfatório em diversas áreas de conhecimento, tais como: classificação, reconhecimento de padrões, mineração de dados, aproximação de funções, processamento de séries temporais. Essas práticas estão se tornando mais presente no nosso dia-a-dia e apresentam-se, atualmente, como uma abordagem alternativa aos métodos estatísticos tradicionais na solução de problemas não triviais [53].

As redes neurais artificiais (RNAs) são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas (normalmente não-lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para considerar a entrada recebida por cada neurônio da rede. Em Haykin [30], uma rede neural pode ser definida como um processador distribuído maciçamente paralelo feito de simples unidades de processamento, que tem a propensão natural de guardar conhecimento experimental e torná-lo disponível para uso. Uma rede neural assemelha-se ao cérebro em dois aspectos:

1. O conhecimento de um domínio é adquirido pela rede durante o processo de aprendizado;

2. As forças de conexões entre neurônios, conhecidos como pesos sinápticos, são usadas para guardar o conhecimento adquirido".

Em Mêuser [90], as redes neurais artificiais podem ser definidas como um sistema constituído por elementos de processamento interconectados, chamados de neurônios, os quais estão dispostos em camadas e são responsáveis pela não-linearidade e pela memória da rede. Em Braga [11], "as RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas (normalmente não lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede e o funcionamento destas redes é inspirado em uma estrutura física concebida pela natureza: o cérebro humano".

3.1.1 O Neurônio Biológico

A célula nervosa ou neurônio biológico é composta por uma fina membrana e tem como função biológica normal processar informações. A partir do corpo celular ou soma, o centro dos processos metabólicos da célula nervosa projeta-se extensões filamentosas, os dendritos e o axônio [53]. Os dendritos cobrem o volume muitas vezes maior do que o próprio corpo celular e formam uma árvore dendrital. A outra, a projeção filamentar do corpo celular, o axônio, também chamado de fibra nervosa, serve para conectar a célula nervosa com outras do sistema nervoso, conforme figura 3.1.

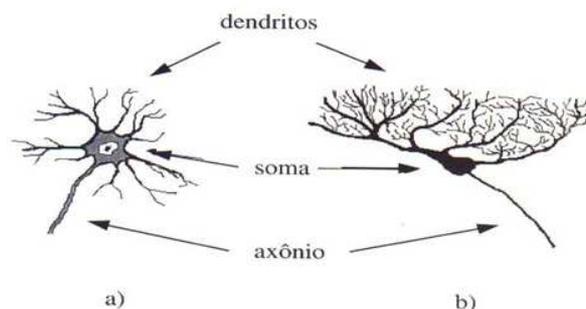


Figura 3.1 Neurônios do sistema nervoso central dos vertebrados: a) neurônio motor da medula espinhal; b) célula de Purkinje (apresenta dendritos extensamente ramificados) encontrada no cerebelo [53]

O neurônio recebe sinais (impulsos) de outros neurônios através de seus dendritos (receptores) e transmite os sinais gerados pelo corpo celular através do axônio (transmissor). Os dendritos e o axônio possuem ramificações em cujos terminais são realizadas as sinapses. As sinapses são unidades estruturais e funcionais elementares que promovem as interações entre os neurônios. Assim, nas descrições tradicionais da organização neural, uma sinapse é assumida pela conexão simples a qual pode impor ao neurônio receptivo o processo de excitação ou

inibição e a partir dela a efetividade da sinapse pode ser ajustada pelo sinal que passa através dela estabelecendo o processo de aprendizagem através da atividade a qual ela participa.

3.1.2 O Neurônio Artificial

O neurônio artificial pode ser considerado uma unidade de processamento de informação e possui fundamental importância para a operação de uma rede neural. O diagrama de blocos exposto pelo Haykin [30] na figura 3.2, apresenta a base do projeto das redes neurais artificiais, a qual se constitui em três elementos básicos do modelo neuronal:

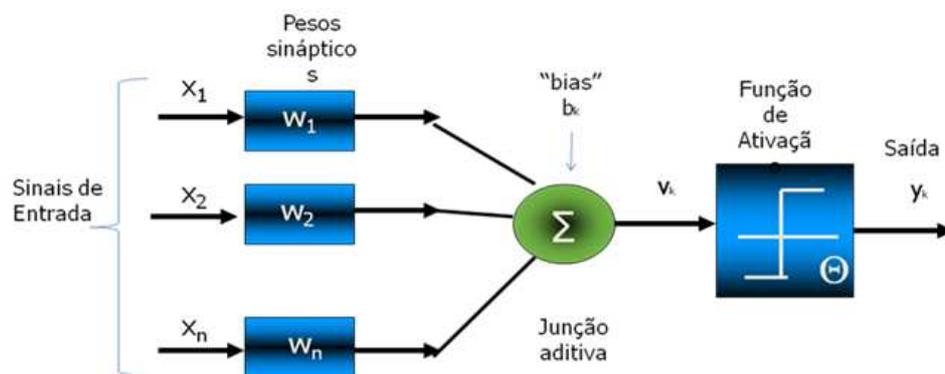


Figura 3.2 Modelo não-linear de um neurônio artificial [30]

1. Um conjunto de sinapses ou elos de conexão, cada uma caracterizada por um peso ou força própria. Especificamente, um sinal x_j na entrada da sinapse j conectado ao neurônio k é multiplicado pelo peso sináptico $w_{k,j}$. O primeiro índice se refere ao neurônio em questão e o segundo se refere ao terminal de entrada da sinapse a qual o peso se refere.
2. Um somador para somar os sinais de entrada, ponderado pelas respectivas sinapses do neurônio.
3. Uma função de ativação (ver figura 3.3) para restringir a amplitude da saída de um neurônio. A função de ativação (limita) o intervalo permissível de amplitude do sinal de saída a um valor finito. Geralmente, o valor normalizado da amplitude da saída de um neurônio pode ser escrito como o intervalo unitário fechado $[0,1]$ ou alternativamente $[-1,1]$.

As topologias mais comuns das RNAs são as de múltiplas camadas (figura 3.4) e as recorrentes, onde os neurônios recebem diretamente as entradas da rede (camada de entrada) e posteriormente são definidas as demais camadas (intermediárias) até a camada final (saída) e as camadas internas são chamadas de camadas ocultas.

Uma rede neural multicamadas de K camadas, terá como entrada um vetor x de dimensão j_0 de componentes x_{j_0} , onde $j_0 = 1, 2, 3, \dots, j_0$. Estas se conectam às entradas dos j_1 neurônios numa primeira camada. As saídas u_{1j_1} , $j_1 = 1, 2, 3, \dots, j_1$, destes, formando as componentes de um novo vetor u_1 de dimensão j_1 , conectando-se as entradas J_2 neurônios da camada seguinte

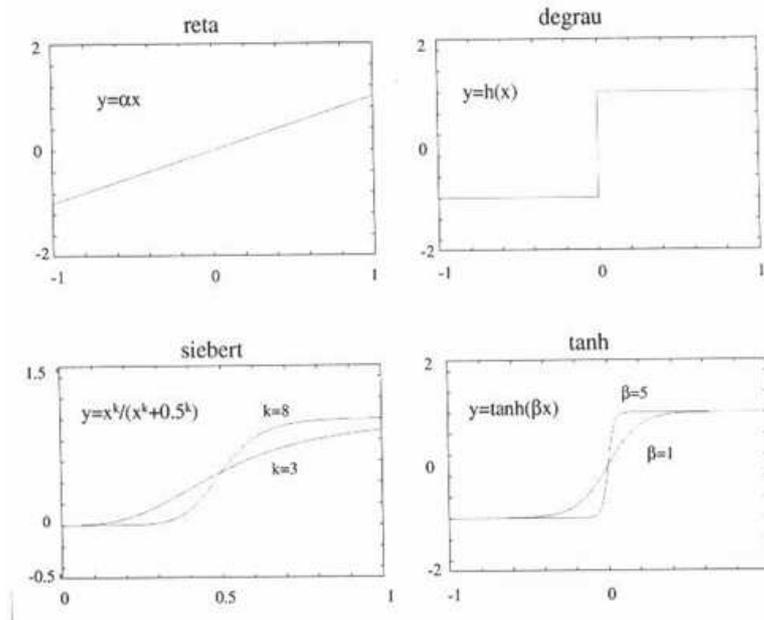


Figura 3.3 Gráficos das Funções de Ativação empregadas na modelagem de neurônios [53]

e assim sucessivamente até a última camada que consistirá de J_k neurônios, fornecendo como saída da rede um vetor $y = u_k$ de dimensão k . Assim, u_{kj} denota a saída do j -ésimo neurônio na k -ésima camada, sendo que para $k = 0$ representa a J_k -ésima entrada da rede, e para $k = K$ a j_k -ésima saída da rede.

As características mais importantes das RNAs em relação ao cérebro humano está relacionado a capacidade de aprender e que são responsáveis pela modificação dos pesos sinápticos, em função dos exemplos de entrada que lhe são apresentados. As regras de aprendizados podem ser divididas em:

- a) **Aprendizado supervisionado:** o aprendizado da rede é feito com o conhecimento prévio do resultado desejado, ou seja, são fornecidos para a rede, o conjunto de exemplo de entrada e as respectivas respostas.
- b) **Aprendizado não supervisionado:** a rede aprende com os próprios dados de entrada (somente os exemplos de entrada são mostrados à rede), ou seja, não requer conhecimento das saídas.

A combinação das características expostas no parágrafo anterior, juntamente com a capacidade de aprender a partir das experiências realizadas durante o processo de treinamento, o *perceptron* de múltiplas camadas deriva seu poder computacional. A teoria das redes neurais artificiais vem se consolidando mundialmente, como uma nova eficiente ferramenta para se lidar com a ampla classe dos problemas complexos, onde extensas massas de dados devem ser modeladas e analisadas em um contexto multidisciplinar, envolvendo, simultaneamente, tanto

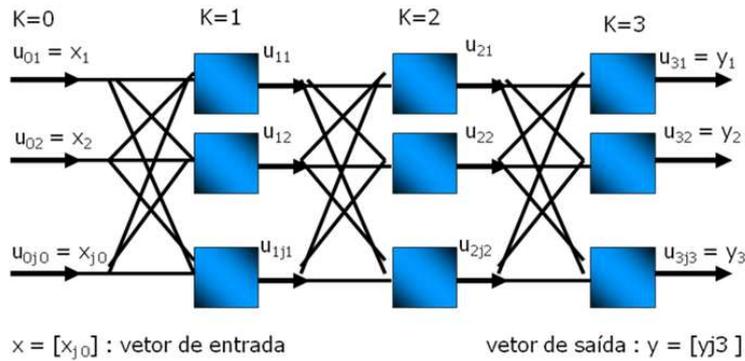


Figura 3.4 RNA de multicamadas [53]

os aspectos estatísticos e computacionais como os dinâmicos e de otimização [53].

3.1.3 Variações de RNAs

Segundo Jain [39], as redes neurais artificiais podem ser agrupadas em duas categorias:

1. Redes alimentadas adiante (*feed-forward*), nas quais os grafos não possuem loops e os sinais se propagam em um único sentido, isto é, da camada de entrada para camada de saída;
2. Redes recorrentes (*feedback*), nas quais os loops ocorrem por causa das conexões de retroalimentação.

A figura 3.5, apresenta modelos de arquitetura das RNA's, conforme a seguir:

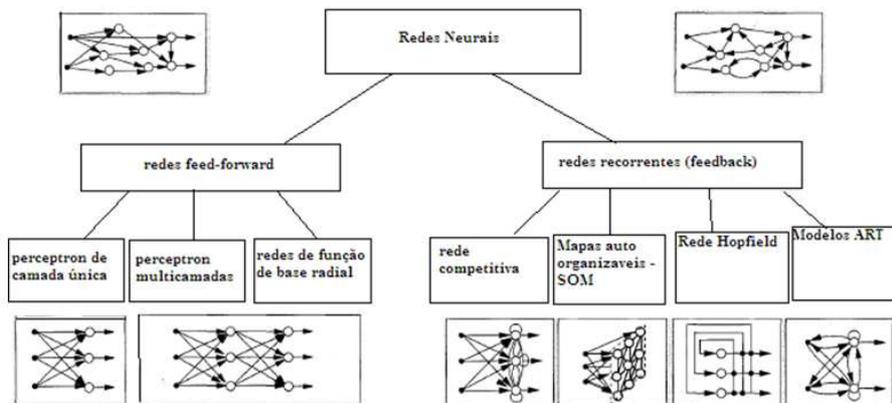


Figura 3.5 Modelo de Arquiteturas de RNAs [39]

- **Perceptron de camada única:** forma mais simples de uma rede neural, consiste de um

único neurônio com pesos sinápticos ajustáveis e "bias".

- **Perceptrons multicamadas:** a mais popular classe de RNAs proveniente de múltiplas camadas, considerada uma generalização do perceptron de camada única, também conhecidas com redes *MultiLayer Perceptrons* (MLP), tornaram-se populares devido ao surgimento do algoritmo de aprendizagem conhecido como algoritmo *backpropagation*, o qual tem como objetivo a busca de um mínimo global através de ajustes nos pesos sinápticos por um valor que é proporcional ao sentido contrário da derivada (o gradiente) do erro fornecido pelo neurônio em relação ao valor do peso.

- **Redes RBF (Radial Basis Function):** constitui-se de uma rede neural onde as unidades ocultas fornecem um conjunto de "funções" que forma uma "base" arbitrária de padrões (vetores) de entrada, quando eles são expandidos sobre o espaço oculto. Geralmente, sua forma mais básica, envolve três camadas: a camada de entrada que é constituída por nós fonte (unidades sensoriais) que conectam a rede ao seu ambiente; a segunda camada oculta a rede, a qual aplica uma transformação não-linear do espaço de entrada para o espaço oculto; na maioria das aplicações, o espaço oculto é de alta, dimensionalidade; a camada de saída é linear, fornecendo a resposta da rede ao padrão (sinal) de ativação aplicado à camada de entrada [30].

- **Mapas auto organizáveis:** classe especial de redes neurais baseada em grades, a qual se constitui da aprendizagem competitiva, onde os neurônios de saída da grade competem entre si buscando a maior saída. Assim, os neurônios vencedores são colocados em nós de uma grade que normalmente pode ser uni ou bidirecional, bastante utilizada em mineração de dados.

- **Redes de Hopfield:** consiste de um conjunto de neurônios e um conjunto correspondente de atrasos unitários, formando um sistema de realimentação de múltiplos laços, onde o número de laços de realimentação é igual ao número de neurônios.

- **Modelos "Adaptive Resonance Theory "(ART):** tipo de rede neural desenvolvida para solucionar o dilema da estabilidade e plasticidade, características mais importantes das RNAs voltada a sua habilidade de generalização, ou seja, produzir respostas para padrões de entrada que são similares, mas não idênticos, aos padrões apresentados à rede durante o seu treinamento. A arquitetura básica da ART envolve duas camadas de nodos; uma camada de entrada que processa os dados de entrada e uma camada de saída que agrupa os padrões de treinamento clusters. Estas camadas estão conectadas por meio de dois conjuntos de conexões que conectam cada nodo de uma camada a todos os nodos da outra. O primeiro conjunto é representado pelas conexões *feedforward* que assume valores reais e segue da camada de entrada para a camada de saída; o segundo conjunto que contém as conexões *feedback*, assume os valores binários e conecta os nodos da camada de saída aos nodos de entrada. Nesta dissertação, a rede SOM foi escolhida como técnica de mineração de dados, por ser considerada um modelo adaptativo para análise de agrupamento em grandes bases de dados.

3.1.4 Redes SOM

As redes SOM (*Self - organizing Maps*), foram desenvolvidas por Teuvo Kohonen [52] na década de 80 e pertencem à classe de algoritmos de codificação vetorial. Estas redes produzem um mapeamento topológico que localiza otimamente um número fixo de vetores em um espaço de dimensionalidade mais elevada e desse modo facilita a compressão de dados. Possuem também uma forte inspiração neurofisiológica baseado no mapa topológico presente no córtex cerebral exposto na figura 3.6. O cérebro dos animais mais sofisticados possui áreas que são responsáveis por funções específicas, por exemplo: áreas responsáveis pela fala, visão, controle motor etc. Cada uma dessas áreas, possui subáreas que mapeiam internamente respostas do órgão sensorial representado por ela. Como forma representativa, podemos citar: córtex auditivo: o mapeamento reflete as diferenças frequências sonoras; córtex visual: o mapeamento é definido pelas características visuais primitivas, como intensidade de luz, orientação e curvatura de linhas [11], [29], [30], e [52].

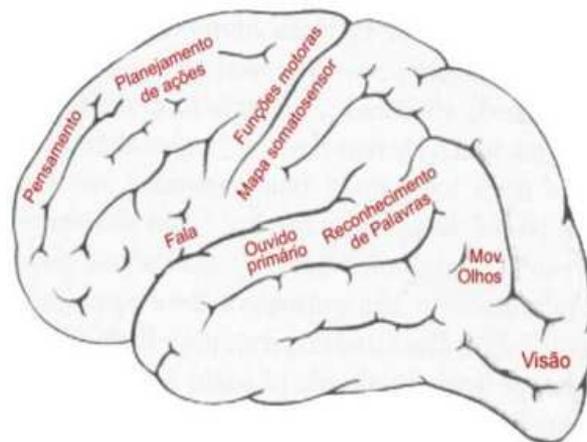


Figura 3.6 Mapa citoarquitetural do córtex cerebral simplificado [52]

A idéia de auto-organização, são motivadas por considerações neurobiológicas para derivar o modelo, que é uma abordagem tradicional. Alternativamente, podemos usar uma abordagem de quantização vetorial que usa um modelo envolvendo um codificador e um decodificador, que é motivada por considerações da teoria de comunicações [30]. O principal objetivo da rede neural SOM é transformar um padrão de sinal incidente de dimensão arbitrária em um mapa discreto uni ou bidimensional e realizar esta transformação adaptativamente de uma maneira topologicamente ordenada.

A figura 3.7, mostra as principais características da arquitetura convencional da rede SOM a partir de uma grade bidimensional de neurônios integrada com um conjunto de vetores de entrada. Assim, quando um padrão de entrada é apresentado à rede, a mesma procura a unidade mais parecida com e durante o seu processo de treinamento, a rede aumenta a semelhança do nodo escolhido e de seus vizinhos ao padrão, construindo um mapa topológico onde os nodos que estão topologicamente próximo respondem de forma semelhante aos padrões de entrada. A rede SOM utiliza um algoritmo de aprendizado competitivo, em que os nodos da camada de

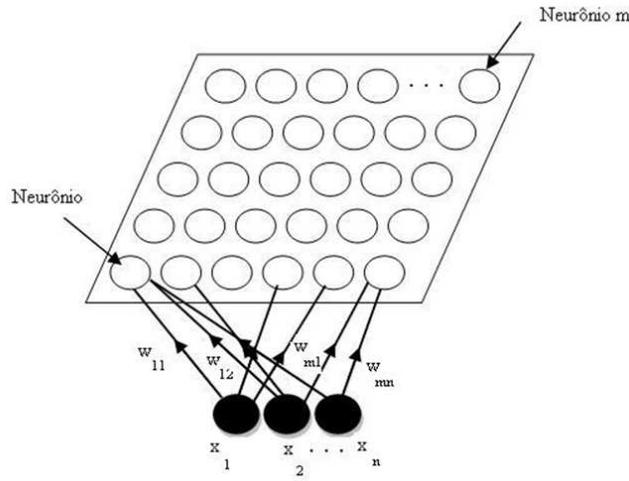


Figura 3.7 Arquitetura do mapa auto organizável [29]

saída competem entre si para se tornarem ativos, ou seja, para ver quem gera a maior saída. Para cada padrão de entrada, apenas um nodo de saída ou nodo por grupo se torna ativo, esta competição é chamada "o vencedor leva tudo"[11] e [52]. Uma maneira de implementar esta competição é a utilização de conexões laterais inibitórias entre nodos de saída. A função de chapéu de mexicano representa bem este processo de acordo com o que segue:

As conexões laterais da rede podem ser representadas a partir da função de chapéu de mexicano, onde os neurônios que estão espacialmente ordenados dentro das áreas de raio (R_1, R_2 e R_3) tendem a responder a padrões de estímulos semelhantes a partir do resultado obtido através da competição. Assim, quando um nodo vence uma competição e produz a maior saída para uma dada entrada, não apenas ele mas também os nodos localizados na sua vizinhança topológica tem seus pesos ajustados. Essa ordenação topológica é resultado do uso do *feedback* lateral entre as células do córtex cerebral. Na função de chapéu de mexicano cada neurônio influencia no estado de ativação de seus neurônios vizinhos de três formas possíveis: excitatória (se os vizinhos de neurônios estão próximos ao neurônio vencedor e dentro de uma área de raio R_1); inibitória (se os vizinhos estão fora da área anterior R_1 , mas dentro de uma segunda área R_2 , onde $(R_2 > R_1)$); levemente excitatória (se os vizinhos estão fora das áreas anteriores mas dentro de uma terceira área R_3 , mas fora das áreas de raios R_1 e R_2 , onde $(R_3 > R_2 > R_1)$) [11].

Desta forma, quando um nodo vence uma competição produzindo a maior saída para uma dada entrada, não apenas ele mas também os nodos localizados na sua vizinhança têm seus pesos ajustados. Com isso, os mapas de Kohonen consegue um resultado semelhante aquele obtido com a utilização da função chapéu de mexicano, proporcionando menor custo computacional.

As entradas da rede exposta na figura 3.7 podem ser representadas na forma de vetor, onde:

$$X = [x_1, x_2, x_3, \dots, x_n]^T \quad (3.1)$$

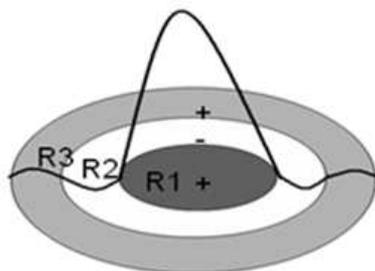


Figura 3.8 Tipos de estímulos produzidos de acordo com a função chapéu de mexicano [11]

Os vetores de pesos dos neurônios, podem ser representados pelo plano bidimensional (2-D).

$$w_i = [w_{i1}, w_{i2}, w_{i3}, \dots, w_{in}]^T, i = 1, 2, 3, \dots, m \quad (3.2)$$

Onde l é o total de neurônios na saída no plano bidimensional (2-D), a melhor saída do vetor de entrada com os vetores de pesos w_i e pode ser determinado por:

$$i(x) = \operatorname{argmin}_j \|x - w_j\|, j = 1, 2, \dots, l. \quad (3.3)$$

Assim, $i(x)$ pode ser representado pelo índice para o neurônio de saída (neurônio vencedor) conhecido como neurônio *BMU* (*Best Matching Unit*) e $\|\bullet\|$ é a norma Euclidiana. Portanto, pode-se usar a equação 3.3, para mapear um espaço contínuo de entrada sobre um arranjo discreto de neurônios. A resposta provém da atualização do vetor de pesos associado ao neurônio vencedor segundo a sua vizinhança topológica. O neurônio BMU é adaptado, seu vetor de pesos sinápticos é alterado no sentido de se aproximar ainda mais do dado apresentado, aumentando a probabilidade de que este mesmo neurônio volte a vencer numa subsequente apresentação do mesmo dado (6), (14).

3.1.5 Principais processos do algoritmo de Kohonen

No mapa auto-organizável de Kohonen, os neurônios são dispostos em um arranjo discreto finito, geralmente unidimensional ou bidimensional e totalmente conectados com a entrada, possibilitando mapear um conjunto de dados. O algoritmo responsável pela formação do mapa auto organizável começa primeiramente inicializando os pesos sinápticos na rede. Isto pode ser feito atribuindo-lhes valores pequenos tomados a partir de um gerador de números aleatórios; fazendo desta forma, nenhuma organização prévia a qual é imposta ao mapa características [29] e [52]. De acordo com Haykin [30], Ham [29], Kohonen [52] e Zuchini [100], a formação de um mapa auto-organizável pode ser dividida em três principais processos:

-Competição. Para cada padrão de entrada, os neurônios da grade calculam seus valores em relação a ele e competem entre si pelo direito de representá-lo. O neurônio que melhor representa

os dados é aquele que utiliza alguma métrica pré-estabelecida, o qual é considerado o vencedor.

- **Cooperação.** O neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo assim a base para a cooperação entre os neurônios vizinhos.

- **Adaptação Sináptica.** Este último processo permite que os neurônios excitados (incluindo os vencedores) aumentem seus valores individuais em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos. Os ajustes são feitos de tal forma que a resposta do neurônio vencedor à aplicação subsequente de um padrão de entrada similar é melhorada.

3.1.5.1 O Processo Competitivo

Dado um conjunto de entradas X (vide equação 3.1), onde n representa a dimensão do espaço de entrada. O vetor de pesos sinápticos de cada neurônio tem a mesma dimensão do espaço de entrada. O conjunto de vetores de pesos sinápticos é denotado por (vide equação 3.2), onde l é o número total de neurônios no arranjo.

3.1.5.2 Maximização do Produto Interno

O critério do melhor casamento (*best matching*), baseia-se na maximização do produto interno como na aprendizagem competitiva, ao determinarmos a localização onde a vizinhança topológica dos neurônios excitados devem ser centrada. Assim, quanto mais semelhante a entrada for do vetor de pesos de um nodo, maior o valor de sua saída. Segundo Kohonen [52] e Haykin [30], isto é matematicamente equivalente a minimizar a entre o vetor de peso w e o vetor de entrada utilizando alguma métrica específica, onde o neurônio vencedor $i(x)$ é escolhido pela equação 3.3. A métrica utilizada para os testes é a distância euclidiana, por ser um procedimento comum quando nenhuma outra informação prévia existe acerca dos dados de entrada. A norma euclidiana tem a propriedade de ser invariante as rotações aplicadas aos dados de entrada, o que pode ser útil no caso de reconhecimento de padrões que sofrem algum tipo de transformação linear. Desta forma, o neurônio vencedor é aquele que possui a menor distância com o dado x (aquele que melhor representa os dados de entrada), assim, o neurônio denotado por $i(x)$ pode ser determinado pela equação 3.3.

3.1.5.3 O Processo Cooperativo

O melhor casamento pode ser baseado na maximização do produto interno de $w_j^T x$. Isto é equivalente a minimizar a distância euclidiana entre os vetores x e w [29], [30] e [52]. Devido ao motivo que o neurônio vencedor é aquele que possui a menor distância como o dado x , este é o que melhor representa os dados. O neurônio vencedor viabiliza o requisito de que os neurônios próximos a ele também tenham seus vetores de pesos sinápticos ajustados na direção dos dados e que, deverão excitar os neurônios pertencentes a sua vizinhança (definindo graus de vizinhança e excitação diferentes, de acordo com alguma função pré-estabelecida). Em particular, um neurônio que está disparando tende a excitar mais fortemente os neurônios na

sua vizinhança imediata entre aqueles distantes de acordo com a especificação da função de chapéu de mexicano (figura 3.8), o que é intuitivamente razoável. Esta observação nos leva a fazer com que a vizinhança topológica em torno do neurônio vencedor decaia suavemente com a distância lateral [29], [30] e [52] e [100].

3.1.5.4 Função de Vizinhança

Seja $h_{j,i}$ a função de vizinhança topológica definida para o SOM, centrada no neurônio vencedor i e um conjunto de neurônios excitados por ele. Onde $d_{j,i}$ é a distância lateral entre o neurônio vencedor i e o neurônio excitado j . Então a função de vizinhança $h_{j,i}$ é uma função da distância unimodal de $d_{j,i}$, desde satisfaça as seguintes condições:

A função de vizinhança é simétrica em relação ao seu ponto de máximo (definido por $d_{j,i} = 0$). Ou seja, o seu valor máximo está no neurônio vencedor i , para qual a distância de $d_{j,i}$ é zero. A amplitude da função de vizinhança $h_{j,i}$ decresce monotonicamente com o aumento da distância lateral $d_{j,i}$, decaindo a zero quando $d_{j,i} \rightarrow \infty$, esta condição é necessária para a convergência. Uma escolha típica de função de vizinhança $h_{j,i}$, satisfaz esses dois requisitos é uma função gaussiana [29], [30] e [52].

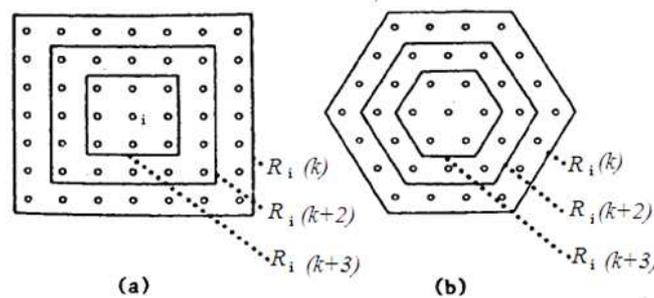


Figura 3.9 Exemplos de função de vizinhança de acordo com o critério de monotonicidade: a) base quadrada e b) base triangular [52] e [29]

Na figura 3.9, o neurônio vencedor poderá disparar estímulos capazes de excitar ou inibir os neurônios vizinhos em função do raio de vizinhança topológica $R_i = R_i(k), R_i = R_i(k+2)$ e $R_i(k+3)$. Desta forma, quando ocorre o processo de decrescimento monotônico na função $h_{j,i} = h_{j,i}(t)$, faz-se necessário especificar os parâmetros da taxa de aprendizagem a partir de uma constante $\alpha(t)$, onde $h_{j,i}(t) = \alpha(t)$, obedece as seguintes regras: se $j \in R_i(k)$ e $h_{j,i}(t) = 0$, se $j \notin R_i(k)$ quando $(0 < \alpha(t) < 1)$. De acordo com Kohonen [52], $\alpha(t)$ e $R_i(k)$ decresce monotonicamente em função do tempo durante o processo de ordenação. A função gaussiana (figura 3.10) é invariante a translação independente da localização do neurônio vencedor [30]. Maiores detalhes são apresentados no Apêndice B. A equação 3.4, mede o grau com que os neurônios excitados na vizinhança do neurônio vencedor participam do processo de aprendizagem e aborda em um sentido qualitativo.

$$h_{j,i} = \exp\left(\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (3.4)$$

A vizinhança topológica gaussiana é mais biologicamente apropriada que a vizinhança retangular, permite que o algoritmo SOM convirja mais rapidamente do que uma vizinhança retangular [30] e [52]. Para que a cooperação entre os neurônios vizinhos se mantenha, é necessário que a vizinhança topológica $h_{j,i}$ seja dependente da distância lateral $d_{j,i}$ entre o neurônio vencedor i e o neurônio excitado j , no espaço de saída, em vez de ser dependente de alguma medida de distância no espaço de entrada original. Assim, a distância entre os neurônios i e j que está na dimensão do arranjo é definida por :

$$d_{j,i} = ||r_j - r_i|| \quad (3.5)$$

Onde r_j é um vetor discreto que define a posição no arranjo do neurônio excitado e r_i define a posição do neurônio vencedor. Ambos os vetores estão contidos no espaço discreto de saída. Outra característica do algoritmo SOM está relacionado na escolha da largura da vizinhança ao longo do treinamento, com o objetivo de alcançar a convergência do mapa. Esse requisito é satisfeito fazendo o tamanho de σ diminuir com o tempo. Uma função bastante utilizada para σ variar de acordo com o tempo discreto n é a função de decaimento exponencial [29] e [30].

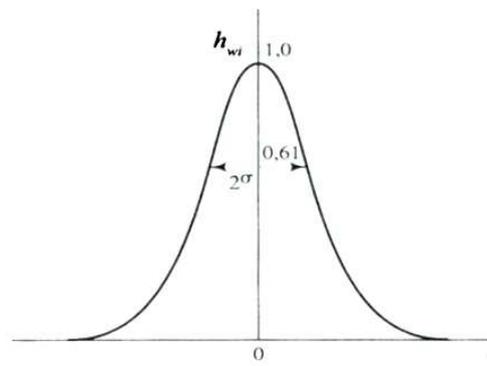


Figura 3.10 Função de vizinhança gaussiana [30]

O tamanho da vizinhança estabelecida pelo algoritmo do SOM poderá diminuir com o tempo. Na medida em que esta exigência é satisfeita faz com que a largura σ da função de vizinhança topológica $h_{j,i}$, diminua. Uma escolha popular para a dependência de σ com o tempo discreto n é o decaimento exponencial, representada por:

$$\sigma(n) = \sigma_0 \exp\left(\frac{-n}{\tau_1}\right), n = 0, 1, 2, \dots, l \quad (3.6)$$

Onde: σ_0 é o valor de σ na inicialização do algoritmo SOM, e τ_1 é uma constante de tempo. Conseqüentemente, a vizinhança topológica assume uma forma variável no tempo, como mostrado por:

$$h_{j,i(x)}(n) = \exp - \left(\frac{d_{j,i}^2}{2\sigma^2(n)} \right), n = 0, 1, 2, \dots \quad (3.7)$$

Onde $\sigma(n)$ é definido pela equação (3.12). Assim, quando o tempo n (número de iterações) aumenta, a largura $\sigma(n)$ decresce a uma taxa exponencial e a vizinhança topológica diminui de uma maneira correspondente [30] e [52].

3.1.6 O Processo Adaptativo

No processo cooperativo, cada vez que um novo padrão de treinamento é apresentado à rede, os nodos competem entre si para ver quem gera a maior saída. Definido o nodo vencedor inicia-se o processo de atualização dos pesos conhecido, como processo adaptativo. Apenas o nodo vencedor e seus vizinhos dentro de um certo raio ou área de vizinhança atualizam seus pesos. Durante o treinamento, a taxa de aprendizado e o raio de vizinhança são continuamente decrementados. Assim, no processo adaptativo, os neurônios excitados têm seus vetores de pesos modificados em direção ao vetor de entrada x . Para tanto, é necessário a existência de alguma regra para que se garanta a convergência do mapa de forma auto-organizada. Segundo Haykin [30], Kohonen [52], Neto [66] e Zuchini [100], o novo valor do vetor de pesos sinápticos do j -ésimo neurônio no instante de tempo $(k+1)$ pode ser definido pela equação (3.8).

$$w_j(k+1) = w_j(k) + \eta(k)h_{j,i(x)} [x - w_j(k)], j = 1, 2, \dots, l. \quad (3.8)$$

Onde $\eta(k)$ define a taxa de aprendizado, $h_{j,i(x)}$ define o grau de adaptação do neurônio em relação ao vencedor. O parâmetro de aprendizado $\eta(k)$, a fim de que haja a convergência do SOM, deve variar gradualmente com o tempo [30]. Normalmente, $\eta(k) \rightarrow 0$ quando $k \rightarrow \infty$ [100]. Esta exigência pode ser satisfeita escolhendo-se um decaimento exponencial para $\eta(k)$, como mostrado na equação (3.9). Onde τ_1 é uma outra constante de tempo do algoritmo SOM.

$$\eta(n) = \eta_0 \exp - \left(\frac{n}{\tau_1} \right), n = 0, 1, 2, \dots, l \quad (3.9)$$

3.1.7 Normalização dos Dados

As variáveis de entrada utilizadas na rede SOM possuem intervalos de variação diferentes. Segundo Kohonen [52] e Zuchini [100], quando pouco ou nada se sabe sobre a importância de cada um dos atributos na expressão da informação contida em uma base de dados, devem-se buscar mecanismos para evitar que um atributo domine o processo de agrupamento. A contribuição de outros que possam ser mais importantes e também que atributos de grande relevância tenham seu papel minimizado. Portanto, todas as variáveis devem ser normalizadas na utilização da rede neural SOM, para assegurar que elas recebam igual atenção durante o processo de treinamento. Além disso, as variáveis têm que ser normalizadas de tal maneira que seus valores sejam proporcionais aos limites das funções de ativação usadas na camada de saída. A normalização dos dados é também importante para a eficiência do algoritmo de treinamento [90].

3.1.8 O valor dos pesos no Melhoramento da Rede

Os pesos iniciais de uma rede do tipo SOM são definidos aleatoriamente, muitos nodos podem apresentar vetores de pesos muito diferentes dos padrões de entrada e poderá não haver o número necessário de nodos utilizáveis (que possam vencer as competições) para definir os *clusters* adequadamente [11], [29], [30] e [52]. Em resultado, a rede pode ou não convergir ou até mesmo apresentar ciclos muito lentos. Diversas alternativas foram propostas na literatura para resolver ou minimizar este problema, tais como: utilização de vetores de pesos iniciais iguais, utilização de um limiar para cada nó e modificação dinâmica de raio de vizinhança. De acordo com Braga [11], a primeira alternativa seria inicializar todos os pesos com o mesmo valor para garantir a atualização do maior número possível de nodos e os padrões de treinamento seriam, no início, tornados semelhantes entre si. A semelhança seria gradativamente reduzida nos primeiros ciclos de treinamento por meio da adição de ruídos aos vetores de entrada e a segunda opção propõe a utilização de um limiar para cada nó. Este limiar seria a consciência do nodo, por isso esta técnica também é chamada de consciência. Os nodos que fossem regularmente selecionados nas competições teriam seu limiar aumentado. Assim, este mecanismo reduz a chance de este nodo ser selecionado novamente, permitindo a utilização de outros nodos. Em Kohonen [52], a redução da vizinhança dos nodos vencedores durante o treinamento em torno do vencedor, faz com que os pesos sejam ajustados, ou seja, define a área de influência do nó vencedor. Quando a solução de reduzir a vizinhança é empregada, utiliza-se inicialmente uma vizinhança grande. Durante o treinamento, a região de vizinhança é progressivamente reduzida até um limite pré-definido, geralmente, a taxa de redução é uma função linear do número de ciclos.

3.1.9 Resumo do Algoritmo

O algoritmo de treinamento para redes SOM poderá ser resumido pelo pseudocódigo a seguir:
SOM

1. inicialize a rede SOM com os vetores de peso $(*)w_i$, e os parâmetros da rede definindo : taxa de aprendizagem, função de vizinhança; inicialize os parâmetros, definido por $k = 0$.
2. verificar a condição $**$). Se for falso, continue; Se for verdadeiro, saia.
3. para cada padrão de treinamento do vetor x faça os passos 4 e 7.
4. computar o melhor casamento com o vetor de pesos em relação ao vetor de entrada

$$i(x) = \arg \min_j \|x - w_j\|^2$$
5. para todas as unidades especificadas pela vizinhança $i \in h_{j,i}(k)$ (i corresponde ao neurônio vencedor), atualize os vetores de peso de acordo com:

$$w_j(k+1) = w_j(k) + \eta(k)h_{j,i}(k) [x - w_j(k)]$$
 Onde : $0 < \alpha(k) < 1$ (parâmetros da taxa de aprendizagem).
6. ajuste a taxa de aprendizagem
7. Em seguida reduza a vizinhança topológica $h_{j,i}(k)$.
8. Ajustar $K \leftarrow K + 1$ e siga para o passo 2.

(*) a inicialização dos pesos poderá ser feita de forma aleatória, ou a partir do conjunto de ponderações que representam o conhecimento sobre os dados de entrada, informações pertinente a distribuição das classes em função da saída.

(**) a condição de parada pode ser declarada com base no número de iterações ou até que o mapa de características não mude.

3.1.10 Treinamento e Convergência

A convergência da rede SOM poderá ser classificada em duas etapas: a primeira no que se refere a fase de ordenação, a qual ocorre a ordenação topológica dos vetores de pesos w_i (com pesos iniciais aleatórios) e poderá durar em torno de 1000 ciclos ou iterações.

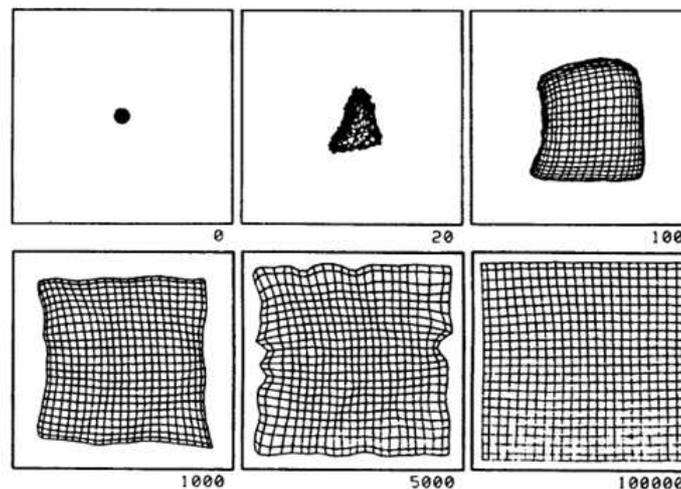


Figura 3.11 Ordenação topológica dos vetores de peso [52]

Na fase de treinamento, o algoritmo busca agrupar os nodos do mapa topológico em *clusters* ou agrupamentos de modo a refletir a distribuição dos padrões de entrada. Assim, a rede descobre quantos *clusters* deve identificar e suas posições relativas no mapa. O mapeamento realizado neste estágio é um mapeamento grosseiro dos padrões de entrada. E, geralmente a taxa de aprendizado é inicialmente alta, próxima de 1 e posteriormente de forma gradual poderá ser reduzida até um valor próximo de 0,1. Nesta fase ocorrem grandes mudanças nos pesos e conseqüentemente a vizinhança é reduzida de modo generalizado envolvendo todos os nodos da rede de forma linear até atingir um raio de um ou dois vizinhos.

A segunda fase o algoritmo (convergência), faz um ajuste mais fino do mapa e requer de 100 a 1000 ciclos que na fase anterior, geralmente utiliza uma taxa de aprendizado baixa, de ordem de 0,01 ou menos e o raio da vizinhança envolve um ou nenhum vizinho. Esta fase sofisticada o mapeamento realizado no estágio anterior, aprimorando o agrupamento realizado [52] e [30]. O treinamento da rede SOM é afetado pela taxa de aprendizado, pela taxa de redução da região de vizinhança e pelo formato da região de vizinhança [11]. Após a fase de treinamento a rede SOM agrupa os padrões de entrada em *clusters* ou agrupamentos.

Em algumas aplicações pode ser necessário rotular os nodos de saída para indicar os *clusters* que representam para uma melhor interpretação dos dados. Esta rotulação irá permitir posteriormente a classificação de padrões desconhecidos. De acordo com Kohonen [52], a rede neural SOM não foi criada para reconhecimento de padrões, mas para agrupamento, visualização e abstração. Ainda assim, as redes SOM podem ser utilizadas para o reconhecimento ou a classificação de padrões, desde que seja utilizada junto com um modelo de aprendizado supervisionado, podendo ser utilizada na saída, uma rede *Adaline*, *Multi Layer Perceptron (MLP)* ou outras técnicas que possibilitem a descoberta desses padrões no caso da lógica nebulosa como uma abordagem híbrida. Também pode ser visto na literatura uma implementação do algoritmo SOM com a técnica de quantização vetorial chamada *Learning Vector Quantization (LVQ)*. De acordo com Braga [11] e Haykin [30], uma vez que a rede SOM tenha sido treinada e rotulada, pode ser necessário ensinar novos padrões. Esta inclusão pode buscar a melhoria do desempenho da rede para certos *clusters*. Um algoritmo LVQ, permite a inclusão de novos padrões em uma rede SOM já treinada é considerada uma técnica de aprendizado supervisionado que utiliza informações sobre a classe para mover levemente os vetores de peso, de modo a melhorar a qualidade das regiões de decisão do classificador, procurando ajustar o mapa de características com objetivo de melhorar o desempenho da rede em circunstâncias modificáveis [30, 52].

3.1.11 Interpretação do Mapa produzido pelo SOM

A utilização de arranjos em 2 dimensões é o modelo mais difundido no processo de mineração de dados. Em Zuchini [100], exibe a possibilidade de construção de uma, duas ou mais dimensões, conforme as necessidades e objetivos. Contudo, para que se consiga interpretar o conteúdo do mapa é necessária a utilização de algum método de visualização que auxilie tal tarefa. Desta forma, muitos métodos são utilizados, entre eles, existe a visualização do mapa como uma grade elástica, onde os neurônios vizinhos estão posicionados mais longe ou mais perto de acordo com a distância de seus vetores de pesos sinápticos [30]. Entre outros métodos, existe a Matriz-U (matriz composta pelas distâncias entre todos os neurônios vizinhos no arranjo) e o mapa contextual, ambos utilizam arranjos bidimensionais.

3.1.12 Matriz-U

A matriz de distâncias unificada (Matriz - U), proposta por Ultsch [88], considerando um arranjo retangular plano de tamanho $L \times C$. O valor da matriz-U sobre os neurônios em si é normalmente obtido pela média aritmética das distâncias entre os vetores de pesos de toda a vizinhança do neurônio e seu próprio vetor de pesos, pode-se avaliar visualmente a existência de "vales" que sugerem onde os vetores de pesos dos neurônios são mais próximos de si, separados por "elevações" (onde os vetores de pesos dos neurônios encontram-se mais distantes). Um vale é associado com a ocorrência de um agrupamento e quanto mais alto uma elevação separando dois vales, tanto mais distintos são estes agrupamentos no espaço de dados [100].

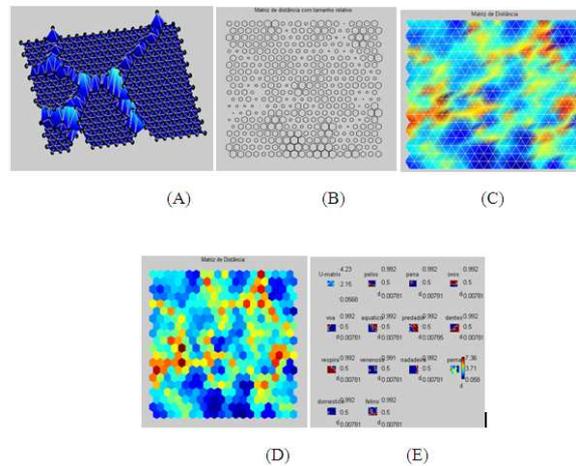


Figura 3.12 Exemplo de Matriz- U : (A) por quantidade de neurônios, (B) modelo por subdivisão de neurônios, (C) matriz de distância, (D) matriz com grade hexagonal e (E) matriz com variáveis no plano [92]

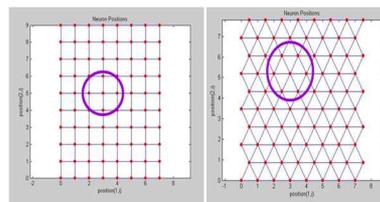


Figura 3.13 Exemplo de um mapa topológico com grades hexagonal e retangular [18]

3.1.13 Mapas Contextuais

A visualização dos mapas auto organizáveis pode ser feita através de mapas contextuais, esses mapas consiste em um mapa onde cada neurônio é "rotulado" com o valor do padrão de teste que ele melhor representa (os padrões de teste são padrões pertencentes ao espaço de entrada, mas não necessariamente ao conjunto de treinamento, que servirão para alimentar o mapa contextual). O rótulo é algum atributo de um elemento pertencente ao espaço de entrada que o distingue, mas não contém informações ou similaridades entre os demais elementos. O mapa contextual tem sua utilização em diversas aplicações e entre elas está relacionada a mineração de dados. O resultado do algoritmo para a formação de um mapa contextual é a geração de um mapa onde os neurônios são rotulados de tal forma que o arranjo seja particionado em regiões coerentes, onde cada grupo de neurônios representa um conjunto distinto de rótulos [30] e [52].

3.1.14 Análise do Mapa

Autores como Kaski [48], Kohonen [52] e Zuchini [100], sugere como critério de configuração básica que vizinhança maior de atualização de pesos representadas pelas épocas (vizinhança regredindo até 1), proporciona um mapeamento mais grosseiro do mapa. Para uma projeção mais lenta e melhorar a qualidade do mapa, as épocas sejam fixadas com vizinhança

$$QE = \frac{1}{N} \sum_n = 1^N \|x - w_j\| \quad (3.11)$$

A segunda medida refere-se ao erro topográfico (ET), o qual quantifica a capacidade do mapa em representar a topologia dos dados de entrada. Cada objeto são calculados seu BMU w_j e o segundo BMU w_z e o erro topográfico é dado por Kiviluoto Guerra [28] e [51]:

$$ET = \frac{1}{N} \sum_n = 1^N u(x) \quad (3.12)$$

Onde $u(x) = 1$ caso, w_j e w_z não seja adjacentes [51].

Abordagens como Zuchini [100], apontam a efetividade de se usar EQ e ET como fontes de medidas para analisar a resolução do mapa, ou seja, quando EQ diminui com o aumento de neurônios no arranjo (resolução aumenta). Se o arranjo possuir um número muito grande de neurônios (maior a quantidade de objetos a representar) poderá sofrer um processo de treinamento onde o raio de vizinhança torna-se menor ou igual a 1 durante muito tempo, pode ocorrer de os neurônios posicionarem-se praticamente sobre os objetos a serem representados, assim, quando $EQ \rightarrow 0$, o arranjo pode estar tão retorcido que a capacidade de representar a topologia dos dados é perdida quando ET aumenta, o comportamento de TE dependerá do número de neurônios disponíveis no arranjo: TE aumenta se há poucos neurônios e diminui se há muitos neurônios no arranjo.

Neste contexto, as medidas de EQ e ET deverão ser analisadas de acordo com o tipo de sinal, uma vez que, quando a base de dados possui valores muito próximos de zero, mesmo que exista um balanceamento do conjunto de neurônios da grade, quando $EQ \rightarrow 0$ e representar adequadamente o conjunto de dados. Autores como Kaski [48], Guerra [28], Zuchini [100], utilizam bases de dados cujos sinais possuem valores superiores ou próximos a 1 fator significativo para o balanceamento de EQ.

3.1.15 Considerações Finais

Devido ausência de fundamentação teórica e matemática do algoritmo SOM principalmente no que se referem as suas propriedades, não existem critérios mensuráveis para os valores dos parâmetros a fim de se obter mapas bem ajustados. Contudo quando o objetivo do SOM está voltado aos princípios de mineração de dados [52] e [100], propõe alguns parâmetros para realização dos testes, porém recomenda realizar diversos testes para verificar o comportamento da rede. Outras limitações foram encontradas no algoritmo de Kohonen no que se refere à rotulação dos dados e no custo computacional, ou seja, uma rotulação superior a 4(quatro) caracteres, poderá influenciar no processo de análise do mapa contextual quando a massa de dados é elevada. Tampouco, recomenda-se rotular os dados com o menor número de caracteres, isto possibilita a melhor visualização e análise dos dados evitando a utilização de neurônios excessivos, o qual poderá influenciar no processamento durante o treinamento. A interpretação do mapa de Kohonen na maioria dos casos é realizada pela matriz-U, considerada subjetiva e por sua vez de bastante difícil de sua interpretação quando a massa de dados é elevada.

Lógica Nebulosa

Este capítulo descreve alguns conceitos necessários para o entendimento da teoria de conjuntos nebulosos e tem como objetivo determinar os princípios de modelagem e mapeamento entre as entradas e saídas do sistema nebuloso que será usado na aplicação dos estudos de casos.

4.1 Introdução

Os modelos matemáticos baseados em lógica *fuzzy* ou nebulosa criada por Zadeh [99], na década de 1960, combinando a lógica clássica com os conceitos de conjuntos com grau de pertinência não binários estabelecidos por Jan Lukasiewicz em 1920 [21], [23] e [63], para representar o conhecimento incerto, impreciso e contraditório, oferecendo um método eficaz para descrever o comportamento inteligente em sistemas complexos que possa apoiar o processo de tomada de decisão. Em Maia [57], define-se que sistemas *fuzzy* como sistemas de inferência nebulosa baseados em regras, ou modelos nebulosos, representam a mais importante ferramenta baseada na teoria dos conjuntos nebulosos. Em Shaw [60], a lógica *fuzzy* pode ser definida como uma técnica que incorpora a forma humana de pensar em um sistema de controle.

A lógica nebulosa têm sido aplicada com sucesso nas mais diversas áreas de conhecimento tais como: controle automático, classificação, reconhecimento de padrões, tomada de decisão, sistemas inteligentes, previsão de séries temporais entre outras [21], [23],[57], [60] e [99]. Apresentando destaque na aproximação de funções buscando aproximar com precisão qualquer mapeamento não-linear contínuo sobre uma região compacta de um domínio [64].

A figura 4.1, apresenta o diagrama de blocos utilizado pelo sistema de controle, a estrutura básica de um sistema nebuloso pode ser dividido por quatro componentes básicos: base de regras (contêm um conjunto de regras nebulosas), fuzzificador (efetua o mapeamento dos dados numéricos *crisp* (não *fuzzy*) de entrada em conjuntos nebulosos), mecanismo de inferência ou raciocínio (processa o raciocínio nebuloso para obter a saída ou conclusão baseados em fatos e na combinação das regras e no mecanismo de fuzzificação) e defuzzificador (converte os conjuntos nebulosos de saída em números *crisp*) [99, 64, 82].

Outra funcionalidade importante da lógica nebulosa está relacionada ao controlador nebuloso, o qual tenta "imitar" o operador humano a partir de uma representação descritiva e experimental do processo, buscando produzir as ações de controle desejadas baseando-se em modelos da operação do sistema, ao invés de utilizar modelos do próprio processo como em outras técnicas clássica de controle [13].

A modelagem clássica de um sistema de controle poderá ser definida a partir da escolha

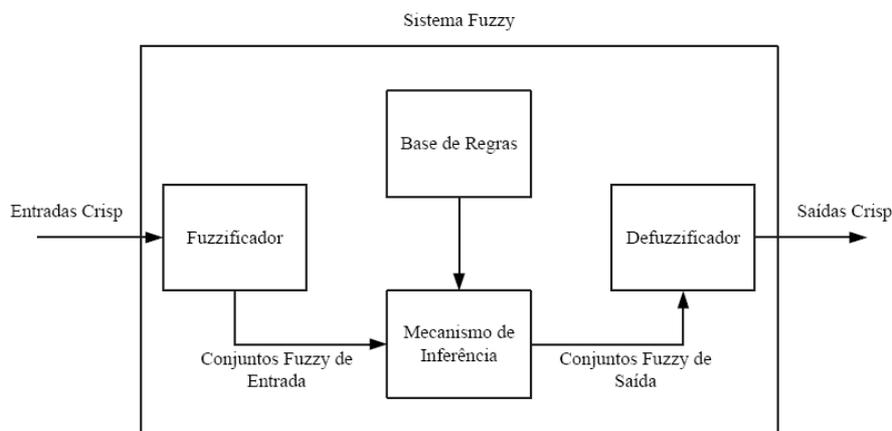


Figura 4.1 Diagrama de blocos de um sistema fuzzy [23] e [82]

das variáveis de entrada e saída do sistema [21] e [99]. As variáveis de entrada são aquelas em que o usuário baseia-se para fazer uma análise e desempenho do processo utilizado para tomar decisões, na maioria dos casos poderá ser escolhida de forma intuitiva por um especialista. As variáveis de saída são aquelas controladas pelo processo, são mais fáceis de identificá-las devido ao relacionamento com os objetivos da modelagem do controlador. Após a definição de todas as variáveis tanto de entrada quanto as de saída para o controlador nebuloso. Neste caso, deve-se especificar a base de conhecimento, o qual formará o núcleo do sistema de controle, se a informação disponível sobre o processo for inadequada, poderá ser pequena a chance de se projetar uma boa base de regras e, em consequência, assegurar um bom desempenho do sistema de controle [36]. Desta forma, o processo de modelagem abrangendo a escolha das variáveis de entrada e saída do sistema é considerado parte fundamental importância na especificação e desenvolvimento do controlador nebuloso.

4.1.1 Conjuntos

Considere φ um elemento do conjunto Γ , onde $\varphi \in \Gamma$ ou $\varphi \notin \Gamma$. Onde $\Gamma = \{-1, 0, 1\}$ ou representado pelo conjunto de inteiros positivos $\Gamma = \{1, 2, 3, \dots, n, \dots\}$ e poderá ser definido por um conjunto de elementos de uma certa propriedade, por exemplo, o conjunto A é composto por todos os elementos φ em Γ que têm a propriedade de \wp :

$$A = \{\varphi \in \Gamma : \wp(\varphi)\} \quad (4.1)$$

Representado pela forma simplificada quando o conjunto Γ está implícito:

$$A = \{\varphi : \wp(\varphi)\} \quad (4.2)$$

Quando Γ não possui essa propriedade pode ser representado por ϕ . Se cada φ é um elemento do conjunto. A é um conjunto em B , tal que, $\varphi \in A \implies \varphi \in B$. Logo A é um subconjunto de B (ou $A \subset B$). Se $A \subset B$ e $B \subset A$, então $B = A$.

Além disso, para qualquer conjunto A que contenha elementos do conjunto ϕ (vazio), os elementos de A podem ser representados por $\phi \subset A$. Por isso, cada conjunto ϕ é um subconjunto de cada conjunto. O conjunto C possui elementos que pertence ao conjunto A , ou seja, um conjunto B pode ser representado pela união de ambos conjuntos e pode ser representados por:

$$A \cup B = C = \{\varphi : \varphi \in A \vee \varphi \in B\} \quad (4.3)$$

Se A e B são subconjuntos de C pode ser definido pela interseção de todos os elementos do conjunto que pertecem a A e B :

$$A \cap B = C = \{\varphi : \varphi \in A \wedge \varphi \in B\} \quad (4.4)$$

Se A for um subconjunto de B , pode ser representado pelo complemento de A , proveniente de uma relação \tilde{A} em B e pode ser definido por um conjunto de elementos que não pertence A :

$$\tilde{A} = \{\varphi \in B : \varphi \notin A\} \quad (4.5)$$

Onde : $\tilde{\tilde{A}} = A, A \cup \tilde{A} = B$ e $A \cap \tilde{A} = \phi$. A lei de *Morgan* [99, 29] pode ser representada por:

$$A \cup \tilde{B} = \tilde{A} \cap \tilde{\tilde{B}} \text{ e } A \cap \tilde{B} = \tilde{A} \cup \tilde{\tilde{B}} \quad (4.6)$$

4.1.2 Funções

Uma função está essencialmente restrita a uma regra [29]. Para uma função f de um conjunto X , o conjunto Y atribuído como regra, onde para cada $x \in X$ a um único elemento $f(x) \in Y$. A coleção de pares ordenados $\langle x, f(x) \rangle$ representa o produto cartesiano de $X \times Y$, chamado como gráfico da função f . Logo f é uma função de X em Y , a qual poderá ser representada por:

$$f : X \rightarrow Y \quad (4.7)$$

O conjunto X é considerado o domínio de f . Caso $A \subset X$, a imagem da função f de A pode ser definido pelo conjunto de elementos de Y , onde $y = f(x)$ para $x \in A$. O conjunto imagem pode ser representado por $f[A]$. Note que f é representada por $f[X]$, e f e $Y = f[X]$. Desta forma, $B \subset Y$ define a inversa do conjunto imagem de $f^{-1}[B]$ de B , para ser um conjunto $x \in X$ aqueles para os quais $f(x)$ está em B , maiores detalhes poderão ser visto em Ham [29].

4.1.3 Conjuntos Nebulosos

Um conjunto clássico A pode ser definido por um conjunto de elementos ou objetos $x \in X$, os quais poderão pertencer ou não ao conjunto A , se $A \subset X$, a partir da definição de uma função característica definida por cada elemento de x em A , o conjunto clássico pode ser formado pelo conjunto de pares ordenados $(x, 0)$ ou $(x, 1)$. Onde 1 pertence totalmente ao conjunto

A e $0 \notin A$. A função característica (função indicadora) é também conhecida como função de pertinência.

$$\mu_A(x) = 1 \begin{cases} \text{se } x \in A \\ \text{se } x \notin A \end{cases} \quad (4.8)$$

A modelagem matemática de subconjunto de um conjunto A é considerado subconjunto *fuzzy* de X , se for descrito como um conjunto de pares ordenados, conforme a seguir:

$$A = \{(x, \mu_A(x)); x \in X(x) \in [0, 1]) \mu_A(x) \in [0, 1] \} \quad (4.9)$$

Onde:

- $\mu_A(x)$ é uma função de pertinência que determina com que grau x está em A ;
- $\mu_A(x) = 1$, x pertence totalmente ao conjunto;
- $0 < \mu_A(x) < 1$, x pertence parcialmente ao conjunto A ;
- $\mu_A(x) = 0$, x não pertence ao conjunto A ;

Um subconjunto *fuzzy* A de x é definido por uma função característica μ , chamada de função de pertinência, que associa a cada elemento de x um grau $\mu_A(x)$, compreendido entre 0 e 1, com o qual x pertence a A :

$$\mu_A : x \rightarrow [0, 1] \quad (4.10)$$

Sendo assim, a álgebra entre os conjuntos *fuzzy* não segue a teoria clássica usual em especial devido ao fato de que as operações de união e de intersecção de um conjunto e seu complementar são diferentes de 1 e 0. Portanto, as operações específicas entre os conjuntos *fuzzy* devem ser estabelecidas [7]. As funções de intersecção ou de união entre os conjuntos A e B são facilmente expressos por uma função indicadora representada por:

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] \quad (4.11)$$

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] \quad (4.12)$$

Do mesmo modo, o indicador para a função complementar de um conjunto $A(\tilde{A})$ é dada por

$$\mu_{\tilde{A}}(x) = 1 - \mu_A(x) \quad (4.13)$$

A condição para A ser subconjunto de B pode ser expresso por:

$$A \subset B \Leftrightarrow \mu_A(x) \leq \mu_B(x), \forall x \in X \quad (4.14)$$

Os conjuntos convencionais, descritos acima, permitem que cada elemento do conjunto universo X é considerado um elemento *fuzzy* A para um determinado grau. Assim, a função característica denotado pelo seu grau que pertença a um determinado elemento em um conjunto *fuzzy* representado pelo valor entre o intervalo de 0 e 1. Se X é denotado pela coleção de objetos genericamente descrito por $\{x\}$, então o conjunto *fuzzy* A poder definido pelo pares ordenados:

$$A = \{(x, m_A(x)) | x \in X\} \quad (4.15)$$

Onde $m_A(x)$ é a característica da função *fuzzy* de X em A que mapeia x no intervalo $[0, 1]$. Quando $m_A(x)$ assume apenas valores 0 e 1, então A é considerado não *fuzzy* (*crisp*) e $m_A(x)$ é idêntica a função característica do conjunto não *fuzzy*. Zadeh [98] definiu a intersecção (Mínimo) e a união (Máximo) de conjuntos *fuzzy* A e B , e do seu complemento de A para expressar padrões entre os conjuntos. Da mesma forma Zadeh [97] propos a definição de A a ser um subconjunto de B usando as características das funções para A e B [29].

$$A \subset B \Leftrightarrow m_A(x) \leq m_B(x), \forall x \in X \quad (4.16)$$

4.1.4 Propriedades

Os operadores *min* e *max*, complemento, união e intersecção satisfazem as seguintes propriedades, como na teoria clássica dos conjuntos nebulosos:

- Comutativa

$$A \cap B = B \cap A \quad (4.17)$$

$$A \cup B = B \cup A \quad (4.18)$$

- Associativa

$$(A \cap B) \cap C = A \cap (B \cap C) \quad (4.19)$$

$$(A \cup B) \cup C = A \cup (B \cup C) \quad (4.20)$$

- Idempotência

$$A \cap A = A \quad (4.21)$$

$$A \cup A = A \quad (4.22)$$

- Distributividade em relação a intersecção

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (4.23)$$

- Distributividade em relação a união:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (4.24)$$

- Conjunto *fuzzy* e o seu complemento:

$$A \cap \tilde{A} \neq \phi \quad (4.25)$$

$$A \cup \tilde{A} \neq X \quad (4.26)$$

- Conjunto *fuzzy* e o conjunto nulo:

$$A \cap \phi = \phi \quad (4.27)$$

$$A \cap \phi = A \quad (4.28)$$

- Conjunto *fuzzy* e o conjunto universal:

$$A \cap X = A \quad (4.29)$$

$$A \cup X = X \quad (4.30)$$

- Involução

$$(\tilde{\tilde{A}}) = A \quad (4.31)$$

- Teorema de Morgan:

$$(A \cap B)^{\sim} = \tilde{A} \cup \tilde{B} \quad (4.32)$$

$$(A \cup B)^{\sim} = \tilde{A} \cap \tilde{B} \quad (4.33)$$

4.1.5 Operadores *Fuzzy*

4.1.5.1 Intersecção

A função de pertinência da intersecção $A \cap B$ pode ser representada pela função de pertinência $\mu_{A \cap B}(x)$, onde $x \in X$, podendo ser definida por:

$$\mu_{A \cap B} = \mu_A(x) \mathbf{t} \mu_B(x) \leq \min[\mu_A(x), \mu_B(x)] \quad (4.34)$$

Onde \mathbf{t} é a norma triangular ou norma- \mathbf{t} de uma intersecção generalizada. A norma- \mathbf{t} corresponde a uma função de duas entradas definidas pelo domínio $\mathbf{t}: [0, 1] \times [0, 1] \rightarrow [0, 1]$, onde X denota o operador de produto cartesiano. O operador- \mathbf{t} indica um mapeamento entre duas funções de pertinência *fuzzy*, cada uma no intervalo $[0, 1]$; a operação- \mathbf{t} é executada entre elas, de acordo com seu formato. Tal função de mapeamento deverá satisfazer as seguintes condições:

-Condições de contorno:

$$x \mathbf{t} 0 = 0, \forall x \in [0, 1] \quad (4.35)$$

Serve para manter a consistência com a observação em um diagrama de *Venn*, onde a intersecção de qualquer conjunto vazio constitui o próprio conjunto vazio.

- Condições de contorno:

$$x \mathbf{t} 1 = x, \forall x \in [0, 1] \quad (4.36)$$

Similarmente, um diagrama de *Venn* mostraria imediatamente que a intersecção de qualquer conjunto *fuzzy* como o conjunto universal, deve ser o conjunto por si só [60].

- Propriedade comutativa : $x \mathbf{t} y = y \mathbf{t} x$

- Propriedade associativa: $x \mathbf{t} (y \mathbf{t} z) = (x \mathbf{t} y) \mathbf{t} z$

- Condições Monotônicas : para $x \leq y$ e $w \leq z$, $x \mathbf{t} w \leq y \mathbf{t} z$

As normas-**t** podem ser: intersecção, produto algébrico, produto logarítmico, produto inverso, produto limitado e produto drástico [60], alguns exemplos poderão ser vistos a seguir:

-Intersecção: $x \mathbf{t} y = \min(x, y)$

-Produto algébrico: $x \mathbf{t} y = xy$

-Produto drástico:

$$x \mathbf{t} y = \begin{cases} x, & \text{quando, } y = 1 \\ y, & \text{quando, } x = 1 \\ 0, & \text{quando, } x, y < 1 \end{cases} \quad (4.37)$$

4.1.5.2 União

A função de pertinência $\mu_{A \cup B}(x), x \in X$, da união $A \cup B$, pode ser definida por:

$$\mu_{A \cup B}(x) = \mu_A(x) \mathbf{s} \mu_B(x) \leq \max[\mu_A(x), \mu_B(x)] \quad (4.38)$$

Onde **s** é a co-norma triangular de uma união generalizada. A norma-**s** constitui-se uma função de duas entradas definida nos domínios **s**: $[0, 1] \times [0, 1] \rightarrow [0, 1]$, onde X denota o produto cartesiano e o operador - **s** indica o mapeamento entre duas funções de pertinência *fuzzy*, onde cada uma delas estará entre o intervalo $[0, 1]$, satisfazendo as seguintes condições:

- Condições de contorno: $x \mathbf{s} 0 = x, \forall x \in [0, 1]$

Essas condições de contorno servem para manter a consistência com a observação em um diagrama de *Venn*, onde a união (contorno comum) de qualquer conjunto com o conjunto vazio

é obviamente o próprio conjunto [60]:

-Condições de contorno: $x \text{ s } 1 = 1, \forall x \in [0, 1]$

Similarmente, um diagrama de *Venn* mostra que a união de qualquer conjunto *fuzzy* com o conjunto universal, deve ser o conjunto universal [60]:

- Propriedade comutativa: $x \text{ s } y = y \text{ s } x$

- Propriedade associativa: $x \text{ s } (y \text{ s } z) = (x \text{ s } y) \text{ s } z$

- Condições Monotônicas: para $x \leq y$ e $w \leq z$, $x \text{ s } w \leq y \text{ s } z$

As normas-s podem ser: união, soma algébrica, soma limitada, soma logarítmica, soma disjunta e soma drástica, para que $x, y \in [0, 1]$ são:

- União: $x \text{ s } y = \max(x, y)$

- Soma algébrica: $x \text{ s } y = x + y - xy$

4.1.6 Fuzzificação

O processo de fuzzificação constitui-se no relacionamento dos conceitos vagos e imprecisos da linguagem natural a ser traduzida em valores de entrada em conjuntos nebulosos, em outras palavras poderá ser descrita por um critério de tradução dos valores de entrada físico em uma forma de descrição nebulosa, tornando-se qualquer numero *crisp* em numero *fuzzy*. A fuzzificação envolve processos de tornar valores numéricos (não *fuzzy*) das variáveis de entrada, proporcionando um mapeamento para condicionar os valores de entrada a universos normalizados, fuzzificando os valores, convertendo-os em conjuntos nebulosos de modo que possam se tornar instâncias de variáveis lingüísticas [36], [61] e [57]. Cada variável de entrada e seus termos lingüísticos deve ser associado a um conjunto *fuzzy* de acordo com uma função de pertinência [7].

4.1.7 Variáveis Lingüísticas

Uma variável lingüística *fuzzy* permite modelar o conhecimento impreciso e vago sobre uma medição quaisquer, cujo valor pode ser desconhecido e seus elementos simbólicos utilizados para descrever o conhecimento [21, 64, 99]. Assim, uma variável lingüística pode ser caracterizada por uma terna (V, X, T_v) , onde V é uma variável definida sobre um conjunto de referência X . O conjunto $T_v = \{M_1, M_2, \dots\}$, finito ou infinito sobre um conjunto de referência X que serve para caracterizar V , utilizando a mesma notação (M_i) para representar os subconjuntos *fuzzy* associados aos seus respectivos valores lingüísticos ("*mínimo*", "*médio*", "*máximo*", *entre outros*), conforme a seguir:

4.1.7.1 Modelagem das Variáveis Linguísticas - Caso Fluxo de Potência

No problema do fluxo de potência (capítulo 6), a variável linguística "Fator de Potência", poderá ser representada pela figura a seguir:

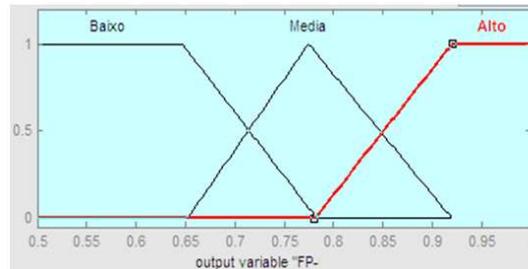


Figura 4.2 Função de pertinência da variável linguística "Fator de Potência".

Considere um intervalo de $[0 \ 1]$ número que indica o quanto da energia elétrica pode ser transformada em outras formas de energia [93]. Esta variável terá três valores linguísticos: "baixo", "medio" e "alto". Estes valores, podem ser caracterizados como subconjuntos *fuzzy*, cujas funções de pertinência são representadas na figura 4.2. Assim, para um fator de potência (FP) $\leq 0,64 =$ "baixo" com valor de pertinência igual a 1,0, "medio" para um FP $\geq 0,65 \leq 0,91$ com valor de pertinência igual a 0,0, "alto" para um FP $\geq 0,92 \leq 1,0$ com valor de pertinência igual a 1,0. Assim, qualquer FP medida será sempre associada aos valores linguísticos da variável com um valor de pertinência entre (0,0) e um (1,0).

4.1.7.2 Modelagem das Variáveis Linguísticas - Caso Sinais Biomédicos para pacientes portadores de Diabetes

Na análise de sinais biomédicos (capítulo 6), a variável linguística "Diagnóstico da Curva Glicêmica", poderá ser representada pela figura a seguir:

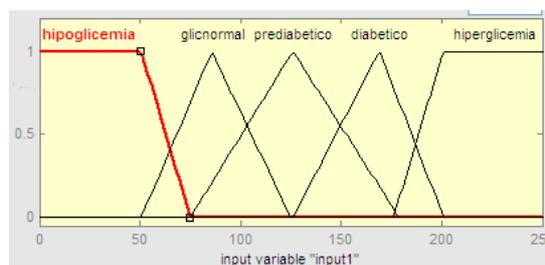


Figura 4.3 Função de pertinência da variável linguística "DiagnósticoCurvaGlicêmica".

Considere um intervalo de $[0 \ 250]$ número que indica os níveis de glicose no sangue. Esta variável terá cinco valores linguísticos pertinente a classificação dos níveis de glicose provenientes dos testes glicêmicos, tais como: "hipoglicemia", "glicemia normal", "pré-diabético", "diabético" e "hiperglicemia" (vista com maiores detalhes no capítulo 6), cujas funções de pertinência são representadas na figura 4.3. Assim, para o Diagnóstico da Curva Glicêmica (DCG)

no caso da glicose (GLI) ≤ 50 mg/dl (definir) de sangue = hipoglicemia com valor de pertinência (VPE) igual 1.0, quando $GLI \geq 50 \leq 120$ mg/dl = glicemia normal com VPE = 0.0, no caso de $GLI \geq 125 < 200$ mg/dl = pré-diabético com VPE = 0.0, quando $GLI = 200$ mg/dl é identificado quadro efetivo de diabetes obtendo VPE = 0.0 e para $GLI \geq 200$ mg/dl = hiperglicemia com VPE = 1.0. Assim, qualquer DCG medido será sempre associada aos valores lingüísticos da variável com um valor de pertinência entre (0.0) e um (1.0).

4.1.8 Regras Fuzzy

As regras *fuzzy* são essenciais a estrutura de conhecimento em um sistema *fuzzy* de inferência, a qual relaciona variáveis fuzzy associadas a um dos seus predicados ou termos lingüísticos ("mínimo", "médio", "máximo"), conectados através de operadores lógicos como: e, ou, então etc, os quais se mostram como conseqüência das operações "e/ou"algébricas, onde "e"designa união e "ou" a intersecção. Os formatos de regras *fuzzy* podem ser divididos em quatro grupos principais a partir dos métodos Mamdani, Takagi-Sugeno e Tsukamoto, os quais correspondem ao modelo de inferência *fuzzy*. A diferença básica entre esses três primeiros modelos recai no tipo de conseqüente e no procedimento de defuzzificação [64].

4.1.9 Método de Mandani

O método Mandani pode ser considerado um dos primeiros métodos utilizando na teoria dos conjuntos *fuzzy* [82]. Os sistemas de controle construídos para utilizar a lógica *fuzzy*, foi proposto por Mamdani (1975)[58] como uma tentativa de controlar um conjunto de turbina a vapor com intuito de sintetizar uma série de variáveis lingüísticas integradas a um controlador, usando regras derivadas de um especialista humano sobre algoritmos *fuzzy* para sistemas complexos e de tomada de decisão [99]. A modelagem matemática da base de regras do método Mandani, propuseram uma relação *fuzzy* binária M entre as entradas x e as saídas u , cuja função de pertinência [7] e será especificada conforme a seguir:

$$\varphi_M(x, u) = \max_{1 \leq j \leq n} \{ \min [\varphi_{A_j}(x), \varphi_{B_j}(u)] \} \quad (4.39)$$

Onde n é o número de regras que compõem a base de regras e $\varphi_{A_j}(x)$ e φ_{B_j} são os graus com que x e u pertencem aos conjuntos *fuzzy* A_j e B_j , respectivamente.

Em cada uma das regras da base de regras *fuzzy*, adota-se o operador matemático mínimo para o conectivo lógico "e" e o operador máximo para o conectivo lógico "ou". As sentenças "se...então..." são modeladas pela aplicação mínimo". Assim, o operador pode ser utilizado para definir a intersecção de dois subconjuntos *fuzzy*, já o operador da equação anterior é utilizado para definir a união de dois subconjuntos *fuzzy*, entretanto existem outros operadores definidos na literatura por exemplo: o produto de $(u * x)$ (o máximo entre $[(u + x - 1)$ e zero, o mínimo entre $[(u + x)$ e o valor 1 e [60].

De acordo com a equação 4.4, podemos representar as funções de pertinência máximo e mínimo a relação *fuzzy* de M para o caso de duas entradas e uma saída u . Sejam as regras compostas por $R_1 : SE x_1 \text{ é } A_1^1 \text{ e } x_2 \text{ é } A_2^1 \text{ então } u \text{ é } B_1$ e $R_2 : SE x_1 \text{ é } A_2^1 \text{ e } x_2 \text{ é } A_1^2 \text{ então } u \text{ é } B_2$.

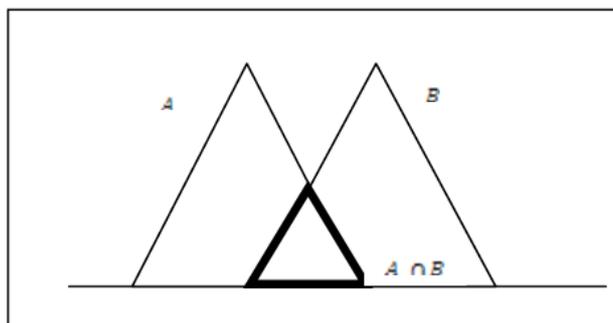


Figura 4.4 Função de pertinência $\mu_{A \cap B}(x) = \min$

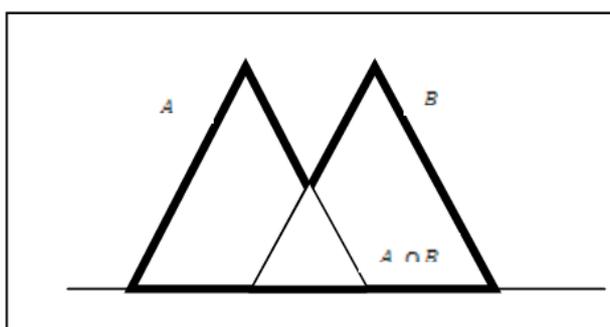


Figura 4.5 Função de pertinência $\mu_{A \cup B}(x) = \max$

A figura 4.6 e 4.7 ilustra como a saída u , de um sistema de inferência do tipo de Mandani é gerada.

A saída precisa (*crisp*) é obtida pela defuzzificação do conjunto fuzzy de saída resultante da aplicação das operações que implementam a interseção (and) e a união (or) entre dois conjuntos nebulosos.

4.1.10 Método de Takagi-Sugeno

No modelo nebuloso de Takagi-Sugeno, a saída de cada regra é uma função das variáveis de entrada. Geralmente, a função mapeia a entrada e saída para cada regra a partir de uma combinação linear das entradas [57]. O modelo criado por Sugeno [42] tem intuito de formalizar um sistema de abordagem nebulosa para gerar regras de entrada-saída de um conjunto de dados, conforme o que segue:

$$R_1 : SE \ x_1 \ A \ e \ y \ B \ entao \ z = f(x,y) \quad (4.40)$$

Onde: A e B são conjuntos nebulosos no antecedente (entradas) e $f(x,y)$ é um polinômio em função das variáveis x e y , contendo regras de produção que associam as variáveis lingüís-

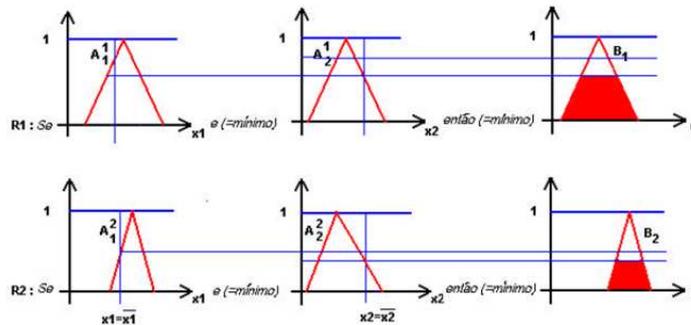


Figura 4.6 Representação esquemática do método de Mandani

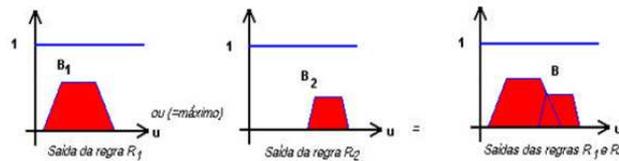


Figura 4.7 Representação esquemática da saída final do método de Mandani

ticas de entrada com as de saída. As regras para modelar o processo podem ser do tipo "i", onde cada regra é uma combinação linear das entradas [13]:

$$R_1 : SE \{x_1 A_1^i(x)\} e \dots \{x_M A_M^i(x)\} \text{ entao } \{y = p_0^i + p_1^i + \dots + p_M^i x_M\} \quad (4.41)$$

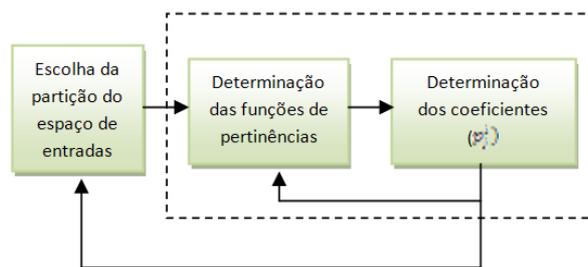


Figura 4.8 Algoritmo de identificação das regras fuzzy [13]

O objetivo deste modelo constitui-se em definir um número de regras N (partição *fuzzy* do universo de entrada do sistema), onde as funções de pertinência $A_j^i(x)$ e os coeficientes p_j^i , aparecem nas conclusões das regras. Essas identificações não são independentes, pelo motivo que, a identificação do número de regras é realizada empiricamente a partir dos conhecimentos dos especialistas ou através de uma otimização heurística e a identificação das funções de pertinência é realizada empiricamente ou por um método de programação não-linear, onde os

coeficientes p_j^i são obtidos pela minimização do erro, isto é, o desvio entre a saída do modelo e a saída do processo [13].

Abordagem paramétrica fornece uma precisão de estimação melhor, por ser inerentemente adaptativo porque os coeficientes de p_j^i podem ser alterados para ajuste do sistema, onde uma implementação adaptativa em tempo real da abordagem paramétrica é possível, mudando-se dinamicamente os coeficientes lineares por meio do algoritmos de mínimo quadrados recursivos, aplicados de modo recorrente, com frequência determinada ou quando necessário em função de alguns parâmetros observados [60].

4.1.11 Método de Tsukamoto

O sistema *fuzzy* de inferência utiliza funções de pertinência monotônica (item 4.1.8) no conseqüente, por exemplo, considere a regra : Se x é A e y é B então z é C monotônica. Segundo Christer [14, 15], o modelo Tsukamoto é similar ao modelo Tagaki-Sugeno, o processo de inferência utiliza-se de um valor não-*fuzzy* induzido pelo nível de ativação da regra e a saída final é obtida mediante a média ponderada das saídas inferidas de cada regra.

4.1.12 A base de Conhecimento

A base de conhecimento representa o modelo do sistema a ser controlado, possui as informações pertinentes ao universo de referência dos valores lingüísticos e das funções de pertinência de todas as variáveis do sistema. A base de dados fornece as definições numéricas necessárias às funções de pertinência usadas no conjunto de regras *fuzzy* [36]. A base de conhecimento de um controlador *fuzzy* é considerada o "cérebro" do sistema, onde está localizada a inteligência do mesmo. Os procedimentos necessários para aquisição da base de conhecimento dar-se-á a parte mais importante, demorada e crítica durante o desenvolvimento de um controlador nebuloso, senão for bem especificada poderá comprometer o futuro controlador [13].

4.1.13 Mecanismo de Raciocínio

Realiza um procedimento de inferência para obter a saída ou a conclusão baseado nas regras e nos fatos conhecidos (operadores: agregação de antecedente, semântica da regra, agregação das regras e o método de defuzzificação), consiste em utilizar regras de controle na presença de variáveis de entrada medidas no processo e que satisfazem parcialmente as condições de utilização das regras (ou antecedentes das regras), isto é, consiste em efetuar um "Modus Ponens generalizado"[13].

4.1.14 Defuzzificação

Uma vez executadas as avaliações das regras *fuzzy*, deve-se determinar o valor real da saída do sistema nebuloso. Como todas as regras do controlador nebuloso são avaliadas em paralelo e de forma independente, e cada regra dá um resultado diferente para a variável de controle,

deve-se ter um modo para ponderar todos estes resultados e escolher a saída do controlador, processo este chamado de decodificação do controlador *fuzzy* ou simplesmente defuzzificação [13]. Muitos métodos de defuzzificação têm sido propostos na literatura, os mais utilizados em sistemas de controle podem ser descritos a seguir:

4.1.15 Método centro da área ou centróide

O método de área é freqüentemente chamado de método centro de gravidade ou centróide, tem por objetivo calcular o centro de gravidade da área composta pelas regras inferidas. É considerado um dos métodos mais utilizados. Seja um universo de discurso discreto, a saída *crisp* z , é produzida pelo cálculo do centro de gravidade do conjunto *fuzzy* conseqüente μ_c obtido a partir da composição das regras representada pela equação:

$$z = \frac{\sum_{i=0}^m \mu_c(z_i) * Z_i}{\sum_{i=0}^m \mu_c(z_i)} \quad (4.42)$$

Onde:

- m é o número de intervalos de quantização da saída;
- Z_i é o valor da variável de saída para o intervalo de quantização "i";
- μ_c seu grau de pertinência.

4.1.16 Método da Média Ponderada

Neste método são usados como conseqüentes das regras *fuzzy*, combina os conseqüentes das regras *fuzzy* como nível de disparo de cada uma delas, gerando uma saída. Assumindo que os valores das funções de pertinência de saída obtidos são ponderações e bastante utilizado na defuzzificação dos modelos *fuzzy* de Takagi-Sugeno de 1ª ordem (onde a saída é a combinação linear das entradas) e a de Tsukamoto que utiliza funções monotônicas nos conseqüentes das regras. A expressão da saída poderá ser obtida a partir da expressão:

$$z = \frac{\sum_{i=1}^n \mu_i * Z_i}{\sum_{i=1}^n \mu_i} \quad (4.43)$$

Onde:

- n é o número de regras *fuzzy*;
- μ_i é o nível de disparo da regra i ;
- Z_i é o valor do "singleton" de i .

4.1.17 Método Centro de Somas

Este método se dá a uma simplificação do método centróide. A grande vantagem de utilizar este método está relacionado ao custo computacional durante o cálculo de saída quando comparado com o método do centro de gravidade:

$$z = \frac{\sum_{i=0}^m C^u (B_i^{\infty i})}{\sum_{i=0}^m \text{area} (B_i^{\infty i})} \quad (4.44)$$

Onde:

- $B_i (i = 1, \dots, n)$, são termos lingüísticos dos conseqüentes; - α_i é o nível de disparo da regra i ; - $(B_i^{\infty i})$ é o conjunto *fuzzy* obtido através do método de implicação de Mandani, ou seja, B_i cortado em α_i como :

$$C^u = (B_i^{\infty i}) = \text{centroide}(B_i^{\infty i}) \times \text{area} (B_i^{\infty i}) \quad (4.45)$$

4.1.18 Considerações finais

Neste capítulo foram abordados os principais aspectos teóricos relacionados aos sistemas nebulosos. Algumas vantagens poderão se adquiridas na literatura no que se refere a especificação e modelagem de um controlador *fuzzy*, devido ao fato de ser uma tecnologia fácil de ser implementada em tempo real e de ser comprovadamente robusta em aplicações práticas e capazes de controlar processos complexos. Embora as funções de pertinência mostradas neste capítulo foram expostas no formato triangulares e trapezoidais, as quais não pode ser considerada modelos obrigatórios. A quantidade de funções em um universo de discurso e seu formato geralmente são escolhidos com base na experiência, na natureza do processo a ser controlado, ou simplesmente numa entrevista com um operador humano especializado, que realize as funções de controle manualmente.

CAPÍTULO 5

Construção do Sistema de Controle *Fuzzy* do Mapa de Kohonen

Os capítulos anteriores apresentaram fundamentos teóricos, metodologias e ferramentas para aplicação técnicas dos métodos de mineração de dados, neste capítulo será apresentado um sistema abordando a interpretação nebulosa do Mapa de Kohonen a partir de um sistema de controle *fuzzy*, desenvolvido em Matlab versão 7.0.

5.1 O Sistema de Controle

O sistema *VisionFuzzy* é composto por dois módulos: no primeiro módulo, o sistema apresenta as configurações do mapa de Kohonen, as quais são inseridas no sistema através da entrada de dados tais como: dimensão do mapa (formada pela rede de neurônios na grade bidimensional), os parâmetros de parada estabelecido pelas épocas (provém dos ciclos da rede pertinente ao intervalo de apresentação do conjunto completo de exemplos a partir do mapeamento fino e grosso do mapa).

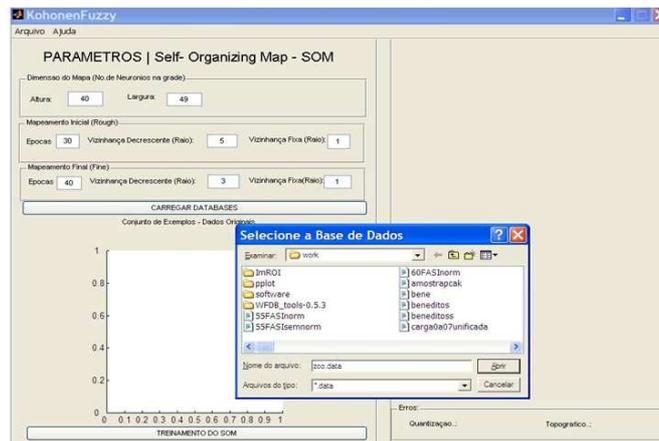


Figura 5.1 Seleção dos parâmetros e base de dados do sistema *VisionFuzzy*

Os parâmetros deverão estabelecer quantas épocas o modelo deverá parar o treinamento. Uma época corresponde à apresentação de todos os exemplos de treinamento na rede SOM. Em seguida, a base de dados deverá ser importada para o sistema com objetivo de iniciar o processo de treinamento da rede conforme figuras 5.1, 5.2 e 5.3.

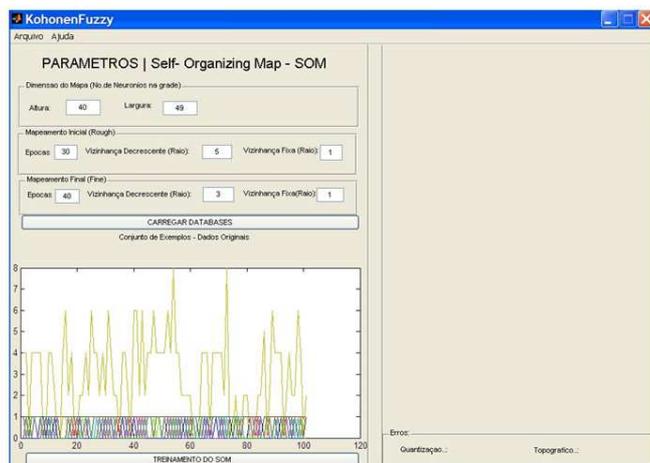


Figura 5.2 Visualização do conjunto de exemplo

As figuras 5.1 e 5.2, apresentam as configurações básicas e os parâmetros da rede a partir da visualização do conjunto de exemplos com objetivo de inicializar o treinamento, conforme figura 5.3.

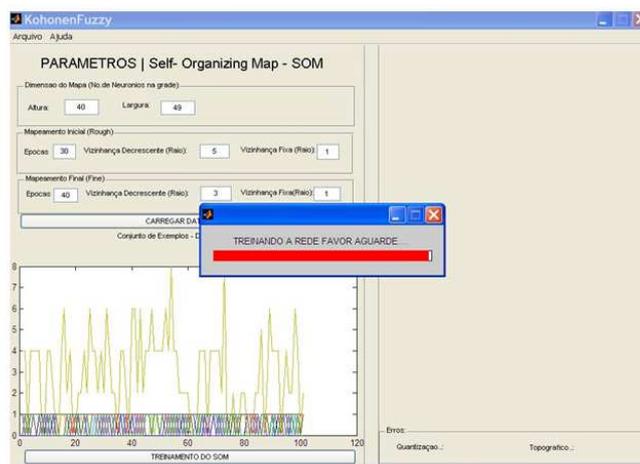


Figura 5.3 Treinamento da rede de acordo com os parâmetros

A configuração dos parâmetros pode ser vistos em Kohonen [52] para uma melhor adaptação do mapa, utilizando conexões laterais inibitórias entre os nodos de saída. Nestes nodos os neurônios que estão espacialmente ordenados dentro das áreas de raio e tendem a responder a padrões de estímulos semelhantes a partir do resultado obtido através da competição.

No segundo módulo do sistema, o mapa de Kohonen é visualizado através da matrix-U e do mapa contextual exibindo a classificação do conjunto de dados de acordo com as respectivas similaridades a partir da análise de agrupamentos. Em seguida, o sistema efetuará a transformação dos dados adotando a transposição da base de dados pré-rotulada com intuito de unificar as variáveis estabelecidas com as entradas do controlador *fuzzy*, ou seja, cada coluna transposta

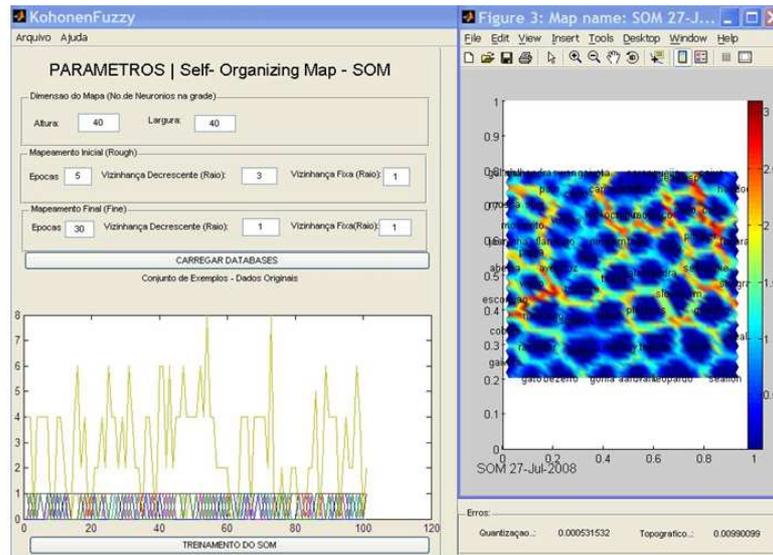


Figura 5.4 Treinamento da rede e visualização do mapa

da matriz original de dados proveniente do mapa de Kohonen corresponde a uma entrada do controlador *fuzzy* e assim, sucessivamente.

No caso do fluxo de potência, a primeira coluna da matriz de dados representa o intervalo da carga no período de 21 às 24 horas (carga média), a segunda coluna representa o intervalo da carga no período de 18 às 21 horas (carga pesada) e assim sucessivamente a partir dos intervalos das demais cargas correspondente aos períodos de 7 às 18 horas (coluna 3 - carga média) e 0 às 7 horas (coluna 4 - carga leve). Esta metodologia também é adotada para o problema dos sinais biomédicos (capítulo 6).

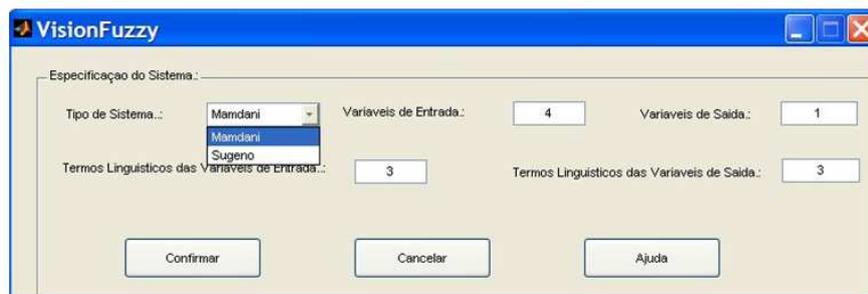


Figura 5.5 Parâmetros do Controlador *Fuzzy*

As configurações básicas do controlador são definidas inicialmente para a modelagem *fuzzy* do mapa, por exemplo, tipo de sistema de inferência (Mamdani), onde as variáveis de entrada, saída e seus termos lingüísticos são definidos. Em seguida, o sistema é projetado para a composição das principais características e ajustes das variáveis de entrada/saída.

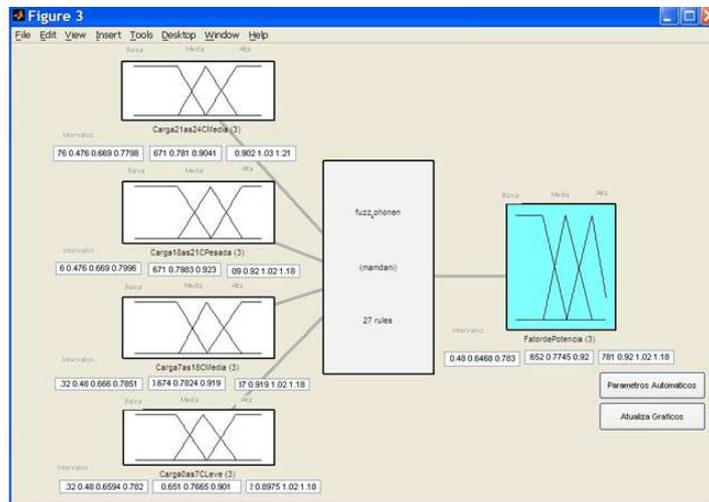


Figura 5.6 Sistema de Controle

Nesta fase, a entrada do sistema fuzzy corresponde a saída do mapa de Kohonen, o qual será encarregado de atualizar as funções de pertinência e gerar o padrão de dados específico aos intervalos das vizinhanças entre as classes.

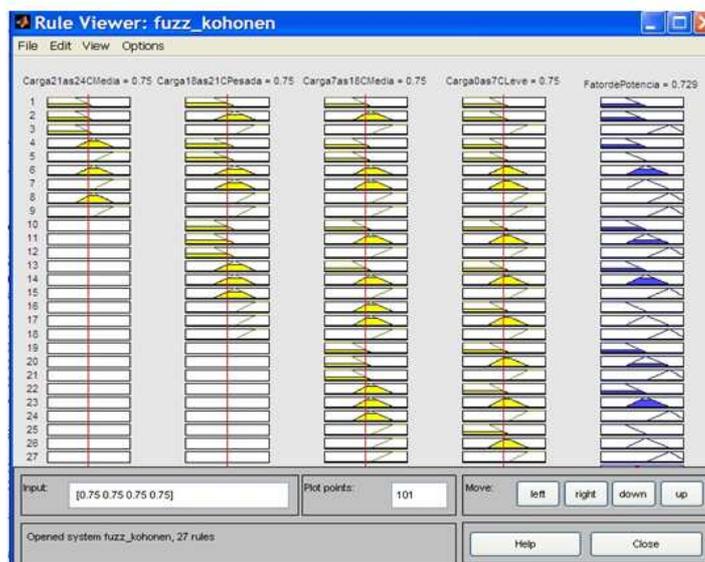


Figura 5.7 Visualização do conjunto de regras do sistema

As regras do sistema podem ser visualizadas a partir da demarcação dos centros de gravidade sobre os conjuntos de variáveis de entrada e saída do controlador.

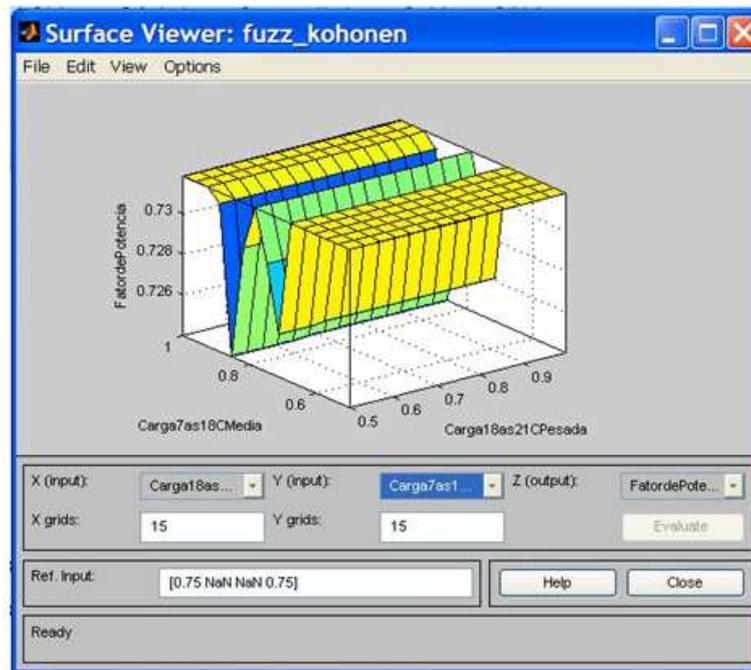


Figura 5.8 Gráfico de controle para o exemplo de cargas

O mapa de regras estabelecido no item 6.7, corresponde a uma implicação *fuzzy* matricial, onde as entradas corresponde a uma determinada saída ou similarmente de saídas associadas com determinadas entradas. Essa matriz pode ser entendida como uma superfície de controle tridimensional onde o eixo vertical é a variável de saída "Fator de Potência", cada eixo horizontal relaciona-se à integração da carga "Média" com a carga "Pesada".

5.1.1 Considerações Finais

Este capítulo abordou as principais funcionalidades sistema de controle *fuzzy* a partir do mapa de Kohonen. A definição dos estudos de casos e da modelagem do sistema de controle serão vistos com maiores detalhes no capítulo 6.

Aplicações em Mineração de Dados

Estudo de Casos

Este capítulo busca aplicar os mapas auto-organizáveis de Kohonen e lógica nebulosa como métodos alternativos para mineração de dados a partir dos métodos apresentados nos capítulos anteriores. Serão utilizados nos testes um conjunto de dados públicos e um conjunto de dados de um caso real sobre a classificação dos sinais provenientes dos sistemas elétricos de potência e sinais biomédicos. Ambos os conjuntos de dados apresentam uma aplicação prática numa área tradicionalmente não explorada pelos métodos de mineração de dados, fluxo de potência e na classificação dos níveis de glicose provenientes dos testes glicêmicos em pacientes portadores de diabetes oferecendo subsídios teóricos e práticos para avaliação e interpretação nebulosa dos mapas de Kohonen como método de mineração de dados.

6.1 Conjunto de Dados - Caso Real: Problema do Fluxo de Potência

6.1.1 Introdução

Os Sistemas Elétricos de Potência (SEP) têm como principal objetivo suprir as necessidades dos consumidores de energia elétrica proporcionando custos mais acessíveis correspondente ao equilíbrio entre a oferta e a demanda da eletricidade objetivando a qualidade de suprimento. De acordo com Kagan [46], os sistemas elétricos de potência têm função precípua de fornecer energia elétrica aos usuários, grandes ou pequenos, com a qualidade adequada, no instante em que for solicitada. Os processos de geração, transmissão e fornecimento de energia tem se tornado cada vez mais significativos na sociedade devido a sua influência na movimentação do mercado, principalmente no que se refere aos aspectos econômicos a partir do recebimento de energia elétrica (as fábricas, hospitais, comércio, escolas etc).

Essa integração do setor elétrico com a sociedade faz com que as concessionárias de energia se preocupem com a confiabilidade dos sistemas elétricos e com o fornecimento de energia de qualidade de acordo com as normas estabelecidas pela Agência Nacional de Energia Elétrica [10] e Operador Nacional do Sistema Elétrico [70], os quais propõem normas aplicadas a estabelecer critérios de análise das condições operacionais dos sistemas elétricos de modo que permita avaliar seu desempenho e definir ações preventivas e corretivas a serem tomadas, para que os consumidores possam ser atendidos com qualidade no fornecimento de energia, sem que os equipamentos das concessionárias de energia elétrica fiquem submetidos a condições operacionais e inadmissíveis.

As concessionárias de energia elétrica buscam cumprir as normas estabelecidas pelos

órgãos reguladores, criando programas de fluxo de potência que permitam analisar e controlar os SEPs. Segundo Carneiro [8] "Em se tratando de sistemas de distribuição de energia elétrica, o desenvolvimento de programas de fluxo de potência que incorporem características típicas como operação desequilibrada, alimentadores com alta relação e geração dispersa, e ainda que permitam a análise de novos conceitos introduzidos com a recente reestruturação do setor elétrico é de vital importância para a operação e planejamento da distribuição de energia elétrica". Assim, o estudo do fluxo de potência da rede, se dá a partir da resolução do circuito elétrico que representa a rede, para o qual se dispõe da topologia, com as constantes elétricas de seus elementos, das demandas das cargas e das tensões dos geradores que o excitam [6, 46]. Os estudos que abordam a análise do fluxo de potência têm como objetivo permitir a simulação de operações fornecidas pela rede e analisar os principais fatores que influenciam no processo de distribuição, transmissão e fornecimento de energia no que se refere a: comparação de demanda, fluxo, desequilíbrio da carga, desempenho da rede [6];

As aplicações mais comuns do fluxo de potência podem ser classificadas: simulação de sistemas elétricos considerando diferentes despachos das usinas geradoras de energia elétrica, de modo a se prever quais as condições operacionais decorrentes desses despachos; simulação de sistemas elétricos operando sobre condições anormais decorrentes da saída de operação de equipamentos como linhas de transmissão, transformadores e unidades geradoras; simulação de sistemas elétricos com finalidade de planejamento de sua expansão [22]. O fornecimento de energia elétrica a consumidores se dá a excepcional importância ao exercício das medidas elétricas em consequência das implicações de caráter comercial e legal, as quais permitem que as medições possam se tornar cada vez mais viáveis economicamente de acordo com as tecnologias de medições disponíveis no mercado.

6.1.2 Sistema Elétrico de Potência

Os SEPs têm principal função o processo de geração no qual transforma a energia (hidráulica, mecânica, térmica etc.) de alguma natureza em energia elétrica; os sistemas de transmissão são responsáveis pelo transporte da energia elétrica dos centros de produção aos de consumo; os de distribuição são capazes de distribuir a energia elétrica recebida do sistema de transmissão aos grandes, médios e pequenos consumidores. Assim, as concessionárias deverão dispor de sistemas de controle da produção de modo que, a cada instante seja produzida a energia necessária para atender à demanda e as perdas na produção e no transporte.

Devido a grande exploração do potencial hídrico no Brasil, a transformação de energia hidráulica em elétrica pode ser considerada a fonte mais usada na produção de energia. Para que os SEPs sejam controlados e analisados de forma eficaz, se faz necessário analisar fatores significativos os quais podem influenciar no atendimento da demanda e nas perdas de produção, por exemplo, a localização geográfica dos centros de produção de energia elétrica poderão ser afastados dos centros de consumo proporcionando de forma imprescindível a existência de um elemento de interligação entre os centros de produção e consumo, os quais sejam aptos a transportar a energia demandada.

As potências demandas durante o processo de transporte poderão ser influenciadas pelas distâncias percorridas pelo sistema e pode se tornar inexecutável na tensão de geração.

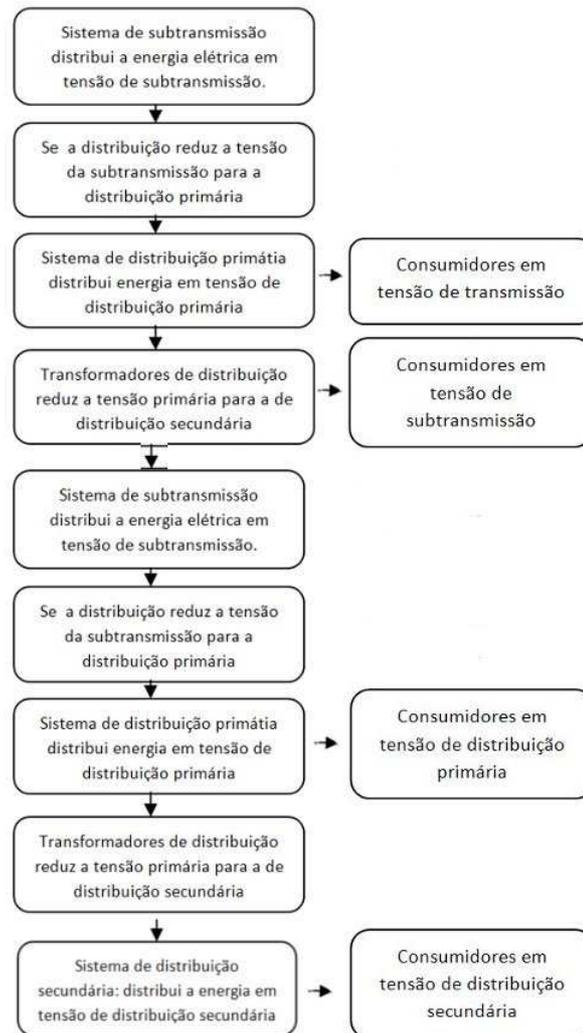


Figura 6.1 Diagrama de blocos do sistema [46]

O diagrama de blocos (figura 6.1) abordado por Kagan [46], apresenta o bloco de geração, o de elevação da tensão, correspondente ao valor no qual foi gerada para o de transporte, "tensão de transmissão", o valor dessa tensão é estabelecido em função da distância a ser percorrida e do montante de energia a ser transportada, pode ser detalhada da seguinte forma: o primeiro abaixamento do nível de tensão para o valor compatível com a demanda dos grandes usuários, "tensão de subtransmissão" é executada a partir das "subestações de subtransmissão", que são supridas através de linhas de transmissão, que por sua vez, as linhas que operam em nível de tensão mais baixo, "tensão de subtransmissão" ou "alta tensão". Esses abaixamentos são adotados em decorrência das características dos consumidores. Assim, o sistema de subtransmissão supre as "subestações de distribuição", que são responsáveis pelo novo abaix-

amento do nível de tensão para a "tensão de distribuição primária" ou "média tensão". A rede de distribuição primária, por sua vez, irá suprir os transformadores de distribuição, dos quais se deriva a rede de distribuição secundária ou rede de baixa tensão, cujo nível de tensão é designado por "tensão secundária" ou "baixa tensão", a qual poderá ser visualizada no diagrama unifilar, conforme figura 6.2.

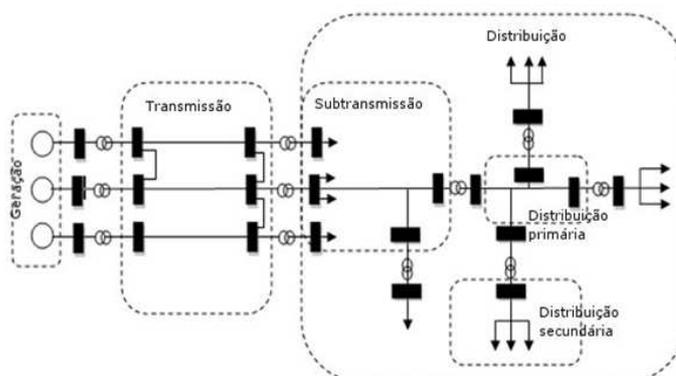


Figura 6.2 Diagrama unifilar do sistema elétrico de potência [46]

A aplicação de técnicas de mineração de dados proveniente dos SEPs pode ser considerado um processo não trivial, principalmente quando as cargas dos consumidores supridos por um sistema de potência possuem várias características que proporcionam a influência direta na qualidade do fornecimento de energia elétrica. Essas características poderão propor um critério de classificação de acordo com o perfil da carga dos consumidores no que se refere: localização geográfica, finalidade a que se destina a energia fornecida, dependência da energia elétrica, perturbações causadas pela carga ao sistema, tarifação, tensão de fornecimento. Os critérios de classificação das cargas abordando a utilização de energia são relevantes para estudos de planejamento, pois permitem identificar hábitos de consumo a partir dos instantes em que há a maior demanda e variações de tensão produzida por esses consumidores, os quais poderão ser visto com maiores detalhes no Apêndice C.

6.1.3 Medidas Elétricas

O avanço tecnológico na fabricação e uso de equipamentos elétricos e eletrônicos, tanto para a satisfação de lazer quanto para o atendimento a parques industriais, provocou pressões junto às concessionárias para o fornecimento de energia elétrica com qualidade. Essas pressões se traduzem em demandas judiciais para indenização, pelas concessionárias pelos danos causados em equipamentos ou perdas de produtos (ou mesmo de produção), reclamações junto aos poderes públicos e demais órgãos responsáveis, faz com que as concessionárias de energia elétrica se preocupem com as medidas elétricas fornecidas pelos SEPs, as quais poderão ser ajustada de acordo com as necessidades de manipulação da carga.

6.1.4 Base de Dados

Os dados disponibilizados pela Companhia Energética de Alagoas (CEAL) para mineração de dados foram coletados a partir das medições elétricas do SEPs, proveniente do sistema SCADA (*Supervisory Control and Data Aquisition*), conhecido como sistema supervisorio, o qual permite o monitoramento do processo produtivo ou da instalação física do SEP.

6.1.5 Análise preliminar da Base de Dados

6.1.5.1 Base de dados do SEP

A base de dados fornecida pela Companhia Energética de Alagoas (CEAL) provém de um SEP trifásico oriundo das variáveis de tensão, corrente, potência, potência ativa, reativa, aparente conforme Apêndice C , a partir de 180.000 (cento e oitenta mil) observações diárias no intervalo de 00 às 24 horas de acordo com as medidas elétricas das subestações de Cruz das Almas, Benedito Bentes, Tabuleiro, Trapiche, localizadas na cidade de Maceió, Alagoas, conforme a seguir:

A	B	C	D	E	F
31/01/05	23:45	393,36	389,88	395,88	0,4117
31/01/05	23:30	400,68	396,24	399,36	0,1646
31/01/05	23:15	412,68	407,76	409,92	0,2744
31/01/05	23:00	422,04	414,96	416,04	0
31/01/05	22:45	418,44	410,16	415,2	0,5067
31/01/05	22:30	425,64	418,32	423,84	0,2744
31/01/05	22:15	426,6	417,84	421,92	0,3742
31/01/05	22:00	418,44	417,24	419,52	0,4117
31/01/05	21:45	423,72	417,12	420,72	0,0549
31/01/05	21:30	420	418,56	422,04	0,4117
31/01/05	21:15	414	408,72	414,48	0,3293
31/01/05	20:45	413,28	405,72	410,64	0,2994
31/01/05	20:30	399,48	393,96	401,04	0,2744
31/01/05	20:15	400,44	391,32	396,48	0,4313

Tabela 6.1 Base de dados Originais: (A e B) período e intervalo das medições; (C,D,E)- valores de corrente das três fases (F) Fator de Potência

6.1.6 Seleção de Variáveis

O fator de potência pode ser considerado uma variável de extrema importância para critério de análise, pelo motivo que, quando existe uma melhoria do fator de potência das instalações poderá conduzir um aumento de produtividade do sistema elétrico podendo reduzir o valor das contas, redução do percentual de perdas, prevalecendo a melhor utilização dos transformadores, permitindo carga adicional sem sobrecargas, o qual poderá reduzir as quedas de ten-

são, passando a observar-se a correta tensão de suprimento de acordo com a legislação em vigor [6, 46, 75].

De acordo com ANEEL [10], o fator de potência é a razão entre a energia elétrica ativa e a raiz quadrada da soma dos quadrados das energias elétrica ativa e reativa, consumida num mesmo período. Para efeito de faturamento, deverá ser verificado pela concessionária por meio de medição apropriada observando os critérios de consumo de energia elétrica e demanda de potência reativas excedentes: quando o fator de potência da unidade consumidora, indutivo ou capacitivo, for inferior a 0,92 (noventa e dois centésimos), deverá ser realizada correções de acordo com a legislação [70]. Entretanto, à análise da carga diária em termos de intervalo de tempo pode ser considerado relevante no que se refere a busca do perfil diferenciado da carga em cada subestação.

Considerando a carga diária da base de dados e intervalo de horários em função dos patamares de carga de acordo com horários estabelecidos pelo ONS [70], onde a carga varia continuamente apresentando um comportamento cíclico, variando entre as condições de carga leve, passando por um período de carga média até alcançar a carga pesada ocorrerá novamente o período de carga média, terminando o ciclo da carga diária, conforme tabela 6.2, faz-se necessário analisar variáveis específicas em termo de qualidade de energia de acordo com a classificação das subestações citadas no parágrafo anterior.

	1	2	3	4
Pesada	18 às 21h		19 às 22h	
Media	07 às 18h	17 às 22h	07 às 19h	18 às 23 h
	21 às 24h		22 às 24h	
Leve	00 às 07h	00 às 17h	00 às 07h	00 às 18 h
		21 às 24h		23 às 24 h

Tabela 6.2 Intervalos de horários dos patamares de carga: (1)segunda/ sábado e (2) domingo/feriados (sem horário de verão); (3)segunda/ sábado e (4) domingo/feriados(com horário de verão) [70]

6.1.7 Divisão dos Dados

Os dados utilizados nas simulações corresponde aos valores diários da carga entre as subestações citadas no item 6.1.4, proveniente de valores do fator de potência do primeiro trimestre de 2005. A base de dados representada pelo conjunto $X = [x_1, x_2, x_3, \dots, x_n]$ refere-se a subestação de Cruz das Almas, onde x_1 denota o primeiro elemento da matriz de dados, x_2 o segundo e assim por diante, calculando-se o intervalo médio diário da carga a partir de:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (6.1)$$

A tabela 6.3, apresenta a descrição dos rótulos na base de dados do mapa de Kohonen com objetivo em proporcionar uma melhor identificação e análise das classes na formação do mapa contextual.

rotulos	subestação	intervaloes	tipo de carga
Cza21a24	Cruz das Almas	21 as 24 h	Media
Cza18a21		18 as 21 h	Pesada
Cza7a18		07 as 18 h	Media
Cza0a07		00 as 07 h	Leve
Bbe21a24	Benedito Bentes	21 as 24 h	Media
Bbe18a21		18 as 21 h	Pesada
Bbe7a18		07 as 18 h	Media
Bbe0a07		00 as 07 h	Leve
Tbm21a24	Tabuleiro	21 as 24 h	Media
Tbm18a21		18 as 21 h	Pesada
Tbm7a18		07 as 18 h	Media
Tbm0a07		00 as 07 h	Leve
Tbd21a24	Trapiche	21 as 24 h	Media
Tbd18a21		18 as 21 h	Pesada
Tbd7a18		07 as 18 h	Media
Tbd0a07		00 as 07 h	Leve

Tabela 6.3 Rotulação dos dados no Mapa de Kohonen

A classificação realizada pelo mapa de Kohonen exposta na figura 6.3, permite visualizar as similiariedades das subestações de acordo com as nove classes representando o conjunto de todas as subestações.

As nove classes apresentadas no mapa contextual proporciona a classificação das subestações de acordo com os níveis de similaridade dos sinais coletados. As classes II e III não possuem características semelhantes, uma vez que, as oscilações no intervalo de carga nos horários de pico são distintos, principalmente quando a carga encontra-se na fase de transição de carga pesada para um novo ciclo de carga média. Nas classes II e VI (subestações cza21as24 e tbd18as21), V e IV (tbd7as18 e tbm18as21) possuem proximidade entre as classes de acordo características semelhantes entre si a partir da classificação dos rótulos de dados no intervalo de carga pesada representados em classes distintas.

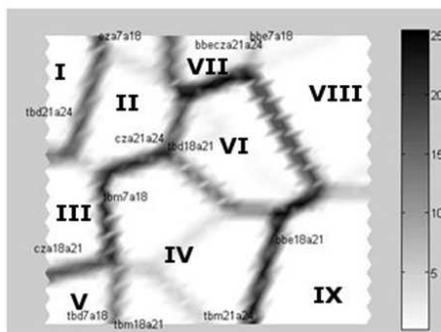


Figura 6.3 Mapa contextual dos intervalos de carga por subestação

No caso da classe VII (bbe7as18 e bbe21as24) a similaridade entre o fator de potência da mesma subestação em horários de cargas distintos são visíveis, devido a proximidade dos rótulos. As regiões mais claras proveniente dos *clusters* II e IV apresentam as possíveis divisões e formação de novos agrupamentos de acordo com o intervalo da separação entre as classes. Nas classes VIII e VII a proximidade dos rótulos entre as classes exibe a similaridade entre os sinais (cza21às24 - carga média) e bbe7as18 quando ocorre a mudança de ciclo entre cargas. Finalizando, a classe IX apresenta a movimentação do rotulo bbe18as21 (carga pesada) similar a tbd21as24 a partir divisão entre as classes. As similaridades entre as cargas constitui-se a proximidade dos valores do fator de potência nos intervalos das cargas entre as subestações.

Neste contexto, a classificação realizada na figura 6.3, não permite avaliar se o fator de potência da referida subestação é transformada integralmente em outras formas de energia dentro o intervalo de 0 a 1 (Apêndice C).

A configuração básica para a projeção do mapa exposto na figura 6.3, baseia-se na utilização do conjunto de 40x40 neurônios, vizinhança hexagonal e dados normalizados, utilizando intervalo de 30 épocas (ver maiores detalhes capítulo 3), reduzindo o estágio de épocas na primeira fase de $3 \rightarrow 1$ e aumentando o refinamento do mapa na segunda fase de $1 \rightarrow 1$ de acordo com a literatura. Os melhores resultados foram obtidos a partir do erro de quantização (EQ) = 0.2236 e erro topográfico (ET) = 0.7575.

A interpretação subjetiva do mapa contextual de Kohonen não permite avaliar de forma adequada os intervalos entre as classes, uma vez que, a rotulação serve apenas para identificar de forma mais clara o posicionamento dos dados entre as classes. Entretanto, existe a possibilidade de não rotular a base de dados como forma de visualizar os valores do conjunto de dados, porém, neste modelo à análise de agrupamentos é prejudicada devido a dificuldade da identificação e classificação das subestações. Outra limitação pode ser vista no que se refere a rotulação adequada para a rede processar a massa de dados, uma vez que, se a rotulação for escolhida de forma indevida poderá dificultar a visualização e interpretação do mapa.

À análise de agrupamentos apresentada pelo mapa de Kohonen não permite buscar um perfil efetivo do limite das classes, ou quando um objeto que pertença uma classe possa migrar de uma classe para outra. Essa característica possui relevância ao processo de análise do fator de potência, uma vez que, um fator de potência inferior a 0,92 permite que as concessionárias fornecedoras de energia possam tomar medidas cabíveis para sua correção de acordo com o

critério de fornecimento de qualidade de energia de acordo com a legislação em vigor. A projeção das nove classes expostas na figura 6.3, permite o acréscimo de variáveis de entrada no controlador *fuzzy*, fato significativo para o aumento do conjunto de regras adotadas pelo métodos Mandani [58, 98]. Considerando que, o conjunto de regras controlador se forma a partir da quantidade de variáveis de entrada combinadas com as variáveis de saída, ou seja, se tivermos um conjunto de entradas muito grande poderá dificultar a formação do mapa de regras. Assim, uma alternativa de minimizar o problema dos rótulos é classificar e analisar as subestações de acordo com o horário dos intervalos de cargas similares entre as subestações. Onde a matrix de dados X exposta pelo mapa de Kohonen (figura 6.3) é transformada para $X^T = A$, maiores detalhes sobre transposição poderá ser visto em Loan [26].

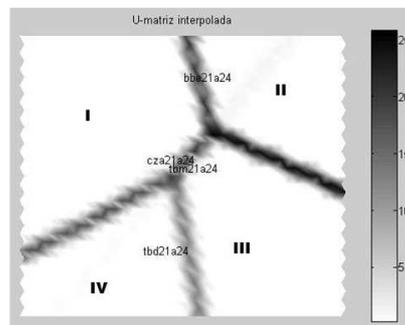


Figura 6.4 Mapa contextual por tipo de carga de acordo com as subestações

Dimensão	Epocas(A)	Raio(A)	Epocas(B)	Raio(B)	EQ	ET
30X35	30	5 → 1	20	1 → 1	0.00	0.75
40X40	20	3 → 1	20	1 → 1	0.00	1.00
21X25	40	3 → 1	30	1 → 1	0.00	0.50
55X55	40	5 → 1	30	1 → 1	0.00	1.00
55X55	3	3 → 1	30	1 → 1	0.00	1.00
24X25	5	5 → 1	40	1 → 1	0.00	1.00
24X25	30	3 → 1	40	1 → 1	0.00	1.00
10X15	5	4 → 1	25	1 → 1	0.00	0.50
34X21	4	3 → 1	25	1 → 1	0.00	1.00
12X12	10	4 → 1	25	1 → 1	0.00	1.00
56X56	10	5 → 1	25	1 → 1	0.00	0.00
5X5	5	5 → 1	25	1 → 1	0.10	0.00
15X15	3	5 → 1	23	1 → 1	0.00	0,50
35X36	15	3 → 1	60	1 → 1	0.00	0.75

Tabela 6.4 Rotulação dos dados no Mapa de Kohonen: (A) refere-se ao mapeamento grosso do mapa considerado fase 1 e (B) refinamento do mapa na fase 2

De acordo com a figura 6.4, utilizando apenas um conjunto de quatro rótulos e variáveis provenientes de intervalos de carga por subestação, observa-se que as classes I (cza21a24) e

III (tbm21a24), possui proximidade relacionadas pelas características dos sinais em *clusters* distintos, as demais classes II e IV não possui nenhum tipo de similaridade entre os sinais de fator de potência. Para projeção do mapa exposto na figura 6.4, utilizou-se um conjunto de neurônios com dimensão 35×36 , vizinhança hexagonal, dados normalizados, 15 épocas na primeira fase de $3 \rightarrow 1$ para um mapeamento mais grosso do mapa e no seu refinamento utilizou-se 60 épocas com vizinhança fixa de $1 \rightarrow 1$, observou-se que o erro de quantização $QE = 0.000$ e $ET = 0.75$.

A tabela 6.4, exibe as configurações básicas para a rede SOM, observa-se que na maioria dos casos EQ e ET $\rightarrow 0$, autores como Kaski [48], Kiviluoto [51], Vesanto [91] e Zuchine [100], discutem a possibilidade da rede entrar no processo de sobre-ajuste, ou seja, quando a rede neural perde a sua capacidade de generalização, fato este, que não é comprovado matematicamente por esses autores, principalmente no que se refere ao intervalo e no tipo de valores de dados no qual poderá ocorrer este processo. Considerando que, uma maior projeção do mapa de Kohonen a partir da definição do número de neurônios do arranjo possibilita a exposição de todos conjuntos de dados, fato comprovado a partir dos conjuntos dos rótulos expostos no mapa contextual, principalmente quando o conjunto de dados possui valores próximo a zero, favorecendo esta tendência de quando $QE \rightarrow 0$, no caso dos sinais de fator de potência, os quais encontra-se nos intervalos entre $[0, 1]$.

6.1.8 O Processo de Fuzzificação da Matriz de Dados pelo Método Mandani

A matriz de dados A , constitui-se o conjunto de dados de entrada do controlador *fuzzy*, definida por uma variável lingüística (V, X, T_v) , onde V é uma variável definida sobre um conjunto de referência X . O conjunto $T_v = [A_1, A_2, \dots]$ finito ou infinito, contém os subconjuntos "fuzzy" normalizados de X que utilizados para caracterizar V . Onde: A_i está associado aos seus valores lingüísticos ("baixo", "médio" e "alto"). Por exemplo, dentro de um intervalo $[0.5, 1]$, pode-se definir uma variável lingüística "Cza21as24CMedia", correspondente ao intervalo de carga da subestação de Cruz das Almas abordando os valores de fator de potência no horário de carga média de 21 às 24 h. Esta variável terá três valores lingüísticos: "baixa", "média" e "alta", caracterizados como subconjuntos *fuzzy* cujas funções de pertinência são representadas por $T_{Cza21as24Media} = \{baixa, media, alta\}$. Para uma carga menor que 0.65 ela será considerada "baixa" com valor de pertinência igual a 1.0, e "média" com valor de pertinência igual a 0.0 e assim, sucessivamente para cada subestação.

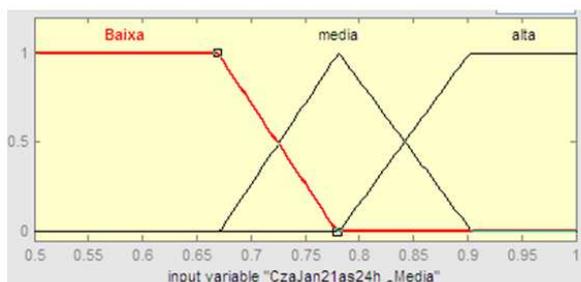


Figura 6.5 Funções de pertinência da variável lingüística de entrada Cza21as24CMedia.

Variáveis Linguísticas Entradas(Subestações)	Função Trapezoidal Baixo	Função Triangular Medio	Função Trapezoidal Alto
Cza21as24	0.2 0.4 0.6 0.7	0.6 0.7 0.9	0.7 0.9 1.0 1.1
Bbe21as24	0.2 0.4 0.6 0.7	0.6 0.7 0.9	0.7 0.9 1.0 1.1
Tbm21as24	0.2 0.4 0.6 0.7	0.6 0.7 0.9	0.7 0.9 1.0 1.1
Tbd21as24	0.2 0.4 0.6 0.7	0.6 0.7 0.9	0.7 0.9 1.0 1.1

Saída (Fator de Potência)

Fator de Potência	0.3 0.4 0.6 0.7	0.6 0.7 0.9	0.7 0.9 1.0 1.1
-------------------	-----------------	-------------	-----------------

Tabela 6.5 Funções de pertinência das subestações :Cruz das Almas, Benedito Bentes, Tabuleiro, Trapiche das cargas residenciais.

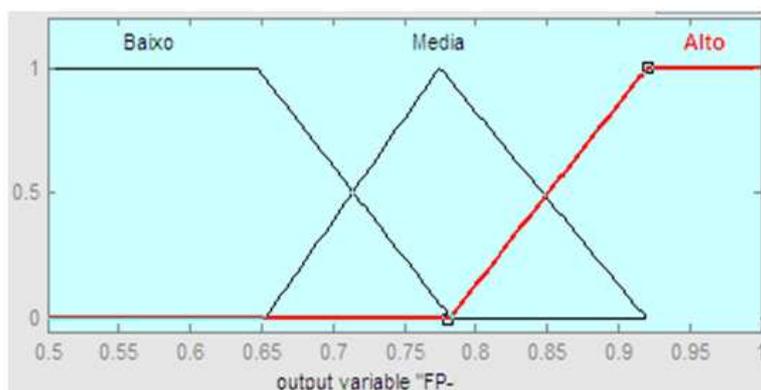


Figura 6.6 Variável linguística de saída Fator de Potência

Na formação da base de regras, um número finito de regras que pode ser estabelecido depois que se determinar o número de entradas e saídas necessários ao sistema de controle, principalmente no que se refere a escolha dos predicados linguísticos (número, distribuição e forma das funções de pertinência) para cada uma destas variáveis. Por exemplo, se o sistema de controle *fuzzy* (variáveis das subestações e fator de potência), possui duas entradas com 3 predicados cada, pode-se definir no máximo $3^4 = 81$ possíveis regras. Uma maneira conveniente para listar todas as possíveis regras, para os casos onde não há muitas entradas para o controlador *fuzzy* usando a tabela para representação das regras, como mostrado na tabela 6.6. As variáveis de entrada "Intervalo de Carga Cza de 21 às 24 h (Média) = variável 1 (V_1)" e "Intervalo de Carga Bbe de 21 às 24 h (Media) = variável 2 (V_2)", "Intervalo de Carga Tbm de 21 às 24 h (Media) = variável 3 (V_3)", "Intervalo de Carga Tbd de 21 às 24 h (Media) = variável 4 (V_4)", são os valores de entrada (premissa das regras) e o "Fator de Potência = variável 5 (V_5)" é o conseqüente da regra:

A construção da tabela pode ser útil para verificar certos detalhes do funcionamento do controlador, para se eliminar predicados redundantes, verificar a consistência das regras, etc.

Entretanto, a construção da superfície de controle gerada pela base de regras é importante para se entender o comportamento do controlador. O mapa (ou tabela) de regras *fuzzy* é idêntico a relação *fuzzy*, onde um produto cartesiano contém, como seus elementos, todos os pares ordenados, onde as entradas na matriz de relação *fuzzy* são valores de pertinência no intervalo [0,1], enquanto as do mapa de regras *fuzzy* são funções de pertinência definidas (B (baixo), M (médio) e A(alto)) [60]. As entradas da matriz de regra são preenchidas durante a identificação do sistema *fuzzy*, quando tipicamente um operador humano auxilia na identificação da operação e controle do processo de acordo com o método Mamdani [58].

Descrição do Conjunto de Regras
1. Se V_1 é B) e (V_2 é B) e V_3 é B) e V_4 é B) então V_5 é B)
2. Se V_1 é B) e (V_2 é M) e V_3 é M) e V_4 é B) então V_5 é B)
3. Se V_1 é B) e (V_2 é A) e V_3 é A) e V_4 é A) então V_5 é A)
4. Se V_1 é M) e (V_2 é B) e V_3 é B) e V_4 é B) então V_5 é B)
5. Se V_1 é A) e (V_2 é B) e V_3 é B) e V_4 é B) então V_5 é B)
6. Se V_1 é M) e (V_2 é M) e V_3 é M) e V_4 é M) então V_5 é M)
7. Se V_1 é A) e (V_2 é M) e V_3 é M) e V_4 é M) então V_5 é M)
8. Se V_1 é M) e (V_2 é A) e V_3 é A) e V_4 é A) então V_5 é A)
9. Se V_1 é A) e (V_2 é A) e V_3 é A) e V_4 é A) então V_5 é A)
10. Se V_2 é B) e (V_3 é B) e V_4 é B) então V_5 é B)
11. Se V_2 é B) e (V_3 é M) e V_4 é M) então (V_5 é M)
12. Se V_2 é B) e V_3 é A) e V_4 é A) então V_5 é A)
13. Se V_2 é M) e V_3 é B) e V_4 é B) então V_5 é B)
14. Se V_2 é M) e V_3 é M) e V_4 é M) então V_5 é M)
15. Se V_2 é M) e V_3 é A) e V_4 é A) então V_5 é A)
16. Se V_2 é A) e V_3 é M) e V_4 é B) então V_5 é B)
17. Se V_2 é A) e V_3 é M) e V_4 é M) então V_5 é M)
18. Se V_2 é A) e V_3 is A) e V_4 is A) então V_5 é A)
19. Se V_3 é B) e V_4 é B) então V_5 é B)
20. Se V_3 é B) e V_4 é M) então V_5 é M)
21. Se V_3 é B) e V_4 é A) então V_5 é A)
22. Se V_3 é M) e V_4 é B) então V_5 é B)
23. Se V_3 é M) e V_4 é M) então V_5 é M)
24. Se V_3 é M) e V_4 é A) então V_5 é A)
25. Se V_3 é A) e V_4 é B) então V_5 é B)
26. Se V_3 é A) e V_4 é M) então V_5 é M)
27. Se V_3 é A) e V_4 é A) então V_5 é A)

Tabela 6.6 Descrição do mapa de regras do controlador *fuzzy* de forma reduzida pelo método "Modus Ponens generalizado"[13]

O mapa de regras constitui-se de uma implicação *fuzzy* matricial, onde as entradas implicam em determinadas saída, ou similarmente saídas estão associadas com determinadas entradas, a qual pode ser entendida como uma superfície de controle, tridimensional [60] e [82].

Na figura 6.8, O eixo vertical é a variável de saída Fator de Potência enquanto que em cada eixo horizontal estão as subestações analisadas, a partir do gráfico de superfície de controle é possível perceber como o controlador se comporta para determinadas entradas.

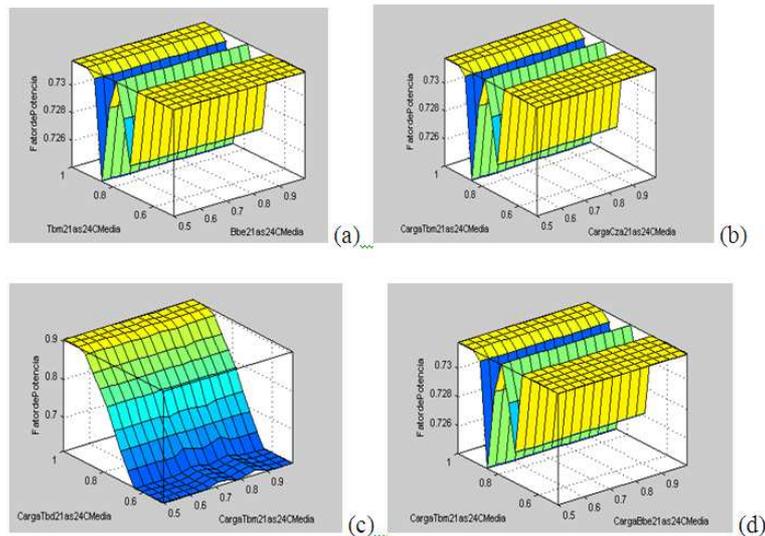


Figura 6.7 Gráfico de superfície das subestações no intervalo de carga de 21 às 24h (carga média)

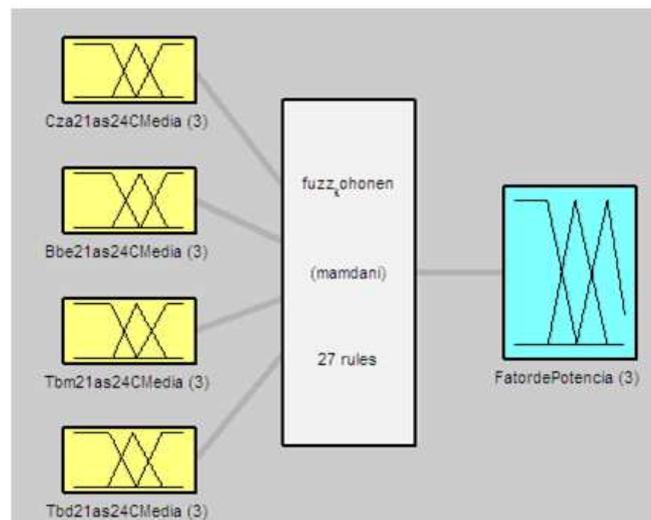


Figura 6.8 Variáveis lingüísticas de entrada e saída do controlador para o intervalo de carga de 21 às 24 horas

A modelagem do controlador *fuzzy* é formado pela composição de 4(quatro) variáveis de entrada, proveniente do intervalo dos horários das cargas (24 às 21 h) das subestações de Cruz das Almas, Benedito Bentes, Tabuleiro e Trapiche. A saída *crisp* do sistema é apresentado na figura 6.10, adota padrões inferiores ao estabelecido pelos órgãos reguladores ANEEL e ONS.

Entretanto, estes valores permite avaliar a saída do sistema de acordo com o perfil do intervalo de carga das subestações auxiliando no processo de tomada de decisão no que se refere a correção do fator de potência visando a qualidade dos serviços prestados no fornecimento de energia.

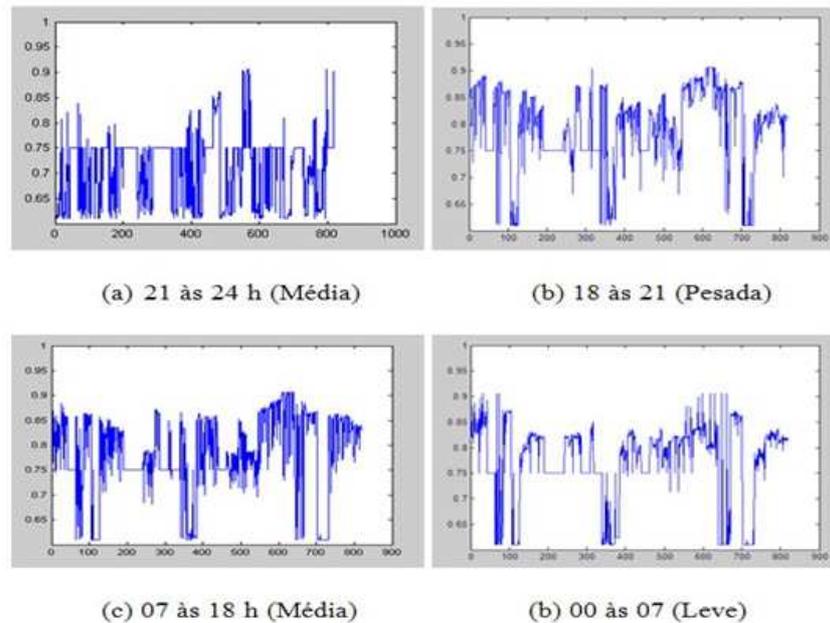


Figura 6.9 Variabilidade do fator de potência de acordo com o perfil da carga de todas as subestações - Método Modus Pones Generalizado

A figura 6.9, apresenta a saída *crisp* do sistema nebuloso, observa-se que as oscilações que ocorre na carga média (21 às 24 h - A), o fator de potência varia em média 0,75, valor inferior ao estabelecido na legislação. Esta caracterização da carga permite que se possa tomar medidas corretivas para elevar o fator de potência dentro do limite estabelecido de 0,92 de acordo com as subestações projetadas no sistema (figura 6.8). Na fase em que a carga é considera pesada (18 às 21 h - B), média (07 às 18 h - C) e Leve (00 às 07 - D), são raros os casos em que o fator de potência encontra-se abaixo de 0,65, porém, na maioria dos casos relaciona-se entre o intervalo de 0,75 a 0,89, número próximo ao valor aceitável. Por outro lado, em todos os casos o fator de potência encontra-se inferior ao padrão esperado de 0,92 estabelecido pela ANEEL e ONS.

O método do Modus Pones generalizado abordado por Saito [13] adota a construção de regras (tabela 6.6) a partir do conhecimento extraído de um especialista. Entretanto, autores como Ming [55] e Prasad [85], demonstram metodologias para redução da base de regras a partir da modelagem hierárquica *fuzzy* para sistemas que possuam mais de 3(três) variáveis utilizando

a composição de regras por acesso aleatório são mais precisos por se extrair a redundância na base de regras obtidos pelo método Modus Ponens generalizado. Por exemplo, seja o mapa de regras construído a partir do acesso aleatório dos termos lingüísticos.

	V1			V2			V3			V4		
FP	B	M	A	B	M	A	B	M	A	B	M	A
B	B	A	M	A	B	A	A	M	A	M	B	M
M	M	A	B	A	B	A	A	B	A	M	B	A
A	B	B	M	B	M	A	A	M	A	B	A	B

Tabela 6.7 Mapa de Regras por acesso aleatório

A tabela 6.7, apresenta a composição da base de regras pelo método do acesso aleatório, formando 9(nove) regras que compõem a base de inferência do controlador *fuzzy* estabelecido pelo método Mandami. As base de dados originais poderão ser visualizadas com maiores detalhes no Apêndice D.

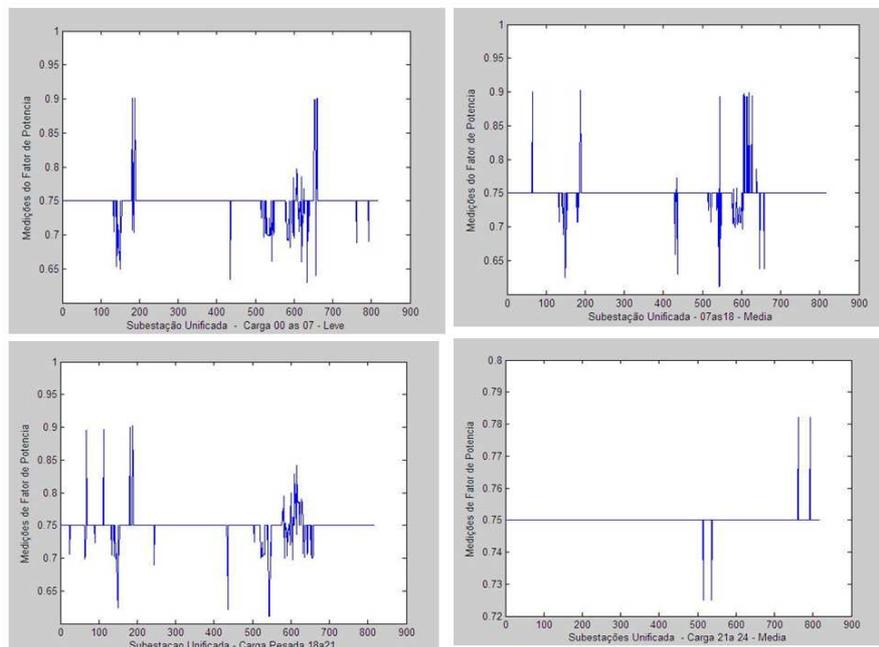


Figura 6.10 Variabilidade da carga de todas subestações após aplicação da lógica *fuzzy*

Observa-se que, a modelagem do conjunto de regras a partir do acesso aleatório permite uma saída *crisp* (figura 6.10), mas nítida em relação a saída exposta na figura 6.9 estabelecido pelo método Modus Ponens Generalizado, o qual possibilita à análise dos sinais de dados

com maiores detalhes de acordo com suas respectivas variações. Entretanto, ambos métodos de modelagem proporcionaram a identificação do Fator de Potência entre as subestações (item 6.1.4), inferior a 0,92 (padrão ANEEL e ONS). Na maioria dos casos, a variabilidade da carga manteve-se constante em 0,75. Nos demais casos, as variações da carga encontra-se nos intervalos [0,76 a 0,90], os quais poderão ser corrigidos para estabelecer um fator de potência de acordo com a legislação.

6.2 Conjunto de Dados Públicos - Análise de Sinais Biomédicos

6.2.1 Introdução

O diabetes corresponde uma das mais comuns doenças não transmissíveis em todo mundo é considerada a quinta causa de morte nos países desenvolvidos[5]. As complicações clínicas proveniente do diabetes poderão causar o aparecimento de doenças cardiovasculares, insuficiência renal, lesões graves, derrame cerebral, impotência, cegueira, úlcera nas pernas e até amputações de membros [37]. Além disso, muitos usuários portadores de diabetes possuem uma redução da esperança de vida proporcionando enormes custos aos hospitais e pode ser considerado um dos maiores problemas de saúde pública mundial e um dos mais desafiadores problemas de saúde no século XXI [5] e [27].

Segundo Mazzaferri [62], o diabetes *mellitus* é um distúrbio do metabolismo dos carboidratos. A perturbação central consiste em uma anormalidade na secreção ou efeito da insulina, ou de ambos. A deficiência de insulina pode ser relativa ou absoluta e que na grande maioria dos casos, o diabetes é genericamente determinado como distúrbios do pâncreas (órgão da cavidade abdominal que produz enzimas digestivos liberados para os intestinos e diferentes tipos de hormônios liberados diretamente para o sangue). A Associação Protetora dos Diabéticos de Portugal, define o diabetes *mellitus* como uma doença metabólica, crônica, caracterizada pelo aumento dos níveis de açúcar (glucose) no sangue, resultado de uma deficiente produção de insulina e/ou pela resistência à ação dessa insulina, o que conduz a uma deficiente capacidade de utilização pelo organismo da nossa principal fonte de energia, a glucose [1].

Geralmente, quando diagnosticada poderá durar a vida toda, provocada pela falta ou resistência à ação de insulina (hormônio produzido pelas células beta do pâncreas) ou causada pelo aumento subitico de açúcar no sangue. De acordo com o Ministério da Saúde[81], o diabetes é uma doença metabólica caracterizada por hiperglicemia e associadas a complicações, disfunções e insuficiência de vários órgãos, especialmente olhos, rins, nervos, cérebro, coração e vasos sanguíneos. Pode resultar de defeitos de secreção e/ou ação da insulina envolvendo processos patogênicos específicos, por exemplo, destruição das células beta do pâncreas (produtoras de insulina), resistência à ação da insulina, distúrbios da secreção da insulina, entre outros.

Muitos estudos abordam indícios concretos da existência uma epidemia em muitos países em desenvolvimento e países industrializados. Calcula-se que, atualmente exista em média 194 milhões de pessoas em todo mundo com diabetes representando 5,1% da população adulta e

Países	Pessoas (milhões) 2003	Pessoas (milhões) 2025
1. Índia	35,5	73,5
2. China	23,8	46,1
3. EUA	16,0	23,1
4. Rússia	9,7	10,7
5. Japão	6,7	7,1
6. Alemanha	6,3	7,1
7. Paquistão	6,2	11,6
8. Brasil	5,7	10,7
9. México	4,4	9,0
10. Egito	3,9	7,8

Tabela 6.8 Os 10 países com a maior taxa de população diabética na faixa etária de 20-79 anos e estimativa para 2025 [5]

que esse percentual aumentará para alcançar os 333 milhões de pessoas diabéticas, representando 6,3% de pessoas adultas para o ano de 2025, estima-se que em um grupo etário entre 40-59 anos existe o maior número de pessoas com diabetes e poderá aumentar até 2025 [5].

A Organização Mundial de Saúde [71], demonstra números ainda maiores, estima-se que o número de pessoas com diabetes irá duplicar ao longo dos próximos 25 anos e que poderá chegar em 366 milhões de habitantes até 2030.

Na maioria dos casos, esse aumento poderá ocorrer em 150% dos países em desenvolvimento, devido ao fato de haver mais pessoas no mundo (população em crescimento) e que haverá mais idosos (envelhecimento da população). Além disso, as tendências de urbanização, fato que, muitas pessoas estão se deslocando das zonas rurais para as cidades, principalmente nos países em desenvolvimento. Essa situação, ainda influencia nas estatísticas, as quais poderão aumentar ainda mais este índice, naquelas pessoas que poderão vir a ter diabetes [71]. Segundo à Organização Mundial de Saúde [71], pessoas que vivem em cidades de países em desenvolvimento tende a ser menos ativos fisicamente e apresentam níveis mais elevados de sobrepeso e obesidade do que as pessoas nas zonas rurais. As tendências atuais sugerem que essas projeções são conservadoras e que o aumento da prevalência do diabetes pode ser ainda maior.

Nos países em desenvolvimento, pessoas que se encontram na fase produtiva de suas vidas são particularmente afetadas pelo diabetes. Nesses países, 3/4 da população que tem menos de 65 anos têm diabetes e 25% dos adultos na faixa etária inferior a 44 anos possuem diabetes. Nos países desenvolvidos, mais da metade de toda população possui diabetes no grupo etário maior de 65 anos e apenas 8% dos adultos com diabetes tem menos de 44 anos [71].

6.2.2 Tipos de Diabetes

O diabetes pode ser classificada em dois tipos: a primeira considerada Tipo 1, geralmente ocorre na infância e atinge na maioria das vezes crianças ou jovens, considerada rara (cerca de 5- 10% do total de casos diagnosticados). Em casos raros, poderá aparecer em adultos e

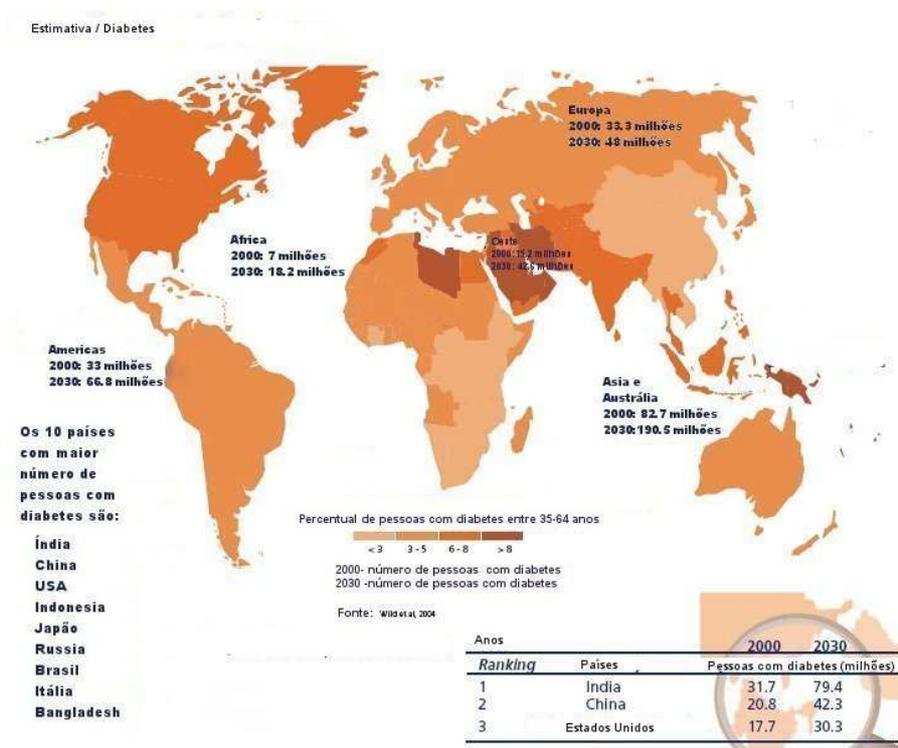


Figura 6.11 Estimativa da população com diabetes para 2030 [71]

em idosos, o diabetes Tipo 2 é a mais comum e está relacionada com a resistência à ação da insulina e com a herança genética. Estes pacientes têm, frequentemente, casos de diabetes na família, hábitos de vida e de alimentação desequilibrada e poderão sofrer de diabetes quando adultos. Os pacientes com diabetes Tipo 2 têm quase sempre peso excessivo e muitas vezes são obesos, praticam pouco exercício físico ou fisicamente inativos e consomem calorias em excesso, em proporção ao que o organismo gasta na atividade física [1, 5, 71].

Estudos têm demonstrado que muitas complicações clínicas podem surgir a partir do diabetes, as quais poderão ser prevenidas. Viver uma vida plena, incluindo uma dieta saudável, atividade física, prevenção de sobrepeso e obesidade, não fumar, são cuidados preventivos. Além disso, o controle da glicose não pode ser considerada a única forma de prevenção, mas também, o controle da pressão arterial, colesterol, triglicerídeos são fatores importantes para se ter uma vida saudável com ou sem diabetes [62]. A supervisão e controle adequados dos sinais vitais de diabetes poderão inibir o aparecimento das complicações clínicas e o paciente diabético poderá ter uma vida praticamente normal [74]

6.2.3 Estágios do diabetes

O diabetes pode-se modifica-se em pacientes individuais, não é raro diagnosticar um diabetes "definido" numa pessoa obesa, e/ou após perda de peso, verificar um retorno completo da tolerância normal à glicose. Assim, torna-se conveniente considerar diabetes genético como progredindo através de diferentes estágios, segundo uma taxa de variação [62]. A interpretação

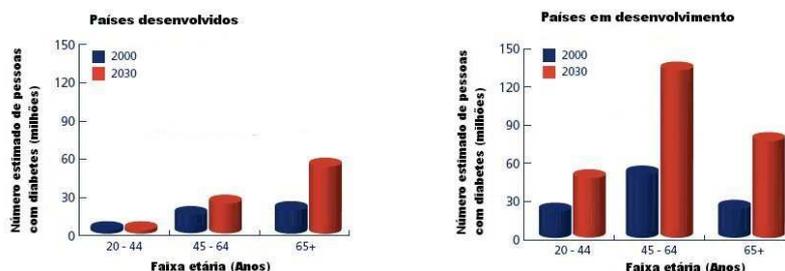


Figura 6.12 Estimativa da população por idade 2030 [71]

dos resultados dos exames diagnósticos para o diabetes e para a regulação glicêmica alterada é apresentada na tabela 6.9.

Classificação	Glicemia de Jejum(mg/dl)	Glicemia 2h após TTG-75g (mg/dl)
Hipoglicemia < 140	≤ 60 Normal	< 110
Hiperglicemia intermediária		
- Glicemia de Jejum alterada	110-125	
- Tolerância a glicose diminuída		140-199
Diabetes Mellitus	≥ 126	> 200

Tabela 6.9 Interpretação dos resultados da glicemia de jejum e do teste de tolerância à glicose [81]

A tabela 6.9, apresenta a interpretação dos resultados da glicemia de jejum com a classificação do estágio clínico de diabetes, tais como:

- Glicemia: quantidade de glicose (níveis de açúcar) existente no sangue [37];
- Hipoglicemia: diminuição dos níveis glicêmicos com ou sem sintomas para valores abaixo de 60 a 70 mg/dl (miligrama por decilitro de sangue), os principais sintomas são : fome, tontura, fraqueza, dor de cabeça, confusão, coma, convulsão e as manifestações de taquicardia, apreensão e tremor [81];
- Hiperglicemia: valor elevado de açúcar no sangue [62];
- Hiperglicemia intermediária: apresentam alto risco para o desenvolvimento do diabetes. São também fatores de risco para doenças cardiovasculares [81].

De acordo com o Ministério da Saúde [81], os testes laboratoriais mais comumente utilizados para suspeita de diabetes ou regulação glicêmica alterada são: glicemia de jejum (nível de glicose sanguínea após um jejum de 8 a 12 horas), teste oral de tolerância à glicose TTG-75g (o paciente recebe uma carga de 75 g de glicose, em jejum, e a glicemia é medida antes e 120 minutos após a ingestão) e glicemia casual (tomada sem padronização do tempo desde a última refeição). Assim, quando a glicemia de jejum situa-se entre entre 110 e 125 mg/dl (glicemia de

jejum alterada), por apresentarem alta probabilidade de ter diabetes, podem requerer avaliação por TTG-75g em 2h. Mesmo quando a glicemia de jejum for normal (< 110 mg/dl), pacientes com alto risco para diabetes ou doença cardiovascular poderão ser submetidos à avaliação por TTG considerados normais, mas não estão suficientemente elevados para caracterizar um diagnóstico de diabetes, os indivíduos são classificados como portadores de hiperglicemia intermediária. Como apresentado na tabela 6.8, quando a glicemia de jejum estiver entre 110-125 mg/dl, a classificação será de glicemia de jejum alterada; quando a glicemia de 2h no TTG-75g estiver entre 140-199 mg/ dl, a classificação será de tolerância à glicose diminuída [81].

6.2.4 Complicações Clínicas

O diagnóstico precoce, controle metabólico (boa compensação) e a vigilância periódica são as principais armas para prevenir ou atrasar o início e a evolução das complicações relacionadas com a diabetes, as quais poderão ser classificadas a seguir:

- Retinopatia Diabética: lesões da retina;
- Nefropatia Diabética: doença renal;
- Neuropatia Diabética: lesões nos nervos;
- Arteriopatia: dor na barriga da perna durante a marcha;
- Macroangiopatia: doença coronária, cerebral e dos membros inferiores;
- Hipertensão-arterial: aumento da pressão arterial;
- Disfunção sexual: problemas de ereção;

- Infecções : caso os níveis de açúcar no sangue não estejam bem controlados, estão mais susceptíveis a infecções no geral e em particular às infecções da boca e das gengivas, às infecções urinárias, dos pés e ainda às infecções de cicatrizes depois de cirurgias.

6.2.5 Técnicas de Mineração de Dados para Análise de Sinais Biomédicos

As técnicas de mineração de dados citadas no capítulo 2 e 3 são bastante utilizadas na classificação e análise de sinais biomédicos. Obras recentes abordam análise de componentes principais (PCA) com redes SOM para remoção da onda P nos sinais de eletrocardiograma (ECG) [4], classificação de sinais de ECG utilizando redes neurais e PCA [69] e reconhecimento de padrões de sinais de eletroencefalograma (EEG) utilizando redes neurais [65]. Além disso, outras propostas utilizando lógica nebulosa poderão ser vista na literatura, as mais recentes encontram-se abordando a lógica nebulosa para análise de sinais de sinais biomédicos, principalmente no que se refere ao controle de glucose. Em Grant [27] identifica uma nova abordagem *fuzzy* para controle de insulina. Em Lascioa [54] adota metodologia para análise de neuropatia diabética baseada em lógica *fuzzy*. Em Man [59], apresenta modelo de sistema para simulação e controle de insulina. Em Dua [20] aborda modelos baseado em programação paramétrica para controle de glucose proveniente de diabetes Tipo 1. Em Owens [72] apresenta sistema de controle de glucose para pacientes portadores de diabetes Tipo 1. A mineração de dados proveniente de medições de glicemia em pacientes portadores de diabetes pode ser considerado um processo não trivial, devido a combinação de variáveis (idade, sexo, cor, profissão,

índice de massa corporal, valor de glicose, colesterol, triglicerídeos, pressão arterial) dos pacientes os quais poderão influenciar no diagnóstico da doença. Este estudo de caso tem como objetivo adotar o sistema proposto nesta dissertação para classificar pacientes portadores de diabetes utilizando mapas de Kohonen e identificar o estágio de diabetes correspondente a curva glicêmica desses pacientes a partir da lógica nebulosa utilizando base de dados pública.

6.2.6 Análise preliminar da Base de Dados

6.2.6.1 Base de dados Diabetes

Os dados utilizados nesta seção foram obtidos do repositório de dados públicos da Universidade da Califórnia do departamento de ciência de computação [67], provêm de medições diárias de testes de tolerância à glucose a partir de medições classificadas tais como: pré-almoço (07:00 às 9:00), Pós-almoço (11:00 às 13:00), Pré-ceia (15:00 às 17:00) e Pós-ceia (19:00 às 21:00), correspondente a amostra de 786 observações.

Pré-almoço	Pós-almoço	Pré-ceia	Pós-ceia
100	192	304	156
216	88	115	119
257	60	125	100
239	81	47	138
67	162	170	89
77	148	176	134
259	220	146	125
109	64	205	75
128	125	325	95
179	187	57	80
86	255	84	94
147	103	206	50
305	105	94	116
133	173	174	182
183	146	253	159
91	202	286	56
121	201	141	0

Tabela 6.10 Amostra da base de dados Originais

A matriz de dados exposta na tabela 6.10, representa os valores originais dos dados por intervalos de medições das taxas de glicose. Os rótulos estabelecidos no mapa de Kohonen (Apêndice E), obedece a mesma classificação do cabeçalho das colunas para que os mesmos possam ser visualizado e identificados no mapa auto-organizável de Kohonen.

A classificação realizada pelo mapa de Kohonen exposta na figura 6.13, permite a caracterização das similaridades dos sinais de glicemia de acordo com os horários estabelecidos no teste de tolerância a glicose (item 6.2.6.1). Inicialmente, observa-se que o mapa contextual

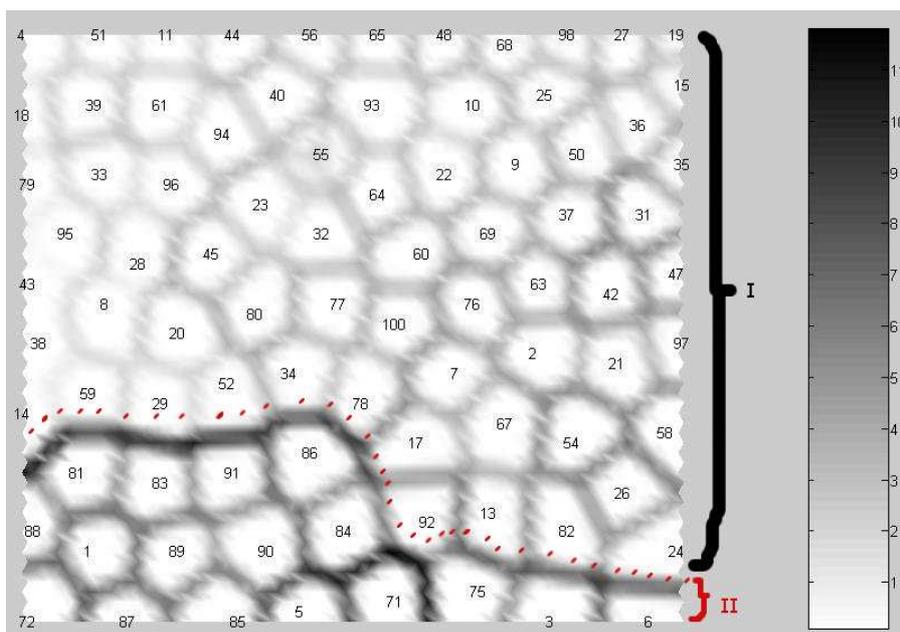


Figura 6.13 Representação dos Sinais de Glicemia no Mapa de Kohonen

demonstra a representação de duas grandes classes (I e II). No corpo central dessas classes concentra-se subclasses menores de acordo com a especificação de cada sinal glicêmico. Por exemplo, na classe I, as divisões mais claras representam sinais com características similares formando pequenas subclasses individualizadas, as quais poderão ser exemplificadas conforme figura 6.14, demonstrando que o sinal 77 proveniente das medições de glicemia estabelecidas no pré-almoço 07:00 às 9:00, possui características similares ao sinal 100 (pré-lanche 22:00), por pertencerem a mesma subclasses individualizadas e caracterizadas pela divisão originadas entre as subclasses selecionadas pela cor cinza claro.

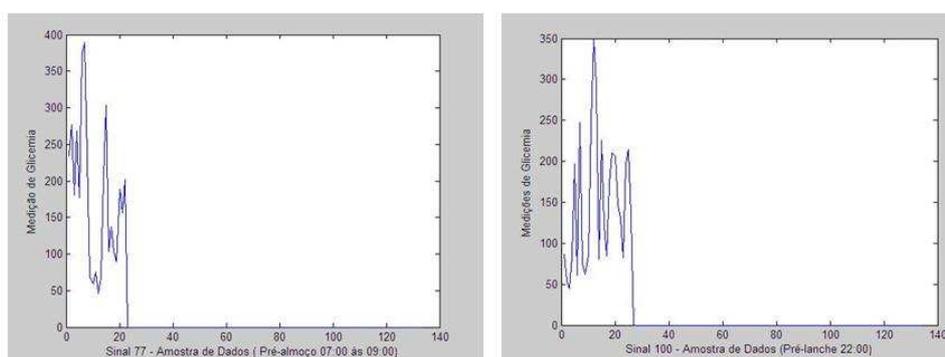


Figura 6.14 Similaridade dos Sinais de Glicemia no mapa de Kohonen - Classe I - Sinal 77 (pré-almoço 07:00 às 9:00) e Sinal 100 (pré-lanche 22:00)

Na figura 6.15, observa-se que o sinal 20 (classe I) proveniente de medições de pós-ceia entre os horários de 19:00 às 21:00 possui características distintas em relação ao sinal 90 (classe

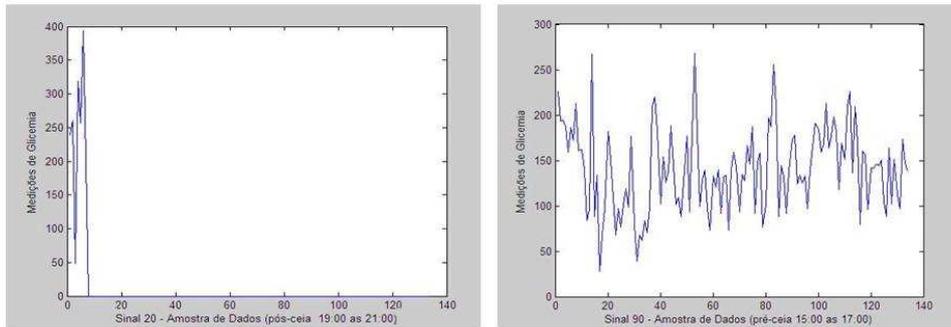


Figura 6.15 Sinais de Glicemia no mapa de Kohonen - Classe I - Sinal 20 (pós-ceia 19:00 às 21:00) e Classe II - Sinal 90 (pré-ceia 15:00 às 17:00)

II) obtido pelas medições de glicemia de pré-ceia 15:00 às 17:00.

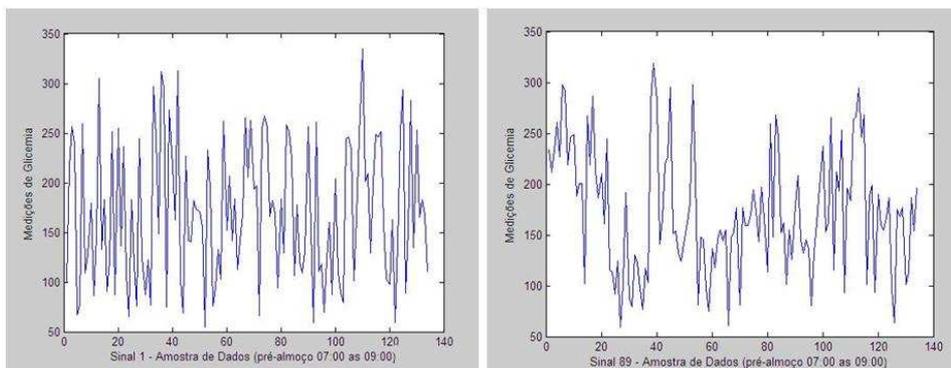


Figura 6.16 Similaridade dos Sinais de Glicemia no mapa de Kohonen - Classe II - Sinal 1 (pré-almoço 07:00 às 09:00) e Classe II - Sinal 89 (pré-almoço 07:00 às 09:00)

Na figura 6.16, a classificação realizada pelo mapa de Kohonen demonstra que o sinal 1 (classe II) proveniente de medições de pré-almoço 07:00 às 09:00, possui características similares ao sinal Sinal 89 (classe II -pré-almoço 07:00 às 09:00), porém, estão separados por subclasses a partir da divisão cinza escuro.

As configurações do mapa foram implementadas a partir das características expostas na tabela 6.11. Para uma projeção mais rápida do mapa utilizou-se uma vizinhança maior de atualização regredindo gradativamente até que possa chegar a 1. Isto permite, uma modelagem mais grosseira do mapa. Na fase de refinamento e projeção mais lenta do mapa, as épocas foram fixadas com vizinhança em 1 [52]. Para projeção do mapa exposto na figura 6.13, utilizou-se um conjunto de neurônios com dimensão 60×60 , vizinhança hexagonal, dados normalizados, 55 épocas na primeira fase de $3 \rightarrow 1$ para um mapeamento mais grosso do mapa e no seu refinamento utilizou-se 35 épocas com vizinhança fixa de $1 \rightarrow 1$, observou-se que o erro de quantização $QE = 0.000$ e $ET = 0.11$.

A tabela 6.11, exibe as configurações básicas para a rede SOM, observa-se que na maioria dos casos EQ e ET $\rightarrow 0$. Considerando as observações expostas no item (6.1.6) adota-se abordagem relacionada quando o conjunto de dados possui valores próximo a zero $QE \rightarrow 0$, no caso dos sinais proveniente de medições de glicemia, especificamente naquelas medições de hipoglicemia, as quais encontram-se próximo a zero.

Dimensão	Epocas(A)	Raio(A)	Epocas(B)	Raio(B)	EQ	ET
30X35	30	5 \rightarrow 1	20	1 \rightarrow 1	0.20	0.00
40X40	20	3 \rightarrow 1	20	1 \rightarrow 1	0.27	1.00
21X25	40	3 \rightarrow 1	30	1 \rightarrow 1	1.02	0.00
55X55	40	5 \rightarrow 1	30	1 \rightarrow 1	0.00	0.29
55X55	3	3 \rightarrow 1	30	1 \rightarrow 1	0.00	0.09
24X25	5	5 \rightarrow 1	40	1 \rightarrow 1	0.950	0.00
24X25	30	3 \rightarrow 1	40	1 \rightarrow 1	0.97	0.00
10X15	5	4 \rightarrow 1	25	1 \rightarrow 1	3.39	0.00
34X21	4	3 \rightarrow 1	25	1 \rightarrow 1	0.53	0.00
12X12	10	4 \rightarrow 1	25	1 \rightarrow 1	3.58	0.00
56X56	10	3 \rightarrow 1	25	1 \rightarrow 1	0.00	0.02
5X5	5	5 \rightarrow 1	25	1 \rightarrow 1	5.12	0.00
15X15	3	5 \rightarrow 1	23	1 \rightarrow 1	2.74	0.00
60X60	60	4 \rightarrow 1	35	1 \rightarrow 1	0.00	0.22
60X60	55	3 \rightarrow 1	35	1 \rightarrow 1	0.00	0.11
35X36	15	3 \rightarrow 1	60	1 \rightarrow 1	0.10	0.00

Tabela 6.11 Rotulação dos dados no Mapa de Kohonen

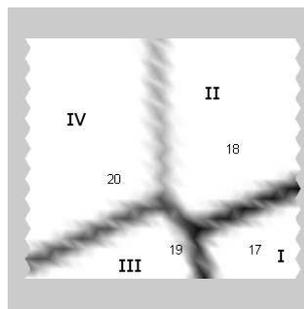


Figura 6.17 Mapa contextual por paciente

A projeção das duas maiores classes (I e II) e suas 98 subclasses expostas na figura 6.14, permite o acréscimo de variáveis de entrada no controlador *fuzzy* de forma proibitiva, o qual poderá aumentar o número de regras adotados pelo método Mandani. Como critério de minimizar o processo de composição de regras, optou-se por analisar o perfil do paciente, buscando verificar as oscilações provenientes das medições de glicemia a partir das quatro classes expostas na figura 6.18 de acordo com os intervalos dos sinais: Classe I (sinal 17 (pré-almoço

(07:00 às 09:00)), Classe II - sinal 18 (Pós-almoço (11:00 às 13:00)], Classe III - sinal 19 (Pré-ceilia (15:00 às 17:00)) e Classe IV - sinal 20 (Pós-ceilia (19 às 21:00)), utilizando um mapa contextual de dimensão 25x25, vizinhança hexagonal, dados normalizados, 30 épocas na primeira fase de $3 \rightarrow 1$ para um mapeamento mais grosso do mapa e no seu refinamento utilizou-se 15 épocas com vizinhança fixa de $1 \rightarrow 1$, observou-se que o erro de quantização $QE = 0.000$ e $ET = 0.80$.

6.2.7 O Processo de Fuzzificação da Matriz

A matriz de dados B , constitui-se o conjunto de dados de entrada do controlador *fuzzy*, definida por uma variável lingüística (W, Z, T_w) , onde W é uma variável definida sobre um conjunto de referência Z . O conjunto $T_w = [B_1, B_2, \dots]$ finito ou infinito, contém os subconjuntos "fuzzy" normalizados de Z que utilizados para caracterizar W . Onde: B_i está associado aos seus valores lingüísticos ("GlicemiaMinima", "GlicemiaNormal" e "GlicemiaAlterada"). Por exemplo, dentro de um intervalo [0 300], pode-se definir uma variável lingüística "Pré-almoço", correspondente as medidas glicêmicas no intervalo de (07:00 às 9:00) da manhã. Esta variável terá três valores lingüísticos: "GlicemiaMinima", "GlicemiaNormal" e "GlicemiaAlterada", caracterizados como subconjuntos *fuzzy* cujas funções de pertinência são representadas por $T_{prealmoco}$. Para uma medida de glicemia menor que 60 mg/dl será considerada "GlicemiaMinima" com valor de pertinência igual a 1.0, Normal com valor de pertinência igual a 0.0 e Glicemia Alterada com valor de pertinência igual a 0.0, sucessivamente para cada conjunto de medição.

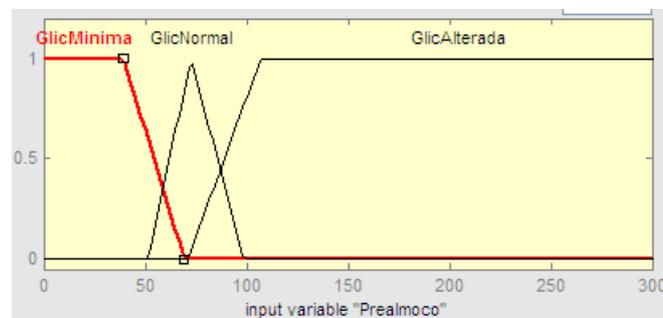


Figura 6.18 Funções de pertinência da variável lingüística de entrada Pré-almoço

Na formação da base de regras, um número finito de regras que pode ser estabelecido depois que se determinar o número de entradas e saídas necessários ao sistema de controle, principalmente no que se refere a escolha dos predicados lingüísticos (número, distribuição e forma das funções de pertinência) para cada uma dessas variáveis. Por exemplo, se o sistema de controle *fuzzy* (variáveis das medições de glicemia no teste de tolerância a glicose e Estágio de Glicemia), possui duas entradas com 3 predicados cada, pode-se definir no máximo $3^4 = 81$ possíveis regras. As variáveis de entrada "Pré-almoço = variável 1 (V_1)" e "Pós-almoço = variável 2 (V_2)", "Pré-ceilia = variável 3 (V_3)", "Pós-ceilia = variável 4 (V_4)", são os valores de entrada (premissa das regras) e o "Estágio de Glicemia = variável 5 (V_5)" é o conseqüente da regra. Para construção da base de regras, utilizou-se o método de acesso ou busca aleatória

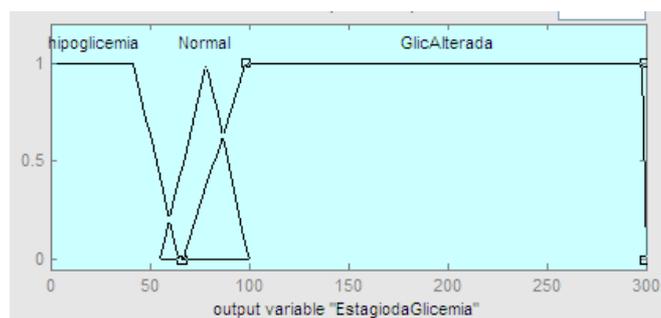


Figura 6.19 Funções de pertinência da variável linguística de saída Estágio de Glicemia

Variáveis Linguísticas Entradas	Função Trapezoidal Mínima	Função Triangular Normal	Função Trapezoidal Alterada
Préalmoço	-3.5 -2.7 39.2 69.4	51.2 72.6 98.8	71 107 303.6 304
Pósalmoço	-3.5 -2.7 39.2 69.4	51.2 72.6 98.8	71 107 303.6 304
Préceia	-3.5 -2.7 39.2 69.4	51.2 72.6 98.8	71 107 303.6 304
Pósceia	-3.5 -2.7 39.2 69.4	51.2 72.6 98.8	71 107 303.6 304
Saída			
Estágio de Glicemia	-3.5 -2.7 39.2 69.4	51.2 72.6 98.8	71 107 303.6 304

Tabela 6.12 Funções de pertinência das medições glicêmicas (Entrada/Saídas)

exposto por Ming [55], conforme a seguir:

A tabela 6.13, apresenta a composição da base de regras pelo método do acesso aleatório, formando 9 (nove) regras que compõem a base de inferência do controlador *fuzzy* estabelecido pelo método Mandami, a qual pode ser entendida como uma superfície de controle responsável pela representação do comportamento do conjunto de regras com as variáveis de entradas. O eixo vertical corresponde a variável de saída do controlador (Estágio de Glicemia), em cada eixo horizontal encontram-se as variáveis de entrada de acordo com seu comportamento representado pelo sistema *fuzzy*.

	V1			V2			V3			V4		
EG	GM	GN	GA									
GM	GM	GA	GN	GA	GM	GA	GA	GN	GA	GN	GM	GN
GN	GN	GA	GM	GA	GM	GA	GA	GM	GA	GN	GM	GA
GA	GM	GM	GN	GM	GN	GA	GA	GN	GA	GM	GA	GM

Tabela 6.13 Mapa de Regras por acesso aleatório composto por: Estágio da Glicemia (EG), Glicemia Mínima (GM), Glicemia Normal (GN), Glicemia Alterada (GA)

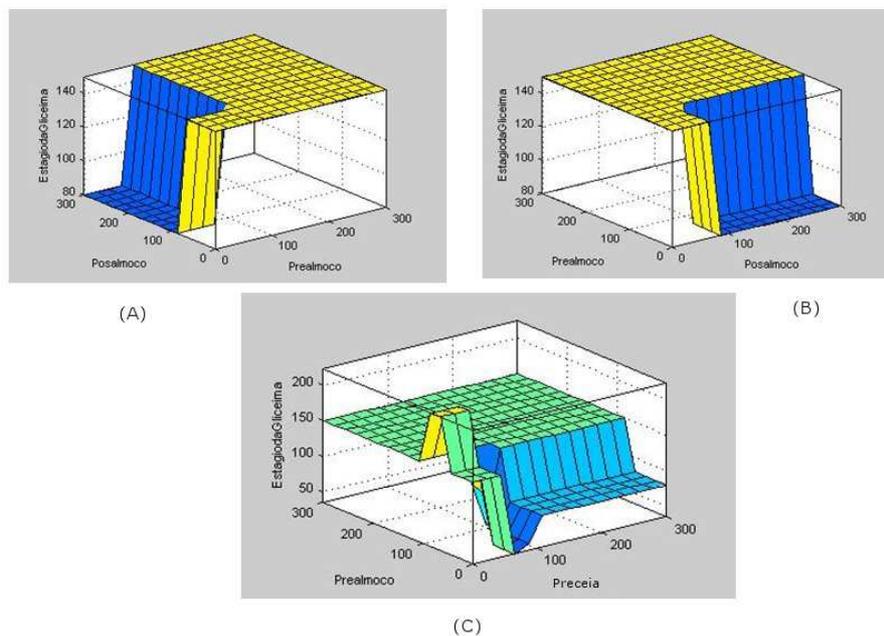


Figura 6.20 Gráfico de superfície de controle dos sinais de glicemia

A figura 6.21, apresenta os sinais de dados proveniente das medições glicêmicas de acordo com o perfil do paciente. Durante o processo de manipulação de dados e treinamento da rede SOM, os dados originais foram submetidos ao processo de normalização (item 3.1.7), após o treinamento os dados foram desnormalizados para a exposição da plotagem dos sinais (figura 6.22). A figura 6.22, apresenta o padrão de dados adotado a partir da saída *Crisp* do controlador *fuzzy*. Observa-se que a variabilidade do sinal glicêmico encontra-se em torno de 217.6 a 227.2 (glicemia alterada), ou seja, paciente com quadro clínico de diabetes *mellitus*.

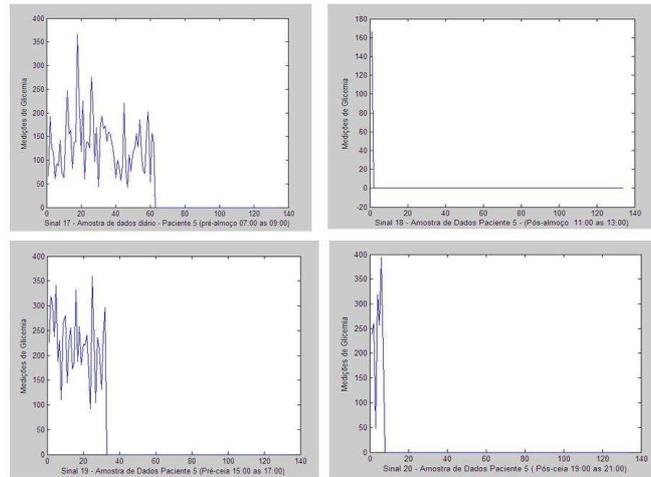


Figura 6.21 Gráficos dos sinais das variáveis de entrada do controlador *fuzzy*, após a execução do mapa de Kohonen

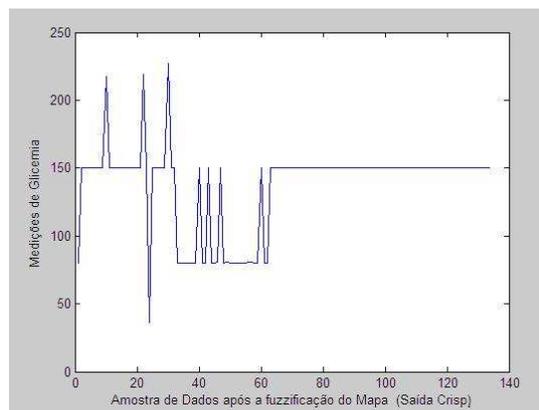


Figura 6.22 Saída *Crisp* do controlador *fuzzy* proveniente do Estágio de Glicemia

6.2.8 Considerações Finais

A integração das redes SOM com modelos *fuzzy* baseados em regras demonstra que a projeção não linear dos dados estabelecida pelos mapas auto organizáveis não permite que os intervalos das classes sejam identificados ou possam estabelecer quaisquer medidas que permita representar um limiar mínimo. Além disso, base de dados com valores muito próximos a zero poderá influenciar na análise do mapa topográfico. A implementação de controladores *fuzzy* baseados em regras apresenta algumas vantagens práticas por ser de fácil compreensão e seu efeito e resultado podem ser facilmente interpretados. Além disso, todas as funções de controle associadas com as regras pode ser testadas individualmente e poderá aumentar a facilidade de manutenção devido a simplicidade das regras, as quais permitem o uso de pessoal menos treinado. Durante o processo de análise do mapa contextual, observou-se que a rotulação da base de dados muitas vezes é descaracterizada devido a alta dimensionalidade dos dados pro-

jetadas na rede, ou seja, quando a base de dados possui uma alta dimensionalidade à análise do mapa é dificultada devido a aglomeração dos rótulos em certas partes do mapa. Este princípio induz a inclusão de um maior número de neurônios para uma projeção mais adequada do mapa, além de aumentar o número de regras no controlador, considerando fator significativo para a elevação do custo computacional da rede SOM.

Neste contexto, sugere-se rotular os dados com o menor quantitativo de caracteres para se obter uma melhor visualização e análise dos dados, evitando assim, a utilização de neurônios excessivos, os quais poderão influenciar no custo computacional durante o treinamento. Desta forma, ambos conjuntos de dados apresentam uma aplicação prática numa área tradicionalmente não explorada pelos métodos de mineração de dados, fluxo de potência e na classificação dos níveis de glicose provenientes dos testes glicêmicos em pacientes portadores de diabetes. Este capítulo não tem a intenção de eleger a "melhor ferramenta" para a atividade de mineração de dados, mas de oferecer subsídios para avaliação e interpretação nebulosa dos mapas de Kohonen a partir dos resultados obtidos pelas ferramentas aplicadas.

Conclusão

Esta dissertação demonstra que os mapas de Kohonen e a lógica nebulosa provém de recursos sofisticados quando aplicado a tarefa de mineração de dados. Uma característica importante, a qual deverá ser observada nesta dissertação relaciona-se ao uso de dados reais para avaliar os resultados obtidos com a aplicação das ferramentas, isto, oferece um forte caráter experimental ao trabalho. Foram também abordados alguns dos principais métodos para aplicação em mineração de dados, os quais foram testados e comentados de forma resumida abordando algumas de suas características e limitações a partir da revisão sobre o estado da arte desta dissertação. As ferramentas pesquisadas mais profundamente neste trabalho possui foco específico na rede SOM e a lógica nebulosa, os quais foram tratadas em capítulos específicos adotando aplicação dessas ferramentas em mineração de dados, incluindo uma análise de influência dos seus parâmetros de controle.

7.1 Contribuições

Resumidamente, as principais contribuições gerais deste estudo são:

- Revisão conceitual das principais ferramentas com possível aplicação em mineração de dados;
- Apresentação das limitações da ferramenta SOM adotando técnicas complementares como lógica nebulosa como critério de nova abordagem ao aspecto de mineração de dados;
- Proposição de um modelo híbrido Som-Fuzzy aplicado na tarefa de mineração de dados;
- Representação da modelagem SOM-Fuzzy proposta nesta dissertação como métodos multidisciplinar, os quais foram aplicados a outros tipos de domínio, fato este, demonstrado no capítulo 6, adotando estudos de casos distintos.

7.2 Extensões

Vários tópicos, relacionados direta ou indiretamente com esta pesquisa, podem ser citados como sugestão para pesquisas futuras:

- A utilização de algoritmo genético para a elaboração e implementação da base de regras do controlador de forma adaptável baseada em programação evolucionária;
- Adotar uma abordagem paramétrica utilizando o método Takagi-Sugeno, considerado um modelo híbrido de um método *fuzzy* baseado em regras e de equações lineares multidimensionais.
- Adotar o modelo exposto nesta dissertação comparando com o modelo de mapeamento topográfico gerativo e a lógica nebulosa paramétrica.
- Utilizar Raciínio baseado em Casos para dirigir os parâmetros da rede de forma automatizada;
- Utilizar a saída *crisp* do controlador nebuloso como critério de análise da série de tempo no que refere a: investigar o mecanismo gerador da série temporal, fazer previsões de valores futuro da série,descrever o comportamento da série; procurar tendências e periodicidades nos dados, modelar periodicidades relevantes nos dados (análise espectral);
- Adotar um estudo da complexidade da rede SOM e da lógica nebulosa e seus eventuais derivados contemplando a complexidade do algoritmo quanto ao seu esforço computacional.

Métodos Estatísticos utilizados em Mineração de Dados

A.1 Tipos de Regressão

A.1.1 Regressão Linear

A regressão linear corresponde à forma mais simples de regressão e os modelos bivariados de regressão linear utilizando variável aleatória X e Y , onde Y é a variável de resposta e X é a variável de predição, pode ser declarada por:

$$Y = \alpha + \beta X \quad (\text{A.1})$$

Onde a variância de Y corresponde a uma constante e os coeficientes de regressão α e β são especificados a partir da interceptação de Y e da inclinação da curva de regressão, respectivamente. Para estimação dos coeficientes existem vários métodos, na literatura, o mais conhecido corresponde ao método dos mínimos quadrados, o qual minimiza o erro entre a linha de separação dos dados e da linha estimada. Dado n de pontos de dados $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ os coeficientes de regressão para o método dos mínimos quadrados poderá ser estimado pelas equações abaixo:

$$\beta = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{A.2})$$

$$\alpha = \bar{y} - \beta \bar{x} \quad (\text{A.3})$$

Onde \bar{x} é a média aritmética de x_1, x_2, \dots, x_n e \bar{y} corresponde a média aritmética de y_1, y_2, \dots, y_n . Os coeficientes α e β podem oferecer boas aproximações para equações mais complexas.

Podemos notar que o salário médio Y observado para um determinado tempo de experiência X . A curva obtida pelos salários médios observados para os respectivos tempo de experiência chama-se de curva de regressão (figura 2.1) e essa dependência de Y é expressa por $E[Y/X]$ (média de dado X).

A.1.2 Regressão Múltipla

Corresponde a uma extensão da regressão linear envolvendo mais de uma variável de predição, permite que a variável de resposta Y a ser modelada como um recurso da função linear a

X (anos de experiência)	Y (salário 1000)
3	30
8	57
9	64
13	72
0 3	36
6	43
11	59
21	90
1	20
16	83

Tabela A.1 Dados de Salário

partir de vetor multidimensional descrita pela equação (4), onde um modelo de regressão está baseado em duas variáveis de predição X_1 e X_2 . O método dos mínimos quadrados poderá ser aplicado para calcular α, β_1 e β_2 [24].

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \tag{A.4}$$

Por exemplo, dado o plano de regressão considerando as seguintes variáveis:

Vendas (Y)	6	7	15	18	20	23
Gastos com TV (X_1)	3	4	8	8	10	11
Gastos com Jornal (X_2)	1	2	3	5	8	6

Tabela A.2 Dados para o plano de regressão

Com base na tabela 2, quanto mais variáveis são consideradas as equações alteram-se, mas os métodos são praticamente os mesmos que a regressão linear. Inicialmente calcula-se os parâmetros a, b_1 e b_2 de tal forma que os quadrados dos desvios dos valores observados em relação aos calculados para Y sejam um mínimo. Onde:

$$M = \sum(Y - \bar{Y})^2 = \sum(Y - a - b_1 X_1 - b_2 X_2) \tag{A.5}$$

Seja um mínimo, para tanto devemos ter:

$$\frac{\delta M}{\delta a} = 0, \frac{\delta M}{\delta b_1} = 0, \frac{\delta M}{\delta b_2} = 0 \tag{A.6}$$

Fornecendo as três equações normais para a determinação de a, b_1 e b_2 :

$$\begin{cases} \sum Y = na + b_1 \sum X_1 + b_2 \sum X_2 (I) \\ \sum Y X_1 = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 (II) \\ \sum Y X_2 = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 (III) \end{cases} \tag{A.7}$$

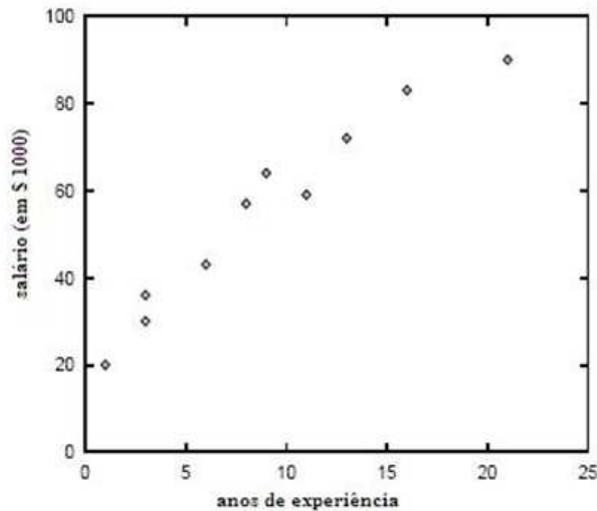


Figura A.1 Plotagem dos dados : embora os pontos não são abrangidos por uma linha reta, o modelo apresentado sugere uma relação linear entre x (anos de experiência) e y (salário)

Resolvendo o sistema da equação (7) e dividindo-se a equação (I) por termos:

$$\frac{\sum Y}{n} = \frac{na}{n} + b_1 \frac{\sum X_1}{n} + \frac{\sum X_2}{n} \quad (\text{A.8})$$

Onde:

$$\bar{Y} = a + b_1 X_1 + b_1 X_2 \quad (\text{A.9})$$

Substituindo-se o valor de a na equação (II):

$$\frac{\sum X_1 \sum Y}{n} - b_1 \frac{(\sum X_1)^2}{n} - b_2 \frac{\sum X_1 \sum X_2}{n} + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \quad (\text{A.10})$$

Colocando-se b_1 e b_2 em evidência,

$$\sum Y X_1 - \frac{\sum X_1 \sum Y}{n} = b_1 \left[\sum X_1^2 - \frac{(\sum X_1)^2}{n} \right] + b_2 \left[\sum X_1 X_2 - \frac{(\sum X_1 \sum X_2)^2}{n} \right] \quad (\text{A.11})$$

Chamando $S_{11} = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$ e $S_{12} = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n}$

Teremos :

$$S Y_1 = b_1 S_{11} + b_2 S_{12} \quad (\text{A.12})$$

Substituindo-se o valor de na equação (III), teremos :

$$S Y_2 = b_1 S_{21} + b_2 S_{22} \quad (\text{A.13})$$

Então, um das maneiras para encontrarmos a, b_1 e b_2 será resolvendo o sistema:

$$\begin{cases} SY_1 = b_1S_{11} + b_2S_{12} \\ SY_1 = b_1S_{11} + b_2S_{12} \end{cases} \quad (\text{A.14})$$

Encontrando b_1 e b_2 e, assim, determinamos o valor de a , substituindo b_1 e b_2 na equação (9), teremos:

$$\bar{X}_1 = \frac{\sum X_1}{n}, \bar{X}_2 = \frac{\sum X_2}{n}, \bar{Y}_1 = \frac{\sum Y}{n} \quad (\text{A.15})$$

A.1.3 Regressão Polinomial

As equações do tipo $Y = \alpha + \beta_1X_1 + \beta_2X_2^2 + \beta_3X_3^3 + \dots + \beta_nX_n^n$, constitui-se um modelo de regressão polinomial de grau n em X . Assim, quando a variável de resposta e as variáveis de predição têm uma relação que pode ser modelada por uma função polinomial, as condições básicas de um modelo linear não tem continuidade transformações nas variáveis para que se possa converter um modelo não linear em um modelo linear e a partir deles poderão ser calculados os coeficientes de regressão pelo método dos mínimos quadrados. A transformação de um modelo de regressão polinomial para um modelo de regressão linear poderá ser aplicado a partir da equação (17).

$$Y = \alpha + \beta_1X_1 + \beta_2X_2^2 + \beta_3X_3^3 \quad (\text{A.16})$$

O processo de conversão da equação (17) de forma linear poderá ser definida a partir das novas variáveis de acordo com a equação (19):

$$X_1 = X, X_2 = X^2, X_3 = X^3 \quad (\text{A.17})$$

Assim, ao substituirmos a equação (17) pela forma linear, a transformação da equação resulta em $Y = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$, o qual poderá ser resolvido pelo método dos mínimos quadrados.

Apontamentos Matemáticos

B.1 Função Gaussiana

A indicação da função gaussiana de modo a satisfazer o critério de vizinhança, está relacionada a simetria entre a distribuição normal e as curvas de densidades de probabilidade de forma unimodais que pode ser caracterizado em formato de sino. Assim, como todas as distribuições normais tem a mesma forma, em função da descrição da média μ e pelo desvio padrão σ , a medição da distribuição, pode ser representada pela função de densidade das distribuições normais por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}e^{-(x-\mu)^2/2\sigma^2}} \quad (\text{B.1})$$

Onde: x , representa cada valor da variável aleatória contínua que $-\infty < x < \infty$; σ , representa o desvio padrão da população; e , base do logaritmo natural; μ é média da população. A padronização da distribuição normal ocorre quando todas as distribuições normais têm os mesmos $\mu = 0$ e $\sigma = 1$. Desta forma, a padronização de uma variável com distribuição normal $N(\mu, \sigma)$, produz uma nova variável com distribuição normal padronizada $N(0, 1)$ que pode ser representada por:

$$Z = \frac{(x - \mu)}{\sigma} \quad (\text{B.2})$$

Onde, z também tem distribuição normal, chamada de variável normal reduzida, normal padronizada ou variável normalizada. A vantagem de se usar a variável exposta na equação B.2, está relacionada onde os valores das áreas poderá ser tabelada, ou as probabilidades, assim, para cada x dado, a área depende de μ e σ .

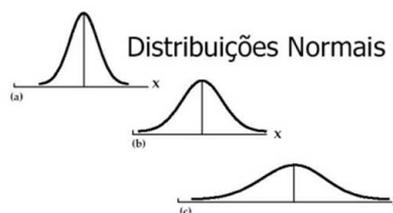


Figura B.1 Evolução da largura da gaussiana nas distribuições normais

Grandezas Elétricas

C.1 Fatores Típicos das Cargas

O termo carga empregado nos estudos dos sistemas elétricos o mais comum considera-se como carga total de um determinado ponto elétrico, a soma de todas as cargas individuais dos diversos equipamentos conectados direta ou indiretamente aos sistemas de transmissão de alta tensão, distribuição de média tensão e baixa tensão, incluindo-se as parcelas referentes aos diversos elementos de transporte e controle do sistema, tais como: cabos, transformadores, reatores, banco capacitores entre outros [6].

C.1.1 Demanda

Os fatores típicos utilizados em sistemas de distribuição em conformidade com as normas técnicas poderão ser definidos a partir da demanda. A demanda de uma instalação é a carga nos terminais receptores tomada em valor médio num determinado intervalo de tempo [46]. Também pode ser definida como o valor médio de potência que o consumidor toma da rede num determinado intervalo de tempo [79]. Desta forma, considera-se "carga", aplicação que está sendo medida em termos de potência, aparente, ativa ou reativa, ou ainda, em termos do valor eficaz da intensidade de corrente, conforme a conveniência.

O intervalo de demanda, corresponde o período no qual o valor médio é designado [70]. Por exemplo, quando a curva da demanda instantânea (intervalo de demanda tender a zero) em função do tempo, obtendo-se a "curva instantânea de demanda no período". Geralmente, ocorre flutuações muito grandes na demanda, sendo, portanto, prática corrente tomar-se a curva de demanda do período considerando-se um intervalo demanda não nulo, usualmente a cada 10 ou 15 minutos [10].

A variabilidade da carga ao longo do dia proporciona um estudo mais detalhado para a condição da demanda máxima, uma vez que, a mesma contribui de forma mais severa de queda de tensão e de aquecimento. A modelagem matemática para um conjunto de cargas n , cuja demanda diária é dada por $D_i(t)$, com $i = 1, 2, \dots, n$, a demanda diversificada D_{div} (soma individuais das cargas, num dado instante) do conjunto de cargas poderá ser expressa por:

$$D_{div}(t) = \sum_{i=1,n} D_i(t) \quad (C.1)$$

Onde a demanda máxima diversificada corresponde ao instante t_a , em que ocorre a demanda máxima (corresponde a maior de todas as demandas que ocorrem num período especi-

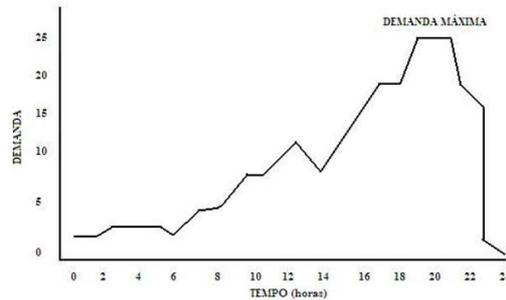


Figura C.1 Curva diária de demanda [46]

ficado de tempo) do conjunto de cargas, isto é:

$$D_{div,max}(t) = D_{div}(t_a) = \sum_{i=1,n} D_i(t_a) \quad (C.2)$$

A demanda diversificada unitária pode ser representada por um conjunto de n , cargas $D_{div}(t)$, como sendo:

$$D_{div}(t) = \frac{1}{n} \cdot \sum_{i=1,n} D_i(t_a) \quad (C.3)$$

C.1.2 Tensão

Os aparelhos eletrodomésticos e equipamentos elétricos são projetados para produzir um determinado trabalho, que terão o seu desempenho e vida útil alterado quanto maior for a diferença entre a tensão de placa ou nominal e a fornecida pela concessionária de serviços de eletricidade. O comportamento da carga tem uma estreita correlação com sua tensão terminal e indiretamente impactam as tensões e controles de um sistema qualquer de transmissão e/ou distribuição, ou seja, diferentes tipos e modelos de carga impactam de forma diferenciada no sistema [6]. Para os fabricantes é interessante que as tensões, entregues pela concessionária, tenham uma gama de variação (diferença entre os valores de tensão máxima e mínima) a mais próxima possível dos valores considerados no projeto, o que conduzirá à produção dos seus equipamentos elétricos a um preço menor. Por outro lado, as concessionárias de serviço de eletricidade desejam que esses equipamentos absorvam o máximo possível de variação, o que resultaria em menores investimentos nos sistemas elétricos [75].

As variações da tensão de fornecimento resultam em alterações no funcionamento dos motores de indução, devido à relação quadrática entre a tensão aplicada ao motor e a sua característica conjugada versus escorregamento poderá influenciar no toque, corrente de partida e na velocidade nominal e em alguns casos poderá fazer com que o motor não funcione [75]. Nos equipamentos eletrônicos as variações de tensão reage de forma diferente, seja acima ou abaixo dos seus valores nominais, podendo ocasionar as subtensões que geralmente conduzem à diminuição da vida útil dos equipamentos, como por exemplo de subtensão (valor abaixo do

nominal), pode-se citar para o caso das televisões, as variações de qualidade de imagens e a redução no quadro da tela.

Muitos estudos podem ser visto na literatura para obtenção de um ponto de equilíbrio entre as necessidades das concessionárias produzirem energia elétrica a tarifa razoável e que os fabricantes possam produzir equipamentos mais acessíveis, com objetivo comum de fornecer aos consumidores de equipamentos eletrodomésticos e eletrônicos com o menor custo. Assim, a tensão nominal é definida como o valor de tensão que se utiliza no projeto de circuitos e aparelhos elétricos [75] e está relacionada ao valor "eficaz da tensão" combinado entre as concessionárias de energia elétrica e o consumidor, no ponto de entrega de energia, nos termos da legislação em vigor, representadas pelos órgãos regulamentadores Aneel[10] e ONS [70]. A regulação de tensão está relacionada a variação existente entre o valor máximo e mínimo de tensão, num determinado ponto do sistema elétrico, que poderá ser representada pela equação a seguir:

$$T_{Reg} = V_0 - V_1 (\text{volts}) \quad (\text{C.4})$$

Onde: T_{Reg} regulação de tensão, V_0 representa a tensão máxima e V_1 a tensão mínima medida em volts. A queda de tensão poderá ser representada para o mesmo instante, pela diferença entre os valores de tensão na entrada e na saída em um dado componente do sistema elétrico, conforme o que segue:

$$T_{tens} = Q_i - Q_f (\text{volts}) \quad (\text{C.5})$$

Onde: Q_{tens} = queda de tensão; Q_i = tensão no início do componente; Q_f = tensão no final do componente.

C.1.3 Fator de Potência

O transporte de energia elétrica, através das linhas de transmissão, da central elétrica para o consumidor, faz com que a mesma seja transformada em outras formas de energia (luminosa, térmica, motora etc.).

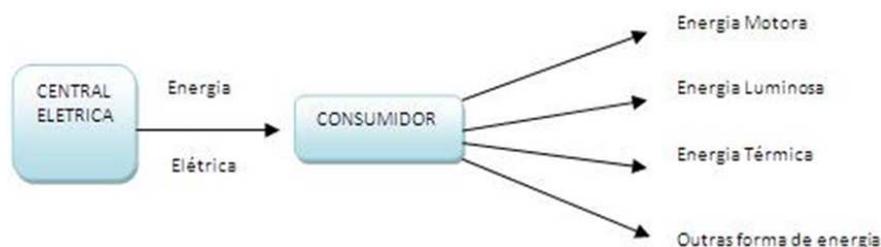


Figura C.2 Modelo de transformação de energia elétrica em outras formas de energia [75]

O fator de potência pode ser definido por um número que indica o quanto a energia elétrica é transformada em outras formas de energia. Este número varia entre zero(0) e um (1), (0 a 100%) definido a nível de percentual [75]. De acordo com a literatura, um fator de potência igual a 100% indicaria que o "aparelho" utilizou toda a energia elétrica consumida, transformando-a em outras formas de energia.

De outra forma, um aparelho que tivesse fator de potência igual a zero não transformaria nenhuma energia elétrica, ou seja, estaria desperdiçando toda a energia recebida. Assim, um $f_{pot} = 1$ (fator de potência), representa que toda energia elétrica consumida é transformada em outras formas de energia. Quando um $f_{pot} < 1$, apenas parte da energia elétrica consumida é transformada. Considerando que, um baixo fator de potência implica em maior custo de operação, possibilitando uma conta de energia mais elevada, sobrecarga no sistema elétrico, aumento das perdas de energia por calor, nível reduzido de tensão, causando mal funcionamento dos motores e iluminação reduzida. Quando este fator passa por um processo de correção poderá obter as seguintes vantagens:

- Redução na conta de consumo de energia elétrica;
- Libera a capacidade elétrica do sistema de distribuição de energia, ou seja, após o processo de correção é possível adicionar novas cargas aos motores, lâmpadas, sem sobrecarregar o sistema de distribuição elétrica;
- Elevar o nível e reduzir as quedas de tensão.

O fator de potência é considerado a relação entre a corrente ativa e a corrente total e indica quanto da corrente total é utilizada para fornecer trabalho e podendo ser representada matematicamente por:

$$F_{pot} = \frac{C_a}{C_t} \quad (C.6)$$

Onde: F_{pot} = Fator de potência; C_a = Corrente ativa; C_t = Corrente total. Quando o $F_{pot} = \cos\theta$ podemos definir o fator de potência em termos de potência (capacidade de produzir trabalho), deduz-se:

$$P = E + T \quad (C.7)$$

Logo: P = Potência, E = Energia e T = tempo.

$$F_{pot} = \cos\theta = \frac{P_{ativa}}{P_{aparente}} \quad (C.8)$$

A representação de P_{ativa} potência ativa (aquela que efetivamente produz trabalho útil, usualmente expressa em kW, $P_{aparente}$ = potência aparente (a soma vetorial das potências ativa, representada por kVA) e $P_{reativa}$ = potência reativa (aquela que produz o fluxo magnético

necessário ao funcionamento dos motores, transformadores usualmente expressa em kVAr), representada pela equação a seguir:

$$P_{aparente}^2 = P_{ativa}^2 + P_{reativa}^2 \quad (C.9)$$

O fator de potência pode ser definido pela relação entre a potência ativa consumida para determinada tensão e corrente e a máxima potência que poderia ser obtida com as mesmas tensões e corrente [75].

APÊNDICE D

Variabilidade da Carga - Dados Originais - Fator de Potência

A variabilidade das cargas nos intervalos de 00 a 24 horas do fator de potência entre as subestações de Cruz das Almas, Benedito Bentes, Tabuleiro e Trapiche localizadas na cidade de Maceió - AL, podem ser visualizadas a seguir:

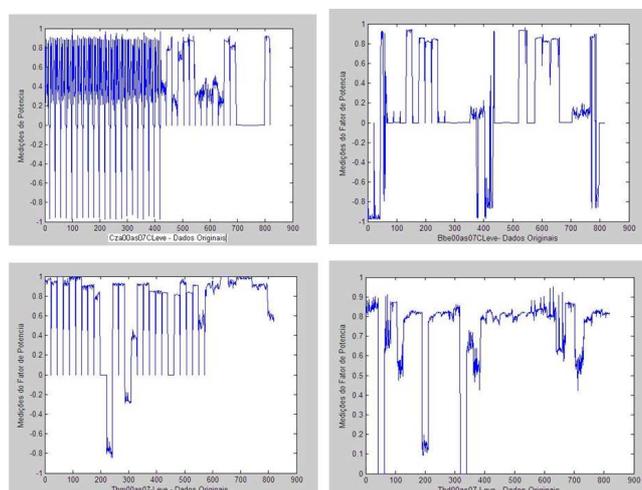


Figura D.1 Variabilidade da carga no intervalo de 00 às 07 - Carga Leve

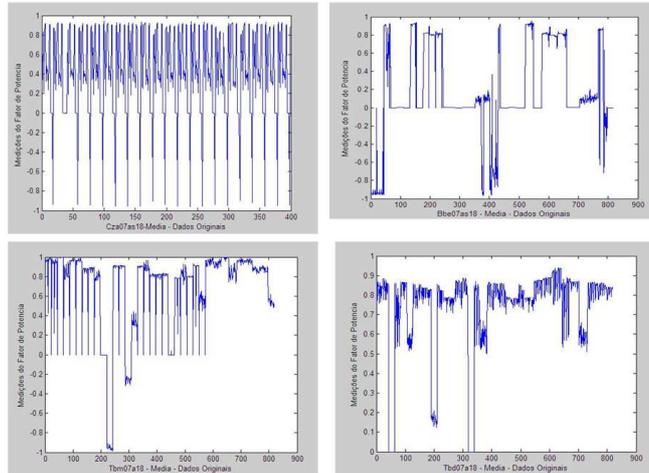


Figura D.2 Variabilidade da carga no intervalo de 07 às 18 - Carga Média

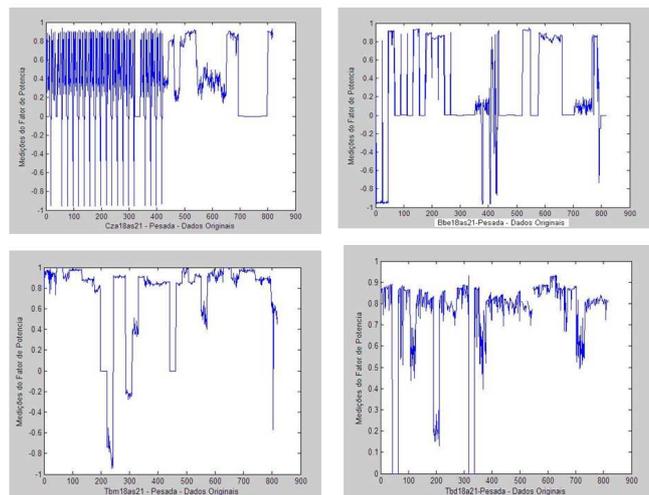


Figura D.3 Variabilidade da carga no intervalo de 18 às 21 - Carga Pesada

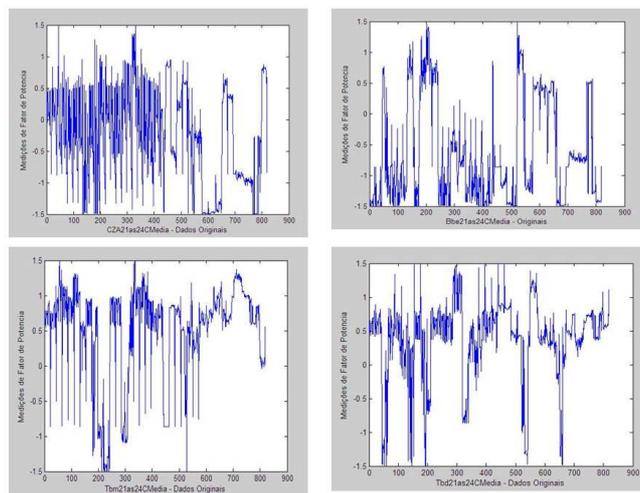


Figura D.4 Variabilidade da carga no intervalo de 21 às 24 - Carga Média

APÊNDICE E

Rótulação dos Sinais de Glicemia

Codificação dos Sinais no Mapa de Kohonen	Descrição Resumida
1	prea1
2	posa1
3	prec1
4	hipo1
5	prea2
6	prec2
7	prea3
8	posa3
9	prec3
10	posc3
11	prel3
12	hipo3
13	prea4
14	posa4
15	prec4
16	hipo4
17	prea5
18	posa5
19	prec5
20	posc5
21	prel5
22	prea6
23	prec6
24	prea7
25	prec7
26	prea8
27	prec8
28	posc8
29	prel8
30	hipo8

Tabela E.1 (A) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen

Codificação dos Sinais no Mapa de Kohonen	Descrição Resumida
30	hipo8
31	prea9
32	prec9
33	posc9
34	prel9
35	prea10
36	prec10
37	prea11
38	posa11
39	prec11
40	prel11
41	hipo11
42	prea12
43	posa12
44	prec12
45	prel12
46	hipo12
47	prea13
48	prel13
49	hipo13
50	prea14
51	prec14
52	prel14
53	hipo14
54	prea15
55	prec15
56	prel15
57	hipo15
58	prea16
59	posa16
60	prec16

Tabela E.2 (B) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen

Codificação dos Sinais no Mapa de Kohonen	Descrição Resumida
61	prel16
62	hipo16
63	prea17
64	prec17
65	prel17
66	hipo17
67	prea18
68	prec18
69	prel18
70	hipo18
71	prea19
72	prec19
73	prea20
74	prec20
75	prea21
76	prec21
77	prea22
78	prec22
79	posc22
80	prel22
81	prea23
82	prec23
83	prea24
84	prec24
85	prea25
86	prec25
87	prea26
88	prec26
89	prea27
90	prec27
91	prea28
92	prec28
93	prea29
94	prec29
95	posc29
96	prel29
97	prea30
98	prec30
99	posc30
100	prel30

Tabela E.3 (C) Mapa de Rótulos dos Sinais de Glicemia no Mapa de Kohonen

Descrição Resumida	Descrição Detalhada
prea1	pré-almoço (número do paciente)
posa1	pós-almoço (número do paciente)
prec1	pré-ceia (número do paciente)
hipo1	hipoglicemia (número do paciente)
posc3	pós-ceia (número do paciente)
prel3	pré-lanche (número do paciente)

Tabela E.4 Descrição detalhada dos rótulos (medições de glicemia)

Referências Bibliográficas

- [1] Dossier diabetes, publicado pela associação protetora dos diabéticos de Portugal, sociedade portuguesa de diabetologia, sociedade portuguesa de endocrinologia, diabetes e metabolismo, Portugal, 2000.
- [2] Pauline M. Mele A and David E. Crowley. Application of self-organizing maps for assessing soil biological quality. *Agriculture, Ecosystems and Environment*, pages 140–145, 2007.
- [3] Pieter Adriaans and Dolf. Zantinge. *Data Mining*. England, 1996.
- [4] D. Aguadoa, T. Montoyaa, L. Borrashb, and J. Ferrera A. Secob. Using SOM and PCA for analysing and interpreting data from a P-removal SBR. *Engineering Applications of Artificial Intelligence*, pages 1–12, 2007.
- [5] Bjørnar Allgot, Delice Gan, Hilary King, Pierre Lefèbvre, Jean-Claude Mbanya, Martin Silink, Linda Siminerio, Rhys Williams, and Paul Zimmet. *Diabetes Atlas*. Belgium, 2003.
- [6] Marcelo Batista. Amaral. Modelagem estática de carga por meio de medição direta: Proposição de uma metodologia sistematizada, seus impactos e benefícios. dissertação de mestrado do programa de pós-graduação em engenharia elétrica. Master's thesis, Universidade Federal de Juíz de Fora. Juiz de Fora - MG, 2004.
- [7] Mariangela Souza Amendola, Anderson Luiz de Barros, and Laercio Carvalho. Manual do uso da teoria fuzzy no matlab 6.5, versão ii. Technical report, FEAGRI & IMECC/ UNICAMP, Campinas - SP, 2005.
- [8] Paulo A.N, Garcia José L.R, and Pereira Sandoval Carneiro Jr. Fluxo de potência trifásico por injeção de corrente. *Revista Controle & Automação n°03, Rio de Janeiro, COPPE/UFRJ.*, 12, 2001.
- [9] Paulo A.N, Garcia José L.R, and Pereira Sandoval Carneiro Jr. Fluxo de potência trifásico por injeção de corrente. *Revista Controle & Automação n°03- COPPE/UFRJ*, 12, 2001.
- [10] ANEEL. *Resolução n° 456,2000/11/29 da Agência Nacional de Energia Elétrica*. URL:<http://www.anel.gov.br>, 2000.
- [11] Antônio de Pádua Braga, Teresa Bernarda Ludemir, and André Carlos Ponce de Leon Ferreira Carvalho. *Redes Neurais Artificiais : teoria e aplicações*. Rio de Janeiro - RJ, 2000.

- [12] Everitt. Brian. *Cluster Analysis - 3rd Edition*. John Wiley, London, 1993.
- [13] Mario Massa Campos and Kaku. Saito. *Sistemas Inteligentes em Controle e Automação de Processos*. Ciência Moderna, Rio de Janeiro, 2004.
- [14] Christer Carlsson and Robert Fuller. Multiobjective linguistic optimization. *Fuzzy Sets and Systems, Elsevier*, 115:5–10, 2000.
- [15] Christer Carlsson and Robert Fuller. Optimization under fuzzy if then rules. *Fuzzy Sets and Systems, Elsevier*, 119:111–120, 2001.
- [16] Herman. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association. No. 342*, 68:361–368., 1973.
- [17] Herman Chernoff and M. Haseeb. Rizvi. Effect on classification error of random permutations of features in representing multivariate data by faces. *Journal of the American Statistical Association. N^o. 351*, 70:548–554, 1975.
- [18] Howard Demuth and Mark. Beale. Neural network toolbox for use with matlab - users guide - version 3.0. url: www.mathworks.com. Technical report, The MathWorks, 1998.
- [19] Carlos Alberto R. Diniz and Francisco Louzada. Neto. *Data Mining : Uma Introdução*. ABE - Associação Brasileira de Estatística, São Carlos, 2000.
- [20] Pink Dua, Francis Doyle, and Efstratios Pistikopoulos. Model - based blood glucose control for type 1 diabetes via parametric programming. *IEEE - Transactions on Biomedical Engineering, n^o 8*, 53:1478–1491, 2006.
- [21] Gomide F. and Pedrycz W. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press, 1998.
- [22] José Carlos Lobato Fasini. Implementação de um simulador de fluxo de potência em regime permanente usando interface gráfica e interativa; dissertação de mestrado, do curso de pós graduação em métodos numéricos em engenharia. Master's thesis, Universidade Federal do Paraná, Curitiba, 1998.
- [23] Rafael Ferrari. Equalização de canais de comunicação digital baseada em filtros fuzzy, dissertação de mestrado do departamento de comunicação da faculdade engenharia elétrica da universidade estadual de campinas. Master's thesis, Universidade Estadual de Campinas, Campinas, SP, 2005.
- [24] Jairo Simon Fonseca, Gilberto de Andrade Martins, and Geraldo Luciano. Toledo. *Estatística Aplicada*. Atlas, São Paulo, 1995.
- [25] G.F.Sudha and P. Jeyashree. Facial expression recognition using self organizing map. *International Conference on Computational Intelligence and Multimedia Applications - IEEE*, pages 219–221, 2007.

- [26] Gene H. Golub and Charles F. Van Loan. *Matrix Computations, Third Edition*. Johns Hopkins University Press, London, 1996.
- [27] Paul. Grant. A new approach to diabetic control: Fuzzy logic and insulin pump technology. *Medical Engineering Physics*, pages 824–827, 2006.
- [28] A. M. Guerra, A Andrés G. C. Piñuela and, B. Galán, and J. R Viguri. Assessment of self-organizing map artificial neural networks for the classification of sediment quality. *Environment International*, pages 1–9, 2008.
- [29] Fredric M. Ham. *Principles of neurocomputing for science and engineering*. McGraw-Hill, New York, 2001.
- [30] S. Haykin. *Neural Networks: a comprehensive foundation*. Hardcover , New York, 2n ed., 2001.
- [31] Wolfgang Härdle and Léopold. Simar. *Applied Multivariate Statistical Analysis*. Berlin, 2003.
- [32] Weiming Hu. Learning activity patterns using fuzzy self-organizing neural network. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. NO. 3*, 34, 2004.
- [33] S. J. Huang and J. M. Lin. Enhancement of anomalous data mining in power system predicting-aided state estimation. *IEEE Transactions on Power Systems, N° 1*, 19:610–619, 2004.
- [34] Mustaq Hussain and John P. Eakins. Component-based visual clustering using the self-organizing map. *Neural Networks*, pages 260–273, 2007.
- [35] W. H. Inmon. *Como construir um Data Warehouse*. Campus, Rio de Janeiro, 1997.
- [36] Josmar Ivanqui. Esteira eletrônica com velocidade controlada por lógica fuzzy, dissertação de mestrado do programa de pós-graduação em engenharia elétrica e informática industrial do centro federal de educação tecnológica do paraná. Master's thesis, Centro Federal de Educação Tecnológica do Paraná, 2005.
- [37] Sylvio Jabloka. *Diabetes Mellitus*. Fundo Editorial, BYK-PROCIENX, São Paulo, 1980.
- [38] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, New Jersey, 1988.
- [39] Anil K. Jain and JianChang. Mao. Artificial neural network: a tutorial. *IEEE*, 1996.
- [40] Anil K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. acm computing surveys, n° 3. *ACM Computing Surveys*, 31:264–323, 1999.
- [41] Javad Jassbi, S. H. Alavi, Paulo J. A. Serra, and Rita A. Ribeiro. Transformation of a mamdani fis to first order sugeno fis. *IEEE*, 2007.

- [42] H. Jiawei and Kamber M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, New York, 2000.
- [43] Jr. Joseph F. Hair. *Análise Multivariada de Dados, 5ª Ed.* Bookman, Porto Alegre, 2005.
- [44] Aldemir. Junglos. Aplicação de data mining em banco de dados do serviço de transplante de medula Óssea - dissertação de mestrado do programa de pós-graduação em métodos numéricos em engenharia. Master's thesis, Universidade Federal do Paraná . Curitiba, 2003.
- [45] Nelson Kagan, Carlos Cesar Barioni de Oliveira, and Ernesto João Robba. *Introdução aos Sistemas de Distribuição de Energia Elétrica*. Edgard Blücherl, São Paulo, 2005.
- [46] A.M Kalteh, P. Hjorth, and R. Berndtsson. Review of the self-organizing map (som) approach in water resources: analysis, modelling and application. *Environmental Modelling & Software*, pages 835–845, 2007.
- [47] S. Kaski. Data exploration using self-organizing maps. acta polytechnica scandinavica, mathematics, computing and management in engineering series n° 82. thesis the doctor : Url: <http://www.cis.hut.fi/sami/thesis.ps.gz>. recuperado em 05/11/07, 1997. Master's thesis, Helsinki University of Technology, Finland, 1997.
- [48] Kyung-Joong Kim and Sung-Bae. Cho. Fuzzy integration of structure adaptive som for web content mining. *Fuzzy Sets and Systems.*, 148:46–60, 2004.
- [49] Ralph. Kimbal. *The Data Warehouse Toolkit: guia completo para modelagem dimensional*. Campus, Rio de Janeiro, 2000.
- [50] Kimmo Kiviluoto and Erkki. Oja. S-map: a simple self-organization algorithm generative topographic mappings. filand : Laboratory of computer of information science helsinki university of technology. Technical report, Helsinki University of Technology, 1995.
- [51] T Kohonen. *Self-Organizing Maps*. Springer, Berlin, 1997.
- [52] Zsolt L. Kovacz. *Redes Neurais Artificiais: fundamentos e aplicações*. Livraria da Física, São Paulo, 2002.
- [53] Luigi Di Lascioa. A fuzzy-based methodology for the analysis of diabetic neuropathy. *Fuzzy and Systems*, pages 203–228, 2007.
- [54] Ming-Ling Lee, Hung-Yuan Chung, and Fang-Ming Yu. Modeling of hierarchical fuzzy systems. *Fuzzy Sets and Systems, Elsevier*, 138:343–361, 2003.
- [55] Eliane Prezepiorski Lemos. Análise de crédito bancário com uso de data mining: Redes neurais e Árvore de decisão, dissertação de mestrado do programa de pós-graduação em métodos numéricos em engenharia do departamento de matemática. Master's thesis, Universidade Federal do Paraná, Curitiba, 2003.

- [56] Renato Dourado Maia. *Projeto de Estruturas Neuro-Fuzzy do Tipo Takagi-Sugeno Utilizando Programação Genética, Dissertação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica da Universidade Federal de Minas Gerais*. PhD thesis, Universidade Federal de Minas Gerais, 2005.
- [57] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 07, nº1:1–13, 1975.
- [58] Chiara Dalla Man, Robert A. Rizza, and Claudio Cobelli. Meal simulation model of the glucose insulin system. *IEEE - Transactions on Biomedical Engineering*, nº 10, 54:1740–1749, 2007.
- [59] Ian S. Shaw Marcelo Godoy Simoes. *Controle e Modelagem Fuzzy, 2ª Edição*. Blucher - Fapesp, São Paulo, 2007.
- [60] Felipe Nascimento Martins. Controle do gerador de indução trifásico baseado em lógica fuzzy, dissertação de mestrado do programa de pós-graduação em engenharia elétrica. Master's thesis, Universidade Federal do Espírito Santo, Vitória, RJ, 2003.
- [61] Ernest L. Mazzaferri. *Endocrinology. A Review of Clinical Endocrinology*. New York - USA, 1978.
- [62] F. Martin McNeill and Ellen Thro. *Fuzzy Logic: A practical Approach*. London, 1994.
- [63] Jerry M. Mendel. Fuzzy logic systems for engineering: A tutorial. *IEEE*, pages 345–377, 1995.
- [64] K. Nagata. Pattern recognition of eeg signals during motor imagery. *SICE-ICASE International Joint Conference, IEEE*, pages 5169–5173, 2006.
- [65] Manoel Jorge Ribeiro. Neto. *Desenvolvimento de Rede Neural SOM: Um Estudo de Caso para Segmentação de Perfis - Trabalho de Conclusão de Curso de Graduação do Instituto de Computação da Universidade Federal de Alagoas*. Maceió - AL, 2007.
- [66] D.J. Newman, S. Hettich, C.L. Blake, and C.J Merz. *UCI Repository of machine learning databases [http://www.ics.uci.edu/mllearn/MLRepository.html]*. Irvine, CA: University of California, Department of Information and Computer Science. 1998.
- [67] I. Nonaka and H. Takeuchi. *Criação de Conhecimento na Empresa : como as empresas japonesas geram a dinâmica da inovação*. Rio de Janeiro, 1997.
- [68] Changmok Oh, Ming-Soeng Kim, and Ju-Jang Lee. Eeg signal classification based on pca and nn. *IEEE - SICE - ICASE International Joint Conference*, pages 1848–1851, 2006.
- [69] ONS. *Consolidação da Previsão de Carga para Elaboração do Programa Mensal da Operação Energética, submodulo 05-06-r0, http://www.ons.org.br/*. Operador Nacional do Sistema Elétrico, São Paulo, 2001.

- [70] World Health Organization. *Diabetes Action Now*, <http://www.who.int/>, visitado 15.06.2008. Suíça, 2004.
- [71] Camella Owens. Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus. *IEEE- Transactions on Biomedical Engineering*, n^o 6, 53:990 – 996, 2006.
- [72] Bigus J. P. *Data Mining with Neural Networks: solving business problems from application development to decision support*. New York, 1996.
- [73] Daniel Panarotto, Allison Roberto Teles, and Marina Verdi Schumacher. Fatores associados ao controle glicêmico em pacientes com diabetes tipo 2. *Associação Médica Brasileira*, pages 314–321, 2008.
- [74] Vladimir Gomes Pinto. *Coleção Distribuidora de Energia Elétrica - Eletrobrás : Controle de Tensão de Sistemas de Distribuição*. Campus, Rio de Janeiro, 1985.
- [75] Newton Roy Pampa. Quispe. Técnicas e ferramentas para a extração inteligente e automática de conhecimento em banco de dados. dissertação de mestrado. Master's thesis, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas. Campinas - SP, 2003.
- [76] Alvin C. Rencher. *Methods of Multivariate Analysis, Second Edition*. Wiley Interscience, Canada, 2002.
- [77] O. S. Resende. *Sistemas Inteligentes: fundamentos e aplicações*. Manole, Barueri - São Paulo, 2005.
- [78] Álvaro Pereira Rizzi. *Medidas Elétricas : Potência, Energia, Fator de Potência e Demanda - Escola Federal de Engenharia de Itajubá*. Livros Técnicos e Científicos Editora, Rio de Janeiro, 1980.
- [79] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, New Jersey, 1995.
- [80] Alexandre José MontAlverne Silva, Amâncio Paulino de Carvalho, Ana Márcia Messeder S. Fernandes, Antônio Luiz Brasileiro, Augusto Pimazoni Netto, Carmem de Simone, Débora Malta, Denizar Vianna Araújo, Dillian Adelaine da Silva Goulart, Dirceu Brás Aparecido Barbano, Edson Aguilar Perez, Flávio Danni Fuchs, Lenildo de Moura, Mário Maia Bracco, Newton Sérgio Lopes Lemos, Regina Maria Aquino Xavier, Renata F. Cachapuz, Rubens Wagner Bressanim, and Victor Matsudo. *Cader-nos de Atenção Básica : Diabetes Mellitus, 1.ª edição*. Brasília-DF, 2006.
- [81] S. N. Silvanandam, S. Sumatri, and S.N. Deepa. *Introduction to Fuzzy Logic Using Matlab*. Springer-Verlag, Berlin, 2007.

- [82] Lindsay I. Smith. A tutorial on principal components analysis, recuperado em 15/05/2007 : Url: csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf. *Technical report*, 2007.
- [83] A.T Stewart. *Intellectual Capital: the new wealth of organizations*. Doubleday Currency, New York, 1997.
- [84] Joe Stufflebeam and Nadipuram R. Prasad. Hierarchical fuzzy control. *IEEE - International Fuzzy Systems Conference Proceedings*, pages 498–503, 1999.
- [85] Pawel Tadejko and Waldemar Rakowski. Mathematical morphology based ecg feature extraction for the purpose of heartbeat classification. In *6th International Conference on Computer Information Systems and Industrial Management Applications, IEEE.*, 2007.
- [86] N.F. Thornhilla. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 2002.
- [87] Ultsch and Vetter. Self-organizing feature maps versus statistical clustering methods: A benchmark. research report n° 9. dep. of mathematics, url. <http://www.mathematik.uni-marburg.de/wina/papers/94.cluster.ps>. Technical report, University of Marburg, 1994.
- [88] M. Fayyad Usama, Piatetsky-Shapiro Gregory, Padhraic Smyth, and Ramasamy Uthurusamy. *Advances In Knowledge Discovery and Data Mining*. AAI Press/ The MIT Press., 1996.
- [89] Mêuser Valença. *Aplicando Redes Neurais: um guia completo*. Livro Rápido, Olinda, 2005.
- [90] Juha Vesanto. Som-based data visualization methods. finland : Laboratory of computer and information science, helsinki university of technology. Technical report, Helsinki University of Technology, 1999.
- [91] J. Vesanto, J. Himberg, E. Alhoniemi, and E. Parhankangas. J. som toolbox for matlab 5 - technical report a57 - helsinki universit of technology - url: <http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf>. recuperado em 05/05/07. finland. Technical report, Helsinki Universit of Technology, 2000.
- [92] Augusto Cesar G. Vieira. *Manual de Correção do Fator de Potência*. São Paulo, 1973.
- [93] C. Apte Weiss and S.M. Data mining with decision trees and decision rules. *Future Generation Computer Systems*, pages 197–210, 1997.
- [94] I.H. Witten and E. Frank. *Data Mining: practical machine learning tools and techniques with Java implementation*. São Francisco, California, 1999.
- [95] Sitao Wu and Tommy W. S. Chow. Self-organizing and self-evolving neurons:a new neural network for optimization. *IEEE TRANSACTIONS ON NEURAL NETWORKS*. N° 2, 18:385–396, 2007.

- [96] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision process. *IEEE Trans. Systems and Cybernetics*, 1973.
- [97] L A. Zadeh. A fuzzy-algorithm approach to the definition of complex or imprecise concepts. *Int. J. Man Machines Studies*, 8:249–295, 1976.
- [98] L A. Zadeh. *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems*:. World Scientific, 1996.
- [99] Márcio Henrique. Zuchini. Aplicações de mapas auto-organizáveis em mineração de dados e recuperação de informações, dissertação de mestrado do departamento de engenharia de computação e automação industrial. Master's thesis, Universidade Estadual de Campinas da Faculdade de Engenharia Elétrica. Campinas - SP, 2003.