

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ANA CARLA MACEDO DA SILVA

**EXTRAÇÃO DE CONHECIMENTO EM FORMA DE REGRAS
DIFUSAS A PARTIR DE MAPAS AUTO-ORGANIZÁVEIS DE
KOHONEN – APLICAÇÃO EM DIAGNÓSTICO DE FALTAS
INCIPIENTES EM TRANSFORMADORES**

TD __/ 2013

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2013

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ANA CARLA MACEDO DA SILVA

**EXTRAÇÃO DE CONHECIMENTO EM FORMA DE REGRAS
DIFUSAS A PARTIR DE MAPAS AUTO-ORGANIZÁVEIS DE
KOHONEN – APLICAÇÃO EM DIAGNÓSTICO DE FALTAS
INCIPIENTES EM TRANSFORMADORES**

TD __/ 2013

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará como requisito parcial para obtenção do Grau de Doutor em Engenharia Elétrica.

Orientador: Profa. Dra. Adriana Rosa Garcez Castro

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2013

Dados Internacionais de Catalogação na Publicação (CIP)
Sistemas de Bibliotecas da UFPA

Silva, Ana Carla Macedo da, 1973-

Extração de Conhecimento em Forma de Regras Difusas a partir de Mapas Auto-Organizáveis de Kohonen – Aplicação em Diagnóstico de Falhas Incipientes em Transformadores / Ana Carla Macedo da Silva. – 2013.

Orientador: Adriana Rosa Garcez Castro.

Tese (Doutorado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2013.

1. Redes Neurais (computação). 2. Sistemas difusos. I. Título.

CDD: 22. ed. 006.32

Agradecimentos

Agradeço a DEUS, por me conduzir e me fortalecer na realização desta jornada.

À Prof. Dra. Adriana Rosa Garcez Castro, pela orientação, pelo auxílio e pela amizade, pelos ensinamentos repassados e pelos caminhos apontados para o sucesso deste trabalho. Por ter me assumido como sua orientanda já com prazo curto para defesa da qualificação.

Ao meu esposo, Jaime Soares, pelo amor, carinho, incentivo e compreensão.

A minha mãe, Ana Maria, pelo incentivo e apoio em todos os sentidos.

Aos meus avós maternos (*in memoriam*).

Ao meu irmão e as minhas tias (em especial à minha tia professora de Português) que me incentivaram durante a realização do curso.

À Pró-Reitora de Planejamento e Desenvolvimento, Raquel Trindade Borges, pelo incentivo, compreensão e liberação.

À Jaciane do Carmo Ribeiro pela ajuda sobre os conceitos estatísticos.

À Míriam Domingues pela ajuda e orientação nos momentos difíceis.

À Lenilda, ao Fábio pelo incentivo.

Aos professores Vladimiro Miranda, Roberto Limão de Oliveira, Rogério Almeida e Ubiratan Holanda Bezerra, membros da banca de qualificação, pelas valiosas revisões, críticas e avaliações.

À Universidade Federal do Pará, pela oportunidade de capacitação profissional oferecida.

Ao Programa de Pós-Graduação em Engenharia Elétrica, pela realização do Curso e pelo apoio.

A todos que de alguma forma contribuíram para a realização desta tese.

So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.

Albert Einstein

Resumo

Apesar das diversas vantagens oferecidas pelas redes neurais artificiais (RNAs), algumas limitações ainda impedem sua larga utilização, principalmente em aplicações que necessitem de tomada de decisões essenciais para garantir a segurança em ambientes como, por exemplo, em Sistemas de Energia. Uma das principais limitações das RNAs diz respeito à incapacidade que estas redes apresentam de explicar como chegam a determinadas decisões; explicação esta que seja humanamente compreensível. Desta forma, este trabalho propõe um método para extração de regras a partir do mapa auto-organizável de Kohonen, projetando um sistema de inferência difusa capaz de explicar as decisões/classificação obtidas através do mapa. A metodologia proposta é aplicada ao problema de diagnóstico de faltas incipientes em transformadores, em que se obtém um sistema classificatório eficiente e com capacidade de explicação em relação aos resultados obtidos, o que gera mais confiança aos especialistas da área na hora de tomar decisões.

Palavras-chave. Mapa auto-organizável, sistemas de inferência difusa, diagnóstico de faltas incipientes em transformadores de potência.

Abstract

Despite the many advantages offered by the artificial neural networks, some limitations still prevent their widespread use, especially in applications that require making decisions essential to ensure safety in environments such as in Power Systems. A major limitation of artificial neural networks with respect to the inability of these networks is to explain how to arrive at certain decisions. This explanation must be humanly understandable. Thus, this paper proposes a method for extracting fuzzy rules from Kohonen self-organizing map, designing a fuzzy inference system capable of explaining the decisions taken by the map. To verify its effectiveness, the method is applied to solve the problem of classification for the diagnosis of incipient faults in power transformers used

Keywords. Self-organized map, fuzzy inference system, dissolved gas analysis.

Lista de Figuras

FIGURA 2.1 Grade bidimensional adaptada de Malone (2006).....	16
FIGURA 2.2 Tipos de organização de grade bidimensional.....	17
FIGURA 2.3 Vizinhanças de tamanhos discretos 0, 1 e 2 da unidade de mapa mais ao centro	19
FIGURA 2.4 Comportamento da função da taxa de aprendizagem.....	20
FIGURA 2.5 Matriz-U e sua escala de cores.....	24
FIGURA 2.6 Componente Plano.....	25
FIGURA 2.7 Localização da unidade de mapa que representa o neurônio 3.....	25
FIGURA 2.8 Tecelagem de Voronoi (KOHONEN, 2001).....	27
FIGURA 2.9 Conjunto difuso convexo (a) e conjunto difuso não-convexo (b) (ROSS, 2004)	29
FIGURA 2.10 Conjuntos difusos para concentração de H ₂	31
FIGURA 2.11 Modificadores para baixa concentração de gás H ₂	32
FIGURA 2.12 Configuração Básica de um sistema difuso com fuzzificador e defuzzificador	36
FIGURA 2.13 Correção de partições difusas incompletas.....	36
FIGURA 2.14 Modelo Takagi-Sugeno de Ordem Zero.....	39
FIGURA 3.1 Critério de transparência.....	44
FIGURA 4.1 Mapa Auto-organizável de Kohonen.....	53
FIGURA 4.2 Exemplo de resultado apresentado pela Matriz-U.....	54
FIGURA 4.3 Função de pertinência triangular.....	56
FIGURA 4.4 Dois grupos são formados após o treinamento.....	60
FIGURA 4.5 Rótulos das unidades de mapa com os 3 grupos formados para cada espécie de flores.....	60
FIGURA 4.6 Sistema difuso extraído para a base dados Iris.....	62
FIGURA 4.7 Sistema difuso extraído para a base dados Iris, após correção (<i>Repairing</i>).....	62
FIGURA 4.8 Sistema difuso extraído para a base dados do banco suíço.....	66
FIGURA 4.9 Três grupos foram formados após o treinamento.....	69
FIGURA 4.10 Rótulos das unidades de mapa com os 3 grupos formados para cada classe de vinho.....	69
FIGURA 4.11 Sistema difuso extraído para a base dados sobre vinhos.....	71
FIGURA 5.1 Triângulo de Durval.....	80
FIGURA 5.2 Grupos formados após o treinamento.....	87
FIGURA 5.3 Rótulos das unidades de mapa com os 5 grupos formados.....	88

FIGURA 5.4 Mapa rotulado dividido em duas partes (A) e (B), indicando o	88
FIGURA 5.5 Funções de pertinência para a regra R1	90
FIGURA 5.6 Sistema difuso extraído para o diagnóstico de cinco faltas	90
FIGURA 5.7 Fuzzificação da saída.....	92

Lista de Tabelas

TABELA 2.1 Modificadores linguísticos.....	31
TABELA 2.2 Normas-S	33
TABELA 2.3 Normas-T	34
TABELA 3.1 Tabela resumida das abordagens estudadas no Capítulo 3.....	51
TABELA 4.1 Composição da base de dados Iris	58
TABELA 4.2 Divisão da base de dados Iris para treinamento e teste.....	58
TABELA 4.3 Treinamento com o SOM para a base de dados Iris	59
TABELA 4.4 Resultados para extração de regras para a base de dados Iris.....	61
TABELA 4.5 Erros de classificação e de equivalência para a base de dados Iris no treinamento	63
TABELA 4.6 Resultados da classificação para a base de dados Iris	63
TABELA 4.7 Estrutura da base de dados do banco suíço.....	64
TABELA 4.8 Divisão da base de dados do banco suíço para treinamento e teste.....	64
TABELA 4.9 Treinamento com o SOM para a base de dados do banco suíço	64
TABELA 4.10 Treinamento com o SOM para a base de dados do banco suíço.....	65
TABELA 4.11 Resultados para extração de regras para o problema de notas falsas.....	65
TABELA 4.12 Erros de classificação e de equivalência para a base de dados suíço no treinamento	66
TABELA 4.13 Resultados da classificação para a base de dados do banco suíço.....	67
TABELA 4.14 Estrutura da base de dados de vinhos	68
TABELA 4.15 Divisão da base de dados de vinhos para treinamento e teste	68
TABELA 4.16 Treinamento com o SOM para a base de dados de vinhos	68
TABELA 4.17 Resultados para extração de regras da base de dados sobre vinhos.....	70
TABELA 4.18 Erros de classificação e de equivalência para a base de dados de vinhos no treinamento	72
TABELA 4.19 Resultados da classificação para a base de dados de vinhos	72
TABELA 5.1 Critério de identificação de falha de Dörnenburg (DÖRNENBURG e STRITTMATTER, 1974).....	77
TABELA 5.2 Concentração dos gases para validação do método de Dörnenburg.....	78
TABELA 5.3 Concentração dos gases para identificação de condição de normalidade do transformador.....	78
TABELA 5.4 Tabela de definição dos códigos do método de Rogers.....	78
TABELA 5.5 Tabela de diagnóstico do método de Rogers em função do código.....	79
TABELA 5.6 Método do gás chave	79

TABELA 5.7 Tabela que substitui o uso do método do IEC	81
TABELA 5.8 Critério IEC 60599 para interpretação de DGA	81
TABELA 5.9 Tipos de faltas da base de dados	85
TABELA 5.10 Divisão da base de dados de treinamento e teste para o transformador de potência.....	85
TABELA 5.11 Treinamentos do SOM para a base de dados de faltas em transformadores....	86
TABELA 5.12 Resultados para extração de regras	89
TABELA 5.13 Resultados de classificação do Sistema Difuso	91
TABELA 5.14 Resultados de Classificação do FIS com saída fuzzificada	92
TABELA 5.15 Resultados de Classificação - Comparação	93
TABELA 5.16 Comparando com abordagens neuro-difusas	94
TABELA 6.1 Avaliação da Metodologia Proposta	97
TABELA 6.2 Resultados de Classificação – Comparação.....	98

Lista de Abreviaturas

DGA	<i>Diagnosis Gas Analysis</i>
IEC	<i>International Electrotechnical Commission</i>
LVQ	<i>Learning Vector Quantization</i> (Quantização Vetorial por Aprendizagem)
RNA	Redes Neural Artificial
SOM	<i>Self-Organizing Map</i> (Mapa Auto-Organizável)
BMU	<i>Best Matching Unit</i> (unidade de mais semelhante)
FIS	<i>Fuzzy Inference System</i> (Sistema de Inferência Difusa)

Sumário

1. INTRODUÇÃO	11
1.1 Objetivos da tese.....	12
1.2 Trabalho publicado	13
1.3 Estrutura da tese.....	13
2. MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN E SISTEMAS DIFUSOS	14
2.1 Mapa Auto-Organizável de Kohonen.....	15
2.1.1 Arquitetura do Mapa	16
2.1.2 Algoritmo Sequencial.....	17
2.1.3 Variações do Algoritmo	22
2.1.4 Algoritmo em Lote (Batch).....	23
2.1.5 Formas de Visualização e Agrupamento.....	24
2.1.6 Quantização Vetorial.....	26
2.2 Sistemas de Inferência Difusa	27
2.2.1 Teoria dos Conjuntos Difusos.....	28
2.2.2 Sistemas Baseados em Regras Difusas	35
2.2.3 Propriedades dos Sistemas de Inferência Difusa	36
2.2.4 Projeto de Sistemas Difusos Orientados a Dados	37
2.2.5 Sistema Difuso do Tipo Takagi-Sugeno	37
2.3 Conclusão do Capítulo	39
3. ESTADO DA ARTE DE EXTRAÇÃO DE CONHECIMENTO DE MAPAS AUTO-ORGANIZÁVEIS	41
3.1 Extração de Regras/Conhecimento de Redes Neurais Artificiais	41
3.2 Extração de Conhecimento de Mapas Auto-Organizáveis de Kohonen – Estado da Arte.....	46
3.2.1 Extração de Regras Proposicionais	46
3.2.2 Extração de Regras Difusas	49
3.3 Conclusão do Capítulo	50
4. METODOLOGIA PROPOSTA PARA EXTRAÇÃO DE REGRAS DIFUSAS A PARTIR DE MAPAS AUTO-ORGANIZÁVEIS.....	52
4.1 Metodologia para Extração de Regras Difusas a partir de Mapas Auto-Organizáveis de Kohonen	52

4.2 Estudos de Caso da Metodologia Proposta	58
4.2.1 Estudo de Caso 1: Problema da Classificação de Flores Íris	58
4.2.2 Estudo de caso 2: Problema da Classificação de Papel-Moeda	63
4.2.3 Estudo de caso 3: Problema da Classificação de Vinhos	67
4.4 Conclusão do Capítulo	72
5. APLICAÇÃO: DIAGNÓSTICO DE FALTAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA	74
5.1 Diagnóstico de Faltas baseado na Análise dos Gases Dissolvidos em Óleo (DGA).....	75
5.1.1 Diagnóstico de Faltas baseados em Métodos Tradicionais.....	77
5.1.2 Diagnóstico de Faltas baseado em Abordagens Não Tradicionais	82
5.2 Sistema proposto para Diagnóstico de Faltas Incipientes em Transformadores de Potência	85
5.2.1 Banco de dados para desenvolvimento/treinamento do sistema de diagnóstico....	85
5.2.2 Treinamento da Rede de Kohonen e Identificação dos grupos.....	86
5.2.3 Extração de Sistema Difuso a partir do SOM.....	89
5.2.3 Processo de Fuzzificação da Saída do Sistema Difuso	91
5.3 Comparação com um Método Tradicional de Análise dos Gases Dissolvidos em Óleo ...	93
5.4 Comparação com outros Sistemas Neuro-Difusos	94
5.5 Conclusão do capítulo	95
6. CONCLUSÕES GERAIS	96
6.1. Contribuições da tese.....	96
6.2 Limitações e Trabalhos Futuros	98
REFERÊNCIAS	100

1. Introdução

Os Mapas Auto-Organizáveis de Kohonen (Redes de Kohonen) surgiram nos anos 1970 com o trabalho sobre simulação computacional, desenvolvido por C. von der Malsburg: “*Self-organization of orientation sensitive cells in the striate cortex*”. Neste período, a pesquisa sobre Redes Neurais Artificiais estava abalada pelo trabalho de Minsky e Papert, intitulado “*Perceptrons*” (1969) (HAYCKIN, 2007), que supunham ter comprovado por meio de cálculos matemáticos que as limitações dos *perceptrons* de camada única eram também limitações dos *perceptrons* de múltiplas camadas. Apenas na década de 1980, o interesse pelas Redes Neurais Artificiais (RNAs) foi renovado devido ao desenvolvimento de um algoritmo de aprendizagem eficiente para redes do tipo *perceptron* de múltiplas camadas (HAYCKIN, 2007).

As RNAs são modelos computacionais não lineares, inspirados na estrutura e operação do cérebro humano, que procuram reproduzir características humanas, tais como: *aprendizado, associação, generalização e abstração*. Elas vêm sendo utilizadas principalmente em problemas de aproximação de funções e classificação de padrões.

A principal vantagem das RNAs é a sua capacidade de aprendizado a partir de exemplos, estando aptas a desenvolver uma representação concisa de conceitos complexos, mesmo quando os dados apresentam ruídos. No entanto, apesar das diversas vantagens oferecidas pelas RNAs, algumas limitações ainda impedem a sua ampla utilização, principalmente em aplicações que necessitam de tomadas de decisão essenciais para garantir a segurança em ambientes controlados por sistemas, como por exemplo, no setor de fornecimento de energia elétrica. Uma das principais limitações das RNAs diz respeito a sua incapacidade de explicar como chegam a determinadas decisões; explicação que seja humanamente compreensível.

Em muitos casos, os resultados apresentados pelas RNAs são suficientes e não existe a necessidade de tornar explícito o conhecimento capturado durante o processo de aprendizado. Por outro lado, existem aplicações em que o especialista da área precisa compreender o modelo criado pela RNA para poder ter confiança nos resultados obtidos. No entanto, explicar o comportamento das RNAs não é uma tarefa simples devido à representação distribuída do conhecimento nos pesos sinápticos da rede.

Nas últimas décadas, algumas pesquisas vêm sendo apresentadas com o objetivo de minimizar o problema relativo à incapacidade de explanação das redes neurais (Da SILVA et al., 2012; HUNG e HUANG, 2011; KAHRAMANLI e ALLAHVERDI, 2009; HUYNH e.

REGGIA, 2009; QUTEISHAT e PENG LIM, 2008, CASTRO et al., 2007; FUNG et al., 2005; BROUWER, 2005). Em geral, os trabalhos apresentados focalizam a atenção para o desenvolvimento de metodologias voltadas para extração de regras a partir do conhecimento escondido nas redes neurais. Atualmente, esta linha de pesquisa vem sendo referenciada como “extração de regras” de redes neurais e tem como objetivo principal a representação das redes neurais por meio de uma linguagem baseada em regras de inferência (CASTRO, 2004).

1.1 Objetivos da tese

Considerando a importância de compreender as decisões tomadas pelas RNAs, este trabalho tem como objetivo apresentar uma nova metodologia para extração de regras difusas a partir de Mapas Auto-Organizáveis de Kohonen. A escolha do mapa auto-organizável de Kohonen se deve ao fato de que este tipo de rede vem sendo utilizado com bastante sucesso em diversas aplicações de classificação e categorização. Entretanto, assim como outras diversas topologias de RNAs, esta rede possui a desvantagem de não apresentar explanação a respeito dos resultados obtidos. Os mapas auto-organizáveis de Kohonen permitem apenas uma forma de visualização dos grupos/classes formados ao final da fase de treinamento, não passando nenhuma informação adicional sobre o motivo da alocação de determinado padrão em uma classe ou em um grupo específico.

A eficiência da metodologia proposta poderá ser verificada através da sua aplicação ao problema de Diagnóstico de Falhas Incipientes em Transformadores baseado na Análise de Gases Dissolvidos em Óleo (DGA). O desenvolvimento deste sistema inteligente de diagnóstico de falhas é o segundo objetivo desta tese.

Alguns trabalhos vêm sendo apresentados na literatura aplicando Redes Neurais Artificiais ao problema de diagnóstico de transformadores, entretanto, na maioria destes trabalhos não existe a preocupação de se obter explicações relativas ao diagnóstico apresentado pelas redes desenvolvidas. Desta forma, esta tese também tem como objetivo desenvolver, utilizando os Mapas Auto-Organizáveis de Kohonen e a metodologia de extração de conhecimento proposta, um sistema de diagnóstico que apresente resultado de classificação superior aos resultados apresentados pelas metodologias convencionais utilizadas pelas concessionárias de energia e que apresente também, e principalmente, a capacidade de explanação em relação aos resultados obtidos para que o sistema possa ser utilizado com maior confiança por parte dos especialistas da área.

1.2 Trabalho publicado

Da SILVA, A. C. M., GARCEZ, A. R. C., MIRANDA, V. Transformer failure diagnosis by means of fuzzy rules extracted from Kohonen Self-Organizing Map. **International Journal of Electrical Power and Energy Systems**, v. 43, i. 1, p. 1034-1042, 2012. DOI: 10.1016/j.ijepes.2012.06.027.

1.3 Estrutura da tese

Além deste capítulo introdutório, a tese é composta de mais cinco capítulos:

Capítulo 2: Este capítulo apresenta uma breve revisão sobre Mapas Auto-Organizáveis de Kohonen e Sistemas de Inferência Difusa, assim como a terminologia utilizada neste trabalho. Mais especificamente, na seção 2.1, será apresentado o Mapa Auto-organizável de Kohonen em termos de sua arquitetura, treinamento, heurísticas e formas de visualização. A seção 2.2 apresenta uma descrição sobre os Sistemas Difusos em termos da teoria dos conjuntos difusos, operações e sistema de inferência, os quais aplicam regras do tipo *Se-Então* e raciocínio difuso.

Capítulo 3: Neste capítulo, uma revisão sobre extração de regras a partir de Redes Neurais Artificiais será apresentada, a fim de fornecer ao leitor uma visão geral sobre a área na qual a proposta deste trabalho se enquadra. Na seção 3.1, a tarefa da extração de regras a partir de Redes Neurais é definida e a taxonomia para avaliar os algoritmos de extração de regras é apresentada. Na seção 3.2, alguns trabalhos publicados sobre extração de regras a partir dos Mapas Auto-Organizáveis de Kohonen são apresentados com o intuito de revelar o estado da arte na área de pesquisa de interesse deste trabalho.

Capítulo 4: Este capítulo apresenta a metodologia proposta para extração de regras difusas do tipo Takagi Sugeno de Ordem Zero a partir de Mapas Auto-Organizáveis de Kohonen. A metodologia é apresentada, testada e avaliada por meio de três estudos de caso.

Capítulo 5: Neste capítulo, a metodologia de extração de conhecimento é aplicada ao problema de diagnóstico de faltas em transformadores de potência, utilizando a análise de gases dissolvidos em óleo. Trata-se do segundo objetivo desta tese. Antes da apresentação do sistema inteligente proposto, será apresentada uma visão geral sobre diagnóstico de faltas em transformadores baseado na análise dos gases dissolvidos em óleo.

Capítulo 6: Neste capítulo são apresentadas as considerações finais (limitações e contribuições da tese).

2. Mapas Auto-Organizáveis de Kohonen e Sistemas Difusos

Redes Neurais Artificiais e Sistemas Difusos são duas técnicas que se destacam dentro da Inteligência Computacional. A Inteligência Computacional é uma área da ciência que desenvolve sistemas com o objetivo de imitar aspectos do comportamento humano tais como: aprendizado, percepção, raciocínio, evolução e adaptação.

Uma Rede Neural Artificial é um paradigma de processamento de informação inspirado na estrutura densamente interconectada e paralela do cérebro humano. As Redes Neurais são modelos matemáticos que tentam imitar algumas das propriedades do sistema nervoso biológico e que se projetam sobre as analogias do aprendizado biológico adaptativo. Dentre as diversas topologias de Redes Neurais Artificiais, os Mapas Auto-Organizáveis de Kohonen (SOMs – acrônimo em inglês para *Self-Organizing Maps*) têm como fundamento os mapas encontrados em certas áreas do cérebro (especialmente o córtex cerebral), que se organizam de acordo com sua tarefa: reconhecimento de voz e análise de sinais sensoriais. Esta tarefa pode ser assimilada geneticamente ou através da experiência devido à capacidade de adaptação de um sistema nervoso ao seu meio ambiente (plasticidade neural) (HAYCKIN, 2007), que pode ser exemplificada pela simples auto-organização que é controlada principalmente por informação sensorial.

Por outro lado, os sistemas baseados em Lógica Difusa ou simplesmente sistemas difusos (FIS, acrônimo em inglês para *Fuzzy Inference System*) apresentam uma abordagem para a criação de modelos mais proximamente relacionados à ciência cognitiva. A lógica difusa é um superconjunto da lógica convencional (Booleana) que foi estendida para tratar o conceito de verdade parcial, uma vez que valores verdadeiros podem não ser completamente verdadeiros, assim como valores falsos podem não ser completamente falsos (EL-HAWARY, 1998). Trata-se de um sistema baseado em conjuntos difusos, que são utilizados para modelar termos linguísticos, e em regras difusas do tipo *Se-Então* que aplicam tais termos linguísticos aos processos de tomada de decisão.

Este capítulo fornece uma visão geral dos Mapas Auto-Organizáveis de Kohonen e Sistemas Difusos e a terminologia utilizada neste trabalho. Mais especificamente, na seção 2.1, será apresentado o Mapa Auto-Organizável de Kohonen em termos de sua arquitetura, treinamento, heurísticas e formas de visualização. A seção 2.2 apresenta uma descrição sobre os Sistemas Difusos em termos da teoria dos conjuntos difusos, das operações e dos sistemas de inferência, os quais aplicam regras do tipo *Se-Então* e raciocínio difuso.

2.1 Mapa Auto-Organizável de Kohonen

Uma Rede Neural Artificial é um modelo computacional geralmente criado com base em uma metáfora sobre o funcionamento do cérebro humano, caracterizando-se por extrair conhecimento a partir das experiências apresentadas pelo ambiente externo. Como o elemento básico do cérebro é o neurônio, também é o neurônio o elemento fundamental da Rede Neural Artificial. O neurônio se constitui em uma unidade de processamento simples, que armazena conhecimento por meio de forças de conexão de entrada ou de ligações entre eles (as sinapses). A intensidade destas conexões é medida por pesos sinápticos, que armazenam o conhecimento adquirido através de um processo de aprendizagem (HAYCKIN, 2007).

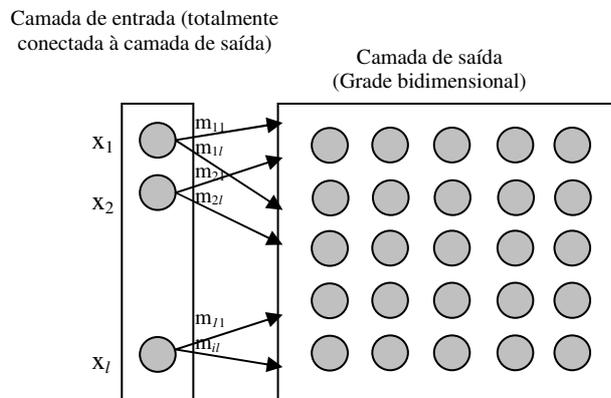
Nos Mapas Auto-Organizáveis de Kohonen, o processo de aprendizagem é baseado na competição entre os neurônios, em que o vencedor é aquele que possui pesos sinápticos mais próximos da informação de entrada da Rede de Kohonen. O neurônio vencedor determina a região, cujos neurônios sofrerão maiores modificações em seu peso, a fim de se aproximarem mais a cada iteração da entrada. Assim, cada neurônio ou grupo de neurônios, ao longo do processo de aprendizado, se torna sensível a um diferente domínio vetorial da entrada, transformando-se em decodificador (es) específico (s) ou detector (es) de seus respectivos domínios de sinal no espaço de entrada. Estes decodificadores são formados na rede em uma ordem significativa, como se algum sistema de coordenadas de características fosse definido sobre a rede (KOHONEN, 2001). A ordenação dos pesos ocorre quando o vetor de pesos sinápticos de cada neurônio é igual à média das amostras alocadas aos neurônios vizinhos por similaridade. Neste momento, os vetores de peso tendem a se aproximar da função de densidade de probabilidade (ou mapeamento de frequências relativas) dos vetores de entrada.

Os Mapas Auto-Organizáveis combinam operações de projeção e de agrupamento, permitindo a realização da análise exploratória de dados, cujo objetivo é produzir descrições sumarizadas de grandes conjuntos de dados. A projeção de dados permite projetar um conjunto de dados de alta dimensão como pontos de uma estrutura de dimensão menor, por exemplo, bidimensional. Pode-se dizer que a finalidade do SOM é permitir a visualização de dados de alta dimensão, comprimindo informação, mais especificamente, convertendo relacionamentos não lineares estatísticos em relacionamentos geométricos simples, quando em um mapa formado por uma grade bidimensional (KOHONEN, 2001).

2.1.1 Arquitetura do Mapa

A arquitetura mais simples do SOM é composta de duas camadas (Figura 2.1). A camada de entrada é composta por nós, nos quais os sinais de um vetor x de dimensão arbitrária l são submetidos à rede. Os nós de entrada apresentam-se totalmente conectados aos neurônios da camada de saída. Tais conexões, as sinapses, possuem valores de pesos que correspondem aos componentes dos vetores de referência m_i , também chamados *modelos* (*codebooks*), que apresentam a mesma dimensão do vetor de entrada x .

FIGURA 2.1 Grade bidimensional adaptada de Malone (2006)



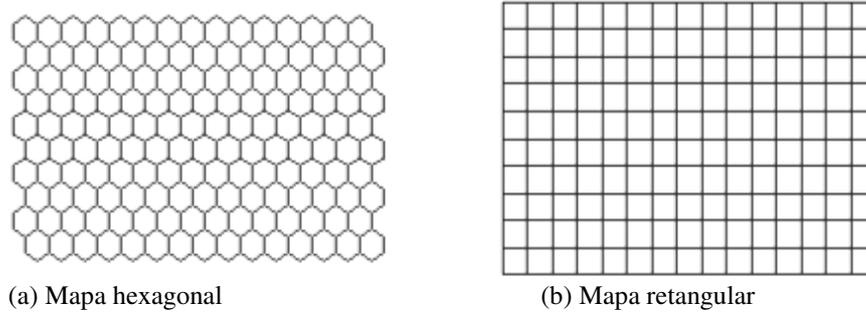
Os neurônios da camada de saída podem estar arranjados em uma grade uni ou bidimensional, que submetida ao algoritmo de aprendizado passa a ter a habilidade de transformar os vetores de referência de dimensão arbitrária l em um mapa discreto uni ou bidimensional por um processo adaptativo e de maneira topologicamente ordenada, o que significa que interações espaciais são definidas entre os neurônios mais próximos ao longo da rede.

Considere $x = [x_1, x_2, \dots, x_l]^T$, $x \in \mathfrak{R}^n$ como um vetor de dados estocásticos, então pode-se dizer mais formalmente que o SOM é uma “projeção não-linear” da função de densidade de probabilidade $p(x)$ do vetor de entrada x de alta dimensão em uma forma de visualização bidimensional (KOHONEN, 2001). De outro modo, $m_i = [m_{i1}, m_{i2}, \dots, m_{in}]^T$, $m_i \in \mathfrak{R}^n$, são os vetores de pesos ou modelos, em que i indica o neurônio ao qual pertencem os pesos sinápticos e n indica a dimensão do vetor x .

A Figura 2.2 apresenta como os neurônios de saída podem ser organizados na grade bidimensional, ou seja, mostra como os neurônios de saída podem-se conectar uns aos outros:

de forma hexagonal ou retangular ou gaussiana. Os neurônios também podem ser arranjados de forma irregular.

FIGURA 2.2 Tipos de organização de grade bidimensional



(a) Mapa hexagonal

(b) Mapa retangular

Quando a grade está organizada de forma hexagonal, cada neurônio apresenta no máximo seis neurônios vizinhos e quando de forma retangular, o número máximo é oito.

2.1.2 Algoritmo Sequencial

O objetivo do algoritmo de treinamento do SOM, na sua modalidade sequencial, é ajustar os pesos sinápticos do mapa de forma que sejam capazes de representar domínios dos valores dos padrões de entrada submetidos. Estes valores podem ou não sofrer um processo de normalização, o qual permite que os componentes dos vetores de entrada sejam colocados em uma mesma escala.

A. Inicialização dos pesos sinápticos

Os vetores de pesos m_i podem ser inicializados de forma randômica, em que os valores para os componentes dos vetores de referência m_i são gerados aleatoriamente. É importante que além de aleatórios, tais valores sejam assimétricos.

Outra forma de inicialização dos pesos é a linear, que se dá com a determinação de dois autovetores da matriz de autocorrelação de x com os maiores autovalores, os quais visam abranger um subespaço linear bidimensional. Um vetor retangular é definido ao longo deste subespaço, sendo que seu centroide coincide com a média de $x(t)$ (vetor de entrada em um tempo t). Este vetor servirá para inicializar os pesos sinápticos, que já estarão ordenados, podendo-se iniciar o treinamento a partir da fase de convergência.

B. Cálculo do Neurônio Vencedor (“the winner neuron”)

Após a inicialização dos pesos sinápticos, os vetores de entrada $x = [x_1, x_2, \dots, x_n] \in \mathfrak{R}^n$ e os vetores de referência m de cada neurônio i da grade são comparados, a fim de encontrar o vetor m , cujos componentes estejam mais próximos dos componentes do vetor de entrada x . Para tanto, é utilizada uma medida de similaridade, geralmente, a distância Euclidiana, definida conforme a equação 2.1:

$$d_{xm} = \|x - m\| = \sqrt{\sum_{i=1}^n (x_i - m_i)^2}, \quad (2.1)$$

onde x_i e m_i são componentes dos vetores x e m respectivamente.

Assim, o neurônio vencedor, identificado pelo índice c , será aquele, cujos componentes apresentarem a menor distância em relação aos componentes do vetor x e será dado por:

$$c = \arg_i \min \{d_{xm}\} \quad (2.2)$$

Outras formas de definir o grau de similaridade entre os vetores de entrada e os pesos sinápticos são possíveis, modificando também a forma como serão realizados os ajustes.

C. Ajuste de Peso

Durante o processo de aprendizagem, os pesos sinápticos do neurônio vencedor e os pesos dos neurônios que estiverem dentro de um raio de vizinhança N_c serão modificados. O ajuste é feito iterativamente a cada passo t e é dado por:

$$\begin{aligned} m_i(t+1) &= m_i(t) + \alpha(t) [x(t) - m_i(t)] \text{ para } i \in N_c(t) \\ m_i(t+1) &= m_i(t) \text{ para } i \notin N_c(t) \end{aligned} \quad (2.3)$$

onde $\alpha(t)$ é o valor da taxa de aprendizagem, calculado a cada iteração t , sendo t o contador do tempo total para submissão de todos os vetores da entrada (uma época). O valor de $\alpha(t)$ deve ser dado por uma função que decresça monotonicamente com o tempo.

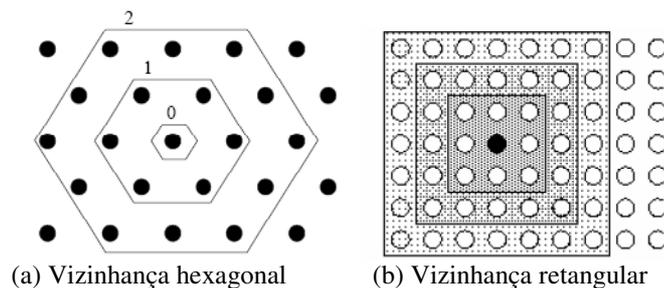
N_c pode ser uma vizinhança retangular como na Figura 2.3 (b) ou pode ser representada por uma função $h_{ci}(t)$, chamada função de vizinhança, escrita em termos da função Gaussiana (KOHONEN, 2001), conforme equação 2.4:

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (2.4)$$

onde $\sigma(t)$ é a largura de $h_{ci}(t)$, calculada a cada iteração t e é dada também por uma função que decresça monotonicamente com o tempo durante o processo de ordenação.

A função de vizinhança deve apresentar o seguinte comportamento: $h_{ci}(t) \rightarrow 0$, quando $t \rightarrow \infty$, a fim de que o processo de aprendizagem seja bem sucedido. Em suma, ela mede o grau de vizinhança topológica entre todos os neurônios (HAYCKIN, 2007). Geralmente, $h_{ci}(t)$ se encontra em função de $\|r_c - r_i\|$, como na equação 2.4, onde $r_c \in \mathfrak{R}^2$ e $r_i \in \mathfrak{R}^2$ são os vetores de localização dos neurônios c e i na grade respectivamente. O aumento da distância $\|r_c - r_i\|$ define o modelo da superfície elástica referida anteriormente, por isso, caso o tamanho da vizinhança N_c seja inicializado com um valor muito pequeno, o mapa não será ordenado globalmente. Ao invés disso, o mapa será dividido em várias partes como se fosse um mosaico, entre as quais a direção da ordenação muda descontinuamente. A Figura 2.3 mostra dois tipos de formatos de vizinhança (hexagonal e retangular) e três tamanhos discretos da vizinhança.

FIGURA 2.3 Vizinhanças de tamanhos discretos 0, 1 e 2 da unidade de mapa mais ao centro



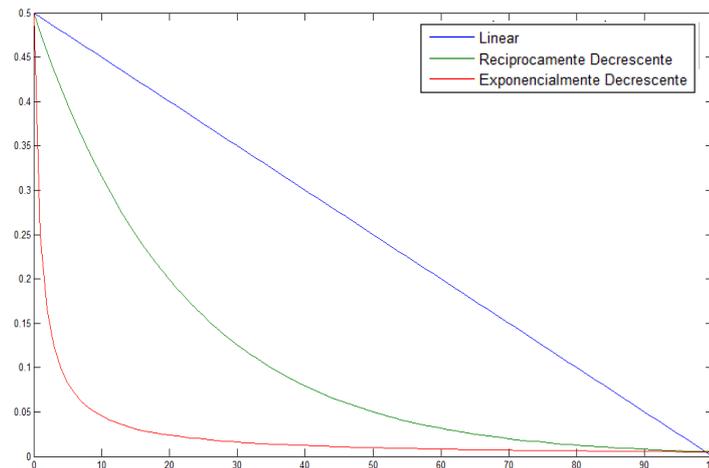
Por fim, o algoritmo passa por duas fases. A primeira fase é de ordenação, que ocorre durante os 1.000 primeiros passos ou mais, nos quais são definidas as vizinhanças entre os neurônios, ou seja, até onde o neurônio vencedor excita os neurônios próximos. A segunda fase, a de convergência, realiza um ajuste fino nos pesos já ordenados dos neurônios mais próximos dos neurônios vencedores para cada padrão de entrada.

D. Heurísticas sobre o algoritmo

O melhor valor do tamanho da vizinhança do neurônio vencedor (N_c) ou do desvio padrão de h_{ci} ($\sigma(0)$) para que o SOM alcance o ordenamento global referido acima, é igual à metade da maior dimensão do mapa, uma vez que evita configurações metaestáveis, para as quais a média da distorção esperada ou erro de quantização médio resulta em um mínimo local e não em um mínimo global (KOHONEN, 2001).

Como não é importante que $\alpha(t)$ seja representado por uma função muito precisa, esta pode ser linear, exponencial ou reciprocamente decrescente a t como mostra a Figura 2.4. Na fase de ordenação, podem ser utilizados valores próximos a 1, já na fase de convergência podem ser usados valores menores, na casa dos centésimos.

FIGURA 2.4 Comportamento da função da taxa de aprendizagem



A fase de convergência deve ser realizada pelos menos 500 vezes o número de unidades do mapa, visto que se trata de um processo estocástico, tal fase deve ser longa para garantir acurácia estatística.

Em Kohonen (2001), são apresentadas as seguintes sugestões para construção de bons mapas topológicos:

1. Para facilitar a inspeção visual, a estrutura hexagonal deve ser escolhida por não favorecer direções verticais ou horizontais. O mapa deve ser retangular e não quadrado, a fim de que os vetores de referência sejam orientados junto com $p(x)$ e estabilizados durante o processo de aprendizagem.
2. Como o processo de aprendizagem requer um grande número de passos e geralmente o número de amostras é pequeno, tal conjunto de dados deve ser

aplicado reiteradas vezes de forma cíclica ou com a ordem alterada randomicamente.

3. Amostras importantes para o aprendizado, quando são raras, podem ser repetidas em um número de vezes que pode ser determinado pelo usuário do mapa.
4. A escala do conjunto de dados de entrada influencia na orientação resultante do vetor de referência, porém se os dados estiverem em diferentes escalas não há regras para colocá-los em uma determinada escala; o que se pode aconselhar é a normalização da variância de cada componente do conjunto de dados e depois checar os erros de quantização médios.
5. Caso se queira forçar a localização de amostras da entrada no mapa, basta usar cópias destas representações como valores iniciais dos vetores de referência, mantendo a taxa de aprendizagem baixa nestes locais durante os ajustes.
6. Diferentes valores iniciais de $m_i(0)$, seqüências de vetores de treinamento $x(t)$ e parâmetros de aprendizagem devem ser testados; o mapa resultante após os treinamentos com o menor erro de quantização é aquele que deve ser escolhido.

Além destas heurísticas, pode-se usar os valores indicados em Hayckin (2007) para as funções com decaimento monotônico de $\sigma(t)$ e $\alpha(t)$. A largura σ da função de vizinhança pode ser dada:

$$\sigma(t) = \sigma(0) \exp\left(-\frac{t}{\tau_1}\right), t = 0, 1, 2, \dots \quad (2.5)$$

onde $\sigma(0)$ é a largura inicial da função de vizinhança com valor igual a 0.1 e τ_1 é constante de tempo com valor igual $\frac{1000}{\log \sigma_0}$

Já a taxa de aprendizagem $\alpha(t)$ pode ser dada por:

$$\alpha(t) = \alpha(0) \exp\left(-\frac{t}{\tau_2}\right), t = 0, 1, 2, \dots \quad (2.6)$$

onde $\alpha(0)$ é a taxa de aprendizagem inicial com valor igual a 0.1 e τ_2 é a constante de tempo com valor igual a 1.000 (número de iterações para a fase de ordenação).

2.1.3 Variações do Algoritmo

A equação da distância Euclidiana ponderada pode ser usada no lugar da equação 2.1 (KANGAS et al., 1990):

$$d_{xm}^w \|x - m\|^w = \sqrt{\sum_{j=1}^n w_{ij}^2 (x - m_{ij})^2} \quad (2.7)$$

onde $w = [w_1, w_2, \dots, w_n]^T$ é um vetor de peso para as entradas da célula i , sendo que $j = 1 \dots n$.

A ideia central é estimar o valor de w_j recursivamente durante o processo de aprendizado, para balancear o efeito dos erros (a disparidade da variância), já que se trata de um processo de aprendizado não supervisionado. Cada célula armazena as médias ponderadas exponencialmente dos valores absolutos de erros $|x - m_{ij}|$, denotando estes valores em cada passo por $\delta_i(t)$:

$$\delta_{ij}(t+1) = (1 - \kappa_1)\delta_{ij}(t) + \kappa_1 \omega_{ij} |x_i(t) - m_{ij}(t)| \quad (2.8)$$

onde a abrangência da média é definida pelo escalar κ_1 .

Os valores de δ_i só serão alterados para as células dentro da vizinhança do neurônio vencedor. Depois, calcula-se a média destes valores:

$$\delta_i(t) = \frac{1}{N} \sum_{j=1}^N \delta_{ij}(t) \quad (2.9)$$

Em cada célula, tenta-se manter a média ponderada de erros no mesmo nível em todas as entradas. Assim, ω é alterado da seguinte maneira:

$$\begin{aligned} \omega_{ij}(t+1) &= \kappa_2 \omega_{ij}(t), 0 < \kappa_2 < 1, \text{ se } \omega_{ij} |x_i(t) - m_{ij}(t)| > \delta_i(t) \\ \omega_{ij}(t+1) &= \kappa_3 \omega_{ij}(t), 1 < \kappa_3, \text{ se } \omega_{ij} |x_i(t) - m_{ij}(t)| < \delta_i(t) \end{aligned} \quad (2.10)$$

Empiricamente, sabe-se que os melhores valores de κ_1 , κ_2 e κ_3 são 0,0001, 0,99 e 1,02 respectivamente (KANGAS et al., 1990). O ajuste do peso é realizado de forma semelhante à equação 2.3.

Outra variação decorre do uso do produto escalar entre x e m no lugar do cálculo da distância Euclidiana, conforme equação 2.11.

$$x^T(t)m_c(t) = \max_i \{x^T(t)m_i(t)\} \quad (2.11)$$

Neste caso, o neurônio que apresentar maior valor para o produto escalar é o que tem maior similaridade em relação a x . O ajuste deve ser realizado da seguinte maneira:

$$m_i(t+1) = \begin{cases} \frac{m_i(t) + \alpha'(t)x(t)}{\|m_i(t) + \alpha'(t)x(t)\|}, & \text{se } i \in N_c(t) \\ m_i(t), & \text{se } i \notin N_c(t) \end{cases} \quad (2.12)$$

A função de aprendizagem deve ser alterada com a equação $\alpha' = \frac{\alpha_0}{t}$, sendo que α_0 deve tomar valores elevados de 10 a 100, por exemplo. Este processo normaliza os pesos e desacelera o treinamento.

2.1.4 Algoritmo em Lote (Batch)

O algoritmo em lote ou paralelo lembra o algoritmo do vetor de quantização. Todas as amostras devem estar disponíveis, quando o treinamento iniciar. Os passos do algoritmo são os seguintes (KOHONEN, 2001):

1. Inicializar os vetores de referência com os padrões de entrada;
2. Para cada unidade do mapa i , coletar uma lista das cópias de todas as amostras de treinamento x , cujo vetor de referência mais próxima seja o da unidade i , ou seja, o conjunto de Voronoi da unidade i ;
3. Atualizar cada vetor de referência com a média da união das listas em N_i , ou seja, na vizinhança da unidade de mapa i ;
4. Repetir o passo 2 umas poucas vezes.

Quando N_i não é uma vizinhança retangular, a equação de ajuste é:

$$m_i^* = \frac{\sum_j n_j h_{ji} \bar{x}_j}{\sum_j n_j h_{ji}}, \quad (2.13)$$

onde \bar{x}_j é a média das amostras do conjunto de Voronoi da unidade j que está na vizinhança de i e n_j é o número de amostras do conjunto de Voronoi de j .

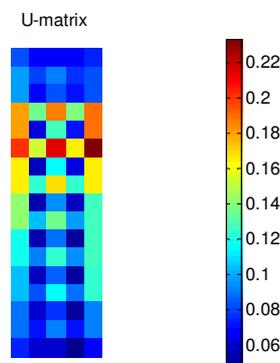
2.1.5 Formas de Visualização e Agrupamento

A seguir, são apresentadas as formas de visualização e/ou agrupamento aplicadas aos mapas de Kohonen.

A. *U-Matrix (Matriz-U)*

As abstrações formadas pelo SOM podem ser visualizadas através de projeções. A Matriz-U (matriz de distância unificada) é a forma mais popular de visualização dos resultados do SOM, a qual interpreta tais abstrações como agrupamentos (VESANTO, 1999). A técnica utilizada para isto é a distância entre matrizes (*distance matrices*) que calcula as distâncias entre os pesos sinápticos dos neurônios em relação a cada um de seus vizinhos, interpretando os seus valores por meio de uma escala de cor (VESANTO, 1999) (Figura 2.5). Em Ultsch e Siemon (1990), a Matriz-U é apresentada como uma forma de visualização de grupos, mesmo para casos em que é difícil destacar as bordas dos mesmos.

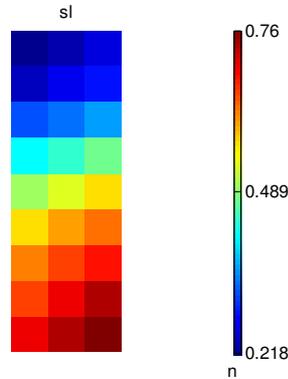
FIGURA 2.5 Matriz-U e sua escala de cores



A Matriz-U é uma matriz com dimensão igual a $\mathfrak{R}^{2l-1 \times 2c-1}$, ou seja, entre cada linha e cada coluna do mapa são acrescentadas uma linha e uma coluna respectivamente. A Matriz-U é implementada pela SOM Toolbox (VESANTO et al., 2000), que produz uma escala de cores ao final do processo de aprendizagem, indicando as distâncias entre os pesos dos neurônios. Ainda, na Figura 2.5, as maiores distâncias entre as unidades de mapa são indicadas pela cor vermelha (bordas de grupos); as menores distância, pela cor azul-escuro, indicando os grupos. As matrizes-u muito pequenas dificultam tal interpretação, porque o SOM gera uma matriz-u complexa. A ferramenta produz também um componente plano para cada componente dos vetores de entrada e por meio de uma escala de cores exibe seu comportamento dentro dos

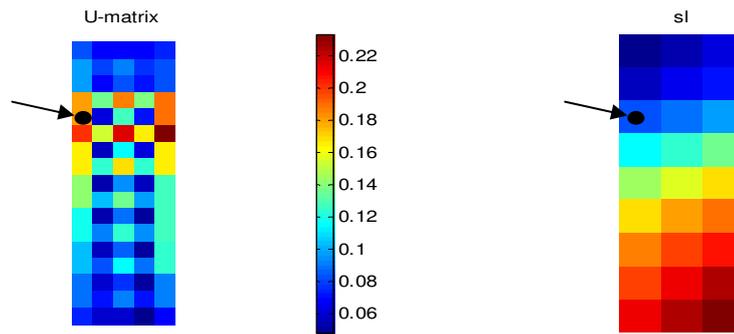
grupos (Figura 2.6). Os componentes planos apresentam os padrões da distribuição de dados do mapa auto-organizável.

FIGURA 2.6 Componente Plano



Na Figura 2.7, a seta aponta para o terceiro neurônio, tanto na Matriz-U quanto no componente plano ao lado, que possui a mesma dimensão do mapa original. Assim, dada a matriz $U = [u_1 \ u_2 \ u_3 \ u_4]$, a Matriz-U é uma matriz $U = [u_1 \ u_{12} \ u_2 \ u_{23} \ u_3 \ u_{34} \ u_4]$, onde u_i é qualquer tipo de média ou outra operação específica entre as distâncias dos neurônios i e j adjacentes, sendo u_{ij} a distância entre os pesos sinápticos dos neurônios i e j , calculada da seguinte maneira: $\|m_i - m_j\|$.

FIGURA 2.7 Localização da unidade de mapa que representa o neurônio 3



B. Algoritmo de Agrupamento Restrito à Vizinhaça

Em Kiang (2001), o algoritmo de agrupamento restrito à vizinhaça (*Contiguity-Constrained Grouping Algorithm*) agrupa neurônios após o processo de aprendizagem com base no critério da variância mínima. Em resumo, é calculada a variância do mapa inteiro, em seguida, a cada fusão de neurônios para formar um novo grupo, calcula-se a nova variância

global após a fusão, que se for menor que a anterior, permitirá que os neurônios continuem formando o grupo. Este método lembra o método estatístico Ward para análise de grupos (WARD JR, 1963), que também já foi aplicado neste sentido (DUTRA e COELHO, 2008).

C. Método Ward

O método de Ward, proposto em 1963, é fundamentado na mudança de variação entre grupos e dentro de grupos formados em um processo de agrupamento, uma vez que quando o nível de fusão dos grupos cresce, a similaridade dentro dos mesmos diminui (WARD JR, 1963). Neste processo, inicialmente, cada *elemento* é considerado um único conglomerado; em cada iteração, calcula-se dentro de cada grupo a somatória do quadrado da distância Euclidiana de cada elemento do grupo em relação ao vetor de médias do grupo (centroide). O valor desta somatória para cada grupo é novamente somado a cada fusão, mantendo-se a fusão caso a nova somatória seja menor que a anterior. Daí a denominação Método da Variância Mínima.

2.1.6 Quantização Vetorial

Os Mapas Auto-Organizáveis se baseiam em um método clássico de aproximação de sinal denominado quantização vetorial, o qual explora a estrutura subjacente de vetores de entrada com o fim de obter a compressão de dados. Este método forma usualmente uma aproximação quantizada para a distribuição de vetores de entrada, usando um número finito de vetores denominados *codebook* $m_i \in \mathfrak{R}^n$, $i = 1, 2, 3, \dots, k$. Uma vez definido o *codebook*, o processo de aproximação de x visa encontrar um *codebook* m_c mais próximo de x (no espaço de entrada), geralmente tomando como base o cálculo da distância Euclidiana, como observado na equação 2.14:

$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \text{ ou } c = \arg \min_i \{\|x - m_i\|\} \quad (2.14)$$

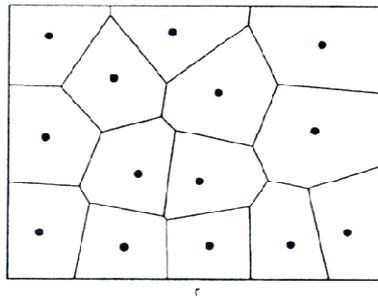
A seleção ótima de m_i minimiza a média quadrática esperada do erro de quantização (medida de distorção), que é definida como (KOHONEN, 2001):

$$E = \int \|x - m_c\|^2 p(x) dx \quad (2.15)$$

onde a integral é tomada sobre todo o espaço de entrada x , dx é uma notação para a diferencial do volume n -dimensional do espaço de integração, e $p(x)$ é a função densidade de probabilidade de x .

Ao aplicar o método, obtém-se um arranjo denominado “tecelagem” de Voronoi (*Voronoi tessellation*) (KOHONEN, 2001), que é um espaço bidimensional onde um número finito de *codebooks* ou vetores de Voronoi são mostrados como pontos relativos as suas coordenadas. Tal espaço é particionado em regiões limitadas por linhas (em geral, hiperplanos), de forma que cada partição (célula de Voronoi) contenha um vetor de Voronoi que é o “vizinho mais próximo” para qualquer vetor dentro da mesma partição. Todos os vetores x em cada partição correspondente da “tecelagem” de Voronoi (Figura 2.8) formam o Conjunto de Voronoi.

FIGURA 2.8 Tecelagem de Voronoi (KOHONEN, 2001)



A quantização vetorial (LVQ, *Learning Vector Quantization*) é uma técnica de aprendizagem supervisionada que usa informação sobre as classes de um problema de classificação para mover os vetores de Voronoi, melhorando a qualidade das regiões de decisão do classificador. Já o algoritmo SOM fornece um método aproximativo para calcular os vetores de Voronoi de uma maneira não-supervisionada.

2.2 Sistemas de Inferência Difusa

Os sistemas difusos (também conhecidos como nebulosos) se baseiam na lógica difusa, concebida por Lotfi Zadeh na década de 1960 (EL-HAWARY, 1998). Este tipo de lógica lida com a imprecisão das variáveis de entrada e saída do problema, definindo números e conjuntos difusos que podem ser expressos por meio de variáveis linguísticas (por exemplo, a variável temperatura pode ter os seguintes valores: baixo, médio e alto). Os conjuntos difusos são utilizados em regras difusas, obtidas através de especialistas humanos ou a partir do domínio de uma determinada área de conhecimento. Assim, eles podem ser usados para

modelar raciocínio aproximado, característico em atividades do cérebro humano como dirigir um carro, jogar futebol e outras.

Os sistemas difusos são baseados em conhecimento expresso por regras SE-ENTÃO difusas. A coleção de regras obtida é combinada em um único sistema. Cada sistema usa princípios diferentes para combiná-las, dando origem a diferentes tipos de sistemas difusos. Os tipos mais comumente usados na literatura são os sistemas Takagi-Sugeno e os sistemas Mamdani (CASTRO, 2004).

2.2.1 Teoria dos Conjuntos Difusos

A teoria clássica dos conjuntos permite que um conjunto possa ser descrito por uma função característica, que tem como únicos resultados possíveis 0 ou 1. Deixe \mathbf{U} ser o universo do discurso, que contém todos os possíveis elementos relativos a cada contexto particular. Um conjunto A definido sobre o conjunto \mathbf{U} é denotado pela função característica $\mu_A(x)$:

$$\mu_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad (2.16)$$

Na transição entre conjuntos crisp e conjuntos difusos, o elemento-chave é a teoria da possibilidade, que permite múltiplas possibilidades de solução definidas através de uma função de distribuição (função de pertinência), que indica a adequação de uma amostra a um conceito ou a possibilidade de um evento ocorrer; que difere da probabilidade de ocorrência de um evento (chance), porque um conjunto difuso não é criado a partir de evidência estatística e sim, de consenso ou opinião de especialistas (ZADEH, 1975; BARKAN e TRUBATCH, 1997).

Definição 2.1 Conjuntos Difusos

Conjuntos difusos são uma generalização de conjuntos *crisp* clássicos, capazes de lidar com verdades parciais ou imprecisas. Trata-se de funções que mapeiam um valor ou um membro do conjunto para um número entre 0 e 1, indicando seu grau de pertinência:

$$\mu_A(x) : x \rightarrow [0 \ 1] \quad (2.17)$$

O conjunto difuso A pode ser também representado por um conjunto de pares ordenados de um elemento genérico x e seu valor de pertinência apresentado:

$$A = \{(x, \mu_A(x)), x \in U\} \quad (2.18)$$

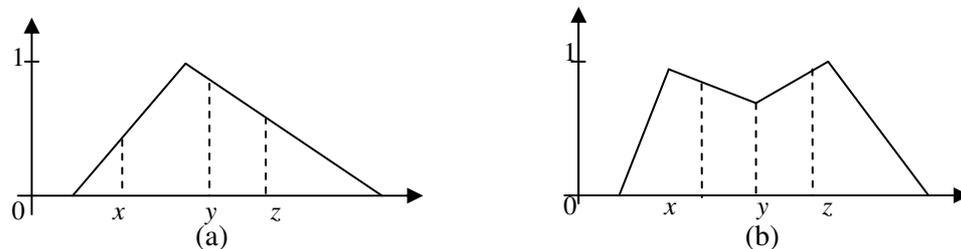
Função de Pertinência

Em virtude de um conjunto difuso ser descrito por uma função de pertinência, é importante descrever algumas de suas propriedades:

- O suporte da função de pertinência de um conjunto difuso A é a região do universo em que $\mu_A(x) > 0$.
- Um conjunto difuso normal é aquele, cuja função de pertinência possui ao menos um elemento x no universo com pertinência igual a 1. Caso nenhum elemento x apresente tal valor de pertinência, o conjunto difuso é dito subnormal.
- Um conjunto difuso convexo é descrito por uma função de pertinência, cujos valores de pertinência comportam-se de três formas: estritamente monotonicamente crescentes ou estritamente monotonicamente decrescentes ou, ainda estritamente monotonicamente crescentes, depois decrescentes e em seguida crescentes novamente. Assim, se A é um conjunto difuso e a relação $x < y < z$ existe em A , então (Figura 2.9) (ROSS, 2004):

$$\mu_A(y) \geq \min[\mu_A(x), \mu_A(z)] \quad (2.19)$$

FIGURA 2.9 Conjunto difuso convexo (a) e conjunto difuso não-convexo (b) (ROSS, 2004)



- Os pontos *crossover* de uma função de pertinência são os elementos do universo para os quais a função de pertinência de um conjunto A é igual a 0.5 ($\mu_A(x) = 0.5$).
- A altura de um conjunto difuso A é o valor máximo da função de pertinência.

- f) Um α -cut de um conjunto difuso A é um conjunto crisp A_α que contém todos os elementos do universo com valores de pertinência em A maiores ou iguais a α .

$$A_\alpha = \{x \in U \mid \mu_A(x) \geq \alpha\} \quad (2.20)$$

A forma geométrica de uma função de pertinência é a caracterização da imprecisão na variável difusa correspondente. A função de pertinência triangular é a função mais frequentemente utilizada, sendo considerada a mais prática, além da trapezoidal, função-s, função-pi e função-z (CASTRO, 2004).

Princípio da Extensão

O princípio da extensão para conjuntos difusos é uma identidade básica que permite que o domínio da definição de um mapeamento ou uma relação sejam estendidas a partir de pontos em U para subconjuntos difusos de U (ZADEH, 1975). Mais especificamente, suponha que f seja um mapeamento de U em V e A seja um subconjunto difuso de U tal que:

$$A = \mu_1 v_1 + \dots + \mu_n v_n \quad (2.21)$$

Então o princípio da extensão afirma que:

$$f(A) = f(\mu_1 u_1 + \dots + \mu_n u_n) \equiv \mu_1 f(u_1) + \dots + \mu_n f(u_n). \quad (2.22)$$

Desta forma, a imagem de A em f pode ser deduzida a partir do conhecimento das imagens de u_1, \dots, u_m em f . Este princípio permite o cálculo de restrições induzidas, de forma que o valor de uma variável linguística possa ser influenciado por outra. Isto quer dizer que por meio do princípio da extensão é possível criar funções, cujas entradas e saídas sejam difusas.

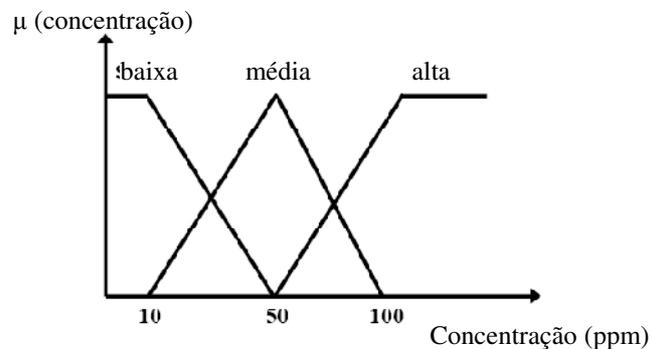
Definição 2.2: Variável Linguística

É uma variável cujos valores são palavras ou sentenças de uma linguagem natural ou artificial, por exemplo, *idade* é uma variável linguística, quando assume valores como jovem, muito jovem, velho, não muito velho, ao invés de valores numéricos (ZADEH, 1975). É caracterizada por (X, T, U, M) , onde X é o nome de uma variável; T é o conjunto de valores linguísticos que X pode assumir; U é o domínio dos valores da variável linguística; M é uma

regra semântica que relaciona cada valor linguístico em T com um conjunto difuso em U. A variável linguística permite a formulação de descrições vagas em linguagem natural em termos matemáticos precisos.

A Figura 2.10 apresenta um exemplo com três conjuntos difusos no universo de discurso U, representando o intervalo de valores possíveis de concentração de gás H₂ dissolvido em óleo de um transformador. “Concentração” é a variável linguística com três termos: “baixa”, “média” e “alta”, representados por conjuntos difusos com funções de pertinência mostradas na figura.

FIGURA 2.10 Conjuntos difusos para concentração de H₂



Definição 2.3: Modificadores Linguísticos

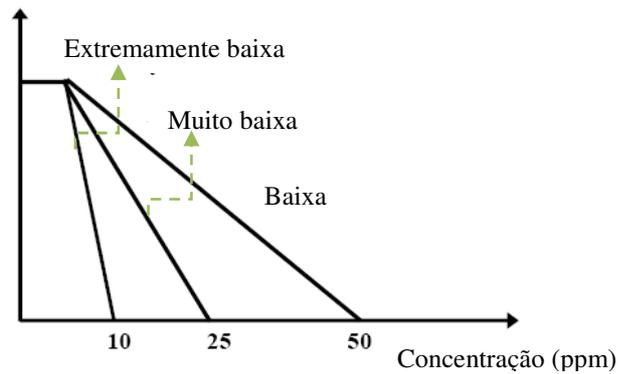
São operadores que alteram as funções de pertinência dos conjuntos difusos associados aos rótulos linguísticos. O significado de um conjunto transformado pode ser facilmente interpretado a partir do significado do conjunto original. Uma pequena lista de modificadores linguísticos e seu padrão em lógica difusa são listados na Tabela 2.1.

TABELA 2.1 Modificadores linguísticos

Modificadores	Função
Muito, extremamente	Concentração
Moderadamente	Diluição
Definitivamente, aproximadamente	Intensificação
Mais ou menos	Relaxação
Não	Negação
Acima, abaixo	Restrição

A Figura 2.11 apresenta um exemplo de modificador linguístico para o conjunto difuso “baixa”.

FIGURA 2.11 Modificadores para baixa concentração de gás H₂
 μ (concentração)



Definição 2.4: União Difusa – Norma-S

Seja $s: [0,1] \times [0,1] \rightarrow [0,1]$ ser um mapeamento que transforma as funções de pertinência de conjuntos difusos A e B na função de pertinência da união de A e B:

$$s[\mu_A(x), \mu_B(x)] = \mu_{A \cup B}(x). \quad (2.23)$$

A função s é qualificada como uma união difusa ou norma-s se satisfaz no mínimo os seguintes requisitos, sendo que a e b denotam a pertinência de algum conjunto difuso, isto é, $a = \mu_A(x)$ e $b = \mu_B(x)$.

- 1) $s(1,1)=1$, $s(0, a) = s(a,0) = a$ (condição limite)
- 2) $s(a,b) = s(b, a)$ (condição comutativa)
- 3) Se $a \leq a'$ e $b \leq b'$, então $s(a,b) \leq s(a', b')$ (condição não-decrescente – um decréscimo nos valores de pertinência em A ou B não pode produzir um acréscimo nos valores de pertinência de A união B).
- 4) $s(s(a,b), c) = s(a, s(b, c))$ (condição associativa).

A Tabela 2.2 lista algumas normas-s já propostas na literatura.

TABELA 2.2 Normas-S

Norma-S	
Soma de Einstein	$s(a,b) = \frac{a+b}{1+a+b}$
Soma Drástica	$s(a,b) = \begin{cases} a & \text{se } b = 0 \\ b & \text{se } a = 0 \\ 1 & \text{de outro modo} \end{cases}$
Soma Algébrica	$s(a,b) = a + b - ab$
Máximo	$s(a,b) = \max(a,b)$

Definição 2.5: Intersecção Difusa – Norma-T

Seja $t: [0,1] \times [0,1] \rightarrow [0,1]$ ser um mapeamento que transforma as funções de pertinência de conjuntos difusos A e B em função de pertinência da intersecção de A e B:

$$t[\mu_A(x), \mu_B(x)] = \mu_{A \cap B}(x). \quad (2.24)$$

A função t é qualificada como uma intersecção difusa ou norma-t se obedecer os seguintes requisitos:

- 1) $t(0,0) = 0$, $t(a,1) = t(1,a) = a$ (condição limite);
- 2) $t(a,b) = t(b,a)$ (condição comutativa)
- 3) Se $a \leq a'$ e $b \leq b'$, então $t(a,b) \leq t(a',b')$ (condição não-decrescente – um decréscimo nos valores de pertinência em A ou B não pode produzir um acréscimo nos valores de pertinência de A intersecção B).
- 4) $t(t(a,b), c) = t(a, t(b,c))$ (condição associativa).

A Tabela 2.3 lista algumas normas-t já propostas na literatura.

TABELA 2.3 Normas-T

Norma-T	
Produto de Einstein	$t(a,b) = \frac{ab}{2 - (a + b - ab)}$
Produto Drástico	$t(a,b) = \begin{cases} a & \text{se } b = 1 \\ b & \text{se } a = 1 \\ 0 & \text{de outro modo} \end{cases}$
Produto Algébrico	$t(a,b) = ab$
Mínimo	$t(a,b) = \min(a,b)$

Definição 2.6: Complemento Difuso

Seja $c:[0,1] \rightarrow [0,1]$ ser um mapeamento que transforma a função de pertinência do conjunto A na função de pertinência do complemento de A:

$$c[\mu_A(x)] = \mu_{\bar{A}}(x). \quad (2.25)$$

A função c é qualificada como um complemento difuso se satisfaz os seguintes requisitos:

- 1) $c(0) = 1$ e $c(1) = 0$ (condição limite)
- 2) Para todo $a, b \in [0,1]$, se $a < b$ então $c(a) \geq c(b)$ (condição de não incremento)

Definição 2.7: Classe Associativa – Lei de DeMorgan

Para cada norma-s, há uma norma-t associada, o que quer dizer que há um complemento difuso tal que os três juntos satisfaçam a Lei de DeMorgan. Especificamente, a norma-s $s(a,b)$, a norma-t $t(a,b)$ e o complemento difuso $c(a)$ formam uma classe associativa se:

$$c[s(a,b)] = t[c(a), c(b)] \quad (2.26)$$

Definição 2.8: Base de Regras Difusas

Uma base de regras difusas consiste de um conjunto de regras difusas SE-ENTÃO que especificam uma relação linguística entre o rótulo linguístico das variáveis de entrada e das variáveis de saída do sistema. É o elemento fundamental do sistema difuso, uma vez que todos os outros componentes são usados para implementar essas regras de uma maneira

razoável e eficiente. Especificamente, a base de regras difusa compreende as seguintes regras SE-ENTÃO difusas:

$$\begin{aligned}
 & \text{Regra } R^1 : \text{SE } x_1 \text{ é } A_1^1 \text{ e } \dots x_n \text{ é } A_n^1 \text{ ENTÃO } y \text{ é } B^1 \\
 & \dots \\
 & \text{Regra } R^m : \text{SE } x_1 \text{ é } A_1^m \text{ e } \dots x_n \text{ é } A_n^m \text{ ENTÃO } y \text{ é } B^m
 \end{aligned} \tag{2.27}$$

onde A_i e B são conjuntos difusos em $U \subset R$ e $V \subset R$, respectivamente, e $X = (x_1, x_2, \dots, x_n) \in U$ e $y \in V$ são as entradas e as saídas do sistema difuso respectivamente.

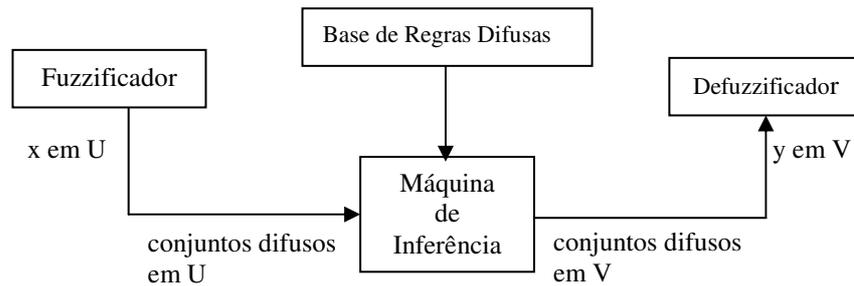
A parte SE da regra é chamada de premissa ou antecedente, enquanto a parte ENTÃO é chamada de conclusão ou consequente da regra.

2.2.2 Sistemas Baseados em Regras Difusas

Um sistema baseado em regras difusas também conhecido como Sistema de Inferência Difusa é composto de quatro blocos funcionais, como mostra a Figura 2.12:

- *Fuzzificação.* Normalmente, as entradas para o sistema difuso são valores crisp, que têm de ser convertidos em conjuntos difusos. O bloco de *fuzzificação* transforma as entradas *crisp* em graus de similaridade relativos a valores linguísticos.
- *Banco de dados e base de regras.* O *banco de dados* define funções de pertinência de conjuntos difusos usados nas regras SE-ENTÃO difusas que compõem a base de regras. Usualmente, a base de regras e o banco de dados são referidos como *base de conhecimento*.
- *Máquina de Inferência.* Realiza o processamento sobre as regras difusas e produz um valor difuso para a saída do sistema.
- *Defuzzificação.* Converte um conjunto de variáveis difusas em valores *crisp*, a fim de permitir que a saída do sistema difuso seja aplicada a outro sistema não-difuso.

FIGURA 2.12 Configuração Básica de um sistema difuso com fuzzificador e defuzzificador

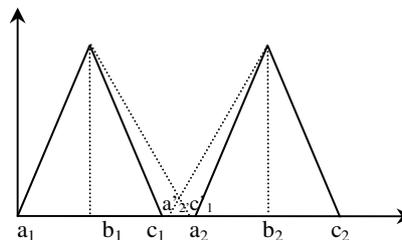


2.2.3 Propriedades dos Sistemas de Inferência Difusa

Definição 2.9: Um conjunto de regras SE-ENTÃO é completo, se para todo e qualquer $x \in U$, existe ao menos uma regra na base de regras difusas que é ativada. Neste caso, o grau de pertinência para o domínio especificado é diferente de zero.

Definição 2.9.1: Quando existem pontos no domínio (x^*), para os quais o grau de pertinência é igual a zero, tais bases de regras difusas são chamadas de esparsas ou incompletas. Quando as abordagens clássicas de inferência não geram uma saída significativa de forma que uma regra seja ativada, utiliza-se o método da interpolação para criar partições inteiras (conjuntos difusos), com intuito de eliminar as faltas no domínio (JOHANYÁK e KOVÁCS, 2006). De acordo com Jin (2003), o método mais simples para resolver o problema da base de regras esparsas é a Correção (*Repairing*), que modifica os extremos de funções de pertinência adjacentes para eliminar faltas de domínio nas partições difusas com alguma sobreposição (Figura 2.13).

FIGURA 2.13 Correção de partições difusas incompletas



Definição 2.10: Um conjunto de regras difusas SE-ENTÃO é consistente se não houver regras com a mesma parte SE para diferentes partes ENTÃO.

Definição 2.11: Um sistema difuso é contínuo, se não existem regras adjacentes, cujos conjuntos difusos da parte ENTÃO não apresentam intersecção.

2.2.4 Projeto de Sistemas Difusos Orientados a Dados

Quando sistemas difusos são projetados a partir de dados, ao invés do conhecimento de um especialista, os mesmos podem apresentar facilidade de interpretação, mas nem sempre, são sistemas transparentes. A transparência é uma medida da validade da interpretação linguística do sistema difuso (RIID e RUSTERN, 2000). Para obter sistemas difusos com bases de regras transparentes, geralmente, é necessário simplificar a base de regras.

Os sistemas podem apresentar redundância entre os conjuntos difusos (SETNES, 1998):

- Tais conjuntos podem ser similares entre si, neste caso, necessitam de uma medida de similaridade para serem unidos ou substituídos um pelo outro;
- Um ou mais conjuntos difusos podem ser similares ao universo do discurso, ou seja, para todo o universo, o valor da função de pertinência é aproximadamente igual a 1, devendo ser removidos;
- Um conjunto difuso pode ser um *singleton* (conjunto com um único valor de entrada); em alguns casos pode ser removido.

A precisão e a cobertura do domínio são os principais aspectos que devem ser considerados quando da simplificação de uma base de regras difusas.

2.2.5 Sistema Difuso do Tipo Takagi-Sugeno

Os sistemas de inferência podem ser categorizados através de duas famílias:

- 1) A família que inclui modelos linguísticos baseados em coleções de regra SE-ENTÃO, cujos antecedentes e consequentes utilizam valores difusos tais como inferência difusa Mamdani e
- 2) A família que usa uma estrutura de regras que tem antecedente difuso e consequente funcional (*crisp*).

Para a segunda categoria, tem-se o sistema difuso do tipo Takagi-Sugeno (TS), que é construído com regras da seguinte forma (TAKAGI e SUGENO, 1985):

$$\text{Regra } R^j : SE \ x_1 \ \acute{e} \ A_{1j} \ E \ x_i \ \acute{e} \ A_{ij} \ ENT\tilde{A}O \ y_j = g(x_1, \dots, x_i) \quad (2.28)$$

onde A_{ij} é um conjunto difuso e x_i é a entrada do sistema. O consequente da regra é uma função linear ou não-linear das variáveis de entrada.

O modelo difuso TS foi proposto por Takagi, Sugeno e Kang para formalizar uma abordagem de sistema para gerar regras difusas a partir de um conjunto de entrada-saída. O modelo TS é também conhecido como modelo Takagi-Sugeno-Kang (TSK). Quando y_j é um polinômio de primeira ordem, tem-se o modelo Takagi-Sugeno de primeira ordem. Quando y_j é uma constante, tem-se o modelo Takagi-Sugeno de ordem zero, que pode ser visto como um caso especial do modelo de inferência de Mamdani com o consequente como um *singleton*. O modelo Takagi-Sugeno de ordem zero é construído com regras da seguinte forma:

$$\text{Regra } R^j : SE \ x_1 \ \acute{e} \ A_{1j} \ E \ x_i \ \acute{e} \ A_{ij} \ ENT\tilde{A}O \ y_j = c_j \quad (2.29)$$

O resultado de cada regra é calculado por:

$$v_j = \bigcap_{i=1}^n \mu_{ij}(x_i) \quad (2.30)$$

onde $\mu_{ij}(x_i)$ é a função de pertinência associada ao conjunto difuso A_{ij} e \cap representa o operador produto (operador e).

A saída do sistema é computada como a média ponderada do y_j de acordo com:

$$f(x) = \frac{\sum_{j=1}^N y_j v_j}{\sum_{j=1}^N v_j} \quad (2.31)$$

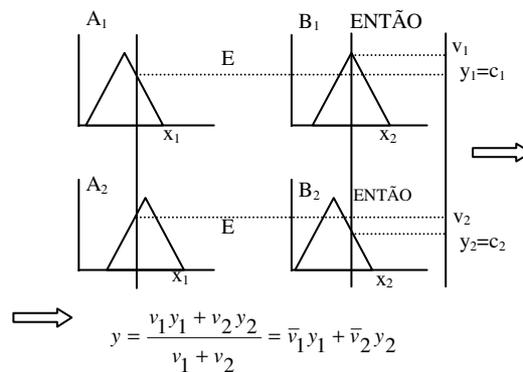
onde N é o número de regras do sistema.

A saída pode ser também calculada por:

$$f(x) = \sum_{j=1}^N y_j \bar{v}_j \quad (2.32)$$

A Figura 2.14 ilustra o mecanismo de raciocínio para o sistema TS de ordem zero, que é o modelo de interesse para esta tese.

FIGURA 2.14 Modelo Takagi-Sugeno de Ordem Zero



2.3 Conclusão do Capítulo

As RNAs são aplicadas a casos em que não há um modelo definido para explicar um determinado comportamento de um domínio de problema. O que se tem é um conjunto de dados extraídos do ambiente através de fatos ou sensores projetados para sondar o ambiente. A RNA adquire conhecimento a respeito destes fatos ou dados pelo uso de um algoritmo de aprendizagem. Tal conhecimento se materializa através das forças de conexões entre neurônios (pesos sinápticos).

As RNAs, no entanto, não possuem a capacidade de explicar as razões para os resultados alcançados, ou seja, para explicar o conhecimento aprendido em decorrência da carência de uma representação estruturada deste conhecimento. Com os SOMs, não é diferente. Eles permitem a compressão de dados, alocando as amostras mais similares em neurônios vizinhos, inclusive para visualizar melhor os agrupamentos, métodos como a Matriz-U são aplicados, não sendo suficientes para indicar como foram tomadas as decisões de agrupamento.

Ao contrário das RNAs, os sistemas difusos têm grande capacidade de explanação, pois são mais inteligíveis para o homem do que as redes neurais, pelo fato de descreverem o

comportamento de variáveis linguísticas. Assim, os sistemas difusos podem ser projetados, utilizando o conhecimento de um especialista sobre um determinado problema. Também é possível projetar sistemas difusos a partir de um conjunto de dados de entrada-saída. Neste caso, a manutenção da facilidade de interpretação pode ser feita através da simplificação de regras obtidas logo após a construção do sistema.

Os sistemas difusos apresentam como principais desvantagens: não possuem a capacidade de generalização, isto é, geram respostas apenas para o que está escrito; e caracterizam-se pela dificuldade em desenvolver o modelo difuso, porque dependem da existência de um especialista para definir as regras do sistema. Desta forma, sistemas híbridos (por exemplo, neuro-difusos) apresentam-se como uma alternativa mais popular para aplicações específicas, uma vez que podem unir a característica de “aprender” a partir de dados das RNAs com a capacidade de explicar de forma explícita o conhecimento por meio de regras SE-ENTÃO dos sistemas difusos.

3. Estado da Arte de Extração de Conhecimento de Mapas Auto-Organizáveis

Redes Neurais Artificiais têm sido aplicadas com sucesso a vários tipos de problemas complexos como reconhecimento de padrões e aproximação de função. Entretanto, o fato de não possuírem capacidade de explicação/explanação em relação aos resultados obtidos tem sido uma barreira para sua aceitação em alguns tipos de aplicação.

Nos últimos anos, muitos trabalhos vêm sendo desenvolvidos buscando a solução para este problema. Novas metodologias vêm sendo criadas principalmente com o objetivo de converter o modelo neural para uma representação que facilite a compreensão dos resultados obtidos (KAHRAMANLI e ALLAHVERDI, 2009; HUYNH e. REGGIA, 2009; QUTEISHAT e PENG LIM, 2008, CASTRO et al., 2007; FUNG et al., 2005; BROUWER, 2005). Estas novas metodologias desenvolvidas fazem parte de uma área de estudo que é referenciada como *extração de regras/conhecimento de redes neurais artificiais*.

A pesquisa sobre extração de regras de redes neurais se originou no final da década de 1980, quando Gallant (2005) publicou um trabalho apresentando uma rotina para extrair regras proposicionais a partir de uma rede simples. Desde então, muitos trabalhos têm sido apresentados na área, com o desenvolvimento de algoritmos que apresentam a saída da RNA como um conjunto de regras, usando lógica proposicional, lógica difusa ou lógica de primeira ordem (Da SILVA et al., 2012; HUNG e HUANG, 2011; KAHRAMANLI e ALLAHVERDI, 2009; HUYNH e. REGGIA, 2009; QUTEISHAT e PENG LIM, 2008, CASTRO et al., 2007; FUNG et al., 2005; BROUWER, 2005).

Neste capítulo, uma revisão sobre extração de regras a partir de redes neurais será apresentada, a fim de fornecer ao leitor uma visão geral sobre a área, na qual a proposta deste trabalho está situada. Na seção 3.1, a tarefa da extração de regra a partir de redes neurais é definida e a taxonomia para avaliar os algoritmos de extração de regras é apresentada. Na seção 3.2, alguns trabalhos publicados sobre extração de regras a partir dos mapas auto-organizáveis de Kohonen são apresentados com o intuito de revelar o estado da arte na área de pesquisa deste trabalho.

3.1 Extração de Regras/Conhecimento de Redes Neurais Artificiais

As redes neurais artificiais seriam mais populares se a capacidade de explicação/explanação dos seus resultados se tornasse parte integral de sua funcionalidade.

Com o objetivo de dar às redes neurais tal capacidade, muitas pesquisas têm sido desenvolvidas no campo da extração de regras a partir de redes neurais. Craven (1996) definiu a tarefa de extração de regra como segue:

Dada uma rede neural treinada e o dado sobre o qual foi treinada, produz-se uma descrição da hipótese da rede que seja compreensível e que se aproxime do comportamento previsível da rede.

A facilidade de explicação do conhecimento obtido, após o treinamento da rede, pode ser considerada a principal motivação para a pesquisa nesta área, porém a extração de regras a partir de redes neurais também tem algumas outras vantagens tais como:

- **Descobrir importantes aspectos da entrada.** Descobrir características de entrada importantes para o resultado ou entradas que acrescentam apenas ruído não são tarefas consideradas fáceis. Com as regras extraídas a partir de redes neurais, pode-se ter uma compreensão mais profunda do relacionamento entrada-saída e pode-se tentar encontrar as características que criam ruído.
- **Melhoria da generalização da RNA.** Através da análise das regras extraídas da RNA, as deficiências do conjunto original de treinamento de dados podem ser identificadas. As regiões que não são representadas propriamente no conjunto de treinamento podem ser encontradas e assim a generalização da rede pode ser melhorada pela adição/aperfeiçoamento de um novo dado representativo do problema.
- **Descoberta de conhecimento.** A RNA é eficiente para descobrir dependências desconhecidas e relacionamentos entre os dados do problema. As regras extraídas a partir da RNA podem revelar esta descoberta, cuja importância não foi previamente reconhecida.
- **Aquisição de conhecimento em sistemas especialistas.** A aquisição de conhecimento para o desenvolvimento de sistemas especialistas não é uma tarefa fácil, já que a base de conhecimento usada neste processo é geralmente adquirida pelo questionamento a um especialista humano. Este especialista nem sempre está apto a esclarecer seu conhecimento sobre o problema na forma de regras crisp. Como a RNA aprende a partir de exemplos, todo conhecimento adquirido sobre o problema, após se extraírem regras das RNAs, pode ser usado para ajudar na construção de sistemas especialistas.

- **Validação.** Analisando as regras extraídas, os usuários podem compreender como a RNA chegou a uma decisão específica e, conseqüentemente, eles podem ganhar mais confiança nos resultados obtidos. Se os usuários podem validar os resultados da RNA, então eles devem estar aptos a interagir de forma competente e eficiente com o sistema.

Andrews et al. *apud* Castro (2004) sugeriu uma taxonomia para categorizar as numerosas contribuições sobre algoritmos de extração de regras. Esta taxonomia pode ser considerada como o *framework* mais aceito nesta área até agora e incorpora os seguintes cinco critérios de classificação:

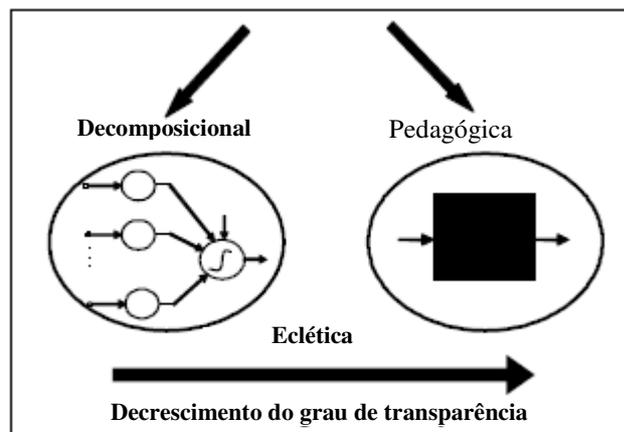
1. **O poder de expressão das regras extraídas.** Este critério está relacionado ao conhecimento simbólico apresentado ao usuário. Três tipos de regras são sugeridos:
 - Regras simbólicas convencionais (Booleana, proposicional);
 - Regras baseadas em conjuntos e lógica difusa e
 - Regras expressas na forma de lógica de primeira ordem.
2. **A qualidade da regra extraída.** A qualidade da regra pode ser considerada um dos mais importantes critérios de avaliação para algoritmos de extração de regras. Quatro medidas para avaliação da qualidade das regras extraídas são sugeridas: fidelidade, acurácia, consistência e facilidade de compreensão.
 - *Fidelidade.* Determina se as regras representam o comportamento da RNA de forma adequada, quando aplicadas às amostras de treinamento e teste. Alta fidelidade é um indicativo de que o sistema de regras capturou toda a informação embutida na RNA e, como consequência, pode responder corretamente para amostras da mesma maneira que a rede neural.
 - *Acurácia.* Determina a habilidade da representação extraída para decidir de forma precisa a respeito de casos desconhecidos. Portanto, a precisão (*accuracy*) é um indicativo da capacidade de generalização das regras extraídas.
 - *Consistência.* Descreve como as regras extraídas em sessões de treinamento distintas produzem o mesmo grau de precisão (*accuracy*).
 - *Facilidade de compreensão.* Determina se as regras extraídas são humanamente compreensíveis. É frequentemente definida pelo número de regras extraídas e pelo número de antecedentes de cada regra. É claro que as estruturas com um

pequeno conjunto de regras e antecedentes são mais compreensíveis para seres humanos do que o contrário.

3. A transparência. Classifica a técnica de extração de regra baseada na granularidade da rede neural. De acordo com a transparência, a extração de regras a partir da RNA pode ser classificada como:

- Abordagem *decomposicional* ou *caixa-branca* considera a extração de regra como um processo de pesquisa que mapeia a estrutura interna de uma rede neural treinada para um conjunto de regras. As regras são extraídas no nível de granularidade mínimo, isto é, a análise dos valores numéricos da rede tal como valores de ativação de neurônios das camadas escondidas e de saída e os pesos de conexões entre eles são usados para extrair as regras diretamente. As regras são extraídas para cada unidade escondida e da camada de saída separadamente, e o sistema de regra para a rede inteira é derivado dessas regras.
- Abordagem *pedagógica* ou *caixa-preta* não desmonta a arquitetura da rede neural treinada. Ao invés disso, considera a RNA como uma entidade e tenta extrair regras que poderiam explicar sua função. A RNA é tratada como uma “caixa-preta”, na qual as regras extraídas descrevem o relacionamento global entre as variáveis da entrada e da saída da rede neural.
- Abordagem *eclética* incorpora elementos dos modelos decomposicional e pedagógico. A Figura 3.1 mostra que há maior transparência na abordagem decomposicional do que na abordagem pedagógica.

FIGURA 3.1 Critério de transparência



- Abordagem *composicional* faz uma extensão na taxonomia de Andrews et al. para acomodar a extração de representações de máquinas de estado finito a partir de redes neurais recorrentes (TICKLE et al., 1998). Esta técnica analisa todos os neurônios recorrentes, a fim de encontrar grupos de ativação de neurônios na camada oculta, ou seja, não trabalha com neurônios individuais, que se agregam para formar um relacionamento global como ocorre na abordagem decomposicional.

4. Complexidade do Algoritmo. O número de cálculos requeridos para a tarefa (complexidade de tempo) e o montante de espaço de armazenamento usado (complexidade de espaço) são geralmente usados como medida da eficiência de um algoritmo. A complexidade de tempo é o mais importante fator na medida da eficiência da extração de regra do que complexidade de espaço. De fato, não se pode levantar a questão da complexidade do espaço, quando a eficiência do algoritmo de extração de regra é medida. A complexidade de tempo é um importante fator, já que os algoritmos de extração de regras são frequentemente baseados no teste de um grande número de combinações de entradas e parâmetros de redes tais como: o número de camadas da rede, neurônios por camada, conexões entre camadas, número de amostras de treinamento, atributos de entrada e valores por atributo de entrada. De qualquer forma, o algoritmo desenvolvido para extração de regra deve ter uma complexidade computacional baixa.

5. Portabilidade ou generalidade. Este critério avalia a extração de regra da RNA em termos do grau de aplicação do algoritmo em um conjunto de arquiteturas e tipos diferentes de treinamento. Craven et al. *apud* Craven e Shavlik (1999) afirmam que, para que os métodos de extração de regra tenham grande impacto, devem ter um nível alto de generalidade, isto é, qualquer que seja o método desenvolvido para extrair regra, este deve ser aplicável a qualquer que seja a rede neural. Em suma, eles devem ser aplicáveis à RNA desenvolvida por outros sem qualquer intenção inicial de aplicar métodos de extração de regras.

De acordo com Neumann (1998), os seguintes aspectos devem ser considerados para uma maior generalidade:

- Nenhuma restrição em relação à arquitetura da RNA;

- Nenhuma restrição em relação ao treinamento da RNA ou hipótese de como a rede foi construída e como os pesos e os *biases* foram ajustados antes da regra de extração;
- Nenhuma modificação da estrutura da RNA e dos parâmetros durante a extração de regra;
- Nenhuma restrição sobre o caráter e tamanho do domínio do problema. Domínios podem conter atributos discretos, contínuos e mistos.

Refinamento de Regras

O refinamento de regras consiste em utilizar conhecimento anterior sobre o domínio do problema para direcionar o aprendizado da rede neural. Isto conduz a um tempo de treinamento menor e à melhoria da generalização da rede (CASTRO et al., 2007). Trata-se na realidade do processo inverso ao da extração de regras.

Segundo Tickle (1998), a utilização de conhecimento anterior pode ocorrer: na arquitetura da rede; na escolha do algoritmo de aprendizado; na função de ativação da camada escondida e unidades de saída e nos pesos sinápticos. O refinamento de regra é um processo de três passos: codificação de uma dada base de regras (conhecimento simbólico) dentro da rede antes de começar o treinamento por meio da programação de pesos; treinamento da rede com exemplos desenvolvidos a partir do domínio do problema e extração do conhecimento refinado após o treinamento ter sido completado.

3.2 Extração de Conhecimento de Mapas Auto-Organizáveis de Kohonen – Estado da Arte

Nesta seção, serão apresentados alguns dos importantes trabalhos já desenvolvidos a respeito de extração de regras a partir de Mapas Auto-Organizáveis de Kohonen.

3.2.1 Extração de Regras Proposicionais

Em Hung e Huang (2011), é apresentado um modelo de extração de regras a partir de um SOM unidimensional, que envolve três subtarefas. A primeira tarefa visa formar grupos, encontrando um limite para cada grupo do SOM, cujos rótulos de classes dominantes sejam idênticos nos neurônios vizinhos. A segunda tarefa consiste em selecionar atributos

significativos para cada grupo. Um valor limite (*threshold*) para medir a importância do atributo é definido para o cálculo da diferença entre a matriz unificada do atributo e a do grupo. A última tarefa é a extração de uma regra para cada grupo, dependendo da configuração do mapa. As regras são classificadas de acordo com o número de atributos significativos, e as bordas são ajustadas. O modelo foi aplicado à base de dados Iris, da qual 3 regras proposicionais foram obtidas com 92,5% de acertos na base de treinamento e 100%, na base de teste.

Em Hung e Huang (2010), um mapa unidimensional é usado para a extração de regras proposicionais, uma vez que determinar limites fechados para grupos em um mapa bidimensional não é uma tarefa trivial. Neste método, o número ótimo de grupos, o centro de cada grupo e as unidades de mapa limites de cada grupo são determinados por meio da técnica de otimização *Particle Swarm*¹ (PSO) com a função de avaliação (*compact-separate measure* (CS)), que é um paradigma de computação evolucionária relativa a algoritmos genéticos. São selecionados atributos significativos para cada unidade de mapa, que é limite do grupo através da Matriz-U, a fim de que as regras sejam geradas sobre estes atributos. É um método bastante semelhante ao de Malone et al. (2006).

Em van Heerden e Engelbrecht (2009), é proposto um *framework* híbrido (*Hybrid*) para extração de regra, que explora características de mapeamento do SOM, permitindo sua combinação com qualquer algoritmo padrão de extração de regra. Assim, qualquer vantagem ou desvantagem do algoritmo selecionado influencia no desempenho do *Hybrid*. A entrada *Hybrid* é um conjunto de registros; cada registro é composto pelos atributos ou variáveis do problema associados aos componentes dos vetores de peso de cada neurônio, enquanto o rótulo do neurônio é indicado como classe daquele registro. Em seguida, um algoritmo de extração de regra é aplicado a este conjunto de registros, gerando uma árvore de decisão. No entanto, a maior parte das configurações do *Hybrid* apresentou resultados de baixo desempenho, levantando a hipótese de que os pesos dos neurônios não são suficientes para descrever o conjunto de dados de treinamento, já que são uma aproximação do mesmo.

Em Malone et al. (2006), a extração de regras SE-ENTÃO proposicionais é realizada a partir dos parâmetros internos do mapa auto-organizável. Primeiro, identificam-se as bordas

¹ Nesta abordagem, um grupo de indivíduos denominados partículas estão sobre o espaço de busca, sendo cada partícula uma solução candidata para o problema da otimização. A posição da partícula é influenciada por sua experiência e pela posição da melhor partícula vizinha. Quando a vizinhança de uma partícula é o *swarm* (grupo) inteiro, a melhor posição na vizinhança é referenciada com a melhor partícula global, e o algoritmo resultante é referenciado como um *gbest* PSO. Quando as vizinhanças menores são usadas, o algoritmo é geralmente referenciado como *lbest* PSO. O desempenho de cada partícula (o grau de aproximação da partícula do ótimo global) é medida usando uma função de avaliação, que varia dependendo do problema de otimização.

de grupos a partir da Matriz-U, calculando o maior valor da diferença limite (BDV – *Boundary Difference Value*). BDV é a diferença entre a média das distâncias entre a unidade corrente e duas outras unidades vizinhas candidatas à borda, dividida pela média da distância das unidades vizinhas restantes. As unidades com maiores BDVs são selecionadas para formarem uma borda. Em seguida, BDVs são calculados para componentes planos. A coincidência entre as bordas apontadas pelo componente plano e pela Matriz-U indica os atributos que serão usados para compor cada regra sobre os limites de cada grupo. A vantagem desta abordagem é que as regras geradas são de fácil compreensão, pois para o caso da base de dados Iris, foram geradas apenas 3 regras para 3 classes. Por ser uma abordagem decomposicional, possui transparência. A desvantagem é a complexidade do cálculo das bordas. Por outro lado, não se pode discutir a respeito da taxa de sucesso, pois foi realizado apenas teste com dados de treinamento. Em Pateritsas et al. (2007), produz-se um aumento da taxa de sucesso de 95,3% para 96,67% para a base dados Iris, ao aplicar um procedimento de simplificação de regras.

Em Darrah et al. (2004), é apresentado um algoritmo que mostra como extrair regras de um tipo de mapa auto-organizável de Kohonen chamado de rede neural com estrutura de célula dinâmica (*Dynamic Cell Structure (DCS) Neural Network*), que inclui aprendizagem Hebbiana e mudança da estrutura da rede, e é aplicada ao controle de vôo adaptativo em tempo real. A técnica consiste na modificação do algoritmo LREX de McGarry et al. *apud* Darrah et al. (2004), usado para extrair regras de uma rede neural de base radial (RBF). Os pesos do DCS, após treinamento, são usados como entrada para o algoritmo por serem os centros das regiões de Voronoi. O BMU (*Best Matching Unit*) correspondente a cada dado de entrada é registrado durante treinamento e também usado como entrada para o algoritmo. Os dados de treinamento são divididos em regiões baseadas no *Best Matching Unit* (BMU). Então para cada região, x_{menor} é o menor valor da variável independente que tem um BMU particular e x_{maior} é o maior valor daquela variável independente que tem aquele mesmo BMU particular. Estes dois números formam limites para os intervalos no antecedente. (Exemplo: $\text{variável} \geq x_{\text{menor}} \text{ AND } \leq x_{\text{maior}}$). Um intervalo é determinado para cada uma das variáveis independentes e as proposições são ligadas por conectivos E para formar um antecedente completo. A vantagem deste algoritmo está na possibilidade de um treinamento adaptativo. A desvantagem é a baixa taxa de sucesso no treinamento em relação as outras abordagens citadas.

Em Hani et al. (2001), a extração de regras proposicionais do SOM começa com o pré-processamento da base de dados composto de: carga de arquivo de dado, normalização e

extração de características. O dado pré-processado é utilizado no treinamento do mapa auto-organizável. Em seguida, o algoritmo k-means é aplicado aos pesos sinápticos resultantes do treinamento para definir os grupos. Para cada grupo, regras proposicionais são extraídas através da atribuição de valores 0 ou 1 para cada sinapse (0 significando a negação do antecedente da regra e 1, o contrário), dependendo se a sinapse que conecta a entrada ao neurônio é ativadora ou inibitória (pesos menores ou iguais a 0.5 tomam valor igual a 0 e acima de 0.5 tomam valor igual a 1). Redundâncias devem ser eliminadas e novos agrupamentos podem ser realizados, caso sejam requeridos pelo usuário. Para testar as regras, um sistema especialista foi desenvolvido. Nota-se que o sistema gera um grande número de regras a partir dos experimentos apresentados, o que prejudica a compreensão das regras geradas.

3.2.2 Extração de Regras Difusas

Em Naresh et al. (2008), regras difusas com função de pertinência Gaussiana são extraídas dos dados resultantes do treinamento de um mapa de Kohonen, para serem aplicadas ao problema da análise dos gases dissolvidos em óleo, considerando 7 gases de entrada. Um conjunto de dados é treinado no mapa de Kohonen com 117 amostras, das quais 30 são destinadas a teste. O algoritmo *subtractive clustering* (CHIU, 1994) é executado sobre o resultado do treinamento, para retornar as amostras que apresentam maior potencial para representarem os centros dos grupos, que serão os centros da função de pertinência Gaussiana na parte antecedente da regra. As regras obtidas devem passar ainda por um processo de otimização para definir os melhores parâmetros da função Gaussiana, utilizando a rede neural de retropropagação. Foi alcançada uma taxa de sucesso de 50%, antes da otimização de parâmetros, e 96,67% após a otimização.

Em Wong et al. (2001), propõe-se uma técnica para extrair regras difusas diretamente dos pares entrada-saída, utilizando uma rede neural auto-organizável e regras de associação para construir uma base de regras difusas. O SOM é usado para classificar os dados de saída. São tomados os valores mínimos e máximos de cada grupo. Para obter regras de associação, definem-se as partições e o número de intervalos apropriados. A partição *equi-depth* é usada. Após a identificação das partições, os conjuntos de itens (*itemsets*) são buscados. Os *itemsets* frequentes descrevem a combinação de itens que tem suporte acima do suporte mínimo especificado pelo usuário. As regras de associação são usadas para encontrar os

relacionamentos entre o espaço de entrada e a classificação da saída, que são subsequentemente convertidos em regras difusas com funções de pertinência triangular, as quais são agrupadas para formar funções de pertinência trapezoidais. Nesse trabalho, é apresentado um experimento com uma função matemática. A restrição do método está no fato de que os algoritmos de regras de associação exigem um alto tempo de processamento, particularmente quando o número de itens é muito alto. A participação do SOM é restrita.

Em Wang (1997), é apresentado um método para projetar sistemas difusos usando técnicas de agrupamento. O algoritmo do vizinho mais próximo é utilizado para determinar os centros das funções de pertinência Gaussiana, que irão compor um sistema difuso do tipo Takagi-Sugeno de ordem zero. A limitação desta abordagem se dá pelo fato de só poder ser aplicada a pequenos conjuntos de dados, caso contrário, a sobreposição das funções de pertinência impedem um bom resultado.

3.3 Conclusão do Capítulo

Neste capítulo, foram apresentados os fundamentos teóricos da área de extração de regras, cujo grande desafio é encontrar uma forma adequada para interpretar os resultados da rede de forma que se possa reproduzi-los através de regras; sejam elas regras simbólicas convencionais, baseadas em conjuntos crisp ou difusos ou expressas na forma de lógica de primeira ordem.

Conforme afirmado por Ultsch e Korus *apud* Fung et al. (2005), poucos trabalhos têm sido apresentados especificamente para extração de regras de Mapas Auto-Organizáveis. Observam-se poucas propostas relacionadas à extração de regras difusas. Com isto, são importantes novos trabalhos neste campo de pesquisa.

A Tabela 3.1 faz um resumo das características apresentadas pelas abordagens decomposicionais e pedagógicas.

TABELA 3.1 Tabela resumida das abordagens estudadas no Capítulo 3

Artigo	Abordagem quanto à transparência	Método de definição de borda	Portabilidade	Formato da Regra	Qualidade das Regras			Complexidade
					Taxa de Sucesso	Fidelidade (treinamento)	Facilidade de Compreensão	
Hung e Huang (2011)	Pedagógica	Matriz-U, inclusive de atributos	Independente	Proposicional	100%	92,5%	3 regras (Iris)	Média
Hung e Huang (2010)	Decomposicional	Matriz-U e PSO	Independente	Proposicional	96,67%	85,83%	3 regras (Iris) - Alta	Alta
Malone et al. (2006)	Decomposicional	Matriz-U e BDV	Independente	Proposicional	-	96,67%	3 regras (Iris) - Alta	Alta
van Heerden e Engelbrecht (2009)	Decomposicional	Hybrid	Independente	Proposicional	85,72%	89,33%	-	Alta
Darrah et al. (2004)	Decomposicional	DCS	Independente	Proposicional	-	82%	-	Alta
Hani et al. (2001)	Decomposicional	K-means	Independente	Proposicional	-	-	65 regras para 15 grupos (diagnóstico de 5 tipos de doenças hepáticas) - Baixa	Alta
Naresh et al. (2008)	Pedagógica	<i>Subtractive Clustering</i>	Independente	Fuzzy	96,67%		30 amostras do problema de DGA	Alta
Wong et al. (2001)	Pedagógica	Regras de Associação	Independente	Fuzzy	-	Erro médio quadrático – 0.25	14 regras (equação matemática) - Alta	Baixa
Wang (1997)	Decomposicional	Qualquer algoritmo de agrupamento	Depende do tamanho do domínio	Fuzzy	-	-	3 regras (Iris) Alta	Baixa

4. Metodologia Proposta para Extração de Regras Difusas a partir de Mapas Auto-Organizáveis

Este capítulo apresenta a metodologia proposta para extração de regras difusas a partir de Mapas Auto-organizáveis de Kohonen. Como discutido no Capítulo 3, atualmente são poucos os trabalhos já apresentados na literatura voltados para o problema de extração de conhecimento a partir deste tipo de rede neural artificial.

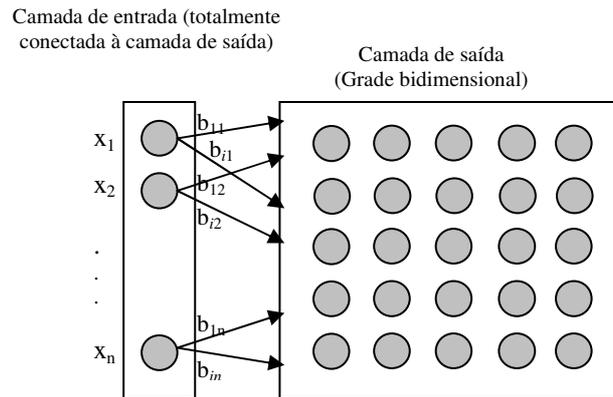
O mapa de Kohonen, assim como outras diversas topologias de RNAs, possui a desvantagem de não apresentar explanação sobre seus resultados. Os SOMs permitem apenas a visualização dos grupos (em problemas de categorização) ou classes (em problemas de classificação) criadas com o auxílio de ferramentas ou algoritmos aplicados ao resultado do treinamento, não informando com clareza o motivo de um padrão ser alocado a uma determinada classe e não, à outra.

A metodologia proposta visa suprir esta carência em relação aos Mapas Auto-Organizáveis de Kohonen, fornecendo a possibilidade de obter conhecimento por meio da extração de regras difusas do tipo Takagi-Sugeno de Ordem Zero. Para demonstração da aplicabilidade da metodologia proposta, serão apresentados três estudos de caso. No capítulo seguinte, a metodologia será aplicada ao problema de diagnóstico de faltas em transformadores de potência.

4.1 Metodologia para Extração de Regras Difusas a partir de Mapas Auto-Organizáveis de Kohonen

Para a apresentação da metodologia proposta de extração de regras difusas do tipo Takagi-Sugeno de Ordem Zero a partir de Mapas Auto-Organizáveis de Kohonen, a rede de Kohonen da Figura 4.1 deve ser considerada.

FIGURA 4.1 Mapa Auto-organizável de Kohonen



Na Figura 4.1, x_n é a n -ésima entrada para o mapa e b_{in} é o peso da sinapse que conecta a entrada n e o neurônio i da grade de saída.

Considerando o processo de treinamento da Rede de Kohonen finalizado (para um determinado padrão de dados), o mapa devidamente formado na grade de saída, o processo de extração de conhecimento por meio de regras difusas pode ser realizado. Para a realização deste processo, três procedimentos devem ser adotados:

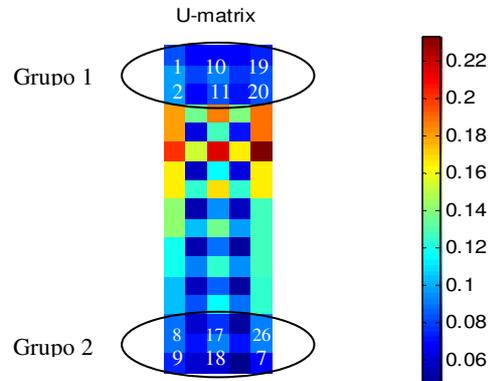
A. Interpretação do mapa resultante após treinamento para determinação das classes ou grupos formados

O primeiro procedimento para extração das regras difusas a partir do mapa obtido durante a fase de treinamento da rede de Kohonen é a determinação dos grupos formados com o auxílio da Matriz-U. Também podem ser considerados os resultados do método de agrupamento restrito à vizinhança (mencionado no Capítulo 2) e do método de rotulação do mapa.

Considerando a Matriz-U, como já exposto no Capítulo 2, este método permite a visualização dos grupos formados após o treinamento da rede de Kohonen. A distância Euclidiana entre os pesos sinápticos de neurônios adjacentes é calculada e apresentada no mapa com diferente coloração entre neurônios adjacentes. De acordo com a escala de cores, o vermelho escuro (parte superior da escala) corresponde à maior distância entre neurônios e a cor azul-escuro (parte inferior da escala) corresponde à menor distância entre neurônios. Áreas no mapa com cores, correspondendo às menores distâncias podem ser analisadas como grupos (*clusters*) e áreas no mapa, correspondendo às maiores distâncias podem ser analisadas como separadores de grupos (bordas).

Para exemplificar, considere a Figura 4.2 que apresenta a Matriz-U para um determinado treinamento, no qual se pode distinguir inicialmente a formação de dois grupos.

FIGURA 4.2 Exemplo de resultado apresentado pela Matriz-U



De acordo com a Figura 4.2, o Grupo 1 é formado pelas unidades de mapa (neurônios) 1, 2, 10, 11, 19 e 20, e o Grupo 2, pelas unidades de mapa 8, 9, 17, 18, 26 e 27. Os dois grupos estão separados por unidades do mapa com cores que tendem, segundo a escala de cores, para o vermelho, o qual representa distâncias maiores entre pesos de neurônios vizinhos, indicando pouca similaridade, correspondendo às bordas dos grupos.

Sabe-se que a Matriz-U acrescenta à grade inicial uma linha entre duas linhas existentes e uma coluna entre duas colunas existentes, por isso sua dimensão é o dobro do número de linhas menos 1 e o dobro do número de colunas menos 1.

B. Determinação do neurônio com maior potencial para representar uma classe/grupo

Com o término do treinamento e os grupos definidos, o procedimento para seleção do neurônio com maior potencial para representação de cada grupo deve ser realizado. O cálculo do potencial de um neurônio para representar uma determinada classe é dado por (CHIU, 1994):

$$P_{ji} = \sum_{l=1}^{N_i} \exp(-\|w_j - x(l)\|^2) \quad (4.1)$$

sendo P_{ji} a medida do potencial do neurônio j que pertence ao grupo i , N_i o número de padrões do vetor de entrada alocados para o grupo i , w_j o vetor de pesos do neurônio j dado

por $w_j = [b_{j1} \ b_{j2} \ \dots b_{jn}]$, com n igual ao número de entradas da rede, $x(l)$ o padrão l do vetor de entradas e $\|\cdot\|$ a distância Euclidiana.

Assim, a medida do potencial de um neurônio é uma função da distância do seu vetor de pesos em relação aos N padrões de entrada alocados ao grupo, ao qual este neurônio pertence, concluindo-se que quanto menor a distância calculada, maior será o potencial do neurônio.

Este procedimento é realizado separadamente para todos os grupos formados, a fim de definir o neurônio p representativo de cada grupo. Isto é proposto porque os vetores de peso das sinapses do SOM tornam-se generalizações da base de dados inteira, uma vez que já no término do processo de ordenação, cada neurônio tem como componentes do vetor aproximadamente o valor das médias dos conjuntos de Voronoi (amostras alocadas) dos neurônios vizinhos. Daí à importância de se evitar configurações metaestáveis para as quais a média da distorção esperada ou erro de quantização médio resulta em um mínimo local e não em um mínimo global, ou seja, quanto menor o erro de quantização, mais bem representadas estarão as amostras da base de dados pelos vetores de peso dos neurônios.

C. Extração das Regras Difusas

Após a definição do neurônio com maior potencial para representar cada grupo, o procedimento de extração do sistema difuso pode ser iniciado.

O conhecimento adquirido pelo SOM será descrito por meio de regras difusas do tipo Takagi-Sugeno de ordem zero. Para cada classe, uma regra pode ser extraída na forma:

$$R_i: \text{SE } x_1 \text{ É } C_1^i \text{ E } x_2 \text{ É } C_2^i \text{ E } \dots x_n \text{ É } C_n^i \text{ ENTÃO } y_i = i \quad (4.2)$$

sendo x_n é o n -ésimo componente do vetor de entrada do mapa de Kohonen e C_n^i a função de pertinência triangular obtida (extraída) a partir do peso sináptico b_{pn} entre a entrada x_n e o neurônio p representativo da classe e a partir dos valores mínimo (l_{\min}) e máximo (l_{\max}) da entrada x_n (considerando apenas os valores de x_n para os padrões que foram classificados na classe i).

A Figura 4.3 apresenta o gráfico da função de pertinência triangular C_n^i representada matematicamente por:

$$C_n^i(x_n) = \begin{cases} \frac{(x_n - a)}{(b_{pn} - a)}, & \text{se } a \leq x_n \leq b_{pn} \\ \frac{(c - x_n)}{(c - b_{pn})}, & \text{se } b_{pn} \leq x_n \leq c \\ 0 & \text{se } x_n > c \text{ ou } x_n < a \end{cases} \quad (4.3)$$

sendo:

b_{pn} - o peso sináptico entre o neurônio p com maior potencial para representar o grupo e a entrada x_n ;

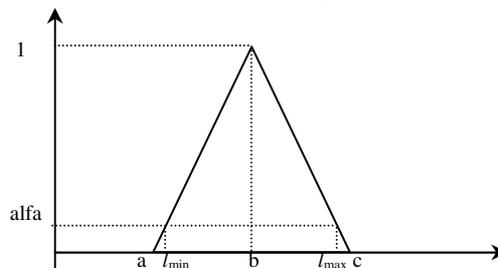
a - valor obtido através do peso b_{pn} e do valor mínimo (l_{min}) do padrão de entrada x_n para o grupo do neurônio p :

$$a = \frac{l_{min} - (alfa * b_{pn})}{1 - alfa} \quad (4.4)$$

c - valor obtido através do peso b_{pn} e do valor máximo (l_{max}) do padrão de entrada x_n para o grupo do neurônio p :

$$c = \frac{l_{max} - (alfa * b_{pn})}{1 - alfa} \quad (4.5)$$

FIGURA 4.3 Função de pertinência triangular



Nos testes realizados, o melhor valor de $alfa$ obtido foi de 0.1, uma vez que tal valor não causou um acréscimo muito significativo ao domínio das funções de pertinência triangulares, evitando uma grande alteração no suporte da função de pertinência.

Considerando a interpretação linguística da função triangular C_n^i como “*está em torno de b_{pn}* ”, embora o ponto b não seja exatamente o ponto médio dos conjuntos difusos, a regra R_i em (4.2) pode ser reescrita como:

$$R_i: SE x_1 \text{ está em torno de } b_{p1} \ E x_2 \text{ está em torno de } b_{p2} \ E \dots \quad (4.6)$$

$$x_n \text{ está em torno de } b_{pn} \text{ ENTÃO } y_i = i$$

A saída do sistema difuso é calculada através da equação 4.7:

$$y = \frac{\sum_{i=1}^m y_i v_i}{\sum_{i=1}^m v_i} \quad (4.7)$$

sendo m , o número de grupos formados correspondente ao número de regras extraídas, y_i , o consequente da regra sendo este o valor de uma constante (i), e v_i , o valor de ativação de cada regra calculado através do operador *min* :

$$v_i = \min(C_1^i(x_1), C_2^i(x_2), \dots, C_n^i(x_n)) \quad (4.8)$$

O sistema difuso obtido a partir deste procedimento deve obedecer à restrição da partição. Assim, a codificação de um conjunto difuso deve obedecer:

$$C_n^i(x_n) = \max \left(0, \min \left(\frac{(x_n - a)}{(b_{pn} - a)}, \frac{(c - x_n)}{(c - b_{pn})} \right) \right) \quad (4.9)$$

Ou seja, tem-se que $a \leq b \leq c$. Para evitar o problema de *gaps*, espaços vazios (*gaps*) nas partições das variáveis de entrada, conjuntos difusos vizinhos devem obedecer à regra a (conjunto difuso da direita) $\leq c$ (conjunto difuso da esquerda). Caso esta condição não seja satisfeita, um ajuste pode ser realizado através de um procedimento denominado *Reparação (Repairing)* (ou por algoritmo genético para delimitar de forma otimizada as partições dos conjuntos difusos).

4.2 Estudos de Caso da Metodologia Proposta

Nesta seção, são apresentados três estudos de caso para utilização da metodologia proposta².

4.2.1 Estudo de Caso 1: Problema da Classificação de Flores Íris

A base de dados de flores Iris, introduzida por Sir Ronald Aylmer Fisher em 1936, traz um problema de classificação bem conhecido, que tem como objetivo reconhecer a espécie de uma flor por meio de 4 atributos (em cm):

Comprimento da pétala do cálice (x_1) \in [4,3 7,9]

Largura da pétala do cálice (x_2) \in [2 4,4]

Comprimento da pétala (x_3) \in [1 6,9]

Largura da pétala (x_4) \in [0,1 2,5]

Esta base de dados é composta por 150 amostras aleatórias de flores divididas em 3 classes de 50 amostras de cada espécie, conforme a Tabela 4.1.

TABELA 4.1 Composição da base de dados Iris

Tipos de flores	Número de Amostras
Iris setosa (SE)	50
Iris versicolor (VE)	50
Iris virginica (VI)	50

A Tabela 4.2 apresenta a divisão da base de dados para treinamento e para teste da Rede de Kohonen, que foi treinada para classificação das flores.

TABELA 4.2 Divisão da base de dados Iris para treinamento e teste

Tipos de flores	Base de Treinamento	Base de Teste
Iris Setosa	33	17
Iris Versicolor	34	16
Iris Virginica	33	17
Total	100	50

² Os bancos de dados dos casos 1 e 2 foram exportados de UC Irvine Machine Learning Repository. Disponível em: <<http://archive.ics.uci.edu/ml/>>, já do caso 3, em: <<http://orion.math.iastate.edu/burkardt/data/martinez/martinez.html>>

Os resultados da melhor topologia observada são apresentados na Tabela 4.3 após o teste de várias topologias. A tabela apresenta alguns dos parâmetros de treinamento da rede, assim como os valores do erro de quantização média e erro topográfico para a topologia de rede escolhida. Para a normalização da entrada, foi utilizado o método do histograma discreto³. O erro de quantização média (resolução) representa a média das distâncias entre cada vetor de dados e o correspondente vetor de pesos do neurônio vencedor (BMU). Assume-se que quanto menor o erro de quantização, mais bem ajustado o neurônio vencedor (BMU) estará aos vetores de entrada. Já o erro topográfico quantifica a capacidade do mapa em representar a topologia dos dados de entrada. Ele é calculado verificando-se a percentagem de vetores de dados para os quais o BMU e um segundo BMU não são unidades vizinhas no mapa.

TABELA 4.3 Treinamento com o SOM para a base de dados Iris

Topologia do mapa	Forma	Função da taxa de aprendizagem	Erro de quantização média	Erro topográfico
[9 x 3]	Retangular	Inv (0.5) ⁴	0.017054	0.03

A arquitetura do SOM possui 4 nós na camada de entrada (para os 4 atributos da flor) e 27 neurônios na camada de saída, organizados de forma bidimensional [9 x 3]. Foi utilizada a função de vizinhança Gaussiana. A topologia escolhida apresentou menor erro de quantização média e apenas 4 erros de classificação.

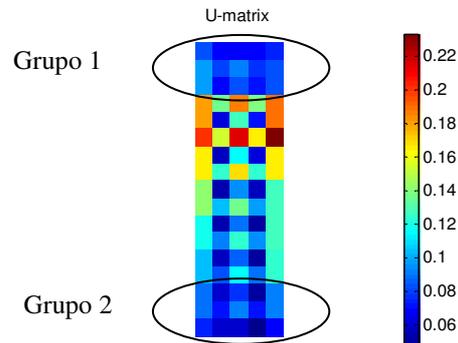
A. Definição dos grupos formados

Após o treino da rede Kohonen com auxílio da ferramenta SOM Toolbox, pode-se analisar os resultados obtidos. A Figura 4.4 apresenta a U-Matriz com os grupos identificados após o treinamento do SOM com a base de dados Iris.

³ É o processo de normalização que escalona os valores linearmente de forma que fiquem entre 0 e 1.

⁴ A Função Reciprocamente Decrescente provoca um decaimento da taxa de aprendizagem acelerado nas primeiras épocas.

FIGURA 4.4 Dois grupos são formados após o treinamento



A U-Matriz permite distinguir a formação de 2 grupos: Grupo 1 (flores do tipo Iris Setosa) e Grupo 2 (flores do tipo Iris Virginica). Neste resultado, deve-se observar que as classes Iris Versicolor e Iris Virginica não são linearmente separáveis da Iris Setosa. Para melhor verificação dos grupos formados, a Figura 4.5 apresenta o mapa rotulado resultante. O rótulo da classe com o maior número de amostras alocadas ao neurônio é atribuído à unidade de mapa correspondente, lembrando que se está trabalhando com um problema de classificação, em que as classes das amostras são conhecidas, o que permite esta rotulação.

FIGURA 4.5 Rótulos das unidades de mapa com os 3 grupos formados para cada espécie de flores

SE(12)	SE(6)	SE(12)
SE(1)		SE(2)
VE(3)		
VE(8)	VE(2)	VE(2)
VE(1) VI(1)	VE(1)	VE(6)
VE(2) VI(1)		VE(5)
VI(7) VE(1)	VI(2) VE(1)	VE(2)
VI(1)		VI(4)
VI(3)	VI(1)	VI(13)

B. Extração de regras do SOM

Após a identificação dos neurônios pertencentes a cada grupo, o processo de extração de regras do mapa de Kohonen pode ser iniciado. A equação (4.1) é aplicada a todos os grupos para selecionar o neurônio de maior potencial em cada grupo. A Tabela 4.4 apresenta os pesos

do neurônio representativo de cada grupo obtido e os valores mínimo e máximo dos padrões de entrada alocados para estas classes.

TABELA 4.4 Resultados para extração de regras para a base de dados Iris

Classes	Neurônios com maior potencial	Pesos dos neurônios com maior potencial	Valores min e max dos padrões de entrada			
			X ₁	X ₂	X ₃	X ₄
SE	10	5.0 , 3.3, 1.5, 0.3	[4.222 5.889]	[2.189 4.522]	[0.9444 1.944]	[0.07778 0.6333]
VE	14	5.9, 2.8, 4.4, 1.4	[4.789 7.122]	[1.911 3.467]	[2.844 5.067]	[0.9556 1.733]
VI	9	6.4, 3.0, 5.4, 2.0	[5.511 8.067]	[2.444 3.889]	[4.733 7.067]	[1.333 2.556]

x₁ = comprimento_da_pétala_do_cálice,
 x₂ = largura_da_pétala_do_cálice,
 x₃ = comprimento_da_pétala e
 x₄ = largura_da_pétala

A partir destas informações, as seguintes regras foram extraídas de acordo com o formato da regra apresentado em (4.6):

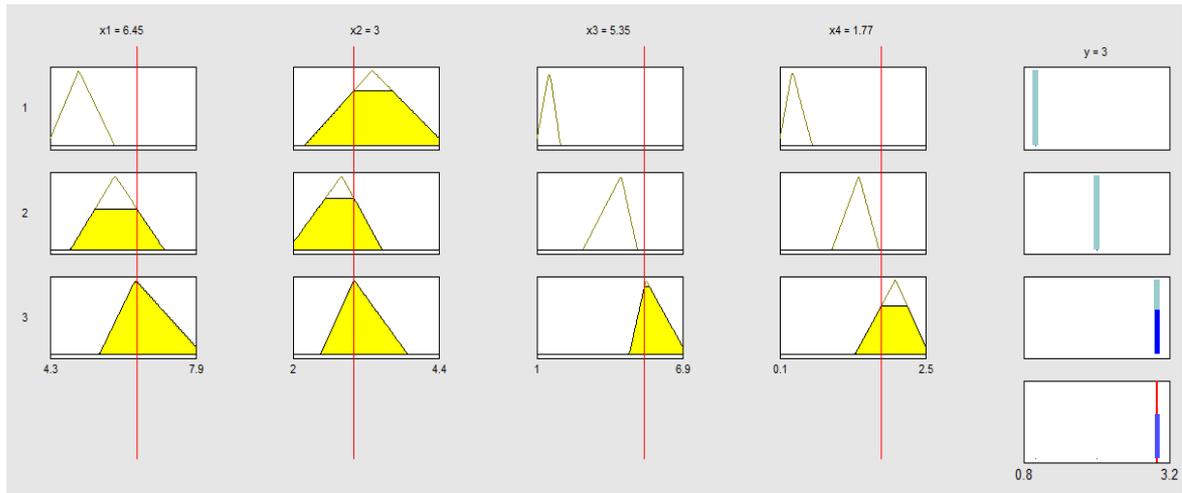
R₁: Se *comprimento_da_pétala_do_cálice* está em torno de 5,0 (GRANDE) e *largura_da_pétala_do_cálice* está em torno de 3,3 (MÉDIO) e *comprimento_da_pétala* está em torno de 1,5 (PEQUENO) e *largura_da_pétala* está em torno de 0,3 (PEQUENO) então classe = 1 (SE)

R₂: Se *comprimento_da_pétala_do_cálice* está em torno de 5,9 (GRANDE) e *largura_da_pétala_do_cálice* está em torno de 2,8 (MÉDIO) e *comprimento_da_pétala* está em torno de 4,4 (MÉDIO) e *largura_da_pétala* está em torno de 1,4 (PEQUENO) então classe = 2 (VE)

R₃: Se *comprimento_da_pétala_do_cálice* está em torno de 6,4 (GRANDE) e *largura_da_pétala_do_cálice* está em torno de 3,0 (MÉDIO) e *comprimento_da_pétala* está em torno de 5,4 (GRANDE) e *largura_da_pétala* está em torno de 2,0 (MÉDIO) então classe = 3 (VI)

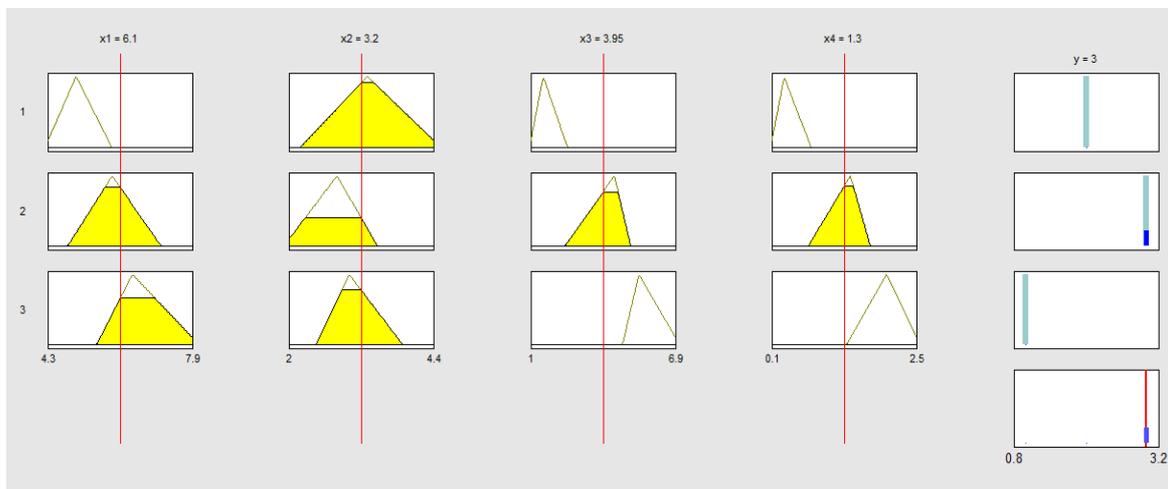
A Figura 4.6, obtida através da Fuzzy Logic Toolbox do Matlab, apresenta as funções de pertinência e as regras extraídas.

FIGURA 4.6 Sistema difuso extraído para a base dados Iris



Como pode ser observado na Figura 4.6, as regras difusas extraídas do mapa de Kohonen resultaram em uma base de regras incompleta (problema de espaços vazios – *gaps*), uma vez que as variáveis x_3 e x_4 apresentam intervalos descobertos, criando não apenas regiões sem sobreposição, mas regiões do domínio em que a função é indefinida. Como mencionado no Capítulo 2, o método mais simples para corrigir uma base de regras difusas incompleta é modificar os parâmetros c_1 e a_2 , fazendo $a_2 < c_1$ nas partições adjacentes para forçar a sobreposição. Outra possibilidade é a interpolação de regras difusas, capaz de criar um novo conjunto difuso no intervalo descoberto.

A Figura 4.7 apresenta as funções de pertinência e as regras extraídas, após alteração dos limites dos conjuntos difusos (*Repairing*) de X_3 de 1.944 para 2.5 e 2.844 para 2.4 e de X_4 , de 0.6333 para 0.75 e 0,9556 para 0.7.

FIGURA 4.7 Sistema difuso extraído para a base dados Iris, após correção (*Repairing*)

As regras difusas extraídas do mapa de Kohonen informam como o sistema chega a uma determinada classificação pela simples observação dos limites dos padrões de entrada em cada função de pertinência, indicando ao usuário as regras que poderão ser disparadas para o padrão apresentado e pelos respectivos valores de ativação destas regras.

Assumindo que o erro de equivalência é o número de saídas do FIS, cujos valores não correspondem às saídas do SOM, a Tabela 4.5 mostra o número de erros de classificação do SOM e do FIS para a base de treinamento juntamente com o número de erros de equivalência do FIS projetado.

TABELA 4.5 Erros de classificação e de equivalência para a base de dados Iris no treinamento

Método	Erros de Classificação	Erros de Equivalência
Kohonen	4	-
FIS	5	1

De acordo com a Tabela 4.5, o FIS apresenta uma taxa de fidelidade de 99%, uma vez que apresentou apenas 1 padrão com classificação diferente da classificação do mapa treinado.

A Tabela 4.6 resume as percentagens de acertos de classificação do SOM e do FIS para o problema da base de dados Iris, tanto para treinamento quanto para validação.

TABELA 4.6 Resultados da classificação para a base de dados Iris

Método	% acerto - Base de Treinamento	% acerto - Base de Validação
Kohonen	96	100
FIS	95	98

Dos resultados apresentados pode-se verificar que o FIS extraído do SOM apresenta uma taxa de sucesso total de 96%, desde que o mesmo, considerando tanto as amostras usadas para treinamento quanto as amostras usadas para validação, conseguiu classificar corretamente 144 padrões dos 150 da base de dados.

4.2.2 Estudo de caso 2: Problema da Classificação de Papel-Moeda

A base de dados tem por objetivo classificar notas monetárias como verdadeiras ou falsificadas por meio de 6 características (atributos):

Comprimento da nota (x_1) \in [213,8, 217,8]
 Largura da nota, medida do lado esquerdo (x_2) \in [129 131]
 Largura da nota, medida do lado direito (x_3) \in [129 131,1]
 Largura da margem menor (x_4) \in [7,2 12,7]
 Largura da margem maior (x_5) \in [7,7 12,3]
 Comprimento da diagonal da direita inferior à direita superior (x_6) \in [137,8 142,4]

Esta base de dados é composta por 200 amostras aleatórias de notas divididas em 2 classes de 100 amostras de cada tipo, conforme a Tabela 4.7.

TABELA 4.7 Estrutura da base de dados do banco suíço

Notas	Número de Amostras
Notas Falsificadas	100
Notas Verdadeiras	100

A Tabela 4.8 apresenta a divisão da base de dados para treinamento e para teste do mapa de Kohonen, que foi treinado para classificação das notas.

TABELA 4.8 Divisão da base de dados do banco suíço para treinamento e teste

Tipos de Notas	Base de Treinamento	Base de Teste
Notas Falsas (BFO)	68	32
Notas Verdadeiras (BGE)	68	32
Total	136	64

Diversas topologias do mapa foram testadas, sendo os resultados da melhor topologia apresentados na Tabela 4.9.

TABELA 4.9 Treinamento com o SOM para a base de dados do banco suíço

Topologia do mapa	Forma	Função da taxa de aprendizagem	Erro de quantização média	Erro topográfico
[3 x 9]	Retangular	Haykin (0.1) ⁵	-	-

A arquitetura do SOM possui 6 nós na camada de entrada (para os 6 atributos das notas) e 27 neurônios na camada de saída, organizados de forma bidimensional [3 x 9]. Foi utilizada a função de vizinhança Gaussiana. A topologia acima não apresentou erro de classificação no treinamento, e um erro no teste.

⁵Função aprendizagem disponível em Haykin (2007)

A. Identificação dos grupos formados

A Tabela 4.10 apresenta os resultados do treinamento, sendo que os grupos foram identificados, utilizando o método do agrupamento por rotulação, já que o agrupamento restrito à vizinhança identificou apenas 1 grupo.

TABELA 4.10 Treinamento com o SOM para a base de dados do banco suíço

11-BFO	15-BFO	7-BFO	5-BFO	5-BGE	7-BGE	6-BGE	6-BGE	9-BGE
2-BFO	2-BFO	10-BFO	6-BFO	2-BGE	6-BGE	4-BGE	3-BGE	
7-BFO		3-BFO	1-BGE	3-BGE	3-BGE	7-BGE	3-BGE	3-BGE

B. Extração de regras do SOM

Após a identificação dos neurônios pertencentes a cada grupo, o processo de extração de regras do mapa de Kohonen pode ser iniciado. A equação (4.1) é aplicada a todos os neurônios para selecionar o neurônio de maior potencial em cada grupo. A Tabela 4.11 apresenta os pesos do neurônio representativo de cada grupo obtido e os valores mínimo e máximo dos padrões de entrada alocados para estes grupos.

TABELA 4.11 Resultados para extração de regras para o problema de notas falsas

Classes	Neurônios com maior potencial	Pesos dos neurônios com maior potencial	Valores min e max dos padrões de entrada					
			X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
BFO	2	214.84, 130.26, 130.13, 11.38, 10.61, 139.5	[213.87 217.85]	[129.57 130.82]	[129.28 131.15]	[7.38 12.75]	[9.05 12.38]	[137.78 140.69]
BGE	17	214.75, 129.68, 129.55, 8.07, 10.15, 141.61	[213.79 216]	[128.96 131.1459]	[128.91 131.15]	[7.19 10.5]	[7.69 11.73]	[139.57 142.5]

x₁. = comprimento_da_nota,
 x₂ = largura_esquerda,
 x₃. = largura_direita,
 x₄. = largura_margem_maior,
 x₅. = largura_margem_menor e
 x₆. = largura_da_diagonal

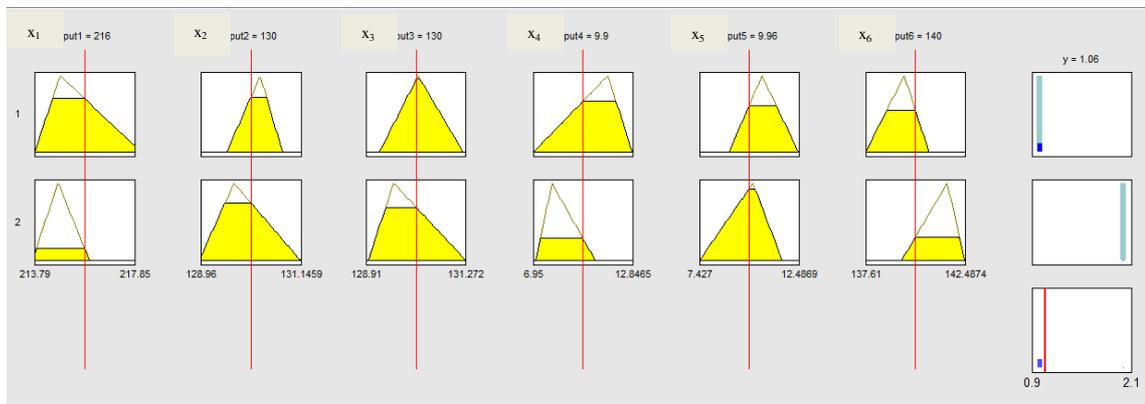
A partir destas informações, as seguintes regras foram extraídas de acordo com o formato da regra apresentado em (4.6):

R₁: Se *comprimento_da_nota* está em torno de 214,84 (MÉDIA) e *largura_esquerda* está em torno de 130,26 (MÉDIA) e *largura_direita* está em torno de 130,13 (MÉDIA) e *largura_margem_maior* está em torno de 11,38 (GRANDE) e *largura_margem_menor* está em torno de 10,61 (GRANDE) e *largura_da_diagonal* está em torno de 139,6 (MÉDIA) então classe = 1 (Nota Falsificada)

R₂: Se *comprimento_da_nota* está em torno de 214,75 (PEQUENA) e *largura_esquerda* está em torno de 129,68 (MÉDIO) e *largura_direita* está em torno de 129,55 (MÉDIA) e *largura_margem_maior* está em torno de 8,07 (PEQUENA) e *largura_margem_menor* está em torno de 10,15 (MÉDIA) e *largura_da_diagonal* está em torno de 141,61 (GRANDE) então classe = 2 (Nota verdadeira)

A Figura 4.8 apresenta as funções de pertinência e as regras extraídas do mapa de Kohonen.

FIGURA 4.8 Sistema difuso extraído para a base dados do banco suíço



A Tabela 4.12 mostra o número de erros de classificação do SOM e do FIS para a base de treinamento juntamente com o número de erros de equivalência do FIS projetado.

TABELA 4.12 Erros de classificação e de equivalência para a base de dados suíço no treinamento

Método	Erros de Classificação	Erros de Equivalência
Kohonen	0	-
FIS	1	1

De acordo com a Tabela 4.12, o FIS apresenta uma taxa de fidelidade de 99,26%, uma vez que apresentou apenas 1 padrão com classificação diferente da classificação do mapa treinado. Este padrão é a amostra 68 que foi alocada ao neurônio 22 (o terceiro neurônio da

quarta coluna da Tabela 4.10, que mostra o resultado do treinamento).

A Tabela 4.13 resume as percentagens de acertos de classificação do SOM e do FIS para o problema da base de dados do banco suíço, tanto para treinamento quanto para validação. Enquanto o mapa de Kohonen classificou 1 padrão de teste incorretamente, o sistema difuso gerado classificou corretamente todos padrões.

TABELA 4.13 Resultados da classificação para a base de dados do banco suíço

Método	% acerto - Base de Treinamento	% acerto - Base de Validação
Kohonen	100	98,4375
FIS	99,26	100

Dos resultados apresentados pode-se verificar que o FIS extraído do SOM apresenta uma taxa de sucesso total de 99,5%, desde que o mesmo, considerando tanto as amostras usadas para treinamento quanto as amostras usadas para validação, conseguiu classificar corretamente 199 padrões dos 200 da base de dados.

4.2.3 Estudo de caso 3: Problema da Classificação de Vinhos

A base de dados tem por objetivo classificar vinhos em 3 classes por meio de 13 características (atributos), com o intuito de determinar a origem dos vinhos:

Álcool (x_1) \in [12,93 14,83]

Ácido Málico (x_2) \in [1,35 4,04]

Cinza (x_3) \in [2,04 2,87]

Alcalinidade das cinzas (x_4) \in [11,20 21,0]

Magnésio (x_5) \in [89,0 128,0]

Fenóis Totais (x_6) \in [2,35 3,88]

Flavonóides (x_7) \in [2,33 3,93]

Fenóis não-flavonóides (x_8) \in [0,17 0,43]

Proantocianidinas (tipo de flavonóide) (x_9) \in [1,25 2,91]

Intensidade da cor (x_{10}) \in [3,52 8,90]

Matiz (x_{11}) \in [0,82 1,28]

Índice de vinhos diluídos (OD280/OD315) (x_{12}) \in [2,57 3,92]

Prolina (x_{13}) \in [680,0 1680,0]

Estes dados são o resultado de uma análise química de vinhos cultivados na mesma região, na Itália, mas de 3 cultivares diferentes. A análise determina as quantidades de 13 constituintes encontrados em cada um dos três tipos de vinhos. Dispôs-se de 178 amostras

aleatórias de vinhos divididas em 3 classes, conforme a Tabela 4.14.

TABELA 4.14 Estrutura da base de dados de vinhos

Tipos de Vinho	Número de Amostras
A	59
B	71
C	48

A Tabela 4.15 apresenta a divisão da base de dados para treinamento e para teste do mapa de Kohonen para classificação de vinhos.

TABELA 4.15 Divisão da base de dados de vinhos para treinamento e teste

Vinhos	Base de Treinamento	Base de Teste
A	40	19
B	50	21
C	32	16
Total	122	56

Diversas topologias do mapa foram testadas, sendo que os resultados da melhor topologia observada são apresentados na Tabela 4.16.

TABELA 4.16 Treinamento com o SOM para a base de dados de vinhos

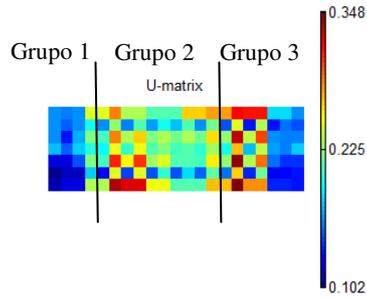
Topologia do mapa	Forma	Função da taxa de aprendizagem	Erro de quantização média	Erro topográfico
[4 x 11]	Retangular	inv (0,5)	0,57153	0,016393

A arquitetura do SOM possui 13 nós na camada de entrada e 44 neurônios na camada de saída, organizados de forma bidimensional [4 x 11]. Foi utilizada a função de vizinhança Gaussiana. Para facilitar a formação do SOM, os dados foram normalizados com o método histograma discreto, devido as grandes variações entre os dados dos atributos.

A. Definição dos grupos formados – SOM-Toolbox

A Figura 4.9 apresenta a Matriz-U com os grupos identificados após o treinamento do SOM com a base de dados de vinhos.

FIGURA 4.9 Três grupos foram formados após o treinamento



A Matriz-U permite distinguir a formação de 3 grupos: Grupo 1 (tipo C), Grupo 2 (tipo B) e Grupo 3 (tipo A). Para melhor verificação dos grupos formados, a Figura 4.10 apresenta o mapa rotulado resultante.

FIGURA 4.10 Rótulos das unidades de mapa com os 3 grupos formados para cada classe de vinho

A(8)	A(2) B(1)	A(2) B(2)	A(1) B(1)	B(7)	B(4)	B(1)	B(4)	C(3)	C(1)	C(8)
A(1)	A(1)	A(1)	A(1)	B(1)	B(1)	B(2)		B(1) C(1)		C(5)
A(1)	A(2)	A(1)	B(2)		B(2)	B(3)	B(3)			C(3)
A(12)	A(3)	A(3)	B(3) A(1)	B(3)	B(3)	B(3)	B(2)	B(1) C(1)	C(2)	C(8)

B. Extração de regras do SOM

Após a identificação dos neurônios pertencentes a cada grupo, o processo de extração de regras do mapa de Kohonen pode ser iniciado. A equação (4.1) é aplicada a todos os neurônios para selecionar o neurônio de maior potencial em cada grupo. A Tabela 4.17 apresenta os pesos do neurônio representativo de cada grupo obtido e os valores mínimo e máximo dos padrões de entrada alocados para estes grupos.

TABELA 4.17 Resultados para extração de regras da base de dados sobre vinhos

	Tipo A	Tipo B	Tipo C
Neurônios com maiores potenciais	12	27	37
Pesos dos neurônios com maiores potenciais	13.52, 1.73, 2.28, 16.5, 101, 2.7, 2.92, 0.28, 1.95, 5.1, 1.07, 3.13, 985	12.36, 1.71, 2.27, 19.1, 91, 2.05, 1.79, 0.4, 1.36, 3.3, 1.07, 2.69, 495	13.24, 3.12, 2.54, 22, 98, 1.83, 0.84, 0.45, 1.35, 6, 0.74, 1.8, 600
X ₁	[11.3422 14.9756]	[10.8822 13.8156]	[12.0844 14.4622]
X ₂	[1.3078 4.2967]	[0.63222 6.2544]	[1.0311 5.9311]
X ₃	[2.0133 3.3356]	[1.2589 2.9922]	[1.9178 2.8956]
X ₄	[10.6111 31.5]	[9.6556 29.5444]	[15.3333 27.1]
X ₅	[84.3333 143.2222]	[69.99 169.8889]	[78 125.7778]
X ₆	[2.19 4.0111]	[1.09 3.6833]	[0.979 2.9078]
X ₇	[2.186 5.32]	[0.43444 3.9678]	[0.33 1.6733]
X ₈	[0.15778 0.51]	[0.11111 0.661]	[0.1699 0.65]
X ₉	[1.1722 3.0167]	[0.30444 3.8267]	[0.46111 2.85]
X ₁₀	[3.1556 9.3222]	[1.0556 6.0222]	[3.1111 13.7778]
X ₁₁	[0.83667 1.3922]	[0.64778 1.7811]	[0.45111 0.98444]
X ₁₂	[2.5078 4.0078]	[1.4678 3.7456]	[1.2111 2.5444]
X ₁₃	[346.1111 1757.2222]	[253.8889 1095]	[346.6667 880.6]

X₁ = álcool,

X₂ = ácido_málico,

X₃ = cinza

X₄ = alcalinidade_das_cinzas;

X₅ = magnésio,

X₆ = fenol_total,

X₇ = flavonóide,

X₈ = fenol_não_flavonóide,

X₉ = proantocianidinas,

X₁₀ = intensidade_da_cor,

X₁₁ = matiz,

X₁₂ = índice_de_vinhos_diluídos e

X₁₃ = prolina;

A partir destas informações, as seguintes regras foram extraídas de acordo com o formato da regra apresentado em (4.6):

R₁: Se *álcool* está em torno de 13,52 (MÉDIO) e *ácido_málico* está em torno de 1,73 (BAIXO) e *cinza* está em torno de 2,28 (MÉDIO) e *alcalinidade_das_cinzas* está em torno de 16,5 (MÉDIO) e *magnésio* está em torno de 101 (MÉDIO) e *fenol_total* está em torno de 2,7 (MÉDIO) e *flavonóide* está em torno de 2,92 (MÉDIO) e *fenol_não_flavonóide* está em torno de 0,28 (MÉDIO) e *proantocianidinas* está em torno de 1,95 (MÉDIO) e *intensidade_da_cor* está em torno de 5,1 (MÉDIO) e *matiz* está em torno de 1,07 (MÉDIO) *índice_de_vinhos_diluídos* está em torno de 3,13 (MÉDIO) e *prolina* está em torno de 985 (MÉDIO) então classe = 1 (A)

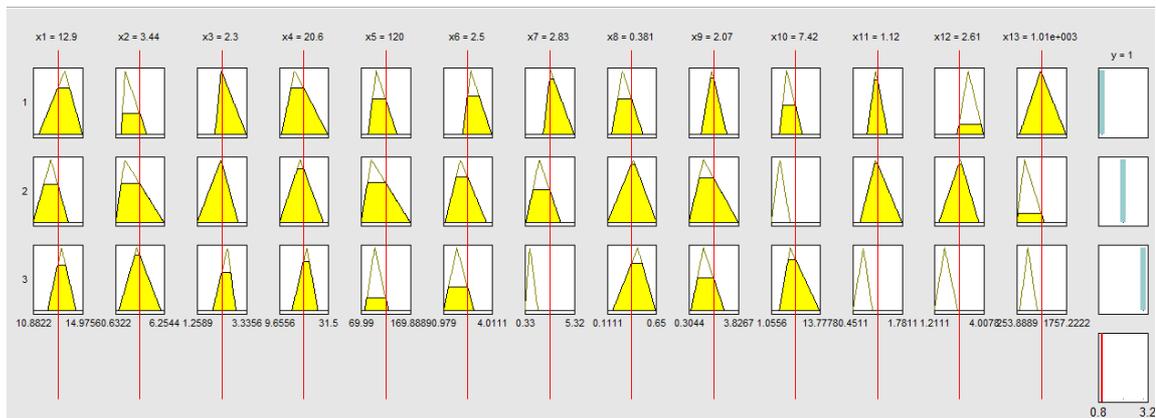
R₂: Se *álcool* está em torno de 12,36 (MÉDIO) e *ácido_málico* está em torno de 1,71 (BAIXO) e *cinza* está em torno de 2,27 (MÉDIO) e *alcalinidade_das_cinzas* está em torno

de 19,1 (MÉDIO) e *magnésio* está em torno de 91 (BAIXO) e *fenol_total* está em torno de 2,05 (MÉDIO) e flavonóide está em torno de 1,79 (MÉDIO) e *fenol_não_flavonóide* está em torno de 0,4 (MÉDIO) e *proantocianidinas* está em torno de 1,36 (M) e *intensidade_da_cor* está em torno de 3,3 (MÉDIO) e *matiz* está em torno de 1,07 (BAIXO) *índice_de_vinhos_diluídos* está em torno de 2,69 (MÉDIO) e *prolina* está em torno de 495 (MÉDIO) então classe = 2 (B)

R₃: Se *álcool* está em torno de 13,24 (MÉDIO) e *ácido_málico* está em torno de 3,12 (BAIXO) e *ash* está em torno de 2,54 (BAIXO) e *alcalinidade_das_cinzas* está em torno de 22 (BAIXO) e *magnésio* está em torno de 98 (BAIXO) e *fenol_total* está em torno de 1,83 (MÉDIO) e flavonóide está em torno de 0,84 (MÉDIO) e *fenol_não_flavonóide* está em torno de 0,45 (MÉDIO) e *proantocianidinas* está em torno de 1,35 (MÉDIO) e *intensidade_da_cor* está em torno de 6 (BAIXO) e *matiz* está em torno de 0,74 (MÉDIO) *índice_de_vinhos_diluídos* está em torno de 1,8 (MÉDIO) e *prolina* está em torno de 600 (BAIXO) então classe = 3 (C)

A Figura 4.11 apresenta as regras extraídas utilizando Fuzzy Logic Toolbox do Matlab.

FIGURA 4.11 Sistema difuso extraído para a base dados sobre vinhos



Assumindo que o erro de equivalência é o número de saídas do FIS, cujos valores não correspondem às saídas do SOM, a Tabela 4.18 mostra o número de erros de classificação do SOM e do FIS para a base de treinamento juntamente com o número de erros de equivalência do FIS projetado.

TABELA 4.18 Erros de classificação e de equivalência para a base de dados de vinhos no treinamento

Método	Erros de Classificação	Erros de Equivalência
Kohonen	8	0
FIS	7	3

De acordo com a Tabela 4.18, o FIS apresenta uma taxa de fidelidade de 97,54%, uma vez que apresentou apenas 3 padrões com classificação diferente da classificação do mapa treinado. Já a Tabela 4.19 resume as percentagens de acertos de classificação do SOM e do FIS para o problema da base de dados de vinhos, tanto para treinamento quanto para validação.

TABELA 4.19 Resultados da classificação para a base de dados de vinhos

Método	% acerto - Base de Treinamento	% acerto - Base de Validação
Kohonen	93,44%	100%
FIS	94,26%	100%

Dos resultados apresentados pode-se verificar que o FIS extraído do SOM apresenta uma taxa de sucesso total de 96,06%, desde que o mesmo, considerando tanto as amostras usadas para treinamento quanto as amostras usadas para validação, conseguiu classificar corretamente 171 padrões dos 178 da base de dados.

4.4 Conclusão do Capítulo

Este capítulo forneceu uma descrição detalhada da metodologia proposta para extração de regras difusas a partir de mapas auto-organizáveis de Kohonen. A metodologia permite que um sistema do tipo Takagi-Sugeno de ordem zero seja extraído de um mapa auto-organizável de Kohonen. O processo de extração se baseia no fato de que após o treinamento da rede, neurônios vizinhos e pertencentes a um mesmo grupo possuem pesos sinápticos similares. Tal metodologia diferente das já apresentadas na literatura é bastante simples, sendo utilizada para projetar sistemas difusos com um número de regras igual ao número de classes, dependendo do agrupamento realizado pelo SOM, não requerendo a simplificação das regras obtidas.

Levando em consideração os critérios de classificação dos métodos de extração de conhecimento de redes neurais apresentados no Capítulo 3, quanto ao *poder de expressão das regras extraídas*, a metodologia fornece aos usuários o conhecimento capturado pelo mapa auto-organizável por meio de regras difusas. Como já mencionado no Capítulo 2, um sistema difuso é uma poderosa ferramenta para representar e inferir conhecimento impreciso ou não

confiável. Diferente das árvores de decisão ou representações simbólicas convencionais, as regras difusas podem tratar a imprecisão das variáveis de entrada e de saída do sistema ao definirem números ou conjuntos difusos expressos em forma de variáveis linguísticas, cujos valores são palavras em linguagem natural, as quais são mais compreensíveis para o ser humano do que uma representação simbólica.

O critério de *qualidade da regra extraída* é um dos mais importantes critérios de avaliação para algoritmos de extração de regras. Possui quatro medidas: fidelidade, taxa de sucesso (*accuracy*), consistência e facilidade de compreensão. A *fidelidade* determina se as regras representam o comportamento da RNA de forma adequada, quando aplicadas às amostras de treinamento e teste. Verifica-se que os erros de equivalência foram baixos entre 1 e 3 amostras classificadas de forma diferente do mapa. Já a *taxa de sucesso*, que é um indicativo da capacidade de generalização das regras extraídas, variou de 98 a 100% para os testes realizados. Os sistemas difusos gerados a partir da metodologia podem ser considerados de *fácil compreensão*, uma vez que possuem um pequeno conjunto de regras e antecedentes, sendo assim bastante compreensível para seres humanos.

De acordo com o critério da transparência, a extração de regras a partir da RNA pode ser classificada como decomposicional, pedagógica e eclética. Conforme a metodologia descrita, pode-se considerar a mesma como uma abordagem *eclética*, uma vez que os valores de pesos das sinapses do mapa treinado são utilizados para compor as regras difusas, assim como os valores limites das amostras de cada grupo. Já a complexidade do algoritmo é baixa, o único cálculo realizado diz respeito ao cálculo do potencial dos neurônios. Quanto à portabilidade ou generalidade pode-se afirmar que esta metodologia só se aplica a mapas auto-organizáveis com duas camadas (entrada e saída), porém a qualquer tipo de domínio.

5. Aplicação: Diagnóstico de Falhas Incipientes em Transformadores de Potência

O transformador é um dos equipamentos elétricos mais caros e importantes de um sistema de energia, logo o seu adequado funcionamento é decisivo para a segurança do sistema. Um transformador em operação está sujeito a uma variedade de falhas elétricas e térmicas que podem provocar defeito no equipamento. Detectar e eliminar as falhas em condição incipiente, ou seja, antes que elas se transformem em falhas graves, é primordial para assegurar o bom funcionamento do sistema de geração de energia.

Quando o transformador está em condição de falha incipiente, alguns gases são desenvolvidos e retidos no óleo isolante (gases dissolvidos no óleo). O valor de concentração destes gases e a relação entre estas concentrações permitem a determinação do tipo e gravidade da falha a qual o transformador está submetido.

Alguns métodos convencionais baseados na análise dos gases dissolvidos em óleo vêm sendo utilizados pelas concessionárias de energia para auxiliar na detecção de falhas incipientes. Entretanto, a análise e interpretação da quantidade destes gases podem ser consideradas não como uma ciência, mas uma arte sujeita à variabilidade (IEEE Std C57.104.1991, 1992), logo a procura por métodos mais eficientes ainda é um tópico de grande interesse por parte das concessionárias de energia.

Alguns trabalhos vêm sendo apresentados na literatura utilizando redes neurais artificiais para o problema de diagnóstico de transformadores (SALAMI e PAHLEVANI, 2008; YANG-JING et al., 2007; THANG e AGGARWAL, 2003). As vantagens oferecidas pelo uso das redes neurais, principalmente no que diz respeito a sua capacidade de adquirir conhecimento, têm levado ao desenvolvimento de sistemas de diagnóstico com resultados promissores. As redes neurais podem aprender o conhecimento dos especialistas que trabalham na área de diagnóstico de transformadores, assim como um conhecimento ainda não adquirido pelos mesmos (mas que pode ser obtido através de banco de dados de transformadores faltosos).

Uma grande desvantagem de alguns dos sistemas inteligentes de diagnóstico baseados em redes neurais desenvolvidos até agora diz respeito à falta de capacidade das redes em explicar como elas chegam a determinado resultado (GOLKHAH et al., 2011; HU et al., 2008). Esta explicação seria importante e necessária para que o especialista da área pudesse compreender os resultados da rede e desta forma pudesse ter confiança no diagnóstico produzido.

Considerando a importância do desenvolvimento de sistemas inteligentes de diagnóstico de faltas em transformadores que apresentem a capacidade de explicação de seus resultados, este capítulo apresenta um sistema de diagnóstico de transformadores baseado em um mapa auto-organizável de Kohonen. Para cobrir a deficiência no que diz respeito à ausência de explanação em relação ao diagnóstico produzido pelas redes treinadas, a metodologia para extração de regras difusas apresentada no Capítulo 4 será utilizada para extração do conhecimento apreendido pelo mapa. Através das regras extraídas, o especialista da área terá a possibilidade de analisar como o sistema inteligente chegou ao diagnóstico de faltas e desta forma ter mais confiança no resultado apresentado.

Antes da apresentação dos sistemas inteligentes propostos será apresentada uma visão geral sobre diagnóstico de faltas em transformadores baseado na análise dos gases dissolvidos em óleo.

5.1 Diagnóstico de Faltas baseado na Análise dos Gases Dissolvidos em Óleo (DGA)

A formação de certos gases no óleo, gerados dentro de uma taxa específica ao longo do tempo, é frequentemente a primeira indicação de mau funcionamento de um transformador em operação.

A cromatografia gasosa é uma técnica utilizada pelas concessionárias para medir a concentração dos gases no óleo. As medidas podem ser obtidas com o transformador em funcionamento, seja por amostragem periódica do óleo ou por monitoramento contínuo. A análise destas medidas permite avaliar a condição de operação do isolamento, identificando eventuais processos de falha que estejam ocorrendo.

Os gases comumente encontrados no óleo dissolvido são: Hidrogênio (H_2), Oxigênio (O_2), Nitrogênio (N_2), Metano (CH_4), Monóxido de Carbono (CO), Dióxido de Carbono (CO_2), Etileno (C_2H_4), Etano (C_2H_6), Acetileno (C_2H_2), Propano (C_3H_8), Propileno (C_3H_6) (DINIZ NETO et al., 2003). Cada tipo de falha afeta o óleo ou papel de uma forma diferente, gerando quantidades relativas características dos gases dissolvidos.

A análise dos gases pode ser feita de duas formas: levando em conta as relações de gases, em que são estabelecidas faixas de valores prováveis para a ocorrência de determinados tipos de faltas, ou levando em conta a formação do gás mais significativo para um tipo de falha, também chamado gás-chave (EL-HAWARY, 1998).

A concentração e relação entre os gases dissolvidos no óleo podem indicar a presença

das seguintes faltas (DINIZ NETO et al., 2003):

- 1) Arco elétrico: vibrações podem provocar perda ou mau contato de conexões do transformador, que culminam com a produção de grande volume de hidrogênio e acetileno e em menores volumes o metano e o etileno. Se o arco envolve a celulose, também podem ser formados monóxido e dióxido de carbono. Gás chave: acetileno.
- 2) Corona: descargas elétricas de baixa energia no enrolamento produzem hidrogênio e metano com pequenas quantidades de etano e etileno. Quantidades significativas de monóxido e dióxido de carbono podem ser devido a descargas na celulose. Gás chave: hidrogênio.
- 3) Sobreaquecimento do óleo: sobrecarga do transformador, aquecimento exagerado do óleo isolante e de outras partes do equipamento, ou óleo em más condições necessitando de reciclagem ou regeneração, resultando em transferência de calor ineficiente. Estes fatos podem provocar a decomposição de produtos, incluindo etileno e metano com pequenas quantidades de hidrogênio e etano. Traços de acetileno podem ser formados se o sobreaquecimento é grave ou envolve contatos elétricos. Gás chave: etileno.
- 4) Degradação da celulose: o papel composto de material isolante enrolado em torno das bobinas torna-se frágil e se deteriora. Grandes quantidades de monóxido e dióxido de carbono estão envolvidas no sobreaquecimento da celulose. Gases hidrocarbonetos, tais como o metano e etileno serão formados se o problema envolver a isolação do enrolamento, normalmente papel isolante impregnado de óleo. Gás chave: monóxido de carbono.

Muitos métodos, baseados na análise dos gases dissolvidos em óleo, têm sido desenvolvidos para detectar faltas incipientes em transformadores (DINIZ NETO et al., 2003). Estes métodos vêm sendo utilizados pelas concessionárias de energia, principalmente devido à simplicidade e ao baixo custo (EL-HAWARY, 1998). As próximas seções deste capítulo apresentam os métodos tradicionais baseados em DGA, que vêm sendo extensivamente utilizados pelas concessionárias de energia. Serão apresentados também alguns dos métodos já apresentados na literatura baseados em DGA e inteligência computacional.

5.1.1 Diagnóstico de Falhas baseados em Métodos Tradicionais

Os Métodos de Dörnenburg (Dörnenburg e Strittmatter, 1974), de Rogers (Rogers, 1978), Triângulo de Duval (Duval, 1974) e IEC (acrônimo em inglês para *International Electrotechnical Commission*) (IEC Std. 60599, 1999) são métodos tradicionais baseados na análise dos gases dissolvidos em óleo, os quais têm sido implementados de forma improvisada ou modificada por várias empresas fornecedoras de energia no mundo.

Essencialmente são métodos baseados nas razões entre os gases dissolvidos em óleo, cujos valores-limite destas razões são utilizados para determinação do tipo de falha. Em geral, estes métodos permitem detectar dois tipos de falhas: elétrica e térmica e se baseiam na comparação das concentrações de gases dissolvidos com um “benchmark”, contendo valores típicos de concentrações de gases. Caso um transformador apresente todos os valores de concentrações de gases dissolvidos abaixo dos valores típicos, então este transformador pode ser considerado em operação normal.

A. Método de Dörnenburg

O método de Doernenburg estabelece as condições de validade para aplicação do método de diagnóstico em função das concentrações individuais dos gases componentes das relações. A Tabela 5.1 apresenta as relações gasosas que classificam o tipo de falha a que o equipamento está submetido.

TABELA 5.1 Critério de identificação de falha de Dörnenburg (DÖRNENBURG e STRITTMATTER, 1974)

Tipo de falha	Relações entre concentrações de gases			
	Relações principais		Relações auxiliares	
	$\frac{CH_4}{H_2}$	$\frac{C_2H_2}{C_2H_4}$	$\frac{C_2H_6}{C_2H_2}$	$\frac{C_2H_2}{CH_4}$
Decomposição térmica	>1	<0.75	>0.4	<0.3
Descarga de baixa energia	<0.1	NS	>0.4	<0.3
Descarga de alta energia	<1 e >0.1	>0.75	<0.4	>0.3

NS - não significativo

De acordo com este critério de identificação de falhas, a Tabela 5.1 deve ser utilizada se no mínimo um dos gases que compõe as relações principais tem uma concentração superior ao dobro do valor da Tabela 5.2 e que para as relações auxiliares, pelo menos um dos gases tem uma concentração superior ao da Tabela 5.2.

TABELA 5.2 Concentração dos gases para validação do método de Dörnenburg

Tipos de gás	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
Concentração PPM (v/v)	200	50	15	60	15

Quando a norma IEEE C57.104-1991 recomendou a utilização do método de Doernenburg com a alteração dos valores da Tabela 5.2 para a Tabela 5.3, este método passou a ser utilizado para identificar as condições de normalidade do transformador. Isto quer dizer que a Tabela 5.1 pode ser associada à Tabela 5.3 da mesma maneira que é associada à Tabela 5.2.

TABELA 5.3 Concentração dos gases para identificação de condição de normalidade do transformador

Tipos de gás	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
Concentração PPM (v/v)	100	120	65	50	35

B. Método de Rogers

Este método permite o diagnóstico baseando-se inicialmente na Tabela 5.4, que apresenta as faixas de valores obtidos para as relações gasosas, as quais determinam um código. A partir dos códigos obtidos e utilizando a Tabela 5.5, obtém-se o diagnóstico final (Rogers, 1978).

TABELA 5.4 Tabela de definição dos códigos do método de Rogers

Relação de gases	Faixa de variação	Código
$\frac{CH_4}{H_2}$	≤ 0.1	5
	$>0.1, <1$	0
	$\geq 1, <3$	1
	≥ 3	2
$\frac{C_2H_6}{CH_4}$	<1	0
	≥ 1	1
$\frac{C_2H_4}{C_2H_6}$	<1	0
	$\geq 1, <3$	1
	≥ 3	2
$\frac{C_2H_2}{C_2H_4}$	<0.5	0
	$\geq 0.5, <3$	1
	≥ 3	2

TABELA 5.5 Tabela de diagnóstico do método de Rogers em função do código

$\frac{CH_4}{H_2}$	$\frac{C_2H_6}{CH_4}$	$\frac{C_2H_4}{C_2H_6}$	$\frac{C_2H_2}{C_2H_4}$	Diagnóstico
0	0	0	0	Deterioração normal
5	0	0	0	Descargas parciais
$\frac{1}{2}$	0	0	0	Sobreaquecimento – abaixo de 150°C
$\frac{1}{2}$	1	0	0	Sobreaquecimento de 150°C-200°C
0	1	0	0	Sobreaquecimento de 200°C-300°C
0	0	1	0	Sobreaquecimento de condutores
1	0	1	0	Correntes de circulação nos enrolamentos
1	0	2	0	Correntes de circulação no núcleo e tanque, sobreaquecimento em conexões
0	0	0	1	Descarga contínua
0	0	$\frac{1}{2}$	$\frac{1}{2}$	Arco com alta energia
0	0	2	2	Descarga contínua de baixa potência
5	0	0	$\frac{1}{2}$	Descarga parcial envolvendo o papel

C. Método do gás chave

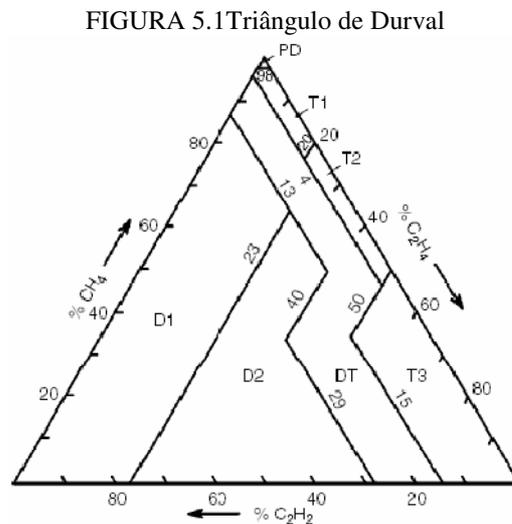
O método do gás-chave foi oficialmente proposto em 1974 (Pugh, 1974). Ele identifica o gás chave para cada tipo de falha e usa o seu percentual para identificá-la. O percentual do gás é baseado no total de gases combustíveis dissolvidos (TDCG) e é um número aproximado. A Tabela 5.6 mostra como se processa a identificação.

TABELA 5.6 Método do gás chave

Falha	Gás chave	Critério	Percentual gás
Arco	C ₂ H ₂	Grande quantidade de H ₂ e C ₂ H ₂ e menor quantidade de CH ₄ e C ₂ H ₄ . CO e CO ₂ podem também existir se a celulose estiver envolvida.	H ₂ : 60% C ₂ H ₂ : 30%
Corona	H ₂	Grande quantidade de H ₂ , algum CH ₄ , com pequena quantidade de C ₂ H ₆ e C ₂ H ₄ . CO e CO ₂ podem ser comparados se a celulose estiver envolvida.	H ₂ : 85% CH ₄ : 13%
Sobreaquecimento do óleo	C ₂ H ₄	Grande quantidade de C ₂ H ₄ , menor quantidade de C ₂ H ₆ , alguma quantidade de C ₂ H ₄ e H ₂ . Traços de CO e CO ₂ .	C ₂ H ₄ : 63% C ₂ H ₆ : 20%
Sobreaquecimento da celulose	CO	Grande quantidade de CO e CO ₂ . Gases hidrocarbonetos podem existir.	CO: 92%

D. Triângulo de Duval

Este método considera apenas o percentual relativo de concentração dos gases acetileno, etano e metano. A Figura 5.1 mostra evolução de gases relacionados a determinadas faltas em um triângulo, desenvolvido a partir de dados históricos na década de 1960 por Michel Duval (Duval, 1974). As coordenadas são definidas pela relação percentual de cada um dos gases em relação ao total de gases gerados. O critério pode identificar três faltas de origem elétrica e três faltas de origem térmica.



As legendas no triângulo da Figura 5.1 representam uma condição de falha, como descrito a seguir:

- PD: descargas parciais;
- T1: falha térmica com temperatura $T < 300^{\circ}\text{C}$;
- T2: falha térmica com temperatura: $300^{\circ}\text{C} < T < 700^{\circ}\text{C}$;
- T3: falha térmica com temperatura $T > 700^{\circ}\text{C}$;
- D1: descargas de baixa energia;
- D2: descargas de alta energia;
- DT: mistura de faltas.

O triângulo de Duval não se aplica à maioria dos transformadores a não ser quando os três gases explosivos citados aparecem juntos, representando perigo iminente. Sua utilização se dá a partir da aplicação das técnicas anteriores ou da Tabela 5.8 a seguir. Antes da confirmação do problema, é necessário que ao menos um dos gases hidrocarbonetos ou

hidrogênio estejam na condição 3 do IEEE e a taxa de geração (G2 da Tabela 5.7) esteja aumentada.

TABELA 5.7 Tabela que substitui o uso do método do IEC

Gás	Limite L1	Limite G1 (PPM/ mês)	Limite G2 (PPM/ mês)
H ₂	100	10	50
CH ₄	75	8	38
C ₂ H ₂	3	3	3
C ₂ H ₄	75	8	38
C ₂ H ₆	75	8	38
CO	700	70	350
CO ₂	7000	700	3500

E. Método do IEC

O critério IEC 60599 (IEC Std. 60599, 1999) é largamente utilizado por empresas fornecedoras de energia para interpretar a análise de gases dissolvidos. A Tabela 5.8 apresenta o critério para a interpretação de DGA de acordo com o IEC 60599.

TABELA 5.8 Critério IEC 60599 para interpretação de DGA

	Tipo de falha	$\frac{C_2H_2}{C_2H_4}$	$\frac{CH_4}{H_2}$	$\frac{CH_4}{C_2H_6}$
PD	Descarga parcial	NS	<0.1	<0.2
D1	Descargas de baixa energia	>1	0.1-0.5	>1
D2	Descargas de alta energia	0.6-2.5	0.1-1	>2
T1	Falha térmica T<300°C	NS	>1 but NS	<1
T2	Falha térmica 300°C< T<700°C	<0.1	>1	1-4
T3	Falha térmica T>700°C	<0.2	>1	>4

NS – Não significativo qualquer que seja o valor

F. Limitações das abordagens tradicionais

Apesar de todos os critérios já desenvolvidos, a busca por métodos mais confiáveis usando DGA é ainda um tópico de grande interesse, visto que estas abordagens apresentam algumas limitações tais como:

- a) De acordo com IEEE Std C57.104.1991 (1992), a análise dos gases dissolvidos em óleo atualmente não pode ser considerada uma ciência, mas sim uma arte sujeita a variabilidade, desde que transformadores de diferentes tamanhos, estruturas e fabricantes apresentam características diferentes e como consequência é difícil obter um consenso quanto a valores limites dos gases e seus significados.

- b) A natureza heurística e empírica dos métodos baseados em razões tem levado a diferentes interpretações; a aplicação de métodos diferentes sobre um conjunto idêntico de dados pode produzir diferentes diagnósticos sobre a condição do transformador.
- c) Devido à imprecisão de um método baseado em DGA, diversos métodos têm de ser utilizados ao mesmo tempo pelas concessionárias para se chegar ao diagnóstico. Nestes casos, os especialistas da área devem utilizar em conjunto seus conhecimentos para chegar à conclusão final.
- d) O diagnóstico é às vezes impossível de ser obtido devido à inabilidade destes métodos para fornecer interpretação para toda combinação possível de valores de razões (problema de “*no decision*”), com exceção do Triângulo de Duval. Com isto, a interpretação pode acabar dependendo do julgamento de um especialista, que pode causar mais controvérsias.
- e) O método do IEC apresenta um problema de sobreposição entre as faltas, podendo ao mesmo tempo diagnosticar dois tipos de falha (descarga de baixa e descarga de alta energia).
- f) Os métodos não possuem meios para mostrar aos engenheiros a evolução de uma falha incipiente após ser detectada.

5.1.2 Diagnóstico de Faltas baseado em Abordagens Não Tradicionais

As abordagens não tradicionais se resumem à aplicação de técnicas da Inteligência Computacional para realizar o diagnóstico do estado dos transformadores baseado nas concentrações de gases dissolvidos em óleo. Elas visam superar as limitações impostas pelas abordagens tradicionais, conseqüentemente objetivam o aumento da taxa de sucesso no diagnóstico.

A. Sistemas de Diagnóstico baseados em Lógica Difusa

Em Su e Li (2006), apresenta-se um sistema especialista para diagnóstico de faltas em transformadores de potência com uma interface para entrada de dados em linguagem natural. As palavras são usadas para identificar a regra a ser ativada. A saída é dada em linguagem natural.

Alguns estudos têm reportado a eficiência e as dificuldades de usar lógica difusa e

RNAs no diagnóstico de transformadores. Em Németh (2009) e Morais et al. (2005), são utilizados sistemas difusos, em que a proporção dos gases é *fuzzificada* para representar a natureza imprecisa do DGA. Sua limitação é que a eficiência do sistema depende do conhecimento integral dos especialistas sobre o assunto.

Em Hmood et al. (2012), os autores criaram modelos difusos para os métodos de Análise de Gases Dissolvidos existentes. O sistema difuso apresentou saídas equivalentes para cada método, sendo que foram testadas 20 amostras. Para evitar respostas com mais de uma falha, na saída do sistema foram colocadas funções de pertinência triangulares sem sobreposição. Os resultados foram compatíveis com os padrões, mas não se apresentaram equivalentes entre si. Um dos objetivos do sistema é eliminar a necessidade de um especialista.

B. Sistemas de Diagnóstico baseados em Redes Neurais Artificiais

Em Salami e Pahlevani (2008), foi implementado um sistema de diagnóstico de faltas de transformadores de potência baseado em análise de gases dissolvidos com uma rede neural, utilizando algoritmo *backpropagation* (retropropagação). As entradas da rede são os 5 gases. A rede apresenta uma baixa taxa de erro para diagnosticar 3 faltas (térmica, corona, arco elétrico).

Em Yang-jing et al. (2007), foi implementado um sistema de diagnóstico de faltas de transformadores de potência baseado em análise de gases dissolvidos (AGD) com o algoritmo *backpropagation* modificado. Esta modificação implica que a taxa de aprendizagem passa a ser uma função das derivadas dos erros sobre os deltas dos pesos consecutivos e adiciona uma fração do ajuste do peso anterior ao valor do ajuste do peso corrente (o *momentum* evita mudanças extremas no gradiente, provocadas por anomalias locais). Com isto, a taxa de aprendizagem e o *momentum* são alterados a cada iteração. Isto reduziu o número de épocas para treinamento. Além disso, o método utiliza o *Redefined Three-Ratio* (RTR) como guia para obter os padrões de treinamento. Após o treinamento, para 10 padrões de teste, o algoritmo convencional *backpropagation* e o RTR, aplicados isoladamente, não conseguiram classificar um padrão, enquanto que a ferramenta desenvolvida em Yang-jing et al. (2007) conseguiu diagnosticar corretamente todos os padrões, utilizando o *backpropagation* modificado.

Em Thang e Aggarwal (2003), uma base de dados de 755 amostras é treinada. Os componentes planos do SOM são utilizados para descobrir os gases-chave nos grupos formados, os quais permitem identificar grupos relativos às faltas térmicas, elétricas, térmicas

e elétricas simultaneamente e à condição de operação normal. O SOM também permitiu a identificação do histórico de faltas.

C. Sistemas de Diagnóstico baseados em Sistemas Neuro-Difusos

Em Pateritsas et al. (2007), trabalho mencionado no Capítulo 3, são extraídas regras difusas do SOM para criar um sistema difuso para diagnosticar faltas incipientes de transformadores de potência. Além do treinamento do SOM com 7 gases de entrada, regras difusas foram extraídas para formação de um sistema difuso capaz de explicar os resultados do SOM. Foi alcançada uma taxa de sucesso de 96,67% de equivalência entre o SOM e o sistema difuso, com uma base de dados pequena: 30 amostras apenas para teste.

Em Castro e Miranda (2005), o sistema de inferência difusa produz melhor taxa de sucesso do que os critérios do IEC 60599 (IEC Std. 60599, 1999). A rede neural artificial apresenta 3 entradas e uma camada escondida, composta por 7 neurônios, cuja função de ativação é a sigmóide positiva. O conceito de *f-duality* foi utilizado para extração de regras difusas, gerando 7 funções de pertinência para cada entrada, o que não possibilitou um sistema transparente. Um processo de aproximação, auxiliado pelo algoritmo dos mínimos quadrados recursivo, converte as funções de pertinência em 5 funções triangulares. Para cada entrada, as 5 funções de pertinência são combinadas, formando um sistema de 5^n regras, onde n é o número de entradas. Quando o sistema ganhou em transparência, o número de regras aumentou de 60 para 125. Assim, sua maior limitação é o grande número de regras, que de alguma forma dificulta a interpretação dos resultados. Foi obtida uma taxa de sucesso de 100% no treinamento e 97,84%, no teste.

Como em Castro e Miranda (2005a), o sistema de diagnóstico apresentado, em Castro e Miranda (2005b), foi também desenvolvido usando a metodologia para extrair sistemas difusos a partir de um rede neural multicamada e o banco de dados TC-10 de transformadores faltosos inspecionados em operação. Neste artigo, como na metodologia apresentada, os autores usaram a *fuzzificação* da saída do sistema de inferência difusa, que levou a 100% de diagnósticos corretos para os dados de treinamento e validação. Entretanto, o sistema extraído também sofre de desvantagens por apresentar um grande número de regras (95 regras).

5.2 Sistema proposto para Diagnóstico de Falhas Incipientes em Transformadores de Potência

Nesta seção, é apresentado o desenvolvimento do sistema para diagnóstico de falhas incipientes em transformadores de potência.

5.2.1 Banco de dados para desenvolvimento/treinamento do sistema de diagnóstico

O sistema de diagnóstico de falhas proposto foi desenvolvido a partir do banco de dados de transformadores faltosos apresentado na Publicação IEC 60599 (IEC Std. 60599, 1999), acrescido de dados fornecidos pela CELPA (Empresa de Energia do Pará, S.A.) e dados retirados da literatura.

O banco de dados é composto por 309 amostras, sendo que 234 foram utilizadas para treinamento do SOM e 75 amostras foram utilizadas para teste do sistema desenvolvido. Cada amostra do banco de dados contém concentrações dos gases hidrogênio (H_2), metano (CH_4), etano (C_2H_6), etileno (C_2H_4) e acetileno (C_2H_2) e a condição do transformador. As 309 amostras incluem 148 casos de falhas térmicas e 161 casos de falhas elétricas. As falhas estão divididas em cinco classes específicas de acordo com a Tabela 5.9.

TABELA 5.9 Tipos de falhas da base de dados

	Tipos de Falhas	Número de amostras
T1	Falta térmica – Temperatura < 300°C	77
T2	Falta térmica – Temperatura > 300°C	71
PD	Descarga Parcial (corona)	29
DL	Descarga de Baixa Energia	30
DH	Descarga de Alta Energia	102

A Tabela 5.10 apresenta a divisão da base de dados para treinamento e para teste do sistema proposto.

TABELA 5.10 Divisão da base de dados de treinamento e teste para o transformador de potência

Tipos de falhas	Base de Treinamento	Base de Teste
T1	61	16
T2	55	16
PD	23	6
DL	25	9
DH	74	28
Total	234	75

As razões utilizadas pelo método IEC foram escolhidas como entradas para o sistema de diagnóstico:

$$\frac{C_2H_2}{C_2H_4} \in [0, 18.53846]$$

$$\frac{CH_4}{H_2} \in [0, 498.33]$$

$$\frac{C_2H_4}{C_2H_6} \in [0, 400.0]$$

É importante enfatizar que baseado em um conhecimento a priori da Tabela 5.8 do método IEC decidiu-se que todo valor de razão maior que 4 seria igualado a 4, obtendo-se assim o novo domínio das razões utilizadas como entrada para a rede:

$$\frac{C_2H_2}{C_2H_4} \in [0, 4]$$

$$\frac{CH_4}{H_2} \in [0, 4]$$

$$\frac{C_2H_4}{C_2H_6} \in [0, 4]$$

Após esta transformação todas as entradas foram normalizadas no intervalo [0 1].

5.2.2 Treinamento da Rede de Kohonen e Identificação dos grupos

Várias configurações de mapas auto-organizáveis foram treinadas, variando-se o número de neurônios na grade, a topologia da grade e a taxa de aprendizagem. A Tabela 5.11 mostra alguns resultados dos treinamentos realizados.

TABELA 5.11 Treinamentos do SOM para a base de dados de faltas em transformadores

Topologia do mapa	Forma	Função da taxa de aprendizagem	Erro de quantização média	Erro topográfico
[4x11]	Hexagonal	Power ⁶ (0.1)	0.1891	0.0304
[4x11]	Retangular	Inv ⁷ (0.1)	0.1834	0.0348
[4x11]	Retangular	Power (0.1)	0.1808	0.0304
[4x16]	Retangular	Inv (0.5)	0.1765	0.0726
[11x4]	Hexagonal	Inv (0.1)	0.178	0.05

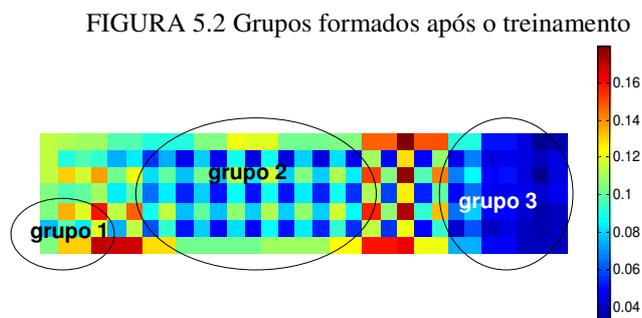
⁶ A função da taxa de aprendizagem é inversamente decrescente.

⁷ A função da taxa de aprendizagem é exponencialmente decrescente .

Como já exposto no Capítulo 3, o erro de quantização média (resolução) representa a média das distâncias entre cada vetor de dados e o correspondente vetor de pesos do neurônio vencedor (BMU). Assume-se que quanto menor o erro de quantização, mais ajustado o neurônio vencedor (BMU) estará aos vetores de entrada. Já o erro topográfico quantifica a capacidade do mapa em representar a topologia dos dados de entrada. Ele é calculado verificando-se a percentagem de vetores de dados para os quais o BMU e um segundo BMU não são unidades vizinhas no mapa.

Após a análise dos diversos resultados de treinamento, a arquitetura do SOM escolhida possui 3 nós na camada de entrada para as 3 razões dos gases (IEC) e 64 neurônios na camada de saída, organizados de forma bidimensional: 4x16. Os dados foram normalizados por histograma discreto. A função de vizinhança selecionada foi a Gaussiana. Esta configuração foi escolhida por apresentar o menor erro de quantização média e menor erro de classificação para os dados de treino.

A Figura 5.2 apresenta a Matriz-U com os grupos formados após o treinamento do SOM com a base de dados de faltas. De acordo com a escala da Matriz-U, a cor tendendo para azul-escuro representa as menores distâncias entre as unidades de mapa, portanto indicam a formação de grupos. Já a cor tendendo para vermelho, representa as maiores distâncias, indicando bordas de grupos.



A Matriz-U permite distinguir a formação de 3 grupos que de acordo com a base de dados são: Grupo 1 (Descargas Parciais), Grupo 2 (Faltas Térmicas) e Grupo 3 (Faltas Elétricas). Para melhor visualização dos grupos, a Figura 5.3 apresenta o mapa rotulado resultante, onde se tem:

- Neurônio 4 pertencente à classe Descarga Parcial (PD);
- Neurônios 25, 29 a 33, 35 a 38, 40 a 42 pertencentes à classe T2 (temperatura > 300° C).
- Neurônios 5 a 6, 9, 13, 15 a 24, 27 a 28 pertencentes à classe T1

(temperatura < 300° C).

- Neurônios 60 e 64 pertencentes à classe DL (Descarga de Baixa Energia).
- Neurônios 48 a 49, 52 a 57,59, 61 a 63 pertencentes à classe DH (Descarga de Alta Energia).

FIGURA 5.3 Rótulos das unidades de mapa com os 5 grupos formados

	T1	T1	T1	T1	T1	T2	T2	T2	T2	T2		DH	DH	DH	DH	
	T1			T1	T1		T2		T2	T2			DH		DH	
			T1	T1	T1	T1	T2	T2					DH	DH	DH	
PD			T1	T1	T1	T1	T2	T2	T2			DH	DH	DH	DL	DL

Os rótulos são atribuídos de acordo com o maior número de elementos das classes que foram atribuídas aos neurônios, lembrando que se trata de um problema de classificação, no qual as classes existentes são conhecidas, o que permite tal rotulação.

No processo de rotulação, a ferramenta *SOM Toolbox* permite que se observe o número de padrões alocados a cada neurônio na grade (Figura 5.4). Assim, o treinamento executado apresentou os seguintes resultados:

- Seis amostras foram classificadas incorretamente, ou seja, alcançou-se uma taxa de acertos em relação ao diagnóstico de faltas de 97,43%;
- As classes T1 e T2 tiveram 98,33% e 98,18% respectivamente de acertos; a classe PD 100% de acertos, a classe DH 98,64% de acertos. A classe DL foi a classe que apresentou maior número de erros, sua taxa de acertos ficou em 88% de acertos.

FIGURA 5.4 Mapa rotulado dividido em duas partes (A) e (B), indicando o número de padrões pertencentes a cada classe

	T1(05)	T1(06)	T1(08)	T1(06)	T1(06)	T1(01) T2(05)	T2(04)
	T1(01)			T1(01)	T1(01)		T2(02)
			T1(06)	T1(03)	T1(02)	T1(02) T2(01)	T2(01)
PD (23)			T1(06)	T1(07)	T1(04)	T1(02)	T2(03)

(A)

T2(11)	T2(09)	T2(03)		DH(08)	DH(10)	DH(07)	DH(15)
	T2(03)	T2(01)			DH(01)		DH(02)
T2(01)					DH(02)	DH(02)	DH(01)
T2(01)	T2(10)		DH(02)	DH(09)	DL(03) DH(14)	DL(01)	DL(17) DH(01)

(B)

5.2.3 Extração de Sistema Difuso a partir do SOM

Após treinamento do SOM e a identificação dos grupos, o processo para extração das regras difusas pode ser iniciado. O primeiro passo é selecionar entre todos os neurônios, que pertencem a um grupo específico, aquele que tem o maior potencial para representar o centro do grupo. A seleção é feita através do uso da equação (4.1) que permite calcular o potencial de um neurônio.

A Tabela 5.12 apresenta os pesos dos neurônios de maior potencial de cada grupo e os valores mínimo e máximo dos padrões de entrada alocados a estes grupos.

TABELA 5.12 Resultados para extração de regras

Classe/Grupo	Pesos dos neurônios com maior potencial	Valores min e max dos padrões de entrada		
		(C ₂ H ₂)/(C ₂ H ₄)	(CH ₄)/(H ₂)	(C ₂ H ₄)/(C ₂ H ₆)
T1	0.0295 1.6707 0.5526	[0 0.3846]	[1.0067 4.0]	[0.0762 1.2857]
T2	0.0652 1.3636 2.766	[0 0.1]	[0.0619 4.0]	[0.9783 4.0]
PD	0.0217 0.02077 0.1071	[0 1.8333]	[0 0.1102]	[0 0.1667]
DL	1.1111 0.27 3.6	[0.9231 4.0]	[0.0833 0.535]	[1.0 4.0]
DH	0.8125 0.3438 3.8333	[0.620 2.6607]	[0.1014 0.9167]	[1.75 4.0]

A partir das informações de cada grupo apresentadas na Tabela 5.12 e de acordo com (4.6) as seguintes regras foram extraídas:

R₁: Se x_1 está em torno de 0,0295 (BAIXO) e x_2 está em torno de 1,6707 (MÉDIO) e x_3 está em torno de 0,5526 (BAIXO) então $y = 1$ (T1)

R₂: Se x_1 está em torno de 0,0652 (BAIXO) e x_2 está em torno de 1,3636 (MÉDIO) e x_3 está em torno de 2,766 (ALTO) então $y = 2$ (T2)

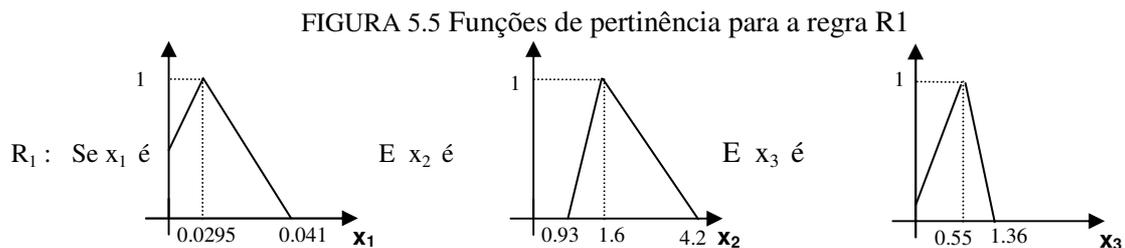
R₃: Se x_1 está em torno de 0,0217 (BAIXO) e x_2 está em torno de 0,02077 (BAIXO) e x_3 está em torno de 0,107 (BAIXO) então $y = 3$ (PD)

R₄: Se x_1 está em torno de 1,1111 (MÉDIO) e x_2 está em torno de 0,27 (BAIXO) e x_3 está em torno de 3,6 (ALTO) então $y = 4$ (DL)

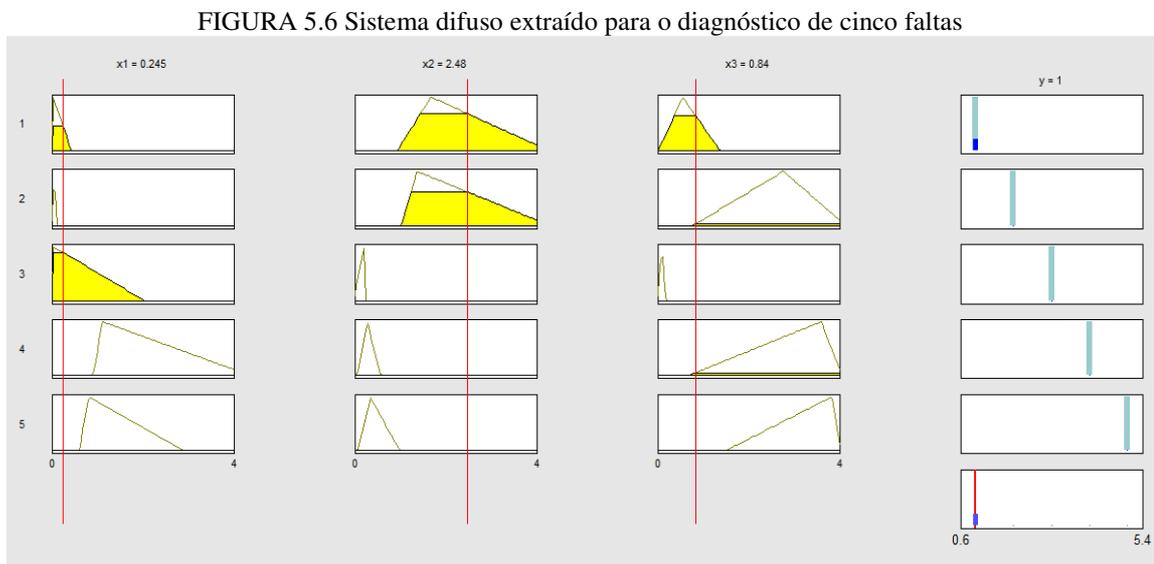
R₅: Se x_1 está em torno de 0,8125 (MÉDIO) e x_2 está em torno de 0,3438 (BAIXO) e x_3 está em torno de 3,833 (ALTO) então $y = 5$ (DH)

onde $x_1 = C_2 H_2 / C_2 H_4$, $x_2 = CH_4 / H_2$ e $x_3 = C_2 H_4 / C_2 H_6$

Os termos linguísticos BAIXO, MÉDIO E ALTO podem também ser usados para representar as funções de pertinência. Estes termos foram definidos intuitivamente considerando os valores mínimo e máximo do padrão de entrada, considerando a base de dados completa, neste caso [0 4]. Para ilustração, a Figura 5.5 apresenta as funções de pertinência extraídas para a Regra 1. Estas funções foram obtidas de acordo com as equações (4.3), (4.4) e (4.5) e com as informações do grupo/classe 1 (T1) apresentado na Tabela 5.12.



A Figura 5.6 apresenta as funções de pertinência e regras extraídas do treinamento do mapa de Kohonen, utilizando a *Fuzzy Logic Toolbox* do Matlab.



A partir das regras extraídas, é possível verificar sua conformidade com a Tabela IEC. A diferença é que com as regras difusas não se tem uma representação *crisp* do domínio de entrada e isto pode trazer melhores resultados de classificação, já que a separação das classes/faltas não é linear. Através destas regras extraídas do SOM, o especialista poderá verificar porque para determinados valores de entrada (razões de gases) o sistema difuso

apresenta como resposta determinado tipo de falta e com seu conhecimento pode validar ou não os resultados apresentados pelo sistema difuso.

A Tabela 5.13 apresenta os resultados de classificação do sistema difuso, considerando dados de treino e dados de validação.

TABELA 5.13 Resultados de classificação do Sistema Difuso

Método	Dados de Treinamento				Dados de Validação			
	% de diagnóstico correto	NI	E	ND	% de diagnóstico correto	NI	E	ND
Sistema Difuso	98.28	0	4	0	94.66	0	4	0

NI- Faltas Não Identificadas

E – Erro de classificação

ND – Não-decisão

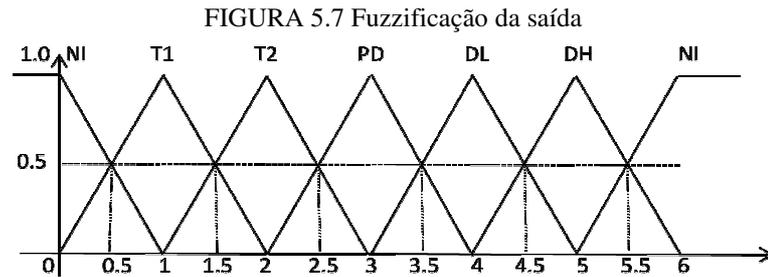
É importante mencionar que ocorreram 3 erros de equivalência, já que o SOM apresentou 6 erros e o FIS, 4.

5.2.3 Processo de *Fuzzificação* da Saída do Sistema Difuso

Para o problema de classificação de faltas o sistema difuso (FIS) extraído trabalha como uma função discriminante, na qual seis linhas discriminatórias podem ser consideradas em 0.5, 1.5, 2.5, 3.5 e 4.5 e com a classificação de acordo com:

- 1) Se a saída do FIS estiver entre]0.5 1.5[então a falha será T1.
- 2) Se a saída do FIS estiver entre]1.5 2.5[então a falha será T2.
- 3) Se a saída do FIS estiver entre]2.5 3.5[então a falha será PD.
- 4) Se a saída do FIS estiver entre]3.5 4.5[então a falha será DL.
- 5) Se a saída do FIS estiver entre]4.5 5.5[então a falha será DH.
- 6) Se a saída do FIS for <0.5 ou >5.5 então a falha não será identificada.

Para se levar em consideração possíveis casos com sintomas não muito claros (casos em que a saída do FIS é um valor próximo ou nos limites do domínio de cada falta), decidiu-se por, de acordo com Castro et al. (2004), *fuzzificar* a saída do sistema difuso extraído usando as funções de pertinência da Figura 5.7. É importante enfatizar, ainda de acordo com Castro et al. (2004), que esta *fuzzificação* é razoável desde que os códigos vizinhos de saída do sistema difuso extraído representem faltas consideradas vizinhas.



Através deste processo de *fuzzificação* seis “áreas de possibilidades” são criadas. Quando a saída do FIS está em alguma destas áreas, o tipo de falta será decidido de acordo com o grau de pertinência no grupo (caso uma decisão “crisp” seja desejada), caso contrário, como alternativa, um resultado difuso pode ser apresentado com duas opções, afetadas pelos seus valores de possibilidade.

Para exemplificar, considerar dois exemplos em que a saída do sistema difuso é 4.55 e 4. A saída do FIS em 4.55 será classificada como falta DL com possibilidade 0.45 ou como DH com possibilidade 0.55. No segundo caso, a falta será classificada como DL com possibilidade 1. Esta *fuzzificação* com funções de pertinência triangulares apresenta casos em que só se formula uma hipótese, e intervalos em que duas hipóteses são formuladas. Isto significa que a hipótese com um menor grau de possibilidade não deve ser descartada como uma possibilidade.

A Tabela 5.14 apresenta os resultados de classificação para o sistema difuso com saída *fuzzificada*.

TABELA 5.14 Resultados de Classificação do FIS com saída fuzzificada

Método	Dados de Treinamento				Dados de Validação			
	% P	NI	E	F	% P	NI	E	F
FIS com saída <i>fuzzificada</i>	100	0	0	29	100	0	0	16

%P – Porcentagem de casos em que a saída *fuzzificada* incluiu a resposta correta.

NI – Faltas Não Identificadas

E – Erro de classificação (número de casos onde a saída indicou duas possibilidades e nenhuma era a correta)

F – número de casos onde a saída difusa indicou duas possibilidades sendo uma delas a correta

É importante notar que, para calcular a percentagem do diagnóstico correto do FIS com a saída *fuzzificada* (o resultado difuso apresenta duas hipóteses, devido aos seus valores de possibilidade), considera-se como um diagnóstico correto qualquer caso em que uma das hipóteses sugeridas pela saída *fuzzificada* do FIS esteja correta, mesmo se nesta hipótese correta o valor da possibilidade seja menor que 0.5 (embora sendo um valor pequeno, é uma possibilidade que não deve ser excluída como correta). O diagnóstico só é considerado

incorreto se as duas hipóteses do FIS *fuzzificado* não são corretas (nenhum caso ocorreu).

Os resultados obtidos evidenciam a vantagem da *fuzzificação* da saída do sistema difuso. Através da *fuzzificação*, verifica-se que os resultados apontam ou no mínimo não excluem a correta causa da falha. A representação crisp da saída apresenta um resultado com casos de erros de diagnóstico, por isso a *fuzzificação* da saída do sistema difuso é mais informativa e mais correta.

5.3 Comparação com um Método Tradicional de Análise dos Gases Dissolvidos em Óleo

O critério IEC 60599 foi aplicado na mesma base de dados usada para o desenvolvimento do sistema proposto. A Tabela 5.15 apresenta os resultados de classificação do método IEC e para comparação, também apresenta os resultados do sistema difuso extraído e do sistema difuso com saída *fuzzificada*.

TABELA 5.15 Resultados de Classificação - Comparação

Método	Dados de Treinamento			Dados de Validação				
	% de diagnóstico correto	NI	E	ND	% de diagnóstico correto	NI	E	ND
IEC 60599	89.31	22	0	3	81.3	8	0	6
FIS	98.28	0	4	0	94.66	0	4	0
FIS com saída <i>fuzzificada</i>	100	0	0	0	100	0	0	0

NI- faltas não identificadas

E – erro de classificação

ND – Não-decisão

A partir dos resultados apresentados, é possível observar que a percentagem de diagnósticos corretos do FIS é maior que o obtido pelo IEC 60599. O problema de “não-decisão”, característico do método IEC e outros métodos tradicionais, foi superado pelo sistema FIS que conseguiu classificar corretamente todos os 9 casos de “ não-decisão” do IEC.

É possível também verificar a partir dos resultados obtidos que em 30 casos de faltas não identificadas pelo IEC, o FIS classifica 22 casos corretamente. A principal razão para esta diferença está no fato de que o IEC é um critério que não cobre o domínio completo da entrada (9 razões dos gases). Outro ponto a destacar é a evidente superioridade dos resultados do FIS devido principalmente ao fato da separação dos grupos não ser linear.

Considerando os resultados do FIS com saída *fuzzificada*, obteve-se 100% de casos identificados corretamente tanto para dados de treino como dados de teste. Em poucos casos o sistema formulou duas hipóteses em que a resposta correta não foi o valor de maior possibilidade, entretanto, esta resposta não foi descartada como uma possível alternativa. Deve ser enfatizado que respostas deste tipo apresentam mais informação que as providenciadas pelo IEC, em que uma grande quantidade de casos não foi identificada, não providenciando alternativas para análise dos engenheiros ou especialistas. Por comparação com os resultados do FIS sem *fuzzificação*, pode-se perceber que a *fuzzificação* leva à melhores resultados de classificação e mais informação para o engenheiro - e o que é mais importante, informações corretas.

5.4 Comparação com outros Sistemas Neuro-Difusos

A Tabela 5.16 apresenta os resultados dos sistemas estudados na Seção 5.1.2-C para comparação a partir de aspectos de interpretabilidade e precisão.

TABELA 5.16 Comparando com abordagens neuro-difusas

Sistema Diagnóstico	% diagnóstico correto (base de treinamento)	% diagnóstico correto (base de teste)	Número de Regras	Tipos de Faltas
Castro e Miranda (2005a)	100	97.84	125	5
Castro e Miranda (2005b)	100	100	95	5
Naresh et al. (2008)	-	96.67	69	5
Trabalho proposto	100	100	5	5

É importante enfatizar que uma comparação da taxa de sucesso entre estes sistemas, não é totalmente apropriada, visto que os sistemas foram testados com diferentes bancos de dados. O que se quer destacar com esta tabela comparativa é a complexidade dos sistemas com relação ao número de regras usadas para representação do conhecimento acerca do problema. A partir da Tabela 5.12, é possível verificar que todos os sistemas têm resultados de classificação satisfatórios; entretanto o sistema proposto representa o conhecimento do problema com apenas 5 regras, enquanto os outros requerem um número de regras significativo. Um grande número de regras torna a interpretação e a validação dos resultados do diagnóstico mais complexas.

5.5 Conclusão do capítulo

Este capítulo apresentou um sistema de diagnóstico de faltas para transformadores de energia elétrica, desenvolvido utilizando uma nova metodologia para extração de regras difusas a partir de mapas auto-organizáveis de Kohonen. O SOM foi treinado para capturar o conhecimento de um banco de dados de transformadores faltosos inspecionado em operação e o conhecimento para detectar o tipo de falha foi transformado em regras difusas. Estas regras difusas extraídas podem ajudar os usuários do sistema a terem mais confiança no diagnóstico de faltas produzido, permitindo que o especialista interprete de forma mais eficiente o sistema físico.

Outro ponto que se deve dar ênfase é que com o uso da nova metodologia apresentada para extrair um sistema de inferência difusa a partir do SOM, e diferente de outros sistemas difusos apresentados na literatura que geram centenas de regras, foi possível desenvolver um sistema difuso com apenas 5 regras para representar o conhecimento a respeito do problema. O resultado mostra que o sistema difuso proposto com estas 5 regras apresenta bons resultados na tarefa do diagnóstico de faltas.

A vantagem da *fuzzificação* da saída dos sistemas foi evidenciada. Na verdade, os exemplos práticos têm mostrado que, quando a *fuzziness* (a dificuldade de distinção, imprecisão) é reconhecida na expressão linguística do conhecimento, o diagnóstico de faltas nos transformadores de potência tem mostrado, ou ao menos (em casos ordinários) não excluído, a causa correta do defeito, enquanto uma resposta *crisp* (verdadeiro ou falso) conduz, em alguns casos, a erros de diagnóstico. Uma resposta difusa é, portanto, mais informativa e mais correta.

Finalmente, tornou-se evidente que a separação de grupos encontrados na análise de gás dissolvido em óleo é não-linear ou, ao menos, não representável por um intervalo de partição do domínio. (Ao tentar construir um sistema com regras proposicionais – uma tabela *crisp* – a partir dos intervalos obtidos com a metodologia, obteve-se uma taxa de sucesso 81,33% com 5 amostras não identificadas e 9 erros.) Esta evidência é decorrente do fato de que ambos, Sistemas de Inferência Difusa extraídos do SOM e sistemas difusos com saída *fuzzificada* produzem resultados com precisão superior à do código IEC. Esta estrutura de dados com intervalos não-separáveis torna mais difícil a captura de conhecimento para humanos e reforça a utilidade de ter uma abordagem envolvendo sistemas inteligentes.

6. Conclusões Gerais

Os mapas auto-organizáveis vêm sendo aplicados em diversas áreas (redes de computadores, medicina, robótica, engenharias, Internet e outras), porém, como outras topologias de RNAs, desperta também o interesse científico pela falta de explanação do conhecimento encapsulado em seus resultados. O foco primário deste trabalho foi tentar solucionar este problema através do desenvolvimento de uma nova metodologia para extração do conhecimento escondido nos mapas auto-organizáveis de Kohonen por meio de regras difusas do tipo Takagi-Sugeno de Ordem Zero.

Algumas metodologias vêm sendo apresentadas na literatura para extração de conhecimento de redes neurais, entretanto poucos trabalhos já foram apresentados com o intuito de extrair conhecimento de mapas auto-organizáveis de Kohonen. Os trabalhos que já foram desenvolvidos apresentam algumas limitações principalmente no que diz respeito ao número excessivo de regras extraídas (o que dificulta a interpretação do sistema).

O segundo foco desta tese, o desenvolvimento de um sistema para diagnosticar faltas em transformadores de potência, atendeu os objetivos definidos. Melhores resultados classificatórios foram obtidos quando comparados a uma das metodologias tradicionais utilizadas pelas concessionárias de energia, além de garantir um sistema inteligente com capacidade de explicação de seus resultados a partir da análise de regras (o que gera mais confiança no resultado de classificação). Os resultados também mostraram as vantagens do sistema de diagnóstico desenvolvido sobre alguns sistemas já desenvolvidos e apresentados na literatura utilizando sistemas neuro-difusos.

Neste capítulo, as conclusões finais sobre esta tese são apresentadas. A seção 6.1 discute as contribuições da metodologia e do sistema desenvolvido para o problema do diagnóstico de faltas incipientes em transformadores de potência. A seção 6.2 discute algumas limitações da metodologia e propõe algumas linhas de pesquisa para trabalhos futuros.

6.1. Contribuições da tese

Conforme discutido em capítulos anteriores, as Redes Neurais Artificiais alcançariam maior aceitação se a capacidade de explicação se tornasse uma parte integral da sua funcionalidade. Portanto, com o objetivo de proporcionar às RNAs tal funcionalidade, nesta tese, foi apresentada uma metodologia simples e original. Trata-se de uma abordagem de extração de regra difusa baseada na equivalência matemática entre a RNA e o sistema difuso

Takagi-Sugeno de Ordem Zero. Diferente de metodologias anteriores, esta metodologia extrai sistemas através de um algoritmo simples e com número de regras pequeno, fornecendo a desejada explicação para o resultado da Rede Neural.

O processo de extração se baseou no fato de que o processo de ordenação compõe um mapa, cujos valores das sinapses de cada neurônio são aproximadamente iguais à média ponderada dos conjuntos de Voronoi dos neurônios vizinhos, segundo T. Kohonen, ou seja, os pesos sinápticos dos neurônios das classes tornam-se generalizações dos dados projetados no mapa. Assim, aplicou-se a fórmula do potencial do neurônio para definir os pontos de maior pertinência para a função triangular.

De acordo com a taxonomia desenvolvida para avaliar os algoritmos de extração de regras, a metodologia pôde ser avaliada. Os resultados da avaliação foram resumidos na Tabela 6.1.

TABELA 6.1 Avaliação da Metodologia Proposta

Abordagem quanto à transparência	Portabilidade	Formato da Regra	Qualidade das Regras (Para a base IRIS)			Complexidade
			Taxa de Sucesso	Fidelidade	Facilidade de Compreensão	
Eclética	Domínio Independente e Restrita ao SOM	Difusa	Alta	Alta	Alta	Baixa

O desenvolvimento de um sistema de diagnóstico de faltas incipientes para transformadores de potência com base na Análise dos Gases Dissolvidos foi o segundo objetivo desta tese. Alguns sistemas de diagnóstico baseados em RNAs têm sido apresentados na literatura, entretanto, poucos são apresentados com o uso do SOM e com um número de regras capaz de facilitar a explanação da solução do problema. Nesta tese, a metodologia desenvolvida tratou deste problema e gerou um sistema difuso classificador com uma certa vantagem sobre o método IEC e sobre alguns métodos apresentados na literatura usando sistemas neuro-difusos.

Considerando os objetivos inicialmente identificados para o trabalho relatado nesta tese, as principais contribuições podem ser resumidas a seguir:

1. Desenvolvimento de um algoritmo que permite estabelecer um mapeamento do resultado do SOM para um sistema do tipo Takagi-Sugeno de Ordem Zero. Isto é possível pela aplicação da fórmula do potencial para descobrir o neurônio que melhor representa cada grupo, levando em consideração que o SOM no processo de ordenação cria no mapa uma

generalização da base de dados inteira, em que cada vetor das sinapses de um neurônio representa as médias das amostras dos neurônios vizinhos.

2. O desenvolvimento de um sistema neuro-difuso para diagnosticar faltas incipientes em transformadores de potência. O novo modelo apresentado nesta tese alcançou bons resultados, mostrando-se eficiente para a tarefa inicialmente definida.
3. O sistema neuro-difuso criado apresenta facilidade de interpretação dos resultados, resumizando os dados por meio de um algoritmo simples.
4. A principal contribuição pode ser vista na Tabela 6.1 (comparação com outros sistemas apresentados na literatura), com relação às taxas de sucesso alcançadas para o número de regras definidas.

TABELA 6.2 Resultados de Classificação – Comparação

Sistema Diagnóstico	% diagnóstico correto (base de treinamento)	% diagnóstico correto (base de teste)	Número de Regras	Tipos de Faltas
Castro e Miranda (2005a)	100	97.84	125	5
Castro e Miranda (2005b)	100	100	95	5
Naresh et al. (2008)	-	96.67	69	5
Trabalho proposto	100	100	5	5
IEC	89.31	81.3	5	5

6.2 Limitações e Trabalhos Futuros

Considerando a extração de regras a partir de RNAs, mostrou-se que há poucas abordagens apresentadas na literatura. As poucas apresentadas tinham limitações principalmente quanto à questão da interpretabilidade. Uma das limitações deste trabalho, no entanto, é a falta de generalidade do método de extração de regras, uma vez que a metodologia só pode ser empregada para mapas auto-organizáveis.

De acordo com Castro e Miranda (2005), um sistema difuso projetado a partir de dados não deve ser analisado apenas do ponto de vista de sua precisão, mas também da transparência. A transparência é uma medida de quão confiável ou válida é a interpretação linguística do sistema. Para ser transparente, um sistema difuso além de apresentar um número moderado de funções de pertinência, entre outros aspectos, deve permitir a distinção entre tais funções, ou seja, o grau de sobreposição deve ser menor que 75%. Porém, segundo Setnes (1998), as modificações feitas em um sistema difuso devem levar em conta o domínio,

para evitar *gaps*, e a imprecisão. Os sistemas apresentaram um pequeno número de regras com pontos de checagem fáceis de serem localizados, porém com alta sobreposição.

Como trabalhos futuros, sugerem-se:

1. Utilizar uma forma mais eficiente para delimitar as funções de pertinência. Os parâmetros das funções de pertinência podem ser otimizados, por exemplo, a partir da aplicação de algoritmos genéticos.
2. Realizar comparações dos sistemas de diagnósticos propostos com outros métodos convencionais, como o de Rogers e Duval, visto que se realizou apenas a comparação com o método do IEC 60599.
3. Outros trabalhos estão relacionados ao algoritmo do mapa de Kohonen com relação à distância empregada para cálculo da similaridade entre os pesos das sinapses dos neurônios e as amostras de treinamento e validação. Além disso, podem-se estabelecer novas maneiras de calcular o neurônio de maior potencial para cada grupo, por exemplo, substituindo a equação 4.1 pelo cálculo para descobrir o neurônio que apresenta os maiores graus de pertinência das amostras alocadas ao grupo.

Referências

- HAYCKIN, S. **Redes Neurais: princípios e prática**. 2ª ed. São Paulo: Bookman, 2007.
- EL-HAWARY, M. E. **Electric Power Applications of Fuzzy Systems**. New York: IEEE Press, 1998.
- Da SILVA, A. C. M.; GARCEZ, A. R. C.; MIRANDA, V. Transformer failure diagnosis by means of fuzzy rules extracted from Kohonen Self-Organizing Map. **International Journal of Electrical Power and Energy Systems**, v. 43, i. 1, p. 1034-1042, 2012. DOI: 10.1016/j.ijepes.2012.06.027.
- HUNG, C.; HUANG, J. J. Mining Rules from One-Dimensional Self-Organizing Map. In: INTERNATIONAL SYMPOSIUM ON INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS (INISTA), 2011. **Proceedings...** Istanbul, Turkey, 2011, p. 292-295 DOI: 10.1109/INISTA.2011.5946078.
- KAHRAMANLI, H.; ALLAHVERDI, N. Rule extraction from trained adaptive neural networks using artificial immune systems. **International Journal Expert Systems with Applications**, v. 36 , i. 2, p. 1513-1522, 2009. DOI: 1016/j.eswa.2007.11.024.
- HUYNH, T. Q.; REGGIA, J. A. Improving rule extraction from neural networks by modifying hidden layer representations. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2009. **Proceedings...** Atlanta, Georgia, USA, 2009. p. 734-739. DOI: 10.1109/IJCNN.2009.5178685.
- QUTEISHAT, A.; PENG LIM, C. A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. **Applied Soft Computing**, v. 8, i. 2, p. 985-995, 2008. DOI: 10.1016/j.asoc.2007.07.013.
- CASTRO, J. L.; FLORES-HIDALGO, L. D.; MANTAS, C. J.; PUCHE, J. M. Extraction of fuzzy rules from support vector machines. **Fuzzy Sets and Systems**, v. 158, i. 18, p. 2057-2077. DOI: 10.1016/j.fss.2007.04.014
- FUNG, G.; SANDILYA, S.; RAO, R. B. Rule extraction from linear support vector machines. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2005. **Proceedings...** Chicago, Illinois, USA, 2005. p. 32-40. DOI: 10.1145/1081870.1081878/
- BROUWER, R. K. Fuzzy rule extraction from a feed forward neural network by training a representative fuzzy neural network using gradient descent. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, v. 13, i.6, p. 673-698, 2005. DOI: 10.1142/S0218488505003746.
- CASTRO, A. R. G. **Knowledge Extraction from Artificial Neural Networks - an Application to Transformer Fault Diagnosis**. Portugal, 2004. Ph.D. Thesis, Ed. FEUP - Faculty of Engineering, University of Porto.
- KOHONEN, T. **Self-Organizing Maps**. 3ª ed., New York: Springer, 2001.

MALONE, J.; MCGARRY, K.; WERMTER, S.; BOWERMAN, C. Data Mining using Rule Extraction from Kohonen Self-Organizing Maps. **Neural Computing & Applications**, v. 15, i. 1, p. 9-17, 2006. DOI:10.1007/s00521-005-0002-1.

KANGAS, J. A.; KOHONEN, T. K.; LAAKSONEN, J. T. Variants of Self-Organizing Maps, **IEEE Transactions on Neural Networks**, v. 1, i. 1, 1990. DOI: 10.1109/72.80208.

VESANTO, J. SOM-Based Data Visualization Methods. **Intelligent Data Analysis**, v.1., i. 2, 1999. DOI: 10.1016/S1088-467X(99)00013-X.

ULTSCH, A.; SIEMON, H. P. Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. In: INTERNATIONAL NEURAL NETWORKS, 1990. **Proceedings...** [S.l.:S.n.], 1990. p. 305–308.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. **SOM Toolbox for Matlab 5**, Technical Report A57. Finland: Helsinki University of Technology, 2000.

KIANG, M. Y. Extending the Kohonen Self-Organizing Map Networks for Clustering Analysis, **Computational Statistics & Data Analysis**, v.38, i. 2, p. 161-180, 2001. DOI: 10.1016/S0167-9473(01)00040-8

WARD JR., J. H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, v. 58, i. 301, p. 236-244, 1963. DOI: 10.1080/01621459.1963.10500845

DUTRA, R. M. O.; COELHO, J. **O Método Ward de agrupamento de dados e sua aplicação em associação com Mapas Auto-Organizáveis de Kohonen**, Brasil: Laboratório de Planejamento de Sistemas de Energia Elétrica, Universidade Federal de Santa Catarina, 2008.

ZADEH, L. The Concept of a Linguistic Variable The Concept of a Linguistic Variable – I, **Information Sciences**, v. 8, i. 3, p. 199-249, 1975. DOI: 10.1016/0020-0255(75)90036-5

BARKAN, R. C.; TRUBATCH, S. L. **Fuzzy Systems Design Principles – Building Fuzzy IF-THEN Rule Bases**. Willey-IEEE Press, 1997.

ROSS, T. J. **Fuzzy Logic with Engineering Applications**, 2^a ed., England: John Wiley and Sons Ltd, 2004.

JOHANYÁK, Z. C; KOVÁCS, S. Fuzzy Rule Interpolation by the Least Squares Method. In: 7TH INTERNATIONAL SYMPOSIUM OF HUNGARIAN RESEARCHERS ON COMPUTATIONAL INTELLIGENCE, 2006. **Proceedings...**[S.l.:S.n.], 2006. p. 495-506.

JIN, Y. **Advanced Fuzzy Systems Design and Applications..** New York: Physica-Verlag, 2003.

RIID, A.; RUSTERN, E. Transparent Fuzzy Systems and Modeling with Transparency Protection, In: SYMPOSIUM ON ARTIFICIAL INTELLIGENCE IN REAL TIME CONTROL THREE CONTROL, 2000. **Proceedings...** [S.l.:S.n.], 2000. p. 229-235. Disponível em: <<http://www.dcc.ttu.ee/andri/teosed/transp.pdf>>.

SETNES, M.; BABUSKA, R.; KAYMAK, U.; van NAUTA LEMKE, H. R. Similarity measures in fuzzy rule base simplification. **IEEE Transactions on Systems, Man, and Cybernetics, Part B**, v. 28, i.3, p. 376-386, 1998. DOI: 10.1109/3477.678632.

TAKAGI, T.; SUGENO, M. Fuzzy Identification of Systems and its Applications to Modeling and Control. **IEEE Transaction on Systems, Man, and Cybernetics**, v. SMC-15, i. 1, p. 116-132, 1985. DOI: 10.1109/TSMC.1985.6313399.

GALLANT, S. I. Connectionist Expert Systems. **Communications of the ACM**, v. 31, i. 2, p. 152-169, 1988. DOI: 10.1145/42372.42377.

CRAVEN, M. **Extracting comprehensible models from trained neural networks**, Ed. Madison, University of Wisconsin, Ph.D. Thesis, 1996.

TICKLE, A. B.; ANDREWS, R.; GOLEA, M.; DIEDERICH, J. The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded Within Trained Artificial Neural Networks. **IEEE Transaction on Neural Networks**, v. 9, n. 6, 1998. DOI: 10.1109/72.728352.

CRAVEN, M.; SHAVLIK, J. **Rule Extraction: Where Do We Go from Here?** Department of Computer Sciences, University of Wisconsin, Machine Learning Research Group Working Paper 99-1, 1999.

NEUMANN, J. **Classification and Evaluation of Algorithm for Rule Extraction from Artificial Neural Networks**, PhD Summer Project, ICCS Division of Informatics, University of Edinburgh, August, 1998.

TICKLE, A.; ANDREWS, R.; GOLEA, M.; DIEDERICH, J. The truth is in there: directions and challenges in extracting rules from trained artificial neural networks. **IEEE Transactions on Neural Networks**, v. 9, p. 1058-1068, 1998. Disponível em: <<http://joachimdiederich.com/assets/IEEESurvey.pdf>>

HUNG, C., HUANG, L. Extracting Rules from Optimal Clusters of Self-Organizing Maps. In: SECOND INTERNATIONAL CONFERENCE ON COMPUTER MODELING AND SIMULATION, 2010. **Proceedings...** Sanya, Hainan, v. 1., 2010. DOI: 10.1109/ICCMS.2010.92.

van HEERDEN, W. S.; ENGELBRECHT, A. P. HybridSOM: A Generic Rule Extraction Framework for Self-Organizing Feature Maps. In: IEEE SYMPOSIUM ON COMPUTATIONAL INTELLIGENCE AND DATA MINING, 2009. **Proceedings...** Nashville, TN, [S.n], 2009. p. 17-24. DOI: 10.1109/CIDM.2009.4938624.

PATERITSAS, C.; MODES, S.; STAFYLOPATIS, A. Extracting Rules from Trained Self-Organizing Maps, In: INTERNATIONAL CONFERENCE APPLIED COMPUTING, 2007. **Proceedings...** [S.l.:S.n], 2007.

DARRAH, M.; TAYLOR, B.; SKIAS, S. Rule Extraction from Dynamic Cell Structure Neural Networks Used in a Safety Critical Application. In: 17TH INTERNATIONAL FLAIRS CONFERENCE, 2004. **Proceedings...** Miami, Florida, USA, 2004. p. 629-634.

HANI, M. K.; NOR, S. M.; HUSSEIN, S.; ELFADIL, N. Machine Learning: the automation of knowledge acquisition using Kohonen Self-Organizing Map Neural Network. **Malaysian**

Journal of Computer Science, v. 14, n. 1, p. 68-82, 2001. Disponível em: <http://e-journal.um.edu.my/filebank/published_article/1781/115.pdf>

NARESH, R.; SHARMA, V.; VASHISTH, M. An Integrated Neural Fuzzy Approach for Fault Diagnosis of Transformers. **IEEE Transactions on Power Delivery**, v. 23, i. 4, 2008. DOI: 10.1109/TPWRD.2008.2002652.

CHIU, S. L. A Cluster Estimation Method with Extension to Fuzzy Model Identification. In: THIRD IEEE CONFERENCE ON FUZZY SYSTEMS, IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE, 1994. **Proceedings...** Orlando, Florida, USA, v. 2, 1994. p.1240-1245. DOI: 10.1109/FUZZY.1994.343644.

WONG, K. W.; GEDEON, T. D.; FUNG, C. C.; WONG, P. M. Fuzzy Rules Extraction Using Self-Organizing Neural Network and Association Rules. In: IEEE REGION 10 INTERNATIONAL CONFERENCE ON ELECTRICAL AND ELECTRONIC TECHNOLOGY, 2001. **Proceedings...** [S.l.], v.1, 2001. p. 403-408. DOI: 10.1109/TENCON.2001.949624.

WANG, L.-X. **A Course in Fuzzy Systems and Control**, Printice-Hall International, p. 192-193, 1997.

ANSI/IEEE Std C57.104.1991, IEEE Guide of gases generated in oil-immersed Transformer, IEEE Power Engineering Society, 1992.

SALAMI, A.; PAHLEVANI, P. Neural Network Approach for Fault Diagnosis of Transformer. In: INTERNATIONAL CONFERENCE ON CONDITIONING MONITORING AND DIAGNOSIS, 2008. **Proceedings...** Beijing, China, 2008. p. 21-24. DOI: 10.1109/CMD.2008.4580518.

YANG-JING, S.; SHEN, Z.; CHANG-XIN, M.; JING-MENG, L. Improved BP Neural Network for Transformer Fault Diagnosis. **Journal of China University of Mining and Technology**, v. 17, i. 1, p. 138-142, 2007. DOI: 10.1016/S1006-1266(07)60029-7.

THANG, K. F.; AGGARWAL, R. K. Analysis of Power Transformer Dissolved Gas Data Using the Self-Organizing Map. **IEEE Transactions on Power Delivery**. v. 18, i. 4, p. 1241-1248, 2003. DOI:10.1109/TPWRD.2003.817733.

GOLKHAH, M.; SHAMSHIRGAR, S. S.; VAHIDI, M. A. Artificial neural networks applied to DGA for fault diagnosis in oil-filled power transformers. **Journal of Electrical and Electronics Engineering Research**, v. 3, i.1, p. 1-10, 2011. Disponível em: <<http://www.academicjournals.org/jeeer/pdf/Pdf2011/Jan/Golkhah%20et%20al.pdf>>

HU, Q.; CHEN, W.; DU, L.; LI, N.; SUN, C. X. The study of variant DGA feature neural network multilayer diagnostic model. In: 7TH WORLD CONGRESS ON INTELLIGENT CONTROL AND AUTOMATION, 2008. **Proceedings...** Chongqing, 2008. p. 8526-8530.

DINIZ NETO, A.; ASSUNÇÃO, T. C. B. N.; ASSUNÇÃO, J. T. **Classificação de transformadores de potência baseado na análise dos gases dissolvidos no óleo isolante**, 2003. Disponível em: <http://www.bibl.ita.br/xivencita/COMP12.pdf>

DÖRNENBURG, E.; STRITTMATTER, W. Monitoring oil-cooled transformers by gas analysis, **Brown Boveri Review**, v. 61, i. 5, p. 238–247, 1974.

ROGERS, R. R. IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis. **IEEE Transaction Dielectrics Electrical Insulation Society**, v. E1-13, i. 5, p. 349–354, 1978.

DUVAL, M. Fault gases formed in oil-filled breathing E.H.V. power transformers - the interpretation of gas analysis data. In: IEEE Power Engineering Society, 1974. **Proceedings...** [S.l.:S.n], 1974. p. 476-478.

IEC Std. 60599. **Mineral Oil-Impregnated Electrical Equipment in Service – Guide to the Interpretation of Dissolved and Free Gases Analysis**, 2 ed., 1999.

PUGH, D. R. Advances in Fault Diagnosis by Combustible Gas Analysis. In: MINUTES OF FORTY-FIRST INTERNATIONAL CONFERENCE OF DOUBLE CLIENTS, 1974. **Proceedings...** [S.l.:S.n], Section 10-1201, 1974.

SU, H. S.; LI, Q. Z. Transformer insulation fault diagnosis method based on fuzzy expert system. In: 8th INTERNATIONAL CONFERENCE ON PROPERTIES AND APPLICATIONS OF DIELECTRIC MATERIALS, 2006. **Proceedings...** Bali, 2006. p. 343-346. DOI: 10.1109/ICPADM.2006.284186.

NÉMETH, B.; LABONCZ, S.; KISS, I. Condition Monitoring of Power Transformers using DGA and Fuzzy Logic. In: 2009 IEEE ELECTRICAL INSULATION CONFERENCE, 2009. **Proceedings...** Montreal, QC, Canada, 2009. p. 373-376. DOI: 10.1109/EIC.2009.5166373.

MORAIS, D. R.; da SILVA, J. R.; ROLIM, J. G. A fuzzy system for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil, diagnostics for electric machines, power electronics and drives. In: 5TH IEEE INTERNATIONAL SYMPOSIUM, 2005. **Proceedings...** Viena, Austria, 2005.

HMOOD, S.; ABU-SIADA, A.; MASOUM, M. A. S.; ISLAM, S. M. Standardization of DGA Interpretation Techniques using Fuzzy Logic Approach. In: IEEE INTERNATIONAL CONFERENCE ON CONDITION MONITORING AND DIAGNOSIS, 2012. **Proceedings...** Bali, Indonesia, 2012. p. 929-932. DOI: 10.1109/CMD.2012.6416305

CASTRO, A. R. G.; MIRANDA, V. Knowledge Extraction in Neural Networks with Application to Transformer Fault Diagnosis, **IEEE Transactions on Power Systems**, v. 20, i. 2, 2005a. DOI: 10.1109/TPWRS.2005.846074.

CASTRO, A. R. G.; MIRANDA, V. Improving the IEC Table for Transformer Failure Diagnosis with Knowledge Extraction from Neural Networks, **IEEE Transactions Power Delivery**, v. 20, i. 4, 2005b. DOI: 10.1109/TPWRD.2005.855423.