



**XXII SNTPEE
SEMINÁRIO NACIONAL
DE PRODUÇÃO E
TRANSMISSÃO DE
ENERGIA ELÉTRICA**

BR/GOP/15
13 a 16 de Outubro de 2013
Brasília - DF

GRUPO - IX

GRUPO DE ESTUDO DE OPERAÇÃO DE SISTEMAS ELÉTRICOS - GOP

**TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS EM SISTEMAS SCADA
DE USINAS HIDROELÉTRICAS**

Ivaldo Ohana
UFPA

João Paulo Abreu Vieira (*)
UFPA

Ubiratan Holanda Bezerra
UFPA

Antonio Augusto Bechara Pardaul
Eletrobrás Eletronorte

RESUMO

Uma das principais ferramentas para gerenciamento de Usinas Hidrelétricas (UHEs) são os sistemas de Supervisão, Controle e Aquisição de Dados (SCADA, do inglês "*Supervisory Control and Data Acquisition*"). Esses sistemas geram e armazenam em banco de dados imensas quantidade de dados representando o histórico da operação das usinas. Um tratamento adequado nesses dados históricos pode levar a descoberta de relações e padrões, no sentido de auxiliar na compreensão de muitos aspectos operacionais importantes e na avaliação do desempenho da usina Hidroelétrica. O processo de Descoberta de Conhecimento em Banco de Dados (KDD, do inglês "*Knowledge Discovery in Database*") 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 da KDD que permite extrair informações úteis em grandes bases de dados. Neste cenário, o presente trabalho apresenta experimentos de mineração de dados no sistema SCADA da UHE-Tucuruí, 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 sistemas SCADA de UHEs.

PALAVRAS-CHAVE

Mineração de dados, Usinas hidroelétricas, Estatística, Árvore de decisão, Associações.

1.0 - INTRODUÇÃO

As usinas hidroelétricas (UHE) são supervisionadas e controladas por meio de sistemas SCADA, os quais geram grandes quantidades de dados discretos e analógicos, que ocorrem durante a operação. Esses dados são registros de valores analógicos de grandezas como tensão, corrente, potência, frequência, velocidade, pressão, temperatura, fluxo de óleo e gás, entre outros; e valores discretos de sinalizações referentes aos diversos equipamentos como bomba de óleo, bomba d'água, ventiladores, disjuntores, seccionadores, relés, etc; alarmes, indicando a ultrapassagem de limites mínimos e máximos, como limites de temperatura, limites de frequência, de potência, etc; *set points*, e comandos emitidos pelo operador ou de forma automática, como partida de gerador, abertura e fechamento de disjuntores, *set points* de tensão, de potência, comandos para inibir variáveis, troca de computadores, entre outros. Essas usinas constituem um ambiente bastante complexo, onde muitas das relações de causas e efeitos entre as variáveis operacionais não são completamente mapeadas, pois faltam ferramentas adequadas para extrair conhecimento e relações úteis a partir da grande massa de dados gerada pelos sistemas SCADA.

O uso de ferramentas adequadas de mineração de dados pode descobrir relações não mapeadas na operação da usina UHE, como, por exemplo: 1) certo tipo de alarme está ocorrendo com maior frequência em uma dada

(*) Rua Augusto Correa, n° 01 – Universidade Federal do Pará, PPGE, CEAMAZON, sala 2 (2º andar).
CEP: 66.075-970 Belém, PA, Brasil. Tel: (+55 91) 3321-8921 – Email: jpavieira@ufpa.br

unidade geradora do que em relação às outras; 2) ocorrem mais emergências na operação noturna do que na operação diurna; 3) o número de ultrapassagem de limites de temperatura em certo período de operação parece excessivo; 4) a usina está operando, na maior parte do tempo, fora da sua característica de maior eficiência; 5) a ocorrência de um alarme específico pode provocar, após decorrido certo tempo, outro alarme em outra unidade geradora; 6) os geradores operam, por alguns momentos, com ultrapassagem de limites de segurança operacional; e 7) outras relações. A extração dessas informações a partir dos dados históricos do sistema SCADA pode fornecer subsídios importantes para o planejamento da operação do sistema elétrico, com maior segurança, qualidade e economia, bem como detectar possibilidades de falhas e defeitos que podem ser corrigidos pela manutenção.

Dentro deste contexto, o presente trabalho apresenta experimentos da utilização de técnicas de mineração de dados na vasta quantidade de dados gerados pelo sistema SCADA da UHE-Tucuruí. A imensa quantidade de dados gerados pelos sistemas SCADA, pela prática atual do setor elétrico, são minimamente utilizados, concentrando-se esta utilização basicamente para as tarefas de supervisão. Com essa prática atual, é desperdiçado o grande potencial que esses dados têm de reter conhecimento implícito sobre as variáveis operacionais do processo. Dessa forma, o presente trabalho pretende mudar este paradigma ao propor o uso de técnicas estatísticas e de mineração de dados, como árvores de decisão e associações, para extrair conhecimento a partir dos dados históricos do sistema SCADA da usina UHE-Tucuruí da Eletrobras Eletronorte.

2.0 - REVISÃO BIBLIOGRÁFICA

Nos últimos anos, várias aplicações de mineração de dados foram desenvolvidas em sistemas elétricos de potência. Em [3], grandezas elétricas, como frequência, tensão, potência ativa e potência reativa, são relacionadas por meio de gráficos para a compreensão do sistema elétrico da região sul da Índia. Em [4], experiências de mineração de dados realizadas na Electricité de France analisam as características dos consumidores através de suas curvas de carga empregando a técnica de agrupamento e realizam predição de falta de dados em banco de dados por meio de regressão logística. Em [5], é proposto um método de análise das condições de um sistema de potência baseado em informações do sistema de supervisão e no algoritmo k-means para analisar as condições de operação e encontrar um valor ótimo para guiar a operação do sistema de potência. Em [6], regras de associação e árvore de decisão são utilizadas para descobrir relacionamentos entre variáveis discretas e analógicas em uma subestação de energia elétrica. Em [7], uma base de dados de possíveis causas de desligamentos não programados é enriquecida por meio de um sistema especialista, a partir do qual é construída uma rede Bayesiana com as prováveis causas dos desligamentos. Em [8], regras de associação são utilizadas para diagnóstico de faltas em transformadores de potência por meio do gás dissolvido no óleo isolante. Pode-se constatar que poucos trabalhos são encontrados na literatura descrevendo a aplicação da mineração de dados em usinas hidrelétricas, o que remete a necessidade da realização de mais estudos a respeito, para a obtenção de informações relevantes oferecidas por essa ferramenta.

3.0 - DESCOBERTA DE CONHECIMENTO E MINERAÇÃO DE DADOS

3.1. Descoberta de Conhecimento

O processo da KDD pode identificar padrões de dados válidos, novos, potencialmente úteis e desconhecidos em grandes bancos de dados [2]. A transformação de dados em informações de alto nível ajuda no processo de tomada de decisões organizacionais por meio do uso de técnicas automáticas de exploração de grandes quantidades de dados. Dificilmente a descoberta de novos padrões e relações seria realizada a olho nu, devido ao enorme volume de dados. A mineração de dados é também conhecida como o processo da KDD. Esse processo tem início com o entendimento do domínio da aplicação e dos objetivos a serem atingidos e é realizado como descrito nas suas etapas:

- a) Limpeza dos dados: retirada de ruídos e dados inconsistentes.
- b) Integração dos dados: Múltiplas fontes de dados são combinadas.
- c) Seleção dos dados: seleção dos dados relevantes para a análise.
- d) Transformação dos dados: transformação ou consolidação dos dados se necessário para a mineração.
- e) Mineração dos dados: aplicação de métodos inteligentes para extração de padrões.
- f) Avaliação dos padrões: avaliação dos padrões descobertos.
- g) Apresentação do conhecimento: apresentação e utilização do conhecimento descoberto.

3.2. Mineração de dados

A mineração de dados é a principal etapa do processo da KDD. É muito comum na literatura sobre mineração de dados, definir tarefas e algoritmos, entretanto, não existe um consenso a respeito das definições. As tarefas podem ser entendidas como tipos de relacionamentos entre os dados que são estabelecidos para a obtenção do

conhecimento e os algoritmos ou técnicas são consideradas como os fundamentos computacionais que propiciam a busca do conhecimento. O processo de mineração inicia com a escolha da tarefa de acordo com o tipo de conhecimento que se espera extrair dos dados e após a escolha da tarefa, opta-se por uma técnica para realizar a tarefa. Em [3], as principais tarefas da mineração de dados são descritas como mostradas na tabela 1. Esta tabela mostra várias técnicas de mineração associada às respectivas tarefas.

Tabela 1 - Tarefas e algoritmos associados

Tarefas	Técnicas
Classificação	Árvores de decisão, Classificadores baseados em regras, Classificadores de vizinho mais próximo, bayesianos, Rede neural artificial (RNA), Máquinas de vetor de suporte (VSM).
Associação	Apriori, FP-Growth, DHP, DIC, min-Apriori, GSP, SPIRIT.
Agrupamento	K-means baseado em protótipos, DBSCAN baseado em densidade, Agrupamento hierárquico aglomerativo, baseados em grafos.
Detecção de Anomalias	Técnicas estatísticas, Técnicas baseadas em proximidade, em densidade, em agrupamento.

3.2.1. Regras de associações

A associação procura descobrir relacionamentos interessantes entre os atributos dos registros escondidos em grandes conjuntos de dados. Os relacionamentos descobertos são representados como regra de associação na forma $X \rightarrow Y$, onde X e Y são conjuntos disjuntos de itens ou atributos, 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 . Os algoritmos mais usados para descobrir as regras de associação são: Apriori e o FP-Growth.

3.2.2. Árvore de decisão

Independentemente do procedimento usado na construção de uma árvore de decisão, sempre existem duas etapas: treinamento e teste. Na primeira etapa, um conjunto de registros de dados com a sua classe rotulada, chamada de conjunto de treino, é utilizado para a construção da árvore de decisão. Uma típica árvore de decisão binária é mostrada na figura 1.

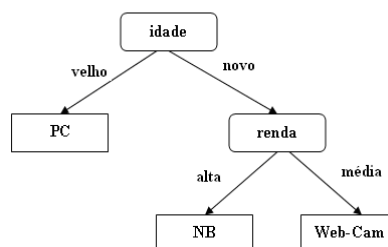


Figura 1 – Típica Árvore de Decisão

Nesta árvore, os retângulos com cantos arredondados são os nós da árvore e neles são realizados os testes nos atributos (variáveis independentes) dos dados. No caso da árvore binária, de cada nó saem duas ligações para outros nós ou folhas. Em cima da ligação é mostrado o valor testado. Os retângulos com cantos retos são as folhas, ou seja, as classes (variável alvo). Esta árvore foi construída a partir de um conjunto de dados com os atributos idade, renda e item comprado, onde se considerou como classe o atributo item comprado. Assim, se a idade for velho então o item comprado é Computador Pessoal (PC) ou se a idade for novo e renda alta então o item comprado é Notebook (NB). Na fase de teste, outro conjunto de dados rotulados, chamado de conjunto de teste é utilizado para testar a árvore criada anteriormente. Cada registro deste conjunto é submetido à árvore que determina uma classe. Esta classe encontrada é comparada com a classe do rótulo. A avaliação de desempenho de uma árvore de classificação é medida pelas contagens dos registros de teste previstos corretamente e incorretamente. Um dos algoritmos mais utilizados para a construção de uma árvore de decisão é o algoritmo de Hunt, que é a base de muitos outros algoritmos como o ID3, C4.5 e CART.

3.2.3. Agrupamentos

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 são regiões de alta densidade, separados por 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 grafos ou protótipo e o DBSCAN, baseado em densidade.

3.2.4. Detecção de Anomalias

Na detecção de anomalias, o objetivo é encontrar objetos que sejam diferentes da maioria dos outros objetos. A detecção de anomalias é também conhecida como detecção de desvios ou mineração de exceções. 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 [Haw80]: “Um *outlier* é um fato que desvia tanto de outros fatos a ponto de gerar suspeitas de que foi gerado por um mecanismo diferente”. Os algoritmos de detecção de anomalias mais utilizados são: NL (*Nested Loop*), *FindAllOutsM* e *FindAllOutsD*.

4.0 - APLICAÇÕES PRÁTICAS

Neste tópico são mostrados, de forma prática e com bastante clareza, exemplos de experimentos de mineração de dados aplicada ao sistema SCADA da UHE-Tucuruí, usando técnicas estatísticas, árvores de decisão e associações. Nos experimentos com associações e árvore de decisão optou-se por utilizar o software *RapidMiner*, o qual é um ambiente de máquina de aprendizado, mineração de dados, mineração de texto e análise preditiva, dentre outros. O *Rapidminer* é usado em pesquisa, educação, treinamento, prototipação rápida e desenvolvimento de aplicações industriais. 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.

4.1, Análise estatística de eventos da usina

O objetivo desse experimento é a realização de uma análise estatística nos eventos ocorridos durante o período 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 2.



Figura 2 - Quantidade de eventos da UHE Tucuruí no período de 17 a 21/01/2010.

Tabela 2 – Descrição dos eventos

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

As médias por hora desses eventos são mostradas na Figura 3 a seguir.

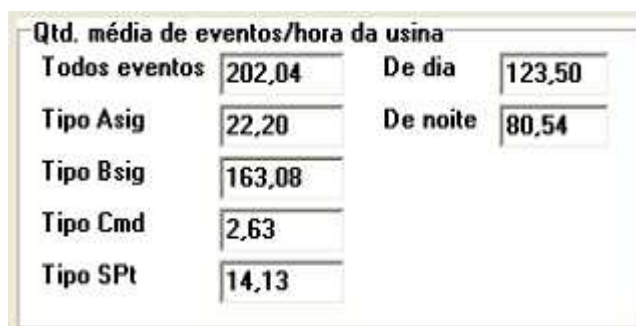


Figura 3 - Média de eventos por hora durante o período de 17 a 21/01/2010.

Houve por hora uma média de 202,04 eventos. Nota-se que ocorreram em média por hora 22,20 eventos tipo Alarme, 163,08 eventos tipo Booleano, 2,63 eventos tipo Comando e 14,13 do tipo *set point*. Uma análise desses valores deve ser realizada. A quantidade de eventos tipo Asig 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. No mesmo período analisou-se a quantidade de eventos discretos que ocorreram nos geradores da UHE Tucuruí, obtendo-se a distribuição total e horária por cada gerador como mostrado na Figura 4.

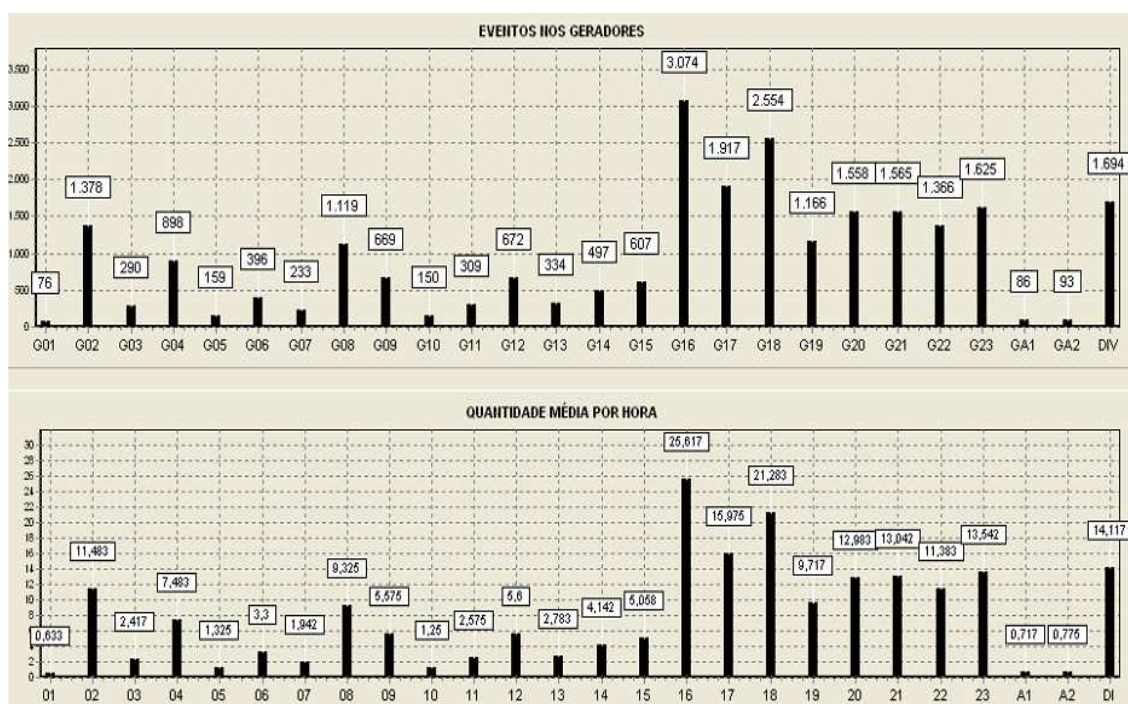


Figura 4 - Distribuição dos eventos entre os hidrogeradores no período de 17 a 21/01/2010.

Verifica-se a não uniformidade de eventos entre os hidrogeradores e a sua grande quantidade. O hidrogerador G02 é antigo e apresenta uma média horária de 11,483 eventos que é menor que a média de 25,617 eventos do hidrogerador G16, que é mais novo. Um fato preocupante que com certeza merece uma atenção especial na sua análise.

4.2. Associações entre Eventos da Usina

Neste experimento foi realizado um pré-processamento no banco de dados de históricos de eventos do período de 17/01/2010 a 21/01/2010, 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 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 *items-sets* frequentes e posteriormente regras de associações foram geradas a partir destes *item-sets*. Usando-se confiança mínima de 0,8, foram obtidas 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 consequências reais e verificar se é possível tirar algum benefício para conduzir melhor a operação.

4.3. 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 (A) dentro de cada hora. O primeiro registro representa a hora 0 do primeiro dia e último representa a hora 23 do último 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. As letras S e N indicam a presença e a ausência de alarme, respectivamente. 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 (unidades geradoras da 1ª etapa da UHE de Tucuruí) tiveram alarmes. Este arquivo foi minerado com o algoritmo *FP-Growth* para gerar os *items-sets* frequentes e então regras de associação foram extraídas dos *items-sets*. Usando confiança mínima de 0,6, duas regras listadas a seguir foram obtidas.

[G02] → [G11] (confiança: 0.625)
 [G08] → [G11] (confiança: 0.667)

As regras mostram que quando os hidrogeradores G02 ou G08 apresentam alarme, na mesma hora o hidrogerador G11 também apresentará 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 G11 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 G11 ou ao menos minimizar as consequências.

4.4. Árvore de decisão para o nível da água na cidade de Tucuruí

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 de Tucuruí (NC) a jusante da usina que é a variável alvo. Os valores desses atributos para o período de 05/01/2008 a 31/12/2010 foram extraídos do banco de dados com informações hídricas da UHE Tucuruí e armazenados em uma planilha. A vazão vertida é a quantidade de água em m³/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 de Tucuruí (NC) é o nível de água em m acima do nível normal em frente à 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 treinamento e 20% para teste. A configuração final da árvore de decisão está apresentada na Figura 5.

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

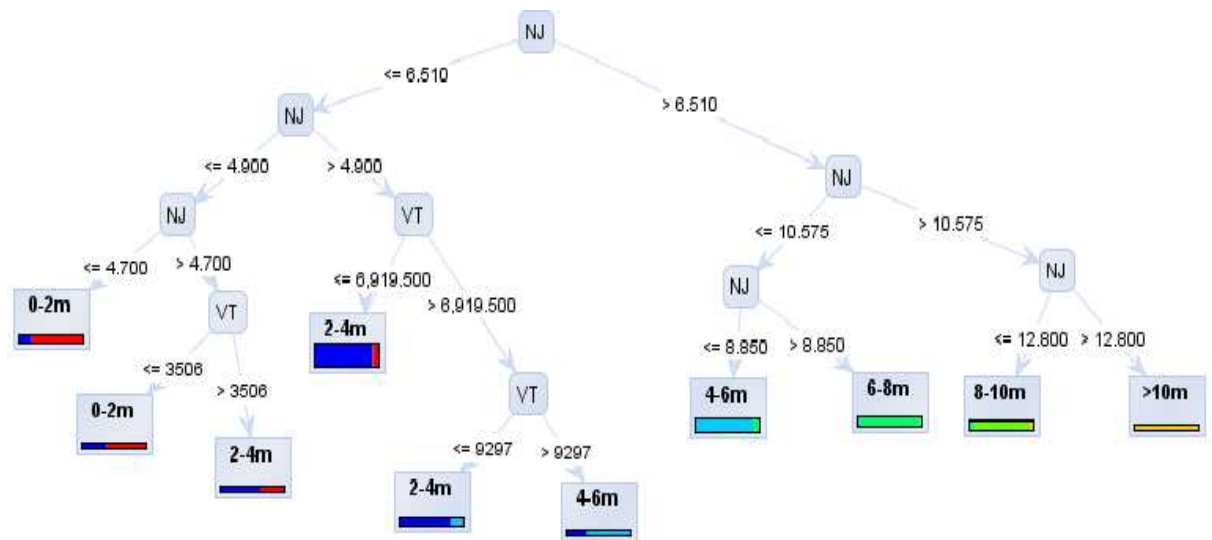


Figura 5- Relacionamento do nível da água na cidade de Tucuruí com o nível da água à jusante da barragem (NJ) e com a vazão turbinada (VT).

Pela observação da Figura 5 pode-se relacionar o nível da água na cidade de Tucuruí, a partir do nível à jusante da barragem (NJ) e da vazão turbina na Usina (VT). Por exemplo, pode-se verificar que o nível em Tucuruí atingirá uma cota no intervalo de (6-8m), se $8,850 \text{ m} < \text{NJ} \leq 10,575 \text{ m}$.

4.5. Limites Operacionais do Hidrogerador

A técnica Árvore de Decisão também poderá ser utilizada com os dados de tempo real do SCADA para caracterizar a segurança operacional dos hidrogeradores. 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 acessar a sua segurança operacional.

Geração de dados simulados para o treinamento da árvore de decisão

Uma grande quantidade de dados é necessária para se treinar uma árvore de decisão a fim de assegurar que ela aprenda todas as condições operacionais possíveis. Como na maior parte do tempo o hidrogerador trabalha em regiões seguras, existem poucos dados de pontos operacionais que tenham ultrapassado os limites operacionais para que se treinar a árvore de decisão. Assim, será necessário gerar dados que simulem a ultrapassagem dos limites impostos pela curva de capacidade e rotulá-los. Neste sentido, foi gerado um arquivo com 12.000 registros de pontos operacionais com e sem ultrapassagem de limites com os seguintes atributos principais:

- 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 2 a seguir.

Tabela 2 - 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

Treino e teste da árvore de decisão

A árvore foi treinada usando 70% do arquivo para treinamento e 30% para teste. Foi obtida uma precisão de 97,67% e um erro de 2,23% usando a métrica índice *gini*. A árvore de decisão projetada é mostrada na Figura 6.

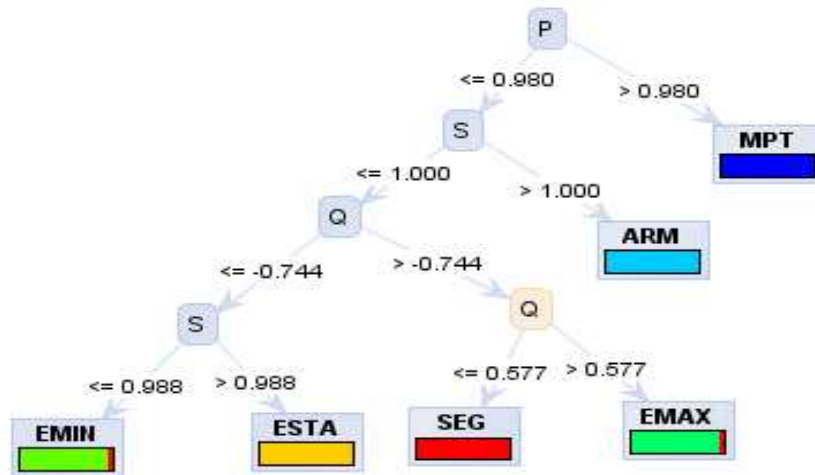


Figura 6 – Árvore de decisão para os limites operacionais.

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 limite 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

Foram usados 17.280 pontos de operação reais obtidos do banco de dados do SCADA da UHE Tucuruí, 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.

As regras da árvore de decisão 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.

5.0 - CONCLUSÃO

A característica multidisciplinar da aplicação das técnicas de mineração de dados, ou seja, a necessidade de conhecer a técnica, bem como, a área de aplicação, aliada a falta de dados rotulados para as mais diversas aplicações, é um forte fator impeditivo na utilização da mineração de dados. O setor elétrico com grandes volumes de dados gerados diariamente é motivo para a aplicação de técnicas de mineração de dados. Neste trabalho foram mostrados experimentos da aplicação de mineração de dados no setor elétrico, utilizando dados reais do sistema SCADA da UHE-Tucuruí. Um experimento de estatística foi realizado usando variáveis discretas da usina. Também foram realizadas duas experiências de associações entre os dados de eventos para se descobrir relacionamentos. Em seguida, foram realizadas duas experiências usando variáveis analógicas e árvore de decisão. Nos cinco experimentos foram utilizados dados reais e verificou-se a aplicabilidade destas técnicas em sistemas elétricos de potência. Além disto, constatou-se a grande potencialidade das árvores de decisão como métodos de classificação. Este artigo mostra aplicações práticas de técnicas de mineração de dados em sistemas de energia elétrica para ajudar a formar uma cultura de sua utilização e assim aplicar esta ferramenta em outros problemas do setor elétrico.

6.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) BOYER, S. A. – SCADA: Supervisory Control and Data Acquisition – The Instrumentation Systems and Automation Society, USA, 1999.
- (2) TAN, P.; STEINBACH, M.; KUMAR, V. – Introdução ao Data Mining: Mineração de Dados – Editora Ciência Moderna, Brasil, 2009.
- (3) DAS, S., RAO, N. – Understanding Power System Behavior through Mining Archived Operational Data – International Journal of Emerging Electric Power Systems, vol. 10, Issue 1, Article 5, 2009.
- (4) HÉBRIL, G. – Practical Data Mining in a Large Utility Company – Compstat 2000, Proceedings, 2000.
- (5) LING, L. B.; JUN, H.; MING, L. G. A Study On Conditions Analysis Of Power Plant Based On Data Mining – 2° International Asia Conference on Informatics in Control, Automation and Robotics, 2010.
- (6) SCHUCH, R.; DILLI, S. L.; SUASEN, P. S.; PADOIN, E. L.; CAMPOS, M. – Mineração de Dados em Uma Subestação de Energia Elétrica – Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications, Brasil, 7-11 junho, pp. 804-810, 2010.
- (7) 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 – Revista Controle & Automação, Vol.21 no.2, pp. 185-200, 2010.
- (8) YANG, Z.; TANG, W. H.; SHINTEMIROV, A.; WUE, Q. H – Association Rule Mining-Based Dissolved Gas Analysis for Fault Diagnosis of Power Transformers – in IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, vol 39, no. 6, pp. 597-610, 2009.

7.0 - DADOS BIOGRÁFICOS



Ivaldo Ohana, 20/01/1950, Belém-PA. Possui Graduação em Engenharia Elétrica Opção Eletrônica pela Universidade Federal do Pará (1976), mestrado em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (1979) e doutorado em Engenharia Elétrica pela Universidade Federal do Pará (2012). Atualmente é professor adjunto da UFPA, atuando nas áreas mineração de dados aplicada a sistemas de potência.



João Paulo Abreu Vieira. Possui Graduação, Mestrado e Doutorado em Engenharia Elétrica pela Universidade Federal do Pará em 2004, 2006 e 2009, respectivamente. Atualmente é Professor Adjunto 2 da Universidade Federal do Pará. Tem experiência na área de Engenharia Elétrica, com ênfase em Sistemas Elétricos de Potência, atuando principalmente nos seguintes temas: Modelagem, simulação, estabilidade, controle e segurança de sistemas elétricos. Geração distribuída, aerogeradores. Medição fasorial sincronizada e suas aplicações. É membro do IEEE PES.



Ubiratan Holanda Bezerra. Possui graduação em Engenharia Elétrica Opção Eletrônica pela Universidade Federal do Pará (1976), mestrado em Engenharia Elétrica pela Universidade Federal de Itajubá (1980) e doutorado em Engenharia Elétrica pela Universidade Federal do Rio de Janeiro (1988). Atualmente é professor titular da Universidade Federal do Pará. Seus interesses em pesquisa estão focados em avaliação estática e dinâmica da segurança de sistemas elétricos, geração distribuída, e qualidade da energia elétrica.



Antonio Augusto Bechara Pardaul. Possui graduação em Engenharia Elétrica pela Universidade Federal do Pará. Atualmente é Superintendente de Produção Hidráulica da Eletrobrás Eletronorte.