

PONTÍFICA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGE

**DETERMINAÇÃO DE CAUSAS DE INTERRUPÇÕES NÃO PROGRAMADAS EM
SISTEMAS ELÉTRICOS UTILIZANDO REDES BAYESIANAS E LÓGICA FUZZY**

RODRIGO LAUX TABBAL

Porto Alegre

2010

RODRIGO LAUX TABBAL

**DETERMINAÇÃO DE CAUSAS DE INTERRUPÇÕES NÃO PROGRAMADAS EM
SISTEMAS ELÉTRICOS UTILIZANDO REDES BAYESIANAS E LÓGICA FUZZY**

Dissertação apresentada como requisito para
obtenção do grau de Mestre pelo Programa de
Pós-graduação da Faculdade de Engenharia
Elétrica da Pontifícia Universidade Católica
do Rio Grande do Sul.

Orientador: Prof. Flávio Antonio Becon Lemos, Dr.

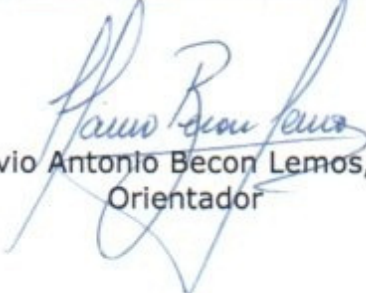
Porto Alegre

2010

“DETERMINAÇÃO DE CAUSAS DE INTERRUPÇÕES NÃO PROGRAMADAS EM SISTEMAS ELÉTRICOS UTILIZANDO REDES BAYESIANAS E LÓGICA FUZZY”

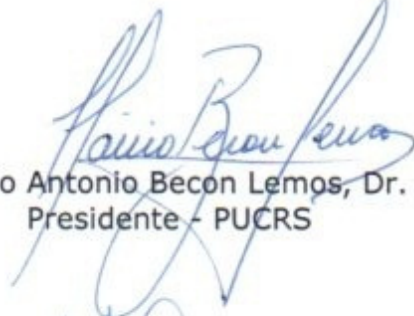
RODRIGO LAUX TABBAL

Esta dissertação foi julgada para a obtenção do título de MESTRE EM ENGENHARIA e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica do Rio Grande do Sul.


Flávio Antonio Becon Lemos, Dr.
Orientador

Rubem Dutra Ribeiro Fagundes, Dr.
Coordenador
Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora:


Flávio Antonio Becon Lemos, Dr.
Presidente - PUCRS


Daniel Ferreira Coutinho, Dr.
PUCRS


Raimundo Celeste Ghizoni Teive, Dr.
UNIVALI

AGRADECIMENTOS

Agradeço a Deus por superações de desafios nesse caminho de persistência e criatividade para o desenvolvimento deste trabalho. A meus pais e minha namorada, pelo apoio e incentivo incondicionais. Aos professores Odilón Duarte, Vinícius Licks e Daniel Coutinho por fornecerem o *start up* para o ingresso no curso de mestrado. Ao professor e orientador Flávio Lemos, me encaminhando e propiciando a concretização desta realização. Ao suporte dos engenheiros Alex Tronchoni e Jayme Filho e ao grupo Rio Grande Energia S.A. (RGE). Agradeço aos demais professores, colegas de mestrado, que marcaram época, e a todas as pessoas que de alguma forma, em algum momento, contribuíram direta ou indiretamente para o desenvolvimento desta dissertação.

SUMÁRIO

Lista de Figuras

Lista de Tabelas

Resumo

Abstract

1 INTRODUÇÃO	14
1.1 REVISÃO BIBLIOGRÁFICA	16
2 IDENTIFICAÇÃO DE CAUSAS	26
2.1 FALHAS EM EQUIPAMENTO	26
2.1.1 Transformadores	26
2.1.2 Redes de distribuição	27
2.1.3 Chaves	28
2.1.4 Para-raios	29
2.1.5 Isoladores e conexões	31
2.2 ANIMAIS	31
2.3 CLIMA ADVERSO	32
2.3.1 Vento	33
2.3.2 Descarga atmosférica	33
2.3.3 Climas quentes	34
2.4 ÁRVORES	35
2.5 INTERFERÊNCIA HUMANA	37
2.6 ANÁLISE DE DESLIGAMENTOS	38
3 LÓGICA FUZZY	40
3.1 FUNDAMENTO	40
3.2 CONCEITOS	41
3.2.1 Conjuntos fuzzy	41
3.2.2 Funções de Pertinência	43
3.2.3 Operadores Lógicos	44
3.2.4 Regras Se-Então	45
3.2.5 Agregação	47
3.2.6 Defuzzyficação	47
3.3 VANTAGENS E DESVANTAGENS DA LÓGICA FUZZY	49
3.3.1 Vantagens	49
3.3.2 Desvantagens	50

4	REDE BAYESIANA	51
4.1	FUNDAMENTO	51
4.2	INFERINDO COM INCERTEZAS	51
4.3	REDES CAUSAIS	52
4.4	CÁLCULO PROBABILÍSTICO	53
4.4.1	Axiomas Básicos	54
4.4.2	Probabilidades Condicionais	54
4.4.3	Regra Fundamental	55
4.4.4	Marginalização	56
4.5	TOPOLOGIAS E D-SEPARAÇÃO	57
4.6	DEFINIÇÃO DE REDES BAYESIANAS	59
4.7	A REGRA DA CADEIA PARA REDES BAYESIANAS	60
4.8	ALGORITMO DE INFERÊNCIA <i>BUCKET ELIMINATION</i>	60
4.9	MÉTODO <i>JUNCTION TREE</i>	62
4.10	OUTROS ALGORITMOS	63
4.11	PARAMETRIZAÇÃO DA REDE BAYESIANA	64
5	METODOLOGIA	68
5.1	INTRODUÇÃO	68
5.2	DEFINIÇÃO DAS VARIÁVEIS	72
5.3	TRATAMENTO DE DADOS	74
5.3.1	Limpeza e Integração	75
5.3.2	Mineração de Dados	77
5.4	ADEQUAÇÃO E ANÁLISE DO BANCO DE DADOS	79
5.4.1	Análise do Banco de Dados	80
5.4.2	Adequação dos Registros	82
5.5	IMPLEMENTAÇÃO DA REDE BAYESIANA	84
5.5.1	Representação da Estrutura Bayesiana	85
5.5.2	Determinação dos Parâmetros da Rede Bayesiana	86
5.5.3	Inferências da RB e critério de saída	86
5.5.4	Análise da saída da Rede Bayesiana	87
5.6	IMPLEMENTAÇÃO DO SISTEMA FUZZY	90
5.6.1	Estabelecimento das Variáveis	90
5.6.2	Funções de Pertinência	91
5.6.3	Conversão dos Valores de Entrada do Bloco fuzzy	93
5.6.4	Base de Regras	94
5.6.5	Análise das saídas do sistema fuzzy	95
5.7	ANÁLISE DO SISTEMA INTEGRADO	96
6	RESULTADOS	98
6.1	RESULTADOS AVALIADOS	98
6.1.1	Análise de agrupamento de dados	98

6.1.2	Resultados dos algoritmos de inferência.....	100
6.1.3	Efeito da etapa de adequação na base de dados	101
6.1.4	Caracterização da base de dados.....	102
6.1.5	Eficácia da base de regras fuzzy	105
6.1.6	Inferência do Sistema Integrado RB / fuzzy.....	109
6.1.7	Inferência de dados incoerentes	112
7	COMENTÁRIOS FINAIS.....	115
7.1	TRABALHOS FUTUROS.....	119
	REFERÊNCIAS.....	120
	ANEXOS	124
	ANEXO A - PROPRIEDADES DE CONJUNTOS FUZZY (MATHWORKS, 1995).....	124

LISTA DE FIGURAS

Figura 2-1– Pára-raios no sistema elétrico. (OSHA, 2010).....	30
Figura 2-2 – Pára-raios do tipo carboneto de silicone e varistor de óxido metálico. (DOBLE, 2010).....	30
Figura 2-3– Anteparos para impedir a interferência de animais. (BROWN, 2002).....	32
Figura 2-4 – Incidências de interrupções para três companhias de energia. (BROWN, 2002)	38
Figura 3-1 – Definição das estações do ano por conjuntos binários. (MATHWORKS, 1995)	42
Figura 3-2 – Definição das estações do ano por conjuntos fuzzy. (MATHWORKS, 1995) ...	42
Figura 3-3 – Funções de pertinência. (MATHWORKS, 1995).....	44
Figura 3-4 – Correspondência entre as operações de lógica booleana e clássica. (MATHWORKS, 1995)	45
Figura 3-5 – Regra determinando a relação entre as variáveis fuzzy <i>volume e densidade</i>	46
Figura 3-6 – Esquema da etapa de <i>agregação</i> . (MATHWORKS, 1995).....	47
Figura 3-7 – Modelagem de uma função f através de regiões de áreas hachuradas compondo f^* . (ZADEH, 1994)	48
Figura 3-8 – Refinamento da função f^* obtida por meio da etapa de <i>defuzzificação</i>	48
Figura 4-1 – Rede causal em que A é Pai de B	53
Figura 4-2 – Rede causal do sistema de ignição de um automóvel.	53
Figura 4-3 – Conexão serial.	58
Figura 4-4 – Conexão divergente.	58
Figura 4-5 – Conexão convergente.....	58
Figura 4-6 – Exemplo de Rede Bayesiana. Evidências $\{D=d, F=f\}$	61
Figura 4-7 – Transformação de um grafo (G) em <i>junction tree</i> (T). (PASKIN, 2003).....	63
Figura 4-8 – Parâmetros de uma Rede Bayesiana. (JENSEN, 2001).....	65

Figura 5-1 – Sistema elaborado por Pretto (2005).	69
Figura 5-2 – Processo de KDD. (TRONCHONI, 2008).....	70
Figura 5-3 – Digrama simplificado proposto.	71
Figura 5-4 – Planilha do Banco de Dados com alguns campos. Observa-se a esparsidade dos registros.	74
Figura 5-5 – Desligamentos descartados, correspondendo a 47.617 eventos.	76
Figura 5-6 – Relação de dados eliminados.....	76
Figura 5-7 – Proporção dos principais registros de causas de desligamento não programados.	77
Figura 5-8 – Processo de Mineração	78
Figura 5-9 – Lista parcial apresentando algumas combinações possíveis do estado <i>Clima Adverso</i> a partir de suas entradas.....	81
Figura 5-10 – Lista parcial de frequência absoluta dos eventos. Observa-se a ordem decrescente do campo Quantidade (Qtde).....	82
Figura 5-11 – Evento com mais de uma saída para a mesma entrada de dados.....	83
Figura 5-12 – Proporção de causas para os 15.994 eventos.....	84
Figura 5-13 – Estrutura da Rede Bayesiana	85
Figura 5-14 – Ilustração da saída da RB para determinadas ocorrências.....	87
Figura 5-15 – Rede Bayesiana com as probabilidades de cada estado <i>a priori</i> , segundo os parâmetros estimados com 7.994 eventos.	88
Figura 5-16 – Inferências das causas para os 7.994 eventos.	89
Figura 5-17 – Topologia do Sistema fuzzy com entradas e saídas.....	91
Figura 5-18 – Funções de pertinência correspondendo aos estados da variável <i>Clima</i>	92
Figura 5-19 - Funções de pertinência de determinada causa, contendo as possibilidades de ocorrência correspondentes a: B- <i>Baixa</i> , M- <i>Média</i> e A- <i>Alta</i>	92
Figura 5-20 – Planilha contendo os estados em destaque para a ocorrência de <i>Descarga Atmosférica</i>	94

Figura 5-21 – Base de regras do sistema fuzzy.	95
Figura 5-22– Fluxograma do tratamento de dados e inferência do sistema integrado.	97
Figura 6-1 - Eventos de sobrecarga agrupados por clima para os 26.282 eventos.	99
Figura 6-2 – Eventos de sobrecarga agrupados por turno para os 26.282 eventos.	99
Figura 6-3 – Análise de <i>clusters</i> conforme o <i>Clima</i> (a) e <i>Turno</i> (b) nos 7.994 eventos.	100
Figura 6-4 – Simulações utilizando-se base de dados com saídas múltiplas e únicas.	102
Figura 6-5 – Conjuntos de dados com 85 eventos de saídas únicas - (a) sem as três variáveis envolvidas; (b) incluindo as três variáveis.	103
Figura 6-6 – Resultados da saída do bloco fuzzy para causas <i>Vegetal</i> . As saídas com o máximo valor foram grifadas em negrito, para cada ocorrência.	105
Figura 6-7 – Mapa de regras da combinação de entradas com a maior saída para a causa <i>Vegetal</i> , com o valor 0,748 (74,8%).	106
Figura 6-8 – Valores de saída pra causa vegetal, utilizando funções de pertinência gaussianas e triangulares.	109
Figura 6-9 – Análise da saída da RB para registros de causas não inferidos do filtro fuzzy.	110
Figura 6-10 – Lista mostrando eventos de Descarga Atmosférica com tempo bom.	112
Figura 6-11 – Percentual dos 730 eventos duvidosos.	113

LISTA DE TABELAS

Tabela 2-1– Principais modos de falha de chaves de proteção. (BROWN, 2002).....	29
Tabela 2-2– Condução de energia em vegetais conforme a tensão, causando falta. (BROWN, 2002).....	36
Tabela 4-1 – Tabela-verdade do tipo subjunção, cuja saída é binária.	52
Tabela 4-2 – Exemplo de probabilidade condicionais dos estados de $P(A B)$	56
Tabela 4-3 – Obtenção da probabilidade conjunta $P(A,B)$ a partir de $P(A/B)$	56
Tabela 4-4 – Obtenção de $P(A)$ a partir de $P(A,B)$	57
Tabela 5-1 – Conversão de datas em estados correspondentes da variável <i>Estação</i>	79
Tabela 5-2 – Conversão de horas em estados correspondentes da variável <i>Turno</i>	79
Tabela 5-3 – Número de estados para cada variável.	80
Tabela 5-4 – Inferência de estados da variável <i>Causa</i> pela RB.	86
Tabela 5-5 – Graduação exemplificada dos diferentes estados da variável <i>Clima</i> , com o valor do universo de discurso correspondente.	93
Tabela 5-6 – Valores críticos obtidos para cada causa na saída do sistema fuzzy.	96
Tabela 6-1 – Erro e tempo de execução dos algoritmos da RB.....	101
Tabela 6-2 – Erros na RB conforme o número de saídas em relação a mesmas entradas.....	102
Tabela 6-3 – Erros para diferentes números de variáveis de entradas em registros de dados sem repetição.	104
Tabela 6-4 – Erros para diferentes números de variáveis de entradas, em registros de dados com repetição.....	104
Tabela 6-5 – Resultados da saída do bloco fuzzy utilizando funções de pertinência Gaussianas.	107
Tabela 6-6 – Resultados da saída do bloco fuzzy utilizando funções de pertinência Triangulares.	108

Tabela 6-7 – Resultados da saída do bloco da Rede Bayesiana.	110
Tabela 6-8 – Erro do sistema integrado fuzzy / RB.	111
Tabela 6-9 – Erro do sistema integrado utilizando dados incoerentes.	113

RESUMO

A correta determinação de causas de desligamentos não programados é fundamental para as empresas de energia definirem prioridades de investimento em melhorias no sistema elétrico. Dessa maneira, é possível evitar perdas e aperfeiçoar a qualidade do serviço de fornecimento. Esta dissertação aborda um método para a identificação de causas de interrupções forçadas utilizando um sistema integrado de modelos baseados em lógica fuzzy e Redes Bayesianas (RB). Para a etapa de treino da RB e estabelecimento de regras no domínio fuzzy, apresenta-se a análise de um banco de dados tratado com técnicas de *Knowledge Discovery in Databases* (KDD), permitindo o mapeamento de variáveis de entradas e saídas dos dados. Este mapeamento compôs uma base de conhecimentos, possibilitando agrupar dados de forma a visualizar o padrão de sobrecarga coerente com um sistema elétrico. O bloco fuzzy avalia eventos com maiores recorrências, utilizando como parâmetro valores mínimos de aceitação da resposta, e o bloco RB estima as causas dos demais eventos, priorizando as respostas que se consideram com baixa incerteza. Com o método *split-half* a partir de 8.000 eventos, o sistema integrado se mostrou eficiente com o erro resultante de 5,94%. Apresentam-se outras análises visando a minimizar o erro de estimação de causas, através da ampliação do número de variáveis de entrada e o descarte de eventos com saídas múltiplas. Após o descarte, obteve-se um total de 15.994 eventos, sendo que para esta amostra, o erro final de identificação pela RB foi reduzido de 15,94% para 4,60%. Considerando 172 registros, com a inclusão de três variáveis de entrada na base de dados, o erro na RB foi reduzido de 5,81% para 1,16%. Também se verifica que o comportamento do sistema ao avaliar dados incoerentes resultou em erro de 84,38%, mostrando-se a aplicabilidade em detectar registros sem características com um sistema elétrico. Por essas análises, a metodologia proposta mostrou-se eficaz na determinação de causas de desligamentos não programados.

ABSTRACT

Correct forced outage diagnosis is fundamental for utilities to establish priorities of investment on improvements in power systems. Thereby it is possible to avoid losses and improve quality of supply service. This dissertation deals a method to diagnose forced outage faults using an integrated system of models based on fuzzy logic and bayesian networks (BN). For the training stage and rule base constitution in fuzzy domain, it is introduced the analysis of a database treated with Knowledge Discovery in Databases (KDD) techniques, allowing map input and output variables of the database. This mapping composed a knowledge base, allowing apply clusters technique on data to visualize overload pattern coherent with power systems. The fuzzy stage evaluates events with more occurrences, using minimal values of acceptance as parameter of answer. The BN part deduces other events faults, prioritizing those considered with low uncertainty. Using 8,000 events obtained by split-half method, the integrated system proved effective with resultant error of 5.94%. Other analysis were studied aiming conditions to minimize outage assess error, using extended input variables and discarding multiple outputs events. After the discard step, it was obtained 15.994 events and, for this sample, the total diagnosis error by BN was reduced from 15.94% to 4.60%. Considering 172 events, with the inclusion of three input variables, the BN error was reduced from 5.81% to 1.16%. The integrated system behavior is also analysed on diagnosis of incoherent data, resulting an error of 84.38%, showing applicable to detect outage records without electric power characteristics. Through these analyses, the proposed methodology showed effective in forced outage diagnosis.

1 INTRODUÇÃO

A distribuição da energia elétrica com o mínimo de interrupções é de importância fundamental para o desenvolvimento econômico de um país. Além de afetar sensivelmente equipamentos industriais, a falta de continuidade no fornecimento elétrico acarreta altas perdas para consumidores e concessionárias, na forma de custos muitas vezes incalculáveis.

No âmbito gerencial das empresas de energia, a Agência Nacional de Energia Elétrica (ANEEL) estabelece às concessionárias padrões e metas na forma de indicadores que definem com rigor a qualidade no fornecimento de energia elétrica. Esses indicadores medem a frequência e duração das interrupções. O não-atingimento dessas metas ocasiona em penalidades financeiras às empresas de energia, estimadas na ordem de 50 a 100 vezes maior que o preço médio do kWh faturado (BORTONI, et al., 2004). Do contrário, o cumprimento da meta garante a satisfação e fidelidade do cliente a uma concessionária em relação as suas concorrentes. Atualmente existe uma forma de minimizar os prejuízos aos consumidores e estimular a melhora de serviço, que consiste na compensação direta na fatura do cliente, caso as interrupções apuradas em mês anterior ultrapassem determinados limites (ANEEL, 2010).

No setor industrial, além da imagem das empresas e sua relação com o consumidor, para aumentar a lucro, existe o consenso da necessidade em qualificar e quantificar as causas das perdas de produção. Desse modo, é possível agir preventivamente nos processos alocando recursos financeiros somente em focos causadores dessas perdas, a fim de minimizá-las. Este conceito da qualidade também se estende ao setor de produção e distribuição de energia elétrica, sobretudo a partir de sua política de privatização, em que é exigido um sistema mais eficiente.

A fim de aumentar a confiabilidade do sistema, as causas de interrupções não programadas, isto é, forçadas, merecem grande destaque na busca da sua identificação pelas concessionárias. De natureza geralmente imprevisível, essas causas são muitas vezes provocadas por eventos externos ao sistema elétrico, tais como: descarga atmosférica, vento, interferências de animais, clima, entre outras variações. As interrupções normalmente caracterizam-se por serem permanentes e necessitam de reestabelecimento manual por parte dos eletricitistas.

Para realizar um controle a respeito de desligamentos forçados, muitas empresas de energia adotam o registro desses eventos em bancos de dados para posterior análise. No

momento do reparo, o eletricitista responsável assinala dados a respeito do sistema elétrico registrando as condições climáticas observadas no ambiente, identificando atuações de proteção, elementos defeituosos, bem como o possível agente causador do defeito.

Entretanto, a sistemática de identificação da causa por parte do eletricitista e o fluxo dessa informação para a base de dados, se tornam muitas vezes ineficientes. Idealmente essa identificação deveria ser realizada por um eletricitista experiente, tendo recebido um treinamento adequado para uma análise minuciosa da falta. Na prática, tais condições nem sempre são possíveis e o resultado final do armazenamento desses dados em estado bruto proporciona registros com diversas falhas humanas, como: o não-preenchimento de determinados campos, interrupções pouco coerentes com as condições locais observadas e atribuição de consequências de interrupções que não são as causas-raiz do desligamento. Exemplo disso ocorre quando é realizado o reparo em um ponto da rede, tendo o eletricitista que preencher a ficha de serviço indicando rapidamente o que foi verificado, a fim de atender a ocorrência emergencial de interrupção em outro segmento da rede elétrica.

Para agravar o problema, em concessionárias que não possuem um sistema de informação implementado para este fim, o processo de transposição das fichas de serviço para a base de dados também está sujeito a erros de digitação. Uma etapa de verificação de incoerências de dados obtidos nessas fichas depende do grau de conhecimento e experiência do funcionário a respeito do sistema elétrico. Assim, um sistema especialista que auxilie na verificação de causas prováveis, tanto no momento do reparo quanto no armazenamento de dados na concessionária é de fundamental importância para a análise e controle das interrupções não programadas, contribuindo para a predição e prevenção de interrupções que geram custos.

1.1 Proposta do Trabalho

Com a hipótese de que em meio a muitos registros falhos alguns sejam verossímeis para o entendimento do ocorrido, é possível restaurar a informação contida na base de dados, utilizando-a para o entendimento da maneira que ocorrem as causas e formando uma base de conhecimento para análise. Este trabalho visa ao desenvolvimento de um módulo que

acompanha a identificação de causas de desligamentos não programados, a fim de apontar o desligamento mais provável a partir de uma base de conhecimento.

Para este propósito, será utilizado um sistema fuzzy contendo a base de conhecimento, que pode ser orientada por um especialista ou extraída de uma base de dados. Neste trabalho, será obtida através da base de registros mais recorrentes de desligamentos por meio da técnica *Knowledge Discovery in Databases* (KDD). A determinação de eventos com menor recorrência será atribuído a um modelo de Redes Bayesianas (RB), sendo utilizada a mesma base de conhecimento do domínio fuzzy.

Para a admissão de uma causa como resposta, as saídas dos modelos devem atingir um nível mínimo de certeza (valor crítico de saída). Serão verificadas também as condições da base de dados para que haja a criação de um sistema integrado fuzzy-RB com a minimização do erro para a estimação de faltas. Considera-se que, para identificar o “estopim” ou causa primária do desligamento é necessário, e não suficiente, ter a disposição o maior número possível de dados de entrada capazes de informar a origem do problema de forma unívoca, ou seja, apontando somente uma saída (causa de interrupção).

1.2 Organização do Trabalho

Na seção 1.3, é realizada uma revisão bibliográfica das diferentes abordagens da detecção e identificação de faltas e o uso de técnicas de inteligência artificial relacionadas ao tema. No capítulo 2, são apresentados os diferentes tipos de desligamento não programados em sistemas de energia. Nos capítulos 3 e 4, são abordados os principais conceitos da lógica fuzzy e Redes Bayesianas, respectivamente. O capítulo 5 apresenta a metodologia proposta para o sistema de identificação de faltas. Na continuidade do método, no capítulo 6 verificam-se os resultados obtidos e suas análises. Finalmente, o capítulo 7 aborda a eficácia do método e apresenta trabalhos futuros para a pesquisa deste tema.

1.3 Revisão bibliográfica

Na literatura científica, existem diferentes abordagens referentes ao estudo de desligamentos não programados. Possivelmente, isso pode ser explicado pela constatação de

que acima de 80% das interrupções do serviço de entrega de energia se devem às falhas nas redes de distribuição (SU et al, 2007).

Na área de pesquisa a respeito de falhas em sistemas elétricos, verifica-se que há uma série de estudos aplicados a técnicas para determinar unicamente a localização da falta na rede, sem investigar a respectiva causa. A exemplo disso, em Johns (1991), dispositivos ao longo do sistema elétrico são empregados a fim de detectar o local da falta a partir da emissão de ruído gerado na rede quando ocorre a falta. Com o intuito de localizar os desligamentos mais rapidamente, no entanto, apenas busca-se o objetivo de reduzir o tempo de reparo da energia elétrica. Contudo, idealizando a redução desse tempo a zero, o que pode ser obtido através da correta identificação e eliminação das causas de desligamentos, as empresas de energia conseguiriam evidentemente obter ganhos muito maiores, evitando o surgimento de perdas e o pagamento de multas às agências reguladoras. Nesta seção, algumas referências abordando a localização de faltas como esta também serão abordadas, entretanto, somente com o intuito de discutir as técnicas de sistemas inteligentes empregadas.

Uma abordagem para melhorar a análise de faltas elétricas consiste em aperfeiçoar os registros de históricos de desligamentos. Objetivando aprimorar a qualidade dos dados coletados em campo por eletricitas de uma concessionária, em Pretto (2003, 2005) o autor descreve um sistema baseado em computador móvel PDA (*Personal Digital Assistant*). O estudo visa também minimizar o registro de eventos improváveis, após a observação de existir a tendência de a pessoa responsável pelo reparo indicar o sintoma de uma falha como sendo a causa. O dispositivo apresenta ao eletricitista um questionário com opções de entrada de dados referentes ao cenário da falta de energia. Desse modo, conseguiu-se caracterizar as condições do desligamento de maneira completa, eliminando registros com campos não preenchidos e incoerentes. Esse sistema também teve como consequência a padronização do registro, o que possibilitou a criação de um aplicativo para realizar uma estimativa a respeito da provável causa do desligamento ou eliminação de causas incoerentes. Para classificar as causas, o autor sugere duas metodologias a partir dos dados tratados: uma baseada em Árvores de Decisão e outra em Matriz de Pertinência. Apesar de os métodos precisarem de um trabalho mais específico para a determinação de causas, já foi percebido o alto ganho que a empresa concessionária obteve somente com um sistema que direciona os dados de modo a extrair informação. Com a utilização do computador móvel, também foi possível realizar o cruzamento dessas informações.

Redes Neurais Artificiais são utilizadas em Chow (1993) para identificar as causas de faltas devido ao contato de animais em sistemas de distribuição, baseando-se em informações do histórico de registro, como: localização geográfica, clima e número de fases do circuito afetadas. A metodologia envolve a escolha das variáveis apropriadas para a entrada no sistema. Modelos com RNA tipicamente fornecem um grande grau de robustez ou tolerância a erros porque possuem muitas unidades de processamento. Além disso, são capazes de se adaptarem e aprenderem com os dados para melhorar sua performance ao longo do processo. Nesse trabalho, após experimentos e análises, os dados de registro foram pré-processados com a finalidade de se obter as probabilidades condicionais dos eventos. Essas probabilidades foram utilizadas como entrada no sistema com RNA, obtendo-se 99% de certeza em reconhecer as causas nos dados de treino e 98% de acerto quando dados desconhecidos foram apresentados. Nesse sentido, o uso das probabilidades condicionais como entrada ao sistema surge como uma proposta interessante, uma vez que alia a capacidade de modelagem das RN's com a natureza estatística da ocorrência de eventos.

A mineração de dados é outro método muito utilizado para construir conhecimentos a partir de registros de dados. De modo geral, consiste na extração automática de informação oculta em bancos de dados. Uma aplicação desta técnica pode ser conferida em Peng (2004). Esse estudo apresenta a teoria dos Conjuntos Aproximados como uma ferramenta de mineração de dados para a representação do conhecimento. O artigo mostra que, através da classificação de dados representativos, é possível encontrar padrões úteis e regras para o diagnóstico e localização de faltas. Por meio de um registro histórico das faltas em alimentadores da Taiwan Power Company, uma análise de mineração foi realizada com os dados, extraindo-se informações das relações entre as variáveis. O sistema foi validado satisfatoriamente com dados de teste. Mostra, com isso, a importância do pré-processamento de um registro histórico, na medida em que filtra dados não significativos para a representação do conhecimento.

Outra abordagem de mineração de dados pode ser verificada em Dola (2005). Nesse trabalho, foi realizada uma análise de faltas elétricas ocasionadas pelo acionamento de chaves de proteção após sobrecorrente. Com os dados de tensão e corrente de nove subestações, foi constatado que a principal causa foi devido ao tipo de curto-circuito (fase-terra). Utilizando-se o software SEL, as correntes e tensões foram observadas e comparadas em gráficos através do software Excel®. A análise possibilitou identificar que as chaves são acionadas devido a

outros dez tipos de causas de sobrecorrente estabelecidas. Os registros redundantes e incoerentes foram suprimidos e registros que não continham valores foram deduzidos através da média global e moda. Para as inferências de causas, foi utilizado um modelo através de Redes Neurais e outro por meio de Árvores de Decisão, através do software WEKA®. Para determinar os parâmetros, ambos modelos foram treinados utilizando a técnica de aprendizado supervisionado, em que recebem um estímulo externo para a identificação da melhor resposta. Ao contrário do aprendizado supervisionado, o não supervisionado não depende de uma avaliação externa, mas das informações constantes na base de conhecimento (TECH, 2006). A partir dos resultados, pôde-se notar que o modelo com Redes Neurais apresentou melhor performance, com exatidão de 93,17%, ao passo que através de Árvores de Decisão a exatidão foi de 91,47%. A análise de mineração de dados permitiu um panorama geral sobre as faltas elétricas, contribuindo para realizar a modelagem dos dados. Além disso, o algoritmo utilizado pelo software WEKA® possibilita a determinação de regras que regem, por meio das variáveis responsáveis, o surgimento de faltas. Essas regras são inteligíveis ao ser humano, o que possibilitaria abstrair conhecimento de como ocorrem as faltas e o que as propiciam. Embora a autora não tenha considerado este último aspecto, foi discutida uma análise comparativa entre as RN's e Árvores de Decisão, concluindo que o sistema neural é mais adequado ao problema.

Em Tso (2004), a fim de detectar o grau de sensibilidade das subestações às variações de distúrbios, o autor também utiliza técnicas de mineração de dados para extrair informações do banco de dados de um sistema de energia em Hong Kong. Por meio de uma análise estatística, são determinados índices de correlação a partir das medidas de carga (tensão) do transformador. Posteriormente, os registros são ordenados formando uma lista de categorias de distúrbios nas subestações, seguindo de forma decrescente aos índices estatísticos. Através de uma lista com os dez maiores fatores de correlação, em geral, é inferido que as subestações com maiores índices indicam estatisticamente maior sensibilidade a distúrbios. Este conhecimento extraído estava de acordo com a experiência comum de engenheiros a cerca da sensibilidade de determinadas subestações. Essa análise mostra-se útil para alertar operadores do sistema onde é necessário guarnecer maior suporte de fornecimento de energia em determinadas áreas. Por meio desta técnica, o autor mostra a possibilidade de extrair-se conhecimento de registros através de um ranking de ocorrências, onde os maiores índices apresentam maiores tendências estatísticas de prevalecerem. Analogamente, verifica-

se que esse conceito pode ser aplicado neste trabalho, substituindo-se os determinados tipos de ocorrências de redução da carga por ocorrências de falta elétrica, seguindo ordem decrescente de algum parâmetro estatístico. Assim, pode-se obter um conjunto de causas e suas variáveis com possibilidade mais significativa de acontecerem.

Com base em um registro de faltas de energia da companhia Duke Energy, contendo seis tipos de dados referentes às condições locais para indicar um desligamento que fora atribuído, no trabalho de Xu (2006) utilizam-se dois métodos de classificação para o diagnóstico das faltas: Regressão Logística (RL) e Redes Neurais Artificiais (RNA). As causas consideradas foram *Vegetal* e *Animal*. O sistema consiste de dois modelos em paralelo: um ramo para classificar os eventos em classes do tipo *Animal* ou *Não-animal* e outro para *Vegetal* ou *Não-vegetal*. A fim de comparar os métodos, os modelos foram testados com as técnicas de RNA e RL. A saída de cada um dos modelos consiste na probabilidade em que a causa assinalada pertença à respectiva classe. Essas saídas são comparadas por um módulo de decisão. Eventualmente era esperado o surgimento de um conflito, quando um ramo indicava a ocorrência de *Vegetal* ao mesmo tempo em que outro indicava *Animal*. Nesses casos, em que totalizaram apenas 0,1% dos eventos, as probabilidades eram comparadas com um valor de referência, sendo estabelecido certo grau que indicava a possibilidade de ocorrência de cada causa. O ramo com maior grau era efetivamente considerado como a causa correspondente. Através de uma análise dos gráficos para a falha do tipo *Animal*, percebe-se que ambos os modelos apresentaram imprecisão quanto às diferentes regiões, mas o modelo com RNA foi ligeiramente mais preciso que a RL. O autor relata que, para dados desbalanceados, designados por registros que são constituídos por grande quantidade de uma classe em relação a outra, o modelo RL é tendencioso para a maioria dos dados do tipo *Não-animal* e *Não-vegetal*. Esse trabalho, portanto, levanta aspectos importantes que podem afetar os resultados. O autor mostra que uma grande desproporção entre as categorias representativas das diferentes causas deve ser levada em consideração no momento de validar um classificador. Contudo, essa pesquisa não aborda a determinação de outros tipos de causas concorrentes, além das supracitadas. Isso inviabiliza afirmar que o sistema de fato classifica de forma satisfatória, uma vez que, na realidade, as outras causas também são relevantes e devem ser consideradas para determinar qual se manifesta preponderante em cada evento.

Um método de inteligência artificial muito bem aceito baseia-se na lógica fuzzy. É designado para incorporar incertezas de dados do mundo real, possuindo a capacidade de

representar a habilidade da mente humana por meio de raciocínios mais aproximados que exatos. As aplicações da lógica fuzzy datam a partir de 1980, como em um forno para cimento de F.L. Smith, e no sistema de controle de metrô na cidade japonesa de Sendai (ZADEH, 1994). Em 1987, surgiu o primeiro chuveiro controlado por lógica fuzzy, seguido pela primeira máquina de lavar baseada nessa lógica, em 1989, por Matsuhita (ZADEH, 1994). Um conceito-chave é o de gráfico fuzzy, apresentado em meados de 1971-1974. A ideia central consiste no fato de que qualquer tipo de função ou relação podem ser representados por um gráfico fuzzy. Com isso, é possível aproximar, por exemplo, distribuições de probabilidade e funções de pertinência, das quais desempenham papel importante em decisões qualitativas e de diagnóstico de faltas.

Uma aplicação da lógica fuzzy pode ser conferida no artigo de Xu (2007). Utiliza-se o algoritmo-E na tentativa de aliviar o efeito de dados desbalanceados. Foram relacionadas três causas para o estudo: *vegetal*, *animal* e *descarga atmosférica*. O algoritmo-E é capaz de extrair um conjunto de regras de inferência, utilizando a informação estatística dos dados, por meio da lógica fuzzy. Desse modo, um classificador fuzzy pôde ser desenvolvido mesmo com um domínio de conhecimento pequeno representado pelos dados. O artigo apresenta ainda duas medidas para a construção das regras do sistema fuzzy, chamadas *suporte* e *serviço*. Segundo o autor, a normalização dessas medidas ocasiona no abrandamento do efeito de dados desequilibrados. A performance do algoritmo-E foi comparada novamente com a de um sistema baseado em RNA, mostrando-se que esse algoritmo apresenta performance superior quando os dados são desequilibrados, mas obtém desempenho pior quando os dados não apresentam esta característica. Portanto, esse trabalho mostra uma desvantagem neste método, que somente apresenta confiabilidade se os dados são desbalanceados, o que na realidade nem sempre ocorre para determinadas causas. Apesar disso, mostra-se também que o algoritmo-E pode ser utilizado para ajudar a explicar sobre os mecanismos da identificação de causas através das regras fuzzy extraídas.

Desenvolvimentos recentes de sistemas AMR (Automated Meter Reading) e o aumento do uso do SCADA® em sistemas de distribuição vêm fornecendo mais medições sobre o sistema de distribuição, contribuindo para obter-se informações a respeito de desligamentos. Contudo, devido à complexidade do sistema de distribuição, nenhuma dessas fontes de dados estão isentas a erros. Geralmente, a informação de fontes múltiplas aumentam a probabilidade de dados conflitantes, falhos e redundantes. Em Liu (2002), utilizando

registro de dados combinados de AMRs, Trouble Calls e sistema SCADA, a referência adota um sistema inteligente, por meio de filtro fuzzy, para inferir a ocorrência de desligamentos através desses registros integrados. Na etapa de validação, assegura-se que todas as informações coletadas contenham somente novos desligamentos e é estabelecida uma ligação entre os dados das diferentes fontes de acordo com suas relações. Essa etapa gera uma lista de faltas que são candidatas a ocorrerem. A partir desta lista, utiliza-se um sistema com filtro fuzzy como “motor de inferência” para inferir na veracidade da falta. Através de regras heurísticas que definem as relações entre as fontes de dados de forma difusa, o sistema fuzzy atribui a possibilidade percentual de incerteza de determinado desligamento. As saídas desse sistema, com incerteza baixa (menor que 50%), são efetivamente consumadas como desligamentos, as demais são descartadas. O sistema foi testado em mais de dez casos, cobrindo entrada de dados redundantes, conflitantes, dados do AMR contendo erros e outras condições possíveis. Após o filtro de dados, o tamanho da lista original foi reduzida de 20% à 30%, diminuindo a quantidade e aumentando a qualidade das informações para a tomada de decisão.

Outra aplicação da lógica fuzzy pode ser verificada em Shridharan (2001). A fim de contornar o erro existente no sinal do medidor AMR para utilizar na detecção de faltas transitórias, foi construído um sistema que modela a incerteza do medidor utilizando-se a lógica fuzzy. Uma vez que o tipo de causa das faltas transitórias depende do instante em que são observadas, os desligamentos puderam ser classificados a partir de conjuntos fuzzy que relacionam o tempo em que foi detectada a falta com a possibilidade de determinada causa ter ocorrido em termos semânticos equivalentes a: *Baixo, Médio e Alto*. Foi constatado que os dados filtrados usando este modelo apresentaram maior qualidade em relação aos dados sem a utilização do filtro.

Com o uso de regras inspiradas no modo de pensar humano, por meio de termos semânticos é possível formar uma base de conhecimento para a modelagem de sistemas inteligentes. Os trabalhos de Liu (2002) e Shridharan (2001) mostram que os sistemas fuzzy apresentam-se bastante eficazes para solucionar problemas com diversas incertezas envolvidas. Sendo o diagnóstico de faltas elétricas com a utilização de variáveis de natureza incertas e dados incompletos, percebe-se que a lógica fuzzy é destacada pela grande aplicabilidade na resolução desses problemas.

Outra forma utilizada para representar modelos de situações cujas variáveis contenham incertezas é verificada com o emprego de sistemas de Redes Bayesianas, que se baseiam na causalidade de eventos condicionalmente dependentes. Essas redes são constituídas por estruturas computacionais chamadas grafos, em que as variáveis de interesse formam nós interligados relacionando as dependências das mesmas. Esses nós agregam informações da distribuição de probabilidades do conjunto das variáveis. A capacidade de instanciar-se a probabilidade de ocorrência de determinadas variáveis, visto que ocorreram outras, pode ser empregada para a determinação de causas de desligamentos não programados.

No artigo de Chien (2002) foi desenvolvida uma RB baseada em conhecimento de especialistas e registro de dados (767 eventos), referentes ao diagnóstico de faltas em redes de distribuição da companhia Taipower. Os conceitos de evidências simples e múltiplas foram pormenorizados por meio de *odd ratios*, que são representações de probabilidades. As causas estudadas foram: *chuva, degradação natural do isolamento do transformador e contato por objetos*. Na análise dos dados, foi montado um grupo de variáveis dependentes e itens sem associação foram eliminados, para fins de simplificação. O objetivo foi relacionar a falha em equipamentos com as condições responsáveis pela falha observadas no ambiente. Essa análise serviu para que fossem determinadas as probabilidades a priori das variáveis da RB. As probabilidades a posteriori foram estimadas através de questionários para especialistas ponderarem valores percentuais referentes a pares de variáveis diretamente dependentes. Esses conjuntos de probabilidades, que representam a parte quantitativa da rede, formam a base de conhecimento do sistema. A partir disso, qualquer hipótese de probabilidade condicional de um ou mais eventos observados pôde ser calculada satisfatoriamente, utilizando formulações desenvolvidas empiricamente. Nesse artigo verifica-se, no momento da construção da rede bayesiana, uma maneira para estabelecer as probabilidades a posteriori que cada nó da rede representa. A estimação desses parâmetros, considerando todas as variáveis ao mesmo tempo, torna-se muitas vezes uma tarefa bastante difícil e complexa para o especialista determinar. Percebe-se que através da utilização dos questionários, é possível realizar a decomposição de um julgamento complexo em diversos julgamentos simples de eventos tomados dois a dois.

Outra forma de estabelecer os parâmetros da rede bayesiana é apresentado em Nassar (2004). Na pesquisa, o sistema probabilístico foi desenvolvido através da associação

entre condições climáticas e o número de atendimento emergencial para restauração do desligamento. O estudo teve a finalidade de criar um modelo para prever o número de faltas elétricas a partir das condições climáticas. O artigo apresentou um método de identificação de padrões climáticos mensais por meio de registros através de técnicas estatísticas e organizando agrupamentos de dados com características em comum, chamados *clusters*. Foi investigada a relação desses padrões com o número de paradas não programadas, possibilitando extrair-se a base de conhecimento necessária para determinar a parte quantitativa da RB. Com isso, verifica-se uma aplicação das técnicas de mineração de dados para compor um sistema inteligente, determinando os parâmetros que caracterizam o modelo a partir de registros que contêm informações implícitas.

Para determinar a causa de falha em transformadores, em Wu (2009) é empregado um conjunto de multi-classificadores bayesianos em paralelo (*Naïve Bayesian classifier*, *Selective Bayesian classifier*, *Tree Augmented Naïve Bayesian*, *BN Augmented Naïve Bayesian classifier* e *General Bayesian Network*), agindo como um sistema de pré-processamento para outro classificador também baseado na teoria estatística, chamado SVM (*Support Vector Machine*). Os diagnosticadores bayesianos são combinados através de determinados pesos obtidos por multiplicadores de Lagrange. O algoritmo SVM assegura que a solução encontrada seja a ótima, por fundamentar-se no conceito de otimização secundária convexa. Resultados mostraram que a taxa de acertos de diagnósticos deste modelo foi maior do que utilizando um modelo simplificado do classificador. Essa técnica apresenta o conceito interessante de combinar os sistemas de classificação, explorando as melhores vantagens de cada método para atingir resultados mais apurados.

No artigo de Yongli (2006), um sistema baseado em conhecimento utilizando três modelos de Redes Bayesianas simplificadas (representando: linhas de transmissão, barras de alimentação e transformadores), foi desenvolvido para determinar se em alguma seção de um sistema de energia ocorreu *blackout*. Utilizando as técnicas de *Noisy-Or* e *Noisy-and* para os modelos, os parâmetros foram estabelecidos por um algoritmo de aprendizado análogo ao *back propagation* de rede neural. Apesar de o método apresentar bons resultados com baixo erro e exigir pouco processamento/armazenamento de dados, a Rede Bayesiana simplificada proposta realiza a inferência para apenas um tipo de falta: o acionamento de chaves fusíveis. Desse modo, os estados das variáveis consistem simplesmente em *ligado* e *desligado*. De maneira geral, esses modelos aparentam ser boas representações das relações de acionamentos

entre chaves de proteção secundária, primária e terciária de cada elemento do sistema. Entretanto, um sistema multi-faltas utilizando o mesmo princípio para determinar o motivo das faltas, ao invés da localização, seria bastante complexo. Não apenas devido à carência de modelos *noisy-and / noisy-or* que retratem de modo adequado os desligamentos, mas também pela complexidade de desenvolver um conhecimento especialista abrangendo as relações entre várias causas ao mesmo tempo, sem a possibilidade de as tratar de modo simples, como as chaves abertas ou fechadas do artigo.

A partir dos estudos de Tronchoni (2007, 2008) são apresentados o desenvolvimento de um sistema de reconhecimento de faltas, aperfeiçoando o módulo de inferências do trabalho realizado por Pretto (2004). O sistema consiste em um classificador com estrutura do tipo *Naïve-bayes* para identificar causas de desligamentos não programados, através de registros de faltas de uma concessionária. Primeiramente, foi utilizado um banco de dados bruto, que não havia ainda passado pelo pré-tratamento idealizado por Pretto (2004). Segundo Tronchoni (2008), os bancos de dados das empresas concessionárias tendem a ser inconsistentes e conter muitas incertezas, o que é devido à inexistência de uma metodologia adequada para a aquisição dos dados referentes ao desligamento. O registro apresentava diversos erros e incoerências, aparentando ser impossível a sua utilização. Entretanto, técnicas de extração de conhecimento - como KDD (*Knowledge Discovery in Databases*) - foram empregadas para explicitar informações, o que possibilitou a utilização dos registros tornando-os mais adequados para atingir os objetivos do trabalho. Posteriormente, a Rede Bayesiana foi criada utilizando a informação desse histórico de eventos tratados para inferir causas, dado um conjunto de condições observadas no ato do reparo de energia. Após a modelagem da estrutura da rede, seus parâmetros foram estabelecidos a partir da aprendizagem de dados de treino por meio do algoritmo *Expectation Maximization*. O sistema foi validado por meio do método *split-half*, apresentando desempenho com erro de 7%, mostrando-se superior a um sistema de Redes Neurais. A grande vantagem do uso de Redes Bayesianas se deve ao fato de o modelo considerar, simultaneamente, as incertezas do problema e evidências incompletas.

2 IDENTIFICAÇÃO DE CAUSAS

O entendimento dos fatores que caracterizam cada desligamento em sistemas de distribuição é de fundamental importância para ser analisado de que maneira é possível prever, registrar e evitar esses acontecimentos futuramente. Devido ao caráter aleatório desses eventos, as causas de desligamento podem ser originadas por diversas razões. Entretanto, fazendo-se uma análise histórica das falhas, existem padrões de interrupção que nos permitem classificá-las nos principais grupos que afetam de forma contundente os índices de confiabilidade de uma concessionária. Conforme Brown (2002), entre esses grupos destacam-se: falhas em equipamento, animais e árvores em contato com a rede, efeitos do clima adverso e interferência humana. Identificando esses problemas como causa raiz, é possível então endereçar recursos de forma eficaz e econômica no intuito de reduzir as interrupções de energia. A seguir serão abordados os modos de falhas causadas por: falhas em equipamento, animais, clima adverso, árvores e interferência humana.

2.1 Falhas em equipamento

Uma vez que determinado equipamento foi adquirido sem defeitos de fábrica e foi corretamente instalado, a falha pode estar relacionada a fatores como: exposição a níveis de tensão e corrente extremos, desgaste térmico/mecânico, término de vida-útil, entre outros. Nas seções seguintes comentam-se as principais características de falha que afetam de maneira significativa os seguintes equipamentos: transformadores, linhas de distribuição, chaves, pára-raios e isoladores.

2.1.1 Transformadores

Transformadores são equipamentos suscetíveis a falhas no momento em que são exigidos a trabalhar com potência além da especificada (sobrecarga). Quando ocorre alguma falha em um transformador, geralmente outros transformadores da rede podem ser solicitados a atender à demanda de energia não suprida pelo equipamento defeituoso, evitando uma

interrupção. Entretanto, esta sobrecarga emergencial nos transformadores vizinhos, pode resultar em futuras interrupções, devido à significativa redução da sua vida-útil.

A vida de um transformador pode ser caracterizada como o tempo necessário para que o seu material isolante perca 50% de suas características físicas devido ao aquecimento. Essa relação entre calor e colapso do transformador ocorre de forma exponencial (BROWN, 2002).

A falha mais incisiva nesses equipamentos, contudo, ocorre se eventualmente surgir um curto-circuito na rede do lado secundário do transformador. A sobrecorrente passa pelo transformador e faz vibrar os enrolamentos com uma força mecânica proporcional ao quadrado da magnitude dessa corrente (BROWN, 2002). Se o isolante do transformador já estiver quebradiço devido à exposição de sobreaquecimento, ocorre uma ruptura desse material, acarretando em uma falha interna do transformador.

Vários estudos têm sido realizados a fim de avaliar as características de falha de transformadores, de modo a determinar o estado de funcionamento do equipamento e identificar as predisposições que levam a essas falhas. Alguns desses métodos são: inspeção visual e sonora, testes de carga térmicos, testes de fator de potência, análise de gases dissolvidos (DGA), entre outros. Alguns desses testes podem ser monitorados a distância, podendo indicar um alarme ao operador da concessionária quando o transformador atingir condições impróprias para o funcionamento. Entretanto, em transformadores de sistemas de distribuição essa medida é anti-econômica.

2.1.2 Redes de distribuição

No Brasil, a maioria das redes de distribuição na região são compostas de cabos nus. Sem isolamento, estas redes ficam expostas a qualquer tipo de interferência ambiental, sujeitas a intempéries, animais e vegetação. Se esses fatores não provocarem fisicamente uma interrupção de energia, é possível que valores altos de correntes tenham causado aquecimento por efeito Joule nos condutores, deixando-os mais arqueados pela excessiva folga causada por dilatação térmica. Esse efeito faz aumentar a possibilidade de haver contato entre os condutores, sucedendo a interrupção. O não funcionamento da proteção pode ainda causar o derretimento dos condutores com sua consequente queda.

Outros problemas não são decorrentes da rede propriamente dita, mas do mau funcionamento de elementos que acompanham a distribuição, como: conexões, chaves, pára-raios, capacitores e reguladores. Ainda incluem os elementos não-energizados: cruzetas e postes.

Para avaliar o estado da rede e evitar futuras interrupções forçadas pela queda de postes danificados, por exemplo, utiliza-se a técnica de bater com uma marreta na base do poste de madeira em busca de um som oco, indicando que o mesmo foi deteriorado pela umidade. Outros métodos utilizados para inspecionar as redes de distribuição incluem:

- inspeção visual;
- interferência de rádio-frequência;
- inspeção infravermelha;
- inspeção aérea;
- teste de comutação de chaves.

No entanto, devido à grande extensão das redes, é impraticável realizar com êxito essas inspeções em todos os elementos para impedir um desligamento. A alternativa é que exista um plano de manutenção preventiva indicando os elementos com maior probabilidade de falha para que sejam o foco da inspeção.

2.1.3 Chaves

O mau funcionamento de uma chave ocorre quando ela espontaneamente abre quando deveria estar fechada ou falha no momento em que deveria comutar. Esses eventos aleatórios devem-se principalmente: à falta de coordenação dos dispositivos de proteção, defeitos com os relés associados às chaves, atuadores descarregados ou emperrados e devido ao rompimento dielétrico de modo semelhante ao do transformador. Com o frequente uso, ocorre ainda uma perda do material por evaporação devido ao arco elétrico gerado, resultando no desgaste da chave. A Tabela 2-1 indica os principais modos de falha desse componente.

Tabela 2-1– Principais modos de falha de chaves de proteção. (BROWN, 2002)

Modo de Falha	% de Falha
Abrir quando não deveria	42
Falha em serviço (exceto abertura ou fechamento)	32
Falha na abertura	9
Danificação enquanto abre corretamente	7
Falha ao fechar	5
Danificação enquanto fecha	2
Falha durante teste ou manutenção	1
Danificação descoberta durante teste ou manutenção	1
Outros	1
Total	100

O desgaste das chaves pode ser estimado através de equações que relacionam valores monitorados de corrente, tempo de abertura, tempo de fechamento e velocidade do contato, com a vida útil do componente (BROWN, 2002).

2.1.4 Pára-raios

Existem, basicamente, duas variedades de pára-raios: carboneto de silicone e varistor de óxido metálico (MOV). O primeiro consiste em uma camada de ar entre dois condutores. Para evitar que as sobrecorrentes percorram a rede durante a operação normal, a corrente excessiva atravessa a camada de ar por meio de um arco elétrico, sendo então desviada para o solo. Este tipo de pára-raios está sujeito a falhas devido ao espaço de ar. Eventualmente pode haver infiltração de umidade, suportando menores valores de sobretensão. Com a presença indesejada de água, pode ocorrer ainda a expansão térmica através do aquecimento, ocasionando em falhas em operações normais ou em casos de sobretensão.

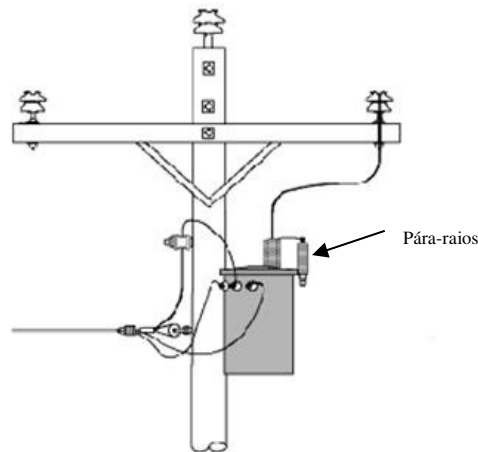


Figura 2-1– Pára-raios no sistema elétrico. (OSHA, 2010)

Os dispositivos do tipo MOV são de tecnologia mais recente e apresentam menos falhas que o anterior, possuindo uma resistência não-linear. A partir de certo valor de tensão, a resistência cai abruptamente, drenando e impedindo a sobretensão na rede. De maneira geral, ambos os dispositivos podem sofrer rupturas devido a um stress por aquecimento. Correntes localizadas de forma concentrada e valores muito altos de corrente resultam em expansão térmica não uniforme do componente.

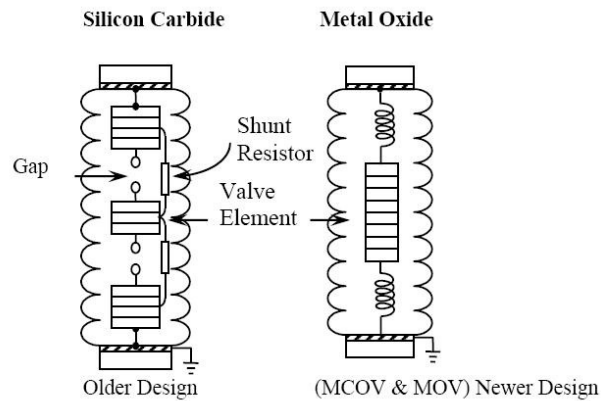


Figura 2-2 – Pára-raios do tipo carboneto de silicone e varistor de óxido metálico. (DOBLE, 2010)

2.1.5 Isoladores e conexões

As falhas dos isoladores estão associadas ao rompimento dielétrico desses materiais, que podem ser de vidro, porcelana e polímero. Esse efeito pode gerar pequenas correntes elétricas que fazem o dispositivo perder sua função de isolar eletricamente. A tendência dessas correntes é se extinguirem ou formarem um arco, o que ocasiona a interrupção. A perda da capacidade de isolamento decorre quando esses elementos são expostos a contaminações, como sal, fertilizantes e poluição industrial.

Estudos em laboratório mostram que os efeitos dessas contaminações se agravam quando combinados com a umidade. Entretanto, a maneira mais plausível de evitar esses desligamentos seria através da retirada do acúmulo de dejetos por meio de lavagens periódicas desses componentes. Além da inviabilidade econômica que existe mesmo com uma manutenção preventiva de forma planejada, existem preocupações do impacto desta ação, de um modo geral. Contudo, avanços tecnológicos na área de materiais, como a nanotecnologia, talvez possam trazer bons resultados a longo prazo.

2.2 Animais

Existem os mais diversos motivos de contatos de animais com a rede que causam faltas elétricas, e isto depende não só dos comportamentos e habitats característicos de cada espécie, quanto da região ou país em que os desligamentos serão analisados. Entretanto, é razoável esperar que a incidência dessas causas por animais de grande porte afetem mais sistemas em áreas rurais, com maior ocorrência em sistemas de transmissão do que de distribuição. É importante afirmar que os pássaros são os animais com maior preponderância de interagir com a rede elétrica, já que os condutores naturalmente passam a fazer parte do habitat aéreo das aves.

Diferentes tipos de pássaros causam diferentes tipos de problemas. Alguns constroem ninhos em treliças de torres, postes e subestações, interferindo na rede e atraindo a presença de predadores. A remoção desses lares de forma permanente é difícil, uma vez que seus instintos os levam para a reconstrução dos ninhos no mesmo lugar. O número de espécies de aves no Brasil é amplo, sendo o terceiro país em diversidade (SAVE, 2010). Entretanto,

algumas espécies em particular, como aves de rapina e pica-paus, tem maiores incidências em países do hemisfério norte, onde causam maiores transtornos. Os primeiros podem ocasionar falhas devido à grande largura das asas e os últimos pelos danos causados em postes de madeira. Em geral, no entanto, todas as espécies utilizam a malha elétrica para descansar e na busca por comida. As falhas ocorrem por meio da ponte que provocam ao fazerem contato das asas com os condutores, bem como através da contaminação pelos excrementos deixados. É possível evitar a incidência dessas aves com dispositivos geometricamente apropriados que impeçam o repouso, como um cone na parte superior do poste ou isolantes em pontos estratégicos, conforme a Figura 2-3. Também existe a possibilidade da utilização de condutores com isolamento parcial na extensão da rede (conhecida como Rede Compacta). Contudo, essa solução está bastante condicionada à relação custo-benefício.

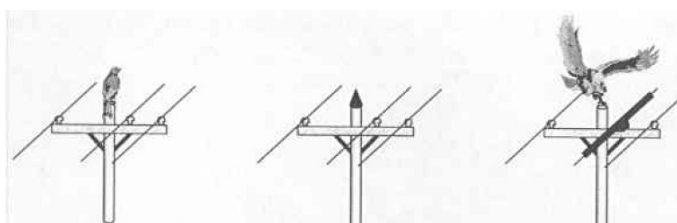


Figura 2-3– Anteparos para impedir a interferência de animais. (BROWN, 2002)

2.3 Clima Adverso

Enquadram-se neste grupo as intempéries causadas por ventos, descargas atmosféricas e ondas de calor. Pode-se observar que essas condições climáticas afetam a confiabilidade da rede de maneira muito significativa. Em condições ideais de clima, a falha em um dispositivo ocorre de forma independente de outro equipamento. Já em sistemas elétricos sob clima severo, verifica-se que a falha nesses equipamentos ocorre simultaneamente, resultando em desligamentos de várias zonas da rede e oferecendo grande dificuldade para as concessionárias realizarem o extenso reparo. Em novembro de 2009, uma companhia de energia elétrica do Rio Grande do Sul (CEEE) registrou um dia de temporal forte, causando 1.017 postes quebrados. Isto resultou em perdas de aproximadamente R\$ 3 milhões naquele mês. Cada poste demandou, em média, 3 horas de trabalho, realizadas por 70 equipes com caminhões, trabalhando 12 horas por dia (ZEROHORA, 2009).

2.3.1 Vento

Os desligamentos causados por vento estão muitas vezes correlacionados com a região em que a rede se encontra, podendo estar sujeita a turbilhões ou túneis de vento e ainda estar perto de vegetações, que contribuam para a interrupção. A pressão exercida nas árvores e postes é proporcional ao quadrado da velocidade do vento, aumentando a probabilidade de desligamento (BROWN, 2002). Além da velocidade, o grau de influência dos ventos está associado à frequência e intensidade das rajadas, direção e extensão da tormenta.

Conforme Brown (2002), a corrente de ar induz ainda diversos tipos de movimento nos condutores que podem afetar a confiabilidade, denominados:

- *balanço*: se refere ao movimento pendular que, tendo uma amplitude grande o suficiente, pode fazer com que um condutor encoste em outro. Esse efeito pode ser reduzido se os cabos forem instalados com um espaçamento e tensionamento adequados;
- *galopada*: ocorre quando gelo se deposita em torres e condutores, criando uma superfície aerodinâmica que possibilita que o vento exerça uma força vertical para baixo. Através de amortecedores e aumentando o tensionamento nos condutores é possível reduzir esse fenômeno;
- *vibração eólica*: ao passar pelo entorno dos condutores, o vento pode criar vórtices com uma frequência que coincida com a frequência natural dos cabos pendurados. Esse efeito gera vibrações mecânicas que acarretam em fadiga nas conexões, causando futuras interrupções. O problema pode ser mitigado diminuindo o tensionamento nos condutores e colocando amortecedores para alterar a frequência natural.

2.3.2 Descarga atmosférica

A corrente de uma descarga atmosférica é da ordem de 30.000A, muitas vezes consistindo de múltiplas descargas com a duração de frações de segundo. Esse fenômeno pode atingir o sistema elétrico diretamente ou indiretamente (através de tensão induzida, após o raio atingir uma árvore ou construção na proximidade). A sobretensão da descarga atmosférica origina inicialmente uma centelha, causando um curto-circuito que passa a gerar

um arco, conduzindo não mais a energia do raio, mas a do sistema elétrico. Conforme Brown (2002), o número de descargas atmosféricas pode ser estimado por meio da equação (2-1).

$$N = \frac{(28 \cdot h^{0,6} + w) \cdot k^{1,25}}{40234} \quad (2-1)$$

onde,

N = descargas atmosféricas (quilômetro/ano)

h = altura do poste (m)

w = largura da cruzeta (m)

k = nível ceraunico (dias de tempestade por ano)

Segundo Brown (2002), a descarga atmosférica que atinge diretamente a rede ocorre com menor frequência. Ainda assim, para evitar os efeitos dessa incidência, a concessionária deve instalar cabos com a função de proteção e pára-raios ao longo de cada poste. Entretanto, devido à grande extensão da rede, é quase impossível evitar, a partir da tensão elevada proveniente da descarga direta, o surgimento da centelha causadora do curto-circuito.

No caso da descarga induzida, as magnitudes de tensão são bem menores e os cabos e pára-raios são geralmente suficientes para anular o efeito, evitando o desligamento.

2.3.3 Climas quentes

Há regiões com climas que apresentam ondas de calor. São períodos do ano extremamente quentes, fazendo com que a população aumente o consumo de energia de forma significativa para arrefecer ambientes residenciais e comerciais. Quando esse consumo não for considerado durante o planejamento do sistema de distribuição, ocorre sobrecarga de condutores e transformadores, que não conseguem dissipar adequadamente o calor excessivo para a atmosfera aquecida. Conforme colocado anteriormente, a sobrecarga em um transformador acarreta na redução de sua vida-útil e muitas vezes na dos transformadores vizinhos, favorecendo as interrupções.

Uma maneira de mitigar esses desligamentos seria superdimensionar os transformadores e condutores, o que é inviável economicamente. Desse modo, a solução mais eficaz é por meio de periódicas previsões de carga, sempre garantindo que a capacidade dos

equipamentos instalados supra a futura demanda em dias quentes. Em alguns estados do Brasil, utiliza-se ainda o artifício de adiantar o horário em uma hora na temporada de verão. Segundo referências da CEEE (2009), é estimada uma redução de 4,7% na demanda de energia em horários de maior consumo (horário de “pico”), que ocorre entre as 18 e 21 horas, possibilitando economia de 2.000MW a cada ano.

2.4 Árvores

Árvores em contato com a rede são, seguramente, uma das causas de maior incidência para a maioria das concessionárias. Por possuírem certa resistência elétrica, ramos de árvores friccionando apenas uma fase, tipicamente não ocasionam em desligamentos. No entanto, as maneiras mais comuns de os vegetais afetarem a rede são:

- choques mecânicos de troncos e ramos nos condutores, causando danos e fragilizando a estrutura;
- crescimento de ramos fazendo dois condutores se encostarem;
- movimento de ramos, causado pelo vento, fazendo dois condutores se encostarem;
- estabelecimento de curto-circuito entre dois condutores por ramo através de dois condutores.

De certo modo, o curto-circuito de um ramo servindo como ponte entre dois condutores nem sempre ocorre e, quando acontece, não é de forma imediata. Para haver a condução de energia, é necessário que o vegetal não seja muito seco e nem muito verde e que a passagem de uma pequena corrente não se extinga, formando um caminho de carbono para conduzir de forma efetiva a corrente. Diâmetros maiores de ramificações de árvore apresentam menor resistência elétrica. O problema se acentua durante tempestades, em que os galhos molhados provocam curtos-circuitos ao tocarem simultaneamente dois condutores, impulsionados pelo vento excessivo.

Além disso, o gradiente de tensão é um fator decisivo para a condutividade no vegetal. A Tabela 2-2 apresenta os níveis de tensão em sistemas de distribuição necessários para que haja incidência de falta.

Tabela 2-2– Condução de energia em vegetais conforme a tensão, causando falta. (BROWN, 2002)

Nível de Tensão	Incidência da Condutividade Elétrica no Vegetal
5kV	~ 0%
15kV	80%
35kV	100%

Desse modo, as medidas para reduzir a quantidade de desligamento devido ao vegetal incluem:

- projetos de rede impedindo condutores muito próximos;
- construção de linhas verticais, evitando o contato entre duas fases por meio do vegetal;
- construção de redes subterrâneas;
- adotar redes compactas;
- avaliação de custo de condutores com isolante.

Uma forma muito utilizada para evitar o contato com as árvores é através da poda periódica adjacente à rede. Na maioria dos sistemas de distribuição realizam-se os desbastes entre duas a seis vezes por ano. Entretanto, essa atividade é uma das mais onerosas para a concessionária, tendo as empresas de energia dos Estados Unidos onerações maiores que \$2 bilhões anualmente (BROWN, 2002). Existe ainda um método chamado *Reliability-centered Maintenance* (RCM) que elenca os locais que devem primeiro ser podados, baseado em índices que avaliam o benefício que a poda irá impactar na confiabilidade (BROWN, 2002).

2.5 Interferência humana

Assim como a interferência dos animais na rede, é interessante analisar as principais causas da falta do sistema elétrico devido à influência humana. Além dos desligamentos programados, em que a concessionária planeja manutenções, troca de equipamentos e expansão da rede, ocorrem fortuitamente muitos desligamentos não programados devido à interferência humana, sendo os principais decorrentes de: erro humano, acidentes automobilísticos e vandalismo.

Os desligamentos ocasionados por erro da concessionária podem decorrer da má reconfiguração de um sistema (por engano ou devido a manobras de contingência) através de chaves ou sistema SCADA, resultando em consumidores sem energia. Outros tipos de falha incluem as que ocorrem durante operações de poda, quando acidentalmente pode cair um ramo da árvore nas linhas de distribuição, ou pelo esquecimento de remover a conexão do terra, para desenergização antes de manutenções.

Os acidentes automobilísticos são muito comuns em decorrência de ambientes urbanos com elevada concentração de veículos, ficando os postes dispostos a sofrerem abaloamentos ou mesmo romperem, causando o desligamento. Ao contrário dos acidentes fortes, que causam interrupção instantânea, os acidentes leves podem levar ao desalinhamento do poste. Isto faz com que os condutores fiquem com excesso de folga, levando à interrupção em curto e médio-prazo. Também é comum a ocorrência de veículos como guindastes que causam contato entre os condutores e, dependendo da capacidade dos guindastes, podem até mesmo arrastar as linhas e postes de distribuição. A frequência desses desligamentos pode ser diminuída através de barreiras e postes de concreto e linhas de distribuição subterrânea.

Ações de vandalismo típicas podem ser observadas quando há calçados amarrados nos condutores, podendo causar a interrupção devido ao choque entre as linhas. Também caracterizam-se como tal os furtos dos condutores para vender o cobre e tiros em elementos como isoladores e conectores cerâmicos, que representam um “alvo” interessante por explodirem ao serem acertados pelo projétil. Esses problemas podem ser mitigados substituindo-se, respectivamente, os condutores por cabos com quantidade reduzida de cobre, de modo a baixar o valor da venda do material, e substituindo-se os isoladores cerâmicos por poliméricos, tornando-os menos atrativos.

2.6 Análise de desligamentos

Devido as diferentes regiões, clima, fauna, flora e consumidores onde os sistemas elétricos das empresas de energia no mundo se localizam, podemos afirmar que cada companhia terá que lidar com as causas de desligamento com características próprias de cada região. Por exemplo, na América do Norte é comum a presença de esquilos como fator de interferência na rede. No Brasil, contudo, a presença desses animais não ocorre. Por isso, é necessário um levantamento fidedigno das causas dos desligamentos em cada concessionária para realizar a identificação dos desligamentos mais influentes. O gráfico da Figura 2-4 apresenta as maiores incidências de interrupções para três diferentes companhias de energia dos Estados Unidos, sendo que uma atende área exclusivamente urbana (*Utility 3*).

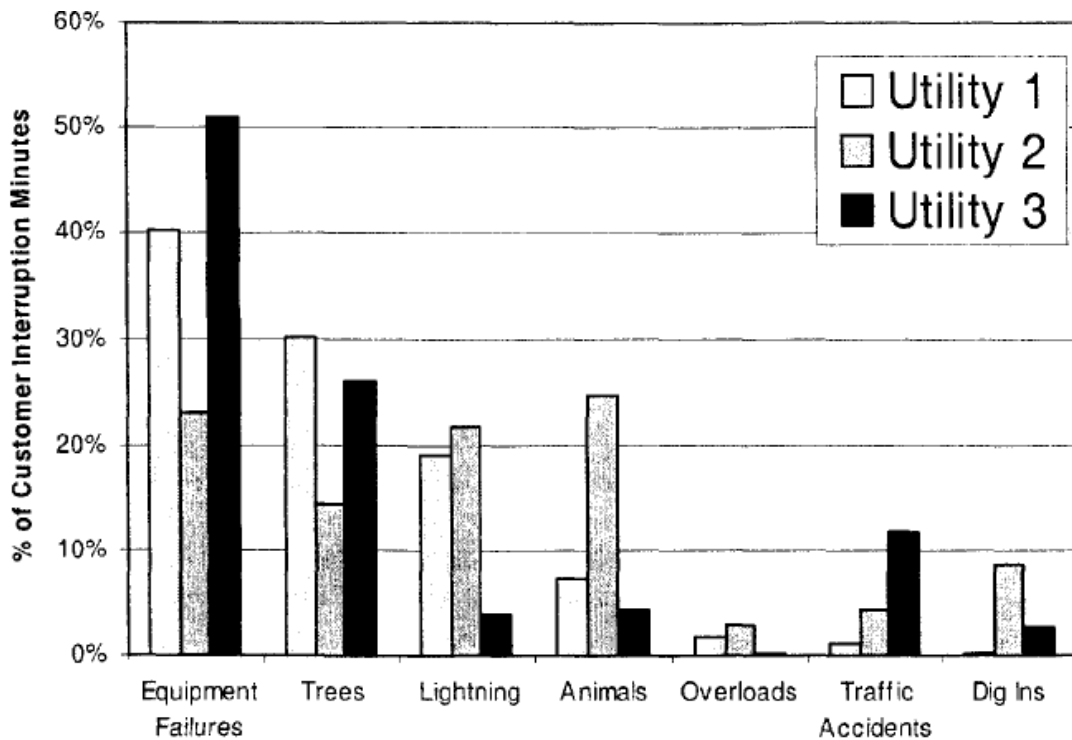


Figura 2-4 – Incidências de interrupções para três companhias de energia. (BROWN, 2002)

A análise da Figura 2-4 permite-nos tirar algumas conclusões interessantes. Observa-se que, em média, as três principais causas são: falhas em equipamentos, árvores e descarga atmosférica. Verifica-se, também, que a empresa que apresenta mais incidências de causas devido a acidentes automobilísticos é a que atende à área urbana, fato que podemos

considerar coerente pela maior concentração de veículos. Além disso, essa mesma companhia 3 não é tão prejudicada com descargas atmosféricas. Conclui-se, também, que as empresas não são afetadas significativamente por sobrecarga, o que não justificaria investir muito, tanto na redução dessa causa de desligamento quanto em pára-raios. Entretanto, sendo a falha por componente a causa predominante, é razoável supor que esses equipamentos não estão sendo corretamente instalados, ou outro motivo que não a redução de vida útil pela sobrecarga.

A associação desses dados com outras variáveis (dia, hora, regiões, temperatura, umidade, etc) possibilitaria ainda estabelecer a identificação de novas informações, contribuindo para o maior entendimento das características em torno das causas de forma conclusiva. A este procedimento de manipulação de dados expondo informações ocultas dá-se o nome de *Data Mining* (Mineração de Dados). Além disso, técnicas estatísticas permitem também uma apuração dos dados, obtendo-se um panorama geral das ocorrências de desligamento. A análise dos principais motivos de interrupção, sobretudo, é a etapa fundamental para o objetivo de estudar economicamente ações que façam reduzir o impacto dessas causas, objetivando melhorar certos índices de confiabilidade da concessionária.

3 LÓGICA FUZZY

Existe muito difundido na cultura do mundo ocidental a ideia de precisão; aquilo que está impecavelmente de acordo com o ideal. Esse conceito tem sua origem na Grécia Antiga, através da lógica clássica de Aristóteles (Harris, 2006). Tal pensamento tem sido a característica essencial do desenvolvimento científico e social, representado pela classificação de objetos reais ou abstratos dentro de conjuntos, em que ou estão perfeitamente membros desses conjuntos ou não o estão.

Por outro lado, a mente humana tem naturalmente a propensão de procurar por abstrações e também a classificar e generalizar. A teoria de conjuntos fuzzy apresenta um conceito de lidar matematicamente com a imprecisão da mesma forma que a mente humana. Segundo Harris (2006), ultimamente a lógica fuzzy tem resultado nas aplicações em uma série de áreas da engenharia, como: determinação de confiabilidade, fadiga/resistência de materiais, controle de processos, logística, diagnósticos e diversos tipos de controles de qualidade. As próximas seções apresentam os principais fundamentos, conceitos, vantagens e desvantagens da lógica fuzzy como um método computacional.

3.1 Fundamento

Na associação da lógica fuzzy aos computadores depara-se com um novo paradigma. Ao contrário da precisão e rigor de outros métodos computacionais, que se baseiam nos extremos da lógica clássica (zero ou um) para determinar a veracidade de uma lógica, na lógica fuzzy, este intervalo entre zero e um é aproveitado, como na colocação:

Lógica Clássica: $\{x \mid x \in \{0,1\}\}$

Lógica fuzzy: $\{x \mid x \in [0,1]\}$

em que x representa o valor lógico de uma sentença. Na lógica fuzzy, este valor possui um grau indicando sua pertinência em um conjunto; quanto maior esse grau, maior sua pertinência (CHEN, 2001).

Parte-se do princípio de que a precisão e certeza geram um elevado custo computacional para representar a mente humana. A fim de contornar este problema, a técnica visa explorar mais as tolerâncias constituídas pela imprecisão e incerteza, a fim de realizar as tomadas de decisão. Conforme observa Zadeh (1994), é com base na habilidade humana de explorar a tolerância, que somos capazes de realizar tarefas diárias, como: compreender sons eventualmente distorcidos de palavras, decifrar manuscritos de diferentes pessoas, resumir textos, reconhecer e classificar imagens e, de forma geral, realizar decisões racionais em ambientes de incerteza. Assim como a mente humana é capaz de solucionar problemas sem natureza precisa, a lógica fuzzy surge como uma técnica eficiente para a modelagem envolvendo incertezas.

A particularidade de utilizar a imprecisão para modelar processos cognitivos humanos permite que nesse método seja possível o uso computacional através de variáveis linguísticas. A partir de conceitos vagos como *quente*, *morno* e *frio*, pode-se, por exemplo, realizar um modelo utilizando essas variáveis semânticas para controlar a temperatura adequada de um chuveiro.

3.2 Conceitos

Utilizando fundamentalmente a imprecisão das variáveis linguísticas, a lógica fuzzy define técnicas para a aplicação da teoria dos conjuntos fuzzy. As próximas seções abordam os principais conceitos empregados na lógica fuzzy para a determinação de um modelo aproximado.

3.2.1 Conjuntos fuzzy

Conjuntos fuzzy são aqueles que não delimitam de forma precisa e rígida se contêm ou não dado elemento. Parte do princípio de que um elemento pode estar contido no conjunto de forma parcial. É o oposto, portanto, do conceito ortodoxo de conjunto que estamos habituados, em que é enquadrado numa das seguintes alternativas: ou o elemento está efetivamente contido em um conjunto clássico ou não o está. Na verdade, a lógica clássica

booleana pode ser tratada como um caso particular da lógica fuzzy, em que são usadas somente as extremidades do intervalo (0,1).

Um exemplo de conjunto clássico consiste na definição das estações do ano conforme os meses em que ocorrem as mudanças, como observa-se na Figura 3-1.

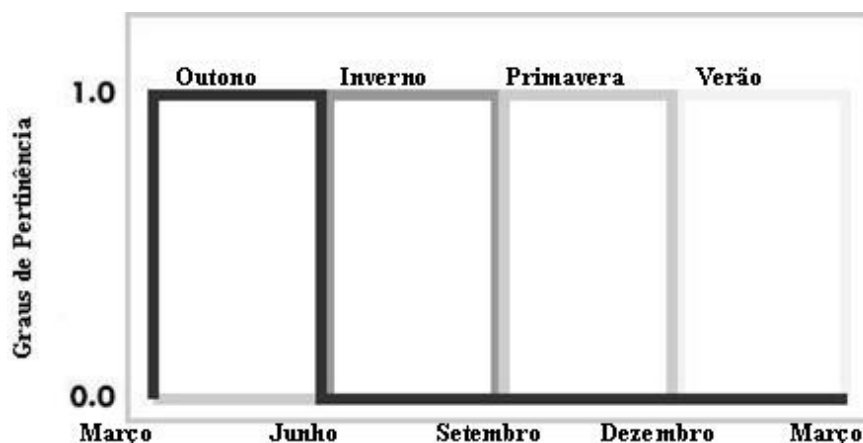


Figura 3-1 – Definição das estações do ano por conjuntos binários. (MATHWORKS, 1995)

De acordo com a figura, o mês de junho divide Outono e Inverno de forma brusca, entre o valor zero ou um. Ou seja, conforme a pequena variação de um dia na metade do mês de junho, teríamos ou um dia típico de Outono (18 °C) ou um dia com frio extremo do Inverno (-2 °C). Sabe-se, evidentemente, que na realidade essa mudança climática ocorre pela forma gradual com que o eixo da Terra aponta em sentido contrário ou a favor do Sol, portanto, de forma suave e contínua. A

Figura 3-2 ilustra esse processo de modo mais adequado.

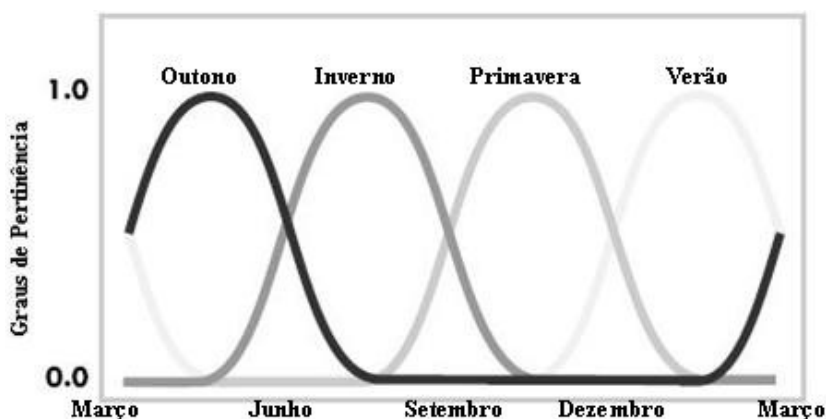


Figura 3-2 – Definição das estações do ano por conjuntos fuzzy. (MATHWORKS, 1995)

Nota-se, agora, a existência de uma graduação para as variáveis do mês pertencerem a determinada estação climática, oferecendo possibilidade destas variarem no intervalo compreendido entre zero e um.

De modo geral, os conjuntos fuzzy permitem a ponderação de pertinência de elementos em conjuntos, permitindo uma modificação necessária nas variáveis para simular eventos da natureza e o raciocínio humano.

3.2.2 Funções de Pertinência

No exemplo citado, as curvas que definem o grau que as estações do ano prevalecem são funções de pertinência, que mapeiam o espaço de entrada (mês) para o espaço de saída (estação do ano). Esse grau é o valor que corresponde ao intervalo entre zero e um (segundo o eixo vertical do gráfico), conferindo a possibilidade que dado elemento *mês* pertença a um determinado conjunto *estação do ano*. O espaço de entrada também é chamado de *Universo de Discurso* (MATHWORKS , 1995).

Desde que respeitem a condição do intervalo compreendido entre zero e um, as funções de pertinência podem assumir diferentes formatos. A Figura 3-3 apresenta os tipos de curva mais usuais: em forma de triângulo, gaussiana, trapezoidal, etc. Esses formatos podem ser optados de acordo com a simplicidade, conveniência e eficiência para tratar certo problema. As funções gaussianas e em formato de sino são as mais utilizadas, por serem suaves e não-nulas em todos os seus pontos (MATHWORKS , 1995). Outras funções de uso comum são as triangulares e trapezoidais.

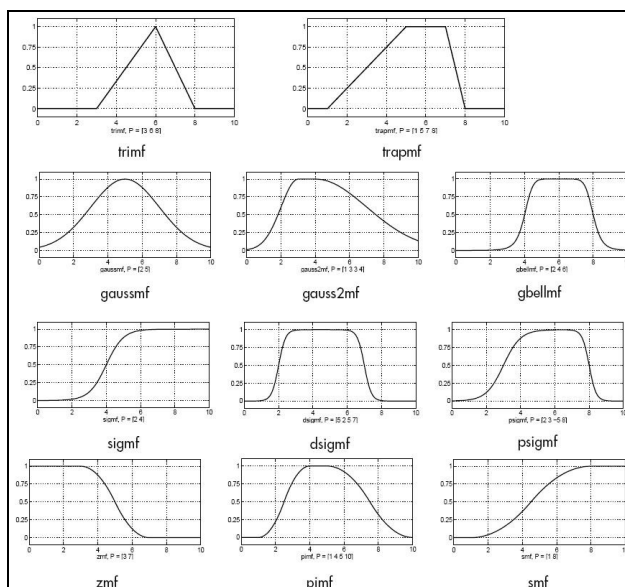


Figura 3-3 – Funções de pertinência. (MATHWORKS, 1995)

Segundo Ross (2004), uma vez que sistemas fuzzy permitem a modelagem de incertezas, não existe uma função de pertinência única para representar valores incertos e o analista deve decidir de forma intuitiva como a função deve se parecer, buscando respeitar o atendimento às propriedades de *normalidade*, *monotonicidade* e *simetria*. Algumas propriedades podem ser verificadas no anexo A. Ainda segundo o autor, os conceitos mais importantes para especificar os conjuntos consistem em: disposição aproximada das funções no universo de discurso, o número de curvas utilizadas e a característica de sobreposição das mesmas.

3.2.3 Operadores Lógicos

Assim como na lógica booleana são utilizadas operações lógicas como AND e OR para valores 0 ou 1, na lógica fuzzy também se faz necessário realizar operações semelhantes, mas desta vez com os valores do intervalo (0,1). Uma maneira de realizar-se isto consiste nas operações *min* (interseção) e *max* (união), respectivamente. As correspondências entre operações de lógica booleana e lógica clássica podem ser conferidas na Figura 3-4.

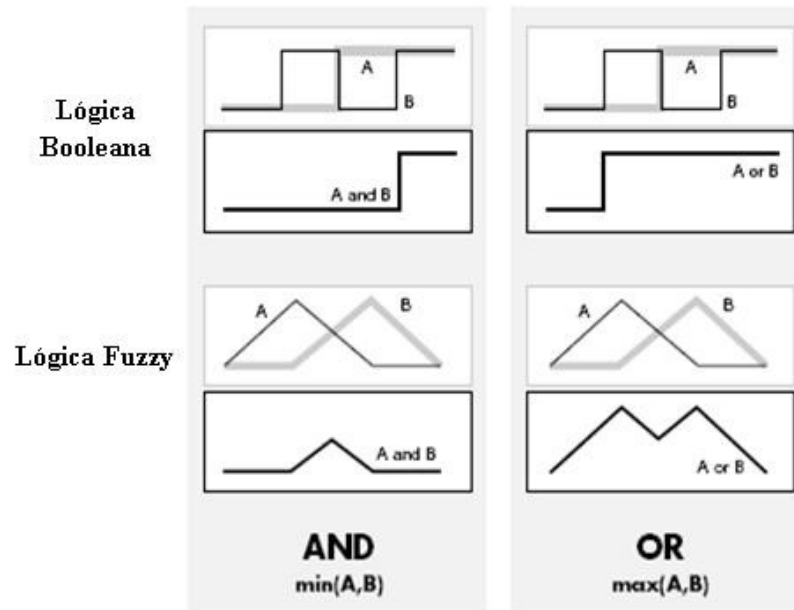


Figura 3-4 – Correspondência entre as operações de lógica booleana e clássica. (MATHWORKS, 1995)

3.2.4 Regras Se-Então

O sistema fuzzy tipicamente compõe um mapa de regras que definem relações condicionais entre os diversos conjuntos (ROSS, 2004). Essas regras apresentam-se na forma:

Se x é A , então y é B

onde A e B são conjuntos fuzzy, nomeados com variáveis linguísticas, que são compostos pelos universos de discurso X e Y , respectivamente. A primeira parte, correspondente a “ x é A ” é denominada *antecedente*, a segunda, “ y é B ”, constitui a *implicação* ou *consequente* (ROSS, 2004).

Observa-se o exemplo de uma regra que modela a relação entre o volume e a densidade de uma bola inflável.

Se Volume é “Alto”, então Densidade é “Baixa”

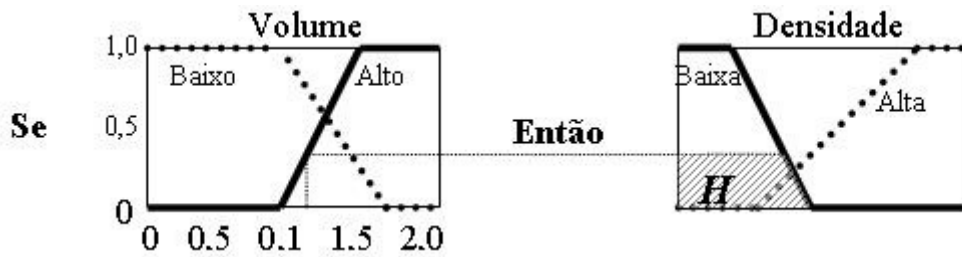


Figura 3-5 – Regra determinando a relação entre as variáveis fuzzy *volume* e *densidade*.

O espaço de entrada desse sistema corresponde aos valores numéricos da variável de entrada *Volume*, de 0 à 2 (u.v.). A regra, portanto, especifica a relação entre esses valores numéricos com um dos conjuntos fuzzy da variável *Volume* (antecedente) e estabelece a correspondência do respectivo grau de pertinência com um determinado conjunto fuzzy da variável *Densidade* (consequente). A área hachurada *H* que se origina na variável de saída da Figura 3-5 compreende um valor que “pesa” apropriadamente as características linguísticas que são atribuídas pela regra, configurando o modelo (MATHWORKS, 1995). A etapa em que o valor numérico é convertido em valor de pertinência, chama-se *fuzzyficação*.

Tanto o antecedente quanto o consequente podem ser constituídos por múltiplas partes, por exemplo:

Se x é A e Se x é F e Se x é P , então $y1$ é B e $y2$ é C (antecedentes com lógicas AND)

Se x é A ou Se x é F ou Se x é P , então $y1$ é B e $y2$ é C (antecedentes com lógicas OU)

Nesses casos é que os operadores lógicos *min* e *max* são aplicados, correspondendo aos operadores booleanos AND e OR, respectivamente. O objetivo é realizar as operações de modo que o resultado do antecedente de cada regra seja apenas um número entre 0 e 1 (grau de pertinência). Na implicação, comumente realiza-se a operação *min* com esse número e o conjunto fuzzy indicado pelo consequente da regra, de modo a obter-se como saída a área hachurada da Figura 3-5.

3.2.5 Agregação

De modo geral, um sistema fuzzy é composto por duas ou mais regras. Os passos descritos anteriormente são aplicados individualmente, para cada regra, obtendo-se como resultado um número de áreas hachuradas equivalente ao número de regras. A etapa de *Agregação* é o processo em que essas áreas são combinadas através do operador *max*, gerando um único conjunto fuzzy.

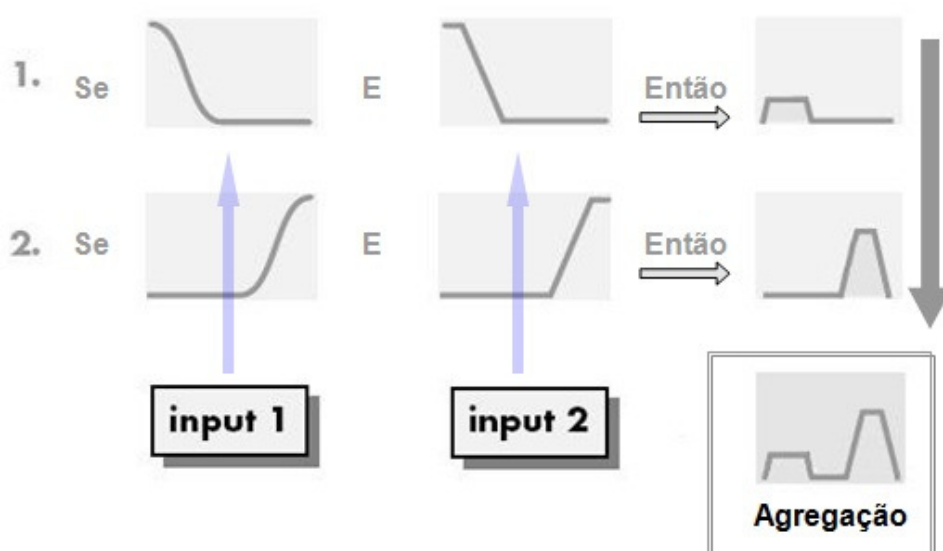


Figura 3-6 – Esquema da etapa de *agregação*. (MATHWORKS, 1995)

Esse conjunto fuzzy resultante ainda contém a informação formada a partir do mapeamento do universo de discurso na variável de saída, utilizando as características qualitativas estabelecidas pelo conjunto de regras.

3.2.6 Defuzzyficação

Pode-se afirmar que o conjunto fuzzy agregado é composto por faixas de valores da variável de saída (áreas hachuradas H) que representam a resposta imprecisa de cada regra. Uma demonstração desse conceito pode ser verificado na Figura 3-7. O grupo de áreas H sobrepostas compõem grosseiramente a modelagem de uma função f^* , que é uma aproximação por lógica fuzzy da função f .

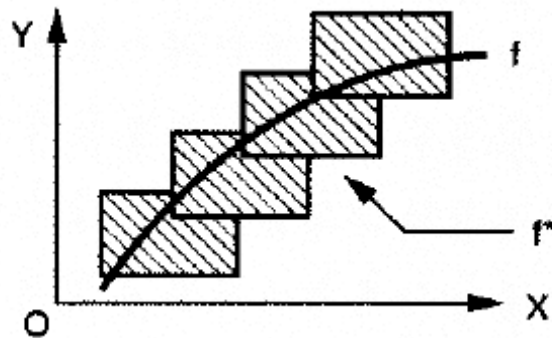


Figura 3-7 – Modelagem de uma função f através de regiões de áreas hachuradas compondo f^* . (ZADEH, 1994)

Para obter uma solução mais refinada, é necessário realizar a etapa de *Defuzzificação*, que consiste em converter o conjunto fuzzy agregado em um único número de saída. Existem diversos tipos de aproximações, sendo talvez o *centróide* o método mais comum (MATHWORKS, 1995). Nesse método calcula-se o centro de área abaixo da curva do conjunto fuzzy agregado, conforme ilustra a Figura 3-8.

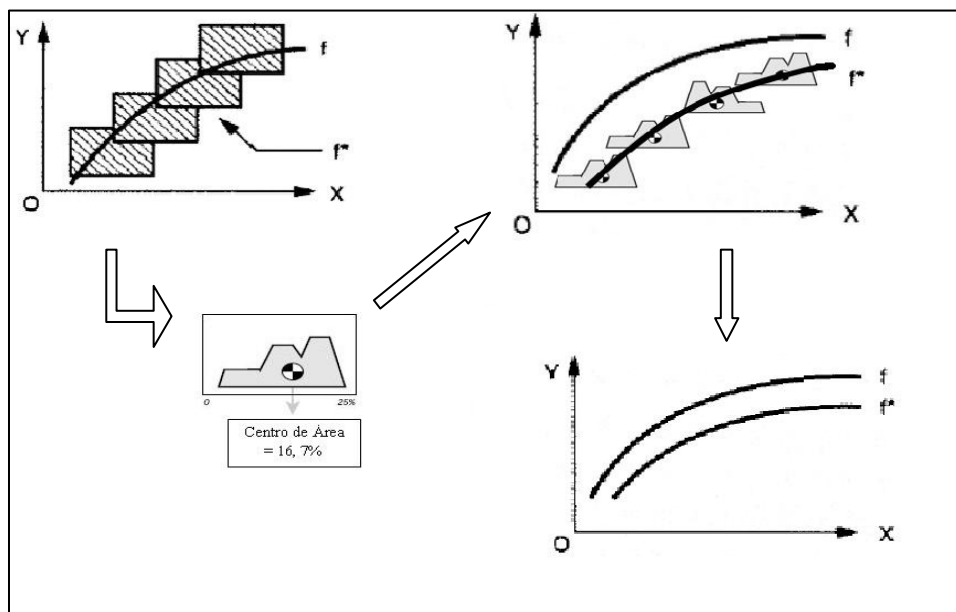


Figura 3-8 – Refinamento da função f^* obtida por meio da etapa de *defuzzificação*.

Outros métodos de aproximação consistem em aproximar as pequenas hachuras através de: *biseção*, *média do máximo*, *maior do máximo* e *menor do máximo*. Para compor um sistema fuzzy, a escolha dos métodos deve estar condicionada ao método que melhor se aproxima da função objetivo f .

3.3 Vantagens e desvantagens da lógica fuzzy

A grande percepção de Zadeh foi que a matemática pode ser utilizada como ligação entre a linguagem e a inteligência humana. Muitos conceitos são melhor definidos por palavras do que pela matemática, e lógica fuzzy com sua “expressão” por conjuntos fuzzy fornece um método que pode construir modelos melhores da realidade. Segundo Neill (1994), em relação a outros métodos, o uso da lógica fuzzy possui diversas vantagens e algumas desvantagens, verificadas nas seções 3.3.1 e 3.3.2.

3.3.1 Vantagens

Entre as vantagens, destacam-se:

- poucos valores, regras e decisões são necessários para mapear as saídas;
- são utilizadas variáveis linguísticas e não numéricas, de modo similar à maneira que os seres humanos raciocinam;
- a lógica fuzzy relaciona saídas com entradas sem precisar da compreensão de todas as variáveis;
- a simplicidade do método permite a solução de problemas sem os ter solucionado antes;
- o desenvolvimento de um protótipo rápido se torna possível, pois um projetista não precisa ter um amplo conhecimento de todo o sistema antes de começar o trabalho;
- são modelos com tendência ao aumento da robustez;
- a aquisição do conhecimento e sua representação são simplificados;
- são necessárias poucas regras para englobar um sistema de grande complexidade.

3.3.2 Desvantagens

Podem ser destacadas as seguintes desvantagens:

- é difícil abstrair o desenvolvimento de um modelo a partir de um sistema fuzzy;
- uma vez que são sistemas fáceis de desenvolver e rápidos de criar protótipos em relação a outras técnicas, são necessárias mais simulações e refinamentos dos sistemas fuzzy, antes de serem utilizados operacionalmente;
- talvez a maior desvantagem seria a tendência da cultura ocidental em dar preferência à criação de sistemas precisos baseados na matemática booleana.

4 REDE BAYESIANA

A modelagem por métodos probabilísticos através de Redes Bayesianas vem sendo empregada desde 1990, com aplicações em áreas como Medicina, Estatística, Física e Engenharia. Talvez a aplicação de Redes Bayesianas mais conhecida são as implementadas em produtos da Microsoft, como o *Answer Wizard*, do Office 95, o Assistente Office (o clipe) do Office 97 e mais de 30 outros serviços de suporte técnicos (MURPHY, 1998). A interface atrativa dos modelos gráficos favorece a modelagem de variáveis com amplas relações causa-efeito entre as mesmas, permitindo o projeto de algoritmos eficientes de uso geral.

4.1 Fundamento

Considera-se o problema da identificação do motivo de um automóvel não ligar seu motor. Através de indicações do que é observado no ambiente (ruído da ignição, mostradores do painel do carro, tempo sem ligar, etc), o ser humano é capaz de inferir se o problema estaria na falta de combustível, ausência de carga na bateria, defeito no motor de arranque, falta de limpeza nas velas, entre outros. O que poderia ser de grande utilidade seria um sistema que representasse a dinâmica de causa e efeito das relações entre as variáveis do automóvel e pudesse simular computacionalmente a conjectura da causa responsável pelo defeito. A Rede Bayesiana pode ser vista como um sistema de Rede Causal, cujos objetivos são representar um problema como este e realizar inferências a partir de certos indícios do sistema.

4.2 Inferindo com Incertezas

O problema do automóvel pode ser melhor representado se considerarmos os graus de certeza das relações entre causa e efeito das variáveis. Do contrário, somente estaríamos restringidos a eventos de natureza binária. Observa-se a sentença simples: “Se chover, com certeza a grama estará molhada”. Segundo Jensen (2001), podemos conceituar essa causa-efeito como um raciocínio lógico simples, em que os resultados das proposições P e Q são

restringidos por valores numéricos 0 ou 1, como o exemplo da tabela-verdade condicional, conforme Tabela 4-1.

Tabela 4-1 – Tabela-verdade do tipo subjunção, cuja saída é binária.

P	Q	$P \rightarrow Q$ (“Se P então Q”)
V	V	1
V	F	0
F	V	1
F	F	1

Ou seja, o resultado das proposições $P \leftrightarrow Q$ são valores binários, sem a presença de incertezas. Entretanto, assim como na lógica fuzzy, ao lidar com eventos incertos, podemos estender melhor essa lógica utilizando resultados com valores que aproveitem o intervalo contido entre 0 e 1. Conforme Jensen (2001), desse modo é possível trabalhar com sentenças como: “Se eu tomar um copo de café no intervalo, ficarei acordado na próxima aula, com incerteza de **0.5**”, ou “Se eu der uma pequena caminhada durante o intervalo, irei ficar acordado, com uma incerteza de **0.8** na próxima aula”.

Muitas vezes, nesses problemas, também é preciso analisar o resultado de múltiplas incertezas. Considere que uma pessoa tenha caminhado, bem como tomado um copo de café no intervalo. Com que certeza então estará essa pessoa acordada na aula? Para responder a isso, é necessário uma regra que combine as incertezas de cada sentença, **0.5** e **0.8**, e retorne um número com a incerteza resultante dessa combinação.

4.3 Redes Causais

Redes Causais consistem de um conjunto de variáveis e conexões (*arcos*) entre as variáveis, que possibilitam representar as associações de incertezas de um sistema. Matematicamente, esse conjunto é chamado de grafo, com as variáveis denotadas por *nós*. Se

existe um arco dirigido da variável A para B , diz-se que A causa B e é *Pai* de B (JENSEN, 2001).

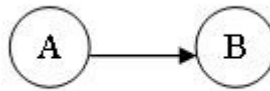


Figura 4-1 – Rede causal em que A é Pai de B .

Um modelo de rede causal para o problema do automóvel de modo simplificado pode ser ilustrado na Figura 4-2. A rede permite o acompanhamento de quanto uma variação de uma determinada variável pode modificar as outras variáveis.

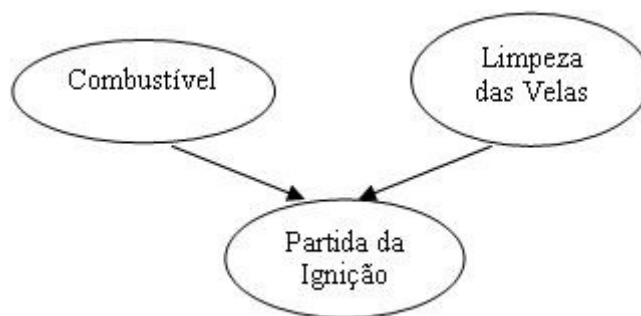


Figura 4-2 – Rede causal do sistema de ignição de um automóvel.

Cada variável representa um evento, que pode ter um ou mais estados (mutuamente exclusivos). Por exemplo, a variável *Limpeza das Velas* assumiria os estados *Sim* ou *Não*, a variável *Combustível*, os estados 0 , $1/4$, $1/2$, $3/4$ e 1 , e assim por diante.

4.4 Cálculo Probabilístico

A teoria de Redes Causais está fundamentada através do estudo do *Cálculo de Probabilidade Clássica*, responsável pela estimação quantitativa da ocorrência de eventos. Se

A é uma variável com estados a_1, a_2, \dots, a_n , então $P(A)$ denota a distribuição de probabilidades desses estados:

$$P(A) = (x_1, \dots, x_n); \quad x_i \geq 0; \quad \sum_{i=1}^n x_i = 1$$

onde x_i é a probabilidade de A estar no estado a_i .

4.4.1 Axiomas Básicos

A probabilidade $P(a)$ é um número no intervalo unitário $[0,1]$, do qual obedece os seguintes axiomas (JENSEN, 2001):

- i) $P(a) = 1$ se, e somente se, a não for incerto (infalível)
- ii) Se a e b são eventos mutuamente exclusivos, então a probabilidade de a ou b ocorrer é determinada como:

$$P(a \vee b) = P(a) + P(b).$$

4.4.2 Probabilidades Condicionais

O principal formalismo que explica o comportamento de Redes Causais baseia-se na *probabilidade condicional*. Considerando que um evento A depende somente de B , a probabilidade condicional $P(A | B) = x$ é interpretada como:

“Dado o evento B , a probabilidade do evento A é x .”

O que não significa dizer que, a partir do conhecimento de que B tenha ocorrido, sempre ocorrerá A . Na verdade, o conhecimento da veracidade de B somente implica então que a probabilidade de A ocorrer é x (JENSEN, 2001).

4.4.3 Regra Fundamental

Conforme Jensen (2001), estipulando-se que existe uma dependência do evento B em relação ao evento A (a causa de A é B), tem-se a *regra fundamental* no Cálculo de Probabilidade, estabelecida como:

$$P(A, B) = P(B) P(A | B) \quad (4-1)$$

onde $P(A,B)$ é a probabilidade da ocorrência simultânea dos eventos A e B , também chamada de *probabilidade conjunta*. De outro modo, considerando-se uma inversão de dependências, desta vez de A em relação a B , obtém-se:

$$P(A, B) = P(A) P(B | A) \quad (4-2)$$

Comparando-se as expressões (4-1 e (4-2), deriva-se a chamada *Regra de Bayes* (JENSEN, 2001):

$$P(B|A) = \frac{P(A, B)}{P(A)} = \frac{P(A|B) P(B)}{P(A)} \quad (4-3)$$

Conforme Murphy (1998), o termo $P(B|A)$ é denotado como a probabilidade *a posteriori* da ocorrência de B . Esta probabilidade na forma invertida, $P(A|B)$, é chamada de *verossimilhança*. O multiplicador $P(B)$ é a probabilidade *a priori* da ocorrência de B . O denominador $P(A)$ pode ser percebido como a constante, obtida a partir de $P(A,B)$, necessária para normalizar o numerador (MURPHY, 1998).

4.4.4 Marginalização

Considere a variável A com os estados a_1 e a_2 e que a variável B tenha os estados b_1 , b_2 , e b_3 . Pode-se esquematizar a relação de estados $P(A|B)$ conforme exemplo da Tabela 4-2:

Tabela 4-2 – Exemplo de probabilidade condicionais dos estados de $P(A|B)$.

	b_1	b_2	b_3
a_1	0.3	0.8	0.1
a_2	0.7	0.2	0.9

Sabendo que $P(B) = (0.2, 0.3, 0.5)$, pode-se aplicar a equação (4-1) para obter a probabilidade conjunta $P(A,B)$:

Tabela 4-3 – Obtenção da probabilidade conjunta $P(A,B)$ a partir de $P(A|B)$.

	b_1	b_2	b_3
a_1	0.06	0.24	0.05
a_2	0.14	0.06	0.45

A probabilidade conjunta caracteriza, portanto, todas as combinações possíveis de estados entre as variáveis relacionadas.

Um conceito muito importante é que, a partir da tabela de probabilidade conjunta, podemos deduzir tanto $P(A)$ quanto $P(B)$. Observa-se que todos os elementos da Tabela 4-3 somam 1. Segundo o axioma (ii), uma vez que cada termo (a_j, b_j) é mutuamente exclusivo, para calcular $P(A)$ é necessário somar todos os elementos em cada linha, obtendo-se $P(A) = (0.35, 0.65)$ independentemente do estado de B :

Tabela 4-4 – Obtenção de $P(A)$ a partir de $P(A,B)$.

	B
α_1	0.35
α_2	0.65

Esse processo de eliminação de uma variável chama-se *marginalização* (JENSEN, 2001), e dizemos que a variável B foi marginalizada em $P(A,B)$, resultando $P(A)$. Este procedimento tem a notação:

$$P(A) = \sum_B P(A,B) \quad (4-4)$$

Do mesmo modo, a partir de $P(A,B)$, pode-se obter $P(B)$ novamente. Para isto, somam-se os elementos em cada coluna da Tabela 4-3. A marginalização é de grande importância para determinar cálculos como o numerador e o denominador da equação de Bayes (4-3) (MURPHY, 1998).

4.5 Topologias e D-Separação

Considere a rede com conexão em série da Figura 4-3. Segundo Jensen (2001), se o estado de B é conhecido (instanciado), então A e C são ditos condicionalmente independentes (ou D-separados). Matematicamente isto pode ser escrito como:

$$P(C|B) = P(C|A,B)$$

ou seja, a probabilidade de C ocorrer depende somente de B e não de A e B .

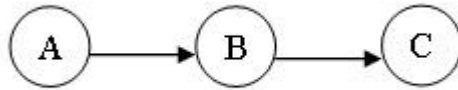


Figura 4-3 – Conexão serial.

Na rede com conexão divergente da Figura 4-4, se *A* é instanciado, então *B*, *C* e *D* são D-separados (JENSEN, 2001). Isto é, sendo *A* instanciado, uma evidência de *B*, por exemplo, não afeta *C* ou *D*.

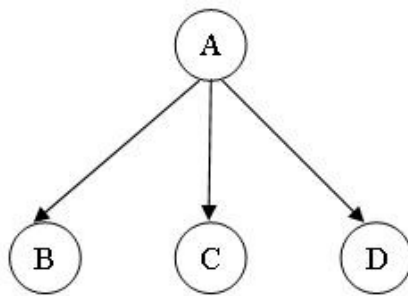


Figura 4-4 – Conexão divergente.

Conforme Jensen (2001), no caso de uma rede com conexão convergente, Figura 4-5 se *A* (ou seus descendentes) não estiverem instanciados, então *B*, *C* e *D* são D-separados e a evidência não se transmite entre os nós *BCD*.

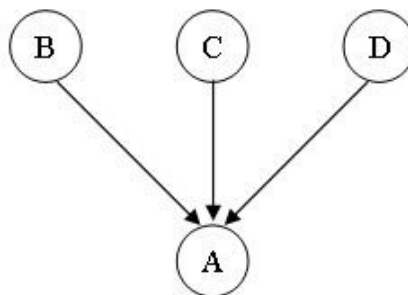


Figura 4-5 – Conexão convergente

O conceito de D-separação tem a importância de definir uma rede como sendo do tipo Bayesiana. Além disso, é uma propriedade muito utilizada em algoritmos de cálculos de

inferência, uma vez que aproveita a vantagem da independência condicional para dispor as variáveis de modo a simplificar o problema.

4.6 Definição de Redes Bayesianas

Conforme Jensen (2001), devido à complexidade de alguns tipos de topologias, que necessitariam de equações diferenciais para a resolução, existem formalmente algumas condições que simplificam uma rede causal, passando a ser denominada Rede Bayesiana:

- consiste em um conjunto de variáveis e um conjunto de arcos entre as variáveis;
- cada variável deve possuir um conjunto finito de eventos mutuamente exclusivos;
- as variáveis, juntamente com os arcos, formam um Grafo Acíclico Dirigido (GAD). (Um grafo dirigido é acíclico se não existe um caminho com realimentação $A_1 \rightarrow \dots \rightarrow A_n$, sujeito a $A_1 = A_n$);
- as propriedades de D-separação devem ser válidas no modelo e ter coerência com o mundo real;
- para cada variável A com pais B_1, \dots, B_n existe uma tabela com a *Distribuição de Probabilidade Condicional* (DPC), denotada por:
 - $P(A \mid B_1, \dots, B_n)$.

Observa-se que se A não possuir pais, a tabela reduz-se à probabilidade incondicional $P(A)$.

O modelo através de Redes Bayesianas também costuma ser chamado de “generativo”, porque ele especifica como uma causa gera efeitos. Redes em que os grafos não são dirigidos (apenas conexões, sem flechas) são chamados de *redes de Markov*, e são comuns nas áreas da Física e visão computacional (MURPHY, 1998).

4.7 A Regra da Cadeia para Redes Bayesianas

Na seção 4.4.4, observou-se a relevância da tabela de probabilidade conjunta. Murphy (1998) afirma que uma maneira de representá-la matematicamente de forma compacta é através da *Regra da Cadeia*. Considere o universo de variáveis $U = \{A_1, \dots, A_n\}$ de uma Rede Bayesiana. Então, a distribuição de probabilidade conjunta $P(U)$ é calculada como:

$$P(U) = \prod_i P(A_i \mid \text{pai}(A_i)) \quad (4-5)$$

onde $\text{pai}(A_i)$ é o conjunto de pais de A_i .

Observa-se que, se tivermos acesso à tabela conjunta $P(A_1, \dots, A_n)$, então pode-se inferir $P(A_i)$, bem como $P(A_i \mid e)$, onde e é uma evidência (JENSEN, 2001). Devido ao grande número de parâmetros que devem ser armazenados, dependendo do número de pais de cada variável, existem diversas técnicas computacionais que permitem representá-las computacionalmente, como através de: função sigmóide, *noisy-OR*, Árvores de Decisão, Árvores de Regressão, Redes Neurais com Realimentação e modelos lineares em geral (MURPHY, 1998). A escolha de uma das técnicas está condicionada à complexidade do modelo e ao custo computacional.

4.8 Algoritmo de inferência *Bucket elimination*

Conforme pode ser notado, é necessário aplicar a regra da cadeia para obter a tabela de probabilidade conjunta para realizar inferências. Apesar de esta tabela disponibilizar diversos estados possíveis entre as variáveis, para um problema com n variáveis, o número de combinações possíveis cresce exponencialmente com n , comprometendo o tempo de processamento computacional. Este problema é classificado como do tipo *NP-Hard* (JENSEN, 2001). Considere a Rede Bayesiana da Figura 4-6, assumindo que todas as variáveis possuam 10 estados. Tem-se a instância $\{D=d, F=f\}$ como evidências e deseja-se inferir $P(A \mid D=d, F=f)$. A partir da equação (4-3), é possível calcular:

$$P(A|D = d, F = f) = \frac{P(A, D = d, F = f)}{P(D = d, F = f)} \quad (4-6)$$

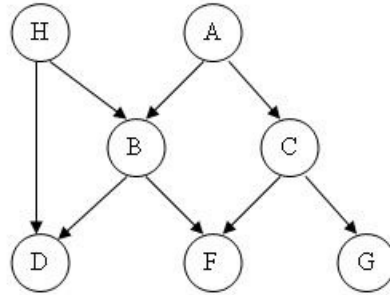


Figura 4-6 – Exemplo de Rede Bayesiana. Evidências $\{D=d, F=f\}$.

Para obter o numerador de (4-6), é preciso recorrer à tabela de probabilidades conjuntas. Visto que são 7 variáveis com 10 estados cada, porém, com 2 delas instanciadas, a tabela terá 10^5 entradas. Entretanto, conforme Jensen (2001), o algoritmo *Bucket elimination* permite obter a probabilidade conjunta sem a necessidade de lidar com uma tabela com as 10^5 entradas. A partir da regra da cadeia, pode-se equacionar:

$$P(A, B, C, d, f, G, H) = P(A) P(H) P(C | A) P(B | H, A) P(d | B, H) P(f | B, C) P(G | C)$$

O numerador da equação (4-6) pode ser obtido através da marginalização de B, C, G e H em $P(A, B, C, d, f, G, H)$:

$$\sum_G P(A, B, C, d, f, G, H) = \sum_G P(A) P(H) P(B | A, H) P(C | A) P(d | B, H) P(f | B, C) P(G | C)$$

Somente o último termo contém G em seu domínio. Devido à lei distributiva (JENSEN, 2001), pode-se realizar:

$$\sum_G P(A, B, C, d, f, G, H) = P(A) P(H) P(B | A, H) P(C | A) P(d | B, H) P(f | B, C) \sum_G P(G | C)$$

Para resolver $\sum_G P(G|C)$, pode-se escrever:

$$\sum_G P(G|C) = \frac{\sum_G P(G,C)}{P(C)}$$

Uma vez que a marginalização de G em $P(G,C)$ é $P(C)$, tem-se que:

$$\sum_G P(G|C) = \frac{P(C)}{P(C)} = 1$$

Assim,

$$\sum_G P(A, B, C, d, f, G, H) = P(A)P(H)P(B|A, H)P(C|A)P(d|B, H)P(f|B, C)$$

Em seguida, marginaliza-se H , obtendo-se:

$$\sum_H P(A, B, C, d, f, G, H) = P(A)P(C|A)P(f|B, C) \sum_H P(H)P(B|A, H)P(d|B, H)$$

Multiplicando-se as tabelas $P(H)$, $P(B|A, H)$ e $P(d|B, H)$ e marginalizando H desse produto, resulta em uma tabela $T(B, A)$, ficando:

$$P(A, B, C, d, f) = P(A) P(C|A) P(f|B, C) T(B, A)$$

A partir disso, calcula-se esse produto e marginaliza-se B e C , resultando em $P(A, d, f)$. Nesse processo de eliminação de variáveis, percebe-se que não se trabalha com uma tabela com mais de 3 variáveis, o que torna bastante eficiente o cálculo computacional.

4.9 Método *Junction Tree*

Segundo Paskin (2003), a técnica através de *junction trees* generaliza o processo *bucket elimination*, de eliminação de variáveis, para a execução simultânea de grandes grupos de variáveis. O algoritmo consiste na transmissão de mensagens em um tipo de grafo

simplificado chamado *junction tree* (Árvore Conjunta), cujos nós são conjuntos de variáveis formando clusters. Dessa forma, é possível simplificar o grafo através da eliminação de ciclos. A Figura 4-7 ilustra um grafo transformado em *junction tree*. Combinando as tabelas de probabilidades de cada cluster com as tabelas recebidas dos clusters vizinhos através de mensagem, é possível marginalizar as variáveis do cluster adequadamente.

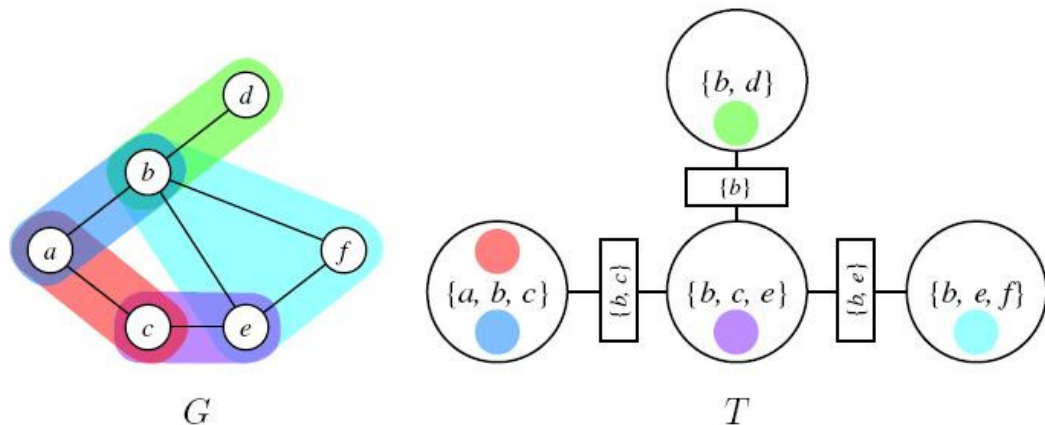


Figura 4-7 – Transformação de um grafo (G) em *junction tree* (T). (PASKIN, 2003)

A complexidade do algoritmo aumenta com a largura da “Árvore”. Dois algoritmos que utilizam a arquitetura de *junction tree* são os chamados Shafer-Shenoy e o Hugin, que evita multiplicações repetidas (PASKIN, 2003).

4.10 Outros algoritmos

Conforme Murphy (1998), os métodos de inferências podem ser classificados em dois grandes grupos: *exatos* e *aproximados*. A inferência exata (por possuir solução analítica) é aplicável principalmente em casos em que todos os estados do nó são discretos. Sistemas inteligentes e modelos de Markov caem nesta categoria. As inferências em Gráficos Acíclicos Dirigidos exploram a decomposição da regra da cadeia da probabilidade conjunta. O método *Bucket elimination*, que segue o princípio de distribuir somas nos produtos é um exemplo de inferência exata, sendo a base para muitas outras técnicas exatas, como: decodificação Viterbi, Transformada Rápida de Fourier e o *Junction Tree*. A Programação Dinâmica

também é utilizada para evitar redundâncias computacionais se for preciso aplicar a eliminação de variáveis de forma repetida. Desta forma são obtidas diversas marginalizações ao mesmo tempo.

No caso da inferência aproximada, o uso se justifica quando o custo computacional da inferência depende de uma grande largura da “árvore” do grafo ou quando não existem equações que permitam solução analítica. Algumas técnicas de algoritmos de inferência aproximada são: Métodos de amostras (Monte Carlo), Métodos por variação, *Loopy belief propagation*, *Bounded cut-set conditioning*, Amostragem de Gibbs e Métodos de aproximação paramétrica (MURPHY, 1998).

4.11 Parametrização da Rede Bayesiana

Segundo Murphy, (1998), na construção de uma Rede Bayesiana, é necessário especificar a estrutura da rede (etapa qualitativa) e seus parâmetros numéricos (etapa quantitativa). A parametrização de um modelo acíclico dirigido envolve a especificação das tabelas de distribuição de probabilidade condicional (DPC) associadas a cada uma das variáveis. Na Figura 4-8 ilustra-se o resultado de uma Rede Bayesiana com suas tabelas DPC.

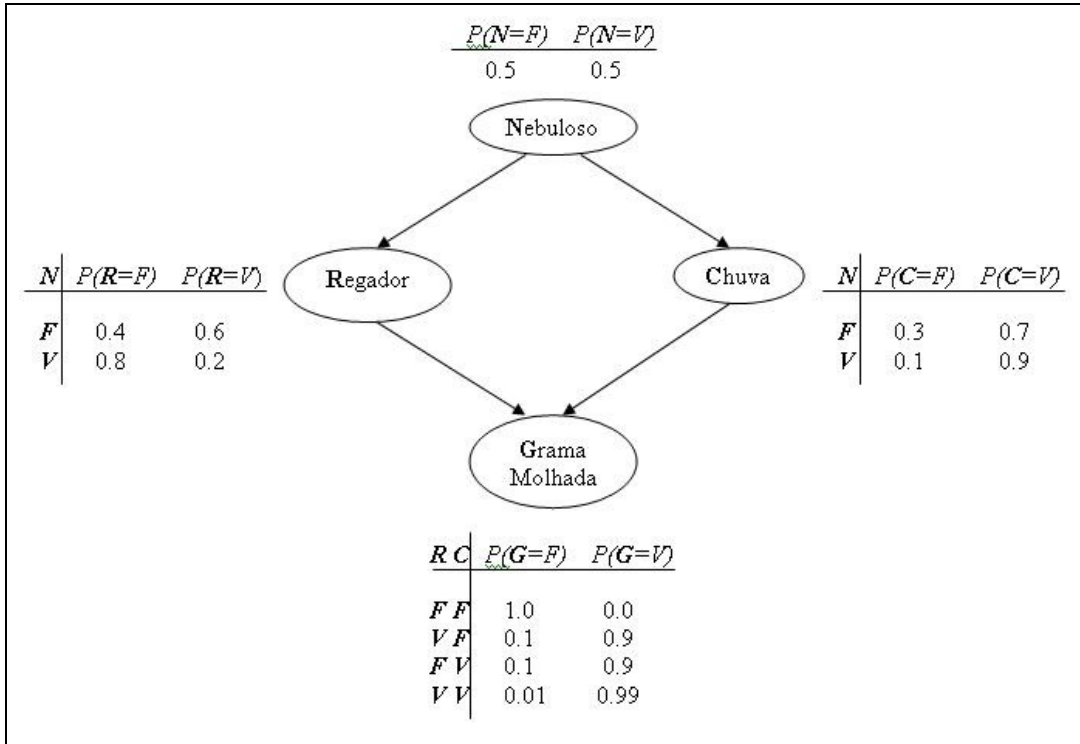


Figura 4-8 – Parâmetros de uma Rede Bayesiana. (JENSEN, 2001)

Em princípio, essas tabelas podem ser determinadas através da análise estatística dos eventos relacionados aos estados dos pais de cada variável. Entretanto, existe a desvantagem de requererem um número de parâmetros exponencial ao número de pais dessas variáveis, e muitas vezes é inviável fazer essa análise, principalmente em modelos muito complexos. Apesar disso, é possível realizar o aprendizado desses parâmetros de forma automática, com base em um histórico de eventos (dados de treino), analisando-se probabilisticamente a ocorrência dos estados de cada variável por meio de uma função de otimização (MURPHY, 1998).

A meta do aprendizado é encontrar os valores dos parâmetros para cada DPC que maximizam a verossimilhança dos dados de treino, que contêm N casos assumidos como independentes. Empregando o logaritmo da verossimilhança, que é mais fácil de trabalhar do que unicamente a verossimilhança, Murphy (1998) afirma que cada nó pode ser calculado como:

$$L = \frac{1}{N} \sum_{i=1}^m \sum_{l=1}^s \log P(X_i | Pai(X_i), D_l) \quad (4-7)$$

onde D é o conjunto dados de treino.

Conforme Murphy (1998), através dos dados de treino, o número de casos que ocorre a mesma combinação de estados são contados. Utilizando a regra de Bayes, estima-se a tabela DPC com essas contagens, obtendo-se a probabilidade condicional $P(X_i | Pai(X_i))$ para cada variável X . No exemplo da Figura 4-8, podemos contar o número de vezes que a grama está molhada quando está chovendo e o regador está ligado, $Num(G=1, R=1, C=1)$, o número de vezes em que a grama está molhada quando está chovendo e o regador está desligado, $Num(G=1, R=0, C=1)$, etc. A partir dessas contagens, obtém-se uma estimativa da DPC através da equação (4-8):

$$P(G = g | R = r, C = c) \approx \frac{Num(G = g, R = r, C = c)}{Num(R = r, C = c)} \quad (4-8)$$

onde o denominador $Num(R=r, C=c) = Num(G=0, R=r, C=c) + Num(G=1, R=r, C=c)$. Por meio da equação (4-7), estima-se então a tabela que possui a máxima verossimilhança (*Maximum Likelihood Estimation* – MLE).

Murphy (1998) indica que, em casos em que alguns valores de nós do treino de dados não estão disponíveis (não-observados), é possível estimar as tabelas através do algoritmo EM (*Expectation Maximization*). A finalidade deste método é encontrar o MLE local dos parâmetros (“etapa M”) por meio de Valores Esperados de todos os nós usando algum dos algoritmos de inferência citados (“etapa E), e então tratando esses dados como se tivessem sido observados. Por exemplo, na variável *Grama Molhada*, substitui-se as contagens observadas de eventos com o número de vezes que esperamos ver cada evento:

$$P(G=g, | R=r, C=c) = E Num(G=g, R=r, C=c) / E Num(R=r, C=c)$$

onde $E Num(x)$ é o número de vezes esperado de um evento x ocorrer em todo o conjunto de treino de dados. Essa contagem esperada pode ser calculada como:

$$E \text{ Num}(\cdot) = E \sum_{k=1}^N I(\cdot | D(k))$$

onde $I(\cdot | D(k))$ é uma função indicativa que retorna 1 se o evento x ocorre em um caso de treino k , ou retorna 0, caso contrário.

Em um processo iterativo, então os parâmetros são maximizados e são recalculadas as contagens esperadas novamente. Nesse procedimento, busca-se convergir para um máximo local da superfície de verossimilhança. Na etapa de otimização, também é possível realizar o método Gradiente Descendente na superfície de verossimilhança (MURPHY, 1998).

5 METODOLOGIA

Com o objetivo de desenvolver um método de auxílio na identificação de causas de falta de energia elétrica, os fundamentos de Redes Bayesianas (RB) e lógica fuzzy foram empregados para criar modelos com variáveis de características incertas. A criação da RB parte de procedimentos equivalentes aos adotados na referência Tronchoni (2008), com a utilização de um banco de dados de uma concessionária, que foram tratados para determinar os parâmetros da RB (etapa de aprendizagem). Desse modo, maiores detalhes desses procedimentos não serão aqui abordados.

Na proposta deste trabalho, apresenta-se um método elaborado com a tentativa de atenuar o erro de inferência de desligamentos não programados ao utilizar registros não confiáveis. Para isto, juntamente com um modelo de Rede Baesiana, um modelo fuzzy será utilizado para auxiliar na correta identificação da causa de interrupção. A associação deste modelo possibilita a inferência da causa a partir de lógicas baseadas em conhecimento humano. A extração de conhecimento foi obtida por meio de uma análise minuciosa do banco de dados, desenvolvendo-se a base de regras para o modelo fuzzy, que neste caso tem atribuição de inferir as causas mais recorrentes em um sistema elétrico. O sistema com Rede Bayesiana fará inferências de causas de acordo com a tendência dos registros. O método proposto consiste em comparar as respostas dos modelos com referências de saídas que orientam a causa mais provável. Os modelos foram construídos e simulados utilizando-se o sistema operacional Windows Vista®, em um *notebook* com processador Centrino, clock de 1.6 Ghz e memória RAM de 1Gb.

5.1 Introdução

Na referência de Pretto (2005), o sistema desenvolvido para o gerenciamento de causas de desligamento é composto por um módulo de aquisição de dados e outro com aplicativos disponíveis para análise de informações, conforme a Figura 5-1.

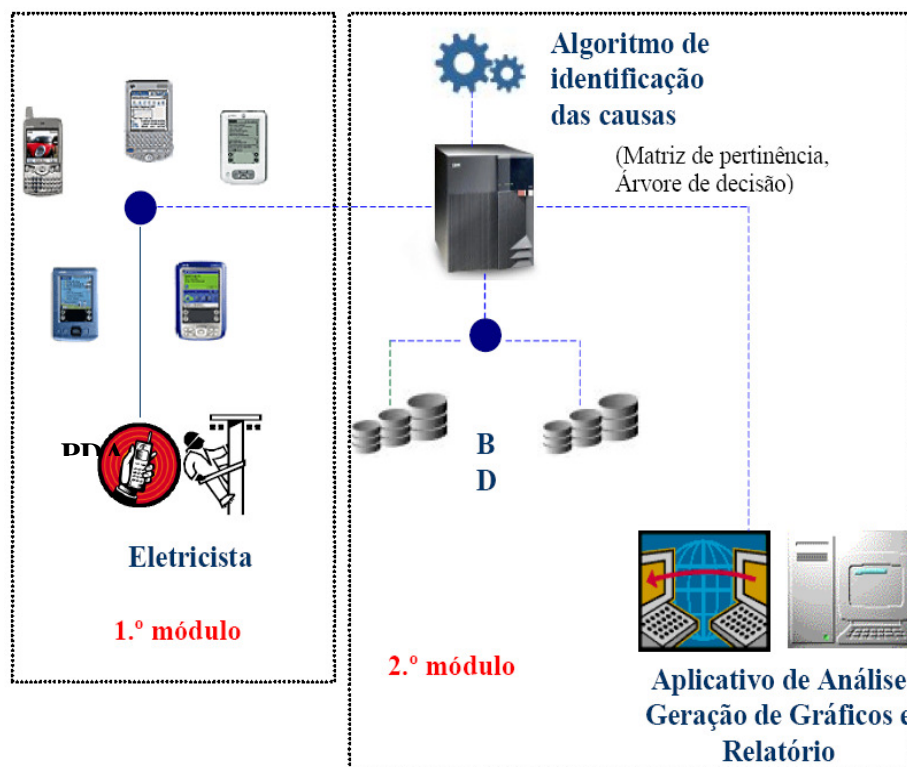


Figura 5-1 – Sistema elaborado por Pretto (2005).

No trabalho de Tronchoni (2008), a pesquisa foi motivada pela evolução de um modelo de inferências de causas promovido pelo segundo módulo da Figura 5-1. Entretanto, os dados não foram coletados através do PDA (1º módulo). A base de dados disponível não havia sido projetada com uma estrutura adequada, tal que embora existisse uma grande quantidade de dados disponíveis, existiam muitos registros incompletos e contraditórios. Apesar do considerável número de registros, apresentando 570.409 eventos, a impureza dessa base de dados não era adequada para a representação necessária a fim de realizar o aprendizado de máquina.

Assim como na pesquisa anterior, no presente trabalho foram aplicados os mesmos tratamentos de dados de Tronchoni (2010) nos registros da mesma concessionária. O número de eventos deste banco de dados é de 803.737 desligamentos, a partir de 384 alimentadores em 76 municípios do RS, no período de abril de 2005 a abril de 2006. O processo de KDD é mostrado na Figura 5-2.

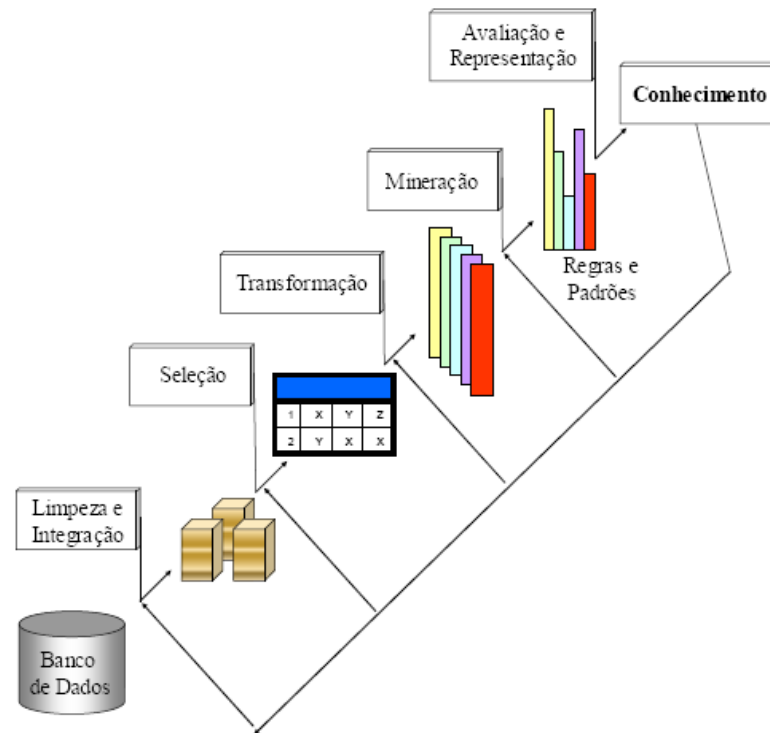


Figura 5-2 – Processo de KDD. (TRONCHONI, 2008)

Após a aplicação do KDD, foi criada uma nova base de dados contendo informações explícitas mais adequadas ao problema com as mesmas variáveis do trabalho de Tronchoni (2008). Contudo, optou-se por realizar duas propostas de alteração nesses registros:

- **adoção de outras três variáveis de entrada:** realizada na etapa de mineração, isto possibilitou maior diferenciação do banco de dados, gerando mais detalhes dos cenários que levam à determinada causa;
- **descarte de eventos com saídas diferentes (causas) para a mesma combinação de entradas (cenários de desligamento):** com esta restrição, espera-se uma caracterização maior das condições de ocorrências para cada tipo de falta, com a finalidade de elaborar um conhecimento com causas mais definidas.

Por meio dessas medidas, foi possível aperfeiçoar o banco de dados de modo a conter dados mais confiáveis e organizados para os propósitos deste trabalho. A análise

posterior da nova base de dados permitiu o entendimento das principais combinações de entrada que geram determinadas saídas. Com isso, foi disponibilizada a extração de informações desses dados para gerar uma base de conhecimento para um sistema de filtro fuzzy. Este sistema irá suplementar um suporte de inferências a um sistema de Rede Bayesiana, conforme ilustrado de forma simplificada na Figura 5-3.

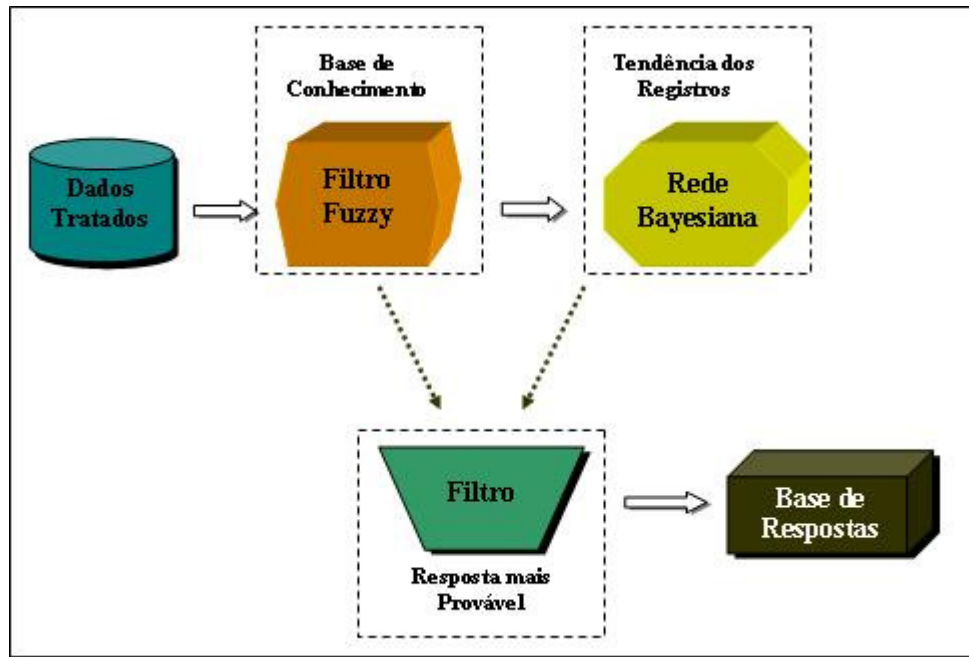


Figura 5-3 – Digrama simplificado proposto.

Desse modo, o filtro fuzzy apresentará uma resposta com base em conhecimento, se esta atingir certo grau de confiança, e a Rede Bayesiana realizará a inferência da causa de interrupção baseando-se na tendência dos dados registrados, priorizando certos eventos quando necessário. A verificação da resposta mais provável entre o fuzzy e a RB é realizada na saída dos modelos, com um filtro adicional. Nas seções a seguir descrevem-se cada uma das seguintes etapas da pesquisa:

- definição das variáveis;
- tratamento de dados;
- adequação e análise;
- elaboração dos sistemas.

5.2 Definição das Variáveis

Seguindo o trabalho de Tronchoni (2008), foram definidas as mesmas variáveis de desligamento não programado, que compõem os principais fatores de interferência na confiabilidade do sistema elétrico encontrados na literatura (BROWN, 2002; OPRISAN, 1991; PRETTO, 2005). Algumas serão obtidas diretamente do banco de dados e outras inferidas através da técnica de mineração (TRONCHONI, 2010).

Os estados da variável de saída **Causa** são as denominações referentes aos diferentes motivos de desligamento. Conforme Tronchoni (2008), a escolha adequada dessas denominações se referem às causas primárias ou causas raíz, visto que as secundárias resultariam em uma manutenção equivocada, não contribuindo para a melhoria na confiabilidade. Os seguintes estados integrarão a variável de saída:

1. **falha no componente:** interrupção motivada por defeito em componentes ou equipamentos, como transformadores, pára-raios entre outros;
2. **sobrecarga:** interrupção causada pelo excessivo consumo, resultando no acionamento da proteção devido ao carregamento;
3. **clima adverso:** desligamento resultante de climas adversos como: chuva, temporal, granizo, vento, temperaturas extremas, entre outras intempéries;
4. **descarga atmosférica:** interrupção devido à descarga atmosférica direta ou indireta sobre a rede, causando a atuação do dispositivo de proteção ou danificando o sistema;
5. **interferência do meio ambiente:** interrupção referente a objetos presos na rede elétrica, incluindo animais, pássaros, entre outros;
6. **vegetal:** desligamento ocasionado pela interferência de árvores e partes do vegetal no sistema elétrico;
7. **interfênciã humana:** interrupção devido a vandalismo, furto, empresas trabalhando no local e demais interferências humanas;
8. **acidente:** desligamento resultante de acidentes automobilísticos;
9. **incêndio/queimada:** Interrupção de energia relacionada a queimadas no local próximo à rede elétrica.

Em contrapartida, as variáveis de entrada são as condições dos equipamentos e ambiente em torno do sistema elétrico que fornecem dados indicativos de fatores relacionados à causa. Estas estão agrupadas como:

1. **elemento de interrupção:** indica a classificação da estrutura física em que ocorreu a falha, classificada a partir do banco de dados. Pode assumir um dos seguintes estados: poste, equipamentos, condutores, isoladores e cruzeta;
2. **clima:** considera-se a condição climática no momento do desligamento. Estados: chuva, temporal, neblina, neve/granizo, bom;
3. **vento:** deriva-se nos estados: excessivo, moderado, sem vento;
4. **fatos associados:** se referem a incidentes ou atividades realizadas próxima à rede com algum potencial de causar desligamento. Compõe os estados: acidente, queimada/incêndio, empresa trabalhando, vandalismo, inundação, erosão, nenhum fato associado (por estar registrado na maioria dos eventos, será simplificado por '*');
5. **vegetação:** fornece indícios sobre a interferência de árvores em regiões próximas da rede elétrica. Constitui os estados: sem poda, podada;
6. **objetos:** relacionado a qualquer objeto preso à rede. Estados: sim, não.

Além dessas variáveis de entrada replicadas da pesquisa de Tronchoni (2008), foram incluídas outras três:

7. **estação:** referente à estação do ano vigente no momento da interrupção de energia: primavera, verão, outono, inverno;
8. **turno:** termo análogo à definição de turno de trabalho das indústrias, sendo neste trabalho considerado como o turno do momento da interrupção de energia (os horários serão definidos na seção seguinte). Estados: 1oTurno, 2oTurno, 3oTurno;
9. **elemento específico:** corresponde ao elemento de interrupção registrado pelo eletricitista, extraído diretamente do banco de dados. Estados: chave, chave

fusível, medidor primário, módulo, religador, seccionizador, transformador e trecho primário.

5.3 Tratamento de Dados

Nesta etapa, os registros foram configurados para torná-los possíveis de serem utilizados a fim de extrair-se informações. Os dados cedidos pela concessionária, obtidos por meio de consulta pelo banco Oracle® através da linguagem SQL (*Structured Query Language*), apresentavam inicialmente 22 campos por registro de desligamento. Cada campo, correspondendo a uma classe de dado, foi armazenado em planilhas do software Excel®. A Figura 5-4 ilustra uma planilha apresentando certos campos de dados, com alguns deles representados por códigos. No campo *WEATHER*, por exemplo, os números representam: *1-Bom, 4-Chuva, 5-Temporal*, entre outros. Tendo em vista que cada linha da planilha registra um desligamento, observa-se que existem campos, de determinados eventos, que não foram preenchidos.

1	A	B	C	D	E	F	G	H	I	J	
EID	FEEDER	SUBSTATION	DEV_NAME	DEV_TYPE	NAME	CAUSE_CODE	RESTORE_DTS	WEATHER	PHASE	ATIVIDADE	
1352	1767703			1762.1-213		0			A	SEM ATIVIDADE	
1353	1767916			192.9-122		0			A	SEM ATIVIDADE	
1354	1767023			338.7-220		0			A	SEM ATIVIDADE	
1355	1756367			15.9-28		0			A	SEM ATIVIDADE	
1356	1767638	FWE-202	144	1512.1-46	Chave Fusivel	304	20050609131000ES		1	A	SEM ATIVIDADE
1357	1767614			6.0-108	Servico					B	SEM ATIVIDADE
1358	1779030	CNL-106	110	91.4-84	Transformador				1	AB	
1359	1764081			2102.4-78		0				A	SEM ATIVIDADE
1360	1764263			7263.0-83		0				ABCN	Abate de ar
1361	1777510	SAR-207	122	6300.2-18	Transformador	304	20050611161600ES		1	ABC	
1362	1779035			10065.0-98	Servico					ABC	Produção n
1363	1778541	TPR1-201	137	28.1-160	Transformador	509	20050611154000ES		5	ABC	
1364	1775220			UNKNOWN		408	20050611151300ES		1		
1365	1778786	VAC-202	120	138.4-235	Transformador				1	ABC	
1366	1777081	VER-204	272	2600.0-230	Transformador	307	20050611143000ES		4	ABC	
1367	1776095	CAX5-106	102	499.5-216	Transformador	509	20050611143500ES		1	ABC	
1368	1776092	CAX3-101	99	260.7-216	Transformador	509	20050611141900ES		1	ABC	
1369	1778443			25.6-235						A	SEM ATIVIDADE
1370	1778501			496.1-82						ABC	Fabricação
1371	1778494			642.4-90						A	SEM ATIVIDADE
1372	1778429			470.7-90						A	SEM ATIVIDADE
1373	1778281	CNL-108	110	1195.9-82	Chave Fusivel				1	ABC	
1374	1777999	SMT-101	115	673.4-187	Transformador	408	20050611113500ES		1	ABC	
1375	1777520			350.6-82						A	SEM ATIVIDADE
1376	1778075	CAX4-104	101	52.3-216	Transformador				1	ABC	
1377	1777986	CAX5-102	102	1343.9-216	Transformador				1	ABC	
1378	1777159			1458.3-216	Servico					ABC	SEM ATIVIDADE
1379	1772388			165.1-82						A	SEM ATIVIDADE
1380	1777825	CAX1-106	97	2104.1-216	Transformador				1	ABC	
1381	1777514	PMI-201	143	5470.4-56	Chave Fusivel	303	20050611104900ES		1	AB	
1382	1776272	BGO2-102	108	1006.5-256	Transformador	408	20050611101500ES		1	ABC	

Figura 5-4 – Planilha do Banco de Dados com alguns campos. Observa-se a esparsidade dos registros.

A partir do software Matlab®, as planilhas foram importadas, reunidas e armazenadas em uma matriz, onde as etapas do Tratamento de Dados conforme o trabalho de Tronchoni (2008) foram realizadas. Os referidos 22 campos correspondiam a dados como: hora e data da interrupção, clima, alimentador, tipo de tarifa, entre outros. Dessas classes, no entanto, apenas 6 foram selecionadas como relevantes para o tratamento de dados:

- data;
- hora;
- alimentador;
- elemento de interrupção;
- clima;
- causa.

Este conjunto de registros formou o grupo de **dados-base** do qual foram extraídas as variáveis de interesse. Para derivar as outras variáveis selecionadas conforme o item 5.2, de início, foi realizada a etapa de limpeza e integração. Após, utilizando-se regras de classificação baseadas em conhecimento de um especialista, essas variáveis foram criadas para compor um cenário detalhado de cada causa.

5.3.1 Limpeza e Integração

A partir dos 803.737 eventos do banco de dados-base, foi verificada a elevada quantidade de determinados campos incompletos devido ao não-preenchimento (campos com o valor 'Null'). Do mesmo modo foram encontrados registros pouco confiáveis, como ocorrências em que a causa havia sido apontada como descarga atmosférica associada com clima bom, além de interrupções denominadas como *Improcedentes*. Assim, uma etapa de filtragem desta base de dados foi necessária para qualificar o banco de dados e torná-lo homogêneo. Eliminando também os registros da concessionária cujas causas foram apontadas como desligamento programado, o número de eventos foi reduzido para 25,6%, ou seja, 205.545 eventos. A Figura 5-5 detalha a proporção entre os desligamentos programados e outras interrupções que não foram consideradas para a análise.

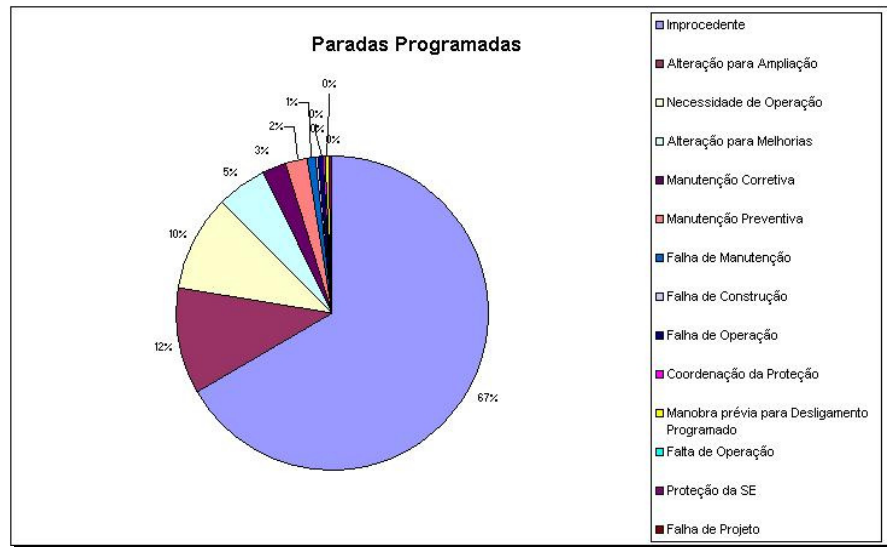


Figura 5-5 – Desligamentos descartados, correspondendo a 47.617 eventos.

Verificou-se também que, dos eventos restantes, um percentual de 90,5% correspondia a causas com clima bom. Segundo Tronchoni (2008), isto contraria a afirmação de Brown (2002), que afirma ser o clima adverso o grande responsável pela maioria dos motivos de interrupção das empresas de energia. Sendo assim, o conjunto de dados foi dividido proporcionalmente de forma a ter-se a totalidade de eventos com clima adverso correspondendo a 70% do banco de dados, sendo o restante dos 30% randomicamente preenchidos com desligamentos registrados com clima bom. A partir disso, o total de eventos passou a integrar 26.282 interrupções de energia, correspondendo a 3,3% dos 807.737 registros iniciais. A Figura 5-6 traduz a grande quantidade de dados descartada.

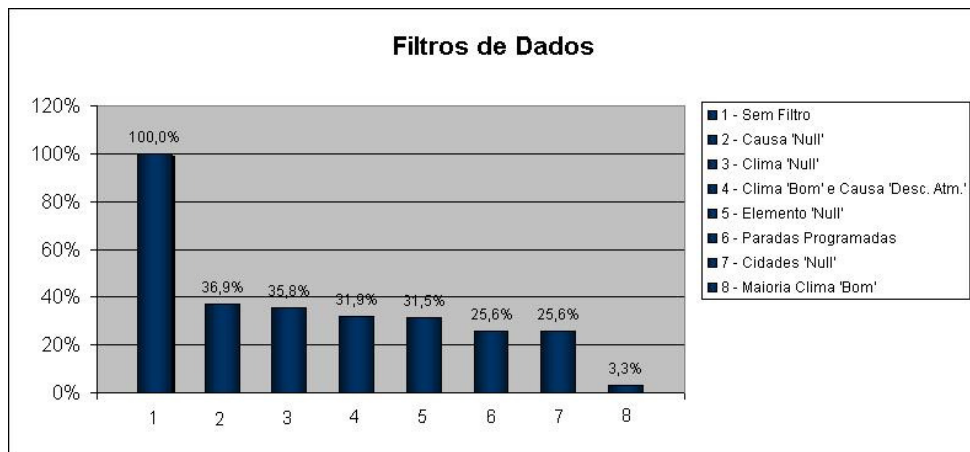


Figura 5-6 – Relação de dados eliminados.

A Figura 5-7 relaciona a proporção das principais causas que integram a base de dados dos 26.282 eventos.

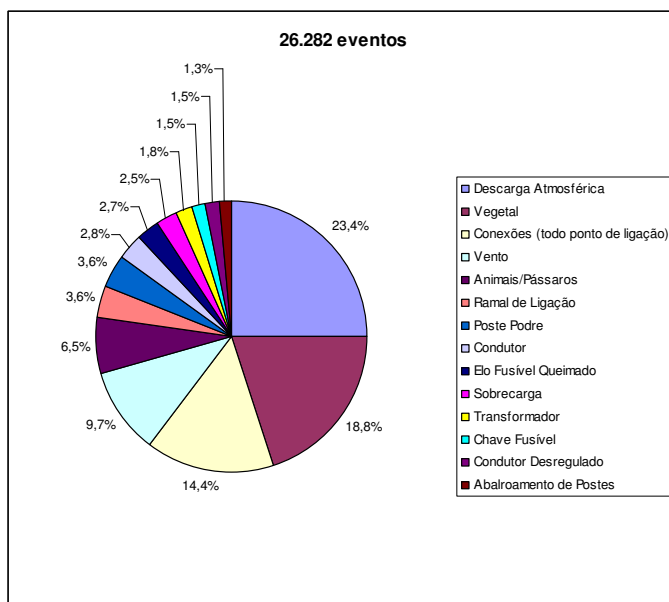


Figura 5-7 – Proporção dos principais registros de causas de desligamento não programados.

5.3.2 Mineração de Dados

Na etapa de Mineração, foi empregado um conjunto de regras de classificação baseadas no conhecimento de especialistas, permitindo-se a predição de variáveis que estavam implícitas no banco de dados original. O resultado disso é o mesmo número de eventos (26.282), mas com as variáveis de entrada e saída atribuídas no item 5.2. Através desta técnica, foram obtidas as variáveis: *Clima*, *Objetos*, *Elemento de Interrupção*, *Vento* e *Causa*. Isto é possível através da associação dos 6 campos do banco de dados-base, permitindo-se a extração de novas informações a partir de outros dados, conforme ilustra a Figura 5-8. O conjunto de regras de classificação é o mesmo adotado na referência de Tronchoni (2008) e não será abordado neste trabalho.

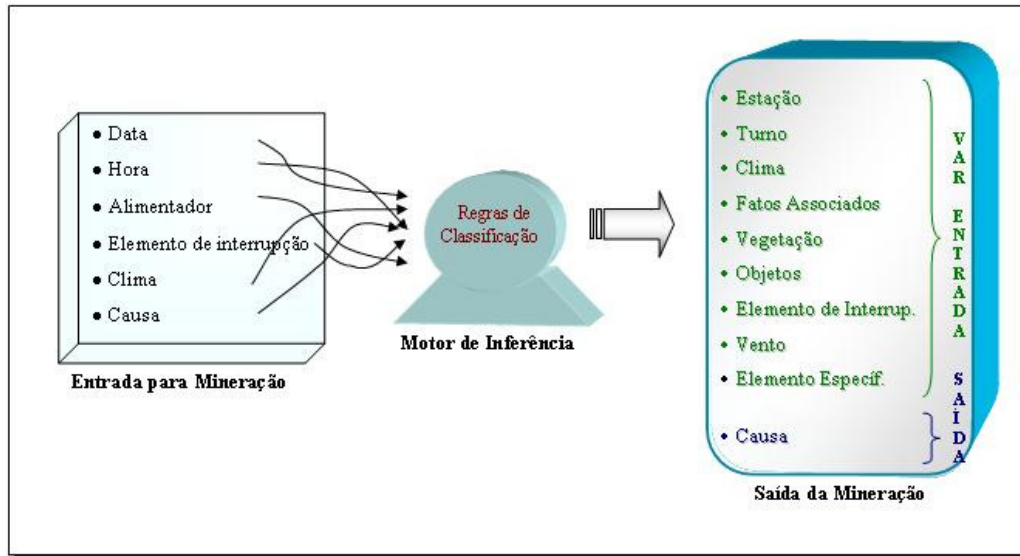


Figura 5-8 – Processo de Mineração

O campo de dados da variável vegetação foi obtido por meio da classe *Alimentador*, que indica o município em que ocorreu a interrupção. De maneira semelhante ao descrito na referência de Tronchoni (2008), por meio do algoritmo *k-means* foi verificado o quão representativa a causa vegetal ocasiona em faltas de energia em determinada região. Desse modo foi estimado se as regiões pertenciam a áreas de vegetação densa de acordo com a localização dos municípios, dada pela variável *Alimentador*, permitindo inferir os estados *Podado* e *Não Podado*. A variável *Elemento Específico* foi obtida apenas por meio da repetição do registro *Elemento de Interrupção*, estabelecido pelo próprio eletricitista no momento do reparo.

As informações contidas nos campos *Data* e *Hora* podem fornecer mais dados de entrada para caracterizar falhas relacionadas com essas variáveis. Conforme o item 2.3.3, em horários de pico de dias quentes, por exemplo, a ocorrência de falha por sobrecarga possui maior chance de ocorrência. Para fazer a correlação aproximada entre temperatura do dia e épocas do ano, os dados do tipo *Datas* foram organizados segundo os seguintes grupos, ou *clusters*:

Tabela 5-1 – Conversão de datas em estados correspondentes da variável *Estação*.

Datas	Estado da Variável
Abril à Maio	Outono
Junho à Agosto	Inverno
Setembro à Novembro	Primavera
Dezembro à Março	Verão

De forma semelhante, os diferentes horários de desligamento podem ser agrupados em 3 *clusters* indicando, por aproximação, os turnos: matinal, vespertino e noturno. Na tentativa de corresponder os horários de maior carregamento, as horas em que cada grupo se enquadram foram definidas a partir de cada turno de uma indústria multinacional, equivalendo aos seguintes horários:

Tabela 5-2 – Conversão de horas em estados correspondentes da variável *Turno*.

Horas	Estados da Variável
06:00 às 14:00	1o Turno
14:00 às 22:00	2o Turno
22:00 às 06:00	3o Turno

Após a criação de todas as variáveis de interesse, considerou-se que o banco de dados estava apto para para um processo de análise de suas informações, descrito nas etapas seguintes.

5.4 Adequação e Análise do Banco de Dados

Na etapa de adequação, foram descartados eventos que continham mais de uma saída. Considerou-se que tais eventos não apresentariam informações suficientes de seus cenários para ser designado apenas um motivo da interrupção não programada de energia.

A etapa de análise do banco de dados possibilitou a obtenção de informações que auxiliaram em determinar quais variáveis de entrada preponderaram em determinada causa. Também possibilitou a verificação de que, após as manipulações do tratamento de dados, a nova base de registros ainda conteria conhecimentos coerentes a respeito do sistema elétrico. A partir disso, um conjunto de regras do modelo fuzzy foi elaborado com base nas informações apresentadas, permitindo o desenvolvimento de inferências de modo análogo a um especialista.

5.4.1 Análise do Banco de Dados

Para esta abordagem, primeiramente foi criada uma lista com todas as combinações possíveis dos estados das 10 variáveis definidas. Pelo Princípio Fundamental da Contagem (HAZZAN, 1977), com n conjuntos contendo k_1 elementos no primeiro, k_2 no segundo e assim por diante, é possível formar C combinações diferentes pela relação:

$$C = k_1 \cdot k_2 \cdot k_3 \cdot k_4 \cdot \dots \cdot k_n. \quad (5-1)$$

Na Tabela 5-3, tem-se a relação da quantidade de estados (elementos) para cada variável (conjunto).

Tabela 5-3 – Número de estados para cada variável.

Variável	Quantidade de Estados
Causa	9
Elemento de Interrupção	5
Clima	5
Vento	3
Fatos Associados	7
Vegetação	2
Objetos	2
Estação	4
Turno	3
Elemento Específico	8

Com isso é possível formar 1.814.400 eventos diferentes, que foram dispostos em uma planilha, como na Figura 5-9.

Causa	Clima	Fatos	Vegetação	Objetos	Elem. Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Transformador	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Chave Fusivel	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Modulo	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Medidor Prima	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Chave	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Religador	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Servico	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Seccionalizadd	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Trecho Primari	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	2oTurno	outono	Regulador	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Transformador	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Chave Fusivel	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Modulo	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Medidor Prima	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Chave	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Religador	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Servico	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Seccionalizadd	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Trecho Primari	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	outono	Regulador	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Transformador	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Chave Fusivel	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Modulo	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Medidor Prima	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Chave	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Religador	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Servico	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Seccionalizadd	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Trecho Primari	0
Clima Adverso	Bom	*	Podado	Não	Equipamento	Excessivo	1oTurno	outono	Regulador	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	1oTurno	outono	Transformador	0
Clima Adverso	Bom	*	Não-Podado	Não	Equipamento	Excessivo	1oTurno	outono	Chave Fusivel	0

Figura 5-9 – Lista parcial apresentando algumas combinações possíveis do estado *Clima Adverso* a partir de suas entradas.

Por meio desta lista, foi contabilizado o número de vezes que essas combinações ocorriam na base de dados, o que corresponde afirmar que foram obtidas as frequências absolutas dos eventos. Na base de dados perfazendo os 26.282 registros, foram encontradas 1.322 combinações diferentes. Essas combinações foram reorganizadas em ordem decrescente das frequências, conforme a Figura 5-10.

Causa	Clima	Fatos	Vegetação	bjeto	Elem. Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.
Vegetal	Bom	*	Não-Podado	Sim	Equipamer	Moderado	2oTurno	outono	Transformador	371
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	2oTurno	verao	Chave Fusivel	308
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	2oTurno	verao	Transformador	281
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	2oTurno	verao	Chave Fusivel	264
Vegetal	Chuva	*	Não-Podado	Sim	Equipamer	Moderado	2oTurno	verao	Transformador	259
Sobrecarga	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	2oTurno	outono	Transformador	231
Vegetal	Bom	*	Não-Podado	Sim	Equipamer	Sem Vento	1oTurno	outono	Transformador	226
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	1oTurno	outono	Transformador	225
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	1oTurno	verao	Transformador	202
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	1oTurno	verao	Chave Fusivel	197
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	1oTurno	verao	Chave Fusivel	190
Vegetal	Temporal	*	Não-Podado	Sim	Equipamer	Excessivo	2oTurno	outono	Transformador	184
Vegetal	Bom	*	Podado	Sim	Equipamer	Sem Vento	2oTurno	outono	Transformador	183
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	1oTurno	outono	Chave Fusivel	178
Vegetal	Bom	*	Podado	Sim	Equipamer	Sem Vento	1oTurno	outono	Transformador	178
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	2oTurno	verao	Transformador	168
Vegetal	Temporal	*	Não-Podado	Sim	Equipamer	Excessivo	2oTurno	verao	Transformador	161
Falha no Componente	Bom	*	Podado	Não	Equipamer	Sem Vento	2oTurno	outono	Transformador	157
Vegetal	Chuva	*	Não-Podado	Sim	Equipamer	Moderado	1oTurno	verao	Transformador	155
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	1oTurno	verao	Transformador	148
Vegetal	Chuva	*	Não-Podado	Sim	Equipamer	Moderado	2oTurno	outono	Transformador	147
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	2oTurno	outono	Chave Fusivel	144
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	2oTurno	outono	Chave Fusivel	143
Falha no Componente	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	1oTurno	outono	Transformador	139
Vegetal	Chuva	*	Não-Podado	Sim	Equipamer	Moderado	1oTurno	outono	Transformador	138
Descarga Atmosférica	Temporal	*	Não-Podado	Não	Equipamer	Excessivo	2oTurno	verao	Chave Fusivel	137
Vegetal	Bom	*	Não-Podado	Sim	Equipamer	Excessivo	2oTurno	outono	Transformador	134
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	2oTurno	outono	Transformador	133
Vegetal	Bom	*	Não-Podado	Sim	Equipamer	Excessivo	1oTurno	outono	Transformador	127
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamer	Moderado	1oTurno	primavera	Transformador	126
Falha no Componente	Bom	*	Podado	Não	Equipamer	Sem Vento	1oTurno	outono	Transformador	122
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamer	Moderado	1oTurno	outono	Transformador	119

Figura 5-10 – Lista parcial de frequência absoluta dos eventos. Observa-se a ordem decrescente do campo Quantidade (Qtde).

Utilizando as frequências absolutas com essa última seleção de eventos, é possível realizar algumas observações agrupando-se os dados em forma de *clusters*, de modo a constatar padrões e comportamentos das variáveis. A técnica utilizada neste trabalho foi dividir a base de dados em *clusters* que continham a mesma variável de saída (causa). Com esses grupos de dados, novas divisões em outras variáveis, como Clima, por exemplo, puderam ser aplicadas, em que objetivou-se identificar padrões de eventos de intempéries relacionados com algum tipo de causa. Assim, as características das variáveis de entrada mais comuns em causas de maior incidência poderiam ser verificadas, favorecendo também a criação da base de regras do sistema fuzzy que suportará a Rede Bayesiana.

5.4.2 Adequação dos Registros

Com a planilha de frequências absolutas foi observada grande quantidade de registros que especificavam mais de um tipo de saída, conforme a Figura 5-11. Essa parcela de

registros corresponde a 39,1% dos 26.282 eventos, o que sugere que as saídas não estão suficientemente caracterizadas pelas combinações das variáveis de entrada, apesar da inclusão das 3 variáveis. Isto se torna um empecilho, tanto para a análise de indicar univocamente que tipo de cenário provoca determinada causa de desligamento, quanto para o aprendizado de máquina, que acaba por realizar inferências ambíguas.



Figura 5-11 – Evento com mais de uma saída para a mesma entrada de dados.

Desse modo, para este trabalho, foi optado adequar a base de dados tornando-a em melhores condições de inferência, de maneira que seja seguramente indicada apenas um tipo de saída para uma combinação de estados de entrada. Visando abranger o maior número possível de eventos, para cada grupo de causas diferentes para a mesma entrada, as causas com maior número de ocorrência foram agregadas ao novo conjunto de registros com saídas únicas, descartando-se as demais. A nova base de dados totalizou 15.994 eventos, correspondendo aos restantes 60,9% das 26.282 ocorrências. A Figura 5-12 apresenta a proporção entre as causas para esta última seleção de dados.

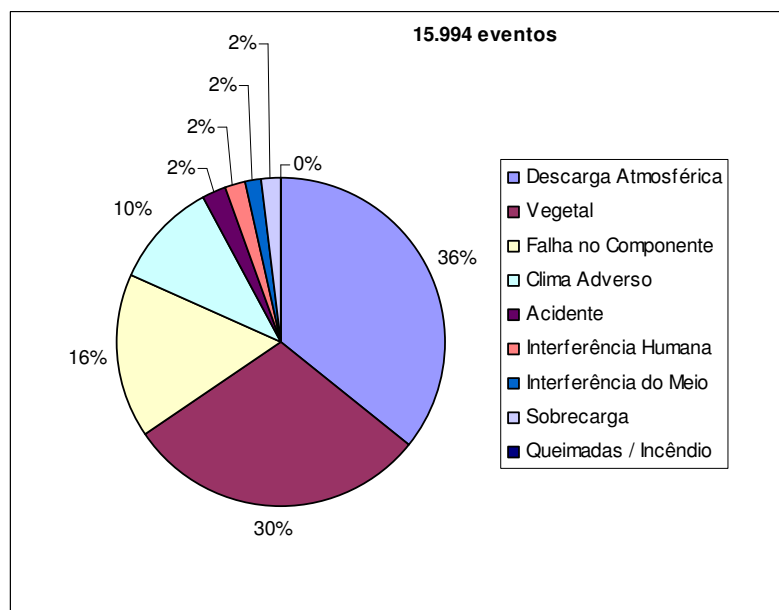


Figura 5-12 – Proporção de causas para os 15.994 eventos.

A partir de então, os sistemas com Rede Bayesiana e Lógica fuzzy foram projetados visando a utilização do novo conjunto de dados definidos, conforme mostrado nas seções 5.5 e 5.6.

5.5 Implementação da Rede Bayesiana

O software Matlab®, juntamente com o pacote BNT (*Bayes Net Toolbox*), foram utilizados para representar a Rede Bayesiana. O software oferece ampla habilidade de programação e simulação, podendo interagir com as planilhas do Excel® e utilitários de outros pacotes, como a representação do modelo fuzzy deste trabalho. Conforme Tronchoni (2008, p.105), o BNT destaca-se, pois:

“... é um pacote de código aberto para Matlab® para modelagem de grafos orientados. O BNT suporta vários tipo de nós (distribuição de probabilidade), inferência exata e aproximada, aprendizado de estrutura e de parâmetros, modelos estáticos e dinâmicos.”

Para representar o modelo Bayesiano no Matlab®, a técnica consiste em determinar a parte qualitativa e quantitativa da rede.

5.5.1 Representação da Estrutura Bayesiana

Para o desenvolvimento da estrutura da Rede Bayesiana, isto é, a definição da topologia com as variáveis selecionadas, tem-se os seguintes procedimentos:

- identificação das variáveis envolvidas e seus estados, que irão compor os nós;
- estabelecimento das relações de dependência entre as variáveis, formando os arcos da rede.

Com base nessas atribuições, as funções do toolbox foram empregadas para configurar a parte qualitativa da RB. Segundo Tronchoni (2008), a estrutura a ser adotada assemelha-se à topologia de um classificador Naïve Bayes, que mostra-se eficaz na realização de diagnósticos em geral. Assim, utilizando as variáveis definidas em 5.2, foi concebida a topologia ilustrada na Figura 5-13, com as variáveis *Estação*, *Turno* e *Elemento Específico* inclusas.



Figura 5-13 – Estrutura da Rede Bayesiana

5.5.2 Determinação dos Parâmetros da Rede Bayesiana

O procedimento utilizado para estimar a parte quantitativa da rede, isto é, deduzir os parâmetros conforme os eventos que se apresentam, foi aplicar o algoritmo *Expectation Maximization*, descrito em 4.11. Através da utilização da função *learn_params_em* no conjunto de eventos observados, o algoritmo estima as probabilidades das tabelas DPC de cada nó da rede e associa em uma variável do algoritmo os parâmetros do modelo.

5.5.3 Inferências da RB e critério de saída

Em contraposição ao trabalho de Tronchoni (2008), que utilizou um algoritmo de inferência baseado na técnica de eliminação de variáveis, o algoritmo escolhido neste trabalho foi o *Junction Tree*, descrito no item 4.9. Tendo-se a estrutura da RB com seus parâmetros estimados, as probabilidades dos estados do nó *Causas* podem ser inferidas pelo algoritmo *Junction Tree*, dado o que é observado nos demais nós (ocorrências). Como exemplo explicativo, a Tabela 5-4 ilustra inferências de probabilidades dos estados da variável *Causa*, dado o que havia sido observado nas variáveis de entrada de 6 eventos.

Tabela 5-4 – Inferência de estados da variável *Causa* pela RB.

	Vegetal	Descarga Atmosférica	Falha no Componente	Clima Adverso	Sobrecarga ...
Ocorrência 1	0%	95%	5%	0%	0%
Ocorrência 2	0%	100%	0%	0%	0%
Ocorrência 3	0%	0%	100%	0%	0%
Ocorrência 4	40%	0%	0%	60%	0%
Ocorrência 5	0%	70%	0%	0%	30%
Ocorrência 6	0%	10%	80%	10%	0%

Também é possível visualizar esse resultado ao plotar-se um gráfico da Tabela 5-4, obtendo-se a Figura 5-14.

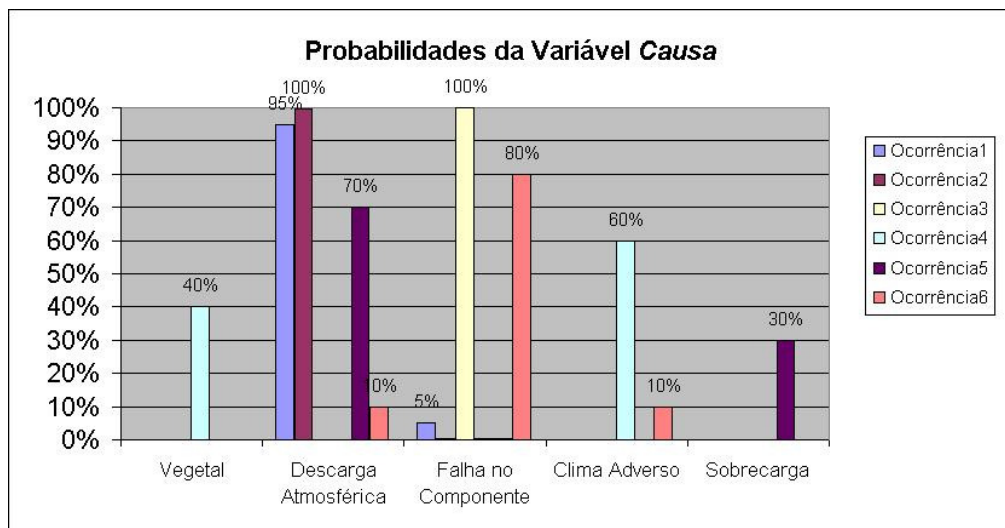


Figura 5-14 – Ilustração da saída da RB para determinadas ocorrências.

Observa-se que a saída da RB para cada ocorrência é normalizada. Para determinar a causa inferida em cada evento, neste trabalho será considerada a maior resposta de probabilidade da Rede Bayesiana para apontar determinado tipo de falta de energia. Assim, para a ocorrência 4, por exemplo, *Clima Adverso* seria a causa selecionada como saída do modelo, por ter a probabilidade de 60% de ter ocorrido.

5.5.4 Análise da saída da Rede Bayesiana

Após o treinamento de uma parcela de 50% dos 15.994 eventos, contendo 7.994 registros, foi obtido o modelo da RB que servirá de ponto de partida para analisar a saída de quaisquer eventos. Como forma de ilustrar a topologia da RB com seus parâmetros estimados, utilizando-se o software Netica®, a Figura 5-15 apresenta as probabilidades *a priori* de cada estado da Rede Bayesiana.

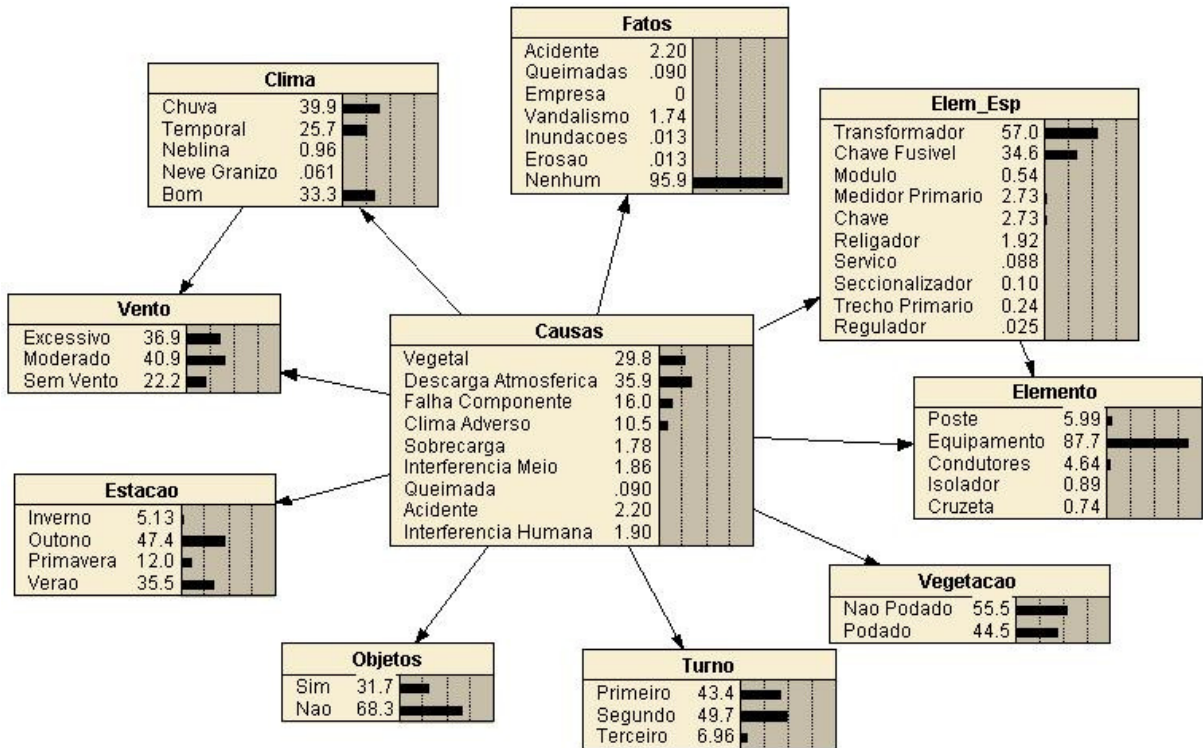


Figura 5-15 – Rede Bayesiana com as probabilidades de cada estado *a priori*, segundo os parâmetros estimados com 7.994 eventos.

Utilizando-se do mesmo conjunto de dados para treino, foi realizada a inferência das causas dado o que foi observado nos demais nós nesse conjunto de dados, em que foram obtidos os gráficos da Figura 5-16.

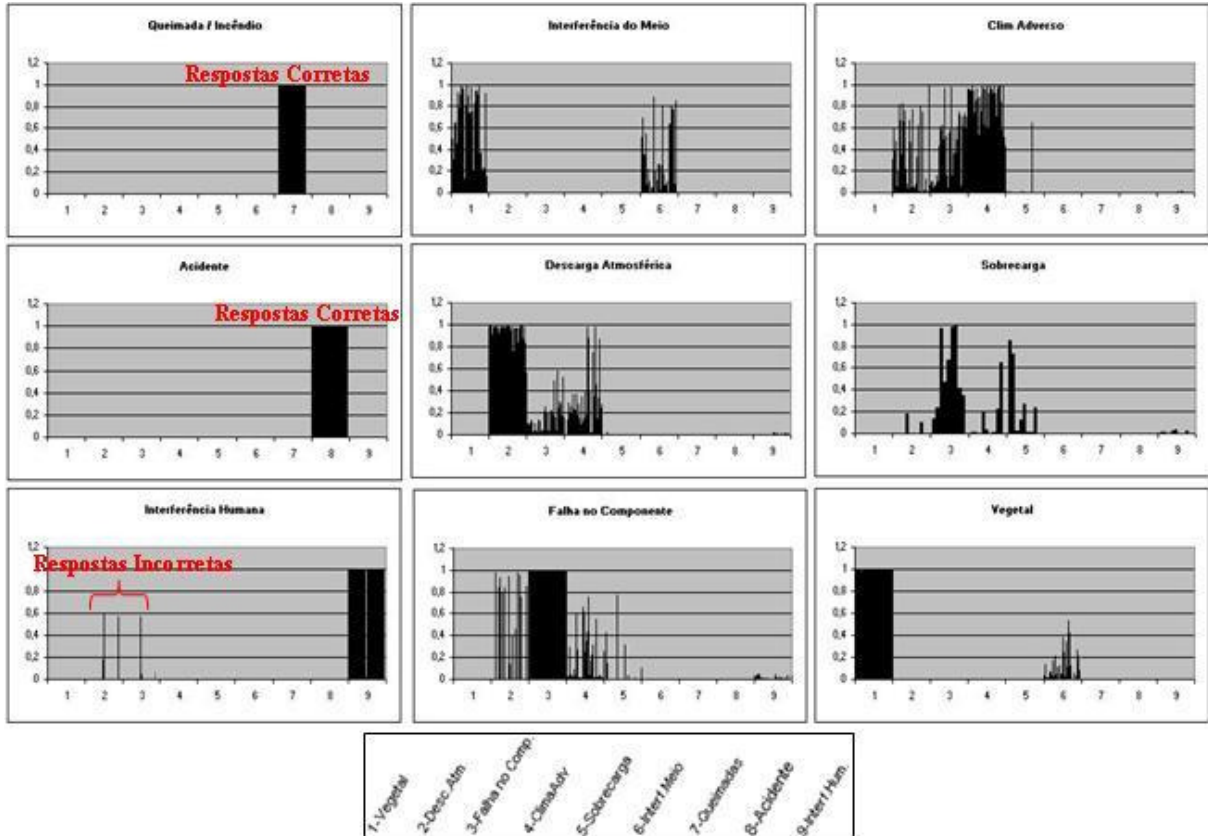


Figura 5-16 – Inferências das causas para os 7.994 eventos.

Na figura, o eixo das ordenadas representa a probabilidade *a posteriori* de ocorrência da causa e o eixo das abscissas corresponde à numeração designada para cada causa de interrupção. Observa-se que, mesmo inferindo-se (estimando-se) as causas a partir do mesmo conjunto de dados de treino, ocorreram erros na RB. Levando em conta o critério de maior valor normalizado para a resposta, verifica-se que, para as causas *Queimada/Incêndio* e *Acidente*, todas as inferências resultaram na resposta correta, com a probabilidade de 100% de os eventos corresponderem aos respectivos desligamentos. A causa *Interferência Humana* apresentou 3 resultados de inferências erradas. Já as demais inferências apontaram causas incorretas em número bastante superior a essa. Assim, levando em consideração a baixa incerteza sobre os desligamentos *Queimada/Incêndio*, *Acidente* e *Interferência Humana*, foi optado que sempre que houver detecção desses tipos de faltas pela RB, o filtro fuzzy não interferirá no diagnóstico. Esta regra irá compor o filtro adicional na saída dos blocos fuzzy e RB, vide Figura 5-3.

5.6 Implementação do sistema fuzzy

Para esta etapa, o Fuzzy Logic ToolboxTM, do *software* Matlab®, foi escolhido por apresentar-se no mesmo ambiente do modelo de Rede Bayesiana, podendo interagir com os resultados da rede de forma a avaliar o resultado mais provável. Dispõe de uma interface gráfica simples para a construção de modelos fuzzy, possibilitando o uso de um editor de regras, diversas configurações e métodos de inferência. Oferece, ainda, recursos de aprendizado neurofuzzy, técnicas de classificação de dados e habilidade para compilar o sistema em linguagem C. A criação do modelo, basicamente, envolve a definição das variáveis de entrada e saída, suas funções de pertinência e a base de regras.

5.6.1 Estabelecimento das Variáveis

As variáveis de entrada e seus respectivos estados a serem modelados no sistema fuzzy são compostas por todas as definidas em 5.2, possibilitando que o máximo de informações da planilha que caracteriza as faltas de energia possam ser extraídas a fim de indicar a causa mais provável.

Para modelar o sistema, foram considerados os desligamentos com maior incerteza sobre a RB, fornecendo um suporte de inferências por meio da base de conhecimentos. Os diferentes desligamentos contêm as quatro maiores causas da base de dados, conforme Figura 5-12, compondo os seguintes:

- *descarga atmosférica;*
- *vegetal;*
- *falha no componente;*
- *clima adverso;*
- *interferência do meio;*
- *sobrecarga.*

O sistema proposto possui a topologia apresentada na Figura 5-17.

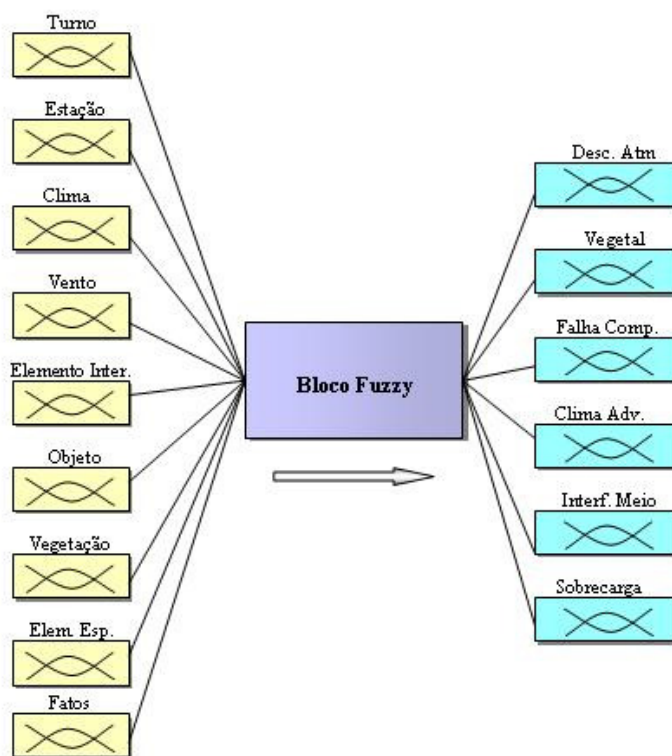


Figura 5-17 – Topologia do Sistema fuzzy com entradas e saídas

5.6.2 Funções de Pertinência

Para a escolha das funções de pertinência visando obter a aproximação de um modelo de inferências, foram estabelecidas as funções do tipo *Gaussian*, devido à característica de mapear o grau de pertinência do universo de discurso de maneira suave e contínua. Conforme afirmado na seção 3.2.2, as funções *Gaussian* e *Bell* são um dos formatos mais usuais para especificar conjuntos fuzzy. Para as variáveis de entrada, cada estado foi definido por uma função de pertinência, de modo a abranger o universo de discurso em um intervalo de zero a um. A Figura 5-18 ilustra a variável *Clima*, com os estados *Neblina*, *Bom*, *Chuva* e *Temporal*.

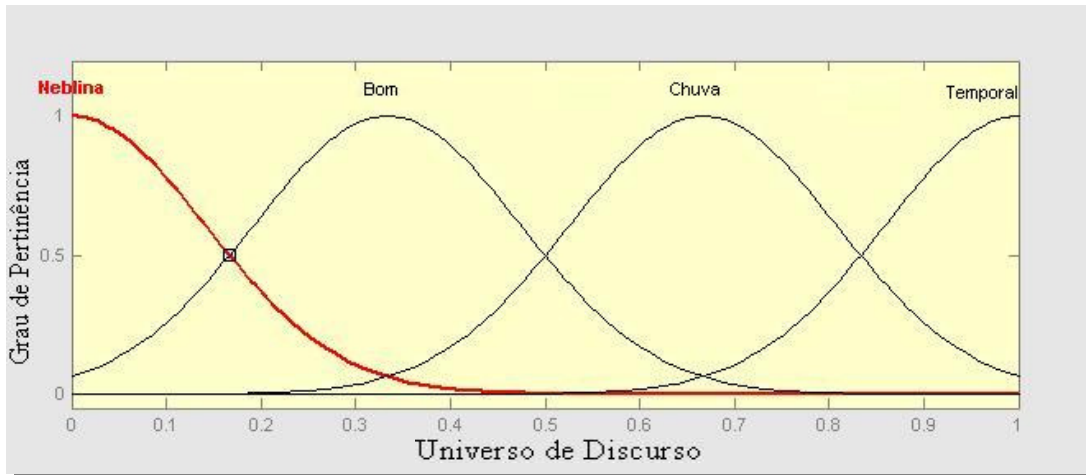


Figura 5-18 – Funções de pertinência correspondendo aos estados da variável *Clima*.

Para a saída do bloco fuzzy, cada causa foi relacionada com três funções de pertinência do tipo *Gaussian*, que indicam a possibilidade da ocorrência de uma causa ocorrer através dos valores semânticos: *Baixa*, *Média* e *Alta*, como mostrado na Figura 5-19.

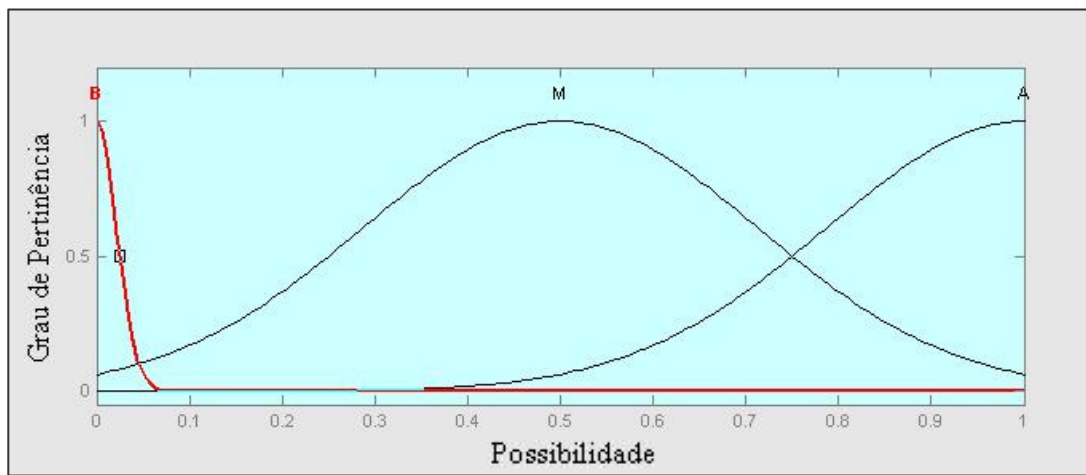


Figura 5-19 - Funções de pertinência de determinada causa, contendo as possibilidades de ocorrência correspondentes a: B-*Baixa*, M-*Média* e A-*Alta*.

Desse modo, a saída com um valor próximo de 1,0 indica alta possibilidade de ocorrência de determinada causa de interrupção. Observa-se que alguns ajustes foram necessários, como no conjunto fuzzy *Baixo* (B), onde forçou-se a resposta de maneira a concentrar um valor de saída muito baixo (próximo de zero).

5.6.3 Conversão dos Valores de Entrada do Bloco fuzzy

Através do universo de discurso de cada variável de entrada, a lógica fuzzy permite a entrada de dados imprecisos. Para o exemplo da variável *Clima*, se o eletricista possuir certeza de que houve predominantemente chuva no momento da interrupção, o valor de entrada associado a este estado é o número do universo de discurso cujo grau de pertinência do conjunto *Chuva* fique tendendo a um de acordo com a disposição da função de pertinência adotada. Por exemplo, adotando-se quatro conjuntos fuzzy igualmente espaçados em um universo de discurso de 0 a 1, a conversão da incerteza do eletricista para o estado correspondente de cada variável poderi ser norteadada através de uma tabela de conversão, conforme ilustra a Tabela 5-5 para a variável *Clima*.

Tabela 5-5 – Graduação exemplificada dos diferentes estados da variável *Clima*, com o valor do universo de discurso correspondente.

	Nebolina	Bom	Chuva	Temporal
Nebolina	0,00	0,15	–	–
Bom	0,15	0,33	0,50	–
Chuva	–	0,50	0,68	0,87
Temporal	–	–	0,87	1,00

Se o eletricista verifica que o clima possui instabilidade, observando clima bom mas eventualmente com pancadas de chuva, a tabela traduz esta incerteza através do valor 0,50 do universo de discurso. Conforme pode ser verificado na Figura 5-18, este valor foi designado por meio do ponto de interseção dos conjuntos fuzzy adjacentes: Bom e Chuva. Vale assinalar que este número deve estar de acordo com a resposta de provável desligamento perante outras variáveis de entrada.

Como neste trabalho utilizam-se os registros da concessionária com dados de entrada com estados únicos, será arbitrado que essas entradas sejam respostas dos técnicos com elevado grau de certeza. Ou seja, para cada variável, se considera que os estados recebam a mínima influência dos outros estados da mesma variável. Sendo assim, os valores de entrada do sistema fuzzy foram associados aos registros de maneira que o grau de pertinência de cada

conjunto corresponda a um. Esses valores correspondem aos elementos da diagonal principal da Tabela 5-5, dos quais não sofrem a influência de outros estados.

5.6.4 Base de Regras

No conjunto de dados obtidos na seção 5.4.1 foi possível explicitar as principais ocorrências das causas de interrupção constadas nos 26.282 eventos. Realizando o mesmo procedimento para o conjunto de 7.994 registros de treino da RB, criados em 5.5.4, os eventos foram dispostos em modo decrescente de frequência absoluta das ocorrências. A Figura 5-20 ilustra um exemplo utilizando os dados referentes à causa *Descarga Atmosférica*.

Causa	Clima	Fatos	Vegetação	Objetos	Elem.Inter.	Vento	Turno	Estação	Elem. Esp.	Qtdd
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	2oTurno	verao	Chave Fusivel	152
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	2oTurno	verao	Transformador	151
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	2oTurno	verao	Chave Fusivel	132
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	outono	Transformador	124
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	verao	Transformador	103
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	outono	Chave Fusivel	95
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	1oTurno	verao	Chave Fusivel	93
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	verao	Chave Fusivel	87
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	2oTurno	outono	Chave Fusivel	80
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	1oTurno	verao	Transformador	78
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	2oTurno	verao	Transformador	74
Descarga Atmosférica	Temporal	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	verao	Chave Fusivel	66
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	primavera	Transformador	61
Descarga Atmosférica	Temporal	*	Não-Podado	Não	Equipamento	Excessivo	2oTurno	verao	Transformador	59
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	2oTurno	outono	Transformador	58
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	2oTurno	outono	Chave Fusivel	58
Descarga Atmosférica	Temporal	*	Não-Podado	Não	Equipamento	Excessivo	1oTurno	verao	Chave Fusivel	55
Descarga Atmosférica	Chuva	*	Podado	Não	Equipamento	Moderado	1oTurno	primavera	Chave Fusivel	54
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	1oTurno	outono	Transformador	54
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	2oTurno	outono	Transformador	54
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	1oTurno	outono	Chave Fusivel	53
Descarga Atmosférica	Temporal	*	Podado	Não	Equipamento	Excessivo	2oTurno	verao	Chave Fusivel	51
Descarga Atmosférica	Temporal	*	Podado	Não	Equipamento	Excessivo	1oTurno	verao	Chave Fusivel	44
Descarga Atmosférica	Chuva	*	Não-Podado	Não	Equipamento	Moderado	1oTurno	primavera	Chave Fusivel	43
Descarga Atmosférica	Temporal	*	Podado	Não	Equipamento	Excessivo	1oTurno	primavera	Transformador	41

Figura 5-20 – Planilha contendo os estados em destaque para a ocorrência de *Descarga Atmosférica*.

Os campos em laranja, partindo do topo da lista, são os estados que ocorrem em maior número e caracterizam o desligamento. Em razão disso, foram selecionados para comporem as regras. Por meio dessas regras, é possível assinalar que, se determinado evento apresenta combinações de entrada com esses estados, isto sugere ao menos alguma possibilidade em resultar desligamento devido à saída com o estado *Descarga Atmosférica*. Realizando o mesmo procedimento para as demais causas foram estabelecidas 20 regras, ilustradas na Figura 5-21.

Regra	Clima	Vegetaçã	Objeto	Elemento	Vento	Turno	Estação	Elem. Esp.	Desc.Atm	FalhaCon	ClimaAdv	Sobrecar	Vegetal	InterfMeio
#1	Temporal	Não-Podad	Não	Equipame	Excessivo	2oTurno	outono	Transformador	B	B	A	B	B	B
#2	Chuva	Não-Podad	Não	Equipame	Moderado	2oTurno	verao	Chave Fusivel	A	B	B	B	B	B
#3	Bom	Não-Podad	Não	Equipame	Sem Vent	1oTurno	outono	Transformador	B	A	B	B	B	B
#4	Bom	Não-Podad	Não	Equipame	Sem Vent	2oTurno	outono	Transformador	B	B	B	A	B	B
#5	Bom	Não-Podad	Sim	Equipame	Sem Vent	2oTurno	outono	Transformador	B	B	B	B	A	B
#6	Bom	Não-Podad	Sim	Equipame	Sem Vent	1oTurno	outono	Chave Fusivel	B	B	B	B	B	A
#7	~Temporal	~Não-Podad	~Não	~Equipame	~Excessivo	~2oTurno	~outono	~Transformador	B	B	B	B	B	B
#8	~Chuva	~Não-Podad	~Não	~Equipame	~Moderado	~2oTurno	~verao	~Chave Fusivel	B	B	B	B	B	B
#9	~Bom	~Não-Podad	~Não	~Equipame	~Sem Vent	~1oTurno	~outono	~Transformador	B	B	B	B	B	B
#10	~Bom	~Não-Podad	~Não	~Equipame	~Sem Vent	~2oTurno	~outono	~Transformador	B	B	B	B	B	B
#11	~Bom	~Não-Podad	~Sim	~Equipame	~Sem Vent	~2oTurno	~outono	~Transformador	B	B	B	B	B	B
#12	~Bom	~Não-Podad	~Sim	~Equipame	~Sem Vent	~1oTurno	~outono	~Chave Fusivel	B	B	B	B	B	B
#13	Bom	-	-	Equipame	Excessivo	-	outono	-	B	B	M	B	B	B
#14	Temporal	-	-	Equipame	Excessivo	-	outono	-	B	B	A	B	B	B
#15	Chuva	-	-	Equipame	Moderado	-	verao	-	A	B	B	B	B	B
#16	Bom	-	-	Equipame	Sem Vent	-	outono	Transformador	B	A	B	B	B	B
#17	-	-	Sim	-	-	-	-	Transformador	B	B	B	B	A	B
#18	-	-	Sim	Equipame	Sem Vent	-	-	Chave Fusivel	B	B	B	B	B	A
#19	-	-	Sim	-	-	-	-	-	B	B	B	B	-	-
#20	-	-	Não	-	-	-	-	-	-	-	-	-	B	B

Figura 5-21 – Base de regras do sistema fuzzy.

As primeiras seis regras correspondem à primeira linha da planilha com frequências absolutas de cada causa, contendo o maior número de eventos com aquelas combinações para cada causa. Em seguida, as seis próximas regras correspondem à negação das anteriores, cujas saídas correspondem ao valor baixo. Desse modo buscou-se mapear os extremos do modelo. Por fim, as últimas oito regras estabelecem as variáveis-chave que preponderam nas respectivas saídas, extraídas conforme exemplo da Figura 5-20. Em cada regra, as variáveis de entrada e saída foram associadas com operadores AND.

5.6.5 Análise das saídas do sistema fuzzy

Para equiparar as saídas do sistema fuzzy com as respostas da RB, as saídas fuzzy foram multiplicadas por 100, de modo a obter-se respostas com valores de 0 à 100. Ao contrário da RB, cuja saída de cada ocorrência está na forma normalizada, no sistema fuzzy proposto isto não ocorre. O critério de resposta da causa inferida por este sistema não será unicamente o maior valor de saída. Além de o valor ser o maior entre as saídas, ele deverá atingir um nível mínimo de aceitação para ser designado como a respectiva causa de desligamento. Este nível mínimo será chamado de valor crítico. A comparação de um valor de referência com a saída do diagnóstico de faltas já foi abordada em Xu(2006), em que a referência permitiu estabelecer um grau indicando a possibilidade de ocorrência de cada causa.

Para determinar o valor crítico em cada causa, utilizando-se os registros de dados de treino de cada causa mais recorrentes - localizados heurísticamente no topo da lista de frequências absolutas conforme exemplo da Figura 5-20 - foram verificadas as saídas do sistema fuzzy. Utilizando-se a função *round* para arredondar as respostas para um número inteiro, foram obtidos os valores da Tabela 5-6. (Ver seção 6.1.5).

Tabela 5-6 – Valores críticos obtidos para cada causa na saída do sistema fuzzy.

Causa	Valor Crítico
Descarga Atmosférica	78
Falha no Componente	82
Clima Adverso	87
Sobrecarga	76
Vegetal	74
Interferência do Meio	82

Para os eventos que não atingirem o valor crítico, a inferência do sistema fuzzy se abstém, restando para a atribuição da Rede Bayesiana a resolução da causa. Da mesma forma que foi atribuído um critério de resposta inferida para a RB, a regra de determinação para a resposta fuzzy foi incluída no filtro adicional da Figura 5-3.

5.7 Análise do sistema integrado

O erro de inferência nos sistemas foi mensurado conforme a relação:

$$Erro = \frac{N^{\circ} de Inferências Erradas}{N^{\circ} de Eventos Considerados} \times 100(\%) \quad (5-2)$$

Observa-se que foi utilizado o método *split-half* para determinar os parâmetros da Rede Bayesiana, dividindo-se os dados disponíveis em duas partes para realizar o treinamento de um conjunto e inferir com outro conjunto de dados. Por isso, para fins de elaboração e verificação dos sistemas, os 15.994 eventos foram divididos em um conjunto para treino e outro para inferência. Procurou-se realizar essa divisão de modo que cada um das nove causas de desligamento ficasse separada em duas partes. Com isso evita-se o risco de uma causa pouco frequente não constar no conjunto de treino. Assim, foram separados 7.994 eventos para as seguintes tarefas: treino da RB, estabelecimento da base de regras do sistema fuzzy e análise das saídas desses sistemas. Os 8.000 eventos restantes foram utilizado para a estimação das causas (inferências) do sistema integrado (RB e fuzzy).

O fluxograma da Figura 5-22 ilustra as principais etapas do sistema proposto.

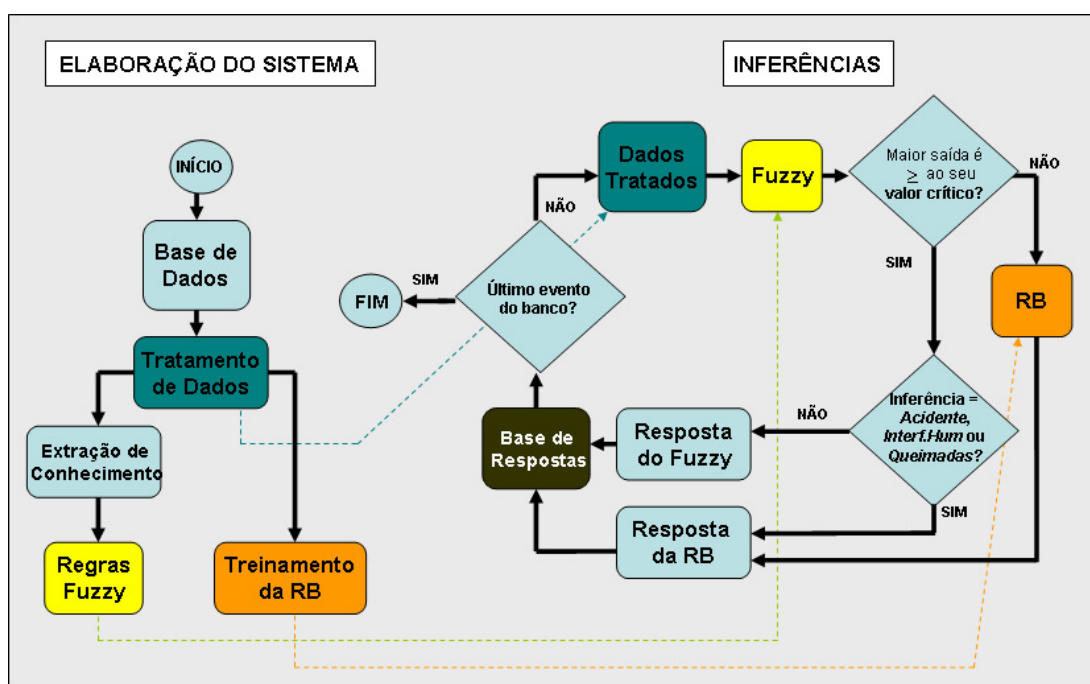


Figura 5-22– Fluxograma do tratamento de dados e inferência do sistema integrado.

Conforme resumido pela Figura 5-22, na etapa de tratamento de dados, descrita em 5.3, foram realizados os processos de limpeza, mineração, análise e adequação da base de dados. Utilizando o método *split-half*, 7.994 registros foram destinados para o treinamento do bloco RB e extração de conhecimento para o bloco fuzzy. Através da análise das saídas dos sistemas com esses registros, foram definidos valores críticos fuzzy e causas com baixa incerteza da RB. Esses critérios de análise constituem um filtro formado pelos blocos de

decisão da Figura 5-3. A outra metade dos dados tratados, constituindo 8.000 eventos, foi utilizada para verificar o sistema de inferências proposto, em que foram obtidos os resultados que serão apresentados na próxima seção.

6 RESULTADOS

6.1 Resultados Avaliados

Por meio dos modelos configurados através das funções do Matlab®, para a RB, e do Fuzzy Logic Toolbox™, para a lógica fuzzy, foram utilizadas as bases de dados a fim de verificar as hipóteses deste trabalho. Para relacionar a eficácia dos métodos, foram calculados os erros perante as técnicas propostas a fim de realizar-se conclusões.

6.1.1 Análise de agrupamento de dados

As observações através do agrupamento de dados na forma de *clusters* nos 26.282 eventos permitiram algumas constatações que caracterizam os desligamentos. Os dados da causa *Sobrecarga* foram subdivididos em *clusters* segundo os estados da variável *Clima*, o que permitiu analisar a proporção de que aproximadamente 70% deste tipo de desligamento ocorreu em presença de tempo *bom*, conforme a Figura 6-1.

Causa	Clima	Fatos	Vegetação	Objetos	Elem. Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	2oTurno	outono	Transforma	231
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	2oTurno	outono	Transforma	90
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	2oTurno	verao	Transforma	25
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	1oTurno	outono	Transforma	24
Sobrecarga	Bom	* ■	Não-Podado	Não	Equipame	Sem Vento	1oTurno	outono	Chave Fus	17
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	2oTurno	verao	Transforma	15
Sobrecarga	Bom	* ■	Não-Podado	Não	Equipame	Sem Vento	3oTurno	outono	Transforma	11
Sobrecarga	Bom	* ■	Não-Podado	Não	Condutore	Sem Vento	1oTurno	verao	Medidor P	4
Sobrecarga	Bom	* ■	Podado	Não	Equipame	Sem Vento	3oTurno	outono	Transforma	3
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	1oTurno	outono	Transforma	6
Sobrecarga	Bom	*	Podado	Não	Equipame	Excessivo	2oTurno	outono	Transforma	5
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	primavera	Transforma	5
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	1oTurno	outono	Chave Fus	4
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	3oTurno	verao	Transforma	4
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	verao	Chave Fus	3
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Sem Vento	2oTurno	outono	Medidor P	3
Sobrecarga	Bom	*	Podado	Não	Equipame	Excessivo	2oTurno	outono	Chave Fus	2
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	1oTurno	primavera	Chave Fus	2
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Excessivo	1oTurno	verao	Medidor P	2
Sobrecarga	Chuva	* ■	Não-Podado	Não	Equipame	Moderado	2oTurno	outono	Transforma	28
Sobrecarga	Chuva	* ■	Não-Podado	Não	Equipame	Moderado	1oTurno	outono	Transforma	20
Sobrecarga	Chuva	*	Não-Podado	Não	Condutore	Moderado	2oTurno	outono	Medidor P	2
Sobrecarga	Temporal	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	verao	Transforma	2
Sobrecarga	Temporal	*	Não-Podado	Não	Equipame	Excessivo	3oTurno	verao	Transforma	1
Sobrecarga	Temporal	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	outono	Transforma	1

}

Clima Bom
69,48%

}

Clima Ruim
30,52%

Figura 6-1 - Eventos de sobrecarga agrupados por clima para os 26.282 eventos.

Outra observação pôde ser obtida subdividindo-se aqueles eventos com clima Bom segundo os estados da variável *Turno*. Foi verificada a proporção que 82% de causas por *Sobrecarga* ocorrem no período vespertino.

Causa	Clima	Fatos	Vegetação	Objetos	Elem. Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	1oTurno	outono	Transformador	24
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	1oTurno	outono	Chave Fusivel	17
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	outono	Medidor Primario	9
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	1oTurno	outono	Transformador	6
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	1oTurno	outono	Chave Fusivel	4
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	verao	Medidor Primario	4
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	1oTurno	primavera	Chave Fusivel	2
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Excessivo	1oTurno	verao	Medidor Primario	2
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	2oTurno	outono	Transformador	231
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	2oTurno	outono	Transformador	90
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	2oTurno	verao	Transformador	25
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	2oTurno	verao	Transformador	15
Sobrecarga	Bom	*	Podado	Não	Equipame	Excessivo	2oTurno	verao	Transformador	7
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	2oTurno	outono	Chave Fusivel	7
Sobrecarga	Bom	*	Podado	Não	Equipame	Excessivo	2oTurno	outono	Transformador	5
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	primavera	Transformador	5
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	verao	Transformador	4
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	outono	Transformador	4
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Excessivo	2oTurno	verao	Chave Fusivel	3
Sobrecarga	Bom	*	Não-Podado	Não	Condutore	Sem Vento	2oTurno	outono	Medidor Primario	3
Sobrecarga	Bom	*	Podado	Não	Equipame	Excessivo	2oTurno	outono	Chave Fusivel	2
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	3oTurno	outono	Transformador	11
Sobrecarga	Bom	*	Não-Podado	Não	Equipame	Sem Vento	3oTurno	verao	Transformador	4
Sobrecarga	Bom	*	Podado	Não	Equipame	Sem Vento	3oTurno	outono	Transformador	3

}

1oTurno
14,0%

}

2oTurno
82,3%

}

3oTurno
3,7%

Figura 6-2 – Eventos de sobrecarga agrupados por turno para os 26.282 eventos.

Essas informações estão de acordo com a referência Brown (2002), na qual é colocado que as interrupções por sobrecarga são mais comuns em dias ensolarados e em

horários de maior consumo (o que foi interpretado como tempo bom e segundo turno no banco de dados). Observa-se também que, mesmo após a adequação dos registros para desconsiderar eventos com mais de uma saída, restando 15.994 eventos, a maioria desses *clusters* continuaram conferindo com a afirmação de Brown (2002), conforme verifica-se na Figura 6-3.

(a)

Causa	Clima	Fatos	Vegetação	Objetos	Elem.Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.	
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Excessivo	1oTurno	verao	Medidor Prima	2	} Clima Bom 96,5%
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	3oTurno	verao	Transformador	4	
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	2oTurno	verao	Transformador	25	
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	2oTurno	outono	Transformador	231	
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	outono	Medidor Prima	9	
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	verao	Medidor Prima	4	
Sobrecarg	Chuva	*	Não-Podado	Não	Condutore	Moderado	2oTurno	verao	Medidor Prima	3	} Clima Ruim 3,5%
Sobrecarg	Neblina	*	Não-Podado	Não	Equipamer	Moderado	1oTurno	outono	Transformador	4	
Sobrecarg	Neblina	*	Não-Podado	Não	Equipamer	Moderado	3oTurno	outono	Transformador	3	

(b)

Causa	Clima	Fatos	Vegetação	Objetos	Elem.Int.	Vento	Turno	Estação	Elem. Esp.	Qtde.	
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Excessivo	1oTurno	verao	Medidor Prima	2	} 1oTurno - 5,5%
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	outono	Medidor Prima	9	
Sobrecarg	Bom	*	Não-Podado	Não	Condutore	Sem Vento	1oTurno	verao	Medidor Prima	4	} 2oTurno - 93,1%
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	2oTurno	verao	Transformador	25	
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	2oTurno	outono	Transformador	231	
Sobrecarg	Bom	*	Não-Podado	Não	Equipamer	Sem Vento	3oTurno	verao	Transformador	4	} 3oTurno - 1,4%

Figura 6-3 – Análise de *clusters* conforme o *Clima* (a) e *Turno* (b) nos 7.994 eventos.

Considerando-se que os registros mantiveram informações coerentes após o tratamento de dados, esse conhecimento pode ser levado para a construção das regras do modelo fuzzy tendo em vista o diagnóstico de interrupções.

6.1.2 Resultados dos algoritmos de inferência

Para comparar a eficácia da inferência pela RB com os algoritmos *Junction Tree* e *Eliminação de Variáveis*, foi utilizado o conjunto de dados de treino, contendo os 7.994 eventos. Após o treinamento, o mesmo conjunto de registros foi utilizado para a inferência, através desses algoritmos. Os erros e o tempo da execução dos mesmos foram avaliados, em que são mostrados na Tabela 6-1.

Tabela 6-1 – Erro e tempo de execução dos algoritmos da RB.

Algoritmo	Inferências erradas	Erro	Tempo de execução
<i>Eliminação de Variáveis</i>	402 eventos	5,02%	1h e 29min
<i>Junction Tree</i>	402 eventos	5,02%	11min

As funções do Matlab® correspondem a *var_elim_inf_engine* e *jtree_inf_engine*. A função que realiza o treinamento através do algoritmo EM (*expectation maximization*) é a *learn_params_em* do Toolbox BNT, em que foi utilizado o número máximo de 10 iterações e limiar (*threshold*) com o valor padrão 0,001. O tempo de execução foi calculado com a função *cputime*, que calcula o tempo com referência à Unidade de Processamento (CPU), em segundos. Verificando-se que a precisão do erro para os dois algoritmos é igual mas a velocidade de execução do *Junction Tree* muito superior, os resultados que apresentam-se a seguir foram realizados por meio deste algoritmo e mesmas configurações.

6.1.3 Efeito da etapa de adequação na base de dados

De maneira a medir o impacto da eliminação de variáveis múltiplas, esta verificação foi realizada treinando-se e inferindo-se um mesmo conjunto de dados. Isto foi abordado com um conjunto de eventos de saídas múltiplas e com outro de saídas únicas. Para fazer uma comparação consistente, primeiro os 26.282 eventos, contendo saídas múltiplas, foram reduzidos para formar um conjunto com 15.994 eventos contendo 10.288 registros com saídas múltiplas. Desse modo, obtiveram-se dois conjuntos de 15.994 registros: um com saídas únicas e outro com saídas múltiplas. Para isso, teve-se o cuidado de manter a mesma proporção de registros de causas, em ambos conjuntos. A Figura 6-4 ilustra o procedimento descrito.

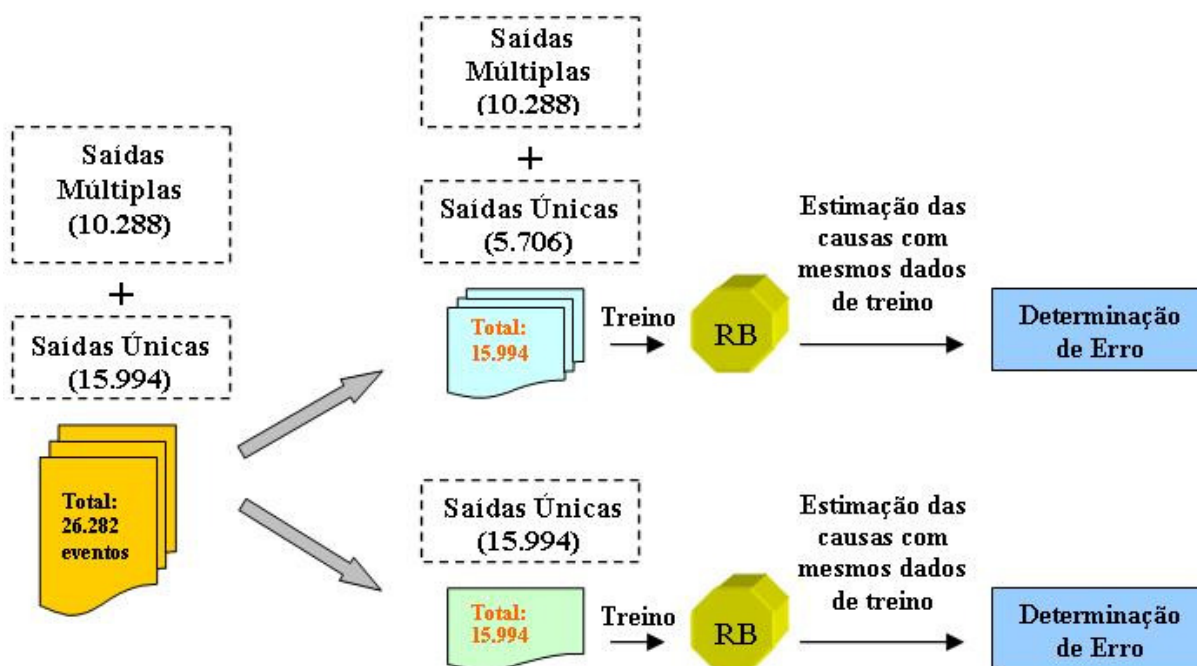


Figura 6-4 – Simulações utilizando-se base de dados com saídas múltiplas e únicas.

Os erros obtidos apresentam-se na tabela Tabela 6-2.

Tabela 6-2 – Erros na RB conforme o número de saídas em relação a mesmas entradas.

Base de dados	Inferências erradas	Erro
Contendo Saídas Múltiplas (15.994 eventos)	2.550 eventos	15,94%
Saídas únicas (15.994 eventos)	736 eventos	4,60%

Para os dados apresentados, verifica-se uma redução de 11,34% de erro ao excluir eventos com saídas múltiplas. Para as respostas de inferências foi utilizado o critério de maior probabilidade de causa, para cada evento.

6.1.4 Caracterização da base de dados

Visando quantificar a contribuição para a inferência correta ao adicionar as variáveis de entrada *Turno*, *Estação* e *Elemento Específico*, primeiramente foram utilizados dois

conjuntos de dados: um com a inclusão das três variáveis e outro não as incluindo. Ambos conjuntos apresentando saídas únicas em relação as suas entradas continham 85 eventos, conforme ilustra a Figura 6-5.

Causa	Clima	Fatos	Vegetaçã	Objetos	Elem.Int.	Vento
Clima Adv	Bom	*	Não-Podad	Não	Equipamer	Excessivo
Clima Adv	Bom	*	Não-Podad	Não	Isolador	Excessivo
Clima Adv	Bom	*	Podado	Não	Condutore	Excessivo
Clima Adv	Bom	*	Podado	Não	Equipamer	Excessivo
Clima Adv	Bom	*	Podado	Não	Isolador	Excessivo
Clima Adv	Chuva	Erosão	Podado	Não	Equipamer	Moderado
Clima Adv	Tempo	*	Não-Podad	Não	Cruzeta	Excessivo
Clima Adv	Tempo	*	Não-Podad	Não	Isolador	Excessivo
Clima Adv	Tempo	*	Podado	Não	Condutore	Excessivo
Clima Adv	Tempo	*	Podado	Não	Cruzeta	Excessivo
Clima Adv	Tempo	*	Podado	Não	Isolador	Excessivo
Clima Adv	Tempo	Inundação	Não-Podad	Não	Equipamer	Excessivo
Inundação	Chuva	*	Podado	Não	Equipamer	Moderado

(a)

Causa	Clima	Fatos	Vegetaçã	Objetos	Elem.Int.	Vento	Turno	Estação	Elem.Esp.
Clima Adv	Bom	*	Não-Podad	Não	Equipamer	Excessivo	2oTurno	outono	Chave Fusivel
Clima Adv	Bom	*	Não-Podad	Não	Isolador	Excessivo	2oTurno	outono	Transformador
Clima Adv	Bom	*	Podado	Não	Condutore	Excessivo	2oTurno	outono	Chave Fusivel
Clima Adv	Bom	*	Podado	Não	Equipamer	Excessivo	2oTurno	primavera	Chave Fusivel
Clima Adv	Bom	*	Podado	Não	Isolador	Excessivo	2oTurno	verao	Transformador
Clima Adv	Chuva	Erosão	Podado	Não	Equipamer	Moderado	2oTurno	outono	Transformador
Clima Adv	Tempo	*	Não-Podad	Não	Cruzeta	Excessivo	2oTurno	inverno	Chave Fusivel
Clima Adv	Tempo	*	Não-Podad	Não	Isolador	Excessivo	2oTurno	inverno	Chave Fusivel
Clima Adv	Tempo	*	Podado	Não	Condutore	Excessivo	2oTurno	outono	Chave
Clima Adv	Tempo	*	Podado	Não	Cruzeta	Excessivo	2oTurno	verao	Chave Fusivel
Clima Adv	Tempo	*	Podado	Não	Isolador	Excessivo	3oTurno	verao	Chave Fusivel
Clima Adv	Tempo	Inundação	Não-Podad	Não	Equipamer	Excessivo	1oTurno	outono	Transformador
Inundação	Chuva	*	Podado	Não	Equipamer	Moderado	2oTurno	verao	Chave Fusivel

(b)

Figura 6-5 – Conjuntos de dados com 85 eventos de saídas únicas - (a) sem as três variáveis envolvidas; (b) incluindo as três variáveis.

A Tabela 6-3 apresenta a relação de causas utilizada nas bases de dados e o resultado de suas inferências na Rede Bayesiana.

Tabela 6-3 – Erros para diferentes números de variáveis de entradas em registros de dados sem repetição.

Causas	Quantidade	Inferências Erradas	
		Sem as 3 Variáveis	Com as 3 Variáveis
<i>Clima Adverso</i>	12	0	0
<i>Acidente</i>	8	0	0
<i>Descarga Atmosférica</i>	5	0	0
<i>Falha no Componente</i>	27	1	1
<i>Interferência Humana</i>	11	0	0
<i>Queimadas / Incêndio</i>	3	0	0
<i>Vegetal</i>	19	0	0
Total:	85	1	1
Erro:		1,17%	1,17%

Em princípio verificou-se que não houve diferença de resultados com a inclusão das variáveis. Entretanto, fazendo-se aleatoriamente a repetição de alguns eventos de forma recíproca em ambos conjuntos, isto é, nos mesmos registros, e procedendo com o treino e inferência dos mesmos, foram obtidos os resultados conforme a Tabela 6-4.

Tabela 6-4 – Erros para diferentes números de variáveis de entradas, em registros de dados com repetição.

Causas	Quantidade	Repetições	Subtotal	Inferências Erradas	
				Sem as 3 Variáveis	Com as 3 Variáveis
<i>Clima Adverso</i>	12	24	36	0	0
<i>Acidente</i>	8	0	8	0	0
<i>Descarga Atmosférica</i>	5	5	10	4	0
<i>Falha no Componente</i>	27	27	54	6	2
<i>Interferência Humana</i>	11	22	33	0	0
<i>Queimadas / Incêndio</i>	3	0	3	0	0
<i>Vegetal</i>	19	9	28	0	0
Total:	85	87	172	10	2
Erro:				5,81%	1,16%

Com a verificação do aumento de eventos inferidos incorretamente, foi possível observar que a proporção de eventos repetidos interfere no resultado do erro, sendo que esta interferência é atenuada ao detalhar mais os dados de entrada. Os resultados confirmam que a inclusão das variáveis de entrada caracterizam mais os eventos, contribuindo para a redução do erro da RB em 4,65% considerando-se os 172 registros.

6.1.5 Eficácia da base de regras fuzzy

A base de regras foi analisada utilizando-se a lista de frequência absoluta do conjunto de inferências (8.000 eventos, obtidos em 5.5.4) e verificando-se as saídas do filtro fuzzy, conforme o exemplo da Figura 6-6 para a causa *Vegetal*.

Banco de Dados			Saída Fuzzy					Resposta	Inferências	
Saída	Entradas	Qtde	Desc.Atm	FalhaCom	ClimaAdu	Sobrecar.	Vegetal	InterfMeit	Inferida	Corretas
Vegetal	Combinação 1	189	12,32433	74,781815	1,430298	37,10261	74,78182	1,408753	Vegetal	189
Vegetal	Combinação 2	132	52,98786	1,408753	1,408753	1,408753	74,78182	1,408753	Vegetal	132
Vegetal	Combinação 3	111	12,32433	74,781815	1,430298	37,07944	74,78182	1,408753	Vegetal	111
Vegetal	Combinação 4	100	12,32433	74,781815	1,430298	37,10261	74,78182	1,408753	Vegetal	100
Vegetal	Combinação 5	92	52,98786	1,408753	1,408753	1,408753	74,78182	1,408753	Vegetal	92
Vegetal	Combinação 6	91	12,32433	74,781815	1,430298	37,07944	74,78182	1,408753	Vegetal	91
Vegetal	Combinação 7	91	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 8	79	20,94864	1,408753	1,408753	1,408753	82,20493	1,408753	Vegetal	79
Vegetal	Combinação 9	73	12,32433	38,481297	35,93947	37,01759	82,20493	1,408753	Vegetal	73
Vegetal	Combinação 10	72	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 11	71	12,32433	38,481297	35,93947	37,01759	82,20493	1,408753	Vegetal	71
Vegetal	Combinação 12	54	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 13	52	20,92331	1,408753	1,408753	1,408753	82,20493	1,408753	Vegetal	52
Vegetal	Combinação 14	51	52,98786	1,408753	1,408753	1,408753	74,78182	1,408753	Vegetal	51
Vegetal	Combinação 15	46	20,92331	1,408753	1,408753	1,408753	82,20493	1,408753	Vegetal	46
Vegetal	Combinação 16	44	12,32433	38,481297	35,93947	37,01759	82,20493	1,408753	Vegetal	44
Vegetal	Combinação 17	36	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 18	36	20,94864	1,432038	1,430285	1,431292	82,20493	1,408753	Vegetal	36
Vegetal	Combinação 19	34	20,94864	1,408753	1,408753	1,408753	82,20493	1,408753	Vegetal	34
Vegetal	Combinação 20	33	52,98786	1,408753	1,408753	1,408753	1,408754	38,54068	Não_Infer.	0
Vegetal	Combinação 21	31	52,98786	1,408753	1,408753	1,408753	1,408754	38,54068	Não_Infer.	0
Vegetal	Combinação 22	6	12,32562	1,408753	36,08563	1,408753	2,734071	2,912784	Não_Infer.	0
Vegetal	Combinação 23	6	12,32562	1,432053	36,08563	1,431307	82,19746	1,408753	Vegetal	6
Vegetal	Combinação 24	5	20,92331	1,408753	1,408753	1,408753	1,408753	82,20493	Interf Meio	0
Vegetal	Combinação 32	5	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 33	5	12,32433	1,432053	1,430298	1,431307	64,21325	2,91267	Não_Infer.	0
Vegetal	Combinação 34	5	12,32433	1,432053	76,03679	1,431307	74,78182	1,408753	Não_Infer.	0
Vegetal	Combinação 35	5	52,98786	1,408753	1,408753	1,408753	1,408754	38,54068	Não_Infer.	0
Vegetal	Combinação 36	5	12,32433	1,432053	1,430298	1,431307	64,21325	2,91267	Não_Infer.	0
Vegetal	Combinação 37	5	20,94864	1,408753	1,408753	1,408753	82,20493	1,408753	Vegetal	5

Figura 6-6 – Resultados da saída do bloco fuzzy para causas *Vegetal*. As saídas com o máximo valor foram grifadas em negro, para cada ocorrência.

Na ilustração da Figura 6-6, verifica-se que os eventos mais numerosos da causa *Vegetal*, no topo da lista, têm em comum a saída **74,78182**. Por ser um valor que predomina em quantidade no conjunto de dados, foi o valor crítico atribuído de forma heurística, conforme a Tabela 5-6. Observa-se que, nos eventos não inferidos, os maiores valores de saída

(grifados em negrito) não atingiram os valores críticos correspondentes das outras causas. Isto significa que, conforme as regras heurísticas, as variáveis de entrada combinadas não possuem características suficientes para serem mapeadas para estimar uma saída com nível alto de confiança. Um mapa de regras do toolbox do Matlab® com as saídas para a combinação de entradas 5 pode ser verificado na Figura 6-7.

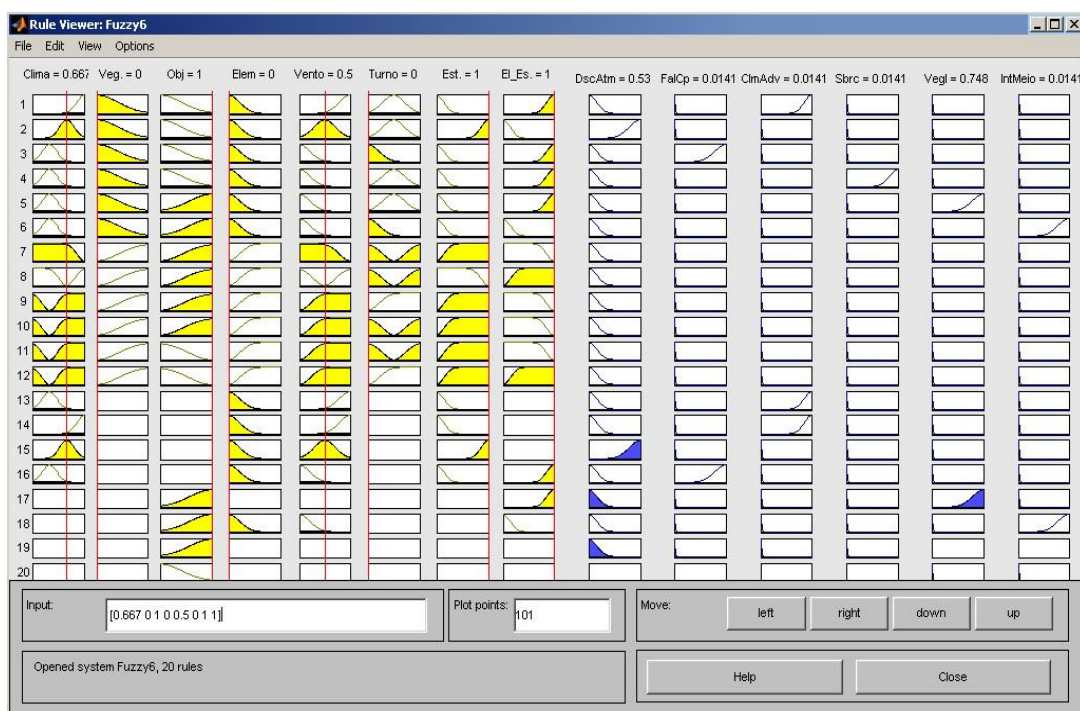


Figura 6-7 – Mapa de regras da combinação de entradas com a maior saída para a causa *Vegetal*, com o valor 0,748 (74,8%).

O mesmo foi realizado para as demais causas, registrando-se os números de inferências corretas e erradas, em que foram obtidos os resultados conforme a Tabela 6-5.

Tabela 6-5 – Resultados da saída do bloco fuzzy utilizando funções de pertinência Gaussianas.

Causa	Quantidade	Eventos Inferidos	Inferências Erradas	Percentual Inferido	Percentual de Erro
<i>Descarga Atmosférica</i>	2.867	1.054	88	36,80%	8,35%
<i>Vegetal</i>	2.384	1.553	22	65,10%	1,42%
<i>Falha no Componente</i>	1.281	282	52	22,00%	18,44%
<i>Clima Adverso</i>	838	420	0	50,10%	0,00%
<i>Interferência do Meio</i>	151	121	30	80,10%	24,79%
<i>Sobrecarga</i>	143	112	0	78,30%	0,00%
<i>Interferência Humana *</i>	153	-	-	-	-
<i>Acidente *</i>	176	-	-	-	-
<i>Queimadas / Incêndio *</i>	7	-	-	-	-
Total:	8.000	3.542	192	Média dos erros:	8,83%
Percentual de inferência total:			44,3%	Erro Total:	5,42%

* causa considerada com incerteza baixa, conforme seção 5.5.4.

Nas inferências foram utilizadas funções gaussianas, com o método de defuzzificação centróide. Nota-se que não foram consideradas as causas *Acidente*, *Queimadas/Incêndio* e *Interferência Humana*, conforme proposto em 5.5.4. O mesmo procedimento foi aplicado em um sistema fuzzy com valores críticos e base de regras iguais, mas com funções de pertinência triangulares, em que se verificam os resultados na Tabela 6-6.

Tabela 6-6 – Resultados da saída do bloco fuzzy utilizando funções de pertinência Triangulares.

Causa	Quantidade	Eventos Inferidos	Inferências Erradas	Percentual Inferido	Percentual de Erro
<i>Descarga Atmosférica</i>	2.867	1.054	88	36,80%	8,35 %
<i>Vegetal</i>	2.384	1.584	22	66,40%	1,39 %
<i>Falha no Componente</i>	1.281	282	52	22,00%	18,44 %
<i>Clima Adverso</i>	838	420	0	50,10%	0,00 %
<i>Interferência do Meio</i>	151	121	30	80,10%	24,79 %
<i>Sobrecarga</i>	143	112	0	78,30%	0,00 %
<i>Interferência Humana</i>	153	-	-	-	-
<i>Acidente</i>	176	-	-	-	-
<i>Queimadas / Incêndio</i>	7	-	-	-	-
Total:	8.000	3.573	192	Média dos erros:	8,83%
Percentual de inferência total:			47,6%	Erro Total Sistema fuzzy:	5,37 %

Observa-se nas tabelas Tabela 6-5 e Tabela 6-6 que as causas *Falha no Componente* e *Interferência do Meio* apresentaram os maiores erros. Isto significa que a base de regras criada não é propícia para caracterizar essas duas causas e um ajuste das regras poderá ser necessário para aperfeiçoar o modelo, atenuando os erros. Outra possibilidade é aumentar o nível de valor crítico aceitável para essas saídas, de modo a exigir maiores saídas do modelo fuzzy antes de estabelecê-las como resposta da causa estimada.

Comparando o resultado das diferentes funções de pertinência, a análise permite verificar que ambos os tipos de função apresentam resultados muito próximos, com diferença de erro total de 0,05%. Esta diferença foi ocasionada devido ao número de causas *Vegetal* inferidas corretamente ser 1,3% maior no modelo com funções triangulares. Esse resultado pode ser explicado pela função gaussiana *Alto* na saída *Vegetal*, cujo formato alongado ocasiona em valores de centro-de-área inferiores aos da função triangular. Na Figura 6-8, verifica-se o exemplo de um evento que teve diferentes respostas do sistema fuzzy conforme as funções de pertinência.

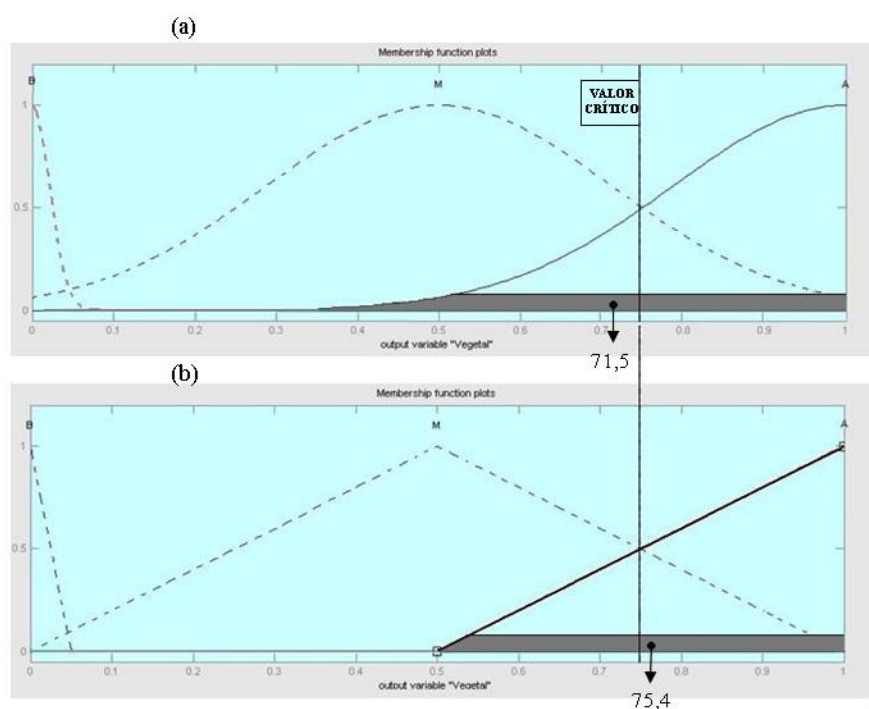


Figura 6-8 – Valores de saída pra causa vegetal, utilizando funções de pertinência gaussianas e triangulares.

Na Figura 6-8-(a), a região fuzzificada possui o centro-de-área (valor 71,5) localizado abaixo do nível crítico (74), não atingindo o critério de resposta confiável. Já na Figura 6-8-(b), o valor 75,4 é suficiente para determinar a causa como sendo vegetal. Contudo, é possível obter resultado semelhante com as funções gaussianas, ajustando-se o posicionamento da função da saída *Vegetal* mais à direita, de maneira que atenda adequadamente o critério do nível crítico. Portanto, considerando-se que a diferença entre os resultados das tabelas Tabela 6-5 e Tabela 6-6, não foi significativa, deu-se o prosseguimento da inferência do sistema integrado utilizando-se o pior caso, por meio do modelo com funções gaussianas.

6.1.6 Inferência do Sistema Integrado RB / fuzzy

O conjunto de dados formado pelos registros não inferidos no modelo fuzzy foi alocado como entrada no bloco da Rede Bayesiana, conforme a Figura 6-9.

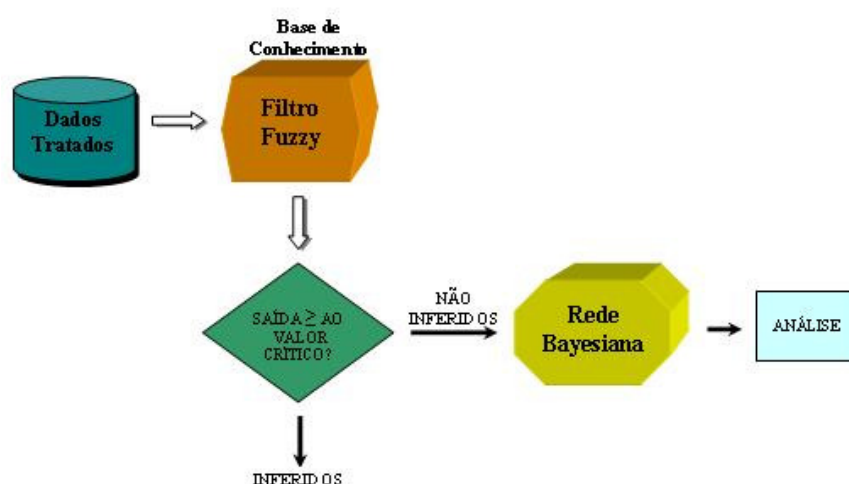


Figura 6-9 – Análise da saída da RB para registros de causas não inferidos do filtro fuzzy.

Da análise da saída da RB foram obtidos os resultados da Tabela 6-7.

Tabela 6-7 – Resultados da saída do bloco da Rede Bayesiana.

Causa	Quantidade	Eventos Inferidos	Inferências Erradas	Percentual Inferido	Percentual de Erro
<i>Descarga Atmosférica</i>	2.867	1.813	14	63,20%	0,77 %
<i>Vegetal</i>	2.384	831	0	34,90%	0,00 %
<i>Falha no Componente</i>	1.281	999	47	78,00%	4,70 %
<i>Clima Adverso</i>	838	418	165	49,90%	39,47 %
<i>Interferência do Meio</i>	151	30	23	19,90%	76,67 %
<i>Sobrecarga</i>	143	31	16	21,70%	51,61 %
<i>Acidente</i>	176	176	5	100,00%	2,84 %
<i>Interferência Humana</i>	153	153	13	100,00%	8,50 %
<i>Queimada/ Incêndio</i>	7	7	0	100,00%	0,00 %
Total:	8.000	4.458	283	Média dos erros:	20,5%
Percentual de inferência total:			55,7%	Erro Total Sist. RB:	6,35 %

Os percentuais de erro da Tabela 6-7 são inerentes aos algoritmos *EM* e *Junction Tree* da RB, não tendo-se encontrada uma explicação consistente para esses erros. Entretanto,

observa-se que os maiores erros, das causas *Clima Adverso*, *Interferência do Meio e Sobrecarga*, estão associados com o número de registros menor dessas causas. Isto é aceitável, uma vez que amostras de dados de treino menores implicam em menos especificações da maneira que as faltas ocorrem.

Verifica-se que para a inferência da causa *Acidente*, considerada isenta de erro nos dados de treino, foram constatadas 5 respostas erradas no conjunto de dados de inferência. Isto ocorre porque o estado *Elemento Específico* da variável de saída não havia permanecido no conjunto de dados de treino da RB, após a divisão dos registros pelo método *split-half*. Para a causa *Interferência Humana* era esperado um erro maior que 0%, pois apesar de ser considerada uma causa com baixa incerteza nos dados de treino, apresentou 3 faltas incorretas, conforme a seção 5.5.4.

A análise do erro total do sistema integrado fuzzy e Rede Bayesiana pode ser realizada calculando-se a razão entre as inferências erradas e os eventos inferidos totais dos sistemas. Nessa relação, verifica-se na Tabela 6-8 um erro de 5,94%.

Tabela 6-8 – Erro do sistema integrado fuzzy / RB.

Sistema	Eventos Inferidos	Inferências Erradas	Percentual de Inferência Total	Erro Parcial
fuzzy	3.542	192	44,3%	5,42%
Rede Bayesiana	4.458	283	55,7%	6,35%
Total:	8.000	475	100 %	
Erro do Sistema Integrado:			5,94%	

Na Tabela 6-8 verifica-se que cada um dos modelos, fuzzy e RB, estimaram aproximadamente metade dos 8.000 registros, apresentando erros de 5,42% e 6,35%, respectivamente. Considerando-se os dois modelos como um sistema único, foi obtido o erro de 5,94%. Verifica-se que este valor é próximo ao erro apresentado somente pela RB na Tabela 6-2, para eventos com saídas únicas. Entretanto, o sistema integrado possui a vantagem de agregar uma base de regras do modelo fuzzy que pode influenciar o resultado da identificação de causas, tendo-se disponível uma base de conhecimento especialista que expresse da melhor maneira as características de falha.

6.1.7 Inferência de dados incoerentes

Considerando que foram utilizados dados típicos que traduzem o comportamento do sistema elétrico, também foi verificado de que maneira os modelos fuzzy e RB se comportariam com eventos incoerentes. Exemplos de eventos duvidosos podem ser a causa Descarga Atmosférica com tempo bom e Sobrecarga com clima de temporal. Para criar esses registros falhos, foi gerada uma lista de combinações com diferentes estados das variáveis de entrada e saída que não constavam no conjunto de 15.994 eventos obtidos anteriormente. Perfazendo 730 eventos, parte da lista é mostrada na Figura 6-10 contendo os percentuais de causas conforme a Figura 6-11.

Causa	Clima	Fatos	Vegetaçã	Objeto	Flem. Int.	Vento	Turno	Estação	Flem.Esp.
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Transformador
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Chave Fusivel
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Modulo
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Medidor Primario
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Chave
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Religador
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Servico
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Seccionalizador
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Trecho Primario
Descarga Atmosférica	Bom	*	Não-Podad	Não	Equipment	Excessivo	2oTurno	outono	Regulador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Transformador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Chave Fusivel
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Modulo
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Medidor Primario
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Chave
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Religador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Servico
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Seccionalizador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Trecho Primario
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	1oTurno	verao	Regulador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Transformador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Chave Fusivel
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Modulo
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Medidor Primario
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Chave
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Religador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Servico
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Seccionalizador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Trecho Primario
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	inverno	Regulador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	verao	Transformador
Descarga Atmosférica	Bom	*	Podado	Não	Equipment	Excessivo	2oTurno	verao	Chave Fusivel

Figura 6-10 – Lista mostrando eventos de Descarga Atmosférica com tempo bom.

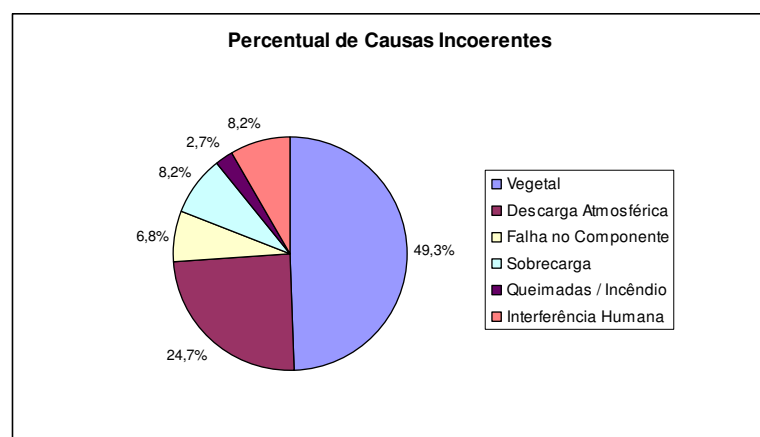


Figura 6-11 – Percentual dos 730 eventos duvidosos.

Utilizando o mesmo conjunto de dados de treino definido em 5.5.4 para a estimação dos parâmetros da RB e estabelecimento da base de regras fuzzy, a inferência das causas dos 730 registros falhos permitiu observar os resultados da Tabela 6-9.

Tabela 6-9 – Erro do sistema integrado utilizando dados incoerentes.

Sistema	Eventos Inferidos	Inferências Erradas	Percentual de Inferência	Erro Parcial
fuzzy	116	116	15,89%	100%
Rede Bayesiana	614	500	84,11%	81,4%
Total:	730	616	100 %	
Erro do Sistema Integrado:			84,38 %	

Apesar do número de eventos incoerentes ser inferior aos 8.000 registros testados nas seções anteriores, os resultados mostram que a proporção de dados inferidos pelo sistema fuzzy não foi aproximadamente igual aos da RB, diferentemente do obtido na Tabela 6-8. Do percentual inferido pelo sistema fuzzy, todas as causas estimadas não estavam de acordo com o estipulado na base de dados, o que demonstra que a base de regras utilizada não possui um domínio de conhecimento para os dados apresentados. Concomitantemente, a estimação das causas com a RB para os demais eventos incoerentes apresentou o erro percentual elevado de 81,4%. Nas inferências dos dois sistemas não foi detectada alguma anomalia nas saídas de

cada modelo, o que possibilitaria indicar que um registro isolado apresenta combinações de variáveis sem sentido.

Com o erro total de 84,38%, conforme a Tabela 6-9, observa-se que o sistema é capaz de detectar a incoerência de um conjunto de dados a partir do elevado erro, pois a base de registros não carrega informações condizentes com o sistema elétrico. Entretanto, com este sistema inicialmente configurado, a ausência de uma característica anormal nas saídas do modelo não permitiriam diagnosticar incoerências de um evento isoladamente no ato de reparo de energia por eletricitistas. Este desempenho é particularmente interessante para aplicar antes do armazenamento dos dados de uma concessionária, tendo disponível um conjunto de registros para comparar o resultado de diversos eventos entre si.

7 COMENTÁRIOS FINAIS

O desenvolvimento de um sistema que realiza diagnóstico de causas de desligamentos não programados foi abordado neste trabalho, tendo como recurso uma base de dados de históricos de interrupções. Inicialmente contendo diversas incertezas, a base de registros foi tratada, possibilitando realizar algumas análises de modo a fornecer informações sobre o sistema elétrico e descartar dados contraditórios. A partir disso, o banco de dados apresentou variáveis de entrada e saída assumidas como coerentes com a realidade. O método proposto foi a utilização desse conjunto de dados tratados para o treinamento de uma Rede Bayesiana, possibilitando a inferência com base na tendência estatística dos registros, e a aplicação das informações em um modelo de inferência fuzzy, simulando uma fonte de conhecimento especialista. Percebe-se que desse modo o sistema de inferências funciona como o comportamento humano, por meio de decisões baseadas no conhecimento heurístico e suposições através da dedução da repetição de eventos conforme experiências anteriores.

Um ponto chave da proposta deste trabalho surgiu na tentativa de explicar por que o treino e inferência do mesmo conjunto de dados na Rede Bayesiana não resultava em erro zero. Seria compreensível a existência de erro se o algoritmo treinasse um grupo de dados e inferisse com outro ligeiramente diferente. Entretanto, considerando que as inferências ocorrem matematicamente de forma probabilística e não determinística, o algoritmo pode ter alguma limitação quanto ao tamanho e natureza da amostra treinada, já que utiliza a otimização de uma função objetivo estatística para obter os parâmetros do modelo. A utilização de outro algoritmo de inferência (*Junction Tree* ao invés de Eliminação de Variáveis), permitiu verificar que o erro permaneceu o mesmo, tendo somente uma variação significativa do tempo de execução. Não dispondo-se de um algoritmo substituto para comparar a estimação de parâmetros da RB (treino), uma forma de analisar o problema foi estudando a natureza e quantidade dos dados, em que foi realizada a tentativa de identificar quais fatores acarretam em erro ao treinar a RB.

Com o objetivo específico de encontrar formas de atenuação do erro de inferência para a Rede Bayesiana, foi proposta a adoção de outras 3 variáveis de entrada para diferenciar mais as causas do banco de dados. Através de um teste, foi verificado que a inclusão de três variáveis em um conjunto de dados confirmou uma redução de erro em 4,65% para 172

eventos envolvidos. Nota-se que a diferenciação faz especificar com mais detalhes de que maneira a falha elétrica ocorreu. Outra medida foi a criação de um método de análise do banco de dados por meio de planilhas específicas, das quais permitem a visualização das combinações possíveis de variáveis envolvidas nos registros (lista da Figura 5-10). A análise permitiu constatar um elevado número de causas múltiplas para as mesmas variáveis de entrada. Estes eventos não significam necessariamente uma falha do processo de registros da concessionária. Pode-se interpretar apenas que as variáveis de entrada estabelecidas não foram suficientes para detalhar a maneira como o sistema elétrico foi afetado. Do ponto de vista da inferência de causas, o descarte desses eventos permitiu a qualificação do banco de dados, contribuindo para a redução do erro da RB de 15,94% para 4,60%, contendo ambos 15.994 eventos. Os resultados provam que a ambiguidade dos eventos é de extrema importância, devendo ser levada em consideração no problema de diagnósticos e determinação de seus erros.

Através do método de análise desenvolvido, também foi realizada a extração de informações que traduzem as características do sistema elétrico. A análise contribuiu para verificar que, mesmo após os diversos tratamentos dos dados, 26.282 registros mostraram a permanência da informação de que interrupções por sobrecarga ocorrem em sua maioria nos horários de pico (período vespertino), com 82,3%, e dias ensolarados (clima bom), com 69,5%. Outras características mais evidentes entre variáveis não foram observadas neste trabalho. Entretanto, da mesma forma que a informação encontrada no banco de dados para sobrecarga estava de acordo com Brown (2002), isto ao menos sugere que existe a possibilidade de as informações a respeito de outras causas estarem coerentes. Essa amostra de análise permite conceber que é possível realizar um estudo mais aprofundado, extendendo a técnica de agrupamento por *clusters* com demais variáveis, com o intuito de encontrar outras relações com entradas e saídas do modelo. Sendo assim, é factível realizar outras deduções de características do sistema que possibilitem justificar investimentos no sistema elétrico de forma preventiva.

A consideração da relevância de informações contidas no banco de dados levou à construção de uma base de 20 regras para reger a relação de variáveis do filtro fuzzy. Observa-se ainda que, utilizando poucas variáveis do modelo fuzzy foi possível definir heurísticamente, isto é, de forma intuitiva, o modelo do comportamento do sistema elétrico. A

legitimidade da informação também proporcionou a crença de que os dados eram coerentes para o treinamento da Rede Bayesiana.

Outro conceito fundamental realizado foi a abordagem de seleção das saídas dos modelos utilizando critérios de grau de confiança alto para a escolha da resposta final inferida. Conforme verifica-se na seção 6.1.6, os resultados apresentam o princípio de que, se determinada saída do sistema fuzzy é a maior, é necessário ainda que esse valor seja maior ou igual a um patamar de confiança aceitável para a inferência. Não sendo aceitável, resta apenas a análise através da tendência histórica de comportamento de dados, orientada pela Rede Bayesiana. O bloco responsável por esse critério de decisão é um filtro que também verifica se determinadas causas da RB têm um nível de confiança previamente analisado como alto, através dos dados disponíveis de treino. Agregando os erros do sistema com filtro fuzzy aos da RB, o erro da rede causal mostrou-se capaz de atenuar com o suporte do modelo baseado em conhecimento especialista, passando de 6,35% para 5,94% para os 8.000 eventos envolvidos.

Vale ressaltar que, devido ao uso da base de conhecimento, em relação à RB, o sistema fuzzy torna possível avaliar mais detalhadamente a natureza das variáveis de entrada cuja saída está sendo inferida. Além disso, dispõe de diferentes variedades de recursos para a modelagem, como: métodos de defuzzificação, tipos de funções de pertinência, tipos de operadores lógicos, entre outros. Exemplo disso é a análise dos resultados de inferências, mostrando que as causas *Falha no Componente e Interferência do Meio* tiveram erros maiores em relação as outras causas (conforme Tabela 6-5 e Tabela 6-6). Esta discrepância é possível de ser resolvida ao ser encontrada uma base de regras que melhor define o modelo. Já para a RB isto se tornaria um inconveniente com difícil solução. Um método prático de criar um mapa de regras é proposto na seção 5.6.4, onde foi realizado o mapeamento dos extremos do modelo. Em seguida, utilizando poucas regras, foi completado o conhecimento a respeito dos desligamentos. Utilizando funções gaussianas e triangulares, também foi comprovado que o formato dessas funções, bem como o método de defuzzificação, não são fatores decisivos para a modelagem deste problema. O que realmente indica determinar o sucesso de acertos são o conjunto de regras e os valores críticos definidos.

Outra abordagem foi a verificação do comportamento dos sistemas ao inferir eventos que não eram coerentes, como interrupções de descarga atmosférica vinculados ao tempo bom. Como os dados obtidos para a pesquisa traziam registros com essas características, um

sistema diagnosticador de faltas deveria detectar tais eventos. O modelo RB realizou a inferência de 84,11% dos eventos, apresentando erro de 81,4%. O modelo fuzzy inferiu 15,89% dos eventos, com erro de 100%, conforme a seção 6.1.7. Essa característica é vantajosa por permitir a uma pessoa sem domínio de conhecimento do sistema elétrico verifique a incoerência de um conjunto de dados, antes do armazená-lo na concessionária. Contudo, o sistema integrado é vulnerável para ser utilizado por eletricitas sem experiências para diagnosticar faltas, uma vez que ele não denunciaria um evento único como incoerente, sem comparar com um conjunto de dados. Esses problema pode ser evitados por meio de um levantamento de possíveis registros que apresentariam diversas incongruências. Após, os modelos deveriam estar treinados para esses eventos de modo a apontá-los como dados incoerentes.

De maneira geral, a grande dificuldade deste trabalho foi a tentativa de extrair e processar informações utilizando um banco de dados bastante aquém do desejado. O problema reside no fato de que apenas o diagnóstico de causas contém em si uma série de incertezas que são inerentes à detecção do “epicentro” da interrupção. Se o banco de dados contém, do mesmo modo, incertezas, é preciso aplicar diversas soluções unicamente para minimizar esta componente. Observando-se por um aspecto mais vantajoso, a qualidade precária dos registros forçou o desenvolvimento de pesquisas concentrando esforços no pior cenário possível, a fim de extrair a melhor solução dentro do que é apresentado. Portanto, a proposta de atenuação do erro foi alcançada da forma apresentada em 6.1.6, através de um paradigma de sistema de inferência que, se houver disponibilidade de dados com maior confiabilidade, o diagnóstico de causas de desligamento terá grande alcance de sucesso.

Cabe ressaltar que as empresas, de modo geral, enfrentam dificuldades para implementarem um sistema de registros de dados com qualidade adequada. Não obstante o fato de que novos sistemas promissores, como os *Smart Grids* venham a trazer muitos benefícios em relação a melhora da aquisição dos dados, a busca por melhorias contínuas com os processos atuais é uma política de gestão que sempre tende a trazer bons resultados diretos ou indiretos. Tais medidas trazem retornos econômicos e sociais inestimáveis para o desenvolvimento da sociedade. Do mesmo modo que neste, os trabalhos de Pretto (2005) e Tronchoni (2008) utilizam recursos tecnológicos e artifícios de diversos domínios para solucionar um problema que é de fundamental importância para agilizar o reparo por parte da

equipe de eletricitas e estabelecer indicadores corretos que justifiquem o investimento em melhorias dos pontos falhos do sistema elétrico.

7.1 Trabalhos Futuros

Paralelamente às contribuições deste trabalho, cabem análises para expandir esta e outras áreas de estudo, dando continuidade de pesquisa aos problemas apresentados. São sugeridas oportunidades de melhoria nas seguintes propostas:

- para eventos de saídas múltiplas com mesma combinação de entradas, realização de diagnósticos dos desligamentos com apresentação de probabilidades de ocorrências das outras causas;
- detalhar um estudo identificando as principais variáveis que têm relação com cada causa, estimando o grau dessas influências e com a tentativa de incluir estas variáveis para evitar registros de múltiplas causas, aparentando mesmos cenários;
- utilização de outras classes de dados para compor as variáveis, de modo a constatar mais informações do comportamento do sistema elétrico e detalhar os modos de falhas da distribuição de energia;
- agregar na base de dados a variável temperatura para verificar sua influência em causas que provavelmente correlacionam-se com esta variação, como sobrecarga e falha em transformadores;
- inclusão do registro de nomes do eletricitista responsável no momento da avaliação do reparo, a fim de realizar um estudo da variabilidade de atribuição de causas envolvendo diferentes pessoas. Esta medida possivelmente também aumentaria a confiabilidade dos registros, por exercer maior controle neste processo;
- estudo com formulações baseadas na estatística justificando o motivo de ocorrência de erro após realizar treino e inferência com o mesmo conjunto de dados;
- teste de outras topologias da Rede Bayesiana a fim de verificar o quanto isto influencia no problema do erro dos conjuntos de dados;
- aplicação de outros algoritmos de estimação dos parâmetros;
- desenvolvimento de outros sistemas inteligentes com métodos de detecção de eventos inverossímeis para um sistema elétrico.

REFERÊNCIAS

- ANEEL (2010). Agência Nacional de Energia Elétrica. Brasília, DF. Disponível em: <http://www.aneel.gov.br/aplicacoes/noticias/Output_Noticias.cfm?Identidade=3253&id_area=>>. Acesso em: 15 fev. 2010.
- BORTONI, E.C.; HASSIN, E.S.; MASSELI, S. (2004). A Regulamentação da continuidade dos serviços de distribuição de energia elétrica no Brasil. In: IV Congresso Brasileiro de Planejamento Energético, Itajubá, 13p.
- BROWN, R.E. (2002). Electric power Distribution reliability. New York, NY. Marcel Dekker, 355p.
- CHIEN, C-F.; CHEN, S-L.; LIN, Y-S. (2002). Using Bayesian Network for fault location on distribution feeder. IEEE Transactions on Power Delivery, jul, v.17, n.13, p.785-793.
- CHOW, M-Y.; YEE, S.O.; TAYLOR, L.S. (1993). Recognizing animal-caused faults in power distribution systems using artificial neural networks. IEEE Transactions on Power Delivery, v.8, n.3, p.1268-1274.
- CHEN, G.; PHAM, T. T. (2001). Introduction to fuzzy Sets, fuzzy Logic and fuzzy Control Systems. CRC Press LLC, Boca Raton, FL, 316p.
- CEEE (2009). Companhia Estadual de Energia Elétrica. Porto Alegre, RS. Disponível em: <<http://www.ceee.com.br/pportal/ceee/component/Controller.aspx?CC=25960>>. Acesso em: 26 nov. 2009.
- DARWICHE, A. (2009). Modeling and Reasoning with Bayesian Networks. Cambridge University Press, New York, p.216-224.
- DOBLE, 2010. Doble Engineerign Company. Watertown, MA, USA. Disponível em: <http://www.doble.com/content/products/ACCA_3.pdf>. Acesso em: 20 jan. 2010.
- DOLA, H.M.; CHOWDHURY, B.H. (2005). Data Mining for Distribution System Fault Classification. Proceeding of the 37th Annual North American Power Symposium, Missouri, 23-25 Oct, p.457- 462.

HARRIS, J. (2006). Fuzzy Logic Applications in Engineering Science – Microprocessor-based and Intelligent Systems Engineering. Vol. 29, Springer. Netherlands, 221p.

HAZZAN, S. (1977). Fundamentos de Matemática Elementar – Combinatória/Probabilidade. 3ª edição, Atual, São Paulo, SP, 149p.

MATHWORKS (1995). Fuzzy Logic Toolbox – User`s Guide, Version 2. Matlab.

JENSEN, F.V. (2001). Bayesian Networks and Decision Graphs. New York: Springer - Verlag. 268 p.

JOHNS, A.T. ; LAI, L.L. ; EL-HAMI, M. ; DARUVALA, D.J. (1991). New approach to directional fault location for overhead power distribution feeders. IEEE Proceedings, July, v.138, n.4, p.351-357.

LIU, Y.; SCHULTZ, N.N. (2002). Integrated fuzzy filter for distribution outage information. Electric Power Systems Research, Elsevier, n. 63, p.177-184.

NEILL, F. M.; THRO, E. (1994). Fuzzy Logic – A practical Approach, Academic Press, United Kingdom Edition, London, 279p.

MURPHY, K. (1998). A Brief Introduction to Graphical Models and Bayesian Networks. University of British Columbia, Vancouver, CA. Disponível em:
<http://people.cs.ubc.ca/~murphyk/Bayes/bayes_tutorial.pdf>. Acesso em: 20 jul. 2009.

NASSAR, S.M.; COELHO, J.; WRONSCKI, V.R.; QUEIROZ, H.; GAUCHE, E. (2004). Identificação de condições climáticas adversas através de redes bayesianas. IX SEPOPE - Symposium of Specialists in Electrical Operational and Expansion Planning, Rio de Janeiro. Anais do IX SEPOPE, SP-132, v. 1, p.1-5.

OPRISAN, M.; FILIPPELLI, F.; CLARK, I.M.; BILLINTON, R. (1991). A Reliability data system for the reporting of forced outages of distribution equipment. In: WESCANEX '91 'IEEE Western Canada Conference on Computer, Power and Communications Systems in a Rural Environment', 29-30 May, p.267-270.

OSHA (2010). Work Hazards and Safety Practices in the Electric Power Industry, United States Department of Labor, Washington, D.C. Disponível em:
<http://www.osha.gov/SLTC/etools/electric_power/illustrated_glossary/substation_equipment/lightning_arresters.html>. Acesso em: 20 jan. 2010.

PASKIN, M.A. (2003). A Short Course on Graphical Models: The Junction Tree Algorithms. Stanford AI Laboratory, Stanford, CA. Disponível em: <<http://ai.stanford.edu/~paskin/gm-short-course/lec3.pdf>>. Acesso em: 15 ago. 2009.

PENG, J.T.; CHIEN, C.F. ; TSENG, T.L.B. (2004). Rough set theory for data mining for fault diagnosis on distribution feeder. Generation, Transmission and Distribution IEEE Proceedings, Nov, v.151, n.6, p.689-697.

PRETTO, C.O.; RANCICH, G.V. ; ROSA, M.A.; LEMOS, F.A.B. (2003). Forced outages information treatment system and cause identification based on mobile computing and neural networks. IEEE PowerTech Conference Proceedings, June, Bologna, v.1, 6p.

PRETTO, C.O. (2005). Sistema de Coleta e Tratamento de Informações sobre Desligamentos Não Programados Baseados em Computadores Móveis. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 87p.

ROSS, T. J. (2004). Fuzzy Logic with Engineering Applications. University of New Mexico, USA, Second Edition, pgs.15; 179.

SAVE B. (2010). Sociedade para conservação de aves do Brasil. São Paulo, SP. Disponível em: <<http://www.savebrasil.org.br/>>. Acesso em: 15 jan. 2010.

SRIDHARAN, K.; SCHULZ, N.N. (2001). Outage management through AMR systems using an intelligent data filter. IEEE Transactions on Power Delivery, v.16, n.4, p.669-675.

SU, C.-L.; TENG, J.-H. (2007) Outage costs quantification for benefit-cost analysis of distribution automation systems. Electrical Power & Energy Systems, Elsevier, June, n.29, p.767-774.

TECH, A.R.B; GUALAZZI, G.A.S. (2006). O processo decisório apoiado em técnicas e modelos de inteligência artificial. XIII Simposio de Engenharia de Produção - SIMPEP, Nov, Bauru, SP, 7 p.

TRONCHONI, A.B.; PRETTO, C.O.; ROSA, M.A.; LEMOS, F.A.B. (2010). Descoberta de Conhecimento em Base de Dados de Eventos de Desligamentos de Empresas de Distribuição. Revista SBA – Controle & Automação, Sociedade Brasileira de Automática. Vol 1.

TRONCHONI, A.B.; PRETTO, C.O.; LICKS, V.; ROSA, M.A.; LEMOS, F.A.B. (2007). Forced outage cause identification based on bayesian networks. Proceedings of 2007 IEEE Lausanne Powertech, Switzerland, v.1., p.1-6.

TRONCHONI, A.B. (2008). Identificação de causas de desligamentos não programados em redes de distribuição. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 131p.

TSO, S.K.; LIN, J.K.; HO, H.K.; MAK, C.M.; YUNG, K.M.; HO, Y.K. (2004). Data mining for detection of sensitive buses and influential buses in a power system subjected to disturbances. *IEEE Transactions on Power Systems*, Nov, v.19, n.1, p.563-568.

WU, Z.; ZHANG, B.; ZHU, Y.; ZHAO, W.; ZHOU, Y. (2009). Transformer fault portfolio diagnosis based on the combination of the multiple Bayesian classifier and SVM. *International Conference on Electronic Computer Technology, Proceedings...*p.370-382.

XU, L.; CHOW, M.-Y. (2006). A classification approach for power distribution systems fault cause identification. *IEEE Transactions on Power Systems*, v. 21, n.1, p.53-60.

XU, L.; CHOW, M.-Y.; TAYLOR, L.S. (2007). Power Distribution Fault Cause Identification with Imbalanced Data Using the Data Mining-Based fuzzy Classification E-Algorithm. *IEEE Transactions on Power Delivery*. v.22, n.1, p.164-171.

YONGLI, Z.; LIMIN, H.; JINLING, L. (2006). Bayesian network-based approach for power systems fault diagnosis. *IEEE Transactions on Power Delivery*, v.21, n.2, p.634-639.

ZADEH, L. A. (1994). *Fuzzy Logic, Neural Networks, and Soft Computing*. IEEE Computer Society Press, v.11, p.48-56.

ZADEH, L.A. (1994). Fuzzy logic, neural network and soft computing. *Communications of the ACM*, v.37, n.3.

ZEROHORA (2009). *Jornal Zero Hora*, Porto Alegre, RS. Disponível em: <<http://zerohora.clicrbs.com.br/zerohora/jsp/default.jsp?uf=1&local=1§ion=Geral&newsID=a2729707.xml>>. Acesso em: 26 nov. 2009.

ANEXOS

ANEXO A - Propriedades de conjuntos fuzzy (MATHWORKS, 1995).

Genericamente, um operador lógico Y pode ser representado por uma função F , que opera dois valores a e b :

$$Y = F(a, b)$$

Se F representar operadores de interseção ou união, deve respeitar certas propriedades, conforme a tabela seguinte:

Tabela 1 – Propriedades de operadores de interseção e união.

	Interseção	União
Fronteira	$F(0,0) = 0, F(a,1) = F(1,a) = a$	$F(1,1) = 1, F(a,0) = F(0,a) = a$
Monotonicidade	$F(a,b) \leq F(c,d)$ se $a \leq c$ e $b \leq d$	
Comutatividade	$F(a,b) = F(b,a)$	
Associatividade	$F(a, F(b,c)) = F(F(a,b), c)$	

A primeira propriedade impõe a correta generalização para conjuntos booleanos. A segunda, implica que um decréscimo em uma das funções de pertinência não pode resultar em um acréscimo na interseção ou união. A terceira, indica que a ordem de operação dos conjuntos fuzzy não altera o resultado. A última propriedade indica que é possível operar com diferentes combinações de pares de conjunto fuzzy, em qualquer ordem.