

**UNIVERSIDADE SANTA CECÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
MESTRADO EM ENGENHARIA MECÂNICA**

LUIZ ANTONIO FERRARO MATHIAS

**MINERAÇÃO DE DADOS EM SISTEMAS DE ENERGIA ELÉTRICA
UTILIZANDO ALGORITMOS FUNDAMENTADOS EM
LÓGICA PARACONSISTENTE ANOTADA - LPA**

SANTOS/SP

2015

LUIZ ANTONIO FERRARO MATHIAS

**MINERAÇÃO DE DADOS EM SISTEMAS DE ENERGIA ELÉTRICA
UTILIZANDO ALGORITMOS FUNDAMENTADOS EM
LÓGICA PARACONSISTENTE ANOTADA - LPA**

Dissertação apresentada à Universidade Santa Cecília como parte dos requisitos para obtenção de título de Mestre no Programa de Pós-graduação em Engenharia Mecânica, sob orientação do Prof. Dr. João Inácio da Silva Filho e co-orientador Prof. Dr. Mauricio da Conceição Mario.

SANTOS/SP

2015

Autorizo a reprodução parcial ou total deste trabalho, por qualquer que seja o processo, exclusivamente para fins acadêmicos e científicos.

Mathias, Luiz Antonio Ferraro.

Mineração de Dados em Sistemas de Energia Elétrica
utilizando algoritmos fundamentados em Lógica Paraconsistente
Anotada - LPA / Luiz Antonio Ferraro Mathias.

-- 2015.

70p.

Orientador: Prof. Dr. João Inácio da Silva Filho.

Coorientador: Prof. Dr. Maurício da Conceição Mario.

Dissertação (Mestrado) -- Universidade Santa Cecília,
Programa de Pós-Graduação em Engenharia Mecânica, Santos,
SP, 2015.

1. Mineração de dados. 2. Energia elétrica. 3. Lógica
Paraconsistente Anotada. I. Da Silva Filho, João Inácio, II.
Mario, Maurício da Conceição. III. Mineração de dados em
Sistemas de energia elétrica utilizando algoritmos
Fundamentados em Lógica Paraconsistente Anotada - LPA.

Elaborada pelo SIBi – Sistema Integrado de Bibliotecas - Unisanta

Dedicatória

Dedico este trabalho a minha esposa e ao meu filho, que me apoiaram e tiveram comigo a paciência necessária nos momentos em que estive ausente, por conta deste.

AGRADECIMENTOS

Ao professor Dr. João Inácio, meu agradecimento especial por sua precisa orientação, pelos ensinamentos na Lógica Paraconsistente e por sua fundamental contribuição para o desenvolvimento deste trabalho de pesquisa.

Ao professor Dr. Maurício Conceição, por suas considerações e recomendações de adequação no presente projeto.

Ao professor Dr. Alexandre Rocco, por suas importantes contribuições ao projeto.

Ao professor Dr. Ederson Rafael Wagner, como membro externo convidado pela UNISANTA para compor a banca de avaliação da apresentação desta dissertação, agradeço pelas preciosas palavras e pela avaliação.

A todo o corpo docente do Programa de Pós-Graduação em Engenharia Mecânica.

À Equipe da secretaria do Programa, pelo auxílio, apoio e suporte em todas as fases do processo de estudo.

RESUMO

Esta dissertação explora a técnica de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases – KDD*) e mais especificamente, o processo de mineração de dados. Tais técnicas foram aplicadas para normalização de sinais de informação armazenados em uma base de dados de uma subestação de distribuição de energia elétrica. Os dados foram coletados a partir de um sistema de supervisão e aquisição de dados (SCADA- *supervisory control and data acquisition*). Tal processo teve por finalidade a preparação destes dados para identificação de padrões e correlações existentes nos mesmos. Para o processo de mineração de dados foi utilizada a Lógica Paraconsistente Anotada – LPA, uma lógica não-clássica que possui a propriedade de suportar sinais contraditórios identificados na base de dados alvo da análise, sem que tal conflito inviabilize as conclusões necessárias. No presente trabalho, uma base de dados foi construída a partir dos dados coletados pelo sistema supervísório da subestação, além da limpeza e tratamento dos dados e realização de inferências com a lógica supracitada, permitindo assim, a descoberta de conhecimento incerto e pouco visível aos operadores da subestação. Como produto da pesquisa foi desenvolvido uma ferramenta computacional estruturada nos algoritmos paraconsistentes que permite consultas e oferece respostas baseadas nos conceitos fundamentais da LPA. Os resultados obtidos nesta primeira versão poderão ser utilizados como referências para as elaborações de estruturas complexas de processamento de dados baseados em técnicas de *Data Mining* que funcionem de forma híbrida com técnicas de tratamento de incertezas fundamentadas em Lógica Paraconsistente.

Palavras-Chave: Mineração de dados. Energia elétrica. Lógica Paraconsistente. Lógica Paraconsistente Anotada.

ABSTRACT

This dissertation explores the Knowledge Discovery in Databases technique – KDD, and more specifically, the data mining process. Such techniques were applied to normalize the information signs stored in a database of an electric energy distribution sub-station. The data were collected from a supervisory and data acquisition system (SCADA – supervisory control and data acquisition). Such process aimed the preparation of these data for identification of existent patterns and co-relations on them. For the data mining process the Paraconsistent Annotated Logic – PAL was adopted, a non-classical logic, that has the property of supporting contradictory signals on the analysis target database, in a way that such conflict won't make the necessary conclusions impracticable. In the present work, a database was built from the data collected by the substation supervisory system, besides the data cleaning and treatment and the execution of inferences with the logic mentioned above, thus enabling the discovery of uncertain and little visible knowledge for the substation team operators. As a product resulting from the research was developed a computational tool that allows paraconsistentes algorithms-structured queries and provides answers based on the fundamental concepts of the PAL. The results obtained in this first version can be used as references for the elaborations of complex structures of data processing based on Data Mining techniques that work in a hybrid with techniques of treatment of uncertainties based on Paraconsistent Logic.

Key-words: Data mining. Electric power. Paraconsistent Logic. Paraconsistent Annotated Logic.

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas do processo de KDD	18
Figura 2 - Etapas do processo de pré-processamento do KDD.	19
Figura 3 - Reticulado Representativo De Hasse.	29
Figura 4 - Representação da análise Paraconsistente LPA2v.	30
Figura 5 - Graus de evidência no Quadrado Unitário do Plano Cartesiano.....	31
Figura 6 - Representação do Reticulado LPA2v com valores extremos	32
Figura 7 - Eixos dos Graus de Contradição e de Certeza.	34
Figura 8 - Quadrado unitário no Plano Cartesiano.	35
Figura 9 - Projeção do Valor da Distância D sobre o eixo dos graus de certeza para obtenção do Grau de Certeza Real G_{CR} no reticulado da LPA2V.	37
Figura 10 - Símbolo do NAP – Nó de Análise Paraconsistente.....	38
Figura 11 - Esquemático de um tipo da Subestação em estudo.	42
Figura 12 - Base de dados original em ACCESS da Subestação em estudo coletados pelo sistema SCADA.....	44
Figura 13 - Base de dados elaborada na ferramenta Microsoft SQL Server.....	45
Figura 14 - Grandeza variando de modo diretamente proporcional relacionada com o Grau de Evidência Normalizado.	46
Figura 15 - Grandeza variando de modo inversamente proporcional relacionada com o Grau de Evidência Normalizado.	47
Figura 16 - Etapas do processo de mineração de dados da Subestação de Energia Elétrica com o emprego da Lógica Paraconsistente Anotada.	49
Figura 17 - Gráfico de normalização da corrente elétrica na fase AV da Barra 1 de disjuntores da alimentação das cargas.	50
Figura 18 - Gráfico de normalização da corrente elétrica na fase AV da Barra 2 de disjuntores da alimentação das cargas.	51
Figura 19 - Gráfico de normalização da corrente elétrica na fase AV da Barra 3 de disjuntores da alimentação das cargas.	51
Figura 20 - Gráfico de normalização da corrente elétrica na fase AV da Barra 4 de disjuntores da alimentação das cargas.	52
Figura 21 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT1.....	53

Figura 22 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT2.....	53
Figura 23 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT3.....	54
Figura 24 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT4.....	54
Figura 25 - Ferramenta desenvolvida com o objetivo de mensurar o processamento dos dados obtidos por medições nos disjuntores com a aplicação da Lógica Paraconsistente.....	55

LISTA DE TABELAS

Tabela 1: Dispositivos da subestação cuja fase AV foi alvo de estudo.	44
Tabela 2: Resultado de inferências realizadas durante o processamento dos dados de corrente na fase AV dos disjuntores da subestação.	57

LISTA DE SIGLAS

LP	Lógica Paraconsistente
LPA	Lógica Paraconsistente Anotada
LPA2v	Lógica Paraconsistente Anotada com anotação de dois valores
QUPC	Quadrado Unitário do Plano Cartesiano;
<i>KDD</i>	<i>Knowledge Discovery in Databases</i>
SCADA	<i>Supervisory Control And Data Acquisition</i>
ANEEL	Agência Nacional de Energia Elétrica

LISTA DE SÍMBOLOS

μ_1	Grau de evidência favorável
μ_2	Grau de evidência desfavorável
T	Inconsistente
$\perp$	Paracompleto ou indeterminado
$\perp \rightarrow f$	Indeterminado, tendendo ao Falso;
$\perp \rightarrow v$	Indeterminado, tendendo ao Verdadeiro;
$T \rightarrow f$	Inconsistente, tendendo ao Falso;
$Qv \rightarrow T$	Quase Verdadeiro, tendendo ao Inconsistente;
$Qf \rightarrow T$	Quase Falso, tendendo ao Inconsistente;
$Qf \rightarrow \perp$	Quase Falso, tendendo ao Indeterminado;
$Qv \rightarrow \perp$	Quase Verdadeiro, tendendo ao Indeterminado.
G_{it}	Grau de Inconsistência;
G_{id}	Grau de Indeterminação;
G_v	Grau de Verdade;
G_f	Grau de Falsidade;
G_{ct}	Grau de Contradição;
G_c	Grau de Certeza;
V_{scc}	Variável superior de controle de certeza (C_1);
V_{icc}	Variável inferior de controle de certeza (C_2);
V_{scct}	Variável superior de controle de contradição (C_3);
V_{icct}	Variável inferior de controle de contradição (C_4).

SUMÁRIO

1. INTRODUÇÃO	12
1.1. JUSTIFICATIVA.....	13
1.2. OBJETIVO DO TRABALHO	14
1.3. ORGANIZAÇÃO DA DISSERTAÇÃO.....	14
1.4. O SISTEMA ELÉTRICO BRASILEIRO E AS SUBESTAÇÕES DE ENERGIA ELÉTRICA	15
1.5. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS.....	16
1.6. ETAPAS DA DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS.....	18
1.7. MINERAÇÃO DE DADOS	20
1.8. PRINCIPAIS TAREFAS DE MINERAÇÃO DE DADOS.....	22
1.8.1. CLASSIFICAÇÃO.....	22
1.8.2. CLUSTERIZAÇÃO	23
1.8.3. REGRESSÃO.....	23
1.8.4. REGRAS DE ASSOCIAÇÃO.....	23
1.8.5. AGRUPAMENTO	24
1.8.6. ESTIMATIVA	24
1.8.7. DESVIO.....	24
1.9. FUNDAMENTOS DA LÓGICA PARACONSISTENTE ANOTADA E ALGORITMOS RELACIONADOS.....	25
1.10. AS LÓGICA NÃO-CLÁSSICAS	26
1.11. A LÓGICA PARACONSISTENTE	27
1.12. A LÓGICA PARACONSISTENTE ANOTADA	28
1.13. A LÓGICA PARACONSISTENTE ANOTADA COM ANOTAÇÃO DE 2 VALORES (LPA2v).....	30
1.14. REPRESENTAÇÃO NO QUADRADO UNITÁRIO NO PLANO CARTESIANO – QUPC	30
1.14.1. O RETICULADO DA LPA COM REGIÕES DELIMITADAS.....	31
1.14.2. O GRAU DE CERTEZA	32
1.14.3. O GRAU DE CONTRADIÇÃO	35
1.14.4. O GRAU DE CERTEZA REAL.....	36

1.14.5.	ALGORITMOS UTILIZADOS NAS ANÁLISES PARACONSISTENTES	37
1.14.6.	NÓ DE ANÁLISE PARACONSISTENTE	37
1.14.7.	ALGORITMO EXTRATOR DE EFEITOS DA CONTRADIÇÃO	39
2.	METODOLOGIA.....	42
3.	RESULTADOS E DISCUSSÕES	50
3.1.	DISCUSSÃO	56
4.	CONCLUSÕES	60
	REFERÊNCIAS BIBLIOGRÁFICAS.....	62
	APÊNDICE A: Implementação do Algoritmo Extrator de Efeitos da Contradição ...	66

1. INTRODUÇÃO

Segundo Martins (2003), uma interrupção no fornecimento de energia aos consumidores é considerada uma condição anormal, originando grandes perturbações, todavia, as mesmas são inevitáveis em decorrência de defeitos de um sistema de potência, sejam por causa de situações internas ou externas, fenômenos elétricos ou ambientais ou ainda humanos.

Após detecção de alguma anomalia, a equipe técnica pode acionar um equipamento de manobra para isolar o componente defeituoso, minimizando assim, o impacto a todo o sistema elétrico da subestação.

A fim de minimizar tal situação, especificamente em subestações de distribuição de energia elétrica, estes locais contam com sistemas de supervisão, proteção e controle aplicados, objetivando assim, o atendimento aos requisitos operacionais de um sistema elétrico. Fazem parte deles, relés de proteção, controladores, redes de comunicação e *gateways* (dispositivo intermediário geralmente destinado a interligar redes, separar domínios de colisão, ou mesmo traduzir protocolos). O objetivo é facilitar as ações de operadores, de forma remota ou local, para realizarem atividades como: manobras, leituras instantâneas, registro de históricos de grandezas, leitura e registros de lista de eventos, alarmes e outros.

Estes sistemas manipulam um grande volume de dados que precisam ser processados e transformados em conhecimento aplicável.

A tecnologia da informação vem viabilizando o processamento destes grandes volumes de dados armazenados, a cada dia, de modo mais rápido e com uma maior precisão. Em tais bases de dados existe um volume de conhecimento bastante substancial, difícil de ser extraído através de métodos tradicionais ou triviais de buscas computacionais. Esta grande disponibilidade de dados armazenados nos meios computacionais permitiu o surgimento de uma nova área de pesquisa e desenvolvimento em Computação: a descoberta de conhecimento em base de dados (*Knowledge Discovery in Databases - KDD*), da qual uma etapa fundamental para esse processo é a de mineração de dados (*Data Mining*) que aplica técnicas e paradigmas de Inteligência Artificial. Esta área tem como objetivo a descoberta e descrição de tendências e padrões contidos neste processo, que

sejam potencialmente úteis e interpretáveis pelo usuário a fim de amparar processos analíticos de tomada de decisão.

Segundo Schenatz (2005), o processo de busca de conhecimento de bases de dados possui várias etapas, e antes da extração da informação é necessária uma manipulação de dados para viabilizar sua mineração. Tais etapas compreendem: seleção de dados, pré-processamento e limpeza, transformação, mineração de dados e interpretação.

O termo mineração de dados (*Data Mining*) tornou-se mais popular que o KDD e é utilizado para se referir ao processo de identificação de padrões a partir de grandes quantidades de dados armazenados em bancos de dados ou outro tipo de banco de armazenamento (HAN e KAMBER, 2000). Especificamente na etapa de mineração de dados, um conjunto de conceitos e métodos são aplicados com o objetivo de encontrar uma informação compreensível e útil para o usuário em um determinado conjunto de dados (RUSSELL, 2011).

A necessidade de armazenar e tratar dados de diversas fontes é um desafio cada vez maior também no setor elétrico que dispõe de sistemas de supervisão e aquisição de dados (*Supervisory Control And Data Acquisition – SCADA*). Estes sistemas são compostos por múltiplas camadas para coleta, processamento, análise e armazenamento de dados analógicos e digitais adquiridos nas subestações de distribuição de energia elétrica. Os dados são coletados a partir da leitura de sensores instalados nos equipamentos das respectivas plantas. Todos os dados coletados por módulos digitais de aquisição e controle e por módulos de aquisição de entradas e saídas analógicas são armazenados em bases de dados, ficando disponíveis para a mineração de dados.

1.1. JUSTIFICATIVA

Devido às suas dimensões, a questão das perdas comerciais tornou-se preocupação constante das empresas de distribuição (MACEDO *et al.*, 1993) e dos órgãos reguladores. Tal fenômeno não está restrito somente a uma determinada empresa ou área de concessão, mas em praticamente todas as distribuidoras de energia (ANEEL, 2005). A Agência Nacional de Energia Elétrica (ANEEL) procura, sistematicamente, estimar e combater as perdas comerciais de forma precisa, buscando soluções para combatê-las (LIMA, 2005).

O fornecimento de energia sem interrupções depende da integridade dos equipamentos da subestação de energia elétrica. Estes equipamentos são silenciosos em seu funcionamento e não emitem sinais avisando a presença de defeitos o que torna difícil a monitoração e controle pela equipe de Operação.

Outro fator que justifica essa pesquisa é que, no caso específico destas subestações, um elevado volume de dados é resultante (objetivo) de leituras periódicas por parte de dispositivos de sensoriamento de rede oferecendo complexidade para o tratamento da informação de forma que este seja inteligível ao operador. O correto processamento destes dados através de técnicas modernas com o processo de *Data Mining* subsidiado por LP como estamos propondo, pode apoiar ações de tomada de decisão por parte de operadores com vistas à minimização das paralisações dos equipamentos na subestação.

1.2. OBJETIVO DO TRABALHO

O objetivo deste trabalho é a aplicação dos conceitos de Busca e Conhecimento em Bases de Dados empregando a Lógica Paraconsistente Anotada para o processo de mineração de dados, visando assim, através de uma ferramenta computacional, métodos eficazes da identificação de correlações entre os dados, alvo da análise, coletados a partir de sinais de uma típica subestação de energia elétrica.

1.3. ORGANIZAÇÃO DA DISSERTAÇÃO

A apresentação desta Dissertação está organizada da seguinte forma: o primeiro capítulo traz a introdução ao assunto alvo da pesquisa deste trabalho, uma breve descrição sobre o sistema elétrico brasileiro e as características das subestações de energia elétrica, uma revisão bibliográfica a respeito do processo de descoberta de conhecimento em bases de dados e as etapas deste, uma descrição dos conceitos a respeito da sistemática de mineração de dados e as principais tarefas o norteiam, os princípios e conceitos da lógica, das diferenças entre as lógicas clássicas e não-clássicas, da Lógica Paraconsistente Aplicada e a representação do Quadrado Unitário do Plano Cartesiano representativo da respectiva lógica. No capítulo 2 são listados os métodos aplicados neste trabalho de

pesquisa com destaque para o detalhamento das inferências realizadas na base de dados da subestação de energia elétrica. A apresentação dos resultados é realizada no capítulo 3, enquanto no capítulo 4 são discutidos tais resultados. O capítulo 5 apresenta as conclusões deste trabalho de pesquisa. Por fim, são expressas as referências bibliográficas utilizadas e inserido o apêndice deste trabalho.

1.4. O SISTEMA ELÉTRICO BRASILEIRO E AS SUBESTAÇÕES DE ENERGIA ELÉTRICA

A energia elétrica pode ser gerada por meio de fontes renováveis de energia (a força das águas e dos ventos, o sol e a biomassa), ou não-renováveis (combustíveis fósseis e nucleares). No Brasil, onde é grande o número de rios, a opção hidráulica é mais utilizada e apenas uma pequena parte é gerada a partir de combustíveis fósseis, em usinas termelétricas.

O sistema elétrico brasileiro é nacionalmente interligado no chamado Sistema Interligado Nacional (SIN), e a maior parte da capacidade instalada no país é composta de usinas hidrelétricas, distribuídas pelas bacias hidrográficas nas diversas regiões do país. Em função desta peculiaridade, as usinas são construídas aproveitando as afluições e desníveis dos rios, nem sempre perto dos centros consumidores. Para transportar a energia gerada para os centros consumidores é necessária uma rede de transmissão que possibilite também o intercâmbio de energia entre as regiões geográficas, aproveitando a mudança de comportamento das vazões dos rios das diversas bacias hidrográficas e, inclusive, países vizinhos. O transporte para os centros de consumo é feito até subestações que se interligam com outras subestações de empresas distribuidoras ou, eventualmente, com consumidores de grande porte (BALTAZAR, 2007).

As subestações são responsáveis pela distribuição da energia elétrica. Antes de chegar aos estabelecimentos comerciais ou residenciais, a eletricidade percorre um sistema de transmissão que começa nas usinas e passa por estas estações, onde equipamentos chamados transformadores fazem o aumento ou a diminuição da tensão.

Quando elevam a tensão elétrica, os transformadores evitam a perda excessiva de energia ao longo do percurso. Quando rebaixam a tensão, já nos

centros urbanos, permitem a distribuição da energia pela cidade. Funcionam, deste modo, como pontos de entrega de energia para os consumidores.

Para se tornar adequada para o consumo, a energia passa por transformadores menores, instalados nos postes das ruas. Eles reduzem ainda mais a tensão que será recebida nas casas e estabelecimentos comerciais. A transmissão e distribuição da energia elétrica são feitas utilizando-se circuitos trifásicos a três fios (3 fases).

O fornecimento de energia é feito por meio de um grande e complexo sistema de subestações, linhas de transmissão e usinas, que constituem o Sistema Interligado Nacional (SIN).

1.5. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O contínuo desenvolvimento e progresso, atualmente verificados, em áreas como a tecnologia da informação, Internet, sistemas gerenciadores de bancos de dados, entre outros, tem proporcionado o armazenamento de dados em heterogêneas e robustas base de dados (RISSINO, 2009).

O termo KDD – *Knowledge Discovery in Databases* (descoberta de conhecimento em bases de dados), foi cunhado em 1989, com o objetivo de representar todo o processo de busca e extração de conhecimento que, em seu nível mais operacional, inclui a aplicação de técnicas e algoritmos de mineração de dados (*datamining*) para manipular e encontrar indícios de correlação ou de implicação em grandes volumes de dados e que envolve para tal fim, a aplicação de diversas áreas do conhecimento humano, como: estatística, matemática, inteligência artificial, etc. Segundo Castanheira (2008), para se iniciar um processo de KDD é necessário o entendimento do domínio da aplicação e dos objetivos finais a serem atingidos.

O KDD é empregado em vários campos de pesquisa com finalidades matemáticas, como por exemplo, estatísticas, aprendizado de máquinas, base de dados, reconhecimento de padrões, econômicas, desenvolvimento para técnicas e soluções para viabilizar a extração de dados de conhecimento em grandes bases de dados (FAYYAD et al., 1996a).

O processo de KDD é imprescindível para as bases devido ao rápido crescimento do volume de dados em tamanho e dimensionalidade; isso acontece

porque, em qualquer tipo de ação, como por exemplo, uma ligação de um aparelho telefônico, a utilização do cartão do banco ou até um simples exame médico são normalmente arquivadas em um computador na forma de dados.

Sem o auxílio do processo de KDD em uma base de dados com um grande volume, um usuário passaria décadas para achar uma informação em um arquivo específico que tenha sido enviado em apenas um dia, já com o auxílio do KDD, nessa mesma base de dados, essa informação seria facilmente encontrada em questão de segundos (FRAWLEY et al., 1992).

Para analisar grandes bases de dados precisamos de ferramentas computacionais apropriadas, pois seria inviável este procedimento de descoberta do conhecimento ser feito pelo homem sem auxílio do KDD. Atualmente, empresas que geram milhares de dados por dia necessitam de ferramentas automatizadas e inteligentes, na tarefa de interpretar, analisar e relacionar os dados da forma que possa ser desenvolvida uma estratégia para cada aplicação (GOLDSCHIMIDT; PASSOS, 2005).

Realizar um processo de KDD é algo que exige algumas considerações técnicas que devem ser levadas em consideração. Na maioria dos casos, um dos principais requisitos para o levantamento inicial é começar com um banco de dados original, que seja normalmente desenvolvido já para as próximas tarefas do KDD.

Após o banco, deve-se ter uma adequada coleta de dados para obter melhores percepções do KDD; esses requisitos para o levantamento inicial são os mais importantes, pois eles formam a base para o resto do processo do KDD (WILLIAMS e HUANG, 1996).

O processo de KDD ilustrado na Figura 1 envolve a avaliação e interpretação dos resultados com o objetivo de aplicar decisões sobre o que foi extraído envolvendo um processo que visa a tomada de decisões sobre as definições do conhecimento.

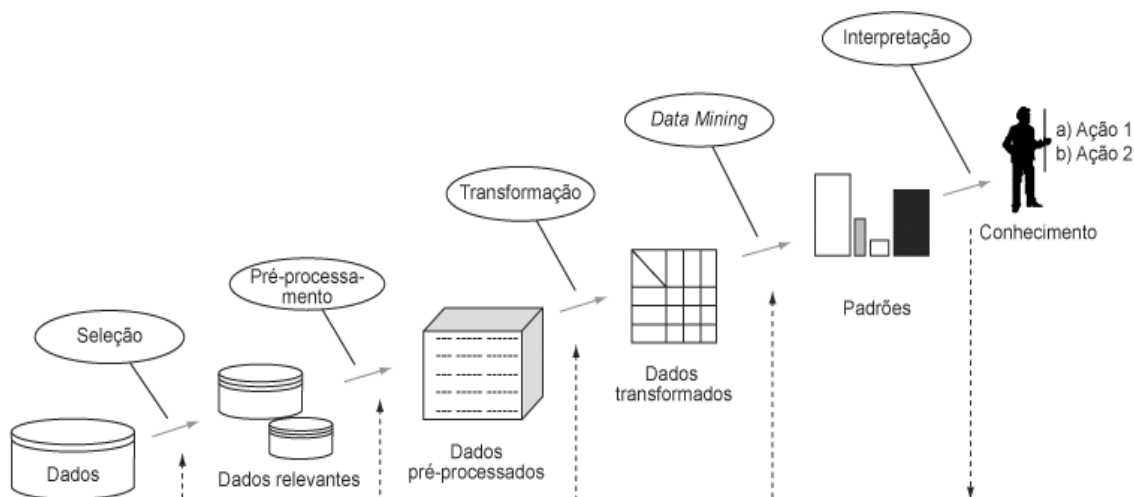


Figura 1 - Etapas do processo de KDD

Fonte: Fayyad et al. (1996)

A busca por conhecimento é um processo iterativo e interativo, que envolve três etapas principais, nas quais cada uma requer a capacidade de análise do usuário e de tomada de decisões. As principais etapas do processo de KDD são: pré-processamento, ligada a todas as funções relacionadas à organização, captação e ao tratamento dos dados; mineração de dados, na qual é realizada a busca por conhecimentos úteis na aplicação do KDD; pós-processamento, no qual é realizado o tratamento do conhecimento obtido através da mineração de dados com o objetivo de facilitar o entendimento do usuário em relação ao conhecimento descoberto (FAYYAD et al., 1996).

1.6. ETAPAS DA DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O processo de KDD é composto das seguintes etapas:

- a) Consolidação de dados: esta etapa tem como objetivo coletar os dados e consolidá-los para o processo de extração do conhecimento. Na mesma são definidos os atributos relevantes para o processo, são definidos os problemas e a forma de tratamento para os dados que não possuem valores e, se necessário, são removidas as exceções (WATSON, 2004).
- b) Seleção de dados: Nesta parte inicial do processo, é feita a seleção e a identificação das informações dos dados nas bases de dados que serão

utilizadas no processo de KDD e as informações selecionadas serão utilizadas durante todo o processo do KDD (GOLDSCHMIDT e PASSOS, 2005);

- c) Pré-processamento: o objetivo desta fase é limpeza dos dados, tanto quanto possível e adequá-los da melhor forma para sua utilização no processamento com as ferramentas de mineração de dados. Os tipos de pré-processamento incluem tradicional projeção do banco de dados, seleção e mapeamento do valor e das funções de classificação e limpeza de dados (WILLIAMS e HUANG, 1996). Nesta fase as etapas a serem seguidas são: limpeza, codificação, enriquecimento e normalização, conforme ilustradas na Figura 2;

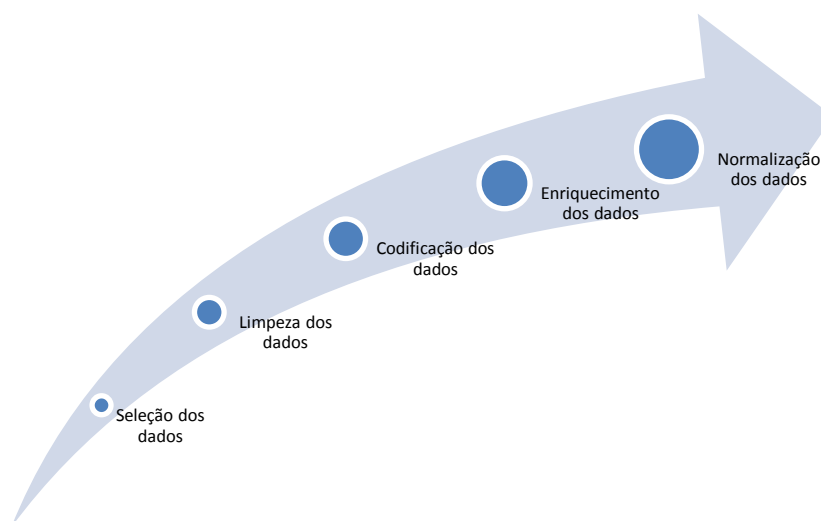


Figura 2 - Etapas do processo de pré-processamento do KDD.

- d) Mineração de dados: a mineração de dados é o processo de descoberta de informações úteis em grandes depósitos de dados (TAN et al., 2009). As técnicas de mineração de dados são planejadas, elaboradas e organizadas para atuar sobre os dados com o intuito de identificar e identificar padrões úteis e recentes que poderiam de outra forma, permanecer ignorados;
- e) Pós-processamento: nesta etapa existem várias medidas para tratar o conhecimento descoberto; como por exemplo a visualização, análise e

interpretação gerada pela etapa de mineração de dados. Envolve a depuração e as sínteses dos padrões nos volumes de dados. O volume de conhecimento dos dados é muito extenso, praticamente é um conjunto de informações relevantes ao processo que auxilia na análise das tomadas de decisões. Para auxiliar, nesse processo são utilizadas técnicas como padrões redundantes, relações irrelevantes, entre outros (CALIL et al., 2008). Nessa etapa é muito comum ver o especialista do domínio da aplicação e o especialista de KDD trabalharem juntos, avaliando os resultados e os modelos de conhecimento e assim, definirem a melhor forma de utilizar os dados minerados.

1.7. MINERAÇÃO DE DADOS

A mineração de dados consiste em um conjunto de conceitos e métodos com o objetivo de encontrar uma informação compreensível e útil para o usuário em um determinado conjunto de dados (RUSSELL, 2011).

De acordo com Klosgen (1992), a descoberta de relações nos dados compreende todas as instâncias de padrões selecionados no espaço das hipóteses que sejam suficientemente interessantes.

Em alguns estudos, pode-se encontrar referências ao termo “mineração de dados” com diferentes nomes tais como *Knowledge Mining from Data* (Mineração de Conhecimento em Bases de Dados), *Knowledge Extraction* (Extração do Conhecimento), *Data/Patterns Analysis* (Análise de Dados/Padrões), *Data Archaeology* (Arqueologia de Dados) entre outros.

Alguns autores consideram como mineração de dados a busca do conhecimento única e exclusivamente em grandes concentrações de dados e utilizando meios computacionais automáticos para tal finalidade.

Definimos mineração de dados como o uso de técnicas automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações que, devido ao volume de dados, não seriam facilmente descobertas a olho nu pelo ser humano. (CARVALHO, 2005).

Porém, estes dados devem ser anteriormente preparados e estruturados para obtermos um melhor resultado com as técnicas de *data mining*, salvo que a origem do dado não precisa necessariamente ser proveniente de uma fonte computacional/digital.

Da mesma forma que um certo tipo de dado pode ser processado na busca de tendências ou padrões, dados textuais também usufruem dessa vantagem. Textos podem ser transformados em uma estrutura mais adequada para um processamento posterior dos dados com métodos de extração do conhecimento. (AMARAL, 2011).

Talvez, a definição que mais contempla o termo *data mining* como sendo este um dos passos dentro do processo de KDD tenha sido elaborada por Usama Fayyad (1996a): "*O data mining é um passo dentro do processo de KDD que consiste na aplicação de análise de dados e algoritmos de descoberta de conhecimento que produzem uma enumeração de padrões ou modelos sobre os dados*". Tal afirmação vem sendo fundamentada com processos e técnicas cada vez mais sofisticadas para a obtenção do conhecimento nas grandes massas de dados.

Amaral (2011) diz que o processo de descoberta do conhecimento é um ato trivial para os humanos visto que os mesmos possuem a capacidade de analisar um fato de diversas perspectivas, indo de um ponto a outro do problema instintivamente para obtenção da conclusão e percebendo significância sobre aquilo que está sendo lido.

De forma resumida, pode-se concluir que a mineração de dados caracteriza-se pela existência de um algoritmo que diante de uma tarefa proposta será eficiente em extrair conhecimento implícito e útil de um banco de dados (CASTANHEIRA, 2008).

Apesar de o desempenho de processamento de dados das máquinas ser superior ao dos humanos em alguns aspectos, esse não tem as principais características humanas na forma de “pensar”, fazendo correlações entre assuntos, definindo relevâncias e olhando um mesmo problema de várias perspectivas levando-as em consideração antes de uma tomada de decisão. Desta forma, o que os programas fazem é simular, grosseiramente, estas características tomando

decisões a partir de critérios previamente estabelecidos e relevantes e ainda com base em suposições plausíveis definidas por um especialista humano.

Sendo assim, temos diante de nós todas as informações armazenadas diariamente, mas torna-se impossível para os humanos em virtude da grande quantidade de dados, uma análise sobre todo o conteúdo em busca de padrões que sejam inerentes a um negócio. Diante deste cenário, o *data mining* é utilizado dentro do processo KDD com o intuito de sintetizar a informação realçando os indicadores dos dados em análise.

“Em *data mining* podemos encontrar a combinação de conjuntos de dados para capturar padrões que são muito sutis ou complexos para serem detectados apenas por analistas de dados devido ao grande volume de dados armazenados” (KREUZE, 2001).

Os métodos de mineração de dados são divididos em aprendizado não supervisionado e supervisionado. A principal diferença entre os mesmos consiste no fato de que os primeiros não necessitam de um método de mineração de dados definido para os registros, ou seja, não é necessário um atributo alvo.

No aprendizado supervisionado, os métodos são definidos através de um conjunto de dados que possuem um atributo alvo pré-definido e os registros são categorizados em relação a esse atributo.

Durante o processo de mineração de dados, diversos métodos podem ser testados e utilizados de forma combinada a fim de viabilizar comparações entre os métodos, permitindo assim a escolha do melhor método. Diversas funcionalidades são usadas para especificar os tipos de padrões que podem ser encontrados na tarefas de mineração de dados. Essas funcionalidades se referem a técnicas que podem ser usadas individualmente ou em conjunto para a descoberta de padrões (DOMINGUES, 2003).

1.8. PRINCIPAIS TAREFAS DE MINERAÇÃO DE DADOS

1.8.1. CLASSIFICAÇÃO

A tarefa de classificação tem como objetivo descobrir a função chave que mapeia um conjunto de registros em um conjunto de classes predefinidas. Quando

os registros são descobertos, a função é aplicada sobre eles e prevê a classe em que esses registros se enquadram.

A classificação consiste em obter um modelo baseado em um conjunto de exemplos que descrevem uma função desconhecida. Esse modelo é utilizado posteriormente para fornecer o valor de atributo para novos exemplos (CASTANHEIRA, 2008).

1.8.2. CLUSTERIZAÇÃO

Assim como ocorre com a tarefa de classificação, a tarefa de clusterização busca identificar grupos de dados com características semelhantes, porém, diferenciando-se da primeira, pelo fato dos grupos de interesse não serem predefinidos (CIGRÉ, 2006).

1.8.3. REGRESSÃO

A tarefa de regressão é muito parecida com a tarefa de classificação, porém restrita apenas para atributos numéricos. Essa tarefa tem como objetivo buscar uma função que mapeie os registros do banco de dados em valores reais. Segundo Uysal e Guvenir (2004), a regressão também é conhecida por predição funcional, predição de valor real, função e aproximação, ou ainda, aprendizado de classes contínuas.

1.8.4. REGRAS DE ASSOCIAÇÃO

Petermann (2006) cita que a técnica de associação é a descoberta de regras associativas em um determinado conjunto de dados que evidenciam condições que ocorrem frequentemente juntas. Em um subconjunto de dados de transações de vendas de um comércio, a aplicação de técnicas associativas poderia evidenciar qual é o tipo de cliente que mais compra determinado tipo de produto e com qual frequência.

Uma regra de associação caracteriza o quanto a presença de um conjunto de itens nos registros de uma base de dados implica na presença de algum outro conjunto distinto de itens nos mesmos registros (AGRAWAL e SPIKANT, 2004).

1.8.5. AGRUPAMENTO

Fayyad (1996) cita que a técnica de agrupamento ou cluster se propõe a identificar um conjunto finito de categorias ou grupos para descrever os dados. Podemos concluir que a técnica de agrupamento ou cluster tem como foco identificar os grupos de informações semelhantes das bases de dados as quais foi aplicada, e que estes grupos descrevem o tipo de dados existentes em tal base.

Trata-se de um processo de partição de uma população heterogênea em vários subgrupos mais homogêneos. No agrupamento, os registros são agrupados de acordo com semelhanças (CASTANHEIRA, 2008).

1.8.6. ESTIMATIVA

Estimativa é a geração de valores contínuos, ou seja, conjunto de valores dentro de um intervalo (escala) para cada tipo de dado em análise e com o fornecimento de alguns dados, usa-se a estimativa para descobrir o valor de alguma variável contínua.

Estimar o número de filhos em famílias de determinada região, estimar a renda total de uma família, estimar o valor em tempo de vida de um cliente ou até mesmo estimar o custo de ligações telefônicas de diferentes grupos de pessoas são exemplos dos resultados que podem ser obtidos pela estimativa.

1.8.7. DESVIO

A tarefa de detecção de desvios tem como objetivo descobrir registros na base de dados com características que não atendam o perfil do contexto. Através dessa tarefa são encontrados registros anormais utilizados para organizar diferentes grupos. Na estatística é possível encontrar ferramentas que utilizam essa tarefa.

1.9. FUNDAMENTOS DA LÓGICA PARACONSISTENTE ANOTADA E ALGORITMOS RELACIONADOS

Ao que tudo indica, a lógica como ciência foi descoberta por Aristóteles (384-322 a.C.) e seus seguidores na Grécia antiga. A lógica pode ser definida como a ciência que estuda as leis dos raciocínios válidos (DA SILVA FILHO e ABE, 2001).

As lógicas como sistemas de inferências, podem ser classificadas, sem muito rigor, da seguinte forma: lógica clássica e lógica não clássica (DA COSTA, 1999). O raciocínio lógico clássico é fundamentado em quatro princípios que são apresentados por meio de símbolos usualmente empregados na Lógica Clássica.

São eles:

1. Princípio da identidade, $\rho = \rho$
Toda proposição ou objeto é idêntico a si mesmo.
2. Princípio da identidade proposicional, $\rho \rightarrow \rho$
Toda proposição implica ela mesma.
3. Princípio do terceiro excluído, $\rho \vee \neg \rho$
De duas proposições contraditórias, isto é, uma nega a outra, uma delas é verdadeira.
4. Princípio da não-contradição, $\neg(\rho \vee \neg \rho)$
Entre duas proposições contraditórias, uma delas é falsa.

Dentre desse raciocínio a Lógica Clássica é binária, portanto uma declaração é falsa ou verdadeira, não admitindo ser ao mesmo tempo parcialmente verdadeira e parcialmente falsa (DA SILVA FILHO, 2008).

Neste cenário, considerando que uma declaração não pode contradizer a outra, a Lógica Clássica cobriu todas as possibilidades, criando assim a base do pensamento lógico ocidental e seus princípios e fundamentos ainda são, nos dias atuais, utilizados no funcionamento de grande parte dos equipamentos e sistemas computacionais, visto que são adequados para representar o conhecimento.

As crescentes demandas atuais do mercado tecnológico exigem equipamentos cada vez mais capazes de controlar e administrar situações nunca antes previstas em um ambiente de elevada complexidade e volumes de dados de

processamento, colocando à prova a qualidade das respostas apresentadas pela aplicação exclusiva dos princípios rígidos da lógica clássica.

1.10. AS LÓGICA NÃO-CLÁSSICAS

Uma lógica não-clássica é toda aquela que de alguma forma apresenta fundamentos que possam contrariar ou ignorar algum princípio da lógica clássica. Como as situações reais não se enquadram inteiramente nas formas binárias da lógica clássica, vários pesquisadores se esforçaram, no sentido de encontrar outras formas que permitissem enquadrar melhor outros conceitos como as indefinições, ambiguidades e inconsistências, com isso surgiram as lógicas não-clássicas (MARTINS, 2003).

No mundo real, seres humanos (ou agentes de forma geral) têm conhecimentos inconsistentes de forma sutil. Podem acreditar em p e $\neg p$ (a negação de p) para alguma proposição p . Isto pode ocorrer, por exemplo, se um agente acredita em p e $\neg q$ sem perceber que p e q são logicamente equivalentes.

As lógicas não-clássicas são divididas em duas categorias principais as complementares e as alternativas.

As complementares são as que não infringem os princípios básicos da lógica clássica e não questionam sua validade universal, apenas ampliam e complementam o seu escopo enriquecendo-a com a introdução de novos operadores.

São exemplos de lógicas complementares: lógicas modais, lógicas deônticas, lógicas do tempo, lógicas epistêmicas, lógicas imperativas, etc.

As lógicas heterodoxas (alternativas, desviantes), rivais da lógica clássica, foram concebidas como novas lógicas, destinadas a substituir a lógica clássica em alguns domínios do saber. Elas anulam alguns princípios básicos da lógica clássica. Hoje, são concebidas infinitas lógicas possíveis, tais como a lógica intuicionista de Brouwer-Heyting, a lógica modal clássica, a lógica modal polivalente, a lógica clássica categorial, a lógica difusa, a lógica paraconsistente, entre outras (MARTINS, 2012).

As inúmeras pesquisas comprovam a relevância dos sistemas não-clássicos, não apenas do ponto de vista prático, mas também teórico, quebrando um

paradigma do pensamento humano, visto que foram obtidas teorias, nas quais, as inconsistências são permitidas sem o perigo da trivialização.

1.11. A LÓGICA PARACONSISTENTE

A Lógica Paraconsistente, que se destaca por infringir o princípio da não-contradição, surgiu dos trabalhos formalizados por volta de 1948 de maneira independente pelo polonês Stanislaw Jaskowski e pelo brasileiro Newton C. A. da Costa, embora houvessem precursores como N. Vasílev e J. Lukasiewicz (MARIO, 2006). Quanto aos sistemas iniciais de Lógica Paraconsistente, contendo todos os níveis lógicos, envolvendo, pois, cálculos proposicionais de predicados e de descrições, bem como lógicas de ordem superior, devem-se a Newton C. A. da Costa (1954 em diante).

Para Bispo e Cazarini (2006), a paraconsistência estará sempre averiguando se existe algum nível de inconsistência, divergência ou contradição nas diferentes análises.

As lógicas paraconsistentes são uma classe de lógica não-clássica e foram edificadas para se encontrar meios de dar tratamento não-trivial às situações contraditórias. Aplicações das lógicas paraconsistentes em diversos campos do conhecimento vêm demonstrando que estas são mais adequadas no enquadramento de problemas ocasionados por situações de contradições que aparecem quando lidamos com descrições do mundo real (DA COSTA, 1999).

Segundo Da Silva Filho (2008), dada uma teoria (dedutiva) T , assentada na Lógica L , diz-se que ela é consistente se entre os seus teoremas não existir um que seja a negação do outro; em hipótese contrária, T denomina-se inconsistente. A teoria T chama-se trivial se todas as sentenças (fórmulas fechadas) de sua linguagem forem teoremas. Se isto não ocorrer, T é não-trivial. Se L for uma das lógicas comuns, tal como a clássica, a teoria T é trivial se e somente se for inconsistente. A Lógica L denomina-se Paraconsistente se puder funcionar como fundamento de teorias inconsistentes e não-triviais.

Outro conceito significativo para o que se segue é o de lógica paracompleta. Uma lógica é chamada paracompleta se pode funcionar como a lógica subjacente de teorias na qual há fórmulas tais que estas fórmulas e suas negações são ambas falsas. Como exemplo, suponha que temos duas informações sendo, uma de crença

indicando em relação à presença de um objeto que este não existe, portanto, apresentando uma falsidade e, outra informação de descrença onde esta também indica falsidade que, neste caso, indicaria a presença de um objeto, essa lógica seria denominada paracompleta pelas duas falsidades apresentadas. Uma teoria é chamada paracompleta se sua lógica subjacente é uma lógica paracompleta.

Aplicações de um tipo de Lógica Paraconsistente denominada de Lógica Paraconsistente Aplicada (LPA), apresentaram trabalhos cujos resultados relevantes estão dispostos na literatura especializada. Entre os trabalhos publicados, destaca-se o algoritmo Para-analisador (DA SILVA FILHO, 1999), no qual, métodos computacionais de análises paraconsistentes podem ser projetados através do algoritmo da Lógica Paraconsistente Aplicada.

Apresentam-se a seguir um resumo da Lógica Paraconsistente e a sua extensão denominada de Lógica Paraconsistente Anotada (LPA).

1.12. A LÓGICA PARACONSISTENTE ANOTADA

As Lógicas Paraconsistentes Anotadas são uma família de lógicas não-clássicas inicialmente empregadas em programação lógica por Subrahmanian (SUBRAHMANIAN, 1987).

Os primeiros estudos sobre os fundamentos da Lógica Paraconsistente Anotada foram efetuados por Da Costa, Vago e Subrahmanian (DA COSTA, VAGO e SUBRAHMANIAN, 1991), por Da Costa, Abe e Subrahmanian (DA COSTA, ABE e SUBRAHMANIAN, 1991) e Abe (ABE, 1992).

Uma Lógica Paraconsistente Anotada (LPA) é uma classe de Lógica Paraconsistente que possui um reticulado associado que desempenha um papel importante na sua representação.

Na LPA as fórmulas proposicionais vêm acompanhadas de anotações. Cada anotação μ , pertencente a um reticulado finito τ , atribui valores à sua correspondente fórmula proposicional ρ . Considerando uma lógica evidencial as anotações vêm representadas por Graus de evidência. A LPA pode ser estudada em um Quadrado unitário no Plano Cartesiano (Figura 3) onde são inseridos os graus de evidência favorável μ e grau de evidência desfavorável λ .

Verifica-se que podemos obter quatro situações possíveis:

1-Inconsistente; no ponto $(1, 1) = T$

2- Verdadeira; no ponto $(1, 0) = V$

3- Falsa; no ponto $(0, 1) = F$

4- Indeterminada ou paracompleta; no ponto $(0, 0) = \perp$

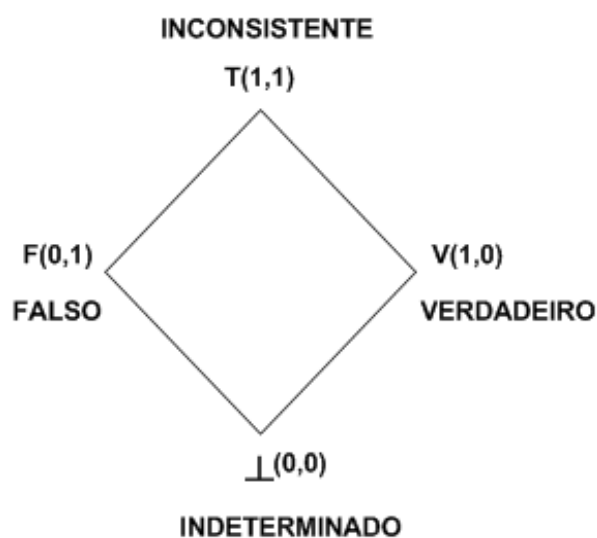


Figura 3 - Reticulado Representativo De Hasse.

Fonte: Da Silva Filho (2001).

Intuitivamente, em tal Reticulado tem-se:

$P_{(\mu,\lambda)} = P_{(1,0)}$: indicando 'existência de evidência favorável total e evidência desfavorável nula', atribuindo uma conotação lógica de *Verdade* à proposição.

$P_{(\mu,\lambda)} = P_{(0,1)}$: indicando 'existência de evidência favorável nula e evidência desfavorável total', atribuindo uma conotação lógica de *Falsidade* à proposição.

$P_{(\mu,\lambda)} = P_{(1,1)}$: indicando 'existência de evidência favorável total e evidência desfavorável total' atribuindo uma conotação lógica de *Inconsistência* à proposição.

$P_{(\mu,\lambda)} = P_{(0,0)}$: indicando 'existência de evidência favorável nula e evidência desfavorável nula', atribuindo uma conotação lógica de *Indeterminação* à proposição.

Quando as proposições vêm acompanhadas de anotações, ou graus de evidência, é possível aplicação real da Lógica Paraconsistente em Sistemas de Análises e tomadas de decisão.

1.13.A LÓGICA PARACONSISTENTE ANOTADA COM ANOTAÇÃO DE 2 VALORES (LPA2v)

Na Lógica Paraconsistente Anotada com dois valores (LPA2v) cada proposição utiliza dois valores de graus denominados; graus de evidência favorável μ e desfavorável λ . O primeiro valor da anotação representa a evidência favorável à proposição p , ou seja, o grau de evidência favorável dessa proposição (denominado μ_1). O segundo valor da anotação representa a evidência contrária à proposição p , ou seja, o grau de evidência desfavorável dessa proposição (denominado μ_2).

Em várias situações reais onde é utilizada a LPA2v, os graus de evidência favorável e desfavorável são considerados como informações de entrada do sistema (exemplo: sinais de entrada de corrente, tensão, etc.) e os estados lógicos representados nos vértices e nas regiões internas do reticulado são as saídas resultantes da análise paraconsistente. Na Figura 4 abaixo é apresentado o diagrama correspondente:

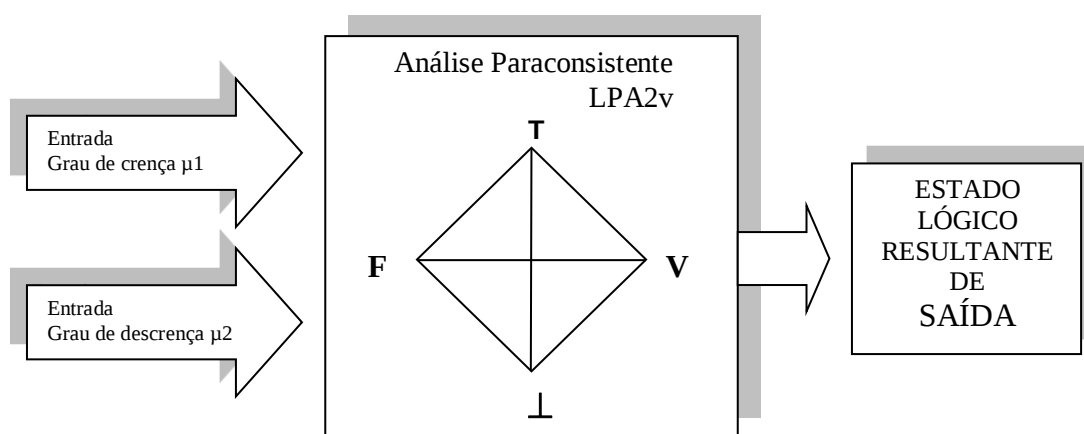


Figura 4 - Representação da análise Paraconsistente LPA2v.

Fonte: Da Silva Filho (2001).

1.14. REPRESENTAÇÃO NO QUADRADO UNITÁRIO NO PLANO CARTESIANO – QUPC

A LPA pode ser estudada em um quadrado unitário no Plano Cartesiano (Figura 5) onde são inseridos os graus de evidência favorável μ e graus de evidência desfavorável λ .

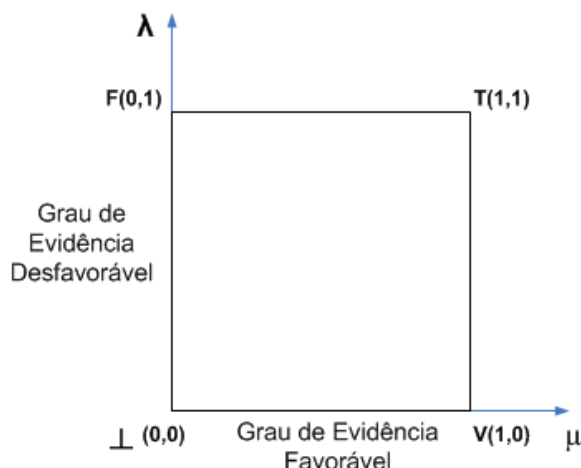


Figura 5 - Graus de evidência no Quadrado Unitário do Plano Cartesiano.
 Fonte: Da Silva Filho (2001).

1.14.1. O RETICULADO DA LPA COM REGIÕES DELIMITADAS

Considerando os valores encontrados no quadrado unitário do plano cartesiano pode-se estender a análise em uma representação de 2 eixos: um com os valores do grau de contradição e outro com os valores do grau de certeza. Estes dois eixos são sobrepostos de tal forma a serem comparados com o reticulado do LPA2v, onde pode-se delimitar regiões as quais serão comparadas a determinados estados lógicos paraconsistentes.

Conforme é visto na Figura 8, para a delimitações das regiões os dois limites externos e arbitrários V_{sc} e V_{ic} determinam quando o grau de Certeza resultante é alto o suficiente para que a proposição analisada seja considerada totalmente verdadeira ou totalmente falsa. Da mesma forma os dois limites externos e arbitrários V_{scct} e V_{icct} determinam quando o grau de Contradição resultante é alto o suficiente para que a proposição analisada seja considerada totalmente consistente ou totalmente inconsistente.

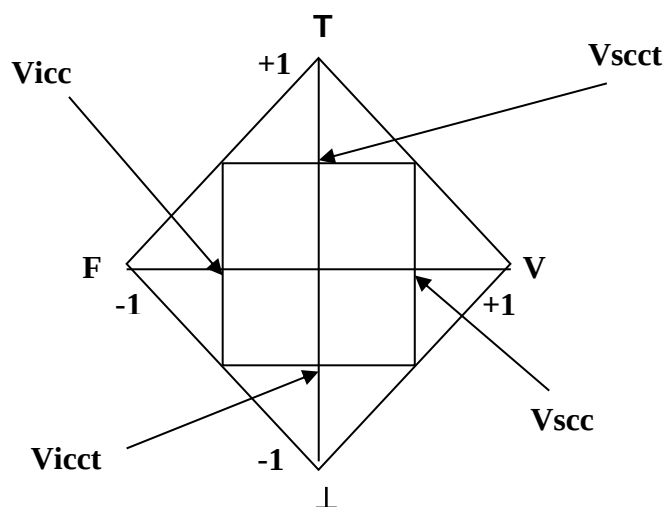


Figura 6 - Representação do Reticulado LPA2v com valores extremos

Fonte: Da Silva Filho (2001).

O reticulado pode ser delimitado conforme a relação abaixo:

Vsc: Valor superior de controle de Certeza (variando entre 0 e +1);

Vic: Valor inferior de controle de Certeza (variando entre 0 e -1);

Vscct: Valor superior de controle de Contradição (variando entre 0 e +1);

Vicct: Valor inferior de controle de Contradição (variando entre 0 e -1).

1.14.2. O GRAU DE CERTEZA

O grau de certeza G_c é definido como sendo o valor que representa a interpolação entre os graus de evidência favorável e desfavorável no quadrado unitário no plano cartesiano representados pelos pontos A e C onde, A = (0,0) - Indeterminado ao ponto C = (1, 1) Inconsistente. Encontra-se no intervalo real fechado [-1,1].

Portanto, o grau de certeza G_c é calculado através da equação:

$$G_c = \mu_1 - \mu_2 \quad (1)$$

Sendo:

$$0 \leq \mu_1 \leq 1$$

$$0 \leq \mu_2 \leq 1$$

Segundo (DA SILVA FILHO, 2008):

O valor $G_c=-1$ é estabelecido no ponto $B = (0,1)$ representando certeza máxima na negação da proposição e o valor $G_c=+1$ é estabelecido no ponto $D = (1,0)$ significando que temos uma certeza máxima na afirmação da proposição. De acordo com os sensores utilizados em situações reais de uso, quando chegam valores que resultam nesses graus de contradição, podemos afirmar que essas informações são consistentes. Quanto mais a interpolação entre os graus de crença e de descrença se aproximarem do segmento de reta AC, representado no QUPC, mais o resultado da subtração dos graus de crença pelo grau de descrença ($\mu_1 - \mu_2$) se aproxima de 0, diminuindo, assim, o valor de G_c . Esta diminuição de G_c representa uma menor certeza entre as informações na entrada porque significa uma maior coincidência entre os Graus de crença e de descrença. Quando os graus de crença e de descrença forem de valores iguais ($\mu_1 = \mu_2$), o grau de certeza é zero e o ponto de interpolação estará sobre a reta AC. (DA SILVA FILHO, 2001).

Com o procedimento básico de interpretar os valores através de um quadrado unitário descrito no plano cartesiano podem-se obter valores diversos de G_{ct} e G_c a partir dos graus evidência recebido todos normalizados e, portanto valorados na faixa entre 0 e 1 inclusive. Os valores dos graus de Certeza e de Contradição podem ser dispostos em dois eixos: “Eixo dos Graus de Contradição G_{ct} ” e “Eixo dos Graus de certeza G_c ” que, sobrepostos e cruzando-os no ponto de Indefinição, onde os graus de evidências valem 0,5 representam o reticulado da LPA com valores. Para qualquer anotação, composta por valores de graus de evidência favorável e graus de evidência desfavorável, pode-se obter um ponto de Interpolação entre valores de Graus de Certeza e Graus de Contradição, que é pertencente ao reticulado da LPA2v. Dessa forma, para qualquer anotação, com valores de graus de evidência favorável e de evidência desfavorável, são obtidos os valores do Grau de Certeza e de Contradição, relacionado a determinada proposição. Fazendo a interpolação dos dois valores pode-se visualizar no Reticulado onde está localizado, o ponto resultante considerado como o estado lógico Paraconsistente ε_r .

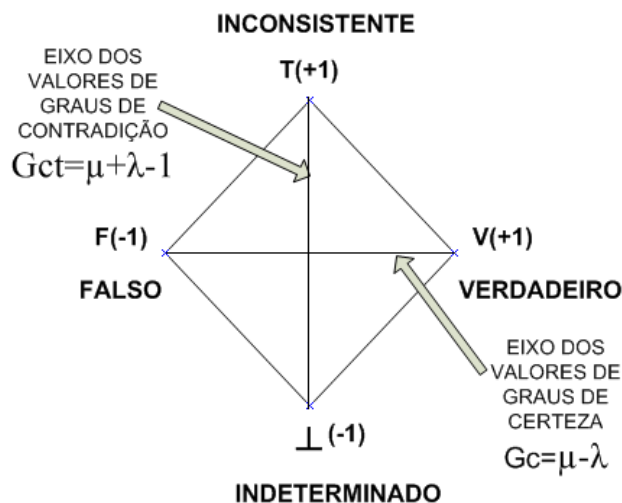


Figura 7 - Eixos dos Graus de Contradição e de Certeza.

Fonte: Da Silva Filho (2001).

Na prática a tomada de decisão pode ser feita relacionando o resultado com proximidade, ou não, de algum dos estados lógicos extremos localizados nos vértices. Por exemplo, intuitivamente, em tal reticulado, (1.0, 0.0) indica ‘evidência favorável total’, (0.0, 1.0) indica ‘evidência desfavorável total’, (1.0, 1.0) indica ‘evidência favorável totalmente inconsistente’ e (0.0, 0.0) indica ‘evidência favorável totalmente indeterminada’.

As representações da proposição com a anotação composta por dois Graus de Evidência são descritas como se segue:

$p_T = p_{(1, 1)} \Rightarrow$ A anotação composta pelos graus de crença = 1 e de descrença = 1 atribui à proposição p uma leitura intuitiva de que p é inconsistente.

$p_1 = p_{(1, 0)} \Rightarrow$ A anotação composta pelos graus de crença = 1 e de descrença = 0 atribui à proposição p uma leitura intuitiva de que p é verdadeira.

$p_0 = p_{(0, 1)} \Rightarrow$ A anotação composta pelos graus de crença = 0 e de descrença = 1 atribui à proposição p uma leitura intuitiva de que p é falsa.

$p_{\perp} = p_{(0, 0)} \Rightarrow$ A anotação composta pelos graus de crença = 0 e de descrença = 0 atribui à proposição p uma leitura intuitiva de que p é indeterminada (DA SILVA FILHO, 2001).

1.14.3. O GRAU DE CONTRADIÇÃO

Conforme visto em (DA SILVA FILHO, 2008) a partir do Quadrado unitário são elaboradas transformações nas quais se podem calcular os valores dos graus de contradição G_{ct} e dos graus de certeza G_c referentes aos dois graus de evidência. Uma análise pode ser feita no Quadrado Unitário no Plano Cartesiano (QUPC) onde os graus de evidência favorável estão dispostos no eixo x e os graus de evidência desfavorável estão dispostos no eixo y, como mostrado na Figura 8:

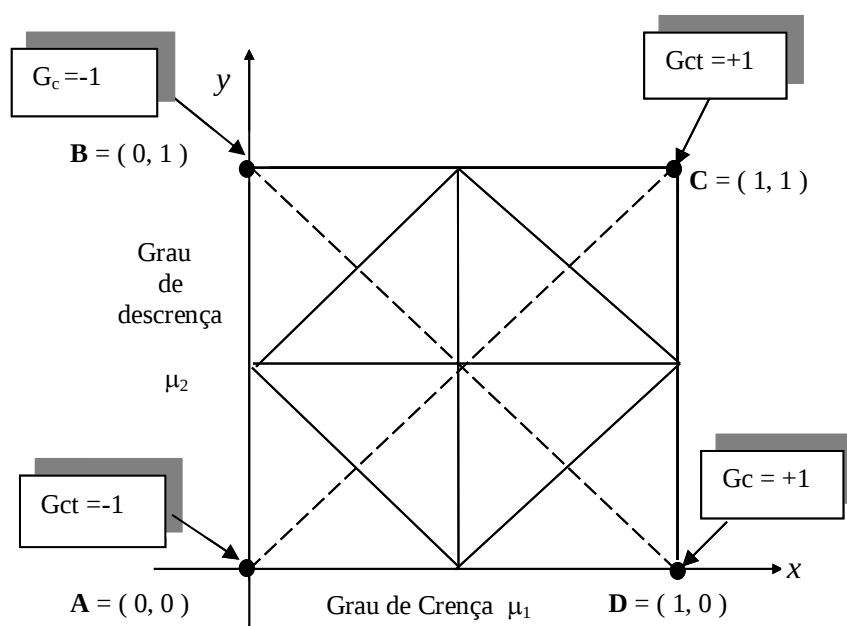


Figura 8 - Quadrado unitário no Plano Cartesiano.

Fonte: Da Silva Filho (2001).

O grau de contradição G_{ct} é definido como sendo o valor que representa a interpolação entre os graus de evidência favorável e de evidência desfavorável no quadrado unitário no plano cartesiano representados pelos pontos B e D onde, B = (0, 1) – Falso e D = (1,0) - Verdadeiro. Encontra-se no intervalo real fechado [-1,1]. Portanto, o grau de contradição G_{ct} é calculado através da equação:

$$G_{ct} = \mu_1 + \mu_2 - 1 \quad (1)$$

$$\text{Sendo: } 0 \leq \mu_1 \leq 1 \quad \text{e} \quad 0 \leq \mu_2 \leq 1$$

Neste contexto Da Silva Filho (2008) afirma que:

O valor $G_{ct}=-1$ é estabelecido no ponto A = (0,0) representando contradição máxima negativa e o valor $G_{ct}=+1$ é estabelecido no ponto C = (1,1) significando que temos uma contradição máxima positiva. De acordo com os sensores utilizados em situações reais de uso, quando chegam valores que resultam nesses graus de contradição, podemos afirmar que essas informações são completamente contraditórias. Em um Sistema de Análise Paraconsistente, quanto mais a interpolação entre os graus de crença e de descrença se aproximarem do segmento de reta BD, representado no QUPC, mais o resultado da soma dos graus de crença e de descrença ($\mu_1 + \mu_2$) se aproxima de 1, diminuindo, assim, o valor de G_{ct} . Esta diminuição de G_{ct} representa uma menor contradição entre as informações na entrada. Quando a soma dos graus de crença e de descrença ($\mu_1 + \mu_2$) for igual a 1, o grau de Contradição é zero e o ponto de interpolação estará sobre a reta BD. Neste caso, $G_{ct} = 0$ e não há contradição entre os sinais. (DA SILVA FILHO, 2008).

1.14.4. O GRAU DE CERTEZA REAL

A partir de um estudo mais detalhado visto em [10] pode-se encontrar o Grau de Certeza Real (G_{CR}) como um valor projetado no eixo dos graus de certeza do Reticulado através das equações:

$$G_{CR} = 1 - \sqrt{(1 - |G_c|)^2 + G_{ct}^2} \quad \text{se } G_c > 0$$

$$G_{CR} = 1 - \sqrt{(1 - |G_c|)^2 + G_{ct}^2} - 1 \quad \text{se } G_c < 0$$

E a partir de G_{CR} pode-se encontrar o seu valor normalizado, denominado de Grau de Evidência Resultante (μ_{ER}). Portanto:

$$\mu_{ER} = \frac{G_{CR} + 1}{2}$$

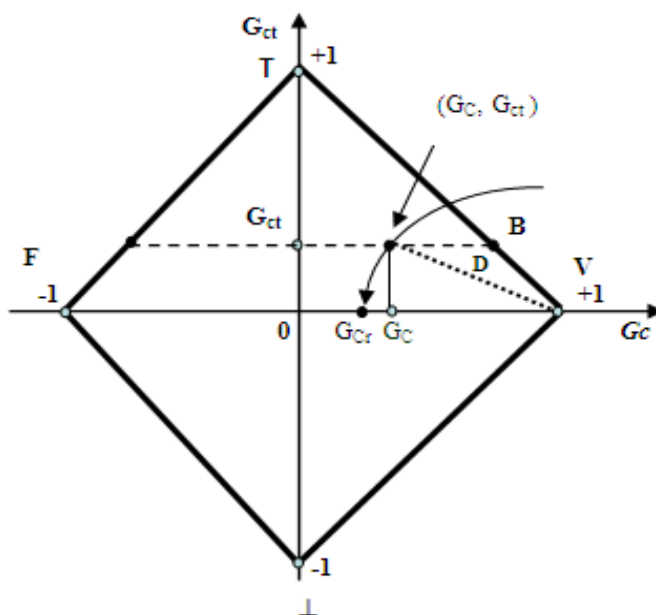


Figura 9 - Projeção do Valor da Distância D sobre o eixo dos graus de certeza para obtenção do Grau de Certeza Real G_{CR} no reticulado da LPA2V.

Fonte: Da Silva Filho (2009).

1.14.5. ALGORITMOS UTILIZADOS NAS ANÁLISES PARACONSISTENTES

Das equações da Lógica Paraconsistente Anotada de dois valores (LPA2v) são extraídos algoritmos para tratamento e análise de sinais de informação aplicados na forma de graus de evidência (DA SILVA FILHO, 2009). Todos os procedimentos relacionados ao tratamento lógico paraconsistente estarão relacionados à análise efetuada por um algoritmo denominado de NAP (Nó de Análise Paraconsistente), conforme descrito a seguir.

1.14.6. NÓ DE ANÁLISE PARACONSISTENTE

Segundo Da Silva Filho (2008), o Nó de Análise Paraconsistente é a denominação dada a um algoritmo extraído da LPA2v para análise de uma proposição através da entrada na forma de anotação composta por um Grau de Evidência Favorável (μ) e um Grau de Evidência Desfavorável (λ). Na saída se obtém o Grau de Evidência Resultante Real (μ_{ER}) (quando normalizado) através de um intervalo de certeza (ϕ_E).

Segue abaixo o descritivo e o símbolo de um NAP típico que será utilizado nas análises.

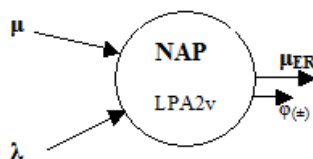


Figura 10 - Símbolo do NAP – Nó de Análise Paraconsistente.

Fonte: Da Silva Filho (2009).

1. Entre com os valores de entrada

μ */ Grau de Evidência favorável $0 \leq \mu \leq 1$

λ */ Grau de Evidência desfavorável $0 \leq \lambda \leq 1$

2. Calcule o Grau de Certeza

$$G_c = \mu - \lambda$$

3. Calcule o Grau de Contradição

$$G_{ct} = (\mu + \lambda) - 1$$

4. Calcule a distância d

$$d = \sqrt{(1 - |G_c|)^2 + G_{ct}^2}$$

5. Calcule o Grau de Contradição normalizado

$$\mu_{ctr} = \frac{(\mu + \lambda)}{2}$$

6. Calcule o Intervalo de Evidência resultante

$$\varphi_E = 1 - |2\mu_{ctr} - 1|$$

7. Determine o sinal da saída

Se $\varphi_E \leq 0,25$ ou $d > 1$, então faça: $S1 = 0,5$ e $S2 = \varphi_{E(\pm)}$: Indefinição e vá para o item 11, senão vá para o próximo item.

8. Determine o Grau de Certeza real

$$\text{Se } G_c > 0 \quad G_{CR} = (1 - d)$$

$$\text{Se } G_c < 0 \quad G_{CR} = (d - 1)$$

9. Determine o Grau de Resultante real

$$\mu_{ER} = \frac{G_{CR} + 1}{2}$$

10. Apresente os resultados na saída

$$\text{Faça } S1 = \mu_{ER} \quad \text{e} \quad S2 = \varphi_{E(\pm)}$$

11. Fim

Os algoritmos NAPs podem ser interligados formando redes paraconsistentes de tratamento de dados. Nesta dissertação será utilizado um algoritmo denominado de Extrator de Efeitos da Contradição que é formado por diversos NAPs. O algoritmo Extrator de Efeitos da Contradição será descrito a seguir.

1.14.7. ALGORITMO EXTRATOR DE EFEITOS DA CONTRADIÇÃO

O algoritmo Paraconsistente Extrator de efeitos da Contradição é capaz de extrair gradativamente os efeitos da contradição em sinais de informação originados de banco de dados de conhecimento incerto. Este algoritmo recebe um grupo de sinais e, independentemente de outras informações externas tem a função fazer uma análise paraconsistente em seus valores subtraindo os efeitos causados pela contradição, de forma a apresentar na saída, um único Grau de Evidência resultante real representativo do grupo.

Na aplicação do respectivo algoritmo, o processo de captura de informações de diversas fontes de dados inicia-se pela obtenção dos Graus de Evidências

favoráveis e desfavoráveis por meio de equações de normalização, formando os grupos que serão analisados por blocos de NAPs compondo o algoritmo. Dessa forma, o algoritmo captura os dados nas fontes e extrai os efeitos das contradições até que na análise final seja obtido um único valor de Grau de Evidência resultante que representará os diversos graus aplicados como evidências da proposição alvo da análise.

O algoritmo utilizado no processo de extração de efeitos da contradição é descrito a seguir:

- 1. Apresente os n valores de Graus de Evidência que compõe o grupo em estudo**

$$G_{\mu} = (\mu_A, \mu_B, \mu_C, \dots, \mu_n) \quad */Graus\ de\ Evid\ancia\ 0 \leq \mu \leq 1 */$$

- 2. Selecione o maior valor entre os Graus de Evidência do grupo em estudo**

$$\mu_{maxA} = \max(\mu_A, \mu_B, \mu_C, \dots, \mu_n)$$

- 3. Considere o maior valor entre os Graus de Evidência do grupo em estudo em Grau de Evidência favorável**

$$\mu_{maxA} = \mu$$

- 4. Selecione o menor valor entre os Graus de Evidência do grupo em estudo**

$$\mu_{minA} = \min(\mu_A, \mu_B, \mu_C, \dots, \mu_n)$$

- 5. Transforme o menor valor em Grau de Evidência desfavorável**

$$1 - \mu_{minA} = \lambda$$

- 6. Faça a análise paraconsistente entre os valores selecionados**

$$\mu_{R1} = \mu_{maxA} \diamond \mu_{minA} \quad */Utilizando\ de\ um\ NAP\ */$$

$$\mu_{R1} = \mu \diamond \lambda$$

- 7. Acrescente o valor obtido μ_{R1} no grupo em estudo, excluindo deste, os dois valores μ_{max} e μ_{min} selecionados anteriormente**

$$G_{\mu} = (\mu_A, \mu_B, \mu_C, \dots, \mu_n, \mu_{R1}) - (\mu_{maxA}, \mu_{minA})$$

8. Retorne ao item 2 até que o Grupo em estudo tenha um único elemento considerado o valor resultante das análises

$$G_{\mu} = (\mu_{ER})$$

2. METODOLOGIA

Neste capítulo são apresentados os métodos aplicados nesta pesquisa com destaque para o detalhamento das inferências realizadas na base de dados de uma subestação de energia elétrica. Primeiramente é mostrado o processo de limpeza dos dados e em seguida, os tratamentos e normalização dos dados coletados para a ação dos algoritmos da LPA2v.

2.1. PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

Para o desenvolvimento da pesquisa foi considerada a arquitetura de uma subestação típica de distribuição de energia elétrica (Figura 11) para servir como fonte de dados secundária.

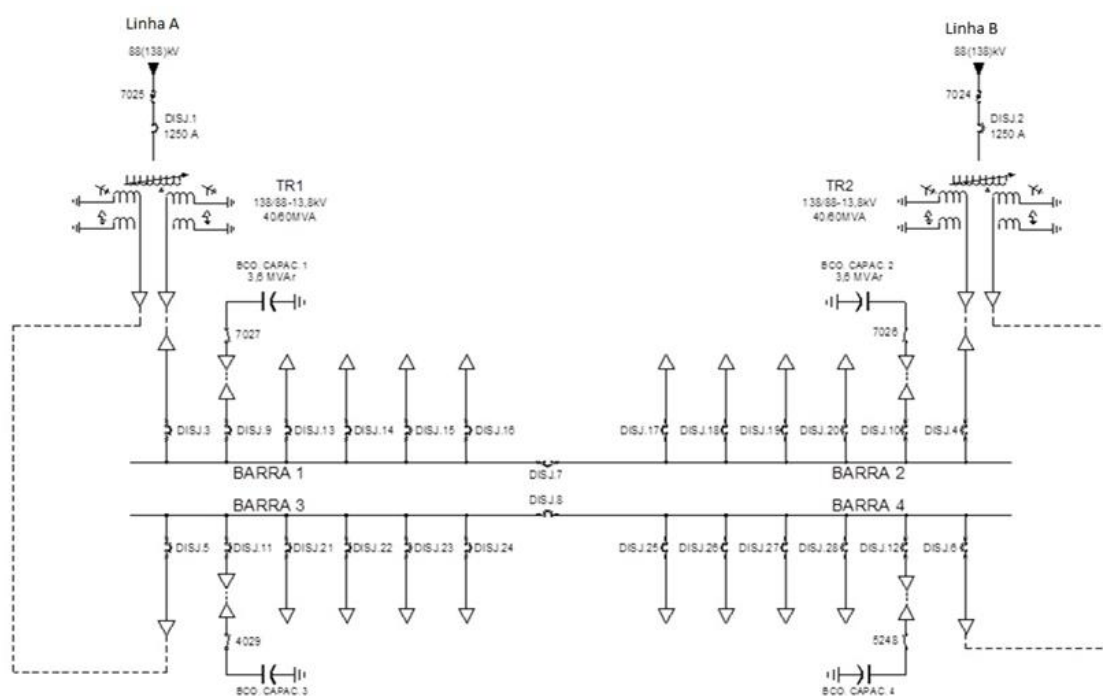


Figura 11 - Esquemático de um tipo da Subestação em estudo.

Fonte: Da Silva Filho et al. (2012).

No primeiro passo da utilização da técnica de KDD, foi realizada a seleção (fase em que os dados relevantes para a análise são recuperados) de uma base de dados de área de trabalho intitulada Microsoft™ Access, contendo as leituras realizadas por um sistema SCADA dentro de um período de 30 (trinta) dias.

Sistemas de Supervisão e Aquisição de Dados, ou abreviadamente SCADA (*Supervisory Control and Data Acquisition*) também chamado de *software* supervisor, são sistemas que utilizam *software* para monitorar e supervisionar as variáveis e os dispositivos de sistemas de controle conectados através de controladores (*drivers*) específicos. Estes sistemas podem assumir topologia mono-posto (na qual existe somente um computador, e este, é responsável por fazer a supervisão e controle do sistema), cliente-servidor (um computador tem a função de prover serviços, chamado de servidor, enquanto o outro se utiliza destes serviços, também chamado de cliente) ou múltiplos servidores-clientes (um único servidor de dados, mas com múltiplos clientes que podem se conectar ao mesmo servidor de dados e realizar sua supervisão).

Foram selecionados os seguintes atributos desta base de dados:

ID_SUBSTN: ID da subestação;
ID_DEVTYP: Tipo do equipamento;
ID_DEVIDE: ID do equipamento;
ID_ANALOG: ID do analógico;
DIS_ANALOG: Valor da medida;
FLDTIME_ANALOG: Instante de tempo da medida.

As leituras armazenadas na base de dados foram realizadas durante 30 dias de um determinado mês, durante intervalos regulares de 5 minutos, de 00h00min às 23h55min para cada um dos equipamentos alvo da leitura do sistema SCADA, totalizando 1.031.281 registros (Figura 12).

ID_MEDIDA	ID_SUBSTN	ID_DEV	ID_DEVICE	NAME_DEVI	ID_ANALOG	DFLAGS_AN	DIS_ANALOG	FLDTIME_ANALOC
1	DDIA	0	0	0	0	0	0	1/1/2001 01:01:01
2	DDIA	TR	TR1_SEC_B1	AV	30	329.96	01/04/2008 00:00	
3	DDIA	TR	TR1_SEC_B1	AB	30	331.14	01/04/2008 00:00	
4	DDIA	TR	TR1_SEC_B1	AA	30	338.75	01/04/2008 00:00	
5	DDIA	TR	TR1_SEC_B1	KVV	30	14.03	01/04/2008 00:00	
6	DDIA	TR	TR1_SEC_B1	KVB	30	14.02	01/04/2008 00:00	
7	DDIA	TR	TR1_SEC_B1	KVA	30	13.91	01/04/2008 00:00	
8	DDIA	TR	TR1_SEC_B1	MW	30	7.33	01/04/2008 00:00	
9	DDIA	TR	TR1_SEC_B1	MQ	30	3.41	01/04/2008 00:00	
10	DDIA	TR	TR1_SEC_B1	MVA	30	8.09	01/04/2008 00:00	
11	DDIA	TR	TR1_SEC_B3	AV	30	614.21	01/04/2008 00:00	
12	DDIA	TR	TR1_SEC_B3	AB	30	606.01	01/04/2008 00:00	
13	DDIA	TR	TR1_SEC_B3	AA	30	624.76	01/04/2008 00:00	
14	DDIA	TR	TR1_SEC_B3	KVV	30	13.81	01/04/2008 00:00	
15	DDIA	TR	TR1_SEC_B3	KVB	30	13.85	01/04/2008 00:00	
16	DDIA	TR	TR1_SEC_B3	KVA	30	13.71	01/04/2008 00:00	
17	DDIA	TR	TR1_SEC_B3	MW	30	13.83	01/04/2008 00:00	
18	DDIA	TR	TR1_SEC_B3	MQ	30	4.95	01/04/2008 00:00	
19	DDIA	TR	TR1_SEC_B3	MVA	30	14.70	01/04/2008 00:00	
20	DDIA	TR	TR2_SEC_B2	AV	30	554.43	01/04/2008 00:00	
21	DDIA	TR	TR2_SEC_B2	AB	30	539.19	01/04/2008 00:00	
22	DDIA	TR	TR2_SEC_B2	AA	30	517.51	01/04/2008 00:00	

Figura 12 - Base de dados original em ACCESS da Subestação em estudo coletados pelo sistema SCADA.

No sentido de viabilizar o processamento das respectivas informações, foi construída uma base de dados operacional (Figura 13) com a ferramenta MicrosoftTM SQL Server 2012 e nesta, realizada a importação de todos os registros supracitados (1.031.281). Este processo de importação dos dados já considerou a etapa de pré-processamento e limpeza do processo de KDD, visto que foram realizadas transformações nos dados, como o formato da leitura dos dispositivos, permitindo assim, a realização de operações matemáticas em procedimentos de tratamentos de incertezas.

Inicialmente o trabalho de processamento se concentrou nos registros pertinentes à fase AV de leitura dos equipamentos, mais especificamente em 189.068 registros.

Tabela 1: Dispositivos da subestação cuja fase AV foi alvo de estudo.

DEVICES	Barramento
13, 14, 15 e 16	1
17, 18, 19 e 20	2
21, 22, 23 e 24	3
25, 26, 27 e 28	4
7 e 8	1, 2 e 3 e 4
LT1, LT2, LT3 e LT4	-

Esta fase permitiu ainda a correlação da numeração dos disjuntores cujos dados estavam registrados na base de dados com a numeração empregada no esquema da subestação, mas especificamente nos barramentos 1, 2, 3 e 4 dos transformadores secundários (veja a Figura 11).

ID_DEV	ID_DEVICE	ID_ANALOG	DIS_ANALOG	FLDTIME_ANALOG
CT100	CT100	AA	0.00	01/02/2008 00:00
CT	CT101	AN	0.00	01/02/2008 00:00
CT	CT101	AV	0.00	01/02/2008 00:00
CT	CT101	AB	0.00	01/02/2008 00:00
CT	CT101	AA	0.00	01/02/2008 00:00
CT	CT102	AN	25.50	01/02/2008 00:00
CT	CT102	AV	126.95	01/02/2008 00:00
CT	CT102	AB	148.92	01/02/2008 00:00
CT	CT102	AA	120.97	01/02/2008 00:00
CT	CT103	AN	0.00	01/02/2008 00:00
CT	CT103	AV	0.00	01/02/2008 00:00
CT	CT103	AB	0.00	01/02/2008 00:00
CT	CT103	AA	0.00	01/02/2008 00:00
CT	CT104	AN	40.69	01/02/2008 00:00
CT	CT104	AV	186.02	01/02/2008 00:00
CT	CT104	AB	208.00	01/02/2008 00:00
CT	CT104	AA	161.05	01/02/2008 00:00
CT	CT105	AN	43.26	01/02/2008 00:00
CT	CT105	AV	254.07	01/02/2008 00:00
CT	CT105	AB	277.98	01/02/2008 00:00
CT	CT105	AA	304.00	01/02/2008 00:00
CT	CT106	AN	15.58	01/02/2008 00:00

Figura 13 - Base de dados elaborada na ferramenta Microsoft SQL Server.

2.2. NORMALIZAÇÃO DOS DADOS COLETADOS

Para a determinação do grau de evidência da intensidade da corrente elétrica em cada disjuntor foram considerados os eventos de menor e maior valores medidos da intensidade de corrente ocorridos em qualquer um dos 4 barramentos no período alvo da análise. Após a conclusão desta etapa, os valores das medidas das leituras dos equipamentos foram alvo de transformação através de modelagem.

Os cálculos foram feitos com variação linear e diretamente proporcional à grandeza medida, resultando em um valor normalizado entre 0 e 1, demonstrando assim, o grau de evidência do valor do objeto do estudo.

Para o processo de normalização constrói-se um universo de discurso com os valores das grandezas e obtêm-se os Graus de Evidência a partir de uma equação da reta.

O Valor limite máximo do Universo de Discurso para o valor da Grandeza será: $VLim_{max}$. O valor limite mínimo do Universo de Discurso para o valor da Grandeza será: $VLim_{min}$.

A forma gráfica de uma grandeza que varia aumentando proporcionalmente e sendo relacionada com o Grau de Evidência Normalizado μ , será representada conforme a Figura 14.

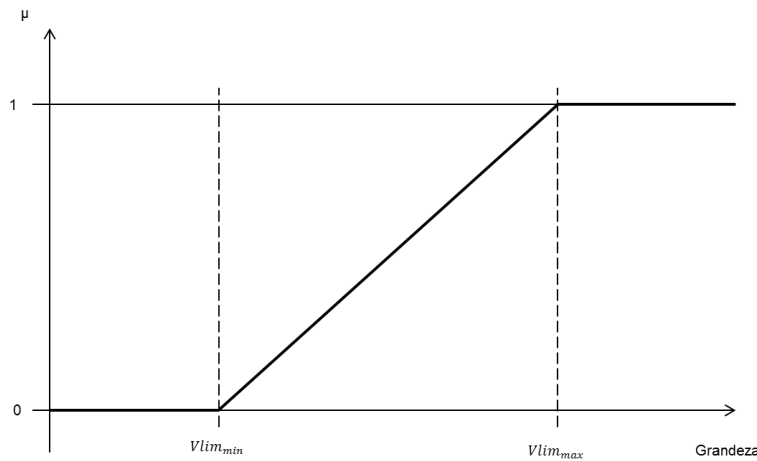


Figura 14 - Grandeza variando de modo diretamente proporcional relacionada com o Grau de Evidência Normalizado.

Fonte: O autor.

Dessa forma o Grau de Evidência para uma variação linear ascendente será encontrado através das equações de normalização a seguir:

$$\mu_{tendRisco} = \begin{cases} \frac{x - VL_{min}}{VL_{min} - VL_{max}} & \text{se } x \in [VL_{min}, VL_{max}] \\ 0 & \text{se } x < VL_{min} \\ 1 & \text{se } x > VL_{max} \end{cases}$$

Onde: x é o valor medido da grandeza.

A forma gráfica de uma grandeza que varia diminuindo proporcionalmente e sendo relacionada com o Grau de Evidência Normalizado μ , será representada conforme a Figura 15.

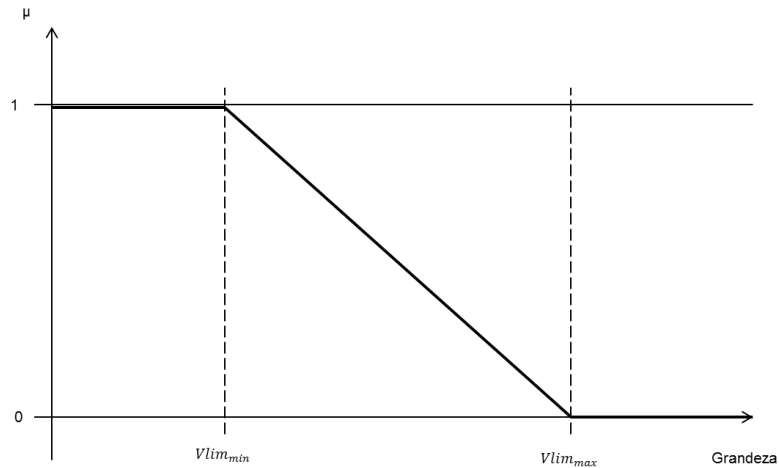


Figura 15 - Grandeza variando de modo inversamente proporcional relacionada com o Grau de Evidência Normalizado.

Fonte: O autor.

O Grau de Evidência para uma variação linear descendente será encontrado através das equações de normalização a seguir:

$$\mu_{tendRisco} = \begin{cases} \frac{VL_{min} - x}{VL_{max} - VL_{min}} & \text{se } x \in [VL_{min}, VL_{max}] \\ 1 & \text{se } x < VL_{min} \\ 0 & \text{se } x > VL_{max} \end{cases}$$

Onde: x é o valor medido da grandeza.

2.3. APLICAÇÃO DO ALGORITMO EXTRATOR DE EFEITOS DA CONTRADIÇÃO

De posse dos graus de evidência da intensidade da corrente elétrica de cada disjuntor foi aplicado o Algoritmo Extrator de Efeitos da Contradição para obtenção do grau de evidência da intensidade da corrente de cada barramento, no qual estavam instalados os respectivos disjuntores. O mesmo algoritmo foi aplicado para

identificação para obtenção do grau de evidência da intensidade da corrente conjunta dos quatro barramentos.

2.4. ELABORAÇÃO DE APLICAÇÃO PARA PROCESSAMENTO DOS DADOS

Objetivando a flexibilização do processamento dos dados importados na base de dados (operacional), da normalização dos dados coletados e da aplicação do Algoritmo Extrator de Efeitos da Contradição, foi elaborada uma ferramenta (cujo código fonte encontra-se no Apêndice 1) desenvolvida com a ferramenta *Microsoft Visual Studio Express 2012 for Windows Desktop*.

O emprego de tal ferramenta flexibilizou não apenas o processamento dos dados já armazenados na base de dados operacional, como também passou a permitir a importação de dados pertinentes a outros meses de leitura dos dispositivos da subestação, como também a seleção da fase (AA, AB, NA e AV) alvo da análise.

A partir da seleção do mês alvo da análise e da respectiva fase de corrente do dispositivo, a ferramenta possui a capacidade de calcular, a partir de inferências (regras de processamento) previamente implementadas na ferramenta:

- a) A situação do disjuntor 7 no momento em que o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 1;
- b) A situação do disjuntor 7 no momento em que o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 2;
- c) A situação do disjuntor 8 no momento em que o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 3;
- d) A situação do disjuntor 8 no momento em que o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 4;
- e) O maior tempo crítico no período nos barramentos 1, 2, 3 e 4;
- f) O valor resultante da aplicação, para os graus de evidência de cada disjuntor dos barramentos 1, 2, 3 e 4, do Algoritmo Extrator de Efeitos da Contradição;
- g) O valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência identificados no item “f” acima.

Uma melhor compreensão de todas as etapas do trabalho realizado de mineração de dados com a aplicação da Lógica Paraconsistente Anotada pode ser verificada com a Figura 16.

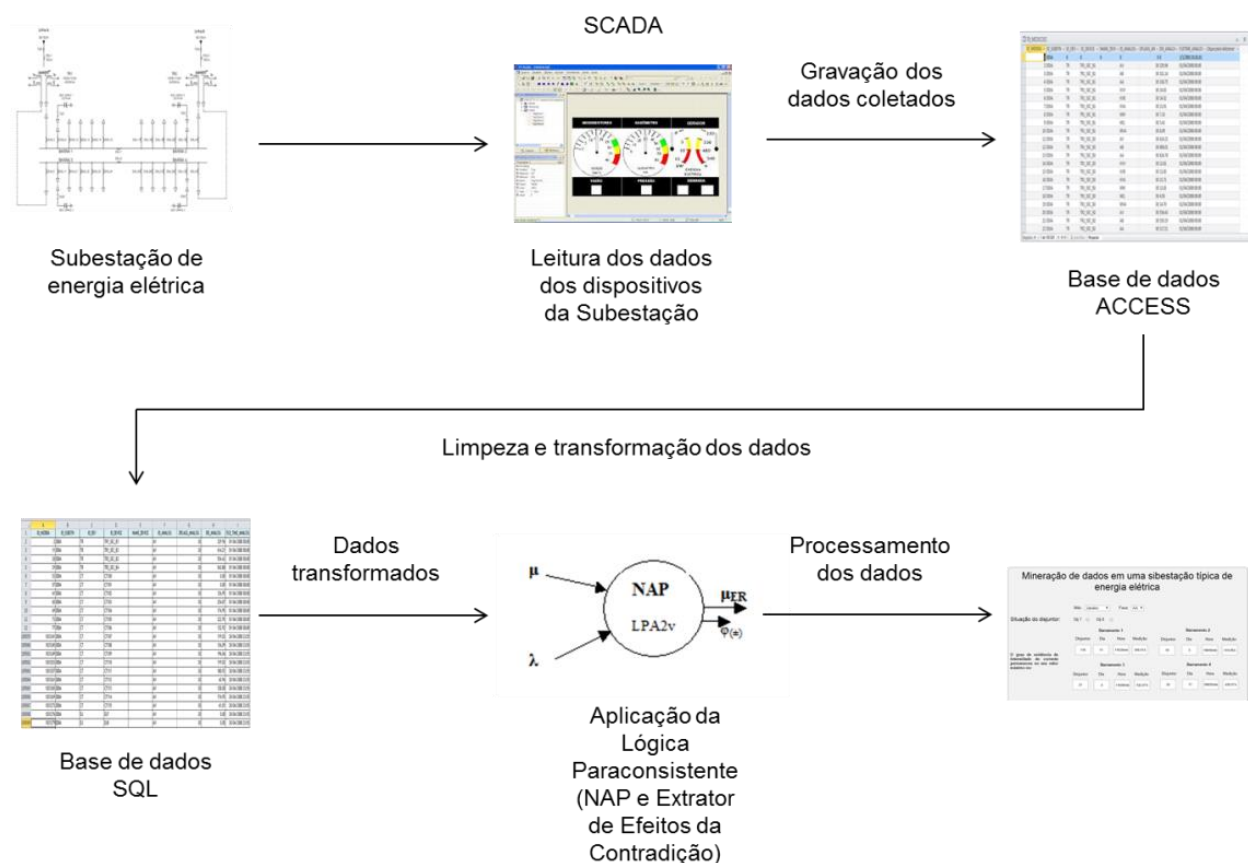


Figura 16 - Etapas do processo de mineração de dados da Subestação de Energia Elétrica com o emprego da Lógica Paraconsistente Anotada.

3. RESULTADOS E DISCUSSÕES

Na fase de mineração de dados do processo de KDD, foram considerados os valores de corrente, na fase AV, dos disjuntores das barras 1, 2, 3 e 4, dos disjuntores seccionadores de interligação das barras 1 e 2, dos disjuntores seccionadores de interligação das barras 3 e 4 dos disjuntores de alimentação dos transformadores de entrada LT1 e LT2 (Veja a Figura 11).

Esta segregação objetivou uma análise crítica do comportamento de cada uma dos disjuntores destes dispositivos, nas medições diárias realizadas pela ferramenta SCADA, no que concerne aos valores máximos e mínimos registrados.

Os gráficos apresentados nas Figuras 17, 18, 19 e 20 apresentam os valores mínimos e máximos de corrente identificados para cada um destes disjuntores dos barramentos 1 a 4, em cada dia do mês alvo da avaliação. A linha azul dos respectivos gráficos representam os valores mínimos da normalização enquanto que a linha vermelha especifica os valores máximos normalizados para os valores diários identificados de corrente elétrica.

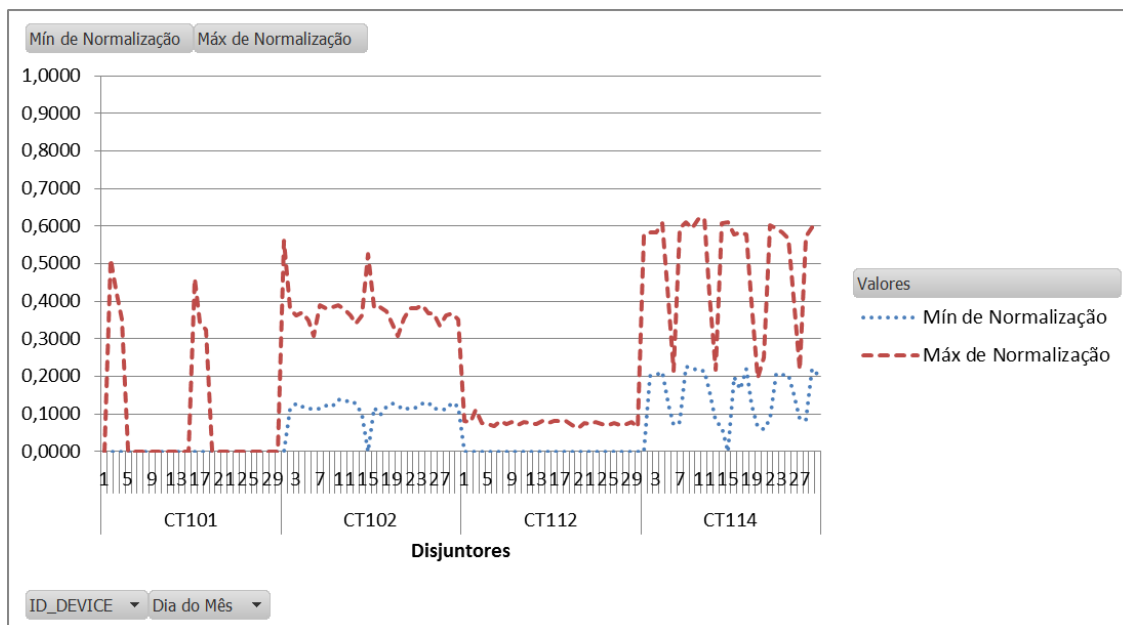


Figura 17 - Gráfico de normalização da corrente elétrica na fase AV da Barra 1 de disjuntores da alimentação das cargas.

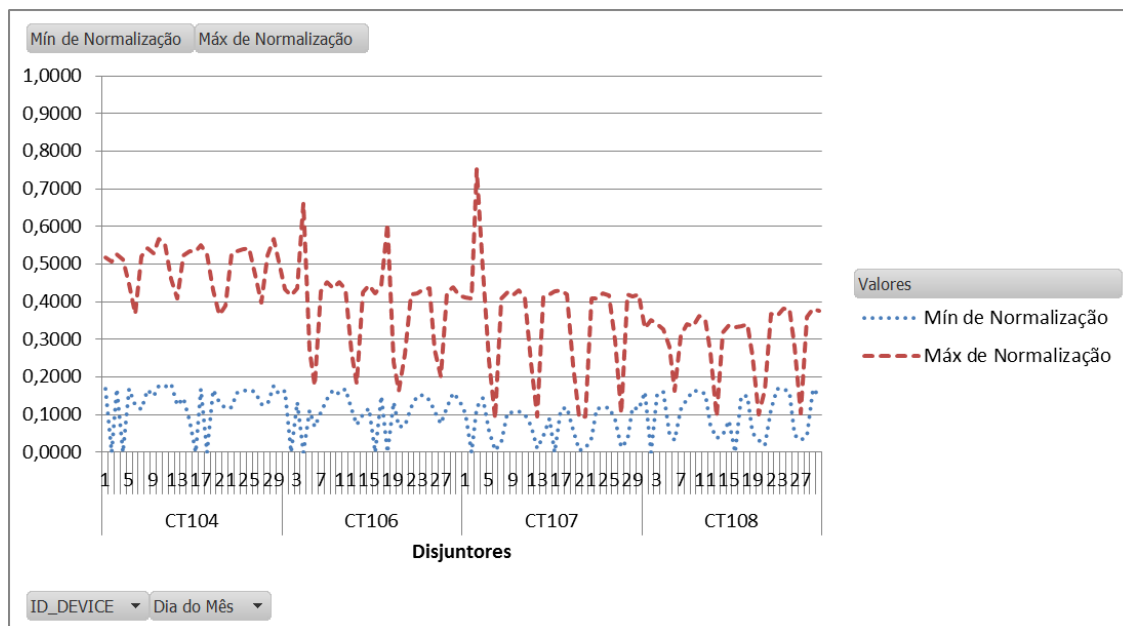


Figura 18 - Gráfico de normalização da corrente elétrica na fase AV da Barra 2 de disjuntores da alimentação das cargas.

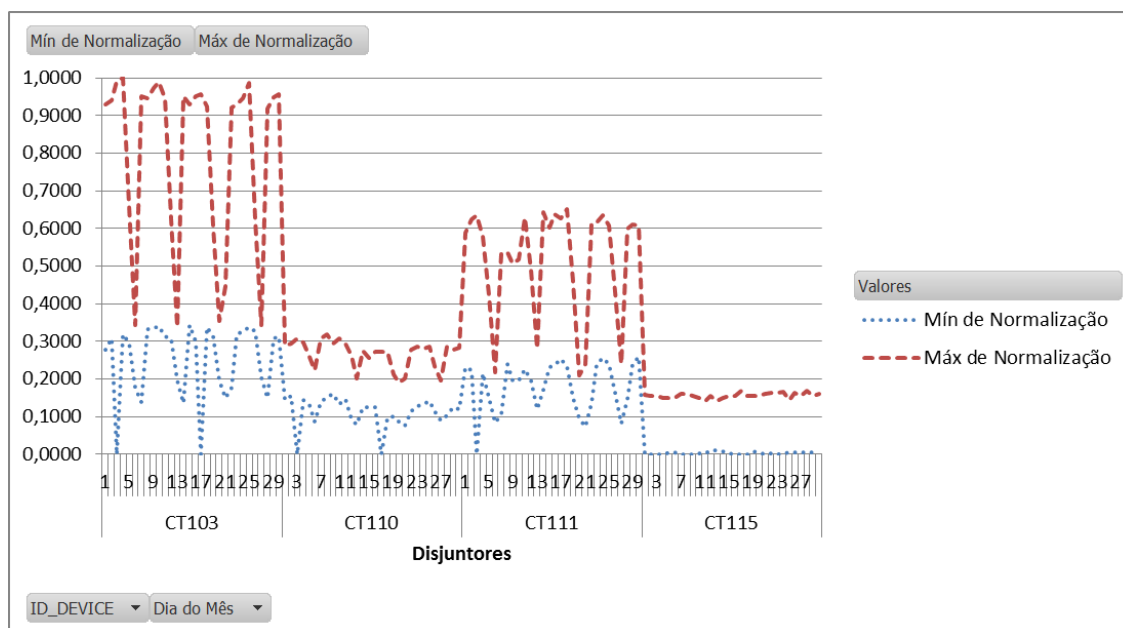


Figura 19 - Gráfico de normalização da corrente elétrica na fase AV da Barra 3 de disjuntores da alimentação das cargas.

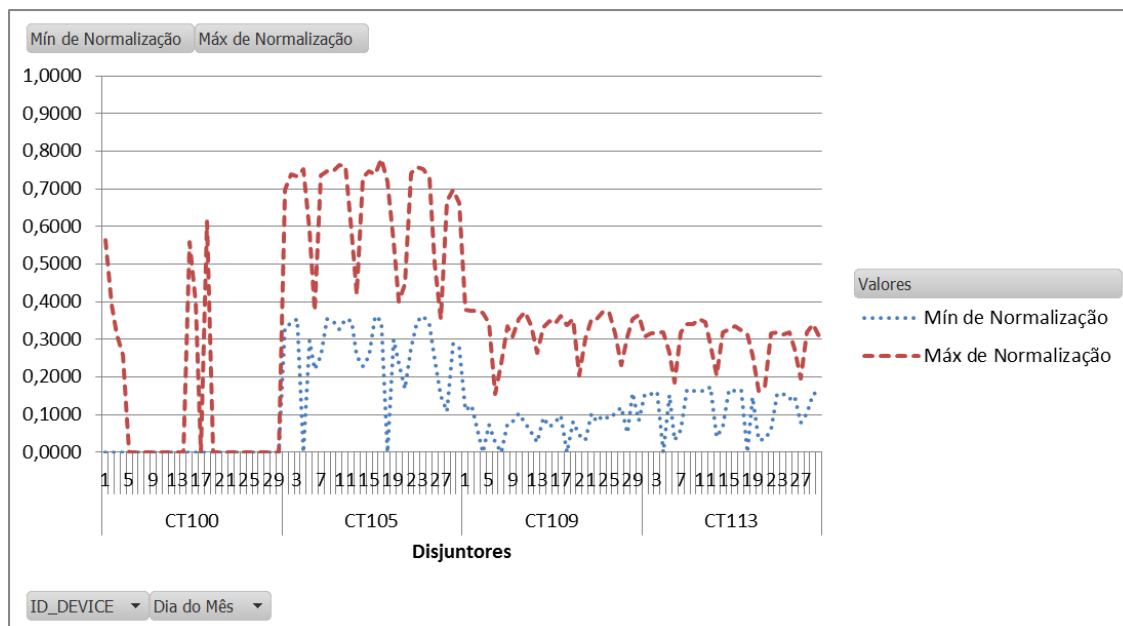


Figura 20 - Gráfico de normalização da corrente elétrica na fase AV da Barra 4 de disjuntores da alimentação das cargas.

Os gráficos apresentados nas Figuras 21, 22, 23 e 24 apresentam os valores mínimos e máximos de corrente identificados para cada um destes transformadores LT1, LT2, LT3 e LT4, em cada dia do mês alvo da avaliação. A linha azul dos respectivos gráficos representam os valores mínimos da normalização enquanto que a linha vermelha especifica os valores máximos normalizados para os valores diários identificados de corrente elétrica.

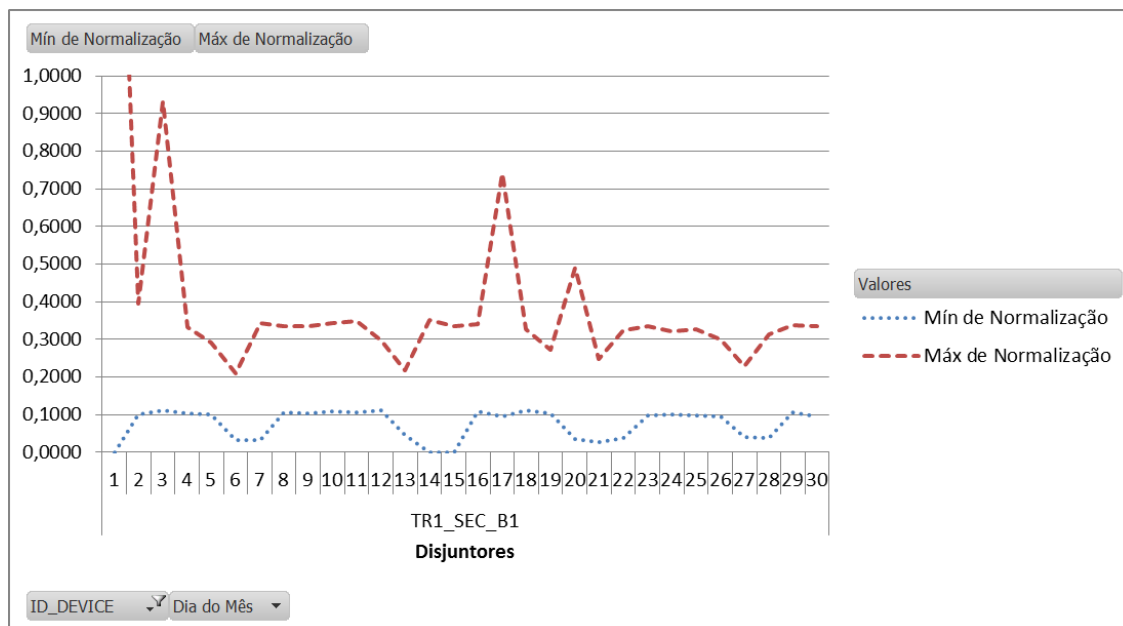


Figura 21 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT1.

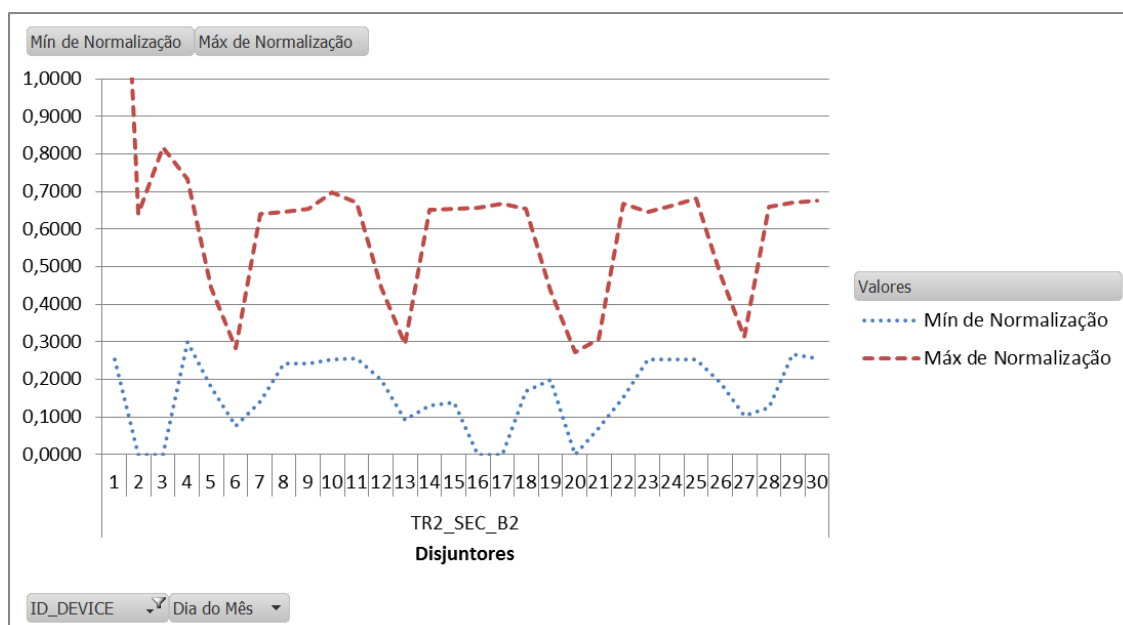


Figura 22 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT2.

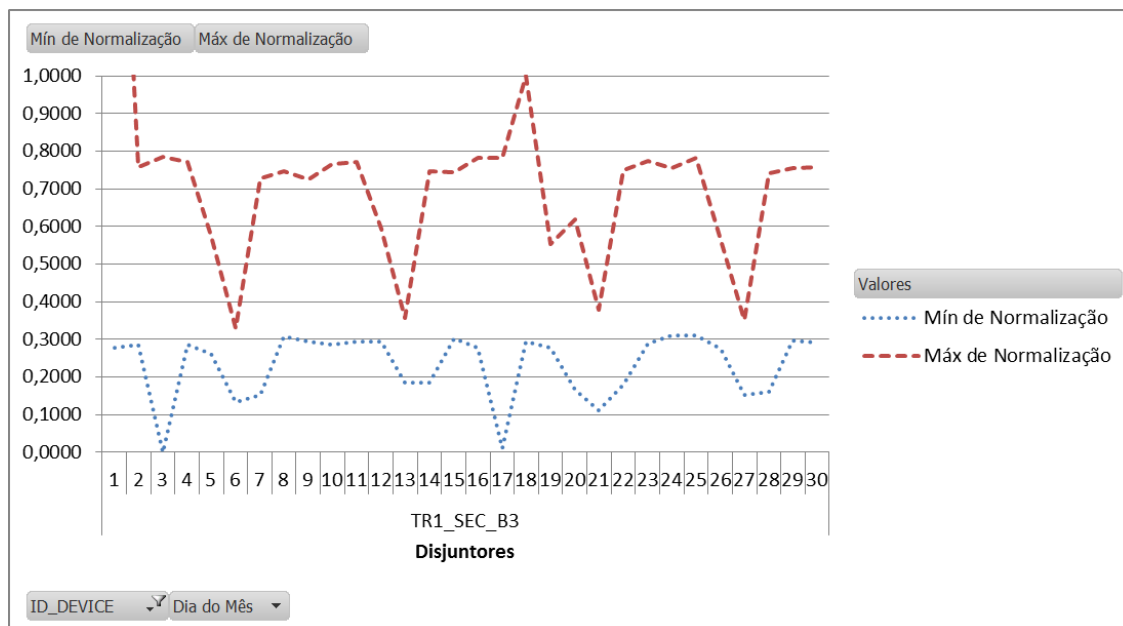


Figura 23 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT3.

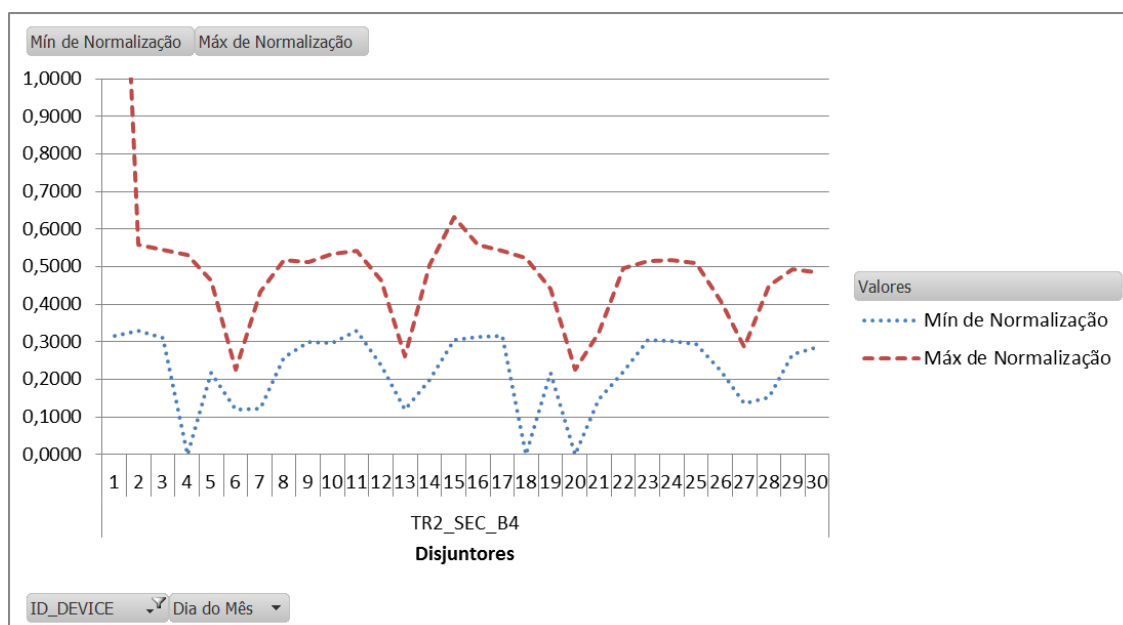


Figura 24 - Gráfico de normalização da corrente elétrica na fase AV no Transformador LT4.

Após esta análise e, considerando que os registros obtidos a partir da leitura do sistema SCADA, foram devidamente importados para a base de dados SQL, foi possível realizar o processamento destes com a ferramenta construída com o objetivo de aplicar a normalização dos dados coletados e aplicar o algoritmo Extrator de Efeitos da Contradição. O resultado deste processamento é apresentado na Figura 25.

Mineração de dados em uma subestação típica de energia elétrica

Mês: Fase:

Situação do Disjuntor:

Disj 7 / Barra 1: Disj 7 / Barra 2:

Disj 8 / Barra 3: Disj 8 / Barra 4:

O grau de evidência de intensidade de corrente permaneceu no seu valor máximo no:

Barramento 1		Barramento 2	
Disjuntor	Dia	Disjuntor	Dia
16	11	19	3
Hora	Medição	Hora	Medição
11h19min	349,01A	16h5min	414,95A

Barramento 3		Barramento 4	
Disjuntor	Dia	Disjuntor	Dia
21	4	26	17
Hora	Medição	Hora	Medição
11h0min	542,07A	9h55min	430,07A

O Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores é:

Barramento 1	Barramento 2	Barramento 3	Barramento 4	Barramento 1,2,3,4
0,4405	0,5315	0,7015	0,5525	0,5516

Figura 25 - Ferramenta desenvolvida com o objetivo de mensurar o processamento dos dados obtidos por medições nos disjuntores com a aplicação da Lógica Paraconsistente.

A aplicação do algoritmo Extrator de Efeitos da Contradição, baseado na Lógica Paraconsistente Anotada com anotação de dois valores, permitiu a extração dos efeitos da contradição dos graus de evidências favoráveis e desfavoráveis (que foram alvo do processo de normalização) dos sinais de intensidade de corrente dos disjuntores alvo do processamento.

3.1. DISCUSSÃO

A partir de inferências realizadas no processamento dos dados da corrente elétrica na fase AV dos disjuntores da subestação, foi possível a identificação de valores para determinados questionamentos conforme exposto na Tabela 2.

Conforme citado anteriormente, os gráficos apresentados nas Figuras 17, 18, 19 e 20 apresentam os valores mínimos e máximos de corrente identificados para cada um destes disjuntores dos barramentos 1 a 4, em cada dia do mês alvo da avaliação.

Com a utilização da ferramenta desenvolvida e que aplica os conceitos da Lógica Paraconsistente Anotada foi possível a realização de inferências sobre os dados coletados pelo sistema supervisório SCADA, operacionalizado pela subestação de energia elétrica alvo da pesquisa, permitindo a identificação dos graus de evidência de intensidade de corrente dos disjuntores instalados nos barramentos 1, 2, 3 e 4, assim como os valores máximos destes ao longo do período selecionado.

Tais inferências permitiram ainda a identificação da situação dos disjuntores seccionadores de interligação das barras 1 e 2, ou seja, a situação do disjuntor 7 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo nos barramento 1 e 2 e situação do disjuntor 8 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo nos barramento 3 e 4.

A aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores das barras 1, 2, 3 e 4, culminando com o cálculo do valor resultante desta análise.

As inferências supracitadas estão dispostas na Tabela 2, cujas ocorrências foram correlacionadas com as ocorrências identificadas nos gráficos citados acima.

Tabela 2: Resultado de inferências realizadas durante o processamento dos dados de corrente na fase AV dos disjuntores da subestação.

Questionamento	Resultado
Situação do disjuntor 7 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 1.	CT 114 – dia 11, às 11h20min – medição: 349,01A. O disjuntor 7 estava com a seguinte corrente: 850,40A.
Situação do disjuntor 7 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 2.	CT 107 – dia 3, às 16h05min – medição: 414,95A. O disjuntor 7 estava com corrente: 840,44A.
Situação do disjuntor 8 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 3.	CT 103 – dia 4, às 11h00min – medição: 542,07A. O disjuntor 8 não apresentava corrente elétrica.
Situação do disjuntor 8 quando o grau de evidência de intensidade de corrente permaneceu no seu valor máximo no barramento 4.	CT 105 – dia 17, às 09h55min – medição: 430,07A. O disjuntor 8 não apresentava corrente elétrica.
Número de vezes que o grau de evidência de intensidade de corrente permaneceu durante n minutos no seu valor máximo.	O valor máximo do grau de evidência de intensidade foi observado em uma única leitura em cada barramento.
Barramentos em que ocorreu o	Maior grau de evidência da

maior tempo crítico no período (maior grau de intensidade de corrente).	intensidade de corrente no barramento 1: 349,01A (dia 11 às 11h20min); Maior grau de evidência da intensidade de corrente no barramento 2: 414,95A (dia 03 às 16h05min); Maior grau de evidência da intensidade de corrente no barramento 3: 542,07A (dia 04 às 11h00min); Maior grau de evidência da intensidade de corrente no barramento 4: 430,07A (dia 17 às 09h55min)
Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores do barramento 1.	0,3511.
Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores do barramento 2.	0,3815
Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores do barramento 3.	0,5424

Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores do barramento 4.	0,4202
Valor resultante da aplicação do Algoritmo Extrator de Efeitos da Contradição sobre os graus de evidência dos disjuntores dos barramentos 1, 2, 3 e 4.	0,4110.

4. CONCLUSÕES

Considerando que o volume de informações relacionadas ao controle das redes de distribuição e de muitas outras fontes fez aumentar significativamente a dimensão das bases de dados das empresas de energia, a utilização adequada dessas informações se tornou fundamental na busca pelo aumento da eficiência na gestão operacional. Desse modo, ficou demonstrado neste trabalho que o processo de KDD e a mineração de dados aliados a técnicas fundamentadas em Lógica Paraconsistente podem ser úteis no processo de busca deste conhecimento, fornecendo padrões e correlações relevantes que podem apoiar o processo decisório da organização. Os resultados indicam que é possível estabelecer processos de tratamento e normalização de dados originários de sistemas de armazenagem de grandes quantidades de informações como os utilizados no SCADA de uma subestação de redes de distribuição de energia elétrica. Criam-se assim possibilidades de se utilizar os dados normalizados para a inserção de processos de KDD funcionando com outros que utilizam lógicas não-clássicas, capazes de gerar apoio a tomadas de decisão nesta área específica de energia elétrica, que é de primordial importância para o País.

Paralelamente, observou-se que a eventual ocorrência de dados inconsistentes pode estar relacionada a diversos motivos, dentre os quais, os ruídos nos dados coletados na subestação de energia elétrica ou mesmo erros decorrentes da coleta dos dados por parte do sistema supervisório SCADA.

Quanto à qualidade dos dados utilizados, o respectivo trabalho foi alicerçado na utilização de técnicas de gerenciamento e tratamento de incertezas com a Lógica Paraconsistente e, desse modo se permitiu um raciocínio mais adequado nas situações em que os dados ou resultados obtidos nem sempre estavam completos ou apresentavam algum tipo de inconsistência.

Verificamos que a utilização da Lógica Paraconsistente Anotada (LPA) no desenvolvimento do processo de mineração de dados permitiu a vinculação, aos respectivos dados coletados da subestação de energia, de fatores evidenciais de crença ou descrença, quantificando assim, o grau de importância deste em relação aos subconjuntos. Desta forma, a contribuição deste trabalho incide em pelo menos duas áreas de interesse da comunidade científica: mineração de dados e aplicações de técnicas de tratamento de incertezas.

A ferramenta computacional estruturado nos algoritmos paraconsistentes permitiu consultas e ofereceu respostas baseadas nos conceitos fundamentais da LPA o que indica ser esta uma relevante contribuição para novas pesquisas que possam ser desenvolvidas nesta área. Portanto, verificou-se neste trabalho que os resultados obtidos nesta primeira versão poderão ser utilizados como referências para as elaborações de estruturas complexas de processamento de dados baseados em técnicas de *Data Mining* que funcionem de forma híbrida com técnicas de tratamento de incertezas fundamentadas em Lógica Paraconsistente.

Este projeto de pesquisa acadêmica poderá ser fruto de trabalhos complementares futuros envolvendo técnicas híbridas de mineração de dados através da aplicação de algoritmos baseados em inteligência artificial baseados em Algoritmos Genéticos, Lógica Bayeasiana, Lógica Fuzzy e Redes Neurais Artificiais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABE, J. M. **Fundamentos da Lógica Anotada**. 1992. Tese (Doutorado). Universidade de São Paulo, São Paulo.
- AGRAWAL, R; SPIKANT, R. **Fast Algorithms for Mining Association Rules**. Chile, 2004. Proceedings of the 20th VLDB Conference.
- AMARAL, F. C. N. **Data Mining Técnicas e Aplicações para o Marketing Direto**. São Paulo: Berkeley, 2011.
- ANEEL. **Apresentação Institucional: Audiência Pública da Comissão de Defesa do Consumidor**. Câmara dos Deputados, Brasília, 2005.
- BALTAZAR, A. C. S. **Qualidade da Energia no Contexto da Reestruturação do Setor Elétrico Brasileiro**. 2007. Dissertação (Mestrado em Energia) - Universidade de São Paulo, São Paulo.
- BISPO, C. A. F.; CAZARINI, E. W. **Avaliação qualitativa paraconsistente do processo de implantação de um sistema de gestão ambiental**. *Gestão e Produção*. v. 13, n. 1, p. 117-127, 2006.
- CALIL, L. A de A. et al. **Mineração de Dados e Pós Processamento em Padrões Descobertos**. UEPG Exact Earth Sci, Ponta Grossa, v 14, n 3. P. 207-215, 2008.
- CARVALHO, L. A. V. **Data Mining. A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Editora Ciência Moderna. Rio de Janeiro, 2005.
- CIGRÉ. **Data Mining Techniques And Applications in the Power Transmission Field**. Working Group D1.11, Task Force 04, 2006.
- CASTANHEIRA, L. G. **Aplicação de Técnicas de Mineração de Dados em Problemas de Classificadores de Padrões**. 2008. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais.
- DA COSTA, N. **Lógica Paraconsistente Aplicada**. São Paulo: Atlas, 1999.
- DA COSTA, N. C.; SUBRAHMANYAN, V. S.; VAGO, C. **The Paraconsistent Logic**. *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik*, vol. 37. Pp. 139-148; 1991.

- DA COSTA, N. C.; ABE, J. M.; SUBRAHMANYAN, V. S. **Remarks on annotated logic.** *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik*, vol. 37, pp. 561-570; 1991.
- DA SILVA FILHO, J. I. da S.; ABE J. M. **Fundamentos das Redes Neurais Artificiais Paraconsistentes.** 1ª. Edição. Editora Arte & Ciência - Wilipress. São Paulo. p. 27-63. 2001.
- DA SILVA FILHO, J. I. da S.; ABE J. M.; TORRES, G. L. **Inteligência Artificial com as Redes de Análises Paraconsistentes: Teoria e Aplicações.** 1ª. Edição. Editora LTC. Rio de Janeiro. 2008.
- DA SILVA FILHO, J. I. **Algoritmo Paraconsistente Extrator de Efeitos da Contradição - ParaExtr_{ctr}.** Artigo técnico, Revista Seleção Documental do GLPA, n. 15, ano 4. 2009.
- DA SILVA FILHO, J. I. da S. **Métodos de interpretação da Lógica Paraconsistente Anotada com anotação com dois valores LPA2v com construção de Algoritmo e implementação de Circuitos Eletrônicos.** 1999. Tese (Doutorado) - EPUSP, São Paulo.
- DOMINGUES, M. L. C. S. **Mineração de Dados Utilizando Aprendizado Não Supervisionado: um estudo de caso para bancos de dados da saúde.** 2003. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Rio Grande do Sul.
- FAYYAD, U. et al. **From Data Mining to Knowledge Discovery in Databases.** *AI Magazine*, California, v. 17, n. 3. P. 37-54, 1996a. The KDD Process for Extracting Useful Knowledge from Volumes of Data. In: *Communications of the ACM*, Nov. 1996b.
- FRAWLEY, W. J. et al. **Knowledge Discovery in Databases: An Overview.** *AI Magazine*, Califórnia, v. 13, n. 3. P. 57-70, 1992.
- GOLDSCHMIDT, R.; PASSOS, E. **Data Mining: um guia prático.** Rio de Janeiro: Elsevier, 2005.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques.** Morgan Kaufmann Publishers, San Francisco, CA, 2000.
- Debugging hospitals.** *Technology Review*, 2001.
- KLOSGEN, W. **Patterns for knowledge discovery in databases. Procedure of Machine Learning.** United Kingdom. P. 1-9, 1992.
- KREUZE, D. **Using Data Mining For**, 2001.

- LIMA, D. A. **Convergência Tarifária: Remédio Regulatório para o Livre Acesso**. ANEEL, 2005.
- MACEDO, W. E. *et al.* **Avaliação das Perdas Elétricas no Sistema de Distribuição da SAELPA**. In: Ciclo de Palestras de Distribuição de Energia Elétrica, 2. ed, Olinda, PE, 1993.
- MÁRIO, M. C. **Modelo de análise de variáveis craniométricas através de Redes Neurais Artificiais Paraconsistentes**. 2006. Tese (Doutorado em Ciências) – Faculdade de Medicina de São Paulo, São Paulo.
- MARTINS, G. B. **Aplicação da Lógica Paraconsistente Anotada Evidencial em Mercados Financeiros**. 2012. Dissertação (Mestrado em Economia) – Universidade Federal de Santa Catarina, Santa Catarina.
- MARTINS, H. G. **A Lógica Paraconsistente Anotada de Quatro Valores – LPA4v aplicada em Sistema de Raciocínio Baseado em Casos para o Restabelecimento de Subestações Elétricas**. 2003. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Itajubá, Minas Gerais.
- PETERMANN, R. J. . **Modelo de mineração de dados para classificação de clientes em telecomunicações**. 2006. Dissertação (Mestrado em Engenharia Elétrica) – Pontífica Universidade Católica do Rio Grande do Sul, Rio Grande do Sul.
- RISSINO, S. das D. **Metodologia de Avaliação da Relevância de Atributos em Grandes Bases de Dados Incompletas Utilizando Conjuntos Aproximados e Lógica Paraconsistente**. 2009. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Itajubá, Minas Gerais.
- RUSSELL, M. A. **Mining the Social Web**. California: O'Reilly, 2011.
- SCHENATZ, B. N. **Utilização de data mining em um sistema de informação gerencial para o diagnóstico da formação de professores da graduação**. 2005. 102 f. Dissertação (Mestrado em Engenharia) – Universidade Federal de Santa Catarina. Florianópolis.
- SUBRAHMANIAN, V. S. **On the semantics of quantitative Logic programs**. Proc. 4th. IEEE Symposium on Logic Programming, Computer Society press, Washington D.C., 1987.
- TAN, P.; STEINBACH, M. e KUMAR, V. **Introdução ao Datamining: Mineração de Dados**. Rio de Janeiro: Ciência Moderna, 2009.
- UYSAL, I.; GUVENIR, H. A. **Instance-Based Regression by Partitioning Feature Projections**. Turkey, 2004. Applied Intelligence.

- WATSON, R. T. **Data Management: Banco de dados e Organizações**. 3^a. Edição. Editora LTC. Rio de Janeiro. p. 100-155. 2004.
- WILLIAMS, G. J.; HUANG, Zhexue. Modelling. **Modelling the KDD Process**. Data Mining Portfolio, Australia, 1996.

APÊNDICE A: Implementação do Algoritmo Extrator de Efeitos da Contradição

```

public string GetDis7Bar1Normalizacao(string fase, string month)
{
    double GrauEvidenciaFavoravel = 0, GrauEvidenciaDesfavoravel = 0,
    GrauCerteza = 0, GrauContradicao = 0, Distancia = 0, GrauCertezaReal = 0,
    GrauEvidenciaResultanteReal = 0;
    List<decimal> listNormalizacao = new List<decimal>();
    List<decimal> auxListNormalizacao = new List<decimal>();
    baseDal = new Base();
    sql =
    "TB_MEDICOES_DISJ7_BAR1_SELECT_NORMALIZACAO_BY_MONTH_FASE";//
    Fazer procedure

    try
    {
        baseDal.conectar();
        sqlCommand = new SqlCommand(sql, cnn);
        sqlCommand.CommandType = CommandType.StoredProcedure;
        sqlCommand.Parameters.Add("@FASE", SqlDbType.NVarChar,
255).Value = fase;
        sqlCommand.Parameters.Add("@MONTH", SqlDbType.Int).Value =
Convert.ToInt32(month);
        sqlDataReader = sqlCommand.ExecuteReader();
        if (sqlDataReader.HasRows)
        {
            while (sqlDataReader.Read())
            {

listNormalizacao.Add(Convert.ToDecimal(sqlDataReader["NORMALIZACAO"]));
            }
        }

        baseDal.desconectar();
    }
    catch (Exception e)
    {

        Console.WriteLine("Erro: " + e);
    }

    decimal[] values = listNormalizacao.ToArray<decimal>();

    while (values.Length > 0)
    {
        //var max2 = listNormalizacao.Max(x => x);
        //var min2 = listNormalizacao.Min(x => x);
        var max = values.Max();
        var min = values.Min();
    }
}

```

```

        GrauEvidenciaFavoravel = Convert.ToDouble(max);
        GrauEvidenciaDesfavoravel = (1 - Convert.ToDouble(min));
        GrauCerteza = GrauEvidenciaFavoravel - GrauEvidenciaDesfavoravel;
        GrauContradicao = (GrauEvidenciaFavoravel +
        GrauEvidenciaDesfavoravel) - 1;
        Distancia = Math.Sqrt(Math.Pow(1 - Math.Abs(GrauCerteza), 2) +
        Math.Pow(GrauContradicao, 2));

        GrauCertezaReal = 0;
        if (GrauCerteza > 0)
            GrauCertezaReal = (1 - Distancia);
        else
            if (GrauCerteza < 0)
                GrauCertezaReal = (Distancia - 1);

        GrauEvidenciaResultanteReal = (GrauCertezaReal + 1) / 2;

    auxListNormalizacao.Add(Convert.ToDecimal(GrauEvidenciaResultanteReal));
    //auxValues = new double[GrauEvidenciaResultanteReal];
    values = values.Where(val => val != max).ToArray();
    values = values.Where(val => val != min).ToArray();
    //listNormalizacao.Remove(max);
    //listNormalizacao.Remove(min);
}

auxListNormalizacao.OrderBy(x => x);

return auxListNormalizacao.Max(x => x).ToString("N4");
}

```