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

Tese de Doutorado

Avaliação da Segurança de Sistemas de Potência para Múltiplas Contingências Usando Árvore de Decisão Multi-Caminhos

WERBESTON DOUGLAS DE OLIVEIRA

TD 09/2017

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
Setembro/2017

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

Tese de Doutorado

Avaliação da Segurança de Sistemas de Potência para Múltiplas Contingências Usando Árvore de Decisão Multi-Caminhos

WERBESTON DOUGLAS DE OLIVEIRA

TD 09/2017

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
Setembro/2017

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

Avaliação Preventiva da Segurança de Sistemas de Potência para Múltiplas Contingências Usando Árvore de Decisão Multicaminhos

Orientador: Prof. Dr. João Paulo de Abreu Vieira

Co-orientador: Prof. Dr. Ubiratan Holanda Bezerra

Tese de doutorado submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica da UFPA para a obtenção do Grau de Doutor em Engenharia Elétrica na área de Sistemas de Energia Elétrica.

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2017

Dados Internacionais de Catalogação - na – Publicação (CIP) Sistema de Bibliotecas da UFPA

Oliveira, Werbeston Douglas de, 1987-

Avaliação da segurança de sistemas de potência para múltiplas contingências usando árvore de decisão multicaminhos / Werbeston Douglas de Oliveira.-2017.

Orientador João Paulo Abreu Vieira

Tese (Doutorado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2017.

1.Mineração de dados. 2. Sistemas de energia elétrica – modelos matemáticos. 3. Àrvores de decisão. I. Títuo.

CDD 22. ed. 005.74

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

"AVALIAÇÃO DA SEGURANÇA DE SISTEMAS DE POTÊNCIA PARA MÚLTIPLAS CONTINGÊNCIAS USANDO ÁRVORE DE DECISÃO MULTI-CAMINHOS"

AUTOR: **WERBESTON DOUGLAS DE OLIVEIRA**

TESE DE DOUTORADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA NA ÁREA DE SISTEMAS DE ENERGIA ELÉTRICA.

APROVADA EM: 15/09/2017

BANCA EXAMINADORA:

Prof. Dr. João Paulo Abreu Vieira
(Orientador – PPGEE/UFPA)

Prof. Dr. Ubiratan Holanda Bezerra
(Co-Orientador – ULHT)

Prof. Dr. Roberto Célio Limão de Oliveira
(Avaliador Interno – PPGEE/UFPA)

Prof. Dr. Djalma Mosqueira Falcão
(Avaliador Externo – UFRJ)

Prof. Dr. Rodrigo Andrade Ramos
(Avaliador Externo – USP)

VISTO:

Prof. Dr. Evaldo Gonçalves Pelaes
(Coordenador do PPGEE/ITEC/UFPA)

AGRADECIMENTOS

Primeiramente a minha família, que mesmo a distância sempre me apoiou e incentivou, à minha mãe, madrinha e irmãos.

Agradeço aos meus orientadores, João Paulo e Ubiratan, por todos os conselhos, recomendações, disponibilidade e prontidão sempre demonstradas, não apenas nas orientações acadêmicas, mas também na amizade, onde esses pequenos detalhes foram de importante relevância em minha preparação. Um agradecimento especial a Daniel Martins, por toda ajuda, amizade e conselhos e aos demais professores do CEAMAZON e PPGEE. Agradecer também aqueles que ajudaram na implementação dos programas e modelos, os amigos Bernard, Rodrigo e Dione.

Aos colegas de CEAMAZON, a todos sem exceção, sempre apoiam e colaboram com este trabalho, em especial, Allan, Felipe, Daniel, Thiago, André, Breno, Juliana, Jamile, Dieigo, Vanderson, Heitor e Diorge. Também é importante ressaltar os professores Djalma, Glauco, Tatiana e Siqueira da COPPE/UFRJ, que no curto período no qual estive lá, sempre me ajudaram na medida do possível.

Ao colegiado de engenharia elétrica da Unifap. Aos amigos da pós-graduação em física, Marcelo, Alberto, Samuel, Júlio, Tercio e Gester.

Agradeço ao PPGEE, a Capes, a CNPq e Eletrobrás Eletronorte, pelo apoio acadêmico, estrutural, financeiro e suporte.

Epigrafe

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”

Arthur Schopenhauer

Sumário

Capítulo 1 - Introdução 1
1.1 Contextualização do Problema 1
1.2 Motivações 4
1.3 Levantamento do Estado da Arte 5
1.4 Objetivo Geral da Tese 13
1.5 Contribuições da Tese 13
1.6 Lista de Publicações 15
1.7 Estrutura da Tese 17
Capítulo 2 - Segurança de Sistemas de Potência 18
2.1 Considerações Iniciais 18
2.2 Segurança Estática de SEPs 19
2.2.1 Estados de Operação dos SEPs 20
2.3 Estabilidade. 24
2.3.1 Estabilidade transitória. 26
2.4 Desafios para a Operação Segura. 28
2.5 Avaliação da segurança em tempo real (on-line). 29
2.5.1 Avaliação da Segurança Estática Online. 32
2.5.2 Avaliação da Segurança Dinâmica em Tempo Real. 34
2.6 Conclusão 37
Capítulo 3 - Mineração de Dados e Árvores de Decisão 38
3.1 Considerações Iniciais: Descoberta de Conhecimento em Base de Dados 38
3.2 Mineração de Dados 39
3.2.1 Etapas do Processo de Mineração de Dados 39
3.3 Tarefas de Mineração 41

3.4 Árvores de Decisão 42
3.5 Construção de Árvores de Decisão 44
3.6 Critérios para seleção de atributos 45
3.7 Algoritmos para construção de árvore de decisão 48
3.8 Estimação de Erro do Classificador 50
3.9 Validação Cruzada (Cross-Validation) 51
3.10 Conclusão 52
Capítulo 4 - Avaliação da Segurança de Sistemas de Potência com Árvore de Decisão Multicaminhos. 53
4.1 Introdução e Considerações Sobre a Metodologia 53
4.2 Criação da Base de Dados para treino e validação baseados em dados de Planejamento – Base de dados estática e dinâmica. 55
4.3 SAGE - EMS/SCADA - Aquisição de Dados Reais e Criação de Base de Dados 60
4.3.1 Criação e Tratamento da Base de Dados 62
4.3.2 Complemento da Base de Dados 64
4.3.3 Inclusão do Atributo Simbólico “Topologia" 68
4.4 Características da Metodologia para Monitoramento da Tensão Via Árvore de Decisão 71
4.5 Avaliação de Segurança Online via DT 76
4.6 Conclusões 79
Capítulo 5 - Resultados 80
5.1 Sistema Elétrico IEEE 39 Barras (New England) 80
5.1.1 Avaliação Estática Via MDT 82
5.1.2 Avaliação Dinâmica via MDT – Estabilidade Transitória 84
5.2 Sistema Elétrico da Eletrobras Eletronorte. 85
5.2.1 Monitoramento da Tensão Operativa 86
5.2.2 – Avaliação da Segurança Dinâmica via MDT 91

5.3 Histórico Operacional 93

Capítulo 6 - Conclusões 98

Referencias 100

Lista de Figuras

Figura 1- Estados de operação para o Sistema de Potência e suas Transições 22
Figura 2- Classificação de Estabilidade de Sistemas de Potência 25
Figura 3 - Resposta no tempo do ângulo do rotor frente a um distúrbio transitório. 27
Figura 4- – Estrutura para Avaliação de SEP em Tempo Real. 31
Figura 5 - Principais funções que um EMS adiciona ao SCADA. 32
Figura 6 - Subproblemas da análise on-line de segurança dinâmica. 36
Figura 7 - Exemplo de uma árvore de decisão 43
Figura 8 - Indução de árvore de decisão por profundidade 45
Figura 9 - Indução de árvore de decisão por largura 45
Figura 10 - Exemplo de divisão de atributos categóricos/simbólicos em uma árvore de decisão 49
Figura 11 - Exemplo de CrossValidation 52
Figura 12 - Modelo de SDSAon-line usando DT. 54
Figura 13 - Fluxograma da metodologia 60
Figura 14 - Distribuição de Topologias Durante um período de Operação 66
Figura 15 - Resumo dos Passos de Geração da DT para SDSA 70
Figura 16 - Estrutura para Mineração de dados no Rapidminer 71
Figura17 - DT Criada pela metodologia proposta 72
Figura 18 - DT Criada de Acordo com as metodologias da literatura 73
Figura 19 - Esquema da Estrutura de Utilização da DT em Ambiente de Operação 74
Figura 20 - Esquemático de uso da DT com múltiplas topologias 76
Figura 21 - Resumo da Etapa Oculta ao Operador 77
Figura 22 - Esquema do Módulo de Ações de Controle Preventivo “On-line” Baseado em Árvore de Decisão 77
Figura 23 - Esquema do Módulo de Ações de Controle Preventivo “Off-line” Baseado em Árvore de Decisão 78
Figura 24 - Tela de Observação do Sistema (SAGE) 78
Figura 25 - Diagrama Unifilar do Sistema teste IEEE 39 barras 81
Figura 26 - Caracteristicas da base de dados do sistema New England 81
Figura 27 - MDT para avaliação estática New England 82

Figura 28 - Comparação dos Modelos para avaliação estática New England.. 83
Figura 29 - MDT para avaliação da segurança dinâmica do sistemaNew England 84
Figura 30 - Comparação dos Modelos para avaliação Dinâmica New England 85
Figura 31 - Sistema Elétrico da Eletrobrás Eletronorte (Regional Pará) localizado na área azul 85
Figura 32 - Perfil de carga dia teste - Planejamento vs carga real 87
Figura 33 - MDT de segurança estática para o sistema Eletronorte para as topologias TOP02 e TOP04 e suas respectivas contingências de saida de linha 88
Figura 34 - Nomograma 2-D representando a região de segurança para a topologia TOP04 89
Figura 35 - Comparação dos Modelos para avaliação estática de tensão para o sistema Eletronorte 90
Figura 36 - Comparação dos Modelos para avaliação estática de fluxo de potência nas linhas para o sistema Eletronorte..... 91
Figura 37 - Ramo da árvore de decisão que se refere à topologia A03 do dia 08/11/2013, para fins de auxílio ao controle preventivo para garantir a estabilidade transitória..... 92
Figura 38 - Comparação dos Modelos para avaliação dinâmica para o sistema Eletronorte 93
Figura 39 -Comparação dos Modelos para avaliação estática detensão para o sistema Eletronorte considerando o histórico operacional como base de dados 94
Figura 40 - Comparação dos Modelos para avaliação estática de fluxo de potência nas linhas para o sistema Eletronorte considerando o histórico como base de dados..... 94
Figura 41 - Comparação dos Modelos para avaliação estática de limites de geração de reativo para o sistema Eletronorte considerando o histórico como base de dados..... 94
Figura 42 - Distribuição de casos por topologias durante o período analisado 95
Figura 43 - Comparação dos Modelos para avaliação dinâmica para o sistema Eletronorte considerando o histórico como base de dados..... 96

Listas de Tabelas

Tabela 1- Descrição das transições dos estados de operação 23
Tabela 2 - Características e potenciais impactos do estado atual da operação de SEP. 29
Tabela 3- Exemplo de Matriz de Confusão 50
Tabela 4 - Aplicações do SAGE/EMS 61
Tabela 5 -Topologias (Horário e Repetições) 67

RESUMO

A busca de formas eficazes para promover a operação segura de sistemas de potência e aumentar a compreensão dos operadores tem encorajado a pesquisa contínua de novas técnicas e métodos que possam ajudar nessa tarefa. Nesta tese, propõe-se uma abordagem para avaliar a segurança da operação de sistemas de potência para múltiplas contingências usando a técnica árvore de decisão multi-caminhos ou (MDT, do inglês "*Multiway Decision Tree*"). A MDT difere de outras técnicas de árvore de decisão por estabelecer, na etapa de treinamento, um valor de atributo categórico por ramo. Essa abordagem propõe o uso de topologias (contingências) como atributos categóricos. Desta forma, a MDT melhora a interpretabilidade em relação ao estado operacional do sistema de potência, pois o operador pode ver claramente as variáveis críticas para cada topologia, de modo que as regras da MDT possam ser usadas no auxílio à tomada de decisão. Essa abordagem proposta foi utilizada para avaliação da segurança em dois sistemas testes, o sistema IEEE 39 barras e o sistema da parte do Norte do Sistema Interligado Nacional (SIN), sendo que este último foi testado com dados reais de um dia de operação. A técnica proposta baseada em MDT demonstrou bom desempenho, utiliando um conjunto de regras simples e claras. Também foi realizada uma comparação dos resultados obtidos com outras técnicas baseadas em árvores de decisão e verificou-se que o MDT resultou em um procedimento mais simples para a classificação de segurança dos sistemas com boa precisão.

Palavras Chaves - Árvore de Decisão, Divisão Multi-caminhos, MDT, Segurança do Sistema de Potência, Múltiplas Contingências.

Abstract

The search for effective ways to promote the secure operation of power systems and to increase its understanding by operators has encouraged continuous research for new techniques and methods that can help in this task. In this paper, it is proposed an approach to assess power system operation security for multiple contingencies using a multiway decision tree (MDT). The MDT differs from other decision tree techniques for establishing, in the training step, one value of the categorical attributes by branch. This approach proposes the use of topologies (contingencies) as categorical attributes. In this way, it improves interpretability regarding the power system operational state, as the operator can see clearly the critical variables for each topology, such that the MDT rules can be used in aiding the decision-making. This proposal was used for security assessment of two test systems, the IEEE 39-bus system and the Northern part of the Brazilian Interconnected Power System (BIPS), and BIPS was tested with real data from one day operation. The proposed MDT-based technique demonstrated superior performance, with a set of simple and clear rules. It was also performed a comparison of the obtained results with other techniques based on DT and it turned out that MDT has resulted in a simpler procedure for power system security classification with good accuracy.

Key Words – Decision Tree, Multiway Split, Power System Security, Multiple Contingencies.

Capítulo 1 - Introdução

1.1 Contextualização do Problema

A crescente demanda por energia elétrica, a proliferação da geração eólica em grande escala nos sistemas de transmissão, os limites operacionais cada vez mais restritivos nos sistemas de transmissão e geração, entre outros fatores, aumentaram a complexidade e a incerteza da operação dos sistemas elétricos de potência (SEPs). Esse fato exige a necessidade do uso de ferramentas de avaliação de segurança mais eficientes e precisas (Liu et al., 2014; Sharifzadeh et al., 2017). Para atender a essas mudanças, o Sistema de Gerenciamento de Energia (EMS, do inglês "*Energy Management System*") é vital para o monitoramento, otimização e despacho do SEP com menor custo e maior segurança. O EMS tradicionalmente usa Sistemas de Controle, Supervisão e Aquisição de Dados (SCADA), para manter a segurança do SEP dentro dos limites especificados.

O uso cada vez maior da tecnologia dos sincrofasores ou de unidades de medição fasorial (PMUs, do inglês "*Phasor Measurament Units*") tende a contribuir para processos automáticos de avaliação de segurança de SEPs, devido à alta taxa de amostragem e a sincronização das medidas via sistema de posicionamento global (GPS, do inglês "*Global Positioning System*"). Porém, sem o devido tratamento nos dados, o uso de PMUs aumentará o tamanho dos cemitérios de dados, de forma que os dados armazenados em concentradores de dados fasoriais (PDC, do inglês "*Phasor Data Concentrator*") tenham pouca ou nenhuma utilização posterior (Miao et al, 2013).

Um dos procedimentos efetivos para aumentar os níveis de qualidade e confiabilidade e reduzir o risco de blecautes na operação de um sistema elétrico é a avaliação da segurança estática e dinâmica (SDSA, do inglês "*Static and Dynamic Security Assessment*"), por meio da qual os operadores podem tomar medidas preventivas e corretivas, a fim de garantir a segurança do sistema. A avaliação de segurança deve fornecer resultados precisos e em tempo hábil para que os centros de operação tomem as ações necessárias para mitigar os problemas nos SEPs (He et al, 2012; Kundur, 1994).

As técnicas utilizadas tradicionalmente para realização da avaliação de segurança estática (*SSA*, do inglês "*Static Security Assessment*") são baseadas na execução de um grande número de casos de fluxo de carga, que consistem de um conjunto contingências para os diversos cenários operativos do sistema. A partir do cálculo de fluxo de carga é possível determinar o estado de operação em regime permanente do sistema elétrico, a partir da monitoração dos módulos e os ângulos das tensões em todas as barras do sistema,

bem como dos fluxos de potências ativa e reativa nas linhas, e das potências ativa e reativa dos geradores, e das perdas de potências ativa e reativa (Saeh etal, 2012; Rodrigues, 2014).

As ferramentas de avaliação da segurança dinâmica (DSA, do inglês "*Dynamic Security Assessment*") utilizam tradicionalmente métodos convencionais de simulação no domínio de tempo para um conjunto de contingências, cujo processamento demanda um tempo considerável. Este fato pode comprometer a tomada de decisão por intermédio de controles preventivos ou corretivos aplicados ao SEP. Outra abordagem tida como tradicional é o uso de métodos diretos baseados em funções energia, porém esses métodos possuem limitações quanto a modelagem, principalmente no que tange as influências dos controles associados aos SEPs. Contudo, o desenvolvimento de aplicativos para SDSA ainda é um grande desafio devido à não linearidade e a alta dimensionalidade do problema, tornando praticamente impossível verificar em um prazo de tempo razoável, sem grandes investimentos em supercomputadores e computação paralela, todas as combinações de contingências em diversos cenários operativos.

No Brasil, o ONS (Operador Nacional do Sistema Elétrico) atua na tentativa de implantação e operacionalização de uma ferramenta computacional de SDSA e auxílio nos processos de programação e planejamento da operação. Os resultados obtidos através de simulações e apresentados na forma de regiões de segurança ou nomogramas vêm auxiliando as equipes de planejamento e de operação, agilizando as análises em diferentes situações operativas e apoiando de maneira segura a tomada de decisão pelas equipes da operação em tempo real (Pássaro, 2013). Contudo, os resultados obtidos pela abordagem da ferramenta adotada atualmente pelo ONS e em outros operadores de rede e concessionárias pelo mundo acabam sendo conservadores, devido ao pequeno número de cenários operativos utilizado na análise.

Outro esforço notável no país é a cooperação entre CEPEL (Centro de Pesquisas de Energia Elétrica), UFRJ (Universidade Federal do Rio de Janeiro) e UFJF (Universidade Federal de Juiz de Fora) para o desenvolvimento de uma ferramenta integrada de SDSA do SIN (Sistema Interligado Nacional) com base em computação paralela, utilizando os dados de planejamento do ONS e dados do SAGE (Sistema Aberto de Gerenciamento de Energia). Contudo, essa abordagem depende de investimentos em estrutura computacional. A UFRJ dispõe de um laboratório chamado NACAD (Núcleo Avançado de Computação de Alto Desempenho), onde estão sendo realizados estudos para a viabilização de uma ferramenta de SDSA por meio de computação paralela.

No contexto mundial, destacam-se as ferramentas de SDSA da DSATools e o projeto iTesla. DSATools é um conjunto de ferramentas comerciais para análise de sistemas de potência que fornece recursos para a avaliação compreensiva da segurança de SEPs, incluindo todas as formas de estabilidade (DSATools, 2017). O projeto iTesla (*Innovative Tools for Electrical System Security within Large Areas*) é um projeto que visa a criação de uma ferramenta integrada para estudos e análise da segurança de sistemas de potência, no sentido de melhorar a operação do sistema frente a condições operacionais cada vez mais incertas. O projeto iTesla visa a utilização de técnicas de árvores de decisão (DTs, do inglês "Decision Tree") para suporte à operação (iTesla, 2017).

Mesmo contando com avanços no processamento paralelo, o que permite fazer uma avaliação mais próxima da operação em tempo real, o custo de implantação e a quantidade de informações a serem analisadas pelos operadores ainda dificultam uma utilização ampla dessa abordagem. Outro fator importante é que os métodos analíticos utilizados em ambientes paralelizados, dificilmente fornecem todas as funcionalidades que os centros de operação gostariam de ter, que são (Moulin et al., 2004):

- Avaliação qualitativa do ponto de operação atual;
- Margens de estabilidade;
- Visualização de regiões de segurança;
- Capacidade de transferência disponível;
- Controles preventivos e/ou corretivos;
- Despacho ótimo de carga.

Entretanto, com o avanço da inteligência artificial, o uso de técnicas de aprendizado de máquinas (ML, do inglês *"Machine Learning"*) como a árvore de decisão (DT, "*Decision Tree*"), máquinas de vetor de suporte (SVM, "*Suport Vector Machine"*) e as RNAs (Redes Neurais Artificiais) é uma proposta de solução promissora no desenvolvimento de ferramentas para SDSA em tempo real. Essas máquinas de aprendizado aproveitam os estudos de simulações *offline* para relacionar os atributos (variáveis do SEP) com o intuito classificar o estado de segurança do mesmo (Diao et al., 2010). Dependendo da técnica utilizada, os resultados podem ser mais facilmente interpretados ou podem possuir maior poder de extrapolação. Com base nisso, é possível descobrir atributos críticos e inerentes ao SEP, relacionados a um objetivo (atributo alvo).

A RNA (Nima, et al. 2003; Fatih et al., 2012), os mapas auto-organizáveis (SOM) (Wehenkel et al., 1997; Singh, 2006), SVM (Kalyani, 2009; Kalyani, 2011) e a DT podem

ser muito eficazes para encontrar limites operacionais automaticamente para avaliar segurança de SEP. Nesses trabalhos, as técnicas ML classificam o estado operacional do sistema a partir de um modelo criado *offline* com base em simulações. As técnicas SVM, RNA e SOM não geram regras ou modos de visualização automaticamente para as variáveis mais relevantes do problema e seus valores limites, impondo uma análise mais complexa para elaborar um possível conjunto de regras para ajudar os operadores no processo de tomada de decisão. No entanto, a DT tem vantagens em comparação com outras técnicas de ML, pois permite relacionar um alvo (atributo de destino) com atributos críticos inerentes à segurança do SEP, por exemplo, com boa transparência (interpretabilidade). Isso é feito automaticamente criando um conjunto de regras do tipo "if-then", composto por algumas variáveis e seus respectivos limites, o que representa um número menor de variáveis para o processo de tomada de decisão,

Outro recurso relevante da DT é a capacidade de trabalhar com atributos categóricos e numéricos sem conversão explícita ou normalização de dados (Harrington, 2012), ao contrário das técnicas SVM, RNA e SOM. Devido a essas características, a DT se torna uma ferramenta rápida e versátil que pode melhorar a avaliação da segurança dos SEPs, de maneira que o operador fique atento apenas aos limites das variáveis críticas extraídas pelas regras DT, os quais representam um subconjunto menor das centenas de variáveis, geralmente analisadas para garantir um estado de operação seguro.

Assim, esse conhecimento obtido dessas máquinas de aprendizado pode ser diretamente utilizado para auxílio na avaliação da segurança e na tomada de ações preventivas, fornecendo um modelo de predição rápido para aplicações em tempo real, como o uso de árvores de decisão (Wehenkel, 1998). Por isso, as DTs têm sido bastante estudadas para aplicações de DSA devido a facilidade de interpretação dos seus resultados, o que beneficia aos operadores de SEPs na avaliação de extração de conhecimento via regras de decisão da DT.

1.2 Motivações

A necessidade de se avaliar a segurança do SEP, de maneira rápida, confiável e com maior facilidade para análise da segurança realizada por engenheiros e operadores durante a operação, justifica estudos contínuos no desenvolvimento de metodologias que combinem estes requisitos.

Os métodos baseados em DTs geralmente conseguem atender de maneira satisfatória os itens de rapidez e confiabilidade no processo de avaliação da segurança.

Contudo, os métodos propostos muitas vezes, não levam em conta a visualização dos resultados para o operador de uma maneira na qual estejam mais habituados, por exemplo a visualização da topologia (que é um importante parâmetro em ambiente de operação) não é tratado de maneira clara, e quando tratada, em geral utiliza-se uma DT para cada topologia, levando o operador a ter diversos modelos de DT disponíveis no ambiente de operação, podendo levá-lo a escolhas erradas em momentos em que haja necessidades.

Além disso, os SEPs armazenam enorme quantidade de dados, que muitas vezes não são reaproveitados. Em conseqüência do anseio por uma operação mais segura, juntamente com a disponibilidade de conhecimento escondido nos históricos operacionais dos SEPs, surge a necessidade de criar metodologias e ferramentas que utilizem o histórico operacional, histórico de eventos, oscilografia e todos os dados armazenados no SEP, para que estes sejam utilizados da melhor forma possível, eficientemente e recebendo tratamento de novas técnicas computacionais. Por exemplo, utilizar dados reais de operação, complementados com dados de simulação, pode ser uma estratégia eficiente para a melhoria do desempenho de máquinas de aprendizado e outras técnicas, a fim de tornar o método mais robusto e com menor necessidade de atualização.

1.3 Levantamento do Estado da Arte

Nos últimos anos, vários trabalhos apresentaram metodologias para avaliação tanto da segurança estática quanto da segurança dinâmica de sistemas de potência com base em técnicas de árvores de decisão, redes neurais e técnicas de ML. A presente revisão da literatura foca nas metodologias de SSA, DSA e SDSA baseadas em árvores de decisão. Trabalhos baseados em outras técnicas também são revisados para melhor evidenciar os potenciais das máquinas de aprendizado na aplicação da SDSA. A revisão aborda os primeiros trabalhos utilizando MD ou ML em SSA, DSA e SDSA e os mais recentes esforços para o uso dessas ferramentas.

Em (Wehenkel et al., 1991), os autores utilizaram as DTs para a avaliação de segurança com foco na estabilidade de sistemas de potência. As principais contribuições do trabalho foram a compactação e organização da DT sobre os dados, através do acúmulo de informações em seus nós. Os autores concluíram que a DT sempre terá partição binária caso os atributos sejam numéricos. Além disso, os autores construíram uma DT para cada contingência testada, e também testaram várias configurações para a criação da DT, com o intuito de encontrar a menor quantidade de atributos, o melhor tamanho do conjunto de treinos, os atributos críticos, bem como o menor tamanho das DTs.

Em 1993, Wehenkel e Pavella apresentaram a técnica árvore de decisão como proposta para solução de problemas de segurança de tensão e de estabilidade transitória, mostrando aspectos importantes como a capacidade de descobrir mecanismos intrínsecos que regem processos físicos, fornecendo clara descrição em termos de parâmetros tratáveis do sistema.

No ano de 1994, Hatziargyriou et al. apresentaram a aplicação de um método de inferência indutiva para avaliação de segurança estática on-line, onde para cada contingência, regras de decisão na forma de árvore de decisão foram construídas *off-line*, a partir de um conjunto de pontos de operação do SEP. As folhas da árvore de decisão mostraram se o estado do sistema foi classificado como seguro ou inseguro, de acordo com os valores testados de cada atributo. Estes atributos são as variáveis de controle do sistema. Desta maneira, para cada contingência, uma árvore de decisão foi gerada para fornecer informações sobre o estado do sistema, além de orientações para as ações de controle preventivo.

Com as possibilidades em aberto para o uso das PMUs em aplicações de controle e proteção de sistemas de potência, (Rovnyak et al., 1994) construíram uma única DT para a predição da estabilidade transitória em tempo real para todos os possíveis locais de faltas. Os autores utilizaram uma pequena quantidade de amostras das medições fasoriais pós-falta na formação da base de dados.

Em (Wehenkel et al., 1994), os autores aplicaram um método de análise da estabilidade transitória com base em DT ao sistema real da França. O trabalho mostra que é possível construir uma DT para diferentes condições, a fim de atender diferentes necessidades operacionais. Os atributos candidatos são divididos em atributos controláveis (variáveis de estado que podem ser utilizadas em tomada de ações), os quais constituem DTs mais simples; atributos observáveis (que são variáveis dependentes), os quais compõem DTs com necessidade de ferramentas de pós-processamento auxiliares para aplicações em controle; e por fim, atributos complexos, que podem ser a combinação de quaisquer variáveis associados ao problema, o que diminui a inteligibilidade da DT. Em (Chien, 1994) é utilizada a DT do tipo ID3, a qual manipula somente dados discretos, para estimar fluxos de potência em linhas e as tensões em barras.

Silva (2004) aplica um classificador não linear SVM na análise da estabilidade transitória de sistemas de potência de grande porte. Os classificadores SVMs apresentaram melhor desempenho quando o conjunto completo de variáveis de entrada foi usado, confirmando na prática que a técnica SVM possui capacidade de seleção de

atributos implícitos. Os autores mostraram que bases de dados de estudos de estabilidade já disponíveis na concessionária, contendo o conhecimento de especialistas, podem na avaliação da estabilidade transitória com base em redes neurais como um bom ponto de partida. Os autores também discutiram o problema da alta dimensão das variáveis de entrada (maldição da dimensionalidade), que é típico em sistemas de grande porte. Porém, SVM não disponibiliza de maneira automática, regras que possam auxiliar os operadores em tomada de decisão.

(Swarup e et al., 2005) propõe o uso das árvores de decisão para classificar os estados de operação dos SEPs, como seguro e inseguro, utilizando para este fim os fluxos de potência ativa (MW) em ramos, as tensões nas barras e o nível de severidade das violações. Ainda em (Alves da Silva, 2005), além de mostrar que é possível utilizar RNAs para DSA, duas limitações críticas para esta aplicação são discutidas: a opacidade da inferência (sistema caixa preta) e a maldição da dimensionalidade. Utiliza-se um problema de estabilidade de tensão para demonstrar como extrair informação qualitativa (extração de regras) através da adaptação do algoritmo VIA (Validity Internal Analysis).

Tem-se uma visão geral da utilização de redes neurais em SEPs em (Haque et al., 2005), apontando-se as vantagens e desvantagens da utilização destas redes, os principais desafios (tempo de treinamento, seleção de variáveis para treinamento, atualização de redes treinadas e integração de tecnologias) são também abordados.

Um esquema para DSA *Online* usando medições fasoriais e árvores de decisão foi proposto em (Sun et al., 2007), baseando-se em três etapas: 1º) Para a previsão no horizonte de 24 horas, uma série de condições de operação que representa a variação projetada na carga diária, juntamente com o padrão de geração baseada na sua disponibilidade, são obtidos os pontos de operação via fluxo de carga. Exaustivas simulações no domínio do tempo para ‘n-1’ contingências e para as prováveis ‘n – k’ contingências são realizadas sobre os casos gerados e armazenados em um banco de dados; 2º) A DT é treinada utilizando a base de dados provenientes das simulações. Posteriormente, os atributos críticos da DT são identificados, os quais caracterizam o desempenho dinâmico do sistema e; 3º) Em cada nó da DT é calculado uma pontuação de insegurança, caso alguma das pontuações ultrapassem os limites pré-estabelecidos, um controle preventivo pode ser projetado e realizado. Os autores propõem a atualização da DT através de previsão de carga de curto prazo (PCCP). Caso considere-se que os pontos de operação de treino não estão próximos o bastante da previsão de curto prazo, atualiza-

se a DT com novas condições operacionais (baseados na PCCP) juntamente com as anteriores.

Em 2009, (Li,Z. et al) é apresentado um algoritmo de mineração de dados para avaliação *online* da segurança de SEPs, utilizando para tanto um banco de dados de conhecimento de possíveis contingências para construção de DTs. Os autores também utilizaram PMU´s (Phasor Measurement Unit) para melhorar o desempenho de classificação das DTs, pois a diferença entre ângulos de tensões de barras adjacentes mostrou-se útil para a melhoria da precisão das DTs.

Em (Diao et al., 2009), é apresentado um método de análise estática para avaliar a segurança de tensão pós-contingência. O colapso de tensão é dado quando uma contingência resulta na divergência da solução do fluxo de potência. Basicamente, a metodologia consiste em usar múltiplas DT's com atributos diferentes para melhorar a avaliação. Caso um caminho da DT não possua um bom nível de classificação, utilizam-se esses dados e mais outros atributos para montar um mini conjunto de treino e refazer o caminho da DT. Esse procedimento tornou maior a complexidade da ferramenta, pois a DT a se adotar deve ser escolhida, os ramos da DT devem ser avaliados e o percentual de acerto de classificação deve ser verificado.

Ainda em 2009, Saeh, I. S at al, são apresentados os resultados comparativos de vários tipos de árvore de decisão para avaliação de segurança estática de SEP para diversos pontos de operação. Os resultados obtidos indicam que o método SSA com base em DT é comparável em termos de precisão e tempo computacional ao método Newton-Raphson para fim de solução de fluxo de carga.

Em (Genc et al., 2010), propõem-se uma ferramenta integrada baseada em DT para avaliação de segurança de SEP em tempo real, considerando quatro importantes problemas de segurança: violação do módulo da tensão, violação dos limites térmicos, estabilidade de tensão e estabilidade transitória. Os autores consideraram o histórico do despacho da geração e o local das contingências para a formação da base de dados.

Em (Diao et al., 2010), os autores desenvolveram esquemas de controles preventivo e corretivo com base em DT para melhoria da segurança dinâmica. O controle preventivo é realizado via o redespacho de geração e o controle corretivo é acionado como corte de carga. As DTs são utilizadas para determinar as regiões de segurança e seus limites para classificar o estado do sistema e fornecer as diretrizes a serem tomadas nas ações de controles preventivo e corretivo contra a instabilidade transitória.

Em 2010 Kamwa et al, utiliza a técnica floresta aleatória ou (RF, do inglês "*Random Forest*") para predição de instabilidades, juntamente com medidas de PMU's. As RFs são basicamente múltiplas DTs que não passam pelo processo de poda. O referido trabalho utiliza diversos índices que são calculados e associados à DT após as contingências. Embora tenham grande poder preditivo, as RFs são de difícil interpretabilidade. (Kamwa et al., 2012) compara a taxa de acerto com a transparência dos modelos baseados em mineração de dados, para tratar de predição de catástrofes baseados em PMUs, comparando DT e DTFuzzy

Uma metodologia para geração de base de dados de treinamento para árvore de decisão aplicada a avaliação da estabilidade de tensão é apresentada em (Krishina et al., 2011). A taxa de acerto da DT se reflete na qualidade dos dados gerados para o treinamento. Os autores afirmam que uma amostragem eficiente dos dados pode melhorar a base de dados de treinamento sem a necessidade de reamostragem. O método de amostragem eficiente baseia-se na técnica de redução de variância de Monte Carlo. Nesse caso, um número menor de simulações é necessário para obtenção de um classificador com desempenho desejado. Contudo, não é mostrado nenhum tratamento para balanceamento da base de dados em relação a casos seguros ou inseguros.

Uma proposta para agrupamento de contingências é apresentada em (Krishina et al., 2012). A proposta visa reduzir a quantidade de árvores de decisão treinadas, de maneira que uma DT possa atender um conjunto de contingências que apresentem comportamento similar. A metodologia baseia-se em verificar o valor da soma das entropias pós-distúrbio para cada contingência, agrupando as que possuem crescimento semelhante. A partir dessas informações, a base de dados de treino é gerada de acordo com a metodologia de (Krishina et al., 2011). A desvantagem dessa proposta de agrupamento é a grande necessidade de simulações para criar esses grupos, alem de não ser testada o quanto variações aleatórias de cargas podem influenciar esse agrupamento.

Em (Verma et al., 2012) avalia-se o status da segurança sistêmica e da coerência entre geradores utilizando redes neurais do tipo RBF (Função Base Radial), sendo composto por uma composição de duas RBFs em paralelo. O módulo RBF-1 é responsável pela classificação do sistema quanto à estabilidade, enquanto que módulo RBF- 2 avalia a coerência entre os geradores. Em (He et al., 2012) utiliza boosting para criação de DT para avaliação de segurança e avalia sua complexidade, com esse método se aumentou o número de nós de decisão.

(Khoshkhoo, 2102), utiliza o algoritmo C4.5 de DT para avaliar a estabilidade da tensão usando características pré-perturbação e pós-perturbações extraídas de medidas de PMUs suportados por um sistema de medição em grandes áreas. Neste trabalho não é explorado as características intrisecas do C4.5. O trabalho de (Amaree et al., 2013), apresenta uma metodologia para a predição de perda de sincronismo entre geradores síncronos, baseadas em DTs do tipo C4.5, utilizando como atributos, variáveis aquisitadas após a eliminação de falta. A metodologia leva em conta uma série de variáveis em diversos cenários de carga e topologias. Tais variáveis são: energia cinética, aceleração média, potência elétrica de saída, entre outras medidas detectadas por métodos diretos para o cálculo da estabilidade transitória. Assim define-se como calcular as variáveis que são usadas para treinar a DT. No entanto, nestes trabalhos, a capacidade do C4.5 para lidar com atributos categóricos e numéricos simultaneamente não é explorada.

(He et al., 2013a) propõe um projeto de DSA com pequenas DTs, com o objetivo de mitigar o impacto da perda de dados provenientes de medições de PMUs. Os autores treinaram pequenas DTs *offline* utilizando um método de subespaço aleátorio, que por sua vez explora a hierarquia de um sistema de monitoramento de área amplas ou (WAMS, do inglês "*Wide Area Monitoring System*"), as informações de localização de atributos e a disponibilidade de medições da PMU, de modo a melhorara robustez devido a falta de dados de PMUs. No modo *online*, um algoritmo para reforço (*boosting*) é aplicado para quantificar os peso dos votos de cada DT. Assim, a decisão de classificação é obtida por meio de uma votação ponderada das DTs viáveis. A tentativa de aumentar a robustez do projeto de DSA frente a falta de medições de PMUs é baseado na construção de diversas DTs, com atributos que são escolhidos aleatoriamente, gerando DTs com nós diferentes. Isso permite que as DTs sejam treinadas para possiveis atributos perdidos, já que a DT é treinada a partir dos dados que são entregues ao algoritmo.

Em (He et al., 2013b), os autores propuseram um esquema de DSA *online* utilizando pequenas DTs para lidar com cenários operativos (variação de pontos de operação e alteração de topologias) que mudam durante o horizonte de operação (24h a frente). As pequenas DTs, com a aplicação do algoritmo *boosting*, são atualizadas periodicamente pelo uso de novos casos de treino que são criados para lidar com qualquer mudança na condição de operação ou na topologia da rede durante o horizonte de operação.

Em (Lin et al., 2013), o autor faz uma comparação entre os algoritmos CART e MLP para o uso em controle preventivo, onde são citadas as vantagens e desvantagens dos algoritmos para esse fim. Afirmando que MLP possuem vantagem no formato das regras que são extraídas para o uso em controle preventivo, pois ele usa regras do tipo "ou" para MLP e regras do tipo "e". Contudo, ele não leva em consideração alguns pontos em consideração, como topologias e facilidade de utilizar as regras da DT.

(Zheng et al., 2013) utiliza árvores de regressão para predizer a margem de estabilidade de tensão e de amortecimento de oscilações no sistema. Foi testada a robustez frente a mudança de topologia, sendo que houve um aumento no erro de predição maior quando se verifica a taxa de acerto da árvore de regressão que avalia a tensão.

(Liu et al., 2014), propõe o uso de duas DTs orientadas por contingências para um módulo cooperativo de DSA e controle preventivo associado. Uma DT utiliza dados de PMUs para avaliar a segurança, enquanto a outra DT é baseada em medidas de variáveis controladas para suporte na decisão de controle preventivo. Os autores também propuseram um algoritmo de amostragem na preparação da base de dados para eficientizar o treino da DT para sistemas de potência com elevada penetração de parques eólicos e geração distribuída. (Rather et al., 2014) propõe que a DT orientada por contingência (uma DT por contigência) considere a importância de avaliar a influência de grandes centrais eólicas.

(Guo et al., 2014) propõe um framework probabilístico genérico para avaliar a precisão da previsão online da estabilidade transitória com base em medidas de PMUs e mineração de dados. DT é usada como exemplo de uma técnica de mineração de dados. Os ângulos do rotor do gerador e as velocidades que são fornecidas pelas PMUs durante a condição pós-falha são escolhidas como preditores. Neste trabalho, a perturbação trifásica é feita nas diferentes barras do sistema somente no nível de carga nominal do sistema teste.

(Zhongbo et al., 2015) usa árvore de decisão ID3 para gerar um modelo operacional baseado em regras para otimizar o sistema de reservatórios ao longo do ano. A base de dados é gerada pelo histórico de dados e decisões tomadas anteriormente. Por usar ID3é necessário fazer discretizações na base de dados. (Mohammadi et al., 2015; Gholami et al., 2016) utilizam análise de componentes principais para reduzir a dimensionalidade dos dados passados a DT na etapa de treinamento. Contudo, as DTs resultantes são binárias sem destaque para qual topologia ou contingência está sendo avaliada pelo modelo de DT.

(Guo et al., 2016), propõe uma metodologia em dois estágios para a identificação on-line da assinatura dinâmica do sistema de potência usando medidas de PMUs e mineração de dados. O primeiro estágio prever o status da estabilidade transtória utilizando a classificação binária tradicional para identificar a estabilidade transitória. O segundo estágio prever a natureza do comportamento dinâmico instável, aplicando agrupamento hierárquico para definir padrões de comportamento dinâmico instável de geradores, em seguida, aplica diferentes técnicas de classificação multiclasses, incluindo árvore de decisão, árvore de decisão de ensembles e máquina de vetor de suporte para identificar o comportamento. Contudo, além de aumentar a complexidade das análises de resultados, surge o problema de desbalanço entre classes no segundo estágio.

(Costa et al., 2016) propõe o uso das regras da DT tipo CART como restrições de segurança para um despacho ótimo de redes elétricas e de gás de modo integrado. É treinada uma DT para uma combinação de contingências simultâneas na rede de gás e elétrica. Sendo treinada uma DT para cada combinação de contingências.

(Papadopoulos et al., 2017) propõe uma estrutura probabilística para a identificação online do comportamento dinâmico pós-perturbação de SEP com energia renováveis. Sendo esta estrutura baseada em árvores de decisão e agrupamento hierárquico para incorporar incertezas associadas a operação. São treinadas múltiplas DTs, esta opção por múltiplas DTs aumenta a complexidade para a avaliação dos resultados por operadores, caso se deseje mais informações além do status de segurança.

(Supreme et al., 2017) desenvolve uma abordagem para detecção de instabilidade dinâmica, baseada em centro de potência para o desenvolvimento de indices de estabilidade que servirão como rótulos para o treino de random forest para a predição da estabilidade pós-contingência.

Os principais algoritmos encontrados na literatura para gerar modelos DTs são ID3, CART, C4.5 e conjuntos de DT. Algumas características desses modelos são: ID3 manipula apenas dados discretos e é muito sensível ao ruído. CART tem uma característica de sempre executar divisões binárias em suas regras criadas, independentemente do tipo de atributo. O CART é o mais utilizado na literatura, uma vez que diversos softwares comerciais o possui implementado, seja o CART simples ou se baseando em ensembles de DT. Contudo, apesar de manter parte das características DT, os ensembles aumentam a complexidade da análise de resultados. Reduzindo a transparência do modelo e, conseqüentemente, aumentando o número de análises para o

operador, uma vez que as divisões DT são binárias, e assim o número de nós (divisões ao longo dos caminhos de DT) aumenta com a aplicação dessas técnicas.

O C4.5 permite dividir um ramo da árvore de decisão para cada atributo categórico no banco de dados, que é chamado de Árvore de Decisão Multi-Caminhos (MDT, do inglês "Multiway Decision Tree"), enquanto que para atributos numéricos é usado um modo binário. Essa característica não é utilizada nos trabalhos relatados na literatura, sendo que esta caracteritíca é explorada e avaliada nesta tese de doutorado.

1.4 Objetivo Geral da Tese

O objetivo principal dessa tese de doutorado é apresentar uma metodologia baseada em árvores de decisão multi-caminhos (MDT), orientada por topologia, para avaliação da segurança estática e dinâmica de sistemas de potência. A metodologia proposta consiste em induzir o modelo da DT, de modo que cada topologia do SEP esteja associada a um único e exclusivo ramo da MDT. A metodologia foi testada no sistema IEEE 39 barras e no subsistema Norte do SIN.

1.5 Contribuições da Tese

A contribuição primária dessa tese de doutorado consiste de uma metodologia para avaliação de segurança de sistemas de potência baseada em árvores multi-caminhos. O objetivo dessa abordagem é dar ao operador um modelo simples para avaliar a segurança do sistema em tempo real, a partir de regras simples e com um número reduzido de variáveis para serem monitoradas. Sendo que estas regras podem também serem utilizadas para fins de controle preventivo do sistema.

A MDT facilita a interpretabilidade do problema de avaliação de segurança do sistema de potência, sem comprometer a precisão da classificação. Desta forma, a MDT permite representar topologias planejadas por um período de tempo e suas respectivas contingências (n-k), em um único DT. As topologias são uma característica física importante dos sistemas de energia e são tratadas como atributos categóricos. A MDT é orientada por topologia, o que permite que o operador avalie o ponto de operação em tempo real em um ramo do MDT e o status de segurança, devido à ocorrência de uma contingência pré-definida, por exemplo, em outro ramo. Esta estratégia simplifica o processo de verificação SSA e DSA, devido à simplicidade de visualização dos limites para cada variável contida na MDT. Esta abordagem é interessante porque, de acordo com

especialistas em ciência da computação, os usuários confiam nos resultados de DTs quando as regras criadas fazem sentido em seus domínios de especialização, especialmente nos casos em que os atributos mais relevantes, no ponto de vista do usuário, aparecem em posições mais próximas ao nó raiz da DT (Pener, 2013; Pener, 2014).

Outra característica importante da abordagem proposta nessa tese é o uso de amostragem estratificada na formação do banco de dados. Isso garante uma distribuição relativamente simétrica de casos seguros e inseguros para cada topologia, o que melhora os resultados de MDT.

Essa abordagem é testada no sistema teste de 39 barras do IEEE e no subsistema Norte do Sistema Interligado Nacional (SIN) e sua eficácia é avaliada usando inclusive dados reais do de um dia de operação do subsistema Norte do SIN.

Os produtos e contribuições secundárias desta tese são divididas em contribuição para o modo offline e modo online. Sendo que estão são listadas a seguir.

Contribuições do modo *Offline:*

- Desenvolvimento de metodologia para aproveitamento de dados reais do histórico operacional provenientes do estimador de estado.
- Desenvolvimento de metodologia e de ferramentas para criação de uma assinatura topológica, a partir dos dados do configurador de rede. A topologia do SEP é representada por um código (variável simbólica inserida na base de dados para treino da árvore de decisão), que é o estado lógico dos equipamentos do sistema.
- Desenvolvimento de estratégia baseada em dados de planejamento para a criação da base de dados para treino da MDT.
- Desenvolvimento de estratégias para a criação da base de dados para treino da DT, onde os dados do passado recente da operação são utilizados como pontos de operação iniciais que juntamente com os dados de planejamento (quando disponíveis). A partir desses dados iniciais, são gerados novos casos com variações de cargas para se cobrir um maior número de possíveis casos, possibilitando assim prever diversos cenários operativos. Respeitando a condição de conter a mesma proporção de casos para cada topologia apresentada para treino da DT.
- Aproveitamento dos casos reais e complementares para a criação de casos com saídas de linha, gerador, transformadores, etc.

- Desenvolvimento de estratégias e ferramentas para execução das simulações dinâmicas, a partir dos casos de fluxo de carga, a fim de verificar a segurança do sistema frente a grandes perturbações, executados no programa ANATEM.
- Desenvolvimento de ferramentas para conversão dos relatórios dos programas ANAREDE e ANATEM, em arquivos que possam ser utilizados pelo programa minerador.

Contribuições do modo *Online:*

- Para o modo *Online*, a principal contribuição com a utilização da DT é a velocidade de resposta do modelo, ideal para a operação *online* de sistemas de potência, devido a sua simplicidade, na forma de "regras se então", facilitando assim os processos de decisão em ambiente de operação das empresas de energia elétrica.
- Uso de uma única DT para o monitoramento *online* do módulo da tensão e para a avaliação preventiva da segurança estática frente a contingências críticas que possam ocorrer durante a operação normal do SEP. Essa estratégia dinamiza o processo de operação.
- A orientação por topologias permite uma redução ainda maior do número de variáveis que é avaliado pela DT. Essa redução não impacta de forma negativa no modelo, pois as áreas de busca na geração das regras da DT ficam confinadas a cada topologia. Essas características ficam ocultas ao operador, onde a informação que ele receberá é apenas a chave topológica e as variáveis críticas e seus limites para manter o sistema em segurança.
- Possibilidade de atualizar somente apenas um ramo da DT, caso haja a necessidade, sem que seja necessário treinar a DT por completo. Essa atualização pode ser feita para topologias não vistas ou casos no qual a previsão de carga não seja satisfatória. A atualização pode ser feita por meio do treino de uma DT para a topologia necessária.

1.6 Lista de Publicações

Alguns resultados relacionados ao desenvolvimento da presente tese foram publicados em conferencias nacionais e internacionais, tais trabalhos são:

- Oliveira, W. D., Gaia, D. S., Monteiro, F., Rodrigues, B. G. D., Vieira, J. P. A., Bezerra, U. H. “Comparação Dos Algoritmos C4.5 E MLP Aplicados A Avaliação Da Segurança Dinâmica e ao Controle Preventivo No Contexto Da Estabilidade Transitória”, The 10th CLAGTEE, Vina Del Mar, 2013.

- Rodrigues, B. G. D., Gaia, D. S., Bernardes, B. C., Vieira, D. J. A., Oliveira, W. D., Vieira, J. P. A., Bezerra, U. H., Castro, V. ”Monitoramento da Segurança Estática em Tempo Real de Sistemas Elétricos Utilizando Árvore de Decisão”, SIMPASE, Belo Horizonte, 2013.

- Werbeston D. Oliveira, Bernard C. Bernardes, Dione A. Vieira, João P. A. Vieira, Daniel A. Martins, Ubiratan H. Bezerra, Vilson Castro - Ferramenta Integrada Em Ambiente De Operação Em Tempo Real: Uma Abordagem Baseada em Árvore de Decisão para Avaliação da Segurança Estática e Dinâmica de Sistemas de Potência – XXIII SNPTEE, Foz Iguaçu, 2015.

- Werbeston D. Oliveira, Bernard C. Bernardes, Dione A. Vieira, João P. A. Vieira, Daniel A. Martins, Ubiratan H. Bezerra, Benedito Das G. Rodrigues, Vilson Castro - A Decision Tree-Based Power System Static and Dynamic Security Assessment Integrated in a Real-Time Operation Environment, SBSE, Natal, 2016.

- Ubiratan H. Bezerra, João P. A. Vieira, Werbeston D. Oliveira, Daniel A.Martins, Dione J. A. Vieira, Bernard C. Bernardes, Benedito G. D. Rodrigues, Vilson Castro - Metodologia De Controle Preventivo Baseada Em Árvore De Decisão Para A Melhoria Da Segurança Estática E Dinâmica Do Sistema Interligado Da Eletronorte - Ix Congresso De Inovação Tecnológica Em Energia Elétrica – Ix Citenel, João Pessoa, 2017.

A presente tese de doutorado conta com uma publicação em periódico internacional Qualis Capes A1:

- Werbeston D. Oliveira, João P.A. Vieira, Ubiratan H. Bezerra, Daniel A. Martins, Benedito das G. Rodrigues. - Power System Security Assessment for

Multiple Contingencies Using Multiway Decision Tree, Electric Power Systems Research, Volume 148, 2017, Pages 264-272, ISSN 0378-7796.

1.7 Estrutura da Tese

No capítulo 1 é realizada a contextualização do problema da avaliação de segurança de sistemas de potência, mostrando o panorama atual da avaliação da segurança on-line, bem como uma revisão da literatura. Conseqüentemente são apresentados os objetivos da proposta e principais contribuições pretendidas.

No capítulo 2 são apresentados os aspectos de segurança dos sistemas elétricos de potência, com ênfase nos conceitos relacionados ao problema de segurança, bem como aos estados operacionais e suas transições que podem ocorrer durante a operação normal. São também abordados neste capítulo os principais desafios e perspectivas da avaliação da segurança dinâmica.

No capítulo 3 são apresentados os conceitos de mineração de dados e descoberta de conhecimento em base de dados (KDD), com uma descrição mais detalhada da técnica árvore de decisão, bem como os aspectos mais relevantes e a sua estrutura de formação para aplicação neste trabalho.

No capítulo 4 é apresentada a metodologia de avaliação da segurança estática e dinâmica de SEPs com base em árvore de decisão multi-caminhos, com o detalhamento do modelo utilizado para a criação da base de dados rotulada utilizada para indução das árvores de decisão.

No capítulo 5 são apresentados resultados referentes a aplicação das MDTs para avaliação de segurança, no sistema teste do IEEE 39 barras e a partir dos dados reais do sistema elétrico da Eletrobras Eletronorte (Regional Pará).

No capítulo 6 são apresentadas as conclusões deste trabalho e sugestões de trabalhos futuros.

Capítulo 2 - Segurança de Sistemas de Potência

2.1 Considerações Iniciais

A segurança de um sistema de potência é entendida como a capacidade do sistema de energia para sobreviver a contingências plausíveis sem interrupção do serviço ao cliente, isto é, suportar a ocorrência de um conjunto de perturbações. Segurança e estabilidade de um SEP são termos altamente relacionados, já que estabilidade é um importante fator para segurança dos SEPs. Contudo, a segurança não se restringe apenas a estabilidade, também engloba a integridade de um sistema de energia e a avaliação do estado de equilíbrio do ponto de vista de sobrecargas, violações de tensão e variação da frequência do sistema (Machowski et al, 2008). A segurança é influenciada pela condição operativa como patamar de carga, despacho e intercâmbio de geração, além da topologia da rede (KUNDUR, 2004).

Os SEPs são normalmente interligados e de dimensões continentais. Por isso, eles não operam em estado de segurança absoluta, isto é, com total ausência de risco, haja vista que na prática este objetivo não é possível devido aos altos investimentos em equipamentos, que resultam no aumento dos custos operacionais que seriam necessários para alcançar níveis elevados de confiabilidade, além do crescimento exponencial do número de possíveis combinações de falhas. Compatibilizar a segurança e a economia impõe grandes desafios às equipes de engenharia na busca por projetos cada vez mais seguros e com menor custo.

Abordar o problema de maneira mais flexível, possibilita operar os sistemas elétricos de maneira mais econômica, com um nível de confiabilidade aceitável de forma a atender as necessidades dos consumidores e as exigências dos órgãos reguladores. Assim os sistemas elétricos são projetados para operar com uma margem de segurança aceitável para ocorrência de determinadas contingências, sem causar grandes transtornos para os clientes.

Entretanto, mudanças nos cenários de operação podem ocorrer mesmo em condições normais de operação, causadas por variações de carga, mudanças na topologia da rede, mudança nos fluxos de intercâmbio entre áreas. Tais mudanças demandam atenção permanente dos operadores nos centros de operação e controle, já que elas podem afetar a qualidade da energia elétrica fornecida por violação de qualquer um dos limites operacionais pré-estabelecidos. Por esse motivo o monitoramento e a adoção de medidas preventivas são necessários para que sua operação fique dentro das faixas normais de

operação. Essa análise de cenários operacionais, ou seja, das condições do sistema em regime permanente é chamada de avaliação de segurança estática (SSA, do inglês Static Security Assessment). Por outro lado, a segurança dinâmica (DSA, do inglês Dynamic Security Assessment) engloba outros aspectos relativos a estabilidade do sistema, tais como estabilidade transitória, estabilidade a pequenas perturbações e estabilidade de tensão, que também demandam grandes esforços das equipes de engenharia.

A grande complexidade dos atuais SEPs exige das ferramentas de avaliação da segurança estática e dinâmica (SDSA), a necessidade de permanente supervisão quanto ao estado de operação por meio do monitoramento contínuo das grandezas elétricas da rede. A avaliação de segurança pode ser realizada a partir de uma lista de contingências prováveis, onde é simulada a ocorrência de cada uma dessas contingências, e ao final do processamento informações ou diretrizes podem ser oferecidas ao operador, no sentido de aumentar o nível de segurança do sistema.

A SDSA geralmente tem como base critérios pré-definidos no que se refere às análises e métodos utilizados na determinação do estado de segurança (seguro ou inseguro), bem como seleciona quais contingências devem ser avaliadas. Esta avaliação pode ser classificada como estática ou dinâmica e depende da modelagem e das técnicas utilizadas para se enquadrar em um dos tipos. Na presente proposta são realizadas abordagens dos problemas relacionados à avaliação da segurança estática (violação do módulo da tensão) e dinâmica (estabilidade transitória) de SEPs.

2.2 Segurança Estática de SEPs

A segurança estática refere-se à habilidade de um SEP de operar dentro dos limites especificados e com qualidade de suprimento após uma contingência depois que os dispositivos de controle automático de ação rápida restabelecerem o equilíbrio de carga do sistema (Pal, 2010).

O sistema é considerado estaticamente seguro se os níveis de tensão em todas as barras (ou em um conjunto pré-definido de barras) da rede estiverem dentro da faixa aceitável e se os fluxos nos circuitos de transmissão e nos transformadores se encontrarem dentro dos seus limites térmicos (Carpentier, 1993) ou outros limites estabelecidos de acordo com a conveniência operativa. A SSA trata da análise, classificação, seleção e avaliação de contingências onde, para um dado ponto de operação, os critérios de segurança de regime permanente devem ser respeitados, mesmo quando houver perda de algum elemento do sistema.

As ferramentas de SSA de SEPs, normalmente, se resumem aos algoritmos de fluxo de carga. Essas rotinas de fluxo de carga possuem como desvantagem a impraticabilidade de analisar todas as possibilidades de saída de circuitos (contingências), devido à restrição de tempo em um ambiente tempo-real. Para superar estas dificuldades, nas últimas décadas, várias técnicas de inteligência computacional vêm sendo estudadas com o objetivo de possibilitar a realização da avaliação de segurança sem grandes esforços computacionais, tornando as análises muito mais rápidas e compatíveis para uso em um ambiente de tempo real.

2.2.1 Estados de Operação dos SEPs

A partir do modelo da rede elétrica e dos fasores de tensão em cada barra, um ponto de operação do SEP pode ser calculado verificando o seu estado estático operacional. Um sistema elétrico pode assumir três possíveis estados, chamados de Estado Normal, Emergência ou Restaurativo, que dependem da natureza da mudança ocorrida nas suas condições de operação (Dy Liacco, 1974).

Adicionalmente a esta classificação tem-se o chamado Estado de Alerta. A esses estados, foram associados três conjuntos de equações genéricas (Fink & Carlsen, 1978), sendo uma equação diferencial e duas equações algébricas, que governam a operação dos SEPs. O conjunto de equações diferenciais representa o comportamento dinâmico dos componentes do sistema, enquanto que os dois conjuntos de equações algébricas compreendem as restrições de carga e as restrições de operação.

As restrições de carga representam as injeções de potência ativa e potência reativa nas barras de carga, assim como as injeções de potência ativa nas barras de geração. Elas estão relacionadas ao equilíbrio entre a carga e a geração. O termo "restrição de carga" significa que o sistema elétrico deve atender a toda a demanda, e corresponde às equações de fluxo de carga em regime permanente, e portando, é uma restriçãode igualdade. As restrições d carga são representadas pela equação:

$$g(x,u) = 0 \qquad (1)$$

Onde.

g é um vetor de funções algébricas não lineares;

x as variáveis de estado e u as variáveis de controle.

As restrições de operação representam a necessidade de que os valores operacionais das variáveis operacionais do sistema, como as magnitudes das tensões nodais, os fluxos de potência aparente nas linhas de transmissão, transformadores, e as injeções de potência reativa nas barras de geração não podem exceder os limites permitidos de operação e, portanto, são restrições de desigualdade representadas pela inequação:

$$h(x,u) \leq 0 \qquad (2)$$

Onde,

h - vetor contendo o conjunto de restrições de operação.

Adicionalmente, pode-se ainda acrescentar ao conjunto de restrições de carga e operação, as denominadas *restrições de segurança*, que possibilitam avaliar a robustez do sistema elétrico frente a ocorrência de um conjunto preestabelecido de contingências mais prováveis.

Os limites operacionais do sistema elétrico e de seus equipamentos, representados pelas restrições de operação, em muitos casos não são rígidos pois, se necessário, podem ser violados durante certo intervalo de tempo. Em situações de emergência, por exemplo, onde as variáveis possam permanecer temporariamente fora dos limites.

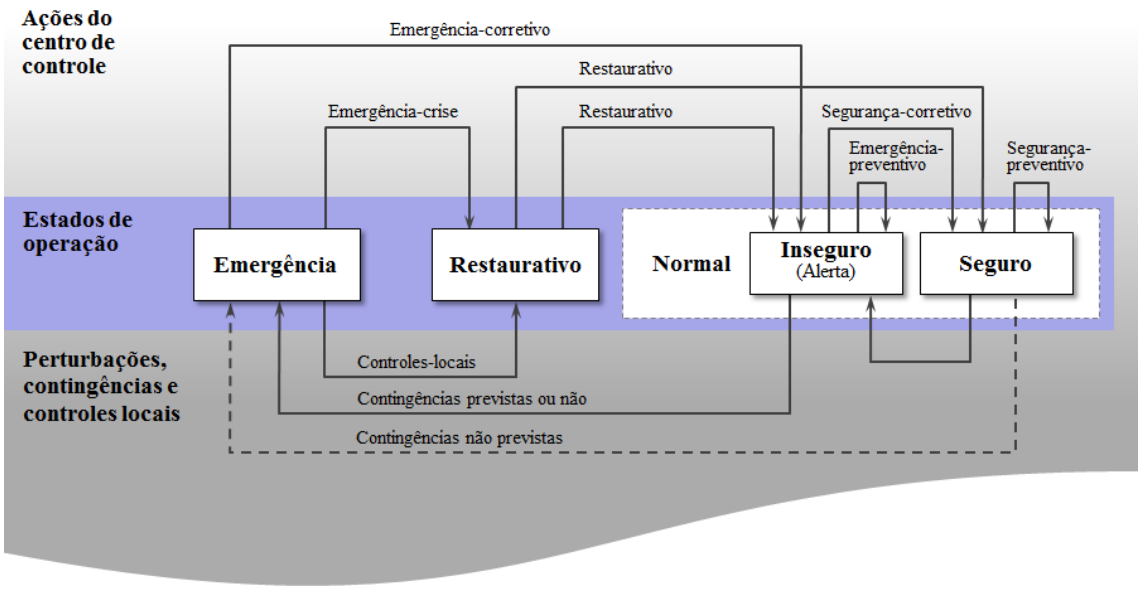
Com as definições das restrições de carga, de operação e de segurança, quatro estados de operação dos sistemas elétricos de potência podem ser definidos de acordo com o nível de atendimento dessas restrições, são eles: Estado Seguro, Alerta, Emergência e Restaurativo.

- Seguro - Também chamado de normal seguro. Todas as restrições de operação e de segurança são respeitadas, isto é, o sistema está operando normalmente, suprindo toda a carga e sem violação dos limites de operação. Mesmo que aconteça uma das possíveis contingências listadas, o sistema continuará suprindo as cargas sem violar nenhum limite operacional. É comum adotar o critério n-1 como lista de contingência.
- Alerta - Neste estado, também chamado de normal alerta, o sistema continua suprindo a demanda de carga e são atendidas as restrições de carga e operação, porém, nem todas as restrições de segurança são atendidas caso haja a ocorrência de alguma contingência da lista.

- Emergência - Este estado é caracterizado pela violação das restrições de operação. A emergência pode ser provocada por uma contingência e consequente desligamento de um ou mais componentes do sistema (linhas de transmissão, geradores, transformadores, etc). O sistema permanece operando a exceção dos componentes que provocaram a emergência.
- Restaurativo - Este estado é atingido quando uma emergência é eliminada por meio do desligamento manual ou automático de partes do sistema efetuados pelo centro de controle ou por dispositivos locais. Neste estado as restrições de operação são atendidas. Porém, o sistema não estará intacto pois restarão cargas não atendidas, ilhamento, etc.

Pode-se ainda observar as transições entre os estados em um SEP. Essas transições são decorrentes de perturbações ou mesmo resultado de ações de controle, sendo que algumas delas são funções dos centros de supervisão e controle. Os operadores realizam as ações de controle nos centros operação, com o suporte de instruções de operação e programas computacionais, necessários para que sejam tomadas as medidas mais adequadas que cada caso exige. Na tabela 1 estão descritas as várias transições possíveis e a figura 1 ilustra essas transições.

Figura 1- Estados de operação para o Sistema de Potência e suas Transições



(Adaptado, Rodrigues 2014)

Tabela 1- Descrição das transições dos estados de operação

Transição:		**Descrição da transição**
De	**Para**	
Seguro	**Alerta**	Pode ocorrer devido a variações na demanda de carga do sistema ao longo do dia ou por uma contingência.
Alerta	**Seguro**	É feita pela execução da função de controle de segurança (*corretivo*), realizada pelo operador do sistema, com uma ação de controle, como redespacho de geração.
Seguro	**Seguro**	O estado do sistema poderá ser alterado para um ponto de operação de carga futura (por exemplo, 1 hora), prevenindo que ele venha a entrar em estado de alerta.
Alerta	**Emergência**	Ocorre quando o sistema está no estado de alerta e uma das contingências responsável por este estado realmente ocorre.
Emergência	**Alerta**	O sistema pode sair do estado de emergência para alerta. Para isso devem ser eliminadas as violações de limites por meio de ações de controle, sem que ocorram cortes de carga.
Emergência	**Restaurativo**	Esta ação é tomada quando não é possível corrigir as violações com redespacho de geração ou controle de tensão, ou seja, quando é necessário o desligamento de cargas ou circuitos. Ocorre quando não é possível a transição do sistema do estado de emergência para alerta.
Alerta	**Alerta**	O ponto de operação do sistema é alterado para que não entre em estado de emergência.
Restaurativo	**Alerta ou Seguro**	Possui a função de religar cargas e circuitos desligados durante uma emergência, levando o sistema de volta ao estado seguro ou, pelo menos, ao estado de alerta.

2.3 Estabilidade.

Apesar de todas as precauções tomadas na etapa de planejamento e concepção de um SEP, a ocorrência de algumas contingências e posterior sequência de acontecimentos pode conduzir a situações de interrupção parcial ou total do sistema, consequências de problemas de estabilidade, que estão associados à segurança dinâmica do sistema.

Entre os múltiplos eventos que podem provocar o colapso do sistema, destacam-se os seguintes: o aumento súbito de carregamento numa ou em várias linhas de interligação de uma rede elétrica com redes vizinhas; produção insuficiente de potência reativa, que provoca a degradação do nível de tensão; produção deficitária de potência ativa originada pela saída de serviço de uma unidade geradora do sistema por uma queda do valor da frequência; defeitos severos em algumas linhas ou barramentos da rede; e situações de instabilidade que conduzam à perda do sincronismo. Todos estes fenômenos, direta ou indiretamente, conduzem a problemas de estabilidade na rede elétrica (Ferreira, 2005).

A elevada dimensionalidade e a complexidade dos SEPs tornam a simplificação e análise dos problemas específicos com grau de detalhe adequado uma tarefa essencial.

Os estudos de estabilidade transitória analisam o comportamento do sistema durante os primeiros segundos após a ocorrência de grandes perturbações. O sistema é dito transitoriamente estável caso mantenha o sincronismo entre as máquinas, retornando ao ponto de operação seguro estaticamente, muitas vezes próximo ao ponto de operação pré-falta.

O problema da análise da estabilidade de um SEP, atualmente, é dividido em três grandes classes (Kundur, 1994), figura 2. Esta classificação tem por base os seguintes fatores: o fenômeno que caracteriza o tipo de instabilidade e as causas físicas que conduzem à sua ocorrência.

A estabilidade angular é conceituada como a capacidade que as máquinas síncronas interconectadas a um SEP possuem em manter o sincronismo, quando o sistema é submetido a uma perturbação.

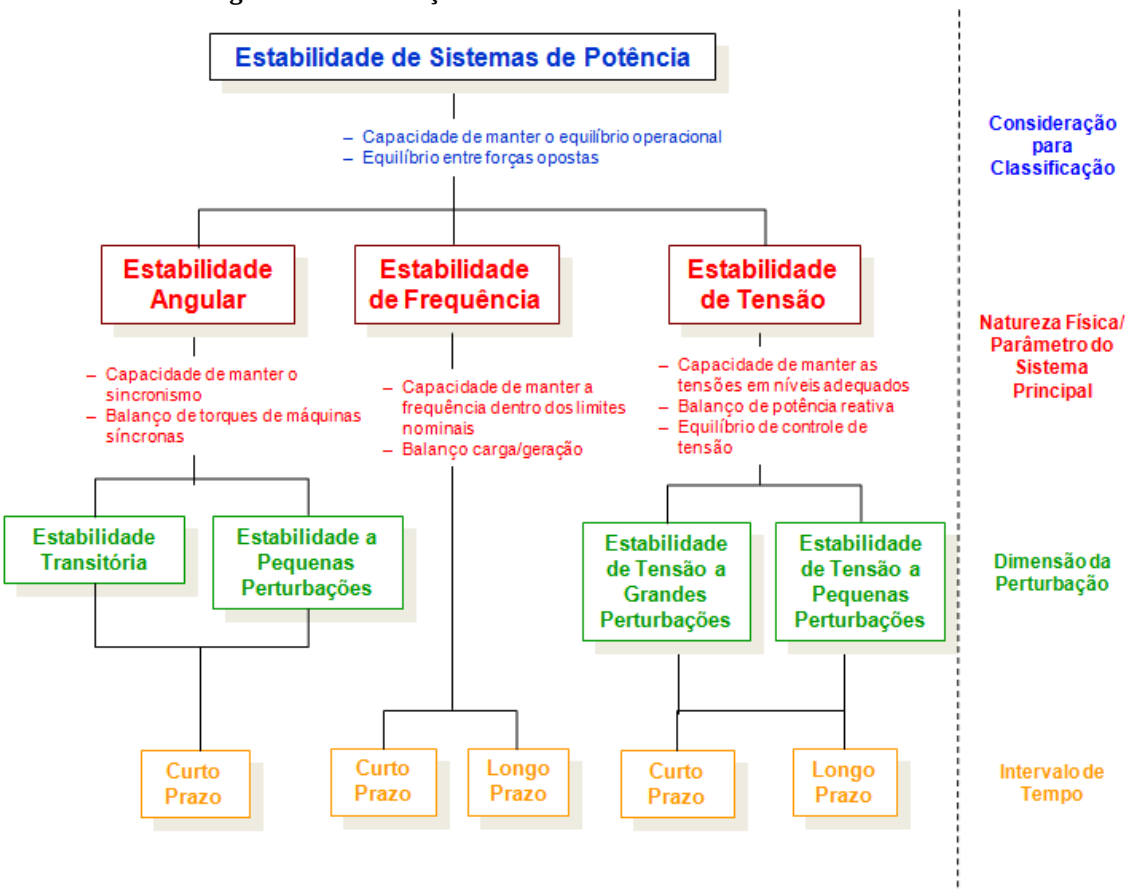
A estabilidade a pequenas perturbações é a capacidade do sistema de potência de manter o sincronismo quando submetido a uma pequena perturbação. Atualmente, esse fenômeno é amplamente tratado como um problema de insuficiência de amortecimento de oscilações. Os modos de oscilação eletromecânica de interesse são:

1. Modos Locais: Modos associados com a oscilação de unidades geradores de uma usina contra o resto do sistema de potência.

2. Modos Inter-áreas: Modos associados com a oscilação de um grupo de geradores em uma área do sistema contra outro grupo de geradores em outra área.

A estabilidade transitória é a habilidade do SEP de manter o sincronismo quando submetido a uma grande perturbação, como um curto-circuito.

Figura 2- Classificação de Estabilidade de Sistemas de Potência



(Kundur et al., 2004)

A estabilidade de tensão é a habilidade de manter as tensões em níveis adequados em todas as barras do SEP após o mesmo ter sido submetido a uma perturbação em uma dada condição operacional inicial. A instabilidade de tensão ocorre quando uma perturbação, como um aumento de carga ou uma mudança da condição do sistema ocasiona uma queda ou elevação progressiva e descontrolada de tensão em algumas barras. A instabilidade de tensão está ligada a incapacidade do SEP em manter um equilíbrio adequado de balanço de potência reativa e ações de controle de tensão. A instabilidade de tensão pode ou não estar associada a um problema de estabilidade

angular. Quando os problemas estão associados, a queda de tensão ocorre de maneira rápida, na escala de segundos (Kundur et al, 2004).

A estabilidade de frequência é a habilidade que o SEP possui de manter a frequência em níveis adequados, ou seja, dentro dos limites nominais após uma perturbação que resulta em um significativo desequilíbrio entre a geração e a carga. Esse fenômeno pode ocorrer na forma de oscilações sustentadas de frequência, ou com queda monotônica, levando ao desligamento de geradores e cargas, sendo determinado pela resposta global do sistema, tal como evidenciado pela sua frequência média, ao invés dos movimentos relativos dos rotores dos geradores (Kundur et al., 2004).

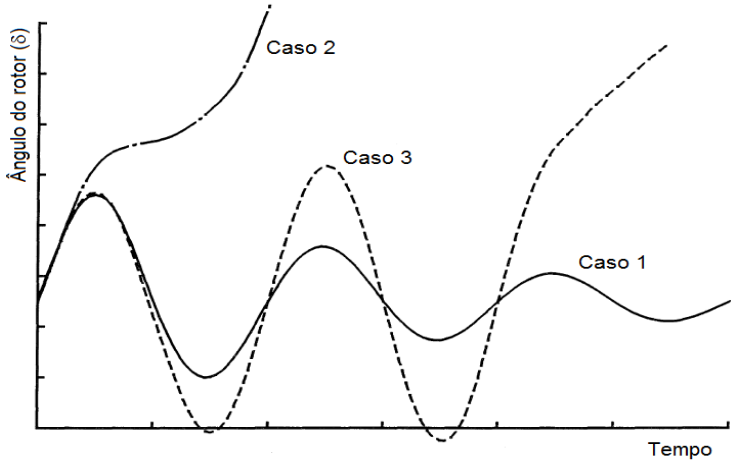
2.3.1 Estabilidade transitória.

Também denominada estabilidade de ângulo a grandes perturbações, a estabilidade transitória depende da capacidade que o SEP possui em manter e/ou restaurar o equilíbrio entre o torque eletromagnético e o torque mecânico de cada máquina síncrona interligada ao sistema.

A avaliação da estabilidade transitória, devido a sua importância, representa um grande esforço nos estudos sobre SEP, a qual é tradicionalmente realizada através de simulações no domínio do tempo. A simulação dinâmica no domínio do tempo, também conhecida como “simulação completa”, oferece a maior precisão com relação ao comportamento dinâmico do sistema.

A resposta do sistema a um curto-circuito envolve grandes excursões dos ângulos dos rotores dos geradores. Se a separação entre as máquinas interconectadas ao sistema cresce aperiodicamente, o sistema perde o sincronismo. A perda do sincronismo devido a instabilidade transitória geralmente se evidencia na primeira oscilação, nos primeiros segundos após a perturbação. A figura 3 apresenta a variação do ângulo do rotor de um gerador síncrono no decorrer do tempo após uma perturbação, considerando três casos.

Figura 3 - Resposta no tempo do ângulo do rotor frente a um distúrbio transitório.



(Kundur, 1994)

No Caso 1 da figura 3 o sistema é dito estável (seguro), pois mesmo que o ângulo do rotor oscile após a perturbação, esse ângulo se estabiliza em um ponto de operação próximo do ponto de operação pré-falta, em regime permanente.

No Caso 2 o ângulo do rotor cresce aperiodicamente logo na primeira oscilação, caracterizando assim um caso de instabilidade transitória.

O Caso 3 apresenta uma situação de instabilidade angular devido à falta de torque de amortecimento. Neste caso, o ângulo do rotor não cresce aperiodicamente na primeira oscilação. Diferentemente do caso 1, essa oscilação começa a crescer indefinidamente levando o sistema a instabilidade. Essa situação pode ocorrer devido a grandes ou pequenas perturbações.

Para a "simulação completa" o processo de cálculo consiste na solução de um conjunto de equações diferenciais e algébricas não lineares que descrevem o comportamento dinâmico dos elementos do sistema (equação 3).

$$\begin{cases} \dot{x} = f(x, r, u) \\ 0 = g(x, r, u) \end{cases} \quad (3)$$

Onde:

x é o vetor de variáveis de estado (velocidade da máquina, ângulo do rotor, etc.);

r é o vetor de variáveis algébricas (módulo e ângulo das tensões terminais, etc.);

u é o vetor de variáveis de entrada (controles);

f é o vetor de funções não lineares das equações diferenciais dos modelos dos dispositivos dinâmicos do sistema (geradores, motores de indução, os diversos controles, etc.);

g é o vetor de funções não lineares que descreve as equações algébricas da rede e de alguns modelos de dispositivos presentes no sistema.

A solução do sistema de equações descrito em (3) é realizada através de técnicas de integração numérica no domínio do tempo. Normalmente, utiliza-se entre 10 e 20 segundos de simulação após o distúrbio, sendo que esse tempo é suficiente para indicar problemas de estabilidade transitória. Por esse motivo, o termo "estabilidade de curto-prazo" também é utilizado. Uma vez detectada a possibilidade de ocorrência de instabilidade, medidas devem ser adotadas para garantir um ponto de operação seguro. Tais medidas podem incluir o redespacho de potência ativa ou até mudanças dos esquemas de proteção do sistema.

Em aplicações onde o tempo de simulação é crítico, como por exemplo, a operação em tempo real, a solução numérica do sistema (Equação 3) pode ser inviável, devido ao tempo consumido pelo processo iterativo. Este fato tem fortalecido linhas de pesquisa que investigam os chamados métodos diretos baseados em funções de energia, sendo capazes de avaliar a estabilidade transitória sem a necessidade da integração numérica. Além disso, o uso de funções de energia tem como grande atrativo o fato de possibilitar a obtenção do grau de estabilidade do sistema com base na margem de energia.

Embora os métodos diretos apresentem grande vantagem na redução no tempo computacional, sua utilização requer limitações na modelagem dos elementos dinâmicos do sistema. Os chamados métodos híbridos, que combinam as vantagens da simulação no domínio do tempo e dos métodos diretos, também vêm sendo bastante explorados. Uma alternativa para a melhora do desempenho computacional, mantendo-se a modelagem detalhada dos elementos dinâmicos, é o uso de métodos de integração com passo e ordem variáveis. Outra maneira de acelerar um estudo de estabilidade transitória em tempo real é a execução concorrente de vários casos em sistemas multiprocessados (*clusters*).

2.4 Desafios para a Operação Segura.

O aumento da demanda por energia elétrica e a necessidade das concessionárias manterem os padrões de qualidade de serviço, aliado há desregulamentação do setor elétrico certamente são fatores desfavoráveis à operação segura do sistema e acabam acarretando a enormes desafios para a operação dos sistemas elétricos. Esse aumento progressivo do consumo traz o aumento no carregamento cada vez maior das linhas de transmissão existentes e obriga as unidades geradoras operarem cada vez mais próximas

dos seus limites máximos, tornando o sistema susceptível há uma condição de colapso frente a distúrbios. Algumas soluções para esse problema e seus respectivos impactos são mostrados na tabela 2 (Morinson et al, 2004):

Tabela 2 - Características e potenciais impactos do estado atual da operação de SEP.

Características	Potenciais Impactos
Envelhecimento das Linhas de Transmissão.	• Aumento da probabilidade de falhas ou mau funcionamento de equipamentos.
Dificuldade para instalações de novas Linhas de Transmissão seja por seu alto custo ou questões ambientais.	• Sobrecarga nas linhas das atuais instalações de transmissão que levam a atuação da proteção, contribuindo para fenômenos como o colapso de tensão. • Gargalos nos principais corredores de transmissão.
Cortes na manutenção do sistema.	• Falhas em componentes e distúrbios como flashovers.
Maior dependência de controles e proteções especiais.	• Aumento da probabilidade de operação incorreta de proteções. • Aumento da imprevisibilidade de eventos em cascata.
Grande número de pequenos geradores distribuídos.	• Maior dificuldade na concepção do sistema adequado devido à incerteza de planejamento. • Incerteza no despacho
Intercambio de Potência	• Fluxos de potência menos previsíveis. • sobrecargas nas principais linhas. • Problemas de instabilidade de tensão ou instabilidade a pequenas perturbações.
Aumento da dependência de sistemas computadorizados e de comunicação.	• Falhas provindas de Software / Hardware podem deixar grande parte do sistema não observável aos operadores.
Planejamento do sistema integrado limitado.	• Utilização de recursos inadequados ou mesmo a falta de recursos para geração de energia.
Tendência de interconexão	• Exposição a distúrbios em cascata provocados por eventos em sistemas vizinhos.
Novas tecnologias, como sistemas avançados de controle, geração eólica, geração fotovoltaica, etc.	• Falta de experiência do operador com tecnologias que podem ter características dinâmicas. • Comportamentos imprevisíveis.

2.5 Avaliação da segurança em tempo real (*on-line*).

Pode-se definir a avaliação da segurança como a análise realizada para determinar o quanto um sistema elétrico de potência está imune a perturbações que podem

comprometer o seu perfeito funcionamento. Para possibilitar a avaliação da segurança, a operação dos sistemas elétricos em tempo real é realizada por meio da execução periódica de funções de análise e controle de rede (Balu, 1992). Isto é feito a partir de dados disponíveis, que representam o estado de operação, sendo possível determinar o nível de segurança atual e também prever, caso venham a ocorrer contingências, o estado de segurança das transições futuras.

A análise de segurança é classificada em tempo real ou *online* quando a determinação do estado e nível de segurança faz parte das tarefas rotineiras de monitoramento. Na análise *online*, as reais condições operativas do sistema podem ser levadas em conta, e os limites operativos e ações de controle podem ser atualizados periodicamente (CASTRO, 1995).

Nos estudos ditos *offline*, opta-se por avaliar somente os piores casos, tendo em vista o tamanho da rede elétrica e o grande número de contingências a analisar. Contudo, esta atitude conservadora leva ao estabelecimento de limites operativos imprecisos e conservativos, válidos somente para situações genéricas de carga leve, média ou pesada, por exemplo. Porém, muitas vezes o sistema opera em uma condição diferente da antes avaliada. Portanto, as instruções operativas disponíveis para dados momentos podem não ser as mais adequadas para a execução pelo operador. Com a análise *online* a operação do sistema se dá de modo muito realista e confiável, pois se baseia nas informações mais recentes possíveis.

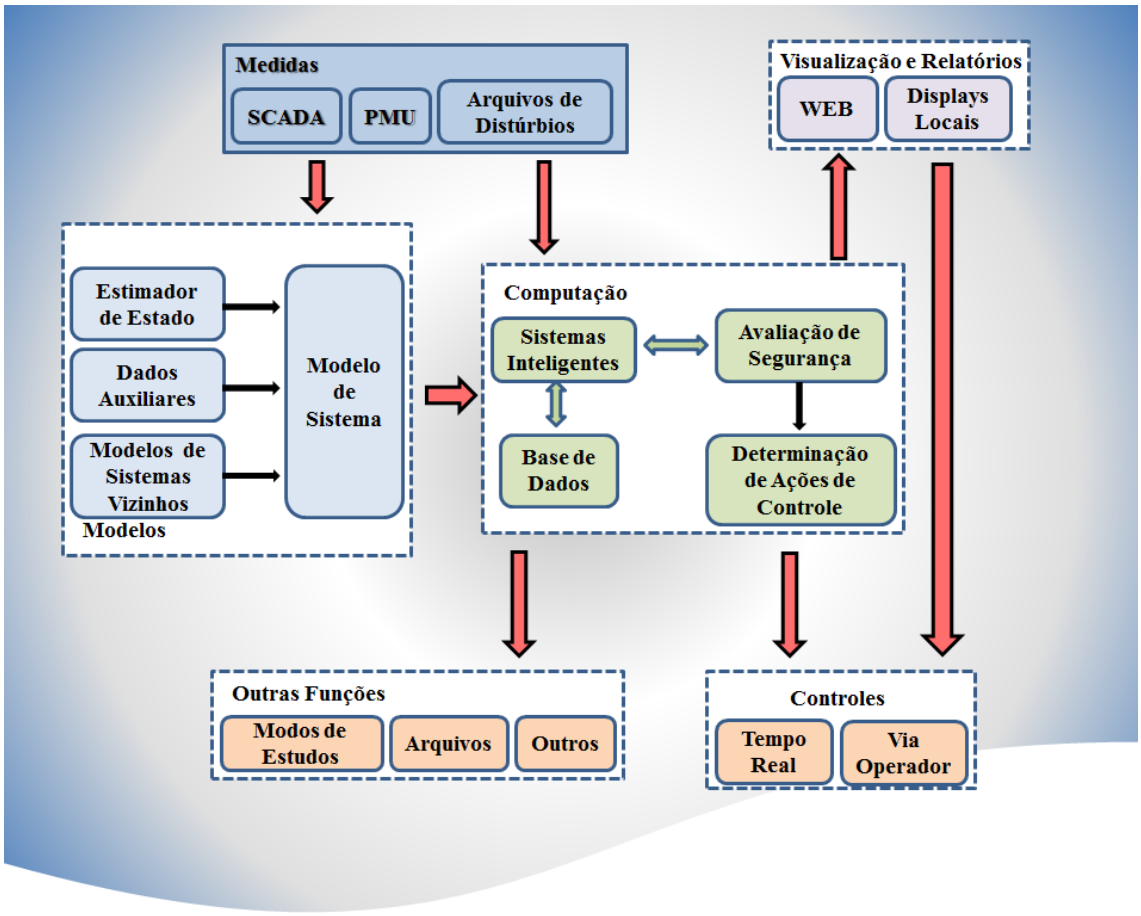
As ações básicas da análise de segurança *online* são: monitoramento, avaliação e melhoria da segurança, como ilustrado na figura 4:

a) Monitoramento da segurança: Usando medições do sistema em tempo real, é identificado se o sistema está em estado normal ou não. Se ele estiver em estado de emergência precisará de um controle de emergência. Se ocorrer a perda de carga será necessário um controle restaurativo.

b) Avaliação da Segurança: Se o sistema está em estado normal, é determinado se o sistema está seguro ou inseguro com respeito a um conjunto de contingências, bem como são estabelecidos índices ou margens de segurança;

c) Melhoria da Segurança: Se inseguro, são determinadas ações corretivas que devem ser tomadas para que o sistema se torne seguro. Se seguro, com pouca margem, são determinadas as ações preventivas para evitar que o sistema se torne inseguro.

Figura 4- – Estrutura para Avaliação de SEP em Tempo Real.



(Morinson et al., 2004)

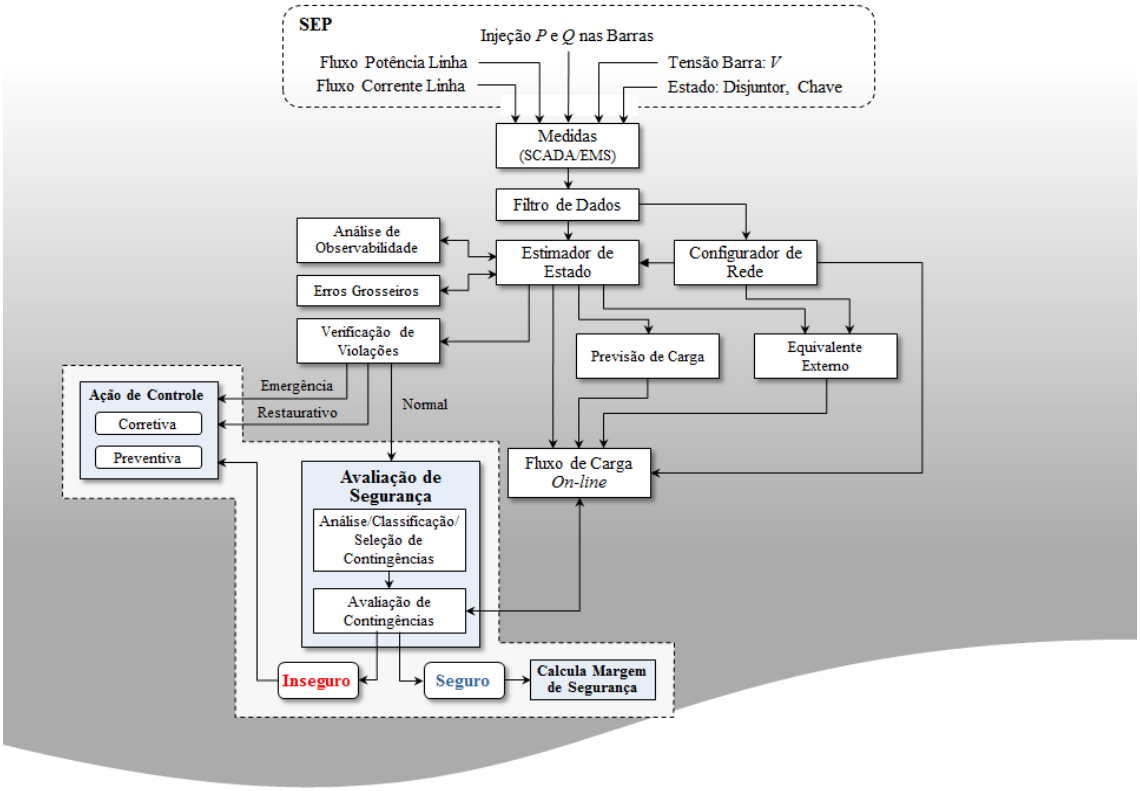
A realização da função de segurança em tempo real está sujeita a um sério problema de restrição de custo computacional, isto é, de tempo de execução da análise, principalmente quando se trata da análise dinâmica em sistemas de grande porte, onde são necessárias as análises de uma extensa lista de contingências críticas. Por esse motivo, avaliação de segurança estática *online* já é aplicada de maneira mais estruturada. Aqui, o termo "estruturado" significa que as tarefas são realizadas em tempo hábil. No entanto, segurança dinâmica ainda consiste em um enorme desafio se avaliada de modo *online* devido a necessidade de maior poder de processamento, para que as tarefas sejam executadas em tempo hábil.

Contudo, devido à natureza desacoplada das análises realizadas pela função de avaliação de segurança, uma opção complementar é a utilização de técnicas de inteligência computacional ou mineração de dados, cujas execuções sejam suficientemente rápidas, sem, no entanto, perder a precisão requerida para o resultado da avaliação.

2.5.1 Avaliação da Segurança Estática O*nline.*

Nos modernos SEPs, ao tradicional ambiente SCADA, de supervisão e controle, são adicionadas ferramentas computacionais de apoio aos operadores nas atividades diárias de operação da rede elétrica, visando aumentar a qualidade do serviço e reduzir as interrupções no fornecimento de energia e a um menor custo possível, ou seja, auxiliando na manutenção do SEP dentro das suas margens de segurança. O ambiente então passa a ser denominado de Sistema de Gerenciamento de Energia (Energy Management System – EMS). Para evitar situações extremas os EMSs executam processamentos adicionais que buscam garantir que as análises de segurança sejam realizadas de modo preventivo, antes da eventual ocorrência de contigências.

Figura 5 - Principais funções que um EMS adiciona ao SCADA.



(Adaptado, Rodrigues 2014)

A partir da estimação do estado da rede elétrica pelo Estimador de Estado, é executada a função de controle da segurança. Essa função é baseada na solução de um Fluxo de Carga Ótimo com Restrições de Segurança e fornece como resultado um conjunto de ações de controle que devem ser tomadas pelo operador para que o sistema seja mantido em estado *seguro*. A Figura 5 apresenta uma visão geral destas funções executadas pelos centros de controle.

É possível determinar o estado de operação do sistema elétrico de potência, bem como do seu nível de segurança, através dos programas computacionais que tradicionalmente constituem as funções de supervisão e controle. Utilizando para isso o processamento dos valores medidos em campo, provenientes dos elementos da rede elétrica, que chegam ao centro de supervisão e controle e do modelo matemático que o descreve. Segundo Monticceli (1983), as principais funções de supervisão e controle, que fazem de um SCADA um EMS, são:

Configurador de Redes – A partir dos dados das medidas de estado dos disjuntores, das chaves seccionadoras e da disponibilidade dos equipamentos da rede, a função determina a topologia atual da rede elétrica. Para cada alteração que ocorre na topologia da rede, devido a ação do operador ou a contingências, o configurador atualiza os dados resultando em um novo modelo eletro-topológico da rede.

Análise de Observabilidade – Esta função consiste em avaliar se o número e a localização das medidas enviadas ao centro de operação e controle fornecem informações suficientes que permite a estimação dos estados do sistema. Esta é uma função que, em geral, se encontra embutida na função Estimador de Estado.

Estimador de Estado –O Estimador de Estado, a partir da topologia informada pelo Configurador de Estado e das equações diferenciais que descrevem o modelo elétrico utilizando as propriedades elétricas dos diversos componentes que compõem o SEP, tais como reatâncias, susceptâncias de linhas, transformadores, etc e, utilizando algoritmos de minimização de erros, estabelece uma nova matriz de estado entre aquele estado apontado pelo modelo matemático e aquele informado pelo SCADA. Esta nova matriz é denominada estado estimado.

Equivalente externo – Com os estados obtidos e as informações sobre a parte não observável do sistema é construído um modelo que representa o comportamento do sistema externo. Este modelo é utilizado pelo estimador de estado para seu processamento das equações diferenciais que descrevem o SEP.

Fluxo de carga *on-line* – De posse dos dados fornecidos pelas funções descritas anteriormente, o modelo final é obtido pela execução do fluxo de potência *on-line*. Esta função possibilita conhecer os efeitos resultantes das ações de controle antes de realizá-las no sistema elétrico, avaliando quais os efeitos das contingências, e identificando medidas de controle preventivo para evitar que o sistema transite para estado inseguro.

Previsão de carga - Esta função realiza a previsão de carga por barra do sistema, a partir dos dados de demanda previstas para o sistema e do seu comportamento histórico.

Avaliação de segurança ou também chamada de Análise de Contigências – Esta função realiza simulações de um conjunto predefinido de contingências e verifica as violações dos limites operacionais. Ela apresenta, a partir do estado atual fornecido pelo Estimador de Estado os estados futuros se as contigências simuladas vierem a ocorrer. A partir destas informações os operadores podem realizar ações de controle preventivo para garantir a condição normal de operação.

Conforme o esquema representado na área destacada da figura 5, se o sistema for classificado como *seguro*, podem ser calculadas as margens de segurança das tensões nodais; por outro lado, se o sistema for classificado como *inseguro*, podem ser definidas e calculadas as ações de controle corretivo ou preventivo necessárias para corrigir as variáveis que estão sendo violadas.

Essas funções de análise de rede são executadas de maneira periódica, a intervalos de tempo de alguns segundos, o que implica em grande restrição de tempo para a execução dos complexos programas computacionais necessários para realização dessas funções, além de resultar em uma grande massa de dados para análise.

2.5.2 Avaliação da Segurança Dinâmica em Tempo Real.

A busca por uma operação segura e econômica de um SEP é sempre prioritária. No decorrer dos anos os centros de controle das empresas de energia elétrica vêm evoluindo e incorporando os recursos computacionais que surgiram nesse ínterim, o que têm aberto o caminho para a implantação de ferramentas *on-line*, com as funções de monitoramento, análise e controle. Atualmente, vários centros de controle realizam análise de segurança estática (análise de contingências) em modo *online*. Já a implementação da avaliação segurança dinâmica *online* está em um estágio bem menos avançado devido a três motivos principais:

1. A complexidade do modelo dinâmico;
2. A capacidade computacional insuficiente dos centros de controle para realizar a análise em tempo hábil.
3. A falta de métodos alternativos para análise de segurança que sejam eficientes e confiáveis;

Na análise *online*, o objetivo é determinar se um ponto de operação atual ou futuro é seguro ou não, partindo-se de medidas em tempo real do sistema, detecção de configuração e estimação de estado que fornece o caso base presente. Em estudos *offline* os casos base são previamente selecionados para representar os cenários de pior caso.

Análise de segurança só faz sentido se o sistema estiver no estado normal do ponto de vista estático, ou seja, toda a carga atendida e nenhuma restrição operativa violada.

A análise *on-line* de segurança dinâmica pode ser decomposta em vários subproblemas que serão descritos a seguir e interagem entre si na forma mostrada na figura 6.

Previsão de Carga: A análise da segurança dinâmica *on-line* é uma análise prospectiva, isto é, visa determinar se um ponto de operação futuro será estável (sendo, portanto, um estado operativo seguro), ou instável (levando o estado do sistema a ser classificado como inseguro). Sua finalidade é dar aos operadores informações atualizadas sobre limites operativos de transferência e ações preventivas de controle que os auxilie na tomada de decisões de forma antecipada. Deste modo, a determinação da previsão de carga torna-se parte essencial na análise.

Preparação de Cenários: Este módulo recebe informações de vários tipos e prepara convenientemente as bases de dados que serão utilizadas durante o cálculo da estabilidade transitória. Fornece o caso base futuro. Como funciona como uma espécie de gerente da análise de segurança pode ser adotada uma certa dose de inteligência artificial.

Cálculo da Estabilidade Transitória: Este subproblema corresponde ao que normalmente faz um programa de cálculo da estabilidade transitória. Embora qualquer método possa ser usado (direto, híbrido ou simulação), porém é mais confiável utilizar a simulação no domínio do tempo por integração numérica. Desta maneira, evitam-se restrições na modelagem de geradores, seus controles e outros componentes. Por outro lado, a simulação passo a passo pode ser muito demorada, levando este módulo a ser excessivamente oneroso em termos computacionais, ou seja, esse é um ponto que muitas vezes inviabiliza a DSA *online*.

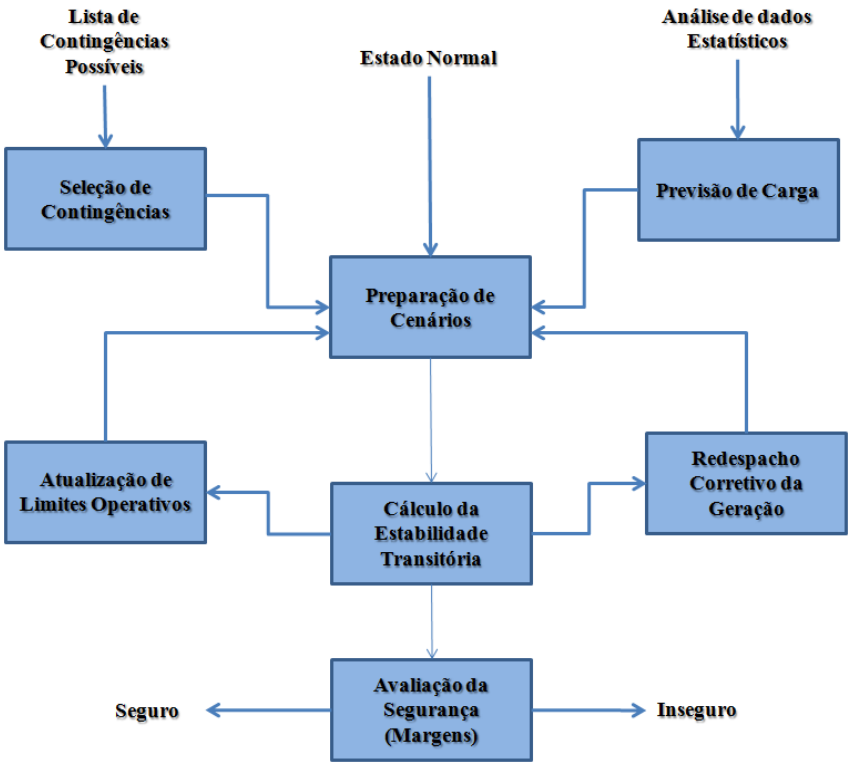
Avaliação da Segurança: os resultados das simulações são analisados e o sistema é classificado como seguro ou inseguro. O sistema (ponto de operação) é considerado inseguro se pelo menos uma contingência levar um ou mais geradores à instabilidade. Dentro deste módulo também se inclui a obtenção de margens ou índices de estabilidade ou instabilidade, dependendo da situação. A margem de estabilidade quantifica a distância que um sistema estável está da instabilidade, enquanto que a margem de instabilidade quantifica a distância que um sistema instável está da estabilidade.

Atualização de Limites Operativos: Em geral, as redes elétricas têm limites operativos dinâmicos expressos através do máximo carregamento de um conjunto de linhas de transmissão ou máxima geração de um grupo de usinas (fluxo de interface). Se

o sistema estiver dinamicamente seguro pode-se estudar o relaxamento dessas restrições e verificar até onde os limites podem ser relaxados. Frequentemente, isto é realizado através de sucessivos redespachos de geração incrementais em pontos localizados até que um estado inseguro seja alcançado, em modo simulação, naturalmente.

Redespacho Corretivo de Geração: Se for detectado que o sistema está dinamicamente inseguro pode-se estudar medidas preventivas ou corretivas para levá-lo ao estado seguro. Uma ação corretiva (que atua após a ação corretiva de uma contingência) que tem se mostrado efetiva e consiste no corte de potência dos geradores que tendem à instabilidade (perda de sincronismo). Portanto, deve-se obter a mínima quantidade de geração capaz de levar o sistema a um estado seguro. Isto corresponde a realizar redespachos incrementais de geração até se obter um esquema de corte de geração, que deverá ser implantado nos mecanismos automáticos de controle de geração.

Figura 6 - Subproblemas da análise on-line de segurança dinâmica.



(Castro, 1995).

Entretanto, todos esses passos necessitam ser automatizados para que possam ser utilizados de maneira confortável em ambiente de operação, o que exige, ainda hoje, pesquisas mais aprofundadas. Como destacado na revisão bibliográfica deste trabalho, muitos estudos são realizados para que se utilizem sistemas baseados em inteligência computacional, métodos diretos ou computação paralela de DSA *online*. Busca-se nesse

trabalho, contribuir com a SDSA *online* por meio de árvores de decisão utilizando dados reais do histórico de operação de SEP.

2.6 Conclusão.

Neste capítulo foram apresentados os conceitos de segurança de sistemas de potência, bem como a forma básica de como sua avaliação é realizada nos centros de operação. É notável a contribuição dos sistemas computacionais no auxílio aos operadores. Enfatiza-se, finalmente, a necessidade de ferramentas ágeis e confiáveis para melhorar as condições de operação de sistema, aumentar sua segurança operativa, reduzir riscos bem como facilitar as condições de trabalho dos operadores.

Capítulo 3 - Mineração de Dados e Árvores de Decisão

3.1 Considerações Iniciais: Descoberta de Conhecimento em Base de Dados

Nas últimas décadas, com os avanços dos recursos computacionais, uma grande quantidade de dados foi armazenada, sendo que esse tamanho cresce devido à coleta diária de dados dos mais diversos seguimentos. Contudo, essa quantidade de dados produzida e armazenada em larga escala, acabou se tornando inviável de se analisar por meio de métodos manuais tradicionais. Essas bases de dados por sua vez, podem apresentar grande valia para o planejamento, gestão e tomadas de decisão. A necessidade dessa análise, exigida por mercados mais competitivos, motivou o desenvolvimento de ferramentas com o intuito de descobrir conhecimento desses dados. Dessa forma é possível observar que a mineração de dados é o resultado natural da evolução da tecnologia da informação (Han, et al, 2012).

A Mineração de Dados (MD), ou DataMining (DM), é um processo de extração de informações, que consiste em analisar dados e aplicar algoritmos que produzem uma enumeração particular de padrões ou modelos sobre um banco de dados. Seu grande potencial para descoberta de informações que se encontram implícitas em bases de dados, faz da mineração de dados uma técnica utilizada em diversas áreas. Assim como a evolução da tecnologia da informação entrou em simbiose com diversos campos da ciência, a mineração de dados começa a se tornar parte integrante de diversos campos de aplicação, tais quais mineração na web, decisões que envolvam julgamento (avaliação de créditos), triagem de imagens, previsão de carga, diagnósticos médicos, diagnósticos de falhas, mercado financeiro e vendas (Witten, 2011).

Entretanto, o processo de mineração de dados é apenas uma parte de um processo maior, conhecido como “Knowledge Discovery in Databases” (KDD) em português, Descoberta de Conhecimento em Bases de Dados, que permite a extração não trivial de conhecimento previamente desconhecido e potencialmente útil de um banco de dados. O processo de KDD foi proposto em 1989 para referir-se às etapas que produzem conhecimentos a partir dos dados. Este processo envolve encontrar e interpretar padrões nos dados, de modo iterativo e interativo, através da repetição dos algoritmos e da análise de seus resultados. Esse processo contém diversas fases: definição do problema; seleção dos dados; limpeza dos dados; pré-processamento dos dados; codificação dos dados; enriquecimento dos dados; mineração dos dados e a interpretação dos resultados (Han, 2012; Kumar, 2006).

Neste capítulo, serão apresentados os aspectos mais importantes da mineração de dados, onde serão mostradas com maiores detalhes as técnicas de classificação por árvores de decisão.

3.2 Mineração de Dados

A mineração de dados tem origem na interseção de diferentes áreas. Sendo que as áreas mais relacionadas são: Machine Learning (aprendizado de máquina), Inteligência Computacional, Estatística e Visualização dos dados (Kumar et al, 2006). Na área de Inteligência Computacional, em particular, as técnicas mais utilizadas são: Redes Neurais Artificiais, Sistemas Fuzzy e Algoritmos Genéticos.

A Mineração de Dados, ou Data Mining, é um processo de extração de informações, que consiste em analisar dados e aplicar algoritmos que produzem uma enumeração particular de padrões ou modelos sobre um banco de dados. Seu grande potencial para descoberta de informações que se encontram implícitas em bases de dados, faz da mineração de dados uma técnica utilizada em diversas áreas. Portanto, empresas, indústrias, instituições de pesquisas, órgãos governamentais, entre outros, utilizam esse processo automático de análise de grandes bases de dados, extraindo assim o conhecimento que pode ser utilizado para auxílio a algo de errado que ocorreu no passado, assim como predizer comportamentos futuros.

3.2.1 Etapas do Processo de Mineração de Dados

Para que a mineração de dados possa ser bem-sucedida, é necessário seguir um conjunto de passos, ou processos. Uma metodologia para tal, pode ser a CRISP-DM (CRoss Industry Standard Process for Data Mining) (The CRISP-DM Consortium, 2000), que nos primórdios do mercado de mineração dados, definiu alguns padrões com processos genéricos, para realizar MD.

O atual processo para mineração de dados propõe uma visão cíclica no projeto. Ele contém as fases correspondentes de um projeto, suas respectivas tarefas e relacionamentos entre essas tarefas. Onde as etapas são ligadas umas às outras, sendo elas:

Entendimento do Negócio - Essa fase inicial tem o foco no entendimento do escopo de negócio, isto é, conhecer as informações iniciais e fundamentais sobre a área, visando obter conhecimento sobre os objetivos do negócio e seus requisitos, e então converter esse conhecimento em uma definição de um problema de mineração de dados, e um plano preliminar designado para alcançar esses objetivos.

Seleção dos Dados - Consiste no entendimento dos dados, que visa à familiarização com o banco de dados pelo grupo de projeto, utilizando-se de conjuntos de dados "modelo". Uma vez definido o domínio sobre o qual se pretende executar o processo de descoberta, o próximo passo é selecionar e coletar o conjunto de dados ou variáveis necessárias. Nesta etapa, costuma-se identificar possíveis problemas de qualidade, ou detectar subconjuntos interessantes para formar hipóteses.

Limpeza dos Dados - A fase de preparação de dados consiste na preparação dos dados que visa a limpeza, transformação, integração e formatação dos dados da etapa anterior. É a atividade pela qual os ruídos, dados estranhos ou inconsistentes são tratados. Esta fase abrange todas as atividades para construir o conjunto de dados final, coma utilização de Data Warehouses (repositórios organizados), facilitando esta etapa do processo de mineração de dados, que costuma ser a fase que exige mais esforço, correspondendo geralmente a mais de 50% do trabalho.

Modelagem dos Dados – Etapa que visa à aplicação de técnicas de modelagem sobre o conjunto de dados preparado na etapa anterior, ou seja, são aplicadas as técnicas baseadas em conceitos de aprendizagem de máquina, reconhecimento de padrões, estatística, classificação e clusterização, onde os parâmetros para cada modelo são calibrados para se obter valores otimizados. Geralmente, existem várias técnicas para o mesmo tipo de problema de mineração, com requerimentos específicos na forma dos dados, o que muitas vezes faz com que se volte para a etapa de preparação de dados.

Avaliação do processo - A avaliação do processo visa garantir que o modelo gerado atenda às expectativas da organização. Os resultados do processo de descoberta do conhecimento podem ser mostrados de diversas formas. Porém, estas formas devem possibilitar uma análise criteriosa para identificar a necessidade de retornar a qualquer um dos estágios anteriores do processo de mineração, avaliando mais detalhadamente o modelo, e rever as etapas executadas para construir o modelo, para se certificar de que ele conseguirá alcançar os objetivos de negócio. Deve se determinar se houve algum importante objetivo do negócio que não foi suficientemente alcançado. No fim desta fase, uma decisão sobre o uso dos resultados da mineração deve ser tomada.

Execução - Esta fase consiste na definição das fases de implantação do projeto de Mineração de Dados. A criação do modelo não é o fim do projeto. Mesmo se a finalidade do modelo for apenas aumentar o conhecimento dos dados, o conhecimento ganho necessitará ser organizado e apresentado em uma maneira que o cliente possa usar. Dependendo das exigências, a fase de execução pode ser tão simples quanto à geração de

um relatório, ou tão complexa quanto executar processos de mineração de dados repetidamente. Em muitos casos será o cliente, não o analista dos dados, que realizará as etapas da execução. Entretanto, mesmo se o analista não se encarregar da execução é importante que ele faça o cliente compreender que medidas deverão ser tomadas a fim de empregar efetivamente os modelos criados.

3.3 Tarefas de Mineração

Tarefas são classes de problemas solucionáveis usando técnicas de mineração de dados. Baseada na experiência prática de exploração de dados, estas classes de problemas foram definidas. Para cada tipo de tarefa existe uma técnica ou um conjunto de técnicas para a sua resolução. A mineração pode ser preditiva e descritiva. A preditiva trata do uso de variáveis no banco de dados para predizer valores de outras variáveis. A descritiva identifica padrões em dados históricos.

Uma tarefa de mineração de dados pode ser classificada de acordo com o tipo de padrão descoberto. Dentre os diversos tipos de padrões que podem ser descobertos em uma tarefa de mineração de dados, destacam-se as seguintes (Kumar,2006):

Descrição - É a tarefa utilizada para descrever os padrões e tendências revelados pelos dados. A descrição geralmente oferece uma possível interpretação para os resultados obtidos. A tarefa de descrição é muito utilizada em conjunto com as técnicas de análise exploratória de dados, para comprovar a influência de certas variáveis no resultado obtido.

Predição - A tarefa de predição visa descobrir o valor futuro de um determinado atributo, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe ou qual valor pode ser associado, a fim de "aprender" como prever um novo registro. Exemplos:

• Predizer o valor de uma ação três meses adiante;

• Predizer o percentual que será aumentado de tráfego na rede se a velocidade aumentar;

• Predizer o vencedor do campeonato baseando-se na comparação das estatísticas dos times;

Alguns métodos de classificação e regressão podem ser usados para predição, com as devidas considerações. Neste trabalho, o foco será dado em predição, mais especificamente na classificação.

Essa tarefa tem como objetivo derivar uma regra que possa ser usada para classificar, de forma otimizada, uma nova observação a uma classe já rotulada. Essa tarefa é preditiva, pois a partir da construção de um modelo, pode automaticamente a classe de um novo dado. Para isso, devem ser considerados doistipos de atributos: atributos preditivos, cujos valores irão influenciar no processo de determinação da classe e atributos objetivos, que indicam a classe a qual o objeto pertence. As principais técnicas para a tarefa de classificação são: Árvores de decisão, Regras de classificação, Redes Neurais, Máquinas de Vetor e Suporte.

3.4 Árvores de Decisão

As árvores de decisão (DT, Decision Tree) são algoritmos de classificação de dados, muito utilizados no ambiente da chamada Mineração de Dados.Com uma característica importante, baseada em particionar de modo recursivo um conjunto complexo de dados, até que cada subconjunto assim obtido contenha apenas os casos de uma única classe possibilitando uma melhor análise (Han, 2012). Após a construção da árvore de decisão, os resultados obtidos, são formados por dados organizados de maneira simples e de fácil entendimento e podem servir como importante ferramenta para tomada de decisão.

As árvores de decisão são construídas baseadas no modelo hierárquico *top-down*, isto é, do nó raiz em direção às folhas. Embora haja diferenças na forma de realizar os passos, os algoritmos utilizam a técnica dividir para conquistar. Esta técnica está baseada nas sucessivas divisões do problema estudado em vários subproblemas de menores dimensões, até que seja encontrada uma solução mais simples para cada um dos problemas.

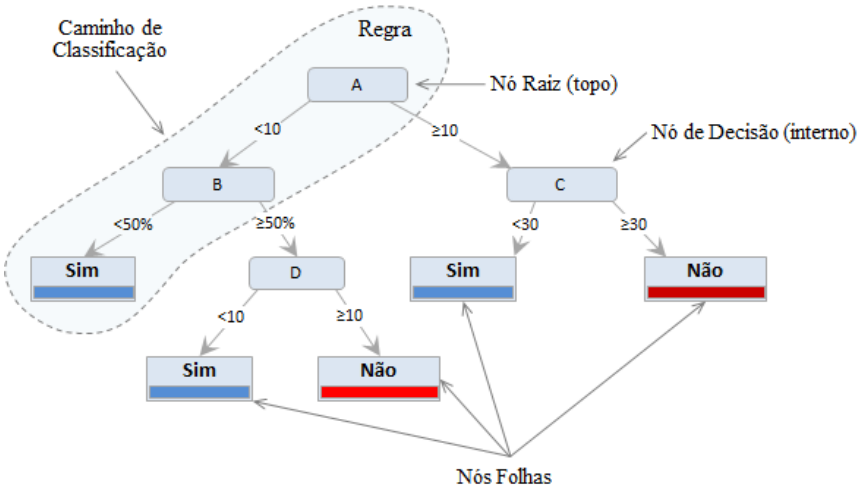
Os primeiros classificadores baseados em árvore de decisão surgiram no final da década de 50, a partir do trabalho de Hunt, foram apresentados vários experimentos para indução de regras. Posteriormente, foi desenvolvido o algoritmo CART (*Classification And Regression Trees*) por BREIMAN, (1984), e QUINLAN, (1986) desenvolveu o algoritmo ID3 (*Iterative Dichotomiser 3*) e como seus sucessores surgiram os algoritmos C4.5 (1993).

Uma árvore de decisão é essencialmente uma série de declarações *if-then*, que quando aplicados a um registro de uma base de dados, resultam na classificação daquele registro. A Figura 3.1 representa um exemplo de uma árvore hipotética, e é constituída dos seguintes elementos:

- Raiz: é o nó do topo da árvore.
- Nós: são todos os elementos que estão conectados por ramos;
- Ramos: Ligações entre nós;
- Folhas: são os últimos nós da árvore e representam as classes do conjunto de dados.

Na figura 7, tem-se um exemplo de DT, onde os nós da árvore são representados pelos nomes dos atributos A, B, C e D dispostos na árvore de acordo com o seu nível de informação. Saindo dos nós tem-se os ramos de ligação onde são testados os valores dos atributos correspondentes, até chegar a classificação final no nó folha que representam as diferentes classes a que pertence o dado. A classificação ocorre seguindo o caminho que vai do nó raiz até as folhas, conforme indicado na área em destaque da figura 7, onde o atributo denominado A tem valor < 10 e o atributo B valor <50%.

Figura 7 - Exemplo de uma árvore de decisão



Resumidamente pode-se afirmar que sobre uma árvore de decisão:

- Cada nó interno representa um teste em um atributo;
- O ramo que sai de um nó interno representa o resultado do teste no atributo;
- A folha da árvore representa um rótulo de classe (*Sim*/*Não*);
- Uma nova observação é classificada seguindo um caminho na árvore, da raiz até a folha;

Da analise da estrutura do exemplo da figura 7, nota-se que é possível extrair regras do tipo "*if-then*" para uma melhor compreensão e interpretação dos resultados. As

regras de decisão seguem de acordo com o trajeto dos nós sucessivos até um nó folha com a sua respectiva classe atribuída e geralmente essas regras são utilizadas juntamente com a árvore, em alguns casos as regras substituem as árvores de decisão. Assim, podem ser derivadas as seguintes regras para o exemplo em questão:

- *If* A < 10 e B < 50% *Then* Sim.
- *If* A < 10 e B ≥ 50% e D < 10 *Then* Sim.

Devem ser aplicados testes em todos os nós definidos e um conjunto de exemplos, para que seja escolhido o caminho a percorrer na árvore durante o processo de classificação. Os resultados destes testes dependem da natureza dos atributos, que podem ser numéricos, categóricos ou simbólicos. Deve-se utilizar apenas um nó para cada teste, pois, isso torna a árvore bem mais simples.

3.5 Construção de Árvores de Decisão

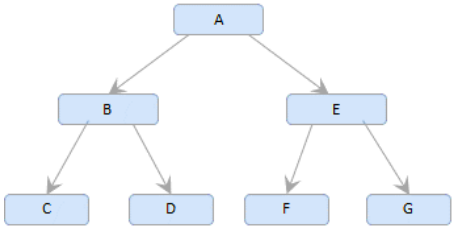
Na literatura, é dado o nome de indução de árvores de decisão ao processo no qual se constrói um modelo de DT, onde esse processo de indução de árvores de baseia-se em particionar recursivamente um conjunto de treinamento até que cada subconjunto obtido deste particionamento contenha casos de uma única classe.A partição é definida relacionando um teste que tenha um pequeno conjunto de saídas, criando uma ligação para cada uma possível, passando cada exemplo para sua correspondente ligação e tratando cada bloco de partição como um subproblema para o qual uma subárvore é construída recursivamente. Como critério de parada é comum adotar quando todos os exemplos de uma ligação pertencem a uma mesma classe.

A estrutura de uma DT é formada por nós que representam os atributos, por ramos (que são as ligações) provenientes dos nós e os nós folha, que representam as diferentes classes do conjunto de exemplos. Geralmente, o atributo mais efetivo dos dados é selecionado como o primeiro nó da árvore e seus valores limites são usados para criar os nós sucessores.

Para realizar o processo de indução de árvore de decisão é utilizada a estratégia *top-down*, onde inicia a geração da árvore a partir do nó raiz, percorrendo os ramos até o próximo nó interno. Este processo de indução pode ser realizado de duas maneiras distintas, conforme descritas a seguir:

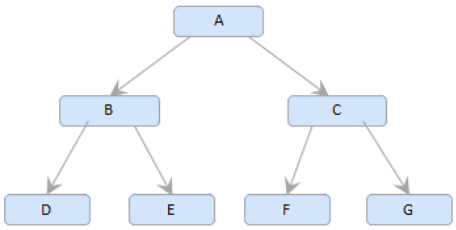
Por profundidade – Conforme mostrado na figura 8, a partir de um nó inicial, é feita ligação até um novo nó, caso este nó não seja uma folha, são acrescentados uma nova ligação e um novo nó, e assim sucessivamente. Quando chegar a uma folha, retornar até o nó que gerou a folha e acrescentar um novo ramo e um novo nó. Caso novamente não seja possível adicionar um novo nó, retornar até o nó que seja possível incluir uma nova ligação. O processo termina somente quando todos os caminhos terminam por uma folha e nenhum nó puder ser encontrado.

Figura 8 - Indução de árvore de decisão por profundidade



Por largura – Faz todas as ligações de um nó chegando a novos nós, e para cada nó resultante, o critério é testado e é adicionada uma folha ou novos nós, e assim, até que não haja nenhum nó a ser testado. Este método de indução não retorna ao nó raiz.

Figura 9 - Indução de árvore de decisão por largura



3.6 Critérios para seleção de atributos

No processo de indução de uma árvore de decisão, a escolha correta do atributo definirá o sucesso do algoritmo de indução. Vários critérios foram desenvolvidos para escolha do melhor atributo que poderá ser utilizado em um nó, isto é, qual a melhor divisão dos registros da base de dados (Kumar, 2006). Este critério verifica cada atributo candidato e seleciona aquele que maximiza (ou minimiza) alguma função heurística sobre os subconjuntos.

Entre os vários critérios para escolha de um atributo candidato a nó, os mais usuais são: o *Ganho de Informação, Critério de Gini e Erro de Classificação*.

O critério de ganho de informação está fundamentado em uma medida conhecida como *Entropia da Informação de Shannon*.

A entropia pode ser definida como a medida de informação calculada pelas probabilidades de ocorrência de eventos combinados ou individuais. Pode-se afirmar que a entropia é uma medida do grau de impureza em um conjunto arbitrário de amostras, ou seja, é a medida da desordem de um conjunto de amostras. O critério da entropia mede a quantidade de informação necessária para codificar a classe do nó. Dado um atributo classe *A*, de um conjunto de amostras *S*, em que *A* pode assumir v_i valores de classes diferentes, então a entropia de *A* relativa a esta classificação é definida como:

$$Entropia(A) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{4}$$

Onde *m* é o número total de classes e $p_i = p(A = v_i)$ é a probabilidade do atributo classe *A* ser igual à classe cujo índice é *i*, isto é, a proporção do número de amostras com valor v_i em relação ao número total de amostras de *S*.

Quanto maior a entropia de um atributo, mais uniforme é a distribuição dos seus valores. Entropia igual à zero significa que ocorreu apenas uma classe no conjunto de dados, e será igual a um se o número de amostras de cada classe for igual. Para uma entropia próxima de zero, indica que as classes são pouco uniformes.

Desenvolvido por Quinlan, (1986), o ganho de informação é dado pela soma das entropias individuais menos a entropia conjunta, sendo uma medida de correlação entre duas variáveis. É uma propriedade estatística que mede como um determinado atributo separa as amostras de treinamento de acordo com sua classificação.

A construção de uma árvore de decisão visa reduzir o valor da entropia. A medida do ganho de informação representa a redução esperada na entropia de um atributo preditivo, considerando que o atributo classe já foi determinado. No processo de construção da árvore, o atributo que possuir o maior ganho de informação deve ser colocado como nó raiz da árvore, pois será este atributo que fornecerá a maior redução na entropia, possibilitando a classificação dos dados de forma mais rápida.

Para poder determinar o valor do ganho de informação devem ser realizados os seguintes cálculos:

- Calcular a entropia conjunta, isto é, para todo o conjunto de dados;
- Calcular a entropia individual para cada atributo do conjunto de dados.

Considere um conjunto de amostras, contendo um atributo classe definido como *A* e um dos atributos preditivos definido como *B*. O ganho de informação (*GI*) do atributo preditivo *B* é definido como sendo a diferença entre a entropia do atributo classe *A* (*Entropia(A)*) menos a entropia condicional do atributo preditivo *B*, tendo sido definido o valor do atributo classe *A* (*Entropia(B|A)*). O ganho de informação será dado por (5).

$$GI(B,A) = Entropia(A) - Entropia(B \mid A) \tag{5}$$

Onde o segundo termo da equação é a entropia condicional, definida como a entropia de um atributo preditivo *B*, sendo conhecido o atributo classe *A*, é dada por:

$$Entropia(B \mid A) = -\sum_{i=1}^{m} p_i . Entropia(B \mid A = v_i) \tag{6}$$

Onde *m* é o número total de classes do conjunto de amostras, *B* é o atributo preditivo que está sendo considerado. *A* é o atributo classe assumindo o valor v_i . O termo $Entropia(B \mid A = v_i)$ é a entropia do atributo preditivo *B* sendo dado o valor do atributo classe $A = v_i$.

$$Entropia(B \mid A = v_i) = -\sum_{i=1}^{m} p(B \mid A = v_i) \log_2 p(B \mid A - v_i) \tag{7}$$

Onde m é o número de classes que o atributo classe *A* pode assumir, $p(B \mid A = v_i)$ é a probabilidade condicional do atributo *B*, isto é, a proporção dada pela razão entre o número de exemplos de *B* com $A = v_i$ e o número total de amostras na classe $A = v_i$.

Outro método desenvolvido foi o *GINI*, por Breiman, (1984) que visa minimizar a impureza de cada nó. Onde a impureza do nó é máxima quando todas as classes possuem igual distribuição e mínima quando existe apenas uma classe. Considerando um conjunto de dados S, que contém n registros, cada um com uma classe A, o índice Gini de S é dado pela equação (8).

$$Gini(S) = 1 - \sum_{i=1}^{m} p_i \left[\frac{A}{n} \right]^2 \tag{8}$$

Onde:

p_i - é a probabilidade relativa da classe *A* em *S*.

n - é o número de registros em *S*.

m - é o número de classes.

Se S for particionado em dois subconjuntos S_1 e S_2, um para cada ligação, o índice Gini dos dados particionados será dado pela equação (9),

$$Gini(S \mid \mathrm{A}) = \frac{n_1}{n} Gini(S_1) + \frac{n_2}{n} Gini(S_2) \quad (9)$$

Onde:

n_1 - é o número de exemplos de S_1.

n_2 - é o número de exemplos de S_2.

Por fim o método de divisão pelo erro de classificação, baseia-se em reduzir o erro de classificação por folha, levando em conta a probabilidade de ocorrência do registro para a classe.

$$erro\ de\ classificação\ (t) = 1 - \max[p(i|t)] \quad (10)$$

3.7 Algoritmos para construção de árvore de decisão

A tecnologia que possibilitou o aparecimento das árvores de decisão foi desenvolvida pelo Professor *R. Quinlan* (Universidade de Sidney). Em 1983 ele elaborou o algoritmo chamado *ID3* (*Iterative Dichotomiser 3*). A seguir um breve resumo das características dos algoritmos *ID3, CART e C4.5*.

O algoritmo *ID3* avalia a informação contida nos atributos segundo o seu ganho de informação. O atributo mais importante, (com o maior valor de ganho de informação) é colocado na raiz e, de forma *top-down*, a árvore é construída recursivamente. Se um atributo classifica perfeitamente um conjunto de treinamento o algoritmo pára; senão ele opera recursivamente nos subconjuntos particionados para conseguir o melhor atributo. Uma forte limitação do algoritmo *ID3* é que ele não trabalha com atributos do tipo contínuo.

O algoritmo *CART* tem como principais características a grande capacidade de gerar árvores de decisão dimensões reduzidas, elevado desempenho, e com grande capacidade de generalização.

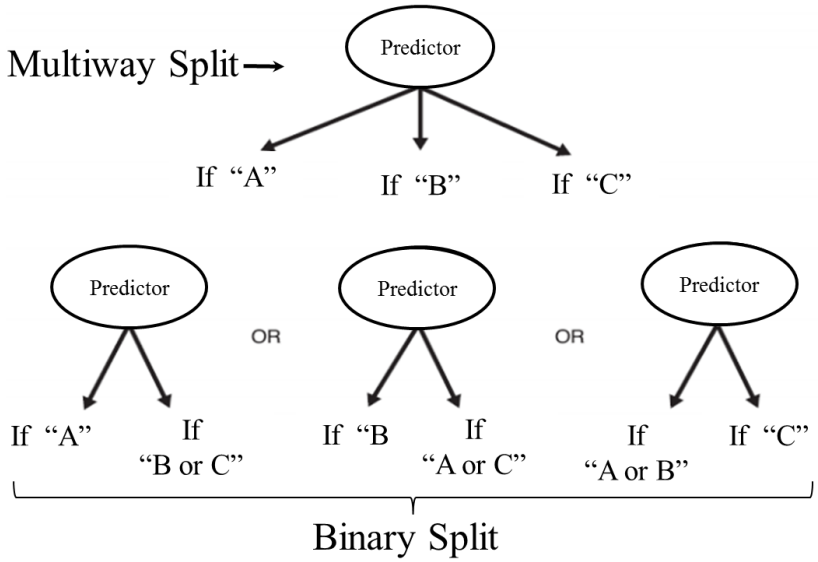
A árvore gerada é baseada na técnica recursiva de divisão binária. O processo binário consiste em que cada nó é dividido em dois subconjuntos e à medida que se percorre a árvore, do nó raiz até as folhas, são respondidas questões simples como sim ou não. A recursividade ocorre a cada subconjunto gerado, até que não seja mais possível efetuar partições da árvore. Neste algoritmo a escolha do atributo é feito normalmente pelos critérios de ganho de informação ou Gini.

Outro algoritmo muito utilizado na construção de árvore de decisão, é o *C4.5*. Durante o processo de construção da árvore de decisão no *C4.5*, os atributos categóricos

podem ser particionados de duas maneiras: um ramo distinto a cada valor do atributo ou a formação de agrupamentos de valores em vários conjuntos. Para as partições efetuadas com base em atributos contínuos é utilizado o método de pesquisa exaustiva do ponto de divisão, gerando árvores binárias. Neste algoritmo a escolha do atributo pode ser feita pelo critério de ganho de informação.

Os algoritmos de DT mais usados são o CART e C4.5. Uma vantagem desses algoritmos, em comparação com outras técnicas de aprendizagem de máquinas, são suas habilidades para lidar com atributos numéricos e categóricos. Existem várias maneiras de gerar divisões de atributos. Essas divisões influenciam em precisão, complexidade, eficiência computacional e interpretação de resultados. Essas divisões são binárias para atributos numéricos, e para atributos simbólicos podem ser binários, que sempre dividem o nó em dois ramos ou multi-caminho, que dividem o nó em tantos ramos como categorias em um atributo.

Figura 10 - Exemplo de divisão de atributos categóricos/simbólicos em uma árvore de decisão



A característica principal que diferencia esses algoritmos é o fato de que CART faz a divisão binária para atributos numéricos e simbólicos. Por sua vez, C4.5 permite divisões múltiplas para atributos simbólicos. A figura 10 ilustra duas estruturas de DTs genéricas, onde é destacado duas formas de divisão para um atributo simbólico que funciona como um preditor. Uma DT com divisão múltipla tem um ramo para cada valor do atributo preditor, ou seja, uma regra de decisão se o valor do preditor for "A", outra regra se for "B" e uma terceira regra se o preditor O valor é "C" enquanto que em um DT com divisão binária, esses valores são divididos em dois subconjuntos, pois é necessário encontrar uma boa regra de particionamento para melhor agrupar esses valores. No

entanto, esse agrupamento pode conter inconsistências ou pouco significado em problemas reais.

A Figura 10 apresenta exemplos de divisões genéricas para uma divisão multi-partes gerada por C4.5 e uma divisão binária gerada pelo algoritmo CART. Nesta tese, os atributos simbólicos são usados para representar topologias e contingências no sistema de potência. Para representar quantidades elétricas como tensão e potência, os valores numéricos resultantes de estudos de simulação e medições reais obtidas a partir da operação do sistema são usados na etapa de treinamento DT. Devido às características do problema abordado, a divisão multi-caminho é usada como atributos categóricos, ou seja, em cada ramo MDT são representadas as variáveis e limites que correspondem a uma topologia de rede. A divisão binária é usada para atributos numéricos. Essas escolhas são justificadas como uma maneira conveniente de implementar a avaliação de segurança e controle preventivo.

3.8 Estimação de Erro do Classificador

É de fundamental importância estimar a qualidade do classificador utilizado, na forma de percentual de erro do modelo gerado. Pode definir-se o erro de um classificador como sendo.

$$P_{e\%} = \frac{N_{erros}}{N_{casos}} \tag{11}$$

Onde:

$P_{e\%}$ é o percentual de erro;

N_{erros} é o números de casos classificados erroneamente;

N_{casos} é o número total de casos testados.

Uma importante ferramenta para a análise de desempenho de um classificador é a chamada de matriz de confusão, tabela 3.

Tabela 3- Exemplo de Matriz de Confusão

Matriz de Confusão		Classe Real	
		Classe A	Classe B
Classe Predita	Classe A	PV	PF
	Classe B	NF	NV

A matriz de confusão é uma matriz que contém os dados dos testes realizados no modelo induzido, tais como:

1. PV – Positivos verdadeiros, ou seja, o número de exemplos classificados como “Classe A” que são verdadeiramente “Classe A”.
2. PF – Positivos falsos, ou seja, o número de exemplos classificados como “Classe A” que são do tipo “Classe B”.
3. NV – Negativos Verdadeiros, ou seja, o número de exemplos classificados como “Inseguros” e são verdadeiramente “Classe B”.
4. NF – Negativos Falsos, ou seja, o número de exemplos classificados como “Classe B” que são do tipo “Classe A”.

A partir da matriz de confusão é possível calcular alguns índices de desempenho para o classificador, estes são:

A taxa de acerto que é definida pela equação 12

$$P_{a\%} = \frac{PV + NV}{PV + PF + NV + NF} \tag{12}$$

A precisão de classe, que é a proporção de positivos verdadeiros do total de exemplos classificados como positivos dados pela equação 13.

$$P_{c\%} = \frac{PV}{PV + PF} \tag{13}$$

E a sensibilidade que é a proporção de exemplos positivos corretamente classificados, e equação 14 define a sensibilidade positiva.

$$P_{s\%} = \frac{PV}{PV + NV} \tag{14}$$

Para que seja possível avaliar o poder de classificação de um modelo é necessário que haja um conjunto de dados para testá-lo, verificando se as regras do classificador conseguem classificar de forma correta esse conjunto.

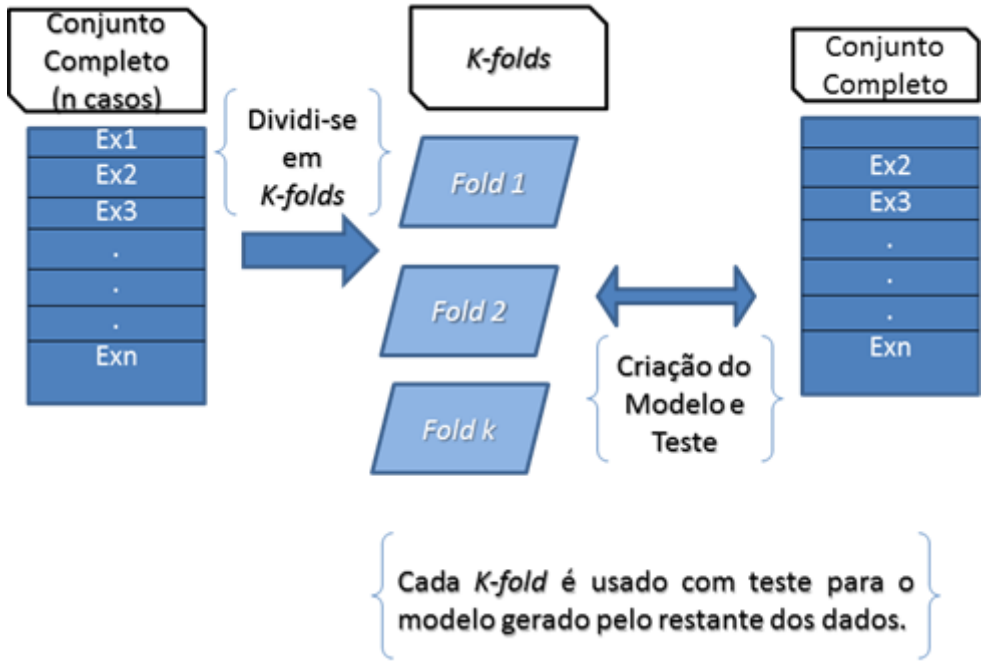
Acontece que muitas vezes não existem dois ou três bancos de dados separados, sendo, por exemplo, um para treino e outro para teste e um terceiro para validação. Quando ocorre essa situação, lança-se mão de algoritmos que particionam o conjunto de dados que seria usado apenas para treino, permitindo assim que os modelos gerados possam ser testados, mesmo havendo apenas um conjunto de dados.

3.9 Validação Cruzada (*Cross-Validation)*

É um dos algoritmos mais utilizados para conjuntos de dados que não possuem muitos exemplos. Ele consiste em usar os mesmos dados, de forma repetida, porém

divididos de forma diferente, ou seja, o conjunto de dados é dividido aleatoriamente em *k* partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual (*n/k*).

Figura 11 - Exemplo de CrossValidation



Os exemplos de (*k*-1) *folds* são independentemente usados no treinamento e os classificadores obtidos são testados com o *fold* remanescente. O processo é repetido r vezes, e a cada repetição um *fold* diferente é usado para tese de modelo.

A estimativa do erro verdadeiro é obtida como a média dos erros de cada experiência, equação 15:

$$E = \frac{1}{k}\sum_{i=1}^{k} E_i \qquad (15)$$

3.10 Conclusão

A partir de técnicas de mineração de dados e aprendizado de maquinas, muitas tarefas e análises em diversas áreas do conhecimento, ganharam agilidade e abordagens que promove a interação das informações, que muitas vezes eram negligenciadas pelas equipes de trabalho.

A importância de analisar a base de dados vem se mostrando uma ferramenta poderosa. Desta forma o setor elétrico não pode ficar à margem desses avanços.

Capítulo 4 - Avaliação da Segurança de Sistemas de Potência com Árvore de Decisão Multicaminhos.

4.1 Introdução e Considerações Sobre a Metodologia

O presente capítulo apresenta uma metodologia para avaliação da segurança estática e dinâmica de sistemas de potência (SDSA), com foco no monitoramento dos níveis de tensão e na estabilidade transitória, utilizando árvores de decisão multicaminhos (MDT). A metodologia consiste das etapas de composição da base de dados, a construção das árvores de decisão e análise para operação *online.* Quanto à criação da base de dados, duas metodologias são propostas, sendo elas: uma baseada em dados de planejamento e outra baseada no histórico operacional recente.

Os modernos sistemas elétricos de potência possuem uma vasta gama de informações provindas de seus sistemas de medição, gerando assim enormes bancos de dados que contém um grande número de informações sobre as condições operacionais do sistema em geral. Bases de dados comumente encontradas são:

- Atuação de alarmes e dispositivos de proteção, ou data logging ou SOE – sequential of events;
- Níveis de carregamento das linhas;
- Medidas elétricas como tensão, corrente, frequência, potência ativa e reativa, posição de TAPs de transformadores, além de temperatura e etc.
- Topologias;

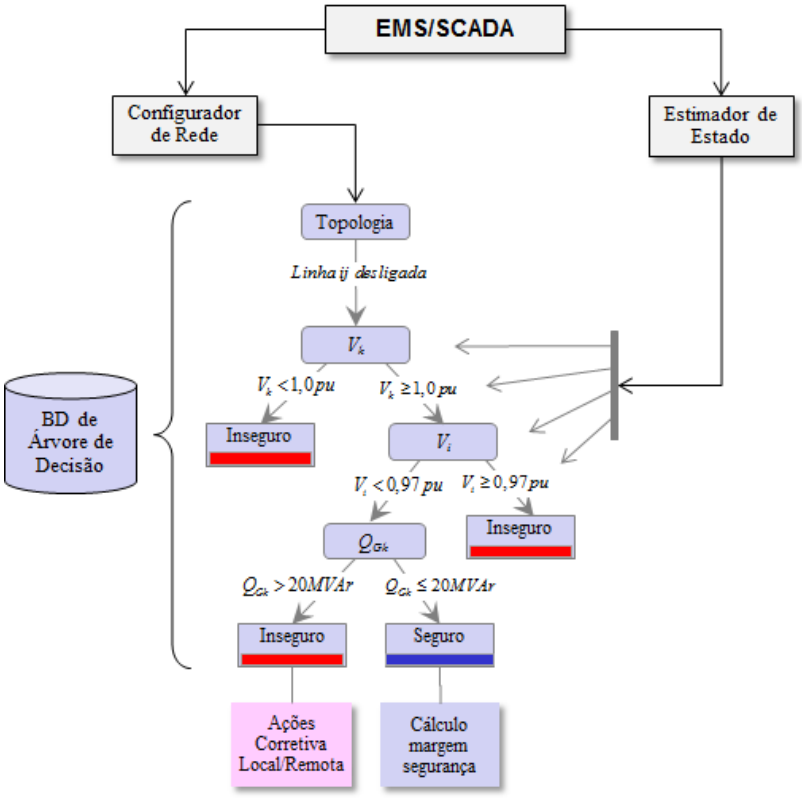
Em geral armazena-se uma enorme gama de variáveis que podem ser úteis para as diversas análises no sistema. Essas informações podem ser tratadas por diversos processos analíticos e estatísticos, que são métodos mais tradicionais. Com o avanço de técnicas computacionais, a mineração de dados pode-se descobrir associações que até então não eram percebidas pelos engenheiros e operadores das empresas. Assim, esse conhecimento pode ser muito útil para análises de planejamento, manutenção, além de outras atividades do setor.

Das diversas aplicações da MD em SEPs, a avaliação da segurança vem ganhando destaque. Para tal função, a tarefa de classificação na mineração de dados é uma alternativa que cumpre de maneira satisfatória esse papel. Sendo possível utilizar diversas técnicas, tais como: redes neurais, máquinas de vetor e suporte, classificadores bayesianos e árvores de decisão. Contudo, as árvores de decisão atendem os requisitos de alto

desempenho de classificação, com grande vantagem frente às outras técnicas por sua fácil interpretabilidade.

A finalidade básica da utilização da MDT, é classificar o ponto de operação do sistema como “SEGURO” ou “INSEGURO”, de acordo com um modelo de tomada de decisão, que converte o processo tradicional de avaliação de segurança em um conjunto de instruções lógicas do tipo “Se-Então”, em termos dos limites dos atributos de entrada. Onde tais atributos são avaliados, a fim de verificar possíveis combinações dos mesmos que se traduzam em regras para auxilio da operação e controle preventivo dos SEPs. Para ser utilizado como uma ferramenta de SDSA em um centro de controle e operação de uma concessionária ou um operador de rede é necessária que seja feita a integração do conjunto de instruções logicas da MDT com o sistema de supervisão e controle existente.

Figura 12 - Modelo de SDSAon-line usando DT.



(Rodrigues, 2014)

A figura 12 ilustra um esquemático de como pode ser realizada a integração das instruções logicas da DT com o sistema EMS (do inglês, “*Energy Management System*”) de uma concessionária, onde a BD de árvore de decisão contém as instruções logicas da DT, a qual serão submetidos os valores provenientes do estimador de estados, juntamente com dados do configurador de redes. A partir do teste dos limites das variáveis vindas do estimador pela DT, são classificados como “SEGURO” ou “INSEGURO”, podendo

serem encaminhados a um módulo de ação corretiva, caso seja classificado como inseguro. Caso seja classificado como seguro, pode-se calcular as margens de segurança.

Para que seja possível a análise ou até mesmo o desenvolvimento de ferramentas para a avaliação de um sistema de potência, o uso de simulações é de fundamental importância, pois nelas são modelados os sistemas e são aplicadas as contingências desejadas, além do que, dificilmente, encontram-se disponíveis dados rotulados nas múltiplas bases de dados das empresas do setor elétrico, que podem ser utilizadas para tal finalidade. Entretanto, um dos aspectos importantes da metodologia proposta neste capitulo, é a utilização de topologias e pontos de operação reais, como casos bases na concepção da base de dados para treino da DT, e não apenas dados de planejamento, como é usualmente realizado na literatura. Contudo, é importante ratificar que quase na sua totalidade do tempo, os SEPs operam em condições consideradas normais, ou seja, sem grandes perturbações no modo de operação do mesmo. Logo, obter uma base de dados, suficientemente grande e eficiente, que leve em consideração todos os possíveis estados operacionais dentro de todas as possíveis contingências, previsíveis ou não, a que o sistema está sujeito, podem demandar anos e anos de coleta de dados nos históricos operacionais.

A seguir são apresentadas as ferramentas e técnicas utilizadas, assim como as características de simulações para a criação da base de dados para treinamento da árvore de decisão para SDSA.

4.2 Criação da Base de Dados para treino e validação baseados em dados de Planejamento – Base de dados estática e dinâmica.

Projetos de ferramentas para SSA e DSA usando ML, seja DT, RNA, SVM ou SOM dependem da eficiência, qualidade e grau de representatividade da base de dados usada para seu treino. Os passos seguidos nos trabalhos relatados na literatura são normalmente os seguintes:

1. Previsão de carga de 24 horas a frente, disponibilidade de geradores e topologias e modelos do sistema.
2. Execução de fluxo de carga.
3. Realização da análise de contingências em regime permanente, simulação no domínio do tempo ou função energia, dependendo do foco da aplicação.
4. Associação de cada cenário operativo criado em 1 e 2 com as respostas pós-contingências.

5. Selecionar e tratar os atributos desejados para o treino da ML.
6. Treinar e avaliar o modelo gerado pela ML.

De posse de dados sobre previsão de carga de 24 horas à frente, e disponibilidades de equipamentos (linhas, transformadores, compensadores, geradores) é possível montar cenários operativos em carga leve, média e pesada onde, na maioria das vezes, para cada um desses cenários tem-se uma topologia padrão para operação. Levando isso em conta, seleciona-se alguns cenários de operação para criação de 24 pontos de operação bases (POB), um para cada hora. Estes pontos de operação podem ser gerados por fluxo de carga ótimo para fins de despacho econômico.

Neste trabalho, partir dos POBs são gerados novos patamares de carga/geração para todas as barras de carga do sistema, por meio de uma distribuição gaussiana, com variações no valor de carga previsto, incorporando um grau de incerteza para a previsão de carga. Além disso, uma lista de contingências pré-determinadas é aplicada a cada um dos OPs, gerando N novos casos, a serem estudados para cada topologia, ou seja, novos cenários de operação (NCO), dessa forma tem-se um significativo número de cenários operativos, com diversas topologias previstas para operação e suas respectivas contingências críticas. Neste trabalho, a escolha das contingências não obedece a nenhum método detalhado. A discussão sobre os métodos de seleção está fora do escopo do trabalho.

Os fluxos de carga são executados para os NCOs. Todas as simulações são realizadas no software ANAREDE do Cepel. As simulações dinâmicas no dominio do tempo são realizadas no software ANATEM, também do Cepel, a partir dos casos convergidos de fluxo de carga executados. Em seguida, é realizada a etapa de tratamento dos relatórios gerados pela execução do fluxo de carga, essa etapa é conhecida em mineração de dados como pré-processamento, onde são selecionados e tratados os atributos que serão apresentados ao algoritmo de MDT que avaliará a segurança estática, ou seja, as variáveis elétricas que são calculadas pelo fluxo de carga e são possíveis de serem coletadas do campo por sistemas SCADA ou PMU’s. As variáveis extraídas do relatório são:

- Módulo da tensão na barra i (Vi);
- Defasagem de tensões entre as barras i e j (θij);
- Potência ativa do gerador síncrono n (Pgn);
- Potência reativa do gerador/compensador síncrono n (Qgn);

- Fluxo de potência ativa entre as barras i e j (Pij)
- Fluxo de potência reativa entre as barras i e j (Qij);
- Estado de segurança (rótulo/label)

Os códigos desenvolvidos para a tarefa de pré-processamento extraem seis rótulos que podem ser utilizados para avaliação estática e uma para segurança dinâmica. São eles os rótulos simbólicos (seguro/inseguro) para violação de tensão, violação de fluxo de potência nas linhas e transformadores e um para violação de limites de geração de potência reativa nos geradores. Os códigos de pré-processamento também extraem os indíces de severidade das contingêncicas para os três tipos de avaliação estática. A metodologia baseada em MDT utiliza os rótulos simbólicos. Os indíces de severidade podem ser utilizados em trabalhos futuros.

Como exemplo de aplicação, a seguir são descritos os passos para a avaliação de tensão. O estado de segurança aqui é dado como inseguro, caso algum valor de módulo de tensão em qualquer barramento do sistema viole os limites pré-estabelecidos. Em geral, esses limites variam de acordo com o tipo da barra. Em seguida é inserida, como atributo em cada caso, uma identificação de topologia (TIK), que consiste de um atributo simbólico correspondente à combinação de todos os elementos ligados no sistema. Essa TIK é obtida dos dados do configurador de rede presente no EMS, dessa forma, cada TIK é um nome dado para uma topologia ou contingência, para facilitar a identificação do operador na MDT. Esses dados são organizados com numa matriz NxD, que contém os seguintes elementos:

$$\left[TIK \cdots V_i \cdots \theta_{ij} \cdots P_g \cdots Q_g \cdots P_l \cdots Q_l \cdots P_{ij} \cdots Q_{ij} \cdots Status\right] \quad (16)$$

Para que modelos de MDTs tenham bom desempenho de classificação é fundamental que as distribuições entre classes sejam iguais, seguro ou inseguro neste caso. Isto é, aproximadamente 50% do total de casos para cada um deles. Contudo, gerar simulações em que todos os casos gerados respeitem esse critério de maneira automática é bastante difícil, tendo isso em vista, simula-se uma grande quantidade de casos e em seguida utiliza-se a amostragem aleatória estratificada, a fim de se obter uma base de dados que respeite a 2 critérios para garantir um melhor desempenho da MDT. O primeiro critério é um número igual de simulações para cada TIK simulada. O segundo critério é que dentro de cada TIK o número de casos seguros e inseguros tenham a mesma proporção, o que aumenta a confiança do resultado de classificação.

A amostragem estratificada garante que tais critérios sejam respeitados, pois identifica subgrupos significativos (estratos), que podem ser definidos de acordo com o objetivo do estudo. Em seguida, calcula-se ou define-se o peso relativo (%) de cada estrato na população, isto é a proporção de cada estrato em relação ao total, uma vez que a representação de cada estrato na amostra deve ser feita na mesma proporção em que esse subgrupo está representado na população. Esse estágio é repetido, levando em consideração o status de segurança como um segundo estrato, a fim de se obter uma base de dados uniforme em quantidade de casos e que represente bem as variações de condições operacionais do sistema.

A etapa de seleção atributos (pré-processamento) inclue uma fase de escolha de variáveis que são apresentadas ao algoritmo de ML. De maneira geral, uma MDT mapeia e relaciona os dados de entrada com o atributo alvo. Contudo, para uma melhor compreensão e aplicação dos resultados obtidos, a escolha dos atributos usados é de fundamental importância.

Isto significa afirmar, no caso da avaliação de segurança, primeiramente, que as variáveis escolhidas estarão disponíveis para o operador quando o modelo baseado em ML estiver sob operação em tempo real. Um segundo aspecto a se analisar em relação a disponibilidade das variáveis, é o grau de liberdade para controle dessas variáveis, isto é, a variável é controlável por ação direta de controles locais, é possível controlá-la por meio de uma ação remota, ou ela é apenas uma consequência de fatores operacionais da rede.

Neste trabalho buscou-se utilizar variáveis que traduzam com maior fidelidade possível as condições da rede disponíveis ao operador, por meio de sistemas EMS, que contenham funções de estimação de estado baseadas em medidas de SCADA ou PMU, configuradores de redes e previsão de carga. Outro fator importante na escolha dos atributos é sua disponibilidade de controle, mesmo que a proposta aqui seja uma MDT para monitoramento e avaliação da segurança de tensão, é importante que as variáveis da MDT sejam passiveis de ação no ambiente de operação. Desta forma, os atributos e suas justificativas para escolha, após diversos testes de combinação são listados abaixo.

- Tensões em todas as barras do sistema teste – Estas são as variáveis mais importantes a serem avaliadas pela MDT, pois é a partir delas que é gerado o rótulo de segurança. Testes utilizando apenas barras com controle local de tensão não apresentaram MDTs com boas performances, porém sabe-se que é possível tomar ações de controle remoto para barras sem possibilidade de controle local.

- Potências ativas e reativas geradas – A geração de potência ativa e a compensação de potência reativa, e suas respectivas reservas são fatores que influenciam na margem de estabilidade de tensão e na capacidade de manter as tensões dentro de uma faixa de operação segura. Com a possibilidade de re-despacho via fluxo de carga ótimo, estas variáveis de rede são excelentes candidatas a atributos da MDT.
- Valores da carga nas barras dos sistemas – Essas são importantes para situar o cenário operativo, além de servir como variáveis para esquemas de corte de carga de emergência.
- TIK – Atributo simbólico que representa a topologia em operação e suas respectivas contingencias n-1.

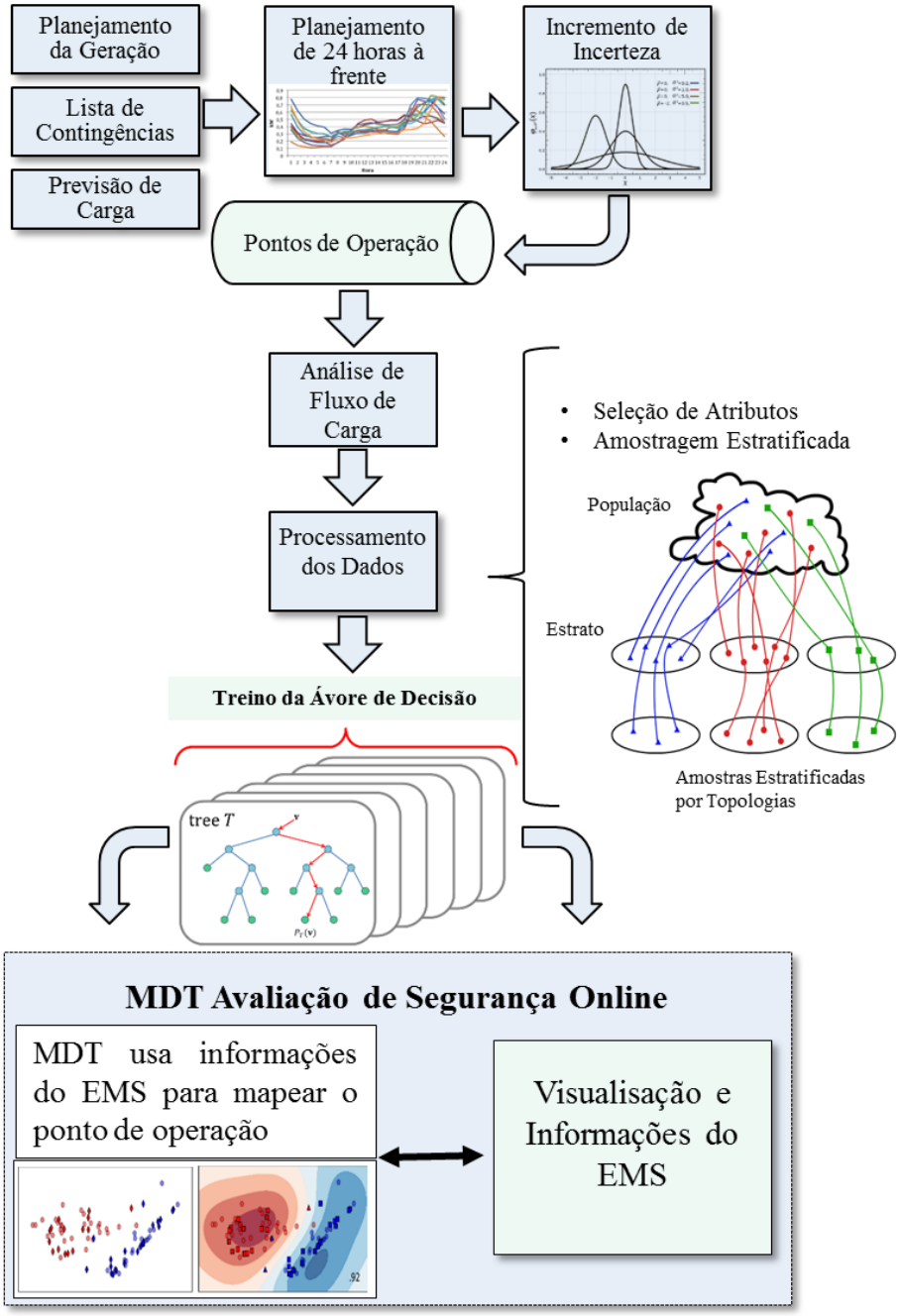
As variáveis θ_{ij}, P_{ij} e Q_{ij} não apresentaram participação nas MDT's nos testes realizados e são variáveis de difícil controle, por esses motivos são removidas da base de dados de treino das MDT's, para aumentar a eficiência computacional. A base de dados que é utilizada no treino da MDT é representada por uma matriz MxK e que contém os dados contido em (17).

$$\left[TIK \cdots V_i \cdots P_g \cdots Q_g \cdots P_l \cdots Q_l \cdots Status\right] \tag{17}$$

A partir da base de dados criada seguindo a proposta descrita em 4.2, é realizado o treino e teste da MDT. O software utilizado para os treinos e testes dos modelos de DT foi o Rapidminer, onde se utilizou os modelos pré-definidos do software para amostragem da base de dados e avaliação do modelo de MDT obtido, bem como uma extensão chamada WHIBO (Delibasic, 2013; Vukićević, 2012), onde é possível escolher os parâmetros para treino da DT com maior flexibilidade que o modelo padrão do Rapidminer. Esta extensão permite usar técnicas de remoção de atributos insignificantes, escolher o tipo de divisão para atibutos categóricos, métodos para avaliar as divisões, critérios de parada e de poda.

Para criação das divisões foram escolhidos Multiway (Multicaminhos) para atributos categóricos e Binary (binário) para atributos numéricos, para avaliação da divisão foi escolhido Gain Ratio. Como critério de parada foram usados a profundidade da MDT, um tamanho minimo para o nó e confiança no nó. Para a validação do modelo gerado usou-se a técnica validação cruzada K-fold, que é aplicado com eficiência em base de dados extensas. A figure 13 apresenta o fluxograma que sintetiza a seção 4.2.

Figura 13 - Fluxograma da metodologia



4.3 SAGE - EMS/SCADA - Aquisição de Dados Reais e Criação de Base de Dados

Para ser utilizado como uma ferramenta de SDSA em um centro de controle e operação de um SEP, é necessário que seja feita a sua integração com o sistema de supervisão e controle existente ou implantação de um novo, o que pode ser inviável por diversos fatores. A abordagem proposta pode ser implementada dentro de um sistema EMS produzido pelo CEPEL

O SAGE (Sistema Aberto de Gerenciamento de Energia), é um sistema computacional que tem por objetivo executar as funções típicas, de um sistema SCADA,

de gerenciamento em sistemas elétricos de potência (CEPEL, Sistemas SCAD/EMS Visão Geral). As principais características do SAGE são:

1 - SAGE/SCADA - Supervisão e controle de redes elétricas.
2 - SAGE/EMS - Análise de Redes em tempo real e de estudos.
3 - SAGE/GBH - Gerenciamento da base de dados históricos.
4 - SAGE/SIA - Subsistema de Inteligência artificial.
5 - SAGE/AGC - Controle Automático de Geração.
6 - SAGE/SIM - Subsistema de Treinamento e Simulação.

A função EMS permite a monitoração confiável da condição operativa do sistema, informando sobre situações indesejáveis. É possível também produzir estratégias de controle buscando melhor ponto de operação. O EMS é composto de diversas aplicações. Elas são divididas em módulos de aplicações de tempo real, que processam dados provenientes do subsistema de aquisição e comunicação de dados do SAGE. São funções executadas periodicamente de forma automática ou por solicitação do operador. As outras aplicações de modo de estudo são executadas a pedido do operador. Estas simulações ou análises de uma condição operativa são classificadas como passada (estudos de pós-operação), presente (proveniente do modo de tempo real) ou futura (simulação de novos pontos de operação).

Tabela 4 - Aplicações do SAGE/EMS

Aplicações de tempo real:	Aplicações de modo de estudo:
• Configurador da Rede	• Fluxo de Potência Convencional
• Estimador de Estado	• Análise de Contingências
• Análise de Contingências	• Análise de Sensibilidade
• Controle de Emergências	• Equivalente de Redes
• Análise de Estabilidade de Tensão	• Fluxo de Potência Ótimo

A tabela 4 contém algumas funções que o sistema SAGE oferece ao seu usuário, sendo que estas podem ser utilizadas em modo de operação tempo real, ou mesmo, como modo de estudos. Por se tratar de um sistema aberto, é possível implementar algumas aplicações dentro do SAGE. Baseado nessa possibilidade, o estimador de estados ganha uma nova funcionalidade, a partir do armazenamento dos resultados de sua execução. A cada execução do estimador, obtém-se um ponto de operação real do SEP, assim, é

possível armazenar tal ponto de operação em um banco de dados ou criar um arquivo com os dados para execução de fluxos de cargas posteriores. No entanto, o próprio SAGE possui uma funcionalidade que fornece um arquivo próprio para execução de fluxos carga, arquivos com extensão "PWF", esse arquivo pode ser executado via a ferramenta ANAREDE, também do Cepel. O SAGE pode exportar arquivos "PWF" a cada um minuto, ou quando ocorre mudança topológica na rede (abertura ou fechamento de chaves). Contudo, para esse trabalho, foram utilizados arquivos exportados a cada 5 minutos, o que resultou em um número de 288 arquivos a cada 24 horas.

Além do histórico operacional, outra informação coletada do SAGE é o estado topológico do sistema em cada um dos momentos que são exportados os arquivos PWF. Com isso é possível extrair o atributo simbólico para representar o estado topológico da rede elétrica.

Essas informações devem estar presentes em quaisquer sistemas SCADA/EMS, portanto uma abordagem que se baseie no de histórico operacional pode ser adotada em qualquer centro de operação, necessitando apenas de processos de customização, aproveitando assim todo conhecimento armazenado em suas bases de dados históricas. Contudo, é importante salientar que a ferramenta será para SDSA online, então, informações somente do passado podem ser insuficientes para avaliar a segurança do SEP em determinados pontos de operação futuros. Isso é devido as caracteristicas exigidas de uma base de dados para treinamento de uma árvore de decisão, isto é, a necessidade de haver casos suficientes de pontos de operação seguros e inseguros distribuídos pela base de dados.

4.3.1 Criação e Tratamento da Base de Dados

De posse dos dados do estimador de estado, são necessárias algumas etapas de pré-processamento nos mesmos, que por sua vez devem fornecer uma boa representação do sistema sob estudo, bem como colocá-los em um formato necessário para a execução do programa minerador.

Como descrito no item 4.3, o sistema SAGE conta com a possibilidade de exportação do ponto de operação atual já no formato "PWF", utilizado por outros programas do Cepel, para que possam serem executados nas suas aplicações de modo de estudo. Contudo, os arquivos no formato "PWF" contêm basicamente informações para que seja executado o fluxo de carga. Tais dados denotam as variáveis do sistema:

- Dados de Barras (DBAR) – Informações como nome e número da barra, tensão (módulo e fase), valores de carga, geração, capacitores *SHUNT*e outras informações de controles.
- Dados de Linhas (DLIN) – Informações sobre as linhas (impedância e conexões do sistema, transformadores).
- Dados *Shunt* de Linhas (DSHL) – Capacitâncias *SHUNT* entre as linhas.
- Dados de Grupos de Limites de Tensão (DGLT) – Onde é possível colocar diversas possibilidades de variações nos limites de tensão.
- Dados de Geração (DGER) – Com limites máximos e mínimos de cada grupo gerador, bem como fator de potência de operação.
- O arquivo "PWF" contém outras informações para execução, que são melhores tratadas em seus manuais.

O programa ANAREDE possui uma funcionalidade que permite executar automaticamente diversos fluxos de cargas em sequência. Assim é possível alterar, por exemplo, carga/geração, topologias de redes e ajustes de controles. Baseado nessa funcionalidade, os arquivos que contém os pontos de operação selecionados são agrupados em um único arquivo "PWF", que contém um vasto conjunto cenários operativos. Desta forma, é possível executar de uma única vez, diversos fluxos de carga, gerando relatórios com todas as informações para a avaliação de segurança estática e um conjunto de arquivos (.SAV) que serão utilizados para a simulação dinâmica pelo programa ANATEM, juntamente com os arquivos "STB" e "BLT.

A partir das execuções dos fluxos de cargas e simulações dinâmicas, os relatórios gerados pelo ANAREDE são tratados, extraindo as variáveis desejadas e colocando-as no formato no qual o programa de mineração de dados suporta. As variáveis pré-selecionadas a partir dos relatórios são:

- Data (dia/mês/ano);
- Hora (hora:minuto:segundo).
- Módulo da tensão na barra i (Vi);
- Defasagem de tensões entre as barras i e j (θij);
- Potência ativa do gerador síncrono i (Pgi);
- Potência reativa do gerador/compensador síncrono i (Qgi);
- Fluxo de potência ativa entre as barras i e j (Pij)
- Fluxo de potência reativa entre as barras i e j (Qij);

- Estado de segurança estático (Label indicando se houve ou não violação de tensão);
- Estado de segurança dinâmico (Label indicando se houve ou não problemas de instabilidade transitória)

As variáveis de data e hora são úteis para fazer a inclusão da chave topológica da rede, enquanto as demais variáveis entraram na base de dados afim de serem associadas pela árvore de decisão com os rótulos de segurança estática e segurança dinâmica.

4.3.2 Complemento da Base de Dados

Durante um dia de operação, são produzidos um número de aproximadamente 280 casos PWF, ou seja, são gravados 280 cenários operacionais. Para se treinar uma DT que avalie a segurança de SEP, essa quantidade de cenários não é suficiente para se obter bons resultados de classificação da MDT. Para contornar essa deficiência e preparar a base de dados para atender não só as topologias usuais do sistema, mas também para as contingências mais críticas na qual o SEP possa vir a enfrentar, são adotadas 3 estratégias que visam melhorar as características da base de dados e consequentemente o resultado de classificação via árvore de decisão. Estas estratégias são sequidas sequencialmente como apresentado abaixo:

Passo 1 - Uso de Dados de dias anteriores;

Passo 2 - Novos casos a partir de previsão ou aleatórios;

Passo 3 - Simulação das contingencias (n-1) a partir dos casos do passo 2;

Cada um desses passos é detalhado a seguir.

4.3.2.1 Prolongamento do Histórico Utilizado

Com apenas os 280 casos que são criados durante um dia de operação, a partir dos dados exportados do SAGE, uma MDT não conseguiria obter resultados satisfatórios frente a novos pontos de operação. Duas possíveis soluções podem minimizar esse problema: incremento da base de dados por simulação ou uso de mais dados históricos.

Os dados históricos de curto prazo são bastante úteis para mostrar as tendências de operação, tanto do ponto de vista de carga, topologias, ajustes de controles como instruções para operadores. Então, uma abordagem que incorpore esse conhecimento aumenta a confiabilidade da MDT, já que a base de dados é o alicerce de um bom modelo de classificação.

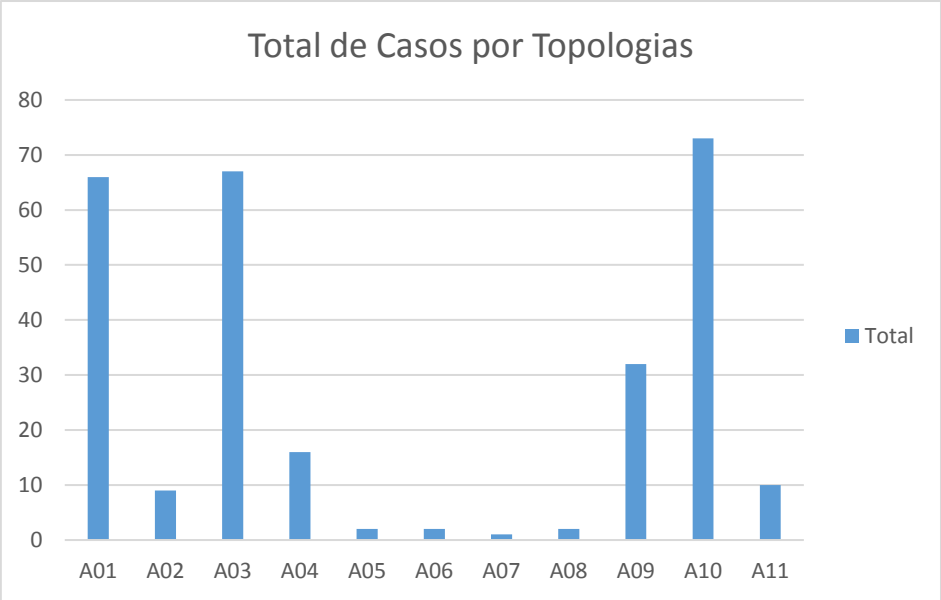
É de conhecimento de especialistas em operação, que de maneira geral, o modo de operação do sistema, é bem conhecido em condições normais, isto é, é possível fazer previsões com elevado nível de precisão. Essas previsões baseiam-se em amostras do passado e algumas informações futuras. Contudo, essas informações do passado, são essencialmente os valores de carga nas barras de carga do sistema em estudo. O diferencial da proposta é utilizar não somente os valores de somatório de cargas ou seus valores individualizados por barras e sim os pontos reais de operação. Desta forma, uma multiplicidade de pontos de operação em torno de um ponto tido como caso base, para uma dada topologia é a base das informações para a geração do modelo pela MDT.

É também conhecido o comportamento sazonal dos SEPs. Diante disso se classificam as curvas de carga em dois grupos: curvas de carga para dias úteis, que são bastante semelhantes, e as curvas de sábados, domingos e feriados. De posse dessas informações adotou-se a seguinte estratégia para selecionar qual o período de dados passados a ser utilizado:

- Utilização de histórico operacional de 5 dias uteis para geração da base de dados de treino da MDT para dias uteis, isto é, usa-se o passado recente para planejar o dia futuro.
- Utilização de 2 finais de semanas (2 sábados e 2 domingos) para utilização em dias não uteis (feriados).

Assim respeita-se a sazonalidade das cargas, as topologias corriqueiras, sazonalidade de geração e intercâmbios de potência do sistema estudado. Esse passado recente tende a ser muito próximo de uma condição de operação futura de maneira geral (carga/geração). Com essa estrutura de utilização de 5 dias passados (para dias uteis), passa-se a ter algo em torno 1400 pontos de operação reais e 1120 para os outros dias (sábados, domingos e feriados), com suas diversas topologias. Contudo, a distribuição desses pontos de operação não é uniforme entre as topologias encontradas no histórico do sistema, a figura 14 apresenta o número de ocorrência das topologias ocorridas em um dia de operação (08/11/2013), onde A1 até A11 são topologias que ocorreram no dia, deixando claro que a distribuição das configurações topológicas não é homogênea.

Figura 14 - Distribuição de Topologias Durante um período de Operação



(Autor)

Entretanto, essa não homogeneidade do número de pontos de operação nas diversas topologias criam um problema para a metodologia, pois a mesma se baseia em MDT orientada por topologias, isto é, cada atributo simbólico (chave de identificação topológica) possui um ramo e em seguida as variáveis associadas (tensões e potências) mais significativas dentro daquele cenário. Para que uma MDT tenha esse comportamento, o número de pontos de operação em todas as topologias deve ser igual ou muito próximo, para que a probabilidade de ocorrência na base de dados e entropias das mesmas tenham valores próximos, forçando-as a serem variáveis de para a divisão nas regras da MDT.

4.3.2.2 Complemento da Base de dados

Simulações de novos pontos de operação são realizadas, para solucionar o problema da não homogeneidade no número de casos em cada topologia, de modo a igualar o número de pontos de operação por topologia. Esses novos pontos de baseiam em uma variação de +/- 5% em torno dos valores de carga dos pontos de operação selecionados, a escolha dos pontos é realizada de acordo com a distribuição das topologias ao longo do período sob estudo.

O número de simulações para cada topologia varia de acordo com sua quantidade de casos reais existentes. Contudo, o número total de pontos de operação para cada topologia é controlado, sendo no valor de 500 pontos de operação para cada uma delas. Assim, para esse dia da figura 14 com 11 topologias, teriam 5500 pontos de operação. No

entanto, as topologias podem ser utilizadas em vários horários diferentes, de acordo com a demanda, por exemplo, a topologia identificada como A01. A tabela 5 apresenta detalhadamente os horários em que ocorreu a operação com a topologia, o número de casos exportados pelo SAGE e o número de simulações que devem ser realizadas *off-line* para se manter a homogeneidade. Portanto, de acordo com a necessidade, são gerados novos pontos de operação em torno de cada um dos casos reais.

Tabela 5 -Topologias (Horário e Repetições)

Topologia	Horário	Exportados SAGE	Off-line
A01	(00:02 até 01:34) - (09:32 até 13:27)	66	434
A02	(01:39 até 02:21)	9	491
A03	(02:26 até 08:04)	67	433
A04	(08:10 até 09:27)	16	484
A05	(13:33 até 13:38)	2	498
A06	(13:44 até 13:49)	2	498
A07	(13:54)	1	499
A08	(13:59 até 14:04)	2	498
A09	(14:10 até 16:48)	32	468
A10	(16:54 até 23:03)	73	427
A11	(23:09 até 23:55)	10	490
Número total de casos (SAGE/Simulados)		280	5.220
		Total	5.500

4.3.2.3 Simulação de contingências e Simulação Dinâmica

A partir da base de dados composta de casos reais e casos simulados em torno deles, é possível realizar mais estudos relacionados à segurança. Portanto, uma lista de contingências é aplicada a cada caso da base de dados, multiplicando assim o tamanho da base de dados e o número de topologias a qual o sistema pode vir a operar.

Tomando em consideração essa lista de contingências, que é criada a partir de estudos *off-line* e experiências da operação, tem-se uma base de dados para criar uma DT para o monitoramento das tensões no sistema e contingencias n-1.

Por fim, para cada cenário operativo gerado, são efetuadas simulações dinâmicas para avaliar a segurança dinâmica da rede. Rotulando cada caso como seguro ou inseguro frente as perturbações, avaliando se as máquinas são mantidas em sincronismo.

A base de dados para DSA é constituída pelos mesmos atributos da base para SSA, ou seja, focada em avaliar se o ponto de operação é robusto a uma perturbação do sistema, caso ela venha a ocorrer.

4.3.3 Inclusão do Atributo Simbólico “Topologia"

A partir do dos disjuntores, das chaves seccionadoras e da disponibilidade dos equipamentos da rede, a função “configurador de redes” de SAGE determina a topologia atual da rede elétrica. Para cada alteração que ocorre na topologia da rede, devido à ação do operador ou a contingências, o configurador atualiza os dados resultando em um novo modelo da rede. Essas mudanças são armazenadas, juntamente com a data e a hora em que ocorreram.

Para o aproveitamento dessas informações da topologia, faz-se uma consulta à base de dados históricos a partir do SAGE. Onde é informada a hora de início até a evolução do estado de todos os equipamentos. Por exemplo, para saber como se deu a evolução dos estados dos equipamentos para o dia 1 de dezembro de 2013 até o instante 12:30:00, executa-se: “estado_equipamentos 2013_12_01 15:30:00” (a hora é fornecida em UTC). O resultado seria o seguinte:

bh_dthr		/ Id	/ Estado
2013-12-01	00:13:19-03|	TUTF7-09|	0
2013-12-01	00:13:19-03|	TUBR2-UG9|	0
2013-12-01	00:45:55-03|	TUTF7-09|	1
2013-12-01	00:45:55-03|	TUTF7-11|	0
2013-12-01	00:45:55-03|	TUBR2-UG11|	0
2013-12-01	00:50:00-03|	MCRE7-03|	1
2013-12-01	00:50:00-03|	TUTF7-11|	1
2013-12-01	01:30:19-03|	TUTF7-08|	0
2013-12-01	01:30:19-03|	TUBR2-UG8|	0
2013-12-01	01:54:21-03|	PXRE7-01|	1
2013-12-01	01:54:21-03|	TUTF7-08|	1
2013-12-01	02:05:18-03|	TUTC-LT7-01|	0

Onde, estado_equipamentos é um serviço em ANSI C desenvolvido para este trabalho.

No sistema elétrico da Eletrobrás Eletronorte que abrange os estados do Pará, Maranhão e Tocantins, o número de equipamentos elétricos em uso na época em que este documento foi descrito era de 482. O arquivo lista, em ordem alfabética, todos os equipamentos envolvidos no desenho topológico utilizado pelo programa configurador de estados do SAGE do SEP supervisionado.

Um arquivo que contém os estados desses equipamentos é obtido da mesma maneira que a lista de equipamentos existentes, realizando-se uma consulta à base de dados histórica do SAGE. Desta vez a consulta é mais elaborada e pode demorar algum tempo. Ele realiza um SQL QUERY do estado de todos os equipamentos desde o dia de início desejado, até o dia do último arquivo exportado, utilizado como amostra de dados para mineração. Uma amostra do arquivo resultante, estados_sql.txt é mostrado abaixo:

bh_dthr | id | estado
2013-10-15 03:00:00-03| CJBC6-01 | 0
2013-10-15 03:00:00-03| GMBC6-01 | 0
2013-10-15 03:00:00-03| GMBC6-02 | 1
2013-10-15 03:00:00-03| LDBC6-01/02 | 0
2013-10-15 03:00:00-03| TEBC6-02 | 1
. . . . .

Onde a primeira coluna informa a data e hora segundo o formato adotado pela base de dados histórica do SAGE, a segunda coluna informa o ID do equipamento e a terceira coluna informa 1 para o estado disponível ou ligado e 0 para o estado desligado ou indisponível.

O arquivo intermediário é obtido por uma transposição do arquivo de estados dos equipamentos. Ele transforma o arquivo para a forma:

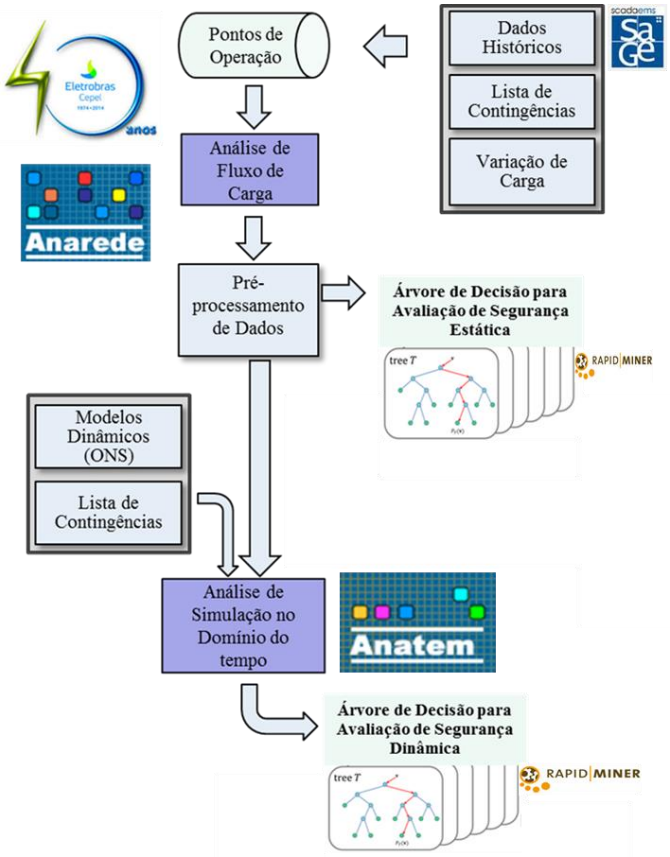
DATAS_PWF|DATAS_EQP|ABVC-LI6-01|ABVC-LI6-02|AC----500|ACCL7-02|ACCL7-|...
2013-11-07 01:48:44-03|0|0|00|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|00|0|0|0|...
2013-11-07 03:00:00-03|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|...
2013-11-07 06:08:17-|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|0|1|0|1|1|1|1|1|1|1|...
2013-11-07 06:09:44-|1|1|1|1|0|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|0|1|0|1|...
. . . . .

Cada conjunto binário, alfabeticamente ordenado diferente é utilizado como uma assinatura topológica, assim, tem-se um conjunto de topologias [A1, A2,...,An]. O arquivo resultado Final.txt é o arquivo contendo os casos Anarede, onde está incluída, uma coluna inicial contendo a assinatura topológica de cada caso exportado. Este arquivo está pronto para sofrer um tratamento de mineração de dados. Um extrato deste arquivo é apresentado a seguir:

Topologia	Data	Hora	V_119	V_118	V_117	V_116
A01	08/11/2013	00:02	1,052	1,048	1,037	1,003
A01	08/11/2013	00:07	1,053	1,049	1,037	1,004
A01	08/11/2013	00:12	1,052	1,048	1,036	1,002
A01	08/11/2013	00:17	1,052	1,048	1,036	1,002
A01	08/11/2013	00:22	1,052	1,048	1,037	1,003
A01	08/11/2013	00:27	1,052	1,048	1,037	1,002
A01	08/11/2013	00:32	1,051	1,047	1,038	1,002
A01	08/11/2013	00:37	1,053	1,049	1,039	1,002
A01	08/11/2013	00:43	1,06	1,044	1,037	1

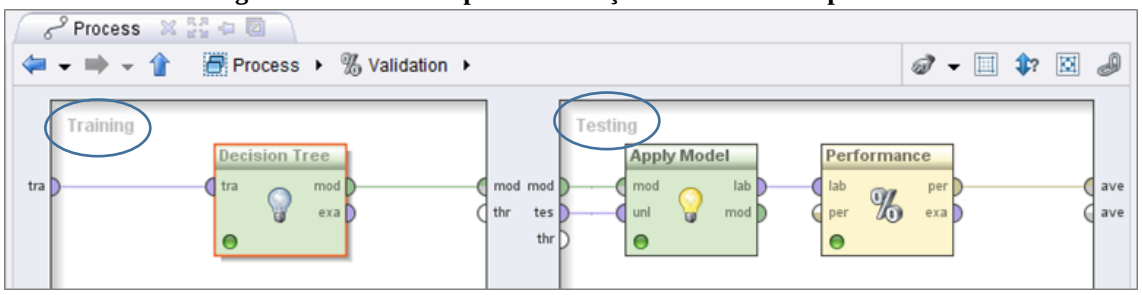
A assinatura topológica é assim, a sequência de bits, cada um contendo o estado de cada equipamentoordenados alfabéticamente. Desta forma, se pode garantir a univocidade da relação lista de equipamentos x assinatura topológica. Esta assinatura só deve ser revista quando da alteração da topologia do sistema, provavelmente pela expansão deste onde, novos equipamentos sejam adicionados aos existentes, bem como a eventual retirada de alguns equipamentos. Assim, este procedimento deve ser revisto e totalmente renovado a cada alteração do SEP supervisionado. A figura 16 apresenta os passos para a criação da base de dados que será entregue ao minerador de dados:

Figura 15 - Resumo dos Passos de Geração da DT para SDSA



De posse das bases de dados rotuladas é possível gerar uma DT para a avaliação de segurança estática e outra para avaliação dinâmica. No programa Rapidminer é montada a estrutura mostrada na figura 17:

Figura 16 - Estrutura para Mineração de dados no Rapidminer



(Autor)

A figura 17 apresenta o bloco onde é realizada a função de validação do modelo, é utilizado *XValidation*, na área à esquerda está o algoritmo de treino, que é alimentado pela base de dados. A região da direita da figura 17 contêm a função de aplicação do modelo e avaliação de performance do classificador, em função da estrutura da *XValidation*. É utilizado 10 *folds* na validação cruzada. Para os parâmetros de criação da DT é utilizado o critério de divisão *accuracy* com uma profundidade máxima de 20 folhas. O Rapidminer fornece como saída a forma gráfica da DT, o conjunto de regras de classificação derivado da mesma, os resultados de classificação. É possível plotar também algumas características estatísticas da base de dados.

4.4 Características da Metodologia para Monitoramento da Tensão Via Árvore de Decisão

De posse da base de dados criada seguindo os passos descritos é possível gerar uma DT de avaliação da segurança estática com foco no monitoramento de tensão. Existem duas características no modelo de DT resultante na proposta que a difere e melhora as características encontradas em outras abordagens com o uso de DTs. Sendo elas

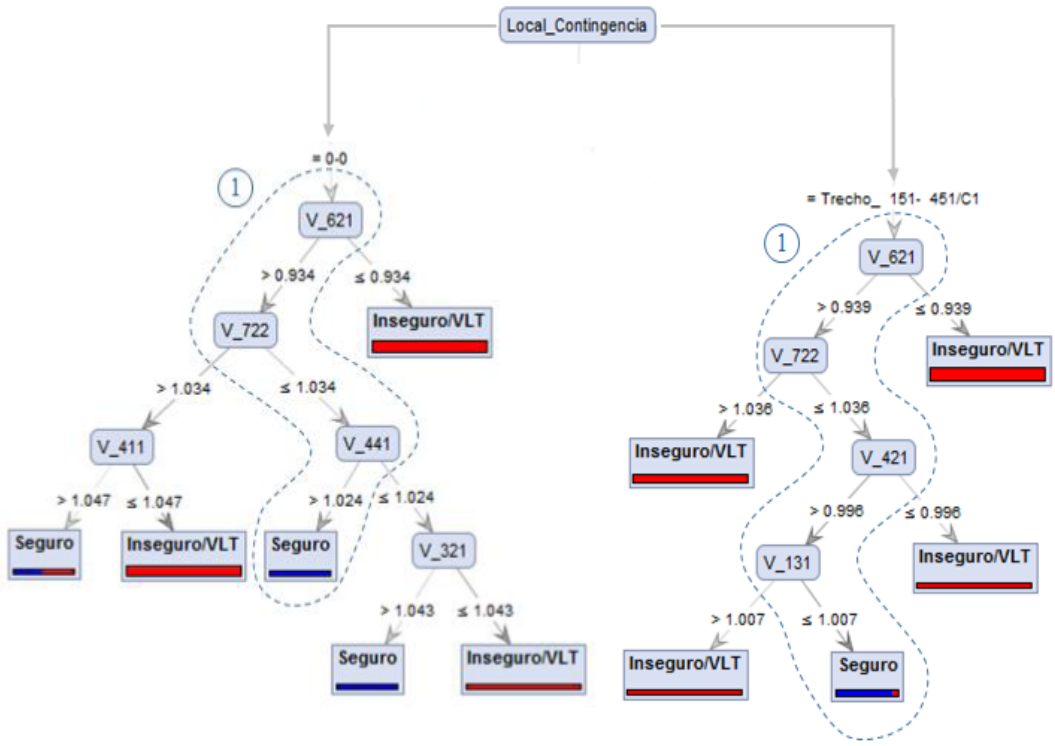
- Orientação da DT por topologia – Cada ramo da DT representa uma topologia;
- Única DT para monitoramento da tensão;

A literatura apresentada na revisão bibliográfica indica uma grande flexibilidade no uso de DTs para o problema de avaliação de segurança. Essa flexibilidade se traduz na escolha de atributos, combinações de atributos, tipos de medição, diferentes técnicas na divisão da DT, entre outras. Contudo, estes trabalhos propõem metodologias que se baseiam em DTs para monitoramento e avaliação de segurança como problemas

separados, atacando somente um dos problemas ou fazendo um uma DT para monitoramento e uma para segurança. Pelo prisma dos atributos da base de dados, o uso de topologias e contingências é pouco explorado, sendo o mais comum o uso de uma DT para cada topologia e/ou contingência.

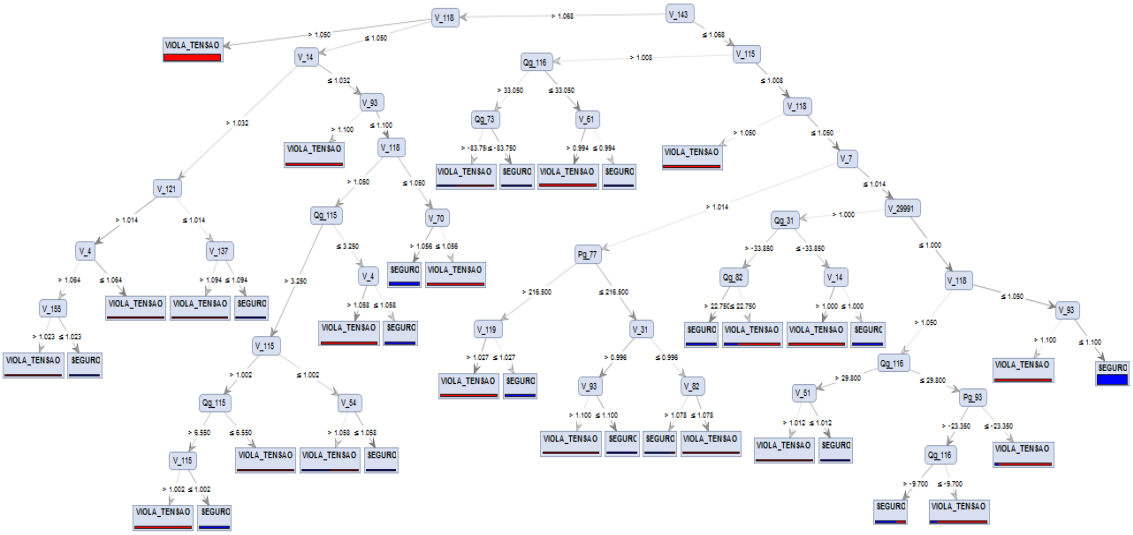
A figura 18 apresenta um trecho de uma DT treinada com as características da metodologia proposta onde o local da contingência é a variável escolhida pela DT como nó raiz, onde o ramo "=0-0" (a esquerda da figura) representa o sistema teste com todas as linhas em operação, chamado de topologia completa, e o ramo da direita representa a saída de uma linha de operação, ou seja, a topologia completa e sua respectiva contingencia n-1. Uma característica importante a se ressaltar é a profundidade da DT, isto é, com poucas variáveis a DT classifica com boa performance, gerando regras simples, de fácil intepretação e implementação, quando comparados a DT da figura 19, que usa a mesma base de dados, com a única diferença, a não inclusão do atributo topológico.

Figura17 - DT Criada pela metodologia proposta



(Autor)

Figura 18 - DT Criada de Acordo com as metodologias da literatura



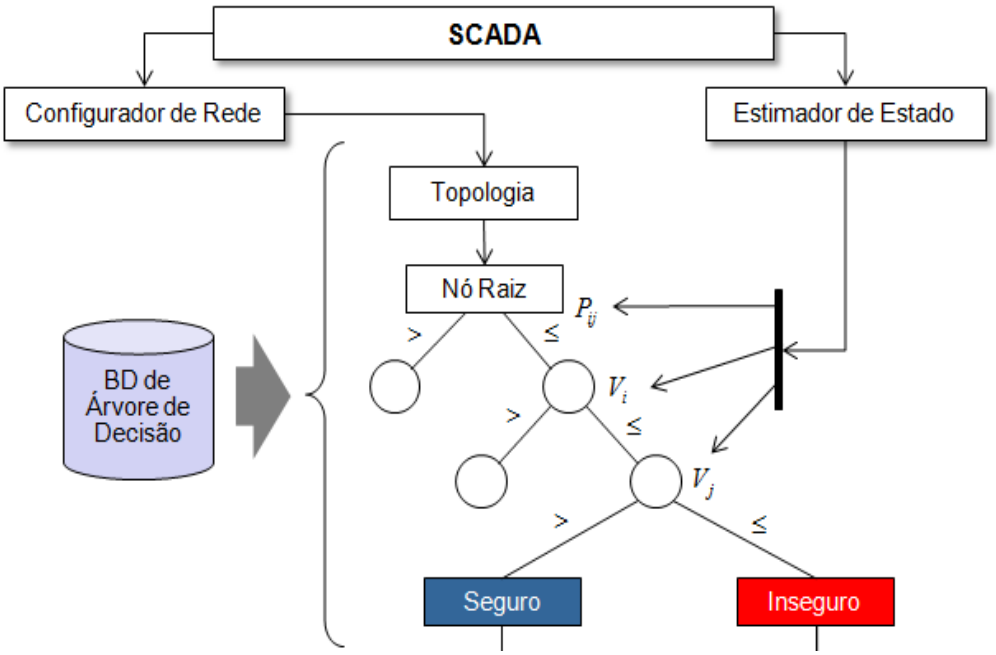
(Autor)

Pode-se levantar dúvidas em relação à classificação do modelo proposto, devido à baixa profundidade da DT. Entretanto, é importante salientar que os métodos mais utilizados na literatura (entenda-se como não orientado por topologias) não restringem a área de busca na base de dados, isto é, há uma grande quantidade de cenários disponíveis, dificultando o relacionamento das variáveis. Enquanto que com a orientação por topologia, tem-se uma diversidade de pontos de operação mais próximos entre si, assim a probabilidade de bom desempenho de classificação para casos não vistos, que estejam próximos dos dados de treinamento da DT é bastante elevada. O que é uma hipótese factível, haja vista que as topologias de operação são normalmente escolhidas de acordo com as condições de carregamento do sistema.

Outro ponto de destaque é a utilização de uma única DT para diversas topologias possíveis para operação do sistema. Pois, de acordo com os métodos utilizados na literatura, cada topologia ou contingencia teria uma árvore de decisão, o que torna o processo de implementação e utilização em centros de operação mais complexos, devido a maior quantidade de linhas de programação e testes para verificação de topologias e bases de dados de arvores de decisão. Sem contar que árvores com grandes profundidades têm maiores chances de terem ramos repetidos. Assim, a simplicidade dos caminhos de classificação para DT da figura 18 ganha destaque, pois o crescimento da DT é em número de ramos, ou seja, a cada topologia apresentada na etapa de treinamento aumenta-se um ramo da DT, o que na prática não será mostrado ao operador e do ponto de vista de

programação seria apenas mais uma condição do tipo “if” ou “case”. A figura 20 ilustra o comportamento da ferramenta baseada em DT proposta.

Figura 19 - Esquema da Estrutura de Utilização da DT em Ambiente de Operação



(Adaptado, Rodrigues 2014)

As medidas provenientes do SCADA (Configurador de Redes e Estimador de Estado) são as entradas para disparar o modelo de DT. A partir da base de dados do configurador, faz-se o processo de identificação da topologia, que será a primeira variável a ser testada (nó raiz da DT), assim uma topologia já treinada tem seu caminho de classificação ativado (conjunto de “if”) através dos dados do estimador de estado. O uso do dado estimado evita eventuais problemas por falta de medição para a variável escolhida, além de ser bastante robusto à ruídos. De acordo com os valores dos dados estimados nas variáveis, a MDT classifica como seguro ou inseguro o ponto de operação da topologia atual, sendo também possível o operador verificar a robustez e segurança do sistema para contingências típicas previamente consideradas na base de dados de treino.

No entanto, uma questão que pode ser crítica para qualquer metodologia de avaliação de segurança é quando ocorre um desligamento forçado de uma unidade geradora, linha de transmissão ou transformador, que resulta em uma configuração topológica da rede não prevista no modelo de classificação. Nesse caso, os trabalhos reportados na literatura são unanimes quanto à necessidade de atualização da árvore de decisão logo após o desligamento forçado. Pois bem, a atualização do modelo deve ter seu custo computacional considerado, pois dependendo do problema e da geração da base

de dados (avaliação estática ou avaliação dinâmica), do modelo a ser treinado (RNA, DT, SVM) e forma de treino (crossvalidation. holdout, boostraping) o tempo para a geração desse novo modelo pode ser crítico. Contudo, o modelo proposto já abrange uma maior quantidade nas possíveis topologias, com a estrutura da ferramenta averiguando a situação topológica em tempo real, estratégias de atualização podem ser desenvolvidas, a fim de mitigar o problema da atualização.

Na metodologia proposta, um fator interessante para a mitigação no problema de atualização do modelo, é que podem ser treinadas DTs somente para as novas topologias, integrando-as à DT já utilizada sem alterá-la nos outros ramos. Isso significa em outras palavras, que uma atualização em um ramo da DT, mesmo que esse ramo seja na realidade uma nova DT exclusiva para aquela nova topologia não necessita do treinamento de uma DT com uma grande base de dados, isto é, as outras topologias não precisam ser treinadas novamente.

Supondo, por exemplo, que a base de dados original possua 5 topologias, mais 2 contingencias n-1 para cada topologia, totalizando 15 possíveis topologias de operação com 200 simulações para cada, ter-se-ia 3000 pontos de operações, ou seja, 3000 registros na base de dados. Nas abordagens utilizadas na literatura, tem-se 2 possibilidades, caso ocorra o surgimento de uma topologia não prevista:

1º Possibilidade - Uma nova base de dados, com 200 casos seria criada e uma nova DT seria treinada, assim passariam a ter 16 árvores de decisão (16 topologias).

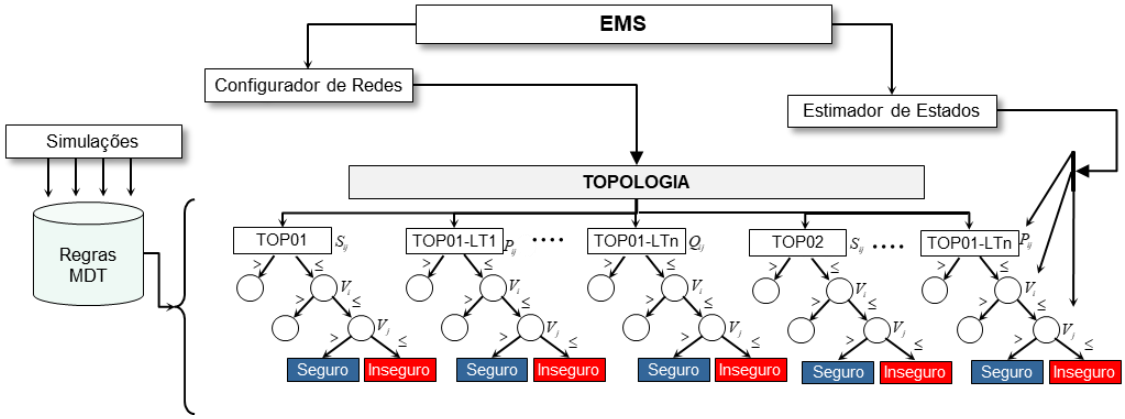
2º Possibilidade – Acrescentar-se-iam mais casos, 200 novos casos, na base original, que ficaria então com 3200 registros. Em seguida a DT seria treinada, alterando todo o código de regras de classificação, o que levaria a uma grande modificação no programa para avaliação de segurança.

Essas estratégias para as abordagens padronizadas, tornam o custo computacional elevado em todas as etapas do processo para a geração de uma ferramenta de SDSA *online*. A figura 21 ilustra uma possível forma de funcionamento de uma ferramenta baseada na metodologia proposta. Como o nó raiz da DT é a topologia (chave topológica), cada possível configuração e contingência prevista. Assim a ferramenta ficaria apresentando ao operador a situação da tensão e robustez frente a instabilidade transitória no modo online para o ponto de operação atual, podendo também dar acesso à verificação da robustez para dadas contingências. Caso haja a necessidade de atualizar algum ramo

da DT, entraria apenas mais um conjunto de “IFs” a partir do nó raiz, não alterando nenhum outro ramo sem necessidade, diminuindo assim o custo computacional.

Essas características mostram o quanto a metodologia pode contribuir no aumento da segurança dos SEPs já que fornece ao operador diversas informações sobre o estado atual do sistemautilizando poucas variáveis, o que diminui a fadiga do operador em momentos de operação normal, além de auxiliar com sugestões de operação para casos de contingência.

Figura 20 - Esquemático de uso da DT com múltiplas topologias



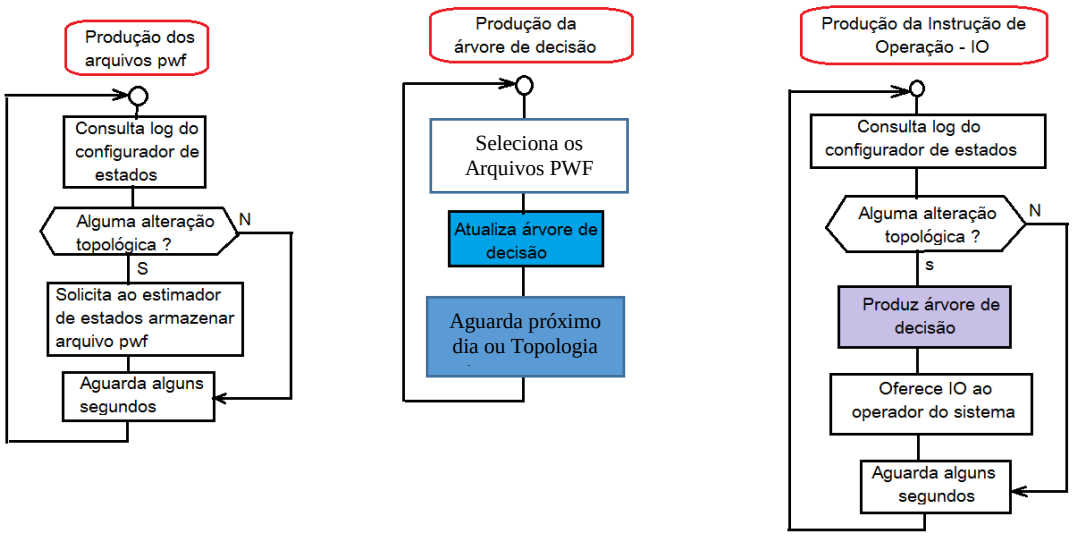
(Autor)

4.5 Avaliação de Segurança *Online* via DT

Com o monitoramento das variáveis críticas para cada topologia é possível manter o sistema dentro de uma faixa segura de operação, tanto em relação aos níveis de tensão nas barras, quanto frente a perturbações que poderiam levar a problemas de instabilidade transitória. Toda a questão de criação da base de dados, avaliação do modelo de DT gerado e atualizações podem ficar invisíveis para os operadores, que estariam monitorando o sistema sem se preocupar com esses detalhes da ferramenta. A figura 22 ilustra de maneira resumida o conjunto de passos que ficam ocultos ao operador.

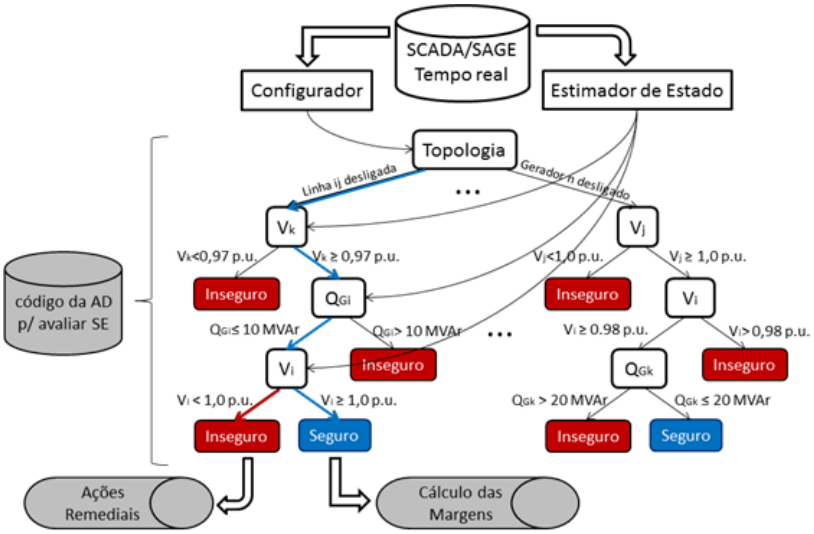
Para o ambiente de operação, os resultados da DT podem ser apresentados basicamente de duas formas: Com a forma gráfica da DT ou apenas como instrução de operação, ou ainda como combinação de ambas. A figura 23 ilustra o esquema do módulo de avaliação da segurança estática online. A partir dos dados adquiridos do configurador de redes e do estimador de estados, ambos integrados ao sistema SCADA/SAGE, do centro de controle do sistema, o modelo da árvore de decisão classifica de maneira rápida se o sistema é *seguro* ou *inseguro* do ponto de vista estático.

Figura 21 - Resumo da Etapa Oculta ao Operador



(Autor)

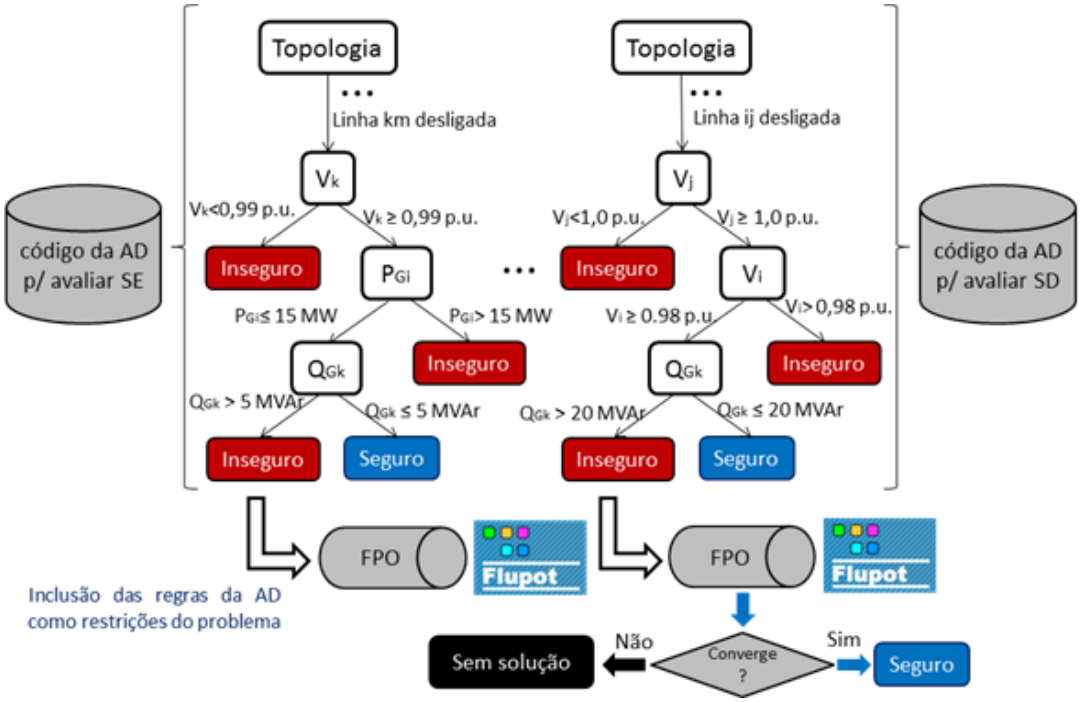
Figura 22 - Esquema do Módulo de Ações de Controle Preventivo "*On-line*" Baseado em Árvore de Decisão



(Bezerra, 2017)

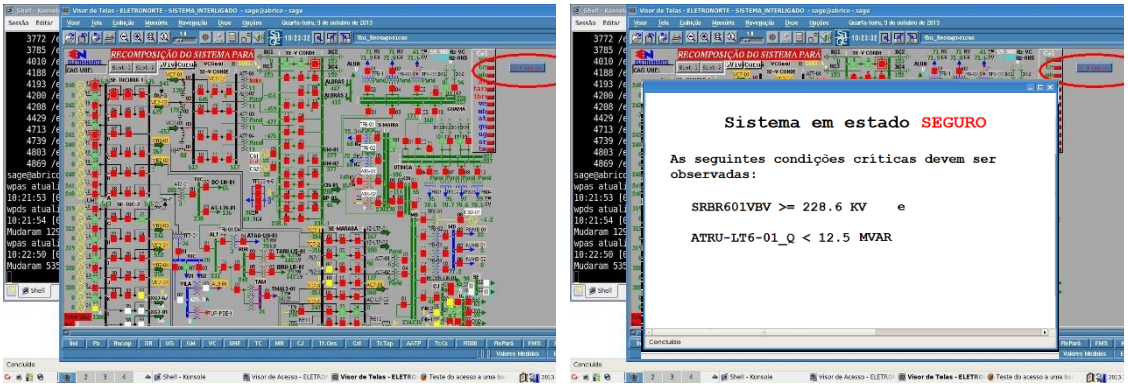
Após a classificação, realizada por este processo, caso o sistema esteja no estado *seguro*, pode-se implementar e realizar o cálculo das margens de segurança do sistema para o ponto de operação em tempo real, indicando a robustez do sistema. Caso seja classificado como inseguro, pode-se usar os limites das variáveis para executar ajustes no sistema ou executar um fluxo de carga ótimo, adicionando os limites das variáveis da DT para dada topologia, como indicado na figura 24.

Figura 23 - Esquema do Módulo de Ações de Controle Preventivo "*Off-line*" Baseado em Árvore de Decisão



(Bezerra, 2017)

Figura 24 - Tela de Observação do Sistema (SAGE)



(Bezerra, 2017)

A figura 24 ilustra como pode ser utilizado um modulo de ações preventivas, onde as regras da DT alimentam um fluxo de carga ótimo (Flupot do Cepel). O modo de apresentação como instrução de operação no SAGE pode ser da seguinte maneira: Na tela principal do operador, há apenas um pequeno botão contendo as instruções de operação, destaque na figura 25 à esquerda. Quando acionado, este acionaria uma tela de pop-up como apresentado na figura 25 à direita, indicando o estado de operação (seguro/inseguro) e quais as variáveis e seus respectivos limites que devem ser monitoradas com mais cautela, para manter o sistema seguro.

4.6 Conclusões

A abordagem proposta para ser a base de uma ferramenta de SDSA mostra características que aproveitam informações do SAGE e do histórico operacional de maneira sistemática. Essa sistematização parte da geração da base de dados, com o aproveitamento do passado recente, pontos de operação e topologias para a uma melhor representação dos cenários operativos.

A melhor visualização e representatividade da DT orientada por topologia, gerando regras menores e mais simples que DTs treinadas nos métodos abordados na literatura, facilitando até mesmo processos para atualização. Além disso, o uso de uma única DT para monitoramento online e avaliação de segurança para contingências (n-1) reduz o esforço computacional e o estresse do operador em ambiente de trabalho.

Capítulo 5 - Resultados

Neste capítulo são apresentados estudos de casos para dois sistemas testes. Inicialmente é avaliada a metodologia prosposta usando o sistema teste de 39 barras do IEEE para fins de avaliação da segurança do perfil de tensão e da estabilidade transitória. Posteriormente, a metodologia é avaliada com dados reais de um estimador de estado do subsistema norte da Eletrobrás Eletronorte. Nos estudos com o sistema da Eletronorte, são avaliados a segurança de tensão e estabilidade transitória.

O processo de criação das bases de dados descrito no item 4.2 é reliazado e testado para o sistema de 39 barras e para o sistema Eletronorte. A MDT é testada para a metodologia de criação de base de dados do item 4.3, na qual são aplicados os dados históricos do sistema Eletronorte.

5.1 Sistema Elétrico IEEE 39 Barras (New England)

Para testar e validar a metodologia de avaliação de segurança de SEPs com base em MDT utilizou-se o sistema teste IEEE 39 barras (figura 25), que é um sistema benchmark do IEEE para estudos de estabilidade e controle de SEPs, disponível para download no site do IEEE. Foram treinadas duas MDTs, a primeira para avaliar o perfil de tensão nas barras do sistema. A segunda MDT foi treinada com foco na estabilidade transitória. Estas MDTs são comparadas com: a DT do tipo CART, um ensemble DT e um Random Forest. São avaliados os percentuais de acerto de treino e teste, o número de nós de decisão, o tempo de criação dos modelos, bem como a clareza e confiabilidade das regras de decisão geradas.

A criação da base de dados seguiu os passos apresentados no item 4.2 desse texto. Após a amostragem da base dados, extraiu-se 30000 casos. Seguindo a regra de manter a distribuição de casos seguros e inseguros, bem como, o número de casos por topologias iguais. Em seguida estes 30000 casos foram separados em duas bases de dados. A primeira para treino e teste com 25000 casos e a segunda para validação com 5000 casos.

A figura 26 apresenta algumas características da base de dados, os histogramas de potência gerada e média tensões nas barras com controle de garação.

Figura 25 - Diagrama Unifilar do Sistema teste IEEE 39 barras

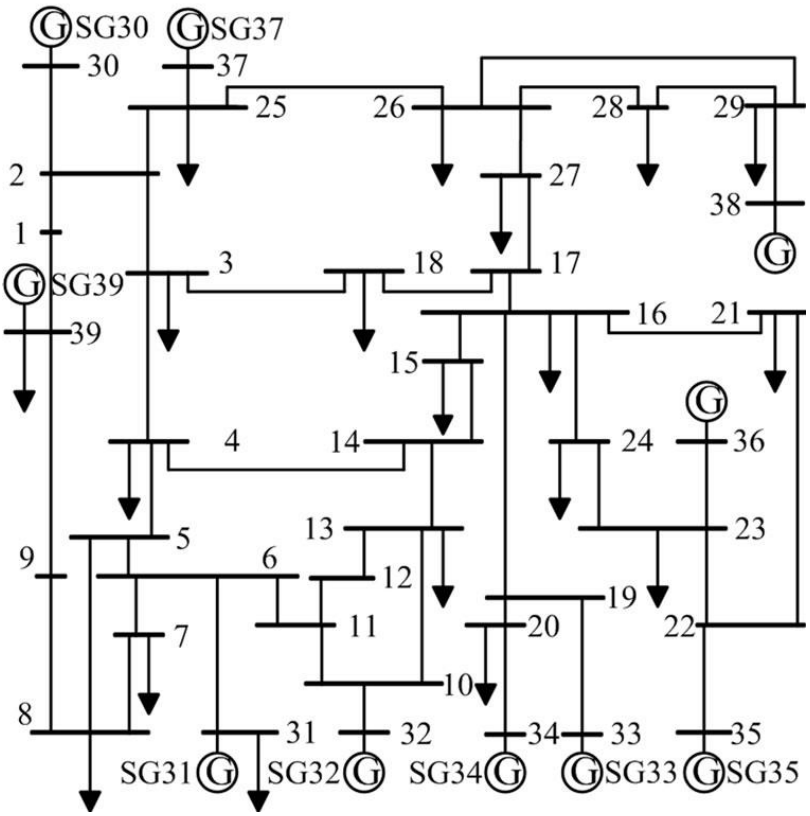
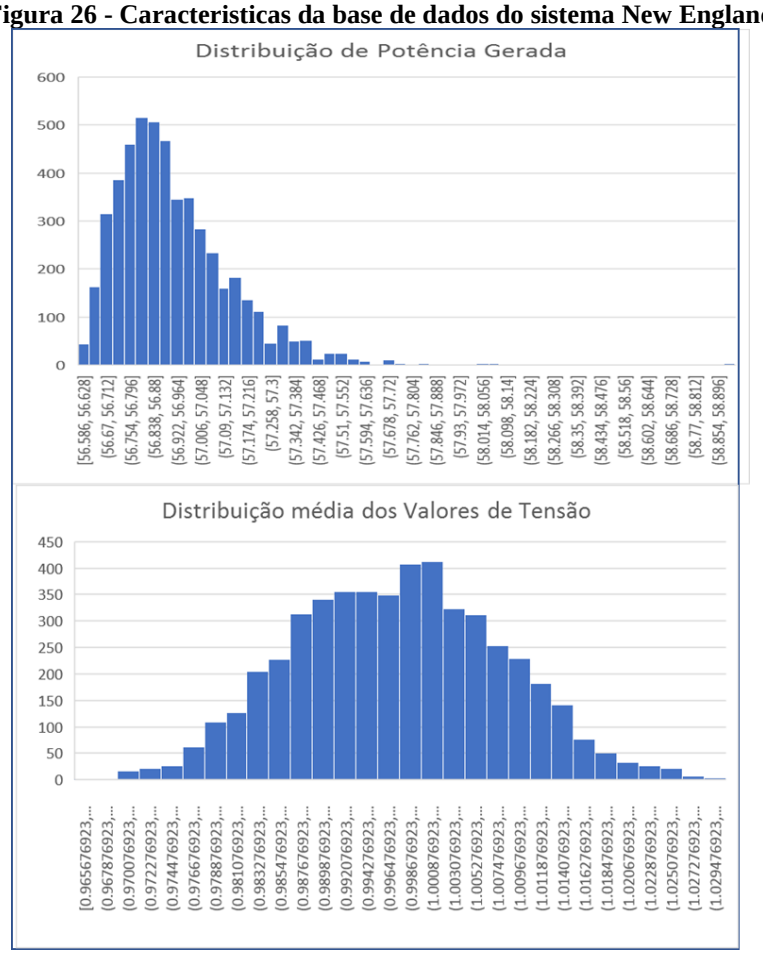


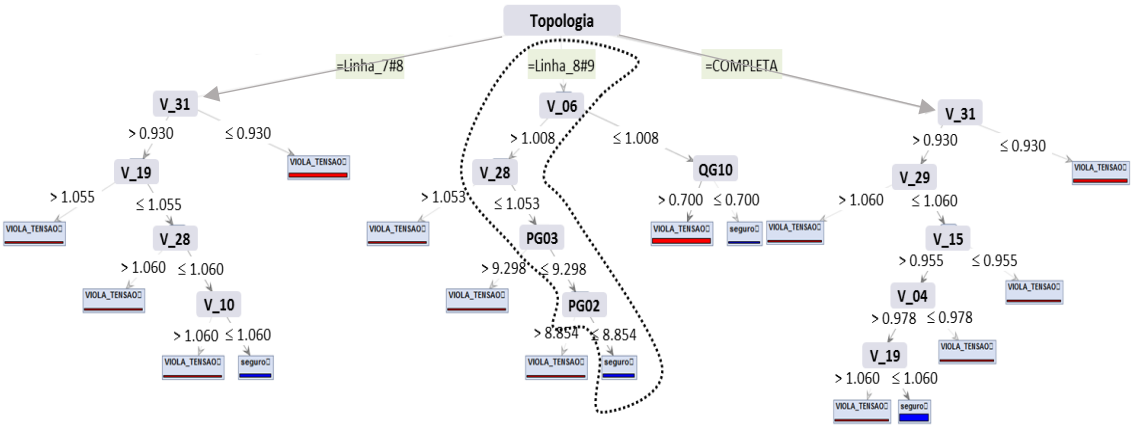
Figura 26 - Caracteristicas da base de dados do sistema New England



5.1.1 Avaliação Estática Via MDT

A MDT treinada para a avaliação estática é apresentada na figura 27. Note que é possível identificar com clareza as regras para manutenção da segurança do perfil de tensão do sistema, para a topologia completa do sistema e para 2 contingências de saída de linha (linha que conecta as barras 7 e 8, linha que conecta as barras 8 e 9).

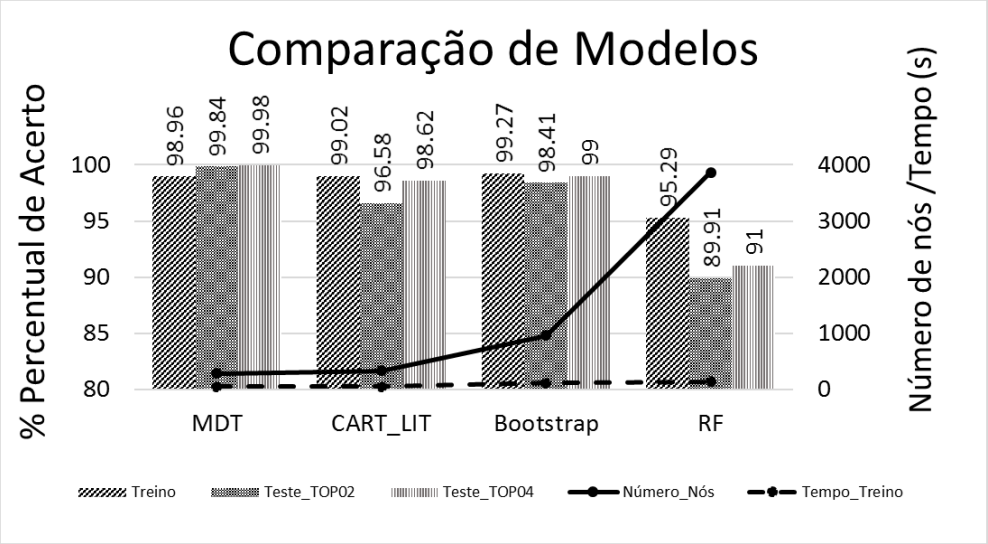
Figura 27 - MDT para avaliação estática New England



Como exemplo da saída da linha que conecta as barras 8 e 9, mantendo os limites das variáveis apresentadas na regra dentro dos limites que a MDT determina, é possível manter o sistema sem violações de tensão. Para tanto deve-se manter a tensão na barra 8 maior que 1,008 p.u. e a tensão na barra 28 menor ou igual a 1,053 p.u. e a geração na barra 3 menor ou igual a 9,298 p.u. e a geração na barra 2 menor ou igual a 8,854 p.u.

A figura 28 apresenta o comparativo entre os quatro modelos de avaliação treinados. É possível notar a elevada taxa de classificação dos 4 modelos nas etapas de treino e teste. Contudo é notável a grande diferença de tempo computacional gasto para a criação de cada um dos modelos. Outra variável importante é o numero de nós de decisão, uma vez que isso implica diretamente no número de variáveis elétricas a serem monitoradas por operadores.

Figura 28 - Comparação dos Modelos para avaliação estática New England



Mesmo levando em consideração que qualquer algoritmo de árvore de decisão possui a característica de redutor de dimensionalidade. Muitas vezes, algumas características desses algoritmos fazem agrupamentos de atributos categóricos para reduzir ainda mais a quantidade de variáveis que "aparecem" no modelo criado. O CART, por exemplo, sempre faz agrupamentos de modo que suas divisões sejam sempre binárias, o que implica em alguns casos no agrupamento de atributos com pouco ou nenhum sentido fisico. No caso de topologias de sistemas elétricos, esses agrupamentos podem muitas vezes ser danosos nas análises que serão feitas por operadores e engenheiros.

O modelo baseado em Random Forest e Boosting são ensembles, isto é, são conjuntos de classificadores que tem o intuito de aumentar o poder de classificação dos modelos. Contudo, esse aumento do poder de classificação acaba levando a um aumento da complexidade do modelo, isto é, a facilidade de visualização das regras da DT acaba sendo bastante reduzida ou mesmo sendo perdida. Isso é evidenciado pelo número de nós de decisão que estes modelos acabam gerando.

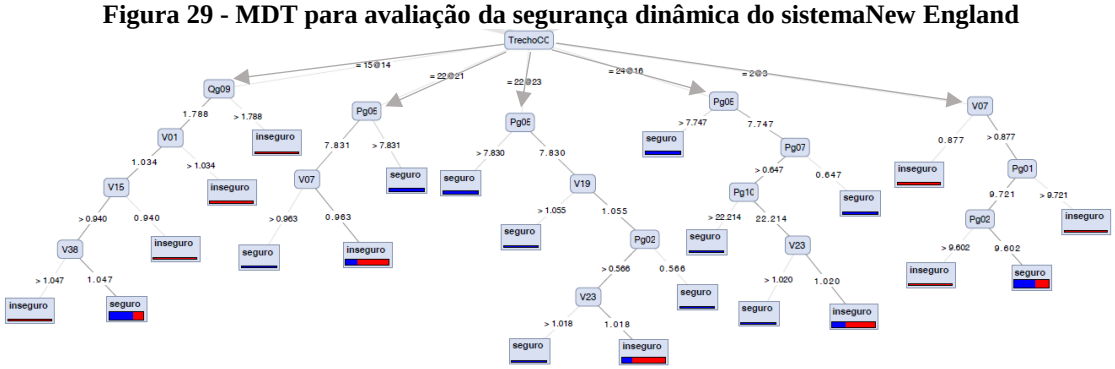
Dessa forma, a avaliação de segurança do perfil de tensão via MDT, bem como as DTs tipo CART e essemble conseguem grandes taxas de acerto de classificação. Entretanto, a MDT possui características que lhe são benéficas, vista que cada ramo da MDT apresenta um conjuto de regras claras para cada topologia. Permitindo que o operador disponha não apenas da classificação do estado operativo do sistema *online,* como também permite que seja avaliado um conjunto de contingências pré-definidas com alta taxa de acerto e quais variáveis e seus respectivos limites para manter o estado de segurança do sistema. Além disso, o conjunto de regras podem ser facilmente adotadas para tomada de ações preventivas, devido a sua clareza.

5.1.2 Avaliação Dinâmica via MDT – Estabilidade Transitória

A avaliação de segurança dinâmica de SEP via inteligência computacional é de bastante interesse. Uma vez que em tempo real muitas vezes é muito difícil tomar ações rápidas o suficiente manter o sistema seguro após uma grande perturbação. Dessa forma, uma maneira que vem sendo trabalhada é manter o ponto de operação em condição pré-falta dentro de faixa de operação que seja robusta a uma lista de contingências possíveis.

Os métodos de avaliação de segurança dinâmica via DTs se baseiam nesse princípio. Contudo, executa-se uma quantidade maior de simulações e treinam-se os modelos de DTs, que conseguem extrair características que podem classificar *online* o ponto de operação para uma lista de contingências, mesmo antes da ocorrência da perturbação. Em vista disso, a MDT, como apresentado na avaliação estática, possui características interessantes para a aplicação de avaliação de segurança, pois apresenta elevadas taxas de acerto na classificação, clareza nas regras e baixo número de variáveis a serem monitoradas.

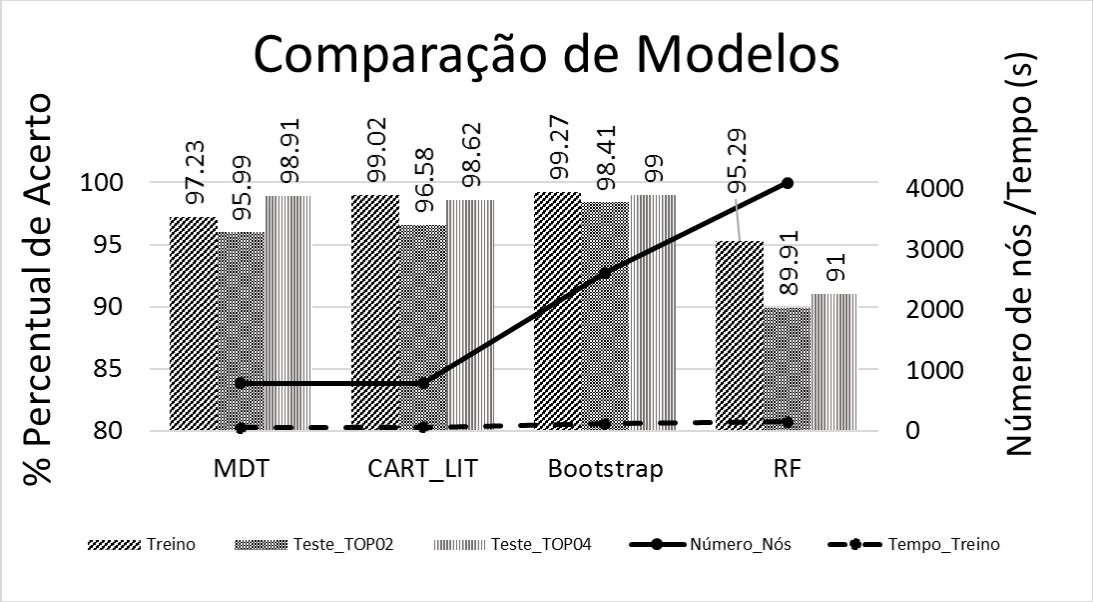
Figura 29 - MDT para avaliação da segurança dinâmica do sistemaNew England



A figura 29 apresenta uma MDT para 5 contingências de curto circuito trifásico de 100 ms próximos as barras específicas do sistema, e a consequente retirada de linha, emulando assim a ação da proteção. Nota-se que para o problema de segurança dinâmica, o número de variáveis a serem monitoradas são maiores que o caso de segurança estática, mesmo sendo essecialmente utilizada a mesma base de dados para treino e teste nos dois casos.

Quando comparado ao treino e teste dos outros modelos, é possível observar que a MDT apresenta taxa de acerto similar aos demais. Porém mantém as características de baixo numero de nós de decisão e clareza nas regras. A figura 30 apresenta o comparativo dos resultados de classificação e características dos modelos treinados.

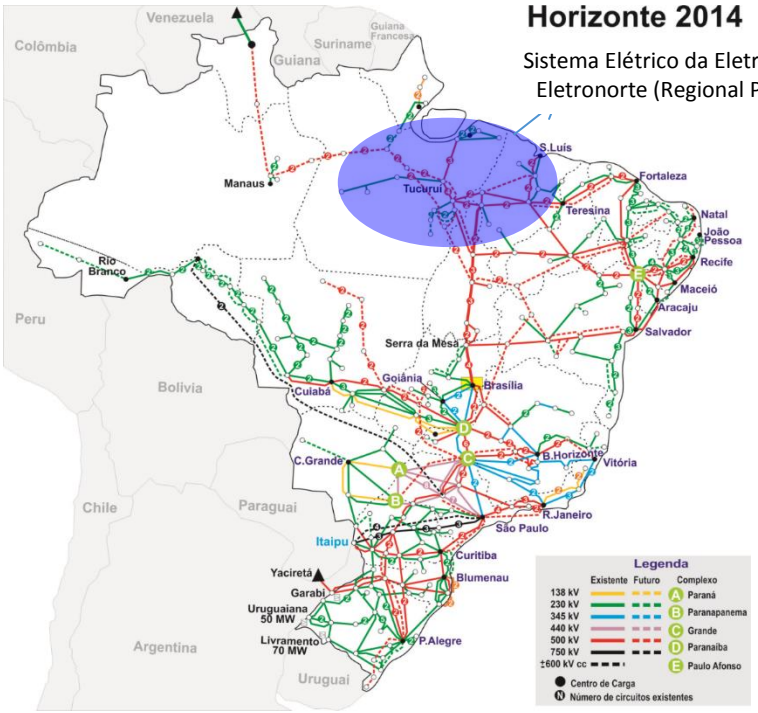
Figura 30 - Comparação dos Modelos para avaliação Dinâmica New England



5.2 Sistema Elétrico da Eletrobras Eletronorte.

Outro teste para validação da metodologia proposta utilizou informações da base de dados do histórico operacional do Sistema Elétrico da Eletrobrás Eletronorte (regional Pará), ilustrado em azul na Figura 31.

Figura 31 - Sistema Elétrico da Eletrobrás Eletronorte (Regional Pará) localizado na área azul



A geração na regional Pará é constituída pelas usinas hidrelétricas Tucuruí e Curuá-Una, sendo que a usina de Tucuruí pode fornecer até 8370 MW, distribuídas entre 12 unidades geradoras de 350 MW, 11 unidades de 375 MW e 2 unidades de geração

auxiliar de 22,5 MW. A usina de Curuá-Una possui 2 unidades de geração de 2 MW e uma de 10,3 MW, totalizando 30,3 MW. Essa energia é escoada por mais de 2539 Km de linhas de transmissão nas tensões de 69 kV, 138 kV, 230 kV e 500 kV. Operando em 12 subestações, alimentando Grande Belém, o Baixo Tocantins e as regiões nordeste, sudeste, leste e oeste do Estado. A rede está conectada ao Sistema Interligado Nacional - SIN e eventualmente é suprida pelos subsistemas Nordeste e Sudeste/Centro-Oeste.

Por meio de um projeto de P&D, teve-se acesso aos dados do histórico operacional, de um período de operação. Esses dados são adquiridos do sistema SAGE-SEM. Em seguida são executados os passos de seleção e tratamentos de dados (etapas de pré-processamento na mineração de dados).

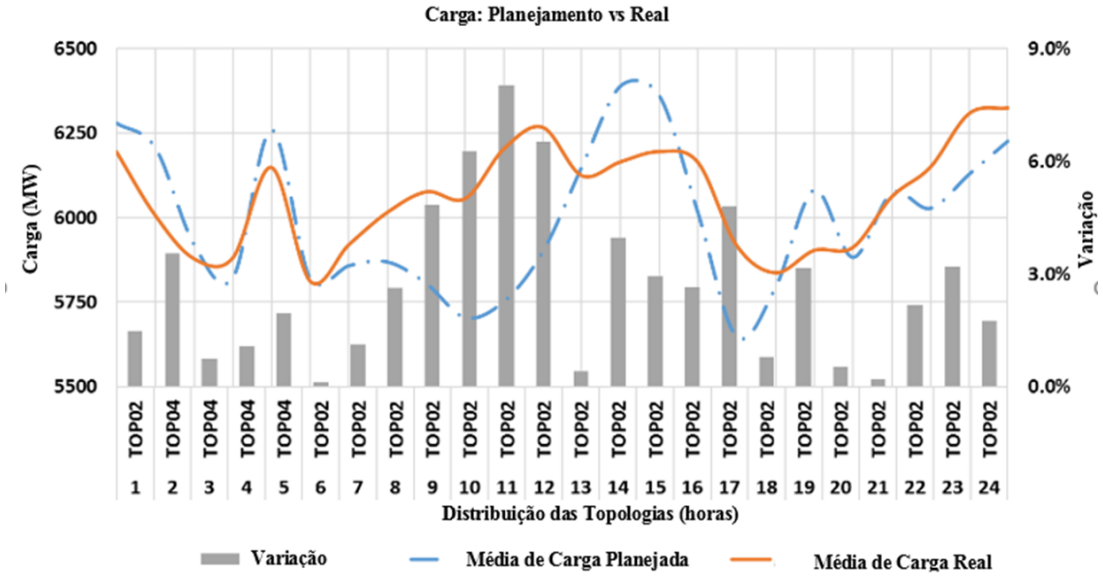
Foram selecionados alguns dias para o teste da metodologia baseada em MDT. Esses dados foram escolhidos dentre o período que se teve acesso aos dados por conter os históricos do estimador de estados, no formato PWF para execução no ANAREDE, e os dados do configurador de redes.

Foram aplicados no sistema Eletronorte duas técnicas para a criação da base de dados de treino para as MDT. Uma baseada no item 4.2, com dados de planejamento e uma segunda técnica, na qual se aproveita o histórico operacional recente.

5.2.1 Monitoramento da Tensão Operativa

A partir dos dados de previsão de carga de 24 horas a frente, e as respectivas topologias de operação previstas, e considerando a rede elétrica definida pelo sistema EMS, 24 pontos de operação base foram criados. A partir dessas cargas e topologias definidas, novos pontos de operação foram gerados variando as cargas em +/- 1%, +/- 3% e +/- 5% em relação às condições de carregamento informadas nos 24 casos base, considerando uma distribuição gaussiana para alocar a variação de carga em cada barra de carga. Considerando esses novos cenários de carregamento, foram executados 3.500 casos de fluxo de carga para cada uma das topologias planejadas, bem como suas respectivas contingências (n-1), sendo escolhidas 5 contingências para cada topologia.

Figura 32 - Perfil de carga dia teste - Planejamento vs carga real



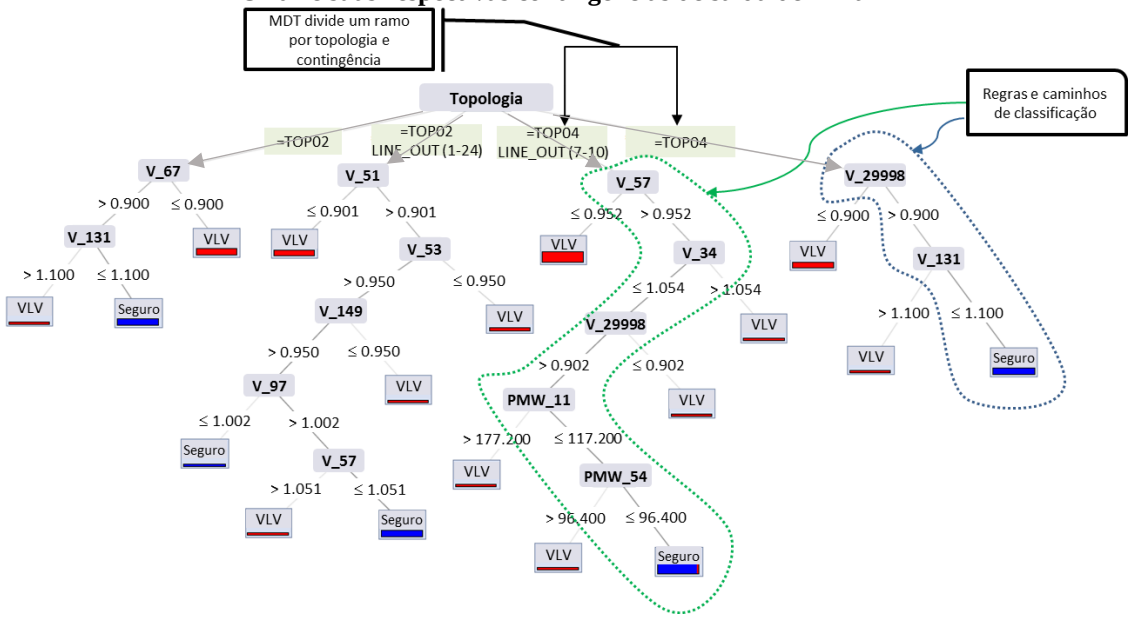
Em seguida, realizou-se um passo de pré-processamento para filtrar os casos não convergentes. A Figura 32 apresenta no eixo x, as curvas de carga de operação planejada e real que são expressas por valores médios diários, e também as topologias de operação. A diferença entre TOP02 (carga pesada) e TOP04 (carga leve) é o desligamento de dois geradores e seus respectivos transformadores. A variação da carga real em relação aos valores máximos e mínimos da carga prevista é representada no eixo y. Observe que a variação máxima é de aproximadamente 8,5%, mas próxima do valor máximo de + / -5% de variação.

Após o pré-processamento dos relatórios, a amostragem estratificada é usada para selecionar os casos que pertencerão ao banco de dados de treinamento. Nessa tese foi considerado que o número de casos selecionados para cada topologia é de aproximadamente 1.250 casos, sendo 625 classificados como seguros e os restantes como inseguros. Os dados de operação para testar a metodologia foram obtidos a partir do estimador do estado correspondente a 288 pontos de operação calculados a cada intervalo de 5 min, durante um ciclo de operação de 24 horas. Em seguida, em cada um desses pontos de operação é aplicada as contingências pré-selecionadas e os resultados do fluxo de carga são rotulados como seguros e inseguros para fazer o teste das regras de MDT.

A MDT criada contém 12 ramos, sendo 2 ramos para cada topologia de operação básica e suas 5 topologias de contingências. A Figura 33 apresenta parte da MDT criada com suas regras, mostrando os ramos para as topologias TOP02 e TOP04 e 2 saidas de linha representando uma contingência para cada topologia, respectivamente. As regras

para a topologia TOP04 e uma contingência envolvendo a desconexão de uma linha de transmissão de 138 kV que fornece uma grande região de extração mineral são destacados.

Figura 33 - MDT de segurança estática para o sistema Eletronorte para as topologias TOP02 e TOP04 e suas respectivas contingências de saida de linha



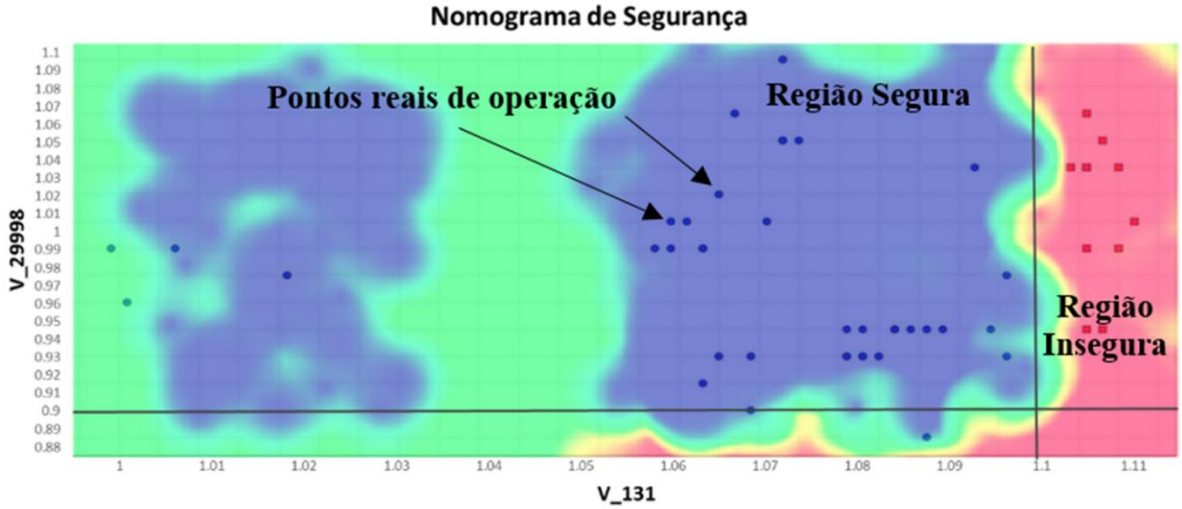
Tomando por exemplo a topologia TOP04 em destaque na figura 33, a MDT indica as variáveis tensão nas barras V_29998 e V_131 e seus respectivos limites. Neste caso, o sistema opera com uma topologia planejada, em tese, opera dentro de uma região de segurança quanto ao perfil de tensão, isso dá suporte ao pequeno número de nós de decisão que aparecem nesse ramo da árvore. Enquanto o ramo com a saída de linha 7_10, mais variáveis são apresentadas para monitoramento do perfil de tensão.

O desempenho da classificação MDT atingiu uma precisão total na etapa de treinamento de 99,4%, considerando todas as topologias. Os testes realizados com dados reais de operação são apresentados divididos por topologias e suas contingências, ou seja, o teste TOP02 contém todas as topologias relacionadas com seu homônimo. Desta forma, obteve-se precisão de 98,56% para TOP02 e 100% para TOP04. Esses resultados são obtidos pelo MDT criado com dados de previsão 24 horas, ou seja, sem atualizações com base em horizonte de previsão de 1 hora. Isso mostra a eficácia da MDT e a qualidade do processo de criação de banco de dados para avaliar o status de segurança do sistema de energia.

A Figura 34 mostra o nomograma 2-D para a topologia TOP04. A área azul é a região segura encontrada pelo MDT enquanto a região em vermelho é considerada a região insegura. Na figura 34 os pontos são os valores de tensão representados nos

barramentos 29998 (eixo y) e 131 (eixo dos x), respectivamente, que ocorreram durante o período TOP04 estava em operação. Observa-se que a maioria dos pontos de operação estão na região segura, mas o sistema também operou com violações momentâneas da tensão. O período em que o sistema operou dentro da região insegura foi principalmente na 5ª hora mostrada na curva de carga da Figura 32.

Figura 34 - Nomograma 2-D representando a região de segurança para a topologia TOP04



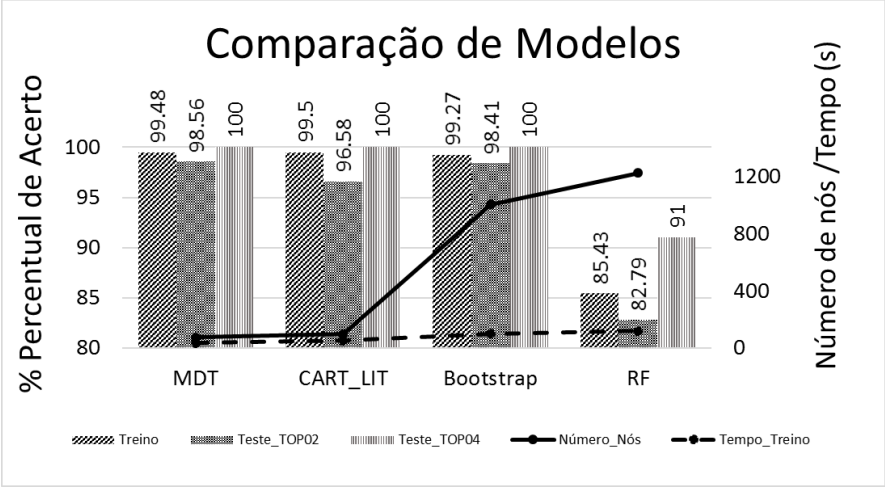
A Figura 34 apresenta um nomograma gerado com base nas regras da MDT. Ajustando as variáveis especificadas nos ramos MDT é possível mover os pontos de operação localizados na região insegura para a região segura. Ações como esta são incluídas em um fluxo de carga ótimo. No entanto, em aplicações em tempo real, muitas vezes não há tempo para executar um algoritmo de fluxo de carga ótimo para obter a solução restrita para mover o estado da operação do sistema para uma região segura. Usando MDT, essa informação já está disponível para cada topologia.

Dois pontos devem ser abordados em relação ao MDT da Figura 33. Primeiro: Sua capacidade de classificar os estados da operação de segurança do sistema, com base em um pequeno conjunto de variáveis críticas, nomeadamente valores de tensão rms e injeções de potência ativa em algumas barras; Segundo: as regras de MDT podem ser usadas para apoiar a tomada de decisões e implementação de ações de controle preventivo.

Em geral, as DTs analisadas apresentaram resultados satisfatórios em relação à classificação de segurança do sistema, independentemente do tipo deles, MDT, CART, RF ou qualquer outro conjunto de DT, conforme mostrado na Figura 35. No entanto, alguns recursos como: sensibilidade a variações em base de dados, esforço computacional para criação e execução, tamanhos e interpretabilidade dos modelos criados, são

parâmetros que devem ser avaliados. Do ponto de vista da sensibilidade às variações, a MDT tem a vantagem de impor um ramo para cada topologia no banco de dados (atributo categórico), o que garante que não haverá ramos distintos (muitas vezes indesejados) para a mesma topologia que pode ocorrer no CART. A possibilidade de que uma topologia seja incluída em mais de um ramo, pode reduzir o interesse e a confiança dos operadores no uso de DTs. Os algoritmos dos ensembles DTs tentam reduzir o problema da sensibilidade. No entanto, o esforço computacional é aumentado consideravelmente. Para o mesmo banco de dados, o tempo gasto para criar os modelos apresentados na Figura 35 foi de 25 a 35 segundos para MDTs; 25-40 segundos para CARTs sem considerar topologia; 40-50 segundos para CARRINHOS binários; 90-120 segundos para RF; E aproximadamente 100 segundos para o Bootstrap.

Figura 35 - Comparação dos Modelos para avaliação estática de tensão para o sistema Eletronorte



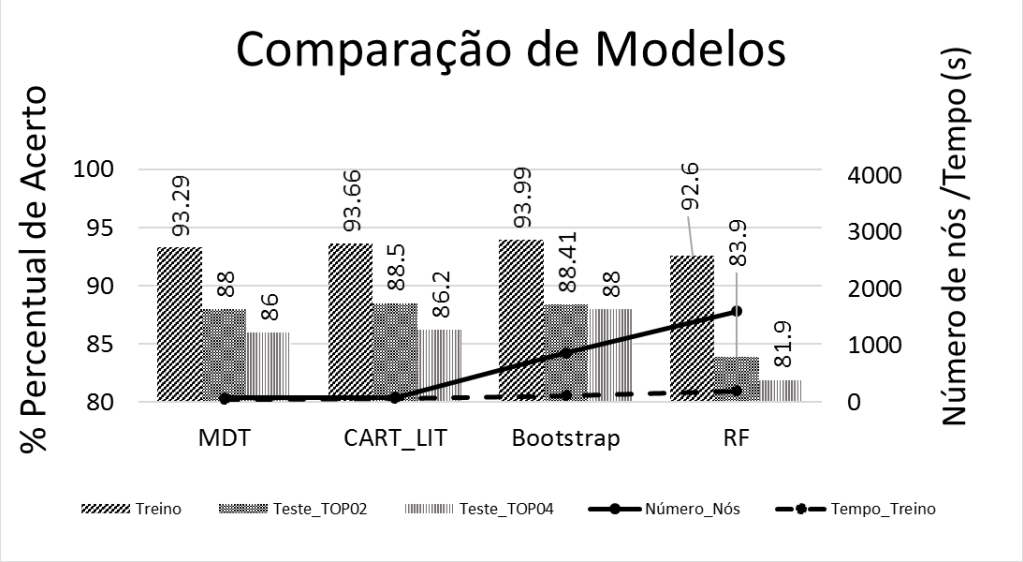
A RF e outros conjuntos são penalizados pelo alto número de nós de decisão. A MDT tem 73 nós, o conjunto bootstrap com 10 DTs tem de 83 a 106 nós para cada DT, o CART possui 97 nós e RF 1222 nós de decisão. As regras da MDT são bastante claras, indicando variáveis e seus respectivos limites para cada contingência e topologia, tendo o número de nós de decisão amplamente reduzidos. Desta forma, as regras determinadas pela MDT na Figura 33 podem ser usadas para avaliar a segurança do sistema e serem facilmente usadas para assistência de tomada de decisão.

A metodologia adotada para criar banco de dados para treinamento dos modelos DTs, que visava equilibrar o número de casos seguros e inseguros, bem como a inserção de incertezas em relação ao ponto de operação, provou ser muito eficiente. Como conseqüência, o treinamento de todos os modelos de DTs apresentou bons resultados tanto na fase de treinamento quanto na fase de validação de modelos com dados reais. Desta forma, a Figura 35 mostra não apenas a eficácia da MDT na classificação do status

de segurança, mas também permite visualizar indiretamente a eficiência da metodologia adotada de criação de banco de dados, uma vez que os modelos testados apresentaram alta precisão na taxa de classificação.

Como teste, treinou-se também uma MDT para monitorar o fluxo de reativo nas linhas de transmissão. A figura 36 mostra os resultados de classificação da MDT para o monitoramento de fluxo.

Figura 36 - Comparação dos Modelos para avaliação estática de fluxo de potência nas linhas para o sistema Eletronorte



É possível notar uma redução na taxa de acerto da MDT, isso ocorre devido ao fato da base de dados ter sido balanceada com o foco no monitoramento de tensão. Dessa forma o equilíbrio de casos seguros e inseguros do ponto de vista de monitoramento de fluxo não foi alcançado. Ficando com aproxidamante 65% de casos seguros e 35% de casos inseguros, isso imapacta diretamente na performance de treino dos classificadores. Este resultado mostra a necessidade do baleanceamento realizado na base de dados para o caso de avaliação de tensão. Desta forma, para melhorar o resultado da MDT basta fazer a amostragem da base de dados de treino de maneira semelhante a realizada para o caso de tensão.

5.2.2 – Avaliação da Segurança Dinâmica via MDT

Para exemplificar o potencial da MDT na DSA, testou-se o comportamento do sistema elético da EletrobrasEletronorte frente a uma grande perturbação, curto circuito trifásico em uma linha de transmissão. Treino e teste da árvore de decisão foram realizados com a base de dados criada para a avaliação estática. Foi aplicado um curto-

circuito trifásico de 100ms próxima à linha de transmissão TUVC-LT, em seguida a linha foi retirada.

Todos os modelos dinâmicos utilizados são fornecidos pelo ONS. A figura 37 ilustra um ramo da árvore da decisão referente a configuração topológica TOP02, estas regras da árvore de decisão podem ser utilizadas para evitar que o sistema seja levado para uma instabilidade, caso a perturbação ocorra.

Os atributos escolhidos para constar na base de dados foram os módulos das tensões em todas as barras do sistema, as potências ativa e reativa das barras de geração, a configuração topológica, e o rótulo (seguro ou inseguro, do ponto de vista da estabilidade transitória). Assim como nos estudos de casos anteriores, o nó raiz da árvore de decisão é a configuração topológica do sistema.

Caso as regras da árvore de decisão circuladas pelo caminho tracejado em azul da figura 37 sejam adotadas para aumentar a margem de estabilidade transitória do sistema, em caso de ocorrência de curto-circuito, a estabilidade do sistema será mantida. A figura 38 mostra o comparativo entre os modelos testados. Assim como nos estudos anteriores, a MDT apresenta características em seus resultados que estimulam o uso desse modelo para avaliação de segurança.

Figura 37 - Ramo da árvore de decisão que se refere à topologia TOP02, para fins de auxílio ao controle preventivo para garantir a estabilidade transitória

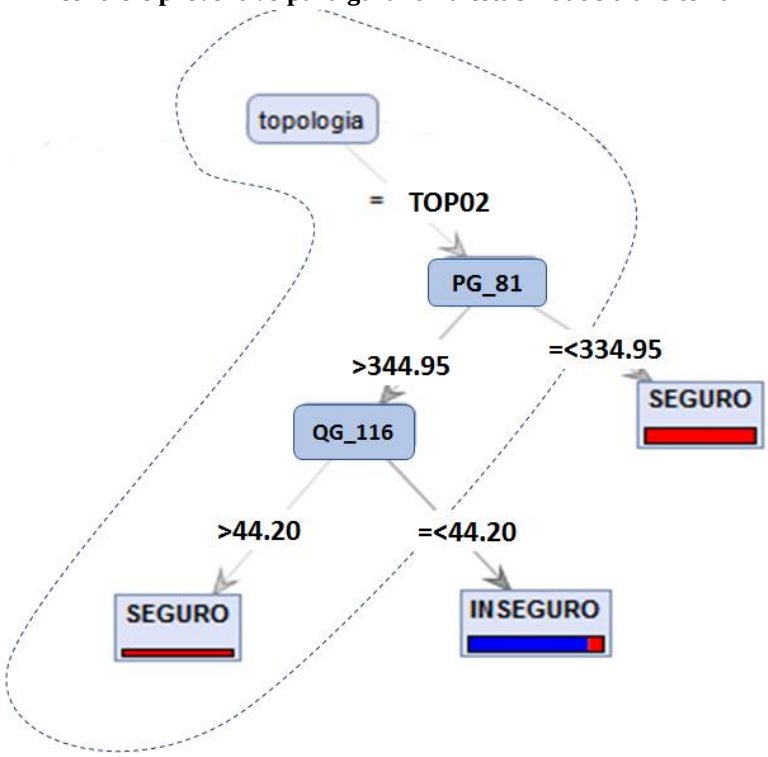
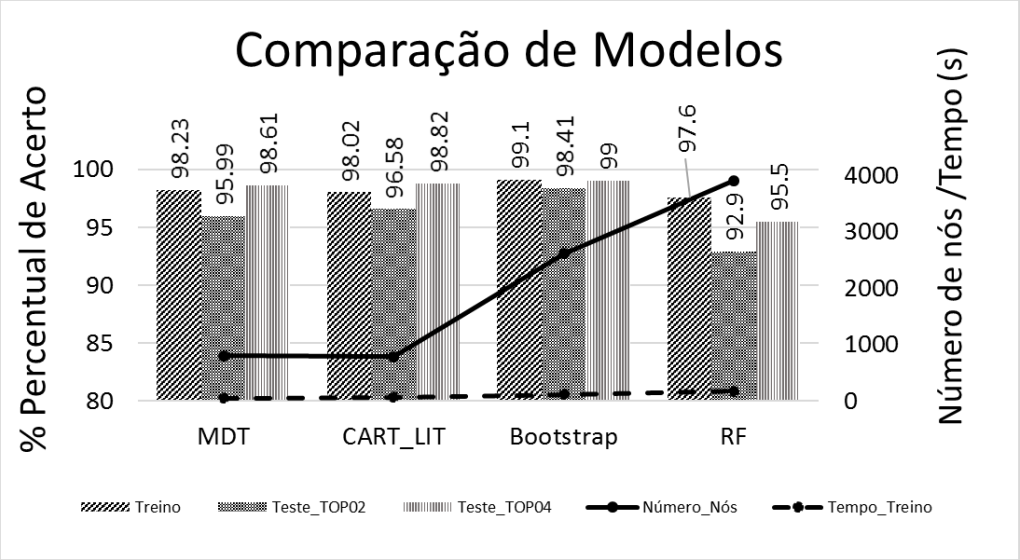


Figura 38 - Comparação dos Modelos para avaliação dinâmica para o sistema Eletronorte



5.3 Histórico Operacional

Dentro de uma condição de normalidade operacional, os SEPs operam próximos do que se planejou, isto é, os pontos de operação que ocorrem durante um determinado período tendem a não se afastar tanto dos dados de planejamento. Baseado nisso, foi proposto o uso e avaliação do histórico operacional recente como base de dados de treino e teste. Com isso se pretende avaliar o quanto os padrões operacionais se repetem.

Tomou-se como parâmetro o comportamento de carga/operação sazonal, isto é, usou-se uma semana de histórico recente para treino da MDT para ser utilizada em um dia de operação. Por exemplo, considerando que uma MDT será treinada para monitorar a segurança operacional de uma segunda-feira, para a criação da base de dados de treino dessa MDT serão tomados os pontos operacionais ocorridos de segunda a sexta da semana anterior. Para a terça-feira, por exemplo, pega-se de a partir da terça da semana anterior e os dados operacionais da segunda (véspera do dia em que será monitorado pela MDT) e assim sucessivamente.

Foram testados 4 tipos de estudos de caso, sendo eles: monitoramento de tensão, monitoramento de fluxos de potência nas linhas, monitoramento de geração de reativo e avaliação da instabilidade transitória. Os resultados de taxas de acerto da MDT para os casos de monitoramento de tensão, monitoramento de fluxo e violação de geração de reativo serão apresentados na figura 39, figura 40 e figura 41 respectivamente. Porém, por se tratarem de dados operacionais da empresa, as regras da MDT apresentadas neste trabalho não serão apresentadas.

Figura 39 -Comparação dos Modelos para avaliação estática detensão para o sistema Eletronorte considerando o histórico operacional como base de dados

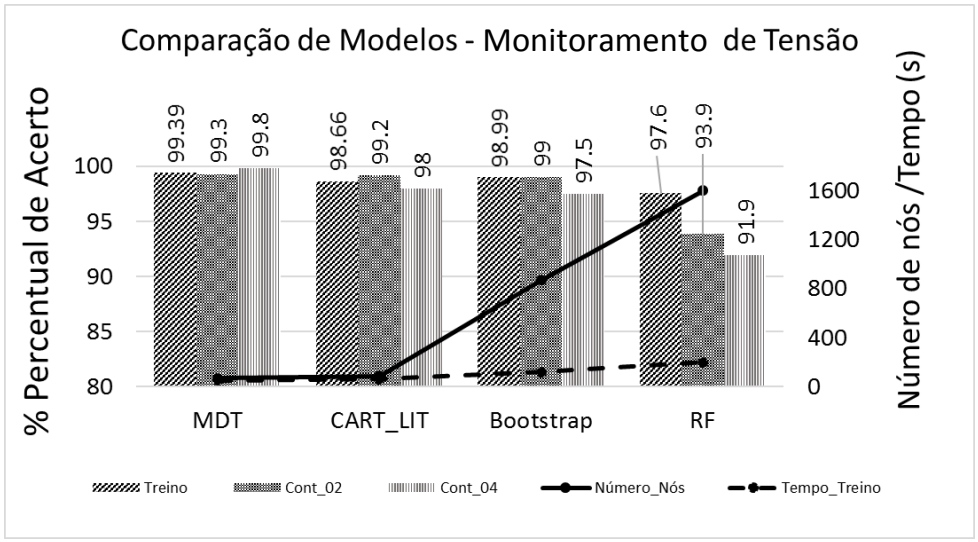


Figura 40 - Comparação dos Modelos para avaliação estática de fluxo de potência nas linhas para o sistema Eletronorte considerando o histórico como base de dados

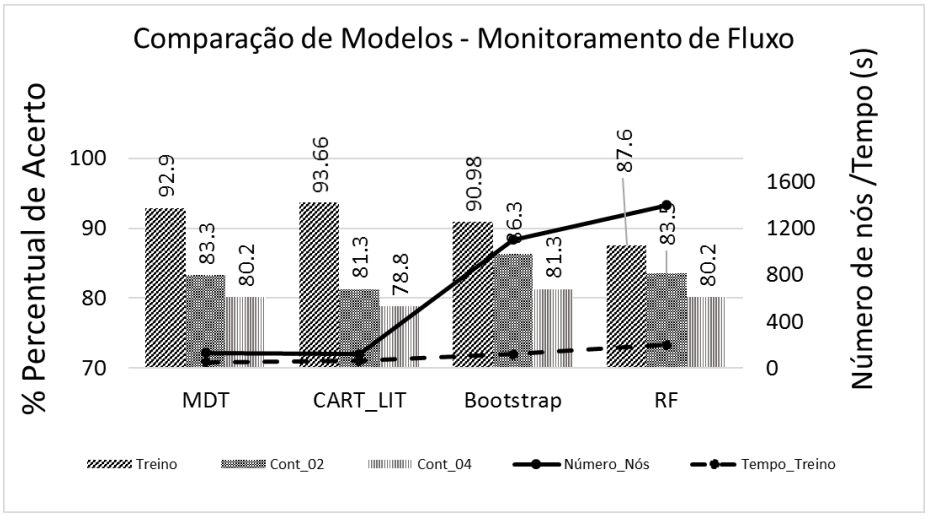
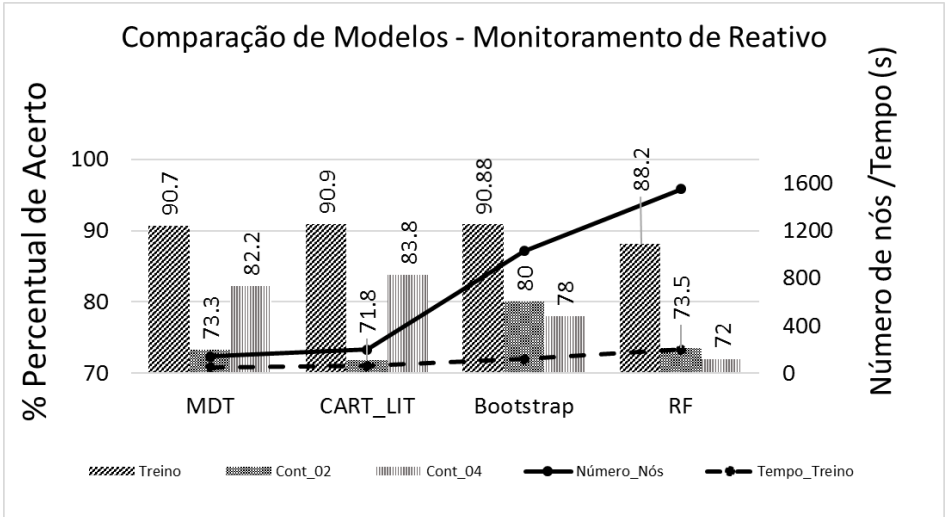


Figura 41 - Comparação dos Modelos para avaliação estática de limites de geração de reativo para o sistema Eletronorte considerando o histórico como base de dados



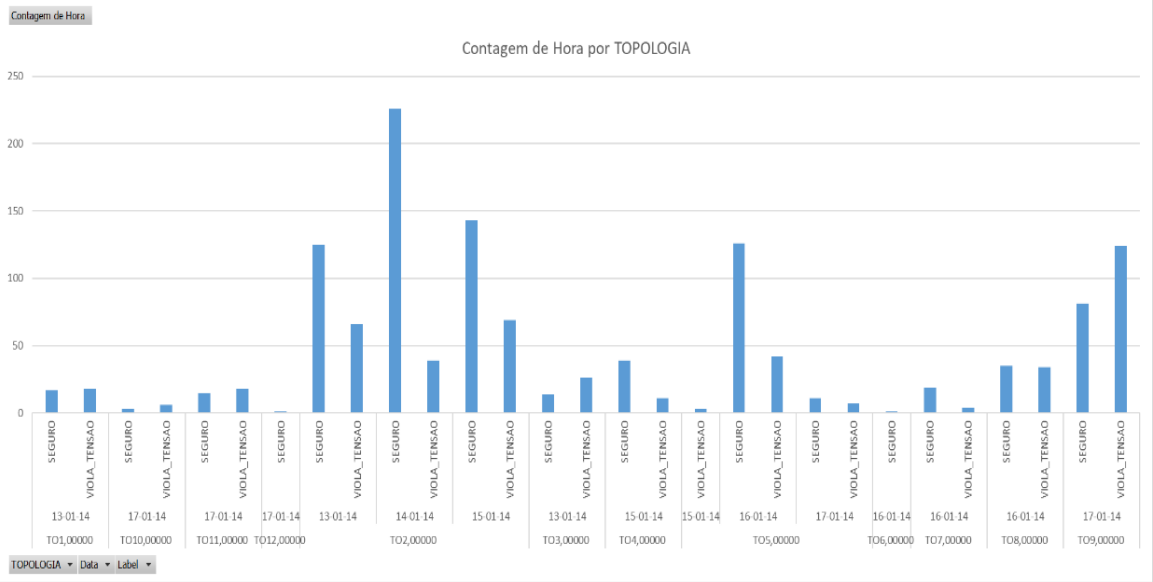
Estes resultados indicam que o uso do histórico operacional pode ser uma opção interessante para a criação da base de dados de segurança estática. Embora os resultados alcançados para o monitoramento de fluxo e geração de potência reativa terem percentuais de acerto menores que os casos apresentados anteriormente. É importante salientar 3 aspectos fundamentais para que esses resultados tenham sido alcançados.

Primeiramente, como já dito antes, o SEP opera essencialmente em pontos de operação sem violações, mesmo quando elas ocorrem são violações pequenas e que tendem a durar pouco tempo. Dessa forma, a distribuição de casos seguros e inseguros não é homogênea. Assim, a base de dados é constituída de uma maioria considerável de casos seguros.

Outra característica que se apresentou na análise dos resultados é que as violações aparecem em sua maioria nos valores de módulo de tensão. Enquanto as violações de fluxo e geração de reativo aparecem em numero bastante reduzido.

A terceira característica é a distribuição de casos por topologias, que também não são homogêneas. A figura 42 apresenta as características da base de dados, distribuição de casos e topologias, para uma semana de operação. Nota-se que as distribuições de casos são bastante variadas, o que difere do que se propõe nesta tese. Uma vez que para garantir resultados confiáveis, é necessário manter uma distribuição igualitária dos casos.

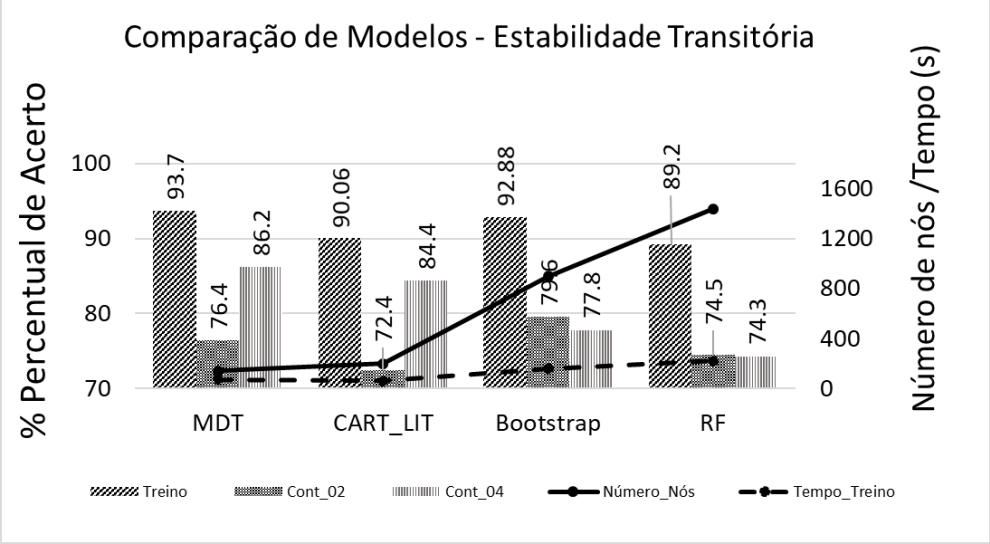
Figura 42 - Distribuição de casos por topologias durante o período analisado



Para o caso de instabilidade transitória, nos casos reais quando aplicado curto-circuito de 100 ms ou 200 ms, o sistema não apresentava problemas de instabilidade

transitória. Para que se pudesse apresentar resultados para a metologia, usou-se os casos onde se aplicava a retirada de linha nos casos de simulação estática. Assim, com o sistema já operando sob contigência, foi possível criar algumas situações que levaram o sistema a instabilidade. A partir disto, foi criada uma base de dados para avaliação da segurança dinâmica. Os resultados são apresentados na figura 43.

Figura 43 - Comparação dos Modelos para avaliação dinâmica para o sistema Eletronorte considerando o histórico como base de dados



Para instabilidade transitória, assim como nos casos anteriores baseados em histórico, nota-se uma taxa de acerto reduzida. Pelo mesmo motivo dos casos estáticos, a distribuição não homogênea das topologias e também dos casos seguros/inseguros.

É importante salientar, que estes resultados para a utilização dos dados históricos como base de dados são obtidos a partir da metodologia proposta de uso de MDT. Todo o processo de criação de base de dados, treino e teste dos modelos de máquinas de aprendizado seguem um padrão único. Desta forma, uma maneira para se melhorar os resultados nas etapas de treinos destes modelos é a utilização de função custos para se contrapor aos desbalanços nas bases de dados. Contudo, os resultados de algumas regras criadas pelos modelos podem não atingir o grau de aceitabilidade aceitável, isto é, um pequeno número de casos na base de dados criando uma regra de segurança.

Essas características não inviabilizam o uso do histórico para a criação da base de dados. Contudo, há necessidade de se complementar essa base de dados com novos pontos de operação. Os dados do histórico podem ser tratados da mesma forma que os dados de planejamento, isto é, em torno de um conjunto de pontos de operação cria-se novos casos com variações de carga, tensões em barras do tipo PV e geração. Essa estratégia deve

seguir o intuito de alcançar uma distribuição similiar de casos seguros e inseguros por topologias avaliadas.

Capítulo 6 - Conclusões

Este trabalho apresentou uma metodologia para avaliação da segurança de sistemas de potência para múltiplas contingências baseada em árvores de decisão multi-caminhos (MDT). Também foi apresentada uma metodologia para criação de banco de dados que garantiu a eficácia do modelo da MDT. Os resultados apresentados mostraram a capacidade da MDT na classificação do status de segurança, com taxas de sucesso equivalentes às técnicas mais usadas para avaliação de segurança com base em DT citadas na literatura, sendo em alguns casos ligeiramente melhores. Os benefícios do uso de MDT são o uso de um pequeno número de nós de decisão, quando comparada com outras técnicas de DTs, principalmente os ensembles. Outro benefício é a facilidade da MDT para extrair regras para a tomada de decisões, que é notável em comparação com os ensembles DT e o próprio CART, além do fato de que a separação por topologia é uma forma de visualização intuitiva para operadores de sistemas. Isto é, a MDT induz regras de segurança por meio de sentenças do tipo "Se-então", colocando em cada ramo da MDT uma topologia e suas respectivas variáveis críticas.

Os resultados obtidos pela aplicação da metodologia proposta ao sistema teste do IEEE 39 barras apresentaram elevados percentuais de acerto no uso da MDT para classificar o status de segurança estática e dinâmica. Além do mais, a MDT apresentou regras claras para cada topologia avaliada, com elevadas taxas de acerto tanto para as etapas de treino como para quando ela foi testada com casos que não participam da base de dados de treino.

Os resultados dos testes realizados com dados de operação reais do subsistema norte do SIN também mostraram elevados percentuais de acerto de classificação do estado operacional do sistema, superiores a 99% na avaliação da segurança de tensão. Isto demonstra que a MDT pode ser aplicada na operação de sistemas de maior porte em tempo real, a qual pode auxiliar os operadores na tomada de decisões em situações de contingência.

Com os resultados apresentados pelos 2 sistemas testes, pode-se concluir também que a metodologia para a criação da base de dados apresenta excelentes resultados. Não apenas por garantir altos percentuais de acerto para a classificação de segurança com a MDT, mas também pelo fato por apresentar taxas elevadas de acerto com outros modelos testados. Isso é ratificado quando se propõe o uso dos dados históricos. Uma vez que a etapa de amostragem estratificada da base não é realizada no

histórico, dessa forma não se mantém a proporcionalidade dos casos por topologia e influência diretamente no resultado final do classificador, seja MDT, CART ou random forest.

As conclusões que podem ser tomadas em relação ao uso do histórico operacional são as seguintes: Dentro da janela de dados que foram analisados, os classificadores de um modo geral apresentam novamente características semelhantes quanto ao potencial de classificação acima dos 90%. Contudo, quando se aplica os dados para validação, isto é, os dados reais do dia em que se deseja avaliar, o percentual de acerto para avaliação de tensão continua elevado, enquanto que para o monitoramento de fluxo nas linhas, por exemplo, esse percentual de acerto decaí para valores inferiores a 90%. Desta forma, o uso direto do histórico para criação da base de dados para avaliação de tensão demonstra bom potencial de utilização.

No entanto, quando se aplica a metodologia proposta de criação de base de dados para homogeneizar os dados nas topologias, os resultados melhoram significativamente. Sendo que do ponto de vista de potencial de classificação é equivalente ao uso dos dados de planejamento.

Sugestões para trabalhos futuros

- A eficiência da MDT pode ser melhorada, principalmente reduzindo o número de ramos de topologias por meio de agrupamento das mesmas. Essencialmente, o CART agrupar dados categóricos, mas esse agrupamento não leva em conta características específicas dos SEPs. Dessa forma, uma proposta para agrupamento de topologias/contingências que possuam características elétricas semelhantes pode melhorar ainda mais os resultados da MDT, isto é, reduzindo ainda mais o número de ramos da MDT. Uma vez que a cada topologia apresentada ao algoritmo, um novo ramo é criado, indicando que para um elevado número de topologias será obtida uma MDT maior.
- Implementação da simulação de Monte Carlo e Hipercubo Latino para a geração dos novos casos na etapa de criação da base de dados.
- Coleta e investigações de dados históricos atualizados.
- Uso de MDT com um método de ensemble Adaboost, visando utilizar os benefícios dos ensembles juntamente com a MDT.

Referencias

Alves da Silva, A.P., 2005,“Overcoming Limitations of NN´s for On-line DSA”. In:Proceedings of the IEEE PES General Meeting, San Francisco, USA, v.3,pp. 2653- 2660,(June).Doi:10.1109/PES.2005.1489392.

Alves, F.R.M. Henriques, R.M. Passos Fº, J.A. Gomes Jr., S. Borges, C.L.T. Falcão, D.M. Aveleda, A. Taranto, G.N. Assis, T.M.L. - Integrating Solution Engines Under a Distributed Processing Environment: An Alternative Approach for Static and Dynamic Security Assessment of Large Scale Power Systems (ieee pes general meeting, 2014

Amraee T., Ranjbar S. - Transient Instability Prediction Using Decision Tree Technique, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 28, NO. 3, AUGUST 2013

Amraee, A. and S. Ranjbar, "Transient Instability Prediction Using Decision Tree Technique," in IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 3028-3037, Aug. 2013. doi: 10.1109/TPWRS.2013.2238684

Assis, Tatiana Mariano Lessa de, “Cálculo da Capacidade de Transmissão Dinâmica em Sistemas de Potência através de Ferramentas Integradas e Sistemas Inteligentes”, Tese de doutorado, COPPE-UFRJ, 2007.

BALU, N.; et al. On-line power system security analysis. Proceedings of the IEEE, Volume: 80. 1992.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. Classification andRegression Trees. London, U.K.: Chapman and Hall/CRC, 1984.

C. FatihKucuktezcan, V.M. IstemihanGenc, A new dynamic security enhancement method via genetic algorithms integrated with neural network based tools, Electric Power Systems Research, Volume 83, Issue 1, February 2012, Pages 1-8, ISSN 0378-7796, http://dx.doi.org/10.1016/j.epsr.2011.09.004.

C. Liu et al., 2014, "A Systematic Approach for Dynamic Security Assessment and the Corresponding Preventive Control Scheme Based on Decision Trees," in IEEE Transactions on Power Systems, vol. 29, no. 2, pp. 717-730, March 2014.

C. Singh, X. Luo and H. Kim, "Power System Adequacy and Security Calculations Using Monte Carlo Simulation incorporating Intelligent System Methodology," 2006 International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, 2006, pp. 1-9.doi: 10.1109/PMAPS.2006.360224

CARPENTIER, J. Static Security Assessment and Control: a Short Survey. Athens Power Tech, 1993. APT 93. Proceedings. Joint International Power Conference.

CARPENTIER, J. Static Security Assessment and Control: a Short Survey. Athens Power Tech, 1993. APT 93. Proceedings. Joint International Power Conference.

Chien YC, Yuafi HY. Estimation of line flows and bus voltages using decision trees. IEEE Trans. Power Syst., 1994 9(3):1569-1574

CIHLAR, T. C.; et al. Eletric Utility System Security. In: Proceedings of the American Power Conference, Vol. 31, pp. 891-908, 1969.

CRISP-DM 1.0 Step-by-step data mining guide, CRoss Industry Standard Process for Data Mining, SPSS.[The CRISP-DM Consortium, 2000

Delibasic, B.; Vukicevic, M.; Jovanovic, M.; Suknovic, M., "White-Box or Black-Box Decision Tree Algorithms: Which to Use in Education?," in Education, IEEE Transactions on , vol.56, no.3, pp.287-291, Aug. 2013 doi: 10.1109/TE.2012.2217342.

Denis C.L. Costa, Marcus V.A. Nunes, João P.A. Vieira, Ubiratan H. Bezerra, Decision tree-based security dispatch application in integrated electric power and natural-gas networks, Electric Power Systems Research, Volume 141, 2016, Pages 442-449, ISSN 0378-7796

Diao R., Sun K., Vittal V., O'Keefe R., Richardson M., Bhatt N, Stradford D, Sarawgi S., Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements - IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 24, NO. 2, MAY 2009

Diao, R. Vittal, V. ; Logic, N. - Design of a Real-Time Security Assessment Tool for Situational Awareness Enhancement in Modern Power Systems, IEEE transaction on Power System, 2010

DIAO, R.; VITTAL, V.; LOGIC, N. Design of a real-time security assessment tool for situational Awareness Enhancement in Modern Power Systems. IEEE Trans. Power Syst., vol. 25, no. 2, pp. 957–965, May 2010.

DIAO, R.; VITTAL, V.; LOGIC, N. Design of a real-time security assessment tool for situational Awareness Enhancement in Modern Power Systems. IEEE Trans. Power Syst., vol. 25, no. 2, pp. 957–965, May 2010.

DY LIACCO, T. E. Real-Time Computer Control of Power Systems. Proceedings of the IEEE, Vol. 62, no 7, July 1974, pp.884-891.

Fayyad, U. Piatetsky-Shapiro, G. ,SmythFrom, P - Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, 1996

FINK, L. H.; CARLSEN, K. Operating under stress and strain. IEEE Spectrum, Vol. 15, March 1978, pp. 48-53.

FINK, L. H.; CARLSEN, K. Operating under stress and strain. IEEE Spectrum, Vol. 15, March 1978, pp. 48-53.

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 IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 25 AUGUST 2010

GENC, I.; DIAO, R.; VITTAL, V. Computation of Transient Stability Related Security Regions and Generation Rescheduling Based on Decision Trees. 07/2010, IEEE PES General Meeting, Minneapolis, Minnesota, 25.07.2010 - 29.07.2010

GENC, I.; DIAO, R.; VITTAL, V. Computation of Transient Stability Related Security Regions and Generation Rescheduling Based on Decision Trees. 07/2010, IEEE PES General Meeting, Minneapolis, Minnesota, 25.07.2010 - 29.07.2010

Genc, R. Diao, V. Vittal, S. Kolluri, S. Mandal,"Decision Tree-Based Preventive and Corrective Control Applications for Dynamic Security Enhancement in Power Systems," IEEE Trans. Power Systems, vol. 25, no. 3, pp. 1611–1619, Aug. 2010.

Gholami, M., G.B. Gharehpetian, M. Mohammadi, Online Decision Tree based strategy for power system static security margin improvement using wind farms, International Journal of Electrical Power & Energy Systems, Volume 83, 2016, Pages 15-20, ISSN 0142-0615, http://dx.doi.org/10.1016/j.ijepes.2016.03.030.

Guo, T. and Milanović, J. V., "Online Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining," in IEEE Transactions on Power Systems, vol. 31, no. 3, pp. 1760-1768, May 2016.

Guo, T. and J. V. Milanović, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability," in IEEE Transactions on Power Systems, vol. 29, no. 1, pp. 377-385, Jan. 2014.

Gutierrez-Martinez, Victor J.; Cañizares , Claudio A.; Fuerte-Esquivel, Claudio R.; Pizano-Martinez, Alejandro; Gu, Xueping, “Neural-Network Security-Boundary Constrained Optimal Power Flow”, 2011.

Han, J., Kamber, M., Pei, J. - Data Mining Concepts and Techiniques, Ed. Elsevier, 2012

Hanieh Mohammadi, Maryam Dehghani, PMU based voltage security assessment of power systems exploiting principal component analysis and decision trees, International Journal of Electrical Power & Energy Systems, Volume 64, 2015, Pages 655-663, ISSN 0142-0615

Haque, M.T., Kashtiban, A.M., 2005,“Application of Neural Networks in Power Systems, A Review”. In:Proceedings of World Academy of Science, Engineering And Technology, v. 6. ISSN: 1307-6884.

Harrington, P., 2012; “Machine Learning in Action”,

Hatziargyriou, N. D.; Contaxis, G. C.; Sideris, N. C. A decision tree method for on-line steady state security assessment, Power Systems, IEEE Transactions on, Volume: 9, 1994, Page: 1052-1061.

Hatziargyriou, N.D.; Contaxis, G.C.; Sideris, N.C., "A decision tree method for on-line steady state security assessment," in Power Systems, IEEE Transactions on, vol.9, no.2, pp.1052-1061, May 1994. doi: 10.1109/59.317626

HAYKIN, S., . “Redes Neurais princípios e prática”. Editora ARTMED, 2008.

He, Miao; Junshan Zhang; Vittal, V., "Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning," in Power Systems, IEEE Transactions on, vol.28, no.4, pp.4089-4098, Nov. 2013 doi: 10.1109/TPWRS.2013.2266617

HERNANDEZ, EMÍLIO DEL MORAL, Inteligência Computacional e Redes Neurais em Engenharia Elétrica, USP, São Paulo, 2005.

I. Kamwa, S. R. Samantaray and G. Joos, "Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices," in IEEE Transactions on Smart Grid, vol. 1, no. 2, pp. 144-158, Sept. 2010. doi: 10.1109/TSG.2010.2052935

Junior, Antonio carlos leal de castro, "Estudo de Controle Preventivo para Análise do Colapso de Tensão", 2009.

K. Morison, L. Wang and P. Kundur, "Power System Security Assessment", IEEE Power & Energy Magazine, September/October, 2004.

Kamwa I., Samantaray S., Joos G., - Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices IEEE TRANSACTIONS ON SMART GRID, VOL. 1, NO. 2, SEPTEMBER 2010

Kamwa I., Samantaray S., Joos G., - On the Accuracy Versus Transparency Trade-Off of Data-Mining Models for Fast- Response PMU-Based Catastrophe Predictors- IEEE TRANSACTIONS ON SMART GRID, VOL. 3, NO. 1, MARCH 2012

Khoshkhoo H, Shahrtash SM. On-line dynamic voltage instability prdiction based on decision tree supported by a wide-area measurement system. IET Gener. Transm. Distrb., 2012, 6(11):1143-1152.

Krishnan, Venkat; McCalley ,James D.; Henry ,Sebastien; Issad, Samir Issad, "Efficient Database Generation for Decision Tree Based Power System Security Assessmen"t, IEEE Transactions On Power Systems, Vol. 26, No. 4, November 2011

Krishnan, Venkat; McCalley, James D., Progressive entropy based contingency grouping for deriving decision trees for multiple contingencies, 2012.

Kumar, V., Tan, P. Steinbach, M. - Introduction to Data Mining, Ed. Pearson, 2006

Kundur P – Power System Stability and Control, 1994

KUNDUR, P., et al. Definition and classification of power system stability. IEEE/CIGRE joint task force on stability terms and definitions. 2004.

Li, Z., Wu, W. Phasor Measurements-Aided Decision Trees for Power System Security Assessment. College of Electrical Engineering, Zhejiang University, Hangzhou, China, 2009. ICIC '09. Second International Conference on Information and Computing Science.

Lin, Yu-Jen “Comparison of CART- and MLP-based power system transient stability preventive control”, 2012

Liu C., Sun K., Rather Z., Chen Z., LethBak C., Thøgersen P., Lund P., A Systematic Approach for Dynamic Security Assessment and the Corresponding Preventive Control Scheme Based on Decision Trees IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 29, NO. 2, MARCH 2014

M. He, J. Zhang, and V. Vittal, “A data mining framework for online dynamic security assessment: Decision trees, boosting, and complexity analysis,” in Proc. 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), Jan. 2012, pp. 1–8.

Machowski, J., Bailek, J. W., Rumby, J. R. - Power Systen Dynamics Stability and Control, Ed. Wiley 2008

Mariotto, Lenois, “Avaliação da Segurança da Operação de Sistemas Elétricos de Potência Considerando os Limites de Estabilidade Angular e de Tensão”, 2008.

Mauricio Campos Passaro, 2013 CONTROLE PREVENTIVO DE ESTABILIDADE ATRAVÉS DE SENSIBILIDADE BASEADA EM REDES NEURAIS, Tese de Doutorado, COPPE/UFRJ

Miao He, Vittal V, Zhang J., Online Dynamic Security Assessment With Missing PMU Measurements: A Data Mining Approach - IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 28, NO. 2, MAY 2013

Miao He, Vittal V, Zhang J., Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning - IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 28, NO. 4, NOVEMBER 2013

Milan Vukićević, Miloš Jovanović, Boris Delibašić, Sonja Išljamović, And Milija Suknović "Reusable Component-Based Architecture for Decision Tree Algorithm Design" International Journal on Artificial Intelligence Tools 2012.

Mitchell, T. M., “Machine Learning”. McGraw Hill, 1997

MONTICCELI, A. Fluxo de carga em redes de energia elétrica. Edgard Blücher Ltda, 1a Edição, São Paulo, 1983.

MORISON, K.; WANG, L.; KUNDUR, P. Power System Security Assessment. IEEE Power & Energy Magazine, September/October, 2004.

MORISON, K.; WANG, L.; KUNDUR, P. Power System Security Assessment. IEEE Power & Energy Magazine, September/October, 2004.

Moulin, L.S., Alves da Silva, A.P., El-Sharkawi, M.A., et al, 2004. “Support Vector Machines for Transient Stability Analysis of Large-Scale Power Systems”,IEEE Transactions on Power Systems, v. 19, nº 2,pp. 818-825,(May), 2004. Doi: 10.1109/TPWRS.2004.826018.

Nima Amjady, Dynamic voltage security assessment by a neural network based method, Electric Power Systems Research, Volume 66, Issue 3, September 2003,

Pages 215-226, ISSN 0378-7796, http://dx.doi.org/10.1016/S0378-7796(03)00048-8.

OLIVEIRA, A. R., 2005. “Redes Neurais Artificiais Aplicadas na Detecção, Classificação e Localização de Defeitos em Linhas de Transmissão”. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Elétrica, UFJF, 2005.

PAL, J. Power System Static Security Analysis and Control: Use of hybrid expert system for MW and voltage security assessment of power system in an on-line environment. LAP Lambert Academic Publishing AG & Co. KG, 2010.

PAL, J. Power System Static Security Analysis and Control: Use of hybrid expert system for MW and voltage security assessment of power system in an on-line environment. LAP Lambert Academic Publishing AG & Co. KG, 2010.

Papadopoulos, P. N. ; T. Guo; J. V. Milanovic, "Probabilistic Framework for Online Identification of Dynamic Behavior of Power Systems with Renewable Generation," in IEEE Transactions on Power Systems , vol.PP, no.99, pp.1-1. 2017

Passaro, Mauricio - CONTROLE PREVENTIVO DE ESTABILIDADE ATRAVÉS DE SENSIBILIDADE BASEADA EM REDES NEURAIS, Tese de Doutorado, 2013

PEREIRA, A. L. S, 2009. “Utilização de Redes Neurais Artificiais no Ajuste de Controladores Suplementares e Dispositivo Facts STATCOM para a Melhoria da Estabilidade a Pequena Perturbações do Sistema Elétrico de Potência”. Tese de Doutorado, Faculdade de Engenharia - UNESP, 2009.

Perner, P., “How to compare and interpret two learnt Decision Trees from the same Domain”. 27th International Conference on Advanced Information Networking and Applications Workshops, 2013

Perner, P., “A Method for Supporting the Domain Expertby the Interpretation of Different Decision Trees Learnt from the Same Domain”. Qual. Reliab. Engng. Int. 2014, 30 985–992.

QUINLAN, J. R. Machine Learning. Springer, Vol.01, Issue 01, 1986.

QUINLAN, J. R. Machine Learning. Springer, Vol.01, Issue 01, 1986.

R. Diao, K. Sun, V. Vittal, R. O’Keefe, M. Richardson, N. Bhatt, D. Stradford, and S. Sarawgi, “Decision tree-based online voltage security assessment using PMU measurements,” IEEE Trans. Power Syst., vol. 24, no. 2, pp. 832–839, May 2009.

Rather Z., Lui C., Chen Z., LethBak C, Thøgersen P - Dynamic Security Assessment of Danish Power System Based on Decision Trees: Today and Tomorrow

RODRIGUES, B. G. D - Uso de Árvore de Decisão para Avaliação da Segurança Estática em tempo real de Sistemas Elétricos de Potência. Dissertação de Mestrado, PPGEE, UFPA 2014

ROVNYAK, S.; et al. Decision Trees for Real-Time Transient Stability Prediction. IEEE Transactions on Power Systems, Vol. 9, No. 3, August 1994, pp. 1417-1426.

ROVNYAK, S.; TAYLOR, C.; SHENG, Y. Decision Trees Using Apparent Resistance to Detect Impending Loss of Synchronism. IEEE Transactions on Power Delivery, Vol. 15, No.4, October 2000, pp. 1157 - 1162.

S. Kalyani and K. S. Swarup, "Binary SVM Approach for Security Assessment and Classification in Power Systems," 2009 Annual IEEE India Conference, Gujarat, 2009, doi: 10.1109/INDCON.2009.5409433.

S. Kalyani and K. Shanti Swarup, "Classification and Assessment of Power System Security Using Multiclass SVM," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 5, pp. 753-758, Sept. 2011.doi: 10.1109/TSMCC.2010.2091630

S. R. Samantara, I. Kamwa and G. Joos, "Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame," in IET Generation, Transmission & Distribution, vol. 4, no. 12, pp. 1334-1348, December 2010. doi: 10.1049/iet-gtd.2010.0201

Saeh, I. S., Mustafa, M. W., Machine Learning Classifiers for Steady State Security Evaluation in Power System. International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012.

Saeh, I.S.; Khairuddin, A., "Decision Tree for Static Security Assessment Classification," in Future Computer and Communication, 2009. ICFCC 2009.

Santa Rosa, Arthur da Silva, “Influência da não representação de dispositivos de controle na análise de segurança de tensão”, 2007.

Sharifzadeh, H. et al, 2017, “Multi-period stochastic security-constrained OPF considering the uncertainty sources of wind power, load demand and equipment unavailability, Electric Power Systems Research, Volume 146, May 2017, Pages 33-42, ISSN 0378-7796,

Sun K. , Likhate S., Vittal V., Kolluri V., Mandal S., - Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees - IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 22, NO. 4, NOVEMBER 2007

SUN, K.; et al. An online dynamic security assessment scheme using phasor measurements and decision trees. IEEE Trans. Power Syst., vol. 22, no. 4, pp. 1935–1943, Nov. 2007.

SUN, K.; et al. An online dynamic security assessment scheme using phasor measurements and decision trees. IEEE Trans. Power Syst., vol. 22, no. 4, pp. 1935–1943, Nov. 2007.

Supreme, H. ; Dessaint, l. A.; Kamwa, I. ; A. Heniche-Oussedik, "Development of New Predictors Based on the Concept of Center of Power for Transient and Dynamic Instability Detection," in IEEE Transactions on Smart Grid , vol.PP, no.99, pp.1-1

Swarup, K. S.; Mastakar, R.; Reddy K. V. P.; Decision Tree for Steady State SecurityAssessment and Evaluation of Power Systems. – Procedings of ICISIP, 2005.

Swarup, K.S.; Mastakar, R.; Prasad Reddy, K.V., "Decision tree for steady state security assessment and evaluation of power systems," in Intelligent Sensing and Information Processing, 2005. Proceedings of 2005 International Conference on, vol., no., pp.211-216, 4-7 Jan. 2005. doi: 10.1109/ICISIP.2005.1529450

Task Force IEEE - Benchmark Systems for Small-Signal Stability Analysis and Control, 2015

Ubiratan H. Bezerra, João P. A. Vieira, Werbeston D. Oliveira, Daniel A.Martins, Dione J. A. Vieira, Bernard C. Bernardes, Benedito G. D. Rodrigues e Vilson Castro Metodologia de Controle Preventivo Baseada em Árvore de Decisão para a Melhoria da Segurança Estática e Dinâmica do Sistema Interligado da Eletronorte, IX CITENEL, 2017

Verma, K., Niazi, K. R., 2012," Generator Coherency Determination in a Smart Grid Using Artificial Neural Network" IEEE Power Engineering Society General Meeting, San Diego, EUA, pp. 1-7. Doi: 10.1109/PESGM.2012.6345255.

VILLANUEVA, W. J. P, 2011. "Síntese Automática de Redes Neurais Artificiais com Conexões à Frente Arbitrárias". Tese de Doutorado, Faculdade de Engenharia Elétrica e Computação, Campinas, São Paulo, 2011.

Wehenkel L. Machine learning approaches to power-system security assessment. IEEE Intelligent Systems and Their Applications, 1997, 12(5):60-72.

Wehenkel, L. A. 1998, Automatic Learning Techniques in Power Systems. Norwell, MA: Kluwer.

Wehenkel, L. A., "Automatic Learning Techniques in Power Systems". Norwell, MA: Kluwer, 1998.

Wehenkel, L., Pavella, M., Decision tree approach to power systems security assessment. – University of Liege- Institut Montefiore, Department of Electrical Engineering, Sart-Tilman B28, B-4000 Liege. Belgium, February 1993.

WEHENKEL, L.; PAVELLA, M. Decision Trees and Transient Stability of Electric Power Systems - Automatica, VoL 27, No. 1, pp. 115-134, 1991

WEHENKEL, L.; et al. Decision tree based transient stability method - A case study. IEEE Trans. on PS, Vol. 9, No. 1, 1994, pp. 459-469.

Witten I. H., Frank E., Hall M. A. – Data mining Practical Machine Learning Tools and Techiniques, Elsevier 2011.

www.dsatools.com/ (2017) – site de feramentas comerciais de análise de sistemas de potência

www.itesla-project.eu/ (2017) - The iTesla project aimed at improving network operations with a new security assessment tool able to cope with increasingly uncertain operating conditions and take advantage of the growing flexibility of the grid.

www.powertechlabs.com/ (2017) – Site da empresa de energia.

Zhiyong Li; Weilin Wu, "Phasor Measurements-Aided Decision Trees for Power System Security Assessment," in Information and Computing Science, 2009. ICIC '09. Second International Conference on, vol.1, no., pp.358-361, 21-22 May 2009. doi: 10.1109/ICIC.2009.98

Zhongbo Zhang, Shuanghu Zhang, Simin Geng, Yunzhong Jiang, Hui Li, Dawei Zhang, Application of decision trees to the determination of the year-end level of a carryover storage reservoir based on the iterative dichotomizer 3, International Journal of Electrical Power & Energy Systems, Volume 64, 2015, Pages 375-383, ISSN 0142-0615,