



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**MÁRCIO ANDRÉ BAIMA AMORA**

**EXTRAÇÃO DE REGRAS INTERPRETÁVEIS PARA O  
DIAGNÓSTICO EFICIENTE DE TRANSFORMADORES DE  
POTÊNCIA ISOLADOS A ÓLEO A PARTIR DO AUMENTO DO  
ESPAÇO DE ATRIBUTOS**

**FORTALEZA**

**2013**

**MÁRCIO ANDRÉ BAIMA AMORA**

**EXTRAÇÃO DE REGRAS INTERPRETÁVEIS PARA O  
DIAGNÓSTICO EFICIENTE DE TRANSFORMADORES DE  
POTÊNCIA ISOLADOS A ÓLEO A PARTIR DO AUMENTO DO  
ESPAÇO DE ATRIBUTOS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Ceará (PPGEE/UFC) como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

**Orientador:** Dr. Arthur Plínio de Souza Braga

**Co-Orientador:** Dr. Otacílio da Mota Almeida

**FORTALEZA**

**2013**

*Oh Maria concebida sem pecado, rogai por nós  
que recorremos a Vós*

*Para a minha família*

*"Mantenha-se interessado na própria carreira, mesmo que humilde, é o que realmente você possui na sorte incerta dos tempos..."*

*"Acima da benéfica disciplina, seja bondoso consigo mesmo. Você é filho do universo não menos que as árvores e as estrelas. Você tem o direito de estar aqui. E esteja claro ou não para você, sem dúvida, o Universo se desenrola como deveria. Portanto, esteja em paz com Deus, qualquer que seja sua forma de concebê-lo. E sejam quais forem sua vida e suas aspirações, na barulhenta confusão da vida, mantenha-se em paz com sua alma..."*

(trechos extraídos do *Desideratu*, 1692)

## AGRADECIMENTOS

- A Deus, por me acompanhar e me ajudar em todos os momentos da minha vida.
- A minha esposa Anna Carolina e aos meus filhos Pedro André e Lucas André pelo amor, incentivo e compreensão.
- A meus pais, Amarílio e Virgínia, pelo amor e apoio sempre demonstrados, e por serem sempre uma lição de vida para mim.
- A meus irmãos Paulo e Juliana pelo companheirismo e amizade.
- Ao meu orientador Prof. Dr. Arthur Plínio de Souza Braga e co-orientador Prof. Dr. Otacílio da Mota Almeida pela dedicação, incentivo e acima de tudo amizade.
- Aos professores do Programa de Pós-Graduação em Engenharia Elétrica pelos ensinamentos e lições de vida.
- Aos meus colegas do Programa de Pós-Graduação em Engenharia Elétrica pela amizade e ajuda sempre demonstrados.
- Aos funcionários e técnicos do Departamento de Engenharia Elétrica pela disponibilidade e ajuda em todos os momentos.

## RESUMO

Este trabalho propõe um novo método de diagnóstico de falhas incipientes em equipamentos de potência com isolamento a óleo baseado em regras interpretáveis extraídas a partir das concentrações dos gases presentes e dissolvidos no óleo isolante e de relações entre essas concentrações não exploradas por outros métodos. Utilizando uma árvore de decisão (*Decision Tree* - DT) com modificações na dimensão dos atributos de entrada para a obtenção de regras, o método proposto: (i) permite que o usuário conheça as regras que geraram o diagnóstico, (ii) contorna limitações comuns a outros métodos (ex.: existência de regiões de “não-diagnóstico”, existência de mais de um diagnóstico possível) e (iii) apresenta percentual elevado de acerto no diagnóstico de dados de DGA (*Dissolved Gas Analysis*) disponíveis na literatura. O método proposto é comparado com outros métodos de DGA descritos em normas: Gás Chave, Rogers, Dörnenburg e Triângulo de Duval. Na comparação dos métodos foram utilizados bancos de dados com informações publicadas, entre eles os dados do IEC TC 10 que apresentam informações e diagnóstico de um comitê de especialistas. Os índices de acerto observados com o método proposto estiveram acima de 94%. Comparativamente, o Método do Triângulo de Duval apresenta índices de acerto entre 83% e 88%. O método também foi avaliado com relação à sensibilidade na ocorrência de erros nas medidas de laboratório, apresentando resultados comparativamente superiores (índice de acerto de 83% em comparação com 78% do Método do Triângulo de Duval). Também foi desenvolvido neste trabalho um método para o diagnóstico simplificado de equipamentos de potência com isolamento a óleo utilizando dados físico-químicos – os índices de acerto obtidos foram inferiores aos alcançados com o DGA proposto, porém próximos aos observados em métodos de DGA citados na literatura, como Rogers IEEE C57.104 – 2008, com a vantagem dos dados não dependerem da disponibilidade de um cromatógrafo.

**Palavras-chave:** Óleo isolante, Dados Físico-Químicos, Diagnóstico de falhas, Transformadores de potência, Árvore de Decisão.

## ABSTRACT

This work proposes a new method for diagnosis of incipient faults in power equipment with insulation oil. It consists on a set of interpretable rules extracted from the concentrations of the dissolved gases in insulating oil, and the relationships between these concentrations. Using a Decision Tree (DT) with changes in the size of the input attributes for obtaining rules, the proposed method: (i) allows the user to know the rules that generated the diagnosis, (ii) bypasses limitations common to other methods (e.g. the existence of "no diagnosis" regions, and the occurrence of more than one possible diagnosis) and (iii) the accuracy in diagnosis of DGA (Dissolved Gas Analysis) data available in the literature. The proposed method is compared with other DGA standard methods: Gas Key, Rogers, Dornenburg and Duval's Triangle. The performed comparison considered databases available in technical literature, including data from the IEC TC 10. The observed accuracy of the proposed method is above 94%. Comparatively, the Duval's Triangle method presents accuracy between 83% and 88% for the considered databases. The method was also evaluated with respect to sensitivity to measurements errors of laboratory – the results are very promising, the proposed method shows a hit rate of 83% in comparison with a hit rate of 78% for the Duval's Triangle method. This work has also been developed a method for simplified diagnostic of power equipments with insulated oil using physico-chemical data - the hit rates obtained were lower than those achieved with the proposed DGA, however close to those observed in DGA methods cited in the literature, as Rogers IEEE C57.104 - 2008, with the advantage of the data not depend on the availability of a gas chromatograph.

**Keywords:** Insulation Oil, Physicochemical Data, Fault Diagnosis, Power Transformers, Decision Tree.

# SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. Motivação .....	3
1.2. Desenvolvimento da Pesquisa .....	3
1.3. Contribuições da Tese .....	5
1.4. Organização do Documento .....	6
1.5. Produção Científica .....	6
2. DIAGNÓSTICO DE FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA .....	9
2.1. Falhas em Transformadores .....	9
2.1.1. Falhas Térmicas .....	10
2.1.2. Falhas Elétricas.....	11
2.2. Formação de Gases Dissolvidos no Óleo .....	12
2.3. Métodos Indicados nas Normas para Diagnóstico Baseado na Concentração de Gases no Óleo Isolante.....	13
2.3.1. Método do Gás Chave .....	14
2.3.2. Método das Razões de Dörnenburg.....	16
2.3.3. Método das Razões de Rogers.....	18
2.3.4. Triângulo de Duval.....	19
2.4. Confiabilidade das Medições dos Gases Dissolvidos no Óleo Isolante .....	20
2.5. Limitações das Metodologias Tradicionais para o Diagnóstico de Transformadores através de DGA .....	22
2.6. Utilização de Técnicas de Aprendizagem de Máquina para o Diagnóstico de Transformadores através de DGA .....	25
2.6.1. Utilizando uma Rede Neural Artificial (RNA).....	26
2.6.2. Adotando um Sistema de Regras Fuzzy (SRF) .....	27
2.6.3. Dificuldades na Utilização de uma RNA e/ou SRF para o Diagnóstico de Transformadores.....	29
2.7. Diagnóstico do Estado do Óleo Isolante.....	30
2.8. Análise das Propriedades Físico-Químicas e Diagnóstico de Falhas em Transformadores .....	32
2.9. Considerações Parciais .....	36

3. ÁRVORES DE DECISÃO .....	38
3.1. Definições.....	38
3.2. Introdução à Árvore de Decisão .....	40
3.3. Etapas de Construção de uma Árvore de Decisão.....	42
3.4. Estrutura Básica de um Algoritmo de Construção de uma Árvore de Decisão.....	43
3.5. Método C4.5 para Construção de uma Árvore de Decisão .....	46
3.6. Árvores de Decisão Oblíquas .....	47
3.7. Vantagens e Desvantagens no Uso de uma Árvore de Decisão .....	49
3.8. Considerações Parciais .....	51
4. CLASSIFICAÇÃO UTILIZANDO ÁRVORE DE DECISÃO COM AUMENTO DA DIMENSÃO DO ESPAÇO DOS ATRIBUTOS DE ENTRADA .....	52
4.1. Limitações da Utilização de uma Árvore de Decisão em Problemas de Classificação e Possibilidades de Correções .....	53
4.2. Teoremas de Cover.....	54
4.2.1. Função de Contagem .....	55
4.2.2. Separação por Superfícies Arbitrárias .....	57
4.2.3. Separação de Padrões Randômicos .....	61
4.3. Metodologia Proposta.....	62
4.4. Exemplos de Utilização.....	64
4.5. Considerações Parciais .....	69
5. DIAGNÓSTICO APRIMORADO DE FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA.....	70
5.1. Dados de Concentração de Gases Utilizados .....	70
5.2. Levantamento das Regras do Método DGA Proposto.....	72
5.3. Avaliação Comparativa das Regras do Método DGA Proposto.....	75
5.4. Sensibilidade a Erros de Medida do Método DGA Proposto.....	79
5.5. Método Simplificado de Diagnóstico de Transformadores Utilizando Dados Físico-Químicos.....	80
5.6. Considerações Parciais .....	83
6. CONCLUSÕES E PERSPECTIVAS .....	85
6.1. Trabalhos Futuros .....	87
REFERÊNCIAS .....	88
APÊNDICES .....	103
APÊNDICE A – Métodos de Extração de Regras .....	103

APÊNDICE B – Abertura de Nós, Poda e Indução de uma Árvore de Decisão .....	105
B.1. Critérios de Abertura de Nós .....	105
B.1.1. Critérios Baseados em Impureza .....	105
B.1.2. Critérios Baseados na Impureza Normalizada.....	107
B.1.3. Outros Métodos de Divisão Univariantes .....	107
B.1.4. Comparação dos Métodos Univariantes .....	107
B.1.5. Critérios Multivariáveis de Construção de Nós .....	107
B.2. Métodos de Poda.....	108
B.2.1. Outros Métodos.....	109
B.2.2. Comparação dos Métodos de Poda.....	109
B.3. Métodos de Indução de uma Árvore de Decisão .....	111
B.3.1. ID3 .....	111
B.3.2. CART .....	111
B.3.3. CHAID.....	112
B.3.4. QUEST.....	112
B.3.5. C4.5.....	112

## LISTA DE FIGURAS

Figura 1- Separação de dados: a) separação complexa, b) separação através de uma reta .....	2
Figura 2- Formação de gases: a) pressão x temperatura (IEEE, 2008), b) concentração dos gases x temperatura .....	13
Figura 3- Composição dos gases para a falha térmica no óleo.....	14
Figura 4- Composição dos gases para a falha térmica na celulose.....	15
Figura 5 - Composição dos gases para descarga parcial .....	15
Figura 6 - Composição dos gases para descarga de alta energia .....	16
Figura 7 - Método do Triângulo de Duval.....	20
Figura 8 - Espectroscopia dielétrica de óleo isolante envelhecido.....	33
Figura 9 - Variação da condutividade do óleo em função da temperatura.....	34
Figura 10 - Exemplos de separação não linear (a) e linear (b) do conjunto de dados.....	40
Figura 11 - Exemplo de um plano a partir de um sistema em três dimensões .....	40
Figura 12 - Partição do espaço de variáveis e regras obtidas de uma árvore de classificação binária .....	42
Figura 13 - Algoritmo para construção de uma DT.....	44
Figura 14 - Árvore de decisão univariável .....	48
Figura 15 - Árvore de decisão oblíqua .....	48
Figura 16 - Hiperplano ótimo para padrões não linearmente separáveis e aumento de dimensão.....	54
Figura 17 - Hiperplano que separa padrões das classes $Z^+$ (círculos brancos) e $Z^-$ (círculos pretos).....	56
Figura 18 - Exemplos de superfícies $\phi$ que separam dicotomias de cinco pontos em duas classes: (a) linearmente separáveis, (b) esférico separáveis, (c) quádrlica separáveis .....	61
Figura 19 -Saída de um dispositivo XOR.....	65
Figura 20 - Solução do XOR .....	67
Figura 21 - Classificação da Íris em 2D .....	68
Figura 22 - Classificação da Íris em 3D .....	68
Figura 23 - Distribuição normal .....	79

## LISTA DE TABELAS

Tabela 1 - Razões entre as Concentrações dos Gases .....	16
Tabela 2 - Concentrações Limite dos Gases Conforme (IEEE, 2008) .....	17
Tabela 3 - Concentrações Limite dos Gases Conforme (IEEE, 1991) .....	17
Tabela 4 - Concentrações Limite dos Gases Conforme Dörnenburg .....	17
Tabela 5 - Valores para as Razões de Gases para o Método de Dörnenburg .....	18
Tabela 6 - Valores das Razões de Gases para o Método de Rogers .....	19
Tabela 7 - Desvios para as Medidas de Repetição dos Laboratórios Avaliados .....	21
Tabela 8 - Desvios para as Medidas de Precisão dos Laboratórios Avaliados.....	22
Tabela 9 - Resultados de Aplicação dos Métodos para o IEC TC 10 .....	24
Tabela 10 - Taxa de acerto quando da possibilidade de erro nas medidas (IEC TC 10).....	25
Tabela 11 - Propriedades físico-químicas do óleo mineral e normas.....	31
Tabela 12 – Recomendações de verificação das condições do óleo isolante .....	32
Tabela 13 - Resultados dos ensaios físico-químicos comparados aos valores normatizados ..	35
Tabela 14- Entradas e saída para um XOR.....	65
Tabela 15 - Problema do XOR com atributos expandidos .....	66
Tabela 16 - Matriz de confusão, dados de treino e entradas típicas .....	73
Tabela 17 - Matriz de confusão, dados de teste e entradas típicas .....	73
Tabela 18 - Erro de classificação, considerando os casos para validação cruzada .....	74
Tabela 19 - Regras extraídas .....	75
Tabela 20 - Erro percentual para os conjuntos de testes.....	75
Tabela 21 - Percentagens de acerto para diversos conjuntos de dados .....	76
Tabela 22 - Percentagens de acerto, comparando diversos métodos e utilizando IET TC 10..	76
Tabela 23 - Matriz de confusão utilizando Dörnenburg (ZIRBES, ROLIM e ZÜRN, 2005)..	77
Tabela 24 - Matriz de confusão utilizando Dörnenburg IEEE C57.104 - 1991 .....	77
Tabela 25 - Matriz de confusão utilizando Dörnenburg IEEE C57.104 - 2008 .....	77
Tabela 26 - Matriz de confusão utilizando Rogers IEEE C57.104 - 2008 .....	78
Tabela 27 - Matriz de confusão utilizando o método do Triângulo .....	78
Tabela 28 - Matriz de confusão utilizando as regras propostas.....	78
Tabela 29 – Taxa de acerto quando da possibilidade de erro nas medidas (IEC TC 10).....	80
Tabela 30 - Percentual de erro, considerando os casos para validação cruzada (dados físico-químicos) .....	81
Tabela 31 - Regras obtidas utilizando dados físico-químicos .....	82
Tabela 32 - Erro percentual para os conjuntos de testes (dados físico-químicos).....	82
Tabela 33 - Taxa de acerto quando da possibilidade de erro nas medidas físico-químicas .....	83

## LISTA DE SÍMBOLOS

$\tan \delta$	Fator de perdas de um transformador
$a_i$	Atributo avaliado
$\theta$	Operação lógica testada ( $=, \neq, \leq, \geq, <, >$ )
$\alpha$	Valor limite
$c1$ e $c2$	"Caminhos" distintos na árvore de decisão
$t$	Nó avaliado
$X$	Conjunto de treinamento para uma árvore de decisão
$X_t$	Subconjunto específico do conjunto de treinamento $X$
$X_{tY}, X_{tN}$	Subconjuntos disjuntos descendentes de $X_t$
$A$	Atributos do conjunto de entrada
$Y$	Valor ou classe esperado do atributo
$E$	Entropia
$G$	Ganho de informação
$N_t$	Número de pontos em $t$
$t_S, t_N$	Nós que descendem do nó $t$
$N_{tS}$	Número de pontos no nó descendente $t_S$
$N_{tN}$	Número de pontos no nó descendente $t_N$
$d$	Dimensão do espaço euclidiano
$\omega_0$	Valor que intercepta o eixo do atributo $a_j$
$\omega_m$	Direção do hiperplano divisor ( $d$ -dimensional)
$\omega_{m0}$	Valor limitante quando $a_j$ está projetado sobre $\omega_m$
$E^d$	Padrões representados por um conjunto de vetores em um espaço euclidiano de dimensão $d$
$z$	Vetor de dados no espaço $d$
$w$	Vetor de pesos no espaço $d$
$N$	Número de pontos
$Z$	Conjunto de vetores $z$ no espaço $d$
$Z^+$ ou $Z$	Classes de $Z$ divididas por um hiperplano
$\phi$	Função de medida
$P$	Probabilidade
$C$	Número de dicotomias linearmente separáveis
$m$	Nova dimensão, sendo $m > d$
$h'$	Vetor com dados de gases com a presença de erros
$H$	Vetor original sem erros dos valores de gases
$r$	Vetor de ruídos obtido de uma distribuição normal

## LISTA DE ABREVIATURAS E SIGLAS

DGA	<i>Dissolved Gas Analysis</i> – Análise de gases dissolvidos
ML	<i>Machine Learning</i> - Aprendizagem de Máquina
DT	<i>Decision Tree</i> – Árvore de decisão
AG	Algoritmo Genético
RNA	Rede Neural Artificial
SRF	Sistema de Regras <i>Fuzzy</i>
MVS	Máquina de Vetor de Suporte
DP	Descargas Parciais
D1	Descargas de baixa energia
D2	Descargas de alta energia
T1	Falhas térmicas de temperatura $< 300\text{ }^{\circ}\text{C}$
T2	Falhas térmicas de temperatura $300\text{ }^{\circ}\text{C} < T < 700\text{ }^{\circ}\text{C}$
T3	Falhas térmicas de temperatura $> 700\text{ }^{\circ}\text{C}$
OLTC	<i>On Load Tap Changer</i>
ppm	Partes por milhão
IEC	<i>International Electrotechnical Commission</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
ABNT	Associação Brasileira de Normas Técnicas
ASTM	<i>American Society for Testing and Materials</i>
MLP	<i>Multilayer Perceptron</i>
CIGRE	<i>Council on Large Electric Systems</i>
AC	Árvore de Classificação
AR	Árvore de Regressão
ACB	Árvore de Classificação Binária
Ac	Acidez
RD	Rigidez Dielétrica
TA	Teor de Água
TI	Tensão Interfacial
D	Densidade
FP100	Fator de Perdas dielétricas a $100^{\circ}\text{C}$

# 1. INTRODUÇÃO

Transformadores de potência estão entre os equipamentos mais caros e críticos de uma planta elétrica de transmissão e distribuição. Dessa forma, um grande esforço deve ser empregado para garantir o bom funcionamento desses equipamentos no que concerne a evitar possíveis falhas que venham a ocorrer devido ao seu ciclo natural de vida ou regimes elétricos a que são submetidos.

As principais falhas que podem ocorrer em um transformador têm suas origens no envelhecimento natural do equipamento ou no resultado da incidência de sobrecarga e/ou curtos-circuitos no regime de trabalho submetido (BARBOSA, 2013).

Os métodos de diagnóstico de falhas em equipamentos com isolamento a óleo, como os transformadores, baseados na análise de gases dissolvidos no óleo isolante (DGA - *Dissolved Gas Analysis*) são amplamente utilizados. Esses métodos consideram a análise de tipo, concentração e taxa de produção de gases gerados e dissolvidos no óleo do transformador, e procuram associar o tipo de falha ao gás presente. Por exemplo, (WANG, 2000) cita que descargas elétricas levam à geração de acetileno enquanto a presença de dióxido de carbono está associada ao sobreaquecimento da celulose.

Para o diagnóstico adequado de equipamentos com óleo isolante, por meio da DGA, o especialista normalmente emprega um ou mais métodos padronizados, que são descritos nas normas (IEEE C57.104, 2008; IEC 60599, 2008). No entanto, tais métodos apresentam sérias limitações quanto a precisão ou mesmo inexistência de diagnóstico conforme indicam os estudos de (WANG, 2000; LIMA, 2005). Nessas situações, o conhecimento de um perito será inestimável para atingir o diagnóstico correto.

Muitos estudos (AMORA *et al.*, 2009; BARBOSA *et al.*, 2009; SPURGEON *et al.*, 2005; MIRANDA e CASTRO, 2005; WANG, 2003; ESP e McGRIL, 1999) propõem a utilização de técnicas de Aprendizagem de Máquinas (*Machine Learning* - ML)<sup>1</sup>, associadas com extração de regras de conhecimento, para obtenção de um diagnóstico mais eficiente de equipamentos com isolamento a óleo, como os transformadores. Entretanto, várias das

---

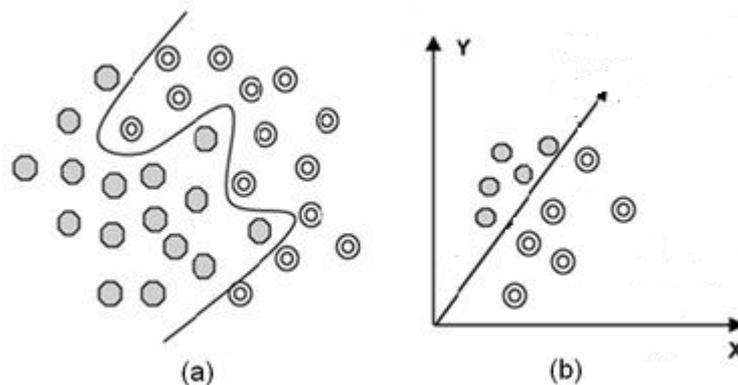
<sup>1</sup> Técnicas de aprendizagem de máquinas abrangem: AG – Algoritmo Genético (MITCHELL, 1999), RNA – Rede Neural Artificial (HAYKIN, 2001), SRF – Sistema de Regras Fuzzy (WOLKENHAUER, 2001), MVS - Máquina de Vetor de Suporte (CRISTIANI e SHAW-TAYLOR, 2000); entre outras.

técnicas empregadas falham na obtenção de tais regras, em questões como precisão e/ou capacidade de interpretação de tais regras obtidas.

A capacidade de interpretação de uma regra está relacionada com a facilidade de um técnico de manutenção de transformadores em interpretar as informações de gases dissolvidos no óleo isolante de maneira intuitiva e prever as consequências na integridade do equipamento. Para os fins deste trabalho, considera-se como um conjunto de regras simples (interpretável) aquela que apresenta um pequeno número de regras que utilizem operadores tradicionais (E ou OU) e com saídas (diagnósticos) definidas.

Matematicamente, uma regra obtida apresentará uma maior simplicidade de representação e, portanto, uma maior capacidade de interpretação intuitiva quando a separação puder ser representada por uma reta, como no caso *b* da Figura 1. Entretanto, muitos métodos de classificação associados a certos conjuntos de dados constroem superfícies complexas de separação, que quando representadas matematicamente não permitem uma representação simples e compacta, dificultando também a interpretação da separação obtida, caso *a* da Figura 1.

**Figura 1- Separação de dados: a) separação não-linear, b) separação linear**



Fonte: (RAMACHANDRA *et al.*, 2012).

Esse trabalho propõe um novo método de diagnóstico de falhas incipientes em equipamentos de potência com isolamento a óleo, como os transformadores de potência. O novo método utiliza uma árvore de decisão (*Decision Tree* - DT) (ROKACH e MAIMON, 2008; RUSSELL e NORVIG, 2004; DUDA, HART e STORK, 2001; QUINLAN, 1993; e

QUINLAN, 1986) com modificações propostas na dimensão dos atributos de entrada, permitindo obter um conjunto de regras com uma precisão maior no diagnóstico que os métodos descritos em normas, e de fácil utilização.

Esse capítulo introdutório é dividido nas seguintes seções: a motivação da Tese é indicada na seção 1.1, a metodologia adotada no desenvolvimento do trabalho é explicitada na seção 1.2, os artigos científicos já publicados em congressos e periódicos relevantes e relacionados com a pesquisa realizada são comentados na seção 1.3, as principais contribuições obtidas dos estudos realizados são citadas na seção 1.4, e a disposição geral do documento é apresentada na seção 1.5.

## **1.1. Motivação**

Os métodos tradicionais de diagnóstico de transformadores de potência por meio de DGA apresentam várias limitações: (i) existência de situações que não são consideradas pelos critérios impostos pelos métodos; (ii) possibilidade da existência de mais de um diagnóstico, gerando dificuldades em chegar a conclusões; ou (iii) resultados errôneos. Uma alternativa para superar as dificuldades dos métodos de diagnóstico de transformadores é a utilização de técnicas de ML. No entanto, como também já ressaltado, nem sempre esses métodos baseados em ML permitem a obtenção de regras de diagnóstico que sejam suficientemente precisas e de fácil entendimento para os técnicos de operação e manutenção.

As motivações principais desse trabalho são buscar soluções para os problemas citados com aplicações em equipamentos com isolamento a óleo, como os transformadores, conduzindo a um novo método de diagnóstico, mais eficiente que os métodos indicados nas normas, utilizando uma técnica de ML com modificações que permitam suplantar as dificuldades normais observadas no uso de técnicas de extração de conhecimento, de modo que sejam obtidas regras simples e fáceis de usar para o diagnóstico de transformadores. O novo método de diagnóstico desenvolvido será proposto como uma possibilidade de melhoria nas normas vigentes.

## **1.2. Desenvolvimento da Pesquisa**

Essa pesquisa tem como objetivo principal a proposição de um novo método de diagnóstico de falhas incipientes em transformadores de potência. Para a obtenção desse

método, diversas etapas foram seguidas no desenvolvimento da Tese. A seguir essas etapas serão comentadas.

Inicialmente foram realizados estudos sobre os métodos Gás Chave, Rogers, Dörnenburg e Triângulo de Duval, indicados nas normas técnicas (IEEE C57.104, 2008; IEC 60599, 2008) para o diagnóstico de equipamentos com isolamento a óleo, considerando as vantagens e desvantagens desses métodos. Também, foram pesquisados métodos mais recentes de diagnóstico que utilizam técnicas de ML para a análise de tais equipamentos. Entretanto, esses métodos também apresentam limitações, normamente associadas à precisão e grau de interpretação das regras de conhecimento extraídas dos resultados obtidos.

Numa fase posterior, devido às dificuldades observadas com a utilização de técnicas de ML para o diagnóstico de falhas incipientes em transformadores e outros equipamentos com isolamento a óleo, foram estudadas técnicas para a extração de regras de conhecimento a partir de tais métodos. Ao final dessa fase, foi escolhido como método a ser utilizado nesse trabalho uma DT, entretanto esse método pode apresentar problemas de precisão ou obtenção de um número elevado de regras devido a sua característica de construir intervalos de separação paralelos aos eixos da dimensão do problema. Entretanto, a escolha de tal método é devido às suas importantes vantagens, como: capacidade de seleção de requisitos e obtenção direta de regras de conhecimento com operadores convencionais.

De forma a superar as eventuais dificuldades no uso de uma DT, é proposta nessa pesquisa uma nova forma de construir as entradas do problema de diagnóstico de equipamentos com isolamento a óleo. Essa técnica também foi testada com outros problemas, mostrando ser promissora como forma alternativa de representação de um problema a ser solucionado por uma DT. A ideia é baseada no Teorema de Cover (COVER, 1965), onde inicialmente tenta-se apresentar a maior quantidade possível de atributos que podem ser significativos para a solução de um problema de classificação para uma DT, depois durante a construção da árvore, a própria DT, devido a sua capacidade intrínseca de seleção, determina-se quais atributos são mais importantes, com isso é possível obter regras com maior precisão, maior capacidade de generalização e em um número menor, quando comparado com uso tradicional de uma DT.

Com o método desenvolvido, foi aplicado o mesmo para o diagnóstico de transformadores de potência imersos em óleo, e os resultados comparados com os métodos indicados nas normas técnicas. Como resultado, os índices de acertos foram significativamente melhores dos que os apresentados tais métodos clássicos, também o número de regras obtidas foi pequena. Houve, ainda, o cuidado de ser realizada uma validação cruzada de forma a

garantir uma maior consistência estatística na utilização dos dados adotados para desenvolver e testar as novas regras de diagnóstico propostas.

Também foi realizado um estudo do impacto nos métodos de diagnóstico da possibilidade de ocorrência de erros nas medidas de DGA. Sendo avaliados e comparados os resultados obtidos com as regras propostas nesse trabalho e com os métodos clássicos (Gás Chave, Razões de Dörnenburg, Razões de Rogers e Triângulo de Duval). As regras propostas apresentam novamente o maior índice de acerto.

### **1.3. Contribuições da Tese**

Como principais contribuições frutos dos estudos desenvolvidos e apresentados nesta Tese podem ser citadas:

- Desenvolvimento de um novo método de diagnóstico, a partir de medidas cromatográficas, de equipamentos de potência com isolamento a óleo, como os transformadores de potência, baseado em regras simples, extraídas de uma DT, e com elevada precisão.
- Proposta de uma nova metodologia de utilização de uma DT, baseado no aumento inicial da dimensão do problema pelo acréscimo de atributos e de relações dos atributos originais. Metodologia utilizada neste trabalho para desenvolver as novas regras propostas de diagnóstico de transformadores.
- Desenvolvimento de um método simplificado de diagnóstico de equipamentos com isolamento a óleo baseado em medidas físico-químicas do equipamento, com resultados superiores a vários métodos indicados nas normas e utilizando dados físico-químicos que normalmente são mais fáceis de serem obtidos que os dados de DGA.

Também podem ser citadas, como contribuições secundárias:

- Estudo do estado da arte sobre métodos de extração de regras a partir de técnicas de ML, apresentado no Apêndice A.

- A realização de um estudo simplificado de sensibilidade a erros nos dados de entrada comparando os métodos de DGA citados nas normas e as novas regras propostas nesta Tese.

## **1.4. Organização do Documento**

O restante desse documento está organizado em quatro capítulos. No Capítulo 2, é inicialmente apresentado o problema da identificação de falhas incipientes em transformadores de potência. Ao final do capítulo são apresentados diversos métodos clássicos para a análise de transformadores. As limitações desses métodos clássicos também são discutidas no capítulo.

As características, métodos de treinamento e problemas de utilização associados com uma DT são comentados no Capítulo 3.

No Capítulo 4 é apresentado um novo método, proposto nesse trabalho, de utilização de uma DT para a solução de problemas e obtenção de regras de classificação, baseado no aumento de dimensão dos atributos de entrada da árvore, permitido construir uma melhor solução para problemas de difícil tratamento. Exemplos de utilização da metodologia também são apresentados.

A utilização de uma DT com entradas modificadas e treinada para a identificação de falhas incipientes em equipamentos com isolamento a óleo, como os transformadores de potência, é consolidada no Capítulo 5. Também é realizado um estudo comparativo dos resultados obtidos das regras extraídas com os resultados de métodos clássicos, utilizando dados indicados nas normas e em bibliografias.

No Capítulo 6 são apresentadas as principais conclusões desse trabalho e proposições de trabalhos futuros.

## **1.5. Produção Científica**

Durante o desenvolvimento da pesquisa, foram escritos seis (6) artigos científicos apresentados em sessões orais de congressos científicos nacionais e internacionais, e dois (2)

artigos em revistas internacionais com *qualis* Capes A1, relacionados com os temas abordados na Tese. Esses artigos são citados a seguir:

1. **AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S.; BARBOSA, PONTES, R. S. T. An improved DGA method based on rules extracted from a high-dimension input space.** *Electronic Letters*, Volume 48, Issue 17, 16 August 2012, p. 1048-1049.
2. **BARBOSA, F. R; ALMEIDA, O. M.; BRAGA, A. P. S.; AMORA, M. A. B.; CARTAXO, S. J. M. Application of an Artificial Neural Network in the Use of Physicochemical Properties as a Low Cost Proxy of Power Transformers DGA Data.** *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 1, February 2012, p. 239 – 246.
3. **BARBOSA, F. R; ALMEIDA, O. M.; BRAGA, A. P. S.; AMORA, M. A. B.; Diagnóstico de Transformadores de Potência Utilizando a Dualidade entre os Ensaio Físico-Químico e Cromatográfico.** XV Encontro Regional Ibero-Americano do CIGRÉ (ERIAN), Foz do Iguaçu, PR, 2013.
4. **SOUSA, R.; AMORA, M.; BARBOSA, F. R.; PONTES, R. T.; BRAGA, A.P.S.; ALMEIDA, O. M. Análise de Métodos para o Diagnóstico de Falhas Incipientes em Transformadores com Base na Concentração de Gases no Óleo Isolante.** IV Simpósio Brasileiro de Sistemas Elétricos - SBSE 2012, Goiás, 2012.
5. **AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S., BARBOSA, F. R., LIMA, S. S., LISBOA, L. A. C. Decompositional Rule Extraction from Artificial Neural Networks and Application in Analysis of Transformers.** In The 15th International Conference on Intelligent System Applications to Power Systems. Curitiba, Brazil, November 8 - 12, 2009.
6. **AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S., BARBOSA, F. R., LIMA, S. S., LISBOA, L. A. C. Aplicação de Extração de Regras para Análise de Transformadores de Potência.** In THE 8th LATIN-AMERICAN CONGRESS ON ELECTRICITY GENERATION AND TRANSMISSION - CLAGTEE 2009, October, 18 to 22nd, 2009, Ubatuba, São Paulo.
7. **AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S., BARBOSA, F. R., LIMA, S. S., BENIGNO, V. B. L., LISBOA, L. A. C. Extração de Conhecimento de Redes**

**Neurais Artificiais e Aplicação na Análise de Transformadores.** In IX SBAI – Simpósio Brasileiro de Automação Inteligente, Brasília, DF, 20 a 23 de setembro de 2009.

- 8. AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S., BARBOSA, F. R., LIMA, S. S., BENIGNO, V. B. L., LISBOA, L. A. C. Estudo de Métodos para Obtenção de Conhecimento de Redes Neurais.** In IX SBAI – Simpósio Brasileiro de Automação Inteligente, Brasília, DF, 20 a 23 de setembro de 2009.

## 2. DIAGNÓSTICO DE FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA

As condições de funcionamento e integridade de um equipamento de potência imerso em óleo isolante, como os transformadores de potência, podem ser estabelecidas a partir da análise do óleo. Os métodos de diagnósticos com base na análise de gases dissolvidos no óleo isolante (DGA - *Dissolved Gas Analysis*) são amplamente utilizados. Esses métodos baseiam-se na análise da concentração e taxa de produção de gases gerados e dissolvidos no óleo do transformador, e procuram associar o tipo de falha aos gases presentes. Vários desses métodos são definidos em normas técnicas que regem a operação e manutenção de equipamentos isolados a óleo (IEEE, 2008; IEC, 1999).

Na seção 2.1 são descritos os tipos de falhas internas que podem surgir durante o funcionamento de um equipamento de potência com isolamento a óleo, como no caso dos transformadores. Na seção 2.2 é discutida a formação de gases durante a ocorrência dessas falhas e como as concentrações dos gases formados podem servir como indicativo para o diagnóstico do transformador. Uma revisão dos métodos indicados nas normas para o diagnóstico de transformadores baseados na análise dos gases dissolvidos no óleo isolante é apresentada na seção 2.3. A ocorrência de erros nas medidas de concentrações dos gases pode afetar o diagnóstico de equipamentos isolados a óleo e é comentada na seção 2.4. Os problemas associados à utilização dos métodos citados nas normas são evidenciados na seção 2.5, com um comparativo dos métodos. A possibilidade da utilização de técnicas de *Learning Machine* (ML) para o diagnóstico de transformadores é discutida na seção 2.6. Nas seções 2.7 e 2.8 são discutidas a avaliação da integridade do óleo isolante através de dados físico-químicos, e a possibilidade de correlação entre os dados físico-químicos e o diagnóstico de falhas no equipamento que utiliza o óleo isolante. A seção 2.9 apresenta a conclusão do capítulo.

### 2.1. Falhas em Transformadores

Na isolação de equipamentos e componentes, tais como transformadores de potência, buchas, reatores e transformadores de medição, é normalmente utilizado o papel impregnado com óleo isolante. O isolamento de papel impregnado é um isolamento composto de celulose e óleo isolante, em que cada um dos componentes possui seus processos normais de

degradação e que, quando em contato, interagem entre si, podendo alterar parcialmente suas características individuais. Os principais fatores primários externos que levam à degradação da celulose e do óleo são o aquecimento, umidade e oxigênio. Os métodos de avaliação da degradação podem medir diretamente características intrínsecas do isolamento, produtos de sua degradação, ou ainda seus efeitos em parâmetros físicos ou químicos (ZIRBES, ROLIM, e ZÜRN, 2005).

Quando o isolamento não tem mais integridade para realizar sua função tem-se a falha do isolamento. Os processos que colaboram para a aceleração da incapacidade do sistema de isolamento, quando em fase principiante, são conhecidos como falhas incipientes (ZIRBES, ROLIM, e ZÜRN, 2005). As falhas incipientes podem ser classificadas em termos dos esforços térmicos, onde sobreaquecimentos são os agentes principais, e esforços elétricos, relacionados às descargas internas no equipamento (SOUSA *et al.*, 2012).

### **2.1.1. Falhas Térmicas**

Quando submetidos a temperaturas entre 150 °C e 500 °C, os óleos minerais utilizados para isolamento sofrem decomposição nos gases hidrogênio (H<sub>2</sub>) e metano (CH<sub>4</sub>), e apenas traços de outros gases, tais como etileno (C<sub>2</sub>H<sub>4</sub>) e etano (C<sub>2</sub>H<sub>6</sub>). À medida que a temperatura se eleva, na vizinhança da falha ocorre o aumento da produção dos gases etileno e etano. Inicialmente, a taxa de produção de etano eleva-se mais rapidamente, seguida pelo etileno. Entretanto, quando uma falha promove o aumento substancial da temperatura, manifesta-se a produção de acetileno (C<sub>2</sub>H<sub>2</sub>). Daí, a presença deste gás indicar para especialistas uma condição extrema dentro do equipamento, considerada como o maior nível de alarme em se tratando de falhas incipientes (LIMA, 2005; BARBOSA, 2008).

A decomposição térmica da celulose e outros isolamentos sólidos produzem óxidos de carbono (CO e CO<sub>2</sub>) em quantidades relevantes já em temperaturas muito baixas. Portanto, a relação entre monóxido (CO) e dióxido de carbono (CO<sub>2</sub>) carrega em si a importância de indicar a decomposição térmica da celulose em termos quantitativos. Sendo assim, o IEEE Std. C57.104-2008 (IEEE, 2008) utiliza-se da razão CO<sub>2</sub>/CO e estabelece que sete (7,0) é um valor típico, muito embora alguns pesquisadores contestem (KAN e MIYAMOTO, 1995; IEEE, 2008). De forma mais geral, uma razão CO<sub>2</sub>/CO na faixa de 3,0 a 10,0 é considerada normal. Também, de modo semelhante, a norma IEC 60599 (IEC, 1999) propõe a avaliação da degradação da celulose através do monitoramento dos óxidos de carbono, no entanto,

utiliza a razão  $CO/CO_2$ , sendo definida uma faixa de valores ( $0,07 < CO/CO_2 < 0,30$ ) para uma condição normal de envelhecimento (WANG, 2000; BARBOSA, 2013).

Em uma situação da celulose sob temperatura extremamente alta, como na ocorrência de arcos elétricos, a razão  $CO/CO_2$  aproxima-se de 1:1, pois a geração de monóxido de carbono (CO) aumenta muito rapidamente em comparação à produção de dióxido de carbono ( $CO_2$ ). Porém, em uma situação de leve sobrecarga ou problemas de ventilação, onde ocorre leve sobreaquecimento, o  $CO_2$  cresce muito mais rapidamente que o CO, portanto a taxa de  $CO/CO_2$  fica na faixa entre 1:10 e 1:20 (WANG, 2000; LIMA, 2005; BARBOSA, 2008).

### **2.1.2. Falhas Elétricas**

Quando há a ocorrência de descargas de baixa intensidade de energia, como descargas parciais ou arcos intermitentes de baixa intensidade, ou descargas de alta intensidade de energia, como arcos elétricos, é provável que ocorra a decomposição do óleo devido à temperatura que pode ser atingida por estas falhas. Na persistência da falha ou na maior duração da mesma, a temperatura do óleo pode se elevar acima de  $1500^{\circ}C$  (WANG, 2000; BARBOSA, 2013).

Para as descargas elétricas de baixa intensidade de energia, a produção de hidrogênio apresenta-se elevada, com pequena quantidade de metano e apenas possíveis traços de acetileno. À medida que a intensidade de energia das descargas aumenta, e, conseqüentemente, aumenta a temperatura na vizinhança da falha, ocorre uma significativa aceleração da produção de etileno e acetileno. Se a quantidade de gases dissolvidos é elevada e ocorrer uma alteração na temperatura que modifique o valor da solubilidade de saturação, os gases antes dissolvidos podem ser liberados na forma de bolhas. A existência de bolhas de gases não dissolvidos altera a condição de rigidez dielétrica, levando a uma ruptura do isolamento (LIMA, 2005).

No caso da intensidade das descargas elétricas atingirem valores de alta energia, ou seja, a formação de descargas contínuas, como arcos elétricos, a produção de acetileno torna-se relevante, uma vez que a temperatura se eleva entre  $800^{\circ}C$  e  $2800^{\circ}C$ . Possíveis causas de formação destes pontos quentes são: centelhamentos promovidos por más conexões ou interrupção de correntes no comutador, e descargas de contorno ou arcos de potência

decorrentes da ruptura dielétrica do óleo entre os enrolamentos, entre espiras ou entre espiras e massa (LIMA, 2005; BARBOSA, 2008).

## **2.2. Formação de Gases Dissolvidos no Óleo**

O óleo mineral usado como isolante líquido dos transformadores de potência apresenta decomposição de acordo com o tempo de envelhecimento. Entretanto, um processo de aceleração desta decomposição pode ser desencadeado quando o equipamento é submetido a esforços térmicos e elétricos. Essa decomposição geralmente é caracterizada por formação de hidrogênio ativo e fragmentos instáveis de hidrocarbonetos ou radicais livres advindos da quebra de algumas ligações químicas do tipo hidrogênio-carbono (H-C) ou carbono-carbono (C-C). Em meio a essas condições, pode surgir o que se conhece por oxidação do óleo. A consequência é a liberação de oxigênio resultante da formação de hidroperóxidos instáveis pela ação catalítica de metais, como o cobre (MILASCH, 1984). Com o avanço deste processo, são formados produtos quimicamente ativos, como ácidos. Isto resulta em aumento de acidez e do fator de dissipação do óleo. Pode ser verificado, também, um aumento da tensão interfacial e um aumento da capacidade de dissolução de água. A rigidez dielétrica é afetada pelo aumento da quantidade de água espalhada no óleo. Uma oportuna elevação de temperatura permite que esses radicais livres se combinem formando gases, hidrogênio molecular, metano, etano e até mesmo podem se recombinar para a formação de novas moléculas estáveis (ZIRBES, ROLIM, e ZÜRN, 2005).

Os principais fatores primários externos que levam à degradação da celulose e do óleo isolante são o aquecimento, umidade e oxigênio. Alguns dos produtos da degradação, assim como a umidade e o aquecimento podem ter seus efeitos ampliados pela aplicação de um campo elétrico (ZIRBES, ROLIM, e ZÜRN, 2005).

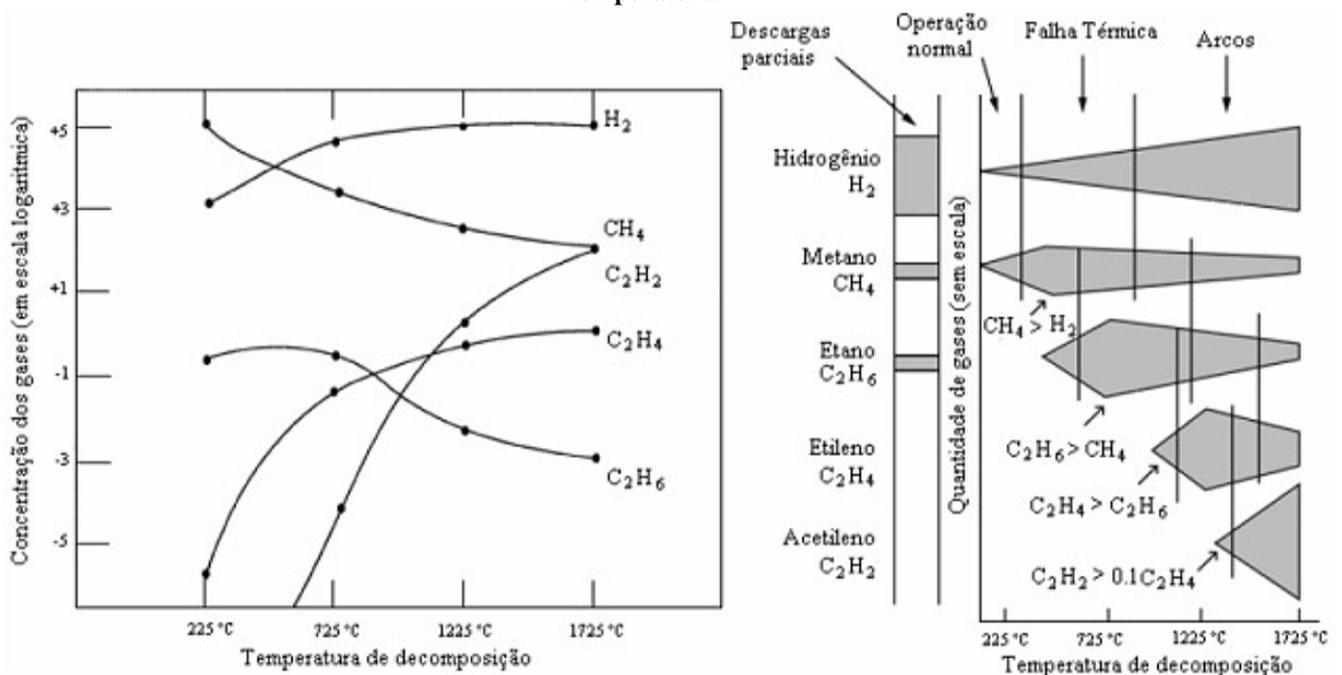
Seria muito oneroso quantificar os compostos presentes no processo e as condições de energia do ambiente da reação. Consequentemente, há dificuldade na aplicação dos conhecimentos da teoria cinética das reações químicas para explanar de forma trivial a decomposição de tais hidrocarbonetos.

Entretanto, um modelo termodinâmico proposto por (HALSTEAD, 1959) permite conhecer a pressão parcial de cada gás em função da temperatura. Esta proposta admite que todos os hidrocarbonetos do óleo são decompostos nos mesmos produtos e que estes estão em

equilíbrio entre si. Idealmente, toma-se que as reações ocorrem em temperatura constante. Podem-se considerar também constantes de equilíbrio conhecidas das reações de decomposição relevantes (BARBOSA, 2013). O modelo está apresentado na Figura 2.

De acordo com o modelo na Figura 2, a taxa de evolução de cada gás pode ser calculada em qualquer temperatura, tal que uma relação entre geração de gás e temperatura pode ser obtida para cada gás. Os estudos dessas relações mostram que os gases são gerados na seguinte ordem com um aumento de temperatura:  $H_2 \rightarrow CH_4 \rightarrow C_2H_6 \rightarrow C_2H_4 \rightarrow C_2H_2$ . O gás hidrogênio ( $H_2$ ) é gerado em baixas temperaturas e sua quantidade aumenta de forma constante, enquanto o acetileno ( $C_2H_2$ ) é gerado em temperaturas muito altas (próximo a 1000 °C) e também apresenta aumento constante de sua quantidade (WANG, 2000).

Figura 2- Formação de gases: a) pressão x temperatura (IEEE, 2008), b) concentração dos gases x temperatura



Fonte: (WANG, 2000; SANDERSON, 2005).

### 2.3. Métodos Indicados nas Normas para Diagnóstico Baseado na Concentração de Gases no Óleo Isolante

As várias técnicas indicadas em normas (IEEE, 2008; IEEE, 1991; IEC, 2008; IEC, 1999; ABNT, 1982) para interpretação da análise cromatográfica visam diagnosticar a condição atual do sistema de isolamento do transformador. Dentre os métodos mais conhecidos estão: método do gás chave (IEEE, 2008; IEEE, 1991; IEC, 2008; IEC, 1999; ABNT, 1982), método das razões de Dörnenburg (IEEE, 2008; IEEE, 1991; IEC, 2008; IEC,

1999; ABNT, 1982), método das razões de Rogers (IEEE, 2008; IEEE, 1991; IEC, 2008; IEC, 1999; ABNT, 1982), e o triângulo de Duval (IEC, 2008; IEC, 1999). Estes métodos são baseados no modelo termodinâmico de Halstead (seção 2.2).

Cada um destes métodos será comentado a seguir, com base nas referências (SOUSA *et al.*, 2012; BARBOSA, 2013; LIMA, 2005; WANG, 2000).

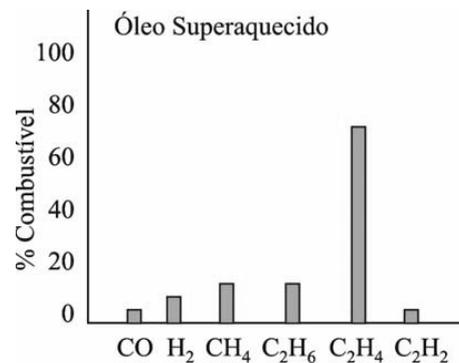
### 2.3.1. Método do Gás Chave

O método do gás chave faz uso do valor percentual dos gases para diagnosticar falhas incipientes em transformadores. A essência do método considera o conhecimento da temperatura onde um gás é predominantemente gerado e a correlação dessa temperatura, e conseqüentemente do gás, a uma determinada falha. O gás predominante é o gás chave ou principal (LIMA, 2005).

As Figuras 3, 4, 5 e 6 ilustram esse método e indicam as proporções relativas dos gases para quatro tipos de falhas.

- *Falha Térmica no óleo:* Produtos da decomposição incluem etileno ou eteno ( $C_2H_4$  – cerca de 63%) e metano ( $CH_4$  – 16%), junto com menor quantidade de hidrogênio ( $H_2$  – 2%) e etano ( $C_2H_6$  – 19%). Traços de acetileno ( $C_2H_2$ ) podem ser formados se a falha for severa ou envolver contatos elétricos. O gás principal indicativo, como mostrado pelas proporções, é o etileno.

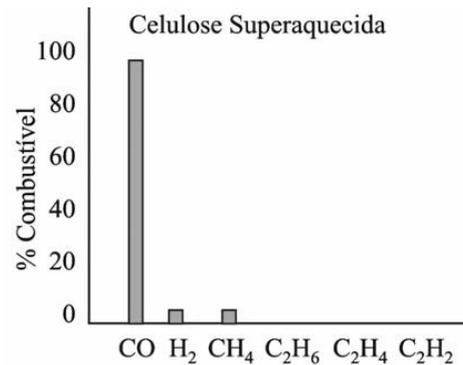
Figura 3- Composição dos gases para a falha térmica no óleo



Fonte: (LIMA, 2005).

- *Falha Térmica na Celulose*: Grandes quantidades de monóxido de carbono (CO – 92%) e dióxido de carbono (CO<sub>2</sub>) são desenvolvidas do sobreaquecimento da celulose. Gases hidrocarbonetos, tais como etileno e metano, serão formados se a falha envolve uma estrutura impregnada de óleo. O gás indicativo é o monóxido de carbono.

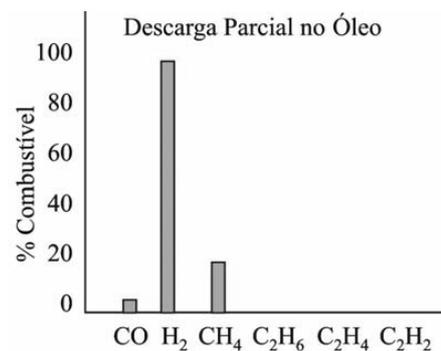
**Figura 4- Composição dos gases para a falha térmica na celulose**



Fonte: (LIMA, 2005).

- *Descarga parcial ou corona*: Produz hidrogênio (85%) e metano (13%), com pequenas quantidades de etano e etileno. Quantidades comparáveis de monóxido e dióxido de carbono podem resultar de descargas na celulose. O gás indicativo é o hidrogênio.

**Figura 5 - Composição dos gases para descarga parcial**

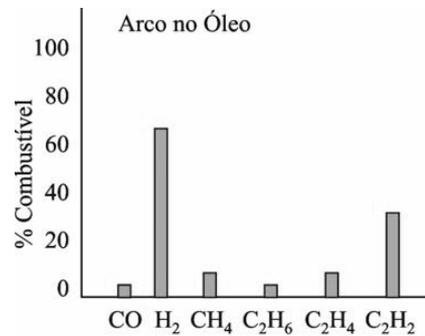


Fonte: (LIMA, 2005).

- *Descarga de alta energia ou arco*: Grandes quantidades de hidrogênio (60%) e acetileno (30%) são produzidas, com menores quantidades de metano (5%) e etileno (3%). Dióxido e monóxido de carbono podem também ser formados se a

falha envolve a celulose. Óleo pode ser carbonizado. O gás indicativo é o acetileno.

**Figura 6 - Composição dos gases para descarga de alta energia**



Fonte: (LIMA, 2005).

### 2.3.2. Método das Razões de Dörnenburg

O uso de razões entre as concentrações dos gases para indicar um tipo possível de falha é um processo empírico baseado na experiência de cada pesquisador de forma individual, correlacionando as análises de gases de muitas unidades com o tipo de falha posteriormente determinada como a causa para o distúrbio ou falha quando a unidade foi examinada. Esse processo foi atribuído a Dörnenburg e posteriormente confirmado por Rogers em sistemas Europeus, dos quais a maior parte da correlação do diagnóstico é obtida (SOUSA *et al.*, 2012).

As teorias de diagnósticos baseadas nos princípios de degradação térmica empregam um conjunto de relações de certos gases chave como os indicadores do tipo de falha (SOUSA *et al.*, 2012). As cinco relações são mostradas na Tabela 1.

**Tabela 1- Razões entre as Concentrações dos Gases**

Abreviação	R1	R2	R3	R4	R5
Razão	CH <sub>4</sub> /H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub> /CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub> /C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub>

Fonte: (WANG, 2000).

O método de Dörnenburg utiliza as razões R1, R2, R3 e R4 para indicar uma determinada falha dentre três tipos gerais sugeridos: sobreaquecimento, descargas de baixa energia (corona) e descargas de alta energia (arco elétrico).

Se no mínimo uma das concentrações dos gases em  $\mu\text{L/L}$  (p.p.m.), para  $\text{H}_2$ ,  $\text{CH}_4$ ,  $\text{C}_2\text{H}_2$ ,  $\text{C}_2\text{H}_4$  e  $\text{C}_2\text{H}_6$  excederem duas vezes os valores para o limite L1 (Tabela 2) e um dos outros gases excederem os valores para o limite L1, a unidade é considerada defeituosa. Se no mínimo um dos gases em cada relação exceder o limite L1, o procedimento das razões é considerado válido, caso contrário, as razões não são significativas e a unidade deveria ter novas amostras retiradas e ser investigada por processos alternativos (SOUSA *et al.*, 2012).

Os valores de concentrações apresentados pela Tabela 2, que seguem a norma (IEEE, 2008), diferem dos valores encontrados na versão anterior de (IEEE, 1991), apresentados na Tabela 3, e dos valores originais de Dörnenburg para validação do seu método (Tabela 4).

**Tabela 2 - Concentrações Limite dos Gases Conforme (IEEE, 2008)**

<b>Gás Chave</b>	<b>Concentrações L1 (<math>\mu\text{L/L}</math> - p.p.m.)</b>
Hidrogênio ( $\text{H}_2$ )	100
Metano ( $\text{CH}_4$ )	120
Acetileno ( $\text{C}_2\text{H}_2$ )	1
Etileno ( $\text{C}_2\text{H}_4$ )	50
Etano ( $\text{C}_2\text{H}_6$ )	65

Fonte: (IEEE, 2008).

**Tabela 3 - Concentrações Limite dos Gases Conforme (IEEE, 1991)**

<b>Gás Chave</b>	<b>Concentrações L1 (<math>\mu\text{L/L}</math> - p.p.m.)</b>
Hidrogênio ( $\text{H}_2$ )	100
Metano ( $\text{CH}_4$ )	120
Acetileno ( $\text{C}_2\text{H}_2$ )	35
Etileno ( $\text{C}_2\text{H}_4$ )	50
Etano ( $\text{C}_2\text{H}_6$ )	65

Fonte: (IEEE, 1991).

**Tabela 4 - Concentrações Limite dos Gases Conforme Dörnenburg**

<b>Gás Chave</b>	<b>Concentrações L1 (<math>\mu\text{L/L}</math> - p.p.m.)</b>
Hidrogênio ( $\text{H}_2$ )	200
Metano ( $\text{CH}_4$ )	50
Acetileno ( $\text{C}_2\text{H}_2$ )	15
Etileno ( $\text{C}_2\text{H}_4$ )	60
Etano ( $\text{C}_2\text{H}_6$ )	15

Fonte: (ZIRBES, ROLIM e ZÜRN, 2005).

Como pode ser observado nas tabelas, entre os valores das duas versões da norma, há diferença apenas no valor para o acetileno, enquanto que para os valores originais de Dörnenburg em relação às normas, há diferenças em todos os limites.

As faixas das razões R1, R2, R3 e R4, definidas no método de Dörnenburg, e a respectiva falha associada são apresentadas na Tabela 5.

**Tabela 5 - Valores para as Razões de Gases para o Método de Dörnenburg**

<b>Falha Incipiente</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>	<b>R4</b>
Falta térmica	> 1,0	< 0,75	< 0,3	> 0,4
Descarga de baixa enérgia	< 0,1	-	< 0,3	> 0,4
Descarga de alta enérgia	> 0,1 e <1,0	> 0,75	> 0,3	< 0,4

Fonte: (IEEE, 2008).

### 2.3.3. Método das Razões de Rogers

O método das razões de Rogers segue o mesmo procedimento geral utilizado no método de Dörnenburg, exceto pelo fato que as razões utilizadas são agora: R1, R2 e R5. A validade desse método é baseada na correlação dos resultados de um grande número de testes de falhas com a análise dos gases para cada caso. A Tabela 6 apresenta os valores para as três relações de gases e o correspondente diagnóstico sugerido (SOUSA *et al.*, 2012 ).

Os primeiros trabalhos de Rogers utilizavam quatro relações de gases. Posteriormente, as normas adotaram o método de Rogers com apenas três relações de concentração de gases e cinco gases. Sendo eliminada a relação  $C_2H_6/CH_4$  que era utilizada para identificação de sobreaquecimento de baixa temperatura (ZIRBES, ROLIM, e ZÜRN, 2005).

É possível observar que as faixas de variação das relações apresentam superposições, não permitindo uma interpretação única. Também em muitos casos não existe um diagnóstico definido, pois as combinações de variações que podem ocorrer nas relações são maiores que aquelas definidas na tabela do método. Esse fato ocorre com todas as variantes do método (ZIRBES, ROLIM, e ZÜRN, 2005; SOUSA *et al.*, 2012).

Tabela 6 - Valores das Razões de Gases para o Método de Rogers

Caso	R1	R2	R5	Diagnóstico
0	> 0,1 a < 1,0	< 0,1	< 1,0	Unidade Normal
1	< 0,1	< 0,1	< 1,0	Descarga de baixa energia
2	0,1 a 1,0	0,1 a 3,0	> 3,0	Descarga de alta energia
3	> 0,1 a < 1,0	< 0,1	0,1 a 3,0	Falha térmica de baixa temperatura
4	> 1,0	< 0,1	0,1 a 3,0	Falha térmica < 700 °C
5	> 1,0	< 0,1	> 3,0	Falha térmica > 700 °C

Fonte: (IEEE, 2008).

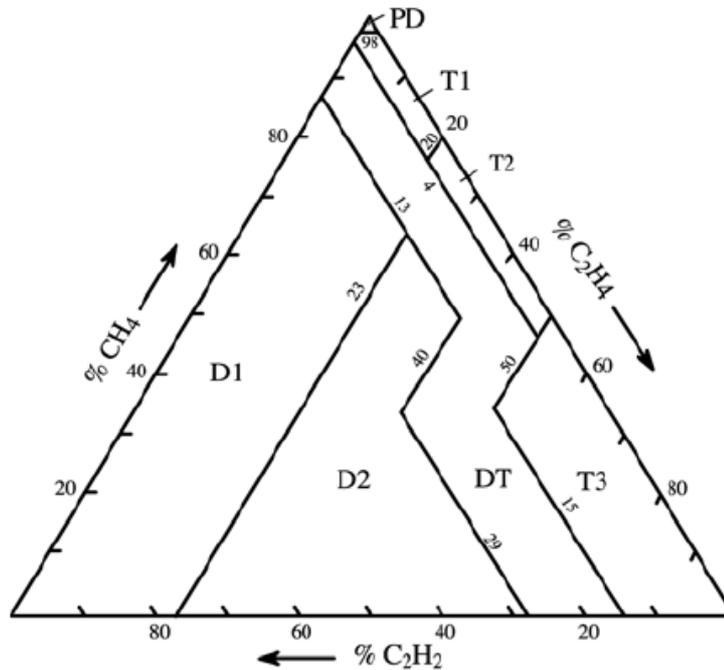
### 2.3.4. Triângulo de Duval

Este método foi desenvolvido por Michel Duval em 1974 e está descrito no apêndice B do IEC 60599 (IEC, 2008). O procedimento de aplicação deste método se inicia com o cálculo das percentagens dos gases metano ( $CH_4$ ), etileno ( $C_2H_4$ ) e acetileno ( $C_2H_2$ ) em relação à soma das concentrações desses gases gerados em p.p.m. – partes por milhão – ( $CH_4 + C_2H_4 + C_2H_2$ ). Uma vez obtidos os percentuais ( $\%CH_4$ ,  $\%C_2H_4$ ,  $\%C_2H_2$ ), estes servem para identificação de um ponto em um sistema de coordenadas triangulares (SOUSA *et al.*, 2012). O triângulo de Duval é geograficamente subdividido em zonas que representam as falhas. As coordenadas triangulares correspondentes ao resultado da análise de gases dissolvidos em p.p.m. podem ser calculadas como segue (DUVAL, 2002):  $\%C_2H_2 = 100x/(x+y+z)$ ;  $\%C_2H_4 = 100y/(x+y+z)$ ;  $\%CH_4 = 100z/(x+y+z)$ , com  $x = (C_2H_2)$ ;  $y = (C_2H_4)$ ;  $z = (CH_4)$ , em p.p.m..

A Figura 7 ilustra de forma gráfica o método do triângulo de Duval, bem como a composição das coordenadas e as zonas das respectivas falhas, de códigos (DUVAL, 2002): DP = descargas parciais, D1 = descargas de baixa energia, D2 = descargas de alta energia, T1 = falhas térmicas de temperatura < 300 °C, T2 = falhas térmicas de temperatura 300 °C < T < 700 °C, e T3 = falhas térmicas de temperatura > 700 °C.

O código DT, contido no triângulo, está relacionado à combinação de falhas térmicas e elétricas associadas principalmente com transformadores com OLTC - *On Load Tap Changer* (comutador de tap com carga) (SOUSA *et al.*, 2012).

Figura 7 - Método do Triângulo de Duval



Fonte: (DUVAL, 2002).

## 2.4. Confiabilidade das Medições dos Gases Dissolvidos no Óleo Isolante

Os parâmetros de qualidade das medidas de laboratório de DGA, como: capacidade de repetição, reprodutibilidade e precisão, são definidos na norma IEC 60567 (IEC, 1992). A capacidade de repetição está relacionada com as diferenças que podem ser observadas ou não quando múltiplas amostras do mesmo óleo retirado do equipamento são analisadas pelo mesmo laboratório ao longo de um curto intervalo de tempo. A reprodutibilidade é definida em função de diferenças que podem ser observadas quando múltiplas amostras do mesmo óleo são analisados por laboratórios diferentes (reprodutibilidade inter-laboratórios) ou são analisadas pelo mesmo laboratório (reprodutibilidade intra-laboratório) ao longo de períodos de tempo longos. A precisão está relacionada com as diferenças que são observadas entre os valores medidos e os valores nominais contidos numa amostra de gás em óleo para validação, preparados de acordo com o procedimento padrão.

A capacidade de repetição e a reprodutibilidade intra-laboratório asseguram que as condições de medição dos laboratórios são estáveis e não variam aleatoriamente. No entanto, os resultados de DGA podem ter uma boa capacidade de repetição e ainda assim a reprodutibilidade ainda ser inexata, devido a um desvio sistemático nas análises que não é detectada até que uma aferição de precisão seja feita (DUVAL e DUKAM, 2005).

Se a capacidade de repetição e a reprodutibilidade são pobres, o diagnóstico da tendência de falhas no equipamento será comprometido, afetando a qualidade e a segurança da análise. Resultados imprecisos de DGA podem resultar em uma detecção pobre de falhas incipientes ou mesmo numa detecção errada, podendo ocasionar graves consequências para os equipamentos em serviço (DUVAL e DUKAM, 2005).

Em (DUVAL e DUKAM, 2005) os autores indicam que é conhecido na indústria que muitos laboratórios oferecem resultados razoavelmente precisos de DGA, mas que muitos outros fornecem resultados muito imprecisos. Pesquisas (DUVAL, 2003) foram realizadas pelo CIGRE TF 15-01-07 e TC10 IEC (MT25) para avaliar os quesitos capacidade de repetição, reprodutibilidade intra-laboratório e precisão de 25 laboratórios de 15 países diferentes, considerando-se os casos de níveis de concentração médio (10 a 100 ppm) e níveis baixos (1 a 10 ppm) para as concentrações de gases.

A avaliação de precisão dos laboratórios realizada pelo IEC/CIGRE foi efetuada através de testes do tipo Round-Robin (RRTs) (BOND e LASHLEY, 1996) utilizando amostras padrões conhecidas de gases dissolvidos em óleo. O desvio médio dos valores para cada gás preparado (erro de medida) foi utilizado como uma estimativa da precisão média de cada laboratório individual. A precisão média de todos os laboratórios IEC/CIGRE pesquisados foi então calculada como sendo a média das precisões individuais de cada laboratório.

Para os estudos realizados, a capacidade de repetição e a precisão média são indicadas nas Tabelas 7 e 8, respectivamente.

**Tabela 7 - Desvios para as Medidas de Repetição dos Laboratórios Avaliados**

<b>Laboratório</b>	<b>Concentrações médias de gás</b>	<b>Concentrações baixas de gás</b>
Melhor	$\pm 1\%$	$\pm 7\%$
Média	$\pm 7\%$	$\pm 27\%$
Pior	$\pm 15\%$	$\pm 65\%$

Fonte: (DUVAL e DUKAM, 2005).

O valor da reprodutibilidade intra-laboratórios para os laboratórios avaliados pelo CIGRE foi estimada como sendo de  $\pm 10\%$  para uma concentração média, e o melhor valor de  $\pm 2\%$  (DUVAL e DUKAM, 2005).

**Tabela 8 - Desvios para as Medidas de Precisão dos Laboratórios Avaliados**

<b>Laboratório</b>	<b>Concentrações médias de gás</b>	<b>Concentrações baixas de gás</b>
Melhor	± 3%	± 22%
Média	± 15%	± 30%
Pior	± 65%	± 64%

Fonte: (DUVAL e DUKAM, 2005).

Na seção a seguir e no Capítulo 5 serão demonstrados os resultados de um estudo simplificado do impacto da utilização de medidas com erros na precisão do diagnóstico de falhas incipientes em equipamentos com óleo isolante, inclusive comparando métodos de diagnóstico.

## **2.5. Limitações das Metodologias Tradicionais para o Diagnóstico de Transformadores por meio de DGA**

A base das técnicas de análise de gases dissolvidos é o processo de cromatografia laboratorial, a qual a amostra do óleo isolante do transformador é submetida. Esse processo tem como produto resultante uma estratificação das concentrações dos gases dissolvidos no óleo e, a partir desses valores quantificadores, aplica-se uma técnica de interpretação dos dados. Esses dados podem ser expressos na forma de concentrações individuais de cada gás, através de razões entre as concentrações dos gases ou ainda como porcentagens em relação à concentração total de gases (SOUSA *et al.*, 2012).

O processo de falha está intrinsecamente relacionado com as temperaturas esperadas no óleo em função de seus mecanismos. São estabelecidas faixas de temperatura para as quais existe uma maior probabilidade da ocorrência de uma determinada falha. Estabelecida a correlação entre a temperatura e a formação de gases no óleo isolante (Figura 2), o processo de falha e sua intensidade podem ser estimados. Essa é a forma usual de estabelecer um diagnóstico do transformador.

A interpretação de uma análise individual pode não ser muito efetiva, pois, mais de uma falha pode estar ocorrendo ao mesmo tempo ou ainda um tipo de falha pode progredir para outro, tal como ocorre com alguns problemas elétricos que têm origem em problemas térmicos. Portanto, é de fundamental importância estabelecer uma base de dados para servir como valores de referência, de forma tal que se torna possível determinar a evolução temporal

das concentrações dos gases, tomando-se como base os valores históricos. Isso é implementado por intermédio de coletas periódicas de amostras de óleo, análise cromatográfica para determinar as concentrações dos gases e formação de um banco de dados com o histórico do transformador (SOUSA *et al.*, 2012).

O tipo de equipamento e seu projeto têm influência no volume de gases gerados. Tradicionalmente, a análise de gases é utilizada em transformadores de potência. Entretanto, é uma técnica recomendada para qualquer equipamento que utilize isolamento de papel impregnado em óleo. A recomendação da IEC 60599 (IEC, 2008) para análise de gases dissolvidos reconhece um comportamento diferenciado entre os equipamentos quanto à geração de gases, estabelecendo diferentes limites por tipo de equipamento. É importante salientar que para baixas concentrações de gases a incerteza é significativa, e no cálculo das relações entre os gases pode representar uma variação de até 40% (ZIRBES, ROLIM, e ZÜRN, 2005).

Em (IEEE, 2008; IEC, 2008) é especificado que quando as relações ou concentrações de gases excedem seus respectivos valores típicos, uma tentativa de diagnóstico pode ser feita aplicando um dos métodos padrões de diagnóstico de análise de gases dissolvidos (métodos de Dörnenburg e Rogers). Esses métodos levam a um número significativo de casos onde nenhum diagnóstico pode ser dado, porque os valores das relações ficam fora da faixa de falhas (SOUSA *et al.*, 2012; LIMA, 2005). O método do gás chave também leva a um grande número de diagnósticos equivocados.

O método triângulo de Duval é uma abordagem gráfica, com menos erros, que gera coordenadas normalizadas em um sistema de coordenadas triangulares, tal que os pontos de dados caem dentro de uma zona correspondendo a um tipo de falha (DUVAL e PABLO, 2001).

Foram realizados testes com dados de transformadores e de outros equipamentos imersos em óleo isolante indicados no banco de dados IEC TC 10 (DUVAL e PABLO, 2001) e que sofreram falhas incipientes, utilizando os métodos indicados nas normas. Os resultados obtidos de cada método foram então comparados com o diagnóstico obtido de um comitê de especialistas após inspeção visual realizada no equipamento. Os resultados percentuais de acerto de cada método são apresentados na Tabela 9.

**Tabela 9 - Resultados de Aplicação dos Métodos para o IEC TC 10**

<b>Método</b>	<b>Acerto (%)</b>
Dörnenburg IEEE C57.104 - 1991	66,67
Dörnenburg IEEE C57.104 - 2008	67,52
Rogers IEEE C57.104 - 2008	56,41
Triângulo de Durval (IEC 599 - 1999)	88,03

Fonte: Elaborada pelo autor.

Como pode ser observado na tabela, o método do triângulo de Duval é o que apresenta o maior número de acertos dentre todos os métodos indicados em normas, entretanto ainda apresentado quase 12% de erro. Os outros métodos apresentam um percentual de acerto significativamente menor.

Outros estudos de comparação apresentam resultados parecidos (AMORA *et al.*, 2012; SOUSA *et al.*, 2012; LI e WU, 2011) utilizando o mesmo banco de dados e outros valores também disponíveis.

Os valores indicados na Tabela 9 além de indicar diagnósticos equivocados por parte do método avaliado, também consideram o fato que muitos dos métodos tradicionais (métodos de Dörnenburg e Rogers) não permitem o diagnóstico para todas as combinações possíveis de entrada, apresentado um resultado de não diagnóstico, e, portanto, também gerando um erro em relação ao diagnóstico correto do equipamento. Então, fica evidente que existem situações que não são cobertas completamente pelas relações de entrada e saída providenciadas pelos métodos clássicos. Esses métodos clássicos são baseados em parte no modelo termodinâmico simplificado de Halstead (HALSTEAD, 1959), mas apresentam também um forte caráter empírico.

Como já destacado na seção 2.4, podem ocorrer erros no processo de obtenção das medidas de DGA. Na Tabela 10 são apresentadas a percentagem de sucesso nos diagnósticos de falhas incipientes utilizando o método do Triângulo, o método da razão de Rogers e o método de Dörnenburg, em situações de ocorrência de erros nas medições dos gases dissolvidos no óleo isolante do transformador. O banco de dados utilizado para comparação é novamente o IEC TC 10.

O erro é modelado por meio de uma distribuição normal, considerando um desvio padrão de 15% e de 65%, respectivamente relacionados com as variações médias e máximas

registradas para um conjunto de medidas em estudos realizados de verificação de laboratórios que realizam DGA (seção 2.4). A distribuição normal é então somada com os valores corretos, gerando um novo banco de dados de teste, no caso com a presença de um ruído de erro. Analisando os resultados apresentados na Tabela 10, observa-se que o método do Triângulo apresenta um menor valor de erro.

**Tabela 10 - Taxa de acerto quando da possibilidade de erro nas medidas de gás (IEC TC 10)**

Método	Sem erro	Erro com 15%	Erro com 65%
Rogers IEEE C57.104 - 2008	56,41	55,56	52,14
Dörnenburg IEEE C57.104 - 2008	67,52	67,52	64,96
Triângulo	88,03	85,47	77,78

Fonte: Elaborada pelo autor.

Também, deve ser salientado que nenhum método indicado em normas permite o diagnóstico de falhas simultâneas e nem avaliar diretamente a evolução histórica de uma possível falha. Sendo recomendado ao utilizador, montar um banco de dados histórico com os dados de DGA de um determinado equipamento, de forma a identificar a evolução de falhas ao longo do tempo.

A questão do diagnóstico de transformadores e de outros equipamentos imersos em óleo ainda representa um problema aberto a melhorias, como indicam os problemas expostos: métodos com possibilidade de diagnósticos não conclusivos, com erros significativos quando comparado com os resultados obtidos de equipamentos vistoriados, e ainda a possibilidade de erros no processo de obtenção das informações de DGA. Uma possibilidade é a utilização de métodos de ML na tentativa de superar os problemas encontrados nos métodos clássicos e capturando relações ainda não exploradas por tais métodos, entretanto como será exposto na próxima seção, nem sempre a utilização das técnicas de ML possibilita a obtenção de um método de diagnóstico eficiente e de fácil utilização.

## **2.6. Utilização de Técnicas de Aprendizagem de Máquina para o Diagnóstico de Transformadores através de DGA**

Nesta seção serão discutidas as vantagens e desvantagens da utilização de técnicas de ML para o diagnóstico de transformadores. A maioria dos trabalhos indicados na literatura, e que serão citados a seguir, utilizam Redes Neurais Artificiais (RNA) ou Sistemas de Regras Fuzzy (SRF), ou ainda uma combinação dos dois, para o diagnóstico de falhas incipientes em

transformadores de potência. Como será comentado a seguir, essas ferramentas apresentam a capacidade de revelar não linearidades que os métodos convencionais não conseguem lidar, também os sistemas *fuzzy* conseguem modelar as mudanças dos estados de diagnóstico de uma maneira mais suave.

### 2.6.1. Utilizando uma Rede Neural Artificial (RNA)

O relacionamento entre os produtos gasosos da degradação do sistema de isolamento de um transformador e as falhas que aceleraram a produção destes gases pode ser realizado por meio de uma RNA (HAYKIN, 2001). Dessa forma, uma RNA pode ser interpretada como uma ferramenta que fornece alternativas às metodologias convencionais de ensaios, produzindo resultados motivadores, principalmente devido às características intrínsecas da técnica, tais como a sua capacidade de generalização e a facilidade de integração com outras ferramentas computacionais (SILVA, SOUZA e SOUZA, 1999).

A aplicação de uma RNA para o diagnóstico de falhas incipientes em transformadores é particularmente interessante, pois, as redes neurais são capazes de adquirir conhecimento diretamente dos dados de entrada (concentrações dos gases) e assim podem revelar relações não lineares entre as entradas e saídas que ainda são desconhecidas pelos especialistas (WANG, 2000). Diversos pesquisadores têm utilizado RNA para detectar e identificar falhas incipientes em transformadores de potência (PATEL e KHUBCHANDANI, 2004; ZHANG *et al.*, 1996; GUARDADO *et al.*, 2001; NARESH, SHARMA e VASHISTH, 2008; SOUSA *et al.*, 2012). Mas, em geral, as redes neurais são utilizadas em conjunto com sistemas especialistas e *fuzzy* em uma ferramenta híbrida de diagnóstico (CHEIM *et al.*, 1999; HUANG, YANG e HUANG, 2002 ; WANG, LIU e GRIFFIN, 1998; AMORA *et al.*, 2009; MIRANDA e CASTRO, 2005; HOOSHMAND *et al.*, 2012).

Em (GUARDADO *et al.*, 2001) os autores realizaram um estudo comparativo entre diferentes topologias de rede. Foram utilizados cinco métodos para realizar o diagnóstico preliminar dos dados de treinamento: Dörnenburg, Rogers, Rogers modificado, IEC e CSUS (IEEE, 1991; IEC, 1999; MYERS, KELLY e PARRISH, 1981). A eficiência das redes ficou entre 87 e 100% e depende do método de diagnóstico utilizado.

Também, em (SOUSA *et al.*, 2012) foi desenvolvida uma RNA para o diagnóstico de transformadores. Os dados para o treinamento e testes da rede foram baseados no banco de dados gerais de diagnósticos de casos testados da IEC TC 10 (DUVAL e PABLO, 2001),

contendo 117 casos. Foram utilizados 70% dos dados para treinamento e 30% para testes da rede neural. A RNA implementada foi do tipo MLP – *Multilayer Perceptron* (perceptron de múltiplas camadas) (HAYKIN, 2001), com apenas uma camada escondida contendo 10 neurônios e um neurônio na camada de saída. As entradas da rede foram às razões de gases R1, R2 e R5. Para os casos considerados, tanto no treinamento quanto nos testes, a rede neural conseguiu atingir 100% de acerto.

A maioria de tipos de RNA apresenta a possibilidade apenas do mapeamento de entradas e de uma respectiva solução, não apresentando uma capacidade de representação das soluções a partir de regras de conhecimento.

Em muitas situações a capacidade de mapeamento direto de entradas e de valores de saída (soluções) é o suficiente e na verdade um dos maiores atrativos da utilização de uma RNA. Entretanto, em outros problemas, é importante, ou pelo menos seria útil, a possibilidade da representação de soluções a partir de regras de conhecimento, como por exemplo, em problemas de diagnóstico médico e em problemas de engenharia, onde os resultados precisam passar pela análise de um especialista devido às particularidades da análise. Em tais casos, uma possível solução é a utilização de alguma técnica de extração de regras associada à RNA. Entretanto, a maioria dessas técnicas não permitem a obtenção de regras realmente interpretáveis e de fácil utilização. No Apêndice A desse trabalho é apresentada uma breve revisão bibliográfica sobre métodos de extração de regras a partir de técnicas de ML.

### **2.6.2. Adotando um Sistema de Regras Fuzzy (SRF)**

Os métodos convencionais de interpretação de análises cromatográficas para diagnosticar a natureza da deterioração do sistema de isolamento de um transformador têm obtido grande aplicabilidade na indústria. Entretanto, tais métodos convencionais falham em alguns casos. Isso acontece, geralmente, quando há mais de uma falha ocorrendo ao mesmo tempo no interior do transformador. Os métodos convencionais definem intervalos fixos para determinadas razões entre as concentrações dos gases, assim, quando uma razão ultrapassa um determinado limiar ocorre uma mudança súbita da faixa à qual ela pertence (0, 1 ou 2). Na realidade, tal transição é suave ao invés de abrupta. Portanto, uma representação *fuzzy* é capaz de modelar corretamente esse tipo de transição, especialmente quando estão ocorrendo mais de um tipo de falha (LIMA, 2005). Neste sentido, diversos pesquisadores têm desenvolvido sistemas nebulosos. Tais sistemas, em geral, integram ferramentas híbridas para diagnosticar a

ocorrência de falhas incipientes em transformadores (DUKARM, 1993; ISLAM, WU e LEDWICH, 2000; MORAIS, ROLIM e SILVA, 2005; NETO e CHIGANER, 1999; WANG, LIU e GRIFFIN, 1998; XU *et al.*, 1997; HUANG, YANG e HUANG, 1997; AMORA *et al.*, 2009; MIRANDA e CASTRO, 2005; HOOSHMAND *et al.*, 2012).

Em (HUANG, YANG e HUANG, 1997) foi desenvolvido um sistema *fuzzy*, aliado a um algoritmo evolucionário, para o diagnóstico de falhas em transformadores. A definição da base de regras e das funções de pertinência está baseada no método de Rogers, formando um espaço padrão em três dimensões (uma para cada variável de entrada), com conjuntos nebulosos trapezoidais. Tais conjuntos nebulosos definem uma pertinência para os valores das razões entre as concentrações dos gases classificando em pequenos, médios e grandes. Assim, a base de regras deste sistema *fuzzy* apresenta 27 regras, com 24 parâmetros das funções de pertinência a serem determinados. Foram utilizados exemplos para formar um padrão para cada tipo de falha. Com dados novos, utiliza-se um algoritmo de programação evolucionária para buscar a minimização da função de erro em relação aos padrões definidos. Para 711 amostras analisadas, foi obtido um índice de acerto de 92% (LIMA, 2005).

Em (AMORA, 2009) é utilizado uma RNA para o diagnóstico de transformadores, juntamente com o método descrito em (CASTRO, MANTAS e BENÍTEZ, 2002) para a extração de regras fuzzy. Apesar de a metodologia obter um alto índice de acerto, próximo a 100%, e as regras obtidas conseguirem reproduzir exatamente a saída numérica da RNA, estas regras não permitem uma fácil interpretação e utilização, pois adotam um operador lógico não convencional, operador *i-OU*, que não é intuitivo e que combina propriedades de uma *t*-norma e uma *t*-conorma.

Também, em (MIRANDA e CASTRO, 2005) é proposto uma RNA para o diagnóstico de transformadores, e é utilizada uma nova forma de extração de regras propostas pelos autores que permite a representação do diagnóstico da RNA por regras simples e com operadores convencionais, entretanto o método exige uma forma especial de treinamento da RNA e também o número de regras obtidas é excessivamente grande, sendo, então, obtido um conjunto de regras que não é intuitivo.

### 2.6.3. Dificuldades na Utilização de uma RNA e/ou SRF para o Diagnóstico de Transformadores

Os problemas expostos, nas subseções anteriores, resumem as dificuldades encontradas na utilização de uma RNA ou SRF associado com algum processo de extração de conhecimento para o diagnóstico de transformadores (AMORA *et al.*, 2012): (i) soluções baseadas em RNA apresentam uma grande precisão, mas o raciocínio que justifique o diagnóstico não é interpretável, e (ii) sistema de diagnóstico que utiliza um SRF pode ser interpretável, mas normalmente exigem a utilização de operadores especiais (ex.: operador *i-OU*) ou são obtidos um número elevado de regras, prejudicando a facilidade de utilização do método.

Uma RNA representa uma ferramenta eficiente de obtenção de solução a partir de um conjunto de dados, inclusive com capacidade de inferência, entretanto o método não permite a obtenção direta de regras de decisão que permitam interpretar os dados de entrada para uma determinada solução. Uma alternativa é a utilização de algum método de extração de regras, no entanto, as regras obtidas normalmente vão utilizar operadores especiais e também muitas vezes não vão conseguir representar de maneira precisa a solução original obtida da RNA. Como exemplo, a seguir é apresentada uma das regras extraída de uma RNA treinada para realizar a classificação da planta Íris (AMORA *et al.*, 2009) utilizando o método de extração de regras indicado em (BENÍTEZ, CASTRO e REQUENA, 1997). Fica evidente a dificuldade na utilização das regras obtidas pela RNA treinada.

Regra: SE o *comprimento da sépala* não é maior que aproximadamente 4,6373 *i-OU largura da sépala* não é maior que aproximadamente 4,6375 *i-OU comprimento da pétala* não é maior que aproximadamente 4,6374 *i-OU largura da pétala* não é maior que aproximadamente 4,638 ENTÃO  $Y=10,3314$ .

A utilização de uma SRF para o diagnóstico pode muitas vezes também resultar na utilização de operadores especiais associados às regras, não permitindo a interpretação intuitiva das regras. Também muitas vezes as regras obtidas são em número elevado, dificultando a utilização. Por exemplo, em (CASTRO, 2004), a partir de uma RNA exemplo com duas entradas quaisquer  $x_1$  e  $x_2$ , dois neurônios na camada intermediária e uma saída  $y$ , e onde os valores dos pesos das ligações entre os neurônios e os valores de *bias* já são

previamente definidos, na conversão para um SRF foram obtidas 25 regras, como as listadas a seguir. Também fica claro nesse último exemplo a dificuldade da utilização de tais regras.

R<sub>1</sub>: SE ( $x_1$  é *extremamente pequeno*) E ( $x_2$  é *muito pequeno*) ENTÃO  $y_1 = 0,3266$

⋮

R<sub>25</sub>: SE ( $x_1$  é *alto*) E ( $x_2$  é *alto*) ENTÃO  $y_{25} = 0,0357$

## 2.7. Diagnóstico do Estado do Óleo Isolante

A caracterização de um óleo mineral para utilização como isolante é feita por uma série de parâmetros físico-químicos. Assim, o estabelecimento de limites para esses parâmetros tem como objetivo alcançar uma uniformidade de comportamento como isolante e referências de qualidade para sua utilização (ZIRBES, ROLIM e ZÜRN, 2005).

A análise das características físico-químicas, bem como seus valores limites, são indicativos importantes da boa adequação do óleo isolante de acordo com procedimentos padronizados por entidades tais como ABNT, IEC e ASTM. Através destas análises, pode-se acompanhar a evolução de algumas características e determinar se o óleo isolante mostra-se adequado. Estes atributos sofrem mudanças ao longo do tempo de operação do transformador, modificando assim a capacidade de transferir calor e a própria estabilidade térmica do óleo (BARBOSA, 2013).

Diversas características físico-químicas podem ser utilizadas para avaliar a condição do óleo isolante, entretanto as mais utilizadas são (BARBOSA, 2013): aparência, densidade máxima, viscosidade, ponto de fluidez, ponto de fulgor, índice de neutralização, tensão interfacial, cor, teor de água, rigidez dielétrica e fator de potência. Quando um óleo apresenta valores fora dos limites pré-estabelecidos, existe a necessidade de realização de uma regeneração ou mesmo a substituição do óleo mineral.

A verificação da condição do isolamento do óleo é de suma importância para garantir o funcionamento correto e seguro dos transformadores. Testes físico-químicos em amostras do óleo podem revelar a integridade dielétrica e refrigerante do mesmo e são atualmente aplicados na gestão de manutenção de transformadores de potência.

Para uma análise adequada do estado e qualidade do óleo alguns procedimentos são padronizados e seguem normas estabelecidas, conforme mostrado na Tabela 11 (IMAMURA,

SILVA, e SOUZA, 2000; CARGOL, 2005; MILASCH, 1984; THE ELECTRICITY FORUM, 2005; BARBOSA, 2013).

**Tabela 11 - Propriedades físico-químicas do óleo mineral e normas**

Propriedade Físico-Química do Óleo	Descrição	Norma Aplicada
Cor	A cor muda e escurece na medida em que o óleo vai se deteriorando. O óleo novo tem uma cor amarelo-pálida e é transparente. Útil na avaliação sobre o estado de oxidação do óleo isolante.	ABNT-MB – 351 ASTM D - 1500
Densidade Relativa	Qualquer variação no valor pode indicar alteração no óleo. A densidade do óleo está, normalmente, em torno de 0,9 na temperatura de 15 °C.	NBR – 7148 ASTM D - 1298
Viscosidade	É a resistência que o óleo oferece ao escoamento contínuo sem turbulência, inércia ou outras forças. A quantidade de calor que o óleo é capaz de transferir, por hora, do transformador para o meio ambiente depende desta propriedade.	NBR-10441 ABNT-MB - 293
Índice de Neutralização	Quantifica a presença de contaminantes polares ácidos, normalmente produtos de oxidação do óleo isolante.	NBR-14248 ASTM D-974
Tensão Interfacial	Na superfície de separação entre o óleo e a água forma-se uma força de atração entre as moléculas dos dois líquidos que é chamada de tensão interfacial, medida em N/m. Uma diminuição da tensão interfacial indica o início da deterioração do óleo.	NBR - 6234 ASTM D - 971
Teor de Água	A água pode existir no óleo sob a forma dissolvida, não dissolvida (em suspensão) ou livre (depositada). A quantidade de água em solução no óleo depende da temperatura e do grau de refino. Quanto mais alta a temperatura, maior a quantidade de água dissolvida no óleo. Quanto mais bem refinado for o óleo, menor será a solubilidade da água.	NBR - 5755 NBR - 10710 ASTM D - 1533
Rigidez Dielétrica	É a tensão alternada na qual ocorre a descarga disruptiva na camada de óleo situada entre dois eletrodos. A rigidez dielétrica do óleo é pouco afetada pela água nele dissolvida. Mas a água livre em suspensão no óleo diminui acentuadamente sua rigidez dielétrica. Em um óleo deteriorado, a água livre tem maior possibilidade de ficar em suspensão que no óleo novo. Também contribui para a redução da rigidez dielétrica as partículas sólidas em suspensão (fibras celulósicas, poeira, etc.).	IEC 156 NBR - 6869 NBR - 10859 ASTM D - 877 ASTM D - 1816
Fator de Potência	O fator de potência do óleo mineral isolante é igual ao cosseno do ângulo de fase ou o seno do ângulo de perdas do mesmo. Este aumenta de valor na medida em que a deterioração do óleo progride. O fator de potência significa o quanto de corrente flui pelo óleo e que é uma medida de sua contaminação e de sua deterioração.	IEC 247

Fonte: (BARBOSA, 2013).

Os resultados das análises físico-químicas indicam o real estado do óleo utilizado como isolante, possibilitando a orientação de procedimentos para regeneração ou até mesmo substituição do óleo.

A Tabela 12 apresenta um guia, fornecido por fabricante de transformadores, para a verificação das condições do óleo isolante mediante as análises físico-químicas, propiciando a obtenção de recomendações para manutenção de transformadores (WEG, 2004).

**Tabela 12 – Recomendações de verificação das condições do óleo isolante**

FP a 90° C (%) ou FP a 100° C (%) (Fator de Perdas dielétricas a 90° C ou 100° C)	Rigidez	Teor de Água	Acidez	TIF > 20 mN/m a 25° C (TIF - Tensão Interfacial)	Recomendações			
Atende	Atende	Atende	Atende	Atende	Nenhuma			
			Não atende	Não atende	Regeneração ou troca do óleo			
	Não atende	Atende	Atende	Atende	Atende	Regeneração ou troca do óleo e limpeza da parte ativa		
				Não atende	-	Filtragem do óleo		
		Não atende	Atende	Atende	Atende	Atende	Regeneração ou troca do óleo	
					Não atende	-	Regeneração ou troca do óleo	
			Não atende	Atende	Atende	Atende	Atende	Secagem da parte ativa e de óleo
						Não atende	Não atende	Secagem da parte ativa e de óleo e regeneração ou troca de óleo
Não atende	-	-	-	-	Secagem da parte ativa e de óleo e regeneração ou troca de óleo			
					Regeneração ou troca do óleo			

Fonte: (WEG, 2004).

## 2.8. Análise das Propriedades Físico-Químicas e Diagnóstico de Falhas em Transformadores

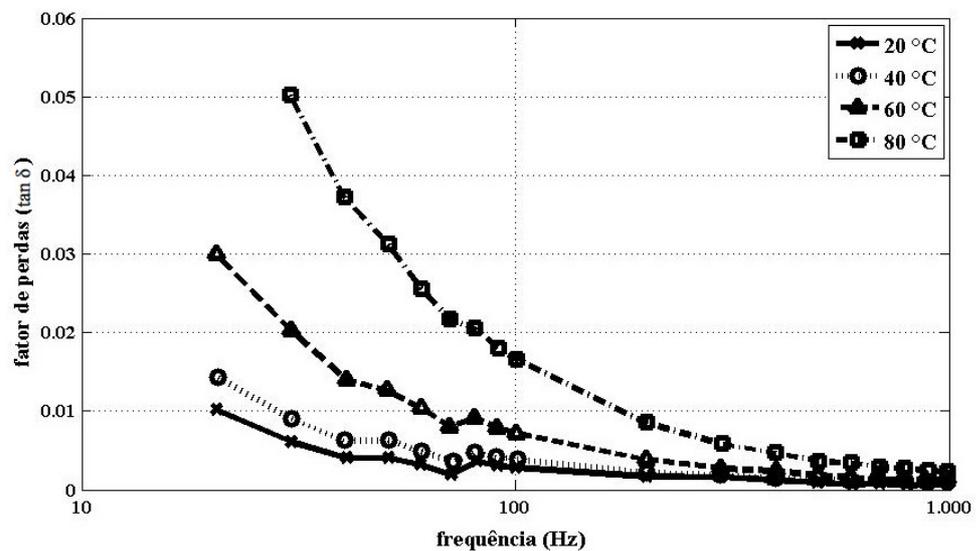
O óleo mineral isolante utilizado em transformadores de potência mantém contato com todas as partes ativas internas do transformador. As falhas que ocorrem no interior do transformador causam mudanças nas características do óleo e podem acelerar seu processo de envelhecimento e desgaste. Devido a isso, a verificação da integridade do óleo pode refletir as condições de funcionamento do equipamento e ainda detectar falhas que estejam ocorrendo no transformador (BARBOSA, 2013).

Apesar da importância das informações advindas das análises físico-químicas e cromatográficas e das possibilidades extraídas da existência de uma associação entre elas, esta associação não é muito bem definida na literatura. Porém, algumas pesquisas mencionam a influência das características dielétricas anormais do óleo ao aparecimento de falhas internas. Tais anormalidades podem estar refletidas na presença de radicais livres e de oxigênio dissolvido sob o efeito catalítico do cobre, reconhecidos como estopim do processo de

degradação do óleo na medida do seu envelhecimento (BARBOSA, 2013; FERGUSON, LOBEIRAS e SABOU, 2002).

A correlação entre envelhecimento do óleo e o fator de perdas ( $\tan \delta$ ) foram determinadas, segundo amostras de óleos ensaiadas com métodos de espectroscopia dielétrica (BARBOSA, 2013). Amostras com características físico-químicas alteradas apresentam fator de perdas dependente da temperatura, conforme Figura 8 (FERGUSON, LOBEIRAS e SABOU, 2002).

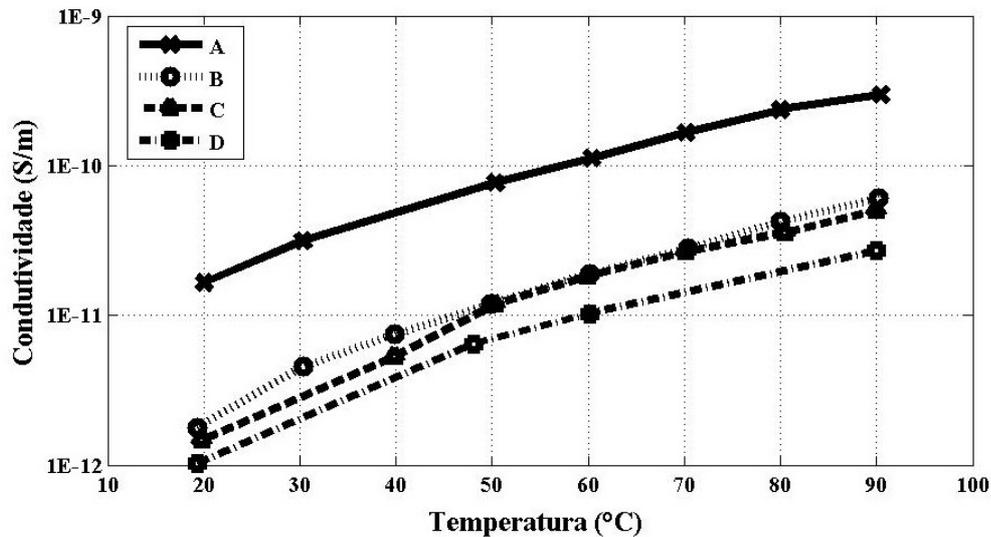
Figura 8 - Espectroscopia dielétrica de óleo isolante envelhecido



Fonte: Adaptado por (BARBOSA, 2013) de (FERGUSON, LOBEIRAS e SABOU, 2002).

Os estudos de (PEYRAQUE, BEROUAL e BURET, 1998) demonstram que a condutividade do óleo, que é um parâmetro complementar à rigidez dielétrica, mantém uma relação crescente com a temperatura. Os resultados são apresentados na Figura 9, utilizando quatro tipos de amostras de óleo: A – óleo fortemente envelhecido em laboratório; B – óleo com envelhecimento moderado; C – óleo em transformador novo e D – óleo novo.

Figura 9 - Variação da condutividade do óleo em função da temperatura



Fonte: Adaptado por (BARBOSA, 2013) de (FERGUSON, LOBEIRAS e SABOU, 2002).

Em (ABDI, BOUBAKEUR e HADDAD, 2008) são apresentados resultados de estudos que relacionam o envelhecimento do óleo isolante às modificações nas características físico-químicas como: tensão de ruptura, acidez, teor de água, cor e viscosidade cinemática. Os resultados mostraram importante redução da tensão de ruptura do dielétrico a temperaturas elevadas de sobreaquecimento. Sobre a acidez, o envelhecimento térmico conduziu a aumento no grau de acidez. O teor de água também sofreu incremento em seus valores com o aumento da temperatura. Houve também alteração na cor do óleo testado após o envelhecimento controlado, provavelmente devido à oxidação que pode levar a formação de produtos ácidos. Em relação à viscosidade cinemática, não foram observadas alterações significativas durante o processo de envelhecimento controlado do óleo.

De forma geral, os estudos de (ABDI, BOUBAKEUR e HADDAD, 2008) concluem que leves sobreaquecimentos não levam a alterações muito significativas no óleo isolante. Entretanto, é verificada a influência do envelhecimento por efeito térmico nas propriedades físico-químicas do óleo (BARBOSA, 2013).

Na Tabela 13 são apresentados os resultados dos ensaios físico-químicos efetuados por (PARASKEVAS, VASSILIOU e DERVOS, 2006) em amostras classificadas segundo as condições de óleo: 1 - condição satisfatória para uso continuado; 2 - necessário recondicionar para eliminar efeitos do envelhecimento, garantindo serviço prolongado e confiável; 3 - condição insuficiente para uso, podendo ser descartada. Baseado nos resultados destes testes,

algumas conclusões podem ser evidenciadas acerca da relação entre as grandezas físico-químicas ensaiadas e o estado do óleo.

**Tabela 13 - Resultados dos ensaios físico-químicos comparados aos valores normatizados**

Ensaio físico-químico	Limites (ASTM)	Condição do óleo		
		1	2	3
Rigidez Dielétrica (kV)	26 mín. (D877)	31,48±5,9	28,70±6,9	27,49±7,24
Tensão Interfacial (mN/m)	24 mín. (D971)	34,60±6,18	24,33±3,8	20,13±3,65
Acidez (mg KOH / g óleo)	0,2 máx. (D974)	0,10±0,016	0,124±0,048	0,19±0,091
Teor de Água (p.p.m.)	35 máx. (D1533)	13,4±7,5	14,6±7,7	17,4±11,0
Densidade Relativa	0,840-0,9 (D1298)	0,8718±0,0011	0,8720±0,0087	0,8752±0,0098
Cor	0,5-8 (D1500)	1,37±0,56	2,13±0,71	3,19±0,83

Fonte: (PARASKEVAS, VASSILIOU e DERVOS, 2006).

Analisando a Tabela 13, algumas observações podem ser feitas (BARBOSA, 2013): apesar do patamar de normalidade padronizado não ter sido ultrapassado pelos conjuntos de dados, há uma tendência de queda no valor da rigidez dielétrica à medida que aumenta o valor do fator de dissipação e a instabilidade térmica da permissividade; também é observada uma tendência de decréscimo da tensão interfacial em direção às amostras de condição deteriorada, as quais apresentaram valores reprováveis; o índice de acidez cresce em proporção direta ao fator de perdas e à instabilidade térmica, ou seja, amostras na condição 3 apresentam níveis de acidez elevados em comparação com as outras condições de amostras, muito próximas do limiar da norma; a densidade relativa também apresenta comportamento de crescimento com relação ao aumento do valor de  $\tan \delta$ ; e o índice de cor também sofre incremento em relação à deterioração do óleo, porém, esta medida não é totalmente confiável como indicador da qualidade do óleo.

Importante também observar que mesmo dentro da faixa de normalidade sugerida, as amostras da condição 3 apresentam alta taxa de elevação em comparação com as outras duas condições (BARBOSA, 2013).

Outro ponto que merece enfoque é a influência do envelhecimento do óleo sobre a degradação das características dielétricas do óleo. Ou seja, quanto maior o tempo de operação do isolante, menor a probabilidade da amostra apresentar boa qualidade dielétrica. Podendo a estabilidade térmica do óleo, inclusive, ser agravada por decorrência de oxidação e esforço

elétrico, influenciando no acréscimo do fator de dissipação do óleo (FERGUSON, LOBEIRAS e SABOU, 2002).

Sabe-se que o óleo isolante tende a ter má qualidade com o decréscimo da rigidez dielétrica e da tensão interfacial e com o aumento do teor de água e da acidez (PALMER *et al.*, 2000).

Portanto, diversos estudos apontam uma possível correlação entre o diagnóstico de falhas incipientes em transformadores de potência e os valores das análises físico-químicas do óleo isolante, porém, ainda não explorada profundamente.

Em (BARBOSA, 2013) são apresentados os resultados da utilização de uma RNA para o diagnóstico de falhas incipientes em um transformador de potência a partir de dados físico-químicos, com resultados promissores. No estudo realizado foram utilizadas 135 amostras, com 94 dados sendo utilizados para o treino e 41 para a validação. Como entradas para a RNA foram utilizados os seguintes dados físico-químicos: acidez, rigidez dielétrica, teor de água, tensão interfacial, densidade e fator de perdas dielétricas a 100° C. Na saída, a RNA poder indicar os seguintes diagnósticos: estado normal, falha térmica ou falha elétrica.

Adotando várias rodadas de simulação para o treino e validação da RNA que realiza o diagnóstico de transformadores utilizando dados físico-químicos, nos melhores casos de simulação o índice de acerto chegou a 100% tanto para o conjunto de treino como para os dados de validação, para as piores rodadas de simulação o acerto foi de 100% para o treino e de 80,49% para o conjunto de validação.

Utilizando a metodologia proposta neste trabalho, uma árvore de decisão com a dimensão aumentada dos dados de entrada, e os dados de (BARBOSA, 2013) foram obtidas regras de decisão que permitem o diagnóstico simplificado de falhas incipientes em equipamentos com isolamento à óleo, como os transformadores. Esses resultados são apresentados no Capítulo 5.

## **2.9. Considerações Parciais**

Nesse capítulo foram discutidos os tipos de falhas que podem ocorrer em um transformador com isolamento a óleo. Também foram apresentados os fenômenos de

formação de gases no interior do transformador quando na ocorrência de falhas, e que podem servir de indicativo da ocorrência de tais falhas. A questão da possibilidade da existência de erros de medidas também foi comentada, tais erros podem impactar a precisão no diagnóstico de falhas incipientes. Os métodos de diagnóstico de falhas incipientes adotados nas normas foram citados. As imprecisões e incertezas relacionadas à utilização de tais métodos clássicos de diagnóstico são também comentadas. Foi discutida a utilização de técnicas de ML para aprimorar o diagnóstico de transformadores e os problemas específicos que podem surgir do uso de tais técnicas. Ao final do capítulo, foram apresentadas informações quanto ao uso de dados físico-químicos para avaliar a qualidade do óleo isolante, e a possível correlação que pode existir entre esses dados e o diagnóstico de falhas incipientes em transformadores de potência.

Nesse trabalho para superar os problemas de eficiência no diagnóstico de falhas incipientes em transformadores de potência apresentados pelos métodos clássicos, foi utilizada também uma técnica de ML, no caso uma DT com modificações no espaço de entradas do problema. Isso permitiu a obtenção de regras simples, em número reduzido, e com uma eficiência maior que os métodos indicados na norma. A utilização de uma DT será discutida com mais detalhes no próximo capítulo. E as modificações no espaço de entradas propostas neste trabalho para o uso da DT serão discutidas com mais detalhes no Capítulo 4.

O uso de DT permite a obtenção de regras interpretáveis, em contraste com o uso de métodos como RNA e SRF para o diagnóstico de transformadores, que não permitem normalmente a obtenção de regras de diagnóstico interpretáveis e de fácil utilização.

O método de diagnóstico de transformadores desenvolvido nessa Tese também será avaliado na questão da possibilidade de ocorrência de erros de medição dos gases. Essas avaliações de robustez serão apresentadas no Capítulo 5.

Também no Capítulo 5, serão apresentadas e avaliadas regras de diagnóstico simplificado de falhas em equipamentos com isolamento a óleo utilizando dados físico-químicos.

### 3. ÁRVORES DE DECISÃO

Métodos de classificação associam um rótulo ou identificador de classe a um conjunto de dados (THEODORIDIS e KOUTROUMBAS, 2009). A associação ocorre em função das características do objeto a ser classificado. Uma árvore de decisão (*Decision Tree* - DT) (ROKACH e MAIMON, 2008; RUSSELL e NORVIG, 2004; DUDA, HART e STORK, 2001; QUINLAN, 1993; e QUINLAN, 1986) é considerada um dos métodos de classificação mais tradicionais, que apresenta a vantagem de ser um método não paramétrico e também permitir a obtenção direta de regras de decisão, facilitando a compressão dos resultados obtidos.

Na seção 3.1 são apresentadas definições básicas associadas a métodos de *data mining* como a DT. As características gerais de uma DT são apresentadas na seção 3.2. As etapas para a construção de uma DT e o algoritmo básico de construção são comentadas nas seções 3.3 e 3.4, respectivamente. Na seção 3.5 é apresentado o método C4.5 de construção de uma DT, que utiliza um processo de construção de nós segundo um critério de otimização e, após a formação inicial da DT, um processo de poda, de forma a eliminar partes menos importantes da árvore, aumentando a capacidade de generalização da DT desenvolvida. As características de uma DT oblíqua são discutidas seção 3.6. Na seção 3.7 são citadas as vantagens e desvantagens da utilização de uma DT de classificação. Finalmente, na seção 3.8 é apresentada a conclusão do capítulo.

#### 3.1. Definições

*Data mining* (mineração de dados) é a ciência e tecnologia associada à exploração de conjuntos grandes e complexos de dados, a fim de descobrir padrões úteis (ROKACH e MAIMON, 2008). Pesquisadores estão continuamente à procura de melhores técnicas para tornar o processo de *data mining* mais eficiente e preciso. Uma das abordagens mais promissoras e mais populares é o uso de DT. As árvores de decisão são técnicas simples, mas bem-sucedidas para prever e explicar a relação entre algumas medições sobre um item e seu valor-alvo. Além de seu uso em mineração, uma DT é uma ferramenta altamente eficaz em outras áreas, como a identificação de texto, extração de informações, aprendizagem de máquina e reconhecimento de padrões.

Vários métodos de *data mining* como, por exemplo, uma DT são denominados métodos supervisionados, que representam métodos que tentam descobrir a relação entre os atributos de entrada (às vezes chamados de variáveis independentes) e um atributo de destino (por vezes referido como uma variável dependente). A relação que é descoberta está representada na estrutura referida como um modelo. Usualmente os modelos podem descrever e explicar fenômenos que estão escondidos no conjunto de dados, e que pode ser usado para prever o valor do atributo alvo quando os valores dos atributos de entrada são conhecidos (ROKACH E MAIMON, 2008).

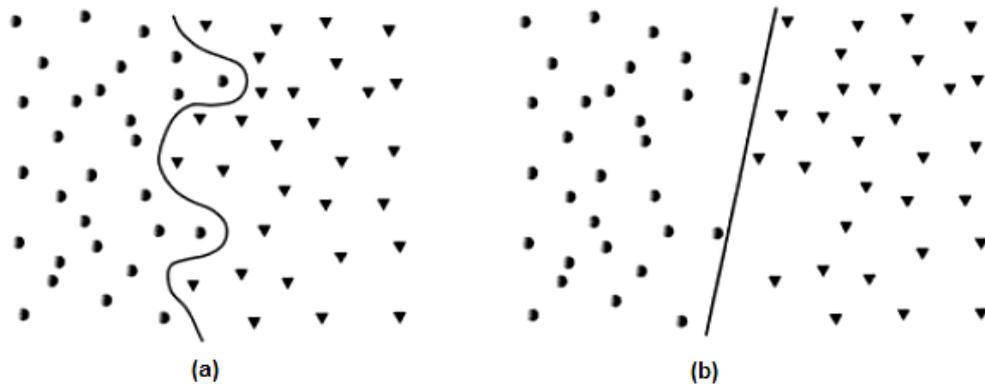
Os métodos não supervisionados tentam realizar a descoberta de relações entre os dados sem a utilização de variáveis dependentes, ou seja, sem a utilização de valores de saída ou rótulos pré-estabelecidos (KOHAVI e PROVOST, 1998).

Os modelos supervisionados podem ser divididos em modelos de classificação e modelos de regressão. Os modelos de regressão promovem o mapeamento do espaço de entrada em um domínio de valor real. Por exemplo, um modelo de regressão pode prever a procura (quantidade) de um determinado produto, dadas as suas características. Por outro lado, os classificadores promovem o mapeamento do espaço de entrada em classes pré-definidas. Por exemplo, os classificadores podem ser usados para classificar os credores de hipotecas como bons (hipoteca paga no tempo estipulado) e ruins (atraso de pagamento). Entre as muitas alternativas para representar classificadores, há, por exemplo, MVS, DT, função algébrica, e outros (ROKACH E MAIMON, 2008).

Durante o processo de classificação, os métodos podem realizar a separação dos dados de entrada através da determinação de uma fronteira representada por uma reta, situação onde os dados serão considerados linearmente separáveis (caso *a* na Figura 10). Caso a fronteira não possa ser representada por uma reta, os dados serão considerados como não linearmente separáveis e a fronteira será representada por curvas (caso *b* na Figura 10) na dimensão normal do problema.

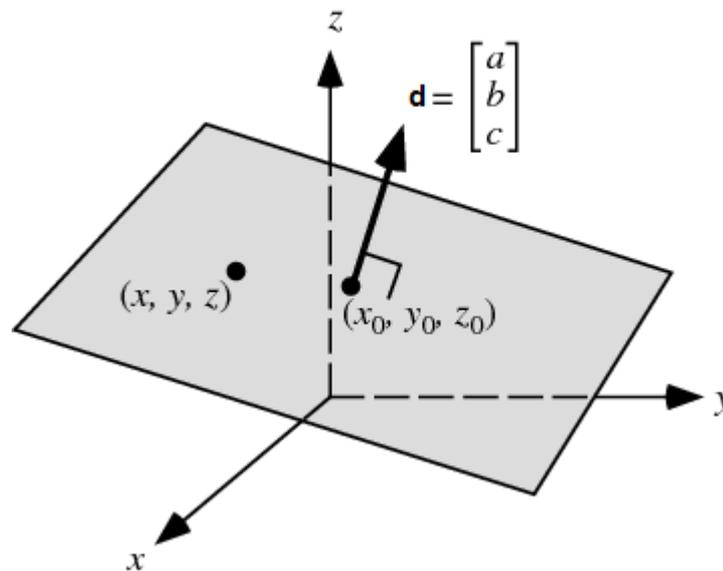
Na definição de separação linear, foi utilizado o conceito de reta, mas em problemas de dimensões maiores, o conceito de reta pode ser generalizado para um hiperplano. Um hiperplano é a generalização do plano em diferentes números de dimensões. Na geometria, um hiperplano pode ser um espaço vetorial, transformação afim ou o subespaço de dimensão  $d-1$ . Em particular, num espaço tridimensional um hiperplano é um plano habitual (Figura 11). Num espaço bidimensional, um hiperplano é uma reta. Num espaço unidimensional, um hiperplano é um ponto (PALHARES, 2011).

Figura 10 - Exemplos de separação não linear (a) e linear (b) do conjunto de dados



Fonte: Adaptado de (ROKACH E MAIMON, 2008) pelo autor.

Figura 11 - Exemplo de um plano a partir de um sistema em três dimensões



Fonte: Adaptado de (WEISSTEIN, 2013) pelo autor.

### 3.2. Introdução à Árvore de Decisão

Uma DT é um modelo preditivo (é capaz de aprender o mapeamento entre um conjunto de variáveis de entrada de dados e uma variável de resposta ou de destino) que pode ser utilizado para representar tanto um modelo de classificação como também um modelo de regressão (ROKACH e MAIMON, 2008). Quando uma DT é utilizada como classificador é normalmente denominada Árvore de Classificação (AC), e quando utilizada para regressão como Árvore de Regressão (AR).

Uma AC é utilizada para classificar um objeto ou instância dentro de um conjunto pré-definido de classes, baseados nos atributos da instância. As árvores de classificação são frequentemente utilizadas em problemas nas áreas de finança, marketing, engenharia e medicina (ROKACH e MAIMON, 2008).

Uma AC representa um sistema de decisão multiestágios onde as classes são sequencialmente rejeitadas até ser alcançada uma classe final de aceitação, durante a apresentação de uma instância (vetor de atributos) a ser classificada. No final, o espaço de entrada é dividido em regiões distintas, correspondendo às classes, de maneira sequencial. Durante a apresentação de um vetor para classificação, a pesquisa da região a ser associada a um parâmetro do vetor é obtida através da pesquisa de uma sequência de decisões ao longo de um caminho de nós, numa árvore apropriadamente construída (THEODORIDIS e KOUTROUMBAS, 2009).

Uma DT típica separa o espaço de entradas em hiperplanos com retas paralelas aos eixos, como exemplificado na Figura 12. A sequência de decisões é aplicada para cada atributo da instância apresentada à árvore, com os testes de decisão associados aos nós sendo na forma:

$$\text{Se } a_i \theta \alpha \text{ Então } c_1 \text{ Senão } c_2 \quad (3.1)$$

onde:  $a_i$  representa o atributo avaliado;

$\theta$  a operação lógica testada ( $=, \neq, \leq, \geq, <, >$ );

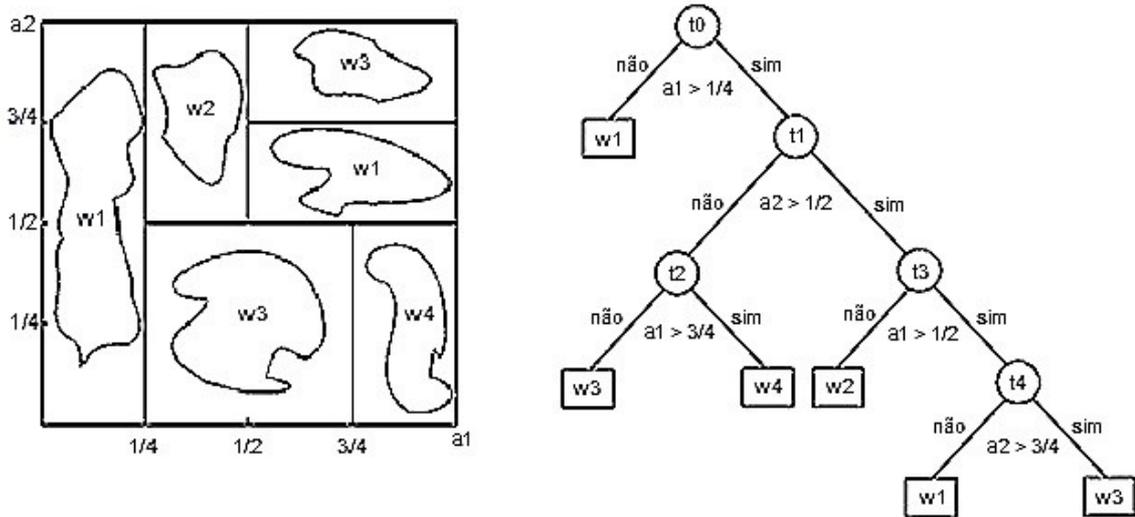
$\alpha$  é um valor limite; e

$c_1$  e  $c_2$  representam “caminhos” distintos na árvore que levam a outros nós na árvore que podem representar um outro nó de teste ou então um nó de “folha” que representa uma classe de classificação.

Uma árvore com nós descritos pela relação (3.1) é também denominada de Árvore de Classificação Binária (ACB), pois existem apenas dois caminhos que podem derivar de cada nó de teste. Na Figura 12, além da partição do espaço, também é apresentada a ACB obtida, indicando as regras de classificação.

A seguir são descritos as etapas na construção de uma ACB. No restante do capítulo e do documento uma ACB, que representa o tipo mais comum de árvore, será apenas denominada como DT.

**Figura 12 - Partição do espaço de variáveis e regras obtidas de uma árvore de classificação binária**



Fonte: Adaptado de (THEODORIDIS e KOUTROUMBAS, 2009).

### 3.3. Etapas de Construção de uma Árvore de Decisão

Num caso geral, para desenvolver uma DT, devem ser considerados os seguintes elementos na fase de treinamento (THEODORIDIS e KOUTROUMBAS, 2009; ROKACH e MAIMON, 2008):

- Durante a construção da árvore, um atributo deve ser testado na construção do nó de divisão interno, gerando uma divisão binária, em dois nós descendentes como pode ser observado na Figura 12. Para cada nó,  $t$ , é associado com um subconjunto  $X_t$  específico do conjunto de treinamento  $X$ . O processo de abertura de um nó de teste é equivalente a dividir  $X_t$  em dois subconjuntos disjuntos descendentes,  $X_{tS}$ ,  $X_{tN}$ . O primeiro subconjunto consiste dos vetores em  $X_t$  que correspondem à resposta positiva da questão lógica válida, e o segundo subconjunto corresponde à resposta negativa. O primeiro nó da árvore é associado com o conjunto total de treinamento  $X$ . Para cada divisão, as seguintes relações devem ser verdadeiras:

$$X_{tS} \cap X_{tN} = 0 \quad (3.2)$$

$$X_{tS} \cup X_{tN} = X_t \quad (3.3)$$

- Um critério de abertura deve ser adotado de forma a permitir a melhor divisão dos atributos candidatos durante a formação dos nós.
- Um critério de parada na formação de nós, e, portanto, no crescimento da árvore, deve ser utilizado. Os nós terminais então definem as classes de classificação para os dados e são denominados de nós folhas.
- Também deve ser adotado um critério de poda de modo a diminuir o tamanho da árvore, retirando nós folhas com menor impacto na precisão da árvore.
- A partir da árvore formada, as regras podem ser extraídas, associando os atributos com cada nó folha (classe).

Na próxima seção, vários destes pontos, que formam um algoritmo completo de construção de uma DT, serão comentados com mais detalhes.

### **3.4. Estrutura Básica de um Algoritmo de Construção de uma Árvore de Decisão**

Métodos de indução de uma DT são algoritmos que automaticamente constroem a árvore a partir de conjunto fornecido de dados. Tendo, tipicamente, como objetivo reduzir o erro na generalização da classificação dos dados, ou seja, na classificação de novos dados após a construção da DT utilizado dados de treinamento. Entretanto, outros objetivos também podem ser considerados, como: minimizar o número de nós ou minimizar a altura (profundidade) da árvore (ROKACH e MAIMON, 2008).

O problema de obtenção de uma DT ótima a partir de um conjunto de dados é considerado um problema NP-hard<sup>2</sup> (HANCOCK *et al.*, 1996; HYAFIL e RIVEST, 1976). Em (NAUMOV, 1991) é afirmado que somente é realizável a obtenção de uma DT ótima em problemas pequenos, com pequenas quantidades de dados. Portanto, métodos heurísticos devem ser utilizados para resolver o problema de construção de uma DT.

---

<sup>2</sup> Representa a classe de complexidade de problemas de decisão que são intrinsecamente mais difíceis do que as que podem ser resolvidos por uma máquina de Turing não determinística em um tempo polinomial (ATALLAH, 1999).

Na Figura 13 é apresentado um algoritmo típico para indução de uma DT para classificação de dados. As entradas do algoritmo são  $X$  (conjunto de treinamento),  $A$  (atributos do conjunto de entrada) e  $y$  (valor ou classe esperado do atributo).

**Figura 13 - Algoritmo para construção de uma DT**

```

Procedimento ArvoreIndutor( $X,A,y$ )
     $T=ArvoreCrescimento(X,A,y)$ 
    Return ArvoreCrescimento( $X,A,y$ )
Procedimento ArvoreCrescimento( $X,A,y$ )
    Criar uma árvore  $T$ 
    Se um dos Critérios de Parada é alcançado Então
        Faça o nó raiz em  $T$  como um nó folha associado ao valor mais comum de  $y$  em
         $X$ , como classe.
    Senão
        Encontre uma função discreta  $f(A)$  dos valores dos atributos de entrada que
        divida  $X$  conforme uma métrica de divisão adotada.
        Se métrica de divisão  $\geq$  Limiar Então
            Rótulo do nó raiz em  $T$  é  $f(A)$ .
            Para cada resultado  $v_i$  de  $f(A)$  faça
                 $Sub-arvore_i = ArvoreCrescimento(\sigma_{f(A)=v_i}X,A,y)$ .
                Conecte o nó raiz de  $T$  a Subarvore $_i$ , com uma aresta que é
                denominada como  $v_i$ .
            Fim do para
        Senão
            Marque o nó raiz em  $T$  como uma folha com o valor mais comum de  $y$ 
            em  $X$  como a classe.
        Fim do se
    Fim do se
    Retorna  $T$ 
Procedimento ArvorePoda( $X,T,y$ )
    Repita até
        Selecione um nó  $t$  em  $T$  onde a poda é máxima segundo o critério de evolução.
        Se  $t \neq 0$  Então
             $T=Poda(T,t)$ 
        Fim do se
    Até t=0.
    Retorna  $T$ 

```

Deve ser observado que os algoritmos de construção de árvores são considerados do tipo gulosos<sup>3</sup> (*greedy*) devido à natureza recursiva destes métodos, utilizando uma abordagem de dividir e conquistar. Em cada iteração, o algoritmo considera a partição do conjunto de treinamento utilizando uma função discreta dos atributos de entrada. A seleção da função mais apropriada é feita de acordo com alguma medida para o processo de criação dos nós. Após a divisão de um nó avaliado, o conjunto de dados é dividido em dois subconjuntos menores, o processo continua até que a divisão não retorne nenhum ganho na medida de abertura considerada ou se algum critério de parada for satisfeito.

Uma alternativa é relaxar os critérios de parada na construção da DT, e aplicar um método de poda após a construção da árvore, de forma a reduzir o tamanho dessa árvore, eliminando nós e ramos com menor impacto na precisão de classificação realizada pela árvore. A ideia de poda foi originalmente proposta por (BREIMAN *et al.*, 1984). Estudos demonstram que a poda permite a melhoria na capacidade de generalização da árvore, mesmo em situações de ruído nos dados (ROKACH e MAIMON, 2005).

Os métodos de poda permitem a obtenção de uma DT com uma mescla de precisão em relação ao treino e uma forma compacta, que por sua vez ajuda na generalização e interpretação da árvore.

Existem vários métodos de poda propostos na literatura (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008), considerando um determinado critério para aperfeiçoar a poda da DT. Esses critérios de poda podem ser utilizados associados ao algoritmo básico descrito na Figura 13, apenas adotando as equações e indicações correspondentes na sub-rotina Poda chamada no procedimento ArvorePoda. O procedimento ArvorePoda é executado após a formação da DT pela rotina ArvoreCrescimento.

A seguir será comentado com mais detalhes o algoritmo C4.5 utilizado nesse trabalho para o desenvolvimento de uma DT. Maiores informações sobre outros algoritmos de construção de uma DT, e os critérios de construção de nós e de poda adotados por tais algoritmos, são apresentadas no Apêndice B.

---

<sup>3</sup> Algoritmo que realiza uma busca local para a solução de um problema (BLACK, 2005).

### 3.5. Método C4.5 para Construção de uma Árvore de Decisão

Um dos algoritmos mais utilizados para construção de uma DT é o método C4.5 (QUINLAN, 1993). O método C4.5 é uma extensão do algoritmo ID3 (QUINLAN, 1986), com vantagens significativas (ROKACH e MAIMON, 2008): trata tanto com atributos categóricos (ordinais ou não) e com atributos contínuos, pode trabalhar com valores desconhecidos, utiliza a medida da relação de ganho (Apêndice B) para selecionar o atributo que melhor divide os nós na construção da árvore, pode lidar com problemas em que os atributos têm custos diferentes, e apresenta um método de poda para a árvore gerada, a Poda Baseada no Erro (Apêndice B).

Algumas premissas guiam a utilização do algoritmo para a construção de uma DT:

- Se todos os casos do treino são de uma mesma classe, a árvore terá um único nó folha, com o rótulo dessa classe;
- Para cada atributo, deve ser calculado o potencial de informação, considerando esse atributo para o teste lógico associado a um nó intermediário a ser construído na árvore. Escolhendo o atributo que resulta num maior ganho de informação e realizando a ramificação da árvore.
- Caso o nó não possa ser mais dividido, é considerado como sendo um nó folha, indicando a classe predominante.
- O processo continua até que todos os vetores de entrada sejam classificados pelos nós folha da árvore construída durante o processo.

O cálculo do ganho de informação, durante a construção dos nós, usa o conceito de "entropia", isto é, a medida da desordem dos dados. Considerando  $P(\omega_i|t)$  como a probabilidade que um vetor do subconjunto  $X_t$ , associado com um nó  $t$ , divida a classe  $w_i$ , com  $i = 1, 2, \dots, M$ ; a entropia  $E$  associada ao nó  $t$  é calculado por (THEODORIDIS e KOUTROUMBAS, 2009):

$$E(t) = -\sum_{i=1}^M P(\omega_i|t) \log_2 P(\omega_i|t) \quad (3.4)$$

e finalmente, o ganho é definido por (THEODORIDIS e KOUTROUMBAS, 2009):

$$G(t) = E(t) - \frac{N_{tS}}{N_t} \cdot E(t_S) - \frac{N_{tN}}{N_t} \cdot E(t_N) \quad (3.5)$$

onde  $N_t$  é o número de pontos em  $t$ ;  $N_{tS}$  representa o número de pontos que são separados e enviados para o nó descendente que representa a resposta *sim* para o teste lógico no nó original  $t$ ;  $N_{tN}$  representa os pontos associados a resposta *não* do teste lógico; também  $E(t_S)$  e  $E(t_N)$  são os valores de entropia para os nós descendentes  $t_S$  e  $t_N$ , respectivamente.

A entropia é usada para determinar qual o nó seguinte a ser formado e a divisão dos ramos, durante a construção da árvore. O objetivo é maximizar o ganho de informação  $G$ .

Após a criação completa da árvore, que deve classificar todas as instâncias no conjunto de treinamento, a mesma é podada. Isso é realizado para reduzir os erros de classificação causados pela especialização no conjunto de treinamento, aumentando a capacidade de generalização da árvore (MARTIN, 1997). No processo de poda, o algoritmo C4.5 realiza uma busca na árvore, de baixo para cima, e retira os nós folhas que não representam ganho significativo na redução do erro de classificação.

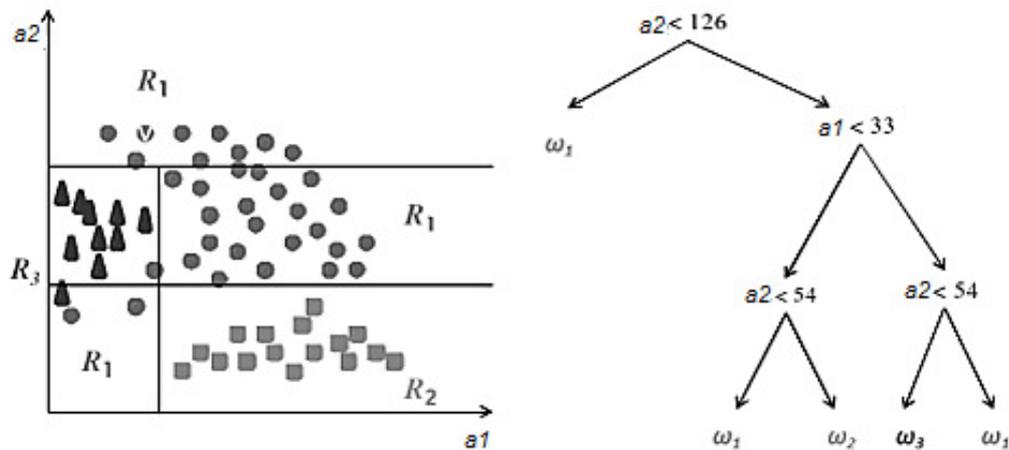
O método C4.5 é muito popular, e foi adotado nas simulações realizadas nessa pesquisa na obtenção de regras para o diagnóstico de falhas incipientes em transformadores de potência. As simulações foram realizadas através do *software* livre SIPINA, que além do C4.5 permite a utilização de outros métodos de construção árvores e de mineração de dados.

### 3.6. Árvores de Decisão Oblíquas

Uma DT oblíqua representa uma árvore com análise multivariável nos nós de decisão, ao contrário da forma convencional de construção e uso de uma DT com apenas análise univariável nos nós.

A Figura 14 ilustra o caso de uma árvore de decisão univariável, suas regras de classificação (ou testes) e a respectiva representação da classificação dos dados em um espaço de atributos. Este é o caso mais simples de DT, sendo chamada de univariável em função de suas regras estarem associadas somente a um único atributo (CASTRO FILHO, 2010).

Figura 14 - Árvore de decisão univariável



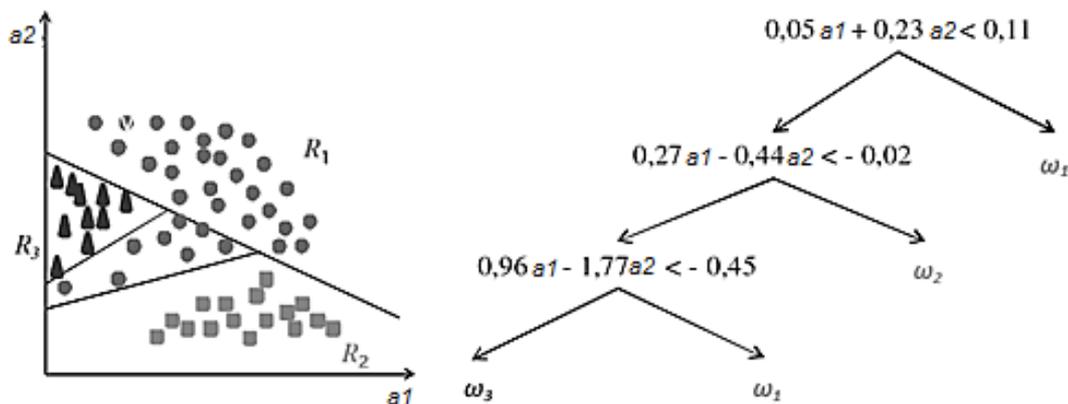
Fonte: Adaptado de (DUDA, HART e STORK , 2001) por (CASTRO FILHO, 2010).

A regra geral da DT univariável apresentada na Figura 14 é dada pela equação (3.6). Nesse tipo de árvore, cada reta que divide os dados é perpendicular ao respectivo eixo de atributo  $a_j$ , interceptando-o no valor  $\omega_0$  (CASTRO FILHO, 2010):

$$f(a / j, \omega_0) = a_j + \omega_0 \quad (3.6)$$

Já na Figura 15 a DT é chamada de oblíqua, ou multivariada linear (YILDIZ e ALPAYDIN, 2005), porque os testes são equivalentes a hiperplanos com orientação oblíqua aos eixos cartesianos.

Figura 15 - Árvore de decisão oblíqua



Fonte: Adaptado de (DUDA, HART e STORK , 2001) por (CASTRO FILHO, 2010).

Na respectiva representação da classificação no espaço dos atributos  $x_1$  e  $x_2$  observa-se nas Figuras 14 e 15 uma divisão das classes através de retas (ou hiperplanos, em casos de dimensões maiores) cujas equações são testes matemáticos. Caso um dado observado tenha

um valor verdadeiro ao do teste ele será conduzido para um dos “galhos” da árvore, caso contrário será conduzido para o outro. Ao término do processo, ao chegar em uma “folha”, aquele dado será associado à uma classe. Tanto no caso da Figura 14 como no da Figura 15 as árvores de decisão podem ser classificadas como binárias. No caso de árvores binárias, cada nó possui como saída somente dois galhos seguindo uma função booleana do tipo  $f_m(a|\Phi) > 0$  com duas saídas possíveis, verdadeiro ou falso (YILDIZ e ALPAYDIN, 2005). Esta função é parametrizada por  $\Phi$ , com duas saídas, verdadeiro ou falso, referentes aos galhos da direita e da esquerda. Cada  $f_m(a)$  define uma função discriminante e uma respectiva forma de modelo. No caso oblíquo o modelo é dado por (CASTRO FILHO, 2010):

$$f(a / \omega_m, \omega_{m0}) = \sum_{j=1}^d \omega_{mj} a_j + \omega_{m0} = \omega_m^T a + \omega_{m0} \quad (3.7)$$

sendo  $\omega_m$  a direção do hiperplano divisor ( $d$ -dimensional) e  $\omega_{m0}$  o valor limitante quando  $a$  está projetado sobre  $\omega_m$ .

Nos critérios multivariáveis diversos atributos devem ser avaliados durante a abertura/divisão dos nós em uma DT. O problema de construção da árvore agora se torna mais complexo. Também, o uso desses critérios pode afetar significamente o rendimento da DT. Por esses motivos, as árvores oblíquas acabam não sendo muito utilizadas. No Apêndice B são citados alguns critérios de construção de nós de uma DT com análise multivariável, a maioria desses critérios é baseada na combinação linear dos atributos de entrada.

### 3.7. Vantagens e Desvantagens no Uso de uma Árvore de Decisão

Podem ser citadas como vantagens da utilização de uma DT como ferramenta de classificação (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008):

- Uma DT é autoexplicativa, pois pode ser convertida para um conjunto de regras.
- São capazes de lidar com valores nominais e numéricos nos atributos de entrada.
- A representação de um problema através de uma DT é rica o suficiente para classificar qualquer valor discreto.

- São capazes de lidar com dados com ruído (erros de leitura na entrada).
- As árvores de decisão são consideradas como métodos não paramétricos. Isto significa que as árvores não assumem hipóteses sobre a distribuição espacial dos dados ou sobre a estrutura do classificador.

As principais desvantagens que podem ser citadas são (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008):

- A maioria dos algoritmos de construção de uma DT, como o ID3 e o C4.5, exigem que o atributo de destino (valor de saída da classe) tenha apenas valores discretos.
- Como os métodos de indução de uma DT utilizam a técnica gulosa de "dividir e conquistar", a mesma vai apresentar um bom comportamento se atributos muito relevantes fizerem parte dos vetores de entrada, mas a eficiência será menor se muitas interações complexas existirem entre esses atributos.
- A característica gulosa de uma DT também apresenta outra desvantagem importante: sua sensibilidade excessiva ao conjunto de treino para atributos irrelevantes e para uma quantidade significativa de ruído.
- O conjunto de regras obtido pode ser grande, prejudicando a sua utilização e capacidade de generalização.

Vale ressaltar também, que as regras de conhecimento obtidas por meio de uma DT podem apresentar problemas de generalização, caso cuidados não sejam tomados durante o treino e teste da árvore. A utilização de técnicas de poda favorece a capacidade de generalização da árvore, e, portanto, das regras obtidas.

Também, as regras extraídas são baseadas em fronteiras construídas de forma paralela aos eixos do problema, que dependendo da característica específica de um problema modelado pode resultar numa solução ineficiente, isto ficou evidenciado na seção 3.6, quando foi discutida a questão da árvore oblíqua, que constrói uma solução não necessariamente paralela aos eixos, e que em muitas situações será mais eficiente.

Nesse trabalho é proposta uma alternativa para obter uma solução mais eficiente de uma DT pela modificação dos atributos de entrada, acrescentando novos atributos e/ou relações dos já existentes. Portanto, aumentando a dimensão do problema, na expectativa de ser construída uma fronteira melhor de separação entre as classes. Essa nova metodologia será discutida com detalhes no próximo capítulo, junto com exemplos.

### **3.8. Considerações Parciais**

Nesse capítulo foram apresentadas as características, vantagens e desvantagens das árvores de decisão. Foi dado destaque ao algoritmo C4.5 de construção de uma DT, comentando características e apresentando equações relacionadas ao algoritmo.

As características das árvores oblíquas também foram citadas, sendo apresentada a sua maior vantagem que é a possibilidade da construção de uma superfície de separação linear, mesmo quando trabalhando com um conjunto complexo de dados, e, portanto, gerando um número menor de regras. Entretanto, as regras obtidas serão multivariáveis, dificultando a interpretação intuitiva de tais regras.

Aproveitando a capacidade de uma DT em gerar diretamente regras interpretáveis, a mesma foi utilizada nesse trabalho para o diagnóstico de falhas incipientes em transformadores de potência. De forma a contornar os problemas relacionados com a obtenção de um número elevado de regras pela árvore, são propostas nessa pesquisa modificações nos atributos de entrada de forma a aumentar inicialmente a dimensão do problema e deixar a árvore selecionar os atributos mais importantes. Esse procedimento será comentado no próximo capítulo.

## **4. CLASSIFICAÇÃO UTILIZANDO ÁRVORE DE DECISÃO COM AUMENTO DA DIMENSÃO DO ESPAÇO DOS ATRIBUTOS DE ENTRADA**

Uma árvore de decisão tradicional resolve problemas de classificação por meio da construção de superfícies de separação perpendiculares aos eixos. No caso de uma distribuição complexa dos dados de entrada do problema, a superfície de separação ainda pode ser construída, entretanto, a complexidade da árvore será maior, comprometendo a solução encontrada. Uma alternativa seria a utilização de uma árvore oblíqua que constrói os hiperplanos de separação oblíquos aos eixos do sistema, entretanto nem sempre a utilização de tais árvores será suficiente para construir a separação adequada de uma determina superfície de dados, pois trabalha com relações lineares dos atributos de entrada durante a construção dos nós da árvore, dificultando a interpretação das regras obtidas da construção da DT.

Neste capítulo é proposta uma metodologia alternativa para a solução de problemas de classificação adotando DT, pela utilização, inicialmente, de um processo simplificado de aumento de dimensão no modelo do sistema e, posteriormente, o uso da capacidade de seleção de atributos durante a construção da própria DT, para então serem extraídas regras de classificação. Essa metodologia utiliza conceitos de aumento de dimensão proposto por Cover (COVER, 1965). Será demonstrada a eficácia desta metodologia em problemas de classificação como o XOR e a planta Iris.

Inicialmente, na seção 4.1, serão discutidas as limitações do uso de uma DT para classificação. Na seção 4.2 serão apresentados teoremas propostos em (COVER, 1965) que relacionam o aumento na dimensão do espaço de atributos em um problema de classificação com o aumento da probabilidade dos pontos neste espaço aumentado serem linearmente separáveis. Na seção 4.3, os conceitos apresentados na seção 4.2 serão utilizados para a formalização da metodologia proposta neste trabalho. O método proposto será testado com vários casos exemplos na seção 4.4. Por fim, a conclusão do capítulo é apresentada na seção 4.5.

#### **4.1. Limitações da Utilização de uma Árvore de Decisão em Problemas de Classificação e Possibilidades de Correções**

A maioria dos algoritmos de construção de uma DT realiza a classificação por meio da análise univariável em cada nó de decisão. Isso implica na construção de superfícies de separação perpendiculares aos eixos. Portanto, no caso de uma distribuição complexa dos dados de entrada do problema, a superfície ainda pode ser construída, entretanto a complexidade da árvore será maior, inclusive podendo comprometer a sua capacidade de generalização (ROKACH e MAIMON, 2008).

Uma alternativa seria utilizar uma DT oblíqua, onde as avaliações nos nós de decisão são realizadas através de relações lineares dos atributos. Um problema potencial em relação a essa abordagem refere-se ao fato de ter que se utilizar agora, associado ao algoritmo de construção da árvore, uma rotina para a obtenção das relações lineares, aumentando o esforço computacional requerido. Também, em várias situações, a construção de retas oblíquas ainda não vai permitir a divisão do espaço de dados de forma adequada. Mas talvez o maior problema seja a obtenção de regras multivariáveis quando da utilização de uma árvore oblíqua, dificultando a interpretação intuitiva das regras.

Esses problemas com a utilização de uma DT normal ou oblíqua foram enfatizados no Capítulo 3, onde as características, utilidades, vantagens e desvantagens do uso de uma DT foram discutidas.

Uma DT é influenciada pela seleção de atributos que podem gerar um conjunto melhor para a realização do treinamento/construção da árvore. Também, os algoritmos de divisão de nós em uma DT têm certa capacidade de seleção, podendo as árvores serem utilizadas para esse intuito (ROKACH e MAIMON, 2005).

Mesmo utilizando seleção de atributos associado a uma DT, em muitas situações somente isso ainda não é suficiente para uma indução adequada de um classificador utilizando DT, pois os dados ainda podem apresentar uma complexidade elevada em função da divisão de espaço promovida pela árvore (ROKACH e MAIMON, 2008).

No Teorema de Cover é discutida a possibilidade da determinação de uma solução linear de um problema de classificação quando aumentamos a dimensão desse problema.

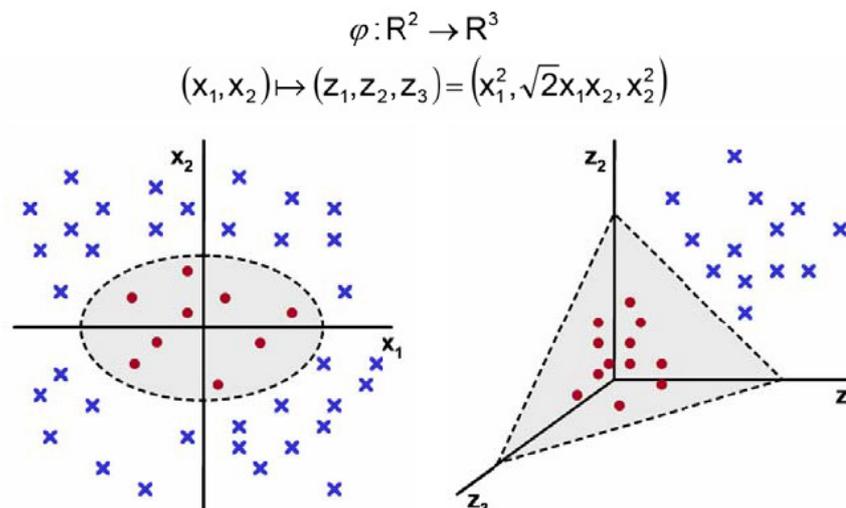
Neste capítulo, é proposta uma metodologia alternativa para a solução de problemas de classificação com DT pela utilização, inicialmente, de um processo simplificado de aumento de dimensão no modelo do sistema e, posteriormente, a utilização de um algoritmo de construção de uma DT que, devido à sua capacidade intrínseca de seleção de atributos durante a criação dos nós, vai utilizar as informações mais relevantes na formação da árvore. O aumento de dimensão ocorre pelo acréscimo de atributos, acrescentando informações novas ou relações não lineares dos atributos originais. As regras extraídas por essa metodologia permitem a classificação dos dados com maior eficiência quando comparado com outros métodos, ou com o uso de uma DT sem aumento de dimensão.

Na seção a seguir deste capítulo será apresentado o Teorema de Cover e discutida a questão do aumento de dimensão de um problema. Posteriormente, esses conceitos apresentados serão utilizados na formalização do método proposto.

## 4.2. Teoremas de Cover

Cover apresenta em seu artigo matemático de 1965 (COVER, 1965), um conjunto de teoremas e definições fundamentais relacionados com a solução de problemas complexos de classificação de padrões, demonstrando que tais problemas quando dispostos em um espaço de alta dimensão apresentam uma maior probabilidade de ser linearmente separáveis do que em um espaço de baixa dimensionalidade. Graficamente esta ideia é representada pela Figura 16.

**Figura 16 - Hiperplano ótimo para padrões não linearmente separáveis e aumento de dimensão**



Fonte: (LORENA e CARVALHO, 2003).

A seguir alguns destes teoremas e definições fundamentais relacionados ao trabalho de Cover, e relacionados à nova metodologia proposta neste capítulo, serão apresentados.

Os teoremas, condições e lema apresentados a seguir apresentam uma demonstração matemática formal indicada em (COVER, 1965).

#### 4.2.1. Função de Contagem

Considere o problema genérico em que é dado um conjunto  $Z$  de  $N$  padrões (vetores),  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ , no espaço euclidiano  $d$ -dimensional  $E^d$ , e atribui-se cada um desses padrões a uma entre duas classes (subconjuntos)  $Z_1$  e  $Z_2$ , desejando-se obter uma função que realize essa separação. Diz-se dessa **partição binária** do espaço  $E^d$  que se trata de uma **dicotomia** (Haykin, 2001), e uma função muito utilizada para realizar essa separação de padrões é a **função de limiar homogênea e linear** (*homogeneous linear threshold function*)  $f_w : E^d \rightarrow \{-1, 0, 1\}$  definida em termos de um parâmetro, ou vetor de pesos  $\mathbf{w}$ , para cada vetor  $\mathbf{z}$  no espaço  $d$ , da seguinte forma (COVER, 1965):

$$f_w(\mathbf{z}) = \begin{cases} \mathbf{1}, & \mathbf{w} \cdot \mathbf{z} > 0 \\ \mathbf{0}, & \mathbf{w} \cdot \mathbf{z} = 0 \\ -\mathbf{1}, & \mathbf{w} \cdot \mathbf{z} < 0 \end{cases} \quad (4.1)$$

com  $\mathbf{w} \cdot \mathbf{z}$  sendo o produto interno de  $\mathbf{w}$  e  $\mathbf{z}$ .

Assim cada função  $f_w$  divide naturalmente  $E^d$  em dois subconjuntos, o conjunto de vetores  $\mathbf{z}$  para qual  $f_w(\mathbf{z}) = 1$  e os vetores para  $f_w(\mathbf{z}) = -1$ . O lugar geométrico dos pontos que define essa separação do espaço  $E^d$  é a superfície de separação. No caso da Equação 4.1, a superfície de separação é um hiperplano definido pelos pontos que satisfazem (COVER, 1965):

$$\{\mathbf{z} : f_w(\mathbf{z}) = 0\} = \{\mathbf{z} : \mathbf{z} \cdot \mathbf{w} = 0\} \quad (4.2)$$

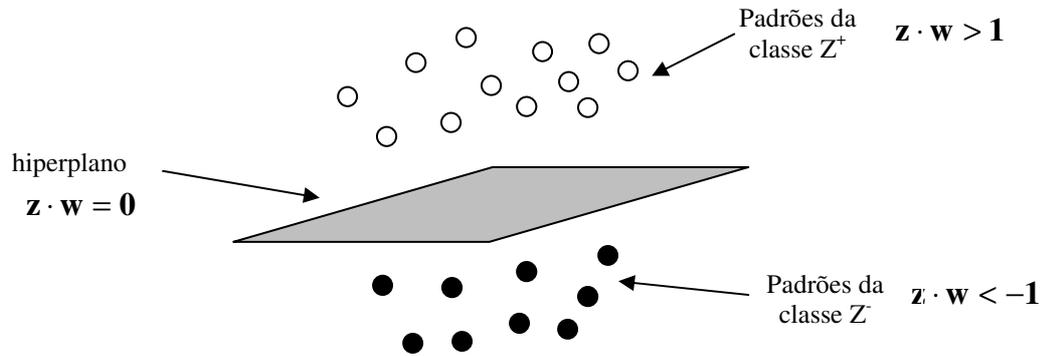
A Figura 17 exemplifica graficamente o hiperplano com relação aos padrões.

Considerando  $Z$  como sendo um conjunto arbitrário de vetores em  $E^d$ . A dicotomia  $\{Z^+, Z^-\}$  de  $Z$  será linearmente separável se, e somente se, existe um vetor de pesos  $\mathbf{w}$  em  $E^d$  e uma escalar  $t$  de tal forma que (COVER, 1965):

$$\mathbf{z} \cdot \mathbf{w} > t, \quad \text{Se } \mathbf{z} \in Z^+$$

$$\mathbf{z} \cdot \mathbf{w} < t, \quad \text{Se } \mathbf{z} \in Z^- \quad (4.3)$$

**Figura 17 - Hiperplano que separa padrões das classes  $Z^+$  (círculos brancos) e  $Z^-$  (círculos pretos)**



Fonte: Elaborada pelo autor.

A dicotomia  $\{Z^+, Z^-\}$  é dita como sendo homogênea e linearmente separável se for linearmente separável com  $t = 0$ . Neste caso, a solução da separação dos padrões é obtida ao se obter um vetor  $\mathbf{w}$  que satisfaça (COVER, 1965):

$$\mathbf{z} \cdot \mathbf{w} > 0, \quad \text{Se } \mathbf{z} \in Z^+$$

$$\mathbf{z} \cdot \mathbf{w} < 0, \quad \text{Se } \mathbf{z} \in Z^- \quad (4.4)$$

Este vetor  $\mathbf{w}$  é chamado de vetor de solução, pois define o hiperplano de separação dos padrões  $\{\mathbf{z} : \mathbf{z} \cdot \mathbf{w} = 0\}$  (Equação 4.2). Se os  $N$  padrões do conjunto  $Z$  estiverem dispostos de forma genérica em  $E^d$  (são linearmente independentes quando tomados em subconjuntos de  $d$  ou menos vetores), é possível definir o número de dicotomias homogêneas e linearmente separáveis pelo Teorema 1 (COVER, 1965), a seguir:

**Teorema 1 (Teorema da Função de Contagem):** Existem  $C(N, d)$  dicotomias homogêneas e linearmente separáveis de  $N$  pontos em uma posição genérica no espaço euclidiano de dimensão  $d$ , onde:

$$C(N, d) = 2 \sum_{k=0}^{d-1} \binom{N-1}{k} \quad (4.5)$$

O coeficiente binomial  $(N, d)$  é definido para todos valores reais  $s$  e inteiros  $k$  por:

$$\binom{s}{k} = \frac{s(s-1)\dots(s-k+1)}{k!} \quad (4.6)$$

Por outro lado, caso se disponha de um conjunto de  $N$  hiperplanos que passem pela origem do espaço  $E^d$ , é possível calcular o número de divisões deste espaço gerados por estes hiperplanos através do Teorema 2 abaixo (COVER, 1965):

**Teorema 2:**  $N$  hiperplanos em posições genéricas que passem pela origem de um espaço de dimensão  $d$  dividem este espaço em  $C(N, d)$  regiões.

Do Teorema 1 tem-se que a partir de  $N$  padrões linearmente independentes é possível estabelecer o número de possíveis hiperplanos, que realizam separações de dados linearmente separáveis. Porém, caso os dados não sejam linearmente separáveis para a dicotomia  $\{Z^+, Z^-\}$ , um único hiperplano não conseguirá separar totalmente os padrões. Como o Teorema 2 aponta que a partir de  $N$  hiperplanos é possível estabelecer o número de divisões do espaço  $E^d$ , é possível buscar tratar a dicotomia  $\{Z^+, Z^-\}$  para padrão não linearmente separáveis a partir do aumento do número de hiperplanos para gerar uma quantidade de regiões suficiente que permita realizar a separação dos dados nas classes  $Z^+$  e  $Z^-$ .

#### 4.2.2. Separação por Superfícies Arbitrárias

Nesta seção, o número de dicotomias de um conjunto de pontos que pode ser separado por uma família arbitrária de superfícies de separação vai ser determinado. Este número se origina diretamente do Teorema da Função de Contagem quando a família de superfícies de separação e os conjuntos de pontos que podem ser separados são cuidadosamente definidos.

Considerando uma família de superfícies, cada uma dividindo um dado espaço em duas regiões distintas, e uma coleção de  $N$  pontos neste espaço, com cada um destes pontos sendo associado a uma das duas classes  $Z^+$  ou  $Z^-$ . Esta dicotomia dos pontos é tomada como separável pela família de superfícies, se existe uma superfície da família que separa os pontos entre  $Z^+$  e  $Z^-$ . Para o conjunto de  $N$  objetos  $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ , cada padrão  $\mathbf{z} \in Z$  poderá ser

associado a um valor real obtido de uma função de medida  $\phi_1, \dots, \phi_d$  que compreende um vetor de medidas  $\Phi: Z \rightarrow E^d$  onde  $\Phi(\mathbf{z}) = [\phi_1(\mathbf{z}), \dots, \phi_d(\mathbf{z})]$ ,  $\mathbf{z} \in Z$ .

A dicotomia  $\{Z^+, Z^-\}$  de  $Z$  é separável em  $\Phi$  se existe um vetor  $\mathbf{w}$  que (COVER, 1965):

$$\mathbf{w} \cdot \Phi(\mathbf{z}) > 0, \quad \mathbf{z} \in Z^+$$

$$\mathbf{w} \cdot \Phi(\mathbf{z}) < 0, \quad \mathbf{z} \in Z^- \quad (4.7)$$

Deve ser observado que a superfície de separação no espaço de medidas é o hiperplano  $\mathbf{w} \cdot \Phi = 0$ . A imagem inversa deste hiperplano é a superfície de separação  $\{\mathbf{z}: \mathbf{w} \cdot \Phi(\mathbf{z}) = 0\}$  no espaço de padrões.

**Definição:** Considerando uma função de medida de valores reais  $\Phi$  que mapeia os padrões  $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$  em  $E^d$ . O conjunto  $Z$  é dito como sendo de posição genérica em  $\Phi$  se a Condição 1 for atendida (COVER, 1965).

**Condição 1:** Cada elemento  $k$  de um subconjunto do conjunto de dimensão  $d$  de vetores de medidas  $\{\Phi(\mathbf{z}_1), \dots, \Phi(\mathbf{z}_N)\}$  é linearmente independente para todo  $k \leq d$  (COVER, 1965).

**Condição 1':** Cada elemento  $d$  de um subconjunto do conjunto de dimensão  $d$  de vetores de medidas  $\{\Phi(\mathbf{z}_1), \dots, \Phi(\mathbf{z}_N)\}$  é linearmente independente (COVER, 1965).

**Condição 1'':** Cada submatriz  $d \times d$  da matriz  $N \times d$  (COVER, 1965):

$$\Phi = \begin{bmatrix} \phi_1(\mathbf{z}_1) & \dots & \phi_d(\mathbf{z}_1) \\ \vdots & & \vdots \\ \phi_1(\mathbf{z}_N) & \dots & \phi_d(\mathbf{z}_N) \end{bmatrix} \quad (4.8)$$

apresenta um determinante não nulo.

**Condição 1''':** Nenhum padrão  $d+1$  repousa sobre qualquer superfície  $\Phi$  no espaço de padrões (COVER, 1965).

A Condição 1'' é consequência algébrica da Condição 1'. Note que a posição geral é relacionada com o posto da matriz  $\phi$  ( $\phi$  tem um posto máximo  $d$  se pelo menos uma submatriz  $d \times d$  apresenta um determinante não nulo). A Condição 1''' relaciona a posição genérica no espaço de medidas com a posição no espaço de padrões.

O lema proposto a seguir estabelece uma extensão do Teorema 1 para o caso em que a família de superfícies de decisão é restrita a passar através de um determinado conjunto de pontos.

**Lema 1:** Tomando  $Z^+$  e  $Z^-$  como subconjuntos de  $E^d$ , e  $\mathbf{v}$  um ponto diferente da origem em  $E^d$ , então a dicotomia  $\{Z^+ \cup \{\mathbf{v}\}, Z^-\}$  e  $\{Z^+, Z^- \cup \{\mathbf{v}\}\}$  são ambas homogêneas e linearmente separáveis por um subespaço de dimensão  $(d-1)$  contendo  $\mathbf{v}$  (COVER, 1965).

Em termos geométricos, o Lema 1 afirma que um novo ponto pode ser agregado a qualquer uma das metades de uma dicotomia separável para formar duas novas dicotomias separáveis, se e somente se existir um hiperplano de separação passando através deste novo ponto e que separa a antiga dicotomia. Isto é razoável, pois se um hiperplano existe, pequenos deslocamentos do hiperplano vão permitir a classificação arbitrária do novo ponto sem afetar a separação da antiga dicotomia.

A seguir será apresentado o Teorema 3 que generaliza o Teorema da Função de Contagem para certas classes de funções não lineares sobre certas restrições. Em particular, afirma que as  $k$  restrições independentes sobre a classe de superfícies de separação reduzem os graus de liberdade da classe por  $k$ .

**Teorema 3:** Se uma superfície  $\phi$ ,  $\{\mathbf{z}: \mathbf{w} \cdot \phi(\mathbf{z}) = 0\}$ , é restrita de forma a conter os pontos  $V = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ , onde  $\phi(\mathbf{v}_1), \dots, \phi(\mathbf{v}_k)$  são linearmente separáveis, e onde as projeções de  $\phi(\mathbf{z}_1), \dots, \phi(\mathbf{z}_N)$  para o subespaço ortogonal ao espaço gerado  $\phi(\mathbf{v}_1), \dots, \phi(\mathbf{v}_k)$  estão numa posição genérica, então existem  $C(N, d-k)$  separações  $\phi$  para a dicotomia de  $Z$  (COVER, 1965).

Uma generalização natural da separação linear é a separação por polinômio. Na discussão a seguir, considere os padrões como vetores em um espaço de dimensão  $m$ . A

função de medida  $\phi$  então realiza o mapeamento dos pontos de um espaço  $m$  em pontos em um espaço  $d$ .

Considere uma classe natural de mapeamentos obtidos por produtos adjacentes das coordenadas do vetor de padrões. As superfícies de separação de tais mapeamentos são conhecidas como variedades racionais de ordem  $q$ . Uma variedade racional de ordem  $q$  num espaço de  $m$  dimensões é representada por uma equação de grau  $q$  homogênea nas coordenadas  $(z)_i$  (COVER, 1965):

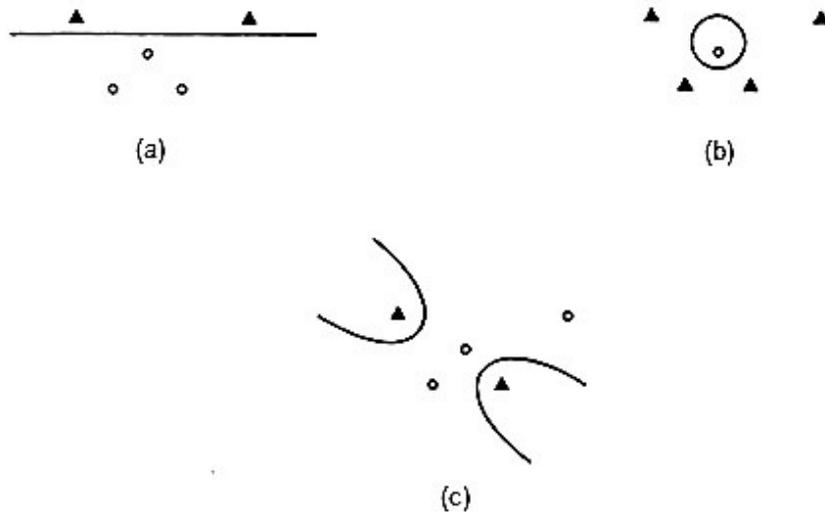
$$\sum_{0 \leq i_1 \leq \dots \leq i_q \leq m} a_{i_1 i_2 \dots i_q} (z)_{i_1} \dots (z)_{i_q} = 0 \quad (4.9)$$

onde  $(z)_i$  é componente  $i$  de  $\mathbf{z}$  em  $E^d$  e  $(z)_0$  é definido como igual a 1 de modo a escrever a expressão de forma homogênea. Deve ser notado que existem  $(m-q)!/m!q!$  coeficientes em (4.9).

Exemplos de superfícies desta forma são hiperplanos (variedades racionais de primeira ordem), quádricas (variedades racionais de segunda ordem), e hiperesferas (quádricas com certas restrições lineares nos coeficientes). A Figura 18 ilustra três dicotomias de uma mesma configuração de pontos. Das 32 dicotomias de cinco pontos na Figura 18, precisamente  $C(5, 3)=22$  são linearmente separáveis,  $C(5, 4)=30$  esfericamente separáveis, e  $C(5, 5) = 32$  são separáveis por quádricas. Fica claro que uma separação linear é contida em uma separação esférica, que por sua vez implica em uma separação com quádrica, o inverso nem sempre é verdade (COVER, 1965).

A função  $\phi$  permite o mapeamento dos dados originais em um espaço  $\mathbf{M}$  de dimensão maior. Métodos baseados em *kernel* utilizam essa idéia associada a funções *kernel*, denominadas de funções de *kernel*, que representam o produto  $kernel(z_1, z_2) = \phi(z_1) \cdot \phi(z_2)$ . Observe que podem existir diversas funções de *kernel* associadas ao espaço  $\mathbf{M}$  mapeado por  $\phi$  (RIESEN e BUNKE, 2010). Na metodologia apresentada nesta Tese é utilizada a idéia da representação do problema original em um espaço de maior dimensão, entretanto não são utilizadas funções de *kernel*.

**Figura 18 - Exemplos de superfícies  $\phi$  que separam dicotomias de cinco pontos em duas classes: (a) linearmente separáveis, (b) esférico separáveis, (c) quádrlica separáveis**



Fonte: (COVER, 1965).

#### 4.2.3. Separação de Padrões Randômicos

Existem dois tipos de variabilidades que são considerados num problema de dicotomia de padrões (COVER, 1965; HAYKIN, 2001):

- 1) Os padrões são fixados em uma posição, mas são classificados independentemente com igual probabilidade entre as duas classes.
- 2) Os padrões são distribuídos randomicamente no espaço, com a classificação binária desejada podendo ser aleatória ou fixa.

Dentro dessas condições, a separação do conjunto de vetores de padrões torna-se um evento randômico dependendo da dicotomia escolhida e da configuração dos padrões. A probabilidade deste evento randômico e o número máximo de padrões randômicos que podem ser separados por uma dada família de superfícies de decisão podem ser determinados.

Suponha que os padrões  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  são escolhidos independentemente, de acordo com uma medida de probabilidade  $\mu$  do espaço de padrões. Suponha também que todas as dicotomias possíveis sejam equiprováveis. A probabilidade  $P(N, d)$  de que os  $N$  padrões sejam separáveis no espaço de dimensão  $d$  por um hiperplano pode ser obtida a partir do valor  $C(N,$

$d$ ) que representa o número de dicotomias homogêneas e linearmente separáveis de  $N$  pontos em uma posição genérica no espaço euclidiano de dimensão  $d$ . A equação a seguir apresenta a formulação de  $P(N, d)$  (COVER, 1965):

$$P(N, d) = \left(\frac{1}{2}\right)^N C(N, d) = \left(\frac{1}{2}\right)^{N-1} \sum_{k=0}^{d-1} \binom{N-1}{k} \quad (4.10)$$

A equação (4.10) demonstra a essência da separação proposta por Cover para padrões aleatórios. Ela expressa o fato de que a distribuição binomial cumulativa corresponde à probabilidade de  $(N-1)$  arremessos de uma moeda honesta resultarem em  $(d-1)$  ou menos caras (HAYKIN, 2001).

A partir de (4.10), observa-se que ao aumentar o valor de  $d$  aumenta-se a probabilidade  $P(N, d)$ . Isso implica que ao aumentar a dimensão dos padrões estes tem maior probabilidade de serem linearmente separáveis. Porém, para que o Teorema 1 e o Teorema 3, permaneçam válidos, os padrões na nova dimensão  $m > d$  devem permanecer linearmente independentes quando tomados em subconjuntos de  $m$  ou menos vetores.

### 4.3. Metodologia Proposta

Nesta seção, é proposta uma metodologia alternativa para a resolução de problemas de classificação utilizando uma DT (Capítulo 3) e conceitos apresentados nos estudos do trabalho de Cover (seção 4.2).

Nas árvores de decisão (Capítulo 3), cada nó particiona de forma binária o espaço de atributos, correspondendo a um hiperplano. Um ramo de uma DT corresponde a  $N$  hiperplanos, onde  $N$  é o número de nós presentes neste ramo. Assim, a partir do Teorema 2 (seção 4.2), tem-se que o ramo de uma DT pode dividir o espaço de atributos em  $C(N, d)$  regiões, onde  $d$  seria o número de atributos considerados no espaço de entrada. Isso implica para padrões não separáveis linearmente que: (i) deve-se gerar DTs com ramos longos (grande quantidade de nós) e (ii) muitos ramos devem ser gerados (uma vez que há muitos nós).

A nova metodologia propõe a solução de problemas de classificação inicialmente realizando um aumento das variáveis do problema, utilizando novos atributos originalmente não utilizados e relações entre os atributos do espaço de entrada original. Com isso, é obtida

uma nova representação do problema em uma dimensão maior e, conforme indicado na equação (4.10), passa-se a ter uma probabilidade maior de levar os padrões do novo espaço de atributos a serem linearmente separáveis, e a uma divisão mais favorável desse novo espaço de entrada por hiperplanos.

Os novos atributos propostos neste trabalho devem atender às exigências da Subseção 4.2.2 e, portanto, devem ser linearmente independentes dos atributos originais. Nesta pesquisa, os atributos incorporados no modelo do problema, além de informações novas originalmente não utilizadas, são relações não lineares entre os atributos originais. Dessa forma, como exemplo, considerando como atributos originais  $x_1$  e  $x_2$ , os novos atributos propostos têm os formatos de  $x_3$ ,  $x_4$  e  $x_5$  apresentados em (4.11).

$$\left\{ \begin{array}{l} x_3 \rightarrow \text{Atributo físico ainda não utilizado} \\ x_4 = x_1 \cdot x_2 + b \\ x_5 = \frac{x_1}{x_2 + b} \end{array} \right. \quad (4.11)$$

onde  $b$  representa um valor constante ou uma soma de outros atributos.

As expressões de  $x_4$  e  $x_5$  fornecem novos atributos a partir de relações não lineares dos atributos originais, garantindo uma independência linear entre os novos atributos e os antigos, conforme as exigências indicadas na Subseção 4.2.2.

Deve ser salientado que os novos atributos formados pelo produto ou pela razão de atributos originais devem ter significado físico no problema a ser tratado ou representarem uma relação adimensional (normalização), de forma a garantir um maior grau de significância para as regras obtidas da DT construída. Isso fica mais bem salientado a partir do exemplo do Problema da Planta Iris (Subseção 4.4.2).

A obtenção de uma solução linear em uma dimensão maior permitirá a obtenção de uma solução de classificação por parte de uma DT com uma melhor construção das fronteiras de separação entre as classes. Na verdade, a utilização de atributos baseados em relações de outros atributos originais, representa uma abordagem semelhante à utilização de uma árvore oblíqua, entretanto não utilizando uma relação linear, mais sim agora uma razão ou produto

entre atributos. Assim é obtido um hiperplano de separação linear numa dimensão maior, mas que numa projeção numa dimensão menor, será uma reta não mais necessariamente paralela aos eixos.

#### 4.3.1. Seleção de Atributos e Extração de Regras

O passo seguinte à aplicação do aumento de dimensão é a utilização direta dos dados modificados em uma DT. Durante a construção dos nós pelo algoritmo de construção da DT, o próprio algoritmo já realiza uma seleção de atributos, de forma a ser obtida uma representação compacta suficiente para o problema, facilitando a capacidade de generalização da árvore, e ao mesmo tempo adequada para uma solução com maior precisão do problema. Evitando a utilização de métodos não lineares para a solução do problema original (agora modelado por uma DT), com uma maior complexidade computacional e que podem ser facilmente polarizados em um mínimo local, resultando em problemas na obtenção de uma solução ótima.

A seguir serão apresentados alguns exemplos simples de utilização do método proposto. No Capítulo 5, o método será utilizado para o caso da análise de falhas incipientes em transformadores, sendo comparados os resultados obtidos com os métodos indicados nas normas para análise de transformadores.

### 4.4. Exemplos de Utilização

Nesta seção são apresentados dois exemplos simples para ilustrar a metodologia proposta na seção anterior. Esses exemplos (dispositivo XOR e subclassificação da planta Iris) não permitem uma solução trivial, pois originalmente são problemas não lineares.

#### 4.4.1. Problema do XOR

Considerando o problema da representação de um dispositivo XOR:

$$v_1 = [-1 \ -1], \quad s_1 = -1$$

$$v_2 = [1 \ -1], \quad s_2 = 1$$

$$v_3 = [-1 \ 1], \quad s_3 = 1$$

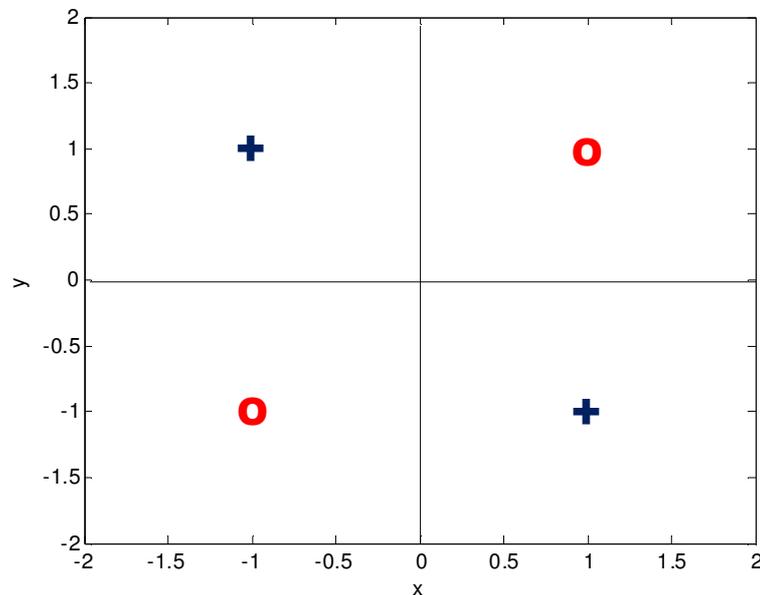
$$v_4 = [1 \ 1], \quad s_4 = -1$$

$$V=[v_1, v_2, v_3, v_4]$$

$$Out=[s_1, s_2, s_3, s_4]$$

O problema do XOR pode ser descrito graficamente pela Figura 19, onde as cruces azuis representam a ativação do dispositivo e os círculos vermelhos a não ativação. Também, a Tabela 14, define as entradas e saídas possíveis para o dispositivo. É fácil perceber que a construção da superfície de separação deve ser oblíqua aos eixos, não fácil de ser construída por vários dos métodos de identificação tradicionais.

**Figura 19 -Saída de um dispositivo XOR**



Fonte: Elaborada pelo autor.

**Tabela 14- Entradas e saída para um XOR**

<b>x1</b>	<b>x2</b>	<b>s</b>
-1	-1	-1
1	-1	1
-1	1	1
1	1	-1

Fonte: Elaborada pelo autor.

Utilizando árvores de decisão construídas através dos métodos ID3 e C4.5, não foi possível a obtenção de uma solução para o problema do XOR, mesmo alterando os ajustes dos algoritmos.

Isso já era esperado, já que os algoritmos univariáveis para construção de nós em uma DT são considerados como algoritmos *gulosos*, e esses algoritmos apresentam problemas de convergência quando na solução do sistema é necessária a observação de iterações entre os atributos de entrada (PAPPA, 2002). Suponhamos que utilizemos o método C4.5, que considera apenas um atributo por vez na construção dos nós, para determinar a função do XOR. Se a condição ( $x1 = 1$ ) é selecionada para tentar iniciar a construção da árvore, ela não será útil, uma vez que, considerando todos os dados, a distribuição de classes é igual (50% de exemplo com  $s = -1$  e 50% com  $s = 1$ ). O mesmo acontecerá se qualquer outro par atributo/valor for selecionado, pois a distribuição das classes permanece constante (PAPPA, 2002).

Considerando, agora, o problema do XOR acrescentado relações entre as entradas, conforme indicado na Tabela 15, foi aplicado o método de construção de árvores C4.5, sendo obtido uma taxa de 100% de acerto na classificação.

**Tabela 15 - Problema do XOR com atributos expandidos**

<b>x</b>	<b>y</b>	<b>x/y</b>	<b>s</b>
1	1	1	NA
-1	-1	1	NA
1	-1	-1	A
-1	1	-1	A

Fonte: Elaborada pelo autor.

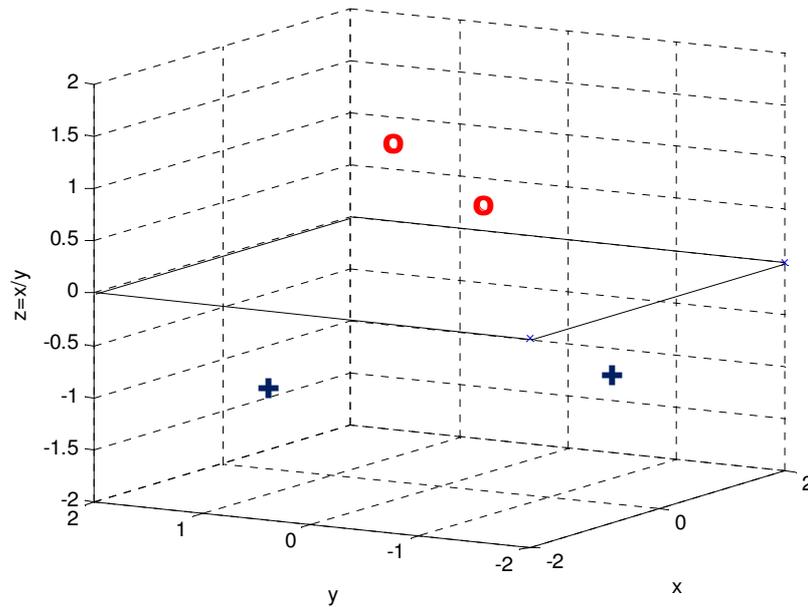
As regras obtidas a partir da solução do C4.5 são apresentadas a seguir, onde podemos perceber que o método faz uma seleção pela utilização apenas do atributo  $x/y$ , relação entre as variáveis originais, portanto um atributo válido para ser acrescentado ao problema, conforme a metodologia descrita na seção 4.3. Graficamente a solução da classificação é feita pelo plano com origem em  $x/y=0$ , como indicado na Figura 20.

$$\text{Se } x/y < 0 \text{ Então XOR ativa} \quad (4.12)$$

$$\text{Se } x/y \geq 0 \text{ Então XOR não ativa}$$

Portanto, é demonstrada a metodologia proposta, onde inicialmente a dimensão do problema é expandida para depois ser selecionado um ou mais atributos que permitam uma solução simples e compacta do problema.

**Figura 20 - Solução do XOR**



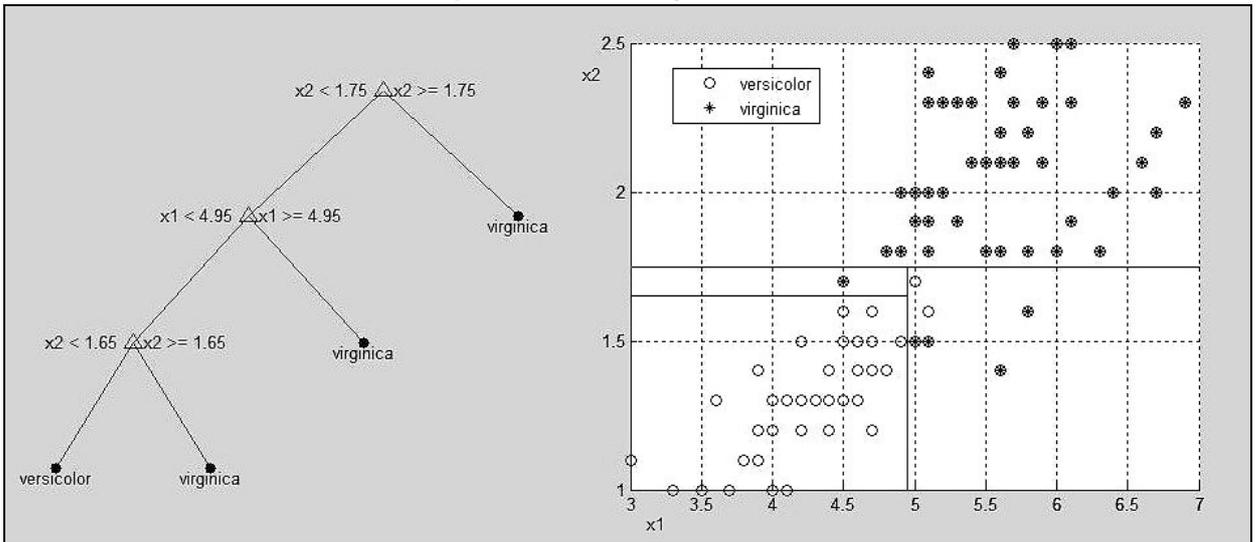
Fonte: Elaborada pelo autor.

#### 4.4.2. Problema da Planta Íris

Considerando-se agora como exemplo, um sub-problema de classificação entre as espécies Versicolor e Virginica da flor Íris, tendo como entradas o comprimento ( $x_1$ ) e a largura ( $x_2$ ) da pétala em centímetros (cm). Inicialmente, uma DT com o algoritmo C4.5 foi utilizada considerando as entradas originais apresentadas no problema. Os resultados são mostrados na Figura 21, com a árvore e o gráfico obtidos representando os dados e as linhas de separação indicadas pela DT.

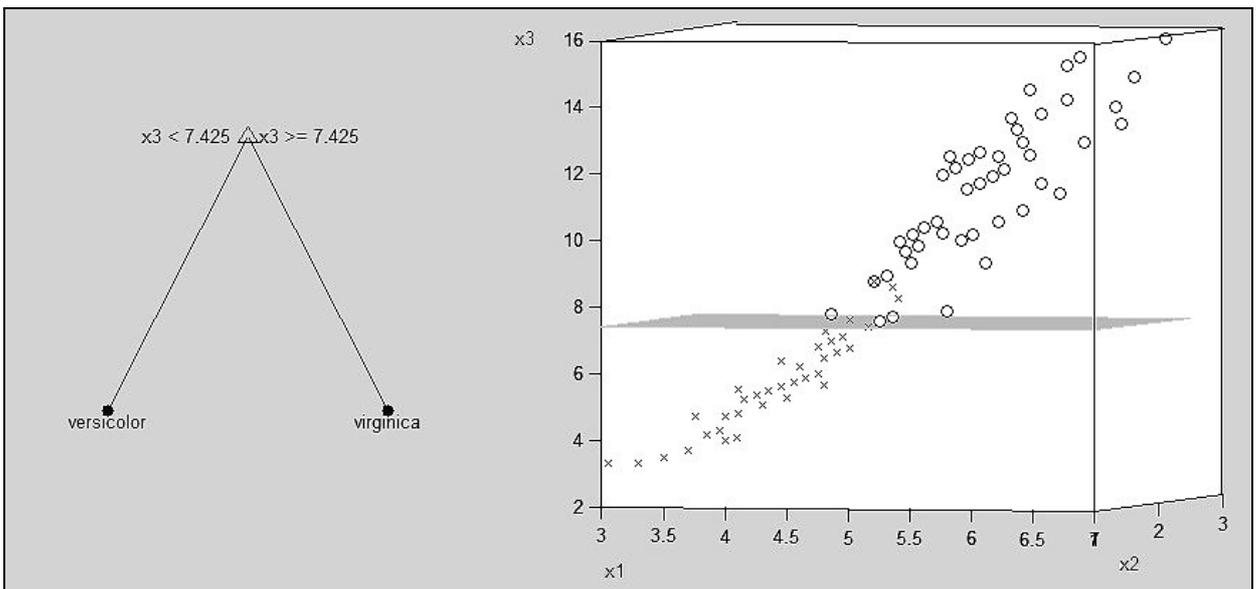
Adotando a metodologia proposta, foi adicionada mais uma entrada ao problema, variável ( $x_3$ ), representando o produto das variáveis originais ( $x_1$  e  $x_2$ ). Portanto, aumentando a dimensão do problema original. A nova variável, representando o produto das variáveis anteriores, apresenta um significado físico (área da pétala) e também atende os requisitos do Teorema de Cover. Nesta nova configuração, a solução com a mesma precisão da construção anterior é obtida com apenas uma única regra, representando um plano no espaço de três dimensões (Figura 22).

**Figura 21 - Classificação da Íris em 2D**



Fonte: Elaborada pelo autor.

**Figura 22 - Classificação da Íris em 3D**



Fonte: Elaborada pelo autor.

Portanto, mais uma vez foi demonstrada a utilidade da metodologia proposta neste capítulo associado ao uso de uma DT para a solução de um problema de classificação.

Nesta seção foram apresentados problemas simples, no próximo capítulo será utilizado o método proposto para a solução de um problema relevante de engenharia, a análise de falhas

incipientes em equipamentos elétricos com isolamento a óleo, como os transformadores de potência.

#### **4.5. Considerações Parciais**

Neste capítulo foi apresentada uma metodologia simplificada para a utilização de uma DT para a solução de problemas de classificação. Esse método utiliza conceitos do Teorema de Cover que foram discutidos no capítulo.

O novo método propõe uma nova representação de um problema, utilizando novos atributos ou relações não lineares das variáveis originais do problema, e, portanto, aumentando a dimensão do sistema. Após isso, é utilizada uma DT que através de sua capacidade intrínseca de seleção escolherá as variáveis mais relevantes para a solução do problema. A própria topologia determinada para a DT construída, também, permite a extração direta de regras de classificação.

Os exemplos simples apresentados neste capítulo demonstraram a utilidade e eficiência do método. No próximo capítulo, a metodologia será utilizada para a análise de transformadores.

## **5. DIAGNÓSTICO APRIMORADO DE FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA**

Como ressaltado no Capítulo 2, os métodos de diagnóstico de falhas incipientes definidos em normas apresentam limitações que podem comprometer a qualidade dos programas de manutenção das concessionárias de energia elétrica. Entre estas limitações listam-se: diagnósticos incorretos, indecisão de diagnóstico para certas faixas de valores dos atributos de entrada e grande sensibilidade à imprecisão nas medidas das concentrações dos gases. Utilizando a ideia proposta no Capítulo 4 para gerar novos métodos de diagnóstico de equipamentos de potência isolados a óleo realizando um aumento do espaço de atributos (a partir das informações cromatográficas e físico-químicas do óleo isolante), analisa-se neste capítulo o método desenvolvido ao realizar sua comparação com métodos clássicos de análise de transformadores, observando vantagens e limitações.

Na organização deste capítulo, analisou-se separadamente a aplicação da metodologia proposta no Capítulo 4 para cada tipo de informação disponível do óleo isolante (cromatográfico ou físico-químico). Na seção 5.1 são descritos os conjuntos de dados cromatográficos (concentrações de gases) utilizados: (i) para gerar as regras do novo método de análise de transformadores utilizando análise de gases dissolvidos no óleo isolante e (ii) para realizar a comparação deste novo método com métodos clássicos. O novo método de DGA é apresentado formalmente na seção 5.2, junto com comentários das simulações realizadas no seu desenvolvimento. A eficiência e a confiabilidade do novo método DGA proposto são discutidas nas seções 5.3 e 5.4, onde esse método é confrontado com as principais limitações de outros métodos baseados em análise dos gases indicados em normas técnicas. Na seção 5.5 é apresentado o desenvolvimento e os testes de outro método desenvolvido para o diagnóstico simplificado de transformadores utilizando informações de medidas físico-químicas realizadas no óleo isolante do transformador. A conclusão do capítulo é apresentada na seção 5.6.

### **5.1. Dados de Concentração de Gases Utilizados**

No método desenvolvido, utilizando uma DT para o diagnóstico de equipamentos com isolamento a óleo como os transformadores, foi adotado um banco de dados com 162 amostras. Sendo 117 casos do conjunto de dados IEC TC 10 (DUVAL e PABLA, 2001). Também são utilizados 39 casos retirados das Tabelas de I a III da referência (DUVAL, 2002)

e 6 casos indicados em (NARESH, SHARMA e VASHISTH, 2008). O diagnóstico associado a esses dados são considerados confiáveis, com a maioria obtida após a abertura do equipamento e análise de especialistas.

Portanto, foi considerado um conjunto de dados com acesso público e de fácil verificação. Os dados do IEC TC 10 serviram para o aprimoramento dos métodos citados na norma IEC 60599 (IEC, 2008).

Como a quantidade de dados, 162 amostras, representa um conjunto não muito elevado. No desenvolvimento das simulações realizadas para a construção de uma DT para o diagnóstico de transformadores, foi adotada a técnica de Validação Cruzada (*k*-partições) (MITCHELL, 1997) com 10 subconjuntos (*k* = 10 partições) para uma análise estatística.

A validação cruzada é utilizada em análises onde o conjunto original de dados apresenta um número reduzido de exemplos, não apresentando um número grande de dados para o treino e teste (verificação de generalização) da solução construída. Na técnica da validação cruzada o conjunto de dados original é dividido em subconjuntos (pastas), tipicamente sendo utilizados 10 subconjuntos. Em cada ciclo de simulação, 9 pastas são utilizadas para o treino da DT e uma pasta para o teste de validação. De modo a reduzir a variabilidade, múltiplos ciclos de simulação são realizados, trocando em cada simulação o subconjunto de validação.

Foram adotados como atributos propostos para as entradas do problema: a concentração de sete gases ( $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $CO$ , e  $CO_2$ ) normalmente utilizados nos métodos normatizados (IEC, 2008; IEEE, 2008) em partes por milhão (ppm), as razões entre os gases ( $CH_4/H_2$ ,  $C_2H_2/C_2H_4$ ,  $C_2H_4/C_2H_6$ ,  $C_2H_4/CH_4$ ,  $C_2H_2/CH_4$ ,  $C_2H_2/H_2$ , e  $CO/CO_2$ ), e as porcentagens relativas ( $CH_4\% = 100.x / (x+y+z)$ ,  $C_2H_2 \% = 100.y / (x+y+z)$ ,  $C_2H_4 \% = 100.z / (x+y+z)$ ,  $H_2\% = 100.H_2 / (H_2+C_2H_6+CO+CO_2)$ ,  $C_2H_6\% = 100.C_2H_6 / (C_2H_6+x+y+z)$ ,  $CO\% = 100.CO / (C_2H_6+x+y+z+CO+CO_2)$ , e  $CO_2\% = 100.CO_2 / (C_2H_6+x+y+z+CO+CO_2)$ ), onde  $x=CH_4$ ,  $y=C_2H_2$ , e  $z=C_2H_4$ ).

As concentrações dos gases ( $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $CO$ , e  $CO_2$ ) e as relações ( $CH_4/H_2$ ,  $C_2H_2/C_2H_4$  e  $C_2H_4/C_2H_6$ ) já foram utilizadas nos métodos de diagnóstico indicados nas normas (IEC, 2008; IEEE, 2008), os valores percentuais ( $CH_4\%$ ,  $C_2H_2\%$  e  $C_2H_4\%$ ) no método do Triângulo de Duval (DUVAL, 2002) e o valor ( $H_2\%$ ) em (LI e WU, 2011). Excetuando os valores das sete concentrações de gases, os demais valores representam razões

entre gases ou de um gás em relação a um conjunto de gases, em razões simples ou percentuais, portanto, em de acordo com as restrições para novos dados no aumento de dimensão do problema, definidas na Seção 4.3.

Este número elevado de entradas está de acordo com a metodologia proposta neste trabalho de considerar inicialmente um aumento da dimensão do problema, de forma a possibilitar na construção da DT a seleção dos requisitos mais significativos, resultando na obtenção de um conjunto de regras que permite o diagnóstico mais eficiente dos transformadores.

Também, as seguintes hipóteses foram feitas na composição dos dados de entrada: (i) se a concentração de um determinado gás não estiver disponível então se considera o valor zero, (ii) uma razão de 0/0 é definida como de valor nulo (LI e WU, 2011), (iii) uma proporção dada do tipo  $v/0$  é definida como sendo de valor 20 (LI e WU, 2011), considerando que  $v$  não é nulo, e (iv) concentrações indicadas como "<1" nos conjuntos de dados são consideradas como de valor 0,5 (LI e WU, 2011).

Os rótulos de saída para a DT construída podem indicar os seguintes diagnósticos possíveis: descargas parciais (DP), descargas de baixa energia (D1), descargas de alta energia (D2), falhas térmicas com temperatura  $<700$  °C (T1/T2) e falhas térmicas com temperatura  $>700$  °C (T3). Os diagnósticos de falhas térmicas de baixa temperatura foram unificados em T1/T2, conforme utilizado pela base de dados IEC TC 10. Rótulos compatíveis com os métodos indicados nas normas, principalmente com o método do Triângulo de Duval.

Em cada rodada de simulação utilizando validação cruzada, todas as pastas são subconjuntos formados por 10% do total de dados, onde os casos envolvendo os cinco rótulos considerados (PD, D1, D2, T1/T2 e T3) no problema são uniformemente distribuídos.

Na próxima seção será demonstrada a obtenção de regras de diagnóstico utilizando DT com os conjuntos de dados citados nesta seção.

## 5.2. Levantamento das Regras do Método DGA Proposto

Nas simulações realizadas para a construção de uma DT neste trabalho foi utilizado o *software* de uso livre SIPINA (RAKOTOMALALA, 2012) que incorpora diversas técnicas de

aprendizado, com destaque para o uso de árvores de decisão. Entre os diversos algoritmos de construção de uma DT, o *software* permite o uso do C4.5 que é um dos algoritmos mais eficientes e populares (Capítulo 3). O C4.5 utiliza a medida da Razão de Ganho (Apêndice B) para seleccionar o atributo que melhor divide os nós na construção da árvore e utiliza como método de poda para a árvore gerada, a Poda Baseada no Erro (Apêndice B).

Para comparação, inicialmente foi realizada a construção de uma DT considerando como entradas as razões:  $R_1=CH_4/H_2$ ,  $R_2=C_2H_2/C_2H_4$ , e  $R_5=C_2H_4/C_2H_6$ , atributos normalmente utilizados nos métodos indicados nas normas IEEE C57.104 (IEEE, 2008) e IEC 60599 (IEC, 2008). As classes associadas às folhas da árvore construída seriam os diagnósticos, indicando os rótulos já citados (DP, D1, D2, T1/T2 e T3) na seção anterior. Do total de dados, 162 amostras, 10% foram considerados para teste e o restante para construção da árvore.

Utilizando o *software* SIPINA, com a opção do algoritmo C4.5, foi construída a árvore, os erros obtidos para o treino e teste da DT foram, respectivamente: 8,9% (num universo de 146 casos) e de 12,5% (em relação a 16 casos). As matrizes de confusão, indicando os casos de acerto e insucesso na classificação para os dados de treino e de teste, são mostradas nas Tabelas 16 e 17, respectivamente.

**Tabela 16 - Matriz de confusão, dados de treino e entradas típicas**

	D1	D2	T1/T2	T3	PD
D1	30	4	1	0	0
D2	1	45	0	0	0
T1/T2	0	0	34	0	0
T3	0	0	7	11	0
PD	0	0	0	0	13

Fonte: Elaborada pelo autor.

**Tabela 17 - Matriz de confusão, dados de teste e entradas típicas**

	D1	D2	T1/T2	T3	PD
D1	3	1	0	0	0
D2	0	5	0	0	0
T1/T2	0	0	4	0	0
T3	0	0	1	1	0
PD	0	0	0	0	1

Fonte: Elaborada pelo autor.

Utilizando a ideia de aumentar o número de atributos, foi construída uma nova árvore com o mesmo banco de dados, mas agora utilizando o conjunto completo de entradas indicadas na seção anterior (21 entradas). Os novos atributos foram escolhidos, tais como novas relações, ainda não utilizadas, entre os gases obtidos da análise cromatográfica; valores em p.p.m. desses gases; e valores percentuais de um determinado gás em relação ao conjunto de gases. Esses novos atributos são compatíveis com a metodologia de aumento de dimensão do problema original associado com a construção de uma DT, demonstrada no Capítulo 4. Também foram realizados testes de validação cruzada que são apresentados na Tabela 18. Também nessa tabela, são apresentados valores considerando o método do Triângulo de Duval para comparação.

A linha "Pasta" na Tabela 18 representa o número do subconjunto considerado para teste na atual rodada de validação cruzada. Em cada rodada, todas as pastas são subconjuntos formados por 10% do total de dados, onde os casos envolvendo os cinco rótulos considerados (PD, D1, D2, T1/T2 e T3) são uniformemente distribuídas.

**Tabela 18 - Percentual de erro de classificação, considerando os casos para validação cruzada**

Pasta \ Conjunto	1	2	3	4	5	6	7	8	9	10
Treinamento (Regras extraídas AD)	2,7	1,4	4,1	3,4	1,4	2,7	2	2,7	4,1	2
Treinamento (Método do Triângulo)	13,7	13,7	13,7	13,7	13,7	13,7	13,7	13,7	13	13
Teste (Regras extraídas AD)	25	6,25	6,25	25	25	12,5	12,5	12,5	12,5	6,25
Teste (Método do Triângulo)	12,5	12,5	12,5	12,5	12,5	12,5	12,5	12,5	18,75	18,75

Fonte: Elaborada pelo autor.

A Tabela 18 apresenta os erros relacionados aos conjuntos de treinamento e de testes para cada pasta da validação cruzada, usando regras obtidas a partir de uma DT e o método do Triângulo de Duval. A percentagem de erro de dados é calculada com base em 146 casos no conjunto de treino, e 16 casos no conjunto de validação. Analisando-se os resultados na Tabela 18, a DT associada com a pasta de número 2 demonstra o melhor desempenho em termos dos conjuntos de treino e de validação, inclusive também comparando com o método do Triângulo. Assim, o conjunto de regras obtido pela DT para o caso número 2 de validação cruzada é considerado como o conjunto de regras final que permite a análise dos transformadores. A Tabela 19 resume as 14 regras extraídas para esse melhor caso.

A Tabela 20 mostra uma comparação entre o erro de classificação quando são aplicadas as regras indicadas na Tabela 19 para cada grupo de dados considerados nas pastas de validação cruzada, e o erro obtido pelo método do Triângulo de Duval para os mesmos dados. Notadamente, percebe-se a maior eficiência quando comparado com o método do Triângulo.

**Tabela 19 - Regras extraídas**

Diagnóstico	C2H2%	C2H4%	C2H4/CH4	CH4/H2	C2H2/H2	C2H2/C2H4	CH4	C2H6
PD	< 12,9		< 0,01				>= 6,75	
T1 ou T2	< 12,9	< 23,35	< 0,01				< 6,75	
T1 ou T2	>= 12,9	< 23,35						>= 2345
T1 ou T2	< 12,9		> 1,25	< 0,08				
T1 ou T2	< 12,9		>= 0,01 e < 1,25	< 135	< 0,02			
T1 ou T2	< 12,9		>= 0,01 e < 1,25	< 135	>= 0,02	>= 0,03		
T3	< 12,9		>= 0,01	> 135				
T3	< 12,9		>= 1,25	>= 0,08 e < 135				
T3	< 12,9		>= 0,01 e < 1,25	< 135	>= 0,02	< 0,03		
D1	>= 12,9	< 23,35						< 2345
D1	>= 12,9	>= 23,35		< 0,05				
D1	>= 12,9	>= 23,35		>= 0,36				< 3
D2	>= 12,9	>= 23,35		>= 0,05				>= 3
D2	>= 12,9	>= 23,35		> 0,05 e < 0,36				< 3

Fonte: Elaborada pelo autor.

**Tabela 20 - Erro percentual para os conjuntos de testes**

Conjunto \ Pasta	Pasta									
	1	2	3	4	5	6	7	8	9	10
Teste (Regras extraídas AD)	0	6,25	0	0	0	0	6,25	6,25	0	0
Teste (Método do Triângulo)	12,5	12,5	12,5	12,5	12,5	12,5	12,5	12,5	18,75	18,75

Fonte: Elaborada pelo autor.

Na próxima seção os resultados obtidos com as regras desenvolvidas (Tabela 19) serão comparados com vários dos métodos indicados nas normas, além do método do Triângulo de Duval já comparado nesta seção.

### 5.3. Avaliação Comparativa das Regras do Método DGA Proposto

Observando a Tabela 19, pode-se verificar que apesar de 21 atributos terem sido inicialmente considerados na formação das árvores de decisão, apenas 8 deles permaneceram

nas regras geradas. Isso ocorre porque o algoritmo C4.5 considera apenas os atributos que têm maior influência na saída desejada. Na verdade, o C4.5 determinou que, a partir dos casos de treinamento considerados, os atributos ( $C_2H_2\%$ ,  $C_2H_4\%$ ,  $C_2H_4/CH_4$ ,  $CH_4/H_2$ ,  $C_2H_2/H_2$ ,  $C_2H_2/C_2H_4$ ,  $CH_4$  e  $C_2H_6$ ) são os mais significativos para o diagnóstico de falhas incipientes em equipamentos isolados a óleo.

Na Tabela 21, são apresentadas as percentagens de sucesso no diagnóstico de falhas utilizando as regras extraídas e o método do Triângulo para cada subconjunto de dados utilizados no desenvolvimento deste trabalho. Já na Tabela 22, são mostradas, a título de comparação, as percentagens de sucesso no diagnóstico de falhas utilizando as regras extraídas e diversos métodos listados nas normas para a base de dados IET TC 10.

**Tabela 21 - Percentagens de acerto para diversos conjuntos de dados**

Base de Dados	Triângulo de Duval (%)	Regras Extraídas (%)
IET TC 10 Database (117 casos)	88,03	99,15
Tabelas I - III (39 casos) em (Duval, 2002)	82,05	94,87
6 casos de (NARESH, SHARMA e VASHISTH, 2008)	83,33	100

Fonte: Elaborada pelo autor.

**Tabela 22 - Percentagens de acerto, comparando diversos métodos e utilizando IET TC 10**

Método	Acerto (%)
Método proposto (Regras DT)	99,15
Dörnenburg IEEE C57.104 - 1991	66,67
Dörnenburg IEEE C57.104 - 2008	67,52
Rogers IEEE C57.104 - 2008	56,41
Triângulo de Duval (IEC 599 - 1999)	88,03

Fonte: Elaborada pelo autor.

Ao analisar os resultados nas Tabelas 21 e 22, pode-se concluir que as regras descritas na Tabela 19 apresenta uma taxa de acerto mais elevada do que a demonstrada por outros métodos, apesar do pequeno número de regras e atributos utilizados. Considerando somente os dados da IEC TC 10, a taxa de sucesso atinge 99,15% contra 88,03% para o método do Triângulo.

Outra vantagem importante das regras extraídas da DT é a de que não existem intervalos sem diagnóstico, como acontecem em versões dos métodos de Rogers e de Dörnenburg citados nas normas, isso fica mais evidente construindo as tabelas de confusão para cada método.

Nas Tabelas 23, 24, 25, 26, 27 e 28, são apresentadas as matrizes de confusão utilizando o método proposto e os demais métodos adotados pelas normas e indicados na Tabela 22, utilizando o banco de dados IET TC 10. A primeira coluna nas tabelas representa o diagnóstico esperado utilizando os dados do IET TC 10, e a primeira linha representa os rótulos de diagnóstico possíveis de serem obtidos pelo método específico.

Em cada tabela, as legendas dos diagnósticos foram adaptadas conforme a nomenclatura adotada por cada método. Assim, para o método de Dörnenburg (ZIRBES, ROLIM e ZÜRN, 2005) os diagnósticos do IEC TC 10 relacionados à T1/T2 e T3 são denominados “Ponto quente”, para DP são “Descarga parcial” e para D1 e D2 são “Outras descargas”. Para o método Dörnenburg IEEE C57.104 nas versões 1991 e 2008, os diagnósticos do IEC TC 10 para T1/T2 e T3 são denominados “Decomposição Térmica”, já para DP são “Descargas de baixa energia” e para D1 e D2 são “Descargas de alta energia”. Em relação ao método de Rogers IEEE C57.104 – 2008, os diagnósticos tipo DP são rotulados como “Descargas de baixa energia”, já D1 e D2 são “Descargas de alta energia”, os valores tipo T1/T2 são “Falha térmica com temperatura < 700 °C”, e T3 são “Falha térmica com temperatura > 700 °C”. Os métodos do Triângulo e as regras propostas apresentam nomenclatura já compatível com os rótulos do IEC TC 10.

**Tabela 23 - Matriz de confusão utilizando Dörnenburg (ZIRBES, ROLIM e ZÜRN, 2005)**

	Ponto quente	Descarga parcial	Outras descargas	Ñ Classificado
Ponto quente	20	2	0	12
Descarga parcial	0	2	0	7
Outras descargas	0	0	52	22

Fonte: Elaborada pelo autor.

**Tabela 24 - Matriz de confusão utilizando Dörnenburg IEEE C57.104 - 1991**

	Decomposição térmica	Descargas de baixa energia	Descargas de alta energia	Ñ Classificado
Decomposição térmica	20	2	0	12
Descargas de baixa energia	0	2	0	7
Descargas de alta energia	0	0	51	23

Fonte: Elaborada pelo autor.

**Tabela 25 - Matriz de confusão utilizando Dörnenburg IEEE C57.104 - 2008**

	Decomposição térmica	Descargas de baixa energia	Descargas de alta energia	Ñ Classificado
Decomposição térmica	20	2	0	12
Descargas de baixa energia	0	2	0	7
Descargas de alta energia	0	0	52	22

Fonte: Elaborada pelo autor.

**Tabela 26 - Matriz de confusão utilizando Rogers IEEE C57.104 - 2008**

	Normal	Descarga baixa	Descarga alta	F. Térmica < 700°C	F. Térmica > 700°C	Ñ classificado
Descarga baixa	3	2	0	0	0	4
Descarga alta	0	0	45	0	0	29
F. Térmica < 700°C	0	1	0	5	3	7
F. Térmica > 700°C	0	0	1	2	11	4

Fonte: Elaborada pelo autor.

**Tabela 27 - Matriz de confusão utilizando o método do Triângulo**

	PD	D1	D2	T1 / T2	T3	DT
PD	9	0	0	0	0	0
D1	0	21	5	0	0	0
D2	0	1	47	0	0	0
T1 / T2	0	0	0	14	2	0
T3	2	0	0	3	12	1

Fonte: Elaborada pelo autor.

**Tabela 28 - Matriz de confusão utilizando as regras propostas**

	PD	D1	D2	T1 / T2	T3
PD	9	0	0	0	0
D1	0	26	0	0	0
D2	0	1	47	0	0
T1 / T2	0	0	0	16	0
T3	0	0	0	0	18

Fonte: Elaborada pelo autor.

Comparando as Tabelas de 23 a 28, mais uma vez fica evidente a maior eficiência do método proposto nesta pesquisa. Na Tabela 28, a matriz de confusão para as regras propostas apresenta apenas um valor fora da diagonal principal, indicando, portanto apenas um erro. Também, para as regras propostas não existem situações de não classificação (não diagnóstico) como acontecem nos métodos das Tabelas de 23 a 26.

Portanto, as regras obtidas neste trabalho além de apresentarem uma eficiência maior do que outros métodos, também apresenta um número reduzido de regras, 14 regras e 8 atributos, utilizando operadores lógicos tradicionais E e OU, sendo de fácil interpretação.

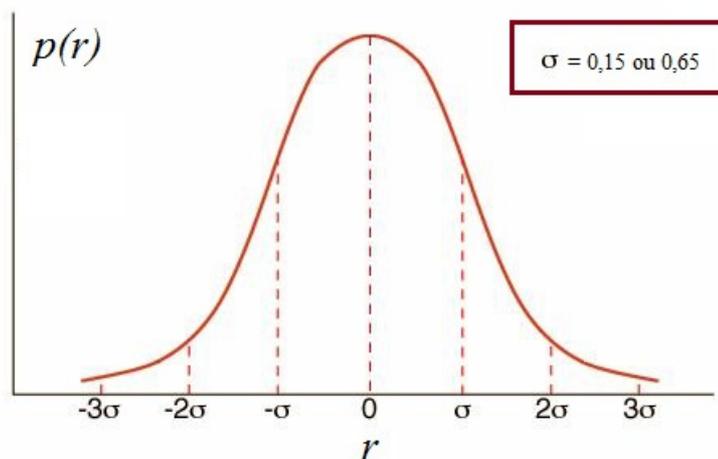
Na próxima seção será examinada a robustez do método proposto em relação a dados de entrada com erros.

#### 5.4. Sensibilidade a Erros de Medida do Método DGA Proposto

Conforme comentado no Capítulo 2, no processo de coleta e análise de gases em testes cromatográficos existe a possibilidade da ocorrência de erros que resultam em valores distorcidos para os gases medidos no óleo isolante. Nesta seção será realizado um estudo simplificado para avaliar a sensibilidade das regras de diagnóstico propostas neste trabalho quando da ocorrência de tais erros, também será realizando uma comparação com alguns métodos indicados nas normas.

O erro nas medidas dos gases é modelado através de uma distribuição normal (Figura 23), considerando um desvio padrão de 15% e de 65%, respectivamente relacionados com as variações de erro médias e máximas registradas para um conjunto de medidas, obtidas em estudos realizados de verificação de laboratórios que realizam DGA (seção 2.6, Capítulo 2). A distribuição normal é então somada com os valores corretos, gerando um novo banco de dados de teste, no caso com a presença de erro de medida simulado.

Figura 23 - Distribuição normal



Fonte: Adaptado de (LEBENSZTAYN, 2012).

A equação (5.1) apresenta a fórmula utilizada para representar o novo valor de medida com o acréscimo de um ruído, representando a existência de erros nos dados:

$$\mathbf{h}' = \mathbf{h} + \mathbf{r} \cdot \mathbf{h} \quad (5.1)$$

sendo  $\mathbf{h}'$  o novo vetor com dados de gases com a presença de erros,  $\mathbf{h}$  é o vetor original sem erros dos valores de gases e  $\mathbf{r}$  é o vetor de ruídos obtido de uma distribuição normal.

A Tabela 29 apresenta a percentagem de sucesso no diagnóstico de falhas incipientes usando as regras obtidas a partir da DT desenvolvida, o método do Triângulo, o método de Dörnenburg IEEE C57.104 – 2008 e o método da razão de Rogers IEEE C.57.104 - 2008, em situações de ocorrência de erros nas medições dos gases dissolvidos no óleo isolante do transformador. O banco de dados utilizado nos testes foi o IEC TC 10.

**Tabela 29 – Taxa de acerto quando da possibilidade de erro nas medidas de gás (IEC TC 10)**

Método	Sem erro	Erro com 15%	Erro com 65%
Rogers IEEE C57.104 - 2008	56,41	55,56	52,14
Dörnenburg IEEE C57.104 - 2008	67,52	67,52	64,96
Triângulo	88,03	85,47	77,78
Regras extraídas	99,15	95,73	82,91

Fonte: Elaborada pelo autor.

Analisando os resultados apresentados na Tabela 29, observa-se que o método desenvolvido nesta Tese apresenta uma melhor robustez para uma distribuição de erros nos dados de entrada, com uma taxa de acerto de 82,91% no pior caso (erro com desvio padrão de 65 %). Entretanto, os métodos de Rogers e Dörnenburg apresentam a menor variação percentual, comparando com o valor sem erro.

Na próxima seção será apresentado outro método de diagnóstico, entretanto agora mais simplificado nas indicações de diagnóstico e que utiliza dados físico-químicos, que também foi desenvolvido adotando a mesma metodologia indicada no Capítulo 4.

## **5.5. Método Simplificado de Diagnóstico de Transformadores Utilizando Dados Físico-Químicos**

No desenvolvimento do método simplificado de diagnóstico de transformadores utilizando dados de análises físico-químicas do óleo isolante foi utilizada a mesma metodologia apresentada no Capítulo 4, adotada nas seções anteriores no desenvolvimento do método de diagnóstico utilizando dados de cromatografia.

Os dados utilizados para o desenvolvimento e teste das regras de diagnóstico utilizando informações físico-químicas são os mesmos utilizados em (ROCHA, 2013), onde foi desenvolvida uma RNA para o diagnóstico de transformadores utilizando dados físico-químicos. Do total de 135 amostras utilizadas com exemplos de análises físico-químicas realizadas em transformadores de potência e com o diagnóstico do estado do transformador realizado por um especialista, 33 apresentam diagnóstico de estado normal, 30 o diagnóstico de falhas térmicas e 72 indicam falhas elétricas.

Na obtenção de uma DT com a metodologia de inicialmente aumentarmos os atributos de entrada, foram utilizados como atributos de entrada seis informações físico-químicas: Acidez (Ac), Rigidez Dielétrica (RD), Teor de Água (TA), Tensão Interfacial (TI), Densidade (D) e Fator de Perdas dielétricas a 100° C (FP100), e mais as seguintes relações:  $Ac/RD$ ,  $Ac/TA$ ,  $Ac/TI$ ,  $Ac/D$ ,  $Ac/FP100$ ,  $RD/TA$ ,  $RD/TI$ ,  $RD/D$ ,  $RD/FP100$ ,  $TA/TI$ ,  $TA/D$ ,  $TA/FP100$ ,  $TI/D$ ,  $TI/FP100$ ,  $D/FP100$ . Outras relações foram testadas, mas não demonstraram a obtenção de melhorias. Também, esses novos atributos, representam relações entre os atributos originais, atendendo as exigências indicadas na Seção 4.3. Então inicialmente foram considerados 22 atributos de entrada a serem apresentados ao algoritmo de construção da DT.

Utilizando validação cruzada, foram construídas 10 árvores de decisão através do *software* SIPINA utilizando os dados e rótulos citados anteriormente. Em cada caso de árvore (rodada de validação) foram utilizados 90% dos dados (122 dados) para o desenvolvimento da árvore e os 10% restantes (13 dados) para validar a árvore construída. Ressaltando que os dados foram distribuídos de maneira proporcional quanto ao diagnóstico indicado pelo especialista. Os resultados são apresentados na Tabela 30. Para comparação também são apresentados na tabela, os resultados para o caso das regras obtidas utilizando apenas como entradas as informações físico-químicas iniciais: Ac, RD, TA, TI, D e FP100.

**Tabela 30 - Percentual de erro, considerando os casos para validação cruzada (dados físico-químicos)**

Pasta	1	2	3	4	5	6	7	8	9	10
Treino Regras Ent. Expandidas	18,85	11,48	5,74	14,75	0,82	4,92	4,1	0	4,1	10,67
Valid. Regras Ent. Expandidas	46,15	53,85	46,15	53,85	30,77	38,46	53,85	46,15	53,85	38,46
Treino Regras Ent. Normais	14,75	6,56	5,74	13,93	3,28	16,39	12,3	5,74	5,74	8,2
Valid. Regras Ent. Normais	38,46	53,85	46,15	53,85	46,15	61,54	38,46	38,46	46,15	46,15

Fonte: Elaborada pelo autor.

Analisando a Tabela 30, foi determinado que o melhor caso foi o da pasta de número 5 utilizando 22 atributos de entrada (entradas expandidas) pois apresenta a maior percentagem de acerto para o conjunto de validação (69,23%) e o segundo melhor resultado para o conjunto de treino (99,2%). Sendo um caso melhor que os observados para as situações de utilização de entradas normais ou expandidas. As regras obtidas para o caso escolhido são apresentadas na Tabela 31.

**Tabela 31 - Regras obtidas utilizando dados físico-químicos**

Diagnóstico	FP100	RD	TI	TA	Ac	Ac/TI	TI/FP100	RD/TI	D	RD/TA	TA/TI
Normal	$\geq 4,31$	$\geq 83,90$									
Normal	$\geq 4,31$	$< 83,90$	$< 26,6$	$< 18$							
Normal	$< 4,31$		$< 29,75$	$< 32,6$		$< 0$					
Normal	$< 4,31$		$\geq 29,75$				$\geq 20,8$	$\geq 2,31$	$\geq 0,87$		
Normal	$< 4,31$		$\geq 29,75$				$\geq 20,8$ e $< 23,67$	$< 2,31$	$< 6,77$		
Normal	$< 4,31$		$\geq 29,75$				$\geq 23,67$	$< 2,31$	$< 6,77$	$< 0,27$	
Normal	$< 4,31$	$\geq 29,75$ e $< 37,45$	$< 0,03$				$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,49$	
Normal	$< 4,31$	$\geq 80,25$	$\geq 37,45$	$< 0,03$			$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,49$	
F. Térmica	$< 4,31$		$\geq 29,75$				$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,46$ e $< 0,49$	
F. Térmica	$< 4,31$		$\geq 29,75$	$\geq 0,03$ e $< 0,035$			$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,49$	
F. Térmica	$\geq 4,31$	$< 83,90$	$\geq 26,6$								
F. Térmica	$\geq 4,31$	$< 83,90$	$< 26,6$	$\geq 18$	$\geq 0,15$						
F. Térmica	$< 4,31$	$\geq 88,40$	$\geq 29,75$ e $< 38,25$				$< 20,8$	$\geq 2,31$			
F. Térmica	$< 4,31$	$< 83,90$	$\geq 38,25$				$< 20,8$	$\geq 2,31$			
F. Térmica	$\geq 0,32$ e $< 4,31$		$\geq 46,55$				$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,27$ e $< 0,46$	
F. Elétrica	$\geq 4,31$	$< 83,90$	$< 26,6$	$\geq 18$	$< 0,15$						
F. Elétrica	$< 4,31$		$< 29,75$	$\geq 32,6$							
F. Elétrica	$< 4,31$		$< 29,75$	$< 32,6$		$\geq 0$					
F. Elétrica	$< 4,31$		$\geq 29,75$				$\geq 20,8$	$\geq 2,31$	$< 0,87$		
F. Elétrica	$< 4,31$	$< 88,40$	$\geq 29,75$ e $< 38,25$				$< 20,8$	$\geq 2,31$			
F. Elétrica	$< 4,31$	$\geq 83,90$	$\geq 38,25$				$< 20,8$	$\geq 2,31$			
F. Elétrica	$< 4,31$		$\geq 29,75$				$\geq 20,8$	$< 2,31$	$\geq 6,77$		
F. Elétrica	$< 4,31$		$\geq 29,75$	$\geq 0,035$			$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,49$	
F. Elétrica	$< 4,31$	$< 80,25$	$\geq 37,45$	$< 0,03$			$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,49$	
F. Elétrica	$< 4,31$		$\geq 29,75$ e $< 46,55$				$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,27$ e $< 0,46$	
F. Elétrica	$< 0,32$		$\geq 46,55$				$\geq 23,67$	$< 2,31$	$< 6,77$	$\geq 0,27$ e $< 0,46$	

Fonte: Elaborada pelo autor.

Analisando as regras obtidas, é verificado que do conjunto inicial de 22 atributos, a DT gerou 26 regras avaliando apenas 11 atributos de entrada que foram selecionados pela DT como mais significativos.

Na Tabela 32 são apresentados os valores percentuais de erro de classificação quando são aplicadas as regras indicadas na Tabela 31 para cada grupo de dados considerados nas pastas de validação cruzada.

**Tabela 32 - Erro percentual para os conjuntos de testes (dados físico-químicos)**

Pasta	1	2	3	4	5	6	7	8	9	10
Validação (%)	0	7,69	0	0	30,77	0	0	0	0	0

Fonte: Elaborada pelo autor.

Realizando um estudo de sensibilidade de erros, conforme já demonstrado na seção 5.4, as regras indicadas na Tabela 32 foram submetidas a erros nas medidas físico-químicas modelado por uma distribuição normal de 15% e de 65%. Os resultados são apresentados na Tabela 33, junto com o comparativo do resultado obtido através de uma DT construída com as seis entradas físico-química originais. Foram utilizados todas as 135 amostras do conjunto considerado (BARBOSA, 2013). Também é apresentado nessa tabela o percentual de acerto para o total de dados sem considerar a ocorrência de erros nas medidas.

**Tabela 33 - Taxa de acerto quando da possibilidade de erro nas medidas físico-químicas**

Método	Sem erro	Erro com 15%	Erro com 65%
Regras com entradas expandidas	96,30	58,52	40
Regras com entradas originais	92,60	57,04	45,19

Fonte: Elaborada pelo autor.

Apesar do resultado para o conjunto de dados sem erro nas medidas de entrada ter apresentado um valor elevado de acerto, quando da ocorrência de erros nas medidas a percentagem de acerto diminui de forma significativa para as regras obtidas tanto para o caso do uso de atributos de entrada estendidos como no caso do uso dos atributos originais. As regras com a entrada de atributos ampliada apresenta um melhor resultado para os casos sem erro e com erro modelado por um desvio padrão de 15%.

Os resultados apresentados, apesar de não serem tão significativos quanto os verificados para o caso da utilização de informações cromatográficas (Tabela 29), se mostraram promissores. Ainda mais considerado que a intenção é o desenvolvimento de um método simplificado de diagnóstico que utiliza dados físico-químicos que são mais fáceis de serem medidos, inclusive podendo ter as medições realizadas por profissionais treinados das equipes de manutenção das empresas concessionárias de energia elétrica sem a dependência de um laboratório externo.

## 5.6. Considerações Parciais

Neste capítulo é apresentada a aplicação da proposta do Capítulo 4 de obtenção de Regras interpretáveis a partir do aumento do espaço de atributos no Problema de Diagnóstico de falhas incipientes em equipamentos elétricos que utilizam óleo isolante. Os atributos

originais considerados são de dois tipos: medidas cromatográficas (concentrações de gases) e medidas físico-químicas do óleo isolante. Do ponto de vista metodológico, a ideia de usar um grande número de atributos de entrada que explora relações entre as concentrações dos gases (ou as medidas físico-químicas) permite a obtenção de métodos de diagnóstico de alta eficiência baseado em DT que produz regras interpretáveis.

As regras obtidas com o Método DGA proposto alcançaram uma grande precisão no diagnóstico, com um erro inferior a 1%, conforme indicado na Tabela 22, no conjunto completo de dados (Seção 5.1) – e precisão comparável ou superior ao Método de Duval nos subconjuntos de dados considerados na validação cruzada (Tabelas 18 e 20). Considerando a possibilidade da ocorrência de erros nas medições de gás para um cenário de desvio padrão de 65%, o método proposto conseguiu quase 83% de acerto, sendo então considerado superior aos outros métodos DGA no quesito sensibilidade a erros para os dados utilizados.

As regras obtidas para o Método de Diagnóstico Proposto a partir dos atributos físico-químicos também foram analisadas neste capítulo. Como esperado, pela natureza menos informativa das medidas físico-químicas (FQ) do óleo isolante, a precisão do diagnóstico FQ mostrou-se inferior ao diagnóstico cromatográfico (Tabela 30). Porém, o estudo inicial apresentado reforça a observação de melhoria na classificação ao se adotar um aumento no espaço de atributos baseado em relações entre os atributos originais: a Tabela 33 resume a comparação entre o diagnóstico realizado com regras obtidas com entradas expandidas e com entradas originais sobre todo o conjunto de dados disponível, comprovando a melhoria no percentual de acertos. A maior sensibilidade a erros de medida observada no método proposto de diagnóstico baseado em medidas FQ (Tabela 33), em comparação com o método DGA proposto (Tabela 29), aponta a necessidade de uma maior investigação para definir a causa dessa sensibilidade – o reduzido conjunto utilizado no estudo para obtenção do DT pode ser uma das causas.

## 6. CONCLUSÕES E PERSPECTIVAS

Para o diagnóstico de equipamentos com isolamento a óleo, como os transformadores, são amplamente utilizados os métodos baseados na análise de DGA, que consideram na análise o tipo, a concentração e a taxa de produção de gases gerados e dissolvidos no óleo do equipamento, e procuram associar o tipo de falha ao gás presente.

Na realização do diagnóstico adequado de equipamentos com óleo isolante, por meio de análise de DGA, o especialista normalmente emprega um ou mais métodos padronizados, que são descritos nas normas (IEEE C57.104, 2008; IEC 60599, 2008). No entanto, tais métodos apresentam severas limitações: (i) existência de situações que não são abrangidas pelos critérios, (ii) a existência de mais de um diagnóstico possível, ou (iii) resultados incorretos.

Uma alternativa é utilizar técnicas de ML associadas com extração de regras de conhecimento, para obtenção de um diagnóstico mais eficiente de equipamentos com isolamento a óleo. Entretanto, várias das técnicas de ML empregadas falham na obtenção de tais regras, em questões como precisão e/ou capacidade de interpretação das regras obtidas.

Este trabalho de pesquisa apresentou um novo método de diagnóstico de falhas incipientes em equipamentos de potência com isolamento a óleo, como os transformadores de potência. O novo método utiliza uma árvore de decisão (*Decision Tree* – DT) com modificações propostas na dimensão dos atributos de entrada, permitindo obter um conjunto de regras com uma precisão maior no diagnóstico que os métodos descritos em normas, e de fácil utilização.

Durante o desenvolvimento do trabalho, inicialmente foi realizada uma pesquisa bibliográfica extensa sobre os temas: diagnóstico de equipamentos elétricos através de DGA e extração de regras a partir de métodos de ML. A ideia foi utilizar alguma técnica de ML (RNA – Rede Neural Artificial, SRF – Sistema de Regras Fuzzy, MVS - Máquina de Vetor de Suporte – MVS, e DT) para superar as dificuldades apresentadas pelos métodos clássicos no diagnóstico de transformadores de potência. Como várias das técnicas de ML, como RNA, apresentam dificuldades próprias de tais métodos relacionados à compreensão dos resultados obtidos, foram pesquisados vários métodos de extração de regras a serem utilizadas junto com as técnicas de ML no diagnóstico de transformadores.

Optou-se por utilizar uma DT que apresenta as vantagens: de efetuar a seleção de requisitos e a obtenção direta de regras de conhecimento. Entretanto, as simulações iniciais não apresentavam resultados superiores aos métodos tradicionais de diagnóstico de transformadores. Optou-se então por utilizar uma estratégia de aumentar a dimensão inicial do problema através da inclusão de novas variáveis relacionadas ao problema ou de relações dessas variáveis novas e das já tradicionalmente utilizadas no diagnóstico de equipamentos utilizando DGA. Essa nova estratégia é baseada no trabalho de Cover (COVER, 1965) que argumenta que o aumento de dimensão de um problema aumenta a probabilidade da obtenção de soluções lineares de problemas não lineares, também é baseado no fato que entre as novas variáveis existem relações entre variáveis originais, o que acaba permitindo que a árvore construa nós pela avaliação de relações lineares de variáveis, numa estratégia parecida com a das árvores obliquas. O resultado é a obtenção de uma DT, projetada para o diagnóstico de transformadores, com alta eficiência e apresentando um número reduzido de regras.

Ainda no desenvolvimento das regras utilizando DT teve-se o cuidado de utilizar validação cruzada, de forma a obter uma solução (conjunto de regras) com maior grau de confiança estatística. Também, os resultados de comparação mostraram uma eficiência muito superior das regras desenvolvidas quando comparado com as regras indicadas nas normas técnicas. Os dados utilizados tanto no desenvolvimento das regras como também na comparação com os métodos normatizados, são de conjuntos públicos, e o maior conjunto já é indicado pelo IEC no aprimoramento dos métodos clássicos.

Também foram realizados estudos relacionados à possibilidade de ocorrência de erros nas medidas de DGA. Foram desenvolvidos testes comparando as regras propostas com os métodos clássicos. O erro foi modelado utilizando uma distribuição de forma Gaussiana.

Como vantagens mais importantes das regras desenvolvidas podem ser citadas:

- Facilidade de utilização, pois apresenta um número reduzido de regras (14 regras e 8 atributos) utilizando apenas os operadores lógicos clássicos E e OU.
- Alta eficiência no diagnóstico de equipamentos, alcançando uma eficiência de 99% contra 88% do método do Triângulo, utilizando na comparação o banco de dados IEC TC 10.

- Apresenta um valor reduzido de erro, com uma eficiência de 82%, mesmo utilizando dados de DGA associados com erros de valores significativos.

Como ressalva apresentada, pode ser citado o fato que nos testes com erros associados às medidas de DGA, outros métodos foram melhores no quesito de variação relativa. Entretanto, como já citado antes, as regras propostas nesse trabalho ainda apresentam o maior valor absoluto de precisão.

Ao final do trabalho foi apresentado um estudo inicial para o desenvolvimento de um método de diagnóstico simplificado de transformadores utilizando dados físico-químicos. Os resultados apresentados a pesar de não serem mostrarem tão bons quanto os verificados para o caso da utilização de informações de gases dissolvidos no óleo isolante do equipamento, mostraram-se promissoras. Ainda mais considerado que a intenção foi o desenvolvimento de um método simplificado de diagnóstico que utiliza dados físico-químicos que são mais fáceis de serem obtidos regularmente, inclusive podendo as medições serem realizadas por profissionais treinados das equipes de manutenção das empresas de energia elétrica.

## **6.1. Trabalhos Futuros**

Como proposta de trabalho futuro, relacionado à Tese apresentada, é indicada o aprimoramento dos estudos para obtenção de regras de diagnóstico de equipamentos com óleo isolante através de medidas físico-químicas, utilizando novas entradas com razões de valores normalizados e aumentando o número de medidas do banco de dados de treino e teste. A vantagem da utilização de dados físico-químicos para o diagnóstico de equipamentos com óleo isolante deve-se ao fato que os valores físico-químicos são normalmente mais fáceis de serem obtidos do que os dados de cromatografia dos gases.

Outra proposta é a introdução de informações históricas, além das medidas de DGA e/ou físico-químicas, através de informações de diagnósticos anteriores do equipamento avaliado, de forma a considerar a evolução temporal dos fenômenos no interior dos equipamentos elétricos de potência com isolamento a óleo, como os transformadores. A intenção é permitir, através dessas informações históricas, a identificação da evolução dos tipos de falha no interior do equipamento ou mesmo permitir o diagnóstico da ocorrência de falhas simultâneas, aumentando a qualidade do monitoramento e diagnóstico do equipamento avaliado.

## REFERÊNCIAS

ABDI, S.; BOUBAKEUR, A.; HADDAD, A. Influence of thermal ageing on transformer oil properties, IEEE International Conference on Dielectric Liquids (ICDL 2008), pp. 1-4, 2008.

ABNT NBR 7274: Interpretação da Análise dos Gases de Transformadores em Serviço. Rio de Janeiro, 1982.

AIZERMAN, M. A.; BRAVERMAN, E. M.; e ROZONOER, L. I. Theoretical foundations of the potential function method in pattern recognition learning. *Automatika i Telemekhanika*, vol 25, Jun 1964; translation published Jan 1965, pp 821-837.

ALMUALLIM, H. An efficient algorithm for optimal pruning of decision trees. *Artif. Intell.*, vol. 83, no. 2, pp. 347–362, 1996.

AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S.; BARBOSA, F. R.; LIMA, S. S.; LISBOA L. A. C. Extraction of Knowledge from Artificial Neural Networks to Application in Analysis of Transformers. The 15th International Conference on Intelligent System Applications to Power Systems ISAP 2009, Curitiba, 2009.

AMORA, M. A. B.; ALMEIDA, O. M.; BRAGA, A. P. S.; BARBOSA, PONTES, R. S. T. An improved DGA method based on rules extracted from a high-dimension input space. *Electronic Letters*, Volume 48, Issue 17, 16 August 2012, p. 1048-1049.

ATALLAH, M. J. *Algorithms and Theory of Computation Handbook*. CRC Press LLC, 1999.

BARAKAT, N.; e BRADLEY, A. Rule Extraction from Support Vector Machines: A Sequential Covering Approach. *IEEE Trans. Knowledge and Data Eng.*, vol. 19, no. 6, pp. 729-741, June 2007.

BARAKAT, N.; e DIEDERICH, J. Eclectic rule-extraction from support vector machines. *International Journal of Computational Intelligence*, 2(1):59-62, 2005.

BARBOSA, F. R. Diagnóstico de falhas incipientes a partir das propriedades físico-químicas do óleo isolante em transformadores de potência como método alternativo à análise de gases

dissolvidos. Tese (Doutorado em Engenharia Elétrica) – Centro de Tecnologia. Universidade Federal do Ceará, Fortaleza, 2013.

BARBOSA, F. R. Monitoramento e diagnóstico inteligente da qualidade dielétrica do isolamento líquido de transformadores de potência. Dissertação (Mestrado em Engenharia Elétrica) – Centro de Tecnologia. Universidade Federal do Ceará, Fortaleza, 2008.

BARBOSA, F. R.; ALMEIDA, O. M.; BRAGA, Arthur P. S.; AMORA, M. A. B.; CARTAXO, S. J. M. Application of an Artificial Neural Network in the Use of Physicochemical Properties as a Low Cost Proxy of Power Transformers DGA Data. IEEE Transactions on Dielectrics and Electrical Insulation, vol. 19, no. 1, pp. 239 – 246, February 2012.

BELLMAN, R. Adaptive Control Processes: A Guided Tour. Princeton University Press, 1961.

BENGTSSON, C.; Status and Trends in Transformer Monitoring, IEEE Transactions on Power delivery, vol. 11, No. 3, 1996.

BENÍTEZ, J. M.; CASTRO, J. L.; e REQUENA, I. Are artificial neural networks black boxes? IEEE Trans. Neural Networks, vol. 8, pp. 1156–164, Sept. 1997.

BENNETT, P.; e MANGASARIAN, O. L. Multicategory discrimination via linear programming. Optimization Meth. Softw., vol. 3, pp. 29–39, 1994.

BISHOP, A. B. Adaptive pattern recognition, 1963 WESCON Rept of Session 1.5, unpublished.

BLACK, P. E. Dictionary of Algorithms and Data Structures. U.S. National Institute of Standards and Technology (NIST).

BOND, C. F.; LASHLEY, B. R. Round-robin analysis of social interaction: Exact and estimated standard errors. Psychometrika, 61, 303-311, 1996.

BOZ, O. Converting A Trained Neural Network To A Decision Tree DecText - Decision Tree Extractor. PhD thesis, Lehigh University, Department of Computer Science and Engineering, 2000.

BRATKO, I.; e BOHANEK, M. Trading accuracy for simplicity in decision trees. *Mach. Learn.*, vol. 15, pp. 223–250, 1994.

BREIMAN, L.; FRIEDMAN, J.H.; OLSEN, R.A.; e STONE, C.J. *Classification and Regression Trees*. Wadsworth and Brooks, 1984.

CARGOL, T. A. Overview of On-Line Oil Monitoring Technologies. Fourth Annual Weidmann-ACTI Technical Conference, San Antonio, USA, 2005.

CASTRO FILHO, C. A. P. *Árvore de Decisão Oblíqua*. X WORCAP. Instituto Nacional de Pesquisas Espaciais (INPE). São José dos Campos, SP, Brasil, 2010.

CASTRO, J. L.; FLORES-HIDALGO, L. D.; MANTAS, C. J.; e PUCHEA, J. M. Extraction of fuzzy rules from support vector machines. *Elsevier Fuzzy Sets and Systems* 158 (2007) 2057 – 2077.

CASTRO, J.L.; MANTAS, C.J.; BENÍTEZ, J.M. Interpretation of artificial neural networks by means of fuzzy rules, *IEEE Trans. Neural Networks* 13 (1) (2002) 101–116.

CASTRO, R. G. C. *Knowledge Extraction from Artificial Neural Networks: Application to Transformer Incipient Fault Diagnosis*. Tese (Doutorado em Engenharia Elétrica). Universidade do Porto, Faculdade de Engenharia, Departamento de Engenharia Eletrotécnica, Porto – Portugal, 2004.

CAVALEIRO P.M., ARAÚJO, F.M., FERREIRA, L. A., SANTOS, J.L., FARAHI, F. Simultaneous Measurement of Strain and Temperature Using Bragg Gratings Written in Germanosilicate and Boron-Codoped-Germanosilicate Fibres, *IEEE Photonics Technology Letters*, 2003.

CHEIM, L.; VARELLA, V.; DUPONT, C.; PÁAL, R. Sistema Inteligente de Diagnóstico de Transformadores de Potência. In: XV SNPTEE, GSE 21. Anais. Paraná, Brasil, out. 1999.

CHEN, F. Learning accurate and understandable rules from SVM classifiers. *Dissertação de Mestrado*, Simon Fraser University, 2004.

CLARK, P.; e NIBLETT, T. The CN2 induction algorithm. *Machine Learning*, 3:261-283, 1989.

COVER, T.M. (1965). Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers EC-14*: 326–334.

CRAVEN, M.W.; e SHAVLIK, J. W. Extracting tree-structured representations of trained networks. In David S. Touretzky, Michael C. Mozer and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pp. 24-30. The MIT Press, 1996.

CRAVEN, M.W.; e SHAVLIK, J. W. Using sampling and queries to extract rules from trained neural networks. In *International Conference on Machine Learning*, pp. 37-45, 1994.

DIETTERICH, T. G.; KEARNS, M., e MANSOUR, Y. Applying the weak learning framework to understand and improve C4.5. *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 96-104, San Francisco: Morgan Kaufmann, 1996.

DUDA, R.; e HART, P. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

DUDA, R.O.; HART, P.E.; e STORK, D.G. *Pattern Classification*. 2nd ed. John Wiley & Sons. NY, 2001.

DUKARM, J. J. Transformer Oil Diagnosis Using Fuzzy Logic and Neural Networks. *Conference in Electrical and Computer Engineering, Canadian*, v. 1, p. 329-332, 1993.

DUVAL, M. A Review of Faults Detectable by Gas-in-Oil Analysis in Transformer. *IEEE Electrical Insulation Magazine*, Vol. 18, N° 3, pp. 8-17, May/June 2002.

DUVAL, M. New techniques for dissolved gas in oil analysis. *IEEE Elect. Insul. Mag.*, vol.19, no.2, pp. 6–15, 2003.

DUVAL, M.; DUKAM, J. Improving the Reliability of Transformer Gas-in-Oil Diagnosis. *IEEE Elect. Insul. Mag.*, vol.21, no.4, pp. 21–27, 2005.

DUVAL, M.; e PABLO, A. de. Interpretation of Gas-in-Oil Analysis Using New IEC Publication 60599 and IEC TC 10 Databases. *IEEE Electrical Insulation Magazine*, Vol. 17, N° 2, pp. 31-41, March/April 2001.

ESP, D.G.; McGRAIL, A. J. Data mining applied to transformer oil analysis data, IEE Digest, Vol. 1999 / no. 119, pp. 9/1–9/7.

FAYYAD, U. M.; e IRANI, K. B. The attribute selection problem in decision tree generation. in Proc. 10th Nat.Conf. Artificial Intelligence, Cambridge, MA, 1992, pp. 104–110.

FERGUSON, R.; LOBEIRAS, A.; SABOU, J. Suspended Particles in the Liquid Insulation of Aging Power Transformers. IEEE Electrical Insulation Magazine. v. 18, n. 4, pp. 17-23, 2002.

FERREIRA, L. A., SANTOS, J.L., FARAHI, F., DIATZIKIS, E.V.; Demodulation of Fibre Bragg Gratings Sensors Based on Dynamic Tuning of a Multimode Laser Diode", Applied Optics, 38, 4751, 1999.

FREITAS, A. A. C. (2000). Redes Neurais Artificiais Aplicadas em Transformadores de Distribuição Imersos em Óleo, Dissertação de Mestrado, UNESP/FE/DEE, Bauru-SP.

FRIEDMAN, J. H. Arecursive partitioning decision rule for nonparametric classifiers. IEEE Trans. Comput., vol. C26, no. 4, pp. 404–408, Apr. 1977.

FU, L. Rule learning by searching on adapted nets. In Ninth National Conference on Artificial Intelligence, pages 590-595, 1991.

FUNG, G.; SANDILYA, S.; e RAO, R. B. Rule extraction from linear support vector machines. In 11th ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 32-40, 2005.

GELFAND, S. B.; RAVISHANKAR, C. S.; e DELP, E. J. An iterative growing and pruning algorithm for classification tree design. IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, no. 2, pp. 163–174, Feb. 1991.

GREENBERG, H. J., e KONHEIM, A. G. Linear and nonlinear methods in pattern classification, IBM J. Res. Develop., vol 8, Jul 1964, pp 299-307.

GUARDADO, J. L.; NAREDO, J. L.; MORENO, P.; FUERTE, R. A Comparative Study of Neural Network Efficiency in Power Transformers Diagnosis Using Dissolved Gas Analysis. IEEE Transactions on Power Delivery, v. 12, n. 4, p. 643-647, 2001.

HALSTEAD, W. D. A thermodynamic assessment of the formation of gaseous hydrocarbons in faulty transformers. *J. Inst. Petrol*, Vol. 59, September 1959, PP. 239-241.

HANCOCK, T. R.; JIANG, T.; LI, M.; e TROMP, J. Lower bounds on learning decision lists and trees. *Inform. Comput.*, vol. 126, no. 2, pp. 114–122, 1996.

HAYKIN, S. *Redes neurais: princípios e práticas*. Bookman, 2001.

HOOSHMAND, R.A.; PARASTEGARI, M.; FORGHANI, Z. Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers. *Electrical Insulation Magazine*, IEEE Volume: 28. Issue: 5. Publication Year: 2012 , Page(s): 32- 42.

HUANG, Y. C, YANG, K., T; HUANG, K. Y. Abductive Network Model Based Diagnosis System for Power Transformer Incipient Fault Detection. *IEE Proceedings - Generation, Transmission and Distribution*. v. 149, n. 3, p. 326-330, May, 2002.

HUANG, Y. C.; YANG, H. T.; HUANG, C. L. Developing a New System Transformer Fault Diagnosis System through Evolutionary Fuzzy Logic. *IEEE Transactions on Power Delivery*, v. 12, n. 2, p. 761-767, 1997.

HUYSMANS, J.; BAESENS, B.; e VANTHIENEN, J. ITER: an algorithm for predictive regression rule extraction. In *8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006)*. Springer Verlag, Incs, 2006.

HUYSMANS, J.; BAESENS, B.; e VANTHIENEN, J. Using rule extraction to improve the comprehensibility of predictive models. *Katholieke Universiteit Leuven. Department of Decision Sciences and Information Management*. Leuven, Belgium, 2006.

HYAFIL, L.; e RIVEST, R. L. Constructing optimal binary decision trees is NP-complete. *Inform. Process. Lett.*, vol. 5, no. 1, pp. 15–17, 1976.

IEC Guide for the Sampling of Gases and of Oil from Oil-Filled Electrical Equipment and for the Analysis of Free and Dissolved Gases, IEC Publication 60567 (1992-07), 1992.

IEC Mineral Oil-Impregnated Electrical Equipment in Service – Guide to the Interpretation of Dissolved and Free Gases Analysis, IEC 60599 – 2008.

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

IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, IEEE Standard C57.104-2008, Sept. 2008.

IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, IEEE Std C57.104-1991, 1991.

IMAMURA, M. M.; SILVA, I. N.; SOUZA, A. N. Uma Abordagem Neural para Análise dos Gases Dissolvidos em Óleo Isolante em Função das Grandezas Físico-Químicas. Anais do XIII CONGRESSO BRASILEIRO DE AUTOMÁTICA (CBA 2000), Florianópolis, pp. 2288-2293, 2000.

ISLAM, S. M.; WU T.; LEDWICH, G. A Novel Fuzzy Logic Approach to Transformer Fault Diagnosis, IEEE Trans. on Dielectrics and Electrical Isolation, v. 7, n. 2, p. 177-186, 2000.

JOHANSSON, U.; KÄONIG, R.; e NIKLASSON, L. Rule extraction from trained neural networks using genetic programming. In Joint 13th International Conference on Artificial Neural Networks and 10th International Conference on Neural Information Processing, ICANN/ICONIP 2003, pp. 13-16, 2003.

JOHN, G. H. Robust linear discriminant trees. In D. Fisher and H. Lenz, editors, Learning From Data: Artificial Intelligence and Statistics V, Lecture Notes in Statistics, Chapter 36, pp. 375-385. Springer-Verlag, New York, 1996.

KAN, H.; MIYAMOTO, T. Proposals for an Improvement in Transformer Diagnosis Using Dissolved Gas Analysis (DGA). IEEE Electrical Insulation Magazine, vol. 11, no. 6, pp. 15-21, 1995.

KASS, G. V. An exploratory technique for investigating large quantities of categorical data. Appl. Statist., vol. 29, no. 2, pp. 119–127, 1980.

KEARNS, M.; e MANSOUR, Y. A fast, bottom-up decision tree pruning algorithm with near-optimal generalization. in Proc. 15th Int. Conf. Machine Learning, J. Shavlik, Ed., 1998, pp. 269–277.

KEARNS, M.; e MANSOUR, Y. On the boosting ability of top-down decision tree learning algorithms. *Journal of Computer and Systems Sciences*, 58(1): 109-128, 1999.

KOFORD, J. Adaptive network organization, Rept SEL-63-009, Stanford Electronics Laboratories Quarterly Research Review, no 3, 1962, 111-6.

KOHAVI R., and PROVOST F., Glossary of Terms, *Machine Learning* 30(2/3): 271-274, 1998.

KOVACEVIC, S. and DOMINELLI, N.; Transformer Life Extension by Continuous On-Line Oil Conditioning; Proceedings of Electrical Insulation Conference and Electrical Manufacturing & Coil Winding Technology Conference. ,pp.23-25, 2003

LEBENSZTAYN, E. Exercícios de Probabilidade. Unicamp. Livro publicado pelo próprio autor, 2012.

LI, X.; e DUBES, R. C. Tree classifier design with a Permutation statistic, *Pattern Recognition* 19:229-235, 1986.

LI, X.; e WU, H. DGA interpretation scheme derived from case study. *IEEE Trans. Power Delivery*, vol. 26, no. 2, pp. 1292 – 1293, April 2011.

LIMA, S. E. U. de, Diagnóstico Inteligente de Falhas Incipientes em Transformadores de Potência Utilizando a Análise dos Gases Dissolvidos em Óleo. Dissertação de Mestrado, PPGEE/DEE/UFC. Fortaleza, Ceará, 2005.

LOH, T.; e SHIH, T. Split selection methods for classification trees. *Statistica Sinica*, vol. 7, pp. 815–840, 1997.

LORENA, Ana C. e CARVALHO, André C.P.L.F. Relatório Técnico: Introdução às Máquinas de Vetores Suporte. ICMC/USP/SC, 2003. Disponível em: <  
[http://www.icmc.usp.br/CMS/Arquivos/arquivos\\_enviados/BIBLIOTECA\\_113\\_RT\\_192.pdf](http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_192.pdf)  
>. Acesso em: agosto de 2013.

LUBINSKY, D. Algorithmic speedups in growing classification trees by using an additive split criterion. in *Proc. AI Statistics*, 1993, pp. 435–444.

MANTRAS, R. L. de. A distance-based attribute selection measure for decision tree induction. *Mach. Learn.*, vol. 6, pp. 81–92, 1991.

MARKOWSKA-KACZMAR, U.; e CHUMIEJA, M. Discovering the mysteries of neural networks. *International Journal of Hybrid Intelligent Systems*, 1(3-4):153-163, 2004.

MARKOWSKA-KACZMAR, U.; e TRELAK, W. Extraction of fuzzy rules from trained neural network using evolutionary algorithm. In *European Symposium on Artificial Neural Networks (ESANN)*, pp. 149-154, 2003.

MARTENS, D.; BAESENS, B.; e Van GESTEL, T. Decompositional Rule Extraction from Support Vector Machines by Active Learning. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, NO. 2, February 2009.

MARTENS, D.; De BACKER, M.; HAESSEN, R.; SNOECK, M.; VANTHIENEN, J.; e BAESENS, B. Classification with Ant Colony Optimization. *IEEE Trans. Evolutionary Computation*, vol. 11, no. 5, pp. 651-665, 2007.

MARTIN, J. K. An exact probability metric for decision tree splitting and stopping. *Mach. Learn.*, vol. 28, no. 2–3, pp. 257–291, 1997.

MEHTA, M.; RISSANEN, J.; AGRAWAL, R. MDL-Based Decision Tree Pruning. *KDD 1995*: pp. 216-221, 1995.

MILASCH M. *Manutenção de Transformadores em Líquido Isolante*. Edgard Blucher, São Paulo, Brasil, 1984.

MINGERS, J. An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4(2):227-243, 1989.

MIRANDA, V.; CASTRO, A. R. G. Improving the IEC Table for Transformer Failure Diagnosis with Knowledge Extraction from Neural Networks. *IEEE Trans. Power Delivery*, vol. 20, no. 4, pp. 2509–2516, Oct. 2005.

MITCHELL, T. M. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.

MORAIS, D. R.; ROLIM, J. G.; SILVA, J. R. A Fuzzy System for Detection of Incipient Faults in Transformers Based on the Dissolved Gas Analysis of Insulating Oil. In: *5th IEEE -*

International Symposium on Diagnostics, Electric Machines, Power Electronics and Drives, 2005. Anais. Viena, p. 1-6, 2005.

MURTHY, S. K.; KASIF, S.; e SALZBERG, S. A system for induction of oblique decision trees. *J. Artif. Intell. Res.*, vol. 2, pp. 1–33, Aug. 1994.

MYERS, S. D.; KELLY, J.; PARRISH, R. H. *A Guide to Transformer Maintenance*, Ohio, USA, ch. 4, p. 323-388, 1981.

NARESH, R.; SHARMA, V.; e VASHISTH, M. An integrated neural fuzzy approach for fault diagnosis of transformers. *IEEE Trans. Power Del.*, vol. 23, no. 4, pp. 2017–2024, Oct. 2008.

NAUMOV, G. E. NP-completeness of problems of construction of optimal decision trees. *Sov. Phys.: Doklady*, vol. 36, no. 4, pp. 270–271, 1991.

NETO, L. B.; CHIGANER, L. Sistema Especialista Fuzzy no Diagnóstico de Falhas em Transformadores, In: XV SNPTEE, GSE 14. Anais. Paraná, Brasil. Outubro, 1999.

NIBLETT, T.; e BRATKO, I. Learning decision rules in noisy domains. in *Expert Systems*. Cambridge, MA: Cambridge Univ. Press, 1986.

NOVIKOFF, A. On convergence proofs for perceptrons, *Symposium on Mathematical Theory of Automata*. Brooklyn, N. Y.: Polytechnic Press, 1963, pp. 615-622.

NÚÑEZ, H.; ANGULO, C.; e CATALÀ, A. Rule extraction from support vector machines. In *European Symposium on Artificial Neural Networks (ESANN)*, pp. 107-112, 2002.

OEMMEN, T. V. Vegetable Oils for Liquid-Filled Transformers, *IEEE Electrical Insulation Magazine*, pp. 6 –11, 2002.

PALHARES, Reinaldo M. *Fund. Controle Robusto via Otimização – Bloco 2*. Disponível em: <[http://www.cpdee.ufmg.br/~palhares/bloco2\\_ftcr.pdf](http://www.cpdee.ufmg.br/~palhares/bloco2_ftcr.pdf)>. Acesso em: 30 de junho de 2011.

PAPPA, G. L. Seleção de atributos utilizando algoritmos genéticos multiobjetivos. Dissertação de Mestrado, Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica do Paraná, 2002.

PARASKEVAS, C. D.; VASSILIOU, P.; DERVOS, C. T. Temperature Dependent Dielectric Spectroscopy in Frequency Domain of High-voltage Transformer Oils Compared to Physicochemical Results, *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 13, no. 3, pp. 539-546, 2006.

PATEL, N. K.; KHUBCHANDANI, R. K. ANN Based Power Transformer Fault Diagnosis. *IE Journal - EL*, [S.l.], v. 85, p. 60-63, June, 2004.

PEYRAQUE, L.; BEROUAL, A.; BURET, F. Static Electrification of Pressboard/Oil Interface and Transient Phenomena. *IEEE Transactions on Dielectrics and Electrical Insulation*, v. 5, n. 3, pp. 443-449, 1998.

POSTON, W. L., MARCHETTE, D. J. Recursive dimensionality reduction using Fisher's linear discriminant. *Pattern Recognition*, v. 31, Issue: 7, pp. 881-888, Julho 1998.

QUINLAN, J. Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, 28(1):71-72, 1996.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

QUINLAN, J. R. Induction of decision trees. *Mach. Learn.*, vol. 1, pp. 81-106, 1986.

QUINLAN, J. R., Simplifying decision trees, *International Journal of Man-Machine Studies*, 27, 221-234, 1987.

QUINLAN, J. R.; e RIVEST, R. L. Inferring decision trees using the minimum description length principle. *Inform. Comput.*, vol. 80, pp. 227-248, 1989.

RABUÑAL, J.R.; DORADO, J.; PAZOS, A.; PEREIRA, J.; e RIVERO, D. A new approach to the extraction of ANN rules and to their generalization capacity through GP. *Neural Computation*, 16(47):1483-1523, 2004.

RAKOTOMALALA, R. SIPINA. Programa computacional. Disponível em: <<http://eric.univ-lyon2.fr/~ricco/sipina.html>>. Acesso em: junho de 2012.

RAMACHANDRA, T.V; SUBASH CHANDRAN, M.D; JOSHI, N.V; PALLAV JULKA; UTTAM KUMAR; BHARATH H. AITHAL; PRAKASH MESTA; RAO GR; e VISHNU

MUKRI. Landslide Susceptible Zone Mapping in Uttara Kannada, Central Western Ghats., ENVIS Technical Report : 28, Energy & Wetlands Research Group, Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560 012, 2012.

RISSANEN, J. Stochastic complexity and statistical inquiry. World Scientific, 1989.

ROCHA, F. B. Diagnóstico de Falhas Incipientes a partir das Propriedades Físico-Químicas do Óleo Isolante em Transformadores de Potência como Método Alternativo à Análise de Gases Dissolvidos. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Ceará, 2013.

ROKACH, L., and MAIMON, O. Data mining with decision trees: Theory and Applications. Series in Machine Perception and Artificial Intelligence - Vol. 69. World Scientific Publishing, 2008.

ROKACH, L., and MAIMON, O. Top-down induction of decision trees classifiers - a survey. IEEE Trans. on Systems, Man, and Cybernetics, Vol. 35, no 4, pp. 476 – 487, Nov. 2005.

ROUNDS, E. A combined nonparametric approach to feature selection and binary decision tree design. Pattern Recognition, vol. 12, pp. 313–317, 1980.

RUSSELL, S.; NORVIG, P. Inteligencia Artificial. Editora Campus, 2004.

SAITO, K.; e NAKANO, R. Law discovery using neural networks. In Proceedings of the 15th International Joint Conference on Artificial Intelligence, pp. 1078-1083, 1997.

SAITO, K.; e NAKANO, R. Extracting regression rules from neural networks. Neural Networks, 15(10):1279-1288, 2002.

SCHMITZ, G.P.J.; ALDRICH, C.; e GOUWS, F.S. ANN-DT: An algorithm for extraction of decision trees from artificial neural networks. IEEE Transactions on Neural Networks, 10(6):1392-1401, 1999.

SETHI, I. K.; e YOO, J. H. Design of multicategory, multifeature split decision trees using perceptron learning. Pattern Recognition, vol. 27, no. 7, pp. 939–947, 1994.

SETIONO, R.; BAESENS, B.; e MUES, C. Risk Management and Regulatory Compliance: A Data Mining Framework Based on Neural Network Rule Extraction. Proc. Int. Conf. Information Systems (ICIS), 2006.

SETIONO, R.; e LEOW, W.K. FERNN: An algorithm for fast extraction of rules from neural networks. Applied Intelligence, 12(1-2):15-25, 2000.

SETIONO, R.; e LIU, H. Neurolinear: From neural networks to oblique decision rules. Neurocomputing, 17(1):1-24, 1997.

SETIONO, R.; e THONG, J.Y.L. An approach to generate rules from neural networks for regression problems. European Journal of Operational Research, 155(1):239-250, 2004.

SETIONO, R.; LEOW, W.K.; e ZURADA, J.M. Extraction of rules from artificial neural networks for nonlinear regression. IEEE Transactions on Neural Networks, 13(3):564-577, 2002.

SILVA, I. N.; SOUZA, A. N.; SOUZA, M. G. Z. N. Avaliação do Nível de Contaminação do Óleo de Transformadores Através de Redes Neurais Artificiais. In: XV SNPTEE, GSE 13. Anais. Paraná, Brasil.Outubro, 1999.

SINGH, S.; BANDYOPADHYAY, M.N. Dissolved gas analysis technique for incipient fault diagnosis in power transformers: A bibliographic survey," Electrical Insulation Magazine, IEEE , vol.26, no.6, pp.41-46, November-December 2010.

SOUSA, R.; AMORA, M.; BARBOSA, F. R.; PONTES, R. T.; BRAGA, A. P. S.; ALMEIDA, O. M. "Análise de Métodos para o Diagnóstico de Falhas Incipientes em Transformadores com Base na Concentração de Gases no Óleo Isolante". IV Simpósio Brasileiro de Sistemas Elétricos - SBSE 2012, Goiás, 2012.

SOUZA, J. A. de. Reconhecimento de padrões usando indexação recursiva". Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 1999.

SPURGEON, K.; TANG, W.H.; WU, Q.H.; RICHARDSON, Z. J.; MOSS, G. Dissolved gas analysis using evidential reasoning. IEE Proceedings Science, Measurement & Technology. Vol. 152, no. 3, pp.110–117, May 2005.

TAHA, I.; e GHOSH, J. Symbolic interpretation of artificial neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 11(3):448-463, 1999.

TAYLOR, P. C.; e SILVERMAN, B.W. Block diagrams and splitting criteria for classification trees. *Statistics and Computing*, vol. 3, no. 4, pp. 147–161, Dec. 1993.

TENBOHLEN, S., FIGEL, F., On-line condition monitoring of power transformers, *IEEE Power Engineering Society Winter Meeting*, pp. 2211 - 2216 vol.3, 2000.

THE ELECTRICITY FORUM By Department of the Interior; Bureau of Reclamation; Hydroelectric Research; Technical Services Group. *Transformer Maintenance, Electrical Transformer Handbook, Electricity Forum*, v. 2, pp. 25-64, Canada, 2005.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 3. ed. San Diego: Academic Press, 2009.

THRUN, S. Extracting provably correct rules from artificial neural networks. Technical report iai-tr-93-5, Universität Bonn, Institut für Informatik III, 1993. Programa computacional. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.2110&rep=rep1&type=pdf>>. Acesso em: agosto de 2013.

TOWELL, G.; e SHAVLIK, J. W. The extraction of refined rules from knowledge-based neural networks. *Machine Learning*, 13(1):71-101, 1993.

UTGOFF, P. E. Incremental induction of decision trees. *Machine Learning*, 4:161-186, 1989.

UTGOFF, P. E.; e CLOUSE, J. A. A Kolmogorov-Smirnoff Metric for Decision Tree Induction, Technical Report 96-3, University of Massachusetts, Department of Computer Science, Amherst, MA, 1996.

WALLACE, C.; e PATRICK, J. Coding decision trees. *Mach. Learn.*, vol. 11, pp. 7–22, 1993.

WANG, M. H, Extension neural network for power transformer incipient fault diagnosis. *IEE Proc., Gener. Transm. Distrib.*, Vol. 150, no. 6, pp. 679–685, Nov. 2003.

WANG, Z. *Artificial Intelligence Applications in the Diagnosis of Power Transformer Incipient Faults*. Thesis (PhD.), Virginia Polytechnic Institute and State University, Blacksburg, 2000.

WANG, Z.; LIU, Y.; GRIFFIN, P. J. Neural net and expert system diagnose transformer faults. *IEEE Computer Applications in Power*. Volume: 13. Issue: 1. Publication Year: 2000. Page(s): 50 – 55.

WANG, Z.; LIU, Y.; GRIFFIN, P. J. A Combined ANN and Expert System Tool for Transformer Fault Diagnosis. *IEEE Transactions on Power Delivery*, v.13, n.4, p.1224-1229, Oct. 1998.

WARD, S. A. Evaluating Transformer Condition Using DGA Oil Analysis. *Electrical Insulation and Dielectric Phenomena, Annual Report*, p. 463 – 468, 2003.

WEG. Manual de Instalação e de Manutenção de Transformadores. Catálogo. 16 p. Blumenau, 2004.

WEISSTEIN, Eric W. MathWorld. Wolfram Research. Disponível em: <<http://mathworld.wolfram.com/Plane.html>>. Acesso em: 18 de abril de 2013.

WENDEL, J. G. A problem in geometric probability, *Mathematica Scandinavica*, vol 11, 1962, pp 109-111.

XU, W., WANG, D., ZHOU, E. Z.; CHEN, H. Fault Diagnosis of Power Transformers: Application of Fuzzy Set Theory, Expert Systems and Artificial Neural Networks. *IEE Proc. – Sci. Meas. Technol.*, v. 144, n. 1, Jan. 1997.

YILDIZ, O. T.; ALPAYDIN, E. Linear discriminant trees. “*International Journal of Pattern Recognition and Artificial Intelligence*”. Vol 19, No. 3, 2005.

ZHANG, Y.; DING X.; LIU, Y.; GRIFFIN, P. J. An Artificial Neural Approach to Transformer Fault Diagnosis. *IEEE Trans. on Power Delivery*, v. 11, n. 4, p.1836-1841, Oct. 1996.

ZHOU, Z. H.; JIANG, Y.; e CHEN, S. F. Extracting symbolic rules from trained neural network ensembles. *AI Communications*, 16(1):3-15, 2003.

ZIRBES, R.; ROLIM, J. G.; e ZÜRN, H. H. Metodologias para avaliação e diagnóstico do estado de isolamentos de papel impregnado com óleo mineral. *Revista Controle & Automação*, Vol. 16, Nº 3, pp. 318-330, Jul./Ago./Set. 2005.

## APÊNDICES

### APÊNDICE A – Métodos de Extração de Regras

Na Tabela A é apresentada um panorama cronológico de vários métodos de extração de regras, indicando também o uso (classificação e/ou regressão), o tipo de dependência com um método de solução “caixa-preta” adotado, e um breve comentário sobre as características do método.

**Tabela A - Métodos de extração de regras**

Algoritmo	Referências	Tipo	Uso	Descrição
CART (1984)	BREIMAN et al. (1984)	I	C e R	Indução de AD.
CN2 (1989)	CLARK e NIBLETT (1989)	I	C	Indução de regras.
KT (1991)	FU (1991)	D	C	Encontra subconjuntos de pesos que asseguram a ativação da unidade de saída.
C4.5 (1993)	QUINLAN (1993)	I	C	Indução de AD.
SUBSET (1993)	TOWELL e SHAVLIK (1993)	D	C	Encontra subconjuntos de pesos que asseguram a ativação da unidade de saída.
N-of-M (1993)	TOWELL e SHAVLIK (1993)	D	C	Cria regras do tipo M-de-N, baseado no agrupamento de pesos.
VIA (1993)	THRUN (1993)	D	C	Cria regras compatíveis com a RNA através da Análise de Intervalo.
REAL (1994)	CRAVEN e SHAVLIK (1994)	D	C	Cria uma regra específica que abrange um exemplo, iterativamente remove condições e verifica as regras.
TREPAN (1996)	CRAVEN e SHAVLIK (1996)	I	C	Indução de AD. Abertura M de N dos nós.
Neuro Linear (1997)	SETIONO e LIU (1997)	D	C	Cria regras oblíquas, para RNAs com uma única camada oculta.
RFS (1997)	SAITO e NAKANO (1997)	D	R	Predecessor do RN2, aceita apenas entradas numéricas.
BIO-RE (1999)	TAHA e GHOSH (1999)	I	CB	Cria uma tabela verdade completa, apenas aplicável para alguns tipos de problemas.
PARTIAL-RE (1999)	TAHA e GHOSH (1999)	D	C	Semelhante ao SUBSET, mas usa ordenação por pesos.
FULL-RE (1999)	TAHA e GHOSH (1999)	D	C	Semelhante ao SUBSET, mas usa programação linear.
ANN-DT (1999)	SCHMITZ, ALDRICH e GOUWS (1999)	I	C e R	Indução de AD, semelhante ao TREPAN.
FERNN (2000)	SETIONO e LEOW (2000)	D	C	Cria regras oblíquas, utiliza o C4.5 para os valores de ativação das unidades ocultas e os seus valores alvo.
DecText (2000)	BOZ (2000)	I	C	Indução de AD.
RN2 (2002)	SAITO e NAKANO (2002)	D	R	Utiliza análise polinomial. Assume as unidades escondidas como produtos e os valores de ativação como clusters.
REFANN (2002)	SETIONO, LEOW e ZURADA (2002)	D	R	Aproxima as funções de ativação da RNA por funções seccionalmente lineares.
SVM+Prototypes (2002)	NÚÑEZ, ANGULO e CATALÀ (2002)	D	C	Utiliza clustering.
STARRE (2003)	ZHOU, JIANG e CHEN (2003)	I	C	Utiliza pesquisa em profundidade, prefere variáveis categóricas ao invés de variáveis contínuas.
G-REX (2003)	JOHANSSON, KÄONIG e NIKLASSON (2003)	I	C e R	Utiliza programação genética. Cria diferentes tipos de regras.
REX (2003)	MARKOWSKA-KACZMAR e TRELAK (2003)	I	C	Utiliza algoritmo genético. Cria regras fuzzy.
GEX (2004)	MARKOWSKA-KACZMAR e CHUMIEJA (2004)	I	C	Utiliza algoritmo genético. Cria regras proposicionais.
NNRULES (2004)	SETIONO e THONG (2004)	D	R	Aproxima as funções de ativação da RNA por funções seccionalmente lineares.
Rabunál (2004)	RABUÑAL et al. (2004)	I	C	Utiliza programação genética.
BUR (2004)	CHEN (2004)	I	C	Baseado em Gradient Boosting Machines.
Barakat (2005)	BARAKAT e DIEDERICH (2005)	D	C	Utiliza uma RNA com os vetores de suporte e seus identificadores.
Fung (2005)	FUNG, SANDILYA e RAO (2005)	D	CB	Aplicável apenas para classificadores lineares.
Miranda e Castro (2007)	MIRANDA e CASTRO (2005)	D	C	Utiliza os vetores de suporte na formação de regras <i>fuzzy</i>
ITER (2006)	HUYSMANS, BAESENS e VANTHIENEN (2006)	I	C e R	Crescimento iterativo de hipercubos.
Re-RX (2006)	SETIONO, BAESENS e MUES (2006)	I	C	Conj. de regras hierárquicas: primeiras divisões são baseadas em atributos discretos.
SQRex-SVM (2007)	BARAKAT e BRADLEY (2007)	D	C	Cobertura sequencial.
AntMiner+ (2007)	MARTENS et al. (2007)	I	C	Indução de regras baseado em colônia de formigas.
ALBA (2008)	MARTENS, BAESENS e Van GESTEL (2008)	D	C	Aprendizado ativo com os vetores de suporte.

Fonte: adaptado de (MARTENS, BAESENS e Van GESTEL, 2008; e HUYSMANS, BAESENS e VANTHIENEN, 2006).

Legenda: D - Dependente, I - Independente, C - Classificação, CB - Classificação Binária e R - Regressão.

O “princípio de incompatibilidade” de Zadeh (CASTRO, MANTAS, e BENÍTEZ, 2002) estabelece que *a complexidade de um sistema e a precisão com que ele pode ser*

*analisado apresentam uma relação aproximadamente inversa.* Esse princípio pode ser aplicado para vários métodos de extração de conhecimento a partir de técnicas de ML.

De um modo geral, os métodos de extração de regras que utilizam Algoritmo Genético (AG) apresentam como a principal vantagem a flexibilidade a alterações. Alterações na função de *fitness* permite ao pesquisador controlar o paradigma precisão-compreensibilidade, comum aos métodos de extração de regras, e também modificações dessas funções permite alterar o formato das descrições (regras) extraídas. A principal desvantagem de todos os algoritmos genéticos são os requisitos computacionais de tempo para a realização das sucessivas iterações. A segunda desvantagem diz respeito à consistência das descrições extraídas, devido ao aspecto de probabilidade durante a criação das populações, as regras extraídas podem ter diferenças significativas quando o método é executado várias vezes sobre o mesmo conjunto de dados.

Em relação aos métodos de extração de regras a partir de uma Rede Neural Artificial (RNA) ou de uma Máquina de Vetor de Suporte (MVS), normalmente estes métodos irão apresentar limitações quanto à precisão e/ou grau de compreensão das regras extraídas. Em relação à compreensibilidade das regras, esses problemas ocorrem principalmente na extração de regras do tipo *fuzzy*, muitas vezes utilizando operadores especiais, como em (BENÍTEZ, CASTRO, e REQUENA, 1997), (CASTRO, MANTAS, BENÍTEZ, 2002), e (CASTRO *et al.*, 2007) que necessitam do cálculo matemático do grau ativação das regras, prejudicando o uso direto por inspeção de tais regras. Em (MIRANDA e CASTRO, 2007) é proposto um método de extração de regras com grande precisão, entretanto exige o treino de uma RNA com funções de ativação especiais para os neurônios da camada escondida, dificultando o treinamento e a convergência da RNA, também o número de regras obtidas é elevado, dificultando o uso e a interpretação de tais regras.

As regras de conhecimento obtidas através da construção de uma árvore de decisão (*Decision Tree - DT*) podem apresentar problemas de generalização, caso cuidados não sejam tomados durante o treino e teste da árvore. A utilização de técnicas de poda favorece a capacidade de generalização da árvore, e, portanto, das regras obtidas. Também, as regras extraídas são baseadas em fronteiras construídas de forma paralela aos eixos do problema, que dependendo da característica específica de um problema modelado pode resultar numa solução ineficiente. Maiores detalhes sobre estes problemas serão discutidos no Capítulo 3.

## APÊNDICE B – Abertura de Nós, Poda e Indução de uma Árvore de Decisão

### B.1. Critérios de Abertura de Nós

Na maioria dos casos, as funções utilizadas para a divisão de nós durante a construção da árvore são univariantes, ou seja, é avaliado somente um único atributo do vetor de entrada. Portanto, o algoritmo de construção precisa determinar qual o melhor atributo para realizar a construção do nó e divisão dos ramos. Existem vários critérios para realizar a construção, os principais critérios utilizados serão comentados a seguir, com base nas referências (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008). Estes critérios podem ser utilizados associados ao algoritmo básico descrito na Figura 8 (Capítulo 3), apenas adotando as equações e indicações correspondentes na sub-rotina chamada *ArvoreCrescimento*.

#### B.1.1. Critérios Baseados em Impureza

Dado uma variável  $x$  com  $k$  valores discretos, distribuídos de acordo com  $P=(p_1, p_2, \dots, p_k)$ , a medida de impureza será a função  $\phi:[0,1]^k \rightarrow \mathfrak{R}$  que satisfaz as seguintes condições (ROKACH e MAIMON, 2008):

- $\phi(P) \geq 0$ ;
- $\phi(P)$  é mínimo se  $\exists i$  para que o componente  $P_i=1$ ;
- $\phi(P)$  é máximo se  $\forall i, 1 \leq i \leq k, P_i=1/k$ ;
- $\phi(P)$  é simétrico com respeito aos componentes de  $P$ ; e
- $\phi(P)$  é totalmente diferenciável em sua faixa de valores.

Deve ser observado que se o vetor de probabilidade tem uma componente 1, ou seja a variável  $x$  apresenta somente um valor, então a variável é considerada como pura. O oposto seria se todas as componentes são iguais, com o nível de impureza sendo máximo.

Considerando um conjunto  $S$  de treinamento, o vetor de probabilidade para o valor esperado do atributo  $y$  é definido por (ROKACH e MAIMON, 2008):

$$P_y(S) = \left( \frac{|\sigma_{y=c_1} S|}{|S|}, \dots, \frac{|\sigma_{y=c_{|dom(y)|}} S|}{|S|} \right) \quad (\text{B.1})$$

O ganho devido a realização da divisão do nó em função do atributo  $a_i$  é definido como a redução de impureza após a partição de  $S$  de acordo com os valores  $v_{i,j} \in \text{dom}(a_i)$  (ROKACH e MAIMON, 2008):

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|\text{dom}(a_i)|} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}} S)) \quad (\text{B.2})$$

Vários critérios baseados em impureza são citados na Tabela B.1, junto com as principais equações e a referência.

**Tabela B.1. Critérios Baseados em Impureza**

Método	Equações	Referências
Ganho de Informação	$G(a_i, S) = E(y, S) - \sum_{v_{i,j} \in \text{dom}(a_i)} \frac{ \sigma_{a_i=v_{i,j}} S }{ S } \cdot E(y, \sigma_{a_i=v_{i,j}} S)$ $E(y, S) = \sum_{c_j \in \text{dom}(y)} - \frac{ \sigma_{a_i=c_j} S }{ S } \cdot \log_2 \frac{ \sigma_{a_i=c_j} S }{ S }$	(QUINLAN, 1987)
Índice Gini	$Gini(y, S) = 1 - \sum_{c_j \in \text{dom}(y)} \left( \frac{ \sigma_{a_i=c_j} S }{ S } \right)^2$ $G_{Gini}(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in \text{dom}(a_i)} \frac{ \sigma_{a_i=v_{i,j}} S }{ S } \cdot Gini(y, \sigma_{a_i=v_{i,j}} S)$	(BREIMAN <i>et al.</i> , 1984; GELFAND <i>et al.</i> , 1991)
Razão de Semelhança	$G^2(a_i, S) = 2 \cdot \ln(2) \cdot  S  \cdot G(a_i, S)$	(ROKACH e MAIMON, 2008)
Critério DKM	$DKM(y, S) = 2 \cdot \sqrt{\left( \frac{ \sigma_{y=c_1} S }{ S } \right) \cdot \left( \frac{ \sigma_{y=c_2} S }{ S } \right)}$	(DIETTERICH, KEARNS e MANSOUR, 1996; KEARNS e MANSOUR, 1999)

Fonte: Elaborada pelo autor.

### **B.1.2. Critérios Baseados na Impureza Normalizada**

A utilização do critério de impureza com normalização nos dados evitar que os resultados sejam afetados por valores elevados no domínio, que resultaria em preferência na utilização de certos atributos do que em outros no processo de divisão, inclusive podendo contribuir para uma capacidade pobre de generalização da árvore formada. Por estas razões, é indicado o uso de critérios baseados em medidas de impurezas com normalização.

A seguir vários destes critérios normalizados são apresentados na Tabela B.2.

### **B.1.3. Outros Métodos de Divisão Univariantes**

Outros critérios de divisão de nós na formação de uma DT são encontrados na literatura (ROKACH e MAIMON, 2005), como: a permutação estatística (LI e DUBES, 1986), a melhoria pela média posterior (TAYLOR e SILVERMAN, 1993), e a medida de distribuição hipergeométrica (MARTIN, 1997).

### **B.1.4. Comparação dos Métodos Univariantes**

Segundo (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008) vários estudos de comparação foram realizados tentando definir quais critérios seriam melhores e em que situações. E apesar de alguns resultados localizados, o que se concluiu é que nenhum dos métodos citados se destaca em relação aos demais.

### **B.1.5. Critérios Multivariáveis de Construção de Nós**

Nas subseções anteriores foram apresentados vários critérios de construção de nós com análise univariável, representando os principais critérios utilizados na construção de uma DT. Entretanto, também podem ser utilizados critérios multivariáveis.

Nos critérios multivariáveis diversos atributos podem ser avaliados durante a abertura dos nós em uma DT. Evidentemente, o problema de se determinar a melhor opção de divisão de nós para o caso multivariável é mais complicado que no caso univariável. Além disso, o uso desses critérios pode afetar drasticamente o rendimento da DT. Portanto, estes critérios acabam não sendo muito adotados.

A maioria destes critérios multivariáveis é baseada na combinação linear dos atributos de entrada (ROKACH e MAIMON, 2008). A determinação da melhor combinação linear pode ser realizada utilizando uma pesquisa gulosa (BREIMAN *et al.*, 1984; MURTHY, KASIF, e SALZBERG, 1978), programação linear (DUDA e HART, 1973; BENNETT e

MANGASARIAN, 1994), análise de discriminante linear (FRIEDMAN, 1977; JOHN, 1996) e outros (UTGOFF, 1989; LUBINSKY, 1993; SETHI e YOO, 1994).

**Tabela B.2. Critérios Baseados em Impureza Normaliza**

<b>Método</b>	<b>Equações</b>	<b>Referências</b>
Razão de Ganho	$R_G(a_i, S) = \frac{G(a_i, S)}{E(a_i, S)}$	(QUINLAN, 1993)
Medida de Distância	$MD(a_i, S) = \frac{\Delta\Phi(a_i, S)}{-\sum_{v_i, j \in \text{dom}(a_i)} \sum_{c_k \in \text{dom}(y)} b \cdot \log_2 b}$ onde $b = \frac{ \sigma_{a_i=v_i, j} \text{ e } S_{y=c_k} }{ S }$	(MANTRAS, 1991)
Critério Binário	$\beta^*(a_i, d_1, d_2, S) = \max \beta(a_i, d_1, d_2, S)$ com $d_1$ e $d_2$ sendo os correspondentes subdomínios da divisão do nó pelo atributo $a_i$ . Restrições: $d_1 \cup d_2 = \text{dom}(a_i)$ e $d_1 \cap d_2 = \emptyset$ .	(ROKACH e MAIMON, 2008)
Critério Twoing	$Twoing(a_i, d_1, d_2, S) = 0,25 \cdot \frac{ \sigma_{a_i \in d_1} S }{ S } \cdot \frac{ \sigma_{a_i \in d_2} S }{ S } \cdot \left( \sum_{c_k \in \text{dom}(y)} \left  \frac{ \sigma_{a_i \in d_1} \text{ e } S_{y=c_k} }{ \sigma_{a_i \in d_1} S } - \frac{ \sigma_{a_i \in d_2} \text{ e } S_{y=c_k} }{ \sigma_{a_i \in d_2} S } \right  \right)^2$	(BREIMAN <i>et al.</i> , 1984)
Critério de Ortogonalidade	$ORT(a_i, d_1, d_2, S) = 1 - \cos\theta(P_{y,1}, P_{y,2})$ sendo $\theta(P_{y,1}, P_{y,2})$ o ângulo entre os vetores de distribuição $P_{y,1}$ e $P_{y,2}$ para o valor de saída do atributo $y$ nos dois subconjuntos $\sigma_{a_i \in d_1} S$ e $\sigma_{a_i \in d_2} S$ , respectivamente	(FAYYAD e IRANI, 1992)
Critério de Kolmogorov-Smirnov	$KS(a_i, d_1, d_2, S) = \left  \frac{ \sigma_{a_i \in d_1} \text{ e } S_{y=c_1} }{ \sigma_{y=c_1} S } - \frac{ \sigma_{a_i \in d_2} \text{ e } S_{y=c_2} }{ \sigma_{y=c_2} S } \right $ considerando um domínio binário para a saída desejada, $\text{dom}(y) = \{c_1, c_2\}$ ,	(FRIEDMAN, 1977; e ROUNDS, 1980)

Fonte: Elaborada pelo autor.

## B.2. Métodos de Poda

A utilização de critérios de parada mais rígidos na formação de uma DT leva a resultados de árvores menores e não sobreajustadas em relação aos dados de treino, entretanto isto pode ter impactos na precisão dos resultados. Enquanto, a utilização de critérios de parada menos rígidos, resultam em árvores maiores e com possibilidade de estarem sobreajustadas aos dados de treino, limitando a capacidade de generalização da árvore.

Uma alternativa é relaxar os critérios de parada na construção da DT, e aplicar um método de poda após a construção da árvore, de forma a reduzir o tamanho dessa árvore, eliminando nós e ramos com menor impacto na precisão de classificação realizada pela árvore. A ideia de poda foi originalmente proposta por (BREIMAN *et al.*, 1984). Estudos demonstram que a poda permite a melhoria na capacidade de generalização da árvore, mesmo em situações de ruído nos dados (ROKACH e MAIMON, 2005).

Os métodos de poda permitem a obtenção de uma DT com uma mescla de precisão em relação ao treino e uma forma compacta, que por sua vez ajuda na generalização e interpretação da árvore.

Existem vários métodos de poda propostos na literatura, considerando um determinado critério para aperfeiçoar a poda da DT. A seguir os métodos mais populares serão comentados, com base nas referências (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008). Estes critérios de poda podem ser utilizados associados ao algoritmo básico descrito na Figura 8 (Capítulo 3), apenas adotando as equações e indicações correspondentes na sub-rotina *Poda* chamada no procedimento *ArvorePoda*. O procedimento *ArvorePoda* é executado após a formação da DT pela rotina *ArvoreIndutor*.

Na Tabela B.3 são apresentados e comentados os principais métodos de poda utilizados pelos algoritmos de indução de uma DT.

### **B.2.1. Outros Métodos**

Existem outros métodos indicados na literatura (ROKACH e MAIMON, 2008), como em: (WALLACE e PATRICK, 1993) que propõem um método baseado no cumprimento mínimo para a informação, e (KEARNS e MANSOUR, 1998) que propõem um algoritmo com justificativa teórica, já em (MINGERS, 1989) é calculado um valor crítico de poda.

### **B.2.2. Comparação dos Métodos de Poda**

Segundo (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008) vários estudos de comparação foram realizados tentando definir quais critérios seriam melhores. E também, como no caso dos métodos de criação de nós, o que se concluiu é que nenhum dos métodos se destaca dos demais.

Tabela B.3. Critérios de Poda

Método	Procedimento	Referências
Poda Através do Custo de Complexidade	Esse método realiza a poda em dois estágios. Primeiro, uma sequência de árvores $T_0, T_1, \dots, T_k$ é construída utilizando os dados de treino, onde $T_0$ é a árvore original antes da poda e $T_k$ é a árvore com poda máxima, apenas com o nó raiz (primeiro nó da árvore). No segundo estágio, uma das árvores é escolhida como a árvore treinada, baseado na estimação do erro de generalização.	(BREIMAN <i>et al.</i> , 1984)
Poda Através da Redução do Erro	Representa um método de redução simples: é feita uma avaliação dos nós do final para o início da árvore, substituindo nós e subárvores pela classe mais comum entre os dados separados pelos mesmos. O procedimento continua enquanto a poda não resultar em erros significativos.	(QUINLAN, 1987)
Poda Através do Erro Mínimo	Esse método realiza uma busca de baixo para cima na árvore e cada nó é comparado com uma taxa de erro de probabilidade com e sem a poda: $\varepsilon'(t) = 1 - \max_{c_i \in \text{dom}(y)} \frac{ \sigma_{y=c_i} S  + l \cdot p_{pri}(y = c_i)}{ S_t  + l}$ onde $p_{pri}(y = c_i)$ representa a probabilidade priori de $y$ obter o valor $c_i$ , e $l$ é o peso dado para essa probabilidade. Um nó só é podado se esta poda não aumentar a taxa de erro de probabilidade $m$ . O conjunto $S_t$ representa as instâncias associadas ao nó $t$ .	(NIBLETT e BRATKO, 1986)
Poda Pessimista	Representa um método rápido de poda que avalia a eliminação de nós e subárvores relacionadas de baixo para cima. Este método evita a necessidade de utilizar um conjunto de dados para a poda ou o uso de validação cruzada, usando um teste de correlação estatística pessimista. Se o erro calculado utilizando o conjunto de treino não é adequado, uma medida será realizada com correlação contínua através de uma distribuição binomial: $\varepsilon'(T, S) = \varepsilon(T, S) + \frac{ \text{folhas}(T) }{2 \cdot  S }$ dentro da faixa: $\varepsilon'(\text{podada}(T, t), S) \leq \varepsilon'(T, S) + \sqrt{\frac{\varepsilon'(T, S) \cdot (1 - \varepsilon'(T, S))}{ S }}$	(QUINLAN, 1993)
Poda Baseada no Erro	Seria uma melhoria no método da poda pessimista, onde o erro é calculado utilizando agora: $\varepsilon_{sup}(T, S) = \varepsilon(T, S) + Z_\alpha \sqrt{\frac{\varepsilon(T, S) \cdot (1 - \varepsilon(T, S))}{ S }}$ com $\varepsilon(T, S)$ significando a taxa de classificação errada de uma árvore $T$ em relação ao conjunto de treino $S$ , $Z$ é a inversa da distribuição normal acumulada e $\alpha$ é o nível de importância desejado.	(QUINLAN, 1993)
Poda Ótima	Algoritmo de poda ótima baseado em programação dinâmica.	(BOHANEK e BRATKO, 1986)
Poda com Comprimento Mínimo	Esse tipo de poda mede o tamanho de uma DT através do número de <i>bits</i> necessários para codificar a árvore, dando preferência em versões da árvore com menor número de <i>bits</i> .	(QUINLAN e RIVEST, 1989; RISSANEN, 1989)

Fonte: Elaborada pelo autor.

### **B.3. Métodos de Indução de uma Árvore de Decisão**

Nas seções B.1 e B.2 foram comentadas e apresentadas equações e procedimentos relacionados com os principais critérios de abertura de nós durante a construção de uma DT, e de poda para obter uma melhor capacidade de generalização da árvore. Essas técnicas são adotadas por vários dos métodos de construção de uma DT, que serão comentados a seguir (ROKACH e MAIMON, 2005; ROKACH e MAIMON, 2008).

Esses métodos de construção de árvores podem ser implementados a partir do algoritmo básico descrito na Figura 8 (Capítulo 3), substituindo as rotinas de construção de nós e de poda pelos procedimentos e equações relacionados com os critérios de construção de nós e de poda adotados pelos métodos de indução.

#### **B.3.1. ID3**

Proposto por (QUINLAN, 1986), é considerado como um algoritmo simples de construção de uma DT. O método utiliza o ganho de informação (ver a Tabela B.1) como critério para a construção dos nós. O crescimento é interrompido quando todas as instâncias pertencem a um único valor de função objetivo ou quando o ganho de informação obtido já não é maior do que zero. Na sua forma normal, o ID3 não utiliza procedimentos de poda. Também, não pode lidar com atributos numéricos e nem valores faltando nos atributos de entrada.

#### **B.3.2. CART**

Pode ser utilizado para árvores de classificação e regressão. Foi desenvolvido por (BREIMAN *et al.*, 1984), sendo caracterizado pelo fato construir árvores binárias, ou seja, cada nó interno tem exatamente dois ramos de saída. As divisões são selecionadas segundo o critério Twoing (ver Tabela B.2), e a árvore obtida é podada pelo critério do custo de complexidade (Tabela B.3).

Uma característica importante do CART é sua capacidade de gerar árvores de regressão. Em caso de regressão, o método procura por divisões que minimizem o erro quadrático de previsão. A previsão de cada folha é determinada a partir da média ponderada para o nó.

### **B.3.3. CHAID**

Esse método (KASS, 1980) foi desenvolvido originalmente para utilização em estatística aplicada. O CHAID foi originalmente concebido para lidar com atributos nominais apenas. Para cada atributo de entrada  $a_i$ , o método encontra o par de valores em  $V_i$  que menos difere do valor de saída do atributo. Essa diferença é medida pelo um valor  $p$  obtido a partir de um teste estatístico, que depende do tipo de atributo de destino (se é contínuo ou não).

### **B.3.4. QUEST**

Proposto por (LOH e SHIH, 1997) este método suporta a criação de nós através de análise univariável ou de uma combinação linear. Para cada divisão, a associação entre cada atributo de entrada e o atributo de destino é calculada. Se o atributo de destino é multinomial, dois centros, considerando cada metade do agrupamento, são usados para criar duas superclasses. O atributo que obtém a maior associação com o atributo alvo é selecionado para a separação. A Análise Discriminante Quadrática (ADQ) é aplicada para encontrar o ponto ideal para dividir o atributo de entrada. O método utiliza validação cruzada no processo de poda da árvore.

### **B.3.5. C4.5**

Representa um dos algoritmos mais utilizados para construção de uma DT, proposto originalmente em (QUINLAN, 1996). O método C4.5 é uma extensão do algoritmo ID3 (QUINLAN, 1986), com vantagens significativas (ROKACH e MAIMON, 2008): trata tanto com atributos categóricos (ordinais ou não) e com atributos contínuos, pode trabalhar com

valores desconhecidos, utiliza a medida da relação de ganho (Tabela B.2) para selecionar o atributo que melhor divide os nós na construção da árvore, pode lidar com problemas em que os atributos têm custos diferentes, e apresenta um método de poda para a árvore gerada, Poda Baseada no Erro (ver Tabela B.3).