

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

EXPERIMENTOS DE MINERAÇÃO DE DADOS APLICADOS A SISTEMAS SCADA  
DE USINAS HIDRELÉTRICAS

IVALDO OHANA

TD 05/2012

UFPA / ITEC / PPGEE  
Campus Universitário Guamá  
Belém-Pará-Brasil  
2012



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

EXPERIMENTOS DE MINERAÇÃO DE DADOS APLICADOS A SISTEMAS SCADA  
DE USINAS HIDRELÉTRICAS

IVALDO OHANA

TD 05/2012

UFPA / ITEC / PPGEE  
Campus Universitário Guamá  
Belém-Pará-Brasil  
2012

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

IVALDO OHANA

EXPERIMENTOS DE MINERAÇÃO DE DADOS APLICADOS A SISTEMAS SCADA  
DE USINAS HIDRELÉTRICAS

Tese submetida à Banca Examinadora  
do Programa de Pós-Graduação em  
Engenharia Elétrica da UFPA para a  
obtenção do Grau de Doutor em  
Engenharia Elétrica

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

2012

---

O36e

Ohana, Ivaldo

Experimentos de mineração de dados aplicados a sistemas scada de usinas hidrelétricas / Ivaldo Ohana; orientador, Ubiratan Holanda Bezerra. – 2012.

Tese (Doutorado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2012.

1. Mineração de dados (computação). 2. Usinas hidrelétricas – processamento de dados. 3. Sistemas de energia elétrica – planejamento. I. Orientador. II. Título.

CDD 22. ed. 005.74

---

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

EXPERIMENTOS DE MINERAÇÃO DE DADOS APLICADOS A SISTEMAS  
SCADA DE USINAS HIDRELÉTRICAS

AUTOR: IVALDO OHANA

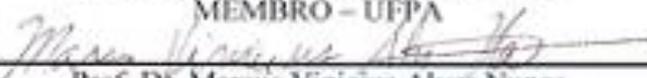
TESE DE DOUTORADO SUBMETIDA À AVALIAÇÃO DA BANCA EXAMINADORA  
APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL DO PARÁ E JULGADA  
ADEQUADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA  
ELÉTRICA NA ÁREA DE SISTEMAS DE ENERGIA ELÉTRICA

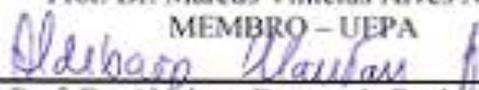
APROVADA EM : 13/04/2012.

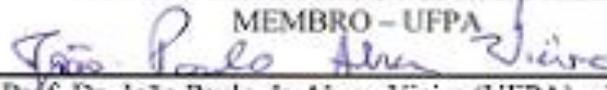
BANCA EXAMINADORA

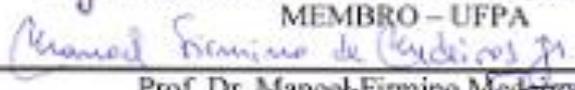
  
Prof. Dr. Ubiratã Holanda Bezerra  
ORIENTADOR - UFPA

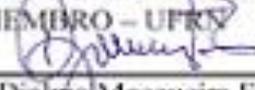
Prof. Dr. Carlos Renato Lisboa Francês  
MEMBRO - UFPA

  
Prof. Dr. Marcus Vinicius Alves Nunes  
MEMBRO - UFPA

  
Prof. Dr. Aldebaro Barreto da Rocha Klautau  
MEMBRO - UFPA

  
Prof. Dr. João Paulo de Abreu Vieira (UFPA) - Membro  
MEMBRO - UFPA

  
Prof. Dr. Manoel Firmino Medeiros Júnior  
MEMBRO - UFPA

  
Prof. Dr. Djalma Mosqueira Falcão  
MEMBRO - COPPE/UFRJ

VISTO:

  
Prof. Dr. Marcus Vinicius Alves Nunes  
COORDENADOR DO PPGE/TEC/UFPA

## DEDICATÓRIA

*Dedico esta Tese a todos meus familiares e em especial a meu pai.*

## AGRADECIMENTOS

- Agradeço primeiramente a Deus;
- Aos professores Ubiratan Holanda Bezerra e João Paulo de Abreu Vieira pelas orientações, ensinamentos e conselhos no desenvolvimento deste trabalho;
- Aos professores Marcus Vinicius Alves Nunes, Aldebaro Barreto da Rocha Klautau, Renato Francês, Manoel Firmino de Medeiros Júnior e Djalma Mosqueira Falcão, que com suas recomendações ajudaram a melhorar ainda mais este trabalho;
- Aos demais professores, colegas e funcionários do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará;
- A minha esposa Maria da Conceição Pamplona Ohana, aos meus filhos Alan Pamplona Ohana e Alex Pamplona Ohana, pelo incentivo e apoio em todos os momentos do desenvolvimento desta Tese;
- A todos que contribuíram, de forma direta ou indireta, na elaboração deste trabalho.

## RESUMO

O atual modelo do setor elétrico brasileiro permite igualdade de condições a todos os agentes e reduz o papel do Estado no setor. Esse modelo obriga as empresas do setor a melhorarem cada vez mais a qualidade de seu produto e, como requisito para este objetivo, devem fazer uso mais efetivo da enorme quantidade de dados operacionais que são armazenados em bancos de dados, provenientes da operação dos seus sistemas elétricos e que tem nas Usinas Hidrelétricas (UHE) a sua principal fonte de geração de energia. Uma das principais ferramentas para gerenciamento dessas usinas são os sistemas de Supervisão, Controle e Aquisição de Dados (*Supervisory Control And Data Acquisition - SCADA*). Assim, a imensa quantidade de dados acumulados nos bancos de dados pelos sistemas *SCADA*, muito provavelmente contendo informações relevantes, deve ser tratada para descobrir relações e padrões e assim ajudar na compreensão de muitos aspectos operacionais importantes e avaliar o desempenho dos sistemas elétricos de potência. O processo de Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery in Database - KDD*) é o processo de identificar, em grandes conjuntos de dados, padrões que sejam válidos, novos, úteis e compreensíveis, para melhorar o entendimento de um problema ou um procedimento de tomada de decisão. A Mineração de Dados (ou *Data Mining*) é o passo dentro do *KDD* que permite extrair informações úteis em grandes bases de dados. Neste cenário, o presente trabalho se propõe a realizar experimentos de mineração de dados nos dados gerados por sistemas *SCADA* em UHE, a fim de produzir informações relevantes para auxiliar no planejamento, operação, manutenção e segurança das hidrelétricas e na implantação da cultura da mineração de dados aplicada a estas usinas.

**PALAVRAS CHAVES:** Mineração de dados, Descoberta de Conhecimento em Banco de Dados, Usina Hidrelétrica, Hidrogerador, Associações, Árvore de Decisão, Alarmes, Eventos.

## ABSTRACT

The current model of the Brazilian electric sector allows equal terms to all actors and reduces the role of the State in this sector. This model forces the electrical utilities to improve the quality of their products and, as a prerequisite for this purpose, they should make more effective use of the enormous amount of operational data that are stored in databases, acquired from the operation of their electrical systems which use the hydroelectric power plants as their main source of energy generation. One of the main tools for managing the operation of these plants are the Supervisory Control and Data Acquisition systems (SCADA). Thus, the large amount of data stored in databases by SCADA systems, certainly containing relevant information, should be treated to discover relationships and patterns that would help in the understanding of many important operational aspects as well as in the evaluation of operational performance of the electric power systems. The process of Knowledge Discovery in Database (KDD) is the process of identification of patterns in large data sets, that are valid, new, and useful to improve the understanding of a problem or a decision-making procedure. Data Mining is the step within KDD that extracts useful information from large databases. In this scenario, the present study objective is to perform data mining experiments on data generated by power plants SCADA systems, to produce relevant information to assist in planning, operation, maintenance and security of hydro power plants and also contribute to the implementation of the culture of using data mining techniques applied to these plants.

**KEY WORDS:** Data Mining, Knowledge Discovery in Databases, Power Plant, Hydrogenerator, Associations, Decision Tree, Alarms, Events.

## SUMÁRIO

1	INTRODUÇÃO	17
1.1	Descrição Geral do Problema.....	17
1.2	Objetivos e Contribuições.....	19
1.3	Revisão Bibliográfica.....	20
1.4	Estrutura da Tese.....	28
2	USINAS HIDRELÉTRICAS E SISTEMAS SCADA.....	30
2.1	Introdução.....	30
2.2	Usinas Hidrelétricas.....	30
2.2.1	Unidades Hidrogeradoras.....	31
2.2.2	Turbina.....	32
2.2.3	Gerador Elétrico.....	32
2.2.4	Casa de Força.....	33
2.3	Sistema SCADA.....	33
2.3.1	Arquitetura de um Sistema SCADA.....	34
2.3.2	Funcionalidades de um Sistema SCADA.....	35
2.3.3	Variáveis Simples.....	36
2.3.4	Variáveis Compostas.....	36
2.3.5	Gestão de Arquivos e Banco de Dados.....	37
2.3.6	Visualização.....	37
2.3.7	Alarmes e Eventos.....	38
2.3.8	Históricos e Relatórios.....	38
2.3.9	Gráficos de Tendência.....	39
2.3.10	Receitas.....	40
2.3.11	Tratamentos Estatísticos.....	40
2.4	Conclusões.....	41
3	DESCOBERTA DE CONHECIMENTO E MINERAÇÃO DE DADOS.....	42
3.1	Introdução.....	42
3.2	Descoberta de Conhecimento em Banco de Dados.....	43
3.3	Mineração de Dados.....	45
3.3.1	Tarefas Preditivas.....	48
3.3.2	Tarefas Descritivas.....	51
3.4	Conclusões.....	53
4	MINERAÇÃO DE DADOS EM USINAS HIDRELÉTRICAS.....	54

4.1	Introdução.....	54
4.2	Análises Estatísticas.....	55
4.2.1	Variáveis Discretas.....	56
4.2.1.1	Análise de Eventos na Usina.....	57
4.2.2	Variáveis Analógicas.....	59
4.2.2.1	Comportamento de Grandezas do Hidrogerador.....	60
4.2.2.2	Correlações entre as Variáveis.....	61
4.2.2.3	Análise de Frequência.....	62
4.2.2.4	Ponto de Operação em Relação a Curva de Capacidade.....	63
4.3	Regras de Associações.....	66
4.3.1	Associações entre Eventos da Usina.....	67
4.3.2	Associações entre Alarmes dos Hidrogeradores.....	68
4.3.3	Associações entre Alarmes de Frequência.....	69
4.4	Árvore de Decisão.....	70
4.4.1	Potência Gerada na Usina.....	70
4.4.2	Nível de Água na Cidade a Jusante da Barragem.....	72
4.4.3	Limites Operacionais do Hidrogerador.....	74
4.5	Conclusões.....	78
5	CONCLUSÕES.....	79
5.1	Conclusões Gerais.....	79
5.2	Perspectivas de Futuros Trabalhos.....	80
	REFERÊNCIAS.....	81
	ANEXO I.....	91
	ANEXO II .....	93

## LISTA DE ILUSTRAÇÕES

Figura 2.1 - Principais componentes de uma usina hidrelétrica.....	31
Figura 2.2 - Turbina hidráulica.....	32
Figura 2.3 - Gerador elétrico de hidrogeradores.....	32
Figura 2.4 - Arquitetura básica de um sistema SCADA.....	34
Figura 2.5 - Exemplo de um unifilar em vídeo e um sinótico.....	37
Figura 2.6 - Exemplo de alarmes e eventos em vídeo.....	38
Figura 2.7 - Exemplo de um relatório de alarmes e eventos.....	39
Figura 2.8 - Exemplo de um gráfico de tendência.....	39
Figura 2.9 - Exemplo de um gráfico estatístico dos eventos.....	40
Figura 3.1 - Pirâmide do conhecimento.....	43
Figura 3.2 - Etapas que constituem o processo KDD.....	43
Figura 3.3 - Posição do KDD na solução de um problema.....	45
Figura 3.4 - Relação da mineração de dados com outras áreas.....	47
Figura 3.5 - Tarefas da mineração de dados.....	48
Figura 3.6 - Princípio básico da classificação.....	49
Figura 3.7 - Abordagem para o modelo de classificação.....	49
Figura 4.1 - Quantidade de eventos da usina no período de 17 a 21/01/2010.....	58
Figura 4.2 - Média de eventos por hora durante o período.....	58
Figura 4.3 - Distribuição dos alarmes entre os hidrogeradores no período.....	59
Figura 4.4 - Análise das grandezas analógicas.....	61
Figura 4.5 - Correlação entre a potência ativa e a pressão da caixa espiral .....	62
Figura 4.6 - Análise de desempenho de frequência.....	63
Figura 4.7 - Curva de capacidade do gerador síncrono.....	63
Figura 4.8 - Pontos de operação do hidrogerador (18/01/2010).....	65
Figura 4.9 - Análise dos pontos de operação do hidrogerador no período.....	66
Figura 4.10 - Árvore para a potência gerada.....	72
Figura 4.11 - Árvore para o nível de água na cidade.....	73
Figura 4.12 - Árvore para os limites operacionais.....	76
Figura 4.13 - Interface tempo real para os hidrogeradores.....	77

## LISTA DE TABELAS

Tabela 3.1 - Matriz de confusão.....	50
Tabela 3.2 - Tarefas e técnicas de mineração dados.....	53
Tabela 4.1 - Exemplos de registros de variáveis discretas.....	56
Tabela 4.2 - Tipos de eventos discretos.....	57
Tabela 4.3 - Tipos de urgências dos eventos discretos.....	57
Tabela 4.4 - Exemplos de registros de variáveis analógicas.....	60
Tabela 4.5 - Regiões da curva de capacidade.....	64
Tabela 4.6 - Características do hidrogerador.....	65
Tabela 4.7 - Regras para a árvore de decisão da potência gerada.....	71
Tabela 4.8 - Regras para a árvore de decisão do nível da água na cidade.....	73
Tabela 4.9 - Rótulos da classe C.....	75
Tabela 4.10 - Regras geradas para os limites operacionais do hidrogerador.....	75

## LISTA DE SIGLAS

SCADA	- Supervisory Control And Data Acquisition Controle Supervisório e Aquisição de Dados
DM	- Data Mining Mineração de Dados
MD	- Mineração de Dados
KDD	- Knowledge Discovery in Databases Descoberta de Conhecimento em Banco de Dados
SEP	- Sistemas Elétricos de Potência
PMU	- Phasor Measurement Units Unidade de Medição Fasorial
IEEE	- Institute of Electrical and Electronics Engineers Instituto de Engenheiros Elétricos e Eletrônicos
GRNN	- Generalized Regression Neural Networks Rede Neural Com Regressão Generalizada
NN	- Neural Networks Rede Neural
TDS	- Time Domain Simulation Simulação no Dominio do Tempo
PNN	- Probabilistic Neural Network Rede Neural Probabilística
MLP	- Multi Layered Perpectron
OLAP	- On-Line Analytical Processing Processamento Analítico em Tempo Real
SVM	- Support Vector Machine Máquinas de Vetor de Suporte
DBSCAN	- Density-Based Spatial Clustering of Applications with Noise Aplicações de Agrupamento com Ruído Baseada em Densidade Espacial
ONS	- Operador Nacional do Sistema
ANEEL	- Agência Nacional de Energia Elétrica
SAGE	- Sistema Aberto de Gerenciamento de Energia
CLP	- Controladores Lógicos Programáveis
UTR	- Unidades Terminais Remotas
CBM	- Condition-Based Maintenance Manutenção Baseada na Condição

- LASSO - Least Absolute Shrinkage and Selection Operator
- SEPOPE - Simpósio dos Especialistas em Planejamento da Operação e Expansão Elétrica
  
- SBSE - Seminário Brasileiro de Sistemas de Energia
- ATP-EMTP - Alternative Transients Program - Electromagnetic Transients Program
- DFT - Discrete Fourier Transform  
Transformada Discreta de Fourier
- VTCD - Variação de Tensão de Curta Duração
- WEKA - Waikato Environment for Knowledge Analysis

## LISTA DE SÍMBOLOS

$P$	- Potência ativa
$Q$	- Potência reativa
$V_t$	- Tensão terminal da máquina
$I_{amax}$	- Máxima corrente de armadura
$P_{mec\_max}$	- Potência mecânica máxima
$X_d$	- Reatância síncrona do eixo direto
$X_q$	- Reatância síncrona do eixo em quadratura
$E_{max}$	- Tensão interna máxima, determinada pela tensão máxima de campo
$E_{min}$	- Tensão interna mínima, determinada pela tensão mínima de campo
$\delta$	- Ângulo da tensão interna da máquina
$\varnothing$	Ângulo da tensão terminal da máquina

# CAPÍTULO 1

## Introdução

### 1.1 Descrição Geral do Problema

O crescimento mundial da demanda de energia elétrica em velocidade maior que a geração de energia, tem provocado preocupações a nível nacional e internacional. Tal crescimento, basicamente é devido ao crescimento da sociedade. No Brasil, se instala a preocupação, na medida em que o sistema elétrico brasileiro vem operando com níveis de geração próximos a demanda dos usuários. Mesmo com investimentos estatais e privados, ainda assim, a preocupação permanece, pois o nível de crescimento atual supera os investimentos sendo realizados e já os compromete quando finalizados. Os órgãos reguladores do mercado de energia elétrica preocupados com este cenário exigem das empresas geradoras a garantia do fornecimento para garantir o consumo crescente. Neste contexto é de fundamental importância a otimização dos processos de planejamento, operação e manutenção das unidades geradoras e para tal é necessário aprofundar o conhecimento sobre os sistemas geradores de energia.

Um sistema de Supervisão, Controle e Aquisição de Dados, tecnicamente conhecido como SCADA é um sistema utilizado para coletar, armazenar, analisar e apresentar os dados de uma planta industrial, sob supervisão e controle. Em alguns desses sistemas, em função dos dados coletados e analisados, algumas decisões são tomadas e enviadas aos atuadores que interferem no processo controlado em tempo real. O sistema SCADA é um processo comum de aplicação de controle, que adquire dados do processo por meio de estações remotas e os envia para processamento por um computador central. Um sistema desse tipo pode variar de uma concepção mais simples, até grandes aplicações mais complexas de coleta e análise de dados, e comando de grandes processos industriais. Os dados obtidos (estados de equipamentos, valores de variáveis, alarmes, ultrapassagem de limites, etc) são armazenados em meios magnéticos e, com o passar do tempo, formam grandes bancos de dados cheios de eventos e valores de variáveis ocorridos no processo sob supervisão e controle. Esses históricos de dados formam a base para a análise do comportamento do processo. O crescimento do volume de dados provoca a necessidade de novas técnicas e ferramentas

capazes de transformar, de forma inteligente e automática, esses dados em informações significativas e em conhecimento. Essas informações, de grande valia para o planejamento, gestão e tomadas de decisão, poderão estar, implícitas ou escondidas sob uma montanha de dados, e não podem ser descobertas ou, no mínimo, facilmente identificadas utilizando-se sistemas convencionais de gerenciamento de banco de dados. Em resposta a essa necessidade, surgiu a Mineração de Dados (MD) que é uma tecnologia que emergiu da interseção de três áreas: estatística clássica, inteligência computacional e aprendizado de máquina, sendo a primeira a mais antiga delas. Observa-se que a MD é parte de um processo maior conhecido como Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases - KDD*), que permite a extração de conhecimento potencialmente útil e previamente desconhecido de um banco de dados.

As Usinas Hidrelétricas (UHE) brasileiras em geral são supervisionadas e controladas por meio de sistemas SCADA, gerando grandes quantidades de dados (discretos e analógicos) que ocorrem durante as suas operações. Essas usinas formam um ambiente bastante complexo em que as relações de causas e efeitos entre as variáveis são conhecidas – mas não completamente mapeadas. Assim, a análise de correlações entre variáveis, a identificação das causas de paradas de equipamentos e processos, a adoção de medidas e procedimentos para a melhoria da qualidade da energia, a programação da operação e o aumento da eficiência da produção de energia elétrica, podem ter ações de melhorias a partir do maior conhecimento do processo, pela extração de conhecimento a partir dos dados de supervisão e controle. De uma forma geral, somente os dados referentes às contingências mais severas ocorridas no processo, como por exemplo, a parada de um hidrogenador, é analisado em pós-operação para se conhecer as causas e os impactos sobre o sistema.

O grande volume de dados referentes à operação normal do sistema em geral não é analisado de forma profunda e a sua análise poderia levar a se obter informações relevantes. Por outro lado, falhas no sistema elétrico produzem a conhecida “avalanche de alarmes”, na qual uma grande quantidade de eventos associados a vários equipamentos (por exemplo: disjuntores, bombas, ventiladores, etc) e alarmes associados a grandezas elétricas (como: frequência, potências ativa e reativa geradas, tensões e correntes, etc) são disparados em curto espaço de tempo. Nesta situação, o operador tem enormes dificuldades para diagnosticar corretamente à causa da falha, em tempo real e, restabelecer a normalidade operativa. Normalmente, ocorre o restabelecimento do sistema e a causa deverá ser analisada em pós-operação.

Além desses aspectos, mesmo em operação normal, ocorrem com certa frequência casos em que aparentemente tudo parece controlado, porém algumas grandezas sob supervisão podem estar próximas a limites perigosos e o operador do sistema deve ser avisado em tempo hábil para tomar as providências necessárias. Neste contexto, por exemplo, se enquadra o caso em que o ponto de operação do hidrogerador em relação às potências ativa e reativa geradas pode estar próximas dos limites operacionais impostos pela curva de capacidade.

A aplicação de técnicas de mineração de dados, como estatísticas, associações, descoberta de padrões, regressão, agrupamentos, redes neurais artificiais, árvores de decisão e outras, nos dados gerados pelos sistemas SCADA nas UHE, podem vir a ajudar na solução das dificuldades apresentadas anteriormente. O resultado da aplicação das técnicas citadas poderão a vir contribuir de forma significativa no planejamento, na operação, na manutenção e na segurança operacional dos sistemas elétricos de potência, em especial das UHE.

## **1.2 Objetivos e Contribuições**

O objetivo desse trabalho é contribuir para o avanço do estado da arte da mineração de dados, aplicada aos dados de sistemas SCADA de Usinas Hidrelétricas, com a realização de experimentos em dados de uma usina hidrelétrica real do sistema Brasileiro.

A principal contribuição desta tese é demonstrar como aplicar a mineração de dados para se obter informações que possam ser relevante para o planejamento, operação, manutenção e segurança das usinas hidrelétricas e, assim, fomentar a cultura da utilização desta técnica para gerar subsídios ao processo de tomada de decisão.

O ineditismo dessa tese reside na aplicação das técnicas de mineração de dados em sistemas SCADA de usinas hidrelétricas, mostrando uma outra forma de análise dos dados em pós-operação, para avaliar os aspectos operacionais da usina, e assim, ajudar no planejamento, operação, manutenção e segurança da mesma.

Como contribuições tecnológicas desta tese destacam-se:

- Desenvolvimento de aplicações de mineração de dados usando estatísticas, associações e árvore de decisão, para extração de conhecimento, a partir dos dados históricos da operação de uma usina hidrelétrica brasileira, usando variáveis discretas e analógicas;
- Desenvolvimento de um software para geração automática de uma base de dados rotulada, a ser aplicada na detecção dos limites operacionais de um hidrogerador usando a curva de capacidade;

- Desenvolvimento de uma árvore de decisão, a partir da base de dados gerada, para detectar os limites operacionais do hidrogerador em relação a sua curva de capacidade;
- Desenvolvimento de *interface* para avaliar em tempo real a segurança da operação do hidrogerador, baseada na árvore de decisão gerada.

Vale ressaltar que, ao longo do desenvolvimento dessa tese alguns trabalhos foram motivados, os quais são citados no Anexo I.

### 1.3 Revisão Bibliográfica

Poucos artigos a nível nacional e até mesmo internacional são encontrados na literatura especializada, abordando as aplicações de mineração de dados em sistemas SCADA em usinas hidrelétricas, especialmente usando estatísticas, associações e árvore de decisão. Contudo, uma revisão bibliográfica da mineração de dados aplicada em sistemas elétricos de potência foi realizada,

LI et al., (2008) usando dados históricos da operação de uma turbina, fazem uma análise de diagnóstico de faltas baseada em tipos de alarmes de vibração para encontrar relacionamentos entre eles. Além disso, uma mineração de regras de associação quantitativa é efetuada para descobrir parâmetros importantes da operação de uma usina movida a carvão, para eficiência de seu desempenho.

NIU et al., (2008) baseados no histórico operacional de uma unidade de 300 MW, usando dados de potência, pressão, temperatura, quantidade de carvão e outros, realizam uma correlação para estimar o coeficiente entre eles. Baseados nesta correlação realizaram uma mineração de regras de associação quantitativa para encontrar os valores ótimos da operação de uma usina movida a carvão.

MAHMOODIANFARD et al., (2009) usaram árvore de decisão para avaliação da segurança de tensão em sistemas de energia. Os valores das variáveis para determinar o ponto de operação foram obtidos, através de Unidades de Mediação Fasorial, para determinar se este ponto é seguro ou inseguro. Com este conjunto de dados é gerada a árvore de decisão que avalia a segurança de tensão. O método foi testado no sistema IEEE de 39 barras.

MORAIS et al., (2009), descrevem uma visão da mineração de dados aplicada a sistemas de potência. Posteriormente uma classificação de série temporal que representam curtos circuitos em linhas de transmissão é realizada usando árvore de decisão. Usam um

conjunto de dados simulados gerados pelo Alternative Transients Program (ATP), representando 5.500 faltas.

YUNYAN (2009) usa a mineração de regras de associação em um sistema de distribuição, para gerenciamento da demanda, a fim de produzir informações para melhor o planejamento e implementação de uma série de medidas para o fornecimento de energia. O método utiliza a teoria dos Conjuntos Aproximados (*rough set theory*) para simplificar os atributos relevantes e, posteriormente do conjunto de dados simplificado se extrair regras de associação.

YANG et al., (2009) apresentam uma metodologia para aplicação de regras de associação para análise de gases emitidos com a finalidade de detecção de falhas em transformadores de potência. Dentre as melhores regras geradas, uma é escolhida para detectar as falhas. Usam o algoritmo *Apriori* com algumas modificações.

DIAO et al., (2009) fazem uma avaliação *on-line* da segurança de tensão através de árvores de decisão que são automaticamente atualizadas com medidas de PMU (*Phasor Measurement Unit*). Primeiramente a árvore é treinada *off-line* usando dados passados de 24 horas antes da operação. Após estar em uso, as árvores são atualizadas a cada hora, incluindo, se necessário novos atributos para melhorar a robustez do sistema. Este desenvolvimento foi testado no sistema da *American Electric Power*, com mais de 2.400 barras ???, 100 geradores e 2.400 linhas com níveis de tensão variando de 4 a 765 kV. Para gerar as condições de operação foram usados os *softwares Powerflow & Short-circuit Analysis Tool (PSAT)* e o *Voltage Security Assessment Tool VSAT*, ambos componentes do *Dynamic Security Analysis (DSA)*.

DAS & NAGENDRA, (2009) usando dados reais do sistema elétrico do sul da Índia, analisam esses dados com técnicas de estatísticas para encontrar relacionamentos entre eles e compreender melhor o sistema elétrico. A análise pode ser vista como uma intenção de mineração de dados. A tensão, a frequência, a potência ativa, a potência reativa e a demanda do sistema são analisadas através de gráficos correlacionando as variáveis. Os dados foram coletados do sistema SCADA do *Southern Regional Load Dispatch Center*.

ZHANG et al., (2009) propuseram um novo método de detecção de faltas baseado em mineração de dados. Usando análise de grupos tentam descobrir rapidamente os componentes que falharam e a seção da falta e com isso detectá-las rapidamente. O método foi testado no sistema IEEE de 9 barras.

VALE et al., (2009) trabalhando com dados reais de um sistema de distribuição, apresentam dois exemplos de aplicação da mineração de dados. No primeiro caso, usando

agrupamentos e com dados dos consumidores tentam descobrir o perfil destes para responder melhor as suas demandas. No segundo caso, usando um sistema multi agente, tentam descobrir melhores modelos e estratégias para o mercado de eletricidade.

KAMWA et al., (2009) propuseram um classificador baseado em regras *fuzzy* para rapidamente avaliar a estabilidade do sistema *Hydro-Quebec*. Esse classificador é construído com árvore de decisão. Usando PMU coletam os dados que são processados nos domínios do tempo e frequência para extrair as medidas que serão à base da classificação, tais como: diferença angular, tensão, frequência.

MEJÍA-LAVALLE et al., (2009) apresentaram cinco experiências de mineração de dados. Na primeira realizaram diagnóstico de falhas nos geradores usando um sistema especialista e uma rede neural. Na segunda, realizam a previsão de descargas em isoladores de alta tensão usando os algoritmos ID3 e vizinho mais próximo. Na terceira experiência, usando árvore de decisão tentam descobrir relacionamentos entre os dados das usinas termoeletricas e hidroelétricas de uma empresa mexicana. Na quarta experiência, procuram descobrir usuários que usam a energia de forma ilícita e finalmente apresentam um sistema de diagnóstico para uma planta de energia baseada em combustível fóssil.

LI et al., (2009) sugerem um modelo de mineração de dados para ser utilizado na exploração e produção de gás e óleo, com o objetivo de descobrir conhecimentos para beneficiar estas produções. Nos testes do modelo foram utilizados quatro conjuntos de dados: dados petrofísicos, registros de dados históricos da exploração e produção, dados sísmicos e dados geológicos.

LI & WU, (2009) descrevem uma aplicação de mineração de dados na avaliação da segurança de sistema de potência. Usando um banco de dados de conhecimento que abrange todas as possíveis condições pré-faltas, regras de decisão sob a forma de árvores hierárquicas são desenvolvidas para a avaliação. Além disso, unidades de medição fasorial (PMU) são levadas em consideração para um melhor desempenho da árvore. O sistema sob estudo com dados reais é o sistema de potência da província de Zhejiang, na China.

FILHO et al., (2009) usam técnicas de mineração de dados para prever o preço da energia elétrica no mercado brasileiro. O preço da energia é a variável dependente e dados mensais de séries de variáveis como carga, energia armazenada e geração térmica são as variáveis independentes. Inicialmente, estes dados são agrupados por atributos semelhantes e, posteriormente, um algoritmo de árvore de decisão é aplicado para extrair as regras e assim identificar os atributos que mais influenciam o preço da energia em curto prazo.

JUNJIE & QUNLI, (2009) usam a correlação de dados para obter os valores ótimos dos parâmetros de controle de uma caldeira para operar uma unidade termoelétrica de 300MW. Usando regras de associação determinam os valores ótimos que podem melhorar a eficiência na unidade, diminuir a taxa de consumo de carvão e minimizar a emissão de gases poluentes.

WEI-HONG et al., (2009) usam regras de associação para determinar as variáveis que influenciam o consumo de energia elétrica nas residências urbanas e assim poderem ter informações para o planejamento do consumo de energia dessas residências. O método foi testado usando quinze variáveis relacionadas ao consumo de energia, do período de 1992 a 2006, de nove cidades da China.

SUN et al., (2009) propõe um novo método para detecção de falhas em unidades turbina-gerador, usando a teoria dos conjuntos aproximados. Uma tabela com informações de vibrações para todas as condições de falhas é montada usando a teoria dos conjuntos aproximados. Posteriormente esta tabela é reduzida usando algoritmo genético e finalmente um conjunto de regras de decisão é gerado da tabela reduzida. O método foi testado com 300 casos de falhas de vibração de uma unidade turbina-gerador.

MATOS, (2009) apresenta uma metodologia para detecção e classificação de curtos-circuitos em alimentadores de sistemas de distribuição de Energia Elétrica. Esta metodologia está baseada na análise de registros oscilográficos através da DFT (*Discrete Fourier Transform*) e de RNAs (Redes Neurais Artificiais). A caracterização de cada tipo de curto-circuito é obtida mediante a análise do comportamento dinâmico das correntes nas três fases durante o período transitório das faltas. A detecção e classificação dos curtos-circuitos são efetuadas por meio de uma rede neural artificial de múltiplas camadas. Um modelo de um alimentador real de Sistema de Distribuição de grande porte, composto por 836 barras, foi utilizado na obtenção dos dados referentes aos curtos-circuitos, com simulações via software ATP-EMTP (*Alternative Transients Program - Electromagnetic Transients Program*).

PIRES, (2009) propõe uma metodologia para classificação de faltas em linhas de transmissão. Apresenta a análise das variações de tensão de curta duração (VTCD) oriundas de faltas do tipo curto-circuito. Foram empregadas técnicas e tarefas de mineração de dados, bem como algoritmos de inteligência computacional. Simularam-se formas de onda de tensão e corrente geradas através do software *Alternative Transient Program* baseado modelo de uma linha de transmissão real, do Sistema Tramoeste, da empresa ELETRONORTE.

BAOLING et al., (2010) usando dados operacionais de uma usina térmica de geração de energia e o algoritmo de agrupamento *k-means*, propõem um método para analisar as

condições operacionais da usina, bem como, encontrar os valores ótimos de algumas variáveis para direcionar a operação da usina. A taxa líquida de consumo de carvão, eficiência da caldeira, a temperatura ambiente e a potência elétrica, são as principais variáveis usadas. O método foi testado em uma unidade de 600MW na província de Jiangxi na China.

KAMWA et al., (2010) discutem um sistema de manutenção baseado na condição para unidades hidrogeradoras e seus ambientes de operação. Mostram as principais características que estes sistemas devem ter. Afirmam que existem problemas e sugerem pesquisas para resolvê-los estando à mineração de dados dentre estas.

GENC et al., (2010) mostram um método baseado em árvores de decisão para determinar as regiões de segurança e os seus limites para um sistema de energia, afim de prever o estado do sistema e para fornecer orientações para as ações necessárias de controle preventivo ou corretivo contra instabilidade transitória. Os testes com sucesso foram realizados no sistema *Entergy* que é uma parte da interligação oeste da América do Norte.

TRONCHONI et al., (2010) mostram o uso de uma Rede Bayseana para auxiliar no diagnóstico de desempenho das redes elétricas, promovendo uma identificação indireta de causas de desligamentos forçados. Para o desenvolvimento, foi utilizada uma base de dados real, na qual são armazenados milhares de ocorrências vinculadas a desligamentos não programados na rede de distribuição de energia.

DIAO et al., (2010) mostram o uso de unidade de medição fasorial e árvore de decisão a fim de desenvolver uma ferramenta em tempo real para avaliação de segurança de sistemas de energia. Quatro condições de pós-contingência são avaliadas para garantir a segurança, as quais são: violação do módulo de tensão, violação do limite térmico das linhas, estabilidade de tensão e estabilidade transitória. O esquema proposto foi testado no sistema de energia do Projeto *Salt River* com uma grande quantidade condições operacionais de um dia representativo da operação do referido sistema.

HUANG et al., (2010) mostram o uso de mineração de dados com árvore de regressão para encontrar a relação dinâmica entre os limites de intercâmbio e a sensibilidade desses limites devido à variação de parâmetros para o sistema de energia da *Hydro-Quebec*, com base na informação da topologia. O método gera grandes quantidades de casos com muitos parâmetros e calcula os limites de intercâmbios com base em simulações dinâmicas no domínio do tempo. Esses limites de intercâmbio são utilizados para calcular a sensibilidade do limite para cada variável considerada como um parâmetro.

YANG et al., (2010) sugerem uma estratégia para otimização da combustão de uma caldeira, baseada em técnicas de mineração de dados. Dados da operação em regime

permanente são selecionados e agrupados para classificar a operação. Com mineração de associações *fuzzy* nos dados contínuos levantam-se os valores alvos ótimos de alguns parâmetros para aumentar a eficiência e diminuir a emissão de gases da caldeira.

LIN et al., (2010) propõem um método baseado em árvore de decisão para detectar ilhamento em sistemas de geração distribuídos. Utiliza o algoritmo C4.5 de alta eficiência e desempenho.

TABBAL & LEMOS, (2010) apresentam uma metodologia para determinação de causas de falhas em sistemas de distribuição. Utilizam um classificador *fuzzy* que tem 55 regras de saída para identificar a provável causa do evento. Usaram registros de uma base de dados de uma concessionária com 15.994 registros de pontos de operação.

COMANESCU et al., (2010) baseados em uma base de dados operacional de um ano apresentam os passos para determinar o perfil típico de carga de uma hidroelétrica usando o método de agrupamento hierárquico para classificar o perfil diário da hidroelétrica em grupos coerentes.

NI & WEN-YING, (2010) mostram um sistema de alarme inteligente para auxiliar o pessoal de operação no tratamento desses alarmes. Inicialmente os alarmes são filtrados para retirada de ruídos e de informações não importantes por meio de uma árvore de decisão. Posteriormente, com um sistema baseado em regras avaliam os alarmes e rapidamente descobrem as causas dos problemas. O sistema foi testado em um centro de operação com 30 subestações.

SCHUCH et al., (2010) apresentam o uso de mineração de dados em uma subestação de energia elétrica. O objetivo é buscar encontrar relações entre os dados que permitam identificar tendências entre eles, usando associação e árvore de decisão. Foram usados 4.777 registros de dados e o software utilizado foi o WEKA.

XU & WANG (2010), usando uma rede *neuro-fuzzy* propõem um método para detecção de falhas em um hidrogenador. As variáveis do problema são: a temperatura do estator, a temperatura de entrada do sistema de refrigeração e a temperatura de saída do sistema de refrigeração. Como saídas usam três possibilidades de faltas: falha do ventilador devido a curto-circuito na refrigeração, falha de ventilação devido a sobrecarga na refrigeração e falha de ventilação devido a sujeira no sistema de resfriamento. Testaram o sistema na companhia FENGMAN na província chinesa de Jilin.

NIRKHI, (2010) em seu artigo apresenta uma descrição das principais tarefas de mineração de dados e mostra como as redes neurais artificiais podem ser potencialmente

usadas como técnica de mineração de dados. Não apresenta nenhuma aplicação, ficando no nível de teoria.

WENHUI et al., (2010) apresenta um sistema de avaliação dos estados operacionais de um transformador usando mineração de dados e teoria *fuzzy*. Três tipos de dados foram usados no sistema: dados de alta-pressão, de óleo e de históricos da operação. Esses dados são usados para determinar quatro estados possíveis do transformador: normal, atenção, subnormal e grave.

CAI et al., (2010) com o intuito de ajudar os engenheiros na seleção de dados em grandes bancos de dados para identificar falhas em um sistema de distribuição, propõem dois métodos de seleção importantes: hipótese e teste, e regressão e apresentam mais dois: seleção passo a passo pelo Critério de Informação Akaike e LASSO (*Least Absolute Shrinkage and Selection Operator*). Esses quatro métodos são comparados em termos de requisitos de seus modelos, suposições, dados e custo computacional. Os testes com dados reais foram realizado na *Progress Energy Carolinas*, na Carolina do Norte nos USA..

ZHENG et al., (2010) fazem uma descrição geral da metodologia de manutenção baseada na condição (CBM), para as unidades hidrogeradores, usadas na companhia chinesa de desenvolvimento do projeto de *Three Gorges*. Mostram que a mineração de dados é considerada como uma forte linha de pesquisa a ser utilizada na CBM.

MORETO & ROLIM, (2011) mostram um método para automaticamente classificar o estado de unidades geradoras. Usam dados de registradores digitais de faltas em correlação com dados do sistema SCADA. Com os dados de pré-faltas e pós-faltas obtidos dos registradores digitais de perturbações, definem características que serão premissas para designar um fato. Este fato será premissa para designar o estado das unidades geradoras. O esquema proposto foi testado em uma usina térmica de geração no Brasil.

BERNARDES et al., (2011) sugerem um método para avaliar em tempo real a segurança estática de sistemas de potência. O método com uma única árvore de decisão avalia a violação do módulo de tensão, a violação dos limites térmicos das linhas e a estabilidade de tensão. Contingências N-1 e N-2 foram consideradas na geração da base de dados para treino e teste da árvore de decisão. O método foi testado em um sistema com 9 barras e 3 geradores síncronos.

DEVARAJ et al., (2011) mostram um método em tempo real para avaliação da segurança da tensão baseado em rede neural artificial. Usam técnicas para reduzir a dimensionalidade dos dados e melhorar a eficiência da rede. O método foi testado no sistema IEEE-30 barras e em um sistema na Índia com 76 barras em várias condições operacionais.

XU et al., (2011) propõem um método para avaliação em tempo real da estabilidade transitória em sistemas de potência usando a teoria das máquinas de aprendizado extremo, uma rede neural de rápido aprendizado, apresentada em HUANG et al., 2004b. O método melhora a velocidade do aprendizado e permite uma efetiva atualização em tempo real. O método foi testado no sistema *New England* de 39 barras.

HAIDAR et al., (2011) propõem uma rede neural com regressão generalizada (GRNN) como base para avaliação da estabilidade transitória em sistemas de potência. No método proposto, conjuntos de dados de aprendizagem foram gerados através da simulação no domínio do tempo. As entradas para a GRNN representam o módulo de tensão para todas as barras e as potências ativas e reativas das linhas de transmissão, a saída representa o índice de estabilidade transitória. O método proposto foi implementado e testado em um sistema de 9 barras e em um sistema de 39 barras. Os resultados mostraram que a estabilidade do sistema de potência pode ser prevista com alta precisão e menor taxa de erro de classificação.

WAHAB et al., (2011) apresentam uma avaliação da estabilidade transitória para um sistema de 87 barras utilizando uma rede neural probabilística (PNN), com incorporação de métodos de seleção e extração de atributos. O sistema de potência investigado é dividido em subsistemas menores para reduzir a quantidade de dados a serem tratados. A estabilidade transitória do sistema é primeiramente determinada com base nos ângulos do rotor dos geradores obtidos a partir de simulações no domínio do tempo realizadas, considerando faltas trifásicas com condições de carga diferentes. Os dados gerados nas simulações são utilizados como entradas para a PNN. Técnicas de redução da dimensionalidade são incorporadas para reduzir o número de atributos para a PNN que é utilizada como um classificador para determinar se o sistema de potência é estável ou instável. A incorporação de técnicas de redução da dimensionalidade reduziu o tempo necessário para o treino da rede neural sem afetar a precisão dos resultados da classificação.

KRISHNAN et al., (2011) desenvolveram uma estratégia eficiente para maximizar a informação do banco de dados e minimizar os requerimentos computacionais de treinamento de uma árvore de decisão para a avaliação da estabilidade de tensão. O método usando sensibilidade linear baseado no hipercubo Latino gera uma base de dados eficiente que é utilizada para gerar uma árvore de decisão para a extração de regras na avaliação da estabilidade de tensão. Os testes foram realizados na França, região da Brittany. Os resultados mostram que a árvore de decisão gerada melhora o desempenho da classificação e oferece benefícios econômicos quando comparada com técnicas de amostragem simples e reduz o tempo de computação requerido.

KARAMI, A., (2011) propõem uma metodologia para estimar a margem de estabilidade transitória de um sistema de potência, utilizando rede neural MLP com um rápido treinamento. Técnicas de simulações no domínio do tempo geram o banco de dados de treino. O método foi testado no sistema *New England* com 10 máquinas e 39 barras e demonstrou ser uma ferramenta rápida e precisa para avaliar a estabilidade transitória em tempo real com uma precisão aceitável. Em adição, examinando os ângulos dos rotores dos geradores após as faltas, seleciona as condições do sistema de potência que mais afetam a margem de estabilidade para cada falta.

LIN, (2011) mostra um método de controle preventivo para a estabilidade transitória que leva em conta a compensação em série. A compensação série refere-se à modificação da impedância da linha pela inserção de capacitores em série com a linha. A compensação série pode conseguir controlar o fluxo de potência, melhorar a estabilidade transitória e amortecer as oscilações. O método proposto extrai regras "*IF-THEN*" de uma rede neural MultiLayer Perceptron (MLP).

Como pôde ser observado, a revisão bibliográfica aqui apresentada, envolveu somente uma pequena quantidade de artigos aplicando a mineração de dados com sistema SCADA em usinas hidrelétricas. A grande maioria aplica a mineração de dados em outros problemas como: segurança de tensão, estabilidade transitória, detecção de faltas, avaliação de estados operacionais, manutenção baseada na condição, tratamento de alarmes, levantamento de perfil de carga e em outros problemas encontrados em sistemas elétricos de potência.

## **1.4 Estrutura da Tese**

Além do capítulo 1, que apresentou uma descrição do problema, as contribuições tecnológicas geradas e uma revisão bibliográfica sobre o tema, mostrando a aplicação da mineração de dados em sistemas elétricos de potência, esta Tese de Doutorado tem 4 capítulos adicionais, organizados do seguinte modo:

No capítulo 2 é apresentada a fundamentação teórica sobre Usinas Hidrelétricas e Sistemas SCADA, com o objetivo de apresentar os principais componentes da usina e as características e funcionalidades do sistema SCADA.

No capítulo 3 é apresentada a fundamentação teórica sobre Descoberta do Conhecimento em Bancos de Dados e Mineração de Dados mostrando os principais aspectos relacionados a estes assuntos.

No capítulo 4 são mostradas as experiências realizadas sobre os dados reais obtidos através de um sistema SCADA de uma usina hidrelétrica brasileira, para a extração de informações relevantes, através de análises estatísticas e tarefas de mineração de dados como: associação e classificação. Apresentam-se também a construção de uma árvore de decisão baseada na curva de capacidade e uma interface desenvolvida com base nesta curva que poderá ser aplicada em tempo real para avaliar a segurança do hidrogerador.

Finalmente, no capítulo 5, apresentam-se as conclusões gerais do trabalho e as possíveis linhas de atuação para trabalhos futuros.

## CAPÍTULO 2

### Usinas Hidrelétricas e Sistemas SCADA

#### 2.1 Introdução

Este capítulo descreverá de forma abrangente os principais componentes das usinas hidrelétricas, assim como as principais características dos sistemas SCADA.

#### 2.2 Usinas Hidrelétricas

Uma UHE é um conjunto de edificações e de equipamentos, construído para gerar energia elétrica usando o aproveitamento do potencial hidráulico existente em um rio. As UHEs geram impactos ambientais como o alagamento de áreas, aumento no nível dos rios, mudança climática e mesmo que minimamente prejudicam a fauna e a flora da região. Ainda assim, a energia gerada por essas usinas é mais barata do que outras, como a energia nuclear e menos agressiva ambientalmente do que a do petróleo ou a do carvão. A energia hidráulica é convertida em energia mecânica por meio de uma turbina hidráulica, que por sua vez é convertida em energia elétrica por meio de um gerador, sendo a energia elétrica transmitida para uma ou mais linhas de transmissão para os grandes consumidores e para as distribuidoras de energia.

Essas usinas são sistemas bastantes complexos formados por sistemas menores que conjuntamente realizam a geração de energia. Além dos principais componentes da usina como: o vertedouro, a casa de força, a admissão, as linhas de transmissão e outros, existem o conjunto turbina/gerador, os sistemas de automação e proteção, os reguladores de tensão e de velocidade e muitos outros diferentes sistemas eletromecânicos.

Todos esses sistemas devem receber permanente observação por parte das equipes de operação e manutenção das usinas. Os principais componentes destas unidades são analisados através das variáveis discretas e analógicas que representam respectivamente os estados dos equipamentos e o comportamento das grandezas físicas relacionadas. Em operação normal da usina as informações sobre as várias variáveis permitem o acompanhamento da sua operação e a avaliação do seu desempenho, bem como, a qualidade da energia produzida. Em condições de contingências os registros dessas variáveis possibilitam a análise dos problemas ocorridos.

Assim, todos os sistemas envolvidos na geração da energia devem ser monitorados para fornecer informações às equipes de planejamento, operação e manutenção. Dessa forma as equipes podem realizar uma análise do processo de geração e com isso avaliar a qualidade de seu produto, aumentar a disponibilidade da geração, reduzir os custos e planejar melhor as ações futuras. A figura 2.1 mostra, de forma esquemática, os principais componentes que compõem uma UHE.

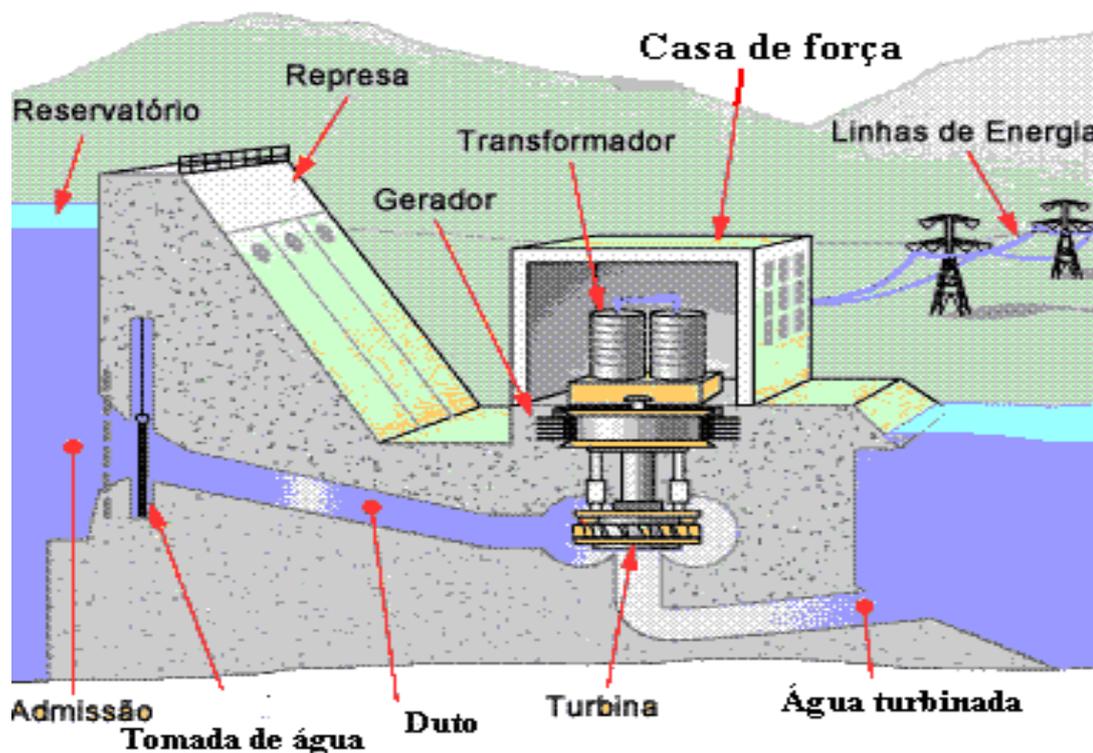


Figura 2.1 – Principais componentes de uma usina hidrelétrica

### 2.2.1 Unidades Hidrogeradoras

O hidrogerador tem como elemento principal o conjunto turbina-gerador que é um dos principais componentes de uma UHE. É neste conjunto que ocorre a transformação da energia hidráulica em energia elétrica. A energia primária de uma hidrelétrica é a energia potencial gravitacional da água contida numa represa elevada. Antes de se tornar energia elétrica, a energia primária deve ser convertida em energia cinética de rotação. O dispositivo que realiza essa transformação é a turbina. O último elemento dessa cadeia de transformações é o gerador, que converte o movimento rotatório da turbina em energia elétrica.

## 2.2.2 Turbina

As turbinas são projetadas especificamente para transformar a energia hidráulica de um fluxo de água em energia mecânica na forma de torque e velocidade de rotação. Em toda turbina hidráulica a água entra vindo do reservatório em nível mais elevado e escapa para um canal de nível mais baixo. A água de entrada é levada através de um duto fechado até um conjunto de lâminas curvas que transferem a energia da água para um rotor. A água que sai da turbina é conduzida por um duto, o tubo de sucção, para o canal ou reservatório na jusante.

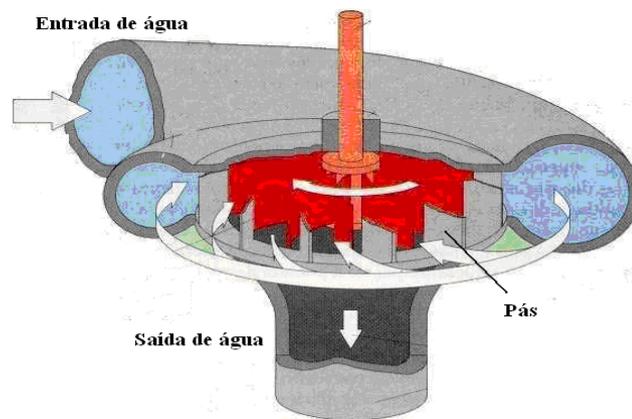


Figura 2.2 – Turbina hidráulica

## 2.2.3 Gerador Elétrico

O gerador elétrico é um dispositivo utilizado para a conversão da energia mecânica proveniente de uma turbina hidráulica em energia elétrica, fornecem corrente alternada e são constituídos de duas partes principais: o rotor, e o estator, como ilustrado na figura 2.3.

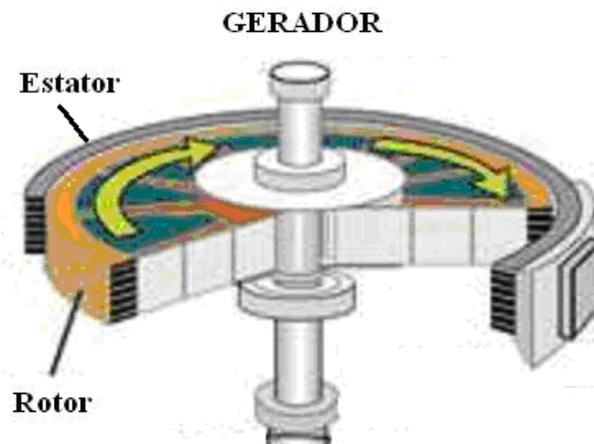


Figura 2.3 – Gerador elétrico de hidrogeradores

O rotor é a parte girante da máquina, constituído por um material ferromagnético envolto num enrolamento designado como enrolamento de campo, que tem como função produzir um campo magnético constante para interagir com o campo produzido pelo enrolamento do estator. O estator é a parte fixa do gerador, montado em volta do rotor de forma que o mesmo possa girar no seu interior, também constituído de um material ferromagnético envolto num conjunto de enrolamentos distribuídos ao longo da sua circunferência.

#### **2.2.4 Casa de Força**

É o local onde está instalado o conjunto turbina-gerador que é o principal equipamento em uma UHE. A definição das principais dimensões da casa de força depende da quantidade e dimensões básicas da turbina e do gerador. As principais elevações da casa de força são definidas em função dos níveis de água notáveis de jusante e da submergência da turbina. A definição dessas elevações é de extrema importância para a localização dos demais equipamentos, como por exemplo, a cota do piso dos transformadores. Esse piso (cota), evidentemente, deve estar a salvo de inundação.

Deverá ser definida, ainda, a cota de fundação da casa de força, que depende da posição do tubo de sucção da turbina. Em seguida, definem-se as cotas e a disposição das galerias de drenagem. As unidades terminais remotas e os centros de controle da UHE normalmente estão localizados na casa de força.

### **2.3 Sistemas SCADA**

Um sistema SCADA (*Supervisory Control and Data Acquisition*) é um sistema que realiza aquisição de dados de um processo, supervisiona esses dados e permite ao operador controlar ou tomar decisões baseados nesses valores, bem como, tomar decisões automaticamente. Existentes desde a metade do século XX, os primeiros sistemas eram simplesmente sistemas de telemetria. Sensores eram colocados em campo e ligados diretamente a medidores onde os operadores poderiam acompanhar em tempo real os valores atuais dessas medidas e tomar decisões baseado nelas (MONTEIRO et al., 2004). No início apenas permitiam informar periodicamente sobre o estado do processo. Atualmente os sistemas SCADA estão evoluindo para sistemas abertos e com uma arquitetura fortemente centrada em conectividade com as *intranets* corporativas e em consequência, com a própria *Internet*. Estes sistemas de *software* são então responsáveis pela coleta de dados em ambientes complexos, pela sua interpretação e respectiva apresentação de modo amigável para o utilizador usando como

recurso as interfaces Homem-Máquina. Os sistemas SCADA melhoram a eficiência da supervisão e controle, disponibilizando em tempo real o estado atual do processo, através de um conjunto de previsões, gráficos e relatórios, de modo a permitir a tomada de decisões operacionais, quer automaticamente, quer por iniciativa do operador. Desta forma, estes sistemas deixaram de ser vistos como meras ferramentas operacionais, ou de engenharia, e passaram a ser considerados como uma importante fonte de informação de crucial importância na estrutura de gestão das empresas.

### 2.3.1 Arquitetura de um Sistema SCADA

A arquitetura básica de um sistema SCADA é mostrada na figura 2.4.

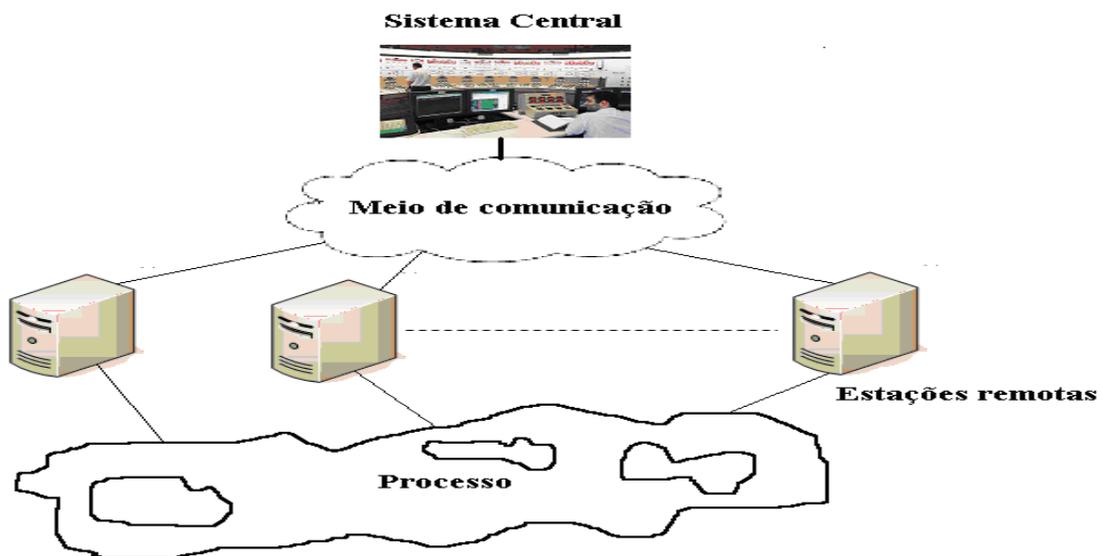


Figura 2.4 - Arquitetura básica de um sistema SCADA

Nesta arquitetura, os componentes básicos de um sistema de supervisão podem ser resumidos, de forma simplificada em: sistema central, meio de comunicação, estações remotas e processo. O sistema central concentra todas as informações do processo recebidas das estações remotas, apresentando-as através de vídeos, painéis, etc. O meio de comunicação interliga o sistema central as várias estações remotas. Estas por sua vez, através de sensores e atuadores interagem com o processo, que é na realidade a instalação física a ser controlada e supervisionada. Os sensores são dispositivos conectados aos equipamentos controlados que convertem parâmetros físicos tais como velocidade, nível de água e temperatura para sinais analógicos e digitais legíveis pela estação remota. Os atuadores são utilizados para atuar sobre o sistema, ligando e desligando, ajustando e regulando os equipamentos.

O processo de controle e aquisição de dados se inicia nas estações remotas, podendo estas ser: CLP (Controladores Lógicos Programáveis) ou UTR (Unidades Terminais Remotas), que adquirem os valores das variáveis, realizam cálculos ou controles e atuam no processo se necessário.

O meio de comunicação é a plataforma por onde a informação flui dos CLP/UTR para o sistema central, levando em consideração os requisitos do sistema e a distância a cobrir, podendo ser implementada através de cabos, redes, fibras ópticas, rádio e etc.

### 2.3.2 Funcionalidades de um Sistema SCADA

Os sistemas SCADA dispõem de três funções independentes e que podem ser utilizadas isoladamente ou em simultâneo.

- A supervisão, que permite o monitoramento de forma contínua do estado real do processo, através de telas gráficas, gráficos de tendência, variáveis analógicas e digitais, alarmes, relatórios, históricos entre outros;
- A operação, responsável pelo ajuste dos parâmetros do processo por parte do responsável do mesmo;
- E finalmente o controle automático, que em função dos parâmetros de entrada (valores de sensores, variáveis, etc), determina o ajuste a fazer no processo ou na operação do sistema.

Atualmente estes sistemas são amplamente utilizados em segmentos do comércio e da indústria, tendo também grande aplicação nas empresas do setor elétrico, na supervisão e controle da geração, transmissão e distribuição da energia elétrica.

Um sistema SCADA gerencia um certo número de variáveis. Muitas destas são estados do processo com seus valores provenientes do campo e outras são variáveis calculadas a partir das anteriores. Assim, basicamente existem dois tipos de variáveis no sistema: as primitivas ou simples e as compostas. As primitivas normalmente são lidas diretamente do processo e são utilizadas sem nenhum ou quase nenhum tratamento; e as compostas que são formadas a partir das primitivas, normalmente por meio de algum tratamento, ou combinação de duas ou mais variáveis simples.

A seguir se mostra as variáveis simples e compostas relativas a um sistema de geração de energia, como por exemplo, uma hidrelétrica.

### 2.3.3 Variáveis Simples

As variáveis simples são variáveis analógicas e discretas adquiridas do campo. Uma variável analógica descreve uma grandeza analógica como tensão, corrente, potência, temperatura, pressão, entre outras. Alguns critérios são aplicados a essas variáveis e associados no sistema de forma a facilitar a sua identificação, dentre eles tem-se: descritivo da variável, unidade de medida, último valor, limites inferiores, superiores e intermediários da variável.

Uma variável discreta é uma variável expressa por somente um bit. Ou seja, 0 ou 1, normalmente utilizado para representar o estado de um equipamento, como por exemplo: indicar estado de aberto ou fechado, ligado ou desligado, atuado ou desatuado e outros.

### 2.3.4 Variáveis Compostas

São aquelas formadas a partir das variáveis simples. Dentre estas, destacam-se:

**Totalizador:** soma os valores obtidos de uma variável analógica ou contabiliza o total de pulsos de uma variável digital.

**Equipamento:** indica um equipamento com várias variáveis que sinalizam seu estado: se pronto para funcionar, ligado ou desligado, modo de operação, chave local/remoto, entre outros.

**Seqüência ou grupo:** É um conjunto de equipamentos inter-travados que podem ser considerados como um único equipamento, pois para o funcionamento correto do processo é necessário que todos esses equipamentos funcionem de forma conjunta como, por exemplo, uma esteira movida por vários motores, pois se todos os motores não estiverem funcionando em conjunto a esteira não irá funcionar adequadamente.

**Cálculo:** Muitos valores não podem ser medidos diretamente no campo ou, em alguns casos, não justificam o investimento em sensores para a obtenção dessas medidas. Então é possível, utilizando-se de outras medidas, calcular a medida desejada, por exemplo, o cálculo da potência elétrica, a partir das medidas de tensão e corrente obtidas em campo.

Baseado nas variáveis simples e compostas pode-se considerar como funcionalidades, a capacidade de gestão de arquivos e banco de dados, a visualização, o tratamento de alarmes e eventos, a elaboração de históricos e relatórios, a elaboração de curvas de tendências, receitas e tratamentos estatísticos.

### 2.3.5 Gestão de Arquivos e Banco de Dados

A capacidade de gerenciar arquivos é em princípio um atributo dos sistemas de supervisão. As funcionalidades oferecidas permitem realizar as operações habituais com os arquivos, ou seja, visualizar o conteúdo, imprimir, copiar, remover ou modificar o seu nome. Permitem ainda visualizar o conteúdo de um diretório. Esta funcionalidade é imprescindível na concretização de um sistema de supervisão. Ela é utilizada para duas finalidades básicas, materializadas na construção de históricos dos processos e na atualização das interfaces gráficas. A capacidade de consulta deve ser disponibilizada.

### 2.3.6 Visualização

Esta funcionalidade consiste na visualização dos valores de determinadas variáveis em tempo real em vídeos e quadro sinóticos. Permite a apresentação de diagramas unifilares das partes do processo. Os vídeos e os quadros sinóticos constituem os meios mais eficazes de apresentação dos dados dos processos supervisionados. Através do vídeo e do quadro sinótico pode-se captar rapidamente informações relevantes do sistema. Os estados dos equipamentos e dispositivos do processo, como válvulas, interruptores, bombas, tanques, canalizações, disjuntores e outros são disponibilizados de forma gráfica. Exemplo dessas visualizações é apresentado na figura 2.5.

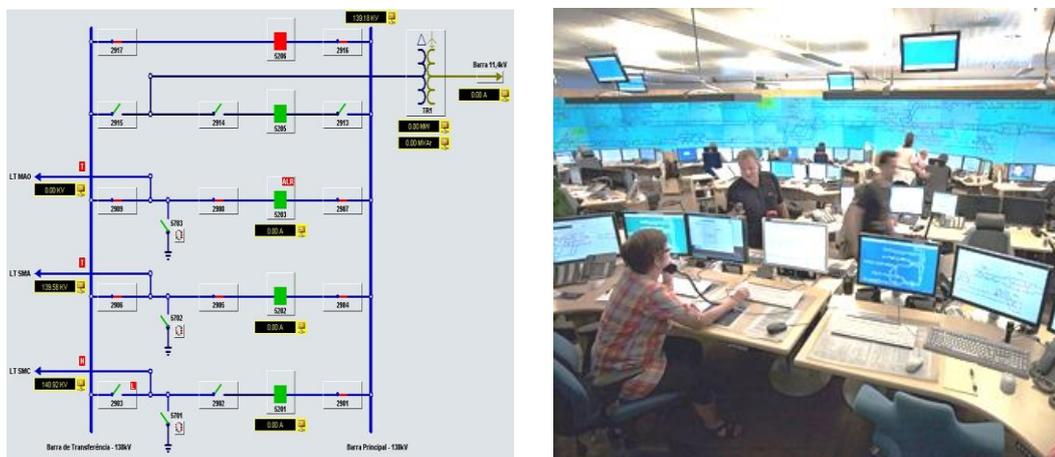
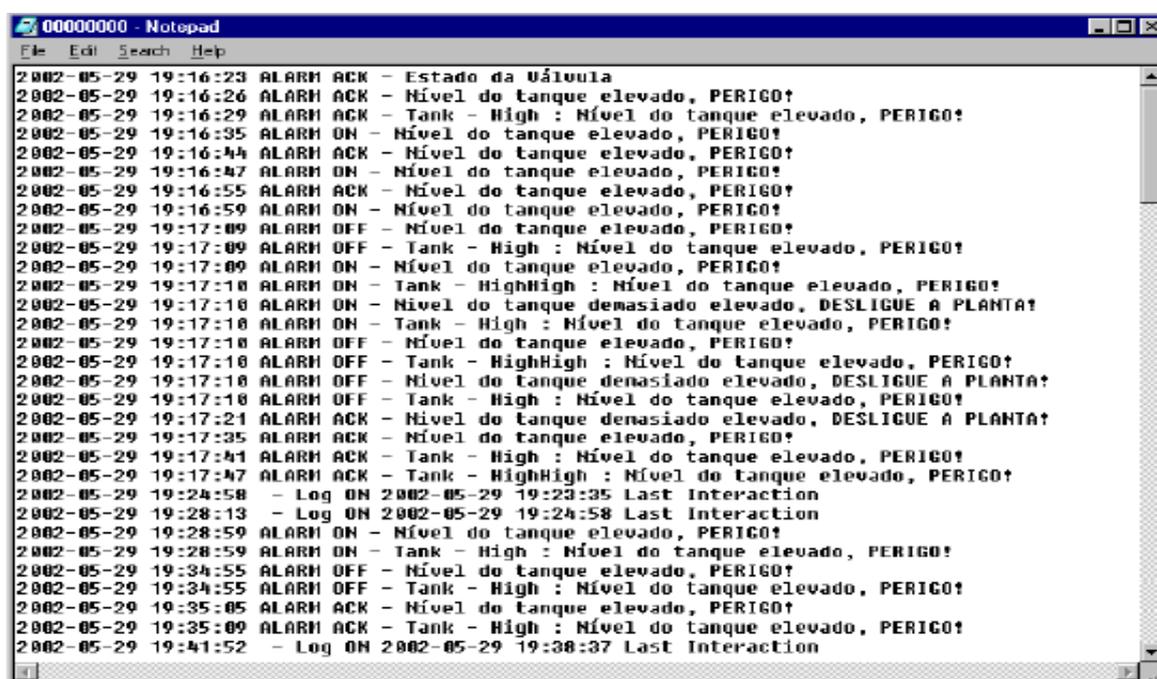


Figura 2.5 - Exemplo de um unifilar em vídeo e um sinótico

### 2.3.7 Alarmes e Eventos

Toda mudança de estado do processo deve ser registrada. A utilização de alarmes e eventos para detecção e aviso de situações anormais é uma funcionalidade disponível em todos os sistemas SCADA. A configuração dos alarmes e dos eventos segue sempre o mesmo princípio, ou seja, é necessário referenciar a variável que será monitorada e indicar as condições ativadoras do alarme. As condições mais comuns de ativação consistem nas igualdades e desigualdades matemáticas, no valor booleano e na alteração de valor. Outras condições como ultrapassagem de limites, mudanças bruscas são também determinantes. Alguns sistemas permitem atribuir prioridades aos alarmes. Exemplo de uma tela de eventos e alarmes é mostrado na figura 2.6..



```

00000000 - Notepad
File Edit Search Help
2002-05-29 19:16:23 ALARM ACK - Estado da U lvula
2002-05-29 19:16:26 ALARM ACK - N vel do tanque elevado, PERIGO!
2002-05-29 19:16:29 ALARM ACK - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:16:35 ALARM ON - N vel do tanque elevado, PERIGO!
2002-05-29 19:16:44 ALARM ACK - N vel do tanque elevado, PERIGO!
2002-05-29 19:16:47 ALARM ON - N vel do tanque elevado, PERIGO!
2002-05-29 19:16:55 ALARM ACK - N vel do tanque elevado, PERIGO!
2002-05-29 19:16:59 ALARM ON - N vel do tanque elevado, PERIGO!
2002-05-29 19:17:09 ALARM OFF - N vel do tanque elevado, PERIGO!
2002-05-29 19:17:09 ALARM OFF - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:09 ALARM ON - N vel do tanque elevado, PERIGO!
2002-05-29 19:17:10 ALARM ON - Tank - HighHigh : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:10 ALARM ON - N vel do tanque demasiado elevado, DESLIGUE A PLANTA!
2002-05-29 19:17:10 ALARM ON - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:10 ALARM OFF - N vel do tanque elevado, PERIGO!
2002-05-29 19:17:10 ALARM OFF - Tank - HighHigh : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:10 ALARM OFF - N vel do tanque demasiado elevado, DESLIGUE A PLANTA!
2002-05-29 19:17:10 ALARM OFF - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:21 ALARM ACK - N vel do tanque demasiado elevado, DESLIGUE A PLANTA!
2002-05-29 19:17:35 ALARM OFF - N vel do tanque elevado, PERIGO!
2002-05-29 19:17:41 ALARM ACK - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:17:47 ALARM ACK - Tank - HighHigh : N vel do tanque elevado, PERIGO!
2002-05-29 19:24:58 - Log ON 2002-05-29 19:23:35 Last Interaction
2002-05-29 19:28:13 - Log ON 2002-05-29 19:24:58 Last Interaction
2002-05-29 19:28:59 ALARM ON - N vel do tanque elevado, PERIGO!
2002-05-29 19:28:59 ALARM ON - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:34:55 ALARM OFF - N vel do tanque elevado, PERIGO!
2002-05-29 19:34:55 ALARM OFF - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:35:05 ALARM ACK - N vel do tanque elevado, PERIGO!
2002-05-29 19:35:09 ALARM ACK - Tank - High : N vel do tanque elevado, PERIGO!
2002-05-29 19:41:52 - Log ON 2002-05-29 19:38:37 Last Interaction

```

Figura 2.6 - Exemplo de alarmes e eventos em v deo

### 2.3.8 Hist ricos e Relat rios

A funcionalidade de produ o de relat rios n o   muito relevante. Na verdade consiste em disponibilizar uma forma auxiliar de apresenta o dos dados, normalmente mais condensada com base em tabelas ou quadros. Os relat rios podem ser produzidos sob a forma visual no v deo ou impressos na impressora. A figura 2.7 mostra um exemplo de um relat rio de alarmes e eventos.

Date	Time	ms	sec	Key	Label	Desc	Urgency	Type			
17/01/2010	00:01:52	977	1263686512		15GRES_PSLLL305T	G15 AR RVT PRESSAO	BAIXA	-> M_BAIXA	2		BSig
17/01/2010	00:01:53	005	1263686513		14GRES_PSLLL305T	G14 AR RVT PRESSAO	BAIXA	-> M_BAIXA	2		BSig
17/01/2010	00:01:53	074	1263686513		13GRES_PSLLL305T	G13 AR RVT PRESSAO	BAIXA	-> M_BAIXA	2		BSig
17/01/2010	00:03:14	271	1263686594		14GRES_PSLLL305T	G14 AR RVT PRESSAO	M_BAIXA	-> BAIXA	0		BSig
17/01/2010	00:03:14	329	1263686594		13GRES_PSLLL305T	G13 AR RVT PRESSAO	M_BAIXA	-> BAIXA	0		BSig
17/01/2010	00:03:14	348	1263686594		15GRES_PSLLL305T	G15 AR RVT PRESSAO	M_BAIXA	-> BAIXA	0		BSig
17/01/2010	00:03:48	056	1263686628		14GRE_LSN301T	G14 TQ RVT NIVEL OLEO	NORMAL	-> ANORMAL	0		BSig
17/01/2010	00:05:38	356	1263686738		14GRE_LSN301T	G14 TQ RVT NIVEL OLEO	ANORMAL	-> NORMAL	0		BSig
17/01/2010	00:06:11	338	1263686771		17GRE_LSN301T	G17 TQ RVT NIVEL OLEO	NORMAL	-> ANORMAL	0		BSig
17/01/2010	00:09:25	989	1263686965		20GRE_LSH302T	G20 TQ RVT NIVEL OLEO	NORMAL	-> ALTO	2		BSig
17/01/2010	00:09:28	837	1263686968		17GRE_LSN301T	G17 TQ RVT NIVEL OLEO	ANORMAL	-> NORMAL	0		BSig
17/01/2010	00:10:15	500	1263687015		13GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.60 kV			SPt
17/01/2010	00:10:24	280	1263687024		14GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.40 kV			SPt
17/01/2010	00:10:29	790	1263687029		15GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.60 kV			SPt
17/01/2010	00:10:51	900	1263687051		18GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.40 kV			SPt
17/01/2010	00:11:15	100	1263687075		19GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.60 kV			SPt
17/01/2010	00:11:20	030	1263687080		20GTAK01U	SET POINT TENSÃO REF. (COU)	=L2/P2=>	13.40 kV			SPt
17/01/2010	00:11:46	988	1263687106		20GRE_LSH302T	G20 TQ RVT NIVEL OLEO	ALTO	-> NORMAL	0		BSig
17/01/2010	00:15:34	804	1263687334		15GRE_FS302T_OK	G15 VALV. RVT FLUXO DESC. AGUA 2	NORMAL	-> BAIXO	2		BSig
17/01/2010	00:15:40	304	1263687340		15GRE_FS302T_OK	G15 VALV. RVT FLUXO DESC. AGUA 2	BAIXO	-> NORMAL	0		BSig

Figura 2.7 - Exemplo de um relatório de alarmes e eventos

### 2.3.9 Gráficos de Tendência

Os gráficos de tendência são um dos instrumentos mais importantes dos sistemas SCADA. Estes gráficos permitem visualizar a evolução temporal do valor de uma ou varias variáveis, fornecendo uma visão clara da tendência evolutiva do processo. Através da observação destes gráficos é possível detectar em um curto espaço de tempo situações de evolução anormais que requerem medidas corretivas. Esses gráficos de tendência são formados a partir dos dados históricos. Esta funcionalidade pode ser muito útil para efeitos de comparação da evolução atual com a tendência registrada ao longo de um largo espaço de tempo. A seguir na figura 2.8 é mostrado um gráfico de tendência.



Figura 2.8 - Exemplo de um gráfico de tendência

### 2.3.10 Receitas

A funcionalidade de criação e carregamento de receitas é utilizada para automatizar os procedimentos de configuração dos próprios processos. De certa forma pode-se considerar que esta é uma funcionalidade estritamente de controle dos processos. Normalmente, uma receita é carregada num dispositivo antes deste ser posto em execução e funciona como uma inicialização. Uma receita pode consistir, por exemplo, em um programa que será executado ou em um conjunto de valores que indicam a quantidade de cada ingrediente que será utilizada no processo de fabricação. Uma característica importante desta funcionalidade é a possibilidade de carregamento de receitas como resposta a eventos gerados no sistema e não apenas por determinação do usuário. Todos os sistemas SCADA devem oferecer esta funcionalidade.

### 2.3.11 Tratamentos Estatísticos

O tratamento estatístico na verdade é a possibilidade de realizar operações de cálculo estatístico e é determinante para a construção de aplicações de gestão de planejamento da produção e do controle de qualidade, entre outras. A importância desta funcionalidade é evidente nos SCADA, pois a sua utilização implica normalmente em custos adicionais pelo fato de não ser incluída no *software* básico, e ter de ser adquirida separadamente. As indicações que podem ser obtidas através da comparação dos dados estatísticos com padrões previamente estabelecidos servem para orientar e melhorar todo o processo produtivo. As ações de processamento estatístico podem ser efetuadas em tempo real sobre dados do processo ou então sobre dados provenientes de bases de dados históricas. A figura 2.9 mostra um gráfico estatístico de eventos que ocorreram nos hidrogeradores de uma usina hidrelétrica, durante um período específico de operação, evidenciando a grande diversidade do número de ocorrência entre eles.

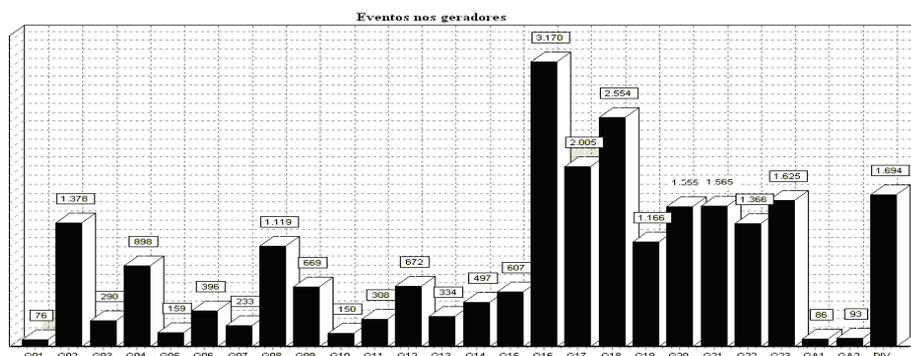


Figura 2.9 - Exemplo de um gráfico estatístico dos eventos

## **2.4 Conclusões**

Neste capítulo apresentou-se as principais características das UHE, enfocando basicamente o hidrogerador com seus componentes e a casa de força. Também, as principais funcionalidades de um sistema SCADA foram apresentadas com a finalidade de situar o leitor quanto ao local e ambiente onde serão aplicadas as tarefas de mineração de dados a serem introduzidas no próximo capítulo.

## CAPÍTULO 3

### Descoberta de Conhecimento e Mineração de Dados

#### 3.1 Introdução

Neste capítulo apresentam-se os fundamentos teóricos da Descoberta de Conhecimento e da Mineração de Dados. Os principais aspectos destes assuntos são explanados, evidenciando as fases da descoberta do conhecimento, as tarefas e as técnicas da mineração de dados.

Os rápidos avanços nas tecnologias de *hardware* e *software* permitiram a geração, coleta e armazenamento de grandes quantidades de dados pelas empresas. Entretanto, muitas vezes estes dados são esquecidos, formando-se grandes bancos de dados que com o passar do tempo terminam sendo apagados. Esses dados necessitam serem analisados, pois, o real valor deles reside na informação que se pode extrair, e a partir desta, a tomada de decisão ou a melhor compreensão dos fenômenos que nos cercam. Uma análise estatística ou uma inspeção nesses dados já permite a descoberta de informações úteis nos dados passados.

Sabe-se atualmente que dentro de vastas quantidades de dados pode-se encontrar conhecimento oculto de grande valia e através de técnicas de mineração de dados pode-se encontrar padrões e relacionamentos entre esses dados, permitindo a criação de modelos e conhecer melhor a realidade.

Os dados compõem o conhecimento, sendo assim, dados são essenciais para qualquer organização e seu acúmulo se tornou relativamente fácil com a tecnologia.

O dado é algo bruto, é a matéria prima da qual se pode extrair informação. Informação é o dado processado, com significado e contexto bem definidos. O computador, em essência, serve para transformar dados em informações. Por fim, o conhecimento é o uso inteligente da informação, é a informação contextualizada e utilizada na prática.

Desta forma, a qualidade da informação sustenta o conhecimento. A figura 3.1 sumariza esses aspectos, interrelacionando-os.

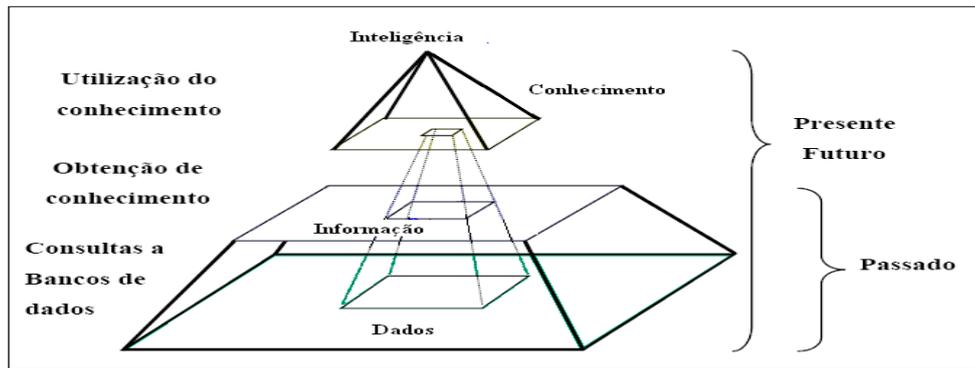


Figura 3.1 - Pirâmide do conhecimento (DIAS, 2002)

### 3.2 Descoberta de Conhecimento

A Descoberta do Conhecimento em Banco de Dados (*Knowledge Discovery in Databases – KDD*) é o processo não trivial de identificar padrões de dados válidos, novos, potencialmente úteis e desconhecidos em grandes bancos de dados (FAYYAD et al., 1996). É a transformação de dados em informações de alto nível para ajudar no processo de tomada de decisões organizacionais através do uso de técnicas automáticas de exploração de grandes quantidades de dados, de forma a descobrir novos padrões e relações, que devido a esse volume não seriam descobertas a olho nu.

O processo KDD é constituído de várias etapas, como ilustrado na figura 3.2, as quais são executadas de forma interativa e iterativa. De acordo com (BRACHNAD & ANAND, 1996), as etapas são interativas porque envolvem a cooperação do responsável pela análise de dados, cujo conhecimento sobre o domínio orientará a execução do processo. Por sua vez, a iteração deve-se ao fato de que, frequentemente esse processo não é executado de forma sequencial, mas envolve repetidas seleções de parâmetros e conjunto de dados, aplicações das técnicas de MD e posterior análise dos resultados obtidos, a fim de refinar os conhecimentos extraídos.

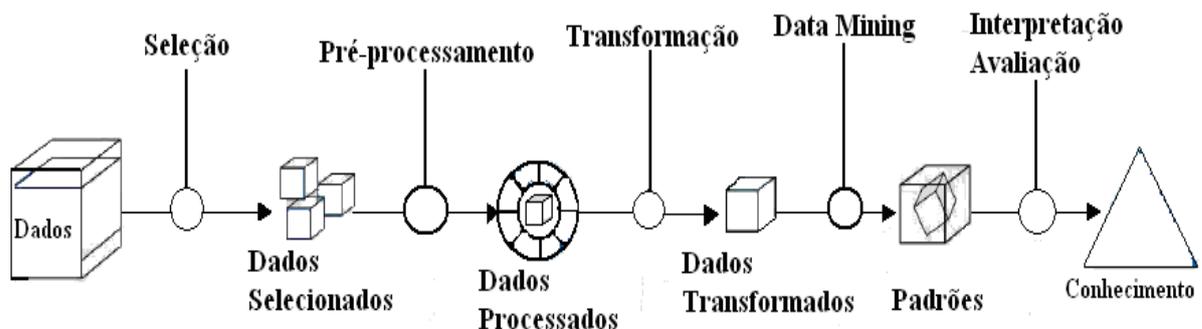


Figura 3.2 - Etapas que constituem o processo KDD

Esse processo tem início com o entendimento do domínio da aplicação e dos objetivos a serem atingidos e é dividido nas etapas mostradas abaixo:

**Seleção:** Os dados são selecionados. Realiza-se um agrupamento organizado da massa de dados alvo da descoberta. Os dados podem ser qualitativos (ou categorizado) e quantitativos (numéricos).

**Pré-processamento:** Como em toda análise quantitativa, a qualidade dos dados é essencial para a obtenção de resultados confiáveis. Portanto, dados limpos e compreensíveis são requisitos básicos para o sucesso da MD, como afirmam (DINIZ & NETO, 2000). A limpeza dos dados é realizada por meio de um pré-processamento, visando assegurar a qualidade dos dados selecionados. Destaca-se que, segundo (MANNILA, 1996), essa etapa pode tomar até 80% do tempo necessário para todo o processo, devido às dificuldades de integração de bases de dados heterogêneas.

**Transformação:** Os dados pré-processados devem passar por outra transformação, que os armazena adequadamente, visando facilitar o uso das técnicas de MD, as quais podem envolver agregação, amostragem, redução da dimensionalidade, seleção de subconjuntos, criação de recursos, discretização, binarização e transformação de variáveis.

**Mineração de dados:** O objetivo principal desse passo é a aplicação de técnicas de mineração nos dados pré-processados, o que envolve ajuste de modelos e/ou determinação de características nos dados. É importante destacar que cada técnica de MD utilizada para conduzir as operações se adapta melhor a alguns problemas do que a outros. Portanto, o sucesso de uma tarefa de MD está diretamente ligado à experiência e à intuição do analista.

**Interpretação e Avaliação (Pós-processamento):** A etapa final do processo de mineração é a interpretação dos padrões descobertos. Assim, a informação extraída é analisada (ou interpretada) em relação ao objetivo proposto, sendo identificadas e apresentadas as melhores informações. Dessa forma, o propósito do resultado não consiste somente em visualizar, gráfica ou logicamente, o rendimento do MD, mas, também, em filtrar a informação que será apresentada, eliminando possíveis ruídos, ou seja, padrões redundantes ou irrelevantes, que podem surgir no processo. A figura 3.3 auto-explicativa, mostra a posição da KDD na solução de um problema.

### Posição da KDD no ciclo de solução de problemas



Figura 3.3 - Posição da KDD na solução de um problema

É importante frisar que a descoberta do conhecimento não ocorre apenas com a mineração de dados. Em muitas situações vastas quantidades de dados manipulados adequadamente podem revelar conhecimento e assim redirecionar processos de tomada de decisão, procedimentos de melhorias na operação de indústrias, estratégias de otimização da produção de produtos e muitas outras.

### 3.3 Mineração de Dados

A obtenção de conhecimento a cada dia torna-se mais difícil como consequência do crescimento das bases de dados em tamanho, complexidade, novos tipos de dados, entre outros aspectos. Imagine a quantidade de dados que são coletados e armazenados pelos sistemas informáticos diariamente. Essas grandes massas de dados suscitam questões como: Existe algo de interessante nesses dados? O que poderá ser feito com eles?

Nos anos 80 e 90 surgiram ferramentas de extração e análises de dados, como por exemplo as ferramentas *OLAP*, *Report*, *Query* (TAN et al., 2009) que agilizaram a busca e visualização de dados pelos usuários finais, e eram capazes de responder indagações do tipo: Qual a quantidade de energia consumida por determinada região? ou Qual a quantidade de defeitos que ocorreram em um gerador em um período especificado de operação?

Mas por outro lado, perguntas como: Quais itens são comprados por pessoas adultas e com renda alta? Ou, o cliente que compra televisão sempre compra refrigerante? Não são facilmente respondidas pelas análises tradicionais e pelas ferramentas existentes na época, pois estas inferências são diferentes de mera análise estatística.

Essas perguntas motivaram um método de extração de conhecimento, a denominada mineração de dados. A MD é a principal etapa da KDD. É uma metodologia que combina

métodos tradicionais de análise de dados com algoritmos sofisticados para processar grandes massas de dados. Há uma grande confusão entre ferramentas tradicionais de consulta e visualização de dados e as ferramentas de mineração de dados. As análises tradicionais normalmente fornecem uma análise tabular. Já a mineração de dados, devido ao seu procedimento científico, identifica todas as possibilidades de correlações existentes nas fontes de dados.

O termo “mineração” pressupõe o garimpo por algo precioso. A “mineração de dados” pressupõe o garimpo em grandes quantidades de dados para obter informações que possa trazer algum valor ou vantagem competitiva para a empresa. Em geral a mineração de dados tem como objetivo descrever ou prever o comportamento futuro de algo.

Descrever tem como foco encontrar algo que faça sentido e que consiga explicar os resultados ou valores obtidos em determinados problemas.

Prever, por outro lado, tem como foco antecipar o comportamento ou o valor futuro de alguma variável de interesse, com base no conhecimento de valores do passado. Diferentes estratégias são utilizadas para buscar indícios que relacionem dados ou fatos. Em todas essas estratégias, o objetivo maior é o de poder generalizar o conhecimento adquirido para novas ocorrências do fenômeno ou para outros contextos ou situações parecidas com aquela utilizada na construção do modelo computacional. A mineração de dados não necessariamente deve encontrar as causas ou solucionar problemas.

A mineração de dados agrupa idéias, e como tal apresenta certas similaridades, em aplicação, com outras técnicas e procedimentos como a amostragem, estimativa e teste de hipóteses, algoritmos de busca, técnicas de modelagem, teorias de aprendizagem da inteligência artificial, reconhecimento de padrões e aprendizagem de máquina. Também, a mineração de dados adotou idéias de outras áreas como: otimização, computação evolucionária, teoria da informação, processamento de sinais, visualização e recuperação de informações.

Outras áreas também desempenham papel chave para que se realizem as tarefas de mineração de dados podendo-se citar: tecnologia de banco de dados para fornecer suporte ao armazenamento e consultas; computação de alto desempenho ou paralela, que muitas vezes é necessária em virtude do alto volume de dados; computação distribuída para ajudar a abordar as questões de tamanho e juntar os dados de locais diferentes.

A Figura 3.4 mostra o relacionamento da mineração de dados com essas outras áreas.

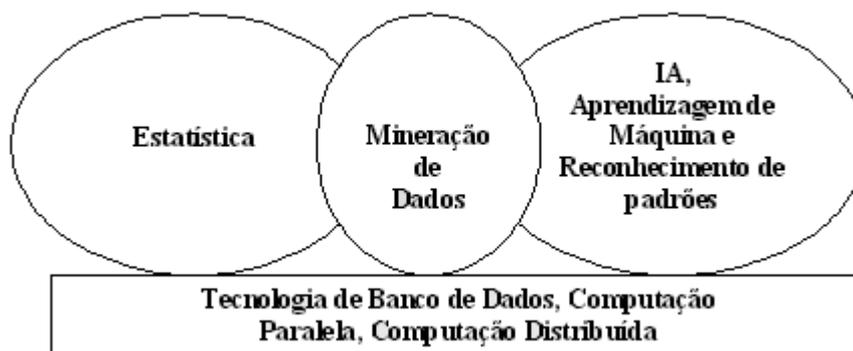


Figura 3.4 - Relação da mineração de dados com outras áreas

Assim, a mineração de dados ajuda a responder perguntas como:

- \* Qual o perfil de um cliente?
- \* Qual a perspectiva de lucro de uma empresa no próximo ano?
- \* Sempre que um hidrogerador apresenta algum defeito, outro também apresentará?
- \* Qual o padrão de defeitos de determinado equipamento?
- \* Sempre que ocorrer um alarme tipo A, após 1 hora ocorrerá o alarme tipo B?
- \* Quão bem se pode prever o início e fim de um período de crescimento?

Com relação às metodologias estatísticas aplicáveis aos procedimentos de MD, registra-se que, embora (HAND, 1998) afirme que o termo MD possa trazer uma conotação simplista para os estatísticos, mostraram a relevância da estatística para o processo de extração de conhecimentos. De acordo com (HAND, 1998), a estatística se preocupa com a análise primária dos dados, no sentido de que eles são coletados por uma razão particular ou por um conjunto de questões particulares. A MD, por outro lado, se preocupa com a análise secundária dos dados, em um sentido mais amplo e mais indutivo do que uma abordagem hipotético-dedutiva, freqüentemente considerada como o paradigma para o progresso da ciência moderna. Assim, MD pode ser vista como o descendente direto da estatística, já que são técnicas metodológicas complementares.

É muito comum na literatura a definição das tarefas e técnicas de MD, entretanto, não existe um consenso nestas definições. As tarefas podem ser entendidas como tipos de relacionamentos entre os dados que serão estabelecidos para a obtenção do conhecimento. As técnicas ou algoritmos são considerados como os fundamentos computacionais que propiciam a busca do conhecimento. O processo de mineração inicia-se com a escolha da tarefa

conforme o tipo de conhecimento que se espera extrair dos dados e após a escolha da tarefa, escolhe-se uma técnica ou algoritmo para realizar a tarefa. Geralmente as tarefas de MD são divididas em duas categorias: Tarefas Preditivas, e Tarefas Descritivas como mostradas na figura 3.5.

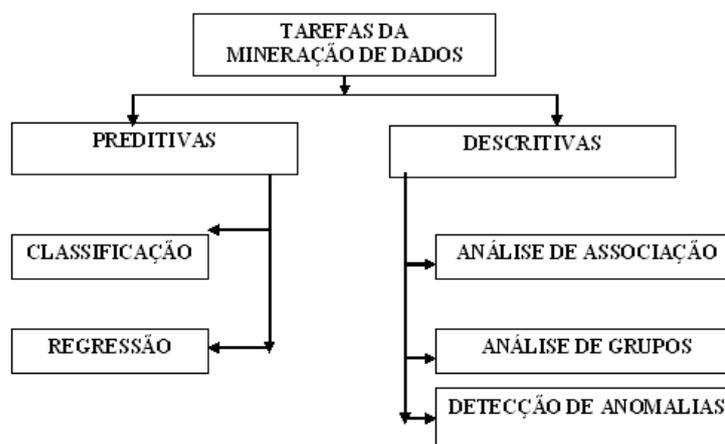


Figura 3.5 - Tarefas da mineração de dados

### 3.3.1 Tarefas Preditivas

O objetivo da tarefa preditiva (ou supervisionada) é prever o valor de um determinado atributo baseado nos valores de outros atributos. É aprender um modelo que minimize o erro entre o valor previsto e o valor real da variável alvo. O atributo cujo valor deve ser descoberto é conhecido como variável **alvo ou dependente** e os outros atributos usados para fazer a predição são as variáveis **independentes ou explicativas**. Essas tarefas podem identificar clientes que responderão a uma campanha de vendas, prever perturbações climáticas, ou julgar se um paciente possui uma doença, baseado em exames médicos. Normalmente na literatura as tarefas preditivas são: Classificação e Regressão.

#### Classificação

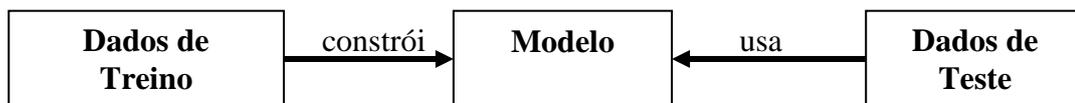
A classificação é a tarefa de classificar objetos em uma dentre várias categorias já conhecidas. Exemplos desta tarefa incluem a classificação de galáxias baseada em seus formatos; classificação do estado de um sistema elétrico em normal e emergência, classificação da operação de um gerador em segura e insegura e outros. A figura 3.6 mostra o princípio básico da classificação.



**Figura 3.6 - Princípio básico classificação**

Os dados de entrada são um conjunto de registros ou instâncias onde cada um é formado por uma dupla  $(x,y)$ , onde  $x$  é o conjunto de atributos e  $y$  é o rótulo da classe, que é a variável alvo. Os atributos podem ser discretos ou contínuos. O rótulo da classe, entretanto, deve ser um atributo discreto. Quando o rótulo da classe é um atributo contínuo, então a tarefa passa a ser a regressão. Esta é a característica chave que distingue a classificação da regressão. Assim, a classificação é a tarefa de aprender uma função alvo  $f$  que relacione cada conjunto de atributos  $x$  para um dos rótulos de classes  $y$  pré-especificados (TAN et al., 2009). A tarefa de classificação é uma abordagem para a construção de modelos classificadores a partir de conjuntos de dados. Um conjunto de dados de entrada (dados de treino) com rótulos conhecidos é usado para a construção do modelo. Após o modelo ser gerado, um outro conjunto de dados (dados de teste) sem rótulos é usado para os testes, onde o modelo deverá rotular estes dados corretamente. Finalmente, o modelo construído e testado pode ser utilizado.

A figura 3.7 mostra a abordagem para a construção de um modelo de classificação.



**Figura 3.7 - Abordagem para o modelo de classificação**

Os testes permitem uma avaliação do desempenho do modelo de classificação. Esta avaliação é medida pelas contagens dos registros de teste previstos corretamente e incorretamente. Estas contagens são tabuladas em uma tabela chamada de Matriz de Confusão.

A tabela 3.1 apresenta uma matriz de confusão para um modelo de classificação binária.

**Tabela 3.1 - Matriz de Confusão**

		Classe prevista	
		Classe = 1	Classe = 0
Classe real	Classe = 1	$f_{11}$	$f_{10}$
	Classe = 0	$f_{01}$	$f_{00}$

As entradas  $f_{11}$ ,  $f_{10}$ ,  $f_{01}$  e  $f_{00}$ , descritas na tabela 2.1 significam respectivamente: registros da classe 1 previstos como classe 1, registros da classe 1 previstos como classe 0, registros da classe 0 previstos como classe 1 e registros da classe 0 previstos como classe 0. O número de previsões corretas é  $f_{11} + f_{00}$  e o de incorretas é  $f_{10} + f_{01}$ . Baseado na matriz de confusão, as métricas de desempenho como a precisão e erro são definidas nas equações (1) e (2) no Anexo II:

As principais técnicas usadas na classificação são: classificadores de árvores de decisão, classificadores baseados em regras, classificadores de vizinho mais próximo, bayesianos, rede neural artificial (RNA), máquinas de vetor de suporte (SVM).

Nesse trabalho optou-se por utilizar a técnica de árvore de decisão devido a sua maior interpretabilidade comparada a outras técnicas. Nas outras técnicas não é óbvio para o usuário interpretar o resultado obtido, ao contrário das árvores de decisão.

Uma árvore de decisão pode ser facilmente transformada num conjunto de regras de classificação e vice-versa.

Uma rede neural também pode ser transformada em um conjunto de regras de classificação, como foi feito com as árvores de decisão. A única diferença é que esta transformação não é tão evidente como no caso das árvores de decisão. Além disto, é difícil para os humanos interpretar o significado simbólico que está por trás dos pesos sinápticos das conexões da rede.

## **Regressão**

A regressão é uma técnica de modelagem preditiva em que a variável alvo é contínua. Na regressão são utilizadas as técnicas matemáticas de Regressão Linear, Regressão não Linear, Regressão Múltipla, Regressão Logística, Regressão de Poisson.

### 3.3.2 Tarefas Descritivas

O objetivo da tarefa descritiva (ou não supervisionada) é derivar padrões (correlações, tendências, grupos, trajetórias e anomalias) que resumam os relacionamentos entre os dados. Essas tarefas podem descobrir páginas da Web que são acessadas juntas, áreas dos oceanos com características semelhantes e detectar fraudes ou intromissões em redes elétricas. Dentre estas tarefas as mais comuns são: Análise de Associação, Análise de Grupos e Detecção de Anomalias.

#### Análise de Associação

A análise de associação é útil para descobrir relacionamentos interessantes escondidos em grandes conjuntos de dados. Os relacionamentos descobertos são representados na forma de regras de associação. As regras podem mostrar que existe um relacionamento forte entre itens comprados juntos, podem revelar relações entre a terra e os oceanos, entre equipamentos elétricos, entre tipos de defeitos e outros. Uma regra de associação é uma expressão na forma  $X \rightarrow Y$ , ( $X$  implica em  $Y$ ) onde  $X$  e  $Y$  são conjuntos disjuntos de itens, isto é,  $X \cap Y = \emptyset$ . A força de uma regra pode ser medida em termos de seu **suporte** e **confiança**. O suporte determina a frequência na qual uma regra é aplicável a um determinado conjunto de dados, enquanto a confiança determina a frequência nas quais os itens de  $Y$  aparecem nas transações que contenham  $X$ . Estas duas medidas são apresentadas no Anexo II pelas equações (3) e (4).

As regras de associação também podem descobrir padrões sequenciais em eventos ocorrendo no tempo. Os algoritmos mais usados para descobrir as regras de associação são: *Apriori* e o *FP-Growth*.

O algoritmo *Apriori*, na medida em que a quantidade de dados a ser minerada aumenta e/ou o valor do suporte mínimo diminui, acaba gerando um número muito elevado de padrões frequentes, o que torna o processamento computacionalmente custoso e, muitas vezes inviável. Por sua vez, o algoritmo *FP-Growth* utiliza uma estrutura de dados baseada em árvore para a compressão e representação dos dados. Tal estrutura possibilita que o processamento seja realizado primariamente em memória, reduzindo as operações que envolvem a leitura da base de dados. Também, o algoritmo *FP-Growth* necessita de apenas duas passagens pelos dados para a construção da árvore. Uma vez construída, todas as demais operações são feitas em memória. Assim, o algoritmo apresenta bom desempenho na maioria dos casos, devido à capacidade de compressão da base de dados pela árvore.

## Análise de Grupos

Segundo (TAN et al., 2009), a análise de grupo agrupa objetos baseados apenas nas informações encontradas nos dados que descrevem os objetos e seus relacionamentos. Os objetos de um grupo (*cluster*) devem ser semelhantes entre si e diferentes dos objetos dos outros grupos. Quanto maior a semelhança em um grupo e maior a diferença entre os grupos, melhor ou mais distinto será o agrupamento. Os agrupamentos podem ser classificados em diferentes tipos: bem separados, baseados em protótipos, baseados em grafos, baseados em densidade e propriedades compartilhadas.

Em um grupo bem separado, cada ponto fica mais próximo de todos os pontos de seu grupo do que de qualquer ponto em outro grupo. Nos grupos baseados em protótipos, cada ponto fica mais próximo do protótipo que define o grupo do que do protótipo de qualquer ponto em outro grupo. Para dados com atributos numéricos o protótipo de um grupo é muitas vezes o centróide (média de todos os pontos do grupo). Quando os atributos dos dados são categorizados, o protótipo é muitas vezes um medóide (o ponto mais representativo do grupo).

Nos grupos baseados em grafos, cada ponto fica mais próximo de pelo menos um ponto do seu grupo do que de qualquer ponto em outro grupo. Nos grupos baseados em densidade, os grupos formam regiões de alta densidade separados das regiões de baixa densidade. Nos grupos baseados em propriedades compartilhadas, os pontos de um grupo compartilham alguma propriedade geral que derivam do conjunto inteiro de pontos.

Os algoritmos para análise de grupos mais utilizados são: *K-means*, baseado em protótipos, o agrupamento hierárquico aglomerativo, baseado em grafos ou protótipo e o DBSCAN, baseado em densidade.

## Detecção de Anomalias

Na detecção de anomalias o objetivo é encontrar objetos que sejam diferentes da maioria dos outros objetos. A maioria dos trabalhos sobre detecção de anomalias (*outliers*) utiliza técnicas estatísticas.

Não existe uma definição formal de anomalias aceita unanimemente por todos que trabalham nesta área, mas sim, uma noção informal proposta por D. Hawkins citada em (HAWKINS, 1980): “Um *outlier* é um fato que desvia tanto de outros fatos a ponto de gerar suspeitas de que foi gerado por um mecanismo diferente”.

A detecção de anomalia pode ajudar a detectar fraudes, intrusão, distúrbios e eventos inesperados. As técnicas para detectar anomalias são as baseadas em modelos, baseadas em proximidade e baseadas em densidade. As técnicas baseadas em modelos constroem um modelo e as anomalias são os objetos que não se enquadram bem no modelo e muitas dessas técnicas usam as abordagens estatísticas. Nas técnicas baseadas em proximidade os objetos anômalos são aqueles que estão distantes da maioria dos objetos. Nas técnicas baseadas em densidade os objetos que estão em regiões de baixa densidade estão distantes dos outros e podem ser considerados como anomalias. Os algoritmos para detecção de anomalias mais utilizados são: NL(Nested Loop), FindAllOutsM e FindAllOutsD. A tabela 3.2 a seguir sumariza as tarefas com suas técnicas associadas (TAN et al., 2009).

**Tabela 3.2 - Tarefas e técnicas de mineração de dados**

<b>Tarefas</b>	<b>Técnicas</b>
Classificação	Árvores de decisão, Classificadores baseados em regras, Classificadores de vizinho mais próximo, bayesianos, Rede neural artificial (RNA), Máquinas de vetor de suporte (VSM).
Regressão	Regressão linear, Regressão não linear, Regressão múltipla, Regressão logística, Regressão de poisson.
Análise de Associação	Apriori, FP-Growth, DHP, DIC, min-Apriori, GSP, SPIRIT.
Análise de Grupos	K-means, Agrupamentos hierárquico, DBSCAN, Baseados em protótipos, Baseados em densidade, Baseados em grafos, Agrupamento escaláveis.
Detecção de Anomalias	Técnicas estatísticas, Técnicas baseadas em proximidade, em densidade, em agrupamento.

### 3.4 Conclusões

Neste capítulo foi abordado o processo de Descoberta de Conhecimento em Banco de dados, onde as diversas fases desse processo foram descritas. Inicialmente mostraram-se as fases desse processo e posteriormente cada fase foi descrita para permitir seu entendimento. Todas as fases são de fundamental importância, sendo, entretanto a fase de mineração a mais importante, pois é nela que o conhecimento é descoberto. Em seguida foi realizada uma explanação sobre a etapa de mineração de dados, sendo abordadas com detalhes mais precisos as principais tarefas da mineração, como: classificação, associação, agrupamento e detecção de anomalias e as técnicas associadas a elas utilizadas para executar a tarefa. Através de uma tabela se mostram os algoritmos utilizados em cada tarefa.

## CAPÍTULO 4

### Mineração de Dados em Usinas Hidrelétricas

#### 4.1 Introdução

Este se constitui o principal capítulo desta tese de doutorado. Usando dados de uma UHE brasileira, algumas tarefas e técnicas de mineração de dados serão aplicadas para possivelmente se obter conhecimento e aumentar ainda mais a compreensão sobre a usina. Com as técnicas estatísticas, árvores de decisão, associações e utilizando dados reais de uma usina hidrelétrica serão mostrados, de forma prática e com bastante clareza, exemplos de experimentos de mineração de dados em usinas hidrelétricas. Uma interface para detecção em tempo real da ultrapassagem dos limites operacionais do hidrogenerador é desenvolvida baseada em árvore de decisão. Os dados reais são as variáveis simples (discretas e analógicas) obtidas através do sistema SCADA relacionadas a equipamentos tais como: alarmes, sinalizações relacionados com a estrutura e núcleo do estator, enrolamento do estator, entreferro e rotor, enrolamento do rotor, refrigeradores, mancais e sistemas auxiliares externos. Também grandezas como: temperaturas, tensões terminais, correntes terminais, frequência, correntes de armadura, velocidade, ângulo de potência, tensão de campo, corrente de campo, descargas parciais, vibrações entre outras são monitoradas.

Nos experimentos com associações e árvore de decisão optou-se por utilizar o software *RapidMiner*. O *RapidMiner*, anteriormente conhecido como *YALE (Yet Another Learning Environment)*, é um ambiente de máquina de aprendizado, mineração de dados, mineração de texto, e análise preditiva. Ele é usado para pesquisa, educação, treinamento, prototipação rápida, desenvolvimento de aplicações e aplicações industriais. É distribuído com licença AGPL. O *RapidMiner* foi projetado no início de 2001 por Ralf Klinkenberg, Ingo Mierswa e Simon Fischer na Unidade de Inteligência Artificial da Universidade de Dortmund. O *RapidMiner* inclui procedimentos de carregamento dos dados, pré-processamento e visualização dos dados, modelagem e avaliação. Este *software* tem interface gráfica para o usuário e está escrito na linguagem de programação Java e é compatível com o software WEKA.

Foram ainda desenvolvidos *softwares* para realizar as análises estatísticas e pré-processamento dos dados para ajustá-los ao *software* minerador, os quais serão descritos nas próximas seções.

## 4.2 Análises Estatísticas

Apesar da grande quantidade de dados disponíveis para uma análise pós-operação, é uma prática comum nos sistemas elétricos de potência só realizar análise pós-operacional quando algo significativo mereça ser analisado, como, por exemplo, a ocorrência de grandes distúrbios na operação do sistema. Com tantos dados disponíveis, pouco se extrai de informações para auxiliar na tomada de decisões operacionais. Assim, análises estatísticas podem ser realizadas nas variáveis discretas (eventos) para apresentar informações como, por exemplo:

- Quantidade de tipos de eventos que ocorrem na usina;
- Quantidade de eventos nos geradores;
- Quantidade de tipos de eventos nos geradores;
- Tipos de eventos por gerador;
- Tempo médio de aparecimento do evento;
- Quantidade de evento por dia e por hora;
- Quantidade de eventos por estação climática;
- Outros.

Com estas informações disponíveis alguns questionamentos podem ser levantados com relação à operação e desempenho do sistema, como por exemplo, pode ser percebido que muitos eventos discretos estão ocorrendo na usina, que a quantidade de eventos nos geradores está muito elevada, que determinado tipo de evento está ocorrendo em demasia em determinado gerador, o tempo médio entre defeitos é muito pequeno, estão ocorrendo muitos eventos por hora e por dia, entre outros.

Assim as análises estatísticas aplicadas corretamente sobre as variáveis discretas podem vir a ajudar na compreensão, manutenção, operação e gerenciamento de uma usina.

### 4.2.1 Variáveis Discretas

Em geral, na grande maioria dos sistemas SCADA as variáveis discretas são utilizadas para monitorar os estados de dispositivos e equipamentos, como por exemplo, abertura de um disjuntor, parada de uma bomba de água, etc e para informar certas condições que podem estar ocorrendo no processo e que são consideradas alarmes, como por exemplo, ultrapassagens de limites de variáveis analógicas. Para o exemplo sobre a aplicação de técnicas estatísticas aos dados da operação da UHE, o banco de dados que foi trabalhado contém informações reais das variáveis discretas geradas pelo sistema SCADA de uma UHE brasileira, de Janeiro a Dezembro do ano de 2010. Este banco de dados tem os seguintes atributos:

- Date: data do evento,
- Time: tempo da ocorrência do evento em hh/mm/ss,
- ms: milisegundo da ocorrência do evento,
- sec: números de segundos desde 1970,
- Key: identificação do evento,
- Label: descrição do evento,
- Desc: estado do evento,
- Urgency: urgência do evento,
- Type: tipo do evento.

Exemplos de alguns registros desse banco de dados são mostrados na tabela 4.1.

**Tabela 4.1 – Exemplos de registros de variáveis discretas**

19/12/2010	00:06:30	570	1292717190	P02KITGUPIND	UNIT ACT.POW. INDIV MODE REQUEST DISAP	-> APARIC 0 B	BSig
19/12/2010	00:06:31	050	1292717191	02GTAP_MODO_MW_IND	CTRL CONJUNTO DE POTENCIA IND SELEZION	Cmd	
19/12/2010	00:06:51	750	1292717211	02GTAK01P	SET POINT CARGA (COU) =L2/P3=>	120.00 MW	Spt
19/12/2010	00:11:25	000	1292717485	02GAL101IYF	FREQUENCIA DO GERADOR NORM -> LOLO[	58.00] 56.75 Hz	1 ASig
19/01/2010	00:19:08	010	1263860348	20GTU_TIS001T	G20 MET TEMPERATURA PATIM HIHI -> NORM[	] 52.00 °C	0 ASig
19/01/2010	00:32:33	900	1263861153	P00KIT_TVCIWPPPOP	PLANT ACTIVE POWER OP. SETPT =L2/P3=>	4650.00 MW	Spt

Um módulo de *software* foi desenvolvido para realizar um pré-processamento no banco de dados para extrair as informações de identificação do hidrogerador, tipo do evento, urgência do evento e período da ocorrência do evento se de “dia” ou “noite”, para posterior tratamento em análises estatísticas.

Os tipos dos eventos são mostrados na tabela 4.2 e os tipos de urgência mostrados na tabela 4.3 a seguir.

**Tabela 4.2 - Tipos de eventos discretos**

<b>Evento</b>		<b>Descrição</b>
Bsig	==>	Eventos Booleanos
Asig	==>	Eventos Alarmes
Cmd	==>	Comandos
Spt	==>	Set Point

**Tabela 4.3 - Tipos de urgências dos eventos discretos**

<b>Urgência</b>		<b>Descrição</b>
0	==>	Muito Baixa
1	==>	Baixa
2	==>	Alta
3	==>	Muito Alta

Os eventos do tipo Bsig estão relacionados às sinalizações referentes aos diversos equipamentos, como: bombas de óleo, bomba de água, ventiladores, disjuntores, seccionadoras, relés, entre outros.

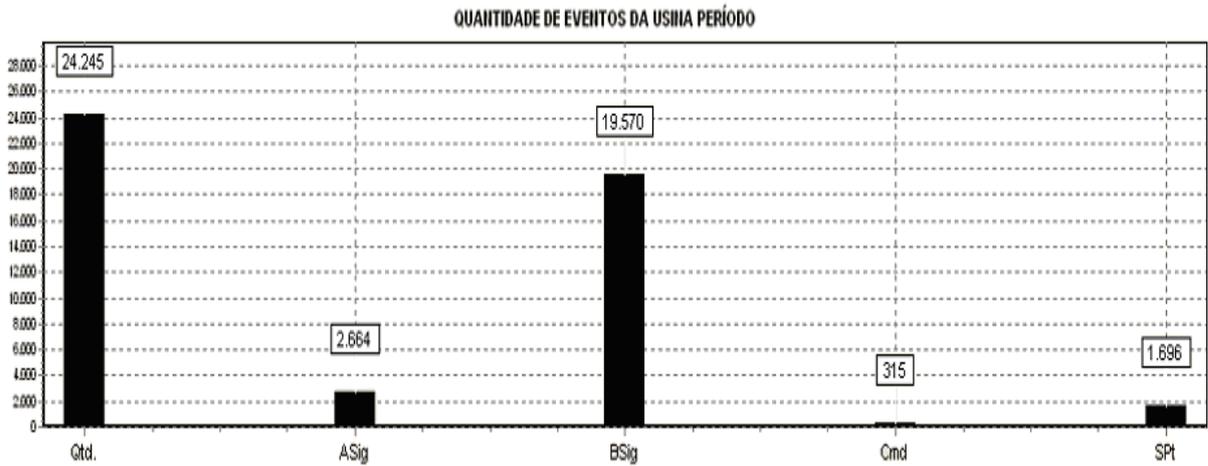
Os eventos do tipo Asig estão relacionados aos alarmes indicando ultrapassagens de limites mínimos e máximos, como: limites de temperaturas, frequência, de potência, entre outros.

Os eventos do tipo Cmd estão relacionados a comandos emitidos pelo operador de forma manual ou automática, como: partida de um hidrogerador, mudança de estado do hidrogerador, abertura ou fechamento de disjuntores e/ou seccionadoras, entre outros.

Os eventos do tipo Spt estão relacionados aos *set points* emitidos pelo operador de forma manual ou automática, como: *set point* de tensão, *set point* de potência ativa, entre outros.

#### **4.2.1.1 Análise de eventos na usina hidrelétrica**

O objetivo desse experimento é a realização de uma análise estatística nos eventos ocorridos durante um determinado período, aqui estabelecido como sendo de 17 a 21 de Janeiro de 2010, para uma avaliação da frequência de ocorrência. Verificou-se a ocorrência de 24.245 eventos distribuídos como mostra a figura 4.1.



**Figura 4.1 – Quantidade de eventos da usina no período de 17 a 21/01/2010**

As médias por hora desses eventos são mostradas na figura 4.2 a seguir.

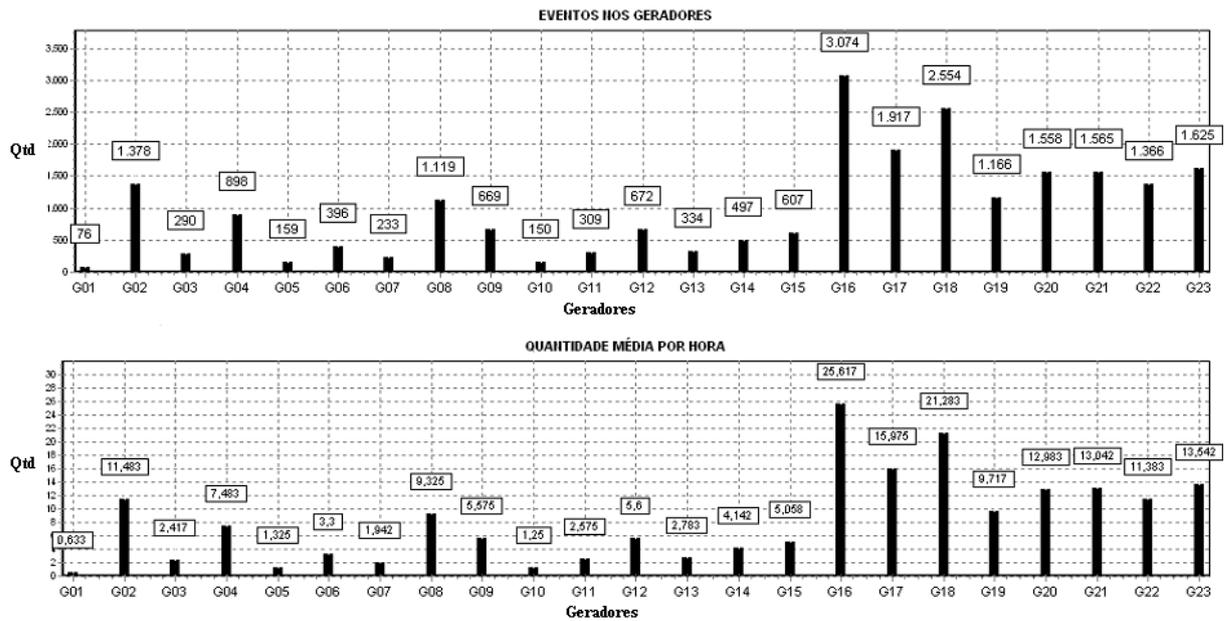
**Qtd. média de eventos/hora da usina**

Todos eventos	202,04	De dia	123,50
Tipo ASig	22,20	De noite	80,54
Tipo Bsig	163,08		
Tipo Cmd	2,63		
Tipo SPt	14,13		

**Figure 4.2 - Média de eventos por hora durante o período**

Houve por hora uma média de 202,04 eventos. Nota-se que ocorreram em média por hora 22,20 eventos tipo A, 163,08 eventos tipo B, 2,63 eventos tipo C e 14,13 *set-points*. Uma análise desses valores deve ser realizada. A quantidade de eventos tipo A pode estar indicando que existem sensores com problemas e/ou a condição operacional do sistema elétrico nesse período variou muito. Por sua vez, eventos do tipo Bsig, podem estar indicando possíveis deficiências de manutenção em equipamentos e sensores. Uma média de 14,13 *set-points* pode estar indicando a necessidade de revisão em ajustes de controladores, por exemplo.

Analisou-se em seguida como os eventos estavam ocorrendo entre os hidrogeradores no mesmo período. A figura 4.3 mostra esta distribuição, para os geradores de G01 a G23.



**Figura 4.3 - Distribuição dos eventos entre os hidrogeradores no período**

Verifica-se a não uniformidade de eventos entre os hidrogeradores e a sua grande quantidade. O hidrogerador G02 é antigo e apresenta uma média de 11,483 eventos que é menor que a média de 25,617 eventos do hidrogerador G16, que é mais novo. Um fato muito preocupante que com certeza merece uma atenção especial na sua análise. Para uma mesma usina e equipe de manutenção, a ocorrência de eventos em condição normal de operação tenderia a ser mais uniforme.

As análises estatísticas nas variáveis discretas são exemplos que remetem a reflexões sobre o comportamento da UHE e podem estar indicando condições não aceitáveis futuramente.

## 4.2.2 Variáveis Analógicas

As variáveis analógicas são utilizadas para monitorar o desempenho, a evolução, a qualidade, entre outros, das grandezas sob supervisão. Geralmente são gravadas nos bancos de dados dos sistemas SCADA em intervalos de alguns segundos. Os dados reais para o estudo provenientes do banco de dados do sistema SCADA, por um período de 01 dia de operação (19/12/2010), foram disponibilizados e contém informações do comportamento de algumas variáveis analógicas de um hidrogerador da usina hidrelétrica. Cada registro do banco de dados representa um ponto de operação, com valores de algumas variáveis, registradas a cada 5 segundos. A tabela 4.4 mostra alguns registros do banco de dados.

**Tabela 4.4 – Exemplos de registros de variáveis analógicas**

0	5.999146e+01	0	1.273755e+01	0	1.270898e+01	0	1.272986e+01	0	1.395835e+01	0	1.398194e+01	0
	1.398784e+01	0	-4.842224e+01	0	3.023962e+02	0	7.140000e+01	0				
5	6.009521e+01	0	1.273755e+01	0	1.275183e+01	0	1.274744e+01	0	1.397267e+01	0	1.397267e+01	0
	1.399627e+01	0	-4.524658e+01	0	3.030261e+02	0	7.140000e+01	0				

Um módulo *software* foi desenvolvido para realizar um pré-processamento no banco de dados para extrair as informações, coloca-las em um padrão mais adequado e de fácil manipulação. Estas informações são gravadas em arquivo de trabalho, que será utilizado nos quatro experimentos realizados adiante. Os atributos desse novo arquivo são mostrados a seguir.

- frege: frequência elétrica do gerador,
- corre: corrente média nas fases na saída do gerador,
- potat: potência ativa gerada,
- potre: potência reativa gerada,
- tensa: tensão média na saída do gerador,
- totpr: temperatura do óleo no transformador principal
- tenca: tensão de campo,
- posdi: posição do distribuidor,
- prece: pressão da caixa espiral,
- veltur: velocidade da turbina.

O mesmo *software* que realiza o pré-processamento contém quatro outros módulos desenvolvidos para realizar os quatro tratamentos descritos a seguir.

#### **4.2.2.1 Comportamento de grandezas do hidrogerador**

Nesse experimento, o objetivo é avaliar o comportamento médio de algumas das grandezas armazenadas de um gerador, durante o período de um dia, amostradas por minuto e apresentá-las por meio de gráficos, os quais são mostrados na figura 4.4.

Frequência média	= 59,961	Potência reativa média	= -114,564
Corrente média nas fases	= 10,681	Tensão média nas fases	= 13,597
Potência ativa média	= 221,693	Temp. trafo principal	= 56,367
Tensão média de campo	= 151,688	Posição média do distribuidor	= 67,822
Pressão média Caixa espiral	= 4,738	Velocidade média da turbina	= 81,868

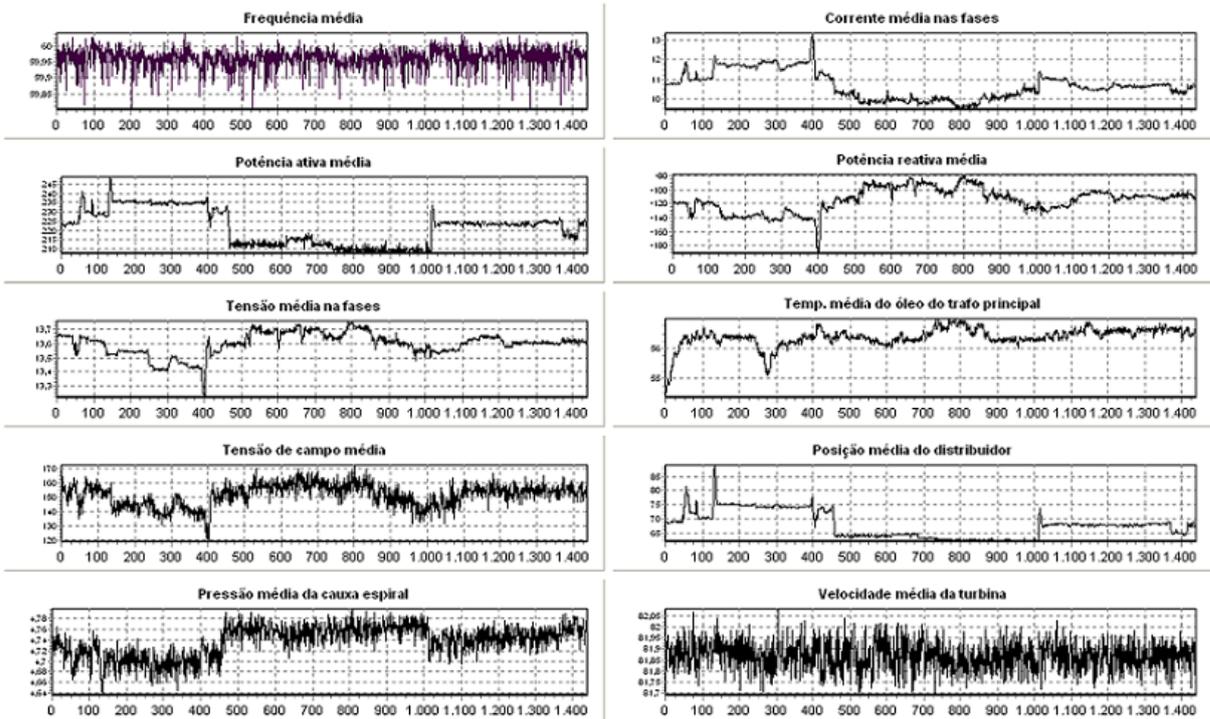


Figura 4.4 - Análise das grandezas analógicas

A representação gráfica das variáveis, por um período de observação, permite que o analista tenha uma visão global do comportamento dessas variáveis, sendo mais fácil identificar pontos críticos como variações abruptas, valores atípicos na operação, ultrapassagem de limiares, entre outros.

#### 4.2.2.2 Correlações entre as variáveis

Nesse experimento o objetivo é realizar correlações entre as diversas grandezas do hidrogerador, correlações estas a serem selecionadas pelo usuário. É possível, por exemplo, correlacionar a potência ativa gerada e a pressão da caixa espiral, ou a frequência e a temperatura do óleo do transformador principal e, descobrir se existe correlação forte entre os comportamentos dessas variáveis. Ao mesmo tempo graficamente é mostrada a evolução das variáveis ao longo do dia, amostradas por minuto. A figura 4.5 a seguir mostra a correlação entre a potência ativa e a pressão da caixa espiral.

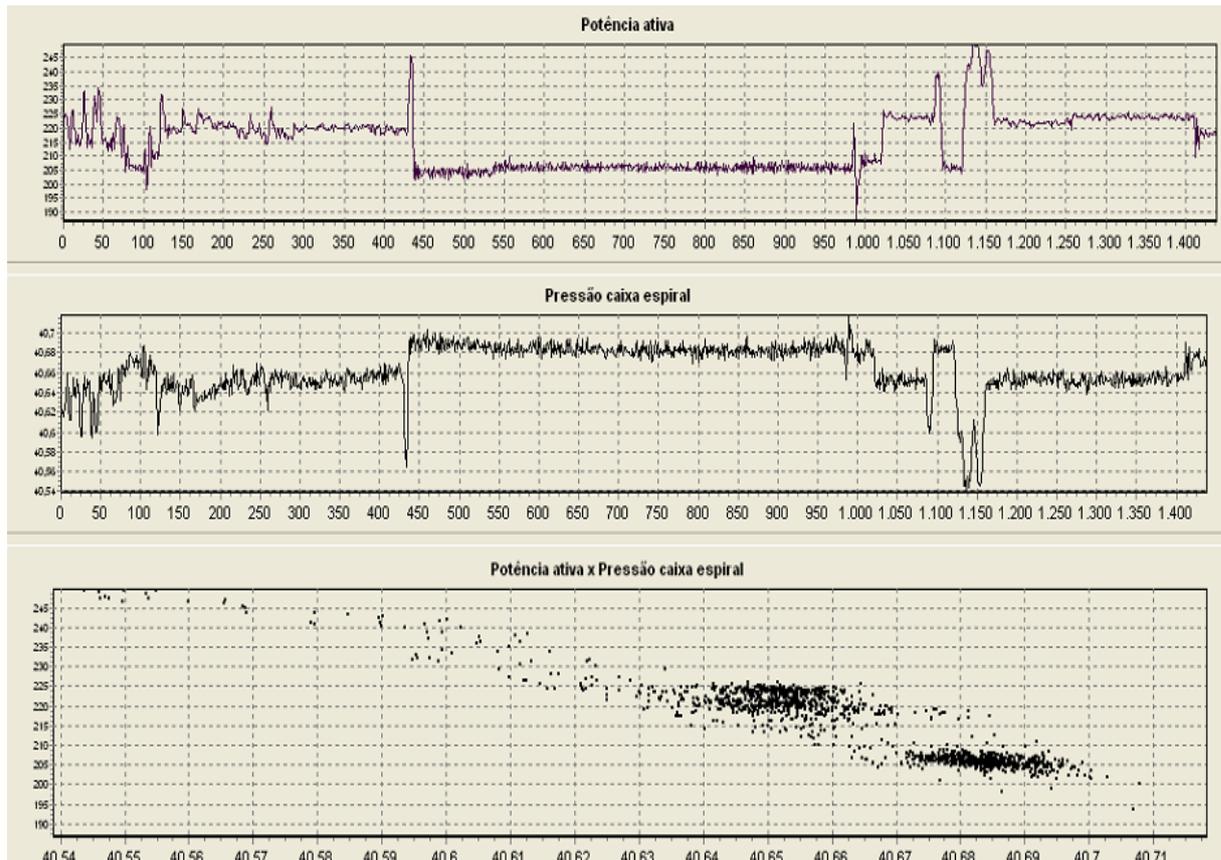


Figura 4.5 – Correlação entre a potência ativa e a pressão da caixa espiral

As correlações entre as variáveis, permitem uma maior compreensão do sistema sob supervisão e controle e assim, vir a ajudar em planejamentos futuros.

#### 4.2.2.3 Análise de Frequência

Nesse experimento o objetivo é avaliar o Desempenho de Frequência em Regime Permanente (DFP) de um hidrogerador segundo as normas estabelecidas pelo Operador Nacional do Sistema – ONS (ONS, 2002). Nesta norma, o indicador DFP avalia as variações de frequência durante a operação do sistema elétrico em regime permanente e são definidas pelas equações (5) e (6) apresentadas no Anexo II.

O indicador DFP não deve apresentar mais que 8 valores da integral do desvio de frequência superior a 0,4 Hz.min a cada dia, e em condições normais de variação de carga, em regime permanente, os desvios da frequência instantânea (valores absolutos sem integralização) em relação ao valor nominal não podem exceder a  $\pm 0,1\text{Hz}$ .

O experimento trabalhou com dados de um hidrogerador do dia 18/01/2010, obtendo-se o desempenho de frequência mostrado na figura 4.6.

Erro médio de frequência =	0,0027	Integral do módulo do desvio de frequência (A) superior a 0,4Hz.mim =	11
Frequência média =	60,0027	Quantidade de vezes que a diferença de frequência foi +/- 0,1Hz =	3638
Valores abaixo de 60Hz =	6316	Desempenho da Frequência em Regime Permanente-DFP =	92,36%
Valores iguais a 60Hz =	0		
Valores acima de 60Hz =	10964		

Figura 4.6 - Análise de desempenho de frequência

A análise do comportamento da frequência mostra que algumas medidas podem ser tomadas para melhorar o seu desempenho. Por exemplo, observando os registros dos valores de frequências acima e abaixo de 60 Hz, nota-se que, no período de observação, ocorreram mais valores acima do que abaixo. Esse fato pode estar contribuindo para um acúmulo de erro de tempo significativo, e neste caso, medidas para melhor ajustar os ganhos do controle de frequência são necessários. Também, pode-se observar que alguns indicadores de desempenho da frequência não estão compatíveis com as normas da ONS, indicando que a qualidade da energia não está como deveria ser e sendo assim, merece uma atenção para resolver esta questão.

#### 4.2.2.4 Ponto de operação em relação à curva de capacidade

Os limites operacionais de um hidrogerador, uma máquina síncrona de pólos salientes, são definidos pela sua curva de capacidade que delimita onde a máquina deve operar de forma segura sem danificar seus enrolamentos (MONTICELLI & GARCIA, 2000), (LOF et al, 1995).

É uma superfície limitada por cinco regiões que depende da potência ativa (P), potência reativa (Q) e tensão terminal (V) como mostrada na figura 4.7. Qualquer ponto de operação fora desta região estará violando condições de operação segura e deve ser evitado.

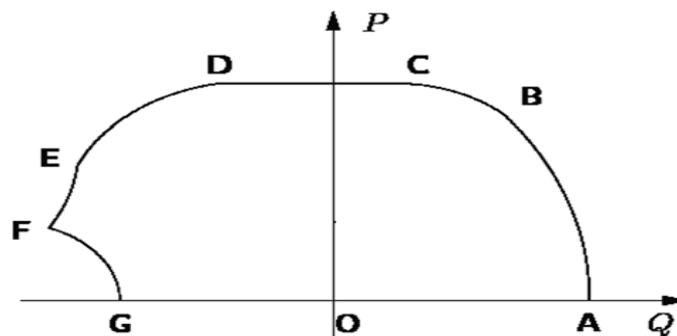


Figura 4.7 – Curva de capacidade do gerador síncrono

Esta superfície é definida em um plano de potência reativa ( $Q$ ) versus potência ativa ( $P$ ) gerada (plano  $P \times Q$ ), com cinco limites mostrados na tabela 4.5 a seguir. Esses limites são funções dos parâmetros básicos da máquina, ou sejam, reatância de eixo direto  $X_d$ , reatância de eixo em quadratura  $X_q$ , tensão terminal  $V_t$ , potência aparente nominal  $S$ , potência mecânica máxima  $P_{mec_{max}}$  e ângulo interno  $\delta$ .

**Tabela 4.5 – Regiões da curva de capacidade**

Limite de excitação máxima, definido pela região AB
Limite de corrente de armadura, definido pelas regiões BC e DE
Limite da máquina primária, definido pela região CD
Limite de estabilidade, definido pela região EF
Limite de excitação mínima, definido pela região FG

Os limites BC e DE, de corrente de armadura ou limite térmico do estator, correspondem a um círculo com centro na origem do plano  $P \times Q$ . É o lugar geométrico dos pontos de operação com corrente estatórica nominal constante. Pode ser obtido graficamente pelo círculo centrado na origem, com raio igual a  $S_n$  (potência aparente). Como este limite é definido pela corrente terminal da máquina, a potência aparente varia diretamente com a tensão terminal. O enrolamento do estator pode suportar uma sobrecarga em regime permanente, mas esta deve ser considerada no seu dimensionamento. A equação deste círculo é dada pela equação (7) mostrada no Anexo II.

O limite C-D, de máxima capacidade da máquina primária é representada no plano  $P \times Q$ , por meio de uma reta paralela ao eixo  $Q$ , e é dada pela equação (8) do Anexo II.

O limite definido por A-B é o limite de excitação máxima que corresponde à máxima corrente que pode ser fornecida ao rotor da máquina sem que haja deterioração dos enrolamentos de campo. Desta forma, este limite é representado pela equação (9) mostrada no Anexo II.

Os limites anteriores definem a operação do hidrogerador na região de sobre-excitação (geração de potência reativa), que é a região onde na maioria das vezes opera o hidrogerador. Entretanto, algumas vezes é necessária a operação na região de sub-excitação (consumo de potência reativa). Assim sendo faz-se necessário a definição dos dois próximos limites: o limite de excitação mínima e o limite de estabilidade prático.

O limite F-G, é uma curva concêntrica com a curva do limite de excitação máxima e é obtida apenas substituindo o termo  $E_{max}$  pelo termo  $E_{min}$  na equação (9) do Anexo II. Desta forma tem-se a equação (10) mostrada no Anexo II.

O limite E-F, de estabilidade em regime permanente, é representado como a potência ativa máxima que pode ser gerada em função da potência reativa. Este limite é dado pela equação (11) mostrada no Anexo II.

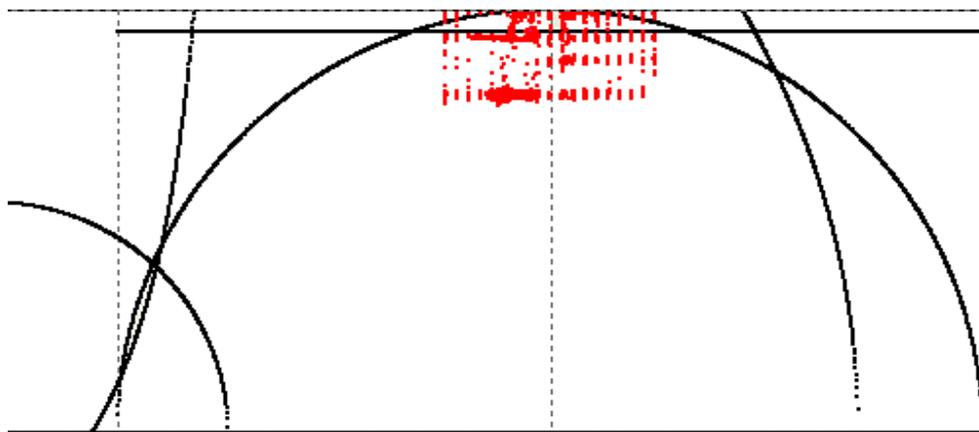
Na prática não é aceito o limite teórico e para se encontrar o limite prático de estabilidade para todos os casos é necessário deixar uma margem de estabilidade disponível de 10% a 20% da potência ativa nominal. A partir das equações que expressam os limites da curva de capacidade obtém-se as restrições impostas para a operação da máquina síncrona.

Neste experimento avalia-se a operação do hidrogerador em relação aos seus limites operacionais impostos pela Curva de Capacidade, com as suas características mostradas na tabela 4.6. Para esta análise, foram utilizados dados da operação de um hidrogerador para o dia 18/01/2010, os quais definem pontos de operação apresentados em vermelho na figura 4.8.

**Tabela 4.6 - Características do hidrogerador**

Tensão terminal da máquina $V_t = 0,95$ pu
Tensão de excitação máxima, $E_{max} = 1,85$ pu
Tensão de excitação mínima, $E_{min} = 0,01$ pu
Potência aparente, $S = 1$ pu
Potência máxima da turbina, $P_{mec_{max}} = 0,98$ pu
Reatância síncrona do eixo direto, $X_d = 1,2$ pu
Reatância síncrona do eixo de quadratura, $X_q = 0,7$ pu

Os dados utilizados foram do dia 18/01/2010.



**Figura 4.8 - Pontos de operação do hidrogerador (18/01/2010)**

As análises das figuras 4.8 e 4.9 mostram o comportamento da operação do hidrogerador em relação à curva de capacidade para o período considerado. A capacidade máxima de geração foi de 375 MW. Nestas condições, neste dia o hidrogerador operou a maior parte do tempo sobre-excitado com 10.583 pontos e sub-excitado com 6.697 pontos. Houve 7.725 pontos em região segura, 9.550 que ultrapassaram a limite da máquina primária e 5 que ultrapassaram a corrente de armadura. A análise destes pontos de operação do hidrogerador permite mostrar o comportamento do mesmo em relação aos seus limites e com isso reavaliar a sua operação para operar com todos os pontos na região segura afim de evitar condições operacionais que possam levar o sistema a uma falha que comprometa a sua integridade.

<b>Limite de aquecimento da armadura</b>	<b>5</b>
<b>Limite da potência da turbina</b>	<b>9.550</b>
<b>Limite de aquecimento do enrolamento do campo</b>	<b>0</b>
<b>Limite de estabilidade</b>	<b>0</b>
<b>Limite de excitação mínima</b>	<b>0</b>
<b>Pontos seguros</b>	<b>7.725</b>
<b>Pontos de operação</b>	<b>17.280</b>
<b>Pontos sobre-excitado</b>	<b>10.583</b>
<b>Pontos sub-excitado</b>	<b>6.697</b>

**Figura 4.9 - Análise dos pontos de operação do hidrogerador**

## **4.2 Regras de Associações**

Com associações procura-se descobrir relacionamentos entre as variáveis discretas buscando-se encontrar regras que possam responder a perguntas como:

- Sempre que um evento tipo A ocorre, a seguir ocorre um do tipo B e outro do tipo C;
- Ocorrendo um evento tipo A, no máximo após 5 minutos ocorre um tipo C;
- Sempre que o disjuntor principal do gerador 03 abre, a bomba de água do gerador 05 pára;
- O evento tipo A sempre ocorre a noite;
- Outras.

Com as associações são realizados três experimentos: o primeiro para descobrir relacionamentos entre os eventos da usina; o segundo para descobrir relacionamentos entre alarmes nos hidrogeradores; e o terceiro para descobrir relacionamentos de alarmes de frequência entre hidrogeradores.

### 4.3.1 Associações entre Eventos da Usina

Com o banco de dados de históricos de eventos do período de 17/01/2010 a 21/01/2010 foi realizado um pré-processamento para gerar um arquivo com 120 registros, onde o primeiro registro indica a hora 0 do primeiro dia e o último registro indica a hora 23 do quinto dia. Os registros possuem os seguintes atributos: data, hora, e seis atributos indicando respectivamente a data, a hora e os tipos de eventos (A, B, C, S) ocorridos na hora. A letra S indica que ocorreu o evento e a letra N indica que ele não ocorreu.

Um registro deste arquivo, por exemplo, poderia ser:

21/01/2010 07 S S N N

indicando que em 21/01/2010 na hora 07 ocorreram os eventos A e B. Este arquivo foi minerado usando também o algoritmo *FP-Growth* para gerar os *item-sets* frequentes e posteriormente regras de associações foram geradas a partir destes *item-sets*. Usando-se confiança mínima de 0,8 obteve-se as regras listadas a seguir.

[A] --> [C] (confiança: 0.824)

[A, S] --> [C] (confiança: 0.920)

[S] --> [C] (confiança: 0.946)

A primeira regra diz que: ocorrendo um alarme na mesma hora ocorrerá um comando. A segunda regra diz que: ocorrendo um alarme e um *set-point* na mesma hora ocorrerá um comando. A terceira regra diz que: ocorrendo um *set-point* na mesma hora ocorrerá um comando. Estas informações devem ser analisadas para avaliar as suas conseqüências reais e verificar se é possível tirar algum benefício para conduzir melhor a operação.

### 4.3.2 Associações entre Alarmes dos Hidrogeradores

Usando o mesmo banco de dados de históricos de eventos ocorridos no período de 17/01/2010 a 21/01/2010, foi criado um arquivo com 120 registros, onde cada registro representa os hidrogeradores que tiveram alarmes (ASig) dentro de cada hora. O primeiro registro representa a hora 0 do primeiro dia e ultimo representa a hora 23 do ultimo dia. Os registros têm os atributos: dia, hora e mais doze atributos, onde cada um indica se os hidrogeradores 1 a 12 tiveram alarmes naquela hora.

A letra S indica a presença de alarme e a letra N a ausência de alarme. Exemplo de um registro deste arquivo é mostrado a seguir.:

01/17/2010 12 SSNNNNNNNNSS

Indicando que no dia 17/01/2010 na hora 12 os hidrogeradores 01, 02, 11 e 12 tiveram alarmes. Este arquivo foi minerado com o algoritmo *FP-Growth* (TAN et al, 2009) para gerar os *item-sets* frequentes e então regras de associação foram extraídas dos *item-sets*. Usando confiança mínima de 0,6, duas regras listadas a seguir foram obtidas.

[G02] → [G11] (confiança: 0.625)

[G08] → {G11} (confiança: 0.667)

A primeira regra mostra que quando o hidrogerador 02 apresenta alarme, na mesma hora o hidrogerador 11 também apresentará alarme.

Similarmente, a segunda regra mostra que quando o hidrogerador 08 alarmar, na mesma hora o hidrogerador 11 irá apresentar um alarme.

De acordo com estas regras o operador saberá com antecedência que quando os hidrogeradores 02 ou 08 tiverem alarmes, existe uma probabilidade de 60% do hidrogerador 11 ter alarme na mesma hora.

Esta informação sugere uma avaliação para determinar as causas desta associação e se este padrão for constante, então o operador poderá realizar as ações necessárias para prevenir o alarme do hidrogerador 11 ou ao menos minimizar as consequências.

### 4.3.3 Associações entre Alarmes de Frequência

Foi observado dos históricos do período de 17/01/2010 a 21/01/2010 que os hidrogeradores 16, 17, 18, 19, 20, 21, 22 e 23 apresentaram alarmes de frequência. Então, com o banco de dados de histórico foi criado um outro arquivo com 120 registros, onde cada registro representa os alarmes de frequência dos hidrogeradores em cada hora do período. A letra S indica a presença do alarme e a letra N a ausência do alarme. Um exemplo de um registro deste novo arquivo é mostrado a seguir.

S      N      S      N      S      N      N      N

Este exemplo indica que o hidrogerador 16, 18 e 20 tiveram alarme de frequência na mesma hora e os hidrogeradores 17, 19, 21, 22 e 23 não. O arquivo foi minerado com o algoritmo *FP-Growth*, com confiança 0,9 e foram obtidas as seis regras listadas a seguir:

- [G18] → [G19] (confiança 1.000)
- [G16] → [G19] (confiança 1.000)
- [G21] → [G19] (confiança 1.000)
- [G22] → [G19] (confiança 1.000)
- [G23] → [G19] (confiança 1.000)
- [G18, G16] → [G19] (confiança 1.000)

A primeira regra mostra que quando o hidrogerador 18 tem alarme de frequência o hidrogerador 19 terá o mesmo alarme na mesma hora. As outras regras têm interpretação similar. Sumarizando, quando os hidrogeradores 16, 18, 21, 22 e 23 apresentam alarmes de frequência o hidrogerador 19 também terá este mesmo alarme na hora. Neste experimento, com confiança 0,6 obteve-se 10 regras. Com confiança 0,7 e 0,8 foram obtidas 7 regras e com confiança 0,9 foram obtidas as seis regras listadas anteriormente. As regras mostram que quando os hidrogeradores 16, 18, 21, 22 e 23 apresentam alarmes de frequência é provável em 100% que o hidrogerador 19 terá o mesmo comportamento na mesma hora. Pode-se observar que estas regras refletem um comportamento consistente que os hidrogeradores devem apresentar, pois todos operam em paralelo, suprindo as variações de carga, que por consequência provocam variações de frequência. Na ocorrência de variação de carga, em um primeiro momento todos os geradores atuam de acordo com as suas regulações primárias, e

em um segundo momento, somente o gerador ou grupo de geradores designados para a corrigir o erro de frequência via o controle secundário, atuarão.

Como se pode verificar conhecendo alguns relacionamentos entre os dados, passa-se a conhecer melhor o sistema e assim, medidas podem ser tomadas para melhorar o gerenciamento da operação das usinas hidrelétricas.

## 4.4 Árvore de Decisão

Dois experimentos foram realizados para descobrir relacionamentos entre variáveis analógicas da usina. O primeiro relacionará variáveis hídricas à potência gerada e a segunda relacionará também variáveis hídricas ao nível de água na cidade a jusante da usina.

### 4.4.1 Potência Gerada na Usina

Neste experimento o interesse é mapear o relacionamento entre vazão turbinada (VT), vazão afluyente (VA) e nível de montante (NM) com a potência gerada (PG) que é a variável alvo. Os valores destes atributos para o período de 10/11/84 a 31/12/10 foram extraídos de um banco de dados com informações hídricas, as quais serão utilizadas como entrada para a árvore de decisão.

A vazão turbinada é a quantidade de água em  $m^3/s$ , que passa através das turbinas, para realizar a geração. A vazão afluyente é a quantidade de água em  $m^3/s$ , que chega a montante da usina. O nível de montante é o nível de água em m na montante da usina.

A variável alvo de uma árvore de decisão deve ser um atributo nominal (não numérico). Nesse caso, sendo a PG uma variável numérica, foi realizada uma discretização nesta variável, onde cada valor dela passa a ser um valor entre as oito faixas: 0-1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000, 5000-6000, 6000-7000 e >7000. Assim, a variável PG deixa de ser atributo e uma nova variável nominal chamada de CL passa a ser o atributo alvo, onde seu conteúdo será um dos oitos valores das faixas.

A árvore foi construída e testada usando o método *Holdout* (TAN et al, 2009) com 70% dos dados para treino, 30% para teste e usando como métrica o ganho de informação para manipulação dos registros.

A árvore criada com precisão de 88,83% e erro de 11,17% gerou as regras mostradas na tabela 4.7 e a árvore mostrada na figura 4.10.

**Tabela 4.7 - Regras para a árvore de decisão da potência gerada**

VT <= 3865	
VT <= 1727,500: 0-1000	1
VT > 1727,500	
VT <= 3388,500	
VT <= 1867,500	
NM <= 71,585: 0-1000	2
NM > 71,585: 1000-2000	3
VT > 1867,500: 1000-2000	4
VT > 3388,500	
NM <= 72,010: 1000-2000	5
NM > 72,010: 2000-3000	6
VT > 3865	
VT <= 6131	
VT <= 5085	
NM <= 61,070	
VT <= 4218,500: 1000-2000	7
VT > 4218,500: 2000-3000	8
NM > 61,070: 2000-3000	9
VT > 5085	
NM <= 65,405: 2000-3000	10
NM > 65,405: 3000-4000	11
VT > 6131	
VT <= 8735	
VT <= 6890: 3000-4000	12
VT > 6890	
NM <= 66,945: 3000-4000	13
NM > 66,945: 4000-5000	14
VT > 8735	
VT <= 10825	
NM <= 65,960: 4000-5000	15
NM > 65,960: 5000-6000	16
VT > 10825	
VT <= 12392,500: 6000-7000	17
VT > 12392,500: >7000	18

As regras geradas são apresentadas através de instruções de decisão (IF's) e mostram o relacionamento entre as variáveis trabalhadas. No conjunto de regras gerado existem 18 regras definidas. As regras 1 e 18 devem ser interpretadas respectivamente como:

- a) se  $VT \leq 3865 \text{ m}^3/\text{s}$  e  $VT \leq 1725,500 \text{ m}^3/\text{s}$  então PG estará entre 0 – 1000 MW.
- b) se  $VT > 3865 \text{ m}^3/\text{s}$  e  $VT > 6131 \text{ m}^3/\text{s}$  e  $VT > 8735 \text{ m}^3/\text{s}$  e  $VT > 10825 \text{ m}^3/\text{s}$  e  $VT > 12392,500 \text{ m}^3/\text{s}$  então PG será  $> 7000$  MW.

As demais regras são interpretadas de forma semelhante.

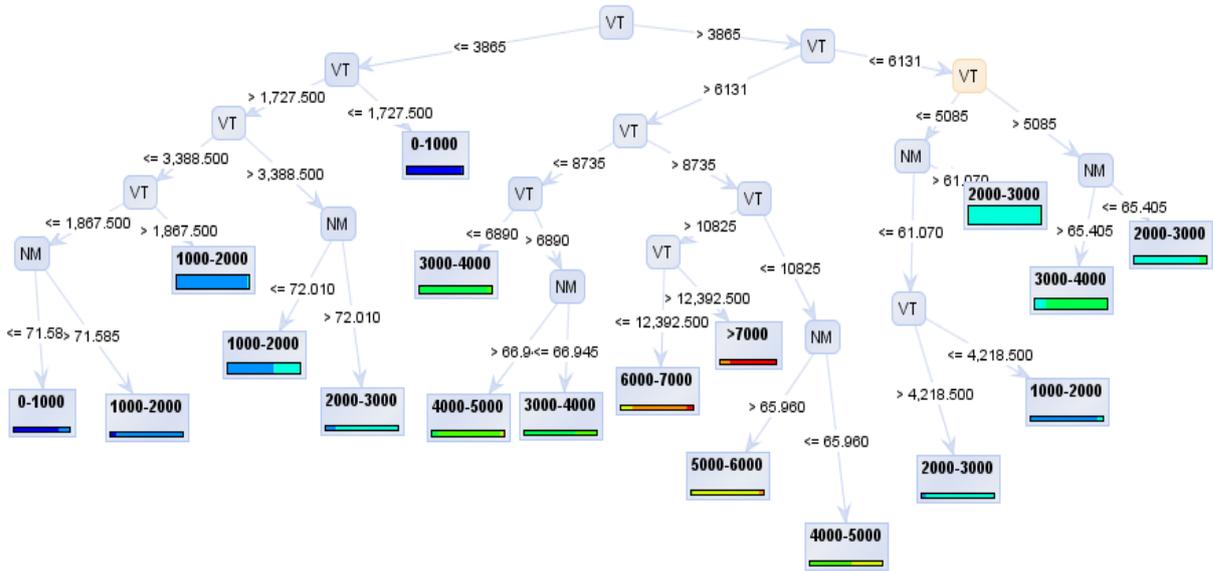


Figura 4.10 – Árvore para a potência gerada

Esses relacionamentos permitem uma melhor compreensão da potência gerada e assim aumentar o conhecimento da geração em relação à vazão turbinada e o nível de montante. Essas informações podem ajudar na qualidade da operação da usina e no planejamento da geração de energia.

#### 4.4.2 Nível de água na Cidade a Jusante da Barragem

Neste experimento o interesse é mapear o relacionamento da vazão turbinada (VT), vazão vertida (VV), nível de jusante (NJ) com o nível de água na cidade (NC) a jusante da usina que é a variável alvo. Os valores destes atributos para o período de 05/01/08 a 31/12/10, foram extraídos do banco de dados com informações hídricas e armazenados em uma planilha. A vazão vertida é a quantidade de água em  $m^3/s$ , que passa pelo vertedouro. O nível de jusante (NJ) é o nível de água em m na saída das turbinas. O nível na cidade (NC) é o nível de água em m acima do nível normal em frente a cidade a jusante da usina. A variável NC é a variável alvo. Sendo esta numérica, foi realizada uma discretização nesta variável, onde cada valor dela passa a ser um valor entre as seis faixas: 0-2m, 2-4m, 4-6m, 6-8m, 8-10m e >10m. Assim, a variável NC deixa de ser um atributo e uma nova variável nominal chamada de CL passa a ser o atributo alvo, onde seu conteúdo será um dos seis valores das faixas.

A árvore foi construída e testada usando o método *Holdout* com 80% dos dados para treino e 20% para teste. A árvore criada apresentou precisão de 89,45% e erro de 10,55% e gerou as regras mostradas na tabela 4.8. A configuração final da árvore de decisão está apresentada na figura 4.11.

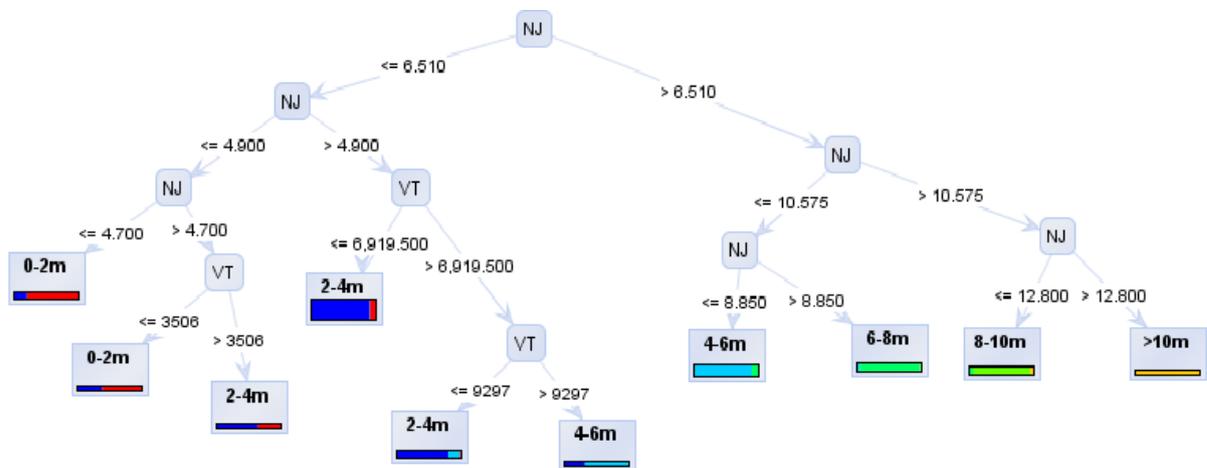
**Tabela 4.8 – Regras para a árvore de decisão do nível da água na cidade**

NJ <= 6.510	
NJ <= 4.900	
NJ <= 4.700: 0-2m	1
NJ > 4.700	
VT <= 3506: 0-2m	2
VT > 3506: 2-4m	3
NJ > 4.900	
VT <= 6,919.500: 2-4m	4
VT > 6,919.500	
VT <= 9297: 2-4m	5
VT > 9297: 4-6m	6
NJ > 6.510	
NJ <= 10.575	
NJ <= 8.850: 4-6m	7
NJ > 8.850: 6-8m	8
NJ > 10.575	
NJ <= 12.800: 8-10m	9
NJ > 12.800: >10m	10

No conjunto de regras gerado existem 10 regras definidas. As regras 1 e 10 devem ser interpretadas respectivamente como:

- se  $NJ \leq 6,510m$  e  $NJ \leq 4,9m$  e  $NJ \leq 4,m$  então NC estará entre 0 – 2 m.
- se  $NJ > 6,510m$  e  $NJ > 10,575m$  e  $NJ > 12,8m$  então NC será  $> 10m$ .

De forma semelhante as demais regras são interpretadas.



**Figura 4.11 – Árvore de Decisão para o nível da água na cidade**

Durante a estação de inverno sempre surge à preocupação das partes baixa da cidade vir a serem inundadas. Assim, é importante conhecer estes relacionamentos para prevenir inundações e avisar a população com antecedência.

Como se constata em ambas as experiências, a construção de árvores de decisão a partir dos dados leva a descoberta de relacionamentos que com certeza ajudarão em um maior conhecimento e compreensão do sistema em estudo.

#### **4.4.3 Limites Operacionais do Hidrogerador**

O hidrogerador, um dos principais componentes de uma usina hidrelétrica, operando em regiões próximas aos seus limites operacionais pode ultrapassar esses limites e vir a comprometer o equipamento e implicar em prejuízos financeiros significativos.

Neste experimento será construída uma árvore de decisão para avaliar o ponto de operação do hidrogerador em relação a sua curva de capacidade e assim avaliar a sua segurança operacional.

#### **Geração de dados simulados e rotulação**

Uma grande quantidade de dados é necessária para se treinar uma árvore de decisão afim de assegurar todas as possibilidades encontradas no problema. Como na maior parte do tempo o hidrogerador trabalha em regiões seguras, existem poucos dados de pontos operacionais que tenham ultrapassado os limites operacionais e em geral estes dados não estão rotulados. Assim, será necessário gerar dados para simular a ultrapassagem dos limites impostos pela curva de capacidade e rotulá-los. Neste sentido, foi gerado um arquivo com dados de pontos de operação, de acordo com as restrições impostas pelas equações (7)-(11) no Anexo II, com os seguintes atributos:

- P – potência ativa,
- Q – potência reativa,
- S – potência aparente,
- C – rótulo do ponto de operação.

O atributo C, sendo a variável alvo, deve ter um dos rótulos mostrados na tabela 4.9 a seguir.

Tabela 4.9 - Rótulos da classe C

Condição operacional	Rótulo
Segura	SEG
Limite de Corrente de Armadura	ARM
Limite de Máxima Potência da Turbina	MPT
Limite de Excitação Máxima	EMAX
Limite de Excitação Mínima	EMIN
Limite de Estabilidade	ESTA

Para cada registro do arquivo com os dados simulados é gerado randomicamente um valor de potência ativa entre 0 e 1 p.u, um valor de potência reativa entre -1 p.u e 1 p.u. Com esses valores calcula-se a potência aparente.

Considerando as características do hidrogerador mostradas na tabela 4.6, aplica-se as equações (7)-(11) para determinar o rótulo relativo a este ponto de operação.

Foram gerados para cada condição operacional 2.000 registros, formando um arquivo de dados simulados com 12.000 registros.

### Treino e teste da árvore de decisão

A árvore foi treinada usando 70% do arquivo para treino e 30% para teste. Foi obtida uma precisão de 97,67% e um erro de 2,23% usando a métrica índice *gini*. As condições de testes geradas para cada limite operacional são mostradas na tabela 4.10 a seguir e a árvore de decisão projetada é mostrada na figura 4.12.

Tabela 4.10 - Regras geradas para os limites operacionais do hidrogerador

P $\leq$ 0.980
S $\leq$ 1.000
Q $\leq$ -0.744
S $\leq$ 0.988: EMIN
S $>$ 0.988: ESTA
Q $>$ -0.744
Q $\leq$ 0.577: SEG
Q $>$ 0.577: EMAX
S $>$ 1.000: ARM
P $>$ 0.980: MPT

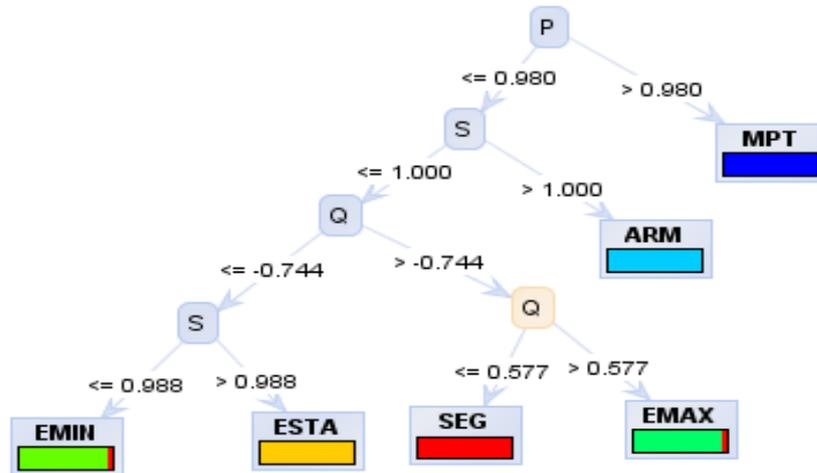


Figura 4.12 - Árvore para os limites operacionais

Foram geradas seis regras as quais serão explicadas a seguir:

- 1) Se  $P > 0,98$  então limite MPT
- 2) Se  $P \leq 0,98$  e  $S > 1$  então limite ARM
- 3) Se  $P \leq 0,98$  e  $s \leq 1$  e  $Q > -0,744$  e  $Q > 0,577$  então limite EMAX
- 4) Se  $P \leq 0,98$  e  $S \leq 1$  e  $Q > -0,744$  e  $Q \leq 0,577$  então SEG
- 5) Se  $P \leq 0,98$  e  $S \leq 1$  e  $Q \leq -0,744$  e  $S > 0,988$  então limite ESTA
- 6) Se  $P \leq 0,98$  e  $S \leq 1$  e  $Q \leq -0,744$  e  $S \leq 0,988$  então limite EMIN

## Testes com dados reais

Foi desenvolvido um software para testar a árvore decisão gerada. Foram usados 17.280 pontos de operação com dados reais, amostrados a cada 5 segundos do dia (18/12/2010) da operação de um hidrogerador. Para cada ponto de operação, foram aplicadas as equações da curva de capacidade e as regras geradas pela árvore de decisão. Comparando os resultados alcançados pela classificação da árvore de decisão com os dados reais obteve-se 100% de acertos.

Dos 17.280 pontos de operação analisados, 7.725 foram classificados como seguros, 5 violaram o limite da corrente de armadura e 9.550 violaram o limite da potência primária da turbina.

Estas regras podem ser facilmente implementadas e integradas ao sistema SCADA para avaliar em tempo real o ponto de operação em relação aos limites operacionais determinados pela curva de capacidade.

Esta experiência mostra que é possível usar uma técnica de mineração de dados baseada em árvore de decisão para detectar as violações aos limites do hidrogerador em tempo real, o que tradicionalmente é realizado por meio das equações algébricas de (7)-(11).

O cálculo de equações que envolvem seno, cosseno, raiz quadrada e potenciação e leva mais tempo computacional do que as instruções IF. Para aplicações em tempo real, a velocidade de processamento é muito crítica, especialmente quando um grande número de hidrogeradores é considerado simultaneamente.

Nesta situação a aplicação de árvore de decisão pode ser uma alternativa atrativa para construir uma interface eficiente para os operadores das usinas, como aquela mostrada na figura 4.13.

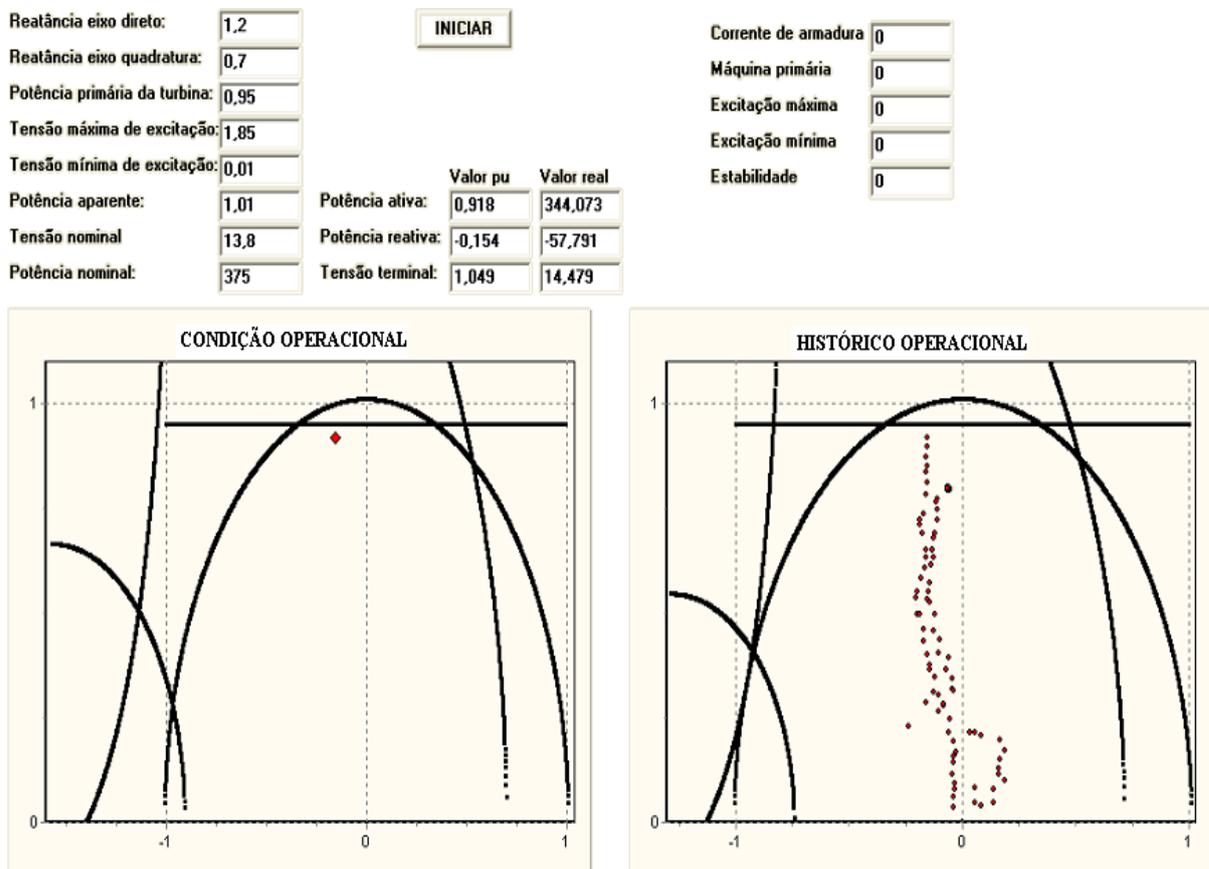


Figura 4.13 - Interface para os limites operacionais dos hidrogeradores

Na figura 13, o gráfico a esquerda apresenta a localização do ponto de operação em tempo real do hidrogerador em relação à curva de capacidade e o gráfico a direita mostra a evolução do ponto de operação do hidrogerador considerado para o dia 17/01/2010.

### 4.3 Conclusões

Este capítulo apresentou algumas experiências de mineração de dados aplicadas em usinas hidrelétricas. Com análises estatísticas, quatro experimentos foram realizados. Com associações usando o algoritmo *FP-Growth* três experimentos foram realizados. Com árvore de decisão, dois experimentos para descobrir relacionamentos entre os dados e um para acessar a ultrapassagem de limites operacionais em tempo real. Estes experimentos demonstram a aplicabilidade da mineração de dados para ajudar na descoberta de conhecimento em massas de dados históricos de usinas hidrelétricas, cujos dados usualmente não seriam mais utilizados para esta finalidade, o que corresponde a desperdiçar uma rica fonte de informações para a gestão operacional e planejamento do sistema.

## CAPÍTULO 5

### Conclusões

#### 5.1 Conclusões Gerais

As hidrelétricas geram imensos bancos de dados formados com dados operacionais obtidos diariamente por seus sistemas SCADA e outras fontes. Esses dados sendo tratados adequadamente caracterizam o estado operacional das usinas ao longo de sua operação e podem vir a revelar aspectos operacionais importantes que ajudem no gerenciamento da geração de energia. Com um mercado altamente competitivo, com seus equipamentos operando próximos aos seus limites devido ao aumento da demanda, a necessidade das empresas do setor elétrico de compreender ainda mais seus sistemas de geração aumenta e as informações adquiridas sobre eles são de vital importância para as empresas do setor. Neste cenário, as técnicas estatísticas e de mineração de dados, como associações, árvore de decisão, rede neural, agrupamentos, entre outras, ganham importância na medida em que, sendo aplicadas aos dados históricos das usinas hidrelétricas podem vir a ajudar na obtenção de conhecimento sobre estas e solucionar problemas. Entretanto, as características multidisciplinares da aplicação das técnicas de mineração de dados, ou seja, a necessidade de conhecer a técnica, bem como, a área de aplicação, aliada à falta de dados rotulados para as mais diversas aplicações, é um forte fator impeditivo na utilização da mineração de dados. Além disso, aproximadamente 60% do tempo gasto no ciclo da KDD está na seleção e preparação dos dados.

Este trabalho mostrou a aplicabilidade da estatística e da mineração de dados nos dados gerados pelos sistemas SCADA de usinas hidrelétricas para ajudar em um melhor gerenciamento da usina e assim vir a fomentar a cultura de sua utilização. Como foi mostrado, uma análise estatística produz informações interessantes sobre a operação das usinas que geralmente não são normalmente disponíveis rotineiramente. Com associações e árvore de decisão foram encontrados relacionamentos entre os dados e demonstraram a sua aplicabilidade na descoberta de conhecimento. Ainda, com árvore de decisão foi desenvolvida uma ferramenta para detecção dos limites operacionais de hidrogeradores para ser aplicada em tempo real. Assim, estes experimentos comprovam a viabilidade das análises pós-operação com técnicas inteligentes. Também em tempo real essas técnicas são perfeitamente

utilizadas e com certeza ajudarão em melhorias para a operação, segurança, manutenção e tomada de decisões nas usinas hidrelétricas, com isso ajudando na fomentação de uma cultura da utilização da mineração de dados nos sistemas elétricos de potência e em especial nas usinas hidrelétricas.

## 5.2 Perspectivas de Futuros Trabalhos

Como trabalhos futuros sugerem-se as linhas indicadas a seguir, para ajudar ainda mais na formação da cultura da aplicação da mineração de dados em usinas hidrelétricas e possíveis descobertas de conhecimentos.

- Mais experimentos para descobrir relacionamentos entre as principais grandezas monitoradas em uma usina hidrelétrica usando associações, classificação e agrupamentos para ajudar na otimização da operação das usinas hidrelétricas;
- A Manutenção Baseada na Condição é uma metodologia que propõe que as manutenções em equipamentos sejam somente realizadas se, o desempenho desses equipamentos estiverem fora dos padrões normais de operação. Assim, uma monitoração constante deve ser realizada para avaliar esse desempenho. Portanto, esta monitoração pode ser realizada com as técnicas de mineração de dados, usando, por exemplo, agrupamentos para auxiliar a manutenção baseada na condição;
- Uso das técnicas de mineração para avaliar a segurança operacional da usina, envolvendo aspectos de segurança estática e de segurança dinâmica em tempo real;
- Definição de uma ontologia para direcionar a aplicação da mineração de dados em sistemas elétricos de potência;
- Definição de uma metodologia para gerar *Data Warehouse* com a finalidade de alimentar a mineração de dados em sistemas elétricos de potência;
- Desenvolvimento de ferramentas para extração do conhecimento em Linguagem Natural

## REFERÊNCIAS

BAOLING, L.; JUN, H.; GUANGMING, L. - **A Study On Conditions Analysis of Power Plant Based On Data Mining** in 2nd International Asia Conference on Informatics in Control, Automation and Robotics, Yuhuan, China, 2010.

BERNARDES, B. C.; OLIVEIRA, W. D.; VIEIRA, J. P. A.; OHANA, I.; BEZERRA, U. H.; NUNES, M. V. A. - **Decision Tree-Based Power System Static Security Assessment Using PMU Measurements** in IEEE PES Trondheim PowerTech, Trondheim, Noruega, 2011.

BRACHNAD, R. J.; ANAND, T. - **The process of knowledge discovery in databases**, In FAYYAD, U. M. et al. *Advances in Knowledge Discovery in Data Mining*. Menlo Park: AAAI Press, 1996.

CAI, Y.; CHOW, M.; LU, W.; LI, L. - **Statistical Feature Selection From Massive Data in Distribution Fault Diagnosis** in IEEE TRANSACTIONS ON POWER SYSTEMS, v. 25, n. 2, pp. 642-648, May, 2010

CIOŚ, K. J.; PEDRYCZ, W.; SWINIARSKI, R. W.; KURGAN, L. A. - **Data Mining A Knowledge Discovery Approach** – Springer, 2007.

COMANESCU, D.; GRIGORAS, G.; CARTINA, G.; ROTARU, F. - **Determination of Typical Load Profiles in Hydro-Power Plant by Clustering Techniques** in 12th International Conference on Optimization of Electrical and Electronic Equipment, Basov, Russia, 2010.

DAS, S.; NAGENDRA, P. S. - **Understanding Power System Behavior through Mining Archived Operational Data** in International Journal of Emerging Electric Power Systems, v. 10, issue 1, Article 5, 2009.

DEVARAJ, D.; J. PREETHA ROSELYN, J. P. - **On-line voltage stability assessment using radial basis function network model with reduced input features** in International

Journal of Electric Power & Energy Systems, v. 33, issue 9, pp. 1550-1555, Nov, 2011.

DIAO, R.; VITTAL, V.; LOGIC, N. - **Design of a Real-Time Security Assessment Tool for Situational Awareness Enhancement in Modern Power Systems** in IEEE Transactions on Power Systems, v. 25, n. 2, pp. 957-965, May, 2010.

DIAO, R.; SUN, K.; VITTAL, V.; O'KEEFE, R. J.; RICHARDSON, M. R.; BHATT, N.; STRADFORD, D.; SARAWGI, S. K. - **Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements** in IEEE Transactions on Power Systems, Vol. 24, n.º. 2, pp. 832-839, May, 2009.

DIAS, C. A. - **Descoberta de Conhecimento em Banco de Dados para Apoio a Tomada de Decisão**- Monografia, Universidade Estadual Paulista, Brasil, 2002.

DINIZ, C.A. e NETO, L. F. - **Data Mining: Uma Introdução** in 14º Simpósio Nacional de Probabilidade e Estatística (SINAPE), Caxambu, MG, Brasil, 2000.

DISSANAYAKA, A.; ANNAKAGE, U. D.; JAYASEKARA, B. BAGEN, B. - **Risk-Based Dynamic Security Assessment** in IEEE Transactions on Power Systems, v. 26, n. 3, pp. 1302-1308, Aug, 2011.

DUARTE, J. L. B. - **Utilização de Técnicas Inteligentes nas Metodologias de Manutenção de Geradores Síncronos** - Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Brasil, 2007.

FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P. - **From data mining to knowledge discovery in databases** in AAAI Press/The MIT Press, 1996.

FERREIRA, G. G. L. - **Desenvolvimento de um Sistema Baseado em Regras para Detecção de Fraude em Unidades Consumidoras Ligadas em Baixa Tensão** – Trabalho de Conclusão de Curso, Universidade Federal de Santa Maria, Brasil, 2007.

FILHO J. R. - **Sistema Inteligente Baseado em Árvore de Decisão, para Apoio ao Combate as Perdas Comerciais na Distribuição de Energia Elétrica** - Dissertação de

Mestrado, Universidade Federal de Uberlândia, Brasil, 2006.

FILHO, J. C. R.; AFFONSO, C. M.; OLIVEIRA, R. C. L. - **Pricing Analysis in the Brazilian Energy Market a Decision Tree Approach** in PowerTech Conference, Bucharest, România, 2009.

FILHO, M. N. S. - **Sistema Inteligente para Tomada Rápida de Decisões nos Sistemas Elétricos** – Tese de Doutorado, Universidade Federal de Itajubá, Brasil, 2006.

GENC, I.; DIAO, R.; VITTAL, V.; KOLLURI, S.; MANDAL S. - **Decision Tree-Based Preventive and Corrective Control Applications for Dynamic Security Enhancement in Power Systems** in IEEE Transactions on Power Systems, v, 25, n. 3, pp. 1611-1619, Aug, 2010.

Haidar, A. M. A.; MUSTAFA, M. W.; IBRAHIM, F. A. F.; AHMED, I. A. - **Transient stability evaluation of electrical power system using generalized regression neural networks** in Applied Soft Computing, journal 11, issue 4, pp. 3558-3570, Jun, 2011.

HAN, J.; KAMBER, M. - **Data Mining: Concepts and Techniques** - Morgan Kaufmann, 2006.

HAND, D. J. - **Data Mining: statistics and more?** in The American Statistician Association, v. 52, n. 2, pp. 112-118, May, 1998.

HAWKINS, D. M. - **Identification of Outliers** - Chapman and Hall, London, 1980.

HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. - **Extreme learning machine: a new learning scheme of feedforward neural networks** in IEEE Int. Joint Conf. Neural Networks, Budapeste, Hungria, 2004.

HUANG, J.A.; VANIER, G.; LOUD, L.; GUILLON, S.; RIZZI, J. C.; Guillemette, F. - **Topology Information Based Decision Trees to Predict Dynamic Transfer limits and Their Sensitivities for Hydro-Quebec's Network** in

HUATUCO, D. N. Z - **Fluxo de Potência Ótimo com restrições da Curva de Capabilidade do Gerador Síncrono** - Dissertação de Mestrado, Universidade Federal do Maranhão, Brasil, 2006.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Guide for Applications of Plant Monitoring for Hydroelectric Facilities – Potential cost benefits of Plant Condition Monitoring (PCM). Janeiro de 1999.

JUNJIE, G.; QUNLI, S. - **School of Energy and Power The Application of Association Rules in Boiler Operation Optimization based on Organizational Evolutionary in Power and Energy Engineering Conference**, Wuhan, China, 2009.

KAMWA, I.; SAMANTARAY, S. R.; JOOS, G. - **Development of Rule-Based Classifiers for Rapid Stability Assessment of Wide-Area Post-Disturbance Records** in IEEE Transactions on Power Systems, v. 24, n. 1, pp. 258-270, Feb, 2009.

KAMWA, I.; SAMANTARAY, S. R.; JOOS, G. - **Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices** in IEEE Transactions on smart grid, v. 1, n. 2, pp.144-158, Sep, 2010.

KARAMI, A. - **Power system transient stability margin estimation using neural networks** in International Journal of Electrical Power and Energy Systems, v. 33, issue 4, pp. 983-991, 2011.

KRISHNAN, V.; MCCALLEY, J. D.; HENRY, S. ISSAD, S. - **Efficient Database Generation for Decision Tree Based Power System Security Assessment** in IEEE Transactions on Power Systems, v. 26, n. 4, pp. 2319-2327, Nov, 2011.

KUNDUR, P - **Power System Stability and Control** - McGraw-Hill, 1994.

LAROSE, D. T. - **Discovering Knowledge in Data - An Introduction to Data Mining** - John Wiley & Sons, Inc, 2005.

LI, J; WANG, S.; NIU, C.; LIU, J. - **Research and Application of Data Mining Technique in Power Plant** in International Symposium on Computational Intelligence and Design, Wuhan, China, 2008.

LI, X.; LI, H.; WU Z. - **Model-Driven Data Mining in the Oil & Gas Exploration and Production** in 2009 Second International Symposium on Knowledge Acquisition and Modeling, Yuhan, China, 2009

LI, Z.; WU, W. - **Phasor Measurements-Aided Decision Trees for Power System Security Assessment** in Second International Conference on Information and Computing Science, Manchester, Inglaterra, 2009.

LIMA, J. M - **Usinas Hidrelétricas - Diretrizes Básicas para Proteção e Controle – Sinergia**, 2009.

LIN, X.; DONG, X.; LU, Y. - **Application of Intelligent Algorithm In Island Detection of Distributed Generation**, in Australian Journal of Electrical & Electronics Engineering, v. 7, n. 3, pp. 203-210, 2010.

LIN, Y. - **Prevention of transient instability employing rules based on back propagation based ANN for series compensation**, International Journal of Electrical Power and Energy Systems, v. 33, n. 10, pp. 1776-1783, Dec, 2011,

LOF, P. A.; ANDERSON, G.; HILL, D. J. - **Voltage Dependent Reactive Power Limits for Voltage Stability Studies** in IEEE Transactions on Power Systems, v. 10, n. 1, pp. 220-228, 1995.

MAHMOODIANFARD, F.; MOHAMMADI, M.; GHAREHPETIAN, G. B.; ABYANEH, H. A. - **Optimal PMU Placement for Voltage Security Assessment using Decision Tree** in IEEE Bucharest Power Tech Conference, Bucharest, Romania, 2009.

MANNILA, H. - **Data mining: machine learning, statistics and databases** in International Conference on Statistics and Scientific Database Management, Stockholm,

Sweden, 1996.

**MATOS, E. R. - Um Método para Detecção e Classificação de Curtos-Circuitos em Redes de Distribuição de Energia Elétrica Baseado na Transformada de Fourier e em Redes Neurais Artificiais**, Dissertação de Mestrado, Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil, 2009.

**MEJÍA-LAVALLE, M.; ARROYO-FIGUEROA, G.; MORALES, E. F. - Innovative Applications of Diagnosis, Forecasting, Pattern Recognition and Knowledge Discovery in Power Systems** in Power & Energy Society General Meeting, Calgary, AB, 2009.

**MINUSSI, M. M. – Metodologia de Mineração de Dados para Detecção de Desvio de Comportamento do Uso de Energia em Concessionária de Energia Elétrica** - Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, Brasil, 2008.

**MONTEIRO, D.; BARRANTES, D. B.; QUIROS, J. M. - Introducción a los sistemas de control supervisor y de adquisición de datos (SCADA)** - Monografia, Universidad de Costa Rica, Costa Rica, 2004.

**MONTICELLI, A., GARCIA, A. - Introduction to Power Systems** - Editora Unicamp, 2000.

**MORAIS, J.; PIRES, Y.; CARDOSO, C.; KLAUTAU, A - An Overview of Data Mining Techniques Applied to Power Systems** in Chapter of Data Mining and Knowledge Discovery in Real Life Applications, InTech, 2009.

**MORETO, M.; ROLIM J. G. - Using phasor data records and sequence of events to automate the classification of disturbances of power generating units** in Electric Power Systems Research, v. 81, n. 7, pp. 1266-1273, Jul, 2011.

**NI, Y.; WEN-YING, L. - Development and Application of Intelligent Alarm System** in Power and Energy Engineering Conference, Chengdu, China, 2010.

NIRKHI, S. - **Potential use of Artificial Neural Network in Data Mining** in 2nd International Conference Computer and Automation Engineering, Singapore, Singapore, 2010.

NIU, C.; LI, J.; LIU, J.; Tan, W. - **Correlation analysis of operation data and its application in operation optimization in power plant** in Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Shandong, China, 2008.

ONS - **Gerenciamento dos indicadores de desempenho da rede básica e de seus componentes, Submódulo 2.8**, 2002.

PIRES, Y. P. - **Mineração de Dados Aplicada a Sistemas Elétricos: Classificação de Falhas de Curto-Circuito em Linhas de Transmissão** - Tese de Doutorado, Universidade Federal do Pará, 2009.

S. A. BOYER, S. A. - **SCADA: Supervisory Control and Data Acquisition** – Instrumentation Society of America, 1993.

SCHUCH, R.; DILL, S. L.; SUASEN, P. S.; PADOIN, E. L.; CAMPOS, M. – **Mineração de Dados em uma Subestação de Energia Elétrica** in Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications, Brasil, 2010.

SOUZA, R. Q. – **Metodologia e Desenvolvimento de um Sistema de manutenção Preditiva Visando a Melhoria da Confiabilidade de Ativos em Usinas Hidrelétricas** - Dissertação de Mestrado, Universidade de Brasília, Brasil, 2008.

SOUZA, Z.; SANTOS, A. H. M.; BORTONI, E. - **Centrais Hidrelétricas - Implantação e comissionamento** – Editora Interciência, 2009.

SQL MAGAZINE - **Data Mining**, Edição 10. Em [http://www.sqlmagazine.com.br/Mat\\_Capa\\_SQL10.asp](http://www.sqlmagazine.com.br/Mat_Capa_SQL10.asp)

ŠTEFAN, Z.; BREZOVEC, M.; MUNAR, D. - **A Decision Support System for Hydro**

**Power Plants in Markets for Energy and Ancillary Services** in 8th International Conference on the European Energy Market, Zagreb, Croácia, 2011.

SUN, H.; HUANG, Y.; HUANG, K.; SU, W. - **Vibration Fault Diagnosis of Rotating Machinery in Power Plants** in Fourth International Conference on Innovative Computing, Information and Control, Kaohsiung, Taiwan, 2009.

SUN, K.; LIKHATE, S.; VITTAL, V.; KOLLURI, V. S.; MANDAL, S. - **An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees** in IEEE Transactions on Power Systems, v. 22, n. 4, pp. 1935-1943, Nov, 2007.

TABBAL, R. L.; LEMOS, F.A.B. – **Determinação de Prováveis Causas de Falhas em Sistemas de Distribuição Utilizando um Sistema Fuzzy** in XVIII Congresso Brasileiro de Automática, Bonito-MS, Brasil, 2010.

TAN, P.; STEINBACH, M.; KUMAR, V. - **Introdução ao Data Mining - Mineração de Dados** - Editora Ciência Moderna, 2009.

TRONCHONI, A. B. - **Identificação de Causas de Desligamentos não Programados em Redes de Distribuição** - Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, Brasil, 2008.

TRONCHONI, A. B.; PRETTO, C. O.; ROSA, M. A.; LEMOS, F. A. B. - **Descoberta de Conhecimento em Base de Dados de Eventos de Desligamentos de Empresas de Distribuição** em Revista Controle & Automação, v. 21, n. 2, pp. 185-200, Março e Abril, 2010.

VALE, Z. A.; RAMOS, C.; RAMOS, S.; PINTO TIAGO - **Data Mining Applications in Power Systems – Case-studies and Future Trends** in Transmission & Distribution Conference & Exposition, Seoul, Coréia do Sul, 2009.

VIANA, M. E. – **Ferramenta para Maximização de Carga na Fase Fluente de Recomposição de Sistemas Elétricos** - Dissertação de Mestrado, Universidade Federal de

Juiz de Fora, Brasil, 2008.

WAHAB, N. I. A.; MOHAMED, A.; HUSSAIN, A. - **Fast transient stability assessment of large power system using probabilistic neural network with feature reduction techniques** in Expert Systems with Applications, v. 38, n. 9, pp. 11112-11119, Sep, 2011.

WEI-HONG, Y.; AI-YING, D.; RUI, F; LI-FANG, Y; YAN-YAN, J. - **Urban Residential Power Load Risk Identification Based on Data Mining** in Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 2009.

WENHUI, Z.; YIXUE, S.; MIN, X.; JINGPING, L. - **State Assessment System of Power Transformer Equipments Based on Data Mining and Fuzzy Theory** in International Conference on Intelligent Computation Technology and Automation, Changsha, China, 2010.

WITTEN, I. H.; FRANK, E.; HALL, M. A. - **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations** - Morgan Kaufmann, 2005.

XU, Y & WANG, Z. - **On a Fault Detection System based on Neuro-Fuzzy Fusion Method** in Chinese Control and Decision Conference, Xuzhou, China, 2010.

XU, Y.; DONG, Z. Y.; MENG, K.; ZHANG, R.; WONG, K. P. - **Real-time transient stability assessment model using extreme learning machine** in IET Generation, Transmission & Distribution, v. 5, issue 3, pp. 314–322, 2011.

YANG, T.; LIU, J.; ZENG, D.; XIE, X. - **Application of Data Mining in Boiler Combustion Optimization** in This paper appears in: The 2nd International Conference on Computer and Automation Engineering, Singapore, Republic of Singapore, 2010.

YANG, Z.; TANG, W. H.; SHINTEMIROV, A.; WU, Q. H. - **Association Rule Mining-Based Dissolved Gas Analysis for Fault Diagnosis of Power Transformers** in IEEE Transactions on Systems, Man, and Cybernetics, v. 39, n. 6, pp. 597-610, Nov, 2009.

YUESHUN, H.; QIULIN, D. - **Fault Mode Analyze of Power System Based on Data Mining** in Proceedings of the International Symposium on Web Information Systems and Applications (WISA'09), Nanchang, China, 2009.

YUNYAN, L. - **Application of Association Rules Mining in Power Demand-Side Management Based on Rough Set** in Power Electronics and Motion Control Conference, Wuhan, China, 2009.

ZHANG, Y.; ZHANG, J.; MA, J.; WANG, Z. - **Fault Detection Based on Data Mining Theory** in International Workshop on Intelligent Systems and Applications, Yuhuan, China, 2009.

ZHENG, Y; CHEN, Q; LI, Q; LI, Y - **Actuality Analysis of Condition-Based Maintenance Technology for Hydroelectric Generating Unit** in Asia-Pacific Power and Energy Engineering Conference (APPEEC), Chengdu, China, 2010.

## ANEXO I TRABALHOS MOTIVADOS PELA TESE

- Trabalho de Conclusão de Curso de Engenharia Elétrica intitulado “Avaliação da Segurança Estática de Sistemas de Potência Utilizando Árvore de Decisão“, desenvolvido na UFPA em 2010, no Campus de Tucuruí, por Bernard Carvalho Bernardes;
- Trabalho de Conclusão de Curso de Engenharia Elétrica intitulado “Avaliação da Segurança Dinâmica de Sistemas de Potência Utilizando Árvore de Decisão“, desenvolvido na UFPA em 2010, no Campus de Tucuruí, por Werbeston Douglas de Oliveira;
- Trabalho de Conclusão de Curso de Engenharia Elétrica intitulado “Classificação de Falhas em Linhas de Transmissão a partir da Análise de Defeitos e Árvore de Decisão“, desenvolvido na UFPA em 2010, no Campus de Tucuruí, por Josias Macedo da Silva;
- Trabalho de Conclusão de Curso de Engenharia Elétrica intitulado “Uma Ferramenta para Geração de Base de Dados Aplicada a Avaliação da Estabilidade Transitória de Sistemas de Potência“ desenvolvido na UFPA em 2011 no Campus de Tucuruí, por Dieigo Sá Gaia;
- Trabalho de Conclusão de Curso de Engenharia Elétrica intitulado “Metodologia para Melhoria da Segurança Dinâmica de Sistemas de Potência Usando Inteligência Computacional“, desenvolvido na UFPA em 2011 no Campus de Tucuruí, por André Luis Barbosa Corrêa ;
- Artigo “Decision Tree-Based Power System Static Security Assessment Using PMU Measurements“, Power Tech 2011, Trondheim, Noruega, por Bernard Carvalho Bernardes, Werbeston Douglas de Oliveira, João Paulo Abreu Vieira,IVALDO OHANA, Ubiratan Holanda Bezerra e Marcus Vinicius A. Nunes.
- Artigo “Uma Metodologia para Avaliação da Estabilidade Transitória em Tempo Real de Sistemas Elétricos de Potência Usando Árvore de Decisão“, SEPOPE 2012, Rio de Janeiro, Brasil, por Dieigo Sá Gaia, André Luis Barbosa Corrêa, Bernard Carvalho Bernardes, Werbeston Douglas de Oliveira, João Paulo Abreu Vieira,IVALDO OHANA e Ubiratan Holanda Bezerra;
- Artigo “Redespacho da Geração para Melhoria da Segurança Dinâmica de Sistemas Elétricos de Potência Usando Inteligência Computacional“, SBSE 2012, Goiás, Brasil, por André Luis Barbosa Corrêa, Bernard Carvalho Bernardes, Werbeston Douglas de Oliveira, João Paulo Abreu Vieira,IVALDO OHANA e Ubiratan Holanda Bezerra;

- Artigo “Data Mining Experiments on a Hydroelectric Power Plant”, IET Generation, Transmission & Distribution, maio 2012, por: Ivaldo Ohana, Ubiratan Holanda Bezerra e João Paulo Abreu Vieira;
- Projeto de Pesquisa e Desenvolvimento (P&D) intitulado “Metodologia de Controle Preventivo Baseado em Árvore de Decisão para a Melhoria da Segurança Estática e Dinâmica do Sistema Interligado da Eletronorte”, iniciado em fevereiro de 2012 junto a Eletrobrás-Eletronorte, com prazo de execução de 24 meses.

**ANEXO II**  
**EQUAÇÕES RELACIONADAS NA TESE**

$$\text{Precisão} = \frac{\text{previsões corretas}}{\text{total de previsões}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (1)$$

$$\text{Erro} = \frac{\text{previsões incorretas}}{\text{total de previsões}} = \frac{f_{01} + f_{10}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2)$$

$$\text{Suporte ( X -> Y )} = \frac{\text{Transações que contém X e Y}}{\text{Quantidade total de transações}} \quad (3)$$

$$\text{Confiança ( X -> Y )} = \frac{\text{Transações que contém X e Y}}{\text{Transações que contem X}} \quad (4)$$

$$\text{DFP} = (1 - (n / 144)) \times 100 (\%) \quad (5)$$

onde, n = número de intervalos de 10 (dez) minutos, considerando o total de 144 intervalos diários, em que a integral do módulo do desvio de frequência (A) foi superior a 0,4 Hz.min.

$$A = \int |\Delta f(t)| \cdot dt \text{ (Hz.min)} \quad (6)$$

com:

A = Integral do módulo do desvio da frequência a cada 10 minutos,

$\Delta f$  = Desvio da frequência =  $f - f_0$ ,

f = Frequência medida (Hz),

$f_0$  = Frequência nominal de 60,00 Hz,

t = tempo.

$$P^2 + Q^2 = V_t I_{a\max} = S^2, \quad \text{com} \quad (7)$$

P = Potência ativa,

Q = Potência reativa,

$V_t$  = Tensão terminal da máquina

$I_{a\max}$  = Máxima corrente de armadura

$$P = P_{mec\_max}, \text{ com} \quad (8)$$

$P_{mec\_max}$  = Potência mecânica máxima

$$P^2 + \left( Q + \frac{V_t^2}{X_q} \right)^2 = \left( \frac{V_t E_{max}}{X_d} + V_t^2 \left( \frac{1}{X_q} + \frac{1}{X_d} \right) \cos(\delta - \theta) \right)^2, \text{ com} \quad (9)$$

$X_d$  = Reatância síncrona do eixo direto

$X_q$  = Reatância síncrona do eixo em quadratura

$E_{max}$  = Tensão interna máxima, determinada pela tensão máxima de campo.

$$P^2 + \left( Q + \frac{V_t^2}{X_q} \right)^2 = \left( \frac{V_t E_{min}}{X_d} + V_t^2 \left( \frac{1}{X_q} + \frac{1}{X_d} \right) \cos(\delta - \theta) \right)^2, \text{ com} \quad (10)$$

$E_{min}$  = Tensão interna mínima, determinada pela tensão mínima de campo

$$P^2 = - \frac{\left( \frac{V_t^2}{X_q} + Q \right)^3}{\frac{V_t^2}{X_d} + Q} \quad \text{com} \quad (11)$$

$$-\frac{V_t^2}{X_q} \leq Q \leq -\frac{V_t^2}{X_d} \quad (12)$$