

UNIVERSIDADE FEDERAL DO PARANÁ

RAFAEL MARTINS

**CLASSIFICAÇÃO DE TRANSFORMADORES DE DISTRIBUIÇÃO DE ENERGIA
ELÉTRICA QUANTO À DHTV USANDO *ROUGH SETS***

CURITIBA

2010

RAFAEL MARTINS

**CLASSIFICAÇÃO DE TRANSFORMADORES DE DISTRIBUIÇÃO DE ENERGIA
ELÉTRICA QUANTO À DHTV USANDO *ROUGH SETS***

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Elétrica, Programa de Pós-Graduação em Engenharia Elétrica – PPGEE, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. Dr. Alexandre Rasi Aoki

Co-Orientador: Prof. Dr. Anselmo Chaves Neto

CURITIBA

2010

TERMO DE APROVAÇÃO

RAFAEL MARTINS

CLASSIFICAÇÃO DE TRANSFORMADORES DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA QUANTO À DHTV USANDO *ROUGH SETS*

Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre no Curso de Pós-Graduação em Engenharia Elétrica, Setor de Tecnologia da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:



Prof. Dr. Alexandre Rasi Aoki

Departamento de Engenharia Elétrica, UFPR

Co-orientador:



Prof. Dr. Anselmo Chaves Neto

Departamento de Estatística, UFPR



Profa. Dra. Thelma Solange Piazza Fernandes

Departamento de Engenharia Elétrica, UFPR



Prof. Dr. Germano Lambert Torres

Instituto de Sistemas Elétricos e Energia, UNIFEI

Curitiba, 26 de Fevereiro de 2010.

AGRADECIMENTO

Algumas pessoas tiveram um papel de grande importância durante o desenvolvimento deste trabalho, pelo apoio científico ou moral. Assim, gostaria de desejar meus agradecimentos:

Ao Professor Doutor Alexandre Rasi Aoki, que teve papel fundamental através de seus ensinamentos, orientações e incentivos durante toda a realização deste trabalho.

Ao Professor Doutor Anselmo Chaves Neto que forneceu considerações importantes durante o desenvolvimento deste projeto através da sua co-orientação.

Ao Instituto de Tecnologia para o Desenvolvimento – LACTEC – que forneceu toda a infra-estrutura necessária para a concretização deste estudo.

Aos meus pais, Adonay Martins e Joceli Regina Martins, que sempre me apoiaram em todos os momentos da minha vida.

Ao meu filho, Lucas Monteiro Martins, que me estimulou a lutar para obter o melhor.

À professora Thelma Solange Piazza Fernandes por suas sugestões durante a pré-defesa e defesa final.

Ao professor Doutor Germano Lambert Torres por sua contribuição durante a defesa final através de seu grande conhecimento à teoria aplicada neste estudo.

Ao professor Leandro dos Santos Coelho por sua contribuição e sugestões durante a fase de pré defesa.

A Deus que me deu força pra chegar até aqui.

E a todos as outras pessoas que de alguma forma contribuíram para o sucesso deste trabalho.

RESUMO

A Qualidade de Energia Elétrica (QEE) tem se tornado um tema importante, nos últimos anos, devido a dois fatores principais: o aumento da competitividade entre as empresas fornecedoras de energia elétrica e aumento de cargas não lineares no sistema elétrico, que gera distorções na forma de onda fundamental fornecida pelas concessionárias. Neste contexto, este trabalho tem como objetivo contribuir com a QEE através da identificação de transformadores de tensão da rede de distribuição que possuem elevada distorção harmônica de tensão (DHTV). Para isso é utilizado um banco de dados obtido através de uma campanha de medições realizada no estado do Paraná e que contém informações das características de demanda de transformadores da rede da Companhia Paranaense de Energia (Copel). Estas informações em conjunto com as características nominais dos transformadores e da DHTV servem de base de dados para a classificação dos mesmos quanto à sua DHTV. Devido ao fato de que a Copel está dividida na sua estrutura em cinco regionais, cada uma atendendo uma parte do estado, o banco de dados também foi disposto em cinco partes, uma para cada regional. Para a classificação dos transformadores, segundo recomendações do Operador Nacional do Sistema Elétrico (ONS) para sistemas de potência com tensão abaixo de 69kV, buscou-se identificar quais transformadores possuem DHTV superior a 6% levando em consideração suas características de consumo e características nominais. Para isso, aplicou-se a teoria dos conjuntos aproximados (*rough sets*). Esta teoria, desenvolvida por Zdzislaw Pawlak na década de 1980 e tem sido cada vez mais explorada e aplicada em sistemas elétricos para a classificação e também para a eliminação de informações irrelevantes de bancos de dados e se mostrou muito eficiente na aplicação deste trabalho. Através desta metodologia é possível extrair regras que podem, posteriormente, ser aplicadas a novas amostras nas quais o objetivo é estimar se a DHTV é superior ou inferior a 6%. Com esta estimativa é possível direcionar uma eventual campanha de medições evitando o desperdício de tempo e mão de obra. Na validação da aplicação desta metodologia, as regras extraídas foram aplicadas ao banco de dados que as originaram, para a confirmação da capacidade de discernibilidade das faixas. Os resultados foram comparados com os obtidos através da aplicação de outras técnicas, tais como regressão logística e escore quadrático. Nesta comparação *rough sets* demonstrou uma melhor capacidade de classificação.

Palavras chave: *Rough Sets*. Qualidade da Energia Elétrica. Distorção Harmônica Total de Tensão. Classificação.

ABSTRACT

Power quality has become an important research field due to two factors: the growing competitiveness in power system and the increase of non-linear loads in the system, serving as a cause of voltage distortion in distribution systems. In this sense, this work has the aim of improve power quality by indentifying power distribution transformers which have voltage total harmonic distortion (VTHD) above the brazilian limits. It was used a database with monitored data and electrical characteristics obtained during a monitoring campaign in Energy Company at Paraná (COPEL). This information was used to classify the transformers by VTHD, according to a physical division in 5 regions in Paraná State. The VTHD limit used was approved by Power System National Operator (ONS) at Brazil and it has the value of 6% for Power Systems with voltage under 69kV. It was applied the Rough Sets Theory (RST), developed by Zdzislaw Pawlak in the 80's, which has increasing usage to Power Systems analysis in classification problems and removal of irrelevant information in databases. The application of RST was successful and it allows the acquirement of rules which makes possible the estimation of VTHD in another power distribution transformers. So, with this VTHD estimated it is possible to guide a monitoring campaign avoiding time and money losses. To validate the application of rough sets Theory the rules obtained were applied to the original database to assure the discernibility capability of created sets. The results were compared with other techniques such as quadratic scores and logistic regression, applied to the same problem, and RST had better classification capability.

Key-words: Rough Sets. Power Quality. Voltage Total Harmonic Distortion. Classification.

LISTA DE ILUSTRAÇÕES

Figura 1 - Exemplo de regiões positiva, de fronteira e negativa.....	15
Figura 2 - Definição das regiões e dos conjuntos de aproximação.....	16
Figura 3 – Sistema elétrico hipotético	17
Figura 4 - Representação gráfica dos conjuntos redução e núcleo básico	20
Figura 5 - Representação da onda fundamental e da componente harmônica.....	35
Figura 6 - Onda fundamental somada à componente harmônica	35
Figura 7 - Harmônica versus corrente harmônica	36
Figura 8 – Período de monitoração e período de análise	48
Figura 9 – Exemplo de DHTV – abaixo de 6%.....	49
Figura 10 – Exemplo de DHTV – acima de 6%.....	49
Figura 11 – Espectro da tensão no período de maior distorção – abaixo de 6% ...	51
Figura 12 – Espectro da tensão no período de maior distorção – acima de 6%	51
Figura 15 - Importação do banco de dados	57
Figura 16 - Geração dos redutos	58
Figura 17 - Geração de regras.....	58
Figura 18 – Regras geradas.....	60
Figura 19 - Resultado da aplicação das regras.....	61

LISTA DE TABELAS

Tabela 1 – Conjunto de casos possíveis.....	18
Tabela 2 – Conjunto de exemplos sem o atributo B.....	22
Tabela 3 – Núcleos básicos dos casos	24
Tabela 4 – Núcleos redução dos casos	25
Tabela 5 – Conjuntos redução dos casos	26
Tabela 6 – Limites globais de distorção harmônica de tensão.....	39
Tabela 8 – Características da Copel para a atuação no Paraná.....	45
Tabela 9 – Características da Copel – dados da distribuição	45
Tabela 10 – Distribuição dos pontos de monitoração	47
Tabela 11 – Distribuição das amostras entre áreas urbanas e rurais	47
Tabela 12 – Valores máximos, mínimos e P95% da DHTV - dentro do limite (6%)50	
Tabela 13 – Valores máximos, mínimos e P95% da DHTV - fora do limite (6%)... 50	
Tabela 14 – Número de transformadores urbanos medidos	52
Tabela 18 – Variáveis associadas às características nominais do transformador .	62
Tabela 19 – Variáveis associadas à característica do consumo	62
Tabela 20 - Dados com número de consumidores em valores absolutos.....	63
Tabela 21 - Dados com número de consumidores em porcentagem.....	63
Tabela 22 – Tabela de classificação.....	65
Tabela 23 – Matriz de classificação e APER - SDC.....	67
Tabela 24 – Matriz de classificação e APER - SDL	68
Tabela 25 – Matriz de classificação e APER - SDN.....	68
Tabela 26 – Matriz de classificação e APER - SDO.....	69
Tabela 27 – Matriz de classificação e APER - SDT	69

Tabela 28 – Matriz de classificação e APER – Agrupamento de todos os bancos de dados	70
Tabela 29 – Comparação entre as metodologias utilizadas para a classificação de transformadores quanto à DHTV	71
Tabela 30 – Índices de acertos de classificação – <i>leave-one-out</i>	72
Tabela 31 – Número de amostras por regional	73

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 CONTEXTO DO PROBLEMA.....	11
1.2 JUSTIFICATIVA.....	12
1.3 OBJETIVOS.....	12
1.3.1 Objetivo Geral	12
1.3.2 Objetivos Específicos	12
1.4 ESTRUTURA DA DISSERTAÇÃO	13
2 <i>ROUGH SETS</i>	14
2.1 ATRIBUTOS DISPENSÁVEIS E INDISPENSÁVEIS.....	19
2.2 CONCEITOS DE CONJUNTO REDUÇÃO E NÚCLEO BÁSICO	19
2.3 ALGORITMO DE CLASSIFICAÇÃO DE DADOS	21
2.4 ESTADO-DA-ARTE SOBRE APLICAÇÕES DE <i>ROUGH SETS</i>	26
2.5 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	31
3 QUALIDADE DA ENERGIA ELÉTRICA (QEE).....	32
3.1 PARÂMETROS DE QEE	32
3.2 DISTORÇÃO HARMÔNICA TOTAL DE TENSÃO (DHTV)	33
3.3 DESEQUILÍBRIO DE TENSÃO	39
3.4 PERFIL DE TENSÃO (LIMITES DE VARIAÇÃO DE TENSÃO)	40
3.5 ESTADO-DA-ARTE SOBRE CLASSIFICAÇÃO DE PARÂMETROS DE QUALIDADE DE ENERGIA ELÉTRICA.....	41
3.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	43
4 MATERIAIS E MÉTODO.....	44
4.1 MATERIAIS	44
4.1.1 Campanha de Medições	44

4.1.2 Área de Atuação.....	44
4.1.3 Medições.....	46
4.1.4 Dados Obtidos.....	52
4.1.5 <i>Software</i>	57
4.2 MÉTODO.....	61
4.2.1 Variáveis.....	61
4.2.2 Tratamento dos Dados.....	63
4.2.3 Forma de Apresentação dos Resultados.....	64
4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	65
5 ANÁLISE DOS RESULTADOS.....	67
5.1 VALIDAÇÃO DO TRATAMENTO DE DADOS.....	67
5.2 COMPARAÇÃO COM OUTRAS METODOLOGIAS DA LITERATURA.....	70
5.3 EFICÁCIA DA METODOLOGIA.....	72
5.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	73
6 CONCLUSÃO E TRABALHOS FUTUROS.....	75
7 REFERÊNCIAS BIBLIOGRÁFICAS.....	77

1 INTRODUÇÃO

1.1 CONTEXTO DO PROBLEMA

A Qualidade de Energia Elétrica - QEE - tem despertado muito interesse dos órgãos ligados ao setor de energia nesses últimos anos. Isto ocorre devido, principalmente, à reestruturação do setor energético que obriga as empresas a alcançarem um mínimo de qualidade, medida através dos índices exigidos pelos órgãos regulamentadores e da busca por melhores resultados devido à livre concorrência existente entre as companhias, a fim de evitar perdas e, com isso, tornar a empresa eficiente e competitiva.

Existem distúrbios na rede de energia elétrica que podem afetar a QEE e causar prejuízos a diversos setores. Isto afeta desde empresas ligadas ao fornecimento de energia até consumidores residenciais e industriais. Tais distúrbios podem danificar cargas ligadas ao sistema elétrico e estes danos causam o mau funcionamento ou até mesmo a parada total e permanente de determinados equipamentos. Alguns fenômenos que conduzem a esses prejuízos são: transitórios (chaveamento de banco de capacitores, cargas etc.), sobretensões, afundamentos de tensão, descargas atmosféricas, distorções harmônicas entre outros.

A distorção harmônica total de tensão - DHTV - neste contexto de distúrbios merece especial atenção, pois é um fenômeno que cresceu muito com o aumento da inserção de equipamentos eletrônicos no sistema, ou seja, cargas não lineares. Este fenômeno provoca diversos problemas na rede, causando um aumento dos custos da empresa fornecedora de energia elétrica. Para diminuir seus efeitos é importante obter o maior número possível de dados do sistema e, para isso, é necessário realizar medições das características elétricas do circuito no local, onde ocorre o fenômeno. A dificuldade está em descobrir onde há distorção harmônica elevada e a única forma de realizar esta tarefa é realizar a medição com analisadores de QEE.

1.2 JUSTIFICATIVA

A aplicação da técnica de *rough sets*, para a estimação da DHTV, pode, previamente, apenas com os dados das características de consumo e de características nominais de transformadores da rede de distribuição, estabelecer quais os locais têm maior chance de fornecer dados relevantes às medições, ou seja, DHTV acima dos limites estabelecidos. Com isso, é possível diminuir os custos envolvidos, principalmente, com o deslocamento e o tempo necessários para o monitoramento e, assim, viabilizar a melhoria do sistema elétrico quanto à QEE, diminuindo os prejuízos causados por estes distúrbios existentes na rede de energia.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é realizar a classificação de transformadores da rede de distribuição, quanto à DHTV, levando-se em consideração suas características de uso e elétricas, através da aplicação da técnica de *rough sets*. Especificamente pretende-se classificar um transformador quanto à DHTV igual ou superior a 6%.

1.3.2 Objetivos Específicos

Este trabalho contempla os seguintes objetivos específicos:

- realizar o levantamento bibliográfico sobre a técnica de *rough sets* e suas aplicações;
- realizar o tratamento dos dados obtidos na campanha de medições;
- aplicar a técnica de *rough sets* através de ferramentas computacionais;
- analisar o desempenho da metodologia aplicada;

- comparar a eficiência da metodologia aplicada com outros resultados da literatura.

1.4 ESTRUTURA DA DISSERTAÇÃO

A estrutura desta dissertação está dividida em seis capítulos. No primeiro estão dispostas a introdução, contexto do problema, justificativa e objetivos. A revisão de literatura abordada neste trabalho é apresentada nos segundo e terceiro capítulos, sendo que no segundo capítulo é apresentada a teoria de *rough sets* com um exemplo ilustrativo e o estado-da-arte sobre suas aplicações. Já o terceiro capítulo contém os principais conceitos de QEE bem como os principais indicadores relacionados com a tensão e o estado-da-arte sobre classificação de parâmetros de QEE.

No capítulo quatro estão descritos os materiais e métodos, destacando-se a metodologia utilizada na campanha de medições que originou o banco de dados utilizado neste trabalho e a metodologia utilizada para a aplicação de *rough sets* para classificação de transformadores quanto à DHTV. No quinto capítulo estão dispostos os resultados obtidos com a aplicação da *rough sets* e também comparações com outras teorias utilizadas encontradas na literatura e no sexto capítulo as conclusões e sugestões de trabalhos futuros. Por fim, estão descritas as referências bibliográficas.

2 ROUGH SETS

2.1 INTRODUÇÃO

A teoria dos conjuntos aproximados (*rough sets*) foi desenvolvida por Pawlak (1982) e hoje é classificada como uma técnica poderosa, convergindo com áreas de grande interesse no campo das Ciências Cognitivas e da Inteligência Artificial.

Assim, a estrutura operacional dos Conjuntos Aproximados é suportada por essa Teoria do Conhecimento e está, fundamentalmente, baseada em dois conjuntos denominados de “aproximação inferior” e “aproximação superior” e que serão aqui denotados por $\underline{R}X$ e $\overline{R}X$, respectivamente.

A ideia central destes dois conjuntos é tentar limitar a solução de um dado problema, onde podem existir duas regiões, uma na qual todos os elementos do conjunto fazem parte da solução (aproximação inferior) e outra em que os elementos possivelmente fazem parte da solução (aproximação superior). Portanto, todos os elementos do conjunto de aproximação inferior fazem parte do conjunto de aproximação superior.

Isto define três regiões de interesse: região positiva, região fronteira e região negativa, com a seguinte notação, respectivamente, $POS_R(X)$, $BN_R(X)$, $NEG_R(X)$.

A região positiva coincide com o conjunto de aproximação inferior, sendo composta, portanto, de elementos que fazem parte da solução do problema. A região fronteira é a diferença entre os dois conjuntos de aproximação e é composta por elementos que podem ou não fazer parte da solução. Finalmente, a região negativa contém elementos que não fazem parte da solução.

Um exemplo dessas três regiões está mostrado na Figura 1, que tem o objetivo de determinar a região de solução gerada pela linha em destaque, isto é, a que delimita as duas regiões, superior e inferior.

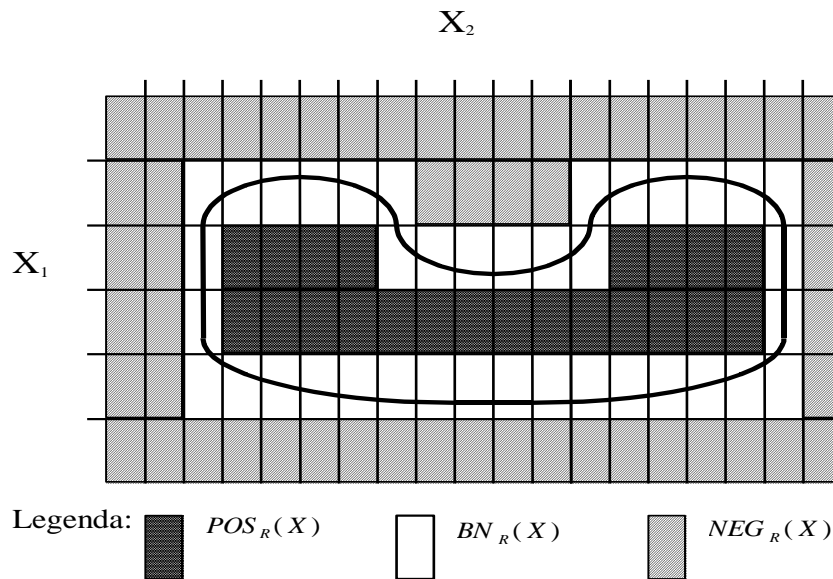


Figura 1 - Exemplo de regiões positiva, de fronteira e negativa

Matematicamente, os conjuntos aproximados podem ser definidos como mostrado a seguir.

Seja o conjunto $X \subseteq U$, e seja R uma relação de equivalência e $K = (U, R)$, uma base de conhecimento, que os associa. Os dois conjuntos aproximados definidos anteriormente podem então ser escritos como sendo:

a) R-inferior: $\underline{R}X = U \{Y \in U \mid R: Y \subseteq X\}$

b) R-superior: $\bar{R}X = U \{Y \in U \mid R: Y \cap X \neq \emptyset\}$

Isto significa que os elementos pertencentes ao conjunto $\underline{R}X$ certamente podem ser classificados como elementos de X ; enquanto os pertencentes a $\bar{R}X$, possivelmente, poderão vir a ser classificados como elementos de X .

Da mesma maneira, as regiões $POS_R(X)$, $BN_R(X)$ e $NEG_R(X)$ podem ser definidas conforme abaixo e ilustradas na Figura 2.

a) $POS_R(X) = \underline{R}(X) \Rightarrow$ certamente membro de X ;

b) $NEG_R(X) = U - \bar{R}(X) \Rightarrow$ certamente não-membro de X ;

c) $BN_R(X) = \bar{R}(X) - \underline{R}(X) \Rightarrow$ possivelmente membro de X .

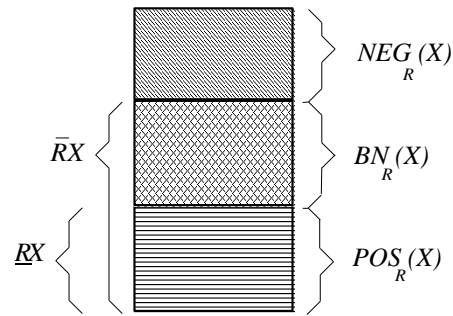


Figura 2 - Definição das regiões e dos conjuntos de aproximação

Evidentemente, essas relações de pertinência assim definidas, apresentam além das propriedades operacionais clássicas, tais como a associativa, a distributiva e a multiplicativa, uma série de outras operações que se encontram relacionadas a seguir:

- a) $\underline{R}X \subseteq X \subseteq \bar{R}X$;
- b) $\underline{R}\phi = \bar{R}\phi = \phi$;
- c) $\underline{R}U = \bar{R}U = U$;
- d) $\bar{R}(X \cup Y) = \bar{R}X \cup \bar{R}Y$
- e) $\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y$;
- f) $X \subseteq Y \Rightarrow \underline{R}X \subseteq \underline{R}Y$;
- g) $X \subseteq Y \Rightarrow \bar{R}X \subseteq \bar{R}Y$;
- h) $\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y$;
- i) $\bar{R}(X \cap Y) \subseteq \bar{R}X \cap \bar{R}Y$;
- j) $\underline{R}(-X) = -\bar{R}X$;
- l) $\bar{R}(-X) = -\underline{R}X$;
- m) $\underline{R}\underline{R}X = \bar{R}\underline{R}X = \underline{R}X$;
- n) $\bar{R}\bar{R}X = \underline{R}\bar{R}X = \bar{R}X$.

O conceito de precisão ou exatidão de uma medida, $\alpha_R(X)$, o qual caracteriza numericamente a imprecisão do conhecimento, pode agora ser apresentado, usando o conceito da cardinalidade dos conjuntos superior ($\bar{R}X$) e inferior ($\underline{R}X$); isto é:

$$\alpha_R(X) = \frac{\text{card } \underline{R}X}{\text{card } \bar{R}X} \quad (1)$$

onde: $\alpha_R(X)$ é definido no intervalo $[0,1]$;

$card \underline{R}X$ - número de elementos contidos em $\underline{R}X$;

$card \bar{R}X$ - número de elementos contidos em $\bar{R}X$.

Quando $\alpha_R(X) = 1$, o conjunto X será denominado de R -definível (R -definable) e, portanto, a região $BN_R(X)$ será vazia. Neste caso, a teoria dos conjuntos aproximados é reduzida à clássica teoria dos conjuntos cantor.

Outra definição necessária é sobre o conceito de “qualidade de aproximação”, que matematicamente é expressa por:

$$\gamma_R(X) = \frac{card \underline{R}X}{card U} \quad (2)$$

onde: $\gamma_R(X)$ - qualidade de aproximação;

$card U$ - número de elementos do universo U .

Este valor representa a porcentagem de objetos que podem ser corretamente classificados pelo atributo R .

Para clareza das ideias, pode-se sugerir um sistema elétrico hipotético, mostrado na Figura 3, no qual sua classificação operativa depende de 4 elementos, a saber: potência transmitida através de duas linhas A e B, da potência gerada na área sob análise, C, e do consumo da carga, D. Estes elementos podem assumir os seguintes valores:

- linha de transmissão \rightarrow {baixo, médio, normal},
- geração da área \rightarrow {baixo, normal, alto},
- carga própria \rightarrow {baixo, médio, alta}.

Enquanto, o estado operativo pode assumir os seguintes valores:

- estado operativo \rightarrow {restaurativo, emergência, normal}.

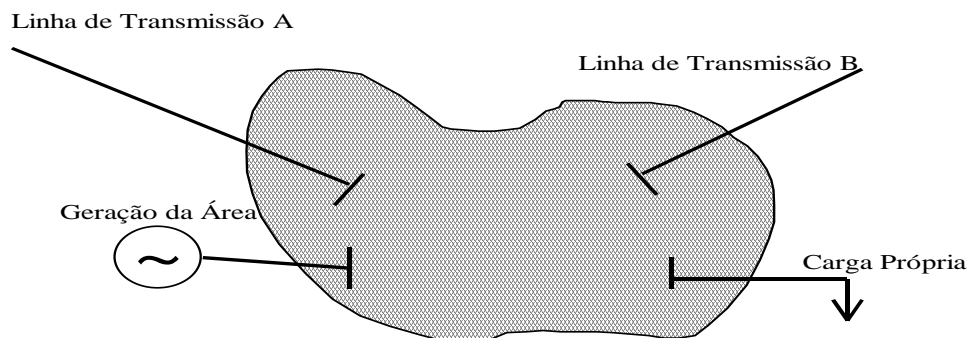


Figura 3 – Sistema elétrico hipotético

Utilizando estes valores, um conjunto de exemplos pode ser gerado. Este conjunto, apresentado na Tabela 1, está incompleto, pois se todas as combinações possíveis fossem realizadas existiriam 27 casos e não somente 8. Na maioria das vezes, os problemas do sistema elétrico são analisados considerando somente conjuntos incompletos de casos. Pode-se observar que, nos sistemas elétricos reais, é quase impossível se obter dados em todos os intervalos existentes ou então produzir todas as combinações entre eles, pois o número de pontos observado é enorme. Portanto, um conjunto de exemplos em um sistema real é sempre incompleto.

Tabela 1 – Conjunto de casos possíveis

Exemplo	Linha A	Linha B	Geração	Carga	Resultado
E1	médio	normal	baixo	médio	restaurativo
E2	médio	médio	baixo	alto	restaurativo
E3	baixo	normal	baixo	médio	restaurativo
E4	baixo	normal	alto	médio	emergencial
E5	médio	baixo	alto	médio	emergencial
E6	baixo	normal	alto	baixo	normal
E7	alto	normal	alto	baixo	normal
E8	alto	normal	normal	baixo	normal

Tomando-se o atributo A (Linha de Transmissão A) como referência, podem-se considerar os seguintes conjuntos elementares de A:

$$U \mid IND(A) = \{ \{E1, E2, E5\}, \{E3, E4, E6\}, \{E7, E8\} \}, \text{ e portanto,}$$

$$\bar{A}X = \{E1, E2, E5\} \cup \{E3, E4, E6\} = \{E1, E2, E3, E4, E5, E6\}$$

$$\underline{A}X = \{E1, E2, E5\}$$

$$BN_A(X) = \{E3, E4, E6\}$$

$$NEG_A(X) = \{E7, E8\}$$

A partir desses dados, pode-se determinar o valor de $\alpha_A(X)$, onde se encontra o valor:

$$\alpha_A(X) = \frac{\text{card } \underline{A}X}{\text{card } \bar{A}X} = 3/6, \text{ assim } = 50\%.$$

E de forma análoga, para a outra definição apresentada, conceito de qualidade de aproximação, para essa relação de atributos, tem-se:

$$\gamma_A(X) = \frac{\sum \text{card } RX}{\text{card } U} = 3/8, \text{ assim } = 37,5 \%$$

2.2 ATRIBUTOS DISPENSÁVEIS E INDISPENSÁVEIS

Na Teoria dos Conjuntos Aproximados, pode-se classificar os atributos de duas formas distintas: os dispensáveis e os indispensáveis. Os atributos dispensáveis são aqueles que se omitidos (ou se não existissem) não trariam nenhum problema de classificação. Aqueles atributos que, com a sua omissão, criam problemas na classificação são chamados indispensáveis.

Existem várias formas para a verificação do tipo do atributo, uma delas é verificar o conjunto induzido com e sem o atributo em questão. Por exemplo, na Tabela 1, observando o atributo B, referente à Linha de Transmissão B, pode-se induzir à seguinte classificação com e sem este atributo:

$$U \mid \text{IND}(R) = \{\{E1\}, \{E2\}, \{E3\}, \{E4\}, \{E5\}, \{E6\}, \{E7\}, \{E8\}\},$$

$$U \mid \text{IND}(R-\{B\}) = \{\{E1\}, \{E2\}, \{E3\}, \{E4\}, \{E5\}, \{E6\}, \{E7\}, \{E8\}\},$$

e pode-se então, verificar que:

$$U \mid \text{IND}(R) = U \mid \text{IND}(R-\{B\}),$$

o que implica dizer que a relação B é dispensável na relação R.

Analogamente, por outro lado, pode-se dizer que a relação C é indispensável sobre R, pois:

$$U \mid \text{IND}(R-\{C\}) = \{\{E1\}, \{E2\}, \{E3, E4\}, \{E5\}, \{E6\}, \{E7, E8\}\} \neq U \mid \text{IND}(R).$$

2.3 CONCEITOS DE CONJUNTO REDUÇÃO E NÚCLEO BÁSICO

Os conceitos de conjunto redução (*reduct*) e núcleo básico (*core*) são importantes com relação as suas aplicações na redução de uma base de conhecimento. Matematicamente, podem-se definir os conjuntos redução e núcleo básico, utilizando uma família de relações equivalentes R . O conjunto redução de R ,

RED (R), é definido como um conjunto reduzido de atributos que conserva a mesma classificação indutiva da família R . O conjunto núcleo básico de R , CORE (R), é aquele que aparece em todos os conjuntos redução de R , isto é, o conjunto de elementos indispensáveis para caracterizar a relação R .

Em outras palavras, o conjunto redução é a estrutura mínima de representação do conhecimento existente em uma base. Ela pode não ser única, pois diversos conjuntos “redução” podem ser conseguidos em uma base de conhecimento, dependendo dos atributos utilizados para a sua representação. Esses atributos correspondem aos parâmetros analisados no sistema monitorado e aos quais serão associados os respectivos valores das medições efetuadas, que se constituirá na base de dados mencionada.

O núcleo básico é composto pelos elementos necessários para compor qualquer conjunto redução, ou seja, em outras palavras é a interseção desses conjuntos.

A Figura 4 mostra, de maneira gráfica, uma base de dados composta pelos seguintes atributos: $\{A, B, C, X, Y, Z, J, K, W\}$, com 3 possíveis conjuntos redução, definidos pelo atributos $\{A-B-C\}$, $\{A-C-X\}$ e $\{A-Y-Z\}$, onde somente o atributo A faz parte do núcleo básico. Com isto, os atributos J, K e W são ditos dispensáveis, pois não fazem parte de nenhum dos conjuntos redução, e, por exemplo, X, Y e Z são ditos dispensáveis em relação ao conjunto redução $\{A-B-C\}$.

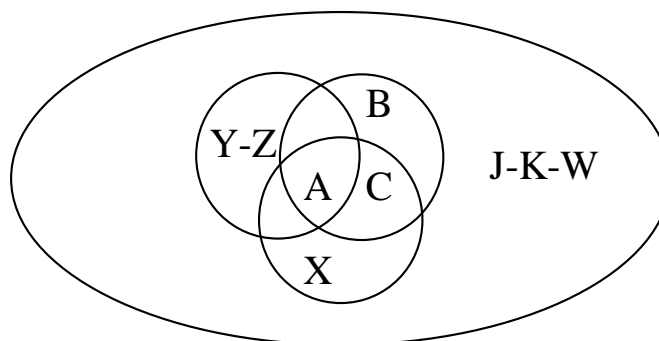


Figura 4 - Representação gráfica dos conjuntos redução e núcleo básico

2.4 ALGORITMO DE CLASSIFICAÇÃO DE DADOS

O algoritmo de redução de uma tabela de decisão, exemplificado na tabela 1, pode ser formulado através de desenvolvimentos algébricos ou por relações lógicas. Neste trabalho, por simplicidade de explanação, utilizou-se o segundo encaminhamento. A ideia central deste algoritmo é retirar o conhecimento de uma base de dados através da redução e compactação de suas informações. Esse algoritmo tem os seguintes passos de processamento:

- Passo 1 - Transformar valores contínuos em faixas;
- Passo 2 - Eliminar atributos idênticos;
- Passo 3 - Eliminar exemplos idênticos;
- Passo 4 - Eliminar atributos dispensáveis;
- Passo 5 - Calcular o conjunto básico da tabela de decisão;
- Passo 6 - Compor a tabela do conjunto redução;
- Passo 7 - Agrupar os exemplos;
- Passo 8 - Compor o conjunto final de regras.

Considerando o mesmo exemplo ilustrativo mostrado na tabela 1, no qual sua classificação operativa depende de 4 elementos, a saber: potência transmitida em duas linhas A e B, da potência gerada na área sob análise, C, e do consumo da carga, D. Estes elementos podem assumir os seguintes valores:

- linha de transmissão → {baixo, médio, normal},
- geração da área → {baixo, normal, alto},
- carga própria → {baixo, médio, alta},
- estado operativo → {restaurativo, emergência, normal}.

Pode-se, então, dizer que o sistema de informações proposto é constituído por um conjunto R, tal que:

$$R = \{A, B, C, D, R \mid \{A,B,C,D\} \Rightarrow \{R\}\},$$

sendo A, B, C, D e R, os atributos associados às linhas de transmissão A e B, à geração própria, à carga existente e o ponto de operação, respectivamente.

Dessa forma, seguindo os passos apresentados anteriormente, o processo de classificação apresenta os seguintes resultados.

Passo 1: Os valores já estão em faixas, ou seja, não são apresentados em valores contínuos.

Passo 2: A Tabela 1 não apresenta atributos repetidos.

Passo 3: A Tabela 1 não apresenta exemplos repetidos.

Passo 4: Observa-se que o atributo *B* é dispensável, pois a sua eliminação não resulta em nenhum problema de classificação. Assim, a tabela inicial é reduzida para a apresentada na Tabela 2 e o problema fica reduzido a $R = \{A, C, D, R \mid \{A,C,D\} \Rightarrow \{R\}\}$.

Tabela 2 – Conjunto de exemplos sem o atributo B

Exemplo	Linha A	Geração	Carga	Resultado
E1	médio	baixo	médio	restaurativo
E2	médio	baixo	alto	restaurativo
E3	baixo	baixo	médio	restaurativo
E4	baixo	alto	médio	emergencial
E5	médio	alto	médio	emergencial
E6	baixo	alto	baixo	normal
E7	alto	alto	baixo	normal
E8	alto	normal	baixo	normal

Neste passo, devem-se eliminar todos os atributos que forem dispensáveis. Entretanto, se isto não for feito, os atributos que são dispensáveis serão eliminados no conjunto redução. A eliminação desses atributos, neste ponto do algoritmo, visa somente reduzir o esforço computacional.

Passo 5: Utilizando-se a Tabela 2, deve-se calcular o núcleo básico de cada caso.

Processamento para determinar o núcleo básico de *E1*:

Para este exemplo, a decisão a ser utilizada é dada por: $X1=\{E1,E2,E3\}$, ou seja, estado operacional restaurativo. Tomando-se os atributos dois a dois têm-se as

relações abaixo. Elas são tomadas utilizando-se os valores do exemplo $E1$, que aparecem também nos outros exemplos, ou seja, $A(E1)$ = médio surge também nos exemplos 2 e 5, $A(E2)$ e $A(E5)$. Daí observa-se a intersecção dos conjuntos e verifica-se se o resultado faz parte do conjunto de decisão procurado.

$$\{E1, E2, E5\}A \cap \{E1, E2, E3\}C = \{E1, E2\} \subset X1,$$

$$\{E1, E2, E5\}A \cap \{E1, E3, E4, E5\}D = \{E1, E5\} \not\subset X1,$$

$$\{E1, E2, E3\}C \cap \{E1, E3, E4, E5\}D = \{E1, E3\} \subset X1.$$

Verifica-se que as intersecções dos atributos $A - C$ e $C - D$ resultam em subconjuntos do conjunto de decisão procurado. Assim, observa-se o atributo que é comum, sendo ele(s) o(s) atributo do núcleo básico. Neste exemplo, o valor do núcleo básico é dado por: $C(E1)$ = baixo.

Analogamente para se determinar o núcleo básico do exemplo $E2$, o conjunto de decisão é dado por $X2 = \{E1, E2, E3\}$ e tem-se as seguintes relações:

$$\{E1, E2, E5\}A \cap \{E1, E2, E3\}C = \{E1, E2\} \subset X2,$$

$$\{E1, E2, E5\}A \cap \{E2\}D = \{E2\} \subset X2,$$

$$\{E1, E2, E3\}C \cap \{E2\}D = \{E2\} \subset X2.$$

Verifica-se que as intersecções dos atributos $A - C$, $A - D$ e $C - D$ resultam em subconjuntos do conjunto de decisão procurado. Isto significa que o valor do núcleo não pode ser expresso por um dos atributos ou por apenas um conjunto deles.

Para a determinação do núcleo básico do caso $E4$, o conjunto de decisão é dado por $X4 = \{E4, E5\}$ e tem-se as seguintes relações:

$$\{E3, E4, E6\}A \cap \{E4, E5, E6, E7\}C = \{E4, E6\} \not\subset X4,$$

$$\{E3, E4, E6\}A \cap \{E1, E3, E4, E5\}D = \{E3, E4\} \not\subset X4,$$

$$\{E4, E5, E6, E7\}C \cap \{E1, E3, E4, E5\}D = \{E4, E5\} \subset X4.$$

Verifica-se que apenas a intersecção dos atributos $C - D$ resulta em um subconjunto do conjunto de decisão procurado. Isto significa que o valor do núcleo é dado pelos atributos C e D , ou seja, $C(E4)$ = alto e $D(E4)$ = médio.

Continuando-se estes cálculos, obtém-se a Tabela 3 com os núcleos básicos de cada caso.

Tabela 3 – Núcleos básicos dos casos

Exemplo	Linha A	Geração	Carga	Resultado
E1	-	baixo	-	restaurativo
E2	-	-	-	restaurativo
E3	-	baixo	-	restaurativo
E4	-	alto	médio	emergencial
E5	-	alto	-	emergencial
E6	-	-	baixo	normal
E7	-	-	-	normal
E8	-	-	-	normal

Passo 6: Utilizando-se a Tabela 3, deve-se calcular o conjunto redução de cada caso. A determinação do conjunto redução de E1 está descrita a seguir:

Para este exemplo, o núcleo básico é $C(E1) = \text{baixo}$. Como o conjunto redução contém o núcleo básico, as seguintes possíveis soluções devem ser testadas: C ou A - C e C - D, pois todas elas resultam no mesmo núcleo básico. Testam-se as possíveis soluções e verifica-se se o resultado faz parte do conjunto de decisão procurado, por exemplo:

$$\{E1, E2, E3\}C = \{E1, E2, E3\} \subset X1.$$

Como somente o atributo C já consegue a classificação devida, o conjunto redução é formado somente por este atributo. Os demais não necessitam ser testados pois são superconjuntos de C.

Para o exemplo da determinação do conjunto redução de E2, o núcleo básico é vazio. Portanto, as possíveis soluções a serem testadas são: (A) ou (C) ou (D) ou (A e C) ou (A e D) ou (C e D) ou (A e C e D) ou (A - C e D) ou (A e C - D) ou (A - D e C), pois todas elas resultam no mesmo núcleo básico vazio. Testam-se as possíveis soluções e verifica-se se o resultado faz parte do conjunto de decisão procurado.

$$\{E1, E2, E5\}A = \{E1, E2, E5\} \not\subset X2,$$

$$\{E1,E2,E3\}C = \{E1,E2,E3\} \subset X2,$$

$$\{E2\}D = \{E2\} \subset X2.$$

Desta forma, verifica-se que o atributo *A* não fornece a classificação devida; entretanto, os atributos *C* e *D* fornecem a solução requerida. As demais soluções não necessitam ser testadas, pois, são superconjuntos de *C* e *D*.

Analogamente, para se determinar o conjunto redução do exemplo *E4*, observa-se o conjunto de decisão, $X4=\{E4,E5\}$, e o núcleo básico, formado pelos atributos C-D. Conclui-se, assim, que a única solução possível para o conjunto redução é o próprio núcleo básico, o que é confirmado pela relação a seguir:

$$\{E4,E5,E6,E7\}C \cap \{E1,E3,E4,E5\}D = \{E4,E5\} \subset X4.$$

Continuando-se estes cálculos, obtém-se a Tabela 4 com os conjuntos redução de cada exemplo.

Tabela 4 – Núcleos redução dos casos

Exemplo	Linha A	Geração	Carga	Resultado
E1	-	baixo	-	restaurativo
E2'	-	baixo	-	restaurativo
E2''	-	-	alto	restaurativo
E3	-	baixo	-	restaurativo
E4	-	alto	médio	emergencial
E5'	médio	alto	-	emergencial
E5''	-	alto	médio	emergencial
E6	-	-	baixo	normal
E7'	alto	-	-	normal
E7''	-	-	baixo	normal
E8'	alto	-	-	normal
E8''	-	normal	-	normal
E8'''	-	-	baixo	normal

Passo 7: Agrupando os exemplos iguais, gera-se a Tabela 5:

Tabela 5 – Conjuntos redução dos casos

Regra	Linha A	Geração	Carga	Resultado	Exemplos
R1	-	baixo	-	restaurativo	E1,E2',E3
R2	-	-	alto	restaurativo	E2''
R3	médio	alto	-	emergencial	E5'
R4	-	alto	médio	emergencial	E4,E5''
R5	-	-	baixo	normal	E6,E7''',E8''''
R6	alto	-	-	normal	E7',E8'
R7	-	normal	-	normal	E8''

Passo 8: Assim, o conjunto final de regras é o seguinte:

R1: Se C =baixo ou D =alto então o sistema está em modo restaurativo.

R2: Se (C =alto e D =médio) ou (A =médio e C =alto) então o sistema está em estado de emergência.

R3: Se A =alto ou C = normal ou D =baixo então o sistema está normal.

Ou ainda,

R1: Se a geração é baixa ou a carga é alta então o sistema está em modo restaurativo.

R2: Se (a geração é alta e a carga é média) ou (a transmissão em A é média e a geração é alta) então o sistema está em estado de emergência.

R3: Se a transmissão em A é alta ou a geração é normal ou a carga é baixa então o sistema está normal.

2.5 ESTADO-DA-ARTE SOBRE APLICAÇÕES DE *ROUGH SETS*

Pawlak (1982) descreveu em seu trabalho a metodologia *rough sets*, citando algumas vantagens e exemplificando com uma aplicação prática desta teoria. O exemplo consistiu na análise de um banco de dados contendo atributos que

caracterizam sintomas médicos de pacientes e a saída, ou variável dependente, foi se o paciente possuía gripe ou não. Com isso, foram realizadas algumas comparações com outros métodos utilizados para extrair conhecimento de um banco de dados. Outros autores, tais como, Kryszkiewicz (1998), Massart e Walczak (1998), Swiniarski e Skowron (2003) realizaram trabalhos que se basearam também na apresentação da teoria de *rough sets* com tutoriais para a sua aplicação.

Jagielska, Matthews e Whitfort (1999) fizeram uma comparação entre diversas técnicas de classificação. As técnicas comparadas foram: redes neurais artificiais, sistemas nebulosos, algoritmos genéticos e *rough sets*. Para a comparação, 3 conjuntos de dados foram utilizados na aplicação de todas as teorias citadas, são eles: identificação de doenças do coração, aprovação de crédito e análise de íris. A teoria de *rough sets* obteve maior índice de acerto no exemplo de aprovação de crédito e obteve o segundo melhor índice nos outros exemplos.

A teoria de *rough sets* pode ser aplicada em diversas áreas. Na Engenharia Biomédica, por exemplo, Chaudhuri e Mitra (2006) utilizaram esta técnica para o auxílio do diagnóstico de doenças através dos parâmetros de tensão e tempo medidos em eletrocardiograma. Através da análise do formato de onda descrito nestes equipamentos, algumas doenças podem ser identificadas através de um criterioso estudo de trechos da forma de onda mostrada no gráfico de tensão (gerada pelo batimento cardíaco) versus o tempo. Para a formação da base de conhecimento foram consultados vinte médicos especialistas que ao analisar a onda, forneceram o seu diagnóstico. Com isso, conseguiu-se gerar 23 regras que podem servir de ferramentas para um diagnóstico mais preciso de doenças ligadas ao coração.

Em recursos humanos, Chen e Chien (2007) realizaram um estudo para definir quais características são relevantes para o recrutamento de funcionários em uma empresa de manufatura de semicondutores. Para o setor de tecnologia, o capital humano é muito importante e isto estimulou o estudo que aplicou *rough sets*. Foram considerados neste estudo parâmetros tais como: escolaridade, idade, sexo, tempo de experiência, entre outros, para identificar potenciais humanos altos de sucesso profissional nesse ramo.

No setor elétrico, apesar de ainda não existir grandes quantidades de

aplicações de *rough sets*, a maioria das aplicações disponíveis se concentram nos estudos de proteção e controle de sistemas de potência.

Han, *et al.* (1999) utilizou *rough sets* em conjunto com redes neurais para a classificação de faltas em sistemas de potência de alta tensão através de anos de dados arquivados. Seu trabalho buscou classificar as faltas através da comparação entre as formas de ondas e as suas respectivas causas. Através de parâmetros já estabelecidos é possível identificar sete tipos de ondas que são resultados de faltas específicas e que serviram de base de conhecimento deste trabalho. As regras foram aplicadas em 26 amostras de faltas e os resultados se mostraram melhores do que quando é aplicada apenas a lógica nebulosa pura, que também foi utilizada para comparação.

Lambert-Torres (2002) realizou um estudo aplicando *rough sets* para diminuir o número de informações a serem analisadas para a determinação do estado de operação de um sistema elétrico. Os operadores usam a experiência para determinar se um sistema está operando em estado normal, anormal ou restaurativo, através da análise de dados e medições fornecidas, porém, devido à grande quantidade de dados, essa análise torna-se complexa. Na aplicação citada, dados como fluxo de potência e tensão nas linhas de transmissão foram utilizados como atributos da base de conhecimento e a saída foi o estado de operação. Com a aplicação de *rough sets* foi possível extrair um conjunto de regras que tornaram o processo de tomada de decisão mais simples do que a metodologia anterior.

Bonaldi *et al.* (2002) aplicou *rough sets* na manutenção preditiva de motores de indução. Para isso, utilizou como base de conhecimento a análise das formas de ondas características de corrente geradas por problemas nos motores, tais como, rotor assimétrico ou desbalanceado, defeitos em rolamentos, excentricidade de *gap*, entre outros. A análise baseou-se nas características espectrais da corrente do estator do motor, pois cada tipo de defeito gera um comportamento específico desta corrente o que se reflete em seu espectro. Após a identificação de quais frequências são geradas por cada tipo de falha, foi então traçada uma curva da amplitude da mesma versus o nível de severidade do defeito e, com isso, foram determinadas as faixas a serem utilizadas nos atributos, como por exemplo, para defeito no rolamento, *Normal (N)*, *Warning (W)* e *Emergency (E)*. Após a aplicação de *rough sets* a saída fornecida foi o tipo de falha e o grau de severidade. Ao final, a

metodologia conseguiu também identificar quais atributos são indispensáveis na análise dos defeitos, contribuindo para facilitar o diagnóstico quando o mesmo é realizado por operadores de manutenção.

Chien, Peng e Tseng (2004) buscaram identificar o local e o equipamento avariado de uma falta através de um banco de dados obtido do histórico de registros da *Taiwan Power Company*, em Taiwan. Esse banco de dados serviu como base de conhecimento para aplicação de *rough sets*. Registros de hora, mês, ano, local, causa ou acidente, equipamento avariado, entre outros, de faltas que ocorreram em anos de arquivos foram utilizados e serviram para a validação da metodologia. A precisão da estimativa foi obtida para cada regra, devido ao fato de que existem vários tipos de saída na aplicação de *rough sets* para este caso, pois as faltas podem ser causadas por falhas em diversos tipos de equipamentos e isto foi levado em consideração. A maioria das regras forneceu estimativas com 100% de acerto e a pior estimativa obtida foi de 82% de acerto.

Cao, Feng e Qiu (2004) utilizaram *rough sets* em conjunto com redes neurais para realizar a previsão a curto prazo de cargas no sistema elétrico. Utilizaram o histórico de uma companhia energética do ano de 2001, com dados de temperatura, demanda etc. A teoria de *rough sets* serviu para extrair o domínio do conhecimento que serviu de base para a aplicação de redes neurais. Em seus trabalhos demonstraram que os resultados obtidos são melhores do que quando aplicada apenas a técnica de redes neurais. Este resultado está se confirmando na prática, visto que as duas técnicas estão sendo aplicadas em um sistema elétrico real, onde este fato vem se confirmando.

Cerchiari *et al.* (2006) utilizou em conjunto com *rough sets* a metodologia *Self-Organizing Maps (SOM)* para a classificação e estimação das curvas diárias de demanda de consumidores de uma companhia de energia elétrica. Esta metodologia foi usada para encontrar o conjunto de curvas que representam o espaço de possíveis curvas de demanda dos consumidores que após processos estatísticos resultam em uma única. Então, na sequência, foi aplicado *rough sets* para a classificação de cada consumidor em sua curva característica usando dados de entrada extraídos de arquivos da companhia, tais como, consumo, tipo de consumidor e número de fases. A validação do método foi feita através da comparação dos resultados obtidos com as medições realizadas em

transformadores da rede. Nesta comparação foi obtido um erro médio de 27,7% em relação aos valores reais medidos. A companhia de energia elétrica obteve uma melhoria de 47% na estimação das curvas de demanda em relação à metodologia anterior aplicada na empresa.

Coutinho (2007) realizou um trabalho para identificar anomalias no sistema de informações da rede elétrica. Tais anomalias podem ser causadas por ruídos ou invasões maliciosas no sistema e poderiam, por exemplo, acusar que um disjuntor está aberto, quando na verdade ele está fechado, mudar o sinal de um valor ou alterar a ordem de grandeza do mesmo. Isto poderia ocasionar diversos erros nas análises dos operadores, inclusive o colapso do sistema. Utilizou como base de conhecimento parâmetros tais como fluxo: de potência, corrente e tensão. A teoria *rough sets* foi aplicada para classificar o sistema em: normal ou anormal. A condição anormal implica em uma possível invasão do sistema. Para a geração das regras foram utilizados diversos exemplos sendo que parte deles foram gerados através da corrupção dos dados do sistema, por exemplo, trocando o sinal da potência ativa que passou de negativa para positiva. Ao final foram geradas as regras que foram aplicadas a novos exemplos. O algoritmo reconheceu como anormal 100% dos casos com dados corrompidos.

Crossley, Hor e Watson (2007) realizaram um trabalho utilizando a técnica para auxiliar na tomada de decisão de operadores do sistema elétrico. Muitas informações disponíveis podem não ser relevantes na análise realizada para executar determinadas operações, porém, com o uso de *rough sets*, foi possível extrair regras que levam em consideração apenas dados importantes e aumentar, com isso, a velocidade de operação do sistema, principalmente diante de situações emergenciais. A base de dados foi obtida através de simulações computacionais de um modelo de subestação onde é possível determinar uma falta em um local específico do circuito e com isso obter os valores de tensão e corrente em vários pontos do sistema. A aplicação da técnica, além de fornecer somente os parâmetros que são relevantes para análise, também serviu para poder estimar a zona de ocorrência de uma falta. Os autores obtiveram uma estimativa correta em 95% dos casos, quando utilizado 50% das amostras para base de conhecimento (treino) e 50% para testes. Quando utilizado para treino 70%, obtiveram sucesso em 100% das estimativas. O mesmo ocorreu quando foi utilizado 90% das amostras para

treinamento.

Chen e Pai (2008) utilizaram *rough sets* para o controle do fluxo de potência em sistemas elétricos, onde, na saída da aplicação da metodologia, obtinha-se os resultados incrementar, manter ou diminuir o fluxo. Porém, seu foco foi a comparação da aplicação de *rough sets* tradicional com a metodologia de *rough sets* utilizando análise discriminante para a redução dos atributos (redutos). Na sua comparação, quando foi utilizada duas saídas (atributos de decisão), incrementar ou diminuir, as duas técnicas apresentaram os mesmos resultados referentes ao grau de confiança da estimativa, porém, o tempo computacional se mostrou melhor com o uso da análise discriminante.

2.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foi descrita a teoria de *rough sets* com alguns exemplos ilustrativos e também o estado da arte desta metodologia com foco em sistemas de potência. É possível perceber que os conceitos dessa metodologia são relativamente simples, porém, a dificuldade está em aplicar este conceito em bancos de dados que possuem muitos atributos. Para isso se faz necessário o uso de ferramentas computacionais, a fim de facilitar a obtenção dos redutos e com isso ser possível a geração das regras.

3 QUALIDADE DA ENERGIA ELÉTRICA (QEE)

3.1 PARÂMETROS DE QEE

Atualmente a energia elétrica pode ser considerada um produto de extrema importância para a sociedade em geral. Isto inclui consumidores residenciais, comerciais, industriais e poder público. O avanço da tecnologia nos últimos anos inseriu em nosso cotidiano equipamentos que se transformaram em itens essenciais que vão desde refrigeradores, para o armazenamento de alimentos, até computadores que gerenciam parte dos sistemas de informação atuais e, a grande maioria destes equipamentos, necessita de energia elétrica para o seu funcionamento. Neste contexto, a QEE vem se tornando cada vez mais importante, pois é através dela que se pode garantir que as cargas do sistema operem corretamente, sem falhas que possam causar desde um funcionamento incorreto até mesmo um dano irreversível. Este problema vem se agravando nos últimos anos, principalmente, devido à inserção de cargas mais sensíveis aos distúrbios elétricos o que pode gerar acentuadas perdas em diversos setores da sociedade.

Um problema de QEE pode ser definido como qualquer alteração na tensão, corrente ou frequência de um sistema elétrico que resulte em uma falha ou mau funcionamento de um equipamento. Isto pode ocorrer em qualquer parte do sistema, ou seja, tanto nas instalações dos consumidores como no sistema de fornecimento elétrico.

O termo Qualidade é abrangente no seu conceito e não há uma definição única, quando aplicado na energia elétrica, este conceito pode ter várias interpretações. Alguns autores consideram QEE como sendo a ausência de distúrbios no sistema elétrico, outros consideram apenas a ausência de distúrbios que possam causar danos em equipamentos, cargas e ao sistema de potência. Um conceito mais abrangente, segundo Dugan *et al.* (2002) considera Qualidade da Energia Elétrica como sendo um conjunto de fatores que determinam a:

- qualidade do atendimento;

- qualidade do serviço;
- qualidade do produto.

A Qualidade do produto está relacionada com a energia elétrica propriamente dita, e os principais fenômenos eletromagnéticos ligados à Qualidade da Energia Elétrica no sistema de fornecimento podem ser agrupados em:

- variações transitórias da tensão;
- variações momentâneas de tensão;
- variações sustentadas de tensão;
- variações momentâneas de frequência;
- ruídos e interferências eletromagnéticas;
- flutuação de tensão, cintilação e desequilíbrio de tensão;
- distorção harmônica total de tensão.

A seguir será dado enfoque na Distorção Harmônica Total de Tensão, por se tratar de um distúrbio que é alvo do estudo aplicado neste trabalho.

3.2 DISTORÇÃO HARMÔNICA TOTAL DE TENSÃO (DHTV)

A distorção, em sistemas elétricos, é uma alteração na forma de onda puramente senoidal de tensão ou corrente, quando o sistema está operando em regime permanente. Existem diversos tipos de distorções que podem ser classificadas em:

- Harmônicas: são correntes ou tensões de frequências múltiplas da fundamental fornecida pela concessionária e que se somam causando deformação na forma de onda puramente senoidal. Estas frequências são provenientes de cargas não lineares instaladas no sistema e geralmente são caracterizadas por equipamentos eletrônicos.

- Inter-harmônicas: componentes de frequência, em tensão ou corrente, que não são múltiplos inteiros da frequência fundamental do sistema supridor (50 ou 60Hz). Elas podem aparecer como frequências discretas ou como uma larga faixa espectral. Os inter-harmônicos podem ser encontrados em redes de diferentes classes de tensão. As suas principais fontes são conversores estáticos de potência,

cicloconversores, motores de indução e equipamentos a arco. Sinais *carrier* em linhas de potência também podem ser considerados como inter-harmônicos.

- Nível CC: a presença de tensão ou corrente CC em um sistema elétrico CA é denominado DC offset. Este fenômeno pode ocorrer como o resultado da operação ideal de retificadores de meia-onda. O nível CC em redes de corrente alternada pode levar à saturação de transformadores, resultando em perdas adicionais e redução da vida útil.

- *Notching*: distúrbio de tensão causado pela operação normal de equipamentos de eletrônica de potência quando a corrente é comutada de uma fase para outra. Este fenômeno pode ser detectado através do conteúdo harmônico da tensão afetada. As componentes de frequência associadas com os *notchings* são de alto valor e, desta forma, não podem ser medidas pelos equipamentos normalmente utilizados para análise harmônica.

- Ruídos: é definido como um sinal elétrico indesejado, contendo uma larga faixa espectral com frequências menores que 200 kHz, as quais são superpostas às tensões ou correntes de fase, ou encontradas em condutores de neutro. Os ruídos em sistemas de potência podem ser causados por equipamentos eletrônicos de potência, circuitos de controle, equipamentos a arco, retificadores a estado sólido e fontes chaveadas e, normalmente, estão relacionados com aterramentos impróprios.

Para o caso das distorções harmônicas pode-se definir que, tecnicamente, uma harmônica é a componente de uma onda periódica cuja frequência é um múltiplo inteiro da frequência fundamental (no caso da energia elétrica, de 60 Hz). A melhor maneira de explicar isto é com a ilustração na figura 5. Nesta figura, veem-se duas curvas: uma onda senoidal normal, representando uma corrente de energia "limpa", e outra onda menor, representando uma harmônica. Esta segunda onda menor representa a harmônica de quinta ordem, o que significa que sua frequência é de 5×60 Hz, ou 300 Hz.

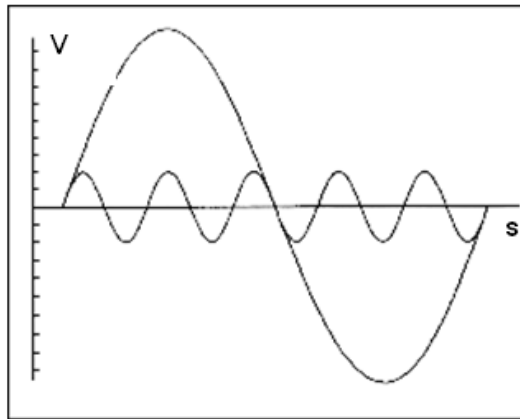


Figura 5 - Representação da onda fundamental e da componente harmônica

Na Figura 6, é possível visualizar como ficaria a soma das duas curvas. Esta curva resultante mostra bem a distorção harmônica da curva de tensão, na presença de harmônicas.

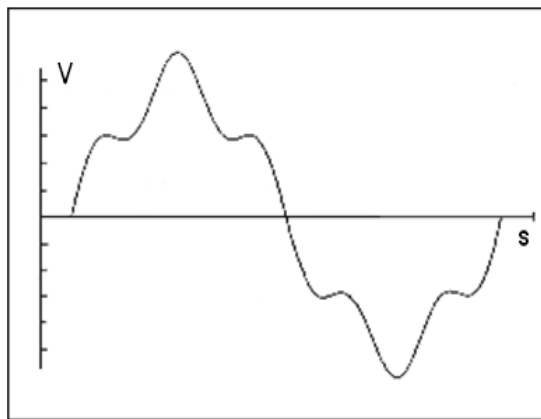


Figura 6 - Onda fundamental somada à componente harmônica

A distorção harmônica de tensão, no geral, pode ser considerada como sendo o resultado da corrente harmônica que circula através da impedância do sistema elétrico. Para cada frequência harmônica da corrente, há uma queda de tensão correspondente, resultando, desta forma, na distorção da tensão na barra. Este fenômeno está ilustrado na Figura 7.

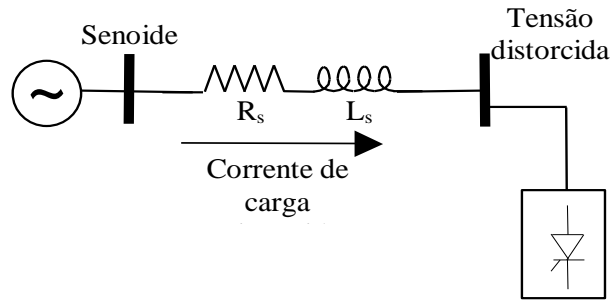


Figura 7 - Harmônica versus corrente harmônica

Fonte: (PENICHE e TEIXEIRA, 2005)

O que determina a magnitude da distorção é a impedância equivalente vista pela carga bem como o valor da corrente solicitada pela mesma. Percebe-se, portanto, que uma mesma carga pode ocasionar diferentes níveis de distorção harmônica de tensão, dependendo da sua localização no sistema elétrico, devido às diferentes impedâncias existentes em cada local.

Para se obter o valor total da distorção harmônica total de tensão deve-se utilizar o método descrito na equação 3. O mesmo conceito pode ser utilizado para a obtenção da distorção harmônica total de corrente, conforme pode ser visto na equação 4.

$$DHTV = \sqrt{\frac{\sum_{n>1}^n V_n^2}{V_1^2}} \times 100 (\%) \quad (3)$$

$$DHTI = \sqrt{\frac{\sum_{n>1}^n I_n^2}{I_1^2}} \times 100 (\%) \quad (4)$$

onde: *DHTV* - distorção harmônica total de tensão;

DHTI - distorção harmônica total de corrente;

V_n - valor eficaz da tensão de ordem n ;

I_n - valor eficaz da corrente de ordem n ;

V_1 - valor eficaz da tensão fundamental;

I_1 - valor eficaz da corrente fundamental;

n - ordem da componente harmônica.

As distorções harmônicas diminuem a qualidade do suprimento promovido por uma concessionária de energia elétrica, a qual deve fornecer aos seus consumidores uma tensão puramente senoidal, com amplitude e frequência

constantes. Entretanto, o fornecimento de energia a determinados consumidores que causam deformações no sistema supridor, prejudica não apenas o consumidor responsável pelo distúrbio, mas também, outros conectados à mesma rede elétrica.

Há alguns anos atrás não havia maiores preocupações com harmônicos, pois cargas com características não lineares eram pouco utilizadas e os equipamentos eram mais resistentes aos efeitos provocados por harmônicas. Contudo, nos últimos anos, com o rápido desenvolvimento da eletrônica de potência, e a utilização de métodos que buscam o uso mais racional da energia elétrica, o conteúdo harmônico presente nos sistemas tem-se elevado, causando uma série de efeitos indesejáveis em diversos equipamentos ligados à rede elétrica.

A seguir, são citados alguns equipamentos que têm contribuído, nos últimos anos, para a distorção harmônica total de tensão e que estão presentes no sistema elétrico brasileiro:

- circuitos de iluminação com lâmpadas de descarga e reatores eletrônicos;
- fornos a arco;
- compensadores estáticos tipo reator saturado;
- motores de corrente contínua controlados por retificadores;
- motores de indução controlados por inversores com comutação forçada;
- processos de eletrólise através de retificadores não-controlados;
- motores síncronos controlados por cicloconversores;
- fornos de indução de alta frequência;
- fornos de indução controlados por reatores saturados;
- cargas de aquecimento controladas por tiristores;
- motores CA controlados por tensão de estator;
- reguladores de tensão a núcleo saturado;
- computadores;
- eletrodomésticos com fontes chaveadas; etc.

Os principais prejuízos que a DHTV causa na indústria são a perda de produtividade devido às paradas de produção causadas por inesperadas falhas em motores, acionamentos, fontes; etc. As falhas mais comuns encontradas nos

equipamentos que causam estes prejuízos são:

- capacitores - queima de fusíveis e redução da vida útil;
- motores - redução da vida útil e impossibilidade de atingir potência máxima;
- fusíveis/disjuntores - operação falsa/errônea, e componentes danificados;
- transformadores - aumento de perdas, causando a redução de capacidade e a diminuição da vida útil;
- medidores - possibilidade de medições errôneas;
- telefones - interferências;
- máquinas síncronas - sobreaquecimento das sapatas polares, causado pela circulação de correntes harmônicas nos enrolamentos amortecedores;
- acionamentos/fontes - operações errôneas devido a múltiplas passagens por zero, e falha na comutação de circuitos;
- carregamento exagerado do circuito de neutro, principalmente em instalações que agregam aparelhos eletrônicos e possuem malhas de terra mal projetadas.

Entre os equipamentos citados, os principais problemas causados por harmônicos se dão junto a bancos de capacitores, que podem originar condições de ressonância, caracterizando uma sobretensão nos terminais das unidades capacitivas.

Em decorrência desta sobretensão, tem-se uma degradação do isolamento das unidades capacitivas, e em casos extremos, uma completa danificação dos capacitores. Além disso, consumidores conectados no mesmo ponto ficam submetidos a tensões perigosas, mesmo não sendo portadores de cargas poluidoras em sua instalação. Mesmo sem uma condição de ressonância, um capacitor é sempre um caminho de baixa impedância para as correntes harmônicas, e sempre estará sujeito a sobrecarga e sobreaquecimento excessivo.

Pensando em todos estes problemas causados pela DHTV, o Operador Nacional do Sistema Elétrico (ONS), Submódulo 2.2 de “Padrões de Desempenho da Rede Básica” recomenda alguns parâmetros a serem seguidos pelos órgãos

ligados ao sistema elétrico, entre estes parâmetros existe uma recomendação para a distorção harmônica total de tensão a ser seguida conforme a Tabela 6.

Tabela 6 – Limites globais de distorção harmônica de tensão

V < 69 kV			
Ímpares		Pares	
Ordem	Valor (%)	Ordem	Valor (%)
3, 5 e 7	5	2, 4 e 6	2
9, 11 e 13	3	≥ 8	1
15 a 25	2	-	-
≥ 27	1	-	-

*DHTV = 6%

A Agência Nacional de Energia Elétrica (ANEEL) através dos “Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional” – PRODIST - Módulo 8 – Qualidade da Energia Elétrica – determina que para redes de baixa tensão, ou seja, tensão inferior a 1000 V, a DHTV deve ser inferior a 10%. Na aplicação deste trabalho foi considerado o valor de 6% do ONS, pois, além do banco de dados utilizado neste trabalho ter sido obtido anteriormente à publicação deste procedimento, outras técnicas aplicadas neste mesmo banco de dados e disponíveis na literatura utilizaram o valor de 6% para classificação. Sendo assim, a manutenção do limite estabelecido pelo ONS, possibilita a realização de uma comparação final dos resultados das diferentes técnicas de classificação aplicadas a uma mesma base de conhecimento.

3.3 DESEQUILÍBRIO DE TENSÃO

Outra recomendação do ONS se refere ao desequilíbrio de tensão. O indicador para avaliar o desequilíbrio de tensão nos barramentos da rede básica é o fator de desequilíbrio de tensão, que exprime a relação entre as componentes de sequência negativa e sequência positiva da tensão, expresso em termos percentuais da componente de sequência positiva. O fator de desequilíbrio de ser menor ou igual

a 2%.

$$K = \frac{V_2}{V_1} \times 100 \quad (5)$$

onde: K – Fator de desequilíbrio de tensão,

V_1 - Componente de sequência positiva,

V_2 - Componentes de sequência negativa.

3.4 PERFIL DE TENSÃO (LIMITES DE VARIAÇÃO DE TENSÃO)

Em relação aos limites de variação de tensão, na Tabela 7 estão descritos os valores conforme a resolução da ANEEL nº 505 de 26 de novembro de 2001, revisada em 28 de janeiro de 2005.

Tabela 7 - Pontos de entrega em tensão nominal igual ou inferior a 1kV

TENSÕES NOMINAIS PADRONIZADAS				
Tensão Nominal (TN)		Faixa de Valores Adequados das Tensões de Leitura (TL) em relação à TN (Volts)	Faixa de Valores Precários das Tensões de Leitura (TL) em relação à TN (Volts)	Faixa de Valores Críticos das Tensões de Leitura (TL) em relação à TN (Volts)
Ligação	Volts			
Trifásico	(220)	(201≤TL≤231)	(189≤TL<201 ou 231<TL≤233)	(TL<189 ou TL>233)
	(127)	(116≤TL≤133)	(109≤TL<116 ou 133<TL≤140)	(TL<109 ou TL>140)
Monofásico	(127)	(116≤TL≤132)	(109≤TL<116 ou 132<TL≤140)	(TL<109 ou TL>140)

3.5 ESTADO-DA-ARTE SOBRE CLASSIFICAÇÃO DE PARÂMETROS DE QUALIDADE DE ENERGIA ELÉTRICA

Diversas técnicas são aplicadas para a classificação de distúrbios de energia elétrica, entre elas, encontra-se principalmente a aplicação de redes neurais artificiais.

Davidson, Fouad e Zhou (1994) buscaram identificar o estado de segurança e de estabilidade de sistemas de potência utilizando redes neurais. O objetivo foi determinar o grau de vulnerabilidade do sistema para o limite de estabilidade de forma dinâmica. Os dados de entrada foram as margens de tensão e ângulos dos geradores e a saída foi o estado do sistema: 1 - vulnerável ou 0 - não vulnerável. O algoritmo de retro propagação de erro foi utilizado para treinar a rede. A aplicação foi realizada utilizando o modelo de teste do *Institute of Electrical and Electronic Engineers* (IEEE). Foram simuladas faltas em nove locais diferentes do modelo e a metodologia obteve sucesso na classificação em 100% dos casos.

Dalstein e Kulicke (1995) em seus trabalhos utilizaram redes neurais para a classificação de faltas no sistema. O objetivo foi identificar, de forma rápida, as faltas como fase – terra, fase - fase – terra, e trifásico – terra. Para treinar a rede neural utilizaram a metodologia *Back Propagation* em mais de 45.000 padrões relevantes através de simulações de faltas derivadas de um sistema de 380 kV e validaram o método em 240 situações de faltas, onde foi identificado corretamente 100 % delas em um tempo médio de 6 ms. Foi utilizada a técnica, também, para identificar se a falta está relacionada com arco elétrico ou não. Isto porque o arco elétrico produz um ruído que possui uma forma de onda retangular em conjunto com a forma de onda da corrente de curto, portanto, foi possível classificar este fenômeno. A técnica se mostrou eficaz, pois realizou a identificação mais rapidamente do que as metodologias convencionais utilizadas na época.

Outra aplicação de redes neurais artificiais utilizada em QEE foi realizada por Ghosh e Lubkeman (1995). Os registros dos eventos nos seus campos de estudo ocorriam sem a discriminação do tipo específico de distúrbio que originou a falta e em seus trabalhos usaram redes neurais artificiais para relacionar as formas de ondas com as suas causas, quando ocorrem distúrbios no sistema elétrico. As

formas de ondas padrões utilizadas para a classificação e base de conhecimento foram aquelas disponíveis na literatura. Foram realizadas simulações no sistema utilizando algumas manobras, tais como: chaveamento de banco de capacitores, manobra de cargas e alguns eventos, tais como, faltas em diferentes locais de ocorrência. Para o treinamento da rede foram utilizadas 50 amostras e para teste foram utilizadas 10 amostras. Foi obtido sucesso nas classificações em mais de 90% das amostras testadas. Outro trabalho realizado por Dash, *et al.* (2000) buscou o mesmo objetivo com sucesso, porém, utilizou uma combinação de sistemas nebulosos com combinação linear de Fourier. Esta última serviu para normalizar a amplitude dos picos de tensão e corrente, para cada instante de amostragem. Após isto, os valores foram inseridos no módulo de diagnóstico, no qual é utilizado o sistema nebuloso, onde, então, foi feita a classificação das ondas conforme os eventos que a originaram, como surtos, afundamentos de tensão, sobretensão e desligamentos.

Keerthipala e Wang (1998) utilizaram, em conjunto com redes neurais, a técnica de sistemas nebulosos para a classificação de faltas em tempo real. Utilizaram, para isso, a análise de componentes simétricas em combinação com as correntes trifásicas para a classificação dos curtos-circuitos em: fase-terra, fase-fase, fase-fase-terra e trifásico-terra, além da identificação da fase. Com isso, conseguiram identificar as ocorrências das faltas corretamente em menos de 10 ms.

Chow e Xu (2006) realizaram uma comparação dos resultados de duas metodologias aplicadas para a classificação de faltas: Regressão Logística e Redes Neurais. A base de dados utilizada em seus estudos foi extraída de um banco de dados do sistema elétrico da empresa *Duke Energy*. Cada falta ocorrida produziu um arquivo com 33 informações a respeito do evento. Para a aplicação das metodologias foram utilizados aqueles dados considerados mais importantes neste contexto por engenheiros com anos de experiência na área, são elas: número do circuito, condições climáticas, estação do ano, hora do dia, número de fases afetadas e equipamento de proteção ativado. A melhor estimativa obtida foi de 93% para testes realizados em situações com parâmetros reais. É levantada a hipótese do banco de dados possuir muitos ruídos, tais como, dados contraditórios, contudo, conclui-se que, para este caso, a metodologia de Redes Neurais se mostrou mais vantajosa por ser mais flexível na sua aplicação.

Choong *et al.* (2007) utilizaram uma combinação de transformada discreta de ondas, redes neurais e sistema nebuloso para a classificação de distúrbios elétricos conforme sua causa. A arquitetura da metodologia consiste no pré-processamento dos dados através da transformada discreta, após este processamento, os dados são processados com redes neurais e então é aplicado o sistema nebuloso. Foi Aplicada, também, somente a transformada discreta com redes neurais e foi feita a comparação dos resultados. No primeiro caso foi obtido sucesso na estimativa de 98,19% dos casos testados, que no total são 10.200 amostras. No segundo caso foi obtido sucesso na classificação de 96,17% dos itens testados.

Souza (2008) utilizou o mesmo banco de dados utilizado neste trabalho para a classificação de transformadores segundo a sua DHTV. O objetivo final de seu trabalho foi o mesmo, porém, foram utilizadas técnicas de análise multivariada, mais especificamente regressão logística e escore quadrático. Os resultados obtidos por ele estão descritos mais adiante, no capítulo 5 deste trabalho, onde são feitas comparações, visto que o banco de dados utilizado, bem como o objetivo principal, são os mesmos da aplicação de rough sets utilizada por essa dissertação de mestrado.

3.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Nesse capítulo foram revisados alguns conceitos de Qualidade de Energia Elétrica, com foco na Distorção Harmônica Total de Tensão. Nota-se que a QEE recebeu maior importância nos últimos anos devido à inserção de cargas não-lineares que são prejudiciais à QEE. O estado-da-arte sobre classificação de parâmetros de QEE também está exposta nesse capítulo.

4 MATERIAIS E MÉTODO

4.1 MATERIAIS

4.1.1 Campanha de Medições

No desenvolvimento deste trabalho foi utilizado um banco de dados obtido através de uma campanha de medições realizada na rede de distribuição da Companhia Paranaense de Energia (Copel). As medições foram realizadas em transformadores situados no estado do Paraná e duraram cerca de 14 meses.

4.1.2 Área de Atuação

A Copel atende cerca de 100% das residências urbanas no Paraná e 90 % das rurais, possuindo no total cerca de 3,39 milhões de consumidores ligados à sua rede (Copel, 2007).

Do total de consumidores, 2,6 milhões são residências, 57 mil indústrias, 283 mil estabelecimentos comerciais e 332 mil propriedades rurais e a cada ano, em média, são realizadas 70 mil novas ligações.

As Tabelas 8 e 9 descrevem algumas características adicionais referentes à Copel:

Tabela 8 – Características da Copel para a atuação no Paraná

Indicadores	Valores
Áreas de concessão (km ²)	194854
Municípios atendidos	393
Localidades atendidas	1111
População atendida PR (Mil Habitantes)	9822
Taxa de atendimento urbano (%)	99
Consumo médio residencial (kWh/mês)	159,6

Fonte: (COPEL, 2007)

Tabela 9 – Características da Copel – dados da distribuição

Dados de Distribuição	Valores
Número de subestações 34,5 kV	238
Potência instalada (MVA)	1576
Redes de distribuição (km)	167535
Total de postes	2.296301
Número de transformadores de distribuição	315587
Potência instalada em transformadores (MVA)	8009

Fonte: (COPEL, 2007)

A estrutura da Copel no estado do Paraná está dividida em cinco superintendências sendo que cada uma delas atua em uma região do estado e são assim denominadas:

- Superintendência de Distribuição Leste – SDL, com sede em Curitiba;
- Superintendência de Distribuição Oeste – SDO, com sede em Cascavel;
- Superintendência de Distribuição Centro-Sul – SDC, com sede em Ponta Grossa;
- Superintendência de Distribuição Noroeste – SDN, com sede em Maringá;

- Superintendência de Distribuição Norte – SDT, com sede em Londrina.

A região que possui o maior consumo no estado é a região leste (SDL-35,6%), seguida da região centro sul (SDC-18,6%), noroeste (SDN-17,3%), norte (SDT-14,8%) e região oeste (SDO-13,7%).

4.1.3 Medições

Devido ao fato de que existem muitos transformadores na rede de distribuição da Copel no estado do Paraná, a medição de todos eles se torna uma prática inviável. Para solucionar este problema buscou-se obter uma amostragem que preservasse as informações do universo estudado, uma das ações seguidas foi a tomada de amostra aleatória dos transformadores que seriam medidos.

Outra ação foi a obtenção do número de amostras que preservariam as informações do universo estudado, para isso, foi consultada a portaria 163/93 do Departamento Nacional de Águas e Energia Elétrica (DNAEE) e obteve-se uma fórmula específica para este fim, conforme equação 6, tal que:

$$C_a = \frac{N.S.\bar{S}}{\frac{(N-1).E^2}{4} + (S.\bar{S})} \quad (6)$$

onde C_a - tamanho da amostra;

N - tamanho do universo considerado = 315587;

S - proporção da população com características consideradas semelhantes, considerado como 0,5 (pior caso);

\bar{S} - proporção da população com características não semelhantes àquelas de S , considerado como 0,5 (1- S);

E - erro amostral = 5%.

Portanto, aplicando-se a equação 6, de um universo total de 315.587 amostras, chegou-se a um tamanho de amostra de 399 transformadores a serem monitorados na campanha de medição. Contudo, prevendo-se as dificuldades que poderiam ser encontradas durante as medições, tais como deslocamento, impossibilidade de medição, qualidade das medidas a serem registradas, perda de

dados, entre outras foi estipulada uma margem de segurança de 10%, com isso, o número definido de amostras a serem medidas foi de 440 transformadores.

Tabela 10 – Distribuição dos pontos de monitoração

Regional	Localização	Região	Amostras (%)	Amostras
SDC	Centro	Ponta Grossa	11%	49
SDL	Leste	Curitiba	33%	146
SDN	Noroeste	Maringá	18%	77
SDO	Oeste	Cascavel	19%	84
SDT	Norte	Londrina	19%	84
Total			100%	440

Fonte: (LACTEC, 2007)

Tabela 11 – Distribuição das amostras entre áreas urbanas e rurais

Regional	Amostras Urbanas	Amostras Rurais
SDC	38	11
SDL	136	10
SDN	68	9
SDO	69	15
SDT	76	8
Total	387	53

Fonte: (LACTEC, 2007)

Na Tabela 10 é possível visualizar a distribuição dos pontos entre as cinco superintendências e na Tabela 11 a distribuição dos pontos entre transformadores rurais e urbanos.

O tempo utilizado para a realização das medições foi de 14 meses, no período de agosto de 2005 a julho de 2007. Foram utilizados 12 aparelhos analisadores de energia, registrador MARH 21, descrito a seguir.

- Fabricante: RMS Sistemas Eletrônicos – Brasil;
- 128 amostras por ciclo;

- Período de integralização de 10 minutos;
- Memória de 4MB;
- Tensão alimentação: CA 70 a 600 V, frequência: 50Hz ou 60Hz;
- CC 11 a 600 V;
- Exatidão: classe 0,5 para tensões e correntes e classe 1 para potências.

As medições foram realizadas com os medidores instalados por um período de oito dias em média para cada transformador, porém, para a análise e composição do banco de dados utilizado neste trabalho foram considerados os dados obtidos em um período de sete dias, devido ao disposto no Submódulo 2.2 do ONS e da Resolução n. 505 da ANEEL, que se referem aos intervalos para análise. Na Figura 8, a seguir, está ilustrado como exemplo o período de monitoração e o período de análise utilizado em uma amostra (transformador).

Os valores obtidos nas medições de DHTV foram dispostos conforme as figuras 9 e 10, onde é possível, no eixo vertical, obter o valor da distorção em relação à fundamental, em cada fase. Estas medições foram realizadas no barramento de 127/220V.

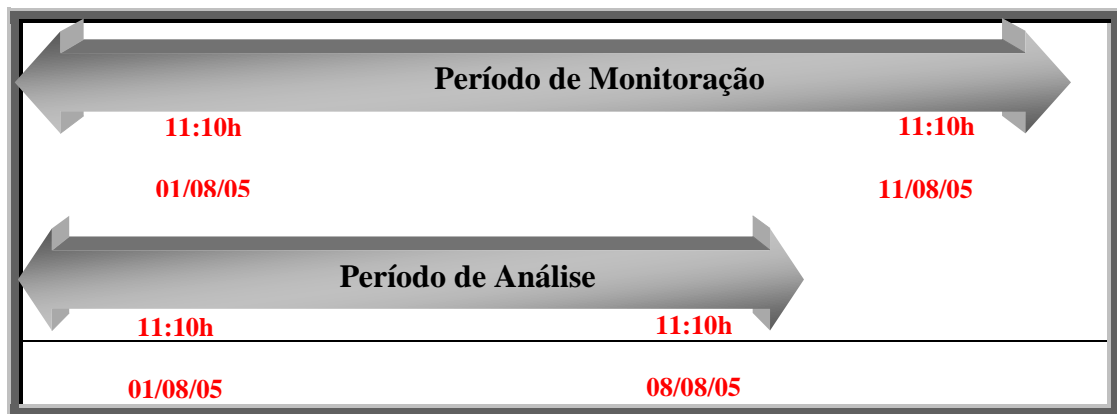


Figura 8 – Período de monitoração e período de análise

Fonte: (LACTEC, 2007)

Nas figuras 9 e 10 tem-se que:

- DHTVa - níveis de distorção harmônica total de tensão na fase A;
- DHTVb - níveis de distorção harmônica total de tensão na fase B;
- DHTVc - níveis de distorção harmônica total de tensão na fase C.

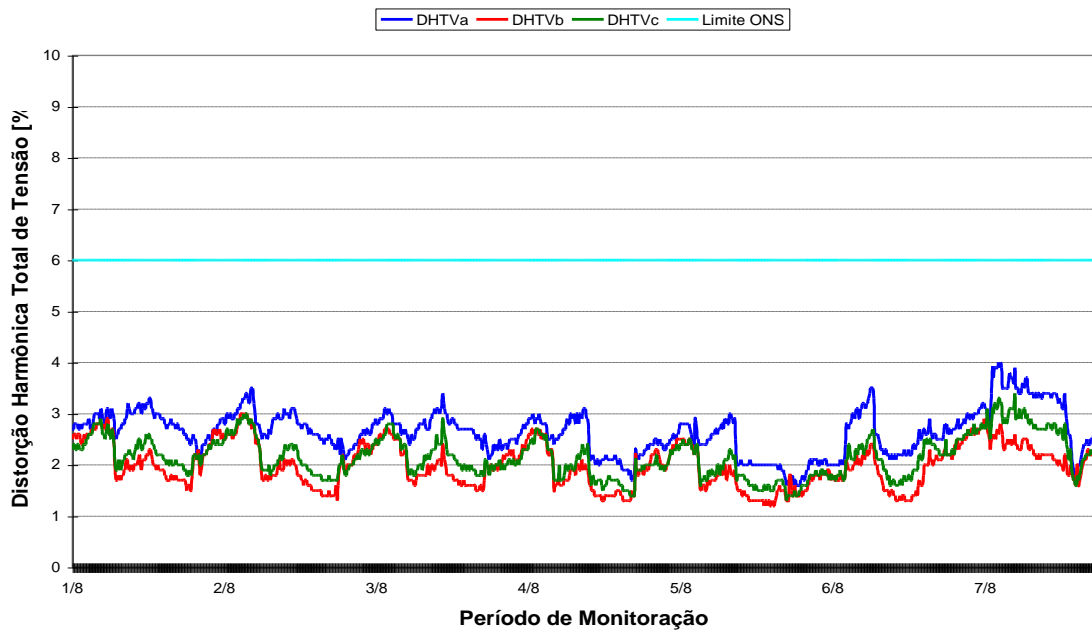


Figura 9 – Exemplo de DHTV – abaixo de 6%
 Fonte: (LACTEC, 2007)

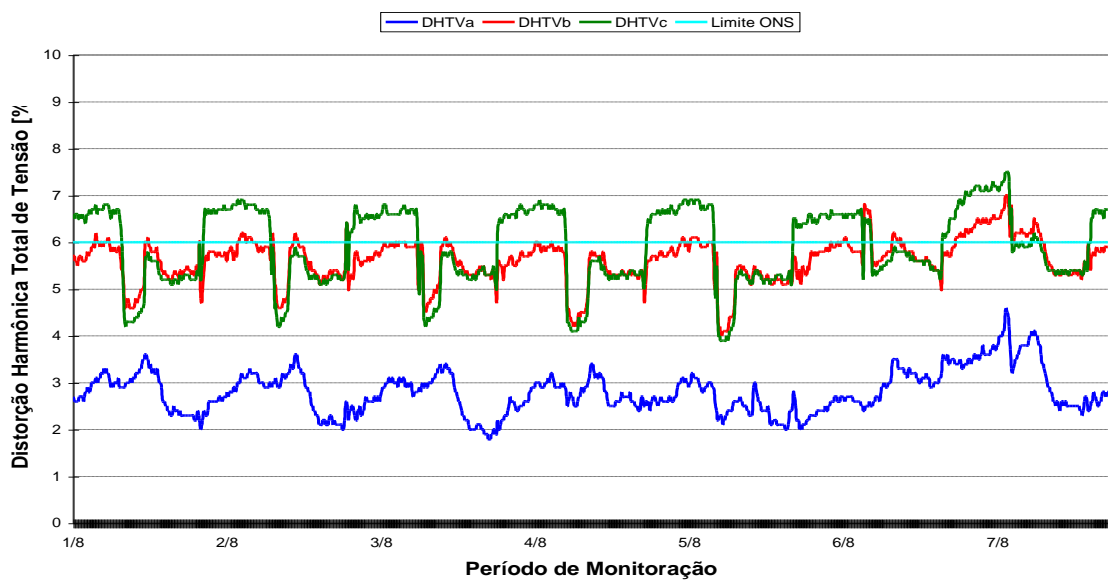


Figura 10 – Exemplo de DHTV – acima de 6%
 Fonte: (LACTEC, 2007)

Algumas informações importantes podem ser extraídas das medições da distorção Harmônica Total de Tensão, tais como valores máximos e mínimos obtidos durante o monitoramento. As Tabelas 12 e 13 a seguir demonstram exemplos

retirados destas informações extraídas das Figuras 9 e 10.

Tabela 12 – Valores máximos, mínimos e P_{95%} da DHTV - dentro do limite (6%)

DHV_T	DHTV_A [%]	DHTV_B [%]	DHTV_C [%]
Máximo	4,70	4,90	4,60
Mínimo	2,30	2,20	2,40
P_{95%}	4,53	4,66	4,35

Fonte: (LACTEC, 2007)

Tabela 13 – Valores máximos, mínimos e P_{95%} da DHTV - fora do limite (6%)

DHV_T	DHTV_A [%]	DHTV_B [%]	DHTV_C [%]
Máximo	4,60	7,00	7,50
Mínimo	1,80	4,00	3,90
P_{95%}	4,09	6,70	7,30

Fonte: (LACTEC, 2007)

Durante o monitoramento foi possível obter também o espectro harmônico, ou seja, o valor em porcentagem de cada frequência harmônica de tensão. Nas figuras 11 e 12 são ilustradas estas medições para o período de maior distorção conforme exemplos citados nas figuras 9 e 10 respectivamente.

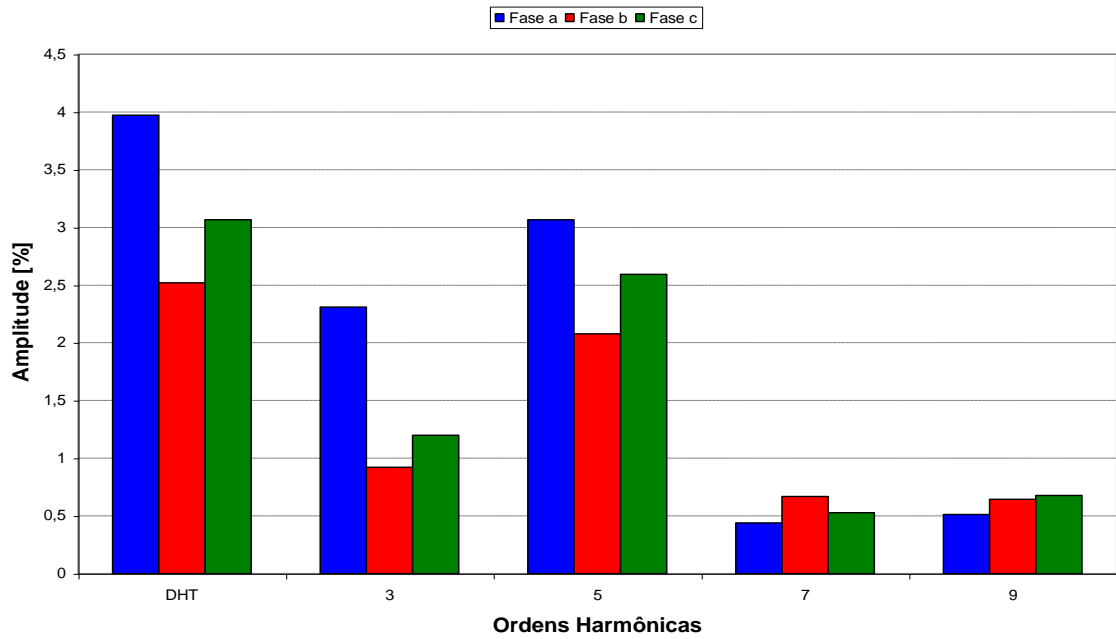


Figura 11 – Espectro da tensão no período de maior distorção – abaixo de 6%

Fonte: (LACTEC, 2007)

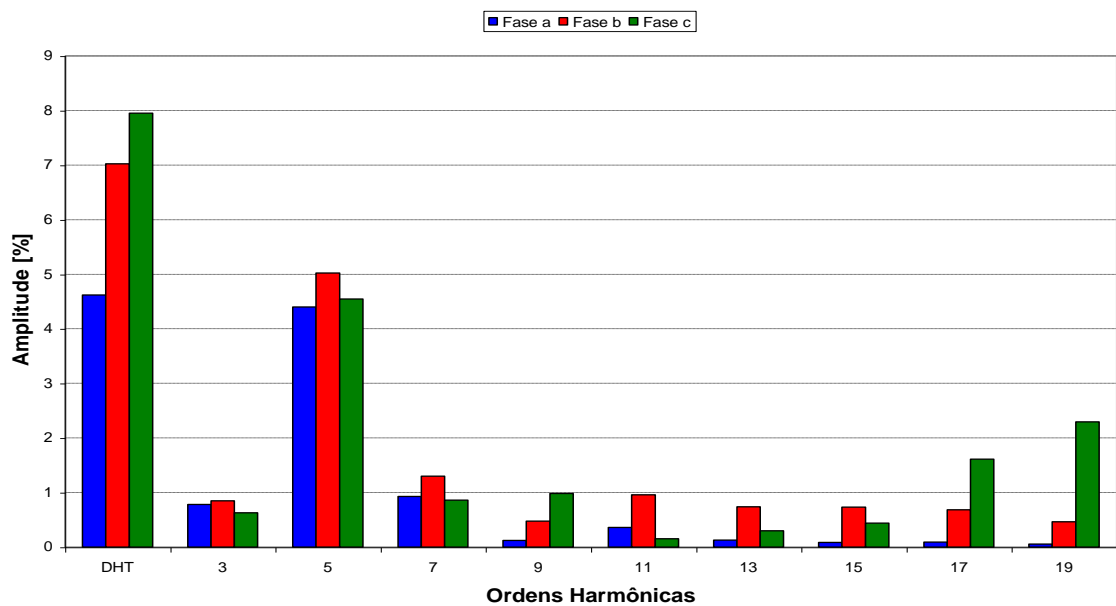


Figura 12 – Espectro da tensão no período de maior distorção – acima de 6%

Fonte: (LACTEC, 2007)

Durante o monitoramento dos transformadores, foram encontradas diversas dificuldades já previstas, tais como: dificuldade de deslocamento, impossibilidade de

medição devido à alguma impossibilidade física, etc. Com isso, ao final da campanha de medições, foram medidos no total 359 transformadores urbanos distribuídos em 186 cidades do estado do Paraná, conforme exposto na Tabela 14.

Tabela 14 – Número de transformadores urbanos medidos

Superintendência	SDL	SDT	SDO	SDC	SDN	TOTAL
Número de cidades visitadas	29	42	55	23	37	186
Quantidade de transformadores	128	74	62	32	63	359

Após a obtenção das medições, os dados foram gerenciados e tratados de forma a obter a formatação necessária para a aplicação de *rough sets*. O *software* utilizado nesta etapa foi o *Microsoft Excel*. Foram gerados cinco bancos de dados distintos, um para cada regional. Então, aplicou-se nesses bancos de dados o método apresentado no capítulo anterior.

4.1.4 Dados Obtidos

As medições realizadas em campo possibilitaram a obtenção de algumas estatísticas, tais como: média, variância, desvio padrão, máximo, mínimo e P_{95%}. Um resumo dos conceitos ligados a estes termos está disposto a seguir.

- Média aritmética simples:

A média aritmética amostral é a mais utilizada no nosso dia a dia. É obtida dividindo-se a soma das observações pelo número delas. Sua função é estimar a verdadeira média da população que originou a amostra, ou seja, o parâmetro estimado é $\mu = E(X)$.

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (7)$$

- Variância:

Define-se como variância amostral, a média dos quadrados dos desvios dos valores em relação à média, isto é:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (8)$$

A função da variância amostral é estimar a verdadeira variância populacional, ou seja, $V(X) = E[(X - \mu)^2]$, onde $\mu = E(X)$.

Uma vez que a variância envolve a soma de desvios quadráticos, a unidade em que se exprime não é a mesma que a dos dados. Assim, para obter uma medida da variabilidade ou dispersão com as mesmas unidades que os dados, tomamos a raiz quadrada da variância e obtém-se o desvio padrão:

O desvio padrão é uma medida que só pode assumir valores não negativos e quanto maior for maior será a dispersão dos dados.

- Desvio Padrão:

Define-se como desvio padrão amostral, a raiz quadrada da variância amostral.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (9)$$

A função do desvio padrão amostral é estimar o verdadeiro desvio padrão populacional σ .

- Percentil 95:

Essa medida de posição expressa que, em 95% do tempo monitorado, as grandezas monitoradas apresentam níveis iguais ou abaixo dos valores preenchidos na coluna $P_{95\%}$, ou seja, foram superadas em apenas 5% dos registros obtidos no período de 1 semana.

Ressalta-se que o valor de cada separatriz ($P_{95\%}$) a ser comparada com o valor padrão é obtido da seguinte forma:

- determina-se o valor que foi superado em apenas 5% dos registros obtidos no período de 1 dia (24 horas), ao longo de 7 (sete) dias consecutivos,
- o valor do indicador corresponde ao maior dentre os sete valores obtidos, anteriormente, em base diária.

- Máximo e Mínimo:

Máximo é o valor mais elevado encontrado em um conjunto de valores analisado e o Mínimo é o menor valor encontrado.

A seguir, na tabela 15 estão descritos os indicadores estatísticos obtidos através das medições, considerando $P_{95\%}$, após análise do banco de dados, para cada regional. São apresentados os valores considerando o valor máximo entre as três fases.

Tabela 15 – Dados estatísticos considerando o máximo entre as três fases

Regional	SDL	SDT	SDC	SDN	SDO	Rurais
Número de Elementos	129	74	36	63	63	45
Média	4,8876	4,90068	5,47083	5,00952	4,8973	6,31378
Variância	4,71038	3,70799	11,8868	10,4524	8,27522	7,96921
Desvio Padrão	2,17034	1,92561	3,44772	3,23301	2,87667	2,82298
Mínimo	2,1	1,8	2,6	1,96	1,99	2,7
Máximo	15,89	10,19	16,78	25,18	22,95	15,89

Observando as médias amostrais de cada uma das regionais, conforme Tabela 16, verifica-se que os valores são muito parecidos. Já o desvio padrão de cada regional está descrito na tabela 17:

Tabela 16 – Médias da DHTV de cada regional

Regional	SDL	SDT	SDC	SDN	SDO
Média	4,8876	4,90068	5,47083	5,00952	4,8973

Tabela 17 – Variâncias da DHTV de cada regional

Regional	SDL	SDT	SDC	SDN	SDO
Desvio Padrão	2,17034	1,92561	3,44772	3,23301	2,87667

As figuras 13 e 14 ilustram as semelhanças estatísticas entre as superintendências.

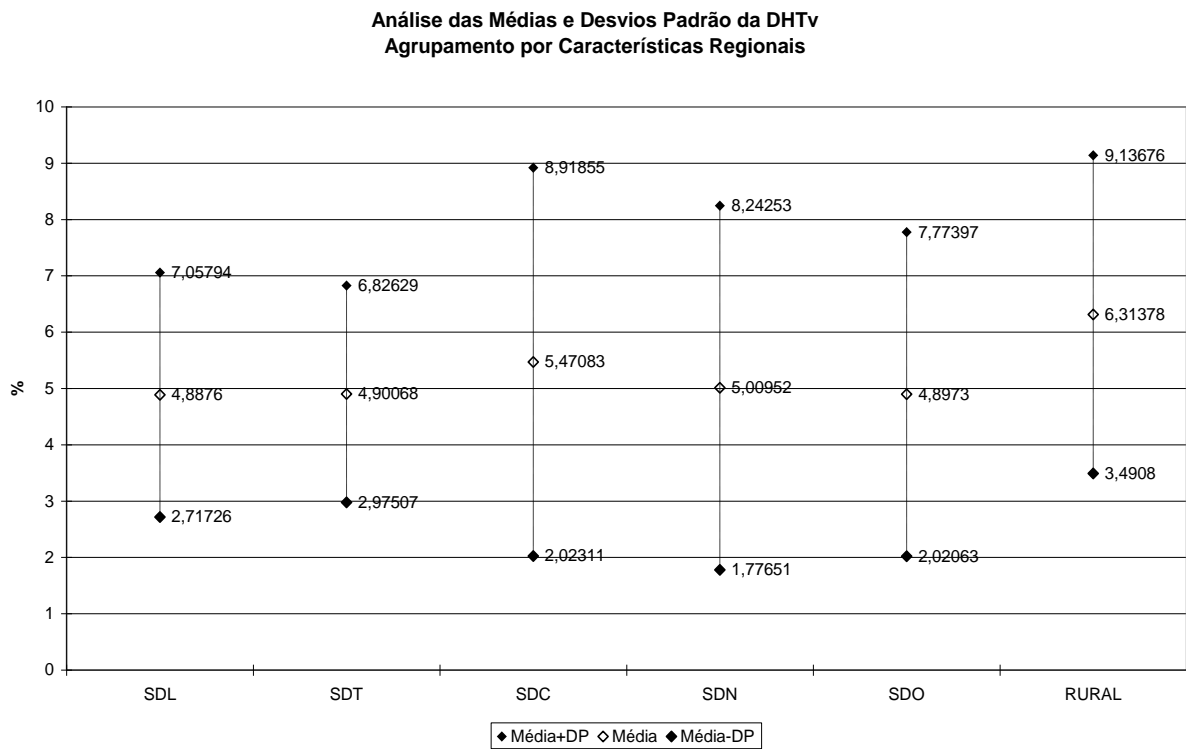


Figura 13 – Análise das médias e desvios padrão da DHTV com agrupamento por características regionais

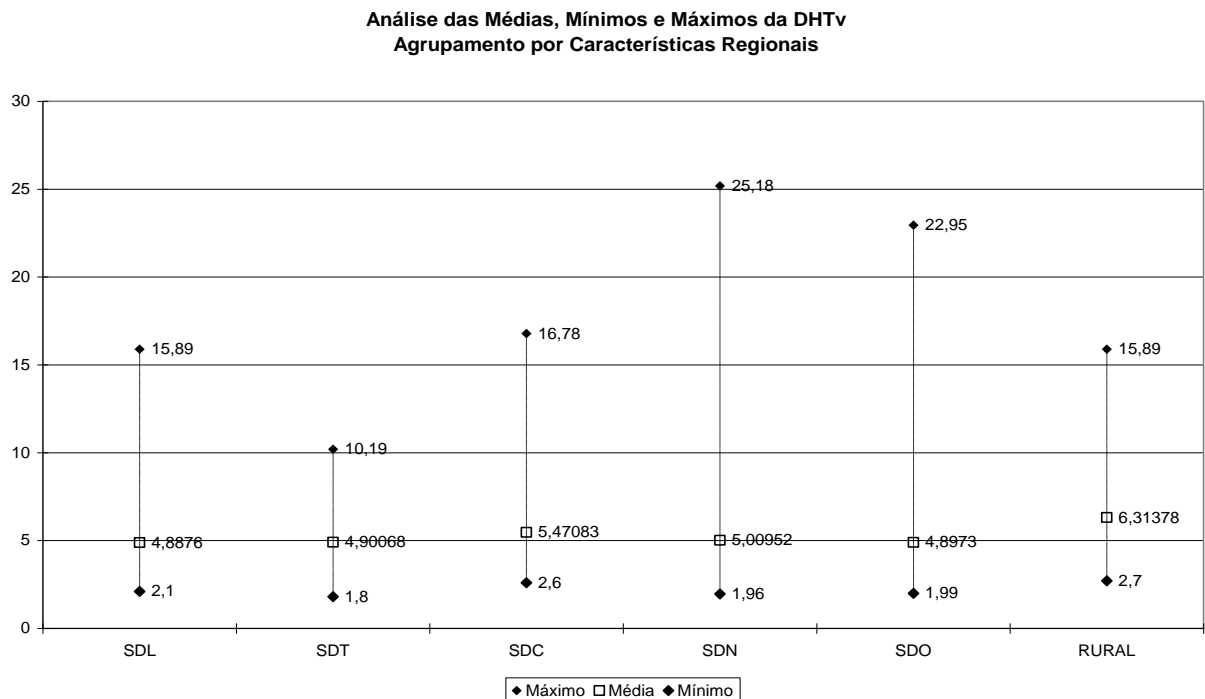


Figura 14 – Análise das médias, mínimos e máximos da DHTV com agrupamento por características regionais

A variância elevada indica que não há homogeneidade nas medições, ou seja, os transformadores medidos fora dos grandes centros apresentam comportamento diferenciado ao dos grandes centros, ou seja, o sistema de distribuição não apresenta o mesmo desempenho em todas as localidades da superintendência.

Destaca-se ainda que o pior resultado é da SDC, com a pior média a pior variância, denotando a fragilidade do sistema dessa superintendência quanto a problemas de distorção harmônica total de tensão.

Com relação aos limites estabelecidos de tensão e DHTV, tem-se que em um total de 366 transformadores (urbanos e rurais) medidos:

- 56 transformadores, ou seja, 15,30% violaram o limite de tensão máximo adequado;
- 10 transformadores, ou seja, 2,73% violaram o limite de tensão mínimo adequado;
- 83 transformadores, ou seja, 22,68% violaram o limite máximo da DHTV;

- 6 transformadores, ou seja, 1,64% violaram o limite máximo do desequilíbrio de tensão.

4.1.5 Software

Ohrn (1999) desenvolveu o *software Rosetta* para a aplicação de *rough sets* em seu trabalho na área da Medicina. Este *software* se mostrou muito eficiente e foi utilizado neste estudo para a classificação dos transformadores. A seguir estão descritos os comandos básicos utilizados para a execução do *software*:

- Após inicializar o *software*, importa-se os dados, já formatados, da planilha do *software Excel*, conforme a seguir:
- Ao clicar no botão *structures* abre-se um menu onde se deve selecionar o item ODBC, então, na janela que se abre, seleciona-se o item *open database*.
- Na próxima janela seleciona-se o item *nova* e então seleciona-se o item *driver do Microsoft Excel*, apresentado na figura 15.
- Após isso, seleciona-se “avançar” para selecionar o arquivo em formato *XLS (Excel)* onde se encontram os dados para análise.

Conforme mencionado anteriormente, foram realizados testes de classificação nas cinco regionais separadamente, portanto, foram inseridas cinco planilhas de dados no *software*.

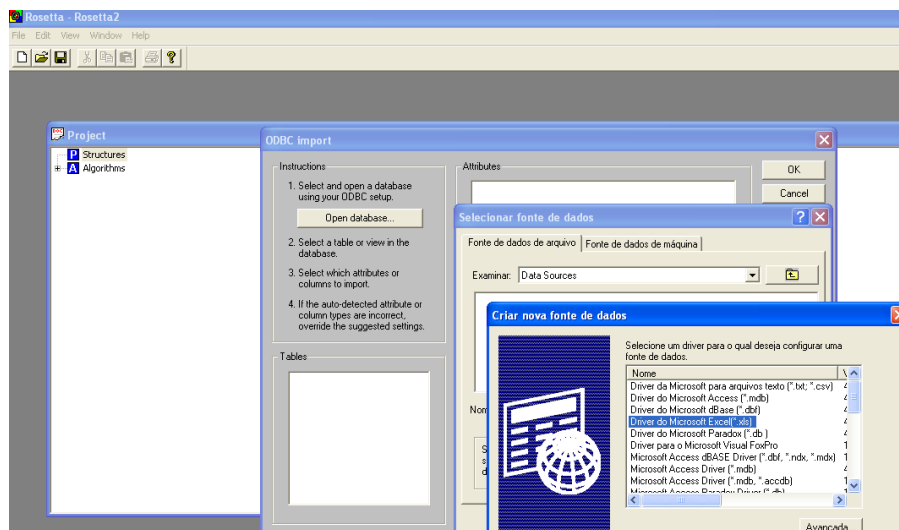


Figura 15 - Importação do banco de dados

Para a geração dos redutos, deve-se selecionar o banco de dados a ser utilizado e então: *reduce>genetic algorithm*, conforme apresentado na Figura 16. Todas as outras metodologias de redução disponíveis foram testadas, porém, para esta aplicação, a que obteve melhores resultados foi a de algoritmos genéticos (*genetic algorithm*).

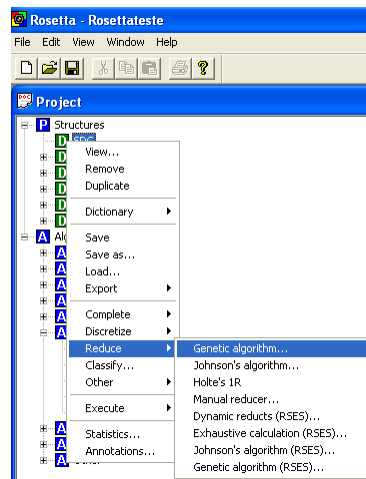


Figura 16 - Geração dos redutos

O passo seguinte consiste na geração de regras. Para isso, seleciona-se o reduto gerado e então *Generate rules*, conforme apresentado na Figura 17.

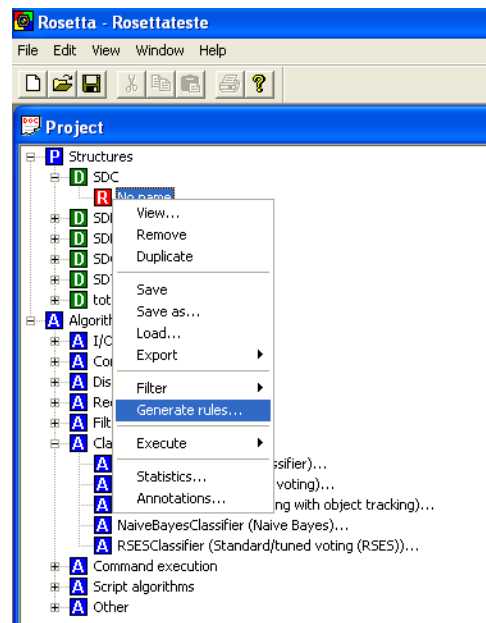


Figura 17 - Geração de regras

Por questões práticas, os atributos do banco de dados, tais como, tensão, potência, número de fases, entre outros, foram denominados conforme a seguir:

P1 – tensão;

P2 – potência;

P3 – número de fases;

P4 – número de consumidores;

P5 – consumo total dos consumidores que consomem entre 0 e 50 kWh;

P6 – consumo total dos consumidores que consomem entre 50 e 100 kWh;

P7 – consumo total dos consumidores que consomem entre 100 e 150 kWh;

P8 – consumo total dos consumidores que consomem entre 150 e 200 kWh;

P9 – consumo total dos consumidores que consomem entre 200 e 300 kWh;

P10 – energia total dos consumidores que consomem entre 300 e 500 kWh;

P11 – energia total dos consumidores que consomem entre 500 e 1000 kWh;

P12 – energia total dos consumidores que consomem mais de 1000 kWh;

P13 – energia total consumido pelo comércio e poder público;

P14 – energia total consumido pela região rural;

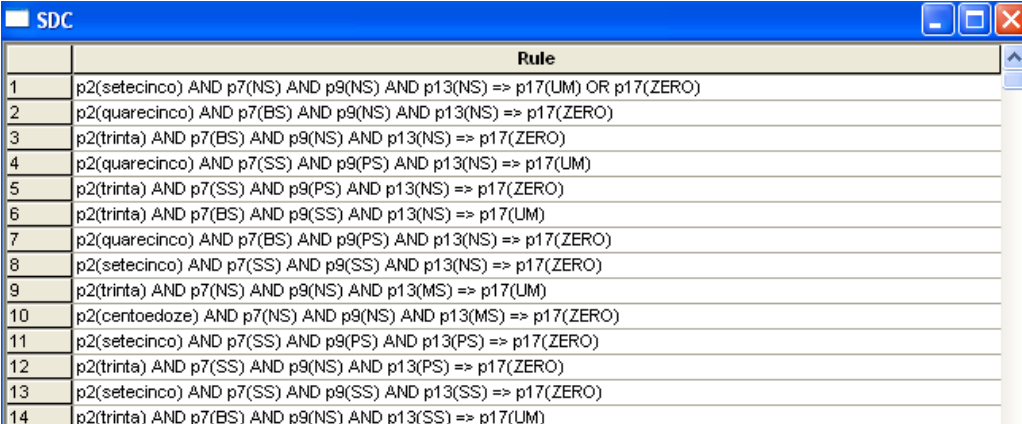
P15 – energia total consumida pela iluminação pública;

P16 – energia total dos consumidores industriais;

P17 – DHTV.

Com esta denominação, observando a Figura 18, pode-se concluir, por exemplo, a regra 1, onde se p2 for igual a “setecinco” e p7 for igual a “NS” e p9 for igual a “NS” e p13 for igual a “NS”, a saída (p17) pode receber valor “UM” (transformador com distorção maior que 6 %) ou valor “0” (transformador com distorção inferior a 6 %). Esta ambiguidade ocorre devido ao fato de que existem transformadores que possuem as mesmas características no banco de dados, porém, possuem saídas diferentes (distorção harmônica total de tensão). O *software Rosetta*, ao invés de excluir a regra devido à incoerência dos dados, gera a regra informando qual a porcentagem de acerto caso um transformador a ser testado se encaixe nesta regra, por exemplo: se no banco de dados existem três transformadores com as mesmas características e com distorção 1, e um transformador com características idênticas possui distorção 0, então o *software* acusa que existe 75% de chance de um transformador a ser testado com as

mesmas características obter classificação 1.



	Rule
1	p2(setecincinco) AND p7(NS) AND p9(NS) AND p13(NS) => p17(UM) OR p17(ZERO)
2	p2(quarecinco) AND p7(BS) AND p9(NS) AND p13(NS) => p17(ZERO)
3	p2(trinta) AND p7(BS) AND p9(NS) AND p13(NS) => p17(ZERO)
4	p2(quarecinco) AND p7(SS) AND p9(PS) AND p13(NS) => p17(UM)
5	p2(trinta) AND p7(SS) AND p9(PS) AND p13(NS) => p17(ZERO)
6	p2(trinta) AND p7(BS) AND p9(SS) AND p13(NS) => p17(UM)
7	p2(quarecinco) AND p7(BS) AND p9(PS) AND p13(NS) => p17(ZERO)
8	p2(setecincinco) AND p7(SS) AND p9(SS) AND p13(NS) => p17(ZERO)
9	p2(trinta) AND p7(NS) AND p9(NS) AND p13(MS) => p17(UM)
10	p2(centoedoze) AND p7(NS) AND p9(NS) AND p13(MS) => p17(ZERO)
11	p2(setecincinco) AND p7(SS) AND p9(PS) AND p13(PS) => p17(ZERO)
12	p2(trinta) AND p7(SS) AND p9(NS) AND p13(PS) => p17(ZERO)
13	p2(setecincinco) AND p7(SS) AND p9(SS) AND p13(SS) => p17(ZERO)
14	p2(trinta) AND p7(BS) AND p9(NS) AND p13(SS) => p17(UM)

Figura 18 – Regras geradas.

Ao final o total de regras geradas por regional foi:

- SDC – 780 regras;
- SDL – 7.270 regras;
- SDN – 718 regras;
- SDO – 3.379 regras;
- SDT – 1938 regras.

Após a geração das regras é possível aplicá-las a um banco de dados que possua as mesmas variáveis analisadas, ou seja, amostras que se deseja prever qual é a distorção harmônica total de tensão através da análise das suas características, com isso é possível realizar a validação do método. Foram utilizadas neste trabalho todas as amostras de cada regional para a geração das regras e, após isso, as mesmas foram aplicadas no mesmo banco de dados, caracterizando a aplicação “todos contra todos”. Esta ação foi aplicada através do *software Rosetta*, conforme a seguir: após selecionar o item *Classification>Bath Classifier*, deve-se então selecionar qual banco de dados deverá ser utilizado para a aplicação das regras. Após isso, o *software* apresenta os resultados da classificação, tal como quantidade de acertos, conforme mostrados na figura 19.

		Predicted		
		ZERO	UM	
Actual	ZERO	51	0	1.0
	UM	2	9	0.818182
		0.962264	1.0	0.967742
ROC	Class	Undefined		
	Area	3.402820e+038		
	Std. error	3.402820e+038		
	Thr. (0, 1)	3.402820e+038		
	Thr. acc.	3.402820e+038		

Figura 19 - Resultado da aplicação das regras

4.2 MÉTODO

4.2.1 Variáveis

Neste trabalho foi considerada a recomendação contida no submódulo 2.2 do ONS, onde é estipulado o limite permissível de distorção harmônica total de tensão de 6% para qualquer fase. Portanto, para a aplicação de *rough sets*, a variável dependente a ser considerada neste estudo, a DHTV, foi utilizada para separar os transformadores em dois grupos distintos, um contendo aquelas amostras que possuem DHTV maior ou igual a 6% em pelo menos uma das fases e o outro contendo as amostras que apresentaram valores inferiores a 6%. Para as variáveis independentes foram consideradas as seguintes características e medições:

- Características nominais dos transformadores;
- Características de consumo.

As variáveis independentes correspondentes às características dos transformadores estão apresentadas na Tabela 18 e as correspondentes às características de consumo dos transformadores estão apresentadas na Tabela 19.

Tabela 18 – Variáveis associadas às características nominais do transformador

Características do Transformador	Classes
Tensão primária (Nível de tensão do alimentador)	<ul style="list-style-type: none"> • 13.800V • 34.500V
Fases (Quantidade de fases no transformador)	<ul style="list-style-type: none"> • 1 fase • 2 fases • 3 fases
Potência (Potência nominal do transformador (kVA))	<ul style="list-style-type: none"> • 112,5 kVA • 75 kVA • 45 kVA • 30 kVA • 15 kVA • 10 kVA • 5 kVA

Fonte: (LACTEC, 2007)

Tabela 19 – Variáveis associadas à característica do consumo

Características do Consumo	Classes
Número de Consumidores	Total de consumidores
Número de Consumidores Segundo o Consumo Médio Mensal	Residencial 0 a 50 kWh
	Residencial 50 a 100 kWh
	Residencial 100 a 150 kWh
	Residencial 150 a 200 kWh
	Residencial 200 a 300 kWh
	Residencial 300 a 500 kWh
	Residencial 500 a 1000 kWh
	Residencial >1000 kWh
	Comércio e Poder Público
	Industrial
	Rural
	Iluminação Pública

Fonte: (LACTEC, 2007)

4.2.2 Tratamento dos Dados

Na aplicação de *rough sets* para a classificação de transformadores segundo sua distorção harmônica total de tensão, foram utilizados os dados levantados na campanha de medições citados anteriormente. Para isso, os mesmos tiveram que ser tratados de forma a favorecer a aplicação da técnica de classificação. Desta forma, algumas ações foram executadas no banco de dados disponível, tais como:

- a) A transformação dos valores absolutos da quantidade de consumidores em porcentagem, conforme exemplificado nas Tabelas 20 e 21.

Tabela 20 - Dados com número de consumidores em valores absolutos

Consumidores	Consumo de 0 - 50	Consumo de 50 - 100	Consumo de 100 - 150	Consumo de 150 - 200
4	0	0	3	1

Tabela 21 - Dados com número de consumidores em porcentagem

Consumidores	Consumo de 0 - 50	Consumo de 50 - 100	Consumo de 100 - 150	Consumo de 150 - 200
4	0	0	75	25

Na tabela 20 pode-se verificar que do total de 4 consumidores, 3 consomem de 100 kWh a 150 kWh e 1 consome de 150 kWh a 200 kWh mensais. Na tabela 21 os mesmos dados são descritos, porém, em porcentagem. Portanto, 75% do total de 4 consumidores consomem de 100 kWh a 150 kWh e 25 % consomem de 150 kWh a 200 kWh mensais.

Esta transformação em porcentagem favorece a análise, pois todos os transformadores podem ser analisados de forma equivalente nas variáveis que levam em consideração a quantidade de consumidores nas faixas de consumo. Vale ressaltar que a informação do total de consumidores não é perdida na análise uma

vez que existe uma variável que possui o valor total de consumidores por transformador.

b) Transformação de todos os valores das variáveis em faixas. Para isso, as variáveis da quantidade total de consumidores de seu consumo por classe de consumidor foram estratificadas em faixas conforme a seguir:

Para a variável do número total de consumidores, os transformadores foram divididos em pequenos (de 0 a 20 consumidores), médios (de 20 a 80 consumidores) e grandes (de 80 a 500 consumidores).

Para as variáveis que levam em consideração o consumo por classe de consumidor (em porcentagem do número total), as faixas foram assim determinadas: Não Significativo – NS – (0 % a 5 % do total de consumo), Pouco Significativo – PS – (5 % a 15 % do total de consumo), Significativo – SS – (15% a 30 % do total de consumidores), Bastante Significativo – BS – (30% a 50 % do total de consumo) e Muito Significativo – MS – (50 % a 100% do total de consumo). Para a transformação dos valores das variáveis em faixas utilizou-se o *software EXCEL*, da *Microsoft*.

Esta mudança de representação dos dados é fundamental para a aplicação de *rough sets*, pois agrupa as variáveis em classes, ou melhor, conjuntos, nos quais a aplicação da técnica se torna viável.

c) Mudança da variável da distorção harmônica total de tensão para uma variável dicotômica. Para isso, todos os transformadores que apresentaram valores iguais ou superiores a 6 % em suas medições, receberam o valor 1, e os transformadores que apresentaram valores inferiores receberam o valor 0. O valor de 6 % foi obtido através da recomendação do Operador Nacional do Sistema Elétrico (ONS). Desta forma, após a aplicação de *rough sets*, é possível classificar os transformadores em “com distorção elevada” classificação 1 e “com distorção dentro do limite” classificação 0.

4.2.3 Forma de Apresentação dos Resultados

Os resultados são apresentados na forma da tabela de confusão

(classificação) conforme exemplificado na Tabela 22, e com a taxa de erro aparente.

Tabela 22 – Tabela de classificação

Classificação Real	Classificação prevista	
	Π_1	Π_2
Π_1	$n_{1/1}$	$n_{1/2}$
Π_2	$n_{2/1}$	$n_{2/2}$

Fonte: (JOHNSON e WICHERN, 1998)

onde: $n_{1/1}$ = número de itens de Π_1 classificadas corretamente como item de Π_1 .

$n_{1/2}$ = número de itens de Π_1 classificados incorretamente como item de Π_2 .

$n_{2/2}$ = número de itens de Π_2 classificados corretamente como item de Π_2 .

$n_{2/1}$ = número de itens de Π_2 classificados incorretamente como item de Π_1 .

Outro dado apresentado é a taxa aparente de erro (*Apparent Error Date* – APER) que é dada por:

$$APER = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2}, \quad (10)$$

onde: n_1 = número total de itens em Π_1 .

n_2 = número total de itens em Π_2 .

Essa taxa deve ser interpretada como a proporção de observações incorretamente classificadas.

4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Nesse capítulo foi descrita a sistemática utilizada para a obtenção dos dados utilizados na classificação dos transformadores através da aplicação de *rough sets*. O campo utilizado para as medições dos transformadores foi a rede de distribuição da concessionária COPEL que atende todo o estado do Paraná e cuja estrutura está dividida em cinco superintendências regionais. Estão descritos os materiais e a metodologia utilizada para a aplicação de *rough sets* no banco de dados obtido através da campanha de medições. São apresentados alguns

indicadores estatísticos obtidos bem como algumas considerações referentes ao banco de dados. A metodologia, as variáveis utilizadas, a transformação de valores contínuos em faixas, mudança de valores absolutos para porcentagem da quantidade de consumidores segundo suas características de consumo e da mudança da variável de saída, DHTV, para “UM” ou “ZERO”, bem como a forma de apresentação dos resultados também foram descritas neste capítulo.

5 ANÁLISE DOS RESULTADOS

5.1 VALIDAÇÃO DO TRATAMENTO DE DADOS

Após o tratamento dos dados optou-se em aplicar a técnica de classificação em cada uma das cinco regionais separadamente e também para o agrupamento de todos os bancos de dados. Para cada teste foram utilizadas todas as amostras (transformadores) para a geração das regras e, então, após isso, as regras foram aplicadas em todos os transformadores, por regional, para o teste de classificação.

A seguir são apresentados os resultados obtidos na aplicação de *rough sets* nos bancos de dados das cinco superintendências regionais da Copel e do agrupamento de todos eles. Os resultados estão dispostos nas Tabelas 23 a 28.

Tabela 23 – Matriz de classificação e APER - SDC

Real	Previsto		
	UM	ZERO	Acerto
UM	9	0	100,00%
ZERO	1	22	95,65%
Acerto	90,00%	100,00%	
APER	3,13%		

Na Tabela 23, pode-se concluir que método obteve 100% de acerto nas previsões dos transformadores que possuíam classificação “UM” (distorção harmônica total de tensão igual ou superior a 6%) e 95,65% de acerto na classificação “ZERO” (distorção harmônica total de tensão inferior a 6%). Outra conclusão que se pode tirar é que das dez classificações realizadas como “UM”, 90% foram corretas e das vinte e duas classificações realizadas como “ZERO”, 100% foram corretas.

Tabela 24 – Matriz de classificação e APER - SDL

Real	Previsto		
	UM	ZERO	Acerto
UM	25	0	100,00%
ZERO	1	102	99,03%
Acerto	96,15%	100,00%	
APER	0,78%		

Na Tabela 24, pode-se concluir que o método obteve 100% de acerto nas previsões dos transformadores que possuíam classificação “UM” (distorção harmônica total de tensão igual ou superior a 6%) e 99,03% de acerto na classificação “ZERO” (distorção harmônica total de tensão inferior a 6%). Outra conclusão é que das 26 classificações realizadas como “UM”, 96,15% foram corretas e das 102 classificações realizadas como “ZERO”, 100% foram corretas.

Tabela 25 – Matriz de classificação e APER - SDN

Real	Previsto		
	UM	ZERO	Acerto
UM	17	0	100,00%
ZERO	0	46	100,00%
Acerto	100,00%	100,00%	
APER	0,00%		

Na Tabela 25, pode-se concluir que o método obteve 100% de acerto nas previsões dos transformadores que possuíam classificação “UM” (distorção harmônica total de tensão igual ou superior a 6%) e 100% de acerto na classificação “ZERO” (distorção harmônica total de tensão inferior a 6%). Outra conclusão é que das 17 classificações realizadas como “UM”, 100% foram corretas e das 46 classificações realizadas como “ZERO”, 100% foram corretas.

Tabela 26 – Matriz de classificação e APER - SDO

Real	Previsto		
	UM	ZERO	Acerto
UM	9	2	81,82%
ZERO	0	51	100,00%
Acerto	100,00%	96,23%	
APER	3,23%		

Na Tabela 26, pode-se concluir que o programa obteve 100% de acerto nas previsões dos transformadores que possuíam classificação “ZERO” (distorção harmônica total de tensão igual ou superior a 6%) e 81,82% de acerto na classificação “UM” (distorção harmônica total de tensão inferior a 6%). Outra conclusão é que das 9 classificações realizadas como “UM”, 100% foram corretas e das 53 classificações realizadas como “ZERO”, 96,23% foram corretas.

Tabela 27 – Matriz de classificação e APER - SDT

Real	Previsto		
	UM	ZERO	Acerto
UM	16	0	100,00%
ZERO	1	57	98,28%
Acerto	94,12%	100,00%	
APER	1,35%		

Na Tabela 27, pode-se concluir que o programa obteve 100% de acerto nas previsões dos transformadores que possuíam classificação “UM” (distorção harmônica total de tensão igual ou superior a 6%) e 98,28% de acerto na classificação “ZERO” (distorção harmônica total de tensão inferior a 6%). Outra conclusão é que das 17 classificações realizadas como “UM”, 94,12% foram corretas e das 57 classificações realizadas como “ZERO”, 100% foram corretas.

Tabela 28 – Matriz de classificação e APER – Agrupamento de todos os bancos de dados

Real	Previsto		
	UM	ZERO	Acerto
UM	63	15	80,77%
ZERO	0	281	100,00%
Acerto	100,00%	94,93%	
APER	1,35%		

Na Tabela 28, pode-se concluir que o programa obteve 80,77% de acerto nas previsões dos transformadores que possuíam classificação “UM” (distorção harmônica total de tensão igual ou superior a 6%) e 100,00% de acerto na classificação “ZERO” (distorção harmônica total de tensão inferior a 6%). Outra conclusão é que das 63 classificações realizadas como “UM”, 100,00% foram corretas e das 296 classificações realizadas como “ZERO”, 94,93% foram corretas.

5.2 COMPARAÇÃO COM OUTRAS METODOLOGIAS DA LITERATURA

Souza (2008) aplicou duas técnicas de Análise Multivariada para a classificação de transformadores utilizando o mesmo banco de dados deste estudo. Com seus dados é possível realizar algumas comparações dos resultados, conforme a Tabela 29.

Tabela 29 – Comparação entre as metodologias utilizadas para a classificação de transformadores quanto à DHTV

Taxas de acertos de classificação (%)			
SDC			
Metodologia	DHTV fora do limite	DHTV dentro do limite	APER
<i>Rough sets</i>	100	95,65	3,13
Regressão Logística	100	91,3	5,88
Escore Quadrático	100	79,91	19,35
Índices de acertos de classificação (%)			
SDL			
Metodologia	DHTV fora do limite	DHTV dentro do limite	APER
<i>Rough sets</i>	100	99,03	0,78
Regressão Logística	84	93,2	8,6
Escore Quadrático	100	88,35	9,75
Índices de acertos de classificação (%)			
SDN			
Metodologia	DHTV fora do limite	DHTV dentro do limite	APER
<i>Rough sets</i>	100	100	0
Regressão Logística	70,59	76,09	25,39
Escore Quadrático	100	78,26	15,87
Índices de acertos de classificação (%)			
SDO			
Metodologia	DHTV fora do limite	DHTV dentro do limite	APER
<i>Rough sets</i>	81,82	100	3,23
Regressão Logística	72,73	86,27	19,23
Escore Quadrático	100	42	48,39
Índices de acertos de classificação (%)			
SDT			
Metodologia	DHTV fora do limite	DHTV dentro do limite	APER
<i>Rough sets</i>	100	98,28	1,35
Regressão Logística	62,5	86,21	18,92
Escore Quadrático	93,75	93,1	6,76

Através das comparações percebe-se que a metodologia *rough sets* obteve o melhor desempenho na previsão em todos os casos comparados.

5.3 EFICÁCIA DA METODOLOGIA

Para a verificação da eficácia da metodologia foi aplicado o método de Lachenbruch (*leave-one-out*), que consiste na classificação de uma amostra retirada do banco de dados (com n-1 amostras) que originou as regras (LACHENBRUCH, 1975). Este procedimento foi realizado para os cinco bancos de dados das regionais da Copel separadamente e também para o banco de dados gerado pelo agrupamento de todos eles. Os índices de acertos obtidos nas classificações das amostras estão descritos na tabela 30 onde também são mostradas a taxa de erro aparente e a porcentagem de amostras indefinidas, ou seja, aquelas que não se enquadraram em nenhuma regra gerada pelo seu respectivo banco de dados.

Tabela 30 – Taxas de acertos de classificação – *Lachenbruch*

Regional	Qtd. Total	DHTV fora do limite	DHTV dentro do limite	Taxa de Lachenbruch	Indefinições
SDC	32	0,0%	55,0%	62,1%	9,4%
SDO	62	0,0%	84,8%	37,8%	27,4%
SDN	63	40,0%	76,0%	34,3%	44,4%
SDT	74	41,7%	79,5%	28,6%	24,3%
SDL	128	9,1%	88,0%	20,4%	19,5%
Agrupado	359	10,5%	71,4%	45,6%	81,1%

A aplicação no banco de dados obtido através do agrupamento das cinco regionais obteve um índice de indefinição relativamente elevado (81,1%), indicando que as cinco superintendências devem ser tratadas em separado.

Ao analisar os resultados de cada regional, percebe-se que quanto maior o número de amostras do banco de dados, menor é a taxa de erro aparente (SDC com 32 amostras obteve APER de 62,1% enquanto que SDL com 128 amostras obteve APER de 20,4%), indicando que a quantidade de amostras influencia o resultado da classificação.

Considerando que a melhor abordagem para classificação é tratar as regionais de forma independente e que a quantidade de amostras influencia a qualidade da classificação, aponta-se para a necessidade de uma amostragem individualizada para cada regional. Isto é confirmado ao se aplicar a equação 6, apresentada no capítulo 3, separadamente para cada regional.

Na tabela 31 são mostradas as quantidades necessárias de amostras quando a equação é aplicada em cada regional, ressaltando-se que aplicando a mesma equação para o número total de amostras, ou seja, a soma de todas as regionais obteve-se um valor de 399 amostras sendo que foram utilizadas neste trabalho 359 amostras (somente transformadores urbanos) distribuídas nas cinco regionais.

Tabela 31 – Número de amostras por regional

	Amostras medidas na campanha de medições	Quantidade de amostras necessárias por regional
SDC	32	397
SDO	62	397
SDN	63	397
SDT	74	397
SDL	128	397

Percebe-se que o número alcançado é bem maior que o utilizado na campanha de medições, logo, ao dividir a quantidade total de amostras proporcionalmente às regionais, perdeu-se representatividade amostral da população por regional.

5.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Os resultados obtidos na aplicação de *rough sets* através do *software Rosetta* para a classificação de transformadores de distribuição segundo sua distorção harmônica total de tensão foram descritos neste capítulo. Na fase de validação, a quantidade encontrada de classificações corretas na aplicação da metodologia foi relativamente elevada demonstrando que a técnica realmente

conseguiu extrair conhecimento do banco de dados com diversas variáveis e que as faixas consideradas garantiram uma boa discernibilidade para as classes. Foi apresentada, também, a taxa de erro aparente (APER) que demonstra a quantidade de erros em relação ao número total de amostras analisadas. Comparações com resultados obtidos através de outras metodologias presentes na literatura foram apresentadas e demonstraram que *rough sets* possui melhor capacidade de classificação nesta aplicação. Por fim, foi realizado o procedimento *leave-one-out* para a classificação das amostras. Os resultados demonstraram um baixo rendimento, devido a segmentação das amostras por regionais, tornando o banco de dados de cada regional insuficiente para se ter uma boa representação estatística do universo em análise.

6 CONCLUSÃO E TRABALHOS FUTUROS

A Qualidade da Energia Elétrica tem se tornado um tema cada vez mais importante nos últimos anos, pois ela pode gerar prejuízos tanto para as empresas fornecedoras de energia como para os seus consumidores. Neste contexto, a distorção harmônica total de tensão tem um papel importante para as concessionárias de energia, Assim, o aumento de cargas não lineares (equipamentos eletrônicos) no sistema é necessária uma quantidade maior de informações sobre o fenômeno para que seja possível uma atuação mais efetiva a fim de diminuir seus efeitos no sistema.

Para a classificação de transformadores de distribuição de energia elétrica quanto a DHTV Aplicou-se nesse trabalho a teoria de *rough sets*, uma vez que após o levantamento bibliográfico constatou-se a potencialidade dessa técnica para a classificação, e que ainda não existe um grande número de aplicações em sistemas de potência.

A aplicação de *rough sets* exigiu um conjunto de dados obtido na campanha de medições de QEE e tratados aqui. Para isso, os dados de consumo dos transformadores, que estavam divididos em classes, foram transformados em porcentagem do consumo total e, dependendo deste valor percentual, este atributo foi caracterizado dentro de uma faixa pré-estabelecida.

Após o tratamento dos dados, foi utilizado o *software Rosetta*, para a aplicação da metodologia proposta. Este programa, que foi desenvolvido por Ohrn (1999) especificamente para a aplicação de *rough sets*, se mostrou eficiente e flexível.

Os testes foram realizados em duas etapas, a primeira teve como objetivo a validação das faixas utilizadas na metodologia e a segunda a avaliação da eficácia de *rough sets* para a classificação de transformadores de tensão segundo à DHTV.

Na fase de validação da aplicação da metodologia, após a geração das regras, as mesmas foram aplicadas ao seu banco de dados e os resultados se apresentaram muito bons, já que foram atingidos índices elevados de acerto,

mostrando que as faixas determinadas estavam coerentes. Foram realizadas comparações com os resultados obtidos com outras metodologias, tais como: regressão logística e escore quadrático. A teoria de *rough sets* se mostrou mais eficiente na classificação do banco de dados disposto do que as demais.

Para a avaliação da eficácia da metodologia, foi utilizado o procedimento *leave-one-out* (Lachenbruch), para o qual o resultado da classificação não teve o mesmo desempenho da fase de validação. Uma análise dos resultados apontou que a quantidade de amostras utilizadas, por regional, influencia diretamente o resultado da classificação e é insuficiente para representar adequadamente o universo de transformadores de cada regional. Foi possível verificar, também, que cada uma das cinco superintendências regionais deve ser tratada separadamente, pois o número de amostras indefinidas no teste realizado no banco de dados do agrupamento das cinco regionais foi muito elevado, sendo superior a 80%.

Em trabalhos futuros, sugere-se realizar uma nova campanha de medições para aumentar o banco de dados de cada regional para a correta avaliação da eficácia da metodologia. Sugere-se, também, realizar testes agrupando algumas regionais que apresentem características semelhantes.

Outra sugestão é a aplicação de *rough sets* em conjunto com outras técnicas de classificação (sistemas híbridos) para melhorar o desempenho de classificação.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL. **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional**. PRODIST, módulo 8, ago. 2007.
- BARKER, P. P.; BURKE, J. J.; MANCAO, R.T.; SHORT, T.A.; WARREN, C.A.; BURNS, C.W.; SIEWIERSKI, J.J.; **Power Quality Monitoring of a Distribution System**. IEEE Transaction on Power Delivery, vol 9, n. 2, apr 1994.
- BONALDI, E. L.; LAMBERT-TORRES, G.; SILVA, L. E. B.; OLIVEIRA, L. E. L.; **A Rough Sets Based Classifier For Induction Motors Fault Diagnosis**; WSEAS Transactions on Systems, Londres, v. 2, n. 2, p. 230-237, 2002.
- CAO, Y.J.; FENG, L.; QIU, J.J.; **Performance of the novel rough fuzzy-neural network on short-term load forecasting**. IEEE Power Systems Conference And Exposition, v.1, p.543-547, Oct 2004.
- CERCHIARI, S. C.; LAMBERT-TORRES, G.; PINTO, J. O. P.; SAUER, L.; TEURYA, A.; ZORZATE, E.H.; **Data Mining in Distribution Consumer Database using rough sets and Self-Organizing Maps**. IEEE PES Power Systems Conference and Exhibition, vol 1, p.38-43, nov 2006.
- CHAUDHURI, B.B.; MITRA, M.; MITRA S.; **A Rough-Set-Based Inference Engine for ECG Classification**. IEEE Transactions on Instrumentation And Measurement, vol.55, n.6, p.2198-2206, dec. 2006.
- CHEN, L. F.; CHIEN, C.F.; **Using Rough Set Theory To Recruit And Retain High-Potential Talents For Semiconductor Manufacturing**. IEEE Transactions on Semiconductor Manufacturing, v.20, n.4, nov 2007.
- CHEN, T. C.; PAI, P. F.; **Rough set theory with discriminant analysis in analyzing electricity loads**; ELSEVIER Expert Systems with Applications, v.36, p.8799-8806, 2009.

- CHIEN C.F.; PENG, J. T.; TSENG T.L.B.; **Rough set theory for data mining for fault diagnosis on distribution feeder**; IEE Proceedings-Generation, Transmission and Distribution, v.151, p.689-697, nov 2004.
- CHOONG, F.; KAMADA, M.; REAZ, M.B.I.; SULAIMAN, M.S.; YASIN, F.M.; **Expert System for Power Quality Disturbance Classifier**. IEE Transactions on Power Delivery, v.22, n.3, p.1979-1988, jul 2007.
- CHOW, L. Y.; XU, L.; **A Classification Approach for Power Distribution Systems Fault Cause Identification**. IEE Transactions on Power Systems v.21, n.1, feb 2006.
- COUTINHO, M. P.; **Detecção de Ataques em Infra-Estruturas Críticas de Sistemas Elétricos de Potência Usando Técnicas Inteligentes**. Universidade Federal de Itajubá. 2007.
- CROSSLEY, P.A.; HOR, C.L.; WATSON S.J.; **Building Knowledge for Substation-Based Decision Support Using Rough SetS**. IEE Transactions on Power Delivery, v.22, n.3, jul 2007.
- DALSTEIN, T.; KULICKE, B.; **Neural Network Approach To Fault Classification For High Speed Protective Relaying**. IEE Transactions on Power Delivery, v.10, n.2, apr 1995.
- DASH, P.K.; LIEW, A.C.; MISHRA, S.; SALAMA, M.A.; **Classification of Power System Disturbances Using a Fuzzy Expert System and a Fourier Linear Combiner**. IEE Transactions on Power Delivery, v.15, n.2, apr 2000.
- DAVIDSON, J; FOUAD, J.J.; ZHOU, Q.; **Application Of Artificial Neural Networks In Power System Security And Vulnerability Assessment**. IEE Transactions on Power Systems, v.9, n.1, feb 1994.
- DUGAN, R. C.; MCGRANAGHAN, M. F.; SANTOSO, S.; BEATY, H. W.; **Electrical Power Systems Quality – 2nd Edition**. McGraw-Hill: New York, 528 p. 2002.
- GHOSH, A. K.; LUBKEMAN, D.L.; **The Classification Of Power System Disturbance Waveforms Using A Neural Network Approach**. IEE Transactions on Power Delivery, v.10, n.1, p. 109-115, jan 1995.

- GRECO, S.; MATARAZZO, B.; SLOWINSKI, R.; **Rough Sets Theory For Multicriteria Decision Analysis**. ELSEVIER European Journal of Operational Research, v. 129, n.1, feb 2001.
- HAN, L.; PETERS, J.F.; RAMANNA, S.; ZHAI, R.; **Classifying Faults in High Voltage Power Systems: A Rough-Fuzzy Neural Computational Approach**. Springer Berlin, v. 1711, p. 47-54, 1999.
- JAGIELSKA, I., MATTHEWS, C., WHITFORT, T. **An Investigation Into The Application Of Neural Networks, Fuzzy Logic, Genetic Algorithms, And Rough Sets To Automated Knowledge Acquisition For Classification Problems**. ELSEVIER Neurocomputing, v.24, p.37-54, 1999.
- KAGAN, N.; FERRARI, E.L.; MATSUO, N.M.; DUARTE, S.X.; CAVARETTI, J.L.; TENORIO, A.; SOUZA, L.R.; **A Methodology for Monitoring Harmonic Distortion in Electric Power Distribution Systems**. IEEE International Conference on Harmonics and Quality of Power , v.1, p.322-327, 2002.
- KEERTHIPALA, W.W.L.; WANG, H.; **Fuzzy-Neuro Approach to Fault Classification for Transmission Line Protection**. IEEE Transactions on Power Delivery, v.13, n.4, p. 1093-1104, oct 1998.
- KRYSZKIEWIEZ, M.; **Rough Set Approach to Incomplete Information System**; ELSEVIER Information Sciences, v.112, p. 39-49, 1998.
- LACHENBRUCH, P. A. **Discriminant Analysis**. New York: Hafner Press. 1975.
- LAMBERT-TORRES, G.; **Application of Rough Sets in Power System Control Center Data Mining**. IEEE Power Engineering Society Winter Meeting, v.1, p.627-631, jan. 2002.
- MASSART, D. L.; WALCZAK, B.; **Rough Sets Theory**. ELSEVIER Chemometrics and Intelligent Laboratory Systems, v.47, p.1-16, 1999.
- OHRN, A. **Discernibility and Rough sets in Medicine: Tools and Applications**. Norwegian University of Science and Technology. 1999.
- OPERADOR NACIONAL DO SISTEMA ELÉTRICO – ONS. **Padrões de Desempenho da Rede Básica**, submódulo 2.2, ago. 2002.

- PATRÍCIO, C. M. M. M; **Rought Sets – Técnica de Redução de Atributos e Geração de Regras para Classificação de Dados**. Universidade Federal do Mato Grosso do Sul. 2003.
- PAWLAK, Z. **Rough sets**. Int. J. Comput. Inf. Sci. 11, 341-356. 1982.
- PAWLAK, Z.; **Rough sets: theoretical aspects of reasoning about data**. Kluwer Academic Publishers, 1991.
- PAWLAK, Z; BUSSE, J.G; SLOWINSKI, R.; ZIARKO, W. **Rought Sets**. Communications of the ACM, v.38, n.6, p.88-95, nov. 1995.
- PENICHE, R. A.; TEIXEIRA, M. D. **Conceitos Gerais sobre Qualidade de Energia Elétrica**. Instituto de Tecnologia para o Desenvolvimento, apostila, 2005.
- SOUZA, W. E.; **Análise das Distorções Harmônicas de Tensão a partir de Características dos Transformadores e de Dados de Consumo**. Universidade Federal do Paraná, 2008.
- SWINIARSKI, R. W.; SKOWRON, A.; **Rough Set methods In Feature Selection And Recognition**. ELSEVIER Pattern Recognition Letters, v.24, p.833-849, 2003.
- YASDI R.; **Learning Classification Rules from Database In The Context Of Knowledge Acquisition And Representation**. IEEE Transactions on Knowledge and Data Engineering, v. 3, n.3, set 1991.